DOCUMENT RESUME

ED 224 689 SE 039 718

AUTHOR	Ross, Steven M.
TITLE	Introductory Statistics. A Conceptual Approach. Modified Preliminary Edition.
REPORT NO	ISBN-0-8134-2248-5
PUB DATE	82
NOTE	403p.
AVAILABLE FROM	Interstate Printers and Publishers, Inc., 19-27 North Jackson Street, Danville, IL 61832 ($14.50, less educational discounts).
PUB TYPE	Guides - Classroom Use - Materials (For Learner) (051)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*College Mathematics; Higher Education; *Individualized Instruction; Instructional Materials; Mathematical Applications; *Mathematical Concepts; Mathematics Instruction; *Statistics; Supplementary Reading Materials; Textbooks

ABSTRACT

This document deals with basic statistical concepts and operations used in the social sciences. The book was written under the philosophy that students enter statistics courses with a variety of aptitudes and experiences, and that traditional teaching approaches can do little to accommodate individual needs. It sets up an individualized course, and is designed to support teaching strategies where pupils do much of the learning on their own. The material is divided into 11 units, and the first part of each unit informs pupils of the general content area and how it relates to other topics. Each unit lists specific learning objectives, and provides instruction which includes a practice test for each stated objective. A major characteristic of the text is its emphasis on concepts rather than computations, with greater stress given on the why of procedures than is thought to be usual in most books. The presentation also attempts to reduce pupil anxiety often associated with statistics. This is done through an informal writing style that intersperses anecdotes, "conversational" prose, and humor. Students seem to like this approach, and have indicated that it seems to make reading easier and helps concepts stand out in memory. (MP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Steven M. Ross

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it
Minor changes have been made to improve reproduction quality
Points of view or opinions stated in this document do not necessarily represent official NIE position or policy

ED224689

INTRODUCTORY STATISTICS

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
Ronald L. McDaniel
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SE037118

A Conceptual Approach

ERIC

INTRODUCTORY

STATISTICS

Modified Preliminary Edition

ERIC

INTRODUCTORY STATISTICS

A Conceptual Approach

Steven M. Ross, Ph.D.

PROFESSOR
DEPARTMENT OF FOUNDATIONS OF EDUCATION
MEMPHIS STATE UNIVERSITY

Order from

THE INTERSTATE Printers & Publishers, Inc.

Danville, Illinois 61832-0594

ERIC Full Text Provided by ERIC

INTRODUCTORY STATISTICS:
A Conceptual Approach

Copyright © 1982, 1979, by The Interstate Printers & Publishers, Inc. All rights reserved. Printed in the United States of America.

ISBN 0-8134-2248-5

Preface

Statistics is hardly the most popular subject one can find in college curriculums; in fact, it scares many students and produces in them many common avoidance and fear responses. Students who are required to take a statistics course often put it off until their last semester, hoping perhaps, that the program will change, that the statistics professor will retire and not be replaced, or that somehow they will succeed in slipping by with the statistics requirement incomplete.

Such negative attitudes, however, seem less the fault of students than of statistics professors and the textbooks they employ. The traditional approach to teaching statistics suggests an image of an old professor (perhaps wearing a black robe and cap) haphazardly filling the blackboard with an uninterpretable array of numbers, equations, and strange symbols. He assumes that students will find themselves drooling with enthusiasm each time a new symbol or equation makes its way into their notebooks. Worst of all, the old professor ignores all the material he has presented on standardizing test scores and constructing "fair" grade distributions, with the result being a preponderance of F's and D's awarded as final grades. (Some professors anyway...)

On a more serious level, it is fairly easy to find faults with the lecture approach to teaching statistics, when it is the only strategy used. Lectures can be helpful, but it is difficult to construct them in such a way so that they will accommodate the needs of most students in attendance. The bright student may find the lecture too elementary; he understands the material, but questions the purpose of hearing it presented again--he finds it much more adaptive to "doze-off." The "slow" student may want to learn, but since he does not have full command of all required skills, finds the lecture almost impossible to understand; he needs more time, more rehearsals, a slower, more systematic presentation, but obviously the instructor is limited in what he can do to accommodate these individual needs.

The present textbook is intended to support teaching strategies where students can do much (or all) of their learning on their own, outside of class. An earlier version was developed and employed specifically for a course taught by the Keller Plan or Personalized System of Instruction (PSI). This version maintains the same spirit, although there is no reason why it could not be employed with virtually any teaching method. To support student self-learning, the following structure is used. The first part of each unit informs students of the general content area and how that content relates to the book's overall coverage. This overview is then followed by a listing of the specific performance objectives for the unit. The instructional portion of the unit then follows, which includes a "lesson" and a "practice posttest" for each of the stated objectives. Following the last unit posttest, a comprehensive "review test" is presented, which covers material relating to all objectives covered. If students are satisfied with their practice test performances they can move to the next major unit. If not satisfied, they can identify the specific objective(s) on which they are weak and then repeat the review and testing activities in the selected area(s). Aside from this structure, two other characteristics distinguish this book from most others being sold. One characteristic is that there is more textual elaboration and explanation than is usual--this is to allow the narrative to "take the place" of the lecturer (should a student-managed teaching approach--without conventional classes--be desired). However, should the book be used along with lecture classes, the increased explanation can only serve to reinforce what is said in class and to provide additional perspectives on the material. So many books seem long on mathematics, tables, and formulas, but short on textual discussion! This one is intended to be different.

A second characteristic is the use of an "informal" writing style, interspersed with anecdotes, conversational-type prose, and "attempted" humor. Thus, along the way, the reader will encounter such characters as "Bubbles the Whale," whose weighing is used to illustrate the concept of "real limits"; "Murray Binet," whose questionable attempt to develop a new IQ test is used to illustrate the need for caution in interpreting standard scores; and "Natasha Bobolinckskismith," for whom the agony of defeat in a gymnastics competition was used to illustrate something or other. It is hoped that these will not be offensive, but rather provide contexts that stand out in your mind and help you to understand and later mentally "reconstruct" the statistical concepts they convey.

Well, enough--perhaps too much--has been said by way of introduction. But before you start plunging into the material to follow, there are three suggestions that may be worthwhile to keep in mind. Here they are:

1. Hand calculators have become very popular (and inexpensive) in recent years. If you have been entertaining the thought of buying one, now would be a very good time to make the purchase. It will enable you to work the problems rapidly and accurately. If you cannot afford one, there is no need to worry since you can get through doing the problems by hand. But a calculator would be very advantageous, mostly in terms of saving time.

2. In solving problems, the general rule is to carry out your work to at least two (2) place decimal accuracy. Sometimes exceptions will be in order; use your judgment.

3. Do all practice problems. They are there for a purpose and you should take advantage of them.

Acknowledgements

The present textbook owes its style, structure, and much of its content to an earlier edition which, in turn, was made possible through helpful assistance of several individuals. First, the author owes a considerable debt of gratitude to Dr. Dennis Roberts, whose successful course at The Pennsylvania State University provided the idea and a procedural framework for the present approach.

The writing of this text was sponsored, in part, by a Seed Grant from the Center for Instructional Service and Research, Memphis State University. The author also wishes to thank Dr. Andrew Bush and Steve Baucum, who made very valuable contributions through their editing and critiquing of parts of the manuscript. Denise Oldham and June Powers provided assistance in proofreading the final version.

Very special thanks are due to Mrs. Sue Douglass, who handled the task of typing the manuscript largely through successfully deciphering the almost illegible scribblings by the author. Mrs. Douglass took a special interest in the project, assuming responsibilities and schedules that went far beyond the call of duty. Writing the book was substantially easier with her support.

The author is grateful to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., to Dr. Frank Yates, F.R.S., and to Longman Group Ltd., London, for permission to reprint portions of Tables III and IV from their book, Statistical Tables for Biological, Agricultural and Medical Research (6th Edition, 1974).

Table of Contents

ERIC

page

UNIT A

REVIEW OF BASIC MATH

Prerequisite Skills

When you turn the page, you will find what is called a "review" unit on basic arithmetic and algebraic operations. Although the present course attempts to emphasize mathematical skills *as little as possible*, there are some rather basic operations which you will need to perform as you move through the "eleven units." All of these basic operations are reviewed in the following section, and it is strongly recommended that you read through these and attempt to diagnose any weaknesses that you might have.

Procedure for Review

Proceed through the review unit, perhaps spending from 30 minutes to one hour on the entire section. Once you feel reasonably comfortable with the various operations covered, take the review test which appears at the end of the unit. Then check your answers in the answer section. If you score 90% or above, math will probably not be a problem for you, but still go over any questions that you missed, and try to identify what the errors involved. If, however, you score less than 80% you definitely need review: reread the review unit and think about stopping in for help. Do not begin working on Unit I until you have completely mastered the various math and algebraic operations that are covered in the practice test; if you begin before mastery is achieved, you will only be making it more difficult for yourself later on in the course.

Begin now...and good luck!

A.1 Addition

Rule A: When positive numbers are added together the answer has a positive sign.

Examples:
$$\begin{array}{r} 15 \\ (+)\ \underline{3} \\ 18 \end{array} \qquad \begin{array}{r} 13X \\ (+)\ \underline{6X} \\ 19X \end{array}$$

Rule B: When negative numbers are added together, the answer has a negative sign.

Examples:
$$\begin{array}{r} -1 \\ (+)\underline{-15} \\ -16 \end{array} \qquad \begin{array}{r} -13Z \\ (+)\underline{-6Z} \\ -19Z \end{array}$$

Rule C: When positive and negative numbers are added together the answer has the sign of the larger quantity.

Examples:
$$\begin{array}{r} 13 \\ (+)\underline{-12} \\ +1 \end{array} \qquad \begin{array}{r} -27 \\ (+)\ \underline{15} \\ -12 \end{array}$$

While the above rule (Rule C) for determining the proper sign is easy, a more basic problem for many students is performing the math operation. Here is a rule for that: (even though we are dealing with an addition problem) subtract the smaller number (whether it is the positive or the negative one) from the larger number. The result is your answer; all that remains is to provide the sign.

Example:
$$\begin{array}{r} -57 \\ (+)\ \underline{32} \\ -25 \end{array}$$

Why? The first rule states that you must use the sign of of the larger number; thus, the answer would be "negative." The second rule states that you should subtract the smaller from the larger (forgetting for a moment about signs); thus, the answer would be "25." Putting both rules together gives the final result, "-25."

A.2 Subtraction

Subtraction, for some people, is much more troublesome than addition, particularly when negative numbers are involved. If this applies to you, it will no longer, provided that you can learn the following rule:

Rule: Given the task of subtracting two numbers, reverse the sign of the number to be subtracted (i.e., the subtrahend) and treat the problem as if it were one of addition.

ERIC

Example:

$$\begin{array}{r} 5 \\ (-)\ \underline{3} \\ 2 \end{array}$$

(Note: this one probably seems very easy, but the important task for now is to learn the rule so that you will be able to apply it to more difficult examples.)

Why? Using the above rule, the subtrahend, 3, is converted to a "-3", and the problem becomes one of addition: 5 (+) -3 = <u>2</u> (remember addition rules given above).

Example:

$$\begin{array}{r} -10 \\ (-)\underline{+12} \\ ? \end{array}$$

For this one, apply the rule in the same manner: (a) Change the sign of subtrahend from a +12 to a -12; (b) then treat the problem as one of addition, -10 (+) -12 = ?

Ans. = -22

(If you do not understand *why*, check Rule "B" under "Addition.")

Example:

$$\begin{array}{r} +15 \\ (-)\underline{-22} \\ ? \end{array}$$

Change the sign of the subtrahend and add: +15 (+) +22 = ?

Ans. = +37

Example:

$$\begin{array}{r} -\ 8 \\ (-)\underline{-\ 6} \\ ? \end{array}$$

Change the sign of the subtrahend and add: -8 (+) +6 = ?

Ans. = -2 (see Rule "C" under "Addition" if you do not see how the answer was derived.)

A.3 Multiplication

Although multiplication can be handled easily by a calculator (or by hand if you are careful with calculations), what gives some people problems is the question of what sign to use in the answer. Here are two simple rules for determining signs with some examples which show how they are applied.

<u>Rule A</u>: If the two numbers being multiplied have the same sign, whether "+" or "-," the answer will have a "+" sign.

<u>Rule B</u>: If the two numbers being multiplied have opposite signs, the answer will have a "-" sign.

Example:

$$\begin{array}{r} 23 \\ (x)\ \underline{4} \\ ? \end{array}$$

Since the two numbers have the same sign, the answer will be positive. Do your multiplication and attach a "+" sign to your answer.

Ans. = +92

Example: $\begin{array}{r} -57 \\ (x)\underline{-\ 4} \\ ? \end{array}$ Since the two numbers have the same sign, do your multiplication and attach a <u>positive</u> ("+") sign to your answer.

Ans. = +228

Example: $\begin{array}{r} -36 \\ (x)\underline{\ 28} \\ ? \end{array}$ Since the two numbers have opposite signs, do your multiplication and attach a <u>negative</u> ("-") sign to your answer.

Ans. = -1008

Example: -27 (x) -3 (x) 4 (x) -10 = ?

There are a number of different systems people use for multiplying a string of numbers of the kind shown. Perhaps the easiest way to stay out of trouble is to multiply only two numbers at a time while applying the above rules to determine the sign of each successive answer. This is how it would be done:

Taking the first two numbers, -27 (x) - 3, we get +81 (signs were the same).

Then multiply the +81 by the next number in the string: +81 (x) +4 = +324 (signs were the same).

Then multiply the +324 by the final number, -10, to yield 324 (x) -10 = ?

Ans. = -3240

A.4 <u>Division</u>

Like multiplication, division can be easily handled by a calculator, but there is sometimes a problem in determining the appropriate sign of the answer. Fortunately, the rules you just learned for multiplication apply to division in the identical manner.

<u>Rule A</u>: If the two numbers have the same sign ("+" or "-"), the answer is "+."

Examples: 9 (÷) 3 = <u>3</u>; -12 (÷) -4 = <u>3</u>

Rule B: If the signs of the two numbers are opposite, the answer is "–."

Examples: 15 (÷) -5 = -3; -21 (÷) 3 = -7

In case you are to do division by hand, here are two more rules to keep in mind.

Rule C: Remember to divide variable labels, such as "X" or "Y," as part of the overall problem: X (÷) X = 1

Z^2 (÷) Z = Z (Note that as a check, Z x Z = Z^2)

Z^4 (÷) Z^2 = Z^2 (Note that Z^2 x Z^2 = Z^4)

Examples: 15X (÷) 3X = ?

Ans. = 5 (X's cancel out)

$39Y^5$ (÷) -3Y = ?

Ans. = $-13Y^4$ (note that answer is negative since signs were opposite; also Y^5 (÷) Y becomes Y^4)

If you're having trouble with problems involving powers (exponents), remember that exponents are added during multiplication and subtracted during division. Thus, a number to the second power times a number to the fifth power yields an answer expressed to the seventh power (2 + 5); or, a number to the eighth power divided by one to the third power yields an answer expressed to the fifth power (8 – 3).

Rule D: In dividing numbers involving decimals (without a calculator), make the divisor a whole number by moving its decimal an appropriate number of places to the right; then do the same operation for the dividend by moving its decimal and/or adding zeroes to equal the total places the divisor's decimal was moved.

Example: 15 (the dividend) ÷.5 (the divisor) = ?

First, make the divisor a whole number by moving its decimal one place to the right.

150 ÷ 5 =

The dividend has no decimal so 1 zero needs to be added, as shown.

Ans. = 30

ERIC

Example: $100.2 \div .505 = ?$

$100200 \div 505 = ?$

Ans. = 198.42 (rounded to 2 decimal places)

Here, a three-place movement is needed in the divisor. We must do the same for the dividend: moving its decimal gives 1 place; adding 2 zeroes gives the total, as shown.

Actually, such problems should be set-up in long division form. The last example will illustrate this procedure.

Example: $-13.1 \div 5.96 = ?$

(a) Set up as: $5.96 \overline{)\,-13.1}$

(b) Move decimal two places in the divisor. Equalize by moving the dividend's decimal the one possible place and adding 1 zero.

$596 \overline{)\,-1310}$

(c) Answer will have a negative sign because the two numbers are mixed. To avoid confusion, put proper sign in the answer space and remove the "-" from the dividend.

$$\begin{array}{r} - \\ 596 \overline{)\,1310} \end{array}$$

(d) Now work the problem.

$$\begin{array}{r} -\,2.197 \\ 596 \overline{)\,1310.000} \\ \underline{1192} \\ 1180 \\ \underline{596} \\ 5840 \\ \underline{5364} \\ 4760 \\ \underline{4172} \end{array}$$

Note: We have added three decimal places to the dividend so that our answer will have two-place accuracy.

Ans. = - 2.20

A.5 Order of Operations

In performing a sequence of operations, do your calculations in the following order of priority.

(1) Do work in parentheses
(2) Raise to powers and/or take roots
(3) Multiply and/or divide
(4) Add and/or subtract

Example:	$(5 - 3) \times 7 = ?$	First, do work in ( ).
	$(2) \times 7 = ?$	Then <u>multiply</u>.
	Ans. = 14	
Example:	$-14\ (-)\ 2 \times -2 + \sqrt{4} = ?$	First, take <u>square root</u>.
	$-14\ (-)\ 2 \times -2 + 2 = ?$	Next, <u>multiply</u> 2 by -2.
	$-14\ (-)\ -4 + 2 = ?$	Then, <u>add/subtract from left to right. Start with</u> -14 (-) -4.
	$-10 + 2 = ?$	
	Ans. = -8	
Example:	$(12 - 3) \times 4^2\ (+)\ -12 + \sqrt{16} = ?$	First, do work in ( ).
	$9 \times 4^2\ (+)\ -12 + \sqrt{16} = ?$	Then do <u>powers</u> and roots.
	$9 \times 16\ (+)\ -12 + 4 = ?$	Then <u>multiply and/or divide from left to right; multiplication</u> is first here.
	$144\ (+)\ -12 + 4 = ?$	
	$144\ (+)\ -3 = ?$	Then <u>add</u>.
	Ans. = 141	

A.6 <u>Fractions</u>

<u>Rule A</u>: For adding or subtracting, convert to common denominator, then reduce to lowest denominator.

Examples: $\frac{3}{12} + \frac{2}{4} = \frac{1}{4} + \frac{2}{4} = \frac{3}{4}$

$$\frac{5}{16} - \frac{1}{12} = \frac{15}{48} - \frac{4}{48} = \frac{11}{48}$$

$$\frac{2}{3} + \frac{2}{4} = \frac{8}{12} + \frac{6}{12} = \frac{14}{12} = \frac{7}{6}$$

<u>Rule B</u>: For multiplication, multiply numerator and denominators together; <u>then</u> reduce.

Examples: $\frac{1}{5} \times \frac{3}{8} = \frac{1\times3}{5\times8} = \frac{3}{40}$

$$\frac{12}{3} \cdot \frac{2}{8} \cdot \frac{X}{Y} = \frac{12\times2\times X}{3\times8\times Y} = \frac{24X}{24Y} = \frac{X}{Y}$$

Rule C: For dividing, invert the fraction in the divisor and then multiply as usual.

Examples: $\frac{1}{6} \div \frac{2}{3} = \frac{1}{6} \times \frac{3}{2} = \frac{3}{12} = \frac{1}{4}$

$$\frac{3}{8} \div \frac{1}{4} = \frac{3}{8} \times \frac{4}{1} = \frac{12}{8} = \frac{3}{2}$$

A.7 Proportions and Percentages

Rule A: A proportion is the decimal equivalent of a fraction. To convert a fraction into a proportion, simply divide its denominator (bottom number) into its numerator (top number).

Examples: $\frac{4}{5} = ?$ $\frac{7}{9} = ?$

```
     .80             .777
  5/4.000          9/7.000
    4 0              6 3
                       70
  Ans. = .80           63
                        70
                        63
```

Ans. = .78 (rounded off).

Rule B: A percentage is a proportion expressed on a scale of 100. To convert a fraction into a percentage, first change it to a proportion (Rule A); then move the decimal point two places to the right.

Examples: Convert the following proportions to percentages.

.80 = ?	.7986 = ?	1.86 = ?
Ans. = 80%	Ans. = 79.86%	Ans. = 186%

A.8 Solving for Unknowns

There is no one way to simplify an equation, and thus, if certain methods learned in previous math courses still work for you, continue to use them. If, however, those methods were never really well learned, you may want to try the conceptual approach illustrated below.

Rule A: In trying to solve for an unknown expression ("X" or "Y," etc.) in an equation, isolate the unknown by moving all known (i.e., numerical) values to the opposite side. The result will be: "*unknown = a certain value.*"

Examples: For $X - 5 = 8$ you want to move the "5" to the right side to isolate, and then solve for, X.

For $\frac{6X - 37}{10} = 197$ you want to isolate X by moving the "6," "37" and "10" to the right side. Once you accomplish this and simplify the right side, the result will be "X = certain value." At that point, the equation is solved!

<u>Rule B</u>: The equation must always be balanced: right side must equal left. Thus, whatever operation you perform on one side must be repeated exactly on the other.

Example: $X + 5 = 10$ As will be discussed below, you would want to isolate X by subtracting "5" from the left side. But if you did <u>only</u> this, the equation would "tip" (i.e., be imbalanced). Thus, "5" needs to be subtracted from the right side as well.

$X + 5 -(5) = 10 -(5)$

$X = 5$

<u>Rule C</u>: The way to remove a number from X's side of the equation is to apply it again to each side of the equation, but in the "*opposite*" way from which it is originally used.

1. Thus, to remove a value being <u>added</u> to X's side, we subtract that number from both sides of the equation.

Example: $X + 15 = 30 - 2$ First, we might as well simplify the right.

$X + 15 = 28$. We want to remove the 15, which is being added, so we subtract 15's. The 15's on the left cancel out, leaving only "X=."

$X + 15 -(15) = 28 - (15)$

$X = 28 - 15$

Ans. $X = 13$

2. To remove a value being <u>subtracted</u> from X's side, we add that number to both sides of the equation.

Example: $X - 27 = 15 + 3$ First, simplify the right.

$X - 27 = 5$ Then remove the -27 by adding 27's.

$X - 27 + (27) = 5 + (27)$ The 27's cancel out on the left.

$X = 5 + 27$

Ans. $X = 32$

3. To remove a value being <u>multiplied</u> by X's side, divide that value into both sides of the equation.

Example: $4X = 40$ Remove the 4 by dividing each side by 4.

$\frac{4X}{4} = \frac{40}{4}$ The 4's cancel out on the left.

$X = 10$

4. To remove a value being <u>divided</u> into X's side, multiply that value by both sides of the equation.

Example: $\frac{X}{2} = 11 - 1$ First simplify the right.

$\frac{X}{2} = 10$ Remove the 2 by multiplying each side by 2.

$\frac{2}{1} \times \frac{X}{2} = 10 \times 2$ The 2's cancel out on the left.

$X = 20$

5. When more than one value needs to be removed, the order in which you should proceed is:
   (a) simplify both sides as much as possible before trying to remove any numbers.

Example: $15X - 4 + 2 + 14 = 10 + 5 + 16$

$15X - 2 + 14 = 2 + 16$

$15X + 12 = 18$

Confused? See Sec. A.5 - Order of Operations for help.

(b) then start your "replacement" with the value on X's side that affects the "whole" expression rather than just one part.

Which value is it in the example, the "15" or the "12"?

Ans.: It is the "12" since it affects the "whole" 15X; the "15" only affects "X".

Thus, we remove the "12" first:

$$15X + 12 - (12) = 18 - (12)$$

$$15X = 6$$

(c) Having "removed" the primary value, then work with the remaining one.

$$15X = 6 \;:\; \frac{15X}{15} = \frac{6}{15} \;:\; X = 6/15$$

Ans: $X = 2/5$

Try these examples of complex equations:

$(X - 12) \div 5 = 10 - 2 \times 2$ — First simplify (the right only is possible here).

$(X - 12) \div 5 = 10 - 4$

$(X - 12) \div 5 = 6$ — Which number do we remove first on the left? The "5" !!!, as it affects both X and 12; the 12 only affects X. Remove by multiplying.

$\frac{X-12}{5} \times \frac{5}{1} = 6 \times 5$ — The 5's cancel out on the left.

$X-12 = 30$ — Then remove "12" through addition.

$X - 12 + (12) = 30 + (12)$ — The 12's cancel out on the left.

Ans.: $X = 42$

$\frac{Y \times 5 + 3}{2} + 1 = 3 \times 2 + 2$ — First simplify both sides.

$\frac{5Y + 3}{2} + 1 = 8$ — Which left-side value goes first? The "1" of course, as it affects the whole expression. The "5" affects Y only; the "3" affects "5Y" only; the "2" affects "5Y + 3" only. Thus, remove the "1" through subtraction.

$\frac{5Y + 3}{2} + 1 - (1) = 8 - (1)$ — The "1's" cancel out on the left; simplify right.

$\frac{5Y + 3}{2} = 7$	Which goes next? The "2" of course as it affects the most values. Multiply to remove it.
$\frac{5Y + 3}{2} \times \frac{2}{1} = 7 \times 2$	2's cancel out on left; simplify right.
$5Y + 3 = 14$	Which goes next? The "3" since it affects the whole "5Y." Subtract to remove it.
$5Y + 3 - (3) = 14 - (3)$	3's cancel out; simplify the right.
$5Y = 11$	Now remove the "5" through division.

$Y = 11/5$ is the Answer (Whew!)

A.9 Exponents and Roots

Rule A: The expression X^2 means raise X to the 2nd power by multiplying it by itself one time: $X \times X =$ ____.

Examples: $10^2 = 10 \times 10 = 100$; $5^2 = 5 \times 5 = 25$

For X^3, multiply X by itself two times: $X \times X \times X =$ ____

Example: $2^3 = 2 \times 2 \times 2 = 8$

Use the same basic logic for all exponent values, X^5, X^{16}, etc.

Rule B: The square root of a number is a number that when multiplied by itself will yield the original value.

Examples: What is $\sqrt{16}$? What we are really asking is what number will yield "16" when multiplied by itself. The answer is 4.

$\sqrt{144} = ?$ Ans.: 12 since $12 \times 12 = 144$

Square roots can be found very easily using a calculator.

A.10 Summation

A very frequently used symbol is the capital Greek letter (Σ) Sigma - it means "sum of" or "add up."

Examples:

A. For the following values of X, what is ΣX?

$X = 2, 4, 6$

List X's in a column, and sum down to obtain answer:

$$\begin{array}{rr} & \underline{X} \\ & 2 \\ & 4 \\ & \underline{6} \\ \Sigma X = & 12 \end{array}$$

B. When X's consist of scores of 1, 4, and -3, what is ΣX^2?

List X's in a column, but note that problem asks for the sum of the "squared X's." Thus, an adjacent column listing the X^2's is needed to obtain the required result.

$$\begin{array}{rrrr} & \underline{X} & & \underline{X^2} \\ & 1 & & 1 \\ & 4 & & 16 \\ & \underline{-3} & & \underline{9} \\ \Sigma X = & 2 & \Sigma X^2 = & 26 \end{array}$$

The answer is 26.

C. For the same data as in "B," what is $(\Sigma X)^2$?
The parentheses makes the problem read: What is the square of the sum of all the X's? Do you see the difference between this type of problem and that represented in "B"?

To derive the correct answer, you sum X's (X), and then square the result. As shown in "B," $\Sigma X = 2$. Thus:

$(\Sigma X)^2 = (2)^2 = \underline{4}$

D. When X = 1, 3, and 2, and Y = 7, 1, and 4, what is ΣXY?

This one may seem a little tricky, but it means: what is the sum of the products of each X multiplied by its corresponding Y? In other words, add the "XY's" together. We start by making an X column, a Y column, and then an XY column. Summing down the latter gives the result.

$$\begin{array}{rrrrrr} & \underline{X} & & \underline{Y} & & \underline{XY} \\ & 1 & & 7 & & 7 \\ & 3 & & 1 & & 3 \\ & \underline{2} & & \underline{4} & & \underline{8} \\ \Sigma X = & 6 & \Sigma Y = & 12 & \Sigma XY = & 18 \end{array}$$

(note: this is formed by X x Y)

The answer is 18

Unit A Review Test

Check your answers in answer section. If you score below 80% you need review, and you should check the appropriate sections in this text.

addition of positive and negative numbers

1. $36X + 23X = ?$
2. $-17X$ (+) $9X = ?$
3. $-22X$ (+) $-16X = ?$

subtraction of positive and negative numbets

4. $13Z$ (−) $-4Z = ?$
5. $-20X$ (−) $6X = ?$
6. -8 (−) $-13 = ?$

multiplication of positive and negative numbers

7. -22 (x) $-4 = ?$
8. $7W$ (x) $-9W = ?$

division of positive and negative numbers

9. $-6X$ (÷) $1.5X = ?$
10. -5 (÷) $-2 = ?$

order of operations

11. $5 + (15 - 2) \times 2 = ?$
12. $4 + 2^2 + 5^2 = ?$
13. $8 \times 2 + 6^2 \div 3 = ?$

equations

14. Solve for X: $X - 10 + 4 = 13$
15. Solve for Y: $4/5 = Y/15$
16. Solve for X: $(X - 6 \div 2) \times 2 = 12$

squaring and square roots

17. $\sqrt{81} = ?$
18. $(14.2)^2 = ?$

fractions and percentages

19. $1/3$ (x) $3/9 = ?$
20. Compute as a percentage: $13/15$

summation

X	Y
11	3
10	5
8	2

21. $\Sigma X = ?$
22. $\Sigma Y = ?$
23. $\Sigma X^2 = ?$
24. $(\Sigma Y)^2 = ?$
25. $\Sigma XY = ?$

ERIC

UNIT I

ORGANIZATION AND PRESENTATION OF DATA

A. General Objectives

The very word "statistics" suggests to many people (and particularly those in statistics courses) a confusing mass of numbers, difficult to work with and impossible to interpret. The objectives of this unit are directed towards teaching the student how to take a set of data and organize it into manageable form. Usually this involves constructing what is called a frequency distribution which is later transformed into a more graphic visual form - a frequency polygon. The present material deals with the construction and use of these organizing devices in psychology and education.

B. Specific Objectives

1.1 Identify the upper and lower limits of scores

1.2 Differentiate between grouped and ungrouped frequency distributions

1.2.1 Construct ungrouped frequency distributions

1.2.2 Construct grouped frequency distributions

1.3 Construct frequency polygons, histograms, cumulative frequency graphs, and define uses of each

1.4 Be able to identify the form of distribution with respect to modality, skewness, and shape

1.1 Instructional Unit: Real Limits of Scores

Whenever we work with numbers, a problem that is frequently encountered concerns whether the measure is continuous or discrete. Discrete measures can only fall at certain separated points on a scale. For example, if we were to measure (determine) the number of students enrolled in different sections of a statistics course, we might end up with numbers such as 5, 21, 16, 32, 7, 18, etc. These are discrete measures for they fall at separate integer points. You can have 5 students in your class, but will rarely find instances where student enrollment is 5.5 or 5.25. Similarly, the number of children in a family, the number of multiple-choice items correctly answered, the number of automobiles in a person's garage, or the number of mice in one's basement can only take on integer values, and thus remain discrete.

Other scale values, however, are not restricted to certain points. The only reason that they may appear to be is related to the limited sensitivity of our measuring instruments. For example, if you were asked to measure pencils, given a ruler that shows only whole inches (not halfs, fourths, or sixteenths, etc.), the scores reported for the first three pencils might be something like 9 in., 12 in., and 5 in. *BUT* is the first pencil really exactly 9 in. on the nose (or on the "tip" to be more exact)? If we gave you a ruler that was sensitive to 1/10 of an inch, your answer might change to, say, 9.2 in. *BUT*, is it really exactly that? We could give you a third ruler that is sensitive to 1/100 of an inch, only to find that you now report the pencil to be 9.23 in. Are you getting the "point?" Whereas *discrete* measures take only certain values, *continuous* measures, such as height, length, time, age, ability, weight, etc., can take on an infinite number of values. For practical reasons, as well as because of physical limitations, we restrict these values according to the sensitivity of the measurement instruments we employ. Age, for example, is commonly measured in year units (but recognize that no one is exactly 31 years old or 57 years old, etc. given the infinite possible values of time); gasoline is commonly measured to the nearest 1/10 of a gallon (but does your car really take exactly 9.2 or 5.7 gallons to the molecule?); and weight is commonly measured to the nearest whole pound (but no one is exactly 130 lbs. or 146 lbs., etc.). Thus, whatever degree of sensitivity our measurements permit, we end up with a series of possible values. Between each pair of successive possible values (let's say 9 in. and 10 in. where length is measured to the nearest whole inch) is a "gap" that divides one value from the other. It is assumed that the area below the halfway point in the gap (which would be 9.5 in. in our example) is associated with the lower possible value (i.e., 9 in.), whereas the area above the halfway point is associated with the higher possible value (i.e., 10 in.). This may be getting a little heavy at this point, so we'll stop the narrative for awhile and consider some concrete examples:

1. Speaking of heavy and concrete, suppose that on a scale that records weight to the nearest whole pound only, "Bubbles the Whale" weighs in at 4,356 lbs. What is her real weight? Nobody knows, but since the scale recorded her at 4,356, we do know that it is closer to that possible (whole pound) value than to other possible whole pound values. More specifically, Bubbles' real weight passed the 4,355 point on the scale. Thus, we know that her real weight has to be somewhere above the halfway point in

the gap between 4,355 and the measured weight 4,356. If it was below the halfway point, the 4,355 - not the 4,356 would have been the value recorded. What is the halfway point? The answer is 4,355.5. If the real weight was below that, the needle on the scale would have stayed at 4,355 rather than moving, as it obviously did, to 4,356. The other side of the coin is that Bubbles never made it to the next possible weight, 4,357 lbs. (although she might after lunch). What that means is that her real weight also has to be below 4,356.5, the halfway point between the measured weight of 4,356 and the next possible higher one, 4,357. If it had been above that point, the needle would have tipped to the 4,357 rather than staying where it was. Getting the picture? Very quickly try this one: Jack, Bubbles' trainer, is recorded on the same scale as weighing 131 lbs. What are the limits of his real weight?

Ans. is 130.5 (lower) and 131.5 (upper). Why? The next possible lowest weight would have been 130 lbs. Given that the needle passed that and went to the measured 131 lbs., we know that Jack has to be somewhere above 130.5. On the other hand, the next possible higher weight would have been 132 lbs. Given that the needle never made it to there, we know that Jack has to be somewhere below 131.5.

2. If you understand the above example, you may want to skip this one and go to #3. Here it goes. Suppose that your boss gives you a box of pencils and a ruler, and asks you to measure each pencil to the nearest inch only (if you reported something like 10.1 in., you would be fired immediately). Well, you take out the first pencil and line up its bottom with the bottom of the ruler. The result is illustrated below:

XXXXXXXXXX PENCIL XXXXXXXXXXXXXXXXXXXXXXXX

1 2 3 4 5 6 7 8

RULER

What would you report? Assuming you wanted to keep your job, you would write down 6 in. for pencil #1. Is the pencil exactly 6 in.? Of course not, but why would you pick "6" instead of "5" or "7"? The obvious answer, "It's closer!!" In the same sense, try this one: How far can a bear go into the woods?

Ans.: Halfway, because once he/she passes that point, he/she is leaving the woods!!

How far can a pencil extend within the gap between 6 and 7 in. and still be a 6 in. pencil (if measurements are restricted to whole inches)?

Ans.: halfway, that is, to 6.5 in., because after that point the pencil becomes a 7 in. pencil.

Final question: What are the limits of the real length of the pencil used in the example?

Ans.: 5.5 (lower) and 6.5 (upper). As far as "real" length is concerned, this boundary is all we can know for sure.

3. The "limits" we have been talking about similarly apply to scores more precise than integers. Suppose that we are racing two turtles in the 10 inch dash. Recognizing the possibility of a new world record, we bring two stop-watches, capable of recording time to 1/10 of a second, to determine the speed of each turtle. The race begins. Turtle A, known to be a sprinter, takes an early lead, but tires at the 8 inch pole. Turtle B rapidly (for turtles) closes the gap, and sensing victory, pulls away during the final stretch and breaks the tape first with his shell. We check the stopwatches for records. There are none, but both finishers are recorded at 9.8 seconds. Assuming no timing errors were made, how is that possible since Turtle B was clearly the winner?

   Ans.: The 9.8's do not represent the turtles' exact times, but rather, only the "best" estimates given the limitations of the timing instruments. What can we say for certain about the real times?

   Ans.: Only that for both "A" and "B" their "real" times were somewhere above 9.75 sec. (if not, the stopwatch would not have made it past 9.7 to the 9.8 actually recorded), and somewhere below 9.85 sec. (if not, the needle would have "slipped on" to show 9.9 rather than the recorded 9.8). Whatever the exact times were, we can assume that Turtle B's was slightly better than Turtle A's, but we probably would have needed a watch calibrated to 1/100 to show it.

Thus, anytime a continuous measure is expressed in terms of *discrete* points, the result is gaps between possible scores. Halfway into the gap is a point which divides one possible score from another. Given a possible score, the dividing point below it is called its lower limit, while that above it is called its upper limit. Together, these two dividing points comprise the real limits for the given score.

(Think about the above for a minute or two, and then try the following examples. Cover the answers as you go so that you can test yourself on the questions presented.)

1. IQ scores are expressed as integer values such as 100, 95, 127, etc. What are the real limits of a score of 97?

   Ans.: 96.5 (lower) and 97.5 (upper). Note that these separate the 97 from the next possible lower and higher scores, 96 and 98, respectively.

   What are the real limits of a score of 100?

   Ans.: 99.5 (lower) and 100.5 (upper).

2. Gasoline is sold to the nearest 1/10 of a gallon. What are the real limits of 10.6 gallons?

   Ans.: 10.55 (lower) and 10.65 (upper). Note, for example, that 10.55 equally divides the recorded 10.6 from the next possible lower score, 10.5.

   What about 20.1 gallons?

   Ans.: 20.05 (lower) and 20.15 (upper)

3. Meat is weighed to the nearest 1/100 of a pound. What are the real limits of 1.29 lbs.?

   Ans.: 1.285 (lower) and 1.295 (upper)

4. Suppose meat is weighed to the nearest 1/1000 of a pound? Give the limits of 3.296 lbs.

   Ans.: 3.2955 (lower) and 3.2965 (upper)

   Having trouble? Note that for the last example the next possible lowest weight would have been 3.295 (remember, the numbers can change in this one by 1/1000). Isn't it true that a 3.2955 is halfway between 3.295 and the recorded 3.296? If you don't see this, add a zero to the two possible scores, which is permissible, as the zero will not change their values:

		3.2960 (recorded)
3.2955 (midpoint in gap)	————	
		3.2950 (next possible lower score)

   Thus, the 3.2955 is the recorded score's lower limit. By the same logic, 3.2965 would be determined as its upper limit.

1.1 POSTTEST

1. Suppose we are measuring the length of pencils to the nearest inch: What are the real limits of scores of 10 in. and 5 in.?

2. Reaction time is measured in 1/1000 of a second. What are the real limits of 1.372 sec.? 1.300?

3. Speed in a race is measured to nearest tenth of a second. What are the real limits of scores of 10.1? 11.0?

See answer section for correct answers. If you missed any, reread instructional unit (or check Yellow Pages for a statistician on call).

ERIC

1.2.1 Instructional Unit: Ungrouped Frequency Distributions

Suppose you are a teacher who has given a test to 20 students. You spend all night grading the papers and when you are finished you end up with the following scores:

1	8	10	4
1	0	6	3
3	7	4	7
9	2	9	3
6	5	2	8

You could leave the scores in a random arrangement as illustrated above, but this would provide you with very little idea of how individuals performed and how the class did as a whole. A much better procedure would be to construct what is called a <u>frequency distribution</u>. The use of a frequency distribution puts the data in a form that makes sense and answers most of the general questions you might have regarding what occurred on the test. A frequency distribution of the above data might look something like this:

<u>X</u>	<u>f</u>	<u>cf</u>
10	1	20
9	2	19
8	2	17
7	2	15
6	2	13
5	1	11
4	2	10
3	3	8
2	2	5
1	2	3
0	1	1

$n=20$

It should be immediately obvious that the frequency distribution puts the data into a more manageable form. The "X" column lists all the possible scores; the "f" column tells us the frequency (or number) of the individual scores obtained; and the _cf_ column indicates the cumulative frequencies of the scores. For example, looking at the _cf_ for a score of 3, we know that 8 scores were obtained equal to or less than 3.5 (remember from Unit 1.1 that the upper limit of a score of 3 would be 3.5).

The steps involved in constructing a frequency distribution are fairly simple:

1. Find the highest and lowest values. In the example the high value is 10 and the low value is 0.

2. Set up the X column going from the high value at the top (10) to the low value (0) at the bottom. Include all possible scores in the column; for example, even if no students earned a score of "3" on the test, you would still usually want to include the category "3" in the X column.

3. Starting at the top (i.e., at the score category, 10) tally how many times each score occurred. This value will be recorded in the _f_ column.

4. Once all the _f_'s are recorded, the next task is to fill-in the cumulative frequency (_cf_) column. The _cf_ for a particular score indicates the number of frequencies _below_ that score. You construct the _cf_ column by starting at the bottom of the distribution and building up, one row (i.e., score) at a time. The first score in our example (on the previous page) is "0". The _f_ for "0" is 1, meaning that one person obtained a 0 on the test. How many scores have accumulated at the "0" category? Only 1 (the same person) therefore, _cf_ is also 1. Moving to the next score, "1", we find an _f_ of 2 (two people scored a "1" on the test). How many scores have accumulated there?

   Ans.: 3 (the two people who scored "1" and the one person who scored "0").

   So, you proceed by row, adding the frequency for the row score to the _cf_ for the previous row. The result is a _cf_ for that row. Check this out by examining the _cf_ listing for the score, "5". Note that "5" has an _f_ of 1. Its _cf_ of 11 was obtained by adding its frequency (i.e., 1) to the _cf_ for the score below it (_cf_ for "4" = 10). Thus, we can say that "11" (10 + 1) frequencies have accumulated through a score of "5".

   Now it is time to throw a monkey wrench of sorts into the conceptual machinery. If you are observant, you may have noted that our description of how to derive _cf_ did not exactly conform to the definition given in the second sentence above; i.e., cf = the number of frequencies _below_ a certain score. Turn once again to the example and examine, for instance, the tabulations for the score, "4". While it is true that 8 people clearly scored <u>below</u> "4" (check the <u>cf</u> for "3") and 2 people scored <u>at</u> "4", can we really say that 10 people scored <u>below</u> "4"? No! That is because <u>cf</u> is really supposed to be associated with the <u>upper limit</u> of the scores. Thus, checking the example (this is the last time - promise), we can say that 1 person scored below ".5" (the upper limit for "0"); 3 people scored below "1.5" (the upper limit for "1"); 5 people scored below 2.5; and, skipping a few, 17 people scored below "8.5".

Question: What score point do all 20 students fall below?

Ans.: If you said "10" you were close. A better answer would be "10.5" because that's the highest possible limit for total score distribution. None of the 20 people could have been above that without necessitating the addition of a higher score (X) listing on the distribution.

Before leaving this section, make sure you understand the following:

(a) <u>f</u>'s are associated with the actual score listings (X's) on the distribution.

(b) <u>cf</u>'s are associated with the upper limits of those listings.

Close your eyes and say that three times, and your book and statistics course will forever disappear.

(Book still around? You were supposed to say "*that three times*" <u>once</u>.)

5. Add up the <u>f</u> column to find <u>n</u>. The symbol "<u>n</u>" is used in statistics to indicate the total number of scores.

1.2.1 POSTTEST (answers in back)

Construct a frequency distribution for the following data. Be sure to include a <u>cf</u> column in the table. Also determine <u>n</u>.

.10	.30	.30	.50
.20	.00	.10	.30
.10	.80	.60	1.00
.70	.90	.90	.00
.60	1.00	.90	.10
.40	.60	.90	.70
.40	.50	.90	.60
.50	.20	.60	.50
		1.00	.20

Final note: The above descriptions pertain to ungrouped frequency distributions because no tampering has been done to the size of the original units. Also, every number still maintains its own identity after the frequency distribution has been made.

1.2.2 Instructional Unit: Grouped Frequency Distributions

Many times it is impractical to attempt to list every possible score in a frequency distribution. This may especially be the case when scores take on many different values. Thus, in order to simplify the recording and interpretation of results, grouped frequency distributions are often employed.

There are no set rules for the construction of grouped frequency distributions, but the following guidelines are generally followed whenever possible:

(1) The use of 8-15 intervals is fairly common. The larger the sample of cases, the larger number of intervals you may wish to employ.

(2) Intervals of the same width (i) are usually preferred.

(3) The interval limits should be chosen so that the interval midpoint tends to correspond with the observed values.

More specific rules are as follows:

(1) Order the scores from highest to lowest.

(2) Find the approximate range of scores by subtracting the lowest score from the highest score.

(3) Divide the range into equal units so that you end up with 8-15 intervals. For example, if your highest score is 100 and your lowest score is 30, the range would be 70. You therefore might decide upon 8 intervals of 9 units, 10 intervals of 7 units, 11 or 12 intervals of 6 units, etc.

(4) Make sure your lowest score will be placed within the lowest interval. For example, if our lowest score was 30 and we decided to use intervals of 9 units, we might start with the interval, 26-34, followed by 35-43, etc. You may note that when the interval width is odd in number (such as 5, 7, 9, etc.) the midpoint of the interval will be a whole number (midpoint of 26-34 would be 30, midpoint of 35-43 would be 39). Thus, odd-numbered interval widths are usually desirable.

(5) Complete the recording of intervals from low to high, and tabulate the frequency of scores that fall into each.

(6) Then record cumulative frequencies (cf) as you did for ungrouped frequency distributions.

Example problem: Construct a group frequency distribution for the following data:

45	52	75	47
64	90	42	57
73	74	59	81
49	58	87	90
68	72	46	63
70	86	79	56
80	75	65	75
87	49	50	67
90	85	84	83
68	69	76	56

1. High score is 90, low score is 42; range is therefore approximately 50. Using intervals 5 units in length would yield 10 separate intervals (which would be fairly desirable). 5 is also an odd number, and therefore the midpoints of the intervals will be whole numbers (again, a desirable feature).

2. Select lowest interval making sure that it will include the lowest score. An interval of 40-44 seems sufficient for a start.

3. Build up from your lowest interval until you reach the point where you can accommodate the highest score (in this case 90). Do not worry if you end up with one interval more (or less) than you originally planned.

4. Tabulate frequencies and cumulative frequencies as follows:

X	f	cf
90-94	3	40
85-89	4	37
80-84	4	33
75-79	5	29
70-74	4	24
65-69	5	20
60-64	2	15
55-59	5	13
50-54	2	8
45-49	5	6
40-44	1	1
	n=40	

Final note: The purpose of using grouped frequency distributions is to condense data. Keep in mind that the example frequency distribution, shown above, is capable of accommodating all possible scores. Take, for instance, the lowest interval in the distribution--40-44. What are the real limits of that interval? The answer is 39.5-44.5 (see Unit 1.1 if you are uncertain about this). What are the real limits of the next highest interval, 45-49? Answer: 44.5-49.5. So, if by some chance, we obtained a score of 44.32, in which interval would it be represented? Answer: We would place it in the interval 40-44, since it falls within the real limits of only that interval. Similarly, any score you could possibly imagine between 39.5-94.5 (the real limits of the whole distribution) falls within one and only one interval in the distribution. What might give you a little more difficulty is the idea of cumulative frequency. Notice on the distribution that we have a cf of 1 for the interval 40-44, a cf of 6 for the interval 45-59, etc. Understanding how we obtained those values should not present much of a problem, but try this: To what <u>specific</u> <u>score</u> does a <u>cf</u> of 1 refer? A reasonable guess would be "42," the midpoint of the corresponding interval, 40-44. A reasonable guess in this case is not a correct one: We really cannot assume that one score "accumulated" that was less than or equal to 42. The score could have been a 43, a 43.57, a 44.05, etc. But there is one thing we positively can assume in talking about our <u>cf</u> of 1, and that is that the score was less than 44.5, the upper limit of the interval. If it was not less than 44.5, it <u>could</u> <u>not</u> have been placed within that interval. Think about it. (And remember that 44.5 is also the lower limit of the interval 45-49).

Now try this: To what hypothetical score would we associate a <u>cf</u> of 13? Looking at the distribution, we immediately realize that the 13th score fell within the interval, 55-59. Is the hypothetical score therefore 57, the interval midpoint? No, we cannot be certain that the 13th score was less than or equal to 57 (it might have been 58, 59.2, etc.) But, we can be absolutely certain that the 13th score is less than 59.5 (the upper limit of the appropriate interval); our answer therefore should be 59.5. One more for practice: We can be absolutely certain that our 24th score falls below what hypothetical cutoff point? Answer: 74.5, the upper limit of the interval 70-74.

1.2.2 POSTTEST (answers in back)

Construct a grouped frequency distribution for the following data:

01	07	40
31	20	18
16	01	07
35	36	05
14	30	07
10	08	36
12	01	36
15	14	30
36	09	10
05	01	07
09	20	18

1.3.1 Instructional Unit: Frequency Polygons

A frequency polygon is used to provide a pictorial or graphic representation of your frequency distribution. It is extremely valuable in that it gives you or your reader a quick, clear picture of the distribution of scores. The procedures involved in constructing a frequency polygon are fairly simple:

(1) Draw a horizontal axis, and label it "X." It will be used to represent all possible scores. Divide the X axis into equal units; each unit will correspond with one of the points on your score scale. (A good length for the X axis is 3" to 5").

(2) Draw a vertical axis, and label it "*f*." It will be used to show the frequency of your scores. In dividing the *f* axis into equal units, start at zero and build up to the highest *f* value in your frequency distribution. You will want the *f* axis to be approximately the same length as the X axis. Therefore, it is not necessary to include every possible *f* value on your graph; you may wish to build up by 2's, 5's, 10's, etc.--whatever seems reasonable considering the amount of scores with which you are dealing.

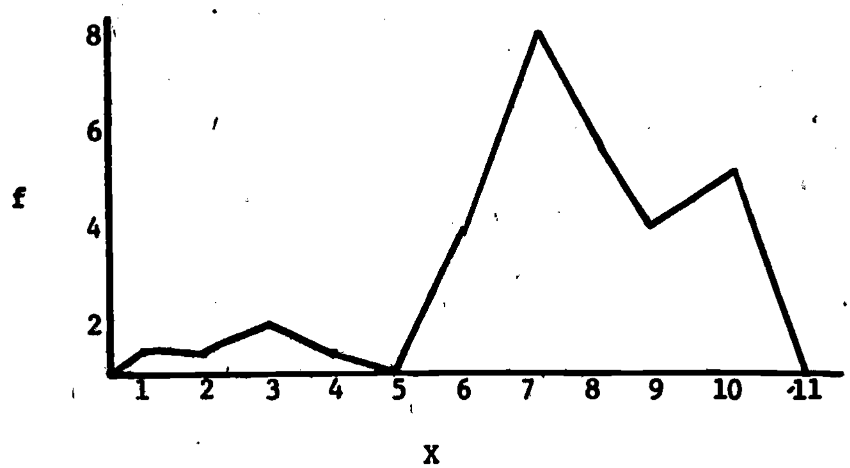
(3) Once the axes are constructed and labeled, place a dot directly above each X value equal to its corresponding *f* value.

(4) Finally, connect the points together and bring both end points down to where the next X value would be on the X axis.

Example problem: Construct a frequency polygon for the following data:

X	f	cf
10	5	32
9	4	27
8	6	23
7	8	17
6	4	9
5	0	5
4	1	5
3	2	4
2	1	2
1	1	1

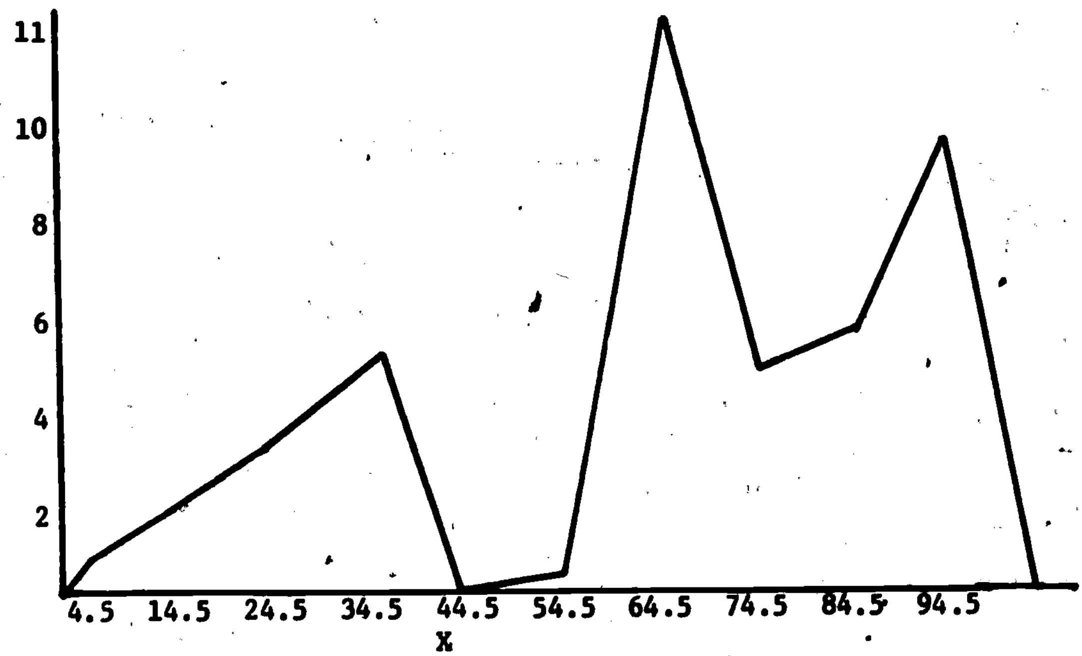
The resulting frequency polygon is:



<u>Final note</u>: The procedures for constructing a frequency polygon of grouped data are essentially the same but there is one major difference. Examine the frequency distribution shown in Unit 1.2.2. Looking at the various intervals, which scores do you plot? Unlike what was discussed to be case for <u>cf</u> (where it was decided at the conclusion of 1.2.2 that the upper limit of the interval comprised the desired score), this time we are looking for the score that is most <u>representative</u> of the whole interval. If you are thinking "midpoint" this time, you are entirely correct. So, the only real difference in plotting grouped data is that the various interval <u>midpoints</u> will be represented on the X axis.

Example problem: Construct a frequency polygon for the following data:

X	f	cf
90-99	10	44
80-89	6	34
70-79	5	28
60-69	11	23
50-59	1	12
40-49	0	11
30-39	5	11
20-29	3	6
10-19	2	3
0-9	1	1



Final note: Notice that when we are dealing with grouped data, the interval <u>midpoints</u> are used on the X axis. Also note that graph is <u>anchored</u> to horizontal axis at points preceding and beyond the score labels.

1.3.1 POSTTEST (answers in back)

Construct frequency polygons for the following distributions:

a) X	f	cf	b) X	f	cf
25	1	19	24-26	4	45
24	2	18	21-23	12	41
23	3	16	18-20	9	29
22	6	13	15-17	10	20
21	3	7	12-14	1	10
20	2	4	9-11	3	9
19	1	2	6-8	2	6
18	1	1	3-5	4	4

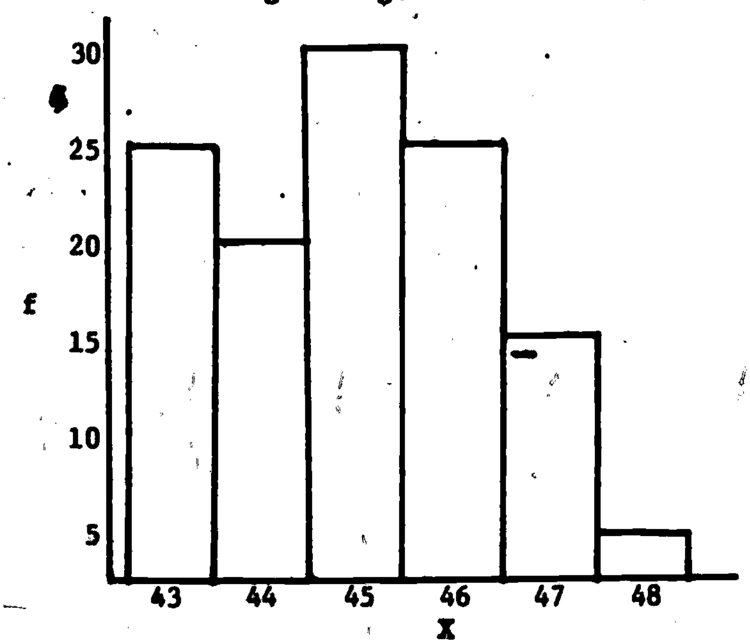
1.3.2 <u>Instructional Unit:</u> Histograms

An alternative way of graphing frequency data is via a histogram. Histograms are virtually the same as the frequency polygons we have just covered, except that vertical bars, extending in width from the lower limit to the upper limit of the score or interval, are used instead of connecting lines. The height of the bars, comparable to the placement of the dot in the frequency polygon, corresponds to the frequencies of the scores or intervals as labeled on the vertical axis. Make your bars even in width; <u>be sure to use the midpoint</u> of the score or interval, <u>positioned in the middle</u> of the bar, as your horizontal axis label.

Consider the following frequency distribution:

X	f	cf
48	5	120
47	15	115
46	25	100
45	30	75
44	20	45
43	25	25

The resulting histogram would look like:



Note that the lowest vertical line is the lower limit for the score of 43 (=42.5) and the uppermost vertical line is the upper limit for the score of 48 (=48.5).

47

The procedures are the same for constructing histograms of grouped data. Just make sure that the midpoints of the "bars" correspond to the midpoints of the intervals.

1.3.2 POSTTEST (answers in back)

Construct histograms using the data provided in POSTTEST 1.3.1.

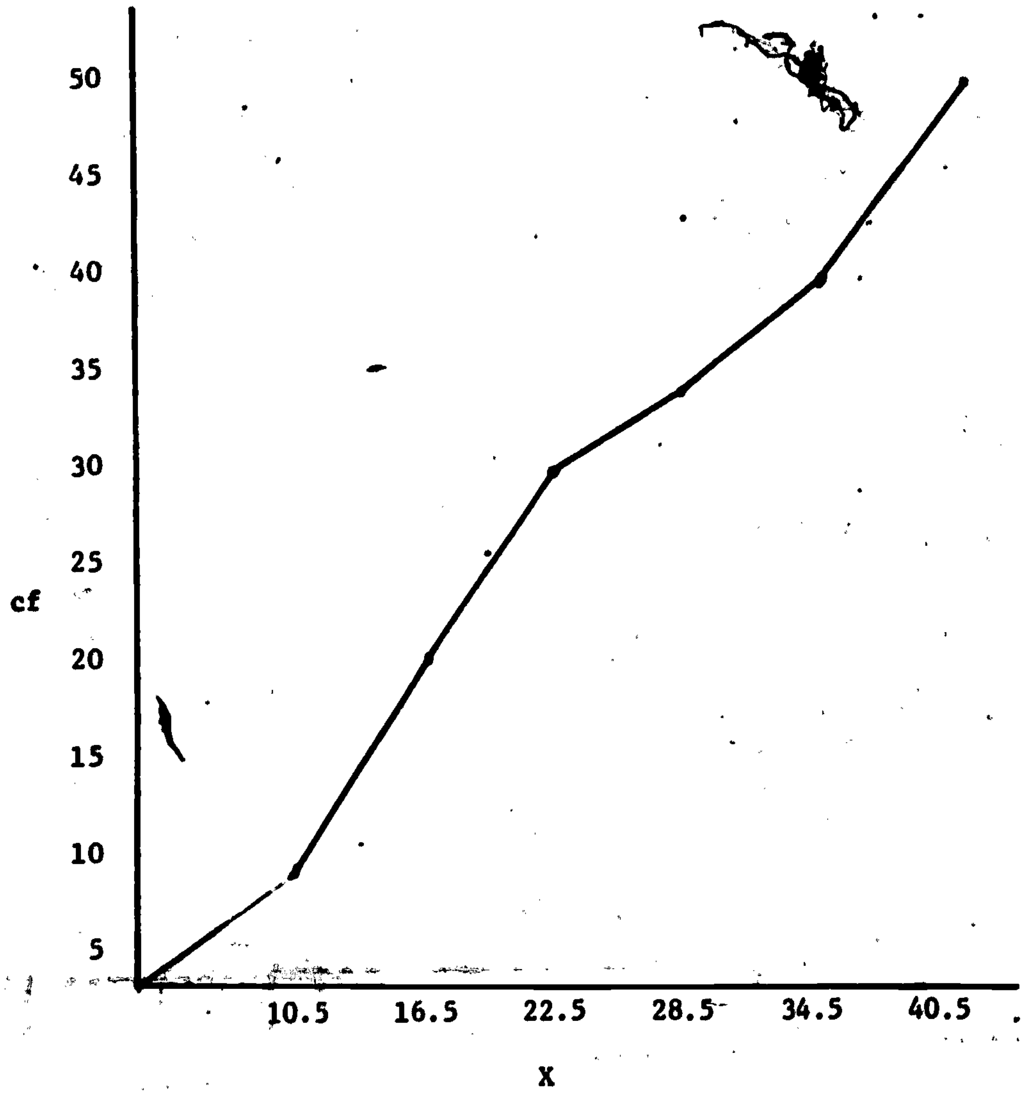
1.3.3 <u>Instructional Unit</u>: Cumulative Frequency Graphs

Cumulative frequency graphs show how many observations fall below a given score or class interval. There is nothing "tricky" about plotting cumulative frequency. The vertical axis of the graph is labeled "<u>cf</u>" with the various points used to represent the cumulative number of observations. The horizontal axis may still be labeled "X" (as in the case of frequency polygons and histograms) but this time the points on the X axis will be expressed in terms of the <u>upper limits</u> of the scores. For review, think about this and remember it: In plotting the cumulative frequency of a class (or score), all observations are associated with the <u>upper limit</u> of the class (or score), not the midpoint.

Consider the following frequency distribution:

<u>X</u>	<u>f</u>	<u>cf</u>
35-40	10	49
29-34	6	39
23-28	5	33
17-22	9	28
11-16	11	19
5-10	8	8

The resultant cumulative frequency graph would look like:



<u>Final comments</u>: Notice that the upper limits of the various score intervals are used on the X axis. Also notice that the higher end of the graph is not brought down to zero. Finally, notice that the steepness of the slopes correspond to the amount of increase from interval to interval: When a large number of observations fall into a particular score interval, the slope becomes steeper.

1.3.3 POSTTEST (answers in back)

Construct cumulative frequency graphs for the data presented in POSTTEST 1.3.1.

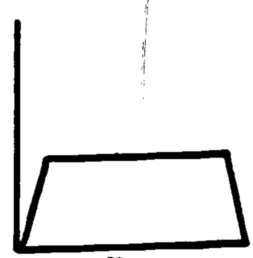
1.4 Instructional Unit: Forms of Frequency Distributions

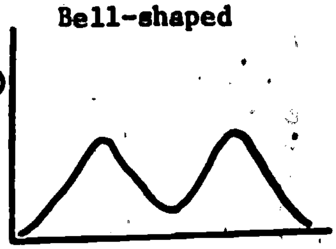
Check back to Unit 1.3.1 and examine the two example frequency polygons. If you are observant you will notice that they are not identical in form. Rarely, in fact, will two frequency polygons ever be identical since the form of each is determined by the number of observations recorded for the specific score points or intervals. Being able to characterize the form of a frequency polygon (and in essence the form of the frequency distribution) is very advantageous in reporting data, especially if space does not permit including the actual polygon in your paper. Generally, there are three types of characteristics that you should become familiar with. These are as follows:

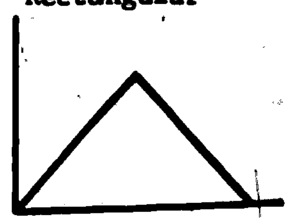
(1) Shape - frequency polygons can normally be described in terms of the "general" shape in which they appear. Some common examples are:

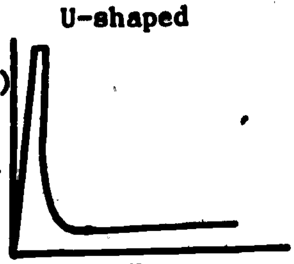
a) f

X
Bell-shaped

b) f

X
Rectangular

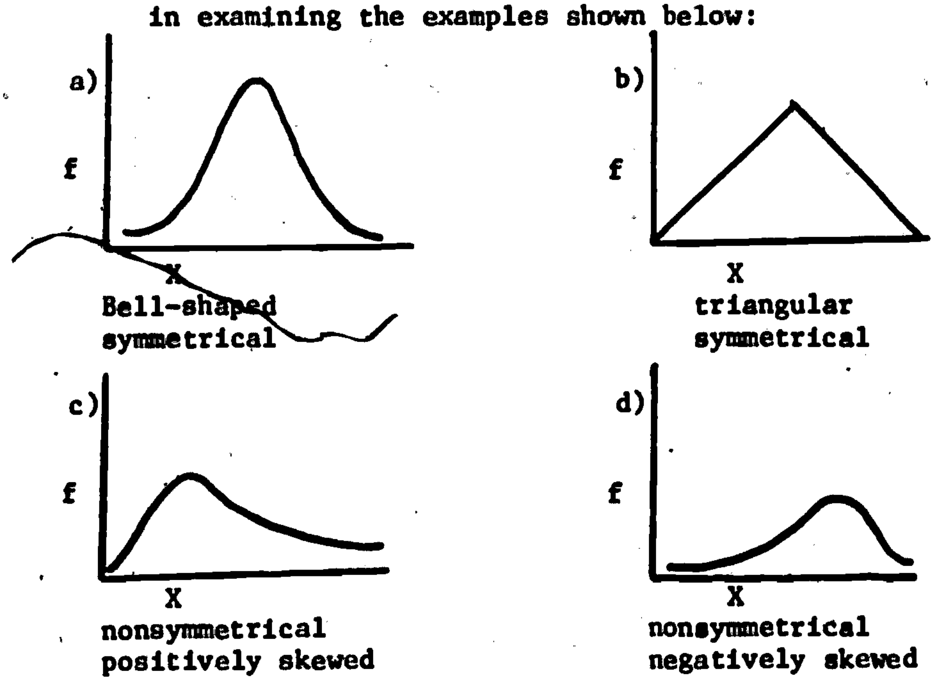
c) f

X
U-shaped

d) f

X
Triangular

e) f

X
J-shaped

(2) Symmetry - another distinction can be made relating to whether your frequency polygon is symmetrical. A distribution is symmetrical if one half of it can be folded over so that it is exactly superimposed over the other half. Distributions that are not distinctly symmetrical are characterized as being skewed. Skewed distributions are labeled by the

direction in which the longer tail is pointing. When the tail points to the right, we say that the distribution is positively skewed. This means that there is a preponderance of "lower" scores and relatively few scores in "higher" intervals. When the tail points to the left, we say that the distribution is negatively skewed. This means a preponderance of "higher" scores and relatively few "lower" scores. Think about these descriptions in examining the examples shown below:



a) Bell-shaped symmetrical

b) triangular symmetrical

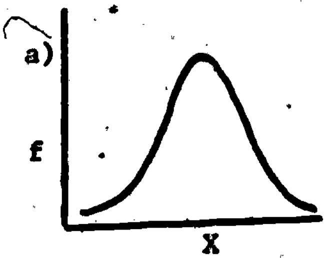
c) nonsymmetrical positively skewed

d) nonsymmetrical negatively skewed

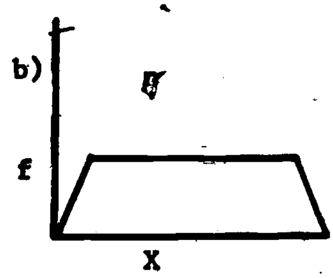
(3) Modality - A final distinction between frequency polygons can be made on the basis of the number of modes (peaks or high points). A unimodal distribution has only one relatively high point, whereas bimodal distributions have two high points or peaks. Similarly, a distribution with three peaks would be trimodal. What does it mean to your reader when you describe your distribution as unimodal, for example? If he is bright and competent in his knowledge of statistics, he immediately knows that, in your distribution, there was a fairly high concentration of scores within one particular interval. If you tell him your distribution

is bimodal, he knows that two score intervals (usually separated to some extent) were relatively popular, etc.

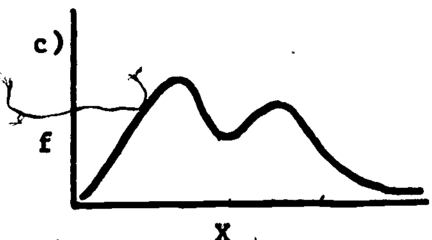
Consider the following:



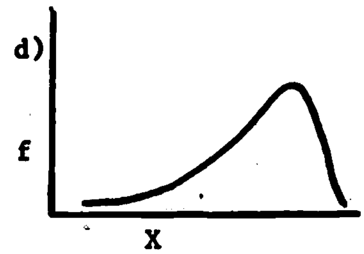
Bell-shaped symmetrical unimodal



Rectangular symmetrical amodal (no Modes)



Positively skewed Bimodal



Negatively skewed unimodal

Final note: It is not always necessary to actually draw a frequency polygon in order to characterize the form of the distribution. For example, consider the following very simplified frequency distribution:

X	f	cf
9-10	1	9
7-8	2	8
5-6	3	6
3-4	2	3
1-2	1	1

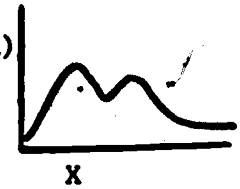
Just by looking at the distribution we should feel confident in characterizing its form as follows: a) bell-shaped, b) symmetrical (as many high scores as low scores), and c) unimodal (really only one mode--interval 5-6).

If you are uncertain, quickly plot the above data, and read the above characterization again.

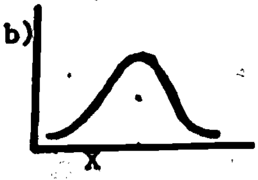
1.4 POSTTEST (answers in back)

Characterize the form of the following distributions in terms of general shape, symmetry, and modality (Note: bimodal distributions typically are u-shaped--no need to say both).

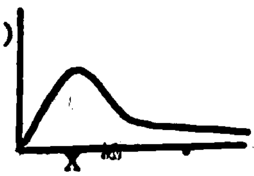
a)



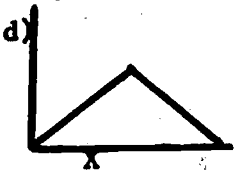
b)



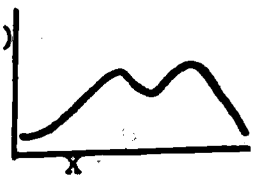
c)



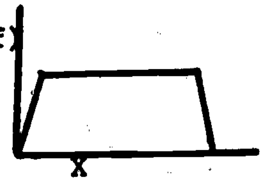
d)



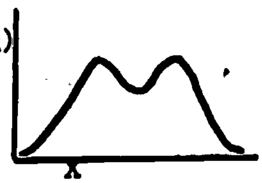
e)



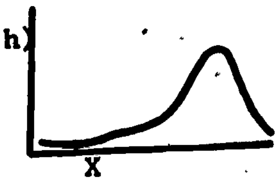
f)



g)



h)



i)

X	f	cf
25	4	32
24	4	28
23	4	24
22	4	20
21	4	16
20	4	12
19	4	8
18	4	4

j)

X	f	cf
44-46	1	72
41-43	2	71
38-40	4	69
35-37	6	65
32-34	8	58
29-31	15	51
26-28	9	36
23-25	11	27
20-22	16	16

Unit I Review Test
(answers in back)

The administrator of an adolescent center lists the ages of her 20 clients as follows:

17	17	18	16	17
18	19	17	15	16
17	16	17	16	19
18	17	18	15	17

Work with these data by doing the following:

1. Construct a frequency distribution of scores, listing f, cf, and n. Why would a grouped distribution be inappropriate?

2. What are the real limits of a score of 18?

3. What score point would be associated with the highest cf?

4. If you were to plot a cf graph, what types of score points would be used to define the horizontal axis?

5. What would be the shape of a frequency polygon of this distribution?

ERIC

UNIT II

PERCENTILE RANKS AND

CENTRAL TENDENCY

A. General Objectives

The preceding unit dealt with methods that are commonly used to facilitate the organization and interpretation of group data. In addition to "summarizing" data by use of frequency distributions, polygons, etc., it is usually desirable for the statistician to provide the reader with more specific information regarding the meaning of individual scores; that is, how individuals performed in comparison to others. One of the general objectives of the present unit involves familiarizing the student with percentile ranks - a descriptive index frequently employed to describe relative standing (individual scores) on aptitude and achievement tests. Another general objective is to facilitate the interpretation of group performance by presenting three numerical indices of "average" or "typical" achievement - the mean, the median, and the mode.

B. Specific Objectives

2.1 Define percentile rank
2.2 Compute percentile ranks for specified raw scores
2.3 Convert specified percentile ranks into raw scores
2.4 Define central tendency

2.4.1 Define and compute means
2.4.2 Define and compute medians
2.4.3 Define and compute modes

2.5 Select the appropriate relative positions of the mean, median, and mode in skewed and non-skewed distributions

2.1 Instructional Unit: Percentile Ranks

A percentile rank is a number that represents the percentage of scores that fall at or below some selected raw score. A percentile rank of 80, for example, means that a certain score has 80% of the other scores equal to or below it. In other words, if the measure involved was a mathematics achievement test, a percentile rank of 80 would mean that 80% of those who took the test scored at or below the selected score. A percentile rank of 50 would be a score where 50% of the scores fall below that point. We can symbolize the reporting of percentile ranks by P_x, where X represents the particular percentile in question; for example, the percentile rank of 80 would be symbolized as P_{80}, the percentile rank of 36 would be symbolized as P_{36}. Thus, when we say that John scored at P_{67}, this can be taken to mean that John scored the same as or higher than 67% of the students in his class (or people taking the same test).

Do not confuse percentile ranks with actual scores. John may earn a score of 80 correct on his basic algebra test. Depending upon how other people in the class performed, John's score could be reported as P_{20}, P_{45}, P_{68}, etc. That is, the percentile rank for a score of 80 will be determined according to what this particular score means in relation to the class performance as a whole. If a score of 80 was higher than 50% of the remaining scores, then the percentile rank would be P_{50}; if it was higher than 95% of the remaining scores, then the percentile rank would be P_{95}, and so on.

ERIC

2.1 POSTTEST (answers in back)

a) Define percentile rank: ____________________

____

____

____

b) What does it mean when we say that a student performed at P_{89}?

____

____

c) John took a math test and scored a 96 which was reported to be P_{79}. Later that day, he took an English test and earned a score of 42. The English score was found to be P_{96}. On which test did John do better relative to the other students in his class? Why?

____

____

____

____

2.2 Instructional Unit: Computing Percentile Ranks

The basic formula for computing percentile ranks is $\frac{cf_x}{N}$, where cf_x is equal to the cumulative frequency of the scores in question, and N is equal to the total number of observations or scores. In Unit I we defined cumulative frequency (cf), and showed how it is presented in tables and graphs of frequency distributions. You will remember (hopefully) that cf shows how many observations fall below a given score or class interval. The cf of a class (or score) is found by adding the class (score) frequency and the frequency of all lower classes (scores). As we move to higher and higher scores on our frequency distribution, cf will likewise become higher and higher.

Now look again at the formula for computing percentile ranks; $\frac{cf_x}{N}$. All that it means is that if you are interested in finding the percentile rank for a particular score, you must find the cumulative frequency for that score (that is, cf_x) and divide it by the total number in your population (that is, N). By the way, if we are using sample ~~data,~~ the total number of scores is symbolized as n; if we are using population data, the total number of scores is symbolized as N. In either case, the procedure for computing percentile ranks is the same.

Let's try a very simple example. Fred earns a score of 56 on his basketweaving competency examination. Fred is concerned about the meaning of his performance, and also about his grade. We cannot make too many inferences about his grade (the teacher may be very nasty or very generous), but by computing the percentile rank for Fred's score of 56, we can, at least, tell him the percentage of people in the class that scored lower than he did. We know that there were 100 people taking the basketweaving competency exam; therefore N=100. When we order the scores from lowest to highest (as in a frequency distribution), we find that the rank of Fred's 56 is 38; therefore cf_x is approximately 38. Using the very simple computational formula, we find that Fred scored higher than 38% of the students in his class. His performance was at P_{38}. Simple!

In a class of 30 students, Mike earns a score of 87 on a geography final. Mike's score is higher than that received by 28 other students. Using the formula, we substitute 28/30 for $\frac{cf_x}{N}$, and find that Mike scored higher than 93% of those taking the test. His P.R. is therefore P_{93} (approximately).

Hopefully by now you see no difficulty with using the formula, $\frac{cf_x}{N}$. And even more hopefully, you fully understand the concept behind the formula; find out how many people you beat on the test or measure, divide that by the number of people who took the test, and you get a percentage. Multiply that percentage by 100 and you get a percentile rank. So technically the formula for percentile ranks is $\frac{cf_x}{N} \times 100$.

Now that you probably feel somewhat confident about your ability to compute percentile ranks, we will throw a "monkey wrench" of sorts into the conceptual machinery. The monkey wrench relates to the computation of cf_x, for if our percentile rank is to be accurate, our estimation of cumulative frequency for the score in question must also be accurate. In the preliminary examples involving the basketweaving and geography scores of Fred and Mike, we have been using approximations of cf_x, and thus, the resultant percentile ranks of P_{38} and P_{93} are also approximations. Examine the following frequency distribution.

X	f	cf
10	2	22
9	3	20
8	6	17
7	2	11
6	5	9
5	3	4
4	1	1

Suppose you are shown the above distribution by a worried student in the class. He tells you that he received a score of 7, and is uncertain about how he performed relative to others. Since you successfully completed Unit I in basic statistics, he regards you as an expert in interpreting test data. You feel extremely confident; not only have you completed Unit I, you also learned something about percentile ranks in Unit 2.1. You know that percentile ranks do provide valuable information about relative performance. Furthermore, you know the formula for computing percentile ranks: $\frac{cf_x}{N}$. You smile, because giving this student information about his relative performance is obviously a very simple matter.

You are ready to begin the simple computational process. N obviously equals 22, the number of observations (people taking the test); cf_x for a score of 7 appears to equal 11 (that is, 11 people scored within or below the score category, 7). You divide 11 by 22 and get .50, the proportion of people who scored below your friend on the test. You multiply by 100 and report this as P_{50}; you tell your friend he was exactly "average" in his performance. Half the class was lower and half was higher than he was. You feel very accomplished and brilliant, until another friend who has completed Unit II in basic statistics tells you that you were <u>wrong</u> in your calculations. Why?

The mistake involves selecting 11 as your cf_x for a score of 7. You will remember, perhaps, that 11 refers to the <u>cf</u> for the <u>upper limit</u> of the score value of 7. Thus, we can say with certainty that 11 people scored below 7.5 (the upper limit); 22 people scored below 10.5; 4 scored below 5.5, etc.

Before proceeding with the explanation, you should be made aware that you are likely to become very confused unless you have mastered the following concepts from Unit I:

(a) <u>real limits</u>: looking at the above frequency distribution, do you understand that a score ("X") listing of, say 8, really refers to an interval encompassing possible scores ranging from 7.5 to 8.5? If you don't, it's time to turn back.

(b) frequency distribution: looking again at the above frequency distribution, do you understand that the "X" column lists the scores that people could receive, the "f" column tells you how many of those scores were actually obtained, and the "cf" column tells you how many scores fell below the upper limits of the specified X's?

(c) cumulative frequency: as just asked, are you comfortable with what cf represents?

If your answer to any of the above is "no," the only way you will survive through this objective is by memorizing the formula, knowledge that may last for a couple of days and then disappear from your memory forever. A better procedure would be to turn back to the appropriate instructional sections, review the concepts that got a "no" response, and then tackle the following through meaningful, rather than rote, learning.

Getting back to the problem at hand: if your friend scored a 7 on the test (see the frequency distribution), why would it be incorrect to assume that he scored higher than 11 people, and therefore stands at P_{50} (11/22 x 100)? The answer is that a cf of 11 is associated with the upper limit of the score value, 7. Thus, a cf of 11 is associated with a score of 7.5 not 7.0. By the same reasoning, a cf of 9 is associated with a score of 6.5 (the lower limit of 7). DO NOT PROCEED UNTIL YOU CAN SEE FROM THE FREQUENCY DISTRIBUTION WHY THE LATTER IS TRUE. Thus, all that we know about your friend's standing is that he "beat" somewhere between 9 and 11 people, as his score of 7 is somewhere between 6.5 and 7.5. Represented as an illustration of sorts:

X (scores)		cf (number "beaten")
7.5	is associated with	11 people
7 (the score of interest)	is associated with	??
6.5	is associated with	9 people

The problem is that although we know that a score of 7 was higher than at least 9 other scores, we cannot say exactly where it stands in terms of a percentile rank. You will note on the frequency distribution that 2 scores fell in the interval for "7" (i.e., 6.5 - 7.5), but were they higher or lower than 7? They could have been (theoretically) 6.96 and 7.49; 6.51234 and 6.7; 7.001 and 7.36; etc. You will note that all of these possibilities would have simply been recorded as "7". So, since we can never know for certain, we must estimate, and please try to follow this conceptually: we assume that the scores associated with each interval are evenly distributed within those intervals; half fall above the midpoint and half fall below the midpoint, and so on. If 2 scores are recorded for the score point 7, we assume that one was greater than 7 and one was less than 7 (remember 7 is the midpoint of the interval, 6.5-7.5). Notice on the table that 6 scores are associated with the score point 8. All we can really assume about these scores is that they are evenly distributed in the interval 7.5 to 8.5 - three fall below the midpoint, 8, and three fall above the midpoint, 8. Getting back to the original problem of determining the <u>cf</u> for a score of 7: How many scores fell on or below the midpoint, 7? Since we are uncertain about two scores, we must use a process called linear interpolation which will estimate where those questionable scores would fall considering: 1) the size of the interval (in this case, 1 unit long - from 6.5-7.5); 2) the number of questionable scores (in this case, two). In using linear interpolation, we <u>assume</u> that all scores are evenly distributed along the interval with which they are associated. The linear interpolation formula we use for finding cf_x is presented below:

$$\frac{X - X_{LL}}{i} = \frac{cf_x - cf_{LL}}{f_w}$$

Where X is the score in question
X_{LL} is the lower limit of X's interval
i is the length of the score interval for X
<u>cf</u> is the cumulative frequency of X
(what we are actually trying to find)
cf_{LL} is the cumulative frequency of the lower limit of X
f_w is the frequency of scores associated with the interval containing X

At this juncture, all of this may seem extremely confusing, so let's backtrack a bit before working on the equation. Our objective, you will recall, involves the computation of <u>percentile ranks</u>. We know that:

1. The formula for computing percentile ranks for given scores (that is, X's) is $\frac{cf_x}{N}$. Thus, if we know the values of cf_x and N, our problem is simply one of long division.

2. We will usually have no problem finding a value for N because N refers to the number of scores (observations, X's, etc.) in our frequency distribution.

3. We may have a problem, however, in determining cf_x because our frequency distribution only tells us the cumulative frequency of the upper limits for the various score intervals. Looking at the example frequency distribution (turn back 2 pages) we know that the cf_x for a score of

7.5 is 11; we know that the cf_x for a score of 9.5 is 20, etc. But we cannot be sure about the cf_x for "midpoint" scores such as 4,5,6,7,8,9, and 10.

4. Therefore, in order to estimate cf_x, we must go through a process called linear interpolation. The particular equation we use is:

$$\frac{X - X_{LL}}{i} = \frac{cf_x - cf_{LL}}{f_w}$$

All of the variables in this equation are defined on the preceding page. Notice that one of the variables is cf_x. Thus, if we can substitute values for all of the other variables we should be able to determine cf_x for any given score.

5. Once we find cf_x, we divide it by N (see step 1), multiply by 100, and we get our percentile rank.

Let's go back to our original problem: find the percentile rank for a score of 7. Check back to example frequency distribution.

1. The equation we use is $\frac{cf_x}{N} \times 100$.

2. $N = 22$, which is the number of scores in our distribution. Therefore, Percentile Rank $= cf_x/22 \times 100$.

3. cf_x for a score of 6.5 is 9; cf_x for a score of 7.5 is 11. We have no direct indication of the value of cf_x for a score of 7 (which is what we need).

4. We must interpolate. We begin by writing down the equation and substituting values. Remember, we are looking for cf_x.

   a) $$\frac{X - X_{LL}}{i} = \frac{cf_x - cf_{LL}}{f_w}$$

   b) X = the score in question
   therefore, X = 7
   therefore, $\frac{7 - X_{LL}}{i} = \frac{cf_x - cf_{LL}}{f_w}$

   c) X_{LL} = the lower limit of X's interval
   therefore, $X_{LL} = 6.5$
   therefore, $\frac{7 - 6.5}{i} = \frac{cf_x - cf_{LL}}{f_w}$

d) i = the length of the score interval for X
therefore, $i = 1$ (7.5 - 6.5)
therefore, $\frac{7-6.5}{1} = \frac{cf_x - cf_{LL}}{f_w}$

e) cf_{LL} = the cumulative frequency of the lower limit of X.
therefore, $cf_{LL} = 9$ (lower limit of 7)
therefore, $\frac{7 - 6.5}{1} = \frac{cf_x - 9}{f_w}$

f) f_w = width of X's interval
therefore, $f_w = 2$ (2 scores fall between 6.5 - 7.5)
therefore, $\frac{7 - 6.5}{1} = \frac{cf_x - 9}{2}$

g) Work through equation:

$$.5 = \frac{cf_x - 9}{2}$$

$$1.0 = cf_x - 9$$

$$cf_x = 10$$

5. Determine percentile rank:

$$cf_x/N \times 100 = 10/22 \times 100$$

$$= .45 \times 100$$

$$= \underline{P_{45}}$$

Try to believe that all we did in using the interpolation procedure was:

a) determine that a score of 7 was half the distance between 6.5 and 7.5, the real limits of the interval in question.

b) conclude that therefore, half of the two scores associated with the interval 6.5-7.5 should fall below the score value, 7. Half of 2 is <u>1</u> (hopefully, you agree with this).

c) we added the one score to the <u>cf</u> for the lower limit of the interval ending up with $cf_x = 9 + 1 = \underline{10}$.

The above description is intended to show that there is nothing "magic" about the interpolation procedure. Examine the interpolation equation and see if you can understand its logic. If you can, you will probably experience little difficulty with this unit.

Example Problem: Using the same frequency distribution as before, compute the percentile rank for a score of 9.

1. $P.R. = cf_x/N \times 100$

2. $N = 22$

3. cf = 17 for 8.5, the lower limit of the score interval, 9.

4. $\dfrac{X - X_{LL}}{i} = \dfrac{cf_x - cf_{LL}}{f_w}$

   $X = 9$ $\quad cf_x = ?$
   $X_{LL} = 8.5$ $\quad cf_{LL} = 17$
   $i = 1$ $\quad f_w = 3$

   Substituting:

   $\dfrac{9 - 8.5}{1} = \dfrac{cf_x - 17}{3}$

   $1.5 = cf_x - 17$ (Note: We subtracted 8.5 from 9, and then
   $cf_x = \underline{18.5}$ cross-multiplied the 3 to get 1.5 on the left.)

5. $P.R. = 18.5/_{22} \times 100$

   $P.R. = 84$

   Answer = P_{84}

Notice that all we did by interpolating was:

a) Find that 9 was 1/2 the distance between 8.5 and 9.5.

b) Determine 1/2 the scores associated with the interval, 8.5- 9.5 (which was found to be 1.5).

c) Add 1.5 to 17, the cf of the lower limit, ending up with $cf_x = \underline{18.5}$.

Example Problem: The exact same procedures are used in determining percentile ranks for grouped data. Using the data provided below, compute the percentile rank for a score of 16.

X	f	cf
21-23	5	100
18-20	8	95
15-17	20	87
12-14	6	67
9-11	10	61
6-8	25	51
3-5	13	26
0-2	13	13

1. $P.R. = cf_x/N \times 100$

2. $N = 100$

3. cf = 67 for the value 14.5, the lower limit of the interval into which 16 falls

4. $$\frac{X - X_{LL}}{i} = \frac{cf_x - cf_{LL}}{f_w}$$

$X = 16$ (the score in question) $cf_x = ?$

$X_{LL} = 14.5$ $cf_{LL} = 67$

$i = 3$ (notice that the interval 14.5 - 17.5 is 3 units long) $f_w = 20$ (notice that $f = 20$ for the interval 14.5 - 17.5)

Substituting: $$\frac{16-14.5}{3} = \frac{cf_x - 67}{20}$$

$$30/3 = cf_x - 67$$

$$cf_x = 77$$

5. $P.R. = 77/100 \times 100$

Answer is P_{77}

Note that what was done can be illustrated as follows:

X (score intervals)		cf (number "beaten")
17.5	is associated with	87 people
16.0	is associated with	?
14.5	is associated with	67 people

By interpolating we found that 16 was 1/2 of the distance between 14.5 - 17.5. We therefore took 1/2 of the scores in that interval (that is 1/2 of 20) and added them to the cf for the lower limit (that is, 67). The result was a cf_x of 77.

Example Problem: Using the same data, determine the P.R. for a score of 6.

1. $P.R. = \frac{cf_x}{N} \times 100$

2. $N = 100$

3. cf = 26 for the lower limit of the interval associated with 6.

4. $\frac{X - X_{LL}}{i} = \frac{cf_x - cf_{LL}}{f_w}$

   $X = 6$ $\quad$ $cf_x = ?$
   $X_{LL} = 5.5$ $\quad$ $cf_{LL} = 26$
   $i = 3$ $\quad$ $f_w = 25$

   Substituting:

   $$\frac{6 - 5.5}{3} = \frac{cf_x - 26}{25}$$

   $4.17 = cf_x - 26$

   $cf_x = 30.17$

5. $P.R. = \frac{30.17}{100} \times 100$

   $P.R. = 30.17$

   Answer: P_{30}

An illustration for this one would look like:

X		cf (number "beaten")
8.5	is associated with	51 people
6.0	is associated with	?
5.5	is associated with	26 people

Since the score of "6" is closer to the lower, than to the upper, limit of the interval, the interpolation process will show its cf to be closer to the lower limit's cf of 26. If you understand this, you have come a long way. Take a two-minute water break as a reward (more exciting things may be promised later).

More specifically, this time by interpolating, we found that a score of 6 was 1/6 of the distance between 5.5 and 8.5. We therefore took 1/6 of the scores in that interval (that is 1/6 of 25) and added them to the <u>cf</u> for the lower limit (that is, 26). The result was $cf_x = 30.\overline{1}7$.

2.2 POSTTEST (answers in back)

For each of the following sets of data, compute percentile ranks for the raw scores indicated.

Set 1

X	f	cf
50	8	38
49	9	30
48	7	21
47	4	14
46	3	10
45	3	7
44	2	4
43	1	2
42	1	1

a) score of 49
b) score of 47
c) score of 44

Set 2

X	f	cf
18-20	2	25
15-17	4	23
12-14	8	19
9-11	3	11
6-8	3	8
3-5	4	5
0-2	1	1

a) score of 17
b) score of 9
c) score of 7

2.3 Instructional Unit: Converting Percentile Ranks into Raw Scores

In the preceding unit, we learned how to convert raw scores into percentile ranks. That is, if a student told us what score he earned on a test, we would be able to use the frequency data for that test to compute his percentile rank (and thereby, tell him the percentage of classmates that he "beat"). But suppose that, after a test or measure has been administered, we are asked to indicate which scores would correspond to various percentile ranks, such as, P_{25}, P_{47}, P_{66}, P_{90}, etc. For example, a friend (or even an enemy) might not know his score on an exam, but be interested in knowing what score he would need in order to surpass two-thirds of the students in his class. In other words, he is asking, "What score corresponds to P_{67}?" If we had the frequency data for his examination, finding the raw score equivalent for a given percentile rank is fairly easy. In fact, it simply involves the "opposite" process from finding the percentile rank for a given raw score. Suppose the test results are as follows:

X	f	cf
10	5	50
9	3	45
8	7	42
7	8	35
6	4	27
5	8	23
4	6	15
3	4	9
2	3	5
1	2	2

We are interested, remember, in determining the raw score equivalent for P_{67}.

1. First, put down the only basic formula you know of that relates percentile ranks to scores in the distribution: P.R. = $cf_x/N \times 100$. Then put down down the interpolation formula: $\frac{X - X_{LL}}{i} = \frac{cf_x - cf_{LL}}{f_w}$

   Keep in mind that we are looking for the value of "X" this time. In the last unit we were trying to compute the value of P.R. Since we are looking for X, the interpolation formula seems like a good place to start--it includes X in the equation, and if we can substitute values for all of the other variables, finding X should be a relatively simple endeavor.

2. Although we would like to start substituting values in the interpolation equation, we really cannot get very far. We don't know the value of X and therefore cannot possibly determine values for X_{LL}, cf_{LL}, f_w, etc. It appears that we are at a standstill, but there is something we can do which will enable things to fall into place; we can determine the value of cf_x. How? By use of the original equation: P.R. = $cf_x/N \times 100$;

   so, we substitute: P.R. = 67 (that was given)
   N = 50 (the number of scores)

   therefore: $67/100 = cf_x/50$

   $.67 = cf_x/50$

   $33.5 = cf_x$

   for convenience (this time only) we will round off and say that $cf_x = 34$.

3. Now we are ready to try the interpolation equation again. If we can substitute values for all variables, we can easily compute X.

$$\frac{X - X_{LL}}{i} = \frac{cf_x - cf_{LL}}{f_w}$$

X = ? (This is what we are looking for.)

cf_x = 34 We determine this in step #2; we now know that X is a score that has a cumulative frequency of 34--looking at the distribution we can further conclude that X will fall within the interval 6.5 (cf=27) to 7.5 (cf=35). It appears that it will be closer to 7.5, but how close?

X_{LL} = 6.5 How do we know this? Since cf_x = 34, X must fall between 6.5 - 7.5. The lower limit of X's interval is therefore 6.5.

i = 1 The score interval associated with X (6.5-7.5) is 1 unit in width.

cf_{LL} = 27 The cumulative frequency of the lower limit of the interval (6.5) is 27.

f_w = 8 The frequency of scores in the interval, 6.5-7.5, is 8.

It looks like we got everything. Now let's find X:

$$\frac{X - 6.5}{1} = \frac{34 - 27}{8}$$

$$X - 6.5 = 7/8$$

$$X = 6.5 + 7/8$$

$$X = 7.38$$

4. Our answer: P_{67} = 7.38

Example Problem: Using the same frequency distribution find P_{25}

We are now looking for the score associated with a P.R. of 25:

1. Put down the basic P.R. equation; put down the interpolation equation.

2. Find cf_x by use of basic P.R. equation.

P.R. = cf_x/N x 100

25 = cf_x/50 x 100

ERIC

$.25 = cf_x/50$

$cf_x = 12.5$ (this time we won't round off)

3. Use interpolation equation to find X.

$$\frac{X - X_{LL}}{i} = \frac{cf_x - cf_{LL}}{f_w}$$

$X = ?$

$cf_x = 12.5$ (see step #2)

$X_{LL} = 3.5$ (if $cf_x = 12.5$, X must be a score between 3.5 and 4.5)

$i = 1$ (the interval, 3.5 - 4.5, is 1 unit in width)

$cf_{LL} = 9$ (the lower limit, 3.5, has a cf of 9)

$f_w = 6$ (6 scores fall between 3.5 - 4.5)

substituting and working through:

$$\frac{X - 3.5}{1} = \frac{12.5 - 9}{6}$$

$$X - 3.5 = \frac{3.5}{6}$$

$$X = \underline{4.08}$$

4. Our answer: $P_{25} = \underline{4.08}$

<u>Example Problem:</u> The procedures are identical when we are dealing with grouped data.

Given the following data, find P_{80}.

X	f	cf
75-79	2	65
70-74	9	63
65-69	15	54
60-64	11	39
55-59	10	28
50-54	7	18
45-49	8	11
40-44	3	3

1. Write down the two equations you will need.

2. Find cf_x.

$P.R. = cf_x/N \times 100$

$.80 = cf_x/65$

$cf_x = 52$

3. Use interpolation equation to find X.

$$\frac{X - X_{LL}}{i} = \frac{cf_x - cf_{LL}}{f_w}$$

$X = ?$

$cf_x = 52$

$X_{LL} = 64.5$ (if $cf_x = 52$, X must be a score between 64.5 - 69.5)

$i = 5$ (the interval 64.5 - 69.5 is 5 units in width)

$cf_{LL} = 39$ (the lower limit, 64.5, has a cf of 39)

$f_w = 15$ (15 scores fall between 64.5 - 69.5)

substituting and working through:

$$\frac{X - 64.5}{5} = \frac{52 - 39}{15}$$

$X = \underline{68.83}$

4. Our answer: $P_{80} = \underline{68.83}$

ERIC

2.3 POSTTEST (answers in back)

For the following sets of data find P_{30}, P_{60}, and P_{90}.

Set 1

X	f	cf
48	1	40
47	1	39
46	3	38
45	4	35
44	10	31
43	8	21
42	6	13
41	5	7
40	2	2

Set 2

X	f	cf
18–20	4	35
15–17	5	31
12–14	6	26
9–11	10	20
6–8	4	10
3–5	3	6
0–2	3	3

2.4 Instructional Unit: Central Tendency

Percentile ranks give us an indication of how an individual performed relative to others. Aside from being interested in individual performance, however, we are usually concerned with how the group did as a whole. Frequency distributions, of course, do supply us with information concerning group performances, but this information is difficult to communicate to our readers. A good way to precisely summarize group achievement would be to determine the score that is most typical or representative of all scores in our frequency distribution. We call these typical or representative scores measures of central tendency.

A measure of central tendency is one of several types of averages. An average is a score that is typical or representative of a group of scores. Three of the most commonly employed central tendency measures are the mode (most frequently occurring score), the median (the middle score), and the mean (the "average" score). Most importantly though - any measure of central tendency is supposed to indicate a "representative" score value for the group being evaluated.

2.4 POSTTEST (answer is provided above - see if you can define it in your own words)

Define Central Tendency ______________________________

____

____

____

ERIC

2.4.1 Instructional Unit: Computation of Means

The most commonly employed measure of central tendency is the mean - symbolized $\overline{X}$. It is obtained by adding together all the numerical values and dividing by N, the number of values. You have probably used this procedure many times in computing your average exam score in high school and college classes. By definition, the mean is the number where the deviations above it equal the deviations below it. That is, if you subtracted $\overline{X}$ from every number above and below $\overline{X}$ and added the deviations together - the two sums of deviations would be equal.

The basic formula for the mean is:

$\overline{X} = \frac{\Sigma X}{N}$ (if you are uncertain about "summation" notation, check back to review on arithmetic operations)

All the formula means is: Add up all your scores; then divide by the number of scores.

Example Problem: Given the following scores, determine the mean.

Scores: 10, 8, 7, 6, 11, 20, 5

$\overline{X} = \frac{\Sigma X}{N}$ $\quad \Sigma X = 67$

$N = 7$

$\overline{X} = 67/7$

$\overline{X} = 9.57$ (two decimal accuracy is sufficient)

Example Problem: Compute the mean of the following scores:

X	f	cf
25	13	30
24	0	17
23	6	17
22	3	11
21	5	8
20	2	3
19	1	1

ERIC

Again, the basic formula is $\frac{\Sigma X}{N} = \bar{X}$

All you have to do is add the scores and divide by N (which you know to be 30). No tricks are involved, but a very common mistake would be to start adding the scores as follows: 19 + 20 + 21 + 22 + 23, etc. Why would this be a mistake?--after all, we are adding the scores to find ΣX. Maybe an important clue would be provided if it was pointed out that the proper adding procedure would be something like this: 19 + 20 + 20 + 21 + 21 + 21 + 21 + 21 + 22, etc. Now, you probably see the light. The "X" column indicates (as always) the possible scores, but the "f" column tells us how many of those scores were actually obtained. It looks as though we will end up with a rather long addition problem in order to determine the correct value of ΣX. Fortunately, however, there is an easier procedure which is clearly appropriate when we have more than one score of the same value. The procedure (which you should definitely learn and try to understand) involves:

1. Multiplying f times the midpoint of each interval. In the above ungrouped distribution, the midpoints are simply the values listed: 19, 20, 21, 22, etc. (remember, the value 22 is the midpoint of the interval 21.5 - 22.5). Multiplying X x f accounts for <u>all</u> the numerical values in the interval. Thus, for each score point, we will end up with an additional column indicating the value of fX.

2. Sum up all the fX's and divide by N. The result will be your mean.

3. Thus, whenever we have a frequency distribution, a more manageable formula for computing the mean is: $\bar{X} = \frac{\Sigma fX}{N}$

Let's work through the formula step-by-step.
Our frequency distribution:

X	f	cf	fX
25	13	30	325
24	0	17	0
23	6	17	138
22	3	11	66
21	5	8	105
20	2	3	40
19	1	1	19
		ΣfX =	693

(note that to facilitate matters, we have added an "fX" column, which shows the products of the various scores times their frequencies. At the bottom, we add all the fX's resulting in ΣfX.)

Now all we do is work the basic formula $\bar{X} = \frac{\Sigma fX}{N}$

$$\bar{X} = 693/30$$

$$\bar{X} = 23.1$$

Note: the formula $\frac{\Sigma fX}{N}$ is really the same as $\frac{\Sigma X}{N}$. When we have many values of the same score, $\frac{\Sigma fX}{N}$ facilitates the computation process.

<u>Example Problem:</u> Compute the mean of the following data:

X	f	X'	fX'
24-26	1	25	25
21-23	2	22	44
18-20	8	19	152
15-17	4	16	64
12-14	6	13	78
9-11	3	10	30
6-8	1	7	7

$$\Sigma fX' = 400$$

$$\bar{X} = \frac{\Sigma fX'}{N}$$

$$\bar{X} = 400/25 = \underline{16}$$

The above problem may appear, at first, to be somewhat complicated, but the procedures are really quite simple. Since we are dealing with a frequency distribution, the formula needed for computing the mean is $\frac{\Sigma fX}{N}$. It would take too long if we decided to add every score individually, so we multiply each score by its frequency, and add the fX's to yield our ΣfX (or ΣX).

The only difference between the above problem, and the previous problem, is that the present data are grouped. Thus, we need a "representative" score for each interval. The most representative score anyone could choose would naturally be the interval midpoint. To avoid confusion, the interval midpoints have been labeled as X' and included on the table.

In summary, in a grouped frequency distribution, the midpoint of the interval (symbolized X') becomes the X score which, in turn, must be multiplied by the corresponding f value. Since the f value indicates how many numbers are in the interval, multiplying f times the midpoint (X') accounts for all numerical values in the interval. Therefore, the formula for $\bar{X}$ using grouped data would be:

$$\bar{X} = \frac{\Sigma fX'}{N} = 400/25 = 16$$

This procedure is really identical to that employed in the previous problems--the only change involves locating the midpoint of each interval. Note that the midpoint of the score interval 15-17 is listed as 16. This makes sense since 16 falls exactly halfway between 14.5 - 17.5, the real limits of the interval.

Concluding note: The important thing is that regardless of whether ungrouped or grouped data are being used, all numbers (all f) must be accounted for. Either you add them up individually (many times an impractical endeavor) or multiply "X" by the appropriate f value and then add up the fX's.

2.4.1 POSTTEST (answers in back)

1. Define Mean ______________________________

____

____

2. For each of the following sets of data, compute the mean. Round off to two decimal places if necessary. (cf has been left out of the tables, since it is really not needed--you will need to add f's to compute N, however)

SET 1		SET 2		SET 3	
X	f	X	f	X	f
18-20	10	95	1	10	3
15-17	12	93	1	9	4
12-14	4	91	1	8	7
9-11	3	87	1	7	5
6-8	1	84	1	6	2
		76	1	5	1
		59	1		
		45	1		

2.4.2 Instructional Unit: Computation of the Median

The median, another measure of central tendency, is the number that corresponds to the middle frequency (that is, the middle score) in a ranked set of data. The median is the value that divides your distribution in half; half of your scores will be higher than the median, and half will be lower than the median.

```
          X          f

         High

                    50%
Median-------------------

                    50%
          Low
```

It is important to remember that the median is the halfway point in the distribution--in terms of frequencies (it corresponds to N/2). For example if N=40 (meaning that you have 40 scores), the median will be your 20th score (in terms of rank); if N=17, the median will be your 8.5 highest score, etc.

Another way of defining the median is to say that it corresponds to P_{50}.

In any distribution, the median will always be the score that corresponds to a percentile rank of 50; it is higher than 1/2 the scores, and lower than 1/2 the scores. If you have mastered Unit 2.3 (converting percentile ranks into raw scores) then computing medians cannot possibly present any difficulty--all you do, given a group of scores or a frequency distribution, is find the raw score for P_{50}. The procedures you mastered in Unit 2.3 are fully applicable for computing medians.

In order to compute medians, use the following guidelines:

1. If you have an odd number of scores, with no repeats (all f's = 1), the P_{50} interpolation formula is not needed. All that is required to find the median is to rank order your scores from lowest to highest; the middle score will be your median.

   <u>Example Problem</u>: Find the median (*Md*) in the following set of data: 26, 24, 23, 28, 31.

   Rank order the scores: <u>X</u>

   31
   28
   26
   24
   23

   What is the middle score?

   <u>Ans</u>. *Md* = 26

2. If you have an even number of scores, with no repeats, the *Md* is 1/2 the distance between the <u>two</u> middle scores. Thus, without worrying about interpolation, rank order the scores from low to high, find the two middle ones, and then determine the point halfway between them. The latter step simply involves averaging the two.

   <u>Example Problems</u>: Find the *Md* in each of the following sets of data:

   SET A: 13, 92, 68, 79, 81, 27

   SET B: 205, 198, 146, 261, 190, 105, 149, 269

   Rank order the scores:

<u>SET A</u>	<u>SET B</u>
92	269
81	261
79	205
68	198
27	190
13	149
	146
	105

   What are the two middle scores?

Ans. 68 and 79 in SET A; and 190 and 198 in SET B

$Md = \frac{68 + 79}{2}$ in SET A; and $\frac{190 + 198}{2}$ in SET B

Md = 73.5 in SET A; and 194 in SET B

3. If you have a large set of data, and/or one in which frequencies are greater than 1, do not use the "short-cut" procedures described above. What you should do is treat the median as P_{50} (which it always is) and follow the identical procedures described in Unit 2.3 for converting percentile ranks into raw scores.

Example Problem: Compute the median for the following set of data.

X	f	cf
56	3	20
55	2	17
54	1	15
53	2	14
52	5	12
51	4	7
50	3	3

We are looking for the median - that is, P_{50}

a. Put down the basic P.R. equation; put down the interpolation equation

b. Find the cf_x associated with P_{50} by use of the P.R. equation

$P.R. = cf_x/N \times 100$

$50 = cf_x/20 \times 100$

$.50 = cf_x/20$

$cf_x = 10$ (We have now determined that the median will be the 10th score in the distribution)

c. Use the interpolation equation to determine the median (X)

$$\frac{X - X_{LL}}{i} = \frac{cf_x - cf_{LL}}{f_w}$$

$X = ?$

$cf_x = 10$ (see step b)

ERIC

X_{LL} = 51.5 (if cf_x = 10, X must be a score between 51.5 - 52.5)

i = 1 (the interval 51.5 - 52.5 is 1 unit in width)

cf_{LL} = 7 (the lower limit 51.5 has a cf of 7)

f_w = 5 (5 scores fall between 51.5 - 52.5)

substituting and working through:

$$\frac{\text{Median (or X)} - 51.5}{1} = \frac{10 - 7}{5}$$

Median - 51.5 = 3/5

Median = 52.1

d. Our answer: P_{50} (or Median) = 52.1

The score 52.1 is the dividing point in our distribution.

Example Problem: Compute the median for the following set of data (this problem will be worked without as much description; if you are having trouble with procedures, reread Unit 2.3)

X	f	cf
72-74	6	30
69-71	10	24
66-68	5	14
63-65	6	9
60-62	3	3

cf_x associated with median = 15 (P.R. = cf_x/N x 100
50 = cf_x/30 x 100)

$$\frac{X - X_{LL}}{i} = \frac{cf_x - cf_{LL}}{f_w}$$

$$\frac{X - 68.5}{3} = \frac{15 - 14}{10} = \frac{X - 68.5}{3} = \frac{1}{10}$$

X - 68.5 = .3

X = 68.8

Answer: Median = 68.8

2.4.2 POSTTEST (answers in back)

1. Define the median ______________________________

____

____

____

2. For each of the following sets of data, compute the median.

Set 1	Set 2	Set 3			Set 4		
X	X	X	f	cf	X	f	cf
12	68	12	2	17	30–32	2	24
10	65	11	4	15	27–29	7	22
9	62	10	3	11	24–26	6	15
8	60	9	6	8	21–23	4	9
7	59	8	1	2	18–20	2	5
	54	7	1	1	15–17	2	3
					12–14	1	1

ERIC

2.4.3 Instructional Unit: Determination of the Mode

The third and last measure of central tendency you will be asked to learn is the mode. The mode is the number that occurs most frequently in a set of data. There is no formula for the mode—it is obtained by simple inspection. If there is more than one number that has the highest frequency (i.e., the highest f value is shared by 2 or more X values), then there are 2 modes; i.e., the distribution is bimodal.

Example Problem: Identify the mode in the following set of data

X
101
101
100
100
100
98
89

the mode is 100—since 100 occurs (three times) more than any other number. In a sense, 100 is the most "popular" score.

Example Problem: Identify the mode in the following set of data

X	f
10	3
9	4
8	7
7	6
6	7
5	4
4	2

Here, both 6 and 8 are modes since they occur seven times—the highest frequency values. "6" and "8" were the most "popular" scores.

Example Problem: Identify the mode in the following set of data

X	f
20–22	9
17–19	12
14–16	19
11–13	18
8–10	24
5–7	16
2–4	7

In a grouped frequency distribution, look for the interval with the highest f value and then use the midpoint of that interval.

In the distribution on the left, the modal interval is 8–10 (7.5 – 10.5) since it has the highest f value (24)

Therefore, the mode is 9 (the interval midpoint)

2.4.3 POSTTEST (answers in back)

1. Define the mode ______________________

____

____

2. Identify the mode in each of the following sets of data.

Set 1 X	Set 1 f	Set 2 X	Set 2 f	Set 3 X
64	1	40-42	14	7
63	3	37-39	6	7
62	4	34-36	3	7
61	2	31-33	2	6
60	6	28-30	4	6
59	4	25-27	1	6
58	6			6
57	2			4
				4
				1

2.5 Instructional Unit: Comparisons between Means, Medians, and Modes in Skewed and Non-Skewed Distributions

Whenever possible, it is always desirable to report all three measures of central tendency since they provide different kinds of information:

1. the Mean is the score point at which the distribution balances.
2. the Median is the score point that bisects the area
3. the Mode is the score point with the highest frequency

In general, however, the mean provides the most useful measure of central tendency. This is because the mean has certain mathematical properties and can be used for a variety of further calculations. The mode is the least useful measure--it is not very sensitive to variations in your distribution. The median is usually not preferred relative to the mean, but there are two special cases where it should be used: The first is a case where the mean cannot be calculated; the second is a case where the median actually gives a better indication of central tendency.

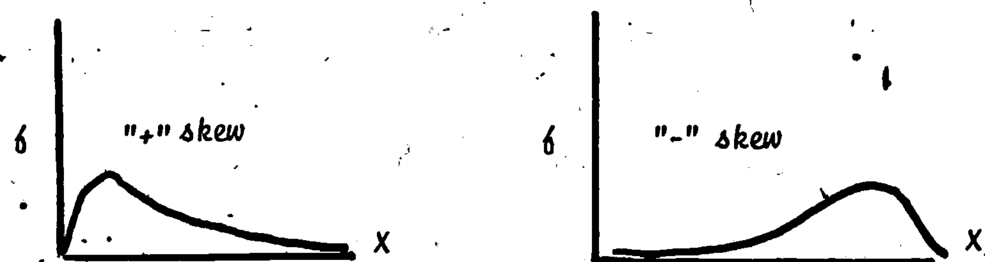
CASE 1 - Use the median if a score interval is open-ended

Salary	f
25,000-over	19
20,000-24,999	23
15,000-19,999	96
10,000-14,999	42
5,000- 9,999	31

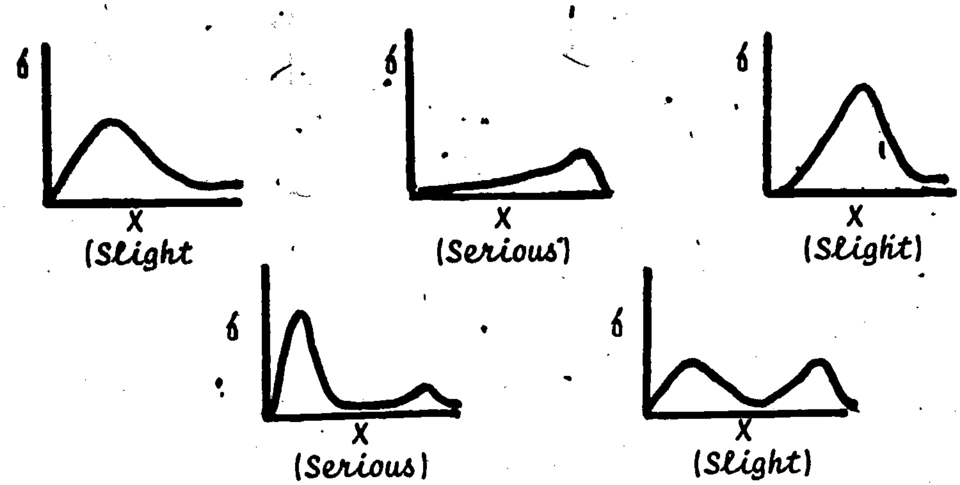
Ordinarily in a grouped distribution, the mean would be calculated by multiplying the midpoint of each interval (X') by the corresponding frequencies (f). However, in the above distribution showing the frequency of various salaries, the midpoint of the uppermost interval is unknown. Therefore, we cannot find "fX'" for that interval - the mean cannot be calculated. The median can, however, by use of the P.R. and interpolation equations.

CASE 2 - Use the median when the group of scores contains extreme values

What do we mean by extreme values? In simple terms, they would encompass scores that are located a considerable distance from where most of the scores seem to cluster. Remember the concept of skewness discussed in the final section of Unit I? To refresh your memory, we can have a symmetrical distribution where there is no skewness, a distribution with a positive skew, or a distribution with a negative skew. As can be seen in the frequency polygons below, the skew represents a "tail" of sorts, encompassing scores that are relatively far from the cluster at the opposite (high or low) end.



Thus, the rule is that when there is a serious skew, either positive or negative, a more representative measure of central tendency can probably be found in the median as opposed to the mean. In the same sense, given a listing of scores that contain some extreme values, the median is likewise the better choice. The frequency polygons shown on the next page give some idea of how serious, as opposed to slight, skewing would be represented.



The distinction between serious and minimal skews should be easy to memorize, but the serious (as opposed to the minimal) statistics student would probably want to know why extreme scores dictate the median. Here is an example: five men (or women) visit a family practice clinic to establish their eligibility for receiving health services. The questionnaire that they are asked to fill out contains a space for them to list their yearly family income. The amounts listed are as follows: \$10,000; \$12,000; \$13,000; \$15,000; and \$400,000 (a former politician who was convicted of accepting bribes and has written a best seller describing his despair). Which would be the "better" measure of central tendency, the mean or the median?

If we calculate the mean, we get: $\frac{\Sigma X}{N} = \frac{450,000}{5} = \$90,000$.

If we determine the median by selecting the middle score, we get: \$13,000.

Which would be the better measure for the clinic to base its estimates of "typical" financial status on: the mean or the median? The median clearly seems better in this instance. Why? The data show a serious positive skew (many "lows"; "few highs").

Here is another example. A teacher gives a math test to students and gets the following results:

X	f	cf
90–99	3	12
80–89	4	9
70–79	3	5
60–69	0	2
50–59	0	2
40–49	0	2
30–39	0	2
20–29	0	2
10–19	0	2
0–9	2	2

He is initially very pleased, since as he records the grades in his book he notes that nearly all of his 12 students did quite well. He expects a high class average (mean), but when he computes it by the formula, $\bar{X} = \frac{\Sigma fX'}{N}$ he gets $\frac{854}{12} = 71.17$, a very disappointing result as he had used his very best teaching techniques in presenting the material. In his despair, he considers the possibility of giving up teaching and enrolling in truck-driving school, but then he notices the two zero scores recorded for the test. This explains the snoring he heard during the exam period. As a former statistics student, he calculates the median through the formulas we have discussed:

$$P.R. = cf_x/N \times 100$$

$$.50 = cf_x/12$$

$$6 = cf_x$$

through interpolation:

$$\frac{Md - 79.5}{10} = \frac{6 - 5}{4}$$

$$Md - 79.5 = 1/4 \times 10$$

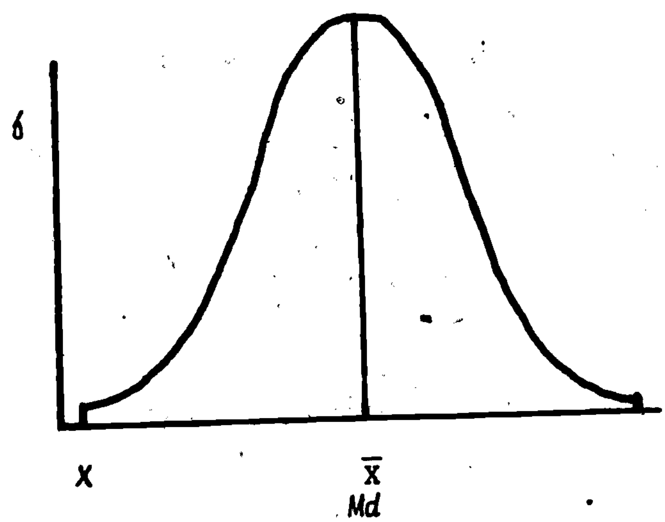
$$Md = 82$$

Due to the extreme scores, the median in this case ($Md = 82$) gives a better picture of the typical student performance than does the mean ($\bar{X} = 71.17$). Knowing this, the math teacher rests easy and even considers signing up to teach summer school.

Before leaving this section, we should point out how the mode may often be the most misleading central tendency statistic of all. Suppose we wish to assess how much money students carry around with them. We question the first five students that we meet and find that they have, respectively: \$.05 (a measly nickel), \$.05, \$2.00, \$3.85, and \$4.05. In this instance, both the mean and median would be \$2.00 (can you determine why?), while the mode would be 5¢, hardly "representative" but still the most frequent score for the group. See the problem?

Concluding Note: The mean is usually the best measure of central tendency to use except when the distribution is open-ended or has extreme scores (i.e., is seriously skewed).

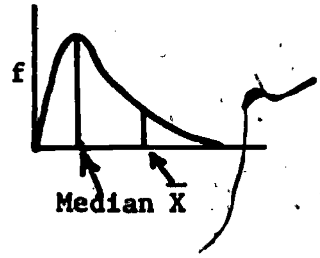
The skewness in a distribution affects whether the mean and the median are approximately the same or different in that distribution. If the frequencies bunch up toward the middle of the score scale or are fairly evenly spread out along the score scale, then there is little or zero skew. In a zero skew, the mean and median are the same.



Note: $\bar{X}$ and *Md* are both the same!

X	f	
6.3	2	
6.2	3	
6.1	4	
		← $\bar{X}$ = Median
6.0	4	
5.9	3	
5.8	2	

If there is a <u>positive</u> skew (frequencies bunch up at low end), the mean tends to shift toward the <u>trailing off</u> end (positive) of the distribution. However, the median remains relatively unchanged. Therefore, in a positively skewed distribution, the mean is <u>larger</u> than the median.



X

10
1
1
1
1

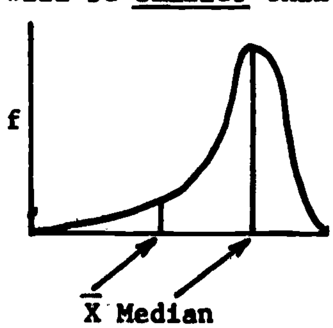
$\Sigma X = 14$

$\bar{X} = \frac{14}{5} = 2.8$

Median = 1 (approximate)

$\bar{X} >$ Median

If there is a <u>negative</u> skew (frequencies bunch up at high end), the mean tends again to shift towards the trailing off end (negative) of the distribution. In this case, the mean moves toward the lower end of the score scale and will be <u>smaller</u> than the median.



X

10
10
10
10
1

$\Sigma X = 41$

$\bar{X} = \frac{41}{5} = 8.2$

Median = 10 (approximate)

$\bar{X} <$ Median

Concluding Note: If a distribution is skewed, the mean is pulled toward the trailing off end--that is, the end <u>opposite</u> from where the bunching up occurs. Positive Skew: $\bar{X} >$ Median. Negative Skew: $\bar{X} <$ Median.

2.5 POSTTEST (answers in back)

1. For each of the following, indicate whether the <u>mean</u> or the <u>median</u> constitutes the more appropriate measure of central tendency.

A.

X	f
10	1
9	2
8	1
7	2
6	9
5	12
4	17

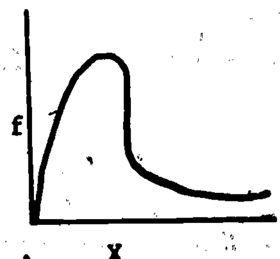
B.

X	f
30	1
27	2
24	3
21	5
18	4
15	2
12	1

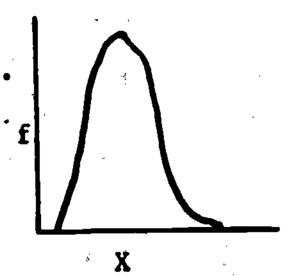
C.

X	f
50-up	7
40-49	8
30-39	11
20-29	8
10-19	7

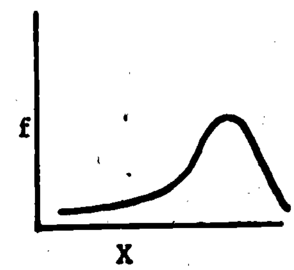
D.



E.

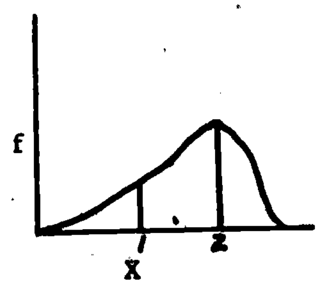


F.

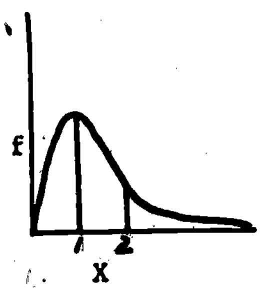


2. For each of the following distributions, select the number of the relative position of the mean and median.

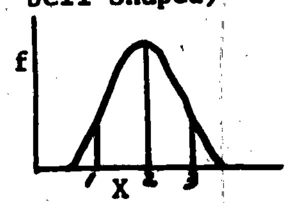
A.



B.



C. (Assume this is perfectly bell-shaped)



Mean # ________ # ________ # ________

Median # ________ # ________ # ________

D.

X	f
10	1
9	2
8	5
7	2
6	1

E.

X
30
30
30
29
4

F.

X
65
4
4
4
4

For "D," "E," and "F" (above), select one of the following answers: (a) $\bar{X} = Md$; (b) $\bar{X} > Md$; or (c) $\bar{X} < Md$.

Unit II Review Test
(answers in back)

1. Listed below are the scores obtained on a history test:

X	f	cf
90-94	21	50
85-89	16	29
80-84	10	13
75-79	2	3
70-74	0	1
65-69	1	1

____a. What is P_{80}?
____b. What is P_{25}?
____c. What is the P.R. for a score of 91?
____d. What is the P.R. for a score of 80?
____e. What is the mean?
____f. What is the median?
____g. What is the mode?
h. Which is probably the more meaningful measure of central tendency for this distribution, the mean or the median? Why?

____

____

2. For each of the following items, match the description to one of the following:

A. mean
B. median
C. mode

____a. the point of balance in a distribution.
____b. the most frequently obtained score.
____c. generally the most useful measure of central tendency.
____d. the halfway point in a distribution.
____e. generally the least reliable measure of central tendency.
____f. the numerical average.
____g. P_{50}

3. Answer each of the following with either A. mean or B. median.

____a. is preferable when the distribution is symmetrical (not seriously skewed).

____b. must be used when the distribution is open-ended.

____c. is larger when the distribution is positively skewed.

____d. is larger when the distribution is negatively skewed.

UNIT III

VARIABILITY

A. General Objectives

A measure of variability is a single number that expresses the extent to which scores in a set of data spread out or disperse around a measure of central tendency—usually the mean. Low variability indicates that most of the scores in a set of data bunch up close to the mean while large variability indicates that most of the scores are spread out a considerable distance from the mean. The present unit is designed to familiarize the student with three variability measures—the range, variance, and standard deviation—and the computational procedures for each.

B. Specific Objectives

3.1 Define variability
3.2 Define and compute the range
3.3 Define and compute the variance
3.4 Define and compute the standard deviation

3.1 Instructional Unit: Variability

In Units I and II you learned (or attempted to learn) how to characterize sets of scores in terms of a) the form of the distribution, and b) central tendency. If someone walked up to you in the street and asked questions about your data, you could reply (for example): "I ended up with a symmetrical, bell-shaped distribution in which the mean was 75, the median 75, and the mode 75." Obviously, the stranger would be highly impressed, and if he too had completed the first two Units, he could immediately formulate a pretty good "picture" of how your data looked. If, however, he had completed Unit III, his reaction to your reply might be: "O.K. buddy, but what about your variability—I am lost until you tell me that!" In a sense, his reaction would be entirely proper: in order to formulate an accurate picture of a set of data one really needs to know about form, central tendency, and variability.

Take, for example, the following two sets of data:

Set 1		Set 2	
X	f	X	f
52	3	100	3
51	4	75	4
50	5	50	5
49	4	25	4
48	3	0	3

Each set consists of 19 scores (N = 19). As far as form is concerned, each constitutes a bell-shaped, symmetrical distribution. As far as central tendency is concerned, the mean, mode, and median are equal to 50 in both sets. But are both sets of data the same even though they do not differ either in form or central tendency? The answer is obviously "no." In Set 1, the scores are bunched together, whereas in Set 2, the scores are spread apart. Thus, we can differentiate between the two sets of data in terms of their variability.

Definition: Variability refers to the extent to which scores in a distribution bunch up close to, or spread out far from, a measure of central tendency--usually the mean. Variability is expressed by a single number and, like central tendency, there are several measures of variability. The most common are the range, variance, and standard deviation. In any case though, a single number expresses how compacted the total set of numbers is around the mean. By compacted is meant "how spread out."

3.1 POSTTEST (answer in back)

Define variability ______________________________

____

____

____

3.2 Instructional Unit: The Range

The range is the simplest measure of variability to compute. However, it is also the most unreliable for reasons to be discussed later. To compute the range, all you have to do is obtain the distance from the lowest score in the distribution to the highest score in the distribution. More simply, subtract the lowest score from the highest score--the result is the range (this is almost as easy as computing modes).

Set 1

X	f
52	3
51	4
50	5
49	4
48	3

Range = highest score minus lowest

= 52.5 (upper limit of top score)
-47.5 (lower limit of bottom score)

5.0

Set 2

X	f
100	3
75	4
50	5
25	4
0	3

Range = 100.5 - (-.5)

Range = 101

*note - lower limit of 0 is -.5

Another Example: Determine the range in the following distribution:

X	f
50-55	3
44-49	2
38-43	3
32-37	8
26-31	1

Range = 55.5 (upper limit of top score)
-25.5 (lower limit of bottom score)

30.0

Final note: The range represents a total variability measure. That is, the range gives you a number that represents a distance where all frequencies above and below the mean have been accounted for. A large number indicates that scores spread out far around the mean while a small number indicates that scores are bunched up close to the mean. Note the differences in the value of the range between Data Sets 1 & 2 on the preceding page.

While the range is the easiest measure of variability to compute, it is also the least stable. Simply changing one number can radically change the range as the following example shows. Therefore, if one can pick the measure of variability, the range typically should not be selected. It is easy to compute, but can lead to considerable misinterpretation.

Example:

X		X
10		15
9		9
8		8
7	but if 10 is changed to 15:	7
6		6
5		5

Range = 10.5 - 4.5 = 6 Range = 15.5 - 4.5 = 11

(or nearly double the previous range)

3.2 POSTTEST (answers in back)

1. Define the range ____________________

____

____

2. For each of the following sets of data, compute the range

Set 1

X
87
83
81
80
68
68
45

Set 2

X	f
106	2
105	3
104	4
100	2
98	1
97	3
96	4

Set 3

X	f
20-24	6
15-19	4
10-14	8
5- 9	3
0- 4	2

3.3 Instructional Unit: The Variance

As indicated above (Unit 3.2), the range does not provide a particularly stable measure of variability since it is fully dependent on the values of only two scores—the highest and the lowest. A more suitable measure, it would seem, should take into account all scores in the distribution—and their relative distances from the mean. Such a measure is the variance.

The variance is a number that represents the average of squared deviations around the mean. The larger the variance is, the farther away scores tend to be (i.e., deviate further) from the mean. The smaller the variance, the closer scores tend to be (i.e., deviate little) from the mean. By subtracting the mean from each score in a distribution, squaring this deviation, and then taking the mean of these squared deviations—the variance is obtained.

Let's run through that again:

1. Subtract the mean from each score in the distribution.

2. Square each deviation score.

3. Compute their average.

We can express this whole operation by the following formula:

$$\sigma^2 = \frac{\Sigma x^2}{N}$$

where: σ^2 = symbol for variance, called sigma squared (σ is the lower case Greek letter "sigma").

x^2 = squared deviations from the mean (small x is used to express a deviation score).

N = number of cases or frequencies (as it always has).

Example Problem: Compute variance for following set of data.

X	f	x	x^2
24	1	4	16
22	1	2	4
20	1	0	0
18	1	-2	4
16	1	-4	16

$\Sigma X = 100$ $\qquad$ $\Sigma x^2 = 40$

$\bar{X} = 20$

$\sigma^2 = \frac{40}{5} = 8$ (which is simply the mean of the x^2 column of numbers)

If you examine the above computational procedures, you should readily understand what we did. We were given our frequency distribution; (1) we computed the mean; (2) we subtracted the mean from each score in the distribution, and called the result "x"—for convenience we added an "x" column to the distribution; (3) we squared each x (or deviation score) to the distribution; (4) we summed all the x^2's and took their average by dividing by N—the result, 8, is the variance, defined as the average of the squared deviations.

It is important that you understand that the computation of the variance, unlike that for the range, takes into account every score in the distribution. Accordingly, it is a much more sensitive and reliable measure of variability.

In looking over the formula for the variance, you may question why it involves *squaring* each deviation (x) score. Perhaps, at first glance, it would seem a lot simpler and more sensible to just add the deviation scores and take their average. In other words, why bother to square?

The answer to the above question may be found in the definition of the mean (Unit 2.4.1). By definition, the mean is the point of balance in the distribution; the sum of deviation scores of all points *above* the mean is *always equal to* the sum of the deviation scores of all points *below* the mean. For example, if you subtracted the mean from all scores above the mean, and added those deviation scores together you might get a sum of +92. Therefore, it follows, that if you subtracted the mean from all scores below the mean in the same distribution, the sum of those scores would be -92.

If you are still disbelieving, go back to the example problem on the preceding page and add the unsquared deviation scores (i.e., the x's). You will find that for the two scores above the mean, $\Sigma x = +6$. For the two scores below the mean, $\Sigma x = -6$. When you add these two partial sums together the result is $+6\ (+)\ -6 = 0$. This will always be the case! Anytime, for any distribution you can imagine, the sum of the unsquared deviation scores (x's) will always be equal to zero. Obviously, then Σx would not be helpful in determining variability, or anything else for that matter. By *squaring* the x's we eliminate positive and negative signs, and thus will never get $\Sigma x^2 = 0$ (unless every single score was equal to the mean). If none of this has made any sense to you, add the x's (without squaring them) in all example problems, and you will prove to yourself that a "zero" result must always be the case.

Take some time now to learn the formula for the variance:

$$\sigma^2 = \frac{\Sigma x^2}{N}$$ Remember, little "x" is a deviation score!!

Using the above formula, it should be relatively easy to compute the variance. But suppose, you computed the mean of your distribution and found it to be a number such as 20.6. You could still use the above formula, but to determine Σx^2, it could get a little messy subtracting 20.6 from each score and squaring the result. You would obviously end up with x's with decimal points, and x^2's with decimals.

Fortunately, by simple algebraic manipulation, an alternative formula has been worked out. This formula should probably be used whenever the mean ($\overline{X}$) is not equal to a whole number. The formula is:

$$\sigma^2 = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N}$$

where X^2 = squared raw scores (not small x, or deviation scores). Actually the numerator of the above formula is equal to the numerator of the original formula, Σx^2. In both cases, you divide the numerator by N.

The alternative formula is called the raw score formula; it yields the same result as the original, but eliminates subtracting and squaring numbers with decimals.

Example Problem: Given the same data used in the preceding example problem find the variance by means of the raw score formula. (Since the mean = 20—a whole number—we would, in actuality, probably be content to stick with the original formula; but this is only for practice).

X	X^2
24	576
22	484
20	400
18	324
16	256

$\Sigma X = 100$ $\quad$ $\Sigma X^2 = 2{,}040$

$(\Sigma X)^2 = 10{,}000$

In order to use the raw score formula, we must find N, ΣX^2, and $(\Sigma X)^2$.

N = 5 (that is easy)

ΣX^2 = squaring each raw score and then adding them together (see column 2)

substituting: $\sigma^2 = \dfrac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N}$

$$\sigma^2 = \frac{2{,}040 - \frac{10{,}000}{5}}{5}$$

98

$$\sigma^2 = \frac{2{,}040 - 2{,}000}{5} = \frac{40}{5} = 8$$

$$\sigma^2 = 8$$

Note that this is the same answer we obtained when the original formula was used.

Two more examples will be given computing the variance using the raw score formula.

<u>Example Problem</u>: Compute the variance using the raw score formula (in this case the mean is not a whole number, so the raw score formula would be considerably easier to use than the original).

X	f	fX	X^2	fX^2
10	3	30	100	300
9	2	18	81	162
8	4	32	64	256
7	2	14	49	98
6	2	12	36	72
	N = 13	$\Sigma fX = 106$		$\Sigma fX^2 = 888$

$$\sigma^2 = \frac{\Sigma fX^2 - \frac{(\Sigma fX)^2}{N}}{N} = \frac{888 - \frac{(106)^2}{13}}{13} = \frac{888 - \frac{11{,}236}{13}}{13} =$$

$$\frac{888 - 864.31}{13} = \frac{23.69}{13} = 1.82$$

Notes: (1) Since there are 13 scores, not 5, all 13 must be added together and all must be squared; multiplying the X and the X^2 by f accomplishes this. Make sure you <u>do not</u> multiply (f)·(X) and <u>then</u> square, $(fX)^2$ is <u>not</u> correct.

(2) 1.82, the variance, represents the average of of a column of squared deviations from the mean.

<u>Example Problem</u>: Compute the variance for the grouped data shown below.

X	f	X'	fX'	X'^2	fX'^2
20-22	1	21	21	441	441
17-19	3	18	54	324	972
14-16	4	15	60	225	900
11-13	6	12	72	144	864
8-10	3	9	27	81	243
5- 7	2	6	12	36	72
2- 4	1	3	3	9	9
	N=20		$\Sigma fX' = 249$		$\Sigma fX'^2 = 3501$

$$\sigma^2 = \frac{\Sigma fX'^2 - \frac{(\Sigma fX')^2}{N}}{N} = \frac{3,501 - \frac{(249)^2}{20}}{20} = \frac{3,501 - \frac{62,001}{20}}{20} =$$

$$\frac{3,501 - 3,100}{20} = \frac{401}{20} = 20.05$$

Notes: (1) Again 20.05 is the average of a column of squared deviations around the mean.

(2) Notice that the X scores in this case are the midpoints (X') of the intervals since, given an interval, the midpoint is our best estimate of the "typical" score in that interval.

<u>Final Notes</u>:

A. Do not try to memorize all the different varieties of the formulas. It will be sufficient to simply learn the raw score formula and realize that in grouped type data, the scores—X or X'—have to be squared and multiplied by f. The important thing is that <u>all</u> frequencies be accounted for.

B. Remember, the variance is an average—actually a mean. It is the average of the <u>squared deviation scores around the original mean</u>.

C. The variance value is expressed in terms of the original raw scores except that it is in <u>squared</u> units. For example, if the original data is in inches, then the variance is expressed in <u>squared inches</u>. This is because the deviation scores are squared in computing the variance.

D. As will be seen later, the variance formula that was given assumes that you have all the data on hand, i.e., the population. If you have <u>sample data</u> and are trying to <u>estimate</u> what the variance would be <u>in the population</u>, then the numerator should be divided by (n-1) instead of (N). That is the only change in the basic formula. Thus, the sample variance, symbolized as s^2, is used to <u>estimate</u> the population variance. Small "n" is used to indicate that the observations (people) are part of a sample.

Later on in this text, we will describe the differences between samples and populations in considerable detail. For now, so that you at least have some understanding of what the "s^2/σ^2" distinction is about, here is a brief explanation. When measurements are taken, whether in the form of test scores, physical characteristics, or something else, they can encompass the entire group of interest (the population) or only part of that group (a sample). Populations are easy to assess when the members are small in number and easily accessible (such as a class of students, members of a football team, etc.). Otherwise, their assessment is extremely difficult, such as when the population of interest consists of all students in an entire school district, all college football players, or all living medical staff who worked for a particular hospital in 1957. In such instances, the statistician will usually be forced to work with a sample of individuals selected from the population. The sample is not interesting in itself, but rather for what it implies about the population. More specifically, it is used to estimate what the population is like when all its members are included. Thus, the variance statistic, s^2, is basically important as an estimate of the population "parameter" (a new term), σ^2. It has been found that, in the long run, using "n-1" in the denominator of the variance formula makes s^2 a better (more accurate) estimate of σ^2 than when "n" is used. And that is what the game of statistics is all about! The game continues below...

The formula for s^2 is:

$$s^2 = \frac{\Sigma x^2}{n-1} \text{ (original formula)} = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n-1} \text{ (raw score formula)}$$

Note that the only change from the formula we have been using is that the denominator is n-1. If you are asked to compute the variance and the problem states that we are dealing with sample data, use n-1 in the denominator. Otherwise, use N, as we have been doing in the preceding example problems.

The reason that the denominator becomes n-1 is that sample variance tends to be systematically smaller than the variance of the population from which the sample has been drawn. This makes sense because a sample is not likely to have the complete range of scores contained in the population. The denominator n-1 produces a somewhat larger variance and is called an unbiased estimate of the population variance. While the difference in the variances obtained by use of n and n-1 is rather trivial in large sized samples, it can be quite significant in smaller samples. Therefore, it is especially important to use the s^2 formula when sample size is small. To be safe, always use s^2 when you are dealing with sample data.

E. If a constant is added to, or subtracted from, a set of data, variability - the variance - will not be altered. Suppose for example, 2 points were added to every score in the distribution. Would the mean be changed? Obviously, it would; since every score is increased by 2, the mean would also become 2 points higher. But, would the spread of the scores (i.e., the variance) be altered? No - since every score would be similarly affected, they would remain just as close together as they were before the 2 points were added. Another way of understanding this is to think about the formula (or definition) for the variance: the average of the squared deviation scores. How do you get a deviation score? Answer: subtract the mean from each score. But - if the mean is increased by 2, and each score is increased by 2, would the deviation scores change from what they were before the points were added? Answer: an emphatic "no"!

Example Problem: We find $\sigma^2 = 25$ in a certain distribution. We then add 1,000 points to each score. What would the new variance be?

Answer: $\sigma^2 = 25$. No change.

F. If a set of data is multiplied or divided by a constant, the variance is affected by the square of that constant. Thus, for example, if all scores are doubled (multiplied by 2), the variance is 2^2 or 4 times as large. If all scores are divided by two, the variance becomes 1/4 as large; if all scores are multiplied by 3, the variance becomes 9 times as large; if all scores are divided by 10, the variance becomes 1/100 as large, etc., etc.

Let's examine how this occurs through the following example.

Example: A teacher gives a 50-item test to three students and obtains the following results.

	X	X^2
Student A	47	2209
Student B	42	1764
Student C	30	900
	119	4873

She computes the variance as follows:

$$\sigma^2 = \frac{4873 - \frac{(119)^2}{3}}{3} = \frac{4873 - 4720.33}{3}$$

$$\sigma^2 = 50.89$$

To make the results comparable to the 100-item tests she has previously given, she decides to multiply each student's score by 2. Thus, she gets:

	X	X^2
Student A	94	8836
Student B	84	7056
Student C	60	3600
	238	19492

She computes the variance:

$$\frac{19492 - \frac{(238)^2}{3}}{3} = \frac{19492 - 18881.33}{3}.$$

$$\sigma^2 = 203.56$$

Compare the original and new variances, and you will see that the latter is 4 times larger, as the rule predicts. For our next trick, we will make a posttest appear from what was previously a completely blank page. Watch below.

3.3 POSTTEST (answers in back)

1. Define the variance ______________________________

____

____

____

2. For each set of data below, compute the variance.

Set 1	Set 2		Set 3 (Sample Data!!)	
X	X	f	X	f
26	42	2	12-14	4
24	40	3	9-11	2
23	38	4	6- 8	1
22	36	4	3- 5	2
20	34	6	0- 2	3
19	32	3		
17	30	1		
15				

3. Suppose that in all three sets of data, each score was decreased by 10. What would the new variances be?

4. Suppose that in all three sets of data, each score was multiplied by 3. What would the new variances be?

3.4 Instructional Unit: The Standard Deviation

The third (and final) measure of variability you will be asked to learn is the standard deviation. If you were successful in learning how to compute the variance, the following should come as very good news: The standard deviation is simply the square root of the variance.

Thus, all you have to do is (1) compute the variance; (2) take its square root; and (3) you end up with the standard deviation. There is really not much new here to learn.

Recall that the symbol for the variance was σ^2 (or s^2 if it is an estimate of the population variance) - so the standard deviation is symbolized as σ (or s if you are estimating the population standard deviation).

Therefore: basic formula raw score formula

$$\sigma = \sqrt{\frac{\Sigma x^2}{N}} \qquad \sigma = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N}}$$

Example Problem: Suppose we find that for a given population, the variance is equal to 64. What is the standard deviation?

Answer: $\sigma = 8$ (simply the square root of the variance)

Example Problem: Compute the standard deviation for the following data.

X	x	x^2	X^2
5	2	4	25
4	1	1	16
3	0	0	9
2	-1	1	4
1	-2	4	1
$\Sigma X=15$		$\Sigma x^2=10$	$\Sigma X^2=55$
$\bar{X}=3$			

Note: little "x" is a deviation score!!

104

Using the basic formula:

$$\sigma = \sqrt{\frac{\Sigma x^2}{N}} = \sqrt{\frac{10}{5}} = 1.41$$

Using the raw score formula:

$$\sigma = \sqrt{\frac{55 - \frac{(15)^2}{5}}{5}} = \sqrt{\frac{55 - \frac{225}{5}}{5}} = \sqrt{\frac{55 - 45}{5}} = \sqrt{\frac{10}{5}} = \sqrt{2} = 1.41$$

Example Problem: Using the raw score formula, compute the standard deviation for the following grouped data.

X	f	X'	fX'	X'^{2}	fX'^{2}
14-16	2	15	30	225	450
11-13	3	12	36	144	432
8-10	3	9	27	81	243
5- 7	2	6	12	36	72
2- 4	1	3	3	9	9

$N = 11$ $\Sigma fX' = 108$ $\Sigma fX'^2 = 1206$

$\Sigma X = \Sigma fX'$ $\Sigma X^2 = \Sigma fX'^2$

$$\sigma = \sqrt{\frac{1{,}206 - \frac{(108)^2}{11}}{11}} = \sqrt{\frac{1{,}206 - \frac{11{,}664}{11}}{11}} = \sqrt{\frac{1{,}206 - 1{,}060.36}{11}} = \sqrt{\frac{145.64}{11}}$$

$$\sqrt{13.24} = 3.64$$

Like the variance, σ is a measure of variability. The larger the σ, the more spreadout the scores are from the mean. The smaller the σ, the more bunched up the numbers are relative to the mean. Also, like the variance, the standard deviation could be 0 if all numbers in a set of data are the same (no variability). The standard deviation can be a positive number, if there is some variability; however, it can never be a negative number.

One useful property of the standard deviation is that—assuming the set of data resembles the general pattern of

ERIC

a bell-shaped curve (or normal distribution), certain percentages of N (the total amount of scores) lie <u>within</u> certain distances or standard deviation units from the mean.

For instance:

a) 68% of the scores fall within $\pm$ 1 standard deviation from the mean. (The 68% is approximate.)

b) 95% of N falls within $\pm$ 2 standard deviations from the mean.

c) over 99% of N falls within $\pm$ 3 standard deviations from the mean.

Thus, for example, if we have a bell-shaped distribution with a mean of 60 ($\bar{X} = 60$) and a standard deviation of 5 ($\sigma = 5$), we can conclude that approximately 68% of the scores fall between 65 (one σ above the mean) and 55 (one σ below the mean); and that approximately 95% of the scores fall between 70 (2 σ above the mean) and 50 (2 σ below the mean). Don't worry too much about this now (if you do not understand it); worry about it when you get to Unit IV.

In any case by knowing $\bar{X}$ and σ we can construct an interval one σ wide around $\bar{X}$ such that 68% or about two-thirds of the scores will fall within that interval. In distributions that have a <u>small</u> standard deviation, the interval around $\bar{X}$ is short since the scores bunch up close to $\bar{X}$. In distributions where the standard deviation is large, the interval around $\bar{X}$ is larger since scores spread out farther from the mean. In either case, the percentage of the total N (68%) is the <u>same</u>. What is different is <u>how far</u> you have to go out from $\bar{X}$ to obtain the 68%: if the $\sigma = 3$, you have to go <u>6</u> points; if $\sigma = 5$, you have to go out <u>10</u> points, etc. The larger the standard deviation, the farther out you have to go to include two-thirds or 68% of the scores.

<u>Final Notes</u>:

A. The standard deviation is the square root of the variance.

B. The raw score formula is usually the best way to compute the standard deviation, especially if the mean <u>is not</u> equal to a whole number.

C. The standard deviation value is expressed in exactly the same terms as the original data. If the raw data is IQ scores, σ <u>is</u> expressed in IQ units; if $\sigma = 15$ for IQ scores, the 15 means 15 points. In contrast, you will recall, the variance would be expressed in <u>squared</u> IQ units. Thus, the value of the variance is not directly comparable to the original raw score units, but the value of the standard deviation can be directly compared.

D. If you are estimating the population standard deviation, the denominator should be n-1 rather than N. Therefore, s, or the estimation of the population is:

$$\underline{s} = \sqrt{\frac{\Sigma x^2}{n-1}} = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n-1}}$$

E. Like the variance, adding or subtracting a constant from a set of data has no effect on variability—i.e., the standard deviation. However, if a set of data is multiplied or divided by a constant, the standard deviation is affected by the amount of the constant. For example, if the constant 4 is added to every score, the standard deviation remains unchanged; but if, every number is multiplied by 4, the standard deviation will be 4 times its original value.

3.4 POSTTEST (answers in back)

1. Define the standard deviation ______________________

____

____

____

2. For each set of data below, compute the standard deviation.

Set 1

X
17
14
13
9
8
3
2

(Sample Data)

Set 2

X	f
10	2
9	3
8	7
7	6
6	4
5	2

Set 3

X	f
21-23	6
18-20	7
15-17	4
12-14	3
9-11	4
6- 8	2

3. Suppose that in all three sets of data, each score was increased by 3. What would the new standard deviations be?

Unit III Review Test
(answers in back)

Consider the following frequency distribution:

X	f
20	1
19	3
18	5
17	4
16	4
15	3
14	2
13	1
12	2
11	0
10	1

Answer the following:

_____ 1. ΣfX

_____ 2. N

_____ 3. $\bar{X}$

_____ 4. ΣfX^2

_____ 5. $(\Sigma fX)^2$

_____ 6. σ^2

_____ 7. $\underline{s}^2$

_____ 8. σ

_____ 9. $\underline{s}$

_____ 10. the range

_____ 11. σ if $\underline{1}$ point is subtracted from every score

_____ 12. σ if every score is divided by $\underline{2}$

_____ 13. σ^2 if every score is divided by $\underline{2}$

_____ 14. $\underline{s}$ if every score is multiplied by $\underline{6}$

_____ 15. $\underline{s}^2$ if every score is multiplied by $\underline{3}$

_____ 16. $\underline{s}^2$ if $\underline{2}$ points are added to every score

ERIC

UNIT IV

POSITION MEASURES AND THE NORMAL CURVE

A. General Objectives

The purpose of this section is to illustrate a basic procedure used to describe the location of *specific scores* within a distribution of data. Measures of central tendency indicate the "typical" value; variability measures indicate the degree of "spread" in the distribution; but *position measures* locate specific points within the data set.

B. Specific Objectives

4.1 Define and compute z scores

4.2 Convert specified z scores into their appropriate raw scores

4.3 Convert raw scores and z scores into T scores

4.4 Describe and make a sketch of the normal curve

4.5 Show the raw score and z score scales on a normal curve diagram

4.6 On a normal curve, label the approximate percentage of scores between $\bar{X}$ and + or - 1 standard deviation, 2 standard deviations, and 3 standard deviations

4.7 Using the "area under the normal curve" table, find the percentage of scores between any two specified z score values

4.8 Using the "area under the normal curve" table, find the percentage of scores between two specified raw scores

4.9 Using the "area under the normal curve" table, convert specified z scores to percentile ranks, and specified percentile ranks into z scores

4.1 Instructional Unit: *z* Scores

Percentile ranks, as discussed in Unit II, do provide a measure of *specific score position*, but cannot indicate unambiguously how much better one person's score is relative to another's. For example, if Bill scored an 87 on a test, the highest score in the class, his percentile rank would be close to P_{100}. His friend Ernie scores 77 which is determined to be P_{80}. Bill has thus scored 20 percentile points higher than Ernie. But suppose an error was made in grading Bill's paper, and his actual score is found to be 97, as opposed to the original 87. Since 97 would obviously be the highest score in the class, Bill's percentile rank would still be P_{100}. In other words, even with 10 points added to his original test score, Bill is still only 20 percentile points higher than Ernie; changes in raw scores do not always produce changes in percentile ranks.

In light of the limitations of percentile ranks, a different position measure - *z* scores - will be introduced and described in this Unit. In addition, relationships among percentile ranks, *z* scores, and the normal curve will be discussed.

A *z* score, like a percentile rank, is a position measure. By position measure is meant a single number that represents a specific location point within a distribution of scores. It is not a distance measure like the standard deviation, or a "representative" number like the mean, mode, and median. Rather, it is a point measure. A percentile rank, for instance, is a specific position measure that states what percentage of scores fall below a certain point in the distribution. Another position measure, called a *z* score, indicates how many standard deviations away from the mean a certain score is. In either case though (percentile ranks or *z* scores), the number given indicates some point in the distribution. Given this "point," different scores in the distribution can be compared in terms of their location. This particular function of *z* scores should become more understandable as you work through this unit.

Definition: A *z* score is a number that indicates how many standard deviations a raw score is away from the mean.

The whole ballgame is in the above sentence. Read it again. So, if a person told you that his/her *z* score was equal to 1, would you know anything interesting? Yes! You would know that he/she scored one standard deviation away from the mean.

Positive z scores indicate that raw scores are a certain distance above the mean, whereas negative z scores indicate that raw scores are a certain number of standard deviations below the mean. A z score of 0 (zero) indicates that the raw score is exactly equal to the mean - in other words, no standard deviations, or part of a standard deviation, above or below the mean.

Read the above paragraph again. Now answer the following question: If a student tells you his/her z score is equal to +1, what does this mean?

Answer: that student scored one standard deviation above the mean.

Question: Suppose the standard deviation, σ, was computed to be 5, what does this mean?

Answer: The student scored 5 points (i.e., one standard deviation) above the mean.

Question: Suppose the mean on the test was 75; what would that indicate?

Answer: The student received a score of 80 points on the test: 80-75 (the mean) = 5. The standard deviation, you will recall, was 5 points. Therefore, z was +1.

Question: Suppose another student receives a z score = 0. What does that suggest given the above test results?

Answer: The student received a score of 75 on the test - because whenever z = 0, the raw score is exactly equal to the mean.

Hopefully, now you are ready for the z score formula. It is perfectly consistent with the logic used in deriving z scores and raw scores in the above example. If you didn't understand the example, go back to it once you feel comfortable with the z score formula.

$$z = \frac{X - \bar{X}}{\sigma}$$

In plain English, the formula says: 1) Subtract the mean ($\bar{X}$) from the raw score (X) for which z is to be computed. 2) Divide the result by the standard deviation (σ) of the distribution. 3) The final product is the z score for that particular raw score (X).

Example Problem: I give a test and find the mean ($\bar{X}$) to be 75 and the standard deviation to be 6. I am interested in computing a z score for a raw score of 69 received by one of the students taking the test. What is z for the raw score, 69? (or phrasing the same question in a slightly different manner) How many standard deviations away from the mean is the raw score, 69?

$$\underline{z} = \frac{X - \overline{X}}{\sigma} = \frac{69 - 75}{6} = \frac{-6}{6} = -1$$

$\underline{z} = -1.00$, indicating that the student scored one standard deviation <u>below</u> the mean.

<u>Example Problem:</u> In the same distribution as above, what would be the $\underline{z}$ score for a raw score of 75?

$$\underline{z} = \frac{75 - 75}{6} = \frac{0}{6} = 0.00$$

the student scored <u>no</u> deviations above or below the mean.

<u>Example Problem:</u> In the same distribution, what would be the $\underline{z}$ score for a raw score of 87?

$$\underline{z} = \frac{87 - 75}{6} = \frac{12}{6}$$

$$\underline{z} = +2.00$$

<u>Example Problem:</u> Assume $\overline{X} = 19.8$ and $\sigma = 3.4$. Find the $\underline{z}$ scores for raw scores of 22.4 and 14.9.

For 22.4: $\underline{z} = \frac{22.4 - 19.8}{3.4}$

$$\underline{z} = \frac{2.6}{3.4}$$

$$\underline{z} = +0.76$$

(in other words, the raw score 22.4, is .76 standard deviations above the mean)

For 14.9: $\underline{z} = \frac{14.9 - 19.8}{3.4}$

$$\underline{z} = \frac{-4.9}{3.4}$$

$$\underline{z} = -1.44$$

(in other words, the raw score 14.9, is 1.44 standard deviations below the mean)

<u>Final Notes:</u>

1. Make sure that you have the correct <u>sign</u> for the $\underline{z}$ score. A negative sign indicates that the score is below the mean, whereas a positive sign indicates that the score is above the mean.

ERIC

2. z scores are comparable from one distribution to another, regardless of the means and standard deviations. All a z score of 0.50 says is that, regardless of the size of σ, a certain score is .5 σ (half a standard deviation) above the mean. If it was determined that your score on a basic algebra test, for example, was equal to a z of 0.56, and that your score on a basketball free-throw shooting test was equal to a z of 0.56, we could say that you performed equally well on both. In each case you scored .56σ above the mean.

3. A z score is a precise position measure in that it locates a raw score within a distribution by indicating how many standard deviations away from the mean that score is.

4.1 POSTTEST (answers in back)

1. Define z score ____

____

____

2. For each of the following, compute the appropriate z scores for the specified raw scores.

A. $\overline{X} = .8$, $\sigma = .03$: $X = .91, .87, .80, .72$

B. $\overline{X} = 92$, $\sigma = 7$: $X = 100, 99, 90, 82, 78$

C. $\overline{X} = 0$, $\sigma = 1$: $X = 1.5, -2, 0, -.4$

4.2 Instructional Unit: Converting z scores into raw score equivalents

This unit should not present any great difficulty, assuming that you understand the material in the preceding section. Given the mean and standard deviation of a distribution, you now should be able to convert various raw scores into their corresponding z scores. The present section simply involves the reverse process: given the mean and standard deviation of a distribution, you will be asked to convert z scores into their corresponding raw scores.

For example, I tell you that on a geography test $\overline{X} = 70$ and $\sigma = 5$. John, a bright student, received $z = 3.00$ on that particular test. What was the actual raw score that he received?

To answer this question, we use the same logic that we used in the preceding section, but in reverse. A $\underline{z} = 3.00$ means that John scored 3 standard deviations above the mean. Looking at the results of the test we know that $\sigma = 5$, and therefore that $3\sigma = 15$ points. John, therefore scored 15 points above the mean. Since the mean was reported as 70, we conclude that John received a raw score of 85 on the geography test. Hooray for John! If you are afraid you will run into trouble by merely using logic (and you will on more difficult problems), the basic z score formula is shown (once again) below. It is the same formula that you used in converting raw scores into z scores.

$$\underline{z} = \frac{X - \overline{X}}{\sigma}$$

But, since now we are looking for "X" rather than "z" a simple rearrangement of the formula should prove desirable:

$\underline{z}\,\sigma = X - \overline{X}$ (cross multiplying)

$X - \overline{X} = \underline{z}\,\sigma$ (switching sides of the equation)

$X = \underline{z}\,\sigma + \overline{X}$ (adding $\overline{X}$ to both sides of the equation)

We still have the same equation, but it is now in a slightly different form. In effect, we have isolated X, thus making it easier to compute its value. Using the above data regarding John and the geography test we can compute his raw score as follows:

$$X = \underline{z}\,\sigma + \overline{X} = 3.00(5) + 70 = (15 + 70) = 85$$

The equation corresponds perfectly with the logic used in the introductory example. Read the example again, but this time follow the equation. Hopefully, it all makes sense, since nothing magical has taken place: the z score formula has simply been used in reverse.

Example Problem: Assume $\overline{X} = 60$, $\sigma = 3$. What is the raw score for $\underline{z} = 0.70$?

$$X = \underline{z}\,\sigma + \overline{X}$$

$$= (.7)\ (3) + 60$$

$$= 2.1 + 60$$

$$= 62.1$$

Example Problem: Assume $\bar{X} = .3$, $\sigma = .06$. What are the raw scores for *z* scores of 2.3 and −0.9?

For *z* = 2.3:

$$X = (2.3)(.06) + .3$$
$$= .138 + .3$$
$$= .438$$

For *z* = −0.9:

$$X = (-0.9)(.06) + .3$$
$$= -.054 + .3$$
$$= .246$$

Some Properties of z Scores

Before turning to the next instructional unit, some properties of *z* that were not listed in the "specific objectives" section (first page of this unit) will be briefly reviewed. Here is the first: Suppose that you take a final exam in your two hardest courses, nuclear physics and advanced basketweaving. You get the results back to find that your score in physics was 75/100, whereas your score in basketweaving was 86/100. When you show these results to your parents (or friends) they conclude that your major strength is in basketweaving and encourage you to turn professional. Is their conclusion correct?

The answer is maybe yes and maybe no. Hopefully, you are now aware that the interpretation must depend on additional information, such as your relative standing to others in the two classes. Your physics score, even though it was a measly 75, could have been the highest score in the class. Your basketweaving score, even though it was an admirable looking 86, could have placed you in the lower echelon of all weavers who ever wove. A clearer basis for comparison would be provided if you computed your *z* score on each test. Let's say you do (all you would need are the mean and the standard deviation for each), and find that for physics, *z* = 2.32; and for basketweaving, *z* = 0.58. Conclusion: you were better than average on both tests, but on that particular day and relative to your classmates in each course, you were a better physics student than you were a basketweaving student. Hopes for a professional basketweaving career take a serious plunge.

The point we are trying to convey is that z scores allow you to make comparisons between performances on different tests. Regardless of raw scores (whether they be in the 80's, 90's, 100's, or 5,000's), the z scores position you relative to others in your group. If $z = 1.46$ in biology, and $z = 1.46$ in English, we can infer that relative to the other students in those classes, your two performances were roughly equivalent. In each case, you scored 1.46 σ above the class mean. The next problem is optional, but you will probably want to give it a try. It could boost your ego tremendously if you get the right answer.

Optional Super-Bonus Question

A student gets a score of 85 on a test where $\overline{X} = 100$ and $\sigma = 15$. What would be a comparable score on a test where $\overline{X} = 40$ and $\sigma = 6$? (Cover the solution steps below, as you should attempt to solve this on your own.)

Solution Steps

The logic used to solve the problem is very straightforward:

(1) For the performance on the two tests to be equivalent, z (Test A) must be equal to z (Test B). In other words, both test scores should be the same relative distance from their respective means.

(2) The only possible way to proceed is to find the z score for Test A:

$$\frac{X - \overline{X}}{\sigma} = \frac{85 - 100}{15} = \frac{-15}{15} = -1.00$$

(3) Now we know that an equivalent score on Test B would have a z score of -1.00. It would be one standard deviation below its mean. We proceed:

$$X \text{ (Test B)} = z\,\sigma + \overline{X}$$

$X = -1.00\ (6) + 40$ (remember, these values are for Test B)

$$X = -6 + 40$$

$$X = 34$$

Thus, equivalent scores on the two tests would be an 85 for Test A and a 34 for Test B. (Both convert to $z = -1.00$.)

Another z Score Property

Any time you have a group of scores, it is usually of interest to find its central tendency ($\bar{X}$) and its variability (σ).
Almost always, that is...

When raw scores have been converted to z scores, central tendency and variability are always the same from one distribution to the next. Here's what happens. Suppose you give a test to a class of 30 students. You compute the raw scores for all students, and then the raw score mean and standard deviation. Then, using the z score formula, you convert each of the raw scores to a z. Looking over the z scores, you begin to question what their mean (average) would be, so you add up the 30 z scores and divide by 30. To your amazement, you get a mean, $\bar{X}$, equal to 0.00 (zero). This arouses your curiosity, so you then enter the 30 z scores into the standard deviation formula to determine their variability. When you are done, you find the standard deviation of the z's to = 1.00. You are disbelieving, as you have learned from all your math courses that problems never work out to give such nice whole numbers like 0 and 1. A more reasonable answer would be somthing like 1.05734. That you could trust. To investigate this further, you wander around the school building, collecting test results from all other teachers. For each test, you convert the scores to z scores, and then compute the z score mean and standard deviation. In each case you get, $\bar{X} = 0$, and $\sigma = 1.00$.

The lesson to be learned is as follows: <u>when all raw scores in a distribution are converted to z scores, the mean of the z's will always be 0 and the standard deviation will be 1.00 and the form will be unchanged</u>.

For practice, try the following problem: The distribution shown below indicates students' scores on a chemistry test.

X	f
95	3
90	10
85	16
80	30
75	17
70	15
65	4

If these scores are all converted to z's, what would (a) the mean of the z's be? (b) the standard deviation of the z's be?

(if this problem takes you more than 5 seconds to answer, you do not understand the concept)

4.2 POSTTEST (answers in back)

Convert the following *z* scores into their appropriate raw scores:

1. $\overline{X} = 100$, $\sigma = 15$: *z* scores of 2.20, 1.00, -2.00, 0.00
2. $\overline{X} = 1.2$, $\sigma = .3$: *z* scores of 2.00, 0.30, -0.40, -1.30
3. $\overline{X} = 50$, $\sigma = 4$: *z* scores of 3.00, 2.00, -2.00, -2.80

4.3 Instructional Unit: Converting *z* scores into *T* scores

There is no question that *z* scores have practical utility. At all large universities (and most colleges) computer programs are available for use in scoring examinations and in analyzing data from research experiments. If you have a class of 30 or 40 students (which is common at the college level) it is usually much more efficient in terms of time and accuracy to use machine scoring on objective tests as opposed to scoring the tests by hand. You also obtain a lot more information from computer scoring such as means, an item analysis, and yes, *z* scores. Some instructors ignore (or de-emphasize) raw scores, and use the *z* scores for determining grades (others, of course, totally ignore the *z* scores, and some have no idea what the *z* scores represent). To give you a general idea of how *z* scores might be used, the instructor knows that a student who received a *z* score of 0.00 was exactly average (scoring right on the mean), and thus might award him or her a C; a student who received a *z* score of +1.00 (pretty good) would probably be awarded an A, etc. There is a little more to it than these examples would indicate, but the point is that it is sometimes easier for the prof to award a grade fairly if he knows, for example, that John Smith received *z* = +2.00 than if he only knows that Smith answered 18 questions out of 20 correctly. *Z* scores can be helpful, but whether they are to be used or ignored is up to the whims and biases of the individual instructor. Ask your statistics instructor to show you (if you are interested) a copy of a print-out from a computer-scored examination. You will discover that the print-out contains most of the statistical indices with which you have become familiar in the preceding units. All of these indices are intended to help the instructor understand how students performed, and suggest ways in which the examination might be improved for subsequent usage.

All right, you say, but what is the objective of the present instructional unit? To introduce the objective, it might be helpful (or relevant, at least) to point out that at many universities, computer print-outs of exam performances use what are called T scores, as opposed to z scores, to indicate students' relative standings. The reason for this relates to several disadvantages of the z statistic: One disadvantage is that in any distribution, approximately half the z scores will be negative (indicating raw scores some distance below the mean). As you probably realize, negative numbers make calculations more difficult, and grade-book summaries less interpretable than would be the case if only positive numbers were employed. Another disadvantage of the z statistic is that nearly all of the scores will involve decimals; for example, $z = 1.50$ (for a raw score 1-1/2 σ above the mean). Rarely, will the z scores end up whole numbers, such as, +2.00, +1.00, etc. The presence of decimals obviously makes calculations more difficult and confusing.

Recognizing these disadvantages of z scores, it would seem desirable to employ an equivalent measure which eliminates the burden of working with negative numbers and decimals. Such a measure is called a T score; it tells you the same thing that a z score does, but is expressed in terms of positive, whole numbers. T scores are so easy to work with that in some university courses grades are determined exclusively in terms of T score averages or totals. You might have heard the term "grading on a curve." T scores offer a way for this to be done.

If you know how to calculate z scores, the computations of T scores should present little problem. You will recall from the preceding unit that $\bar{Z} = 0.00$, and $\sigma_z = 1$. For a T score distribution, however, $\bar{T} = 50$ (that is, the average T score is always = to 50) and $\sigma_T = 10$ (the standard deviation of the T's is = to 10). We derive our T scores directly from z scores using the following formula:

$$T = 10z + 50$$

Thus, if we know a student's z score, finding his T score is a relatively simple endeavor. By changing z to T we eliminate negative numbers and decimals.

Example Problems:

1. John's z score on a test is +1.00. What is his T score?

   Answer: $T = 10\ (1.00) + 50$

   $T = 60$

2. Floyd's z score is −2.06. Convert to a T score.

Answer: $T = 10\,(-2.06) + 50$

$= (-20.6) + 50$

$= 29.4$ (rounding off)

$= 29$

3. Reginald's raw score on a Health 4062 exam is exactly equal to the class mean. What is his T score?

Answer: If he scored on the mean, $z = 0.00$ (Right?)

$T = 10\,(0.00) + 50$

$T = 50$

(a reasonable answer since $\bar{T} = 50$ – see page 104)

4. Edwinna scored 3σ below the mean on her sewing mid-term. What is her T score?

Answer: If she scored 3σ below the mean we know that $z = -3.00$.

$T = 10\,(-3.00) + 50$

$T = 20$ (poor Edwinna)

Also, remember that $\bar{T} = 50$ and $\sigma_T = 10$.

Knowing that Edwinna scored 3σ below the mean translates to 30 points below the T mean, which is always 50. Thus, again we find that for poor Edwinna, $T = 20$. If this logic is confusing stick to the above formula using z's.

5. On a bone-chewing speed test for dogs, $\bar{X} = 40$ seconds and $\sigma = 5$ seconds. Rover chews his bone in 42.5 seconds. What is Rover's T score?

Answer: First compute z.

$$z = \frac{42.5 - 40}{5} = \frac{2.5}{5} = +.50$$

Finding T should be easy.

$T = 10\,(.50) + 50$

$= 55$

To show the relationship between raw scores, z scores, and T scores, some selected performances from a hypothetical examination are shown below. The mean on that examination was 75 and the standard deviation was 5. Keep these figures in mind when viewing the resultant z and T scores.

Student Name	Raw Score	z Score	T Score
Bernie	90	+3.00	80
Petunia	87	+2.40	74
Louise	85	+2.00	70
Arthur	80	+1.00	60
Archie	78	+0.60	56
Flo	75	0.00	50
Boris	74	-0.20	48
Beatrice	70	-1.00	40
Xavier	68	-1.40	36
Sally	65	-2.00	30
King	60	-3.00	20

You will note that each T is 10 times its corresponding z, plus 50. Answer this: if every score in the class was included above, what would the average z score be?

Answer: 0.00

If every score was included, what would the average T be?

Answer: 50

We had mentioned earlier that T scores can come into play for grading on a curve. If the instructor wanted to grade on the curve, he might establish the following cutoffs for letter grades.

T Score	Grade
59 or above	A
54-58	B
47-53	C
42-46	D
41 or below	F

Note that since T scores, like z scores, indicate distances from the class mean, this procedure ensures that some students will get grades below "C" and some will get grades above "C". Assuming that the distribution approximates the "bell-shaped" polygon that you learned to identify in Unit I (and will study in more detail in the immediately following section), there will be a greater proportion of "C's" in the above class than there will any of the other grades. As one moves

ERIC

upward or downward relative to "C," the number of frequencies will decrease (there will be more "B's" than "A's" and more "D's" than "F's"). The number of "B's" will be approximately the same as the number of "D's"; the number of "A's" will be roughly equal to the number of "F's". We are not necessarily recommending this procedure of grading, but rather are just presenting it as a system that some teachers use. In fact, we should make you aware of its major disadvantage, namely, that not on all tests do certain numbers of students deserve to receive a particular grade, such as "A" or "B" or "F". Here is a brief example that should illustrate this problem. On the written test for Volleyball II, all 10 students receive a score of 90 percent or above. The result is a whopping class mean of $\bar{X} = 95$. When a T score is computed for John, who received a score of 91 (4 points below the mean), it turns out to be $T = 43$. Using something similar to the above grading scheme, the result will be a test grade of "D" for John. But he knew 91 percent of the information tested! Perhaps you can see where such a system would be unfair. If a criterion-referenced system was used, which specified particular grades for various test scores (regardless of how the class does as a whole), John would probably receive a "B" or an "A". As far as John and many others would be concerned, one of the latter two grades would be a great deal fairer than the "D" established by the curve.

In summary, the T and the z are standard scores. They place score values relative to the mean of a distribution, and thus provide standard bases for comparisons across different tests. There are other common standard scores besides T's and z's. IQ, for example, is derived through a standard score formula. You probably know that values of 120, 135, etc. are interpreted as representing fairly high IQ's. How are those values obtained (they do not imply having to answer 120 or 135 questions correctly on a test)? The formula is this:

$$\text{IQ score} = 15z + 100$$

Thus, regardless of how many questions are on an IQ test, an appropriate IQ value can be derived by converting the student's test score to a z score and then applying the above formula. (Note: Some IQ formulas use slightly different standard deviation values.)

Example Problem: Murray Binet, an unemployed psychologist who believes that he is a descendant of Alfred Binet, devises what he believes to be a new test of IQ. The test consists of reading five newspaper "advice" columns and answering four multiple-choice questions on each. He assumes that the more questions that one misses (20 misses are possible), the higher his/her IQ. The test is normed using 500 people sampled in various laundromats across the country. The overall test results yield $\bar{X} = 12$ (missed) and $\sigma = 3$. Jeanette Hufnager, a housewife from Milwaukee, misses only 2 out of the 20 items and thus gets a score of 2 (since this test works on the basis of number incorrect). Can you figure out what her IQ would be represented in the standard IQ units?

Solution Steps

(1) First find $\underline{z}$: $\underline{z} = \frac{X - \bar{X}}{\sigma} = \frac{2 - 12}{3} = \frac{-10}{3} = -3.33$

(2) Then convert to an IQ score:

$$IQ = 15\underline{z} + 100 = 15\ (-3.33) + 100$$
$$= -50 + 100$$
$$IQ = 50$$

Thus, Jeanette, as a result of answering 18 out of 20 questions correctly, is determined to have an absolute score of 2 misses and an IQ = 50.

Is this really her IQ? No, unless testing specialists and the measurement community in general regard the test as a valid measure of IQ, which seems highly unlikely. The moral of the story is that any test scores can be converted to the IQ standard by the above procedure. What makes the scores representative of intelligence is agreement by measurement and IQ experts that the test does, in fact, characterize people in terms of that construct.

Here is a final example of standardization. You may have heard people talk about their scores on the Scholastic Aptitude Test (SAT), used in many cases for admission to undergraduate programs, or their scores on the Graduate Record Examination (GRE), used for admission to graduate programs.

The standardization procedure employed in these tests is:

GRE (or SAT) score = $100\underline{z} + 500$

Some quick calculations would show you that an average score ($\underline{z} = 0$) would convert to a GRE standard of 500. If you scored one standard deviation above the mean ($\underline{z} = +1.00$), that would convert to a standard score of 600. Three standard deviations above ($\underline{z} = +3.00$) would yield an 800; three below ($\underline{z} = -3.00$) would yield a 200, etc.

It's now time to present the POSTTEST on $\underline{T}$ scores as was promised many sentences ago.

4.3 POSTTEST (answers in back)

1. Convert the following $\underline{z}$ scores into $\underline{T}$ scores.

   a. +2.00 b. 0.00 c. −1.56 d. +0.72 e. −2.40

2. I give an examination and find that $\overline{X} = 22$ and $\sigma = 8$. Determine $\underline{T}$ scores for the following raw scores.

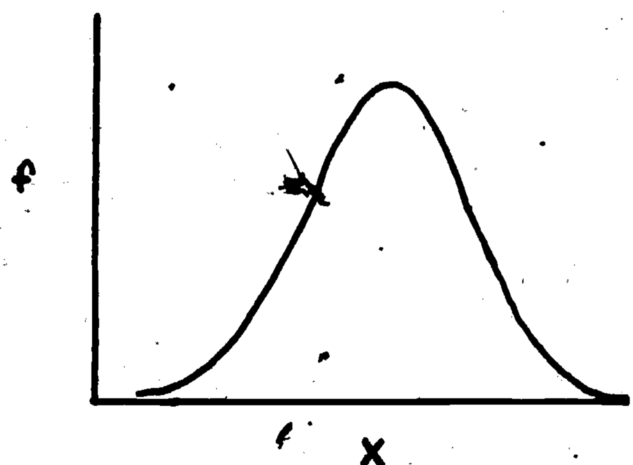
   a. 14 b. 22 c. 26 d. 16 e. 34

3. I give an exam and find that $\overline{X} = 68$ and $\sigma = 7$. Find the raw scores for the following $\underline{T}$ scores.

   a. 50 b. 40 c. 35 d. 62 e. 47

4.4 Instructional Unit: Normal Curve

The normal curve, bell-shaped curve, or – technically – the Gaussian curve, is simply a smoothed out frequency polygon where the bunching of the frequencies occurs at the middle and the stringing out or lower, flatter portions of the curve occur at the ends.



The importance of the normal curve is not that it is the ideal curve, or that it has some magical properties but rather that many human characteristics tend to produce frequency polygons that take on the normal or Gaussian form. Since many things in nature fit this pattern and since there are certain known mathematical properties of the normal curve (as with other curves also), scientists and laymen alike use it for various types of quantitative work.

For this particular unit, the instructional objective is that you gain a general understanding of the properties of the normal distribution. You will not have to derive its properties mathematically. But, can you describe in 10 words or more what the normal curve suggests? Suppose you were told that a distribution of scores falls into the normal curve pattern - What does that mean?

Examine the normal curve diagram and note that the figure represents a frequency polygon of the type you studied (and constructed) in Unit I. Perhaps you recall that frequency polygons may differ in terms of their shape. Some of the patterns we studied were the rectangular, bell-shaped (actually another name for the normal curve), triangular, negatively skewed, positively skewed, etc. So, when we discuss the normal distribution, we are referring to a frequency polygon that is unimodal, symmetrical, and bell-shaped, and more!!

So, if a series of scores fall into a normal distribution, what does that mean? The interpretation is fairly straightforward. Looking at the X and f axis, we can say the following: the highest frequencies (i.e., the most "popular" scores) fall towards the middle of the distribution; the mean, mode, and median are essentially identical; and as you move farther and farther from the mean (or center of the distribution), the frequencies of scores tend to decrease.

Height, for example, tends to be normally distributed. If you measure the height of every male in the United States and plot a frequency polygon, the resultant curve will resemble the bell-shaped (normal). The highest point on the curve will correspond with the mean height in the population, let's say, approximately 5' 9". As you move farther and farther away from the mean in a positive direction, the frequencies will decrease: more men will be 5' 10" than 6' 9"; fewer will be 6' 0" than 5' 11"; and so on. Finally, as you move toward the very extreme positive end of the distribution, (say, for example, heights greater than 7'), frequencies will be extremely low. Since the curve is symmetrical, the same thing happens when you move away from the mean in the negative direction: More men will be measured at 5' 8" than at 5' 7"; more at 5'5" than at 5' 3"' and extremely few at less than 5'. The curve should be symmetrical, which means that half of those measured will be above the mean, whereas half will be below the mean. Even more specifically, there will be the same frequencies of scores at positive and negative points equidistant from the mean: that is, f's should be equal for 5' 11" (+2" from mean) and 5' 7" (-2" from mean); for 6' and 5' 6"; for 6'1" and 5' 5"; etc.

ERIC

4.4 POSTTEST (answers in back)

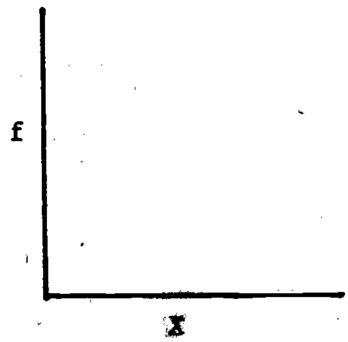
1. Describe the normal curve ______________________________

____

____

____

2a. Make a sketch of the normal curve.



2b. Suppose that the above diagram (assuming you drew it correctly) represents the distribution for the number of hamburgers eaten in a contest held on a college campus. In other words, when the scores for all participants, represented as the number of hamburgers eaten by each, were tabulated and plotted, the result was a normal distribution of scores. The mean of the distribution was 10. Assuming an exact normal distribution, identify each of the following statements as "true," "false," <u>or</u> "can't answer."

_____ 1. The mean score of 10 would be positioned on the "X" axis directly below the highest point on the curve.

_____ 2. There would be more scores above 10 than below 10.

_____ 3. There would be more scores below 10 than above 10.

_____ 4. The probability of finding a student who consumed 13 hamburgers would be greater than finding one who consumed 10 hamburgers.

_____ 5. A score of 15 should be equally common as one of 5.

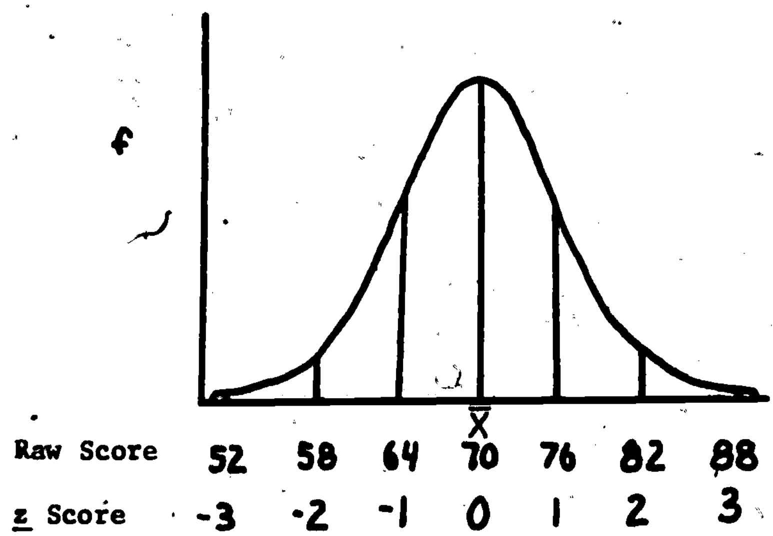
_____ 6. The mode of the distribution should be 10.

_____ 7. The median should be about 11.

_____ 8. The probability of eating 8 cheeseburgers would be greater than that of eating 7 cheeseburgers.

4.5 Instructional Unit: Raw Score and z score scales in Normal Distribution

Normally, on the baseline of a frequency polygon, you have a raw score scale. These scores may be heights, IQ's, learning rates, etc. On this scale we would typically place points that at least represent $\bar{X}$ and several σ's on either side of $\bar{X}$. In addition to the raw score scale, it may be necessary to label some other transformed or converted scale. The most frequently used transformed scale is the z score scale. Given a normally distributed set of data with $\bar{X} = 70$ and $\sigma = 6$, the raw score and z score scales would look something like the diagram below.



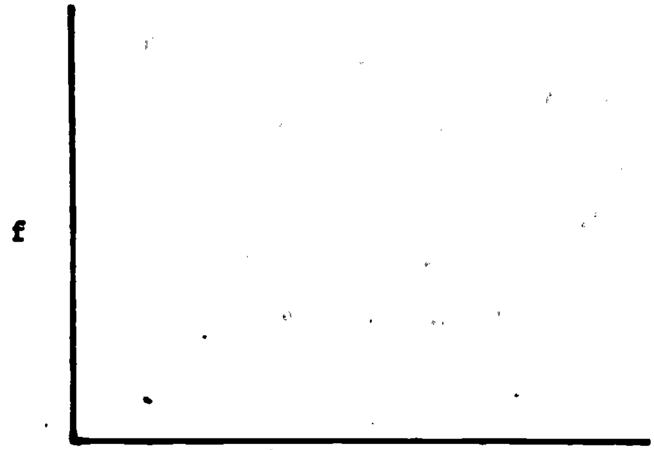
Notice that three σ units have been marked off on either side of $\bar{X}$; while this doesn't take in every possible score, it is rare for any distribution to spread out more than 3σ around $\bar{X}$. Also note that the σ units have been evenly spaced on either side of $\bar{X}$ so that 1σ represents the same distances at all places along the baseline. *It can't be emphasized too strongly that making a simple normal curve diagram with the raw score and z score scales penciled in will facilitate solving a variety of probability and statistical problems.*

Final Note:

The normal curve is not an ideal curve but rather a curve that seems to accurately describe many human characteristics.

4.5 POSTTEST (answers in back)

1. Make a sketch of the normal curve in the space below and label as accurately as possible a raw score scale and $\underline{z}$ score scale where $\bar{X} = 30$ and $\sigma = 4$.



Raw Score ______________________________

$\underline{z}$ Score ______________________________

2a. Assume that IQ scores are normally distributed in Mrs. Gordan's 12th grade class. If we make a plot of this distribution, will the heights of the curve at $+1\sigma$ and -1σ be equal or different? Explain.

2b. Assume that $\bar{X} = 98$ in Mrs. Gordan's class, and give answers to the following:

1. Over which score point will the height of the curve be greatest?

2. Which score should be more common, a 106 or a 90?

3. Is a score of 50 possible? Why or why not?

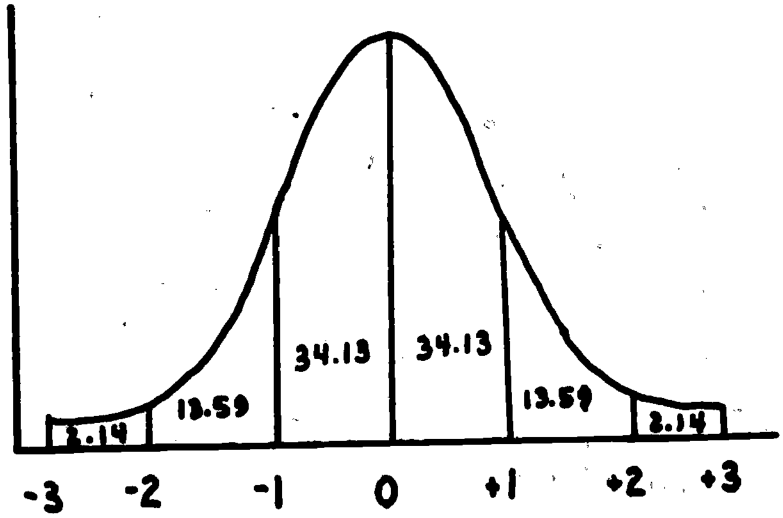
4.6 Instructional Unit: Percentages of scores between standard deviations

One useful property of the normal curve is that a constant percentage of the total area (assume total area = 100%) will fall between fixed $\underline{z}$ scores on the baseline. That is, a certain constant percentage will fall between $\underline{z} = 0$ (i.e., $\bar{X}$) and $\underline{z} = 1$, or between $\underline{z} = -1$ and $\underline{z} = -2$, and so on.

Regardless of the raw score mean and standard deviation of the set of normally distributed data, the percentage of the total area between two fixed $\underline{z}$ scores will be the same.

ERIC

The exact area between any two z scores can be found using a normal curve table, as will be discussed in Units 4.7 - 4.10. For present purposes, though, we will diagram the percentages between whole number z scores. These values are often helpful to memorize, as they provide a ready basis for estimating percentages of scorers in any normally distributed group. For percentage areas between fractional z scores, use the normal curve table (we'll worry about this later).



Thus:

z values	area of curve	rounded off area (memorize these)
-3 to -2	2.14%	2%
-2 to -1	13.59%	14%
-1 to 0	34.13%	34%
0 to +1	34.13%	34%
+1 to +2	13.59%	14%
+2 to +3	2.14%	2%
	99.72% = 100%	

Note:

The reason that the total is not exactly 100% is that, technically, the curve does not stop at the z scores of -3.00 and +3.00. Actually, the curve goes to "+" or "-" infinity. However, only .28 of 1% fall outside of z scores of -3 and +3; therefore, not much information can be gained by extending the curve beyond three standard deviations in either direction.

If you think you understand the material on areas of the normal curve, see if you can answer the following questions:

a) You gave a math test, the results of which were normally distributed. What percentage of students who took the test scored between 0 and +1σ from the mean.

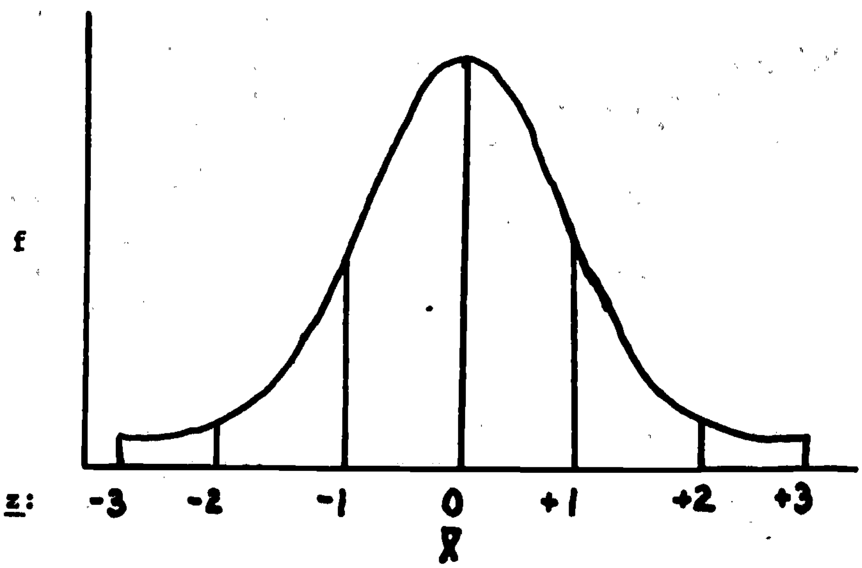
Answer: 34% of the students fall into that range. Why? Anytime data is normally distributed approximately 34% of the distribution will be between a z score of 0 (that is, the mean) and a z score of +1.

b) On the same math test, what proportion of the students scored between z = -1.00 and z = +1.00?

Answer: 68%; anytime data is normally distributed, 68% of the scores will fall within z = "+" or "-"1.00. 34% are between z = -1 and z = 0, and 34% are between z = 0 and z = +1. 34 + 34 = 68. Technically, the answer is 68.26% (34.13 + 34.13).

4.6 POSTTEST (answers can be checked on preceding page)

On the normal curve below, label the approximate percentage values between the vertical lines. (Do this without looking on the previous page; otherwise you will merely be demonstrating that you have mastered the art of copying numbers.)



130

4.7 Instructional Unit: Finding percentage of scores between any two $\underline{z}$'s by use of "normal curve" table

As the title of this unit suggests, we will be working with something called a "normal curve" table, or more specifically, "area under the normal curve" table. Coincidently, one is provided in Table I in the Appendix. That is the table to work from once you are ready for the actual problems. First, however, we'll present an abbreviated version to show you how the table is used.

In the preceding unit, you memorized some percentages which indicated the proportion of scores, in a normal distribution, between whole number $\underline{z}$ values such as 0 to +1; +1 to +2; -2 to -3; etc. Presumably that did not overly tax your memory. You probably would object quite a bit, however, if you were asked to memorize the proportions for all possible combinations of $\underline{z}$ values; for example, -2.78 to -1.03, etc. Thus, to find the percentage of scores between any two $\underline{z}$ values, we make use of an "area under the normal curve" table. If you can learn how to use the table, you can easily determine the percentage of scores between any two $\underline{z}$ values.

The table specifies the area or percentage of the total normal curve between $\bar{X}$ (for which $\underline{z}$ always equals 0) and some other $\underline{z}$ value (whichever is of interest to you). The table contains two types of information: a) $\underline{z}$ values, which are located in the first column on the left (along with hundredths values at the top); and b) the percentage of scores that fall from $\bar{X}$ ($\underline{z}$ = 0) to other $\underline{z}$ values; these are the figures in the main body of the table. The general outline of the table is presented below. (It would probably be valuable to also take a look now at the full picture in Appendix Table I.)

Area Under Normal Curve Table
(hundredths)

$\underline{z}$	.00	.01	.02 09	
0.0	.0000	.0040	.0080	.0359
0.1	.0398	.0438	.0478	.0753
.				
.				
.				
1.0	.3413	.3438	.3461	.3621
1.1	.3643	.3665	.3686	.3830
.				
.				
.				
3.0	.4987	.4987	.4987	.4990

By use of the table we can find the percentage of scores that fall between any two specified _z_ values. Keep in mind that the figures shown on the table represent the area from _z_ = 0 to any specified _z_ value. For example, suppose you are interested in determining the percentage of scores in a normal distribution that fall between the mean (_z_ = 0) and 1.5 standard deviations above the mean (_z_ = 1.5). (You will need a complete table to follow this example.) All you need to do is locate _z_ = 1.5 on the table; since no hundredths are involved (as would be the case if _z_ = 1.51, for example), move down the vertical column until you reach _z_ = 1.5. Immediately adjacent to the _z_ value, 1.5 (in the column corresponding to "no hundredths" — .00), is the area value, .4332. Find this value now, otherwise you will be totally lost (if you aren't already). Move the decimal place on the area value two places to the right and you will end up with 43.32. This means that 43.32% of the scores in a normal distribution fall between _z_ = 0 (the mean) and _z_ = 1.50. Notice on the table that the listed _z_ values range from 0.0 to beyond 3.00. Since the normal curve is symmetrical and has 50% of the area on each side of $\bar{X}$, the normal curve table only has to account for 1/2 or 50% of the area because the other half is identical. Keep in mind that the z values in the table can be either + or –.

The above description may seem pretty heavy, so let's backtrack a bit. Actually all that was said is as follows:

(1) locate the _z_ value you are interested in on "the area under the normal curve" table. To do this you move down the vertical column (indicating whole number _z_'s and tenths) and across the first row (indicating hundredths).

(2) the 4-digit numbers you find indicate the proportion of scores (or area) in a normal distribution between _z_ = 0 and "your" _z_ - the _z_ score you located.

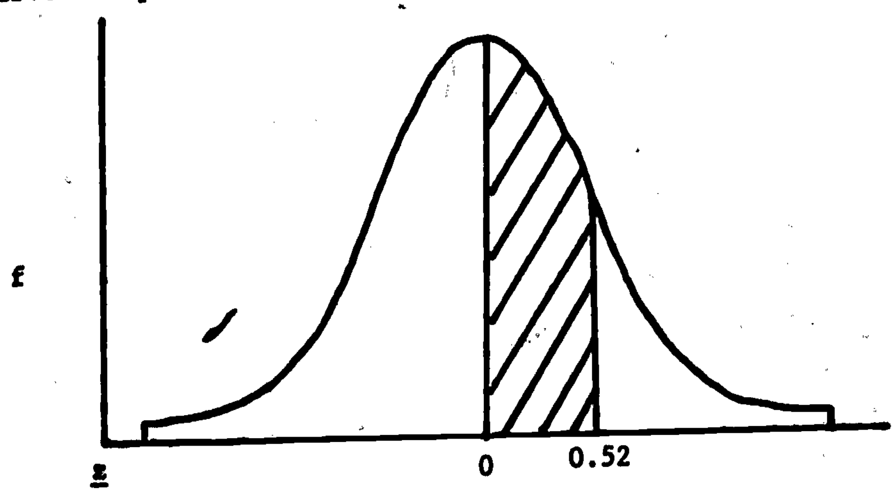
The remaining sections of this unit will present some example problems intended to illustrate different prototype applications of the "normal curve area" table. These basic prototypes will also comprise the focus of the three units that follow. They consist of:

A. Finding the area between _z_ = 0 and some specified ("+" or "–") _z_ score.
B. Finding the area between a "+" and a "–" _z_ score.
C. Finding the area between two "+'s" or between two "–'s."
D. Finding the area above or below a "+" _z_ score.
E. Finding the area above or below a "–" _z_ score.

Given that these types of problems can apply to cases where z's are given, or where raw scores are given which must be converted to z's, there would seem to be at least 10 different applications to learn (or memorize). Take heart, this does not have to be the case at all. From this point onward, the remainder of the material all relates to a limited number of fairly simple concepts. If you understand the concepts (which mostly involve being comfortable with the "table" and "normal curve" diagrams), you should be able to solve any of the prototype problems presented, without having to memorize the specific steps required for each. It is all a logical process. So, "meaningful" learners have a (fairly easy) problem-solving exercise to look forward to. Rote memorizers, however, beware! You have a long road to travel, filled with the usual pitfalls of having to rehearse things that you don't understand, and then confusing the steps memorized for different problems, or forgetting them completely, at the time correct procedures need to be applied. How will you know if you are one of the respected and revered meaningful learners or one of the dastardly rote learners? There is no absolute criterion, but it would seem that you are probably a meaningful learner if, given an area problem, you can represent what it asks in a mental image or, much better, in an actual diagram of the normal curve. You are probably a rote learner if you can derive the correct answers to problems, but cannot represent what you are doing in picture form. Beware of this and also of the full moon, as meaningful learners have been known to turn into rote learners on such nights. The early danger signs are rehearsing things to yourself at meal times and writing formulas down on little pieces of paper which you store in pockets of different shirts. (The recommended cure for this condition, sometimes called "rotewolvery," is to eat a wreath of garlic following each recital; if nothing else, it will keep your statistics instructor away.)

Prototype A: What is the proportion of scores in a normal distribution between the mean and $z = 0.52$?

First, you should draw a rough sketch of the normal curve and pencil in the area you are asked to measure:



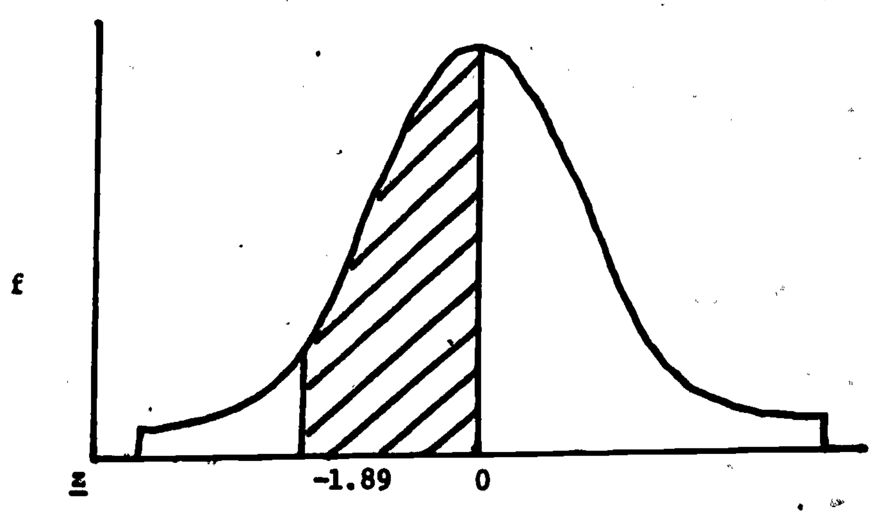
Note that z = .52 is positioned a short distance above z = 0.00, since it is a positive value. Exact placement is not necessary for these diagrams. Also note that the area between the two z's is shaded, as this is what the problem asks you to find. Then you simply locate on the "normal curve" table the specific z value (z = 0.52) and record the four-digit area value.

Hopefully, you agree that for z = 0.52, the area value is .1985. We move the decimal place two digits to the right and end up with 19.85 percent. (If you were unable to locate .1985 for z = 0.52, seek help immediately - you must be totally comfortable with the "normal curve" table in order to proceed further with any chance of success.) In any case, the answer to this problem is 19.85% of the scores in a normal distribution are between z = 0 and z = 0.52.

Another Example: What proportion of scores in a normal distribution fall between the mean and z = -1.89.

The fact that we are working with a negative z score makes no difference since the normal curve is symmetrical. The area between the mean and z = -1.89 is identical to the area between the mean and z = +1.89. Thus, we use the table in the exact same manner for both positive and negative z scores.

First, make a sketch of the normal curve and pencil in the area in question, i.e., the area between z = 0 and z = -1.89 (note that z = -1.89 was positioned "some distance" below the mean, since it is negative).



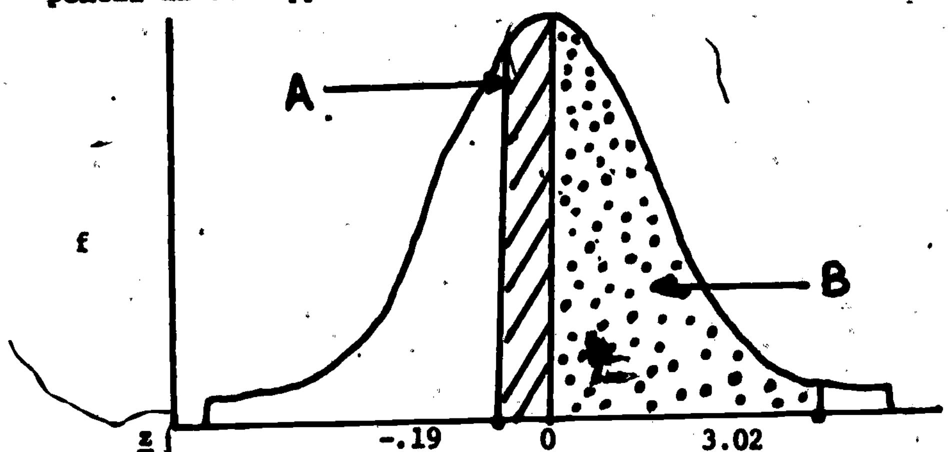
Examining the "table," we find that the proportion listed for z = 1.89 is .4706. Thus, we can conclude that 47.06% of the scores fall between the mean and z = -1.89.

Now, let's finish this instructional unit by examining four more prototypes. Each prototype relates to a specific and unique type of "area" problem. If you try to follow their logic, they all should be fairly easy.

Prototype B: Finding percentage of scores between positive and negative z scores.

What is the percentage of scores between z = -.19 and z = +3.02?

Before giving up, make a sketch of the normal curve and pencil in the approximate area you are trying to find.



First, place the z = -.19 somewhere below the mean, and the z = +3.02 somewhere above the mean. Then read the problem again. Note that it asks for the area between these scores. Perhaps by looking at the diagram you can gain some insight as to the correct procedure.

The diagram suggests that we need to go .19 standard deviations below $\bar{X}$ to get Area "A"; according to the normal curve table, the area for z = -.19 is 7.53%. In addition, we need to go 3.02 standard deviations above $\bar{X}$ to get Area "B"; the table shows this area to be 49.87%.

Since our interest is in the total area between z = -.19 and z = 3.02 (that is, Sections A & B), we add the two areas:

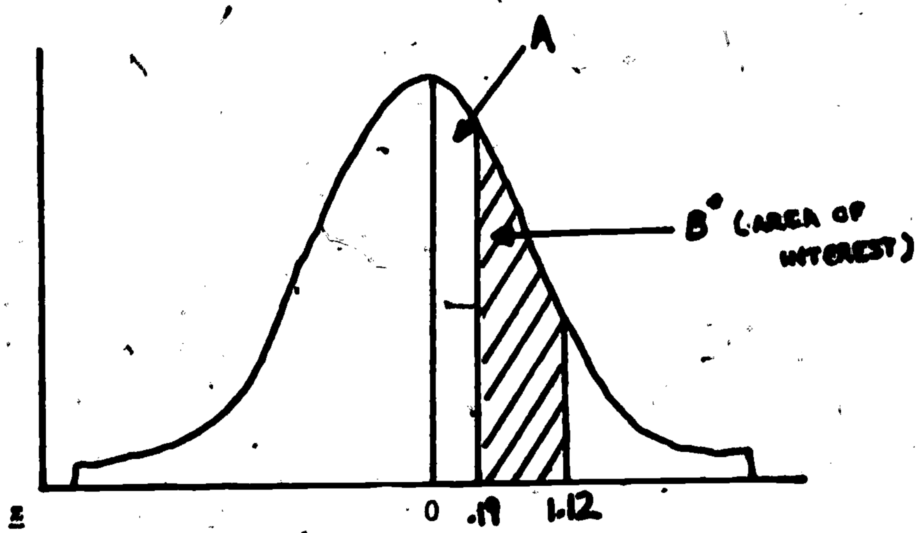
7.53% + 49.87% = 57.40%

Our answer: 57.40% of the scores in a normal distribution fall between z = -.19 and z = 3.02.

Note: With the diagram, this type of problem should be easy; without the diagram it could become confusing. Always use a diagram!

Prototype C: Finding percentage between two positive z scores, or between two negative z scores.

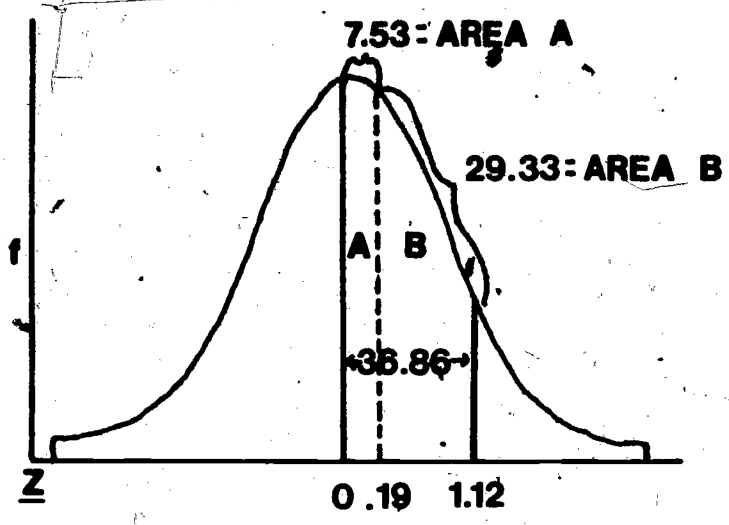
What is the percentage of scores that fall between z = .19 and z = 1.12? Make a sketch of the normal curve, placing both z values on the positive end, but the +1.12 somewhat higher. Then pencil in the area of interest.



To find the appropriate area, first look up the area for z = .19. This tells you that 7.53% of the scores range from $\bar{X}$ to z = .19 (or area in Section A). Then look up area for z = 1.12 which is 36.86%; this represents area from mean to z = 1.12 (or areas in Sections A and B). To solve the problem you must get the value for Section B alone. So, you subtract: 36.86 - 7.53 = 29.33% (A + B - A = B).

Your answer: appropriate area = 29.33%.

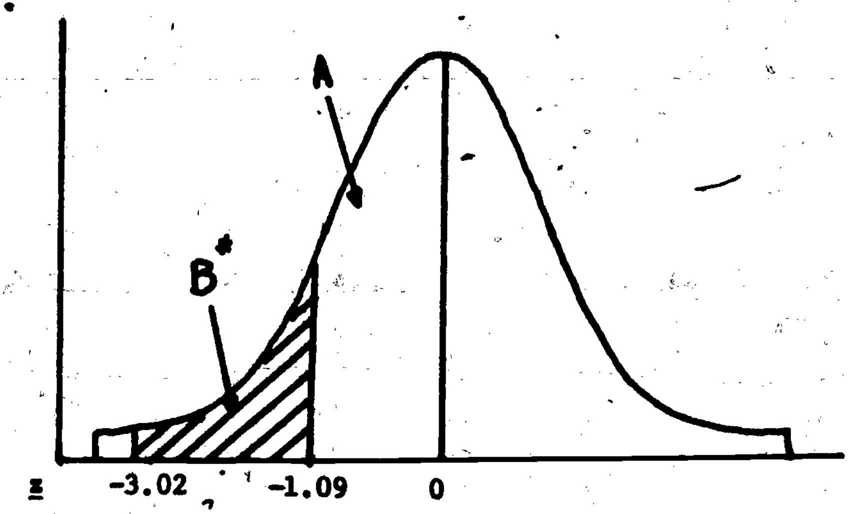
For this particular problem, what we had to do was find the shaded area (B) by treating it as the remainder of the whole (A + B) minus the part (A). It's all a logical process, and there is never any need to panic. Just draw the curve, fill-in the givens, reread what the problem asks, and then attack it step-by-step. As you locate a particular area from the table, it's usually a good idea to mark it in the appropriate place on your diagram. That way, your diagram will continually present an up-to-date picture of where you stand on the problem. For the problem we just solved, the areas would be marked in the manner shown on the next page.



Another Example: Finding percentage between two negative z scores.

What is the percentage of scores that fall between z = -3.02 and z = -1.09?

Make your sketch, being sure to position the -3.02 somewhat lower on the scale than the -1.09. Then shade in the area of interest, i.e., the area between the two z scores.



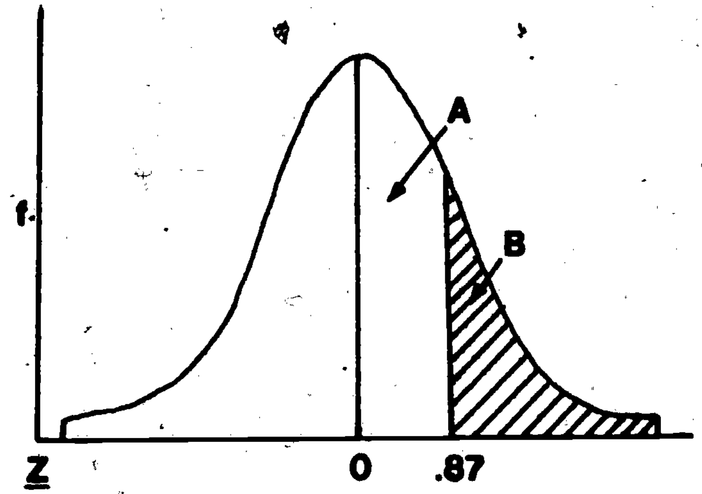
Using the table we find that from 0 to -1.09 (look up 1.09), we have 36.21%. This is Section A. From 0 to -3.02 (look up 3.02), we have 49.87%. This accounts for Sections A and B. To find Section B alone: 49.87% - 36.21% = 13.66%. (A + B - A = B).

Our answer: 13.66% of scores fall between z = -1.09 and z = -3.02. The procedure is exactly the same as for the positive z scores in the preceding example.

Prototype D: Finding the area above or below a "+" z.

What percentage of scores in a normal distribution would be above z = +.87?

First, draw the normal curve sketch, being sure to place the z of .87 somewhere above the mean. Reread the problem; what area is it asking for? You are right if you're thinking the area above the z score. Shade that area in as shown below.



Now the task is to find the percentage that corresponds to the shaded area only. We look up the area value for .87 z, and find it to be 30.78%. Is this the answer? If you say "yes," you are not paying attention to the problem, because the 30.78% belongs between z = 0 and z = .87. We are looking for the area beyond .87z. How do we proceed???

If you are stumped, here is a clue (actually a prompt): How much area is above the mean in every normal curve that ever existed?

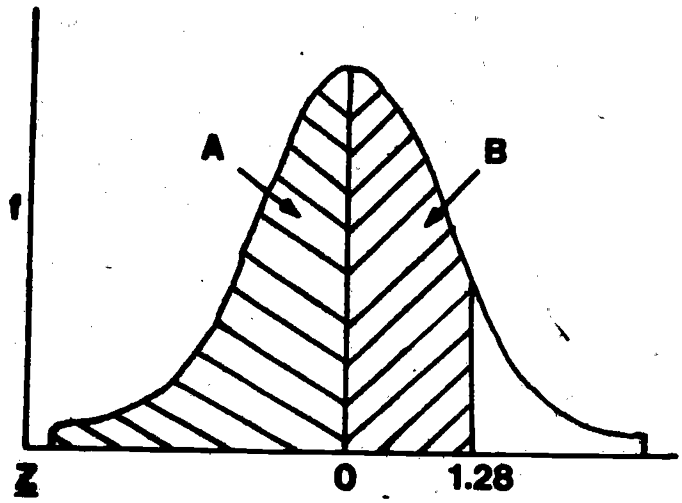
Answer: 50 % above; 50% below.

Thus, the solution to the problem is 50 (A + B) - 30.78 (A) = 19.22% (B). You are taking the part (A) away from the whole (A + B) to find the remainder (B).

Another Example: What percentage of scores would be below z = 1.28?

First, make your sketch, placing the 1.28 somewhere above the mean. Reread the problem and note that it is asking for the percentage below the z score. Shade in that area as shown:

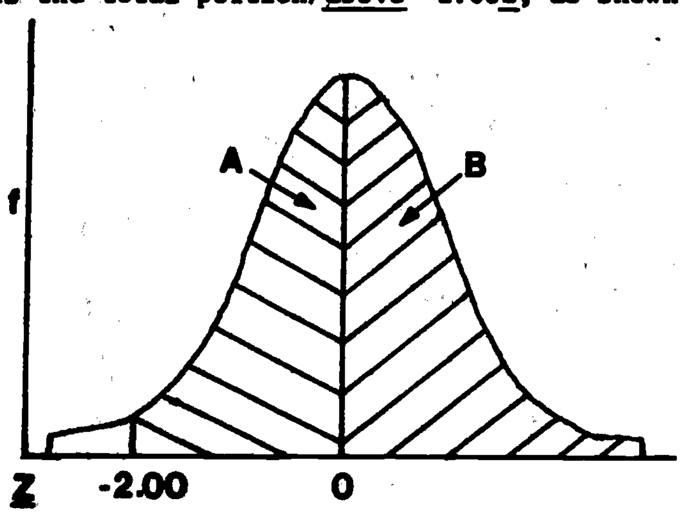
ERIC



How do we proceed? The only logical place to start is with looking up the area for 1.28z, which turns out to be 39.97%. Is this the answer? NO!!!! It is only part of it, as our mission is to find the total percentage below 1.28z. What percent is below z = 0.00? Fifty percent, as always, is below the mean. Thus, to get the total below we simply add 39.97 (area B) and 50.00 (area A) to get 89.97%.

Prototype E: Finding the area above or below a "-" z.

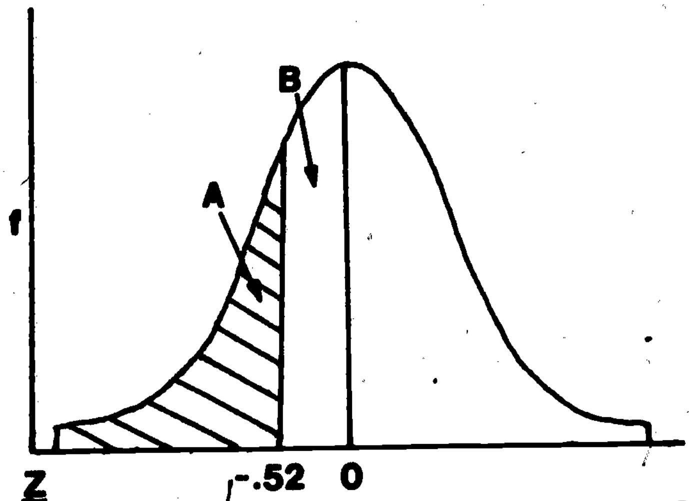
What percentage of scores would be above z = -2.00? Here, we are confronted with simply a mirror image of the "+" z problems. The first step, as usual, is to make a sketch and shade in the area of interest. This time the area is the total portion above -2.00z, as shown.



We look up the area for -2.00z, and find that it is 47.72%. The important thing to note is that it represents area A. Area B, of course, is 50%. Thus, the total percentage above -2z is 47.72 (A) + 50 (B) = 97.72% (A + B).

Last Example Before Posttest

What percentage of scores are below $-0.52z$ in a normal distribution? Diagram looks like:



The shaded area reminds us that we are looking for an area below the z score. The tabled value for $-.52z$ is 19.85%, and on the diagram it represents Area B. The total area below the mean is 50% (Area A + B). Therefore, we proceed as follows: 50 (A + B) - 19.85 (B) = 30.15% (A). We are done!!!

Final Notes:

1. Do not attempt to memorize the five prototypes. Try to understand them, and to see how they all really involve the same concepts.

2. Be sure you can make a sketch of any of the prototypes. Artistry is not important unless you want to use your normal curve sketches for your bulletin board or portfolio.

3. It should be pointed out that the above exercises involve the maximum work ever needed to solve area problems. Some more elaborate normal curve tables not only list the areas between the mean and the z scores, but also provide the total areas both above and below each of the z's. This information would cut short some of the solution steps we went through above. But, if you know the full procedure, you are naturally prepared to work with any normal curve table. A little more work at this stage could be valuable later.

ERIC

4.7 POSTTEST (answers in back)

Assuming a normal curve, find the areas specified by each of the following *z* score boundaries, and represent each problem in a normal curve sketch.

1. between $z = 0.00$ and $z = 2.00$
2. between $z = -.90$ and $z = -.20$
3. between $z = -2.20$ and $z = .70$
4. between $z = 0.00$ and $z = -2.50$
5. between $z = .70$ and $z = 2.10$
6. above $z = .45$
7. below $z = -1.12$
8. above $z = -.70$
9. below $z = +.98$
10. (BONUS) above $z = 1.96$ OR below $z = -1.96$

4.8 Instructional Unit: Finding area between any two raw scores in a normal distribution

This is one of those instructional units that really does not offer anything new; it merely asks you to think again about concepts that presumably have been mastered in previous sections. To get through this instructional unit with ease, you need to know: a) how to convert raw scores to *z* scores given the mean and standard deviation of the distribution (instructional unit 4.1), and b) how to find the area between any two *z* scores in a normal distribution (instructional unit 4.7).

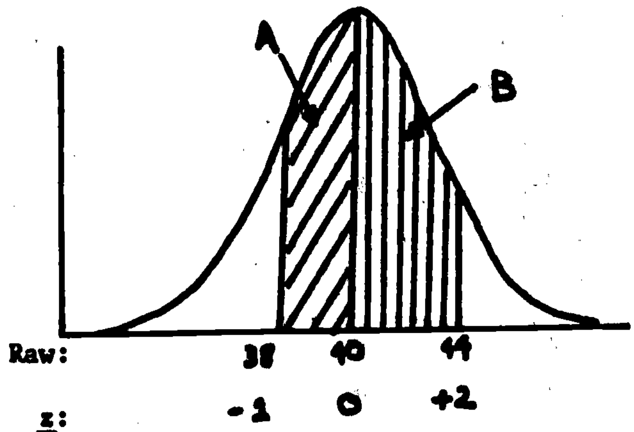
In other words, this unit is going to be very similar to the preceding unit. However, instead of being given two *z* scores, you will be given two raw scores. Your task will be to find the area between the two raw scores. If you remember how to convert the raw scores into *z* scores (if you do not remember you should not be this far ahead in the unit), then you end up with exactly the same kind of problem you attempted to solve in instructional unit 4.7 (turn back one page). It should be great fun.

O.K., let's get through this quickly. To find the area between two raw scores in a normal distribution, first convert the raw scores to *z* values and then use the "area under the normal curve" table to find the area between the two resultant *z* scores. To assist in this process, it is again essential to make a simple sketch of the normal curve. In the sketch, include a raw score scale with (see next page) corresponding *z* values. Then pencil in the area you are trying to measure.

ERIC

Example Problem: What is the area (percentage of scores) between the raw scores of 38 and 44 in a distribution where $\bar{X} = 40$ and $\sigma = 2$?

Make your diagram:



Convert the raw scores to $\underline{z}$ scores and include a $\underline{z}$ score scale on the diagram as shown above and below.

For 38: $\underline{z} = \dfrac{38-40}{2} = -2/2 = -1.00$

For 44: $\underline{z} = \dfrac{44-40}{2} = 4/2 = +2.00$

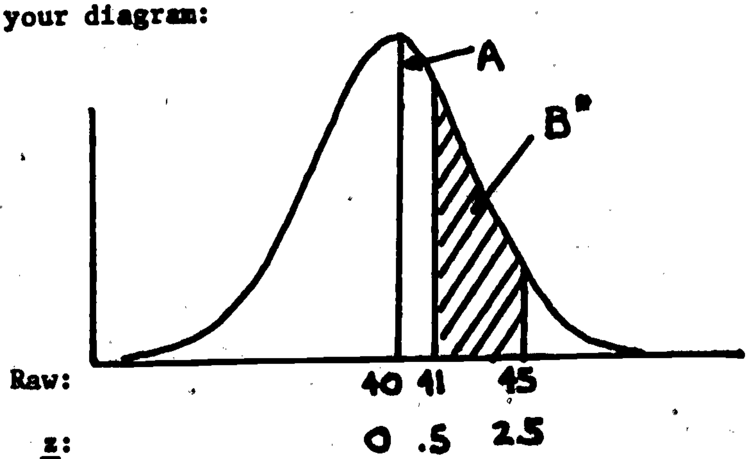
You will note that this is a problem in which you must find the area between a positive z score and a negative z score. Thus, you find the percentages in the "normal curve" table for each $\underline{z}$ score and add them together (A + B = A + B).

The area between $\bar{X}$ and $\underline{z} = -1$ is 34.13%. Between $\bar{X}$ and $\underline{z} = 2$, you get 47.72%. Therefore, between $\underline{z} = -1$ and $\underline{z} = 2$, or raw scores of 38 and 44, you have 34.13% + 47.72% = 81.85%.

Answer: 81.85% fall between scores of 38 and 44 on this particular distribution.

Example Problem: What is the area between raw scores of 45 and 41 in a distribution where $\bar{X} = 40$ and $\sigma = 2$?

Make your diagram:



Convert the raw scores to z scores and include a z score scale on diagram.

For raw score of 41: $z = \frac{41 - 40}{2} = 1/2 = .5$

For raw score of 45: $z = \frac{45 - 40}{2} = 5/2 = 2.5$

Note that this is a problem in which you must find the area between two positive z scores. Thus, find the percentage values for each and subtract the Part (A) from the Whole (A + B) to get the missing Part (B).

Between $\overline{X}$ and z = .5 you get 19.15% (in Section A).

Between $\overline{X}$ and z = 2.5 you get 49.38% (in Section A & B).

For Section B alone:

49.38% - 19.15% = 30.23% (A + B - A = B).

Our answer: 30.23% of the scores fall between the raw scores of 41 and 45 in this particular distribution. Keep in mind that 30.23% of the scores do not always fall between raw scores of 41 and 45. They do in the above example, however, because the corresponding z scores, +.5 and +2.50, happen to include that area. Think about it.

Final Notes:

1. Always sketch out problem. If you cannot, you should return to Unit 4.7 and review the procedure for the diagrams.

2. Given raw scores, first convert to z scores and then find the area between the two z values.

4.8 POSTTEST (answers in back)

For each of the sets of data below, find the percentage of scores specified by the listed raw score boundaries. Make a diagram for each.

1. Assume $\overline{X} = 72$ and $\sigma = 4$. Find the percentage of scores:

   a. between 66 and 71
   b. between 72 and 80
   c. below 68

2. Assume $\overline{X} = 500$ and $\sigma = 100$. Find the area:

a. between 450 and 625
b. above 480
c. above 530

3. Assume $\overline{X} = 22.1$ and $\sigma = 3.4$. Find the area:

a. below 21.3
b. between 22.0 and 24.0
c. between 20.3 and 21.9

4.9 Instructional Unit: Converting *z* scores into percentile ranks, and percentile ranks into *z* scores

As you found out in Unit II, a percentile rank is the percentage of scores (of total N) that fall below some point in the distribution. If Fred's percentile rank is P_{34}, then we can say that 34% of the scores in the distribution fall below Fred's. No problem.

Also in Unit II, you learned to use the interpolation equation to compute percentile ranks...and if you are like most people, you probably disliked doing that immensely. Well, here is some good news: if you know the person's *z* score in a normal distribution, you can rather quickly compute his/her percentile rank by use of the "area under the normal curve" table. As a matter of fact, you have already had experience with this type of problem in the past two units.

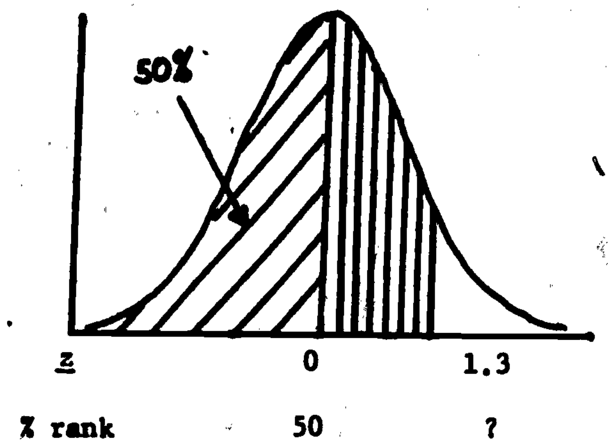
Any time you have been given an area problem that asked you to find "the percentage of scores below...," you were really being asked to determine a percentile rank. Maybe you realized this, but probably, with your attention directed to the different normal curve problems and the making of beautiful diagrams, you did not. To capitalize on your past experiences (assuming you do complete the posttests or at least stare at them for awhile), turn back to Posttest 4.7 and look at problems 7 and 9. Both ask you to find the area below a specified *z* score, a "−" one in #7 and a "+" one in #9. You have come a long way in your study, if you can see that both problems really ask you to specify the percentile ranks of the given *z*'s, i.e., the percentage of scores that would be positioned lower than they. To get a feeling of progression, now turn forward to Posttest 4.8 and examine problems 1c and 3a. It's the same story; both ask you to find the percentage of scores below the given (raw) scores; both ask, in essence, for the specification of the percentile ranks for those scores. It should be obvious to you that the percentile rank for a *z* score or raw score positioned below the mean must be less than 50 (P_{50}); for a z or raw score positioned above the mean, it must be greater than P_{50}. What this means is that if you calculate the percentile rank of a z score equal to, say, +1.00 and determine it to be something like P_{34}, your answer has to be wrong and, in fact, is downright embarrassing. In such an instance, you would have forgotten to add 50% to the tabled value for z = 1.00 of 34% to get a more decent looking P_{84}. Let's briefly review the procedures, although you should know them quite well at this point.

For normal distributions, percentile rank problems are identical to "area below" problems. To find a percentile rank for a z score, you need to find the area below that z score. To find a percentile rank for a raw score, you need to: a) convert the raw score to a z score, and b) then find the area below that z score. The system that we have been using throughout this Unit, but especially on area problems, can be represented as follows:

Score Indices	Conversion Procedure
Raw Score (X)	
↓↑	
Standard Score (z) ⟶	formula (unit 4.1)
↓↑	
Area of Percentage (% or P_x) ⟶	"Normal curve" table (unit 4.6)

What this shows is that any type of problem which we have dealt with can be solved by either one or two conversions. Changing an X to a z, or a z to an X, is a one-step process involving use of the conversion formula. Changing a z to a %, or a % to a z, is also a one-step process, but involves the use of the table. Changing an X to a %, or a % to an X, is a two-step process, involving "going to" z first and then converting the z in the appropriate manner. All problem types presented in this Unit are solved by the steps included in this rather simple scheme.

For our first percentile rank problem, assume that we are trying to find the P.R. for a student named Sidney who has just completed a psychology exam. Sid, as he likes to be called, was determined to have scored at z = 1.30. This z score information is good news for us because it eliminates the need for the "formula" (X-to-z conversion) step. Only the table is needed. We make the following sketch, this time incorporating a % rank scale:



To find the area below z = 1.30 and thus, the corresponding P.R., we need to measure the entire shaded part of the curve. The first half, by definition, encompasses 50%. All we need, then, is the "unknown" area from z = 0 to z = 1.30. The table, of course, comes to our aid by listing 40.32% for that portion. What percentage of the class has Sid beaten?

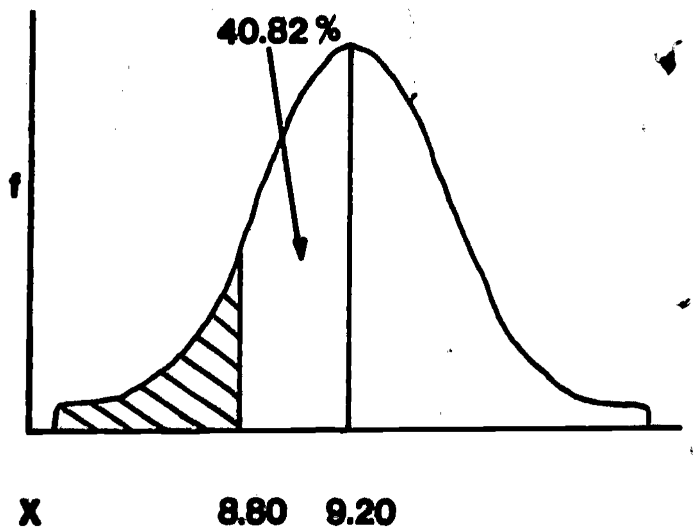
50 + 40.32 = 90.32%

Sid should be proud.

Another Example

In an international women's gymnastics tournament, the average rating for participants on the balance beam is 9.20 with σ = .3. Natasha Bobolinckskismith, from lower Urkansas, received an 8.80. What is her P.R.? (assume a normal distribution).

Looking at the information provided, this has to be one of those two-step problems. To get from raw (actual) scores to percentages, we must go to z first. To keep our sanity, however, it would be a good idea to represent the problem in diagram form.



The diagram shows that Natasha scored below the raw score mean, and thus will receive a negative z (when one is computed) and a P.R. somewhere below 50. The z conversion is the first step:

$$z = \frac{8.80 - 9.20}{.3} = \frac{-.4}{.3} = -1.33$$

We should list the -1.33 on our sketch in place of the "?" in the z scale. Now we are ready for the table. We look up the area for z = 1.33, and find that it is 40.82%. What is Natasha's P.R.?

If you said "40.82" you should be sent back to the salt mines (with Natasha). That figure represents the area <u>between</u> -1.33z and 0. To get the area that is below -1.33z, we need to subtract the 40.82 from the whole (50%). (If you don't see this, check unit 4.7, Prototype E.) The P.R. for Natasha is 50% - 40.82% = 9.18%. Natasha, apparently, will soon be repairing cartwheels, rather than doing them, in the Ukzark mountains.

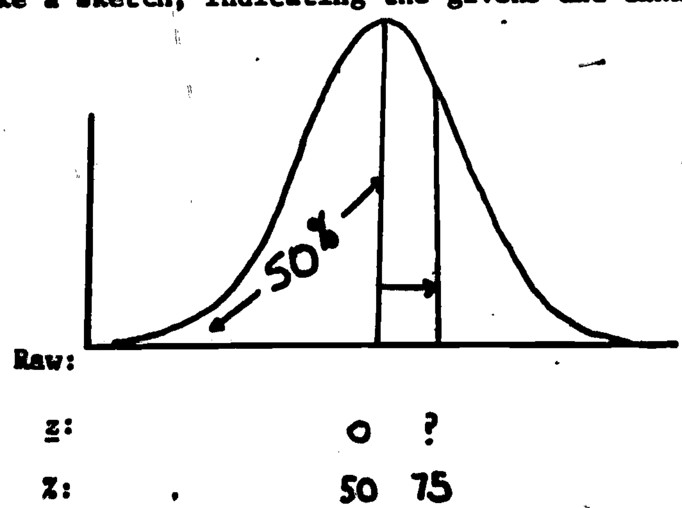
Reversing the Process: % to z's

You can now determine the percentile rank equivalent for a z score in a normal distribution. Based on this knowledge, reversing the process to get a z from a P.R. should not be very difficult. Using the kind of area table that we have, there is only one "trick" involved, and if you can master that, it will be smooth sailing thereafter. Let's see how the reversal process works in an example.

Example

On a nationally given scholastic aptitude test, Tim is told that he scored at P_{75}. Assuming a normal distribution, what would his z score be?

Make a sketch, indicating the givens and unknowns.

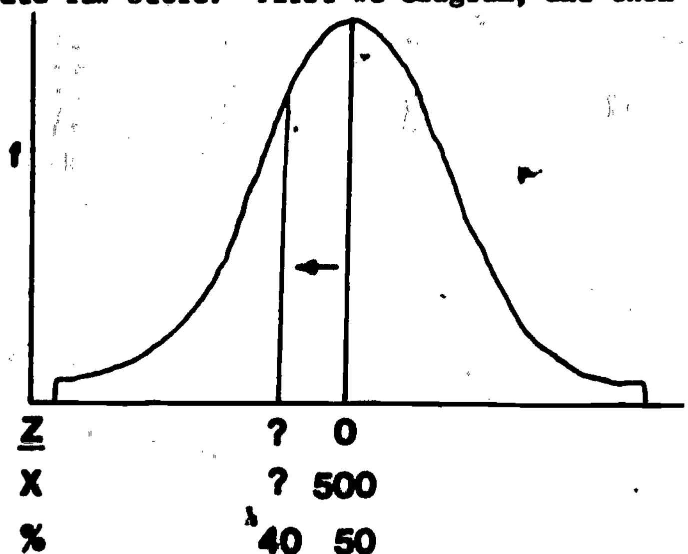


The sketch shows us immediately that Tim's z score is positive (P_{75} defines that). The question is how do we find it. Well, if you remember the conversion scheme we talked about a few pages back, you can probably identify this as a one-step problem: a % to a z. As defined, it requires use of the table only. We must look up an area to find z, but here is the trick. Some people, who are not thinking, try to look up 75% (.7500) on the table, because that is the only area they have been given. We don't want a z score that has 75% of the scores between it and the mean (besides, there isn't such an animal). We want one that "beats" 75%. Since our table works from z = 0 outward, we need to look up the area between the mean and our P.R.: 75% - 50% = 25%. Thus, the z score in question has 25% of the cases between it and the mean. To find it, we scan the areas on the table and identify the one closest to .2500 (25%). It turns out to be .2486. The z that corresponds to 24.86% is .67. We attach a "+" sign to it, since we are dealing with the positive end of the curve, and express the answer as +.67z. It is the only z score below which 75% of the cases fall.

Another Example

On the same aptitude test as in the first example, Sally's performance places her at P_{40}. The overall test mean is 500, with σ = 100. What was Sally's score?

This is one of those two-step problems, as it involves going first from % to z, and then using z to obtain the appropriate raw score. First we diagram, and then identify z.



The diagram reminds us that we will be working on the negative side of the curve. We need to find a z that surpasses 40% of the scores, but to look up .4000 on the table would be an outlandish mistake. Since the table works from the middle (z = 0), we need to define the z score in question in terms of its distance from the middle. The lower half of the curve encompasses 50% of the scores and "our" z score surpasses 40%. Through simple subtraction, 50% - 40% = 10%, it becomes clear that the distance between the z and the mean is 10%. On the table, the z will be the one most closely associated with the area value .1000. A scan of the table values gives .0987 as the closest figure. Its associated z is .25, which must be given a negative sign due to P_{40} falling on the negative side.

Our answer: $z = -.25$ for P_{40}.

We are almost, but not quite, finished. The problem asks for Sally's score on the test, which moves us to the second step: converting a z to a raw score. Now the highly familiar conversion formula should do the trick:

$$X = z\sigma + \overline{X}$$

substituting the givens:

$$X = -.25(100) + 500$$

$$X = -25 + 500$$

$$X = \underline{475}$$

Bonus question: If 1000 people took the aptitude test, how many did Sally beat?

The solution is much simpler than it might first appear. Sally scored at P_{40}. By definition, she scored higher than 40% of the people in her group. We get: 40% of 1000 = .40 x 1000 = 400 people.

Let's return to Tim from the previous example. You may recall that he scored at P_{75}. Suppose that when he took the test, the group consisted of 11,978 students. How many did he beat? We need to take 75% of 11,978: .75 x 11,978 = <u>8,983.50 people</u>. Looking at the answer, you may be worried about the half a person represented by the decimal figure, and wonder whether he/she is very short. No need to worry; a fractional person does not exist, but is almost certainly a "creation" of round-off errors that occurred when Tim's P.R. was computed originally. For accuracy in our present computation, we should keep the ".5" in our answer.

4.9 POSTTEST (answers in back)

1. Convert the following z scores into percentile ranks:

   a. −2.40
   b. −1.20
   c. 0.00
   d. +0.90
   e. +2.30

2. Convert the following percentile ranks into z scores:

   a. 15
   b. 25
   c. 40
   d. 60
   e. 85

3. In a distribution where $\overline{X} = 10$ and $\sigma = 3$, what is the percentile rank of raw scores of:

   a. 14
   b. 5
   c. 9

4. In the same distribution as above, what raw scores correspond with:

   a. P_{50}
   b. P_{80}
   c. P_{43}

Recommendations for Unit IV:

1. Try to understand the logic to the area problems. If you can achieve this understanding, you will be able to solve any of the prototype problems, without having to memorize the specific solution steps for each. For example, if you can solve the problems presented in the preceding section, unit 4.9, you should be able to solve the problems presented in units 4.7 and 4.8, as they deal with the identical concepts.

2. Always make a normal curve sketch, even a very rough one, as it serves as a reminder of what the problem involves, the information provided, etc. It also gives reality to the problem, representing the data as related to a distribution of scores, not just to a series of fairly mechanical computations.

3. Remember that all of the problems that have been presented are solvable through a restricted set of operations. Here is a summary:

Given	Find	Use
Raw Score (X)	Standard Score (*z*)	Formula
z	X	Formula
z	Area (% or P.R.)	Table
%	*z*	Table
X	%	Formula to find *z*; then Table
%	X	Table to find *z*; then Formula

4. Try not to regard the different problems as independent entities. A poor procedure for understanding the normal curve concept is to memorize operations such as "for finding the area between a +*z* and a -*z*, I should add." If you can diagram the problem, the normal curve representation will remind you what to do.

Unit IV Review Test
(answers in back)

1. We give a language skills test and end up with a normal distribution with $\bar{X} = 75$ and $\sigma = 8$.

   ____ a. What percentage of scores are above .34$\underline{z}$?

   ____ b. What percentage are below 2.29$\underline{z}$ (or what is its P.R.)?

   ____ c. What percentage are between +.24$\underline{z}$ and −.24$\underline{z}$?

   ____ d. If 1000 students took the test, how many scored between −1$\underline{z}$ and −.5$\underline{z}$?

   ____ e. What percentage are between raw scores of 83 and 87?

   ____ f. John's score was 70. What is his P.R.?

   ____ g. Sandra's score was 90. What percentage scored above her?

   ____ h. Bob's P.R. was 25. What was his raw score?

   ____ i. BONUS: What percentage scored between $\underline{T} = 35$ and $\underline{T} = 55$?

2. Four students received the following scores on a test where $\bar{X} = 500$ and $\sigma = 100$. Assume a normal distribution, and find the percentile rank for each student.

   ____ a. Mary = 575

   ____ b. Ann = 440

   ____ c. Sue = 620

   ____ d. Joan = 780

   ____ e. Barry = 500

   ____ f. Ed = 510

UNIT V

CORRELATION

A. General Objectives

Up to now, we have concerned ourselves exclusively with the measurement and description of *single* variables, such as scores on a test, heights, weights, etc. Frequently, however, researchers in most sciences are interested in testing hypotheses concerning the relation between two variables for a given population of persons or things; for example, height and weight, IQ and Grade Point Average, smoking and heart disease, exercise and physical health, etc. The statistical tool through which such relations can be objectively determined is called *correlation*. In this unit we will introduce the concept of correlation and show how it may be used by researchers to provide insight into relations among different properties or characteristics that define a given population.

B. Specific Objectives

5.1 Define correlation
5.2 Describe the data collection procedures involved in conducting correlational studies
5.3 Construct and interpret scatterplots of correlational data
5.4 Define and interpret correlation coefficients
5.5 Compute correlations by means of different formulas

5.1 Instructional Unit: Definition of Correlation

In earlier units, we have dealt with variables (such as IQ, test scores, heights, weights) on an individual basis. Specifically, we have shown that through the assessment of such characteristics we can describe a given population in terms of several kinds of indices. For example, we might report that a random sample of high school seniors in Vienna, Illinois has a mean IQ of 102, with a standard deviation of 14.5, and a range of 67 - 174. As will be described in later units, we might then use this information to make inferences about the population that this sample represents. Our results, however, would be restricted to IQ scores as would the conclusions we could then make. Later, we may wish to describe and compare the two populations on the basis of numerous other individual variables.

Such assessments would tell us something more about the particular population(s) in question. But they would not tell us whether the various traits meaningfully relate to one another and can be used as a basis for making predictions about the characteristics of individuals in the population. We may know, for example, that students in a particular school average 101.7 in IQ, and 508 on the verbal portion of the SATs. Having this information, however, is not sufficient to answer questions such as: "Are IQ and SAT scores significantly related for this group?" "Will a student who is high on IQ tend to do well on his SATs?" "How is Johnny, who has an IQ of 78, likely to perform on the SATs?" Questions like these can be asked in reference to numerous other variables: Is height related to weight? Is teacher salary related to teaching effectiveness? Is noise level related to work efficiency?

Such questions lead to the concept of *correlation*. In introducing this concept, we might best begin with a definition.

Definition: Correlation is a technique used to determine the relation between two variables.

Simple enough, but one might appropriately question what is meant by the term, *variable*. When we speak of variables, we are referring to characteristics or events that are likely to differ from measurement to measurement, i.e., from element to element in the population being assessed. Thus, we can differentiate between variables and constants, the latter being a measure that remains the same across a series of assessments. We have been dealing with variables since Unit I when we constructed numerous frequency distributions of test scores. But just to make sure of your understanding of these concepts, try to identify the following measures as either variables or constants (cover the answers on the right).

a) IQ in a 2nd grade class:	clearly a variable since the IQ scores of children will not all be the same.
b) number of inches in a population of yardsticks:	a constant, by definition. (Of course there may be trivial differences in length, but for a yardstick to call itself a yardstick, it would have to be 36 inches long.)
c) shoe size in a group of undergraduates (assume that all wear shoes):	a variable since shoe sizes would differ among individuals.
d) number of teeth in a population of parakeets:	a constant; parakeets do not have teeth and thus all scores would be zero.
e) number of digits in a group of social security numbers:	a constant; all scores would be 9.
f) number of face freckles in a population of adults:	a variable; obviously people differ in the number of freckles they have.

Now that we have illustrated what is meant by variable, let us return to our definition of correlation: the degree of relation between two variables. What does it mean if we say that two variables are related in a given population?

One interpretation is that values on one variable can be used to predict values on the other. Or, by knowing a person's standing on one variable, say height, we can make a reasonable guess about his standing on another, say weight. Or, people at certain levels of one variable tend to be found at certain levels of another. If this is the case, we can say that the two variables are correlated. Let's look at some examples.

Example 1: Height and Weight. Would height and weight tend to be correlated in a population of adults? The answer is "yes." Knowing that a person is 7' tall allows us to be reasonably certain that he is heavier than average (unless his nickname is "stringbean"). Thus, height and weight are likely to be related, that is, correlated. As height increases in a population so does weight, at least on the average.

Example 2: Toes and Hair Color. Would there be a correlation between number of toes and hair color in a population of physically normal adults? This one is designed to trap you, since number of toes would not be a variable and thus, could not be correlated. All "toe" scores would, by definition, be 10. Thus, being told a person's toe score would be absolutely useless in making a prediction about his hair color.

Example 3: IQ and GPA. Would there be a correlation between IQ and Grade Point Average in a population of high school students? Most likely there would: Knowing a student's IQ score would provide some basis for predicting his high school average. On the average, students with high IQs have high grades; students with low IQs have low grades.

Example 4: Length of Name and Salary. Finally, would there be a correlation between the number of letters in people's names and the amount of money they make? We could correlate the two, but would probably obtain little support. That is, knowing that Alexander Kowalski-freehoppin has 27 letters in his name provides little insight into his probable income. Conversely, knowing a person's annual income would provide little basis for making even an educated guess about the length of his name.

Before going on to a more detailed analysis of what correlation implies, we'll leave you with the following thought: You are likely to find a correlation when scores on one variable seem to be associated in some systematic way with scores on the other variable; e.g., height and basketball ability; IQ and annual salary; daily temperature and the amount of clothing worn, etc. Do not read further until you think you understand why there would be a significant relation between these variables.

One more thought: Variables can be correlated in two different ways, positively or negatively. These types of correlations will be discussed in considerable detail later on, but for now, you should at least try to understand the following very general definitions:

1. We obtain a positive correlation when high scores on one variable tend to be associated with high scores on the other variable, and similarly, when low scores on one variable tend to be associated with low scores on the other.

Can you see how this would apply in the case of:
- IQ and grades?
- height and weight?
- annual salary and size of car?
- love of statistics and the closeness of this book to your nightstand?

2. We obtain a <u>negative</u> (or inverse) correlation when <u>high</u> scores on one variable are associated with <u>low</u> scores on another variable, and vice versa.

   Can you see how this would apply in the case of:
   - absences and grades?
   - annual salary and felony convictions?
   - temperature and number of articles of clothing worn?

To focus on the last example of a negative correlation, isn't it true that when temperatures are high (90, 95, 100 degrees) the number of articles of clothing tends to be low and vice versa? Can you imagine a person wearing only 1 article of clothing (a low value) when the temperature outside is 5 degrees (a low value also)? (Okay, stop imagining; it's time again to get back to work.)

5.1 POSTTEST (answers in back)

1. Define correlation: ______________________________

   ____

   ____

2. Which of the following measures could be examined in a correlational study (which are variables)?

   a. number of digits in U.S. area codes
   b. number of freckles in a group of 2nd graders
   c. number of illnesses in 1975 for adults in same group
   d. number of toes in a group of physically normal 3-toed sloths
   e. number of friends possessed by individuals

3. Suppose we determined the <u>Grade Point Averages</u> of all seniors in a particular high school. How would Grade Point Averages probably relate to the following variables? Choose one of the following in giving your answer:

   a) no correlation is likely
   b) a positive correlation is likely
   c) a negative correlation is likely

   _____1. height

   _____2. number of absences from class

_____ 3. pulse rate

_____ 4. IQ

_____ 5. amount of TV watching

_____ 6. number of books read during past year

4. Of the following pairs of variables, which pair would be most likely to correlate positively?

   Pair 1: Age of automobile and miles per gallon.

   Pair 2: Distance north of the equator and average daily temperature for different cities.

   Pair 3: Number of bars and number of schools in different U.S. cities.

5.2 Instructional Unit: Conducting a Correlational Study: Data Collection

Now that you have some understanding of what correlation implies, we will turn to some of the basic procedures involved in conducting a correlational study. The procedures are really quite simple, and, believe it or not, once you've read through this material, and learned it, you should be able to set up a correlational study of your own (wow!). Before starting, you will be relieved to find out that it is not really necessary to memorize the procedures that will be enumerated. What is important is that you acquire a general understanding of the logic and rationale involved. Here goes...

Step #1: Define the Problem

Before conducting a correlational study, you should identify the problem or hypothesis with which you are concerned. Specifically, you should clearly determine which variables you wish to correlate and why.

In addition to defining the variables of interest, you should also specify the type of population to which your hypothesis is directed. Obviously, the relation between variables such as number of showers (or baths) taken on a weekly basis and popularity might be very different for a population of human beings and a population of baboons. For all we know, the most popular baboons might be the ones who take the least number of showers (thus, a negative correlation). We doubt, however, that such would be the case for humans.

Enough about baboons and showers. For illustrative purposes, let's consider a researcher who is interested in determining whether a significant relation exists between IQ and grades for a population of fifth graders in Tupelo, Mississippi. He might define the purpose of his study in the following manner:

__Purpose__: to determine the relation between IQ and grades for fifth grade students in Tupelo.

To make things interesting, this researcher (let's call him A) actually has a hypothesis: IQ is positively related to grades.

Researcher B who specializes in physical education believes that his college students who are overweight will experience difficulty in performing gymnastics (specifically, tumbling). He might define the purpose of his analysis as follows:

__Purpose__: to determine the relation between weight and proficiency at tumbling for males taking college physical education classes.

__Step #2__: Specify the Sample to be Measured

This really involves determining what may be thought of as the "unit of analysis." That is, __who__ or __what__ will be used to provide measures on the two variables to be correlated. Will you be measuring people, chairs, baboons, farms, etc.?

For Researcher A, it is obvious that the "units" will be people, specifically, 5th graders in Tupelo. Is it necessary for this researcher to measure IQ and grades for every fifth-grade student in Tupelo? It is not necessary since it would probably take too long. It would be far more practical to use a sample, say, of 50 students. The larger the sample, the more likely the data will reflect actual, real-life conditions.

So, using a random selection procedure, Researcher A randomly selects 50 students from the population of fifth graders in Tupelo. He is now ready to begin testing.

Turning now to Researcher B - what is his unit of analysis? Given his interests, the units would be college students. For each student he can obtain a measure of weight and a measure (probably a rating) of proficiency in tumbling. He can then try to determine whether the two measures (variables) relate to one another, that is, whether they correlate.

Since it would be too expensive to include every college student he decides to use a random sample of 25 students from which he hopes he can generalize to the entire population.

__Step #3__: Measure Every Member of the Sample on Both Variables

Correlation implies determining whether scores on one variable tend to be associated in some systematic manner with scores on another variable. Thus, we need to obtain two sets of scores from each element in the sample being measured.

For Researcher A, this will involve measuring all 50 students on both IQ and grades.

For Researcher B, this will involve measuring all 25 students on both weight and tumbling skill. For the latter, he has each student perform a sequence of exercises, and then rates each, with the help of an observer, on a 1-10 scale.

Step #4: Tabulate the Resultant Data

This really involves listing and organizing the obtained scores in order to facilitate their interpretation. The goal, you'll remember, is to determine whether the two variables assessed are related. The best way of obtaining a "quick" picture of this possible relation is to list the scores obtained from each individual in the form of a table.

Researcher A measured IQ and Grade Point Average for 50 students. For the sake of brevity and convenience, let's assume that he only used five students in his sample (summarizing the data for 50 will take too long). Having obtained his data, he might tabulate the scores in the following manner:

Sample:	IQ	Grade Point Average
Freddie	120	3.50
Gertrude	90	2.00
Horatio	100	2.50
Leona	70	1.00
Percy	110	2.00

Note that even a casual glance at the resultant scores will permit some interpretation. Specifically, IQ scores and grades are positively correlated. Individuals with high IQs tend to be those who earn high grades (Freddie, for example), while individuals with low IQs tend to be those who earn low grades (Leona, for example). In fact, with only one exception (Percy), as IQ scores increase, grades tend to increase. For all we know, Percy, who has a relatively high IQ (110), may be a person who does not study or spends too much time with extracurricular activities (very possibly Leona). The important point is that, for the group as a whole, IQ and grades appear to be related. Put another way, knowing a person's IQ allows us to make a better prediction about his grades than would be possible if such information were not available; similarly, knowing about a person's grades allows us to make a better prediction about his IQ than would otherwise be possible. Be alert though! DO NOT use knowledge of correlations to prejudge individuals. Use correlations to derive understanding about how variables operate in relation to one another. Do not use them as a basis for categorizing and stereotyping people before they have been given a fair chance to perform.

Let's now turn to Researcher B, who was interested in determining the relation between weight and tumbling skill of male college students. For convenience, we'll pretend that his sample consisted of only six students although we should recognize that for an actual study, this number would be much too small (most authors of statistics texts recommend a sample size of at least 25-30 for correlational analysis). Like Researcher A, he would first tabulate his results. To illustrate this, we'll make up some "hypothetical" scores for the 6 contestants:

Students	Weight	Tumbling Score
Woody	112	9
Wyeth	186	5
Rollo	227	10
Abe	136	8
Grover	250	1
Bert	150	7

Looking at the data, what is your impression regarding the relationship between variables?

Your conclusion was right if it was that a correlation does seem to be present. Give yourself further credit if you concluded that the direction of this relationship is negative. Higher weights tend to be associated with lower scores. Conversely, the lighter guys (e.g., Woody) tended to receive the higher scores. Assuming that anyone cared about this, and that the relationship was reliable (could be replicated), weight could be used as a meaningful predictor of tumbling skill and vice versa.

Returning to the example, would weight be a perfectly accurate predictor of tumbling skill? The data suggest that it would not. Rollo, for example, at a king-sized 227 lbs., received an absolutely perfect tumbling score of 10 points! Apparently, there are other factors besides weight that influence tumbling. (If you want to find out what happened in Rollo's case, check the inset.)

* * * * * * * * * * * *

Some Assault! Rollo's Great Roll

The absence of a perfect correlation suggests that one variable does not completely account for the other. Clearly, there are numerous factors besides weight that can influence one's skill at tumbling...

In Rollo's case, it happened to be the variable of "roundness." What transpired (be glad you weren't there) is that Rollo, being only 5' tall but 227 lbs., tucked himself into a ball which allowed him to gain momentum after each tumble. He took off across the gym like a gigantic bowling ball, knocking over students and apparatus as he went. His form was superb, and the results--36 tumbles and 23 students down--were rather striking to say the least. Had a score of "11" been possible, Rollo would have received it. Besides, if you were the instructor, would you dare to downgrade this tumbling terror?

Conclusion: In this isolated case, weight worked to the individual's advantage. Correlation can be helpful in making general predictions, but could lead to a wrong conclusion in any single case. Weight was shown as accounting for some, but not all of the differences in tumbling skill.

* * * * * * * * * * * *

In summary, after you collect your data, it is usually quite helpful and always appropriate to present it in tabular form. The table will provide a first impression of what the results show.

---

cor<u>RELATION</u> not

. causation

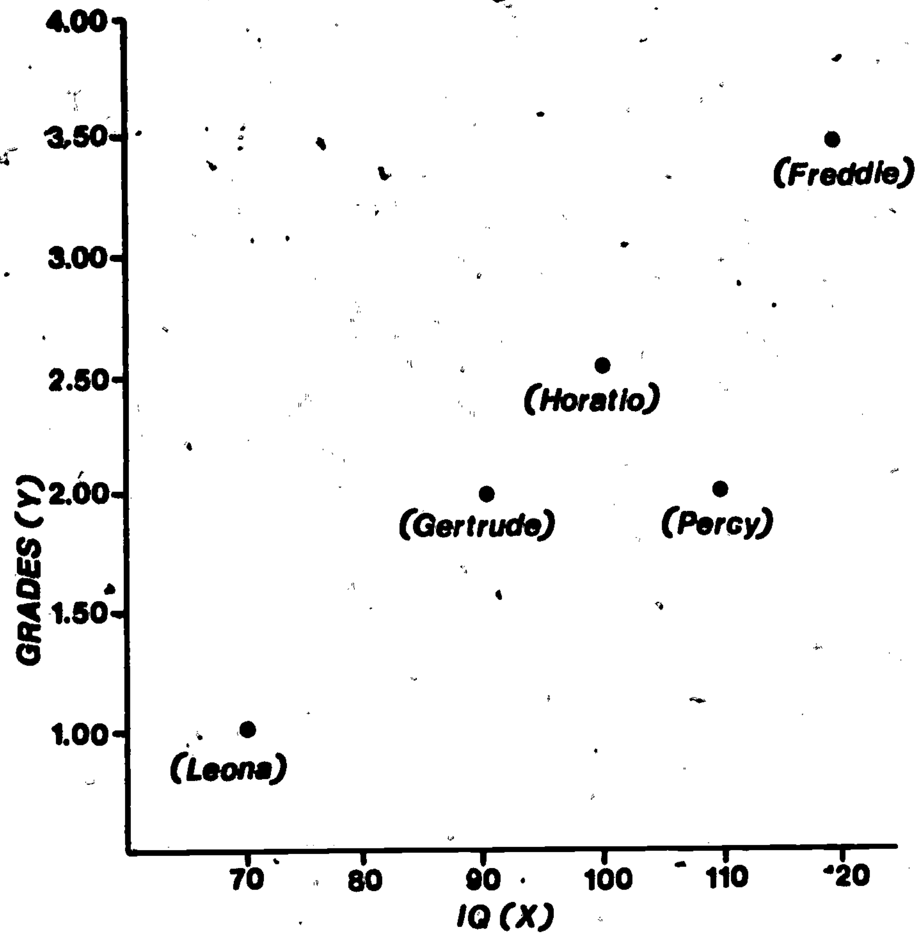
(necessarily)

Finding a correlation merely implies the existence of a relationship between the two variables; *it does not imply that one variable necessarily causes the other.* With regard to Researcher A's study, can you conclude that IQ <u>causes</u> grades or that grades <u>cause</u> IQ? Such a conclusion would be unjustified without much more sophisticated analyses. What usually causes grades, it would seem, is the number of items people get right on tests. The same variable (number right) applied specifically to IQ tests is the direct "cause" of IQ scores. Doing well on an IQ test is something that should typically go along with doing well on academic tests used for grades. Thus, "third factors" such as test-taking skill, problem-solving ability, reading proficiency, etc., can be the actual things that tend to produce the association between measured IQ and obtained grades in a course. By the same reasoning, it would be incorrect to conclude that weight, as used in the second example, is necessarily the <u>cause</u> of tumbling performance. Weight and tumbling could well be associated with third factors, such as balance, speed, coordination, etc. whose influences are the primary determinants of the correlation. In fact, the relationship suggested in the example could be mostly attributable to the instructors' being biased in their perceptions of the tumblers' performance. That is, without being aware of what their biases are, they might tend to downgrade the heavier people and upgrade the lighter people, even though the performances of the two groups might be of equal quality. Here, the "cause" of tumbling skill would not be weight, per se, but instructors' biases.

Step #5: Construct a Scatterplot of the Data

Since constructing and interpreting scatterplots will be the subject of the next instructional unit, the present discussion will be brief. First, imagine that Researcher A actually used 1000 rather than 50 students in his IQ-grades study. He would still want to tabulate the results. After doing so, he would end up with 1000 pairs of numbers. Getting a "first impression" would be rather difficult, especially if he wants to maintain his sanity and his eyesight. In "real life" he would probably look toward the computer to provide the actual correlation, without becoming too concerned with a first impression. But let's suppose there is no computer and he does not know how to calculate a correlation (which you will learn to do in Unit 5.5). In this instance, he would probably find it very helpful and informative to construct what is called a "scatterplot."

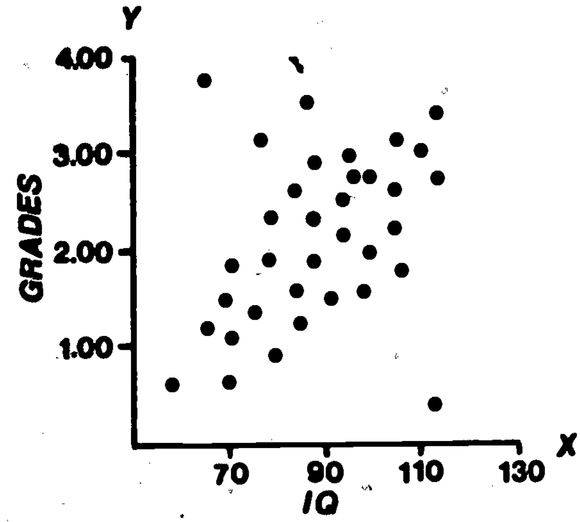
A scatterplot is simply a graphical representation of the scores, with each point on the graph representing the scores obtained for an individual unit (person) on both variables. The graph consists of an X axis (horizontal), showing the range of possible scores on one variable, and a Y axis (vertical), showing the range of possible scores on the other variable. When used together, these axes indicate appropriate positions for representing the scores obtained by each unit (person) in the sample. We represent these scores by placing a point for each person at the point of intersection of straight lines drawn through his score on Variable X perpendicular to the X axis and through his score on Variable Y perpendicular to the Y axis. Sound confusing? It's really quite simple. Look at the following scatterplot for the five people measured in the IQ-grades study (Study A).



Note that, from the scatterplot, we can easily determine how each individual in the sample scored on the two variables. The dot used to represent Freddie's scores is perpendicular to the Y axis (Grades) at the score point 3.50 and perpendicular to the X axis (IQ) at the score point 120. Since the sample included five units, the scatterplot will consist of five dots.

163

Suppose the researcher included every student in the class in his study. The resultant scatterplot might look something like this:

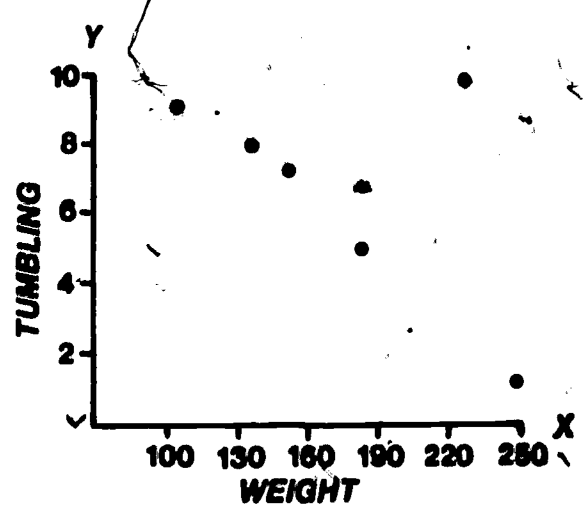


Once again, every dot on the graph represents two scores for the individual units that were measured. How many people were included in the sample? If you feel inclined to count the dots you will find that 35 students were included.

By looking at the scatterplot, anyone who is familiar with correlations would be able to characterize the relationship easily. In this case, grades and IQ appear to show an imperfect, positive relation for this particular sample. By the time you finish reading the next instructional unit, you will be able to do this too (isn't statistics exciting?).

By the way...
... a scatterplot of the "weight-tumbling" data (Study B) would look like:

Which "dot" is Rollo?



5.2 POSTTEST (answers in back)

1. A teacher is interested in whether students' smoking habits relate to their attitudes toward school. Briefly describe a set of procedures through which she could investigate this question. (there is no one correct answer - numerous approaches are possible, depending upon how one interprets her interests.) What would be appropriate units of analysis? What variables would need to be assessed, and how would the assessment be structured?

2. A basketball coach is interested in determining whether height relates to point production for starting players on high school basketball teams. He measures a sample of five players on both variables and obtains the following results:

Name:	Height (in inches)	Points (average per game)
Rimm	83	28
Hook	80	20
Duncker	70	5
Block	76	14
Walker	69	7

Given these data, would you conclude that there is a relation present? If so, what type?

3. An animal researcher believes that the amount of aggression displayed by dogs (biting people, barking, etc.) is significantly related to the amount of punishment they receive. In testing this hypothesis, what would constitute the unit of analysis (that is, of what would the sample to be tested be comprised)?

4. (BONUS) A medical researcher believes that the frequency of daily fainting spells is positively related to the average daily temperature. What would be the appropriate units of analysis in conducting research on this question?

   Suppose the researcher's interest shifted to the question of whether people's tendency to faint on a given day of the year (Oct. 29) is positively related to their body temperature. What would the units of analysis be here?

5.3 Instructional Unit: Construction and Interpretation of Scatterplots

As suggested at the conclusion of the preceding unit, scatterplots provide a helpful and often highly interpretable visual representation of correlational data. By examining the pattern formed by the various points, one can obtain an immediate impression of the strength and direction of the relation between the two variables of interest.

Construction

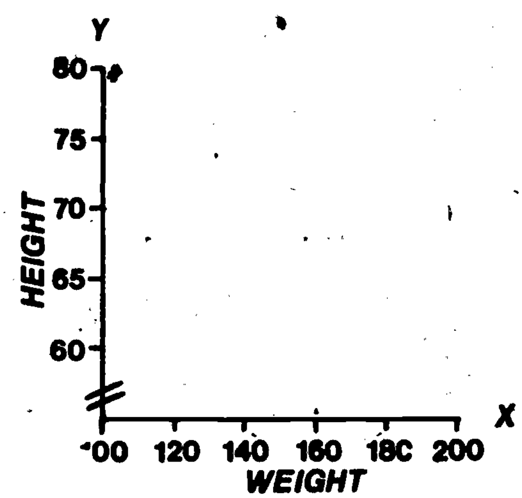
Note: Constructing a scatterplot is very simple, and it is quite possible that you picked up the rules from the previous section...or from another course. Take a quick look at the explanation that follows. If you feel that you already know it, save time by either skimming through or skipping it completely and branching to the section on interpretation.

Explanation

1. Draw a horizontal number line beginning with zero to be used as the X axis. Similarly, draw a vertical number line up from the origin of the X axis. Call the vertical axis the Y axis. Note that you are simply constructing the first quadrant of a typical coordinate system in two dimensions.

2. On each axis, place appropriate numeric labels that will permit you to fully account for the full range of variation for the particular variables being represented. For example, if your Y variable was height (the possible scores for which could range from 60 in. to 80 in.), you might divide your Y axis into equally spaced intervals of five units, starting at 60 and ending at 80. If your X variable was weight, with the range of scores being from 100 to 200, you might divide your X axis into equally spaced intervals of 20 units, starting at 100 and ending at 200. The intervals you finally select should depend upon how much information you want to provide in representing the data. For this particular illustration, an example might be:

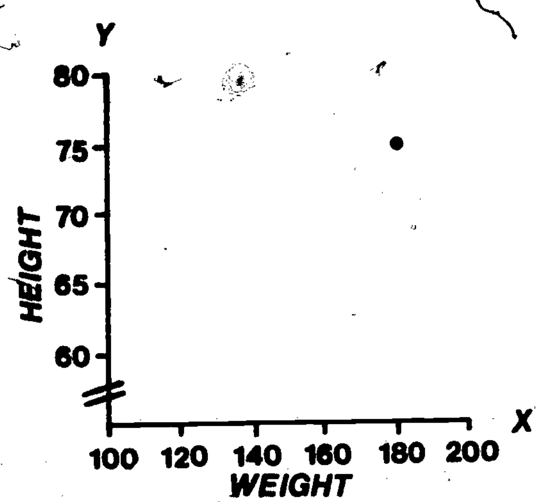
Note that the slash marks are included to show that parts of the X & Y axes are not present in the graph. Without this convention, it would be easy to distort relationships.



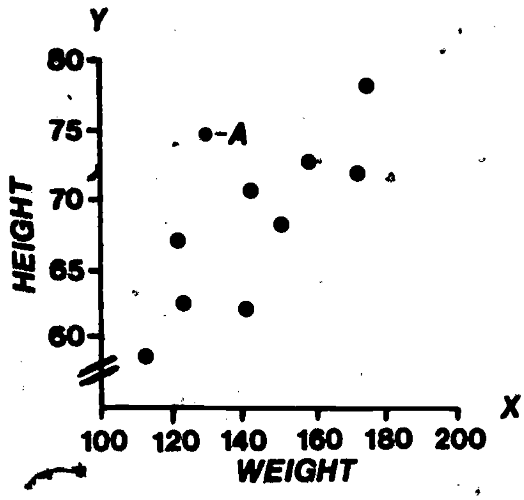
166

3. Once your axes are constructed, plot the individual pairs of scores by representing them in terms of "dots" or "points" appropriately positioned with respect to the X and Y axes. Each point represents a person (or thing) and how he or she performed on the X and Y variables. How to plot these points is easier to demonstrate than to describe, but one way would be to: a) start with the first person and look at his/her score on variable X; b) move along the X axis until you find the appropriate position for that score; c) keeping your pencil at that position, find the same person's score on variable Y; d) then move your pencil vertically upwards until it reaches the Y axis position that corresponds with the just determined Y score; e) place your "dot" where your pencil rests and; f) repeat these procedures for the remaining elements in your sample.

For example, let's suppose that the first person in the height/weight study received scores of 180 lbs. and 75 inches. The dot representing these scores would be positioned as follows: (Note that it corresponds to 180 units on the X axis, and 75 units on the Y axis.)



With ten people in the sample, the scatterplot of the height/weight study might look something like the one shown on the next page. For practice can you determine the height and weight scores of the person represented by dot "A"?



You should be able to determine that person A weighs 130 lbs. and is 75 inches tall (it's probably easier to find him on the graph than it is in real life).

End of Explanation

Interpretation of Scatterplots

By examining a scatterplot, it is usually possible to obtain a fairly accurate first impression of the direction and strength of the obtained relation between variables.

Direction: You will recall that the relation between two variables can be either positive or negative. A positive relation occurs when high scores on Variable X tend to be associated with high scores on Variable Y, and low scores on X tend to be associated with low scores on Y. For negative relations, the converse is true: high scores on X tend to correspond with low scores on Y; low scores on X tend to correspond with high scores on Y. This classification can be summarized as follows:

Positive Correlation (+)		Negative Correlation (-)	
X scores	Y scores	X scores	Y scores
high	high	high	low
low	low	low	high

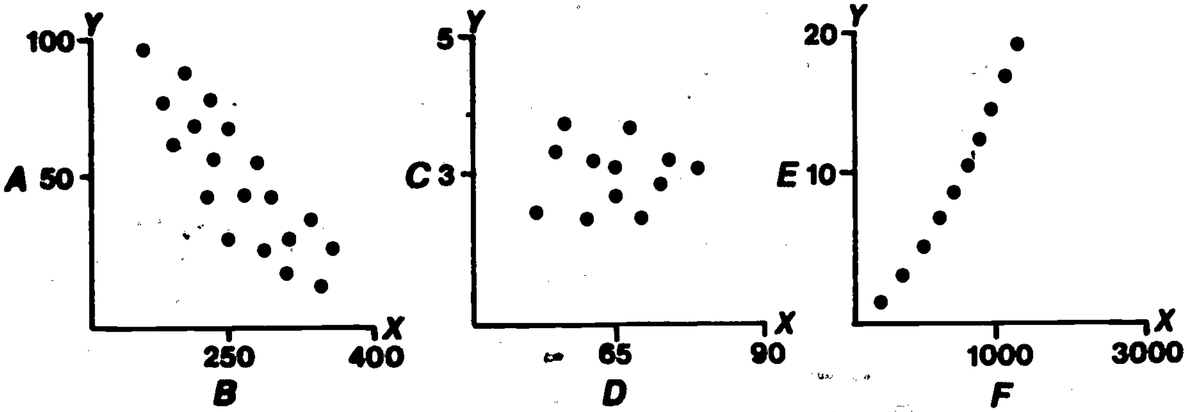
Rule: If two variables are positively correlated, the dots represented on a scatterplot will tend to slope upwards (from the bottom lefthand corner to the top righthand corner).

If two variables are negatively correlated, the dots represented on a scatterplot will tend to slope downwards (from the top lefthand corner to the bottom righthand corner).

Turn back one page and look at the scatterplot for the height/weight study. What type of relation does this appear to represent?

Answer: A positive relation because there is a clear tendency for the points to slope upwards.

Without becoming too concerned about actual interpretation, determine what types of relations appear to exist for the following three pairs of variables, as suggested by the scatterplots:



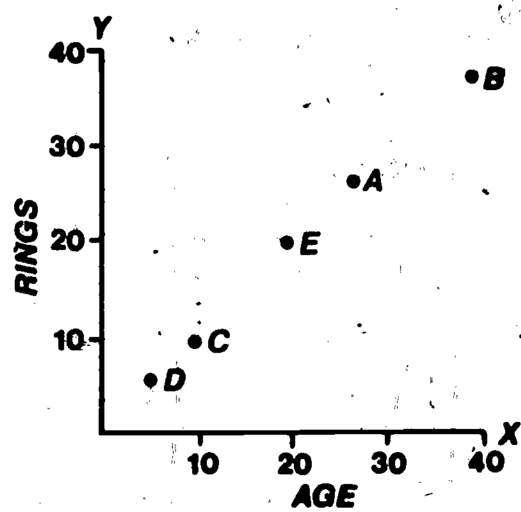
Judging from the scatterplots, variables A and B seem to be negatively related; C and D seem mysterious; and E and F appear to be positively related. But, wait, what about this C and D correlation--the points do not appear to have any slope whatsoever! The only conclusion that one can reach is that C and D are not correlated--there is little or no association between the values obtained on each.

Do you understand why particular slopes follow from negative and positive correlations? When there is a direct association between scores, i.e., a positive correlation, X scores near the origin (0, 0 point) will be associated with Y scores near the origin; X scores far from the origin will likewise be associated with Y scores far from the origin. The result is an upward slope. For a negative correlation, the reverse occurs; the result is a downward slope. If you're having trouble picturing this, you may want to read the additional explanation provided on the following page.

Additional Explanation: Ups and Downs on a Scatterplot

Why, for example, does a positive correlation produce an upward slope in the scatterplot points? The answer lies in the definition of positive correlation: High on X, High on Y; Low on X, Low on Y. Thus, many of the dots to be placed on the scatterplot will represent people whose X and Y scores are related in one of these two ways. Putting this in another way, as the dots are moved farther to the right on the X axis (indicating higher X scores), they are also moved farther upwards on the Y axis (indicating higher Y scores), and vice versa. The result is the general pattern of movement from lower left to upper right on the scatterplot.

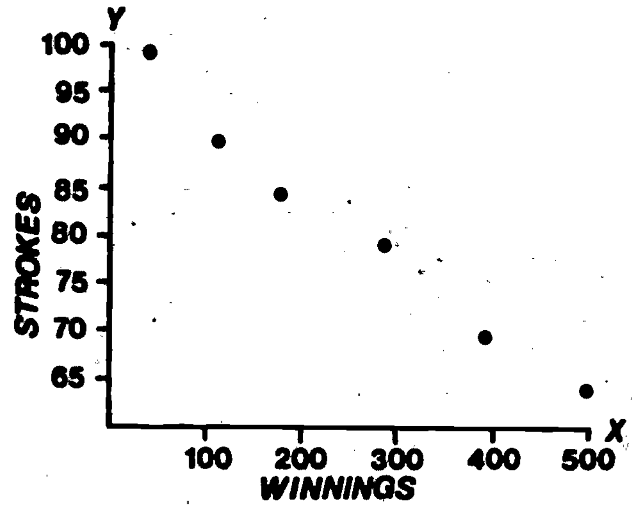
Let's see how this happens. Suppose every tree in a forest is tagged according to the year it was planted (that is, its age). We may then wish to correlate the relation between <u>age</u> (call this X) and the number of <u>annual rings</u> (call this Y). Intuitively speaking, this should produce a positive correlation since trees with low ages would have few annual rings, whereas those with high ages should have many annual rings. In fact, there should be a perfect correspondence, as reflected below.



A somewhat different pattern would emerge, however, if we plotted data that were negatively correlated. Take, for instance, a correlational study designed to determine the relation between <u>number of strokes</u> and <u>money earned</u> for all golfers participating in a local tournament. Hypothetical data from six participants could be as follows:

Golfer	Strokes	Winnings ($)
Potter	80	300
Green	85	200
Wood	70	400
Parr	65	500
Flagg	90	100
Ruff	100	50

Here we get the opposite (a downward slope) since the data are negatively correlated (high-low; low-high). The pattern is shown below. Note, for example, the one X score nearest the origin is associated with a Y score far from the origin. See how the downward slope develops?



In summary, when the points on a scatterplot slope upwards, the trend is indicative of a <u>positive</u> correlation; downwards is indicative of a <u>negative</u> correlation; no slope is suggestive of <u>zero</u> correlations.

<u>Strength of Relation</u>

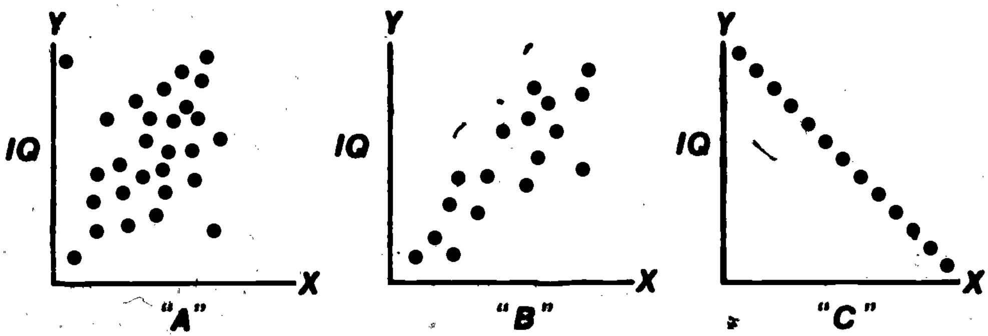
Examination of a scatterplot also provides an indication of the <u>strength</u> of correlations. Keep in mind that correlations do differ in terms of strength; for example, <u>hip width</u> and <u>shoe size</u> would both be correlated with <u>weight</u>. In most instances, however, we would be better off basing our judgments about weight on hip width than on shoe size.

How do you determine the strength of a correlation? As will be discussed in the following instructional unit, correlations can be assessed through very exact, precise means. But, if all one desires is a general overall impression, a quick glance at a scatterplot will usually suffice.

<u>Rule</u>: In examining a scatterplot - the more the points approximate a straight line, the <u>stronger</u> the relation between variables. Whether the slope is positive or negative is not relevant.

Given this simple rule, you are now in a position to evaluate the strength of correlations simply by looking at the corresponding scatterplots. Which of the three variables, A, B, or C, appears to correlate most strongly with IQ (turn page for graphs)?

ERIC



To feel completely accomplished and highly learned, you would have to answer C. Why? The strength of a correlation is independent of its direction. Thus, it is certain that the strongest relation is that between IQ and variable C, since the points more closely approximate a straight line relative to the patterns shown for the other two variables. Of variables A and B, the latter seems more closely related to IQ for the same reason. The best predictor of IQ is probably Variable C, the second best is probably Variable B, while the least effective is probably Variable A.

In summary, looking at the slope of the points in a scatterplot provides insight into the direction of a correlation; that is, whether it is positive or negative. Determining the degree to which the points in a scatterplot fall into straight line pattern provides insight into the strength of the correlation; that is, how closely the two sets of scores are associated.

5.3 POSTTEST (answers in back)

1. A researcher believes that the amount of money people earn (annual salary) is significantly related to their educational background (number of years of schooling). To test this hypothesis, he decides to correlate annual salary with years of schooling for a sample of 30-year-old males who reside in his home town. Although he probably should employ at least 25 subjects in his sample, for purposes of of convenience we will assume that he only tests six subjects.

The data he obtains is as follows:

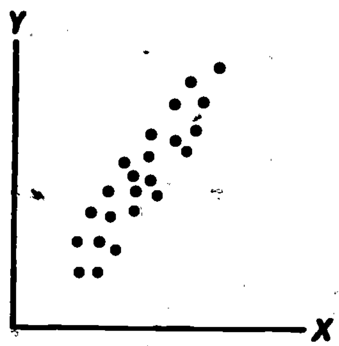
Subjects	Years of Education	Salary (in units of $1000)
#1	12	8
#2	17	17
#3	16	6
#4	18	16
#5	18	20
#6	14	9

a. Construct a scatterplot of these data (treat salary as the Y variable).

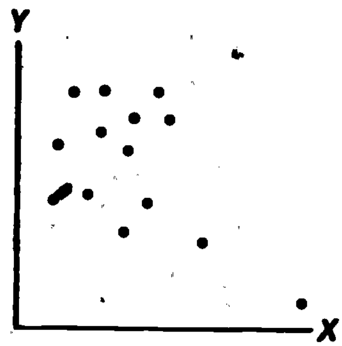
b. Does there appear to be a relation? If so, what type?

c. Suppose an additional subject, whose Education and Salary scores were 20 and 5 respectively, were to be included in the sample. Would this tend to strengthen or weaken the obtained relationship? Why?

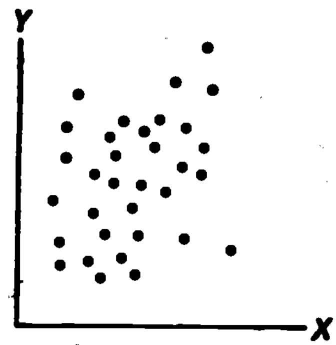
2. Which of the following three scatterplots represents the strongest relation between variables?



PLOT A

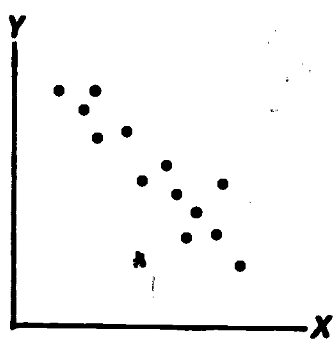


PLOT B

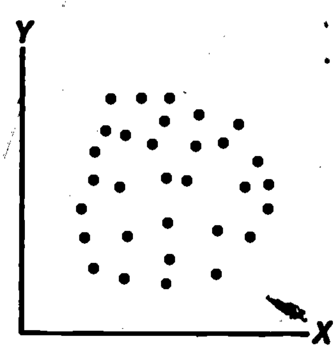


PLOT C

3. Briefly characterize the <u>above</u> scatterplots in terms of the direction and relative strength of the relations they depict. Do the same for the two scatterplots presented <u>below</u>.



PLOT D

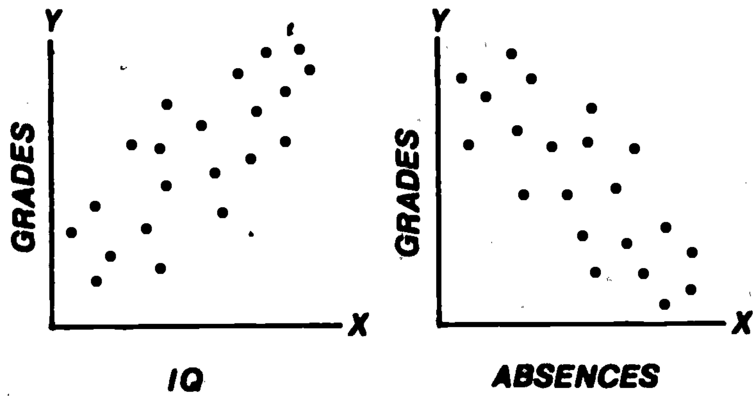


PLOT E

<u>Note</u>: Do not become concerned if your descriptions differ slightly from those used in the answer section. Determining whether a correlation is moderately strong, very strong, etc. involves a lot of subjectivity.

5.4 Instructional Unit: The Correlation Coefficient

Assuming you have understood the preceding sections, you should now be capable of doing many exciting things pertaining to the concept of correlation: You know what correlation implies, what types of relations are possible, and how to assess relations merely by examining the patterns represented on scatterplots. Given all these skills, perhaps you will be able to answer the following question: Of the two relationships shown below, which is probably stronger? Note in one instance we are correlating Grades with IQ and in the other we are correlating Grades with Number of Class Absences.



Perhaps you are inclined to pick IQ as the better predictor of (the more closely related to) grades because the relationship appears to be positive. This rationale would be incorrect since the strength of a relationship is independent of its direction. Then, perhaps you are inclined to pick absences as the more closely related since the scatterplot "appears" to resemble a straight line pattern more closely than is the case for IQ. Here, the rationale would be correct, but can you really be sure about your decision?

The conclusion to be reached is that scatterplots are useful from the standpoint of providing a general impression of what the data <u>probably</u> imply, but they cannot be used for making precise, certain judgments. Recognizing this, statisticians have derived what is called the <u>correlation coefficient</u>, which when computed, defines the direction and strength of a relation in clear, precise terms. More specifically, it is also often called the Pearson-product-moment correlation (since Pearson developed the formula). The correlational coefficient is symbolized as "r" when the data is obtained from samples, and "ρ" when obtained from populations.

<u>Definition</u>: The <u>correlation coefficient</u>, r, is a statistical index used to indicate the precise direction and strength of the relation being examined. It is expressed in the form of a numerical value which can range from +1.00 to -1.00.

ERIC

The sign of the coefficient "+" or "-" indicates the direction of the relationship, positive or negative.

The absolute value of the numerical component indicates the strength of the relationship. By absolute value is meant the value of the number without the sign, i.e., its distance from zero.

Example Problem: On the basis of the above definition, answer the following questions in regard to this series of r's:

-.26; +.09; -.78; -.47; +.51; -1.26; +.01

a) Which of the above coefficients could not possibly be obtained?

b) Which indicates the strongest positive relation?

c) The strongest negative relation?

d) The strongest relation overall?

e) Which, if represented on a scatterplot, would result in a pattern that most closely approximates a straight line?

Answers:

a) the $r = -1.26$ could not be obtained; it is beyond the possible range of -1.00 to +1.00. Somebody goofed in deriving it.

b) the strongest positive relationship is +.51--it has the highest absolute value.

c) the strongest negative relationship is $r = -.78$--it has the highest absolute value.

d) the strongest relationship overall is -.78--it has the highest absolute value of any of the coefficients listed.

e) the coefficient that would most closely approximate a straight line pattern is again, $r = -.78$--it indicates the strongest relationship (see Unit 5.3 for discussion of scatterplots).

A coefficient of +1.00 indicates a perfect positive correlation, which if represented on a scatterplot, would result in all points falling in a straight line pattern with an upward slope. A necessary (although not sufficient) condition for a perfect positive correlation is that the person who scores highest on Variable 1, scores highest on Variable 2; second highest on Variable 1, second highest on Variable 2; and so on. Examples would be the relationship between annual rings and the age of trees; and the relation between height in inches and height in centimeters for all students in a class.

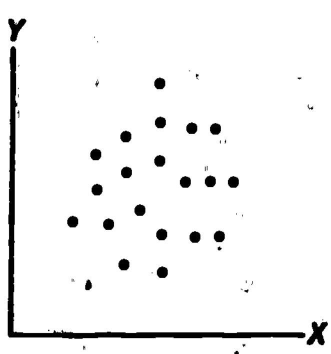
A coefficient of -1.00 indicates a perfect negative correlation, as would be represented by a downward straight line pattern on a scatterplot. A necessary (although not sufficient) condition for this type of relation is that the person who is highest on one variable is lowest on the other; the person who is second highest on one is second lowest on the other; and so on. An example would be the relation between multiple-choice test scores and the number of incorrect answers on the same test (think about it). Whenever you obtain either a perfect positive or perfect negative correlation, you are able to predict one variable from the other without error; that is, predictions will be perfect.

A coefficient of $r = 0.00$ indicates no relation between variables, as would be represented by a scatterplot which has no slope whatsoever. This occurs when there is no systematic correspondence between scores on the two variables; people who score high on one variable obtain all different score values on the other. Examples would be the relation between IQ and freckles; height and grades in Freshman English; length of name and visual acuity. A zero correlation indicates that knowing a person's score on one variable provides absolutely no basis for predicting his score on the other.

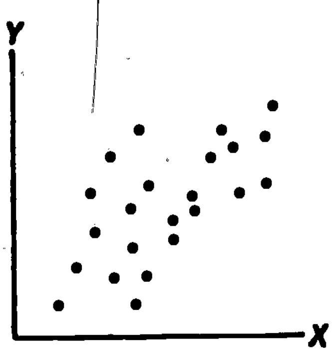
A coefficient greater than $r = 0$ and less than $r = +1.00$ (e.g., +.72, +.07, +.91) indicates an imperfect positive correlation, as would be represented by a scatterplot in which the points slope upward, but do not fall in a straight line. Examples would be the relation between height and weight; IQ and grades; etc.

A coefficient less than $r = 0$ but greater than $r = -1.00$ (e.g., -.68, -.02, -.95) indicates an imperfect negative correlation, as would be represented by a scatterplot in which the points slope downward, but do not fall in a straight line. Examples would be the relation between number of absences and grades; weight and ability to fit through small doors; darkness of skin and susceptibility to sunburn (note that very light people tend to burn more easily); etc.

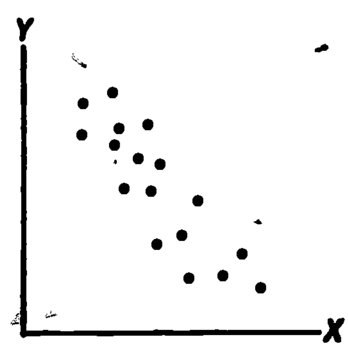
Some Examples:



$r = 0.00$



$r = +.80$



$r = -.90$

Final Note:

Correlation coefficients allow one to make an accurate and precise determination of the degree of relation between the two variables of interest. The sign of the coefficient indicates the direction, + or −; the absolute value of the number indicates the strength. Coefficients of ±1.00 represent a perfect correspondence between scores. By knowing the correlation coefficient, one can infer how accurately one variable can be predicted from the other.

Example:

Suppose that the correlation between high school grade point average (GPA) and college Freshman GPA is +.47. The correlation between ACT scores and Freshman grades is +.61. Which provides the better predictions of Freshman grades, high school grades or ACT scores? Apparently, it is the latter. Will predictions be perfect? No, but they will be more accurate than they would be without the ACT information.

5.4 POSTTEST (answers in back)

1. Briefly describe what is meant by the correlation coefficient. What values can it take? How are those values interpreted?

____

____

____

2. Match the following coefficients to the scatterplots shown below.

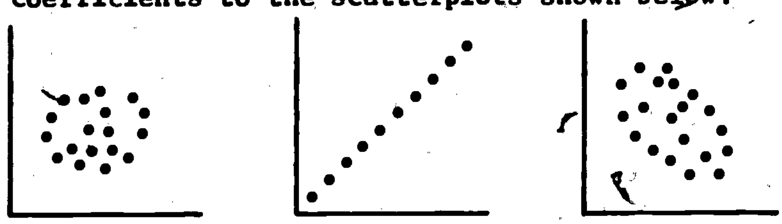
____ a. −1.00

____ b. −.50

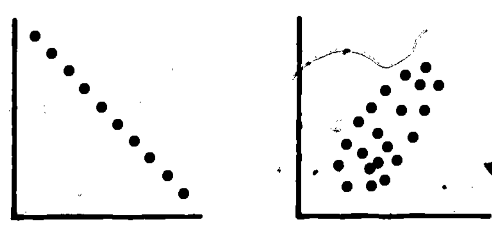
____ c. 0.00

____ d. +.50

____ e. +1.00



1 2 3



4 5

3. A researcher finds the correlation between amount of dating and grades to be $r = -.20$. He also finds that the correlation between number of absences and grades is $r = -.50$.

ERIC

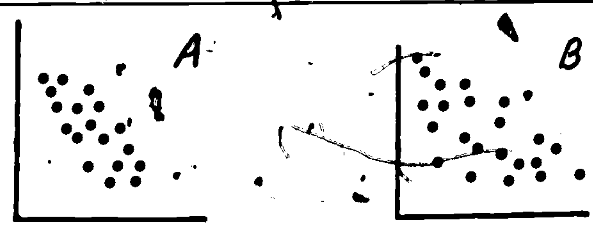
a. How would you characterize the two relations that were obtained (positive, negative, perfect, imperfect, etc.)?

b. Which is a better predictor of grades, dating or absences? How do you know?

c. On the basis of the results, how would you characterize the general grade pattern (high or low) for students who date frequently?

John is a heavy dater, not because he weighs 190 lbs., but because he generally has about 50 dates a week, more than one of which he sometimes absent-mindedly schedules for the same time. Based on the correlation, is it necessarily true that John's grades are among the lowest in his class?

d. Which of the two scatterplots would probably represent the dating-grades relation (as opposed to the absences-grades)? Why?



4. A researcher finds the correlation between years of education and annual income to be $r = .26$ for a random sample of males over 30 years of age, living in different areas of the country. In a separate study, he finds the correlation between distance of residence from a large city and annual income to be $r = -.39$ for the same sample.

a. What types of relations were obtained?

b. Which is likely to be the better predictor of income, years of education or distance from a large city? Why?

c. On the basis of the obtained relations, would income tend to be high or low for people with many years of education? Would income tend to be high or low for people who live far from a large city? Could you be certain that these tendencies would apply to all people? Why or why not?

d. If both relations were represented on scatterplots, which would more closely approximate a straight line pattern?

5.5 Instructional Unit: Computation of r

We have now reached the last step in our treatment of correlation; you know how to interpret correlation coefficients, but the question of how to compute them remains. Two formulas will be presented, one being more valuable from a conceptual than a computational standpoint, and the other the reverse. A third (special) formula will close the unit.

The z score formula

This formula involves computing the correlational coefficient by working with z scores obtained on each variable. Remember z scores (Unit IV)? Each raw score can be converted to a z score by the formula:

$$z = \frac{X - \bar{X}}{\sigma}$$

You should recall that z scores indicate how many standard deviations above or below the mean a certain raw score lies. Thus, a person whose raw score on a spelling test translates into $z = +3.00$ has performed quite well; one with a z score of -1.98, for example, has performed relatively poorly.

The z score formula for computing correlations is fairly easy to understand conceptually. It is cumbersome to use, however, unless the raw scores on each variable are already in z score form.

The formula is as follows:

$$r = \frac{\Sigma z_x z_y}{n}$$

What it means is that to compute the correlation coefficient, you must: (1) multiply each person's z score on variable X by his/her score on variable Y; (2) obtain the sum of these crossproducts;

and (3) divide the sum by n . The result will be r. Easy to use if the z scores are provided; cumbersome to use if only the raw scores are provided. (n= number of observations; i.e., people who were tested.)

Example

A teacher is interested in determining the relation between Math Achievement and Spelling Achievement for her class of 5th graders. She tests all her students ($n = 7$) on both variables. The tests are graded by computer, and the results she obtains include the z score for each raw score value. By having these z scores, she knows that a correlation coefficient can easily be computed by use of the z score formula. The results are as follows:

Child	z score on X (Math)	z score on Y (Spelling)	$z_x z_y$
A	1.40	1.00	1.40
B	1.00	1.40	1.40
C	.70	.70	.49
D	.00	.00	.00
E	- .60	-1.20	.72
F	-1.30	- .60	.78
G	-1.20	-1.30	1.56

$$\Sigma z_x z_y = 6.35$$

$$r = \frac{\Sigma z_x z_y}{n} = \frac{6.35}{7} = +.91$$

Note that the availability of z scores makes the computation of r amazingly simple, although the result, r = .91 (though accurate) might seem a little far-fetched. Do you think you'd obtain such a correlation between math and spelling achievement many times if you checked many classes?

If you think about it, it should be easy to determine (in a general sense) why the z score formula works. If positive z scores are associated with positive z scores, and negatives with negatives, we should end up with a positive correlation, right? Why does this happen with the formula? The answer should be understandable if you think about it: a positive z multiplied by a positive z yields a positive $z_x z_y$; similarly, a negative z multiplied by a negative z yields a positive $z_x z_y$. The result is all positive numbers in the righthand column, which, when added together will yield a positive $\Sigma z_x z_y$ and thus, a positive correlation. Watch how this happens in the sample data shown below, in which performances on Variable X correspond perfectly with performances on Variable Y (in terms of magnitude).

Example

Child	z_x	z_y	$z_x z_y$
A	+1.50	+1.50	+2.25
B	+ .50	+ .50	+ .25
C	.00	.00	.00
D	- .50	- .50	+ .25
E	-1.50	-1.50	+2.25
		$\Sigma z_x z_y$	= 5.00

$$r = \frac{\Sigma z_x z_y}{n} = 5/5 = +1.00$$

When performances on both variables are perfectly matched, we obtain a correlation coefficient of +1.00. What would happen if high scores tended to be paired with high scores (and lows with lows), but the matching was less than perfect? We would expect an imperfect positive correlation--high z scores would be multiplied by "relatively" high z scores, but the resultant products would have somewhat lower values than in the previous example. Similarly, low (negative) z scores would be multiplied by "relatively" low z scores, yielding positive products that are slightly lower in value than those obtained when the correspondence is perfect. Let's see how this occurs.

Example

Student	z_x	z_y	$z_x z_y$
A	+1.50	+ .50	+ .75
B	+ .50	+1.50	+ .75
C	.00	.00	.00
D	- .50	-1.50	+ .75
E	-1.50	- .50	+ .75
		$\Sigma z_x z_y$	= 3.00

$$r = \frac{3.00}{5} = +.60$$

Thus, a fairly strong but imperfect correspondence between scores on both variables yielded a moderately high positive correlation. Note that the z_x and z_y above scores are the same as in the preceding example, but the pairings have been changed.

Why do we obtain a zero correlation coefficient when there is absolutely no association between variables? In terms of the z-score formula, we would be multiplying positive values by negative (or zero) values in some cases, and by positive values in other cases. The result would be some positive $z_x z_y$'s and some negative $z_x z_y$'s which, when summed together, would mostly cancel each other out. Using the same z scores as above, but different pairings, let's see what occurs when performances on X and Y show almost no relation to one another.

Example

Person	z_x	z_y	$z_x z_y$
A	+1.50	- .50	- .75
B	.50	+ .50	+ .25
C	.00	+1.50	.00
D	- .50	-1.50	+ .75
E	-1.50	.00	.00
		$\Sigma z_x z_y$ =	.25

$$r = \frac{.25}{5} = +.05 \text{ (very low)}$$

Finally, why do we obtain a negative correlation when the relationship between variables is inverse (i.e., high with low; low with high)? In terms of the z score formula, we would be multiplying high (positive) scores by low (negative) scores and vice versa. The result would be all (or mostly all) negative values in the $z_x z_y$ column which, when summed, produce a negative value. Watch how this happens in the following example in which the obtained relation shows a perfect inverse correspondence between scores.

Example

Subject	z_x	z_y	$z_x z_y$
A	+1.50	-1.50	-2.25
B	+ .50	- .50	- .25
C	.00	.00	.00
D	- .50	+ .50	- .25
E	-1.50	+1.50	-2.25
		$\Sigma z_x z_y$	-5.00

$$r = \frac{-5.00}{5} = -1.00 \text{ (perfect negative)}$$

The above examples are intended merely to illustrate how the z score formula works. If you understand it conceptually, reconstructing the formula for actual use should present little problem. If you do not understand it conceptually, there is no need to become concerned--such understanding may come in time. The question for now is--if given the formula and pairs of z scores, could you use it to compute the correlation coeffecient? If you answer "yes" feel free to continue further; an answer of "no" implies the need to review the five example problems presented above.

In conclusion, the z score formula constitutes a fairly simple means of computing r, provided that all raw scores have been converted into z scores. If such is not the case, a more appropriate formula to employ is the "raw score" formula to be described below.

Raw Score Formula for Computing r

At first glance the raw score formula looks formidable and threatening. In fact, it has the potential to throw apprehensive students into shock. The effects, however, are reversible--just make sure oxygen is near before you read further. Seriously though, the raw score formula is merely a derivation of the z score formula just explained; thus, for any set of data it will produce the exact same result. Unlike the z score formula, it is designed for use with distributions that contain only the raw score, i.e., <u>actual scores</u>. When you examine the raw score formula below, your first reaction may be: "none of this for me. I'm going to use the z score formula for computing all correlations!" Easy to say, but very difficult to do if you are not given z scores. With a hand calculator using the raw score formula should present very little problem, provided that you carefully derive all the values that must be entered into the equation. Good work habits are very important.

The formula is as follows:

$$r = \frac{n\Sigma XY - \Sigma X \Sigma Y}{\sqrt{n\Sigma X^2 - (\Sigma X)^2}\ \sqrt{n\Sigma Y^2 - (\Sigma Y)^2}}$$

Collect your senses, and try to overcome any inclination to physically abuse this book. Ask yourself what types of values are needed for entry into the formula. If you do this, perhaps you'll realize that the requirements are really quite simple, although the mathematics do leave room for error. To work the formula, all you need to compute are:

a) ΣX : simple to do; add up all the scores on the X variable

b) ΣY : just as simple; add up all Y scores

c) ΣX^2 : add up the <u>squares</u> of all X scores

d) ΣY^2 : add up the <u>squares</u> of all Y scores

e) ΣXY : a little more complicated, but not much; multiply each X score by its corresponding Y score. Then add up the resultant crossproducts.

Once you do these five things, you are ready to work the equation. If you are careful and systematic in entering values and manipulating them arithmetically, there is no reason to anticipate anything but perfection in your final answer.

Example

Compute the correlation coefficient for the following test scores. (Remember, for present purposes, the formula does not have to be memorized; refer to it freely without experiencing guilt.)

Child	Arithmetic (X)	Reading (Y)
A	3	6
B	2	4
C	4	4
D	6	7
E	5	5
F	1	3

Obtain ΣX and ΣY by summing down the appropriate columns. Add new, appropriate columns for X^2, Y^2, and XY as shown. By summing down such columns, we could obtain the desired formula entries. Also, note that some of the formula entries must be multiplied by n, the number of scores (6).

	Arithmetic (X)	Reading (Y)	X^2	Y^2	XY
A	3	6	9	36	18
B	2	4	4	16	8
C	4	4	16	16	16
D	6	7	36	49	42
E	5	5	25	25	25
F	1	3	1	9	3
	$\Sigma X=21$	$\Sigma Y=29$	$\Sigma X^2=91$	$\Sigma Y^2=151$	$\Sigma XY=112$

Now we can begin to simplify the equation; we start with:

$$r = \frac{n\Sigma XY - \Sigma X\Sigma Y}{\sqrt{n\Sigma X^2 - (\Sigma X)^2}\ \sqrt{n\Sigma Y^2 - (\Sigma Y)^2}}$$

Substituting for n, we get:

$$r = \frac{6\Sigma XY - \Sigma X\Sigma Y}{\sqrt{6\Sigma X^2 - (\Sigma X)^2}\ \sqrt{6\Sigma Y^2 - (\Sigma Y)^2}}$$

Substituting for $\Sigma XY, \Sigma X, \Sigma Y, \Sigma X^2$ and ΣY^2 we get:

$$r = \frac{(6)\ (112) - (21)\ (29)}{\sqrt{(6)\ (91) - (\Sigma X)^2}\ \sqrt{(6)\ (151) - (\Sigma X)^2}}$$

Substituting for $(\Sigma X)^2$ and $(\Sigma Y)^2$ we get:
(21x21) (29x29)

$$r = \frac{(6)\ (112) - (21)\ (29)}{\sqrt{(6)\ (91) - 441}\ \sqrt{(6)\ (151) - 841}}$$

$$r = \frac{63}{(10.25)\ (8.06)} = .76 \text{ (a fairly high positive relation)}$$

One More:

A teacher wishes to correlate number of absences during the week with end-of-week quiz scores. The data he/she obtains is shown below. For practice problems we are restricting sample size considerably for the purpose of facilitating computations. Also, you may have noted that in all problems, one-digit numbers have been used for score values. Again, this is only to make the computations easier. In actual practice, you may be dealing with test scores that are expressed in two or even three digits, and thus, computing r should not ever be attempted without a calculator or computer.

Child	Absences (X)	Quiz Score (Y)	X^2	Y^2	XY
A	5	1	25	1	5
B	0	5	0	25	0
C	1	4	1	16	4
D	3	3	9	9	9
E	2	3	4	9	6
	$\Sigma X=11$	$\Sigma Y=16$	$\Sigma X^2=39$	$\Sigma Y^2=60$	$\Sigma XY=24$

$$r = \frac{n\Sigma XY - \Sigma X \Sigma Y}{\sqrt{n\Sigma X^2 - (\Sigma X)^2}\ \sqrt{n\Sigma Y^2 - (\Sigma Y)^2}}$$

$$r = \frac{(5)\ (24) - (11)\ (16)}{\sqrt{(5)\ (39) - 121}\ \sqrt{(5)\ (60) - 256}}$$

$$r = \frac{120 - 176}{\sqrt{74}\sqrt{44}}$$

$$r = \frac{-56}{57.06} = -.98 \text{ (high negative relation)}$$

Correlation Coefficient for Ranked Data: r_s

On occasion, the data you wish to correlate may be presented in the form of ranks rather than "regular" (i.e., interval or ratio) scores. For example, suppose that instead of receiving scores on math and spelling tests, students were rank-ordered by their teachers in each subject. Mary, for example, was ranked #1 in math and #5 in spelling, Charles was #6 in both. Ranks, just like numerical scores, can be correlated. In fact, a British psychologist named Charles Spearman derived a special formula for use with ranks. We call the correlation coefficient that it yields the Spearman correlation coefficient (what else?), and symbolize it as r_s (note the "s" stands for Spearman. Tricky!). For all practical purposes, r_s behaves like, and is directly comparable to, plain old r. The formula is as follows:

$$r_s = 1 - \frac{6\ \Sigma D^2}{n(n^2-1)}$$

where D refers to the difference between each pair of ranks, and 6 is a constant that is always applied. The symbol n, of course, refers to the number of pairs of ranks.

Here's an example of how students were ranked in (1) neatness and (2) popularity by their homeroom teacher, Mrs. Frinkleworth:

Student	Neatness	Popularity	D	D^2
Bluto	6	6	0	0
Gloria	1	2	-1	1
Frank	5	4	1	1
Aaron	2	1	1	1
Sylvia	4	3	1	1
Grace	3	5	-2	4
				$\Sigma D^2=8$

$$r_s = 1 - \frac{6\Sigma D^2}{n(n^2-1)}$$

Substituting:

$$= 1 - \frac{6(8)}{6(36-1)} = 1 - \frac{48}{6(35)}$$

$$= 1 - \frac{48}{210} = 1 - .229 = .77$$

All we did was: (1) take the difference (D) between each person's rankings, (2) square those differences (D^2), and (3) sum them to get ΣD^2. The rest simply involves substituting this value and that for n in the formula. Our result was $r_s = .77$, a fairly strong positive correlation. For this particular class, rankings in popularity tended to be associated with rankings in neatness.

Note: Suppose the teacher couldn't decide who to rank lowest in popularity, Bluto or Frank. If the two boys are tied, what rank should each receive? Answer: Assign each the average of the ranks involved in the tie, in this case (Ranks 5 + 6)/2 = 5.5 for each.

Suppose two people tie for first? What rank do they get? (Ranks 1 + 2)/2 = 1.5 for each.

Bonus: Suppose three people tie for 2nd? Each gets (Ranks 2 + 3 + 4)/3 = 3 for each.

Final Notes:

1. The Pearson correlation coefficient (r) measures the degree of relationship between two variables. Two formulas for its computation were described, the z score formula and the raw score formula.

2. If you wish to correlate ranked data, use the Spearman formula (r_s). The values it gives have the same range (-1.00 to +1.00) as the Pearson and are interpreted in the same way.

3. Speaking of interpretation, what constitutes strong or weak correlations? Well, that really depends on what you're correlating. In educational and in behavioral research, correlations greater than .50 in absolute value are generally considered large. In the physical sciences, where things are supposed to operate in an orderly fashion, correlations as large as .90 may be considered low. Don't think of correlations as proportions or percentages--.80, for example, isn't 80% as strong as its maximum level. As you gain experience in using correlation in your field, you will begin to gain a perspective on what levels to expect. When you get to Unit IX in this book, you'll find that there are ways of testing the statistical significance of correlations, determining the probability that the coefficient you obtain differs from zero, given your sample size and other factors. When you get to higher level statistics courses, you'll find how correlation relates in mathematical and conceptual ways to numerous analytical procedures (multiple regression, discriminant analysis, factor analysis, analysis of covariance). An exciting world awaits you (after you part with us).

5.5 POSTTEST (answers in back)

1. Give a logical explanation for why the z score correlation formula's mathematical properties will yield:

   a) a positive correlation when variables are directly (positively) related.

   b) a negative correlation when variables are inversely related.

   c) no correlation when variables are unrelated.

2. A school administrator is interested in the relation between parents' income and the scores their children obtain on a nationally administered scholastic aptitude test for third graders. He randomly selects nine students from his district and tabulates their parents' annual incomes along with their test scores in z score form: The resultant data is shown below. What is the correlation between these variables? (Carry out this problem to completion.)

Student	Parents' Income (z_x)	Test Score (z_y)
A	-1.50	-1.20
B	-1.20	-1.50
C	- .80	- .40
D	- .40	+ .80
E	.00	.00
F	+ .40	- .80
G	+ .80	+ .40
H	+1.20	+1.50
I	+1.50	+1.20

Feel free to turn back and re-examine the z score formula.

3. The table shown below gives the raw scores 5 students received on reading and spelling tests. Compute the correlation coefficient for these measures. (This problem does not have to be carried out to completion - it will suffice if you merely enter the appropriate numerical values into the raw score equation.)

Student	Reading (X)	Spelling (Y)
J	3	5
K	1	2
L	2	5
M	4	3
N	1	1

4. In the Miss Soybean Contest, the 7 finalists were rank-ordered on the basis of congeniality and muscle tone. Use the Spearman Formula to determine the correlation between the two rankings. What do the results imply?

Contestant	Congeniality	Muscle Tone
A	1	6
B	3	4
C	6	5
D	4	3
E	7	2
F	2	7
G	5	1

Unit V Review Test
(answers in back)

Select one answer for each of the following:

_____1. Correlational analysis is used to
a. determine the extent to which variables can be predicted from constants.
b. determine the effect of the independent variable on the dependent variable.
c. determine the relationship between two variables.
d. determine the relationship between two constants.
e. none of the above.

_____2. Which of the following statements is true given that a correlational study is to be conducted using 30 people (n=30)?
a. Half of the observations (i.e., 15) will provide scores on variable X, whereas the other half will provide scores on variable Y.
b. The number of points on a scatterplot will be 30.
c. The sample size will be 30.
d. The size of the sample is insufficient to permit a correlational analysis.
e. b. and c. are correct.

_____3. A correlation is determined for the length of days (sunrise to sunset) and the daily outside temperature at 12:00 noon. The units of analysis (i.e., the sample from which data are obtained) would consist of
a. people
b. days
c. thermometers
d. clocks
e. none of these

_____4. The pattern represented on a particular scatterplot shows points that slope downward from the upper lefthand side of the graph to the lower righthand side. Many of the points deviate from positions that would make up a straight line pattern. Based on this scatterplot, one could conclude that the correlation between the variables involved was
a. a perfect positive
b. an imperfect positive
c. a perfect negative
d. an imperfect negative
e. zero

18J

_____5. Which of the following sets of variables would yield a correlation of -1.00?

a. Height and gymnastics ability for a sample of eighth-grade girls.
b. Driving speed and gas consumption for a sample of cars traveling the same 10-mile stretch on an interstate highway.
c. Body weight and physical coordination for a sample of adults who consumed 12 ounces of alcohol.
d. Distance already traveled and distance still remaining to finish for a sample of drivers two hours into a 500 mile race.
e. Number of pounds of food consumed during a week and total calorie intake for a sample of high school students.

_____6. For a perfect positive correlation to be obtained, which of the following conditions must prevail?

a. The highest person on variable X must also be the highest on variable Y.
b. When plotted, all points must fall on a straight line.
c. The direction of the scatterplot must be from lower left corner to upper right corner.
d. The lowest person on Y must also be the lowest on X.
e. All of the above.

_____7. If, for a sample of 1000 physically normal adults, right leg length were to be correlated with left leg length, the resultant coefficient would probably be closest to

a. -.60
b. -1.00
c. +.54
d. +.98
e. -.01

_____8. A businessman finds that the correlation between age and productivity of his employees is $r = .50$. A correct conclusion would be

a. Age is a cause of good productivity.
b. When you produce more, you age faster.
c. Age and productivity are related.
d. Young workers tend to be more productive.
e. Older workers should be retired immediately.

_____9. Number of hours devoted to study correlates with high school grades at $r = +.20$. Number of TV shows watched at night correlates with high school grades at $-.40$. If these were "real" results, which variable would be the better predictor of grades?

a. Study hours
b. Number of TV shows
c. Both would be equally effective
d. An answer cannot be determined without additional data.

190

_____10. Suppose that running endurance (i.e., distance one can run at the pace of 9 min. per mile) correlates positively with the amount of time one can hold his breath under water. Tim is a person who can run great distances at 9 min. per mile, but can hold his breath for only a fairly short time. His addition to the sample would

a. weaken the correlation.
b. strengthen the correlation.
c. have no effect on the correlation.
d. an answer cannot be determined without additional data.

Check the answers in the back and then rate yourself.

Number Correct	Scoreboard
9 - 10	Excellent
7 - 8	Fair-Good
5 - 6	Poor
Below 5	Horrendous

UNIT VI

PROBABILITY

A. General Objectives

Remember Unit IV on the normal curve? Whether you realized it at the time, the area problems you were working are probability problems, and the normal curve is, in essence, a probability model. Given the importance of probability to statistics (and you will see this come into play quite strongly in Units VII-VIII), this unit provides an orientation to this subject. We're betting that you'll "probably" benefit from the experience.

B. Specific Objectives

6.1 Define probability

6.2 Compute probabilities on a single trial using (a) addition rule for mutually exclusive events and (b) addition rule for non-mutually exclusive events

6.3 Compute probabilities over multiple trials using (a) multiplication rule for independent trials, and (b) multiplication rule for dependent trials

6.1 Instructional Unit: Probability: What is it?

Let us get right to the point by presenting a definition:

> The probability that a particular event will occur is the relative frequency with which that event can be expected to occur.

Sound a bit circular? The key term here is relative frequency. If there are 30 patients seeking medical attention and 12 are males, the relative frequency of males is 12/30 or 2/5 or .40. We refer to the total group as the sample space, the total number of events in question. So, we determine probability by the formula:

$$P(A) = \frac{n(A)}{n}$$

Where $n(A)$ is the number of instances of Event A, and n is the size of the sample space. $P(A)$, therefore, represents the relative number of times Event A is represented in the sample space, i.e., its relative frequency.

Unlike previous units, this one will de-emphasize discussion, and rely mostly on presenting the material through rules and examples. Hope you like it!

Example 1: To return to our original example, but using different numbers, suppose that a group of 50 patients contains 12 who need immediate medical attention. What is the probability that if a patient is selected at random to be treated first, he/she will be one of those 12? What we're really asking is "what is the relative frequency of patients needing immediate attention (in that group)?"

$$P(\text{Im. Atten.}) = \frac{n(\text{Im. Atten.})}{\text{all patients}}$$

Since n(Im. Atten.) = 12 and since "all patients" or $n = 50$, we substitute to get:

$$P(\text{Im. Atten.}) = 12/50 = 6/25 \text{ or } 24\%$$

Thus, there is a 24% chance (6/25) of selecting one of those patients.

Example 2: Teachers need to be aware of guessing factors on their tests as such could result in invalid results for certain students. Suppose that Mrs. Jones asks a multiple-choice question that has one correct answer and four distractors (incorrect answers). Moose Crenshaw, who is allergic to studying, takes a random guess by picking alternative "b." Since we don't know the actual position of the correct answer, what can be said about the probability of Moose's guess being correct? Set-up is

$$P(\text{correct}) = \frac{n(\text{correct})}{n}$$

Where n(correct) is, in this case, the number of correct alternatives (there's 1) and n is the total number of alternatives (5).

Ans: $P(\text{correct}) = 1/5$ or 20%

Moose's chances aren't good. He'd be better off with a true-false test.

Example 3: Medical personnel always have to worry about the condition of their equipment, which is often beyond their control. In a new shipment of 15 blood pressure gauges, 8 are defective and give slightly inaccurate readings. Nurse Needles selects one at random to take a reading on patient Jones. What is the probability that Needles has picked one of the defective instruments?

Ans: $P(\text{Defect}) = \frac{n(\text{Defect})}{n}$

$P = 8/15$ or 53.33%

6.1 POSTTEST (answers in back)

1. Define what probability is in your own words.

____

____

2. In a class consisting of five students, two are females and three are males. What is the probability of selecting a male student if a last name is called out at random?

3. In a group of 60 people, 10 have type O blood, 30 have type A blood, 10 have type B blood, and 10 have type AB blood. If one of them gets ill and needs a transfusion, what is the probability that the type of blood needed will be type O?

6.2 Instructional Unit: Computing Probabilities on a Single Trial

Before beginning our discussion of single trial probability, it would probably be helpful to call your attention to some key vocabulary that will appear in definitions of the two rules to be covered. We will define them here once and then again when the rules are actually introduced. It may not be that important to study them (or worse, memorize them)now. They are presented mostly to provide an organizer of sorts:

Mutually Exclusive: A condition that prevails when two (or more) events cannot occur together; i.e., cannot take place at the same time.

Example: You purchase a prescription from your pharmacist. The specific medication purchased can be a liquid or a pill, but not both a liquid and a pill (at least at the same time).

Nonmutually Exclusive: Well, would you be satisfied if we said it is the opposite of mutually exclusive? Probably not. Nonmutually exclusive refers to a condition where the events in question can occur together, i.e., can take place at the same time.

Example: You are interested in the results of tests involving the effects of new drugs in treating laryngitis. You are also interested in the effects of drugs for treating bronchitis. You select a report on a particular drug at random. Could it be a drug that is effective for treating both laryngitis and bronchitis? Of course...those conditions are nonmutually exclusive.

Rule 1: Addition Rule for Mutually Exclusive Events

When events A and B are mutually exclusive, the probability that either A or B will occur is $P(A \text{ or } B) = P(A) + P(B)$

Thus, all we are doing is adding the separate probabilities associated with events A and B.

Example 1: In a math class of 20 students, 5 are high achievers, 12 are middle achievers, and 3 are low achievers. If a student is selected at random, what is the probability that he/she will be either a high achiever or a middle achiever?

$$P(\text{High or Middle}) = P(\text{High}) + P(\text{Middle}) = \frac{5}{20} + \frac{12}{20} = \frac{17}{20} \text{ or } 85\%$$

Example 2: A medical staff has 6 wards under its reponsibility, 4 involving geriatric patients and 2 involving pediatrics. If a supervisor randomly selects one ward to inspect, what is the probability of its involving either geriatrics or pediatrics?

$$P(\text{P or G}) = P(P) + P(G) = \frac{4}{6} + \frac{2}{6} = \frac{6}{6} = 1 \text{ or } 100\%$$

(Since all the wards involve either pediatrics or geriatrics, the probability is 1 (100%) that one of those types will be inspected.)

Example 3: There are 12 student desks in the classroom: 3 have broken seats and 5 others have broken tops. If a student selects a desk at random, what is the probability of either its seat or top being broken?

$$P(\text{Seat or Top}) = P(\text{Seat}) + P(\text{Top}) = \frac{3}{12} + \frac{5}{12} = \frac{8}{12} \text{ or } \frac{2}{3} \text{ or } 67\%$$

Example 4: Of every 100 maternity patients who are interviewed by the nursing staff at City Hospital, 50 indicate the desire for natural childbirth,25 prefer general anesthesia, and 25 prefer local anesthesia. A nurse meets a patient and, without knowing her position, makes derogatory comments about using anesthesia. What is the probability of that patient being one who actually prefers either local or general anesthesia?

$$P(\text{Local or Gen.}) = P(L) + P(G) = \frac{25}{100} + \frac{25}{100} = \frac{50}{100} \text{ or } \frac{1}{2} \text{ or } 50\%$$

Example 5: A teacher grades test results such that of her 21 students, 5 receive A's, 6 receive B's, 6 receive C's, 2 receive D's, and 2 receive F's. On this particular test, what was the probability of receiving either a D or a F?

$$P(\text{D or F}) = P(D) + P(F) = \frac{2}{21} + \frac{2}{21} = \frac{4}{21} \text{ or } 19\%$$

Rule 2: Addition Rule for Nonmutually Exclusive Events

When events are nonmutually exclusive, $P(\text{A or B}) = P(A) + P(B) - P(A+B)$

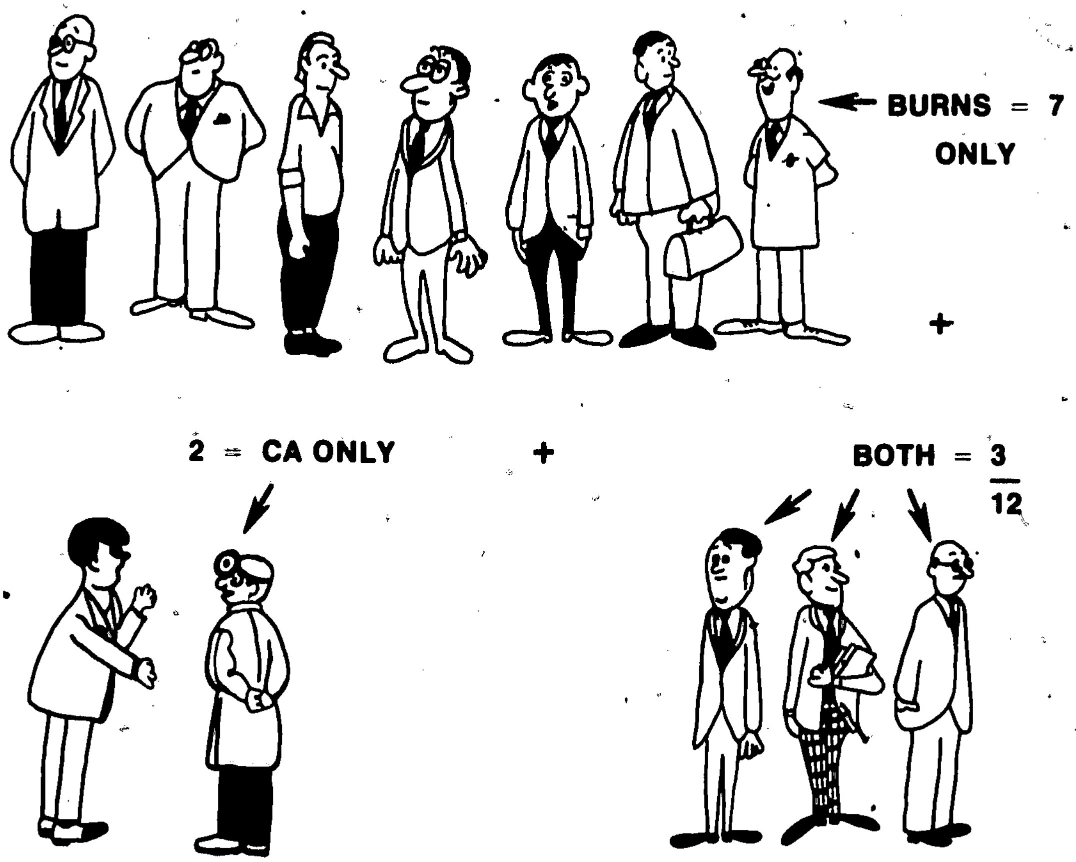
Once again, we are determining the probability of selecting either A or B on a single draw. The change, however, is due to the condition of events A and B not being mutually exclusive; i.e. they can occur together. As a result, a new expression, (A + B), enters into the formula. It refers to the probability of the outcome containing both A and B. The logic of the formula should become clearer as we work some examples.

Example 1: In a staff of 25 physicians, 5 have had special training in cardiac arrest (CA) and 10 have had special training in second and third degree burns. Of these, 3 have had special training in both. If a physician is selected at random, what is the probability that he/she has had training in CA or serious burns or both?

$$P(\text{CA or Burns}) = P(\text{CA}) + P(\text{Burns}) - P(\text{Both})$$

$$= \frac{5}{25} + \frac{10}{25} - \frac{3}{25} = \frac{12}{25} \text{ or } 48\%$$

Why is the P(Both) subtracted? Those 3 physicians are part of the group of 5 with CA training, right? But, they are also part of the group of 10 receiving burn training, right too! So......we are subtracting them so they won't be counted <u>twice</u>. Look below.



Example 2: You have a class of 14 students, 3 of whom have problems in reading and 6 have problems in math. Of these, 2 have problems in both reading and math. What is the probability of selecting a student who has problems in reading or math or both?

$$P(\text{Read. or Math}) = P(\text{Read.}) + P(\text{Math}) - P(\text{Both})$$

$$= \frac{3}{14} + \frac{6}{14} - \frac{2}{14} = \frac{7}{14} \text{ or } \frac{1}{2} \text{ or } 50\%$$

Again, do you see why the P(Both) is subtracted? Those 2 students help make up the group with reading problems and also the group with math problems. We subtract them so that they're <u>not</u> counted twice.

Example 3: There are 20 post-operative patients being monitored. Following rounds, 8 are identified as showing abnormal respiration patterns and 8 are identified as having uneven pulse rates. Of these, 5 are evidencing both types of problems. Naturally, such problems are cause for concern. What is the percentage (probability) of patients having respiratory problems or pulse rate problems or both?

$$P(\text{Resp. or Pulse}) = P(R) + P(P) - P(\text{Both})$$
$$= \frac{8}{20} + \frac{8}{20} - \frac{5}{20} = \frac{11}{20} \text{ or } 55\%$$

Example 4: Four types of bacteria are identified in a sample of city drinking water: 1 is shown to produce acute gastroenteritis and 2 produce allergy-type symptoms. Of these, 1 produces both gastroenteritis and allergy-type symptoms. If you were told that one type of bacterium was concentrated in your water, what is the probability of it being a type that produces gastroenteritis or allergy-type symptoms or both?

$$P(\text{Gastro. or Allergy}) + P(G) + P(A) - P(\text{Both})$$
$$= \frac{1}{4} + \frac{2}{4} - \frac{1}{4} = \frac{2}{4} \text{ or } \frac{1}{2} \text{ or } 50\%$$

Example 5: Of 64 third graders, 5 have a history of frequent absenteeism and 15 have a history or behavior problems. Of these, 3 have a history of both types of problems. For this student population, what is the probability of a student having a history of attendance problems, behavior problems, or both?

$$P(\text{Absentee or Behav.}) = P(\text{Absentee}) + P(\text{Behav.}) - P(\text{Both})$$
$$= \frac{5}{64} + \frac{15}{64} - \frac{3}{64}$$
$$= \frac{17}{64} \text{ or } 27\%$$

6.2 POSTTEST (answers in back)

1. Indicate whether each of the following pairs of categories (events) would be mutually exclusive or nonmutually exclusive for individuals.

   a. The sex of patients in a ward: male or female.
   b. Courses that A's are received in: history or math.
   c. Membership in different medical societies: national or state.
   d. Grade on yesterday's math exam: A or B or C.
   e. Type of class scheduled for 1 p.m. today: English or home economics.

2. Find the probability for each of the following:

   a. Of a D or an F in biology, given that in a class of 25 students, F's = 3, D's = 5, C's = 7, B's = 1, and A's = 9.
   b. Of a student participating in a sport or a club or both after school, given that in a sophomore class of 210, 22 students are in clubs, 60 play sports, and 2 participate in both.
   c. Of a person exposed to a particular pathogen developing anemia, ulcers, or both, given that out of 65 people formerly exposed, 8 developed anemia, 10 ulcers, and 8 developed both.
   d. Of an incorrect temperature reading being taken, given that of 15 available thermometers, 10 are accurate, 3 give overestimates, and 2 give underestimates.

3. Explain in your own words why it is necessary to subtract P(Both) in problems involving nonmutually exclusive events.

____

____

____

6.3 Instructional Unit: Computing Probabilities Over Multiple Trials

This section presents rules for computing the probability that each of a series of events will occur across trials. This description could make a relatively simple idea sound confusing, so let's consider some examples.

In the previous unit we restricted ourselves to one-trial possibilities such as: "Given 20 drugs of which 10 are effective, what is the probability of picking an effective one at random?" One trial (i.e., one pick) is involved.

But, suppose we said that there are two sets of drugs in different rooms each containing 10 effective ones out of 20. What are the chances of randomly picking an effective drug in both? Now two trials (picks) are involved. Based on what we learned before, we could figure out that for Set 1:

$$P(\text{Effective}) = \frac{n(\text{Effect.})}{n(\text{Drugs})} = 10/20 = 50\%$$

The same goes for Set 2. But, still what is the probability that an effective drug will be selected from both?

Or, suppose Moose Crenshaw (remember him?) takes a random guess on two multiple choice items in a row. If an item has one correct answer and four distractors, we know that the chances of a lucky guess on one item is 1/5 or 20%. But, what are his chances of getting both right? The difference here is that we have two trials (or probability situations) going, rather than only one.

Now, just one more idea. Try hard on this one. Trials can be independent or dependent. If independent, what occurs on one does not affect what occurs on the other. The two examples above indicate independent trials. Why? If you happen to pick an effective drug in Set 1, that has no effect on your chances of picking one in Set 2. Given that Moose makes a completely random guess on both items, his success (or lack of it) on one item has no bearing on his fortunes on the second. The two trials are independent.

Dependent trials, on the other hand, involve situations where what occurs on one trial does have an effect on what occurs on the other. Suppose you make two selections from one class of 20 students. What are the chances that both will be male if the class has 10 males and 10 females? For selection #1, we have

$P(\text{Male}) = \frac{10}{20} = 50\%$. But, suppose a male is picked on that trial?

For Selection #2, there are only 19 students left! And, there are only 9 males left! Why? Because one student, a male, has already been picked. So here, what happens on one trial does affect the probability results for the other.

Rule 3: Multiplication Rule for Independent Trials

Let's concentrate now on the independent trials case and return to dependent trials when we cover our last rule.

The formula (Rule) for determining probabilities on independent trials is:

$$P(A) \bullet P(A) \bullet \ldots \bullet P(A) = P(A)^n$$

n = number of trials

Thus, we raise the probability of A to a power equivalent to trials. If $P(A) = 1/5$, what is the chance that A will occur on two trials in a row?

Ans.: $P(A)^2 = \left(\frac{1}{5}\right)^2 = \frac{1}{5} \bullet \frac{1}{5} = \frac{1}{25}$ or 4%

Let's try some more meaningful examples.

Example 1: A student makes a completely random guess on each of three T-F items on a quiz. The probability of making a correct guess on any one item is 1/2. What is the probability that he will make three correct guesses?

$$\text{Probability} = P(\text{Correct Guess})^3$$
$$= \left(\frac{1}{2}\right)^3 = \frac{1}{2} \bullet \frac{1}{2} \bullet \frac{1}{2} = \frac{1}{8} \text{ or } 13\%$$

There is one chance in eight of guessing all three.

Example 2: Based on statistics for a rare form of malaria, the chances of contraction being lethal are one out of three. What is the probability then that two individuals who independently contract the disease will both die from it?

$$\text{Probability} = P(\text{Death})^2$$

$$= \frac{1}{3} \bullet \frac{1}{3} = 11\%$$

Example 3: The probability of event A is 2/3. What is the probability that A will occur in four out of four trials?

$$\text{Probability} = P(A)^4 = \left(\frac{2}{3}\right)^4$$

$$= \frac{2}{3} \bullet \frac{2}{3} \bullet \frac{2}{3} \bullet \frac{2}{3} = \frac{16}{81} \text{ or } 20\%$$

Example 4: The State Board of Education publishes figures indicating that the chances of an applicant being accepted for a teaching position in a typical school district are 2 out of 5 (2/5). Joe Smith submits applications to two typical districts. What is the probability that he will be offered a position in both?

$$\text{Probability} = P(\text{Accept.})^n = \left(\frac{2}{5}\right)^2$$

$$= \frac{2}{5} \bullet \frac{2}{5}$$

$$= \frac{4}{25} \text{ or } 16\%$$

Example 5: There are six cards, one of which is the Queen of Diamonds. A card is selected at random and then put back on a given draw. What are the chances that if three draws are made the card will be the Queen of Diamonds?

$$\text{Probability} = P(\text{Q-D})^n = \left(\frac{1}{6}\right)^3$$

$$= \frac{1}{6} \bullet \frac{1}{6} \bullet \frac{1}{6} = \frac{1}{6\bullet6\bullet6}$$

$$= \frac{1}{216}$$

Rule 4: Multiplication Rule for Dependent Trials

Dependent trials, as described earlier, occur when what happens in one trial affects the probability of the event in question taking place on another trial. Don't panic quite yet...

Suppose there is a class of 28 students. Fifteen of them are A students in your book--a pleasure to teach. If 2 students from the class were to transfer to new schools, what is the probability that both will be among the A group?

Our reasoning is as follows: For the first selection, we have $n = 28$ and A = 15. Thus, the probability of the first student to transfer being one of the A's is:

15/28 or 54%

One A student down, but to finish the probability estimation we need to know the chances of both transfers being A's. Well, if the first is an A, how many A's are left?

Ans.: 14 out of the original 15.

How many students are left following the first transfer?

Ans.: 27 out of the original 28.

What's the probability of A for transfer (or trial) #2?

$\frac{14}{27}$ or 52%

See how the two trials (i.e., probability estimates) are dependent? Your pick on Trial 1 affects the number of critical events and the size of sample space on Trial 2. If an A student is the first transfer, isn't there less chance than an A student will also be the second? Yes--there will be relatively fewer left in the class.

The formula for dependent trials is:

Probability = $P(A) \bullet P(B)$

Think of A in this case as the probability of the event occurring on Trial 1 and B as the probability of its occurring on Trial 2. Thus, the formula refers to the probability of the events happening in both trials 1 and 2.

Thus, in the example: Probability = $P(A) \bullet P(B)$

$$= \frac{15}{28} \bullet \frac{14}{27} = \frac{15 \bullet 14}{28 \bullet 27}$$

$$= \frac{210}{756} \text{ or } 28\%$$

Let's try some others:

Example 1: A matching test has 10 names in column A and 10 titles in column B, with the task being to match one correct title to each name. If a student has not read the materials and makes a completely random title selection for name #1, and then a completely random selection of the remaining titles for name #2, what is the probability that both guesses will be correct? (We'll substitute C-1 and C-2 to indicate "correct" on selections 1 and 2, respectively.)

Let C stand for correct guess.

$$\text{Probability} = P(C\text{-}1) \bullet P(C\text{-}2)$$

Thus, the formula reads:

"The probability of a correct guess on both Trials 1 and 2 equals the probability of a correct guess on Trial 1 multiplied by the probability of a correct guess on Trial 2." What is (C-1)? That is, what are the chances that the student will guess the correct title for the first name? How many titles?

Ans.: 10
$P(C\text{-}1) = 1/10$ or 10%, right?

But now for the <u>dependency</u>.

If he <u>has</u> guessed right on name 1, what is (C-2)? How many titles remain to choose from?

Ans.: only 9 - one has already been used. Thus,
$P(C\text{-}2) = 1/9$ or 11.11%

Does our student have a good chance of guessing the first two correctly?

$$P(C\text{-}1 \text{ and } C\text{-}2) = P(C\text{-}1) \bullet P(C\text{-}2)$$
$$= \frac{1}{10} \bullet \frac{1}{9}$$
$$= \frac{1}{90} \text{ or } 1.11\%$$

Not a very good chance is there?

<u>Example 2</u>: Suppose the above matching test contained five states to be matched to five state capitals. What would the probability be of guessing the first <u>three</u> correctly?

Don't be scared by the "three" probabilities. The logic is still the same; just write the formula (C=correct guess):

$$\text{Probability} = P(C\text{-}1) \bullet P(C\text{-}2) \bullet P(C\text{-}3)$$

We've just expanded it to include the third trial (item).

$P(C\text{-}1) = 1/5$ or 20%
$P(C\text{-}2) = 1/4$ or 25%

Understand why? If he's guessed the first one correctly, there are only 4 remaining to choose from for his second guess.

$P(C\text{-}3) = 1/3$ or 33.33%

and for his 3rd guess, there are only three to choose from.

Ans.: $\frac{1}{5} \bullet \frac{1}{4} \bullet \frac{1}{3} = \frac{1}{60}$ or 1.67%

Not too good a chance of guessing all three.

Example 3: There are 12 heart patients needing bypass operations. Dr. Smith will be performing the operation on 2 of them. Among the 12 are three very difficult cases involving patients of relatively advanced ages. What is the probability that both of Dr. Smith's operations will involve these problem individuals?

Let P stand for "problem."

$P(P\text{-}1 \text{ and } P\text{-}2) = P(P\text{-}1) \bullet P(P\text{-}2)$

What we're saying is that the probability of both operations (1 and 2) involving problem patients (P's!) is equal to the product of the probabilities of a P being picked on two selections. But, those probabilities are dependent, as should become obvious below:

$P(P\text{-}1) = \frac{3}{12}$ or 25%

Only 11 cases remain. Of those, only 2 are problems. Thus,

$P(P\text{-}1 \text{ and } P\text{-}2) = \frac{3}{12} \bullet \frac{2}{11} = \frac{6}{132}$ or $\frac{1}{22}$ or 4.55%.

Not much reason to panic. Odds aren't that great that Dr. Smith will get two problem cases for his operations.

Example 4: Students are told not to put comic books in their desks. Out of the 28 students in class, 20 obey that rule and remove comics from their desks on a particular day. The teacher decides to check desks on that day, but ends up being able to check only two before being called away. What is the probability that both of the ones she checks are proper (i.e., no comics)? Let's just use A to symbolize a "good desk."

Probability $= P(A\text{-}1) \bullet P(A\text{-}2)$

$P(A\text{-}1) = \frac{20}{28} = \frac{5}{7}$ or 71.43%

On first check, remember 20 out of 28 are good. But,

$P(A\text{-}2) = \frac{19}{27}$ or 70.37%

Understand how we get the 19? The 27?

$P(A\text{-}1 \text{ and } A\text{-}2) = \frac{5}{7} \bullet \frac{19}{27} = \frac{95}{189}$ or 50.26%

Thus, there is about a 50% chance that both desks will pass. Students better pray that luck is with them.

Example 5: There are 5 cards in a pile. One is a heart and the rest are clubs. If a card is selected at random and not put back, what is the probability that the first two picks will not be hearts? Let NH stand for "no heart."

$$\text{Probability} = P(\text{NH-1}) \bullet P(\text{NH-2})$$

$$P(\text{NH-1}) = \frac{4}{5} \text{ or } 80\%$$

See that there are 4 other cards available for the first selection. There are 5 cards in the sample space.

$$P(\text{NH-2}) = \frac{3}{4} \text{ or } 75\%$$

Once an NH is selected on first pick, only 3 NH remain in the resultant sample of 4.

$$\text{Probability} = \frac{4}{5} \bullet \frac{3}{4} = \frac{12}{20} \text{ or } 60\%$$

There's a reasonably good chance (60%) that the two cards selected will be other than hearts. To show you that our "heart" is in the right place, here are some questions to practice on.

6.3 POSTTEST (answers in back)

1. Indicate which of the following involve independent trials and which involve dependent trials.
   - _____ a. The probability of making a correct guess on each of a series of multiple-choice items.
   - _____ b. The probability of randomly matching each of a group of newborns to the correct set of parents.
   - _____ c. The probability of picking a diamond from a deck of cards, removing it, and then picking a diamond from the remaining cards.
   - _____ d. The probability of picking a heart from a deck of cards, putting it back, reshuffling, and then picking another heart.

2. Compute the probability for each of the following:
   a. In a class of 20 students, 18 have done their homework. The teacher randomly selects one student to work problems on the board, and then randomly selects another student. What is the probability that both will be ones who did their homework? (Are these independent or dependent trials?)
   b. Suppose that 18/20 students have done their homework. The teacher randomly selects a student to do a problem on the board. After the student sits down, the teacher randomly selects a student to do the next problem (note that the first student can be selected again). What is the probability that both selections will be students who have done their homework? (Are these independent or dependent?)

3. The chance of arresting growth of a particular type of pituitary tumor through chemotherapy is 1/4. What is the probability that in a series of different cases:
   a. the first two will be successful?
   b. the first three?
   c. the first four?

4. Given a conventional deck of 52 cards, what is the probability of:
   a. drawing a heart on Draw 1, putting the card back, reshuffling, and then drawing a heart on Draw 2?
   b. drawing a king on Draw 1, leaving it out of the deck, and then drawing a queen on Draw 2?
   c. drawing a diamond on Draw 1, leaving it out of the deck, and then drawing another diamond on Draw 2?
   d. (Bonus) drawing either a 10 or a jack on Draw 1, leaving it out of the deck, and then drawing either a nine or a king on Draw 2?

Unit VI Review Test
(answers in back)

1. There are 32 hospital centers in a foreign country. Of those, 5 have equipment for renal dialysis and 27 have no such equipment. If a hospital is selected at random,
   a. what is the probability that it will not have renal dialysis equipment?
   b. what is the probability that it will have renal dialysis equipment?
   c. what is the probability that it will either have renal dialysis equipment or not have such equipment?

2. Of 127 items on the State Achievement Test in English, 15 cover antonyms and 20 cover conjugations, two areas you haven't covered well in your course. If an item is selected at random, what is the probability that
   a. it will cover antonyms?
   b. it will cover conjugations?
   c. it will cover either antonyms or conjugations?
   d. if two items are selected, both will cover antonyms (are these dependent or independent trials?)?
   e. if two items are selected, the first will cover antonyms and the second will cover conjugations?

3. Differentiate between
   a. mutually exclusive events and nonmutually exclusive events.
   b. independent trials and dependent trials.

4. Five American students are entered in the World Science Fair contest. Altogether there are 316 contestants. Without knowing anything about the projects (i.e., assuming all contestants are equally likely to win before the contest starts), what is the probability that
   a. the winner will be an American?
   b. both the winner and the second place contestants will be Americans?
   c. the winner will not be American?
   d. both the winner and the second place contestants will not be American?
   e. if separate prizes are given for originality and for interest value, the winners of both prizes will be Americans (note: one individual can win both)?

UNIT VII

MAKING INFERENCES FROM SAMPLES

A. General Objectives

Statistics deals with what we can know about a large group (or population) by examining a smaller group (or sample) from that population. Of course, if complete population data are available, then there is no need to work with samples. However, due to the difficulty of measuring a population, samples are often the only way. One way to think about this is a direct route (working with a population) and an indirect route (using a sample) to the same destination. The concepts of random sample and Sampling Distribution play important roles in taking the indirect route, statistical inference, to our destination.

B. Specific Objectives

7.1 Define random sample and describe how to obtain a random sample
7.2 Explain the relationship between parameters and statistics
7.3 Explain the notion of bias with regard to Sampling Distributions
7.4. Characterize the properties of Sampling Distributions of means
7.5 Compute probabilities for obtaining sample means of various sizes when population parameters are known

7.1 Instructional Unit: Random Samples

As indicated above in the "general objectives" section, there are many instances in which we want to describe a population of individuals in terms of some characteristic: What is the average age of Oklahomans (Bet you've always wondered about that!)? What percentage of women voters in the United States are registered Republicans? These are but two examples of a multitude of questions asked every day about populations. The problem, however, is that to assess a population directly, you would need to collect the information from all members. See any problem with that? In fact, there are immense problems involving time and expense. For the average researcher, collecting complete population data is often impossible.

So, here's a riddle for you. You've been hired to determine the mean and standard deviation of the Grade Point Averages (GPAs) of all full-time students currently attending State University. Given the time and resources at your disposal, it is impossible for you to test the entire population of students and the data are not presently available anywhere. Yet, you're being asked to report population results!! What should you do? Here are some choices:

ERIC

a. get another job.
b. find the GPAs for the first 100 students you see on campus, and infer the population results from those.
c. find the GPAs for a representative (random) sample of 100 students and infer the population results from those.
d. all of the above.

This may surprise you, but the best answer is "c." If you cannot test an entire population, the first step is to (1) obtain a random sample. Then you (2) compute statistics for the sample (e.g., the mean, variance), and then (3) make an inference about the population on that basis. In this section, we will examine the first step--obtaining a random sample.

Why random samples?

Statisticians use samples to estimate parameters (population data). Why do they use random samples? If you do not know the answer, an important clue should be provided if we say that randomness helps to obtain representativeness.

Example

Mrs. Pedogog, a high school guidance counselor, believes that the students in her school are utterly brilliant, heads and shoulders above the kids attending the neighboring schools. She decides that the best way to test this hypothesis would be to compare the IQs of students in her school with the average IQ of all students in the city, which she knows to be 99. She does not have the time or money to test all students in her school, so she decides to use a sample, from which she can generalize to the school population. Since her office is next to the physics laboratory, run by Mr. Molecule, she administers the Stanford-Binet IQ test to everyone of Mr. Molecule's third-period (11:00-11:45) students. The average IQ turns out to be a whopping 125. She runs through the halls shouting hysterically, "The kids in our school are brighter than the average students in the city!"

Why is the above conclusion ridiculous?

Answer: Mrs. Pedogog used a sample that was not representative of the student population in her school. The sample may be representative of Mr. Molecule's brilliant physics students, but it is certainly biased in reference to the total student population at Pedogog's school.

Moral of story: Samples save time and money, but in order to be valid, they must be representative of the population to which they are supposed to relate. Random samples help to engender representativeness.

What exactly is a random sample?
or
What do I have to do to get one?

As the name implies, a random sample is obtained by a method which guarantees randomness (huh?). The question now is what do we mean by "randomness?" Actually, the name refers to how the sample is selected rather than of what it is comprised. Given several samples with no information about how they were obtained gives us no basis for determining whether they are random or not.

The idea of "randomness" requires that every element (person, place, or thing) used in the sample had an equal chance of being selected. Confusing? Let's break it into steps using the example involving Mrs. Pedogog.

You will recall that Mrs. Pedogog was interested in assessing the "typical" IQ of the students in her high school. Rather than test every single student (2,000 in all), she decided to use a sample, and from the sample data generalize to the population.

To draw a random sample she must do the following:

1. Clearly define her population: Mrs. Pedogog does so by saying, "The population with which I am concerned consists of all students in Beaver High School."

2. Decide on an adequate sample size: This one is up for grabs--obviously the larger your sample, the better your estimate of population parameters; Mrs. Pedogog defines her sample size by saying, "I will use 30 students in my sample (30 is generally a pretty good number, but randomness does not directly depend upon sample size)."

3. Select her sample by insuring that every student in the population has an equal chance of being chosen: Mrs. Pedogog does this by writing the name of every student in the high school on a piece of paper, putting all of the names in a drum, mixing the names, and having the principal of the school (with blindfold on) draw 30 names, one-by-one, from the drum.

Mrs. Pedogog's sample of 30 students is a random sample for the simple reason that every element (student) in the population the sample is intended to represent had an equal chance of being selected.

Did Joe Smith, who is repeating the 10th grade for the fifth time, have the same chance of being selected as Boyd Buttrom, the math genius? Yes, both names went into the drum on the same size piece of paper. No reason for believing that any one student had a greater chance of being selected than anyone else.

This example illustrates the basic criterion of random sampling: EVERY ELEMENT IN THE POPULATION FROM WHICH THE SAMPLE IS COMPOSED HAS AN EQUAL CHANCE OF BEING SELECTED.

Another Example

If you wished to determine the mean height of a population consisting of 3,000 college freshmen, you would have two choices--(1) test all 3,000, (2) test a random sample and estimate the population mean.

Being of sound mind, you would elect the latter alternative, keeping in mind that your sample (let's say 100 people) must be random. One possible way to select a random sample would be to sit down and write out all possible combinations of 100 names, and then, closing your eyes, select one of the lists. But, as you hopefully realize, the number of lists containing all possible combinations of 100 names would be so large that it would take a lifetime to write them all out. A more sensible alternative is to: Write down each name on an individual sheet of paper; place the papers into a container; draw out one name at a time, making sure that on the second draw any of the remaining 2,999 names has an equal chance of being drawn, that on the third draw any of the remaining 2,998 names has an equal chance, and so on until you reach 100--the desired sample size.

Question: Is the above sample random?

Your answer: Yes, because every college freshman in the population (3,000 in all) had an equal chance of being selected: the 7' basketball center and the 5' class-clown.

Still Another

Suppose we wanted to know whether college students at Ukaducca University are in favor of the administration allocating more money for the football program. We don't wish to question all of Ukaducca's 27,000 students, so we decide to use a sample. We select our sample by going to the Saturday football game and questioning every tenth student from Ukaducca that we find. We end up with a sample of 100 students from which we can generalize to the population.

Question: Is the above sample random?

Your answer: Clearly, it is not. All of Ukaducca's 27,000 students did not have an equal chance of being selected. Did a student who hates football and spends all her Saturdays in the library, have the same opportunity as one who has never missed a game? Clearly not. Students who attend football games are not representative of the total student population.

A Final One

Suppose a teacher wishes to obtain a random sample from his 8:00 a.m. class of 100 students, and he selects the first 20 students who arrive on Monday morning.

Question: Is the sample random?

Your answer: Obviously, it is not. What about the 15 or so students who cut class after the weekend? What about those students who are usually late to class? They did not have an equal chance: therefore, the teacher has selected a biased (not random) sample.

Final notes

1. Random samples are used in statistics because:

   a) From random samples you can validly generalize to the population with which you are concerned (but can never hope to test in its entirety).

   b) Many of the statistical procedures developed for use in hypothesis testing (to be covered in Unit IX) require random sampling. If your sampling procedures are not random, those statistical tests are invalid.

2. It may have bothered you in the discussion of random sampling procedures that the suggested method involved writing names on pieces of paper and drawing them from a drum. Actually, statisticians do not spend their time tearing up pieces of paper in the eternal quest for random samples. Two alternate sources are available. Just so you know about them, they are: (a) a random number table, available in many statistics texts (not yours!), and (b) computers, which generate random numbers for the asking.

3. The major point in this unit is that whether you write names on pieces of paper or use a random number table, you end up with a random sample if and only if all subjects (people, students, etc.) in the population you are measuring had an equal chance of being selected.

7.1 POSTTEST (answers in back)

1. Describe what is meant by random sample. ______________________

____

____

2. Why do statisticians use random samples? ______________________

____

____

3. Describe briefly how you could obtain a random sample of 50 subjects from a population of 300.

4. A researcher is interested in determining the average yearly income of male citizens over 18 years of age who reside in Catatonia, Nevada (pop. 424). To save time, he uses a random numbers table to select a sample of 50 from the Catatonia phonebook.

Will the resultant sample be <u>random</u>? Why or why not?

7.2 <u>Instructional Unit</u>: The Relationship Between Parameters and Statistics

In introducing this unit, some steps in the route of statistical inference were briefly mentioned. They were: 1) drawing a random sample, and 2) computing statistics of the sample.

You know what to do in step #1: drawing a random sample. You know what to do in step #2: computing statistics of the sample (computing means, modes, medians, variances, standard deviations, etc.). Then, we want to move to a third step--to make inferences from the sample about the population. Remember, as sampling has been described in this unit, we have no interest in the sample <u>per se</u> (i.e., the particular 100 college freshmen that we "happened" to select out of the population of 3000). We are interested in the whole 3000(!) and are merely using the sample to get the best picture that we can. But how do you estimate populations and test hypotheses about those estimates? It's a fairly "giant" step which will take at least another unit (Unit VIII to be exact) to even fully introduce. But we must make a start somewhere, so we will begin by again mentioning that the major function of samples is to provide a basis for making inferences about populations. In instances where <u>it is reasonable to describe all members of a population</u> (e.g., Mrs. Jones wants to find the average age of <u>her</u> 10 fourth-grade students), the use of samples would make little sense. In instances where it is unreasonable to describe an entire population (the average age of <u>all</u> fourth graders!), use of a sample is absolutely necessary.

Obviously, whether scores are from a sample or a population, it is possible to compute the various types of statistics we have discussed in earlier units. A mean or a median, for example, can be

obtained for either a sample or a population, using exactly the same computation procedures. But, perhaps you are getting the picture that it is important to differentiate between the two types of data sources. The indices to which we have referred as "statistics" are actually labeled as such only when computed for samples. When the same indices (means, modes, correlation, etc.) are computed for populations, they are called parameters. Remembering this scheme is easy: (s)amples yield (s)tatistics
(p)opulations yield (p)arameters

Note the use of Greek letters when referring to populations and Roman letters when referring to samples. This way the data source is never in doubt. Here are some examples:

	Population Parameter	Sample Statistic
Mean	μ	$\overline{X}$
Standard Deviation	σ	s
Variance	σ^2	s^2
Correlation	ρ	r

So, to say it again, we must rely on statistics as a means of estimating population parameters. But will the computed statistic necessarily give us the exact parameter of interest? The answer is "hardly ever," but the validity of the estimate (i.e., its closeness to the "true" parameter value) will depend on a number of factors. Here are the most obvious:

1. Randomness of the sample: If your sample isn't random, the accuracy of estimation should be reduced. For example, if the true average height of the entire population of male seniors in North-South High School is 68 in., it's doubtful that you would come even close if the sample you employed for estimation consisted of all male seniors on the basketball team. The probable mean value for that sample would be something like 74 in., with the result being an incorrect conclusion that male students grow especially big at North-South High. A random sample would be more likely to give a mean value much closer to the actual parameter, 68 in.

2. Sample size: Should sample size relate to the accuracy of estimation? The answer is "positively." The larger the sample, the greater your chances of obtaining a valid estimate. To use an extreme example, suppose you were interested in estimating grade point average (GPA) of students at North-South High, and decided to use a random sample of two students. By the "luck of the draw," the two names selected from the drum happen to be Joe Smitherene and Harry Duence, whose best subjects are homeroom and lockeroom, respectively. When you check the records to obtain GPAs, the computed value is 2.50 on a 4.00 scale. But that is for both added together! Dividing by two

gives an average for that sample of 1.25. Unfortunately, this time randomness did not provide a close estimate of the population parameter. Someone else might select a different random sample of two, and by chance, select Herb Einstein and Geraldine Cranium, who spent the night of the Prom comparing equations for determining the stress load on the dance floor. For this sample, the estimate is 3.99. Someone else might use a sample of two and actually get the true population parameter of 2.85. The point is that with such small samples, there is much room for error (chance variation): even one extreme case can radically affect the computed value. On the other hand, if a larger sample, say 50 students, had been used, high extremes would probably balance out low extremes, thus increasing the chances of coming very close to the true parameter.

3. Sample Statistic: As will be discussed in more detail below, certain statistics give better estimates of their corresponding parameters than do others. For example, using a sample mean to estimate a population mean will normally provide more accuracy than using a sample median to estimate a population median. Further, both of these statistics give more accurate estimations than would a sample range for estimating the range for the population. The ways in which different statistics are computed allow for some to vary more (from sample to sample) than others. And some, like the sample range and correlation coefficient, tend to systematically differ from their corresponding population values.

4. Population Form: In using statistics to make inferences about parameters, another important factor is the form of the population. It will be seen later on in this unit and also in the next, that some of the procedures used for statistical inference-making depends on assumptions regarding normality, i.e., scores being distributed in normal curve form. Analytical procedures used when a distribution approaches normality might not be appropriate when it is rectangular, seriously skewed, bimodal, etc. The general rule in the statistical world seems to be that the closer the population is to normality, the easier life becomes.

5. Luck: Statistics are merely estimators of population parameters. Even if you use the most conscientious random sampling techniques (with a gold plated wastepaper basket from which to draw selections), a very large sample, a "reliable" statistic (e.g., the mean as opposed to the range), and a perfect normally distributed population, there is no guarantee that the statistic you get will provide a close estimate of the population parameter. The above factors do guarantee precision in the long run, but in any single sample there will always be a certain margin for error. The hypothesis testing procedures to be dealt with later on take these error margins into account in telling us how "confident" we can be about our estimates. Although you are not expected, at this point, to understand how the following figures would be derived, we might find, for example, that the sample mean we obtain of 106 allows us to be

95% confident that the true population parameter lies between 102.04 and 109.96, and 99% confident that it lies between 100.84 and 111.16. All we are attempting to say at this point is that statistics is a game of probability; it is only possible to forecast what will happen <u>over time</u>. What occurs on any draw is unpredictable--otherwise Las Vegas would be a barren space in the desert and statistics a nonexistent field (shame on you if you cheered).

All of the above discussion naturally leads to our next topic.

<u>What Is a Sampling Distribution?</u>

We have made it clear that a given statistic computed from a single sample may not closely represent the true parameter of the population. But what would happen in the long run? That is, if instead of stopping with one sample, we computed the same statistic for many? Let's take the case of means as an illustration. Here is the situation: We have a population of 15,000 physicians in a particular region of the country and reason to believe that their salaries, whatever they may be, are normally distributed. Since we don't have the time to obtain a salary figure for all 15,000 individuals, we decide to use a random sample of 100 to help answer the pressing question, "How much do physicians representing that population make?" Unbeknownst to us, somewhere in the Internal Revenue Service computer are data that would indicate the actual and true mean salary for that group to be $80,000 (enough to make one think about med school). But not having this information, we must go ahead and compute the mean for our sample. Will it be exactly $80,000? Very unlikely, as samples fluctuate. In this case, we happened to get more than the usual share of "Fat Cat" surgeons, giving us a sample mean of $90,000. We have overestimated, unintentionally of course. Suppose we had an infinite amount of time on our hands and threw the first 100 back into the pool and then drew another random sample, and then another, and then another etc., etc. What mean values would we get? All different ones, of course; some would be higher than the parameter (80,000), some would be lower, and still others would hit it just about on the nose. The distribution of these mean values would comprise what is referred to as a <u>Sampling Distribution</u>. Here's a definition: <u>Sampling Distribution</u>: A Sampling Distribution of a particular statistic (e.g., the mean) refers to the probability distribution of that statistic (e.g., means) for all possible samples of a given size from some population.

Quite a mouthful, but the key to understanding here is not to try to memorize what is said, but rather to sit back and see if you can grasp its meaning. We have, thus far, established the following:

1. Random samples are used to derive statistics (means, correlations, etc.).

2. The statistics provide a basis for estimating population parameters (means, correlations, etc.)

3. Due to chance factors, e.g., who is selected in the sample, samples fluctuate.

4. As a result of sampling fluctuation, the statistic computed from one sample will probably differ to some extent from the same statistic, computed in the same way, from another sample.

5. The distribution of these differing statistical values represents what is called a Sampling Distribution. It tells you what possible values of the statistics can be obtained, with what frequencies, etc.

The notion of a Sampling Distribution suggests an organization of scores that take on different values. This was the very idea with which we were concerned in Unit I, when we organized scores in terms of frequency distributions and frequency polygons. It might be worthwhile to relate this prior learning to the present topic by distinguishing between "ordinary, run-of-the-mill" frequency distributions and this "imposing" new concept of Sampling Distributions.

Here we go:	Frequency Distribution (Unit I)	Sampling Distribution (Unit VI)
Purpose	To portray the numerical values of a set of scores (X)	To portray the expected numerical values of a statistic (e.g., $\overline{X}$'s, r's, Md's, etc.)
Source	A single sample or population	All possible unique samples of a certain size from a population
Horizontal Axis	Score values (X's)	Values of a statistic ($\overline{X}$'s or others)
Vertical Axis	Frequency of scores (f's)	Proportion of values (p)
Central Tendency	The mean ($\overline{X}$ or μ) indicates the average score, interpreted as a "typical" value	Expected Value represents the average value of the statistic; i.e., the "typical" estimate given by the random samples.
Variability	The variance (σ^2 or s^2) or standard deviation indicates the "spread" of scores	The Standard Error indicates the spread of the values of the statistic
Description	Empirical (obtained) set of data	Theoretical probability model

Thus, a frequency distribution tells you what happened in terms of individuals' scores on a particular measure: How many obtained which scores? What was the most common score? What was the average? etc.

A Sampling Distribution tells you what values are likely if a certain statistic is computed from random samples selected from some population: What value of the statistic is most probable? What is the spread of the different values? etc.

As you're reading along, perhaps you're getting worried about the idea of Sampling Distributions representing possible statistical values for all unique samples. "Will I have to construct one of these things?" you wonder. "That could take some time, and this course (as much as I'm enjoying it!!) isn't worth that amount of punishment for any grade!" you mutter (angrily). All of a sudden, dust appears on the horizon and hoofbeats are heard... STATISTICAL THEORY to the rescue!! Sampling distributions do not have to be constructed piece by piece--they are mathematically derived. More on that in the next section....

Final Note: Sampling Distributions must always be based on samples of a given size (whether that size be 3,5,23,56,89,700, etc.), and the samples must always be selected at random from the same population.

7.2 POSTTEST (answers in back)

1. In your own words, explain the difference between statistics and parameters. From what source is each derived? Which is used as the estimate for the other? How are they symbolized?

2. What is meant by sampling fluctuation? How does this concept relate to the idea that samples allow for probabilistic rather than certain estimations?

3. In general terms, what does a Sampling Distribution represent? What are the major factors that determine its properties?

7.3 Instructional Unit: The Notion of Bias in Statistical Inference

Samples fluctuate. If two or more samples, identical in size and selected from the same population, are used to derive the same statistic, the values obtained in each case will vary. The Sampling Distribution for that statistic represents the probability of obtaining each of the different values. But as it turns out, some statistics give "better" estimates of their corresponding parameters than do others. Actually, there are two groups: the unbiased and the biased.

Unbiased: Unbiased statistics, when computed from random samples, give values that differ from the population parameter only on the basis of chance (who was included in the sample, testing conditions, etc.). If you take sample after sample and compute that statistic for each, the individual values will tend to differ from one another (remember sampling fluctuation?) but will cluster around the population (the parameter) value. Examples of unbiased statistics are the mean, the median, and the variance (computed by s^2 formula). If a statistic is unbiased, it is valuable to use as a basis for making an inference about the population.

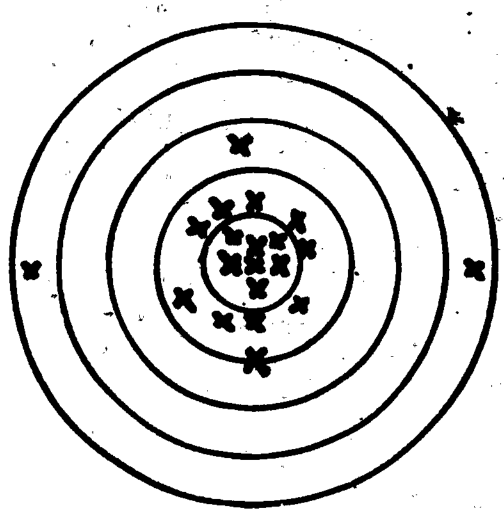
Biased: Biased statistics, when computed from random samples, give values that systematically differ from the population parameter. If you take sample after sample and compute that statistic for each, the individual values will not only differ from each other, due to sampling fluctuation, but will cluster around a value that is different from the population value. Examples of biased statistics are the range and the variance (if computed for a sample using the σ^2 formula). If a statistic is biased, using it to make an inference about the population can be quite misleading.

* * * * * * * * * * * * * * * * *

The Great Living Room Squirtout
or
How to Cross-Examine a Wetness

Pretend there is a watergun (we're intentionally making this example nonviolent) that is positioned 10 feet away from a target. The watergun is perfectly aimed at that target, and to maintain that positioning, is locked in a vise so that it cannot be moved.

You shoot the gun. Will the water hit the bullseye? Given that it is perfectly aimed at the bullseye, that is the most probable destination. In fact, though, many chance factors could prevent the shot from being accurate: a north wind this time, a south wind the next, a little too heavy on the trigger this time, a little too light the next, "heavier" water this time, maybe lighter water the next. If conditions are perfect we're going to hit the target! If they're not, we're going to miss it by being high or low, too far left or too far right. Who can predict? It's a function of chance. Here's what happened in 20 tries.

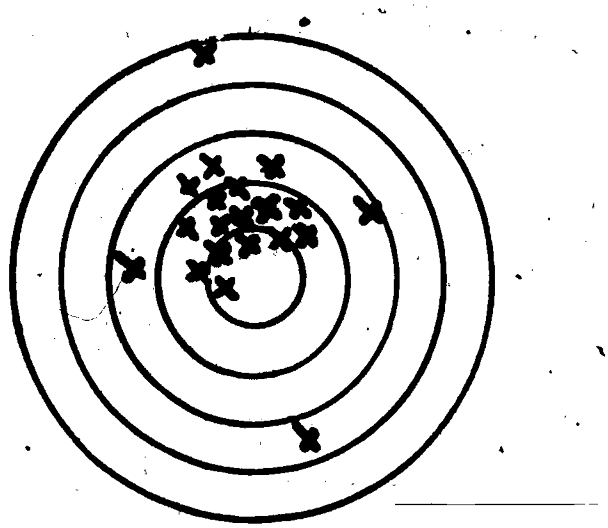


Notice that the bullseye was frequently missed, and some of the misses were rather severe at that. But, if you had to guess where the 21st shot would land, the bullseye would still be the best bet. Chance factors (like wind) resulted in some error, but the shots nonetheless cluster around the bullseye, since that's where the gun is pointed.

* See the relationship between these shots and computing unbiased statistics from random samples?

Shots #21-40

The next day, your clumsy cousin Ferdie stumbles on the vise and moves it! Now the water gun's aim is positioned 1 inch to the right of the target! You shoot the gun. Will the water hit the bullseye? It can, but that's no longer the most probable spot. If conditions are absolutely perfect, the water's going to land--you got it!--exactly 1 inch to the right of the bullseye. If a strong east wind blows, we might get the bullseye that way. Here's what happened in shots 21-40.



Notice that the bullseye was almost always missed, and most of the shots are to the right of it. No surprise! The gun was (mistakenly) aimed to the right--and the shots will cluster around the aiming point. As far as the bullseye is concerned, our aim is <u>biased</u>.

* * * * * * * * * * * * * * * * * * *

Unbiased statistics, like the mean, are analogous to the first part of the above illustration. If your sample is <u>perfectly</u> representative of the population--and nothing goes wrong in the testing-- the sample mean should be exactly equal to the population mean (bullseye!!). But how many times are you going to get a perfectly representative sample? Random samples help to produce representativeness but by no means guarantee it. So, many times, you'll miss the target, but your shots (sample statistics) will cluster around the bullseye (population parameter).

Biased statistics, like the range, are analogous to the second part of the illustration. Due to the way they're computed, the aim is all wrong from the very beginning. The expected value is <u>not</u> equal to the population value. Using the sample range to estimate the population range is like trying to hit the bullseye (see illustration) with your watergun pointed in the wrong direction. All you can do is (s)pray for luck. But the act itself is <u>inadvisable</u>. Here's an example:

<u>Example</u>

Suppose that for the entire population of employees at General Mittens Co. the average number of cups of coffee consumed each day (u_x) equals 2.3 & Range = 20. How do <u>we</u> know this? In real-life we wouldn't, but being allowed to make up hypothetical situations for instructional purposes gives one a great deal of power. Remember that the population mean and range would be derived based on <u>all</u> persons in the group. Range = 20 would be derived by subtracting the lowest value for the population (most likely, many people drink zero cups) from the highest value (in this case, Roderick Di Beaneo, a resident of Seattle, reported drinking 20 cups a day--the highest in the entire population). Suppose we take a sample of 20 workers: will the sample mean, $\bar{X}$, be identical to the real population mean, $u_x = 2.3$? No, unless we are extremely lucky. But, since we sampled randomly, we should be able to regard the statistic, $\bar{X}$, as a fair estimate. It could be too large or it could be too small, but given the way in which the mean is computed, no <u>systematic</u> overestimation or underestimation should have taken place. Any variation from the population mean should be due entirely to chance. If additional random samples were employed, some should yield $\bar{X}$'s that are higher or lower. Over all possible samples, however, the lows should balance out the highs, with the central tendency of the Sampling Distribution represented by the population value of 2.3. The larger the deviation from this central value, in either the positive or negative direction, the lower the probability of our obtaining a sample mean that size. What we are saying is that $\bar{X}$ provides an unbiased estimate of u_X. You may "miss" on any given try, but the long run projection is one of favoring $\bar{X}$'s that approximate u_X.

What about the range in our sample? The range is a <u>biased</u> statistic, as its deviation from the true population value takes place systematically rather than strictly by chance. Specifically, the sample range systematically <u>underestimates</u> the population range; after repeated sampling, you will get sample statistics that cluster around a value that is less than the true range of the population. Can you figure out why? Unlike the mean, the range is based on two scores only--the two most extreme. You may recall that in the whole population, the highest score was Rod Di Beaneo's 20 cups per day. What is the only way that the sample range can equal the population range? Rod Di Beaneo must be one of those selected!! If he is not, the sample range has to be smaller. If we don't get Rod Di Beaneo, we still have a decent chance of selecting some reasonably high score (Tom Perky, for example, who drinks 18 cups a day). But, can you appreciate what could happen if the population range were to be estimated from a very small sample? The low score might be something like 2 cups a day and the high score only 4 cups. The sample range would be only 2, a very substantial deviation from the parameter value of 20. Note, however, that the mean for this very small sample could still provide a close approximation of $u_X = 2.3$.

Let's turn to some other statistics and ask the same question: Which provide unbiased estimates of population parameters and which do not?

1. <u>Mean</u> ($\bar{X}$): Means will differ from sample to sample. But the preponderance of sample means should cluster around the true population value, u_X, whatever it may be. The further the distance of $\bar{X}$ from the parameter, the fewer the number of samples yielding those $\bar{X}$'s. The mean, as we have discussed, is an <u>unbiased</u> statistic.

2. <u>Range</u>: The ranges too will differ from sample to sample, but due to the way in which ranges are computed, they will cluster around a value that is smaller than the actual population range. The largest sample range can only be as great as the parameter; all other sample ranges will be smaller, due to the

failure of the samples to include the highest and lowest (most extreme) scores in the population. The range is a biased statistic.

3. Variance: You may recall that the variance can be computed in two different ways: with n in the denominator (symbolized σ^2), or with $n-1$ in the denominator (symbolized s^2). It was discussed that the former procedure is used for making inferences about populations from samples. Now you should get a better idea of why such a distinction is necessary.

   If you use the formula $\Sigma x^2/N$ (see p. 82) to compute the sample variance, you will get a biased estimate of the population variance, σ^2. Actually, for reasons we won't go into at this point, we will tend to get an underestimate of σ^2.

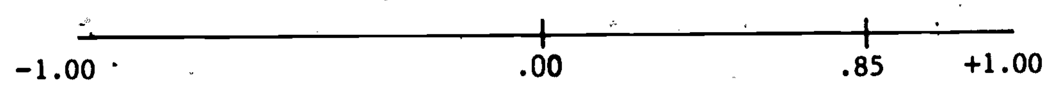
   If we use the formula $\Sigma x^2/n-1$ (see p. 86) to compute the sample variance, we get an unbiased estimate of σ^2. This formula, if you recall, gives a statistic called s^2 (see p. 86). s^2 is an unbiased estimate of σ^2.

4. Median (Md): The median, like the mean, is an unbiased statistic. If computed from different samples, the values obtained would tend to cluster around the actual population Md. What is interesting, however, is that the sample Md's would show greater spread (greater variability) than would the sample $\bar{X}$'s. Thus, while both statistics tend, over repeated sampling, to approach the population parameter, the chances of any single sample providing an exact (or close) estimation is greater for $\bar{X}$ than for Md. Remember when we discussed that the mean is usually the preferred measure of central tendency? One of the reasons is that it is a more precise estimator.

5. Correlation (r): Two variables are needed for correlational analysis, so let's assume that chemistry achievement scores for all high school students are correlated with their junior-year grade point averages (GPA). The question is, will the computed sample r's represent unbiased estimators of the correlation coefficient for the population (ρ)? Unfortunately for those seeking the most straightforward reply, the answer is "maybe, maybe not." For r, the rule is as follows: if the population $\rho = 0$, r does represent an unbiased statistic. That is, the sample r's will tend to distribute around a central value of 0.00. If the population correlation differs from 0.00 (either + or −), r becomes a biased statistic, for which the tendency is to underestimate the absolute value of the parameter. In conclusion, the Sampling Distribution of r may or may not reflect bias in estimating the population parameter, depending upon whether the two variables are correlated (bias) or not correlated (no bias) in the population.

* * * * * * * * * * * * * * * * *

Optional explanation for those interested: Why would a sample r tend to underestimate the population correlation? Suppose the correlation between two variables is +.85 in the population. If you try to estimate this value by use of a sample, the concept of sampling fluctuation suggests that in a single draw, anything is possible. The greatest probability, of course, is that you will obtain a positive correlation that is at least close to the actual population coefficient of .85. But given that correlations can range from -1.00 to +1.00, there is a much greater area <u>below</u> +.85 than there is above +.85. Look below:



Thus, following repeated sampling, you will end up with many values clustered around the +.85, but some will (by chance) fall in the rather <u>extreme areas below .85</u>. These extreme values will tend to pull the Expected Value (i.e., the average sample r) down, and thus yield a Sampling Distribution that systematically underestimates the parameter.

* * * * * * * * * * * * * * * * *

<u>Final note before practice posttest</u>: If sample values tend to cluster around the population parameter, we identified the statistic as unbiased (e.g., $\bar{X}$); if they tended to yield an average equal to some other value, we identified the statistic as <u>biased</u> (e.g., range). Sampling Distributions are theoretical models which are based on what is mathematically determined to occur if a statistic is computed from all unique samples that can be generated from a population.

7.3 POSTTEST (answers in back)

1. Assuming that the population in question is normally distributed, which of the following statistics would comprise <u>unbiased</u> estimators of their corresponding parameters. For any that you characterize as <u>biased</u>, specify the direction that the bias will take (overestimate or underestimate).

   a. Md
   b. variance using $\Sigma x^2/n-1$
   c. variance using $\Sigma x^2/N$
   d. $\bar{X}$
   e. r ($\rho = 0$)
   f. r ($\rho \neq 0$)

2. What is the relation between sample size (n) and the accuracy of estimation?

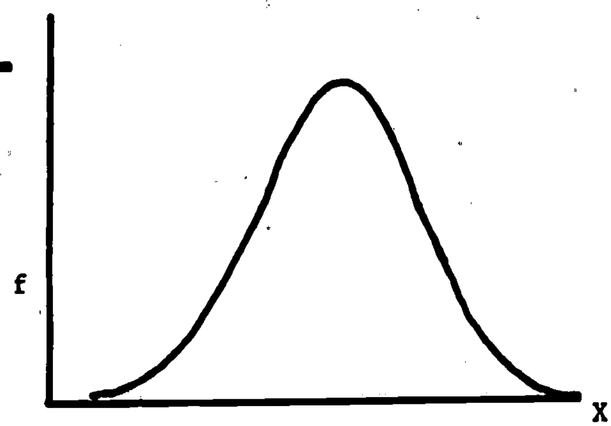
____

____

7.4 Instructional Unit: Sampling Distributions of Means

Given that the mean is probably the most commonly used statistic for inference-making, it seems important to describe the properties of its Sampling Distribution in a little more detail. First, what might a Sampling Distribution look like? At first glance, it can easily be mistaken for a Frequency Polygon (Unit I). Actually it's quite similar, but while the Frequency Polygon tells you the frequency with which different score values (X's) were obtained from all people, the Sampling Distribution tells you the frequency (or probability) with which different **means** ($\bar{X}$'s) were obtained from all unique samples.

SAMPLING DISTRIBUTION



Regular "OLD"

FREQUENCY DISTRIBUTION

Question: What three properties are important in describing distributions of scores?

Answer: (from Units I, II, and III) Form, central tendency, and variability.

Just as we dealt with these properties in discussing frequency distributions, Sampling Distributions have form, central tendency, and variability.

How are the values distributed when computed for all unique samples, of identical size, selected from the same population? Let's deal with form first, assuming random sampling in all examples.

Form

If the population from which sampling takes place is normally distributed, the Sampling Distribution of means will also be normally distributed. Assume that height is normally distributed for a certain population (e.g., all females in Omaha). If means ($\bar{X}$'s) are computed for all unique samples of a certain size selected from that population, the values of those means will be normally distributed!

HOWEVER, if the population scores are not normally distributed, the form of the Sampling Distribution of means will approach normality as sample size increases. What this suggests is that, regardless of the form of the population (rectangular, positively skewed, etc.), we will end up with a Sampling Distribution resembling the normal curve if sample size (n) is large enough. For instance, suppose that the distribution of women's heights in Omaha is positively skewed. An infinite number of very small samples (e.g., all n's = 2) may yield $\bar{X}$'s that do not distribute normally. An infinite number of large samples (e.g., all n's = 30) would yield $\bar{X}$'s that tend to distribute normally. (The advantage of a "normal" Sampling Distribution is that it can be analyzed via the "normal curve probability estimates" with which we dealt in earlier sections--more on this later.)

How big must n be to ensure normality? You do not have to memorize these, but please read them; they provide valuable guidelines for making decisions about sample size.

A. If your population data are normally distributed, a sample size of only 4, (n = 4) is sufficient to insure normality in the Sampling Distribution.

B. If your distribution is not normally distributed but is symmetrical, samples of 10 - 15 cases may be needed.

C. If your distribution is slightly skewed, you may still need at least 10 cases.

D. If population is moderately skewed, 20 cases may be called for.

Example Problem (just for fun):

A Sampling Distribution of means is based on samples of 5 cases selected from a population that is moderately skewed in form. Would it be correct to assume that the Sampling Distribution is normally distributed?

ERIC

Answer: No. For a moderately skewed population, samples of size 20 (see "D") are generally needed to insure normality in the Sampling Distribution.

Suppose the population data were normally distributed. What then?

Answer: Samples of $n = 5$ would be adequate to insure normality (see "A").

Central Tendency

In discussing the concept of bias, we said that if the typical sample statistic was expected to be equivalent to the population parameter, the estimator was identified as being unbiased (e.g., s^2, Md, and, of course, $\overline{X}$.) The mean value of a Sampling Distribution is called its Expected Value, and symbolized as, E( ). Inserted in the parentheses is the symbol for the particular statistic being characterized; thus, we can have $E(Md)$, E(range), etc.

In the case of means, the properties of $E(\overline{X})$ can be very easily identified by the following rule: $E(\overline{X}) = \mu_X$. What this means is that if $\overline{X}$'s are computed for an infinite number of random samples, the mean of all the $\overline{X}$'s (i.e., the average $\overline{X}$) will be equal to the population mean (μ_X).

Example:

Suppose that for all bluejays who nest in a particular park, the mean length of individuals is 9 inches. All unique random samples of 25 bluejays are identified and an $\overline{X}$ is computed for each. The $\overline{X}$'s will naturally differ from one another, but if all are averaged together, what value will be obtained? The answer: $E(\overline{X}) = 9$, the population mean! Suppose that $\overline{X}$'s are computed for all unique samples of 10 individuals. What will these $\overline{X}$'s average? The answer is the same: $E(\overline{X}) = 9$. Does the form of the distribution matter? It does not; in the case of means, $E(\overline{X}) = \mu_X$ regardless of form or of sample size.

Here is the rule again: When means are computed for an infinite number of random samples (of a certain size), the mean of the means [i.e., $E(\overline{X})$] will be equal to the mean of the entire population. Understand what this means? Try your skill on this one:

The average weight of adults in a particular town is 135 lbs. If $\overline{X}$'s are computed for all unique samples of 25 people from that town, what will $E(X) = ?$

Answer: $E(\overline{X}) = 135$; the mean of all the means will be 135.

Suppose the population was positively skewed, would the $E(\overline{X})$ change? No! Suppose the sample size were reduced to n = 4; would that have an effect? Well, it might reduce our chances of obtaining a normally distributed Sampling Distribution - depending on the form of the population (see previous section) - but central tendency would conform to the same rule: $E(\bar{X}) = u_X$.

Here are the symbols again for a final view before we turn to the final Sampling Distribution characteristic, variability.

$\bar{X}$: the mean of an individual sample selected from a population

u_X : the mean of the population

$E(\bar{X})$: the mean of a Sampling Distribution of $\bar{X}$'s for that population

Variability

Thinking way back, you may recall that Unit II (Central Tendency) was followed as closely as was possible by a unit on the subject of variability (Unit III). Does Unit III apply to what we're doing here? There is no question that it does - anytime we have a distribution of scores, whether it consists of raw scores or sample means, it helps to give a characterization of form, central tendency, and variability. Having already covered the first two concepts in reference to Sampling Distributions of Means - we are now ready for the third. What about variability?

The *variance* and the *standard deviation* are the most respected and revered indices of variability. Theoretically, we can compute them for Sampling Distributions using the same formulas employed for frequency distributions in Unit III.

The standard deviation of a Sampling Distribution, like the Expected Value, is so important that it is given a special name--the Standard Error. The Standard Error is simply the standard deviation of all values in the Sampling Distribution (keep in mind that these values are all sample means). To avoid confusion with other types of standard deviations, the Standard Error is symbolized as $\sigma_{\bar{X}}$; this symbol suggests that we are dealing with the standard deviation (σ) of sample mean ($\bar{X}$).

Let's stop for a moment and get our symbols straight.

Question: What do the following symbols mean: σ, s, and $\sigma_{\bar{X}}$?

Answer: σ refers to the standard deviation of all raw scores in the population.

s refers to the standard deviation of all raw scores in an individual sample from the population.

$\sigma_{\bar{X}}$ refers to the standard deviation (called Standard Error) of all scores (sample mean, $\bar{X}$) in the Sampling Distribution.

What does the Standard Error $\sigma_{\overline{X}}$ tell us? It indicates the variability in the Sampling Distribution. It is particularly valuable in allowing us to compare the variability (or sampling fluctuation) in two or more different Sampling Distributions.

Properties of the Standard Error

Suppose we have a population of 100 test scores with $u_X = 57$. We randomly select an infinite number of samples of size 5 and compute their means. The result is a Sampling Distribution of means with an Expected Value, $E(\overline{X})$, of 57. Right?

Use your imagination, and pretend that it would be practical to compute the Standard Error by laboriously subtracting $E(\overline{X})$ from each of the infinite number of sample means, squaring the resultant deviation scores, averaging them, and taking the square root of the result. In practice, this would be absurd, but let's say that we were committed to the task, and that the result was equal to 3.00. Now we have a Sampling Distribution that we can characterize quite well.

$$E(\overline{X}) = 57$$

$$\sigma_{\overline{X}} = 3$$

If you remember the material on z scores, you realize that given the mean and standard deviation of any normal distribution, we can compute the percentage of scores that fall between any two z scores, raw scores, and so on. Knowing the Expected Value and Standard Error of a Sampling Distribution can be very valuable. We'll save this for later, however.

For now, simply consider the above results and think about the following situation. Suppose we went back to the same population of 100 test scores, but this time selected at random an infinite number of samples of size 10. We end up with a Sampling Distribution of means based on 10, not 5, cases.

Question: What would be the Expected Value of this new Sampling Distribution?

Answer: $E(\overline{X}) = 57$, of course! True, we changed sample size, but the basic principle is still in effect: $E(\overline{X}) = u_X$.

The following question, which is similar to one posed in Unit 6.2, is really the "meat" of this section. Before looking at the answer, use your powers of intuition and pure logic and give it a try. If you get the right answer (for the right reasons), the remainder of this unit will be smooth sailing. If you cannot get the answer, give the accompanying explanation your full attention.

THE QUESTION: In the first example, we pretended that the Standard Error of the Sampling Distribution for 5-case samples was 3. Would the Standard Error of the new Sampling Distribution, based on 10 cases, be equal to, greater than, or less than 3? In other words, would the variability between sample means be different when all samples are made up of 5 cases as opposed to 10 cases?

THE ANSWER: The variability (Standard Error) of the 10-case Sampling Distribution would be less than that of the 5-case distribution. Why? The 10-case sample means are likely to fluctuate (from u_X) less than the 5-case sample means simply because more scores are being taken into account in each case.

NEW QUESTION: What is the most accurate way of determining the mean of a population of 1,000 scores?

ANSWER: Take all of the 1,000 scores and average them!!

NEWER QUESTION: What would be the second most accurate way?

ANSWER: Take a sample of 999 scores and average them!!

NEWEST: The third most reliable estimate?

ANSWER: Take a sample of 998 scores and average them.

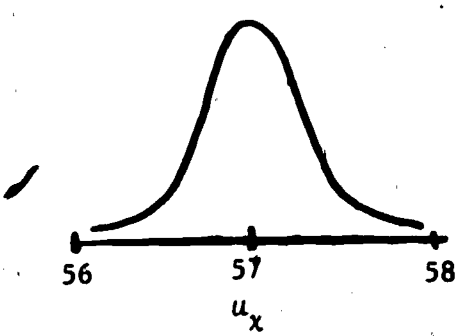
The larger the samples, the closer each mean will probably be to the population mean and to each other!!

For those serious students who are still plagued by doubts and uncertainties, more explanation can be found below.

* * * * * * * * * *

Pretend that from our population of 100 scores with $u_X = 57$, we construct a Sampling Distribution based on samples of size 99. The Sampling Distribution, by definition, would show the proportion of the sample means of different values that we obtained. Understanding that each sample mean is based on 99/100 scores from the population, would there be much variability? Would there be a large Standard Error?

NO! Each sample mean would be extremely close to u_X. After all, each is based on every score from the population save one. Thus, as we draw sample after sample, and compute $\bar{X}$ for each, the resultant values would probably be ones like 57.05, 56.98, 57.01, 57.23, 56.89, etc. See what's happening? With only one member of the population missing in our samples of $n = 99$, each sample mean would almost be identical to $u_X = 57$. Variability or $\sigma_{\bar{X}}$ of the Sampling Distribution, would therefore, be very small. The Sampling Distribution, when plotted, would have, of course, an Expected Value = to 57, and look something like this:



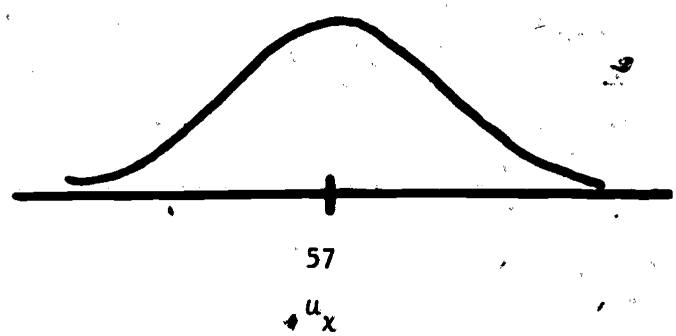
Note: Nearly the entire distribution is between 56 and 58--small variation!!

22J

On the other hand, pretend that from our population of 100 scores with $\mu_X = 57$, we construct a Sampling Distribution of means based on 2-case samples. Understanding that each sample mean is based on only 2/100 scores from the population, would there be much variability? Would there be a large Standard Error?

Yes! Yes! Using only two cases, some of the sample means would be extremely high, and some of the sample means would be extremely low. Thus, as we draw sample after sample, and compute $\bar{X}$ for each, we could get fairly discrepant values such as 42.06, 65.28, 58.01, 39.86, 49.37, etc. With only 2 people being considered in each sample, the chances for variation--from sample to sample--would be relatively large. Therefore, variability of $\sigma_{\bar{X}}$ of the Sampling Distribution would also be large.

The Sampling Distribution, when plotted, would have an Expected Value = to 57, and look something like this:



ANOTHER BASIC LAW TO REMEMBER: As sample size increases, the Standard Error of the Sampling Distribution decreases. The Expected Value, however, is unaffected by sample size.

Computing the Standard Error

We have thus far been able to define with pinpoint accuracy one of a Sampling Distribution's most important properties: $E(\bar{X}) = \mu_X$.

Likewise, we can determine the Standard Error of a Sampling Distribution through a mathematical derivation. Through genius, patience, and sophisticated mathematical procedures, statisticians have come up with the following formula:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

ERIC

In a sentence this equation means, the Standard Error of a Sampling Distribution is equal to the population standard deviation (σ) divided by the square root of the sample size (n).

Example Problem

Given a population of 1,000 scores with $u_x = 97$ and $\sigma = 20$, what is the Expected Value and the Standard Error of a Sampling Distribution of means where sample size is $n = 16$.

The first question is easy: $E(\overline{X}) = u_X$; Expected Value = 97.

The second question is also easy:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} ; \sigma_{\overline{X}} = \frac{20}{\sqrt{16}}; \sigma_{\overline{X}} = 20/4 = 5$$

*note: n = sample size, not population size.

Example Problem

From the same population, we construct a Sampling Distribution of means where sample $n = 4$. What is the Expected Value? Standard Error?

Answer: $E(\overline{X})$ again = 97.

Since sample size is less this time, we expect variability to increase. Using the formula, this turns out to be the case.

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

$$= \frac{20}{\sqrt{4}}$$

$$= \frac{20}{2}$$

$$= 10$$

The formula for computing the Standard Error is entirely consistent with the basic premise that suggests that as sample size increases, variability of the Sampling Distribution will decrease. Can you see why that is the case both logically and mathematically (or either)?

Final Notes

1. The Standard Error is the standard deviation of a Sampling Distribution. In the case of means, which we have been discussing, it is symbolized, $\sigma_{\overline{X}}$. This symbol very simply says that we are representing

the standard deviation, σ, of sample means, $\overline{X}$'s.

If you like analogies: $\sigma_{\overline{X}}$ is to a Sampling Distribution what σ is to a population of scores.

2. If you know σ for a population, you can determine the amount of variation between sample means for all unique samples based on the same number of cases.

3. Descriptive statistics (Units I-V) help us to characterize a given set of scores. Inferential statistics help us to make probability judgments about populations using data from samples. Sampling Distributions are probability models; they tell you what to expect in the long run if you rely on samples of a given size to estimate certain population parameters (means, standard deviations, etc.). In restricting our focus to sample means, we have found that with the availability of certain "givens," such probability models can be derived in terms of form, central tendency, and variability. As will be seen in the next section, these three characteristics are all we need to know to begin the inferential process.

7.4 POSTTEST (answers in back)

1. Describe the relationship between sample size (n) and the form of a Sampling Distribution of means.

____

____

2. How does form of the population influence the above relationship?

____

____

3. For each of the symbols in the lefthand column, match the appropriate description from those listed at the right.

_____ a. $\overline{\overline{X}}$ 1. the variance of a Sampling Distribution

_____ b. σ 2. the mean of an individual sample

_____ c. $\sigma_{\overline{X}}$ 3. the population standard deviation

_____ d. μ_X 4. the population mean

_____ e. s 5. the Standard Error of a Sampling Distribution

_____ f. $E(\overline{X})$ 6. the standard deviation of a sample

7. the Expected Value of a Sampling Distribution

8. the variance of a sample

4. What is the relationship between μ_X and $E(\bar{X})$? How is this relationship influenced by n?

---

5. What is the relationship between σ and $\sigma_{\bar{X}}$? How is this relationship influenced by n?

---

6. A researcher is concerned with a population of test scores where $\mu_X = 70$ and $\sigma = 20$. Determine the Expected Value and Standard Error of Sampling Distributions of means where:

   a) the sample $n = 4$

   b) the sample $n = 16$

   c) the sample $n = 64$.

7.5 Instructional Unit: Computing Probabilities for Obtaining Sample Means

You are about to see how Sampling Distributions can actually be put to use in answering certain types of questions. To get through this section, and thereby finish this Unit, you will need to be comfortable with the following:

1) the general concept of Sampling Distributions;
2) determining the Expected Value and Standard Error of a Sampling Distribution; and
3) the use of the normal probability model ("area under the normal curve" table) presented in the last few units of Unit IV.

Suppose that every fifth grader in Merryville, New Mexico (5,000 in all) was administered the Auto Mechanics Aptitude Test. The mean score for the population was found to be 50, and the standard deviation was found to be 12. The distribution was also found to be normally distributed.

Mr. Greaser, an auto mechanics teacher, asks himself the following question: "If I were to take a random sample of 16 fifth graders and give them the aptitude test, what would be the chances that the mean of my sample would be greater than the population mean ($\mu_X = 50$), but less than 53?"

Given the above information, can Greaser's question be answered?

Yes it can! Sampling Distributions show the probability of obtaining sample means of different values. That statement is definitional. If we know the properties of the Sampling Distribution that defines Mr. Greaser's students, the question he poses can be answered with ease.

Can we characterize the appropriate Sampling Distribution?

It can be done very easily. The mean of the Sampling Distribution of means would be equal to the population mean, 50.

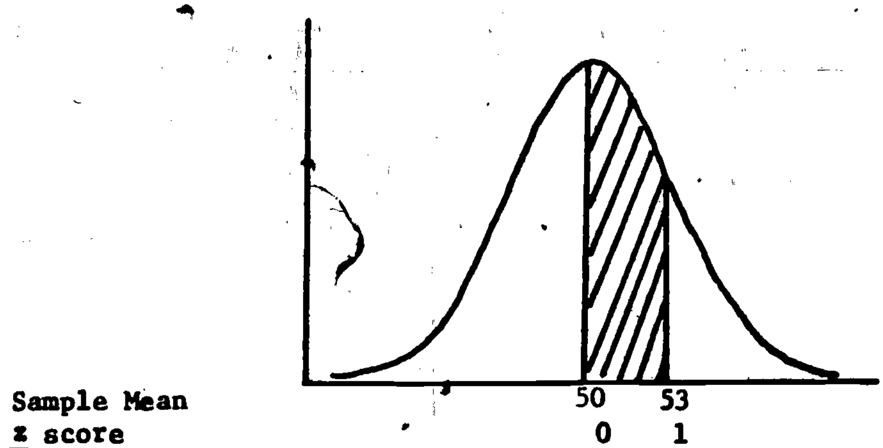
$$E(\overline{X}) = \mu_X \ ; \quad E(\overline{X}) = 50$$

The Standard Error of the Sampling Distribution can also be easily derived:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{16}} = \frac{12}{4} = 3$$

Since the population was normally distributed and the sample size (n=16) is fairly large, we know that the Sampling Distribution will resemble a normal curve.

Given the Expected Value, Standard Error, and form of the Sampling Distribution, we can now actually draw it. Using the "area under the normal curve" table as we did in Unit IV, we can answer questions regarding the probability of selecting samples (of size 16) with means of different values.



In order to determine probabilities, you will need to convert your raw scores (actually, sample means) into z scores.

$$z = \frac{X - \overline{X}}{\sigma} \text{ in this case: } z = \frac{\overline{X} - E(\overline{X})}{\sigma_{\overline{X}}}$$

for a sample mean of 50: $\frac{50 - 50}{3} = z = \underline{0.00}$

for a sample mean of 53: $\frac{53 - 50}{3} = z = +1.00$

Use the "area under the normal curve" table to find the distance between your z scores.

From the table we find that for $z = +1.00$, we have 34.13% of the area between that value and the mean ($z = 0$).

Our Answer: If Mr. Greaser selects a random sample of 16 students from the population, there is a 34.13% chance that the mean of the sample will fall between 50 (the population mean) and 53.

All we are doing is using normal curve probabilities to determine the likelihood of obtaining sample means of certain values. The procedure is identical to that used in Unit IV, except of course, that the z scores which comprise the referencing points for the "normal curve" table are now based on distances from $E(\bar{X})$ divided by $\sigma_{\bar{X}}$. Why these and not our old friends $\bar{X}$ and σ from Unit IV? Shame on you if you asked: We are dealing with a Sampling Distribution of means, and these elements are what define the Sampling Distribution. The concept of Sampling Distributions is now being applied to answer questions like "If a sample of a given size is selected from a certain population, what are the chances that the sample mean will deviate by 'so many points' from the population mean?" "What are 'probable' sample means and what are 'improbable' ones?" etc.

If, in attempting to solve some of the problems on your own, you find that your answers consistently differ from the correct ones shown, the source of your difficulty is probably one of the following:

1. You never read Unit IV and are attempting this section by ambition and continual prayer.

2. You studied Unit IV by rote, and have no understanding of what "normal curve area" problems involve.

(For either of these, the only possible remedy is to review Unit IV now. Consume two sections with a glass of water before going to bed.)

3. You learned Unit IV well, but that was months ago, as you have taken a long break in getting here. Remedy: a quick review of Unit IV should get you healthy again.

4. You know the "area problems" cold, and think that the present stuff is very easy, but glancing at your work, you now notice that in trying to derive z scores for the sample means, you used σ in the denominator instead of $\sigma/\sqrt{n}$. If such were

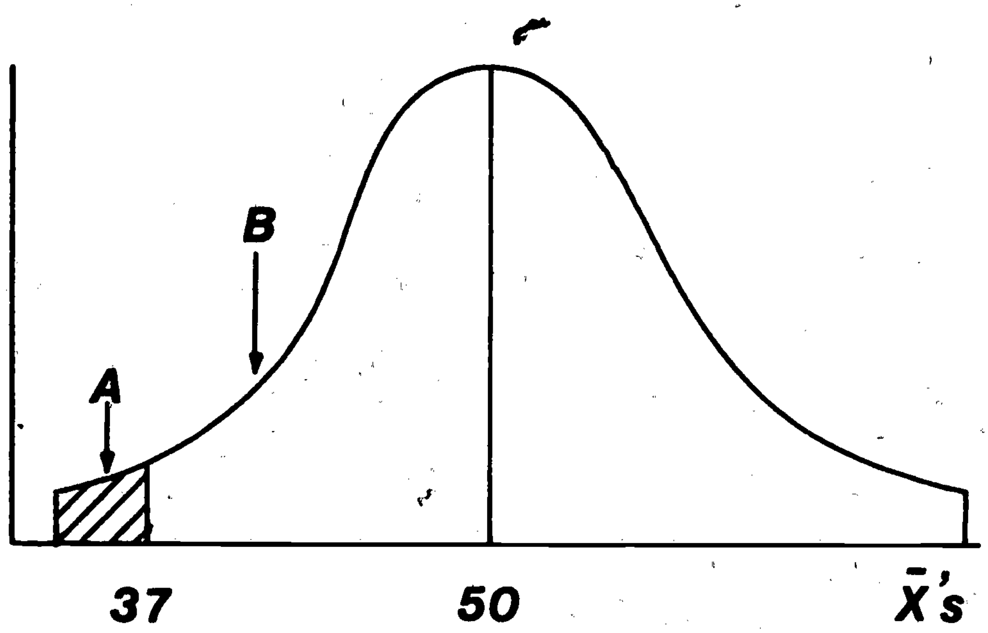
the case, you unjustifiably attempted to mix apples and oranges in your equation, resulting in absolutely no chance of obtaining a correct answer. Remedy: You have been careless. If you have been drinking a beverage while doing your calculations, check its contents for any "strong" substances; replace drink with orange juice and get a night's sleep. Tomorrow start the morning by saying to yourself 100 times, "Sampling Distributions have $\sigma_{\overline{X}}$ not σ as their standard deviation index."

Example Problem

In the fictional town of Feckless, the citizens are famous for flaunting face freckles. Each year, the Feckless Face Freckle Festival is held, at which time a count is made of the face freckles flowering on fellows from Feckless. Descriptive indices obtained for the population at the very last freckle figuring were $\mu_X = 50$ and $\sigma = 10$. Suppose that you are a visitor in Feckless who feels that the face freckles figures forwarded are fraudulent. You think that a mean of 50 freckles a fellow is farfetched and that the real figure is probably more like 37 (a fairer freckle frequency in your mind). Having limited time in Feckless, you are forced to limit your investigation to a random sample of only four people; i.e., $n = 4$. Even though your challenge of the Feckless face freckle figures will be futile (μ_X does equal 50), if you proceed with a count for your randomly formed foursome, what are the chances that the obtained $\overline{X}$ will support your initial hypothesis by being less than 37? In other words, assuming normality, what is the probability of obtaining a mean lower than 37 in a random sample of $n = 4$?

Step #1: You may recall from Unit IV that we called this a two-step problem: (a) use of the "formula" to convert a score to an appropriate z value, and then (b) use of the "normal curve table" to represent the z as a probability index. To use the "formula" as our first step, however, we must determine the parameters of the Sampling Distribution that define our problem (i.e., the Expected Value and the Standard Error are needed). At the next step we should sketch the Sampling Distribution (the normal curve), labeling the givens and shading the area of concern.

$$E(\overline{X}) = \mu_X = 50 \qquad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{4}} = \frac{10}{2} = \underline{5}$$



Step #2: Convert your sample mean to a z score and represent the latter on your diagram.

For a sample mean of 37: $z = \dfrac{37 - 50}{10/\sqrt{4}} = \dfrac{-13}{5} = -2.60$

Step #3: Use "area under normal curve" table to locate the area in question. Remember that the problem asked us to determine the probability of selecting a sample mean below 37. Thus, we need to find the area below z = -2.60. The table shows an area of 49.53% for that particular z. But, remembering that the table works from the middle outward, the 49.53% must be taken to represent Area B on the graph - the area between z = -2.60 and z = 0. We are interested in Area A -- the area below z = -2.60. This is obtained by remembering that half of a normal curve constitutes 50% of the total scores. Area A and B, therefore, equals 50%. To get Area A, we must use subtraction: A + B - B = A; 50 - 49.53 = .47%.

Ans. .47% (less than 1%) represents the probability of selecting four people at random from the Feckless populace and discovering that they average less than 37 in number of face freckles. A sample mean of 37, even with only $n = 4$, would be an unusual result, given that the population parameter is $u_{\chi} = 50$.

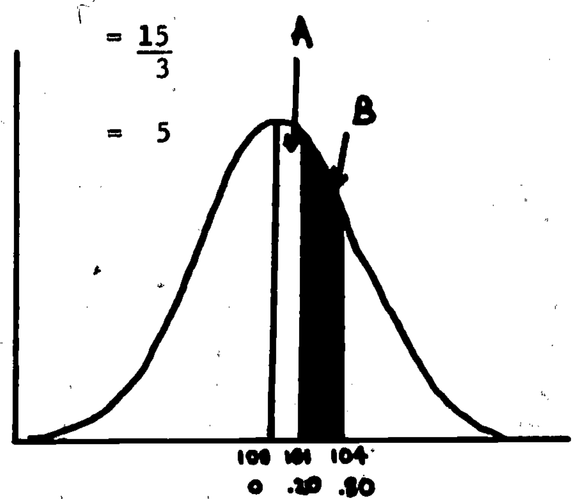
One More Example

From a population of scores where $\mu_X = 100$ and $\sigma = 15$, we select a random sample of size 9. What are the chances that the sample mean, $\bar{X}$, will fall between 101 and 104?

Step #1: Determine the parameters of the Sampling Distribution. Draw the distribution, penciling in area of interest.

$E(\bar{X}) = 100$

$$\sigma_{\bar{X}} = \frac{15}{\sqrt{9}} = \frac{15}{3} = 5$$



Sample Mean
z score

Step #2: Convert your sample means into z scores.

sample mean of 101: $z = \dfrac{101 - 100}{5} = 1/5 = +.20$

sample mean of 104: $z = \dfrac{104 - 100}{5} = 4/5 = +.80$

Step #3: Use "area under normal curve" table to locate area of interest (Section B).

Table shows that 7.93% of area falls between $z = 0$ and $z = .20$. This is Section A.

Table shows that 28.81% of area falls between $z = 0$ and $z = .80$. This is Section A + B.

To find Section B alone:

28.81% - 7.93% = 20.88%

A + B - A = B

Our Answer: There is a 20.88% chance that the sample mean will fall between 101 and 104.

Final notes:
1. Always use normal curve diagram.
2. Keep in mind that these problems are very similar to the ones you attempted to solve in Unit IV. The only difference is that you are dealing with a normal distribution of sample means (the Sampling Distribution), rather than a normal distribution of individual scores.

7.5 POSTTEST

1. A sample of 16 cases is selected from a population where $\mu_X = 25$ and $\sigma = 4$. Determine the probability that the sample mean will be:

   a. greater than 26
   b. between 24 and 26
   c. between 23.5 and 24

2. From the same population as above, a sample of 4 cases is selected. Determine the probability that the sample mean will be:

   a. between 26 and 28
   b. less than 23
   c. between 24 and 26

Recommendations for Unit VII:

1. Unit VII does not place too much emphasis upon computations. But, be sure that you understand the major concepts presented.

2. Do not memorize definitions and concept descriptions. Try to define the concepts, like Sampling Distributions, in your own words. Your own definition is always more meaningful than a memorized one from the text.

Unit VII Review Test
(answers in back)

The following description will be used to generate questions that should cover all (or most) of the major concepts discussed in this unit. Try to answer the questions without looking back; if you must consult the text, that is a good indication that the particular section examined needs some review. Here is the description:

Through careful investigation, it is determined that for a certain population of 1000 dogs, the number of bones buried by individuals in 1977 averages 8.5 with a standard deviation of 2.3. Bone burying scores are normally distributed for the population.

1. Assuming that you didn't know the above parameters and were interested in estimating the population mean:

   a. why would a random sample be important?

   b. how might you select your sample (there is no one correct answer)?

   c. what would be considerations involving sample size (e.g., advantages/disadvantages of large vs. small)?

2. Assume that sampling (in a hypothetical case) was repeated over and over such that all unique samples of size n were characterized in terms of the following statistics. Which of these statistics would have Expected Values equal to their corresponding population parameters? In other words, which would give unbiased estimates?

   a. $\overline{X}$

   b. Md

   c. range

   d. r (population $\rho = +.84$)

   e. r (population $\rho = 0.00$)

   f. variance ($n-1$ in denominator)

   g. variance (N in denominator)

3. What is meant by Sampling Distribution?

4. Describe how the following factors influence the nature of a Sampling Distribution: (a) the statistic, (b) sample size, and (c) the population.

____

____

____

____

____

5. If a Sampling Distribution of means was derived from samples of $n = 16$ in the bone burying example (see initial description) what would be known about its characteristics in terms of: (a) form, (b) Expected Value, and (c) Standard Error?

____

____

____

6. If the Sampling Distribution were based on samples of $n = 25$, which of your answers to the preceding question (#5) would change?

____

____

7. In reference to the bone burying example, suppose a random sample of 4 dogs was selected and characterized in terms of the average number of bones buried by individuals in 1977. What would be the probability of obtaining:

________a. a sample mean greater than 9.0?

________b. a sample mean less than 7.0?

Assume the sample was based on $n = 9$. What would be the chances of obtaining:

________c. a sample mean <u>between</u> 8.3 and 8.7?

________d. a sample mean <u>between</u> 8.5 and 9.5?

________e. (SUPER BONUS) Above what score would 5% of the possible sample means fall?

UNIT VIII

HYPOTHESIS TESTING

A. General Objectives

In education and the behavioral sciences, the need to design and conduct experiments in studying phenomena is obvious. The present study of statistics has been to a large extent an attempt to equip the student with an understanding of how statistics can be used to describe a set of data. In this unit we will go one step further in using statistics to test hypotheses (predictions or "educated guesses") in a logical, objective manner. After completing this unit, the student should be able to understand hypothesis testing whenever he/she confronts it in the literature and be able to employ it in conducting and analyzing simple experiments on his/her own.

B. Specific Objectives

8.1 Describe the rationale for hypothesis testing

8.2 Define null hypothesis and alternative hypothesis

8.3 1. Differentiate between Type I and Type II errors
2. Select level of significance

8.4 Test hypotheses when population parameters are known (z test)

8.5 Test hypotheses when population parameters are unknown (t test)

8.1 Instructional Unit: Rationale and Basic Procedures

The reasoning behind hypothesis testing follows directly from the work we have done with Sampling Distributions. Remember in the last section of Unit VII we used the normal curve model and Sampling Distributions to determine the probability of obtaining certain values of $\bar{X}$. For example, if we know that the population of a certain school has a mean IQ of 100 and a standard deviation of 15, we can use this information to answer questions such as "If I select 10 students at random from that population, what would be the probability that the mean of the sample would be between 95 and 105?" If you do not understand how Sampling Distributions can be used to answer that, you are in need of a review--go back to the last unit in Unit VII now.

We will try to add a little more to the above reasoning and see how to test a hypothesis about the value of a parameter using sample statistics. First, keep in mind that hypothesis testing is a decision-making process. Researchers do not conduct experiments simply for the purpose of "seeing what happens." Good researchers always have a hypothesis (prediction or educated guess) of what the experiment will eventually show. Examples of hypotheses are: "I hypothesize that students learn more under programmed instruction than under conventional instruction," "I hypothesize that the students at my school are brighter than the average student in the city," "I hypothesize that Drug A will be more effective than Drug B in reducing the incidence of heart attacks,"... and so on... Once a hypothesis is formulated, the researcher designs his experiment for the specific purpose of testing that hypothesis.

Hypothesis testing is grounded on the notion of probability. In education and the behavioral sciences, our experiments never conclusively support one hypothesis over another. Given the results of one or more experiments, we can never say that one hypothesis is definitely more valid than another. We can say, however, that on the basis of our results, one hypothesis is probably more valid than another. The more striking the results, the more certainty we attach to our conclusion; the more probable our conclusion becomes.

Suppose one hypothesizes that the students in Genius Junior High School have higher IQs than the average student in the city. To test this hypothesis, nothing fancy is really needed. We administer an IQ test to a sample of Genius Jr. High students and compare the results to the IQ scores for the city as a whole (which we get from the city school administration). Suppose results show that the Genius Jr. High $\bar{X} = 127$ and the city mean $(\mu_X) = 100$. Can we now conclude with all certainty that positively the Genius Jr. High students are brighter (in terms of IQ) than the average student in the city? If you remember and understand the material on Sampling Distributions, you would correctly conclude that such a conclusion would be unjustified.

Sampling Distributions illustrate the concept of sampling fluctuation. If we select all unique samples of the same size from the same population, their means, $\bar{X}$'s, will vary: some will be equal to the population mean (μ_X), some will deviate slightly, and some will be real weird (extremely high or low). How do we know for sure that the sample we tested from Genius Jr. High is a "good" sample? Sure, we used random selection procedures in choosing the sample, but, as we know, the means of even randomly selected samples vary from one sample to the other.

All that we know for sure is that the mean of our sample was equal to 127; and the μ_X (for the city) = 100. We could justifiably make the following conclusion: judging from our results, it appears probable (not certain) that the average IQ of Genius Jr. High students is higher than the average IQ of students in the city. We cannot be certain about this because we have no way of knowing for sure that the sample we tested (although randomly selected) was "perfectly" representative of the Genius Jr. High student population. If we wanted to take the time and trouble to test every Genius Jr. High Student, then we could make a certain conclusion. If μ_X for Genius Jr. High turned out to be 125, we could say positively that this particular school population has a higher average

IQ than the city student population. Thus, in <u>hypothesis testing</u> we generally use samples to make inferences about populations; because we use samples (to save time and money), we lose the capability of being certain in our judgments. We, therefore, become restricted to expressing our conclusions in terms of probability statements.

8.1 POSTTEST (answers in back)

Briefly describe why probability rather than certainty is the rule in hypothesis testing.

____

____

____

____

8.2 <u>Instructional Unit</u>: The null and alternative hypotheses

Whenever you formulate statistical hypotheses, they should be exhaustive, covering all possible outcomes of your experiment or study. For this reason, we refer continually in statistics to what are called the <u>null</u> hypothesis and the <u>alternative</u> hypothesis (hypotheses). The conventional approach in experimentation is to always state a hypothesis of "no differences," called the null. The null hypothesis which we will symbolize as "H_0" calls attention to the possibility (hypothesis) that there are no real differences between the populations you are comparing in your experiment (this idea will be better explained later).

Specifically, in statistical practice, the <u>null</u> hypothesis (i.e., the hypothesis of "no differences") is the one that is actually tested. If there is sufficient evidence to reject it as being false, one can conclude, within certain probability limits, that differences <u>do</u> exist. If there is <u>not</u> sufficient evidence, the null must be retained. As difficult as the logic may at first appear, the objective really is to gather evidence that is convincing enough to "disprove" certain assumed conditions (i.e., the null). If such evidence is present, the null is rejected in favor of the decision that differences (probably) do exist. In the absence of such evidence, we don't "prove" the null, but merely retain it as a condition that still remains possible or acceptable. That is why in reading research papers you will encounter reports of the null being rejected or retained, but never of it being "proven." The null, like the condition of innocence in law, is the "assumed" state of affairs until there is sufficient cause to doubt its validity. In fact, the term "null" derives from the conception in philosophy of science that the orientation of empirical research is directed to <u>nullifying</u> hypotheses. We will proceed in a lighter vein below (take a deep breath or two before continuing).

Given the idea of H_0 (the null), the alternative hypothesis, symbolized as H_a covers all other possible outcomes of the experiment. Given H_0 and H_a, you have exhausted the field in terms of the possible findings of your experiment. Let's try an example.

Example:

We perform a study to determine whether the mean IQ of Genius Jr. High students is higher than the mean IQ of all junior high students in the city. We test this by selecting a random sample of 25 Genius Jr. High students, and determining the mean IQ of the sample on the Stanford Binet IQ test. We compare this score to the mean IQ of all junior high students in the city (that is, the Jr. High population), data which we collect from the School Administration records. Keep in mind that what we are really asking is "Does the Genius Jr. High School population (we'll symbolize this as u_1) differ in terms of IQ from the city population (we'll symbolize this population mean as u_2)?" We can directly determine u_2 from city records, but we must infer u_1 (Genius Jr. High population mean) on the basis of our random sample.

What is the null hypothesis in this study?
What is the alternative hypothesis?

Ans.: First, state your null hypothesis, keeping in mind that the null represents the prediction of "no differences":

H_0: u_1 (Genius Jr. High) = u_2 (City)

Note that the null, by stating $u_1 = u_2$ indicates the possibility that there are really no differences between the average IQs of the two populations.

Then state your alternative hypothesis (hypotheses), keeping in mind that they must represent all remaining possible outcomes. We can state one exhaustive alternative hypothesis very simply as follows:

H_a: $u_1 \neq u_2$

Note that when used in conjunction with the null, the alternative covers all possibilities: that is, you will either find the two population means to be equal or you will find them to be not equal.

You can get more specific, however, if you so desire:

H_a: $u_1 < u_2$ (which suggests that the city mean will be greater than the Genius Jr. High Mean)

But you cannot stop here: hypotheses must be exhaustive. You need at least one more alternative:

H_a': $u_1 > u_2$ (suggesting that the city mean will be less than the Genius Jr. High Mean)

This time our three hypotheses cover all possible outcomes: a) no differences (null); b) city greater than Genius (H_a:); c) city less than Genius (H_a':).

Example:

A researcher believes that males who were born in the month of February are taller than average. From the government census data he finds the mean height of all males over 17 years of age in the U.S. He also uses the census data to contact and measure a random sample of 100 males (over 17) who were born in February. He will use this sample to generalize (estimate) the population of all males born in February.

State the null and alternative hypotheses:

One possible answer: $H_0 : u_1$ (Feb. males) $= u_2$ (all males)

$H_a : u_1 \neq u_2$

Note that these two hypotheses are exhaustive in terms of outcomes.

Another possibility: $H_0 : u_1 = u_2$

$H_a : u_1 < u_2$

$H_a' : u_1 > u_2$

Note that these three are exhaustive and thus correct.

An incorrect answer: $H_0 : u_1 = u_2$

$H_a : u_1 < u_2$

Note that these do not cover all possibilities; what if u_1 turns out to be greater than u_2?

An incorrect answer: $H_0 : u_1 = u_2$

$H_a : u_1 \neq u_2$

$H_a' : u_1 < u_2$

Although these are exhaustive they are not mutually exclusive; suppose you find that u_1 is less than u_2? Do you accept H_a (which is technically correct) or H_a' (which is also correct). You cannot have two alternatives supported at the same time. H_a should be changed to $u_1 > u_2$, or H_a' should be eliminated.

Final Notes:

1. Statistical hypotheses should always be stated in terms of a null hypothesis (symbolized H_0) and one or more alternative hypotheses (symbolized H_a, H_a'). The null hypothesis represents a prediction of "no differences"; for example, $u_1 = u_2$. It should be noted, however, that the null can be expressed in terms of actual parameter values. For instance, a researcher may want to test the hypothesis that the mean IQ of City X is greater than the mean IQ of the U.S. population, which he knows to be 100. Using the procedure we are accustomed to, he can specify his hypotheses as follows:

H_o: u_1 (city) = u_2 (U.S.)

H_a: $u_1 \neq u_2$

But since he is testing the prediction that the city mean will be greater than 100, he could specify his hypotheses in a slightly different manner (one that really means the same thing, however):

H_o: $u_1 = 100$

H_a: $u_1 \neq 100$

Instead of using the parameter symbol for the U.S. population, u_2, he used the parameter value, 100. The choice is pretty much arbitrary.

2. The specifications of hypotheses using the null and alternative(s) should be exhaustive and nonoverlapping (i.e., no two alternatives should cover the same possible finding).

8.2 POSTTEST (answers in back)

1. A school guidance counselor is interested in determining whether the seniors in his high school receive higher SAT scores than average. He knows from Educational Testing that the mean SAT score of all seniors in the United States is 500. He selects 30 seniors at random from his school, and compares their average SAT performance to the United States norm.

   Specify the null hypothesis and the alternative(s). (Note that there are several possibilities.)

   ____

   ____

   ____

2. For each set of hypotheses shown below, indicate whether the specification is incorrect or correct. If incorrect, explain why.

Set 1	Set 2	Set 3
$H_o : u_1 \neq u_2$	$H_o : u_1 = 50$	$H_o : u_1 = u_2$
$H_a : u_1 = u_2$	$H_a : u_1 \neq 50$	$H_a : u_1 < u_2$
		$H_a' : u_1 > u_2$

Set 4	Set 5	Set 6
$H_o : u_1 = u_2$	$H_o : u_1 = 23$	$H_o : u_1 = u_2$
$H_a : u_1 > u_2$	$H_a : u_1 = 22$	$H_a : u_1 = 100$
	$H_a' : u_1 < 23$	

8.3.1 Instructional Unit: Type I and Type II errors

In review, the basic procedure used in hypothesis testing is to:

1. First assume that the null hypothesis is true; that the population mean you are estimating (u_1) is actually equal to the population mean to which it is to be compared (u_2).

2. Having made the above assumption (for example, $u_1 = 100$), consider the Sampling Distribution for u_1, based on samples of size n.

3. Given the above Sampling Distribution, determine whether the sample mean, $\bar{X}$, is "probable" or "improbable" (we will see later how to decide more objectively what is "probable").

4. If the sample mean, $\bar{X}$, turns out to be "probable," the researcher retains the null hypothesis, and, in essence, concludes that there is no difference between u_1 and u_2.

5. If the sample mean, $\bar{X}$, turns out to be "improbable," the researcher rejects the null hypothesis and accepts one of the alternatives. What he has done, in essence, is to reason that given the Sampling Distribution that represents the null condition ($u_1 = 100$), his sample mean, $\bar{X}$, turns out to be improbable--one that would be found for very few of the infinite samples of size n. Therefore, having nothing else to go on excepting his sample mean, he concludes that the above Sampling Distribution probably doesn't apply to the population in question. The null hypothesis is rejected in favor of one of the alternatives.

At this point it is not important for you to fully understand how the researcher decides that his sample mean is "probable." As you will find out later, this is pretty much an arbitrary determination which may differ from researcher to researcher and from study to study. Nor is it important for you to memorize the five steps listed for hypothesis testing. It is essential, however, that you do acquire an understanding of the logic, as reviewed below:

If the null is true, then my sample of n cases will relate to a Sampling Distribution with $E(\bar{X}) = u_X$ and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

Given that particular Sampling Distribution, does my sample mean, $\bar{X}$, appear probable or improbable?

If $\bar{X}$ appears improbable, then there is a good chance that the parameter values, u_X and σ, used to construct the Sampling Distribution do not describe the population from which the sample was drawn. Therefore, it seems reasonable to reject the null hypothesis and accept the alternative. The alternative says, in essence, that there are differences between u_1 and u_2.

* * * * * * * * * * * * * * * * * * * *

Trial of the Tenpins: A Tale About Nullifying the Null

At great risk of confusing the issue, but out of an unwillingness to abandon our "law" analogy, a final attempt will now be made to illustrate the logic of hypothesis testing in terms of deciding between innocence and guilt in trial. Use your imagination and pretend that <u>deviancy</u> in the students served by schools is considered a crime in the fictional country, Untied Stakes. School XYZ is up for trial, accused by a twelfth-grade double agent of serving a student populace that differs in bowling average from the "accepted" national standard of 100 as a school average (with $\sigma = 15$). Based on the convention of assuming innocence until guilt is established, the jury must begin with accepting the premise:

u_1 (School XYZ) = 100

The task of the prosecutor is to present evidence to convince the jury that this condition is <u>improbable</u>, given that actual conditions cannot be enacted (all 5,000 XYZ students cannot be tested). If he succeeds in presenting this evidence, the jury will have no choice but to accept the alternative, $u_1 \neq 100$, and chances are that the judge will pronounce the XYZ student population as <u>deviant</u>! With the original records unavailable and conditions making it impossible to test every student involved, testimony (in this case, bowling scores) <u>is</u> obtained from a random sample of 25 students. The sample averages 108 ($\bar{X} = 108$) on the lanes. The prosecutor then takes this evidence and tries to argue that it makes innocence ($u_1 = 100$; the null) seem extremely improbable. The defense attorney argues that this seemingly incriminating evidence is merely circumstantial (obtained by unusually good rolls); specifically, he states that it doesn't preclude the possibility that if <u>all</u> XYZ students were tested, rather than just 25, u_1 <u>would</u> = 100. Suppose that you were a member of the jury, how would you vote?

<u>Highlights of the Actual Trial Proceedings</u>

The prosecutor's closing speech goes as follows: "Ladies, gentlemen, and others, the Defendant, School XYZ, is obviously guilty, based on the evidence that has been rolled before you on these lanes. If we were to assume that $u_1 = 100$ (with $\sigma = 15$) as my esteemed colleague (the Defense Attorney) claims, then a Sampling Distribution based on all unique combinations of 25 students selected from this population would be characterized by $E(\bar{X}) = 100$ and $\sigma_{\bar{X}} = 15/\sqrt{25} = 3.0$. Using the normal curve probability model which you have before you, this would indicate that roughly 68% of the samples of $n = 25$ would average between 97 and 103 in their bowling scores (between $\pm$ 1 S.E. from the mean); roughly 95% would have averages between 94 and 106 (between $\pm$ 2 S.E. from the mean). Yet the random sample presented as evidence in this trial was shown to average 108. Again, if you will refer to your normal curve table, you will quickly see that after converting the submitted score of 108 to an appropriate z score, $(108-100)/3 = z$, the resultant value of +2.66 is so high that it would be obtained less than 1% of the time (50% - 49.61% = .39%), a highly unlikely occurrence. Thus, I submit that for School XYZ,

u_1 does not equal 100, but represents a value somewhat higher, a value that could more appropriately account for $\bar{X} = 108$. DEVIANCY of this school is definitely indicated, and a vote of GUILTY is the only possible decision."

The defense attorney's speech was rather long, as he tried to show that the conditions of testing the sample were unusual relative to "typical" bowling events (representiveness of conditions should always be an important concern in using sample data to make inferences about "true" states). He argued that the excitement of using official bowling lanes added points to the students' scores (and hoped that this argument would do the same to his case). His main arguments, however, can be summarized as follows: "My colleague has taken you through the normal curve probability table in an attempt to have you believe that $u_1 = 100$ does not apply to the Defendant. It has been stated that if such a parameter did apply, an $\bar{X}$ as high as 108 would be obtained from a random sample of $n = 25$ less than 1% of the time. If you, the members of the jury, view your role as it is defined by law--to assume innocence unless guilt is proven--it is that very figure that must lead you to cast your vote for 'innocence.' Given the infinite range of normal curve values, sample $\bar{X}$'s even higher than 108 are theoretically possible, given a population $u_X = 100$. The obtained $\bar{X}$ of 108, although higher than most of the $\bar{X}$'s we would obtain if other samples had been tested, is merely a product of sampling fluctuation. School XYZ does average 100 overall, and if the trial were repeated, we could just as easily obtain $\bar{X} = 92$ as we did the 108. Innocence cannot be ruled out based on the evidence submitted. Therefore, the only possible decision is for a vote of NOT GUILTY."

How would you vote now?

The Outcome

The most heated discussion among the jury concerned whether or not the bowling scores used as evidence were obtained under representative conditions. After failing to resolve this for 38 successive days, they finally reached unanimous agreement that the data should be regarded as valid. With this decided, the verdict submitted was GUILTY. The judge, feeling that leniency was in order as this was a first offense, sentenced the school to serve six months at a fast food restaurant, after which it could return to providing education if students' bowling averages were lowered to conform with the national average.

Epilogue

The arguments presented by the opposing attorneys are much the same as the reasoning applied by the statistician in deciding whether to reject or retain the null hypothesis. In hypothesis testing, as in court cases, decisions are not based on certainty but rather on probabilistic judgments that weigh the amount of evidence brought into view. In the example, the prosecutor showed that an $\bar{X}$ as high as 108 will occur less than 1% of the time given the assumptions of the null (i.e., assumptions of innocence). The jury, forwarding a verdict of "guilty," presumably used some subjectively selected criteria to decide that the 1% probability value was sufficiently small to reject the validity of those assumptions. In statistics, as we shall see shortly, conventional practice frowns on subjective choices of such decision criteria by establishing ones expected to be applied in all (or most) published work. But, even though in our example trial, the "weight of evidence" seemed to go against the school population averaging 100, can we be certain that the jury made a correct decision? We cannot, as there is still some probability that the 108 was simply an extreme value in the Sampling Distribution that does define $u_X = 100$, $\sigma = 15$, $n = 25$. Thus, since statistics are only estimates of population parameters (and testimony in a courtroom only remembrances of actual happenings), the decision, whether to reject or not reject the null (guilty or not guilty), will always be associated with some degree of doubt.

* * * * * * * * * * * * * * * *

If you understand the logic of hypothesis testing as described, feel free to merely skim the next few pages (up to Type I and Type II errors). If you are still fairly fuzzy and uncertain, do not give up hope - additional explanation will be provided below.

Additional Explanation: (Skip or skim this section if you understood the preceding material.)

Let's continue to work with the example used earlier, but without the analogy to law. A researcher knows that IQs are distributed normally in the U.S. population with $u_X = 100$ and $\sigma = 15$. He hypothesizes that the students in School XYZ are brighter than average, which will be evidenced by their higher IQ scores. He cannot possibly test every single student in the school, so he decides to use a random sample of 25 cases. Keep in mind that he is really trying to compare one population mean (the U.S. average, which we'll call u_2) against another population mean (School XYZ average, which we'll call u_1). But he cannot determine u_1 directly, so he uses a sample to estimate it. He specifies his hypotheses as follows:

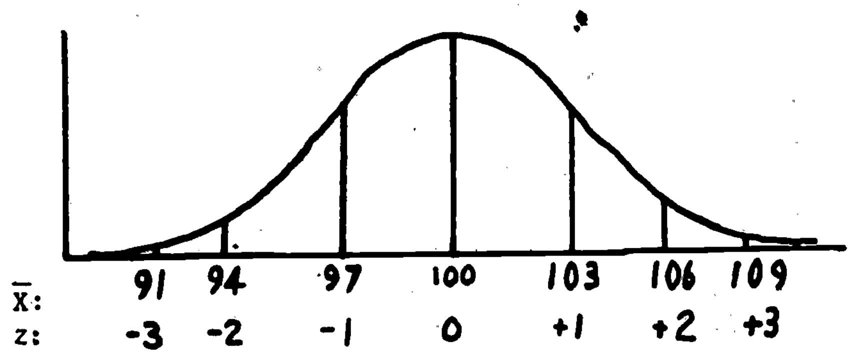
$$H_0: u_1 = u_2 \text{ (or 100)}$$

$$H_a: u_1 \neq u_2 \text{ (or 100)}$$

He goes out and tests this sample of 25 students. He computes their mean IQ and finds that it is $\overline{X} = 109$. Now he must make a decision: "Should I retain the null or reject it? Does $\overline{X} = 109$ suggest that u_1 (school pop.) = u_2 (U.S. pop.), or that $u_1 \neq u_2$?"

To answer the question, he reasons as follows: if the null hypothesis is true, then the population of School XYZ has $\mu_X = 100$ and $\sigma = 15$. If an infinite number of samples of size 25 were drawn from this population and tested on IQ, the Sampling Distribution of means would have $E(\bar{X}) = 100$ and $\sigma_{\bar{X}} = 3$.
(By now you should know how we derived these values.)

The Sampling Distribution would look something like this:



Since the researcher is totally familiar with the normal curve and its assumptions, he regards the Sampling Distribution as something that shows the probability of obtaining size-25 sample means of different values selected from a population where $u_X = 100$ and $\sigma = 15$.

Do those parameters describe School XYZ?

We have no way of knowing for sure, but we can look at our sample mean, $\bar{X} = 109$, and decide whether it is or is not a "probable" occurrence, assuming that those parameters do describe School XYZ. If it turns out to be a "probable" occurrence, then there is no reason to believe that School XYZ does not have a mean IQ of 100; thus, we will retain the null. If it turns out to be an "improbable" occurrence, then we have good reason for suspecting that School XYZ has $u_X \neq 100$; therefore, we reject the null and accept the alternative.

Getting back to the researcher; he looks at the above Sampling Distribution, and knowing about normal curves and probability, he reasons that a sample mean = to or greater than 109 ($z = +3.00$) will only occur about .1% of the time (50.00% - 49.87% = .13%, the latter figure corresponding to the shaded portion of the curve).

Could his sample of 25 students with $\bar{X} = 109$ be one of that .1%? Perhaps, but the researcher decides that probably that is not the case. His sample mean of 109 appears very "improbable," assuming the sample was drawn from a population with $u_X = 100$. Therefore, he concludes that his sample was probably drawn from a population (specifically, School XYZ) where $u_X \neq 100$.

With that conclusion, he has rejected the null and accepted the alternative.

Can he be entirely certain that his conclusion $u_X \neq 100$ is a correct one? No, he can't be certain because some of the sample means (.1% to be exact) selected from a population where $u_X = 100$ will be as high as 109. Maybe the sample he selected was, by chance, exceptionally bright on IQ tests. But given the data, it seems much more reasonable to assume that School XYZ probably has a mean IQ higher than 100, and that this particular sample with $\bar{X} = 109$ was merely reflecting that parameter value.

Another Example (Study, Skim, or Skip, based on need)

A researcher knows that height of males over 17 in Texas is normally distributed with $u_X = 72''$ and $\sigma = 10''$. He wishes to determine whether males over 17 in his home town, Austin, are taller than the state average. He can't possibly test every male over 17 who lives in Austin so he decides to use a random sample of 100 cases. He measures the 100 men in his sample and finds that $\overline{X}$ = 72.5".

He states his hypotheses as follows:

H_0: u_1 (Austin) = u_2 (Texas)

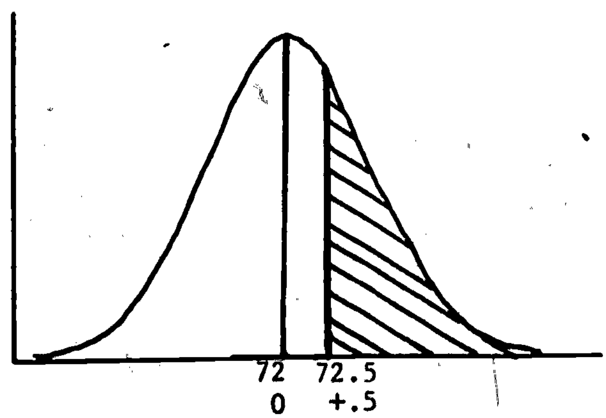
H_a: $u_1 < u_2$

H_a': $u_1 > u_2$

To decide on whether to reject or not reject the null, he reasons as follows: If the null hypothesis <u>is</u> true, then the population of males over 17 in Austin, Texas, has u_X = 72. If an infinite number of samples of size 100 were drawn from this population and measured on height, the Sampling Distribution of means would have $E(\overline{X})$ = 72 and $\sigma_{\overline{X}}$ = 1.

This particular Sampling Distribution would look something like this:

If the null is true:



Sample $\overline{X}$'s
score

The researcher regards the above Sampling Distribution as showing the probabilities of obtaining sample means of different values when an infinite number of size-100 samples are selected from a population where u_X = 72" and σ = 10.

Does that description apply to the population of all males over 17 in Austin, Texas?

The researcher has no way of knowing for sure, but he can look at his sample mean and determine whether or not it is a "probable" occurrence, assuming that the u_X of Austin, Texas <u>is</u> 72". If it turns out to be improbable, then there is good reason for rejecting the null; therefore, there is good reason for suspecting that Austin is characterized by $u_X \neq 72"$.

Okay, so the researcher examines his sample mean, $\bar{X}$ = 72.5", in relation to the Sampling Distribution of all means based on $u_X = 72$. He notes that the Standard Error of the Sampling Distribution is $\sigma_{\bar{X}} = 1$. Therefore, for a sample mean of 72.5", $z = +0.50$ [$z = (72.5 - 72)/1$]. Since he is familiar with the normal curve and probabilities, he concludes that a score within the range $z = +0.50$ is really a very common occurrence, not very improbable at all.

On the basis of this reasoning, he decides that the data does not support rejecting the null. A sample mean, based on 100 cases, which is equal to 72.5" would not be uncommon or unusual if the sample were selected from a population where $u_X = 72$.

His final conclusion is one of "no differences" between the height of males in Austin, Texas, and the height of males in all of Texas.

Note: At this point you should be more concerned with getting a "flavor" of the logic used in hypothesis testing than with trying to understand exactly how "probable" and "improbable" are to be determined.

Type I and Type II Errors

Throughout this unit, we have been trying to stress the point that hypothesis testing is a game of probabilities. If you could measure every single individual in both populations you are comparing, no guesswork would be involved. But since you must use sample data to estimate at least one of your parameters, you can never be certain that the conclusion you reached is absolutely correct. Suppose in the above example the researcher measures his sample of 100 males and finds $\bar{X} = 76"$. He probably would reject the null by reasoning that a sample mean of 76" is exceedingly unlikely to be found if Austin, Texas males really averaged only 72" in height. That may be true, but it is still possible that the sample he happened to select was "one in a million," consisting of 100 very tall males. Thus, although it seems improbable, it is still possible that he would be making an error by rejecting the null. The error would be an unavoidable one, but it is nevertheless important for you to understand that whether you decide to reject or retain the null, the decision is always subject to errors.

There are two types of errors to which our conclusions are subject. Only one type can be made for a given experiment. We will soon find out why.

The first error with which we should be concerned is called a Type I error. We make a Type I error when we falsely reject a true null hypothesis. This means that on the basis of a sample statistic which appears "improbable," we conclude that the null hypothesis is false; i.e., $u_1 \neq u_2$. In actuality, however, the null hypothesis is true--the two populations being compared are really identical; it's just that the sample we selected to estimate the parameters for one population was, by chance, unusual, giving us a false impression.

Example:

A researcher tries to determine whether the students in School XYZ have math scores higher than the U.S. population norm, which is 100. If he had time to actually test the total School XYZ population, he would find (to the school administrators' dismay) that the mean score for the school is exactly equal to 100; XYZ has "typical" students--u_X (School XYZ) = 100. But, as is normally the case, it is simply not practical in terms of time and money to test the IQs of every student at XYZ. Therefore, he uses a sample of 49 randomly selected students from all grades. His results show that $\bar{X} = 115$, causing him to consider it "improbable" that such a sample could have been selected from a population where $u_X = 100$. He rejects the null, and to the school administrators' collective delights, accepts the alternative, $u_X > 100$.

The researcher has unknowingly made a Type I error. Why? The sample he selected was, by chance, an exceedingly unusual one. He did use random selection procedures, and thus, cannot be faulted. Chances are that the vast majority of samples of 49 students selected from the XYZ population would yield $\bar{X}$'s very close to 100. But, for whatever reason (actually the reason is sampling fluctuation), his particular sample of 49 happened to be comprised of very bright students from XYZ. The administrators will now happily report to the community that XYZ is a "bright" school, but if the study is done again, the results will probably show the truth--XYZ is merely average.

Anytime you reject the null, there is some probability that you have made a Type I error.

The second type of error, Type II, occurs when we retain a false null hypothesis. This means that on the basis of a sample statistic that appears "probable," we conclude that there is no evidence to support the rejection of the null.

In actuality, however, the null hypothesis is false -- the populations we are comparing really do differ; it is just that the sample we selected to estimate one population parameter happened to be unusual, giving us a false impression.

Example: A researcher tries to determine whether students in School ABC have higher math scores than the U.S. population norm, which is 100. If he had time to test all 3,000 of ABC's students he would find that this is really the case -- ABC has a whopping μ_X of 127! He doesn't have much time or money so he tests a sample of 49 students, selected at random from all grades. His results show that $\bar{X}$ = 101, causing him to conclude that it is very probable that his sample would have been selected from a population where μ_X = 100. He thus has no evidence with which to reject the null, and must retain it. He sadly reports to the school administrators that their students are merely "average." But, if the same study were to be repeated, the results would probably show that School ABC is really quite a bit above the national average in student math achievements.

This researcher has unknowingly made a Type II error. Why? The only evidence he had for accepting or rejecting the null was his sample statistic, $\bar{X}$ = 101. By chance, the 49 students comprising that sample, although randomly selected, happened to be considerably lower than the actual ABC average, μ_X = 127. The sample data gave him a false impression.

Anytime you accept (retain) the null hypothesis, there is some probability that you have made a Type II error.

The relation between possible outcomes and Type I and II errors is shown in the table below:

Actual Parameter Condition

		H_0 TRUE	H_0 FALSE
Outcome of Experiment	H_0 rejected	Type I error	CORRECT DECISION
	H_0 retained	CORRECT DECISION	Type II error

Note on the table that you cannot make both types of errors at the same time. If you decide to reject the null hypothesis, you have made either a) a Type I error, provided the null is true in actuality; or b) a correct decision, provided the null is false in actuality.

If you decide to retain the null, you have made either a) a correct decision, provided the null is true in actuality; or b) a Type II error, provided the null is false in actuality.

Since, in a given experiment, you can either reject or retain (but not both), you run the risk of either a Type I or Type II error (but not both).

Practice Questions: For each of the following, indicate whether the researcher has made (a) a Type I error; (b) a Type II error; or (c) a correct decision

(use a cover sheet to block answers from view)

1. On the basis of his sample data, a researcher decides to reject the null by concluding that girls in the ABC Day Care Center are better than average in physical skills. In actuality, however, had he the time to measure the whole ABC population, he would have found this population to be merely average. Error?

ans. The researcher has made a Type I error by rejecting a "true" null hypothesis.

2. The average SAT-Verbal score for all high school students is 500. A researcher tests a random sample of 50 students who live in Montana and finds $\bar{X} =$ 498, a "probable" statistic. He decides to retain the null. In actuality, the average SAT-Verbal score for the Montana population is exactly 500.

ans. The researcher has made a correct decision by retaining the null; there are really no differences between u_χ (Montana) and u_χ (U.S. population).

3. Suppose in the above example, the actual Montana average is $u_\chi = 450$. What then?

ans. The researcher has made a Type II error by failing to reject (retaining) a false null hypothesis. In actuality, $u_1 \neq u_2$; his sample data, although appearing to be "probable" was not a true indicator of the Montana average.

Keep in mind that Type I and II errors are totally unintentional, and, for the most part, undetectable in practice. All you can do is decide whether to retain or reject the null on the basis of your sample data. Since your sample is selected at random, you are correct in assuming that it will provide a reasonable estimate of the population parameters in question. Due to the problem of sampling fluctuation, this will not always be the case; sometimes your sample data will be misleading. When this happens, you may end up rejecting a null hypothesis that is actually true (Type I), or retaining a null hypothesis that is actually false (Type II). Unless you replicate the study, you will never be aware that an error was made.

(8.3 Posttest includes the following subsection)

8.3.2 Probability Level

If you have an alert, inquiring mind you have probably been somewhat frustrated by the discussion on hypothesis testing procedures and Type I and II errors. The reason for your frustration might revolve around the question of determining when your sample statistic should be considered "probable" (suggesting retention of the null) and when it should be considered "improbable" (suggesting rejection of the null). You may ask: "Aren't there any rules to the game?" "Are these important decisions left totally up to the whims and moods of the researcher?" "Isn't there a way to determine objectively what is probable or improbable?"

The answer to all of the above questions is "clearly yes and no."

You may recall from our fictitional account of the "Trial of the Tenpins," the jury voted "guilty" on the basis of the prosecutor's argument:

> IF School XYZ $\mu_X = 100$, an $\bar{X}$ as large as 108 would be obtained in less than 1% of all random samples of $n = 25$.
>
> An $\bar{X} = 108$ <u>was</u> obtained for the sample selected.
>
> 1% is a "very low" probability.
>
> Therefore, it is "probable" that μ_X for School XYZ > 100.

The jury subjectively decided that the $\bar{X} = 108$, p (probability) < 1%, was sufficient to reject the null condition of innocence. If a much more conservative jury had been involved perhaps the null would have been retained - stronger evidence (maybe an $\bar{X} = 109$) might have been needed to convince them. At the conclusion of that section, it was implied that such subjectivity in selecting decision criteria, although unavoidable where actual probability estimates cannot be derived (as in <u>real</u> trials), is reduced to some degree in research through the adoption of conventional standards for retaining and rejecting the null.

However, even though there are rules to the game, much in the way of making decisions is still left up to the individual researcher. In the remainder of this section, we will discuss some of these rules and procedures. Sometimes they may seem difficult to understand, and therefore will require careful reading. If you decide to merely skim over this section, you will most likely be wasting your time. Give it your full attention.

Prior to the actual running of an experiment, it is expected that the researcher will specify what is referred to as a <u>level of probability</u>. This specification indicates in clear, precise, and objective terms how different the sample statistic, $\bar{X}$, must be in relation to expected (probable) values in order to justify rejection of the null. The probability level, as you might expect, is specified in probability terms, rather than in score values. But through the use of Sampling Distributions and the normal curve, the researcher <u>will be able</u> to specify his acceptance or rejection levels in terms of real score values: For example, "I will reject the null hypothesis, if my sample mean turns out to be greater than 110 or less than 90 (since both levels appear improbable, assuming that the null were true)." Once these levels are clearly specified, he goes out and tests his sample. The important point is that the criteria for accepting or rejecting the null are selected before the data is collected, rather than on a <u>post hoc</u> basis.

How do you go about specifying ~~a~~ probability level?

The conventional choice of probability is usually the .01 or .05 level, although exceptions are many times entirely appropriate.

What does .01 or .05 mean?

Seemingly the best way to explain the use of probability level is through an example. But for now, we can say that the .01 level means that the sample statistic you obtain is only 1% probable, assuming the null is true; the .05 level means that the sample mean you obtain is only 5% probable, assuming the null is true. If you use a probability level of .01 and find that your sample data is less than 1% probable (assuming the null is true), you reject the null; if your sample data is more than 1% probable, you retain the null. The same thing holds for the 5% level, except that in this case you would be less conservative in your criteria for rejection. All this is probably fairly confusing, so let's apply it to an example situation and see how it works.

Example Study:

An investigator knows that IQ is normally distributed in the U.S. population with $u_X = 100$ and $\sigma = 15$. He wishes to determine whether the adolescent population in the local juvenile hall is average in IQ scores. He can't possibly test all 3,000 juvenile hall residents so he decides to employ a randomly selected sample of 25. He states his hypotheses as follows:

$$H_0: \; u_X = 100$$

$$H_a: \; u_X \neq 100$$

He decides to use the .05 level of probability (also called, level of significance, and symbolized as α alpha).

$$\alpha = .05$$

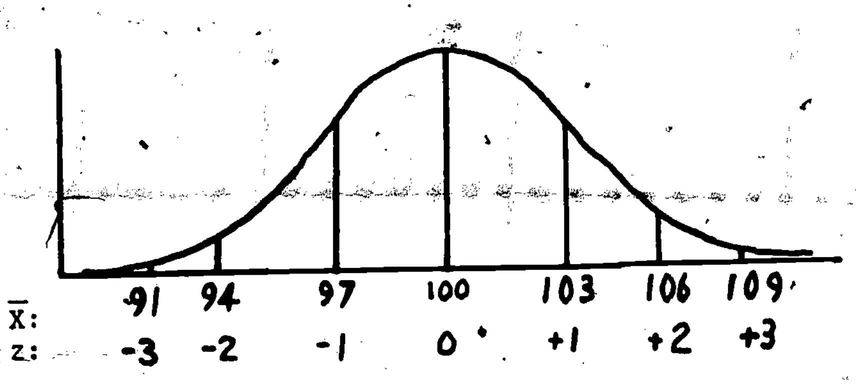
By specifying the significance (probability) level as $\alpha = .05$, he is saying, in essence, that if his sample mean (statistic) turns out to be more than 5% probable (assuming the null is true), he will retain the null; if it turns out to be less than 5% probable (assuming the null is true), he will reject the null. We will see how this works very soon.

First, he will reason that if the null hypothesis is true ($u_X = 100$), the means of an infinite number of samples of 25 cases selected from the juvenile hall population will be described by a Sampling Distribution where:

$$E(\overline{X}) = 100$$

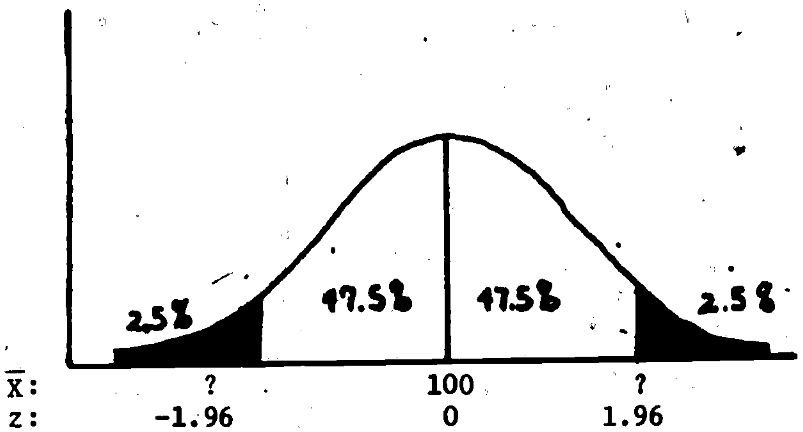
$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = 15/\sqrt{25} = 15/5 = 3$$

A sketch of this Sampling Distribution would look like this:



Since the researcher knows about normal curves and probability, he can examine the Sampling Distribution and reason as follows: "If this Sampling Distribution is valid ($\mu_{\bar{X}} = 100$), then there is a 68% chance that I will obtain a sample mean somewhere between 97 and 103 (between +1z and -1z), there is a 99.7% chance that my sample mean will fall somewhere between 91 and 109 (between +3z and -3z), and so on." If you're uncertain about how these probabilities were derived, check your "area under the normal curve" table. Won't 68% of the cases always fall between +1z and -1z in a normal distribution?

But he has specified his significance level as $\alpha = .05$. What does this mean? In "hypothesis-testing" language it means that the region of rejection for the null hypothesis (the "improbable" region) will include values that are not likely to occur more than 5% of the time. What is this region? We can pencil it in on the sketch shown below:



By examining the "area under the normal curve" table, we determine that 2.5% of the time, values in a normal distribution will fall above +1.96z, and 2.5% of the time, values will fall below -1.96z. Thus, 5% of the time (2.5% + 2.5%), values will occur that are either equal to or greater than +1.96z, or equal to or less than -1.96z. Since in this case $\alpha = .05$, the researcher will consider this extreme 5% region of the curve as improbable. It can be thought of as the "region of rejection." If the sample mean falls within this region (one of the penciled-in areas of the curve) it will be considered "improbable," and justify rejection of the null hypothesis. On the other hand, the area of the curve that falls between -1.96 and +1.96 can be thought of as the "region of acceptance" (it comprises 95% of the area). If the sample mean falls within that region, it will be considered "probable," and lead to retention of the null.

Do you understand how the shaded regions beyond the +1.96 and the −1.96 were selected? Remember that the identified significance level, $\alpha = .05$, defines an extreme sample statistic (one that is "deviant" enough to support rejecting the null) as that which has a probability of occurring 5% or less of the time assuming that the null is true. But an "extreme" value can be high or low, can't it? Thus, the area of rejection, encompassing 5% of the total normal curve area, must be equally divided between the high and low ends. The result has to be 2.5% at the top and 2.5% at the bottom. How do we derive z scores for these regions?

ans.: The same way we would if we were given the same problem in Unit IV.

Which z score has 2.5% of the area in a normal curve above it? The total area above z = 0 is 50%; thus, 50 − 2.5 = 47.5% between "our" z and the mean. Look up 47.5 (actually .4750) on the "normal curve" table and you will find that it most closely corresponds to z = 1.96. The symmetry of the normal curve (or the same mathematics if you've forgotten what symmetry means) gives z = −1.96 to define the 2.5% region at the bottom end. Now, how would the situation change had the researcher decided to use $\alpha = .01$? Here, he's being more cautious, as he's saying that extremity will be defined as statistics likely to occur 1% or less of the time given the null assumptions. Since extremity can be either high or low, the region of rejection must be balanced to leave half of 1% at the top (.5%) and half at the bottom (.5%). What z scores define these regions? The positive z must be one that has .5% of the scores above it and, therefore, 50 − .5 = 49.5% between it and z = 0. Look up .4950 on the table and you get z = 2.58. At the bottom, the region would have to be defined by z = −2.58. Compare these to the ± 1.96z's and it should be obvious that the .01 level is asking for "more convincing" evidence for rejection of the null than is the .05 level.

Without reading further, can you determine what score values would correspond to the 1.96z regions? The answer simply involves converting the z scores in question into raw scores, by use of the basic z score formula.

$$z = \frac{X - \bar{X}}{\sigma}$$

$$X = \sigma z + \bar{X}$$

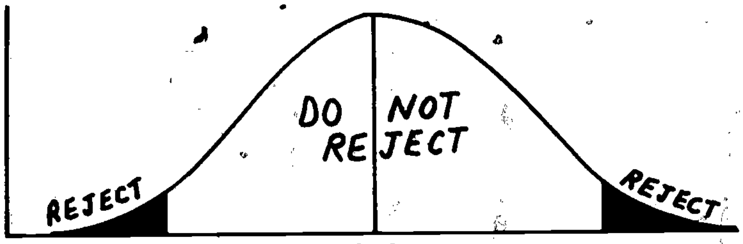
$X_1 = 3\,(1.96) + 100$ (this is the score corresponding to $z = +1.96$; see sketch)

$X_1 = 105.88$

$X_2 = 3\,(-1.96) + 100$ (this is the score corresponding to $z = -1.96$)

$X_2 = 94.12$

We can show these values on the Sampling Distribution as follows:



$\bar{X}$:	94.12	100	105.88
z:	-1.96	0	+1.96

Now the researcher can go out and test his sample of 25 juveniles. If the mean IQ of the sample, $\bar{X}$, turns out to be greater than 105.88 or less than 94.12, it will be considered "improbable."

He tests his 25 juveniles, and finds $\bar{X} = 90.45$. What is his decision?

He rejects the null. Using $\alpha = .05$, his sample mean appears "improbable." Such a value will be found less than 5% of the time when an infinite number of size-25 samples are selected from a population where $\mu_X = 100$. Therefore, there is very good reason to suspect that μ_X (Juvenile Hall) $\neq 100$. The mean of this population is most likely less than 100; otherwise, it seems extremely unlikely that $\bar{X} = 90.45$ would have been obtained.

But suppose, in actuality, μ_X (Juvenile Hall) is equal to 100. What type of error would the researcher be making?

ans. A Type I error: rejecting a true null hypothesis.

Example Study II

Suppose the researcher decided to use $\alpha = .01$ in the exact same study.

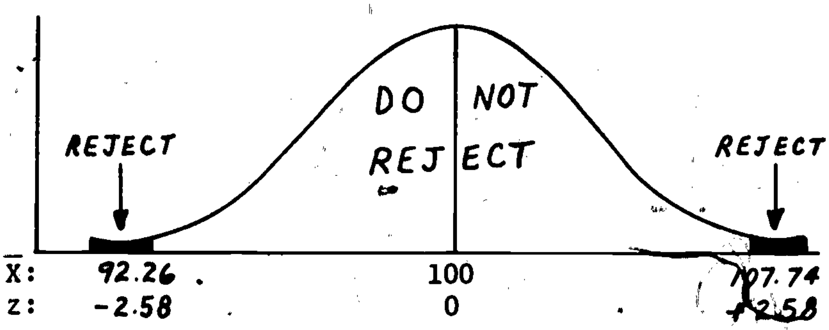
Everything would be the same, but this time the researcher is being more conservative in his criteria for rejecting the null. By using $\alpha = .01$, he is saying that he will reject the null hypothesis, if his sample mean is determined to occur only 1% of the time (or less) assuming the null is true. If he used the .05 level, he would reject the null if the sample mean was determined to occur 5% (or less) of the time.

Can you see how he is being more conservative this time in his criteria for rejection?

If you can't see the difference perhaps it will become more obvious as we continue.

To review from our earlier discussion, since he is using $\alpha = .01$, the region of rejection on the Sampling Distribution will correspond to the area that is only 1% probable in relation to the mean (Expected Value = 100). More specifically, it will correspond to two areas, one that accounts for .5% of the area above the mean, and .5% of the area below the mean (.5% + .5% = 1%).

Looking in the "area under the normal curve" table we find that the area <u>above</u> z = 2.58 is approximately .5% of the total (50.00% - 49.51% = .05), and likewise, the area <u>below</u> z = -2.58 is approximately .5% of the total. Thus, we can illustrate our region of rejection as follows:



A z = +2.58 corresponds to a raw score (actually a sample mean value) of 107.74:

$$X_1 = 2.58\,(\sigma) + \overline{X}$$

$$= 2.58(3) + 100$$

$$= 107.74$$

A z = -2.58 corresponds to a raw score of 92.26

$$X_2 = -2.58(3) + 100$$

$$= 92.26$$

Both of these values are shown on the diagram.

Now, suppose the researcher goes out and tests his sample of 25 juveniles and finds that $\overline{X} = 93$. What is his decision?

He decides to <u>retain</u> the null; $\overline{X} = 93$ falls within the region of acceptance. It is <u>more than</u> 1% probable (assuming the null were true), and thus does not constitute sufficient support for rejection.

Suppose that the mean IQ of the juvenile population is, in actuality, less than 100. What type of error would the researcher be committing?

ans. A Type II error: failure to reject a false null hypothesis.

Suppose he had used the .05 level of significance. Would his sample mean, $\overline{X} = 93$, be used as evidence for the same decision--retention of the null?

ans. No. Had he used the .05 level, $\overline{X} = 93$ would fall in the region of rejection (see diagram one page back) since it is less than $z = -1.96$ ($\overline{X} = 94.12$). Thus, in this particular instance, using the .01 level would lead to a different conclusion than if the .05 level were used.

The .01 level is more conservative than the .05 level. You are more likely to reject the null with the latter level of significance. What if he made a less conventional choice, such as $\alpha = .20$?

With the .20 level he is being even less conservative. Now he will reject the null if his sample mean is determined to be 20% probable. If you bother to compute what $\alpha = .20$ means in terms of z score values and real score values (in this example study) you would find that:

Region of rejection: $z = 1.28$; $z = -1.28$

: $\overline{X} = 103.84$; $\overline{X} = 96.16$

Thus, with $\alpha = .20$, he would reject the null if his sample mean was greater or equal to 103.84 or less than or equal to 96.16. Obviously, the criteria for rejection in this case are much less stringent than if .05 or .01 were used.

Basic rule: The greater your level of significance (.001 < .01 < .05 < .10, etc.) the less stringent (the less conservative) your criteria for rejecting the null hypothesis. Think about it!

<u>Relation between significance level (α) and Type I and II errors</u>

If you can apply the <u>basic rule</u> stated above to the previous information concerning Type I and II errors, the relation between α and these errors should become apparent.

The basic rule suggests that as α increases, your likelihood of rejecting the null also increases.

Therefore, as α increases, the likelihood of committing a Type I error increases. Why? If you are being less stringent in your criteria for rejection, you will be more likely to make a mistake by rejecting a null that is actually true. Samples fluctuate, right? Frequently you are going to collect sample data which deviate from the population mean, either overestimating it or underestimating it. If you use a large α (let's say .10 or .20) you are going to regard a fairly large percentage (10% or 20%) of possible sample means as "improbable" even though the null hypothesis might be true. Thus, a large α means more probability of rejection, and more susceptibility to the error (Type I) of rejecting null hypotheses when they are true.

Using the same reasoning, but in reverse, we can say that as α <u>increases</u>, the likelihood of committing a Type II error <u>decreases</u>. Why? If you are being less stringent in your criteria for rejection (by increasing α), you are going to reject more null hypotheses. Thus, there will be less likelihood that you will fail to reject a null that happens to be false (Type II error).

In summary, the relation between the choice of α and susceptibility to Type I and II errors can be described as one of "give and take." If you use a very conservative (low or small) level of significance (such as .01), you run relatively little risk of making a Type I error, but a high risk of making a Type II error. On the other hand, if you use a fairly liberal level of significance (such as .10), you run a high risk of making a Type I error, but a small risk of making a Type II error. Either way you go, you win and you lose; the choice is up to you and the people who sponsored the experiment.

<u>Unit 8.3.2 Finale</u>

The last point to be made in this particular instructional unit concerns an interesting aspect of the relationship between α and the chance of making a Type I error.

The interesting aspect is this: The chance of committing a Type I error is equal to α.

What does this mean? In "question and answer" form it means this:

Question: A researcher uses α = .05 in testing the null hypothesis. What are his chances of committing a Type I error?

Answer: 5% (If you use the .05 level of significance, 5 times out of 100 you will reject a true null hypothesis.)

Question: What if he used α = .01?

Answer: He would run a 1% chance of committing a Type I error (1/100 times he would end up rejecting a true null hypothesis).

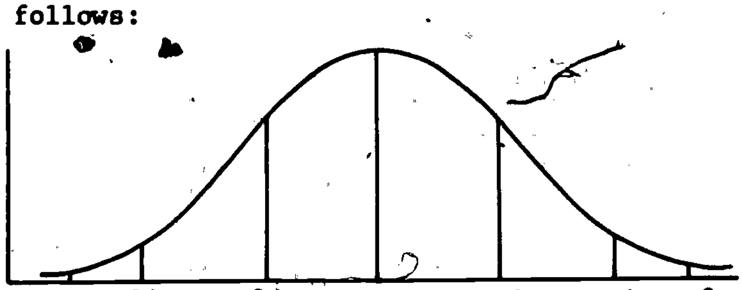
Question: What if he used α = .20?

Answer: He would run a 20% risk of making a Type I error (20/100 times he would end up rejecting a true null hypothesis).

That should be simple enough to memorize, but it will require a little more mental effort to understand why this relationship holds.

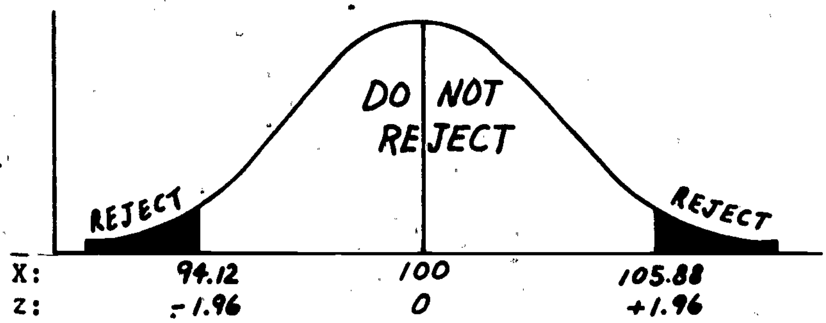
To use our continuing example - Suppose our researcher is trying to determine whether the average math scores of School XYZ students are equal to 100. He knows that those scores are normally distributed in the U.S. population with $\mu_X = 100$ and $\sigma = 15$. He decides to test a sample of 25 students from School XYZ since he simply does not have the resources to measure the whole XYZ population. (Being a conventional person, he specifies his level of significance as $\alpha = .05$.)

The hypothesis testing procedure dictates determining what the Sampling Distribution of means (size 25) from the XYZ population would look like if the null were true (that is, $\mu_{XYZ} = 100$). This is simple enough: $E(\bar{X}) = 100$ and $\sigma_{\bar{X}} = 3$. A sketch can be made of the Sampling Distribution as follows:



$\bar{X}$:	91	94	97	100	103	106	109
z:	-3	-2	-1	0	+1	+2	+3

Since $\alpha = .05$, we know the region of rejection will correspond to the areas beyond $z = +1.96$ and below $z = -1.96$ (this will always be the case whenever the .05 level of significance is used). We can "resketch" the Sampling Distribution to show these areas of rejection.



In order to see why the probability of making a Type I error is = to α, assume that in this particular instance the null hypothesis is actually true; μ_{XYZ} does = 100, and if the researcher had the time to test every single student, he would have found this to be the case. But, since he is restricted to using a sample of 25, there is no guarantee, due to sampling fluctuation, that his sample data will be reflective of this true parameter value (μ_{XYZ} = 100).

Now ask yourself the following question: Given the fact that μ_{XYZ} does = 100 (a fact unknown to the researcher), and the fact that the researcher has chosen the .05 level to test significance, what are the chances that his sample data ($\bar{X}$) will fall within the region of rejection, and thus cause him to reject a true null hypothesis?

A long question, true enough, but the correct answer is fairly brief: The researcher has a 5% chance of rejecting the true null. Why? The explanation can be found within the Sampling Distribution shown above. Doesn't that Sampling Distribution represent the probabilities of obtaining sample means of different values, based on samples of 25 cases, which are selected from a population where μ_X = 100 and σ = 15? Another long question, but one that merely states a definition you should have mastered long ago in Unit VII. Given that definition, what are the chances of obtaining a sample mean that will be greater than 105.88 (z = +1.96) or less than 94.12 (z = -1.96)? Again, the answer is 5%; 5/100 samples of 25 cases will yield means that fall into one of those regions.

Thus, whenever the null is true, the probability of rejecting it and thus committing a Type I error will be equal to the α level. If the researcher had used the .01 level, the chances of error would be 1% and so on. Think about this in reference to the Sampling Distribution. Try to understand why this occurs before resorting to sheer memorization.

Final notes

1. A Type I error refers to the rejection of a true null hypothesis.
   A Type II error refers to the retention of a false null hypothesis.

2. Probability or significance level, symbolized as α (alpha), is used in research to specify how "improbable" a sample mean must be in relation to the parameter value in question ($\mu_{\bar{X}}$) in order to justify rejection of the null. Conventional selections for α are the .05 and .01 levels.

3. The greater the probability level, the greater the chances of rejecting the null, and the greater the chances of committing a Type I error.

   The smaller the probability level, the smaller the chances of rejecting the null, and the greater the chances of making a Type II error.

4. The chances of making a Type I error are actually directly defined by α. If $\alpha = .05$, the chances for a Type I error are 5%; for $\alpha = .01$, the chances for a Type I error are 1%, and so on. *The chances of making a Type II error, although related to α, also depend on a number of other factors, and therefore cannot be determined as readily and directly.

8.3 POSTTEST (answers in back)

1. Define and differentiate between Type I and Type II errors. ________________________________

   ____

   ____

2. Discuss the rationale for specifying a level of significance (α) prior to the conduct of your experiment or study.

   ____

   ____

   ____

   ____

3. Describe the relation between the choice of α and the chances of committing Type I and Type II errors.

4. A researcher uses $\alpha = .05$ to test the hypothesis that children born in November are taller than average.

   a. What is the chance the he will commit a Type I error?
   b. What is the chance that he will commit a Type II error?

   (Note: If you feel that a question cannot be answered, use the notation "NA.")

   The researcher, after measuring a sample of 46 children born in November, decides to reject the null on the basis of the sample data.

   c. Which type of error might he be committing?

   If, before selecting his random sample, he decided to use $\alpha = .10$ instead of $\alpha = .05$, what would be the effect on his chances for:

   d. a Type I error?
   e. a Type II error?

8.4 Instructional Unit: Testing hypotheses when population parameters are known: the z Test

You probably will be pleased to hear that once again, you will be dealing with a "new" unit that really does not involve anything "new," provided that Unit 7.1.2 has been understood (to some extent). In the next few pages will be described, mostly through examples, the exact procedures to be used for testing hypotheses when population parameters (μ_x and σ) are known (as has been the case in all examples given thus far in this unit). For the most part, this will be review, with only very slight changes from what was discussed in Unit 8.3.

THE FOLLOWING TEST IS USED FOR COMPARING A SAMPLE MEAN TO A PARAMETER VALUE.

Steps for hypothesis testing (using the z test)

1. State your hypotheses.

   For example: H_0: $u_X = 79$

   H_a: $u_X \neq 79$

   Note: In this test we are testing the hypothesis that u_X = the parameter value, 79.

2. Select a level of significance.

   For example: $\alpha = .05$

3. Given your level of significance, determine the region of rejection in terms of z score (you may actually sketch the appropriate Sampling Distribution, but, in actuality, such extra effort is not really needed).

   For example: For $\alpha = .05$: region of rejection $> +1.96z$ and $< -1.96z$

   (By the way, this is always the case. Whenever the .05 level is used the "critical" z's will be +1.96 and -1.96. Why? Simply because the areas beyond these two scores will always account for 5% of the area in a normal distribution. If you are disbelieving, check the "area under the normal curve" table.)

   For example: For $\alpha = .01$: region of rejection $\geq +2.58z$ and $\leq -2.58z$.
   (This too is always the case when the .01 level is used. Why? Because the areas beyond these two z scores will always define 1% of the area in a normal distribution. Check the "area under the normal curve" table and discover this basic truth.)

4. Go out and collect your sample data.

5. Using the z test, determine whether your sample mean falls within the region of rejection or the region of acceptance.

   The z test may sound new, but it should be familiar. All that it involves is finding the z value of your sample mean in relation to the parameters of the Sampling Distribution, $E(\overline{X})$ and $\sigma_{\overline{X}}$:

   $$z = \frac{\overline{X}(\text{sample mean}) - E(\overline{X})}{\sigma_{\overline{X}}}$$

   Then compare this z value to the critical z value you determined in step #3.

If the z score for the sample falls within either region of rejection, you will reject the null. If it does not, you will retain the null.

Having completed these steps (which do not need to be memorized -- just understood), you have done all that is required for this test. Let's try some examples:

<u>Example #1</u> (this one will involve a test using pieces of rope, rather than people, as subjects in the study)

A company manufactures rope whose breaking strength averages 300 lb ($u_x = 300$) with a standard deviation of 24 lb ($\sigma = 24$). It is believed that by a newly developed process the mean breaking strength can be increased. A researcher is hired to investigate this hypothesis. Unfortunately, he does not have the time to measure individually each of the 10,000,000 pieces of rope manufactured by the new process, so he decides to use a random sample of 64 pieces.

He conducts his hypothesis test as follows:

1. States the hypotheses

H_0: $u_x = 300$ (Note: We are testing the hypothesis
H_a: $u_x \neq 300$ that H_0 = the parameter value, 300.)

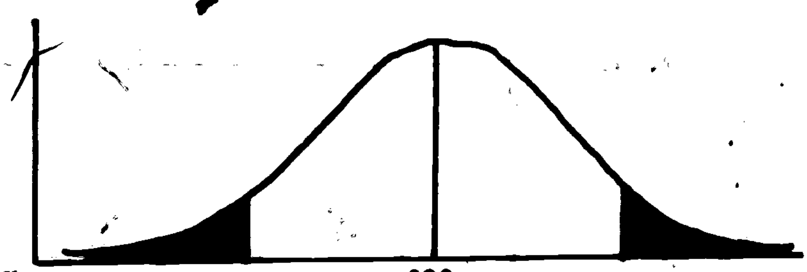
2. Selects level of significance

$\alpha = .05$

3. Determines the region of rejection

For the .05 level: region is greater than $+1.96z$ and less than $-1.96z$

A sketch of this, although not needed in reality, would look like:



X: 300
z: -1.96 0 +1.96

For Sampling Distribution of means, based on 64 cases:

$E(\bar{X}) = \underline{300}$

$\sigma_{\bar{X}} = 24 / \sqrt{64} = \underline{3}$

4. Collect your sample data

The researcher tests the 64 pieces of rope selected for his sample and finds that their average breaking strength, $\bar{X}$, = $\underline{306}$

5. Use the z test to determine whether sample mean falls within region of rejection

$$z = \frac{306 - 300}{3} = 2.00$$

Since the z score associated with $\bar{X} = 306$ is greater than the critical value, $z = 1.96$, the null hypothesis is rejected.

The researcher concludes that the new method produces significantly stronger rope than the original method.

QUESTION: What type of error could have been committed?

ANSWER: A Type I error; rejecting a true null hypothesis

QUESTION: What is the probability that a Type I error was committed?

ANSWER: 5% (.05), since the chances of a Type I equal α.

QUESTION: What would the procedures have been if the .01 level had been used?

ANSWER: The procedures would have been identical with the only change being the region of rejection. When the .01 level is used, the region of rejection involves the areas beyond 2.58z and -2.58z.

The decision, using $\alpha = .01$?

Retain the null: Our computed z value, 2.00, does not fall within the region of rejection for the .01 level.

What type of error is possible?

Type II in this case; failure to reject a false null.

The chances of that error?

Impossible to compute without much more information; only Type I error probabilities can be determined directly.

O. K., then, what are the chances of a Type I in this case?

There is no chance; you cannot make a Type I error when you retain the null.

You should be able to see that it is possible to reach opposite conclusions in the same exact experiment, depending upon which level of significance is used. That is why the researcher should select α before collecting his data (i.e., doing the actual study or experiment).

Last Example (This one will be done much more quickly)

A hospital administrator wants to determine whether the yearly incomes of orderlies in his state compare favorably with the national average, u_X = \$12,000 with σ = \$1,000. He decides to use a random sample of 25 orderlies selected from different hospitals in his state. He also decides to use α = .01 for his significance test.

His hypotheses: H_0: $u_X = 12,000$

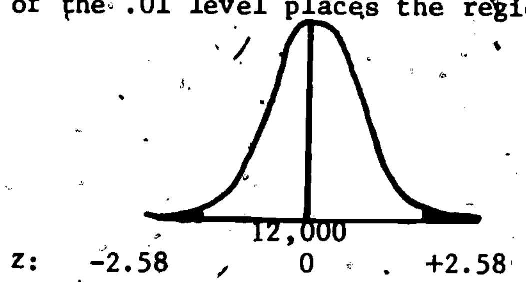
H_a: $u_X \neq 12,000$

The Sampling Distribution of means, based on 25 cases, will have:

$$E(\overline{X}) = 12,000$$

$$\sigma_{\overline{X}} = 1,000/\sqrt{25} = 200$$

Use of the .01 level places the region beyond $\pm$ 2.58z



He goes out and tests his sample. He finds that $\overline{X}$ = \$11,200 for this particular group.

By use of the z test, we determine how this value relates to the parameter (u_X = 8,000) Sampling Distribution of means.

$$z = \frac{11,200 - 12,000}{200} = \frac{-800}{200} = -4.00$$

With this calculation, it is clear that $\bar{X}$ is well within the region of rejection. The administrator is forced to conclude that hospital orderlies in his state are underpaid relative to the national average.

What type of error is possible here?

A Type I: rejection of a true null hypothesis. There is always the possibility that the average in his state is $12,000, but the sample he selected was, by chance, exceptionally low in income. There is always a chance for error whenever sample data, rather than population data, is used. The .01 level of significance sets the chance of a Type I error at 1%.

8.4 POSTTEST (answers in back)

1. The Scholastic Aptitude Test (SAT) yields a national mean of 500 and a standard deviation of 100 for high school seniors. The guidance counselor at Happy Valley High School selects a random sample of 100 seniors to determine whether the school SAT mean is comparable to the national norm. He finds that the mean SAT score of his sample is 482.

   a. State the null hypothesis
   b. What is $\sigma_{\bar{X}}$
   c. What is the z equivalent of the Happy Valley sample mean of 482?
   d. If α = .05, what should the conclusion of this study be?
   e. In drawing the above conclusion, what type of error is possible?
   f. Suppose the mean score of the 100 students was 520. Using the .05 level, what would you conclude?
   g. In drawing the conclusion in f., what type of error do you risk?

2. A manufacturer of flashlight batteries reports that his batteries have a mean life of 25 days with a standard deviation of 5 days. A consumer testing organization purchases a random sample

of 100 batteries and has them used. The mean life of the sample batteries proves to be 23.5 days ($\bar{X}$ = 23.5).

a. State the null hypothesis
b. What is $\sigma_{\bar{X}}$?
c. What is the z equivalent of $\bar{X}$?
d. What is the decision if α = .05?
e. In making the decision in d., what type of error is possible?
f. What is the decision if the significance level is .01?
g. Suppose that a sample of 25 batteries was used. If α = .05, what is the decision?
h. In making the decision in g., what type of error do you risk?
i. BONUS: How is the probability of a Type I error influenced by reducing the sample size from 100 to 25? What about a Type II error?

8.5 Instructional Unit: Testing hypotheses when population parameters are unknown: the t test

In this unit, we will cover briefly another procedure for hypothesis testing. If you have a superhuman memory, you might recall that in every example problem covered in the last few units, population parameters (μ_X and σ) were provided and known to the researcher. For example, if the researcher wanted to compare the IQs of adolescents in Juvenile Hall to the U.S. average, he knew that IQ in the U.S. had μ_X = 100 and σ = 15; if he wanted to test the strength of a new type of rope with the standard type of rope, he knew that the standard population had μ_X = 300 lb. and σ = 24 lb.; if he wanted to determine whether hospital orderlies in his state earned more money than average, he knew that for the whole U.S. population of orderlies, μ_X = \$12,000 and σ = \$1,000. Check these out and you will find that in every single example, population parameters were known. When they are known, the z test described in Unit 8.4 is perfectly appropriate for use in hypothesis testing.

Unfortunately, in most instances, population parameters will not be known. A teacher, for example, may wish to test the hypothesis that the average IQ in her school is 100, but be unaware of the population mean and standard deviation. Can she go ahead and test her sample by use of the z test? The rules dictate that she may not, and if you check the z formula you will realize that its application in this case would be impossible. In order to use z, you must know the $\sigma_{\bar{X}}$ for the Sampling Distribution; in order to know the value of $\sigma_{\bar{X}}$, you must be able to divide σ (the population standard deviation) by the square root of n. Does the teacher know the value of σ? According to the above description, she does not; as far as the z formula is concerned, she is finished before she starts.

Basic Rule: When population parameters are unknown, the z test is inappropriate. Test your hypotheses by use of the t test.

By now you are probably mumbling nasty words to yourself given the thought of being required to learn a whole new procedure. Chin up -- the t test is almost exactly the same as the z test with two very slight changes:

(1) You will have to estimate σ since it is not provided. This is done very simply by computing s for your sample (see Unit III), and using <u>it</u> as the population standard deviation. Actually, given nothing else, it is the best estimate available.

"Will I have to go through all the work to compute s?" you may ask with fear and trepidation.

No. In this unit, s will be provided. In real life it will not; then you will have to compute s using the procedures covered in Unit III.

The formula for t, as compared to z, is as follows:

$$z = \frac{\bar{X} - E(\bar{X})}{\sigma_{\bar{X}}} \qquad \text{where } \sigma_{\bar{X}} = \sigma/\sqrt{n}$$

$$t = \frac{\bar{X} - E(\bar{X})}{s_{\bar{X}}} \qquad \text{where } s_{\bar{X}} = s/\sqrt{n}$$

(2) The second minor change is that you will have to use a t table, instead of the "area under the normal curve" table, to find the critical values (i.e., the cut-off points indicating whether the null should be retained or rejected).

Find the t table, Table II, in the Appendix. Note that the t table is partially arranged in terms of columns corresponding to different probability values (levels of significance). If α = .01, use the .01 column; if α = .05, use the .05 column.

The rows on a t table correspond to what are called degrees of freedom (df), the rationale for which will not be given here. No need to get excited: $df = n - 1$. Thus, if there are 25 subjects in your sample, you find the row corresponding to df = 24; with 20 subjects, look for df = 19; with 60 subjects, look for df = 59, and so on. If a df is not tabled, use the next available lower df value.

Some Examples: Find the critical t value for a study in which n = 4 and α = .01.

Ans: t = 5.84 (df = 3)

In the same study, α = .05.

Ans: t = 3.18

What about when n = 24, and significance level = .01?

Ans: t = 2.81 (df = 23)

If you cannot find the above values on the t table, seek help immediately. Keep in mind that these values will be used to determine the region of rejection, just like in the case of z values. If the t values you compute exceeds the critical t value, you will reject the null, and so on.

O.K., now gather your senses and remember:

1) If population parameters are not given, use the t formula rather than z formula for comparing a sample mean to a parameter value.

2) Using t is almost exactly the same as using z except:

a) the sample standard deviation, s, is used as the denominator in the equation (see formula), and a t value is computed as in the z formula.

b) Instead of using the "area under the normal curve" table, you use the table to find the critical value, against which the computed t score (step a) is compared.

Example Problem: A teacher wishes to determine whether the average IQ in her class of 300 students is above 100. She doesn't have the time to test all 300, so she decides to use a sample of 25 (n = 25). A researcher recommends that she use the .01 level of significance in her hypothesis test.

Question: Which test should she use?

ans. The t test because the population u_{χ} and standard deviation are not available.

Question: What are the chances she will make a Type I error?

ans. 1% because she is using the .01 level; the t test does not change this basic rule.

She states her hypotheses as follows:

H_0: $u_{\chi} = 100$

H_a: $u_{\chi} \neq 100$

(this is always the same whether you use t or z)

She determines the region of rejection using the t table:

$t = 2.80$ (when $n = 25$, $df = 24$)

Thus, if her computed t value is greater than or equal to $t = +2.80$, or is less than or equal to $t = -2.80$, she will reject the null. The areas equal to and beyond $+2.80t$ and $-2.80t$ constitute the region of rejection.

She now goes out and tests her sample of 25, keeping in mind that since σ is unavailable, she will need to compute s (the standard deviation) of the sample data.

ERIC

For the sample tested, $\overline{X} = 104$, and $s = 10$.

Using the t formula:

$$t = \frac{\overline{X} - E(\overline{X})}{s_{\overline{X}}} \quad \text{where} \quad s_{\overline{X}} = s/\sqrt{n} = 10/\sqrt{25} = 2$$

$$= \frac{104 - 100}{2} = 4/2 = 2.00$$

Does she retain or reject?

She must retain since the computed t value, +2.00, does not exceed the critical t value of +2.80. Her sample data falls into the region of acceptance.

What type of error is possible?

ans. Type II

Last Example

A lightbulb manufacturer reads in Consumer Reports that a good bulb should last for 52 hours. He decides to test his bulbs against this recommended standard using a sample of 16 bulbs. Being conservative, he decides to use the .01 level of significance for his hypothesis test.

He states his hypotheses as follows:

H_0: $u_X = 52$

H_a: $u_X \neq 52$

The t test must be used since the population standard deviation, σ, is unavailable.

The critical value for t when $df = 15$ $(n-1)$ and $\alpha = .01$ is 2.95. Thus the area of rejection is that equal to or beyond $+2.95t$ and $-2.95t$.

He tests his sample of 16 bulbs and finds $\overline{X} = 42$ and $s = 12$.

What is his decision?

The decision must rest upon the computed t value:

$$\frac{42 - 52}{12/\sqrt{16}} = \frac{-10}{12/4} = \frac{-10}{3} = -3.33$$

The computed t value falls beyond the critical value, $t =$ -2.95 (or +2.95).

The decision must be to reject the null; the lightbulbs manufactured by this company do not meet the standard of 52 hours of life.

What type of error is possible?

ans. Type I; anytime you reject the null there is some chance of committing a Type I error -- rejecting a "true" null hypothesis.

Notes:

1. Use the t test when population parameters are unknown.

2. The t test is almost identical to the z test except:

   a) You will need to compute s (standard deviation) for the sample data. (On example problems, however, s will be given).

   b) Use s as an estimate of σ in the denominator of the t equation.

   c) Compute your t value the same way as for z. The critical value against which your computed value must be compared is to be found in the t table, by finding the column corresponding to your level of significance, and the row corresponding to your df ($df = n-1$).

8.5 POSTTEST (answers in back)

1. Find the critical t values when:

   a. $n = 23$ and $\alpha = .01$ b. $n = 23$ and $\alpha = .05$

   c. $n = 4$ and $\alpha = .02$ d. $n = 61$ and $\alpha = .001$

   e. $n = 11$ and $\alpha = .05$

2. Looking at the t table, what is the relationship between the critical t and:

   a. n?
   b. α?

3. A school teacher reads in the newspaper that the average second grader should be able to run the 40 yard dash in 15 seconds. She decides to test her own second graders by timing a sample of 9 children.

   a. Which hypothesis-testing procedure should be used, z or t? Why?

   The teacher is told by the school psychologist to use $\alpha = .05$.

   b. What are her chances of making a Type I error? a Type II error?

   c. What critical value defines the region of rejection?

   She goes out and tests her sample, finding $\bar{X}$ = 12.5 and s = 3.

   d. What is her decision regarding the null?

   e. What type of error is she in danger of committing?

   Suppose she had used the .01 level instead of the .05 level of significance.

   f. What would her decision be?

   g. What type of error would she be in danger of committing?

Unit VIII Review Test
(answers in back)

1. A researcher does a study to test the null hypothesis, $u_X = 197$. He obtains $\bar{X} = 230$ which is sufficient to reject the null. Can he conclude that the population mean in question is definitely higher than 197? Explain.

____

____

____

____ 2. What type of error (I or II) might have been made in the above example?

____ 3. Based on the decision in #1, which one(s) of the following alternative hypotheses would be supported?
a. $H_a = 197$ c. $H_a \neq 197$ e. $H_a < 197$
b. $H_a = 230$ d. $H_a > 197$

____ 4. Which of the following α would best protect against a Type II error?
a. .01 b. .10 c. .05 d. .001

____ 5. Which of the above alternatives would best protect against a Type I error?

6. Describe the logic under which hypothesis testing works (Hint: analogy to law).

____

____

____

____ 7. A researcher fails to reject the null using α = .05. He reasons "Had I selected .01 as the level I would have rejected it--too bad!" Does his reasoning make sense?

8. The average daily temperature in July in Newtown (at 12:00 noon) is 80 degrees, with σ = 8--based on town records. Farmer Jones thinks it's getting warmer over the years, so he measures temperature on 100 randomly selected July days in a 4-year period. He obtains $\bar{X} = 86$.

a. What is Farmer Jones' null hypothesis?

____

b. If he decides to test it at $\alpha = .05$, what is the rejection region (i.e., critical values)?

____

c. What is the computed z value? ______________

d. What is the decision? ______________

e. What type of error is possible? ______________

9. Suppose you wish to test $H_0 = 62$ at $\alpha = .01$, using a sample of $n = 25$. Sample statistics are $\bar{X} = 68$ and $s = 14$.

a. Why can't the z test be used? ______________

____

b. What is the computed t value? ______________

c. What is the tabled t value? ______________

d. What is the decision regarding the null? ______________

____

ERIC

UNIT IX

SELECTED INFERENTIAL TECHNIQUES

A. General Objectives

This unit extends the previous one by presenting other (more commonly used) techniques for making statistical inferences. The first two to be presented allow us to test hypotheses regarding the difference between two population means, asking, for example, "Does one population of individuals have higher scores than another (perhaps as a result of a special treatment that one received)?" The third technique extends our hypothesis testing procedure to the correlation coefficient through a popular test for judging its "significance" when computed from samples. The unit closes with a discussion of a somewhat different perspective on inference-making: interval, as opposed to point, estimation using confidence intervals. Be "confident" about your enjoyment of this material.

B. Specific Objectives

9.1 Test the Difference Between Two Means

9.1.1 The t test using independent samples
9.1.2 The t test using dependent samples

9.2 Test the Significance of the Correlation Coefficient

9.3 Use Confidence Intervals for Parameter Estimation

9.1 Instructional Unit: Testing the Differences Between Two Means

The two hypothesis testing procedures you have learned thus far both involve making inferences about a single population in terms of whether its mean conforms to some parameter value. Thus, what basically takes place is that a representative sample is selected and its mean is computed. The mean is then compared to the hypothesized number value; e.g. $H_0 : \mu = 100$, or $H_0 : \mu = 569$, etc. If σ^2 is known for the population, the z statistic is used for the test; if σ^2 is not known, the t statistic is used.

But how many times will the researcher have an exact comparison value in mind? Very rarely will that be the case and typically only when he is dealing with well-known measures for which standards (or norms) have been derived. For example, if we are testing students in SAT scores we can hypothesize that they will do better or worse than the national average which we know to be approximately 500. Similarly, we may wish to compare the mean for a selected group of individuals to known standards for such variables as height, weight, ACT scores, or body temperatures, etc. In such instances, we can safely turn to the z or the t procedures previously examined.

ERIC

Most hypothesis tests, however, involve comparisons between treatments that have never been used before in the exact same fashion. For example, one may wish to compare the effects on achievement in a math class of lecture instruction vs. programmed instruction. There would be no national norms telling us what the achievement means should be using those procedures, with those types of students, in that particular class, on that particular achievement test, etc. What we are really asking is, would the total population of students who receive lecture in that class perform differently than the total population who receive programmed instruction? If the answer is yes, then we have a basis for recommending whichever method came out better to those who teach similar courses. But can we test the total populations who would be eligible to receive those methods? Definitely not..., so as usual, we will use samples to make inferences about the populations. The following sections will illustrate two analytical designs for doing this: the t test for independent samples and the t test for dependent samples.

9.1.1 The t Test Using Independent Samples

In this first case, we wish to compare two means obtained from samples selected completely at random from their respective populations. Who makes it into Sample #1 has no influence on who is selected for Sample #2, and vice versa. We'll see in the next section (9.1.2) how a different design involving "matching" subjects across samples can be employed.

Let's illustrate the independent samples design using the example involving the comparison between lecture and programmed instruction. We randomly select a sample of subjects to represent the total population who might receive programmed instruction (PI). We do the same with regard to lecture. Who is picked for PI has no influence on who is picked for Lecture--it's the luck of the draw. Thus, one sample will be given lecture, and the other programmed instruction. Here is how the null and alternative hypotheses would be set up:

$$H_0 = u_1 - u_2 = 0$$

$$H_a = u_1 - u_2 \neq 0$$

No need to panic about these, as they indicate exactly what we would expect them to. The null (H_0) says that differences between the population that receives lecture instruction and the population that receives programmed instruction are equal to zero (no difference between them!). The alternative (H_a) says that there are differences, meaning that one population performs higher than the other. Based on the results we obtain from our samples, we will either retain the null or accept the alternative. The formula that is used to test these hypotheses (now brace yourself) is as follows:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

The numerator of this equation simply represents subtraction of one sample mean from the other. That part is easy.

The denominator looks quite a bit more complicated, but the mathematics involved, particularly with a calculator handy, is really quite simple:

s_1^2 is the unbiased variance estimate from sample #1. Remember the variance from Unit III? Unbiased means that $n-1$ is used in its derivation rather than n. Check back to Unit III if all this seems mysterious.

s_2^2 is the unbiased variance estimate from sample #2. Remember, we are using samples to make inferences about two populations. Thus, variance estimates are needed from both samples, rather than from only one as was the case the first time the t statistic was used.

n_1 & n_2 are the sizes of samples 1 and 2.

It would be beyond the scope of this text to try to explain how the above formula is derived (hope you're not too upset...). Perhaps it would suffice to say the following: If there was actually <u>no difference</u> between the effects of the two treatments, and samples of sizes n_1 and n_2 were selected from the populations receiving the treatments, the Expected Value of the difference between samples would be zero. But, note that just like in the case of the other hypothesis tests we have covered, some of the pairs of samples would show a difference - due to sampling fluctuation. When all difference scores (from an infinite number of sample pairs) are each divided by the error term shown in the denominator above, the resultant values will be distributed as t with $df = n_1 + n_2 - 2$. The t probability table for that particular df shows the likelihood of obtaining certain t values (under the assumption that the null is true). The logic hasn't changed from that underlying the two hypothesis tests discussed earlier. As an example, suppose that the sample size for the programmed instruction treatment is $n = 33$, while that for the lecture treatment is $n = 29$. To compute df for the hypothesis test, we apply the formula, $n_1 + n_2 - 2$, to get $33 + 29 - 2 = 60$. Looking at the t table, we find that if the null was true (i.e., there are no differences in effectiveness between programmed instruction and lecture for the "population"), the chances of getting a t ratio as extreme as ± 2.00 would be 5/100; for a t ratio as extreme as ± 2.66, they would be 1/100; and so on (make sure that you can see where these estimates come from on the table). So, once again we are asked to obtain a sufficiently "extreme" outcome in order to justify rejection of the null in favor of the conclusion that the treatments (probably) do differ in effectiveness.

Confused? That's very possible, as perhaps we've made a very simple procedure sound rather mysterious and complex. Let's try to clean up any misconceptions by considering a concrete example:

Example

A researcher believes that for 10th grade math students in Fly High, programmed instruction will result in better end-of-year test scores than will lecture instruction. Mrs. Smith's third period section, having an enrollment of 55 students, is randomly assigned to the two treatments; resulting in 27 students being taught by programmed instruction and 28 by lecture instruction. At the end of the year, the same examination is given to both groups to enable the comparison to be made between methods. The results show the following:

	n	$\overline{X}$	s^2
Programmed	27	90	26
Lecture	28	84	32

Before going on, make sure you understand what would be involved in putting the above table together. The derivation of n should be obvious, but if, as a result of some "memory problem," you have forgotten what $\overline{X}$ and s^2 are, make sure that you look them up (Units II and III, respectively); also remember that s^2 is derived by using $n - 1$ in the denominator of the variance formula - it is used to estimate σ^2.

O.K., back to the problem at hand. The results look suggestive, but can the researcher conclude that programmed instruction was significantly more effective than lecture instruction? First, of course, we need a significance level for the test. Sticking with convention, we'll say that the researcher is content to go with $\alpha = .05$. The next step is to compute a t score and determine whether it falls within or outside the ".05" rejection region for $df = 28 + 27 - 2 = \underline{53}$. We'll call the Programmed Instruction sample n_1 and the lecture sample n_2 (which labels are used for the two samples does not affect the workings of the formula). Substituting the obtained values in the t formula shown two pages back gives the following:

$$t = \frac{90 - 84}{\sqrt{\frac{(27-1)\ 26 + (28-1)\ 32}{27 + 28 - 2}\left(\frac{1}{27} + \frac{1}{28}\right)}}$$

$$= \frac{6}{\sqrt{\frac{(26)\ (26) + (27)\ (32)}{53}\left(\frac{1}{27} + \frac{1}{28}\right)}}$$

$$= \frac{6}{\sqrt{\frac{676 + 864}{53}\left(\frac{1}{27} + \frac{1}{28}\right)}}$$

$$= \frac{6}{\sqrt{29.06\left(\frac{1}{27} + \frac{1}{28}\right)}}$$

$$= \frac{6}{\sqrt{\frac{29.06}{27} + \frac{29.06}{28}}}$$

$$= \frac{6}{\sqrt{1.08 + 1.04}}$$

$$= \frac{6}{\sqrt{2.12}} = \frac{6}{1.46}$$

$$t = 4.11$$

Before you continue further, make sure that you know where the numbers that were used in the very first expression came from. Check them with the results of the study listed in the table on page 274 and with the original formula shown on page 272. The different expressions that appear above represent a step-by-step process of simplifying the original. When we are all done, we get $t = 4.11$ for the comparison of the means obtained in the two treatments.

The next step is to compare this value with the table listing for $\alpha = .05$, $df = 53$. The t table contained in the Appendix does not list any values for $df = 53$, so to be conservative (i.e., avoid elevating the chances of a Type I error above 5%), we use the next available lowest df, which on that table is 40. (If our computed t value was very close to the tabled listings, we could get a more accurate critical value by using a more detailed table or by interpolating.) Looking at the values for $df = 40$, we find that a t greater than +2.02 (or less than -2.02) is needed to reject at the .05 level of significance. The computed t therefore falls in the region of rejection. The researcher will conclude that programmed instruction yielded significantly higher test scores than did lecture instruction. What are the chances that a Type I error was made? Give yourself a "gold star" if you were thinking 5%. What are the chances that a Type II error was made? Give yourself another gold star if you were thinking that such a question is ridiculous. Type II's can only be made when the null is retained.

Here's another example:

Mrs. Goldfingers, a school nurse, is into the literature on health foods. She decides on the basis of her reading that having "Crunchy Cornflakes" for breakfast will give kids "go power" in their physical education classes. She convinces the principal to allow her to conduct an experiment using third-grade classes. The experiment involves serving "Crunchy Cornflakes" for breakfast to half of the students in each third grade class and the normal scrambled egg dish to the other half. The significance level is established at $\alpha = .01$. The students are then tested on the number of push-ups they can do in their physical education class that day. When the experiment is completed, the results obtained are as follows:

	n	$\overline{X}$	s^2
Cornflakes	50	10	9
Eggs	50	9	12

Can it be concluded that cornflakes were more effective than eggs?

$$t = \frac{10 - 9}{\sqrt{\frac{(50-1)\ 9 + (50-1)\ 12}{50 + 50 - 2}\left(\frac{1}{50} + \frac{1}{50}\right)}}$$

$$= \frac{1}{\sqrt{\frac{(49)\ 9 + (49)\ 12}{98}\left(\frac{2}{50}\right)}}$$

*Note, since n's are equal we can add the two multipliers in the denominator.

$$= \frac{1}{\sqrt{\frac{441 + 588}{98}\left(\frac{2}{50}\right)}}$$

$$= \frac{1}{\sqrt{\frac{1029}{98}\left(\frac{2}{50}\right)}}$$

$$= \frac{1}{\sqrt{\frac{2058}{4900}}} \quad \frac{1}{\sqrt{.42}} \quad \frac{1}{.65}$$

$$t = 1.54$$

Having +1.54 as the computed t value, what is the decision? To find degrees of freedom, we apply the simple formula 50 + 50 - 2 = 98. The t table has no listing for df = 98, so we must move down to the row associated with df = 60. Moving across the columns, we eventually reach the .01 significance level, which lists the critical t value as 2.660. The decision, then, is to retain the null. The experiment is inconclusive.

9.1.2 The t Test Using Dependent Samples

The type of experiment illustrated in the previous section involved comparing treatments by applying them and analyzing their outcomes using independent samples selected from the same overall population. For example, one group might receive programmed instruction and the other lecture instruction, but there would be no systematic relationship between the members of the two groups. In this section, we will discuss a slightly different type of experiment whose uniqueness lies in its matching of subjects across treatments. For instance, in the case of the programmed instruction vs. lecture experiment considered earlier, one might argue that an intervening variable such as IQ might be so powerful that any differences between the two treatments would be "washed-out." The specific concern might be that, by chance - even though assignment to treatments was done on a random basis, one group would be higher in IQ than the other. Regardless of treatment, then, that group might show the higher performances. To eliminate this possibility, the matched group design to be discussed here would systematically assign subjects to treatments, such that for each "high IQ" student assigned to Treatment A there would be a matched, high IQ student assigned to B, etc. There would be a number of different ways of doing this, but obviously, all would involve first measuring the prospective participants in terms of IQ.

The "dependent samples" design can be advantageous in many instances. The most important criterion affecting the desirability of its use is the relation of the "matching" variable to the outcome variable. With regard to the above example, it should be obvious that under normal circumstances IQ would relate significantly to the outcome variable in question, i.e., achievement on the lesson. Thus, IQ could be considered as a possible matching variable. But, would it do any good to match subjects on the basis of hair color, such that there would be a blond for a blond, etc.? Given that there would probably be no theoretical basis for predicting a relationship to exist between hair color (matching variable) and achievement (outcome variable), one would be doing a lot of extra work for nothing. Thus, the rule is that the stronger the relationship between the matching and outcome variables, the stronger the justification for using a dependent samples design. When the relationship is strong, the payoff is a more powerful test of treatment differences, which is the whole purpose of doing the experiment in the first place (i.e., comparing treatments or conditions). The cost, however, is additional work (it's much easier to assign subjects randomly to independent samples) and a cost of roughly half of the degrees of freedom (df) that would be present were independent samples to be used. The latter "cost" will become evident when you work with the formula to be provided later. But, if you'll look at the t table now, you will see that the greater the df for any α level, the lower the critical value that is needed to reject the null. Thus, a loss in df without any compensation, would hardly be to the experimenter's advantage. Before moving to the formula for the dependent samples design, test yourself on your understanding of the matching idea by identifying each of the following experiments as either justified or unjustified in terms of the probable relationship between its proposed matching and outcome variables (cover the answers as you go along).

Example A:

Treatment	Outcome Variable	Matching Variable
Watching films vs. Active Playing	Shooting ability in basketball	Height

Justified or not?

Ans: This dependent samples design would probably be justified. Height, in most cases, would have a reasonably strong relationship with shooting ability. Thus, a stronger hypothesis test would most likely result from equalizing the two treatment groups such that a tall person in one is matched to a tall person in the other, etc.

Example B:

Treatment	Outcome Variable	Matching Variable
Watching films vs. Active Playing	Knowledge of rules of basketball	Height

Justified or not?

Ans: Here the use of the "dependent" design would be questionable. Height should not show much of a relationship with knowledge of rules following the treatments. One would be going through some extra work and throwing away df, without earning much of a pay-off.

Example C:

Treatment	Outcome Variable	Matching Variable
Aspirin "A" vs. Placebo	Frequency of headaches	Amount of education

Justified or not?

Ans: Probably not. While one could perhaps argue that college grads in "white collar" positions might tend to have more headaches that high school grads in "blue collar" positions, it seems doubtful that such a relationship - assuming it really exists - would be very strong. Matching participants in terms of education would probably be wasted effort in this study, as random sampling would probably achieve a reasonable enough balance on the education dimension.

Example D:

Treatment	Outcome Variable	Matching Variable
Aspirin "A" vs. Placebo	Blood Pressure	Age

Justified or not?

Ans: Such matching would well be worth some consideration, as age should show a fairly consistent correlation with blood pressure. If, by chance, random sampling resulted in one group being older than the other, one could obtain differences in this experiment that do not reflect the actual effects of the aspirin vs. the placebo.

Example E:

Treatment	Outcome Variable	Matching Variable
Film watching vs. Reading	Score on a history exam	grade point average

Justified or not?

Ans: Such matching would seem justified, as grade point average would probably show fairly strong relationship with examination scores.

HAD ENOUGH? Let's move on to the formula for the dependent samples design. Here's how to compute t:

$$t = \frac{\bar{D}}{\sqrt{\frac{\Sigma d^2}{n(n-1)}}}$$

where $\Sigma d^2 = \Sigma D^2 - \frac{(\Sigma D)^2}{n}$

DON'T PANIC! This formula is actually fairly easy to work with once you know the meaning of "D." Here are some definitions:

D: is the difference between the scores of a subject in Treatment-1 and his/her matched counterpart in Treatment-2. This variable should become clearer in the example to be presented below, but for now imagine that John has been matched to Sally in IQ, and each receives a different instructional treatment. John obtains a score of 80 on the criterion measure (the final exam), while Sally obtains a score of 75.

For this particular matched pair, $D = 5$ (or -5, depending on which treatment is listed first).

ΣD^2: is the sum of the squared difference scores (D's) across all matched pairs.

$(\Sigma D)^2$: is the square of the sum of all difference scores (D's).

Σd^2: is needed for the error term; it is obtained by the bottom formula on page 280.

$\bar{D}$: is the average difference between matched pairs.

n: is the number of matched pairs.

If you have a good memory, you might recall that the formula for obtaining Σd^2 is operationally identical to that used for calculating the variance, as was shown in Unit III. No surprise; what we are doing here is the same thing, except we are getting a variance estimate for "difference scores," rather than for plain old X's.

So, at the risk of getting you confused with "words," use of the formula involves computing t by:

(a) determining the difference between the scores of members of each matched pair. The result is a series of D's, one for each pair.

(b) adding all of these D's and dividing by the total number of pairs to get an average. This is labeled as $\bar{D}$, and it goes in the numerator of the formula.

(c) manipulating the D's as shown in the second formula to get Σd^2.

ERIC

This value is then divided by $n(n-1)$, and the square root of the result then provides the denominator of the formula.

(d) dividing numerator by denominator gives us t.

(e) the computed t, as in all other tests we've covered, is then compared to the critical t for the appropriate α and df.

IN THE DEPENDENT SAMPLES ANALYSIS, $df = n - 1$, Where n = number of matched pairs.

Now for an example:

Example A

Let's work with the same experiment outlined in an earlier example. A basketball coach receives an official "NBA" training film that is supposed to teach "all one needs to know" about shooting techniques. He wonders whether showing the film will have much effect on the performance of his 9th grade students. He decides to do an experiment in which half of the 20 students in 11:00 section will watch the film while the other half will spend the same amount of time in "free play." Not wanting to take the chance that height could be a factor, he matches each student in the film condition to a student of similar height in the "free play" condition. During the next period, he gives the students a "free throw" test in which they are allowed 10 chances at the foul line. Results are as follows:

Pair #	Film: Treatment 1		Play: Treatment 2		D (X_1-X_2)	D^2
1	Rick	8	Bob	10	-2	4
2	Mike	8	Fred	8	0	0
3	Tom	4	Alex	8	-4	16
4	Robby	10	Mick	9	1	1
5	Steve	2	Carl	10	-8	64
6	Don	7	Roger	6	1	1
7	Tom A.	6	Bill	10	-4	16
8	Jim	7	Marc	9	-2	4
9	Buster	5	Fran	5	0	0
10	Pete	0	Ed	4	-4	16
					$\Sigma D = -22$	$\Sigma D^2 = 122$
					$\bar{D} = -2.2$	

Before continuing, make sure you understand that members of a given pair were matched in this experiment according to height. Each then received a different treatment, the score under which is shown adjacent to the name. Most important, make sure that you understand how the D's were derived, as this variable is the one required to get the analysis started.

First, we need to get the d^2 value that goes in the error term (denominator of the main equation). Substituting values in the Σd^2 equation, we get:

$$122 - \frac{(22)^2}{10} = 122 - 48.4 = 73.6$$

$$\Sigma d^2 = 73.6$$

Now we are ready for the main equation. The numerator $\bar{D}$ is the average difference between pair members. As computed above it equals -2.2.

$$t = \frac{-2.2}{\sqrt{\frac{73.6}{10(9)}}} = \frac{-2.2}{\sqrt{\frac{73.6}{90}}} =$$

$$\frac{-2.2}{\sqrt{.818}} = \frac{-2.2}{.904} = -2.43$$

If the significance level being applied was $\alpha = .05$, what would the decision be?

Remembering that df in the dependent samples design is equal to $n - 1$ (where n represents the number of pairs), we get $10 - 1 = 9$ degrees of freedom. The t table shows that the critical value for this df is 2.262. We, therefore, have a sufficient basis for rejecting the null, but because the computed t had a negative sign, we want to be sure to conclude that Treatment-2 ("Play") was favored over Treatment-1 ("Film"). (Note that the positioning of treatments was arbitrary - we could have made "Play" the first treatment just as easily, in which case the computed value would have had a "+" sign.)

If the significance level was $\alpha = .01$, what would the decision be?

RETAIN THE NULL, as the critical value for $df = 9$, $t = 3.25$, exceeds the computed value.

The possibility of reaching these opposite conclusions should reconfirm in your mind the importance of the selected α level to the interpretation of research findings.

Given the formula for computing t in a dependent samples analysis, there is really nothing much new to learn here. Thus, we'll go immediately to a practice posttest, which will test your skill at doing the analysis on your own.

9.1 POSTTEST (answers in back)

1. A new sedative is developed and tested in terms of its effects on reducing pulse rate. A random sample of 30 adults is administered the drug, whereas another random sample is administered a placebo. Pulse rate of the two groups is then analyzed. Results are

	n	$\bar{X}$	s^2
Drug	30	79	100
Placebo	30	91	95

_____ a. What is the computed t value?

_____ b. What is the critical t value for $\alpha = .01$?

_____ c. Should the null be rejected or retained?

_____ d. What type of error is possible?

2. A study is planned in which the effects of a new vitamin compound will be compared against those of a placebo on an incidental outcome--the growth of new facial hair. Given the nature of the treatment and outcome variables, which of the following would be the best choice as a matching variable for a dependent samples design? Why?

a. IQ
b. sex
c. height
d. physical fitness
e. popularity

3. Attitudes toward busing are analyzed as a function of whether adults live in a large city or in a small one. Given the nature of the treatment (city size) and outcome variable (attitudes), which of the following would be the best choice as a matching variable? Why?

a. height
b. sex
c. physical fitness
d. race
e. popularity

4. Analyze the data obtained in the following experiment. Students in a fourth-grade class are matched in IQ. One member of each matched pair receives written teacher comments ("Good going," "Nice work," "Try better next time," etc.) on his/her returned answer sheet from Arithmetic Test #1. The other member receives no comments. Comparisons between groups are then made with regard to performance on Test #2. The .05 level of significance is selected for the evaluation of results. Results are as follows:

	Test Scores	
Pair #	Student Receiving Comments	Student Receiving No Comments
1	80	75
2	98	90
3	75	77
4	83	77
5	100	88
6	63	65
7	79	60
8	94	93
9	90.	81
10	70	66

a. What is the computed t?
b. What is the critical t for $\alpha = .05$?
c. What is the decision?
d. What type of error is possible?

ERIC

9.2 Instructional Unit: Testing the significance of the correlation coefficient

In this section, we will illustrate how hypothesis testing can be applied to statistics other than means, $\bar{X}$'s. At the same time, you will learn about a very valuable and commonly used analytical technique--that for testing the significance of the correlation coefficient r.

Example Situation

Dr. Wynn Flynn, a noted clothing researcher, believes that there is a significant relationship between the length of people's toenails and the amount of wear they get from socks. To test this (using $1,000,000 grant from the American Toenail Association) he selects a random sample of 100 adult males. He measures the length of each subject's toenail (right foot, big toe) and the amount of wear obtained from a brand new pair of dress socks (number of days before hole appears). (Out of consideration for the subjects' friends and families, each subject is required to wash the socks each night.) The correlation obtained is $r = -.47$. Can you interpret this? Shame if you cannot and you're this far along in the book!

Dr. Flynn now wants to publish his findings to support the conclusion that longer toenails are associated with less wear from socks. Is that conclusion justified?

Well, the negative correlation obtained does support that trend, BUT any result is possible from a sample. How do we know that this result represents what would happen for an entire population of adult males possessing big toes and similar types of socks?

Answer: We don't! Just like in the case of means ($\bar{X}$'s), all we can do is make probablistic judgments based on the results obtained and statistical theory. We have a null hypothesis here too. It essentially says "no difference--no effect (no correlation!)" for the population. If the weight of evidence makes that hypothesis appear improbable, the conclusion is that "a difference or effect (a correlation!)" does seem likely for the population. Let's illustrate the hypothesis testing procedure, using Dr. Flynn's results as an example:

1. Hypotheses: $H_0 : \rho_{xy} = 0$

   $H_a : \rho_{xy} \neq 0$

The null establishes a condition in which there is no correlation (ρ) between X (toenail length) and Y (sock wear) for the population. If evidence favors "rejection" of the null as improbable, we are left with the alternative that there is a correlation.

2. Random sampling and data collection. These go without saying. Obtain your random sample, as Dr. Flynn did, and collect your data.

3. Compute statistic and test H_0. Since our null concerned correlation, we naturally want to compute the sample correlation r. Dr. Flynn's r was -.47. Is that a sufficient finding to view the null (which states no correlation) as improbable? Just like for our tests involving means, we need a formula:

$$t = \frac{r}{\sqrt{(1-r^2)/(n-2)}}$$

$$df = n - 2$$

Yes, when $\rho_{XY} = 0$, the above outcome is distributed as t. It thus provides a basis for concluding whether the r we obtained is improbable if the population correlation (ρ_{XY}) is actually zero. If so, we reject the null! Note that df for the distribution is n - 2(number of subjects-2). Let's work with Dr. Flynn's results, using .01 as our significance level:

$$t = \frac{-.47}{\sqrt{(1-(-.47)^2)/(100-2)}}$$

$$= \frac{-.47}{\sqrt{.779/98}} = \frac{-.47}{.089}$$

$$= -5.28$$

Now check the t table for df = 98 at α = .01.

4. Conclusion. The computed t value of 5.28 clearly exceeds the tabled one of about 2.66. The sample correlation appears "significant," meaning, "reject the null!" We can infer that a negative correlation between toenail size and sock wear probably exists for the population. Any chances of a Type I error? Yes--1% (assuming the null were actually true). Any chance of a Type II error? None! We rejected the null!

Another Example

Dr. Flynn suspects that there will be a significant negative correlation between head size and time taken to put on sweaters. He employs a sample of 25 women for this, and tests them using appropriate materials and measures. The resultant correlation is r = -.14.

Flynn's feeling pretty good about himself, since the result was just the way he called it--a negative correlation. His head is getting so big that the sweater he's wearing may never come off. But we know about sampling fluctuation. Maybe his result is quite reasonable as a product of sampling fluctuation. Maybe had he tested everyone, the correlation would have been zero! Let's do the test (using α = .05).

$$H_o : \rho_{xy} = 0$$

$$H_a : \rho_{xy} \neq 0$$

$$t = \frac{-.14}{\sqrt{(1-(-.14)^2)/(25-2)}} = \frac{-.14}{\sqrt{.98/23}}$$

$$= \frac{-.14}{.206} = -.678$$

With $df = 23$, the critical region for rejecting the null is defined (see t table) as $\pm$ 2.069 for α = .05. Our computed t value was only -.678. There is not sufficient evidence for rejecting the null. It could well be the case that were all women to be tested, head size would have no relationship to "sweater putting on" time. Flynn is foiled by his foolish forecast.

9.2 POSTTEST (answers in back)

1. A researcher correlates two variables using a sample of n = 30. The obtained r = .36. He wishes to test its significance (relative to $\rho = 0$) using α = .01.

   a. What are the null and alternative hypotheses?

   b. What is the tabled t value?

   c. What is the computed t value?

   d. What is the decision?

   e. What type of error (I or II) could have been made?

2. Answer the same questions as above (a-e), given r = .78, n = 16, α = .05.

303

9.3 Instructional Unit: Using confidence intervals for parameter estimation

The analytical techniques discussed in the two previous sections are used to test hypotheses regarding one or more populations. Example questions: Is Treatment A better than Treatment B (for all people receiving each)? Is there a positive correlation between X and Y for all people in Population A? Again, the formulas allow us to test specific hypotheses. But suppose that we have no hypotheses in mind and instead, simply want to estimate a population parameter that is unknown? Maybe you are a real estate salesman and want to estimate the square footage of a typical house in your community. Or, maybe you're a college instructor and want to know how much time a typical student devotes to studying. No hypotheses or no actual figures may exist for either of these concerns. But the notion of Sampling Distributions, Expected Values, and Standard Errors can still be applied to allow us to estimate the "answers."

Point versus Interval Estimation

A point estimate of a population parameter, let's say u_X, uses a single value as the basis for estimating the unknown. The real estate salesman in the above example simply selects a random sample of houses and measures their square footage. Let's pretend that he obtains $\overline{X} = 14{,}962$. That single value then becomes the basis for estimating u_X. Using the same logic the college professor surveys a random sample of students on the amount of time they devote to studying. He obtains $\overline{X} = 96.34$ min. a night, which becomes his estimate of u_X.

Point estimates are nice and convenient. But, based on your knowledge of sampling fluctuation, can you see any problem? Sure.... if you happen to get a weird sample, the impression you obtain (and maybe rely on) could be horribly misleading. This concern creates a rationale for a different type of estimation, involving intervals.

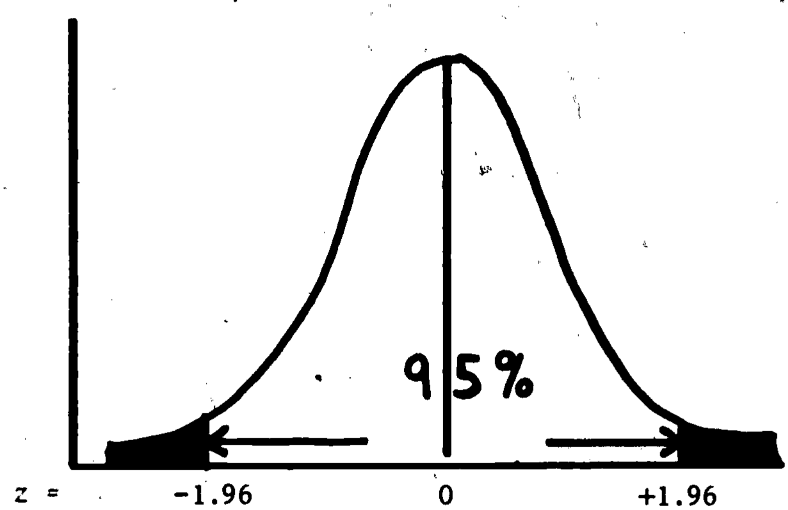
Interval estimation specifies a range of values that include the unknown population parameter a certain percent of the time. The conventional term statisticians use in referring to these ranges is confidence intervals. The level of "confidence" expresses the statistician's confidence that tht true parameter value is contained in the interval of values. For example, we might compute, using methods to be described below, the 95 percent confidence interval for the housing square footage u_X to be from 13,600 to 14,400. We can say "with 95% confidence" that the actual population mean lies somewhere within this range. Let's examine more closely how confidence intervals are constructed. For simplicity and clarity, we'll restrict our examples to instances involving means (confidence intervals, though, can be applied in the identical manner to other statistics as well).

Constructing Confidence Intervals

Let's suppose that scores on a standardized spelling test for second graders are normally distributed with $u_X = 250$ and $\sigma = 50$.

You are unaware of these figures and select a random sample of 100 scores to estimate u_X. Naturally, the basis for your estimate is $\overline{X}$, the mean of your sample. It is highly unlikely that this point estimate will be exactly equal to u_X, due to sampling fluctuation. But you could construct an interval around $\overline{X}$ that will have a certain probability (give you a certain degree of confidence) of containing u_X. A conventional probability or confidence level is .95. What interval will have a 95% chance of containing u_X?

Sampling theory provides the answer. Look below at the Sampling Distribution for $\overline{X}$'s computed from all unique samples of $n = 100$ drawn from that population. As reviewed in Unit VII, $E(\overline{X}) = u_X$, and $\sigma_{\overline{X}} = \sigma/\sqrt{n}$. What portion of the distribution would contain 95% of the samples? Answer: That between +1.96 and −1.96 standard errors (z's). How do we know? Normal curve probability of course! Ninety-five percent of scores always fall between $\pm$ 1.96z.



Now let's reason the same way, but (sort of) in reverse. Given that we have computed a single $\overline{X}$, what interval around it is 95% likely to contain u_X? Answer: $\overline{X} \pm 1.96\sigma_{\overline{X}}$. Think about it and try these questions:

1. What percentage of sample $\overline{X}$'s would be within $\pm$ $1.96\sigma_{\overline{X}}$ from u_X? Answer: 95% (look at the figure)

2. So, if you added $1.96\sigma_{\overline{X}}$ to each $\overline{X}$ to form an upper limit and then subtracted the same value to form a lower limit, what percentage of those intervals (one for each sample), would contain u_X?
Answer: 95%

3. Huh?

Answer: Look at the Sampling Distribution. Isn't it true that on the left side, only the 2.5% in the shaded area would never reach u_X with their upper limit? Isn't it also true that of those on the right side, only 2.5% would never reach u_X with their lower limit?

4. What, then, is the 95% confidence interval for $\overline{X}$?
Answer: $\overline{X} \pm 1.96\sigma_{\overline{X}}$

Thus, if we add and subtract $1.96\sigma_{\overline{X}}$ to $\overline{X}$, we can be 95% confident that the resultant interval will contain u_X. Or, to say it in a slightly different way, if we were to repeat the process for all sample $\overline{X}$'s, 95% of the confidence intervals would "capture" u_X, while 5% would not.

5. Would a 99% confidence interval be wider or narrower than a 95% one?
Answer: If you got this, you're really thinking! WIDER is the only choice! If you want to be more confident of capturing u_X, don't you have to cover more ground??

6. What would the 99% confidence interval be?
Answer: $\overline{X} \pm 2.58\sigma_{\overline{X}}$

7. Huh?
Answer: Isn't it true from normal curve probability that 99% of the scores fall within $\pm$ 2.58z ? By the same reasoning, if limits based on those dimensions were formed around the $\overline{X}$'s of all unique samples, 99% of the resultant intervals would contain u_X.

8. Can one give a general expression that would represent confidence intervals using normal curve probability (i.e., z values)?
Answer: Yes

9. Well, what is the expression?
Answer: $\overline{X} \pm (z_{conf.})\ \sigma_{\overline{X}}$, meaning (a) determine the z value corresponding to the particular level, multiply it by the standard error, add it to $\overline{X}$ for an upper limit and subtract it for a lower limit.

10. Suppose your computed sample $\overline{X}$ in the previous example (spelling scores) was 240. What is the 95% confidence interval? What is the 99% confidence interval?

Answer: For both levels $\sigma_{\overline{X}} = \sigma/\sqrt{n} = 50/10 = 5$

$$Conf_{.95} = 240 \pm (1.96)\ (5)$$

$$= 230.2 \text{ (lower limit) and } 249.8 \text{ (upper limit)}$$

We can be 95% confident that this interval contains u_X. Note, however, that this is one of the 5% that actually doesn't!! We obtained an extreme sample (u_X = 250, see page 289).

$$Conf_{.99} = 240 \pm (2.58)\ (5)$$

$$= 227.1 \text{ (lower limit) and } 252.9 \text{ (upper limit)}$$

We can be 99% confident that this interval contains u_X. (And it does. See how the larger, 99% interval, can be more accurate than the smaller, 95% interval? The trade-off, though, is that the 99% interval gives a less precise estimation.)

Using t as the Position Measure

In the above example, we had the luxury of trying to estimate u_X using known properties of the population it describes (that σ = 50). As was discussed in Unit VIII, when we introduced the t distribution, it is usually the case that σ won't be known. So all we can do is to estimate it using our sample standard deviation, s. But, when Sampling Distributions are constructed with s in the standard error formula, they conform to t rather than z probabilities. If you're not following all of this, no need to worry. Maybe by examining the process, the underlying theory will become a little clearer.

Suppose, for example, that you are interested in determining the average number of daily phone calls received in your office. Your boss supports you on this, but says that he will not permit you to spend time surveying the calls every single day. So, you decide to randomly select 36 days--the boss grants permission. You count the calls and compute your statistics. The results: $\overline{X}$ = 16.5 and s = 4.

Your interest naturally is with estimating the average for all days, of which the 36 examined is only a very small sample. Using point estimation, you would report 16.5 calls as typical--but imagine the room for error! Confidence intervals to the rescue...What is the 95% confidence interval for our estimate? Since we are using s rather than σ, the "general" expression for such an interval is

$$\overline{X} \pm (t_{conf})\ (s_{\overline{X}})$$

If you look back to the previous examples, you'll see that now we're asking for (a) the t score, not the z score, that corresponds to the probability level desired, and to (b) the standard error based on s, not σ.

Let's identify this mysterious confidence interval:

$\overline{X} = 26.5$

$s_{\overline{X}} = 4/\sqrt{36} = .667$

But what about t?

The t distribution varies based on df. Our df, in this case, is $36 - 1 = 35$. Looking at the t table, we find that $t_{.95 \text{ or } .05} = 2.042$ for the closest df. So, we get:

$$Conf_{.95} = 16.5 \pm (2.042)(.667)$$

$$= 16.5 \pm 1.36$$

$$= 15.14 \text{ (lower limit) and } 17.86 \text{ (upper limit)}$$

We can feel 95% confident that the actual mean falls within that interval.

What would the 99% confidence level be? The only change here is that we need the $t_{.99}$ value (same in table as $t_{.01}$), which is 2.75. Thus:

$$Conf_{.99} = 16.5 \pm (2.75)(.667)$$

$$= 16.5 \pm 1.83$$

$$= 14.67 \text{ (lower) and } 18.33 \text{ (upper)}$$

We can feel 99% confident that the actual mean lies within that interval.

Can you see the relation between confidence intervals and hypothesis testing through these examples? In hypothesis testing we compare our sample mean to some hypothesized value; using confidence intervals, we try to provide an estimation of the population mean using the information we obtain from the sample. Both procedures employ the identical information--$\overline{X}$, s, n, and t value to produce the desired outcome. For each hypothesis test we have presented, there is a corresponding procedure for constructing confidence intervals.

Take, for example, the case where we compare two sample means using independent samples (Unit 9.1.1). As introduced in this unit, the interest was with <u>testing the hypothesis</u> that the two populations represented did not differ from one another. On the other hand, we can simply be interested in <u>estimating</u> what the actual difference, if any, is. To avoid burdening ourselves with a whole new series of calculations, let's return to the example on p. 274. What difference really exists between using programmed instruction and lecture for that lesson? Our samples indicated a 6-point advantage for programmed instruction over lecture. So, we could give that as a <u>point estimation</u>--and hope that luck is with us. Or, we can be more cautious and construct a confidence interval around that value? What would the 95% confidence interval be?

The general formula remains the same, but there are some values we need to identify:

$$Conf. = \overline{X}_1 - \overline{X}_2 \pm (t_{conf})\, s_{\overline{X}_1 - \overline{X}_2}$$

$$\overline{X}_1 - \overline{X}_2 = 6$$

$$t_{conf} = t_{.95} \text{ for } df = 53 = 2.02$$

$s_{\overline{X}_1 - \overline{X}_2}$ denominator of the t formula (see p. 275) = 1.46

$$Conf_{.95} = 6 \pm (2.02)(1.46)$$

$$= 3.05 \text{ (lower) and } 8.95 \text{ (upper)}$$

We can be 95% confident that the "true" treatment difference lies between 3.05 and 8.95 points.

Summary

Using the procedures just illustrated, a confidence interval can be constructed around any of the statistics we have examined. All that one needs to know are the t (or z value) and standard error corresponding to the statistic's Sampling Distribution. If you really want to test your understanding of the methods, see if you can identify the 95% confidence interval for $\overline{D}$ obtained in Example A (p. 283). The answer is typed upside down at the bottom of this page.

Confidence intervals provide a way of estimating population parameters using a "probable" range of scores. The more confident you want to be about that range, the wider you must make it. Thus, a 99% confidence interval will be wider than a 95% or 90% interval. The confidence interval procedure contrasts with point estimation which uses a single value to represent the population. Though logically and functionally related to hypothesis testing, confidence intervals are mainly concerned with estimating not comparing. Try to "estimate" your understanding of these ideas by answering the questions below.

-4.24 (lower) and -.16 (upper)

9.3 POSTTEST (answers in back)

1. Differentiate between <u>point</u> estimation and <u>interval</u> estimation.

____

____

____

____ 2. A 95% confidence interval is constructed around a sample mean. Which of the following is true?

a. There is a 95% probability that the interval will contain u_X.
b. If such intervals were constructed around all sample means, 95% of them would contain u_X.
c. Such an interval will be smaller in size than if a 99% confidence interval was used.
d. All are true.

____ 3. Psychometrician Jones obtains a new "Leadership Abilities" test and wants to estimate what an average score would be for all people who might take it. After administering the test to a random sample of 36 people he obtains $\bar{X} = 68$ and $s = 12$. Report the:
a. 95% confidence interval around $\bar{X}$;
b. the 99% confidence interval; and
c. the point estimate of u_X

Unit IX Review Test
(answers in back)

A. Multiple-Choice

_____1. A researcher rejects the null, when in reality the null is true. He/she has made
a. a Type I error
b. a Type II error
c. both Type I and II errors
d. a Type IV supreme error

_____2. If you were doing an experiment for which you wanted to increase your chances of rejecting the null, which of the following procedures would be helpful?
a. increase n (i.e., test more subjects)
b. increase α
c. identify an intervening variable that strongly relates to the outcome variable, and use it for matching in a dependent samples design
d. all of the above would help
e. none of the above would help

_____3. A medical researcher is doing a study involving a comparison between two surgery techniques. Naturally, it would be very dangerous to conclude that a particular technique is better than another, if it actually is not. In selecting a significance level, the researcher would be mostly concerned with
a. minimizing the chances of a Type I error
b. minimizing the chances of a Type II error
c. maximizing the chances of a Type I error
d. maximizing the chances of a Type II error

_____4. A study is conducted in which subjects are assigned strictly at random to two treatments, and outcomes of the two treatments are then compared. The type of hypothesis test that would be appropriate here would be
a. z-test comparison with a numerical parameter value
b. t-test comparison with a numerical parameter value
c. t-test comparison for independent samples
d. t-test comparison for dependent samples

_____5. An experimenter wants to test the hypothesis that the GRE scores of a certain population of students are equivalent to the national average, which he knows to be 500 with $\sigma = 100$.

ERIC

Which one of the tests listed in the alternatives for item #4 would be the most appropriate?

____6. Fifty students are randomly assigned to Treatment A and 48 are randomly assigned to Treatment B. The appropriate df for a comparison of means would be

a. 48 d. 96
b. 50 e. 97 g. 100
c. 49 f. 47

____7. Fifty students in Treatment A are matched on some variable to 50 students in Treatment B. The appropriate df for a comparison between means would be

a. 48 d. 98
b. 49 e. 99 g. none of these
c. 50 f. 100

____8. An experimenter uses the dependent samples design, but selects a matching variable that has a zero correlation with the outcome variable in his experiment. Which of the following best describes the implications of what he has done?

a. even though a "better" matching variable could have been selected, he has still strengthened his analysis by using the dependent samples approach
b. aside from the extra effort involved in matching people, he will come out the same as if he had used the independent samples approach
c. he will lose df without achieving a stronger test; thus, the procedure would have been to his disadvantage
d. the chances of making a Type I error will increase

____9. In the dependent samples analysis, D refers to

a. the difference between the means of Treatments A & B
b. the standard deviation of a particular treatment
c. the variance of a particular treatment
d. the difference between the scores of members of a matched pair

____10. A study is performed in which the effects of drinking different quantities of alcohol are compared with respect to college seniors' ability to roller skate. If considering the possibility of using dependent samples, the most appropriate matching variable would be

a. height d. grade point average
b. age f. amount of prior skating experience
c. IQ

B. A sample of $n = 25$ is tested, with results indicating $\overline{X} = 525$ and $s = 112$. Compute the 99% confidence interval and the 95% confidence intervals around $\overline{X}$.

ERIC Full Text Provided by ERIC

UNIT X

THE CHI-SQUARE TEST

A. General Objectives

The statistical procedures discussed thus far, with one exception, are designed for use with interval or ratio scores--scores indicated by numbers on a continuum having equal distances between them. The exception was the Spearman correlation (Unit V) which is applied to ordinal or ranked data. But there are also many instances in which data will be nominal or categorical. For example, is there an association between people's membership in charitable organizations and how they vote in elections? Are the city parks frequented more by one sex (or racial group or age group) than another? "Scores" here are essentially frequently counts of different subpopulations. How would such nominal data by analyzed? The purpose of this unit is to acquaint you with the uses and computational procedures of a test designed specifically for that purpose--the chi-square test.

B. Specific Objectives

10.1 Identify the type of scores analyzed by the chi-square test

10.2 Use chi-square in a one-way test

- 10.2.1 Test for equal frequencies
- 10.2.2 Test for goodness-of-fit

10.3 Use chi-square in a two-way test

10.1 Instructional Unit: Nominal scores and their analysis through chi-square

Scores indicating examination performances, outside temperatures, yearly incomes, size of houses, populations of cities, and on and on, all constitute interval measures or ratio measures. Each observation can take only one value in a given assessment (John can have only one score on the test), and can be ordered on a continuum with other observations on the basis of that value (John's score was 87 which is higher than Tom's 86). Further, the distances between the possible points on the continuum are equal and have the same meanings (Tom's 1 point deficit relative to John means the same as his 1 point advantage over Ken's 85). Interval measures do not have a natural zero point (e.g., a test score of zero usually does not indicate the total absence of knowledge) while ratio measures do (a "money earned" score of zero means exactly that--a complete absence of earnings!!).

Then we can move from interval and ratio measures to the somewhat less precise ordinal measures which might be thought of as rankings. Although ranks can be arranged in a numerical continuum, distances between points on that continuum may be widely uneven. The person ranked

#1 might be far superior on the trait in question to #'s 2 and 3 who are actually quite close to one another. Is the distance between 1 and 2 the same as that between 2 and 3? Not by a long shot.

Finally, we can move to an even less precise but just as important type of measure--nominal or categorical. Here we get neither a meaningful numerical score nor a rank, but simply an indication of membership or nonmembership in a particular category. Are church-goers more likely to participate in local charity drives than those not attending church? If we take a random sample of people participating in local charity functions, and Glenda Mender is a subject, what will her score be? Certainly not an interval or ratio score like a 97, or a rank like a 28. The only possibility here is a nominal score, indicating to which group, church or non-church, she belongs. When all subjects are "tested" we will simply end up with a frequency count indicating the number of observations (subjects) in each nominal category. On that basis, we can do our comparison: 80 out of 100 were church-goers; 20 were not. Interesting...

A statistical test specifically designed to enable such comparisons is the chi-square (χ^2) test. As will be seen below, general categories of uses pertain to cases in which there is one variable (are political parties equally represented in the Senate?); and two variables (are political parties and seniority related for members of the Senate?). Again, what χ^2 is looking at is nominal data--the frequencies of individuals who fall into different categories. It would allow us to make a statistical judgment as to whether, for example, a count of 54 Democrats and 46 Republicans does suggest unequal representation. Further, half of the Democrats might have 10 or more years of seniority while only one-third of the Republicans do. Is there a significant relationship there? Chi-square would help us to find the answer.

10.1 POSTTEST (answers in back)

1. Differentiate between interval-ratio, ordinal, and nominal measures. To which is the chi-square test applied? Give an example of the type of research question to which it might be applied.

___

___

___

___

10.2 Instructional Unit: The one-way χ^2 test.

In this unit, we'll use the term one-way to describe χ^2 applications in which one variable (e.g., race, sex, blood type, school attended) is being examined. Two related uses will be examined, one that tests for equal frequencies and another that tests for goodness-of-fit. As will be seen, the procedures for both tests are virtually identical. If you can do one you can do both.

10.2.1 Testing for Equal Frequencies

Suppose that at the monthly neighborhood school PTA meeting, we count 120 people in attendance. Further inquiry indicates that 60 of these people are parents, 45 are teachers, and 15 are "others" (students, administrators, maintenance people, etc.). Can we conclude that attendance is evenly distributed across these three categories (parents, teachers, or others)? Having come this far along in the book, you should be ashamed for trying to "conclude" anything. These 120 people are only a sample of all those who might attend such meetings. All that we can do is use this sample as a basis for inferring what might be found if we could characterize the total population of attenders. Just like we did in other inference-making situations, we will state a null hypothesis, run a statistical test, and then make an inference (not a firm conclusion) on that basis.

Since we're dealing with nominal data (frequencies), our test is χ^2. But what would our null hypothesis be? For these types of analyses, the following general statements can be used.

$$H_0 : O = E$$

$$H_a : O \neq E$$

Translation: The null hypothesis says "What is observed (O) will be the same as what is expected (E)--nothing unusual!". The alternative hypothesis, of course, says the opposite. "But what is expected?" you might ask. Well, since the null refers to the "no difference" condition, we'd expect in this study, that out of 120 people attending PTA meetings, one-third (n = 40) would be teachers, one-third parents, and one-third "others"--an even distribution. Let's run our test to see whether the evidence seems consistent with that condition. We need a formula:

$$\chi^2 = \Sigma \frac{(O - E)^2}{E}$$

where O is the obtained frequency and E is the expected frequency in each category. For our example, we can tabulate these as follows:

PTA Study #1

	Categories (Groups)			
Frequency	Parents	Teachers	Others	Total
Observed (O)	60	45	15	120
Expected (E)	40	40	40	120

Remember, the O's are what was actually observed. The E's are what we'd expect if there were no differences in group attendance. Does $O = E$, i.e., does the null appear true? Application of the formula will be the deciding factor.

$$\chi^2 = \frac{(60-40)^2}{40} + \frac{(45-40)^2}{40} + \frac{(15-40)^2}{40}$$

$$= \frac{400 + 25 + 625}{40}$$

$$= \frac{1050}{40} = 26.25$$

So, $\chi^2 = 26.25$, what do we do now? Just like for z or t, used for analyses of means, χ^2 has a probability distribution. The probability distribution, like these other ones, indicates the likelihood of obtaining different χ^2 values were the null to be true (i.e., $O = E$). So, again, we might take the .05 or .01 level as a basis for inferring that our sample represents a population where $O \neq E$ and justifies rejection of the null.

If you'll turn to Table III in the Appendix, you'll see such a chi-square table. Like the t table, it lists only "selected" probability values (.01, .05, .10)—the ones conventionally used in hypothesis testing. But, also like the t table, there's a "catch" in the form of df. You need to know df in order to obtain the critical value. For reasons we won't go into in this introductory book, df is determined in χ^2 one-way (one-variable) cases by the following simple expression: $k-1$, where k = *number of categories or groups*. In our case, $k = 3$; so $df = 2$.

Now we're ready. Suppose we want to test the null at $\alpha = .05$. What is the critical value we need to reach or exceed? For $df = 2$, it is 5.99.

What is the conclusion in our study? Reject the null. Results suggest that the three categories (groups) are not equally represented. Based on the frequency table, a reasonable inference is that fewer "others" attend PTA meetings for this school than do parents and teachers.

Another Example: Let's suppose that we repeat the study at some other school in a different state. Results for 180 people attending the meeting are as follows: 65 parents, 62 teachers, and 53 "others." Let's set up our O and E frequency table:

PTA Study #2

	Categories (Groups)			
Frequency	Parents	Teachers	Others	Total
Observed	65	62	53	180
Expected	60	60	60	180

Note that $f = 60$ becomes the E value for all, since if the $180n$ were evenly divided, 60 people would belong to each group. Also, df will be 2 since $k-1=2$. Let's test at $\alpha = .05$ again.

$$\chi^2 = \frac{(65-60)^2}{60} + \frac{(62-60)^2}{60} + \frac{(53-60)^2}{60}$$

$$= \frac{25 + 4 + 49}{60} = \frac{78}{60}$$

$$= 1.30$$

What is the decision?

Retain the null. The critical value, 5.99, was not reached. There is no basis for rejecting the null hypothesis of $O = E$.

10.2.2 Testing for Goodness-of-Fit

Directly related to the above uses of chi-square (i.e., testing for equal frequencies) is the goodness-of-fit test. The null hypothesis statement ($O = E$) and computational procedures are identical to that described in the previous section. The only difference is that we are comparing our observed (O) frequencies to an expected (E) "pattern" which can take any form, i.e., have unequal frequencies across categories.

For example, suppose that administrators at a certain university are concerned with grade inflation. They distribute a memo to faculty recommending that grades be awarded in roughly the following percentages.

	A	B	C	D	F	Total
%:	10	20	40	20	10	100

Professor Warmglow has the reputation for being an easy grader, so his department chairman calls him in to present his grade distribution from the previous semester. Warmglow had 130 students who received the following grades:

	A	B	C	D	F
f:	25	40	35	20	10

Warmglow does seem more generous than the university would like. Is he a grade inflator? Let's test goodness-of-fit, using $\alpha = .05$. Hypotheses are:

$$H_0 : O = E$$

$$H_a : O \neq E$$

Now we'll construct our frequency table. In doing so, note that the E's here will reflect the expected proportions of students in the different grade categories. For example, based on what the university expects, 10% of Warmglow's 130 students should have been A's: $E_a = .10 \times 130 = 13$; 20% should have been B's: $E_b = .20 \times 130 = 26$; etc. Our frequency table thus becomes:

Grade Inflation Study
Professor Warmglow

Frequency	A	B	C	D	F	Total
			Categories (Grades)			
Observed	25	40	35	20	10	130
Expected	13	26	52	26	13	130

We compute χ^2 in the usual way:

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

$$= \frac{(25-13)^2}{13} + \frac{(40-26)^2}{26} + \frac{(35-52)^2}{52} + \frac{(20-26)^2}{26} + \frac{(10-13)^2}{13}$$

$$= \frac{144}{13} + \frac{196}{26} + \frac{289}{52} + \frac{36}{26} + \frac{9}{13}$$

$$= 11.08 + 7.54 + 5.56 + 1.38 + .69$$

$$= 26.25$$

Here $df = 4$, since number of categories $(k) = 5$. Table III gives the critical χ^2 value as 9.49 for $\alpha = .05$, $df = 4$.

Decision: Reject the null. Warmglow distribution does not provide a very good fit with the University's.

Do you see that the goodness-of-fit test involved exactly the same procedures as the "equal frequencies test," except the expected values (E's) were based on some model. We applied χ^2 to determine how well the O's conformed to that model--testing for equal frequencies were not of concern (unless they were that way in the model!).

Very quickly, we'll test to see how Professor Warmglow's colleague, Dr. Blackhart did in his grading of 50 history students.

Dr. Blackhart's Grades

Frequency	A	B	C	D	F	Total
			Categories			
Observed	5	5	10	15	15	50
Expected	5	10	20	10	5	50

It is important before continuing that you understand how the E's were derived. Remember, the university recommended 10% A's = .10x50 = 5; 20% B's = .20x50 = 10; etc.

$$\chi^2 = \frac{(5-5)^2}{5} + \frac{(5-10)^2}{10} + \frac{(10-20)^2}{20} + \frac{(15-10)^2}{10} + \frac{(15-5)^2}{5}$$

$$= 0 + 2.5 + 5 + 2.5 + 20$$

$$= 30$$

31

Since df hasn't changed, the critical χ^2 value remains at 9.49.

Decision: Reject the null. Blackhart's distribution doesn't provide a very good fit with the University's either. Which distribution would students hope your statistics teacher's is closest to, Warmglow's or Blackhart's? If you answered Blackhart, you could well be one of the people helping to make the distribution that way!

Final Note: Whenever $df = 1$ in the one-way test, it is recommended that an adjustment, officially known as Yates' correction, be made in the formula. When will $df = 1$? Wherever you have only two categories, since $df = k-1$. The correction involves subtracting the constant .5 from each $O-E$ in the numerator of the formula. Why? Mathematical analyses have shown that when there are only two categories, the conventional χ^2 formula tends to overestimate the values that should be obtained if the null ($O=E$) were true. Yates' formula corrects for this. Here it is:

$$\chi^2 = \Sigma \frac{(|O-E| - .5)^2}{E}$$

The special vertical parentheses means that we will ignore the sign of the a difference, and treat all $O-E$'s as positive. Suppose Prof. Blackhart's grades were simply reported as Pass or Fail. Since F is the only failing grade, the new table would look like this.

Dr. Blackhart's Passes/Failures

	Categories		
Frequency	Pass	Fail	Total
Observed	35	15	50
Expected	45	5	50

$$\chi^2 = \frac{(|35-45| - .5)^2}{45} + \frac{(|15-5| - .5)^2}{5}$$

$$= \frac{(10-.5)^2}{45} + \frac{(10-.5)^2}{5}$$

$$= \frac{(9.5)^2}{45} + \frac{(9.5)^2}{5}$$

$$= 2.01 + 18.05$$

$$= 20.06$$

Looking at the table for $\chi^2_{(.05, df=1)} = 3.84$

Reject the null.

10.2 POSTTEST (answers in back)

1. A librarian is interested in determining the popularity of different types of books. She defines a number of categories and does a frequency count of the number of books taken out in each. Results for a total of 108 books are as follows: Novels=40, Nonfiction=25, Children's Books=28, "Other Types"=15.

   a. Given that she wants to test for equal frequencies, state the null hypothesis. Interpret what it means in terms of this situation.

   b. What are the observed and the expected frequencies?

   c. What is the computed χ^2?

   d. What is the tabled χ^2 value, for $\alpha = .01$?

   e. What is the decision?

   f. What type of error is possible?

2. Census data for a particular city indicates that its total population consists of 40% Caucasians, 40% Blacks, 15% Chicanos, and 5% other. The city Zoo Director wants to determine whether the racial make-up of those visiting the zoo conforms to that distribution. He monitors the zoo attendance and finds the visitors to be: 35% Caucasians, 41% Black, 20% Chicanos, and 4% Other.

   a. Fill in the following table:

Zoo Attendance

Frequency	Caucasian	Black	Chicano	Other	Total
Observed					
Expected					

   Note: We can use the percentages as if they were frequencies. All that matters is that the O's and the E's are on the identical score scale (i.e., f's or %'s, etc.)

   b. Test the Zoo Director's hypothesis at $\alpha = .05$. What is the computed χ^2? What is the decision?

3. Suppose the Zoo Director in the above example didn't know the city figures. Instead he wants to test whether the number of visitors to the zoo varies on the basis of race. Test this hypothesis using χ^2. (Again, simply use the %'s as if they were frequencies. Your E's will also have to be in that scale, but note that the E's for this hypothesis will not be the same as those for #2.)

4. Identify the circumstances which dictate using Yates' correction in the χ^2 formula.

10.3 Instructional Unit: Using chi-square in a two-way test

In the last section we looked at uses of χ^2 to test hypotheses concerning one variable. In other instances, we may wish to test hypotheses concerning two (or more) variables. The basic interest in the latter case usually boils down to questions about the relationship or independence of two variables. Are the two variables related to (or dependent on) each other? Are they unrelated or independent of each other? Examples might be, are peoples' ethnic backgrounds related to the way they vote? Is occupation related to where people live in a city? Testing for "relationship" might bring to mind the correlational techniques studied in Unit V. That's an entirely appropriate association--the difference, here, though is that the data are nominal/categorical measures (i.e., f's). Otherwise, the two applications (correlation and χ^2) can be used to answer quite similar research questions.

It seems best at this point to move quickly to examples, and illustrate the two-way design through them. Here we go:

The Rubber Duck Manufacturing Co. has 3 general classes of employees: management, worker (secretarial, assembly line, mechanical, etc.), and custodial. In planning for shifts in personnel due to changes in the economy, company officials want to know if a relationship exists between job category and location of residence: city, suburbia, country. They survey their 150 employees and get the following results:

Residence Location	Manager	Worker	Custodial	Total
	Job Category			
City	7	60	12	79
Suburbia	13	20	3	36
Country	5	20	10	35
Total	25	100	25	150

Is residence location related to job category? χ^2 to the rescue! Our hypotheses:

$$H_0 : O = E$$

$$H_a : O \neq E$$

Surprised? No need to be. We're again asking whether the observed frequencies conform to what would be expected (if the null were true). What is the null saying here? It represents a condition in which there is no relationship between variables; a condition where the variables are independent of one another.

Think hard about what is being said as you read this paragraph. It is the "crux" of the logic behind the χ^2 application. Give yourself an A+ and a pat on the back (and maybe even a cookie!) if you can answer the following question without reading further.

Question: If the null were true, what would be the expected frequencies for the table? (How would one go about determining them?)

Answer: If the null were true, Residence and Job would be unrelated. The same proportions of people living in the three different types of locations would exist for management personnel, workers, and custodians. For example, ignore the separate figures for the three job categories and look only at the Totals in the last column on the right. What percentage of workers, overall, live in the city? 79/150 = 52.67%. If the null is true, then 52.67% of managers, 52.67% of workers and 52.67% of custodians should live in the city. What, for example, is the actual percentage of managers who live in the city? 7/25 = 28%. Not much support for the null, is it? But we need a statistical test to find this out more objectively and systematically.

Believe it or not, the χ^2 formula we need is exactly the same as the one used in the last section. But, what still might present a bit of uncertainty is the task of finding the E's. We need 9 of them because our problem has 3(job categories) x 3(locations) = 9 cells. The formula for identifying each E is shown below:

$$E = \frac{\text{(row tot.) (column total)}}{\text{Overall total}}$$

Applying this to our results:

		Manager	Worker	Custodian	Row Total
City	O :	7	60	12	79
	E :	$\frac{(79)(25)}{150} = 13.17$	$\frac{(79)(100)}{150} = 52.67$	$\frac{(79)(25)}{150} = 13.17$	
Suburbia	O :	13	20	3	36
	E :	$\frac{(36)(25)}{150} = 6$	$\frac{(36)(100)}{150} = 24$	$\frac{(36)(25)}{150} = 6$	
Country	O :	5	20	10	35
	E :	$\frac{(35)(25)}{150} = 5.83$	$\frac{(35)(100)}{150} = 23.33$	$\frac{(35)(25)}{150} = 5.83$	
Column Total		25	100	25	150

Note that E for managers in the city = 13.17. That was derived by multiplying the Row Total for that cell (79) by the Column Total (25) and dividing by the overall total (150). The same procedure was applied to the remaining 8 cells. And now for the χ^2 formula:

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

$$= \frac{(7-13.17)^2}{13.17} + \frac{(60-52.67)^2}{52.67} + \frac{(12-13.17)^2}{13.17} + \frac{(13-6)^2}{6} + \frac{(20-24)^2}{24} +$$

$$\frac{(3-6)^2}{6} + \frac{(5-5.83)^2}{5.83} + \frac{(20-23.33)^2}{23.33} + \frac{(10-5.83)^2}{5.83}$$

$$= 2.89 + 1.02 + .10 + 8.17 + .67 + 1.5 + .12 + .48 + 2.98$$

$$= 17.93$$

Is that result significant, i.e., do we reject the null ($\alpha = .05$)? What we need to know is df.

In the two-way χ^2, df is determined by the following expression:

$$df = (C-1)(R-1)$$

Where C = number of columns and R = number of rows

Thus $df = (3-1)\ (3-1)$

$= 4$

What is the critical value for $\chi^2_{.05},\ df = 4$?

It is 9.49 (make sure you can find this on Table III).

Decision: Reject the null. The two variables do not appear independent since the computed value 17.93 > critical value 9.49. Looking at the table, what is your interpretation of the relationship?

Here's ours: Managers tend to live in suburbia more than is usual, workers more in the city, custodians more in the country.

<u>Final Notes</u>: In case you're developing any doubts about your ability to remember "all" the different χ^2 applications and formulas, here's a summary chart to help you put everything in place.

I. <u>One-Way Analysis (one variable)</u>

A. Test H_0: $O=E$, for either:

1. Equal frequencies

   A. O = the f's observed
   B. E = the expected number if all categories were equal

2. Goodness-of-fit

   A. O = the f's observed
   B. E = the expected f's if the proportions in each category conformed to some hypothesized (or known) distribution.

B. Formula: $\chi^2 = \Sigma \frac{(O-E)^2}{E}$

C. $df = k-1$

D. Special Case:
1. If $df = 1$, use Yates' correction formula (see p. 305).

2. O's and E's can be f's, proportions or percentages, but they must <u>both</u> be the same type of score.

II. <u>Two-Way Analysis (two variables)</u>

A. Test H_0:$O=E$ where

1. O = the frequencies actually observed.

2. E = the frequencies expected if the two variables are independent (unrelated).

$$E = \frac{\text{(row tot.)(column tot.)}}{\text{Overall Total}}$$

B. Formula: $\chi^2 = \Sigma \frac{(O-E)^2}{E}$

C. $df = (C-1)\ (R-1)$

D. Special case:
1. Use regular formula even if $df = 1$.
2. O's and E's must both be the same type of score.

10.3 POSTTEST (answers in back)

1. A researcher hypothesizes that the marital status of women is related to their employment status. After surveying a random sample of 200 women, the following results are collected:

Employment Status

Marital Status	Unemployed	Employed Part-time	Employed Full-time	Total
Single	10	20	30	60
Married	60	50	30	140
Totals	70	70	60	200

Use the two-way χ^2 analysis to test the independence of variables at $\alpha = .01$. What is:

a. the computed χ^2?
b. df ?
c. the tabled χ^2?
d. the decision?

2. The researcher (from #1) repeats the study using men. He finds:

Employment Status

Marital Status	Unemployed	Employed Part-time	Employed Full-time	Total
Single	10	20	40	70
Married	15	40	75	130
Totals	25	60	115	200

What is:

a. the computed χ^2?
b. the tabled χ^2 for $\alpha = .05$?
c. the decision?

Unit X Review Test
(answers in back)

1. Differentiate between a one-way and a two-way χ^2 analysis. Give an example of each.

2. Differentiate between χ^2 tests for equal frequencies and goodness-of-fit. Where do they differ: formula, O's, E's? Explain.

____

____

____

3. The athletic director of a university wonders whether attendance at basketball games is similar for students at different class levels. He collects attendance data and finds the following:

Freshmen	Sophomores	Juniors	Seniors	Total
20	30	25	25	100

Should he conclude that the classes differ?

a. What χ^2 test is appropriate for that hypothesis, equal f's or goodness-of-fit?

b. What is the computed χ^2?

c. What is the tabled χ^2, for $\alpha = .05$?

d. What is the decision?

4. The math instructor at the university wonders whether attendance at the games conforms to the make-up of the student body in the following majors: Arts and Sciences: 15%, Business 30%, Education 20%, Others 35%.

The data he collects indicates the following attendance outcomes:

	Arts and Sciences	Business	Education	Others	Total
%	30	10	30	30	100

Does the attendance distribution conform to how majors are distributed at the university?

a. What is the computed χ^2?

b. What is the tabled χ^2, for $\alpha = .01$?

c. What is the decision?

d. Interpret the meaning of that decision with regard to the research question.

5. Use the following results to test the independence of 120 individuals' preferences for political party and for newspapers in a certain town.

Newspapers	Republican	Democrat	Undecided or other	Total
	Political Preference			
Press Minotaur	30	15	25	70
Daily Planet	10	30	10	50
Totals	40	45	35	120

a. What is the decision based on χ^2 results? ($\alpha = .05$)

b. Interpret what it means with regard to the research question.

UNIT XI

INTRODUCTION TO ANALYSIS OF VARIANCE

A. General Objectives

This unit is designed to provide a transition point from the present book's introductory treatment of statistics to the more advanced treatment likely to be encountered in subsequent courses. The selected area for the transition is "analysis of variance" (ANOVA)--a highly useful tool in statistical analysis, but (depending on instructor preference and student level) not always a topic that can be adequately covered in a beginning course. Thus, with the goal being to provide a transition rather than a thorough treatment, the present unit on ANOVA will be relatively brief, placing emphasis on its rationale, exemplary uses, and interpretation rather than on its mathematical properties.

B. Specific Objectives

11.1 Describe the rationale for analysis of variance (ANOVA) designs

11.2 Describe the general procedures used in computing ANOVA

11.3 Interpret the statistical and research implications of ANOVA results

11.1 Instructional Unit: Using analysis of variance in research

The present unit is intended as a bridge to help you to step more easily from the introductory topics covered thus far to the more advanced ones likely to be encountered in later courses (you were planning to continue your study of statistics, weren't you?). Intermediate courses (textbooks) have the habit of dashing through introductory topics at breakneck speed only to set up camp and travel at a more leisurely pace within a new territory--analysis of variance (ANOVA) designs. Uncomfortable in these new surroundings, students may frantically dust off their old introductory books searching for a friendly road map to connect the new with the old. But one of two things often happens. Either the road map isn't there (ANOVA isn't covered); or the old map looks almost exactly the same as the new one--some explanation, but mostly formulas and math applications requiring too much "decoding" to provide the direction needed. The following approach is intended to be different. Its focus will be on practical, user-oriented concepts that should (a) be understandable even long after you forget symbols and formulas (many years from now), (b) provide a helpful transition to more advanced study, and (c) leave you capable of using/interpreting simple ANOVA designs before initiating that progression.

Why ANOVA? Couldn't They Have Stopped With the *t* Test?

Picture the following situation. A health specialist is concerned with testing the effectiveness of five exercise programs on weight loss. The programs consist of: (1) jogging, (2) swimming, (3) weight lifting, (4) push-ups, and (5) aerobics. So, remembering things that we did in Unit IX, you might recommend obtaining a sample pool of, say, 100 eligible men, and randomly assigning each subject to one of the treatments (20 in each). We would then go ahead and compare each treatment mean to each of the others, using the t test for independent samples. The number of unique combinations of pairs would be 10: Sample 1 vs. 2, 1 vs. 3, 1 vs. 4, 1 vs. 5, 2 vs. 3, etc.). Findings would indicate which sample means were significantly higher than others. We could then make inferences about the associated populations (i.e., which exercise program, if given to all eligible people, would be best).

But suppose that in reality all programs are <u>exactly</u> the same in effectiveness; each should result in an average weight loss of 10 lbs. The researcher, of course, doesn't know this. If he did, there would be no reason to do the experiment!

<u>Big Question</u>: If he runs his analyses, which is the only type of error (I or II) that he could make?

<u>Small (but important) Answer</u>: Type I

Given that there are actually <u>no</u> differences, he'll be making a correct decision if he retains the null. If he rejects it, Type I all the way! Can you see a problem for the Type I error rate when multiple t tests are used? Remember, this research ran 10 tests to compare all unique pairs of means.

To get you primed for identifying the problem, try this question. What is the Expected Value, $E(\overline{X})$ for the Swimming Treatment mean? Answer: 10 lbs. (see above). What, then is $E(\overline{X})$ for the Aerobics Treatment? Answer: 10 lbs. In fact, <u>all</u> five $E(\overline{X})$'s are 10, since all treatments are equally effective (and the parameter value was defined as 10).

Will the actual Treatment $\overline{X}$'s <u>all</u> be equal to 10? No!! Sampling fluctuation and random error will inflate some $\overline{X}$'s and reduce others. <u>Here's the key point (at last!)</u>: Isn't it true, then, that the more treatments (samples) you compare, the greater the chances of obtaining at least one weird (extreme) $\overline{X}$, even though the particular treatment is not different in its effects from any other? Thus, the more multiple t comparisons you make, the greater the chances that at least one Type I error will be made. Think about it one more time in terms of this rather extreme example: If we tested 100 different random samples from the same population, isn't it very likely that some will "appear" to represent different populations (at $\alpha = .01$ even!)? What makes them different? Sampling fluctuation is the culprit. Each t test holds the Type I error rate at .05 for <u>that</u> one comparison. But what isn't held at .05 is Type I error rate across <u>all</u> comparisons (4,950 unique comparisons in the above example). That error rate will be much larger than 5%.

Now for the answer to the question we started out with. Why ANOVA? The reason is that ANOVA allows us to conduct an "overall" test of differences between treatments. ANOVA holds the chances of falsely concluding that treatments differ from one another at α (e.g. .05, .01). The null hypothesis statement for our example thus becomes:

ERIC

$$H_0 : u_1 = u_2 = u_3 = u_4 = u_5$$

If retained, the suggestion is that the population means represented by the sample (treatment) means do not differ. If rejected, the suggestion is that there is at least one difference between population means. What is that difference? ANOVA doesn't say, since it is an "overall" test, but follow-up analyses can be used to probe that issue (those are easy to perform, but describing them would take us beyond the scope of this textbook). How does ANOVA work? We'll turn to that question in the immediately following section. First, try your hand at the following "thought" questions.

11.1 POSTTEST (answers in back)

1. What is the problem with comparing more than two treatments using a t test for each unique pair?

2. What happens to the Type I error rate as additional treatments are added to the design and compared to one another?

3. Four treatment groups are compared through ANOVA. State the null hypothesis.

4. Suppose the null hypothesis (from #3) is retained. What is the implication about the population means in question? Suppose it is rejected. What is the implication then?

11.2 Instructional Unit: Understanding general procedures for computing ANOVA

In previous units we used statistics called z, t, and χ^2 in testing hypotheses. The logic was to compute the given statistic from a mathematical formula and then compare the result to a probability table (Tables I-III in the Appendix). The table showed the probability of obtaining different values of the statistic, <u>given that the null hypothesis was true</u>.

The ANOVA procedure also involves computation of a statistic. In its case, the statistic is called F and is derived from the following ratio:

$$F = \frac{\text{Variability between groups}}{\text{Variability within groups}}$$

Don't give up--give this concept a chance! By "variability between groups" we mean how much the <u>treatments</u> (or groups) differ from one another on the criterion. We would determine this specifically by computing the variance (σ) between the group means ($\overline{X}$'s). But for now, think about it this way. If we were comparing the effectiveness of three study procedures on test performance, we might obtain results like $\overline{X}_1 = 80$, $\overline{X}_2 = 70$, and $\overline{X}_3 = 100$. Someone else might compare three different procedures and get: $\overline{X}_1 = 80$, $\overline{X}_2 = 82$, and $\overline{X}_3 = 80$. It should be obvious that variability between groups is larger in the first study than in the second. Without knowing the actual numerical value of that variance (or the value of the denominator of the F ratio) it seems likely that the first analysis will produce a larger F value than the second. Just like for t, z, and χ^2, the larger the value of the statistic, the less probable it becomes, and the stronger the basis for rejecting the null. In this case:

$$H_0 : u_1 = u_2 = u_3$$

So, at this point, the null looks less secure in the first study than it does in the second.

Turning to the denominator of the ratio, what does it mean? By "variability within groups" we mean how much the scores <u>within</u> the treatments themselves differ from each other. This should be a familiar concept because it has come into play every time we have computed the variance or standard deviation for a set of data. So, the F formula really involves analyzing "differences between groups" relative to "differences between people" within those groups.

Now, let's turn our thoughts to some of the beautiful logic that underlies the F formula. Think about this for a moment: What factors would account for people <u>within</u> treatments obtaining different scores? After all, all receive the same treatment.

If you're thinking along the lines that some are brighter, some try harder, those same descriptions (brighter, trying hard) might fit you as well! People differ, so there's going to be variability in their scores. This is variability over which we have no control. Statisticians call it <u>random error</u>.

So, let's ask the question again, using slightly different words. What factor accounts for the denominator of the F ratio? Answer: Random error.

$$F = \frac{\textit{Variability between groups}}{\textit{Random error}}$$

Ready to take another step in our journey? Try this one. Assume that the null is <u>false</u>, and identify the factors that would account for groups (or treatments) having different means.

One very good answer would be that if the null is false, the groups would differ because their respective populations differ. Another way of saying it, is that there is an honest-to-goodness <u>treatment effect</u>. (The groups differ because they receive different treatments and the treatments differ in effectiveness.) But is this the only factor that would lead to group variability? No--there's one more. Can you identify it?

The second factor is .... random error! Just like people can differ within groups due to uncontrollable factors, groups are affected by the same properties. (We called this sampling fluctuation in earlier discussions.) So, when the null is <u>false</u>, the F ratio is determined by:

$$F = \frac{\textit{Random error + Treatment effect}}{\textit{Random error}}$$

Since there is no reason to assume that the two random error estimates differ significantly, we can begin to make a prediction about the nature of our computed F. What would it be in your opinion: (a) a number equal to 1.00, (b) a number less than 1.00, or (c) a number greater than 1.00? Looking at the expression, the expected value of F would have to be <u>greater</u> than 1.00. So, the larger the treatment effect, the larger the F value will tend to be. And, on F tables (one is <u>not</u> included in this book), you might expect "statistically significant" F's to be ones with values greater than 1.00. That <u>is</u> the case! (How much greater, though, depends on other factors.)

Now let's assume that the null is <u>true</u>. What factor would account for the denominator of the F ratio? <u>Random error</u> again--no change here. What factor(s) would account for the numerator? Random error another time! But what about that treatment effect we talked about earlier? It doesn't appear this time because if the null is true, no treatment effect exists. This leaves us with the following as an expectancy:

$$F = \frac{\textit{Random error}}{\textit{Random error}}$$

So, when the null is true, what is the expected F: (a) a value of about 1.00, (b) a value less than 1.00, or (c) a value greater than 1.00? Looking at the expression, "a" should be the obvious choice. When ANOVA is run, F values close to 1.00 are suggestive of "no treatment effects" and justify retention of the null.

In real life, of course, we don't know whether or not the null is true. All that we can do is conduct our study, collect our results, and analyze them through the F formula. The magnitude of the F value computed will then be used to infer whether or not a treatment effect is likely to be operating.

Sums of Squares and Degrees of Freedom

The above non-mathematical conceptions of the F ratio are useful for conveying the logic involved in the procedure. But what numerical values actually represent the different types of effects? Here we must resort to some mathematical relationships whose logic may not be so obvious to the beginning student. But give them a try. You might surprise yourself.

Remember the opening example involving the five types of exercise treatments? There were 20 subjects randomly assigned to participate in each program. Let's say that after all have participated, we average the scores of the total sample of 100, as if treatments didn't exist, to obtain a grand (overall) mean. In this example we'll pretend that $\bar{X}_{total} =$ 6.2 lbs. (weight loss). Now, continuing to ignore treatment membership, we compute the sum of the squared deviations of each score from the overall mean, just the way we would have in any Unit III problem:

$$\Sigma x^2 = \Sigma(X-\bar{X}_{tot.})^2$$

(Subtract the total mean, 6.2, from each score, square result, and sum all such results.)

or

$$= \Sigma X^2 - \frac{(\Sigma X)^2}{n}$$

where $n = 100$

In ANOVA, the obtained value has a name, sum of squares total, symbolized SS_{total}.

By a surprisingly simple mathematical proof (which we won't demonstrate), it can be shown that SS_{total} can be partitioned into two components, symbolized because of the way they are computed as $SS_{between}$ and SS_{within}.

$SS_{between}$ is derived by summing the squared deviations of group means about the overall mean. The farther the group means are from one another, the larger $SS_{between}$ will be.*

$$\Sigma(\bar{X}_{group} - \bar{X}_{tot.})^2$$

(Subtract the total mean, 6.2, from each group mean, square result, sum all such results.)

*In actually computing $SS_{between}$, the sum of the deviation scores are weighted by the number of subjects in each group: $\Sigma n(\bar{X}-\bar{X}_{total})^2$. (See page 328 for more detail.)

SS_{within} is derived by adding the squared deviations of scores about their respective group means. It is influenced by the extent to which individuals in a group differ from one another (remember random error?),

$$SS_{within} = \Sigma(X-\bar{X}_1)^2 + \Sigma(X-\bar{X}_2)^2 + \ldots + \Sigma(X-\bar{X}_{last})^2$$ (See p. 328 for more detail.)

If you can make sense out of these expressions, fine. If not, don't worry. It's the procedure that's important in this treatment of ANOVA.

Whether or not you can translate the expressions, perhaps you perceived their relationship to the components of the F ratio introduced earlier. Your tendency might be to say, "I got it! You compute $SS_{bet.}$ and SS_{within}, divide the former by the latter, and you got F!" Well, almost, but not quite. Variances, you may recall (from Unit III), are averages. Analysis of Variance (ANOVA), therefore, implies that just "summing along" (as done to get the various SS's) is not sufficient.

Specifically, each SS must be converted into an average (mean) to get the entries for the final ANOVA formula. What was the appropriate divisor (df) used in Unit III to get s, the sample variance? It was $n-1$. The same principle holds here: Since $SS_{bet.}$ is derived from the squared deviations of treatments means (let's call their number a) from the overall mean, $df_{bet.} = a-1$ (# of treatments minus one).

SS_{within} is derived from the squared deviations of scores from their treatment means. In each treatment, $df = n-1$. If there are a treatments, $df_{within} = a(n-1)$.

SS_{tot} is based on the squared deviations of scores from the overall mean: $df_{tot} = an-1$.

Here's a summary:

$df_{bet.} = a-1$ $\qquad a$ = # of treatments

$df_{within} = a(n-1)$ $\qquad n$ = # of subjects in each treatment

$df_{tot} = an-1$

In our example involving 5 treatments and 20 subjects in each:

$$df_{bet} = 5-1 = 4$$

$$df_{within} = 5(20-1) = 95$$

$$df_{tot} = (5)(20) - 1 = 99$$

Note that 95 + 4 = 99.
Remember we said that: $SS_{tot} = SS_{bet} + SS_{within}$
This relationship is also reflected in df's.

In any case, now we are ready for the actual computation formula:

$$F = \frac{SS_{bet.}/df_{bet.}}{SS_{within}/df_{within}}$$

A simpler and more commonly used notation is to express the two components as "Mean Squares" (MS).

$$MS_{bet.} = SS_{bet.}/df_{bet.}$$

$$MS_{within} = SS_{within}/df_{within}$$

So, we compute F by:

$$F = \frac{MS_{bet.}}{MS_{within}}$$

F Summary Tables

When ANOVA is computed, the results are normally presented in a summary table. As a matter of fact, computerized data analysis programs provide these as standard output. They appear as follows:

Source	SS	df	MS	F
Between	620	4	155	35.07*
Within	420	95	4.42	
Total	1040	99		

*Significant at .05 level

The numbers filled in are hypothetical results for our "exercise" study. Note that when $SS_{between}$ and SS_{within} were computed, results were 620 and 420, respectively. $SS_{tot.} = 1040$, which is the sum of the <u>between</u> and <u>within</u> components. Since there were five groups, $df_{bet.} = 4$. By using the formula, $a(n-1)$, $df_{within} = 95$. Dividing each SS by df gives us the values in the MS column. F is the ratio of those MS values. The asterisk next to the 35.07 tells the reader that the F value was significant. (Remember a value close to 1.00 is expected if there is no treatment effect.) This study reflected significant variation between the five treatments. We can infer that a treatment difference occurred. Examination of means (and follow-up tests) will later be needed to tell us exactly where the differences lie.

Shown below is a partially completed summary table for a study in which three different teaching strategies were compared in terms of student achievement on a final exam. The number of students assigned to each treatment was 16. Can you fill in the missing values (indicated by question marks in parentheses)?

Source	SS	df	MS	F
Between	42	(?)	(?)	(?)
Within	(?)	(?)	(?)	
Total	138	(?)		

A lot of question marks, no question about that! But all the information you need to derive every single value is there. First, if we know that $SS_{tot.} = 138$ and that $SS_{bet.} = 42$, then $SS_{within} = 138 - 42 = 96$. The df's are $a - 1 = 2$, and $a(n-1) = 3(16-1) = 45$, for the between and within components. Total $df = 2 + 45 = 47$.

$$MS_{bet.} = \frac{42}{2} = 21$$

$$MS_{within} = \frac{96}{45} = 2.13$$

$$F = \frac{21}{2.13}$$

$$= 9.86$$

You would need an F table to tell you whether that value is significant. In case you're curious, it is significant ($\alpha = .05$). We would conclude that there is a treatment effect.

Again, what we have tried to do in this section is give you the basic logic and procedures involved in ANOVA (very little emphasis has been put on explaining calculations). The goal is to enable you to interpret ANOVA results that you might read in research reports, and to progress to more advanced courses (should you so desire) having had some background with the ANOVA procedures. Try these posttest questions to test your understanding of the material in the section. Then progress to the last section (11.3) which provides a brief review of how to interpret ANOVA results.

11.2 POSTTEST (answers in back)

1. The underlying logic in ANOVA involves the following ratio

$$F = \frac{\text{variability between groups}}{\text{variability within groups}}$$

In your own words, explain what the ratio means.

____

____

____

2. In reference to the expression shown in #1, indicate (a) what factor(s) account for the denominator (regardless of whether the null were true or false).
(b) What factor(s) would account for the numerator if the null were true?
(c) What factor(s) would account for the numerator if the null were false?
(d) What would the Expected Value of F be if the null were false:
(1) > 1.00, (2) < 1.00, or (3) about = 1.00? Explain.

3. What is the interpretation regarding treatments (or groups) if the null hypothesis is rejected?

4. Suppose the four treatment groups of 10 subjects each are compared by ANOVA.
a. What is $df_{bet.}$?
b. df_{within}?
c. $df_{tot.}$?

5. Replace the question marks in the following ANOVA summary table with the correct values.

Source	SS	df	MS	F
Between	200	4	?	?
Within	?	40	?	
Total	1000	?		

11.3 Instructional Unit. Interpreting ANOVA results

This section is designed to boost your confidence that you can look at an ANOVA summary table and interpret the findings. While it is true that the present chapter provides very little exposure to calculating F from original data, it is also true that in practice, hand calculations are rarely done given that there is any access to computers. The computer data analysis package (e.g. SPSS) tells you how to enter the data and what "control" information is needed for the run. The output may consist of many statistical indices, but in the case of ANOVA, the important information is the ANOVA summary table. If you can interpret it, you are well on the way to understanding what occurred in that part of the study.

To give you practice at doing this, some ANOVA summary tables from a hypothetical study will be presented for your consideration.

Description of Study

High school students were randomly assigned to three different learning treatments, 17 per treatment group. One treatment presented statistical material using education-related examples. Another presented the same material using sports examples. A third presented the material using abstract examples (A+B=C, etc.). The groups were compared on a number of different criteria. Below are some results:

Analysis #1

In one analysis, the groups were compared on their ability to solve completely new problems involving higher-order applications of the material (30 item test).

Results for the three groups were:

	$\bar{X}$
Education	23.5
Sports	22.9
Abstract	15.3

ANOVA Summary

Source	SS	df	MS	F
Between	71.76	2	35.88	3.33*
Within	516.47	48	10.76	
Total	588.23	50		

*Significant at .05

(Can you figure out how the df were derived?)

What happened in this analysis?

Ans.: The significant F indicates a significant treatment effect. The null hypothesis of "no treatment differences" is rejected. Looking at the means, the impression we get is that the Education and Sports groups performed about the same. The Abstract group, though, "bombed out." (Although follow-up tests will be needed to confirm that, such a description seems a reasonable account of the data.) Conclusion: The Abstract presentation was a relatively poor teaching device.

Analysis #2

In another analysis, the three groups were compared on their retention of formulas. Results were:

	$\bar{X}$
Education	21.76
Sports	24.12
Abstract	22.35

ANOVA SUMMARY

Source	SS	df	MS	F
Between	5.10	2	2.55	.13 (not significant)
Within	936.47	48	19.51	
Total	941.57	50		

Conclusion? There were no significant differences between groups. The null hypothesis should be retained. If you examine the means, you can get a further idea of why statistical analyses are needed. Without a statistical test, what's to prevent the researcher from concluding that the Sports treatment (which produced the highest mean) was superior for learning formulas? The statistical test (ANOVA) let us know that such "superiority" could easily have been obtained by <u>random error</u> (chance). The value of the computed F is not large enough to suggest a real treatment effect.

Analysis #3

In the last analysis, the three groups were compared on attitudes toward instruction. Scores were derived from a survey, where 8 was the highest (most positive) score and 0 was the lowest.

Results were:

	$\bar{X}$
Education	5.53
Sports	6.71
Abstract	4.47

ANOVA SUMMARY

Source	SS	df	MS	F
Between	42.51	2	21.26	7.39*
Within	138.00	48	2.88	
Total	180.51	50		

*Significant at .05

11.3 POSTTEST (answers in back)

Interpret the results of Analysis #3.

____

____

____

____

____

Unit XI Review Test
(answers in back)

1. Describe the effect on the Type I error rate as additional comparisons between pairs of means are conducted via t tests.

   ____

   ____

2. As differences between treatments increase, the size of the F value in ANOVA tends to ________

   a. increase   b. decrease   c. remain the same

3. As the number of treatments compared via ANOVA increases, $df_{bet.}$ ________

   a. increases   b. decreases   c. remains the same

4. As the number of treatments compared via ANOVA increases, with α held constant, the chances of a Type I error ________

   a. increase   b. decrease   c. remain the same

5. As additional subjects are added to treatments, with everything else held constant, df_{total} ________

   a. increases   b. decreases   c. remains the same

6. As additional subjects are added to treatments, with everything else held constant, $df_{bet.}$ ________

   a. increases   b. decreases   c. remains the same

7. As treatment effects increase, with everything else held constant, the amount of random error ________

   a. increases   b. decreases   c. remains the same

8. Complete the following ANOVA Summary Table for a study involving four treatments with 18 subjects in each.

Source	SS	df	MS	F
Between	60	?	?	?
Within	?	?	?	
Total	650	?		

9. If the F probability table shows the critical value (at α = .05) needed to reject the null as 2.75, what is the decision in #8?

What is the meaning of that decision with regard to treatments?

____

____

____

____

(See pp. 328-329 for "optional" explanations on F computations and critical values.)

* * * * * * * * * * * * *

SOS for SSS

In case you are suffering from Statistics Symbol Shock (SSS) and have no idea what the different *SS* formulas mean (i.e., you have SSS from the *SS*'s), here's an explanation. (Keep in mind that there are simpler to use computational formulas--but these lose visible association with the *SS* definitions in the translation.)

A. $SS_{total} = \Sigma(X-\overline{X}_{tot.})^2$

1. Subtract the total mean, $\overline{X}_{total}$, from each score (regardless of groups)
2. Square each resultant deviation score
3. Add them to get SS_{total}

B. $SS_{bet.} = \Sigma n(\overline{X}_{group}-\overline{X}_{total})^2$

1. Subtract the overall mean, $\overline{X}_{total}$ from each group mean.
2. Square each resultant deviation score and multiply it by n, the number of subjects in the group.
3. Sum the values from Step #2 and you have $SS_{bet.}$.

C. $SS_{within} = \Sigma(X-\overline{X}_1)^2 + \Sigma(X-\overline{X}_2)^2 + \ldots + \Sigma(X-\overline{X}_{last})^2$

1. Subtract the Group 1 mean, $\overline{X}_1$, from each score in Group 1.
2. Square each resultant deviation score.
3. Add the deviation scores to get $\Sigma(X-X_1)^2$
4. Repeat the process for all groups ending with the last one.
5. Add the group *SS*'s to get SS_{within}.

* * * * * * * * * * * * *

* * * * * * * * * * * * *

F TABLE MANNERS

As previous mentioned, your book does not include an F probability table. The reason is simple. No problems are included that necessitate use of the table. Also, the table is quite long, containing much more information than needed for t or χ^2. Why? That reason is also simple. F probabilities vary on the basis of <u>two</u> df values, df_{within} and $df_{between}$. Thus, the F table must show critical (α = .05, .01) values for <u>combinations</u> of the two df's. For example, if $df_{bet.} = 2$, there will be a different critical value for each possible value of df_{within} (which is an infinite number!). F tables, being of sound mind, only report values for selected df's (like every 5th one up to 100), but the result is still a pretty long table.

Should you need to consult an F table, do it with enthusiasm and excitement. Using one is easy! Columns (moving across the page) will correspond to different $df_{bet.}$ values and rows (moving down the page and onto other pages) will correspond to different df_{within} values. Find the ones that correspond to your df's (if not found, use the next available <u>lower</u> ones reported). The section where the column and row cross will report critical F values for commonly used α's (.05, .01, etc.). There is nothing to it!

* * * * * * * * * * * * *

FINAL, FINAL POSTTEST

Select one of the alternatives provided in response to the following question:

QUESTION: Now that you have finished your introduction to statistics (as provided in this book), your reaction is one of

a. "At last it's over!" "A horrible experience that should be avoided at all costs!!"

b. "It was bearable, but about as interesting as reading the Phonebook." "Never again, unless absolutely necessary."

c. "Well, it's not as exciting as reading a mystery novel, but the stuff was far more doable than had been expected." "If I had the opportunity to take a more advanced course, I might even consider it."

d. "It's been the event of my life; I am now happier, look better, and have more friends as a result." "My major will be changed immediately to statistical theory or nuclear physics."

e. all of the above.

We hope that if your response was not "d," it was at least "c." Statistics hardly makes for exciting, provocative reading. But, it does include concepts that are becoming increasingly prevalent in our very data-oriented society. If you feel, as a result of this course, that those concepts are learnable, useable, and need not be threatening to the "non-mathematician" (which may include yourself), we have accomplished our major objective. If your answer was "a" or "b," we're sorry - but it's not because we didn't try. You can still put your book to good use, as we hear it makes a tolerable frisbee or seat pad (for short people).

Answer Section

340

ANSWER SECTION

UNIT A

Unit A Review Test

1.	59x	14.	19
2.	-8x	15.	12
3.	-38x	16.	9
4.	17z	17.	9
5.	-26x	18.	201.64
6.	5	19.	1/9
7.	88	20.	87%
8.	$-63w^2$	21.	29
9.	-04.0	22.	10
10.	2.5	23.	285
11.	31	24.	100
12.	26	25.	99
13.	28		

UNIT I

1.1 POSTTEST

1. 10 in.: 9.5 in. - 10.5 in.
   5 in.: 4.5 in. - 5.5 in.

2. 1.372 sec.: 1.3715 - 1.3725
   1.300 sec.: 1.2995 - 1.3005

3. 10.1: 10.05 - 10.15
   11.0: 10.95 - 11.05

1.2.1 POSTTEST

X	f	cf
1.00	3	34
.90	5	31
.80	1	26
.70	2	25
.60	5	23
.50	4	18
.40	2	14
.30	3	12
.20	3	9
.10	4	6
.00	2	2

n = 34

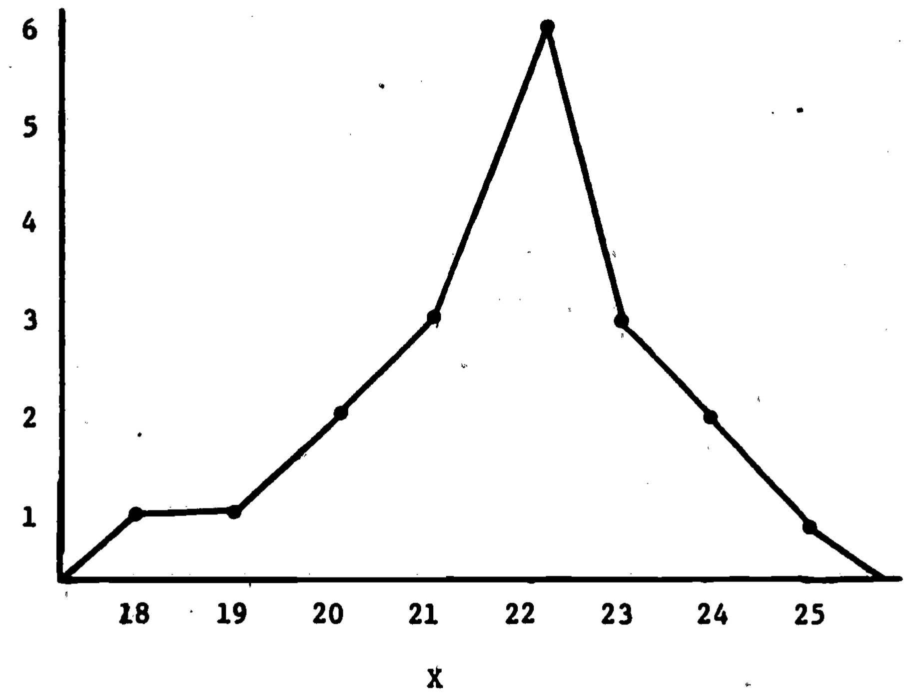
1.2.2 POSTTEST

X	f	cf
40-44	1	33
35-39	5	32
30-34	3	27
25-29	0	24
20-24	2	24
15-19	4	22
10-14	5	18
5- 9	9	13
0- 4	4	4

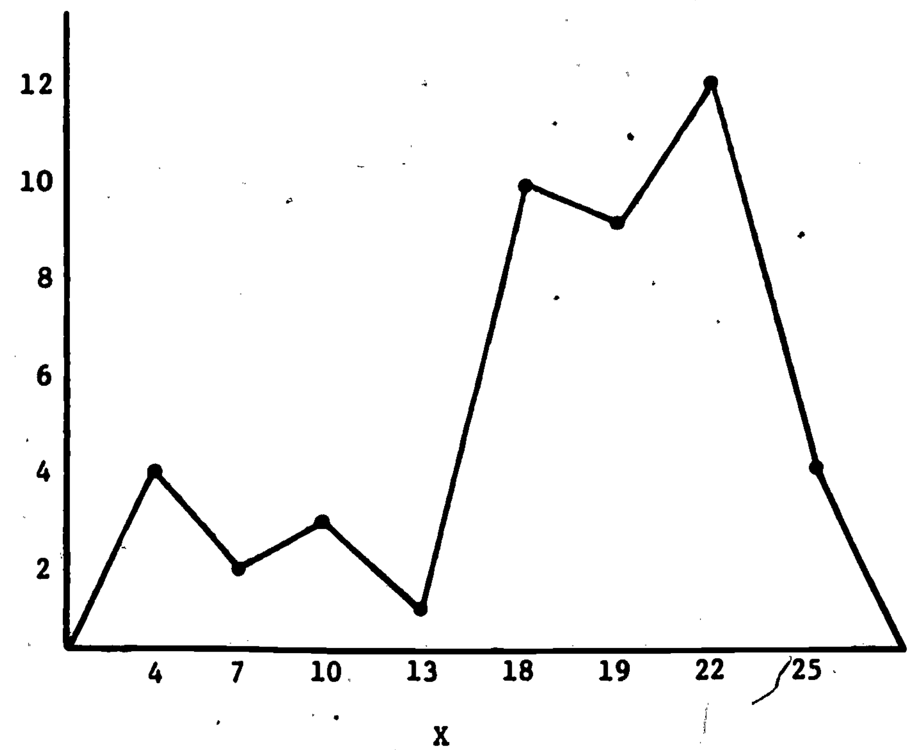
(this is only one of several possibilities)

1.3.1 POSTTEST

a.

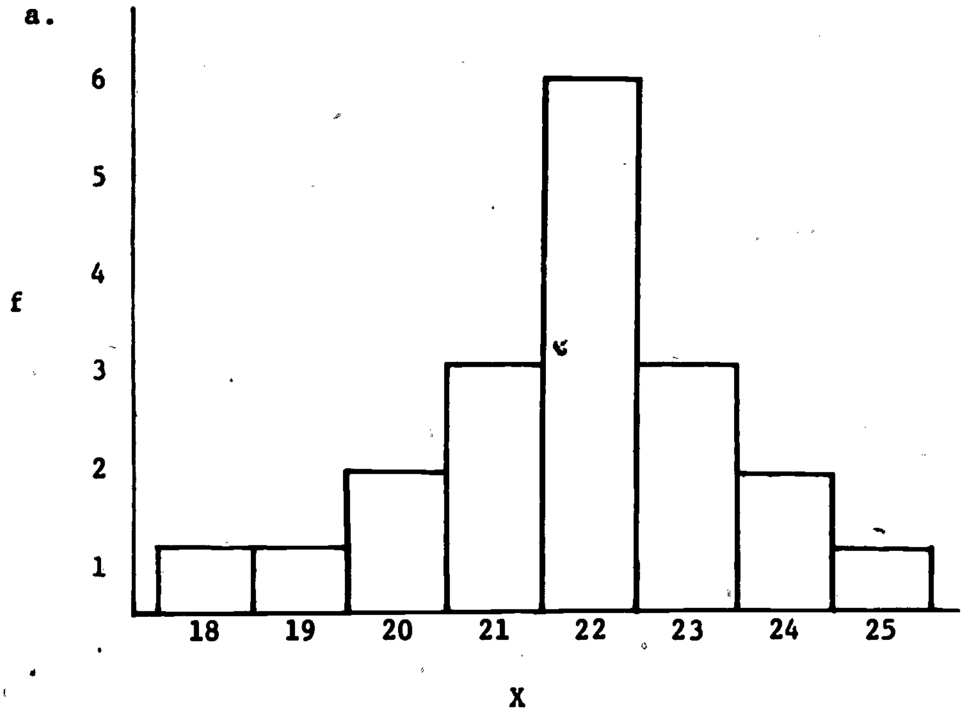


b.

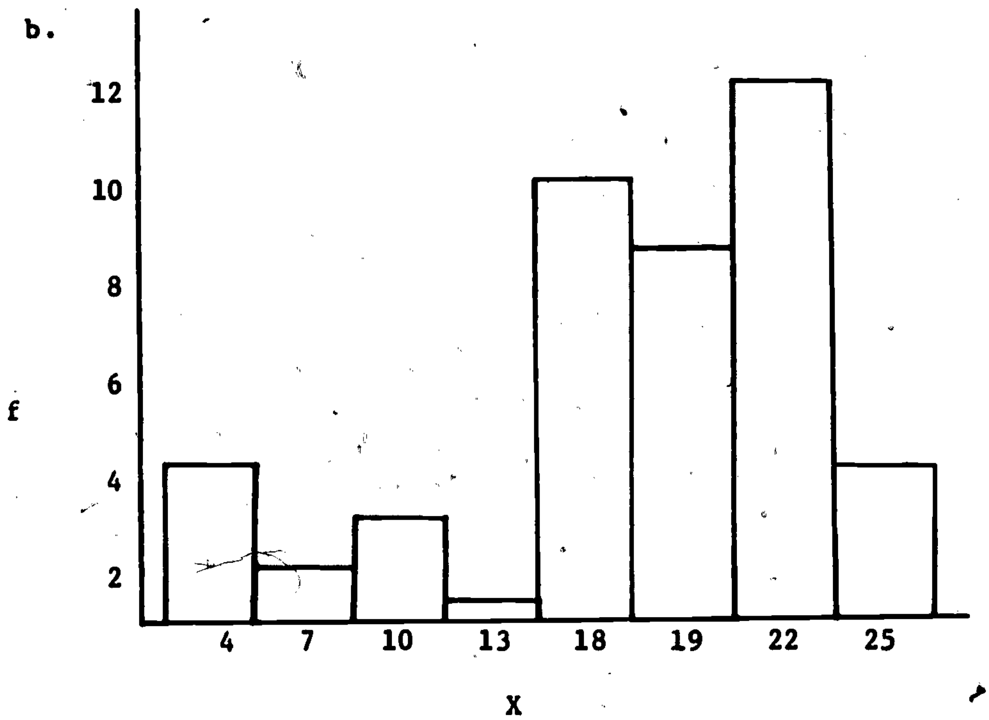


1.3.2 POSTTEST

a.

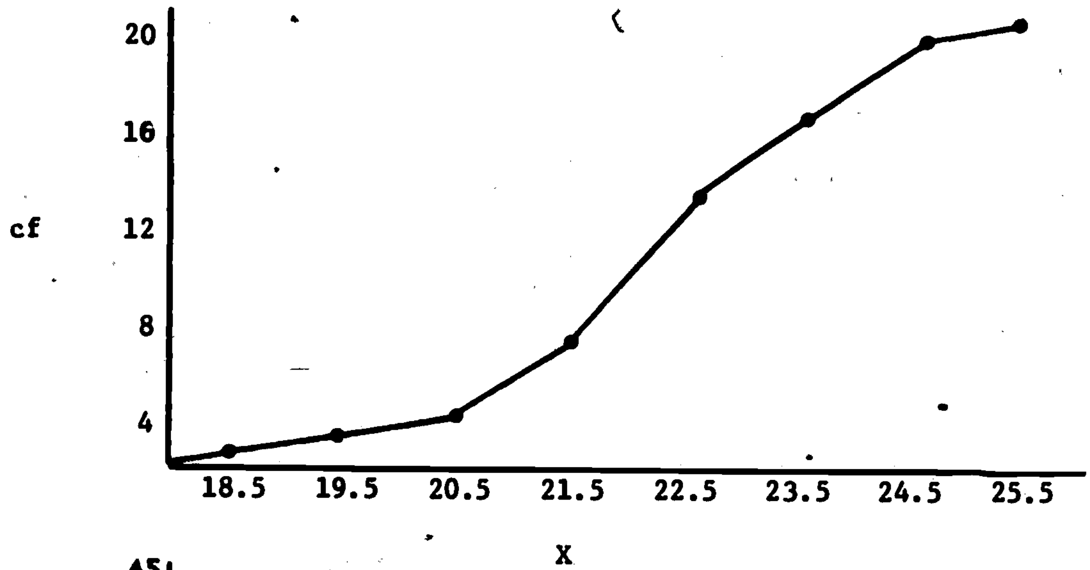


b.

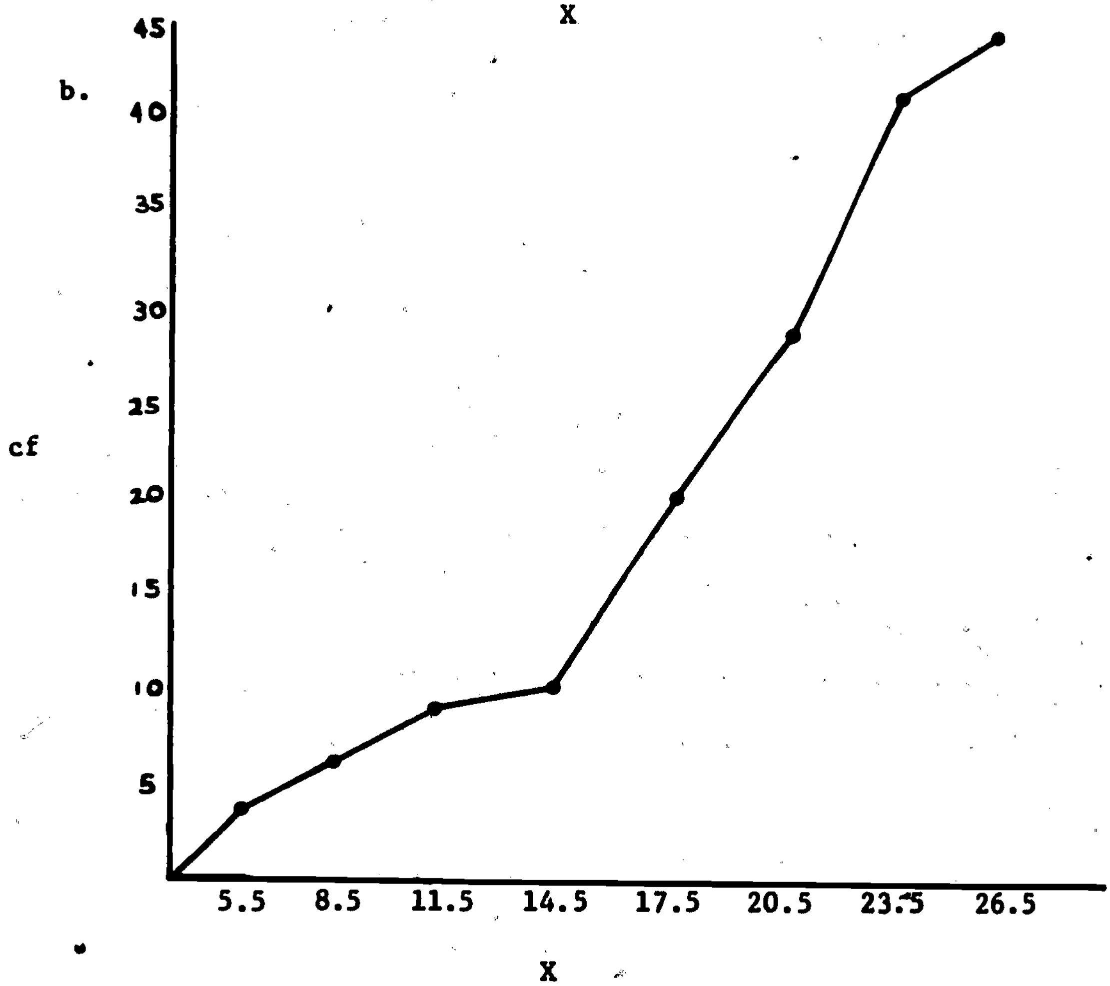


1.3.3 POSTTEST

a.



b.



1.4 POSTTEST

a. bimodal, positively skewed
b. bell-shaped, unimodal, symmetrical
c. unimodal, positively skewed
d. triangular, unimodal, symmetrical
e. bimodal, negatively skewed
f. rectangular, symmetrical
g. bimodal, symmetrical
h. unimodal, negatively skewed
i. rectangular, amodal, symmetrical
j. bimodal (intervals 29 - 31 and 20 - 22), positively skewed

Unit I Review Test

1.

X	f	cf
19	2	20
18	4	18
17	8	14
16	4	6
15	2	2

n = 20

It would be inappropriate to use a grouped frequency distribution since you are only dealing with five scores and grouped distributions are usually reserved for distributions where it is impractical to list every possible score.

2. 17.5 - 18.5
3. 19.5
4. The upper limit of each score would be used to define the horizontal axis (i.e., 15.5, 16.5, 17.5, etc.).
5. Bell-shaped.

UNIT II

2.1 POSTTEST

a. A percentile rank is a number that represents the percentage of scores that fall at or below some specified raw score.

b. When we say that a student performed at P_{89}, we mean that his raw score on the test or measure was equal to or higher than 89% of the remaining raw scores. In other words, this particular student scored higher than (or the same as) 89%.

c. John did relatively better on the English test. Even though his raw score, 96, was higher on the math test, his percentile rank indicates that he only beat 79% of his classmates. On the English test, however, his raw score of 42 was higher than 96% of his classmates' scores.

2.2 POSTTEST

Set 1

a) $\frac{49-48.5}{1} = \frac{cf_x - 21}{9}$

$cf_x = 25.5$

P.R. = 25.5/38 x 100 = 67.1

answer: $P_{67.1}$

b) $\frac{47 - 46.5}{1} = \frac{cf_x - 10}{4}$

$cf_x = 12$

P.R. = 12/38 x 100 = 31.6

answer: $P_{31.6}$

c) $\frac{44 - 43.5}{1} = \frac{cf_x - 2}{2}$

$cf_x = 3$

P.R. = 3/38 x 100 = 7.9

answer: $P_{7.9}$

Set 2

a) $\frac{17 - 14.5}{3} = \frac{cf_x - 19}{4}$

$cf_x = 22.33$

P.R. = 22.33/25 x 100 = 89.3

answer: $P_{89.3}$

b) $\frac{9 - 8.5}{3} = \frac{cf_x - 8}{3}$

$cf_x = 8.5$

P.R. = 8.5/25 x 100 = 34

answer: P_{34}

c) $\frac{7 - 5.5}{3} = \frac{cf_x - 5}{3}$

$cf_x = 6.5$

P.R. = 6.5/25 x 100 = 26

answer: P_{26}

2.3 POSTTEST

Set 1

a. P_{30}:

$.30 = cf_x/N = cf_x/40$

$cf_x = 12$

$$\frac{X - 41.5}{1} = \frac{12 - 7}{6}$$

$X = 42.33$

answer: $P_{30} = 42.33$

b. P_{60}:

$.60 = cf_x/40$

$cf_x = 24$

$$\frac{X - 43.5}{1} = \frac{24 - 21}{10}$$

answer: $P_{60} = 43.8$

c. P_{90}:

$.90 = cf_x/40$

$cf_x = 36$

$$\frac{X - 45.5}{1} = \frac{36 - 35}{3}$$

answer: $P_{90} = 45.83$

Set 2

a. P_{30}:

$.30 = cf_x/N = cf_x/35$

$cf_x = 10.5$

$$\frac{X - 8.5}{3} = \frac{10.5 - 10}{10}$$

$X = 8.65$

answer: $P_{30} = 8.65$

b. P_{60}:

$.60 = cf_x/35$

$cf_x = 21.0$

$$\frac{X - 11.5}{3} = \frac{21-20}{6}$$

answer: $P_{60} = 12$

c. P_{90}:

$.90 = cf_x/35$

$cf_x = 31.50$

$$\frac{X - 17.5}{3} = \frac{31.5 - 31}{4}$$

answer: $P_{90} = 17.88$

2.4 POSTTEST

<u>Central tendency</u> is one of several measures that indicates the most typical value in a set of data.

2.4.1 POSTTEST

1. <u>Mean</u> is the point of balance in a distribution; the deviations below it equal the deviations above it. It is the numerical average of the scores.

2. SET 1: $\overline{X} = 15.7$

   SET 2: $\overline{X} = 78.75$

   SET 3: $\overline{X} = 7.91$

2.4.2 POSTTEST

1. <u>Median</u> is the midway point in a distribution; half of the scores fall above it, and half fall below it.

2. SET 1: Median = 9

   SET 2: Median = 61

   SET 3: Median = 9.67

   SET 4: Median = 25.0

2.4.3 POSTTEST

1. <u>Mode</u> is the most frequently occurring score value.

2. SET 1: Mode = 60 and 58 (distribution is bimodal)

   SET 2: Mode = 41

   SET 3: Mode = 6

2.5 POSTTEST

1.

A. Median (seriously skewed distribution)
B. Mean (slight skew)
C. Median (top interval is open-ended)
D. Median (serious skew)
E. Mean (slight skew)
F. Median (serious skew)

2.

A. Mean # <u>1</u>
Md # <u>2</u>

B. Mean # <u>2</u>
Md # <u>1</u>

C. Mean <u>and</u> Median # <u>2</u>

D. Mean = Median

E. Mean <u>less than</u> Median

F. Mean <u>greater than</u> Median

<u>Unit II Review Test</u>

1.

a. 92.12
b. 84.25
c. $P_{70.6}$
d. P_8
e. 87.3
f. 88.25
g. 92
h. The median is the more meaningful measure because this distribution has a <u>serious</u> negative skew.

2.

a. A
b. C
c. A
d. B
e. C
f. A
g. B

3.

a. A
b. B
c. A
d. B

ERIC

UNIT III

3.1 POSTTEST

Variability refers to the extent to which scores in a distribution bunch up close to a measure of central tendency (usually the mean) or spread out far from a measure of central tendency.

3.2 POSTTEST

1. The range is the distance from the upper limit of the highest score to the lower limit of the lowest score. Or more simply: the range is the distance from the bottom of the distribution to the top of the distribution.

2. Set 1: range = 43
   Set 2: range = 11
   Set 3: range = 25

3.3 POSTTEST

1. The variance is the average of the squared deviations around the mean in a set of data.

2. Set 1: $\sigma^2 = \dfrac{3540 - \dfrac{(166)^2}{8}}{8} = \dfrac{3540 - 3444.5}{8}$

$= \dfrac{95.5}{8}$

$= 11.94$

Set 2: $\sigma^2 = \dfrac{\Sigma fX^2 - \dfrac{(\Sigma fX)^2}{N}}{N} = \dfrac{30196 - \dfrac{(830)^2}{23}}{23}$

$= \dfrac{30196 - 29952.17}{23} = \dfrac{243.83}{23} = 10.6$

*Set 3:

$$s^2 = \frac{\Sigma fX'^2 - \frac{(\Sigma fX')^2}{n}}{n - 1} = \frac{960 - \frac{(90)^2}{12}}{11}$$

$$= \frac{960 - 675}{11} = 25.91$$

*Note change in denominator as a result of using sample data.

3. The variances would remain unchanged.

4. Each new variance would be 3^2 times as large as the originals:

   Set 1: 107.46
   Set 2: 95.4
   Set 3: 233.19

3.4 POSTTEST

1. The standard deviation is the square root of the variance; it is the square root of the average of the squared deviations around the mean.

2. Set 1: Use n - 1 in denominator.

$$s = \sqrt{\frac{812 - \frac{(66)^2}{7}}{6}} = \sqrt{\frac{812 - 622.29}{6}} = \sqrt{31.62}$$

$$= 5.62$$

Set 2:

$$\sigma = \sqrt{\frac{\Sigma fX^2 - \frac{(\Sigma fX)^2}{N}}{N}} = \sqrt{\frac{1379 - \frac{(179)^2}{24}}{24}}$$

$$= \sqrt{1.83} = 1.35$$

Set 3:

$$\sigma = \sqrt{\frac{\Sigma fX'^2 - \frac{(\Sigma fX')^2}{N}}{N}} = \sqrt{\frac{7460 - \frac{(422)^2}{26}}{26}}$$

$$= \sqrt{\frac{7460 - 6849.38}{26}} = \sqrt{23.49} = 4.85$$

3. The new standard deviations would be exactly the same as the originals.

Unit III Review Test

1. 419
2. 26
3. 16.12
4. 6907
5. 175,561
6. 5.95
7. 6.19
8. 2.44
9. 2.49
10. 11
11. 2.44
12. 1.22
13. 1.49
14. 14.94
15. 55.71
16. 6.19

UNIT IV

4.1 POSTTEST

1. A z score indicates the number of standard deviations a raw score is above or below the mean.

2. A. $X = .91$, $z = 3.67$; $.87 = 2.33$; $.8 = 0.00$; $.72 = -2.67$

   B. $X = 100$, $z = 1.14$; $99 = 1.00$; $90 = -.29$; $82 = -1.43$; $78 = -2.00$

   C. $X = 1.5$, $z = 1.50$; $-2 = -2.00$; $0 = 0.00$; $-.4 = -.40$

4.2 POSTTEST

1. For $z = 2.2$: $X = z\sigma + \bar{X} = (2.2)(15) + 100 = 33 + 100 = 133$

   $z = 1.0$: $X = (1)(15) + 100 = 115$

   $z = -2.0$: $X = (-2)(15) + 100 = -30 + 100 = 70$

   $z = 0.00$: $X = (0)(15) + 100 = 100$

2. For $z = 2.00$: $X = (2)(.3) + 1.2 = .6 + 1.2 = 1.8$

   $z = 0.30$: $X = (.3)(.3) + 1.2 = .09 + 1.2 = 1.29$

   $z = -.40$: $X = (-.4)(.3) + 1.2 = -.12 + 1.2 = 1.08$

   $z = -1.3$: $X = (-1.3)(.3) + 1.2 = -.39 + 1.2 = .81$

3. For $z = 3.00$: $X = (3)(4) + 50 = 62$

   $z = 2.00$: $X = (2)(4) + 50 = 8 + 50 = 58$

   $z = -2.0$: $X = (-2)(4) + 50 = -8 + 50 = 42$

   $z = -2.8$: $X = (-2.8)(4) + 50 = -11.2 + 50 = 38.8$

4.3 POSTTEST

1. a. $T = 10z + 50 = 10(2.00) + 50 = 20 + 50 = 70$
   b. $T = 10(0) + 50 = 50$
   c. $T = 10(-1.56) + 50 = -15.6 + 50 = 34.4 = 34$
   d. $T = 10(.72) + 50 = 7.2 + 50 = 57.2 = 57$
   e. $T = 10(-2.40) + 50 = -24 + 50 = 26$

2. a. For $X = 14$: $z = \frac{14 - 22}{8} = \frac{-8}{8} = -1.00$; $T = 10(-1) + 50 = 40$

   b. For $X = 22$: $z = 0$; $T = 50$

   c. For $X = 26$: $z = .5$; $T = 55$

   d. For $X = 16$: $z = -.75$; $T = 43$

   e. For $X = 34$: $z = 1.5$; $T = 65$

3. In order to find the raw scores, you must first convert T to z, and then convert z to X.

   a. For $T = 50$: $z = \frac{T - 50}{10}$ (this is the same equation as $T = 10z + 50$)

   $$z = \frac{50 - 50}{10} = 0.00$$

   $$X = z\sigma + \bar{X}$$

   $$X = (0)(7) + 68$$

   $$X = 68$$

   b. For $T = 40$: $z = \frac{40 - 50}{10} = -1.00$

   $$X = (-1)(7) + 68 = -7 + 68 = 61$$

   c. For $T = 35$: $z = \frac{35 - 50}{10} = -15/10 = -1.5$

   $$X = (-1.5)(7) + 68 = -10.5 + 68 = 57.5$$

ERIC

d. For $\underline{T} = 62$: $\underline{z} = \frac{62 - 50}{10} = 12/10 = 1.20$

$X = (1.2)(7) + 68 = 8.4 + 68 = 76.4$

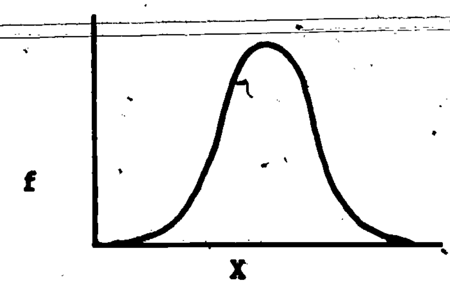
e. For $\underline{T} = 47$: $\underline{z} = \frac{47 - 50}{10} = -3/10 = -.30$

$X = (-.3)(7) + 68 = -2.1 + 68 = 65.9$

4.4 POSTTEST

1. The normal curve resembles a bell where the majority of frequencies are in the middle area with the smaller number of frequencies being at each end of the distribution.

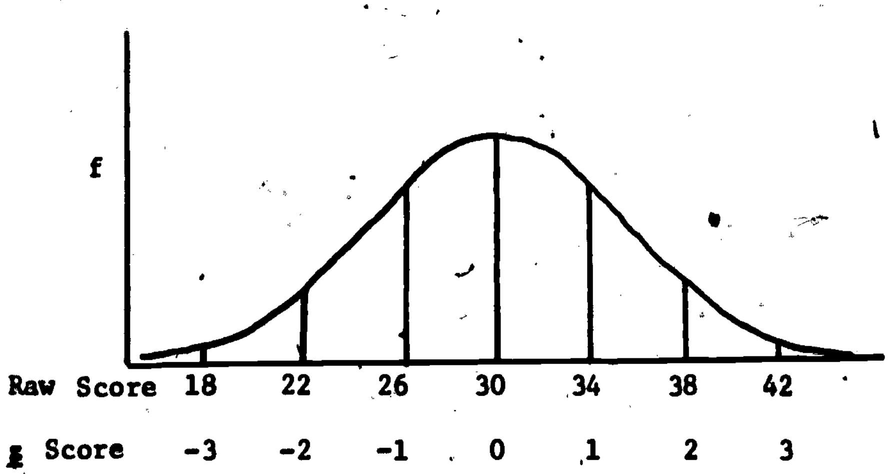
2. a.



b.
1. true
2. false
3. false
4. false
5. true
6. true
7. false
8. can't answer (curve refers to hamburgers not cheeseburgers)

4.5 POSTTEST

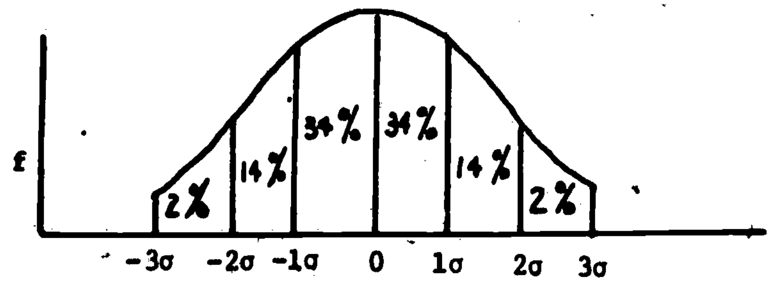
1.



2. a. If we plot the distribution, the heights of the curve at $+1\sigma$ and -1σ will be exactly equal. This is because the normal curve is symmetrical.

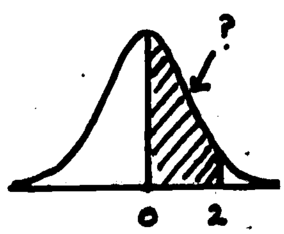
   b. 1. 98
      2. Neither since they are both equidistant from the mean.
      3. Yes, it could be a possible score since you don't know how large the σ around the mean is and σ <u>usually</u> spread out 3σ around the mean.

4.6 POSTTEST

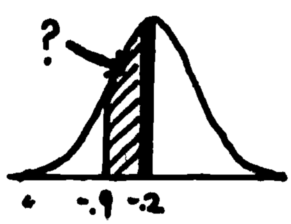


4.7 POSTTEST

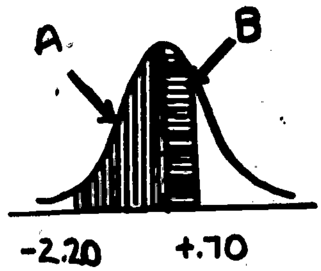
1. 47.72%



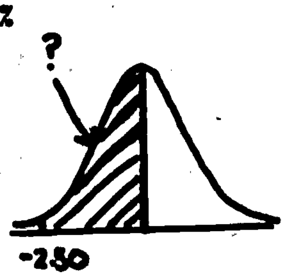
2. 31.59 - 7.93 = 23.66%



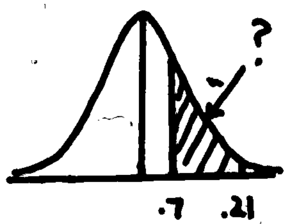
3. 48.61 + 25.8 = 74.41%



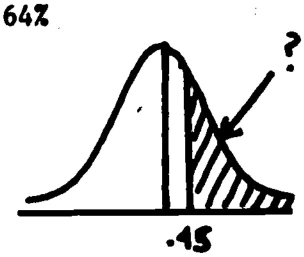
4. 49.38%



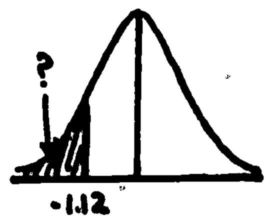
5. 48.21 - 25.8 = 22.41%



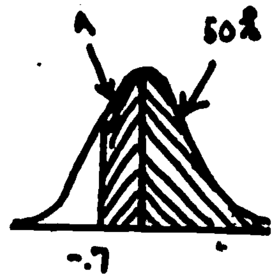
6. 32.64%



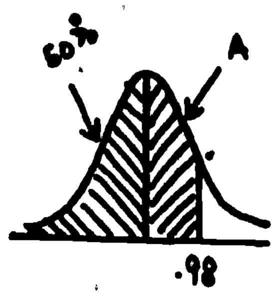
7. 13.14%



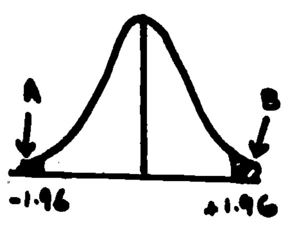
8. 75.8%



9. 83.65%

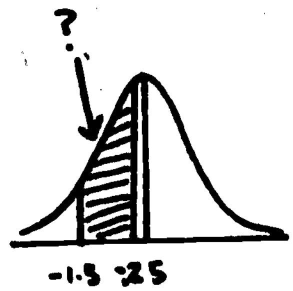


10. (BONUS) 2.5% + 2.5% = 5%

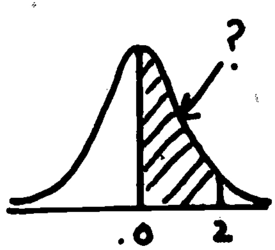


4.8 POSTTEST

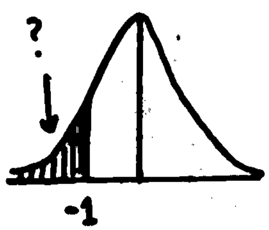
1. a. 66 and 71, $\underline{z}$ = -1.5 to -.25 = 43.32 - 9.87 = 33.45%



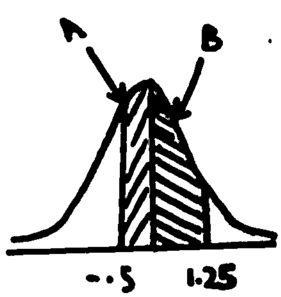
b. 72 and 80, $\underline{z}$ = 0 to 2 = 47.72%



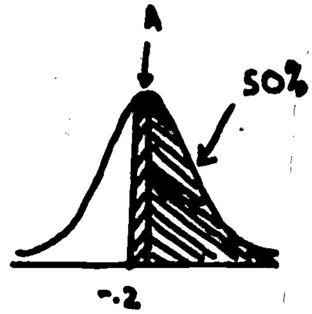
c. below 68, $\underline{z}$ = -1 = 50.00 - 34.13 = 15.87%



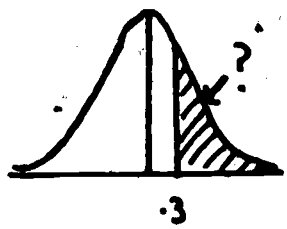
2. a. 450 and 625, $\underline{z}$ = -.5 to 1.25 = 19.15 + 39.44 = 58.59%



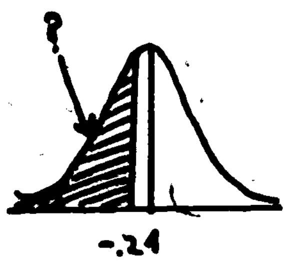
2. b. above 480, $\underline{z}$ = -.2 = 50.00 + 7.93 = 57.93%



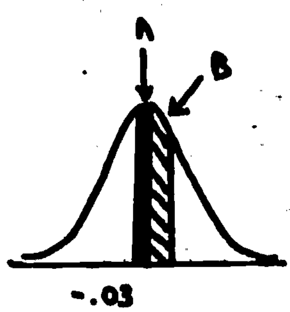
c. above 530, $\underline{z}$ = .3 = 50.00 - 11.79 = 38.21%



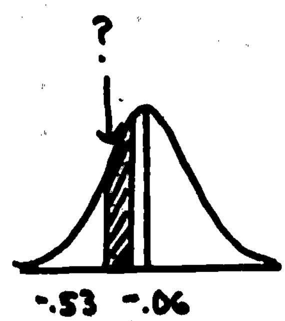
3. a. below 21.3, $\underline{z}$ = -.24 = 50.00 - 9.48 = 40.52%



b. 22 and 24, $\underline{z}$ = -.03 to .56 = 1.20 + 21.23 = 22.43%



3. c. 20.3 and 21.9, z = −.53 to −.06 = 20.19 − 2.39 = 17.80%



4.9 POSTTEST

1. a. For z = −2.4, P.R. = 0.82
   b. For z = −1.2, P.R. = 11.51
   c. For z = 0, P.R. = 50
   d. For z = .9, P.R. = 81.59
   e. For z = 2.3, P.R. = 98.93

2. a. For P.R. = 15, z = −1.04
   b. For P.R. = 25, z = −.67
   c. For P.R. = 40, z = −.25
   d. For P.R. = 60, z = .25
   e. For P.R. = 85, z = 1.04

3. a. For X = 14, P.R. = 90.82
   b. For X = 5, P.R. = 4.75
   c. For X = 9, P.R. = 37.07

4. a. For P.R. = 50, X = 10
   b. For P.R. = 80, X = 12.52
   c. For P.R. = 43, X = 9.46

<u>Unit IV Review Test</u>

1. a. 36.7%
   b. 98.9%
   c. 18.96%
   d. 14.98% of 1000 = 150 (approx.)
   e. 9.19%
   f. 26.43
   g. 3.01%
   h. 69.64
   i. 62.47%

2. a. 77.34
   b. 27.43
   c. 88.49
   d. 99.74
   e. 50.00
   f. 53.98

UNIT V

5.1 POSTTEST

1. Correlation is a method or technique used to determine the relation between two variables.

2. a. number of digits: (a constant) cannot be used
   b. number of freckles: (a variable) can be used
   c. number of illnesses: (a variable) can be used
   d. number of toes: (a constant, assuming sloths are physically normal) can't be used
   e. number of friends: (a variable) can be used

3. __a__ 1. height

   __c__ 2. absences (the more absences, the lower the grade)

   __a__ 3. pulse rate

   __b__ 4. IQ (the higher the IQ, the higher the grade)

   __c__ 5. TV watching (the more TV, the lower the grade)

   __b__ 6. books read (the more books read, the higher the grade)

4. Answer is pair #3, numbers of bars and schools in different cities.

   Pair #1, age of automobile and miles per gallon, should show <u>negative</u> correlation, i.e., higher the age, the <u>lower</u> the miles per gallon (i.e., efficiency).

   Pair #2, distance north and average daily temperature for cities, should also correlate negatively, i.e., the <u>farther</u> north one is, the colder it gets and the <u>lower</u> the temperature.

   Pair #3 should correlate positively due to <u>population of city</u> as an intervening variable. The higher the population, the more bars; the higher the population, the more schools. Thus, cities with many bars will <u>tend</u> to have many schools and vice versa.

5.2 POSTTEST

1. There are numerous possible procedures that could be used. If her students were cooperative, she could have them fill out two questionnaires, one designed to determine how many cigarettes they smoke, and the other designed to assess their attitudes toward school (perhaps a 10-point scale). Thus, her units of analysis would be <u>students</u>. The variables to be examined would be <u>number of cigarettes smoked per day</u> and <u>attitude toward school</u> (with a score of "1" indicating a very poor attitude and a score of "10" indicating a very favorable attitude). After measuring each student in her sample

on both of these variables, she would then tabulate the data by listing each student's cigarette score and attitude score next to his/her name. Finally, she may wish to construct a scatter-plot to obtain an immediate overall impression of the obtained relation.

2. There does appear to be a relation between height and point production for this particular sample. Generally speaking, the taller players tend to score more points than do the shorter players. That is, high scores on one variable tend to be associated with high scores on the other, and low scores on one with low scores on the other. Thus, the relation would be positive.

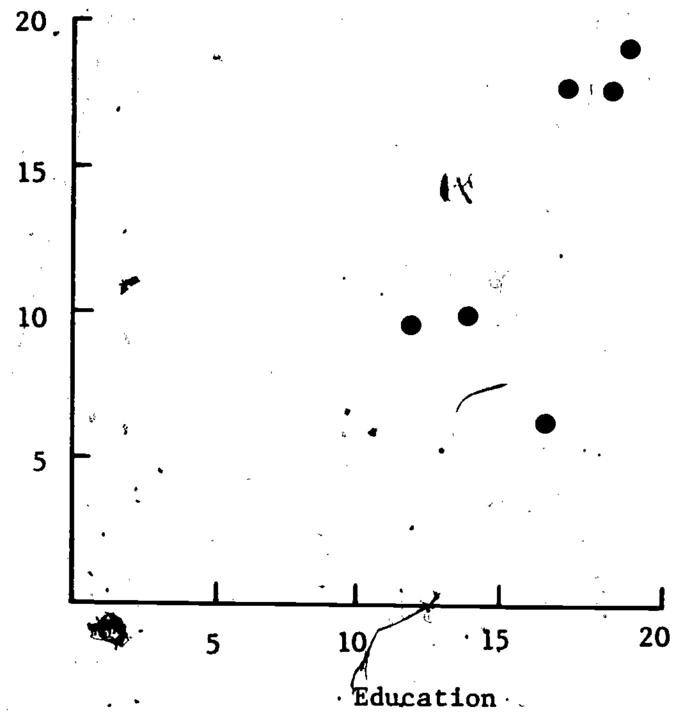
3. The units of analyses would be "dogs." Each dog included in the sample would be measured with respect to aggression tendencies and amount of punishment received. The latter measure might be obtained, of course, through interviews with the dogs' masters (as the dogs themselves may be reluctant to admit anything).

4. This one might be a little difficult to understand at first. The appropriate units would have to be "days." One could randomly select a sample of "days" from a particular calendar year. For each day, we could check existing weather information to determine its average temperature, and medical records to determine the number of fainting spells which occurred. Our measures would, of course, have to be restricted to a particular area location (e.g., New York City, Pittsburgh, our home town, etc.)

   Now, if the interest shifted to fainting by people on a particular day, the units of concern would be people. That is, people could provide a measure of fainting (0 = no; 1 = yes; 2 = more than one time) and a measure of body temperature (98.8, 104, etc.). Do you see the difference between this research question and the one above? The first asks what daily temperature does for daily fainting; the second asks what people's temperature does for people's fainting.

5.3 POSTTEST

1.



b. There appears to be a moderately strong positive relation (the points slope upward, but cannot all be fitted on a straight line).

c. If the new subject's scores were included in the analysis, the relation would be weakened. This is because his scores run contrary to the overall tendency for high salaries to be associated with additional years of education. If these scores were to be represented by a point on the scatterplot, the result would be a reduction in slope and more deviation from a straight line pattern.

2. Plot A: strongest relation because more points could be fitted on a straight line.

3. Plot A: very strong positive relation
Plot B: moderate-strong negative relation
Plot C: weak-moderately strong positive relation
Plot D: very strong negative relation
Plot E: no relation

5.4 POSTTEST

1. The correlation coefficient is a numerical value that clearly specifies the strength and direction of the relation between variables. It ranges from +1.00, indicating a perfect positive relation, to -1.00, indicating a perfect negative relation. The middle value is 0.00, indicating no relation. The sign of the coefficient indicates the direction of the relation, whereas the absolute value of the numerical component indicates the strength of the relation.

2. a. Plot #4
   b. Plot #3
   c. Plot #1
   d. Plot #5
   e. Plot #2

3. a. grades and dating: imperfect negative
   grades and absences: also imperfect negative

   b. absences are the better predictor since their relation to grades is stronger

   c. A person with many dates (high score) should have relatively low grades (low score), as suggested by the direction of the obtained correlation. This would not necessarily apply, however, to all students because the relation is imperfect, suggesting that there are exceptions to the overall tendency. Thus, we can make an educated guess but not be totally certain about John.

   d. Plot B would probably represent the dating-grades relation since it represents a weaker relation than that depicted in Plot A. Plot A, of course, would be more likely to describe the absences-grades relation.

4. a. years of education and income: imperfect positive
   distance from city and income: imperfect negative

   b. <u>distance</u> would be the better predictor since its relation to income is stronger (even though the direction of the relation is negative).

   c. people who have many years of education would tend to have high incomes. People who live far from large cities (high scores on distance variable) would tend to be low in income. These would not necessarily apply to all people: the relations are not particularly strong and thus there are likely to be many exceptions.

   d. the distance-income relation would more closely approximate a straight line because it is stronger. Strength has nothing to do with direction of sign in a coefficient or slope in a scatterplot.

5.5 POSTTEST

1. a. When variables are positively related, positive z scores on the X variable will tend to be associated with positive z scores on Y. The result will be positive cross-multiplication products, $z_X z_Y$. Conversely, negative z_X scores will tend to be associated with negative z_Y scores, also producing positive cross-products.

   b. When variables are inversely related positive z_X scores will tend to be associated with negative z_Y scores, and vice versa. Result will be mostly negative cross-products, a negative sum of cross-products, and finally, a negative r.

   c. When variables are unrelated, some positive z_X scores will be associated with positive z_Y scores, resulting in <u>positive</u> cross-products. Others, however, will be associated with negative z_Y scores, producing <u>negative</u> cross-products. The positive and negative cross-products will tend to cancel each other out, producing a low sum and a low r.

2.

Student	z_X	z_Y	$z_X z_Y$
A	-1.50	-1.20	+1.80
B	-1.20	-1.50	+1.80
C	- .80	- .40	+ .32
D	- .40	+ .80	- .32
E	.00	.00	.00
F	+ .40	- .80	- .32
G	+ .80	+ .40	+ .32
H	+1.20	+1.50	+1.80
I	+1.50	+1.20	+1.80
		$\Sigma z_X z_Y$	= 7.20

$$r = \frac{7.20}{9} = +.80$$

3\. **Student**

Student	X	Y	X^2	Y^2	XY
J	3	5	9	25	15
K	1	2	1	4	2
L	2	5	4	25	10
M	4	3	16	9	12
N	1	1	1	1	1
	$\Sigma X = 11$	$\Sigma Y = 16$	31	64	40

$$r = \frac{(5)\ (40) - (11)\ (16)}{\sqrt{(5)\ (31) - 121 \quad 5(64) - 256)}}$$

$$r = \frac{24}{\sqrt{34}\ \sqrt{64}} = \frac{24}{46.6} = +.52$$

4\.

Contestant	Congeniality	Muscle Tone	D	D^2
A	1	6	-5	25
B	3	4	-1	1
C	6	5	1	1
D	4	3	1	1
E	7	2	5	25
F	2	7	-5	25
G	5	1	4	16
			ΣD^2	= 94

$$r_s = 1 - \frac{6(94)}{7(49-1)} = 1 - \frac{564}{336}$$

$$= 1 - 1.679 = -.68$$

Results imply that congeniality and muscle tone ranks are negatively correlated. The contestants with smaller muscles aren't as friendly (maybe they don't have to be!)

Unit V Review Test

A. _c_ 1. _e_ 6.

_e_ 2. _d_ 7.

_b_ 3. _c_ 8.

_d_ 4. _b_ 9.

_d_ 5. _a_ 10.

6.1 POSTTEST

1. Probability refers to the relative frequency with which an event is expected to occur.

2. 3/5 or 60%

3. 10/60 or 1/6 or .167

6.2 POSTTEST

1. a. mutually exclusive
   b. nonmutually exclusive
   c. nonmutually exclusive
   d. mutually exclusive
   e. mutually exclusive

2. a. $\frac{5}{25} + \frac{3}{25} = \frac{8}{25}$ or 32%

   b. $\frac{22}{210} + \frac{60}{210} - \frac{2}{210} = \frac{80}{210}$ or 38%

   c. $\frac{8}{65} + \frac{10}{65} - \frac{8}{65} = \frac{10}{65}$ or 15%

   d. $\frac{3}{15} + \frac{2}{15} = \frac{5}{15}$ or 33%

3. P(both) is subtracted because the individuals involved in one set of events are also involved in the other set. If you don't subtract P(both), they get counted twice and probability is distorted.

6.3 POSTTEST

1. a. independent
   b. dependent
   c. dependent
   d. independent

2. a. $\frac{18}{20} \bullet \frac{17}{19} = \frac{306}{380}$ or 81% dependent trials!

   b. $\frac{18}{20} \bullet \frac{18}{20} = \frac{324}{400}$ or 81% independent trials!

3. a $\frac{1}{4} \bullet \frac{1}{4} = \frac{1}{16}$ or 6%

   b. $\frac{1}{4} \bullet \frac{1}{4} \bullet \frac{1}{4} = \frac{1}{64}$ or 2%

   c. $\left(\frac{1}{4}\right)^4 = \frac{1}{256}$ or .39%

4\. a. $\frac{13}{52} \cdot \frac{13}{52} = \frac{169}{2704} = \frac{1}{16}$

b. $\frac{4}{52} \cdot \frac{4}{51} = \frac{16}{2652} = \frac{4}{663}$

c. $\frac{13}{52} \cdot \frac{12}{51} = \frac{156}{2652} = \frac{13}{221} = \frac{1}{17}$

d. $\frac{8}{52} \cdot \frac{8}{51} = \frac{64}{2652} = \frac{16}{663}$

Unit VI Review Test

1\. a. 27/32 or 84%

b. 5/32 or 16%

c. 32/32 or 100%

2\. a. 15/127 or 12%

b. 20/127 or 16%

c. 35/127 or 28%

d. $\frac{15}{127} \cdot \frac{14}{126} = \frac{210}{16002} = \frac{105}{8001}$ dependent trials!

e. $\frac{15}{127} \cdot \frac{20}{126} = \frac{300}{16002} = \frac{150}{8001}$

3\. a. Mutually exclusive events refer to outcomes that cannot occur together, e.g., drawing one card and getting a queen and a king. Nonmutually exclusive events are outcomes that can occur together, e.g., drawing one card and getting a queen and a diamond.

b. Independent trials refer to the case where outcomes on one trial have no influence on the outcomes in another trial, e.g., drawing a queen from one deck (Trial 1) and drawing a queen from another deck (Trial 2).

Dependent trials refer to the case where outcomes on one trial influence what occurs on another trial, e.g., drawing a queen from a deck (Trial 1), keeping it, and trying to draw another queen from the remaining cards (Trial 2).

4\. a. 5/316 or 1.6%

b. $\frac{5}{316} \cdot \frac{4}{315} = \frac{20}{99540} = \frac{1}{4977}$

c. 311/316 or 98.42%

d. $\frac{311}{316} \cdot \frac{310}{315} = \frac{96410}{99540}$ or about 97%

e. $\frac{5}{316} \cdot \frac{5}{316} = \frac{25}{99856}$

UNIT VII

7.1 POSTTEST

1. A random sample is representative of the population from which it is selected. Random sampling involves insuring that every element (person or thing) in the population of interest has an equal chance of being selected.

2. Statisticians use samples to estimate population parameters since whole populations usually cannot be tested. A random sample helps insure that the estimate (statistic computed from the sample) will be based on representative data. A random sample is likely to provide a close estimate of the population parameters; a biased (non-random) sample will provide, of course, a biased estimate. Also, statistical procedures developed for hypothesis testing require that the samples employed are random.

3. The most straightforward procedure would be to write the names of all 300 Ss (subjects) on separate pieces of paper. Place all the names in a container, mix them up, and without looking into the container, select the 50 names one-by-one. The result will be a random sample of 50 individuals because each name had an equal chance of being selected. A more efficient procedure would be to use a random numbers table of the type found in almost every statistics textbook.

4. This one is a little tricky. The resultant sample will be random, but not for the population of male citizens who reside in Catatonia, Nevada. It is true that random selection procedures were employed (i.e., the use of a random numbers table). But, did every male citizen meeting the requirements have an equal chance of being selected? The answer is "no," because the phonebook was used to define the population. What about individuals with unlisted numbers? What about individuals who don't have phones? What about individuals who recently moved to the area and are not listed, as yet, in the phonebook? Hopefully, all would agree that the resultant sample of 50 is random with respect to the population of males whose names are listed in the Catatonia phone directory; but the sample is biased with respect to the total population of males over 18 who reside in Catatonia.

7.2 POSTTEST

1. Statistics are indices that are derived from samples; Parameters are indices derived from populations. (Samples yield statistics and populations yield parameters.) Statistics are used as estimates of parameters. Greek letters refer to populations and Roman letters refer to samples.

2. Sampling fluctuation occurs where there is variation between representative samples for a given population. Samples allow for probabilistic estimations because only a portion of the entire population is dealt with and, therefore, is subject to chance variation.

3. Generally speaking, a Sampling Distribution represents the probabilities of obtaining values for a given statistic computed for each of an infinite number of samples that are the same size and that all represent the same population. The three major factors determining its properties are the type of statistic, the size of the sample, and the nature of the population.

7.3 POSTTEST

1. a. Unbiased
   b. Unbiased
   c. Biased, underestimate
   d. Unbiased
   e. Unbiased
   f. Biased, underestimate

2. The larger the size of the sample, the smaller the amount of bias and vice versa - why? Large samples are closer to representing the population, and populations yield true values.

7.4 POSTTEST

1. If the population from which sampling takes place is normally distributed, the Sampling Distribution of means will also be normally distributed. If the population scores are not normally distributed, the form of the Sampling Distribution of means will approach normality as sample size increases.

2. The farther the population is from "normality" the higher the $\underline{n}$ needed to ensure normality in the Sampling Distribution; e.g., a seriously skewed population necessitates a larger sample $\underline{n}$ to ensure normality than does a moderately skewed population.

3. a. 2
   b. 3
   c. 5
   d. 4
   e. 6
   f. 7

4. $E(\overline{X}) = u_X$. That is, that if $\overline{X}$'s are computed for an infinite number of random samples, the mean of all the $\overline{X}$'s will be equal to the mean (u_X) score of the population from which the samples are selected. Sample size ($\underline{n}$) <u>does not</u> influence this relation. The means of all samples of the same size (regardless of their absolute size) would still average out to be the population value.

5. σ refers to the standard deviation of all raw scores in the <u>population</u>; $\sigma_{\overline{X}}$ refers to the standard deviation of all scores (actually sample means, $\overline{X}$'s) in the Sampling Distribution; it expresses the variability in the Sampling Distribution. As sample size increases, the Standard Error of the Sampling Distribution decreases.

6. a. $\underline{n} = 4$, $\sigma = 20$, $u_X = 70$

   Expected Value = 70 $[E(\overline{X}) = u_X]$

   $$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{\underline{n}}} \quad ; \quad \sigma_{\overline{X}} = \frac{20}{\sqrt{4}},$$

   $\sigma_{\overline{X}} = 10$

   b. $\underline{n} = 16$, $\sigma = 20$, $u_X = 70$

   Expected Value = 70 $[E(\overline{X}) = u_X]$

   $$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{\underline{n}}} \quad ; \quad \sigma_{\overline{X}} = \frac{20}{\sqrt{16}}$$

   $\sigma_{\overline{X}} = 5$

c. $n = 64$, $\sigma = 20$, $u_x = 70$

Expected Value = 70 $(E(\bar{X}) = u_x)$

$\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$; $\sigma_{\bar{X}} = \dfrac{20}{\sqrt{64}}$

$\sigma_{\bar{X}} = 2.5$

7.5 POSTTEST

1.a. greater than 26: In this Sampling Distribution, $E(\bar{X}) = 25$, $\sigma_{\bar{X}} = 1$. Thus, $z = \dfrac{26 - 25}{1} = +1.00$

Using the table, the percentage of scores greater than z = +1.00 (50.00 - 34.13 = 15.87) = 15.87%

b. between 24 and 26: For 24, $z = \dfrac{24 - 25}{1} = -1.00$

For 26, z = +1.00 (see above)

Using the table, the percentage of scores between +1z and -1z (34.13 + 34.13) = 68.26%

c. between 23.5 and 24: For 23.5, $z = \dfrac{23.5 - 25}{1} = -1.50$

For 24, z = -1.00 (see above)

Using the table, the percentage of scores between -1.5z and -1z (43.32 - 34.13) = 9.19%

2. If 4 cases are used $E(\bar{X})$ will still = 25, but $\sigma_{\bar{X}}$ is now = 2

a. between 26 and 28: For 26, $z = \dfrac{26 - 25}{2} = +.50$

For 28, $z = \dfrac{28 - 25}{2} = +1.50$

Using the table, the % of scores between +.5 and +1.5z = (43.32 - 19.15) 24.17%

b. less than 23: For 23, $z = \dfrac{23 - 25}{2} = -1.00$

on table, % less than -1z (50.00 - 34.13) = 15.87%

c. between 24 and 26: For 24, z = -.50

For 26, z = +.50

on table, % between +.5z and -.5z (19.15 + 19.15) = 38.30%

Unit VII Review Test

1.a. A random sample would insure that every member of the population has an equal chance of being selected, therefore increasing the chances of accurately representing the whole population involved.

b. One procedure would be to write the scores of all 1000 dogs down on separate slips of paper. Place them in a container, mix them up thoroughly, and then pick out the slips one-by-one until the desired sample quantity (or size) has been obtained.

c. The larger your sample, the more reliable your estimate of population parameters.

2.a. Unbiased
b. Unbiased
c. Biased
d. Biased
e. Unbiased
f. Unbiased
g. Biased

3. A Sampling Distribution represents the probabilities of obtaining values for a given statistic that is selected from an infinite number of samples that are the same size and that all represent the same population.

4.a. The statistic: Which statistic are we interested in? The mean will have one type of Sampling Distribution, and the variance, for example, will have another.

b. Sample size: The larger your sample size, the more accurate your estimation.

c. Population: Sampling Distributions are clearly influenced by the population from which the samples are randomly selected. They will differ on the basis of population form, central tendency, and variability.

5.a. Its form would be a normal distribution because the population is normally distributed.

b. $E(\bar{X}) = \mu_X$

$E(\bar{X}) = 8.5$

c. $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$; $\sigma_{\bar{X}} = \frac{2.3}{\sqrt{16}}$; $\sigma_{\bar{X}} = \frac{2.3}{4} = 0.575 = .58$

6. The Standard Error (c) would change:

$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$; $\sigma_{\bar{X}} = \frac{2.3}{\sqrt{25}} = 0.46$

7.a. greater than 9.0: In this Sampling Distribution, $E(\bar{X}) = 8.5$, $\sigma_{\bar{X}} = 1.15$. Thus, $z = \frac{9.0 - 8.5}{1.15} = +.43$.

Using the table, the % of scores greater than z = +.43 (50.00 - 16.64) = 33.36%

b. less than 7.0: $z = \frac{7.0 - 8.5}{1.15} = -1.30$. Using the table, the % of scores less than $z = -1.30$ (50.00 - 40.32) = 9.68%.

c. between 8.3 and 8.7: The standard error has changed as a result of using $n = 9$. Compute it first.

$$\sigma_{\bar{X}} = \frac{2.3}{3} = .77$$

For 8.3, $z = \frac{8.3 - 8.5}{.77} = -.26$

For 8.7, $z = +.26$

Using the table, we get: 10.26 + 10.26 = 20.52%

d. between 8.5 and 9.5: For 8.5, $z = 0$
For 9.5, $z = 1.30$
Using the table, we get: 40.32%

e. Super Bonus: To determine the score above which 5% of the sample means fall, you need to:

1. Realize that you are actually looking for the score associated with P_{95} (95% below it, 5% above it).
2. Realize that P_{95} is associated with a z score that has 45% (50%-5%) of the scores between it and the mean. From the table, this z would be +1.64.
3. To find the raw score value, use your z score conversion, but remember that you are dealing with sample $\bar{X}$'s and need to use $E(\bar{X})$ and $\sigma_{\bar{X}}$.

$$1.64 = \frac{X - 8.5}{.77}$$

$$1.26 = X - 8.5$$

$$9.76 = X$$

Answer: 9.76

UNIT VIII

8.1 POSTTEST

Since hypothesis testing generally involves the use of statistics (sample data) to estimate parameters (population data), the conclusions reached must be expressed in probability terms. Since parameters usually cannot be computed directly, they must be estimated on the basis of sample results. Samples fluctuate, and therefore, the researcher cannot be certain that the samples he employed actually provide close estimates. For this reason, we are restricted to expressing conclusions as probable, rather than certain.

8.2 POSTTEST

1. One possibility:

   $H_0: u_1 = u_2$

   $H_a: u_1 \neq u_2$

   Another possibility:

   $H_0: u_1 = u_2$

   $H_a: u_1 < u_2$

   $H_a': u_1 > u_2$

2. Set 1: Incorrect--H_0 must be specified in terms of "no differences." ($u_1 = u_2$)

   Set 2: Correct

   Set 3: Correct

   Set 4: Incorrect--alternatives must be exhaustive; no allowance has been made for the possibility that u_1 might be less than u_2.

   Set 5: Incorrect--alternatives must be exhaustive; no allowance has been made for the possibility that u_1 might be greater than 23.

   alternatives must be nonoverlapping; if u_1 turns out to equal 22, more than one alternative would be valid.

   Set 6: Incorrect--alternatives must be exhaustive; no allowance has been made for the possibility that u_1 does not equal <u>either</u> u_2 or the value, 100.

8.3 POSTTEST

1. A Type I error occurs when one rejects a null hypothesis that is true. A Type II error occurs when one retains a null hypothesis that is false.

2. The decision to reject or retain the null hypothesis is directly related to the probability level that is employed. The use of the .05 level, for example, may result in the decision to reject the null, whereas the opposite conclusion may be justified were the .01 level employed in the same study. Thus, to eliminate bias, the probability level should be specified before the sample data is collected. That way, a researcher cannot be criticized for using a probability level specifically designed to yield an outcome that he favors.

3. The larger the probability level, the greater the chance of committing a Type I error.

   The larger the probability level, the smaller the chance of committing a Type II error.

4. a. The chance of a Type I error is 5%.

   b. NA; impossible to compute the probability for a Type II error without further information.

   c. He may be committing a Type I error by his rejection of the null.

   d. Type I: 10%

      Type II: chances would decrease. Exact probability is "NA."

8.4 POSTTEST

1. a. $H_0 : \mu_X = 500$

   b. $\sigma_{\bar{X}} = 100/\sqrt{100} = 10$

   c. $z = \frac{482-500}{10} = \frac{-18}{10} = -1.80$

   d. Retain the null: critical value for the .05 level is $-1.96z$ and $+1.96z$. The computed z, therefore, does not fall within the region of rejection.

   e. A Type II error: failure to reject a false null hypothesis.

   f. $z = \frac{520-500}{10} = 2.00$
      Reject the null

   g. Type I

2. a. $H_0 : \mu_X = 25$

   b. $\sigma_{\bar{X}} = 5/\sqrt{100} = 5/10 = .5$

ERIC

c. $z = \frac{23.5 - 25}{.5} = \frac{-1.5}{.5} = -3$

d. Reject the null; computed z exceeds the critical value, -1.96z

e. Type I

f. Reject the null; computed z value exceeds the critical value, 2.58z

g. If a sample of 25 batteries were used, the Standard Error of the Sampling Distribution would increase:

$\sigma_{\bar{X}} = 5/5 = 1$

$z = \frac{23.5 - 25}{1} = -1.50$

Retain the null

h. Type II

i. BONUS: By reducing sample size, the chances of making a Type I error will remain unchanged. Type I errors are totally dependent upon the level of significance.

The chances of making a Type II error, however, will be increased since the Standard Error of the Sampling Distribution will increase. This will require a greater difference between the sample mean and the parameter value to justify rejection of the null. Thus, there will be a greater probability of failing to reject a null hypothesis that is actually false (i.e., making a Type II error).

8.5 POSTTEST

1. a. 2.82   c. 4.54   e. 2.23
   b. 2.07   d. 3.46

2. a. as n <u>increases</u>, the critical t value decreases.

   b. as <u>alpha increases</u> (say, from .01 to .05 to .10, etc.), the critical t value <u>decreases</u>.

3. a. t must be used; population data are unavailable.

   b. chances for a Type I error are 5%; chances for a Type II cannot be determined from the available data.

   c. for the .05 level with $df = 8$, the critical t value is 2.31.

   d. $t = \frac{12.5 - 15}{s/\sqrt{n}} = \frac{-2.5}{3/3} = -2.50$; Decision would be to <u>reject</u>.

   e. Type I

   f. using .01, critical $t = 3.36$; Decision would be to <u>retain</u>.
   g. Type II

Unit VIII Review Test

1. No, he had to rely on a sample to make inferences about the population. The sample could be composed of unusual people and give a misleading result. The findings are only suggestive, not conclusive.

2. Type I since the null was rejected.

3. Answer is "d" since his sample mean was higher than the parameter value.

4. Answer is "b."

5. Answer is "d."

6. The null is assumed to be true until evidence is obtained which makes its truth seem improbable.

7. No! The .01 level has even stricter criteria (larger critical values) for rejecting the null.

8. a. $H_0 : u_X = 80$

   b. ± 1.96

   c. $\dfrac{86 - 80}{8/\sqrt{100}} = \dfrac{6}{.8} = 7.5$

   d. Reject the null

   e. Type I.

9. a. We don't know the population variance (σ)

   b. $\dfrac{68 - 62}{14/5} = \dfrac{6}{2.8} = 2.14$

   c. for $df = 24$, $t = 2.797$

   d. Retain

ERIC

9.1 POSTTEST

1\. a. $$t = \frac{91-79}{\sqrt{\frac{29(100)+29(95)}{30+30-2}\left(\frac{1}{30}+\frac{1}{30}\right)}}$$

$$= \frac{12}{\sqrt{\frac{2900+2755}{58}\left(\frac{2}{30}\right)}}$$

$$= \frac{12}{\sqrt{\frac{5655}{58}\bullet\frac{2}{30}}} = \frac{12}{\sqrt{\frac{11310}{1740}}} = \frac{12}{\sqrt{6.5}}$$

$$= \frac{12}{2.55}$$

Ans. = 4.71

b. For $df = 58$, $t = 2.70$ at $\alpha = .01$

c. Reject the null

d. Type I

2\. Sex--it is the only variable likely to show a significant relationship to facial hair.

3\. Race--it would show the highest correlation with attitudes in this study.

4\.

Comments	No Comments	D	D^2	D^2
80	75	5	25	25
98	90	8	64	64
75	77	-2	4	4
83	77	6	36	36
100	88	12	144	144
63	65	-2	4	4
79	60	19	361	361
94	93	1	1	1
90	81	9	81	81
70	66	4	16	16

$\Sigma D = 60$ $\quad \Sigma D^2 = 736$

$\overline{D} = 6$

$$d^2 = 736 - \frac{(60)^2}{10}$$

$$d^2 = 736 - 360$$

$$d^2 = 376$$

$$t = \frac{6}{\sqrt{\frac{376}{10(9)}}} = \frac{6}{\sqrt{4.17}}$$

$$t = \frac{6}{2.04} = 2.94$$

a. 2.94
b. 2.26 ($df = 9$)
c. reject null. Comments are more effective than no comments.
d. Type I

ERIC

9.2 POSTTEST

1.a. $H_0 : \rho_{xy} = 0$

$H_a : \rho_{xy} \neq 0$

b. $df = n - 2$
$= 30 - 2$
$= 28$

$t_{.01} = 2.76$

c. $$t = \frac{.36}{\sqrt{(1-.36^2)/28}}$$

$$= \frac{.36}{\sqrt{.87/28}}$$

$= .36/.176$

$= 2.04$

d. Retain the null

e. Type II

2.a. $H_0 : \rho_{xy} = 0$

$H_a : \rho_{xy} \neq 0$

b. $t_{.05}$ for df of $14 = 2.145$

c. $$t = \frac{.78}{\sqrt{(1-.78^2)/14}}$$

$$= \frac{.78}{\sqrt{.392/14}} = .78/.167$$

$= 4.67$

d. Reject the null

e. Type I

9.3 POSTTEST

1. Point estimation uses a single score value, such as one sample mean, as the sole basis for estimating the population parameter (i.e., mean). Interval estimation also uses a value obtained from a single sample as the basis for its estimate of the population. But, it further employs knowledge of the Sampling Distribution (standard error and probability) to set an interval around the estimate and to specify a degree of confidence (based on probability) that the interval will contain the parameter. (Try to use <u>your</u> own words. Memorizing this word for word may produce permanent brain damage and no idea, before the damage sets in, of what it means.)

2. "d": all are true.

3.a. $Conf_{.95} = \bar{X} \pm (t_{conf})\, s_{\bar{X}}$

$df = n - 1 = 35$

$t_{.95} = 2.04$ (approx.)

$s_{\bar{X}} = 12/\sqrt{36} = 2$

$Conf_{.95} = 68 \pm 2.04\ (2)$

$= 63.92$ (lower) + 72.08 (upper)

b. $t_{.99\ or\ .01} = 2.75$

$Conf_{.99} = 68 \pm 2.75\ (2)$

$= 62.5$ (lower) + 73.5 (upper)

c. Point estimate = 68

<u>Unit IX Review Test</u>

A. 1. a 6. d

2. d 7. b

3. b 8. c

4. c 9. d

5. a 10. f

B. $Conf_{.99} = 525 \pm 2.80\ (22.4)$

$= 462.28$ and 587.72

$Conf_{.95} = 525 \pm 2.06\ (22.4)$

$= 478.85$ and 571.14

10.1 POSTTEST

1. Interval-ratio measures are scores that can be ordered on a continuum from low to high, and have equal distances between points. Test scores, height, weight, etc. are examples.

   Ordinal scores can be thought of as rankings. They, too, can be ordered on a continuum from low to high. But one cannot infer that there are equal distances between ranks on the trait being measured. For example, in a beauty contest are all contestants likely to be the same amount "more beautiful" than the person finishing immediately behind them? Probably not: the 3rd and 4th place finishers might have been very close; the 5th place finisher way behind them, etc.

   Nominal measures simply identify a subject as a member or nonmember in a particular category. The scores are f's--frequency counts for the different categories.

   Chi-square is applied to nominal data. An example research question is "Is hair color for school children equally distributed among blonds, brunettes, and redheads?" "Are attitudes toward capital punishment equally divided among supporters, nonsupporters, and undecideds?"

10.2 POSTTEST

1.a. $H_0 : O = E$ It means that the different types of books are selected with equal frequency.

b.

	Novels	Nonfiction	Children	Other	Total
Observed:	40	25	28	15	108
Expected:	27	27	27	27	27

c. $$\chi^2 = \frac{(40-27)^2 + (25-27)^2 + (28-27)^2 + (15-27)^2}{27}$$

$$= \frac{318}{27}$$

$$= 11.78$$

d. for $df = 3$, $\chi^2_{.01} = 11.34$

e. Reject (barely made it!)

f. Type I

2.a.

	Cauc.	Black	Chic.	Other	Total
Observed	35	41	20	4	100
Expected	40	40	15	5	100

b. $\chi^2 = 25/40 + 1/40 + 25/15 + 1/5$

$= 2.52$

Crit. $\chi^2 = 7.82$ (from Table III)

Retain the null.

3.

	Cauc.	Black	Chic.	Other	Total
Observed	35	41	20	4	100
Expected	25	25	25	25	100

$$\chi^2 = \frac{100 + 256 + 25 + 441}{25}$$

$$= \frac{822}{25}$$

$$= 32.88$$

Reject the null.

4. If $df = 1$ in a one-way test, Yates' correction is needed.

10.3 POSTTEST

1.a. $$E\text{'s} = \frac{(60)(70)}{200} = 21; \quad \frac{(60)(70)}{200} = 21; \quad \frac{(60)(60)}{200} = 18$$

$$\frac{(140)(70)}{200} = 49; \quad \frac{(140)(70)}{200} = 49; \quad \frac{(140)(60)}{200} = 42$$

$$\chi^2 = \frac{(10-21)^2}{21} + \frac{(20-21)^2}{21} + \frac{(30-18)^2}{18} + \frac{(60-49)^2}{49} + \frac{(50-49)^2}{49} + \frac{(30-42)^2}{42}$$

$$= 5.76 + .05 + 8 + 2.47 + .02 + 3.43$$

$$= 19.73$$

b. $df = (R-1)\ (C-1)$

$= (2)\ (1)$

$= 2$

ERIC Full Text Provided by ERIC

c. tabled $\chi^2_{.01} = 9.21$

d. reject the null

2.a. E's = 8.75 21 40.25

16.25 39 74.75

$$\chi^2 = \frac{(10-8.75)^2}{8.75} + \frac{(20-21)^2}{21} + \frac{(40-40.25)^2}{40.25} +$$

$$\frac{(15-16.25)^2}{16.25} + \frac{(40-39)^2}{39} + \frac{(75-74.75)^2}{74.75}$$

$$= .18 + .05 + 0.00 + .10 + .03 + 0.00$$

$$= .35 \text{ or } .36$$

b. $\chi^2_{2,.05} = 5.99$

c. Retain the null

Unit X Review Test

1. A one-way test looks at results on one variable only, testing for equal frequencies or goodness-of-fit.
Example: Are equal numbers of automobiles presently using regular, unleaded, and diesel gasoline?

A two-way test looks at the relationship (or independence) of two variables.
Example: Is the type of gasoline used (regular, unleaded, or diesel) related to the number of doors in the vehicle (two or four)?

2. The equal frequencies test examines whether the same number of f's fall into each of the other categories. The goodness-of-fit instead examines whether the observed f's for the categories conform to some hypothesized or known distribution.

Formula: the same for both

O's : Whatever is observed in the particular study. (O's are determined in the same way for both.)

E's : Differ depending on the design. In the equal f case, $E = \frac{\text{Total Obs.}}{k}$

All E's will be the same.

In goodness-of-fit test, E's will be the number (proportion) of δ hypothesized for each category. E's will differ across categories.

3.a. Equal δ's is appropriate, since he is trying to determine whether δ's are <u>similar</u> at different class levels.

b. $\chi^2 = \frac{(20-25)^2}{25} + \frac{(30-25)^2}{25} + \frac{(25-25)^2}{25} + \frac{(25-25)^2}{25}$

$= 1 + 1 + 0 + 0$

$= 2$

c. $\chi^2_{3,.05} = 7.82$

d. Retain the null

4.a. $\chi^2 = \frac{(30-15)^2}{15} + \frac{(10-30)^2}{30} + \frac{(30-20)^2}{20} + \frac{(30-35)^2}{35}$

$= 15 + 13.33 + 5 + .71$

$= 34.04$

b. $\chi^2_{3,.01} = 11.34$

c. Retain the null

d. The distribution of majors' attending games does not conform to the distribution of all majors in the university. Arts and Sciences and Education majors are over-represented at games; Business and "Other" majors are under-represented.

5.a. $E's = \frac{(70)(40)}{120} = 23.33; \quad \frac{(70)(45)}{120} = 26.25; \quad \frac{(70)(35)}{120} = 20.42$

$\frac{(50)(40)}{120} = 16.67; \quad \frac{(50)(45)}{120} = 18.75; \quad \frac{(50)(35)}{120} = 14.58$

$\chi^2 = \frac{(30-23.33)^2}{23.33} + \frac{(15-26.25)^2}{26.25} + \frac{(25-20.42)^2}{20.42} +$

$\frac{(10-16.67)^2}{16.67} + \frac{(30-18.75)^2}{18.75} + \frac{(10-14.58)^2}{14.58}$

$= 1.91 + 4.82 + 1.03 + 2.67 + 6.75 + 1.44$

$= 18.62$

$\chi^2_{2,.05} = 5.99$

Reject the null

b. Political preference and newspaper preference appear to be related. Republicans and "Others" seem more inclined to pick the <u>Press Minotaur</u>. Democrats seem more inclined to pick the <u>Daily Planet</u>.

UNIT XI

11.1 POSTTEST

1. The t test holds the Type I error rate at α for <u>each</u> comparison. If there are more than two treatments, more than one t test will be needed to analyze each unique pairing (e.g., 1 vs. 2, 1 vs. 3, 2 vs. 3). The chances of making a Type I error in <u>at least one</u> of those analyses exceed α since each separate test carries that amount of risk by itself.

2. As more treatments are added to the design, the Type I error rate will increase. More treatments mean more chances of selecting an unusual sample, and concluding that significant differences exist when the null (no differences) is actually true. ANOVA avoids this problem by holding the overall Type I error rate (regardless of the number of treatments) at α.

3. $H_0 : u_1 = u_2 = u_3 = u_4$

4. If the null is retained, one concludes that there are no differences in the populations represented by the sample means.

   If the null is rejected, it means that there is at least one difference between population means. Follow-up tests are then performed to find where that difference lies.

11.2 POSTTEST

1. The F ratio expresses the extent of differences between groups (or treatments) relative to the extent of differences between individuals within those groups.

   If the null hypothesis is false, the expectancy is for "significant" differences to occur between groups, and the F value to reach a significantly high level.

2. a. Random error accounts for the denominator.

   b. Random error only would account for the numerator.

   c. Random error and a treatment effect would account for the numerator.

   d. If the null were false, $E(F) > 1$.
   The reason can be easily demonstrated mathematically. If the null is false, there should be a treatment effect. If so: random error + treatment effect > random error only. The numerator will thus tend to be larger than the denominator, producing an F value > 1.

3. If the null is rejected we would conclude that there is at least one difference between population means (check back to Unit 11.1 for review).

   Another way of saying this is that there is a significant treatment effect.

4. a. $a - 1 = 3$

   b. $a(n-1) = 4(9) = 36$

   c. $an-1 = 40 - 1$

   $= 39$

5.

Source	SS	df	MS	F
Between	200	4	50	2.50
Within	800	40	20	
Total	1000	44		

395

11.3 POSTTEST

The analysis indicated that there was a significant treatment effect. Looking at the means we can get a general impression of the nature of that effect. The Sports treatment produced the highest attitudes, the Abstract treatment produced the lowest and the Education treatment fell in the middle. Follow-up tests could be used to make specific comparisons between the three means.

Note: One type of follow-up test would be to make all unique comparisons using the t test for independent samples. Since we protected the overall Type I error rate by using F, there is less concern about using t than would have been the case otherwise. However, other follow-up tests exist and these should be explored before selecting t as the preferred method.

Unit XI Review Test

1. As additional comparisons between pairs of means are made using t tests, the overall Type I error rate increases. By "overall" is meant the chances of at least one Type I error occurring across all comparisons.
2. a
3. a
4. c
5. a
6. c
7. c
8.

Source	SS	df	MS	F
Between	60	3	20	2.31
Within	590	68	8.68	
Total	650	71		

9. The decision would be to retain the null. The computed F is smaller than the tabled F. There is insufficient basis for inferring that the population means differ from one another; thus there is insufficient basis for concluding that any treatment is more effective than any other.

Appendix

Table I: Normal Curve Area

Table II: Significance Levels of the t Ratio

Table III: Significance Levels of the Chi-Square Test

ERIC

Table I

Normal Curve Area

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.49903									
3.2	.49931									
3.3	.49952									
3.4	.49966									
3.5	.49977									
3.6	.49984									
3.7	.49989									
3.8	.49993									
3.9	.49995									
4.0	.50000									

Table II

Significance Levels of the t Ratio

	Level of Significance				
df	.10	.05	.02	.01	.001
1	6.314	12.706	31.821	63.657	636.619
2	2.920	4.303	6.965	9.925	31.598
3	2.353	3.182	4.541	5.841	12.941
4	2.132	2.776	3.747	4.604	8.610
5	2.015	2.571	3.365	4.032	6.859
6	1.943	2.447	3.143	3.707	5.959
7	1.895	2.365	2.998	3.499	5.405
8	1.860	2.306	2.896	3.355	5.041
9	1.833	2.262	2.821	3.250	4.781
10	1.812	2.228	2.764	3.169	4.587
11	1.796	2.201	2.718	3.106	4.437
12	1.782	2.179	2.681	3.055	4.318
13	1.771	2.160	2.650	3.012	4.221
14	1.761	2.145	2.624	2.977	4.140
15	1.753	2.131	2.602	2.947	4.073
16	1.746	2.120	2.583	2.921	4.015
17	1.740	2.110	2.567	2.898	3.965
18	1.734	2.101	2.552	2.878	3.922
19	1.729	2.093	2.539	2.861	3.883
20	1.725	2.086	2.528	2.845	3.850
21	1.721	2.080	2.518	2.831	3.819
22	1.717	2.074	2.508	2.819	3.792
23	1.714	2.069	2.500	2.807	3.767
24	1.711	2.064	2.492	2.797	3.745
25	1.708	2.060	2.485	2.787	3.725
26	1.706	2.056	2.479	2.779	3.707
27	1.703	2.052	2.473	2.771	3.690
28	1.701	2.048	2.467	2.763	3.674
29	1.699	2.045	2.462	2.756	3.659
30	1.697	2.042	2.457	2.750	3.646
40	1.684	2.021	2.423	2.704	3.551
60	1.671	2.000	2.390	2.660	3.460
120	1.658	1.980	2.358	2.617	3.373
∞	1.645	1.960	2.326	2.576	3.291

Table II is abridged from Table IV of Fisher & Yates, Statistical Tables for Biological, Agricultural and Medical Research, published by Longman Group Ltd., London (previously published by Oliver and Boyd, Ltd., Edinburgh), and by permission of the authors and publishers.

Table III

Significance Levels of the Chi-Square Test

	Level of Significance		
df	.01	.05	.10
1	6.64	3.85	2.71
2	9.21	5.99	4.60
3	11.34	7.82	6.25
4	13.28	9.49	7.78
5	15.09	11.07	9.24
6	16.81	12.59	10.64
7	18.48	14.07	12.02
8	20.09	15.51	13.36
9	21.67	16.92	14.68
10	23.21	18.31	15.99
11	24.72	19.68	17.28
12	26.22	21.03	18.55
13	27.69	22.36	19.81
14	29.14	23.68	21.06
15	30.58	25.00	22.31
16	32.00	26.97	23.54
17	33.41	27.59	24.77
18	34.80	28.87	25.99
19	36.19	30.14	27.20
20	37.57	31.41	28.41
21	38.93	32.67	29.62
22	40.29	33.92	30.81
23	41.64	35.17	32.01
24	42.98	36.42	33.20
25	44.31	37.65	34.38
26	45.64	38.88	35.56
27	46.96	40.11	36.74
28	48.28	41.34	37.92
29	49.59	42.56	39.09
30	50.89	43.77	40.26

Table III is abridged from Table IV of Fisher & Yates, Statistical Tables for Biological, Agricultural and Medical Research, published by Longman Group Ltd., London (previously published by Oliver and Boyd, Ltd., Edinburgh), and by permission of the authors and publishers.

Index