DOCUMENT RESUME

ED 090 036 SE 017 626

AUTHOR Rubner, Victoria Petro; And Others

TITLE STAT-CONCEPT: An Interactive Computer Package Supporting a First Course in Educational Statistics.

PUB DATE Apr 74

NOTE 22p.; Paper presented at the annual meeting of the American Educational Research Association (Chicago, Illinois, April 1974)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE

DESCRIPTORS *College Mathematics; *Computer Assisted Instruction; Graduate Study; Instruction; *Instructional Materials; Laboratories; *Probability; *Statistics

IDENTIFIERS Educational Statistics

ABSTRACT

The Statistical Concepts Package (STAT-CONCEPT) is a coordinated collection of interactive computer programs and printed materials designed to enrich graduate instruction in educational statistics. Probability, descriptive statistics, sampling distributions, and power are four of the thirteen topics included. By doing the computational labor and providing a responsive enrivornment, the computer frees the student to explore statistical concepts in a "what happens if ...," laboratory-type, investigative atmosphere. This paper discusses the rationale, content, and student-program interaction of the STAT-CONCEPT package. Technical aspects as well as the package's usage during the past two years are also discussed. (Author)

HERA

Session 12.19

STAT*CONCEPT

An Interactive Computer Package Supporting a First Course in Educational Statistics

Victoria Petro Rubner
University of Wisconsin-Madison

George E. Behr
Southwest Regional Laboratories

Frank B. Baker
University of Wisconsin-Madison

STAT*CONCEPT

An Interactive Computer Package Supporting a First Course in Educational Statistics

Abstract

The Statistical Concepts Package (STAT*CONCEPT) is a coordinated collection of interactive computer programs and printed materials designed to enrich graduate instruction in educational statistics. Probability, descriptive statistics, sampling distributions, and power are four of the thirteen topics included. By doing the computational labor and providing a responsive environment, the computer frees the student to explore statistical concepts in a "what happens if ...," laboratory-type, investigative atmosphere. This paper discusses the rationale, content, and student-program interaction of the STAT*CONCEPT package. Technical aspects as well as the package's usage during the past two years are also discussed.

STAT*CONCEPT

An Interactive Computer Package Supporting a First Course in Educational Statistics

1. INTRODUCTION

The Statistical Concepts Package (STAT*CONCEPT) is a coordinated package of interactive computer programs and printed materials. The development of STAT*CONCEPT was an outgrowth of a perceived need to modernize laboratory aspects and enrich instruction in a first course in graduate educational statistics. It was developed and implemented by the Laboratory of Experimental Design, Department of Educational Psychology at the University of Wisconsin-Madison. Mr. George Behr developed the computer programs and Ms. Victoria Petro Rubner developed the initial coordinated lesson plans and user guides. A grant from the Knapp Bequest allowed Mr. Frank Baker to prepare a concise user manual which incorporates both the computer program rationale and lesson plans in a thirteen-session format. The purpose of the manual and computer programs is to involve the use of computers in teaching statistics on a university-wide basis. Implementation of STAT*CONCEPT during the past two years has proven it to be a useful tool in augmenting lecture material.

2. CONTENT

STAT*CONCEPT is designed to supplement instruction and not to teach per se. It is not a spectator activity and requires that students become involved in interacting with the computer. At present the package consists of thirteen laboratory sessions which deal with major statistical constructs and relationships. The sessions follow a sequence similar to that found in elementary statistics textbooks. A brief description of each session appears below.

Session 1 Tables	The primary function of this session is to acquaint the student with the computer terminal procedures and formats employed by STAT*CONCEPT. The statistical concept of interest is that of the relationship between the width of the grouping interval employed in constructing frequency distributions and the information conveyed by the distribution.
Session 2 Probability	This session provides the student an opportunity to apply elementary rules of probability to a data set provided in the lesson. Such concepts as mutually exclusive, independence, and joint events are examined.
Session 3 Descriptive Statistics	This session assists the student in developing a correspondence between actual data and the descriptive statistics

which depict certain attributes of that data. The student is able to get a "feel" for infoimation conveyed by the mean, median, range, variance and standard deviation.

Session 4
Random Variables

In this session the student performs coin flipping experiments quickly and efficiently. This facility provides the opportunity to grapple with the concepts of a fair coin, random sequencing, expected value and discrete binomial random variables. By varying the number of coins tossed and the number of trials performed, the student observes the effect on the frequency distribution of the number of successes.

Session 5
Binomial Distribution

The relationship between a particular binomial distribution, defined by its parameters n and p, and the resulting probability distribution is the focus of this session. The student varies the values of n and p observing the effect on the shape of the binomial density.

Session 6
Standard Scores

This session deals with standard scores. The student specifies values for the mean and standard deviation of a normal distribution and practices converting data values to standard scores and conversely, standard scores to data values.

Session 7

Normal Distribution

This session focuses on the development of student skill in using tables of the normal distribution. The student specifies values for the mean and standard deviation of a hypothetical distribution. The computer then generates positive or negative z-values to which the student must respond with the appropriate probability value. Similarly, the computer generates a cumulative probability value from which the student must derive the correct z-value.

Session 8

Sampling Distributions of the Mean

The relationships between the distribution of a random variable in a population. The distribution of a random variable in a sample and the sampling distribution of a statistic are examined in this session. The program is designed to draw samples from a uniform, normal, binomial, χ^2 or F population distribution. The student specifies the distribution type and the size and number of random samples to be drawn. The computer then draws each sample computing its mean and generating an empirical sampling distribution of the mean.

Session 9

Sampling Distribution of the Variance

In this session, the same procedures as those employed in Session 8 are utilized to generate empirical sampling distributions of both the biased and unbiased estimates of the variance. For each empirical sampling distribution, the computer calculates the mean, variance and standard deviation.

Session 10
Hypothesis Testing

The procedures followed in setting up a hypothesis testing situation and the interdependencies of the factors involved are examined. The program assumes a large sample z-test of the hypothesis H_0: $\mu = 100$. The student specifies the significance level, rejection region, population, standard deviation, sample size and population mean under the alternative. The student varies the value of one of these quantities, holding the other four constant and analyzes the effect on the critical value(s).

Session 11
Power Curves

The computer performs the calculations needed for the student to construct power curves for a large sample z-test of the hypothesis H_0: $\mu = 100$. As in Session 10, the student varies the value of one of the five factors, holding the others constant. For each variation, the value of power is calculated by the computer, the student plots the derived values of power and analyzes the effects the manipulations have on the power curve.

Session 12
The t-test

The focus of this session is significance testing using the t-distribution. The computer has been programmed to perform the calculations associated with the single sample t-test, two independent sample t-test and the correlated (paired) t-test. The student is able to explore statistical inference procedures.

Session 13

Analysis of Variance

Concepts involved in ANOVA are examined by the student. The one-way fixed effects ANOVA is used to illustrate sources of variation, breakdown of sums of squares, fixed effects, and the F tests employed.

A proposed expansion of STAT*CONCEPT includes adding sessions to deal with such topics as regression and correlation, concepts usually encountered in a second educational statistics course.

3. STUDENT-PROGRAM INTERACTION

Instructions for each session include a worksheet explanation of how to use the computer via the terminal, a process flow chart which provides an overall diagram for the session, and a set of coordinated exercises designed to provide direction for the exploration of each statistical concept. (See STAT*CONCEPT, Session 5 in the Appendix of this paper.) Students are instructed to study the lesson in detail before interacting with the computer. This allows them to become familiar with console procedures, the sequence of actions involved, the source of the data to be dealt with and how the exercises direct exploration of the concepts. Many sessions employ a prestored set of data that can be used to illustrate the computer procedures as well as provide information needed for the exercises. Some sessions allow the students to provide the data to be analyzed. Various sessions contain extended capabilities that allow interested students to go beyond the requirements of the exercises and explore the concepts in greater depth.

The student generally deals with each session in a "what happens if ..." framework; certain quantities are manipulated over a range of values and the effect of the manipulation is observed and analyzed with respect to a particular variable. For example, in Session 11 power curves are constructed for a large sample z-test of the hypothesis H_0: $\mu = 100$. The student specifies whether the test is two-tailed or one-tailed (and the direction) and fixes values for α, the population standard deviation and sample size. The alternative mean is then systematically varied and the student examines the value of power for each alternative. Similarly, other factors such as α, the population standard deviation or the sample size can be varied and the effect on the power curve observed.

4. TECHNICAL ASPECTS

STAT*CONCEPT software is at present written in FORTRAN-V and operates on the UNIVAC 1110 at the Madison Academic Computing Center (MACC) of the University of Wisconsin-Madison. Each session's computer program is an independent, self-contained package and does not require secondary storage for operation. The 1110 is a word oriented machine in which one word equals 36 bits and 6 bits equal one character.

The programs are designed to operate with either a standard teletypewriter or Hazeltine 2000 alphanumeric display as the input-output medium. Particular advantage has been taken of the Hazeltine's screen control capabilities; e.g., cursor movements, print position, program control to roll up or down selected portions of the display and variation of screen intensity are used to provide effective formatting and information presentation on the alphanumeric display. Graphical displays appear with some of the computer programs. An extension of the display mode could incorporate more sophisticated graphical displays in a larger number of sessions. Availability of a graphics terminal would make such an extension feasible.

The programs are basically exportable despite naturally existing hardware and operating system idiosyncracies. The programs are written in FORTRAN V, a UNIVAC extended FORTRAN, except the Hazeltine control routine which is written in assembly language. The following brief statements describe the nature of the dependencies:

1. Special FORTRAN V capabilities such as ENCODE, DECODE, REREAD, etc. are employed by the programs. However, their functional counterparts do exist or are programmable in other FORTRANS.
2. Format specifications are machine dependent. UNIVAC uses FIELDDATA code, a 6-bit character code. Parenthetically, Hazeltine uses standard 8-bit ASCII code for control characters. Since such a code is not available to FORTRAN users on the UNIVAC, an assembly language code is required. It is quite possible that the Hazeltine command routine could be written in FORTRAN in another system.
3. MACC developed and supported utility routines, e.g., centering character strings, providing date and time of day, are employed. Statistical routines to generate random numbers from specified distributions as well as calculating cumulative areas under such distributions are utilized. Most academic computing centers should contain libraries of equivalent software functions.

Although conversion efforts would be necessary to use STAT*CONCEPT on another computer, the following considerations will aide such endeavors:

1. The source document includes numerous explanatory comments.

2. Each program is built in a modular with a set of subroutines common to many programs.
3. The STATISTICAL CONCEPTS PACKAGE manual includes functional flowcharts for the computer programs which support the thirteen sessions. Sample output is also included, providing a criteria for debugging.

Due to differing hardware and operating systems it is difficult to extrapolate program operating characteristics and costs to other machines. However, a few observations concerning our operational experience may provide some idea of the nature of these issues. Our experience indicates that a student spends about one-half hour per session interacting with the computer. Operationally this amounts to five seconds of CPU time and run-time costs of about $2 under MACC's billing scheme. Storage in terms of space and cost is minimal since the thirteen programs are relatively small. Programs are independent and need not be simultaneously on-line. To further minimize storage requirements, it may be desirable to have those programs on-line which coincide with lecture material and remove them after a suitable time period.

5. CONCLUSION

Students enjoy using the coordinated package and find it a beneficial means of exploring basic statistical concepts. STAT*CONCEPT is significant because of its utility in modernizing the laboratory facet of a first graduate course in statistics. It can be incorporated for use in several ways. The computer supported package can constitute the entire laboratory aspect for the course or the materials can be used to supplement existing laboratory experiences. Students can individually use various sessions to review and explore concepts which are difficult for them to understand. Our experiences with STAT*CONCEPT have shown that nonquantitatively-oriented students find this application particularly useful. In addition to individual use, STAT*CONCEPT sessions have been utilized effectively in group situations. Sessions can be simultaneously coordinated with class presentation to dramatize and visually depict the concepts under consideration. The instructor can perform the mechanics at the console and direct class analysis of the output.

Useful criticisms resulting from completion of evaluation forms by students have been incorporated into the current revision of the original system. Feedback from users continues to be solicited as students employ STAT*CONCEPT individually and in groups to aid their understanding of statistical principles and relationships.

APPENDIX

Session 5

BINOMIAL DISTRIBUTION

Introduction

The binomial distribution is one of several reference distributions that are used in statistics. The binomial distribution is actually a family of distributions whose individual members are defined by the values of the parameters N and P. In the present laboratory session the relationship between the probability distribution and its parameters is examined. For a given set of parameter values the computer will evaluate the binomial function over all possible outcomes. The resulting probability distribution will then be displayed at the computer terminal for your inspection. By manipulating the values of the parameters, you should develop an appreciation for the family of binomial distributions, the role of parameters, and gain an understanding of the relationship of the shape of the binomial density to the values of the parameters.

The computer program evaluates the function rule

$$P(X = r) = \binom{N}{r} P^r (1-P)^{N-r} \quad 0 \leq X \leq N$$

where: N is the number of observations

P is the probability of a success

r is the number of successes in N observations

X is a random variable

N and P are the parameters of the binomial distribution.

Before using the computer terminal, read the complete session description and determine what you are to do.

Computer Terminal Procedures

Follow the instructions in the computer procedures section to turn on the terminal and log-in with the computer. The XQT statement will be

@XQT^STAT*CONCEPT.BDIST

Step 1

COIN TOSS OR BINOMIAL? Respond by typing BVAL.

Step 2

N = ? will be printed. You must provide a value of N, the number of independent observations, where $1.0 \leq N \leq 9.0$. Be sure to include the decimal point.

Step 3

P = ? will be printed. Type in a value of P, the probability of success. The value of P must be a two digit decimal number whose value is $0.00 < P \leq 0.99$.

The computer will then evaluate the binomial function over all possible values of r for the parameters N and P you have specified. The resulting probability distribution will be printed in the form shown below.

For example, if N = 5.0 and P = .38, the corresponding table would read:

R	P(X = R)
0	.091613
1	.280750
2	.344146
3	.210928
4	.064639
5	.007924

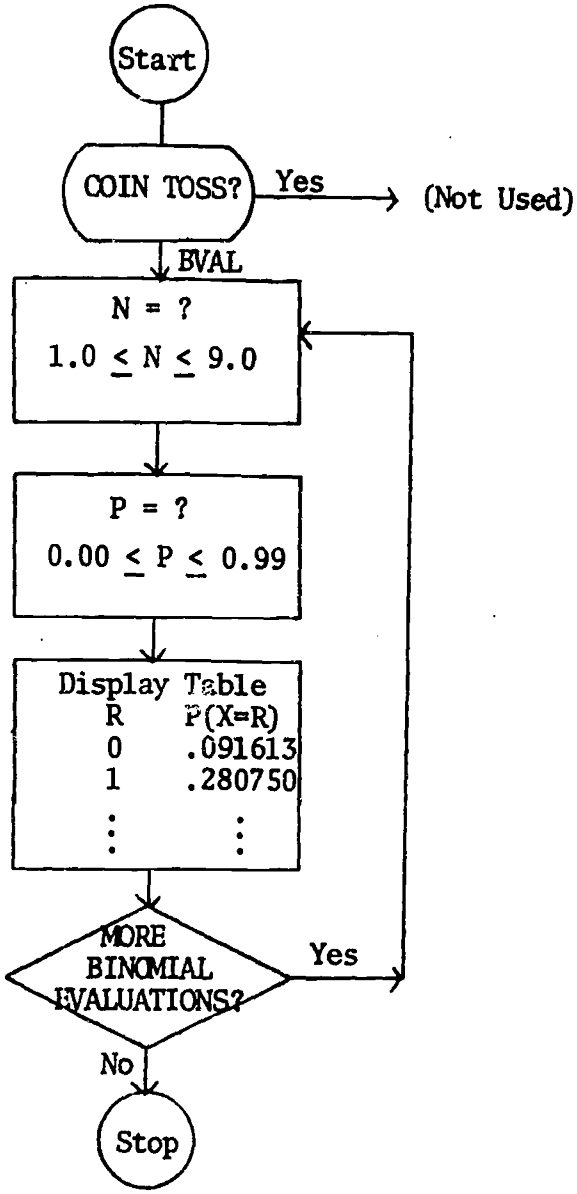
As expected, the sum of the probabilities is 1.000000.

Step 4

MORE BINOMIAL EVALUATIONS? will be printed. A YES response will return you to Step 2 and the whole process repeated. A NO response will result in termination and the end of lesson message will appear.

BINOMIAL DISTRIBUTION

Process Flow

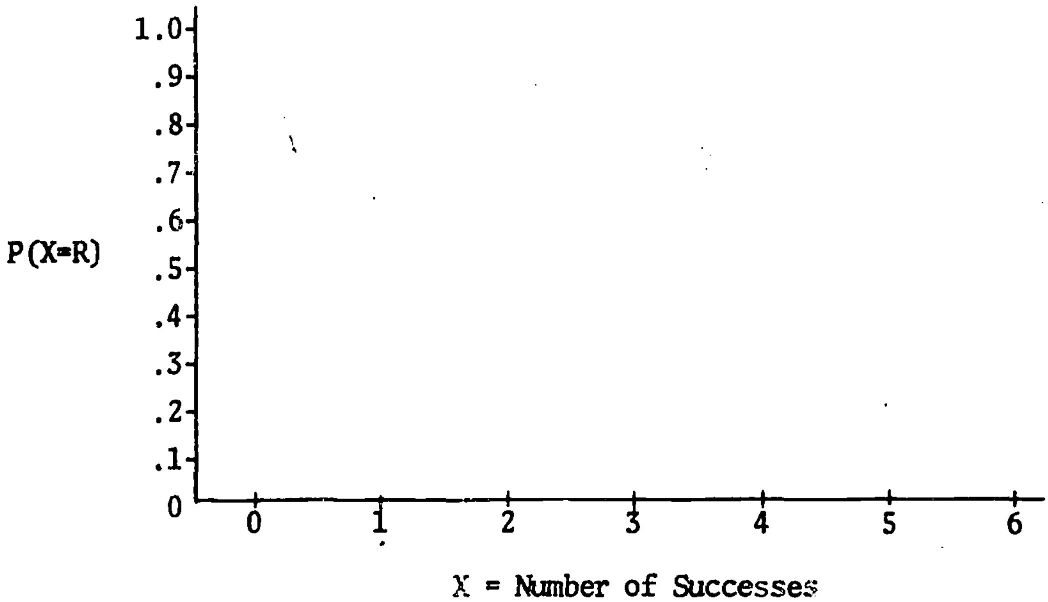


Exercises

1. Examine the effect of varying P for fixed N.

   Let N = 6. be fixed. Set P = .2 initially.

   a) Sketch the resulting binomial distribution on the figure below.



   b) Use the computer to obtain tables for N = 6, P = .5 and N = 6, P = .8. In each case sketch the resulting binomial distribution on the figure above. You should observe three major relationships:

      1) As P increases, the largest probability value occurs near NP.

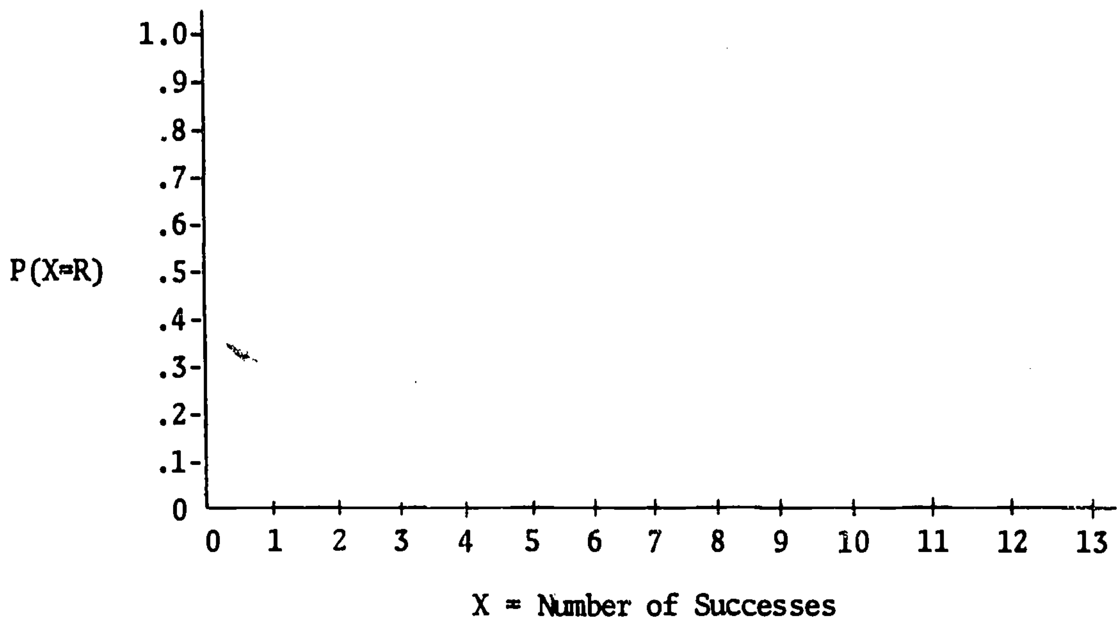
      2) When P = .5, the distribution is symmetrical with its maximum at X = 3.

      3) When P < .5 the "tail" of the distribution is to the right. When P > .5 the tail of the distribution is to the left.

2. Examine the effect of varying N for a fixed value of P.

Let N = 3 initially. Let P = .3 be fixed.

a) Sketch the resulting binomial distribution.



b) The mean of the binomial distribution is NP, which in the present case is (3)(.3) = .9, thus the maximum probability should occur near X = 1.0.

c) Perform a series of binomial evaluations with N = 6., 9. and 12. for P = .3.

1) For each distribution, calculate NP and note whether the maximum probability occurs near this value.

2) Sketch each distribution on the diagram above.

d) You should observe the following:

1) The location of the distribution is a function of NP, hence as sample size increases for fixed P, the distribution is further to the right of the diagram.

2) The ________ distribution tends to smooth out as N is increased, while its shape is generally the same for the values of N involved.

3. Now that you have observed the effect of N and P on the binomial distribution, explore this relationship in greater detail.

   a) Select a value of N and P; then sketch what you think the corresponding binomial distribution should look like.

   b) Enter the value of N and P and compare the obtained distribution with your sketch.

   c) Repeat a,b until you get a good correspondence between the two distributions.

Teletype Output for Session 5

```
@XQT STAT*CONCEPT.BDIST
  ARE YOU USING A HAZELTINE 2000
NO
   COIN TOSS OR BINOMIAL
BVAL
                          BINOMIAL FORMULA EVALUATION N = ?  6.
 P = ?  .2
 R        P(X = R)
  0        .262144
  1        .393216
  2        .245760
  3        .081920
  4        .015360
  5        .001536
  6        .000064    MORE BINOMIAL EVALUATION?  YES
                          BINOMIAL FORMULA EVALUATION N = ?  6.
 P = ?  .5
 R        P(X = R)
  0        .015625
  1        .093750
  2        .234375
  3        .312500
  4        .234375
  5        .093750
  6        .015625    MORE BINOMIAL EVALUATION?  YES
                          BINOMIAL FORMULA EVALUATION N = ?  6.
 P = ?  .8
 R        P(X = R)
  0        .000064
  1        .001536
  2        .015360
  3        .081920
  4        .245760
  5        .393216
  6        .262144    MORE BINOMIAL EVALUATION?  YES
                          BINOMIAL FORMULA EVALUATION N = ?  3.
 P = ?  .3
 R        P(X = R)
  0        .343000
  1        .441000
  2        .189000
  3        .027000    MORE BINOMIAL EVALUATION?  YES
                          BINOMIAL FORMULA EVALUATION N = ?  6.
 P = ?  .3
 R        P(X = R)
  0        .117649
  1        .302526
  2        .324135
  3        .185220
  4        .059535
  5        .010206
  6        .000729    MORE BINOMIAL EVALUATION?  YES
                          BINOMIAL FORMULA EVALUATION N = ?  9.
 P = ?  .3
 R        P(X = R)
  0        .040354
  1        .155650
  2        .266828
  3        .266828
  4        .171532
  5        .073514
  6        .021004
  7        .003858
  8        .000413
  9        .000020    MORE BINOMIAL EVALUATION?  YES
                          BINOMIAL FORMULA EVALUATION N = ?  12.
N MUST BE IN RANGE 1.0 TO 9.0.  TRY AGAIN.  N = ?  5.
P = ?  .3
R         P(X = R)
 0         .168070
 1         .360150
 2         .308700
 3         .132300
 4         .028350
 5         .002430    MORE BINOMIAL EVALUATION?  NO
THE NEXT COMMAND YOU ENTER SHOULD BE EITHER
A @FIN OR A @XQT . . .
```