DOCUMENT RESUME

ED 083 964 JC 730 260

| | |
|---|---|
| AUTHOR | Barton, George M.; Mines, R. F. |
| TITLE | Applied Educational Research and Evaluation. |
| INSTITUTION | Nova Univ., Fort Lauderdale, Fla. |
| PUB DATE | 73 |
| NOTE | 96p. |
| EDRS PRICE | MF-$0.65 HC-$3.29 |
| DESCRIPTORS | Bibliographies; Data Analysis; *Educational Research; *Evaluation Techniques; Guides; *Models; Problem Solving; Research Methodology |

ABSTRACT

A module is presented which instructs the reader in the following activities: (1) identification of a researchable problem, (2) preparation of a research proposal, (3) selection of appropriate research tools and methods, (4) identification of different kinds of variables in research, (5) preparation of a research report, (6) drawing of conclusions from results, (7) evaluation of research of others, and (8) an understanding of the roles of institutional and instructional research. A bibliography and index are given. (CK)

# APPLIED EDUCATIONAL RESEARCH AND EVALUATION

George M. Barton
Nova University

R. F. Mines
Miami-Dade Community College

**National Ed.D. Program for Community College Faculty**

Nova University

1973

**Fort Lauderdale, Florida**

# INTRODUCTORY NOTE

This has been prepared for use with the module entitled Applied Educational Research and Evaluation in the National Ed.D. Program for Community College Faculty of Nova University.

Because of this, subjects given emphasis in this manual are those which we view as being of the greatest importance to the module. This means that subjects which might receive considerable emphasis in the conventional textbook on research and evaluation may be treated only briefly or, in some instances, limited to two or three sentence definitions.

Moreover, since this is intended only as a study guide, even topics of major importance will be treated too concisely to provide the student with comprehensive knowledge of the subjects involved. There obviously must be gaps in any book of this kind -- ones the National Lecturer will fill in through his seminars or which the student will follow up in his readings of other texts.

On the other hand, this manual follows the sequence of topics considered most appropriate for this module and it is hoped that the student will find it a useful springboard to his activities in Applied Educational Research and Evaluation.

G.M.B
R.F.M.

## TABLE OF CONTENTS

## I. THE OBJECTIVES AND SCOPE OF THIS MODULE

This module is entitled Applied Educational Research and Evaluation.

### GOALS

At the completion of this module you should be able to:

(1) Identify a researchable problem

(2) Prepare a research proposal

(3) Select appropriate research tools and methods

(4) Identify different kinds of variables useful in your own research or that of others

(5) Prepare a research report or otherwise communicate research findings

(6) Draw reasonable conclusions from research results

(7) Evaluate research performed and reported by others

(8) Comprehend the roles of institutional research and instructional research in a community college

### SCOPE

Probably the most important point to be made regarding the scope of this module is that it goes well beyond simply involving a study of statistics. (Too frequently seminars or courses of this type are seen as centering almost overwhelmingly about statistical techniques.)

The unit of this module dealing with the treatment of data involves a number of statistical techniques that you will want to consider when planning and undertaking any project in educational research. These statistical techniques are simply tools of research. This module also deals with a number of other facets of research and evaluation, as a glance at the Table of Contents will indicate.

In this module there are included such units as Planning a Research Project; Preparing a Research Proposal; Preparing a Research Report; and Evaluation. This very genuinely is intended as a module in research and evaluation -- not simply a study of statistics alone.

## ORIENTATION

The orientation of this module is toward educational research and evaluation, especially as it occurs in a community college setting.

In this module you will be asked to give emphasis to research problems of which the results can be used by you in your own work or, at the very least, by the institutions with which you are connected.

Defined most simply, research covers the systematic investigation of a problem. Evaluation as defined here may take different forms: it may involve evaluating particular institutions or programs, it may involve evaluating research projects conducted by others or it may involve evaluating the use to which research conducted by yourself or others may be applied to any given situations. In the unit on evaluation included in this module, the matter is considered from these points of view.

In the community college, we are most likely to engage in three broad categories of research: (1) Institutional research -- which has to do with such factors as the role, image and expectations of the college (2) Instructional research -- which involves such matters as the effects of particular teaching strategies or class size (3) Student behavior.

These catergories are by no means all inclusive or mutually exclusive. Examples of each will be given in later sections of this manual.

It should be noted that there are any number of types of research with which this module does not attempt to deal. Some of these -- for instance, market research -- may involve many of the same techniques outlined in this module. But the emphasis here is given entirely to the kinds of institutional, instructional and behavioral research likely to be conducted in a community college.

# II. MAKING THE MOST OF THE GUIDE

## PURPOSE OF THIS GUIDE

This is intended as both a syllabus and a readily-usable manual for this module. To the greatest extent possible, the outline of each unit has been made concise and to-the-point.

This also means that this is not intended in any way as a "text" relating to educational research and evaluation. For fuller explanations of topics covered here, you should refer to other books which cover these subjects in more detail.

## PRESENTATION OF MATERIAL

To make this manual as compact and readily-usable as possible, most narrative material is outlined in point-by-point form. Statistical measures or techniques are outlined in the same way and are accompanied by illustrative examples.

## LIMITATIONS

The obvious limitations of such a manual as this is that it cannot present any topic in great detail and that it must also omit what can be very interesting facets of this overall subject -- historical perspectives, brief biographies of individuals who have made notable contributions to this field, complete descriptions of outstanding studies, and so on.

To counter-balance this factor, suggestions for this kind of background reading are included in the annotated bibliography at the end of this manual.

FORMULAS

Perhaps the principal reason why so many students find textbooks in the field of research and statistics so confusing is that they are filled with what appear to be almost endless pages of formulas. While to some extent, this is unavoidable -- and indeed the absence of formulas in any manual such as this would be unjustifiable -- an effort has been made here to place the emphasis on concepts.

It is hoped that in the long run, this will make the formulas more meaningful to you rather than the reverse being the case.

SEQUENCE OF TOPICS

The topics presented in this manual are arranged in a definite sequence, intended to closely duplicate the way in which these usually will occur in the three seminars of the module itself.

Usually topics will be dealt with in the following sequence:

Seminar 1 - The objectives and scope of this module; the role of research in the community college; identifying researchable problems; planning a research project; preparing a research proposal. (Units I - IV)

Seminar 2 - The selection, application and use of appropriate research tools and methods; the treatment of data; drawing reasonable conclusions from research results. (Units IV - VI)

Seminar 3 - The preparation of a research report; the different types of evaluation; what is involved in evaluation. (Units VII and VIII)

LOCATING ITEMS IN THIS MANUAL

You will notice that this manual has a more detailed index than might usually be found in a manual of this type. You are encouraged to refer to this whenever you need information about a particular item.

SUCCESSFUL USE OF THIS MANUAL

In conclusion, this manual will be most helpful to you if you do not attempt to use it as a textbook but rather as a compact source of ready-to-use information geared to the contents of this particular module. You will use it most successfully if you combine it with other sources listed in the bibliography.

## III. PLANNING A RESEARCH PROJECT

Any educational research project that you plan is likely to fall into one of four categories. These are not the only research types, but they are the ones most likely to occur in the educational field. Knowing the particular category into which your research project falls will help you establish the methodology for it:

### TYPES OF EDUCATIONAL RESEARCH

A. Historical Research:

Usually this is a straight historical study. It differs from the types of research with which we will principally be concerned in this module, in that historical research may well not involve a hypothesis. (The hypothesis is a statement of the outcome which you anticipate in any investigation. It is always in declarative sentence form.)

On the other hand, historical research may formulate certain questions about events of the past and direct the research project toward answering these.

In undertaking historical research, the inexperienced investigator frequently chooses a topic that is too global and is beyond his capabilities and resources in the time available. This is a point to be very carefully considered.

B. Descriptive Research:

Descriptive research describes what currently exists in a given area, institution or other situation. It does not attempt to evaluate whether this is "good" or "bad"; rather it simply describes the current status. (The U. S. Census is an excellent example of this.) In such a study, causal factors will also go uninvestigated; they are part of a different type of study.

C. Experimental and Quasi-Experimental Research:

While these are two different categories of research, they are classified together here since they differ in only one major element. In experimental and quasi-experimental research you have two or more variable. (For example: "If you vary this factor (independent), the other changes (dependent) --"). In experimental research, you can control other variables which might contribute to the effect on the dependent variable. In quasi-experimental research, you nave one or more of these variables which you cannot control.

Much educational research must be designated as quasi-experimental -- especially since we are dealing with groups already established, rather than being randomly selected. (An instance: A study involving two already established classes.) In addition, in educational research we often have to recognize the possible effect of such variables as mental ability and motivation.

While experimental research obviously comes closest to our ideal, a great deal of quasi-experimental research that is of real value is constantly being undertaken in our educational institutions. In stating the limitations of any research project, we must recognize the factors that we cannot control but this nonetheless should not prevent our undertaking any study of this kind that may be helpful to us.

## EDUCATIONAL RESEARCH APPROPRIATE FOR THIS MODULE

It is likely that about ninety percent of the research conducted in educational institutions is of the descriptive type. (This category is broader than the word itself might suggest -- for instance, the Neilsen TV Audience Study, the Gallup Poll, the Harris Public Opinion Studies, etc., are all of this type.)

But in this module we are principally concerned with studies which involve manipulation of data. It is recommended that such studies be of the experimental or quasi-experimental type.

In the research proposal which you will design as part of this module -- and in the resulting study which you will conduct -- you will very probably select an experimental or quasi-experimental study.

## STEPS IN PLANNING A RESEARCH PROJECT

The following steps summarize the stages you are most likely to go through in planning a research project. (They will be better understood if you consider them in conjunction with the subsequent unit in preparing a research proposal):

1. Identify the problem area which you wish to investigate. This will be discussed in more detail in a later section of this unit.
2. Define this problem in clear and specific terms.
3. Survey the literature relating to this problem.
4. Formulate a testable hypothesis (or hypotheses) relating to this problem.

5. Define the basic variables involved in your study.

6. Define the important concepts and terms involved with this study.

7. Define the limitations of your study.

8. Identify the underlying assumptions which will govern the interpretation of your results.

9. Construct a research design which will provide to the greatest degree possible for maximum external and internal validity. (This will include selection of your subjects, control and/or measurement of relevant variables, establishment of criteria to evaluate outcomes, and selection or development of the criteria measures.)

10. Identify data collection procedures.

11. Identify data treatment methodology.

## SELECTION OF A RESEARCH PROBLEM

There are certain important points which must be kept in mind in selecting the problem for your research project.

1. Make certain that it is a "researchable problem." (All too frequently an inexperienced researcher will spend considerable time planning a project for which it later turns out that data are not available. This is only one example of the kind of circumstance that can make a problem unresearchable.)

2. Select a research project that is not too global. In other words, make sure that it is one that you can deal with in the time you have available and in terms of your resources.

3. While you should not lean too greatly toward the simplistic kind of problem, make certain also that the problem is within the sphere of your own competencies as an investigator. The only exception to this would be if you might have unusual resources or experts available to help you.

4. In this doctoral program, you are also expected to select research problems, the investigating of which is likely to be of use to you in your work or at the very least, to your institution. This means that this should be a problem for which a solution is <u>needed</u>.

5. Once you have identified such a problem break it down into its constituent parts. Unless you undertake a research problem appropriate to more than one module (for instance, governance and this module), you are unlikely to have the time to investigate a major problem in its entirety. Breaking a problem down into constituent parts will help you identify any portion of it which you can investigate in the time you have available and which will still be of significance to you and your institution.

## FORMULATION OF A TESTABLE HYPOTHESIS

This matter is covered in some detail in later units dealing with preparing a research proposal and treatment of the data. As a result, it will not be discussed at this juncture. (This is also true of a number of other steps involved in planning a research project; only these not outlined in detail in those sections are discussed here.)

CONSTRUCTION OF THE RESEARCH DESIGN

There are certain fundamental points about the construction of a research design which should be mentioned here, although they cannot be handled in the detail that they deserve and further information about them should be obtained from the reference books listed at the end of this manual.

Most careful consideration should be given to the matter of the population involved with the study, the selection of a sample, and the information which these data are to provide or which is to be ascertained about them. Although these are discussed at some length in a later section, some brief definitions are in order at this point.

In research, the terms population and sample have different meanings. Your population is the total group existing (for instance, all ninth-grade students in a high school); your sample is that smaller group from your population to be included in your study (for instance, every fifth ninth-grade student on the school rolls). The technique utilized to select a sample is critical and will be discussed in a later unit.

As mentioned earlier, it is often difficult to obtain valid samples in educational settings, and this must be expressed as one of the limitations of your study when it occurs.

In educational research, we often obtain our data from school records, test results, semester grades, and so forth. Should we ask people to provide for us certain information, it is important that we survey people only on matters about which they actually can provide useful data. There are other important points to keep in mind in designing questionnaires; it is a more complex matter than the person inexperienced in this field may recognize.

Common errors in questionnaire design include the following: asking imprecisely worded questions; using terms which are not carefully defined and which are open to different interpretations; asking for irrelevant data which will contribute nothing to the purpose of the study; using open-ended questions far too frequently; making open-ended questions so general that it is almost impossible to classify these responses.

In dealing with the results of the questionnaire, there are two other cautionary points which should be mentioned: (1) When people tell us about their plans, this should simply be reported as that, and not listed as something they are certain to do; (2) When people tell us about their behavior, this also should be reported in these terms and not as something that the research has shown they have actually done.

Finally, in planning our research, we must deal with variables -- and in educational research there are three types we are most likely to become involved with: (1) Independent variables (categorized as input, manipulated, treatment or stimulus); (2) Dependent variables (categorized as output, outcome or response); (3) Control variables (categorized as background, classificatory or organismic). There are also intervening variables but they are unlikely to arise in simple research projects. These are cited here for reference only and will be explained later in greater detail.

You will also be asked to deal with matters of validity. Internal validity refers to the effect of the treatment of your data. External validity refers to the generalizability of this treatment to other populations.

There is further information about these matters in the unit on Treatment of Data; but even there it is handled in relatively brief detail and, while these are matters which will be discussed in the seminars which make up this module, you will probably want to read about them as well in the reference books listed at the end of this manual.

## INSTITUTIONAL RESEARCH

The following are some kinds of institutional research about which studies are frequently conducted in community colleges -- and although this list is not complete, it may help stimulate your thinking along these lines:

(1) Faculty Data
- Workload studies
- Faculty characteristics
- Teacher evaluation

(2) Student Data
- Enrollment reports
- Student attrition
- Enrollment projections

(3) Curriculum Data
- Course loads
- Transcript analysis
- Course models

(4) Financial Data
- Program classification
- Income and expenditures
- State system controls

(5) Affirmative Action Data
Collection of data
Analysis
Institutional uses

## INSTRUCTIONAL RESEARCH

The following are some examples of the kinds of instructional research frequently conducted in community colleges:

(1) Comparison of the attrition rates between various departments or courses.

(2) Analyses of academic achievement and classroom attendance.

(3) Analyses of the relationship of grade point averages (or certain course scores) to future success in college (or to success in certain programs.)

(4) Comparison of grades students achieve and how they rate the effectiveness of an instructor.

(5) Analyses of the effects of certain demonstrated expectancies to college success.

As was the case with the topic cited under institutional research, these are cited as examples only -- to provide stimulus for those have difficulty in identifying research topics.

## STUDENT BEHAVIOR RESEARCH

While this is less generally undertaken in community colleges than either institutional or instructional research, in recent years there has been considerable interest in such topics as student activism, drug usage, and alcohol usage. In fact, some of the research conducted in these other two categories has overlapped with this field.

One of the difficulties with research into student behavior is that it is sometimes difficult to attain agreement on the meaning of certain crucial terms (for instance, "habitual drug usage" or "above-the-average alcoholic consumption" or, for that matter, any type of "deviant behavior").

In this doctoral program, it is quite possible that many persons coming from the student personnel services fields will be interested in this kind of research, particularly as it relates to the student and his college setting.

## IV. PREPARING A RESEARCH PROPOSAL

Below there is outlined a format for a research proposal which is appropriate for educational research projects.

When completed in sufficient detail, this will provide to anyone reviewing the proposal all of the information he or she will need to evaluate it.

Increasingly the National Lecturers who teach this module are allowing the same project to serve as both the research activity conducted to meet the seminar requirements as well as the practicum requirement. The evaluator at Nova will accept this format for the practicum proposal for this module as well. Thus, this format will meet both requirements.

The following are the headings used in this format:

I. THE TITLE

This should simply and clearly identify the problem to be investigated. A frequent trouble with titles created by inexperienced researchers is that they are too long, complex and unclear. It is a truism in research that if you cannot state your title in less than twenty words, you may well have not yet thought it through as clearly as you should.

II. THE STATEMENT OF THE PROBLEM

This expands upon the title, subdividing it into its major components or sections (if these are present).

A common error is to include a justification of the problem under this heading. This is inappropriate, since that material comes under a later heading (Background and Significance of the Study).

All that is required under this heading is that you (1) expand upon the title to the extent that the various elements of the problem are clearly identified and (2) answer important questions about the nature of the problem which a reviewer might raise.

## III. THE HYPOTHESIS

The hypothesis is a statement of the outcome which you anticipate in this investigation. It is always in declarative sentence form.

An example of a simple hypothesis would be: "Group study contributes to higher grade achievement." Thus, your hypothesis will represent a statement which you intend to test in your study.

Often this is a statement about the relationships between variables. It is always a simple statement of what you expect to be shown; it will not involve such words as "should" or "ought"; and it will not be stated in terms or morals or ethics.

This brings us to the matter of the Null Hypothesis. In the unit on Treatment of Data, there are a number of research problems outlined, all of which make use of the null hypothesis. You should examine these now to gain an understanding of how such hypotheses are worded; you will see why the null hypothesis frequently is called the "no-difference hypothesis."

To be tested, any hypothesis must be expressed as a null hypothesis. Therefore, each hypothesis you use in this module must be expressed in that way.

In order to state a hypothesis your problem must be clarified to the extent that you can make a statement concerning it which can be tested and which can be shown to be true or false.

The importance of the hypothesis in research literally cannot be underestimated -- since the nature of the hypothesis directs the investigation you will undertake and has a great deal to do with telling you what to do. Many inexperienced researchers spend too little time on the development of the hypothesis. But it deserves considerable thought and contemplation.

As a final point, in any project there can be more than one hypotheses. In any particular project you should use only the number of hypotheses that you can investigate in the time available and with the resources at hand.

## IV. BACKGROUND AND SIGNIFICANCE OF THE STUDY

This is the section in which you justify the importance of undertaking this project. The research problem should be discussed in sufficient detail to indicate to a reviewer who may be unfamiliar with this situation exactly why it is worth investigating and what the uses of the research will be.

With a relatively small research project, it may happen that the major use of the results will be to make a contribution to larger studies or to pave the way for further small ones.

But even a small research project should resolve a significant question (or questions) through the testing of the hypothesis (or hypotheses) involved.

This is also the section of your research proposal where you report your survey of the literature relating to this problem area -- and provide whatever background is available concerning other research conducted in this field.

Too many inexperienced researchers handle this section too lightly. As a result, a reviewer is likely to have a "so what?" reaction to the proposal. Unfortunately, this happens even with very worthwhile projects, simply because their importance is not written up in sufficient detail.

Finally, this is a section that should be handled in very specific terms. Avoid broad generalities; refer to specific studies; provide exact statistics and facts about the problem whenever these are available.

## V. DEFINITION OF TERMS

This section should include a definition of every important term which you will use in your research proposal, as well as every term about which a reviewer may have even the slightest question.

Too frequently we assume that any reader will understand what we mean by such terms as "part time student", "full time student", "adult student", etc. But these can have different meanings in different situations and should always be very carefully defined.

## VI. LIMITATIONS OF THE STUDY

Under this heading you should very clearly state the boundaries and limitations of your project. These are any circumstances or conditions which will place certain restrictions upon the reliability, validity, applicability or general usefulness of your study. Examples of these which may occur in educational research include: a smaller sample than would have been preferable, less-than-desirable response to a questionnaire, less-than-desirable sampling techniques, or inability to generalize beyond the particular program or institution involved.

Sometimes the inexperienced researcher will attempt to minimize these limitations or will hesitate to discuss them in any detail, because of a fear that this will take away from the value of the project. However, if limitations exist, they should be stated among other reasons, because the experienced reviewer will probably recognize them in any event and will wonder why they are being covered up.

There are limitations involved with every research project -- and clearly identifying them contributes to the overall meaning and understanding which you will want your study to achieve.

The inexperienced researcher may particularly overlook that his study will have application only to a particular institution or a particular part of that institution. This should be carefully thought about and where such limitations exist they should be stated. Since the majority of small research studies do have this limitation, this in no way indicates that the study should not be carried out -- if it is important to this institution it is still important enough to undertake.

VII. <u>BASIC ASSUMPTIONS</u>

In undertaking your project you make certain assumptions without actually testing them yourself or even having definitive evidence concerning them.

Typical assumptions of this kind might be that two students groups you are studying are reasonably homogeneous; that a student body involves normal distribution of intelligence; or that certain student groups have equal opportunities to achieve certain standards. What this means is that in the absence of any evidence to the contrary and in view of what is generally assumed about these groups, you are making these assumptions for purposes of this study.

Make certain that <u>each</u> such item is clearly identified under this heading.

VIII. <u>PROCEDURES FOR COLLECTING DATA</u>

This section should very clearly state the exact procedures you will use in assembling the data which will provide the results for your study.

Among the information to be included will be: what data you intend to collect; the sources from which you will obtain the data; the instruments you will use in collecting the data; other details about how you will assemble this data; the time span in which you will obtain it.

This is one of the most significant sections of your research proposal and the more specific the details provided in it, the better.

## IX. PROCEDURES FOR TREATING DATA

In this section you are to indicate what you will do with the data once you have assembled it. These techniques for treating the data obviously must be ones which will test the hypothesis or hypotheses outlined in a previous section of your proposal.

With inexperienced researchers, this treatment too often goes only halfway -- that is, it does not completely test whatever hypotheses are involved. There is no point in using any statistical techniques which provide irrelevant data or which serve mainly to show off your knowledge of statistics. But whatever techniques for treating the data will contribute to the complete testing of your hypothesis must be included here.

Any statistical techniques which are described in a basic statistical textbook usually need only be identified by a name. However, any other techniques should be outlined in more detail -- including the purpose of each of these techniques and the formula involved.

The only acceptable way to outline the contents of this section is to do so in very specific terms.

These are the only headings you need use in your research proposal -- but there are certain other points that should be made about preparing a research proposal.

1. No statement of actual results from the proposed study should appear in the proposal -- since the assumption is that the research has not yet been done and that the actual outcome will not be known at the time you write your proposal.

2. Most of the same headings used in your research proposal can be used in your research report, as can the material appearing under those headings.

3. The more precise you can be in preparing your research proposal -- involving everything from use of language to exact investigative instruments and methods of data treatment -- the more successful your proposal is likely to be.

4. While it is difficult to identify the exact length that such a proposal as this will involve, the usual research proposal prepared for this module -- written in sufficient detail but not at too great a length -- will probably average about ten double-spaced typewritten pages.

5. Whenever a research proposal is approved, it is invariably assumed that the participant will carry out the project exactly as proposed. This will always be an important factor in any evaluation.

6. Most sections of the research proposal should be written in the future tense.

## V. THE TREATMENT OF DATA: DESCRIPTIVE TECHNIQUES

You will want to present your data in the most effective ways possible. There are various descriptive techniques you can use for this -- some more dramatic than others and some better related to particular purposes than others.

While sometimes it will require more effort to present data through these techniques rather than in narrative form, persons reviewing your research projects will usually be able to respond more quickly and with greater understanding if the data are presented in a descriptive manner.

One very important point to keep in mind is that when you utilize descriptive techniques you only describe what currently exists. Usually it is not possible to use these techniques to engage in prediction or inference.

This unit deals with three principal categories: tables and graphs; central tendency; variability and the dispersion.

### TABLES AND GRAPHS

Often your data will consist of a great many numbers. Unless these numbers are presented in some organized way, it may be literally impossible to understand them.

A preliminary step to preparing a descriptive table is to group or summarize your data according to some meaningful plan. An example of a sample table is as follows:

Table I

THE NUMBER OF GRADUATES FOR THE YEARS 1968 THROUGH 1972 ACCORDING TO SEX

| Year | | Male | Female |
|---|---|---|---|
| 1968 | | 246 | 227 |
| 1969 | | 273 | 244 |
| 1970 | | 288 | 253 |
| 1971 | | 289 | 260 |
| 1972 | | 250 | 226 |
| | Total | 1346 | 1210 |

## Frequency Distribution

Very often the data with which you will be concerned will be a group of numbers such as student test scores. An example -- one which we will use in subsequent cases as well -- involves a group of test scores for thirty students. In arranging these test scores according to some logical order, the best idea may be to place them in descending order: 50, 50, 48, 48, 47, 47, 46, 45, 44, 44, 44, 43, 43, 43, 43, 42, 42, 41, 41, 40, 40, 39, 39, 39, 38, 38, 37, 36, 34.

When the data you are working with represent a large number of cases, it is preferable to summarize these by distributing them into classes or intervals. When this is accomplished you will have developed a frequency distribution.

The procedures for developing a frequency distribution is as follows:

1. Determine the interval size.

   A convenient way to establish the interval size is to find the difference between the highest score and the lowest score (50 - 34 = 16), then divide the remainder by the number of intervals you desire (usually between 5 and 20, depending on the data).

Let us assume we wish to have 5 class intervals. In this case, we will divide 16 by 5 and have a resulting quotient of 3.2.

Accept the nearest odd number to this as your class size. In this case, the class size will be 3.

2. Establish the upper and lower limits of the interval. Obviously, you want to choose a top interval that will include the highest score (50). You have three possibilities available:

48 - 50
49 - 51
50 - 52

You should select as the lower limit a number that is a multiple of the interval size. In this case the number 48 is a multiple of 3. Therefore, the choice for the top interval is 48 - 50.

3. Establish all other intervals.

In establishing the other intervals there are three points to remember -- all intervals must be the same size and there must be no overlap and finally, you must have an interval to accomodate each score in the array. This results in the intervals being:

48 - 50
45 - 47
42 - 44
39 - 41
36 - 38
33 - 35

4. Record the frequency (number) of test scores in each interval. Since the top interval contains all scores between 48 and 50, the frequency in this interval is 4 (50, 50, 48, 48). Continue this process and the frequency distribution will result as:

TABLE 2

FREQUENCY DISTRIBUTION OF 30 TEST SCORES

| Interval | | Frequency |
|---|---|---|
| 48 - 50 | | 4 |
| 45 - 47 | | 5 |
| 42 - 44 | | 9 |
| 39 - 41 | | 7 |
| 36 - 38 | | 4 |
| 33 - 35 | | 1 |
| | Total | 30 |

The frequency distribution is a very useful device for allowing the reader to quickly determine characteristics of the distribution of test scores.

Graphs:

These same test scores may be presented by producing "graphs." In fact, the graphical presentation of data is often the most effective way to illustrate differences.

There are two types of graphs commonly included in research reports. They are the histogram and the frequency polygon.

The histogram is constructed by forming a set of rectangles or bars, which have their bases on the horizontal axis of the graph. The center of the bar corresponds to the specific test score -- or, in the case of intervals, the center is the mid-point of the interval (with the interval limits being the left and right edge points of the bar).

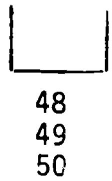


Figure 1. Construction of a bar for a histogram.

The length of the bars equals the frequencies in the intervals. Thus, a histogram describing the 30 test scores appears as:

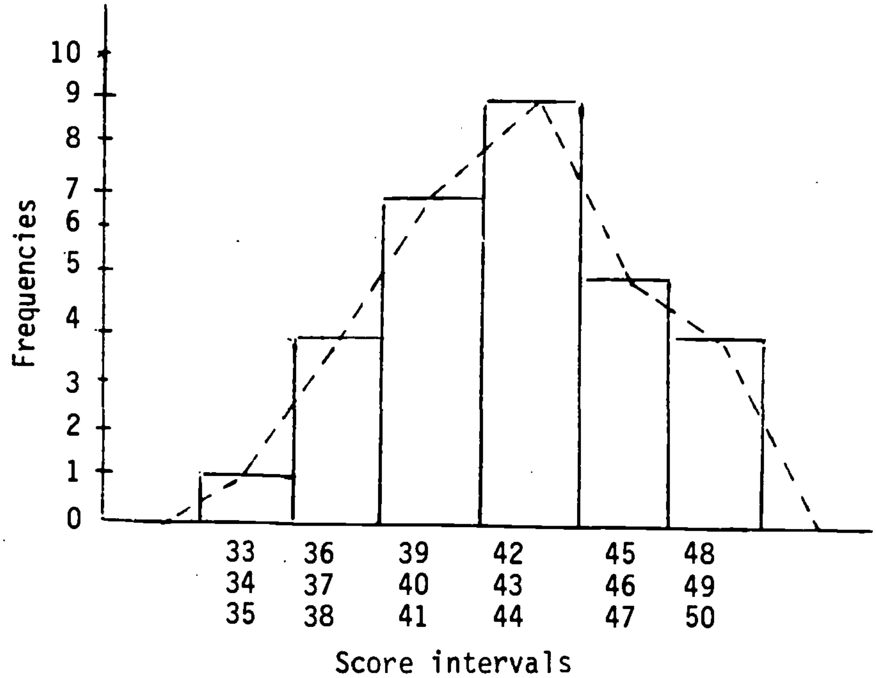


Figure 2. Histogram of the 30 test scores.

In particular, a histogram dramatically illustrates the differences in the frequency of test scores.

The second type of graph often used in research is the frequency polygon. This is created by merely connecting the tops of each bar -- with straight lines -- at the midpoints. This is illustrated by the dotted lines in Figure 2 (on page 29).

The frequency polygon is closed by connecting the midpoints of the left-and right-most bar to the base line, at a point that would be the midpoint of the next interval if it existed.

## MEASURES OF THE CENTRAL TENDENCY

Measures of central tendency allow us to reduce a large group of numbers to a single figure.

These measures are commonly referred to as averages. Averages make possible the easy and rapid comparisons between groups of numbers.

Of measures of central tendency, there are three that you should be especially familiar with: the median, the mode and the mean.

The median is simply the middle number when a group of numbers have been arranged in ascending or descending order. This means that there are as many numbers above the median as there are below it.

Consider these five numbers: 1, 2, 3, 4, 5. The median of this array is 3 because there are two numbers above it (4 and 5) and there are two numbers below it (1 and 2).

If there is an even number of figures in an array, the median is located half-way between the two middle numbers. For example, in the array 1, 2, 3, 4 the median is half-way between 2 and 3 -- thus the median is 2.5.

The median is a very useful measure because its value is not related to the magnitude of the numbers in the array. It is not affected by extreme cases when there are numbers of great or small magnitude. (The median of the array of test scores on page 26 is 43.)

The mode is the number that occurs the most times in any group (or in other words, it is the number that has the greatest frequency). In the array 1, 2, 2, 3, 4, the mode is 2 because it has a greater frequency than any of the other numbers.

Occasionally an array may appear as: 1, 2, 2, 3, 3, 4. In this case there is a tie; here there are the same number of 2's as 3's. This is referred to as being bimodal or having two modes (2 and 3). Obviously, some arrays of distributions may have any number of modes. (The mode of the array on page 26 is 43.)

The mean is the most commonly used of all averages; you know it as the arithmetic average of numbers. For the array on page 26, it is calculated by finding the sum of the scores and dividing this by the number of scores in the array. The sum of the 30 test scores on page 26 is 1276 and when this is divided by 30, the mean is found to be 42.53.

At this point you should begin to acquire familiarity with the symbology of statistics. Most statistical concepts have symbols assigned to them. In the calculation of the mean you were concerned with four statistical concepts which often are designated by symbols:

| | |
|---|---|
| An individual number or case | $X$ |
| The sum ($\Sigma$) of the numbers | $\Sigma X$ |
| The number of cases in an array | N |
| The calculated mean | $\bar{X}$ |

Thus, the formula for the calculation of the mean is:

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

In the above discussion, the calculation of the median, mode and mean was in terms of ungrouped or raw data. There are other techniques for calculating these measures when the data has been grouped as in a frequency distribution. These are not explained in this manual but are dealt with in the reference books listed in the bibliography.

MEASURES OF VARIABILITY AND DISPERSION

Once we have established the central tendency of a group of numbers, we have also established a point of reference relating to other numbers in that distribution.

By way of example, now that you know that the mean of these scores on page 26 is 42.53, you are able to better interpret a single student's score. If you were asked to meaningfully interpret an individual student's score without knowing the mean, you would not be able to do so. At least, when given the mean, you know whether that score is higher or lower than the class average.

But the mean is only part of the information required to interpret a single student's score. When you know only the mean, you have no concept of how much higher or lower than the mean a particular score is, and you are unable to consider it in terms of its relationship to all other scores in the distribution. This relationship depends largely on the dispersion or variability of the scores.

One of the simplest ways of describing this spread of scores is to report the top score, the lowest score and the range. The range is merely the difference between the top score and the lowest score. You will remember the range was calculated for the 30 test scores when it was necessary to establish the interval size: (50 - 34 = 16).

An individual score of 42 has more meaning once one knows the mean, the highest score, the lowest score and the range. But we still do not know the exact relationship of this score to the distribution of scores -- as an instance, you do not know how many students had scores lower than 44.

The statistical concept that it utilized to approximate this characteristic of disperson is the standard deviation.

STANDARD DEVIATION

Standard deviation may be defined as the average of the deviations from the mean. This means that if, at least theoretically, we calculate the difference between each score and the mean and sum these differences, we then can divide by N to establish the average of the deviations from the mean. The statistical symbology for this calculation is:

| | |
|---|---|
| Each individual score | $X$ |
| Mean | $\bar{X}$ |
| The number of scores | N |
| A score deviation form the mean | $X - \bar{X}$ |
| The calculated standard deviation | s |

Above it was stated that, theoretically, this procedure would calculate the standard deviation. Realistically it will not because the sum of the deviations from the mean is always zero. In order to perform this calculation, then, it is necessary to square each deviation -- thus eliminating negative numbers.

Since the deviations have been squared for purposes of calculation, it is necessary that following the division by N, we should take the square root of the quotient.

The formula for this calcuation is:

$$s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{N}}$$

In utilizing this particular formula it is necessary to calculate the deviation for each score -- and when working with a very large N this task is enormous. For this reason, it is customary to find the standard deviation being calculated through the use of a raw score formula, which is:

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

The calculated standard deviation of the scores listed on page 26 is 4.02.

This value has particular meaning in interpreting a single score.

In order to interpret the meaning of the standard deviation, it is necessary to refer to a normal curve. A curve describing an array of scores is created by smoothing out the frequency polygon, thereby eliminating the peaks and straight lines.

The curve for these 30 scores would appear as:

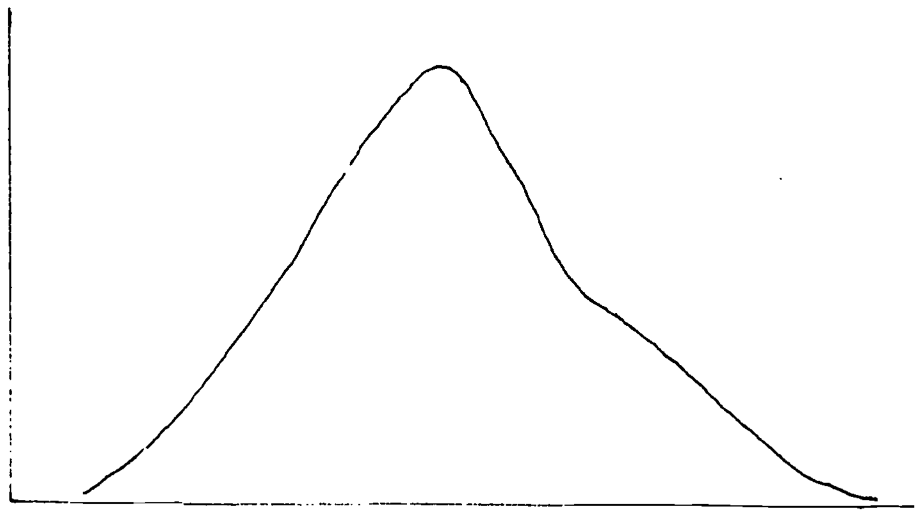

Figure 3. Distribution curve of the 30 test scores

This curve is not a normal curve; rarely do samples of 30 or less form a normal distribution. However, by knowing certain characteristics of a normal curve, it is possible to make approximations about this distribution.

The most meaningful characteristic of a normal distribution is that because it is symmetrical, it is possible to determine the areas under the curve that are above, below or between any points on the base line.

In almost all statistical textbooks there are tables of normal curve areas; the areas under discussion in the previous paragraph appear in these.

In terms of the standard deviation, it is known that 34.13% of the involved area is contained between the mean and one standard deviation. Another way to state this is that 34.13% of the scores we are considering would fall between the mean (42.55) and one standard deviation above the mean (46.55) if this were a normal distribution. This means that we can expect approximately 10 scores to be in this range. There are actually 10 scores between 42.53 and 46.53 in this distribution, even though it is not normal. It was accurate in this instance to assume the percentage of areas known to exist in a normal distribution.

Another fact that we can also establish is that 84.13% of the area is below the score that lies at one standard deviation above the mean. This means that approximately 25 scores will fall below a score of 46.55. (The actual number of scores from page 26 falling in this range is 24.)

At this point you have enough information about the central tendency and dispersion of the distribution of the 30 test scores to interpret the meaning of any student's score. You now know, for example, that a score of 44 is 1.47 points higher than the average score (42.53) and that it was higher than 18 of the scores earned. You are also in a position to make other calculations of a similar kind.

# VI. TREATMENT OF DATA: ANALYSIS

## PART I: CONCEPTS

This unit outlines a number of concepts and statistical techniques important in the analysis of data in such research projects as you are likely to undertake in this module.

As was true with earlier units in this guide, the topics covered here -- inference, sampling, variables, significance, Type I and Type II errors, parametric and non-parametric techniques; and prediction -- do not constitute a comprehensive group of all concepts you may encounter in analysis of data. As was also mentioned earlier, there is a certain amount of information which you must obtain from other books in this field; in addition, your National Lecturer may call your attention to other concepts, either by referring you to texts or by elaborating upon them in the seminars of this module.

Nevertheless, the concepts and techniques outlined in these pages are the ones most important to you in your treatment of data in educational research projects such as those discussed in this manual. Beyond these, the unit also contains one other very important feature -- several problems of the kind one might encounter in educational research, along with the statistical technique most appropriate for the handling of each problem and a step-by-step application of this particular technique. These problems are worked out in such detail that they constitute the entire second section of this unit.

INFERENCE

A great many research studies require that the investigator draw inferences about groups other than that one involved in a study.

As an instance, if an instructor discovers that a group of students experiencing a new educational strategy reach a higher level of proficiency than did another group who experienced a traditional educational program, the instructor might infer from the results that future students would reach a similarly higher level of proficiency with the new method of instruction. Often when we institute new programs system-wide or college-wide we do it on just this basis; from successful results obtained with a smaller group, we infer that the program will have applicability on a broader scale.

Innumerable other instances of a similar kind could be cited from the field of educational research. One of the most common occurs when it is impossible to examine all the cases appropriate to a particular study, as a result of which the investigator selects a sample of these cases and infers about the population from the results of this sample. As an example, if you wished to determine the mean grade point average for all the students who ever attended your college, it might not be possible for you to include the total group in your study. In this case, you could use an appropriate sampling technique to draw a sample of these students and then calculate the mean grade point average of the sample. By inferring from the sample mean, you could then estimate the value of the mean of the population as a whole.

In a previous unit, you were introduced to symbols as used in statistics. At this juncture, it should be mentioned that there are certain differences in the terminology which we use about population and sample -- most notably, that we speak of the statistics of a sample and, on the other hand, the parameters of a population -- but there are also important differences in the symbols involved:

| | Sample Statistics | Population Parameters |
|---|---|---|
| Mean | $\bar{X}$ | $\mu$ |
| Standard Deviation | s | $\sigma$ |
| Number of Cases | n | N |
| Proportion | p | $\pi$ |

So important is this entire matter of inference to statistics and research that there is a category known as Inferential Statistics (or Statistics of Inference). Much of the contents of this unit -- including the problems presented in the second half -- will involve inferential statistics. However, this will not be entirely the case, and you might make a project of establishing which of these concepts fall within this category and which do not, based upon the concepts outlined here and the additional material on the subject which you will find in other books.

SAMPLING

To put it most concisely, we use sample statistics to estimate population parameters.

But in order for these estimates to be acceptable and useful, they must be based on appropriate and proper sampling techniques.

Probably the most important point to remember in this connection is that the greater the size of a sample, the more likely you are to obtain a valid estimate of whatever population characteristics concern you. There are guide lines for determining the size of any sample; space limitations for this manual make it impossible to discuss them here but they are available in any comprehensive book on this subject.

Among the most important sample types -- all of them appropriate for certain kinds of educational research -- are the following:

The Purposive Sample

While this approach to sampling has its uses, it must be admitted that they are fairly limited -- largely because of weaknesses with the method of sample selection involved here and because no precise measure of the results may be made.

This kind of sampling takes place when, for instance, a researcher who knows well the community colleges in a given region selects as a sample of them several which have characteristics he considers "average" or typical." (Hence the alternative name for this kind of sample: a judgement sample.) His assumption is that if these colleges are typical with regard to the important items he can identify, then they will be typical with regard to other important items as well.

The difficulty with this method is that while it can be stated positively that such a sample will not possess the characteristics of the population, there is no way of establishing that it will be representative. Therefore, it is not possible to assess objectively the precision of the estimates made about this population in any study.

Despite its limitations, this method is used in the field of educational research and you will be likely to encounter such studies in your reading. The important qualification with regard to these studies is that the quality of the results can only be viewed in terms of whatever expert knowledge of a given situation someone has exercised in selecting the sample.

The Random Sample

When a random sample is drawn, each member is selected on the basis of chance; there are no determining factors in the selection; and each person in the population has an equal chance of becoming part of the sample.

As an illustration, if you were to draw a sample from the list of currently enrolled students in your college, you could write each student's name on a piece of paper, which you would then place with all other such names in a container. After shaking the container, you would draw enough names to make up the total sample required. This would be a cumbersome procedure, but it would meet the requirements of a random sample.

On the whole, the best way to select a random sample -- and normally, the most efficient -- is to utilize a table of random numbers. When you make use of such a table, each number (or member of the sample) will have genuinely the same possibility of being picked as any other.

The Systematic Sample

This is an approach you may want to use when you seek a sample equal in size to a certain per cent of the population involved. Assuming you wish your sample to include ten or fifteen per cent of all the students enrolled in your institution, you may then simply draw every tenth or fifteenth name from an alphabetical list of such students.

Usually a sample drawn in this way from an alphabetical list is considered a random sample.

The Stratified Random Sample

In educational research, it is often possible to divide the population involved into homogeneous groups. In fact, it is often desirable to do this -- to stratify the population, for instance, so as to obtain particular data about such groups as transfer students, career education students, community service students, and so on. Still further categorization can also occur, so that, for instance, transfer students may be classified by major or career students by program.

Once the population has been divided into what the investigator sees as meaningful groups, a random sample can then be selected from each of these groups. The total sample that results will be a stratified random sample.

The Cluster Sample

There may be occasions when an investigator will find that he cannot efficiently undertake a conventional approach to sampling -- as happens, for instance, when the population is spread over a broad geographic area (when the usual random sample might be so dispersed as to make it an uneconomical venture) or when an up-to-date list of the population might not even be available. (This latter may happen when one wishes to conduct a study of community perceptions concerning an institution).

One way of dealing with this problem is to divide the community into geographic clusters. A random sample can then be drawn for each of these clusters. This approach has acquired considerable use in recent years, especially with large studies and/or those conducted in major metropolitan areas.

In summary, as you read different texts in this field, you will discover various approaches to sampling. There is not space available to discuss other types; but before the subject is left, it should be stressed that whatever the type of sampling upon which you decide for a project, it must be carried out very rigidly and strictly according to whatever conditions are outlined for the method. A project can only be as good as the quality of the sampling involved.

## VARIABLES

Whatever the educational research projects that you consider in this module, the chances are that you will be dealing with variables.

For instance, you may be interested in discovering what relationship might exist between students' entrance test scores and their grade point averages. In this case, you are said to be dealing with two variables -- the entrance test scores and the grade point averages. The entrance test scores are identified as the independent variable; the achievement level (the grade point average) is the dependent variable.

This is the case because of two factors: (1) The value or occurrence of the dependent variable is considered to be dependent on the independent variable. (2) Consequently, as the independent variable changes so does the dependent variable. Both these conditions for being denoted as the independent variable, on the one hand, and the dependent variable, on the other, are met by the entrance test scores and the grade point average.

In this particular case, we do not yet know the extent to which the dependent variable will change as the independent variable changes. To study this relationship will be the purpose of our research study.

But unfortunately, variables are not all this easily observed or controlled. What if students with low entrance test scores consistently achieve high levels of proficiency in course work? Could intelligence or motivation be the variables that are related to high achievement -- and could the low test scores be the result of such factors as invalid tests or "test anxiety" on the part of these students? You can see how complex this kind of circumstance can become.

Variables have been classified in a very convenient way by Bruce W. Tuckman of Rutgers University and others. An adaptation of this categorization follows:

The Independent Variable

The independent variable is one which, when manipulated or observed, causes a change in the observed results. Independent variables are input variables related to a future phenomenon that may be observed or measured.

The Dependent Variable

The dependent variable is the response or output phenomenon that changes as the independent variable is manipulated.

The Moderator Variable

It may happen that the relationship between the independent variable and the dependent variable will not be as directly related as we would prefer. This may be the result of moderator variables affecting the dependent variable.

As an instance, in considering the relationship between entrance test scores (the independent variable) and achievement scores (the dependent variable), it may be found that for certain students reading ability or intelligence may have more effect on the dependent variable than the identified independent variable. Moderator variables do not necessarily affect all members of a group.

The Control Variable

In order to reach the most meaningful conclusions possible in any study, it is necessary to control as many variables as one can. Thus, it may be desirable to insure that all the students in a particular study are of the same sex, of approximately the same intelligence, or of about the same achievement level. When you do this you are attempting to control variables. These variables (sex, intelligence, achievement) are the control variables.

## The Intervening Variable

The intervening variable is one that cannot be observed or measured. (Those identified previously in this unit can be.) It is important for the person undertaking research to recognize intervening variables and to identify them in both the research proposal and the research report.

An example of an intervening variable may be the interest (or lack of it) of a student as he completes a test, the score of which is that student's observed behavior in a research study.

## SIGNIFICANCE

All statistical analysis rests heavily upon the concept of significance.

Put most simply, significance deals with the probability that something could occur, in either of two ways; (1) It will occur because of chance. (2) It will occur because of something else taking place and thereby bringing on this result.

There are a variety of ways to establish significance -- and there are numerous tests of analysis to seek it. (There are a number of these tests included in the second section of this unit.)

Basically, all significance is related to the area contained under a curve. This may be the normal curve or the curve involved may be of a different kind -- for instance, the curve involved in the t distribution, or the chi square distribution, or the F distribution. (Each of these is dealt with in the second section of this unit also.) But in any case, every time we seek significance in a statistical analysis, it is related to the area contained under such a curve.

Along these lines, one of the most readily comprehended tests is the z - test. In such a test, you establish what is called the critical z by deciding what level of error you will tolerate. This error becomes the level of significance. Thus, if you decide that five per cent of the time you could allow error to happen by chance and yet still accept the results as meaningful, a statistician would say that you are testing this at the.05 level of significance.

Even when you plan your research study with the utmost care, there is always the possibility that errors will occur -- and at this juncture, we should probably mention the Type I Error and Type II Error. These are widely known in research and frequently referred to in statistics textbooks. They also frequently crop up in connection with our decisions regarding the level of significance.

Simply stated, if we reject a hypothesis which should be accepted, we say that a Type I Error has been made. On the other hand, if we accept a hypothesis which should be rejected, we say that a Type II Error has been made. In either case an error or wrong judgement has occurred.

There are more elaborate ways of expressing the Type I Error and the Type II Error -- and you will find them in the statistics texts which you consult during this module -- but this summarizes what is involved with these two kinds of error.

Even when you plan your research study with the utmost care, there remains the possibility that each type of error may occur, although your research design should be planned to reduce the possibility of either to the greatest extent possible.

As was mentioned earlier, when you plan your research project, you will state a null hypothesis which you will test. Although null hypotheses may be stated in many terms, it is most common for them to be put in one of two ways: (1) There is no difference between two observations or (2) One observation is greater or less than another.

When attempting to establish significance, it is important to understand the differences between these two types of statements. For example, if you attempt to determine if two means are significantly different you will be concerned with a completely different amount of area under the curve than you will be if you are testing whether one mean is greater than another. In the former case -- the difference between means -- you will engage in what is called a two-tailed test. When you engage in the latter -- whether one mean is greater than another -- you are involved in a one-tailed test.

Whether or not a particular statistical test is a one-tailed or two-tailed determines in what portion of the curve the area representing your acceptable error will be located. By way of example: if you have elected to do a two-tailed analysis at the .05 level of significance, the area under a curve with which you will be concerned will be:

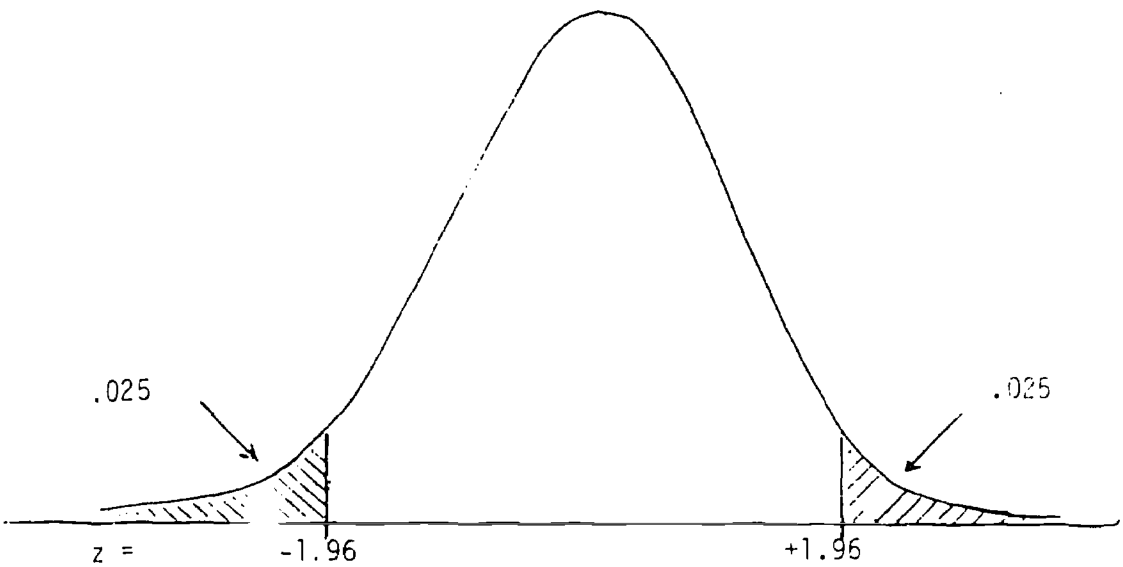


Figure 4. Two-Tailed Test

Thus, you accept the .025 of the area under the curve at both ends. Corresponding to the point on the base line that designates .025 of the area will be a quantitative value.

In the case of the normal curve this value is called z, and in the case of .025 the z value is 1.96. This, of course, means + 1.96 and -1.96, as indicated in Figure 3 . If in your calculations you find a z value that exceeds the critical z (1.96) in either direction, you will reject your null hypothesis -- and interpret your results as meaning that the difference is significant and that this difference could only occur 5% of the time due to chance.

If you had elected to do a one-tailed test to determine if one mean is greater than another you would have been concerned with the area under the curve as indicated on the next page:

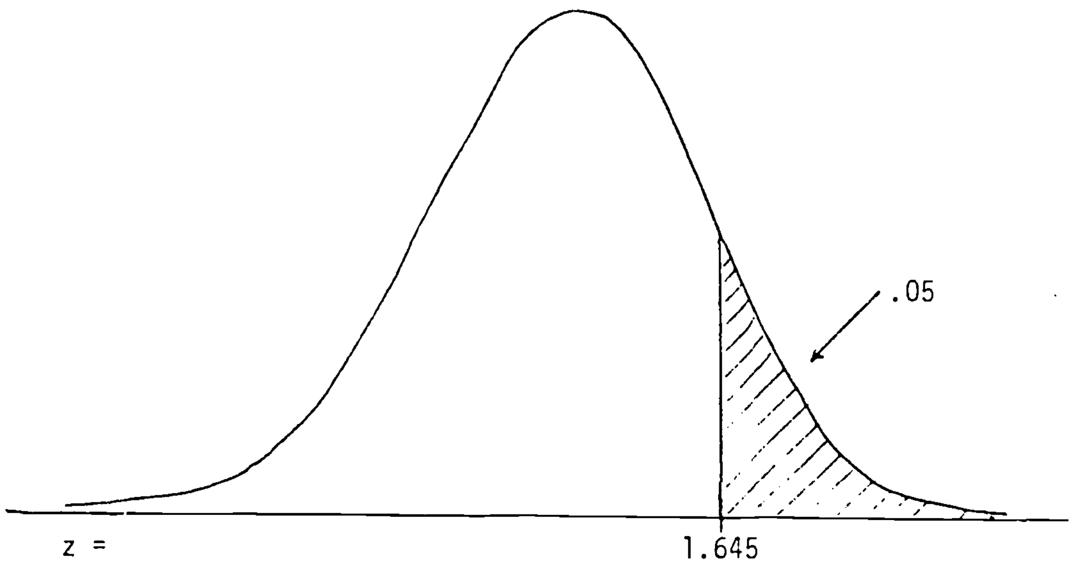


Figure 5. One-Tailed Test

The corresponding z value in this case would be 1.645. Therefore, if the calculated z exceeds the critical z you would again reject the null hypothesis and conclude that the fact that one mean is greater than another is significant. Since it could only occur 5% of the time due to chance.

In either of the above instances the calculated z might not exceed the critical z. If this had occurred you would have stated that you had failed to reject the null hypothesis and had accepted an alternative hypothesis representing the opposite of the null hypothesis.

The above example may not be entirely clear to the student who has not previously been exposed to research methodology.

Fortunately, other examples of this kind are likely to be presented in your seminars; they are also available in any elementary statistics textbooks; and there are further examples of this kind in the second section of this unit.

MEASUREMENT SCALES

A measurement scale is simply a means of assigning quantitative values to any variable. All data are reported in terms of some measurement scale.

There are four measurement scales: nominal, ordinal, interval and ratio. They are defined as follows:

Nominal Scales

The nominal scale is the least accurate of all the measurement scales, since there is no measurement involved except categorization. Thus, a dependent variable is expressed in terms of nominal values when a student's achievement is simply rated as "high" or "low" -- or "pass" or "fail" -- or "successful" or "unsuccessful."

To summarize, a nominal scale is merely a classification technique and there is no measurement involved.

Ordinal Scales

The ordinal scale is somewhat more precise than the nominal scale, since it does it does at least denote an "order of things." An instance of this will occur when the test scores of students are reported in terms of rank: first, second, third, fourth, and so on.

Interval Scales

The interval scale introduces further precision through showing the distance between two observations as well as denoting rank order. A familiar example will be that of test raw scores since we know that a score of 80 is higher than one of 65 and we also know how much higher it is (15 points).

In all probability, most of the measurement with which you are concerned will be expressed in ordinal form.

Ratio Scales

The ratio scale is primarily used in the physical sciences and will rarely be encountered in educational research. In addition to having all the precision of the interval scale, it also has a zero point, providing it with such characteristics as proportion. (Our number system is a ratio scale.)

By no means does this exhaust the list of characteristics of these scales; but these are sufficient for our purposes at this juncture. In turn, measurement scales lead us to the topic of parametric and nonparametric techniques.

## PARAMETRIC AND NONPARAMETRIC TECHNIQUES

All statistical tests are classified as either parametric or nonparametric. Primarily this classification is based upon the type of data in which the dependent variable is expressed. Since in any research undertaking you are expected to select the appropriate test, a knowledge of these will be important to you at this point.

It is when certain conditions of data are met that a statistical test is defined as being parametric. Similarly, when other conditions exist, the test will be nonparametric. These conditions may be summarized as follows:

Parametric Techniques

1. The further the curve of the distribution of the data is from normal, the less valid will be the application of a parametric test.

2. The more equal the variance of groups involved, the more applicable is a parametric test. (This is known as homogeneity of variance -- which may be determined by comparing the standard deviations or variances of the groups included in the study.)

3. The dependent variable must be described by data expressed in terms of an interval scale.

Nonparametric Techniques

1. The data describing the dependent variable does not have to involve a normal distribution.

2. It is not necessary to have homogeneity of variance between the groups included in the study.

3. The primary assumption when using nonparametric tests is that the measurement scales involved will be of the ordinal or nominal level.

## CORRELATION TECHNIQUES

In our research undertakings, we frequently wish to establish the predictability of an occurrence -- that is, the chances of it actually taking place.

To establish predictability, we use what are known as correlation techniques. Thus, in a community college we may make the assumption that more effective counseling of students will take place if we can establish a relationship between entrance test scores and student achievement. To establish the relationship between these, we use a correlation technique.

Basically, correlation is involved with two or more variables. Most often two variables will be studied: a predictor (or independent variable) and a criterion (or dependent variable) -- and our purpose in investigating these will be to determine if there is a relationship between them. If the predictor variable is increased, does the criterion variable also increase? In a study of entrance test scores and achievement test scores, is it the case that the higher the former the better chance the student has of demonstrating success through high scores on the latter?

The most common parametric correlation technique is the Pearson-Product-Moment correlation. Formulas for calculating this may be found in any elementary statistic textbook.

One of the most widely used nonparametric approaches to correlation is the Spearman rank-order correlation. This approach involves studying sets of ranks in a college setting, the Spearman rank-order correlation might be used in an attempt to predict the rank of any student on a post test from his rank on a certain pre-test.) Calculations involved here are somewhat more simple than in a parametric counterpart.

The result of any correlation study is a <u>coefficient of correlation</u>. All coefficients of correlation are expressed in terms of numerical values. They range between -1.00 and + 1.00.

Both the numerical values of -1.00 and +1.00 represent perfect relationships. In the case of a coefficient of correlation of -1.00, the criterion variable decreases as the predictor variable increases. In the case of +1.00, the criterion variable increases as the predictor variable increases. In both cases, there is a perfect correlation.

What of a correlation of zero? This indicates there is no relationship whatsoever. But usually the correlation coefficient will be a fractional number somewhere between -1.00 and +1.00.

Broadly speaking, the interpretation of positive correlation coefficients is as follows:

| | |
|---|---|
| Less than .20 | Very slight relationship |
| .20 to .40 | Definite but small relationship |
| .40 to .70 | Moderate relationship |
| .70 to .90 | High relationship |
| .90 to 1.00 | Very high and extremely dependable relationship |

As a final point, it must be stressed that correlation studies only measure that -- the relationship between two variables. They establish nothing about casual relationships; in point of fact, none may exist.

<u>SUMMARY</u>

This unit has attempted to introduce the various approaches available to you once you have gathered data for any research project. Since a manual of this length cannot include every possible stage in the analysis of such data, the emphasis has been upon the most important concepts associated with data analysis in educational evaluation and research.

To place these concepts in the proper perspective, a summary may be of help to you at this point.

This section opened with a discussion of inference. As was mentioned then, many of the research projects you will encounter will be concerned with inference. In order for you to draw valid inferences, you must have a valid sample. Several sampling techniques were discussed in this unit.

Variables were then introduced. In order to design a research project of an inferential nature -- once you have selected a sampling technique if required -- you must define and identify the variables in your study. You are then ready to select a level of significance -- or what margin of error are you willing to tolerate in accepting your results.

Next, measurement scales were introduced and defined. The purpose of introducing measurement scales at this point was that when selecting a statistical test your decision will be based primarily on the level of measurement at which your identified variables fall.

It was mentioned that the statistical tests are divided into two categories: parametric and nonparametric. It was pointed out that the selection of one or another type must be based on whether or not your variables are expressed in terms of certain measurement scales. ( No attempt was made to present the various types of statistical test available under each of these categories. However, there are five selected statistical tests contained in the next section.)

The final topic covered in this section involved correlation techniques -- a widely used technique in educational research. This presentation was to merely familiarize you with the concept as its total ramifications are too great to be covered in a manual of this type.

It is hoped that this brief introduction will provide you with some basis for your study of statistical analysis in educational research.

# VI. TREATMENT OF DATA: ANALYSIS

## PART TWO: PROBLEMS

This second section of Unit VI contains the step-by-step application of selected statistical tests to particular research problems.

While these represent only a few of the statistical tests available to us, these do represent commonly used problems in the field of educational research. These outlines are are intended to supplement what you will learn about these techniques in the seminars of this module and in statistical texts listed in the bibliography at the end of this volume.

The problems included here are entitled:

| | |
|---|---|
| Problem I | Testing The Significance Of The Difference between Two Sample Proportions |
| Problem II | Testing The Significance Of The Difference Between The Sample Means |
| Problem III | Testing The Significance Of The Difference Between Two Sample Proportions |
| Problem IV | Test of Independence |
| Problem V | Testing The Significance Of The Difference Between Four Means |

## PROBLEM ONE

## TESTING THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN TWO SAMPLE PROPORTIONS

### z - Test

Problem: A particular community college wants to determine if faculty and students agree or disagree concerning the need for student evaluation of faculty. Of 800 students queried, 475 responded in the affirmative and of 65 faculty, 22 responded in the same way. Was there a significantly higher proportion of students desiring this evaluation procedure than faculty desiring it?

Procedure:

1. State the null hypothesis ($H_o$): The difference between the proportions of students and faculty is zero.

$$H_o: P_1 = P_2$$

2. State the alternative hypothesis ($H_a$):

$$H_a: P_1 > P_2$$

3. Establish the desired level of significance:

$$\alpha = .01$$

4. Establish the critical z value: By referring to a table of areas of the normal curve, it is found that the critical z value for a one-tailed test is 2.58. Therefore, reject $H_o$ and accept $H_a$ if $z > +2.58$.

5. Apply appropriate formula:

$$z = \frac{P_1 - P_2}{\sqrt{\frac{\overline{pq}}{n_1} + \frac{\overline{pq}}{n_2}}}$$

Where:

$$P_1 = \frac{475}{800} = .5937 \qquad P_2 = \frac{22}{65} = .3384$$

$$\bar{p} = \frac{475 + 22}{800 + 65} = .5745 \qquad \bar{q} = 1 - \bar{p} = 1 - .5937 = .4225$$

$$n_1 = 800 \qquad n_2 = 65$$

$$z = \frac{.5937 - .3384}{\sqrt{\frac{(.5745)(.4225)}{800} + \frac{(.5745)(.4225)}{65}}}$$

$$z = \frac{.2553}{\sqrt{.0041}} = 3.99$$

6. State your conclusion:

Since the calculated value of z exceeds the critical value of z, the null hypothesis can be rejected. There is a significant difference between the proportion of students favoring faculty evaluation by students and the proportion of faculty favoring it.

## PROBLEM TWO

## TESTING THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN THE SAMPLE MEANS

### THE t - Test

Problem:

A teacher arranged for ten students to experience a video-tape presentation as a supplement to the program of traditional instruction. On an achievement test administered at the end of the four week experiment, the teacher found the mean test score of the students to be 45.5 and the standard deviation to be 3.5. Ten other students who experienced only the traditional instruction had a mean score of 42.5 and a standard deviation of 3.4 on the same test after this same four week period. Is there a significant difference between the mean test scores of these two groups?

Procedures:

1. State the null hypothesis ($H_o$):

   The two sample means do not differ significantly: In other words, they are samples from the same population.

   $$H_o : \overline{X}_1 = \overline{X}_2$$

2. State the alternative hypothesis ($H_a$):

   $$H_a : \overline{X}_1 \neq \overline{X}_2$$

3. Establish the desired level of significance:

   $$\alpha = .05$$

4. Calculate the degrees of freedom:

   $$df = n_1 + n_2 - 2$$

   $$df = 10 + 10 - 2 = 18$$

5. Establish the critical t-value:

   By referring to a table of critical t-values, it is found that for 18 degrees of freedom the critical t-value at $\alpha = .05$ is 2.101 for a two-tailed test. Therefore, reject $H_o$ and accept $H_a$ if $t < +2.101$ or $t > -2.101$.

6. Apply appropriate formula:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_1^2}{n_2}}}$$

where: $\overline{X}_1 = 45.5 \quad s_1 = 3.5 \quad n_1 = 10$

$\overline{X}_2 = 42.5 \quad s_2 = 3.4 \quad n_2 = 10$

$$t = \frac{45.5 - 42.5}{\sqrt{\frac{3.5^2}{10} + \frac{3.4^2}{10}}}$$

$$t = \frac{3.0}{1.54} = 1.95$$

7. State your conclusion:

Since the calculated value of t does not exceed the critical value of t, the null hypothesis cannot be rejected; there is no significant difference between the mean test scores of the two groups of students.

# PROBLEM THREE

## TESTING THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN TWO SAMPLE PROPORTIONS

### Chi Square ($X^2$)

Problem:

At the close of fall registration there were 641 male students and 442 female students enrolled in a particular institution. During this term 143 male students and 101 female students withdrew. Is there any significant difference between the percentage of males withdrawing and females withdrawing during this term?

Procedure:

1. State the null hypothesis ($H_o$):

   The bases of classification are independent; or the difference between the proportion of males withdrawing and the proportion of females withdrawing is zero.

   $$H_o: \pi_1 = \pi_2$$

2. State the alternative hypothesis ($H_a$):

   $$H_a: \pi_1 \neq \pi_2$$

3. Establish dividend of significance:

   $$\alpha = .05$$

4. Calculate the degrees of freedom:

   $$df - (r - 1)\ (c - 1)$$

   $$df = (1)\ (1)$$

   $$df = 1$$

5. Establish the critical $X^2$ value:

   By referring to a table of critical values of Chi Square, it is found that for 1 degree of freedom the critical $X^2$ value at $\alpha = .05$ is 2.71. Therefore, reject $H_o$ and accept $H_a$ if $X^2$ is $>$ 2.71.

6. Apply appropriate formula:

$$X^2 = \sum \left[ \frac{(f_o - f_e)^2}{f_e} \right]$$

Where:

$f_o$ = observed frequency

$f_e$ = expected frequency

7. The procedure for calculating the $X^2$ value is as follows:

. Construct a contingency table of observed frequencies:

CLASSIFICATION OF OBSERVED WITHDRAWAL RECORDS OF MALES AND FEMALES

$f_o$

| Classification | Withdrawal | Persisted | Total |
|---|---|---|---|
| Male | 143 | 498 | 641 |
| Female | 101 | 341 | 442 |
| Total | 244 | 839 | 1083 |

. Construct a contingency table of <u>expected</u> frequencies:

EXPECTED WITHDRAWAL RECORDS OF MALES AND FEMALES ASSUMING INDEPENDENCE OF CLASSIFICATION

$f_e$

| Classification | Withdrawal | Persisted | Total |
|---|---|---|---|
| Male | 144.2 | 496.6 | 641 |
| Female | 99.6 | 342.4 | 442 |
| Total | 244 | 839 | 1083 |

. Calculate $X^2$

CALCULATION OF CHI SQUARE

| Cell | Observed Frequencies $f_o$ | Expected Frequencies $f_e$ | $f_o - f_e$ | $(f_o - f_e)^2$ | $\frac{(f_o - f_e)^2}{f_e}$ |
|---|---|---|---|---|---|
| Male Withdrew | 143 | 144.2 | -1.2 | 1.44 | .01 |
| Male Persisted | 498 | 496.8 | 1.2 | 1.44 | .003 |
| Female Withdrew | 101 | 99.5 | 1.5 | 2.25 | .02 |
| Female Persisted | 341 | 342.6 | -1.5 | 2.25 | .01 |
| | | | | $X^2$ = | .04 |

8. State your conclusion:

Since the calculated value of $X^2$ does not exceed the critical value of $X^2$, the null hypothesis cannot be rejected; there is no significant difference between the proportion of males withdrawing and the proportion of females withdrawing.

## PROBLEM FOUR

## TEST OF INDEPENDENCE

Chi Square ( $X^2$ )

Problem:

The director of the humanities division in a community college developed concern about add-and-drop transactions occurring in this division. Included in the division were four departments, designated here as Departments A, B, C, and D. The director decided to maintain records of all drop-and-add transactions occurring in these departments throughout one particular academic year, consisting of the Fall Term, the Winter Term, the Spring Term, Summer Term I and Summer Term II. What the director was most anxious to learn was whether or not there was any relationship between drop-and-add transactions of the different departments according to the term involved.

Procedure:

1. State the null hypothesis ( $H_0$ ):

   The drop-and-add transactions in the different departments are independent of the different terms.

2. State the alternative hypothesis ( $H_a$ ):

   Drop-and-add transactions in the different departments are related to the terms.

3. Establish the desired level of significance:

   $\alpha = .05$

4. Calculate the degrees of freedom:

   $$df = ( r-1 ) ( c-1 )$$

   $$df = ( 4-1 ) ( 5-1 )$$

   $$df = ( 3 ) ( 4 )$$

   $$df = 12$$

5. Establish the critical value of $X^2$:

   By referring to a table of the distribution of $X^2$ it is found that for 12 degrees of freedom the critical $X^2$ value at $\alpha = .05$ is 21.026. Therefore, reject $H_0$ and accept $H_a$ if $X^2$ is > 21.026.

6. Apply the appropiate formula:

$$X^2 = \sum \left[ \frac{(f_o - f_e)^2}{f_e} \right]$$

Where: $f_o$ = observed frequency

$f_e$ = expected frequency

7. The procedure for calculating the $X^2$ value is as follows:

   . Construct a contingency table of observed frequencies:

CLASSIFICATION OF DROP-AND-ADD TRANSACTIONS
BY TERM FOR FOUR DEPARTMENTS

$f_o$

| Department | Fall | Winter | Spring | Summer I | Summer II | T |
|---|---|---|---|---|---|---|
| A | 550 | 450 | 525 | 155 | 125 | 1805 |
| B | 475 | 350 | 400 | 75 | 65 | 1365 |
| C | 390 | 400 | 250 | 90 | 70 | 1200 |
| D | 510 | 525 | 400 | 105 | 95 | 1635 |
| T | 1925 | 1725 | 1575 | 425 | 355 | 6005 |

   . Construct a contingency table of expected frequencies:

EXPECTED DROP-AND-ADD TRANSACTIONS ASSUMING
INDEPENDENCE OF CLASSIFICATION

$f_e$

| Department | Fall | Winter | Spring | Summer I | Summer II | |
|---|---|---|---|---|---|---|
| A | 578.6 | 518.5 | 473.4 | 127.8 | 106.7 | 1805 |
| B | 437.6 | 392.1 | 358.0 | 96.6 | [illegible]0.7 | 1365 |
| C | 384.7 | 344.7 | 314.7 | 84.9 | 1.0 | 1200 |
| D | 524.1 | 469.7 | 428.9 | 115.7 | 96.6 | 1635 |
| T | 1925.0 | 1725.0 | 1575.0 | 425.0 | 355.0 | 6005 |

. Calculate $X^2$

CALCULATION OF CHI SQUARE

| $f_o$ | $f_e$ | $f_o - f_e$ | $(f_o - f_e)^2$ | $\frac{(f_o - f_e)^2}{f_e}$ |
|---|---|---|---|---|
| 550 | 578.6 | -28.6 | 817.96 | 1.41 |
| 475 | 437.6 | 37.4 | 1398.76 | 3.20 |
| 390 | 384.7 | 5.3 | 28.09 | .07 |
| 510 | 524.1 | -14.1 | 198.81 | .38 |
| 450 | 518.5 | -68.5 | 4692.25 | 1.33 |
| 350 | 392.1 | -42.1 | 1772.41 | 4.52 |
| 400 | 344.7 | 55.3 | 3058.09 | 8.87 |
| 525 | 469.7 | 55.3 | 3058.09 | 6.51 |
| 525 | 473.4 | 51.6 | 2662.56 | 5.62 |
| 400 | 358.0 | 42.0 | 1764.00 | 4.93 |
| 250 | 314.7 | -64.7 | 4186.09 | 13.30 |
| 400 | 428.9 | -28.9 | 835.21 | 1.95 |
| 155 | 127.8 | 27.2 | 739.84 | 5.79 |
| 75 | 96.6 | -21.6 | 466.56 | 4.83 |
| 90 | 84.9 | 5.1 | 26.01 | .31 |
| 105 | 115.7 | -10.7 | 114.49 | .99 |
| 125 | 106.7 | 18.3 | 334.89 | 3.14 |
| 65 | 80.7 | -15.7 | 246.49 | 3.05 |
| 70 | 71.0 | - 1.0 | 1.00 | .01 |
| 95 | 96.6 | - 1.6 | 2.56 | .03 |
| | | | $X^2 =$ | 70.24 |

8. State your conclusion:

Since the calculated value of $X^2$ exceeds the critical value of $X^2$, the null hypothesis is rejected and the alternative hypothesis is accepted; drop-and-add transactions in the different departments are related to the terms.

## PROBLEM FIVE

## TESTING THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN FOUR MEANS

### Analysis of Variance

Problem: In an attempt to determine if there was a difference between four different individual directions which are given to students prior to them engaging in a learning experience, a teacher divided a class, consisting of 20 students, into four groups of five students each. Each group was given a different set of directions and then proceeded to complete the learning experience. An evaluation instrument, consisting of 25 questions, was used to measure the amount of learning experienced by the four groups of students. The results are as follows:

| Group | Test Scores |
|---|---|
| A | 24, 20, 23, 22, 19 |
| B | 23, 19, 20, 20, 18 |
| C | 24, 25, 20, 20, 21 |
| D | 24, 22, 21, 23, 19 |

1. State the null hypothesis ($H_0$):

   The four sets of student scores do not differ significantly from one another.

2. State the alternative hypothesis ($H_a$):

   The four sets of student scores do differ significantly from one another.

3. Establish the desired level of significance:

   $$\alpha = .01$$

4. Calculate the degrees of freedom:

   Between sets df = number of sets - 1

   df = 4 - 1

   df = 3

Within sets df = number of observations - number of sets

$$df = 20 - 4$$

$$df = 16$$

5. Establish the critical value of F:

   By referring to a table of the distribution of F, it is found that for 3 and 16 degrees of freedom, the critical F value at $\alpha = .01$ is 5.29. Threfore, reject $H_0$ and accept $H_a$ if $F > 5.29$.

6. The procedure for calculating F is as follows:

Calculate sums and means of the observation sets, the grand total $\Sigma X$ and the grand mean $\overline{X}_t$.

The Observations (X)

| | I | II | III | IV | | |
|---|---|---|---|---|---|---|
| | 24 | 23 | 24 | 24 | | |
| | 20 | 19 | 25 | 22 | | |
| | 23 | 20 | 21 | 19 | | |
| | 22 | 20 | 22 | 21 | | |
| | 19 | 18 | 21 | 23 | | |
| $\Sigma X$ | 108 | 100 | 113 | 109 | 430 | $\Sigma X$ |
| $\overline{X}_s$ | 21.6 | 20.0 | 22.6 | 21.8 | 21.5 | $\overline{X}_t$ |

Calculate the deviations of each score from the set means $(x - \overline{X})$.

Deviations Within Sets $(X - \overline{X})_s$

| I | II | III | IV |
|---|---|---|---|
| 2.4 | 3.0 | 1.4 | 2.2 |
| -1.6 | -1.0 | 2.4 | 0.2 |
| 1.4 | 0.0 | -1.6 | -2.8 |
| 0.4 | 0.0 | -0.6 | -0.8 |
| -2.6 | -2.0 | -1.6 | 1.2 |

Squares of Deviations Within Sets $(X - \overline{X})_s^2$

Square the deviations within each set and calculate the sum.

| | I | II | III | IV |
|---|---|---|---|---|
| | 5.74 | 9.00 | 1.96 | 4.84 |
| | 2.56 | 1.00 | 5.74 | 0.04 |
| | 1.96 | 0.00 | 2.56 | 7.84 |
| | 0.16 | 0.00 | 0.36 | 0.64 |
| | 6.76 | 4.00 | 2.56 | 1.44 |
| T | 17.18 | 14.00 | 13.18 | 14.80 |

$59.16 = \Sigma(X - \overline{X})_s^2$

Deviations of Set Means From Grand Mean ( d )

Calculate d for each set and square each d, and find $n\Sigma d^2$.

| | I | II | III | IV |
|---|---|---|---|---|
| $d$ | .01 | -1.50 | 1.10 | .30 |
| $d^2$ | .0001 | 2.25 | 1.21 | .09 |
| $nd^2$ | .0005 | 11.25 | 6.05 | .45 |

$3.55 = \Sigma d^2$

$17.75 = n\Sigma d^2$

The Total Variance Subdivided Into Two Components

Calculate the variances by dividing the sums of squares by their degrees of freedom.

| Components | df | sums of squares | Variance |
|---|---|---|---|
| Between Sets | 3 | 17.75 | 5 83 |
| Within Sets | 16 | 14.76 | .92 |
| Total | 19 | 32.51 | |

. At this point the F - ratio can be calculated. This is accomplished by dividing the between sets variance by the within sets variance:

$$F = \frac{5.83}{.92}$$

$$F = 6.34$$

7. State your conclusion:

Since the calculated value of F exceeds the critical value of F, the null hypothesis can be rejected. There is a significant difference; the four sets of student scores do differ significantly from one another.

## VII. PREPARING A RESEARCH REPORT

In preparing your research report, your first headings will be identical with those you have used in your research proposal:

I. THE TITLE
II. THE STATEMENT OF THE PROBLEM
III. THE HYPOTHESIS
IV. BACKGROUND AND SIGNIFICANCE OF THE STUDY
V. DEFINITION OF TERMS
VI. LIMITATIONS OF THE STUDY
VII. BASIC ASSUMPTIONS
VIII. PROCEDURES FOR COLLECTING DATA
IX. PROCEDURES FOR TREATING DATA

(The major difference between the research proposal and the research report is that you will have written in the future tense in the former, whereas in the research report much will be written in the past tense.)

### RESULTS

Following the above headings, you should present your research results. The exact headings which you will use here will depend very largely upon the procedures you have used both in collecting and treating the data. The two simplest ways of reporting your research results are to use the headings "Data Resulting from the Study" (the pure data) and "Significance of the Data" (interpretation).

Should there be a great deal of material under each of these headings, you may wish to further subclassify them -- and how you will do this will depend upon the kinds of data that you have to report.

PRESENTING THE DATA

As a rule, you should not present your data totally in narrative form. This only makes it more difficult for the reader or reviewer to pick out salient details. Instead, data should be presented through tables, graphs and other means which will allow the reader to grasp significant details.

In presenting your data, do not use computer printouts. Usually this is considered the laziest way of presenting data; it puts most of the burden of picking out significant points upon the reader. Naturally, it is appropriate to present tables selected from such printouts, and in some rarer instances you may want to use tables almost totally based upon them. But this latter should only be indulged in when it actually makes a contribution to the reader's understanding of the project. It should never be done simply for padding or to make the report appear more impressive.

When you use tables, diagrams or graphs, you usually need not repeat this information in the narrative. Rather, the narrative will supplement what already appears in such charts.

As a final point, only relevant data should be reported at this juncture. Interesting data which has been turned up in the study but which contributes nothing to the problem at hand can be reported under a later heading. (This is discussed later in this unit.)

CONCLUSIONS AND SIGNIFICANCE

Once your data has been reported, you should present the conclusions which can be derived from these data, the implications of them for the program or institution involved, and their significance for similar kinds of programs or systems.

At this point, the inexperienced researcher may find himself indulging in such expressions as: "The investigator feels that the data show..." or "Personally, the investigator thinks that..." This is totally unacceptable in any kind of research report. Conclusions reached should be based strictly on the data. No other conclusions should be included in a research report.

It is important also that only those conclusions that can be supported by the data which you have developed should be presented. This applies equally to any inferences which you may draw from these -- especially in terms of other institutions or programs to which you may suggest that the conclusions will also apply. When developing this section, it may be helpful to review the statement regarding the limitations of your project; in fact, a review of these limitations might well be appropriate at this point in your report.

On the other hand, your report should include at this juncture a detailed statement of the significance that your study does have for your own institution or program, since this is a major point of emphasis both in this particular module and in all of the research studies that you will conduct as part of this doctoral program.

## RESIDUAL FINDINGS

From time to time, a research project will turn up data which make little or no contributions to the questions about which the project itself revolves.

These findings may impress the researcher as being sufficiently interesting or worthwhile to deserve reporting. They should not be reported as part of the results of the study but rather, should be outlined under a separate heading. An appropriate designation for this would be: Residual Findings or Auxilliary Results. This identifies such findings as not relating to the purposes of the study but nonetheless, resulting from the study itself.

## FURTHER STUDIES

Should your project suggest that further studies are indicated, it is also appropriate to report these -- and, in fact, the more specific you can be about them, the better.

As one final word, the research report is always expected to stand by itself. You should not have to depend upon a covering letter or conversations with the reviewer to explain or clarify your study. Everything that will make your research project meaningful and comprehensible should be included in the final report.

# VIII. EVALUATION

The term evaluation has more than one meaning in the field of education. Among the most important of these are:

## RESEARCH AS EVALUATION

In a number of books that have been written in recent years, what we have called research in the foregoing sections has there been termed evaluation. In fact, there are books surveying little more than the applications of various statistical techniques which are termed by their publishers as "evaluation textbooks."

This is perfectly legitimate, providing one understands what evaluation actually means when used in this context -- that is, applied statistics.

Similarly, in many graduate schools courses in research and evaluation cover much this same ground. Here again, this is perfectly legitimate, providing one understands the context in which these terms are used.

## RESEARCH: THE TOOL OF EVALUATION

A more accepted viewpoint -- at least in terms of the lip service given it -- is that evaluation actually is a broader field with research as one of its tools. The same will apply to any statistics involved.

This viewpoint would hold, for instance, that when one undertakes a research project, the statistics one uses will only be a part of the research project and will be one facet contributed to the decision-making process. Further, when one has completed the research project, one has still not completed the evaluation process -- this requires yet further consideration of how the results of the research project can be applied to given situations within an educational institution or system.

EVALUATING THE RESEARCH OF OTHERS

Another aspect of this involves our evaluating the research of others. In fact, one of the objectives of this module is that you should develop skills of this type. Undoubtedly the most successful way for you to develop these skills is to acquire a real understanding of what constitutes a successful research project (or an unsuccessful one) when you yourself undertake it. The criteria that you apply to your own research will be the same as those which you apply to projects undertaken by others.

By way of summary, some of the most important points to keep in mind in evaluating research projects are these:

(1) Does the title of the research project actually reflect what it undertakes?

(2) Does the research problem under investigation have the significance which its sponsors claim for it?

(3) Does the research project actually measure what it is supposed to measure?

(4) Are the research techniques or investigative instruments involved appropriate to the problem?

(5) Are the conclusions reached from the study supported by the data?

(6) Would the person supporting the research have any particular bias toward the kinds of results which should be produced?

Too frequently when we read of research projects in the press, they actually involve failure to meet certain of these criteria but we are not informed of this.

To cite a few instances:

Recently a widely publicized study indicated that much network TV news broadcasting revealed certain kinds of political bias. What was not revealed was that the study was supported by a foundation with well-established political bias in the opposite direction.

For years, newspapers reported that the Neilsen TV Audience Studies measure "TV viewing." In point of fact they measure only TV sets turned on -- at which an entire family, one person, the household cat or no one may be looking.

A number of studies purported to reveal startling facts about I.Q.'s of certain population groups have been widely publicized -- but not so widely publicized has been the fact that sometimes the I.Q. tests involved have been of a type not well accepted in psychological circles.

The more one engages in research, the more one develops a questioning attitude toward research in general. This is a very healthy sign.

EVALUATION PROJECTS

In large educational institutions or systems, one can scarcely move today without encountering yet another kind of evaluation project.

These may be undertaken by the research personnel of an institution, by outside consultants, by department heads, or by indivduals pursuing advanced degrees or supported by some kind of government or foundation grant.

Similarly, these evaluation projects may focus on a wide variety of matters: for instance, the extent to which an institution is living up to community expectations; the extent to which an institution is living up to student expectations; the extent to which individual programs are living up to community or student expectations; the degree to which faculty attitudes support the goals of the institution; the degree to which faculty members and students participate in important decisions about the institution; the extent to which new instructional strategies are reaching goals set for them; and so on.

In such broad scale evaluation projects, research as described in this manual may constitute only a preliminary fact-finding stage. Following this there may be workshops, position papers, and other activities or presentations -- sometimes defined as research, but which are not of the formal, systemized type described in this manual.

In recent years educational institutions and systems utilizing such evaluation studies -- involving not only administrators but as well, faculty, students and representatives of the community -- have broadened participation in the process of evaluation to an extent not previously visualized.

Today we have certain types of evaluations built into ongoing college activities that are representative of this: student evaluations by faculty members (a traditional undertaking but now often handled in non-traditional ways); faculty evaluations by administrators; faculty evaluations by peer members; faculty evaluations by students; administrator evaluations by peer administrators; and perhaps most innovative of all, evaluations of administrators by faculty and students.

These developments have raised a number of interesting questions about who should do evaluations and what techniques should be involved in this process. These are questions which obviously cannot be answered in a manual of this kind. But it is important that we recognize that evaluation has gone far beyond what is suggested by the term in some of our more traditional texts and is now engaged in by persons in the institution and community to whom this role would not have been assigned only a few years ago.

SUMMATIVE EVALUATIONS

With many developmental or innovative programs -- particularly those on a large scale basis -- a summative evaluation is required at the completion of the project. Often this is written into the contract for the project by the sponsoring government agency or foundation.

In this connection the term "summative" can be readily understood -- since it simply means that at the conclusion of the project (or at a certain significant stage in it), a study is made summarizing achievements to date; the extent to which the objectives of the project have been reached; acceptance of the project by those it has been intended to serve; problems which have developed with the project or the degree to which it has not met its goals; probable reasons for such developments; and so on.

Almost invariably, such studies are conducted by persons or agencies totally outside the project itself. Among other reasons, this is to ensure objectivity in the conclusions reached. But sometimes these summative evaluative studies are so under-funded that there is not sufficient time for the persons conducting them to become familiar enough with all of the implications of developments within the project. A second criticism leveled against such studies is that they are too of ted at a point where they have little ability to influence the project or bring about what would have been desirable changes in it.

FORMATIVE EVALUATIONS

Evaluations of this type are conducted while a project is in process. Necessarily they must be undertaken by people more closely identified with the project than will be the case with agencies or persons conducting summative studies. On the other hand, they have the advantage of being able to influence the project itself through not only identifying problems but also doing so while there . still time to effect change.

In recent years some of the larger projects funded by government agencies or foundations have made provisions for both formative and summative types of evaluation.

----------

"If we had to take in our hand any volume...let us ask, Does it contain any abstract reasoning concerning quantity or number? No. Does it contain any experimental reasoning concerning matter of fact and existence? No. Commit it then to the flames: for it contains nothing but sophistry and illusion!"

David Hume: An Enquiry Concerning
Human Experience (1777)

# IX. BIBLIOGRAPHY

These are only a few of the books which can provide you with the kinds of information you will need to supplement the lectures of this module and this particular manual. Books listed below which are astericked (*) are included in the official bibliography for this module:

*Anastasi, Anne: Psychological Testing (3rd Edition). New York: Macmillan, 1968.

This is about as comprehensive a book as you can expect in this field. It runs to 664 pages and in addition to covering most types of psychological tests available today, it also deals with the interpretation of test scores, norms and their uses, reliability, validity, and item analysis.

Arkin, Herbert & Colton, Raymond R.: Tables For Statisticans (2nd Edition). New York: Barnes & Noble, 1963.

This contains 27 of the most commonly used tables in statistics -- not all of them having that much applicability to educational research but others being very useful in this field. Among the contents: Tables of Areas of the Normal Curve; Table of t; and Table of Chi Square. There are 168 pages of tables in all.

*Bloom, B. S., Hastings, J. T. and Madans, G.: Handbook on Formative & Summative Evaluation of Student Learning.

This 923 page book is divided into two parts. The first deals with the evaluation of learning objectives from both a

summative and a formative approach. The second presents methods of formative and summative evaluation for various levels of education as well as for different discipline areas.

*Campbell, Donald T. and Stanley, Julian C.: Experimental And Quasi-Experimental Designs for Research (8th Edition) Chicago: Rand McNally, 1972.

This is an 81 page booklet which, considering its length, is remarkably detailed and comprehensive. The emphasis is heavily in the direction of educational research.

*Dressel, Paul L. & Associates: Institutional Research In The University. San Francisco: Jossey-Bass, 1971.

This is a 347 page book relating to institutional research and with nine different contributors writing the different chapters.

Elzey, Freeman F.: A Programmed Introduction To Statistics (2nd Edition). Belmont, Calif.: Brooks/Cole, 1971.

This 385 page book could be the answer to the prayers of anyone faced with undertaking statistical operations without having had any formal courses in this field. The 25 chapters cover every topic found in most elementary statistics texts and there are also sections devoted to formulas, tables, and a glossary.

*Games, P. A. and Klare, G. R.: Elementary Statistics. New York: McGraw-Hill, 1967.

This book is so comprehensive an introduction to the field of statistics that it goes well beyond what most texts in this field contain. Especially valuable are the many problems which

are worked out in detail and which illustrate the application of statistics to a variety of situations.

*Glass, Gene V. and Stanley, Julian C.: Statistical Methods In Education and Psychology. Englewood Cliffs, N. J.: Prentice-Hall, 1970.

This 596 page volume goes well beyond what one might expect in an introductory statistics text and includes a good deal of intermediate material as well. There are particularly useful sections on such statistical tests as analysis of variance and multiple comparion procedures.

Gottman, John W. and Clasen, Robert E.: Evaluation In Education: A Practitioner's Guide. Itasca, Ill.: F. E. Peacock, 1972.

This book represents a very comprehensive treatment of the process of evaluation. Of the 512 pages, a majority contain worksheets and problems designed to assist the student in acquiring an understanding of evaluative techniques. It contains a brief but easily understood section on elementary statistics.

Hayslett, H. T., Jr.: Statistics Made Simple. New York: Doubleday, 1968.

This 192 page book is one of a series intended to provide us with relatively simple but comprehensive courses for self-study and review. This particular work dealing with elementary statistics does not differ that much from the conventional text in this field but it is well-organized and covers a good deal of territory.

Huff, Darrell: How To Lie With Statistics (23rd Printing). New York: W. W. Horton, 1954 (1972).

This 142 page book is a near-classic and probably constitutes the most fun filled introduction to statistics currently available. Despite its satiric approach, it also provides sound insight into the uses -- and the misuses -- of any number of statistical concepts.

Isaac, Stephen and Michael, William B.: Handbook In Research And Evaluation. San Diego: Robert R. Knapp, 1971.

This 186 page book is described as a collection of principles, methods, and strategies useful in the planning, design, and evaluation of studies in education and the behavioral sciences. It is remarkably comprehensive and well-written, which makes it all the more a pity that it is not more handily organized

*Kerlinger, Fred N., Foundations of Behavioral Research (2nd Edition). New York: Holt. Rinehart & Winston, 1973.

This is a 741 page book which covers a considerably broader range than its title might initially suggest. The major divisions of it are: The Language & Approach of Science; Sets, Relations & Variance; Probability, Randomness & Sampling; Analysis, Interpretation, Statistics & Inference; Analysis of Variance; Design of Research; Measurement; Methods of Observation and Data Collection; Multiple Regression; and Factor Analysis.

Spiegel, Murray R.: Theories & Problems of Statistics. New York: McGraw-Hill, 1961.

This 358 page book is one of the Schaum Outline Series that McGraw-Hill publishes. Although perhaps oriented more toward business than education, it nonetheless contains a number of units of interest to the person moving into educational research. These are clearly written, well organized and illustrated with innumerable problems and examples.

*Tuckman, Bruce W.: Conducting Educational Research. New York: Harcourt, Brace & Jovanovich, 1973.

This 402 page book is not a statistical text, but rather it is a handbook to be used when undertaking educational research. The organization of it is fairly unique, in that it contains learning objectives for the student and worksheets accompany selected statistical tests. This constitutes a concise but comprehensive approach to conducting educational research.

*Winer, B. J.: Statistical Principles in Experimental Design (2nd Edition). New York: McGraw-Hill, 1962.

Although many of the 907 pages included in this book deal with more advanced statistics than will be true of the other books included here, it is nonetheless very comprehensive and worth consulting with relation to specific statistical techniques. In each chapter there is a discussion of certain principles, along with illustrative examples, explanations and relevant computational procedures.

## X. INDEX OF MAJOR TOPICS

## SPECIAL NOTE

This is the first edition of this manual. It is very likely that for use in future years, it will undergo considerable revision.

The authors would appreciate your comments and suggestions regarding such revisions. Because of necessary space limitations, they have been faced with considerable conflict about which topics to cover in considerable detail, which to handle only briefly, and which to mention simply for purposes of identification while leaving the major responsibility for them to the National Lecturers.

Any comments you have about these matters should be addressed to:

George M. Barton
Nova University
College Avenue
Ft. Lauderdale, Florida 33314

or

R. F. Mines
Miami-Dade Community College
11011 S. W. 104 Street
Miami, Florida 33143