MMC 500
Researching Communication I

Professor Matthew Lombard

Supplementary Course Materials:

NOTES
(Special thanks to Professor Tricia Jones)

Table of contents

| I. Overview of the research process | 3 |
| II. Paradigms and research formats | 6 |
| III. Concept explication | 11 |
| IV. The literature review | 13 |
| V. Measurement issues across methodologies |
| A. Sampling | 16 |
| B. Reliability | 19 |
| C. Validity | 22 |
| VI. Research methodologies |
| A. Qualitative/interpretive research methodologies |
| 1. Qualitative/interpretive research | 25 |
| 2. Field observation | 30 |
| 3. Qualitative interviewing | 33 |
| 4. Conversation and discourse analysis | 38 |
| B. Quantitative/Qualitative research methodologies |
| 1. Content analysis | 41 |
| 2. Survey research | 45 |
| 3. The experiment | 53 |
| VII. Data analysis |
| A. Overview | 62 |
| B. Using the computer |
| 1. Computers and computing | 64 |
| 2. Preparing the data | 65 |
| C. Descriptive statistics | |
1. Displaying the data  
2. Levels of measurement  
3. Measures of central tendency  
4. Measures of dispersion  

D. Inferential statistics  
1. The logic of inferential statistics  
2. The t-test  
3. One-way analysis of variance (ANOVA)  
4. Two-way analysis of variance (ANOVA)  
5. Interpreting a significant F value  
6. Correlation  
7. Regression  
8. Chi-square  
9. Interpreting a significant Chi-square value  

E. Summary materials  
1. Guide to common statistical symbols  
2. Overview of statistics tools and techniques  
3. Choosing the right inferential statistics technique  

VIII. Suggested references
OVERVIEW OF THE RESEARCH PROCESS

This section of notes contains first a short summary, and then a more complete summary, of the steps involved in conducting research.

Short overview

I. Develop a research question or hypothesis
   A. Find topic area
   B. Literature review
   C. Conceptualization, explication, and theory-building
   D. Begin operationalization

II. Select a research paradigm, format, and methodology
   A. Decide how you believe we come to know things
   B. Determine goal of research project (exploratory, descriptive, explanatory, predictive)
   C. Determine appropriate research methodology (e.g., participant observation, content analysis, discourse analysis, survey, experiment)

III. In accord with the methodology selected, consider issues in each of the following areas:
   A. Measurement
   B. Sampling
   C. Design
   D. Procedures
   E. Data analysis
   F. Schedule
   G. Before finalizing plans also consider ethics, practicality and necessary compromises (time and money constraints), and potential criticisms of method

IV. Conduct study (gather data and analyze it (specific analysis techniques vary according to methodology chosen))

V. Write a paper that summarizes the project

VI. Submit paper to academic conference and then revise paper

VII. Submit paper to journal for publication

Long overview
I. Develop a research question or hypothesis
   A. Identify a research or topic area that is interesting to you
   B. Identify a phenomenon that is interesting to you (from observations, other theories, policy concerns)
   C. Decide whether phenomenon is important
   D. Literature review. Find out:
      1. What theories exist concerning phenomenon
      2. How research has been conducted and what results have been found
      3. What interesting and important theoretical, practical, and/or methodological questions remain
   E. Conceptualization, explication, and theory-building (along with reviewing the literature this is an iterative process).
      1. Identify key concept(s)
      2. Review previous conceptual definitions
      3. Develop your conceptual definition
      4. Repeat for relationships between concepts if applicable
   F. Refine research question or hypothesis
   G. Operationalization of concepts in theory (begin to move from level of concepts to level of measurement)
   H. Evaluate feasibility of research on question or hypothesis (are necessary expertise, time and resources, and access to necessary information/data available?)

II. Select a research paradigm, format, and methodology
   A. Consider alternative research paradigms (this varies across researchers, but seldom from study to study)
      1. Humanism (qualitative/interpretive research and critical analysis; assumes multiple, subjective, socially constructed realities)
      2. Rationalism (quantitative research; assumes objective, measurable reality).
   B. Determine format of research based on research goals (refer to literature review):
      1. Exploration (formulative research)
      2. Description (descriptive research)
      3. Explanation (experimental research)
   C. Select methodology in accord with earlier choices (look at what others have done)
      1. Humanism (qualitative/interpretive research and critical analysis).
         a. Field observation (exploration, description)
         b. Interviewing (exploration, description)
         c. Conversational/discourse analysis (exploration, description)
      2. Rationalism (quantitative research):
         a. Content analysis (description)
         b. Survey (description)
         c. Experiment (explanation)

III. In accord with the methodology selected, consider these issues:
Measurement: What category or type of object, event, or person(s) will be observed and how will observations be recorded? What is the operational definition of the concepts involved (how will they be measured)? Are the definitions valid (do they measure what they should?) and reliable (will they be consistent)?

Sampling: Out of the whole universe of this type of object, event, or person(s), which (and how many) will be selected for study? On what basis will they be selected?

Design: A series of choices specific to the methodology selected concerning how the study will accomplish its goal (for example, will the research be inductive (data to theory) or deductive (theory to data)? Interpretive or functional? Controlled or naturalistic? Conducted in the field or in an artificial setting (laboratory)? Will you be a participant in the social setting or an observer? Will your presence be overt or unobtrusive (will the subjects know they are being studied)? Will it involve qualitative or quantitative data? Will it be cross-sectional or longitudinal?

Procedures: What specific steps will you follow to complete the study? Over what period of time? If applicable, what specifically will happen to participants in the study?

Data analysis: Of course specific techniques vary according to methodology chosen. Whichever techniques will be used, have a plan BEFORE conducting the study! Know in advance how you will evaluate each piece of data you gather to answer the research question or test the research hypothesis. Will analysis be qualitative or quantitative? What measures will be compared? Will you look for differences between groups or relationships between variables?

Schedule: How long will each part of the study take and when will you begin and complete each?

NOTE: Before settling on answers to all of these questions, consider: ethics, practicality and necessary compromises (time and money constraints), potential criticisms of method.

IV. Conduct the study (gather data and analyze it)

V. Write a paper that summarizes the project.

VI. Submit paper to academic conference and then revise paper.

VII. Submit paper to journal for publication.
PARADIGMS AND RESEARCH FORMATS

In the field of communication there are a variety of theoretical orientations and research methods available to the beginning scholar. Unfortunately, the plethora of choices makes clear distinctions and choices sometimes difficult. The sources for this section of notes include Rubin, Rubin and Piele (1990), Yin (1984), and Tucker, Weaver, & Berryman-Fink (1981).

I. Paradigmatic preference (or philosophical assumptions guiding your theory and research).

Although Smith (1988) does an admirable and lengthy job of discussing paradigms in Ch. 15, we'll break her discussion down into more bite-size pieces. First, a paradigm is basically a world view, or set of beliefs about the nature of things. In our field, the most significant paradigmatic distinction is between humanism and scientism. Essentially (and this is incredibly oversimplified), a humanist perspective believes in the presence of multiple realities, in the ability of individuals and groups to create their own realities and truths, and in the importance of discovering those truths through an investigation of the subjective realities of individuals. A rationalist perspective (more associated with scientism) believes in objectively verifiable truths or Truth where the purpose of research is the discovery of that Truth through objective methods. As a result, people operating from a humanistic paradigm tend to engage in research that is interpretive and/or critical (both of these will be explained in more detail below). People operating from a rationalist (or what some call a functionalist) paradigm tend to engage in research that follows a more traditional scientific model, i.e., theory testing.

A. Humanistic approaches: There are two primary humanistic approaches to research in our field: Critical analysis (often referred to as rhetorical criticism) and interpretive analysis (borrowed primarily from sociology, anthropology, and philosophy [phenomenology and hermeneutics]).

1. Critical analysis is a method that places the researcher in the role of the critic, examining a communication situation or event and critically analyzing that event in terms of its meaning or effect. The purpose of critical analysis is to add a means of understanding the event; either in terms of its significance or the level or manner of oppression present, etc. It involves the application of some critical method (for example, Burkean analysis). The most important aspect of critical analysis is that it adds some sort of insight, from the researcher's perspective. This insight does not have to be consensually shared among people who examine this event, or "provable" in any objective sense.

2. Interpretive analysis is an approach that seeks to understand and interpret the social realities of social actors. The main emphasis in this type of approach is to present an interpretation of this reality from the social actor's point of view rather than from the view of the researcher.

B. Rationalist (scientific) approaches: These involve the assessment of reality from an objective orientation, i.e., there is a reality out there and if we can describe it and find out how it is related to other phenomena, then we can come to understand and predict it. For some people, but not all, the ultimate goal of this type of approach is the identification of patterns of regularity in the way the phenomenon behaves. Some people call these laws, some people call these rules. Some people don't bother to call them anything.
II. Approaches to theory building. Essentially, a theory is an expectation for a relationship between variables or phenomena. There are two basic approaches to building or developing a theory: the data-to-theory (inductive) approach, and the theory-to-data (deductive) approach.

A. Data-to-theory (inductive) approach. This is more associated with a humanistic paradigm, but not exclusively so. It assumes that the best way to identify a theory (or set of expectations about the relationships between things) is to immerse yourself in the phenomenon itself, and then let your increasing knowledge of the thing lead you to expectations about its nature or relationships. The grounded theory method is a key example of this type of thinking.

B. Theory-to-data (deductive) approach. The second approach to theory (more associated with rationalism/scientism, but again, not exclusively so) assumes that you generate a theory from your own speculation, existing literature in the area, dreams, etc.; and then you test that theory on selected data, or cases of the phenomenon of interest.

III. Basic research purposes or formats: All research is intended to either build or test theory in some sense. That is why you do it. There are three basic purposes of research (some people call these research formats): exploration/formulative, description/descriptive, and explanation/experimental.

A. Exploration/formulative research: The basic purpose of research in this case is to explore the nature of the phenomenon, to formulate a better understanding of its nature. As a result, this type of research generally asks the question "What is it?". It is most useful when there has not been significant research on the phenomenon before, and when the researcher needs to have a better understanding of the characteristics of this thing in order to relate it to something else or test the effect of it on something else.

B. Description/descriptive research: This purpose of research is intended to answer two general questions: a) Who possesses this attribute, thing, or phenomenon?, and b) How is this related to other phenomena? In the first case, the researcher is usually interested in finding out how this attribute or thing occurs in a population; for example: Who has more of this, men or women? In what contexts does this occur most often? In terms of the second question, the main purpose is to correlate or find an association between the occurrence of the phenomenon of interest and other things (e.g., Is communication apprehension related to business success?)

C. Explanation/experimental research: This purpose is to explain how the phenomenon is affected by or affects other things. In short, this is the focus on causation. The main interest is to examine how this thing is caused by changes in something else, or how changes in this thing cause some reaction in another phenomenon.

IV. Sources of data: In general research methods can be thought of as ways to generate data to be used for a particular research purpose. There are several ways to generate or collect data, but the most common are: interviewing, observation, and questionnaires.

A. Interviewing involves asking people who have some knowledge of or experience with the phenomenon about their experiences with, knowledge of, or attitudes about it.

B. Observation involves actually observing people, places, events, or documents.
C. **Questionnaires** are a method of asking questions in a written form to obtain information from people about their experiences with, knowledge of, or attitudes about the phenomenon. In our field, the questionnaire, either in the form of surveys or as used in other methodologies to measure attitude or behavior, is the most common way of generating data.

V. Dangers of inflexibility. It is important to note that certain kinds of data generation are more common to some research purposes than others, just as some kinds of research purposes are more common to some orientations to theory-building than others, and just as some orientations to theory-building are more common to some paradigmatic beliefs than others. However, be careful of excessive rigidity in your understanding of these relationships; don't limit yourself unnecessarily. In general, one can use any source of data for any research purpose.

VI. More on research formats. The choice of research format (exploration/formulative, description/descriptive, and explanation/experimental) depends upon three general factors: 1) the type of research question proposed, 2) the extent of control a researcher has/desires over the actual behavior or events under study, and 3) the degree of focus on contemporary as opposed to historical phenomenon. These factors can be broken down into a simple chart (and please remember that this is very simplified):

<table>
<thead>
<tr>
<th>Format</th>
<th>Question(s)</th>
<th>Control</th>
<th>Time Focus</th>
<th>Paradigm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Historical/critical</td>
<td>How, why</td>
<td>No</td>
<td>Past</td>
<td>Humanism</td>
</tr>
<tr>
<td>Exploratory/formulative</td>
<td>What</td>
<td>No</td>
<td>Present</td>
<td>Humanism</td>
</tr>
<tr>
<td>Descriptive</td>
<td>Who, where, how much, how many</td>
<td>No</td>
<td>Present</td>
<td>Humanism and Rationalism</td>
</tr>
<tr>
<td>Explanatory/experimental</td>
<td>Why</td>
<td>Yes</td>
<td>Present</td>
<td>Rationalism</td>
</tr>
</tbody>
</table>

A. **Historical/critical research**: This is the only time that we'll discuss historical/critical research in this class (there is a CORE graduate course devoted to this topic). Historical/critical research is a reconstruction of the past in a systematic and objective manner.

1. Functions:
   a. To ascertain the meaning and reliability of facts concerning the past.
   b. To appraise or evaluate past events.
   c. To study trends and the mechanics by which past events have occurred.
   d. To make comparisons of similarities or differences between past events.

2. Types of historical/critical research in communication
a. Biographical studies  
b. Movement or idea studies  
c. Regional studies  
d. Institutional studies  
e. Case histories (focusing on the social setting of a single event)  
f. Selected studies (focusing on a particular element in a complex process, e.g., the use of humor in a presidential campaign)  

B. Exploratory/formulative research: This research format is traditionally focused on the identification or development of constructs, or the development of theory. Basically, research questions that seek to provide an initial or exclusive description of a phenomenon, that ask "what" questions, are appropriate for this format. The analysis of uncontrolled, contemporary communication events are the focus.  

C. Descriptive research: This involves the collection of contemporary information directly from social actors to describe people, situations, events, or relationships.  
   1. General functions.  
      a. To answer questions about who (in a group or population) possesses or demonstrates certain attributes or characteristics and the extent of those attributes or characteristics.  
      b. To demonstrate or describe relationships (associational rather than causal) between variables of interest.  
   2. Specific purposes: The two general functions may be clarified by looking at more specific purposes of descriptive research in communication.  
      a. To describe individuals, events or groups (e.g., what are the characteristics of communication graduate students at Temple?)  
      b. To compare individuals, events, or groups (e.g., how are communication graduate students at Temple similar to/different from communication graduate students as Penn State?)  
      c. To determine needs, desires, or problems (e.g., what communication challenges are faced by communication students at Temple?)  
      d. To determine preferences to make predictions or projections (e.g., who will people vote for in the Philadelphia mayoral election?)  
      e. To determine relationships or associations (e.g., is communication apprehension related to success in an entry level organizational position?)  
      f. To link attitudes to behavior (e.g., how do parents' attitudes about television violence relate to their behaviors regarding the monitoring of their children's viewing?)  
   3. General modes of descriptive research: There are three general modes, each of which will be discussed in more detail later in the class -- surveys, observational studies, and correlational studies.  

D. Explanatory/experimental research: Usually referred to as experimental research, this format involves the manipulation of situations (or control) to determine causal relationships between or among contemporary phenomena. A more formal definition of
an experiment is "a recording of observations, quantitative or qualitative, made by defined and recorded operations and in defined conditions, followed by examination of the data using appropriate statistical and mathematical rules, for the existence of significant causal relations." It is fairly easy to see from this description, that "why" questions are the focus of experimental research.
CONCEPT EXPLICATION

One of the functions of a literature review is to assist the researcher in a process called concept explication. This involves identifying the key concepts that are part of the phenomenon or question to be studied, carefully defining them, and determining how they can be measured. This is a VERY oversimplified summary; for a much more complete description of this process see Chaffee (1991).

I. The basics.
   A. Definitions contain primitive and derived terms. Concepts are ideas, which we define with words. A concept is only useful if everyone agrees on what it is, which means that the researcher must carefully and exactly define the concepts with which she is working. Some of the words used to define any concept are called primitive. These are words that one can assume any reader will understand (e.g., person, object, time). Others are derived, or words that are created from the meanings of primitive terms and which we cannot assume are understood in the same way by other people (e.g., attribute, college freshmen). A useful conceptual definition must therefore include a definition of any of the terms in the definition itself which are derived.
   B. Conceptual and operational definitions. A concept can be defined at two different "levels," the conceptual and the operational. The conceptual level refers to the actual meaning of the concept regardless of how it is measured, while an operational definition describes a specific method for actually measuring the concept. They should not be confused. For example, intelligence might be defined at the conceptual level as the ability to learn or understand (these terms would have to be carefully defined as well). At the operational level intelligence is typically defined as a score on the IQ test. Note that one concept can be measured in many different ways by different researchers, even though each researcher uses the exact same conceptual definition.
   C. Validity of a concept. In the most general sense a concept is valid if it represents what we want it to. See the notes titled "Validity" for more information on types of validity. An operational definition of the concept is valid if the definition captures the essence of the concept. Note that there is not usually one true, objectively valid definition. Validity is defined by "intersubjective" agreement among scholars.

II. Steps in the explication process (for one concept, but expandable to multiple concepts and relationships between them (i.e., theory). Note that the process is iterative -- the researcher must repeat steps and sets of steps throughout the explication process.
   A. Identify focal concept of interest
      1. Does it vary or is it a constant?
      2. What is the unit of analysis (what is one of the concept?) Be aware of the ecological fallacy -- making conclusions about individuals based on aggregates (e.g., "public" opinion) and vice versa
      3. How does the concept vary with respect to time (cross-sectional versus process)?
      4. How might we use the variation -- alone or with respect to other concepts or antecedents or effects or as intervening variables?
B. Literature search (see notes titled "The literature review"). Look for meanings, operational definitions, findings of others.
   1. What different conceptual meanings have been used?
   2. For what different purposes has the concept been used?
   3. What different operational definitions exist?
   4. What different names have been used for the concept?
   5. What operational contingencies (specific times, places, paradigmatic views) have shaped uses of the concept?
   6. In light of your purpose (see above), which usages in literature seem best?
   7. What has been discovered about concept?
      a. What values does it take in different populations?
      b. How much does it vary (what is its range) across units of observation?
      c. How does it vary across time? (increase, decrease, stable)
      d. What other concepts are correlated with (related to) it in the literature?
      e. What are apparent antecedents and consequences of the concept and its variation?

C. Define the concept (a nominal definition (a name) and a conceptual definition).
   1. Conduct a "meaning analysis" by developing lists of objects, events, people, etc. and distillations of definitions to identify the "essential nature" of the concept.
   2. Conduct an "empirical analysis" by developing rules for deciding if a given object, event, person, etc. is an instance of the concept. What are the necessary and sufficient conditions for inferring an object is in this class? What observations, of what units of analysis, at what times, under what conditions, are required?

D. Review (evaluate) the conceptual definition
   1. Is it specific enough?
   2. Is it a reification, a name for something that doesn't really exist (e.g., attitude, public opinion)?
   3. Is it possible or likely that other scholars will agree on and use this definition?

E. Modify conceptual definition as needed

F. Develop operational definition (operationalize the concept)
   1. What is the unit of observation? (Remember fallacies)
   2. Can the concept be observed passively (do representatives of the concept exist already in enough quantity to measure/observe?), or do you need to create the phenomenon?
   3. How should the units of observation be sampled? (Don't do more than necessary)

G. Evaluate operational definition by doing actual observations (doesn't typically get done as separate step)
   1. Develop an empirical description (Is variation, shape of distribution, presence of missing or ambiguous information as expected?)
   2. Compare to prior literature
   3. Evaluate number of dimensions: Are there too many dimensions and therefore multiple concepts?
   4. Reliability and validity:
      a. Without reliability who cares about validity! Check it with test-retest, split-half, inter-item correlation, etc. (see notes titled "Reliability")
      b. Increase reliability with multiple indicators
THE LITERATURE REVIEW

Locke, Spirduso, and Silverman (1987) present considerable information on literature review guidelines. These notes are taken from Rubin, Rubin, and Piele (1990) and Leedy (1985).

There are two main purposes to a literature review: 1) to summarize relevant research in the area, and 2) to evaluate that literature. There are also two basic types of literature reviews: 1) exemplary or representative reviews, which only cover key references and assume that the reader has a general knowledge of the topic area, and 2) exhaustive literature reviews, which review all available information on the topic. Generally, the literature review in a proposal is the former type and the literature review expected in a dissertation project is the latter type.

I. Conducting the search
   A. Select the research area.
   B. Narrow the research area so that you can begin to formulate a specific research question (and identify its components).
      1. The area should be defined as narrowly as possible. For example, if you are interested in studying the effect of television violence on children, you would want to clarify several aspects (television = entertainment shows? news? movies? network? cable?; violence = physical? verbal? both?; children = toddlers? youngsters? adolescent? all?)
      2. Your research area will undergo revision and narrowing as you become familiar with the literature that already exists in the area.
   C. Adjust the topic, if necessary, based on initial exposure to the literature.
   D. Choose a search strategy for more comprehensive review (either general to specific or specific to general).

II. Using effective search procedures
   A. Begin by locating all references that are relevant to your search.
      1. Important resources to check are:
         b. Indexes (e.g., Social Science Citation Index (SSCI), Index to Business Periodicals, Index to Legal Periodicals, Index to Communication Journals, Matlon Index)
         c. Annual reviews and series
         d. Handbooks (e.g., Handbook of Communication Science, Handbook of Interpersonal Communication)
         e. Communication Journals (especially if you know that a specific journal tends to publish articles in your areas of interest) (e.g, Journal of Communication, Journal of Broadcasting, Human Communication Research, Communication Monographs, Communication Theory)
         f. Textbooks
         g. General book searches (usually using the on-line catalog system at the library)
2. Create a bibliography card file to keep track of each reference collected. Each card should contain the full citation in APA style (see Appendix). One reference per card. These can then be rearranged at the end and used to type your bibliography.

3. Keep a careful record of the sources, topics, and key words you have searched (e.g., "searched Psychological Abstracts 1982-present under the key words mediation and conflict")

B. Photocopy all important references for later review. Remember that the copy machine is your friend.
   1. Examine the reference lists in each article/book reviewed to find additional important references for your use.
   2. Abstract critical studies and note critical quotes (be sure to record the precise location of quotes, including page number)

III. General tips to remember during search
   A. Focus only on relevant information
   B. Set a search stopping point (including recency limit)
   C. Practice good note-taking skills

IV. Writing the literature review
   A. Parts of the review
      1. Introduction
      2. General statement of the problem
      3. Summary of literature
      4. Critical evaluation
      5. Research questions/hypotheses
   B. Tips for effective writing (these are just a few of many)
      1. Use the 3rd person
      2. Use active voice
      3. Avoid jargon
      4. Use gender neutral language
      5. Emphasize brevity of expression (short sentences)
      6. Be consistent in your use of tense
      7. Include effective transitions, previews, summaries to add to internal coherence.

V. Formulating research questions and hypotheses: Remember that the research questions or hypotheses presented at the end of the literature review serve as the guiding focus for all further methodological decisions in your study and proposal. Therefore, it is critical that you invest considerable time in formulating exactly the research question you want (based, of course, on your knowledge of the phenomenon from your literature review).

A. Nature of research questions and hypotheses: The basic difference between a research question and a hypothesis is that the latter indicates an expected effect or relationship while the former does not. For example, "Persons from Southern European countries will differ from persons from Northern European countries in terms of the amount of touch they use in dyadic conversations" is a hypothesis. It indicates an expected relationship or difference. "Do touch differences in dyadic conversations exist between Southern and
Northern Europeans?" is an example of a research question. Furthermore, hypotheses can be directional or nondirectional. The aforementioned hypothesis is nondirectional -- it hypothesizes that a difference exists but does not predict the direction of the difference. A directional hypothesis might be "Southern Europeans will engage in more touch during dyadic conversations than will Northern Europeans."

B. When to use research questions versus hypotheses: The general rule is to use research questions when there is not enough previous information to be able to safely or confidently predict a specific relationship or effect.
SAMPLING

Sampling is a process of selecting observations (people, objects, events, documents, etc.). Sampling strategies are classified as either probability sampling or as nonprobability sampling. For several reasons the former is the more preferred.

The information in this section of notes was obtained from both Smith (1988) and Babbie (1989).

I. Probability sampling. This is a method of sampling that allows the researcher to calculate the probability that any member of the population may or may not be included in the sample. Whenever we are dealing with a population there will be some degree of heterogeneity (or difference) in members of the population. We are most concerned with the extent to which members of our population differ on variables that are of potential interest to us in our investigation. The main goal in probability sampling is to select a sample that is representative of the population. A sample can be said to be representative of the population if the aggregate characteristics of the sample closely approximate the occurrence of those same aggregate characteristics in the population. Probability theory allows us to estimate the accuracy or representativeness of the sample in terms of the population. Random selection is the key to probability sampling. More specific information about how probability theory and probability sampling work can be found in the notes titled "The logic of inferential statistics."

A. A summary of probability sampling strategies
1. **Simple random sampling**: Each member of the population has an equal chance of being included in the sample. This requires a list of all of the elements (people, marriages, TV shows, newspapers, etc.) in the population. Each element is assigned a number, and specific numbers (or members of the population) are selected using a table of random numbers.
2. **Systematic random sampling**: Every kth element in the total list of elements in the population is selected. If you had a population of 100 elements and you were selecting every 10th one, your sampling interval, or the standard difference between selected elements would be 10; and, your sampling ratio, or proportion of elements in the population that are selected, would be 10/100 = .10 = 1/10th of the total population. When using systematic random sampling beware of periodicity, or the possibility of selecting a nonrepresentative sample because the list of elements was arranged in such a way that your selection of every kth element would give you a skewed sample (e.g., the features included in newspapers and the cycle of news in the world vary systematically throughout a week, so sampling every seventh issue of a newspaper would not be a good idea).
3. **Stratified random sampling**: This is a method for obtaining a greater degree of representativeness by dividing your population into more homogeneous subgroups and then randomly selecting an equal number of subjects from each subgroup. Usually, to do this, you decide on critical strata, or characteristics in your population that you want to have represented.
4. **Probability proportionate to size (PPS) sampling:** This can be used as a variation of stratified random sampling (then it is called proportional stratified random sampling) or with multistage cluster sampling (discussed next). Sometimes you have a situation where your subgroups or strata are very unequal and you want the members selected in your sample to accurately reflect the occurrence of that characteristic in the population. Say you are sampling men and women (your two subgroups), and men comprise 65% of the population and women 35%. Then you would (for a sample of 100) randomly select 65 men from the group of men and randomly select 35 women from the group of women.

5. **Multistage cluster sampling:** All of the preceding sampling strategies have assumed that you have a complete list of all of the elements in the population. Often you can't manage that. In those cases, MSCS can be used. This procedures involves repetition of 2 steps: 1) listing the sampling units and 2) randomly selecting units. For example, imagine you want to produce a sample of Christians in Colorado. Your first step might be to identify all of the denominations of Christian churches. Next you randomly select 10 of those. Then you repeat the two steps: For each of those denominations selected in the first "sampling stage" list all of the churches belonging to the denomination and randomly sample 10 of these churches. Then (for those selected churches) list all of the members of each church, and randomly select 10 members for each church. Your final list of selected church members would comprise your sample. Please note that sampling error is introduced for every stage of your process. So you want to maximize the number of clusters you select at each stage and decrease the number of elements selected in each cluster.

II. **Nonprobability sampling:** This is much less preferred than probability sampling because it does not allow the researcher to estimate the representativeness of her sample. Nonetheless, if you have to do it, here are the three most common methods.

A. **Purposive or judgmental sampling:** Purposively selecting particular subjects (e.g., go to the mall and pick people who appear knowledgeable, cooperative, attractive, etc.).

B. **Accidental or convenience sampling:** Taking whoever is available (e.g., interviewing friends and family).

C. **Quota sampling:** Same as PPS sampling (see above) only the people within each subgroup are not randomly selected.

III. **Summary table.** Another breakdown of the types of sampling is presented on the following page.
<table>
<thead>
<tr>
<th>Sampling Type</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Random sampling</strong></td>
<td>Avoids systematic bias in the sample; large sample size is important for making generalizations.</td>
</tr>
<tr>
<td>1. Simple random sample</td>
<td>Achieve a representative sample that permits generalizations to the whole population.</td>
</tr>
<tr>
<td>2. Stratified random and cluster samples</td>
<td>Increase confidence in making generalizations to particular subgroups or areas.</td>
</tr>
<tr>
<td><strong>B. Purposeful sampling</strong></td>
<td>Increase the utility of information obtained from small samples; sampling criteria based on specific desired attributes.</td>
</tr>
<tr>
<td>1. Sampling extreme or deviant cases</td>
<td>Provide information about unusual cases that may be particularly enlightening</td>
</tr>
<tr>
<td>2. Sampling typical case(s)</td>
<td>Avoid studying an entity where the results would be dismissed outright because THAT entity is known to be special, deviant, unusual, extreme, etc.</td>
</tr>
<tr>
<td>3. Maximum variation sampling - picking three or four cases that represent a range on some dimension</td>
<td>Increase confidence in common patterns that cut across different entities.</td>
</tr>
<tr>
<td>4. Sampling critical cases</td>
<td>Permits LOGICAL generalization and maximum application of information to other cases because if it's true of this one case, it's likely to be true of all other cases.</td>
</tr>
<tr>
<td>5. Sampling politically important or sensitive cases</td>
<td>Attracts attention to the study (or avoids attracting undesired attention by purposefully eliminating from the sample politically sensitive cases).</td>
</tr>
<tr>
<td>6. Convenience sampling -- take the easy cases</td>
<td>Saves time, money, and effort.</td>
</tr>
</tbody>
</table>
RELIABILITY

In all measurement there is error. Some error is systematic. For example, say you use a thermometer to measure temperature and the thermometer is always 2 degrees off. The error in this measure is systematic because the same amount of error is always introduced by the flaw in your measure. Error can also be random, i.e., the temperatures you take with a thermometer differ because you are tired and don't read the thermometer correctly. To the extent that this error in measurement is slight, or to say it another way, the "measurement error" is slight, a measure is reliable.

I. The classical view of measurement error. The argument here is that each person has a TRUE score, one that would be obtained if there were no measurement error at all. We can graph a pair of hypothetical true scores for person A and person B:

But of course, when we obtain a score for each person there will be some measurement error. Because of this, we can imagine that over several measurement trials we will obtain a distribution of scores for person A and for person B which are a mixture of the true scores and the measurement error:
The wider the spread of obtained scores around the true scores, the more measurement error is present. An index of the amount of error in the measurement and therefore the measure's reliability, would be the typical or average difference between obtained scores and the true scores. Since we can't ever know what a true score is, we assume that it is the average ("mean") score in the distribution of obtained scores and we estimate the reliability of the measure by the typical or average difference between obtained scores and the average (mean) obtained score. In short, we judge a measure to be more reliable if the "spread" of scores we obtain with it, across different trials with the same person and/or across trials with different people, is smaller.

II. The domain sampling theory of measurement error. This theory basically argues that a test is composed from all of the possible measurement items in the domain of interest; since all of the items can't be used the goal is to select a good, i.e. random, sample of them to use. For example, if we were to construct a spelling test for fourth graders, and we could create a perfect random sample of all the words that could be assumed to be known at that age level, then we would have a good test. Further, the theory argues that the degree to which we obtain similar or consistent answers to the items we use is a good indication of the reliability of the set of items. The number that represents this degree of similarity is a correlation or correlation coefficient (see notes on "Correlation"), and in this context it is usually called a reliability coefficient. Because it is nearly impossible to have an actual random sample of all possible items in a domain, the reliability coefficient is really an estimate. In truth, most researchers don't even try to approximate this random sample; they merely pick and choose items for inclusion that seem to make some sense to them.

III. Methods for the investigation of reliability of measures
A. Inter-item correlation methods: These are used to examine the internal consistency of the set of items. This means that you will be investigating the extent to which items on the overall measure are highly correlated with each other. In a set of items designed to measure one concept (a unidimensional scale) all items should be highly correlated with all other items. For a set of items designed to measure a number of distinct components of a concept (a multidimensional scale) one usually assesses the internal consistency for each dimension as well as overall.

Tests for internal consistency are usually done using Cronbach's alpha for interval level items (see notes titled "Levels of measurement") and Kuder-Richardson 20 for dichotomous or nominal level items. Nunnally (1978) argues that alpha should be obtained before anything else since it sets an upper limit. In other words, even if the other reliability tests are performed, the alpha will always generate the highest reliability coefficient.
B. **Alternative forms (AF) test**: The researcher constructs two parallel versions of the instrument; in other words, two overall measures that contain different items but are thought to be measuring the same thing. The researcher selects a group of subjects and gives them one form and then later (usually about two weeks later) gives these same subjects the second form. The scores on both measures are then correlated. If the reliability coefficient obtained from an alternative forms test is markedly lower than Cronbach's alpha, say by about 20% lower, it indicates a considerable amount of measurement error. In addition, if the correlation between the two forms is low, but the correlation between individual items on the test is high, it indicates that the two forms reliably measure somewhat different traits or constructs.

Sometimes people compare an AF test given two weeks apart with an AF test given 1 day (or relatively immediately) after the first. By comparing the correlations obtained from the AF-2 weeks and the AF-Immediate you can conclude some interesting things:
1. If the correlation of AF-2 weeks is low and the correlation for the AF-Immediate is low it indicates that the two AF forms are measuring different types of content.
2. If the correlation between the AF-2 weeks is low but the correlation for the AF-Immediate is high, it indicates that the trait being measured fluctuates significantly over time.

C. **Split-half technique**: In this approach the researcher constructs the measurement instrument and administers it to a group of subjects. Then the researcher takes the resulting data and splits the measure into two equal halves (for example, all odd number questions versus all even number questions). A correlation between the two halves is calculated which results in the reliability coefficient. The problem with this method is that the coefficient then depends upon how the measure has been divided.

D. **Test-retest method**: In this approach, the researcher administers the very same test to the same group of subjects at two different points in time and then correlates the scores on the first administration with the scores on the second administration. The main problem with this method is that the first administration of the instrument can sensitize subjects so that their answers on the second administration are affected. This method should be reserved for a measure that has a large number of items (say around 200) and when the test is administered 2 weeks to 1 month apart.
VALIDITY

In a general sense, a measure is valid if it does what it is intended to do. Validation always requires empirical investigations, with the nature of the evidence required dependent upon the type of validity being assessed. Validity is usually a matter of degree rather than an all or nothing property.

Most measures in the social sciences serve one of three functions: 1) the establishment of a statistical relationship with a particular variable, 2) the representation of a specified universe of content, and 3) measurement of psychological or social traits. Corresponding to these three general functions are the three types of validity: 1) predictive, 2) content, and 3) construct.

I. Predictive validity: This type of validity is important when your measure is intended to estimate or predict some external behavior. The external behavior is referred to as the criterion behavior. For example, the LSAT (Law School Aptitude Test) is assumed to predict someone's success in law school. If we choose as our criterion a person's class rank at the end of her law program, her score on the LSAT should be highly related to her class rank, if the LSAT has predictive validity.

As a result, establishing predictive validity is a three step process: 1) select the criterion behavior (and give this careful attention since this is where you can really go wrong), 2) administer your measure, which you can call the predictor, and 3) assess the relationship between the predictor and the criterion (this is usually done through correlation between the two).

Predictive validity is especially important to assess when you are planning to use the results of your measure to make policy decisions, such as letting someone into law school. It is critical that you select an appropriate criterion.

II. Content validity: This is the type of validity you use when the adequacy of the test itself is in question in terms of two standards: 1) whether the test is a representative collection of items from the universe of items pertaining to this phenomenon, and 2) whether the test demonstrates sensible methods of test construction. In the case of the first standard, we want to try and assess whether all critical dimensions or possible attributes of the phenomenon are assessed by the items on the test. For the second standard we are more concerned with whether the questions are asked in a comprehensible manner, that they are unbiased in their use of language and phrasing, and in general, that there are no artifactual results likely from the violation of sensible test construction methods.

Content validity usually rests on appeals to reason regarding the adequacy with which a measure meets these two standards. One common form of establishing content validity is having a panel of experts review the measure.
III. **Construct validity**: This type of validity is critical when we are attempting to define a construct and are interested in determining whether our measure is a valid representation of the construct as we have defined it.

To the extent that a variable is abstract, such as credibility, we call it a construct. Constructs are ideas, arbitrarily defined, that have no necessary rooting in objective reality. Constructs vary in the extent to which the domain of related variables is: 1) large or small, and 2) specifically or loosely defined. For example, communication competence is a construct that is relatively large in scope and has not been specifically well defined in terms of all the things that make up or comprise communication competence. Obviously, the larger the domain of variables or observables related to a construct, the more difficult it is to define variables that do or do not belong in the domain.

A measure is said to have construct validity to the extent that: 1) the measure represents the dimensions of and internal structure of the construct, 2) the measure is correlated with other measures of the construct that have been previously proven to be valid, and 3) the measure is not related to the measurement of constructs that have been proven to be similar to but fundamentally different from your construct. The first type of construct validity is usually determined through the use of some sort of factor analytic method. The second sort requires you to administer your measure and other measure(s) of the construct to a sample of people and then to correlate the scores from the measures. To the extent that the scores are highly and positively correlated, you have demonstrated convergent construct validity. The third method requires you to administer your measure and other measures that represent different constructs to a group of people and to correlate the results. To the extent that there is a low or negative correlation between the measures, you have demonstrated divergent construct validity.

IV. **A summary of the more common types of validity**. (Overview: Measurement is a tool of research, and validity is the attempt to determine whether a type of measurement actually measures what it is presumed to measure. Various disciplines rely upon various types of validity to verify the effectiveness of their measurement procedures.) Here are the principal types of validity.

A. **Face validity**. This type of validity relies basically upon the subjective judgment of the researcher. It asks two questions which the researcher must finally answer in accordance with best judgment: 1) Is the instrument measuring what it is supposed to measure? and 2) Is the sample being measured adequate to be representative of the behavior or trait being measured?

B. **Criterion validity**. Criterion validity usually employs two measures of validity; the second, as a criterion, checks against the accuracy of the first measure. The essential component in criterion validity is a reliable and valid criterion -- a standard against which to measure the results of the instrument which is doing the measuring. The data of the measuring instrument should correlate highly with equivalent data of the criterion.
C. **Content validity.** This type of validity is sometimes equated with face validity. Content validity is the accuracy with which an instrument measures the factors or situations under study; i.e. the "content" being studied. If, for example, we are interested in the content validity of questions being asked to elicit familiarity with a certain area of knowledge, content validity would be concerned with how accurately the questions asked tend to elicit the information sought.

D. **Construct validity.** A construct is any concept, such as honesty, which cannot be directly observed or isolated. Construct validation is interested in the degree to which the construct itself is actually measured. To this end a significant procedure has been developed by Campbell and Fiske known as the Multitrait-Multimethod Matrix Method. It makes use of the traits of convergence and discriminability. **Convergence** looks to the focal effect of various methods of measuring a construct. Different methods of measurement of the same construct should "converge" or "focus" in their results. **Discriminability** means that the measuring instrument should be able to discriminate, or differentiate, the construct being studied from other similar constructs.

E. **Internal validity.** This term, and the one following, should not be confused with internal and external criticism, which are tests of validity in historical research. Internal validity is the freedom from bias in forming conclusions in view of the data. It seeks to ascertain that the changes in the dependent variable are the result of the influence of the independent variable rather than the manner in which the research was designed.

F. **External validity.** This type of validity is concerned with the generalizability of the conclusions reached through observation of a sample to the universe; or, more simply stated: Can the conclusions drawn from a sample be generalized to other cases?
QUALITATIVE/INTERPRETIVE RESEARCH

Qualitative/interpretive research usually consists of research that uses inductive analyses of qualitative data in order to build theory (referred to as grounded theory), and research that uses inductive analyses of qualitative data to describe situations (referred to generically as interpretive research). This section of notes, based on information from Patton (1980), Anderson (1987), Kirk and Miller (1986), Marshall and Rossman (1989), and Strauss and Corbin (1990), concentrates on the underlying philosophical assumptions of qualitative research, general measurement methods and concerns, sampling concerns, design concerns, and data analysis concerns.

I. Philosophical assumptions and traditions: (For those who are interested, Anderson [1987] presents a nice overview of the philosophical bases of qualitative/interpretive research). Drawing on the works and traditions of Kant, Husserl, Weber, and Shutz, there are certain basic concepts underlying this type of research:

A. Knowledge and meaning are socially constructed: Basically, humans are interpretive and active creators of meaning and knowledge; therefore research should attempt to understand the social construction of reality (or realities).

B. Qualitative research shares certain normative ideals:
   1. Inductive: It is "presuppositionless" research that is based on a specific-to-general process of understanding.
   2. Wholistic/eidectic: The researcher must understand the totality of the situation or context before the meaning of its various parts can be understood.
   3. Subjective: The research is directed toward understanding the subjectively held realities of the social actor.
   4. Contextual/naturalistic: Since meaning structures are formed within a context and are affected by that context, research should be conducted in that context and with an appreciation for that context.

II. Grounded Theory: Developed originally by Glaser and Strauss (1967), Grounded Theory is the process of developing theory in an inductive manner.

A. Three assumptions:
   1. The adequacy of a theory cannot be divorced from the process by which it was generated. Theories should be "grounded" in a deep understanding of the phenomenon in action.
   2. Generating a theory from data means that research decisions (including the focus of the research question, the research design, the nature of observations, the nature of the sample) emerge from the data and are constantly and systematically worked out in relation to the data during the course of the research.
   3. Verification of the resulting theory must be based on testing the theory against the data from which it was derived.

B. The adequacy of the resulting theory can be assessed using four central criteria:
1. **Fit**: The theory matches the data.
2. **Comprehensibility**: The theory is understandable to social actors and outsiders.
3. **Generality**: The theory applies equally well to a variety of contexts related to the essential focus of the theory.
4. **Control**: The theory should enable control of or management of change in the phenomenon of interest.

III. Major research decisions in qualitative research. Even though there are differences between Grounded Theory and interpretive research, the similarities outweigh the differences. Thus, we will focus this discussion on the similarities and note the differences where they occur.

A. **Measurement**: Qualitative research involves the collection of qualitative data, or observations that are expressed in nonquantitative ways. Qualitative data provides depth of detail and description usually lost in numerical data.

1. Types of data. There are two general types of qualitative data, each of which is described in considerably more detail in notes regarding qualitative/interpretive research methods.
   a. Observation (including observation of documents and artifacts)
   b. Interviewing

2. Reliability and validity of qualitative data: All measurement should be reliable and valid. However, these concepts are more commonly used in reference to quantitative rather than qualitative measures.
   a. Reliability: Loosely defined as the extent to which a measurement procedure yields the same answer however and whenever it is carried out. Kirk and Miller (1986) discuss three types of reliability in qualitative data:
      1. **Quixotic reliability**: Refers to the circumstances in which a single method of observation continually yields an unvarying answer. They suggest that in field research, this may actually suggest that the researcher is only tapping into the superficial information and is missing the nuances.
      2. **Diachronic reliability**: Refers to the stability of an observation over time. This type of reliability is only appropriate when the phenomenon observed is not assumed to change over time.
      3. **Synchronic reliability**: Refers to the similarity of results from the use of multiple measures within the same time period. They suggest that this is the most useful form of reliability for qualitative research.
   b. Validity: This is the extent that the measure or observation provides the "correct answer" or the answer that most closely approximates the "reality" of the phenomenon of interest. They discuss three types of validity in qualitative data:
      1. **Apparent validity** (analogous to face validity): A measure so closely linked to the phenomenon that it "obviously" measures it. Apparent validity suggests the presence of the other two types, but is not a sufficient condition for them.
      2. **Instrumental validity** (analogous to pragmatic or criterion validity): This is when observations match those from other measures that are assumed to be valid.
3. **Theoretical validity** (analogous to construct validity): This is when there is substantial evidence that the conceptual definition of the phenomenon corresponds to the observations.

4. They argue that the best way to protect validity is to use a diversity of apparently valid measures and to take context into account.

B. **Sampling**: Sampling decisions differ for grounded theory studies and interpretive studies. Both types of research depend more on purposive sampling than representative sampling.

1. Grounded Theory uses purposive, theoretical sampling. This means that sources of observation, or sampling units, are chosen based on which units can provide relevant information concerning the phenomenon. Two guidelines help identify relevant information: 1) it is repeatedly present or notably absent, and 2) it seems to have significance with the study.

   a. Theoretical sampling means that you sample until no new information is being obtained. Only a careful continual analysis of the data enables the researcher to know when theoretical sampling is completed (known as theoretical saturation).

   b. Sampling decisions are constantly changing during the course of a grounded theory study. Just as research questions and decisions about observation and design change as new information is gathered, so do sampling decisions.

   1. Initial sampling decisions involve:
      a. Choosing a site or group that seems relevant.
      b. Deciding upon the kinds of data to be used.
      c. Deciding whether to study the group cross-sectionally or longitudinally.

   2. Sampling decisions consistent with types of coding: When we discuss data analysis we'll present the three types of coding and their purposes. For now, realize that each stage of coding has different implications for sampling strategies.

      a. Sampling in open coding: The purpose is to uncover as many relevant categories as possible. Therefore, the sampling is open to those sampling units that provide the greatest opportunity to produce relevant data. This usually means concentration on a heterogeneous sample.

      b. Sampling in axial coding: Sampling now focuses on finding cases that uncover or validate expected relationships between categories. Thus, sampling now focuses on finding variation by looking at heterogeneous groups.

      c. Sampling in selective coding: At this stage you are trying to verify the broad theory being developed. Therefore, sampling decisions here require you to be very purposive and choose sampling units or cases that provide critical cases where theory can be applied.

2. Interpretive research: Usually, sampling decisions are made a priori. In other words, a sampling strategy is decided upon at the outset and carried through without
revision. Sampling may be purposive or representative, but it is usually purposive with an emphasis on sampling the most relevant sources of information. Types of purposive sampling strategies are:
   a. Extreme cases
   b. Typical cases
   c. Heterogeneous cases to maximize comparison

C. Design: The notion of design in qualitative research concerns the ability to dictate important decision choices and place them in a logical plan of research action.
   1. General components of qualitative design: In many ways, all of the factors we have already discussed are components of design.
      a. Site and sample selection
      b. Researcher's role management
      c. Data collection techniques
      d. Managing and recording data
      e. Data analysis strategies
      f. Research action plan and timeline.
   2. Critical tradeoffs:
      a. Boundary problem: whether to study one or a few questions in depth or to study many questions more superficially.
      b. Prioritization problem: what is most important and should be studied first?
   3. General design logic: Andersen (1987) suggests that the process of qualitative research involves four general steps.
      a. Encountering the Other (immersion in the context and focusing of research)
      b. Creating the research text (collection of data and preparation of field notes)
      c. Composing the episode (providing a descriptive, written interpretation of the event)
      d. Presenting the ethnographic argument (writing the entire interpretation across episodes)

D. Data analysis: For most people, this is the most difficult part of qualitative research. Strauss and Corbin (1990) provide one of the most thorough descriptions of the process. Please remember that, although they were discussing grounded theory (which means that
the analysis process is continually interacting with data collection), their general approach is usable in all qualitative data analysis.

1. **Open coding** (creating categories): the process of naming and categorizing observations through a close, inductive reading of the data. There are certain stages of this process and tests for adequacy of categorizations.
   a. Phases of the open coding process:
      1. Labeling observations: assigning conceptual labels to small units of the data. Labels should be descriptive and capture the social actors' meanings.
      2. Discovering categories: grouping similarly labeled units
      3. Naming a category
      4. Developing categories in terms of their properties, where a property is a characteristic or attribute of a category. For example, the category "color" can be discussed in terms of the lightness versus darkness of a component.
   b. Tests for adequacy of categories:
      1. Internal homogeneity: Review items that have been assigned to that category to make sure that they make sense together.
      2. External heterogeneity: Compare categories with one another to make sure that there are clear and obvious differences between them. This is also referred to as the test of mutual exclusivity.
      3. Exhaustiveness: Test to make sure that all data can be assigned to one of the categories.
      4. Reproducibility: the ability of a knowledgeable judge or other to reproduce the same category system by reviewing the same data
      5. Credibility/validity: the extent to which a social actor who contributed data assesses the category system as a valid representation.

2. **Axial coding**: This is the coding process that concentrates on identifying subcategories within a main category. It is done using the paradigm model. The paradigm model can be described as "specifying a category and the conditions that give rise to it: the context in which the category is prevalent, the action/interaction strategies by which it is managed, and the consequences of those categories."

3. **Selective coding** (usually more for grounded theory): the process of generating relationships between categories in order to develop a core theory.
FIELD OBSERVATION

One of the major means of measurement or data collection in qualitative/interpretive research is through the use of observational techniques, usually referred to as "participant observation". In a sense, this term may be misleading since not all observers actually act as participants of the social groups or events that are being observed. The information here on participant observation, taken from Patton (1980), Anderson (1987), and Douglas (1976), addresses qualitative observation in general.

I. Five dimensions of variation in approaches to observation. Any observational study varies in terms of these five basic dimensions.

A. Role of the evaluator-observer: This is the extent to which the observer is an active participant versus a detached observer.
   1. Complete participant: the observer is a long-term and invested member of the social action/group being observed. This role provides a complete "insider" view.
   2. Participant as observer: the observer is an active participant in the social action but attempts to "step back" from the situation in order to appreciate the nature of the action.
   3. Observer as participant: the observer becomes involved in the social action that she was previously uninvolved in, to better understand the action.
   4. Complete observer: the observer does not become involved in participation in any way, remaining on the outside looking in.

B. Portrayal of the role to others: This is the extent to which the observer allows social actors to know that they are being observed.
   1. Completely overt: all social actors being observed know that they are being observed
   2. Partially overt: some social actors are informed that they are being observed, but others aren't.
   3. Completely covert: none of the social actors know that they are being observed.

C. Portrayal of purpose of evaluation to others: This is the extent to which the observer allows the social actors to know the reasons for observation.
   1. Completely explained: the true purpose of the evaluation is completely explained to all actors.
   2. Partially explained: part of the true purpose of observation is shared with the social actors (usually a very general statement of purpose only)
   3. Completely unexplained: none of the social actors is informed of the purpose of the observation
   4. Falsely explained: the social actors are deceived into thinking that the observation is for a particular purpose, when, in fact, it is for another purpose.

D. Duration of observation: This is the amount of time over which the observer conducts observations. It can range from a single, limited observation session to a very long-term, multiple observation situation.
E. **Focus of observations:** This is the amount and type of information that is being observed. In general, it differs on a dimension of specific elements being observed to wholistic observation of the social context. The observer must decide what to observe and why. Different things that may be observed include:
1. Physical environment
2. Social environment (how people organize themselves into groups, subgroups, or relationships)
3. Social activities (group activities, rites, rituals, etc.)
4. Participant behaviors (including the entire range of possible individual verbal and nonverbal behaviors)
5. Documents and artifacts
6. Nonoccurrences

II. **Process of observation.** This really involves basic design considerations in conducting participant observation.

A. **Entering the field:** This is the initial stage of immersion in the context, especially for nonparticipant observers. In this stage, the observer usually attempts to be as wholistic as possible, not discarding possible modes of observation in an a priori fashion. Some suggestion are:
1. Remain passive and unobtrusive.
2. Do not challenge behaviors of social actors.
3. Avoid allowing others to define your role for you.
4. Concentrate on the most explicit parts of the social environment.

B. **Establishing rapport:** As the observer becomes more familiar with the context and more focused on the nature of the desired observations, it is imperative that she gain the support/comfort of the social actors. Some suggestions are:
1. Attempt to emphasize commonalities.
2. Begin to participate slowly, always testing for the social actors' acceptance of your participation.
3. Don't react to hostility of social actors in a hostile manner.
4. Learn their language and behavior norms.
5. Avoid establishing overly close relationships with a few members.

C. **Completing observation:** By this time, you should have successfully negotiated your research presence and should be able to concentrate completely on your observations of interest.

III. **Field notes.** The quality of your analysis will ultimately depend upon the quality of the field notes or documentation you construct. Writing field notes is an extremely difficult and time-consuming activity and will yield a great deal of data for a project of any size.

A. **General guidelines for field notes:**
1. Detailed as possible
2. Complete and comprehensible
3. Accurate

B. **Components of field notes:**
1. Observational notes: descriptions of activities or behaviors that were observed. 
2. Methodological notes: descriptions of events or factors during the observation that may have affected one of the three guidelines above (especially accuracy). 
3. Theoretical notes: ideas and beginning interpretations about the observations. 
C. Preparation of field record: After each observation session, the observer should carefully prepare field notes while the information is freshly in mind.
QUALITATIVE INTERVIEWING

One of the primary means of obtaining information in a qualitative/interpretive study is by interviewing social actors. This section of notes discusses qualitative interviewing in general, and presents two types of specialized qualitative interviews: the focus group interview and the critical incident interview. The material here was gleaned from Patton (1980), Krueger (1988), Morgan (1988), McCracken (1988).

I. General characteristics of qualitative interviews:
   A. **Nondirective.** This means that the interviewer allows the interviewee a fair amount of control over what is actually discussed in the interview. There are gradations of this, however; most qualitative interviews fall decidedly on the nondirective end of the directive/nondirective dimension.
      1. This increases the amount of information obtained.
      2. It may decrease the reliability or standardizations of information obtained across interviews.
   B. **General-to-specific format.** Most qualitative interviews follow a funnel sequence, beginning with rather broad questions and slowly narrowing to more specific questions and information.
   C. **Lengthy.** Most qualitative interviews are relatively lengthy, whether they are individual or group interviews. They may range from 30-45 minutes to three or four hours in length.

II. General concerns in interviewing. As in all interviews, it is helpful to remember that there are several possible barriers to effective interviewing that arise in each situation. The five most common are:
   A. Language influences
   B. Psychological and background influences
   C. Behavioral influences
   D. Role influences
   E. Environmental influences

III. Individual interviews.
   A. There are three types:
      1. **Informal conversational interviews.** These are "interviews" that occur in the field without the "interviewees" being aware that they are being interviewed. Basically, this type of interview resembles a normal conversation, although the responses of the other are being mentally noted by the interviewer. In most cases, informal conversational interviews are spontaneous, unstructured, very nondirective, and covert. They are most useful for uncovering information that is potentially sensitive, or that may be altered if a respondent realized he or she were being monitored.
2. **General interview guide.** This interview is a moderately structured interview that occurs with the knowledge of the respondent. It consists of a small number of generally oriented topic questions and allows the researcher to probe at will into comments or areas that are raised during the interview. Thus, an interviewer may use the same general guide to interview a number of people in the field, but the results of each interview may be very different depending upon the direction taken during the interview.

3. **Standardized open-ended interview.** This interview is highly structured and is administered in the same way to each interview subject in the study. Highly structured means that the specific questions (both primary and secondary) are carefully worded and sequenced prior to the interview, and that deviations from the plan are discouraged. This type of interview is most important when comparability of information is critical.

**B. Target question areas:** According to Patton (1980), there are several types of questions that may prove useful. They are:

1. **Experience/behavior questions** -- about what a person does or has done.
2. **Opinion/value questions** -- aimed at understanding the cognitive and interpretive processes of the interviewee. These questions tell what the person thinks about a particular thing.
3. **Feeling questions** -- ask about the emotional reaction of the person to objects, people, or events.
4. **Knowledge questions** -- asked to find out what factual information the person has.
5. **Sensory questions** -- ask how the interviewee experienced something (its touch, taste, smell, etc.).
6. **Background/demographic questions** -- relatively routine questions about age, occupation, education, social status, etc.

**IV. Focus group interviews:** A focus group interview (traditionally related to marketing research but applied in general qualitative research as well) is a relatively nondirective, unstructured discussion with people who are in some way representative of social actors of interest. The purpose of the focus group can be two-fold: 1) to gain information for the development of an individual interview schedule, or 2) to gather information in a more expedient manner. The intent of a focus group interview is to promote self-disclosure of attitudes and opinions in an informal discussion context.

**A. Characteristics of focus groups**

1. They involve (usually 7 to 10) people
2. The participants typically share certain characteristics
3. The data gathered is of a qualitative nature
4. The group enters a focused discussion (meaning that topics of discussion are predetermined and sequenced although the range of discussion on the topics is relatively free-flowing).

**B. Criteria for effective focus group interviews:**
1. Should cover a maximum range of relevant topics
2. Should provide data that is as specific as possible
3. Should foster interaction between group members that explores the participants' feelings in depth
4. Should take into account the personal context operating for participants

C. Advantages of focus group interviews:
   1. Naturalistic
   2. Group interaction encourages thought
   3. Format allows facilitator to probe
   4. High face validity
   5. Relatively inexpensive
   6. Time efficient
   7. Enables increased sample size in qualitative studies

D. Limitations of focus group interviews:
   1. Researcher has less control than in individual interviews
   2. Data are more difficult to analyze
   3. Requires carefully trained interviewers
   4. Groups can vary considerably in nature and ability
   5. Groups may be difficult to schedule/assemble

E. The process of conducting focus group interviews:
   1. Conceptualizing
      a. Purpose
         1. What information is needed?
         2. Why is the information needed?
         3. Who will use the information?
      b. Participants
         1. Who will/should they be?
         2. How will they be contacted?
         3. How many should there be?
      c. Timeline
         1. How long will the entire interview take?
         2. How long will the entire study take?
   2. Asking questions in focus group interviews
      a. Use open-ended questions
         1. Begin with individual reactions to question
         2. Proceed to open discussion of question
      b. Use feelings/opinions/experiences questions
      c. Pilot test the interview
   3. Conducting the focus group interview
      a. Welcome
b. Overview and topic explanation  
c. Explain ground rules  
d. First question  
e. Probe/question/probe  
f. Involve all participants in discussion  
4. Analyzing the data  
1. Inductive analyses  
2. Content analysis  

V. Critical incident interviews: This is a specialized form of interviewing (it has even been translated into open-ended survey formats) usually used in individual qualitative interviews (although it can also be converted to group interviews). It asks people to comment on critical positive and/or negative incidents they have experienced that are relevant to the phenomenon of interest. The underlying assumption is that people's recall of events that they define as critical (either positively or negatively) indicates extremely salient information and provides excellent contrast of positive and negative aspects of the phenomenon.  
A. Basic format of critical incident interviews  
1. Ask for recall of a positive (then negative) incident  
2. Allow them to elaborate with as much detail as possible on the nature of the incident  
3. Ask for their reactions to the incident  
4. Ask for their assessment of consequences of the incident  
5. Then repeat process with other pole of questioning  

B. General question guidelines  
1. Phrase questions as open-ended  
2. Phrase questions neutrally  
3. Phrase questions specifically  
4. Probe for specifics throughout  

VI. A summary of evaluation research interview instrumentation.  
A. Informal conversational interview  
1. Characteristics: Questions emerge from the immediate context and are asked in the natural course of things; there is no predetermination of question topics or wording.  
2. Strengths: Increases the salience and relevance of questions; interviews are built on and emerge from observations; the interview can be matched to individuals and circumstances.  
3. Weaknesses: Different information collected from different people with different questions. Less systematic and comprehensive if certain questions don't arise "naturally." Data organization and analysis can be quite difficult.  

B. Interview guide approach
1. Characteristics: Topics and issues to be covered are specified in advance, in outline form; interviewer decides sequence and working of questions in the course of the interview.

2. Strengths: The outline increases the comprehensiveness of the data and makes data collection somewhat systematic for each respondent. Logical gaps in data can be anticipated and closed. Interviews remain fairly conversational and situational.

3. Weaknesses: Important and salient topics may be inadvertently omitted. Interviewer flexibility in sequencing and wording questions can result in substantially different respondents, thus reducing the comparability of responses.

C. Standardized open-ended interview

1. Characteristics: The exact wording and sequence of questions are determined in advance. All interviewees are asked the same basic questions in the same order.

2. Strengths: Respondents answer the same questions, thus increasing comparability of responses; data are complete for each person on the topics addressed in the interview. Reduces interviewer effects and bias when several interviewers are used. Permits decision makers to see and review the instrumentation used in the evaluation. Facilitates organization and analysis of the data.

3. Weaknesses: Little flexibility in relating the interview to particular individuals and circumstances; standardized wording of questions may constrain and limit naturalness and relevance of questions and answers.

D. Closed quantitative interview

1. Characteristics: Questions and response categories are determined in advance. Responses are fixed; respondent chooses from among these fixed responses.

2. Strengths: Data analysis is simple; responses can be directly compared and easily aggregated; many questions can be asked in a short time.

3. Weaknesses: Respondents must fit their experiences and feelings into the researcher's categories; may be perceived as impersonal, irrelevant, and mechanistic. Can distort what respondents really mean or experienced by so completely limiting their response choices.
CONVERSATION AND DISCOURSE ANALYSIS

In a broad sense, both conversational and discourse analysis are methods used to discover patterns of interaction in ongoing conversation. As such, some argue that they fall under the more general headings of interaction analysis, or even content analysis (although this is also debated by others). The material in this section of notes is taken from six sources (see suggested references): Smith's (1988) chapter on conversational analysis (Ch. 12); McLaughlin's (1984) book on conversation (both her chapter and the chapter by Hopper, Koch, and Mandelbaum); Sigman, Sullivan, and Wendell's chapter in C. Tardy's (Ed.) book (1988), and Heritage's and Wetherall and Potter's chapters from C. Antaki's (Ed.) book (1988).

I. Nature of conversational analysis and discourse analysis
   A. Conversational analysis. A systematic method for describing, explaining, and evaluating the structure and function of rules governing conversations. It is a subdiscipline of ethnomethodology which is primarily concerned with social action and its underlying reasoning. The focus of conversational analysis is descriptive and naturalistic rather than experimental or explanatory.
      1. Five themes or underlying assumptions of conversational analysis:
         a. Conversation as a structured social activity. The assumption is that talk is a structured activity and therefore patterns are discernible.
         b. Need for an interactional vocabulary. Conversational analysis should focus on the interactional elements in conversation rather than on the social or social role elements in the context.
         c. Nondeterministic understanding. The assumption is that conversation is a rule-governed behavior and that behaviors can serve multiple functions and therefore can be interpreted in a number of ways. Essentially, people have choices to make about how they engage in conversation and each conversational element may reveal a number of different choices for different considerations.
         d. Local orientation. The assumption is that conversation occurs in a context and is emergent from that context. Each conversational utterance is shaped by prior talk and, in turn, shapes subsequent talk. Thus, the emphasis is on analysis of naturally occurring interaction with a focus on the sequencing of conversational elements.
         e. Continuous and emergent categorization. Conversational analysis relies on inductive analyses to make claims about conversational processes. The emphasis is on indigenous rather than researcher-derived constructs of meaning. Concomitantly, no order of detail in the interaction is dismissed a priori as irrelevant.
      2. Basic concepts in conversational analysis:
         a. Conversational rule: An "if-then" assumption that defines the common behaviors required to achieve desired outcomes or objectives in a social situation. Rules can be broadly defined as constitutive (also called interpretive) that allow conversationalists to make sense of utterances, and regulative (also called behavioral) that allow conversationalists to determine behavioral alternatives that are appropriate (that are required, preferred, prohibited, or permissible in the situation).
b. **Conversational structure:** The way people organize meaning by producing interrelated utterances. Structural coherence is the extent to which meanings created by utterances coincide with the general goal of the conversation, and thus, is a major focus of conversational analysis.

c. **Conversational functions:** What people "do to" each other when they talk (e.g., complimenting, etc.). A primary objective of conversational analysis is to assess a conversation's functional coherence, or the extent to which a sequence of utterances produces the actions desired by the participants.

B. **Discourse analysis.** Like conversational analysis, discourse analysis analyzes sequential structures of talk. However, discourse analysis usually involves coding utterances into categories of function or illocutionary force. Then, discourse analysts search for statistical relationships among these categorical variables. As such, discourse analysis is more consistent with quantitative interaction analyses.

1. **Assumptions of discourse analysis**
   a. Discourse is action-oriented. Discourse serves basic functions, whether they are intended or not, explicit or implicit, overt or covert. The understanding of functions is one of the goals of discourse analysis.
   b. Discourse is rule-governed. This is essentially the same as the assumption of nondeterministic understanding above.
   c. Study of variation is critical. Since discourse varies with purpose, context, and time, the attention to variation helps us to understand different structures and their functions.

C. **Differences between discourse analysis (DA) and conversational analysis (CA).** Although overstated, the differences can be summarized as follows:

1. CA focuses more heavily on inductive analysis while DA is more deductive in approach.
2. CA is more concerned with preserving the nuance of conversation through microanalysis of conversation, while DA is more concerned with identifying broad patterns of relationships.

II. **Methodological considerations in conversational analysis.** The remainder of this section of notes will focus exclusively on CA, its methods and procedures.

A. **What counts as data?** In a word, conversation. The source of conversations can be any or all of the following:

1. Naturalistic conversation gathered in the field
2. Naturalistic conversation gathered in the lab or some other controlled environment
3. Specific speech elicited by the researcher (asking people to engage in role plays)
4. "Spoken" conversation from literary or historical sources
5. Researcher-constructed examples

B. **How is data collected?**

1. Audio recording
2. Videotaping
C. Sampling: Usually purposive and consists of small amounts of conversation from a relatively few number of interactants

D. How is data analyzed?
1. Preparation of transcripts: All taped interaction must first be transcribed. The method of transcription used for conversational analysis is very specific and extremely time-consuming.
   a. Repeated listenings to tapes
   b. Requires notation of paralinguistic cues concurrent with linguistic content (pitch, loudness, inflection, and in the case of videotaping perhaps also notation of gestural and gaze behaviors)
   c. Requires notation of the precise moments that utterances begin and end (especially important for assessment of turn taking, interruptions, overlaps, etc.)
2. Analytic purposes may be one or more of the following:
   a. To describe recurring patterns of interaction
   b. To explain observed patterns by referring to constitutive or regulatory rules
   c. To evaluate or critique the extent to which an interaction has attained conversational coherence.
3. Four types of evidence (things that data can be analyzed in terms of):
   a. Conversational regularity: showing that a particular pattern of utterances recurs on a persistent basis.
   b. Rule knowledge: showing that communicators have knowledge of certain rules that are operating and the impact of rules on conversation.
   c. Context effects: showing that people's conversational rules change depending upon context.
   d. Consequence effect: showing that adhering to or ignoring a rule results in positive or negative consequences.
4. Analytic procedures:
   a. Use multiple listenings to tapes with transcripts.
      1. Focus on very brief segments (5 seconds).
      2. Listen to each segment at least 20 times.
      3. Note line-by-line how speakers are performing conversational activities.
   b. Reasoning from conversations.
      1. Create a formal description of "how" conversation is being accomplished.
      2. Test these instances against others to demonstrate patterns or variation.
CONTENT ANALYSIS

Content Analysis is a research technique for categorizing observational data for the purposes of describing the phenomenon or testing relationships between the phenomenon and some other variable of interest. Content analysis may be used with a variety of types of observational data (conversations, statements, media artifacts, behaviors, etc.). Content analysis is usually unobtrusive, meaning that the producers of the messages do not know that their behaviors or communication acts are being (or will be) studied. This section of notes, derived from Krippendorff (1980) will discuss content analysis.

I. Types of data. Almost anything can serve as data for content analysis. Data from observations of behavior, documents or artifacts, and interviews are all amenable to this method. What is needed is a permanent record of the data in the form of a tape recording, transcript, or copy of the document.

II. Measurement. The basis of content analysis is the application of a category system to "measure" the data. In this sense, the category system is the measurement. In an earlier section of notes we discussed the development of category systems in interpretive research. In quantitative/descriptive content analysis, the assumption is that the researcher selects a category system that already exists (either one that has been developed by someone else, or one that was developed earlier by the researcher). There are several guidelines for an acceptable category system. Such a system should be:

A. Valid for the phenomenon of interest
   1. It should have been developed specifically for the assessment of this phenomenon.
   2. It should be tailored to fit this phenomenon.

B. Relatively exhaustive, meaning that all (or almost all) instances of the data should fit into one of the categories (i.e., be categorizable).

C. Mutually exclusive, meaning that each observation should fit into one and only one of the categories.

D. Workable, meaning that it should have enough clarity to be understood and applied by coders and should be simple enough to produce acceptable levels of reliability.

E. Pilot-tested using samples of observations that will not be included in the final data analysis. This will enable the researcher to work out the bugs.

III. Sampling. In content analysis, there are actually two different sampling processes. The first concerns the selection of the entire sample of observations and the second concerns the division of sampling elements into codable units.

A. Sampling units (units of analysis). These are separable by physically identifiable boundaries and are completely independent of each other. Examples include the books on a certain topic, newspapers in big cities, and prime-time television programs.

B. Recording units (units of observation). These are the segments within the sampling units that will actually be coded, the separately analyzable parts of sampling units. There are a
variety of types of recording units that are possible and their appropriateness depends
t upon the nature of the data and the nature of the research question. Some common types
of recording units are:
1. Words
2. Phrases
3. Complete sentences
4. Thought units
5. Themes
6. Utterances
7. Time units (seconds, hours, days, weeks, etc.)
8. Pictures
9. Video frames
10. Scenes
11. Program segments
12. Commercials
13. Behaviors (nonverbal)

C. Context units. These are ways of defining the context in which a recording unit occurs,
and thus help to understand the meaning of the recording unit. An easy example is the
use of an interact as a contextual unit for an act, or the use of a show as a contextual unit
for a scene.

IV. Design. In a content or interactional analysis study, the research design consists of the
series of logical steps that the researcher must follow in order to select, code, and analyze
the data. We have already discussed the first basic steps of selecting a coding system and
sampling data. From this point on, there are several more design considerations:
A. Coders. Coders are the people we will train to actually apply the category system to the
data. There are several things to consider here:
1. Who should the coders be? Most people argue that coders should fulfill several
criteria:
   a. Coders should be somewhat knowledgeable about the phenomenon but not too
      knowledgeable.
   b. Coders should be serious and dedicated.
   c. Coders should be accessible for training and coding time required.
2. How many coders? Usually people suggest that we use two or three coders. The
   more coders we have, the shorter the time it will take to actually code the data.
   However, the more coders we have, the more difficult it will be to establish
   acceptable levels of reliability.

B. Training the coders. All coders have to be trained to apply the category system to our
data in a reliable and valid fashion. Training involves several steps.
1. Create a coding sheet. This is a sheet that coders use to actually record categorization for recording units. The coding sheet should contain spaces for the provision of the following information:
   a. Identification of coder
   b. Identification of sampling unit
   c. Identification of recording units
   d. Identification of variables, and values of variables, to be coded
2. Create a codebook: The codebook is the manual that coders will use to code the data. It should include the following:
   a. A copy of the complete category system. All categories, their conceptual definition and examples of their operationalization should be provided.
   b. An explanation of unitizing rules, which are guidelines that help coders decide how to code unusual (e.g., incomplete or complex) recording units.
   c. Prioritization of categories. In the case of a recording unit that may be codable into more than one category, this tells the coder which category to give preference to. The prioritization should be based on sound theoretical arguments and should be congruent with the purposes of our research.
   d. Decision rules. These are rules that coders should follow in the case of coding difficulties such as interruptions, poor quality data copying, etc.
   e. Schedule of coding. This describes what sampling units the coder should be coding and in what order. In addition, it should suggest a general time frame for the completion of the coding.
3. Select training materials. We will need some samples of data to use in training the coders before we set them loose on our actual data. If we have thought ahead, we will have oversampled data and can use some of that for training purposes. If there isn't enough data in our sample for this, we can "create" similar data for training.
4. Select reliability materials. We will need to test for reliability among our coders using data that is as similar as possible to the actual data they will be coding. Follow the same general guidelines suggest above.
5. Establish a training schedule. Usually we should plan on training coders in two to three sessions per week for a maximum of 2 to 3 hours per session. The number of weeks it will take to train is dependent upon the complexity of the category system we are using and the complexity of our data. Be patient. It always takes longer than you think it will.
6. Train coders. Training should follow a basic process:
   a. In the initial meeting(s) provide them with copies of the codebook and have them carefully read through the entire book. Then discuss its content with them.
   b. Next, present them with a small number of recording units and, as a group, discuss how they would categorize them and why. Continue this process until you feel that all coders have a good grasp of the entire category system.
   c. Third, present them with recording units and have them individually, and without any discussion, code them. Then have them discuss their coding decisions and deal with disagreements.
   d. Finally, present them with a group of recording units, have all coders privately code all units, and discuss afterward. If this goes well, you are ready to test reliability.
C. Testing reliability. Determine whether the coders are reliably coding the data before turning them loose on the data set. As in everything else, there are general steps to follow:

1. Create a reliability data set. This should consist of approximately 150-200 recording units of data.
2. Have coders simultaneously and privately code unit by unit. There should be no discussion between coders.
3. Calculate reliability using one of the various formulas (Cohen's Kappa, Scott's pi, Holsti's formula, Krippendorff's coincidence matrix formula, etc.).
   a. Reliability coefficients should be at least .80.
   b. Calculate reliabilities for general categories as well as an overall reliability if the coding system is complex.
   c. If reliabilities are good, then proceed with coding.
   d. If reliabilities are not satisfactory, then go back to the training table.

D. Monitor coding process. Keep in touch with each coder as they work their way through coding the data. Help them with any troubles that arise.

V. Data analysis: Once the coding is complete and the data are gathered, subject it to a variety of statistical analyses to test your hypotheses or answer your research question(s).
SURVEY RESEARCH

This section of notes loosely follows the outline of Chapter 11 in Smith (1988). However, additional material has been added from Fowler (1984), and Converse and Presser (1987).

Survey research is a method of constructing a series of questions in the form of a questionnaire to examine sample(s) from a population for the purposes of description and/or explanation. In the case of description, you are trying to define the pertinent characteristics of a population. In the case of explanation, you are trying to supply reasons for those characteristics. In addition, there are three more reasons for deciding to do a survey:

1) Probability sampling allows you to have confidence that the sample is truly representative and allows you to calculate an index of operative error.
2) Standardized measurement allows you to compare data across samples and across members within a sample.
3) Survey questionnaires often allow you to collect data that would be unavailable from any other source.

I. Types of surveys. Smith (1988) argues that there are three general types of surveys:
   A. Cross-sectional surveys collect information from one sample at a single point in time.
   B. Longitudinal surveys collect several samples at different points in time. She notes two types:
      1. Trend studies. Here you draw more than one representative sample at different times. These are recommended only if your population is sufficiently large and if representative samples are needed for your research purposes.
      2. Panel studies. Here you follow the same sample across time. This is recommended if a purposive sample is needed and if you do not expect to experience mortality problems in your sample.
   C. Multiple sample surveys. These describe and explain relationships among two or more populations.

II. Five steps in survey research. These steps have been significantly embellished through the use of information from the two additional sources indicated above.
   A. Step 1: Select a representative survey sample
      1. Define your population.
      2. Specify sampling elements.
      3. Secure an adequate sampling frame or list of elements in the population.
      4. Select a sampling method.
      5. Determine your required sample size.

   B. Step 2: Design a survey questionnaire.
      1. Determine the data collection method (i.e., self-administered questionnaire, personal interview questionnaire, mail questionnaire, or telephone interview).
2. Define your objectives. The purpose of your survey is the critical basis for designing the actual questionnaire:
   a. Write a paragraph about what the survey is supposed to accomplish.
   b. Make a list of all critical variables or concepts that should be measured.
   c. Conduct focused discussions with members of the population and/or experts in the area to identify specific topics for inclusion.

3. Review the types of questions that are possible (see below for specialized types of questions)
   a. Open (the respondent can answer in any way she chooses) versus closed (the respondent must choose one of the options the researcher provides)
      1. Rating scales. A closed question type in which the respondent rates objects, people, etc., usually on a 5 or 7 point scale
      2. Ranking scales. A closed question type in which the respondent ranks or orders a series of objects, events, etc.
      3. Inventory questions. A closed question type in which the respondent selects as many options as apply.
   b. Leading/loaded (e.g., "Don't you think . . . ?" or "How bad is . . . ?") versus neutral
   c. Factual (i.e., a measure of respondent's knowledge) versus opinion (i.e., a measure of respondent's perceptions)
   d. Primary (initial) versus secondary (repeated or follow-up)
      1. Mirror questions. A question asked in a manner opposite that of the primary question to see if respondent's answers are consistent.
      2. Confrontive probes. A follow-up question that is asked in a confrontational manner to test the strength and consistency of respondent's views.
      3. Reactive probes. A follow-up question that challenges respondent to reconsider her earlier answer.
   e. Vignettes/hypotheticals versus informational
   f. Contingency and filter questions. Questions which are asked or not asked depending on the respondent's earlier answers.

4. Criteria for constructing questions
   a. Questions should be clear and unambiguous.
   b. Questions should be simple and brief (usually no more than 20 words unless you are trying to stimulate their recall, in which case the introduction to the question itself can be longer).
   c. Questions should not be leading.
   d. Questions should not be double-barreled (i.e., asking more than one thing in a question).
   e. Questions should be as nonthreatening as possible.
   f. Questions should not include single and especially double negatives.
   g. Questions should not include implicit negatives (like control, forbid, restrict).
   h. Questions should not have response lists that are too long for the respondent to remember (more than four or five responses) unless you are going to use a show card.
i. Questions should be specific rather than general. Specific questions have been shown to increase recall, reduce misinterpretation of the question, and be better predictors of behavior.

j. Questions should usually be closed unless not enough is known to write appropriate response options or unless you are asking about sensitive or disapproved behavior.

k. Questions should offer no-opinion, "Don't know," and/or "Not applicable" option(s).

l. Questions should omit the middle alternative and use follow-up questions that gauge the respondents' intensity of attitude (usually through the use of forced-choice questions rather than agree-disagree questions, since the latter are prone to acquiescence response bias where respondents will agree regardless of the topic of the question).

5. Special problems with questions

a. Recall of the past. Questions often ask respondents to recall events that have occurred in the past. This can be difficult for them. Some ideas to help effective recall are:
   1. Ask them questions about topics that are salient to them; the more salient the better the recall.
   2. Narrow the reference period (limit it to the past 6 months or less if possible).
   3. Ask respondents to "average" over time.
   4. Use landmark events as guides to help put respondents in the right frame of time.
   5. Ask more than one question about the topic.
   6. Give them a second chance to think about and revise answers.
   7. Interview the same person more than once if necessary.
   8. Ask them to keep a diary or jot down notes.

b. Sensitive topics. Some topics can threaten respondents or make them uncomfortable so that they avoid answering or try to give the socially appropriate response. To counteract this:
   1. Minimize a sense of judgment.
   2. Use self-administered questions.
   3. Insure confidentiality and anonymity if possible.

6. Choose the appropriate ordering of questions. This has to do with the overall format and flow of the questionnaire.

   a. Adopt a general pattern that complements the survey's research objective
      1. Funnel (general to specific)
      2. Inverted funnel (specific to general)
   
   b. Topically related questions should be grouped together.
   
   c. Very difficult or complex questions should not be placed at the beginning of the survey.
   
   d. Demographic questions, which are easy-to-answer, should be placed at the beginning of a survey that is administered by an interviewer, but in a self-administered survey (where the interviewer is not there to encourage the respondent to finish), the demographic questions should be placed at the end.
e. All contingency and filter questions (or skip patterns) should be clearly marked
with clear instructions (for either the interviewer or the respondent) concerning
how to proceed.

f. For administered interviews use several types of questions to keep up the
respondent's interest.

g. For self-administered surveys only use a few question forms to reduce the
confusion of the respondent (but to avoid boredom, do not use just one question
form).

h. When using hypothetical questions, ask two or three and then follow with an
"actual experience" question. Also, ask respondents to indicate the frame of
reference they used to answer a hypothetical question.

i. Important: Carefully consider the effect that a preceding question may have on the
response to the succeeding question.

j. Important: Don't make the survey too long. Remember that the respondent has
limited time and is doing the researcher a favor. Even if she completes an
excessively long survey, disinterest and frustration may affect the validity of the
data provided by the respondent.

C. Step 3: Pretest the questionnaire. Pretesting is critical for any questionnaire.

1. Purposes of pretests.
   a. To determine how much variation in response to a question there is in your
      population
   b. To determine how respondents interpret the meaning of a question
   c. To determine how difficult it is for respondents to answer specific questions
   d. To gauge respondent interest and attention throughout a survey
   e. To determine the flow of the questionnaire (is it in need of better transitions, does
      it logically proceed from one topic to another)
   f. To determine how long it takes to complete the questionnaire
   g. To determine if there are any problems with respondent well-being, i.e., are
      respondents protected in their answers?

2. Forms of pretesting.
   a. Informal peer or convenience sample pretest. Especially appropriate for testing
      the first version of a new survey, this involves having colleagues and friends and
      family review, take, or administer the survey.
   b. For an administered interview: Have interviewers interview approximately 10-20
      members of the population, telling the respondents the nature of the pretest and
      soliciting their input in terms of ideas and specific feedback about questions.
      These pretests should be taped and systematically analyzed.
   c. For a self-administered questionnaire (requires more pretesting because
      interviewers won't be there to fix problems as they crop up): Ask a group of
      subjects (25-75) to complete the questionnaire, then lead a discussion about
      difficult or confusing questions, sensitive questions, adequacy of instructions, and
      language level/reading level.
3. Evaluating a pretest can be done by several different methods, most of which require cooperation from your interviewers or administrators:
   a. Marginal comments on schedule
   b. Oral debriefing
   c. Written reports
   d. Field observation of questionnaire in use
   e. Coding of answers and tallying of marginal frequencies
   f. Complete data analysis. This is especially useful because it forces the researcher to develop an analysis plan for each question before the data is actually collected.

D. **Step 4**: Administer the survey to the sample.
   1. Procedures
      a. Mail surveys
         1. Enclose an explanatory (cover) letter
         2. When the perception of anonymity is not a major concern consider putting identification numbers on the surveys. This allows the researcher to keep track of which surveys are returned (which can be especially useful if they are mailed over a period of time).
         3. Postage alternatives include self-addressed stamped envelopes, bulk mailing permits, and no postage but a survey that folds closed and is already addressed.
         4. Follow-up phone calls to encourage higher response rate (see below).
      b. Phone and personal interviews
         1. Finding, training, and paying interviewers can be difficult.
   2. Response rates. Watch out for response rate problems. As Smith (1988) argues, there is total nonresponse or item nonresponse. In either case the big worry is whether or not the nonresponse is biased in some way such that the people who are not answering (overall or specific items) are significantly different than other members of your sample.
      a. The nature of nonresponse differs with mail, telephone and personal interviews.
         1. Gauge nonresponse bias in mail surveys by comparing those who answer immediately with those who answer after follow-up and those who don't answer at all. The main problem here is that people with a high interest in the topic are more likely to respond.
         2. In telephone and personal interviews the big nonresponse bias problem is in access and availability -- who you can reach and when.
      b. To reduce nonresponse rates several suggestions are provided on the attached pages.
E. **Step 5: Prepare survey data for analysis**
   1. Format or organize data
   2. Design a code (rules to change responses into numbers). Note that some surveys include these codes, but for self-administered surveys this can be distracting to the respondent.
   3. Code the responses
   4. Data entry
   5. Data cleaning (check data file for accuracy and consistency prior to analysis)

III. Specialized types of survey questions or instruments. Many researchers construct their own measurement instruments (questionnaire items or other means of eliciting survey-type responses). Others use existing ones. In either case, the researcher needs to know the logic behind the instrument, the most common types of instruments in communication research, as well as the benefits and advantages of each.

A. **Semantic differential.** This is a procedure involving standard scaling procedures and a variety of analytical methods. Usually, it involves bipolar adjectives on seven-step interval (assumed) scales.
   1. Examples of scales:
      a. Good ___:___:___:___:___:___:___ Bad
      b. Ugly ___:___:___:___:___:___:___ Pretty
      c. Poor ___:___:___:___:___:___:___ Rich
   2. How they are created. These scales may be specially constructed for a particular task, or selected from existing sets by any number of criteria. Most commonly, and unfortunately, scales are selected to represent results of previous factor analyses.
   3. Criticisms.
      a. The researcher provides the scales. She cannot assume and assert that these scales represent the respondents' way of thinking about the focal object.
      b. It is assumed that the adjectives are bipolar. This is not necessarily true.
      c. It is assumed that the intervals of the scale are equal. This is not necessarily true, and it is almost impossible to prove.

B. **Likert scales** (summated ratings). These are often confused with semantic differential scales. The major difference between the two is that the Likert scales are summated rather than treated individually.
   1. Examples of scales:
      a. "I hate this program."
         SA A U D SD
      b. "I get annoyed with statistics."
         YES Yes yes ? no No NO
   2. How they are created. The scales are usually constructed by the researcher. The researcher selects a large number of statements about the phenomenon under study. Each statement is accompanied by some 5 or 7 interval scale that indicates the levels
of agreement/disagreement with the statement. Subjects complete the scales for each statement. Responses are summed to give a composite score for each subject. (Some statements are usually positive and some are usually negative in terms of the way they are worded. Therefore, the researcher has to use reverse scoring.)

3. Criticisms. Same as with semantic differential scales with the exception of the question about the bi-polarity of adjectives. Additionally, since Likert scales are summed, an additional problem is that the items should reflect a unidimensional construct or at least adequately reflect single dimensions of a multi-dimensional construct with scales being summed within but not across dimensions.

C. Thurstone scales (equal appearing intervals). Thurstone scales are attitude-measurement scales. There are two distinct phases to their development: 1) development of items, and 2) application. This is a very involved procedure. Think carefully about whether it is really necessary for your project before using it.

1. Development of items:
   a. Select 100 or more attitude statements.
   b. Using the Q-sort methodology, have a group of respondents sort the cards into 11 piles according to how much each person agrees or disagrees with each attitude statement.
   c. From the results obtained, select 22 (or 33) attitude statements which represent each step of your 11 step continuum. With 22 statements you would be selecting 2 statements from each step. One way to select statements is to take a mean ranking for each of the 100+ statements. In other words, add the values of the positions that various respondents assigned to each statement and divide by the number of respondents. That would give a general idea of where each statement landed on the continuum. Then, toss out statements with a standard deviation of +/- 2.00 (or 3.00, whichever works better).

2. Application:
   a. Choose a totally different group of respondents.
   b. Present them with the 22 statements. Ask them to indicate which statements they agree with and which statements they disagree with. Examples:
      1. I like cookies. Agree / Disagree
      2. I like ice cream. Agree / Disagree
   c. Now, from your pre-test you have a value of 1-11 for each statement, depending upon which group the statement fell into. So, you can compute an "agree" and a "disagree" score for each respondent. You can compare one person with another, or one group with another.

3. Criticisms. Aside from taking a lot of time and effort, the quality of the scale depends upon your precision in developing the pool of items and the decision criteria for selection. Thus, the criticisms are:
   a. It is an awful lot of work.
b. You still chose the original statements.
c. You selected the decision rules for the final test, i.e., how many statements included, and on what basis.

D. Q-Sort. This refers to a specific method of eliciting responses by having subjects sort cards. It is a form of ordering statements, adjectives, or objects by rank. Here are the procedures for this type of survey instrument.
1. Select the statements, words, or objects to be ranked. For the purposes of statistical reliability and stability have 50 to 100 items. The items can be derived from theory or selected at random. If items represent or are derived from theory it is a structured sort; if not, it is unstructured.
2. Place the selected items on cards, one to a card, and shuffle.
3. Instruct respondents to place cards into sets of piles (usually 7 to 11) according to the extent of agreement with the statements/objects. When respondents are required to place a pre-determined number of cards in each pile it is a forced sort; when not specified, it is unforced.
4. The purpose of the Q-Sort is to get a picture of the respondent's attitude. So record which cards the person puts in which piles.
5. Use statistical techniques to determine the relationship between respondents by comparing their individual ratings. You are really looking for a way to correlate the responses of different people and classify the people into groups on the basis of similar/dissimilar perceptions as revealed in the Q-sort.

E. Guttman scaling. This is perhaps the "strongest" type of scaling available. A Guttman scale is one in which the levels within the scale represent a hierarchy of ability/attitude so that the higher levels of the scale include abilities/attributes/competencies included in all lower levels of the scale. The best example of this is weight lifting. Imagine a Guttman scale of 1001 pounds, 901 pounds, 801 pounds, 701 pounds, and 601 pounds. Any person who can lift the 1001 pound weight is presumed to be able to lift all lesser weight values on the scale. Guttman scales are extremely rare in the social sciences and almost nonexistent in communication research.
THE EXPERIMENT

Experimentation constitutes a large portion of research done in the communication field. The primary focus of experimentation is the manipulation of one or more variables, while controlling for the effects of other variables, to determine their impact on one or more other variables. The variables that are manipulated are called independent variables. The variables that are (hopefully) changed by the manipulation are called dependent variables.

This section of notes provides an overview of the experimental method by summarizing, in a step-by-step fashion, key decisions made by the researcher when planning and conducting an experiment.

I. Start, as usual, with a research question.
   A. Identify key concepts
   B. Identify one or more specific hypotheses
   C. Identify appropriate unit of analysis and observation (e.g., Are you interested in an individual or a group process/effect? Do you plan to examine individuals or groups to test your hypothesis?)

II. Identify independent and dependent variables based on key concepts, and decide on tentative operationalizations of each.
   A. Stimulus preparation choices: sample, alter, create
   B. Dependent variable choices: "pencil-paper versus action" continuum of involvement for subject
   C. Degree of impact: experimental reality

III. Select the appropriate venue/format for study
   A. Traditional lab experiments. More control but also more artificiality (less external validity)
   B. Field experiments. Manipulations are by researcher but in a natural environment rather than the lab. This offers less control but is also less artificial (greater mundane realism, external validity, generalizability).
   C. Natural experiments. These take advantage of manipulations "done" by nature or people other than the researcher (e.g., natural disasters, training programs). They typically allow no control and no random assignment of subjects to groups, so have less internal validity.

IV. Consider key experimental design issues
   A. Within or between subject design?
      1. Within subject designs. These have the same group of subjects in different treatment, or treatment and control, groups (for example, all subjects watch both a violent and a neutral TV program).
      2. Between subject designs. These have different groups of subjects in different treatment, or treatment and control, groups (for example, half the subjects watch only the violent program and the other half watch only the neutral show).
3. Guidelines. Within subject is generally better (because each subject serves as his/her own control) and should be used unless there will be too much interference or contamination between treatment(s), or treatment(s) and control(s), if they are experienced by the same person or if there will be significant effects of practice or fatigue.

B. Is a control group necessary?
   1. What it is and why it's used. Typically used in between subject designs, a control group is a group of subjects who do not receive the treatment. Having a control group permits a comparison of the treatment effects with the effects of all other variables except the treatment (for example, a community that receives a long-term health promotion campaign is compared to an identical community that does not receive the campaign).
   2. Guidelines. Control groups are useful if, regardless of their placement in treatment or control groups, subjects' responses on the measures of the dependent variable(s) are likely to be affected by the attention of being in an experiment (this is known as the Hawthorne Effect) or if the effects of events outside the experiment might be confounded with those of the experimental treatment.

C. Is a pretest necessary or possible?
   1. What it is and why it's used. A pretest determines a baseline value on the dependent variable(s) before the treatment (independent variable) is administered. The baseline can then be compared with values on the dependent variable(s) after the treatment, eliminating the possibility that the treatment didn't actually cause (and is only correlated with) the posttest results. For example, if an instructor tests students' knowledge before a lecture (pretest) and again after it (posttest), higher scores on the posttest indicate that the lecture was effective, whereas without the pretest it could be argued that the students didn't learn from the lecture but already knew the material.
   2. Guidelines. The use of a pretest may be impractical because of demand characteristics: subjects may guess the hypothesis and act to help or hurt the experimenter or to present themselves favorably. For example, asking subjects to complete a questionnaire about their attitudes toward members of ethnic groups (pretest), showing them an anti-prejudice film, and repeating the questionnaire (posttest) may arouse suspicions among subjects, who then may respond in the "politically correct" manner even if it doesn't represent their true attitudes.

D. Is there a danger of subject and/or experimenter bias?
   1. What it is and why it's important. Although the problem is more typical with medical experiments (such as tests of new drugs), it is sometimes important in social experiments to prevent the subject and even the person who administers the experiment from unconsciously biasing results because of knowledge about the experiment. For example, in a test of mediation strategies, both subjects and experiment administrators might try harder to make one strategy work if they know it represents the treatment rather than the control condition.
   2. Solutions.
a. Blind experiments. Here subjects don't know whether they're in a treatment or control group (e.g., a placebo can be used: something that appears to be a treatment but is really a control). This controls for attention effects or other subject motivations/effects due to the experiment itself.
b. Double-blind experiments. Here neither subjects nor the administrator of the experiment know which subjects are in which groups, to prevent biases from both. The more objective the dependent variable measure, the less likely there will be experimenter bias (note similarity to manifest versus latent issue here: see Babbie p. 241)

E. What are the ethical issues? (see Aronson & Carlsmith, 1968)
1. Deception. Will it be necessary to deceive subjects regarding the purpose of the experiment and what they will experience if they participate?
2. Will a debriefing following the experiment reverse any negative effects on subjects?
3. Remember that a human subjects review committee at your university must be given a brief summary of your proposed study and will examine it from an ethical standpoint and may require changes.
4. What special issues are relevant because of your subject population? For example, if you have young children watch violent cartoons or commercials for different toys, you may need permission from their parents.
5. All subjects must sign a "release" form which informs subjects about the confidentiality and/or anonymity of their responses, their right to not answer any specific question(s) and to withdraw their participation at any time, and how subjects can register complaints to the human subject committee on campus.

F. Identify potential threats to internal validity. There are two general types of validity to be concerned with in experimental situations. The first is internal validity, which concerns the extent to which you can claim that results from your experiment are due to the manipulation of your independent variable (or variables) rather than some other cause. Cook and Campbell (1979) describe 12 types of threat to internal validity.
1. History. When historical events occur during the course of an experiment and affect the results. For example, in a study of the effects of a course on drug abuse in changing students' attitudes on drugs, a major sports figure dies of a drug overdose halfway through the course.
2. Maturation. The natural growth and change that occurs in subjects that could affect results. This is a more serious problem in long-term experiments than in short-term ones. For example, in a study of the effects of diet on growth in seventh grade boys, considerable growth unrelated to diet will occur naturally if the study lasts 6 months.
3. Testing. When you increase the subjects' sensitivity to the testing instrument due to repeated applications. For example, subjects fill out a pretest that measures their knowledge about the connection between nutrition and heart disease, watch a TV
show that teaches them about basic nutrition concepts, and answer the same questions again. Their answers the second time will be affected not just by the film but by the fact that they were sensitized to the issues in the test before they even saw the film.

4. **Instrumentation.** Using different measures (or different forms of the same general measure) to measure the dependent variable(s) at different times. Using the above example, assume that the questions included before and after the film are different. The findings could be due to the difference in the measures rather than a true change in knowledge.

5. **Statistical regression or "regression to the mean."** When subjects who have extreme initial scores on a measure are used there is a strong tendency for subsequent scores to regress to the mean score, regardless of any real change in knowledge or attitude. This can be seen as due to measurement error (the initial score is extreme and so more likely to be inaccurate) as well as a case of floor or ceiling effects (the only way to go from a low score is up; the only way to go from a high score is down).

6. **Selection biases.** When groups are not comparable due to selection of nonrepresentative samples or because of lack of random assignment to conditions. For example, if there are significantly more males in a treatment group that views an R-rated "slasher" film than in a control group that watches a nature film, results may be due to a relationship between gender and violence rather than any real effects of violent portrayals.

7. **Experimental mortality.** When subjects in your sample are lost for various reasons during the course of the experiment and the subjects remaining are substantively different than those who left. For example, in a study of a new math curriculum for high school students, the 30% of the students who drop out before the end of the study all had low initial scores in math. Final results based only on those students who remain would be based on a nonrepresentative sample that favors students who started out with moderate to good math skills.

8. **Causal time-order.** When causality can't be demonstrated because the time order of cause and effect cannot be determined, or is seen to actually be "effect-cause" rather than "cause-effect." For example, in a study of the effectiveness of three mediation strategies on disputant behavior, it may be that the mediator wasn't causing a change in the disputants' behaviors but the behaviors caused a change in the mediator.

9. **Diffusion or imitation of treatments.** When members of the treatment group talk with members of the control group, influencing the experimental responses of both groups. For example, subjects in the treatment group may learn about the purpose of the experiment by comparing their experiences with those of control subjects and then act (consciously or unconsciously) on that information.
Diffusion of treatments may be a problem in a within subject design where the same subjects are exposed to different treatments or both treatment and control. For example, in a study of the effects of violent and sexual program content, subjects who watch the violence first might respond differently to the sexual content than subjects who see the sexual content first. A variety of techniques are available to "balance" and "counterbalance" the effects of the order of treatments for different subjects to maintain internal validity.

10. **Compensation.** When offering some form of compensation (money, extra credit, etc.) to subjects changes their behavior. That is, they're getting paid somehow so they try harder and thus their scores are due to this rather than to the effects of the treatment.

Also: When the administrator of the experiment "feels sorry for" members of the control group who do not receive the treatment and tries to compensate for this by behaving toward them in a more friendly manner, etc.

11. **Compensatory rivalry.** When subjects in a control group know they are not receiving a treatment and work harder to try and overcome this missing stimulus. For example, a group of supervisors in an organization serve as a control group and do not receive training to increase their leadership effectiveness. As a result, they work harder on their own to increase their leadership effectiveness scores.

12. **Demoralization.** The reverse of compensatory rivalry, where members of the control group realize they are not being given a valued treatment and become less motivated or try less hard than they would have if they hadn't been aware of their status as members of the control group. This unnaturally increases the difference between control and treatment group scores.

G. Identify potential threats to external validity. The second type of validity to consider is external validity, the extent to which the results of an experiment are generalizable to other groups and other contexts. Usually when people talk about the results of an experiment being generalizable to other groups outside of a sample they refer to this as external validity. When they talk about the results of an experiment being generalizable to other contexts (say you did your study in a university and you want to be able to generalize to military situations) they refer to this as ecological validity, a subset of external validity. Some of the threats to external validity identified by Cook and Campbell (1979) are described here.

1. **Interaction between testing and treatment.** When the manipulation may only produce an effect after the subjects have been sensitized to the nature of the study through the testing instrument. For example, in a test of the effects of a video about the heart, subjects first try to identify each part of the heart on a diagram. This sensitizes them and leads them to pay attention to the parts of the video that identify the parts of the heart, so the final results depend not just on the video itself, but on the sensitization provided by the pretest.

2. **Interaction between selection biases and treatment.** When selection biases are closely related to the nature of the treatment condition. For example, in a study of the effects of a video about AIDS, several members of the treatment group are also members of a Gay and Lesbian rights organization. This selection bias then interacts with the film to influence the conclusions of the study.
3. **Reactive effects of experimental arrangements.** When the conditions under which the experiment or treatment is administered influence the results and prevent them from being generalized to other contexts. For example, subjects watch a TV program in a small laboratory room containing several pieces of electronic equipment and answer questions about their emotional responses to what they see. The same responses may or may not occur at home in the subjects' living room. (In rare cases the experimental arrangements correspond exactly to the nature of the experimental treatment, such as in the Milgram studies on obedience.)

4. **Interaction of history and treatment.** When an experiment takes place on a day or in a time during which something significant that is related to the study occurs, making it difficult to generalize to other past or future situations. For example, a study on the effects of a series of anti-drug public service announcements conducted during the "Just Say No" campaign or during Woodstock or immediately after John Belushi died all would make results more difficult to generalize.

V. Based on the key considerations above, select an appropriate experimental design (these are from Campbell & Stanley, 1963) (see Babbie, p. 245).

A. 3 pre-experimental designs (These are not recommended, although they are used):
   1. **One-shot case study:** no pretest, no control group. No pretest means a true demonstration of causality is not possible, only correlation of treatment and the dependent variable. No control group means there will be only an intuitive standard for values of the dependent variable.
   2. **One-group pretest-posttest:** pretest, but no control. Doesn't eliminate the possibility that influences other than the treatment are responsible for results.
   3. **Static-group comparison:** control group, but no pretest. Allows an absolute standard of comparison of the dependent variable across groups, but still just correlation of treatment and dependent variable, not causation. (The lack of random assignment of subjects to groups here differentiates this from the posttest-only control group design below.)

B. 3 true experimental designs (These are the most preferred experimental designs):
   1. **Pretest-posttest control group design** (Babbie calls this the "classical" experimental design): There is a control group and an experimental group. Each is given both a pretest and a posttest, but only the experimental group is given the treatment.
   2. **Posttest-only control group design:** There is a control group and an experimental group. One group is given a treatment and the other is not. Then both groups are posttested.
   3. **Solomon four-group design:** A combination of a. and b. Group 1 receives a pretest, treatment, and posttest and group 2 receives a pretest, no treatment, and a posttest (this is the Pretest-posttest control group design). Group 3 receives a treatment and a posttest and group 4 receives no treatment and a posttest (this is the Posttest-only control group design). (See Babbie, p. 251)

C. Quasi-experimental designs (used in social sciences where only partial control is feasible).
1. **Time series experiments**: a series of measurements are taken and a treatment is administered during the series. The pattern of responses before and after the treatment are then compared. For example, blood pressure of patients is measured regularly over nine months, and in the fifth month the patients are given a large one-time dose of a new blood pressure medicine.

2. **Equivalent time samples**: a version of time series experiments, but the treatment is administered more than once. For example, in the blood pressure study above, the medicine would be given in alternating months and the pattern of blood pressure measurements would be evaluated.

3. **Nonequivalent control group designs**: a pretest-posttest control group design (see above) in the case where the experimental and control groups are not equivalent. This design is often used in educational research where, for example, the groups are different classes.

D. Complex factorial designs. In a complex factorial design there are two or more independent variables and there is interest in the effects of combinations of these variables on the dependent variable(s). Each independent variable is referred to as a factor. If there are two IVs the design is referred to as a two-way factorial design. Further, the design is usually referred to as the number of levels of one IV by the number of levels of the other IV (for example, a "2 X 3 factorial design" means there are two IVs, one with two levels, the other with 3). The combination of the levels of each factor result in what are called cells (for example, if the factors are subject (math and English) and gender (male and female) the cells are math/male, math/female, English/male, and English/female).

1. **Between subject factorial designs**: In some designs all of the factors are between subject (sometimes called independent group designs, because each cell contains an independent group of subjects).

2. **Within subject factorial designs**: In other designs all of the factors are within subject (sometimes called repeated measures designs because responses of the same group of subjects are measured repeatedly). Here, the same group of subjects is exposed to all levels of all factors.

3. **Mixed designs** are designs in which different subjects are randomly assigned to the levels of one IV but are exposed to all levels of another IV.

4. **Nested designs** are designs in which a level of one IV occurs only within a level of another IV (one IV is nested in the other). For example, a researcher is interested in the potential impact of teacher gender and subject of class (math versus English) on student satisfaction; the researcher also believes that characteristics of a specific class may influence student satisfaction as well. In a standard fully-crossed factorial design (one in which the independent variables are fully "crossed") the same class would be taught both math and English and the teacher would be both male and female:

<table>
<thead>
<tr>
<th>Teacher gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Since the former is rare and the latter impossible, different classes would have to be used in each of the four conditions in the experiment, so that class would be nested within subject and within teacher gender.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Class 1</th>
<th>Class 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>Class 1</td>
<td>Class 1</td>
</tr>
</tbody>
</table>

Teacher gender

<table>
<thead>
<tr>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>Class 1</td>
</tr>
<tr>
<td>English</td>
<td>Class 2</td>
</tr>
</tbody>
</table>

VI. Refine operationalizations of independent and dependent variables.

VII. Determine sampling procedures. It is less of a problem to have a sample that does not represent an overall population (for example, a sample of only freshmen college students) in experiments than in surveys, because patterns of cause and effect are more generalizable and stable than specific characteristics of a population. However, it is still important. Further, it is vital that experimental and control groups be equivalent in every way possible in order to make valid comparisons between the groups. How can the researcher accomplish all of this?

A. Probability sampling. For each group of subjects take one sample of the population, insuring that each person has an equal chance of being included. This can be difficult, since it may require several samples, each of which is large enough to represent the overall population.

B. Randomization/random assignment. Randomly assign each person in one subject pool to one of the groups. For example, say you are testing the effect of teaching strategy (lecture versus role-playing) on students' learning. You would want to randomly select one sample of students, say 50. Then you would want to randomly assign these 50 subjects so that 25 ended up in the lecture condition and 25 ended up in the role-playing condition. The members of each group should be similar to each other even if all together they don't represent the whole population accurately. Of course the number in the initial group is key -- the worst case would be where the initial group is just two
people, where 1 is the control and 1 is the treatment, and this is clearly inadequate. The very minimum number of people in each group is usually 20.

C. Matching. Select variables that might be related to the dependent variable(s) and divide subjects into these subgroups (for example, if gender is important create subgroups of males and females). Then randomly assign members of each of these subgroups to the treatment and control groups. This is similar to stratified sampling (if it is not done randomly it's more like quota sampling).

1. Another way of matching. Matching can also be done by oversampling to equalize the number of subjects with specific characteristics who end up in treatment and control groups, or by giving a pretest and eliminating those subjects who can't be matched off into the two groups.

2. Problems with matching. It is hard to know which are the important variables to match on and most statistics assume randomization.

3. Matching and equal variance. The goal of matching is not just to get the average member of each group to be similar; it is also to create groups with similar variance on the selected variables (despite Babbie on p. 243).

VIII. Write out procedures to administer experiment, including specifically what each person who administers the experiment will do and say, from before the subject arrives to after he or she leaves. Special emphasis should be given to what will be said during the debriefing period.

IX. Conduct a pretest to find out the following.

A. Do the written procedures work smoothly? What revisions need to be made?

B. How do the subjects feel about the experience? Did they understand the instructions? Did they enjoy it? Did it take too long? Did they guess the hypothesis? Did they believe the deception? Did they take it seriously?

C. Look at the data gathered from the pretest subjects. Do they seem to correspond to the hypothesis?

X. Administer (run) the experiment, analyze the data, and write it up for submission for presentation at a conference, then for publication.
DATA ANALYSIS: OVERVIEW

I. Overview of discussion
   A. A lot of specific information in this part of the course.
   B. Goal is to provide at least a flavor of both the conceptual and the practical sides of data
      analysis (see also MMC 503 and statistics courses offered in other departments)
      1. The conceptual side involves the reasoning behind specific techniques. We'll
         illustrate these by doing "by hand" calculations.
      2. The practical issues involve data preparation and using the computer to calculate
         results.

II. Statistics, math, and computer phobia
   A. It sounds scary, but it isn't if, as with most things, you take it one step at a time
   B. Statistics to use comes logically from what data you have gathered.
   C. There isn't really much math, and none beyond addition, subtraction, division,
      multiplication, squaring and square roots.
   D. Computers can be a pain, but when things work, there is a great satisfaction, and when
      they don't you can be frustrated like the rest of us.
   E. You are encouraged to help each other, and tell me when ANY part of the discussion is
      unclear. It takes many repetitions in different ways for different people, for statistics and
      data analysis to become clear and doable.

III. Organizing concepts and important terms
   A. Descriptive (describe the data, often from whole population) versus inferential (typically
      sampling, probability, explanation) statistics
   C. Univariate, bivariate, multivariate statistics
   D. Parametric and nonparametric statistics
   E. Levels of measurement (see notes titled "Levels of measurement")
   F. Use of symbols (see notes titled "Guide to common statistical symbols")
   H. Other
      1. Variables and values
      2. Null and research hypotheses (see notes titled "The experiment")
      3. Scales and indexes/indices
      4. Outliers
      5. Interactions between variables (see notes titled "Two-way analysis of variance
         (ANOVA)")
COMPUTERS AND COMPUTING

I. How computers work
   A. Binary code and IBM punch cards
   B. Higher level programming languages
      1. C
      2. Pascal
      3. Fortran
      4. Basic and Visual Basic

II. Elements of a computer (hardware)
   A. Central Processing Unit (CPU)
   B. Input devices
      1. Keyboard
      2. Mouse
      3. Joystick
      4. Microphone
   B. Storage devices: memory
      1. Active or working memory
      2. Permanent memory
         a. Hard drives
         b. Floppy disks
         c. CD-ROM
   C. Output devices
      1. Monitor (screen)
      2. Printer
      3. Speakers

III. Size, capacity, power, speed, and compatibility considerations
   A. Mainframe computers
   B. Work stations
   C. Personal computers (PCs)
   D. Laptop computers
   E. MAC versus IBM

IV. Computer software
   A. Systems software (Operating systems)
      1. For mainframes and workstations
         a. CM/ESA (e.g., the Temple IBM mainframe)
         b. Unix (e.g., the Temple DEC mainframe -- Astro)
      2. For IBM and IBM compatible computers
         a. DOS (Disk Operating System)
         b. Windows
         c. OS/2
      3. For Apple/MacIntosh computers
a. System

B. Applications software
1. Word processing
   a. WordPerfect
   b. Microsoft Word
2. Database management
   a. Accounting spreadsheets
      1. Lotus 1-2-3
      2. Quattro Pro
   b. Text-based (flexible format) databases
      1. FileMaker Pro
      2. Specialized bibliographic databases (e.g., Notebook, ProCite, Endnote)
3. Communications (access to each other and the internet)
   a. Kermit
   b. ProComm
   c. Telnet
   d. Hyperterminal
   e. QWS3270
4. Statistics/data analysis
   a. SPSS -- Statistical Package for the Social Sciences
   b. SAS -- SAS, Inc.
   c. BMDP -- Biomedical Data Processing
   d. Minitab
   e. Others: Systat, Jmp!
5. Desktop publishing, graphics
   a. Pagemaker
   b. Photoshop
6. Authoring programs (for presentations)
   a. Toolkit

C. Utilities software
1. Backup and restore
2. Data recovery and memory management
   a. Norton Utilities
3. Virus protection
   a. Symantic Norton
   b. McAfee
4. Data compression
   a. Stacker
PREPARING THE DATA

I. Preparing qualitative data for analysis

II. Preparing quantitative data for analysis
   A. Codebooks (see example in Babbie)
      1. Identifying values and variables within the data
      2. Cases, records, lines, and columns
      3. How to use vertical and horizontal space
   B. Getting the data into the computer
      1. Direct data entry
         a. Advantages of working in pairs
      2. Optical scanners
      3. Using a word processing program and uploading

III. Data cleaning and checking
   A. Missing data -- problems with using empty spaces
   B. Spot checking for data accuracy and error rates
   C. Using spatial patterns in printed data to check for errors
   D. Using possible values for each variable to check for errors
   E. Using possible values based on contingency questions to check for errors
DISPLAYING THE DATA

I. Display techniques
   A. Histograms (vertical and horizontal)
   B. Stem and leaf plots
   C. Bar graphs
   D. Plots on X and Y axes
   E. Pie chart graphs
   F. Other methods: Ear cues, 3D, etc.

II. Display issues
   A. Detail versus manageability
   B. Collapsing response categories
   C. What to do with "Don't know"s
LEVELS OF MEASUREMENT

There are four basic levels of measurement: Nominal, ordinal, interval, and ratio. These levels are often referred to as "strong" versus "weak" (nominal being the weakest, ratio the strongest) because of the utility of the data obtained from them. You can do a lot more statistically with ratio level data than with nominal level data. However, this does not mean that there is no benefit to using the "weaker" levels.

I. **Nominal**: At this level of measurement "scales" contain only names of categories into which someone or something being studied falls. The only information it provides is names for the observations, therefore the label "nominal." Some nominal scales are dichotomous, with only two values (for example, male versus female, yes versus no, etc.). Others have several categories (for example, colors). In general, categories for a nominal scale should be exhaustive and mutually exclusive. Exhaustive means that every observation (object, event, person, etc.) can be placed in a category. Mutually exclusive means that each observation fits into one and only one of the categories.

II. **Ordinal**: An ordinal level of measurement is a rank ordering of observations in terms of how much or how little of a characteristic they possess. Instead of just having names, the categories also have an order or sequence. For example, a group of individuals can be ranked in terms of height. The tallest person is placed first on the scale, the second tallest person is placed second, and so on. Unfortunately, this does not provide information about how much taller the first person is than the second, how much taller the second person is than the third, etc.

III. **Interval**: At the interval level of measurement it is assumed that there are distinct and equal differences between observations. Therefore, not only can the observations or categories be ordered, we can assess how much more of something one observation has than another. The system of whole numbers represents an interval level scale. We know that the difference from 1 to 2 is equal to the difference from 3 to 4. An important feature of the interval level of measurement is that the point at zero is only arbitrary; it has no counterpart in the real world. When respondents rate their need for excitement on a standard 5 point scale, with 5 being "a lot" and 1 being "none," zero is meaningless. Because of this, interval level data allows us to compare observations using addition and subtraction (e.g., "That one was 3 levels higher than the other one") but not using multiplication and division (e.g., "That one was three times as high as the other one").

IV. **Ratio**: The difference between ratio and interval level data is that ratio data has a meaningful zero point. Age, weight, height, speed, length, and time are examples of ratio level scales of measurement because the zero point for each has a real world counterpart that represents the absence of something. Data at this level can therefore be compared using multiplication and division (i.e., it permits the calculation of ratios, such as "a twenty year old person has lived twice as long as a ten year old").

V. A summary of measurement scales, their characteristics and their statistical implications follows.
<table>
<thead>
<tr>
<th>Characteristics of the scale</th>
<th>Statistical possibilities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nominal measurement scale</strong></td>
<td>A scale which &quot;measures&quot; in terms of names or designations of discrete units or categories. Can be used for determining the mode, the percentage values, or the chi-square.</td>
</tr>
<tr>
<td><strong>Ordinal measurement scale</strong></td>
<td>A scale which &quot;measures&quot; in terms of such values as &quot;more&quot; or &quot;less,&quot; &quot;larger&quot; or &quot;smaller,&quot; but without specifying the size of the intervals. Can be used for determining the mode, percentage, chi-square, median, percentile rank, or rank correlation.</td>
</tr>
<tr>
<td><strong>Interval measurement scale</strong></td>
<td>A scale which measures in terms of equal intervals or degrees of difference, but whose zero point, or point of beginning is arbitrarily established. Can be used for determining the mode, the mean, the standard deviation, the t-test, the F-test, and the product moment correlation.</td>
</tr>
<tr>
<td><strong>Ratio measurement scale</strong></td>
<td>A scale which measures in terms of equal intervals and an absolute zero point of origin. Can be used for determining the geometric mean, the harmonic mean, the percent variation and all other statistical determinants.</td>
</tr>
</tbody>
</table>
MEASURES OF CENTRAL TENDENCY

As the name indicates, the measures of central tendency tell you where the center of your data is. Most statistics are based on the assumption that when objects, events, people, etc. are placed in order based on one (or more) characteristic, the resulting graph will be bell-shaped (see below). This is called the "normal distribution" or the "Gaussian curve of normal frequency." (Most parametric statistics use this bell-shaped curve as the "parameter." A parameter is a function, characteristic, or quality of a population which is constant.) Measures of central tendency attempt to describe the center of the bell-shaped curve (the peak of the bell).

Suppose there were nine students in this class. If each of the nine students reported how many people were in her family, the resulting data, placed in order, might be: 1, 2, 2, 3, 3, 3, 4, 4, 5. If we graphed this data it would look like this:

```
Number of family members
0  1  2  3  4  5  6
Number of students
0  X
1  X
2  X
3  X
4  X
```

A measure of central tendency would identify the center or peak of this curve. Note that usually the mean is the preferred measure of central tendency; however, in significantly skewed distributions, use the median since it provides a more accurate reflection of the center of the data.

I. The most common measures of central tendency are:
   A. Mode. This is the most popular value or score; the score that is reported the most times.
      1. CALCULATIONS: It is easiest to identify the mode if the values are first placed in rank order. Then simply count the number of times each score appears and identify the score that appears the most frequently.

      2. EXAMPLE: What is the mode of these data: 4, 25, 51, 59, 62, 63, 63, 66, 66, 68, 70, 71, 71, 71, 73, 75, 76, 77, 77, 80, 88, 93? Since 71 appears three times and none of the other values appears this many times, 71 is the mode.
3. **EXERCISE:** What is the mode of these data: 
   0, 0, 1, 2, 3, 3, 4, 4, 4, 5, 6, 6?

4. **Additional information: types of distributions.**
   a. Bimodal data has two most popular scores (e.g., ~/\_\_/~\)
   b. Trimodal data has three most popular scores
   c. Amodal data has no one most popular score (e.g., |~~~~~~~|)
   d. Unimodal data has one most popular score. A bell-shaped curve is always unimodal. BUT, having a unimodal distribution DOES NOT mean that you have a bell-shaped curve.

B. **Median.** The median is that score in a ranked distribution of scores (i.e., arranging the scores from lowest to highest or vice versa) at which 50% of the scores are higher and 50% of the scores are lower.
   1. The median can be represented by a score that does not actually appear in the data. For example, the median of the scores 1, 2, 3, and 4 is 2.5 (half the scores are above and half are below this point) even though 2.5 is not one of the scores.
   2. The median is not sensitive to extreme values in the data. For example the median of (1, 2, 3, 4, 5, 6, 7) is 4; the median of (-1000, -999, -998, 4, 998, 999, 1000) is also 4.

3. **CALCULATIONS:** To calculate the median of a set of data, first put the data points in order. If the number of datum is odd find this number:

   \[
   \frac{N + 1}{2}
   \]

   where \(N\) is the number of data points. Starting at one end of the ranked scores, count this number of scores. The score you end at is the median.

   In a data set where the number of datum is even find

   \[
   \frac{N}{2}
   \]

   again where \(N\) is the number of data points. Starting at one end of the ranked scores, count this number of scores.

   Take the score you end at and the next score, add them together, and divide by 2. The result is the median.

4. **EXAMPLE:** The scores are 1, 2, 3, 5, 8, 9, and 10. First note that there are 7 scores.

   \[
   \frac{N + 1}{2} = \frac{7 + 1}{2} = \frac{8}{2} = 4
   \]
The fourth score from either end of the ranked scores is 5, the median.

5. EXERCISE: The data are 2, 3, 3, 3, 4, 5, 6, 7, 8. What is the median?

C. **Mean.** This is often called the average and is calculated by finding the sum of all of the scores and dividing that sum by the number of scores.

1. This is the ONLY measure of central tendency that is responsive to all of your scores, including those with extreme values. Therefore, it is typically the most useful of the three methods. When in doubt, rely on the mean rather than the median or the mode.

2. **CALCULATIONS:** The formula for the mean is simple:

\[
\bar{x} = \frac{\text{sum of scores}}{\text{number of scores}}
\]

3. **EXAMPLE:** The mean of 3, 4, 5, 6, and 7 is:

\[
\bar{x} = \frac{3 + 4 + 5 + 6 + 7}{5} = \frac{25}{5} = 5
\]

4. **EXERCISE:** What is the mean of this data set: 5, 6, 17, 2, 3, 18, 19?
MEASURES OF DISPERSION

The measures of central tendency (mode, median, and mean) only provide the researcher with limited information about the nature of the data. They tell her where the center of the data is. To understand the distribution, spread, or dispersion of the data the researcher uses measures of dispersion.

I. **Range.** This is the most basic measure of dispersion. It is the highest value minus the lowest value, or the range of values.
   A. **CALCULATIONS:** To be mathematically correct it is necessary to add one to the difference between the highest and lowest values. So the formula for computing range is:
      \[ \text{Range} = (\text{high value} - \text{low value}) + 1 \]
   B. **EXAMPLE:** What is the range of the following data set: 3, 2, 4, 6, 1, 1, 8, 9?
      \[
      \text{Range} = (\text{high value} - \text{low value}) + 1 \\
      (9 - 1) + 1 = (8) + 1 = 9
      \]
   C. **EXERCISE:** This is the data set: 1, 2, 3, 4, 5, 6, 7, 11, 13, 24, 0. What is the range?

II. **Kurtosis.** This measure, also sometimes called "skewness," describes the shape of the distribution curve that summarizes the values or scores. Here are some examples of different types of kurtosis (see also Williams, p. 36):

<table>
<thead>
<tr>
<th>Mesokurtic</th>
<th>Platykurtic</th>
<th>Leptokurtic</th>
</tr>
</thead>
<tbody>
<tr>
<td>(&quot;normal&quot;)</td>
<td>(&quot;flat&quot;)</td>
<td>(&quot;tall and skinny&quot;)</td>
</tr>
</tbody>
</table>

Leptokurtic data is called skewed (as in uneven). Skewed data is referred to as either positively or negatively skewed depending upon the direction of the tail of the skew.
III. **Standard deviation.** This measure provides a number to describe how much difference there is between a typical value or score in a set of data and the mean of all the values or scores.

A. **CALCULATIONS:** The standard deviation is calculated by finding the mean, finding the deviation of each score from that mean, squaring each of these deviations, and then finding the mean value of these squared deviations (by summing them and dividing by the number of squared deviations). The squaring is necessary to give equal "voice" in the formula to positive and negative deviations from the mean (e.g., if the mean is 3, \((7 - 3) = +4\) and \((-1 - 3) = -4\) should be considered equal and squaring the deviations makes both +16). The formula for calculating a standard deviation is therefore:

\[
S = \sqrt{\frac{\sum (x - \bar{x})^2}{N - 1}}
\]

where:
- \(S\) = Standard deviation
- \(\sum\) = "The sum of each" of what follows
- \(x\) = Each value or score
- \(\bar{x}\) = The mean of all the values or scores
- \(N\) = The total number of values or scores

B. **EXAMPLE:** This is the data set: 72, 81, 86, 69, 57. What is the standard deviation?

1. First find the mean:

\[
\bar{x} = \frac{\sum x}{n} = \frac{365}{5} = 73
\]

2. Then calculate how much each score deviates from the mean; i.e.,

\((x - \bar{x})\).

\[
\begin{align*}
(x - \bar{x}) &= (72 - 73) = -1 \\
(x - \bar{x}) &= (81 - 73) = 8 \\
(x - \bar{x}) &= (86 - 73) = 13 \\
(x - \bar{x}) &= (69 - 73) = -4 \\
(x - \bar{x}) &= (57 - 73) = -16
\end{align*}
\]
3. Then square the deviations; i.e.,

\[(x - \bar{x})^2.\]

\[(x - \bar{x})^2 = ( -1 * -1) = 1\]
\[(x - \bar{x})^2 = ( 8 * 8) = 64\]
\[(x - \bar{x})^2 = ( 13 * 13) = 169\]
\[(x - \bar{x})^2 = ( -4 * -4) = 16\]
\[(x - \bar{x})^2 = ( -16 * -16) = 256\]

4. Then sum the squared deviations. This is known as a "sum of squares" or SS.

\[\sum ((x - \bar{x})^2) = (1 + 64 + 169 + 16 + 256) = 506\]

5. Then divide the sum of squared deviations by the number of values (N) minus 1, or \((N - 1)\).

\[\frac{\sum ((x - \bar{x})^2)}{N - 1} = \frac{506}{(5 - 1)} = \frac{126.5}{4}\]

6. Finally take the square root of this figure.

\[S = \sqrt{\frac{\sum ((x - \bar{x})^2)}{N - 1}} = \sqrt{126.5} = 11.24\]

7. What does this result mean? It means that the standard deviation of any score in the data set from the mean is 11.24 units.

C. EXERCISE: This is the data set: 3, 6, 7, 12, 13, 16. What is the standard deviation?

\[(x) \quad (x - \bar{x}) \quad (x - \bar{x})^2\]

\begin{tabular}{lrr}
3 & & \\
6 & & \\
7 & & \\
12 & & \\
13 & & \\
16 & & \\
\end{tabular}
\[ \sum (x - \bar{x}) = \quad \sum (x - \bar{x})^2 = \quad \]

\[ \bar{x} = \quad \]
\[ N - 1 = \quad \]
Standard deviation \( (S) = \quad \)

IV. **Variance**. This is a measure of dispersion that is directly related to standard deviation. Variance \( (S^2) \) is calculated exactly the same way as the standard deviation \( (S) \) except that the square root is not taken. Therefore, standard deviation is the square root of variance and standard deviation squared is variance.

A. **EXAMPLE**: The variance of the data in the standard deviation example above is 126.5 (see step 6).

B. **EXERCISE**: What is the variance of the data in the standard deviation exercise above?

Variance \( (S^2) = \quad \)
THE LOGIC OF INFERENTIAL STATISTICS

Inferential statistics let the researcher do much more than merely describe a set of data. They allow her to use the small amount of data that has been gathered in a study (data from a sample) to make inferences about all of the data that exist everywhere (data about an entire population). Such inferences are possible because of the laws of probability as set out in probability theory. Because these laws can only be applied when a sample represents well the population from which it comes, the explanation of the logic behind inferential statistics begins with an examination of how to select such a sample.

I. Probability sampling theory and the sampling distribution. Using some exemplary figures from Babbie (1995), we can see how probability theory explains the formation and utility of a sampling distribution, a distribution of some characteristic of a sample (such as its mean) from many separate samples.

Assume we have a group of ten people, each possessing a certain amount of money ranging from $0 to $9, so that the first person has $0, the second person has $1, etc. Assume that we wanted to know the average (mean) amount of money held by people in our population, but that we could not simply ask all of them how much money they had. Instead we could only ask one person (that is, we could take a sample of \( n = 1 \) from our population). In this situation there are just ten possible results, as shown in figure 8-4 in Babbie (p. 198). As we can see, this method of sampling doesn't provide us with very good information in terms of estimating the true mean of the population. But suppose that we could take a sample that contained two of the people in the population (\( n = 2 \)). In a population of 10 that would mean that there are 45 different possible samples (persons 1 and 2, 1 and 3,..., 2 and 4, 2 and 5,... and so on). If we calculate the mean for each of these samples of 2 (number of dollars of first person plus number of dollars of second person, all divided by two) and then plot those we would get the distribution in figure 8-5 in Babbie (p. 198). Note that there are more samples that are close to the true mean of the ten people than when \( n = 1 \). This method would give us more information from which to estimate the true mean of the population. However, it is still not as accurate as it could be.

As we review the sampling distributions (the distributions of all of the possible samples of that size) obtained when the number of members of the population in each sample (\( n \)) is 3, 4, 5, and 6 (figure 8-6 in Babbie, p. 199) we can easily see how taking samples with larger numbers of subjects, which also means taking more samples, moves us closer and closer toward a sampling distribution that gives us better chances of accurately estimating the true mean of the population. This example is, of course, unrealistic; the population we're interested in is never this small (or we'd just "sample" the whole population!) and we rarely can expect to select even one sample that represents such a large proportion of the population.

But this example demonstrates one of the fundamental "truths" about sampling distributions:

IF MANY SEPARATE INDEPENDENT RANDOM SAMPLES OF SUFFICIENT SIZE ARE SELECTED FROM A POPULATION, THE SAMPLE STATISTICS PROVIDED BY
THOSE SAMPLES WILL BE DISTRIBUTED AROUND THE POPULATION PARAMETER (A CHARACTERISTIC, HERE THE MEAN) IN A KNOWN WAY.

Under more realistic circumstances than the above example the sampling distribution has a "bell" shape and is called the "normal" distribution. Babbie presents such a sampling distribution in figure 8-9 (p. 201). This distribution is important because we know how many of all of the possible sample estimates fall under different parts of the normal curve. Using an increment on the horizontal axis of the distribution called a standard error, we know that 68 percent of the sample statistics fall within +/- 1 standard error, 95 percent of the sample statistics fall within +/- 2 standard errors, and 99.9 percent of the sample statistics fall within +/- 3 standard errors from the true population value at the curve's center. This relationship is constant for any random sampling procedure provided that a large number of samples are selected and that the samples are large enough.

The key to using this information is to determine how many of the units we're interested in (in the example above it was "dollars" but it might be "hours of TV viewing" "attitudes about the president" or anything else) constitute one of these increments called a standard error. We'll see exactly how to use this information below, but note that we would like the number of "our" units that equal a standard error to be small: A big number (e.g., 10 hours of TV viewing per week) means that the population value is within a larger range of the sampling distribution (+/- 10 hours, 20 hours, 30 hours) while a small standard error (e.g., 2 hours of TV viewing per week) means that the population value is within a smaller range of the sampling distribution (+/- 2 hours, 4 hours, 6 hours) and we are more likely to be able to make an accurate estimate of the population value.

II. Calculating a standard error. The standard error is the standard deviation of the sampling distribution, or the average difference of each sample statistic from the population parameter. Standard errors can be calculated for any type of sampling distribution (i.e., means, medians, percentages, etc.) but most statistics rely on the standard error of the mean or the standard error of proportions (e.g., in survey research where people are more concerned with the percentage or proportion of people who respond to an item in one of two ways such as "yes" or "no").

The standard error of the mean can be expressed as follows:

\[
\frac{\sigma}{\sqrt{n}}
\]

where:

\(\sigma\) = Standard deviation of values in the population

\(\bar{x}\) = Standard error of the mean

\(n\) = Number of observations (cases, subjects) in each sample
The standard error of proportions can be expressed as follows:

\[ \sigma_p = \sqrt{\frac{pq}{n}} \]

where:
- \( \sigma_p \) = Standard error of proportions
- \( p \) = True proportion of people, cases or subjects in population who respond in a given way
- \( q \) = True proportion of people, cases or subjects in population who respond in the only other way
- \( n \) = Number of observations (cases, subjects) in each sample

Please note that you do not have to know how to calculate these population values; in fact only in our artificially small example could these be calculated.

But note that the numerators (top parts) of both formulas serve as a measure of the homogeneity/heterogeneity of the population, while the denominators (bottom parts) of both formulas serve as a measure of the sample size. Why is this important? BECAUSE THE STANDARD ERROR IS ALWAYS A FUNCTION OF TWO CRITICAL FACTORS: 1) HOMOGENEITY/HETEROGENEITY AND 2) SAMPLE SIZE. As heterogeneity in the population increases, the size of the standard error will always increase. As the size of the sample increases, the size of the standard error will always decrease. Another way to say this is that standard error is a direct function of heterogeneity and an inverse function of sample size. So in general we want to have larger samples from homogeneous populations so that our standard error will be small.

III. How to use the sampling distribution when we only have one sample. The reason it is so important to know the standard error of a sampling distribution is that in practice researchers can't and don't take all of the possible samples; they just take one. What they need to know is whether the statistic they get from the one sample they take is from a sample at the very end of one of the tails of the sampling distribution or somewhere near the center. Because we can assume that the distribution is "normal," we can assume that 68 percent of the sample statistics will fall within +/- 1 standard error, 95 percent of the sample statistics will fall within +/- 2 standard errors, and 99.9 percent of the sample statistics will fall within +/- 3 standard errors from the population parameter.

So we can say that we are 68 percent confident that our single random sample falls within +/- 1 standard error from the population parameter, 95 percent confident that it falls within +/- 2 standard errors, and 99.9 percent confident that it falls within +/- 3 standard errors. This is known as our confidence level.
Further, if we know the range of scores on our measure that corresponds to one standard error, we can state that our confidence interval is the range of scores that should fall between +/- 1 standard error, or +/- 2 standard errors, or +/- 3 standard errors (e.g., "I am 95% confident that people in the population from which I sampled watch TV between 32 and 40 hours each week because my sample reports watching TV 36 hours each week and my standard error is 2 hours.")

Note that the researcher can decide in advance on the standard error and therefore the range of values in which she can be confident the true population value falls; a smaller range is clearly preferable. As demonstrated by the formulas above, if she wants a smaller standard error she can increase her sample size or select a more homogeneous population (usually done through stratification [discussed in the notes titled "Sampling"], or through redefining the population).

As noted, in reality we rarely know the population parameter and we rarely take more than one sample from the population. Since the formulas above for calculating standard error require information about the population we have a problem. The solution is to use the information from our one sample to estimate the information we need but do not have about the whole population. We substitute the standard deviation of our sample as an estimate of the standard deviation in the whole population or we substitute the actual proportions of responses in our sample to estimate the population proportions.

IV. Hypothesis testing. In addition to using probability theory to ask, "What is the population value?" we can use it to test a prediction, or hypothesis, concerning what the population value is, and then test this prediction. The process actually involves two kinds of hypotheses: Research hypotheses and Null hypotheses.

A. An example: Suppose we think people in the Philadelphia area watch TV, on average, more than 40 hours each week. Our research hypothesis is then:

\[ H_r: \mu > 40 \]

where:
\[ \mu = \text{Population mean} \]

To determine the validity of this hypothesis we try to contradict the null hypothesis, which describes the state of affairs in which the research hypothesis is not true (null means "nothing," so think of the null hypothesis as the hypothesis that "nothing interesting" is going on). In this case the null hypothesis is:

\[ H_0: \mu \leq 40 \]
Next we go out and randomly select and administer a survey to 400 people in Philadelphia and ask them how many hours of TV they watched last week. We calculate the mean of the 400 answers and get 42 hours; because we'll use this value to test our hypotheses it is called a test statistic. The standard deviation (\( \sigma \)) of the 400 answers is 10 hours. Does this mean our research hypothesis is correct, or did we just happen to sample an extreme or unusual group of Philadelphians while the whole population actually watches only 40 hours (or less) of TV a week?

It's safer to assume that our sample statistic is just a fluke, that our null hypothesis is valid, that "nothing interesting" is going on. We'd like to reject this null hypothesis, so we assume that it is true and determine the likelihood (probability) that with a true population mean of 40 our one random sample would yield a mean of 42. If this is unlikely enough (admittedly a subjective judgment), we will reject the null hypothesis.

First we need to calculate the standard error of the mean:

\[
\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}
\]

\[
\sigma_{\bar{x}} = \frac{10}{\sqrt{400}} = \frac{10}{20} = 0.5
\]

Now we need to know how many standard errors the sample mean we found (42) is from the population mean under the null hypothesis (40). Since 42 - 40 = 2 and 2 divided by 0.5 = 4 we conclude that if the population mean is actually 40, the sample statistic we obtained is 4 standard errors higher than the population mean. What is the probability that we would have selected such an unusual sample? The table on p. A30 in Babbie (or the similar table in Williams) indicates that the probability of obtaining a test statistic that is 3.09 or more standard errors above the true population value in a normal sampling distribution, is 50% - 49.9% = 0.1% = .001 or 1 in a thousand (and remember that our test statistic is actually 4 standard errors away from the hypothesized population value). This .001 is called a probability value or "p-value" (\( p \)).

We therefore reject the null hypothesis in favor of the research hypothesis and conclude that if all of the people in Philadelphia were asked how many hours of TV they watch per week, the average answer would be greater than 40 hours.

Rather than finding the specific \( p \)-value for a given sample test statistic, researchers typically use arbitrary cut-offs (called alpha levels, denoted by \( \alpha \)) such as .05, .01, and .001. They then can check whether or not their test statistic value would be expected to appear, if the null hypothesis is true, either more or less than 5 out of 100 times, 1 out of 100 times, etc.
B. One and two tailed tests. The example above contained null and research hypotheses that called for a one-tailed test, because the null hypothesis could be rejected only if the sample mean was greater than 40. If the hypotheses had been:

\[ H_r: \mu < 40 \]
\[ H_0: \mu \geq 40 \]

the appropriate test would still have been one-tailed, because the null hypothesis could be rejected only if the sample mean was less than 40. In contrast, a two-tailed test would be appropriate if the hypotheses tested simply whether 40 is the true population mean:

\[ H_r: \mu \text{ does not equal } 40 \]
\[ H_0: \mu \text{ does equal } 40 \]

In this case, the probability of finding an extreme sample mean under the null hypothesis must be divided into two halves, one under each tail of the normal sampling distribution.

C. Type I (alpha) error (\( \alpha \)). Making this type of error means rejecting the null hypothesis when it really should be retained. The p-value refers to the probability of making this type of error.

D. Type II (beta) error (\( \beta \)). The instance of affirming the null hypothesis when it should be rejected. If we wish to be cautious we prefer to increase the probability of Type II error rather than Type I error.

E. Degrees of freedom (df). This appears as an element in the calculation of many inferential statistics. It represents the degree to which different arrangements of elements are possible. For example, we know that the sum of all deviations of scores in a set of data from the mean of the scores must equal 0 (by the definition of a mean the negative ones counter the positive ones). Therefore if we know all of the deviations except one, we can calculate the last deviation -- there is no "freedom" concerning what that value is. This means that before we know any of the deviations there are \( N - 1 \) degrees of freedom.
THE T-TEST

The t-test is an inferential statistic. This means that it allows us to make inferences from our sample to a population, rather than just describing the data in our sample. The purpose of the t-test is to determine whether the means of the data from two separate samples are different enough from each other to make the claim that they represent two different populations.

I. Assumptions and requirements
   A. To use the t-test you must have interval level data
   B. Both samples must represent scores on the same variable. For example, you can use communication apprehension scores for two samples, but you cannot compare communication apprehension scores for one sample with communication competence scores for another sample.
   C. The populations from which the samples were drawn are assumed (we can not know for sure) to be normally distributed. If there is reason to believe this assumption is not warranted do not use this procedure.
   D. The variances of the two populations are assumed (we can not know for sure) to be equal. If there is reason to believe this assumption is not warranted do not use this procedure.

II. Rationale. Three factors determine whether or not the statistical difference between the means of the two samples is significant (at the alpha level you set, i.e. p < .05).
   A. The magnitude of the difference between the two means.
   B. The variability within each sample. We are more likely to believe there actually is a difference between means if all of the scores within each sample are about the same than if the scores vary a great deal.
   C. The number of observations (subjects) in each sample. The larger the number of observations, up to a point, the more confidence we have that the relationship of differences between the means actually holds in the population.

III. Two types of t-test based on type of sample. There are two types of t-test, each with their own formula, based on the fact that there are two types of samples that can be compared.
   A. Samples are independent when the people in the two samples are different; they haven't been matched in any way. For example, a sample of the men in this class and a sample of the women.
   B. Samples are matched when the people in the two samples are matched in some way. For example, one sample may contain the scores from a test given before an experimental manipulation and the other sample may contain the scores from the same test given following the manipulation.
IV. Two types of t-test based on type of difference being tested. A researcher can test for two types of differences between the means of two samples. These two types of t-tests require different use of a table of $t$ statistics and probability values.

A. One type of difference is nondirectional: the researcher wants to know whether the mean of sample 1 is different (either bigger or smaller) than the mean of sample 2.

B. A difference between means may also be directional: the researcher wants to know whether the mean of sample 1 is BIGGER than the mean of sample 2 (OR whether the mean of sample 1 is SMALLER).

V. CALCULATIONS:

A. The formula for the t-test with independent samples is:

$$ t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} $$

where:

- $t$ = $t$ test statistic
- $\bar{x}_1$ = Mean of sample 1
- $\bar{x}_2$ = Mean of sample 2
- $n_1$ = Number of observations in sample 1
- $n_2$ = Number of observations in sample 2

$$ s^2 = \frac{\sum (x - \bar{x}_1)^2 + \sum (x - \bar{x}_2)^2}{n_1 + n_2 - 2} $$

B. $s^2$ is an estimate of the variance among both samples obtained by pooling the variance of the distribution in each of the two samples. It represents the variance of the distribution of differences between the two means if we drew a large number of samples of this size.

C. Note that because the probability distribution used to evaluate it is symmetrical, the direction of the value of the $t$ statistic is not important (i.e., a $t$ statistic of $-2$ is the same as a $t$ statistic of $2$).

D. To determine whether the value of the $t$ statistic we obtain represents a statistically significant difference between the means of the two samples we must consult a table of $t$ statistic values and probability values (found in most statistics books and typically called a t-table).
1. To use the table first calculate the appropriate degrees of freedom. In this case the formula is

$$df = n_1 + n_2 - 2$$

Note that this is the same as the denominator in the formula for $S^2$. This is the case with many calculations of $df$.

2. See the instructions that accompany the t-table to determine the $p$ value associated with the obtained value of the $t$ statistic. The larger the obtained value of the $t$ statistic the more certain the researcher can be that the difference between the means of the two samples is the result of the samples being taken from different populations (rather than merely a fluke resulting from limited samples of one population).

3. The t-table allows the researcher to test nondirectional (called two-tailed) or directional (one-tailed) t-tests. A smaller $t$ statistic value is required for the same level of significance (or confidence) with a one-tailed test than with a two-tailed one. See the instructions accompanying the published t-table.

VI. EXAMPLE: Here are two sets of data gathered from independent samples: 1, 2, 3, 4, 5 and 3, 4, 5, 6, 7. Are the means of these two samples significantly different?

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
</tr>
</tbody>
</table>

$$\sum (x) = 15 \quad \sum (x) = 25$$

$$\bar{x}_1 = \frac{\sum (x)}{n_1} = \frac{15}{5} = 3 \quad \bar{x}_2 = \frac{\sum (x)}{n_2} = \frac{25}{5} = 5$$

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$(x - \bar{x}_1)$</td>
<td>$(x - \bar{x}_1)^2$</td>
<td></td>
<td>$(x - \bar{x}_2)$</td>
<td>$(x - \bar{x}_2)^2$</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1 - 3 = -2</td>
<td>4</td>
<td>3</td>
<td>3 - 5 = -2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2 - 3 = -1</td>
<td>1</td>
<td>4</td>
<td>4 - 5 = -1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3 - 3 = 0</td>
<td>0</td>
<td>5</td>
<td>5 - 5 = 0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4 - 3 = 1</td>
<td>1</td>
<td>6</td>
<td>6 - 5 = 1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5 - 3 = 2</td>
<td>4</td>
<td>7</td>
<td>7 - 5 = 2</td>
</tr>
</tbody>
</table>

$$\sum (x - \bar{x}_1)^2 = 10 \quad \sum (x - \bar{x}_2)^2 = 10$$

$$S^2 = \frac{(\sum (x - \bar{x}_1)^2) + (\sum (x - \bar{x}_2)^2)}{n_1 + n_2 - 2}$$
\[
\begin{align*}
\frac{10}{5 + 5 - 2} + \frac{10}{5 + 5 - 2} &= \frac{20}{8} = 2.5 \\
\end{align*}
\]

\[
\begin{align*}
\bar{x}_1 - \bar{x}_2 &= \frac{3 - 5}{\sqrt{2.5} + \sqrt{2.5}} = \frac{-2}{1 + 1} = \frac{-2}{2} = -1 \\
\end{align*}
\]

The degrees of freedom for this problem are

\[
\begin{align*}
n_1 + n_2 - 2 &= 5 + 5 - 2 = 8 \\
\end{align*}
\]

Turning to a t-table, we find that for a two-tailed test of significance, where we set our criteria for statistical significance at \( p < .05 \) (i.e., we can assume that we would only obtain a \( t \) statistic this large 5 times out of 100 if the two samples are actually from just one population), with degrees of freedom = 8, our obtained \( t \) statistic value must be 2.306 or greater to obtain statistical significance.

Since our \( t \) statistic is only 2, we can not say that in a nondirectional or two-tailed test, these two means are significantly different and represent different populations. Note however that if we lower our criterion for significance to \( p < .1 \), our \( t \) statistic does surpass the table value of 1.860. Also note that if we had conducted a directional, one-tailed test (e.g., we expected the mean of sample 1 to be smaller than the mean of sample 2 and we wanted to know whether that difference was significant at \( p < .05 \)), we would compare our \( t \) statistic of 2 with this same 1.860 and conclude that the difference is indeed significant.

VII. EXERCISE: Here are two sets of data gathered from independent samples: 12, 13, 15, 16, 18 and 14, 16, 18, 19 20. Are the means of these two samples significantly different? Conduct a nondirectional test and use \( p < .05 \) as your criterion for statistical significance.

<table>
<thead>
<tr>
<th>Group 1</th>
<th>((x - \bar{x}_1))</th>
<th>((x - \bar{x}_1)^2)</th>
<th>Group 2</th>
<th>((x - \bar{x}_2))</th>
<th>((x - \bar{x}_2)^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td></td>
<td></td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td>20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\sum(x - \bar{x}_1)^2 &= \ldots \\
\sum(x - \bar{x}_2)^2 &= \ldots \\
\end{align*}
\]
\bar{x}_1 = \quad \bar{x}_2 = \\

s^2 = \\

t = \\

df = \\

\text{t statistic needed to reach significance:} \\

Conclusion:
ONE-WAY ANALYSIS OF VARIANCE (ANOVA)

The t-test compares the means of two samples to see if they are different and therefore likely to represent different populations; the analysis of variance (ANOVA) does the same thing but it can compare the means of not just two samples but many. Most frequently used in experimental designs, the different samples represent different conditions or levels of the same variable. For example, we might have three groups, or levels, of a self-disclosure variable and want to find out whether the means of these three groups differ significantly; an ANOVA would be the appropriate statistical technique.

Of course, it is possible to conduct a t-test between groups 1 and 2, a t-test between groups 2 and 3, and a t-test between groups 1 and 3, but the results of each of these tests would be based on probabilities that assume only one test is being conducted. The more tests we calculate, the more likely we will obtain an extreme t statistic value even if it is just a fluke and does not represent a true difference between means. By conducting the three separate tests we increase the chances of believing we have found a difference when we really have not.

The ANOVA uses the $F$ statistic, named for its creator, Sir R. A. Fisher. If an obtained $F$ statistic is statistically significant, this means that there is a statistically significant difference among the means of the 2 or more groups being compared. What the $F$ statistic does NOT tell the researcher is where that difference is, i.e., which means are significantly different from which other means (see below).

I. Assumptions and requirements.
   A. As with the t-test, the ANOVA requires you to compare scores on the same (and only one) variable (the dependent variable in experiments).
   B. It is assumed that the populations from which the samples are taken are normally distributed.
   C. It is assumed that the samples are normally distributed.
   D. It is assumed that the variance within one group is approximately equal to the variance of every other group.

II. Conceptual rationale. Although the purpose is to determine whether there is a significant difference between means, this is done by computing a ratio of variances. The logic is that if there are differences in the means of the various groups, the variance between the groups should be greater than the variance within each group. That is, the average difference between the mean of one group and the mean of another group, should be greater than the average difference between the score of one member of a (typical) group and the mean of that group. So:

$$F = \frac{s_b^2}{s_w^2}$$

where:

$s_b^2 = $ The variance between groups
$s_w^2 = \text{The variance within a group}$

Note that this logic is actually the same as that used in the calculation of the t-test, in which the numerator represents the difference between the (2) means and the denominator represents the average variance within the (2) groups.

To illustrate this graphically:

---

**III. Factors which influence whether or not a significant $F$ statistic is found:**

A. Magnitude of the differences between means. The greater the differences, the more likely they will be significant (this is the numerator of the formula for the $F$ statistic).

B. Variability within each group. The less variability within groups, the greater the chances for significance (this is the denominator of the formula for the $F$ statistic).

C. Number of observations within the groups. More observations means more degrees of freedom; and more degrees of freedom increases the likelihood of significance. However, this effect levels off after $df = 40$.

**IV. Two types of ANOVA based on type of samples.** Just as with the t-test there are two types of ANOVA, each with its own formula, based on the fact that there are two types of samples that can be compared, independent and matched.

A. The standard type of ANOVA assumes that the different groups being compared are made up of different (independent) people, objects, etc.

B. When the samples are matched, as when they are the different levels of a within subject factor in an experiment (e.g., all subjects watch and respond to three types of television program; responses to the first program constitute the scores in group 1, responses to the second program (by the same subjects) constitute the scores in group 2, etc.), the ANOVA is called a "repeated measure" ANOVA.

**V. CALCULATIONS:**

---
A. Again the overall formula is:

$$F = \frac{S_b^2}{S_w^2}$$

where:

$S_b^2 = $ The variance between groups

$S_w^2 = $ The variance within a group

B. To calculate $S_b^2$:

$$S_b^2 = \frac{\sum_{j=1}^{k} n_j (\bar{x}_j - \bar{X})^2}{(k - 1)}$$

where:

$k = $ Total number of groups

$j = $ The numerical identifier of a specific group (e.g., group 1, 2, etc.)

$n_j = $ Number of observations (subjects) in a specific group (e.g., in group 1, 2, etc.)

$\bar{x}_j = $ Mean of a specific group

$\bar{X} = $ Grand mean (mean of all scores in all groups)

C. To calculate $S_w^2$:

$$S_w^2 = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_i} ((x_{ij} - \bar{x}_j)^2)}{(N - k)}$$

where:

$n_i = $ The numerical identifier of a specific observation or subject in a group (e.g., subject 1, 2, etc.)

$x_{ij} = $ The numerical identifier of a specific observation or subject in a specific group (e.g., subject 1 in group 2, subject 2 in group 3, etc.)
\( N \) = The total number of scores in all groups combined

D. To determine whether the value of the \( F \) statistic we obtain represents a statistically significant difference among the means of the samples we must consult a table of \( F \) statistic values and probability values (found in most statistics books and typically called an F-table).

1. To use the table first calculate the appropriate degrees of freedom. In this case the formulae are

\[
\text{df}_1 = k - 1 \\
\text{df}_2 = N - k
\]

Note that \( \text{df}_1 \) is the same as the denominator in the formula for \( S_{b}^2 \), and \( \text{df}_2 \) is the same as the denominator in the formula for \( S_{w}^2 \). Since \( S_{b}^2 \) is the numerator in the formula for \( F \) and \( S_{w}^2 \) is the denominator in the formula for \( F \), \( \text{df}_1 \) is sometimes called \( \text{df}_{\text{numerator}} \) and \( \text{df}_2 \) is sometimes called \( \text{df}_{\text{denominator}} \).

2. See the instructions that accompany the F-table to determine the p value associated with the obtained value of the \( F \) statistic. The larger the obtained value of the \( F \) statistic the more certain the researcher can be that the differences among the means of the samples is the result of the samples being taken from different populations (rather than merely a fluke resulting from limited samples of one population).

VI. EXAMPLE: Below are scores from three independent groups. Are there significant differences among the means of these groups?

A. First we find \( S_{b}^2 \):

1. Calculate the grand mean and the mean of each group.
2. Take deviation of each group mean from the grand mean.
3. Square this deviation.
4. Multiply this squared deviation by the number of subjects in that group.
5. Repeat steps 1-3 for all groups, and add results.

6. Divide this sum by the number of groups minus 1.
<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>15</td>
<td>20</td>
<td>25</td>
</tr>
</tbody>
</table>

\[
\bar{x}_1 = \frac{15}{5} = 3 \quad \bar{x}_2 = \frac{20}{5} = 4 \quad \bar{x}_3 = \frac{25}{5} = 5
\]

\[
\bar{X} = \frac{15 + 20 + 25}{5 + 5 + 5} = \frac{60}{15} = 4 \quad \bar{X} = \frac{3 + 4 + 5}{3} = 4
\]

\[
n_1 = n_2 = n_3 = 5
\]

\[
S_{b2} = \sum_{j=1}^{k} n_j ((x_j - \bar{X})^2) \quad \frac{1}{(k - 1)}
\]

\[
= \frac{n_1((\bar{x}_1 - \bar{X})^2) + n_2((\bar{x}_2 - \bar{X})^2) + n_3((\bar{x}_3 - \bar{X})^2)}{3 - 1}
\]

\[
= \frac{5 ((3 - 4)^2) + 5 ((4 - 4)^2) + 5 ((5 - 4)^2)}{2}
\]

B. Next we find \(S_{w2}\):

1. Take the deviation of each score from the mean of the group that score belongs to.
2. Square these deviations.
3. Add the squared deviations.
4. Divide this sum by the total number of observations minus the number groups.
<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( (x_{11} - \bar{x}_1)^2 )</td>
<td>( (x_{12} - \bar{x}_2)^2 )</td>
<td>( (x_{13} - \bar{x}_3)^2 )</td>
</tr>
<tr>
<td>1 - 3 = -2</td>
<td>2 - 4 = -2</td>
<td>3 - 5 = -2</td>
</tr>
<tr>
<td>2 - 3 = -1</td>
<td>3 - 4 = -1</td>
<td>4 - 5 = -1</td>
</tr>
<tr>
<td>3 - 3 = 0</td>
<td>4 - 4 = 0</td>
<td>5 - 5 = 0</td>
</tr>
<tr>
<td>4 - 3 = 1</td>
<td>5 - 4 = 1</td>
<td>6 - 5 = 1</td>
</tr>
<tr>
<td>5 - 3 = 2</td>
<td>6 - 4 = 2</td>
<td>7 - 5 = 2</td>
</tr>
</tbody>
</table>

\[
N = 5 + 5 + 5 = 15
\]

\[
S_w^2 = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_i} ((x_{ij} - \bar{x}_j)^2)}{(N - k)}
\]

\[
= \frac{( (x_{11} - \bar{x}_1)^2 + (x_{21} - \bar{x}_1)^2 ) + (x_{31} - \bar{x}_1)^2 ) + (x_{41} - \bar{x}_1)^2 + (x_{51} - \bar{x}_1)^2 ) + (x_{12} - \bar{x}_2)^2 + (x_{22} - \bar{x}_2)^2 + (x_{32} - \bar{x}_2)^2 ) + (x_{42} - \bar{x}_2)^2 + (x_{52} - \bar{x}_2)^2 ) + (x_{13} - \bar{x}_3)^2 + (x_{23} - \bar{x}_3)^2 + (x_{33} - \bar{x}_3)^2 ) + (x_{43} - \bar{x}_3)^2 + (x_{53} - \bar{x}_3)^2 ) }{(N - k)}
\]

\[
= \frac{( (1 - 3)^2 + (2 - 3)^2 + (3 - 3)^2 ) + (x_{4} - 3)^2 + (5 - 3)^2 ) + (2 - 4)^2 + (3 - 4)^2 + (4 - 4)^2 ) + (5 - 4)^2 + (6 - 4)^2 ) + (3 - 5)^2 + (4 - 5)^2 + (5 - 5)^2 ) + (6 - 5)^2 + (7 - 5)^2 ) }{(N - k)}
\]

\[
= \frac{4 + 1 + 0 + 1 + 4 + 4 + 1 + 0 + 1 + 4 }{15 - 3}
\]

\[
= 30 = 2.5
\]
C. Then we calculate the value of $F$:

$$F = \frac{5}{2.5} = 2.00$$

D. Interpretation. Turning to an F-table, we find that if we set our criteria for statistical significance at $p < .05$ (i.e., we can assume that we would only obtain an $F$ statistic this large 5 times out of 100 if all of the samples are actually from just one population), with $df_1 = (k - 1) = (3 - 1) = 2$, and $df_2 = (N - k) = (15 - 3) = 12$, our obtained $F$ statistic value must be 3.89 or greater to obtain statistical significance. Since our $F$ statistic is only 2, we cannot say that these means are significantly different and represent different populations. Note that because the table indicates that with these degrees of freedom the probability of obtaining an $F$ statistic value of 1.46 would be .25 (25 times out of 100, or one quarter of the time), we can tell that the p-value associated with our $F$ statistic is less than .25 (but more than the criterion of .05).

VII. Another way to conceptualize and calculate an ANOVA. The formula for the $F$ statistic is a ratio of variances. The ANOVA can also be thought of as "partitioning" (i.e., dividing up) the variance in a set of data. This approach provides a short-cut way to calculate the $F$ statistic by hand, is the basis for the standard display of ANOVA results, and also provides an important foundation for understanding more complex ANOVA designs.

A. Rationale. Recall that the main part of the formula for $S$, the standard deviation of a sample of data, is:

$$\sum \frac{(x - \bar{x})^2}{N - 1}$$

The numerator of this formula is called a "sum of squares," (as discussed in the notes titled "Measures of dispersion").

The formula for $F$ contains two sums of squares. One is in the numerator of the calculation for $S_{b^2}$ (the sum of the squared deviations of each group mean from the grand mean). The other sum of squares is in the numerator of the calculation for $S_{w^2}$ (the sum of the squared deviations of each individual observation from the mean of the group to which it belongs). Remember that the denominator of both $S_{b^2}$ and $S_{w^2}$ are degrees of freedom. When a sum of squares is divided by its degrees of freedom, the result is called a Mean Squares (MS). Therefore, the formula for $F$ is often defined as
The sum of the between group sum of squares ($SSB$), and the within group sum of squares ($SSW$), is a quantity called the total sum of squares ($SST$). So $SST = SSB + SSW$ (this is the basis of the short cut formula for calculating $F$).

The means being compared in an ANOVA are typically the mean responses obtained from subjects who are exposed to different levels of a variable being manipulated in an experiment, i.e, the independent variable (IV). Because a manipulation is involved, a significant difference between groups suggests an impact of the manipulation (the IV). Therefore, the calculations and statistics of ANOVA that involve differences between groups are identified by referring to the independent variable.

As discussed, the bigger the variance within groups ($MSW$), the smaller the $F$ statistic will be, and the less likely it will be significant. In an idealized world, there would be little or no difference between scores within each group; the only differences would be between the groups. For this reason, the calculations and statistics of ANOVA that involve differences within groups are identified by referring to them as "error."

B. The ANOVA table. The standard way to present the results of an ANOVA is based on the terms and ideas discussed here. It looks like this:

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>Sum of squares</th>
<th>Degrees of freedom</th>
<th>Mean squares</th>
<th>$F$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between group (The IV)</td>
<td>SSB</td>
<td>$k - 1$</td>
<td>$S_b^2 = \frac{SSB}{df_1}$</td>
<td>$S_b^2$</td>
<td>$p$</td>
</tr>
<tr>
<td>Within group (Error)</td>
<td>SSW</td>
<td>$N - k$</td>
<td>$S_w^2 = \frac{SSW}{df_2}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>SST</td>
<td>$N - 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C. The short cut calculation of an $F$ statistic is this:
\[
\text{SST} = ( \sum_{ij} x_{ij}^2 ) - \frac{\left( \sum_{ij} (x_{ij}) \right)^2}{N}
\]

\[
\text{SSB} = \sum_{i} \left( \frac{T_i^2}{n_i} \right) - \frac{\left( \sum_{ij} (x_{ij}) \right)^2}{N}
\]

\[
\text{SSW} = \text{SST} - \text{SSB}
\]

where:

\( T_i \) = The total of all observations in a specific group (group i)

This looks much worse than it is. An example follows.

VIII. EXAMPLE: Here is the information we are given:

Independent Variable = Types of instruction

Levels of IV:

1. Lecture
2. Computer
3. Combination

Dependent Variable = Number of sales per week

Subjects (Ss) = 12 trainees from a total trainee population. The subjects have been randomly assigned to levels of the independent variable.

The data:

<table>
<thead>
<tr>
<th>Lecture</th>
<th>Computer</th>
<th>Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

TOTAL 10 18 31

Grand total = \( \sum_{ij} (x_{ij}) = ( 10 + 18 + 31 ) = 59 \)

A. Computation of sums of squares

1. Sum of squares total (SST).
a. Square all raw scores and sum them; i.e.,
\[ \sum_{ij} (x_{ij}^2) \]

SQUARED RAW SCORE MATRIX

<table>
<thead>
<tr>
<th></th>
<th>Lecture</th>
<th>Computer</th>
<th>Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>36</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>49</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

TOTAL: \( 1 + 25 + 1 + 9 + 9 + 36 + \ldots + 16 = 395 \)

b. Take the grand total and square it; i.e.,
\[ (\sum_{ij} (x_{ij}))^2 \]

\[ 59 \times 59 = 3481 \]

c. Divide the above number by total number of Ss (i.e., \( N \)).

\[ 3481 / 12 = 290.0833 \]

This number (290.0833) is known as the correction term.

d. Calculate \( SST \) by subtracting the correction term from the sum of the squared raw data.

\[ 395.00 - 290.0833 = 104.9167 \]

2. Sum of squares between (SSB).

a. Take the treatment group totals \( (T_i) \), square each total \( (T_i^2) \), divide each by the number of observations in that group

\[ \frac{\sum_{i} (T_i^2)}{n_i} \]

and then sum them

\[ \sum_{i} \left( \frac{T_i^2}{n_i} \right) \]

\[ SSB = (10)^2 + (18)^2 + (31)^2 \]
\[
\begin{align*}
\frac{4}{4} + \frac{4}{4} + \frac{4}{4} &= \frac{100}{4} + \frac{324}{4} + \frac{961}{4} = \frac{1385}{4} = 346.25 \\
\end{align*}
\]

b. Calculate \(SSB\) by subtracting the correction term (the same number used to calculate \(SST\)) from this number.

\[
346.25 - 290.0833 = 56.1667
\]

3. Sum of squares within (\(SSW\)).
   a. Calculate \(SSW\) by subtracting \(SSB\) from \(SST\):

\[
SST - SSB = SSW
\]

\[
104.9167 - 56.1667 = 48.75
\]

4. Degrees of freedom. Now that we have the sums of squares we have to figure the degrees of freedom and divide the SSs by the corresponding degrees of freedom to get the mean squares (\(MS\)).
   a. \(Df\) for \(SSB\) = \(df_1\) = \((k - 1)\), or the number treatment levels of the independent variable minus 1 = 3 - 1 = 2.
   b. \(Df\) for \(SSW\) = \(df_2\) = \((N - k)\), or the total number of observations (subjects) minus 3 = 12 - 3 = 9.
   c. \(Df\) for \(SST\) = \((k - 1) + (N - k)\) = \((k - 1 + N - k)\) = \((N - 1)\) = 12 - 1 = 11.

5. Mean squares.

\[
\begin{align*}
MSB &= \frac{SSB}{df_1} = \frac{56.1667}{2} = 28.0833 \\
MSW &= \frac{SSW}{df_2} = \frac{48.7500}{9} = 5.4166
\end{align*}
\]

6. The \(F\) statistic

\[
F = \frac{MSB}{MSW} = \frac{28.0833}{5.4166} = 5.18
\]

7. The ANOVA table:

<table>
<thead>
<tr>
<th>Source of</th>
<th>Sum of</th>
<th>Degrees of</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
8. Interpretation. To discover whether or not the $F$ value is statistically significant at the probability level we have set, we turn to an $F$-table. We find that if we set our criteria for statistical significance at $p < .05$ (i.e., we can assume that we would only obtain an $F$ statistic this large 5 times out of 100 if all of the samples are actually from just one population), with degrees of freedom 2 and 9, our obtained $F$ statistic value must be 4.26 or greater to obtain statistical significance.

Since our $F$ statistic is 5.18, we can say that these means are significantly different and represent different populations.

IX. EXERCISE: Here's what we know:

Independent variable = Levels of disclosure
Levels of IV:
1. High
2. Moderate
3. Low

Dependent variable = Number of friends
Number of subjects = 15

DATA MATRIX

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Moderate</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Is there a statistically significant (at the $p < .05$ level) difference among the mean number of friends for subjects who are high, moderate, and low in level of disclosure?
TWO-WAY ANALYSIS OF VARIANCE (ANOVA)

A one-way analysis of variance (ANOVA) can tell the researcher whether there is any significant difference among the means of two or more samples, but in a one-way ANOVA all of the samples represent different values or levels of just one (independent) variable (e.g., method of instruction (lecture, computer, combination)). The one-way ANOVA is generally used in experiments to assess the impact of varying the independent variable on some dependent variable (e.g., number of sales per week). A two-way ANOVA is used to assess the impact of varying not just one but two independent variables on a dependent variable (a three-way ANOVA involves three independent variables, and so on).

I. Conceptual rationale.
   A. Variance components. In a two-way ANOVA, the total variance is divided not just between between group variance and within group variance. Total variance is the sum of variance due to the first independent variable (A), the second independent variable (B), the interaction of the two independent variables (AB), and within group or error variance (E).
   B. Interactions. An interaction between two independent variables (IVs) means that the impact on a dependent variable (DV) of changing the value of one IV is influenced by the value of the other IV. Rather than having separate, "independent" impacts on a dependent variable, the two variables "interact" with each other to produce an effect. An example will help explain: High credibility generally produces more attitude change than low credibility; and weak fear appeals generally produce more attitude change than strong fear appeals. So IV A is source credibility (high, low), IV B is strength of fear appeal (weak, strong), and the DV is attitude change. If there is no interaction between these two IVs a graph of the results of an experiment would look like this:

```
   10 |
     9
     8      X
     7------X  Strong
  Amount of
     6  X
     5  X
     4  X
     3
     2
     1
     0

   |  High  Low |

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

   Source credibility
```

Strength of fear appeal
Note the parallel lines in this graph, indicating that regardless of level of source credibility, a weak fear appeal is associated with attitude change 4 units greater than a strong fear appeal, and that regardless of strength of fear appeal, a high credibility source is associated with attitude change 2 units greater than a low credibility source.

If there was an interaction between the two IVs in this example, the graph might look like this:

![Graph](image)

In this graph the lines are not parallel, indicating the presence of an interaction. In these results the difference in amount of attitude change associated with high versus low credibility depends on the strength of the fear appeal. If the fear appeal is weak, a high credibility source is associated with attitude change that is 4 units (10 - 6) greater than a low credibility source. But if the fear appeal is strong, the difference between attitude change for the high and low credibility source is only 2 (4 - 2).

Similarly, the difference in amount of attitude change associated with weak versus strong fear appeals depends on the credibility of the source. If the source's credibility is high, a weak fear appeal is associated with attitude change that is 6 units (10 - 4) greater than a strong fear appeal. But if the source's credibility is low, the difference between attitude change for the weak and strong fear appeals is only 4 (6 - 2).

II. Formulas for calculation. The short cut calculations for a two-way ANOVA are only slightly different than those for the one-way ANOVA:

\[
SST = \left( \sum_{ij} (x_{ij}^2) \right) - \frac{\left( \sum_{ij} (x_{ij}) \right)^2}{N}
\]
SSA = $\sum \left( \frac{\left( TA_i \right)^2}{n_{Ai}} \right) - \frac{\left( \sum_{ij} \left( x_{ij} \right) \right)^2}{N}$

SSB = $\sum \left( \frac{\left( TB_i \right)^2}{n_{Bi}} \right) - \frac{\left( \sum_{ij} \left( x_{ij} \right) \right)^2}{N}$

SSAB = $\sum \left( \frac{\left( TAB_{ij} \right)^2}{n_{ABi}} \right) - SSA - SSB - \frac{\left( \sum_{ij} \left( x_{ij} \right) \right)^2}{N}$

SSE = SST - SSA - SSB - SSAB

where:
SST = Sum of Squares Total.
SSA = Sum of Squares for independent variable A.
SSB = Sum of Squares for independent variable B.
SSAB = Sum of Squares for the interaction between independent variables A and B.
SSE = Sum of Squares for Error (same as SSW, Sum of Squares Within).
TA_i = The total of all observations in the ith level of IV A.
TB_i = The total of all observations in the ith level of IV B.
TAB_{ij} = The total of all observations in the ith level of IV A and the ith level of IV B.
n_{Ai} = The number of observations in the ith level of IV A.
n_{Bi} = The number of observations in the ith level of IV B.
n_{ABi} = The number of observations in the ith level of IV A and the ith level of IV B.
III. EXAMPLE:
A. The problem:
   Independent variable A = Method of instruction
   3 levels:
   1. Lecture
   2. Computer
   3. Combination

   Independent variable B = Gender of the subject
   2 levels:
   1. Female
   2. Male

   Dependent variable: = Number of sales per week

   Total number of subjects (N) = 24

   Number of subjects in each condition (n) = 4

   DATA MATRIX

   Method (A)

<table>
<thead>
<tr>
<th>Lecture</th>
<th>Computer</th>
<th>Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(B)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>TOTALS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>40</td>
<td>59</td>
</tr>
</tbody>
</table>
B. Computations:

1. Sum of squares total (SST)
   a. Square all the raw scores and sum them.

   **MATRIX FOR RAW SCORES SQUARED**

   Method (A)

<table>
<thead>
<tr>
<th>Lecture</th>
<th>Computer</th>
<th>Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>49</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>81</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>81</td>
</tr>
</tbody>
</table>

   Gender (B)

<table>
<thead>
<tr>
<th>Lecture</th>
<th>Computer</th>
<th>Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>36</td>
</tr>
</tbody>
</table>

   Male

<table>
<thead>
<tr>
<th>Lecture</th>
<th>Computer</th>
<th>Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>49</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>64</td>
<td>49</td>
</tr>
</tbody>
</table>

   **TOTALS**

   Sum of squared raw scores = 20 + 224 + 457 = 701.

   b. Now find the correction term. Take the grand total and square it, then divide by total number of subjects \( N = 3 \times 2 \times 4 = 24 \).

   \[
   \text{Grand total} = 1 + 2 + 1 + 2 + \ldots + 6 + 6 + 5 + 7 = 111
   \]

   \[
   \frac{111^2}{24} = \frac{12321}{24} = 513.375
   \]

   c. Now calculate SST

   \[
   701.00 - 513.375 = 187.625
   \]

2. Sum of squares for independent variable A (SSA)
   a. Take the total of the observations at each level of IV A, square them, divide them by the number of observations, and sum them.

   \[
   = \frac{(12)^2}{8} + \frac{(40)^2}{8} + \frac{(59)^2}{8}
   \]

   \[
   = \frac{144}{8} + \frac{1600}{8} + \frac{3481}{8}
   \]
\[
\frac{5225}{8} = 653.125
\]

b. Then subtract the correction term.

\[
SSA = 653.125 - 513.375 = 139.75
\]

3. Sum of squares for independent variable B (\(SSB\))
   a. Take the total of the observations at each level of IV B, square them, divide them by the number of observations, and sum them.

\[
\frac{(56)^2}{12} + \frac{(55)^2}{12} = \frac{3136}{12} + \frac{3025}{12} = \frac{6161}{12} = 513.4167
\]

b. Then subtract the correction term.

\[
SSB = 513.4167 - 513.375 = .0417
\]

4. Sum of squares for the interaction between independent variables A and B (\(SSAB\))
   a. Take the total of the observations at each combination of level of IV A and level of IV B, square them, divide them by the number of observations, and sum them.

\[
\frac{(6)^2}{4} + \frac{(15)^2}{4} + \frac{(35)^2}{4} + \frac{(6)^2}{4} + \frac{(25)^2}{4} + \frac{(24)^2}{4} = \frac{2723}{4} = 680.75
\]

b. Then subtract \(SSA\), \(SSB\), and the correction term.

\[
SSAB = 680.75 - 139.75 - .0417 - 513.75 = 27.578
\]
5. Sum of squares for error (SSE). The error term is what's left over:

\[ SSE = SST - SSA - SSB - SSAB \]
\[ = 187.625 - 139.75 - .0417 - 27.578 = 20.2553 \]

6. Degrees of freedom
   a. For SSA, df = (a - 1) = (3 - 1) = 2
   b. For SSB, df = (b - 1) = (2 - 1) = 1
   c. For SSAB, df = (a - 1)(b - 1) = (3 - 1)(2 - 1) = 2 * 1 = 2
   d. For SSE, df = ab(n - 1) = 3 * 2 * (4 - 1) = 6 * 3 = 18
   e. For SST, df = abn - 1 = 3 * 2 * 4 - 1 = 24 - 1 = 23

7. Calculating the \( F \) statistic for IV A, IV B, and the interaction of IV A and IV B.
   a. For A: MSA / MSE
   b. For B: MSB / MSE
   c. For AB: MSAB / MSE

8. The standard ANOVA table:

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>Sum of squares</th>
<th>Degrees of freedom</th>
<th>Mean squares</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method of instruction (A)</td>
<td>139.7500</td>
<td>2</td>
<td>69.8750</td>
<td>62.10</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Gender (B)</td>
<td>.0417</td>
<td>1</td>
<td>.0417</td>
<td>.04</td>
<td>&lt;.85</td>
</tr>
<tr>
<td>Interaction (AB)</td>
<td>27.5780</td>
<td>2</td>
<td>13.7890</td>
<td>12.25</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Within group (Error)</td>
<td>20.2553</td>
<td>18</td>
<td>1.1252</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>187.6250</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9. Interpreting the interaction effect
   a. ALWAYS EXAMINE SIGNIFICANT INTERACTION EFFECTS BEFORE EXAMINING SIGNIFICANT MAIN EFFECTS. If the interaction is significant, it may influence any interpretation of main effect results.
b. Make graphs of cell means to *visually* represent findings.

**MATRIX OF CELL MEANS**

<table>
<thead>
<tr>
<th></th>
<th>Lecture</th>
<th>Computer</th>
<th>Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>1.50</td>
<td>3.75</td>
<td>8.75</td>
</tr>
<tr>
<td>Male</td>
<td>1.50</td>
<td>6.25</td>
<td>6.00</td>
</tr>
</tbody>
</table>

The next task is to isolate the specific differences among the means that are responsible for the significant findings for the interaction of A and B, and for the "main effect" of A. See the notes titled "Interpreting a significant F-value."

IV. **EXERCISE:** Here's the problem:

Independent variable A = Gender of Subject
2 levels:
1. Female
2. Male

Independent variable B = Level of Self-Disclosure
3 levels:
1. High
2. Moderate
3. Low

Dependent variable = Satisfaction with conversation

Total number of subjects (N) = 30

Number of subjects per cell = 5

<table>
<thead>
<tr>
<th>DATA MATRIX</th>
</tr>
</thead>
</table>
| Gender 
| Female | Male |
| --- | --- | --- | --- |
| High | 3 | 5 | 
| 4 | 5 | 
| 2 | 6 | 
| 5 | 7 | 
| 3 | 5 | 
| Self- | 2 | 7 | 
| dis- | 1 | 5 | 
| closure | 3 | 7 | 
| Medium | 6 | 8 | 
| 1 | 11 | 
| Low | 4 | 4 | 
| 5 | 3 | 
| 6 | 2 | 
| 5 | 5 | 
| 9 | 4 | 

Calculate a two-way ANOVA and fill in the appropriate ANOVA table:
<table>
<thead>
<tr>
<th>variance</th>
<th>squares</th>
<th>freedom</th>
<th>squares</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(B)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Interaction (AB)

Within group
(Error)

Total
INTERPRETING A SIGNIFICANT F VALUE

When we find a significant F value (either for a main effect or interaction effect), the significant F only tells us that there is a statistically significant difference between groups. It does not tell us which means are different from each other. As a result we may have to perform follow-up tests to determine two things: 1) where the differences are, and 2) the practical significance, usually defined as the amount of variance explained, associated with the statistically significant F value.

For a variable with 2 levels within any ANOVA it is easy to determine where the significant difference between cells is because there is only one difference: the difference between the means of the two treatment groups is, by definition, significant or we would not have found a main effect for that variable.

However, if there are more than two levels for the variable associated with a significant F value, and/or a significant interaction effect, we need to conduct multiple comparison tests.

There are many different multiple comparison tests. It is beyond the scope of this course to discuss them in detail. However, some of the important concepts concerning multiple comparisons are presented below.

I. Pairwise and compound comparisons (contrasts).
   A. **Pairwise comparisons.** Usually researchers perform pairwise comparisons. A pairwise comparison is a comparison between individual means. For example:

   \[ l = 1 \times (\bar{X}_1) - 1 \times (\bar{X}_2) \] (Lecture versus Computer)

   \[ l = 1 \times (\bar{X}_2) - 1 \times (\bar{X}_3) \] (Computer versus Combination)

   \[ l = 1 \times (\bar{X}_1) - 1 \times (\bar{X}_3) \] (Lecture versus Combination)

   B. **Compound comparisons.** Sometimes we may want to use compound comparisons, or comparisons of the means of two or more groups with another group. For example:

   \[ l = \frac{((\bar{X}_1) + (\bar{X}_2)) - \bar{X}_3}{2} \] (Lecture + Computer versus Combination)

   which is written as

   \[ l = (.5 \bar{X}_1) + (.5 \bar{X}_2) + (-1 \bar{X}_3) \]

II. Overall or experiment-wise error rate. The reason that there are specific follow-up procedures for a finding of a significant F in an ANOVA is that we want to discover where the significant differences are occurring without committing what is called a Type I error: rejecting the null hypothesis (here, that there are no significant differences) when it should be retained. When we conduct multiple comparisons there is a risk of committing Type I error unless we calculate our error rate over the entire collection of comparisons that are to
be made. This is particularly important if some of our tests are nonindependent (i.e., the same information, such as a cell mean, is used in more than one comparison).

The overall error rate, the probability of falsely rejecting the null hypothesis that there are no differences among a group of means for at least one of a group of comparisons is:

\[ 1 - (1 - \alpha)^c \]

where:
\[ \alpha = \text{Probability of making a Type I error on an individual test} \]
\[ c = \text{Total number of comparisons being made} \]

The more comparisons that are made, the greater the probability of making at least one Type I error.

<table>
<thead>
<tr>
<th>c, number of comparisons</th>
<th>( \alpha = .1 )</th>
<th>( \alpha = .05 )</th>
<th>( \alpha = .01 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.100</td>
<td>.050</td>
<td>.010</td>
</tr>
<tr>
<td>2</td>
<td>.190</td>
<td>.097</td>
<td>.020</td>
</tr>
<tr>
<td>3</td>
<td>.271</td>
<td>.143</td>
<td>.030</td>
</tr>
<tr>
<td>4</td>
<td>.344</td>
<td>.185</td>
<td>.039</td>
</tr>
<tr>
<td>5</td>
<td>.410</td>
<td>.226</td>
<td>.049</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>10</td>
<td>.651</td>
<td>.401</td>
<td>.096</td>
</tr>
</tbody>
</table>

[Note that this table applies only to what are called orthogonal contrasts, in which the weights (the numbers each mean is multiplied by) fall in a certain pattern.]

This table shows that if an experimenter wanted to compare 20 population means by making 10 comparisons, the probability of falsely concluding that a difference was significant for at least one of the comparisons could be as high as .401 when each test is performed with the \( \alpha < .05 \) criterion.

III. A priori and post hoc comparisons
A. A priori multiple comparisons. These comparisons are planned before the running of the study and are supported by existing theory and research. Since these comparisons are posited in advance, the study design often reflects these expected differences. Obviously, researchers usually plan relatively few a priori comparisons. There are different types of a priori comparisons, each requiring a different analytic technique.

B. Post hoc comparisons are those the researcher decides to make after the completion of the study and the finding of a significant \( F \) value.
IV. Specific techniques. Some of the common statistical techniques for making multiple comparisons include Fisher's Least Significant Difference, Tukey's W, Student-Newman-Keuls, Duncan's Multiple Range test, and Scheffe's S method. Each technique has pros and cons, and researchers and statisticians often differ concerning which technique is most appropriate in a given situation. For further details consult Ott (1984), Kennedy (1978; includes a comprehensive decision tree for choosing an appropriate technique), or another statistics text.
CORRELATION

Correlation is a statistic that indicates the extent to which two or more variables (sets of observations) vary systematically, i.e., when a change in one is matched by some change in the other. For example, if we ask a group of people how tall they are and how much they weigh, we should find that the taller a person is, the more he weighs. Note that the existence of this systematic variation does mean that ALL tall people weigh more than people who are shorter. And it certainly does not demonstrate a causal relationship (that is, that height causes weight or the reverse).

Although the correlation of any two variables can be calculated, correlational statistics are generally used when the researcher has two or more variables that contain scores or evaluations that are matched or linked in some way (e.g., the height and weight of the same people, the height of fathers and of sons, etc.), and when the researcher is interested in examining the relationship between the variables rather than differences between groups on one or more variables.

I. Interpreting correlations. Correlation coefficients (r) range from -1 to +1. The coefficient tells the researcher two things: 1) the magnitude of the relationship between the variables, and 2) the direction of the relationship.

A. When we speak of magnitude we mean the strength of the correlation, or how systematically or closely the variables covary. As Smith (1988) indicates, the following cutoffs are usually used to evaluate the strength of the relationship:

| +/-.01 - .25 | Weak relationship |
| +/- .26 - .55 | Moderate relationship |
| +/- .56 - .75 | Strong relationship |
| +/- .76 - .99 | Very strong relationship |
| +/- 1.00 | Perfect relationship |

B. The direction of the relationship is indicated by whether the correlation coefficient is positive or negative. In a positive relationship both variables vary the same way, i.e., when one increases the other increases or when one decreases the other decreases. In a negative relationship the variables vary in opposite ways, i.e., when one increases the other decreases or vice versa.

C. The correlation of two variables is often apparent when the relationship is seen graphically:

<table>
<thead>
<tr>
<th>A</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable B</td>
<td>Variable B</td>
</tr>
</tbody>
</table>
### D. Note that the correlation coefficient describes only a linear relationship, not a curvilinear one as in the last three figures, which show that there is in fact a strong relationship between the variables.

### II. CALCULATIONS: The most common correlation statistic is Pearson's Product Moment Correlation. The formula for Pearson's $r$ is:

\[
r = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \cdot \sqrt{\sum (y_i - \bar{y})^2}}
\]

### III. EXAMPLE: Assume we have 10 subjects who are given a communication competence measure and a communication apprehension measure. Thus, we have scores for each subject's communication competence ($X$) and communication apprehension ($Y$).

<table>
<thead>
<tr>
<th>Subject</th>
<th>$x_i$ ($x_i - \bar{x}$)</th>
<th>($x_i - \bar{x})^2$</th>
<th>$y_i$ ($y_i - \bar{y}$)</th>
<th>($y_i - \bar{y})^2$</th>
<th>$(x_i - \bar{x}) (y_i - \bar{y})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>7</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>(-1)</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

| Total   | 70                      | 26                   | 20                      | 6                    | 8                                |
A. The first thing we need to do is calculate the mean for each measure.

\[
\bar{X} = \frac{70}{10} = 7 \quad \bar{Y} = \frac{20}{10} = 2
\]

B. Next find the deviation of each subject's score from the mean of that measure; i.e.,

\[(x_i - \bar{X}) \quad \text{and} \quad (y_i - \bar{Y})\]

C. Then square the deviation scores for each measure; i.e.,

\[(x_i - \bar{X})^2 \quad \text{and} \quad (y_i - \bar{Y})^2\]

D. Then sum the squared deviation scores for each measure; i.e.,

\[\sum (x_i - \bar{X})^2 = 26 \quad \text{and} \quad \sum (y_i - \bar{Y})^2 = 6\]

E. Then find the cross products of the deviation scores for measures X and Y (i.e.,

\[(x_i - \bar{X})(y_i - \bar{Y})\]

F. Then sum the cross products of the deviation scores (here the total is 8).

G. Now we have the data we need to plug into the formula:

\[
r = \frac{\sum (x_i - \bar{X}) * (y_i - \bar{Y})}{\sqrt{\sum (x_i - \bar{X})^2} \ast \sqrt{\sum (y_i - \bar{Y})^2}} = \frac{8}{\sqrt{26} \ast \sqrt{6}} = \frac{8}{12.48} = .64
\]

H. Interpretation. What we have discovered is that our correlation coefficient is .64, indicating a strong positive relationship between the two variables.

IV. Follow-up analysis. In addition to finding the correlation coefficient, there are usually two other things the researcher wants to do: 1) determine if the correlation is statistically significant, and 2) determine the proportion of variance in one variable that is shared by variation in the other (sometimes phrased as "the amount of variance explained," even though correlation does not imply causation and therefore does not imply explanation).
A. To determine if the correlation is statistically significant the researcher usually uses the following \( t \) formula:

\[
\frac{r}{\sqrt{N-2}} = \frac{.64}{\sqrt{10-2}} = \frac{.64}{\sqrt{8}}
\]

\[
\frac{1 - r^2}{\sqrt{1 - .64^2}} = \frac{1 - .59}{\sqrt{1 - .59}} = \frac{.77}{.77}
\]

\[
= \frac{(.64)(2.82)}{.77} = \frac{1.80}{.77} = 2.34
\]

Next consult a table of \( t \) values to determine whether the \( t \) score is statistically significant. As always, pick an alpha level (.05) and figure the degrees of freedom (\( df = \) number of pairs of scores - 2). Here the \( t \) statistic must be equal to or greater than 2.306 for a two-tailed test when the alpha level is .05 and \( df = 10 - 2 = 8 \). Therefore our correlation is statistically significant at the .05 level.

B. To determine the amount of variance "explained" by the correlation between the two variables we simply convert our correlation coefficient (\( r \)) into a coefficient of determination, or \( r^2 \). In other words, we square the correlation coefficient. The relationship between \( r \) and \( r^2 \) is such that except for \( r^2 = 0 \) or 1, the value of \( r^2 \) is always lower than that of \( r \). Remember, the \( r^2 \) tells us the proportion of variance shared by the two variables.

V. EXERCISE: What is the correlation between the following two sets of scores? Is the correlation coefficient statistically significant at \( p < .05 \)?

<table>
<thead>
<tr>
<th>Subject</th>
<th>( x_i )</th>
<th>( (x_i - \overline{X}) )</th>
<th>( (x_i - \overline{X})^2 )</th>
<th>( y_i )</th>
<th>( (y_i - \overline{Y}) )</th>
<th>( (y_i - \overline{Y})^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
REGRESSION

A researcher often wants to know the relationship between two or more variables not just to describe that relationship but to use it to predict unknown values of one variable based on the known values of the other. For example, if we know the relationship between weight and height in a group of subjects, we may be able to use that knowledge to predict the weight of another set of people based on knowledge only of their height. To make such predictions, researchers use a statistical procedure called regression (of which there are several variations).

I. Conceptual rationale. To see how prediction using regression works, examine the graph below that summarizes the relationship between variable X and variable Y. The correlation between X and Y is .6, indicating a fairly strong, positive relationship.

\[ Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \]

where:

\( Y_i \) = The actual (not the predicted) value of variable Y for a specific score or observation (i.e., the \( i \)th score or observation)

\( \beta_0 \) = A regression coefficient, this is the "Y intercept" or the point on the graph at which the regression line intercepts the vertical axis (i.e., where \( X = 0 \))

\( \beta_1 \) = Another regression coefficient, this one representing the slope of the regression line, or the number of units the value of Y goes up or down when the value of X changes.
\( X_i \) = The value of variable X for a specific score or observation (i.e., the \( i \)th score or observation)

\( i = 1, ..., n \) (\( n \) is the total number of scores, which is the same for both X and Y)

\( \varepsilon_i \) = Random error

B. Missing information in the model of a regression line. Note that the researcher does not actually know the values of any of the terms in this description of the regression line except \( X_i \). The values exist in the population and the researcher can only use her sample data to estimate these terms. The best prediction of what the value of \( Y_i \) actually is for a given and known value of \( X_i \) is typically called y-hat.

Note also that the best estimate of the slope of the regression line (\( \beta_1 \)) is closely related to \( r \), the correlation of the two variables.

C. The error term. The existence of random error (\( \varepsilon_i \)) is important. At each point on the regression line (i.e., the point on the regression line at each value of X on the horizontal axis), one can imagine a vertical bell shaped probability distribution. The center of the distribution is the actual point on the regression line, which is the best (most likely) prediction of the value of Y given the specific value of X and knowledge of the relationship between X and Y. But random error means that not all predictions will be accurate; most will be relatively close to the regression line, others will be further away.

II. Evaluating the regression fit. The researcher will want to know the likely degree of accuracy of the predictions made by a regression procedure. This involves finding the same two pieces of information as in the follow-up to simple correlation: 1) Is the estimate of the regression coefficient associated with the predictor variable statistically significant?, and 2) What proportion of variance in the variable being predicted is shared by variation in the predictor variable? (again the phrase "amount of variance explained" is often used, but is misleading).

A. Significance of regression coefficients. Mathematically derived formulae are used to determine whether the regression coefficients are statistically significant. A significant coefficient means that the coefficient is unlikely to be 0, and that therefore the predictor variable associated with it plays a role in predicting the value of Y.

B. Coefficient of determination or r-squared. To find out what proportion of variance of Y is shared by the predictor variable is to determine the degree to which knowing the values of X helps describe variation in the values of Y. For this the researcher usually looks at the coefficient of determination (\( r^2 \), or "r-squared"). This statistic ranges from 0 to 1. If \( r^2 \) is 0 it means that X has not helped predict Y at all; the regression line is completely horizontal and at the mean of Y, so that regardless of the value of X, the predicted value of Y is always simply the mean of all values of Y. If \( r^2 \) is 1 it means that every point in the data falls exactly on the regression line, so that knowing the value of X always allows us to identify the corresponding value of Y. Of course, \( r^2 \) is usually somewhere
between 0 and 1.

The $r^2$ statistic is the squared correlation, so the square root of $r^2$ is $r$.

III. An ANOVA interpretation of regression. An interpretation of regression in the terms used to understand analysis of variance is possible, and helps to show that all basic statistical techniques are related. Just as total sums of squares ($\text{SST}$) equals the sum of between group sum of squares ($\text{SSB}$) and within group sum of squares ($\text{SSW}$ or $\text{SSE}$), the total variation in the variable $Y$ is $(Y_i - \bar{Y})$, and this variation can be broken down into component parts. First there is the total of the differences between the regression line value and the mean value of $Y$ ($y_{-\text{hat}_i} - \bar{Y}$). Then there is the differences due to random error, between the true value of $Y_i$ and the predicted value of $y_{-\text{hat}_i}$. When these deviations are squared and summed, the relationship continues to hold, so that an ANOVA table, with sums of squares and mean squares, can be created.

IV. Different types of regression. We have been discussing simple linear regression. There are many variations on this type of procedure.

A. Multiple regression. Here there is not just one predictor variable ($X$), but two or more ($X_1, X_2, \text{etc.}$). All of these predictor variables are then used to predict values of $Y$. The choice of predictor variables is important, because results for any given predictor are dependent on the other predictor variables in the regression (a fact researchers often forget!).

B. Stepwise regression. In multiple regression, the different predictor variables can be allowed to simultaneously predict values of $Y$, or they can be used one at a time, or in small or large groups (called blocks) to predict $Y$. For example, a researcher might want to see how well a set of demographic variables predict a certain variable before determining how much the prediction can be improved by adding other predictors.

C. Best subset regression. This is a technique used to identify the best group of predictor variables for a given variable $Y$. The procedure produces statistics to evaluate all possible combinations of predictor variables so the researcher can choose an appropriate group.
**CHI-SQUARE**

Chi-square ($\chi^2$) is a statistic that is used only with nominal level data. While statistics that are used for interval and ratio level data are typically based on means (as in t-tests and ANOVAs) or variance and covariance (as in correlation and regression), the chi-square is based on counts or frequencies of observations which fall into different (nominal) categories.

Chi-square can be used to examine differences between the number of observations in the levels of one variable (a one-way chi-square) OR to determine whether there is a statistically significant relationship between two or more nominal variables (multi-way chi-square).

I. **CALCULATIONS:** The formula for all applications of the chi-square statistic is based on a comparison of observed frequencies with the frequencies expected if there are no differences between levels of a variable or relationships between variables. The formula is:

\[
\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}
\]

where:

- $\chi^2$ = Chi-square test statistic value.
- $O_i$ = The observed number of observations in category $i$.
- $E_i$ = The expected number of observations in category $i$ assuming no differences (or relationship).

**NOTE:** The chi-square formula assumes that all categories or cells contain at least 5 cases or observations.

II. **EXAMPLE** of a one-way chi-square to examine differences between levels of a nominal variable.

A researcher is interested in students' choices of academic major. She wants to find out if there is a statistically significant difference between the observed frequencies of students who have chosen science, math, and English majors. We have the following data:

<table>
<thead>
<tr>
<th>Major</th>
<th>Science</th>
<th>Math</th>
<th>English</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of students:</td>
<td>13</td>
<td>7</td>
<td>10</td>
<td>30</td>
</tr>
</tbody>
</table>
A. The first step is to determine the expected cell values (E) if there were no difference between the number of students who choose each of these majors. In a one-way chi-square this is easily calculated by taking the total number of subjects and dividing by the number of levels of the variable, or: \(30/3 = 10\). In this case, the expected cell frequency for each cell would be 10.

B. Next we have to determine the difference between the observed and expected cell frequencies, or \((O - E)\).

\[
\begin{align*}
\text{Science} & : (13 - 10) = 3 \\
\text{Math} & : (7 - 10) = -3 \\
\text{English} & : (10 - 10) = 0
\end{align*}
\]

C. Now square those differences to eliminate negative numbers and exaggerate where true differences are occurring, or \((O - E)^2\).

\[
\begin{align*}
\text{Science} & : 3^2 = 9 \\
\text{Math} & : (-3)^2 = 9 \\
\text{English} & : 0^2 = 0
\end{align*}
\]

D. In order to correct for the squaring and correct for expected probability we divide each \((O - E^2)\) by E.

\[
\begin{align*}
\text{Science} & : 9/10 = .9 \\
\text{Math} & : 9/10 = .9 \\
\text{English} & : 0/10 = 0
\end{align*}
\]

E. Finally, to get our chi-square value we simply sum up the indices of difference. Thus:

\[\chi^2 = .9 + .9 + 0 = 1.8\]

F. To determine whether our chi-square value of 1.8 is statistically significant at our chosen alpha level (say \(p < .05\)) we must determine the degrees of freedom. In a one-way chi-square, the degrees of freedom are equal to the number of levels of the nominal variable \((k)\) minus 1, or \((k - 1)\). In this case we have three levels of the variable, so \(3 - 1 = 2\).

We then consult a distribution table of chi-square values and if our value is greater than or equal to that listed with 2 degrees of freedom and at .05 alpha level, then our value is statistically significant. Our chi-square is not significant at the error level we've set because it is less than the table value of 5.991.

III. EXAMPLE of a two-way chi-square or contingency table analysis.
We want to discover the relationship (if any) between a student's gender (male versus female) and her or his choice of academic major (science, math, English). If there were no relationship between gender and major, the percentage of men in each major would be the same as the percentage of women in each major. A chi-square determines the existence of a relationship by comparing the observed (actual) number of cases in each cell with the expected number of cases (that is, the number expected if there was no relationship).

This is the data matrix:

<table>
<thead>
<tr>
<th></th>
<th>Science</th>
<th>Math</th>
<th>English</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>50</td>
<td>10</td>
<td>40</td>
<td>100</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>10</td>
<td>60</td>
<td>10</td>
<td>80</td>
</tr>
<tr>
<td>TOTAL</td>
<td>60</td>
<td>70</td>
<td>50</td>
<td>180</td>
</tr>
</tbody>
</table>

Note: The letters in the data matrix identify for convenience the "cells" (combinations of values of the two variables).

A. First we want to find out what the expected probability for each cell is. To do this, we take the total number of cases in the row divided by the total number of cases in all cells, multiply it by the total number of cases in that column divided by the total number of cases in all cells, and finally, multiply that by the total number of cases in all cells.

Thus for the 6 cells we have:

a) \( E = \frac{100}{180} \times \frac{60}{180} \times 180 = 33.3 \)

b) \( E = \frac{100}{180} \times \frac{70}{180} \times 180 = 38.9 \)

c) \( E = \frac{100}{180} \times \frac{50}{180} \times 180 = 27.8 \)

d) \( E = \frac{80}{180} \times \frac{60}{180} \times 180 = 26.7 \)

e) \( E = \frac{80}{180} \times \frac{70}{180} \times 180 = 31.1 \)

f) \( E = \frac{80}{180} \times \frac{50}{180} \times 180 = 22.2 \)

B. Next, we find the difference between the observed \((O)\) and expected \((E)\) frequency for each cell.

\[
\begin{align*}
a) \; (O - E) & = 50 - 33.3 = 16.7 \\
b) \; (O - E) & = 10 - 38.9 = -28.9 \\
c) \; (O - E) & = 40 - 27.8 = 12.2 \\
d) \; (O - E) & = 10 - 26.7 = -16.7 \\
e) \; (O - E) & = 60 - 31.1 = 28.9 \\
f) \; (O - E) & = 10 - 22.2 = -12.2 \\
\end{align*}
\]

C. Next, we square the \((O - E)\) figures.
a) \((O - E)^2\) = 16.72 = 278.9
b) \((O - E)^2\) = -28.92 = 835.2
c) \((O - E)^2\) = 12.22 = 148.8
d) \((O - E)^2\) = -16.72 = 278.9
e) \((O - E)^2\) = 28.92 = 835.2
f) \((O - E)^2\) = -12.22 = 148.8

D. Then we divide each \((O - E)^2\) by the expected cell frequency.

a) \(\frac{(O - E)^2}{E}\) = \(\frac{278.9}{33.3}\) = 8.4
b) \(\frac{(O - E)^2}{E}\) = \(\frac{835.2}{38.9}\) = 21.5
c) \(\frac{(O - E)^2}{E}\) = \(\frac{148.8}{27.8}\) = 5.4
d) \(\frac{(O - E)^2}{E}\) = \(\frac{278.9}{26.7}\) = 10.4
e) \(\frac{(O - E)^2}{E}\) = \(\frac{835.2}{31.1}\) = 26.9
f) \(\frac{(O - E)^2}{E}\) = \(\frac{148.8}{22.2}\) = 6.7

E. Then sum all of these to find the chi-square statistic.

\(\chi^2 = 8.4 + 21.5 + 5.4 + 10.4 + 26.9 + 6.7 = 79.3\)

F. Interpretation: After computing the chi-square statistic, consult a table of chi-square values to determine if the statistic is significant. First, choose an alpha level, say .05. Degrees of freedom = the number of rows minus 1 times the number of columns minus 1, so \(df = (3 - 1)(2 - 1) = 2 \times 1 = 2\). In this example, the chi-square test statistic is significant at the .05 level and even at the .001 level (since 79.3 is greater than 13.82). This tells us that there is a strong relationship between gender and major.

IV. EXERCISE: Here is the data for the two nominal variables hair color, and gender. What is the chi-square statistic? Is it significant at the .05 level? If so, what does this mean?

<table>
<thead>
<tr>
<th>Hair color</th>
<th>Red</th>
<th>Black</th>
<th>Blonde</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>a</td>
<td>b</td>
<td>c</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>60</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>d</td>
<td>e</td>
<td>f</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>30</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>40</td>
<td>90</td>
<td>80</td>
<td>210</td>
</tr>
</tbody>
</table>

CELL O E \((O - E)^2\) \(\frac{(O - E)^2}{E}\)
a) 3  
b) 6  
c) 2  
d) 1  
e) 3  
f) 6
INTERPRETING A SIGNIFICANT CHI-SQUARE VALUE

If we have obtained a statistically significant chi-square value this only tells us that there is a statistically significant difference between the observed and expected frequencies or that there is a statistically significant relationship between two nominal levels variables. It does not tell us how strong that relationship is or which differences between observed and expected values are responsible for the significant chi-square. In order to answer these two questions we need to calculate further.

I. Measures of association. In order to determine the strength of the relationship between the variables we will need to calculate a measure of association. There are three standard measures of association used for the interpretation of significant chi-square results: the phi coefficient, the contingency correlation coefficient, and Cramer's V coefficient.

A. The Phi Coefficient ($\phi$) formula is:

$$\phi = \frac{\chi^2}{N}$$

where:

$N$ = the total number of cases or observations in all categories.

1. This statistic corrects for the effects of sample size on the value of chi-square (note the $N$ in the denominator).
2. The use of the phi coefficient should be reserved for studies involving 2 by 2 designs (designs with two nominal variables, each with only two values, such as gender). In those cases, the square root of the phi coefficient is equal to a correlation coefficient ($r$).
3. In cases where there are more than two variables or more than two categories for one or more variables, the phi coefficient does not have an upper limit, making it very difficult to interpret.
4. Even in the case of 2 x 2 tables, the phi coefficient is subject to skewing by severely uneven marginal distributions (the totals in the "margins") such as the following:

```
  | 60 |
-+----+---
  | 52 | 7 |
  | 15 |   |
```

With a 2 x 2 table with severely skewed marginals use Cramer's V (see below).
B. The Contingency Correlation Coefficient formula is:

\[
C = \sqrt{\frac{\phi^2}{\phi^2 + 1}} = \sqrt{\frac{\chi^2}{\chi^2 + N}}
\]

1. This statistic controls for sample size as well as the magnitude of the chi-square.
2. This statistic is appropriate for use with square matrices, i.e., 2 x 2, 3 x 3, 4 x 4, etc. With a square matrix the contingency correlation coefficient has a range of 0 to 1. Thus it is easily interpreted. With a nonsquare matrix, the contingency correlation coefficient does not have a maximum value and is therefore very difficult to interpret.

C. The Cramer's V Coefficient formula is:

\[
V = \frac{\phi^2}{\sqrt{\frac{\phi^2}{S}} = \sqrt{\frac{\chi^2}{(N)(S)}}}
\]

where:

\[
S = \text{The smaller of the number of rows minus 1 and the number of columns minus 1.}
\]

1. This statistic controls for sample size as well as the number of levels in your variables. It is the most versatile statistic of the three.
2. Cramer's V coefficient works for square and nonsquare matrices. In either case it has a range of 0 to 1 and is very easy to interpret.

II. Which cells are responsible for significance? In order to determine which categories of which variables (i.e., which cells in the contingency table) are critical contributors to a significant chi-square statistic we can use two different methods: 1) examine the components of the chi-square and 2) partition the chi-square. While these are often used separately, it makes the most sense to use them together to get the fullest picture of where critical differences are occurring.

A. Examining the components of the chi-square involves looking back to the \[ \frac{(O - E)^2}{E} \] calculations for each cell.

Recall that these components indicate the extent of difference that a particular cell contributes to the overall chi-square, since the chi-square value is merely the sum of these components.
1. Looking back to the example with major and gender (in the previous section of notes, titled "Chi-square") we can identify the components as:

   cell a) \(\frac{(O - E)^2}{E} = 8.4\)

   cell b) \(\frac{(O - E)^2}{E} = 21.5\)

   cell c) \(\frac{(O - E)^2}{E} = 5.4\)

   cell d) \(\frac{(O - E)^2}{E} = 10.4\)

   cell e) \(\frac{(O - E)^2}{E} = 26.9\)

   cell f) \(\frac{(O - E)^2}{E} = 6.7\)

2. From this review we can tell that the largest components are associated with cells b, d, and e.

   We want to determine the extent to which this combination of cells contributes to the overall chi-square, or what percentage of the overall chi-square is due to the contribution of these cells. In order to do that we merely add these components together:

   \[21.5 + 10.4 + 26.9 = 58.8\]

   And then find the percentage of the overall chi-square value represented by this amount:

   \[8.4 + 21.5 + 5.4 + 10.4 + 26.9 + 6.7 = 79.3\]

   \[\frac{58.8}{79.3} = .74\]

   Thus, 3/4ths of the overall chi-square is due to the contribution of these three cells, or we can say that the majority of the significant difference is found in the areas of men in math, women in math, and women in science.

3. When trying to identify components for this analysis try to follow the rule of efficiency, i.e., select the fewest number of cells that collectively contribute the most difference. Usually we want to try to select the fewest number of cells that contribute at least 50 percent of the overall chi-square.

B. Partitioning the chi-square involves dividing the whole chi-square matrix into smaller, independent subtables and calculating component chi-squares for those subtables. A component chi-square is indicated by the symbol \(L^2\).

   1. First, divide the table into independent subtables (in which cells contained in one table are not contained in any other tables) based on theoretical interest, previous identification of components contributing to the overall chi-square, or both (both is better).
2. Calculate component chi-squares on each subtable using the appropriate degrees of freedom. The calculation of the component chi-square is done the exact same way as the calculation of the overall chi-square, except it is on a smaller number of cells and uses different degrees of freedom, depending upon the size and shape of the subtable examined.

3. Using the major and gender example and the results of examining the components suppose we are interested in the following subtable:

<table>
<thead>
<tr>
<th></th>
<th>Science</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>Females</td>
<td>10</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>130</td>
<td></td>
</tr>
</tbody>
</table>

We then calculate the component chi-square:

<table>
<thead>
<tr>
<th></th>
<th>E</th>
<th>(O - E)</th>
<th>(O - E)^2</th>
<th>[(O - E)^2] / E</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>(60/130)(60/130)(130)</td>
<td>27.7</td>
<td>497.3</td>
<td>18.0</td>
</tr>
<tr>
<td>10</td>
<td>(70/130)(60/130)(130)</td>
<td>32.3</td>
<td>497.3</td>
<td>15.4</td>
</tr>
<tr>
<td>10</td>
<td>(60/130)(70/130)(130)</td>
<td>32.3</td>
<td>497.3</td>
<td>15.4</td>
</tr>
<tr>
<td>60</td>
<td>(70/130)(70/130)(130)</td>
<td>37.7</td>
<td>497.3</td>
<td>13.2</td>
</tr>
</tbody>
</table>

\[ L^2 = 62.0 \]

We find that the component chi-square for this subtable is 62.0, which is statistically significant at \( p < .001 \) with 1 degree of freedom. In addition, this component chi-square explains 78 percent of the original chi-square value \((62.0 / 79.3 = .78)\). Thus, we can say that the majority of significant relationship between major and gender is due to the relationship between women and men in science and math such that there are significantly more men in science and more women in math than would be expected.
GUIDE TO COMMON STATISTICAL SYMBOLS

A "statistic" is a characteristic of a sample. There are several common symbols used to represent individual statistics, or elements involved in a statistical manipulation. Below are some of the most common.

\( x \) or \( X \) Usually refers to a score, a datum. When it is used with a subscript (e.g., \( x_1 \)), the subscript refers to the group that the score belongs to.

\( n \) or \( N \) The number of members in a sample or population. In most statistics, \( N \) refers to the total number of subjects, and \( n \) refers to the total number of subjects in a particular group.

\( m_d \) Median.

\( m_o \) Mode.

\( \bar{x} \) Mean of a sample.

\( \mu \) Mean of a population.

\( s \) Standard deviation of a sample.

\( \sigma \) Standard deviation of a population.

\( s^2 \) Variance of a sample.

\( \sigma^2 \) Variance of a population.

\( p \) Probability or "p-value."

\( df \) Degrees of freedom.

\( \alpha \) Type I (alpha) error. The instance of rejecting the null hypothesis when it should be retained.

\( \beta \) Type II (beta) error. The instance of affirming the null hypothesis when it should be rejected.

\( \sum \) "The sum of..." So, any time this symbol appears, it means "add the values that follow." For example, \( \sum x \) means the sum of all the scores for variable \( x \).

\( > \) Greater than.

\( < \) Less than.
$\geq$   Greater than or equal.

$\leq$   Less than or equal.
OVERVIEW OF STATISTICS TOOLS AND TECHNIQUES

I. Descriptive (all univariate)

A. Measures of central tendency (see Babbie pp. 377-380, Williams p. 38, and course notes). Only appropriate for interval or ratio level data

1. **Mean**: Add up all values and divide by number of values.
   - a. The mean is like the balancing point (fulcrum) on which the distribution of values rests
   - b. This value can be strongly influenced by outliers

2. **Median**: rank order all values and then divide the list into two halves: this dividing point is the median (also called the 50th percentile).
   - a. A further division of each half results in the first and third quartiles (25th and 75th percentiles).
   - b. EXAMPLE. SAT and GRE scores are reported in specific percentile values that show a person's rank among all those who have taken the test.

3. **Mode**: rank order all values and identify the one that occurs most frequently
   - a. This is the least often used measure of central tendency

B. Measures of dispersion (see Williams p. 40 and course notes)

1. **Range**: The highest value - the lowest value (plus one for mathematical correctness)
   - a. There are two types of range: the theoretically possible range, and the range actually found in the data.

2. **Variance**: Calculate the difference (deviation) between each value and the mean value. Square each of these deviation scores to remove the impact of the direction (or sign) of the deviations. Now take the mean of these values to find the "average squared deviation."
   - a. Note that to calculate this value one adds up the squared deviation values, producing what is called a "sum of squares."

3. **Standard deviation**: The square root of the variance
   - a. This puts the variance calculation of average squared deviation back on the scale of the data itself and provides just an "average deviation" of the values from the mean value.

4. **Kurtosis**: This term refers to how the values are grouped, or put another way, how the distribution of values is "skewed." (see notes)

II. Inferential (primarily bivariate and multivariate)

A. The logic of inferential statistics: Reasoning from information about a sample of data to population. (see Williams pp. 51-64 and Babbie pp. 430-437)
1. Univariate case: How confident can we be that the sample mean accurately represents the population mean?

2. EXAMPLE. Ask a sample of 100 people how many hours of TV they watch each day. Find the mean value. How confident can we be that this mean value applies to the whole population and not just our sample?

B. T-test (see Williams pp. 83-89 and course notes)
1. Question addressed. Are the means of two different samples different enough from each other that we can be confident that they represent two different populations?

2. EXAMPLE. Some subjects watch a videotaped lecture while others attend the lecture in person. All the subjects complete a questionnaire that measures knowledge gain. How confident can we be that the difference between the means for the two groups represents the fact that the two groups now come from different populations? (This is usually shortened to "How confident can we be that the two groups are significantly different?")

3. Other. There are different formulas for independent groups (between subject designs, as in the example) and matched groups (within subject designs)

C. Analysis of Variance (ANOVA) (see Williams pp. 90-116 and course notes)
1. Question addressed. Just like a t-test but with more than two groups: Are the means of three or more different samples different enough from each other that we can be confident that they represent different populations?

2. EXAMPLE. A study measures the number of violent incidents during films of the 1970s, 1980s, and 1990s. A mean number of violent incidents is calculated for each decade. How confident can we be that the differences among the means for the three decades represent the fact that the decades are significantly different with respect to violence?

3. EXAMPLE. The same analysis would be used if instead of decades the study measured the number of violent incidents observed among groups of subjects who had viewed violent, violent and sexual, and nonviolent movies.

4. One way and multi-way anova. When there is one independent variable, as in the example, the appropriate technique is a one-way anova. When there are two IVs the appropriate technique is a 2-way anova, which evaluates differences among the means of groups corresponding to the levels of one IV (e.g., Type of content (violent, violent and sexual, nonviolent)) and mean differences for groups that correspond to the levels of another IV (e.g., Age of subject (young, medium, old).

5. Omnibus test and planned or post hoc comparisons. The results of an anova tell us only that there are significant differences among the means for the groups in the analysis. They do not tell us which means are different. For this we must either have planned certain mean comparisons (planned comparisons) or be guided by the actual mean values to conduct post hoc comparisons.

6. Multivariate analysis of variance (MANOVA). A multi-way anova involves multiple independent variables. When there is interest in more than one DEPENDENT variable, the appropriate technique is a multivariate anova: Are sets of two or more means found for three or more different samples, different enough from each other that we can be confident that the sample groups represent different populations?
a. EXAMPLE. Subjects watch a pro-health or neutral video and their health-related attitudes, knowledge, and behavior are evaluated. How confident can we be that the differences among the mean attitudes, knowledge and behaviors of the three groups represent the fact that the groups are significantly different with respect to health?

b. Other. Manova is a controversial procedure

7. Analysis of covariance (ANCOVA, MANCOVA)
   a. Question addressed. The question asked by an ancova or mancova is identical to that asked by an anova or manova. However, an analysis of covariance allows a more precise answer. If a continuous independent variable that is not of interest to the researcher and therefore is not being used in an anova or manova, is systematically related to the dependent variable(s) in the analysis, an ancova can be used to remove the effects of that independent variable from the analysis, leaving a more "pure" or "accurate" set of means to be compared.

   b. EXAMPLE. Subjects are given a low, medium, or high dose of a new drug to reduce anxiety. An anova would compare the mean anxiety score for each group. An ancova might include the continuous variable age in the analysis to remove the effect on the scores due to the fact that older patients are known to need larger doses of this type of drug to reach the same level of effectiveness as younger patients. In effect, the ancova then controls for the effects of age on the dependent variable.

D. Correlation (see Williams pp. 131-146 and course notes)
   1. Question addressed. This is a measure of association (co-relation) between two continuous, interval or ratio level variables. It answers the question: How confident can we be that the magnitude and direction of association found in the sample accurately represents the relationship in the population? Another way to think of correlation is: when one variable "wiggles," does the other wiggle, and if so, in what direction?

   2. EXAMPLE. A study measures the amount of television people watch every day and their estimates of the frequency of violent incidents in society. A significant, high, positive correlation (e.g., .6) indicates that when amount of viewing is high, perceived violence in society is also likely to be high, both in the sample and in the general population.

   3. Other.
      a. Remember that correlation or association does not imply causation!
      b. Other types of correlation include multiple, partial, and canonical correlation
      c. The phi statistic to measure strength of association between two dichotomous (2 values) variables

E. Chi-square (see Babbie pp. 437-440, Williams pp. 117-128, and course notes)
   1. Question addressed. When data is nominal (i.e., categorical) correlation can not be used to measure association, and chi-square is used instead. By measuring the discrepancy between the number of cases expected to fall into each of several categories in the event that no association exists, with the number of cases in a sample that actually do fall into each category we can answer the same question as
above: How confident can we be that the magnitude and direction of association found in the sample accurately represents the relationship in the population?

2. EXAMPLE. A one-way chi-square: Do the programs on one television network feature more gender stereotyped characters than the others? We could use chi-square to compare the frequencies actually found for each network with the theoretical frequencies that should occur if there is no difference among the networks, and determine whether a significant difference does exist.

3. EXAMPLE. A two-way chi-square: Is gender associated with skills in different types of tasks? The numbers of males and females in a sample who do well at mathematical, verbal, and analytic tasks could be subjected to chi-square analysis to determine whether an association exists.

4. Log-linear models. Just like chi-square but with more than two variables: How confident can we be that the magnitude and direction of association among three or more variables found in the sample accurately represents the relationships in the population?

F. Regression (see Williams pp. 147-167, Babbie pp. 420-424, and course notes)

1. Question addressed. An extension of correlation; the question here is: What is the mathematical relationship between two variables and how confident can we be that this relationship can accurately predict values on one variable in the population based on the other?

2. EXAMPLE. In the example above, a regression analysis can help us predict the level of societal violence a person will perceive if she watches a given number of hours of television each day. A test of significance will tell us how much confidence to have in this prediction.

3. Other.
   a. Remember that regression is based on correlation and therefore does not permit conclusions concerning causality!
   b. Other forms of regression
      1. Linear regression: Used to predict values of variable Y based only on values of variable X.
      2. Multiple regression: Used to predict values of variable Y based on values of variables X1, X2, etc. (Note that results depend very much on which predictor variables (X1, X2, etc.) are included)
      3. Time-series regression analysis uses regression to predict changes in a dependent variable over time.

G. Path analysis (see Babbie pp. 424-426)

1. Question addressed. An extension of regression; the question here is: What are the associations among a set of three or more variables, and how confident can we be about each of these associations?

2. EXAMPLE. See Babbie p. 442

3. Other. This is not considered a valid technique by many statisticians (correlation is not causation, and that is how this is often interpreted)
H. **Structural equation modeling.**
   1. **Question addressed.** An extension of path analysis designed to account for some of that technique's limitations. The question is the same here, but the degree of measurement error associated with each variable in the model is included in the analysis, producing a more "believable" model.
   2. **EXAMPLE.** See Computer As Social Actor study
   3. **Other.** This is a very complex statistical analysis, still somewhat controversial in communication research.

I. **Factor analysis** (see Babbie pp. 427-430 and Williams pp. 176-194)
   1. **Question addressed.** An extension of correlation; the question here is: What dimensions (factors) can be revealed by the intercorrelations among a large set of measures? (And, as usual, how confident can we be that these dimensions exist in the general population as well?)
   2. **EXAMPLE.** See Williams p. 177 and Babbie pp. 428-429.

J. **Cluster analysis** (see Williams pp. 171-175).
   1. **Question addressed.** Another extension of correlation; the question here is: Can we use the responses of a large group of people to a set of measures to separate the group of people into relatively homogeneous subgroups?
   2. **EXAMPLE.** A survey might contain numerous questions concerning media use. The researcher might use cluster analysis to divide the respondents into subgroups whose media use habits are similar (e.g., not much TV or film but lots of newspapers and books). These groups could then be analyzed separately to determine the causes or effects of such patterns.

K. **Discriminant analysis** (see Williams pp. 195-204).
   1. **Question addressed.** This is the opposite of cluster analysis in the sense that the researcher here divides the respondents or subjects into groups based on an attribute she considers important (e.g., high and low TV viewers, successful and unsuccessful teachers, etc.); the question then is: Can we uncover patterns of responses on a large set of measures that define or discriminate between the predetermined groups. (The patterns can also be used to classify new respondents or subjects.)
   2. **EXAMPLE.** A health study reveals two groups of respondents: those who are very concerned about their health and those who are not. We might perform a discriminant analysis on these two groups to see what distinguishes or discriminates one from the other.

L. **Multidimensional Scaling**
   1. **Question addressed.** Based on a set of scores that represent the degree of perceived similarity or difference between pairs of objects or entities, can we develop a spatial representation of the relationships among all the objects or entities?
   2. **EXAMPLE.** See Computer As Social Actor survey
CHOOSING THE RIGHT INFERENTIAL STATISTICS TECHNIQUE

The overview of major inferential statistical techniques below is organized not based on a list of the techniques but based on the goal of the researcher preparing to analyze specific types of data. Note that the researcher should decide before the data is collected which technique(s) she will eventually use.

DIFFERENCES AMONG MEANS:

I want to compare the means of 2 groups on 1 interval or ratio level variable (DV): I use T-TEST

I want to compare the means of 3 or more groups (levels of 1 IV) on 1 interval or ratio level variable (DV): I use 1 WAY ANOVA

I want to compare the means of 3 or more groups (levels of 2 or more IV) on 1 interval or ratio level variable (DV): I use 2 WAY ANOVA

I want to compare the means of 3 or more groups (levels of 1 or more IV) on 2 or more interval or ratio level variables (DV): I use MANOVA

ASSOCIATION AMONG VARIABLES:

I want to know whether there is an association (a "moving together") between 2 interval or ratio level variables: I use CORRELATION

I want to know whether there is an association (a "moving together") between 2 sets of interval or ratio level variables: I use CANONICAL CORRELATION

I want to know whether there is an association (a "moving together") between 2 interval or ratio level variables, with the value of 1 or more additional interval or ratio level variable(s) held constant: I use PARTIAL CORRELATION

I want to know whether there is an association (a "moving together") between 2 nominal or ordinal level variables: I use CHI-SQUARE (Also: I want to know if there is a significant pattern among the values of 1 nominal or ordinal variable)

I want to know whether there is an association (a "moving together") between 2 dichotomous nominal or ordinal level variables: I use a PHI COEFFICIENT

I want to know whether there is an association (a "moving together") between 3 or more nominal or ordinal level variables: I use a LOG-LINEAR MODEL
I want to predict the values of 1 interval or ratio level variable based on the values of another interval or ratio level variable: I use LINEAR REGRESSION

I want to predict the values of 1 interval or ratio level variable based on the values of 2 or more other interval or ratio level variables: I use MULTIPLE REGRESSION (Also: I want to predict the values of 1 interval or ratio level variable based on the values of 2 or more nominal or ordinal level variables: use dummy coding)

I want to know how 3 or more variables are interrelated so I can use the values on some of them to predict the values on others: I use a PATH MODEL or STRUCTURAL EQUATION MODELING (the latter is more complex but potentially more accurate; both are often misinterpreted as revealing causality)

I want to know what dimension or dimensions underlie the patterns of values on 3 or more interval or ratio level variables: I use FACTOR ANALYSIS

I want to separate a group of people into relatively homogeneous subgroups based on their responses to a set of measures: I use CLUSTER ANALYSIS

I have divided a group of people into relatively homogeneous subgroups based on their responses to one or more measures that I consider important and now I want to discriminate between the different groups based on their responses to a separate set of measures (I may then want to use responses to these measures to classify the members of a new group of people): I use DISCRIMINANT ANALYSIS

I have data that quantify the differences or similarities between all pairs of elements in a set of 3 or more objects, places, events, people, etc. and I want to develop a spatial representation of the relationships among all of the elements: I use MULTIDIMENSIONAL SCALING (MDS)
SUGGESTED REFERENCES
(Special thanks to Professor Tricia Jones)

GENERAL RESEARCH METHODS:


QUALITATIVE/INTERPRETIVE METHODS:


**QUANTITATIVE/QUALITATIVE RESEARCH:**


**LITERATURE REVIEW AIDS:**

