February 14, 2005

MMC 503
Advanced Quantitative Research Methods

Spring 2005
Matthew Lombard

Supplementary Course Materials
(Note: All materials here were originally in the supplementary course materials for MMC 500/BTMM 411

Contents

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):

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

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.

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.

Sampling Type	Purpose
A. Random sampling	Avoids systematic bias in the sample; large sample size is important for making generalizations.
1. Simple random sample	Achieve a representative sample that permits generalizations to the whole population.
2. Stratified random and cluster samples	Increase confidence in making generalizations to particular subgroups or areas.
B. Purposeful sampling	Increase the utility of information obtained from small samples; sampling criteria based on specific desired attributes.
1. Sampling extreme or deviant cases	Provide information about unusual cases that may be particularly enlightening
2. Sampling typical case(s)	Avoid studying an entity where the results would be dismissed outright because THAT entity is known to be special, deviant, unusual, extreme, etc.
3. Maximum variation sampling - picking three or four cases that represent a range on some dimension	Increase confidence in common patterns that cut across different entities.
4. Sampling critical cases	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.
5. Sampling politically important or sensitive cases	Attracts attention to the study (or avoids attracting undesired attention by purposefully eliminating from the sample politically sensitive cases).
6. Convenience sampling -- take the easy cases	Saves time, money, and effort.

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. <u>The classical view of measurement error</u>. 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. <u>Predictive validity</u>: 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. <u>Content validity</u>: 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. <u>Construct validity</u>: 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?

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").

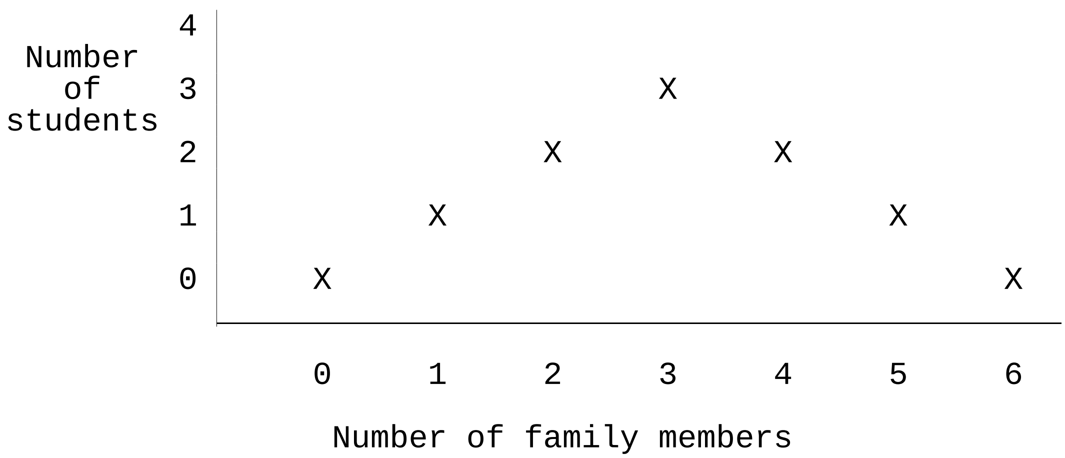
V. A summary of measurement scales, their characteristics and their statistical implications follows.

Characteristics of the scale	Statistical possibilities
Nominal measurement scale	
A scale which "measures" 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.
Ordinal measurement scale	
A scale which "measures" in terms of such values as "more" or "less," "larger" or "smaller," but without specifying the size of the intervals.	Can be used for determining the mode, percentage, chi-square, median, percentile rank, or rank correlation.
Interval measurement scale	
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.
Ratio measurement scale	
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.

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 bellshaped 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:



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. <u>Mode</u>. 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 bellshaped 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:

$$x = \frac{\text{sum of scores}}{\text{number of scores}}$$

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

$$\bar{x} = \frac{\text{sum of scores}}{\text{number of scores}} = \frac{3 + 4 + 5 + 6 + 7}{5} = \frac{25}{5} = 5$$

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:

```
Range  =  (high value - low value) + 1
```

B. EXAMPLE: What is the range of the following data set: 3, 2, 4, 6, 1, 1, 8, 9?

```
Range  =  (high value - 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):

Mesokurtic ("normal")	Platykurtic ("flat")	Leptokurtic ("tall and skinny")

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.

- Positive skew + ("tail on right")	- Negative skew + ("tail on left")

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 of 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:

$$\bar{X} = \frac{(72 + 81 + 86 + 69 + 57)}{5} = \frac{365}{5} = 73$$

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

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

$$\begin{aligned}
(x - \bar{X}) &= (72 - 73) = -1 \\
(x - \bar{X}) &= (81 - 73) = 8 \\
(x - \bar{X}) &= (86 - 73) = 13 \\
(x - \bar{X}) &= (69 - 73) = -4 \\
(x - \bar{X}) &= (57 - 73) = -16
\end{aligned}$$

3. Then square the deviations; i.e.,

$(x - \bar{X})^2$.

$$\begin{aligned}
(x - \bar{X})^2 &= (-1 * -1) = 1 \\
(x - \bar{X})^2 &= (8 * 8) = 64 \\
(x - \bar{X})^2 &= (13 * 13) = 169 \\
(x - \bar{X})^2 &= (-4 * -4) = 16 \\
(x - \bar{X})^2 &= (-16 * -16) = 256
\end{aligned}$$

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).

$$\frac{\Sigma\ ((x - \bar{X})^2)}{N - 1} = \frac{506}{(5 - 1)} = \frac{506}{4} = 126.5$$

6. Finally take the square root of this figure.

$$S = \sqrt{\frac{\Sigma\ ((x - \bar{X})^2)}{N - 1}} = \sqrt{126.5} = 11.24$$

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)	$(x - \bar{X})$	$(x - \bar{X})^2$
3		
6		
7		
12		
13		
16		

$\Sigma(x - \bar{X})$ = ____ $\Sigma(x - \bar{X})^2$ = ____

$\bar{X}$ =

N - 1 =

Standard deviation (S) =

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) =

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:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

where:

$\sigma_{\bar{x}}$ = Standard error of the mean

σ = Standard deviation of values in the population

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:

σ_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:
μ = 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_o: $\mu <= 40$

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 (S) 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:

$$\sigma_{\bar{x}} = \frac{S}{\sqrt{n}}$$

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

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 α) 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 <u>greater than</u> 40. If the hypotheses had been:

H_r: μ < 40

H_o: μ >= 40

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

H_r: μ does not equal 40

H_o: μ 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 (α). 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 (ß). 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 <u>any</u> 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

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

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?

Group 1	Group 2
1	3
2	4
3	5
4	6
5	7
$\Sigma(x) = 15$	$\Sigma(x) = 25$

$$\bar{X}_1 = \frac{\Sigma(x)}{n_1} = \frac{15}{5} = 3 \qquad \bar{X}_2 = \frac{\Sigma(x)}{n_2} = \frac{25}{5} = 5$$

Group 1	$(x - \bar{X}_1)$	$(x - \bar{X}_1)^2$	Group 2	$(x - \bar{X}_2)$	$(x - \bar{X}_2)^2$
1	1 - 3 = -2	4	3	3 - 5 = -2	4
2	2 - 3 = -1	1	4	4 - 5 = -1	1
3	3 - 3 = 0	0	5	5 - 5 = 0	0
4	4 - 3 = 1	1	6	6 - 5 = 1	1
5	5 - 3 = 2	4	7	7 - 5 = 2	4

$$\Sigma(x - \bar{X}_1)^2 = 10 \qquad \Sigma(x - \bar{X}_2)^2 = 10$$

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

$$= \frac{(\ 10\ ) + (\ 10\ )}{5 + 5 - 2} = \frac{20}{8} = 2.5$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} = \frac{3 - 5}{\sqrt{\frac{2.5}{5} + \frac{2.5}{5}}} = \frac{-2}{\sqrt{1}} = \frac{-2}{1} = -2$$

The degrees of freedom for this problem are

$$n_1 + n_2 - 2 = 5 + 5 - 2 = 8$$

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 $\underline{p} < .1$, 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 $\underline{p} < .05$), 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 $\underline{p} < .05$ as your criterion for statistical significance.

Group 1	$(x - \bar{X}_1)$	$(x - \bar{X}_1)^2$	Group 2	$(x - \bar{X}_2)$	$(x - \bar{X}_2)^2$
12			14		
13			16		
15			18		
16			19		
18			20		

$\Sigma(x - \bar{X}_1)^2 =$ ____ $\Sigma(x - \bar{X}_2)^2 =$ ____

$\bar{X}_1 =$ $\bar{X}_2 =$

$s^2 =$

$t =$

$df =$

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 = 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:

Between group variance is greater than within group variance	Between group variance is the same as within group variance

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:

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

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:

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

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

      $$df_1 = k - 1$$

      $$df_2 = N - k$$

      Note that df_1 is the same as the denominator in the formula for s_b^2, and 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, df_1 is sometimes called $df_{numerator}$ and df_2 is sometimes called $df_{denominator}$.

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.

Group 1	Group 2	Group 3
1	2	3
2	3	4
3	4	5
4	5	6
5	6	7
15	20	25

$$\bar{x}_1 = \frac{15}{5} = 3 \qquad \bar{x}_2 = \frac{20}{5} = 4 \qquad \bar{x}_3 = \frac{25}{5} = 5$$

$$\bar{X} = \frac{15 + 20 + 25}{5 + 5 + 5} = \frac{60}{15} = 4$$

OR THE MEAN OF MEANS:

$$\bar{X} = \frac{3 + 4 + 5}{3} = 4$$

$$n_1 = n_2 = n_3 = 5$$

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

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

$$= \frac{5\,((3 - 4)^2) + 5\,((4 - 4)^2) + 5\,((5 - 4)^2)}{(3 - 1)}$$

$$= \frac{5 + 5}{2} = 5$$

B. Next we find s_w^2:

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.

Group 1		Group 2		Group 3	
$(x_{i1} - \bar{x}_1)$	$(^2)$	$(x_{i2} - \bar{x}_2)$	$(^2)$	$(x_{i3} - \bar{x}_3)$	$(^2)$
1 - 3 = -2	=4	2 - 4 = -2	=4	3 - 5 = -2	=4
2 - 3 = -1	=1	3 - 4 = -1	=1	4 - 5 = -1	=1
3 - 3 = 0	=0	4 - 4 = 0	=0	5 - 5 = 0	=0
4 - 3 = 1	=1	5 - 4 = 1	=1	6 - 5 = 1	=1
5 - 3 = 2	=4	6 - 4 = 2	=4	7 - 5 = 2	=4

$N = 5 + 5 + 5 = 15$

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

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

$$= \frac{\begin{array}{l}(((1 - 3)^2) + ((2 - 3)^2) + ((3 - 3)^2) + \\ ((4 - 3)^2) + ((5 - 3)^2) + \\ ((2 - 4)^2) + ((3 - 4)^2) + ((4 - 4)^2) + \\ ((5 - 4)^2) + ((6 - 4)^2) + \\ ((3 - 5)^2) + ((4 - 5)^2) + ((5 - 5)^2) + \\ ((6 - 5)^2) + ((7 - 5)^2))\end{array}}{(N - k)}$$

$$= \frac{(4 + 1 + 0 + 1 + 4 + 4 + 1 + 0 + 1 + 4 + 4 + 1 + 0 + 1 + 4)}{15 - 3}$$

$$= \frac{30}{12} = 2.5$$

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 can not 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:

$$\Sigma \frac{((x - \overline{X})^2)}{N - 1}$$

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

$$\frac{\dfrac{\text{Between group sum of squares}}{\text{deg. of freedom}_1}}{\dfrac{\text{Within group sum of squares}}{\text{deg. of freedom}_2}} = \frac{\dfrac{SSB}{df_1}}{\dfrac{SSW}{df_2}} = \frac{S_b^2}{S_w^2} = \frac{MSB}{MSW}$$

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 <u>between</u> 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 <u>within</u> 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:

Source of variance	Sum of squares	Degrees of freedom	Mean squares	F	p-value
Between group (The IV)	SSB	k - 1	$S_b^2 = \dfrac{SSB}{df_1}$	$\dfrac{S_b^2}{S_w^2}$	p
Within group (Error)	SSW	N - k	$S_w^2 = \dfrac{SSW}{df_2}$		
Totals	SST	N - 1			

C. The short cut calculation of an F statistic is this:

$$SST = \left(\sum_{ij} (x_{ij}^2)\right) - \frac{\left(\sum_{ij} (x_{ij})\right)^2}{N}$$

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

$SSW = SST - SSB$

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:

	Lecture	Computer	Combination
	1	3	9
	5	6	10
	1	7	8
	3	2	4
TOTAL	10	18	31

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

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

Lecture	Computer	Combination
1	9	81
25	36	100
1	49	64
9	4	16

TOTAL: 1 + 25 + 1 + 9 + 9 + 36 + ... + 16 = 395

b. Take the grand total and square it; i.e.,

$$\left(\sum_{ij} (x_{ij})\right)^2$$

59 x 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

$$\left(\frac{T_i^2}{n_i}\right),$$

and then sum them

$$\sum_i \left(\frac{T_i^2}{n_i}\right)$$

$$SSB = \frac{(10)^2}{4} + \frac{(18)^2}{4} + \frac{(31)^2}{4}$$

$$= \frac{100}{4} + \frac{324}{4} + \frac{961}{4} = \frac{1385}{4} = 346.25$$

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.

$$MSB = \frac{SSB}{df_1} = \frac{56.1667}{2} = 28.0833$$

$$MSW = \frac{SSW}{df_2} = \frac{48.7500}{9} = 5.4166$$

6. The F statistic

$$F = \frac{MSB}{MSW} = \frac{28.0833}{5.4166} = 5.18$$

7. The ANOVA table:

Source of variance	Sum of squares	Degrees of freedom	Mean squares	F	p-value
Between group (The IV)	56.1667	2	28.0833	5.18	
Within group (Error)	48.7500	9	5.4166		
Total	104.9167	11			

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

High	Moderate	Low
5	8	6
6	8	2
3	10	1
1	3	6
7	9	2

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).

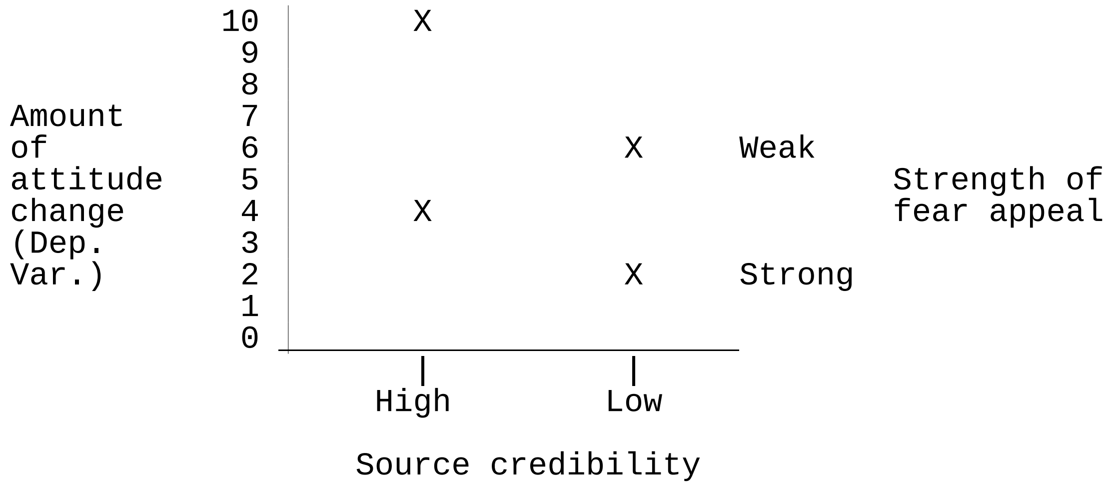
B. Interactions. An interaction between two independent variables (IVs) means that the impact on a dependent variable (DV) of changing the value of one IV is influenced by the value of the other IV. Rather than having separate, "independent" impacts on a dependent variable, the two variables "interact" 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
Amount         7
of             6                X     Weak
attitude       5                               Strength of
change         4      X                        fear appeal
(Dep.          3
Var.)          2                X     Strong
               1
               0
                      |         |
                     High      Low

                   Source credibility
```

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:



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:

$$SST = \left(\sum_{ij}(x_{ij}^2)\right) - \frac{\left(\sum_{ij}(x_{ij})\right)^2}{N}$$

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

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

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

$SSE = SST - SSA - SSB - SSAB$

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)

		Lecture	Computer	Combin-ation
Gender (B)	Female	1	3	7
		2	4	9
		1	5	10
		2	3	9
	Male	2	6	6
		1	4	6
		1	7	5
		2	8	7
TOTALS		12	40	59

B. Computations:

1. Sum of squares total (SST)

   a. Square all the raw scores and sum them.

MATRIX FOR RAW SCORES SQUARED

Method (A)

		Lecture	Computer	Combin-ation
Gender (B)	Female	1	9	49
		4	16	81
		1	25	100
		4	9	81
	Male	4	36	36
		1	16	36
		1	49	25
		4	64	49
TOTALS		20	224	457

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 x 2 x 4 = 24).

Grand total = 1 + 2 + 1 + 2 + ... + 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:

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

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 <u>visually</u> represent findings.

MATRIX OF CELL MEANS

	Lecture	Computer	Combination
Female	1.50	3.75	8.75
Male	1.50	6.25	6.00

```
9                               X        Female
8
7
6                X              X        Male
5
4                X
3
2    X
1
  ------------------------------------
   LECTURE    COMPUTER    COMBINATION
```

```
9        X
8
7
6                  X     Computer
5                        Combination
4        X
3
2        X         X     Lecture
1
  -------------------------
       FEMALE     MALE
```

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

DATA MATRIX

		Gender: Female	Gender: Male
Self-disclosure	High	3 4 2 5 3	5 5 6 7 5
	Medium	2 1 3 6 1	7 5 7 8 11
	Low	4 5 6 5 9	4 3 2 5 4

Calculate a two-way ANOVA and fill in the appropriate ANOVA table:

Source of variance	Sum of squares	Degrees of freedom	Mean squares	F	p-value
(A)					
(B)					
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:

$$l = 1 \times (\bar{X}_1) - 1 \times (\bar{X}_2) \quad \text{(Lecture versus Computer)}$$

$$l = 1 \times (\bar{X}_2) - 1 \times (\bar{X}_3) \quad \text{(Computer versus Combination)}$$

$$l = 1 \times (\bar{X}_1) - 1 \times (\bar{X}_3) \quad \text{(Lecture versus Combination)}$$

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:

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

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:
α = Probability of making a Type I error on an individual test

c = Total number of comparisons being made

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

c, number of comparisons	α, probability of Type I error on an individual test: .1	.05	.01
1	.100	.050	.010
2	.190	.097	.020
3	.271	.143	.030
4	.344	.185	.039
5	.410	.226	.049
.	.	.	.
.	.	.	.
10	.651	.401	.096

[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 $p < .05$ 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.

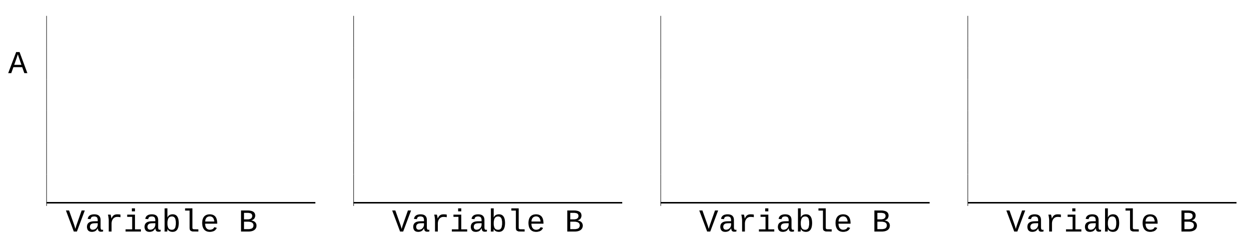
I. Interpreting correlations. Correlation coefficients (r) range from -1 to +1. The coefficient tells the researcher two things: 1) the magnitude of the relationship between the variables, and 2) the direction of the relationship.

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:

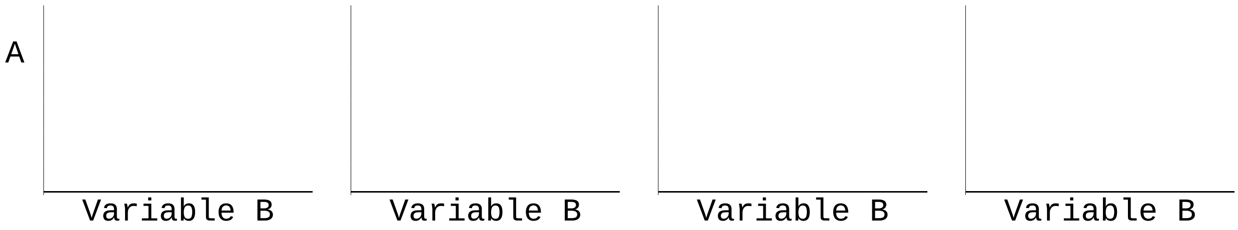
+/- .01 - .25	Weak relationship
+/- .26 - .55	Moderate relationship
+/- .56 - .75	Strong relationship
+/- .76 - .99	Very strong relationship
+/- 1.00	Perfect relationship

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:



r = + .80 r = - .80 r = + 0.0 r = + 1.0



r = - 1.0   r = + 0.1   r = + 0.1   r = + 0.1

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:

$$r = \frac{\Sigma\left((x_i - \bar{X}) * (y_i - \bar{Y}) \right)}{\sqrt{\Sigma\left((x_i - \bar{X})^2\right) * \Sigma\left((y_i - \bar{Y})^2\right)}}$$

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).

Subject	x_i	$(x_i - \bar{X})$	$(x_i - \bar{X})^2$	y_i	$(y_i - \bar{Y})$	$(y_i - \bar{Y})^2$	$(x_i - \bar{X})(y_i - \bar{Y})$
1	9	9 - 7 = 2	4	2	2 - 2 = 0	0	(2)(0) = 0
2	8	8 - 7 = 1	1	3	3 - 2 = 1	1	(1)(1) = 1
3	9	9 - 7 = 2	3	2	2 - 2 = 0	0	(2)(0) = 0
4	7	7 - 7 = 0	0	1	1 - 2 = -1	1	(0)(-1) = 0
5	6	6 - 7 = -1	1	2	2 - 2 = 0	0	(-1)(0) = 0
6	4	4 - 7 = -3	9	1	1 - 2 = -1	1	(-3)(-1) = 3
7	5	5 - 7 = -2	4	1	1 - 2 = -1	1	(-2)(-1) = 2
8	6	6 - 7 = -1	1	2	2 - 2 = 0	0	(-1)(0) = 0
9	8	8 - 7 = 1	1	3	3 - 2 = 1	1	(1)(1) = 1
10	8	8 - 7 = 1	1	3	3 - 2 = 1	1	(1)(1) = 1
	70		26	20		6	8

A. The first thing we need to do is calculate the mean for each measure.

$$\bar{X} = \frac{70}{10} = 7 \qquad \bar{Y} = \frac{20}{10} = 2$$

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

$(x_i - \bar{X})$ and $(y_i - \bar{Y})$

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

$(x_i - \bar{X}^2)$ and $(y_i - \bar{Y})^2$

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

$\Sigma(x_i - \bar{X})^2 = 26$ and $\Sigma(y_i - \bar{Y})^2 = 6$

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:

$$r = \frac{\Sigma\left((x_i - \bar{X}) * (y_i - \bar{Y}) \right)}{\sqrt{\Sigma\left((x_i - \bar{X})^2\right) * \Sigma\left((y_i - \bar{Y})^2\right)}} = \frac{8}{\sqrt{(26)(6)}}$$

$$= \frac{8}{\sqrt{156}} = \frac{8}{12.48} = .64$$

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:

$$t = \frac{r\sqrt{N - 2}}{\sqrt{1 - r^2}} = \frac{.64\sqrt{10 - 2}}{\sqrt{1 - .64^2}} = \frac{.64\sqrt{8}}{\sqrt{1 - .59}}$$

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

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.

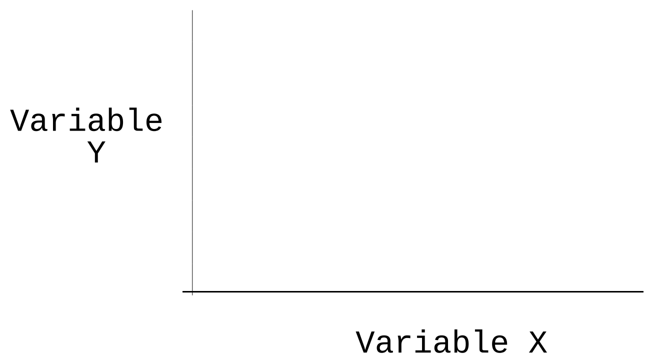
V. EXERCISE: What is the correlation between the following two sets of scores? Is the correlation coefficient statistically significant at $p < .05$?

Subject	x_i	$(x_i - \bar{X})$	$(x_i - \bar{X})^2$	y_i	$(y_i - \bar{Y})$	$(y_i - \bar{Y})^2$	$(x_i - \bar{X})(y_i - \bar{Y})$
1	2			4			
2	3			6			
3	6			5			
4	8			4			
5	4			7			
6	5			8			
7	6			4			
8	7			3			
9	9			2			
10	9			2			

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.



A. The regression line. In order to use the knowledge about the strength and direction of this relationship to predict values of Y given known values of X, the regression procedure fits a line, called a regression line, over this data. The goal in fitting such a line is to minimize the distance of each point from the line. For a given value of X, the best prediction for the associated value of Y can then be found on this line. A mathematical model used to describe the regression line is:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

where:

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

β_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)

β_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 ith score or observation)

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

ε_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 ($ß_1$) is closely related to r, the correlation of the two variables.

C. The error term. The existence of random error (ε_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 (SST) equals the sum of between group sum of squares (SSB) and <u>w</u>ithin group sum of squares (SSW or SSE), the total variation in the variable Y is (y_i - 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-hat_i - Y). Then there is the differences due to random error, between the true value of Y_i and the predicted value of y-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, 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 (χ^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:

χ^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).

<u>NOTE</u>: 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:

	Major			
	Science	Math	English	TOTAL
Number of students:	13	7	10	30

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 $(O - E)$.

Science	Math	English
$(13 - 10) = 3$	$(7 - 10) = -3$	$(10 - 10) = 0$

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

Science	Math	English
$3^2 = 9$	$-3^2 = 9$	$0^2 = 0$

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

Science	Math	English
$9/10 = .9$	$9/10 = .9$	$0/10 = 0$

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:

		Major			
		Science	Math	English	TOTAL
Gender	Male	a 50	b 10	c 40	100
	Female	d 10	e 60	f 10	80
TOTAL		60	70	50	180

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  =  (100/180) x (60/180) x (180)  =  33.3
b) E  =  (100/180) x (70/180) x (180)  =  38.9
c) E  =  (100/180) x (50/180) x (180)  =  27.8
d) E  =  (80/180)  x (60/180) x (180)  =  26.7
e) E  =  (80/180)  x (70/180) x (180)  =  31.1
f) E  =  (80/180)  x (50/180) x (180)  =  22.2
```

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

```
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
```

C. Next, we square the (O - E) figures.

a) $(O - E)^2 = 16.7^2 = 278.9$
b) $(O - E)^2 = -28.9^2 = 835.2$
c) $(O - E)^2 = 12.2^2 = 148.8$
d) $(O - E)^2 = -16.7^2 = 278.9$
e) $(O - E)^2 = 28.9^2 = 835.2$
f) $(O - E)^2 = -12.2^2 = 148.8$

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

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

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?

		Hair color			
		Red	Black	Blonde	TOTAL
Gender	Male	a 30	b 60	c 20	110
	Female	d 10	e 30	f 60	100
TOTAL		40	90	80	210

CELL	O	E	(O - E)	$(O - E)^2$	$(O - E)^2 / E$
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 (ϕ) 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
		7
52	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 O 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:

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

where:
S = 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 O 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-sqaure") we can identify the components as:

cell a) $(O - E)^2 / E = 8.4$
cell b) $(O - E)^2 / E = 21.5$
cell c) $(O - E)^2 / E = 5.4$
cell d) $(O - E)^2 / E = 10.4$
cell e) $(O - E)^2 / E = 26.9$
cell f) $(O - E)^2 / E = 6.7$

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:

$8.4 + 21.5 + 5.4 + 10.4 + 26.9 + 6.7 = 79.3$

$58.8 / 79.3 = .74$

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:

	Science	Math	
Males	50	10	60
Females	10	60	70
	60	70	130

We then calculate the component chi-square:

O		E	(O - E)	$(O - E)^2$	$\frac{(O - E)^2}{E}$
50	(60/130)(60/130)(130) =	27.7	22.3	497.3	18.0
10	(70/130)(60/130)(130) =	32.3	-22.3	497.3	15.4
10	(60/130)(70/130)(130) =	32.3	-22.3	497.3	15.4
60	(70/130)(70/130)(130) =	37.7	22.3	497.3	13.2
				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.
$\overline{X}$	Mean of a sample.
μ	Mean of a population.
S	Standard deviation of a sample.
σ	Standard deviation of a population.
S^2	Variance of a sample.
σ^2	Variance of a population.
$\underline{p}$	Probability or "p-value."
df	Degrees of freedom.
α	Type I (alpha) error. The instance of rejecting the null hypothesis when it should be retained.
ß	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.
$>=$	Greater than or equal.
$<=$	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 X_1, X_2, etc. (Note that results depend very much on which predictor variables (X_1, X_2, 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. <u>Multidimensional Scaling</u>

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)

APPENDIX
SUGGESTED REFERENCES
(Special thanks to Professor Tricia Jones)

QUANTITATIVE/QUALITATIVE RESEARCH:

Brewer, J., & Hunter, A. (1989). *Multimethod research: A synthesis of styles*. Beverly Hills, CA: Sage.

Brinberg, D., & McGrath, J. (1985). *Validity and the research process*. Beverly Hills, CA: Sage.

Budd, R. W., & Thorp, R. K. (1963). *An introduction to content analysis*. ??, Iowa: University of Iowa School of Journalism.

Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand-McNally.

Chaffee, S. H. (1991). *Communication Concepts 1: Explication*. Newbury Park, CA: Sage.

Converse, J. M., & Presser, S. (1987). *Survey questions: Handcrafting the standardized questionnaire*. Beverly Hills, CA: Sage.

Cook, T., & Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton-Mifflin.

Fowler, F. J., Jr. (1984). *Survey research methods*. Beverly Hills, CA: Sage.

Kennedy, J. J. (1978). *An introduction to the design and analysis of experiments in education and psychology*. ??: University Press of America.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills, CA: Sage.

Lavrakas, P. (1987). *Telephone survey methods: Sampling, selection, and supervision*. Beverly Hills, CA: Sage.

Long, J. (Ed.). (l988). *Common problems/proper solutions: Avoiding error in quantitative research*. Beverly Hills, CA: Sage.

Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York: McGraw Hill.

Tardy, C. (Ed.). (1988). *A handbook for the study of human communication: Methods and instruments for observing, measuring and assessing communication processes*. Norwood, NJ: Ablex.

Stempel, G. H., & Westley, B. H. (Eds.) (1989). *Research methods in mass communication*. Englewood Cliffs, NJ: Prentice-Hall.

SPSS DATA ANALYSIS PROGRAMS

1. Descriptive statistics:

```
TITLE     MMC 500 - DESCRIPTIVE STATISTICS EXAMPLE
SUBTITLE  SPRING 2000  LOMBARD [CHANGE THIS TO *YOUR* NAME!!]

/*  ANYTHING THAT FOLLOWS "/*" AT THE BEGINNING OF A LINE
/*  IS IGNORED BY THE COMPUTER.  IT IS INCLUDED HERE TO DESCRIBE
/*  TO THE READER WHAT IS GOING ON.

/*  RESTRICT OUTPUT TO 80 CHARACTERS FOR EASY PRINTING
/*  AND READING
SET WIDTH=80

/*  TELL PROGRAM WHAT COLUMN OF DATA BELOW REPRESENTS
DATA LIST
  /  GENDER   1
     AGE      3-4

/*  TELL PROGRAM WHAT THE SHORT NAMES OF THE VARIABLES STAND FOR
VARIABLE LABELS
     GENDER   'GENDER OF PERSON IN GROUP'
     AGE      'AGE OF PERSON IN GROUP'

/*  TELL PROGRAM WHAT EACH VALUE ON EACH VARIABLE STANDS FOR
VALUE LABELS
     GENDER   1  'MALE'
              2  'FEMALE'
/*   AGE IS SELF EXPLANATORY!

/*  GIVE PROGRAM ACTUAL DATA TO BE ANALYZED BY THE COMMANDS ABOVE
/*  THE COLUMNS ARE GENDER AND AGE
BEGIN DATA
1 25
1 22
1 33
1 35
1 29
1 30
1 22
1 11
1 38
1 22
2 23
2 27
2 35
2 29
2 38
2 23
2 28
2 40
2 29
2 31
```

```
END DATA

/*  TELL PROGRAM TO PROVIDE DESCRIPTIVE STATISTICS FOR BOTH
    VARIABLES
DESCRIPTIVES  VARIABLES=ALL

/*  TELL PROGRAM TO GIVE FREQUENCIES AND HISTOGRAMS OF THE DATA
FREQUENCIES   VARIABLES=ALL
   /HISTOGRAM

/*  TELL THE PROGRAM WE'RE DONE!!
FINISH
```

2. T-test example:

```
TITLE     MMC 500 -- T-TEST EXAMPLE
SUBTITLE  SPRING 2000  LOMBARD [CHANGE THIS TO *YOUR* NAME!!]

/*  ANYTHING THAT FOLLOWS /* AT BEGINNING OF LINE IS IGNORED BY
/*  THE COMPUTER.  IT IS INCLUDED TO DESCRIBE TO READER WHAT
/*  IS GOING ON.

/*  RESTRICT OUTPUT TO 80 CHARACTERS FOR EASY PRINTING
/*  AND READING
SET WIDTH=80

/*  TELL PROGRAM WHAT COLUMNS OF DATA BELOW REPRESENT
DATA LIST
  /  GENDER   1
     SCORE    3

/*  TELL PROGRAM WHAT THE SHORT NAMES OF THE VARIABLES STAND FOR
VARIABLE LABELS
     GENDER   'GENDER OF GROUP MEMBERS'
     SCORE    'SCORE ON SOME TEST OR MEASURE'

/*  TELL PROGRAM WHAT EACH VALUE OF EACH VARIABLE STANDS FOR
VALUE LABELS
     GENDER   1  'MALE'
              2  'FEMALE'

/*  TELL PROGRAM TO CALCULATE THE T-TEST ON THE TWO GROUPS (MALES
/*  AND FEMALES) BASED ON THE MEAN SCORES IN EACH GROUP
T-TEST
    GROUPS=GENDER(1,2)
    /VARIABLES=SCORE

/*  GIVE PROGRAM THE DATA TO BE ANALYZED BY THE COMMANDS ABOVE
BEGIN DATA
1 1
1 2
1 3
1 4
1 5
2 3
2 4
2 5
2 6
2 7
END DATA

/*  TELL THE PROGRAM WE'RE DONE!!
FINISH
```

3. One-way ANOVA example:

```
TITLE     MMC 500 -- ONE-WAY ANOVA EXAMPLE
SUBTITLE  SPRING 2000  LOMBARD [CHANGE THIS TO *YOUR* NAME!!]

/*  RESTRICT OUTPUT TO 80 CHARACTERS FOR EASY PRINTING
/*  AND READING
SET WIDTH=80

/*  TELL PROGRAM WHAT COLUMNS OF DATA BELOW REPRESENT
DATA LIST
  /  INSTRCT  1
     SALES    3-4

/*  TELL PROGRAM WHAT THE SHORT NAMES OF THE VARIABLES STAND FOR
VARIABLE LABELS
     INSTRCT  'TYPE OF INSTRUCTION'
     SALES    'NUMBER OF SALES'

/*  TELL PROGRAM WHAT EACH VALUE ON EACH VARIABLE STANDS FOR
VALUE LABELS
     INSTRCT  1  'LECTURE'
              2  'COMPUTER'
              3  'COMBO'

/*  TELL PROGRAM TO CALCULATE A ONE-WAY ANOVA ON THE DISTRIBUTION
/*  OF ALL RESPONDENTS WITH RESPECT TO INSTRCT
ONEWAY  SALES BY INSTRCT(1,3)

/*  GIVE PROGRAM ACTUAL DATA TO BE ANALYZED BY THE COMMANDS ABOVE
/*  THE COLUMNS ARE INSTRCT, SALES
BEGIN DATA
1  1
1  5
1  1
1  3
2  3
2  6
2  7
2  2
3  9
3 10
3  8
3  4
END DATA

/*  TELL THE PROGRAM WE'RE DONE!!
FINISH
```

4. Two-way ANOVA example:

```
TITLE     MMC 500 -- TWO-WAY ANOVA EXAMPLE
SUBTITLE  SPRING 2000  LOMBARD [CHANGE THIS TO *YOUR* NAME]

/*  RESTRICT OUTPUT TO 80 CHARACTERS FOR EASY PRINTING
/*  AND READING
SET WIDTH=80

/*  TELL PROGRAM WHAT COLUMNS OF DATA BELOW REPRESENT
DATA LIST
  /  GENDER   1
     INSTRCT  4-5
     SALES    8-9

/*  TELL PROGRAM WHAT THE SHORT NAMES OF THE VARIABLES STAND FOR
VARIABLE LABELS
     GENDER   'GENDER OF RESPONDENT'
     INSTRCT  'TYPE OF INSTRUCTION'
     SALES    'NUMBER OF SALES'

/*  TELL PROGRAM WHAT EACH VALUE ON EACH VARIABLE STANDS FOR
VALUE LABELS
     GENDER   1  'MALE'
              2  'FEMALE'
   / INSTRCT  1  'LECTURE'
              2  'COMPUTER'
              3  'COMBO'

/*  TELL PROGRAM TO CALCULATE A TWO-WAY ANOVA ON THE DISTRIBUTION
/*  OF RESPONDENTS WITH RESPECT TO BOTH GENDER AND INSTRCT.
ANOVA  SALES BY INSTRCT(1,3) GENDER(1,2)
   / STATISTICS=MEAN

/*  GIVE PROGRAM ACTUAL DATA TO BE ANALYZED BY THE COMMANDS ABOVE
/*  THE COLUMNS ARE GENDER, INSTRCT, SALES
BEGIN DATA
1   1   1
1   1   2
1   1   1
1   1   2
1   2   3
1   2   4
1   2   5
1   2   3
1   3   7
1   3   9
1   3  10
1   3   9
2   1   2
2   1   1
2   1   1
2   1   2
2   2   6
```

```
2   2   4
2   2   7
2   2   8
2   3   6
2   3   6
2   3   5
2   3   7
END DATA

/*  TELL THE PROGRAM WE'RE DONE!!
FINISH
```

5. Correlation example:

```
TITLE     MMC 500 -- CCORRELATION EXAMPLE
SUBTITLE  SPRING 2000  LOMBARD [CHANGE THIS TO *YOUR* NAME]

/*  RESTRICT OUTPUT TO 80 CHARACTERS FOR EASY PRINTING
/*  AND READING
SET WIDTH=80

/*  TELL PROGRAM WHAT COLUMNS OF DATA BELOW REPRESENT
DATA LIST
  /  CASENO   1-2
     X        6
     Y        10

/*  TELL PROGRAM WHAT THE SHORT NAMES OF THE VARIABLES STAND FOR
VARIABLE LABELS
     CASENO   'CASE NUMBER'
     X        'VARIABLE X'
     Y        'VARIABLE Y'

/*  TELL PROGRAM TO CALCULATE THE CORRELATION BETWEEN X AND Y
CORRELATIONS  VARIABLES=X Y

/*  GIVE PROGRAM THE DATA TO BE ANALYZED BY THE COMMANDS ABOVE
/*  THE COLUMNS ARE CASENO, X, Y
BEGIN DATA
01   9   2
02   8   3
03   9   2
04   7   1
05   6   2
06   4   1
07   5   1
08   6   2
09   8   3
10   8   3
END DATA

/*  TELL THE PROGRAM WE'RE DONE!!
FINISH
```

6. One-way chi-square example:

```
TITLE     MMC 500 - ONE-WAY CHI-SQUARE EXAMPLE
SUBTITLE  SPRING 2000  LOMBARD [CHANGE THIS TO *YOUR* NAME]

/*  ANYTHING THAT FOLLOWS "/*" AT THE BEGINNING OF A LINE
/*  IS IGNORED BY THE COMPUTER.  IT IS INCLUDED HERE TO DESCRIBE
/*  TO THE READER WHAT IS GOING ON.

/*  RESTRICT OUTPUT TO 80 CHARACTERS FOR EASY PRINTING AND
    READING
SET WIDTH=80

/*  TELL PROGRAM WHAT COLUMNS OF DATA BELOW REPRESENT
DATA LIST
  /  CASENO   1-2
     MAJOR    6

/*  TELL PROGRAM WHAT THE SHORT NAMES OF THE VARIABLES STAND FOR
VARIABLE LABELS
     CASENO   'CASE NUMBER OF RESPONDENT'
     MAJOR    'MAJOR OF RESPONDENT'

/*  TELL PROGRAM WHAT EACH VALUE OF EACH VARIABLE STANDS FOR
VALUE LABELS
     MAJOR    1  'SCIENCE'
              2  'MATH'
              3  'ENGLISH'

/*  TELL PROGRAM TO CALCULATE A ONE-WAY CHI-SQUARE TEST ON THE
/*  DISTRIBUTION OF ALL RESPONDENTS WITH RESPECT TO MAJOR.
/*  THE "EXPECTED" COMMAND TELLS THE PROGRAM THAT THE EXPECTED
/*  DISTRIBUTION IS THAT EACH CATEGORY WILL CONTAIN AN EQUAL
/*  NUMBER OF PEOPLE.
NPAR TESTS CHISQUARE = MAJOR(1,3)
   / EXPECTED = 5 5 5

/*  GIVE PROGRAM THE DATA TO BE ANALYZED BY THE COMMANDS ABOVE
/*  THE COLUMNS ARE CASENO AND MAJOR
BEGIN DATA
01   1
02   1
03   1
04   1
05   1
06   1
07   1
08   1
09   1
10   1
11   1
12   1
13   1
14   2
```

```
15   2
16   2
17   2
18   2
19   2
20   2
21   3
22   3
23   3
24   3
25   3
26   3
27   3
28   3
29   3
30   3
END DATA

/*  TELL THE PROGRAM WE'RE DONE!!
FINISH
```

7. Two-way chi-square example:

```
TITLE     MMC 500 - TWO-WAY CHI-SQUARE EXAMPLE
SUBTITLE  SPRING 2000  LOMBARD [CHANGE THIS TO *YOUR* NAME]

/*  ANYTHING THAT FOLLOWS "/*" AT THE BEGINNING OF A LINE
/*  IS IGNORED BY THE COMPUTER.  IT IS INCLUDED HERE TO DESCRIBE
/*  TO THE READER WHAT IS GOING ON.

/*  RESTRICT OUTPUT TO 80 CHARACTERS FOR EASY PRINTING AND
    READING
SET WIDTH=80

/*  TELL PROGRAM WHAT COLUMNS OF DATA BELOW REPRESENT
DATA LIST
  /  CASENO   1-3
     MAJOR    7
     GENDER   11

/*  TELL PROGRAM WHAT THE SHORT NAMES OF THE VARIABLES STAND FOR
VARIABLE LABELS
     CASENO   'CASE NUMBER OF RESPONDENT'
     MAJOR    'MAJOR OF RESPONDENT'
     GENDER   'GENDER OF RESPONDENT'

/*  TELL PROGRAM WHAT EACH VALUE OF EACH VARIABLE STANDS FOR
VALUE LABELS
     MAJOR    1  'SCIENCE'
              2  'MATH'
              3  'ENGLISH'
   / GENDER   1  'MALE'
              2  'FEMALE'

/*  TELL PROGRAM TO CALCULATE A TWO-WAY CHI-SQUARE TEST ON THE
/*  DISTRIBUTION OF RESPONDENTS WITH RESPECT TO BOTH MAJOR AND
/*  GENDER.
CROSSTABS
   / VARIABLES=MAJOR(1,3) GENDER(1,2)
   / TABLES=GENDER BY MAJOR
   / CELLS=COUNT ROW COLUMN EXPECTED RESID
   / STATISTICS=CHISQ

/*  GIVE PROGRAM THE DATA TO BE ANALYZED BY THE COMMANDS ABOVE
/*  THE COLUMNS ARE CASENO, MAJOR, AND GENDER
BEGIN DATA
001   1   1
002   1   1
003   1   1
004   1   1
005   1   1
006   1   1
007   1   1
008   1   1
009   1   1
```

010	1	1
011	1	1
012	1	1
013	1	1
014	1	1
015	1	1
016	1	1
017	1	1
018	1	1
019	1	1
020	1	1
021	1	1
022	1	1
023	1	1
024	1	1
025	1	1
026	1	1
027	1	1
028	1	1
029	1	1
030	1	1
031	1	1
032	1	1
033	1	1
034	1	1
035	1	1
036	1	1
037	1	1
038	1	1
039	1	1
040	1	1
041	1	1
042	1	1
043	1	1
044	1	1
045	1	1
046	1	1
047	1	1
048	1	1
049	1	1
050	1	1
051	1	2
052	1	2
053	1	2
054	1	2
055	1	2
056	1	2
057	1	2
058	1	2
059	1	2
060	1	2
061	2	1
062	2	1
063	2	1

064	2	1
065	2	1
066	2	1
067	2	1
068	2	1
069	2	1
070	2	1
071	2	2
072	2	2
073	2	2
074	2	2
075	2	2
076	2	2
077	2	2
078	2	2
079	2	2
080	2	2
081	2	2
082	2	2
083	2	2
084	2	2
085	2	2
086	2	2
087	2	2
088	2	2
089	2	2
090	2	2
091	2	2
092	2	2
093	2	2
094	2	2
095	2	2
096	2	2
097	2	2
098	2	2
099	2	2
100	2	2
101	2	2
102	2	2
103	2	2
104	2	2
105	2	2
106	2	2
107	2	2
108	2	2
109	2	2
110	2	2
111	2	2
112	2	2
113	2	2
114	2	2
115	2	2
116	2	2
117	2	2

118	2	2
119	2	2
120	2	2
121	2	2
122	2	2
123	2	2
124	2	2
125	2	2
126	2	2
127	2	2
128	2	2
129	2	2
130	2	2
131	3	1
132	3	1
133	3	1
134	3	1
135	3	1
136	3	1
137	3	1
138	3	1
139	3	1
140	3	1
141	3	1
142	3	1
143	3	1
144	3	1
145	3	1
146	3	1
147	3	1
148	3	1
149	3	1
150	3	1
151	3	1
152	3	1
153	3	1
154	3	1
155	3	1
156	3	1
157	3	1
158	3	1
159	3	1
160	3	1
161	3	1
162	3	1
163	3	1
164	3	1
165	3	1
166	3	1
167	3	1
168	3	1
169	3	1
170	3	1
171	3	2

```
172   3   2
173   3   2
174   3   2
175   3   2
176   3   2
177   3   2
178   3   2
179   3   2
180   3   2
END DATA

/*  TELL THE PROGRAM WE'RE DONE!!
FINISH
```