Fundamentals of Statistics

This course is intended to teach students about the application of statistical theory and statistical methodology. Students will learn to use statistical procedures to examine differences or relationships among variables in a set of data. This course is, perhaps, one of the most important courses that a student can take in their educational career. The principles and procedures that students will learn in this course will enable them to analyze data for course projects and for their thesis. But, more importantly, the principles and procedures taught in this course will be useful in both your personal and professional life.

Consider the fact that statistics are everywhere. Whether it is the weather forecast—that tells you about the temperature, barometric pressure, humidity—or results of sporting events—like batting averages in baseball games, percentage of free throws completed in a basketball game, total rushing yardage in a football game, or the medicines that we rely on to stay healthy—prescription doses reflect the average amount of a given medicine that is required for an individual to effectively treat a condition with minimal side effects, statistics are pervasive in our lives.

Let me give you some examples:

- 9 Million People could not find work in June
- The jobless rate climbed to 6.4%
- 64% of Adult Americans are overweight
- 30% of Adult Americans are obese
- 4.5 Million Individuals claim multiracial identity
- 16.67% babies are born to multiracial heritages
- An average of 13 attacks/day has been launched within the past 45 days (Lt.Gen. Sanchez)
- 23% of Americans polled in July, believe the military effort is going well compared with 61% in polled in April
- 66% of Americans are committed to rebuilt and stabilize Iraq
- 3.9 Million/month is the cost of military operations in Iraq
- Shuttle Program Funding has steadily decline in the past decade from $5.15 billion in '93 to $3.34 billion in '02
- DC Metropolitan area highly educated as compared with rest of US 39% vs. 24% of DC Metro Area have a college degree—Whites (81%) and Asians (58%) are most likely to hold college degree than Black (18%) and Hispanics (25%).

A fundamental understanding of statistical principles and procedures can be useful as we navigate the sea of numbers we are bombarded with on a daily basis. After this course, students will become informed consumers of research. Students will be able to read various books and articles containing the results of statistical analysis with some degree of discrimination. Students will be able to evaluate the strength and weaknesses of an empirical study based on the statistical evidence that is presented.

The Origin of Statistics

Statistics came into use more than 200 years ago. It traces its origin to two quite (dis)similar human interests, politics and gambling. In the mid-sixteenth century Girolamo Cardano, an Italian mathematician, physician and gambler wrote the *Liber de Lude Aleae* (The Book on Games of Chance) that contained the first known study of the principles of probability. In the 17^{th} century, Pascal and Fermat proposed the a set of theorems that would help to establish the mathematics of probability and in the 18^{th} century Laplace published the *Theorie Analytiques des Probabilites* (Analytical Theory of Probability) that laid the foundation for the formal study of statistics. By the mid 18^{th} century, the term statistics was born as a word describing "the political arrangement of the modern states of the known world". The term "statistics" was used to describe a country's political characteristics including population, taxes and area. Statistics referred to "state numbers".

More recently the term statistics has been used to refer to a mass of figures or a collection of data. Statistics also refers to a certain kind of measure used to evaluate a selected property of the collection of items under investigation. As a science, statistics refers to the process of assembling, analyzing, characterizing and interpreting collections of data. It is this process that we will be learning about during this course.

Why Statistics?

Students have always been concerned with statistics. The inevitable questions that student's always ask is "Why do we need this course?" This may be especially relevant for some of you for whom "formal statistical theories and methods" may not have been a part of your professional activities. Guilford (1956) has provided a good summary of reasons for students to take statistics:

1. *A student must be able to read professional literature.* As part of your educational process you are required to read and digest research. Some of this research may contain statistical information that you must decipher to evaluate the merits of an empirical claim. You cannot read much of the literature in any field, particularly in the social and behavioral sciences, without encountering statistical symbols, concepts and ideas. "Skipping" the sections of an empirical article that present such information is not a useful strategy.
2. *A student must master statistical techniques for advanced courses.* Full appreciation of the subject matter of advanced course sometimes requires an understanding of statistical concepts.
3. *Statistics is an essential part of professional training.* To some extent, statistical logic, statistical thinking and statistical operations are a necessary part of any profession.
4. *Statistics are a basic part of research.* All disciplines within the social and behavioral sciences (e.g., leadership, human resource management) rely on research that employs statistical principals and procedures to derive inferences about their subject matter.

Statistics allow us to summarize a result in a meaningful way. Statistics allows us to describe our observation and enable us to make predictions about the relationship among a set of variables in a set of data. Statistics enable us to analyze causal relations among variables and force us to be precise in both our thinking and our interpretation. Statistics enable us to draw conclusions that can be challenged and verified by others. Finally, statistics are particularly important for research associated with the Master's thesis.

Consider the following examples from real life:

Egas Moniz

In 1949, Dr. Egas Moniz was awarded the Nobel Prize in Physiology and Medicine for the development of prefrontal lobotomy---a surgical procedure in which the connections between the prefrontal lobes and the rest of the brain are cut--as a treatment for mental illness.

Moniz's discovery was based on the report that Becky, a chimpanzee that frequently became upset when she made errors during the performance of a food-rewarded task, did not do so following the creation of a bilateral lesion (an area of damage to both sides of the brain) of her prefrontal lobes. After hearing about this isolated observation at a scientific meeting in 1935, Moniz persuaded neurosurgeon Almeida Lima to operate on a series of psychiatric patients; Almeida cut out six large cores of prefrontal tissue with a surgical device called a leuctome.

Following Moniz's claims that prefrontal surgery was therapeutically successful and had no significant side effects, there was a rapid proliferation of various forms of prefrontal psychosurgery. One such variation was transorbital lobotomy, which was developed in Italy and then was popularized in the United States by Walter Freedman in the late 1940's.

It involved inserting an ice-pick like device under the eyelid, driving it through the orbit (the eye socket) with a few taps of a mallet, and pushing it into the frontal lobes, where it was waved back and forth to sever the connections between the prefrontal lobes and the rest of the brain. This operation was frequently performed in the surgeon's office.

As it incredible as it may seem, Moniz's program of psychosurgery was largely based on the observation of a single chimpanzee in a single situation. What problems can you see with this type of research?

- Animal and human brains are different. The lack of appreciation for brain differences between and within species underscores one of the major flaws of Moniz's assertions.

- There was no systematic assessment of the consequences of the surgery. The first reports about the operations therapeutic success was from physicians who had prescribed the surgery--least objective people. Patients were frequently judged as improved if they were more manageable, and little effort was made to evaluate more important aspects of their psychological functioning (e.g., cognitive abilities) or adverse effects of the surgery.

Eventually, it became clear that prefrontal lobotomies had little therapeutic benefit. Further it resulted in a number of side effects including amorality, lack of foresight, emotional unresponsiveness, epilepsy and urinary incontinence. These and other discoveries led to the abandonment of frontal lobotomies--but not before over 40,000 patients had been lobotomized in the US alone. Today frontal lobotomies continue to be performed in some countries. Some of these individuals disregard the fundamentals of scientific protocol as unnecessary obstacles in the path of patients seeking treatment and therapists striving to provide it.

Although this example may appear to be dramatic, it underscores several important points. First, an appreciation and understanding of research methodology is important for both practitioners and consumers of research. Secondly, an appreciation and understanding of statistical principles and procedures is essential for researcher to evaluate the findings reported in the literature. Finally, research methodology and statistical principles and procedures go hand in hand.

What is the relationship between design and statistics?

The major goal of any research endeavor is to build an organized body of knowledge about a particular subject matter and to develop valid and reliable explanations for particular phenomena. The principal method for acquiring knowledge and uncovering causes of behavior is research. You identify a problem and then systematically set out to collect information about the problem. Research is an essential part of any science and is important for a number of reasons:

1. Research allows us to test ideas empirically.
2. Research allows us to specify the limitations of our results.
3. Research allows us to establish universal laws about behavior.

What distinguishes a science from nonscience (and pseudoscience) involves the quality of the explanations offered for the phenomena, and the methods used to uncover those explanations.

The Scientific Method

The scientific method is a cycle of steps used to pursue the solution to a scientific problem. This cycle involves (1) observing the phenomena, (2) forming a tentative explanation or statements of cause and effect, (3) further observing or experimenting (or both) to rule out alternative explanations and (4) refining and retesting the explanations.

Observing phenomena

The starting point for using the scientific method is to observe the behavior of interest. Through the process of observation, you identify variables that appear to be important in explaining behavior. A **variable** is any characteristic or quantity that can take on two or more values.

Formulating tentative explanations

After identifying interesting phenomena to study, your next step is to develop one or more tentative explanations that seem consistent with your observations. In science these tentative explanations often include a statement of the relationship between two or more variables. These tentative statements are called hypotheses. Research **hypotheses** often take the form of a statement of how changes in one variable will affect the value of another variable.

Further observing and experimenting.

After you formulate a hypothesis, you design a research study to test the relationship you proposed. This study can take a variety of forms. It could be a correlational study in which you simply measure two or more variables and look for a relationship among them, a quasi experimental study where you take advantage of some naturally occurring event or preexisting conditions, or it could be an experiment where you systematically manipulate a variable and look for changes in the value of another that occurs as a result.

Refining and Retesting explanations

The final step in the process is refining and retesting. This entails generating new, more specific, hypotheses in light of the previous results.

There are eight specific steps involved in the research process.

> Developing a research idea. The first step in the research process is to identify an issue you want to study. Once you have identified a behavior to study, you must then state a research question in terms that will allow others to test it empirically.
>
> Choosing a research design. Next you must decide on a design or plan of attack for your research. This includes considering where to conduct your research, the equipment needed, and so on.
>
> Choosing subjects. After deciding on the design, you must consider whether to use human or animal subjects. In addition, you must consider how to obtain your subjects, how they will be treated or handled and the cost related to using them.
>
> Deciding on what to observe and appropriate measures. You must decide exactly what it is that you want to observe. Along with this, you must decide the most appropriate way to measure the behavior of interest.
>
> Conducting your study. After making the necessary decisions and preparations you are now ready to conduct your study. You observe and measure the behavior of interest and record the data.
>
> Analyzing your results. After you have collected your data, you must summarize and analyze it. This process involves making decisions about the types of analysis to be carried out and the way to present the data.
>
> Reporting your results. After analyzing your data, the last step involves preparing a report of your research.

Using Statistics to Describing Data

As I have alluded in the description of the scientific method, statistics figures prominently in this process. Once you have conducted your research, the next step is to organize, summarize and describe your data. How you organize your data depends on your research design, on how many variables were observed and recorded, and on how observations were grouped or subdivided.

Descriptive statistics are commonly used to address the specific questions you had in mind. They can also be employed to help you discover important but perhaps hidden patterns in your data. The search for such patterns in your data is termed exploratory data analysis. A **descriptive statistic** is a numerical index that is used to summarize or describe a set of data. Most of you are familiar with at least one form of this definition. In your classes after exam results are given, most of your professors tell you the class average or mean. The mean is a number that indicates the central tendency of a collection of numbers. Usually, most teachers use the mean to determine whether or not to grade on the curve. So a **descriptive statistic** is a number, which expresses a single characteristic of a set of data.

Inferential statistics are another set of statistics that can be used to answer questions of interest in your research. Inferential statistics were developed as a way to account for the effects of chance that go with any sample. When we collect information on people's attitudes toward a particular topic, we want to ascertain the extent to which our statistical estimates reflect those of the general population or target group. That is, we want to determine whether our result is descriptive of the phenomena we are studying and not a byproduct of the individuals in our sample. To evaluate our ability to make inferences about our sample to a population, we employ inferential statistics. Inferential statistics allow us to account for the effects of chance that go with any sample.

Inferential statistics is a method that takes chance factors into account when samples are used to reach conclusions about populations. This is particularly important since most of the time in research we seldom have the resources, time or manpower available to conduct studies with entire populations. Inferential statistics allows us to draw a sample from a population and determine the extent to which our findings are/are not due to chance. Inferential statistics can be used to assess the "statistical significance" of the findings and to request only those descriptive statistics related to the analysis, such as group means, standard deviations, and standard errors.

Most inferential statistical tests make certain crucial assumptions about the populations from which the scores in your data set were drawn. If these assumptions are violated, the results of the statistical analysis may be misleading.

Suppose we want to know how satisfied service members are with their military careers, their unit commanders and their installations. How would we go about determining which service members to survey to determine individual attitudes toward these topics? The most accurate estimate would be obtained if we were to ask every service member in the military. However, for obvious reasons we cannot do that—both practical and economic costs make it unreasonable to try to survey every individual member of the armed forces. Instead, we may draw random representative samples from the population of military personnel and make statistical inferences about the larger population.

Suppose I asked all members of the USNA to complete a survey assessing their satisfaction with the military life. What would we be able to say about satisfaction with military life? Can you think of some of the problems related to sampling from only USNA personnel? Inferential statistics would enable us to examine the extent to which our results were at least not due to some chance factor. We often hear of this in the form of "significant results". Inferential statistics works by giving you the probability of obtaining the difference you actually observed if, in fact the difference is present.

In sum, statistics allows to summarize large amounts of numerically coded data and enable us to describe patterns and make inferences to members of other groups and sometimes to a larger population. Now let me introduce you to some important concepts that you need to know as we begin our statistical odyssey.

TERMINOLOGY

Population. A population consists of all members of a specified group. A sample is a subset of a population. For example:

Depressed Midshipmen	US Population	Anglo Americans
USNA Students	Annapolis Residents	Asian Americans
USNA Graduates	Maryland Residents	African-Americans
US Military Personnel	East Coast Residents	Hispanic Americans
USN Personnel	West Coast Residents	Native Americans

Investigators are always interested in populations. However, as you can imagine we seldom have the financial resources to conduct a study with ALL the members of a given population. Also, some populations are so large that not all the members can be measured (e.g., US population). Think of US population. How many people would be necessary to estimate any type of statistics for this population? What are some of the variables that you would want to consider?

As a result of this very real practical concern, we often have to resort to measuring a sample that is small enough (and thereby affordable enough) yet large enough to be representative of the population. Resorting to the use of samples introduces uncertainty into the conclusions because different samples from the same population nearly always differ from one another in some respects. Inferential statistics are used to determine whether or not such differences should be attributed to chance.

Example: Attitudes toward women in the Swedish Military

As part of a series of studies I have been involved with on the topic of gender integration in the military, several colleagues and I wanted to understand how male officers in the Swedish Armed

Forces felt about having women in the military. Unlike the US military, women make up nearly 5% of the Armed Forces in Sweden. We surveyed over 1,320 male officers to assess their attitudes toward the participation of women in the armed forces. We found that the male officer's in our sample scored 3.89 (where 3 represented neutral scores and higher scores were more positive) on a 10 items scale designed to assess attitudes toward women in the military. If we had stopped at this point we would be describing our data. We were interested in determining whether the observed result of 3.89 was reliably different from 3.0, the neutral point of the scale. In essence we wanted to determine whether our results was reliable or whether it was simple due to chance factors in our sample. We performed a one-sample t-test to examine the probability of obtaining this result beyond chance. We found that that the difference was indeed significant at the $p < .001$ level.

By convention, scientist have established that the minimum criteria for determining the significance of a particular results is a $p < .05$. That means that we would expect to obtain a statistic of XX by chance only 5% of the time. In our case, the results of the statistical analysis indicated that our mean was reliably different from 3.0 at the $p < .001$. Thus, we can be 99.999% sure that our findings are in fact different from 3.0 the neutral point of the scale. So now we can interpret our result and say that men's attitudes were slightly positive—recall that the score was 3.89 where 3 is neutral, 4 would reflect positive attitudes and 5 would reflect very positive attitudes.

As illustrated by this example, when we resorting to the use of samples, we introduce uncertainty into the conclusions. Different samples from the same populations nearly always differ from one another in some respect—what we want to know is whether the difference we obtain is reliably different and not due to chance factors, like fluctuation in our sample. Inferential statistics is used to determine whether or not such differences should be attributed to chance.

Parameter & Statistics

Another important term you need to become familiar with is a parameter and a statistic. A **parameter** is a numerical (number) or nominal (name) characteristic of a population. For example, the mean IQ of college freshmen in the US is a parameter. The mean score on the SAT verbal and quantitative sections for all US college freshmen is a parameter. A parameter is constant and it does not change unless the population itself changes.

A **statistic** is a numerical or nominal characteristic of a sample. For example, the mean IQ of USNA freshmen engineering majors is a statistic. The mean score on the SAT verbal and quantitative sections for all USNA freshmen is a statistic. Note that in both examples we are describing a numerical characteristic of a small group not an entire group as was the case for a parameter.

Remember that we seldom work with populations. The majority of the time we include samples from a particular population and use inferential statistics to estimate the probability that our particular result is due to chance at some predetermined value ($p < .05$). Also, often we cannot compute a population parameter for some research questions. So we use a statistic as an estimate of a parameter.

VARIABLES

Variables are something that exists in more than one amount or in more than one form. A variable is a synonym for a characteristic we are interested in investigating. For example, height, weight, gender, political ideology, intelligence, ethnicity, are all types of variables. There are two types of variables quantitative and qualitative variables.

Quantitative variables are ones that we can assign a numerical value to it. For example, heights, weight, intelligence all have numbers that indicate the degree to which an individual possess that variable. So, for weight some people possess lots of it while others less of it—recall that 64% of adult Americans are overweight and 30% are obese. Similarly with height and IQ. When a quantitative variable is measured, the scores tell us something about the amount or degree of the variable.

Qualitative variables do not have the continuous nature that quantitative variables have. That is the numbers that may be associated to these variables do not indicate the degree to which an individual possesses a particular characteristic. For example, ethnicity, political party membership, sex.

SCALES OF MEASUREMENT

Having talked about types of variables and their definitions we are now ready to think about how to measure or assess characteristics or variables of samples or populations. There are four different types of scales of measurement. Each scale uses numbers, but the information that can be inferred from the numbers differs. The kind of statistics we can use depends on the scale of measurement that we employ. The kind of descriptive statistics you can compute from your numbers depends, in part, on the scale of measurement the numbers represent. Scales of measurement are also important in choosing the kind of inferential statistic that is appropriate for a set of data.

Nominal Scale

Nominal scales used numbers simply as names. The number has no real numerical value. The number on sports uniform is an example of a nominal scale. The numbers differ from one another but the difference has no real value. So, player number 14 is different from player number 2 but they are not necessarily better or worse or anything else based on their number. Another way to think about nominal variables is as categories. For example, when we are asked about our ethnic background we are often given a list of groups and asked to write the number of the group to which we belong.

Ordinal Scales

Ordinal scales use numbers to name a variable and the number has the characteristic of indicating greater than or less than. The number reflects the degree to which a property of a given variable is present. For example, using our sports example, most professional sports leagues have rankings based on a team's performance. At the end of a season we find out who the number one team in the league is and the number 2 and so on. In this case the numbers are used as labels and they also indicate that one team is greater than (better than) the others on the league. Ordinal scales are characterized by rank order.

Interval Scales

Interval Scales have the properties of both nominal and ordinal scales. That is the numbers are used to represent a name (quantity), they also represent an order (greater than or less than). In addition, an interval scale has the additional property that intervals between the numbers are equal. Equal interval means that the distance between the things represented by say the number 2 and 3 is the same as the distance between say 3 and 4, 4 and 5 and so on. For example, we know that the difference between 79 and 89 degrees is the same as that between 59-69, 99-109. We also know that the difference between an IQ score of 50 and 60 and 80-90 is the same (10 points). However, it is important to remember that we cannot make statements concerning the relative difference between these sets of temperatures or scores. So we cannot say that 100 is twice as hot as 50 or that a person scoring 40 is half as intelligent as a person scoring 80 on an IQ test. The absolute value of the variable does not represent the absolute quantity of the attribute being measured.

Ratio Scales

Ratio scales have all of the properties of all the previous scales, nominal, ordinal, interval, and an additional one. The ratio scale has a true zero point which indicates the complete absence of something. You will remember from our previous example with temperature 0 degrees does not mean we have not temperature. In this case 0 degrees indicates an actual temperature. On a ratio scale the 0 point means that the characteristic is not present. For example, height, weight, and time all use ratio scales—0 indicates the absence of something. With a true zero point we can make inferential statements about the meaning of particular numbers. So for example, we can say that a person weighting 200 lbs is twice as heavy as a person weighting 100 lbs.

As you can see from our on-going discussion, scales of measurements are important since they dictate the type of statistics we can use to analyze our data.

Lets summarize, statistics involves the manipulation of numbers and the conclusions based on those manipulations. Statistics has its foundation in mathematics. Statistics describes the mathematical procedures that we use to analyze data.

Main Concepts

Explain the difference between descriptive and inferential statistics.
Define population, sample, parameter, and statistic.
Define quantitative and qualitative variables.
Identify the four scales of measurements.

Multivarite statisics are increasingly popular techniques used for analyzing complicated data sets. They provide analysis when there are many IVs and/or many DVs, all correlated with one another to varying degrees.

Multivariate models afford researchers with the ability to handle more complex data analysis. Multivariate statistical methods are an extension of univariate and bivariate statistics. Multivariate are the complete or general case.

If your design has many variable, multivariate techniques often let you perform a single analysis instead of a series of univariate or bivariate tests.

Variables can be thought of as independent (predictor) variables and dependent (criterion) variables. Independent variables are the differing conditions to which you expose your subjects or characteristics that the subjects themselves bring into the research situation. IVs are usually considered predictors or causal variables because they predict or cause a DV—a response or outcome variable. It is important to note that a IV and DV must be defined by the context of the research. A variable can serve as DV in one analysis but a IV in another.

Univariate statistics refers to analyses in which there is a single DV. There may be however, more than one IV.

Bivariate statistics refer to analysis of two variables neither of which is experimentally manipulated and thus not an IV in the formal sense. Bivariate analysis simply determine the linear relationship between variables.

Multivariate statistics – refers to the case where there are multiple IV and DV.

Experimental vs. Nonexperimental Research

A critical distinction between experimental and nonexperimental research is whether the research manipulates the levels of the IVs. In an experiment, the researcher has control over the levels of at least one IV to which a person is exposed. The experimenter also randomly assigns subjects to levels of the IV and controls all other influential factors by holding them constant, counterbalancing or randomizing their influence. Scores on the DV are expected to be the same, within random variation, except for the influence of the IV. The value of the DV depends on the manipulated level of the IV. The IV is manipulated by the experimenter, and the score on the DV depends on the level of the IV.

In nonexperimental (correlational/survey) research, the levels of the IV(s) are not manipulated by the researcher. The research can define the IV but has not control over the assignment of subjects to levels of it. In nonexperimental reseach, it is very difficult to attribute causality to an IV, if there is a systematic difference in a DV associated with the levels of an IV, the two variables are said to be related, but the cause of the relationship is unclear.

One of the few considerations not relevant to choice of statistical technique is whether the data are experimental or correlational.

Nonexperimental research takes many forms. One of the most common forms is the survey. In a survey, a respondent provides answers to many questions, producing a large number of variables. These variables are usually interrelated in highly complex ways, but univariate and bivariate statistics are not sensitive to this complexity. Bivariate correlations between all pairs of variables for example, could not reveal that the 20 to 25 variables measures equally represent only 2 or 3 supervariables.

Although most multivariate techniques were developed for use in nonexperimental research, they are also useful in experimental research in which there may be multiple IVs and DVs. Until recently, experiments with multiple DVs were unsual. The risk of running an experiment with only one DV and risk missing the impact of the IV because the most sensitive DV is not measured. Multivariate statistics help the experimenter design more efficient and more realistic experiments by allowing measurement of multiple DVs without violation of acceptable levels of TYPE I errors.

The trick in multivariate statistics is not in computation but in the selection of reliable and valid measures that assess the variables of interest and allow for the use of computers to analyze data and interpret the output.

Variables

Continuous variables are measured on a scale that changes values smoothly rather than in steps. Continuous variables can take on any value within the range of a scale and the size of the number reflects the amount of the variable. Examples include age, temperature, GPA.

Discrete variables take on a finite and usually small number of values and there is no smooth transition from one value or category to the next. Examples include religious affiliation, political affiliation, region of country.—recategorization of a discrete variable into a series of dichotomous ones is called dummy variable coding.

The property of variables is not crucial to application of multivariate procedures but the nature of their distribution.