Statistics for Political Scientists

by

Christopher Dennis

Note: If you download all 6 files of this manuscript you will have discussions of all the topics listed ahead. There may, occasionally, be discontinuity in page numbering. As I made revisions, I shortened one particular section. Due to changing from Word Perfect to Word, the "cost" of redoing the entire manuscript became "high." I don't know how to type equations in Word (i.e., like the equation boxes/equation editor in Word Perfect). Also, the manuscript has a number of hand drawn diagrams. So, I can't just "run off" the entire manuscript in Word. This means that many of the pages are "scanned." This is why the files are so large and the manuscript is broken into 6 files. All the page numbers refer to "the text" (i.e., do not include either the title page or the table of contents). So, use the page numbers below as "approximate." This also accounts for any page discontinuity. The font is changed in order to have all the information necessary appear on the appropriate page. This also requires explanation. This manuscript is cross-referenced within itself. Thus it may tell you to go back to a particular page. This makes revisions difficult. It's easier to change fonts than all the cross-references. If you would like the "Word" files that contain the text, please email me (cdennis@csulb.edu) and I'll send them to you.

Table of Contents

The Scientific Method	1-4
Hypothesis	1-2
Research Design	4-8
Measurement	8-11
Levels of Measurement	9-10
Generalizing Results	11-13
Unit of Analysis	12-13
Ecological Fallacy	12
Descriptive Statistics	14-27
Frequency Distribution	14
Measures of Central Tendency (Mean, Median and Mode)	16-17, 64, 66
Measures of Dispersion	17-25, 64, 66-67
(Standard Deviation)	19-24
Coefficient of Variation	23

Normal Distribution (or Normal Curve)	24
Tchebysheff's Theorem	25
Z Scores (or Standard Scores)	25-27
Cross Tabulation	28-30
Measures of Association	31-34, 69, 75
Weaknesses of Cross Tabulation and Measures of Association	35-38
Probability Theory	39-63
Random Sample and Population	40
Null Hypothesis	48, 53-54
Type I and Type II Errors	49-50, 53-56
One and Two-Tailed Tests	51-52
Infinite Population	57-60
Regression	
Computing and interpreting "b"	76-77, 109-110
Y Intercept or "a"	78-79, 109-110
Residual or "e"	79-81
Explained Variance or R^2	82-84, 109-110
Least Squared Errors Principle	85-87
Statistical Significance of "b" (t ratio)	88-92
Nonlinear Models	93-105
Logarithmic Transformation	93-100
Polynomial Transformation	101-105
Multiple Regression	106-114
Multicollinearity	115-123
Explained Variance Test	120-121
Deletion Test	121-122

Dummy Variables	124-126
Standardized Coefficients	127-129, 154-155
Interaction Terms	130-132
Heteroscedasticity	133-139
Autocorrelation	140-149
Distributed Lag Models	147-149
Pooled Time Series Analysis	149
Discrete Dependent Variable Models	150-164
Logit	151-159
Magnitude Assessments in Logit	155-158
Likelihood Ratio Index (LRI)	158-159
Probability and Odds	159-160
Probit	160-161
Ordered/Multinomial Logit/Probit	161-162
Hazard Models	162-163
Event Count Models	163
Causal Models	165-167
Simultaneous Equation Models	168-178
Vector Autoregressive Models (VAR)	178

The Scientific Method in Political Science

Political scientists are interested in a wide variety of different topics. Some political scientists study the causes and consequences of war. Other political scientists study why voters support a particular candidate and what difference it makes which candidate or party is victorious. For political scientists studying these and a vast array of other topics, quantitative application of the scientific method offers one of the most useful approaches to increasing knowledge. While the scientific method is not useful in answering questions posed by normative political philosophers (e.g., What is justice?), it is extremely valuable in understanding how and why political phenomena occur. As a first step in this process, we need to understand what the scientific method is. Let me suggest that science is defined by its methodology, not its subject matter. Thus, it is how one studies something, not what is studied, that determines whether or not the researcher is using the scientific method. Throughout this course I will use the following definition of science: a communicable (can communicate to those who do not "know"), falsifiable (possibility of non-confirmation), and logical (conclusions follow from the facts) method of pursuing knowledge involving the recognition and formulation of a problem, the collection of data through observation and experiment, and the formulation and testing of hypotheses. That is a long and involved definition!

In order to fully understand the above definition of science, it is useful to break it into parts. As stated in the definition, "communicable" means that your study can be understood by those who are not already part of it. Thus, you need to be able to explain your method to other researchers so that they can both check your work for accuracy and apply it to new circumstances. For example, if you were testing a new vaccine, future researchers might wish to replicate (repeat) your work on the same type of subjects and then administer it to a different group. Therefore, other researchers would have to know both the composition of your vaccine and how the tests of it were conducted.

The second portion of the above definition of science concerns "falsifiability." Falsifiability means that there must be some outcome which would countermand what we expect to happen. For example, suppose you said that if God wants you to go to Pittsburgh he will provide you with the plane tickets. In terms of falsifiability, this would not be a scientific test of the existence of God. Regardless of whether you received the tickets or not, you would not doubt the existence of God. No possible outcome would lead to a rejection of the premise that God exists.

The next portion of the definition of science concerns logic. As stated above, your conclusions must follow from the facts. You cannot conclude that the Democratic party favors a greater tax burden on the wealthy than the Republican party when the facts suggest otherwise.

The next sections of the definition of science are rather straightforward. Certainly, you must formulate a problem or else there is nothing to study. Furthermore, you collect data which bear on the problem (or topic) you are studying. For example, you collect data on the differences between the tax proposals offered by the Democratic and Republican parties.

The final portion of the definition of science concerns the formulation and testing of hypotheses. Let me define a hypothesis as follows: a relational statement between two, or more, concepts which is deductively plausible and empirically generalizable. I think it would be useful to begin by defining a concept. A concept is an abstraction representing an object, a property of an object, or a certain phenomena. For example, "poverty" could be a concept. Should we conceptualize "poverty" in "relative" or "absolute" terms? A person with an income of $20,000 could be thought of as impoverished relative to someone who had an income of $500,000. On the other hand one could argue (as conservatives generally do) that poverty is absolute. Thus, the determination of poverty would not concern how much income you had relative to someone else, but rather whether you could attain a particular standard of living (e.g., avoid hunger). My point is that a researcher measuring poverty has to use one of these conceptualizations. Obviously, it matters which one they choose. One of the great advantages of the scientific method is that the researcher must state and defend their choice. By so doing, other researchers can then use the same, or different conceptualizations, and see to what extent the results are affected by the conceptualization employed.

A concept that can assume different values is called a variable. For example, since all governments do not have the same degree of liberalism, governmental liberalism is a variable. Two particular types of variables are fundamental to the scientific method (particularly hypothesis testing). The presumed "causal" factor is referred to as the "independent" (or "predictor") variable and the effect is referred to as the "dependent" variable. Suppose we hypothesize that the degree of governmental liberalism alters the percentage of income going to the poor. Since governmental liberalism is presumed to effect the percentage of income going to the poor, governmental liberalism is the independent variable and the percentage of income going to the poor is the dependent variable. Alternatively, you might think of it this way: the score on the dependent variable depends upon the score on the independent variable (not the other way around).

An operationalization is the measurement of a concept. For example, how do you quantify whatever conceptualization of "income" you are using? If you receive medical benefits from the government does this count as "income"? Should the use of a company car count as "income"? The researcher must explain how they measure "income" and defend why their measure is appropriate.

A relational statement is a causal or associational link between concepts. For example, suppose we hypothesize that the liberalism of the federal government is "positively" associated with the percentage of income going to the poor. The previous statement is relational because it depicts an association between the concepts (liberalism of the federal government and the percentage of income going to the poor). Notice the use of the directional term "positive." A "positive" relationship means that higher scores on one variable are associated with higher scores on the other variable. For example, if the liberalism of the federal government and the percentage of income going to the poor are "positively" related (i.e., associated) it would mean that if the score on federal government liberalism were to increase from say 50% to 75% (e.g., by electing many more Democratic congressmen and senators) the percentage of the national income going to the poor might then

increase from 10% to 12%.

Relationships between variables can also be "negative." A "negative" relationship means that higher scores on one variable are associated with lower scores on the other variable. For example, if the liberalism of the federal government and the percentage of income going to the poor are "negatively" related it would mean that if the score on federal government liberalism were to increase from 50% to 75% the percentage of the national income going to the poor might then decrease from 10% to 8%.

Deductive plausibility means that the researcher deduces (reasons from) something else which is plausible. Thus, if we observe that Democrats tend to be more liberal than Republicans we may reasonably deduce from this that a particular Democratic candidate is likely to be more liberal than their Republican opponent (this is what we hypothesize and will be testing).

Empirically generalizable means that our findings are applicable to much of the observable (empirical) world. We want to generalize as far as we can. From Democrats and Republicans in California to Democrats and Republicans in the United States as a whole. Any theory (a theory is just a more certain hypothesis) is more valuable the wider its applicability. For example, isn't the anti-crime argument for the death penalty (that the death penalty will lower the murder rate) actually just an application of the basic economic theory that the more something costs (here "costs" would refer to the penalty) the less of it will be sold (each murder would be an occurrence, i.e., a "sale")?

Basically, the scientific process is just a continual testing of hypotheses in order to find their limits (i.e., how far they can be generalized) and then to modify the theory in light of the findings. For example, in most western democracies the poor are more supportive of liberal governments than conservative governments. Since it is logical to hypothesize that a government will pursue policies that disproportionately benefit its supporters, it would seem logical to hypothesize that the degree of liberalism of the government is positively associated with the percentage of income going to the poor. Thus, higher scores on our measure of governmental liberalism should be associated with higher scores on our measure of the percentage of income going to the poor. In testing this hypothesis we may find that government today has either a greater, or lesser, impact on the distribution of income than during the 1950s.

Another benefit of the scientific method is that the user must make their model explicit. For example, it is possible that the liberalism of the government has little "direct" effect on the percentage of income going to the poor. Since governments often control policy instruments (e.g., the money supply), as opposed to policy outcomes (e.g., the percentage of income going to the poor), it is likely that much of the effect of government on the percentage of income going to the poor would be "indirect" (i.e., through other factors). For example, a more liberal government could increase the money supply. A larger supply of money lowers interest rates which, in turn, make borrowing less expensive. The reduced cost of borrowing money generally causes plants to expand which, in turn, lowers the unemployment rate. As the unemployment rate decreases, the percentage of income going to the poor typically increases. My point is that the user of the scientific method must explain

which variables effect which other variables (i.e., they must make their "model" explicit).

Because users of the scientific method must make their models and measures explicit, other researchers can replicate (i.e., repeat) and expand on the original study. Over the past two decades, political scientists have tested the governmental liberalism hypotheses I have been mentioning in most all major industrialized democracies in the world. They have used an impressive group of alternative income measures, time periods, and models. For example, in addition to studying the effect of governmental liberalism on the money supply, political scientists have also examined the effects of governmental liberalism on the amount and distribution of the tax burden over various income groups, numerous measures of social welfare spending and the amount of economic growth.

Users of the scientific method usually have two goals in mind. Typically, the first goal of a user of the scientific method is explanation. In our example we are trying to explain why the percentage of income going to the poor varies (i.e., is not always the same - hence a "variable"). Our hypothesis is that variation in the liberalism of the government is what causes variation in the percentage of income going to the poor. A large literature (to which political scientists have greatly contributed) has rather firmly established that governmental liberalism is positively associated with the percentage of income going to the poor. However, while governmental liberalism is likely to positively influence the percentage of income going to the poor, other factors (i.e., independent variables) are also likely to influence the percentage of income going to the poor (e.g., international economic trends). The result of incorporating these additional independent variables in the data analysis is a richer explanation of why the percentage of income going to the poor varies.

A second goal of users of the scientific method is prediction. Applied to our hypothesis this would mean to predict how much the percentage of income going to the poor would increase, or decrease, depending upon a particular amount of change in the liberalism of the government. Often these two goals are related. As our ability to "explain" a process improves, our predictions are likely to become more accurate. However, prediction is more difficult than explanation. The impact of some of the independent variables may change in the future. Consequently, accurate predictions are difficult. Nevertheless, political scientists have formulated relatively accurate forecasts of the share of the vote American political parties will receive (the dependent variable) based upon changes in various economic and non-economic variables (the independent variables). However, typically the major goal of contemporary quantitative political science is explanation.

Research Design

Before continuing, make sure you understand that pages 2-5. The topics dealt with over pages 2-5 are the foundation of every reading in this course. The first quiz (coming the day this reading assignment is due) may well ask you to define a variable, abstract a hypothesis from written material, and/or to explain the difference between a "positive" and "negative" relationship. You will need the information from

the aforementioned lecture on the scientific method for quizzes 1-3 and the final examination.

The main purpose of pp. 6-14 is to discuss the early stages of a quantitative research project. The first decision any researcher must make is what topic to study. A political scientist should be able to defend their choice of a topic on normative grounds. Thus, why is the topic important? For example, why should we study the causes of war? I think one could make an excellent case that war is undesirable and, consequently, that determining why war starts is a logical pre-condition to minimizing its occurrence. Although a normative defense of a topic is important, it is typically handled in several sentences. The central contribution of quantitative research is to explain what takes place and why, not what is "good" or "bad."

In quantitative research (the topic of this course), we are usually testing a theory of behavior. Whether it is the behavior of governments or individuals, we will probably be examining the causes (and/or consequences) of some form of political behavior.

Any quantitative (i.e., empirical) study is trying to perform two fundamental tasks. First, we are trying to test and refute hypotheses. For example, is the liberalism of a government positively associated with the amount of government support for the poor? Second, we are trying to estimate the magnitude of the relationships between the variables (Hanushek and Jackson, Statistical Methods for Social Scientists, pp. 2-3). For example, the replacement of a Republican President with a Democratic President would result in how much more support for the poor?

After formulating the hypotheses (defined on pages 3-5), we need to begin thinking about how we will test them. The strategy by which one tests their hypotheses is called a research design.

While this may be a bit of an oversimplification, there are two basic types of research designs. The first type of research design is called an experimental design. With an experimental design, the researcher can adjust the level (i.e., amount/scores) on each of the independent variables. For example, supposing a biologist formulates a new plant growth additive and wishes to test its effectiveness. The amount of the plant growth additive each plant receives would be the independent variable. The growth rate of the plant would be the dependent variable. The biologist would probably think that factors other than the amount of the plant growth additive would alter the rate of plant growth. Thus, the plant biologist would want additional independent variables. For example, such factors as the type of plant, plant condition, water quality and the amount of sunlight could all affect the growth rate of a plant. Each of these factors, in addition to the growth additive, is an independent variable. The advantage of using an experimental research design is that the researcher can set the level of each of the independent variables. For example, the plant biologist can determine what types of plants will be used, the amount of the growth additive each plant will receive and the amount of sunlight each plant is exposed to. Being able to set the level (i.e., amount) of each of the independent variables is an extremely useful capability. If all conditions (i.e., independent variables) other than the independent variable in which the plant biologist is most interested (the growth additive) are set at the same level (i.e., "controlled" - each plant is of the same type, receives the same amount of sunlight, etc.) and if plants

that receive more of the plant growth additive grow faster, we are on rather sound ground in thinking that the plant growth additive matters. Since the plants do not differ on any factors that could conceivably alter their growth rates except the amount of the plant growth additive, it makes sense to think that the growth additive increased plant growth rates.

By contrast, a political scientist will almost invariably have to use what is termed a nonexperimental research design. With a nonexperimental research design the researcher is not able to set the levels of the various independent variables. The inability of the researcher to set the levels of the various independent variables is important because it is possible (in some circumstances likely) that the independent variables will be related to each other, as well as to the dependent variable. We refer to the situation where the independent variables are strongly related to each other as "multicollinearity."

For example, suppose we are trying to test a model of partisan affiliation (the dependent variable). Thus, our model will be trying to explain why individual voters register as Democrats, Republicans, or Independents. Please note that we have three categories of responses (i.e., Democrat, Republican or Independent) on one dependent variable. Let us say that two of our hypotheses are that the more highly educated a voter is the more likely they are to register Republican and the higher the voter's income the more likely they are to register Republican. Note that both education and income are independent variables. Since occupations requiring a higher level of education generally pay higher salaries than occupations with lower educational requirements, education and income are likely to be related. If it turns out, as is likely, that education and income are strongly related to each other (hence we have "multicollinearity"), and both education and income are related to partisan affiliation, it can be difficult to determine the impact of either education or income on partisan affiliation. In the worst case situation, where all voters with high levels of education have high incomes and vice versa, it would be impossible to determine the contribution of either education or income to partisan affiliation.

A political scientist would like to assign various levels of education to high income voters. Thus, some high income voters would have low levels of education (e.g., through the tenth grade), others would have a somewhat higher level of education (e.g., high school graduate) and others a still higher level of education (e.g., college graduate). As all voters with high incomes would not have the same level of education, this would eliminate the multicollinearity between income and education. Needless to say, "assigning" levels of education is not possible. For example, how could a political scientist remove four years of education from a voter? While a biologist can often set the level of each independent variable for each observation (i.e., each plant) and hence eliminate multicollinearity, a political scientist is unlikely to be in a similar situation. However, as we will see later, political scientists using nonexperimental research designs can still "control" for the impact of each independent variable on the dependent variable. We just do it statistically rather than by setting the level of each independent variable. Furthermore, suppose no voter in our sample had a doctorate in medicine. While a political scientist might like to study the effect of having a doctorate of medicine on someone's partisan affiliation, unless some members of our study have such a

degree, we will be unable to estimate the impact.

<u>A political scientist frequently encounters one additional problem: Did change in the independent variable precede change in the dependent variable</u>? In order for a change in income to "cause" a change in partisan affiliation (i.e., a voter's income increases from $40,000 annually to over $200,000 so they change from being a Democrat to a Republican), the change in income would have to occur <u>before</u> the change in partisan affiliation. The fact that most voters with an annual income of over $200,000 are Republicans may, or may not, mean that if a Democrat's income changes from $40,000 to over $200,000 they will become a Republican. The assumption of our model <u>is</u> that income change <u>precedes</u> partisan change. While this is plausible, it may not be accurate. It would be preferable to test according to the assumptions of our model. In this case that would literally mean we would have to change a voter's income and <u>then</u> see what, if anything, happened to their partisan affiliation. Obviously, we can <u>not</u> do this. As previously mentioned, the plant biologist is in a preferable situation because s/he can <u>first</u> administer the plant growth additive and <u>then</u> see how fast the plant grows.

The situations I have just described are the crux of the differences between an experimental and a nonexperimental research design. <u>To recap briefly, the previous analysis suggested three weaknesses of the nonexperimental research design relative to the experimental research design: (1) more severe multicollinearity (e.g., voters with high incomes were also likely to have high levels of education); (2) an absence of some possible levels of an independent variable (e.g., no one in our study of partisan affiliation with a doctorate of medicine); and (3) less confidence that change in the level of one of the independent variables preceded change in the level of the dependent variable (e.g., did a voter's partisan affiliation change before, or after, a change in their income</u>?). You might well have gotten the impression that since political scientists typically have to use a nonexperimental research design their findings are not very useful. Fortunately, this is <u>not</u> the case. Furthermore, the situation is improving.

<u>Let me now address each of the three problems mentioned above. First, in many studies the interrelationships between the independent variables are actually quite low (i.e., multicollinearity is quite low). Additionally, even when multicollinearity is rather high, we can often accurately estimate the impact of the interrelated independent variables</u>. For example, in a study of voting in the U.S. Senate the principle independent variable in which the researcher may be interested (the senator's political philosophy) is highly related to some of the <u>other</u> independent variables (e.g., the senator's partisan affiliation). Nevertheless, the findings concerning the impact of political philosophy are quite strong and reliable. Hence, even though multicollinearity <u>appears</u> to be a major problem, it is not. Furthermore, later in the course we will discuss strategies to deal with severe multicollinearity. A major topic of this course is how we "control" (i.e., set, or hold constant) the level of various independent variables. While our approach to isolating the unique impact of <u>each</u> independent variable on the dependent variable is not as desirable as that offered by the experimental design, it <u>is</u> nonetheless quite useful.

<u>The second problem of a nonexperimental research design is that we may not have observations on some scores for one, or more, of the independent variables</u>. For example, perhaps no voters in our sample have a doctorate of medicine degree. While potentially important, this problem is usually <u>not</u> catastrophic (terrible pun!). With large sample sizes we usually have several cases of each interesting score. In the partisan affiliation study, political scientists typically have samples of 2,500, or more, respondents. Even if we have few medical doctors in such a study, we probably have enough individuals with similar educational backgrounds (e.g., dentists) for useful statistical analysis. Furthermore, in many instances the omission of a particular category is <u>not</u> of critical importance. For example, it may not be important that we have no respondents with zero dollars in income. Even the poor have some income. It is probably not important to be able to generalize one's findings to situations which are extremely unlikely to ever occur.

<u>The third problem of a nonexperimental research design concerns the degree of confidence we can have that change in the independent variable(s) precedes change in the dependent variable</u>. Thus, did a change in the voter's income precede a change in their partisan affiliation? This is a serious problem. However, like the preceding two problems, the situation is <u>far</u> from hopeless. <u>In many practical research situations our theory is strong enough to be reasonably certain that change in the independent variable preceded change in the dependent variable</u>.

Suppose we are interested in the impact of a senator's political philosophy (the independent variable) on the probability that the senator will vote in favor of shifting the federal tax burden more toward higher income earners (the dependent variable). We can feel quite certain that the senator's political philosophy was formed <u>prior</u> to the time they voted on the tax shift. Few senators either enter the Senate without a political philosophy, or significantly change their political philosophy after they begin serving in the Senate. <u>Much</u> prior research has established the preceding point. In such situations, we can be quite confident that the level of the independent variable (e.g., the senator's political philosophy) was established <u>prior</u> to the score on the dependent variable (i.e., the direction of the senator's vote on the tax shift). <u>Additionally, political scientists have minimized this "time of change" problem through the increasing use of time series studies. A time series study means that the data are collected over time (e.g., annually - each year from 1950 to the present). By contrast, a study in which all data are collected at the same time point (e.g., all nations of the world in the year 2008) is called a cross-sectional study</u>.

For example, if we study the effect of political party control of the executive (i.e., whether the president is a Democrat or a Republican) on such economic outcomes as the unemployment rate or the growth rate, we would probably collect our data annually for a number of years. Obviously, we would know what the level of the independent variable was (i.e., the political party of the president) <u>prior</u> to any change in the dependent variable (e.g., the unemployment rate). Thus, with time series data we can often be more confident of our conclusions than with cross-sectional data. <u>A time series style study done by collecting data on the same individuals at several time points is called a panel study</u>. For example a famous panel study of political socialization (i.e., how people learn about politics) was done

interviewing the same people over several decades (M. Kent Jennings and Richard G. Neimi, Generations and Politics). Multiple interviews of the same person at several different time points (e.g., in 1986, 1996 and 2006) can give us a more valid view of the learning process than by interviewing someone as an adult and asking them to "recall" their youth.

Measuring Variables in Political Science

After formulating our research design, we need to measure our variables. Thus, we will need measures for each independent variable and the dependent variable. Our purpose is to classify outcomes on each variable. For example, in an international relations study we may need to measure the balance of power. How does the researcher do this? First, we need a "concept" of power. Second, we need an "operationalization" of power. For this discussion I will assume that we already have both a concept and an operationalization of power (see page 2 on the meaning of "concept" and "operationalization").

Once we have concepts and operationalizations for each variable, we can proceed to assigning mathematical values for each possible category on every variable. For example, if we conceptualize power as money and operationalize military power in terms of the defense budget, then we need categories for each possible outcome. In this case that would likely mean the amount of money (probably in U. S. dollars) spent by each nation over some specified period of time (probably annually). Perhaps, we should have conceptualized power differently. For example, if economic power is part of "defense" (or "offense"!), then perhaps the value of the gross national product would be a better indicator of "power" than military spending. While critical, the answers to such questions are often unique to a particular topic. Our consideration here is to assign mathematical values of outcomes on a particular variable.

There are four different "levels" of measurement. In the presentation that follows, each succeeding "level" retains all of the desirable properties of the preceding "level(s)," but adds some useful properties not contained in the preceding "level(s)." The weakest (i.e., "lowest") level of measurement is called nominal measurement. A nominal measure classifies each possible outcome. For example, the dependent variable in international relations studies is frequently whether or not war occurred. Let us say we scored "peace" as zero and "war" as one. This would be an example of a nominal level measure. Each outcome (peace or war) is mutually exclusive (i.e., can be only one category). Thus, peace can not be classified as war and vice versa. Furthermore, every outcome has a category. Thus, the only possible outcomes are peace or war. Hence, our measure is collectively exhaustive. When a variable has only two categories of responses (e.g., peace or war), it is termed a "dummy" variable. Occasionally, we will have a nominal level measure with more than two categories. For example, suppose we are studying voter attitudes on foreign policy (the dependent variable), one of our independent variables might be the respondent's race. Race would obviously have many more than two categories. Additionally, there is no inherent ordering to the categories. The coding would be entirely arbitrary. For example, would it make any more sense to code Asian-

American as zero, White as one, African-American as two than White as zero, Asian-American as one and African-American as two. No!! Thus race is inherently a nominal level variable.

<u>The ability to rank (i.e., order) categories of responses on a variable is a feature of the second level of measurement, ordinal level measurement</u>. For example, in a study of the foreign policy attitudes of voters (the dependent variable), political party affiliation of the voter might be a logical independent variable. Political party affiliation is often measured by what is termed a Likert scale (named for psychologist Rene Likert). Such a scale of partisan affiliation could be formulated as follows: (1) strong Democrat, (2) weak Democrat, (3) Independent, (4) weak Republican, (5) strong Republican. The preceding scale has an underlying order (hence "ordinal") or continuum. The continuum might be thought of as being from the most Democratic (category #1) to the least Democratic (category #5). Thus, a strong Republican is the least likely to support a Democratic candidate. Alternatively, a strong Democrat is the "most" Democratic orientation. <u>Ordinal measures add the ability to rank (or "order") to the mutually exclusive and collectively exhaustive traits of nominal level measurement. While such an "advance" is useful, we do not know the difference between the categories</u>. For example, is the difference between a strong Democrat and a weak Democrat the same as between a weak Democrat and an Independent? We have <u>no</u> way of knowing.

<u>The third level of measurement, interval level measurement, retains the mutually exclusive and collectively exhaustive properties of nominal level measurement, and the rank (or order) capabilities of ordinal level measurement. However, interval level measures also contain an equal mathematical difference between categories</u>. For example, supposing a political scientist were trying to explain the likelihood of someone voting. Weather might be a useful predictor variable. While some conceptualizations of weather would be nominal (e.g., it either rained or it did not), temperature would be an interval level measure. Temperature is an interval level measure because there is a constant unit of measure (a degree). Thus, the difference between 39 and 40 degrees is the same as the difference between 70 and 71 degrees. This equal unit of measure is lacking in ordinal level measures.

<u>A measure that has all the properties of an interval level measure (e.g., rank ordering and equal mathematical difference between categories) and has the added property that a score of zero indicates the absence of the phenomena being measured, is called a ratio level measure</u>. For example, since a temperature of zero degrees does <u>not</u> indicate the absence of temperature (i.e., a temperature of zero degrees does <u>not</u> mean that there is no temperature but rather a very cold temperature), temperature is <u>not</u> a ratio level measure (just keep reading). However, if we measure income by dollars, a score of zero <u>does</u> indicate the <u>absence</u> of money (i.e., no dollars). Thus, a score of zero dollars indicates the complete absence of dollars. For this reason, income measured by dollars is a ratio level measure whereas temperature is an interval level measure (the next paragraph will make it clearer).

Ratio level measures are even more useful than interval level measures because, as the name implies, we can form ratios. For example, we can say that someone with an income of $40,000 has twice as much income as someone with an income of $20,000 (i.e., a "ratio" of 2 dollars to 1 dollar). The reason we can say this is that a score of zero implies the absence of money (i.e., no money). However, since a temperature of zero degrees does not mean the absence of temperature we can not say that a temperature of 70 degrees is twice as high (or as warm) as a temperature of 35 degrees.

While ratio level measures are quite common in political science, interval level measures are relatively rare. Political scientists often measure variables in either percentages (e.g., percentage of times a nation resolves its' disputes with other nations by peaceful means - where zero indicates no conflict was resolved peacefully) or other scales in which zero indicates the absence of the phenomena (e.g., a person with zero years of education indicates they have no formal schooling).

A basic rule of statistical analysis is that any statistical technique usable with a lower level of measurement can be used with a higher level of measurement, but not the reverse. For example, if a statistical technique is acceptable for use with nominal level measurement, then it can be also be used with ordinal, interval or ratio level measures. However, if a statistical technique requires an interval level of measurement, then it can not be used with nominal or ordinal level measures.

While the difference between each level of measurement is important, the primary distinction is between the interval and ordinal levels. Thus, we may usefully think of measures as either interval (i.e., interval or ratio) or sub-interval (i.e., nominal or ordinal). As you will read in future assignments, the statistical techniques that require at least an interval level of measurement are much more desirable than those that require only nominal or ordinal levels of measurement. The increased precision that interval or ratio level measurement provides is very useful.

Not surprisingly, researchers have tried using techniques designed for interval level data with ordinal level data. This raises an important question: How serious are the consequences of treating ordinal level data as interval level data? As is often the case, the seriousness of the consequences differ depending upon the gravity of the violation. A good rule in this regard is: The more categories of responses and the more uniform the distribution of responses, the less serious the consequences of violating the interval assumption. For example, it would be preferable to have five categories of responses (e.g., strong Democrat, weak Democrat, Independent, weak Republican, strong Republican) as opposed to three categories of responses (e.g., Democrat, Independent, Republican). We can further minimize the severity of violating the interval assumption by having an approximately uniform distribution of responses. For example, with five categories of responses it would be desirable to have each category contain approximately 20% of the responses. Thus, if 50% of our sample selected response "A" but only 5% selected response "E" we could potentially have serious problems in treating ordinal data as interval data. However, if we have several categories of responses and a relatively uniform distribution of responses, it appears that we can relax the interval assumption without grave consequences (Herbert Asher, Causal Modeling, second edition, pp. 37, 90). In such situations, the advantages of interval level techniques probably outweigh the

consequences of violating the interval assumption.

Two important considerations in evaluating a measure of a variable are validity and reliability. Validity assesses how accurately we are measuring what we claim to be measuring. For example, if our measure of the balance of power says that two nations have the same degree of power, is this really true? A second, and related concept, is reliability. A reliable measure is one which, if applied time after time, will yield the same results (assuming no change in the level, or score, on a particular variable). For example, a reliable measure of unemployment will report the same incidence of unemployment in 2007 as in 2008 if indeed unemployment was the same in those two years.

It is useful to be able to distinguish between validity and reliability. Whereas a valid measure is always reliable, a reliable measure may not always be valid. For example, a gas gauge scale that consistently reports that your car has three more gallons of gas than it actually does is reliable, but not valid. Your car always has three less gallons of gas than the gauge suggests. However, since the gas gauge always reports your car as having three more gallons of gas than it actually does, the gauge is reliable.

Generalizing Our Results

As mentioned previously (pages 2 - 5), one of the tenets of the scientific method is to try and generalize our results. Thus, is what occurs in a city similar to what occurs in a state? At this point it would be useful to discuss what is termed the "unit of analysis." The "unit of analysis" is what we collect data on. For example, if we survey individual voters the "unit of analysis" is the individual. On the other hand, if we are using the unemployment rate for the entire United States, the nation is the "unit of analysis." Suppose we are interested in explaining variation in statewide voter turnout rates. As literate individuals are more likely to read political information (and hence be more political "involved"), a reasonable hypothesis might be that literacy (the independent variable) and voter turnout rates (the dependent variable) are "positively" associated. Thus, as literacy increases, voter turnout rates would be expected to increase. Suppose we have the following statewide data on the percentage of adults who are literate in the state and the percentage of the registered voters who voted in the last election:

State	Percent Literate	Percent Voted
California	90%	90%
Kansas	70%	70%

While tempting, you should not interpret the above data to conclusively support a hypothesis that the more literate an individual is, the more likely they are to vote. For example, the 90% "voted" figure for California could have occurred by having 100% of the 10% illiterate in California vote and only approximately 80% of the 90% of adult Californians who are literate vote. In order to infer to the behavior of

individuals the data should be collected on individuals. This would mean that <u>individuals</u> were surveyed in both of the above states and we knew whether or not <u>each</u> individual was literate <u>and</u> whether that <u>same</u> individual voted. The above data are statewide. <u>Inferring from one "unit of analysis" (here statewide) to another "unit of analysis" (here individuals) is called the ecological fallacy</u>.

<u>Make sure you do not confuse the "unit of analysis" with other concepts previously discussed. For example, do not confuse the "unit of analysis" with the "level of measurement." There is no necessary relationship. We could collect nominal level data on either an individual, a state, or a nation. Furthermore, do not confuse the "unit of analysis" with either a "variable" or the "number of observations</u>". If we are collecting data on the education and political philosophy of 100 individuals, there is one "unit of analysis" (the individual), two "variables" (the individual's level of education and their political philosophy) and 100 "observations" (100 scores on each of the two variables). <u>Moreover, do not confuse the "unit of analysis" with the subject of the study</u>. The subject of the study might be the impact of education (the independent variable) on political philosophy (the dependent variable). The "unit of analysis" is still the individual because the data are collected on individuals. <u>Finally, there is no relationship between the "unit of analysis" and the type of research design the researcher uses</u>. Regardless of whether the researcher uses an experimental or a non-experimental research design, there is always a "unit of analysis." In the above example, if the research could set the level of an individual's education, they would be using an experimental research design. If the researcher could not set the level of an individual's education, they would be using a non-experimental research design. In either event, the individual is still the unit of analysis. Since it is highly unlikely a researcher would be able to either add or subtract years of education from an individual, such a study would invariably use a non-experimental research design.

<u>On quizzes I often ask people to write a hypothesis. Do not write something such as: liberal senators do not support the rich. If you only include "liberal senators," then we have a constant instead of a variable because all the senators you studied would be liberals (just keep reading). Remember that in order to be a variable, a concept (such as political philosophy) must be able to assume more than one value or score (such as: liberal or conservative - two different values or scores). Thus, a better formulation would be: liberal senators are less supportive of wealthy taxpayers than conservative senators. The use of "less" (or "more") conveys the notion of probability</u>. While a liberal senator may support wealthy taxpayers, they are less likely to support them as frequently as conservative senators. <u>Better still, think in terms of a continuum (or gradation) of scores. Thus, all liberals are not equally liberal just as all conservatives are not equally conservative. Therefore, a better phrasing of the above hypothesis would be: the more liberal the senator, the less supportive they are of wealthy taxpayers. The degree of liberalism of the senator is the independent variable while the senator's degree of support for wealthy taxpayers is the dependent variable</u>. This phrasing allows for multiple categories of both liberalism (a senator could be very liberal, somewhat liberal, or not very liberal - i.e., rather conservative, etc.) and support for wealthy taxpayers (much support, some support or perhaps no support).

Descriptive Statistics

The purpose of this section is to introduce you to descriptive statistics. While a quiz on this material may involve some calculation, you do not need to memorize any formulas or bring a calculator to class. Do not panic! The small amount of math that is involved is explained step by step. Honestly, if you can add "5" and "3" you should have little trouble.

The purpose of descriptive statistics is to summarize and illuminate some important characteristics of the data. For example, suppose you listen to a Presidential debate and hear several candidates discuss tax rates. Perhaps one of the candidates mentions tax rates in both the United States and various foreign countries. Perhaps the discussion sparks an interest on your part. So, you decide to assess tax rates in the approximately 160 (or more) nations of the world. What is the average tax rate of all nations of the world? How much variation (difference) is there in tax rates among nations of the world? These are the types of questions that descriptive statistics seek to answer. So the ensuing discussion will make more sense, let me mention that a nation's annual gross national product (GNP) is the total value of goods (e.g., cars, houses, etc.) and services (e.g., teachers' salaries, nurses' salaries, etc.) produced in that nation in a particular year. Descriptive statistics can summarize the data in that they would allow us to make a statement such as the following: In 1995, taxes represented approximately 35% of the value of the gross national product for the average nation. This certainly conveys information in a more useful form than would 160 slips of paper each containing the percentage of the gross national product represented by taxes for a different nation. Thus, we have summarized the data. This becomes even more imperative with larger data sets (e.g., 3000 respondents to a national opinion poll).

A first step in uncovering some salient characteristics of our data is to assess how many and what proportion of the scores are between two points. For example, how many nations have taxes equal to between 20%-29% of their gross national product? Additionally, this number of nations is what proportion of the total number of nations? A frequency distribution is a method of helping us answer such questions. To construct a frequency distribution of national tax rates for 160 nations in the year 2008, we would need to divide the total dollar amount of taxes collected in a nation in 2008, by the dollar value of the gross national product in the same year for the same nation. For example, if all governments in the U.S. collected one trillion dollars in taxes in 2008, and the value of the U.S. gross national product was three trillion dollars in 2008, the tax rate for the U.S. in 2008, would be .33 (1 trillion dollars / 3 trillion dollars = 1/3 = .33). Since .33 is a proportion, we need to multiply it by 100 in order to obtain a percentage (percentage means "per hundred"). As (.33)(100) = 33, this would mean that in 2008 taxes in the U.S. were equal to 33% of our gross national product. Using this same procedure for the additional 159 nations might produce a frequency distribution as follows (if the table on page 15 does not make sense, just keep reading - there is a discussion of the table immediately after the table appears):

Table 1
Frequency Distribution of International Tax Rates
(Hypothetical Data)

National Taxes as a Percentage of National GNP	Percentage of Nations	Number of Nations
70% or more	0%	0
60 - 69	10	16
50 - 59	15	24
40 - 49	40	64
30 - 39	31	50
20 - 29	4	6
0 - 19	0	0
	100%	160

There are several important points to remember in constructing a frequency distribution. First, have a descriptive title. The title should convey to the reader what information the table contains. Second, use appropriately sized categories. For example, suppose the above table had only two categories, less than 30 percent and 30 percent or greater. If so, 96% of the nations would have been in the same category (30 percent or greater). As the data in Table 1 show, such a categorization scheme would have concealed a large amount of variation. Notice the percentage of nations in each category above 39 percent. All this information would have been concealed if one category was used for all nations with taxes equal to 30 percent, or more, of the gross national product. Third, always present both the percentage (column 2) and the frequency (column 3). Percentages "standardize" the data (just keep reading - it will be clear shortly). For example, suppose that we wanted to compare how likely a nation was to have taxes consume between 50% and 59% of that nation's gross national product in 2008 and in 1958. According to the table above, in 2008 there were 24 nations out of the total of 160 nations, or 15%, (24 is 15% of 160) that fit this criteria. Suppose we found that in 1958, 24 nations out of the total of 120 nations, or 20%, (24 is 20% of 120) fit this same criteria. While the number of nations, 24, is the same in both 2008 and 1958, the percentage of nations in which taxes consumed between 50% and 59% of the gross nation product was lower in 2008 than in 1958 (because 15% is lower than 20%). If we did not adjust for the different number of nations in the two time periods (160 vs. 120), we would not have known this. Expressing the number 24 as a percentage placed both 2008 and

1958 on a common measuring scale which then revealed a difference between the two years that would not have been apparent had we just examined the number of nations in the 50% to 59% category in both years. As the reader may want to reorganize the data, it is also important to show the frequency (i.e., number of nations) in each category.

Measures of Central Tendency and Measures of Dispersion

While frequency distributions are important, they do not convey all the salient characteristics of a variable. Two additional questions that would be useful to answer are: (1) What is the average score; and (2) How representative is the average score of all the scores?

We often think in terms of an "average." You may assess how well you scored on a test by comparing your score with "the average." There are several different measures of "the average." The statistics pertaining to the average are termed "measures of central tendency." There are three commonly used measures of central tendency. The appropriate measure depends upon three factors: (1) the question you want to answer; (2) the level of measurement of your data; and (3) the distribution of the scores. Remember from the notes on the level of measurement that while a nominal level measure categorizes the data (e.g., pear, oak and elm are three categories of trees), it neither orders the data (e.g., are pear trees "higher" than oak trees?) nor provides an equal unit of measure between categories (e.g., is the difference between pear and oak the same as between oak and elm?). Given these limitations, the only method of measuring the average for a nominal level measure is to see which category occurs the most frequently. Such a measure is called "the mode." For example, the mode for the following scores: 2, 3, 4, 5, 5, and 6 is 5. If two different scores are equally numerous then we have what is termed a "bi-modal" distribution (just keep reading - the next sentence will make it clear). For example, suppose we had the following scores: 1, 1, 2, 3, 4, 5, 5. Since "1" and "5" both appear twice and no other score appears more than once, both "1" and "5" are modes. As there are two modes, the scores are "bi-modal."

If our data are ordinal (on "ordinal" see page 11), in addition to the mode, we may also employ a second measure of central tendency, the median. The median is the number that divides the distribution into two equal parts. For the following 25 scores: 0, 1, 1, 2, 2, 2, 2, 3, 3, 3, 4, 4, 5, 6, 6, 6, 6, 6, 7, 8, 9, 75, 100, 125 and 300 the median is the "middle" score (i.e., if 25 scores then the 13th score from the left) which is 5. Note also that the mode for this data is 6. If we had an even number of scores the median would be the mid-point between the two middle scores. For example, with ten scores the median would be midway between the fifth and sixth scores.

Look again at the 25 scores that we just used in the example of the median. Notice that the last four scores (75, 100, 125 and 300) are much higher than the other scores. However, if a score is the same (or higher) than the middle score, the median is unaffected (e.g., if the last four scores were 14, 15, 16 and 17 the median would still be 5). It is this feature of the median that makes it usable with ordinal level data. Remember that ordinal level data tell the rank, but not the precise numerical

difference between the scores. Thus, if we just want to select the middle score, all we have to know is the ranking of the scores, <u>not</u> the degree of difference between the scores. However, if we have interval or ratio level data (see pp. 11-12), we also know the numerical difference between the categories. It would seem natural to develop a measure of the average that took account of the numerical differences between the scores. This is the idea behind the mean. <u>The mean of a group of scores is the total of the scores divided by the number of scores. For example, the mean of the following scores: 10, 20 and 60 is 30 (because 10 + 20 + 60 = 90 and 90/3 = 30). Hence, the mean tells us the average per score</u>.

In the calculation of the median, we used the 25 scores listed above. The mean of these 25 scores is 27.44 (because if you add the 25 scores above together they total 686 and 686/25 = 27.44). Obviously, the median (5) is quite different than the mean (27.44) for this data. This great difference between the median and the mean occurs because the highest four scores are so much greater than the other 21 scores. While the size of the numerical difference between scores has no effect on the median, it has a great impact on the mean. Remember that one of the criteria for choosing a measure of central tendency is the distribution of the data. Because the extreme scores distort the mean value of the 25 scores listed on the previous page, I would recommend using the median as the measure of central tendency for the above data set. However, note that if we had not had interval or ratio level data, we could not have calculated the mean. Hence, we would have had to use either the median or the mode as our measure of central tendency. Additionally, if you wanted to know which score occurred the most frequently, you would use the mode as your measure of central tendency.

<u>By providing a notion of "the average," measures of central tendency reveal important characteristics of the data. However, a measure of central tendency does not tell us how representative the average is of all the scores in the distribution</u>. For example, if you were a professor and found that the mean score on an exam was 50, I think you would want to know if the mean of 50 occurred because nearly all students scored approximately 50, or because perhaps half the students scored 0 and the other half scored 100. In both circumstances the mean would be the same, but the ramifications for how you taught the course would be entirely different. This example illustrates the need for what are termed "measures of dispersion." <u>The purpose of measures of dispersion is to assess how representative the average is of all the scores in the distribution</u>.

<u>The simplest measure of dispersion is called "the range." The range is the difference between the highest and lowest score</u>. As the range implies order, we would need at least an ordinal level of measure to compute the range. While the range is simple, it does not convey as much information as we might like. For example, suppose we return the glorious 1980s when American life was good, pure and Donald Trump was in financial ascendence. Relaxing in the plush surroundings of Trump Tower, I joyously calculate Sir Donald's income for the current year. When I compare his income to that for each other American family, I discover that not only is Trump's income higher than that for any other American family, but that no other American family is within $1,000,000 of Trump's income. If we were to use the range as a measure of dispersion, we would simply take the difference between Trump's

income and the lowest family income. However, by using only two scores we have masked the important information that Trump's income was by far the highest. Thus, the range is insensitive to any scores except the two most extreme (highest and lowest). We can minimize this problem somewhat by using several additional data points. For example, in addition to the highest and lowest scores we could include the score wherein 1/4 of the families were above and wherein 1/4 of the families were below. Such a measure is referred to as the "interquartile" range.

While the interquartile range conveys more information than the range, it would be desirable to have a measure of dispersion based upon all the scores in the distribution. Intuitively, we might think that we could measure dispersion by subtracting the mean from each score. The difference between a single score and the mean would indicate the amount of "deviation," or "dispersion" of that particular score from the mean. If we summed (added) each of these "dispersions" we would have the total amount of dispersion present in our data. We could then divide this total by the number of observations in order to obtain a typical deviation (i.e., the deviation per score). While there is a mathematical "problem" in this method, the procedure we have just outlined is the basis of the approach we will ultimately use.

The mathematical "problem" with the approach we just formulated is that it must result in an answer of zero. This is because one of the properties of the mean is that the total amount of "distance" below the mean must be equal to the total amount of "distance" above the mean. For example, the mean of the following scores: 4, 6, and 8 is 6 (because 4 + 6 + 8 = 18 and 18/3 = 6). The "average" of the deviations of these same scores would be "0" [because 4 - 6 = -2, 6 - 6 = 0, 8 - 6 = 2; adding these deviations equals 0 (i.e., -2 + 0 + 2 = 0) and then dividing this total by the number of scores yields "0" (i.e., 0/3 = 0)]. The value of the "positive" deviations from the mean (i.e., scores higher than the mean) equals the value of the "negative" deviations from the mean (i.e., scores lower than the mean) and thus the "total" deviation from the mean must equal zero. Therefore, we are left with the impression that there is no deviation (i.e., no variation) in our data. Thus, we would mistakenly conclude that the mean occurred because every score was the same.

This problem can be rectified by taking the "absolute" value of each deviation from the mean. Hence, a deviation of -2 would be treated as a deviation of 2. As we would be adding a series of positive numbers, the cancellation problem I just discussed would not occur (i.e., in the example above remember that -2 + 0 + 2 = 0 and hence our deviation measure ended up as 0/3 = 0; taking "absolute" values would have instead produced a typical deviation of 1.33 because 2 + 0 + 2 = 4 and 4/3 = 1.33) This revised procedure (i.e., using "absolute" values) is called the average deviation. While such a measure tells us the average amount of deviation from the mean for a typical score, the result is still not as useful as we might prefer. For example, suppose we calculated the average deviation and found that it was 5.7. What statement(s) could we make? We could say that the typical score deviated 5.7 units (in whatever units the variable was measured, dollars, percentage points, etc.) from the mean.

We could further amplify the preceding approach by taking the average deviation as a percentage of the mean (even if you are confused, just keep reading through the end of this paragraph). Thus, an average deviation of 5.7 would seem

<u>small if the mean were 1,000, but large if the mean were 10</u>. Therefore, an obvious approach would be to divide the average deviation by the mean (just keep reading). This would yield a "proportion" which we could then multiply by 100 in order to obtain a "percentage" (just keep reading). For example, an average deviation of 5.7 is 57% of a mean of 10 (5.7/10 = .57 and (.57)(100) = 57). However, an average deviation of 5.7 is only .57% (approximately 1/2 of 1%) of a mean of 1000 (5.7/1000 = .0057 and (.0057)(100) = .57). If the average deviation is 57% of the mean, it tells us that the mean was achieved by many scores being quite far from the mean (e.g., the mean of 0 and 10 is 5 because (0 + 10)/2 = 10/2 = 5 but neither 0 nor 10 is very close to 5). On the other hand, if the average deviation is only 1/2 of 1% of the mean, this tells us that virtually all the scores are quite close to the mean.

While the aforementioned procedure is a definite improvement over the range, we can still do better. <u>A critical question to ask with the result from any statistic is: How can you interpret the answer</u>? In the last paragraph I outlined an approach to interpreting the average deviation. However, we could improve upon my approach if, in addition to taking the average deviation as a percentage of the mean, we could also compare the average deviation to a known distributional formula. For example, suppose the mean of a group of scores was 50 and the average deviation was 5. Using my approach we could say that the average deviation was 10% of the mean [5/50 = .10 and (.10)(100) = 10]. This would suggest that the mean was achieved by most scores falling pretty close to the mean. Put differently, when we consider the average deviation relative to the mean, dispersion seems low. <u>However, it would be even more informative to give a percentage distribution of the scores. For example, it would be desirable to be able to say that approximately 68% of the scores were within one average deviation of the mean</u>. Given our results (mean = 50, average deviation = 5) this would mean that approximately 68% of the scores would fall between 45 and 55. <u>Unfortunately, we can not make such a statement with the average deviation. In order to obtain a percentage distribution of the scores we need to calculate the standard deviation</u>.

<u>While our purpose is to illustrate the calculation and use of the standard deviation, this is a good point to introduce some statistical symbols and mathematical procedures. Do not panic! While a quiz on this material may involve some calculation, you do not need to memorize any formulas or bring a calculator to class</u>. <u>I will provide any formulas which are needed on the quiz. You only must know how to work the formulas on pages 21-24. Just follow how I work the computations on pages 21-24 (i.e., how I calculate the various answers). Finally, you will not have to calculate a square root</u>. The quizzes on this material will have you working largely with single digit numbers. <u>For example, you might have to subtract 3 from 5 (i.e., 5 - 3 = 2). Wasn't that difficult? If we are feeling "adventurous" later on we might actually try multiplying 2 times 2!! Wow!!! Do you really need a calculator to do these computations? I hope not</u>! When people have trouble on a computational quiz it is almost always because they do not know the order of mathematical operations. For example, do you add and then multiply or multiply and then add? A calculator can not help you with the order of mathematical operations. The calculator assumes you know the order of operations. That is why a calculator is of such little value on the type quiz you will take.

<u>The standard deviation is undoubtedly the most often used measure of dispersion in political science</u>. As just mentioned, the standard deviation will permit us to make what I will term "percentage distribution" statements. Just keep reading!! For example, let us say that you are studying international relations. International relations scholars have often quantitatively tested interesting models concerning the factors (i.e., independent variables) that explain why some nations are more likely to resolve their conflicts peacefully than other nations. The dependent variable in this type of study might be the percentage of times a nation resolves its disputes peacefully. Peaceful resolution would mean that a dispute was resolved without violence. Typically, international relations scholars pretty much agree on what constitutes a "dispute." So, through historical records we construct the number of disputes that each nation has been involved in over the last 100 years. While the name and boundaries of many nations have changed over the past 100 years, we could still obtain data for a large number of nations. <u>Using the historical record,</u> **<u>(continued on the next page</u>)**

<u>NOTE</u>: Odd page breaks will occur when there are formulas. This material was originally written in Word Perfect and the equations do not automatically transfer from one word processing package to another. I don't know how to use equation "boxes" in Microsoft Word. Additionally, in some cases there are drawings that I paid to have made that I want to transfer to this edition of the material. So, just be prepared for an occasional page fragment!

we calculate the percentage of disputes which each of these nations resolved in a peaceful manner. To simplify the presentation, let us examine just two regions of the world. If each of the two regions contained 5 nations, the data might look something like the table below.

Percentage of Disputes Resolved Peacefully

Region 1		Region 2	
Nation A	50%	Nation F	68%
Nation B	60%	Nation G	69%
Nation C	70%	Nation H	70%
Nation D	80%	Nation I	71%
Nation E	90%	Nation J	72%

Like the average deviation, the standard deviation is based upon how each score deviates (hence "deviation") from the mean (i.e., the average score). Accordingly, before we can calculate the standard deviation, we must first calculate the mean. Do not "freak" when you see "strange" symbols! It is all explained over the next page and a half. So, just keep reading! The symbol and formula for the mean are shown immediately below.

$$\text{Mean of Variable } X = \overline{X} = \frac{\Sigma X}{N}$$

$\overline{X}$ is the symbol for the mean of variable X. If we had designated this variable as variable Y, then the symbol for the mean would, not surprisingly, be $\overline{Y}$. The Σ symbol above is called a "summation operator." The "summation operator" tells us to add the scores. "N" is the symbol for the "number of scores." To obtain the mean score for region 1, the computation would be as follows:

$$\text{Mean of Region 1} = \overline{X_1} = \frac{\Sigma X_1}{N} = \frac{50 + 60 + 70 + 80 + 90}{5} = \frac{350}{5} = 70$$

To obtain the mean score for region 2 the computation would be as follows:

$$\text{Mean of Region 2} = \overline{X_2} = \frac{\Sigma X_2}{N} = \frac{68 + 69 + 70 + 71 + 72}{5} = \frac{350}{5} = 70$$

The mean score for both regions is the same, 70. This means that in both regions the average nation resolved disputes peacefully 70% of the time. However, while the mean score for both regions is the same, it is obvious from the data [redacted] that the scores varied more (i.e., were more different) in region 1 than in region 2. Put another way, the mean in region 2 is more representative of all the scores in region 2 than the mean in region 1 is of all the scores in region 1 (thus the mean of 70 is closer to numbers like 71 and 72 as in region 2 than 70 is to numbers like 50 and 90 as in region 1). The standard deviation is a method of summarizing how much a group of scores differ. For example, how much do the scores in region 1 differ from each other? The standard deviation helps us answer such a question. Do not "freak" when you see the following formula for the standard deviation. Each step will be carefully explained.

$$\text{Standard Deviation of } X = S_x = \sqrt{\frac{\text{Variation of } X}{N}} = \sqrt{\frac{\Sigma(X-\bar{X})^2}{N}}$$

Do not worry if you get "lost" on the operations below. Just keep reading! You will see them performed shortly. Here is the order of operations in the above formula:

1. The order of mathematical operations says to perform the computations inside the parentheses first. It is clear that we are to subtract the mean ($\bar{X}$) from something. Obviously, we would first have to know what the mean was in order to subtract it. Therefore, our first step is to calculate the mean. We did this on page 22. We know that the mean score for both regions is 70.

2. The symbols ($X-\bar{X}$) tell us to subtract the mean (70) from each score.

3. After working inside the parentheses, we move outside the parentheses. The symbol "2" in the upper right hand side of the expression $(X-\bar{X})^2$ tells us to square each of the entries in step 2. Since there are five nations in region 1, we would have 5 "entries" for step 2 (i.e., subtracting the mean from nation A, then subtracting the mean from nation B, etc.). We would now square each of these five entries. Squaring means to multiply a number times itself. Thus, take nation A. Look at the table on page 20. The score for nation A is 50. If we are subtracting the mean (70) from this score it reduces to 50-70 which is -20. To square -20 we multiply -20 times -20. This is also stated as (-20) (-20). A negative number multiplied, or divided, by a negative number yields a positive answer. So (-20) (-20) will produce a positive answer, 400. The answer, 400, could be more simply attained by decomposing one of the -20's. Thus, (-20)(-20) = (-20)(-10)(2) = (200)(2) = 400. So, our first squared deviation is 400. We repeat this process for each of the other four nations in region 1.

4. The summation operator tells us to add each of the five square deviations we obtained in step 3. It is critically important that we square before we sum. Watch for a quiz the first day this material is due that will test whether you know to square before you sum.

5. We divide the total we obtained in step 4 by the number of scores we added (i.e., 5 - since there are five nations in each region).

6. Take the positive square root of the answer in step #5. Squares and square roots are opposites. The sign for a square root is: $\sqrt{\ }$ Thus 4 "squared" is 16 because (4) (4) = 16. Additionally, 4 is the square root of 16. Note that (-4) (-4) also equals 16. Thus, both (4) and (-4) are square roots of 16. In calculating the standard deviation always take the positive square root (i.e., 4 instead of -4).

The table below shows the computations necessary to calculate the standard deviation of region 1 and region 2. The scores in the column marked "X" are the same scores as appeared on page ~~20~~ 22. The symbols at the top of each column are those discussed on pp. ~~20-21~~ 22-23.

Standard Deviation of Region 1

Column 1	Column 2	Column 3	Column 4
Nation	X	$(X - \bar{X})$ the above says subtract the mean (70) from each score in column 2	$(X - \bar{X})^2$ this column is the square of the answer in column 3
A	50	(50 - 70) = -20	400 [because: (-20)(-20) = 400]
B	60	(60 - 70) = -10	100
C	70	(70 - 70) = 0	0
D	80	(80 - 70) = 10	100
E	90	(90 - 70) = 20	400

Σ = 1,000

(Σ means to add the numbers in column 4, thus 400 + 100 + 0 + 100 + 400 = 1,000)

$$\text{Standard Deviation of Region 1} = S_{x_1} = \sqrt{\frac{\Sigma(X-\bar{X})^2}{N}} = \sqrt{\frac{1000}{5}} = \sqrt{200} = 14.14$$

5 nations

Note: 14.14 is the positive square root of 200 because 14.14 is a positive number (i.e., 14.14 instead of -14.14) and (14.14) (14.14) = approximately 200. You will not have to calculate a square root on the quiz.

Standard Deviation of Region 2

Column 1	Column 2	Column 3	Column 4	
Nation	X	$(X - \bar{X})$	$(X - \bar{X})^2$	
F	68	(68 - 70) = -2	4	[i.e., (-2)(-2) = 4]
G	69	(69 - 70) = -1	1	
H	70	(70 - 70) = 0	0	
I	71	(71 - 70) = 1	1	
J	72	(72 - 70) = 2	4	
			Σ = 10	

$$\text{Standard Deviation of Region 2} = S_{X_2} = \sqrt{\frac{\Sigma(X-\bar{X})^2}{N}} = \sqrt{\frac{10}{5}} = \sqrt{2} = 1.41$$

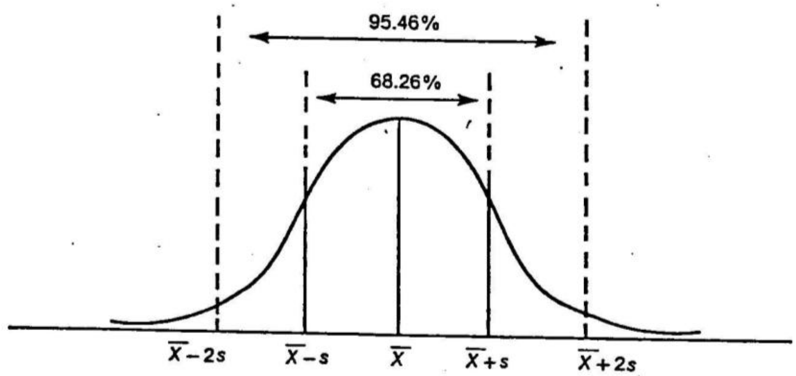
Notice that the standard deviation for region 1 (14.14) is 10 times the size of the standard deviation for region 2 (1.41). One useful method of interpreting the standard deviation is as a percentage of the mean. In region 1 the standard deviation is 20% of the size of the mean (i.e., 14.14/70 = .20 and .20 x 100 = 20). In region 2 the standard deviation is a mere 2% of the mean (i.e., 1.41 is 2% of 70). We could say that the scores were more "dispersed" (i.e., on average further from the mean) in region 1 than region 2. Expressing the standard deviation as a percentage of the mean is called the coefficient of variation. If the standard deviation is 15%, or less, of the size of the mean, we could say dispersion is low (i.e., the typical score is rather close to the mean). If the standard deviation is from 16% to 35% of the size of the mean we could say there is a moderate amount of dispersion. If the standard deviation is greater than 35% of the size of the mean we could say there is a high degree of dispersion (i.e., the typical score is somewhat far from the mean or, alternatively, the mean is not very representative of the typical score).

Another measure of dispersion that is very similar to the standard deviation is the "variance." The variance is literally the square of the standard deviation. For example, the variance for region 2 is 2 because 1.41 (the standard deviation of region 2) squared (i.e., 1.41 times 1.41) equals approximately 2.
Put another way, when we calculated the standard deviation, the result of the next to last step (i.e., just before we took the square root) was the variance. The formula for the variance appears below. Except for the square root sign, isn't it identical to the formula for the standard deviation? Yes!

$$\text{Sample Variance of } X = S_x^2 = Var(X) = \frac{\text{Sample Variation of } X}{N} = \frac{\Sigma(X-\bar{X})^2}{N}$$

Since I mentioned that you would not have to calculate a square root and that you would not need a calculator, a possible quiz question might be to give you the formula for the variance and a small group of scores that were easy to work with (e.g., numbers like 2, 4, etc.) and ask you to calculate the variance.

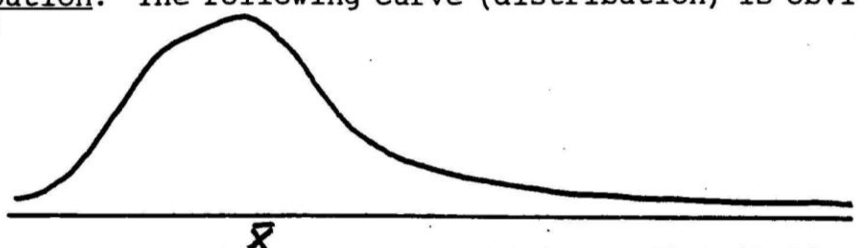
As I mentioned previously, the standard deviation can also be interpreted by using what I termed a "percentage distribution" capability (e.g., 68% of the scores are within - plus or minus - one standard deviation of the mean). Just keep reading it will become clear shortly!!! In order to understand this capability, let us introduce one of the very most important concepts in statistics, the normal distribution. A distribution is a group of scores. For example, because the readings in this course are so exciting (sure!), suppose you decide to go to graduate school in political science. Most doctoral programs in political science require you to take the Graduate Record Examination (GRE). Let us say we took all of the scores on the Graduate Record Examination for a particular year. A diagram of the scores might well look like the drawing below. Note that the height of the line indicates the frequency of occurrence. Thus, in the diagram below, the mean ($\bar{X}$) is the most frequently occurring score (i.e., the "mode") because the curve is highest directly over the mean score.



Several other properties of the normal distribution are also important to mention. First, the distribution is "symmetrical" (i.e., the portion to the left of the mean is of the identical shape of the portion to the right of the mean). Second, the mean, median and mode are all at the same point (i.e., same score). Third, approximately 68% of the scores are between one standard deviation above the mean (i.e., "$\bar{X}$ + s" in the above diagram) and one standard deviation below the mean (i.e., "$\bar{X}$ - s" in the above diagram). For example, if the distribution of scores on the GRE was normal (i.e., shaped as the drawing above), then if the mean on the GRE was 1,100 and the standard deviation was 100, approximately 68% of the scores would be between 1,000 and 1,200 (because 1,100 - 100 = 1,000 and 1,100 + 100 = 1,200). Fourth, approximately 95% of the scores would be between two standard deviations below the mean and two standard deviations above the mean. In our example, this would mean that 95% of the scores on the GRE would be between 900 and 1,300 (because 1,100 - 100 - 100 = 900 and 1,100 + 100 + 100 = 1,300). Furthermore, over 99% of the scores would fall within plus or minus three standard deviations of the mean. This is the "percentage distribution" capability of the standard deviation. The average deviation, which was discussed on pages 18-19, does not have such a capability. This is one of the major reasons why the standard deviation is by far the most commonly used measure of dispersion in political science.

Tchebysheff's Theorem

The scores on a variable may not always conform to a normal distribution. The following curve (distribution) is obviously non-normal:



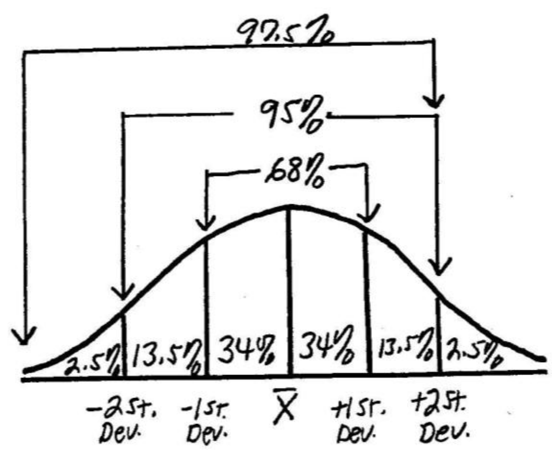
For example, the above curve could be a distribution of wages (e.g., many low - to the left of the "mean," and a few high - to the right of the "mean"). The curve is said to be "skewed" to the right. Note that the total amount of area on each side of the mean is the same. It is just "stretched" more to the right of the mean (or more "dense" to the left of the mean). By contrast, notice that the normal curve (page 26) is symmetrical (i.e., the portion of the curve on each side of the mean has the same shape).

In order to use the percentage distribution capabilities of the standard deviation when dealing with a non-normal distribution, we use Tchebysheff's Theorem. Tchebysheff's Theorem states that 75%, or more, of the scores must be within two standard deviations of the mean and 88%, or more, of the scores must be within three standard deviations of the mean. Thus, regardless of the shape of the distribution of the scores, we can use the percentage distribution capabilities of the standard deviation. As many variables are not normally distributed, Tchebysheff's Theorem makes the standard deviation much more useful. Note that Tchebysheff's Theorem does not imply that a distribution is symmetrical. For example, as in the curve drawn above, we may have more scores below than above the mean. In general, the scores above the mean are further from the mean than the scores below the mean. Thus, few scores further from the mean represent as much "area" under the curve as many scores closer to the mean. Note that the normal curve is consistent with Tchebysheff's Theorem. Tchebysheff's Theorem states that 75%, or more, of the scores must be within two standard deviations of the mean. With a normal curve 95% of the scores are within two standard deviations of the mean (see page 24). Isn't 95% either equal to, or greater, than 75%? Yes, 95% is greater than 75%.

Z (or Standard) Scores

Occasionally, you may want to compare scores from distributions with different means and standard deviations. For example, suppose that you apply to graduate school in political science and one of the schools you apply to gives the option of submitting either a score on the Graduate Record Exam or the Miller Analogies Test. Let us say that you scored 1,200 on the Graduate Record Exam and 50 on the Miller Analogies Test. Which is the "better" score? Thus, which score should you submit? It might be a mistake to "automatically" assume that the "higher" score (i.e., the Graduate Record Exam score) is the "better" score. A good way to see which is the "better" score is to compare each score to the mean score on that particular test. This is the essence of what is called a "Z" (or "standard") score. For example, if scores on the Graduate Record Exam are normally distributed with a mean of 1,100 and a standard deviation of 100, your score of 1,200 is one standard deviation higher than the mean (i.e., 1,200 = 1,100 + 100). However, if the Miller Analogies Test has a mean of 40 and a standard deviation of 5, your score of 50 is two standard deviations above the mean (i.e., 50 = 40 + 5 + 5). Let us think of this in terms of the normal distribution. As we discovered on page 24, approximately 68% of the scores in a normal distribution are within either one standard deviation above or below the mean. Since, the normal curve is symmetrical (meaning each half of the curve is of the same shape), half of this 68% of the scores (or 34%) are below the mean and half are above the mean. Thus, 34% of the scores are between the mean and one standard deviation above the mean. Remember that your score on the Graduate Record Exam was also one standard deviation above the mean.

To assess how well you scored on the Graduate Record Exam we could use the following analysis: (1) since you scored above the mean, we know that at least 50% of the scores were lower than yours (i.e., the 50% that were either at the mean level or below); (2) as 34% of the scores are between the mean and one standard deviation above the mean, then approximately 84% of the scores were below yours (i.e., the 50% that were at the mean level and below plus the 34% that were between the mean and one standard deviation above the mean. Since your score on the Miller Analogies Test was two standard deviations above the mean, the analysis would be as follows: (1) since you scored above the mean, we know that at least 50% of the scores were below yours; (2) since 95% of the scores are between two standard deviations above and below the mean and the normal distribution is symmetrical, then approximately 47.5% of the scores (i.e., half of 95%) are between the mean and two standard deviations above the mean; (3) therefore, approximately 97.5% of the scores on the Miller Analogies Test are below your score (i.e., the 50% that were at the mean level and below plus the 47.5% that are between the mean and two standard deviations above the mean). The diagram below pictures the aforementioned analysis.



Although your "raw" score on the Graduate Record Exam (1,200) is much higher than your "raw" score on the Miller Analogies Test (50), the Miller Analogies Test was your "better" score relative to those who took each test. The formula for the "Z" (or standard score) utilizes the reasoning we just followed. A "Z" score is simply how many standard deviations a score is above the mean (if the Z score is positive) or below the mean (if the Z score is negative). The formula is shown below.

$$Z = \frac{X - \overline{X}}{\text{Standard Deviation of } X}$$

In this example, your Z score on the Graduate Record Exam was 1.0 [because (1,200 - 1,100)/100 = 100/100 = 1] and your Z score on the Miller Analogies Test was 2.0 [because (50 - 40)/5 = 10/5 = 2]. Since 2.0 is greater than 1.0, your Miller Analogies score is your better score. Here are some important points to remember: a positive Z score means the score is above the mean (i.e., above average); a Z score of zero means the score equals the mean (i.e., average); a negative Z score means the score is below the mean (i.e., below average). Just keep reading, examples are ahead.

Suppose, for example, that your Miller Analogies score was 35 (instead of 50). This would have resulted in a Z score of -1.0 [because (35 - 40)/5 = -5/5 = -1] and placed you in the 16th percentile [i.e., in the diagram on page 26 notice that only 16% (2.5% + 13.5% = 16%) of the area under a normal curve is further than one standard deviation below the mean]. A score of 40 on the Miller Analogies Test would have yielded a Z score of 0 [(40 - 40)/5 = 0/5 = 0)] and placed you at the 50th percentile (i.e., half scored higher and half scored lower - sounds like the median - which is also the same point as the mean and mode in a normal distribution). A score of 45 on the Miller Analogies Test would have resulted in a Z score of 1.0 [(45 - 40)/5 = 5/5 = 1] and would have placed you in the 84th percentile (above both the 34% between your score and the mean score and the 50% who score at, or below, the mean: 34% + 50% = 84%).