Political Science 104, Spring 2009
Final Review

The major topics covered in class are listed below. You should also take a look at the readings listed on the class website.

Studying Politics Scientifically (Week 2: April 6)

Major Points Covered:

- Our goal is to uncover the truth, not advocate a particular position.
- We will be conducting *inference*: using known facts about the world to learn about facts we don't know. Descriptive inference is learning something about the state of the world we didn't know before. Causal inference is learning something about how the world works that we didn't know before.
- We use the scientific method to make our inferences. (1) Begin with a theory, (2) form a hypothesis based on the theory, (3) gather data to test the hypothesis, (4) test the hypothesis, (5) reject or fail to reject (accept) the hypothesis, (6) integrate the results of our test into our theory.
- The above point is the deductive approach – going from the general theory to a specific test. There is also the inductive approach – going from specific observations to a general theory. Both are valid approaches to science.
- Quantitative research is focused on gathering a large number of observations and analyzing them statistically. Qualitative research is focused on in-depth description of one or a few observations.
- The characteristics of scientific knowledge: (1) empirical (based on data), (2) replicable (if I explain my procedure, you can replicate my study), (3) explanatory (about the state of the world or how the world works), (4) probabilistic (we can never have 100% certainty about our findings).
- Gathering data: An observation is a single unit or thing you are studying. We combine observations into data. The unit of analysis is the type of thing you are studying. A variable is a characteristic of an observation that can vary across observations.

Hypothesis Formation (Week 2: April 8)

Major Points Covered:

- Most of our hypotheses are focused on causal relationships ("X causes Y" – education causes voter turnout to increase, for example). Know how to identify the *dependent* variable in the hypothesis (the variable being affected) and the *independent* variable in the hypothesis (the variable influencing the dependent variable). Be ready to read an abstract or two and identify the hypothesis and its constituent parts.

- The 9 characteristics of a good hypothesis. The two most important is the hypothesis must be *testable* and *falsifiable* – we can gather data to test our hypothesis, and it must be possible for our hypothesis to be wrong.
- Two other important characteristics of good hypotheses is they are: *empirical* (focused on relationships in the real world rather than normative concerns) and *plausible* (there is a plausible *causal mechanism* that leads us to expect the hypothesized relationship).
- Other more minor characteristics of good hypotheses are they are: *in declarative form*, *about an expected relationship*, *guided by theory and past work*, *brief and direct*, and *general without being unambiguous*. We defined these terms more precisely in class.

Determining Causality (Week 3: April 13)

Major Points Covered:

- "Causality" means that changes in our independent variable cause changes in our dependent variable. *Deterministic causality* means that changes in the independent variable *always* cause changes in the dependent variable. *Probabilistic causality* means that changes in the independent variable *usually* cause changes in the dependent variable. In the social sciences we usually talk about probabilistic causality.
- The *treatment effect* is the difference in the dependent variable we would see under two different values of the independent variable. The *fundamental problem of causal inference* is that each observation has only one value on the independent variable at any given time, so it is impossible to know with complete certainty if things would have been different if the independent variable was different.
- There are four necessary (but not sufficient) conditions we can observe that might tell us if one variable has a causal influence on another variable. (1) A *correlation* between the dependent and independent variables, (2) *temporal ordering*, or observing our hypothesized cause happening before our hypothesized effect, (3) *a plausible causal mechanism*, or a reasonable story as to why the independent variable should cause the dependent variable to change, and (4) *ruling out alternative explanations* through careful research design.
- There are four kinds of research designs to consider: (true) experiments, natural experiments, quasi-experiments, and observational studies.

Experiments (Week 3: April 15)

Major Points Covered:

- Experiments use *random assignment* to place observations in *treatment* and *control* groups. Then a treatment (the independent variable) is applied to the treatment group, and the values of the dependent variable are compared across the groups. Alternative explanations for any differences observed should be ruled out by the random assignment to groups.

- The characteristics of experiments are that they have *random assignment*, *treatment and control groups*, and *control over the values of the independent variable.*

- Experiments are high on *internal validity*, or the ability to determine if there is a causal relationship between the independent and dependent variables in the study. They are lower on *external validity*, or the ability to generalize the results of the study to the population we really care about.

- There are *laboratory* experiments (that take place in a completely controlled environment) and *field* experiments (that take place in the real world).

- Threats to internal validity: (1) history (something uncontrolled happens between the treatment and measurement of the dependent variable), (2) maturation (subjects are changing over time), (3) testing (the experiment itself changes the dependent variable), and (4) demand characteristics (subjects change their behavior to please/annoy the researcher).

- Threats to external validity: (1) testing interaction effects (people act differently because they are being observed), (2) unrepresentative subjects (i.e., undergraduate students versus all voters), (3) spurious measures (the treatment only works in the experimental setting).

Observational Studies (Week 4: April 20)

Major Points Covered:

- Natural experiments are like observing an experiment run by nature. The characteristics of natural experiments are that they have *"as-if" random assignment*, treatment and control groups, but *no* control over the values of the independent variable by the researcher.

- In natural experiments observations are assigned to treatment and control groups by a random or *"as-if" random* process. "As-if" random assignment means that the assignment process is not related to any alternative explanations you hope to rule out.

- Quasi-experiments have treatment and control groups, but assignment to either group is not necessarily random or "as-if" random. Further, the researcher has no control over the values of the independent variable.

- In other observational studies there are no treatment and control groups, no random assignment, and no control over the independent variable. We simply gather data on something that has happened and test it. Observational studies are lowest on internal validity, and highest on external validity.

- Observational studies use *control variables* to try to rule out alternative explanations.

- Many of the same ideas apply to qualitative studies – careful research design can help rule out alternative explanations even when there is not enough data to use statistical controls for these alternative explanations.

Measurement Part 1 (Week 4: April 22)

Major Points Covered:

- Pay attention to the unit of analysis in your hypothesis. A mismatch between the unit of analysis in a hypothesis and the unit of analysis in the data used to test the hypothesis is known as *cross-level inference*, which can be problematic. The best example of this is the *ecological fallacy* – trying to study individual behavior with aggregate data can be misleading. For instance, rich people tend to vote Republican, but rich states tend to vote Democratic. If we test a hypothesis on individuals using state-level data we will get it wrong.
- Defining a concept means we take a vague term in a hypothesis like "democracy" and provide a more precise definition, like "a country with respect for free speech." Operationalizing the concept means providing some rules that will allow people to recognize your definition when looking at the real world, for example "a country with respect for free speech is one that has no laws forbidding demonstrations or speech against the government."
- A measure is *valid* if it actually captures the concept you are interested in. GNP growth is probably a valid measure of economic progress in a country, but it is probably not a valid measure of respect for human rights. Two kinds of validity were discussed in detail in class: *face validity* (does the measure appear to capture the concept you care about?), *content validity* (does the measure capture all aspects of the concept – for instance, does a measure of democracy capture competitive elections, respect for the rule of law, etc.?).

Measurement Part 2 (Week 5: April 27)

Major Points Covered:

- A measure is *reliable* if it doesn't have much measurement error. For example, an NGO might measure the number of people killed in a genocide – this is a valid measure (it captures the scale of the atrocity), but it won't be reliable (there will be a lot of variation, error, uncertainty, and conflicting reports). Another example comes from survey questions: do you approve/disapprove of the president (more reliable) versus rating the president on a 100-point scale (less reliable).
- Levels of measurement: a *nominal* measure just has unordered categories (example: a survey respondent's race) – all we know is the categories are different. An *ordinal* measure has ordered categories (example: a 5-point survey question on ideology running from most to least liberal) – all we know is that one category has more of something that another. An *interval* measure has actual numerical values that have meaning (example: percent of GDP spent on national defense).

- Methods of data collection can influence the validity and reliability of our data. For instance, watch for misleading or poorly designed survey questions, data collection methods that are unlikely to produce a representative set of data, and so on.

- Missing data can affect our hypothesis tests. There is *item nonresponse* (an observation is missing some data — example: a country doesn't keep statistics on some ethnic groups), and *unit nonresponse* (a potential observation doesn't wind up in our data – example: some people refuse to take an exit poll after the 2004 election).

Descriptive Statistics (Week 5: April 29)

Major Points Covered:

- Univariate frequency distributions show how frequently each value of a variable occurs. They are used with nominal, ordinal, and grouped interval level variables (grouped interval level just means we pick ranges for the interval level variable — for instance, how many people between $10,000 and $20,000, etc.).

- The *mode* is the most commonly occurring value of a variable. It is usually used with nominal level data.

- The *median* is the "middle" value of a variable (if we arranged the data from lowest to highest values, the media is in the middle). It is usually used with ordinal level data.

- The *mean* is the sum of the values of a variable divided by the number of observations. It is usually used with interval level data.

- Medians are more resistant to *outliers* (one or a few extreme values in the data) than means. Example: housing prices where there are a bunch of regular homes and a mansion – the median home price will be a lower number than the mean. In general, watch for outliers as a sign of a problem with the data or an unusual case to examine more closely.

- Measures of dispersion: quartiles are related to the median – again line our data up from lowest to highest values, and look at the value of the observations 1/4, 1/2, and 3/4 of the way along the line. Can also do this with other fractions (quintiles: every 1/5, deciles: every 1/10, percentiles: every 1/100, etc.).

- The variance is a measure of dispersion for interval level variables. The standard deviation is the square root of the variance.

- A variable can be *standardized* by subtracting off the mean and dividing by the standard deviation. This restates the value of the variable in terms of the number of standard deviations each observation is from the mean value of the variable. This is known as a *z-score.*

Descriptive Statistics (Week 6: May 4)

Major Points Covered:

- Bivariate frequency distributions show how frequently combinations of values between two variables occur. They are used with nominal, ordinal, and grouped interval level variables.
- *Correlations* are a useful way to examine the relationship between two interval level variables. The sign of a correlation tells us if the relationship between two variables is positive or negative. The strongest positive correlation possible is 1. The strongest negative correlation possible is -1. The correlation of a variable with itself is 1. If there is no relationship between the variables the correlation will be about 0. However, a correlation of 0 is *not* a guarantee that there is no relationship between the variables – correlations don't measure non-linear relationships. Correlations for groups of variables can be computed and presented in a correlation matrix.
- In order to examine the underlying relationship between two variables we often calculate a *regression line*. We do this through *ordinary least squares (OLS)* – this minimizes the squared distance between the line and each data point.
- The equation for the regression line is $y = a + bx + e$, where y is the dependent variable in our hypothesis, x is the independent variable in our hypothesis, a is the intercept, b is the slope, and e is random error.
- The intercept (or constant term) tells us what value of the dependent variable our regression line predicts when the independent variable is equal to zero.
- The slope tells us how much the dependent variable is expected to change when our independent variable increases by 1 unit (you'll need to know how the independent variable is coded to know what "1 unit" is).
- Know how to interpret a basic regression line and related it to a hypothesis.

Midterm (Week 6: May 6)

Multiple Regression (Week 7: May 11)

Major Points Covered:

- Multiple regression is similar to the bivariate regression discussed before the midterm, except there are multiple independent variables. The intercept tells us what value of the dependent variable is expected when all independent variables are equal to zero, and the slope coefficients tell us how the dependent variable is expected to change when that independent variable increases by 1 unit, holding all other independent variables constant.
- R^2 (R-squared) tells us the proportion of the variance in the dependent variable explained by the regression line. R-squared is always between 0 and 1, with higher numbers meaning more variance explained. Adjusted R-squared is a measure of the goodness of fit of the regression line controlling for having multiple independent variables.

Sampling and Statistical Inference (Week 7: May 13)

Major Points Covered:

- The *population* is the group of things that we care about (voters, countries, etc.).
- In most cases we can't observe the entire population, so we take a *sample* — a selected subgroup of the population. Even in some cases where we might think we have the entire population we might actually have a sample — for instance, we might observe all countries today, but our hypothesis is about a general causal rule that will apply to all countries (even those that don't exist yet, no longer exist, etc.) — in this case we actually have a sample.
- A *parameter* is some characteristic of the population we care about — it is what our hypothesis is about. It could be a number (the mean education in the U.S.) or a relationship (the relationship between education and voter turnout).
- A *statistic* is a number that we calculate in our sample to approximate the parameter we care about. We then use *statistical inference* to infer something about the population parameter using the statistic we calculated in our sample.
- Statistical inference depends on drawing a sample that is representative of the population and large enough to make reliable inferences from.
- Proper *sample selection* techniques are important to make the sample as representative of the population as possible. We usually want a *random sample*: a random sample is one in which every unit in the population had an equal chance of being selected to be in the sample. Some sampling techniques are better than others at this (example: random telephone surveys versus internet surveys).
- *Sample size* is also important. It is risky to make inferences from small samples. Larger samples are better, but the benefit of each additional observation shrinks as sample size increases.
- A *sampling distribution* is the distribution of estimates we would get if we drew a large number of samples from the population and calculated our statistics in each sample. Sampling distributions are almost always normal distributions (Central Limit Theorem).

Hypothesis Testing (Week 8: May 18)

Major Points Covered:

- Setting up hypothesis tests. A hypothesis is a statement about a population parameter. When we test our hypothesis using the data in our sample, we are unlikely to get exactly the number we hypothesized. Our task is to determine if our sample statistic differs from our hypothesized value because of sampling variability or because the hypothesis is wrong.

- We set up two opposing hypotheses in hypothesis testing. We have a *null hypothesis* — there is *no* difference between the population parameter and the hypothesized value. We also have an *alternative hypothesis* — there *is* a difference between the true population parameter and the hypothesized value.

- If our null hypothesis is correct, there will be a normally distributed sampling distribution around that number. If we know the variance of this normal sampling distribution, we can calculate how likely it is that we would have calculated the sample estimate we have if our hypothesis was true. If this probability is high, we will probably accept (fail to reject) our hypothesis. If this probability is low, we will probably reject our hypothesis. Put another way, we want to know if our estimate is "close" or "far" from our hypothesized value in terms of standard deviations of the sampling distribution — if it is "close" we accept the hypothesis, if it is "far" we reject the hypothesis.

- We set a *significance level* for our hypothesis tests. This is the amount of probability in the tail of the sampling distribution we are "cutting off." The standard level of significance in political science hypothesis tests is 5%, which tells us that we're cutting off the last 5% in the tails of the sampling distribution. Imagine a normal distribution centered on our hypothesized value — there will be two other values (known as *critical values*) that cut off the last 5% of the normal distribution. There is only a 5% chance that sampling variability would cause you to estimate a population parameter further from your hypothesized value than the critical values *if your hypothesis is true* — thus, we reject the hypothesis if our estimate falls outside the critical values.

Interpreting Hypothesis Tests (Week 8: May 21)

Major Points Covered:

- In most cases we do not know the variance of the sampling distribution, so we must estimate it. This means we switch from using a normal distribution to a *t distribution*, and thus the actual hypothesis tests we usually see in the social sciences are *t-tests*.

- The actual hypothesized value we are testing depends on the context. We saw two examples. In one case we could hypothesize the mean of a variable is some specific number – in this case the null hypothesis is that the true parameter is that specific number, and the alternative is that it is not. In the other case (and the most common case) we just have a hypothesis that a parameter should be positive or negative. This is commonly seen in regression, where we anticipate a positive or negative relationship between the independent variable and the dependent variable, but we don't have a specific number in mind. In that case, the null hypothesis for out test is *zero* — we can look at the slope coefficient to see if it is positive or negative, so we test to rule out zero. The alternative hypothesis of course is that the slope coefficient is not zero. If we can rule out zero, we call that regression slope *statistically significant*.

- We also pay attention to *substantive significance*, or the size of the effect of the independent variable on the dependent variable. A variable could be statistically significant, but have such a tiny substantive effect that it is unimportant in the real world.

- The results of hypothesis tests are often reported as *p-values*, which tell us how much probability is beyond the test statistic in the tail of the sampling distribution. p-values of less that 0.05 are usually regarded as indicating statistical significance.

Interpreting Hypothesis Tests, Specifying Regression Models (Week 9: May 28)

Major Points Covered:

- *Confidence intervals* are related to hypothesis tests. For instance, a 95% confidence interval is a range between two numbers, and we would be 95% confident that this range contains the true population parameter. We often see confidence intervals in public opinion polls — for instance, the level of support for candidate X is 50%, +/- 3%. The range 47% to 53% is a 95% confidence interval, meaning there is a 95% chance the candidates true level of support falls on that range.
- Interpretation of hypothesis tests in SPSS.
- Interpretation of hypothesis tests in journal articles.

Dummy Variables (Week 10: June 1)

Major Points Covered:

- In order to use nominal level variables as independent variables in regression we create *dummy variables*. All categories of the nominal variable are coded as zero except for one category, which is coded a one. These dummy variables can then be included in the regression as control variables or independent variables.
- The coefficient on the dummy variable tells us how to adjust the constant term for the category in the nominal variable coded as a one.
- Examples of interpreting dummy variables in SPSS regressions.
- Examples of interpreting dummy variables in regressions in journal articles.

Interaction Terms (Week 10: June 3)

Major Points Covered:

- Dummy variables adjust the intercept of the regression line for one group compared to other groups. We can also change the slope coefficient for an independent variable in the regression for a group by interacting the dummy variable for that group with the other independent variable. We do this by just multiplying these variables together to create a new variable, which we then include in the regression. This is known as an *interaction term*.

- The coefficient on an interaction term tells us how we should adjust the slope coefficient on the variable interacted with the dummy variable for the category in the nominal variable coded as a one. For instance, if we have an interaction term that is a dummy variable on gender (coded 0 for men, 1 for women) times education, the coefficient on that interaction term would tell us how much more positive or negative the effect of education is for women as compared to men.
- Examples of interaction terms in regressions.