David Youngberg

Acct/Busa/Econ 222—Bethany College

**Lecture 06: Hypothesis Testing**

1. Hypotheses, revisited
	1. *Null Hypothesis*—assertion corresponding to the default position, where there is no significant different, or where nothing is happening
		1. Example: Income ***doesn’t*** predict how much you spend
	2. *Alternative Hypothesis*—assertion that claims there is a significant difference.
		1. Example: Income ***does*** predict how much you spend
	3. Alternative hypothesis can be either one-tailed or two-tailed.
		1. Example: More income predicts that you spend less (one-tailed), that you spend more (one-tailed), or that you spend either more or less (two-tailed).
		2. We usually focus on two-tailed tests.
	4. *Level of significance*—determines a cut-off point when the null hypothesis is rejected or failed to be rejected. The standard is 95% (or 1.96).
	5. There are two types of mistakes you can make when working with a null hypothesis.
		1. *Type I error*—false positive; a non-match is declared a match—or when you reject the null hypothesis and you should fail to reject it
		2. *Type II error*—false negative; a good match is not detected—or when you fail to reject the null hypothesis and you should reject it

|  |
| --- |
| **Examples** |
| *Type I* | *Type II* |
| Convicting an innocent person | Letting the guilty go free |
| Approving a damaging drug | Rejecting a beneficial drug |
| Befriending a jerk | Ignoring a nice person |
| Funding a poor investment | Passing on a good investment |

* 1. Type I and Type II errors are equally undesirable, but Type II errors are insidious because they are harder to notice when they happen.
		1. *In general*, Type I errors are self correcting; Type II errors are not. But precisely because Type I errors are self correcting, the fact that one made an error at all is evident thus there is a tendency for people to favor Type II errors.
1. **iClicker**
2. Simple tests of hypothesis
	1. Suppose you take a sample and you’re interested if its difference from the mean (higher or lower) is because of random chance or because there’s something special about the sample you took.
		1. This is something you should remember form statistics, so we are going to do quick overview.
	2. You know the population standard deviation:

$$z=\frac{\overbar{x}-μ}{{σ}/{\sqrt{n}}}$$

* 1. You don’t know the population standard deviation:

$$t=\frac{\overbar{x}-μ}{{s}/{\sqrt{n}}}$$

* 1. You are using a proportion:

$$z=\frac{p-π}{\sqrt{\frac{π(1-π)}{n}}}$$

* + 1. Where π is the estimate of the population’s proportion and p is the sample estimate.
1. Two sample tests
	1. To compare if two samples are truly different, or their differences come from sample error, use this method:
	2. For sample means between two independent populations (known standard deviations):

$$z=\frac{\overbar{x}\_{1}-\overbar{x}\_{2}}{\sqrt{\frac{σ\_{1}^{2}}{n\_{1}}+\frac{σ\_{2}^{2}}{n\_{2}}}}$$

* 1. For two population proportions:

$$z=\frac{p\_{1}-p\_{2}}{\sqrt{\frac{p\_{c}(1-p\_{c})}{n\_{1}}+\frac{p\_{c}(1-p\_{c})}{n\_{2}}}}$$

* + 1. Where,

$$p\_{c}=\frac{x\_{1}+x\_{2}}{n\_{1}+n\_{2}}$$

1. **iClicker**
2. Paper
	1. As we approach your paper’s due date, let’s take some time to discuss implementation and expectations.
	2. Your paper should involve running a single multivariable regression, one dependent variable with three independent variables. I’ll accept other possibilities but you should come see me first so we could discuss your alternative.
		1. Exactly how you run regressions is something we’ll be discussing after break. But don’t worry: running a regression is really easy. It’s the interpretation that’s tricky. (That’s something else we’ll be discussing.)
3. **STEP ONE: NAIL DOWN YOUR VARIABLE RELATIONS**
	1. By now you should have one variable that you’re curious as to why it is what it is. What determines a country’s average income? What determines how healthy people are? What determines a company’s stock price? This (income, health, or price), is your *dependent variable*.
	2. You should also have some theories as to what influences the dependent variable you selected. Perhaps more patent activity makes people wealthier. Perhaps better access to AIDS medication makes people healthier. Perhaps better access to clean water makes people healthier. Perhaps better sales increase a stock price. These are the *independent variables*. (Even if the correlation is obvious, another interesting question is *how much* these variables matter. If sales increase by $10,000, how much does the stock price rise? The regression will tell us.)
	3. You should also make sure your variables are adjusted for population; that’s why I included population in the data; it is not a good independent variable.
		1. *Per capita* is a Latin phrase meaning “per person.”
		2. You want everything adjusted for population because it’s the only way to make sense of the analysis. Without adjusting, you’ll get strange results.
		3. For example, some of you might have noticed that the U.S. has a large number of students who don’t go to school. This is largely because the U.S. is so large; adjusted for population, it’s not that different from other countries.
		4. I will show you in lab how to adjust for population.
	4. You’re also going to want to take a moment and make sure you’re not committing any obvious errors with causation. For example, it’s unlikely that a rising stock price caused profits to rise. Similarly, watch out for confounding variables: you don’t want to end up arguing that ice cream causes drowning. In cases where causation *could* run either way or confounding variables could be or not be an issue, you’ll just have to make a strong case for the side you’re on. And that leads us to step two.
4. **STEP TWO: RUN THE NUMBERS**
	1. This part’s pretty quick so I suggest you do it a lot in many various ways. Run the regression, run the descriptive statistics, check for more subtle data problems (we’ll talk about those more later). Record the results, which will be included in your paper.
	2. Three things can happen for a variable:
		1. You reject the null hypothesis. This is often the desired result since it suggests something interesting is going on.
		2. You fail to reject the null hypothesis, even though every bit of research says you shouldn’t. While its possible you did something wrong, it is also possible that you found something interesting! If you discovered, for example, that rising incomes didn’t lead to rising consumption, that would be very surprising and worth reporting.
		3. You fail to reject the null hypothesis and that result is not ground-breaking. You will probably want to look at other independent variables and go again. Back to Step One.
5. **STEP THREE: READ!**
	1. You’ll have to make a strong argument that the inclusion of your independent variables make sense. This involves reading and summarizing the arguments of others. It involves research. This can take some time (the rest, honestly, goes pretty quickly) so make sure you set aside a good hunk of time for it.
	2. In the research process you might find out a better independent variable. If you can get the data, don’t be afraid to go back to Step 1 and include it. This is often how writing papers go; don’t assume these steps are linear.
6. **STEP FOUR: ADDRESS EXOTIC PROBLEMS**
	1. Regressions are built around a few assumptions, such as homoskedasticity and no multicollinearity. We will talk more about these later and how to check for them. In your paper, you’ll either need to point out that these are not present or that they are present but not a big deal.
	2. If they are present and a big deal, go back to Step One.
	3. You’ll also need to provide proof that you checked for these exotic problems.
7. **STEP FIVE: CRAFT YOUR ARGUMENT**
	1. Yes, you theoretically are supposed to figure out your argument first, but regressions can give you surprising results and the whole point of the regression is to figure out what’s going on.
	2. So in practice, combine your research on what other people said with your regression results and make an argument. This is your thesis statement: it is what the paper is arguing.
		1. *All papers need a thesis statement*.
	3. Your thesis statement doesn’t have to directly involve all your variables—most don’t actually—since you can argue your other variables are there for controls. Of course, you’ll have to justify them as controls.
	4. Make sure to calculate the impact of the variable you’re focusing on. This basically involves looking at its beta value. We’ll talk more about how to do this later, but you’ll want to include that “punch line” in your conclusion.
8. **STEP SIX: WRITE YOUR PAPER**
	1. Now you write your paper. Here’s my suggested organization:
		1. *Argument*. Why should I care about the determinants of this dependent variable? What are you concluding is important?
		2. *Literature review.* This is what others have said about this dependent variable. Make sure to cite your sources*.*
		3. *Description of data*. Because of the literature review, state the variables you are using, where you got them from (state original source, not gapminder.org), and what exactly they measure. Include descriptive statistics. Make sure to justify the inclusion of each variable and argue why causation is running as you describe. Do this even if you think it is obvious.
		4. *Regression*. Include a table of your regression output as well as a description of your regression, highlighting the dependent variable you crafted your thesis statement around. State your punch line to highlight how correct your thesis statement is.
		5. *Complications*. Note that you considered issues under Step 5 and everything’s fine. If everything’s not fine (because you couldn’t find a good substitute), then acknowledge that. Note these weakens your paper significantly and you should avoid this problem. Include the proof that you checked for these complications here.
		6. *Conclude*. Restate your thesis statement and summarize your argument, with evidence.
	2. Then proofread and correct your mistakes.
	3. Then proofread again and correct your mistakes again.

**Lab Section**

1. Chapter 10
2. Chapter 11
	1. Per capita
		1. To turn data from total to per capita, divide by population.
		2. Create a new column and type “=A1/B1” where A1 is the total and B1 is the population.
		3. This will give you “per capita.” So if A1 is the total number of houses in a country, “=A1/B1” will give you the number of houses per person.
	2. Per 100, 1,000, 10,000, etc.
		1. Note that homicide is done as “per 100,000.” This is because homicide is rare; recording it per person would result in a very small number with a lot of zeros after the decimal. It makes it hard to read.
		2. You can adjust for this by multiplying the result by any number. This will give you the number of A1’s per that number.
		3. If we multiply the number of homes per person by 10,000, that will give us the number of homes per 10,000. “=A1/B1\*10000”
	3. Copy & Paste
		1. Once you have your equation as you want it, copy it into the other cells. Make sure it updates cell references as appropriate.
3. Homework
	1. Chapter 10: 1,4,7
	2. Chapter 11: 1,6