David Youngberg
BSAD 210—Montgomery College

LECTURE 11: HYPOTHESES AND TYPES OF ERROR

I. Introduction
   a. Confidence intervals are all about trying to figure out what the population average is. But we can take it one step farther—let's estimate what the population average is *and then compare* that estimate to some pre-existing standard (such as what the population average was) and ask ourselves: is any difference we find the result of randomness or because something has changed? This is hypothesis testing.
   b. Hypothesis testing is the bread and butter of statistics. We know that samples can be unusually high or unusually low and, by The Central Limit Theorem, we can get an idea of how likely it is for them to be high or low by random chance. (That's a big deal.)
   c. But to understand hypothesis testing, we first have to cover some foundational concepts—the number of tails of a test, the null and alternative hypotheses, and types of error.

II. One-Tail or Two?
   a. In statistics, a "tail" refers to the low parts of a distribution, one that tappers off like a tail of an animal. A normal distribution has two tails.
   b. Some hypothesis tests only care about one of the tails, others care about both of them.
      i. We think of the tail value as α; it can be split between the tails (for two-tailed tests) or concentrated in one tail (for one-tailed tests).
   c. One-tail tests are for tests of improvement, where doing worse is effectively the same as doing average; neither is impressive. One-tail tests also used to refute points of view (someone might say something is popular so the alternative is that it's not popular).
      i. Examples: Longer battery life; faster acceleration time.
   d. Two-tail tests are for tests of unusualness. In other words, there's a "sweet spot" that the null hits but the alternative doesn't. The question is *not* if the value is more or less than the average but if the value is *different*.
      i. Examples: the accuracy of a machine putting ketchup in a bottle (could be putting too much or too little in); the length of time you must stand in line at a store (both too little or too much is

noteworthy); if a new employee is unusually good or bad at the job.

III. Hypotheses

a. *Null Hypothesis*—assertion corresponding to the default position, where there is no significant different, or where nothing is happening
   i. The null hypothesis captures the state where nothing makes a difference (even if the intuition is that it should).
   ii. Example: Income ***doesn't*** predict how much you spend
   iii. Example: An hour of exercise each day for a year ***won't*** cause people to lose weight.

b. *Alternative Hypothesis*—assertion that claims there is a significant difference.
   i. Example: Income ***does*** predict how much you spend
   ii. Example: An hour of exercise each day for a year ***will*** cause people to lose weight.

c. Alternative hypothesis can be either one-tailed or two-tailed.
   i. 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).
   ii. Example: An hour of exercise each day for a year will cause people to lose weight (one-tailed), or their weight will change (two-tailed).

d. This brings us back to the different tails of a test. We use different notation not just for the kind of hypotheses but also how many tails the test has.

	One tail	*Two tails*
Null (H_0)	$\geq$ or $\leq$	$=$
Alternative (H_a)	$<$ or $>$	$\neq$

   i. Example: determining if a machine is putting 50 lbs of rice in a bag as it is supposed to, you would write H_0: $\mu = 50$ lbs and H_a: $\mu \neq 50$ lbs
   ii. Example: determining if a factory redesign increased output from the normal 9,000 units, you would write H_0: $\mu \leq 9{,}000$ units and H_a: $\mu > 9{,}000$ units

IV. Types of error

a. There are two types of mistakes you can make when working with a null hypothesis.

i. *Type I error*—when you reject the null hypothesis and you should've failed to reject it
ii. *Type II error*—when you fail to reject the null hypothesis and you should've rejected 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

b. Type I and Type II errors are equally undesirable, but Type II errors are insidious because they are harder to notice when they happen.
   i. *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 commit Type II errors.

V. The Critical Values

a. Let's revisit critical values from earlier. Notice the expansion.

Confidence	α	$z_{\alpha/2}$	z_{α}
95%	0.05	1.960	1.645
99%	0.01	2.576	2.326
99.9%	0.001	3.291	3.090

b. The additional column is for when the significance level is concentrated on just one side of the distribution—it's for one-tailed tests. Note that instead of $\alpha/2$ (two-tailed test) in the subscript for z, there's just α (one-tailed test).

VI. Simple tests of hypothesis

a. Terminology
   i. The calculated value is the result of your calculation.
   ii. The critical value is based on the confidence level (as above).
b. You'll calculate a z-value and compare a critical z-value. All values will be positive (take the absolute value first if they are not).
   i. If your calculated value is greater than the critical value, you reject the null hypothesis. The difference is probably not due to chance. We call this result "statistically significant."

1. It can <u>still</u> be due to chance. We could be committing type I error: 95% isn't 100%. The phrase "statistically significant" comes from this idea that passing the threshold means the results are interesting, even if nothing's been "proven."
2. <u>Fun fact</u>: α is the chance of type I error.

ii. If your calculated value is less than the critical value, you fail to reject the null hypothesis. It is not statistically significant.

1. Why "fail to reject" rather than "accept?" Because that sounds like we're saying the null hypothesis is true; it may not be. Null hypotheses are *never* "accepted."
2. This sounds like a trivial distinction but it's not. "Accepting" the null hypothesis is declaring that the null is correct. But, as confidence intervals remind us, it may not be correct. We have only failed to find a difference; failing to find a difference is not the same thing as showing there is no difference!

iii. Note that as α decreases, the critical value increases. Thus, if you reject the null hypothesis at a particular α, you should reject the null at a higher α (because the critical value will be lower).

c. What should matter when determining if you should reject or fail to reject the null hypothesis?

i. How different the sample average is from the population average. If the difference is really large, then the sample mean is far from the center of the distribution and thus the likelihood this sample is different by chance falls.

ii. If values for the variable tend to fall in a narrow range, it should be less likely (all things other being equal) the less likely the null hypothesis is correct. The lower the standard deviation, the lower the likelihood that the sample mean is different by chance.

iii. If you have a lot of observations, you know more about the population you sampled from. The bigger the sample, the less likely the sample mean is different by chance.