David Youngberg
BSAD 210—Montgomery College

LECTURE 18: HYPOTHESIS TESTING II

I. Unknown σ

a. You don't know the population standard deviation:

$$t_{\bar{x}} = \left|\frac{\bar{x} - \mu_{H_0}}{s/\sqrt{n}}\right|$$

i. Where $t_{x\text{-bar}}$ is the z-test statistics; and
ii. s is the standard deviation of the sample.

b. Example. Let's do the same problem again but assume that we don't know the population standard deviation. If that's the case, we have to rely on the sample's standard deviation; it's our best estimate of σ.

i. Assume s is exactly two; in other words, assume our sample standard deviation happens to be the same as what σ was before.

$$t_{\bar{x}} = \left|\frac{\bar{x} - \mu_{H_0}}{s/\sqrt{n}}\right| = \left|\frac{13.75 - 15}{2/\sqrt{16}}\right| = \left|\frac{-1.25}{2/4}\right| = \left|\frac{-1.25}{0.5}\right| = 2.50$$

ii. Note that, equation-wise, nothing's changed because none of the numbers have changed. Our calculated value is identical.
iii. What's changed is what "2" represents. Because it represents s instead of σ, we have to use the t-distribution to find our critical values. But what are our critical values?

c. We can use the T.INV and T.INV.2T functions to determine critical values for one and two tailed tests, respectively.

i. Regardless, Excel will ask for two pieces of information: probability and degrees of freedom:
ii. *Probability* is the alpha value.
iii. *Deg_freedom* is the degrees of freedom: n – 1.

d. Let's do a quick example before turning back to the problem. Suppose you want to know the critical t-value for an alpha of 0.05 with six degrees of freedom. Then you'd type "=T.INV.2T(0.05,6)" and press ENTER.

i. You should get about 2.447, the same as in the table from the last notes under 95% confidence with 6 degrees of freedom.

ii. If you think the table says that the result is 1.943 (or -1.943), it's because you're looking at the one-tailed version and not the two-tailed version. (In the two-tailed version, that 0.05 is equally split between two sides of the distribution, thus a single tail would have a value of 0.025.)

e. Now to the actual problem. This one is a one-tailed test so we'll use T.INV. We want to know the critical values at 95%, 99%, and 99.9%, or alphas of 0.05, 0.01, and 0.001, respectively. With fifteen degrees of freedom (16-1) this is what you should get:

0.05	1.753
0.01	2.602
0.001	3.733

i. Note in all cases, I took the absolute value of the result. This is because critical values are always about distance.

f. With a calculated value of 2.5, the result is still significant at the 95% level, but no longer at the 99% level. 2.602 > 2.5.

II. Proportion

a. Proportions are always fractions, with values ranging from zero to one. They represent a portion of a group that contains some value of interest. Like the portion of people who got a cold or support a candidate or returned what they bought.

i. Percent values are proportions ***but only*** if they are a percent of the observations having or lacking a quality. For example, suppose 50 students took a class. Their average grade is not a proportion (even if that grade is expressed as a percent) but the ***percent of students*** who passed is a proportion.

b. If you are using a proportion, the equation for the calculated value looks like this:

$$z_p = \left| \frac{\bar{p} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \right|$$

i. Where π is the estimate of the population's proportion; and

ii. p-bar is the sample estimate.

c. Example. Suppose you work at a growing trucking business and you want to know how its recent expansion has affected the business's operations. Imagine before the expansion, 20% of deliveries were late to their destination. After the expansion, 28% (out of 100 deliveries) were late. Is this a statistically significant difference?

i. Your null hypothesis is that the proportion of late deliveries hasn't changed: p-bar = π
ii. Your alternative hypothesis is that it has changed: p-bar ≠ π.
iii. Here's what your equation should look like:

$$z_p = \left| \frac{\bar{p} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \right| = \left| \frac{0.28 - 0.2}{\sqrt{\frac{0.2(1-0.2)}{100}}} \right| = \left| \frac{0.08}{\sqrt{\frac{0.2(0.8)}{100}}} \right| = \left| \frac{0.08}{\sqrt{\frac{0.16}{100}}} \right| = \left| \frac{0.08}{\frac{0.4}{10}} \right| = 2.0$$

iv. Note that because this is a two-tailed test, our critical z-scores are 1.960 (95%); 2.576 (99%); and 3.291 (99.9%). Because the absolute value of 2.0 is greater than 1.96, we have evidence that expanding the business impacted the rate of late deliveries.
v. It is statistically significant at the 95% level *and no more*. There's a 5% chance (technically, slightly less than a 5% chance) that our difference is merely a coincidence.