David Youngberg
BSAD 210—Montgomery College

LECTURE 12: HYPOTHESIS TESTING I

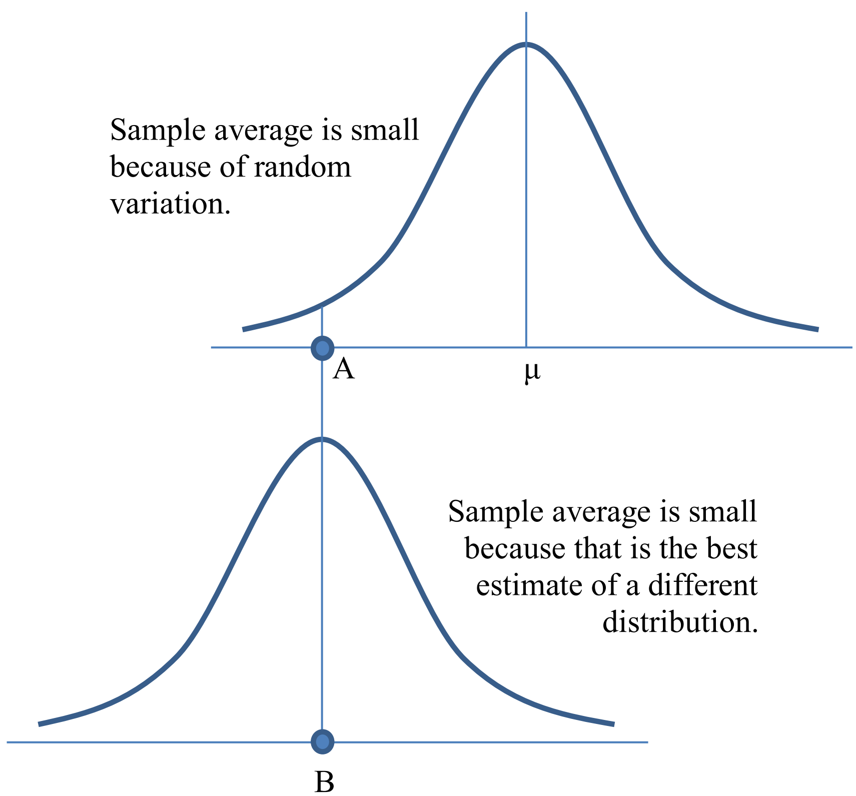
I. The Nature of Hypothesis Testing

a. Let's remember what we're doing here.

   i. If the sample mean is close what we normally see, it stands to reason that we got a different mean because of random chance.

   ii. But if the sample mean is far from what we normally see, then there's probably something unusual about the sample. In other words, the sample means is *from a different distribution* than the population mean.

b. Graphically, hypothesis testing is about answering this question: Is the sample mean from A (unusual due to randomness) or B (unusual due to a genuine difference in populations).



c. If it's A or if it's B depends on how far away the sample mean is from the population mean, the standard deviation and the sample size.

i. The bigger the difference between the sample mean and the population mean, more likely your sample mean is different because of some genuine difference.
ii. The smaller the standard deviation, the more likely your sample mean is different because of some genuine difference.
iii. The larger the sample size, the more likely your sample mean is different because of some genuine difference.

d. All of these considerations are combined into an equation to find the calculated value, but it also turns there are three different equations. We'll start with the most straightforward.

II. Known σ

a. If you know the population standard deviation:

$$z = \left|\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right|$$

i. Where z is the z-test statistics;
ii. x-bar is the sample mean;
iii. μ is the mean of the population distribution, which is assumed to be true for the null hypothesis;
iv. σ is the population standard deviation; and
v. n is the sample size.
vi. Note that we take the absolute value; this only for purposes of comparing to critical values.

b. <u>Example</u>. Nurses typically have a 12-hour shift, which is a grueling pace. Suppose, on average, a nurse makes 8.3 mistakes per eight-hour period, with a population standard deviation of 2.4 mistakes. Also suppose that 16 nurses were given 8-hour shifts; based on that sample of nurses, the average number of mistakes per eight-hour period fell to 6.8 mistakes. Is this a statistically significant improvement?

i. As this is a one-tailed test, our null hypothesis is that it does not reduce the number of mistakes. H_0: $\mu \geq 8.3$ mistakes
ii. As this is a one-tailed test, our alternative hypothesis is that it reduces the number of mistakes: H_a: $\mu < 8.3$ mistakes.
iii. Here's what your equation should look like:

$$z = \left|\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right| = \left|\frac{6.8 - 8.3}{2.4/\sqrt{16}}\right| = \left|\frac{-1.5}{2.4/4}\right| = \left|\frac{-1.5}{0.6}\right| = 2.50$$

iv. Note that because this is a one-tailed test, our critical z-scores are 1.645 (95%); 2.326 (99%); and 3.090 (99.9%). Because calculated value is greater than 2.326, we have evidence that our sports drink improves performance. It is statistically significant at the 99% level. There's a 1% chance (technically, less than a 1% chance) that our difference is merely a coincidence.

v. However, the calculated value is not greater than 3.090, our threshold for 99.9% confidence. There is greater than a 0.1% chance that our difference is merely a coincidence.

c. You might be tempted to say that we've "proved" reducing the length of the shift reduces mistakes but we haven't "proved" anything. Because there's always a chance of luck, statisticians state that we have evidence but claims of "proof" would not be accurate.

III. P-values

a. Another way to do this same thing is to calculate what's called a "p-value." This is an optional way to find statistical significance but it's helpful to get an understanding of it now because the concept will come up a lot next unit.

b. A *p-value* is the small α you can claim and still have statistical significance. In other words, if the p-value is **smaller** than our standard thresholds of 0.05, 0.01, and 0.001, then there's statistical significance at the 95%, 99%, and 99.9%, respectively.

c. To find a **p-value** for a calculated z-value, follow these steps:

i. Make your calculated value negative.

ii. In Excel, type in NORM.S.DIST(X,1), where X is the negative calculated value. (Note the cumulative function is **on**.)

iii. If this is a one-tailed test, then the result is the p-value. If this is a two-tailed test, **multiply** the result by two to get the p-value.

d. <u>Example</u>: In the above example, we would enter into Excel: =NORM.S.DIST(-2.5,1) and get about 0.0062.

i. This <u>is less</u> than 0.01, so it <u>is</u> statistically significant at the 99% level, but it <u>is not less</u> than 0.001, so it <u>is not</u> statistically significant at the 99.9% level.

ii. If the nursing example was a two-tailed test, we'd double 0.0062 to be 0.0124, which is statistically significant at 95% but nowhere else.

IV. Connection to Confidence Intervals

a. To better understand what we're doing, note there's a connection between confidence intervals and hypothesis testing.

b. Consider a confidence interval at 95% confidence. That means there's a 95% chance that the true population mean is in that interval. That also means that there's a 5% chance the population means is above or below that interval.
c. Suppose the true population mean is outside the interval. That would be surprising (as it's unlikely; 5% is a small chance) and would suggest that maybe that true population mean isn't really the true population mean, or that the sample mean and the population mean are actually from two different distributions.
    i. Suppose we're estimating the costs of wind vs solar energy. Solar seems more expensive—the difference averages 5 cents per kilowatt-hour (meaning solar is somewhat cheaper)—but there's a confidence interval which ranges from -1 cent (wind is slightly cheaper) to 11 cents (solar is *much* cheaper).
    ii. Note this range includes our null hypothesis: zero, or that there's no cost difference between them. In other words, there's less than a 95% chance that the null hypothesis is included in our range and the 5-cent difference is not statistically significant.
d. Here's some algebra to show you the connection:

$$z = \left|\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right| = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \quad or \quad -\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$
$$z\,\sigma/\sqrt{n} = \bar{x}-\mu$$
$$\mu = \bar{x} - z\,\sigma/\sqrt{n}$$

$$z = -\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$
$$z\,\sigma/\sqrt{n} = \mu - \bar{x}$$
$$\mu = \bar{x} + z\,\sigma/\sqrt{n}$$

$$\mu = \bar{x} \mp z\,\sigma/\sqrt{n}$$

e. Look, it's a confidence interval!

V. Unknown σ

a. If you don't know the population standard deviation:

$$t = \left|\frac{\bar{x}-\mu}{s/\sqrt{n}}\right|$$

    i. Where t is the t-test statistic; 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 2.4; in other words, assume our sample standard deviation happens to be the same as what σ was before.

$$t = \left|\frac{\bar{x} - \mu}{s/\sqrt{n}}\right| = \left|\frac{6.8 - 8.3}{2.4/\sqrt{16}}\right| = \left|\frac{-1.5}{2.4/4}\right| = \left|\frac{-1.5}{0.6}\right| = 2.50$$

   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 "2.5" represents. Because it represents t instead of z, 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-tailed and two-tailed tests, respectively. Regardless, Excel will ask for two pieces of information:
   i. *Probability* is the alpha value.
   ii. *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 (n=7). 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—just a hair under the threshold!

g. Alternatively, you can calculate the **p-value**.

i. Make your calculated value negative (like before):

ii. In Excel, type in T.DIST(X,1), where X is the negative calculated value. (Note the cumulative function is **on**.)

iii. If this is a one-tailed test, then the result is the p-value. If this is a two-tailed test, **multiply** the result by two to get the p-value.

1. In our example, we type in: =T.DIST(-2.5,15) and get 0.0123. Note this is just a hair over what's needed for statistical significance at the 99% level—the result we had when comparing critical and calculated values.

VI. 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:

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

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.

VII. Practical significance
a. Suppose our company ran a study on using wind power and we found that after considering construction costs, inconsistent power, and other considerations, we'd save $42 a year and that savings is statistically significant to the 99.9% level. Should we switch to wind?
i. No! $42 is hardly worth the effort. Just because it's a statistically significant difference, doesn't mean it's a meaningful difference.
b. *Practically significant* is a statistically significant result that is also meaningful or noteworthy.
i. The numerator in all these equations is the difference between the population average and the sample average and it's the most important bit of information—*if* this is a statistically significant difference, that doesn't mean the difference is noteworthy in a practical sense. Small differences can be statistically significant differences (especially when n is very large).
ii. Note that all any difference that isn't statistically significant can't then be practically significant. Practical significance is a subset of statistical significance.