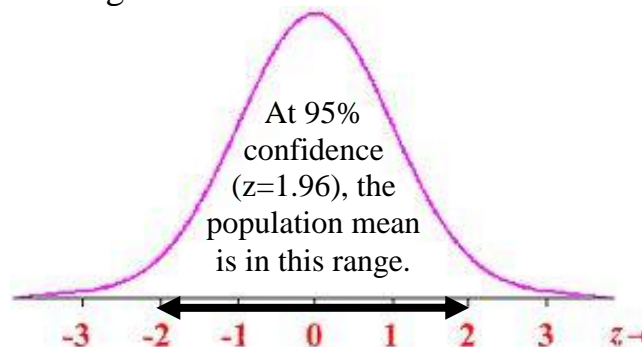


## LECTURE 12: HYPOTHESIS TESTING I

### I. The Nature of Hypothesis Testing

- a. Let's begin on familiar ground: The Central Limit Theorem. We know that if you take a bunch of samples, the means of those samples form a normal distribution. In practice, this implies that for any sample we pull, we don't know where on the normal distribution we fall. It could be in the middle, it could be unusually small, unusually large, etc.
  - i. But we do know where the *most* likely location is: the middle. In other words, our sample mean—assuming the sample isn't biased—is the best estimate of the population.
  - ii. Note the connection between this and the confidence interval: the margin of error revolves around the sample mean.



- iii. A large margin of error and/or a high confidence level means the margin of error is probably in the interval you calculated. In other words, your sample is about spot on.
- b. But let's go a different direction. The world is in a constant state of change and there are many good reasons that a sample taken from unusual circumstances, such as a subgroup of a population at a different point in time, will not capture the population mean.
  - i. This is a problem if you're trying to figure out what's going on under normal circumstances but what if you're trying to figure out if something is unusual? That's what hypothesis testing is for.
  - ii. Recall what the null hypothesis is: the usual result. The alternative hypothesis is the unusual result.
- c. If the sample mean is close what we normally see, it stands to reason that we got a different mean because of random chance. But if the

sample mean is far from what we normally see, then there's probably something unusual about the sample.

- d. The question becomes, "What's close? What's far away?" That depends on standard deviation and sample size.
  - i. The larger the standard deviation, the more likely your sample mean is different by chance.
  - ii. The larger the sample size, the more likely your sample mean is different because of some interesting reason.
- e. When the calculations are complete, we get a result which, after taking the absolute value, compare to a critical z-score.
  - i. If our result is less than the critical value, we *fail to reject the null hypothesis*. It's not unusual enough to be noteworthy.
  - ii. If our result is greater than the critical value, we *reject the null hypothesis*. This result is probably not due to chance.
- f. It can still be due to chance—95% confidence isn't 100% confidence—but probably not. The phrase "statistically significant" comes from this idea that passing the threshold means the results are interesting.
  - i. It's a controversial mindset. The process states if your result is just below the critical score, nothing interesting is going on. But just above the score is suddenly interesting. The lack of a gray area is frustrating.
  - ii. That said there is still some gray area: samples significant at the 95% level may not be significant at the 99% or 99.9% level.

## II. Known $\sigma$

- a. If you know the population standard deviation:

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

- i. Where  $z_{\bar{x}}$  is the z-test statistics;
  - ii.  $\bar{x}$  is the sample mean;
  - iii.  $\mu_{H_0}$  is the mean of the sample distribution, which is assumed to be true for the null hypothesis;
  - iv.  $\sigma$  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. Example. You want to know if your new sports drink improves athletic performance over what people normally do. Suppose the average athlete can run a mile in exactly fifteen minutes with a

standard deviation of exactly two minutes. You give 16 athletes your sports drink and time their run. Your sample mean is exactly 13.75 minutes.

- i. Your null hypothesis is that it does nothing to reduce the time it takes to run a mile:  $\bar{x} \geq \mu$
- ii. Your alternative hypothesis is that it reduces the time it takes to run a mile:  $\bar{x} < \mu$ .
- iii. Here's what your equation should look like:

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

- iv. Note that because this is a one-tailed test, our critical z-scores are 1.645 (95%); 2.330 (99%); and 3.090 (99.9%). Because the absolute value of -2.5 is greater than 2.330, we have evidence—strong 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" our sports drink make a positive difference. We haven't "proved" anything. Because there's always a chance of luck, statisticians state that we have evidence for something. "Proof" is not an option.
  - d. The example also highlights the difference between statistical significance and practical significance.
    - i. Yes, our sports drink really does improve performance (statistical significance).
    - ii. But is the one minute difference really that big of a deal? If the answer is no, then it's not really interesting and people might not be willing to spend the money on it (practical significance).

### III. Unknown $\sigma$

- 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  $t_{\bar{x}}$  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  $\sigma$ .
- i. Assume  $s$  is exactly two.

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

- 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  $\sigma$ , 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. So if you want to know the critical t-value for an alpha of 0.05 with six degrees of freedom, 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 under 95% confidence with 6 degrees of freedom.
  - ii. If you think the table claims the result should be 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 0.25.)
- e. But this is a one-tailed test so we'll just 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

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

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

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

- i. Where  $\pi$  is the estimate of the population's proportion; and
- ii.  $p$  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:  $x\text{-bar} = \mu$
- ii. Your alternative hypothesis is that it has changed:  $x\text{-bar} \neq \mu$ .
- iii. Here's what your equation should look like:

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

- 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.