Lesson 51: Significance Tests: The Basics

Significance Testing (Hypothesis Testing)

While confidence intervals are a common type of statistical inference – used when the goal is to estimate a population parameter using sample data – another form of statistical inference is that of significance testing (A.K.A. Hypothesis Test).

Significance tests are used when one wants to assess whether the data provides enough evidence of some claim about the population.

The Key Concept:

The claim made is the **Null Hypothesis**. We know need evidence to prove that the claim is not true, so a sample is taken. If we took many samples and the claim was true, then what is the probability of getting the resulting statistic from the sample? This probability is called the **p-value**. If the probability is low enough, we say the claim is rejected.

Example:

Mr. E claims that teachers at NWHS tend to inflate their grades in order to keep students from complaining. The class disagrees and thinks that grades are not inflated and state that the average GPA in the school is only 3.0. In order to test Mr. E's claim of grade inflation, an SRS of 40 students was taken and the average GPA from that sample was found to be $\bar{x} = 3.3$ (GPA's at NWHS have a standard deviation $\sigma = 0.9$). Is this enough evidence to suggest that the average NWHS grade is in fact above 3.0 (thus constituting grade inflation)?

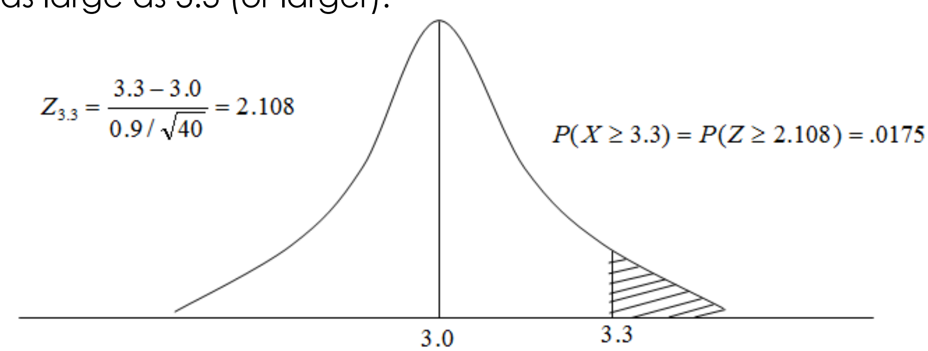
Solution:

Keep in mind, that our sample is only one of many possible samples. So if we were to take another sample of the same size our $\bar{x}$ might be different. The sampling distribution of all sample means will be approximately normal (due to CLT) with a mean of $\mu_{\bar{x}} = \mu$ and a standard deviation of

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.9}{\sqrt{40}}.$$

If, in fact, the mean GPA at NWHS is 3.0 (as the student's claim) and the standard deviation is 0.9, how likely are we to get an $\bar{x}$ that is 3.3? Is the likelihood so small that we can't attribute it to chance variation? Or is 3.3 a "reasonable" sample mean based on our data?

We can answer this question by sketching the distribution of sample means below and finding the shaded area. This shaded area represents the probability that we would get a mean as large as 3.3 (or larger):



The p-value is .0175. If the true mean is 3.0, then the probability of getting a sample mean of 3.3 from a sample of 40 is .0175.

This means that there is less than a one in 50 chance that our $\bar{x}$ was 3.3 by chance. Because this probability is so small it suggests that the true population mean GPA is in fact not 3.0, but some higher number. It appears that Mr. E is right after all.

How do we know when we have enough evidence to reject the claim?

The **p-value** measures the probability of getting the test statistic if the null hypothesis is true.

Most of the time, there is a **significance level** (called the **α-level**) set before sample is collected.

If the p-value is **below** the significance level:

- Then we reject the null hypothesis H_0 in favor of the alternate hypothesis H_a
- We say the results are statistically significant

If the p-value is **above** the significance level:

- Then we fail to reject the null hypothesis H_0 and reject the alternate hypothesis H_a
- We say the results are NOT statistically significant

Most α levels are set at 0.05, which means that there would less than a 5% probability that the sample result would have occurred by chance.

From the original example:

- Using a α levels of 0.05, we would reject the null hypothesis since our p-value = 0.0175 is less than the α level.
- Using a α levels of 0.01, we would fail to reject the null hypothesis since our p-value = 0.0175 is greater than the α level.

Daily Data Collection

Think of your ideal city to begin your life after college.

My Claim: The average distance away from our school where students plan to begin their lives is 100 miles away.

Calculate the distance from our school to your city and report the number of miles.

Hypothesis

Data

Sample size	
Sample mean	
Sample st. dev.	
St. Dev. of the mean	
T =	
p-value (two sided test with alpha = 0.05)	

Graph

Conclusion

Example

A golfer wants to improve his play. A friend suggests getting new clubs and lets him try out his 7-iron. Based on years of experience, he has established that the mean distance that balls travel with his old 7-iron mean = 175 yards with standard deviation 15 yards. He is hoping that this new club will make his shots with a 7-iron more consistent (less variable), so he goes to the driving range and hits 50 shots.

a. Describe the parameter of interest

b. State the appropriate hypotheses for performing a significance test

c. If the p-value for the hypotheses in part b is .035. What does this mean?

d. If the significance level is $\alpha = .01$. What is the conclusion?

Example

Zeb performs a study to see if students prefer name brand chips or generic chips. He randomly selects 50 students. Overall, 34 of the 50 students preferred the name brand chips. Zeb performed a significance test using the hypotheses H_0: $p = 0.5$ H_a: $p > 0.5$ where p = the true proportion of students at his school who prefer name-brand chips. The p-value was 0.0055. What conclusion would you make at each of the following significance levels?

$\alpha = .05$

$\alpha = .01$

HW 51 from Section 9-1: 1,3,5,7,9,11,13