Chapter Nineteen: More about Tests and Intervals

Recall: The definition of a p-value is the probability of getting the observed test statistic (whether it is z or t) or one that is more extreme if the null hypothesis (H_o) is true.

We note, that a p-values is not the probability that the null hypothesis is true.

Relationship between H_o and H_a

To demonstrate the relationship between the null and the alternative hypothesis, we might compare a hypothesis test to a criminal trial. In the U.S judicial system, we assume that the defendant is innocent until proven otherwise. Therefore, we might say the null hypothesis in this case is:

H_o: the defendant is innocent

Therefore the prosecutor must convince the jury of the defendant's guilt by presenting evidence of a crime. We might say that this is the alternative hypothesis, it is the slaim that we are trying to find evidence to support.

H_a: the defendant is guilty

- If **enough evidence** is presented to suggest that the defendant is guilty ...
  - conclude the defendant is guilty
  - There is evidence against the null hypothesis H_o
- If **not enough evidence** is presented to suggest that the defendant is guilty ...
  - conclude there is not evidence that the defendant is guilty
  - There is little to no evidence against the null hypothesis (H_o)
  - we do not conclude the defendant is innocent (that H_o is true)
  - We were trying to show evidence of guilt, not innocence.

But a question might arise, how much evidence is enough evidence.

Practical vs. Statistical Significance

A test is statistically significant if ...

↳ There is a statistical difference between the null hypothesis value of the parameter and its true value (p-value < some cut-off)

↳ An indication of the strength of evidence.

A test is practically significant if ...

↳ The difference between the null hypothesis value of the parameter and its true value is meaningful in context.

↳ Not necessarily based on statistical analysis

statistical significance ≠ practical significance

For large samples, even small deviations from the null hypothesis could be statistically significant. But even if these differences are statistically significant, they may not be practically significant.

For small samples, small but impactful differences may not end up being statistically significant. Hypothesis tests can only detect a very large difference between H_o and the true value of the parameter. But in these cases, lack of sitatistically significant evidence does not mean that a signficant relationship does not exist.

Errors in Hypothesis Testing

Recall from Chapter 17 handouts part two that there are four different options for hypothesis tests:

- Little to no evidence
- weak evidence
- Moderate evidence

- Strong evidence

But just like how courts can sometimes wrongfully convict an innocent person or let a guilty person walk free, we can make errors in our hypothesis testing. For this class we will focus on two different types of hypothesis errors.

Type I and Type II Errors

	Evidence Against H_o	**No Evidence Against H_o**
H_o **is Actually True**	Type I Error	Correct
H_a **is Actually True**	Correct	Type II Error

The probabilities of Type I and Type II Errors are inversely related:

- If you increase the probability of a Type I error, you decrease the probability of a type II error.
- If you decrease the probability of a type I error, you increase the probability of a type II error.

So ultimately the goal of hypothesis testing is to minimize both the probability of a Type I Error and the probability of a Type II Error. But because these two concepts are inversely related, these goals conflict. So often we choose to minimize our Type I error at the expense of our Type II error

Hypothesis Tests & Confidence Intervals for μ

Consider a two-sided hypothesis test:

$$H_o : \mu = \mu_o \quad H_a : \mu \neq \mu_o$$

This test has a direct relationship with a confidence interval for μ:

- little to no evidence against the null, then the confidence interval will contain μ_o.
- some strength of evidence against the null hypothesis test, then the CI will not contain μ_o

(doesnt necessarily apply for inference on p)

Example *A study was done to determine the average commute time to work in Atlanta, Georgia. A random sample of 500 residents of metropolitan Atlanta was taken. The sample mean was* $\bar{y} = 29.11$ *minutes with a standard deviation of* $s = 20.7$ *minutes. A 90% confidence interval is computed to be (27.58 minutes, 30.64 minutes). Consider the following hypotheses:*

$$H_o : \mu = 31 \ minutes \quad H_a : \mu \neq 31 \ minutes$$

Based on the above confidence interval, what can we say about the strength of he evidence against the null hypothesis.

μ_0 is not in (27.58, 30.64) strong evidence against the null.

Important Take-Aways

- inference does not fix bad data.
- Good data comes from random samples & randomized experiments.
- Assumptions must hold in order to perform a hypothesis test & construct a confidence interval.