Section 8.1 Hypothesis Testing

In Statistics, a hypothesis proposes a model for the world. Then we look at the data. If the data are consistent with that model, we have no reason to disbelieve the hypothesis. Data consistent with the model lend support to the hypothesis, but do not prove it. But if the facts are inconsistent with the model, we need to make a choice as to whether they are inconsistent enough to disbelieve the model. If they are inconsistent enough, we can reject the model.

Think about the logic of jury trials:

To prove someone is guilty, we start by assuming they are innocent. We retain that hypothesis until the facts make it unlikely beyond a reasonable doubt. Then, and only then, we reject the hypothesis of innocence and declare the person guilty.

The null hypothesis, which we denote H_0, specifies a population model parameter of interest and proposes a value for that parameter.

We might have, for example,

$$H_0 : \mu = 7.23.$$

We want to compare our data to what we would expect given that H_0 is true. We can do this by finding out how many standard deviations away from the proposed value we are. We then ask how likely it is to get results like we did if the null hypothesis were true. To perform a hypothesis test, we must first translate our question of interest into a statement about model parameters. In general, we have

$$H_0 : \text{population parameter} = \text{hypothesized value}.$$

The alternative hypothesis, H_A, contains the values of the parameter we accept if we reject the null. There are three possible alternative hypotheses:

$$H_A : \text{parameter} < \text{hypothesized value}$$

$$H_A : \text{parameter} \neq \text{hypothesized value}$$

$$H_A : \text{parameter} > \text{hypothesized value}$$

The tests are either two-tailed or one-tailed depending upon the alternative hypothesis.

The statistical twist that makes this different from jury trials is that we can quantify our level of doubt. We can use the model proposed by our hypothesis to calculate the probability that the event we've witnessed could happen. That's just the probability we're looking for; it quantifies exactly how surprised we are to see our results. This probability is called a **P-value**. When the data are consistent with the model from the null hypothesis, the P-value is high and we are unable to reject the null hypothesis. In that case, we have to "retain" the null hypothesis we started with. We can't claim to have proved it; instead we "**fail to reject the null hypothesis**" when the data are consistent with the null hypothesis model and in line with what we would expect from natural sampling variability. If the P-value is low enough, we'll "**reject the null hypothesis**," since what we observed would be very unlikely were the null model true.

We can also compute the **test statistic (**which looks very much like a z-score**)**. When testing averages, the test statistic is

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

The test statistics measures how far from the sample data is from the null hypothesis claim. If the test statistic is too far from the claim (relative to the critical value) then we "**reject the null hypothesis**." The critical values depend on the level of confidence and the number of tails in the test.

level of confidence	α	1-tailed test	2-tailed test
90%	0.10	1.28	1.645
95%	0.05	1.645	1.96
98%	0.02	2.05	2.33
99%	0.01	2.33	2.575

Example 1 *Twenty years ago, the average number of miles KSU students commuted to campus was 12.8 miles. A student group thinks that on campus housing and new apartment construction new campus has reduced commutes for students. Formulate as H_0 and H_A.*

$$H_0 : \mu = 12.8$$

$$H_A : \mu < 12.8$$

Example 2 *During a world series game, Mighty Casey comes up to bat. The announcers highlight his .310 batting average (which is really a percentage) for the regular season but then go on to say that those numbers don't apply to the playoffs. Formulate H_0 and H_A for the announcers comments.*

$$H_0 : p = .310$$

$$H_A : p \neq .310$$

In the previous section on confidence intervals we encountered the question of the number of hours KSU students watch Dodgeball. We answered the question there with a confidence interval. We can also tackle the problem with hypothesis testing.

Example 3 *The KSU Dodgeball exploratory committee claims that KSU students watch an average of 10 hours of Dodgeball per week during the season. A study group collects data from 50 KSU students which yields an average of 9.5 hours watching Dodgeball with a known population standard deviation of 3.5 hours. Has the KSU Dodgeball exploratory committee overestimated the average number of hours KSU students watch Dodgeball or is this just chance variation in the sample? Test at 95% confidence. Formulate H_0 and H_A.*

$$H_0 : \mu = 10.$$

$$H_A : \mu < 10$$

We use a one-tail test since the question indicates a direction in which we disagree with the null hypothesis.

Compute the test statistic for our KSU Dodgeball exploratory committee problem.

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{9.5 - 10}{\frac{3.5}{\sqrt{50}}} = -1.010\,2$$

The critical value for this test is 1.645. We compare the absolute value of the test statistic to the critical value. If the test statistic is larger then we reject the null in favor of the alternative. If the critical value is larger than the test statistic we fail to reject the null hypothesis. In this case, $1.645 > 1.0102$ and we fail to reject the null hypothesis. We can also use P-values to reject or fail to reject the null hypothesis. In this case the P-value is .1562. This P-value is the probability that we obtained our sample mean given that the null hypothesis is true. Since $.1562 > .05$, it is reasonable to get the sample results that we did. Again, we fail to reject the null hypothesis.

Section 8.2 Hypothesis Testing for a Mean with known Standard Deviation

On the TI 83/84 series of calculators, use the **Z-Test** command to conduct a hypothesis test for a mean with known population standard deviation.

Exercise 1 *The average GPA at a particular university is 2.37 with a known population standard deviation of 1.2. The 40 member fraternity Kappa Epsilon Gamma has an average GPA of 1.89. Dean Wormer believes the fraternity is a disgrace to academic standards with their low GPA. The fraternity believes their GPA is just an example of chance variation. At 99% significance, with whom do you agree?*

Exercise 2 *Josh's lifetime average word score in Scrabble is 15.5 points per turn with a standard deviation of 7 points. However, in 30 games with Brenda, his average score is 12.7 points per turn Brenda thinks Josh is distracted by her great beauty. Josh claims chance variation. At 90% significance, with whom do you agree?*

Exercise 3 *A particular network claims the average length of its commercial breaks is 2 minutes with a standard deviation of 30 seconds. Kay thinks the breaks are much longer than that. Over a period of two weeks, Kay tracks 267 commercial breaks that last an average of 2 minutes and 10 seconds. At 95% confidence, do you agree with Kay?*

1 Types of Errors

No level of confidence is perfect and so we may fail to make the correct decision. There are two types of errors that can creep in during hypothesis testing. We might reject the null when it is true (Type I error). We might fail to reject the null when we should (Type II error). Which type of error do you think is more dangerous to make? Which type of error does α measure?

	Reject H_0	Do not reject H_0
H_0 is true	Type I error	Correct!
H_0 is false	Correct!	Type II error

Which error is more dangerous? The type I error is the more dangerous of the two because one is going against the standard convention. Consider the process of a bank loaning money to a customer. A bank doesn't loan money to everyone who walks in off the streets. Thus, the null hypothesis is that a customer will not pay back a loan. If the bank fails to reject the null when the alternative is true then the bank has lost the opportunity to make interest from the amount of the loan. In this case the type II error is a missed opportunity. On the other hand, if the bank does not reject the null when it is true then the bank loses the amount of the loan. In this case the type I error is the loss of capitol. As usual, perspective comes into play as to which error is truly more dangerous. What is the borrower's perspective on the type I and II errors?

The probability of committing a type I error is α. Thus, the level of confidence directly tells you the probability of committing a type I error. The probability of committing a type II error is β. The techniques for computing β are beyond the scope of this class.

2 Exercises

1. Navidi/Monk Section 8.1: 7, 8, 13-24
2. Navidi/Monk Section 8.2: 35-39, 43, 44, 45, 51-55, 61, 62