4.3.1 Hypothesis testing framework:

Students from the 2011 YRBSS lifted weights (or performed other strength training exercises) 3.09 days per week on average. We want to determine if the yrbss samp data set provides strong evidence that YRBSS students selected in 2013 are lifting more or less than the 2011 YRBSS students, versus the other possibility that there has been no change.15 We simplify these three options into two competing hypotheses:

H0: The average days per week that YRBSS students lifted weights was the same for 2011 and 2013.

HA: The average days per week that YRBSS students lifted weights was di↵erent for 2013 than in 2011.

TIP: Hypothesis testing framework

The skeptic will not reject the null hypothesis (H0), unless the evidence in favor of the alternative hypothesis (HA) is so strong that she rejects H0 in favor of HA.

Guided Practice 4.18

A US court considers two possible claims about a defendant: she is either innocent or guilty. If we set these claims up in a hypothesis framework, which would be the null hypothesis and which the alternative?

The jury considers whether the evidence is so convincing (strong) that there is no reasonable doubt regarding the person’s guilt; in such a case, the jury rejects innocence (the null hypothesis) and concludes the defendant is guilty (alternative hypothesis).

Example 4.19

In the sample of 100 students from the 2013 YRBSS survey, the average number of days per week that students lifted weights was 2.78 days with a standard deviation of 2.56 days (coincidentally the same as days active). Compute a 95% confidence interval for the average for all students from the 2013 YRBSS survey. You can assume the conditions for the normal model are met.

The general formula for the confidence interval based on the normal distribution is

x¯ ± z\*SEx¯

We are given ¯x13 = 2.78, we use z\* = 1.96 for a 95% confidence level, and we can compute the standard error using the standard deviation divided by the square root of the sample size:

SEx¯ = s13/√n = 2.56/√100 = 0.256

Entering the sample mean, z?, and the standard error into the confidence interval formula results in (2.27, 3.29). We are 95% confident that the average number of days per week that all students from the 2013 YRBSS lifted weights was between 2.27 and 3.29 days.

Because the average of all students from the 2011 YRBSS survey is 3.09, which falls within the range of plausible values from the confidence interval, we cannot say the null hypothesis is implausible. That is, we fail to reject the null hypothesis, H0.

Guided Practice 4.20

Colleges frequently provide estimates of student expenses such as housing. A consultant hired by a community college claimed that the average student housing expense was $650 per month. What are the null and alternative hypotheses to test whether this claim is accurate?

H0: The average cost is $650 per month, µ = $650.

HA: The average cost is di↵erent than $650 per month, µ ≠ $650.

Guided Practice 4.21

The community college decides to collect data to evaluate the $650 per month claim. They take a random sample of 175 students at their school and obtain the data represented in Figure 4.11. Can we apply the normal model to the sample mean?

Applying the normal model requires that certain conditions are met. Because the data are a simple random sample and the sample (presumably) represents no more than 10% of all students at the college, the observations are independent. The sample size is also suciently large (n = 175) and the data exhibit strong skew. While the data are strongly skewed, the sample is suciently large that this is acceptable, and the normal model may be applied to the sample mean.

Example 4.22

The sample mean for student housing is $616.91 and the sample standard deviation is $128.65. Construct a 95% confidence interval for the population mean and evaluate the hypotheses of Guided Practice 4.20.

The standard error associated with the mean may be estimated using the sample standard deviation divided by the square root of the sample size. Recall that n = 175 students were sampled.

SE = s/√n = 128.65/√175 = 9.73

You showed in Guided Practice 4.21 that the normal model may be applied to the sample mean.

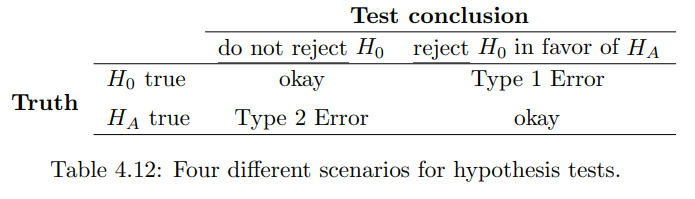
This ensures a 95% confidence interval may be accurately constructed:

x¯ ± z?SE ! 616.91 ± 1.96 ⇥ 9.73 ! (597.84, 635.98)

Because the null value $650 is not in the confidence interval, a true mean of $650 is implausible and we reject the null hypothesis. The data provide statistically significant evidence that the actual average housing expense is less than $650 per month.

4.3.3 Decision errors

Hypothesis tests are not flawless, since we can make a wrong decision in statistical hypothesis tests based on the data. For example, in the court system innocent people are sometimes wrongly convicted and the guilty sometimes walk free. However, the di↵erence is that in statistical hypothesis tests, we have the tools necessary to quantify how often we make such errors. There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a statement about which one might be true, but we might choose incorrectly. There are four possible scenarios, which are summarized in table 4.12

A Type 1 Error is rejecting the null hypothesis when H0 is actually true.

A Type 2 Error is failing to reject the null hypothesis when the alternative is actually true.

Guided Practice 4.23

In a US court, the defendant is either innocent (H0) or guilty (HA). What does a Type 1 Error represent in this context? What does a Type 2 Error represent? Table 4.12 may be useful.

If the court makes a Type 1 Error, this means the defendant is innocent (H0 true) but wrongly convicted. A Type 2 Error means the court failed to reject H0 (i.e. failed to convict the person) when she was in fact guilty (HA true).

Guided Practice 4.24

How could we reduce the Type 1 Error rate in US courts? What influence would this have on the Type 2 Error rate?

To lower the Type 1 Error rate, we might raise our standard for conviction from “beyond a reasonable doubt” to “beyond a conceivable doubt” so fewer people would be wrongly convicted.

However, this would also make it more difficult to convict the people who are actually guilty, so we would make more Type 2 Errors.

Guided Practice 4.25

How could we reduce the Type 2 Error rate in US courts? What influence would this have on the Type 1 Error rate?

To lower the Type 2 Error rate, we want to convict more guilty people. We could lower the standards for conviction from “beyond a reasonable doubt” to “beyond a little doubt”.

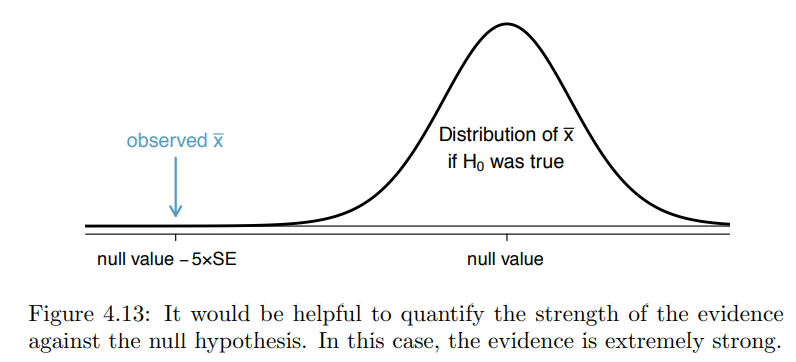
Lowering the bar for guilt will also result in more wrongful convictions, raising the Type 1 Error rate.

Consider the following two scenarios:

• The null value (the parameter value under the null hypothesis) is in the 95% confidence interval but just barely, so we would not reject H0. However, we might like to somehow say, quantitatively, that it was a close decision.

• The null value is very far outside of the interval, so we reject H0. However, we want to communicate that, not only did we reject the null hypothesis, but it wasn’t even close. Such a case is depicted in Figure 4.13.

In Section 4.3.4, we introduce a tool called the p-value that will be helpful in these cases. The p-value method also extends to hypothesis tests where confidence intervals cannot be easily constructed or applied.



TIP: One-sided and two-sided tests

When you are interested in checking for an increase or a decrease, but not both, use a one-sided test. When you are interested in any difference from the null value – an increase or decrease – then the test should be two-sided.

Guided Practice 4.27

In the sleep study, the sample standard deviation was 1.75 hours and the sample size is 110. Calculate the standard error of ¯x.

The standard error can be estimated from the sample standard deviation and the sample size: SEx¯ = sx/√ n = 1.75/√110 = 0.17

Guided Practice 4.26

A poll by the National Sleep Foundation found that college students average about 7 hours of sleep per night. Researchers at a rural school are interested in showing that students at their school sleep longer than seven hours on average, and they would like to demonstrate this using a sample of students. What would be an appropriate skeptical position for this research?

A skeptic would have no reason to believe that sleep patterns at this school are different than the sleep patterns at another school.

The researchers at the rural school conducted a simple random sample of n = 110 students on campus. They found that these students averaged 7.42 hours of sleep and the standard deviation of the amount of sleep for the students was 1.75 hours. A histogram of the sample is shown in Figure 4.14.

Before we can use a normal model for the sample mean or compute the standard error of the sample mean, we must verify conditions. (1) Because this is a simple random sample from less than 10% of the student body, the observations are independent. (2) The sample size in the sleep study is suciently large since it is greater than 30. (3) The data show strong skew in Figure 4.14 and the presence of a couple of outliers. This skew and the outliers are acceptable for a sample size of n = 110. With these conditions verified, the normal model can be safely applied to ¯x and we can reasonably calculate the standard error.

The hypothesis test for the sleep study will be evaluated using a significance level of ↵ = 0.05. We want to consider the data under the scenario that the null hypothesis is true. In this case, the sample mean is from a distribution that is nearly normal and has mean 7 and standard deviation of about SEx¯ = 0.17. Such a distribution is shown in Figure 4.15.

The shaded tail in Figure 4.15 represents the chance of observing such a large mean, conditional on the null hypothesis being true. That is, the shaded tail represents the p-value. We shade all means larger than our sample mean, ¯x = 7.42, because they are more favorable to the alternative hypothesis than the observed mean.

We compute the p-value by finding the tail area of this normal distribution, which we learned to do in Section 3.1. First compute the Z-score of the sample mean, ¯x = 7.42

Z = x¯- null value/SEx¯ = 7.42-7/0.17 = 2.47

Using the normal probability table, the lower unshaded area is found to be 0.993. Thus the shaded area is 1-0.993 = 0.007. If the null hypothesis is true, the probability of observing a sample mean at least as large as 7.42 hours for a sample of 110 students is only 0.007.

That is, if the null hypothesis is true, we would not often see such a large mean.

We evaluate the hypotheses by comparing the p-value to the significance level.

Because the p-value is less than the significance level (p-value = 0.007 < 0.05 = α), we reject the null hypothesis.

What we observed is so unusual with respect to the null hypothesis that it casts serious doubt on H0 and provides strong evidence favoring HA.

p-value as a tool in hypothesis testing

The smaller the p-value, the stronger the data favor HA over H0. A small p-value (usually < 0.05) corresponds to sufficient evidence to reject H0 in favor of HA.

The ideas below review the process of evaluating hypothesis tests with p-values:

• The null hypothesis represents a skeptic’s position or a position of no difference. We reject this position only if the evidence strongly favors HA.

• A small p-value means that if the null hypothesis is true, there is a low probability of seeing a point estimate at least as extreme as the one we saw. We interpret this as strong evidence in favor of the alternative.

• We reject the null hypothesis if the p-value is smaller than the significance level, α, which is usually 0.05. Otherwise, we fail to reject H0.

• We should always state the conclusion of the hypothesis test in plain language so non-statisticians can also understand the results.

The p-value is constructed in such a way that we can directly compare it to the significance level (α) to determine whether or not to reject H0. This method ensures that the Type 1 Error rate does not exceed the significance level standard.

Guided Practice 4.28

If the null hypothesis is true, how often should the p-value be less than 0.05?

About 5% of the time. If the null hypothesis is true, then the data only has a 5% chance of being in the 5% of data most favorable to HA.

Guided Practice 4.29

Suppose we had used a significance level of 0.01 in the sleep study. Would the evidence have been strong enough to reject the null hypothesis? (The p-value was 0.007.) What if the significance level was α = 0.001?

We reject the null hypothesis whenever p-value < α. Thus, we would still reject the null hypothesis if α = 0.01 but not if the significance level had been α= 0.001.

Guided Practice 4.30

Ebay might be interested in showing that buyers on its site tend to pay less than they would for the corresponding new item on Amazon. We’ll research this topic for one particular product: a video game called Mario Kart for the Nintendo Wii. During early October 2009, Amazon sold this game for $46.99. Set up an appropriate (one-sided!) hypothesis test to check the claim that E-bay buyers pay less during auctions at this same time.

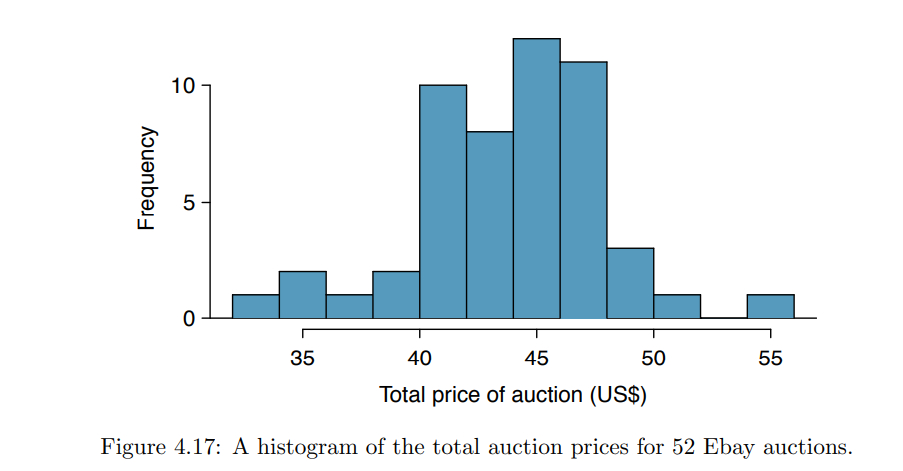
The skeptic would say the average is the same on Ebay, and we are interested in showing the average price is lower.

H0: The average auction price on Ebay is equal to (or more than) the price on Amazon. We write only the equality in the statistical notation: µebay = 46.99.

HA: The average price on Ebay is less than the price on Amazon, µebay < 46.99.

Guided Practice 4.31

During early October 2009, 52 Ebay auctions were recorded for Mario Kart. 28 The total prices for the auctions are presented using a histogram in Figure 4.17, and we may like to apply the normal model to the sample mean. Check the three conditions required for applying the normal model: (1) independence, (2) at least 30 observations, and (3) the data are not strongly skewed.



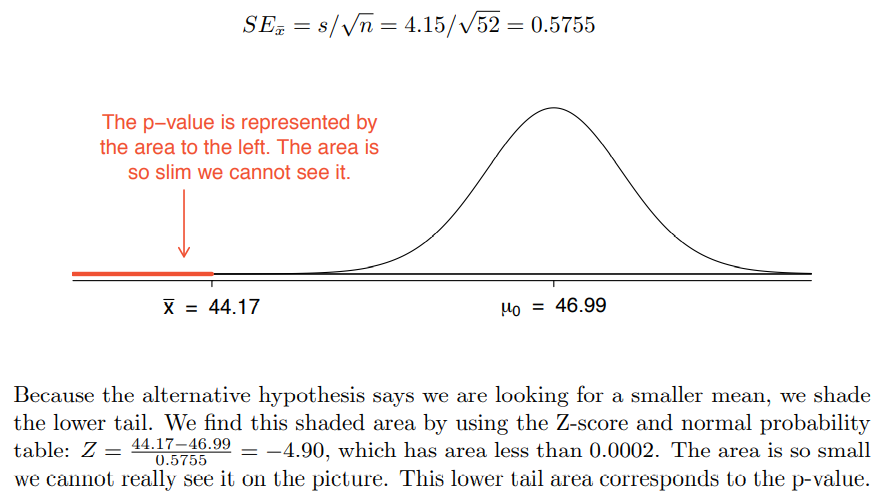
(1) The independence condition is unclear. We will make the assumption that the observations are independent, which we should report with any final results.

(2) The sample size is sufficiently large: n = 52 30.

(3) The data distribution is not strongly skewed; it is approximately symmetric.

Example 4.32

The average sale price of the 52 Ebay auctions for Wii Mario Kart was $44.17 with a standard deviation of $4.15. Does this provide sufficient evidence to reject the null hypothesis in Guided Practice 4.30? Use a significance level of α = 0.01. The hypotheses were set up and the conditions were checked in Exercises 4.30 and 4.31. The next step is to find the standard error of the sample mean and produce a sketch to help find the p-value.



Because the p-value is so small – specifically, smaller than ↵ = 0.01 – this provides suciently strong evidence to reject the null hypothesis in favor of the alternative. The data provide statistically significant evidence that the average price on Ebay is lower than Amazon’s asking price.

