This Z score is so large it isn't even in the table, which ensures the single tail area will be 0.0002 or smaller. Since the p-value corresponds to both tails in this case and the normal distribution is symmetric, the p-value can be estimated as twice the one-tail area:

$$\text{p-value} = 2 \times (\text{one tail area}) \approx 2 \times 0.0002 = 0.0004$$

Because the p-value is less than 0.05, we reject the null hypothesis. We have found convincing evidence that Amazon is, on average, cheaper than the UCLA bookstore for UCLA course textbooks.

Exercise 5.3 Create a 95% confidence interval for the average price difference between books at the UCLA bookstore and books on Amazon.[3]

5.2 Difference of two means

In this section we consider a difference in two population means, $\mu_1 - \mu_2$, under the condition that the data are not paired. The methods are similar in theory but different in the details. Just as with a single sample, we identify conditions to ensure a point estimate of the difference $\bar{x}_1 - \bar{x}_2$ is nearly normal. Next we introduce a formula for the standard error, which allows us to apply our general tools from Section 4.5.

We apply these methods to two examples: participants in the 2012 Cherry Blossom Run and newborn infants. This section is motivated by questions like "Is there convincing evidence that newborns from mothers who smoke have a different average birth weight than newborns from mothers who don't smoke?"

5.2.1 Point estimates and standard errors for differences of means

We would like to estimate the average difference in run times for men and women using the `run10Samp` data set, which was a simple random sample of 45 men and 55 women from all runners in the 2012 Cherry Blossom Run. Table 5.5 presents relevant summary statistics, and box plots of each sample are shown in Figure 5.6.

	men	women
$\bar{x}$	87.65	102.13
s	12.5	15.2
n	45	55

Table 5.5: Summary statistics for the run time of 100 participants in the 2009 Cherry Blossom Run.

The two samples are independent of one-another, so the data are not paired. Instead a point estimate of the difference in average 10 mile times for men and women, $\mu_w - \mu_m$, can be found using the two sample means:

$$\bar{x}_w - \bar{x}_m \;=\; 102.13 - 87.65 \;=\; 14.48$$

[3]Conditions have already verified and the standard error computed in Example 5.2. To find the interval, identify $z^\star$ (1.96 for 95% confidence) and plug it, the point estimate, and the standard error into the confidence interval formula:

$$\text{point estimate} \;\pm\; z^\star SE \quad\to\quad 12.76 \;\pm\; 1.96 \times 1.67 \quad\to\quad (9.49, 16.03)$$

We are 95% confident that Amazon is, on average, between \$9.49 and \$16.03 cheaper than the UCLA bookstore for UCLA course books.

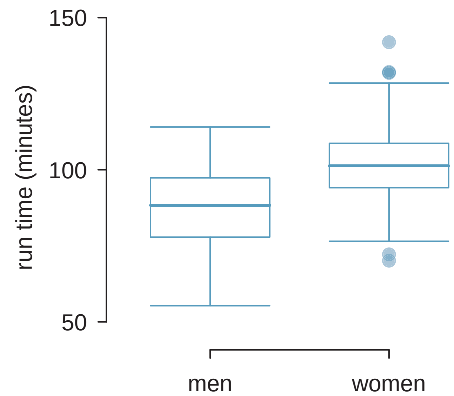


Figure 5.6: Side-by-side box plots for the sample of 2009 Cherry Blossom Run participants.

Because we are examining two simple random samples from less than 10% of the population, each sample contains at least 30 observations, and neither distribution is strongly skewed, we can safely conclude the sampling distribution of each sample mean is nearly normal. Finally, because each sample is independent of the other (e.g. the data are not paired), we can conclude that the difference in sample means can be modeled using a normal distribution.[4]

> **Conditions for normality of $\bar{x}_1 - \bar{x}_2$**
> If the sample means, $\bar{x}_1$ and $\bar{x}_2$, each meet the criteria for having nearly normal sampling distributions and the observations in the two samples are independent, then the difference in sample means, $\bar{x}_1 - \bar{x}_2$, will have a sampling distribution that is nearly normal.

We can quantify the variability in the point estimate, $\bar{x}_w - \bar{x}_m$, using the following formula for its standard error:

$$SE_{\bar{x}_w - \bar{x}_m} = \sqrt{\frac{\sigma_w^2}{n_w} + \frac{\sigma_m^2}{n_m}}$$

We usually estimate this standard error using standard deviation estimates based on the samples:

$$\begin{aligned} SE_{\bar{x}_w - \bar{x}_m} &= \sqrt{\frac{\sigma_w^2}{n_w} + \frac{\sigma_m^2}{n_m}} \\ &\approx \sqrt{\frac{s_w^2}{n_w} + \frac{s_m^2}{n_m}} = \sqrt{\frac{15.2^2}{55} + \frac{12.5^2}{45}} = 2.77 \end{aligned}$$

Because each sample has at least 30 observations ($n_w = 55$ and $n_m = 45$), this substitution using the sample standard deviation tends to be very good.

[4]Probability theory guarantees that the difference of two independent normal random variables is also normal. Because each sample mean is nearly normal and observations in the samples are independent, we are assured the difference is also nearly normal.

Distribution of a difference of sample means

The sample difference of two means, $\bar{x}_1 - \bar{x}_2$, is nearly normal with mean $\mu_1 - \mu_2$ and estimated standard error

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad (5.4)$$

when each sample mean is nearly normal and all observations are independent.

5.2.2 Confidence interval for the difference

When the data indicate that the point estimate $\bar{x}_1 - \bar{x}_2$ comes from a nearly normal distribution, we can construct a confidence interval for the difference in two means from the framework built in Chapter 4. Here a point estimate, $\bar{x}_w - \bar{x}_m = 14.48$, is associated with a normal model with standard error $SE = 2.77$. Using this information, the general confidence interval formula may be applied in an attempt to capture the true difference in means, in this case using a 95% confidence level:

$$\text{point estimate} \ \pm\ z^\star SE \quad \to \quad 14.48 \ \pm\ 1.96 \times 2.77 \quad \to \quad (9.05, 19.91)$$

Based on the samples, we are 95% confident that men ran, on average, between 9.05 and 19.91 minutes faster than women in the 2012 Cherry Blossom Run.

⊙ **Exercise 5.5** What does 95% confidence mean?[5]

⊙ **Exercise 5.6** We may be interested in a different confidence level. Construct the 99% confidence interval for the population difference in average run times based on the sample data.[6]

5.2.3 Hypothesis tests based on a difference in means

A data set called `baby_smoke` represents a random sample of 150 cases of mothers and their newborns in North Carolina over a year. Four cases from this data set are represented in Table 5.7. We are particularly interested in two variables: `weight` and `smoke`. The `weight` variable represents the weights of the newborns and the `smoke` variable describes which mothers smoked during pregnancy. We would like to know, is convincing evidence that newborns from mothers who smoke have a different average birth weight than newborns from mothers who don't smoke? We will use the North Carolina sample to try to answer this question. The smoking group includes 50 cases and the nonsmoking group contains 100 cases, represented in Figure 5.8.

[5] If we were to collected many such samples and create 95% confidence intervals for each, then about 95% of these intervals would contain the population difference, $\mu_w - \mu_m$.

[6] The only thing that changes is $z^\star$: we use $z^\star = 2.58$ for a 99% confidence level. (If the selection of $z^\star$ is confusing, see Section 4.2.4 for an explanation.) The 99% confidence interval: $14.48 \ \pm\ 2.58 \times 2.77 \to (7.33, 21.63)$. We are 99% confident that the true difference in the average run times between men and women is between 7.33 and 21.63 minutes.

	fAge	mAge	weeks	weight	sexBaby	smoke
1	NA	13	37	5.00	female	nonsmoker
2	NA	14	36	5.88	female	nonsmoker
3	19	15	41	8.13	male	smoker
⋮	⋮	⋮	⋮	⋮	⋮	
150	45	50	36	9.25	female	nonsmoker

Table 5.7: Four cases from the `baby_smoke` data set. The value "NA", shown for the first two entries of the first variable, indicates that piece of data is missing.

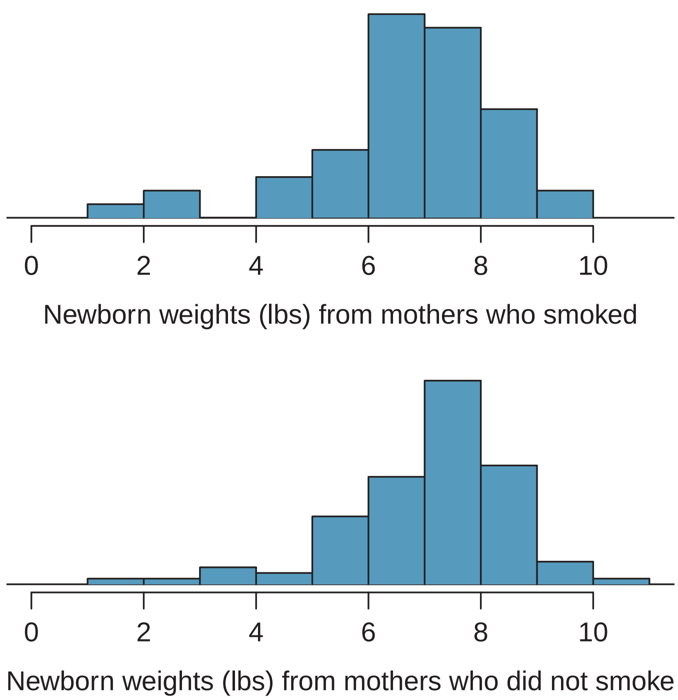


Figure 5.8: The top panel represents birth weights for infants whose mothers smoked. The bottom panel represents the birth weights for infants whose mothers who did not smoke. Both distributions exhibit strong skew.

Example 5.7 Set up appropriate hypotheses to evaluate whether there is a relationship between a mother smoking and average birth weight.

The null hypothesis represents the case of no difference between the groups.

H_0: There is no difference in average birth weight for newborns from mothers who did and did not smoke. In statistical notation: $\mu_n - \mu_s = 0$, where μ_n represents non-smoking mothers and μ_s represents mothers who smoked.

H_A: There is some difference in average newborn weights from mothers who did and did not smoke ($\mu_n - \mu_s \neq 0$).

Summary statistics are shown for each sample in Table 5.9. Because the data come from a simple random sample and consist of less than 10% of all such cases, the observations are independent. Additionally, each group's sample size is at least 30 and the skew in each sample distribution is strong (see Figure 5.8). The skew is reasonable for these sample sizes of 50 and 100. Therefore, each sample mean is associated with a nearly normal distribution.

	`smoker`	`nonsmoker`
mean	6.78	7.18
st. dev.	1.43	1.60
samp. size	50	100

Table 5.9: Summary statistics for the `baby_smoke` data set.

Exercise 5.8 (a) What is the point estimate of the population difference, $\mu_n - \mu_s$? (b) Can we use a normal distribution to model this difference? (c) Compute the standard error of the point estimate from part (a).[7]

Example 5.9 If the null hypothesis from Example 5.7 was true, what would be the expected value of the point estimate? And the standard deviation associated with this estimate? Draw a picture to represent the p-value.

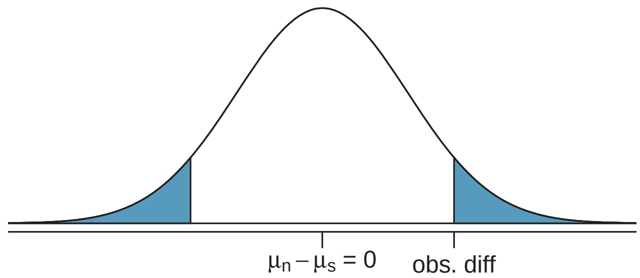
If the null hypothesis was true, then we expect to see a difference near 0. The standard error corresponds to the standard deviation of the point estimate: 0.26. To depict the p-value, we draw the distribution of the point estimate as though H_0 was true and shade areas representing at least as much evidence against H_0 as what was observed. Both tails are shaded because it is a two-sided test.

---

[7](a) The difference in sample means is an appropriate point estimate: $\bar{x}_n - \bar{x}_s = 0.40$. (b) Because the samples are independent and each sample mean is nearly normal, their difference is also nearly normal. (c) The standard error of the estimate can be estimated using Equation (5.4):

$$SE = \sqrt{\frac{\sigma_n^2}{n_n} + \frac{\sigma_s^2}{n_s}} \approx \sqrt{\frac{s_n^2}{n_n} + \frac{s_s^2}{n_s}} = \sqrt{\frac{1.60^2}{100} + \frac{1.43^2}{50}} = 0.26$$

The standard error estimate should be sufficiently accurate since the conditions were reasonably satisfied.



Example 5.10 Compute the p-value of the hypothesis test using the figure in Example 5.9, and evaluate the hypotheses using a significance level of $\alpha = 0.05$.

Since the point estimate is nearly normal, we can find the upper tail using the Z score and normal probability table:

$$Z = \frac{0.40 - 0}{0.26} = 1.54 \quad \rightarrow \quad \text{upper tail } = 1 - 0.938 = 0.062$$

Because this is a two-sided test and we want the area of both tails, we double this single tail to get the p-value: 0.124. This p-value is larger than the significance value, 0.05, so we fail to reject the null hypothesis. There is insufficient evidence to say there is a difference in average birth weight of newborns from North Carolina mothers who did smoke during pregnancy and newborns from North Carolina mothers who did not smoke during pregnancy.

Exercise 5.11 Does the conclusion to Example 5.10 mean that smoking and average birth weight are unrelated?[8]

Exercise 5.12 If we made a Type 2 Error and there is a difference, what could we have done differently in data collection to be more likely to detect such a difference?[9]

5.2.4 Summary for inference of the difference of two means

When considering the difference of two means, there are two common cases: the two samples are paired or they are independent. (There are instances where the data are neither paired nor independent.) The paired case was treated in Section 5.1, where the one-sample methods were applied to the differences from the paired observations. We examined the second and more complex scenario in this section.

When applying the normal model to the point estimate $\bar{x}_1 - \bar{x}_2$ (corresponding to unpaired data), it is important to verify conditions before applying the inference framework using the normal model. First, each sample mean must meet the conditions for normality; these conditions are described in Chapter 4 on page 168. Secondly, the samples must be collected independently (e.g. not paired data). When these conditions are satisfied, the general inference tools of Chapter 4 may be applied.

For example, a confidence interval may take the following form:

$$\text{point estimate} \ \pm \ z^{\star} SE$$

[8] Absolutely not. It is possible that there is some difference but we did not detect it. If this is the case, we made a Type 2 Error.

[9] We could have collected more data. If the sample sizes are larger, we tend to have a better shot at finding a difference if one exists.

When we compute the confidence interval for $\mu_1 - \mu_2$, the point estimate is the difference in sample means, the value $z^\star$ corresponds to the confidence level, and the standard error is computed from Equation (5.4) on page 217. While the point estimate and standard error formulas change a little, the framework for a confidence interval stays the same. This is also true in hypothesis tests for differences of means.

In a hypothesis test, we apply the standard framework and use the specific formulas for the point estimate and standard error of a difference in two means. The test statistic represented by the Z score may be computed as

$$Z = \frac{\text{point estimate} - \text{null value}}{SE}$$

When assessing the difference in two means, the point estimate takes the form $\bar{x}_1 - \bar{x}_2$, and the standard error again takes the form of Equation (5.4) on page 217. Finally, the null value is the difference in sample means under the null hypothesis. Just as in Chapter 4, the test statistic Z is used to identify the p-value.

5.2.5 Examining the standard error formula

The formula for the standard error of the difference in two means is similar to the formula for other standard errors. Recall that the standard error of a single mean, $\bar{x}_1$, can be approximated by

$$SE_{\bar{x}_1} = \frac{s_1}{\sqrt{n_1}}$$

where s_1 and n_1 represent the sample standard deviation and sample size.

The standard error of the difference of two sample means can be constructed from the standard errors of the separate sample means:

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{SE_{\bar{x}_1}^2 + SE_{\bar{x}_2}^2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{5.13}$$

This special relationship follows from probability theory.

⊙ **Exercise 5.14** Prerequisite: Section 2.4. We can rewrite Equation (5.13) in a different way:

$$SE_{\bar{x}_1 - \bar{x}_2}^2 = SE_{\bar{x}_1}^2 + SE_{\bar{x}_2}^2$$

Explain where this formula comes from using the ideas of probability theory.[10]

5.3 One-sample means with the t distribution

The motivation in Chapter 4 for requiring a large sample was two-fold. First, a large sample ensures that the sampling distribution of $\bar{x}$ is nearly normal. We will see in Section 5.3.1 that if the population data are nearly normal, then $\bar{x}$ is also nearly normal regardless of the

[10] The standard error squared represents the variance of the estimate. If X and Y are two random variables with variances σ_x^2 and σ_y^2, then the variance of $X - Y$ is $\sigma_x^2 + \sigma_y^2$. Likewise, the variance corresponding to $\bar{x}_1 - \bar{x}_2$ is $\sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2$. Because $\sigma_{\bar{x}_1}^2$ and $\sigma_{\bar{x}_2}^2$ are just another way of writing $SE_{\bar{x}_1}^2$ and $SE_{\bar{x}_2}^2$, the variance associated with $\bar{x}_1 - \bar{x}_2$ may be written as $SE_{\bar{x}_1}^2 + SE_{\bar{x}_2}^2$.