

Solutions to Homework 9

Statistics 302 Professor Larget

Textbook Exercises

In Exercise 6.160, situations comparing two proportions are described. In each case, determine whether the situation involves comparing proportions for two groups or comparing two proportions from the same group. State whether the methods of this section apply to the difference in proportions.

6.160 (a) In a taste test, compare the proportion of tasters who prefer one brand of cola to the proportion who prefer the other brand.

(b) Compare the proportion of males who voted in the last election to the proportions of females who voted in the last election.

(c) Compare the graduation rate (proportion to graduate) of students on an athletic scholarship to the graduation rate of students who are not on an athletic scholarship.

(d) Compare the proportion of voters who vote in favor of a school budget to the proportion who vote against the budget.

Solution

(a) This compares two proportions (one brand of cola vs the other brand) drawn from the same group (tasters). The methods of this section do not apply to this type of difference in proportions.

(b) This compares the proportion who voted using two different groups (males vs females). The methods of this section are appropriate for this type of difference in proportions.

(c) This compares the proportion who graduate using two different groups (athletes and non-athletes). The methods of this section are appropriate for this type of difference in proportions.

(d) This compares two proportions (proportion in favor vs proportion opposed) drawn from the same group (voters). The methods of this section do not apply to this type of difference in proportions.

6.164 Is Argentina or Bolivia More Rural? We see in the **AllCountries** dataset that the percent of the population living in rural areas is 8.0 in Argentina and 34.3 in Bolivia. Suppose we take random samples of size 200 from each country, and compute the difference in sample proportions $\hat{p}_A - \hat{p}_B$, where \hat{p}_A represents the sample proportion living in rural areas in Argentina and \hat{p}_B represents the proportion of the sample that lives in rural areas in Bolivia.

(a) Find the mean and standard deviation of the distribution of differences in sample proportions, $\hat{p}_A - \hat{p}_B$.

(b) If the sample sizes are large enough for the Central Limit Theorem to apply, draw a curve showing the shape of the sampling distribution. Include at least three values on the horizontal axis.

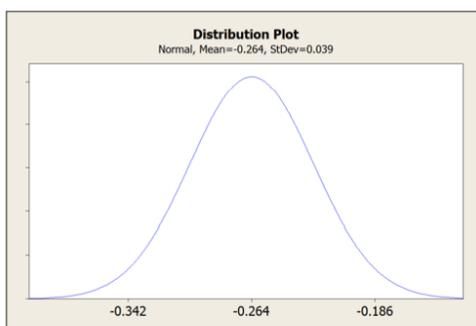
(c) Using the graph drawing in part (b), are we likely to see a difference in sample proportions as large in magnitude as -0.4 ? As large as -0.3 ? Explain.

Solution

(a) The differences in sample proportions will be centered at the difference in population proportions, so will have a mean of $p_A - p_B = .080 - .344 = -0.264$, and a standard deviation equal to the standard error SE. We have

$$SE = \sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}} = \sqrt{\frac{0.08(0.92)}{200} + \frac{0.344(0.656)}{200}} = 0.039$$

(b) The sample sizes of 200 are large enough for the normal distribution to apply. The distribution of differences of proportions will be $N(-0.264, 0.039)$. We use technology to sketch the graph, or we can sketch it by hand, noting that the differences in sample proportions will be centered at -0.264 and roughly 95% of the distribution lies within two standard deviations on either side of the center. This goes between $-0.264 \pm 2(0.039)$ or -0.342 to -0.186 .



c) We see from the figure that -0.4 is a very unlikely result for the difference in sample proportions (it's more than 3 SE's from the center) whereas -0.3 is plausible (it's less than one SE from the center).

6.176 Has Support for Capital Punishment Changed over Time? The General Social Survey (GSS) has been collecting demographic, behavioral, and attitudinal information since 1972 to monitor changes within the US and to compare the US to other nations. Support for the capital punishment (the death penalty) in the US is shown in 1974 and in 2006 in the two-way table below. Find a 95% confidence interval for the change in the proportion supporting capital punishment between 1974 and 2006. Is it plausible that the proportion supporting capital punishment has not changed?

Year	Favor	Oppose	Total
1974	937	473	1410
2006	1945	870	2815

Solution

Letting \hat{p}_1 and \hat{p}_2 represent the proportion supporting capital punishment in 2006 and 1974, respectively, we have

$$\hat{p}_1 = 1945 / 2815 = 0.691 \text{ and } \hat{p}_2 = 937 / 1410 = 0.665$$

The sample sizes are both large, so it is reasonable to use a normal distribution. For 95% confidence the standard normal endpoint is $z^* = 1.96$. This gives

$$\begin{aligned}
 (\hat{p}_1 - \hat{p}_2) &\pm z^* \cdot \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \\
 (0.691 - 0.665) &\pm 1.96^* \cdot \sqrt{\frac{0.691(1 - 0.691)}{2815} + \frac{0.665(1 - 0.665)}{1410}} \\
 0.026 &\pm 0.030 \\
 -0.004 &\text{ to } 0.056
 \end{aligned}$$

We are 95% sure that, between 1974 and 2006, the percent change in support of the death penalty is between a decrease of 0.4% and an increase of 5.6%. Since a difference of zero (no change) is within this interval, it is plausible that there have been no change in support or opposition to the death penalty in this 22-year period.

6.178 Mental Tags on Penguins and Breeding Success Data 1.3 on page 10 discusses a study designed to test whether applying metal tags is detrimental to penguins. One variable examined is the survival rate 10 years after tagging. The scientists observed that 10 of the 50 metal tagged penguins survived, compared to 18 of the 50 electronic tagged penguins. Construct a 90% confidence interval for the difference in proportion surviving between the metal and electronic tagged penguins ($p_M - p_E$). Interpret the result.

Solution

We have $\hat{p}_M = 10/50 = 0.20$ and $\hat{p}_E = 18/50 = 0.36$. The 95% confidence interval for $p_M - p_E$ is

$$\begin{aligned}
 (\hat{p}_M - \hat{p}_E) &\pm z^* \cdot \sqrt{\frac{\hat{p}_M(1 - \hat{p}_M)}{n_M} + \frac{\hat{p}_E(1 - \hat{p}_E)}{n_E}} \\
 (0.20 - 0.36) &\pm 1.645 \cdot \sqrt{\frac{0.20(0.80)}{50} + \frac{0.36(0.64)}{50}} \\
 -0.16 &\pm 0.145 \\
 -0.305 &\text{ to } -0.015
 \end{aligned}$$

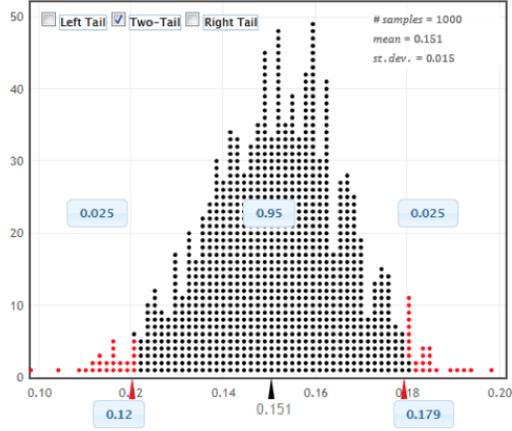
We are 90% sure that the survival rate for metal tagged penguins is between 0.305 and 0.015 less than for electronic tagged penguins. This shows a significant difference at a 10% level.

Comparing Normal and Bootstrap Confidence Intervals In Exercise 6.185, find a 95% confidence interval for the difference in proportions two ways, using *StatKey* or other technology and percentiles from a bootstrap distribution and using the normal distribution and the formula for standard error. Compare the results.

6.185 Difference in proportion who use text messaging, using $\hat{p}_t = 0.87$ with $n = 800$ for teens and $\hat{p}_a = 0.72$ with $n = 2252$ for adults.

Solution

We use *StatKey* or other technology to create a bootstrap distribution with at least 1000 simulated differences in proportion. We find the endpoints that contain 95% of the simulated statistics and see that this 95% confidence interval is 0.120 to 0.179.



Using the normal distribution and the formula for standard error, we have

$$(0.87 - 0.72) \pm 1.96 \cdot \sqrt{\frac{0.87(0.13)}{800} + \frac{0.72(0.28)}{2252}} = 0.15 \pm 0.030 = (0.12, 0.18)$$

The two methods give very similar confidence intervals.

6.195 Babies Learn Early Who They Can Trust A new study indicates that babies many choose not to learn from someone they don't trust. A group of 60 babies, aged 13 to 16 months, were randomly divided into two groups. Each baby watched an adult express great excitement while looking into a box. The babies were then shown the box and it either had a toy in it (the adult could be trusted) or it was empty (the adult was not reliable). The same adult then turned on a push-on light with her forehead, and the number of babies who imitated the adult's behavior by doing the same thing was counted. The results are in the table below. Test at a 5% level to see if there is evidence that babies are more likely to imitate those they consider reliable.

	Imitated	Did not imitate
Reliable	18	12
Unreliable	10	20

Solution

The hypotheses are

$$H_0 : p_1 = p_2$$

$$H_a : p_1 > p_2$$

where p_1 represents the proportion of babies imitating an adult considered reliable and p_2 represents the proportion of babies imitating an adult considered unreliable. There are at least 10 in each group (just barely) so we may use the normal distribution. We compute the sample proportions and the pooled sample proportion:

$$\hat{p}_1 = \frac{18}{30} = 0.60$$

$$\hat{p}_2 = \frac{10}{30} = 0.333$$

$$\hat{p} = \frac{28}{60} = 0.467$$

The standardized test statistic is

$$\frac{\text{Sample statistic} - \text{Null parameter}}{SE} = \frac{(0.60 - 0.333) - 0}{\sqrt{\frac{0.467(0.533)}{30} + \frac{0.467(0.533)}{30}}} = 2.073.$$

This is an upper-tail test, so the p-value is the area above 2.073 in a standard normal distribution. Using technology we see that the p-value is 0.0191. This p-value is less than the 0.05 significance level, so we reject H_0 . There is evidence that babies are more likely to imitate an adult who they believe is reliable. (And this effect is seen after one instance of being unreliable. Imagine the impact of repeated instances over time. It pays to be trustworthy!)

6.202 Green Tea and Prostate Cancer A preliminary study suggests a benefit from green tea for those at risk of prostate cancer. The study involved 60 men with PIN lesions, some of which turn into prostate cancer. Half the men, randomly determined, were given 600 mg a day of a green tea extract while the other half were given a placebo. The study was double-blind, and the results after one year are shown in the table below. Does the sample provide evidence that taking green tea extract reduces the risk of developing prostate cancer?

Treatment	Cancer	No Cancer
Green Tea	1	29
Placebo	9	21

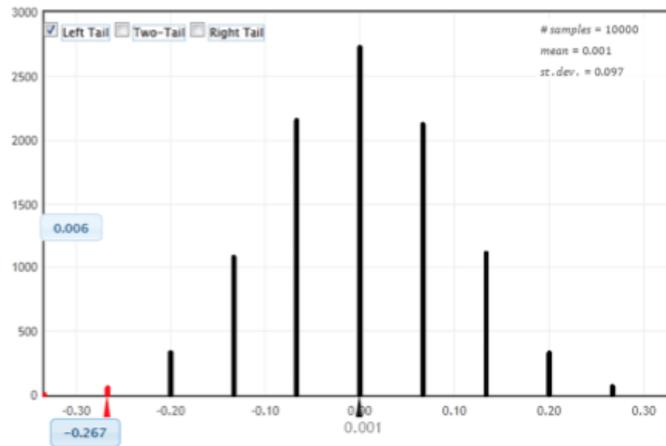
Solution

We cannot use the normal distribution and a standardized test statistic to conduct this test! Notice that two of the numbers in the cells are less than 10, so this data does not satisfy the conditions of the Central Limit Theorem for proportions. Luckily, randomization tests for experiments do not have any assumptions, so we conduct the test using a randomization test. Using p_T to denote the proportion of men with PIN lesions who get prostate cancer after taking green tea extract for a year and p_C to denote the proportion of men with PIN lesions who get prostate cancer after taking a placebo for a year, we have as our hypotheses:

$$H_0 : p_T = p_C$$

$$H_a : p_T < p_C$$

Using StatKey or other appropriate technology, we conduct a randomization test for this difference in proportions using the data in the table where the original difference is $\hat{p}_T - \hat{p}_C = 1/30 - 9/30 = -0.267$. The dotplot below shows results for 10,000 differences in proportions, generated by re-assigning the “Green Tea” and “Placebo” groups at random to the cancer results.



The p-value is very small, 0.006 from these randomizations, so we reject the null hypothesis and find strong evidence that green tea extract offers some benefit against prostate cancer.

6.223 Mathematics Scores by Native Language The distribution of sample means $\bar{x}_N - \bar{x}_E$, where \bar{x}_N represents the mean Mathematics score for a sample of 100 people whose native language is English, is centered at 10 with a standard deviation of 17.41. Give notation and define the quantity we are estimating with these sample differences. In the population of all students taking the test, who scored higher on average, non-native English speakers or native English speakers?

Solution

The quantity we are estimating is $\mu_N - \mu_E$, where μ_N represents the average score on this test by all non-native English speakers who took the test and μ_E represents the average score on the test by all native English speakers who took the test. The distribution of differences in sample means, $\bar{x}_N - \bar{x}_E$, is centered at 10 which means that the difference in the population means is $\mu_N - \mu_E = 10$. Since $\mu_N - \mu_E$ is positive, we must have $\mu_N > \mu_E$ and find that non-native English speakers scored higher, on average, on the Mathematics test. (This is quite impressive, since the test is given in English!)

6.238 Does Red Increase Men’s Attraction to Women? Exercise 1.89 on page 42 described a recent study which examines the impact of the color red on how attractive men perceive women to be. In the study, men were randomly divided into two groups and were asked to rate the attractiveness of a women on a scale of 1 (not at all attractive) to 9 (extremely attractive). Men in one group were shown pictures of women on a white background while the men in the other group were shown the same pictures of women on a red background. The results are shown in the table below and the data for both groups are reasonably symmetric with no outliers. To determine the possible effect size of the red background over the white, find and interpret a 90% confidence interval for the difference in mean attractiveness rating.

Color	n	\bar{x}	s
Red	15	7.2	0.6
While	12	6.1	0.4

Solution

We want to estimate the size of the difference in the mean rating with the red background (μ_R) compared to the white background (μ_W). We estimate the difference in population means using the difference in sample means $\bar{x}_R - \bar{x}_W$, where \bar{x}_R represents the mean rating in the sample using red and \bar{x}_W represents the mean rating in the sample using white. For a 90% confidence interval with degrees of freedom equal to 11 (the smaller sample size minus one), we use $t^* = 1.80$. We have:

$$\begin{aligned} \text{Sample statistic} &\pm t^* \cdot SE \\ (\bar{x}_R - \bar{x}_W) &\pm 1.80 \sqrt{\frac{s_R^2}{n_R} + \frac{s_W^2}{n_W}} \\ (7.2 - 6.1) &\pm 1.80 \sqrt{\frac{0.6^2}{15} + \frac{0.4^2}{12}} \\ 1.1 &\pm 0.35 \\ 0.75 &\text{ to } 1.45 \end{aligned}$$

We are 90% confident that men's average rating of women's attractiveness on a 9-point scale will be between 0.75 and 1.45 points higher when the picture is displayed on a red background rather than a white background.

6.239 Light at Night and Weight Gain A study described in Data A.1 on page 136 found that mice exposed to light at night gained substantially more weight than mice who had complete darkness at night, despite the fact that calorie intake and activity levels were the same for the two groups. How large is the effect of light on weight gain? In the study, 27 mice were randomly divided into two groups. The 8 mice with darkness at night gained an average of 5.9 grams in body mass, with a standard deviation of 1.9 grams. The 19 mice with light at night gained an average of 9.4 grams with a standard deviation of 3.2 grams. We see in the figure in the book that there is no extreme skewness or extreme outliers, so it is appropriate to use a t-distribution. Find and interpret a 99% confidence interval for the difference in mean weight gain.

Solution

We estimate the difference in population means using the difference in sample means $\bar{x}_L - \bar{x}_D$, where x_L represents the mean weight gain of the mice in light and x_D represents the mean weight gain of mice in darkness. For a 99% confidence interval with degrees of freedom equal to 7, we use $t^* = 3.50$. We have:

$$\begin{aligned} \text{Sample statistic} &\pm t^* \cdot SE \\ (\bar{x}_L - \bar{x}_D) &\pm t^* \sqrt{\frac{s_L^2}{n_L} + \frac{s_D^2}{n_D}} \\ (9.4 - 5.9) &\pm 3.50 \sqrt{\frac{3.2^2}{19} + \frac{1.9^2}{8}} \\ 3.4 &\pm 3.48 \\ 0.02 &\text{ to } 6.98 \end{aligned}$$

We are 99% confident that mice with light at night will gain, on average, between 0.02 grams and 6.98 grams more than mice in darkness.

6.259 Diet Cola and Calcium Exercise B.5 on page 305 introduces a study examining the effect of diet cola consumption on calcium levels in women. A sample of 16 health women aged 18 to 40 were randomly assigned to drink 24 ounces of either diet cola or water. Their urine was collected for three hours after ingestion of the beverage and calcium excretion (in mg) was measured. The summary statistics for diet cola are $\bar{x}_C = 56.0$ with $s_C = 4.93$ and $n_C = 8$ and the summary statistics for water are $\bar{x}_W = 49.1$ with $s_W = 3.64$ and $n_W = 8$. Figure 6.26 shows dot plots of the data values. Test whether there is evidence that diet cola leaches calcium out of the system, which would increase the amount of calcium in the urine for diet cola drinkers. In Exercise B.5, we used a randomization distribution to conduct this test. Use a t-distribution here, after first checking that the conditions are met and explaining your reasoning. The data are stored in **ColaCalcium**.

Solution

The hypotheses are $H_0 : \mu_C = \mu_W$ vs $H_a : \mu_C > \mu_W$, where μ_C and μ_W are the mean calcium loss after drinking diet cola and water, respectively. The sample sizes are quite small, so we check for extreme skewness or extreme outliers. We see in the dotplots that the data are not too extremely skewed and don't seem to have any extreme outliers, so a t-distribution is acceptable. The t-test statistic is:

$$t = \frac{\text{Sample statistics} - \text{Null parameter}}{SE} = \frac{(\bar{x}_C - \bar{x}_W) - 0}{\sqrt{\frac{s_C^2}{n_C} + \frac{s_W^2}{n_W}}} = \frac{56.0 - 49.1}{\sqrt{\frac{4.93^2}{8} + \frac{3.64^2}{8}}} = 3.18.$$

This is an upper-tail test, so the p-value is the area above 3.18 in a t-distribution with $df = 7$. We see that the p-value is 0.0078. This is a very small p-value, so we reject H_0 , and conclude that there is strong evidence (even with such a small sample size) that diet cola drinkers do lose more calcium, on average, than water drinkers. Another reason to drink more water and less diet cola!

6.288 Pheromones in Female Tears? On page 11 in Section 1.1, we describe studies to investigate whether there is evidence of pheromones (subconscious chemical signals) in female tears that affect sexual arousal in men. In one of the studies, 50 men had a pad attached to the upper lip that contained either female tears or a salt solution dripped down the same female's face. Each subject participated twice, on consecutive days, once with tears and once with saline, randomized for order, and double-blind. Testosterone levels were measured before sniffing and after sniffing on both days. While normal testosterone levels vary significantly between different men, average levels for the group were the same before sniffing on both days and after sniffing the salt solution (about 155 pb/mL) but were reduced after sniffing the tears (about 133 pg/mL). The mean difference in testosterone levels after sniffing the tears was 21.7 with standard deviation 46.5.

- (a) Why did the investigators choose a matched pairs design for this experiment?
- (b) Test to see if testosterone levels are significantly reduced after sniffing tears.
- (c) Can we conclude that sniffing female tears reduces testosterone levels (which is a significant indicator of sexual arousal in men)?

Solution

a) A matched pairs design is appropriate because testosterone levels vary greatly between different men, and the matched pairs design eliminates this variability as a factor.

(b) We are testing for a difference in two means from a matched pairs experiment. Using the differences, we have

$$H_0 : \mu_d = 0$$

$$H_a : \mu_d > 0$$

where μ_d is the mean of differences, where the differences are computed as (testosterone level after salt solution - testosterone level after sniffing tears). Notice that this is a one-tail test since we are specifically testing to see if testosterone levels are reduced after sniffing female tears. The t-test statistic is:

$$t = \frac{\bar{x}_d - 0}{s_d/\sqrt{n_d}} = \frac{21.7}{46.5/\sqrt{50}} = 3.30.$$

Using the t-distribution with degrees of freedom 49, we find a p-value of 0.0009. This provides strong evidence to reject H_0 and conclude that sniffing female tears significantly reduces testosterone levels in men.

(c) Yes, we can conclude that sniffing female tears reduces testosterone levels since the results are significant and come from a double-blind, placebo-controlled, randomized experiment.

Computer Exercises

For each R problem, turn in answers to questions with the written portion of the homework. Send the R code for the problem to Katherine Goode. The answers to questions in the written part should be well written, clear, and organized. The R code should be commented and well formatted.

R problem 1 Earlier in the semester, we learned to use the bootstrap to find confidence intervals for many parameters. In this problem, we will compare multiple methods for finding a 95% confidence interval between two population proportions. The methods are:

1. Bootstrap using difference in sample proportions +/- 2 times bootstrap SE.
2. Bootstrap using quantiles.
3. The function `prop.test()` in R with the Yates correction.
4. The function `prop.test()` in R without the Yates correction.
5. The formula in the textbook.

Let's reconsider the data in Exercise 6.176, which we can represent in a 2 by 2 matrix.

```
x = matrix(c(937, 1945, 473, 870), nrow = 2, ncol = 2)
rownames(x) = c("1974", "2006")
colnames(x) = c("Favor", "Oppose")
x
##      Favor Oppose
## 1974   937   473
## 2006  1945   870
```

1. Create 10,000 bootstrap samples from each of the 1974 and 2006 populations and compute the difference between sample proportions. Construct 95% confidence intervals by the first two methods above. Graph a histogram or density plot of the bootstrap distribution and comment on

its shape, center, and spread (as measured by standard deviation).

Now that we know about the binomial distribution, we can use this to apply the bootstrap efficiently with the `rbinom()` function. Specifically, the code snippet

```
rbinom(10000,size=1410,prob=937/1410)/1410
```

will generate 10,000 sample proportions from the 1974 population using the bootstrap by taking 10,000 random binomial random variables and dividing each by the sample size.

Solution

We start by creating the 10,000 bootstrap samples. We could do this two ways. The first involves the previous methods we used. Below is the code for my bootstrap.

```
diff <- numeric(10000)
for(i in 1:10000)
{
  p.1 <- sum(sample(c(rep(1,937),rep(0,473)),937+473,replace=TRUE))/(937+473)
  p.2 <- sum(sample(c(rep(1,1945),rep(0,870)),1945+870,replace=TRUE))/(1945+870)
  diff[i] <- p.1-p.2
}
```

With our bootstrap samples, we are able to obtain the standard error of the bootstrap distribution and compute the confidence interval using the first method.

$$\begin{aligned} \hat{p}_1 - \hat{p}_2 \pm 2 * SE \\ -0.02640238 \pm 2 * 0.01531484 \end{aligned}$$

Thus, our confidence interval is

$$(-0.0570, 0.0042).$$

We also use the 2.5 and 97.5 percentiles to obtain a confidence interval.

$$(-0.0566, 0.0034)$$

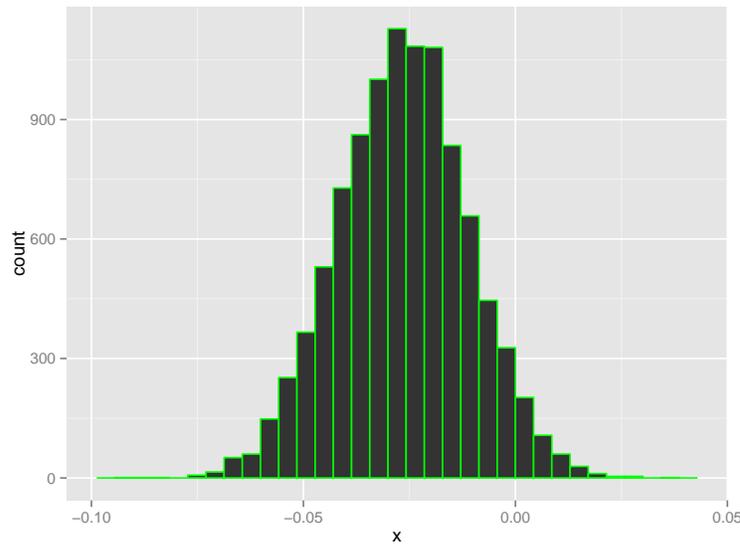
The following R code was used to obtain these intervals.

```
boot.se <- sd(diff)
diff.obs <- (937/(937+473))-(1945/(1945+870))
CI.1 <- diff.obs+c(-2,2)*boot.se
CI.2 <- quantile(diff,c(0.025,0.975))
```

Below is a histogram showing the bootstrap distribution. We use the following code to create this graph.

```
ggplot(data.frame(x = diff), aes(x = x))+geom_histogram(color="green")
```

From looking at the histogram, we see that the distribution is symmetric and centered around -0.025.



The other method we could have used to create the bootstrap distribution is the method the professors suggests at the beginning of this problem, which uses the binomial distribution. Using this method, the two confidence intervals we obtain are

$$(-0.0568, 0.0040) \text{ and } (-0.0562, 0.0030).$$

The code we used to obtain these intervals is as follows.

```
boot <- rbinom(10000,size=1410,prob=937/1410)/1410-rbinom(10000,size=2815,prob=1945/2815)/2815
se <- sd(boot)
CI.3 <- diff.obs+c(-2,2)*se
CI.4 <- quantile(boot,c(0.025,0.975))
```

2. The R function `prop.test()` uses a different method for constructing a confidence interval. When the first argument is a matrix, the function treats each row as a population. Find the confidence intervals from the output using this code.

```
prop.test(x)
prop.test(x, correct = FALSE)
```

Solution

We now use the `prop.test` in R with and without the correction to obtain confidence intervals. We obtain the following output from R.

```
> x = matrix(c(937, 1945, 473, 870), nrow = 2, ncol = 2)
> rownames(x) = c("1974", "2006")
> colnames(x) = c("Favor", "Oppose")
> x
      Favor Oppose
1974   937   473
2006  1945   870
> prop.test(x)
```

2-sample test for equality of proportions with continuity correction

```
data: x
X-squared = 2.8997, df = 1, p-value = 0.0886
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.056913895  0.004109139
sample estimates:
  prop 1    prop 2
0.6645390 0.6909414
```

```
> prop.test(x,correct=FALSE)
```

2-sample test for equality of proportions without continuity correction

```
data: x
X-squared = 3.0202, df = 1, p-value = 0.08223
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.056381666  0.003576909
sample estimates:
  prop 1    prop 2
0.6645390 0.6909414
```

Thus, with the correction applied, we obtain the following confidence interval.

$$(-0.0569, 0.0041)$$

Without the correction applied, we obtain this confidence interval.

$$(-0.0564, 0.0036)$$

3. Finally, compare these confidence intervals with that you found using the textbook formula from earlier in the assignment. Are these confidence intervals all about the same, or are the values from some methods markedly different? For each interval, find the center (mean of the endpoints) and width (difference between the endpoints) and display these values in a table.

Method	Center	Width
Bootstrap $\pm 2 \times SE$		
Bootstrap percentiles		
<code>prop.test()</code> with Yates correction		
<code>prop.test()</code> without Yates correction		
Textbook formula		

Solution

We begin by calculating a confidence interval that uses the formula from the textbook.

$$\begin{aligned} \hat{p}_1 - \hat{p}_2 &\pm z^* \sqrt{\frac{(\hat{p}_1)(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \\ &\Rightarrow (-0.0564, 0.0036) \end{aligned}$$

We are now able to fill in the table.

Method	Center	Width
Bootstrap $\pm 2 \times SE$	-0.0264	0.0613
Bootstrap percentiles	-0.0266	0.0599
<code>prop.test()</code> with Yates correction	-0.0264	0.0610
<code>prop.test()</code> without Yates correction	-0.0264	0.0600
Textbook formula	-0.0264	0.0600

We see that all of the intervals have centers and widths that are relatively similar.

R Problem 2 In this exercise, we will reconsider the data from the light experiment on mice from page 130 that is summarized in Exercise 6.239. The data is in the textbook data frame `LightatNight`. In this data set, the variable `Light` has three categories, but for this exercise, we combine the data from the LL and DM groups for those mice that had some light at night and leave the LD group alone for the group with darkness at night. Here is some R code to create two separate groups for the `BMGain` variable.

```
library(Lock5Data)
data(LightatNight)
light = with(LightatNight, BMGain[(Light == "DM") | (Light == "LL")])
dark = with(LightatNight, BMGain[Light == "LD"])
```

1. Use the bootstrap by sampling with replacement from each variable separately 10,000 times to estimate the bootstrap sampling distribution of the difference between sample means. Construct 95% confidence intervals using the $\pm 2SE$ and the percentile methods.

Solution

We use R to create a bootstrap sample with 10000 resamples. We use this distribution to create two confidence intervals. The confidence interval using the $\pm 2SE$ method is

(1.5381, 5.3126),

and the confidence interval from the percentile method is

(1.5811, 5.3093).

This is the code used in R for this problem.

```
library(Lock5Data)
data(LightatNight)
light = with(LightatNight, BMGain[(Light == "DM") | (Light == "LL")])
dark = with(LightatNight, BMGain[Light == "LD"])

miceboot <- numeric(10000)
for(i in 1:10000)
{
  lightrs <- mean(sample(light,length(light),replace=TRUE))
  darkrs <- mean(sample(dark,length(dark),replace=TRUE))
```

```

  miceboot[i] <- lightrs-darkrs
}

se <- sd(miceboot)
obsmean <- mean(light)-mean(dark)
CI.1 <- obsmean+c(-2,2)*se
CI.2 <- quantile(miceboot,c(0.025,0.975))

```

2. Use the R function `t.test()` to find a 95% confidence interval for the difference in population means

```
t.test(light, dark)
```

Solution

Using the command `t.test(light, dark)` in R, we obtain the following output.

Welch Two Sample t-test

```

data: light and dark
t = 3.4463, df = 21.653, p-value = 0.002341
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.362164 5.488494
sample estimates:
mean of x mean of y
 9.351579  5.926250

```

Thus, the 95% confidence interval for the difference in population means is

(1.3622, 5.4885).

3. Use the textbook formula for the confidence interval.

Solution

We now use the textbook formula to compute the confidence interval. The formula we use is

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

From our data, this interval comes out to be

(1.0751, 5.7756).

The code we use in R is

```

SE <- sqrt((sd(light)^2/length(light))+(sd(dark)^2/length(dark)))
CI.3 <- obsmean+c(-1,1)*qt(0.975,length(dark)-1)*SE

```

4. For each interval, find the center and width and tabulate these. How do they compare?

Solution

Below is a table of the centers and widths for each of the four intervals calculated in the previous parts of the problems. We see that the bootstrap center and widths are much different than the other two methods.

Method	Center	Width
Bootstrap $\pm 2 \times SE$	3.4254	3.7745
Bootstrap percentiles	3.4254	3.7282
<code>t.test()</code>	3.4253	4.1263
Textbook formula	3.4253	4.7005

5. Now, use a randomization distribution of size 10,000 to test the hypothesis $H_0 : \mu_{Light} = \mu_{Dark}$ versus the two-sided alternative. Find the test statistic and p-value.

Solution

Using R, we create a randomization distribution from which we obtain a p-value of 0.0081 by considering the number of values that are as extreme or more extreme than our observed value of 3.425329. Thus, we are able to reject the null hypothesis that $H_0 : \mu_{Light} = \mu_{Dark}$ and conclude that there is a difference in the means.

Below is the R code that was used to create the randomization distribution and obtain the p-value.

```
allvalues <- c(light,dark)
observed <- mean(light)-mean(dark)
N <- 9999
result <- numeric(N)
for (i in 1:N)
{
  index <- sample(length(light)+length(dark), size=length(light), replace=FALSE)
  result[i] <- mean(allvalues[index])-mean(allvalues[-index])
}

(sum(result<=-observed)+sum(result>=observed)+1)/(N+1)
observed
```

6. Find the p-value from `t.test()`. What degrees of freedom are used for the t distribution?

Solution

Referring back to the output from our t-test in R, we see that the p-value is 0.0023, which implies that the null hypothesis should be rejected. This agrees with the result we found in the previous part. Note that 21.653 degrees of freedom were used to do this t test.

7. Use the method from the textbook to find the t test statistic and p-value. What degrees of freedom are used?

Solution

In order to use the method from the textbook, we need to use the following formula for the test statistics.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Based on the data from this situation, we find that

$$t = 3.446.$$

We determine the p-value by considering a t-distribution with 8-1 degrees of freedom. This value is

$$\text{p-value} = 0.01074.$$

In this case, we would also reject the null hypothesis.

8. Compare the p-values from these three methods. Would conclusions be about the same or quite different depending on the method used?

Solution

We find that the p-value using the formula in the book is higher than the other two, but we would still be able to reject the null hypothesis in all cases.