

Textbook Exercises

6.160, 6.164, 6.176, 6.178, 6.185, 6.195, 6.202, 6.223, 6.238, 6.239, 6.259, 6.288

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.

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)
```

- 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		

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"])
```

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

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

- Use the textbook formula for the confidence interval.
- For each interval, find the center and width and tabulate these. How do they compare?

Method	Center	Width
Bootstrap $\pm 2 \times SE$		
Bootstrap percentiles		
<code>t.test()</code>		
Textbook formula		

- Now, use a randomization distribution of size 10,000 to test the hypothesis $H_0: \mu_{\text{Light}} = \mu_{\text{Dark}}$ versus the two-sided alternative. Find the test statistic and p-value.
- Find the p-value from `t.test()`. What degrees of freedom are used for the *t* distribution?
- Use the method from the textbook to find the *t* test statistic and p-value. What degrees of freedom are used?
- Compare the p-values from these three methods. Would conclusions be about the same or quite different depending on the method used?