Chapter 8: Statistics with R - 2nd Edition

Robert Stinerock

Student Exercises

The csv data sets used in these exercises can be found on the website:

1. `cafe_ratings.csv`
2. `dining.csv`
3. `insurance.csv`
4. `benefits.csv`

1. Suppose a random sample of size $n = 100$ has been selected and the sample mean is found to be $\overline{x} = 67$. The population standard deviation is assumed to be $\sigma = 12$. Please answer the following questions.

   (a) What is the standard error of the mean $\sigma_{\overline{x}}$?

   Answer: 1.2

   $$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{100}} = 1.2$$

   (b) What is the margin of error if the confidence level is $(1 - \alpha) = 0.95$?

   Answer: 2.352

   If $(1 - \alpha) = 0.95$, then $z_{\alpha/2} = z_{0.025} = 1.96$

```r
qnorm(0.025, lower.tail = FALSE) * 12 / sqrt(100)

## [1] 2.351957
```

   Therefore, $z_{\alpha/2}\sigma_{\overline{x}} = (1.96)(1.2) = 2.352$

   (c) What is the 95% confidence interval estimate of μ?

   Answer: 67 ± 2.352 or $[64.648, 69.352]$

   $$\overline{x} \pm z_{\alpha/2}\sigma_{\overline{x}}$$

$$67 \pm 2.352$$

$$[64.648, 69.352]$$

```r
# To find the upper bound of the confidence interval
# estimate, add the margin of error to the sample mean.

67 + qnorm(0.025, lower.tail = FALSE) * 12 / sqrt(100)

## [1] 69.35196

# To find the lower bound of the confidence interval
# estimate, subtract the margin of error from the sample mean.

67 - qnorm(0.025, lower.tail = FALSE) * 12 / sqrt(100)

## [1] 64.64804
```

2. Suppose a random sample of size $n = 36$ has been selected from a population with $\sigma = 30$ and a sample mean of $\overline{x} = 205$

   (a) What is the 90% confidence interval estimate of μ?

   Answer: 205 ± 8.225

   If $(1 - \alpha) = 0.90$, then $z_{\alpha/2} = z_{0.050} = 1.645$, then

```r
qnorm(0.05, lower.tail = FALSE)

## [1] 1.644854
```

$$\overline{x} \pm z_{\alpha/2}\sigma_{\overline{x}}$$

$$205 \pm (1.645)\frac{30}{\sqrt{36}}$$

$$205 \pm 8.225$$

$$[196.775, 213.225]$$

```r
# To find the upper bound of the confidence interval
# estimate, add the margin of error to the sample mean.

205 + qnorm(0.05, lower.tail = FALSE) * 30 / sqrt(36)
```

```r
## [1] 213.2243
# To find the lower bound of the confidence interval
#estimate, subtract the margin of error from the sample mean.

205 - qnorm(0.05, lower.tail = FALSE) * 30 / sqrt(36)

## [1] 196.7757
```

(b) What is the 95% confidence interval estimate of μ?

Answer: 205 ± 9.80

If $(1-\alpha) = 0.95$, then $z_{\alpha/2} = z_{0.025} = 1.96$, then

```r
qnorm(0.025, lower.tail = FALSE)

## [1] 1.959964
```

$$\overline{x} \pm z_{\alpha/2}\sigma_{\overline{x}}$$

$$205 \pm (1.96)\frac{30}{\sqrt{36}}$$

$$205 \pm 9.80$$

$$[195.20, 214.80]$$

```r
# To find the upper bound of the confidence interval
# estimate, add the margin of error to the sample mean.

205 + qnorm(0.025, lower.tail = FALSE) * 30 / sqrt(36)

## [1] 214.7998

# To find the lower bound of the confidence interval
# estimate, subtract the margin of error from the sample mean.

205 - qnorm(0.025, lower.tail = FALSE) * 30 / sqrt(36)

## [1] 195.2002
```

(c) What is the 99% confidence interval estimate of μ?

Answer: 205 ± 12.880

If $(1-\alpha) = 0.99$, then $z_{\alpha/2} = z_{0.005} = 2.576$, then

```r
qnorm(0.005, lower.tail = FALSE)
```

```
## [1] 2.575829
```

$$\overline{x} \pm z_{\alpha/2}\sigma_{\overline{x}}$$

$$205 \pm (2.576)\frac{30}{\sqrt{36}}$$

$$205 \pm 12.880$$

$$[192.12, 217.88]$$

```r
# To find the upper bound of the confidence interval
# estimate, add the margin of error to the sample mean.

205 + qnorm(0.005, lower.tail = FALSE) * 30 / sqrt(36)
```

```
## [1] 217.8791
```

```r
# To find the lower bound of the confidence interval
# estimate, subtract the margin of error to the sample mean.

205 - qnorm(0.005, lower.tail = FALSE) * 30 / sqrt(36)
```

```
## [1] 192.1209
```

3. If a 99% confidence interval is [228, 232] for a population with $\sigma = 10$, what is n?

   Answer: $n = 166$

   Margin of error=$(232 - 228)/2 = 2$

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$2 = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$n = (\frac{z_{\alpha/2}\sigma}{2})^2 = (\frac{(2.576)(10)}{2})^2 \approx 166$$

```r
qnorm(0.005, lower.tail = FALSE)
```

```
## [1] 2.575829
```

Checking the margin of error when $n = 166$, $(1 - \alpha) = 0.99$, and $\sigma = 10$

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = (2.576)(\frac{10}{\sqrt{166}}) = 2$$

4. Provide the interval estimate at the 90% confidence level for each case below.

   (a) For $n = 40$, the sample mean is $\overline{x} = 70$; assume it is known that $\sigma = 6$.

   Answer: 70 ± 1.56

$$\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

```r
qnorm(0.05, lower.tail = FALSE)

## [1] 1.644854
```

$$70 \pm (1.645)\frac{6}{\sqrt{40}}$$

$$70 \pm 1.56$$

$$[68.44, 71.56]$$

```r
# To find the upper bound of the confidence interval
# estimate, add the margin of error to the sample mean.

70 + qnorm(0.05, lower.tail = FALSE) * 6 / sqrt(40)

## [1] 71.56045

# To find the lower bound of the confidence interval
# estimate, subtract the margin of error from the sample mean.

70 - qnorm(0.05, lower.tail = FALSE) * 6 / sqrt(40)

## [1] 68.43955
```

   (b) For $n = 120$, the sample mean is $\overline{x} = 44$; assume it is known that $\sigma = 2$.

   Answer: 44 ± 0.30

$$\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

```r
qnorm(0.05, lower.tail = FALSE)

## [1] 1.644854
```

$$44 \pm (1.645)\frac{2}{\sqrt{120}}$$

$$44 \pm 0.30$$

$$[43.70, 44.30]$$

```r
# To find the upper bound of the confidence interval
# estimate, add the margin of error to the sample mean.

44 + qnorm(0.05, lower.tail = FALSE) * 2 / sqrt(120)

## [1] 44.30031

# To find the lower bound of the confidence interval
# estimate, subtract the margin of error from the sample mean.

44 - qnorm(0.05, lower.tail = FALSE) * 2 / sqrt(120)

## [1] 43.69969
```

(c) For $n = 81$, the sample mean is $\overline{x} = 0$; assume it is known that $\sigma = 9$.

Answer: 0 ± 1.645

$$\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$0 \pm (1.645)\frac{9}{\sqrt{81}}$$

$$0 \pm 1.645$$

$$[-1.645, 1.645]$$

```r
# To find the upper bound of the confidence interval
# estimate, add the margin of error to the sample mean.

0 + qnorm(0.05, lower.tail = FALSE) * 9 / sqrt(81)

## [1] 1.644854
```

```r
# To find the lower bound of the confidence interval
# estimate, subtract the margin of error from the sample mean.

0 - qnorm(0.05, lower.tail = FALSE) * 9 / sqrt(81)

## [1] -1.644854
```

(d) For $n = 30$, the sample mean is $\overline{x} = -12$; assume it is known that $\sigma = 7$.

Answer: -12 ± 2.10

$$\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$-12 \pm (1.645)\frac{7}{\sqrt{30}}$$

$$-12 \pm 2.10$$

$$[-14.10, -9.90]$$

```r
# To find the upper bound of the confidence interval
# estimate, add the margin of error to the sample mean.

-12 + qnorm(0.05, lower.tail = FALSE) * 7 / sqrt(30)

## [1] -9.897845

# To find the lower bound of the confidence interval
# estimate, subtract the margin of error from the sample mean.

-12 - qnorm(0.05, lower.tail = FALSE) * 7 / sqrt(30)

## [1] -14.10215
```

5. In a continuing study of the amount MBA students are spending each term on textbooks, data were collected on $n = 81$ students. In previous studies, the population standard deviation has been $\sigma = \$24$.

(a) What is the margin of error at the 99% confidence level?

Answer: 6.87

```r
qnorm(0.005, lower.tail = FALSE) * 24 / sqrt(81)

## [1] 6.868878
```

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = (2.576)\frac{24}{\sqrt{81}} = 6.87$$

(b) If the mean from the most recent sample was $\overline{x} = \$288$, what is the 99% confidence interval estimate of the population mean μ?

Answer: $\$288 \pm \$6.87$

$$\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$288 \pm (2.576)\frac{24}{\sqrt{81}}$$

$$288 \pm 6.87$$

$$[281, 295]$$

```
# To find the upper bound of the confidence interval
# estimate, add the margin of error to the sample mean.

288 + qnorm(0.005, lower.tail = FALSE) * 24 / sqrt(81)

## [1] 294.8689

# To find the lower bound of the confidence interval
# estimate, subtract the margin of error from the sample mean.

288 - qnorm(0.005, lower.tail = FALSE) * 24 / sqrt(81)

## [1] 281.1311
```

6. A study based on a random sample of $n = 30$ North American educational institutions indicates that the average financial endowment is $\$313,182,000$. If the population standard deviation is known to be $\sigma = \$64,500,000$, what is the 80% confidence interval estimate of the population mean μ?

Answer: $\$313,182,000 \pm \$15,091,596$

$$(1 - \alpha) = 0.80$$

$$z_{\alpha/2} = z_{0.10} = 1.281552$$

```r
qnorm(0.10, lower.tail = FALSE)

## [1] 1.281552
```

```r
# The margin of error.

qnorm(0.10, lower.tail = FALSE) * 64500000 / sqrt(30)

## [1] 15091596
```

$$\overline{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$313182000 \pm (1.281552) \frac{64500000}{\sqrt{30}}$$

$$313182000 \pm 15091596$$

$$[298090404, 328273596]$$

```r
# To find the upper bound of the confidence interval
# estimate, add the margin of error to the sample mean.

313182000 + qnorm(0.10, lower.tail = FALSE) * 64500000 / sqrt(30)

## [1] 328273596

# To find the lower bound of the confidence interval
# estimate, subtract the margin of error from the sample mean.

313182000 - qnorm(0.10, lower.tail = FALSE) * 64500000 / sqrt(30)

## [1] 298090404
```

7. Referring to the previous exercise, if the population (of all North American universities) from which the random sample has been drawn is highly skewed because of the presence of a number of financially well-endowed universities (also known as outliers), does the above procedure lead to reliable estimates? If not, comment on what might be done to improve the estimation procedure.

   Answer: If the sample size is $n = 30$, the sampling distribution of $\overline{x}$ approaches normal. However, when the data are highly skewed, it is a good practice to increase the size of the sample if possible. In this, we make it more likely that the sampling distribution of $\overline{x}$ approaches normal.

8. Find the following probabilities using the t distribution with $n = 23$.

(a) What is the probability that t will be greater than or equal to 0? That is, what is $p(t \geq 0, df = 23 - 1 = 22)$?

Answer: 0.5000

```
# Use 1 minus the pt() function.

1 - pt(0, 22)

## [1] 0.5

# Or using the "lower.tail=FALSE" argument.

pt(0, 22, lower.tail =  FALSE)

## [1] 0.5
```

(b) What is the probability that t will be greater than or equal to 2.074? That is, what is $p(t \geq 2.074, df = 23 - 1 = 22)$?

Answer: 0.02499

```
# Use 1 minus the pt() function.

1 - pt(2.074, 22)

## [1] 0.02499358

# Or using the "lower.tail=FALSE" argument.

pt(2.074, 22, lower.tail =  FALSE)

## [1] 0.02499358
```

(c) What is the probability that t will be less than or equal to 2.819? That is, what is $p(t \leq 2.819, df = 23 - 1 = 22)$?

Answer: 0.995

```
# Use the pt() function.

pt(2.819, 22)

## [1] 0.9950028
```

(d) What is the probability that t will be less than or equal to -1.321? That is, what is $p(t \leq -1.321, df = 23 - 1 = 22)$?

Answer: 0.10

```r
# Use the pt() function.

pt(-1.321, 22)

## [1] 0.1000388
```

(e) What is the probability t will fall between -0.858 and 1.717? That is, what is the $p(-0.858 \leq t \leq 1.717)$?

Answer: 0.7499

```r
# Subtract pt(-0.858,22) from pt(1.717,22).

pt(1.717, 22) - pt(-0.858, 22)

## [1] 0.7499147
```

(f) What is the probability t will fall between -1.717 and -0.858? That is, what is the $p(-1.717 \leq t \leq -0.858)$?

Answer: 0.1501

```r
# Subtract pt(-1.717,22) from pt(-0.858,22).

pt(-0.858, 22) - pt(-1.717, 22)

## [1] 0.1500585
```

9. Find the value of t for each of the following questions.

(a) What is the value of t which has an area to its right of 0.05 when the sample size is $n = 28$.

Answer: 1.703

```r
# Use the qt() function.

qt(0.95, 27)

## [1] 1.703288

# Include the "lower.tail=FALSE" argument.

qt(0.05, 27, lower.tail =  FALSE)

## [1] 1.703288
```

(b) What is the value of t which has an area to its right of 0.025 when the sample size is $n = 41$.

Answer: 2.021

```r
# Use the qt() function.

qt(0.975, 40)

## [1] 2.021075

# Include the "lower.tail=FALSE" argument.

qt(0.025, 40, lower.tail = FALSE )

## [1] 2.021075
```

(c) What is the value of t which has an area to its right of 0.500 when the sample size is $n = 11$.

Answer: 0.000

```r
# Use the qt() function.

qt(0.5000, 10)

## [1] 0
```

(d) What is the value of t which has an area to its left of 0.01 when the sample size is $n = 76$.

Answer: -2.377

```r
# Use the qt() function.

qt(0.01, 75)

## [1] -2.377102
```

(e) What is the value of t which has an area to its left of 0.10 when the sample size is $n = 33$.

Answer: -1.309

```r
# Use the qt() function.

qt(0.10, 32)

## [1] -1.308573
```

(f) What is the value of t which has an area to its left of 0.20 when the sample size is $n = 100$.

Answer: -0.8453

```r
# Use the qt() function.

qt(0.20, 99)

## [1] -0.845267
```

10. A sample of gasoline prices was taken at $n = 70$ service stations across the U.S., and the sample mean price per gallon was found to be $\overline{x} = \$3.67$, the sample standard deviation $s = \$0.21$.

    (a) What is the 90% confidence interval estimate of the population mean price per gallon μ?

    Answer: 3.67 ± 0.0418

$$\overline{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

```r
# The value of t providing area of 0.05 in upper tail.

qt(0.05, 69, lower.tail = FALSE)

## [1] 1.667239

# The margin of error.

qt(0.05, 69, lower.tail = FALSE) * 0.21 / sqrt(70)

## [1] 0.04184736
```

$$3.67 \pm (1.667)(0.0251)$$

$$3.67 \pm 0.0418$$

$$[3.63, 3.71]$$

```r
# To find the upper bound of the confidence interval
# estimate, add the margin of error to the sample mean.

3.67 + qt(0.05, 69, lower.tail = FALSE) * 0.21 / sqrt(70)

## [1] 3.711847

# To find the lower bound of the confidence interval
# estimate, subtract the margin of error from the sample mean.

3.67 - qt(0.05, 69, lower.tail = FALSE) * 0.21 / sqrt(70)

## [1] 3.628153
```

(b) What is the 95% confidence interval estimate of the population mean price per gallon μ?

Answer: 3.67 ± 0.0501

$$\overline{x} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$$

```
# The value of t providing area of 0.025 in upper tail.

qt(0.025, 69, lower.tail = FALSE)

## [1] 1.994945

# The margin of error.

qt(0.025, 69, lower.tail = FALSE) * 0.21 / sqrt(70)

## [1] 0.05007273
```

$$3.67 \pm (1.995)(0.0251)$$

$$3.67 \pm 0.0501$$

$$[3.62, 3.72]$$

```
# To find the upper bound of the confidence interval
# estimate, add the margin of error to the sample mean.

3.67 + qt(0.025, 69, lower.tail = FALSE) * 0.21 / sqrt(70)

## [1] 3.720073

# To find the lower bound of the confidence interval
# estimate, subtract the margin of error from the sample mean.

3.67 - qt(0.025, 69, lower.tail = FALSE) * 0.21 / sqrt(70)

## [1] 3.619927
```

(c) What is the 99% confidence interval estimate of the population mean price per gallon μ?

Answer: 3.67 ± 0.0665

$$\overline{x} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$$

```r
# The value of t providing area of 0.005 in upper tail.

qt(0.005, 69, lower.tail = FALSE)

## [1] 2.648977

# The margin of error.

qt(0.005, 69, lower.tail = FALSE) * 0.21 / sqrt(70)

## [1] 0.06648879
```

$$3.67 \pm (2.649)(0.0251)$$

$$3.67 \pm 0.0665$$

$$[3.60, 3.74]$$

```r
# To find the upper bound of the confidence interval
# estimate, add the margin of error to the sample mean.

3.67 + qt(0.005, 69, lower.tail = FALSE) * 0.21 / sqrt(70)

## [1] 3.736489

# To find the lower bound of the confidence interval
# estimate, subtract the margin of error from the sample mean.

3.67 - qt(0.005, 69, lower.tail = FALSE) * 0.21 / sqrt(70)

## [1] 3.603511
```

11. According to a survey, the average British household is expected to spend £868 on holiday-related expenses during the approaching Christmas period. This amount covers not only gifts, but food, beverages, and decorations. Assume the study is based on $n = 94$ randomly sampled households throughout Great Britain; assume also that the sample standard deviation is s=£162.

    (a) What is the 60% confidence interval estimate of the population mean amount-to-be-spent μ during the next Christmas holiday period?

    Answer: $\pounds 868 \pm \pounds 14.13$

$$\overline{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

```r
# The value of t providing area of 0.20 in upper tail.

qt(0.20, 93, lower.tail = FALSE)

## [1] 0.8455033

# The margin of error.

qt(0.20, 93, lower.tail = FALSE) * 162 / sqrt(94)

## [1] 14.12753
```

$$868 \pm 14.13$$

$$[854, 882]$$

```r
# To find the upper bound of the confidence interval
# estimate, add the margin of error to the sample mean.

868 + qt(0.20, 93, lower.tail = FALSE) * 162 / sqrt(94)

## [1] 882.1275

# To find the lower bound of the confidence interval
# estimate, subtract the margin of error from the sample mean.

868 - qt(0.20, 93, lower.tail = FALSE) * 162 / sqrt(94)

## [1] 853.8725
```

(b) What is the 80% confidence interval estimate of the population mean amount-to-be-spent μ during the same period?

Answer: £868 ± £21.57

$$\overline{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

```r
# The value of t providing area of 0.10 in upper tail.

qt(0.10, 93, lower.tail = FALSE)

## [1] 1.290721

# The margin of error.

qt(0.10, 93, lower.tail = FALSE) * 162 / sqrt(94)

## [1] 21.56669
```

$$868 \pm (1.291)(16.71)$$

$$868 \pm 21.57$$

$$[846, 890]$$

```
# To find the upper bound of the confidence interval
# estimate, add the margin of error to the sample mean.

868 + qt(0.10, 93, lower.tail = FALSE) * 162 / sqrt(94)

## [1] 889.5667

# To find the lower bound of the confidence interval
# estimate, subtract the margin of error from the sample mean.

868 - qt(0.10, 93, lower.tail = FALSE) * 162 / sqrt(94)

## [1] 846.4333
```

(c) What is the 90% confidence interval estimate of the population mean amount-to-be-spent μ during the same period?

Answer: £868 ± £27.76

$$\overline{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

```
# The value of t providing area of 0.05 in upper tail.

qt(0.05, 93, lower.tail = FALSE)

## [1] 1.661404

# The margin of error.

qt(0.05, 93, lower.tail = FALSE) * 162 / sqrt(94)

## [1] 27.76043
```

$$868 \pm (1.661)(16.71)$$

$$868 \pm 27.76$$

$$[840, 896]$$

```r
# To find the upper bound of the confidence interval
# estimate, add the margin of error to the sample mean.

868 + qt(0.05, 93, lower.tail = FALSE) * 162 / sqrt(94)

## [1] 895.7604

# To find the lower bound of the confidence interval
# estimate, subtract the margin of error from the sample mean.

868 - qt(0.05, 93, lower.tail = FALSE) * 162 / sqrt(94)

## [1] 840.2396
```

12. The `cafe_ratings.csv` data consist of a sample of $n = 50$ New York restaurants. The variables include `cuisine` (type of cuisine: American, Chinese, French, Italian, and Japanese), `rating` (on thirty-point scale), and `price` (average price of a meal). Import `cafe_ratings.csv` into the R Workspace and name it `E8_1`.

```r
cafe_ratings <- read.csv('cafe_ratings.csv')
```

```r
E8_1 <- cafe_ratings
```

   (a) What are the point estimates of the mean and standard deviation of price μ?

   Answer: the sample mean is \$93.46, the sample standard deviation is \$75.65.

```r
# Use the names() function to identify the variable names.

names(E8_1)

## [1] "cuisine" "price"   "rating"

# Use mean() to find mean of price.

mean(E8_1$price)

## [1] 93.46

# Use sd() to find the standard deviation of price.

sd(E8_1$price)

## [1] 75.64935
```

(b) What is the margin of error at the 95% level of confidence?

Answer: $21.50

```
qt(0.025, 49, lower.tail = FALSE) * sd(E8_1$price) /
  sqrt(length(E8_1$price))

## [1] 21.49931
```

(c) Find the 95% confidence interval estimate of the population mean μ.

Answer: The 95% confidence interval estimate is [$71.96, $114.96].

```
# For the upper bound of the interval estimate, add
# the margin of error to the sample mean of E8_1.

mean(E8_1$price) + qt(0.025, 49, lower.tail = FALSE) *
  sd(E8_1$price) / sqrt(length(E8_1$price))

## [1] 114.9593

# For the lower bound of the interval estimate,
# subtract the margin of error from the sample mean of E8_1.

mean(E8_1$price) - qt(0.025, 49, lower.tail = FALSE) *
  sd(E8_1$price) / sqrt(length(E8_1$price))

## [1] 71.96069
```

13. Referring to the preceding exercise, find the 95% confidence interval estimate using the `t.test()` function. Does the answer using this approach square with the answer in the preceding exercise?

Answer: Yes, it does. The 95% confidence interval estimate is [$71.96, $114.96].

```
# Use the t.test() function.

t.test(E8_1$price, conf = 0.95)

##
##  One Sample t-test
##
## data:  E8_1$price
## t = 8.7359, df = 49, p-value = 1.465e-11
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   71.96069 114.95931
## sample estimates:
## mean of x
##     93.46
```

14. Find the 90% confidence interval estimate of `price` from the `cafe_ratings.csv` data. Use both methods.

    Answer: Both methods produce the same 90% confidence interval: [$75.52, $111.40].

```
# For the upper bound of the interval estimate, add
# the margin of error to the sample mean of E8_1.

mean(E8_1$price) + qt(0.05, 49, lower.tail = FALSE) *
  sd(E8_1$price) / sqrt(length(E8_1$price))

## [1] 111.3965

# For the lower bound of the interval estimate, subtract
# the margin of error from the sample mean of E8_1.

mean(E8_1$price) - qt(0.05, 49, lower.tail = FALSE) *
  sd(E8_1$price) / sqrt(length(E8_1$price))

## [1] 75.52353

# Use the t.test() function for the 90% interval estimate.

t.test(E8_1$price, conf = 0.90)

##
##  One Sample t-test
##
## data:  E8_1$price
## t = 8.7359, df = 49, p-value = 1.465e-11
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##   75.52353 111.39647
## sample estimates:
## mean of x
##     93.46
```

15. Find the 99% confidence interval estimate of `price` from the `cafe_ratings.csv` data. Use both methods.

    Answer: Both methods produce the same 99% confidence interval: [$64.79, $122.13].

```
# For the upper bound of the interval estimate, add
# the margin of error to the sample mean of E8_1.
```

```r
mean(E8_1$price) + qt(0.005, 49, lower.tail = FALSE) *
  sd(E8_1$price) / sqrt(length(E8_1$price))

## [1] 122.1313

# For the lower bound of the interval estimate,
# subtract the margin of error from the sample mean of E8_1.

mean(E8_1$price) - qt(0.005, 49, lower.tail = FALSE) *
  sd(E8_1$price) / sqrt(length(E8_1$price))

## [1] 64.78871

# Use the t.test() function for the 99% interval estimate.

t.test(E8_1$price, conf = 0.99)

##
##  One Sample t-test
##
## data:  E8_1$price
## t = 8.7359, df = 49, p-value = 1.465e-11
## alternative hypothesis: true mean is not equal to 0
## 99 percent confidence interval:
##   64.78871 122.13129
## sample estimates:
## mean of x
##     93.46
```

16. A university's director of research wanted to get an idea of the amount being spent on dining by faculty attending academic conferences over the course of a year. Accordingly, she directed her staff to randomly sample the annual expense reimbursement reports from 100 faculty members during 2021. Import `dining.csv` data (from the website) into the R Workspace and name it `E8_2`. Use the `t.test()` function to find the 90% confidence interval estimate of μ; the variable name is `annual.cost`.

    Answer: The 90% confidence interval estimate of μ is [\$891.25, \$947.45].

```r
dining <- read.csv('dining.csv')
```

```r
E8_2 <- dining
```

```r
# Use the names() function to identify the variable name.

names(E8_2)

## [1] "annual.cost"

# Use the t.test() function.

t.test(E8_2$annual.cost, conf = 0.90)

##
##  One Sample t-test
##
## data:  E8_2$annual.cost
## t = 54.326, df = 99, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  891.2516 947.4484
## sample estimates:
## mean of x
##    919.35
```

17. The manager of an insurance office wishes to gain a better understanding of the dollar value of the newly-purchased automobiles his firm has insured over the previous twelve months. To this end, he randomly selects $n = 70$ insurance applications from the previous year which specified the total cost of each vehicle insured, including market value, taxes, and licensing fees. Import `insurance.csv` (see the website) into the R Workspace and name it `E8_3`; the variable name is `automobile`.

    (a) Find the 90% confidence interval estimate of the mean dollar value of newly purchased automobiles μ.

    Answer: The 90% confidence interval estimate of μ is [\$17027, \$20221].

    ```r
    insurance <- read.csv('insurance.csv')
    ```

    ```r
    E8_3 <- insurance
    ```

    ```r
    # Use the names() function to identify the variable name.

    names(E8_3)

    ## [1] "automobile"

    # Use t.test() function.

    t.test(E8_3$automobile, conf = 0.90)
    ```

```
##
##  One Sample t-test
##
## data:  E8_3$automobile
## t = 19.443, df = 69, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  17027.03 20220.97
## sample estimates:
## mean of x
##     18624
```

(b) Find the 95% confidence interval estimate of the mean dollar value of newly purchased automobiles μ.

Answer: The 95% confidence interval estimate of μ is [\$16713, \$20535].

```
# Use t.test() function.

t.test(E8_3$automobile, conf = 0.95)

##
##  One Sample t-test
##
## data:  E8_3$automobile
## t = 19.443, df = 69, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  16713.13 20534.87
## sample estimates:
## mean of x
##     18624
```

(c) Find the 99% confidence interval estimate of the mean dollar value of newly purchased automobiles μ.

Answer: The 99% confidence interval estimate of μ is [\$16087, \$21161].

```
# Use t.test() function.

t.test(E8_3$automobile, conf = 0.99)

##
##  One Sample t-test
##
## data:  E8_3$automobile
## t = 19.443, df = 69, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 99 percent confidence interval:
```

```
##  16086.66 21161.34
## sample estimates:
## mean of x
##     18624
```

18. What sample size n is required to provide a margin of error of 5 at the 95% confidence level. Assume the population standard deviation is $\sigma = 25$.

    Answer: $n = 97$

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{(PME)^2}$$

$$n = \frac{(1.96)^2 (25)^2}{(5)^2} = 96.04 \approx 97$$

```r
qnorm(0.025, lower.tail = FALSE) # For 95% confidence.

## [1] 1.959964
```

```r
(qnorm(0.025, lower.tail =  FALSE) * 25 / 5) ^ 2 # Sample size.

## [1] 96.03647
```

check:

$$z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = (1.96) \frac{25}{\sqrt{97}} = (1.96)(2.55) \approx 5$$

```r
qnorm(0.025, lower.tail = FALSE) * 25 / sqrt(97)

## [1] 4.975105
```

19. What sample size n is required to provide a margin of error of 5 at the 99% confidence level. Assume the population standard deviation is $\sigma = 25$.

    Answer: $n = 166$

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{(PME)^2}$$

$$n = \frac{(2.576)^2 (25)^2}{(5)^2} = 165.77 \approx 166$$

```r
qnorm(0.005, lower.tail = FALSE) # For 99% confidence.

## [1] 2.575829
```

```r
(qnorm(0.005, lower.tail =  FALSE) * 25 / 5) ^ 2 # Sample size.

## [1] 165.8724
```

check:

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = (2.576)\frac{25}{\sqrt{166}} = (2.575)(1.94) \approx 5$$

```r
qnorm(0.005, lower.tail = FALSE) * 25 / sqrt(166)

## [1] 4.998078
```

20. The director of a large medical college wishes to estimate the mean student age μ of the most recent entering class of aspiring physicians pursuing an M.D. degree. A quick pilot study reveals that 3 years might be used as a planning value estimate of σ. If a 95% confidence interval estimate with a margin of error of 1 is desired, what sample size should we recommend?

    Answer: $n = 35$

$$n = \frac{(z_{\alpha/2})^2\sigma^2}{(PME)^2}$$

$$n = \frac{(1.96)^2(3)^2}{(1)^2} = 34.57 \approx 35$$

```r
qnorm(0.025, lower.tail = FALSE) # For 95% confidence.

## [1] 1.959964
```

```r
(qnorm(0.025, lower.tail =  FALSE) * 3 / 1) ^ 2 # Sample size.

## [1] 34.57313
```

check:

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = (1.96)\frac{3}{\sqrt{35}} = (1.96)(0.51) \approx 1$$

```r
qnorm(0.025, lower.tail = FALSE) * 3 / sqrt(35)

## [1] 0.9938831
```

21. The commute-to-work time for the residents of the world's large cities has been investigated extensively. A pilot study involving an SRS of residents of Toronto is used to provide a planning value estimate of 10 minutes for the population standard deviation σ.

    (a) If we want to estimate the population mean commute-to-work time μ for the residents of Toronto with a margin of error of 2 minutes, what sample size n should we recommend? Assume 90% confidence.

    Answer: $n = 68$

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{(PME)^2}$$

$$n = \frac{(1.645)^2 (10)^2}{(2)^2} = 67.65 \approx 68$$

```r
qnorm(0.05, lower.tail = FALSE) # For 90% confidence.

## [1] 1.644854
```

```r
(qnorm(0.05, lower.tail =  FALSE) * 10 / 2) ^ 2 # Sample size.

## [1] 67.63859
```

check:

$$z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = (1.645) \frac{10}{\sqrt{68}} = (1.645)(1.213) \approx 2$$

```r
qnorm(0.05, lower.tail = FALSE) * 10 / sqrt(68)

## [1] 1.994678
```

    (b) If we want to estimate the population mean commute-to-work time μ for the residents of Toronto with a margin of error of 1 minute, what sample size n should we recommend? Assume 90% confidence.

    Answer: $n = 271$

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{(PME)^2}$$

$$n = \frac{(1.645)^2 (10)^2}{(1)^2} = 270.6 \approx 271$$

```r
(qnorm(0.05, lower.tail =  FALSE) * 10 / 1) ^ 2 # Sample size.

## [1] 270.5543
```

check:

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = (1.645)\frac{10}{\sqrt{271}} = (1.645)(0.607) \approx 1$$

```r
qnorm(0.05, lower.tail = FALSE) * 10 / sqrt(271)

## [1] 0.9991774
```

22. A sample of 64 MBA students at the London Business School (LBS) reported they spent an average of £252.45 on textbooks and case materials per semester. Assume the population standard deviation is σ=£74.50.

    (a) What is the 95% confidence interval estimate for the mean amount spent μ per term by an MBA student at LBS?

    Answer: £252.45 ± £18.25 or [£234.20, £270.70]

$$\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$252.45 \pm 1.96\frac{74.50}{\sqrt{64}}$$

$$252.45 \pm 18.25$$

$$[234.20, 270.70]$$

```r
# For the lower bound of the interval estimate,
# subtract the margin of error from the sample mean.

252.45 - qnorm(0.025, lower.tail = FALSE) * 74.50 / sqrt(64)

## [1] 234.1978

# For the upper bound of the interval estimate, add
# the margin of error to the sample mean.

252.45 + qnorm(0.025, lower.tail = FALSE) * 74.50 / sqrt(64)

## [1] 270.7022
```

    (b) What sample size n would be needed to estimate μ with a margin of error of £10 at 90% confidence?

    Answer: $n = 151$

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{(PME)^2}$$

$$n = \frac{(1.645)^2 (74.50)^2}{(10)^2} = 150.19 \approx 151$$

```
qnorm(0.05, lower.tail = FALSE) # For 90% confidence.

## [1] 1.644854
```

```
(qnorm(0.05, lower.tail =  FALSE) * 74.50 / 10) ^ 2 # Sample size

## [1] 150.1644
```

check:

$$z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = (1.645) \frac{74.50}{\sqrt{151}} = (1.645)(6.06) \approx 10$$

```
qnorm(0.05, lower.tail = FALSE) * 74.50 / sqrt(151)

## [1] 9.972294
```

23. In a pilot study, $n = 80$ respondents have been interviewed, 56 of whom have answered "yes" to a particular survey question. What sample size n is required if we wish to estimate the population proportion p of "yes" answers with a margin of error of 0.02 at the 95% level of confidence?

    Answer: $n = 2017$

$$\dot{p} = \frac{56}{80} = 0.70$$

$$n = \frac{(z_{\alpha/2})^2 \dot{p}(1 - \dot{p})}{(PME)^2}$$

$$n = \frac{(1.96)^2 (0.7)(0.3)}{(0.02)^2} = 2016.84 \approx 2017$$

```
qnorm(0.025, lower.tail = FALSE) # For 95% confidence.

## [1] 1.959964
```

```r
(1.96^2) * (0.70) * (1 - 0.70) / (0.02 ^ 2) # Sample size.

## [1] 2016.84
```

check:

$$z_{\alpha/2}\sqrt{\frac{\dot{p}(1-\dot{p})}{n}} = (1.96)\sqrt{\frac{(0.70)(0.30)}{2017}} = 0.02$$

```r
qnorm(0.025, lower.tail = FALSE) * sqrt((0.70) * (1 - 0.70) / (2017))

## [1] 0.01999884
```

24. Referring to the previous exercise, what sample size should we recommend if we had no pilot study data to go on? Would n be greater or less than the n recommended in the previous exercise? Assume the study is investigating a question never before researched.

    Answer: The sample size should be greater than $n = 2017$ since we would use $\dot{p} = 0.50$, the most conservative value when determining sample size. In fact, when $\dot{p} = 0.50$, $n = 2401$.

$$n = \frac{z^2_{\alpha/2}\dot{p}(1-\dot{p})}{(PME)^2}$$

$$n = \frac{(1.96)^2(0.5)(0.5)}{(0.02)^2} = 2401$$

```r
(1.96^2) * (0.50) * (1 - 0.50) / (0.02 ^ 2)

## [1] 2401
```

25. In a study of consumer confidence among middle-class Indian families, 450 heads-of-households were interviewed in the Mumbai metropolitan area. When asked about having to cut back on discretionary purchases of big-ticket items, 212 of 450 responded that their families had done so over the previous six months.

(a) What is the 90% confidence interval estimate of the population proportion p of this type of household cutting back on discretionary spending?

Answer: 0.47 ± 0.0387

$$\overline{p} = \frac{212}{450} = 0.47$$

$$\overline{p} \pm z_{\alpha/2}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

$$0.47 \pm 1.645\sqrt{\frac{(0.47)(0.53)}{450}}$$

```
qnorm(0.05, lower.tail = FALSE)

## [1] 1.644854
```

$$0.47 \pm 0.0387$$

$$[0.4313, 0.5087]$$

```
# For the lower bound of the interval estimate, subtract
# the margin of error from the sample proportion.

0.47 - qnorm(0.05, lower.tail = FALSE) * sqrt((0.47) *
      (1 - 0.47) / (450))

## [1] 0.4313003

# For the upper bound of the interval estimate, add
# the margin of error to the sample proportion.

0.47 + qnorm(0.05, lower.tail = FALSE) * sqrt((0.47) *
      (1 - 0.47) / (450))

## [1] 0.5086997
```

(b) What sample size should we recommend to achieve a margin of error of 0.03?

Answer: $n = 749$

$$\dot{p} = \frac{212}{450} = 0.47$$

$$n = \frac{(z_{\alpha/2})^2\dot{p}(1-\dot{p})}{(PME)^2}$$

$$n = \frac{(1.645)^2(0.47)(0.53)}{(0.03)^2} = 748.97 \approx 749$$

```
(1.645^2) * (0.47) * (1 - 0.47) / (0.03 ^ 2) # Sample size.

## [1] 748.9676
```

check:

$$z_{\alpha/2}\sqrt{\frac{(\dot{p})(1-\dot{p})}{n}} = 1.645\sqrt{\frac{(0.47)(0.53)}{749}} = (1.645)(0.0182) = 0.03$$

```
qnorm(0.05, lower.tail = FALSE) * sqrt((0.47) *
      (1 - 0.47) / (749))

## [1] 0.02999668
```

26. Referring to the previous exercise, answer the following questions.

   (a) How large a sample size n would be required to achieve a 99% confidence interval estimate of p with a margin of error of 0.03?

   Answer: $n = 1837$

$$n = \frac{(z_{\alpha/2})^2\dot{p}(1-\dot{p})}{(PME)^2}$$

$$n = \frac{(2.576)^2(0.47)(0.53)}{(0.03)^2} = 1836.6 \approx 1837$$

```
qnorm(0.005, lower.tail = FALSE)

## [1] 2.575829
```

```
(2.576^2) * (0.47) * (1 - 0.47) / (0.03 ^ 2)

## [1] 1836.635
```

check:

$$z_{\alpha/2}\sqrt{\frac{(\dot{p})(1-\dot{p})}{n}} = 2.576\sqrt{\frac{(0.47)(0.53)}{1837}} = (2.576)(0.0116) = 0.03$$

```
qnorm(0.005, lower.tail = FALSE) * sqrt((0.47) *
       (1 - 0.47) / (1837))

## [1] 0.02999503
```

(b) How large a sample size n would be required to achieve a 99% confidence interval estimate of p with a margin of error of 0.025?

Answer: $n = 2645$

$$n = \frac{(z_{\alpha/2})^2 \dot{p}(1-\dot{p})}{(PME)^2}$$

$$n = \frac{(2.576)^2(0.47)(0.53)}{(0.025)^2} = 2644.8 \approx 2645$$

```
(2.576^2) * (0.47) * (1 - 0.47) / (0.025 ^ 2)

## [1] 2644.755
```

check:

$$z_{\alpha/2}\sqrt{\frac{(\dot{p})(1-\dot{p})}{n}} = 2.576\sqrt{\frac{(0.47)(0.53)}{2645}} = (2.575)(0.0097) = 0.025$$

```
qnorm(0.005, lower.tail = FALSE) * sqrt((0.47) *
       (1 - 0.47) / (2645))

## [1] 0.02499719
```

27. A human resources manager at a small university has been considering a change to the structure of employee benefits. To get an idea of how receptive faculty, administrators, and staff members might be to the proposed changes, she has decided to conduct a survey in which $n = 188$ respondents could register their support or opposition. Import `benefits.csv` (see the website) into the R Workspace and name it `E8_4`; the variable name is `agree`. Note: 0 denotes opposition, 1 support.

```
benefits <- read.csv('benefits.csv')
```

```
E8_4 <- benefits
```

(a) What is the 90% confidence interval estimate of the population proportion p of university employees who favor the change to their benefits package?

Answer: $[0.5697, 0.6857]$

```r
# Find variable name in benefits.

names(E8_4)

## [1] "agree"

# Find sample proportion; assign value to object pbar.

pbar <- sum(E8_4$agree) / length(E8_4$agree)

# What is the sample proportion?

pbar

## [1] 0.6276596

# For the lower bound of the interval estimate, subtract
# the margin of error from the sample proportion.

pbar - qnorm(0.05, lower.tail = FALSE) * sqrt((pbar) * (1 - pbar)
      / (length(E8_4$agree)))

## [1] 0.5696659

# For the upper bound of the interval estimate, add the
# margin of error to the sample proportion.

pbar + qnorm(0.05, lower.tail = FALSE) * sqrt((pbar) * (1 - pbar)
      / (length(E8_4$agree)))

## [1] 0.6856532
```

(b) What is the 95% confidence interval estimate of the population proportion p of university employees who favor the change to their benefits package?

Answer: $[0.5586, 0.6968]$

```r
# For the lower bound of the interval estimate,
# subtract the margin of error from the sample proportion.

pbar - qnorm(0.025, lower.tail = FALSE) * sqrt((pbar) * (1 - pbar)
      / (length(E8_4$agree)))

## [1] 0.5585559

# For the upper bound of the interval estimate,
# add the margin of error to the sample proportion.

pbar + qnorm(0.025, lower.tail = FALSE) * sqrt((pbar) * (1 - pbar)
      / (length(E8_4$agree)))

## [1] 0.6967633
```

(c) What is the 99% confidence interval estimate of the population proportion p of university employees who favor the change to their benefits package?

Answer: $[0.5368, 0.7185]$

```
# For the lower bound of the interval estimate,
# subtract the margin of error from the sample proportion.

pbar - qnorm(0.005, lower.tail = FALSE) * sqrt((pbar) * (1 - pbar)
      / (length(E8_4$agree)))

## [1] 0.5368419

# For the upper bound of the interval estimate,
# add the margin of error to the sample proportion.

pbar + qnorm(0.005, lower.tail = FALSE) * sqrt((pbar) * (1 - pbar)
      / (length(E8_4$agree)))

## [1] 0.7184772
```

28. Use the `t.test()` function to confirm that the confidence interval estimates from the previous exercise (parts (a), (b), and (c)) are correct.

(a) Answer: the 90% confidence interval estimate is [0.5692, 0.6861].

```
t.test(E8_4$agree, conf.level = 0.90)

##
##  One Sample t-test
##
## data:  E8_4$agree
## t = 17.755, df = 187, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  0.5692216 0.6860976
## sample estimates:
## mean of x
## 0.6276596
```

(b) Answer: the 95% confidence interval estimate is [0.5579, 0.6974].

```
t.test(E8_4$agree, conf.level = 0.95)

##
##  One Sample t-test
##
## data:  E8_4$agree
## t = 17.755, df = 187, p-value < 2.2e-16
```

```
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.5579200 0.6973991
## sample estimates:
## mean of x
## 0.6276596
```

(c) Answer: the 99% confidence interval estimate is [0.5357, 0.7197].

```
t.test(E8_4$agree, conf.level = 0.99)
```

```
##
##  One Sample t-test
##
## data:  E8_4$agree
## t = 17.755, df = 187, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 99 percent confidence interval:
##  0.5356609 0.7196582
## sample estimates:
## mean of x
## 0.6276596
```

Note that the answers to Exercises 27 and 28 diverge very slightly. But because the magnitude of the difference is so small, we will usually prefer to use the `t.test()` function when finding confidence interval estimates of the population proportion. While the approach taken in Exercise 27 provides answers that are exactly correct, it is easier and quicker to use the `t.test()` function. For all practical purposes, the results are the same.