Chapter 7: Statistics with R - 2nd Edition

Robert Stinerock

Student Exercises

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

1. `tv_hours.csv`
2. `exit.csv`

1. Draw a random sample of $n = 9$ from the `tv_hours.csv` data set (from the website). Use the `data[sample(nrow(data),n ),]` function. Name the random sample `E7_1`.

   (a) List nine elements of the random sample taken from `tv_hours.csv` . As a first step, be sure to import the data set into the R Workspace.

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

```
# Use the tv_hours[sample(nrow(tv_hours),9),] function to select
# a random sample of n = 9 from tv_hours and name it E7_1.

E7_1 <- tv_hours[sample(nrow(tv_hours), 9), ]
```

```
E7_1

##      X hours
## 63 63 13.18
## 66 66 13.42
## 17 17  7.50
## 94 94 17.37
## 99 99 20.22
## 3   3  3.29
## 51 51 11.73
## 27 27  9.35
## 74 74 14.14
```

   (b) Using this sample, what is the point estimate of the population mean μ?

```
# Use the mean() function to find the sample mean.

mean(E7_1$hours)

## [1] 12.24444
```

(c) What is the point estimate of the population standard deviation σ?

```
# Use the sd() function to find sample standard deviation.

sd(E7_1$hours)

## [1] 5.087551
```

2. Draw a second random sample of $n = 9$. Use `data[sample(nrow(data),n),]`. Name this random sample E7_2.

(a) List the nine elements of the random sample.

```
# Use the tv_hours[sample(nrow(tv_hours),9),] function to select
# a random sample of n=9 from tv_hours and name it E7_2.

E7_2 <- tv_hours[sample(nrow(tv_hours), 9), ]

# Exam the contents of E7_2.

E7_2

##     X hours
## 56 56 12.59
## 28 28  9.50
## 38 38 10.55
## 90 90 16.30
## 75 75 14.19
## 58 58 12.63
## 22 22  8.53
## 46 46 11.35
## 74 74 14.14
```

(b) Using this sample, what is the point estimate of the population mean μ?

```
# Use the mean() function to find the sample mean.

mean(E7_2$hours)

## [1] 12.19778
```

(c) What is the point estimate of the population standard deviation?

```
# Use the sd() function to find sample standard deviation.

sd(E7_2$hours)

## [1] 2.476756
```

3. Assuming that the tv_hours data set includes the entire population of interest, what is the population mean? Do the two point estimates (from the two random samples above) equal the population mean? Do they equal one another?

   Answer: In general, point estimates do not equal the population parameter they are intended to estimate because they are derived from samples that do not include all the elements of the population; nor can point estimates formed on different samples be expected to equal one another.

```
# The mean of the first random sample, E7_1.

mean(E7_1$hours)

## [1] 12.24444

# The mean of the second random sample, E7_2.

mean(E7_2$hours)

## [1] 12.19778

# The mean of the population from tv_hours.

mean(tv_hours$hours)

## [1] 11.6397
```

4. Referring to E7_1, E7_2, and the tv_hours data set, answer the following questions.

   (a) Assuming tv_hours includes data on the entire population of interest, what is the population mean μ?

```
# Use mean() for mean of entire population.

mean(tv_hours$hours)

## [1] 11.6397
```

   (b) What is the sampling error $|\overline{x} - \mu|$ for both random samples (that is, from both E7_1 and E7_2)?

```
# Find the absolute value of the difference between
# the sample mean of E7_1 and the population mean.

abs(mean(E7_1$hours)-mean(tv_hours$hours))

## [1] 0.6047444
```

```r
# Find the absolute value of the difference between
# the sample mean of E7_2 and the population mean.

abs(mean(E7_2$hours)-mean(tv_hours$hours))

## [1] 0.5580778
```

It is clear that the sampling error varies from one sample to the next. This should not be surprising since any given sample almost certainly includes different elements from any other sample, and thus have different sample means. In the computation of the sampling error, the only constant term is the population mean.

5. During the 2019 U.K. general election, 1500 voters were interviewed upon exiting a Manchester polling station where they had just cast their votes. The data are recorded as `1` for a Conservative Party vote and `0` for a Labour Party vote. Draw a random sample of $n = 25$. Apply function `data[sample(nrow(data),n),]`, and use the `exit` data set from the website. Name the random sample `E7_3`.

   (a) List the 25 elements of the random sample taken from `exit`. As a first step, import the `exit.csv` data into the R Workspace.

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

   ```r
   # Use function exit[sample(nrow(exit),25),] to select
   #a random sample of n=25 from exit and name it E7_3.

   E7_3 <- exit[sample(nrow(exit), 25), ]

   # Examine contents of E7_3.

   E7_3

   ##  [1] 1 0 1 1 0 1 1 1 0 1 1 0 1 0 1 1 1 1 1 0 0 1 1 0 1
   ```

   (b) Using this random sample, what is the point estimate of the population proportion p of Conservative Party voters?

   ```r
   # Use the mean() function to find the sample proportion.

   mean(E7_3)

   ## [1] 0.68
   ```

6. Draw a second random sample of $n = 25$ from `exit` and name it E7_4.

(a) List the 25 elements of the random sample taken from `exit`.

```
# Use function exit[sample(nrow(exit),25),] to select
# a random sample of n=25 from exit and name it E7_4.

E7_4 <- exit[sample(nrow(exit), 25), ]

# Examine the contents of E7_4.

E7_4

##  [1] 1 1 1 1 0 0 1 0 1 0 1 1 0 1 1 1 0 1 0 1 0 0 1 1 0
```

(b) Using this random sample, what is the point estimate of the population proportion p of Conservative Party voters?

```
# Use the mean() function to find the sample proportion.

mean(E7_4)

## [1] 0.6
```

7. Do the two point estimates (found in the preceding two exercises) equal the population proportion? Do they equal one another?

Answer: In general, point estimates do not equal the population parameter they are intended to estimate because they are derived from samples that do not include all the elements of the population; nor can point estimates formed on different samples be expected to equal one another. What is true for sample means is also true for sample proportions.

```
# The proportion of the first random sample, E7_3.

mean(E7_3)

## [1] 0.68

# The proportion of the second random sample, E7_4.

mean(E7_4)

## [1] 0.6

# The proportion of the population from exit.

mean(exit$obama)

## [1] 0.62
```

8. How many votes did each party receive from those in this sample of $n = 1500$

   Answer: Of the 1500 voters in sample, the Conservative Party received 930 votes while Labour received 570.

```
# Use the table() function to provide counts of 0 and 1.

table(exit)

## exit
##   0   1
## 570 930
```

9. Now for a bit of practice. Suppose a random sample consists of the following elements: 37, 14, 54, 91, 13, 88, 4, 16, 62, 18, 88, and 99. Copy and paste these values into the R Console and store them in an object named `E7_5`. Once this has been done, add the variable name `values` and create a data frame named `E7_6`. Answer the following questions using `E7_6`.

   (a) What is the point estimate of the population mean μ?

   Answer: $\overline{x} = 48.67$

$$\overline{x} = \frac{\sum_{i=1}^{12} x_i}{n} = 48.67$$

```
# (1) Use c() function to create the object E7_5.

E7_5 <- c(37, 14, 54, 91, 13, 88, 4, 16, 62, 18, 88, 99)

# (2) Examine contents of E7_5.

E7_5

##  [1] 37 14 54 91 13 88  4 16 62 18 88 99

# (3) Use data.frame() function to create the data frame
# named E7_6.  The variable name is values.

E7_6 <- data.frame(values = E7_5)

# (4) Examine the contents of E7_6.

E7_6
```

```
##    values
## 1      37
## 2      14
## 3      54
## 4      91
## 5      13
## 6      88
## 7       4
## 8      16
## 9      62
## 10     18
## 11     88
## 12     99

# (5) What is the point estimate of the population mean?

mean(E7_6$values)

## [1] 48.66667
```

(b) What is the point estimate of the population standard deviation σ?

Answer: $s = 35.98$

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{(n-1)}} = 35.98$$

```
# Use the sd() function for the sample standard deviation.



sd(E7_6$values)

## [1] 35.97811
```

10. When an Iberian tourism authority wanted to know from where travelers on one of their superhighways were coming, they monitored the bridge traffic connecting Castro Marim, Portugal with Ayamonte, Spain (crossing the Guadiana River). They found that for a random sample of 1062 vehicles, 377 had Portuguese license plates while 418 had Spanish plates. The remaining 267 vehicles had plates from a country other than Spain or Portugal.

(a) What is the point estimate of the population proportion p from Portugal?

Answer: $\overline{p} = 0.3550$

$$\overline{p} = \frac{377}{1062} = 0.3550$$

(b) What is the point estimate of the population proportion p from Spain?

Answer: $\overline{p} = 0.3936$

$$\overline{p} = \frac{418}{1062} = 0.3936$$

(c) What is the point estimate of the population proportion p from countries other than Portugal and Spain?

Answer: $\overline{p} = 0.2514$

$$\overline{p} = \frac{267}{1062} = 0.2514$$

11. A random sample of size $n = 36$ is drawn from a population with a mean of $\mu = -17$ and a standard deviation of $\sigma = 6$.

(a) What is $p(\overline{x} \leq -16)$?

Answer: 0.8413

$$p(\overline{x} \leq -16) = p(z \leq 1) = 0.8413$$

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

pnorm(-16, -17, 6 / sqrt(36))

## [1] 0.8413447
```

(b) What is $p(\overline{x} > -19)$

Answer: 0.9772

$$p(\overline{x} > -19) = p(z > -2) = 0.9772$$

```r
# Use 1 minus the pnorm() function.

1 - pnorm(-19, -17, 6 / sqrt(36))

## [1] 0.9772499
```

(c) What is $p(-18 \leq \overline{x} \leq -16)$?

Answer: 0.6827

$$p(-18 \leq \overline{x} \leq -16) = p(-1 \leq z \leq 1) = 0.6827$$

```
# Subtract pnorm(-18,-17,6/sqrt(36)) from pnorm(-16,-17,6/sqrt(36))

pnorm(-16, -17, 6 / sqrt(36)) - pnorm(-18, -17, 6 / sqrt(36))

## [1] 0.6826895
```

(d) What is $p(-19 \leq \overline{x} \leq -15)$?

Answer: 0.9545

$$p(-19 \leq \overline{x} \leq -15) = p(-2 \leq z \leq 2) = 0.9545$$

```
# Subtract pnorm(-19,-17,6/sqrt(36)) from pnorm(-15,-17,6/sqrt(36)).

pnorm(-15, -17, 6 / sqrt(36)) - pnorm(-19, -17, 6 / sqrt(36))

## [1] 0.9544997
```

12. The mean level of debt carried by students graduating from U.S. universities has now reached $\$27,000$. Use this value as the population mean μ and assume that the population standard deviation is $\sigma = \$4,500$. If a random sample of $n = 121$ graduating students is selected, answer the following questions.

(a) What is the probability that the sample mean $\overline{x}$ will fall within $\pm$ \$500 of the population mean μ? That is, what is $p(26500 \leq \overline{x} \leq 27500)$?

Answer: 0.7784

$$p(26500 \leq \overline{x} \leq 27500) = p(-1.22 \leq z \leq +1.22) = 0.7784$$

```
# Subtract pnorm(26500,27000,4500/sqrt(121) from
# pnorm(27500,27000,4500/sqrt(121)

pnorm(27500, 27000, 4500 / sqrt(121)) -
  pnorm(26500, 27000, 4500 / sqrt(121))

## [1] 0.7783764
```

(b) What is the probability that the sample mean $\overline{x}$ will fall within ± \$250 of the population mean μ? That is, what is $p(26750 \leq \overline{x} \leq 27250)$?

Answer: 0.4589

$$p(26750 \leq \overline{x} \leq 27250) = p(-0.61 \leq z \leq +0.61) = 0.4589$$

```
# Subtract pnorm(26750,27000,4500/sqrt(121) from
# pnorm(27250,27000,4500/sqrt(121)

pnorm(27250, 27000, 4500 / sqrt(121)) -
  pnorm(26750, 27000, 4500 / sqrt(121))

## [1] 0.458874
```

13. An administrator at a university in the U.K. wishes to estimate the mean age for its 3700 faculty members, and decides to draw a random sample of size $n = 37$ to derive the sample mean $\overline{x}$.

(a) Should the finite population correction factor be used in the computation of the standard error of the mean σ_x?

Answer: No, since $n/N = 37/3700 = 0.01 \leq 0.05$, the finite population correction factor is unnecessary.

(b) Calculate the standard error of the mean both with and without the finite population correction factor. Assume the population standard deviation is $\sigma = 11.1$ years. How far apart are the two values?

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{11.1}{\sqrt{37}} = 1.8248$$

$$\sigma_{\overline{x}} = \sqrt{\frac{N-n}{N-1}}\frac{\sigma}{\sqrt{n}} = \sqrt{\frac{3663}{3699}}\frac{11.1}{\sqrt{37}} = (0.9951)(1.8248) = 1.8159$$

Introducing the finite population correction factor reduces the value of $\sigma_{\overline{x}}$ by less than one-half of one percent. Clearly, when the size of the sample is small relative to the size of the population, the inclusion of this term makes almost no difference.

14. A study reports that teenagers spend an average of 31 hours a week online and texting. Assume that this is the population mean μ. Assume also that the population standard deviation is $\sigma = 7$ hours.

(a) If a random sample of $n = 64$ teenagers is selected, what is the probability that $\overline{x}$ is no more than 30? That is, what is $p(\overline{x} \leq 30)$?

Answer: 0.1265

$$p(\overline{x} \leq 30) = p(z \leq -1.14) = 0.1265$$

```
# Use the pnorm() function.

pnorm(30, 31, 7 / sqrt(64))

## [1] 0.126549
```

(b) What is the probability that $\overline{x}$ is greater than 33? That is, what is $p(\overline{x}>33)$?

Answer: 0.01114

$$p(\overline{x}>33) = p(z>2.29) = 1 - p(z \leq 2.29) = 0.01114$$

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

1 - pnorm(33, 31, 7 / sqrt(64))

## [1] 0.01113549
```

(c) What is the probability $\overline{x}$ will fall in the interval from 30 to 33 hours? That is, what is $p(30 \leq \overline{x} \leq 33)$

Answer:0.8623

$$p(30 \leq \overline{x} \leq 33) = p(-1.14 \leq z \leq 2.29) = 0.8623$$

```
# Subtract pnorm(30,31,7/sqrt(64)) from pnorm(33,31,7/sqrt(64)).

pnorm(33, 31, 7 / sqrt(64)) - pnorm(30, 31, 7 / sqrt(64))

## [1] 0.8623156
```

15. Referring to the previous exercise, would it be appropriate to include the finite population correction factor? Are there are circumstances where we might include it?

    Answer: No, we would not use the finite population correction factor. The reason is that although we cannot calculate the ratio n/N (because we do not know the size of N), we can assume that N is very large. Would we ever introduce the term in this question? Yes, if the population were defined as the population of all students at the local high school and if $N \leq 1280$. As we know, the decision to include the term depends on the ratio of the size of the sample n to the size of the population N.

16. Suppose that in a study of faculty salaries at US-based graduate schools of management, the standard error of the mean is $\sigma_{\overline{x}} = \$75$ but the population standard deviation is $\sigma = \$4875$.

    (a) What is the sample size n?

    Answer: 4225

    Since

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

$$n = (\frac{\sigma}{\sigma_{\overline{x}}})^2 = (\frac{4875}{75})^2 = 4225$$

    (b) What is the probability that the sample mean $\overline{x}$ will be within $\pm$ \$150 of the population mean μ?

    Answer: 0.9545

    Since

$$\frac{150}{75} = 2$$

    then

$$p(-2 \leq z \leq 2) = 0.9545$$

```
pnorm(2) - pnorm(-2)

## [1] 0.9544997
```

17. Referring to the previous exercise, demonstrate empirically that the probability that $\overline{x}$ is within $\pm$ \$150 of μ is 0.9545. Do not formally prove the result but rather substitute some value (any value) for μ and work out the result.

Let $\mu = \$100,000$, a value selected at random. Then

$$p(99850 \leq \overline{x} \leq 100150) = p(\frac{99850 - 100000}{75} \leq z \leq \frac{100150 - 100000}{75}) =$$

$$p(-2 \leq z \leq 2) = 0.9545$$

```
pnorm(100150, 100000, 75) - pnorm(99850, 100000, 75)

## [1] 0.9544997
```

Note: this result applies for any value of μ we might select.

18. Suppose a random sample of size $n = 200$ is drawn from a population with population proportion $p = 0.55$.

   (a) What is the expected value of $\overline{p}$?

   Answer: $E(\overline{p}) = p = 0.55$

   (b) What is the standard error of the proportion $\sigma_{\overline{p}}$?

   Answer: 0.0352

$$\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.55)(0.45)}{200}} = 0.0352$$

   (c) What is the sampling distribution of $\overline{p}$?

   Answer: the sampling distribution of $\overline{p}$ is the probability distribution of all possible values of the sample proportion $\overline{p}$.

19. A random sample of size $n = 100$ is selected from a population with $p = 0.60$.

   (a) What is the probability that the sample proportion $\overline{p}$ will be within $\pm$ 0.02 of the population proportion? That is, what is $p(0.58 \leq \overline{p} \leq 0.62)$?

   Answer: 0.3182

$$\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.60)(0.40)}{100}} = 0.0490$$

$$p(0.58 \leq \overline{p} \leq 0.62) = p(-0.41 \leq z \leq 0.41) = 0.3182$$

```r
# Subtract pnorm(-0.41) from pnorm(0.41).

pnorm(0.41) - pnorm(-0.41)

## [1] 0.3181941
```

(b) What is the probability that the sample proportion $\overline{p}$ will be within $\pm$ 0.05 of the population proportion? That is, what is $p(0.55 \leq \overline{p} \leq 0.65)$?

Answer: 0.6923

$$p(0.55 \leq \overline{p} \leq 0.65) = p(-1.02 \leq z \leq 1.02) = 0.6923$$

```r
# Subtract pnorm(-1.02) from pnorm(1.02).

pnorm(1.02) - pnorm(-1.02)

## [1] 0.6922715
```

(c) What is the probability that the sample proportion $\overline{p}$ will be within $\pm$ 0.10 of the population proportion? That is, what is $p(0.50 \leq \overline{p} \leq 0.70)$?

Answer: 0.9586

$$p(0.50 \leq \overline{p} \leq 0.70) = p(-2.04 \leq z \leq 2.04) = 0.9586$$

```r
# Subtract pnorm(-2.04) from pnorm(2.04).

pnorm(2.04) - pnorm(-2.04)

## [1] 0.9586497
```

20. A population proportion is $p = 0.50$. What is the standard error of the proportion for the following sample sizes?

(a) If $n = 50$, what is $\sigma_{\overline{p}}$?

Answer: 0.0707

$$\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.50)(0.50)}{50}} = 0.0707$$

(b) If $n = 200$, what is $\sigma_{\overline{p}}$?

Answer: 0.0354

$$\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.50)(0.50)}{200}} = 0.0354$$

(c) If $n = 800$, what is $\sigma_{\overline{p}}$?

Answer: 0.0177

$$\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.50)(0.50)}{800}} = 0.0177$$

(d) If $n = 3200$, what is $\sigma_{\overline{p}}$?

Answer: 0.0088

$$\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.50)(0.50)}{3200}} = 0.0088$$

21. What can we conclude about the relationship between the size of the sample n and the magnitude of the standard error of the proportion $\sigma_{\overline{p}}$?

    Answer: Larger sample sizes result in smaller standard errors and greater precision. However, there are diminishing returns characterizing this relationship. In fact, for each quadrupling of the sample size, we reduce the standard error by only half.

22. Assuming that the population proportion is $p = 0.50$, find $p(0.49 \leq \overline{p} \leq 0.51)$ for each of the sample sizes below.

    (a) What is $p(0.49 \leq \overline{p} \leq 0.51)$ if $n = 50$?

Answer: 0.1124

$$p(0.49 \leq \overline{p} \leq 0.51) = p(-0.1414 \leq z \leq +0.1414) = 0.1124$$

```
pnorm(0.1414) - pnorm(-0.1414)

## [1] 0.112446
```

(b) What is $p(0.49 \leq \overline{p} \leq 0.51)$ if $n = 200$?

Answer: 0.2227

$$p(0.49 \leq \overline{p} \leq 0.51) = p(-0.2828 \leq z \leq +0.2828) = 0.2227$$

```
pnorm(0.2828) - pnorm(-0.2828)

## [1] 0.2226698
```

(c) What is $p(0.49 \leq \overline{p} \leq 0.51)$ if $n = 800$?

Answer: 0.4284

$$p(0.49 \leq \overline{p} \leq 0.51) = p(-0.5657 \leq z \leq +0.5657) = 0.4284$$

```
pnorm(0.5657) - pnorm(-0.5657)

## [1] 0.4284023
```

(d) What is $p(0.49 \leq \overline{p} \leq 0.51)$ if $n = 3200$?

Answer: 0.7421

$$p(0.49 \leq \overline{p} \leq 0.51) = p(-1.1314 \leq z \leq +1.1314) = 0.7421$$

```
pnorm(1.1314) - pnorm(-1.1314)

## [1] 0.7421132
```

23. The percentage of people who are left-handed is not known with certainty but it is thought to be about 12%. Assume the population proportion of left-handed people is $p = 0.12$.

(a) If a sample of $n = 400$ people is chosen randomly, what is the probability that the proportion of left-handers will be within $\pm$ 0.02 of p? In other words, what is $p(0.10 \leq \overline{p} \leq 0.14)$?

Answer: 0.7813

$$p(0.10 \leq \overline{p} \leq 0.14) = p(-1.23 \leq z \leq 1.23) = 0.7813$$

```r
pnorm(1.23)-pnorm(-1.23)

## [1] 0.7813029
```

(b) If a sample of $n = 800$ people is chosen randomly, what is the probability that the proportion of left-handers will be within $\pm$ 0.02 of p? In other words, what is $p(0.10 \leq \overline{p} \leq 0.14)$?

Answer: 0.9181

$$p(0.10 \leq \overline{p} \leq 0.14) = p(-1.74 \leq z \leq 1.74) = 0.9181$$

```r
pnorm(1.74)-pnorm(-1.74)

## [1] 0.918141
```

24. Referring to the previous exercise where we take a random sample of $n = 400$, suppose we learn that the population from which we are drawing the sample consists of the entire student body of a grammar school in Birmingham where the entire student body is made up of 1200 students altogether. The school administration is in the process of building a new auditorium, and it wants to make sure that there are a sufficient number of seats to accommodate the left-handed students. What is the standard error of the proportion?

Answer: 0.0133

We must use the Finite Population Correction Factor since 400/1200 = 0.3333>0.05

$$\sigma_{\overline{p}} = \sqrt{\frac{N-n}{N-1}}\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{1200-400}{1200-1}}\sqrt{\frac{0.12(0.88)}{400}} = 0.0133$$

25. A quality control inspector is always on the lookout for substandard parts and components provided to her manufacturing company by outside suppliers. Because most shipments contain some defective items, each must be subjected to inspection. Naturally, some shipments contain more defectives than others, and it is the job of the inspector to identify the most defective-laden shipments so that they can be returned to the supplier. Suppose the inspector selects a sample of $n = 100$ items from a given shipment for testing. Unbeknownst to the inspector, this particular shipment includes 9% defective components. If the policy is to return any shipment with at least 5% defectives, what is the probability that this bad shipment will be accepted as good anyway?

    Answer: 0.0885

$$\overline{p} = \frac{9}{100} = 0.09$$

$$\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.09)(0.91)}{100}} = 0.0286$$

$$p(\overline{p} \leq 0.05) = p(z \leq \frac{0.05 - 0.09}{0.0286}) = p(z \leq -1.40) = 0.08076$$

```
pnorm(-1.40)

## [1] 0.08075666
```

    Thus, there is nearly a 0.081 probability that this *bad* shipment will sneak in as *good*. It should be clear that by increasing the sample size n, the inspector can reduce the probability of accepting a shipment with too many defective components. The downside to testing large samples, however, is that it is expensive and time-consuming to test large numbers of items.