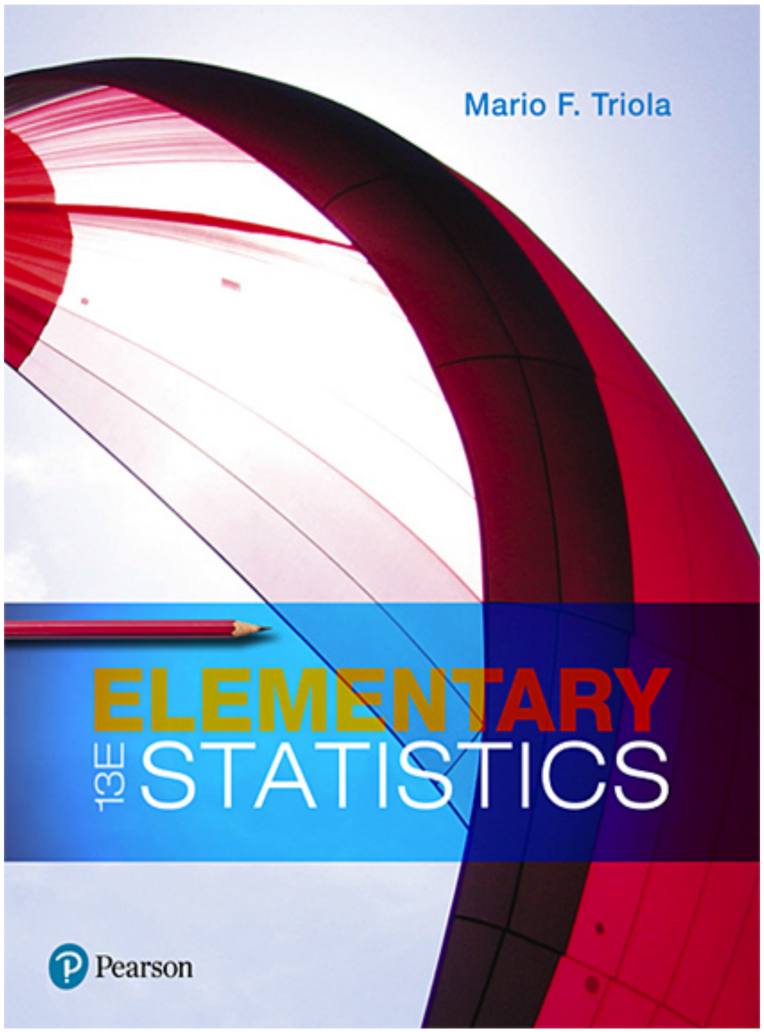
Elementary Statistics

Thirteenth Edition



Chapter 7

Estimating Parameters and Determining Sample Sizes

Estimating Parameters and Determining Sample Sizes

7-1 Estimating a Population Proportion

7-2 Estimating a Population Mean

7-3 Estimating a Population Standard Deviation or Variance

7-4 Bootstrapping: Using Technology for Estimates

Key Concept

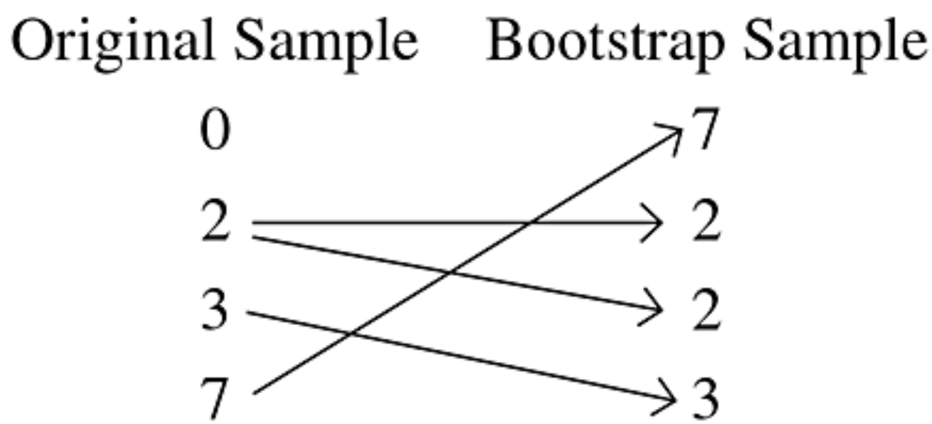
The preceding sections presented methods for estimating population proportions, means, and standard deviations (or variances). All of those methods have certain requirements that limit the situations in which they can be used. When some of the requirements are not satisfied, we can often use the bootstrap method to estimate a parameter with a confidence interval. The bootstrap method typically requires the use of software.

Bootstrap Sample

- Bootstrap Sample
  - Given a simple random sample of size *n,* a **bootstrap sample** is another random sample of *n* values obtained **with replacement** from the original sample.

Example: Bootstrap Sample of Incomes (1 of 2)

When the author collected annual incomes of current statistics students, he obtained these results (in thousands of dollars): 0, 2, 3, 7.



Example: Bootstrap Sample of Incomes (2 of 2)

The sample of {7, 2, 2, 3} is one bootstrap sample obtained from the original sample. Other bootstrap samples may be different.

Incomes tend to have distributions that are skewed instead of being normal, so we should not use the methods of Section 7-2 with a small sample of incomes. This is a situation in which the bootstrap method comes to the rescue.

Bootstrap Procedure for a Confidence Interval Estimate of a Parameter (1 of 2)

1. Given a simple random sample of size *n*, obtain many (such as 1000 or more) bootstrap samples of the same size *n*.
2. For the parameter to be estimated, find the corresponding statistic for each of the bootstrap samples. (Example: For a confidence estimate of μ, find the **sample mean** $\bar{x}$ from each bootstrap sample.)
3. Sort the list of sample statistics from low to high.

Bootstrap Procedure for a Confidence Interval Estimate of a Parameter (2 of 2)

4. Using the sorted list of the statistics, create the confidence interval by finding corresponding percentile values. Procedures for finding percentiles are given in Section 3-3. (Example: Using a list of sorted sample means, the 90% confidence interval limits are P_5 and P_{95}. The 90% confidence interval estimate of μ is $P_5 < \mu < P_{95}$.)

Proportions

When working with proportions, it is very helpful to represent the data from the two categories by using 0's and 1's.

Example: Eye Color Survey: Bootstrap CI for Proportion (1 of 11)

In a survey, four randomly selected subjects were asked if they have brown eyes, and here are the results: 0, 0, 1, 0 (where 0 = no and 1 = yes). Use the bootstrap resampling procedure to construct a 90% confidence interval estimate of the population proportion *p*, the proportion of people with brown eyes in the population.

Example: Eye Color Survey: Bootstrap CI for Proportion (2 of 11)

Solution

Requirement Check

The sample is a simple random sample. (There is no requirement of at least 5 successes and at least 5 failures or $np \geq 5$ and $nq \geq 5$. There is no requirement that the sample must be from a normally distributed population.)

Example: Eye Color Survey: Bootstrap CI for Proportion (3 of 11)

Solution

Step 1:

In the table on the next slide, we created 20 bootstrap samples from the original sample of 0, 0, 1, 0.

Example: Eye Color Survey: Bootstrap CI for Proportion (4 of 11)

Solution

Bootstrap Samples for *p*.

Bootstrap Sample				$\hat{p}$	Sorted $\hat{p}$	
1	0	0	1	0.50	0.00	→ $P_5 = 0.00$
1	0	1	0	0.50	0.00	
0	1	1	1	0.75	0.00	
0	0	0	0	0.00	0.00	
0	1	0	0	0.25	0.25	
1	0	0	0	0.25	0.25	
0	1	0	1	0.50	0.25	
1	0	0	0	0.25	0.25	
0	0	0	0	0.00	0.25	
0	0	1	1	0.50	0.25	**90% Confidence Interval:**
0	0	0	1	0.25	0.25	$0.00 < p < 0.75$
0	0	1	0	0.25	0.25	
1	1	1	0	0.75	0.50	
0	0	0	0	0.00	0.50	
0	0	0	0	0.00	0.50	
0	1	1	0	0.50	0.50	
0	0	1	0	0.25	0.50	
1	0	0	0	0.25	0.75	
1	1	1	0	0.75	0.75	→ $P_{95} = 0.75$
0	0	0	1	0.25	0.75	

Example: Eye Color Survey: Bootstrap CI for Proportion (5 of 11)

Solution

Step 2:

Because we want a confidence interval estimate of the population proportion *p*, we want the sample proportion $\hat{p}$ for each of the 20 bootstrap samples, and those sample proportions are shown in the column to the right of the bootstrap samples on the next slide.

Example: Eye Color Survey: Bootstrap CI for Proportion (6 of 11)

Solution

Bootstrap Samples for *p*.

Bootstrap Sample				$\hat{p}$	Sorted $\hat{p}$
1	0	0	1	0.50	0.00
1	0	1	0	0.50	0.00
0	1	1	1	0.75	0.00
0	0	0	0	0.00	0.00
0	1	0	0	0.25	0.25
1	0	0	0	0.25	0.25
0	1	0	1	0.50	0.25
1	0	0	0	0.25	0.25
0	0	0	0	0.00	0.25
0	0	1	1	0.50	0.25
0	0	0	1	0.25	0.25
0	0	1	0	0.25	0.25
1	1	1	0	0.75	0.50
0	0	0	0	0.00	0.50
0	0	0	0	0.00	0.50
0	1	1	0	0.50	0.50
0	0	1	0	0.25	0.50
1	0	0	0	0.25	0.75
1	1	1	0	0.75	0.75
0	0	0	1	0.25	0.75

$P_5 = 0.00$

90% Confidence Interval:

$0.00 < p < 0.75$

$P_{95} = 0.75$

Example: Eye Color Survey: Bootstrap CI for Proportion (7 of 11)

Solution

Step 3:

The column of data shown farthest to the right is a list of the 20 sample proportions arranged in order ("sorted") from lowest to highest.

Example: Eye Color Survey: Bootstrap CI for Proportion (8 of 11)

Solution

Bootstrap Samples for *p*.

Bootstrap Sample				$\hat{p}$	Sorted $\hat{p}$
1	0	0	1	0.50	0.00
1	0	1	0	0.50	0.00
0	1	1	1	0.75	0.00
0	0	0	0	0.00	0.00
0	1	0	0	0.25	0.25
1	0	0	0	0.25	0.25
0	1	0	1	0.50	0.25
1	0	0	0	0.25	0.25
0	0	0	0	0.00	0.25
0	0	1	1	0.50	0.25
0	0	0	1	0.25	0.25
0	0	1	0	0.25	0.25
1	1	1	0	0.75	0.50
0	0	0	0	0.00	0.50
0	0	0	0	0.00	0.50
0	1	1	0	0.50	0.50
0	0	1	0	0.25	0.50
1	0	0	0	0.25	0.75
1	1	1	0	0.75	0.75
0	0	0	1	0.25	0.75

$P_5 = 0.00$

90% Confidence Interval:

$0.00 < p < 0.75$

$P_{95} = 0.75$

Example: Eye Color Survey: Bootstrap CI for Proportion (9 of 11)

Solution

Step 4:

Because we want a confidence level of 90%, we want to find the percentiles P_5 and P_{95}. Recall that P_5 separates the lowest 5% of values, and P_{95} separates the top 5% of values.

Using the methods from Section 3-3 for finding percentiles, we use the *sorted* list of bootstrap sample proportions to find that $P_5 = 0.00$ and $P_{95} = 0.75$. The 90% confidence interval estimate of the population proportion is $0.00 < p < 0.75$.

Example: Eye Color Survey: Bootstrap CI for Proportion (10 of 11)

Solution

Bootstrap Samples for *p*.

Bootstrap Sample				$\hat{p}$	Sorted $\hat{p}$
1	0	0	1	0.50	0.00
1	0	1	0	0.50	0.00
0	1	1	1	0.75	0.00
0	0	0	0	0.00	0.00
0	1	0	0	0.25	0.25
1	0	0	0	0.25	0.25
0	1	0	1	0.50	0.25
1	0	0	0	0.25	0.25
0	0	0	0	0.00	0.25
0	0	1	1	0.50	0.25
0	0	0	1	0.25	0.25
0	0	1	0	0.25	0.25
1	1	1	0	0.75	0.50
0	0	0	0	0.00	0.50
0	0	0	0	0.00	0.50
0	1	1	0	0.50	0.50
0	0	1	0	0.25	0.50
1	0	0	0	0.25	0.75
1	1	1	0	0.75	0.75
0	0	0	1	0.25	0.75

$P_5 = 0.00$

90% Confidence Interval:

$0.00 < p < 0.75$

$P_{95} = 0.75$

Pearson

Example: Eye Color Survey: Bootstrap CI for Proportion (11 of 11)

Interpretation

The confidence interval of $0.00 < p < 0.75$ is quite wide. After all, every confidence interval for every proportion must fall between 0 and 1, so the 90% confidence interval of $0.00 < p < 0.75$ doesn't seem to be helpful, but it is based on only four sample values.

Means

Earlier in this chapter we noted that when constructing a confidence interval estimate of a population mean, there is a requirement that the sample is from a normally distributed population or the sample size is greater than 30. The bootstrap method can be used when this requirement is not satisfied.

Example: Incomes: Bootstrap CI for Mean (1 of 10)

When the author collected a simple random sample of annual incomes of his statistics students, he obtained these results (in thousands of dollars): 0, 2, 3, 7. Use the bootstrap resampling procedure to construct a 90% confidence interval estimate of the mean annual income of the population of all of the author’s statistics students.

Example: Incomes: Bootstrap CI for Mean (2 of 10)

Solution

Requirement Check

The sample is a simple random sample and there is no requirement that the sample must be from a normally distributed population. Because distributions of incomes are typically skewed instead of normal, we should not use the methods of Section 7-2 for finding the confidence interval, but the bootstrap method can be used.

Example: Incomes: Bootstrap CI for Mean (3 of 10)

Solution

Step 1:

In the table on the next slide, we created 20 bootstrap samples (with replacement!) from the original sample of 0, 2, 3, 7. (Here we use only 20 bootstrap samples so we have a manageable example that doesn’t occupy many pages of text, but we usually want at least 1000 bootstrap samples.)

Example: Incomes: Bootstrap CI for Mean (4 of 10)

Solution

Bootstrap Samples for μ.

Bootstrap Sample				$\bar{x}$	Sorted $\bar{x}$
3	3	0	2	2.00	1.75
0	3	2	2	1.75	1.75
7	0	2	7	4.00	1.75
3	2	7	3	3.75	2.00
0	0	7	2	2.25	2.00
7	0	0	3	2.50	2.25
3	0	3	2	2.00	2.50
3	7	3	7	5.00	2.50
0	3	2	2	1.75	2.50
0	3	7	0	2.50	2.75
0	7	2	2	2.75	3.00
7	2	2	3	3.50	3.25
7	2	3	7	4.75	3.25
2	7	2	7	4.50	3.50
0	7	2	3	3.00	3.75
7	3	7	2	4.75	4.00
3	7	0	3	3.25	4.50
0	0	3	7	2.50	4.75
3	3	7	0	3.25	4.75
2	0	2	3	1.75	5.00

$P_5 = 1.75$

90% Confidence Interval:

$1.75 < \mu < 4.875$

$P_{95} = 4.875$

Example: Incomes: Bootstrap CI for Mean (5 of 10)

Solution

Step 2:

Because we want a confidence interval estimate of the population mean μ, we want the sample mean $\bar{x}$ for each of the 20 bootstrap samples, and those sample means are shown in the column to the right of the bootstrap samples.

Example: Incomes: Bootstrap CI for Mean (6 of 10)

Solution

Bootstrap Samples for μ.

Bootstrap Sample				$\bar{x}$	Sorted $\bar{x}$	
3	3	0	2	2.00	1.75	
0	3	2	2	1.75	1.75	$P_5 = 1.75$
7	0	2	7	4.00	1.75	
3	2	7	3	3.75	2.00	
0	0	7	2	2.25	2.00	
7	0	0	3	2.50	2.25	
3	0	3	2	2.00	2.50	
3	7	3	7	5.00	2.50	
0	3	2	2	1.75	2.50	
0	3	7	0	2.50	2.75	**90% Confidence Interval:**
0	7	2	2	2.75	3.00	$1.75 < \mu < 4.875$
7	2	2	3	3.50	3.25	
7	2	3	7	4.75	3.25	
2	7	2	7	4.50	3.50	
0	7	2	3	3.00	3.75	
7	3	7	2	4.75	4.00	
3	7	0	3	3.25	4.50	
0	0	3	7	2.50	4.75	
3	3	7	0	3.25	4.75	
2	0	2	3	1.75	5.00	$P_{95} = 4.875$

Example: Incomes: Bootstrap CI for Mean (7 of 10)

Solution

Step 3:

The column of data shown farthest to the right is a list of the 20 sample means arranged in order ("sorted") from lowest to highest.

Example: Incomes: Bootstrap CI for Mean (8 of 10)

Solution

Bootstrap Samples for μ.

Bootstrap Sample				$\bar{x}$	Sorted $\bar{x}$
3	3	0	2	2.00	1.75
0	3	2	2	1.75	1.75
7	0	2	7	4.00	1.75
3	2	7	3	3.75	2.00
0	0	7	2	2.25	2.00
7	0	0	3	2.50	2.25
3	0	3	2	2.00	2.50
3	7	3	7	5.00	2.50
0	3	2	2	1.75	2.50
0	3	7	0	2.50	2.75
0	7	2	2	2.75	3.00
7	2	2	3	3.50	3.25
7	2	3	7	4.75	3.25
2	7	2	7	4.50	3.50
0	7	2	3	3.00	3.75
7	3	7	2	4.75	4.00
3	7	0	3	3.25	4.50
0	0	3	7	2.50	4.75
3	3	7	0	3.25	4.75
2	0	2	3	1.75	5.00

$P_5 = 1.75$

90% Confidence Interval:

$1.75 < \mu < 4.875$

$P_{95} = 4.875$

Example: Incomes: Bootstrap CI for Mean (9 of 10)

Solution

Step 4:

Because we want a confidence level of 90%, we want to find the percentiles P_5 and P_{95}. Recall that P_5 separates the lowest 5% of values, and P_{95} separates the top 5% of values. Using the methods from Section 3-3 for finding percentiles, we use the **sorted** list of bootstrap sample means to find that $P_5 = 1.75$ and $P_{95} = 4.875$. The 90% confidence interval estimate of the population mean is $1.75 < \mu < 4.875$, where the values are in thousands of dollars.

Example: Incomes: Bootstrap CI for Mean (10 of 10)

Solution

Bootstrap Samples for μ.

Bootstrap Sample				$\bar{x}$	Sorted $\bar{x}$
3	3	0	2	2.00	1.75
0	3	2	2	1.75	1.75
7	0	2	7	4.00	1.75
3	2	7	3	3.75	2.00
0	0	7	2	2.25	2.00
7	0	0	3	2.50	2.25
3	0	3	2	2.00	2.50
3	7	3	7	5.00	2.50
0	3	2	2	1.75	2.50
0	3	7	0	2.50	2.75
0	7	2	2	2.75	3.00
7	2	2	3	3.50	3.25
7	2	3	7	4.75	3.25
2	7	2	7	4.50	3.50
0	7	2	3	3.00	3.75
7	3	7	2	4.75	4.00
3	7	0	3	3.25	4.50
0	0	3	7	2.50	4.75
3	3	7	0	3.25	4.75
2	0	2	3	1.75	5.00

$P_5 = 1.75$

90% Confidence Interval:

$1.75 < \mu < 4.875$

$P_{95} = 4.875$

Standard Deviations

In Section 7-3 we noted that when constructing confidence interval estimates of population standard deviations or variances, there is a requirement that the sample must be from a population with normally distributed values. Even if the sample is large, this normality requirement is much stricter than the normality requirement used for estimating population means. Consequently, the bootstrap method becomes more important for confidence interval estimates of σ or σ^2.

Example: Incomes: Bootstrap CI for Standard Deviation (1 of 4)

Use these same incomes (thousands of dollars) from the previous example: 0, 2, 3, 7. Use the bootstrap resampling procedure to construct a 90% confidence interval estimate of the population standard deviation σ, the standard deviation of the annual incomes of the population of the author's statistics students.

Example: Incomes: Bootstrap CI for Standard Deviation (2 of 4)

Solution

Requirement Check

The same requirement check used in the previous example applies here.

Example: Incomes: Bootstrap CI for Standard Deviation (3 of 4)

Solution

The same basic procedure used in the previous example is used here. The previous example already includes 20 bootstrap samples, so here we find the **standard deviation** of each bootstrap sample, and then we sort them to get this sorted list of sample standard deviations:

1.26	1.26	1.26	1.41	1.41	2.22	2.31	2.38	2.63	2.63
2.87	2.87	2.89	2.94	2.99	3.30	3.32	3.32	3.32	3.56

Example: Incomes: Bootstrap CI for Standard Deviation (4 of 4)

Solution

The 90% confidence interval limits are found from this sorted list of standard deviations by finding P_5 and P_{95}. Using the methods from Section 3-3, we get $P_5 = 1.26$ and $P_{95} = 3.44$. The 90% confidence interval estimate of the population standard deviation s is $1.26 < \sigma < 3.44$, where the values are in thousands of dollars.