edX　　　BerkeleyX: Data8.2x
　　　　　　Foundations of Data Science: Inferential Thinking by Resampling

Help　　sandipan_dey ▾

Course > Week 5 > Reading and Practice for Section 10 > Reading and Practice for Section 10

Reading and Practice for Section 10

This guide assumes that you have watched the videos for Section 10.

This corresponds to textbook section:

Chapter 13.1: Percentiles ⬀

Chapter 13.2: The Bootstrap ⬀

In section 10, we developed a way to estimate an unknown parameter. A statistic based on a random sample can be a reasonable estimate of an unknown parameter in the population. Bootstrapping is a technique that generates new samples by resampling from an original sample. In order to bootstrap, the original sample must be drawn at random and be large enough so that it resembles the population. In this case, the distribution of a large random sample will likely resemble the distribution of the population it was drawn from.

Try the following practice question.

Practice for Section 10

0 points possible (ungraded)
Suppose we have a sample of the heights of 100 Data 8.2x students, contained in a table `heights`, drawn as a simple random sample from the entire population of Data 8.2x students. We would like to estimate the median height of the Data 8.2x population. To do so, we will use the median height in our sample as our estimate.

Here are the first 5 rows of the table `heights`:

Height

67

72

78

65

66

Let's identify the population parameter and the sample statistic for this experiement.

The median height of all Data 8.2x students.

Population Parameter ▼   ✔ **Answer:** Population Parameter

The median height of Data 8.2x students in our sample.

Sample Statistic   ▼   ✔ **Answer:** Sample Statistic

Is the population parameter random or not random?

Not Random   ▼   ✔ **Answer:** Not Random

Is the sample statistic random or not random?

Random   ▼   ✔ **Answer:** Random

We would like to use the bootstrap to generate an empirical distribution for our statistic and calculate a **90%** confidence interval. Fill in the code blanks below that does exactly that. Please use the `sample` method.

```
medians = []

for i in np.arange(5000):
    resample = ____A____.____B____
    median = ____C____(____D____.column('Height'))
    medians.append(median)

[percentile(5, ____E____), ____F____(____G____, medians)]
```

Blank A

heights   ✔ **Answer:** heights

Blank B

sample()   ✔ **Answer:** sample() **or** sample(100) **or** sample(100, with_replacement=True)

Blank C

np.median   ✔ **Answer:** np.median

Blank D

resample   ✔ **Answer:** resample

Blank E

medians   ✔ **Answer:** medians

Blank F

percentile   ✔ **Answer:** percentile

Blank G

95   ✔ **Answer:** 95

Submit

ℹ   Answers are displayed within the problem

Discussion

Topic: Week 5 / Section 10 (Reading and Practice)

Hide Discussion

Add a Post

Show all posts ▼	by recent activity ▼
💬 **Textbook Links** The links to the textbook sections are not working.	2