Comprehension Check: Bootstrap

Q1

1/1 point (graded)

The `createResample` function can be used to create bootstrap samples. For example, we can create 10 bootstrap samples for the `mnist_27` dataset like this:

```
set.seed(1995)
indexes <- createResample(mnist_27$train$y, 10)
```

How many times do `3`, `4`, and `7` appear in the first resampled index?

Enter the number of times 3 appears:

1 ✔ **Answer:** 1

1

Enter the number of times 4 appears:

4 ✔ **Answer:** 4

4

Enter the number of times 7 appears:

0 ✔ **Answer:** 0

0

Explanation

You can find the number of times each digit appears using this code:

```
sum(indexes[[1]] == 3)
sum(indexes[[1]] == 4)
sum(indexes[[1]] == 7)
```

Submit You have used 1 of 10 attempts

Answers are displayed within the problem

Q2

1/1 point (graded)

We see that some numbers appear more than once and others appear no times. This has to be this way for each dataset to be independent. Repeat the exercise for all the resampled indexes.

What is the total number of times that 3 appears in all of the resampled indexes?

11 ✔ **Answer:** 11

11

Explanation

You can find the number of times 3 appears using this code:

```
x=sapply(indexes, function(ind){
        sum(ind == 3)
})
sum(x)
```

Submit You have used 1 of 10 attempts

Answers are displayed within the problem

Q3

2/2 points (graded)

Generate a random dataset using the following code:

```
y <- rnorm(100, 0, 1)
```

Estimate the 75th quantile, which we know is `qnorm(0.75)`, with the sample quantile: `quantile(y, 0.75)`.

Set the seed to 1 and perform a Monte Carlo simulation with 10,000 repetitions, generating the random dataset and estimating the 75th quantile each time. What is the expected value and standard error of the 75th quantile?

Expected value

0.6654465 ✔ **Answer:** 0.666

0.6654465

Standard error

0.1351181 ✔ **Answer:** 0.135

0.1351181

Explanation

The following code can be used to run the simulation and calculate the expected value and standard error:

```
set.seed(1)
B <- 10000
q_75 <- replicate(B, {
        y <- rnorm(100, 0, 1)
        quantile(y, 0.75)
})

mean(q_75)
sd(q_75)
```

Submit You have used 1 of 10 attempts

Answers are displayed within the problem

Q4

2/2 points (graded)

In practice, we can't run a Monte Carlo simulation. Use the sample:

```
set.seed(1)
y <- rnorm(100, 0, 1)
```

Set the seed to 1 again after generating `y` and use 10 bootstrap samples to estimate the expected value and standard error of the 75th quantile.

Expected value

0.7312648 ✔ **Answer:** 0.731

0.7312648

Standard error

0.07419278 ✔ **Answer:** 0.0742

0.07419278

Explanation
The following code can be used to take 10 bootstrap samples and calculate the expected value and standard error:

```
set.seed(1)
indexes <- createResample(y, 10)
q_75_star <- sapply(indexes, function(ind){
        y_star <- y[ind]
        quantile(y_star, 0.75)
})
mean(q_75_star)
sd(q_75_star)
```

Submit You have used 1 of 10 attempts

Answers are displayed within the problem

Q5

1/1 point (graded)
Repeat the exercise from Q4 but with 10,000 bootstrap samples instead of 10. Set the seed to 1.

Expected value

0.6737512 ✔ **Answer:** 0.674

0.6737512

Standard error

0.0930575 ✔ **Answer:** 0.0931

0.0930575

Explanation
The following code can be used to take 10,000 bootstrap samples and calculate the expected value and standard error:

```
set.seed(1)
indexes <- createResample(y, 10000)
q_75_star <- sapply(indexes, function(ind){
        y_star <- y[ind]
        quantile(y_star, 0.75)
})
mean(q_75_star)
sd(q_75_star)
```

Submit You have used 1 of 10 attempts

Answers are displayed within the problem

Q6

1/1 point (graded)
When doing bootstrap sampling, the simulated samples are drawn from the empirical distribution of the original data.

True or False: The bootstrap is particularly useful in situations in which a tractable variance formula does exist.

- True
- False ✔

Explanation
False. The bootstrap is particularly useful in situations in which a tractable variance formula does NOT exist.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Ask your questions or make your comments about Bootstrap here! **Remember, one of the best ways to reinforce your own learning is by explaining something to someone else, so we encourage you to answer each other's questions (without giving away the answers, of course).**

Some reminders:

- Search the discussion board before posting to see if someone else has asked the same thing before asking a new question.
- Please be specific in the title and body of your post regarding which question you're asking about to facilitate answering your question.
- Posting snippets of code is okay, but posting full code solutions is not.
- If you do post snippets of code, please format it as code for readability. If you're not sure how to do this, there are instructions in a pinned post in the "general" discussion forum.

Discussion: Bootstrap

Show Discussion

Topic: Section 4: Distance, Knn, Cross-Validation, and Generative Models / 4.2.2: Bootstrap