BIOL 220 Problem Set 02: Sample estimates and confidence intervals

Answer Key

How much caffeine is in your coffee?

The table below (not shown in answer key) contains data on the amount of caffeine in a 16 oz. cup of coffee obtained from various vendors. For context, doses of caffeine over 25 mg are enough to increase anxiety in some people, and doses over 300 to 360 mg are enough to significantly increase heart rate in most people. A can of Red Bull contains 80mg of caffeine.

Use the data table of coffee caffeine concentration to answers Questions 1–8

You can copy the data by clicking the "Copy" button. After you click the copy button, go to Google Drive and create a new google sheet. Paste the data into the google sheet and then download it as a CSV. To do this, follow the directions in Lab 3 Section 3.2.2 Creating a data file. Work all the way through the steps including uploading your new data to *Posit Cloud*

Read the data into an R session in Posit Cloud and use those data to answer the following questions.

Answer

```
caf <- read.csv("data/caffeine.csv")
```

1. What is the mean amount of caffeine in the sample of 16 oz. coffees? Round your answer to an appropriate digit. [1 point]

Answer

```
ybar <- mean(caf$caffeine_mg_16oz)
ybar
```

```
[1] 188.0643
```

2. What is the 95% confidence interval for the mean? Use the 2 SE rule and round your answer to an appropriate digit. Report your answer like "lower - upper". For example, if the confidence interval were -1.1 to 2.1, then I would report "-1.1, - 2.1". [1 point]

Answer

```
se <- sd(caf$caffeine_mg_16oz) / sqrt(nrow(caf))

ci95 <- c(ybar - 2 * se, ybar + 2 * se)
ci95
```

```
[1] 169.0485 207.0801
```

3. Is the amount of caffeine in a cup of coffee relatively consistent from one vendor to another? To help answer that, calculate the standard deviation of caffeine level and round your answer to an appropriate digit. [1 point]

Answer

```
sd(caf$caffeine_mg_16oz)
```

```
[1] 35.57535
```

With a standard deviation of 35.58, the amount of caffeine is not relatively consistent

4. The table below (not shown in answer key) has data on six 16 oz. cups of Breakfast Blend coffee sampled on six different days from a Starbucks location. Calculate the the 95% confidence interval for the mean using the 2 SE rule for these data. Report your answer in the same format as Question 2. [1 point]

Answer

```
sb <- read.csv("data/caffeineStarbucks.csv")

se <- sd(sb$caffeine_mg_16oz) / sqrt(nrow(sb))

ci95 <- mean(sb$caffeine_mg_16oz) + 2 * c(-se, se)
ci95
```

```
[1] 268.7885 475.1448
```

5. Compare these results to the data taken on the broader sample of vendors in the first file. Describe the difference in 1-2 sentences. [2 point]

Answer

The range of the 95% CI for Starbucks is substantially bigger (spanning roughly 200 mg per 16 oz) compared to the 95% CI from multiple brands (which spans roughly 40 mg per 16 oz). Two factors could contribute to this difference in the widths of the CIs: there are fewer data points for Starbucks, there could also be a genuinely huge discrepancy in how much caffine is in the Starbucks coffee.

6. Calculate the approximate 99% confidence interval for the mean caffeine level in the first table by using 2.5 SE's rather than 2 SE. Again, round your answer to an appropriate digit and report your answer in the same format as Question 2 and 4. [1 point]

Answer

```
se <- sd(caf$caffeine_mg_16oz) / sqrt(nrow(caf))

ci99 <- c(ybar - 2.5 * se, ybar + 2.5 * se)
ci99
```

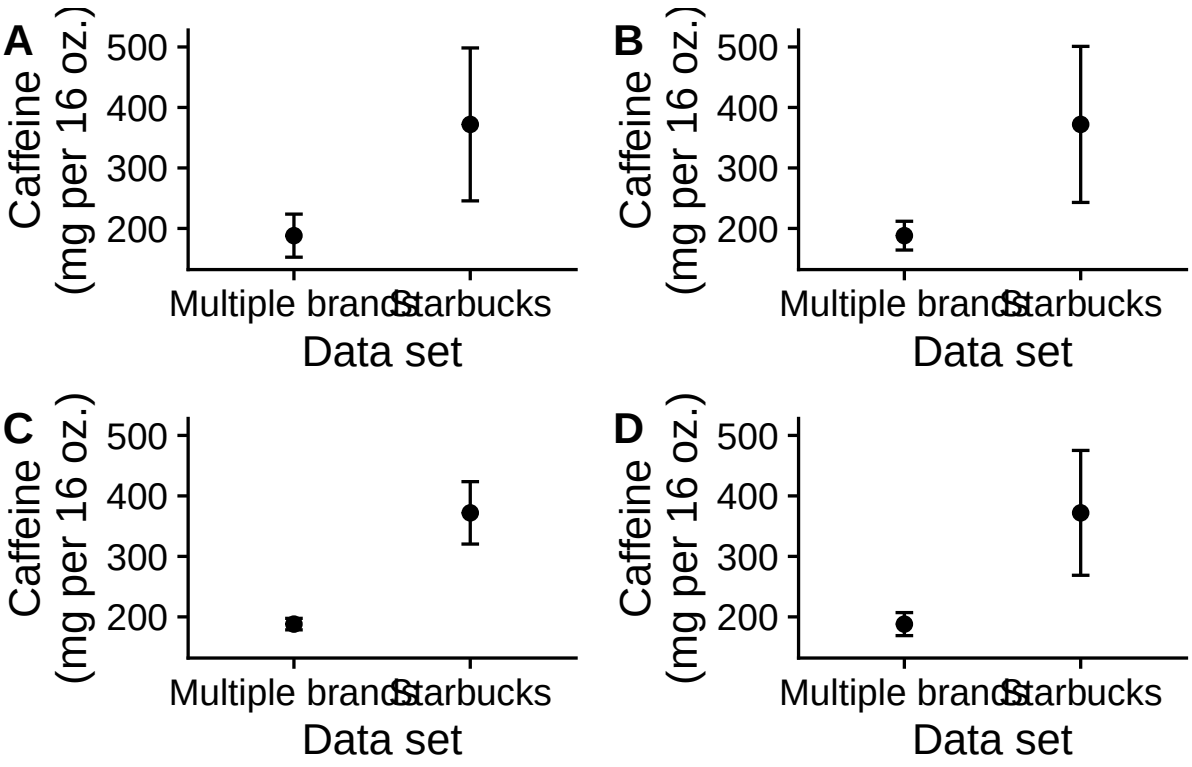
```
[1] 164.2945 211.8341
```

7. Compare the 99% confidence interval calculated in Question 6 to the 95% confidence interval you calculate in question 2. Which confidence interval is wider (i.e., spans a broader range) and why? [1 point]

Answer

The 99% CI is wider because it covers more of the sampling distribution

8. Based on the two data sets above (Multiple brands and Starbucks), which of the figures below depicts the mean caffeine concentration with error bars equal to ± 1 standard error of the mean? [1 point]



Answer

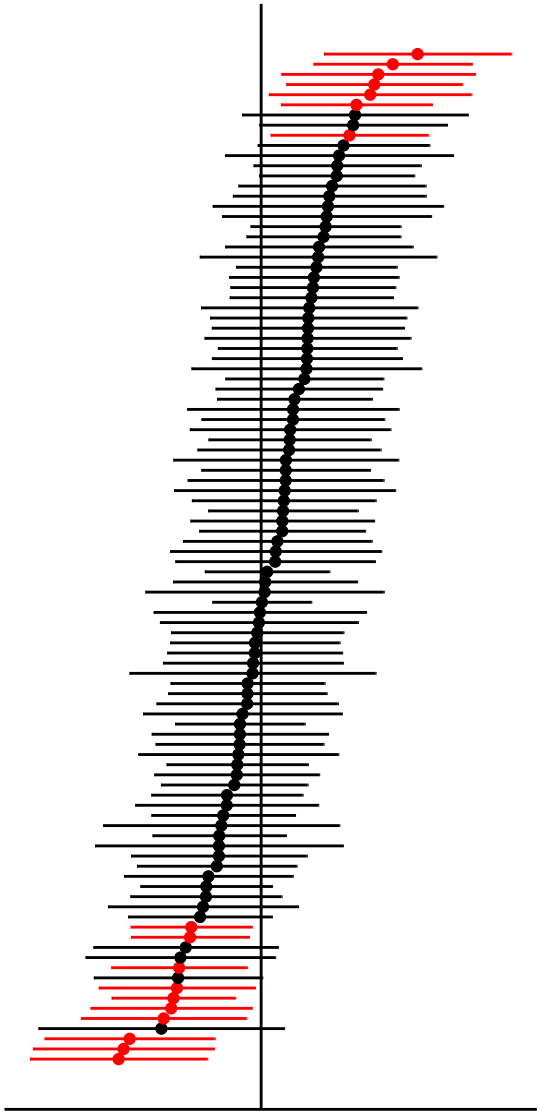
C.

Understanding confidence intervals

The figure below shows sample means (points) and associated confidence intervals of the mean (horizontal lines) from 100 different samples of the same population. In each case, the sample size is exactly $n = 16$. Assume that all samples were conducted using an identical procedure.

83 of 100 have confidence intervals that overlap the true population mean (vertical black line). These are shown with **black** points and lines. 17 of 100 have confidence intervals that do **not**

overlap the true population mean (vertical black line). These are shown with red points and lines.



Use the figure above to answer Question 9

9. Which of the following statements most likely explains what you observe in this figure? [1 point]

- These are 95% confidence intervals from 100 random samples
- These are 95% confidence intervals from 100 nonrandom samples
- These are 95% confidence intervals but some confidence intervals do not overlap

the population mean because the sample size is too small
- These are 95% confidence intervals but but some confidence intervals do not overlap the population mean because the samples are biased

Answer

- These are 95% confidence intervals but some confidence intervals do not overlap the population mean because the sample size is too small