Confidence Intervals

In this data analysis, you will investigate confidence intervals to gain a deeper understanding of how and why they're constructed. You may find it useful to review the notes from classes 9 and 11 before proceeding with this assignment.

If you haven't already done so, work through the tutorial provided on the Data Analysis 5 Canvas page. Once you've worked through the tutorial, write up your responses to the questions listed throughout the tutorial. The same questions are included below to help you format your submissions.
Submit a PDF copy of your responses to Gradescope by the deadline stated on Canvas.

Part 1: Confidence Interval for a Proportion

Question 1 (1 point)

What proportion of the adults in your sample think climate change affects their local community? Hint: Just like we did with the population, we can calculate the proportion of those in this sample who think climate change affects their local community.

0.625 of the adults in my sample think climate change affects their local community.

Question 2 (2 points)

Would you expect another student's sample proportion to be identical to yours? Would you expect it to be similar? Why or why not?

No, I would not expect another student's sample proportion to be identical to mine because they are taking a sample from the same population but with a different set of individuals. However, I would expect it to be similar because we are both sampling from the same population and the proportion of individuals who think climate change affects their local community should remain relatively constant across different samples. However, there is always some random variation in the sample, so the sample proportion may not be the same as mine.

Question 3 (2 points)

Check that the success-failure conditions needed to apply the Central Limit Theorem to the sample proportion are met in this case. Show your work.

In order to apply the CLT, we need to check if the success-failure to meet the following condition. Np >= 10 and n(1-p)>= 10

Np = 120 *0.625 = 75 which is greater than 10

N(1-p) = 120(1-0.625) = 45 which is greater than 10

Thus, success-failure conditions needed to apply the Central Limit Theorem to the sample proportion are met in this case.

Question 4 (3 points)

Suppose you instead wanted to construct a confidence interval with a different confidence level. How can you determine the necessary critical values in R for the following confidence levels? Your answers must include the R code used to find the critical value.

a. A 95% confidence interval

R code: qnorm(0.025) [1] -1.959964
qnorm (0.975,0,1) [1] 1.959964

b. An 80% confidence interval

R code: qnorm(0.9,0,1) [1] 1.281552
qnorm(0.1) [1] -1.281552

c. A 99% confidence interval

R code: qnorm (0.005) [1] -2.575829
qnorm (0.995,0,1) [1] 2.575829

Question 5 (2 points)

Using the point estimate from the sample of size n=120 you generated earlier, construct a **90%** confidence interval for the true population proportion p. Show your work.

$$0.625 + 1.645 \cdot \sqrt{\frac{0.625(1-0.625)}{120}}$$ = 0.69769

$$0.625 - 1.645 \cdot \sqrt{\frac{0.625(1-0.625)}{120}}$$ = 0.5523

Thus the 90% interval is (0.5523, 0.69769) for the true population proportion.

Question 6 (1 point)

Does your confidence interval capture the true population proportion of US adults who think climate change affects their local community, p=0.62?

Yes, my confidence interval captures the true population proportion of US adults who think climate changes affects their community, p = 0.62, since the true population proportion fall within the calculated interval (0.5523,0.69769)

Question 7 (1 point)

There are 200 students enrolled in this course. Each student should have gotten a slightly different 90% confidence interval due to random sampling. How many of 200 confidence intervals constructed do you expect to have captured the true population proportion p=0.62?

90%* 200 = 180

180 of 200 confidence intervals constructed I expect to have captured the true population proportion.

Question 8 (2 points)

What proportion of your simulated intervals captured the true population parameter p=0.62? Is this proportion what you expected? Why or why not?

0.8918*200 = 178

A proportion of 0.8918 I simulated intervals captured the true population parameter p = 0.62
From my perspective, this proportion is what I expected, because the calculated proportion of 0.8918 indicated that approximately 89.18% of these intervals captured the true population parameter p = 0.62. The expected confidence intervals are 90%, so, the calculated proportion of 0.8918 is very close to the expected value. Therefore, the simulation approach is reliable in capturing the true parameter within the 90% confidence level.

Part 2: Confidence Interval for a Mean

Question 9 (1 point)

Start by calculating the point estimate for the parameter of interest. You can do this by wrapping the mean function around the appropriate variable from the salinity data set. You'll need to use this value later on, so it is recommended that you store this value in the object called samp_mean. Report the value of point estimate, $\bar{x}$.

R code:

```
library(openair)
salinity <- read_csv(file.choose())
class(salinity$salinity_ppt)
salinity$salinity_ppt <- as.numeric(salinity$salinity_ppt)
samp_mean <- mean(salinity$salinity_ppt)
samp_mean
```

$\bar{x}$ = 38.60033

Question 10 (2 points)

We're working towards constructing a confidence interval for the true population mean. To do this, we'll need to estimate the standard error associated with the point estimate. Calculate this value. Hint: use the sd() function to determine the sample standard deviation, a necessary piece of the standard error estimate.

sample_sd <- sd(salinity$salinity_ppt)

sample_sd = 1.282957

standard_error <- sample_sd / sqrt (30)

standard_error = 0.2360423

Question 11 (2 points)

Return to the problem of interest: estimating the average salinity for the entire Bimini Lagoon. Determine the appropriate critical value that will be used to construct a 99% confidence interval. Include the numerical value of the critical value and the R code you used to determine this value.

Critical value: 2.76

R code I used to determine this value: qt (0.995, 29)   [1] 2.756386

Question 12 (2 points)

Using your answers to questions 9, 10, and 11, construct the 99% confidence interval for the true average salinity for the entire Bimini Lagoon. Report the lower and upper bounds of the interval and show your work.

Upper bound: 38.60033+2.756386*0.2360423 = 39.251   lower-bound:38.6003-2.756386*0.2360423 = 37.9497

Therefore, the interval is (37.9497,39.251)

Question 13 (2 points)

Interpret the 99% confidence interval in the context of the problem.

We are 99% confident that the true average salinity for the entire Bimini Lagoon is between 37.9497 and 39.251 with a point estimate of 38.60033

Gradescope Page Matching

When you upload your PDF file to Gradescope, you will need to match each question on this assignment to the correct pages. Video instructions for doing this are available in the Start Here module on Canvas on the page "Submitting Assignments in Gradescope". Failure to follow these instructions will result in a 2-point deduction on your assignment grade. Match this page to outline item "Gradescope Page Matching".