

Let's return to the penguins dataset in the `stat20data` package once more. For today's questions, you can consider our penguins to be a simple random sample (SRS) from the broader population of Antarctic penguins.

1. Add a column to the penguins data frame called `is_chinstrap` which is TRUE when the penguin is of the "Chinstrap" species and FALSE otherwise.
2. Use `dplyr` code to calculate the proportion of penguins that are Chinstrap and save the result into the object `point_estimate`. Write the code you used below, and write down also the value of the point estimate.
3. Visualize the `is_chinstrap` variable using an appropriate plot type. Write the code you used here below.
4. Generate one bootstrap samples of penguins (specifically, their responses to the `is_chinstrap` variable) using the `infer` library and save it into the object `bootstrap_sample`. Write the code you used below.
5. Visualize the responses in `bootstrap_sample` using an appropriate plot type. Write the code you used here below.
6. What does the visualization of the responses in `bootstrap_sample` look like in comparison to the visualization of the responses in `is_chinstrap` in **Question 3**? Explain in at least one sentence.

7. Now use the `infer` library to take 500 bootstrap samples (save this into `bootstrap_samples`). Then, calculate and plot the bootstrap sampling distribution on a histogram. Write the code you used below.

8. Describe the shape and center of the bootstrap sampling distribution in at least one sentence.

9. Out of the following three distributions: the population distribution, the one visualized in **Question 3**, and the one visualized in **Question 7**, which are observed in real life settings?

10. Using the `infer` library and `bootstrap_samples`, calculate a 95 percent bootstrapped percentile confidence interval for the proportion of all penguins on Antarctica that are Chinstrap. Write the code you used below.

11. Interpret the interval in the context of the problem in at least one sentence.

12. Tinker with the `level` argument in the `get_ci` function to calculate two more confidence intervals: a 90 percent confidence interval, and a 99 percent confidence interval. You do not need to copy the code down for this portion. Once you're done, fill out the lower and upper bounds for each of the three intervals you've calculated in the table below.

Confidence Level	Lower Bound	Upper Bound
90		
95		
99		

13. How is the center of the interval impacted (or not impacted) by changing the confidence level? What about the bounds of the interval? Answer in at least one sentence.