Z-tests and Confidence Intervals: Homework 6

NAME:

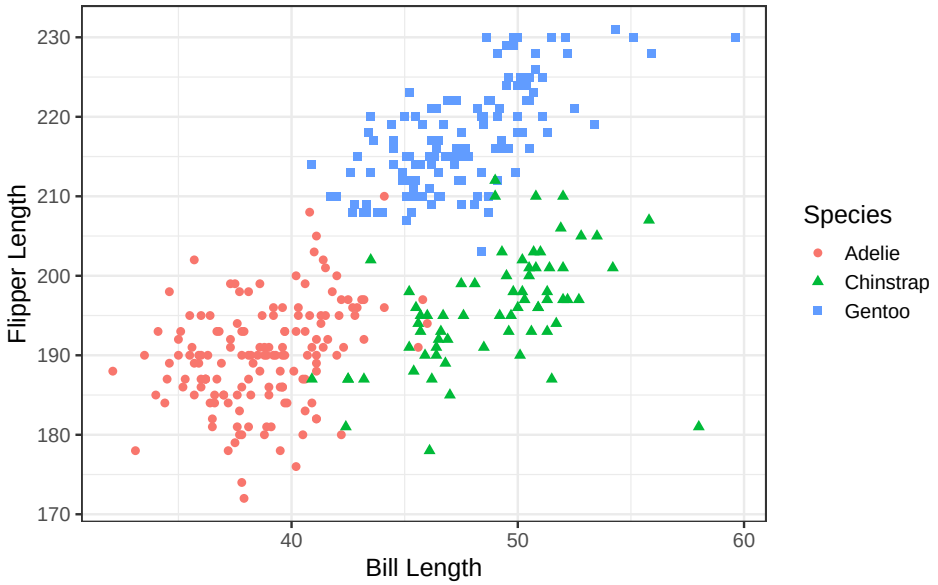
CLASS START TIME:

1 Background

We will be working with the penguin data again. In particular we are going to be running a couple z-tests and confidence intervals around the mean bill lengths for the different species of penguins.

As background, in the early 2000's 344 penguins were randomly sampled from some islands off the coast of Antarctica. The penguins came in three distinct species: Adelie (152 penguins), Chinstrap (68 penguins), and Gentoo (124 penguins). The number after the species is the number of that type of penguin in our sample.

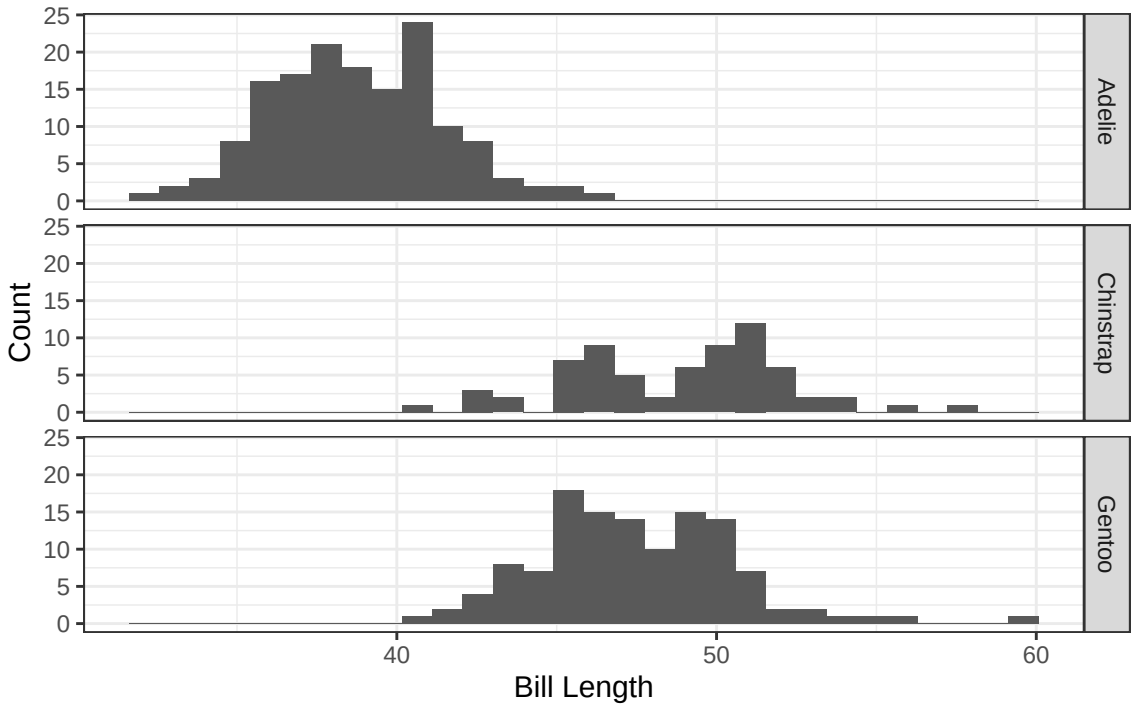
1. Please, off on the side, comment on the form, direction, strength, and if there are outliers in the below scatterplot. It may be useful to discuss the data by species.



2. Based on the above graph do you believe...
   A. **TRUE** or **FALSE**: Gentoo have the highest flipper and bill length on average
   B. **TRUE** or **FALSE**: The average bill length for the Chinstrap and Gentoo penguins are approximately the same
   C. **TRUE** or **FALSE**: Adelie penguins seem to have the shortest flipper and bill length on average.
   D. **TRUE** or **FALSE**: Species doesn't seem to have an effect on bill or flipper length.

2 Confidence Intervals

I talked with a friend who really knows penguins and was assured that, for the length of penguin's bills, the population variance, σ^2, was 10 (units are mm^2 but don't worry about that). Assume it's the same variance regardless of the species. Below are the histograms of the bill length for each species of penguin ploted as a histogram.



3. Please describe the three histograms below. Be sure to mention if they are symmetric, how many modes, and if there are outliers. Also identify if you think it is approximately normally distributed (bell shape).

4. Time for assumptions: For each of the three species of penguins, please indicate if the population distribution is normally distributed or if the sample size is sufficiently large.

5. Time for assumptions: For each of the three species, was the data randomly collected? We will tackle this in parts below.

Time for the assumptions: Was the data independent and identically distributed?

6. Is the data independent? That is, do we believe one penguins bill length should be independent of another penguins bill length?

7. Is the data identically distributed? HINT: focus on the number of modes for each species.

8. If there is reason to believe there are subpopulations within each species, what variable could be driving the discrepancies we are observing?

9. Assume we passed the assumptions. Please build a 95% confidence interval for the mean for each penguin species. Below is a table that gives you relevant information you may want.

Species	Sample Mean	Variance	Sample Size
Chinstrap	48	10	68
Gentoo	47.5	10	124
Adelie	38	10	152

10. Please interpret the confidence interval you made for the mean of the Gentoo penguins.

11. Note how some confidence intervals overlap and some do not. What do you think this implies about our ability to distingish the two groups based on their average bill length?

12. Do you think the earlier assumption about the penguins having the same population variance (spread of the data) is tenable? HINT: Use the three histograms from earlier.

3 Z-tests

Going back to my friend, he also says he knows for a fact the true bill length of chinstrap penguins is actually 48.32mm (which we take as true). Let's run a z-test to see if the mean bill length of Gentoo penguins is also 48.32mm or if it is different from that number. NOTE: We are NOT testing anything with Chinstrap penguins but instead are testing Gentoo penguins' average bill length.

13. Please write down the null and alternative hypothesis. Use μ for the population mean of Gentoo penguins since we are talking about means please.

14. Let's assume we pass the assumptions (which we already checked in the confidence interval section). Please write down the sampling distribution for the sample mean bill length for Gentoo penguins. Please use the formal notation like in the slides please.

15. Please find your test statistic (i.e. standardize your sample mean on page 3 using the above sampling distribution).

16. Using the normal calculate online, please find the p-value for this test. That is, please find the probability we would see something more extreme than what we saw if the null hypothesis is true.

17. Using the above p-value, please state your decision for this test. NOTE: If you write "reject" or "fail to reject" you are doing it wrong and will receive no credit. See my notes for a discussion.

18. Given your earlier confidence intervals, does this surprise you?

19. Does Adelie penguins have the same average bill length as Chinstrap as 48.32mm? Answer this question without using a z-test. HINT: Use the graphs from earlier.