BEE552 Biometry Week11

Maria Feiler

04/12/2022

My Learning Journey

Over the last week, I participated in Biometry in the following ways:

- I asked / answered **** questions posed in class.
- I asked **** questions in Slack.
- I answered **0** questions posed by other students on Slack.
- I came to Heather's office hours: **Yes**
- I came to Jose's office hours: **No**
- I met with Heather or Jose separately from office hours: **No**

Anything not falling into one of the above categories?

> **No**

On a scale of 1 (no knowledge) to 10 (complete expert), how would I rate my comfort with R programming after this week?

> **7**

Any topics from last week that you are still confused about?

Problem Set

The data for this problem set come from the very recent paper by Cordero et al. (2022) that has been provided. I cleaned up the raw data and have provided that to you in the file BC-measurements.xls.

```
bc <- read_xlsx("BC-measurements.xlsx")

# Cleaning column names
dimnames(bc)[[2]] <- c("Site", "BCEst", "Region", "Depth")
```

In this problem set, you will be examining differences in Black Carbon in different regions of Antarctica.

a. Considering only Regions 1 and 2, use a two sample t-test to test for a statistically significant difference in the amount of Black Carbon measured in the snow in each region. Calculate the test statistic (manually – using the formula you learned), state the distribution of the test statistic under the null hypothesis, and calculate the associated p-value. Explicitly check whether the data for these two groups conform to the equal variance assumption and be sure to use the appropriate t-test.

b. Why is this an unpaired two-sample t-test? Assuming infinite power to sample when and where you want to, how would you have designed this study to have used a paired two-sample t-test? (This will probably require reading enough of the paper to get the gist of it.)

Confirm your results using the R function "t.test". Set the options so the test matches that calculated in (A).

c. Now considering all three Regions in a single "omnibus" test for differences, complete the following ANOVA table by calculating the sum of squares manually (in other words, you can use R to do the summing, but I want you to work out the sums of squares yourself).

TABLE

d. What is the null hypothesis for the full ANOVA model? What is the alternative hypothesis? Are the data consistent with the null hypothesis?

e. What is the difference between the ANOVA test and a series of pairwise comparisons using a t-test?

f. Do these data actually meet the assumptions of ANOVA? If not, what assumption(s) are violated?

g. To review GLMs, we will re-fit this data, but with the Black Carbon measurements rounded to the nearest integer. We will examine whether Black Carbon measurements are a function of measurement depth. Fit a Poisson GLM to these data for the following model:

$$Y_i \sim Pois(\lambda_i)$$
$$log(\lambda_i) = \beta_0 + \beta_1 Depth_i$$

Include in your answer the results of the model and a plot of the data with the best-fitting model curve (and confidence interval) overlaid on top.