Biostatistics

Problem Set 3 – Introductory Statistics

25 points
Due Friday, November 15, 9:20 am.

This problem set is an opportunity for you to
**demonstrate everything you have learned about basic
statistical analysis using R.** I am looking to see that you can
apply the workflow approach presented in the book: plot your data
(and guess the results) → build the statistical model → Check the
model assumptions → Interpret the model → Make a final plot.

Figure 1. Big brown bat in flight.

If you are stuck or need a hint, feel free to reach out to me for help via Teams or in person.

Instructions:
Please complete the following tasks on your own (you may **NOT** work together with another
person or with AI on this assignment).

The data for this assignment tell a bit of a story. It goes like this:

Aria, a friend of yours, is a SLU undergraduate Biology major who got a summer research
opportunity working with a bat biologist (her position was an REU position, something you
could apply for!). Aria's project was to look at the rate of tick parasitism in 5 species of bat
native to northeastern North America. The 5 species were the big brown bat, *Eptesicus fuscus*,
the little brown bat, *Myotis* lucifugus, the silver-haired bat, *Lasionycteris noctivagans*, the
eastern pipistrelle *Perimyotis subflavus*, and the hoary bat, *Lasiurus cinereus*. With her mentor,
Aria spent the summer netting bats. They captured 25 bats of each of the 5 species and
recorded the following information about each individual bat (variable names in parentheses):
- Bats were determined to be either adults or subadults based on physical examination
  (age)
- Bats were assigned as either male or female (sex)
- Based on species, bats were assigned as either cave bats, species who migrate to caves
  and hibernate there during the winter, or tree bats, species who overwinter by
  burrowing down in the forest leaf litter and who spend the summer in tree roosts (type)
- Body mass measured in grams (mass_g)
- The number of ticks found on the wings and body of each bat (tick_num)
- Body length of each bat, in millimeters, measured from the tip of the nose to the tip of
  the tail (body_length_mm)
- Tail length of each bat, measured in millimeters from the base of the tail to the tip of
  the tail (tail_length_mm)

Biostatistics

During the school year, Aria's mentor asks her to do some data analysis with the data they collected in the summer, and her mentor would like Aria to answer the following questions from the data:

A. Does body mass differ between these five species of bats and, if so, how does body mass differ between the species?

B. Does body length differ between these five species of bats and, if so, how does body mass differ between the species?

C. Is the number of ticks found on the bats associated with their sex or age?

D. Disregarding species, is there a relationship in bats between tail length and body length?

Unfortunately, Aria has never taken biostats, and her mentor would like her to do all this analysis using RStudio following a reproducible workflow. Her mentor set up a GitHub repository with the data. Aris is at a loss about what to do, when she remembers that YOU are taking biostatistics and she reaches out to you for help. Her mentor is in support of her asking you to perform the data analysis, as long as you write up your code and results so clearly that she and Aria can understand and follow along with everything that you did and understand why you made all the choices that you did.

Here's what you need to do:

1. Fork the Github repository with Aria's data, found at https://github.com/SLU-Biostats-Fall-2024/PS3-Applying-Stats

2. Create a brand new RStudio project called "Lastname-PS3" using version control and the forked repository.

3. Inside your RStudio project, create a new .qmd file in a Code folder called Lastname-PS3.qmd. I will expect that you can now work independently on a .qmd file without me having to tell you everything you need to include. **Remember – your job is to DEMONSTRATE your ability to implement the material we have been learning in this class. That demonstration is the whole purpose of this assignment.**

4. Create the code that is necessary to analyze Aria's data. Remember to think about everything we learned BEFORE Chapter 5 of GSWR and include those steps in your workflow in addition to steps from Chapter 5.
   a. For each analysis that you perform, be sure to explain the nature of the predictor and response variables and why you selected the analysis method in your .qmd file.

5. When you have finished, commit your changes and open a pull request in the forked repository to indicate proposed changes to the project code. Then push your project to your own GitHub repository.

6. Submit your assignment by pasting the link to **your** forked GitHub repository on the assignment page on Canvas.

Biostatistics

Grading Rubric
- Successfully fork a GitHub repository and create new RStudio project from fork (1 pt)
- Set up project and workspace, pull in and examine data (2 pts)
- Analyze each question about bats, making sure to document each step in your process. At the end of each analysis, provide a more polished figure and a clear, written statement of the results of the analysis in the same way you would write it to include in a published paper. (4 points for each question – 16 points total).
- Thought processes are well documented outside of code blocks, code is well-documented inside of code blocks, and you have addressed all needed steps that should happen prior to beginning a data analysis (4 pts).
- Successfully open a pull request to add your changes to the forked repository (1 pt)
- Code represents material we have covered in Chapters 3 - 5 of GSWR and not material learned elsewhere (1 pt).

Figure 2. Hoary bat perched on a branch.