Biostatistics: Problem Set 3 – Introductory Statistics

Name: __Andrew Bazuro__   Score = __21.25__ /25

GitHub repo: __https://github.com/Andrew-Baz/Bazuro-PS3.git__

Submitted on time?   ● Y   ○ N

Project element	Value	Pts earned	Comments
Successfully fork a GitHub repository and create a new RStudio project from fork • Project called "Lastname-PS3"	1	1	good
Set up project and workspace, pull in and examine data, fix mistakes • Lastname-PS3.qmd • Use at least 2 functions • Assign data types • Error checking	2	1	You didn't do any error checking on the data.
Analyze Q1: Does body mass differ b/w these 5 species of bats, and if so, how does body mass differ b/w species? • Nature of P and R vars • Analysis method explained • More polished figure • Clear, written interpretation	4	3.25	What kind of predictor and response vars do you have and thus why do you run a particular test? I need to see the thought processes. Line 53 - outlier analysis is ok here, but should be more formal in step above including plotting histograms. Add ncol =1 to facet wrap and will show better any diffs b/w means. Guesstimates for means? Boxplot is not a good choice for final plot and needs axes repaired.
Analyze Q2: Does body length differ b/w species and, if so, how ? • Nature of P and R vars • Analysis method explained • More polished figure • Clear, written interpretation	4	3	You should not be removing these outliers, and, had you plotted the histograms, you would see that the points are reasonable. Also similar feedback to Q1 esp w/regard to final plot.
Analyze Q3: Is the number of ticks found on the bats associated with their sex or age? • Nature of P and R vars • Analysis method explained • More polished figure • Clear, written interpretation	4	4	Chi-square test - good. Better final plot.

Analyze Q4: Disregarding species, is there a relationship in bats b/w tail length and body length? • Nature of P and R vars • Analysis method explained • More polished figure • Clear, written interpretation	4	3.5	Why is regression the correct test? Line 325 - looking at the regression plot is not the best way to find outliers. Line 333 "there is a significant difference in the data" is imprecise - we are not testing for a difference in means. Also, you didn't change axis labels in final plot.
Thought processes are well documented outside of code blocks, code is well commented, all steps prior to data analysis	4	3.5	Overall good, but some mistatements or imprecise statements throughout.
Successfully open a pull request to add your changes to the forked repository • Commit changes • Open PR • Link pasted in Canvas	1	1	good
Code represents material we have covered in GSWR Chs 3-5 and not elsewhere	1	1	good
Additional feedback			