Jessica Bonin
Analysis of Environmental Data
Final Project: Part 2: Data Analysis
December 16, 2021
Worked with Juliana Berube and Julia Vineyard

Q1: Qualitatively describe the relationship between body mass and length.
Does the relationship seem linear, curved, nonexistent?

The relationship between body mass and body length seems linear. As body mass increases, body length increases.

Q2: Qualitatively describe the shapes of the histograms.
Do the data appear normally distributed? Explain why or why not.

The histogram of body mass looks normally distributed showing a bell shape and even distribution. The histogram of the body length looks slightly skewed to the left with the "hump" to the right of the middle of the center.

Q3: Using both the histograms and normality tests, do you think the (unconditioned) body masses and body length are normally distributed? Contrast your visual assessment of normality to the results of the numerical normality tests.

I do not think the data are normally distributed because for both body mass and body length, the shapiro test resulted in a p-value less than 0.05. If the p-value is less than 0.05 then the null hypothesis that the data are normally distributed is REJECTED. This contrasts with what I thought based on the histogram of body mass but makes sense for my assessment of the histogram of body length.

Q4: Examine the conditional boxplots. Describe any graphical evidence you see for body mass differences based on species and/or sex.

The boxplots show that dorsalis has a higher body mass than sublineatus overall and that males have a higher body mass than females overall. The Boxplot conditioned by both factors makes it very clear that males are heavier than females for both species.

Q5: What do you conclude about residual normality based on the numerical and graphical diagnostics?

While the histograms look like they are normally distributed, all of the shapiro test results resulted in a p-value less than 0.05. If the p-value is less than 0.05 then the null hypothesis that the data are normally distributed is REJECTED. I would conclude that the residuals are not normally distributed.

Q6: Are violations of the normality assumption equally severe for all the models?

The p-value from the shapiro test for fit2 is much less negative than the others. This means that the distribution is closer to normal (although it does not mean it is normal).

Q7: What is the magnitude of the mass/length relationship?

0.8754988 (the slope)

Q8: What is the expected body length of an animal that weighs 100g?

163.6745

Q9: What is the expected body length of an animal that weighs 0g?

76.1246565

Q10: What is the base level for sex?

female

Q11: What is the base level for binomial?

Delomys dorsalis

Q12: Which sex is heavier?

male

Q13: Which species is heavier?

Delomys dorsalis

Q14: Are sex and species significant predictors for body mass?

Yes, both the p-values are less than 0.05 for the anova for fit4.

Q15: Is there a significant interaction?

No, the p-value is larger than 0.05 for the anova for fit5.

Q16: How does the significance (as measured by p-value) of either of the main effects (sex and species) differ between the single-predictor models, the additive model, and the interactive model?

The single-predictor and additive models are significant (p-value lower than 0.05), but the interactive model is not (p-value greater than 0.05).

Q17: Which two models have the lowest AIC?

Except for fit1, fit4 and fit5 have the lowest AIC.

Q18: Which of the two models with lowest AIC scores would you select?
Explain your decision based on model fit and the complexity/understanding tradeoff.

I think fit4 is the best model of body mass. It has the lowest AIC and is easier to interpret than the fit5 model. Although fit5 is more complex, it has more components making it more complicated to interpret.