Keegan Moynahan

Data Analysis Final

1. Looking at the scatterplot between body mass and body length it seems as if the data is linear with some outliers. It appears that as the body length increases the body mass will also increase, the scatterplot represents a positive correlation.
2. The histogram of body mass seems to be slightly skewed towards the left and is unimodal. This histogram shows that the data has more large body mass delomys than small body mass delomys. The histogram of body length is skewed left with a noticeable outlier (longer than rest of data). This data is also unimodal and seems like the data has longer delomys than shorter delomys.
3. After viewing the histograms and not seeing much normal distribution and the running the Shapiro test to observe the p-value for both body mass (4.33e-05) and body length (< 2.2e-16). Where both p-values are small enough to reject the null hypothesis (data is normally distributed). It is safe to say the data isn’t normally distributed.
4. When first viewing the boxplot of body mass conditioned on sex there isn’t a huge difference between male and female, but there is a little difference. It shows that male delomys tend to have a greater body mass than female delomys. When looking at the boxplot of body mass conditioned on species, one species has a greater body mass than the other. Delomys dorsalis is clearly heavier than delomys sublineatus when viewing the boxplot.
5. I would say that all the residuals do not meet the normality criteria from looking at the histograms and numeric exploration. The residuals of fit2 had the best histogram and p-value (0.0001541) which is by far the least significant p-value.
6. 1 p-value < 2.2e-16, 2 p-value = 0.0001541, 3 p-value = 8.103e-05, 4 p-value = 6.805e-05, 5 p-value = 6.816e-05. These were the p-values for each residual test of normality. As you can see fit4 and fit5 residuals had an equally severe violation, while fit 3 had the most severe, fit1 had the second-best p-value and fit2 had the best p-value.
7. 0.8754988 g
8. 0.8754988\*100 + 76.1246565 = 163.6745365 g
9. 76.1246565 g
10. Base level is female
11. Delomys Dorsalis is the base level
12. Males are heavier than females because the coefficient table shows that for every unit (gram) of male body mass you add 3.795395 to the intercept which is female delomy dorsalis.
13. Delomys Dorsalis is heavier than Delomys Sublineatus because the coefficient table shows that for every unit (gram) of delomy sublineatus female you subtract 8.193492 from the intercept which is female delomy dorsalis.
14. I would say that sex is a significant predictor of body mass. It can be seen throughout my fit models anova tables where fit 2 p-value = 0.0001951.
15. There doesn’t seem to be major significance between sex and binomial because in the anova table for fit5 the p-value is 0.9504424.
16. In fit model 2 the only value with a p-value is sex which has a consistent p-value of 0.0001951. In fit model 3 there is a p-value of 0. In fit model 4 there seems to be slight difference in the p-value for sex (p-value =0.0001144), for binomial the p-value stays at 0. In fit model 5 the p-value for sex changes to 0.0001150, again a slight difference but not much and the p-value for binomial stays at 0. This helps show that sex has a bigger impact on body mass than binomial.
17. The two models with AIC are fit4 (12896.73) and fit5 (12898.72).
18. I would go with fit4 as my model because it is less complex than fit5 and has a better AIC score than fit5, but it also includes the four different criteria sex (male & female) & species (delomy dorsalis & delomy sublineatus) in a better way. This model would be easier to display and explain than fit5. In the end both models are good models but fit5 is less complex and has the lower AIC value. Fit4 creates the best fit for the models (AIC) and is also the easiest to understand out of the two.