

Kaitlyn Valentinetti

November 9, 2022

ECo 634: Michael France Nelson

Lab 9: Modeling Data II

Q1 (1 pt.): State the null hypothesis of the Chi-square test.

The null hypothesis is that there is no association between the presence/absence of Brown creepers and edge/interior habitats.

Q2 (2 pts.): Consider the results of your test and explain whether you think that Brown Creepers show a significant habitat preference.

The results of my Chi-squared test yield a p-value of 1.386e-06, so it is suggested that Brown Creepers show a significant habitat preference. Since a p-value of less than or equal to 0.05 is considered statistically significant, our p-value (1.386e-06) fits this assumption and demonstrates a rejection of the null hypothesis.

Q3 (1 pt.): Show the R-code you can use to create a model fit (call it fit\_species) of penguin body mass as predicted by penguin species.

```
fit_species =  
  lm(  
    formula = body_mass_g ~ species,  
    data = penguins)
```

Q4 (1 pt.): Show the R-code you can use to create a model fit (call it fit\_sex) of penguin body mass as predicted by sex.

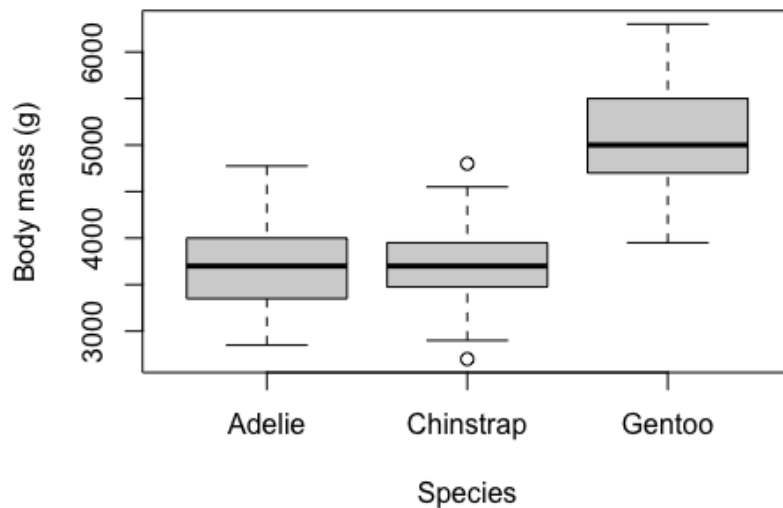
```
fit_sex =  
  lm(  
    formula = body_mass_g ~ sex,  
    data = penguins)
```

Q5 (1 pt.): Show the R-code you can use to create a model fit (call it `fit_both`) of penguin body mass as predicted by species and sex.

```
fit_both =  
lm(  
  formula = body_mass_g ~ sex:species,  
  data = penguins)
```

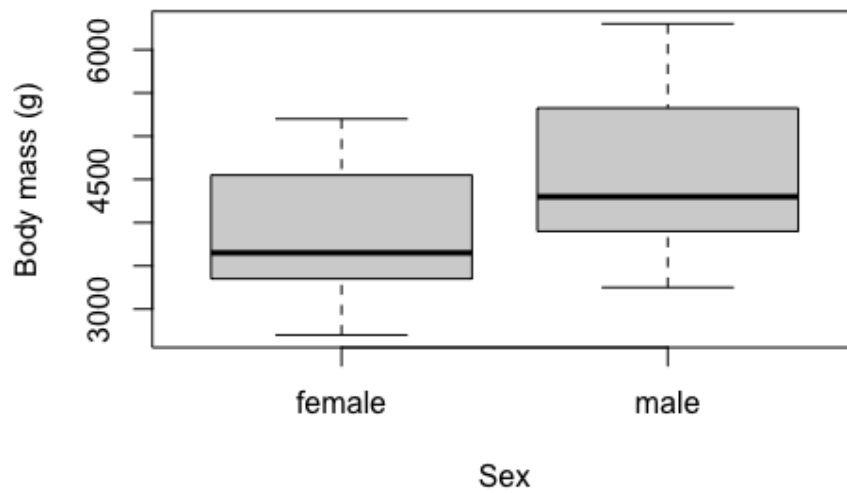
Q6 (1 pt.): Include a conditional boxplot corresponding to the grouping structure in your `fit_species` model.

### Conditional Plot of Penguin Species and Body Mass



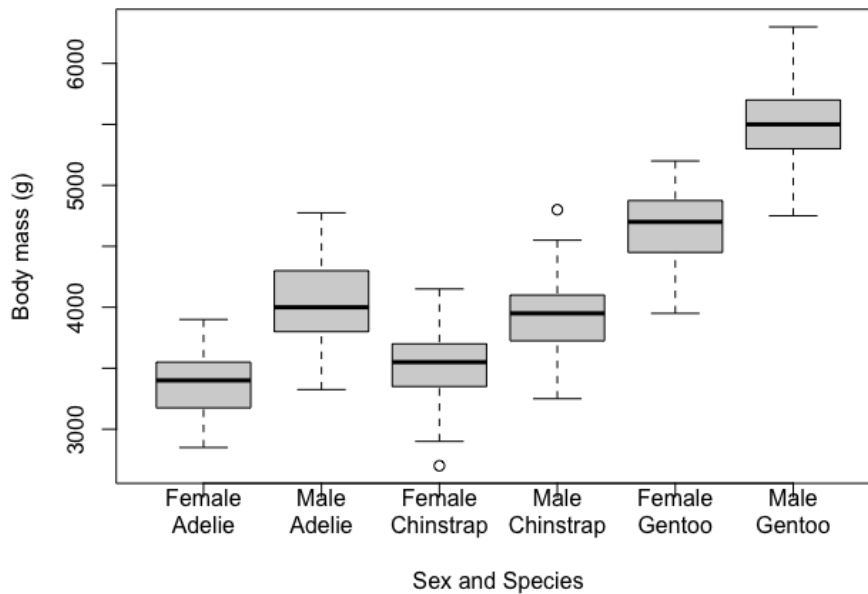
Q7 (1 pt.): Include a conditional boxplot corresponding to the grouping structure in your `fit_sex` model.

### Conditional Plot of Penguin Sex and Body Mass



Q8 (3 pts.): Include a conditional boxplot corresponding to the grouping structure in your fit\_both model.

### Double Conditioned Boxplot of Species and Sex



Q9 (3 pts.): Based on the shapes of the boxes, which of the models (if any) do you think may have problems fulfilling the homogeneity assumption?

Based on the shapes of the boxes, the conditional plot for species fit and the double conditional boxplot may have problems fulfilling the homogeneity assumption because the size of the boxes varies. If the boxes appear different sizes, then they also have different variances, so I would think the homogeneity assumption would be problematic in these models.

Q10 (1 pt.): State the null hypothesis of the Bartlett test.

The null hypothesis of the Bartlett test is that the variance in all species and sex is homogeneous.

Q11 (1 pt.): What was the p-value from the Bartlett test of homogeneity for observations grouped by species?

You can round your answer to 4 decimal digits.

p-value = 0.0501

Q12 (1 pt.): What was the p-value from the Bartlett test of homogeneity for observations grouped by sex?

You can round your answer to 4 decimal digits.

p-value = 0.0319

Q13 (1 pt.): What was the p-value from the Bartlett test of homogeneity for observations grouped by both factors?

You can round your answer to 4 decimal digits.

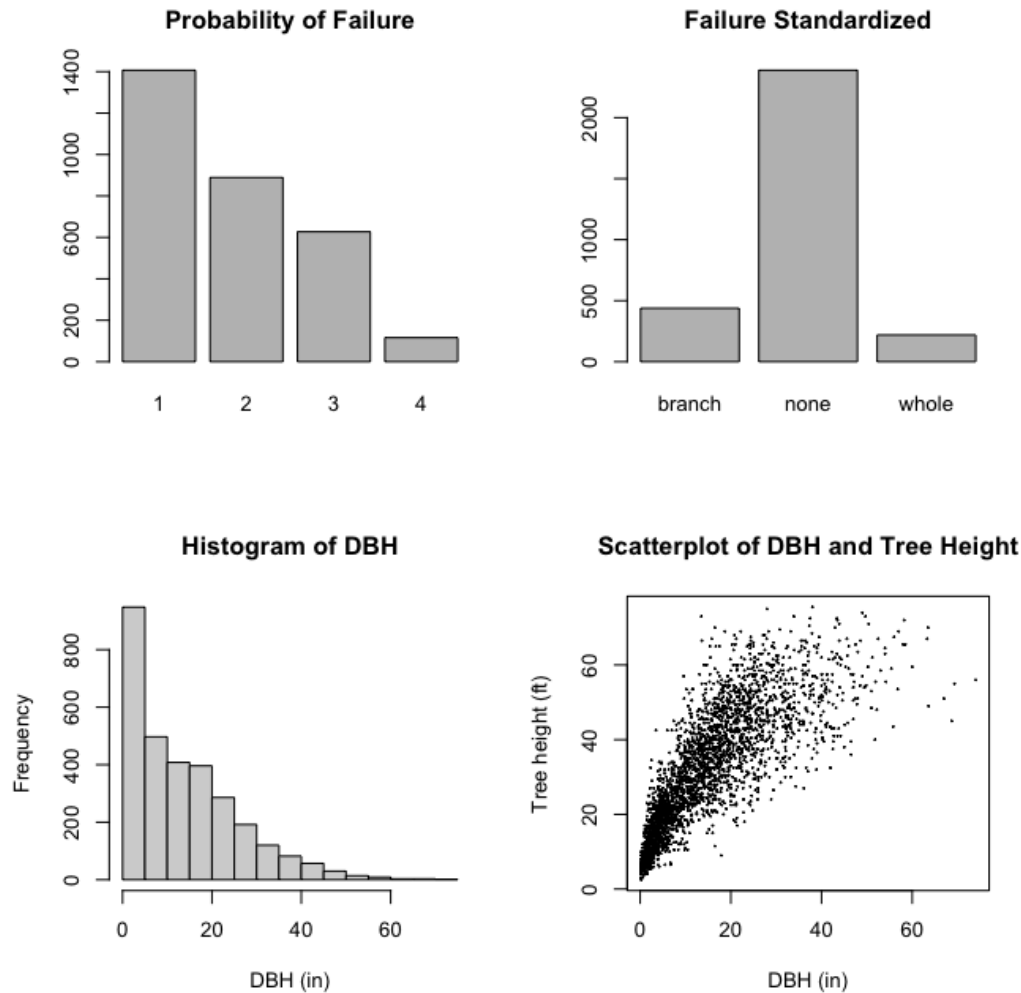
p-value = 0.1741

Q14 (3 pts.): Based on the results of the Bartlett tests, do you anticipate any issues with heterogeneity in any of the models?

Make sure you justify your response with the results of your tests.

Based on the results of the Bartlett tests, there will be issues of heterogeneity with the model grouped by sex (p-value = 0.0319). Since the null hypothesis of the Bartlett test supports homogeneity of variance, if the p-value is less than 0.05, there is support against homogeneity. With a p-value of 0.0319 for the model grouped by sex, the null hypothesis of homogeneity is rejected.

Q15 (5 pts.): Perform a graphical exploration of the dataset. Create the following plots and include them in your report. You may create separate figures or combine them into one multi-panel figure.



Q16 (1 pt.): State the null hypothesis for the Kolmogorov-Smirnov test. Your answer should be in terms of the DBH of the two groups of trees.

The null hypothesis for the Kolmogorov-Smirnov test is that the DBH of the two groups of trees is from the same sampling distribution.

Q17 (1 pt.): What was the p-value of the test? Based on the evidence, do you think the distribution of DBH is the same for the two groups?

The p-value of the test is 0.02125. Based on the p-value, I don't think that the distribution of DBH is the same for the two groups because the p-value is less than 0.05.

Q18 (1 pt.): Qualitatively describe the shape of the relationship between DBH and height. Is it linear? Curved? Monotonic?

The shape of the scatterplot demonstrating the relationship between DBH and height is slightly curved and resembles a megaphone. The correlation therefore is weaker as the DBH increases.

Q19 (1 pt.): Given your answer to the previous question, which type of correlation coefficient is most appropriate?

Given the previous answer, the Spearman correlation coefficient is most appropriate.

Q20 (1 pt.): What is the p-value? Do you conclude that the two variables are significantly correlated?

The p-value is  $< 2.2e-16$ . Yes, I do conclude the two variables are significantly correlated.

Q21 (2 pts.): What was the value of the test statistic (X-squared)? What was the corresponding p-value?

The value of the test statistic is 202.65, and the corresponding p-value is  $< 2.2e-16$ .

Q22 (1 pt.): What is the value of the chi-square residual (rounded to the nearest whole number) for the count of failures in probability category 1?

The chi-squared residual for the count of failures in probability category #1 is -136.

Q23 (1 pt.): Were there more, or fewer, tree failures than expected by chance in failure probability category #1?

There were fewer tree failures than expected by chance in the failure probability category #1.

Q24 (1 pt.): Were there more, or fewer, tree failures than expected by chance in failure probability category #4?

There were more tree failures than expected by chance in failure probability category #4.

Q25 (2 pts.): Given your answers to the previous two questions, do you conclude that the probability of failure rating system is effective?

I would conclude that the probability of failure rating system is ineffective because the answers from the previous two questions are inconsistent.