Analysis of Environmental Data – Lab 9
Olivia Dinkelacker

Q1 (1 pt.): State the null hypothesis of the Chi-square test.
Make sure you state the null hypothesis in terms of Brown Creeper presence/absence and edge/interior habitats.

There is no difference in the frequency of presence and absence of Brown creepers in edge and interior habitats

Q2 (2 pts.): Consider the results of your test and explain whether you think that Brown Creepers show a significant habitat preference.
Make sure your use the output of your statistical test to support your answer.

We reject the null hypothesis, because there is a significant preference for habitat type, since the p-value is smaller that 0.05 (p-value = 1.386e-06).

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 = penguins$body_mass_g ~ penguins$species,
    data = penguins)
```

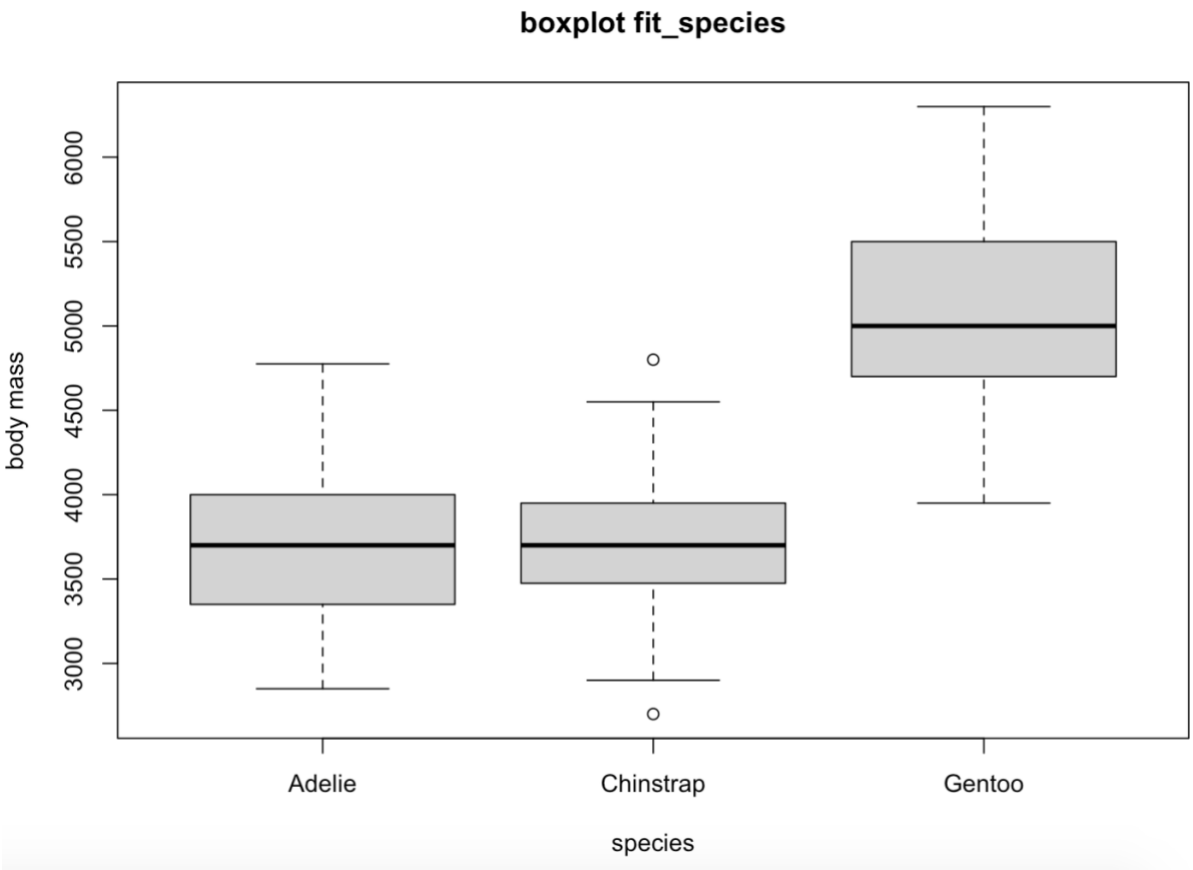
Q4 (1 pt.): Show the R-code you can use to create a model fit (call it `fit_sex`) of penguin **body mass** as predicted by **sex**.

```
fit_sex =  lm(
  formula = penguins$body_mass_g ~ penguins$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**. This should be an *interactive* model, i.e. it should include a sex:species interaction.

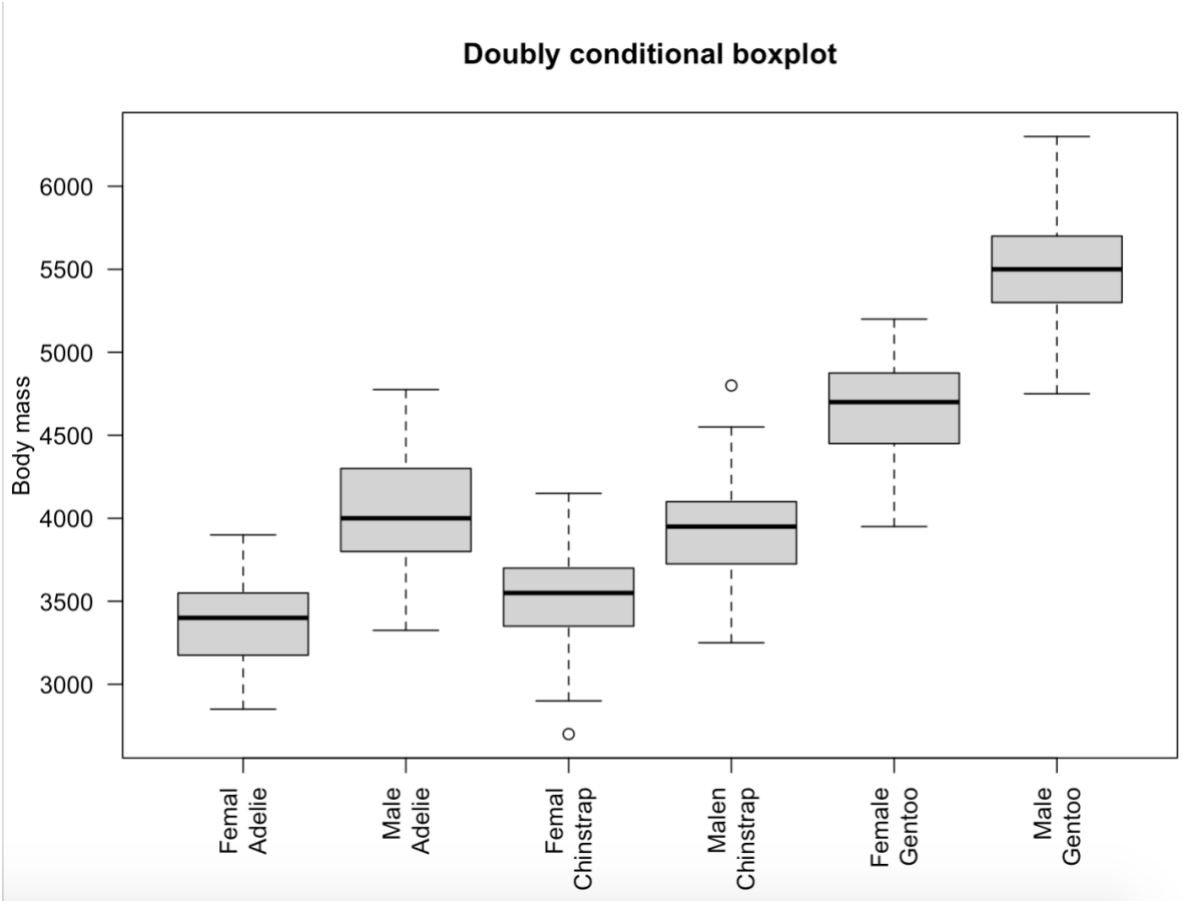
```
fit_both =  lm(
  formula = penguins$body_mass_g ~ species:sex,
  data = penguins)
```

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



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

Q8 (3 pts.): Include a conditional boxplot corresponding to the grouping structure in your `fit_both` model.
Your group labels must all correspond to the correct box, be visible, and sensible.



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?

The fit_species and fit_both models may have problems fulfilling the homogeneity assumption, since the variability differs between variables. This can be inferred by the different length/shapes of the plots. Fit_sex seems to be similar in shape.

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

There is homogeneity of variances among all treatment levels (species and sex).

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 = 0.0505

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

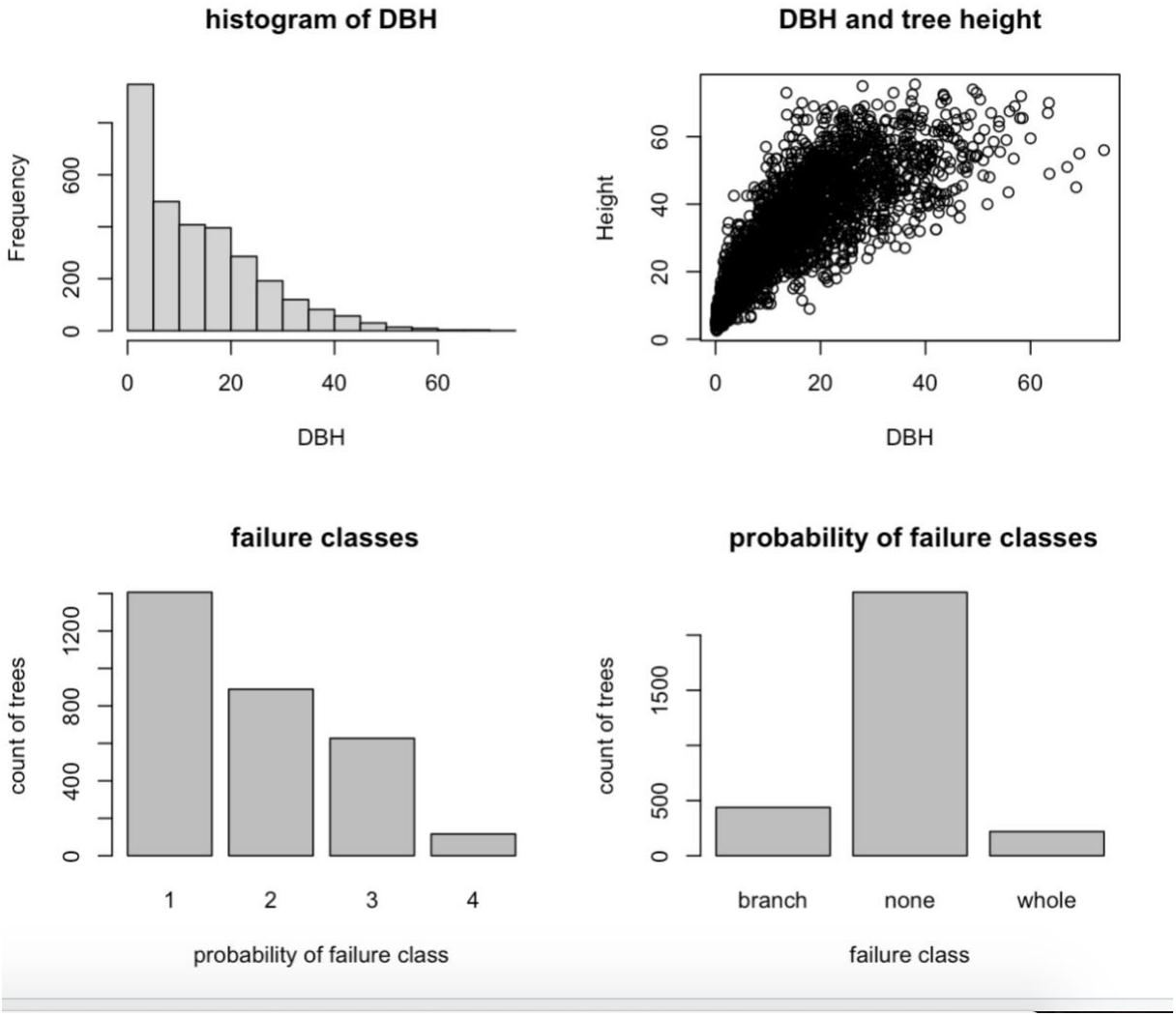
Q13 (1 pt.): What was the p-value from the Bartlett test of homogeneity for observations grouped by both factors?
You can round your answer to 4 decimal digits.

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.

Only for sex there is no homogeneity in variances, since the p-value is less than 0.05.

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.
A barplot of counts of trees in each probability of failure class (column `ProbabilityofFailure`.
A barplot of the counts of trees in each of the failure classes (column `Failure_Standardized`)
A histogram of DBH
A scatterplot of DBH (x-axis) and tree height (y axis)



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 DBH from the two groups of trees are drawn from the same continuous 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?

p-value = 0.02125
Based on the test the two group of trees are not drawn from the same distribution, since 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 relationship is positive linear and curved. It has a megaphone shape meaning that the correlation is getting less.

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

A correlation coefficient around 1 would be most appropriate because it appears to be a positive relationship.

Q20 (1 pt.): What is the p-value? Do you conclude that the two variables are significantly correlated?
p-value < 2.2e-16

There is a significant correlation because the p-value is very small.

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

X-squared = 202.65, df = 3, p-value < 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?
-136

Q23 (1 pt.): Were there more, or fewer, tree failures than expected by chance in failure probability category #1?
There were fewer failures than expected in 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 failures than expected in 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 not very effective, since the observed values vary a lot from the expected values.