

Matthew Jusino

Lab 10: ANOVA

- **Q1 (8 pts.):** Submit the code you used to build your ANOVA by hand. Make sure you use the code template so that you use the same variable names as those which we'll use for the grading.

```
rm(list = ls())
```

```
rope = read.csv(here("data", "rope.csv"))
```

```
rope$rope.type = factor(rope$rope.type)
```

```
levels(rope$rope.type)
```

```
rope$rope.type
```

```
summary(rope$rope.type)
```

```
n_obs = 121
```

```
n_groups = 6
```

```
boxplot(p.cut ~ rope.type, data = rope)
```

```
grand_mean = mean(rope$p.cut)
```

```
grand_resids = rope$p.cut - grand_mean
```

```
sum_grand_resids = sum( (rope$p.cut - grand_mean)^2)
```

```
ss_tot = sum_grand_resids
```

```
df_total = n_obs - 1
```

```
aggregate(
```

```
  x = rope$p.cut,
```

```
  by = list(rope$rope.type),
```

```
  FUN = function (x) mean(x))
```

```
agg_resids = aggregate(  
  x = rope$p.cut,  
  by = list(rope$rope.type),  
  FUN = function(x){  
    (x - mean(x))  
  })  
str(agg_resids)
```

```
agg_sum_sq_resids = aggregate(  
  x = rope$p.cut,  
  by = list(rope$rope.type),  
  FUN = function (x){  
    sum( (x - mean(x))^2)  
  })  
str(agg_sum_sq_resids)
```

```
agg_sum_sq_resids$x  
ss_within = sum(agg_sum_sq_resids$x)  
ss_within  
df_within = n_obs - n_groups
```

```
ss_among = ss_tot - ss_within  
df_among = n_groups - 1
```

```
ms_among = ss_among / df_among  
ms_within = ss_within / df_within
```

```
f_ratio = ms_among / ms_within  
f_pvalue = pf(f_ratio, df_among, df_within, lower.tail = FALSE)
```

- **Q2 (1 pt.):** Examine the conditional boxplot in the *Partitioning Variance: Within-Group* section of the walkthrough. Based on the figure, do you think there are equal variances among the groups?

No, I do not. They all seem to have different variances, especially stark in the contrast between BLAZE and PI rope types.

- **Q3 (1 pt.):** Conduct a Bartlett test to assess the homogeneity of variances of the percent cut among the rope type groups.

```
bartlett.test(rope$p.cut, g = rope$rope.type)
```

data: rope\$p.cut and rope\$rope.type

Bartlett's K-squared = 19.687, df = 5,

p-value = 0.00143

- **Q4 (2 pts.):** Given your graphical assessment (question 2) and the Bartlett test, do you think an ANOVA-type analysis is appropriate on the raw data? Explain why or why not.

I do think that an ANOVA-type analysis is appropriate for the raw data because ANOVA's are good for comparing the means and differences, and the low p-value in the Bartlett test leads to a conclusion that there is homogeneity in the variances.

- **Q5 (1 pt.):** Which rope type is the base case?

The best case is the BLAZE rope

- **Q6 (1 pt.):** What is the mean percent cut of the base case rope? Show your calculation using value(s) from the model coefficient table.

The mean percent cut of the base case rope is 36.7%.

0.36714×100

- **Q7 (1 pt.):** What is the mean percent cut rope type XTC? Show your calculation using value(s) from the model coefficient table.

The mean percent cut of rope type XTC is 26.55%

$(0.36714 - 0.10164) \times 100$

- **Q8 (1 pt.):** Use the `residuals()` function to retrieve the residuals from your model and perform an overall normality test. Report the p-value.

p-value = 7.238e-07

- **Q9 (1 pt.):** Do your model residuals meet the normality assumption, and how do you know?

The residuals do not meet the normality assumption, because the p-value is significantly less than the threshold of 0.05

- **Q10 (4 pts.):** Perform normality tests on the residuals within each group. How many groups meet the normality assumption?

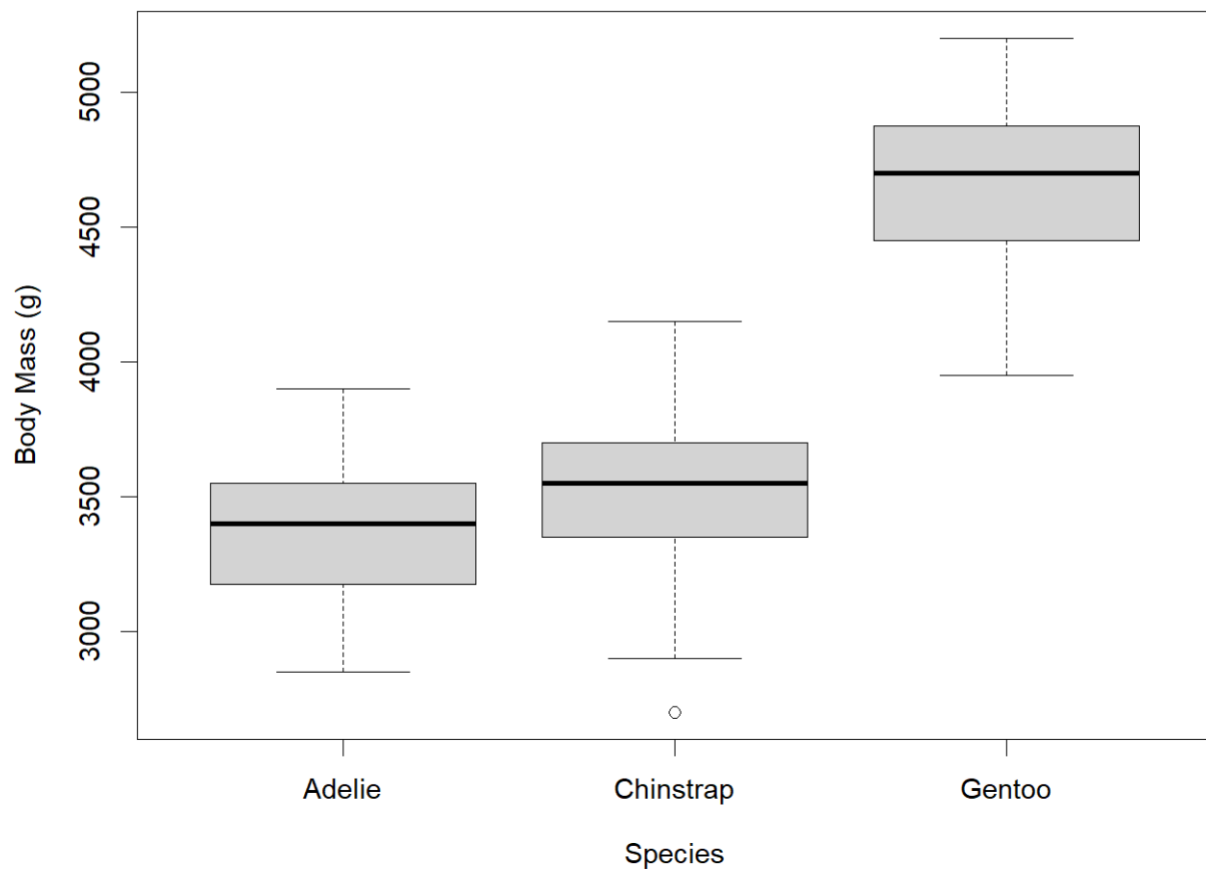
After performing the normality test on all six groups, no rope type meets the normality assumption. They range from a p-value of 1.228e-07 to a p-value of 9.508e-09

- Optional challenge: identify which rope types meet the assumption.
- **Q11 (1 pt.):** Given the results of your tests for residual normality, do you think that a one-way Analysis of Variance is appropriate for this dataset?

No, I do not believe that a one-way ANOVA would be appropriate for this dataset, as a one-way ANOVA relies on equal variance, which we clearly do not have.

- **Q12 (2 pts.):** Create a conditional boxplot of the female penguins: body mass conditioned on species.

Female Penguins Body Mass vs Species Comparison



- **Q13 (1 pt.):** Based on the boxplot, do you anticipate any problems with residual normality, or homogeneity of variances? Why or why not?

I do, because the Chinstrap species appears to have slightly longer tails than the Adelie species, and the Gentoo species has unequal tail lengths, leading me to believe that they will not have homogeneous variances.

- **Q14 (2 pts.):** Conduct a Bartlett test for homogeneity of variances of body mass grouped by species. Hint: use the formula notation. Report the p-value. Is the homogeneity assumption met? Why or why not?

```
bartlett.test(fpens$body_mass_g ~ fpens$species)
```

p-value = 0.9056

The homogeneity assumption is not met, because the p-value is significantly higher than the threshold of 0.05

- **Q15 (2 pts.):** Fit a linear model of body mass (the response) and species (the predictor) using the female penguin data. Conduct a test for normality of the residuals. Report the p-value. Is the residual normality assumption met? Why or why not?

The p-value is 0.3639

Thus the normality assumption is met, because the p-value is significantly larger than the threshold of 0.05 (which would suggest the residuals do not come from a normal distribution).

- **Q16 (2 pts.):** Conduct a Tukey HSD post-hoc test on your model. Which pair or pairs of species have significantly different body masses?

(Gentoo and Adelie) and (Gentoo and Chinstrap) both have body mass differences of over 1100 g.

- **Q17 (2 pts.):** Describe how your HSD test results match, or do not match, the graphical insight from the conditional boxplot.

The HSD results do match my graphical insights, because I predicted issues with the homogeneity of variance, which the assumption for is not met. Though normality is assumed based on the results of the Shapiro test, which my graphical insights did not align with.