Analysis of Environmental Data - Lab 10
Olivia Dinkelacker

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.

```
rope =
read.csv("/Users/oliviadinkelacker/Documents/ECo/environmental_data/data/rope.csv")
rope$rope.type = factor(rope$rope.type)
levels(rope$rope.type)

n_obs = nrow(rope)
n_obs
n_groups = length(unique(rope$rope.type))
n_groups

ss_tot = sum(resids^2)
ss_tot
grandmean = mean(rope$p.cut)
obs = rope$p.cut
resids = obs - grandmean
df_tot = n_obs-1

agg_mean = aggregate(
  x = rope$p.cut,
  by = list(rope$rope.type),
  FUN = mean)

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(y) y-mean(y))

str(agg_resids)

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

str(agg_sum_sq_resids)
```

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

#Partitioning Variance: Among Groups
ss_among = ss_tot - ss_within
ss_among
df_among = n_groups-1
df_among

ms_among  =  ss_among / (n_groups - 1)
ms_within = ss_within / (n_obs - n_groups)
ms_among
ms_within

#The Test Statistic: F
f_ratio = ms_among /ms_within
f_ratio

f_pval = 1-pf(f_ratio,df_among,df_within)
f_pval
```

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?

Because the plots differ in sizes there is no homogeneity in variances.

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

```
> bartlett.test(p.cut~rope.type, data = rope)

        Bartlett test of homogeneity of variances

data:  p.cut by 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.

It is not appropriate, because the assumption for homogeneity in variances is not met. You would need a group 2 model here.

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

BLAZE is the base case.

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.

We are just reading the Estimate for the Intercept: 0.36714.

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

We are just subtracting the XTC from the Estimate of the intercept . 0.2655

Q8 (2 pts.): Use the `residuals()` function to retrieve the residuals from your model and perform an overall normality test. Report the p-value.

```
shapiro.test(residuals(fit_rope_1))

        Shapiro-Wilk normality test

data:  residuals(fit_rope_1)
W = 0.91144, p-value = 7.238e-07
```

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

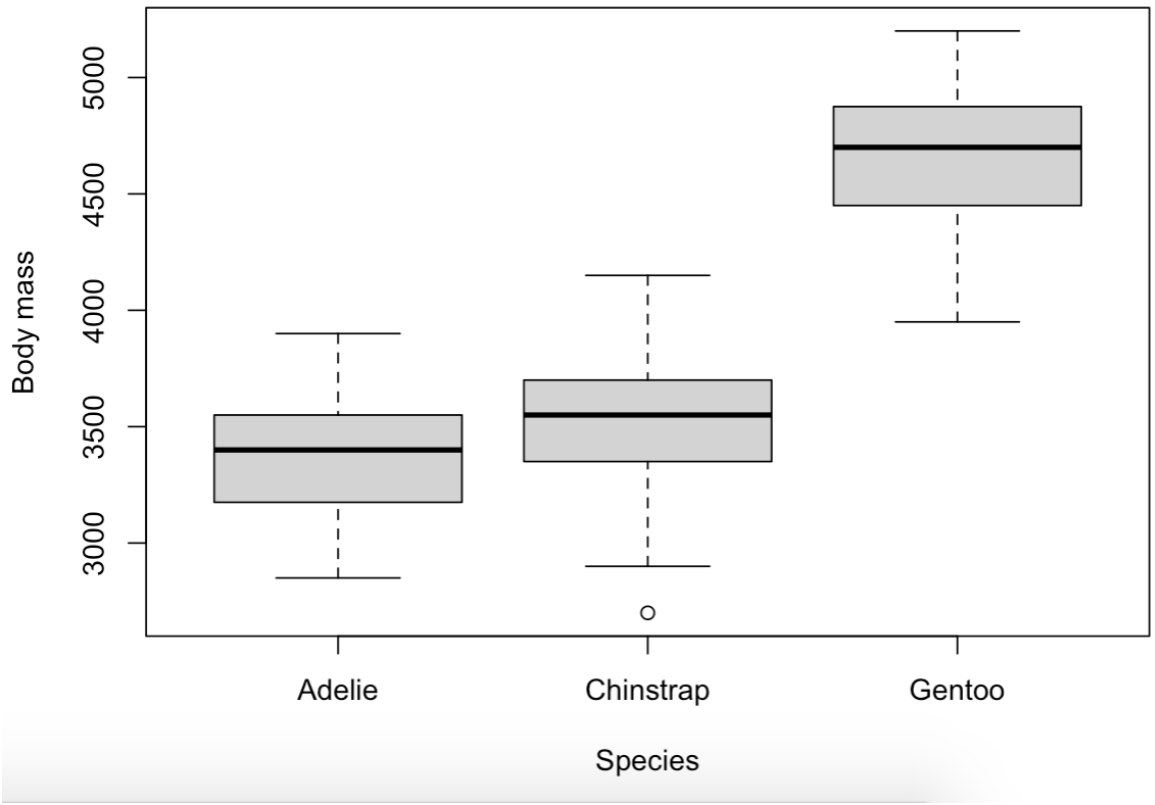
They do not meet the normality assumption since the p-value is less than 0.05.

Q10 (4 pts.): Perform normality tests on the residuals within each group. How many groups meet the normality assumption?
Optional challenge: identify which rope types meet the assumption.

```
Three groups meet the normality assumption:
[1] 0.0005471159
[[2]]
[1] 0.1127864
[[3]]
[1] 0.1289522
[[4]]
[1] 0.07460241
[[5]]
[1] 0.04439502
[[6]]
[1] 0.02306614
```

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?
I think it is not appropriate since the normality assumption is not met in every group.

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



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 not anticipate problems with homogeneity since the shapes of the plots look quite similar, indicating no variance in homogeneity.

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?
p-value = 0.9056
There is homogeneity in variances. The assumptions are met, since the p-value is higher than 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?
p-value = 0.3639
The residual normality function is met, since the p-value is higher than 0.05.

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

	diff	lwr	upr	p adj
Chinstrap-Adelie	158.3703	22.32078	294.4197	0.0179471
Gentoo-Adelie	1310.9058	1195.64908	1426.1624	0.0000000
Gentoo-Chinstrap	1152.5355	1011.00620	1294.0648	0.0000000

All pairs of species have significant difference in body mass.

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

Based on the boxplot you would think that Adelie and Chinstrap are not significantly different in body mass. The difference to Gentoo is not surprising based on the boxplot.