Keegan Moynahan　　　　**** Worked on with Steph ****

Lab 10

1)

```
require(here)
rm(list = ls())


rope = read.csv(here("data", "rope.csv"))
rope$rope.type = factor(rope$rope.type)
## check##
length(levels(rope$rope.type))
##


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


grand_mean = mean(rope$p.cut)
resid = (rope$p.cut - grand_mean)
ss_tot = sum(resid^2)
df_tot = n_obs - 1


test_agg = aggregate(
  x = rope$p.cut,
  by = list(rope$rope.type),
  FUN = function(x) mean(x))
agg_resid = aggregate(
  x = rope$p.cut,
  by = list(rope$rope.type),
  FUN = function(x) x - mean(x))
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)

ss_within = sum(agg_sum_sq_resids$x)

df_within = n_obs - 6


ss_among = ss_tot - ss_within

df_among = n_groups - 1


ms_within = ss_within / (n_obs - n_groups)

ms_among  = ss_among / (n_groups - 1)


f_ratio = ms_among / ms_within

f_pval = 1 - pf(f_ratio, df_among, df_within)
```
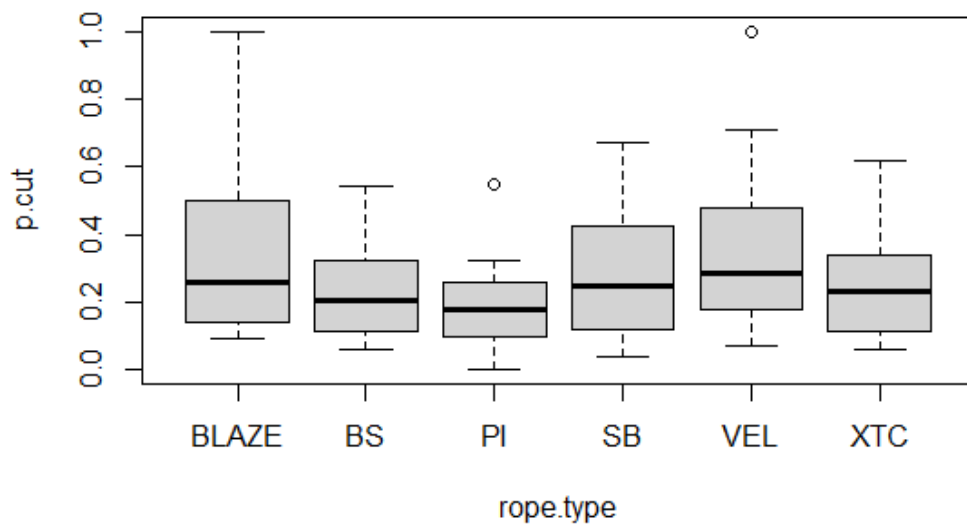
2)   There is a low amount of variance between most of the rope types, this can be seen by the compact boxplots and similar sized boxes and whiskers. The group with the least amount of variance is Pl and the most variance is Blaze.

3) bartlett.test(p.cut ~ rope.type, rope)
4) The boxplot showed the pl rope type had the least variance and the p-value from the bartlett test (0.00142) indicates that one group had a different variance than the rest. From this I would say I would reject the null hypothesis of having a equal variance therefore the ANOVA test wouldn't be appropriate for this data.
5) Blaze rope type is the base case in the table.
6) The mean percentage cut for the base case is 0.36714 and no calculations are needed, and I read it off the table
7) 0.36714 – 0.10164 = 0.2655 for mean percentage cut for XTC rope.
8) p-value = 7.238e-07
9) The p-value rejects the normality assumptions because it is way smaller than the acceptable p-value.
10) Three groups meet the normality assumption
11) I don't think this test is good for this dataset because all groups don't meet the normality assumption which the test states all groups do.
12)

Conditional BoxPlot of Female Species
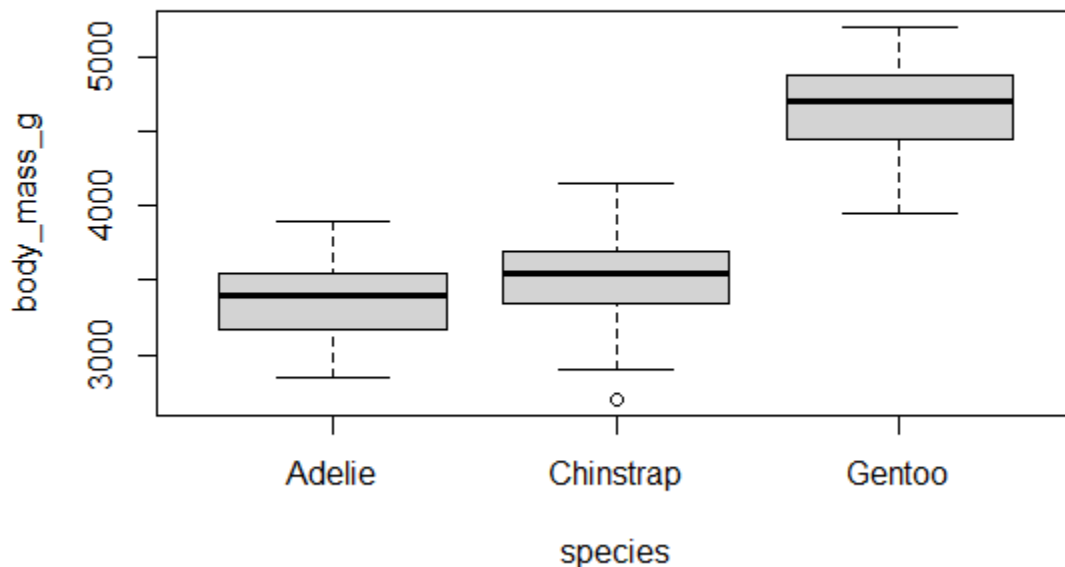
13) I don't think this group of data has good normality because the boxes aren't condensed together and have different widths (Chinstrap is less variance and Gentoo and Adelie have more variance). The homogeneity of the variance look better than the normality because the whiskers are close to similar size, but looking closer you can see small differences in the spread of the data, but in the end the homogeneity looks good.
14) p-value = 0.9056, I would say the homogeneity assumption is met because the p-value of the bartlett test is a lot greater than the value to confirm the null hypothesis.
bartlett.test(body_mass_g ~ species, data = pen_fem)

15) p-value: < 2.2e-16, I would say the normality assumption is not met because the p-value is lower than the p-value to confirm the null hypothesis.
fit_fempen = lm(body_mass_g ~ species, data = pen_fem)
summary(fit_fempen)

16) Gentoo and Adelie female penguins have the biggest difference in body mass (1310.9058 g) which is very significant. Right behind is the difference between Gentoo and Chinstrap female penguin body mass (1152.5355 g) which is also very significant.
fit_fempen_2 = lm(body_mass_g ~ species, data = pen_fem)
fempen2_hsd = TukeyHSD(aov(fit_fempen_2))
class(fempen2_hsd)
round(fempen2_hsd$species, digits = 4)

17) These results match the conditional boxplot I created, you can see there is a big difference in body mass between Gentoo female penguins and the other two species. You can also see there is a difference between Adelie and Chinstrap penguins but not as significant as Gentoo to the other two.