Julia Vineyard

Worked with: Andrew Gordon, Jessica Bonin

Lab 10 ANOVA By Hand

1. require(here)

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

rm(list = ls())

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

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

levels(rope$rope.type)

n\_obs = nrow(rope)

n\_obs

n\_groups = length(levels(rope$rope.type))

n\_groups

mean\_cut= mean(rope$p.cut)

resids\_cut= rope$p.cut - mean\_cut

ss\_tot= sum(resids\_cut^2)

ss\_tot

df\_tot = n\_obs -1

df\_tot

agg\_resids= aggregate(x=rope$p.cut, by=list(rope$rope.type),

FUN= function(x) {x-mean(x)})

str(agg\_resids)

agg\_sq\_resids = aggregate(x=rope$p.cut, by=list(rope$rope.type),

FUN= function(x) {sum((x-mean(x))^2)})

agg\_sq\_resids

str(agg\_sq\_resids)

ss\_within = sum(agg\_sq\_resids$x)

ss\_within

df\_within = n\_obs - n\_groups

df\_within

ss\_among = ss\_tot - ss\_within

ss\_among

df\_among = n\_groups - 1

df\_among

ms\_within =ss\_within / df\_within

ms\_among =ss\_among / df\_among

f\_ratio = ms\_among / ms\_within

f\_ratio #if a big number then reject the null that all means are the same

f\_pval = pf(f\_ratio, df\_among, df\_within, lower.tail= FALSE)

f\_pval

fit\_1= lm(p.cut ~ rope.type, data=rope)

anova\_fit\_1 =anova(fit\_1)

str(anova\_fit\_1)

anova\_fit\_1$`Sum Sq`

2. Based on the boxplot I think that there are not equal variances among the groups.

3. p-value = 0.00143

4. No, an ANOVA analysis is not appropriate for the raw data since there is not homogeneity among the variances. This is supported by the boxplot and the p-value of the Bartlett test. The significant p-value means that we can reject the null hypothesis that the variance is the same amongst the groups.

5. BLAZE is the base case.

6. 0.36714. The base case rope is just the Estimate value of the Intercept.

7. 0.36714 - 0.10164 = 0.2655 (Intercept Estimate + rope.typeXTC Estimate)