Lab 10: ANOVA

Ethan Rutledge

11/28/2021

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 I'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)
```

```
## [1] "BLAZE" "BS"    "PI"    "SB"    "VEL"   "XTC"
```

```
length(rope$rope.type)
```

```
## [1] 121
```

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

```
## [1] 6
```

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

```
## BLAZE    BS    PI    SB   VEL   XTC
##    21    20    20    20    20    20
```

```
n_obs = 121
n_groups = 6

ss_tot = sum((rope$p.cut - mean(rope$p.cut))^2)
ss_tot
```

```
## [1] 5.34757
```

```
df_tot = n_obs - 1
df_tot
```

```
## [1] 120
```

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

```
##   Group.1         x
## 1   BLAZE 0.3671429
## 2      BS 0.2370000
## 3      PI 0.1870000
## 4      SB 0.2720000
## 5     VEL 0.3500000
## 6     XTC 0.2655000
```

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

str(agg_resids)
```

```
## 'data.frame':    6 obs. of  2 variables:
##  $ Group.1: Factor w/ 6 levels "BLAZE","BS","PI",..: 1 2 3 4 5 6
##  $ x      :List of 6
##   ..$ : num  0.633 0.633 0.623 0.173 0.143 ...
##   ..$ : num  0.303 0.223 0.193 0.183 0.093 ...
##   ..$ : num  0.363 0.133 0.113 0.103 0.083 0.063 0.053 0.053 0.033 -0.007 ...
##   ..$ : num  0.398 0.238 0.178 0.168 0.168 0.138 0.118 0.118 0.048 0.038 ...
##   ..$ : num  0.65 0.36 0.3 0.22 0.16 ...
##   ..$ : num  0.3545 0.3145 0.2745 0.2545 0.0745 ...
```

```r
agg_sq_resids = aggregate(x = rope$p.cut,by = list(rope$rope.type),FUN = function(x) sum((x - mean(x))^

str(agg_sq_resids)
```

```
## 'data.frame':    6 obs. of  2 variables:
##  $ Group.1: Factor w/ 6 levels "BLAZE","BS","PI",..: 1 2 3 4 5 6
##  $ x      : num  1.808 0.405 0.312 0.633 1.129 ...
```

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

ss_within
```

```
## [1] 4.874684
```

```r
df_within
```

```
## [1] 115
```

```r
ss_among = ss_tot - ss_within
df_among = df_tot - df_within

ss_among
```

```
## [1] 0.4728867
```

```
df_among
```

```
## [1] 5
```

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

ms_within
```

```
## [1] 0.04238855
```

```
ms_among
```

```
## [1] 0.09457734
```

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

f_ratio
```

```
## [1] 2.2312
```

```
f_pval
```

```
## [1] 0.05582134
```

Self Check: I'll use the following script to test your answers. You can use it as a self-test prior to submitting your answer.

```
# number comparison tolerance
digits_check = 5

# Build the reference model using R functions
fit_1 = lm(p.cut ~ rope.type, data=rope)
anova(fit_1)
```

```
## Analysis of Variance Table
##
## Response: p.cut
##            Df Sum Sq  Mean Sq F value  Pr(>F)
## rope.type   5 0.4729 0.094577  2.2312 0.05582 .
## Residuals 115 4.8747 0.042389
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova_fit_1 = anova(fit_1)

# Check degrees of freedom
anova_fit_1$Df == c(df_among, df_within)
```

```
## [1] TRUE TRUE
```

```r
# Check sums of squares
round(anova_fit_1$`Sum Sq`, digits = digits_check) == round(c(ss_among, ss_within), digits = digits_che
```

```
## [1] TRUE TRUE
```

```r
# Check mean squares
round(anova_fit_1$`Mean Sq`, digits = digits_check) == round(c(ms_among, ms_within), digits = digits_ch
```

```
## [1] TRUE TRUE
```

```r
# Check the F-ratio
round(anova_fit_1$`F value`[1], digits = digits_check) == round(f_ratio, digits = digits_check)
```

```
## [1] TRUE
```

```r
# Check the F test statistic p-value
round(anova_fit_1$`Pr(>F)`[1], digits = digits_check) == round(f_pval, digits = digits_check)
```

```
## [1] TRUE
```

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?

There does not appear to be equal variance among the groups because the sizes of the boxes are different.

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

```r
bartlett.test(agg_resids$x)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  agg_resids$x
## Bartlett's K-squared = 19.687, df = 5, p-value = 0.00143
```

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.

No, the data violates the homogeniety assumption according to the box plots and the significant value in the bartlett test.

```r
fit_rope_1 = lm(p.cut ~ rope.type, data = rope)
summary(fit_rope_1)
```

```
##
## Call:
## lm(formula = p.cut ~ rope.type, data = rope)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.2800 -0.1500 -0.0355  0.1030  0.6500
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.36714    0.04493   8.172 4.45e-13 ***
## rope.typeBS  -0.13014    0.06433  -2.023  0.04538 *
## rope.typePI  -0.18014    0.06433  -2.800  0.00599 **
## rope.typeSB  -0.09514    0.06433  -1.479  0.14186
## rope.typeVEL -0.01714    0.06433  -0.266  0.79033
## rope.typeXTC -0.10164    0.06433  -1.580  0.11683
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2059 on 115 degrees of freedom
## Multiple R-squared:  0.08843,    Adjusted R-squared:  0.0488
## F-statistic: 2.231 on 5 and 115 DF,  p-value: 0.05582
```

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

BLAZE

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.

0.36714, it is what it is because it is the base case there was no calculation.

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

```
0.36714 + -0.10164
```

```
## [1] 0.2655
```