Data Assignment 6

Yash Patel

29 April 2022

Question 1

There are two ways to get the data for the one way ANOVA. Use the either, depending on your preference.

```r
anv_data_df <- data.frame(
  group = rep(
    c("A", "B", "C", "D"),
    each = 50),
  DV = c(runif(50, 0, 3), runif(50, 0, 5), runif(50, 1, 7), runif(50, 1, 7)),
  stringsAsFactors=T)

class(anv_data_df)
```

```
## [1] "data.frame"
```

```r
str(anv_data_df)
```

```
## 'data.frame':    200 obs. of  2 variables:
##  $ group: Factor w/ 4 levels "A","B","C","D": 1 1 1 1 1 1 1 1 1 1 ...
##  $ DV   : num  1.471 1.624 0.126 0.891 0.061 ...
```

Run one_way_anova

```r
anv_output = aov(DV ~ group, data = anv_data_df)
summary(anv_output)
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## group          3  290.6   96.86   44.73 <2e-16 ***
## Residuals    196  424.4    2.17
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Get group means

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```
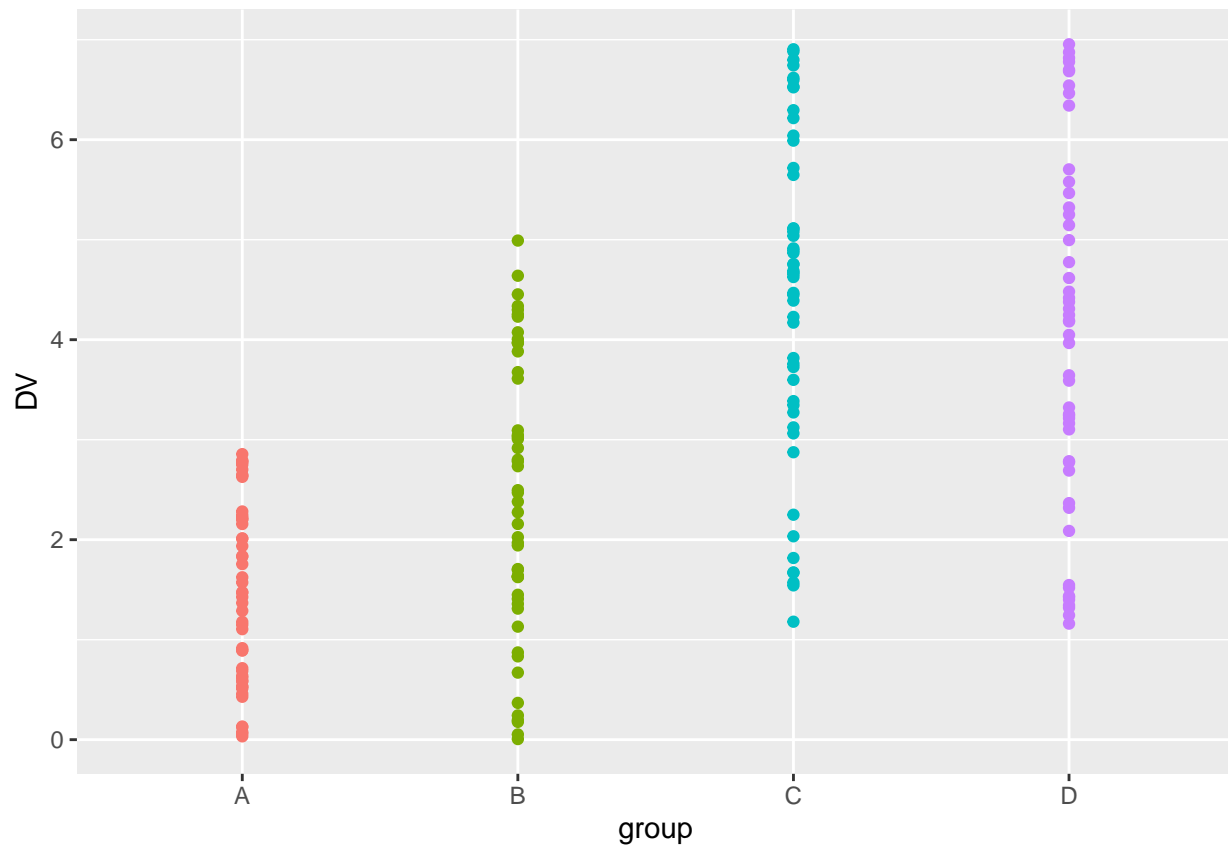
```
anv_data_df %>% group_by(group) %>% summarise(mean = mean(DV))
```

```
## # A tibble: 4 x 2
##   group  mean
##   <fct> <dbl>
## 1 A      1.45
## 2 B      2.39
## 3 C      4.44
## 4 D      3.98
```
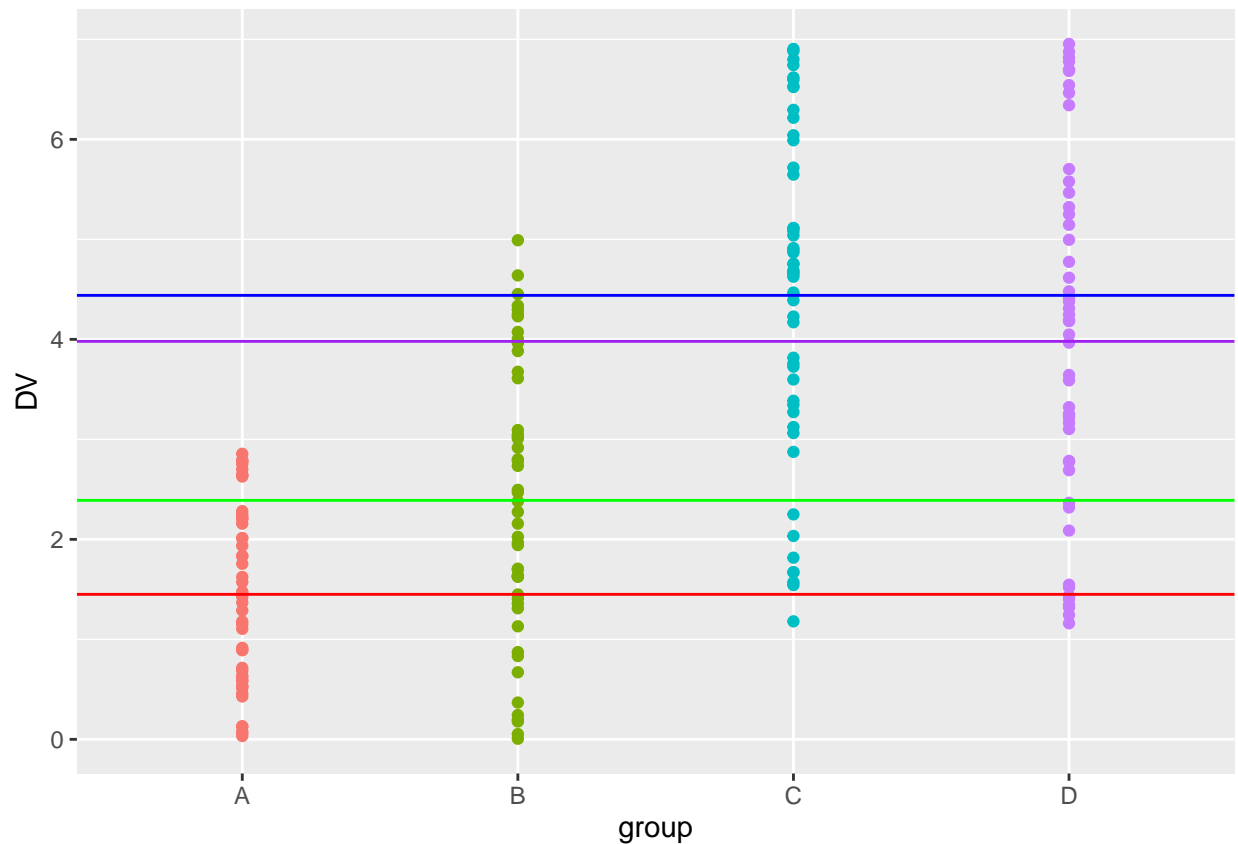
Plot data

```
library(ggplot2)
ggplot(anv_data_df) + aes(x = group, y = DV, color = group) + geom_point() + theme(legend.position = "n
```


Redo the plot, but adding lines to indicate group means

```
ggplot(anv_data_df) + aes(x = group, y = DV, color = group) + geom_point() + theme(legend.position = "no
```


What is the F value for the ANOVA? (6 pts)

44.73

Is it statistically significant? (6 pts). At what level (what p value)? (6 pts)

Yep! p < 2e-16

What are the between group and within group degrees of freedom? (6pts)

Between group: 3 Within group: 196

What are the total degrees of freedom? (5 pts)

199

Interpret the graph (either one). What can you say about the different group data? (5 pts)

The means of group A and B seem to be different compared to each other and groups C and D (C and D, while different, are relatively close together, though it can be argued that they are different. It depends on

what further statistical tests say (Tukey's HSD)!). The within-group variance of B, C, and D seem to be higher than A's since A's plot seems to be a bit tighter.

Question 2

There are two ways to get the data for the two-way ANOVA. Use the either, depending on your preference.

```r
anv2_data_df <- data.frame(
  IV1 = rep(c("A", "B", "C"), each = 60),
  IV2 = c(rep(c("short", "long"), each =30),
          rep(c("short", "long"), each =30),
          rep(c("short", "long"), each =30)),
  DV = c(runif(60, 0, 3), runif(60, 0, 5), runif(60, 1, 10)),stringsAsFactors=T)

class(anv2_data_df)
```

```
## [1] "data.frame"
```

```r
str(anv2_data_df)
```

```
## 'data.frame':    180 obs. of  3 variables:
##  $ IV1: Factor w/ 3 levels "A","B","C": 1 1 1 1 1 1 1 1 1 1 ...
##  $ IV2: Factor w/ 2 levels "long","short": 2 2 2 2 2 2 2 2 2 2 ...
##  $ DV : num  0.82 0.203 0.788 2.482 2.921 ...
```

Run two_way_anova

```r
anv2_output = aov(DV ~ IV1*IV2, data = anv2_data_df)
summary(anv2_output)
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## IV1            2  700.8   350.4 107.987 <2e-16 ***
## IV2            1    0.3     0.3   0.082  0.775
## IV1:IV2        2    6.9     3.5   1.069  0.346
## Residuals    174  564.6     3.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Plot data

```r
interaction.plot(x.factor = anv2_data_df$IV2, #x-axis variable
  trace.factor = anv2_data_df$IV1, #variable for lines
  response = anv2_data_df$DV, #y-axis variable
  fun = mean, #metric to plot
  ylab = "DV",
  xlab = "IV2",
  col = c("green", "blue", "red"),
  lty = 1, #line type
  lwd = 2, #line width
  trace.label = "IV1")
```
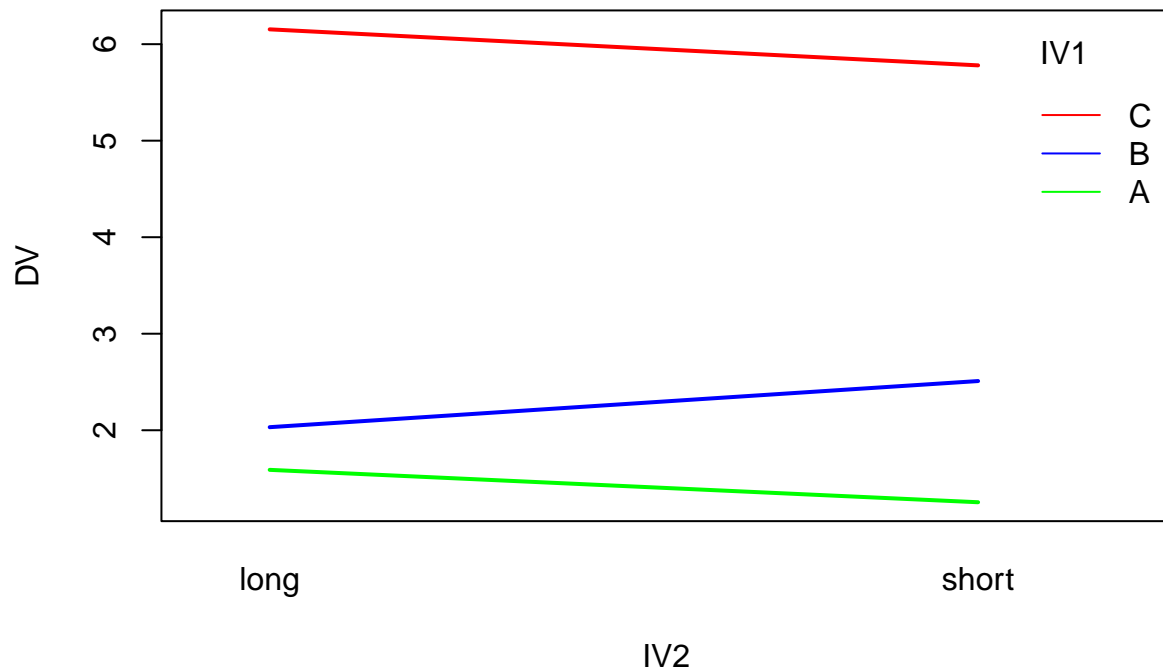
What are the three F values? (6 pts)

IV1 - 107.987 IV2 - 0.082 IV1:IV2 - 1.069

Which F values are statistically significant? (6 pts). At what level (what p value)? (6 pts)

IV1's 107.987 is statistically significant (p < 2e-16)

What are the degrees of freedom for the two independent variables and the interaction? (6pts)

IV1 - 2 IV2 - 1 IV1:IV2 - 2

What are the total degrees of freedom? (5 pts)

179

Interpret the graph (either one). What can you say about the different independent variables (as indicated by the plots)? (5 pts)

The graph suggests that IV2 (long v. short) does not have much of an effect on the DV (lines are "parallel-ish" with the x-axis). IV1 (A, B, C), however, does have an effect, for group C is much higher than groups A and B. There does not seem to be any interaction effects as well since there are no intersections.