Advanced Psychological Statistics
PSYCH-UA.11
Department of Psychology
Spring 2022

Assignment 6

1. **One Way ANOVA**

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

a) Import data
Download the data from Brightspace (for this to work, you'll need to know in which directory the downloads are stored)

Once the data has been downloaded import the data using the following steps -

Step 1. import ggplot2 (This package, "ggplot2" is graphics plug-in for R
install.packages("ggplot2")

Step 2. load ggplot2
library(ggplot2)

Step 3. Data
Option 1 - import data sets
a) Use command "read.csv" to import data
anv_data_df = read.csv("one_way_data.csv", stringsAsFactors=T)

b) Check what kind of object is "anv_data_df" with "class"
class(anv_data_df) (it should be a data frame)

c) Check the item types (e.g. numbers, characters...) using "str")
str(anv_data_df) There are two columns. The column "group" should be factor
The column DV should be "num" (number)

d) If needed, change "group" column from character ("chr") to factor with as.Factor
anv_data_df$group = as.factor(anv_data_df$group))

e) Check results using "str" again
str(anv_data_df)

Option 2 - create data (instead of loading data)

a) Run this code -

```
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)
```

b) Check what kind of object is "anv_data_df" with "class"
class(anv_data_df) (it should be a data frame

c) Check the item types (e.g. numbers, characters...) using "str")
str(anv_data_df) There are two columns. The column "group" should be factor
The column DV should be "num" (number)

d) If needed, change "group" column from character ("chr") to factor with as.Factor
anv_data_df$group = as.factor(anv_data_df$group))

e) Check results using "str" again
str(anv_data_df)

Step 4. Run one_way_anova

a) anv_output = aov(DV ~ group, data = anv_data_df) - Note dependent variable is written first, then independent variable

b) Look out summary ANOVA table using "summary" command
summary(anv_output)

Step 5. Get group means

Option 1 - import data sets

a) group_a = read.csv("group_a.csv")
b) group_b = read.csv("group_b.csv")
c) group_c = read.csv("group_c.csv")
d) group_d = read.csv("group_d.csv")

Step 6. Get the means
a) group_a_mean = mean(group_a[[2]])
b) group_a_mean = mean(group_b[[2]])
c) group_a_mean = mean(group_c[[2]])
d) group_a_mean = mean(group_d[[2]])

Option 2 – Getting means with plugin “dplyr”

a) Install dplyr

```
install.packages(“dplyr”)
```

b) load dplyr

```
library(dplyr)
```

c) get means using “summarise” command from dplyr package

```
anv_data_df %>% group_by(group) %>% summarise(mean = mean(DV))
```

Step 7. Plot data

```
ggplot(anv_data_df)+aes(x = group, y = IV, color = group)+ geom_point() +
theme(legend.position = "none")
```

Step 8. Redo the plot, but adding lines to indicate group means

```
ggplot(anv_data_df)+aes(x = group, y = IV, color = group)+ geom_point() +
theme(legend.position = "none")+ geom_hline(yintercept= ms,color=c("red", "green",
"blue", "purple"))
```

Questions -

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

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

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

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

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

2. **Two Way ANOVA**

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

a) Import data

Download the data from Brightspace (for this to work, you'll need to know in which directory the downloads are stored)

Once the data has been downloaded import the data using the following steps -

Step 1. import ggplot2 (This package, "ggplot2" is graphics plug-in for R

```
install.packages("ggplot2")
```

-----Not needed if RStudio hasn't been closed after performing the one way ANOVA

Step 2. load ggplot2

```
library(ggplot2)
```

-----Not needed if RStudio hasn't been closed after performing the one way ANOVA

Step 3. Data

Option 1 - import data sets

a) Use command "read.csv" to import data

```
anv2_data_df = read.csv("two_way_data.csv", stringsAsFactors=T)
```

b) Check what kind of object is "anv2_data_df" with "class"

```
class(anv2_data_df) (it should be a data frame)
```

c) Check the item types (e.g. numbers, characters...) using "str")

```
str(anv_data2_df)  There are two columns.  The column "group" should be factor
The column DV should be "num" (number)
```

d) If needed, change "group" column from character ("chr") to factor with as.Factor

```
anv_data_df$group = as.factor(anv2_data_df$group))
```

e) Check results using "str" again

```
str(anv2_data_df)
```

Option 2 - create data (instead of loading data)

a) Run this code -

```
anv2_data <- 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)
```

b) Check what kind of object is "anv2_data_df" with "class"
class(anv2_data_df) (it should be a data frame

c) Check the item types (e.g. numbers, characters...) using "str")
str(anv2_data_df) There are two columns. The columns "IV1" and “IV2” should be factor
The column DV should be "num" (number)

d) If needed, change "IV1" and “IV2” columns from character ("chr") to factor with as.Factor
anv2_data_df$group = as.factor(anv_data_df$group))

e) Check results using "str" again
str(anv2_data_df)

Step 4. Run two_way_anova

a) anv2_output = aov(DV ~ IV1*IV2, data = anv2_data_df) - Note dependent variable is written first, then independent variable

b) Look out summary ANOVA table using "summary" command
summary(anv2_output)

Step 7. Plot data

```
interaction.plot(x.factor = anv_data$IV2, #x-axis variable
      trace.factor = anv_data$IV1, #variable for lines
      response = anv_data$DV, #y-axis variable
      fun = mean, #metric to plot
      ylab = "DV",
      xlab = "IV2",
      col = c("green", "blue"),
      lty = 1, #line type
      lwd = 2, #line width
      trace.label = "IV1")
```

Questions -

a) What are the three F values? **(6 pts)**

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

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

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

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