Statistics challenge

Wifred Mambina

5/23/2020

#Loading tidyverse package that contains ggplo2 for data importation and data vizualization

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.0     v purrr   0.3.4
## v tibble  3.0.1     v dplyr   0.8.5
## v tidyr   1.0.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lsr)
```

#Task 1:Data loading and inspection;since the data is a CSV file file I used read_csv() function which is related to tidyverse package

```
statistics.data<-read_csv("C:/Users/Fred/Desktop/PU7_exercise2_ads.csv")
```

```
## Parsed with column specification:
## cols(
##   gender = col_character(),
##   age = col_double(),
##   animation = col_character(),
##   tagline = col_character(),
##   mean_att_overall = col_double()
## )
```

```
head(statistics.data)
```

```
## # A tibble: 6 x 5
##   gender   age animation tagline mean_att_overall
##   <chr>  <dbl> <chr>     <chr>              <dbl>
## 1 male      20 no        no                  2
## 2 male      21 no        no                  2.93
## 3 male      24 no        no                  3.4
## 4 male      45 no        no                  3.93
## 5 male      27 no        no                  2.47
## 6 male      20 no        no                  2.47
```

#Data inspection;since I have used the above head() to view the data ,I can add two more as shown below

```
aggregate(mean_att_overall~animation+tagline,statistics.data,mean)
```

```
##   animation tagline mean_att_overall
## 1        no      no         3.003509
## 2       yes      no         2.950476
## 3        no     yes         3.169231
## 4       yes     yes         3.677778
```

```
tail(statistics.data)
```

```
## # A tibble: 6 x 5
##   gender   age animation tagline mean_att_overall
##   <chr>  <dbl> <chr>     <chr>              <dbl>
## 1 female    20 no        yes                 2.87
## 2 female    31 no        yes                 3.47
## 3 female    20 no        yes                 3.27
## 4 female    42 no        yes                 2.4
## 5 female    23 no        yes                 2.47
## 6 female    19 no        yes                 4.93
```

```
attach(statistics.data)
```

The tail() gives the overview of the data values on the cells while attach() attaches the five varibles of the dataset so that it can be available when recalling them in exercise. #Task 2: Part a;we are to convert the the variables tagline and animation into factors.Using the head() as shown above it indicates the animation and tagline are in form

```
tagline_v1<-as.factor(statistics.data$tagline)
head(tagline_v1)
```

```
## [1] no no no no no no
## Levels: no yes
```

```
animation_v1<-as.factor(statistics.data$animation)
head(animation_v1)
```

```
## [1] no no no no no no
## Levels: no yes
```

The above output shows the conversion of tagline and animation to factors and its levels. #In order to confirm the conversion we can apply the function below

```
class(tagline_v1)
```

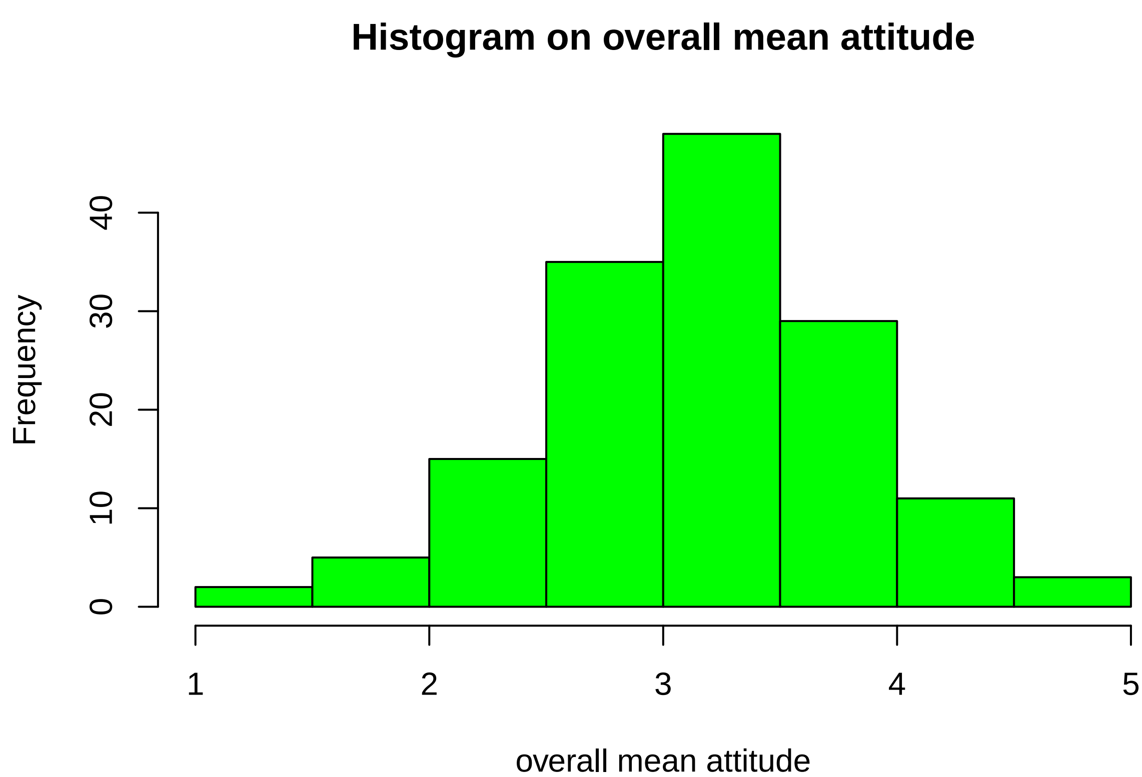
```
## [1] "factor"
```

```r
class(animation_v1)
```

```
## [1] "factor"
```

#Pqrt b:Data visualization on mean_att_all using histogram

```r
hist(x=statistics.data$mean_att_overall,
     main="Histogram on overall mean attitude",
     col = "green",
     xlab = "overall mean attitude")
```



#Task 3;checking on the count levels by using xtab()

```r
xtabs(~animation_v1+tagline_v1,data = statistics.data)
```

```
##             tagline_v1
## animation_v1 no yes
##          no  38  39
##          yes 35  36
```

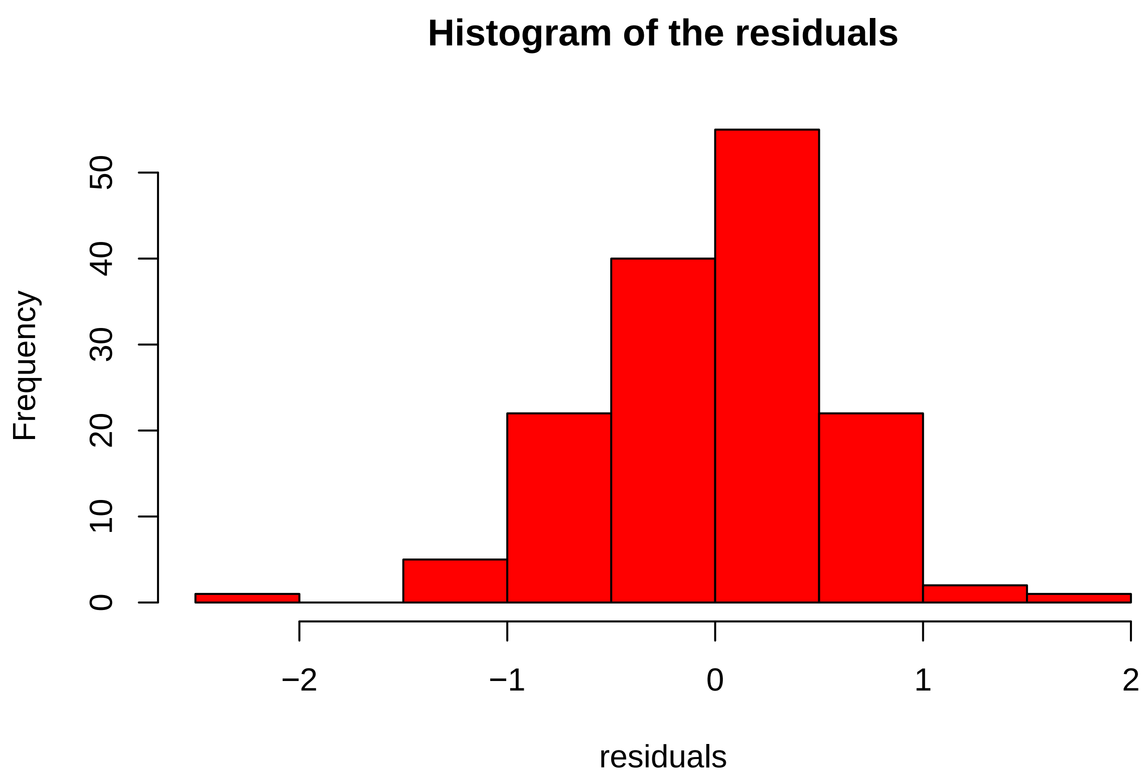
#Task 4:In this case ,the first scenario is to run the ANOVA

```r
anova.t4<-aov(formula = mean_att_overall~animation*tagline,data = statistics.data)
summary(anova.t4)
```

```
##                    Df Sum Sq Mean Sq F value   Pr(>F)    
## animation           1   1.98   1.985   5.757  0.01771 *  
## tagline             1   7.00   7.004  20.314 1.35e-05 ***
## animation:tagline   1   2.91   2.912   8.445  0.00424 ** 
## Residuals         144  49.65   0.345                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

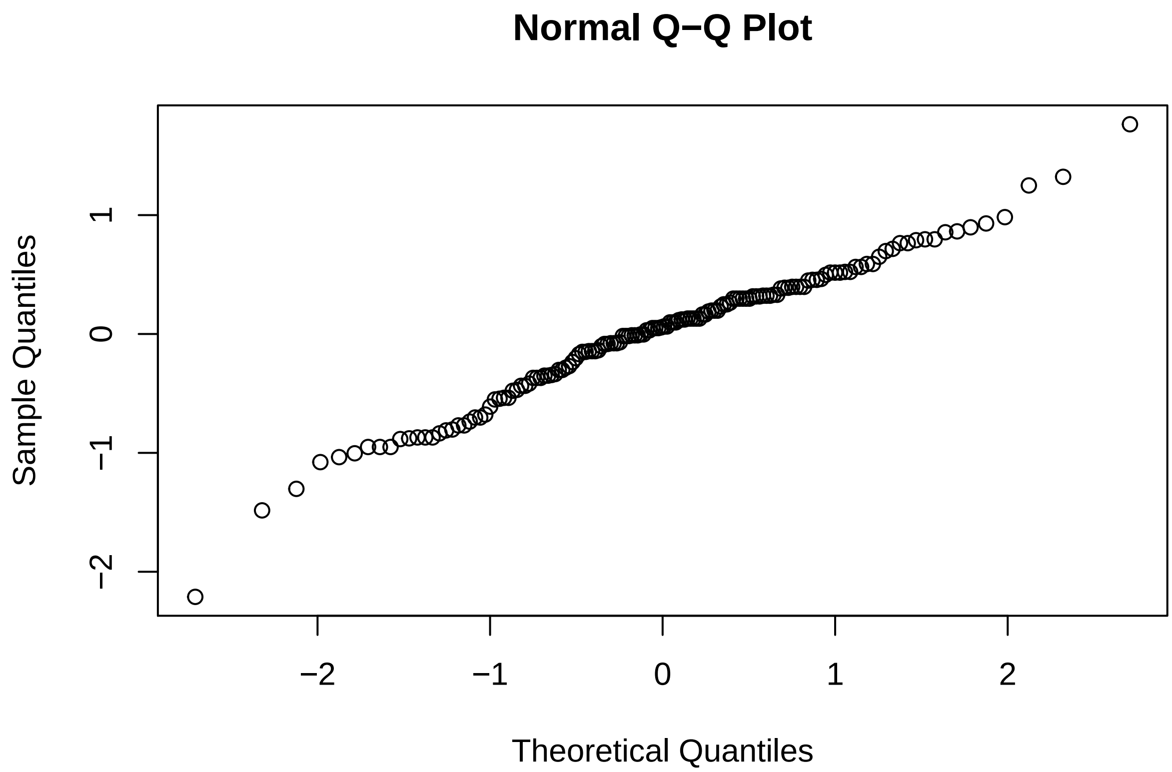
Before interpretation and conclusion on the output I have to check whether the model residuals are normally distributed by the use of the "QQ plot"and "Histogram" #Histogram

```r
my.anova.residuals<-residuals(anova.t4)
hist(my.anova.residuals,
     main = "Histogram of the residuals",
     col = "red",
     xlab = "residuals")
```



From the above we can see the residuals are not normally distributed since data points are negatively skewed. #Using QQ Plot

```r
qqnorm(my.anova.residuals)
```



In QQ plot the assumption is that the data is assumed normally distrbuted if the points on the graph forms a line Which is not the case above.Hence,its not normally distributed.

#Task 5:Conducting Shapiro test and calculating effect size of the ANOVA test.

```
shapiro.test(my.anova.residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  my.anova.residuals
## W = 0.97996, p-value = 0.02928
```

From our output we found that (W=0.97966,P-value=0.02928<0.05) Therefore the residuals are not normally distributed since the P-value=0.02928<0.05 #To check the assumption of homogeneity of variance ,I will conduct a Lev ene test

```
library(car)
```

```
## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```r
leveneTest(anova.t4)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   3  0.2859 0.8355
##       144
```

The test reveals a P-value>0.05 which indicates that there is no significant difference between the group variance.

Further I will perform regular factorial ANOVA since the normality assumption is violated.In this case I will apply Anova(),that is associated with te car package #Factorial ANOVA

```r
Anova(anova.t4,type=2)
```

```
## Anova Table (Type II tests)
##
## Response: mean_att_overall
##                   Sum Sq  Df F value    Pr(>F)
## animation          1.981   1  5.7448  0.017821 *
## tagline            7.004   1 20.3138 1.351e-05 ***
## animation:tagline  2.912   1  8.4452  0.004239 **
## Residuals         49.650 144
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I also run a Kruskal-Wallis test on the main effect to see if there is still significant in a non-parametric test

```r
kruskal.test(formula=mean_att_overall~animation,data=statistics.data)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  mean_att_overall by animation
## Kruskal-Wallis chi-squared = 5.8547, df = 1, p-value = 0.01554
```

```r
kruskal.test(formula=mean_att_overall~tagline,data = statistics.data)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  mean_att_overall by tagline
## Kruskal-Wallis chi-squared = 17.677, df = 1, p-value = 2.617e-05
```

From the ANOVA output,we can see a significant effect for animation($F(1)=5.7,p<.05$),a significant effect for tagline ($F(1)=20.3,P<0.001$),and a significant interaction effect ($F(1)=8.4,P<.01$).Also in Kruskal-Wallis test indicates significant effect for animation(Chi-squared(1)=5.9,$p<.05$) and for tagline(Chi-squared(1)=17.7,$P<.001$) Lastly,to calculate effect sizes,I used etaSquared() with a type II sum of squares

```r
etaSquared(anova.t4,type = 2)
```

```
##                        eta.sq eta.sq.part
## animation          0.03218073  0.03836371
## tagline            0.11379282  0.12362809
## animation:tagline 0.04730816  0.05539856
```

I also used Tukey's HSD as a posthoc test

```r
TukeyHSD(anova.t4)
```

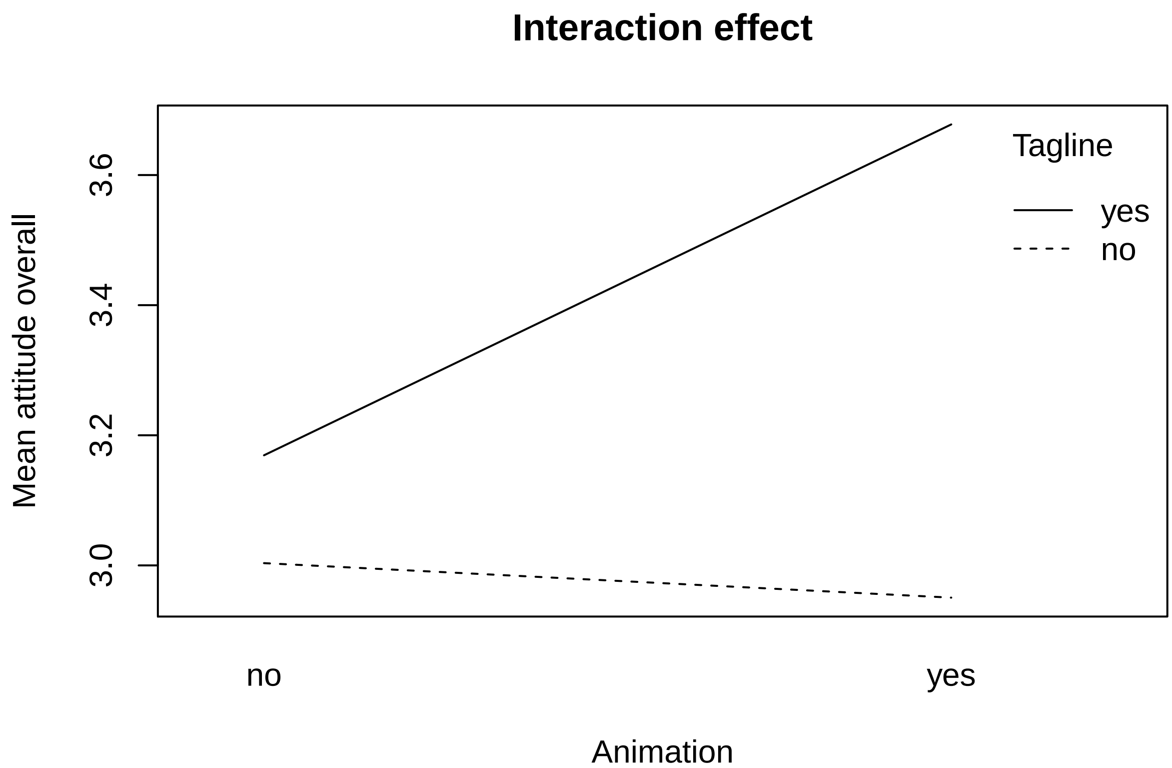
```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = mean_att_overall ~ animation * tagline, data = statistics.data)
##
## $animation
##             diff        lwr       upr     p adj
## yes-no 0.2318029 0.04084028 0.4227656 0.0177057
##
## $tagline
##             diff      lwr       upr    p adj
## yes-no 0.4351241 0.244301 0.6259472 1.35e-05
##
## $`animation:tagline`
##                        diff        lwr       upr     p adj
## yes:no-no:no   -0.05303258 -0.4106085 0.3045433 0.9804523
## no:yes-no:no    0.16572200 -0.1821776 0.5136216 0.6036877
## yes:yes-no:no   0.67426901  0.3192877 1.0292503 0.0000128
## no:yes-yes:no   0.21875458 -0.1366166 0.5741257 0.3818582
## yes:yes-yes:no  0.72730159  0.3649948 1.0896084 0.0000037
## yes:yes-no:yes  0.50854701  0.1557867 0.8613073 0.0014591
```

#Task 6:Reporting the ANOVA test result using APA format

A factorial ANOVA was conducted to compare the effect of tagline and animation on overall mean attitude.There was a significant effect of animation(F(1)=5.7,p<.05),a significant effect for tagline (F(1)=20.3,P<0.001),and a significant interaction effect (F(1)=8.4,P<.01).Also in Kruskal-Wallis test indicates significant effect for animation(Chi-squared(1)=5.9,p<.05) and for tagline(Chi-squared(1)=17.7,P<.001) Taken together ,these results suggest that the animation real affects the overall mean attitude as compared to tagline.

#Task 7;Visualising the data by plotting the results and looking for an interaction effect

```r
interaction.plot(x.factor = statistics.data$animation,trace.factor = statistics.data$tagline,
                 response = statistics.data$mean_att_overall,
                 main="Interaction effect",
                 xlab = "Animation",
                 ylab = "Mean attitude overall",
                 trace.label = "Tagline")
```



From the plot above we can see there is a significant interaction effect for the level“no”.