Problem Set 2, Fall 2021

Ben Karabinus

```
# load required libraries
library(ggplot2)
library(ggpubr)
library(car)

## Loading required package: carData

library(lawstat)

##
## Attaching package: 'lawstat'

## The following object is masked from 'package:car':
##
##     levene.test

library(dplyr)

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:car':
##
##     recode

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

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(ggeasy)
```

Donna is the owner of a boutique doughnut shop. Because many of her customers are conscious of their fat intake but want the flavor of fried doughnuts, she decided to develop a doughnut recipe that minimizes the amount of fat that the doughnuts absorb from the fat in which the doughnuts are fried.

She conducted a factorial experiment that had a similar procedures as Lowe (1935). Like Lowe, she used four types of fats (fat_type). She also used three types of flour (flour_type): all-purpose flour, whole wheat flour, and gluten-free flour. For each combination of fat type and flour type, she cooked six identical batches of doughnuts. Each batch contained 24 doughnuts, and the total fat (in grams) absorbed by the doughnuts in each batch was recorded (sim_tot_fat).

Question 1 - 5 points

You may need to process your data before you begin your analysis. Specifically, you will need to make sure that the variable type is set to 'factor' for both of your grouping variables and 'num' for your outcome variable.

```
doughnuts.factorial <- read.csv("doughnutsfactorial.csv", header=TRUE, sep=",
") # Loads the CSV file into memory. You may need to adapt this line to work
on your computer
```

Like in Problem Set 1, please create two new variables in the doughnuts.factorial data set. The first new variable will be called fat_type_factor and will contain the same values as in the fat_type variable but will have a variable type of factor. The second new variable will be called flour_type_factor and will contain the same values as in the flour_type variable but will also have a variable type of factor.

```
# create factors for flour_type and fat_type
doughnuts.factorial$fat_type_factor <- as.factor(doughnuts.factorial$fat_type
)
doughnuts.factorial$flour_type_factor <-
as.factor(doughnuts.factorial$flour_type)
```

Check your work by running the following code chunk. Be sure that fat_type_factor and flour_type_factor are factor-type variables before you complete the rest of the problem set.

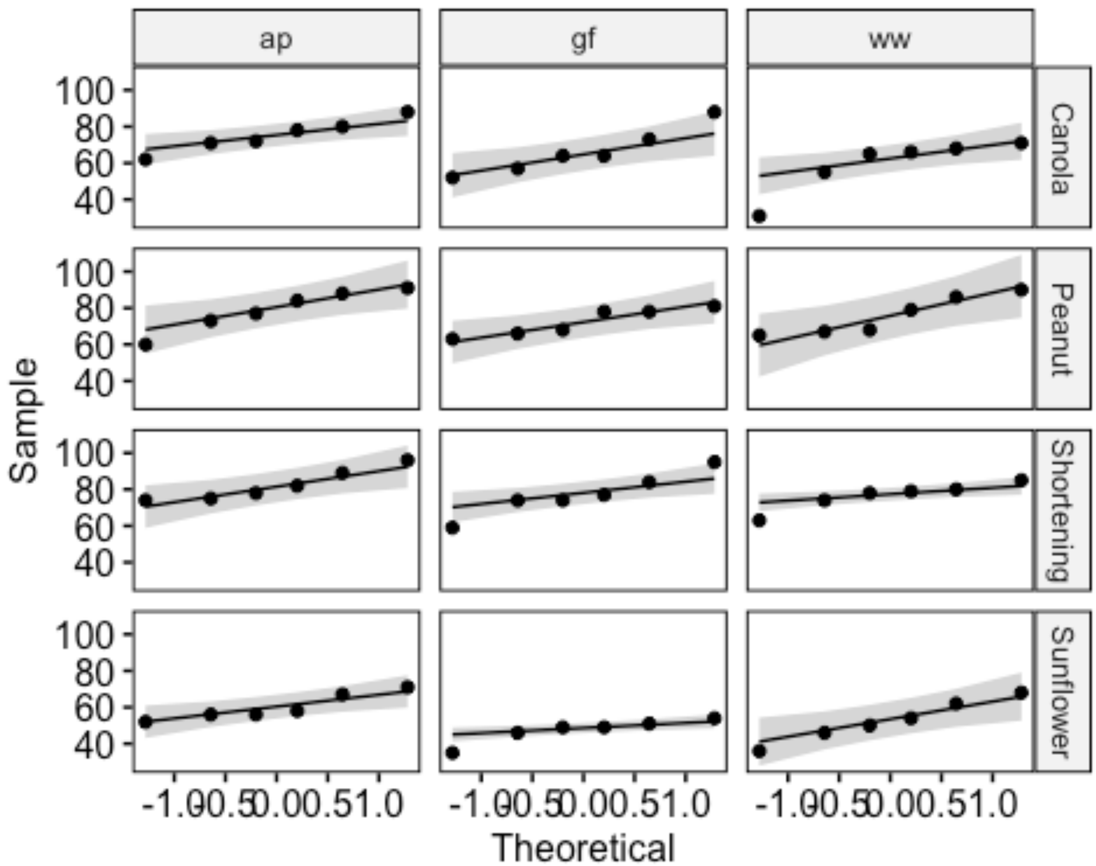
```
#check the structure of the dataset
str(doughnuts.factorial)

## 'data.frame':    72 obs. of  5 variables:
##  $ fat_type         : chr  "Canola" "Canola" "Canola" "Canola" ...
##  $ flour_type       : chr  "ap" "ap" "ap" "ap" ...
##  $ sim_tot_fat      : int  78 71 80 88 62 72 78 75 89 74 ...
##  $ fat_type_factor  : Factor w/ 4 levels "Canola","Peanut",..: 1 1 1 1 1 1
3 3 3 3 ...
##  $ flour_type_factor: Factor w/ 3 levels "ap","gf","ww": 1 1 1 1 1 1 1 1 1
1 ...
```

Question 2 - 5 points

Provide a visual assessment and a quantitative assessment for the assumption of *normality* for each cell. Hint: Remember that a cell contains the observations that make up a particular combination of two factors. Therefore, there will be as many graphs/quantitative tests as are unique combinations of flour and fat types.

Code for your visual assessment of normality

```
# Create ggqqplot for flour_type, fat_type combinations
donut_plt <- ggqqplot(doughnuts.factorial, x="sim_tot_fat",
        facet.by =c("fat_type_factor","flour_type_factor"))
donut_plt
```



Code for your quantitative assessment of normality

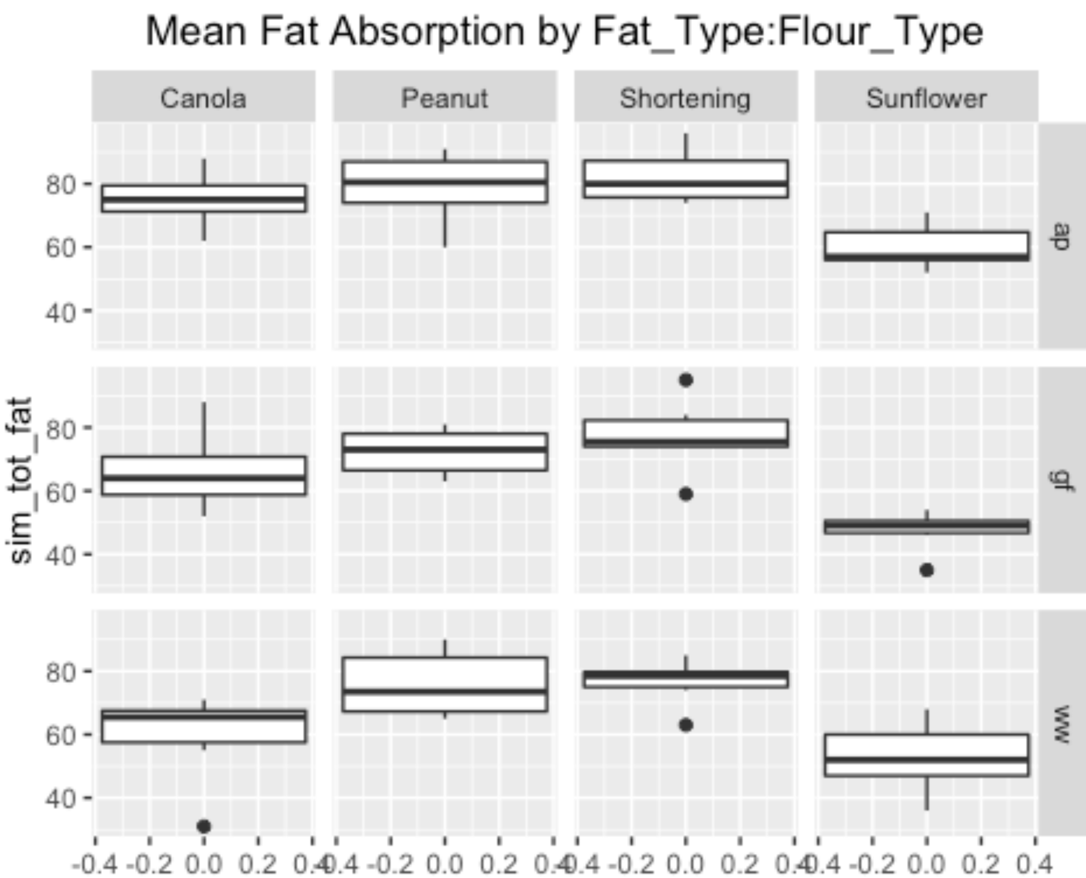
```
# Run the Shapiro-Wilk test for each fat_type, flour_type combination and dis
play as tibble
sw_doughnuts.factorial <- doughnuts.factorial %>%
                group_by(fat_type_factor,
                         flour_type_factor)%>%
                summarize(pval=shapiro.test(sim_tot_fat)$p)

## `summarise()` has grouped output by 'fat_type_factor'. You can override us
ing the `.groups` argument.

# print results
sw_doughnuts.factorial

## # A tibble: 12 × 3
## # Groups:   fat_type_factor [4]
##    fat_type_factor flour_type_factor   pval
##    <fct>           <fct>              <dbl>
##  1 Canola          ap                0.974
##  2 Canola          gf                0.616
##  3 Canola          ww                0.0404
##  4 Peanut          ap                0.675
##  5 Peanut          gf                0.258
##  6 Peanut          ww                0.257
##  7 Shortening      ap                0.434
##  8 Shortening      gf                0.832
##  9 Shortening      ww                0.375
## 10 Sunflower       ap                0.306
## 11 Sunflower       gf                0.168
## 12 Sunflower       ww                0.987
```

Question 3 - 5 points

Provide a visual assessment and a quantitative assessment for the assumption of *equality of variances* for each cell.

Code for your visual assessment of equality of variances

```
# create ggplot geom_boxplot to assess variance of groups visually
donut_box_plt <- ggplot(data=doughnuts.factorial, aes(y=sim_tot_fat))+
       geom_boxplot()+facet_grid(rows=vars(flour_type_factor),
       cols=vars(fat_type_factor))+
       ggtitle("Mean Fat Absorption by Fat_Type:Flour_Type")+
       ggeasy::easy_center_title()

donut_box_plt
```



Code for your quantitative assessment of equality of variances

```
# perform the Brown-Forsythe aka Levene test
# create 12 factor levels for fat_type:flour_type
doughnuts.factorial.levene <- doughnuts.factorial %>%
  mutate(fat_flour = case_when(
  fat_type_factor == "Canola" & flour_type_factor == "ap" ~ 1,
  fat_type_factor == "Canola" & flour_type_factor == "ww" ~ 2,
  fat_type_factor == "Canola" & flour_type_factor == "gf" ~ 3,
  fat_type_factor == "Shortening" & flour_type_factor == "ap" ~ 4,
  fat_type_factor == "Shortening" & flour_type_factor == "ww" ~ 5,
  fat_type_factor == "Shortening" & flour_type_factor == "gf" ~ 6,
  fat_type_factor == "Sunflower" & flour_type_factor == "ap" ~ 7,
  fat_type_factor == "Sunflowwer" & flour_type_factor == "ww" ~ 8,
  fat_type_factor == "Sunflower" & flour_type_factor == "gf" ~ 9,
  fat_type_factor == "Peanut" & flour_type_factor == "ap" ~ 10,
  fat_type_factor == "Peanut" & flour_type_factor == "ww" ~ 11,
  fat_type_factor == "Peanut" & flour_type_factor == "gf" ~ 12))

# as factor
doughnuts.factorial.levene$fat_flour <-
                    as.factor(doughnuts.factorial.levene$fat_flour)

# levenTest from car library (deprecated)
# leveneTest(sim_tot_fat~fat_type_factor*flour_type_factor, data = doughnuts.
factorial)

# perform levene test using lawstat library
levene.test(doughnuts.factorial.levene$sim_tot_fat,
            doughnuts.factorial.levene$fat_flour)

##
##  Modified robust Brown-Forsythe Levene-type test based on the absolute
##  deviations from the median
##
## data:  doughnuts.factorial.levene$sim_tot_fat
## Test Statistic = 0.41383, p-value = 0.9342
```

Question 4 - 10 points

Before conducting your two-way ANOVA, start by conducting one-way ANOVAs for each of your factors. You wouldn't do this in practice - you would just conduct the two-way ANOVA - but you'll do it here to allow you to make some comparisons between one-way ANOVA and two-way ANOVA in Question 7. You do not need to interpret these ANOVAs, but be sure to display the output in your knitted document.

Your one-way ANOVA for testing if the means in total fat (sim_tot_fat) are the same across fat types:

```
# one-way ANOVa sim_tot_fat~fat_type_factor
fat.aov <- aov(sim_tot_fat~fat_type_factor,
               data=doughnuts.factorial)
# ANOVA summary
summary(fat.aov)

##                 Df Sum Sq Mean Sq F value   Pr(>F)    
## fat_type_factor  3   6967  2322.5   20.17 1.86e-09 ***
## Residuals       68   7831   115.2                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Your one-way ANOVA for testing if the means in total fat (sim_tot_fat) are the same across flour types:

```
# one-way ANOVa flour_type_factor
flour.aov <- aov(sim_tot_fat~flour_type_factor, data=doughnuts.factorial)
# ANOVA summary
summary(flour.aov)

##                   Df Sum Sq Mean Sq F value Pr(>F)  
## flour_type_factor  2   1063   531.3   2.669 0.0765 .
## Residuals         69  13736   199.1                 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 5 - 10 points

Conduct a two-way ANOVA with an interaction between fat type and flour type. Use sim_total_fat as the outcome and fat_type_factor and flour_type_factor as the grouping variables. Please be sure to display your ANOVA results using the summary() function.

```
# create two-way ANOVA with interaction term
fat_flour_int.aov<-aov(sim_tot_fat~fat_type_factor*flour_type_factor,
                        data=doughnuts.factorial)
# ANOVA Summary
summary(fat_flour_int.aov)

##                                    Df Sum Sq Mean Sq F value   Pr(>F)
## fat_type_factor                     3   6967  2322.5  21.976 1.01e-09 ***
## flour_type_factor                   2   1063   531.3   5.028  0.00958 **
## fat_type_factor:flour_type_factor   6    427    71.2   0.674  0.67095
## Residuals                          60   6341   105.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 6 - 10 points

Be sure to have completed the two-way ANOVA with an interaction analysis before answering the following four questions.

Main effects hypotheses - two questions to answer

A) Please select the statement that is the best interpretation of the p-value associated with the main effect of fat type.

Statement A: I reject the null hypothesis and conclude that at least one fat type has a statistically significantly different mean fat absorption than the other groups.

Statement B: I fail to reject the null hypothesis and conclude that there is no statistically significant difference in the mean amount of fat absorbed among fat types.

Your answer here:

A.

B) Please select the statement that is the best interpretation of the p-value associated with the main effect of flour type.

Statement A: I reject the null hypothesis and conclude that at least one flour type has a statistically significantly different mean fat absorption than the other groups.

Statement B: I fail to reject the null hypothesis and conclude that there is no statistically significant difference in the mean amount of fat absorbed among flour types.

Your answer here:

B.

Interaction hypothesis - 2 questions to answer

C) Please select the statement that is the best interpretation of the p-value associated with the interaction between fat type and flour type.

Statement A: The interaction between fat type and flour type is statistically significant.

Statement B: The interaction between fat type and flour type is not statistically significant.

Your answer here:

D) Based on your response to the previous question about the interaction, can you interpret the main effects in a straightforward fashion?

Your answer here (yes or no):

Yes

Question 7 - 5 points

You conducted 2 one-way ANOVAs in Question 4 and 1 two-way ANOVA with an interaction in Question 5. In this question, you will answer four questions comparing the results of these analyses.

A) Look at the lines for fat_type_factor in both the one-way ANOVA with fat_type_factor used as the grouping variable and the two-way ANOVA with an interaction. Is there any difference in the degrees of freedom or the sums of squares between these lines?

Your answer here (yes/no):

No

B) Looking at the same lines as the previous question, is there a difference between the F test statistic or the p-values?

Your answer here (yes/no):

Yes

C) Look at the lines for flour_type_factor in both the one-way ANOVA with flour_type_factor used as the grouping variable and the two-way ANOVA with an interaction. Is there any difference in the degrees of freedom and the sums of squares between these lines?

Your answer here (yes/no):

No

D) Looking at the same lines as the previous question, is there a difference between the F test statistic or the p-values?

Your answer here (yes/no):

Yes