Problem Set 2, Winter 2022

Krista Miller

knitr::opts\_chunk$set(echo = TRUE)  
# Load any packages, if any, that you use as part of your answers here  
  
# For example:   
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

library(tidyverse)

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

## v ggplot2 3.3.3 v purrr 0.3.4  
## v tibble 3.1.0 v stringr 1.4.0  
## v tidyr 1.1.3 v forcats 0.5.1  
## v readr 1.4.0

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(readxl)  
library(nortest)  
library(ggpubr)

## Warning: package 'ggpubr' was built under R version 4.0.5

library(lawstat)

## Warning: package 'lawstat' was built under R version 4.0.5

library(ggplot2)

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.

# Complete the lines to properly create the two new variables  
  
doughnuts.factorial$fat\_type\_factor <- as.factor(doughnuts.factorial$fat\_type)# Complete this line  
  
doughnuts.factorial$flour\_type\_factor <- as.factor(doughnuts.factorial$flour\_type)# Complete this line

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.

str(doughnuts.factorial)

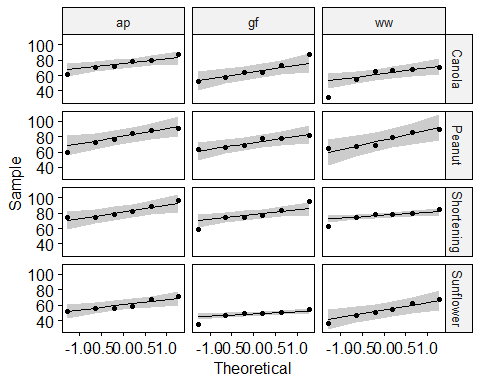
## 'data.frame': 72 obs. of 5 variables:  
## $ fat\_type : chr "Canola" "Canola" "Canola" "Canola" ...  
## $ flour\_type : chr "ap" "ap" "ap" "ap" ...  
## $ sim\_tot\_fat : int 78 71 80 88 62 72 78 75 89 74 ...  
## $ fat\_type\_factor : Factor w/ 4 levels "Canola","Peanut",..: 1 1 1 1 1 1 3 3 3 3 ...  
## $ 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

# Code for visual assessment of normality for each cell  
  
ggqqplot(doughnuts.factorial,x="sim\_tot\_fat",facet.by = c("fat\_type", "flour\_type"))



# Be sure to display your visual assessment in your knitted document!

Code for your quantitative assessment of normality

# Code for quantitative assessment of normality for each cell  
  
doughnuts.factorial%>%group\_by(fat\_type, flour\_type)%>%summarize(pval=shapiro.test(sim\_tot\_fat)$p)

## `summarise()` has grouped output by 'fat\_type'. You can override using the `.groups` argument.

## # A tibble: 12 x 3  
## # Groups: fat\_type [4]  
## fat\_type flour\_type pval  
## <chr> <chr> <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

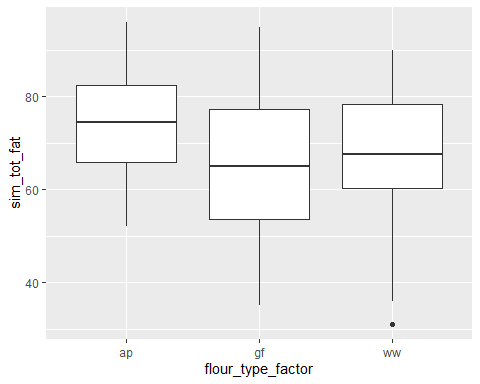
# Be sure to display your quantitative assessment in your knitted document!

## Question 3 - 5 points

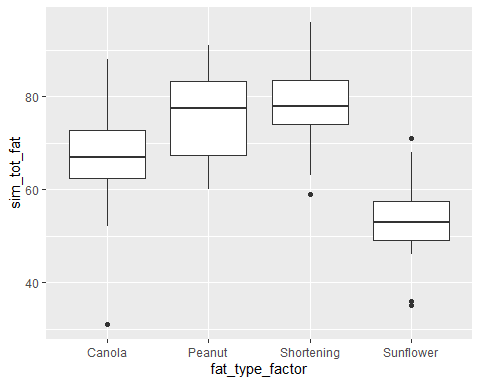
Provide a visual assessment and a quantitative assessment for the assumption of *equality of variances* for each cell.

Code for your visual assessment of equality of variances

# Code for visual assessment of equality of variances for each cell  
  
ggplot(doughnuts.factorial, aes(x=flour\_type\_factor, y= sim\_tot\_fat)) + geom\_boxplot()



ggplot(doughnuts.factorial, aes(x=fat\_type\_factor, y= sim\_tot\_fat)) + geom\_boxplot()



# Be sure to display your visual assessment in your knitted document!

Code for your quantitative assessment of equality of variances

# Code for quantitative assessment of equality of variances for each cell  
levene.test(doughnuts.factorial$sim\_tot\_fat,doughnuts.factorial$fat\_type)

##   
## Modified robust Brown-Forsythe Levene-type test based on the absolute  
## deviations from the median  
##   
## data: doughnuts.factorial$sim\_tot\_fat  
## Test Statistic = 0.60943, p-value = 0.6112

levene.test(doughnuts.factorial$sim\_tot\_fat,doughnuts.factorial$flour\_type)

##   
## Modified robust Brown-Forsythe Levene-type test based on the absolute  
## deviations from the median  
##   
## data: doughnuts.factorial$sim\_tot\_fat  
## Test Statistic = 0.50196, p-value = 0.6075

# Be sure to display your quantitative assessment in your knitted document!

## 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\_total\_fat) are the same across fat types:

fat.aov <- aov(sim\_tot\_fat~fat\_type\_factor, data=doughnuts.factorial)  
  
# Don't forget to display your results!  
   
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\_total\_fat) are the same across flour types:

flour.aov <- aov(sim\_tot\_fat~flour\_type\_factor, data=doughnuts.factorial)  
  
# Don't forget to display your results!  
   
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.

fat\_flour\_int.aov <- aov(sim\_tot\_fat~flour\_type\_factor\*fat\_type\_factor,data=doughnuts.factorial)  
  
summary(fat\_flour\_int.aov)

## Df Sum Sq Mean Sq F value Pr(>F)   
## flour\_type\_factor 2 1063 531.3 5.028 0.00958 \*\*   
## fat\_type\_factor 3 6967 2322.5 21.976 1.01e-09 \*\*\*  
## flour\_type\_factor:fat\_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

1. 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

1. 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: A

# Interaction hypothesis - 2 questions to answer

1. 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: B

1. 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.

1. 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

1. 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

1. 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

1. 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