R Assignment 12

Code segments provided for each question, full R script for the assignment at the bottom

> 1: Using tidyverse commands discussed in class, subset just the rows corresponding to food items sold at Arby's, Subway, and Taco Bell. Share your code here.

```
small_fastfood = filter(fastfood, restaurant == "Arbys" | restaurant == "Subway"
                        | restaurant == "Taco Bell")
```

> 2: Using the mutate() function on the full dataset (not the one created in Question 1), create a column that subtracts the calories from fat from the total calorie content of each food item.

```
fastfood = mutate(fastfood, cal_wo_fat = calories - cal_fat)
```

> 3: Share what commands you would use to select just the restaurant, item, and calories columns AND only include the food items with calorie counts > 1000. Do this all using piping.

```
subset = fastfood %>%
  select(restaurant, item, calories) %>%
  filter(calories > 1000)
```

> 4: Using piping, the group_by function, and the summarise function, compute the average calorie content, the standard deviation of the calorie content, and the sample size for Arby's, Subway, and Taco Bell separately. Use the dataset you created in Question 1. Share your code here and fill in the following table with your results:

```
small_fastfood %>%
  group_by(restaurant) %>%
  summarise(avg_cal = mean(calories), sd_cal = sd(calories),
            sample_size = length(calories))
```

Restaurant	Average Calorie Content	Standard Deviation of Calorie Content	Sample Size
Arby's	533	210	55
Subway	503	282	96
Taco Bell	444	184	115

Performing an ANOVA:

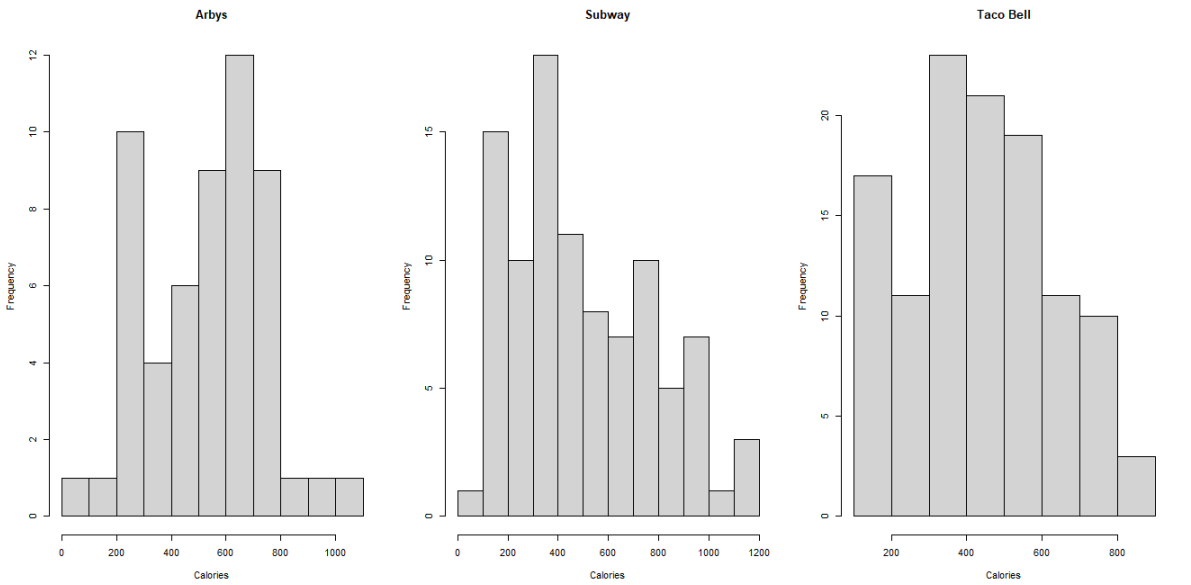
> 5: What are the hypotheses for this ANOVA test?

H_0: There is no difference in the average calorie content of foods between Arby's (μ_{arbys}), Subway (μ_{subway}), and Taco Bell (μ_{taco})
H_0: $\mu_{arbys} = \mu_{subway} = \mu_{taco}$

H_A: At least one of the average calorie contents of foods between Arby's (μ_{arbys}), Subway (μ_{subway}), and Taco Bell (μ_{taco}) is different.

> 6: Create histograms of the calorie contents of each restaurant. Evaluate how normal the distribution of calories looks for each restaurant.

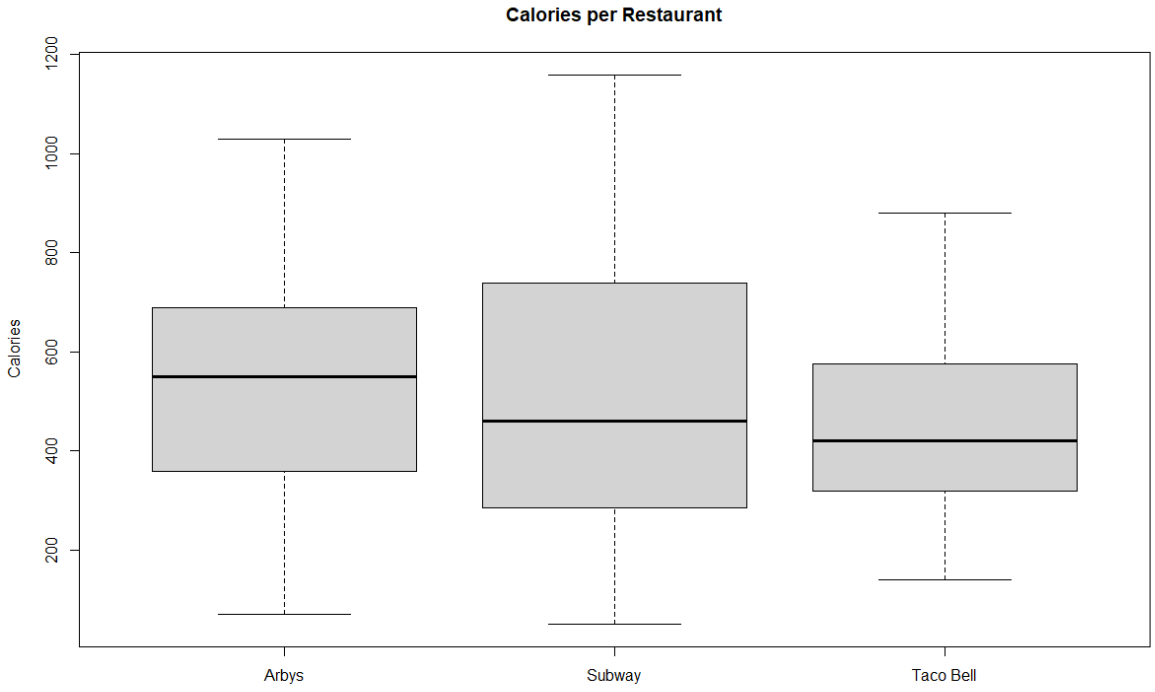
```
par(mfrow=c(1,3))
hist(small_fastfood[which(small_fastfood$restaurant == "Arbys"), ]$calories,
     main = "Arbys", xlab = "Calories")
hist(small_fastfood[which(small_fastfood$restaurant == "Subway"), ]$calories,
     main = "Subway", xlab = "Calories")
hist(small_fastfood[which(small_fastfood$restaurant == "Taco Bell"), ]$calories,
     main = "Taco Bell", xlab = "Calories")
par(mfrow=c(1,1))
```



All three distributions are bimodal, while the distributions for *Subway* and *Taco Bell* are also right-skewed. The distribution for *Taco Bell* is the closest to a normal distribution but it is not appropriate to approximate any of the distributions with a normal distribution.

> 7: Create side-by-side boxplots of the calorie contents of each restaurant. Evaluate the constant variance assumption based on these boxplots. Feel free to reference the standard deviations you computed in Question 4.

```
b1 = small_fastfood[which(small_fastfood$restaurant == "Arbys"), ]$calories
b2 = small_fastfood[which(small_fastfood$restaurant == "Subway"), ]$calories
b3 = small_fastfood[which(small_fastfood$restaurant == "Taco Bell"), ]$calories
boxplot(b1, b2, b3, main = "Calories per Restaurant", ylab = "Calories",
        names = c("Arbys", "Subway", "Taco Bell"))
```



The variance for the mean calories for *Subway* is the largest and much larger than the variance for mean calories for *Taco Bell*. Therefor, the constant variance assumption is not met.

> 8: Using the above code (but filling in the appropriate variable names), carry out an ANOVA test using a significance level of 0.05. Fill in the following table with your results AND state your conclusions.

```
results <- aov(calories ~ restaurant, data = small_fastfood)
summary(results)
```

	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	F-Statistic	P-Value
restaurant	2	352468	176234	3.351	0.0365
residuals	263	13829781	52585		

With a p-value of 0.0365, a f-statistic of 3.351 and a significance level of 0.05, we reject the null hypothesis and conclude, that there is a difference for at least one restaurant in average calories per food items.

> 9: Carry out a pairwise t-test using a Bonferroni correction. Share your results here and if you rejected any of the pairwise t-tests.

```
pairwise.t.test(small_fastfood$calories, small_fastfood$restaurant,
                p.adj = "bonferroni")
```

	Arbys	Subway
Subway	1.0	-
Taco Bell	0.056	0.187

With a significance level of 0.05, we fail to reject any of the pairwise t-tests and conclude that the data is inconclusive. As the Anova-test suggests that there is a difference for at least one restaurant in average calories per food item, the pairwise test should be repeated with a larger sample.

Complete Script for R Assignment 12

```
library(tidyverse)

load("~/R_631/fastfood.rda")
view(fastfood)

# subset data
small_fastfood = filter(fastfood, restaurant == "Arbys" | restaurant == "Subway"
                        | restaurant == "Taco Bell")

# mutate full data set
fastfood = mutate(fastfood, cal_wo_fat = calories - cal_fat)

# subset with piping
subset = fastfood %>%
  select(restaurant, item, calories) %>%
  filter(calories > 1000)

# piping with group_by
small_fastfood %>%
  group_by(restaurant) %>%
  summarise(avg_cal = mean(calories), sd_cal = sd(calories), sample_size =
            length(calories))

# histogram for calories by restaurant
par(mfrow=c(1,3))
hist(small_fastfood[which(small_fastfood$restaurant == "Arbys"), ]$calories,
     main = "Arbys", xlab = "Calories")
hist(small_fastfood[which(small_fastfood$restaurant == "Subway"), ]$calories,
     main = "Subway", xlab = "Calories")
hist(small_fastfood[which(small_fastfood$restaurant == "Taco Bell"), ]$calories,
     main = "Taco Bell", xlab = "Calories")
par(mfrow=c(1,1))


# boxplot for calories by restaurant
b1 = small_fastfood[which(small_fastfood$restaurant == "Arbys"), ]$calories
```

```r
b2 = small_fastfood[which(small_fastfood$restaurant == "Subway"), ]$calories
b3 = small_fastfood[which(small_fastfood$restaurant == "Taco Bell"), ]$calories
boxplot(b1, b2, b3, main = "Calories per Restaurant", ylab = "Calories",
        names = c("Arbys", "Subway", "Taco Bell"))

# anova test
results <- aov(calories ~ restaurant, data = small_fastfood)
summary(results)

# pairwise t-test
pairwise.t.test(small_fastfood$calories, small_fastfood$restaurant,
                p.adj = "bonferroni")
```