MATH 338

EXAM 2 – LAB PORTION

TUESDAY, JULY 18, 2017

Your name: _____**SOLUTIONS**_____

Your scores (to be filled in by Dr. Wynne):

Problem 1: ____/13

Problem 2: ____/6

Total: ____/19

You have 30 minutes to complete this exam. This exam is open book, open notes, open help menus, and open labs.

For full credit, include all R code (if using RStudio), graphs, and output. Save your answers as a .docx or .pdf file and upload the file to Titanium.

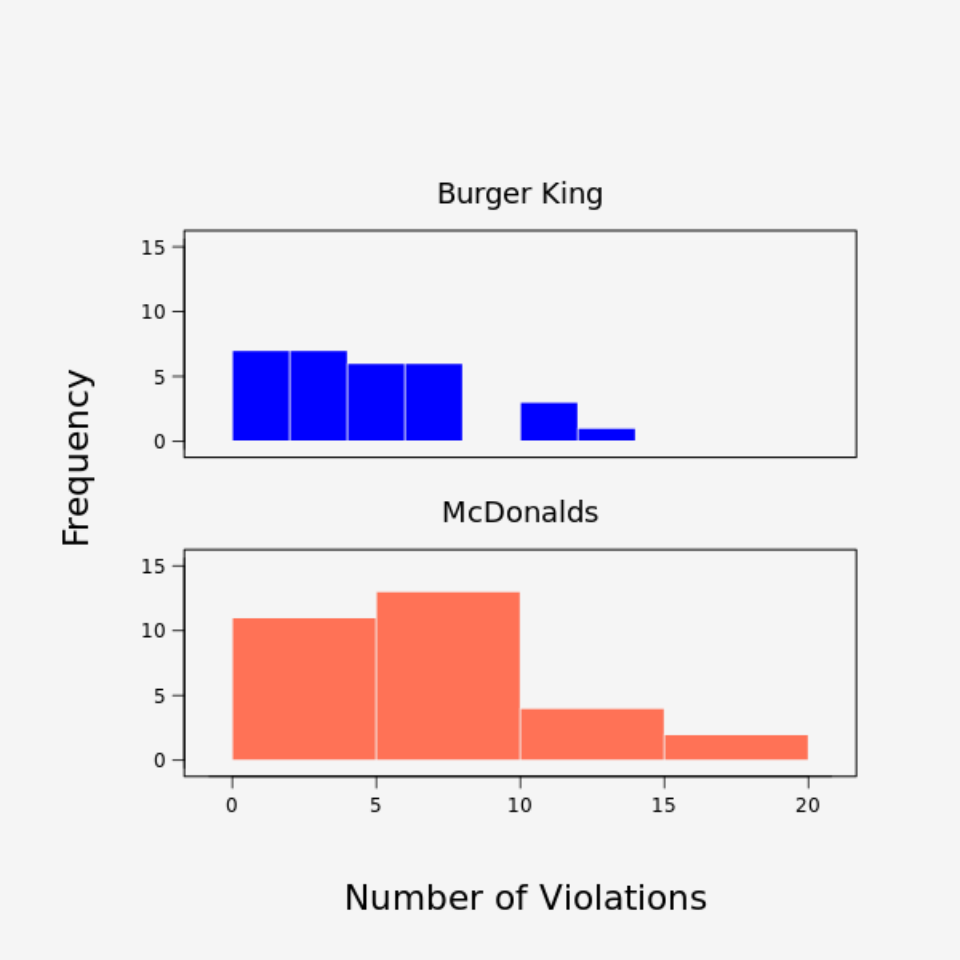
The florida_mcd_bk.csv file on Titanium contains a simple random sample of 30 McDonald's restaurants and 30 Burger King restaurants in the state of Florida, and the number of violations found in each restaurant during a recent inspection. You can assume that there are many more than 30 of each restaurant chain in the state of Florida, and that the restaurants are independent. Upload the dataset to Rguroo, or import the dataset in RStudio.

1. In this problem we will assess the claim that McDonald's and Burger King restaurants have a different mean number of violations.

A) [1 pt] Which type of hypothesis test should be used to assess that claim?

Two independent samples (two-sample) t hypothesis test

B) [4 pts] Create a set of two histograms showing the distribution of the number of violations for each restaurant chain. Make sure that your axes have informative, relevant labels, and that the shape of the distribution is clearly shown in each histogram. Paste the histograms, and any code used to create them, below.



C) [2 pts] Are we okay to perform the hypothesis test from part (A)? Why or why not?

Yes. We are told we have independent simple random samples, so we only need to check whether the sample size is large enough. Although both distributions are clearly skewed to the right, we have a total of 60 observations between the two groups, and our rule of thumb says that in the presence of skew we need at least 40 combined observations to produce a reasonably accurate test. Since 60 > 40, we should be okay to perform t-procedures even though the distributions are skewed.

D) [6 pts] Regardless of your answer to part (C), perform a hypothesis test to assess the claim that McDonald's and Burger King restaurants have a different mean number of violations. Paste all relevant code and output from Rguroo/RStudio, and don't forget to state your conclusion.

H_0: $\mu_{BK} = \mu_{MD}$, H_a: $\mu_{BK} \neq \mu_{MD}$, $\alpha = 0.05$ by default

Test of hypothesis (t-test): 'violations (Burger King) - violations (McDonalds)', Unequal Population Variances

Research Hypothesis H1: Mean of 'violations (Burger King) - violations (McDonalds)' is not equal to 0

Diff Means	Standardized Obs Stat	DF	P-value	95% Lower CL	95% Upper CL
-2.16667	-2.18529	54.6573	0.0331718	-4.15391	-0.179422

Test is significant at 5% level.

The t-statistic is -2.18529 and the corresponding p-value is 0.033. Therefore, the result is significant at the 5% level, and we reject H_0 in favor of H_a. We conclude that McDonald's and Burger King restaurants in Florida do indeed have a different mean number of violations.

I needed to see the table above, or one of the p-value and critical value graphs along with a statement of the t-statistic and the p-value, for full credit.

2. Now we will focus only on McDonald's restaurants (RStudio hint: you can use the **`filter()`** command from the **`dplyr`** package to subset your data set).

A) [5 pts] Report and interpret a 99% confidence interval for the population mean number of violations in McDonald's restaurants in Florida. Paste all relevant code and output from Rguroo/RStudio. (Rguroo hint: you can change the confidence level in the *Details* menu.) (RStudio hints: you will need to use the **`$`** sign to select the appropriate variable; for example, **`CI_df$low`** in Lab 4 selects the variable **`low`** from the **`CI_df`** data frame. Also, in the command you use to create the confidence interval, you will need to include the argument **`conf.level = 0.99`**)

Confidence Interval - t Distribution

99% Confidence interval

Variable	DF	Lower CL	Upper CL	Mean	Margin of Error
violations (McDonalds)	29	5.27513	9.59154	7.43333	2.15821

We are 99% confident that the true population mean number of violations in McDonald's restaurants in Florida is between 5.28 and 9.59.

I needed to see the table above for full credit.

B) [1 pt] If we did part (A) 1000 times (with 1000 different random samples of 30 McDonald's restaurants), what percentage of the resulting confidence intervals would you expect to contain the true population mean number of violations?

We would expect 99% of our 99% confidence intervals to contain the true population mean number of violations.