Name: Helen Gutierrez

UmichID: 6337-1628

What dataset are you working with: US\_births\_2000\_2014

List 3 questions that you can ask with your dataset.

Q1: Are the number of births on the 13th of each month less than the 14th of each month?

Q2: Are the number of births on Friday the 13th less than the number of births on all other Fridays?

List the associated null hypothesis for each question:

Q1: The births on the 13th of each month are higher than the births on the 14th of each month.

Q2: The number of births on Friday the 13th are less than the number of births on all other Fridays.

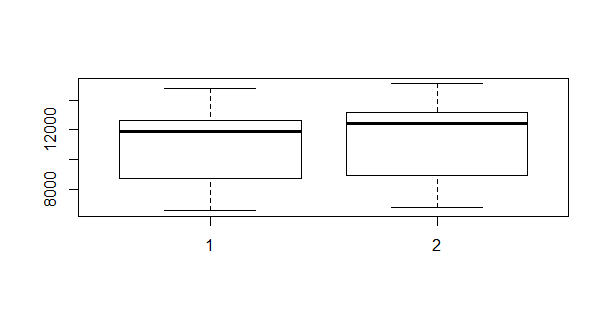
What statistical test(s) will you use to answer each of the questions:

Q1: Two sample, One tailed t-test

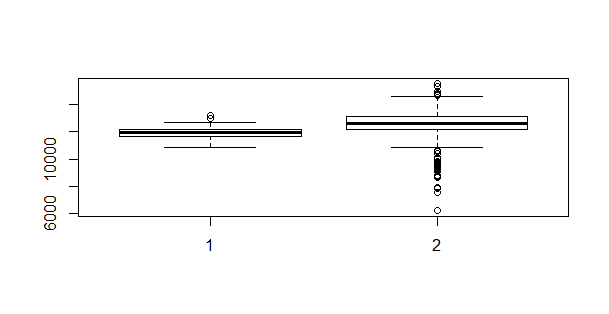
Q2: One sample, one-tailed t-test

Make a visual plot showing the relationship that you will analyze statistically (e.g. boxplot for t-test or ANOVA; scatterplot for regression; table for chi-square).

Q1:



Q2:



Do your data meet the assumptions required for the statistical test you want to run? Please state the assumptions you examined and whether or not your data meet those assumptions:

Q1: I tested for normality and variance. To test normality, I ran the Shapiro test and also made a QQplot, and for variance I ran a variance (F-test) test. The data is not normal, however, the sample size was large, so I assumed normality. The F-test indicates that there is no variance between samples.

Q2: I tested for normality and variance. To test normality, I ran the Shapiro test and also made a QQplot, and for variance I ran a variance (F-test) test. The data is almost normal, and because the sample size is large, I assumed normality. The F-test indicates that there is no variance between samples

Run the statistical test! Put your results here:

Q1:

Welch Two Sample t-test

data: thirteenth[, "births"] and not.13[, "births"]

t = -1.8096, df = 356.83, p-value = 0.0712

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-883.7129 36.7462

sample estimates:

mean of x mean of y

11111.47 11534.95

Q2:

Welch Two Sample t-test

data: fri.13[, "births"] and friday[, "births"]

t = -5.3283, df = 28.625, p-value = 1.059e-05

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-894.3832 -398.0216

sample estimates:

mean of x mean of y

11949.96 12596.16

Interpret your results!

Q1: H0 with 95% significance is not rejected , because the p-value is 0.07. This

#Interpretation of test results:

#H0 is rejected with a p-value of 0.00014

# the average violent crimes in CA per 1000000 people is 877.48

# The upper limit of the CI is 926.17, which tells us that the true mean lies below this limit and is less than 1000.

Q2:

Q3: