Test #1



The scatterplot shows how the two variables aren’t strongly rated, but they do have some relation when the average quality rating of the instructor went up the number of publications the instructor had went up slightly too.

b. Commands: TeachResearchDataSet<-read.csv("C:/Users/kurti/OneDrive/Documents/R/TeachResearchDataSet.csv", header = TRUE)

(Entered Data Manually through Excel)

c. Commands: line<-ggplot(TeachResearchDataSet, aes(class\_pubs, rating\_1))

line + stat\_summary(fun.y = mean, geom = "point") + stat\_summary(fun.y = mean, geom = "line", aes(group = 1), colour = "Blue", linetype = "dashed") + stat\_summary(fun.data = mean\_cl\_boot, geom = "errorbar", width = 0.2) + labs(x = "Publication Class", y = "Average Quality Rating of the Instructor")



The line graph shows that the mean quality rating of instructors increases as the publication class goes from Low to Medium to High. However, the error bar charts shows the average quality ratings between the publication classes overlaps quite a lot.

2.

a. Categorical variable

b. Continuous, interval variable

c. Continuous, ratio variable

d. Continuous, ratio variable

e. Categorical, binary variable

f. Categorical, ordinal variable

3. Commands: bigMacData<-read.csv("C:/Users/kurti/OneDrive/Documents/R/BigMacData.csv", header = TRUE)

stat.desc(bigMacData$time, basic = TRUE, norm = TRUE)

a. mean = 671.05s , standard deviation = 88.527s , median = 672s , minimum = 478s , maximum = 889s , upper quartile = 729.5s , lower quartile = 601s

(Counted the median because the true median is the average of 672 + 672 which is of course 672. So both 672’s are relevant to the lower and upper quartile.)

b. 95% CI upper boundary = 699.36s

lower boundary = 642.74s

c. Smallest Data Value Deviance = -193.05s

Deviance of Smallest Magnitude = -.95s

d. Error Bar Chart

Commands: bigMbar<-ggplot(bigMacData, aes(weight, time))

bigMbar + stat\_summary(fun.y = mean, geom = "bar", fill = "White", colour = "Black") + stat\_summary(fun.data = mean\_cl\_normal, geom = "errorbar", position = position\_dodge(width=0.90), width = 0.4) + labs(x = "Weight Type of Person", y = "Time in Seconds")



The error bar chart clearly shows that overweight people ate Big Mac value meals faster than normal weight people. The error bars do not overlap each other either, meaning the means relating to the two groups of people would most likely not overlap in future observations.

e. Simple Boxplot

bigBoxplot<-ggplot(bigMacData, aes(weight, time))

bigBoxplot + geom\_boxplot() + labs(x = "Weight Type of Person", y = "Time in Seconds")



The boxplot shows that overweight people certainly ate faster than normal weight people within this dataset, which reaffirms the error bar chart.

f. Commands for f.: nonFattyData<-subset(bigMacData, bigMacData$weight=="normal weight")

fattyData<-subset(bigMacData, bigMacData$weight=="overweight")

nonFattyHist<-ggplot(nonFattyData, aes(time)) + geom\_histogram(aes(y = ..density..), colour = "black", fill = "white", binwidth = 35) + labs(x = "Time to Eat Mac Meal in Seconds", y = "Density") + stat\_function(fun=dnorm, args=list(mean = mean(nonFattyData$time, na.rm = TRUE), sd = sd(nonFattyData$time, na.rm = TRUE)), colour = "red", size=1)

fattyHist<-ggplot(fattyData, aes(time)) + geom\_histogram(aes(y = ..density..), colour = "black", fill = "white", binwidth = 35) + labs(x = "Time to Eat Mac Meal in Seconds", y = "Density") + stat\_function(fun=dnorm, args=list(mean = mean(fattyData$time, na.rm = TRUE), sd = sd(fattyData$time, na.rm = TRUE)), colour = "red", size=1)

qqNotF<-qplot(sample=nonFattyData$time)

qqFatt<-qplot(sample=fattyData$time)

i. Histogram Normal Weight



The histogram of normal weight people indicates a slightly negative skew, which means there was a slightly higher concentration of normal weight people eating slower than the rest of the group.

Q-Q Plot Normal Weight



The Q-Q Plot of normal weight people shows a slightly skewed dataset with the plot deviating from a diagonal normal distribution with some twists and turns in it. Also, notice the two outliers within the group that don’t really affect much, but they are still interesting to note.

Histogram Overweight



The histogram of overweight people indicates a slightly positive skew, which means there was a slightly higher concentration of overweight people eating faster than the rest of the group.

Q-Q Plot Overweight



The Q-Q Plot shows the overweight data is skewed, but I think the sample size for the plot is really too small to make any significant decision on it.

ii. Commands: by(bigMacData[, c("time")], bigMacData$weight, stat.desc, basic = FALSE, norm = TRUE)

Normal Weight Group

Kurtosis: .271 Kurtosis/2SE: .163

Skew: -.2102 Skew/2SE: -.246

Both the skew and kurtosis divided by two standard errors are greater than one which means there is not significant skew or kurtosis.

Overweight Group

Kurtosis: -1.254 Kurtosis/2SE: -.469

Skew: .458 Skew/2SE: .333

Both the skew and kurtosis divided by two standard errors are greater than one which means there is not significant skew or kurtosis.

iii. Commands: by(bigMacData$time, bigMacData$weight, shapiro.test)

(Normal Weight p=.871, Overweight p=.258)

Shapiro-Wilk Test

The normal weight big mac meal eaters, W = .982, p>.05, the overweight big mac meal eaters, W = .907, p>.05.

Both p-values are above .05 which indicates that both datasets of the food consumers are normally distributed.

Overall, after completing all the tests associated with normality within the data, the tests suggest that the datasets of the two groups of food consumers are normally distributed.

g. Commands: leveneTest(bigMacData$time, bigMacData$weight, center = mean)

i. Levene’s Test for Homogeneity of Variance (center = mean)

For the speed at which the two types of people ate, the variances were similar for both overweight and normal weight people, F(1, 38) = 2.74, with p=.1058. The high p-value indicates that variances of the two types of people are not significantly different.

ii.

Normal Weight Variance = 6,880.52

Overweight Variance = 1,816

Variance Ratio: 6,880.52/1,816

= 3.79

This Fmax implies that variances between the two types of people is borderline between being significantly different and not significantly different. However, I will side with the variances not being significantly different because the Levene Test suggested that the variances are not significantly different.

Overall, the variance tests indicate that the variances between the two types of food consumers are not significantly different from each other.