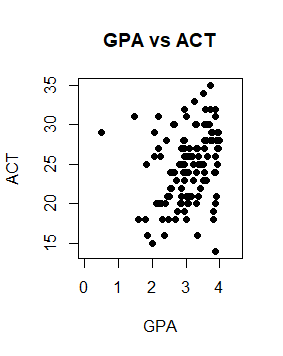
**Mini Project #4**

**Harrison Jansma**

**April 11,2019**

**Section 1 Analysis**



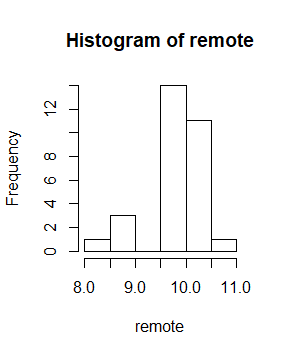
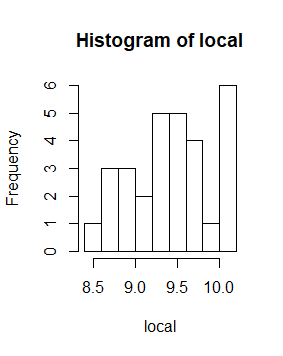
There seems to be a positive linear relationship between ACT scores and first semester GPA. If ρ is the population correlation coefficient between GPA and ACT, then a point estimate of ρ based on our data is . Bootstrapped bias and standard error estimates for are -0.013 and 0.115 respectively. Furthermore, based on a bootstrap resampling, we compute the following estimates.

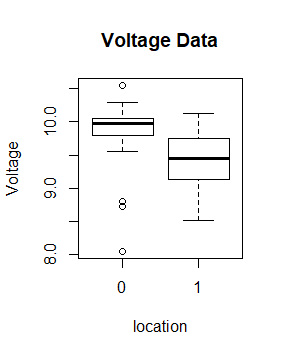
Std Error of : 0.1145

Bias of :

95% confidence interval for : [0.0169, 0.5029].

The data seems to suggest a positive correlation between GPA and ACT, but this may be due to random chance. More data is needed to determine how strongly the two variables are correlated. At the very least, we know from this data that the two variables are highly likely (95%) to be positively correlated.





Note that remote and local are indicated as 0 and 1, respectively. From the above visuals, one could come to the conclusion that the distribution of voltage at the remote location is different than the voltage distribution at the local facility. The remote facility appears to have a less varied distribution, but this is probably due to the small sample size.

To test if there is no difference between the mean voltage of the two populations, we will be performing a two-sample t-test. Before we begin, we must check if a few conditions hold. The two assumptions we must check are as follows.

1. Samples are independent, with random sampling. -confirmed by problem statement.
2. Sampling distribution is approximately normal. -Large sample size is evidence of normality.

With both assumptions verified, we will conduct the test. Our null hypothesis being that the two distributions are the same , and the alternative being the opposite. That the means are not equal. We will have a confidence level of 95%. We will use a 95% confidence interval of to determine if these two values are the same. If the confidence interval contains 0, we cannot reject the null hypothesis.

The following 95% confidence interval was obtained from our data.

[0.1115719 0.6510948]

Since this confidence interval does not contain zero, we must reject the null hypothesis and conclude that the two distributions are different. Thus, we cannot establish the manufacturing process locally.

This conclusion matches the observations from the exploratory data analysis.

3.

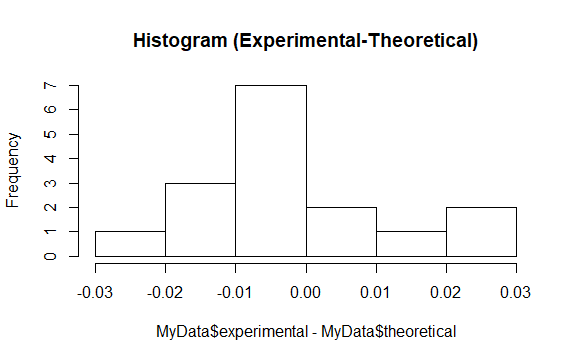
The dataset contains 16 observations of experimental and theoretical vapor pressures at different temperatures. Because each observation of vapor pressure is linked to a temperature, we will be performing a paired t-test of theoretical and experimental data paired by temperature. We will be testing whether the two distributions are different (two-sided test) at a 95% confidence level.

• The null hypothesis (H0) assumes that the true mean difference is equal to zero.

• The two-tailed alternative hypothesis (H1) assumes that μd is not equal to zero.

If we find that there is little difference between both distributions’ means, we can conclude that the theoretical model is sound.

Before we can perform a statistical test, we must verify that a t-test would be appropriate for this data. First, we need to check that the distribution of the difference between the two samples is normally distributed. Normality can be checked with a histogram of the data.



From the above histogram, we can justify that the random variable is normally distributed. We also can tell from a quick inspection of the data that the data is measured on a continuous scale, and that there are no major outliers. Lastly, a paired test is appropriate because observations in the data can be paired by temperature.

Now that assumptions have been checked and a t-test seems appropriate, lets work on constructing a confidence interval for . If a 95% confidence interval contains 0, we can infer that the theoretical model is valid based on experimental observations.

Based on our data, we calculate that a 95% confidence interval for is,

[-0.315711 0.314336]

Based on this we can infer that the experimental and theoretical distributions are not different, and therefore the theoretical model may be a sound representation of the natural process.

**Section 2 Code**

library(boot)

###############################################################

#QUESTION 1

#Reading in the data

MyData <- read.csv(file="C:/Users/Harrison/Desktop/gpa.csv", header=TRUE, sep=",")

attach(MyData)

#Scatterplot + Correlation

plot(gpa, act, main="GPA vs ACT", xlab="GPA", ylab="ACT", pch=19, xlim=c(0,4.5))

cor(gpa,act)

#Bootstrapping

correl = function(data, indices){

return(cor(data$gpa[indices], data$act[indices]))

}

results = boot(data=MyData,statistic = correl,R=10000)

plot(results)

#Bootstrapped StdError

sd(results$t)

#Bootstrapped bias

print(mean(results$t-results$t0))

#Bootsrap Percentile ConfInterval

boot.ci(results, type="perc")

#####################################

# QUESTION 2

MyData <- read.csv(file="C:/Users/Harrison/Desktop/MP4/VOLTAGE.csv", header=TRUE, sep=",")

#Seperating the data into remote/local

remote=MyData[1:30,]$voltage

local = MyData[31:60,]$voltage

#Visuals to compare distributions

hist(remote)

hist(local)

boxplot(voltage~location,data=MyData, main="Voltage Data",

xlab="location", ylab="Voltage")

#Normality Check

hist(remote-local)

#Confidence Interval

stderror = sqrt((var(remote)/29)+(var(local)/29))

mu = mean(remote-local)

c(mu+qt(.025,29)\*stderror,mu+qt(.975,29)\*stderror)

##################################################

# QUESTION 3

MyData <- read.csv(file="C:/Users/Harrison/Desktop/MP4/vapor.csv", header=TRUE, sep=",")

# Normality check

diffs = MyData$experimental-MyData$theoretical

hist(diffs,main="Histogram (Experimental-Theoretical)")

# Confidence interval calculation

mu = mean(diffs)

stderror = sqrt((var(MyData$experimental)/15)+(var(MyData$theoretical)/15))

c(mu+qt(.025,15)\*stderror,mu+qt(.975,15)\*stderror)