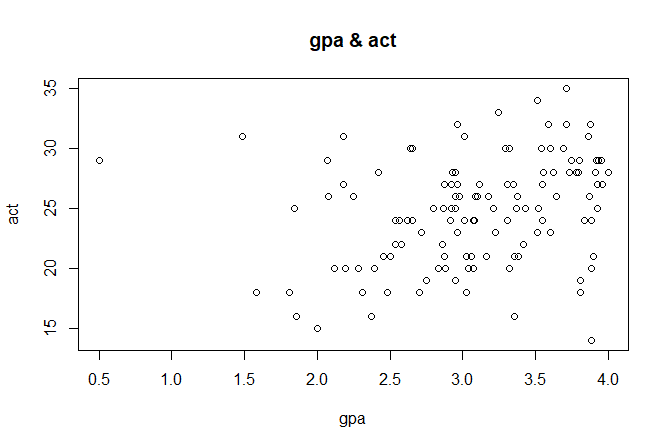
**Q1.** Scatterplot of ‘gpa’ against ‘act’ can be drawn using the plot function as below:

plot(gpa$gpa, gpa$act, main="gpa & act", xlab = "gpa", ylab = "act")



From the plot seen above, we can say that there appears to be positive(not so strong) covariance & correlation among ‘gpa’ and ‘act’ since the slope seems to be positive.

Sample correlation, which is an estimator of population correlation, can be calculated as below:

cor(gpa$gpa, gpa$act)

[1] 0.2694818

Now, using bootstrap method to find point estimate of correlation, bias and its standard error. We write the functions in R as below:

#Below function calculates and returns the correlation for each resample

>cor.npar <- function(x, indices){

result <- cor(x$gpa[indices], x$act[indices])

return (result)

}

#using the boot function to get bootstrap values for correlation

> set.seed(123)

> cor.npar.boot

ORDINARY NONPARAMETRIC BOOTSTRAP

Call:

boot(data = gpa, statistic = cor.npar, R = 999, sim = "ordinary",

stype = "i")

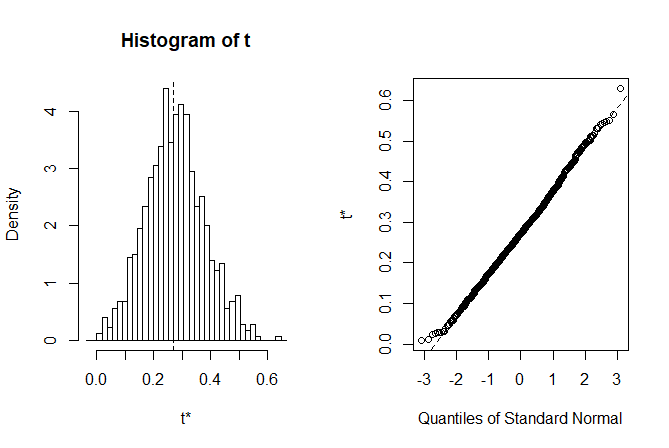
Bootstrap Statistics :

original bias std. error

t1\* 0.2694818 0.005322198 0.103147

So, we can see that the bootstrap estimate for correlation is 0.2694818, for bias is 0.005322198 and for standard error is 0.103147. Bootstrap distribution of correlation estimate can be obtained as below:

>plot(cor.npar.boot)



95% confidence interval can be obtained using the boot.ci function in R, as below:

> boot.ci(cor.npar.boot)

BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS

Based on 999 bootstrap replicates

CALL :

boot.ci(boot.out = cor.npar.boot)

Intervals :

Level Normal Basic

95% ( 0.0620, 0.4663 ) ( 0.0491, 0.4627 )

Level Percentile BCa

95% ( 0.0762, 0.4899 ) ( 0.0567, 0.4753 )

Calculations and Intervals on Original Scale

We can also calculate 95% confidence interval using percentile bootstrap as below:

> sort(cor.npar.boot$t)[c(25,975)]

[1] 0.0762461 0.4899069

So, 95% confidence interval using percentile bootstrap is (0.0762, 0.4899).

**Q2. a)** We can do the exploratory analysis by using the boxplot and histogram for the voltages at remote location and voltages at local locations. Below code shows how we do it in R:

> vol <- read.csv(file = "VOLTAGE.csv", header=TRUE, sep=",")

> vol\_remote <- vol$voltage[vol$location == 0]

> vol\_local <- vol$voltage[vol$location == 1]

> boxplot(vol\_remote,vol\_local)

Figure 2.1 shows the result of the boxplot, left one for voltages at remote location. We can see that the two distributions are not similar (different variance, median). Also, the boxplot for voltages at remote location is left skewed. Below we see how to get QQplots for the distributions of voltages at remote and local location:

> qqnorm(vol\_remote,main="QQ-norm for remote voltage")

> qqline(vol\_remote)

> qqnorm(vol\_local,main="QQ-norm for local voltage")

> qqline(vol\_local)

Figure 2.2 and 2.3 show QQplots for the remote and local voltages respectively, which again verifies dis-similarity.

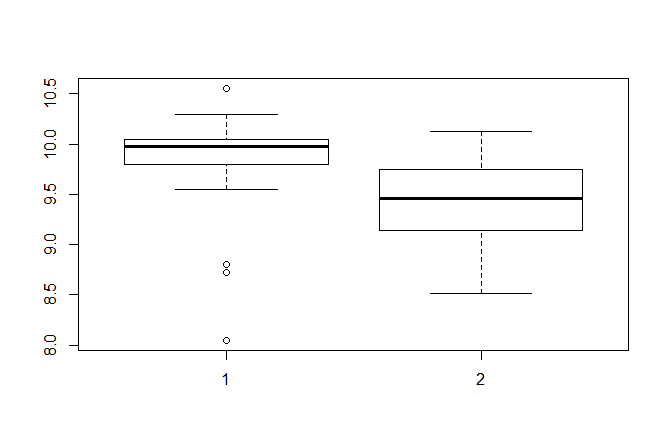


Figure 2.1

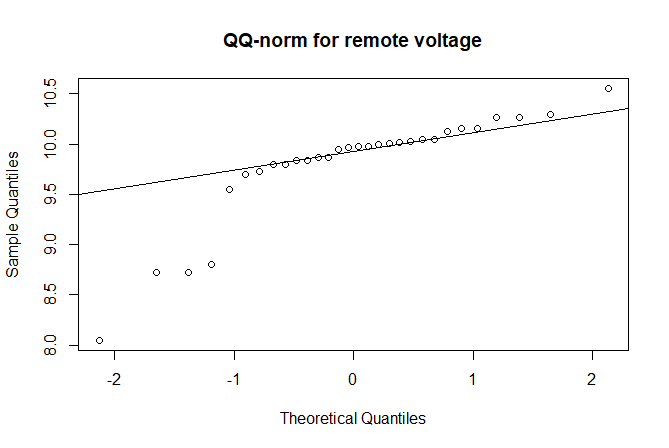


Figure 2.2

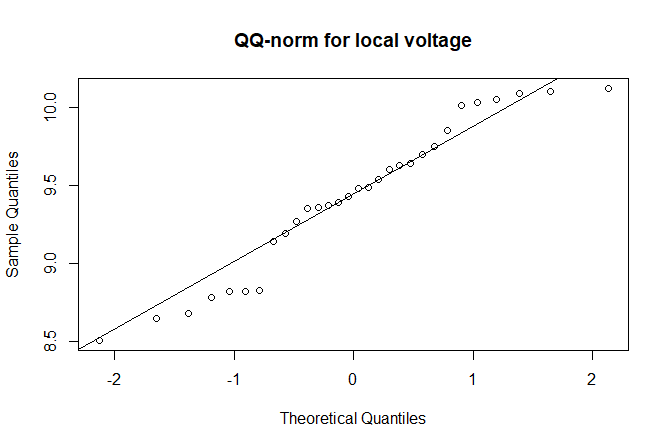


Figure 2.3

**b)** Now we have the case where the two population variances are not equal (verified below). So, we find the confidence interval for the difference of population means of voltages at the two locations. We are assuming the normal distribution and hence will be using the Welch’s two sample t-test for finding the CI as below:

> var(vol\_local)

[1] 0.229322

> var(vol\_remote)

[1] 0.2925895

>t.test(vol\_remote, vol\_local, alternative = "two.sided", conf.level = 0.95,

+ var.equal = FALSE)

Welch Two Sample t-test

data: vol\_remote and vol\_local

t = 2.8911, df = 57.16, p-value = 0.005419

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

95 percent confidence interval:

0.1172284 0.6454382

sample estimates:

mean of x mean of y

9.803667 9.422333

Since zero doesn’t lie in calculated confidence interval above, we can say that the difference in population means of voltages at two locations will not be zero. Hence, the manufacturing process cannot be established locally.

To verifying the assumptions, we can do the manual calculations of the confidence interval for difference of population means in case of normal distributions, as below:

> ci <- (mean\_vol\_remote - mean\_vol\_local) +c(-1,1)\*qt(0.025,58)\*sqrt((var(vol\_local) + var(vol\_remote))/30)

> ci

[1] 0.6453556 0.1173110

So, the assumption seems to be correct.

**c)** Our analysis in a) showed the two distributions of voltages are not similar, which could lead to the estimate that the population means would be different. That is exactly what we get from part b), so, the manufacturing cannot be established locally.

**Q3.** The analysis of data (vapor.dat) provided on theoretical (calculated) and experimental values of the vapor pressure

for dibenzothiophene, a heterocycloaromatic compound similar to those found in coal tar, at given values of

temperature is as follows:

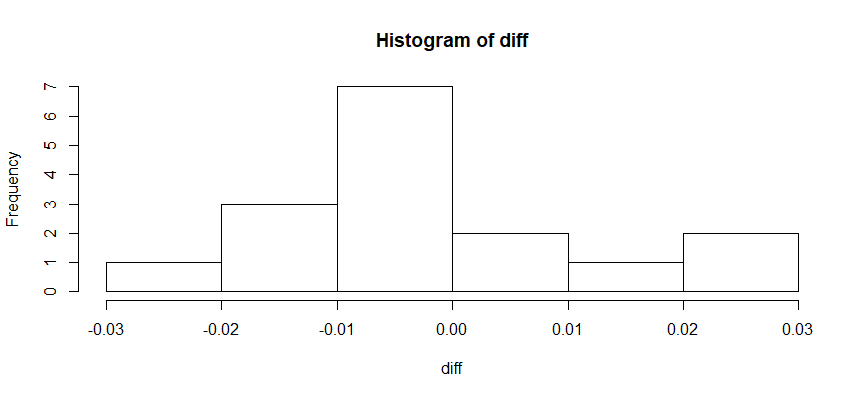
> vaporData <- read.csv(file="vapor.csv", header=TRUE, sep=",")

> diff <- vaporData$theoretical - vaporData$experimental

In the problem, we are interested in the difference between the experimental values and the calculated theoretical

values at given set of temperatures. Thus, we calculate the difference data from the given data as above (like a paired sample).

> hist(diff)



From the histogram we can draw inference that we can assume normal distribution for the difference data. This is the case of two variable theoretical and experimental values calculated upon same set of temperatures.

Thus, it is the case of paired sampling. We will now calculate the confidence interval of the mean of the difference in

the data observation in both the cases and if we get an interval that includes zero in it then we can say that the theoretical model for vapor pressure is a good model of reality.

> ci <- mean(diff) + c(1,-1)\*qt(0.975,15)\*(sd(diff)/sqrt(16))

> ci

[1] -0.006887694 0.008262694

We can see from the R snippet that we get confidence interval which includes 0 in it. Thus, we can say that the theoretical model is a good model of reality.