Assignment 9 (S-520)

FNU Anirudh

November 5, 2015

1.

(a)   The experimental unit is a person. i.e. aerobic students.

(b)   The experimental units belong to one population, i.e., aerobic students. (1- sample location)

(c)   Two measurements were taken on each experimental unit:
i.   Number of watts expended during protocol S (30-minute ride on the first week)
ii.   Number of watts expended during protocol D (30-minute ride on the second week)

(d) Let Si be the score on protocol S for student i, and let Di denote score on protocol D for student i.м

Then, Xi = Di -Si is the random variable of interest. We are interested on drawing inferences about м.

(e) м > 0 iff Di > Si. Thus, to test the theory in favor of dynamic stretches we might want to test H0 : м≤ 0

vs. H1 : м> 0.

2.C-1: 2-sample location problem

(a)   The experimental unit is a middle-aged man.

(b) The experimental units belong to one of two populations:

i. Type A heavy men.

ii. Type B heavy men.

(c) One measurement (cholesterol level) were taken on each experimental unit.

(d) Let Xi denote the cholesterol level for man i (Type A).

Let Yj denote the cholesterol level for man j (Type B).

Then, $X_1; X_2\ldots; X_{n1} \sim P_1; Y_1; Y_2\ldots..; Y_{n2} \sim P_2$.

We are interested on drawing inferences about $\Delta = \mu_1 - \mu_2$

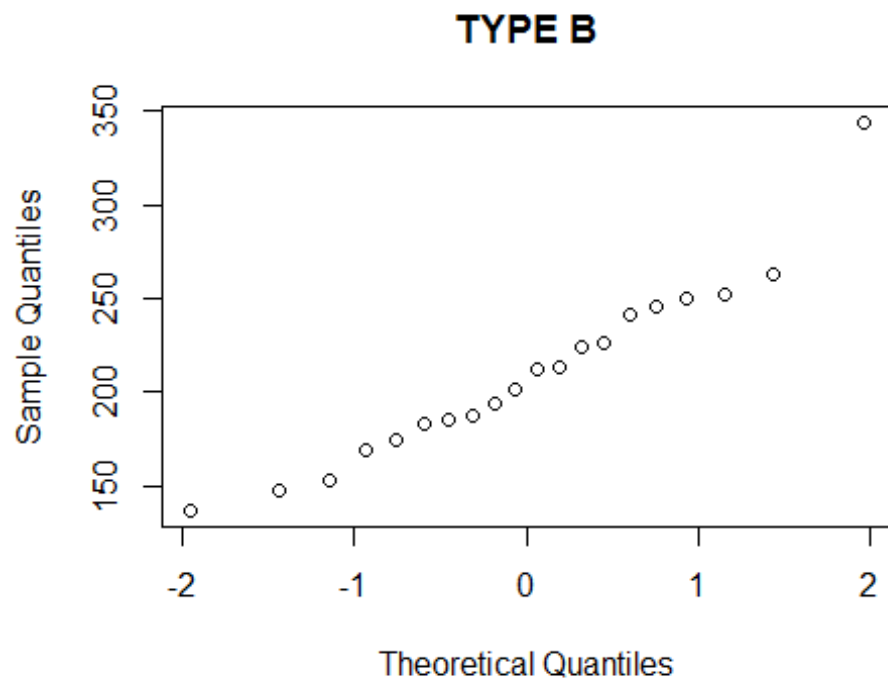(e) $\Delta > 0$ iff $\mu_1 > \mu_2$. Thus, to document that Type A have higher cholesterol than Type B, we might want

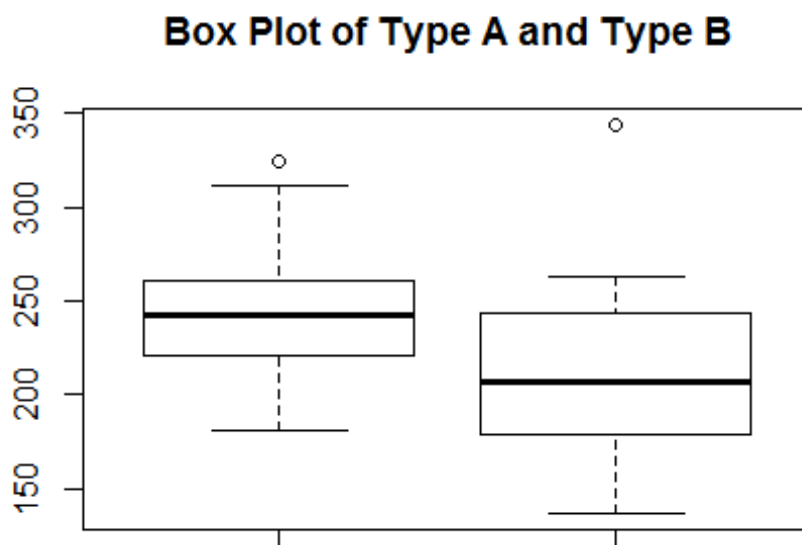to test $H_0 : \Delta \le 0$ vs. $H_a : \Delta > 0$.

C2

```
typea<- c(233,291,312,250,246,197,268,224,239,239,254,276,234,181,248,252,
          202,218,212,325)
typeb<- c(344,185,263,246,224,212,188,250,148,169,226,175,242,252,153,183,
          137,202,194,213)
qqnorm(typea,main = "TYPE A")
```
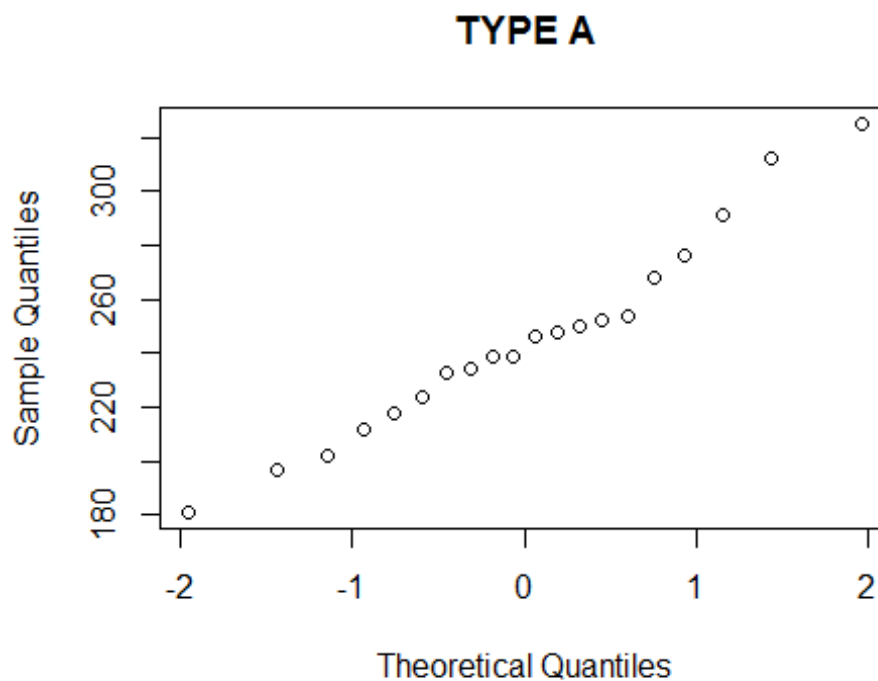

TYPE A

```
qqnorm(typeb,main = "TYPE B")
```

TYPE B


```r
boxplot(typea,typeb,main="Box Plot of Type A and Type B")
```

Box Plot of Type A and Type B


```r
# QQplot for both Type A and Type B suggests some values may be inconsistent
# with normal distribution specially largest in each set as seen in boxplot.
```

```r
a=IQR(typea)/sqrt(var(typea))
a
```

```
## [1] 0.9552842
```

```r
b=IQR(typeb)/sqrt(var(typeb))
b
```

```
## [1] 1.282584
```

```r
# Ratio for Type B suggest sample more close to normal distribution but
# also has large outlier hence I would not assume data was drawn from
# normal distribution although there is slight chance of being picked up
# from normal distribution.
delta<- mean(typea)- mean(typeb)
n1=length(typea)
n2=length(typeb)
va=var(typea)/n1
vb=var(typeb)/n2
se=sqrt(va+vb)
nu<- (va+vb)^2/(va^2/(n1-1)+vb^2/(n2-1))
# If we let alpha= 0.05 then
1- pt(2.5621,nu)
```

```
## [1] 0.00740548
```

```r
# 0.007405 < 0.05 = alpha -> reject H0
# b) We want 90% confidence interval for delta,
qt=qt(0.95,nu)
lower=delta-qt*se
upper=delta+qt*se
lower
```

```
## [1] 11.84155
```

```r
upper
```

```
## [1] 57.65845
```

3.

Let μ_d be the mean urinary β-thromboglobulin excretion in diabetic patients.
Let μ_n be the mean urinary β-thromboglobulin excretion in normal patients.
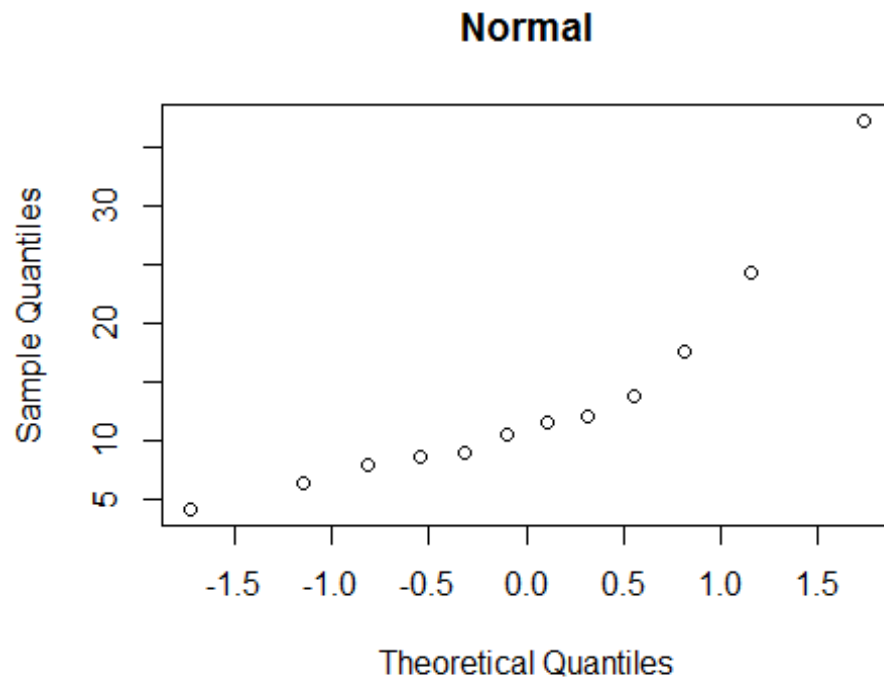$\delta = \mu_d - \mu_n$ be the difference in mean urinary β-thromboglobulin excretion of diabetic and normal patients respectively.

Hypothesis Test
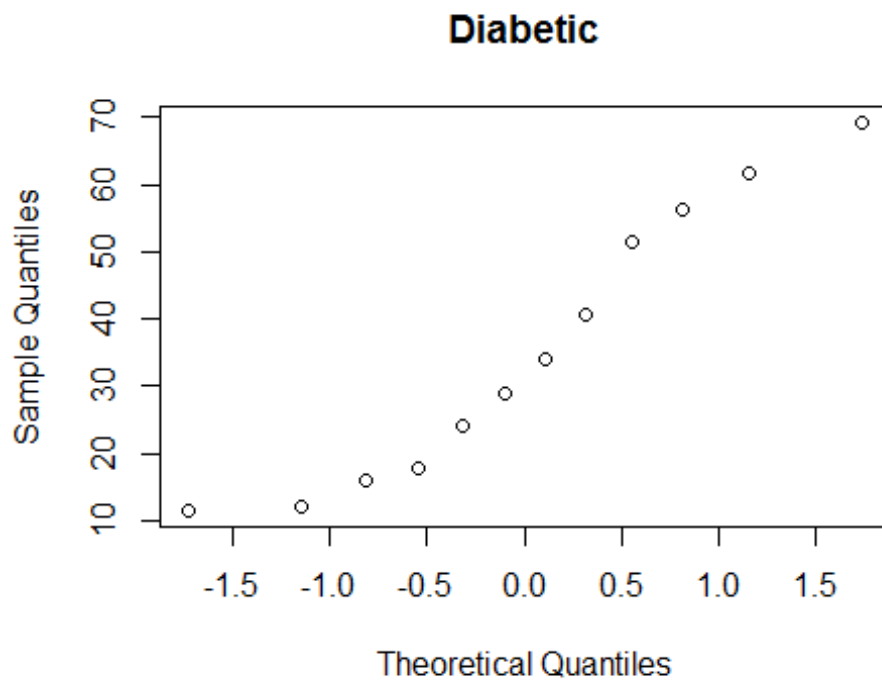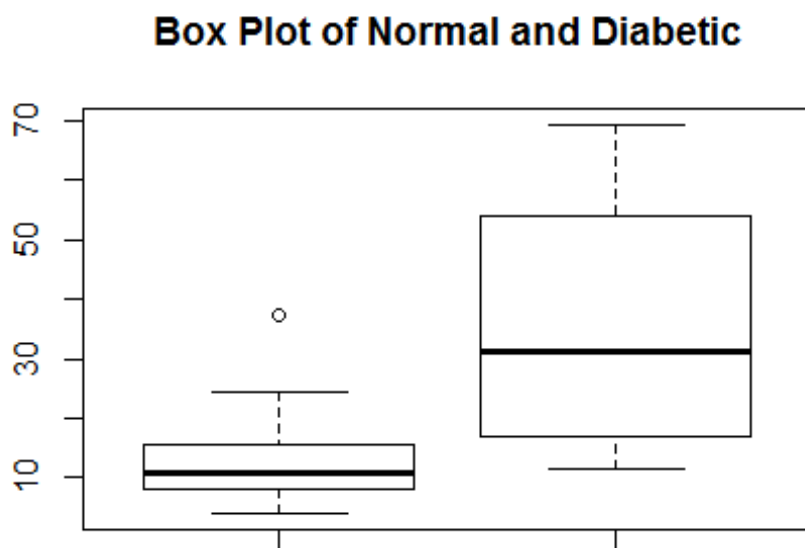$H_0 : \delta <= 0$
$H_a : \delta > 0$

```
normal<- c(4.1,6.3,7.8,8.5,8.9,10.4,11.5,12.0,13.8,17.6,24.3,37.2)
diabetic<- c(11.5,12.1,16.1,17.8,24.0,28.8,33.9,40.7,51.3,56.2,61.7,69.2)
qqnorm(normal,main="Normal")
```
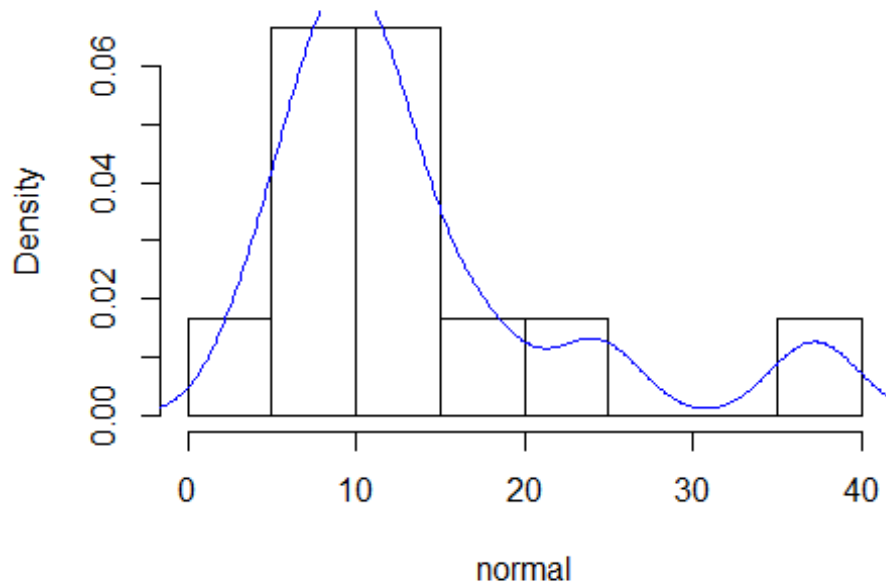
Normal


```
qqnorm(diabetic,main = "Diabetic")
```

Diabetic


```
boxplot(normal,diabetic,main="Box Plot of Normal and Diabetic")
```
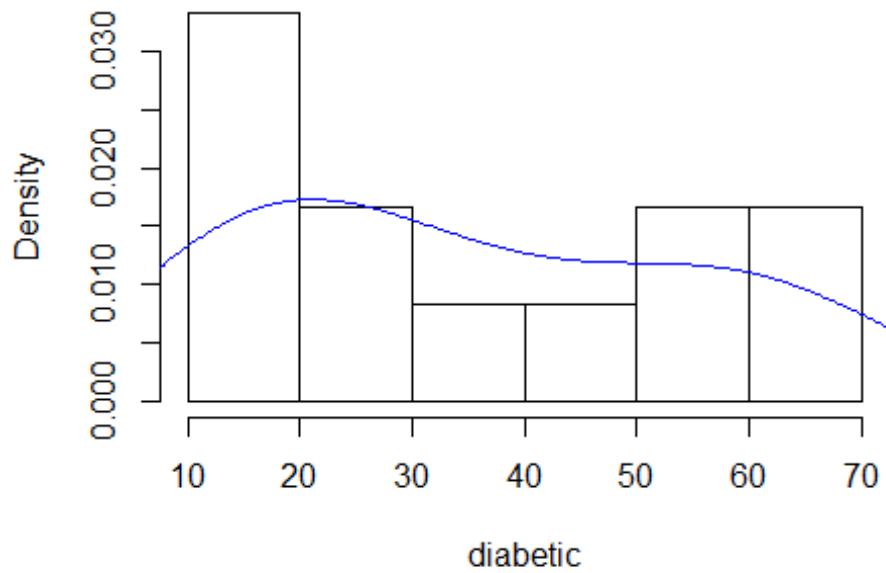
Box Plot of Normal and Diabetic


```
hist(normal,prob=TRUE)
lines(density(normal),col="blue")
```
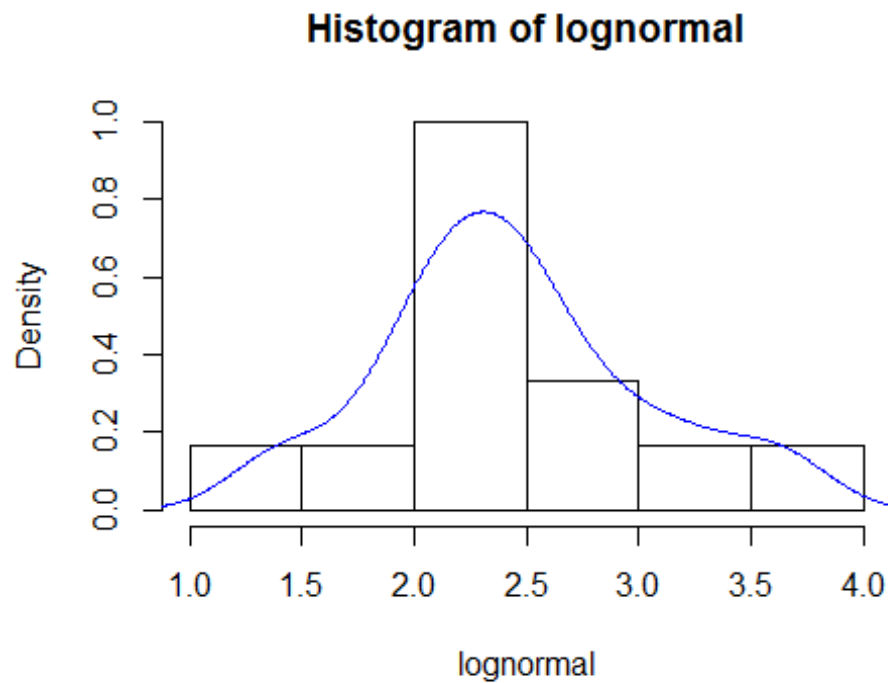
Histogram of normal


```r
hist(diabetic,prob=TRUE)
lines(density(diabetic),col="blue")
```
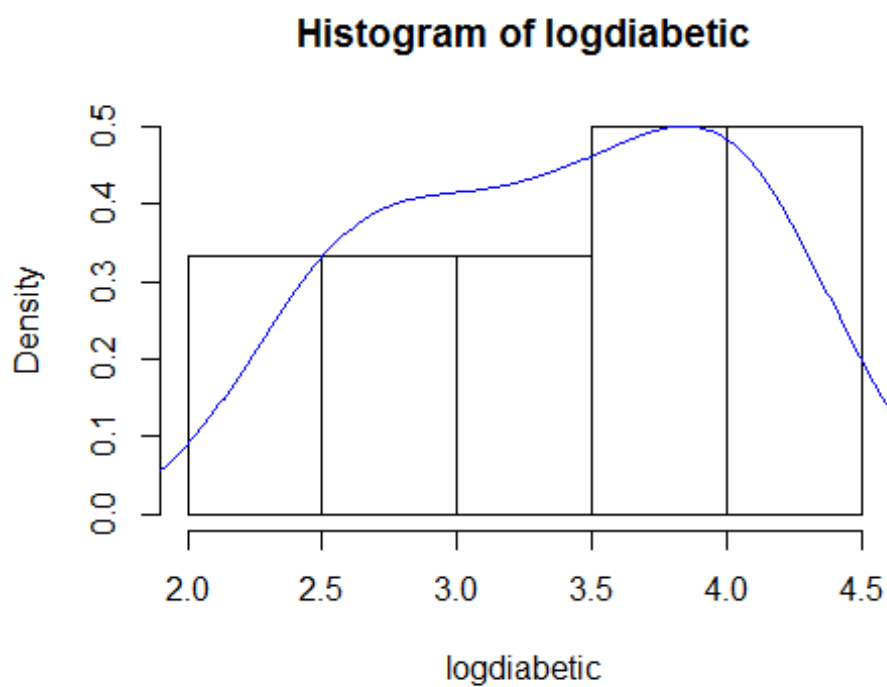
Histogram of diabetic

```
# 1 )
#After seeing qqplot,boxplot and histogram we can say that samples are
# not drawn from normal distribution.
# 2)
# (a) Natural Logarithm
lognormal<- log(normal)
logdiabetic<- log(diabetic)
hist(lognormal,prob=TRUE)
lines(density(lognormal),col="blue")
```
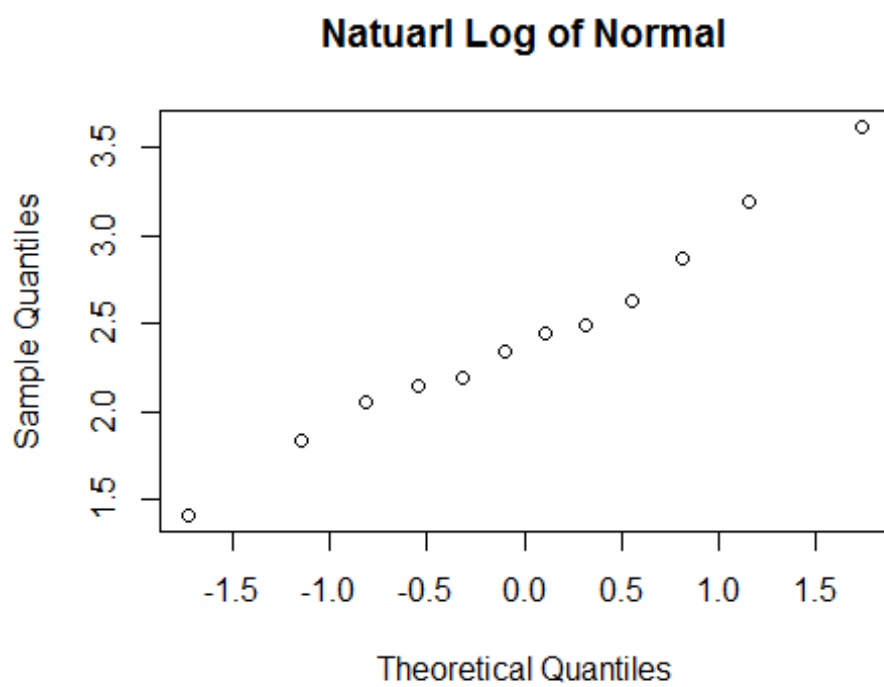
Histogram of lognormal


```
hist(logdiabetic,prob=TRUE)
lines(density(logdiabetic),col="blue")
```

Histogram of logdiabetic


```
qqnorm(lognormal,main="Natuarl Log of Normal")
```

Natuarl Log of Normal


```
qqnorm(logdiabetic,main ="Natural Log of Diabetic")
```
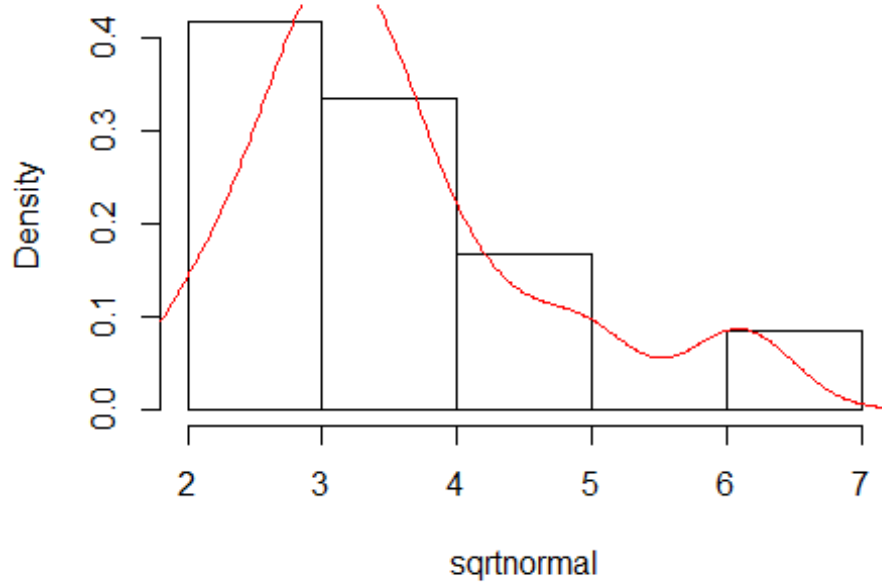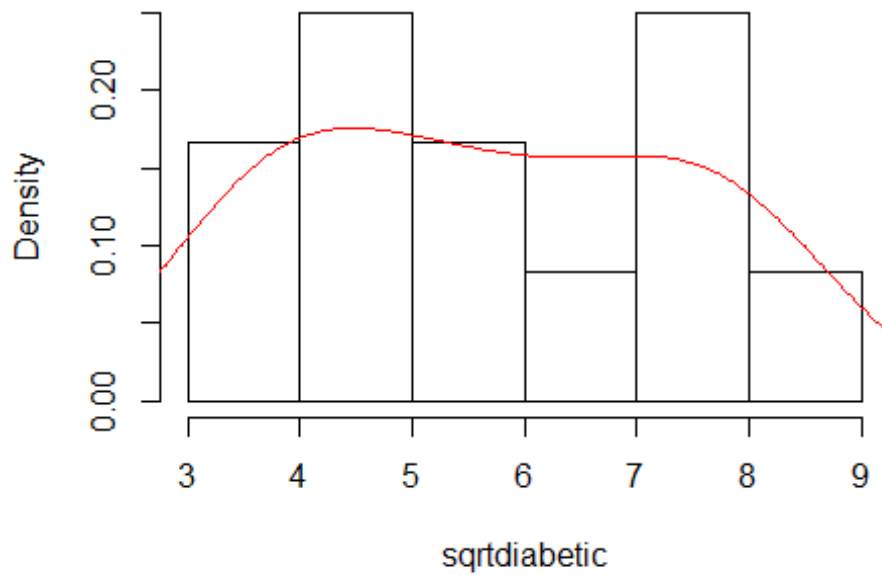
Natural Log of Diabetic
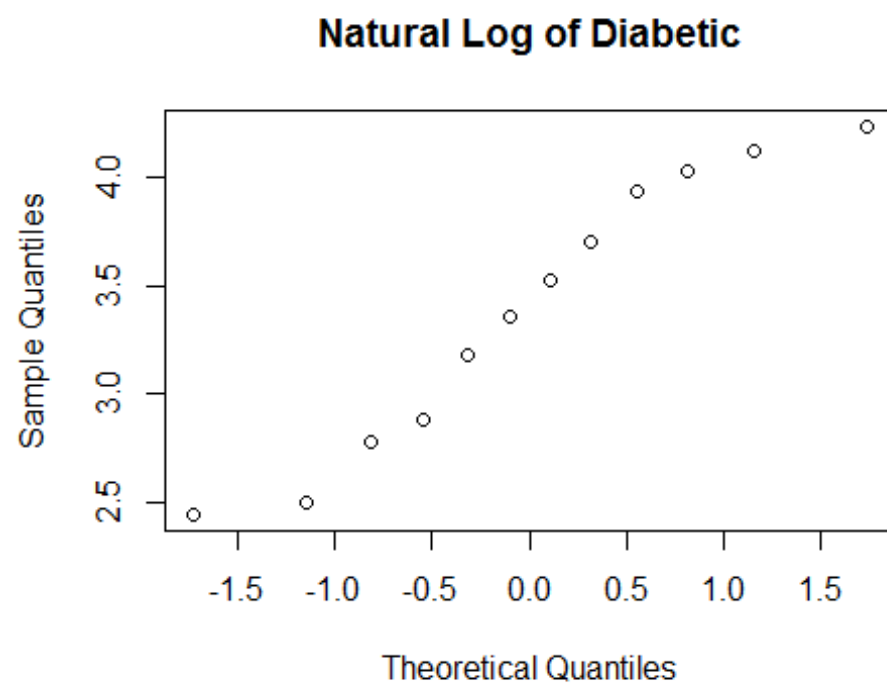

```
# (b) Square Root
sqrtnormal<-sqrt(normal)
sqrtdiabetic<-sqrt(diabetic)
hist(sqrtnormal,prob=TRUE)
lines(density(sqrtnormal),col="red")
```
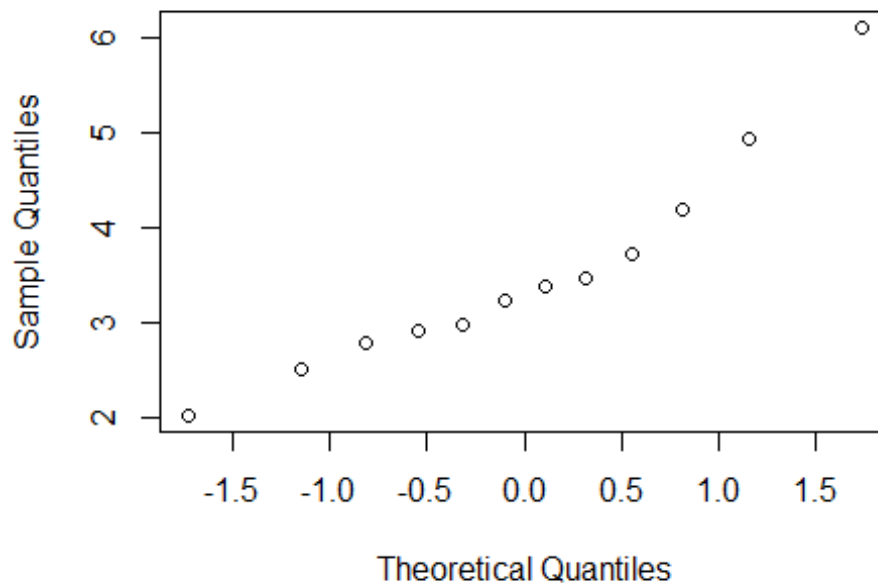
Histogram of sqrtnormal


```
hist(sqrtdiabetic,prob=TRUE)
lines(density(sqrtdiabetic),col="red")
```
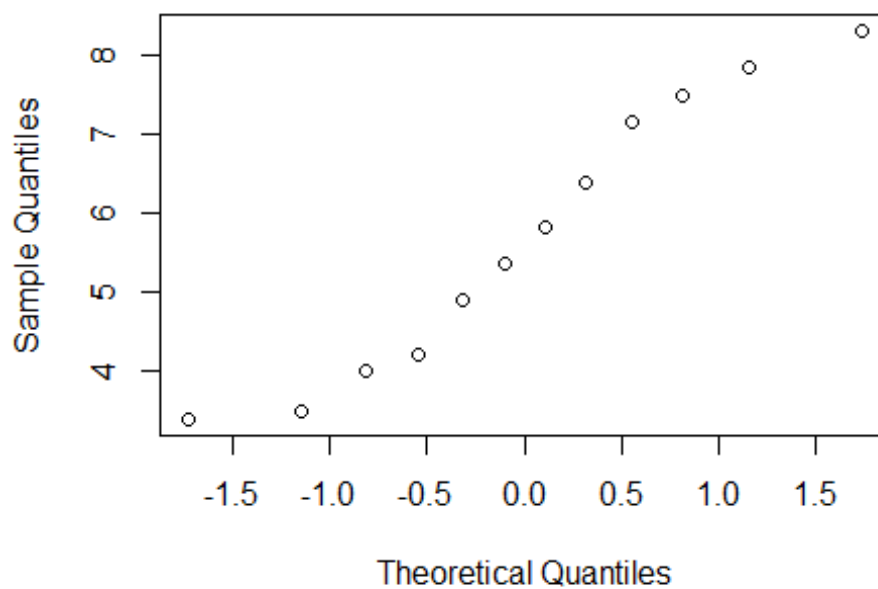
Histogram of sqrtdiabetic


```
qqnorm(sqrtnormal,main="Square root of Normal")
```

Square root of Normal


```
qqnorm(sqrtdiabetic,main ="Square root of Diabetic")
```

Square root of Diabetic


```
# I would prefer log transformation over square root transformation since
# log transformation is more symmetric to normal distribution.
```

```r
# 3)
# As seen from histograms, density plots and qqplots, log transformed
# measurements appear closer to normal distribution.
# 4)
# Welch's t-test
Delta = mean(logdiabetic) - mean(lognormal)
se = sqrt(var(logdiabetic)/12 + var(lognormal)/12)
Tw = Delta/se
nu = (var(logdiabetic)/12+var(lognormal)/12)^2/((var(logdiabetic)/12)^2/11+(v
ar(lognormal)/12)^2/11)
Pvalue = 2*(1-pt(abs(Tw),df=nu))
Pvalue
```

[1] 0.0009776127

```r
# Welch 95% confidence interval
q = qt(0.975, df=nu)
lower = Delta - q*se
upper = Delta + q*se
CI<-c(lower,upper)
CI
```

[1] 0.4352589 1.4792986

```r
# Since P-value is quite low we can reject H0 in favor of Ha.
```

4.

Let μ_e be the mean movie length in ealier years (1956)

Let μ_t be the mean movie length in Todays years (1996)

$\delta = \mu_t - \mu_e$ be the difference in mean movie length time respectively. **Hypothesis Test**
H_0 be the hypothesis that $\delta <= 0$
H_a be the hypothesis that $\delta > 0$

```r
earlier<-c(74,114,114,87,92,55,67,118,79,83,79,92,99,87)
now<-c(70,98,90,95,88,108,110,96,91,88,120,96,90,90)
hist(earlier,prob=TRUE)
lines(density(earlier),col="red")
```
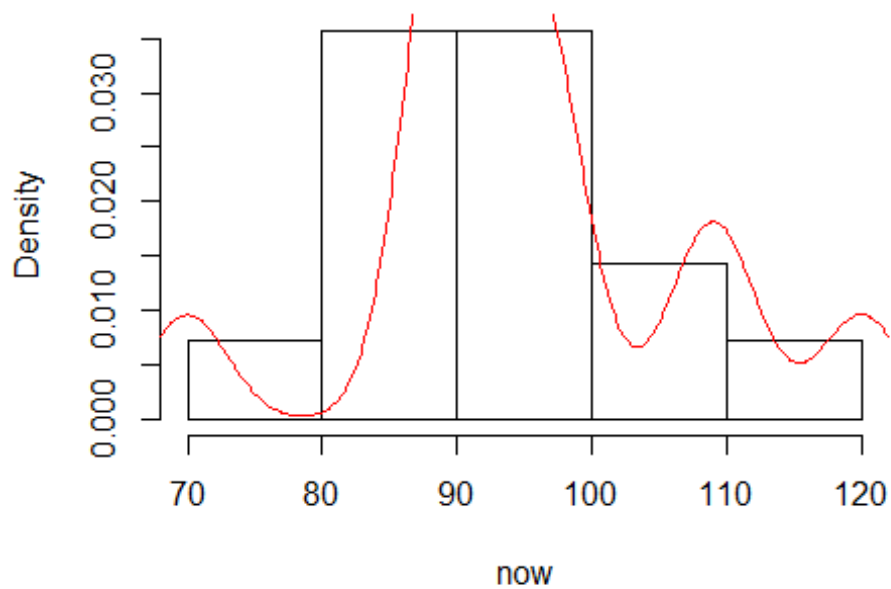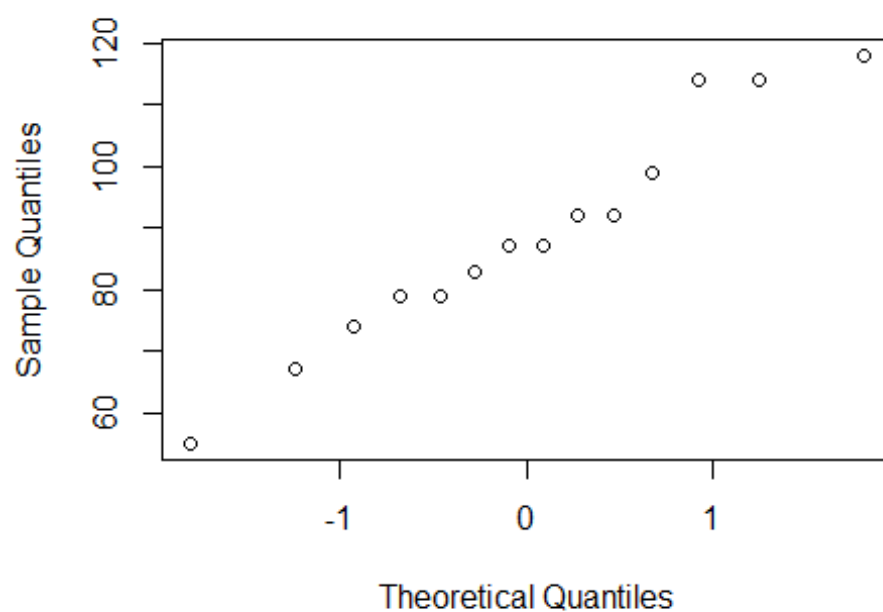
Histogram of earlier
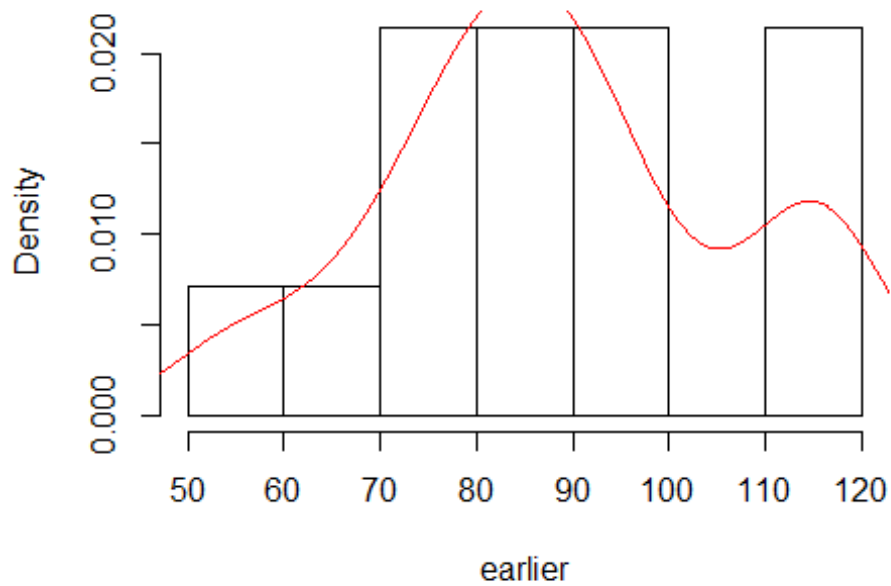

```r
hist(now,prob=TRUE)
lines(density(now),col="red")
```
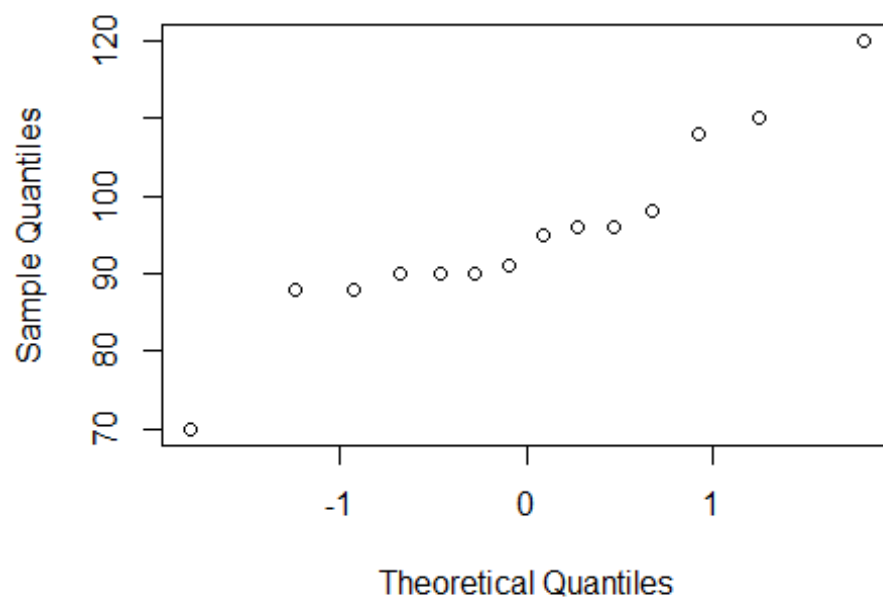
Histogram of now


```r
qqnorm(earlier,main="Feature Films made in 1956")
```

Feature Films made in 1956


```
qqnorm(now,main="Feature Films made in 1996")
```

Feature Films made in 1996


```
t.test(now,earlier)
```

```
##
##   Welch Two Sample t-test
##
## data:  now and earlier
## t = 1.105, df = 22.395, p-value = 0.2809
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   -5.623821 18.480963
## sample estimates:
## mean of x mean of y
##   95.00000  88.57143
```

```
# P-value =0.2809 which shows that mean length of movies in 1956 is more
# than mean length of movies in 1996 with 28% probability assuming null
# hypothesis is true.
# 95% Confidence Interval is (-5.623,18.480) shows that difference in movie
# is not necessarily above 0.
# 0.01 and 0.05 are ideal significance values in hypothesis testing which
# is read as probability of null hypothesis being true.
# After conducting welch's t-test which is based on normality of samples
# datasets are normally distributed as seen from plots to conduct experiment.
```