Introduction of R

Assignment 2

Tirth Parmar

Section B

Roll No. 52

Q1)

```
set.seed(10221)
x = rlnorm(1000,5,2.5)
a = quantile(x , c(0.25,0.5,0.75,1))
b = qlnorm(c(0.25,0.5,0.75,1),5,2.5)
a
b
```

```
> a
        25%         50%         75%        100%
   31.08172   159.36563   850.77350 618183.97963
> b
[1]  27.48872 148.41316 801.29116       Inf
```

The empirical and theoretical values of the quartiles are very close so we can consider it a good approximation.

Q2)

a)

model_0 = lm(mpg ~ 1, data = mtcars)

model_0

```
Call:
lm(formula = mpg ~ 1, data = mtcars)

Coefficients:
(Intercept)
      20.09
```

b)

model_1 = update(model_0, .~. + wt)

model_1

```
Call:
lm(formula = mpg ~ wt, data = mtcars)

Coefficients:
(Intercept)           wt
     37.285       -5.344
```

C)

model_2 = update(model_0, .~. + log(wt) + log(disp))

model_2

```
Call:
lm(formula = mpg ~ log(wt) + log(disp), data = mtcars)

Coefficients:
(Intercept)      log(wt)    log(disp)
     57.128       -8.369       -5.232
```

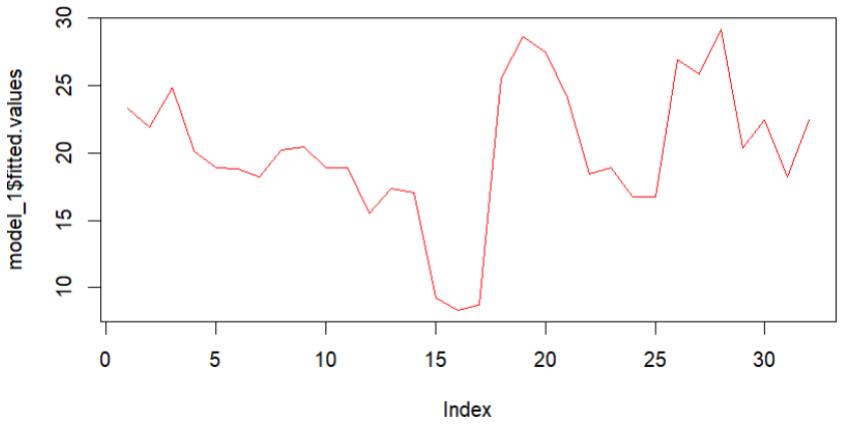
d)

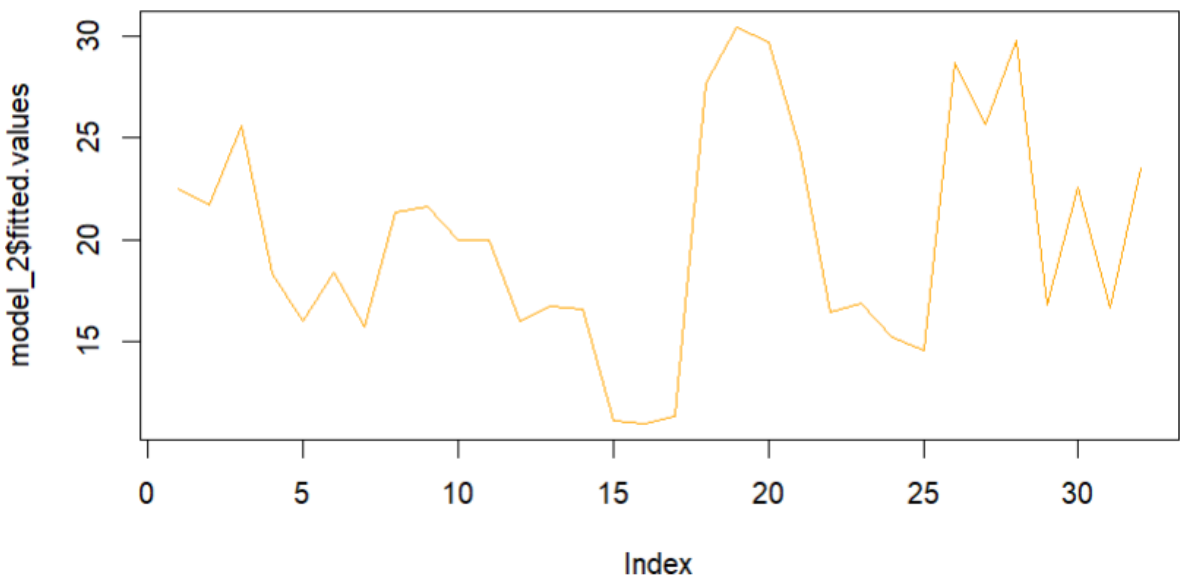
```
> summary(model_0)

Call:
lm(formula = mpg ~ 1, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-9.6906 -4.6656 -0.8906  2.7094 13.8094

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   20.091      1.065   18.86   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.027 on 31 degrees of freedom
```

```
> summary(model_1)

Call:
lm(formula = mpg ~ wt, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-4.5432 -2.3647 -0.1252  1.4096  6.8727

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  37.2851     1.8776  19.858  < 2e-16 ***
wt           -5.3445     0.5591  -9.559 1.29e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.046 on 30 degrees of freedom
Multiple R-squared:  0.7528,    Adjusted R-squared:  0.7446
F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10

> summary(model_2)

Call:
lm(formula = mpg ~ log(wt) + log(disp), data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-3.0257 -1.5407 -0.6875  0.7324  4.7111

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   57.128      5.765   9.910 8.11e-11 ***
log(wt)       -8.369      3.014  -2.777  0.00952 **
log(disp)     -5.232      1.627  -3.216  0.00318 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.331 on 29 degrees of freedom
Multiple R-squared:  0.8601,    Adjusted R-squared:  0.8504
F-statistic: 89.12 on 2 and 29 DF,  p-value: 4.132e-13
```

e) plot(model_1$fitted.values, type = "l", col = "red")

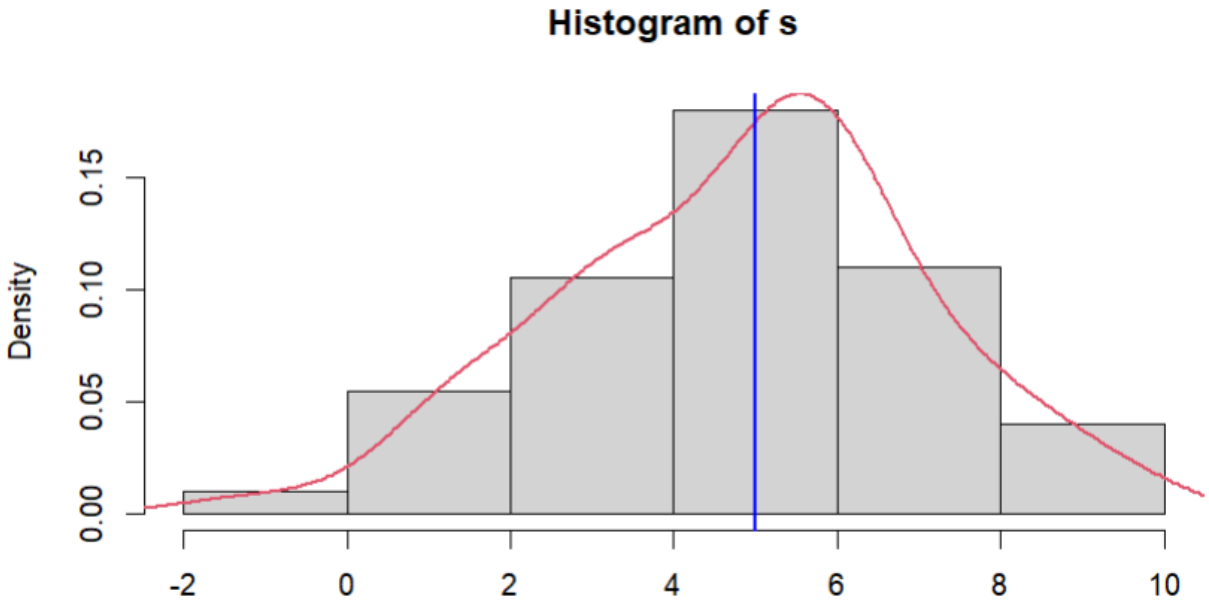
plot(model_2$fitted.values, type = "l", col = "orange")




model_2$fitted.values
15
20
25
30
0
5
10
15
20
25
30
Index

Q3)

```
set.seed(2919)
s = rnorm(100,5,2.5)
x = density(s)
hist(s, freq = F)
lines(x$x , x$y , col = 2, lwd = 2)
abline(v = 5 , col = "blue", lwd = 2)
```



Q4)

```
x = c(0:6)
f = c(739280,185309,23349,1937,114,1,3)
m = sum(f*x)/sum(f)
p = (dpois(0:5,m))
chisq.test(f , correct = T ,p = c(p , 1 - sum(p)))
```

```
> chisq.test(f , correct = T ,p = c(p , 1 - sum(p)))

        Chi-squared test for given probabilities

data:  f
X-squared = 33.475, df = 6, p-value = 8.493e-06
```

Q5)

library(MASS)

x = Animals

a) cor(x)

```
              body        brain
body   1.000000000 -0.005341163
brain -0.005341163  1.000000000
```

b) cor(log(x))

```
> cor(log(x))
          body     brain
body  1.0000000 0.7794935
brain 0.7794935 1.0000000
```

c) cor.test(Animals$body,Animals$brain)

Q6)

obs = matrix(c(40,20,10,30,35,15,30,45,25),3,3,byrow = TRUE)

chisq.test(obs)

```
        Pearson's Chi-squared test

data:  obs
X-squared = 13.415, df = 4, p-value = 0.009416
```

Q7)

```
c = read.csv(file.choose())

c$Spam = as.factor(c$Spam)

model = glm( Spam ~ ., data = c,binomial("logit"),)

model
```

```
Call:  glm(formula = Spam ~ ., family = binomial("logit"), data = c)

Coefficients:
(Intercept)          the           to          and           of            a          in.
   0.332279    -0.164586    -0.024602     0.219322     0.181715     0.162628     0.056937
        for.         you           is           on         this         that            s
  -0.101996     0.173672    -0.074714    -0.393930     0.213083    -0.135693    -0.003784
      enron           be         with            i         your         have           as
 -21.098030    -0.182656    -0.020856    -0.170753     0.522094    -0.293822     0.068130
       will           it           we          are
  -0.332093     0.169437    -0.144296     0.067673

Degrees of Freedom: 4999 Total (i.e. Null);  4975 Residual
Null Deviance:      6931
Residual Deviance: 4065         AIC: 4115
```

```
Call:
glm(formula = Spam ~ ., family = binomial("logit"), data = c)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.3006  -0.6469   0.0000   0.7918   3.9864

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.332279   0.050606   6.566 5.17e-11 ***
the          -0.164586   0.018233  -9.027  < 2e-16 ***
to           -0.024602   0.019595  -1.255 0.209300
and           0.219322   0.024289   9.030  < 2e-16 ***
of            0.181715   0.027009   6.728 1.72e-11 ***
a             0.162628   0.027082   6.005 1.91e-09 ***
in.           0.056937   0.030434   1.871 0.061369 .
for.         -0.101996   0.028459  -3.584 0.000338 ***
you           0.173672   0.023386   7.426 1.12e-13 ***
is           -0.074714   0.033662  -2.220 0.026449 *
on           -0.393930   0.037303 -10.560  < 2e-16 ***
this          0.213083   0.035686   5.971 2.36e-09 ***
that         -0.135693   0.040352  -3.363 0.000772 ***
s            -0.003784   0.041093  -0.092 0.926640
enron       -21.098030 243.275658  -0.087 0.930890
be           -0.182656   0.046072  -3.965 7.35e-05 ***
with         -0.020856   0.048726  -0.428 0.668628
i            -0.170753   0.025139  -6.792 1.10e-11 ***
your          0.522094   0.038343  13.616  < 2e-16 ***
have         -0.293822   0.048427  -6.067 1.30e-09 ***
as            0.068130   0.044897   1.517 0.129148
will         -0.332093   0.044793  -7.414 1.23e-13 ***
it            0.169437   0.047200   3.590 0.000331 ***
we           -0.144296   0.037909  -3.806 0.000141 ***
are           0.067673   0.046464   1.456 0.145262
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6931.4  on 4999  degrees of freedom
Residual deviance: 4065.3  on 4975  degrees of freedom
AIC: 4115.3

Number of Fisher Scoring iterations: 21
```