Homework 3

Megha Pandit

March 8, 2019

Problem 1

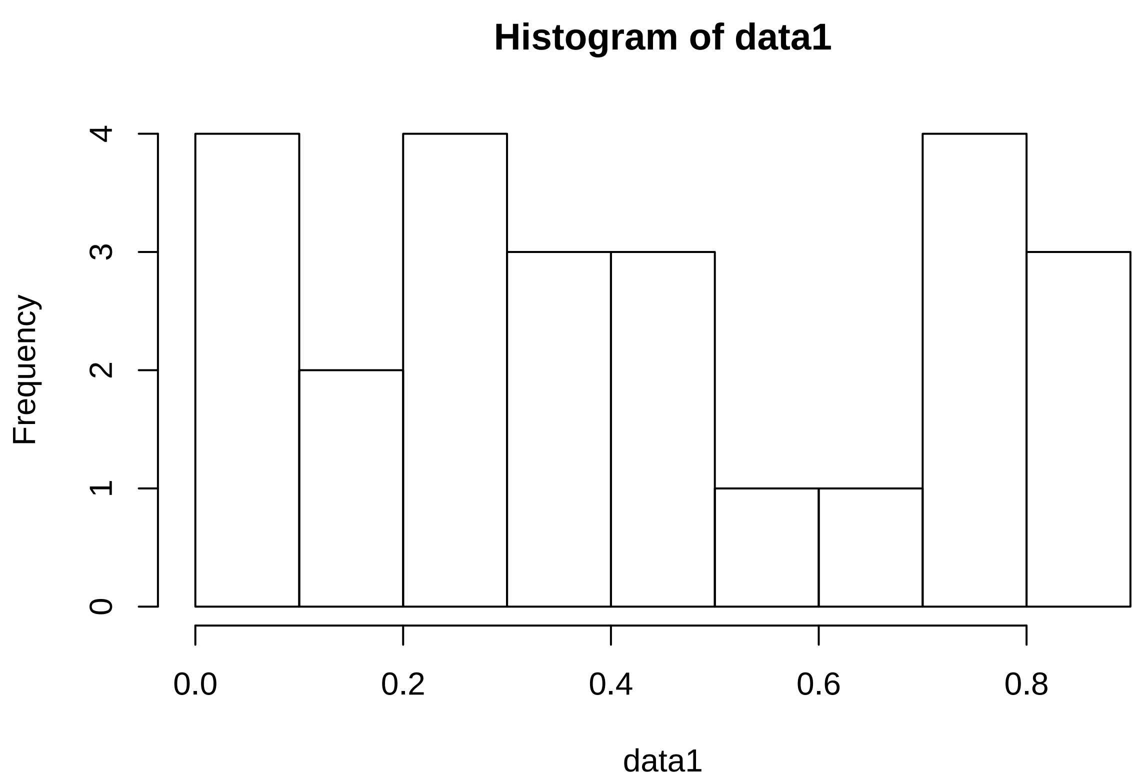
```
x1 <- read.table("maybe_uniform.txt")
```

```
## Warning in read.table("maybe_uniform.txt"): incomplete final line found by
## readTableHeader on 'maybe_uniform.txt'
```

```
data1 <- c(x1[1,], x1[2,], x1[3,], x1[4,], x1[5,])
data1 <- as.numeric(data1)
ks.test(data1, "punif")
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  data1
## D = 0.18, p-value = 0.3501
## alternative hypothesis: two-sided
```

```
hist(data1, breaks = 10)
```



Given the p-value of 0.35 from the KS test, we cannot reject the null hypothesis that the data comes from a uniform distribution.

```
x2 <- seq(-20, 20, by = 0.01)
fx2 <- ifelse(0 < x2 & x2 <= 0.5, 3/2, ifelse(0.5 < x2 & x2 < 1, 1/2, 0))
ks.test(fx2, data1)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  fx2 and data1
## D = 0.97526, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

The resulting p-value suggests that we reject the null hypothesis that the above two samples come from the same distribution. The D statistic is much larger for this model than for the uniform distribution, implying that the above model is not better than uniform distribution.

```
ks.test(data1, "pgamma", shape = 1, rate = 2)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  data1
## D = 0.1653, p-value = 0.4535
## alternative hypothesis: two-sided
```

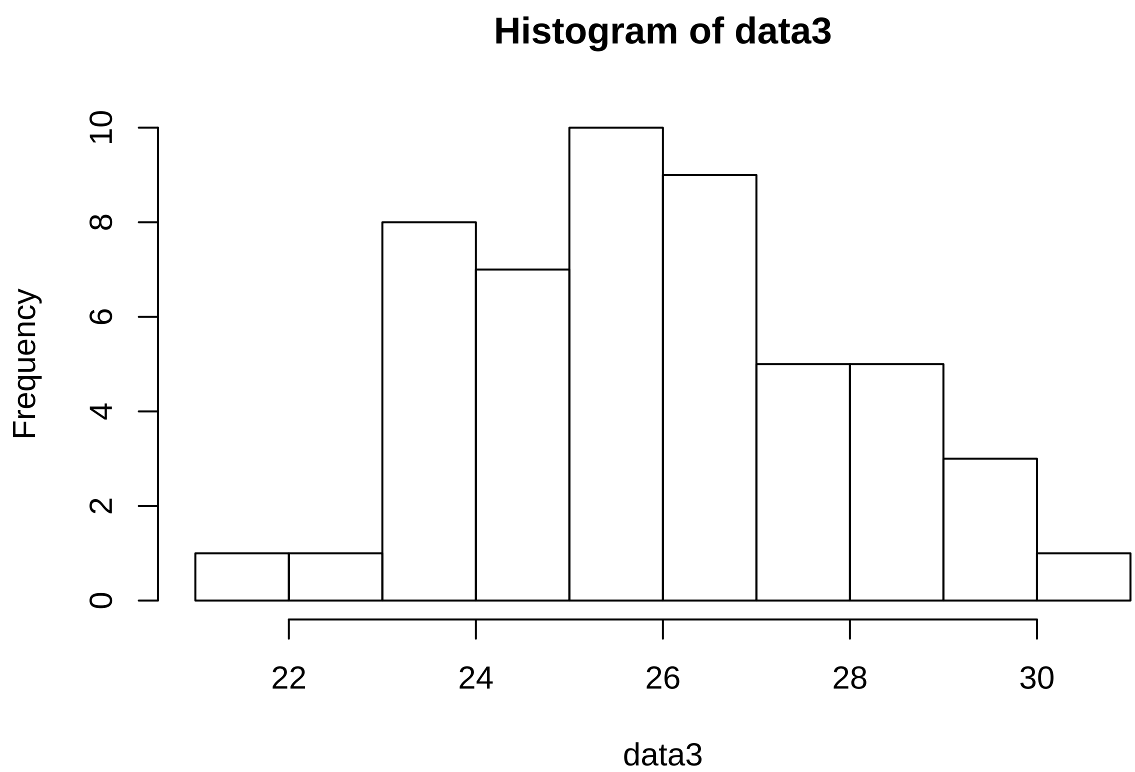
The D statistic is smaller for the above test, and the p-value also suggests that we cannot reject the null hypothesis that the data is drawn from a gamma distribution. This may be a better model than the previous two.

Problem 2

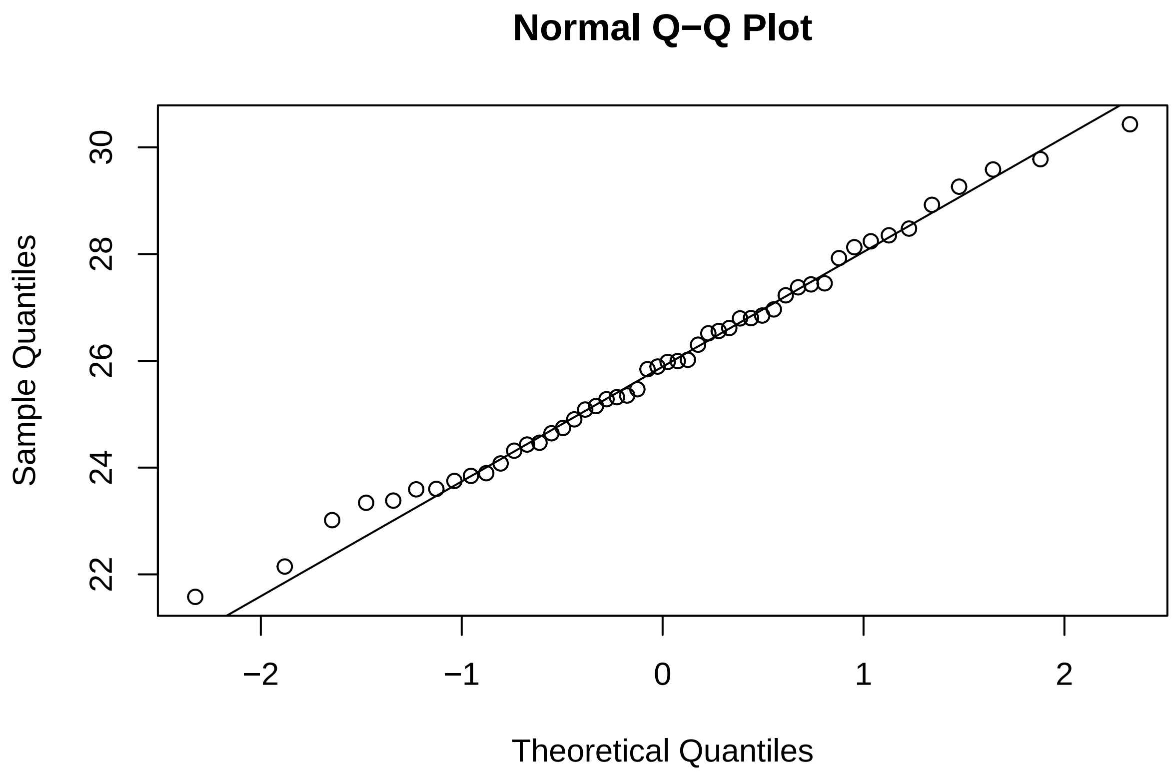
```
x3 <- read.table("maybe_normal.txt")
data3 <- c(x3[1,], x3[2,], x3[3,], x3[4,], x3[5,], x3[6,], x3[7,], x3[8,], x3[9,], x3[10,])
data3 <- as.numeric(data3)

ks.test(data3, "pnorm", mean = 26, sd = 2)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  data3
## D = 0.06722, p-value = 0.9663
## alternative hypothesis: two-sided
```

```r
hist(data3)
```




```r
qqnorm(data3)
qqline(data3)
```



From the above results, and with a p-value of 0.9663, we cannot reject the null hypothesis that the data comes from normal distribution.

Problem 3

```r
txt1 <- read.table("maybe_same_1.txt")
X <- c(txt1[1,], txt1[2,], txt1[3,], txt1[4,], txt1[5,])
X <- as.numeric(X)

txt2 <- read.table("maybe_same_2.txt")
Y <- c(txt2[1,], txt2[2,], txt2[3,], txt2[4,], txt2[5,])
Y <- as.numeric(Y)

ks.test(X, Y)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  X and Y
## D = 0.25, p-value = 0.491
## alternative hypothesis: two-sided
```

```r
ks.test(X+2, Y)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  X + 2 and Y
## D = 0.65, p-value = 0.0001673
## alternative hypothesis: two-sided
```

From the two-sample KS test for X and Y, we cannot reject the null hypothesis that X and Y come from the same distribution. However, from the two-sample test for X+2 and Y, the p-value is 0.0001673 suggesting that we could reject the null hypothesis that X+2 and Y are from the same distribution.

Problem 4

```r
norm1 <- readRDS("norm_sample.RData")
norm1 <- data.frame(norm1)
ns <- ecdf(norm1$norm1)

norm1$ecdf <- ns(norm1$norm1)
norm1$nm <- pnorm(norm1$norm1)
norm1$D <- norm1$ecdf - norm1$nm

print(paste0("The D statistic is: ", max(abs(norm1$D))))
```

```
## [1] "The D statistic is: 0.13724272684825"
```

```r
ks.test(norm1$norm1, "pnorm", mean = 0, sd = 1)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  norm1$norm1
## D = 0.17724, p-value = 0.3683
## alternative hypothesis: two-sided
```

The calculated D statistic is 0.137 and the D statistic given by the KS test is 0.177. From the results of the KS test, we fail to reject the null hypothesis that the data comes fom a standard normal distribution.

Problem 5

```r
fiji <- read.table("fijiquakes.dat", header = T)

fiji_ecdf <- ecdf(fiji$mag)

#Finding the 95% CI
```

```r
ci_data <- sum(fiji$mag <= 4.9 & fiji$mag > 4.3)

print(paste0("The 95% confidence interval for F(4.9) - F(4.3): "))
```

```
## [1] "The 95% confidence interval for F(4.9) - F(4.3): "
```

```r
binconf(ci_data, length(fiji$mag), method = "wilson", 0.05)
```

```
##  PointEst     Lower     Upper
##     0.526 0.4950118 0.5567892
```

```r
faith <- read.table("faithful.dat", header = T, skip = 25)

faith_ecdf <- ecdf(faith$waiting)

#90% confidence interval for mean waiting time
mean_faith <- mean(faith$waiting)
sd_faith <- sd(faith$waiting)/sqrt(length(faith$waiting))

print(paste0("The 90% confidence interval for the mean waiting time is: [", mean_faith - 1.64*sd_faith,
             ", ", mean_faith + 1.64*sd_faith, "]"))
```

```
## [1] "The 90% confidence interval for the mean waiting time is: [69.5451799826703, 72.2489376643885]"
```

```r
summary(faith_ecdf)[3]
```

```
## Median
##     70
```

```r
print(paste0("The estimated median waiting time is: ", summary(faith_ecdf)[3], " ", "min"))
```

```
## [1] "The estimated median waiting time is: 70 min"
```