MA677_HW3

Sky Liu

3/8/2019

```r
mu = read.table("maybe_uniform.txt")
```

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

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

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  mu
## D = 0.18, p-value = 0.3501
## alternative hypothesis: two-sided
```
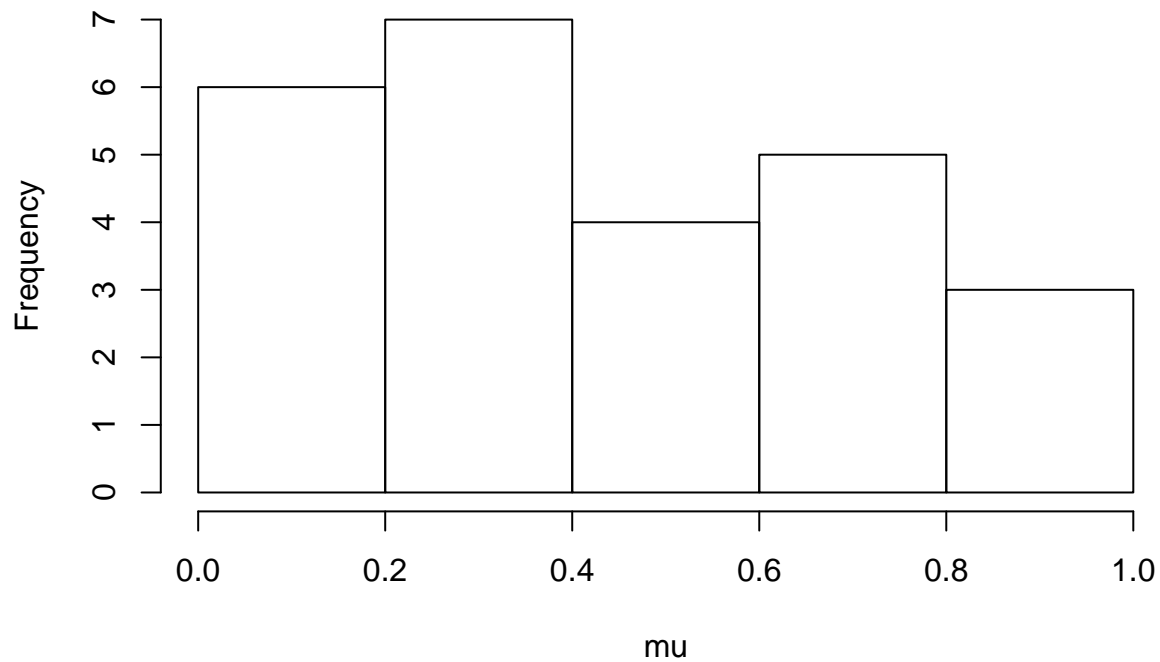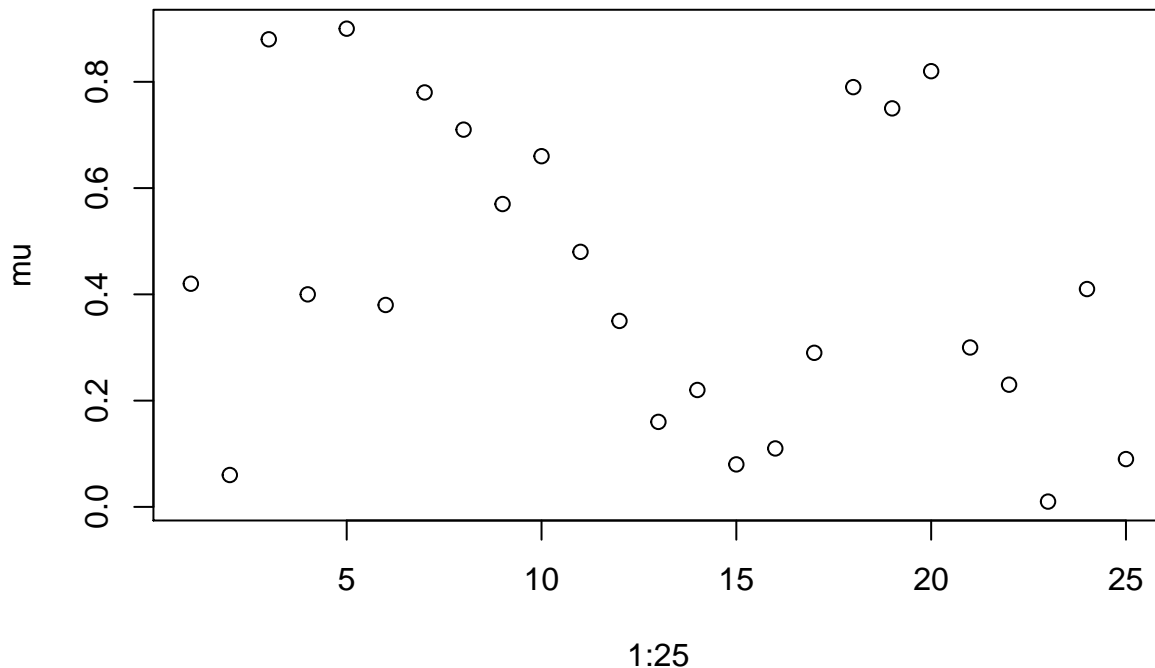
```r
hist(mu)
```

Histogram of mu


```r
plot(1:25,mu)
```

Given the p value of KS test being 0.35, we fail to reject the null hypothsis that the data from maybe_uniform txt is uniformlly distributed.

```
x<-seq(-10, 10, by=0.001)
fx <- ifelse(x > 0 & x <=0.5, 2/3,
    ifelse(x > 0.5 & x < 1,  1/2, 0))
ks.test(fx,mu)
```

```
## Warning in ks.test(fx, mu): p-value will be approximate in the presence of
## ties
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  fx and mu
## D = 0.95005, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

From the KS test, we could see that the data does not have matched distribution with the piecewise model, with very small p value and much higher D value than the previous one. Thus, the model is not better than uniform distribution.

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

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

The KS test of gamma distribution with rate = 2 and shape = 1 gives even smaller D value than uniform distribution, showing the gamma distribution is a bit better.

Problem 2

```r
mn = read.table("maybe_normal.txt")
mn<-c(mn[1,],mn[2,],mn[3,],mn[4,],mn[5,])
mn<-as.numeric(mn)

n25 <- rnorm(25, 26, 4)
ks.test(mn, n25)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  mn and n25
## D = 0.28, p-value = 0.285
## alternative hypothesis: two-sided
```
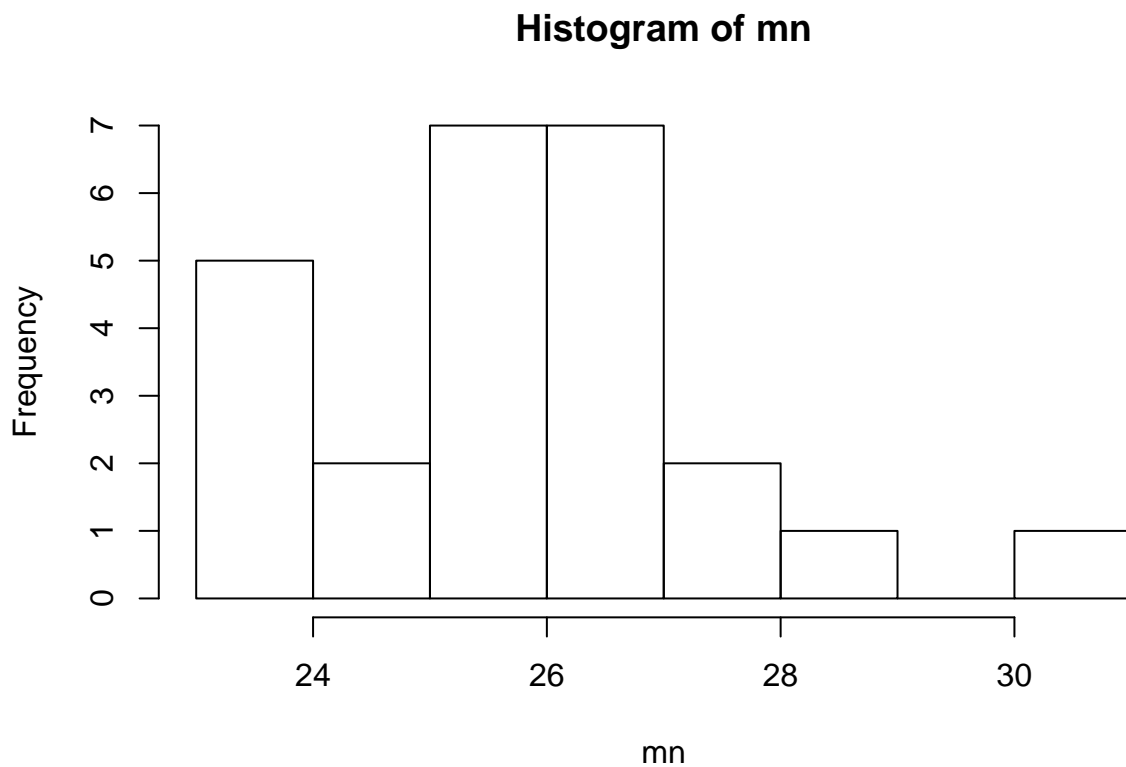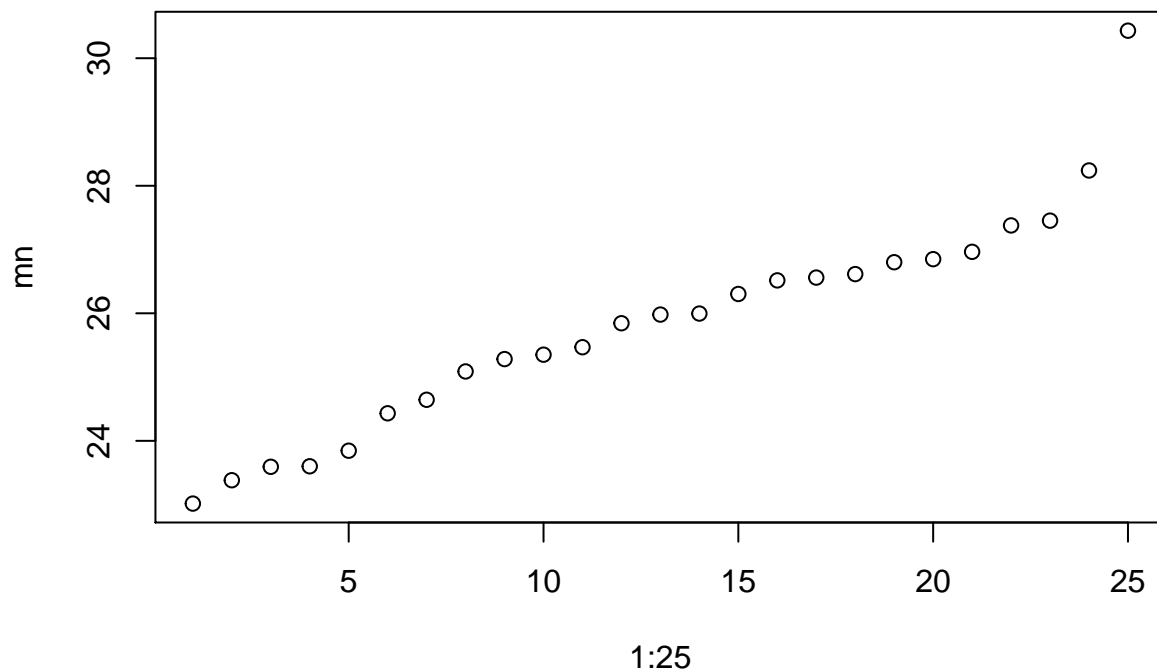
```r
mean(mn)
```

```
## [1] 25.82532
```

```r
var(mn)
```

```
## [1] 2.898906
```

```r
hist(mn)
```

Histogram of mn

```r
qqplot(1:25,mn)
```

Given the p value of KS test being 0.285, we fail to reject the null hypothsis that the data from maybe_normal txt is normally distributed.

Problem 3

```
X = read.table("maybe_same_1.txt")

## Warning in read.table("maybe_same_1.txt"): incomplete final line found by
## readTableHeader on 'maybe_same_1.txt'
X<-c(X[1,],X[2,],X[3,],X[4,],X[5,])
X<-as.numeric(X)

Y = read.table("maybe_same_2.txt")

## Warning in read.table("maybe_same_2.txt"): incomplete final line found by
## readTableHeader on 'maybe_same_2.txt'
Y<-c(Y[1,],Y[2,],Y[3,],Y[4,],Y[5,])
Y<-as.numeric(Y)
X2<-X+2

ks.test(X, Y)

## Warning in ks.test(X, Y): cannot compute exact p-value with ties

##
##  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)
```

```
## Warning in ks.test(X + 2, Y): cannot compute exact p-value with ties
```

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

Given the p value of KS test between X and Y being 0.491, we fail to reject the null hypothsis that X and Y are from the same distribution.

The p value of KS test between X+2 and Y is 0.0001673, it is likely we could reject the null hypothesis that X+2 and Y are from the same distribution.

Problem 4

```r
NS <- readRDS("norm_sample.Rdata")
NS <- as.data.frame(NS)
colnames(NS) <- "x"
NS_d <- ecdf(NS$x)
NS$ecdf <- NS_d(NS$x)
NS$nm <- pnorm(NS$x)
NS$D <- NS$ecdf - NS$nm
max(abs(NS$D))
```

```
## [1] 0.1372427
```

```r
ks.test(NS$x, "pnorm", 0,  1)
```

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

The calculated D value is 0.137, the KS test D value is 0.177. From the test, we fail to reject the null hypothesis that the normal sample data is normally distributed.

Problem 5

fujiquakes

```r
FJ <- read.table("fijiquakes.dat", header = T)
FJ_d <- ecdf(FJ$mag)
# 95% CI for F(4.9) - F(4.3)
CI<-sum((FJ$mag<=4.9) & (FJ$mag>4.3))
binconf(CI,length(FJ$mag),method="wilson",0.05)
```

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

faithful

```r
FF <- read.table("faithful.dat", header = T, skip = 25)
f_d <- ecdf(FF$waiting)
# 90% confidence interval for mean
m <- mean(FF$waiting)
se <- sd(FF$waiting)/sqrt(length(FF$waiting))
print(paste0("90% CI for mean waiting time is [",
             round(m-1.64*se,2),", ", round(m+1.64*se,2),"]"))
```

```
## [1] "90% CI for mean waiting time is [69.55, 72.25]"
```

```r
# median waiting time
summary(f_d)
```

```
## Empirical CDF:     51 unique values with summary
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   43.00   56.50   70.00   69.67   82.50   96.00
```