# 677hw

*Kaiyu Yan*

*March 8, 2019*

## R Markdown

```r
library(ACSWR)
```

```
## Warning: package 'ACSWR' was built under R version 3.5.2
```

```r
library(readr)
library(sfsmisc)
```

```
## Warning: package 'sfsmisc' was built under R version 3.5.2
```

```r
A <- read.delim("maybe_uniform.txt",header = F, sep= "",dec = " ")
```
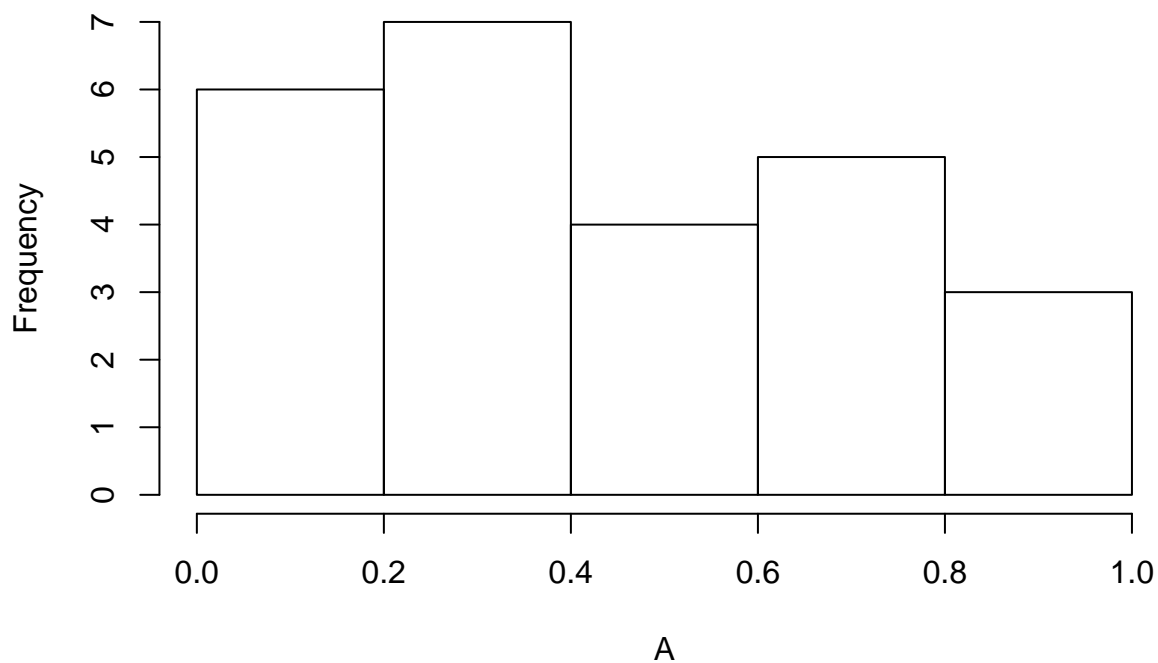
```
## Warning in read.table(file = file, header = header, sep = sep,
## quote = quote, : incomplete final line found by readTableHeader on
## 'maybe_uniform.txt'
```
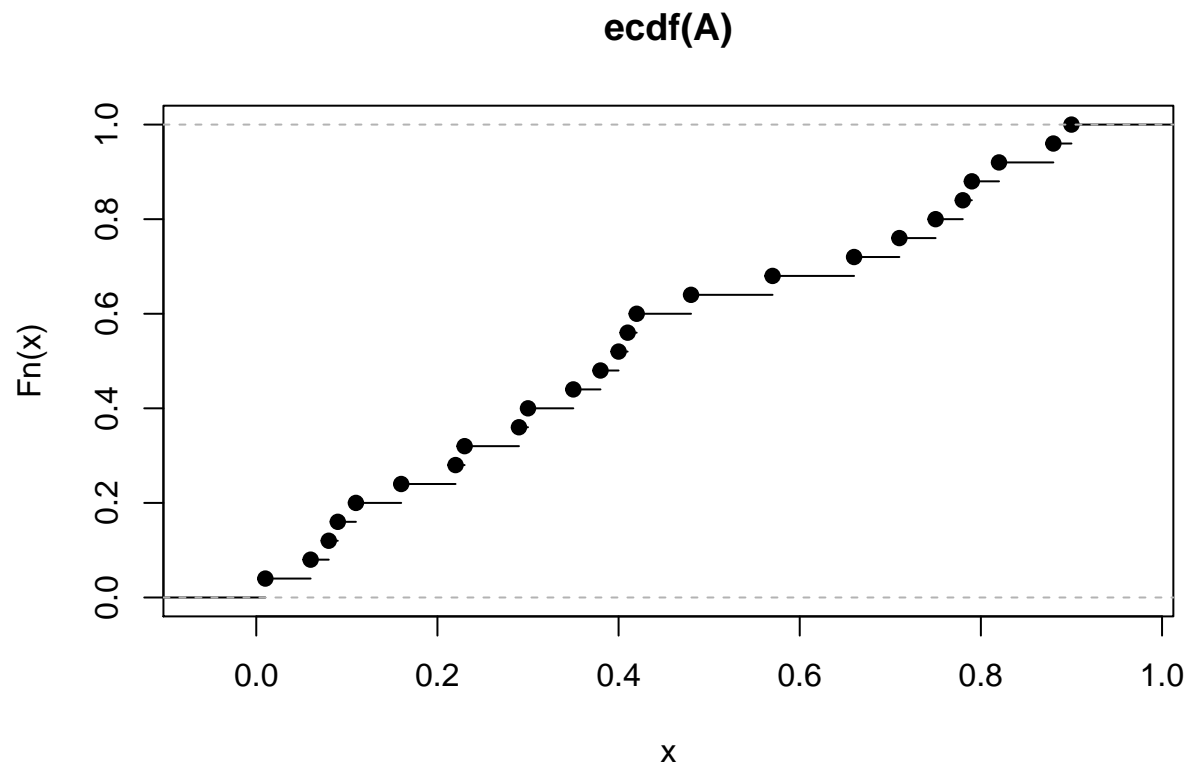
```r
A <-  cbind(A[1,], A[2,],A[3,],A[4,],A[5,])
A <-  as.numeric(as.vector(t(A)[,1]))
hist(A)
```
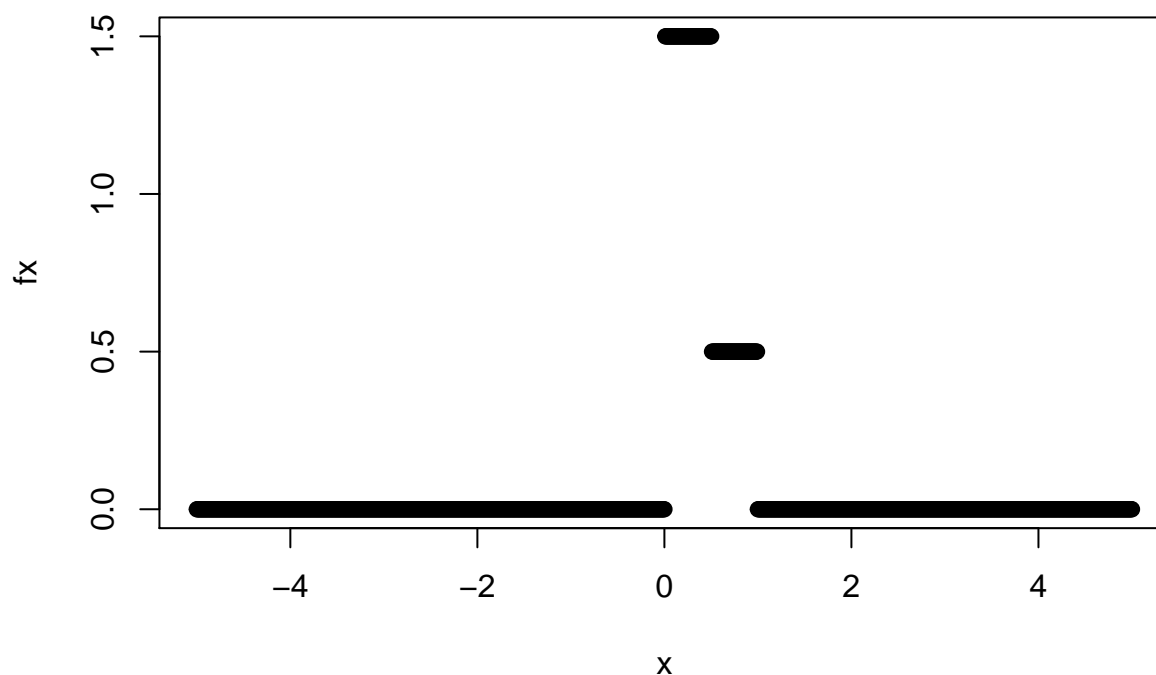
**Histogram of A**

```r
dist_A <- ecdf(A)
plot(dist_A)
```

**ecdf(A)**



```r
ks.test(A,"punif")
```

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

```r
x<-seq(-5, 5, by=0.01)
fx <- ifelse(x > 0 & x <=0.5, 3/2,
   ifelse(x > 0.5 & x < 1,  0.5, 0))
plot(x,fx)
```

```r
ks.test(fx,"punif")
```

```
## Warning in ks.test(fx, "punif"): ties should not be present for the
## Kolmogorov-Smirnov test
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  fx
## D = 0.9011, p-value < 2.2e-16
## alternative hypothesis: two-sided
```
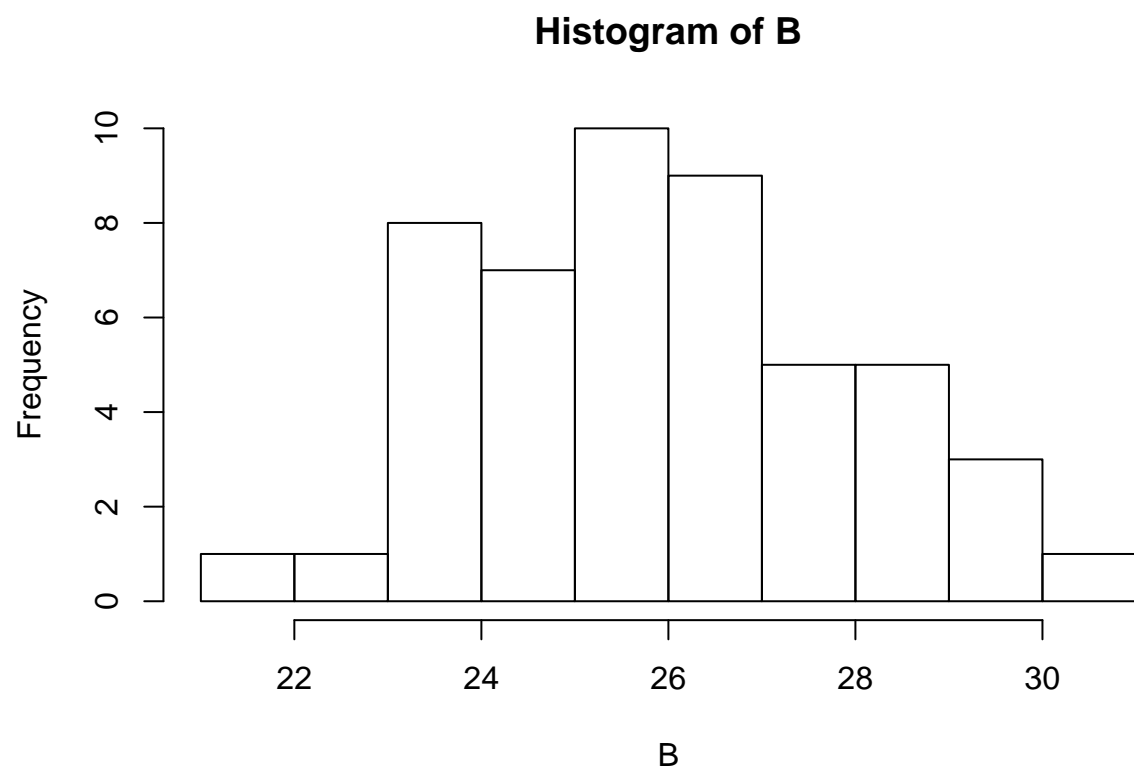
I think the distribution of maybe_uniform belongs to uniform distribution on [0,1]. Also, there should be better fit model.

```r
B <- read.table("maybe_normal.txt")
B <-  cbind(B[1,], B[2,],B[3,],B[4,],B[5,],B[6,], B[7,],B[8,],B[9,],B[10,])
B <-  as.numeric(as.vector(t(B)[,1]))
qqnorm(B)
```
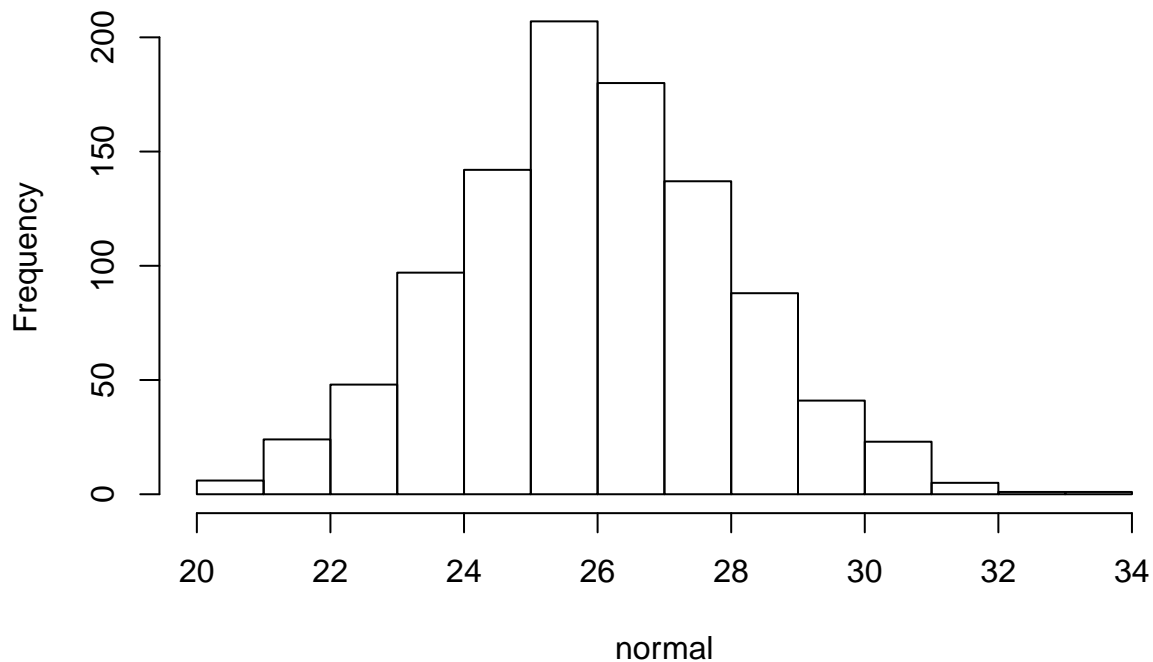
## Normal Q–Q Plot



```
hist(B)
```
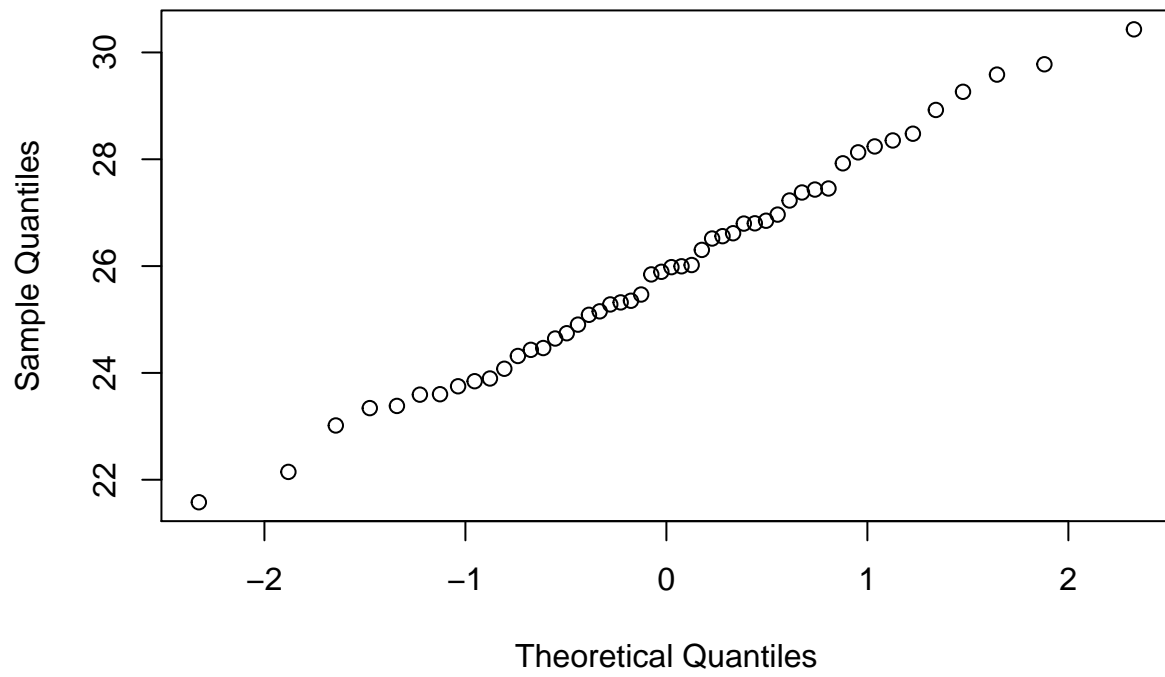
## Histogram of B



```
normal <- rnorm(1000,26,2)
hist(normal)
```

## Histogram of normal



```r
ks.test(B,"pnorm")
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  B
## D = 1, p-value = 8.882e-16
## alternative hypothesis: two-sided
```

**According to the ks-test result, the maybe_normal belongs to normal distribution.**

```r
C1=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'
```

```r
x=c(C1$V1,C1$V2,C1$V3,C1$V4,C1$V5)
C2=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'
```

```r
y=c(C2$V1,C2$V2,C2$V3,C2$V4,C2$V5)
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
```

For the first question. Since p-value of of ks.test is much larger than 0.05, we cannot to reject the null hypotheses, C1,C2 are from the same distribution. Also, for the second qeustion. Since p-value of ks-test is much smaller than 0.05.Then we reject the null hypotheses, therefore C1+2 and C2 are not from the same distribution
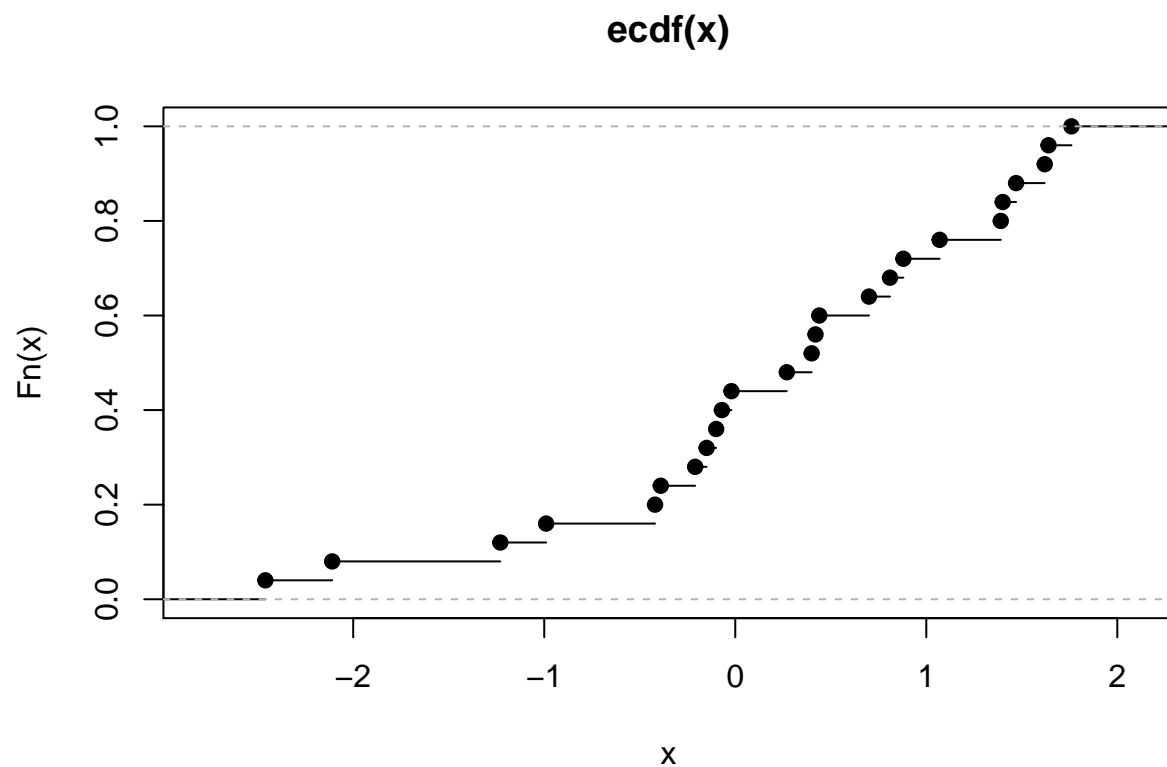
```r
D <- readRDS("norm_sample.Rdata")
ks.test(D,"pnorm")
```

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

```r
D_ecdf <- ecdf(D)
summary(D_ecdf)
```

```
## Empirical CDF:      25 unique values with summary
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.4600 -0.2100  0.4000  0.2448  1.0700  1.7600
```

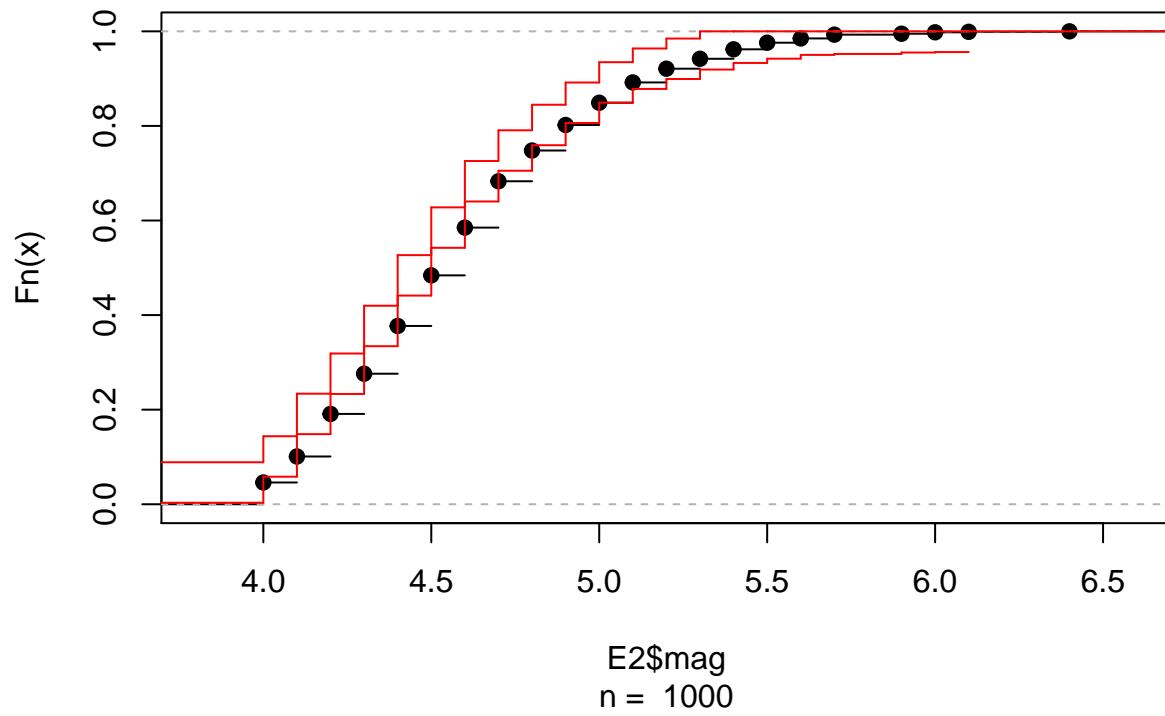```r
plot.ecdf(D)
```

## ecdf(x)



```r
D_norm <- rnorm(1000)
ks.test(D,D_norm)

##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  D and D_norm
## D = 0.194, p-value = 0.3177
## alternative hypothesis: two-sided

E1 <- read.table("faithful(1).dat",header = T,skip = 25)
E2 <-  read.table("fijiquakes(1).dat",header = T)

ecdf.ksCI(E2$mag)
```
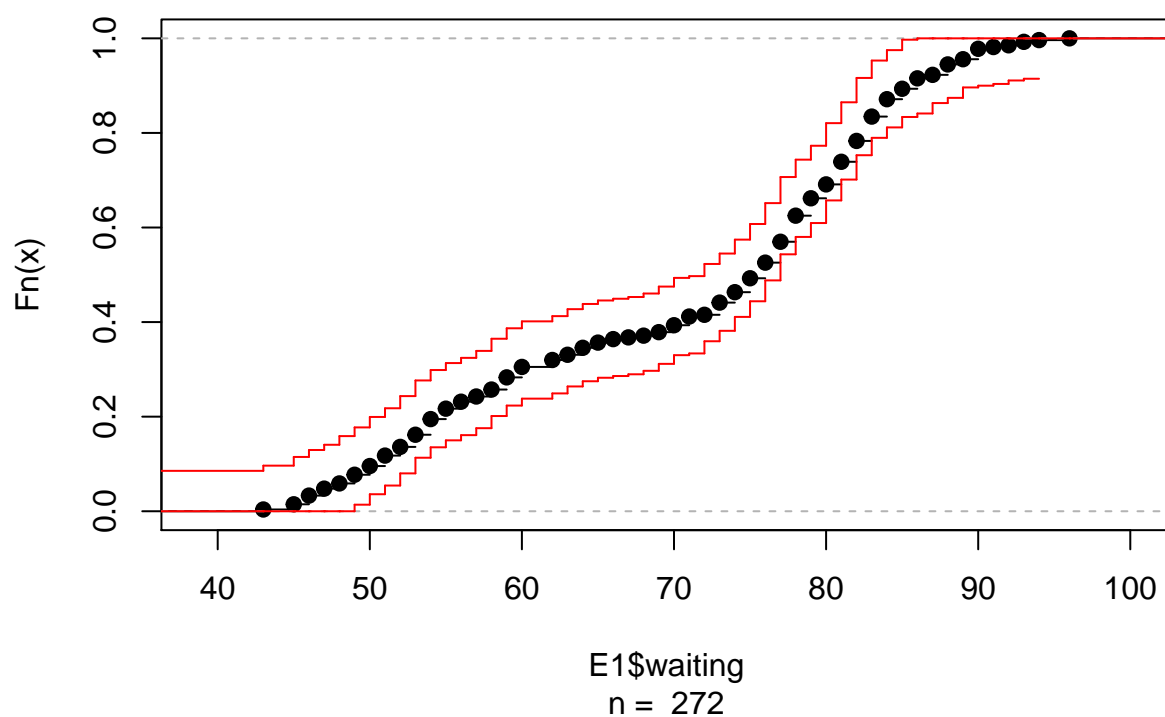
## ecdf(E2$mag) + 95% K.S. bands



E2$mag
n = 1000

```
ecdf.ksCI(E1$waiting)
```

## ecdf(E1$waiting) + 95% K.S. bands



E1$waiting
n = 272

```r
mag <- ecdf(E2$mag)
mag(4.9) - mag(4.3)
```

```
## [1] 0.526
```

```r
s = sd(E2$mag)
n = length(E2$mag)
error <- qt(0.95,df=n-1)*s/sqrt(n)

up <- mean(E1$waiting)+error
low <- mean(E1$waiting) -error

sprintf("The CI is %e - %e",low,up)
```

```
## [1] "The CI is 7.087609e+01 - 7.091803e+01"
```