# Homework Empirical distn and KS test

*Xiang XU*

*March 7, 2019*

1.

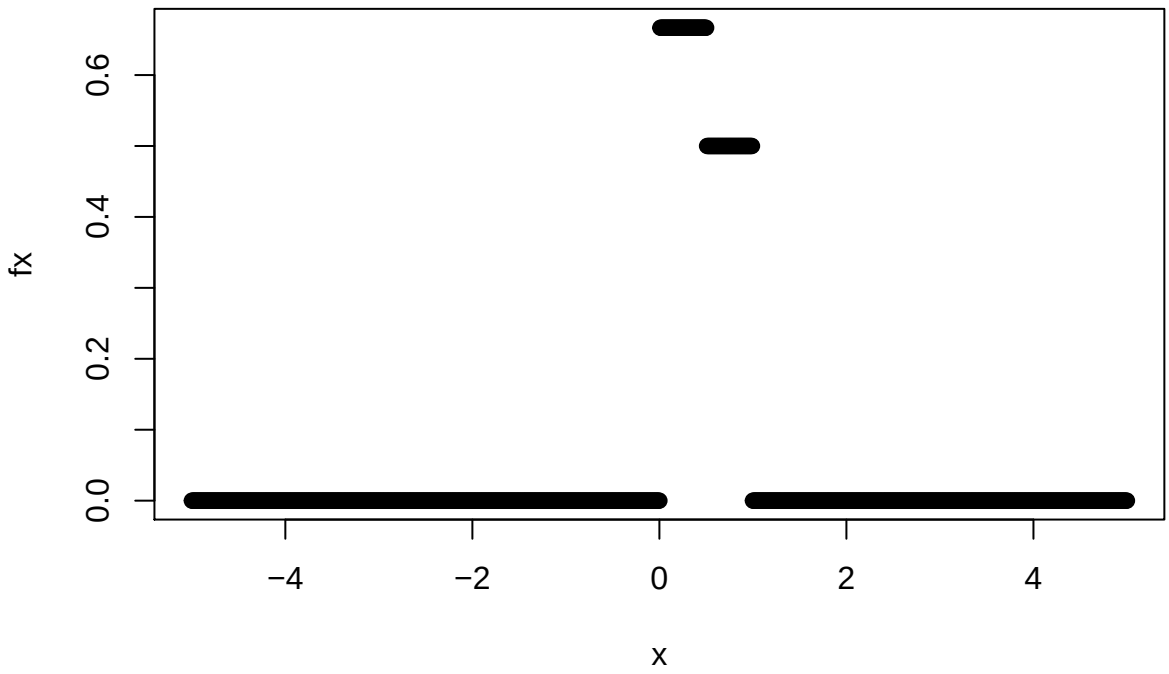
```
Df1=read.table("maybe_uniform.txt")
```

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

```
test1=c(Df1$V1,Df1$V2,Df1$V3,Df1$V4,Df1$V5)
ks.test(test1,"punif")
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  test1
## D = 0.18, p-value = 0.3501
## alternative hypothesis: two-sided
```

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

According to result of ks-test(P-value bigger than 0.05, fail to reject), therefore, these data is distributed as a Uniform distribution on [0,1]. Since in ks-test, the D value of fx is bigger than that of data, the model below is not better than the Uniform???

2.

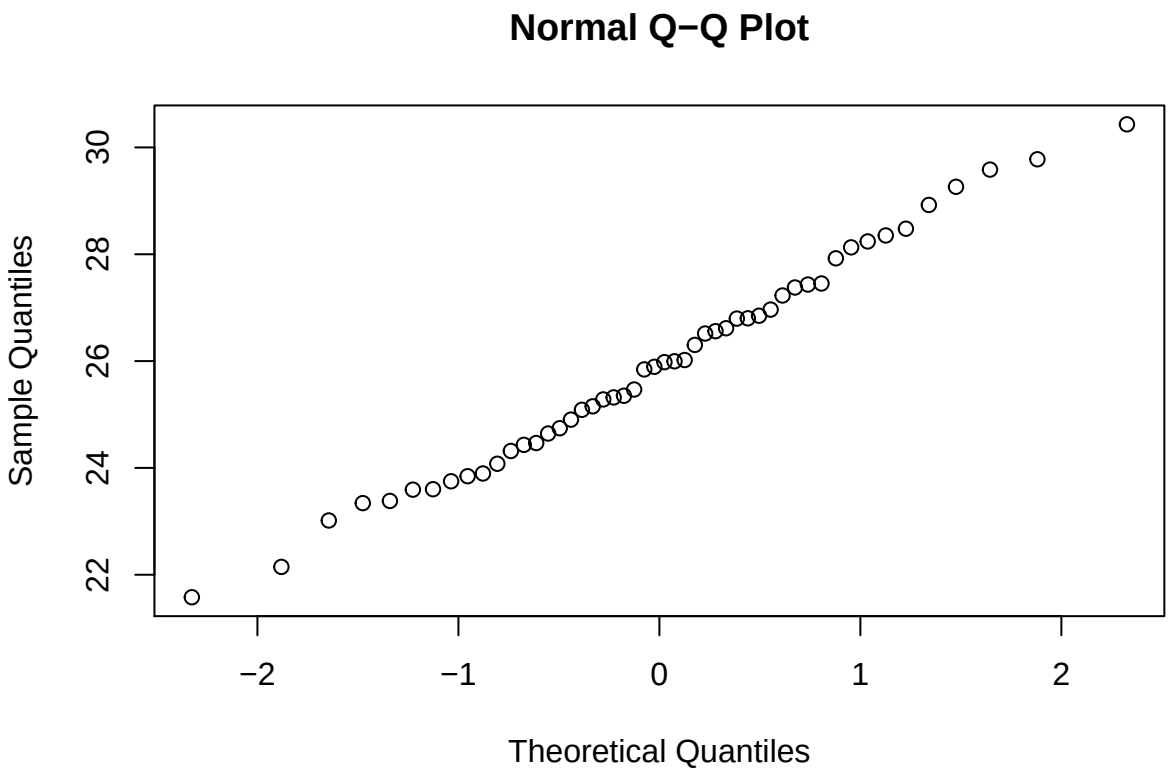
```r
Df2=read.table("maybe_normal.txt")
d=c(Df2$V1,Df2$V2,Df2$V3,Df2$V4,Df2$V5)
test2=(c(Df2$V1,Df2$V2,Df2$V3,Df2$V4,Df2$V5)-rep(26,4))/2
```

```
## Warning in c(Df2$V1, Df2$V2, Df2$V3, Df2$V4, Df2$V5) - rep(26, 4): longer
## object length is not a multiple of shorter object length
```

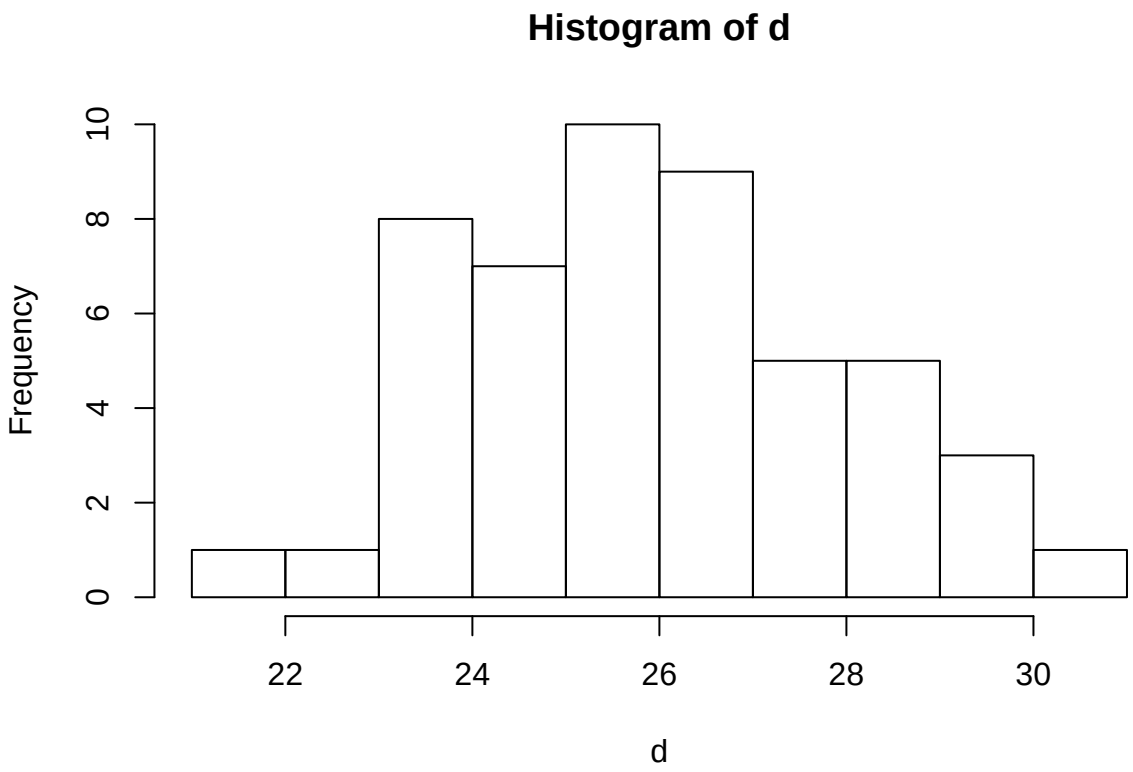
```r
ks.test(test2,"pnorm")
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  test2
## D = 0.06722, p-value = 0.9663
## alternative hypothesis: two-sided
```

```r
qqnorm(d)
```


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
22
24
26
28
30
−2
−1
0
1
2

```
hist(d)
```




According to Ks-test(p smaller than 0.05 fail to reject) and qqnorm, the data generally follow normal disatribution.

3.

```
Df31=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(Df31$V1,Df31$V2,Df31$V3,Df31$V4,Df31$V5)
Df32=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(Df32$V1,Df32$V2,Df32$V3,Df32$V4,Df32$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
```

```
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
```

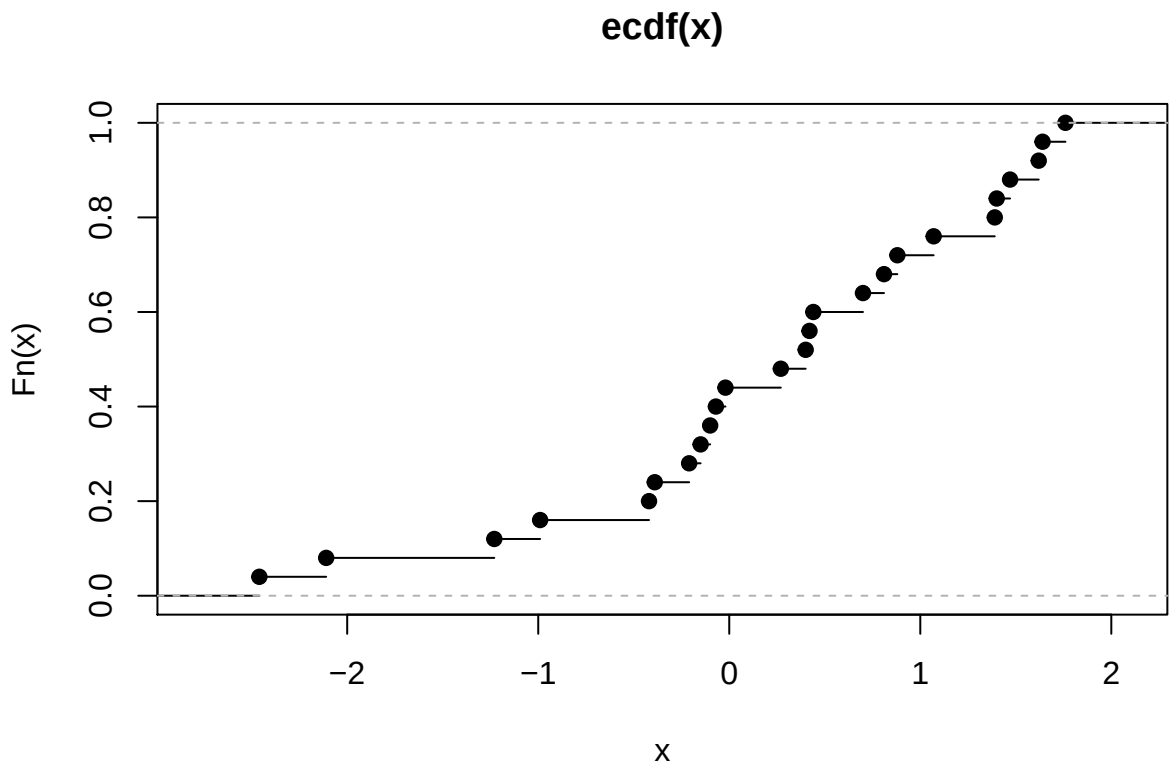
Since p-value of of ks.test(X,Y) is much larger than 0.05. Then we fail to reject the null hypotheses, thus X and Y are from the same distribution. Since p-value of of ks.test(X+2,Y) is much smaller than 0.05.Then we reject the null hypotheses, therefore X+2 and Y are not from the same distribution

4.

```
Df4=readRDS("norm_sample.Rdata")
Df4_ecdf=ecdf(Df4)
summary(Df4_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
```

```
plot.ecdf(Df4)
```



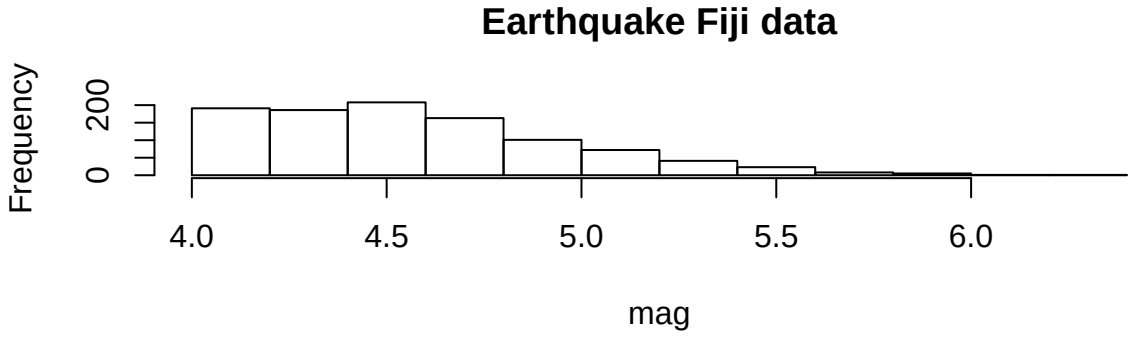

```
us=rnorm(25)
ks.test(Df4,us)
```

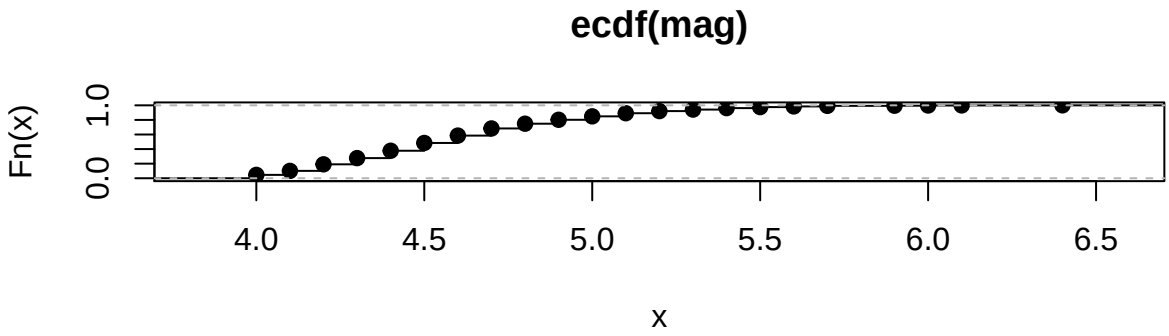
```
##
##  Two-sample Kolmogorov-Smirnov test
##
```

```
## data:  Df4 and us
## D = 0.2, p-value = 0.7102
## alternative hypothesis: two-sided
```

5.

```
#Find a 95for F(4.9) ??? F(4.3).
Df51=read.table("faithful.dat",header=TRUE,skip=25,sep="")
Df52=read.table("fijiquakes.dat",header=T,sep="")
 mag<-Df52$mag
par(mfrow=c(2,1))
hist( mag,main="Earthquake Fiji data")
u<-ecdf(mag)
plot(u)
```

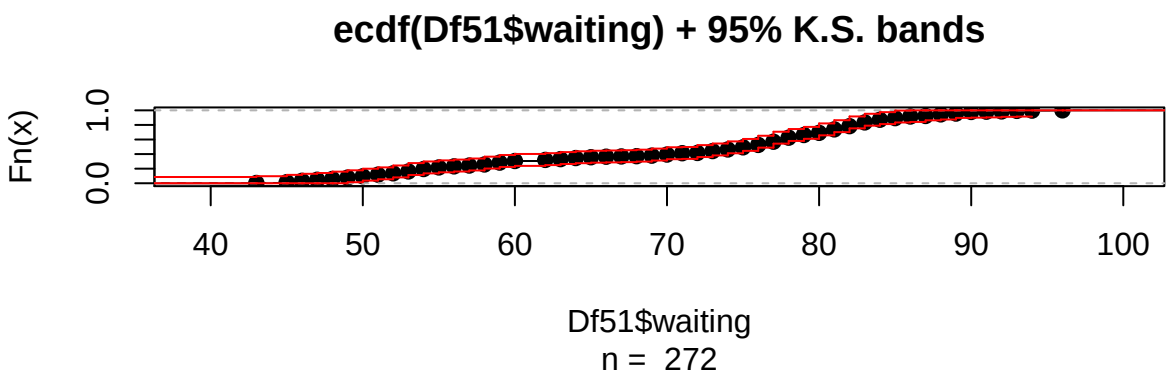



```
print(th<-u(4.9)-u(4.3))
```

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

```
# Estimate a 90 percent confidence interval for the mean waiting time
ecdf.ksCI(Df51$waiting)
```


ecdf(Df51$waiting) + 95% K.S. bands
Fn(x)
0.0
1.0
40
50
60
70
80
90
100
Df51$waiting
n = 272