

MA677_HW3

Chaoqun Yin

3/8/2019

1.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
library(plotly)
```

```
##
## Attaching package: 'plotly'
## The following object is masked from 'package:ggplot2':
##
##   last_plot
## The following object is masked from 'package:stats':
##
##   filter
## The following object is masked from 'package:graphics':
##
##   layout
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.2.1 --
## v tibble  1.4.2    v purrr   0.2.4
## v tidyr   0.8.0    v stringr 1.3.0
## v readr   1.1.1    v forcats 0.3.0
## -- Conflicts ----- tidyverse_conflicts() --
## x plotly::filter() masks dplyr::filter(), stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
## The following object is masked from 'package:purrr':
##
##   set_names
```

```
## The following object is masked from 'package:tidyr':  
##  
##   extract
```

```
library(readr)  
library(RCurl)
```

```
## Loading required package: bitops
```

```
##
```

```
## Attaching package: 'RCurl'
```

```
## The following object is masked from 'package:tidyr':
```

```
##
```

```
##   complete
```

```
library(ACSWR)  
df=read.table("maybe_uniform.txt")
```

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

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

```
##
```

```
## One-sample Kolmogorov-Smirnov test
```

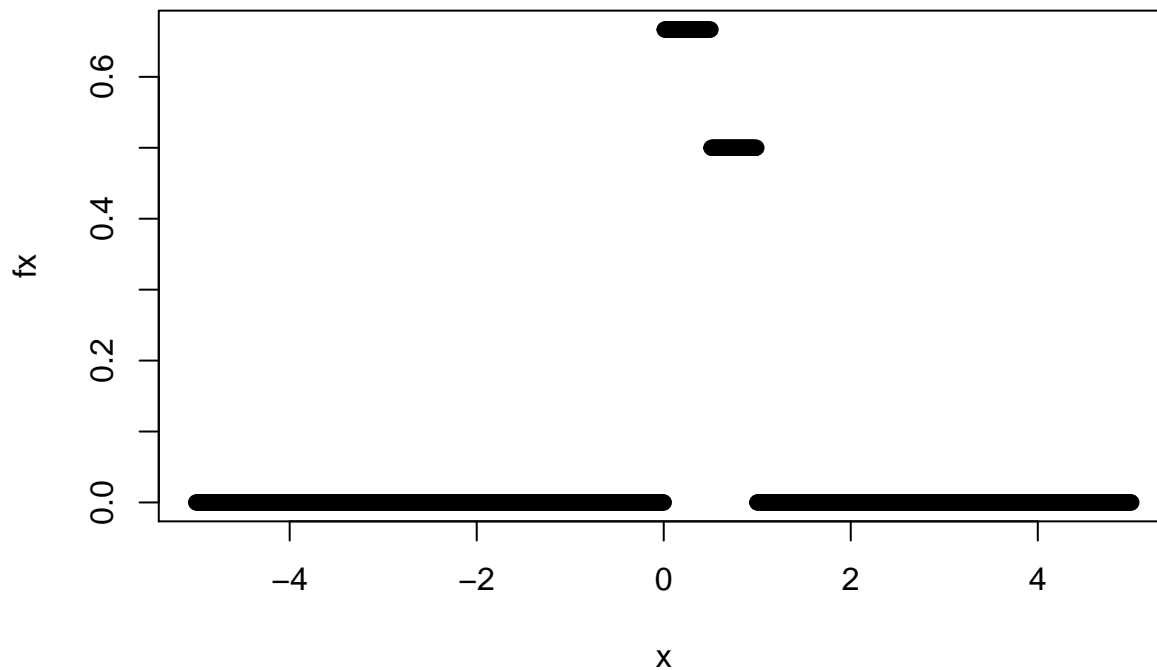
```
##
```

```
## 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)
```



```
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, these data is distributed as a Uniform distribution on $[0,1]$. Since in ks-test, the D value of fx is less than that of data, the model below is better than the Uniform.

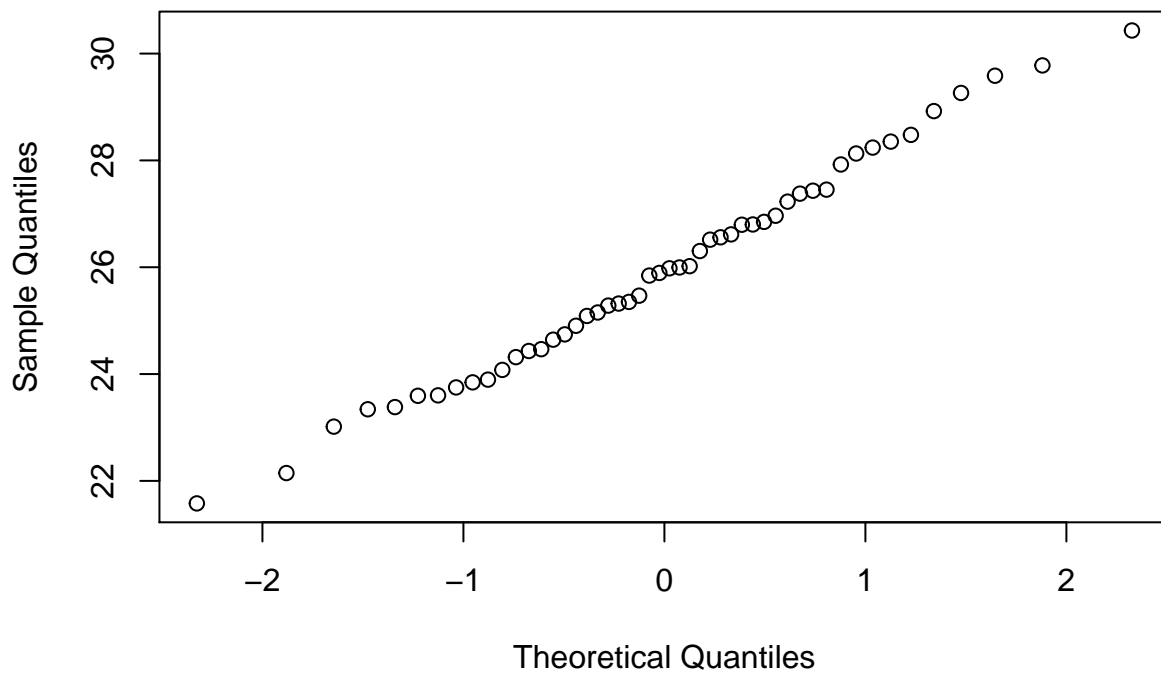
2.

```
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(26,50))/2  
ks.test(test2,"pnorm")
```

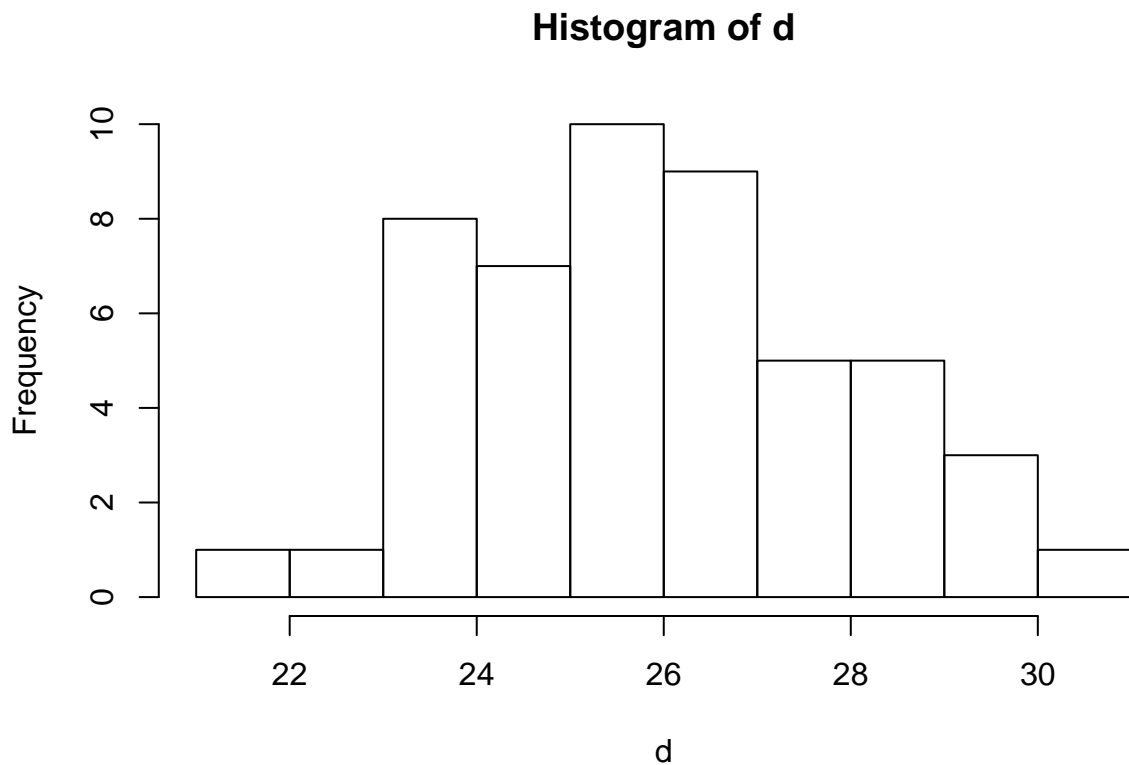
```
##  
## One-sample Kolmogorov-Smirnov test  
##  
## data: test2  
## D = 0.06722, p-value = 0.9663  
## alternative hypothesis: two-sided
```

```
qqnorm(d)
```

Normal Q-Q Plot



```
hist(d)
```



According

to Ks-test and qqnorm, the data generally follow normal distribution.

3.

```
df3=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(df3$V1,df3$V2,df3$V3,df3$V4,df3$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
```

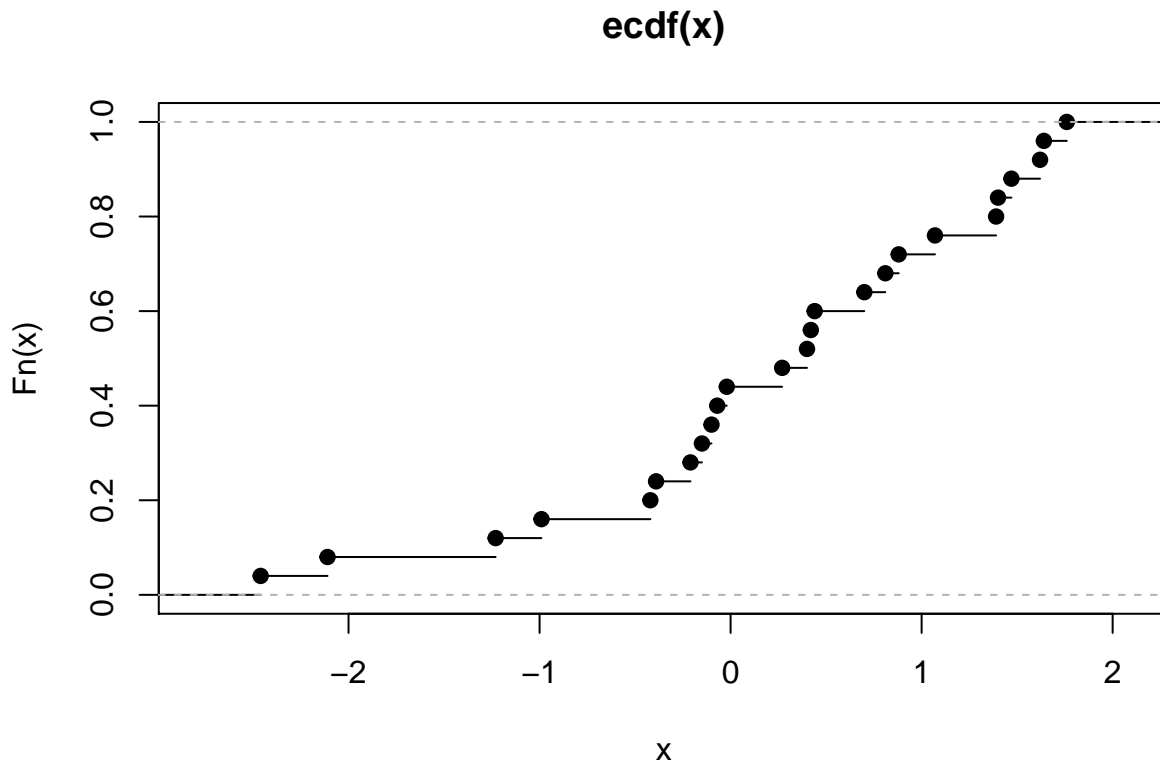
Since p-value of $ks.test(X,Y)$ is much larger than 0.05. Then we fail to reject the null hypotheses, thus X and Y are from the same distribution. Since p-value of $ks.test(X+2,Y)$ is much smaller than 0.05. Then we reject the null hypotheses, therefore $X+2$ and Y are not from the same distribution

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 = 0.24, p-value = 0.4755
## alternative hypothesis: two-sided
```