Homework 4

Wenjia Xie

March 8, 2019

Problem One

(1)

```
uni <- c(0.42,0.38,0.48,0.11,0.30,0.06,0.78,0.35,0.29,0.23,0.88,0.71,0.16, 0.
79,0.01,0.40,0.57,0.22,0.75,0.41,0.90,0.66,0.08,0.82,0.09)
ks.test(uni,"punif")

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

                                                                  Since
 the p-value here is 0.3501, we can not reject the null hypothesis. The data
is distributed as a uniform distribution.
```

(2)

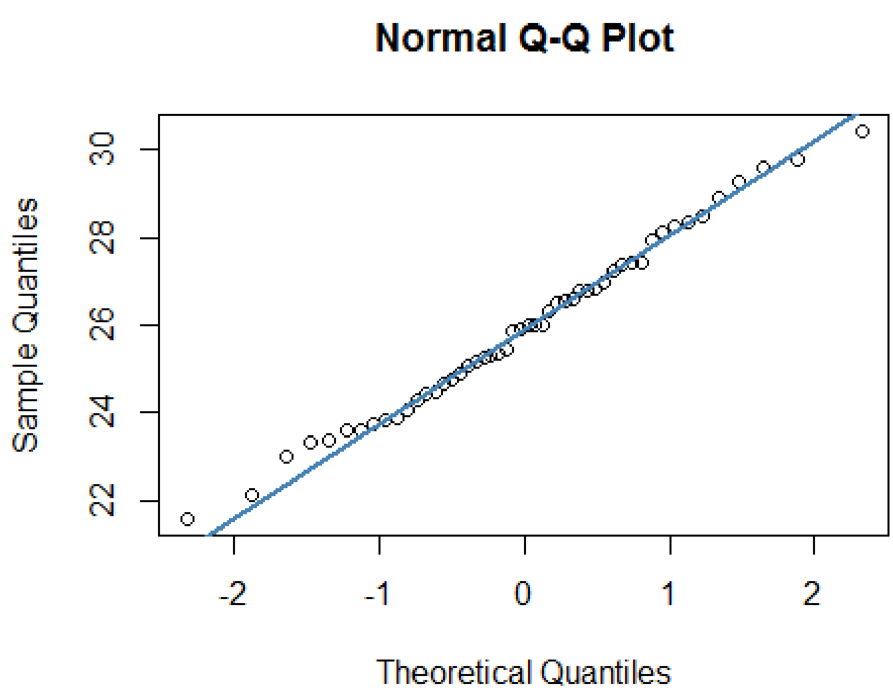
```
x <- seq(-1,1,0.01)
fx <- ifelse(x > 0 & x <=0.5, 2/3,
   ifelse(x > 0.5 & x < 1,  0.5, 0))
ks.test(uni,fx)

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

##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  uni and fx
## D = 0.50746, p-value = 2.127e-05
## alternative hypothesis: two-sided
```

The p-value is so small that we reject the null hypothesis.Thus,the data is not distributed as fx.

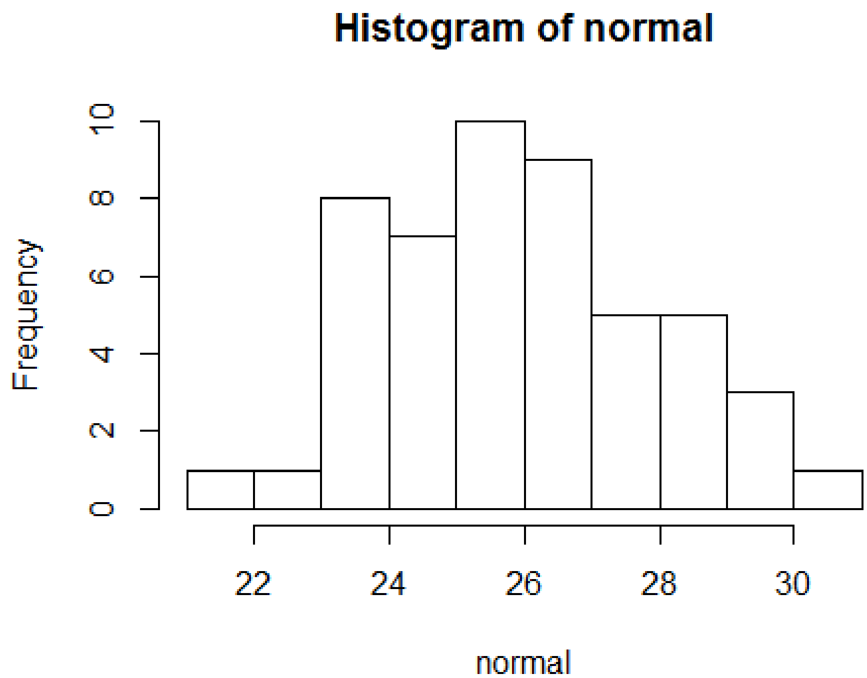
(3)

Problem Two

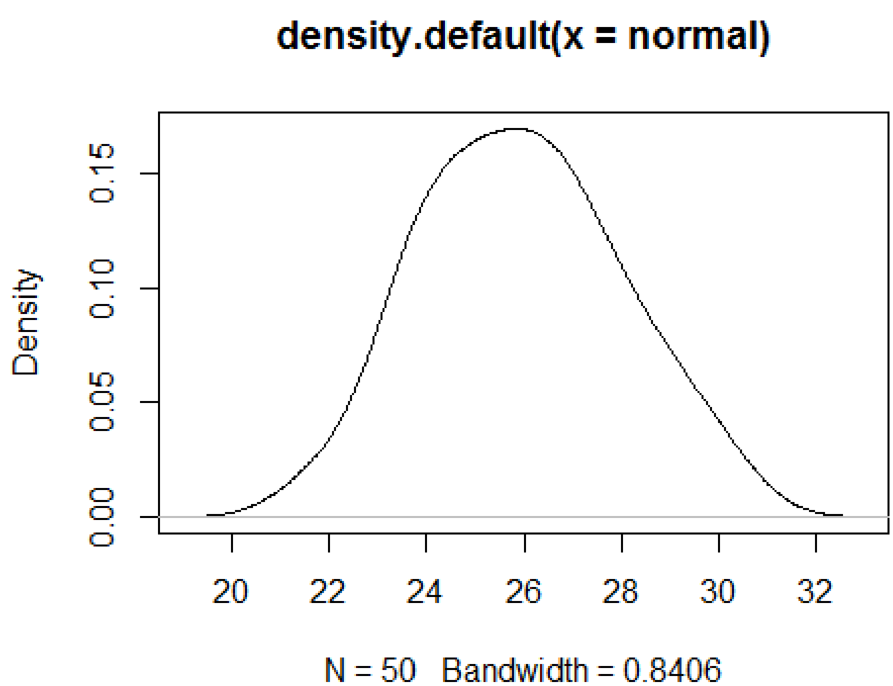
```
normal <- c(25.088,26.615,25.468,27.453,23.845,
25.996,26.516,28.240,25.980,30.432,
26.560,25.844,26.964,23.382,25.282,
24.432,23.593,24.644,26.849,26.801,
26.303,23.016,27.378,25.351,23.601,
24.317,29.778,29.585,22.147,28.352,
29.263,27.924,21.579,25.320,28.129,
28.478,23.896,26.020,23.750,24.904,
24.078,27.228,27.433,23.341,28.923,
24.466,25.153,25.893,26.796,24.743)
ks.test(normal,"pnorm",26,2)

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

qqnorm(normal)
qqline(normal, col = "steelblue", lwd = 2)
```



```r
hist(normal)
```




```r
plot(density(normal))
```



The p-value is large enough and we do not have sufficient evidence to reject the null hypothesis.From the qqplot,histgram and density plot, we can also draw the same conclusion that these data come from normal distribution.

Problem Three

```
X <- c(0.61,0.29,0.06,0.59,-1.73,
-0.74,0.51,-0.56,-0.39,1.64,
0.05,-0.06,0.64,-0.82,0.31,
1.77,1.09,-1.28,2.36,1.31,
1.05,-0.32,-0.40,1.06,-2.47)

Y <- c(2.20,1.66,1.38,0.20,
0.36,0.00,0.96,1.56,
0.44,1.50,-0.30,0.66,
2.31,3.29,-0.27,-0.37,
0.38,0.70,0.52,-0.71)

ks.test(X,Y)

##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  X and Y
## D = 0.27, p-value = 0.3357
## 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 = 0.56, p-value = 0.001881
## alternative hypothesis: two-sided
```

Using ks.test, we can find that X and Y are from the same distribution. However, X+2 and Y are not from the same distribution.

Problem Four

```
data4 <- readRDS("norm_sample.Rdata")
ks.test(data4,"pnorm")

##
##  One-sample Kolmogorov-Smirnov test
##
## data:  data4
```

```
## D = 0.17724, p-value = 0.3683
## alternative hypothesis: two-sided
```

```
set.seed(1)
ecdf(data4)
```

```
## Empirical CDF
## Call: ecdf(data4)
##  x[1:25] =  -2.46,  -2.11,  -1.23,  ...,   1.64,   1.76
```

```
standnorm <-rnorm(n = length(data4),mean = 0,sd = 1)
diff <- sort(data4)-sort(standnorm)
D <- max(abs(diff))
```

problem Five

```
fiji<-read.table("fijiquakes.dat",header = T)
mag<-fiji$mag
Fn <- ecdf(mag)
library(Hmisc)
```

```
## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
total<-sum( (mag<=4.9) & (mag>4.3))
binconf(total,length(mag),method="wilson",0.05)
```

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

The 95% for F(4.9) – F(4.3) is [0.50,0.56].

```
faith<-read.table("faithful.dat",skip = 25)
waiting<-faith$waiting
avg<-mean(waiting)
var<-var(waiting)
n<-length(waiting)
L<-round(avg-qnorm(0.95)*sqrt(var/n),2)
U<-round(avg+qnorm(0.95)*sqrt(var/n),2)
```

```r
print(paste("the 90% CI for mean waiting time is:[",L,",",U,"]"))

## [1] "the 90% CI for mean waiting time is:[ 69.54 , 72.25 ]"

median(waiting)

## [1] 76
```

For the faithful data, the 90 percent confidence interval for the mean waiting time is [69.54,72.25]. The median of the waiting time is 76.