Week 3

3.1

```
s <- scan()
74.3 78.8 68.8 78.0 70.4 80.5 80.5 69.7 71.2 73.5
79.5 75.6 75.0 78.8 72.0 72.0 72.0 74.3 71.2 72.0
75.0 73.5 78.8 74.3 75.8 65.0 74.3 71.2 69.7 68.0
73.5 75.0 72.0 64.3 75.8 80.3 69.7 74.3 73.5 73.5
75.8 75.8 68.8 76.5 70.4 71.2 81.2 75.0 70.4 68.0
70.4 72.0 76.5 74.3 76.5 77.6 67.3 72.0 75.0 74.3
73.5 79.5 73.5 74.7 65.0 76.5 81.6 75.4 72.7 72.7
67.2 76.5 72.7 70.4 77.2 68.8 67.3 67.3 67.3 72.7
75.8 73.5 75.0 73.5 73.5 73.5 72.7 81.6 70.3 74.3
73.5 79.5 70.4 76.5 72.7 77.2 84.3 75.0 76.5 70.4
outline <- function(x){
  n <- length(x)                  #长度
  m <- mean(x)                    #样本均值
  v <- var(x)                     #样本方差
  s <- sd(x)                      #样本标准方差
  me <- median(x)                 #中位数
  cv <- 100*s/m                   #变异系数
  css <- sum((x-m)^2)             #样本校正平方和
  uss <- sum(x^2)                 #样本未校正平方和
  R <- max(x)-min(x)              #样本极差
  R1 <- quantile(x,3/4)-quantile(x,1/4)    #样本四分差（半极差）
  sm <- s/sqrt(n)                 #样本标准误
  u3  <- sum((x-m)^3)/n           #样本三阶中心距
  g1 <- n^2/((n-1)*(n-2))*u3/s^3  #偏度系数
  u4 <- sum((x-m)^4)/n            #样本四阶中心距
  g2 <- ((n^2*(n+1))/((n-1)*(n-2)*(n-3))*u4/s^4 - (3*(n-1)^2)/((n-
2)*(n-3)))     #峰度系数
  data.frame(N=n, Mean=m, Var=v, std_dev=s,
    Median=me, std_mean=sm, CV=cv, CSS=css, USS=uss,
    R=R, R1=R1, Skewness=g1, Kurtosis=g2, row.names=1)
}
d <- outline(s)
d[c('Mean', 'Var', 'std_dev', 'R', 'std_mean', 'CV', 'Skewness', 'Kurtosis')]
```

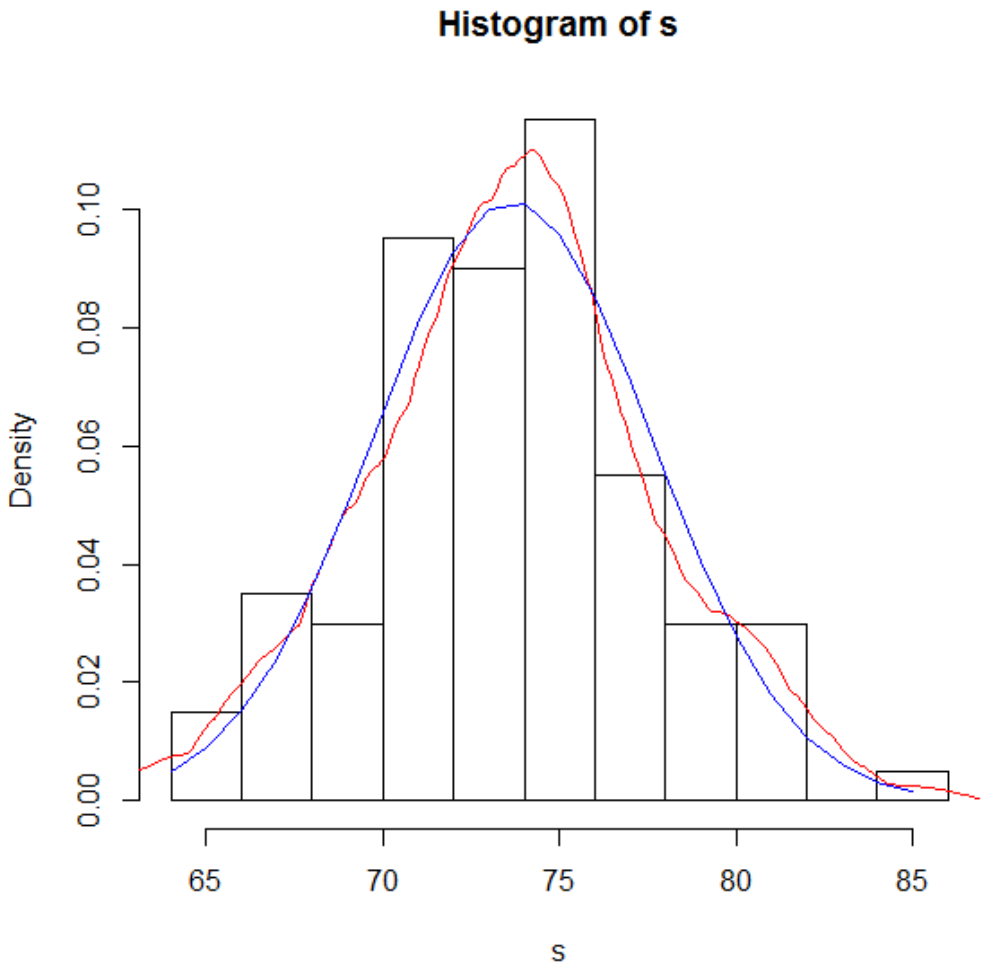
最后返回结果：

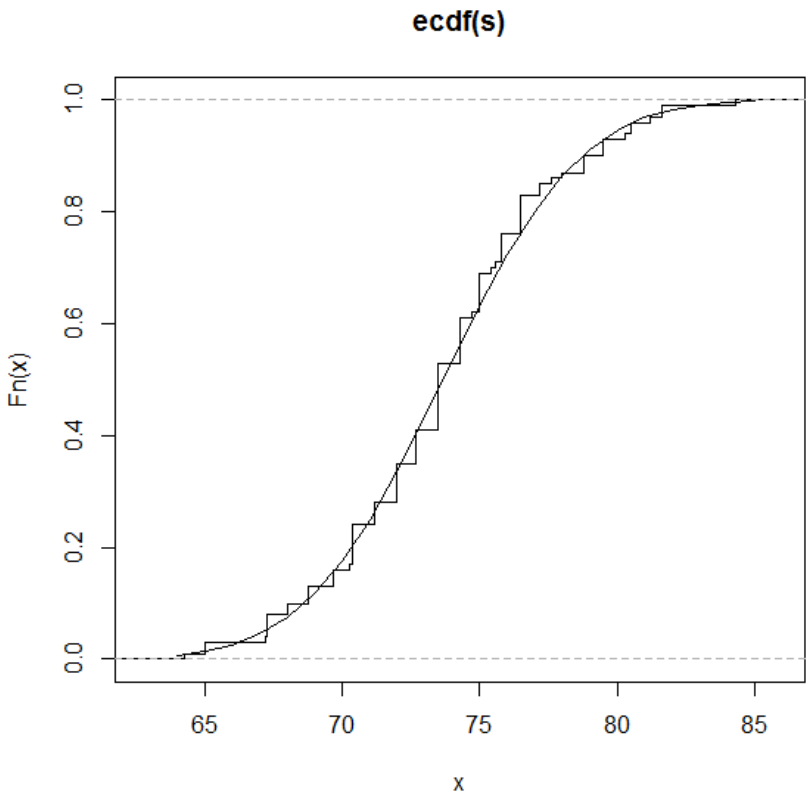
Mean	Var	std_dev	R	std_mean	CV	Skewness	Kurtosis

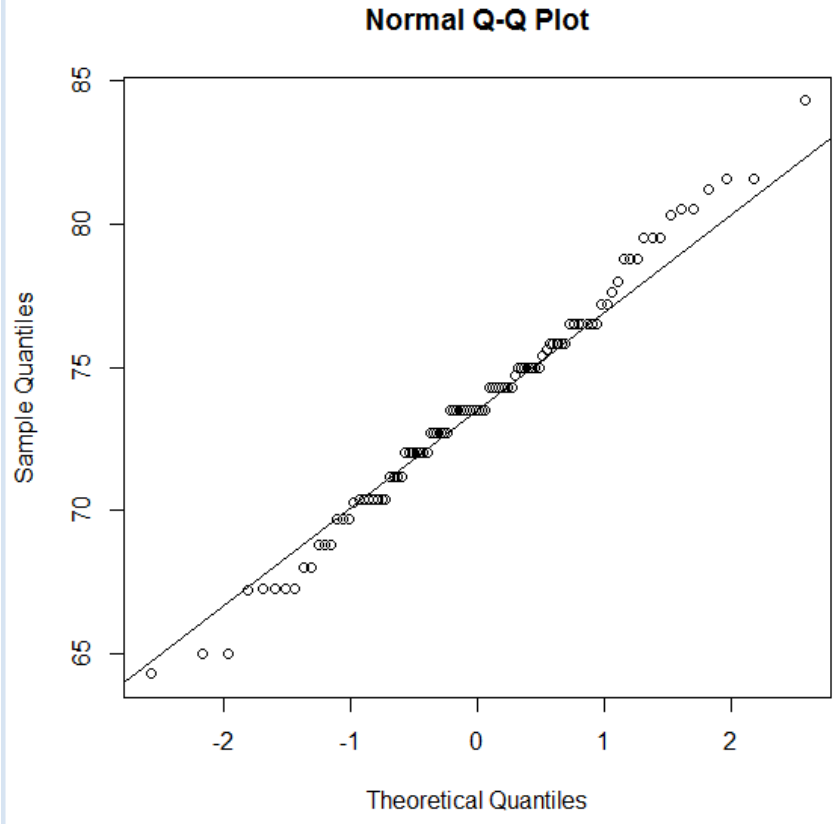
1 73.668 15.51513 3.938925 20 0.3938925 5.34686 0.0540593 0.03702249

3.2

```
#QQ 图
hist(s, freq = F)     #直方图
lines(density(s, kernel = "epanechnikov"), col = "red")     #密度估计曲线
lines(64:85, dnorm(64:85, mean(s), sd(s)), col = "blue")     #正态密度曲线
plot(ecdf(s),verticals = TRUE, do.p = FALSE)     #经验分布图
lines(64:85, pnorm(64:85, mean(s), sd(s)))     #正态分布曲线
qqnorm(s);qqline(s)                                  #QQ 图
```




ecdf(s)
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0
65
70
75
80
85
x


Normal Q-Q Plot
Sample Quantiles
65
70
75
80
85
-2
-1
0
1
2
Theoretical Quantiles

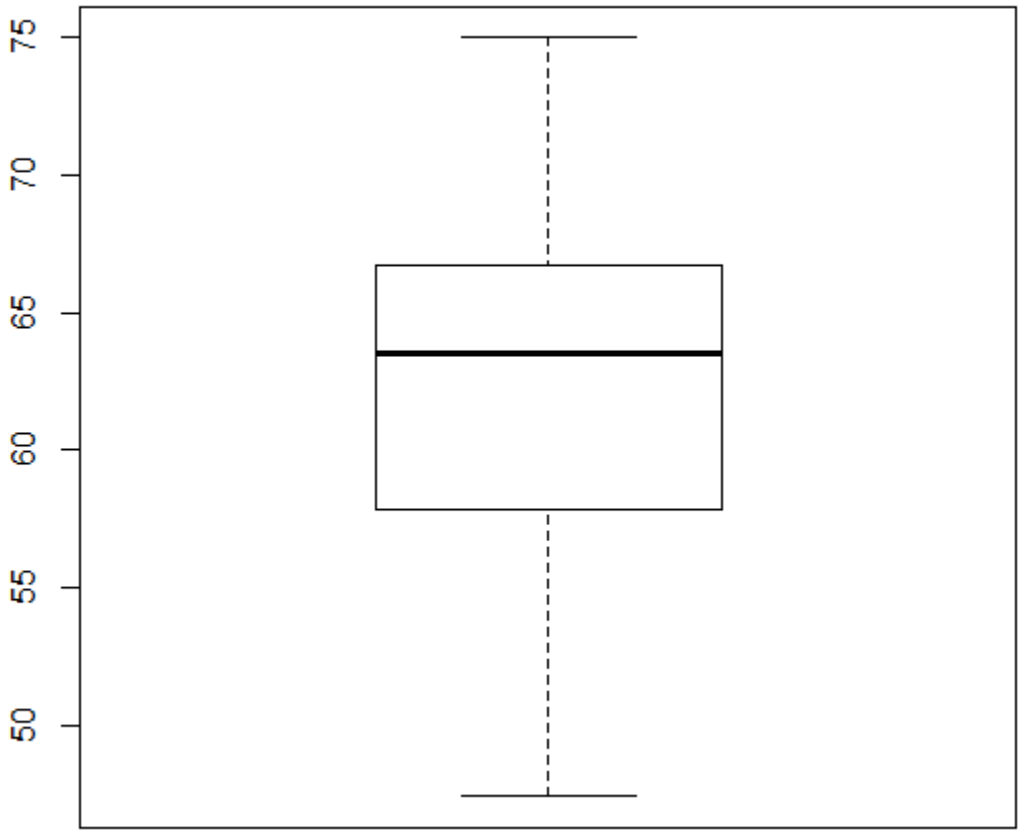
3.3

stem(s, scale=0.5)　#茎叶图

```
The decimal point is 1 digit(s) to the right of the |

6 | 4
6 | 557777788999
7 | 000000000001111222222333333444444444444444444
7 | 555555556666667777777788999
8 | 0000111224
```

boxplot(w)



fivenum(s)

3.4

shapiro.test(s)

ks.test(s, 'pnorm')

3.9

```
> Sys.setlocale(,"CHS")
[1] "LC_COLLATE=Chinese (Simplified)_People's Republic of
China.936;LC_CTYPE=Chinese (Simplified)_People's Republic of
China.936;LC_MONETARY=Chinese (Simplified)_People's Republic of
China.936;LC_NUMERIC=C;LC_TIME=Chinese (Simplified)_People's
Republic of China.936"
> s = read.table('student.data', header=T)
> cor(s[c('年龄','身高','体重')], method='pearson')
          年龄       身高       体重
年龄 1.0000000 0.8114343 0.7408855
身高 0.8114343 1.0000000 0.8777852
体重 0.7408855 0.8777852 1.0000000
> cor.test(~身高+体重, data=s)

        Pearson's product-moment correlation

data:  身高 and 体重
t = 7.5549, df = 17, p-value = 7.887e-07
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.7044314 0.9523101
sample estimates:
      cor
0.8777852
```

6.1

```
s <- read.table('6.1.data', header=T)
(1)plot(s$X, s$Y)
(2)
X <- s$X
Y <- s$Y
lm.snow<-lm(Y ~ 1+X)
(3)summary(lm.snow)
从分析结果（***）看，说明线形回归极为显著
(4)
new <- data.frame(X = 7)
lm.pred<-predict(lm.snow, new, interval="prediction", level=0.95)
lm.pred
```

结果如下：

	fit	lwr	upr
1	2690.227	2454.971	2925.484

预测值 2690.227，区间估计[2454.971，2925.484]