

R (1)

1.文件flowering.txt为100株玉米的开花时间，试利用read.table()读入数据：

- (1) 计算基本描述统计值，包括平均数、中值、方差、标准差。
- (2) 测验数据是否服从正态分布，作出说明。
- (3) 试作直方图，箱线图 boxplot，正态分布 QQ 图，并将3个图放在一个图中。
- (4) 试测验该数据与理论均值 $\mu=68$ 是否存在显著差异，并对结果给以说明。

执行代码如下：

```
y=read.table("C:\\Users\\czh\\Desktop\\flowering.txt",sep="\t",header=T)
mean(y[,1])
median(y[,1])
var(y[,1])
sd(y[,1])
shapiro.test(y[,1])
par=(mfrow=c(1,3))
hist(y[,1])
boxplot(y[,1])
qqnorm(y[,1])
qqline(y[,1],col="red")
t.test(y[,1],mu=68)
```

结果如下：

其中第一行为平均数，第二行为中值，第三行为方差，第四行为标准差。

```
PS C:\Users\czh> Rscript "c:\Users\czh\Desktop\Code\try.r"
[1] 71.24752
[1] 71
[1] 8.668119
[1] 2.944167

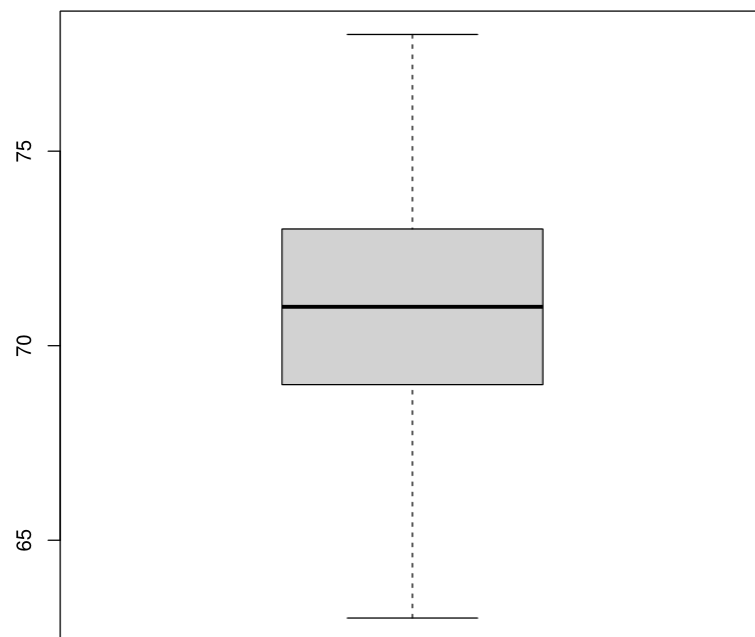
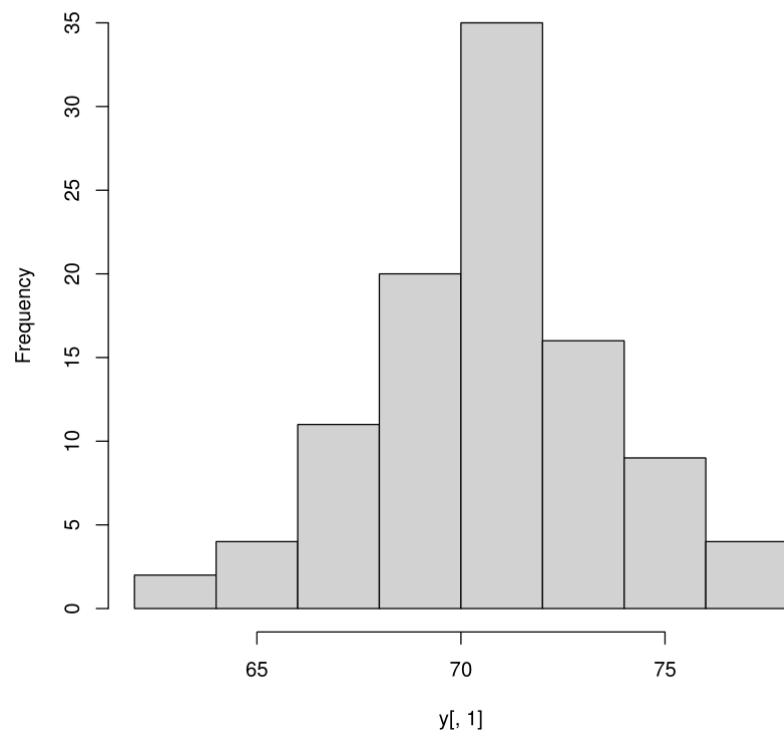
      Shapiro-Wilk normality test

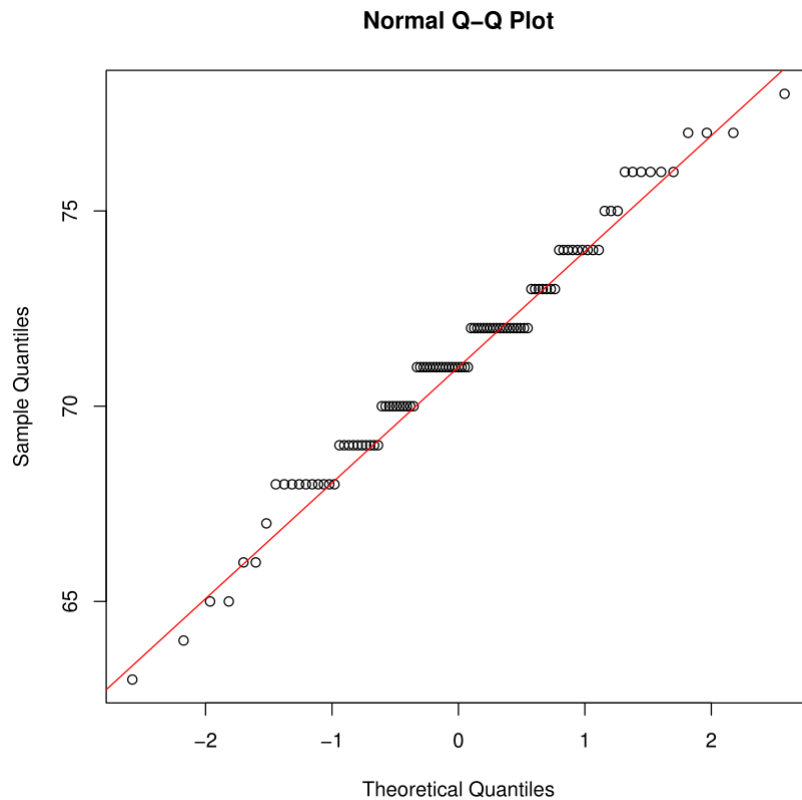
data:  y[, 1]
W = 0.97952, p-value = 0.1181

      One Sample t-test

data:  y[, 1]
t = 11.085, df = 100, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 68
95 percent confidence interval:
 70.66631 71.82874
sample estimates:
mean of x
 71.24752
```

Histogram of $y[, 1]$





分析略.

2.文件 `fertilizer.txt` 为新肥料(new)与原肥料(old)产量比较试验数据, 试测验两样本方差是否差异显著, 并分析新肥料与原肥料的产量平均值是否有差异. 并利用 `boxplot` 显示新肥料和原肥料的表现.

执行代码如下:

```
a=read.table("C:\\Users\\czzh\\Desktop\\fertilizer.txt",sep="\t",header=T)
y=a[1:30,2]
z=a[31:60,2]
var.test(y,z)
t.test(y,z)
boxplot(y,z,col=c("red","blue"),ylab="field",xlab="treatment",names=c("new","old"))
```

结果如下:

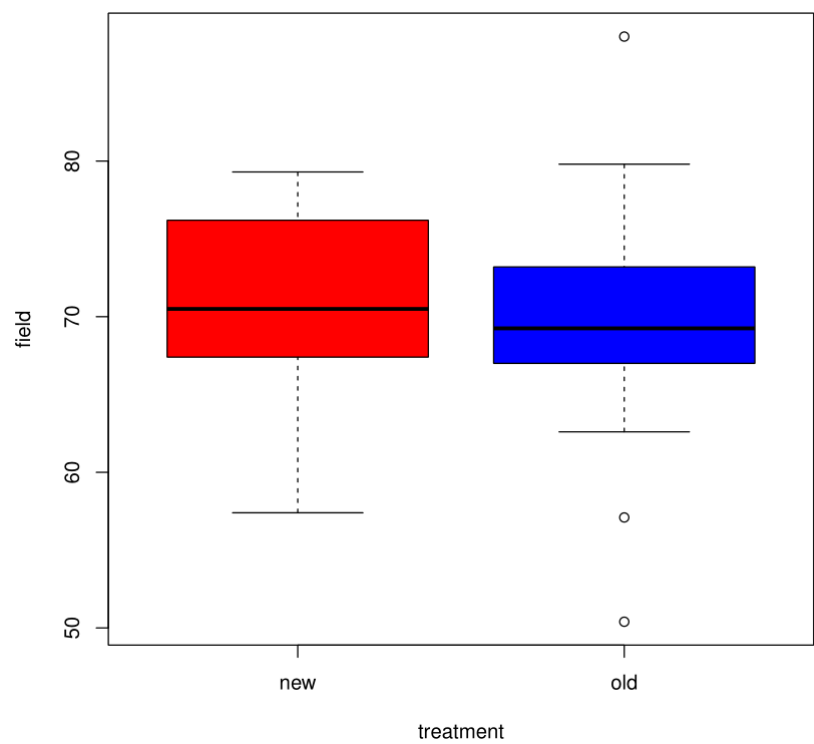
```
PS C:\Users\czzh> Rscript "c:\Users\czzh\Desktop\Code\try.r"

      F test to compare two variances

data:  y and z
F = 0.70793, num df = 29, denom df = 29, p-value = 0.3577
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3369484 1.4873522
sample estimates:
ratio of variances
 0.7079273

      Welch Two Sample t-test

data:  y and z
t = 0.81847, df = 56.352, p-value = 0.4165
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.895851  4.515851
sample estimates:
mean of x mean of y
 70.71667  69.40667
```



分析略.