

## 《多元统计分析及 R 语言建模》第 2 章

王斌会

2020.2.1

```
rm(list=ls())           #清理内存
options(digits=4)        #输出结果位数
par(mar=c(4,4,2,1))     #设置图片输出位置
library(openxlsx)
library(knitr)
```

### 2.1

对下面的相关系数矩阵，试用 R 语言求其逆矩阵、特征根和特征向量。要求写出 R 语言计算函数。

$$R = \begin{bmatrix} 1.00 & 0.80 & 0.26 & 0.67 & 0.34 \\ 0.80 & 1.00 & 0.33 & 0.59 & 0.34 \\ 0.26 & 0.33 & 1.00 & 0.37 & 0.21 \\ 0.67 & 0.59 & 0.37 & 1.00 & 0.35 \\ 0.34 & 0.34 & 0.21 & 0.35 & 1.00 \end{bmatrix}$$

```
R=matrix(c(1.00,0.80,0.26,0.67,0.34,0.80,1.00,0.33,0.59,0.34,0.26,0.33,
1.00,0.37,0.21,0.67,0.59,0.37,1.00,0.35,0.34,0.34,0.21,0.35,1.00),nrow=
5,ncol=5);R           #生成矩阵R
```

```
      [,1] [,2] [,3] [,4] [,5]
[1,] 1.00 0.80 0.26 0.67 0.34
[2,] 0.80 1.00 0.33 0.59 0.34
[3,] 0.26 0.33 1.00 0.37 0.21
[4,] 0.67 0.59 0.37 1.00 0.35
[5,] 0.34 0.34 0.21 0.35 1.00
```

```
R.=solve(R);R.
```

```
      [,1] [,2] [,3] [,4] [,5]
[1,] 3.3881 -2.1222 0.23706 -1.0685 -0.10623
[2,] -2.1222 2.9421 -0.33593 -0.1331 -0.16164
[3,] 0.2371 -0.3359 1.20699 -0.3764 -0.08812
[4,] -1.0685 -0.1331 -0.37637 2.0091 -0.21562
[5,] -0.1062 -0.1616 -0.08812 -0.2156 1.18505
```

```
R.e=eigen(R,symmetric = T);R.e
```

```
eigen() decomposition
$values
[1] 2.7923 0.8263 0.7791 0.4206 0.1818
```

```
$vectors
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] -0.5255  0.34022 -0.1665  0.15938  0.74494
[2,] -0.5187  0.23435 -0.1778  0.50823 -0.62142
[3,] -0.3131 -0.90308 -0.2287  0.14943  0.10844
[4,] -0.4966  0.03869 -0.1186 -0.83116 -0.21673
[5,] -0.3318 -0.11084  0.9350  0.05616  0.01355
```

## 2.2

某厂对 50 个计件工人某月份工资进行登记，获得以下原始资料（单位：元）。试按组距为 300 编制频数表，计算频数、频率和累计频率，并绘制直方图。

```
1465 1760 1985 2270 2980 1375 1735 1940 2220 2670 1405 1755 1965 2240
2820 1295 1645 1880 2110 2550 1355 1710 1910 2190 2600 1265 1625 1865
2095 2520 1225 1605 1845 2040 2430 1175 1835 2030 2370 1000 1535 1810
2010 2290 1125 1575 1815 2030 2320
```

### 2.2.1

写出 R 语言程序。

```
d2.2=read.xlsx('mvexer5.xlsx',sheet='E2.2');tail(d2.2) #读取数据并显示后
六行
```

```
      X
45 2290
46 1125
47 1575
48 1815
49 2030
50 2320
```

先对数据进行简单的统计分析：

```
min(d2.2)
```

```
[1] 1000
```

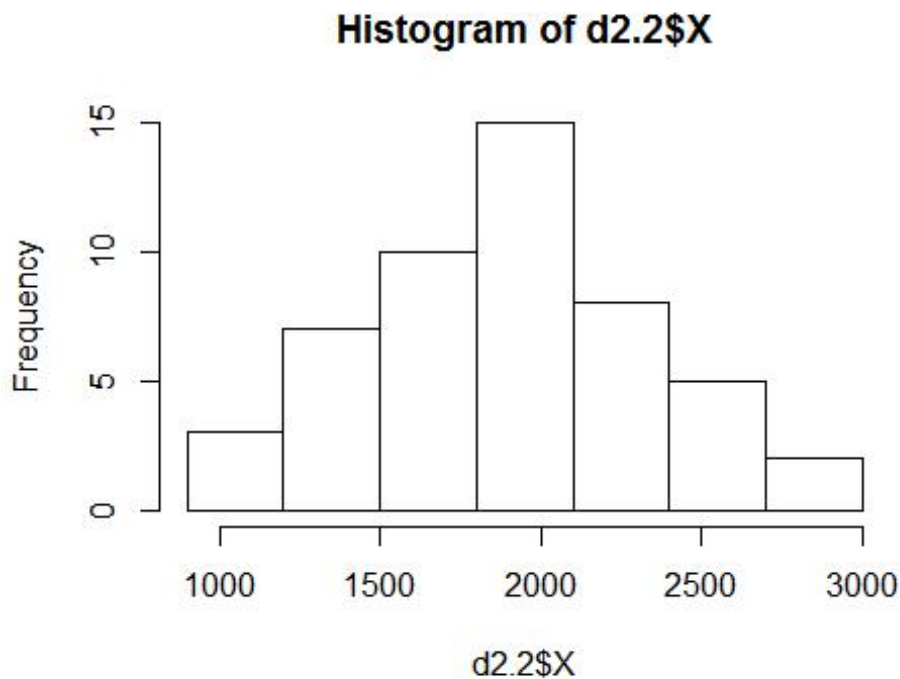
```
max(d2.2)
```

```
[1] 2980
```

可以得到某月份最低工资为 1000 元,最高工资为 2980 元。

画出组距为 300 的直方图：

```
H1=hist(d2.2$X,breaks=seq(900,3000,300)) #以组距为 300 画直方图
```



编制频数表，列出频数、频率和累计频率：

```
names(H1)
```

```
[1] "breaks" "counts" "density" "mids" "xname" "equidist"
```

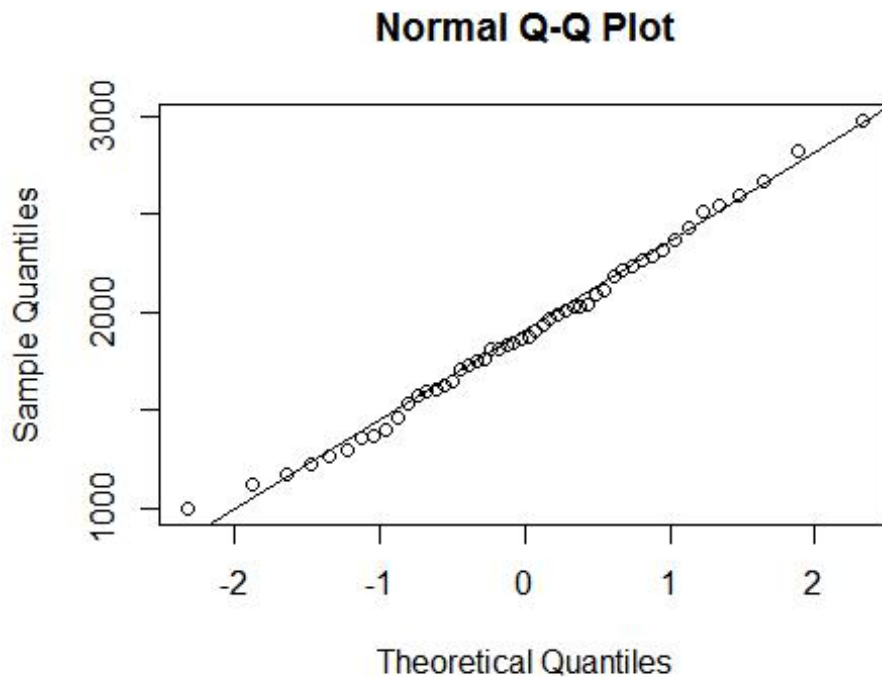
```
data.frame('组中值'=H1$mids, '频数'=H1$counts, '频率'=H1$density*300, '累计频率'=cumsum(H1$density*300))
```

	组中值	频数	频率	累计频率
1	1050	3	0.06	0.06
2	1350	7	0.14	0.20
3	1650	10	0.20	0.40
4	1950	15	0.30	0.70
5	2250	8	0.16	0.86
6	2550	5	0.10	0.96
7	2850	2	0.04	1.00

### 2.2.3

用 R 语言作正态概率图并分析之：

```
qqnorm(d2.2$X);qqline(d2.2$X)
```



由正态概率图可以看出数据点基本是一条直线，因此数据可以近似看作正态分布。

## 2.3

以下是一份关学生是否抽烟与每天学习时间长短关系的调查数据，试用 R 语言对其进行基本统计分析。

```
d2.3=read.xlsx('mvexer5.xlsx',sheet='E2.3');d2.3 #读取数据
```

编号	是否抽烟	每天学习时间
1	1	是 少于 5 小时
2	2	否 5-10 小时
3	3	否 5-10 小时
4	4	是 超过 10 小时
5	5	否 超过 10 小时
6	6	是 少于 5 小时
7	7	是 5-10 小时
8	8	是 少于 5 小时
9	9	否 超过 10 小时
10	10	是 5-10 小时

对数据进行简单的统计分析：

```
dim(d2.3);names(d2.3)
```

```
[1] 10 3

[1] "编号"          "是否抽烟"      "每天学习时间"

str(d2.3)

'data.frame': 10 obs. of 3 variables:
 $ 编号      : num  1 2 3 4 5 6 7 8 9 10
 $ 是否抽烟   : chr  "是" "否" "否" "是" ...
 $ 每天学习时间: chr  "少于 5 小时" "5-10 小时" "5-10 小时" "超过 10 小时" ...

summary(d2.3)

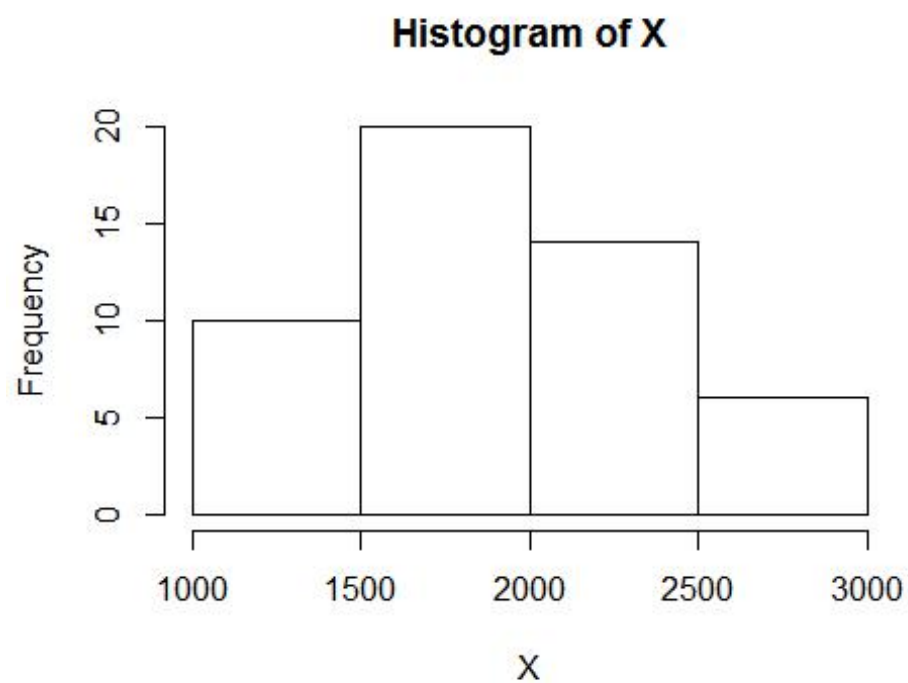
      编号      是否抽烟      每天学习时间
Min.   : 1.00   Length:10      Length:10
1st Qu.: 3.25   Class :character  Class :character
Median : 5.50   Mode  :character  Mode  :character
Mean    : 5.50
3rd Qu.: 7.75
Max.    :10.00
```

可以看到抽取学生中 4 位不抽烟，6 位抽烟，每天学习时间小于 5 小时的有 3 位，5-10 小时的有 4 位，超过 10 小时的有 3 位，抽烟情况和学习时间分别的分布比较均匀。

## 2.4

试编制进行计量数据频数表分析的 R 语言函数。

```
ch=function(X){
  H=hist(X)
  n=H$breaks[2]-H$breaks[1]
  c=data.frame('组中值'=H$mids,'频数'=H$counts,'频率'=H$density*n,'累计
频率'=cumsum(H$density*n))
  c
}
ch(d2.2$X)
```



	组中值	频数	频率	累计频率
1	1250	10	0.20	0.20
2	1750	20	0.40	0.60
3	2250	14	0.28	0.88
4	2750	6	0.12	1.00