多元数据分析第一次作业

应数 2101　杨嘉昱　2216113458

2024 年 3 月 9 日

【题目 1】 代码如下

```
library(e1071)
a = read.table("exercise.txt")
X_1 <- as.numeric(a[,2])
X_2 <- as.numeric(a[,3])
mean(X_1)
mean(X_2)
var(X_1)
var(X_2)
sd(X_1)
sd(X_2)
sd(X_1) / mean(X_1) * 100
sd(X_2) / mean(X_2) * 100
skewness(X_1)
skewness(X_2)
kurtosis(X_1)
kurtosis(X_2)
quantile(X_1,c(0.5, 0.25, 0.75))
quantile(X_2,c(0.5, 0.25, 0.75))
quantile(X_1, 0.75) - quantile(X_1, 0.25)
quantile(X_2, 0.75) - quantile(X_2, 0.25)
hist(X_1)
hist(X_2)
plot(ecdf(X_1), verticals = T, main = "EDF of X_1")
plot(ecdf(X_2), verticals = T, main = "EDF of X_2")
X <- (a[,2:3])
data_Pearson = round(cor(X, method = "pearson"), 2)
data_Pearson
data_Spearman = round(cor(X, method = "spearman"), 2)
data_Spearman
```

1. X_1 的均值、方差、标准差、变异系数、偏度、峰度分别为

   19.16645　392.0308　19.79977　103.3043　2.277166　6.182618

   X_2 的均值、方差、标准差、变异系数、偏度、峰度分别为

   246.1932　54276　232.9721　94.62978　1.734529　3.107709

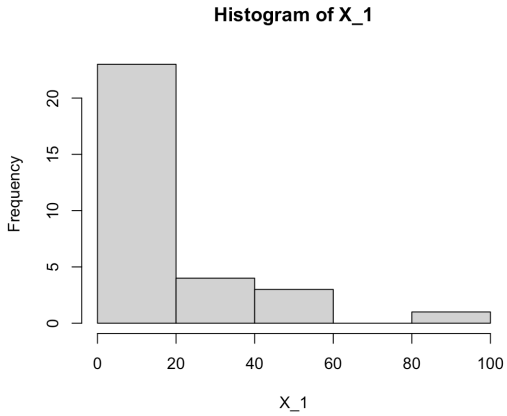
2. X_1 的中位数，上、下四分位数，四分位极差分别为

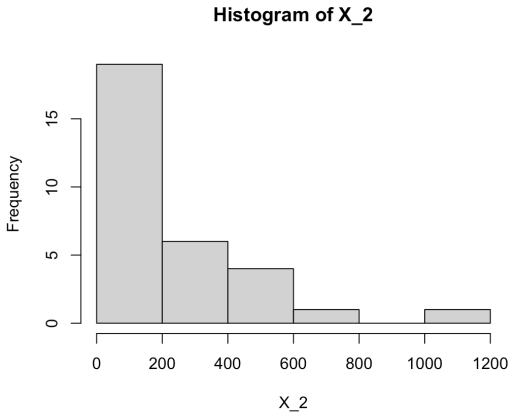
$$14.770 \quad 8.265 \quad 20.080 \quad 11.815$$

X_2 的中位数，上、下四分位数，四分位极差分别为

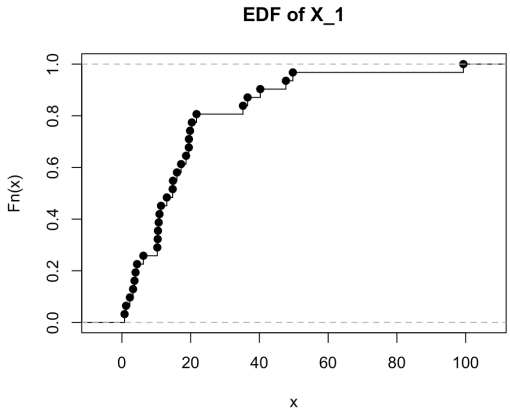
$$179.410 \quad 105.350 \quad 270.745 \quad 165.395$$

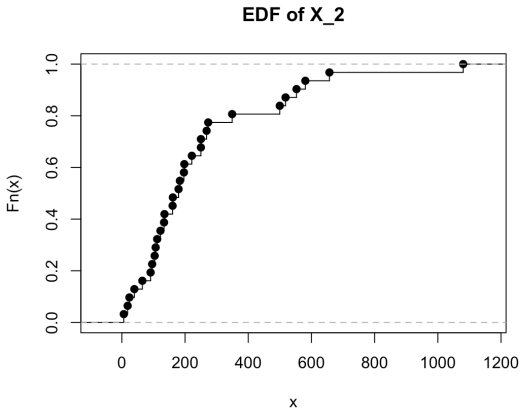
3. 直方图





4. 经验分布函数图





5. Pearson 相关系数与 Spearman 相关系数分别为

$$\begin{pmatrix} 1.00 & 0.98 \\ 0.98 & 1.00 \end{pmatrix} \quad \begin{pmatrix} 1.00 & 0.93 \\ 0.93 & 1.00 \end{pmatrix}$$

【题目 2】 代码如下

```
a = read.table("exercise1_5.txt")
mu = c()
M = c()
for (i in 1:4) {
    mu <- append(mu, round(mean(a[,i]), 1))
```

```
    M <- append(M, quantile(a[,i], 0.5))
}

p <- dim(a)[2]
p.value <- matrix(0,p,p)
for (i in 1:(p - 1)){
    for (j in (i + 1):p){
    test.obj <- cor.test(a[,i],a[,j],method = 'pearson')
    p.value[i,j] <- test.obj$p.value
    }
}


q <- dim(a)[2]
q.value <- matrix(0,q,q)
for (i in 1:(q - 1)){
    for (j in (i + 1):q){
    test.obj <- cor.test(a[,i],a[,j],method = 'spearman',exact=FALSE)
    q.value[i,j] <- test.obj$p.value
    }
}

mu
M
cov(a)
cor(a, method = "pearson")
cor(a, method = "spearman")
p.value
q.value
```

总体均值向量，总体中位数向量，总体协方差矩阵为

$$\boldsymbol{\mu} = \begin{pmatrix} 18.2 & 27.9 & 4.5 & 33.8 \end{pmatrix} \qquad \boldsymbol{M} = \begin{pmatrix} 18.1 & 27.4 & 4.8 & 34.1 \end{pmatrix}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} 13.508619 & 2.707167 & 1.019405 & 1.265667 \\ 2.707167 & 3.559333 & 1.138667 & 1.289333 \\ 1.019405 & 1.138667 & 1.998476 & 1.739667 \\ 1.265667 & 1.289333 & 1.739667 & 4.032333 \end{pmatrix}$$

Pearson 相关矩阵和 Spearman 相关矩阵为

$$\boldsymbol{R} = \begin{pmatrix} 1.0000000 & 0.7660596 & 0.3849719 & 0.3364907 \\ 0.7660596 & 1.0000000 & 0.4269360 & 0.3403319 \\ 0.3849719 & 0.4269360 & 1.0000000 & 0.6128276 \\ 0.3364907 & 0.3403319 & 0.6128276 & 1.0000000 \end{pmatrix}$$

$$\boldsymbol{S} = \begin{pmatrix} 1.0000000 & 0.7896983 & 0.4339915 & 0.4305367 \\ 0.7896983 & 1.0000000 & 0.5111078 & 0.4884056 \\ 0.4339915 & 0.5111078 & 1.0000000 & 0.6911813 \\ 0.4305367 & 0.4884056 & 0.6911813 & 1.0000000 \end{pmatrix}$$

Pearson 检验的显著性 p 值为

$$\begin{pmatrix} 0 & 5.152838e-05 & 0.08483807 & 0.135839682 \\ 0 & 0.000000e+00 & 0.05357904 & 0.131150557 \\ 0 & 0.000000e+00 & 0.00000000 & 0.003140558 \\ 0 & 0.000000e+00 & 0.00000000 & 0.000000000 \end{pmatrix}$$

可见 p_{14}, p_{24} 的 p 值大于 $\alpha = 0.10$，可认为相关性不显著。

Spearman 检验的显著性 p 值为

$$\begin{pmatrix} 0 & 2.070355e-05 & 0.04933616 & 0.0513801108 \\ 0 & 0.000000e+00 & 0.01788785 & 0.0246757038 \\ 0 & 0.000000e+00 & 0.00000000 & 0.0005210014 \\ 0 & 0.000000e+00 & 0.00000000 & 0.0000000000 \end{pmatrix}$$

可见均小于 $\alpha = 0.10$，认为均有相关性。

【题目 3】 代码如下

```
a = read.table("exercise1_7.txt")
for (i in 1:3) {
    mu <- append(mu, round(mean(a[,i]), 1))
    M <- append(M, quantile(a[,i], 0.5))
}


p <- dim(a)[2]
p.value <- matrix(0,p,p)
for (i in 1:(p - 1)){
    for (j in (i + 1):p){
    test.obj <- cor.test(a[,i],a[,j],method = 'pearson')
    p.value[i,j] <- test.obj$p.value
    }
}

q <- dim(a)[2]
q.value <- matrix(0,q,q)
for (i in 1:(q - 1)){
    for (j in (i + 1):q){
    test.obj <- cor.test(a[,i],a[,j],method = 'spearman',exact=FALSE)
    q.value[i,j] <- test.obj$p.value
    }
}

mu
M
cor(a, method = "pearson")
cor(a, method = "spearman")
p.value
q.value

```

总体均值向量，总体中位数向量为

$$\boldsymbol{\mu} = \begin{pmatrix} 14.4 & 16.0 & 4.2 \end{pmatrix} \qquad \boldsymbol{M} = \begin{pmatrix} 15 & 15 & 4 \end{pmatrix}$$

Pearson 相关矩阵和 Spearman 相关矩阵为

$$\boldsymbol{R} = \begin{pmatrix} 1.0000000 & 0.6193020 & 0.5195171 \\ 0.6193020 & 1.0000000 & 0.4614949 \\ 0.51951717 & 0.4614949 & 1.0000000 \end{pmatrix}$$

$$\boldsymbol{S} = \begin{pmatrix} 1.0000000 & 0.5455093 & 0.5066848 \\ 0.5455093 & 1.0000000 & 0.5295098 \\ 0.5066848 & 0.5295098 & 1.0000000 \end{pmatrix}$$

Pearson 检验和 Spearman 检验的显著性 p 值为

$$\begin{pmatrix} 0 & 1.629059e-06 & 0.0001105698 \\ 0 & 0.000000e+00 & 0.0007427133 \\ 0 & 0.000000e+00 & 0.0000000000 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 4.187896e-05 & 1.735774e-04 \\ 0 & 0.000000e+00 & 7.684268e-05 \\ 0 & 0.000000e+00 & 0.000000e+00 \end{pmatrix}$$

可见均小于 $\alpha = 0.10$，认为均有相关性。