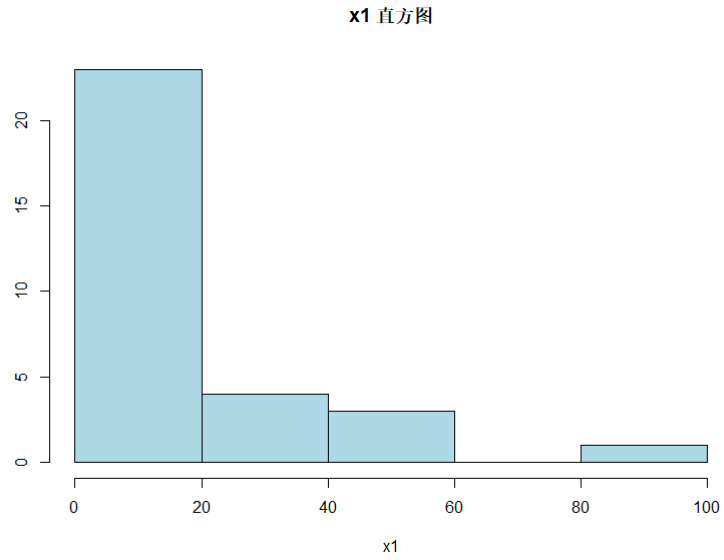
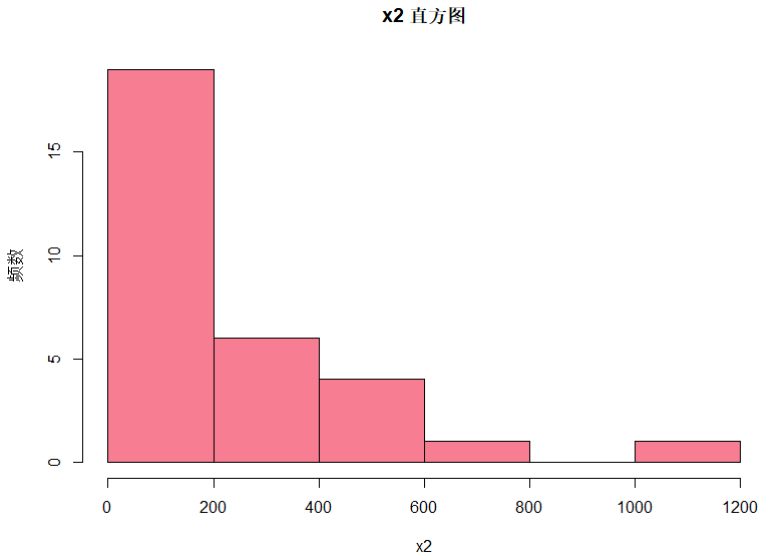
日期	科目	班级	姓名	学号
2023 年 3 月 7 日	数据分析	强基数学 002	吴天阳	2204210460

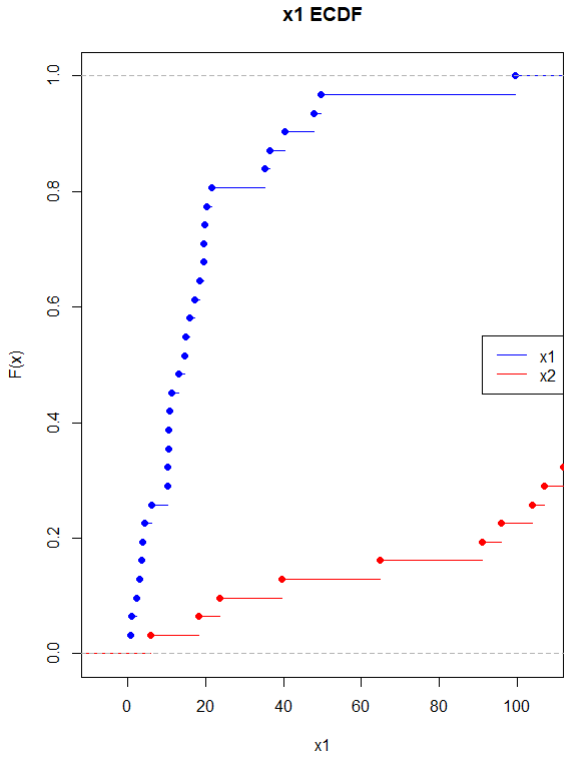
第一章作业

1.4 解答.

```
x1: 均值 = 19.16645, 方差 = 392.0308, 标准差 = 19.79977, 变异系数 = 103.3043, 偏度 =
↪  2.391967, 峰度 = 9.804995
x2: 均值 = 246.1932, 方差 = 54276, 标准差 = 232.9721, 变异系数 = 94.62978, 偏度 =
↪  1.821974, 峰度 = 6.521676
x1: 上四分位数 = 8.265, 中位数 = 14.77, 下四分位数 = 20.08
x2: 上四分位数 = 105.35, 中位数 = 179.41, 下四分位数 = 270.745
Pearson 相关系数 = 0.9762474, Spearman 相关系数 = 0.9278226
```







1.5 解答. 均值向量为

```
       V1        V2        V3        V4
18.219048 27.866667  4.504762 33.766667
```

协方差矩阵为

```
         V1       V2       V3       V4
V1 3.508619 2.707167 1.019405 1.265667
V2 2.707167 3.559333 1.138667 1.289333
V3 1.019405 1.138667 1.998476 1.739667
V4 1.265667 1.289333 1.739667 4.032333
```

1.6 解答.

```
中位数向量
  V1   V2   V3   V4
18.1 27.4  4.8 34.1

Pearson 相关矩阵
          V1        V2        V3        V4
V1 1.0000000 0.7660596 0.3849719 0.3364907
V2 0.7660596 1.0000000 0.4269360 0.3403319
V3 0.3849719 0.4269360 1.0000000 0.6128276
V4 0.3364907 0.3403319 0.6128276 1.0000000

Spearman 相关矩阵
          V1        V2        V3        V4
V1 1.0000000 0.7896983 0.4339915 0.4305367
V2 0.7896983 1.0000000 0.5111078 0.4884056
V3 0.4339915 0.5111078 1.0000000 0.6911813
V4 0.4305367 0.4884056 0.6911813 1.0000000

Pearson 检验
             [,1]          [,2]          [,3]        [,4]
[1,] 0.000000e+00  5.152838e-05  8.483807e-02 0.135839682
[2,] 5.152838e-05 3.532105e-150  5.357904e-02 0.131150557
[3,] 8.483807e-02  5.357904e-02 3.532105e-150 0.003140558
[4,] 1.358397e-01  1.311506e-01  3.140558e-03 0.000000000

Spearman 检验
             [,1]         [,2]         [,3]          [,4]
[1,] 0.000000e+00 2.070355e-05 0.0493361598  5.138011e-02
[2,] 2.070355e-05 0.000000e+00 0.0178878482  2.467570e-02
[3,] 4.933616e-02 1.788785e-02 0.0000000000  5.210014e-04
[4,] 5.138011e-02 2.467570e-02 0.0005210014 2.557517e-147
```

1.4 代码：

```r
setwd("E:/Coding/R/习题一")
data <- read.table("exercise1_4.txt", fileEncoding = "GB2312")
x1 <- data[, 2]
x2 <- data[, 3]
cat("x1 =", x1, "\nx2 =", x2, "\n")

# 计算均值
mean_x1 <- mean(x1)
mean_x2 <- mean(x2)

# 计算方差
var_x1 <- var(x1)
var_x2 <- var(x2)

# 计算标准差
sd_x1 <- sd(x1)
sd_x2 <- sd(x2)

# 计算变异系数
cv_x1 <- sd_x1 / mean_x1 * 100
cv_x2 <- sd_x2 / mean_x2 * 100

# 计算偏度
skewness_x1 <- moments::skewness(x1)
skewness_x2 <- moments::skewness(x2)

# 计算峰度
kurtosis_x1 <- moments::kurtosis(x1)
kurtosis_x2 <- moments::kurtosis(x2)

# 输出结果
cat("x1: 均值 = ", mean_x1, ", 方差 = ", var_x1, ", 标准差 = ", sd_x1,
    ", 变异系数 = ", cv_x1, ", 偏度 = ", skewness_x1, ", 峰度 = ", kurtosis_x1,
    "\n", sep = "")
cat("x2: 均值 = ", mean_x2, ", 方差 = ", var_x2, ", 标准差 = ", sd_x2,
    ", 变异系数 = ", cv_x2, ", 偏度 = ", skewness_x2, ", 峰度 = ", kurtosis_x2,
    "\n", sep = "")


# 计算上下四分位数及中位数
q1_x1 <- quantile(x1, probs = 0.25)
median_x1 <- quantile(x1, probs = 0.5)
q3_x1 <- quantile(x1, probs = 0.75)

q1_x2 <- quantile(x2, probs = 0.25)
median_x2 <- quantile(x2, probs = 0.5)
q3_x2 <- quantile(x2, probs = 0.75)

# 计算四分位极差
iqr_x1 <- q3_x1 - q1_x1
iqr_x2 <- q3_x2 - q1_x2

```

```r
cat("x1: 上四分位数 = ", q1_x1, ", 中位数 = ", median_x1, ", 下四分位数 = ", q3_x1,
    "\n", sep = "")
cat("x2: 上四分位数 = ", q1_x2, ", 中位数 = ", median_x2, ", 下四分位数 = ", q3_x2,
    "\n", sep = "")

png("x1_histogram.png", width = 800, height = 600, res = 96)
hist(x1, main = "x1 直方图", xlab = "x1", ylab = " 频数", col = "lightblue")
dev.off()
png("x2_histogram.png", width = 800, height = 600, res = 96)
hist(x2, main = "x2 直方图", xlab = "x2", ylab = " 频数", col = "#f77e92")
dev.off()

# 计算两组样本的 ECDF 函数
ecdf_x1 <- ecdf(x1)
ecdf_x2 <- ecdf(x2)

# 绘制 ECDF 图
png("ecdf.png", width = 600, height = 800, res = 96)
plot(ecdf_x1, main = " 经验分布函数图", xlab = "x", ylab = "F(x)", col = "blue")
lines(ecdf_x2, col = "red")
legend("right", legend = c("x1", "x2"), col = c("blue", "red"), lty = 1)
dev.off()

# Pearson 相关系数和 Spearman 相关系数
cor_pearson <- cor(x1, x2, method = "pearson")
cor_spearman <- cor(x1, x2, method = "spearman")

cat("Pearson 相关系数 = ", cor_pearson, ", Spearman 相关系数 = ", cor_spearman,
    "\n", sep = "")
```

1.5 和 1.6 代码：

```r
setwd("E:/Coding/R/习题一")
data <- read.table("exercise1_5.txt")

# 总体均值向量
data_mean <- colMeans(data)
print(data_mean)
# 总体协方差矩阵
data_cov <- cov(data)
print(data_cov)

# 中位数向量
data_median <- apply(data, 2, median)
cat(" 中位数向量\n")
print(data_median)

# Pearson 相关矩阵
R <- cor(data, method = "pearson")
cat("Pearson 相关矩阵\n")
print(R)
# Spearman 相关矩阵
Q <- cor(data, method = "spearman")
```

```r
cat("Spearman 相关矩阵\n")
print(Q)

# 计算 Pearson 两两列做相关性分析
pearson_values <- matrix(nrow = ncol(data), ncol = ncol(data))
spearman_values <- matrix(nrow = ncol(data), ncol = ncol(data))
for (j in 1:ncol(data)) {
    for (k in 1:ncol(data)) {
        pearson_test <- cor.test(data[, j], data[, k])
        pearson_values[j, k] <- pearson_test$p.value
        spearman_test <- cor.test(data[, j], data[, k], method = "spearman")
        spearman_values[j, k] <- spearman_test$p.value
    }
}
cat("Pearson 检验\n")
print(pearson_values)
cat("Spearman 检验\n")
print(spearman_values)
```