多元统计分析第四次作业

学习交流，无限进步

2024 年 10 月 18 日

Exercise 1

4. 设总体 $X \sim N_p(\mu, \Sigma)$，其中 $\mu = (\mu_1, \cdots, \mu_p)'$ 和 $\Sigma > 0$。假设 $x_1, \cdots, x_n$ 为来自 p 元正态总体 X 的一组独立同分布的简单随机样本，且 $n > p$。记 C 为 $k \times p$ 的常数矩阵和 r 为已知的 k 维向量，且要求 $k < p$ 和秩 $\text{rank}(C) = k$。试给出检验 $H_0 : C\mu = r$ 的检验统计量及其分布。

证明. $Y = CX \sim N_k(C\mu, C\Sigma C')$ ，则 $y_i = Cx_i$ 可以看作来自总体 Y 的一组独立同分布简单随机样本，原零假设等价于 $H_0 : \mu' = r$ ，其中 μ' 为总体 Y 的均值向量。
构造检验统计量：

$$t^2 = n(n-1)(\bar{y} - r)' V_y^{-1} (\bar{y} - r) \sim T^2(k, n-1)$$

其中

$$\begin{aligned} \bar{y} =& C\bar{x} \\ V_y =& \Sigma_{i=1}^n (y_i - \bar{y})(y_i - \bar{y})' \\ =& C[\Sigma_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})']C' \\ =& CV_xC' \end{aligned}$$

因此检验统计量及其分布为：

$$t^2 = n(n-1)(C\bar{x} - r)'(CV_xC')^{-1}(C\bar{x} - r) \sim T^2(k, n-1)$$

□

答案：

想法一：采用拉格朗日乘子法求解极大似然估计

全参数空间的极大似然估计容易得出，故这里只给出零假设参数空间下的极大似然估计

$$\begin{aligned} &max \quad \frac{1}{(2\pi)^{n/2}|\Sigma|^{n/2}} exp\{-\frac{1}{2}tr[\Sigma^{-1}(V+n(\bar{x}-\mu)(\bar{x}-\mu)'\} \\ &s.t. \quad c\mu = r \end{aligned}$$

考虑对数似然函数拉格朗日函数

$$L(\mu,\Sigma,\lambda) = -\frac{n\ln 2\pi}{2} - \frac{n\ln|\Sigma|}{2} - \frac{tr(\Sigma^{-1}V)}{2} - \frac{n(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)}{2} - \lambda'(c\mu - r)$$

运用以下矩阵/ 向量求导法则:（小写字母为向量，大写字母为矩阵）

$$\frac{\partial x'Ax}{\partial x} = 2Ax \tag{1}$$

$$\frac{\partial \ln|AXB|}{\partial X} = A'(AXB)^{-1}B' \tag{2}$$

$$\frac{\partial tr(AXB)}{\partial X} = A'B' \tag{3}$$

对 μ 求导

$$\begin{aligned} \frac{\partial L}{\partial \mu} &= \frac{\partial[-\frac{n(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)}{2} - \lambda'(c\mu - r)]}{\partial \mu} \\ &= -\Sigma^{-1}(\bar{x}-\mu) - c'\lambda \qquad (a) \end{aligned}$$

对 λ 求导

$$\frac{\partial L}{\partial \lambda} = C\mu - r \qquad (b)$$

对 Σ 求导, 此处为了方便，转变为对 Σ^{-1} 求导

$$\begin{aligned} \frac{\partial L}{\partial \Sigma^{-1}} &= \frac{\partial[-\frac{n\ln|\Sigma|}{2} - \frac{tr(\Sigma^{-1}V)}{2} - \frac{n(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)}{2}]}{\partial \Sigma^{-1}} \\ &= \frac{\partial[\frac{n\ln|\Sigma^{-1}|}{2} - \frac{tr(\Sigma^{-1}V)}{2} - \frac{ntr[(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)]}{2}]}{\partial \Sigma^{-1}} \\ &= \frac{n\Sigma - n(\bar{x}-\mu)(\bar{x}-\mu)' - V}{2} \qquad (c) \end{aligned}$$

令 (a),(b),(c) 式为 0 得:

$$\begin{aligned} \mu =& \bar{x} + \Sigma c' \lambda & (4) \\ c\mu =& r & (5) \\ \Sigma =& \frac{1}{n}[V + n(\bar{x} - \mu)(\bar{x} - \mu)'] & (6) \end{aligned}$$

将 (4) 代入 (5), 并将 (6) 左乘 c ，右乘 c' ，并代入 (5) 得:

$$\begin{aligned} \lambda =& -(c\Sigma c')^{-1}(c\bar{x} - r) & (7) \\ c\Sigma c' =& \frac{1}{n}[cVc' + n(c\bar{x} - r)(c\bar{x} - r)'] & (8) \end{aligned}$$

可以得到

$$\lambda = -nM^{-1}(c\bar{x} - r) \tag{9}$$

其中 $M = cVc' + n(c\bar{x} - r)(c\bar{x} - r)'$
将 (6),(9) 代入 (4) 可以得到关于 μ 的方程

$$(1 - na)\mu = (1 - na)\bar{x} - Vc'M^{-1}(c\bar{x} - r) \tag{10}$$

其中 $a = (c\bar{x} - r)'M^{-1}(c\bar{x} - r)$, 是一个常数
于是: $\mu = \bar{x} - \frac{Vc'M^{-1}(c\bar{x}-r)}{1-na}$
为了简化表达，我们考虑 $\frac{M^{-1}(c\bar{x}-r)}{1-na} = \frac{M^{-1}(c\bar{x}-r)}{1-n(c\bar{x}-r)'M^{-1}(c\bar{x}-r)}$
结论：Sherman-Morrison 公式
A 为可逆矩阵，u, v 为两个向量

$$(A + uv')^{-1} = A^{-1} - \frac{A^{-1}uv'A^{-1}}{1 + u'A^{-1}v}$$

证明过程可见 https://zhuanlan.zhihu.com/p/635551477
在本题中令 $A = cVc', u = n(c\bar{x} - r), v = (c\bar{x} - r)$ 于是:

$$M^{-1} = (cVc'(cvc')^{-1}(c\bar{x} - r)c')^{-1} - \frac{(cVc')^{-1}n(c\bar{x} - r)(c\bar{x} - r)'(cVc')^{-1}}{1 + n(c\bar{x} - r)'(cVc')^{-1}(c\bar{x} - r)}$$

于是

$$\begin{aligned} na =& n(c\bar{x}-r)'(cVc')^{-1}(c\bar{x}-r) - \frac{n^2(c\bar{x}-r)'(cVc')^{-1}(c\bar{x}-r)(c\bar{x}-r)'(cVc')^{-1}(c\bar{x}-r)}{1+n(c\bar{x}-r)'(cVc')^{-1}(c\bar{x}-r)} \quad (11)\\ =& \frac{n(c\bar{x}-r)'(cVc')^{-1}(c\bar{x}-r)}{1+n(c\bar{x}-r)'(cVc')^{-1}(c\bar{x}-r)} \quad (12) \end{aligned}$$

于是, 记 $\alpha = (c\bar{x}-r)$

$$\begin{aligned} \frac{M^{-1}(c\bar{x}-r)}{1-na} =& \frac{M^{-1}(c\bar{x}-r)}{1-n(c\bar{x}-r)'M^{-1}(c\bar{x}-r)} \\ =& \frac{(cVc')^{-1}[I-\frac{n\alpha\alpha'(cVc')^{-1}}{1+n\alpha'(cVc')^{-1}\alpha}]\alpha}{\frac{1}{1+n\alpha'(cVc')^{-1}\alpha}} \\ =& (cVc')^{-1}[I+n\alpha'(cVc')^{-1}\alpha I - n\alpha\alpha'(cVc')^{-1}]\alpha \\ =& (cVc')^{-1}[\alpha+n\alpha'(cVc')^{-1}\alpha I - n\alpha\alpha'(cVc')^{-1}\alpha] \\ =& (cVc')^{-1}[\alpha+n\alpha'(cVc')^{-1}\alpha I - n\alpha'(cVc')^{-1}\alpha\alpha] \\ =& (cVc')^{-1}\alpha \end{aligned}$$

回代得: $\mu = \bar{x} - Vc'(cVc')^{-1}(c\bar{x}-r)$;
$\Sigma = \frac{1}{n}[V + n(Vc'(cVc')^{-1}(c\bar{x}-r))(Vc'(cVc')^{-1}(c\bar{x}-r))']$
最后算出来的似然比统计量会和我自己的做法相同。
想法二: 拉格朗日乘子与代入消元相结合
得到 (6)(即先给定了 μ 求出 Σ 的最优解) 后要优化的对数似然函数可以消元为:

$$L(\mu) = ln\frac{1}{(2\pi)^{n/2}|\Sigma|^{n/2}}exp\{-\frac{np}{2}\}$$

这等加于最小化 $|n\Sigma| = |V + n(\bar{x}-\mu)(\bar{x}-\mu)'|$
而

$$\begin{aligned} |V + n(\bar{x}-\mu)(\bar{x}-\mu)'| =& |V||I + V^{-1/2}n(\bar{x}-\mu)(\bar{x}-\mu)'V^{-1/2}| \\ =& |V|(1+n(\bar{x}-\mu)'V^{-1}(\bar{x}-\mu)) \end{aligned}$$

于是在 $c\mu = r$ 的情况下最小化 $y = (\bar{x}-\mu)'V^{-1}(\bar{x}-\mu)$
求导:

$$\frac{\partial[y-\lambda'(c\mu-r)]}{\partial x} = 2V^{-1}(\mu-\bar{x}) - c'\lambda$$

$$\frac{\partial[y-\lambda'(c\mu-r)]}{\partial\lambda}=-c\mu+r$$

联立很容易解得：$\lambda=2(cVc')^{-1}(r-c\bar{x})$

$\mu=\frac{1}{2}Vc'\lambda+\bar{x}=\bar{x}-Vc'(cVc')^{-1}(c\bar{x}-r)$

代入可以解得 Σ 与上面结果相同。

Exercise 2

5. 设 $x_1,\cdots,x_n$ 为来自总体 $X\sim N_p(\mu,\Sigma)$ 的独立同分布的简单随机样本，其中 $\mu=(\mu_1,\cdots,\mu_p)'$，$\Sigma>0$ 且 $n>p$。记样本均值为 $\bar{x}$，样本离差阵为 V。考虑下面的假设检验问题：

$$H_0:\mu_1=\mu_2=\cdots=\mu_p,\quad H_1:\mu_1,\mu_2,\cdots,\mu_p\text{ 至少有一对不相等.}$$

令 C 为 $(p-1)\times p$ 的矩阵，记为

$$C=\begin{pmatrix}1&-1&0&\cdots&0\\1&0&-1&\cdots&0\\\vdots&\vdots&\vdots&&\vdots\\1&0&0&\cdots&-1\end{pmatrix}$$

则上面的假设等价于

$$H_0:C\mu=0_{p-1},\quad H_1:C\mu\neq 0_{p-1},$$

其中 0_{p-1} 为 $p-1$ 维零向量。试求检验 H_0 的似然比统计量及其分布。

证明. 任对总体作变换 $Y=CX\sim N_{p-1}(C\mu,C\Sigma C')$，则 $y_i=Cx_i$ 可以看作来自总体 Y 的一组独立同分布简单随机样本，原零假设等价于 $H_0:\mu_y=0$ 。

则对 $(y_1, y_2, \cdots, y_n)$ 的似然函数为

$$\begin{aligned}
L_y(\mu,\Sigma) =& \frac{1}{(2\pi)^{n/2}|\Sigma_y|^{n/2}} exp\{-\frac{1}{2}tr[\Sigma_y^{-1}(V_y + n(\bar{y}-\mu_y)(\bar{y}-\mu_y)')]\} \\
=& \frac{1}{(2\pi)^{n/2}|C\Sigma C'|^{n/2}} exp\{-\frac{1}{2}tr[(C\Sigma C')^{-1}(CVC' + n(C\bar{x}-C\mu)(C\bar{x}-C\mu)')]\}
\end{aligned}$$

在零假设成立时，μ_y, Σ_y 的极大似然估计为 $0_{p-1}, \frac{V_y + n\bar{y}\bar{y}'}{n}$ 。对于全参数空间，μ_y, Σ_y 的极大似然估计为 $\bar{y}, \frac{V_y}{n}$ ，代入得似然比检验统计量为：

$$\lambda = \frac{|V_y|^{n/2}}{|V_y + n\bar{y}\bar{y}'|^{n/2}}$$

而 $|V_y + n\bar{y}\bar{y}'| = |V_y||I_p + nV_y^{-1/2}\bar{y}\bar{y}'V_y^{-1/2}| = |V_y||1 + n\bar{y}V_y^{-1}\bar{y}'|$
于是

$$\lambda = \left(\frac{1}{1 + n\bar{y}V_y^{-1}\bar{y}'}\right)^{n/2}$$

于是 λ 随 $n\bar{y}V_y^{-1}\bar{y}'$ 单调递减
故可以 $t^2 = n(n-1)\bar{y}V_y^{-1}\bar{y}'$ 为检验统计量。

$$t^2 = n(n-1)\bar{y}V_y^{-1}\bar{y}' = n(n-1)(C\bar{x})'(CVC')^{-1}(C\bar{x}) \sim T^2(p-1, n-1)$$

$\square$

Exercise 3

8. 对两个 p 元正态总体 $N_p(\mu_1, \Sigma)$ 和 $N_p(\mu_2, \Sigma)$ 均值向量的检验问题，试用似然比原理导出检验 $H_0 : \mu_1 = \mu_2$ 的似然比统计量及其分布。

证明. 设 $x_1,x_2,\cdots,x_n$ 为来自第一个总体的独立样本，$y_1,y_2,\cdots y_m$ 为来自第二个总体的独立样本，且两组样本相互独立。

似然函数:

$$\begin{aligned}L(\mu_1,\mu_2,\Sigma)=&\frac{1}{(2\pi)^{n/2}|\Sigma|^{n/2}}exp\{-\frac{1}{2}tr[\Sigma^{-1}(V_x+n(\bar{x}-\mu_1)(\bar{x}-\mu_1)']\}\times\\&\frac{1}{(2\pi)^{m/2}|\Sigma|^{m/2}}exp\{-\frac{1}{2}tr[\Sigma^{-1}(V_y+m(\bar{y}-\mu_2)(\bar{y}-\mu_2)']\}\\=&\frac{1}{(2\pi)^{(m+n)/2}|\Sigma|^{(m+n)/2}}\times\\&exp\{-\frac{1}{2}tr[\Sigma^{-1}(V_x+V_y+n(\bar{x}-\mu_1)(\bar{x}-\mu_1)'+m(\bar{y}-\mu_2)(\bar{y}-\mu_2)']\}\end{aligned}$$

在零假设成立的条件下，取似然函数中与 $\mu_1=\mu_2=mu$ 有关的部分, 并取对数得

$$\begin{aligned}f=&-\frac{1}{2}tr[\Sigma^{-1}(n(\bar{x}-\mu_1)(\bar{x}-\mu_1)'+m(\bar{y}-\mu_2)(\bar{y}-\mu_2)')]\\=&-\frac{1}{2}[n(\bar{x}-\mu_1)'\Sigma^{-1}(\bar{x}-\mu_1)+m(\bar{y}-\mu_2)'\Sigma^{-1}(\bar{y}-\mu_2)]\\=&-\frac{1}{2}[n(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)+m(\bar{y}-\mu)'\Sigma^{-1}(\bar{y}-\mu)]\\\frac{\partial f}{\partial\mu}=&-\frac{1}{2}[-2n\Sigma^{-1}(\bar{x}-\mu)-2m\Sigma^{-1}(\bar{y}-\mu)]\\=&n\Sigma^{-1}(\bar{x}-\mu)+m\Sigma^{-1}(\bar{y}-\mu)\end{aligned}$$

可以解得 μ 的极大似然估计为: $\mu_0=\frac{n\bar{x}+m\bar{y}}{m+n}$，代入原似然函数可以得到 Σ 的极大似然估计为 $\Sigma_0=\frac{V_x+V_y+\frac{nm}{m+n}(\bar{x}-\bar{y})(\bar{x}-\bar{y})'}{m+n}$。

对于全参数空间，极大似然估计为 $\mu_1=\bar{x},\mu_2=\bar{y},\Sigma=\frac{V_1+V_2}{m+n}$

似然比检验统计量:

$$\begin{aligned}\lambda=&\frac{L(\mu_0,\mu_0,\Sigma_0)}{L(\mu_1,\mu_2,\Sigma)}\\=&\left(\frac{|V_1+V_2|}{|V_1+V_2+\frac{nm}{m+n}(\bar{x}-\bar{y})(\bar{x}-\bar{y})'|}\right)^{(m+n)/2}\end{aligned}$$

而

$$\begin{aligned}&|V_1+V_2+\frac{nm}{m+n}(\bar{x}-\bar{y})(\bar{x}-\bar{y})'|\\=&|V_1+V_2||I_p+\frac{nm}{m+n}(V_1+V_2)^{-1/2}(\bar{x}-\bar{y})(\bar{x}-\bar{y})'(V_1+V_2)^{-1/2}|\\=&|V_1+V_2|(1+\frac{nm}{m+n}(\bar{x}-\bar{y})'(V_1+V_2)^{-1}(\bar{x}-\bar{y}))\end{aligned}$$

于是:

$$\lambda = \frac{1}{(1 + \frac{nm}{m+n}(\bar{x} - \bar{y})'(V_1 + V_2)^{-1}(\bar{x} - \bar{y}))^{(m+n)/2}}$$

由于 $\bar{x}, \bar{y}, V_1, V_2$ 相互独立，于是:

$$t^2 = \frac{mn(m+n-2}{m+n}(\bar{x} - \bar{y})'(V_1 + V_2)^{-1}(\bar{x} - \bar{y})) \sim T^2(p, m+n-2)$$

□

Exercise 4

11. 某项研究确定运动或膳食补充是否会减缓妇女的骨质流失，研究人员通过光子吸收法源量了实验前和实验 1 年后骨髓中的矿物质含量，表 6.8 是时参与该项目实验前 25 个个体和参与该项目实验 1 年后 24 个个体骨髓中的矿物质含量数据，记录了 3 个骨酪上主力侧和非主力侧上矿物质含量，其中 X_1 表示主力侧的桡骨、X_2 表示桡骨、X_3 表示主力侧的念骨、X_4 表示脑骨、X_5 表示主力侧的尺骨、X_6 表示尺骨中矿物质含量。假设 $X = (X_1, \ldots, X_6)' \sim N_p(\mu, \Sigma)$。表格略。

1. 分别制作实验前数据和实验 1 年后数据的矩阵散点图;
2. 给定显著性水平 $\alpha = 0.05$，检验经过实验后骨图中的矿物质含量是否有流失?
3. 构造均值差 90% 的同时置信区间和 Bonferroni 同时置信区间;
4. 给定显著性水平 $\alpha = 0.05$，分别对实验前和实验后的数据进行独立性检验，首先对随机向量 X 和协方差矩阵 Σ 进行如下分解:

$$X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \\ X^{(3)} \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \\ \mu^{(3)} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{pmatrix},$$

其中 $X^{(1)}=(X_1,X_3)',X^{(2)}=(X_2,X_4)',X^{(3)}=(X_5,X_6)'$。考虑如下的假设检验问题:

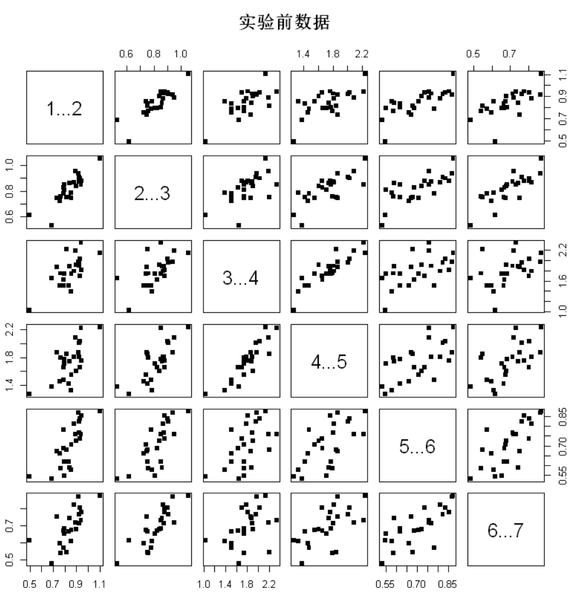
$$H_0:\Sigma_{12}=0,\Sigma_{13}=0,\Sigma_{23}=0,\quad H_1:\Sigma_{12},\Sigma_{13},\Sigma_{23}\text{不全为 }0.$$

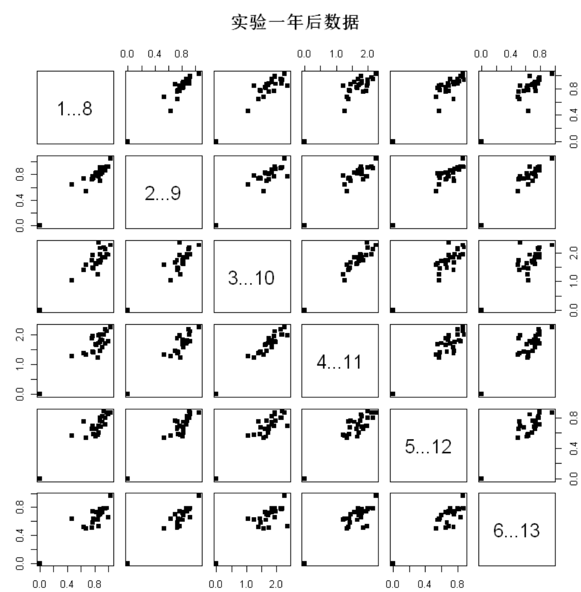
证明. (1)

```
library(readxl)
data <- read_excel("tabel6_8.xlsx")
data_before <- data[, c(2, 3, 4, 5, 6, 7)]
data_after <- data[, c(8, 9, 10, 11, 12, 13)]

# 绘制矩形散点图
pairs(data_before,
    pch = 15, col = "black",
    main = "实验前数据"
)
pairs(data_after,
    pch = 15, col = "black",
    main = "实验一年后数据"
)
```

散点图如下:





(2) 构造假设检验问题:

$$H_0: \mu_1 = \mu_2 \qquad H_1: \mu_1 > \mu_2$$

采用检验统计量:

$$t^2 = \frac{mn(m+n-2)}{m+n}(\bar{x}-\bar{y})'(V_1+V_2)^{-1}(\bar{x}-\bar{y})) \sim T^2(p, m+n-2)$$

为 P 值计算方便，根据 Hotelling T^2 分布的性质，可将检验统计量改为:

$$f^2 = \frac{m+n-p-1}{(n+m-2)p}t^2 = \frac{(m+n-p-1)nm}{p(m+n)}(\bar{x}-\bar{y})) \sim F_{p,n+m-p-1}$$

此处 $m = 24, n = 25, p = 6$，$\bar{x}, \bar{y}$ 分别表示实验前后的样本均值，V_1, V_2 表示实验前后的样本离差阵。

```
# (2)
library(DescTools)
results = HotellingsT2Test(data_before, data_after, test = "f")
print(results)
#单侧检验
T2_statistic <- results$statistic
p_value <- 1 - pf(T2_statistic * (25 + 24 - 6 - 1) /
                       ((25 + 24 - 2) * 6),
                    df1 = 6,
                    df2 = 24 + 25 - 6 - 1)

cat("Pvalue=", p_value)
```

输出单侧检验 P 值为 0.9999827，显著大于 0.05，于是接受原假设，认为矿物质没有流失。

(3) Σ 的最佳估计 $S_0 = \frac{V_1+V_2}{m+n-2} = \frac{(n-1)S_1}{m+n-2} + \frac{(m-1)S_2}{m+n-2}$ 其中 S_i 表示实验前后的样本协方差矩阵。

令 $c = \frac{(m+n-2)p}{n+m-p-1}F_{p,m+n-p-1}(\alpha)$ ，其中 $F(\alpha)$ 表示 F 分布的上 α 分位数，$\alpha = 0.1$。

则 $\mu_{1k} - \mu 2k$ 的 90% 置信区间为:

$$(x_k - y_k) \pm c\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right) + S_{0,kk}} \quad k = 1, 2\cdots, 6$$

各个分量差的 Bonferroni 同时置信区间为:

$$(x_k - y_k) \pm t_{n+m-2}(\frac{\alpha}{2p})\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right) + S_{0,kk}} \quad k = 1, 2\cdots, 6$$

```r
#(3)
m <- 24
n <- 25
p <- 6

#实验前后的样本均值
x <- colMeans(data_before)
y <- colMeans(data_after)
#实验前后样本协方差
S1 <- cov(data_before)
S2 <- cov(data_after)
#Sigma 的最优估计
S0 <- (n - 1) * S1 / (m + n -2) + (m - 1) * S2 / (m + n - 2)

f <- qf(0.9, p, n + m - p - 1)
c <- ((m + n - 2) * p / (m + n - p - 1)) * f
t <- qt(1 - (0.1 / (2 * p)), n + m - 2)

# 同时置信区间
for (k in 1:6){
    upper <- x[k] - y[k] + c * sqrt((1 / n + 1 / m) * S0[k, k])
    lower <- x[k] - y[k] - c * sqrt((1 / n + 1 / m) * S0[k, k])
    cat("k=", k, "lower:", lower, "upper:", upper, "\n")
}
# Bonfferrni同时置信区间
for (k in 1:6){
    upper <- x[k] - y[k] + t * sqrt((1 / n + 1 / m) * S0[k, k])
    lower <- x[k] - y[k] - t * sqrt((1 / n + 1 / m) * S0[k, k])
    cat("k=", k, "lower:", lower, "upper:", upper, "\n")
}
```

输出结果为:

(4) 无偏纠正后的似然比统计量为:

表 1: 置信区间

Variable	Lower	Upper
X1	-0.4365019	0.4421853
X2	-0.3882227	0.4045294
X3	-1.105136	: 1.134329
X4	-1.002386	1.038233
X5	-0.4038639	0.3873305
X6	-0.3887357	0.4029157

(a) 实验前

Variable	Lower	Upper
X1	-0.08179797	0.0874813
X2	-0.06820859	0.08451526
X3	-0.2011201	0.2303134
X4	-0.1786395	0.2144861
X5	-0.08447855	0.06794522
X6	-0.0691659	0.0833459

(b) 实验一年后

$$\lambda = \left(\frac{|V|}{\prod_{i=1}^{3} |V_{ii}|} \right)^{(n-1)/2}$$

自由度为: $df = \frac{p^2 - \Sigma_{i=1}^{3} q_i^2}{2}$ 其中 q_k 表示各个块的变量维数，此处均为 2。
为提高精度，调整参数为:

$$\rho = 1 - \frac{\alpha}{n}, \quad \alpha = \frac{2\left(p^3 - \sum_{k=1}^{m} q_k^3\right) + 9\left(p^2 - \sum_{k=1}^{m} q_k^2\right)}{6\left(p^2 - \sum_{k=1}^{m} q_k^2\right)}.$$

渐进 P 值为: $P = \mathbf{P}(\chi^2_{df} \geq -2\rho \ln \lambda)$

```
#(4)
#数据划分
#实验前
n <- 25
data_cur <- data_before
X1 <- data_cur[, c(1, 2)]
X2 <- data_cur[, c(3, 4)]
X3 <- data_cur[, c(5, 6)]
#计算自由度
q <- c(2, 2, 2)
df <- (p * p - (q[1]^2 + q[2]^2 + q[3]^2)) / 2
#调整参数
alpha <- (2 * (p ^3 - (q[1]^3 + q[2]^3 + q[3]^3)) +
          9 * (p ^2 - (q[1]^2 + q[2]^2 + q[3]^2))) /
```

```
          (6 * (p ^2 - (q[1]^2 + q[2]^2 + q[3]^2)))
rho <- 1 - alpha / n
#似然比统计量
V <- (n - 1) * cov(data_cur)
V11 <- (n - 1) * cov(X1)
V22 <- (n - 1) * cov(X2)
V33 <- (n - 1) * cov(X3)

lambda <- (det(V) / (det(V11) * det(V22) * det(V33)))^((n - 1) / 2)
#P值
Pvalue <- 1 - pchisq(-2 * rho * log(lambda), df = df)
cat("Pvalue_before=", Pvalue, "\n")

#实验后
n <- 24
data_cur <- data_after
X1 <- data_cur[, c(1, 2)]
X2 <- data_cur[, c(3, 4)]
X3 <- data_cur[, c(5, 6)]
#计算自由度
q <- c(2, 2, 2)
df <- (p * p - (q[1]^2 + q[2]^2 + q[3]^2)) / 2
#调整参数
alpha <- (2 * (p ^3 - (q[1]^3 + q[2]^3 + q[3]^3)) +
         9 * (p ^2 - (q[1]^2 + q[2]^2 + q[3]^2))) /
          (6 * (p ^2 - (q[1]^2 + q[2]^2 + q[3]^2)))
rho <- 1 - alpha / n
#似然比统计量
V <- (n - 1) * cov(data_cur)
V11 <- (n - 1) * cov(X1)
V22 <- (n - 1) * cov(X2)
V33 <- (n - 1) * cov(X3)

lambda <- (det(V) / (det(V11) * det(V22) * det(V33)))^((n - 1) / 2)
#P值
Pvalue <- 1 - pchisq(-2 * rho * log(lambda), df = df)
```

```
cat("Pvalue_after=", Pvalue, "\n")
```

输出实验前 $P = 1.191535e - 07$ ，实验后 $P = 7.602601e - 07$ 显著小于 0.05，故拒绝原假设，认为各组变量不独立。

□

Exercise 5

自己设计一个模拟例子，编写程序，对球形检验问题进行研究

证明. 由两个正态总体生成两组正态分布样本

$$X_1 \sim N_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}20 & 5\\5 & 10\end{pmatrix}\right)$$

$$X_2 \sim N_2\left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}2 & 0\\0 & 1\end{pmatrix}\right)$$

对假设检验问题：

$$H_0 : \Sigma = \sigma^2 I_p \quad H_1 : \Sigma \neq \sigma^2 I_p$$

进行检验。

采用无偏修正后的似然比检验统计量

$$\lambda = \frac{|V|^{(n-1)/2}}{(tr(V)/p)^{(n-1)p/2}}$$

引入调整参数 $\rho = 1 - \frac{2p^2+p+2}{6p(n-1)}$ ，自由度 $df = \frac{p(p+1)}{2} - 1$。

渐进 P 值 $P = \mathbf{P}(\chi^2_{df} \geq -2\rho \ln \lambda)$

```r
#生成多元正态模拟样本
library(MASS)
library(lava)
# 生成100个样本
set.seed(123)  # 设置随机数种子以获得可重复的结果
sample1 <- mvrnorm(n = 100, mu = c(0, 0),
        Sigma = matrix(c(20, 5, 5, 10), nrow = 2))
sample2 <- mvrnorm(n = 100, mu = c(1, 0),
        Sigma = matrix(c(4, 0, 0, 4), nrow = 2))

#对于第一组样本
n <- 100
p <- 2
V1 <- (n - 1) * cov(sample1)
lambda1 <- (det(V1))^((n - 1) / 2) / (tr(V1) / p)^((n - 1) * p / 2)
rho <- 1 - (2 * p^2 + p + 2) / (6 * p * (n - 1)) #调整参数
df <- p * (p + 1)  / 2 - 1 #自由度
Pvalue <- 1 - pchisq(-2 * rho * log(lambda1), df = df)
cat("Pvalue_1=", Pvalue, "\n")

#对于第二组样本
n <- 100
p <- 2
V2 <- (n - 1) * cov(sample2)
lambda2 <- (det(V2))^((n - 1) / 2) / (tr(V2) / p)^((n - 1) * p / 2)
rho <- 1 - (2 * p^2 + p + 2) / (6 * p * (n - 1)) #调整参数
df <- p * (p + 1)  / 2 - 1 #自由度
Pvalue <- 1 - pchisq(-2 * rho * log(lambda2), df = df)
cat("Pvalue_2=", Pvalue, "\n")
```

对于第一个总体 $P = 5.015499e - 05$ 拒绝原假设；

对于第二个总体 $P = 0.612357$ 接受原假设。 □