Mean Vector and Covariance Matrix Analysis

Josh Balingit

2023-05-09

1) a)

Given that Σ is unknown and n_1 does not equal n_2, we should use T^2

$$T^2 = (\bar{x}_1 - \bar{x}_2)' \left(\frac{S_p}{n_1} + \frac{S_p}{n_1} \right)^{-1} (\bar{x}_1 - \bar{x}_2) \text{ where } S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

1) b)

```
y11 = c(189,192,217,221,171,192,213,192,170,201,195,205,180,192,200,192,200,181,192)
y12 = c(245,260,276,299,239,262,278,255,244,276,242,263,252,283,294,277,287,255,287)
y13 = c(137,132,141,142,128,147,136,128,128,146,128,147,121,138,138,150,136,146,141)
y14 = c(163,217,192,213,158,173,201,185,192,186,192,192,167,183,188,177,173,183,198)
df1 = data.frame(y11,y12,y13,y14)
x_bar_1 = colMeans(df1)
S_1 = cov(df1)
n_1 = nrow(df1)
y21 = c(181,158,184,171,181,181,177,198,180,177,176,192,176,169,164,181,192,181,175,197)
y22 = c(305,237,300,273,297,308,301,308,286,299,317,312,285,287,265,308,276,278,271,303)
y23 = c(184,133,166,162,163,160,166,141,146,171,166,166,141,162,147,157,154,149,140,170)
y24 = c(209,188,231,213,224,223,221,197,214,192,213,209,200,214,192,204,209,235,192,205)
df2 = data.frame(y21,y22,y23,y24)
x_bar_2 = colMeans(df2)
S_2 = cov(df2)
n_2 = nrow(df2)
S_p = ((n_1-1)*S_1 + (n_2-1)*S_2)/(n_1+n_2-2)
T_sq = crossprod(x_bar_1-x_bar_2,solve((S_p/n_1)+(S_p/n_2)))%*%(x_bar_1-x_bar_2)
p = 4
T_sq_crit = (p*(n_1+n_2-2)/(n_1+n_2-p-1))*qf(1-0.05,p,(n_1+n_2-p-1))
```

$T^2 = 133.4873031$

1) c)

Since $T^2 = 133.4873031$ is greater than $T^2_{0.95,4,37} = 11.5348328$, we reject H_0 and conclude that group 1 and 2 differ.

2) a)

We should use the modified likelihood ratio statistic u

$$u = (n-1)\left(\sum_{i=1}^{p}(\lambda_i - ln\lambda_i) - p\right) \text{ where } \lambda_1, ...\lambda_p \text{ are eigenvalues of } S\Sigma_0^{-1}$$

2) b)

```
height = c(69,74,68,70,72,67,66,70,76,68,
           72,79,74,67,66,71,74,75,75,76)
weight = c(153,175,155,135,172,150,115,137,200,130,
           140,265,185,112,140,150,165,185,210,220)
df = data.frame(height,weight)
S = cov(df)
sigma_0 = matrix(c(20,100,
                   100,1000),byrow = T, nrow = 2)
A = S%*%solve(sigma_0)
eigen_A = eigen(A)
eigenval = eigen_A$values
u = (20-1)*(sum(eigenval-log(eigenval))-2)
u_crit = qchisq(1-0.05,0.5*2*(2+1))
```

$u = 11.0937431$

2) c)

Since $u = 11.0937431$ is greater than $\chi^2_{0.95,3} = 7.8147279$, we reject H_0 and conclude that the population's variances for height and weight does not equal hypothesized measures for height and weight.