**Exercise 7.6** In the U.S. companies data set, test the equality of means between the

energy and manufacturing sectors, taking the full vector of observations x1 to *x*6.

Derive the simultaneous confidence intervals for the differences.

The data set consists of measurements for 79 U.S. companies. The abbreviations in

this section are as follows:

*x*1: A Assets (USD),

*x*2: S Sales (USD),

*x*3: MV Market Value (USD),

*x*4: P Profits (USD),

*x*5: CF Cash Flow (USD), and

*x*6: E Employees.

Sol.

根據題目假設檢定 ;

首先假設和服從多變量常態，且。

*變數間兩兩互相獨立。*

設

經由r計算得：

,

因此，由公式 ，得

**拒絕域：**

因為，不拒絕表示沒有足夠證據證明 。

但若以90%的信賴水準評估，則結果為，拒絕，表示有足夠證據證明之間存在不同。

**Confidence Region:**

根據信賴區間估計公式得六種變數在energy與manufacturing兩種產業的公司間皆無足夠證據說明兩者存在差異性，因此不拒絕，可能不獨立。

透過mvn套件也可以檢驗出每個變數個別p-value與多變量的p-value皆小於0.05，亦即該資料不符合多變量及單變量的常態分佈。

$multivariateNormality

Test Statistic p value Result

1 Mardia Skewness 242.233307803641 5.21767710767055e-25 NO

2 Mardia Kurtosis 9.74673862485956 0 NO

3 MVN <NA> <NA> NO

$univariateNormality

Test Variable Statistic p value Normality

1 Shapiro-Wilk V2 0.7720 1e-04 NO

2 Shapiro-Wilk V3 0.7016 <0.001 NO

3 Shapiro-Wilk V4 0.8191 5e-04 NO

4 Shapiro-Wilk V5 0.7827 1e-04 NO

5 Shapiro-Wilk V6 0.8260 6e-04 NO

6 Shapiro-Wilk V7 0.5853 <0.001 NO

**R Code:**

*# clear variables and close windows*

*rm(list = ls(all = TRUE))*

*graphics.off()*

*# Load data*

*x = read.table("C:/Users/user/Desktop/多變量11101/MVA-ToDo-master/QID-1659-MVAsimcidif/uscomp2.dat")*

*y = data.frame(x)*

*# Create subsets for Energy and Manufacturing*

*yE = subset(y, y$V8 == "Energy")*

*yM = subset(y, y$V8 == "Manufacturing")*

*# Calculate means of groups*

*exE = cbind(apply(yE[, 2:7], 2, mean)) # 1:by row;2:by column*

*exM = cbind(apply(yM[, 2:7], 2, mean))*

*# https://kemushi54.github.io/R-basic/apply\_family.html*

*# Estimating variance of the groups observations within the groups and overall*

*nE = length(yE[, 1])*

*nM = length(yM[, 1])*

*n = nE + nM*

*# number of groups*

*p = length(exE)*

*sE = ((nE - 1)/nE) \* cov(yE[, 2:7]) # S1*

*sM = ((nM - 1)/nM) \* cov(yM[, 2:7]) # S2*

*s = (nE \* sE + nM \* sM)/(nE + nM)*

*sinv = solve(s)*

*k = nE \* nM \* (n - p - 1)/(p \* (n^2))*

*# Computing the test statistic*

*(f = k \* t(exE - exM) %\*% sinv %\*% (exE - exM))*

*# Computing the critical value*

*(critvalue = qf(1 - 0.05, 6, 18))*

*# ALPHA=0.05 95%CI 右尾 IF f>critvalue 拒絕HO*

*(critvalue = qf(1 - 0.1, 6, 18))*

*# Computes the simultaneous confidence intervals*

*deltau = (exE - exM) + sqrt(qf(1 - 0.05, p, n - p - 1) \* (1/k) \* diag(s))*

*deltal = (exE - exM) - sqrt(qf(1 - 0.05, p, n - p - 1) \* (1/k) \* diag(s))*

*(confit = cbind(deltal, deltau))*

*# extrapoints*

*library(MVN)*

*result\_uni <- mvn(yEM, mvnTest = "mardia", univariateTest = "SW", showOutliers = TRUE)*

*result\_multi <- mvn(yEM, mvnTest = "mardia", multivariateOutlierMethod = "quan", showOutliers = TRUE)*