hw2

LianZuo

2023-10-04

# Stat 577 Homework 2

# Question 1

set.seed(1234)  
z1 <- rnorm(50, 0,1)  
z2 <- rnorm(50, 0,1)  
mu\_x <- 10  
mu\_y <- 20  
sigma\_x <- 1  
sigma\_y <- 2  
rho <- 0.8  
X <- sigma\_x \* z1 + mu\_x  
Y <- sigma\_y \* (rho \* z1 + sqrt(1 - rho^2) \* z2) + mu\_y  
D <- cbind(X, Y)  
mean\_vec<- colMeans(D)  
mean\_vec

## X Y   
## 9.546947 19.442551

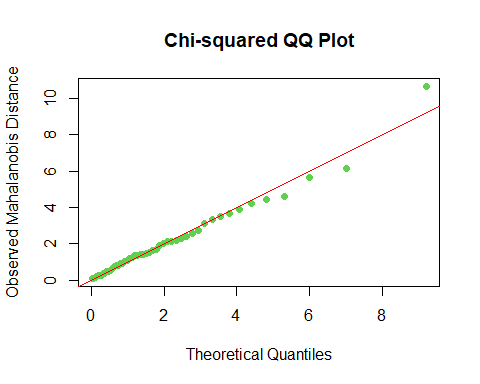
cov\_matrix <- cov(D)  
cov\_matrix

## X Y  
## X 0.783302 1.184362  
## Y 1.184362 3.333838

# Calculate Mahalanobis distance  
mahal\_D <- mahalanobis(D, center = mean\_vec, cov = cov\_matrix)  
mahal\_D

## [1] 4.46180633 1.09769812 4.22148135 6.13390377 1.04602018 1.40197320  
## [7] 2.11684143 0.80060634 2.00019522 1.89299034 0.24853128 5.61906646  
## [13] 0.16993800 0.89695843 2.54905044 2.72986893 1.53896021 1.59994494  
## [19] 1.44814029 10.66227834 0.44759412 0.33525574 0.23831919 1.37260582  
## [25] 3.48941431 1.38923546 3.32390022 1.17989427 0.26605505 0.52359089  
## [31] 3.12691700 0.09101159 2.27026216 0.09665552 2.12961079 0.89049892  
## [37] 3.88276074 1.35081254 0.12742870 1.66339155 4.62371030 0.48842832  
## [43] 2.41102342 0.75183392 0.79898930 0.36884816 2.17770666 1.24650904  
## [49] 0.64649157 3.65499115

# Degrees of freedom for Chi-squared distribution  
df <- 2 # Since we have 2 variables  
  
# Calculate the expected quantiles  
theo.quan<- qchisq(ppoints(50), df)  
  
# Create the QQ plot  
qqplot(theo.quan, mahal\_D,  
 xlab = "Theoretical Quantiles",  
 ylab = "Observed Mahalanobis Distance",  
 main = "Chi-squared QQ Plot",  
 pch=16,col=3)  
  
# Draw a reference line  
abline(0, 1, col = "red")



# Show the plot  
#Install. packages("mvnormtest)  
library(mvnormtest)  
mshapiro.test(t(D))

##   
## Shapiro-Wilk normality test  
##   
## data: Z  
## W = 0.95719, p-value = 0.06778

#The multivariate Shapiro-Wilk p-value is not significant.Therefore, we accept the null hypothesis and conclude that the data is coming from a bivariate normal distribution.

# Question 2

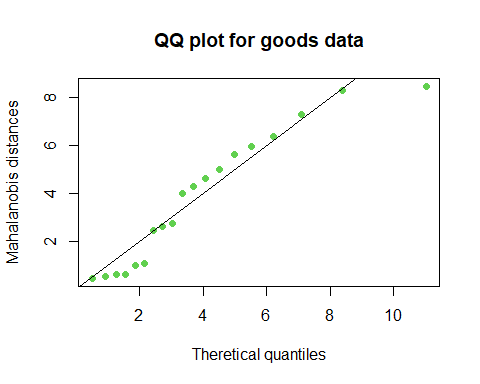
goods <- read.table("C:/Users/lian/Desktop/goods.txt",header=F)  
goods <- goods[,-1]  
colnames(goods) <- c("x1","x2","x3","x4")  
smean <- colMeans(goods)  
scov <- cov(goods)  
print (list(smean,scov))

## [[1]]  
## x1 x2 x3 x4   
## 70.55263 51.52632 14.39474 1.00000   
##   
## [[2]]  
## x1 x2 x3 x4  
## x1 1021.302632 -75.057018 101.8808480 1.1888889  
## x2 -75.057018 70.818713 -2.5248538 -1.0444444  
## x3 101.880848 -2.524854 27.5160819 0.9083333  
## x4 1.188889 -1.044444 0.9083333 0.1922222

n <- nrow(goods)  
mhd.goods <- mahalanobis(goods,smean,scov)  
mhd.goods

## [1] 2.7575021 0.4432065 0.6059117 1.0522412 4.9932445 5.9728333 4.2853229  
## [8] 2.4623656 7.2895720 0.5379913 2.6062002 8.3091260 4.6109191 0.9813463  
## [15] 0.5978303 5.6317653 8.4727766 4.0087419 6.3811032

theo.quan2 <- qchisq(ppoints(n),4)# df=number of variables = 4  
qqplot(theo.quan2,mhd.goods,xlab = "Theretical quantiles",  
 ylab="Mahalanobis distances",pch=16,col=3,main="QQ plot for goods data")  
abline(0,1)# drawing a reference line



#install.packages("mvnormtest")  
library(mvnormtest)  
mshapiro.test(t(goods))

##   
## Shapiro-Wilk normality test  
##   
## data: Z  
## W = 0.89405, p-value = 0.03803

# Formal test for checking normality:Shapiro-Wilk test  
attach(goods)  
shapiro.test(x1)

##   
## Shapiro-Wilk normality test  
##   
## data: x1  
## W = 0.94327, p-value = 0.3017

shapiro.test(x2)

##   
## Shapiro-Wilk normality test  
##   
## data: x2  
## W = 0.94188, p-value = 0.285

shapiro.test(x3)

##   
## Shapiro-Wilk normality test  
##   
## data: x3  
## W = 0.90661, p-value = 0.06417

shapiro.test(x4)

##   
## Shapiro-Wilk normality test  
##   
## data: x4  
## W = 0.94433, p-value = 0.3151