MVA_HW2

陈子睿 15220212202842

4/5/24

1 The generalized variance and the total variance

By the definition of generalized variance and total variance we have

$$|\Sigma_1| = 1, |\Sigma_2| = 4, tr(\Sigma_1) = 20, tr(\Sigma_2) = 15$$

We know that a larger variance indicates that the data points are more dispersed, whereas a smaller variance suggests that the data points are more concentrated. The generalized variance of y_1 is larger than of y_2. However, the total variance of y_1 is smaller than of y_2. One possible explanation is that there's a significant positive covariance between the original data variables, indicating positive correlations, the Generalized Variance calculated via the covariance matrix might be large.

2 Find a value for ρ

We define

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \end{pmatrix}.$$

Then $Ay \sim N_p(0, A\Sigma A')$ with covariance matrix

$$A\Sigma A' = \begin{pmatrix} 3+4\rho & -1-2\rho \\ -1-2\rho & 3 \end{pmatrix}.$$

Then we can see that $Y_1 + Y_2 + Y_3$ and $Y_1 - Y_2 - Y_3$ are independent if and only if $-1 - 2\rho = 0$, i.e. $\rho = -1/2$.

3 Transformation of Mutivariate Normal Vector

We're given that $y \sim N_3(\mu, \Sigma)$, then

(a) We define $a = (2, -1, 3)'$, then $Z = a'y \sim N(a'\mu, a'\Sigma a) = N(16, 21)$.

(b) We define

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \end{pmatrix},$$

then

$$(Z_1, Z_2)' = Ay \sim N_2(A\mu, A\Sigma A') = N_2(\begin{pmatrix} 9 \\ 9 \end{pmatrix}, \begin{pmatrix} 29 & -1 \\ -1 & 9 \end{pmatrix}).$$

(c) It's trivial that $Y_2 \sim N(2, 13)$.

(d) The joint distribution of (Y_1, Y_3) is $(Y_1, Y_3) \sim N_2(\begin{pmatrix} 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 6 & -2 \\ -2 & 4 \end{pmatrix})$.

(e) We define

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix},$$

then,

$$\begin{pmatrix} Y_1 \\ Y_3 \\ \frac{1}{2}(Y_1 + Y_3) \end{pmatrix} = Ay \sim N_3(Ay, A\Sigma A') = N_3(\begin{pmatrix} 3 \\ 4 \\ \frac{7}{2} \end{pmatrix}, \begin{pmatrix} 6 & -2 & 2 \\ -2 & 4 & 1 \\ 2 & 1 & \frac{3}{2} \end{pmatrix}).$$

(f) First we need to obtain the joint distribution for (Y_1, Z_1, Y_2, Z_2). We define

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & -1 & 2 \end{pmatrix}.$$

Then we have

$$\begin{pmatrix} Y_1 \\ Z_1 \\ Y_2 \\ Z_2 \end{pmatrix} = Ay \sim N_4(Ay, A\Sigma A') = N_4(\begin{pmatrix} 3 \\ 9 \\ 2 \\ 9 \end{pmatrix}, \begin{pmatrix} 6 & 5 & 1 & 1 \\ 5 & 29 & 18 & -1 \\ 1 & 18 & 13 & -4 \\ 1 & -1 & -4 & 9 \end{pmatrix}).$$

Then $E\left[\begin{pmatrix} Y_1 \\ Z_1 \end{pmatrix} | \begin{pmatrix} Y_2 \\ Z_2 \end{pmatrix}\right] = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}((Y_2, Z_2)' - \mu_2) = \begin{pmatrix} \frac{13}{101}Y_2 + \frac{17}{101}Z_2 + \frac{124}{101} \\ \frac{158}{101}Y_2 + \frac{59}{101}Z_2 + \frac{62}{101} \end{pmatrix}$,

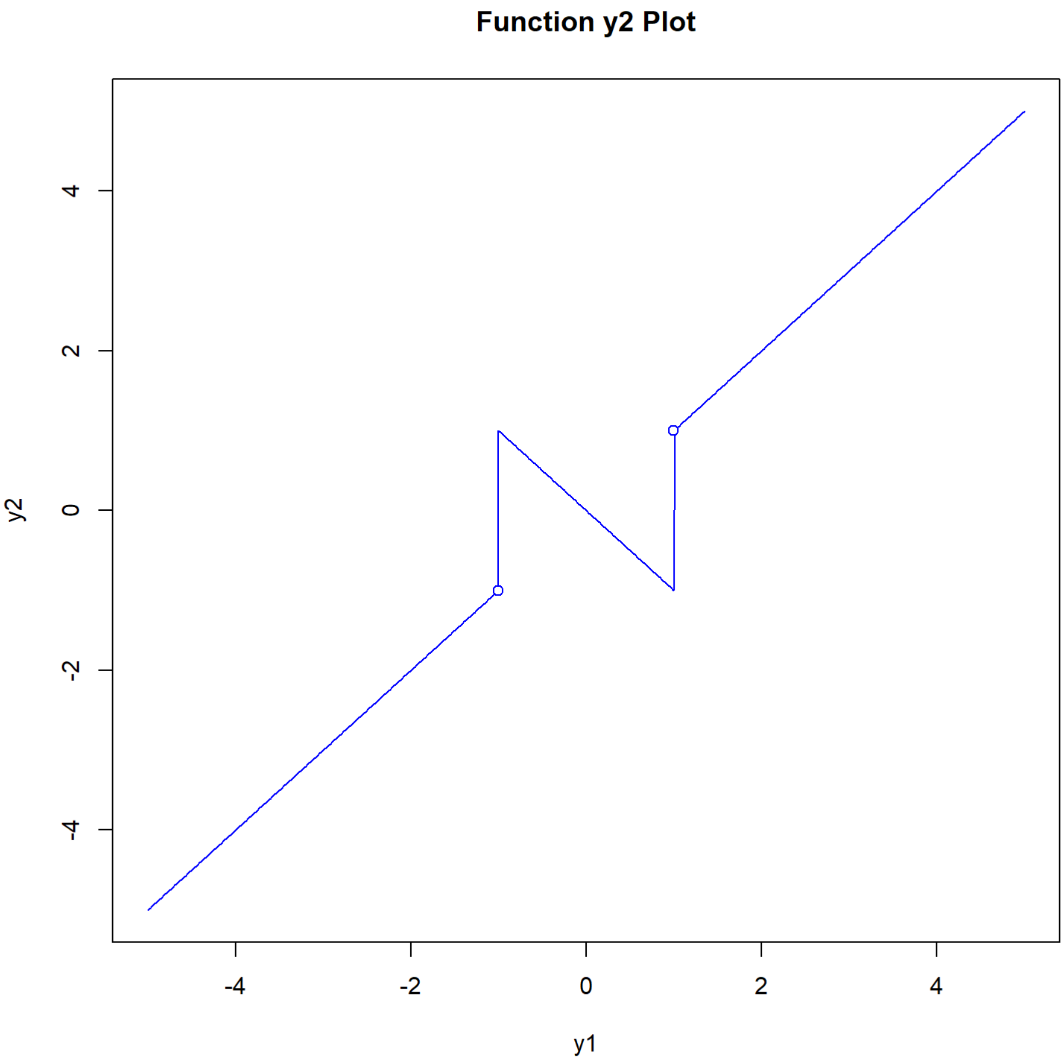
$Var\left[\begin{pmatrix} Y_1 \\ Z_1 \end{pmatrix} | \begin{pmatrix} Y_2 \\ Z_2 \end{pmatrix}\right] = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \begin{pmatrix} \frac{576}{101} & \frac{288}{101} \\ \frac{288}{101} & \frac{144}{101} \end{pmatrix}$.

Hence

$$\begin{pmatrix} Y_1 \\ Z_1 \end{pmatrix} | \begin{pmatrix} Y_2 \\ Z_2 \end{pmatrix} \sim N_2(\begin{pmatrix} \frac{13}{101}Y_2 + \frac{17}{101}Z_2 + \frac{124}{101} \\ \frac{158}{101}Y_2 + \frac{59}{101}Z_2 + \frac{62}{101} \end{pmatrix}, \begin{pmatrix} \frac{576}{101} & \frac{288}{101} \\ \frac{288}{101} & \frac{144}{101} \end{pmatrix}).$$

4 The distribution for Y_1 and Y_2

(a) First we can plot Y_1 and Y_2, we have



Then we have

$$P(Y_2 \leq y) = \begin{cases} P(Y_1 \leq y) & y \leq -1 \\ P(Y_2 \leq -1) + P(-1 < Y_2 \leq y) = P(Y_1 \leq -1) + P(-y \leq Y_1 < 1) = P(Y_1 \leq y) & -1 < y < 1 \\ = P(Y_2 \leq -1) + P(-1 < Y_2 \leq 1) + P(1 < Y_2 \leq y) = P(Y_1 \leq y) & y > 1. \end{cases}$$

Hence, Y_1 and Y_2 are identically distributed, i.e. $Y_2 \sim N(0,1)$.

(b) By the property of mutivariate normal data, if $(Y_1, Y_2)'$ are bivariate normal, then the sum $Y_1 + Y_2$ must be normal. However,

$$P(Y_1 + Y_2 = 0) = P(-1 \leq Y_1 \leq 1) \neq 0.$$

Contradiction E.q..

5 Property after partition

5. (a) Check

$$(y-\mu)'\Sigma^{-1}(y-\mu)=[y_1-\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(y_2-\mu_2)]'$$
$$\times\left(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)^{-1}[y_1-\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(y_2-\mu_2)]+(y_2-\mu_2)'\Sigma_{22}^{-1}(y_2-\mu_2)$$

Proof. We have

$$\Sigma^{-1}=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}^{-1}=\begin{pmatrix}(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} & -(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{12}\Sigma_{22}^{-1}\\ -\Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} & \Sigma_{22}^{-1}+\Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{12}\Sigma_{22}^{-1}\end{pmatrix}.$$

Denote $V=(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}$, then

$$(y-\mu)'\Sigma^{-1}(y-\mu)=(y_1'-\mu_1',y_2'-\mu_2')\begin{pmatrix}V & -V\Sigma_{12}\Sigma_{22}^{-1}\\ -\Sigma_{22}^{-1}\Sigma_{21}V & \Sigma_{22}^{-1}+\Sigma_{22}^{-1}\Sigma_{21}V\Sigma_{12}\Sigma_{22}^{-1}\end{pmatrix}\times$$
$$(y_1-\mu_1,y_2-\mu_2)'$$

We try to decompose Σ^{-1} as

$$\Sigma^{-1}=\begin{pmatrix}V & -V\Sigma_{12}\Sigma_{22}^{-1}\\ -\Sigma_{22}^{-1}\Sigma_{21} & \Sigma_{22}^{-1}\Sigma_{21}V\Sigma_{12}\Sigma_{22}^{-1}\end{pmatrix}+\begin{pmatrix}0 & 0\\ 0 & \Sigma_{22}^{-1}\end{pmatrix}.$$

i.e. It's trivial to check

$$(y_1'-\mu_1',y_2'-\mu_2')\begin{pmatrix}0 & 0\\ 0 & \Sigma_{22}^{-1}\end{pmatrix}(y_1-\mu_1,y_2-\mu_2)'=(y_2'-\mu_2')\Sigma_{22}^{-1}(y_2-\mu_2)'.$$

(b) Since

$$f(y_1|y_2)=\frac{f(y_1,y_2)}{f(y_2)}$$

and

$$f(y)=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\right\},$$
$$f(y_2)=\frac{1}{(2\pi)^{p-r/2}|\Sigma_{22}|^{1/2}}\exp\left\{-\frac{1}{2}(y_2-\mu_2)'\Sigma_{22}^{-1}(y_2-\mu_2)\right\}$$

Thus

$$f(y_1|y_2)=\frac{f(y)}{f(y_2)}=\frac{1}{(2\pi)^{r/2}|\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}|^{1/2}}$$
$$\times\exp\left\{-\frac{1}{2}\left(y_1-\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(y_2-\mu_2)\right)'V\left(y_1-\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(y_2-\mu_2)\right)\right\}$$

6 Prove that LRT leads to Hotelling's T^2 test

Suppose the sample $X_1, \dots, X_n$ is from a multivariate normal distribution, where

$$X_i \sim N_P(\mu, \Sigma) \quad i = 1, 2, \dots, n$$

Our goal is to test the null hypothesis $H_0 : \mu = \mu_0$ against the alternative hypothesis $H_1 : \mu \neq \mu_0$. The likelihood function for the multivariate normal distribution is:

$$\mathcal{L}(\mu, \Sigma) = (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} exp\left(-\frac{1}{2}\sum_{i=1}^{n}(X_i - \mu)'\Sigma^{-1}(X_i - \mu)\right)$$

The statistic for the likelihood ratio test is:

$$\Lambda = \frac{sup_{\mu,\Sigma}\mathcal{L}(\mu, \Sigma|H_0)}{sup_{\mu,\Sigma}\mathcal{L}(\mu, \Sigma)}$$

The unrestricted maximum likelihood is:

$$\begin{aligned}
\mathcal{L}(\Theta) &= sup_{\mu,\Sigma}\mathcal{L}(\mu, \Sigma|H_0) \\
&= \mathcal{L}(\hat{\mu}, \hat{\Sigma}) \\
&= (2\pi)^{-\frac{np}{2}} |\hat{\Sigma}|^{-\frac{n}{2}} exp\left(-\frac{1}{2}\sum_{i=1}^{n}(X_i - hat\mu)'\hat{\Sigma}^{-1}(X_i - hat\mu)\right) \\
&= (2\pi)^{-\frac{np}{2}} |\hat{\Sigma}|^{-\frac{n}{2}} exp\left(-\frac{1}{2}\sum_{i=1}^{n} tr[\hat{\Sigma}^{-1}(X_i - \hat{\mu})(X_i - \hat{\mu})']\right) \\
&= (2\pi)^{-\frac{np}{2}} |\hat{\Sigma}|^{-\frac{n}{2}} exp\left(-\frac{1}{2}tr\left[\hat{\Sigma}^{-1}\sum_{i=1}^{n}(X_i - \hat{\mu})(X_i - \hat{\mu})'\right]\right) \\
&= (2\pi)^{-\frac{np}{2}} |\hat{\Sigma}|^{-\frac{n}{2}} exp\left(-\frac{np}{2}\right).
\end{aligned}$$

Notice that since the MLE estimator for Σ is $\hat{\Sigma}_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})(X_i - \hat{\mu})'$, then $\hat{\Sigma}^{-1}\sum_{i=1}^{n}(X_i - \hat{\mu})(X_i - \hat{\mu})' = nI$.

In the same case, the restricted maximum likelihood is:

$$\mathcal{L}(\Theta_0) = sup_{\mu,\Sigma}\mathcal{L}(\mu, \Sigma) = sup_{\Sigma}\mathcal{L}(\mu_0, \Sigma) = \mathcal{L}(\mu_0, \hat{\Sigma}_0) = (2\pi)^{-\frac{np}{2}} |\hat{\Sigma}_0|^{-\frac{n}{2}} exp\left(-\frac{np}{2}\right),$$

where $\hat{\Sigma}_0 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)(X_i - \mu_0)'$.

Therefore, the generalized likelihood ratio is:

$$\begin{aligned}
\lambda &= \frac{sup_{\mu,\Sigma}\mathcal{L}(\mu, \Sigma|H_0)}{sup_{\mu,\Sigma}\mathcal{L}(\mu, \Sigma)} \\
&= \frac{\mathcal{L}(\Theta)}{\mathcal{L}(\Theta_0)} \\
&= \left[\frac{|\hat{\Sigma}_0|}{|\hat{\Sigma}|}\right]^{\frac{n}{2}} \\
&= \exp\left(-\frac{1}{2}n(\bar{X} - \mu_0)^T S^{-1}(\bar{X} - \mu_0)\right)
\end{aligned}$$

We know that the likelihood ratio test statistic $-2\ln\Lambda$ asymptotically follows a chi-squared distribution in large samples. For Hotelling' s T^2 test, its statistic is defined as:

$$T^2 = n(\overline{X} - \mu_0)^T S^{-1}(\overline{X} - \mu_0)$$

It is evident that the $-2\ln\Lambda$ obtained from the likelihood ratio test is directly related to Hotelling' s T^2 statistic. In fact, by appropriate scaling (multiplying by the sample size n), the Hotelling' s T^2 statistic can be transformed into a form that follows an F-distribution, further confirming the equivalence of these two methods under the condition of multivariate normal distribution.

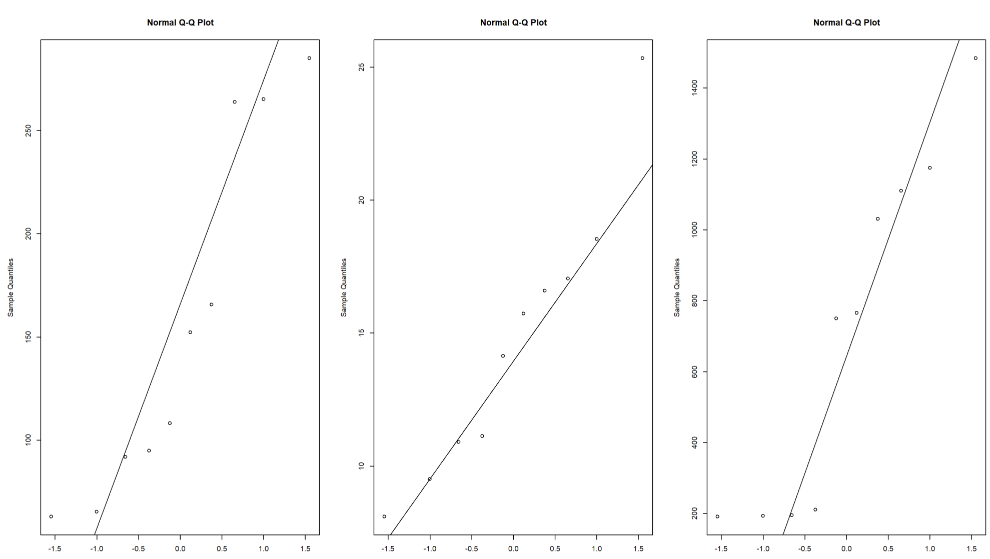
7 R Exercise: Company

First we shall read the data by

```
y <- read.table(paste('C:/Users/Ray Chen/Desktop/MVA/company.txt'), header = TRUE)
```

(a) The Q-Q plots can be plotted by

```
par(mfrow = c(1,3))
qqnorm(y$x1)
qqline(y$x1)
qqnorm(y$x2)
qqline(y$x2)
qqnorm(y$x3)
qqline(y$x3)
```



We can see that most points are closed to the corresponding Q-Q line, however, X_1 and X_3 seemed to have a S-shape, indicated that they actually has a thinner tail than normal. For X_2, it has an outlier.

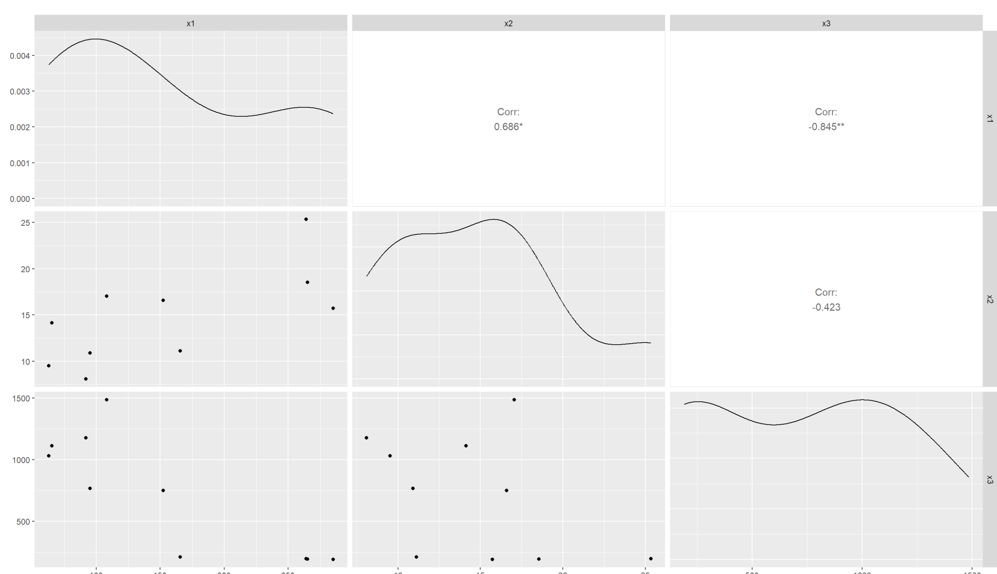
(b) We use Shapiro-Wilks test to check the individual normality,

```
> shapiro.test(y$x1)
##Shapiro-Wilk normality test
data:  y$x1
W = 0.85909, p-value = 0.07444

> shapiro.test(y$x2)
##Shapiro-Wilk normality test
data:  y$x2
W = 0.94221, p-value = 0.5778

> shapiro.test(y$x3)
##Shapiro-Wilk normality test
data:  y$x3
W = 0.86969, p-value = 0.09914
```

By the result of the Shapiro-Wilks tests, the normality of all three variables cannot be rejected at the significance level of 0.05. However, the normality of X_1 and X_3 can be rejected at the significance level of 0.1. And we can see that a single outlier won't effect much to the result of the Shapiro-Wilks test by the result of X_2.

(c) The pairwise scatterplot can be demonstrated by

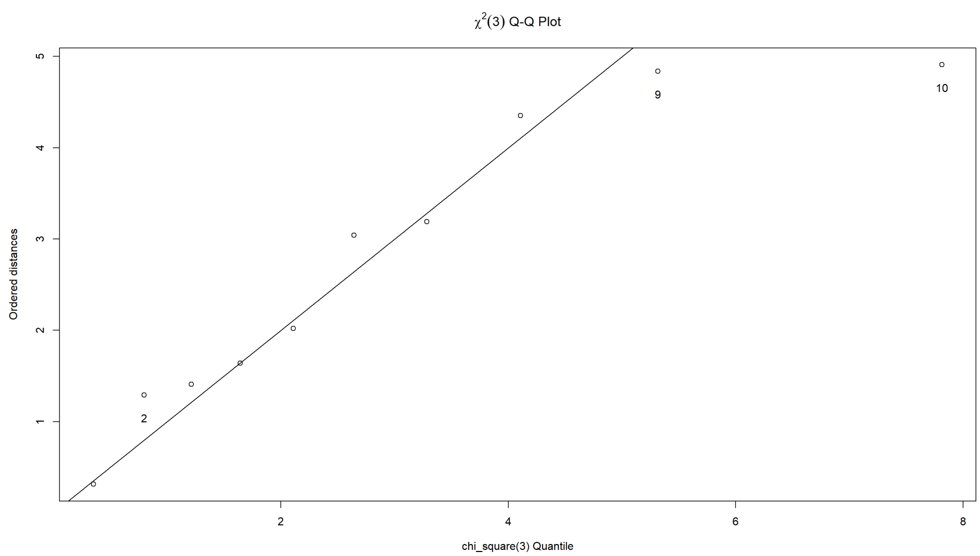
```
library(GGally)
ggpairs(y)
```



And the χ^2 Q-Q plot is given by

```
cM <-colMeans(y)
S <- cov(y)
```

```r
d <- apply(y,1,function(y) t(y-cM) %*% solve(S) %*%  (y-cM ))
plot(qc <- qchisq((1:nrow(y) - 1/2) / nrow(y), df = 3), sd <- sort(d),
     xlab = expression(paste("chi_square(3) " ,"Quantile")), ylab = "Ordered
     ↪  distances")
oups <- which(rank(abs(qc - sd), ties.method = "random") > nrow(y) - 3)
text(qc[oups], sd[oups] - 0.25, oups)
abline(0, 1)
title(expression(paste(chi^2,(3), " Q-Q Plot")))
```



By the pairwise scatterplot and the χ^2 Q-Q plot, there's actually no significant evidence to say claim that $(X_1, X_2, X_3)'$ are not jointly normal.

8 Confidence Region

(a) Recall the tests on μ when Σ is unknown, we use the Hotelling's T^2 test statistic:

$$T^2 = n(\bar{y} - \mu_0)' S^{-1} (\bar{y} - \mu_0)$$

Under $H_0 : \mu = \mu_0$, T^2 follows Hotelling's T^2 distribution $T^2(p, n-1)$. Since the Hotelling's T^2 test statistic has the propert that:

$$\frac{n-p}{p(n-1)} T^2(p, n-1) = F(p, n-p)$$

Therefore, the $1-\alpha$ confidence region can be given by:

$$\left\{ \mu : (\bar{y} - \mu)' S^{-1} (\bar{y} - \mu) \le \frac{p(n-1)}{n(n-p)} F_{1-\alpha}(p, n-p) \right\}.$$

(b) By the procedure mentioned in (a),

```
> rm(list = ls())
> y <- read.table(paste('C:/Users/Ray Chen/Desktop/MVA/sweat.DAT'), header = FALSE)
> y_partial <- y[,1:2]
> n <- nrow(y_partial)
> p <- ncol(y_partial)
> cM <- colMeans(y_partial)
> cM
   V1    V2
 4.64 45.40
> S <- cov(y_partial)
> S_inv <- solve(S)
> S_inv
            V1           V2
V1  0.42055021 -0.021070829
V2 -0.02107083  0.006061007
> RHS_quantile <- p * (n-1) / (n * (n-p)) * qf(0.95,p,n-p)
> RHS_quantile
[1] 0.3752033
```

Therefore, an 95% confidence region for μ is given by

$$\left\{\mu : (4.64-\mu_1, 45.40-\mu_2)' \begin{pmatrix} 0.42055021 & -0.021070829 \\ -0.02107083 & 0.006061007 \end{pmatrix} (4.64-\mu_1, 45.40-\mu_2) \leq 0.3752033 \right\}.$$

(c) First we need to compute the rotated angle,

```
> S_eig <- eigen(S)
> S_eig
eigen() decomposition
$values
[1] 200.295978   2.371812

$vectors
          [,1]       [,2]
[1,] 0.0506399 -0.9987170
[2,] 0.9987170  0.0506399

> phi <- acos(S_eig$vectors[1]) * 180 /pi
> phi
[1] 87.09731
```

Then the spectral decomposition of S is given by

$$S = LDL', \quad D = \begin{pmatrix} 200.295978 & 0 \\ 0 & 2.371812 \end{pmatrix}, L = \begin{pmatrix} cos\varphi & -sin\varphi \\ sin\varphi & cos\varphi \end{pmatrix}, \varphi = 87.09731°.$$

The semi-axes are given by:

```
> semiaxes <- sqrt(S_eig$values * RHS_quantile)
> semiaxes
[1] 8.6690082 0.9433512
```

Therefore, the confidence region of μ is

$$\left\{\mu : (\bar{y} - \mu)' L D^{-1} L' (\bar{y} - \mu) \leq \frac{p(n-1)}{n(n-p)} F_{1-\alpha}(p, n-p)\right\},$$

i.e. which is actually an 87.09731° rotated ellipse centered at $\bar{y} = (4.64, 45.40)'$ whose semi-axes are 8.6690082 and 0.9433512.

(d) To construct the 95% confidence interval for each variable,

```
> t_quantile <- qt(0.975, n -1 )
> L1 <- cM[1] - t_quantile * sqrt(S[1] / n)
> U1 <- cM[1] + t_quantile * sqrt(S[1] / n)
> L2 <- cM[1] - t_quantile * sqrt(S[4] / n)
> U2 <- cM[1] + t_quantile * sqrt(S[4] / n)
> L1
     V1
3.84584
> U1
     V1
5.43416
> L2
      V2
38.78478
> U2
      V2
52.01522
```

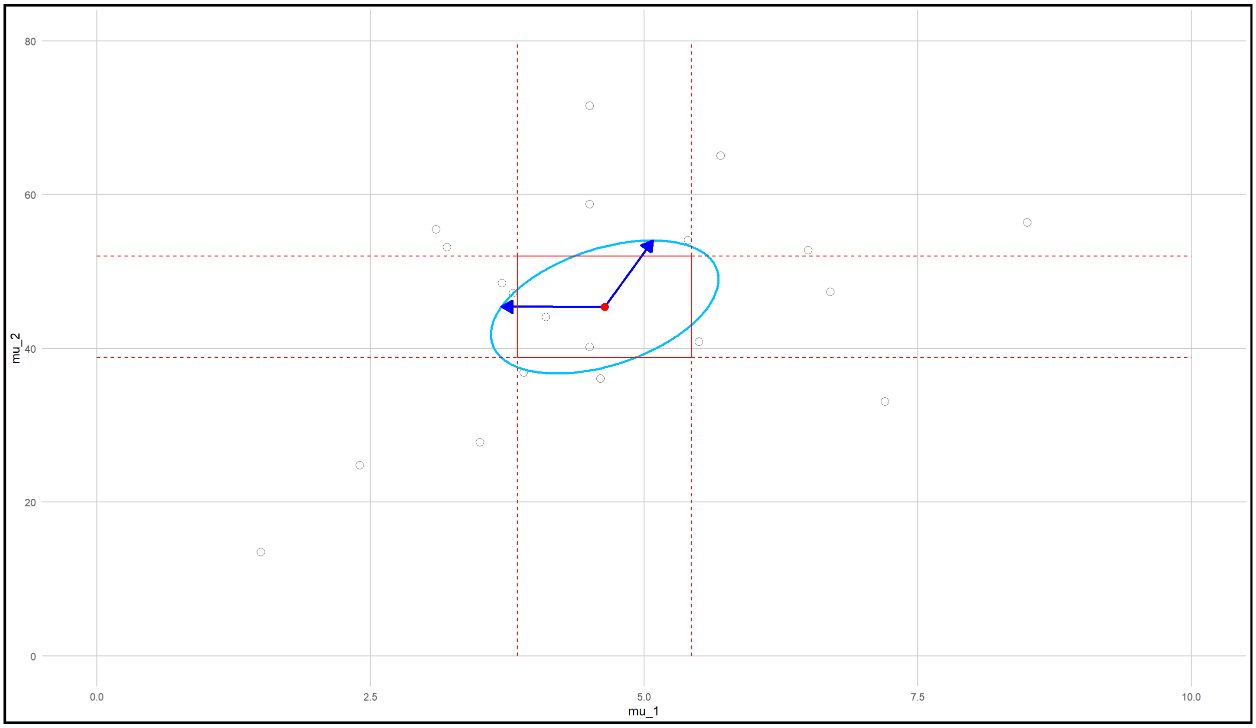
The 95% confidence interval for μ_1 is $[3.84584, 5.43416]$, for μ_2 is $[38.78478, 52.01522]$.

(e) To find such a μ_0 such that the multivariate test reject H_0 but both univariate tests fails to,

```
library(ggplot2)
library(ggforce)
df <- data.frame(y_partial)
p <- ggplot(df, aes(x = V1, y = V2)) +
```

```
geom_point(shape = 1, color = "black", size = 3, alpha = 0.6) +
geom_ellipse(aes(x0 = cm[1], y0 = cm[2], a = semiaxes[1], b = semiaxes[2], angle =
↪  atan(S_eig$vectors[2,1] / S_eig$vectors[1,1])),
             color = "deepskyblue", alpha = 0.3, linetype = "solid", size = 1) +
geom_segment(aes(x = cm[1], y = cm[2], xend = cm[1] + semiaxes[1] *
↪  S_eig$vectors[,1][1], yend = cm[2] + semiaxes[1] * S_eig$vectors[,1][2]),
             arrow = arrow(type = "closed", length = unit(0.15, "inches")), color
             ↪  = "blue", size = 1) +
geom_segment(aes(x = cm[1], y = cm[2], xend = cm[1] + semiaxes[2] *
↪  S_eig$vectors[,2][1], yend = cm[2] + semiaxes[2] * S_eig$vectors[,2][2]),
             arrow = arrow(type = "closed", length = unit(0.15, "inches")), color
             ↪  = "blue", size = 1) +
geom_point(aes(x = cm[1], y = cm[2]), color = "red", size = 3) +
geom_segment(aes(x = L1, y = 0, xend = L1, yend = L2), col = "red", linetype = 2)
↪  +
geom_segment(aes(x = L1, y = L2, xend = L1, yend = R2), col = "red") +
geom_segment(aes(x = L1, y = R2, xend = L1, yend = 80), col = "red", linetype = 2)
↪  +
geom_segment(aes(x = R1, y = 0, xend = R1, yend = L2), col = "red", linetype = 2)
↪  +
geom_segment(aes(x = R1, y = L2, xend = R1, yend = R2), col = "red") +
geom_segment(aes(x = R1, y = R2, xend = R1, yend = 80), col = "red", linetype = 2)
↪  +
geom_segment(aes(x = 0, y = L2, xend = L1, yend = L2), col = "red", linetype = 2)
↪  +
geom_segment(aes(x = L1, y = L2, xend = R1, yend = L2), col = "red") +
geom_segment(aes(x = R1, y = L2, xend = 10, yend = L2), col = "red", linetype = 2)
↪  +
geom_segment(aes(x = 0, y = R2, xend = L1, yend = R2), col = "red", linetype = 2)
↪  +
geom_segment(aes(x = L1, y = R2, xend = R1, yend = R2), col = "red") +
geom_segment(aes(x = R1, y = R2, xend = 10, yend = R2), col = "red", linetype = 2)
↪  +
theme_minimal() +
theme_minimal() +
labs(x = "mu_1", y = "mu_2") +
theme(plot.title = element_text(hjust = 0.5),
      plot.background = element_rect(fill = "white", colour = "black", size = 2),
      panel.grid.major = element_line(color = "gray80"),
```

```
        panel.grid.minor = element_blank())
print(p)
```



To fid such a μ_0, we need to find a point that is outside the ellipse but inside the rectangular. One possible choice is $\mu_0 = (5.4, 39)'$.