

Nama : M. Mahbubbillah

NIM : 222011569

Kelas : 3SI1

Mata Kuliah : Analisis Peubah Ganda (P)

Tugas Pertemuan 3 (Penugasan 3.5)

Keberhasilan suatu pembangunan di suatu daerah dapat dilihat dari beberapa indikator, diantaranya adalah Indeks Pembangunan Manusia (IPM) dan laju pertumbuhan ekonomi. Pemerintah RI menetapkan target dalam RPJMN 2020-2024 IPM sebesar 63,94 dan laju pertumbuhan ekonomi di Provinsi Papua sebesar 4,7 persen. Tahun 2020 diperoleh data sebagai berikut.

Kab/Kota	IPM	Laju	Kab/Kota	IPM	Laju
Merauke	70.09	0.07	Sarmi	63.63	4.38
Jayawijaya	58.03	-2.10	Keerom	66.40	2.49
Jayapura	71.69	-4.57	Waropen	64.94	-0.28
Nabire	68.83	1.04	Supiori	62.30	3.26
Kepulauan Yapen	67.66	-2.16	Mamberamo Raya	51.78	4.77
Biak Numfor	72.19	-4.63	Nduga	31.55	4.49
Paniai	56.31	2.50	Lanny Jaya	47.86	5.82
Puncak Jaya	48.37	-1.11	Mamberamo Tengah	47.57	4.40
Mimika	74.19	15.62	Yalimo	48.34	5.27
Boven Digoel	61.53	-0.25	Puncak	43.04	2.66
Mappi	58.15	1.01	Dogiyai	54.84	3.23
Asmat	50.55	4.09	Intan Jaya	47.79	2.01
Yahukimo	49.37	3.48	Deiyai	49.46	3.56
Pegunungan Bintang	45.44	2.87	Kota Jayapura	79.94	-0.51
Tolikara	49.50	5.03			

a. Apakah data mengandung outlier? Buatlah chi-square plot untuk mendeteksinya.

Dengan menggunakan software R, diperoleh hasil berikut ini.

Test <chr>	Statistic <fctr>	p value <fctr>	Result <chr>
Mardia Skewness	37.3538973049009	1.52286713326e-07	NO
Mardia Kurtosis	5.68021034001387	1.34529198891897e-08	NO
MVN	NA	NA	NO

Test <chr>	HZ <dbl>	p value <dbl>	MVN <chr>
Henze-Zirkler	1.097582	0.006932237	NO

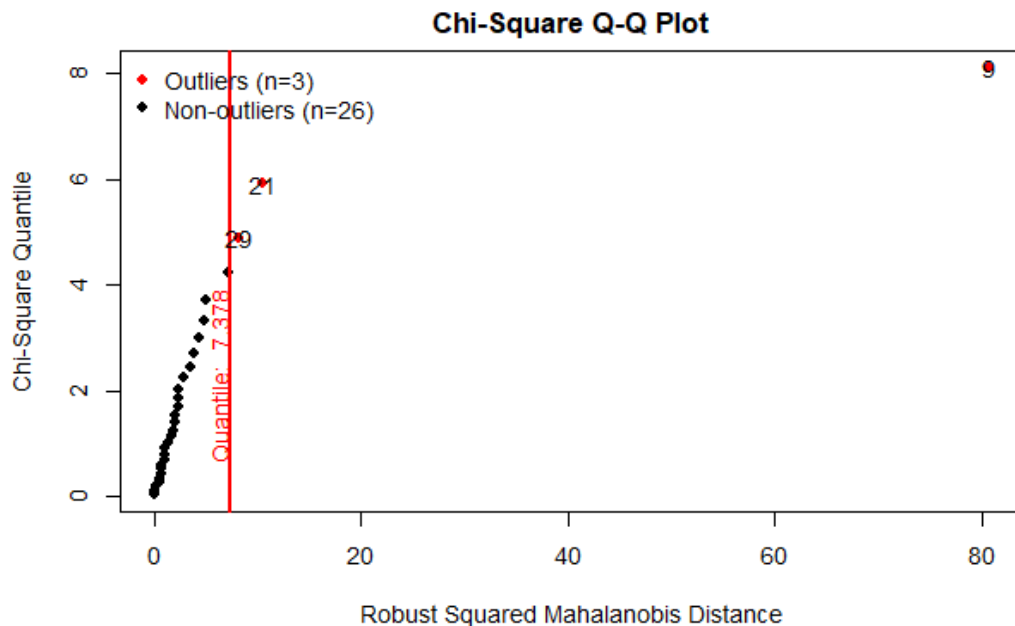
Test <chr>	H <dbl>	p value <dbl>	MVN <chr>
Royston	10.14344	0.00623429	NO

Test <chr>	E <dbl>	df <dbl>	p value <dbl>	MVN <chr>
Doornik-Hansen	29.68167	4	5.681798e-06	NO

Test <chr>	Statistic <dbl>	p value <dbl>	MVN <chr>
E-statistic	1.479482	0	NO

Berdasarkan output R di atas, dapat diketahui bahwa dari seluruh jenis uji multivariate menghasilkan keputusan tolak H0. Dengan tingkat signifikansi 5%, belum cukup bukti bahwa dataset berdistribusi normal.

Sedangkan, untuk pendeteksian outlier didapatkan hasil sebagai berikut.



Berdasarkan output R diatas, dapat diketahui bahwa dataset ini memiliki 3 pencilan yaitu data nomor 29, 21, dan 9. Hal ini juga bisa menjadi penyebab dataset ini tidak berdistribusi normal. Untuk mengoreksi data, dapat dilakukan dengan menghilangkan pencilan pada dataset ini.

b. Lakukan uji multivariate normal dan apa yang terjadi jika data outlier dibuang?

Setelah mempertimbangkan beberapa hal, saya memutuskan untuk hanya membuang data nomor 9. Hal ini dikarenakan data nomor 9 sangat jauh dari sebaran data lainnya apabila dilihat dari grafik ellips diatas. Dan, berikut merupakan hasil uji multivariate setelah data outlier dibuang menggunakan software R.

Test <chr>	Statistic <fctr>	p value <fctr>	Result <chr>
Mardia Skewness	2.66300501394376	0.615703603468754	YES
Mardia Kurtosis	-0.722188768789149	0.470178427488964	YES
MVN	NA	NA	YES

Test <chr>	HZ <dbl>	p value <dbl>	MVN <chr>
Henze-Zirkler	0.6309832	0.1439167	YES

Test <chr>	H <dbl>	p value <dbl>	MVN <chr>
Royston	3.431584	0.1248133	YES

Test <chr>	E <dbl>	df <dbl>	p value <dbl>	MVN <chr>
Doornik-Hansen	71.70624	4	9.900616e-15	NO

Test <chr>	Statistic <dbl>	p value <dbl>	MVN <chr>
E-statistic	0.8300293	0.144	YES

Berdasarkan output R di atas, dapat diketahui bahwa dari seluruh jenis uji multivariate menghasilkan keputusan gagal tolak H0. Dengan tingkat signifikansi 5%, cukup bukti bahwa dataset berdistribusi normal.

Berdasarkan pernyataan di atas, dapat disimpulkan bahwa perlakuan untuk dataset yaitu membuang data outlier berhasil dalam mengatasi permasalahan normalitas data.

- c. Setelah permasalahan outlier ditangani, tuliskan hipotesis penelitiannya, lakukan prosedur pengujian hipotesisnya dan buatlah ellipsoid daerah penerimaan H0 pada taraf $\alpha=5\%$

Diketahui :

H0 : myu = (63.94, 4.7)

H1 : myu != (63.94, 4.7)

Alpha = 5%

Dikarenakan pada kasus kali ini tidak memiliki varians dan memiliki sampel yang kecil (kurang dari 30), maka rumus T uji yang digunakan :

$$T^2 = n(\bar{X} - \mu_0)'(S)^{-1}(\bar{X} - \mu_0) \sim \frac{(n-1)p}{n-p} F_{(p, n-p)}$$

Dengan syarat (Gagal Tolak H0) :

$$n(\bar{x} - \mu)'(S)^{-1}(\bar{x} - \mu) \leq \frac{(n-1)p}{n-p} F_{(p, n-p); \alpha}$$

Output R :

T2 <dbl>	Fstat <dbl>	df1 <int>	df2 <int>	p.value <dbl>
112.3645	54.10141	2	26	5.416916e-10

Berdasarkan output di atas, diketahui bahwa T uji bernilai 112,3645 dan F Statistik bernilai 54.101. Hal ini dapat disimpulkan bahwa hipotesis nol ditolak karena $T \text{ Uji} > F \text{ Stat}$ ($P\text{-Value} < 0,05$).

Dengan tingkat signifikansi 5%, belum cukup bukti bahwa nilai rata-rata dari masing-masing variabel sama dengan nilai rata-rata pada masing-masing pasangan hipotesisnya.

d. Buatlah simultan confidence intervalnya.

Dengan menggunakan Confidence Interval Bonferroni dengan software R, didapatkan hasil berikut.

Variabel 1 (IPM)

```
      [,1]  
[1,] 49.47697  
      [,1]  
[1,] 63.89089
```

Berdasarkan output R di atas, diketahui bahwa batas bawah dari variabel 1 adalah 49,47697 dan batas atas dari variabel 1 adalah 63,89089. Hal ini berarti bahwa nilai rata-rata hipotesis (63,94) berada diluar confidence interval yang berarti bahwa tolak H_0 .

Variabel 2 (Laju Perumbuhan Ekonomi)

```
      [,1]  
[1,] -0.07844248  
      [,1]  
[1,] 3.708442
```

Berdasarkan output R di atas, diketahui bahwa batas bawah dari variabel 2 adalah -0.078 dan batas atas dari variabel 2 adalah 3,708442. Hal ini berarti bahwa nilai rata-rata hipotesis (4,7) berada diluar confidence interval yang berarti bahwa tolak H_0 .

e. Apa kesimpulan Anda?

Berdasarkan analisis yang telah dilakukan di atas, dapat disimpulkan bahwa target pemerintah RI dalam RPJMN 2020-2024 IPM sebesar 63,94 dan laju pertumbuhan ekonomi di Provinsi Papua sebesar 4,7 persen tidak terpenuhi meskipun angka IPM hampir mencapai target. Namun, laju pertumbuhan ekonomi papua masih jauh dari target yang telah ditetapkan pemerintah.

Selain itu, terdapat data pencilan yaitu data miliki kabupaten mimika. Data tersebut terletak sangat jauh dari sebaran data lainnya. Hal ini yang menyebabkan gangguan normalitas pada data. Hal ini juga harus menjadi perhatian pemerintah setempat karena terjadinya anomali data ini.

Lampiran

Tugas Pertemuan 3

M.Mahbubillah

2022-09-10

a. Apakah data mengandung outlier? Buatlah chi-square plot untuk mendeteksinya.

```
library(MVN)

## Warning: package 'MVN' was built under R version 4.1.3

library(RVAideMemoire)

## Warning: package 'RVAideMemoire' was built under R version 4.1.3

## *** Package RVAideMemoire v 0.9-81-2 ***

library(car)

## Warning: package 'car' was built under R version 4.1.3

## Loading required package: carData

## Warning: package 'carData' was built under R version 4.1.2

IPM <- c(70.09, 58.03, 71.69, 68.83, 67.66, 72.19, 56.31, 48.37, 74.19, 61.53, 58.15, 50
.55, 49.37, 45.44, 49.50, 63.63, 66.40, 64.94, 62.30, 51.78, 31.55, 47.86, 47.57, 48.34, 4
3.04, 54.84, 47.79, 49.46, 79.94)
laju <- c(0.07, -2.10, -4.57, 1.04, -2.16, -4.63, 2.50, -1.11, 15.62, -0.25, 1.01, 4.09,
3.48, 2.87, 5.03, 4.38, 2.49, -0.28, 3.26, 4.77, 4.49, 5.82, 4.40, 5.27, 2.66, 3.23, 2.01, 3
.56, -0.51)

dataP3 <- data.frame(IPM, laju)
dataP3

##      IPM  laju
## 1  70.09  0.07
## 2  58.03 -2.10
## 3  71.69 -4.57
## 4  68.83  1.04
## 5  67.66 -2.16
## 6  72.19 -4.63
## 7  56.31  2.50
## 8  48.37 -1.11
## 9  74.19 15.62
## 10 61.53 -0.25
## 11 58.15  1.01
```

```
## 12 50.55 4.09
## 13 49.37 3.48
## 14 45.44 2.87
## 15 49.50 5.03
## 16 63.63 4.38
## 17 66.40 2.49
## 18 64.94 -0.28
## 19 62.30 3.26
## 20 51.78 4.77
## 21 31.55 4.49
## 22 47.86 5.82
## 23 47.57 4.40
## 24 48.34 5.27
## 25 43.04 2.66
## 26 54.84 3.23
## 27 47.79 2.01
## 28 49.46 3.56
## 29 79.94 -0.51

mshapiro.test(dataP3)

##
## Multivariate Shapiro-Wilk normality test
##
## data: (IPM,laju)
## W = 0.73189, p-value = 6.242e-06

mvn(dataP3, mvnTest = "mardia",alpha = 0.05)

## $multivariateNormality
##          Test          Statistic          p value Result
## 1 Mardia Skewness 37.3538973049009 1.52286713326e-07 NO
## 2 Mardia Kurtosis 5.68021034001387 1.34529198891897e-08 NO
## 3          MVN          <NA>          <NA> NO
##
## $univariateNormality
##          Test Variable Statistic p value Normality
## 1 Anderson-Darling IPM 0.5212 0.1699 YES
## 2 Anderson-Darling laju 0.7890 0.0359 NO
##
## $Descriptives
##          n          Mean Std.Dev Median Min Max 25th 75th Skew
## IPM 29 57.287586 11.284393 56.31 31.55 79.94 48.37 66.40 0.03806064
## laju 29 2.291034 3.825124 2.66 -4.63 15.62 -0.25 4.38 0.97210195
##          Kurtosis
## IPM -0.7497457
## laju 3.0491227

mvn(dataP3, mvnTest = "hz",alpha = 0.05)
```

```

## $multivariateNormality
##           Test      HZ      p value MVN
## 1 Henze-Zirkler 1.097582 0.006932237 NO
##
## $univariateNormality
##           Test Variable Statistic    p value Normality
## 1 Anderson-Darling   IPM      0.5212    0.1699    YES
## 2 Anderson-Darling   laju      0.7890    0.0359    NO
##
## $Descriptives
##           n      Mean   Std.Dev Median   Min   Max  25th  75th      Skew
## IPM  29 57.287586 11.284393  56.31 31.55 79.94 48.37 66.40 0.03806064
## laju 29  2.291034  3.825124   2.66 -4.63 15.62 -0.25  4.38 0.97210195
##           Kurtosis
## IPM  -0.7497457
## laju  3.0491227

mvn(dataP3, mvnTest = "royston",alpha = 0.05)

## $multivariateNormality
##           Test      H      p value MVN
## 1 Royston 10.14344 0.00623429 NO
##
## $univariateNormality
##           Test Variable Statistic    p value Normality
## 1 Anderson-Darling   IPM      0.5212    0.1699    YES
## 2 Anderson-Darling   laju      0.7890    0.0359    NO
##
## $Descriptives
##           n      Mean   Std.Dev Median   Min   Max  25th  75th      Skew
## IPM  29 57.287586 11.284393  56.31 31.55 79.94 48.37 66.40 0.03806064
## laju 29  2.291034  3.825124   2.66 -4.63 15.62 -0.25  4.38 0.97210195
##           Kurtosis
## IPM  -0.7497457
## laju  3.0491227

mvn(dataP3, mvnTest = "dh",alpha = 0.05)

## $multivariateNormality
##           Test      E df      p value MVN
## 1 Doornik-Hansen 29.68167  4 5.681798e-06 NO
##
## $univariateNormality
##           Test Variable Statistic    p value Normality
## 1 Anderson-Darling   IPM      0.5212    0.1699    YES
## 2 Anderson-Darling   laju      0.7890    0.0359    NO
##
## $Descriptives
##           n      Mean   Std.Dev Median   Min   Max  25th  75th      Skew
## IPM  29 57.287586 11.284393  56.31 31.55 79.94 48.37 66.40 0.03806064
## laju 29  2.291034  3.825124   2.66 -4.63 15.62 -0.25  4.38 0.97210195

```

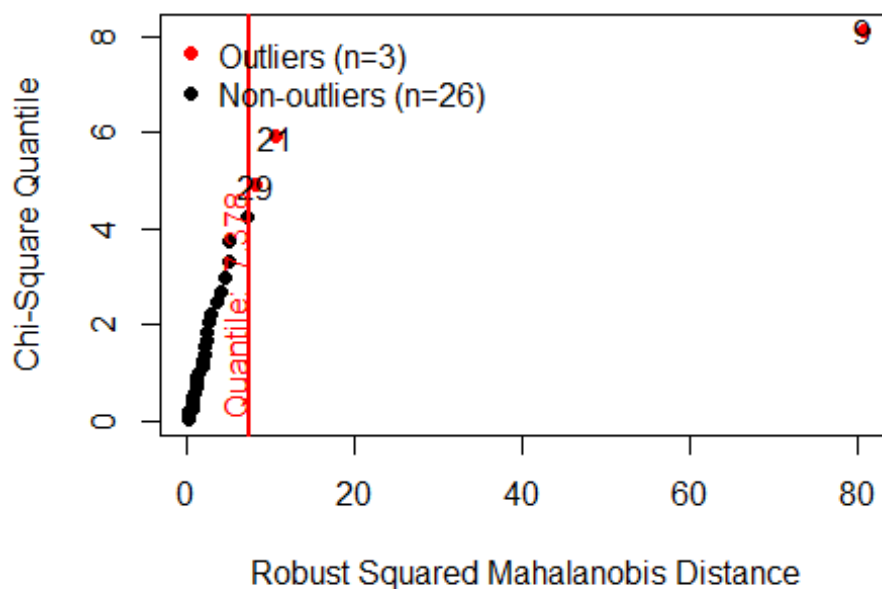
```
##          Kurtosis
## IPM    -0.7497457
## laju    3.0491227

mvn(dataP3, mvnTest = "energy", alpha = 0.05)

## $multivariateNormality
##          Test Statistic p value MVN
## 1 E-statistic  1.479482      0 NO
##
## $univariateNormality
##          Test Variable Statistic  p value Normality
## 1 Anderson-Darling    IPM      0.5212    0.1699    YES
## 2 Anderson-Darling    laju      0.7890    0.0359    NO
##
## $Descriptives
##          n      Mean   Std.Dev Median   Min   Max  25th  75th      Skew
## IPM   29  57.287586  11.284393   56.31  31.55  79.94  48.37  66.40  0.03806064
## laju  29   2.291034   3.825124    2.66  -4.63  15.62  -0.25   4.38  0.97210195
##          Kurtosis
## IPM    -0.7497457
## laju    3.0491227

hasilUji <- mvn(data = dataP3, multivariateOutlierMethod = "quan", showOutliers = TRUE)
```

Chi-Square Q-Q Plot



hasilUji


```
## $multivariateNormality
##           Test           HZ      p value MVN
## 1 Henze-Zirkler 1.097582 0.006932237 NO
##
## $univariateNormality
##           Test Variable Statistic    p value Normality
## 1 Anderson-Darling    IPM      0.5212    0.1699    YES
## 2 Anderson-Darling    laju      0.7890    0.0359    NO
##
## $Descriptives
##           n      Mean   Std.Dev Median   Min   Max  25th  75th      Skew
## IPM  29 57.287586 11.284393  56.31 31.55 79.94 48.37 66.40 0.03806064
## laju 29  2.291034  3.825124   2.66 -4.63 15.62 -0.25  4.38 0.97210195
##           Kurtosis
## IPM  -0.7497457
## laju  3.0491227
##
## $multivariateOutliers
## Observation Mahalanobis Distance Outlier
## 9           9              80.699    TRUE
## 21          21              10.465    TRUE
## 29          29               8.191    TRUE
```

b. Lakukan uji multivariate normal dan apa yang terjadi jika data outlier dibuang?

```
dataP3_2 <- dataP3[-c(9),]
mvn(dataP3_2, mvnTest = "mardia", alpha = 0.05)

## $multivariateNormality
##           Test           Statistic      p value Result
## 1 Mardia Skewness    2.66300501394376 0.615703603468754    YES
## 2 Mardia Kurtosis   -0.722188768789149 0.470178427488964    YES
## 3           MVN              <NA>          <NA>    YES
##
## $univariateNormality
##           Test Variable Statistic    p value Normality
## 1 Anderson-Darling    IPM      0.5127    0.1781    YES
## 2 Anderson-Darling    laju      0.6710    0.0712    YES
##
## $Descriptives
##           n      Mean   Std.Dev Median   Min   Max  25th  75th      Skew
## IPM  28 56.68393 11.004302 55.575 31.55 79.94 48.3625 65.3050 0.07194461
## laju 28  1.81500  2.891096  2.580 -4.63  5.82 -0.2575  4.1625 -0.68716526
##           Kurtosis
## IPM  -0.6233012
## laju -0.5471259

mvn(dataP3_2, mvnTest = "hz", alpha = 0.05)

## $multivariateNormality
##           Test           HZ      p value MVN
```

```
## 1 Henze-Zirkler 0.6309832 0.1439167 YES
##
## $univariateNormality
##           Test Variable Statistic    p value Normality
## 1 Anderson-Darling   IPM      0.5127    0.1781     YES
## 2 Anderson-Darling   laju      0.6710    0.0712     YES
##
## $Descriptives
##      n      Mean   Std.Dev Median   Min    Max   25th   75th      Skew
## IPM  28 56.68393 11.004302 55.575 31.55 79.94 48.3625 65.3050 0.07194461
## laju 28  1.81500  2.891096  2.580 -4.63  5.82 -0.2575  4.1625 -0.68716526
##           Kurtosis
## IPM  -0.6233012
## laju -0.5471259

mvn(dataP3_2, mvnTest = "royston",alpha = 0.05)

## $multivariateNormality
##           Test      H    p value MVN
## 1 Royston 3.431584 0.1248133 YES
##
## $univariateNormality
##           Test Variable Statistic    p value Normality
## 1 Anderson-Darling   IPM      0.5127    0.1781     YES
## 2 Anderson-Darling   laju      0.6710    0.0712     YES
##
## $Descriptives
##      n      Mean   Std.Dev Median   Min    Max   25th   75th      Skew
## IPM  28 56.68393 11.004302 55.575 31.55 79.94 48.3625 65.3050 0.07194461
## laju 28  1.81500  2.891096  2.580 -4.63  5.82 -0.2575  4.1625 -0.68716526
##           Kurtosis
## IPM  -0.6233012
## laju -0.5471259

mvn(dataP3_2, mvnTest = "dh",alpha = 0.05)

## $multivariateNormality
##           Test      E df      p value MVN
## 1 Doornik-Hansen 71.70624  4 9.900616e-15 NO
##
## $univariateNormality
##           Test Variable Statistic    p value Normality
## 1 Anderson-Darling   IPM      0.5127    0.1781     YES
## 2 Anderson-Darling   laju      0.6710    0.0712     YES
##
## $Descriptives
##      n      Mean   Std.Dev Median   Min    Max   25th   75th      Skew
## IPM  28 56.68393 11.004302 55.575 31.55 79.94 48.3625 65.3050 0.07194461
## laju 28  1.81500  2.891096  2.580 -4.63  5.82 -0.2575  4.1625 -0.68716526
##           Kurtosis
```

```
## IPM -0.6233012
## laju -0.5471259

mvn(dataP3_2, mvnTest = "energy", alpha = 0.05)

## $multivariateNormality
##          Test Statistic p value MVN
## 1 E-statistic 0.8300293  0.144 YES
##
## $univariateNormality
##          Test Variable Statistic  p value Normality
## 1 Anderson-Darling   IPM      0.5127    0.1781    YES
## 2 Anderson-Darling   laju      0.6710    0.0712    YES
##
## $Descriptives
##          n      Mean   Std.Dev Median   Min    Max   25th   75th      Skew
## IPM  28 56.68393 11.004302 55.575 31.55 79.94 48.3625 65.3050 0.07194461
## laju 28  1.81500  2.891096  2.580 -4.63  5.82 -0.2575  4.1625 -0.68716526
##          Kurtosis
## IPM  -0.6233012
## laju -0.5471259

mshapiro.test(dataP3_2)

##
## Multivariate Shapiro-Wilk normality test
##
## data: (IPM,laju)
## W = 0.98143, p-value = 0.8832
```

c. Setelah permasalahan outlier ditangani, tuliskan hipotesis penelitiannya, lakukan prosedur pengujian hipotesisnya dan buatlah ellipsoid daerah penerimaan H_0 pada taraf $\alpha=5\%$

```
T.test <- function(X, mu=0){
  X <- as.matrix(X)
  n <- nrow(X)
  p <- ncol(X)
  df2 <- n - p
  if(df2 < 1L) stop("Need nrow(X) > ncol(X).")
  if(length(mu) != p) mu <- rep(mu[1], p)
  xbar <- colMeans(X)
  S <- cov(X)

  T2 <- n * t(xbar - mu) %*% solve(S) %*% (xbar - mu)

  Fstat <- T2 / (p * (n-1) / df2)
  pval <- 1 - pf(Fstat, df1=p, df2=df2)
  data.frame(T2=as.numeric(T2), Fstat=as.numeric(Fstat),
    df1=p, df2=df2, p.value=as.numeric(pval), row.names="")
}
```

```

myu <- c(63.94,4.7)
myu

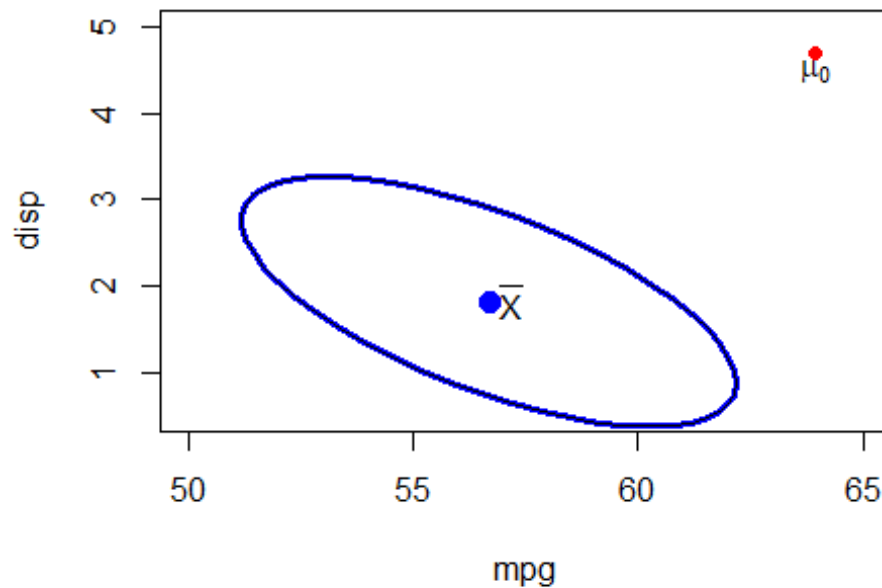
## [1] 63.94  4.70

T.test(dataP3_2, myu)

##           T2      Fstat df1 df2      p.value
## 112.3645 54.10141    2   26 5.416916e-10

n <- nrow(dataP3_2)
p <- ncol(dataP3_2)
ybar <- colMeans(dataP3_2)
s <- cov(dataP3_2)
alpha <- 0.05
tconst <- sqrt((p/n) * ((n-1)/(n-p)) * qf(1-alpha,p,n-p))
id <- c(1,2)
plot(ellipse(center=ybar[id],
              shape=s[id,id],
              radius=tconst, draw=F),
      type="n", xlab="mpg", ylab="disp", xlim = c(50,65), ylim = c(0.5, 5))
lines(ellipse(center=ybar[id],
              shape=s[id,id],
              radius=tconst,
              lwd=3), xlab="mpg", ylab="disp")
points(63.94,4.7,col="red", pch=16)
text(63.94,4.5, expression(mu[0]))
text(ybar[1]+0.5, ybar[2], expression(bar(X)))

```



d. Buatlah simultan confidence intervalnya.

```
a1 <- c(1,0)
ta1 <- t(a1)
a2 <- c(0,1)
ta2 <- t(a2)
ttab <- qt(1-(0.05/(58)),28)

pembilang1 <- sqrt(ta1%%s%%a1)
penyebut1 <- sqrt(n)
b1 <- pembilang1/penyebut1
lower1 <- ta1%%ybar - ttab * b1
upper1 <- ta1%%ybar + ttab * b1
lower1

##           [,1]
## [1,] 49.47697

upper1

##           [,1]
## [1,] 63.89089

pembilang2 <- sqrt(ta2%%s%%a2)
penyebut2 <- sqrt(n)
b2 <- pembilang2/penyebut2
lower2 <- ta2%%ybar - ttab * b2
```

```
upper2 <- ta2*%ybar + ttab * b2  
lower2
```

```
##           [,1]  
## [1,] -0.07844248
```

```
upper2
```

```
##           [,1]  
## [1,] 3.708442
```