

Handout 2  
Sebaran Normal Ganda  
(*Multivariate Normal Distribution*)

Variabel  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p$  dikatakan berdistribusi normal ganda dengan parameter  $\mu$  dan  $\Sigma$  jika mempunyai *probability density function* :

$$f(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{p/2}} e^{-\frac{1}{2}(\mathbf{X}-\mu)' \Sigma^{-1} (\mathbf{X}-\mu)}$$

jika  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p$  berdistribusi normal multivariat maka  $n(\mathbf{X} - \mu)' \Sigma^{-1} (\mathbf{X} - \mu)$  berdistribusi  $\chi_p^2$ .

Berdasarkan sifat ini maka pemeriksaan distribusi normal multivariat dapat dilakukan pada setiap populasi dengan cara membuat *q-q plot* atau *scatter-plot* dari nilai  $d_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$ ,  $i = 1, 2, \dots, n$ .

Tahapan dari pembuatan *q-q plot* ini adalah sebagai berikut :

1. Mulai
2. Tentukan nilai vektor rata-rata:  $\bar{\mathbf{X}}$
3. Tentukan nilai matriks varians-kovarians:  $\mathbf{S}$
4. Tentukan nilai jarak *mahalanobis* atau kuadrat *general* setiap titik pengamatan dengan vektor rata-ratanya  $d_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$ ,  $i = 1, 2, \dots, n$ .
5. Urutkan nilai  $d_i^2$  dari kecil ke besar:  $d_{(1)}^2 \leq d_{(2)}^2 \leq d_{(3)}^2 \leq \dots \leq d_{(n)}^2$ .
6. Tentukan nilai  $p_i = \frac{i-1/2}{n}$ ,  $i = 1, 2, \dots, n$ .
7. Tentukan nilai  $q_i$  sedemikian hingga  $\int_{-\infty}^{q_i} f(\chi^2) d\chi^2 = p_i$  atau  $q_{i,p}(p_i) = \chi_p^2((n-i+1/2)/n)$ .
8. Buat *scatter-plot*  $d_{(i)}^2$  dengan  $q_i$
9. Jika *scatter-plot* ini cenderung membentuk garis lurus dan lebih dari 50% nilai  $d_i^2 \leq \chi_p^2(0,50)$ , artinya data berdistribusi normal multivariat.
10. Selesai

## LATIHAN

Berikut data diambil dari Garperz (1992) dalam simatupang (2002).

$X_1$  = Kontribusi industri manufaktur dalam produk domestik regional bruto(%)

$X_2$  = Banyaknya tenaga kerja dalam sektor industri manufaktur (%)

No.	$X_1$	$X_2$
1	8.8	2589
2	8.5	1186
3	7.7	291
4	4.9	1276
5	9.6	6633
6	10	12125
7	11.5	36717
8	11.6	43319
9	11.2	10530
10	10.7	3931
11	10	1536
12	6.8	61400

Akan di uji kenormalan dari data dengan menggunakan excel. Macro SAS. dan R.

### 1. Menggunakan Excel

- a. Tentukan nilai vektor rata-rata:  $\bar{X}$

No.	$X_1$	$X_2$
1	8.8	2589
2	8.5	1186
3	7.7	291
4	4.9	1276
5	9.6	6633
6	10	12125
7	11.5	36717
8	11.6	43319
9	11.2	10530
10	10.7	3931
11	10	1536
12	6.8	61400
$\bar{X}$	9.275	15127.75

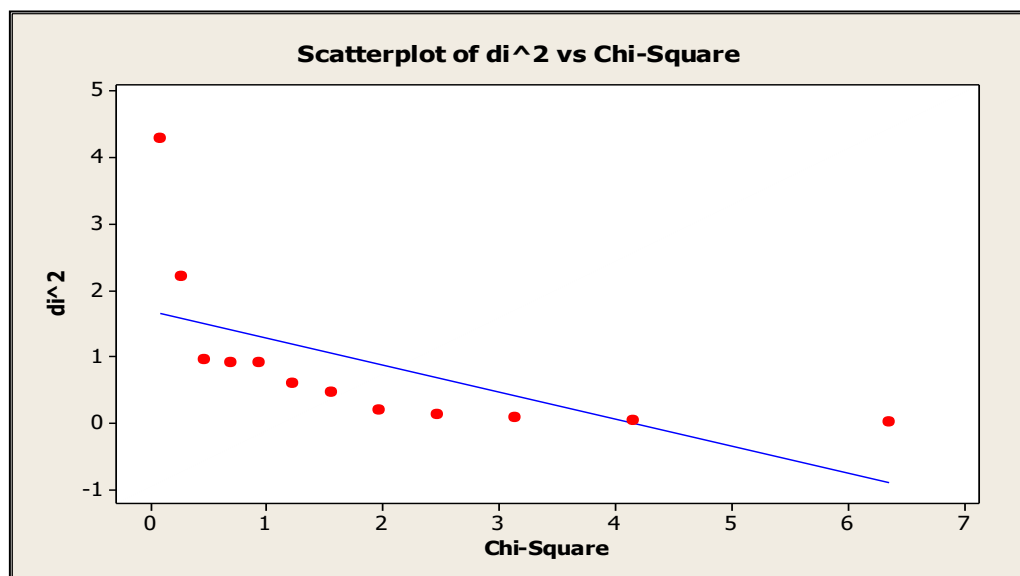
- b. Tentukan nilai matriks varians-kovarians:  $S^{-1}$

0.244275	-0.000003
-0.000003	0

- c. Jarak *mahalanobis* yang sudah dari kecil ke besar,  $p_i = \frac{i-1/2}{n}$ ,  $i = 1, 2, \dots, n$  dan nilai  $q_i$  sedemikian hingga  $\int_{-\infty}^{q_i} f(\chi^2) d\chi^2 = p_i$  atau  $q_{i,p}(p_i) = \chi_p^2((n-i+1/2)/n)$ .

$d_i^2$	$i$	$(i-0.5)/n$	$\chi^2$
0.018188	12	0.958333	6.35610766
0.079728	10	0.791667	3.13723184
0.461074	7	0.541667	1.56031716
4.299847	1	0.041667	0.08511923
0.042918	11	0.875	4.15888308
0.141894	9	0.708333	2.46428736
0.91149	5	0.375	0.94000726
0.914082	4	0.291667	0.68968098
0.960066	3	0.208333	0.4672297
0.594954	6	0.458333	1.22620895
0.189492	8	0.625	1.96165855
2.206385	2	0.125	0.26706279

- d. Buat *scatter-plot*  $d_{(i)}^2$  dengan  $q_i$



Dari plot diatas terlihat bahwa ada 1 pengamatan berada jauh dari garis normal. sehingga untuk meyakinkan hasil dari plot tersebut maka akan dikorelasikan nilai  $d_i^2$  dengan  $\chi^2$  dengan pengujian hipotesis berikut :

$H_0$  : Data berdistribusi normal

$H_1$  : Data tidak berdistribusi normal

Hasil korelasi person dari  $d_i^2$  dengan  $\chi^2$  adalah sebesar -0.612. artinya ada hubungan negatif yang kuat antara  $d_i^2$  dengan  $\chi^2$ . Dapat diambil keputusan bahwa dari nilai P-Value sebesar  $0.035 < \alpha = 0.05$  yang berarti menolak  $H_0$ . Kesimpulan yang dapat diambil bahwa data tidak berdistribusi normal.

## 2. Menggunakan Macro SAS

✓ Sintaks untuk memanggil Macro SAS sebagai berikut :

```

Data NORMALGANDA;
Input x1 x2;
Cards;
8.8 2589
8.5 1186
7.7 291
4.9 1276
9.6 6633
10 12125
11.5 36717
11.6 43319
11.2 10530
10.7 3931
10 1536
6.8 61400
;
%inc "D:\Data\Bahan kuliah S2\MAKRO UJI NORMAL GANDA MARDIA.sas";
%multnorm(data= NORMALGANDA. var=x1 x2. plot=mult. hires=no)

```

✓ Output SAS :

MULTNORM macro: Univariate and Multivariate Normality Tests  
The MODEL Procedure

Normality Test			
Equation	Test Statistic	Value	Prob
x1	Shapiro-Wilk W	0.93	0.3598
x2	Shapiro-Wilk W	0.74	0.0016
System	Mardia Skewness	11.12	0.0252
	Mardia Kurtosis	0.59	0.5552
	Henze-Zirkler T	2.64	0.0083

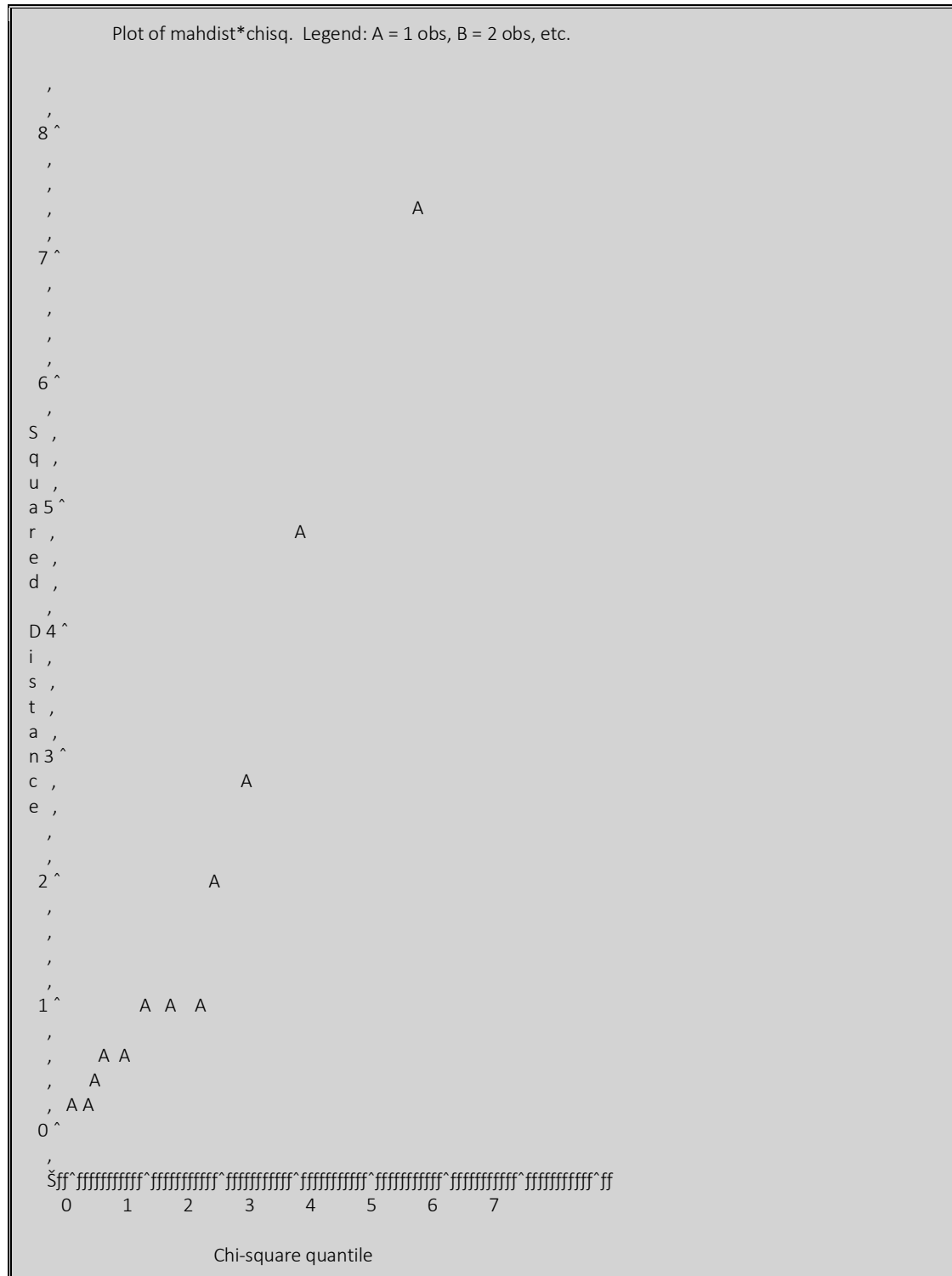
Dari hasil output SAS diatas akan dilakukan pengujian hipotesis terhadap mardia skewness test sebagai berikut :

$H_0$  : Peubah ganda mengikuti distribusi normal

$H_1$  : Peubah ganda tidak mengikuti distribusi normal

Sehingga disimpulkan  $p\text{-value} = 0.0252 < \alpha = 0.05$  yaitu menolak  $H_0$ . Artinya bahwa peubah ganda tidak mengikuti distribusi normal. Dari Q-Q plot yang dihasilkan dari output SAS juga menunjukkan bahwa sebaran dari data tidak mengikuti distribusi normal.

#### MULTNORM macro: Chi-square Q-Q plot



### 3. Menggunakan R

```
x<-c(8.8, 8.5, 7.7, 4.9, 9.6, 10, 11.5, 11.6, 11.2, 10.7, 10, 6.8, 2589, 1186, 291, 1276, 6633, 12125,  
      36717, 43319, 10530, 3931, 1536, 61400)
```

```
data<-matrix(x, nrow=12, ncol=2)
```

```
data
```

```
      [,1] [,2]  
[1,] 8.8 2589  
[2,] 8.5 1186  
[3,] 7.7 291  
[4,] 4.9 1276  
[5,] 9.6 6633  
[6,] 10.0 12125  
[7,] 11.5 36717  
[8,] 11.6 43319  
[9,] 11.2 10530  
[10,] 10.7 3931  
[11,] 10.0 1536  
[12,] 6.8 61400
```

```
#Uji normalitas ganda menggunakan Mardia's Skewness and kurtosis
```

```
mardia<-mardiaTest(data,cov=TRUE,qqplot=TRUE)
```

```
mardia
```

Mardia's Multivariate Normality Test

-----  
data : data

```
g1p      : 3.764213  
chi.skew  : 7.528426  
p.value.skew : 0.1104623  
g2p      : 9.362561  
z.kurtosis : 0.5900063  
p.value.kurt : 0.5551865
```

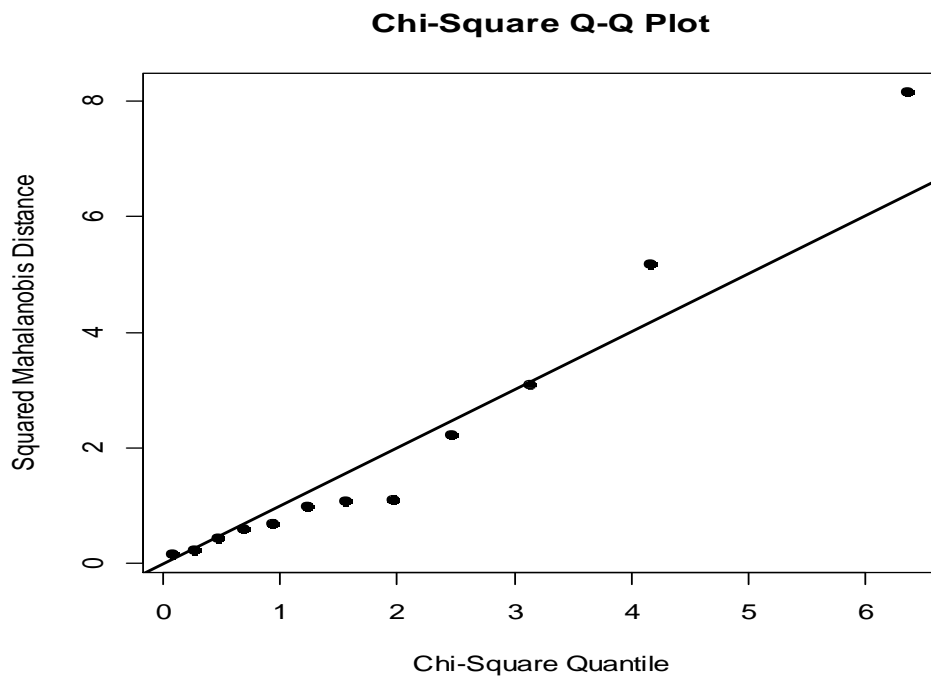
```
chi.small.skew : 11.12154  
p.value.small : 0.02523152
```

Dari hasil R diatas akan dilakukan pengujian hipotesis terhadap mardia skewness test sebagai berikut :

$H_0$  : Peubah ganda mengikuti distribusi normal

$H_1$  : Peubah ganda tidak mengikuti distribusi normal

Sehingga dikesimpulan  $p\text{-value} = 0.02523 < \alpha = 0.05$  yaitu menolak  $H_0$ . Artinya bahwa peubah ganda tidak mengikuti distribusi normal. Dari Q-Q plot yang dihasilkan dari output R dibawah ini juga menunjukan bahwa sebaran dari data tidak mengikuti distribusi normal.



#Uji normalitas ganda menggunakan Henze-Zikler Test

```
henze<-hzTest(data, cov=TRUE,qqplot=FALSE)
```

```
henze
```

Henze-Zirkler's Multivariate Normality Test

-----

data : data

HZ : 0.9995781

p-value : 0.004125476

Dari Henze-Zirkler's Multivariate Normality Test menghasilkan nilai p-value = 0.004125 <  $\alpha$  = 0.05. Hal ini berarti data mendukung untuk menolak  $H_0$ , dengan demikian dapat dikatakan bahwa peubah-peubah tersebut tidak mengikuti distribusi normal ganda.

```
#Uji Normalitas ganda menggunakan
royston<-roystonTest(data,qqplot=FALSE)
royston
Royston's Multivariate Normality Test
```

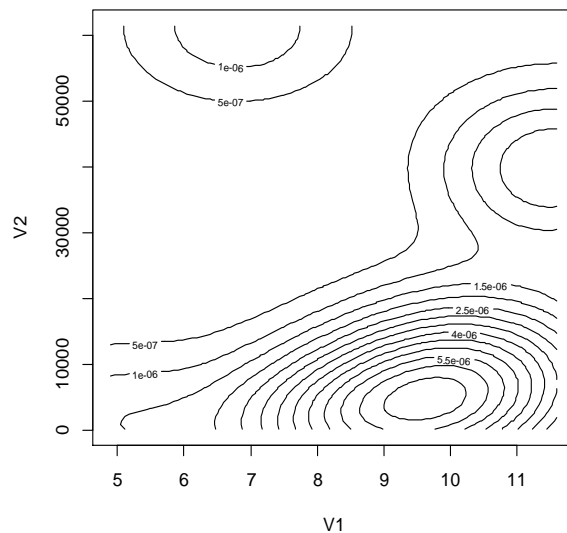
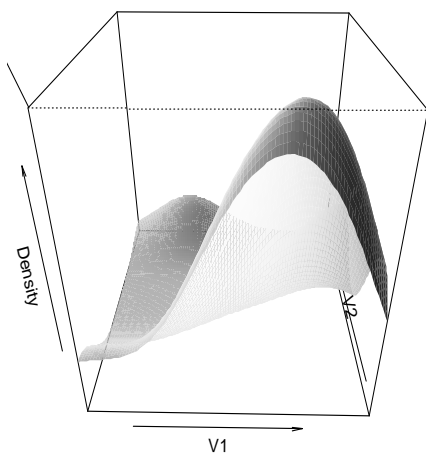
```
-----
data : data
```

```
H      : 10.10069
```

```
p-value : 0.006407917
```

Dari Royston Test menghasilkan nilai  $p\text{-value} = 0.0064 < \alpha = 0.05$ . Hasil uji ini juga menunjukkan data mendukung untuk menolak  $H_0$ , dengan demikian dapat dikatakan bahwa peubah-peubah tersebut tidak mengikuti distribusi normal ganda. Berikut ini akan ditampilkan bentuk dari countur data peubah ganda :

```
#Membuat kurva normal ganda dua
kurva=roystonTest(data)
mvnPlot(kurva,type="persp", default=TRUE)
mvnPlot(kurva,type="contour", default=TRUE)
```



Berdasarkan beberapa uji normalitas ganda di atas, dapat disimpulkan bahwa data tidak menyebar multivariat normal. Sehingga dapat dilakukan penanganan lanjutan terhadap data tersebut, dapat menggunakan transformasi normal atau metode lainnya.



## DAFTAR PUSTAKA

Johnson RA, Wichern DW. 2007. *Applied Multivariate Statistical Analysis*. Ed ke-6. New Jersey: Prentice Hall, Inc