

# ALGORITMA PENENTUAN DISTRIBUSI MULTIVARIAT NORMAL ANALISIS PEUBAH GANDA

**Audhi Apriliant (G14160021)**

Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Institut Pertanian Bogor, Jl. Meranti Wing 22 Lantai 4,  
Dramaga, Bogor, Jawa Barat 16680

Fungsi kepekatan Multivariat Normal (Normal ganda) adalah generalisasi dari fungsi kepekatan univariat Normal dengan  $p \geq 2$ . Fungsi kepekatan bersama dari  $p$  peubah acak yang menyebar Normal dan saling bebas adalah sebagai berikut:

$$f(x_1, x_2, \dots, x_p) = \frac{1}{(2\pi)^{\frac{p}{2}} \cdot (\sigma_1, \sigma_2, \dots, \sigma_p)} \exp \left[ -\frac{1}{2} \sum_{i=1}^p \left( \frac{x_i - \mu_i}{\sigma_i} \right)^2 \right]$$

Keterangan:

- Bentuk  $\left( \frac{x_1 - \mu}{\sigma} \right)^2$  dari eksponen fungsi sebaran Normal mengukur jarak kuadrat dari  $x_1$  ke  $\mu$  dalam unit simpangan baku
- Bentuk  $\left( \frac{x_1 - \mu}{\sigma} \right)^2$  dapat digeneralisasikan untuk vektor  $\mathbf{x}$  dari pengamatan beberapa peubah sebagai  $(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$

Secara umum, fungsi kepekatan peluang Normal bersama untuk  $p$  peubah dapat ditulis sebagai berikut:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}} \cdot |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp \left[ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] \text{ dimana } -\infty < x_i < \infty, i = 1, 2, \dots, p$$

Sifat-sifat sebaran Multivariat Normal

- Kombinasi linear dari semua komponen peubah  $x$  juga menyebar Normal. Jika  $X_p \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , maka kombinasi linear  $\mathbf{a}'\mathbf{x} = a_1x_1 + a_2x_2 + \dots + a_px_p$  menyebar  $N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$
- Jika  $X_p \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , maka semua anak gugus dari  $X$  juga menyebar Normal
- Jika  $\mathbf{x}_1$  dan  $\mathbf{x}_2$  saling bebas dan menyebar  $X_1 \sim N_{q_1}(\mu_1, \Sigma_{11})$  dan  $X_2 \sim N_{q_2}(\mu_2, \Sigma_{22})$ , maka sebaran bersyarat  $\frac{x_1}{x_2}$  adalah Multivariat Normal yang menyebar

$$N_{q_1+q_2} \left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix} \right)$$

Untuk mengevaluasi apakah data yang dimiliki menyebar Multivariat Normal dapat ditelusuri secara eksploratif. Seperti hanya untuk kasus univariat, penelusuran sebaran Multivariat Normal dapat juga memanfaatkan plot kuantil-kuantil (*Quantil Chi-Square*).

### Algoritma pemeriksaan Multivariat Normal (Johnson 1990)

Peubah  $X_1, X_2, \dots, X_p$  dikatakan menyebar Multivariat Normal, maka  $(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$  berdistribusi  $\chi^2_p$ . Berdasarkan sifat ini, maka pemeriksaan Multivariat Normal dapat dilakukan dengan cara membuat quantil plot dari  $d_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$  dimana  $i = 1, 2, \dots, n$

1. Tentukan nilai vektor rata-rata  $\bar{\mathbf{x}}$
2. Tentukan nilai matriks ragam-peragam (*variance-covariance*)  $\mathbf{S}$
3. Tentukan nilai jarak Mahalanobis setiap titik pengamatan dengan vektor rata-rata dari  $d_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$  dimana  $i = 1, 2, \dots, n$
4. Urutkan nilai  $d_i^2$  dari terkecil hingga terbesar  $d_{(1)}^2 \leq d_{(2)}^2 \leq d_{(3)}^2 \leq \dots \leq d_{(n)}^2$
5. Tentukan nilai  $p_i = \frac{i-0.5}{n}$ ,  $i = 1, 2, \dots, n$
6. Tentukan nilai  $q_i$  sedemikian sehingga  $p_i = \int_{-\infty}^{q_i} f(\chi^2) d\chi^2$
7. Buat *scatterplot*  $d_{(i)}^2$  dengan  $q_i$
8. Apabila *scatterplot* ini cenderung membentuk garis lurus dan lebih dari 50% nilai  $d_{(i)}^2 \leq \chi^2_{(0.05)(p)}$  dimana  $p$  adalah banyaknya peubah

## LAMPIRAN ALGORITMA PENENTUAN DISTRIBUSI MULTIVARIAT NORMAL ANALISIS PEUBAH GANDA

Diberikan data 17 negara Asia dengan peubah DENSITY, URBAN, LIFEEXPF, LIFEEXPM, LITERACY, BABYMORT, dan GDP CAP.

COUNTRY	X1	X2	X3	X4	X5	X6	X7
Afganistan	25	18	44	45	29	168	205
Bangladesh	800	16	53	53	35	106	202
Cambodia	55	12	52	50	35	112	260
China	124	26	69	67	78	52	377
Hong Kong	5494	94	80	75	77	5,8	14641
India	283	26	59	58	52	79	275
Indonesia	102	29	65	61	77	68	681
Japan	330	77	82	76	99	4,4	19860
Malaysia	58	43	72	66	78	25,6	2995
North Korea	189	60	73	67	99	27,7	1000
Pakistan	143	32	58	57	35	101	406
Philippines	221	43	68	63	90	51	867
South Korea	447	72	74	68	96	21,7	6627
Singapore	4456	100	79	73	88	5,7	14990
Taiwan	582	71	78	72	91	5,1	7055
Thailand	115	22	72	65	93	37	1800
Vietnam	218	20	68	63	88	46	230

Keterangan:

X1	: DENSITY	X5	: LITERACY
X2	: URBAN	X6	: BABYMORT
X3	: LIFEEXPF	X7	: GDP CAP
X4	: LIFEEXPM		

1. Tentukan nilai vektor rata-rata  $\bar{x}$

802,4706   44,76471   67,41176   63,47059   72,94118   53,88235   4263

2. Tentukan nilai matriks ragam-peragam (*variance-covariance*)  $S$

2541171,515	32032,81	7505,544	6615,327	5596,717	-30378,8	6496493,875
32032,80515	815,4412	248,5404	204,1801	420,6728	-1016,39	154273,8125
7505,544118	248,5404	118,5074	94,10662	245,0257	-499,011	48464,5625
6615,327206	204,1801	94,10662	76,13971	186,4669	-393,766	40633
5596,716912	420,6728	245,0257	186,4669	633,4338	-1039,24	70965,125
-30378,77868	-1016,39	-499,011	-393,766	-1039,24	2156,8	-189249,55
6496493,875	154273,8	48464,56	40633	70965,13	-189250	39577266

3. Tentukan nilai jarak Mahalanobis setiap titik pengamatan dengan vektor rata-rata dari  $d_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})'S^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})$  dimana  $i = 1, 2, \dots, n$

<b>X1</b>	<b>X2</b>	<b>X3</b>	<b>X4</b>	<b>X5</b>	<b>X6</b>	<b>X7</b>	<b>di^2</b>
25	18	44	45	29	168	205	11,01081
800	16	53	53	35	106	202	5,499957
55	12	52	50	35	112	260	6,388696
124	26	69	67	78	52	377	9,519827
5494	94	80	75	77	5,8	14641	9,718033
283	26	59	58	52	79	275	5,289502
102	29	65	61	77	68	681	1,506263
330	77	82	76	99	4,4	19860	13,88204
58	43	72	66	78	25,6	2995	4,697955
189	60	73	67	99	27,7	1000	5,04582
143	32	58	57	35	101	406	9,74309
221	43	68	63	90	51	867	2,568987
447	72	74	68	96	21,7	6627	4,467219
4456	100	79	73	88	5,7	14990	6,387216
582	71	78	72	91	5,1	7055	3,667225
115	22	72	65	93	37	1800	8,149869
218	20	68	63	88	46	230	4,45749

4. Urutkan nilai  $d_i^2$  dari terkecil hingga terbesar  $d_{(1)}^2 \leq d_{(2)}^2 \leq d_{(3)}^2 \leq \dots d_{(n)}^2$

<b>X1</b>	<b>X2</b>	<b>X3</b>	<b>X4</b>	<b>X5</b>	<b>X6</b>	<b>X7</b>	<b>di^2</b>	<b>K*</b>
25	18	44	45	29	168	205	11,01081	16
800	16	53	53	35	106	202	5,499957	9
55	12	52	50	35	112	260	6,388696	11
124	26	69	67	78	52	377	9,519827	13
5494	94	80	75	77	5,8	14641	9,718033	14
283	26	59	58	52	79	275	5,289502	8
102	29	65	61	77	68	681	1,506263	1
330	77	82	76	99	4,4	19860	13,88204	17
58	43	72	66	78	25,6	2995	4,697955	6
189	60	73	67	99	27,7	1000	5,04582	7
143	32	58	57	35	101	406	9,74309	15
221	43	68	63	90	51	867	2,568987	2
447	72	74	68	96	21,7	6627	4,467219	5
4456	100	79	73	88	5,7	14990	6,387216	10
582	71	78	72	91	5,1	7055	3,667225	3
115	22	72	65	93	37	1800	8,149869	12
218	20	68	63	88	46	230	4,45749	4

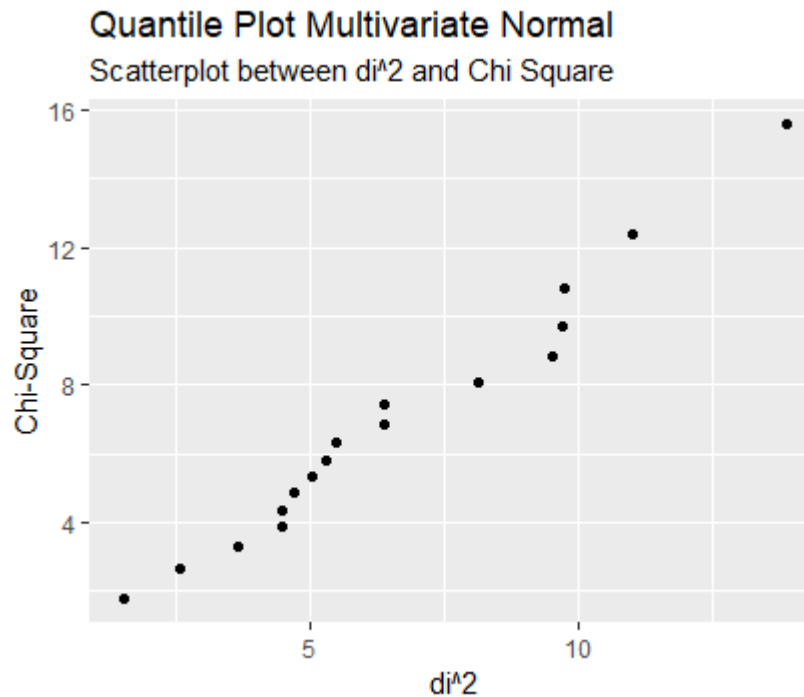
5. Tentukan nilai  $p_i = \frac{i-0,5}{n}, i = 1,2, \dots n$

X1	X2	X3	X4	X5	X6	X7	di^2	K*	(k-0,5)/n
25	18	44	45	29	168	205	11,01081	16	0,911765
800	16	53	53	35	106	202	5,499957	9	0,5
55	12	52	50	35	112	260	6,388696	11	0,617647
124	26	69	67	78	52	377	9,519827	13	0,735294
5494	94	80	75	77	5,8	14641	9,718033	14	0,794118
283	26	59	58	52	79	275	5,289502	8	0,441176
102	29	65	61	77	68	681	1,506263	1	0,029412
330	77	82	76	99	4,4	19860	13,88204	17	0,970588
58	43	72	66	78	25,6	2995	4,697955	6	0,323529
189	60	73	67	99	27,7	1000	5,04582	7	0,382353
143	32	58	57	35	101	406	9,74309	15	0,852941
221	43	68	63	90	51	867	2,568987	2	0,088235
447	72	74	68	96	21,7	6627	4,467219	5	0,264706
4456	100	79	73	88	5,7	14990	6,387216	10	0,558824
582	71	78	72	91	5,1	7055	3,667225	3	0,147059
115	22	72	65	93	37	1800	8,149869	12	0,676471
218	20	68	63	88	46	230	4,45749	4	0,205882

6. Tentukan nilai  $q_i$  sedemikian sehingga  $p_i = \int_{-\infty}^{q_i} f(\chi^2)d\chi^2$

X1	X2	X3	X4	X5	X6	X7	di^2	K*	(k-0,5)/n	Chi-Square
25	18	44	45	29	168	205	11,01081	16	0,911765	12,39703
800	16	53	53	35	106	202	5,499957	9	0,5	6,345811
55	12	52	50	35	112	260	6,388696	11	0,617647	7,462588
124	26	69	67	78	52	377	9,519827	13	0,735294	8,835197
5494	94	80	75	77	5,8	14641	9,718033	14	0,794118	9,705556
283	26	59	58	52	79	275	5,289502	8	0,441176	5,837697
102	29	65	61	77	68	681	1,506263	1	0,029412	1,789059
330	77	82	76	99	4,4	19860	13,88204	17	0,970588	15,56405
58	43	72	66	78	25,6	2995	4,697955	6	0,323529	4,864698
189	60	73	67	99	27,7	1000	5,04582	7	0,382353	5,34741
143	32	58	57	35	101	406	9,74309	15	0,852941	10,81142
221	43	68	63	90	51	867	2,568987	2	0,088235	2,694362
447	72	74	68	96	21,7	6627	4,467219	5	0,264706	4,378501
4456	100	79	73	88	5,7	14990	6,387216	10	0,558824	6,882871
582	71	78	72	91	5,1	7055	3,667225	3	0,147059	3,329498
115	22	72	65	93	37	1800	8,149869	12	0,676471	8,103807
218	20	68	63	88	46	230	4,45749	4	0,205882	3,874466

7. Buat *scatterplot*  $d_{(i)}^2$  dengan  $q_i$



Kesimpulan:

Quantil-quantil plot dari nilai  $d_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})'S^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})$ ,  $i = 1, 2, 3 \dots n$  cenderung membentuk garis lurus, sehingga dapat disimpulkan bahwa  $x_1, x_2, \dots, x_p$  menyebar Multivariat Normal. Berdasarkan sifat sebaran Multivariat Normal jika  $X_p \sim N_p(\mu, \Sigma)$ , maka semua anak gugus dari  $X$  juga menyebar Normal.

8. Pemeriksaan peubah  $X_1, X_2, X_3, X_4, X_5, X_6, X_7$  mengikuti sebaran Univariat Normal  
Hipotesis:

$H_0 : X_i$  mengikuti sebaran Univariat Normal

$H_1 : X_i$  tidak mengikuti sebaran Univariat Normal

**Peubah  $X_1$  (DENSITY)**

Anderson-Darling normality test

```
data: data.mvn.full$DENSITY
A = 3.8675, p-value = 4.062e-10
```

Kesimpulan:

Berdasarkan hasil uji *Kolmogorov Semirnov* di atas, diperoleh  $P\text{-value} = 4.062e-10$  dengan taraf nyata yang digunakan ( $\alpha$ ) sebesar 5%, sehingga  $P\text{-value} = 4.062e-10$  lebih kecil dibandingkan  $\alpha = 0.05$  yang artinya tolak  $H_0$ . Berdasarkan hasil uji tersebut, dapat disimpulkan cukup bukti untuk menyatakan bahwa  $X_1$  (DENSITY) tidak menyebar Univariat Normal.

### **Peubah X2 (URBAN)**

Anderson-Darling normality test

```
data: data.mvn.full$URBAN
A = 0.76895, p-value = 0.03645
```

**Kesimpulan:**

Berdasarkan hasil uji *Anderson-Darling* di atas, diperoleh  $P\text{-value} = 0.03645$  dengan taraf nyata yang digunakan ( $\alpha$ ) sebesar 5%, sehingga  $P\text{-value} = 0.03645$  lebih kecil dibandingkan  $\alpha = 0.05$  yang artinya tolak  $H_0$ . Berdasarkan hasil uji tersebut, dapat disimpulkan cukup bukti untuk menyatakan bahwa  $X_2$  (URBAN) tidak menyebar Univariat Normal.

### **Peubah X3 (LIFEEXPF)**

Anderson-Darling normality test

```
data: data.mvn.full$LIFEEXPF
A = 0.39617, p-value = 0.3312
```

**Kesimpulan:**

Berdasarkan hasil uji *Anderson Darling* di atas, diperoleh  $P\text{-value} = 0.3312$  dengan taraf nyata yang digunakan ( $\alpha$ ) sebesar 5%, sehingga  $P\text{-value} = 0.3312$  lebih besar dibandingkan  $\alpha = 0.05$  yang artinya tak tolak  $H_0$ . Berdasarkan hasil uji tersebut, dapat disimpulkan tidak cukup bukti untuk menyatakan bahwa  $X_3$  (LIFEEXPF) tidak menyebar Univariat Normal. Sehingga dapat dikatakan bahwa peubah  $X_3$  (LIFEEXPF) menyebar Univariat Normal.

### **Peubah X4 (LIFEEXPM)**

Anderson-Darling normality test

```
data: data.mvn.full$LIFEEXPM
A = 0.24197, p-value = 0.7297
```

**Kesimpulan:**

Berdasarkan hasil uji *Anderson Darling* di atas, diperoleh  $P\text{-value} = 0.7297$  dengan taraf nyata yang digunakan ( $\alpha$ ) sebesar 5%, sehingga  $P\text{-value} = 0.7297$  lebih besar dibandingkan  $\alpha = 0.05$  yang artinya tak tolak  $H_0$ . Berdasarkan hasil uji tersebut, dapat disimpulkan tidak cukup bukti untuk menyatakan bahwa  $X_4$  (LIFEEXPM) tidak menyebar Univariat Normal. Sehingga dapat dikatakan bahwa peubah  $X_4$  (LIFEEXPM) menyebar Univariat Normal.

### **Peubah X5 (LITERACY)**

Anderson-Darling normality test

```
data: data.mvn.full$LITERACY
A = 1.2646, p-value = 0.001897
```

Kesimpulan:

Berdasarkan hasil uji *Anderson Darling* di atas, diperoleh  $P\text{-value} = 0.001897$  dengan taraf nyata yang digunakan ( $\alpha$ ) sebesar 5%, sehingga  $P\text{-value} = 0.001897$  lebih kecil dibandingkan  $\alpha = 0.05$  yang artinya tolak  $H_0$ . Berdasarkan hasil uji tersebut, dapat disimpulkan cukup bukti untuk menyatakan bahwa  $X_5$  (LITERACY) tidak menyebar Univariat Normal.

### **Peubah $X_6$ (BABYMORT)**

Anderson-Darling normality test

```
data: data.mvn.full$BABYMORT
A = 0.54077, p-value = 0.1405
```

Kesimpulan:

Berdasarkan hasil uji *Anderson Darling* di atas, diperoleh  $P\text{-value} = 0.1405$  dengan taraf nyata yang digunakan ( $\alpha$ ) sebesar 5%, sehingga  $P\text{-value} = 0.1405$  lebih besar dibandingkan  $\alpha = 0.05$  yang artinya tak tolak  $H_0$ . Berdasarkan hasil uji tersebut, dapat disimpulkan tidak cukup bukti untuk menyatakan bahwa  $X_6$  (BABYMORT) tidak menyebar Univariat Normal. Sehingga dapat dikatakan bahwa peubah  $X_6$  (BABYMORT) menyebar Univariat Normal.

### **Peubah $X_7$ (GDP.CAP)**

Anderson-Darling normality test

```
data: data.mvn.full$GDP.CAP
A = 2.2749, p-value = 4.732e-06
```

Kesimpulan:

Berdasarkan hasil uji *Anderson Darling* di atas, diperoleh  $P\text{-value} = 4.732e-06$  dengan taraf nyata yang digunakan ( $\alpha$ ) sebesar 5%, sehingga  $P\text{-value} = 4.732e-06$  lebih kecil dibandingkan  $\alpha = 0.05$  yang artinya tolak  $H_0$ . Berdasarkan hasil uji tersebut, dapat disimpulkan cukup bukti untuk menyatakan bahwa  $X_7$  (GDP.CAP) tidak menyebar Univariat Normal.