



**PENGUJIAN DISTRIBUSI MULTIVARIAT NORMAL DAN VEKTOR RATAAN PADA JUMLAH PENDUDUK MENURUT JENIS KELAMIN, TINGKAT PENDIDIKAN DAN KEMISKINAN DI PROVINSI KALIMANTAN SELATAN DAN KALIMANTAN BARAT**

Oleh

**Gieta Sonya<sup>1)</sup>, Nur Azmi Khairinda<sup>2)</sup> & Nur Salam<sup>3)</sup>**

<sup>1,2,3</sup>Universitas Lambung Mangkurat

Email: [gietasn@gmail.com](mailto:gietasn@gmail.com), [azminurnur12@gmail.com](mailto:azminurnur12@gmail.com) & [nursalam2011@gmail.com](mailto:nursalam2011@gmail.com)

**Abstrak**

Penelitian ini dilakukan pada 2 (dua) Provinsi yaitu Kalimantan Selatan yang memiliki 13 Kabupaten/Kota dan Kalimantan Barat yang memiliki 14 Kabupaten/Kota. Pada artikel ini akan dilakukan pendekatan statistik menggunakan uji asumsi distribusi multivariat normal yakni Uji T<sup>2</sup>-Hotteling untuk data dua populasi serta menguji vektor mean untuk satu populasi dalam melihat perbandingan pola demografi kedua provinsi tersebut. Hasil analisis dengan pengujian asumsi normal multivariat menggunakan pendekatan plot *chi-square*, dan uji korelasi memberikan kesimpulan bahwa data jumlah penduduk menurut tingkat pendidikan di bawah SMA, jumlah penduduk miskin, jenis kelamin laki-laki dan perempuan di kedua provinsi berdistribusi normal multivariat dan pendekatan uji T<sup>2</sup>-Hotteling menyimpulkan bahwa terdapat perbedaan rata-rata dari variabel yang mempengaruhi jumlah penduduk di kedua provinsi.

**Kata Kunci : Normal Multivariat, Vektor Mean & Penduduk**

**PENDAHUALUAN**

Pemakaian metode statistik dalam membantu mengetahui tentang jumlah penduduk, pendidikan dan kemiskinan. Salah satu metode statistika yang digunakan adalah statistika multivariat.

Statistik multivariat dipergunakan apabila variabel penelitian terdiri dari dua variate atau lebih, dan antar variabel tersebut akan diteliti korelasi, perbedaan, pengaruh antara dua variabel atau lebih. Salah satu uji statistik multivariat yang sering digunakan adalah Uji Hotelling's T<sup>2</sup>. Uji Hotelling's T<sup>2</sup> berfungsi untuk melihat perbedaan antara dua kelompok percobaan, yang masing-masing kelompok terdiri dari dua variate atau lebih, dan akan dilakukan analisis statistik pada variate tersebut secara serentak.

Pendidikan merupakan hal terpenting dan salah satu tanggung jawab yang utama bagi bangsa. Pendidikan merupakan kebutuhan dasar untuk mengembangkan potensi sumber daya manusia.

Permasalahan kemiskinan merupakan salah satu persoalan yang mendasar yang dihadapi oleh

negara-negara berkembang tidak terkecuali Indonesia. Tingginya tingkat kemiskinan pada suatu negara mencerminkan rendahnya tingkat kesejahteraan masyarakat pada daerah tersebut. Fenomena kemiskinan telah berlangsung lama, akan tetapi sampai saat ini pemerintah masih belum bisa mengatasi masalah yang menjadi dasar bagi pembangunan negara.

Tingkat kemiskinan di Kalimantan Selatan hingga Maret 2018 turun menjadi 4,54% dengan jumlah penduduk miskin sebesar 189 ribu lebih orang. Jika dibandingkan dengan periode sama di tahun sebelumnya, jumlah penduduk miskin di Kalimantan Selatan mencapai 193 ribu lebih orang.

Pada penelitian ini, akan diuji asumsi distribusi multivariat normal satu populasi serta menguji nilai vektor Mean satu populasi untuk 2 populasi menggunakan pengamatan Uji T<sup>2</sup>-Hotteling. Variabel penelitian yang digunakan dalam penelitian ini adalah Persentase jumlah penduduk dengan tingkat pendidikan di bawah SMA ( $X_1$ ), Persentase jumlah penduduk



miskin( $X_2$ ), Persentase jenis kelamin laki-laki( $X_3$ ), Persentase jenis kelamin perempuan( $X_4$ )

## LANDASAN TEORI

### Analisis Multivariat

Analisis Multivariat Analisis multivariat merupakan salah satu teknik statistik yang digunakan untuk memahami struktur data dalam dimensi tinggi. Variabel-variabel itu saling terkait satu sama lain. Pendekatan multivariat melibatkan multivariabel tetapi tidak sebaliknya. Multivariabel yang saling berkorelasi yang dikatakan multivariat.

Analisis multivariat merupakan analisis lanjutan dari analisis univariat maupun bivariat. Menurut Santoso (2004), analisis multivariat dapat didefinisikan secara sederhana sebagai metode pengolahan variabel dalam jumlah banyak untuk mencari pengaruhnya terhadap suatu objek secara simultan. Metode memungkinkan untuk menganalisis lebih dari satu variabel dependen untuk mengetahui apakah rata-rata dua populasi berbeda secara signifikan.

### Uji Asumsi Normal Multivariat

Variabel  $X_1, X_2, \dots, X_p$  dikatakan berdistribusi normal multivariat dengan parameter  $\mu$  dan  $\Sigma$  jika mempunyai probability density function :

$$f(X_1, X_2, \dots, X_p) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{p/2}} e^{-\frac{1}{2}(X-\mu)^T \Sigma^{-1}(X-\mu)}$$

Jika  $X_1, X_2, \dots, X_p$  berdistribusi normal multivariat maka  $(X - \mu)^T \Sigma^{-1}(X - \mu)$  berdistribusi  $X_p^2$ . Berdasarkan sifat ini maka pemeriksaan distribusi normal multivariat dapat dilakukan dengan cara membuat  $Q-Q$  plot dari nilai  $d_j^2 = (X_i - \bar{X})^T S^{-1}(X_i - \bar{X})$ ;  $i = 1, 2, \dots, n$

Tahapan dari pembuatan  $Q-Q$  plot ini adalah sebagai berikut :

1. Tentukan nilai vektor rata-rata :  $\bar{X}$
2. Tentukan nilai matriks varians-kovarians :  $S$
3. Tentukan nilai jarak Mahalanobis setiap titik pengamatan dengan vektor rata-ratanya ( $d_j^2$ ).
4. Urutkan nilai  $d_j^2$  dari kecil ke besar:  $d_{(1)}^2 \leq d_{(2)}^2 \leq d_{(3)}^2 \leq d_{(4)}^2 \leq d_{(5)}^2 \leq \dots \leq d_{(n)}^2$ .
5. Tentukan nilai  $p_i = \frac{i-0.5}{n}$ ;  $i = 1, 2, 3, \dots, n$

6. Tentukan nilai  $q_i$  sedemikian hingga  $\int_{-\infty}^{q_i} f(x^2) dx^2 = p_i$

7. Buat scatter-plot  $d_j^2$  dengan  $q_i$ .

8. Jika scatter-plot ini cenderung membentuk garis lurus dan lebih dari 50% nilai  $d_j^2 \leq X_{p(0.5)}^2$  maka  $H_0$  diterima artinya data berdistribusi normal multivariat.

Pengujian asumsi normal multivariat juga dapat dilakukan secara inferensia, yaitu dilakukan dengan pengujian koefisien korelasi dengan hipotesis sebagai berikut:

$H_0$  : Data berdistribusi normal multivariat.

$H_1$  : Data tidak berdistribusi normal multivariat.

Statistik Uji :

$$r_Q = \frac{\sum_{j=1}^n (x_j - \bar{x})(q_j - \bar{q})}{\sqrt{\sum_{j=1}^n (x_j - \bar{x})^2} \sqrt{\sum_{j=1}^n (q_j - \bar{q})^2}}$$

Daerah kritis : tolak  $H_0$  jika  $r_Q > \text{critical point}$

### Pengujian Vektor Mean Satu Populasi

Uji  $T^2$ -Hotteling digunakan untuk menguji mean vektor dari beberapa variabel yang bersifat multivariat.  $T^2$ -Hotteling merupakan salah satu metode yang digunakan untuk mengontrol pergeseran mean proses untuk kasus multivariat berdasarkan pengamatan individual. Selain pergeseran mean proses, pergeseran variabilitas proses juga perlu dikontrol.

1. Uji Vektor Mean Satu Populasi

Hipotesis yang digunakan :

$H_0$  :  $\mu = \mu_0$

$H_1$  :  $\mu \neq \mu_0$

Statistik Uji :

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

dimana :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$S = \frac{1}{n} \sum_{i=1}^n (\bar{X} - \mu_0)(\bar{X} - \mu_0)^T$$

Keterangan :

$n$  : banyaknya data

$x$  : vektor rata-rata dari sampel  $x$

$\mu_0$  : vektor rata-rata dari populasi

$S$  : matriks kovarian

$x_i$  : nilai data

Daerah kritis :

Tolak  $H_0$  jika  $T^2\text{-Hotteling} > \frac{(n-1)p}{(n-p)} F_{p, n-p, \alpha}$



### Uji Hipotesis Vektor Mean Satu Populasi

Misalkan  $X_1, X_2, \dots, X_n$  merupakan sampel random dari populasi normal  $N_p(\mu, \Sigma)$ . Akan diperlihatkan pengujian  $H_0 : \mu = \mu_0$  dibandingkan dengan  $H_1 : \mu \neq \mu_0$ , dimana  $\mu_0$  adalah sebuah vektor yang diketahui. Uji hipotesis dilakukan dengan langkah-langkah sebagai berikut:

Hipotesis:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

Taraf Signifikansi:  $\alpha$

Statistik uji:

$$T^2 > (S - \mu_0)' \left(\frac{S}{n}\right)^{-1} (S - \mu_0)$$

Kriteria uji:

$H_0$  ditolak jika  $T^2 > \frac{(n-1)}{n-p} F_{p, n-p(\alpha)}$  atau jika p-value  $< \alpha$ . Sementara itu kriteria uji untuk n besar, yaitu  $H_0$  ditolak jika  $T^2 > X_p^2(\alpha)$ .

## METODE PENELITIAN

### Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari Badan Pusat Statistika Provinsi Kalimantan Selatan yang digunakan berupa data persentase jenis kelamin, tingkat pendidikan dan kemiskinan dengan jumlah pengamatan 13 kabupaten/kota pada tahun 2018. Variabel-variabel yang digunakan pada penelitian inidisajikan pada Tabel 1 berikut.

**Tabel 1. Variabel Penelitian yang Digunakan**

Notasi	Variabel
Populasi 1	Kalimantan Selatan
Populasi 2	Kalimantan Barat
$X_1$	Persentase jumlah penduduk dengan tingkat pendidikan di bawah SMA (%)
$X_2$	Persentase jumlah penduduk miskin(%)
$X_3$	Persentase jenis kelamin laki-laki(%)
$X_4$	Persentase jenis kelamin perempuan(%)

### Metode Analisis Data

Penelitian ini, akan diuji asumsi distribusi multivariat normal satu populasi serta menguji nilai vektor Mean satu populasi untuk populasi 1 menggunakan pengamatan Uji  $T^2$ -Hotteling.

### Langkah Analisis

1. Melakukan uji normalitas multivariat, jika tidak memenuhi asumsi distribusi normal multivariat, maka data diasumsikan berdistribusi normal multivariat dan homogen sehingga dapat dilakukan pengujian vector meansatu populasi.
2. Melakukan analisis statistika deskriptif pada persentase data kependudukan di Provinsi Kalimantan Selatan menurut jenis kelamin, tingkat pendidikan dan kemiskinan.
3. Melakukan uji vektor mean satu populasi yang diambil dari data kependudukan di Provinsi Kalimantan Selatan dan Kalimantan Barat menurut persentase jumlah penduduk menurut persentase tingkat pendidikan di bawah SMA, persentase jumlah penduduk miskin, persentase jenis kelamin laki-laki, persentase jenis kelamin perempuan.

## HASIL DAN PEMBAHASAN

### Statistika Deskriptif

Secara deskriptif data Persentase jumlah penduduk di Provinsi Kalimantan Selatan dan Kalimantan Barat tahun 2018 menurut tingkat pendidikan di bawah SMA ( $X_1$ ), Persentase jumlah penduduk miskin ( $X_2$ ), Persentase jenis kelamin laki-laki ( $X_3$ ), Persentase jenis kelamin perempuan ( $X_4$ ) disajikan dalam tabel berikut.

**Tabel 2. Deskriptif Variabel (a) Provinsi Kalimantan Selatan; (b) Provinsi Kalimantan Barat**

	Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Penduduk_Miskin	13	3,34	7,76	6,0538	1,25128	1,566
Pendidikan_dibawah_SM A	13	2,93	13,64	7,6931	2,53336	6,418
Perempuan	13	47,75	51,08	49,3823	,95693	,916
Laki_laki	13	48,91	63,12	51,4462	3,60241	12,977
Valid N (listwise)	13					

(a)



Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Penduduk_Miskin	14	11,13	54,86	27,6479	14,28295	204,003
Pendidikan_dibawah_SMA	14	1,50	11,64	7,0321	2,55792	6,543
Perempuan	14	47,71	50,69	48,8264	,87085	,758
Laki_laki	14	49,30	52,28	51,3157	,84514	,714
Valid N (listwise)	14					

(b)

Dari Tabel 5a dan 5b di atas dapat diketahui bahwa untuk variabel persentase jumlah penduduk miskin, Provinsi Kalimantan Barat rata-rata lebih tinggi dibandingkan dengan Provinsi Kalimantan Selatan, dan juga memiliki standar deviasi yang lebih tinggi itu berarti. Sedangkan, untuk variabel persentase pendidikan di bawah SMA Kalimantan Barat rata-rata lebih tinggi dibandingkan dengan Provinsi Kalimantan Selatan, dan juga memiliki standar deviasi yang lebih rendah untuk variabel jenis kelamin perempuan di Provinsi Kalimantan Selatan rata-rata lebih tinggi dibandingkan dengan Provinsi Kalimantan Barat, dan juga memiliki standar deviasi yang lebih tinggi. Sedangkan untuk variabel jenis kelamin laki-laki Provinsi Kalimantan Selatan memiliki nilai rata-rata lebih tinggi dibandingkan dengan Provinsi Kalimantan Barat, dan juga memiliki standar deviasi yang lebih tinggi.

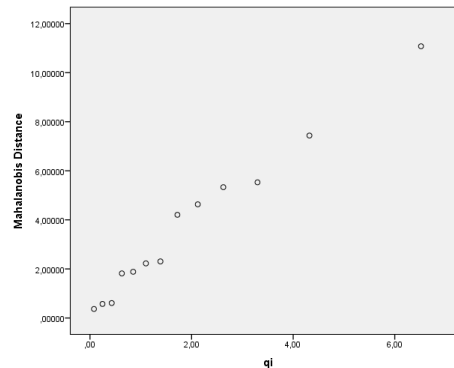
**Pemeriksaan Normal Multivariat**

Untuk dapat dilanjutkan pada analisis Hotelling’s  $T^2$  (uji vektor mean) maka data mengikuti distribusi normal multivariat (diasumsikan). Data jumlah penduduk menurut tingkat pendidikan di Provinsi Kalimantan Selatan dan Kalimantan Barat Tahun 2018, selanjutnya akan dianalisis apakah data tersebut berasal dari populasi yang berdistribusi normal multivariat atau tidak. Pengujian dilakukan dengan

1. Plot Chi-square

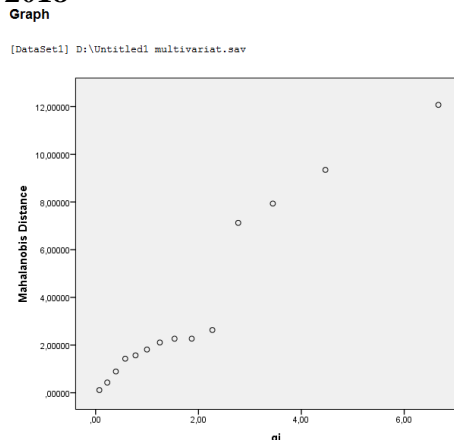
Pemeriksaan distribusi normal multivariat juga dapat dilakukan secara visual dengan melihat pola sebaran data pada scatterplot antara nilai Square Distance dengan  $d^2_j$  dengan nilai

$q_{c,p}$  untuk output  $q_{c,p}$  dapat dilihat pada lampiran 3. Berikut ini adalah hasil pengujian distribusi normal multivariat menggunakan scatterplot pada data jumlah penduduk menurut tingkat pendidikan di bawah SMA, persentase jumlah penduduk miskin, persentase jenis kelamin laki-laki, Persentase jenis kelamin perempuan di Provinsi Kalimantan Selatan dan Kalimantan Barat Tahun 2018.



Gambar 1. Ouput Scatterplot pada data jumlah penduduk di Kalimantan Selatan Tahun 2018 Berdasarkan Gambar 1 terlihat bahwa titik-titik pada scatterplot mengikuti garis normal (membentuk garis lurus), sehingga dapat disimpulkan bahwa data persentase jumlah penduduk menurut persentase tingkat pendidikan di bawah SMA, persentase jumlah penduduk miskin, persentase jenis kelamin laki-laki, Persentase jenis kelamin perempuan di Provinsi Kalimantan Selatan Tahun 2018 berdistribusi normal multivariat.

**Gambar 2. Ouput Scatterplot pada data jumlah penduduk di Kalimantan Barat Tahun 2018**





Berdasarkan Gambar 2 di atas terlihat bahwa titik-titik pada scatterplot mengikuti garis normal (membentuk garis lurus), sehingga dapat disimpulkan bahwa data jumlah penduduk menurut persentase tingkat pendidikan di bawah SMA, persentase jumlah penduduk miskin, persentase jenis kelamin laki-laki, Persentase jenis kelamin perempuan di Provinsi Kalimantan Barat Tahun 2018 berdistribusi multivariat normal.

## 2. Uji Koefisien Korelasi

Uji korelasi digunakan untuk menguji tingkat signifikansi dengan cara mengkorelasikan nilai Square Distance dengan nilai  $d^2_j$  dengan nilai  $q_{c,p}$

Hipotesis :

$H_0$ : Data jumlah penduduk menurut tingkat pendidikan, jumlah penduduk miskin, jenis kelamin laki-laki, jenis kelamin perempuan di Provinsi Kalimantan Selatan Tahun 2018 mengikuti distribusi normal multivariat.

$H_1$ : Data jumlah penduduk menurut tingkat pendidikan, jumlah penduduk miskin, jenis kelamin laki-laki, jenis kelamin perempuan di Provinsi Kalimantan Selatan Tahun 2018 tidak mengikuti distribusi normal multivariat.

### Gambar 3 Koefisien Korelas

		Mahalanobis Distance	qi
Mahalanobis Distance	Pearson Correlation	1	,989**
	Sig. (2-tailed)		,000
	N	13	13
qi	Pearson Correlation	,989**	1
	Sig. (2-tailed)	,000	
	N	13	13

\*\* Correlation is significant at the 0.01 level (2-tailed).

Dari Gambar 3 di atas dapat di lihat bahwa koefisien korelasi yang diperoleh 0,989 menunjukkan koefisien korelasi yang sangat tinggi. Besarnya koefisien korelasi antara -1 sampai dengan +1. Apabila koefisien korelasi  $> r$  tabel atau nilai sig.  $< 0,05$  maka terdapat korelasi yang signifikan. Dalam scatter-plot ini berarti

data berasal dari sampel yang mengikuti distribusi normal multivariat.

Hipotesis :

$H_0$ : Data jumlah penduduk menurut tingkat pendidikan, jumlah penduduk miskin, jenis kelamin laki-laki, jenis kelamin perempuan di Provinsi Kalimantan Barat Tahun 2018 mengikuti distribusi normal multivariat.

$H_1$ : Data jumlah penduduk menurut tingkat pendidikan, jumlah penduduk miskin, jenis kelamin laki-laki, jenis kelamin perempuan di Provinsi Kalimantan Barat Tahun 2018 tidak mengikuti distribusi multivariat normal.

### Gambar 4 Koefisien Korelas

Correlations

		Mahalanobis Distance	qi
Mahalanobis Distance	Pearson Correlation	1	,968**
	Sig. (2-tailed)		,000
	N	14	14
qi	Pearson Correlation	,968**	1
	Sig. (2-tailed)	,000	
	N	14	14

\*\* Correlation is significant at the 0.01 level (2-tailed).

Dari Gambar 4 di atas dapat di lihat bahwa koefisien korelasi yang diperoleh 0,968 menunjukkan koefisien korelasi yang sangat tinggi. Besarnya koefisien korelasi antara -1 sampai dengan +1. Apabila koefisien korelasi  $> r$  tabel atau nilai sig.  $< 0,05$  maka terdapat korelasi yang signifikan. Dalam scatter-plot ini berarti data berasal dari sampel yang mengikuti distribusi normal multivariat.

### Pengujian Vektor Mean Satu Populasi

Pengujian vektor mean satu populasi akan dilakukan pada data jumlah penduduk di Provinsi Kalimantan Selatan Tahun 2018, dengan variabel yang meliputi persentase tingkat pendidikan di bawah SMA ( $X_1$ ), persentase jumlah penduduk miskin ( $X_2$ ), persentase jenis kelamin laki-laki ( $X_3$ ), persentase jenis kelamin perempuan ( $X_4$ )

Hipotesis:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

Dengan menggunakan taraf signifikan 5% , maka dapat diperoleh daerah kritis atau  $H_0$  di tolak  $T^2$



$> c^2$  9,24732 berdasarkan hitungan nilai  $T^2$  diperoleh 73,6729, sehingga dapat disimpulkan bahwa minimal ada satu rata-rata dari variabel-variabel yang mempengaruhi jumlah penduduk di Provinsi Kalimantan Selatan .

## PENUTUP

### Kesimpulan

Pengujian asumsi distribusi multivariat normal dengan pendekatan secara visual dari plot Chi-Square, dan uji korelasi memberikan kesimpulan bahwa data persentase jumlah penduduk menurut tingkat pendidikan di bawah SMA, persentase jumlah penduduk miskin, persentase jenis kelamin laki-laki, persentase jenis kelamin perempuan di Provinsi Kalimantan Selatan dan Kalimantan Barat Tahun 2018 berdistribusi Normal Multivariat.

Berdasarkan uji T2-Hotelling untuk variabel persentase jumlah penduduk menurut tingkat pendidikan di bawah SMA , persentase jumlah penduduk miskin, persentase jenis kelamin laki-laki, persentase jenis kelamin perempuan dengan mean dari populasi dapat disimpulkan bahwa minimal ada satu rata-rata dari variabel-variabel yang mempengaruhi jumlah penduduk di Provinsi Kalimantan Selatan

## DAFTAR PUSTAKA

- [1] Badan Pusat Statistik. (2008). Indeks Pembangunan Manusia 2006 - 2007. Badan Pusat Statistik. Jakarta.
- [2] Johnson, R. A., & Winchurn, D. W. (2007). Applied Multivariat Statistical Analysis : Sixth Edition. New Jersey : Pearson Prentice Hall
- [3] Marfai, M.A., King, L., Sartohadi, J., Sudrajat, S., Budiani, S.R., and Yulianto, F, (2008). The impact of tidal flooding on a coastal community in Semarang, Indonesia. *Environmentalist*, 28: p. 237-248.
- [4] Perraudeau, M., (1988). Luminance models. In *National Lighting Conference*. Cambridge, UK, March 27-30.
- [5] International Daylight Monitoring Programme, [Online], Diakses di: <http://idmp.entpe.fr/> [16 Juni 2008].