

Isma Muthahharah, S.Si., M.Si.
Zakiyah Mar'ah, S.Si., M.Si.



STATISTIKA MULTIVARIAT

APLIKASI SPSS DAN RSTUDIO



Editor:
Andika Isma, S.Pd., M.M

STATISTIKA MULTIVARIAT APLIKASI SPSS DAN RSTUDIO

Isma Muthahharah, S.Si., M.Si
Zakiyah Mar'ah, S.Si., M.Si



TAHTA MEDIA GROUP

UU No 28 tahun 2014 tentang Hak Cipta

Fungsi dan sifat hak cipta Pasal 4

Hak Cipta sebagaimana dimaksud dalam Pasal 3 huruf a merupakan hak eksklusif yang terdiri atas hak moral dan hak ekonomi.

Pembatasan Pelindungan Pasal 26

Ketentuan sebagaimana dimaksud dalam Pasal 23, Pasal 24, dan Pasal 25 tidak berlaku terhadap:

- i. penggunaan kutipan singkat Ciptaan dan/atau produk Hak Terkait untuk pelaporan peristiwa aktual yang ditujukan hanya untuk keperluan penyediaan informasi aktual;
- ii. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk kepentingan penelitian ilmu pengetahuan;
- iii. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk keperluan pengajaran, kecuali pertunjukan dan Fonogram yang telah dilakukan Pengumuman sebagai bahan ajar; dan
- iv. penggunaan untuk kepentingan pendidikan dan pengembangan ilmu pengetahuan yang memungkinkan suatu Ciptaan dan/atau produk Hak Terkait dapat digunakan tanpa izin Pelaku Pertunjukan, Produser Fonogram, atau Lembaga Penyiaran.

Sanksi Pelanggaran Pasal 113

1. Setiap Orang yang dengan tanpa hak melakukan pelanggaran hak ekonomi sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf i untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 1 (satu) tahun dan/atau pidana denda paling banyak Rp100.000.000 (seratus juta rupiah).
2. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf c, huruf d, huruf f, dan/atau huruf h untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 3 (tiga) tahun dan/atau pidana denda paling banyak Rp500.000.000,00 (lima ratus juta rupiah).

STATISTIKA MULTIVARIAT APLIKASI SPSS DAN RSTUDIO

Penulis:

Isma Muthahharah, S.Si., M.Si

Zakiah Mar'ah, S.Si., M.Si

Desain Cover:

Tahta Media

Editor:

Andika Isma, S.Pd., M.M

Proofreader:

Tahta Media

Ukuran:

vi, 129, Uk: 15,5 x 23 cm

ISBN: 978-623-147-341-7

Cetakan Pertama:

Maret 2024

Hak Cipta 2024, Pada Penulis

Isi diluar tanggung jawab percetakan

Copyright © 2024 by Tahta Media Group

All Right Reserved

Hak cipta dilindungi undang-undang

Dilarang keras menerjemahkan, memfotokopi, atau
memperbanyak sebagian atau seluruh isi buku ini
tanpa izin tertulis dari Penerbit.

PENERBIT TAHTA MEDIA GROUP
(Grup Penerbitan CV TAHTA MEDIA GROUP)
Anggota IKAPI (216/JTE/2021)

KATA PENGANTAR

Puji syukur kehadiran Allah SWT dan junjungan Nabi besar Muhammad SAW. Atas rahmat dan karunianya sehingga penulis bisa menyusun buku STATISTIKA MULTIVARIAT dengan tujuan untuk memperlancar proses pengajaran. Buku ini berisi materi antara lain Konsep Dasar Multivariat, Distribusi Normal Multivariat, Analisis Regresi Lienear Berganda, Analisis Komponen Utama, Analisis Faktor, Analisis Cluster dan Analisis Diskriminan.

Kami menyadari ada banyak kesalahan dalam penyusunan buku ini. Karena itu, kami tim penyusun sangat mengharapkan masukan, kritik dan saran yang sifatnya membangun demi perbaikan ke depan. Akhir kata, dengan segenap kerendahan hati, kami berharap semoga buku ini dapat bermanfaat dan bernilai ibadah di sisi Allah SWT.

Makassar, Maret 2024

Penulis

DAFTAR ISI

KATA PENGANTAR.....	iv
DAFTAR ISI.....	v
BAB I KONSEP DASAR MULTIVARIAT.....	1
A. Pengertian Variat.....	1
B. Jenis Data Multivariat.....	3
C. Karakteristik Analisis Multivariat.....	3
D. Kegunaan Analisis Multivariat.....	4
E. Klasifikasi Teknik-Teknik Analisis Multivariat.....	4
F. Teknologi Informasi Dalam Pengolahan Data Multivariat.....	5
G. Latihan Soal.....	6
BAB II DISTRIBUSI NORMAL MULTIVARIAT.....	7
A. Konsep Distribusi Normal Multivariat.....	7
B. Sifat-Sifat Tertentu Dari Distribusi Normal.....	9
C. Distribusi Sampel Multivariat Normal dan Maximum Likelihood Estimasi.....	10
D. Uji Asumsi Untuk Data Multivariat.....	12
E. Aplikasi SPSS.....	13
F. Aplikasi RStudio.....	16
G. Latihan Soal.....	18
BAB III MODEL REGRESI LINEAR MULTIVARIAT.....	19
A. Konsep Analisis Regresi.....	19
B. Model Regresi Linear Berganda.....	19
C. Asumsi-Asumsi Model Regresi Linear Berganda.....	21
D. Estimasi Parameter Model Regresi Linear Berganda.....	21
E. Pengujian Parameter Model Regresi Linier Berganda.....	23
F. Pelanggaran - Pelanggaran Terhadap Asumsi Regresi Linier Berganda.....	25
G. Aplikasi SPSS.....	26
H. Aplikasi RStudio.....	31
I. Latihan Soal.....	37
BAB IV ANALISIS KOMPONEN UTAMA.....	38
A. Konsep Analisis Komponen Utama.....	38
B. Kegunaan Analisis Komponen Utama.....	39
C. Uji Kedekatan Hubungan Antar Variabel.....	40

D.	Penentuan Komponen Utama Berdasarkan Nilai Eigen	41
E.	Kriteria Pemilihan Komponen Utama	41
F.	Aplikasi SPSS.....	42
G.	Aplikasi RStudio	48
H.	Latihan Soal.....	53
BAB V ANALISIS FAKTOR.....		55
A.	Konsep Analisis Faktor	55
B.	Pengujian Analisis Faktor.....	56
C.	Kegunaan Analisis Faktor	57
D.	Persyaratan Analisis Faktor	57
E.	Asumsi Analisis Faktor	58
F.	Rotasi Faktor	58
G.	Aplikasi SPSS.....	59
H.	Aplikasi RStudio	69
I.	Latihan Soal.....	74
BAB VI ANALISIS CLUSTER.....		76
A.	Konsep Analisis Cluster	76
B.	Metode Pengelompokkan	76
C.	Karakteristik Analisis Cluster.....	79
D.	Tujuan Analisis Cluster	80
E.	Analisis SPSS	80
F.	Aplikasi RStudio	97
G.	Latihan Soal.....	105
BAB VII ANALISIS DISKRIMINAN		107
A.	Konsep Dasar Analisis Diskriminan.....	107
B.	Tujuan Analisis Diskriminan.....	108
C.	Asumsi Analisis Diskriminan.....	108
D.	Keputusan Dalam Pengujian	109
E.	Aplikasi SPSS.....	110
F.	Aplikasi RStudio	118
G.	Latihan Soal.....	124
DAFTAR PUSTAKA.....		126
BIODATA PENULIS.....		129

BAB I

KONSEP DASAR MULTIVARIAT

A. PENGERTIAN VARIAT

Sebelum membahas tentang variat maka terlebih dahulu mengetahui apakah itu variabel. Variabel berasal dari kata “*vary*” dan “*able*” yang berarti “berubah” dan “dapat”. Jadi, secara harfiah variabel berarti dapat berubah, sehingga setiap variabel dapat diberi nilai dan nilai itu berubah-ubah. Nilai tersebut bisa kuantitatif (terukur dan atau terhitung, dapat dinyatakan dengan angka) juga bisa kualitatif (jumlah dan derajat atributnya yang dinyatakan dengan nilai mutu). Variabel merupakan faktor penting dalam penelitian. Dalam statistika, variabel diartikan sebagai konsep, ciri, struktur atau sifat suatu benda (orang, benda, tempat, dan lain-lain) yang nilainya berbeda-beda antara satu objek dengan objek lainnya dan diidentifikasi oleh peneliti untuk dipelajari dan ditarik kesimpulan (Irwan, 2021).

Contoh variabel yang sering diteliti adalah jenis kelamin yang mempunyai dua nilai yaitu laki-laki dan perempuan, Agama merupakan suatu variabel yang mempunyai nilai-nilai, misalnya Islam, Kristen, Katolik, Hindu, dan Budha. Tinggi badan merupakan suatu variabel yang mempunyai nilai 150 cm, 165 cm, dan seterusnya. Demikian pula dengan berat badan merupakan suatu variabel, karena mempunyai nilai misalnya 45 kg, 50 kg, dan seterusnya.

Dalam konteks penelitian, variabel dapat dibagi menjadi kategori, seperti di bawah ini:

1) Variabel Independen

Variabel independen sering disebut stimulus, prediktor, anteseden atau lebih sering disebut variabel independen. Variabel bebas merupakan variabel yang mempengaruhi atau menjadi akibat perubahan atau terjadinya variabel terikat (terikat).

2) Variabel Terikat

Variabel ini sering disebut dengan variabel keluaran, kriteria, konsekuensi, disebut juga dengan istilah variabel terikat. Variabel terikatnya adalah variabel yang dipengaruhi oleh variabel bebas.

3) Variabel Moderator

Variabel moderator adalah variabel yang mempengaruhi (memperkuat atau memperlemah) hubungan antara variabel bebas dengan variabel terikat. Variabel moderator ini disebut juga dengan variabel bebas kedua.

4) Variabel Intervening

Variabel Intervening adalah variabel yang secara teoritis mempengaruhi hubungan antara variabel bebas dengan variabel terikat menjadi hubungan yang tidak langsung, tidak dapat diobservasi, dan mungkin dapat diukur. Variabel ini merupakan variabel perantara/intermiten yang terletak diantara variabel bebas dan variabel terikat, sehingga variabel independen tidak langsung mempengaruhi berubahnya atau timbulnya variabel dependen.

5) Variabel Kontrol

Variabel kontrol adalah variabel yang dikendalikan atau dibuat konstan agar faktor luar yang tidak terkendali. Variabel kontrol sering digunakan peneliti jika ingin melakukan penelitian komparatif.

Secara sederhana, variat bisa didefinisikan sebagai suatu kombinasi linear dari variabel-variabel dengan bobot variabel yang ditentukan secara empiris. Sebagai contoh ada persamaan regresi berganda

$$\text{Nilai variat} = k_1x_1 + k_2x_2 + k_3x_3 + \dots + k_nX_n \quad (1.1)$$

Di sini X_n adalah variabel yang telah ditentukan oleh peneliti, sedangkan k_n adalah hasil dari proses multivariat. Nilai variat adalah hasil dari proses perkalian dan penjumlahan x , yang menghasilkan nilai variat tertentu. Nilai variat di atas dapat dikembangkan dengan menggunakan notasi x_{jk} untuk menunjukkan nilai tertentu dari variabel ke- k yang diamati pada butir ke- j , atau percobaan. x_{jk} = pengukuran variabel ke- k pada item ke- k . Akibatnya, n pengukuran pada p variabel dapat ditampilkan sebagai berikut (Iverson & Dervan, 2014):

	Variabel 1	Variabel 2	...	Variabel k	...	Variabel p
Item 1	x_{11}	x_{12}	...	x_{1k}	...	x_{1p}
Item 2	x_{21}	x_{22}	...	x_{2k}	...	x_{2p}
⋮	⋮	⋮		⋮		⋮
Item j	x_{j1}	x_{j2}	...	x_{jk}	...	x_{jp}
⋮	⋮	⋮		⋮		⋮
Item n	x_{n1}	x_{n2}	...	x_{nk}	...	x_{np}

Selanjutnya data dapat di tampilkan sebagai array persegi panjang, yang disebut X , yang terdiri dari n baris dan p kolom:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2k} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{i1} & x_{i2} & \dots & x_{ik} & \dots & x_{ip} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} & \dots & x_{np} \end{bmatrix} \quad (1.2)$$

B. JENIS DATA MULTIVARIAT

Banyak metode multivariat, seperti regresi berganda, analisis diskriminan, dan lainnya, mensyaratkan ciri-ciri data tertentu. Dalam pemahaman statistik termasuk statistika Multivariat, data dapat dibagi menjadi data metrik dan non metrik.

- Data metrik data yang didapat dengan jalan *mengukur* dan bias mempunyai desimal. Seperti tinggi badan, yang bias saja bernilai 170 cm atau 178,45 cm (desimal). Data metrik akan dikategorikan lagi menjadi data interval atau data rasio.
- Data non metrik adalah data yang didapat dengan jalan *menghitung* tidak mempunyai desimal, dan pada data dapat dilakukan kategorisasi. Seperti jenis kelamin, yang dapat diberi kode "1" untuk pria diberi kode "2" untuk wanita. Data non metrik dibagi lagi menjadi data nominal dan data ordinal.

C. KARAKTERISTIK ANALISIS MULTIVARIAT

Analisis statistik multivariat adalah metode statistik yang memungkinkan peneliti melakukan penelitian terhadap lebih dari dua variabel dalam waktu

BAB II

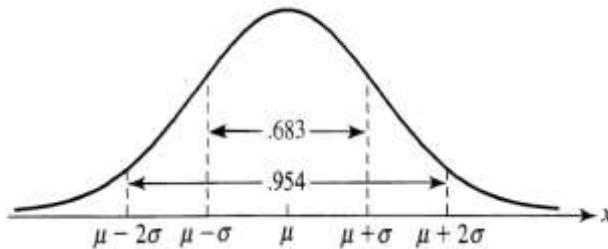
DISTRIBUSI NORMAL MULTIVARIAT

A. KONSEP DISTRIBUSI NORMAL MULTIVARIAT

Distribusi berguna dalam praktiknya karena dua alasan: Pertama, distribusi normal berfungsi sebagai model populasi yang bonafid dalam beberapa kasus; kedua, distribusi sampling dari banyak statistik multivariat yang mendekati normal, apa pun kondisinya bentuk populasi induk, karena efek batas pusat. Ringkasnya, banyak permasalahan dunia nyata yang secara alami masuk dalam kerangka tersebut teori normal. Pentingnya distribusi normal terletak pada peran gandanya sebagai baik model populasi untuk fenomena alam tertentu maupun pendekatan pengambilan sampel distribusi untuk banyak statistik.

Distribusi normal disebut pula distribusi Gauss. Distribusi ini merupakan distribusi yang paling banyak digunakan di dalam statistika. Fungsi densitasnya diberikan oleh formula berikut:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{[(x-\mu)/\sigma]^2}{2}}, -\infty < x < \infty \quad (2.1)$$



Gambar 1. Densitas Normal dengan mean μ dan varian σ^2 dan pemilihan area di bawah kurva

Varibel dikatakan berdistribusi normal multivariat dengan parameter μ dan Σ jika mempunyai *probability density function*

$$f(X_1, X_2, \dots, X_n) = \frac{1}{(2\pi)^{1/2} |\Sigma|^{1/2}} \left\{ e^{-\frac{1}{2}(X - \mu)' \Sigma^{-1} (X - \mu)} \right\} \quad (2.2)$$

Jika X_1, X_2, \dots, X_n berdistribusi normal multivariat maka $(X - \mu)' \Sigma^{-1} (X - \mu)$ berdistribusi χ_p^2 .

Contoh:

Misalkan kita akan menghitung fungsi padat normal dengan $p = 2$ dengan parameternya $\mu_1 = E(X_1), \mu_2 = E(X_2), \sigma_{11} = \text{Var}(X_1), \sigma_{22} = \text{Var}(X_2)$ dan

$$\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}$$

Jawab:

Dari soal kita peroleh matriks kovariansi

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

dengan

$$\Sigma^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}$$

Koefisien korelasi dapat dihitung dengan

$$\sigma_{12} = \rho_{12} \sqrt{\sigma_{11}} \sqrt{\sigma_{22}}$$

Diperoleh $\sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22} - (1 - \rho_{12}^2)$ dan jarak kuadrat menjadi

$$(X - \mu)' \Sigma^{-1} (X - \mu)$$

$$\begin{aligned}
&= \left[x_1 - \mu_1, x_2 - \mu_2 \right] \frac{1}{\sigma_{11}\sigma_{22}(1 - \rho_{12}^2)} \\
&\quad \begin{bmatrix} \sigma_{22} & -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \\ -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{11} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} \\
&= \frac{\sigma_{22}(x_1 - \mu_1)^2 + \sigma_{11}(x_2 - \mu_2)^2 - 2\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_{11}\sigma_{22}(1 - \rho_{12}^2)} \\
&= \frac{1}{(1 - \rho_{12}^2)} \left[\left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}} \right)^2 + \left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}} \right)^2 - 2\rho_{12} \left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}} \right) \left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}} \right) \right]
\end{aligned}$$

B. SIFAT-SIFAT TERTENTU DARI DISTRIBUSI NORMAL

Sifat-sifat tertentu dari distribusi normal akan diperlukan berulang kali dalam penjelasan kami model dan metode statistik. Properti ini memungkinkan untuk memanipulasi distribusi normal dengan mudah yang dapat dinyatakan lebih tepat secara sederhana.

Hal berikut ini berlaku untuk vektor acak \mathbf{X} yang memiliki normal multivariat distribusi:

1. Kombinasi linier komponen \mathbf{X} terdistribusi normal.
2. Semua himpunan bagian komponen \mathbf{X} mempunyai distribusi normal (multivariat).
3. Kovarian nol menyiratkan bahwa komponen-komponen yang bersesuaian bersifat independen didistribusikan.
4. Distribusi kondisi komponen (multivariat) normal.

Pernyataan-pernyataan ini direproduksi secara matematis dalam hasil yang diilustrasikan dengan contoh. Bukti-bukti yang disertakan akan membantu meningkatkan pemahaman tentang manipulasi matriks dan juga membawa pada apresiasi untuk cara di mana hasil-hasil tersebut secara berturut-turut berkembang dengan sendirinya

Contoh:

Untuk \mathbf{X} adalah distribusi dari $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, tentukan distribusi dari

$$\begin{bmatrix} X_1 - X_2 \\ X_2 - X_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \mathbf{A}\mathbf{X}$$

BAB III

MODEL REGRESI LINEAR MULTIVARIAT

A. KONSEP ANALISIS REGRESI

Regresi adalah ukuran hubungan antara dua variabel atau lebih yang dinyatakan sebagai suatu hubungan atau fungsi. Untuk menentukan bentuk hubungan (regresi) perlu adanya pemisahan yang tegas antara variabel bebas yang biasanya diwakili dengan simbol dari variabel lain dan sebaliknya. Kedua variabel ini seringkali bersifat kausal atau mempunyai hubungan sebab akibat, artinya saling mempengaruhi. Jadi regresi merupakan suatu bentuk fungsi antara variabel terikat Y dan variabel bebas X atau dapat dikatakan regresi adalah fungsi $Y = f(X)$. Bentuk regresi bergantung pada fungsi pendukungnya atau persamaannya (Yuliara, 2016).

Analisis/pengujian regresi adalah studi tentang hubungan antara suatu variabel dengan satu variabel atau lebih. Apabila variabel bebasnya hanya satu maka uji/analisis regresi disebut regresi linier sederhana. Jika terdapat lebih dari satu variabel bebas maka uji/analisis regresi disebut regresi linier berganda. Dikatakan linier berganda karena paling sedikit terdapat dua variabel bebas yang mempengaruhi variabel terikat. Analisis regresi digunakan untuk memberikan dasar dalam membuat prediksi. Variabel yang satu dapat diprediksi oleh variabel yang lain. Variabel yang memprediksi disebut variabel kriteria atau variabel terikat, sedangkan variabel yang berperan sebagai prediktordisebut variabel prediktor atau variabel bebas. Tujuan analisis regresi adalah untuk menggambarkan garis regresi dengan menggunakan persamaan regresi dan memperoleh dasar prediksi dengan kesalahan atau residu prediksi sekecil mungkin. Variabel kriteria harus dapat diprediksi oleh variabel prediktor. Syarat pertama adalah regresi linier hanya dapat digunakan pada skala interval dan rasio. Syarat kedua adalah variabel terikat dan variabel bebas harus mempunyai koefisien korelasi yang signifikan.

B. MODEL REGRESI LINEAR BERGANDA

Regresi linear berganda digunakan untuk mengetahui seberapa besar pengaruh antara lebih dari satu variabel bebas (x) terhadap variabel terikatnya

yaitu (y). Data yang terkumpul diolah dengan menggunakan model regresi linier berganda. Model regresi linier berganda sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_m X_{im} + \varepsilon_i \quad (3.1)$$

atau

$$Y_i = \beta_0 + \sum_{i=1}^n \sum_{j=1}^m \beta_j X_{ij} + \varepsilon_i \quad (3.2)$$

dengan:

- Y_i = Variabel terikat
- β_0 = Intersep
- β_j = Koefisien Regresi untuk peubah prediktor ke- j
- X_{ij} = nilai ke- i untuk peubah prediktor ke- j
- ε_i = nilai ke- i peubah acak galat

Dalam bentuk matriks persamaan (3.1) adalah:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1m} \\ 1 & X_{21} & X_{22} & \dots & X_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nm} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

atau

$$Y = X\beta + \varepsilon \quad (3.3)$$

dengan:

- Y = vektor peubah respon berukuran $n \times 1$
- X = matriks peubah prediktor berukuran $n \times (m + 1)$
- β = vektor koefisien regresi berukuran $(m + 1) \times 1$
- ε = vektor peubah acak galat berukuran $n \times 1$

Asumsi yang melandasi peubah acak galat adalah:

$$\varepsilon_{ii} \sim NIID (\mathbf{0}, \sigma^2 \mathbf{I}) \quad (3.4)$$

Menerapkan asumsi (3.4) pada persamaan (3.1)

I. LATIHAN SOAL

1. Tentukan nilai penaksir $\hat{\beta}$ dengan metode OLS pada data berikut

Bulan	X_1	X_2	Y (Rp 1000)
	Jam kerja pegawai	Jam kerja mesin/komputer	Pajak aktual yang tidak dibayar
Januari	45	16	29
Pebruari	42	14	24
Maret	44	15	27
April	45	13	25
Mei	43	13	26
Juni	46	14	28
Juli	44	16	30
Agustus	45	16	28
September	44	15	28
Oktober	43	15	27

2. Tentukan persamaan regresi linear berganda berdasarkan data menggunakan R Studio dan SPSS

No.	Disiplin Kerja	Beban kerja	Kinerja
1	54	85	62
2	42	82	58
3	47	87	60
4	45	75	58
5	41	71	58
6	38	78	43
7	28	78	40
8	47	87	59
9	56	86	65
10	22	72	35

BAB IV

ANALISIS KOMPONEN UTAMA

A. KONSEP ANALISIS KOMPONEN UTAMA

Analisis komponen utama (PCA) ditemukan oleh Karl Pearson pada tahun 1901 dan digunakan dalam bidang biologi. Pada tahun 1947, teori ini ditemukan kembali oleh Karhunen dan kemudian dikembangkan lebih lanjut oleh Loeve pada tahun 1963, sehingga teori ini dikenal juga dengan nama transformasi Karhunen-Loeve dalam bidang ilmu telekomunikasi. Analisis komponen utama (PCA) adalah teknik statistik untuk mengubah variabel asli yang paling banyak digunakan dan berkorelasi satu sama lain menjadi kumpulan variabel baru yang lebih kecil dan independen satu sama lain. Oleh karena itu, analisis komponen utama (PCA) berguna dalam mereduksi data, memfasilitasi interpretasinya (Noya van Delsen et al., 2017).

Analisis komponen utama merupakan suatu teknik statistik yang dapat digunakan untuk menjelaskan struktur varians suatu himpunan variabel pada beberapa variabel baru dimana variabel-variabel baru tersebut tidak bergantung satu sama lain dan membentuk suatu kombinasi linier sifat-sifat dari variabel aslinya. Variabel baru tersebut disebut komponen utama. Secara umum, tujuan analisis komponen utama adalah mereduksi dimensi besar dari data berkorelasi menjadi dimensi data kecil yang tidak berkorelasi, yang dilakukan untuk tujuan interpretasi.

Analisis Komponen merupakan suatu teknik statistika untuk mengubah dari sebagian besar variabel asli yang digunakan yang saling berkorelasi satu dengan yang lainnya menjadi satu set variabel baru yang lebih kecil dan saling bebas (tidak berkorelasi). Jadi analisis komponen utama berguna untuk mereduksi data, sehingga lebih mudah untuk menginterpretasikan data-data tersebut. Secara aljabar linear komponen utama adalah kombinasi-kombinasi linear tertentu dari p peubah acak $x_1, x_1, x_1 \dots, x_p$ yang memiliki sebaran peubah ganda dengan vektor rataan μ dan matrik peragam Σ yang dapat ditulis:

$$\underline{Y} = A\underline{X} \quad \underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix} \quad A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pp} \end{bmatrix} \quad (4.1)$$

Sehingga komponen utama pertama dapat ditulis sebagai:

$$\begin{aligned} Y_1 &= a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ \underline{Y}_1 &= \underline{a}_1\underline{X} \end{aligned} \quad (4.2)$$

Yang memiliki ragam sebesar:

$$\begin{aligned} S^2Y_1 &= \sum_{i=1}^p \sum_{j=1}^p a_{i1} a_{j1} S_{ij} \\ &= \underline{a}_1 S \underline{a}_1 \end{aligned} \quad (4.3)$$

Persamaan dan ragam yang dimiliki pada komponen satu hampir sama dengan komponen ke dua sampai komponen p .

B. KEGUNAAN ANALISIS KOMPONEN UTAMA

Analisis Komponen Utama biasanya digunakan untuk:

1. Identifikasi variabel baru yang mendasari data multivariat.
2. Kurangi dimensi kumpulan variabel yang sering kali menyertakan banyak variabel berkorelasi dengan mempertahankan keragaman sebanyak mungkin dalam kumpulan data.
3. Dengan menghilangkan variabel-variabel asli yang kontribusi informasinya relatif rendah, maka variabel baru tersebut di atas disebut komponen utama dengan ciri-ciri sebagai berikut:
 - a. Ini adalah kombinasi linier dari variabel asli.
 - b. Jumlah kuadrat koefisien kombinasi linier sama dengan satu.
 - c. Itu tidak berkorelasi dan menunjukkan variasi dari yang terbesar hingga yang terkecil. Jika pendekatan Pearson dapat dikaitkan dengan masalah ruang vektor, khususnya mencari ruang vektor optimal. Pendekatan Hotelling dapat dikaitkan dengan masalah variabel acak, khususnya variabel acak baru dari keragaman terurut dan tidak berkorelasi, sehingga pendekatan lain adalah dari sisi komputasi.

Analisis komponen utama digunakan untuk menjelaskan struktur matriks varians-kovarians suatu himpunan variabel dengan kombinasi linier variabel-variabel tersebut. Secara umum, komponen pokok dapat digunakan untuk mereduksi dan menginterpretasikan variabel. Misalnya, ada variabel p yang terdiri dari objek. Kami juga berasumsi bahwa dari variabel δ , dihasilkan komponen utama p (dengan $k \leq p$) yang merupakan kombinasi linier dari

BAB V

ANALISIS FAKTOR

A. KONSEP ANALISIS FAKTOR

Analisis faktor merupakan salah satu metode multivariate yang digunakan untuk menganalisis variabel-variabel yang diduga memiliki keterkaitan satu sama lain sehingga keterkaitan tersebut dapat dijelaskan dan dipetakan atau dikelompokkan pada faktor yang tepat. Tujuan dari analisis faktor adalah untuk menggambarkan hubungan-hubungan kovarian antara beberapa variabel yang mendasari tetapi tidak teramati, kuantitas random yang disebut faktor. Vektor random teramati X dengan p komponen, memiliki rata-rata μ dan matrik kovarian Σ . Model analisis faktor adalah sebagai berikut :

$$X_1 - \mu_1 = \ell_{11}F_1 + \ell_{12}F_2 + \dots + \ell_{1m}F_m + \varepsilon_1 \quad (5.1)$$

$$X_p - \mu_p = \ell_{p1}F_1 + \ell_{p2}F_2 + \dots + \ell_{pm}F_m + \varepsilon_p \quad (5.2)$$

Atau dapat ditulis dalam notasi matrik sebagai berikut :

$$X_{pxl} = \mu_{(pxl)} + L_{(pxm)}F_{(m \times l)} + \varepsilon_{pxl} \quad (5.3)$$

dimana:

μ_i : rata-rata variabel i

ε_i : faktor spesifik ke $-i$

F_j : *common faktor* ke- j

ℓ_{ij} : loading dari variabel ke $-i$ pada faktor ke- j

Bagian dari varian variabel ke $-i$ dari m *common faktor* disebut komunalitas ke $-i$ yang merupakan jumlah kuadrat dari loading variabel ke $-i$ pada m *common faktor* dengan rumus:

$$h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \dots + \ell_{im}^2 \quad (5.4)$$

B. PENGUJIAN ANALISIS FAKTOR

- Uji Kaiser Meyer Olkin (KMO)

Uji KMO bertujuan untuk mengetahui apakah semua data yang telah terambil telah cukup untuk difaktorkan. Hipotesis dari KMO adalah sebagai berikut :

Hipotesis

H_0 : Jumlah data cukup untuk difaktorkan

H_1 : Jumlah data tidak cukup untuk difaktorkan

Statistik uji

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (5.5)$$

$i = 1, 2, 3, \dots, p$ dan $j = 1, 2, \dots, p$

r_{ij} = Koefisien korelasi antara variabel i dan j

a_{ij} = Koefisien korelasi parsial antara variabel i dan j

Apabila nilai KMO lebih besar dari 0,5 maka terima H_0 sehingga dapat disimpulkan jumlah data telah cukup difaktorkan.

- Uji Bartlett (Kebebasan Antar Variabel)

Uji Bartlett bertujuan untuk mengetahui apakah terdapat hubungan antar variabel dalam kasus multivariat. Jika variabel X_1, X_2, \dots, X_p *independent* (bersifat saling bebas), maka matriks korelasi antar variabel sama dengan matriks identitas. Sehingga untuk menguji kebebasan antar variabel ini, uji *Bartlett* menyatakan hipotesis sebagai berikut:

$H_0 : \boldsymbol{\rho} = \mathbf{I}$

$H_1 : \boldsymbol{\rho} \neq \mathbf{I}$

Statistik Uji :

$$\bar{r}_k = \frac{1}{p-1} \sum_{i=1}^p r_{ik}, \quad k = 1, 2, \dots, p \quad (5.6)$$

$$\bar{r} = \frac{2}{p(p-1)} \sum \sum_{i < k} r_{ik} \quad (5.7)$$

$$\hat{\gamma} = \frac{(p-1)^2 [1 - (1 - \bar{r})^2]}{p - (p-2)(1 - \bar{r})^2} \quad (5.8)$$

dengan :

\bar{r}_k = rata-rata elemen diagonal pada kolom atau baris ke k dari matrik R
(matrik korelasi)

\bar{r} = rata-rata keseluruhan dari elemen diagonal

Daerah penolakan :

tolak H_0 jika

$$T = \frac{(n-1)}{(1-\bar{r})^2} [\sum \sum_{i < k} (r_{ik} - \bar{r})^2 - \hat{\gamma} \sum_{k=1}^p (\bar{r}_k - \bar{r})^2] > \chi^2_{(p+1)(p-2)/2; \alpha}$$

Maka variabel-variabel saling berkorelasi hal ini berarti terdapat hubungan antar variabel. Jika H_0 ditolak maka analisis multivariat layak untuk digunakan terutama metode analisis komponen utama dan analisis faktor.

C. KEGUNAAN ANALISIS FAKTOR

Kegunaan analisis faktor antara lain :

- 1) Mereduksi sejumlah variabel asal menjadi variabel baru dengan jumlah lebih sedikit, variabel baru disebut faktor atau variabel laten atau konstruk
- 2) Mempermudah interpretasi hasil analisis, sehingga didapatkan informasi yang realistik dan bermanfaat.
- 3) Pengelompokkan dan pemetaan obyek (mapping dan clustering) berdasarkan karakteristik yang terkandung di dalam faktor.
- 4) Pemeriksaan validitas dan reliabilitas instrumen penelitian (berupa kuisioner).
- 5) Dengan diperolehnya skor faktor, maka analisis faktor merupakan data input dari berbagai metoda analisis data yang lain, misalnya analisis diskriminan, analisis regresi, analisis cluster, anova, analisis jalur dan model struktural.

D. PERSYARATAN ANALISIS FAKTOR

- 1) Skala pengukuran variabel adalah interval atau rasio
- 2) Besar sampel cukup memadai, kurang lebih setiap variabel memerlukan $n = 4-5$ responden
- 3) Besar sampel untuk k (jumlah) variabel $\times 4-5$ responden

BAB VI

ANALISIS CLUSTER

A. KONSEP ANALISIS CLUSTER

Analisis *cluster* merupakan metode analisis untuk mengelompokkan objek-objek pengamatan menjadi beberapa kelompok, sehingga akan diperoleh kelompok dimana objek-objek dalam satu kelompok memiliki banyak persamaan sedangkan dengan anggota kelompok lain memiliki banyak perbedaan. Jadi, tujuan pengelompokan adalah agar objek-objek yang bergabung dalam sebuah kelompok merupakan objek-objek yang mirip (atau berhubungan) satu sama lain dan berbeda (tidak berhubungan) dengan objek dalam kelompok yang lain. Ukuran jarak yang sering digunakan adalah ukuran jarak Euclidean. Jarak Euclidean merupakan jarak antar objek, misalkan dua objek ke- i dan ke- j yang berada pada p dimensi. Persamaanya dapat dituliskan sebagai berikut:

$$d_{ij} = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2}$$

(6.1)

dimana:

d_{ij} = Jarak antara objek ke- i dan objek ke- j

P = Jumlah variabel kluster

X_{ik} = Data dari subjek ke- i pada variabel ke- k

X_{jk} = Data dari subjek ke- j pada variabel ke- k

B. METODE PENGELOMPOKKAN

1. Metode Hierarki

Metode hierarki adalah suatu metode pada analisis *cluster* yang membentuk tingkatan tertentu seperti pada struktur pohon karena proses *cluster* dilakukan secara bertahap dan bertingkat. Hasil dari metode hierarki dapat disajikan dalam bentuk dendogram. Ada empat metode aglomeratif dalam pembentukan *cluster* yaitu *Single Linkage*, *Complete Linkage*, *Average Linkage*, dan *Ward's Method* dengan penjelasan sebagai berikut:

- *Single Linkage (Nearest Neighbor)*

Metode Single Linkage atau di sebut juga dengan metode pautan tunggal, jarak antara dua cluster (kelompok) dapat ditentukan dari dua abyek berpasangan yang memiliki kesamaan atau memiliki jarak terdekat (Nearest Neighbor) dalam cluster yang berbeda. Untuk menghitung jarak cluster (XY) dengan cluster-cluster yang lain dapat dirumuskan dengan:

$$d_{(XY)Z} = \text{Max}(d_{(XZ)}, d_{(YZ)}) \quad (6.2)$$

dimana:

$d_{(XZ)}$ = jarak antara *cluster* X dan Z

$d_{(YZ)}$ = jarak antara *cluster* Y dan Z

$d_{(XY)Z}$ = jarak antara *cluster* XY dan Z

- *Complete Linkage (Furthest Neighbor)*

Metode pautan lengkap didasarkan pada jarak maksimum. Jarak antara satu *cluster* dan *cluster* lain diukur berdasarkan objek yang mempunyai jarak terjauh. Prosedur ini dimulai dengan pengelompokkan dua individu yang mempunyai jarak terjauh (lebih melihat ketaksamaan). Misalkan individu X dan Y mempunyai jarak (d_{XY}) terjauh. Bagaimanakah dengan individu Z? Disini harus dicari jarak maksimum XZ dan XY sehingga dapat dituliskan sebagai berikut:

$$d_{(XY)Z} = \text{Max}(d_{(XZ)}, d_{(YZ)}) \quad (6.3)$$

dimana:

$d_{(XZ)}$ = jarak antara *cluster* X dan Z

$d_{(YZ)}$ = jarak antara *cluster* Y dan Z

$d_{(XY)Z}$ = jarak antara *cluster* XY dan Z

- *Average Linkage*

Jarak antara cluster pada metode Average Linkage atau disebut juga pautan rata-rata, di tentukan dari rata-rata seluruh objek suatu cluster terhadap seluruh objek pada cluster lainnya. Pada berbagai keadaan, metode ini dianggap stabil dibandingkan dengan kedua metode di atas. Untuk menghitung jarak cluster (XY) dengan cluster-cluster yang lain dapat dirumuskan dengan:

$$d_{(XY)Z} = \text{rata - rata}(d_{(XZ)}, d_{(YZ)}) \quad (6.4)$$

dimana:

$d_{(XZ)}$ = jarak antara *cluster* X dan Z

$d_{(YZ)}$ = jarak antara *cluster* Y dan Z

$d_{(XY)Z}$ = jarak antara *cluster* XY dan Z

- *Ward's Method*

Ward's Method merupakan suatu metode pembentukan *cluster* yang didasari oleh hilangnya informasi akibat penggabungan objek menjadi *cluster*. Hal ini diukur dengan menggunakan jumlah total dari deviasi kuadrat *mean cluster* untuk setiap pengamatan. *Error sum of square* (SSE) digunakan sebagai fungsi objektif. Dua objek akan digabungkan jika mempunyai fungsi objektif terkecil diantara kemungkinan yang ada. Dengan persamaan sebagai berikut:

$$SSE = \sum_{j=1}^p \left(\sum_{i=1}^n X_{ij}^2 - \left(\frac{1}{n} \sum_{i=1}^n X_{ij} \right)^2 \right) \quad (6.5)$$

dimana:

X_{ij} = nilai untuk obyek ke-i pada *cluster* ke-j

p = banyaknya variabel yang diukur

n = banyaknya objek dalam *cluster* yang terbentuk

2. Metode Non-Hirarki

Metode Non-Hirarki yang digunakan adalah K-Means yang bertujuan untuk menjelaskan algoritma dalam penentuan suatu objek ke dalam gerombol tertentu berdasarkan rata-rata yang terbentuk. K means adalah metode pengelompokan data non-hierarki yang mencoba membagi data yang ada menjadi satu atau lebih *cluster*/grup. Metode ini membagi data ke dalam *cluster*/kelompok sehingga data dengan karakteristik yang sama dikelompokkan menjadi *cluster* yang sama. *K-Means* merupakan algoritma clustering dengan proses iteratif. Huruf K diartikan sebagai jumlah grup yang akan dibuat. Selanjutnya nilai K ditentukan secara acak. Sedangkan resource memiliki nilai sementara sebesar yang merupakan pusat *cluster* atau disebut juga centroid. Jarak ke setiap centroid dihitung untuk setiap datum yang ada dengan menggunakan rumus Euclid sehingga diperoleh jarak terdekat ke centroid untuk setiap datum (Zaki et al., 2022).

BAB VII

ANALISIS DISKRIMINAN

A. KONSEP DASAR ANALISIS DISKRIMINAN

Pemodelan antara variabel prediktor/independen dengan variabel respon (independen) direpresentasikan dengan suatu fungsi, yang dibentuk dengan cara memaksimalkan jarak antar kelompok sehingga mampu membedakan antar kelompok. Berdasarkan fungsi tersebut, kelompok observasi yang tidak diketahui dapat ditentukan. Oleh karena itu, analisis diskriminan dapat digunakan sebagai metode klasifikasi. Dalam penerapannya, analisis ini mengasumsikan bahwa data dalam setiap kategori/kelompok bersifat multivariat berdistribusi normal dan memiliki struktur matriks varians-kovarians yang sama.

Analisis diskriminan adalah teknik statistika untuk mengelompokkan individu-individu ke dalam kelompok-kelompok yang saling bebas dengan tegas berdasarkan segugus peubah bebas. Analisis juga dapat dipergunakan untuk mengalokasikan pengamatan baru ke dalam kelompok yang telah didefinisikan sebelumnya. Dua asumsi yang harus diperhatikan dalam analisis diskriminan adalah:

1. P peubah bebas menyebar mengikuti sebaran normal ganda.
2. Matriks peragam menyebar mengikuti sebaran normal (homogen).

Sebelum dilakukan analisis Diskriminan, perlu dilakukan pengujian kesamaan vektor-vektor nilai tengah. Ada beberapa statistic uji yang dapat digunakan untuk menguji kesamaan vektor-vektor nilai tengah yaitu Lamda Wilk, Teras lawley-Hoteling, dan Teras Pillai. Model Analisis Diskriminan termasuk dalam multivariate dependence method adalah:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p \quad (7.1)$$

dimana:

X_p = variabel independen, yaitu data berjenis interval atau rasio, seperti usia seseorang, tinggi pohon, kandungan zat besi dalam tubuh.

b_p = koefisien fungsi diskriminan yang dihasilkan

Y = variabel dependen, yaitu data kategorikal atau nominal, seperti golongan tidak tertinggal (kode 0), golongan tertinggal (kode 1), dan sebagainya.

Jika data kategorikal tersebut hanya terdiri atas dua kode saja, maka

model bisa disebut *Group Discriminat Analysis*. Sedangkan jika kode lebih dari dua kategori disebut dengan *Multiple Discriminant Analysis*.

B. TUJUAN ANALISIS DISKRIMINAN

Tujuan analisis diskriminan: (1) Menghasilkan fungsi diskriminan, atau kombinasi linier, dari prediktor atau variabel bebas yang dapat membedakan atau membeda-bedakan kategori variabel atau kriteria atau kelompok terikat, yaitu mampu membedakan subjek (responden) yang mana. kelompok/kategori yang mana, (2) Periksa apakah terdapat perbedaan yang signifikan antar kategori/kelompok, dikombinasikan dengan variabel independen atau prediktor, (3) Tentukan prediktor/variabel independen mana yang memberikan kontribusi paling besar terhadap perbedaan (Susongko & Sari, 2016).

C. ASUMSI ANALISIS DISKRIMINAN

Asumsi penting yang harus dipenuhi agar model diskriminan dapat digunakan antara lain:

- 1) Variabel bebas harus terdistribusi normal (adanya normalitas).
- 2) Matriks kovarians semua variabel bebas harus sama (equal).
- 3) Tidak terjadi multikolinearitas (tidak berkorelasi) antar variabel bebas.
- 4) Tidak terdapat data yang ekstrim (outlier).

Untuk mengetahui kekuatan fungsi diskriminan yang telah terbentuk pada masing-masing observasi dalam proses pengelompokkan dapat dilakukan dengan melihat nilai dari korelasi kanonik, akar ciri (eigen value) dan group centroid (Astuti et al., 2021):

- 1) Korelasi Kanonik (Canonical correlation)

Canonical Correlation (R) merupakan satuan ukuran hubungan keterkaitan antara kelompok yang terbentuk oleh variabel respon dengan fungsi diskriminan yang ada. Ketika nilai R adalah nol, maka dapat disimpulkan bahwa tidak ada hubungan antara kelompok yang ada dengan fungsi yang terbentuk. Sebaliknya apabila nilai R mendekati 1, menunjukkan bahwa terdapat korelasi yang tinggi antara fungsi diskriminan dengan kelompok yang ada. Nilai R digunakan untuk menjelaskan seberapa besar fungsi yang terbentuk berguna dalam menentukan perbedaan kelompok.

1 1 0

2 0 3

Accuracy : 1

95% CI : (0.3976, 1)

No Information Rate : 0.75

P-Value [Acc > NIR] : 0.3164

Kappa : 1

Mcnemar's Test P-Value : NA

Sensitivity : 1.00

Specificity : 1.00

Pos Pred Value : 1.00

Neg Pred Value : 1.00

Prevalence : 0.25

Detection Rate : 0.25

Detection Prevalence : 0.25

Balanced Accuracy : 1.00

'Positive' Class : 1

Interpretasi:

Output 5 pada tabel *Confusion Matrix and Statistics*, 1 laki-laki yang masuk sesuai ke dalam Kelompok Laki-Laki dan 0 orang yang masuk ke dalam kelompok lainnya. Selanjutnya, 3 perempuan masuk sesuai ke dalam Kelompok Perempuan dan 0 orang yang masuk ke dalam kelompok lainnya. Selain itu, diperoleh nilai akurasi sebesar 1 atau 100%, yang artinya tidak terdapat kesalahan dalam mengklasifikasikan jenis kelamin responden.

G. LATIHAN SOAL

Buatlah Klasifikasi Kecamatan di Kabupaten Gowa berdasarkan jenjang Pendidikan dengan menggunakan 2 aplikasi

Kecamatan	TK	SD	SMP	PKBM
Somba Opu	94	71	51	30
Parangloe	13	27	14	12
Manuju	12	21	30	20
Bungaya	38	38	34	13
Tinggimoncong	23	82	66	37
Parigi	25	14	24	17
Tompobulu	29	39	53	33
Biringbulu	38	39	43	18
Bontomarannu	30	30	39	26
Pattalassang	26	28	44	27
Bajeng	16	9	10	9
Pallangga	11	9	20	16
Bajeng Barat	30	14	25	15
Barombong	50	29	22	8
Bontolempangan	0	0	0	0
Bontonompo	51	21	31	7
Bontonompo Selatan	34	21	24	10
Tombolopao	21	14	9	1

|| DAFTAR PUSTAKA

- Achim, Z. Torsten, H. (2002). Diagnostic Checking in Regression Relationships. *R News*: 2(3), 7-10. URL <https://CRAN.R-project.org/doc/Rnews/>
- Astuti, C. C., Wiguna, A., & Ariyanti, N. (2021). Pendekatan Analisis Diskriminan Pada Pembelajaran Daring Di Masa Pandemi Covid-19. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 15(1), 137–146. <https://doi.org/10.30598/barekengvol15iss1pp137-146>
- Departemen Statistika, (2004). Buku Teori: Pelatihan Analisis Multivariat. Fakultas Matematika dan Ilmu Pengetahuan Alam. Institut Pertanian Bogor.
- Fox, J. Weisberg, S. (2019). *An R Companion to Applied Regression*, Third edition. Sage, Thousand Oaks CA.
<https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- Haeruddin, M. F. (2016). *Pemodelan Regresi Linier untuk Data Deret Waktu Linear Regression Modeling for Time Series Data. 2.*
- Haumahu, G., & Lewaherilla, N. (2020). Penerapan Analisis Komponen Utama Dalam Mereduksi Faktor-Faktor Penyebab Diare Di Provinsi Maluku the Application of Principal Component Analysis To Reduce Diarrhea Factors in Maluku Province. *Mathematics & Applications (MAP Journal)*, 2(1), 41–46.
- Hennig, C. (2023). fpc: Flexible Procedures for Clustering. R package version 2.2-10. <https://CRAN.R-project.org/package=fpc>
- Irwan, A. S. (2021). *Statistika Multivariat* (M. S. P. Khalilah Nurfadilah, S.Si. (ed.)). Alauddin University Press.
- Iverson, B. L., & Dervan, P. B. (2014). *Applied Multivariate Statistical Analysis* (Sally Yagan (ed.); fifth). Prentice Hall, Upper saddle River, New Jersey.
- Johnson, R. A. dan Winchern, D.W (2007). *Applied Multivariate Statistical Analysis*. Sixth Edition. New Jersey: Person International Edition.
- Kadek, N. I., Putri, L., Asih, N. I. M., Eka, D. P., Wayan, N. I., Cahyani, Y., Ayu, I. G., Srinadi, M., Ariasih, M. P., Jayanegara, K., Widana, I. N., Putu, I., Kencana, E. K. A. N., Maziyya, P. A. Y. U., Sukarsa, I. K. G. D. E., Asih, N. I. M., D, P. M. P., & Sukarsa, K. G. D. E. (2015). Mengatasi

- Heteroskedastisitas Pada Regresi Dengan Menggunakan Weigted Least Square. *E-Jurnal Matematika*, 4, 20–25.
- Kassambara, A. Mundt, F. (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7. <https://CRAN.R-project.org/package=factoextra>
- Korkmaz, S. Goksuluk, D. Zararsiz, G. (2014). Mvn: An r package for assessing multivariate normality. *The R Journal*, 6(2):151–162.
- Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28(5), 1–26. <https://doi.org/10.18637/jss.v028.i05>
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2022). cluster: Cluster Analysis Basics and Extensions. R package version 2.1.4.
- Malikhatin, H., Rusgiyono, A., & Maruddani, D. A. I. (2021). Penerapan K-Modes Clustering Dengan Validasi Dunn Index Pada Pengelompokan Karakteristik Calon TKI Menggunakan R-Gui. *Jurnal Gaussian*, 10(3), 359–366. <https://doi.org/10.14710/j.gauss.v10i3.32790>
- Michael, F. John, F. (2021). candisc: Visualizing Generalized Canonical Discriminant and Canonical Correlation Analysis. R package version 0.8-6. URL <https://CRAN.R-project.org/package=candisc>
- Noya van Delsen, M. S., Wattimena, A. Z., & Saputri, S. (2017). Penggunaan Metode Analisis Komponen Utama Untuk Mereduksi Faktor-Faktor Inflasi Di Kota Ambon. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 11(2), 109–118. <https://doi.org/10.30598/barekengvol11iss2pp109-118>
- Supriyadi, E., Mariani, S., & Sugiman. (2017). Perbandingan Metode Partial Least Square (Pls) Dan. *Unnes Journal of Mathematics*, 6(2), 117–128.
- Susongko, P., & Sari, I. (2016). Analisis Diskriminan dalam Penelitian Ekonomi. *Cermin*, 047, 241–218. <https://media.neliti.com/media/publications/241218-analisis-diskriminan-dalam-penelitian-ek-a92272b4.pdf>
- Venables, W. N. Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
- Wickham, H. Averick, M. Bryan, J. Chang, W. McGowan, L. D. François, R. Grolemund, G. Hayes, A. Henry, L. Hester, J. Kuhn, M. Pedersen, T. L.

- Miller, E. Bache, S. M. Müller, K. Ooms, J. Robinson, D. Seidel, D. P. Spinu, V. Takahashi, K. Vaughan, D. Wilke, C. Woo, K. Yutani, H. (2019). "Welcome to the tidyverse." *Journal of Open Source Software*: 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- William, R. (2023). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 2.3.9, <https://CRAN.R-project.org/package=psych>
- Yuliara, I. M. (2016). Buku Regresi Linier Berganda. *Universitas Udayana*, 2(2), 18.
- Zaki, A., Irwan, I., & Sembe, I. A. (2022). Penerapan K-Means Clustering dalam Pengelompokan Data (Studi Kasus Profil Mahasiswa Matematika FMIPA UNM). *Journal of Mathematics Computations and Statistics*, 5(2), 163. <https://doi.org/10.35580/jmathcos.v5i2.38820>

|| BIODATA PENULIS



Isma Muthahharah, S.Si, M.Si.

Dosen Program Studi Statistika FMIPA Universitas Negeri Makassar, dilahirkan di Sinjai 05 April 1990. Pendidikan SD dijalani di SD Negeri 5 Sinjai Utara dan SMP di SMP Negeri 1 Sinjai Utara. Setamat SMP, ia meneruskan pendidikan SMA di SMA Negeri 01 Sinjai Utara. Ia meraih gelar S.Si (2011) Matematika dari Universitas Negeri Makassar (UNM). Gelar M.Si (2015) Statistika dari Institut Teknologi Sepuluh Nopember (ITS). Dengan latar belakang tersebut, Isma Muthahharah aktif melakukan penelitian di bidang Statistika Multivariat dan Analisis Time Series.



Zakiyah Mar'ah, S.Si., M.Si.

Dosen Program Studi Statistika FMIPA Universitas Negeri Makassar, dilahirkan di Ujung Pandang (Makassar) 1 Januari 1992. Pendidikan SD dijalani di SD Inpres Mallengkeri Bertingkat I Makassar dan SMP di SMP Negeri 3 Makassar. Setamat SMP, ia meneruskan pendidikan SMA di SMA Negeri 11 Makassar. Ia meraih gelar S.Si (2014) Statistika dari Universitas Hasanuddin (Unhas). Gelar M.Si (2017) Statistika dari Institut Pertanian Bogor (IPB). Dengan latar belakang tersebut, Zakiyah Mar'ah aktif melakukan penelitian di bidang Statistika Spasial dan Statistika Multivariat.

STATISTIKA MULTIVARIAT

APLIKASI SPSS DAN RSTUDIO



CV. Tahta Media Group
Surakarta, Jawa Tengah
Web : www.tahtamedia.com
Ig : tahtamedia group
Telp/WA : +62 896-5427-3996

ISBN 978-623-567-361-7 (PDF)



9 786231 473417