MODUL 3

Sampel Random

Prof. Dr. Sri Haryatmi Kartiko, M.Sc.

PENDAHULUAN

Pada umumnya populasi yang merupakan himpunan semua objek pembicaraan tidak dapat diketahui secara keseluruhan. Untuk melakukan inferensi pada populasi digunakan sampel, sehingga harus dipunyai sampel yang representatif yaitu sampel yang mewakili populasi dengan baik. Sampel yang memenuhi syarat tersebut adalah sampel random yang sesuai digunakan untuk populasi yang homogen. Dalam pengambilan sampel random mensyaratkan bahwa setiap elemen dalam populasi mempunyai kemungkinan yang sama untuk terambil ke dalam sampel.

Suatu observasi multivariat adalah koleksi pengukuran pada p variabel yang berbeda, diukur pada objek atau trial yang sama. Seperti telah disebutkan pada Modul 1, bila didapat n observasi, keseluruhan data dapat ditempatkan pada matriks bertipe $p\times n$. Setiap kolom dari X merupakan satu observasi multivariat. Karena keseluruhan pengukuran merupakan realisasi observasi maka data ini sering dikatakan sebagai sampel pengukuran n dan populasi p variat. Sampel terdiri dari n pengukuran, masing-masing dengan p komponen.

Dalam Kegiatan Belajar 1, Anda akan memepelajari interpretasi geometris dari statistik deskriptif $\underset{\sim}{\bar{x}}, \underset{\sim}{S}_n$, dan $\underset{\sim}{R}$. Penjelasan yang disampaikan menggunakan representasi baris-baris X sebagai p titik dalam dimensi n. Dalam Kegiatan Belajar 2 diperkenalkan anggapan bahwa suatu observasi merupakan sampel random. Secara sederhana sampel random memenuhi (1) pengukuran diambil pada objek atau trial berbeda, tidak berkorelasi antara satu dengan yang lain, (2) distribusi bersama untuk p variabel pada masing-masing objek adalah sama. Struktur sampel random ini memberikan justifikasi pemilihan jarak tertentu yang menghasilkan representasi geometri dimensi n dari data. Lebih lanjut, data yang diperlakukan sebagai sampel random memberikan dasar yang kuat dalam inferensi statistika yang dilakukan.Dalam kegiatan belajar ini juga disajikan *generalized variance*, yang mendeskripsikan ukuran variabilitas serta hubungan antara $\underset{\sim}{\overline{x}}$, $\underset{\sim}{S}_n$ dan mean dan kovariansi kombinasi linear variabel.

Secara umum setelah mempelajari modul ini, Anda diharapkan dapat menjelaskan sifat-sifat sampel random dalam kasus multivariat.Secara khusus setelah mempelajari modul ini, Anda diharapkan dapat:

1. memahami struktur sampel random;
2. menghitung *generalized variance* baik untuk satu variabel maupun variabel terstandarisasi; dan
3. menghitung dan memahami sifat-sifat kombinasi linear beberapa variabel.

KEGIATAN BELAJAR 1

Sifat Geometri Sampel

Suatu observasi multivariat adalah koleksi pengukuran pada p variabel berbeda diukur pada item atau trial yang sama. Seperti telah disebutkan pada Modul 2, untuk n observasi keseluruhan data set dapat ditulis sebagai matriks $\underset{\sim}{X}$ bertipe $p \times n$, yaitu:

$$\underset{p\times n}{\underset{\sim}{X}} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix}$$

Setiap kolom dari matriks $\underset{\sim}{X}$ merupakan satu observasi multivariat.

Data dapat diplot dalam 2 cara. Pertama, dengan diagram pencar (*scatter plot*) 2 dimensi, kolom-kolom X menunjukkan n titik dalam ruang berdimensi p.

$$\underset{p\times n}{\underset{\sim}{X}} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix} = \left[\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_n\right]$$

$\underset{\sim}{x}_1$ merupakan observasi multivariat pertama,

$\underset{\sim}{x}_2$ merupakan observasi multivariat kedua,

$\vdots$

$\underset{\sim}{x}_j$ merupakan observasi multivariat ke j,

Jadi, vektor kolom $\underset{\sim}{x}_j$ menunjukkan observasi ke j, membuat kolom koordinat titik tersebut.

Diagram pencar untuk n titik dalam ruang berdimensi p memberikan informasi tentang lokasi variabilitas mereka. Bila titik-titik menunjukkan lingkaran padat, vektor mean sampel $\overline{\underset{\sim}{x}}$ merupakan pusat keseimbangan. Variabilitas terjadi pada lebih dari 1 dimensi, yang besarnya ditunjukkan oleh matriks variansi dan kovariansi sampel $\underset{\sim}{S}_n$, satu ukuran variabilitas tunggal diberikan oleh determinan dari matriks variansi dan kovariansi sampel.

Contoh 3.1

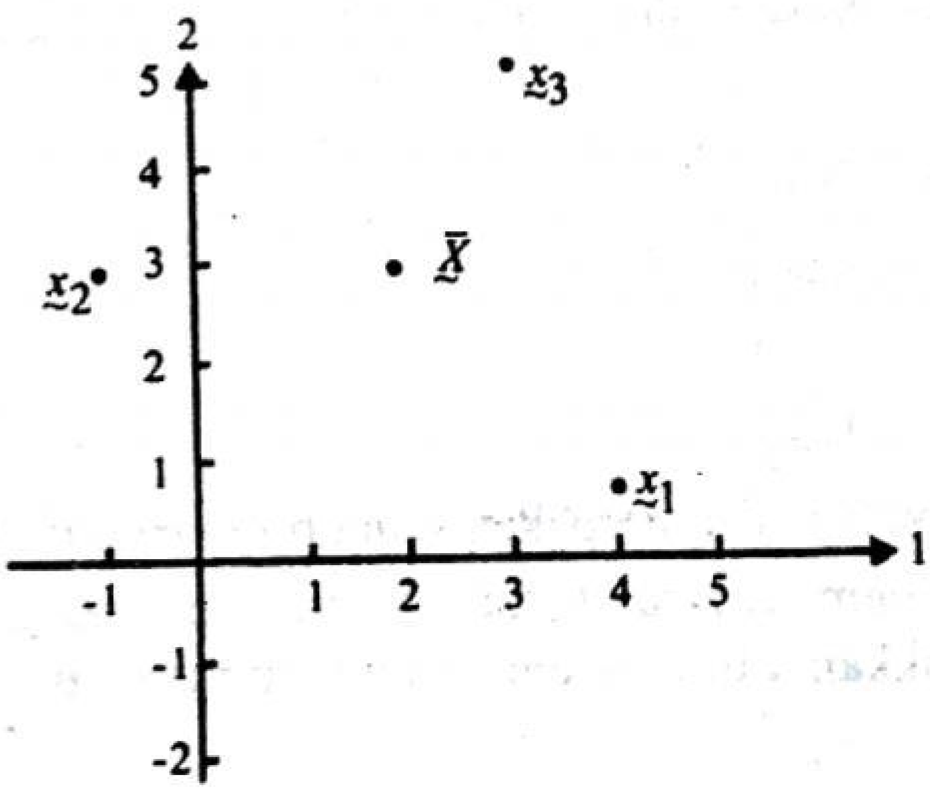
Hitung mean $\bar{\underset{\sim}{x}}$ dari matriks $\underset{\sim}{X}$, plot ke-3 titik data pada ruang berdimensi $p = 2$ dan tentukan letak $\bar{\underset{\sim}{x}}$ pada diagram tersebut.

$$\underset{\sim}{X} = \begin{bmatrix} 4 & -1 & 3 \\ 1 & 3 & 5 \end{bmatrix}$$

Penyelesaian:

$\underset{\sim}{x}_1' = [4,1], \underset{\sim}{x}_2' = [-1,3]$, dan $\underset{\sim}{x}_3' = [3,5]$

$$\bar{\underset{\sim}{x}} = \begin{bmatrix} \dfrac{4-1+3}{3} \\ \dfrac{1+3+5}{3} \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$



Gambar 3.1
Plot Matriks data X dengan n = 3 dalam ruang dimensi p = 2

Gambar 3.1 menunjukkan bahwa $\bar{\underset{\sim}{x}}$ merupakan titik keseimbangan diagram pencar.

Kedua, representasi geometris lain dikontruksikan dengan memandang data sebagai p titik dalam ruang berdimensi n. Dalam hal ini elemen-elemen dari baris-baris matriks data merupakan titik koordinat, yaitu

$$\underset{p\times n}{\underset{\sim}{X}} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{y}_1' \\ \underset{\sim}{y}_2' \\ \vdots \\ \underset{\sim}{y}_p' \end{bmatrix}$$

Koordinat titik pertama $\underset{\sim}{y}_1' = [x_{11}, x_{12}, ..., x_{1n}]$ adalah n pengukuran pada variabel pertama. Secara umum titik ke-i $\underset{\sim}{y}_i' = [x_{i1}, x_{i2}, ..., x_{in}]$ ditentukan oleh n-tupel pengukuran variabel ke-i.

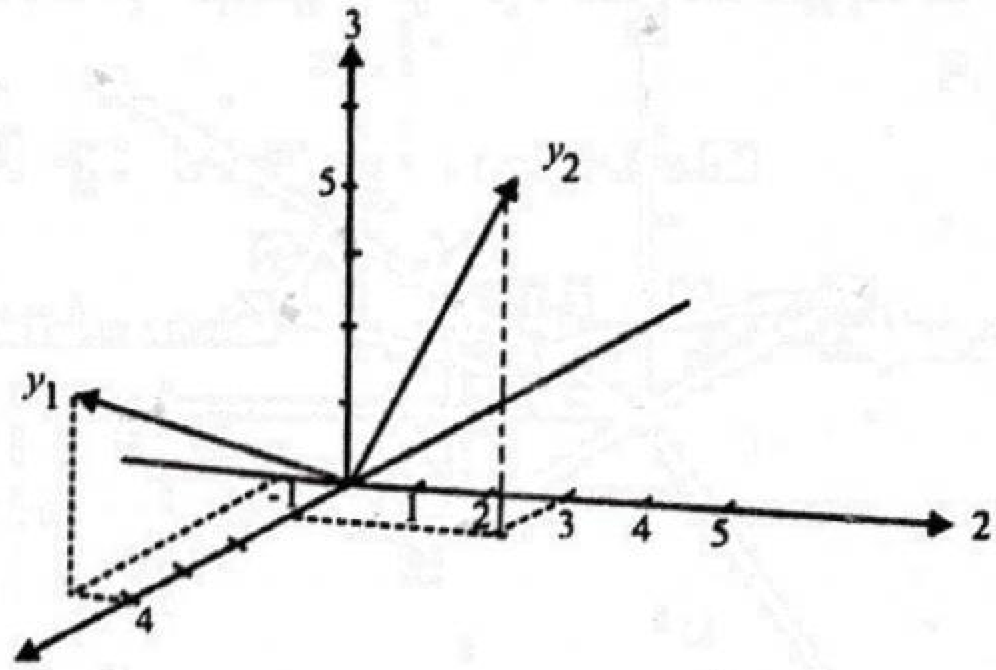
Contoh 3.2

Untuk data $\underset{\sim}{X}$ dalam Contoh 3.1 gambarlah sebagai vektor berdimensi p = 2 dalam ruang n = 3.

Penyelesaian:

$$\underset{\sim}{X} = \begin{bmatrix} 4 & -1 & 3 \\ 1 & 3 & 5 \end{bmatrix}$$

$\underset{\sim}{y}_1' = [4, -1, 3]$, $\underset{\sim}{y}_2' = [1, 3, 5]$

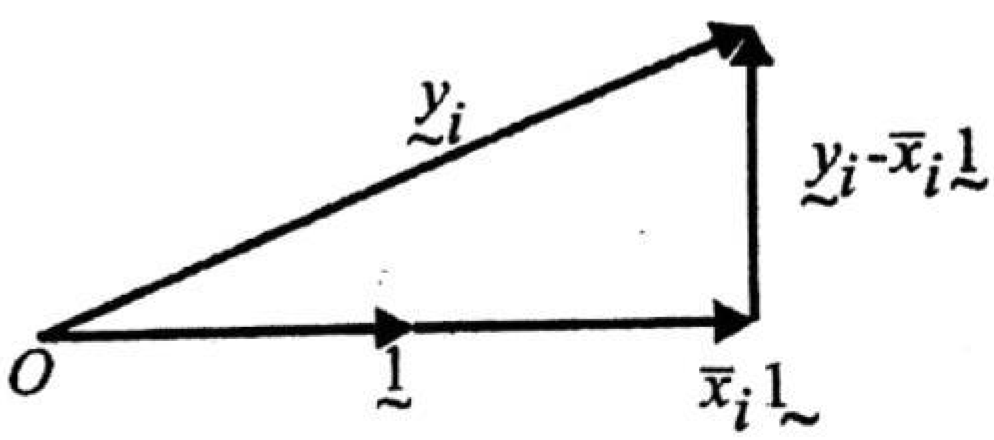


Gambar 3.2
Plot matriks data X sebagai vektor p = 2 dalam ruang ber dimensi n = 3

Sekarang Anda akan mempelajari interpretasi geometris tentang proses mendapatkan mean sampel. Didefinisikan vektor bertipe $n \times 1, \underset{\sim}{1}_n' = [1,1,...1]$, yang selanjutnya ditulis sebagai $\underset{\sim}{1}$. Vektor $\underset{\sim}{1}$ membentuk sudut sama dengan setiap koordinat sehingga vector $\left(\frac{1}{\sqrt{n}}\right)\underset{\sim}{1}$ mempunyai panjang 1unit dan sudut arah sama. Pandang vektor $\underset{\sim}{y}_i' = [x_{11}, x_{12}, ..., x_{1n}]$.Proyeksi $\underset{\sim}{y}_i'$ pada vektor satuan $\left(\frac{1}{\sqrt{n}}\right)\underset{\sim}{1}$ adalah

$$y_i'\left(\frac{1}{\sqrt{n}}\underset{\sim}{1}\right) = \frac{1}{\sqrt{n}}\underset{\sim}{1} = \frac{x_{i1}, x_{i2}, ..., x_{in}}{n}\underset{\sim}{1} = \overline{x}_i\underset{\sim}{1}$$

Dengan demikian setiap $\underset{\sim}{y}_j'$ mempunyai dekomposisi

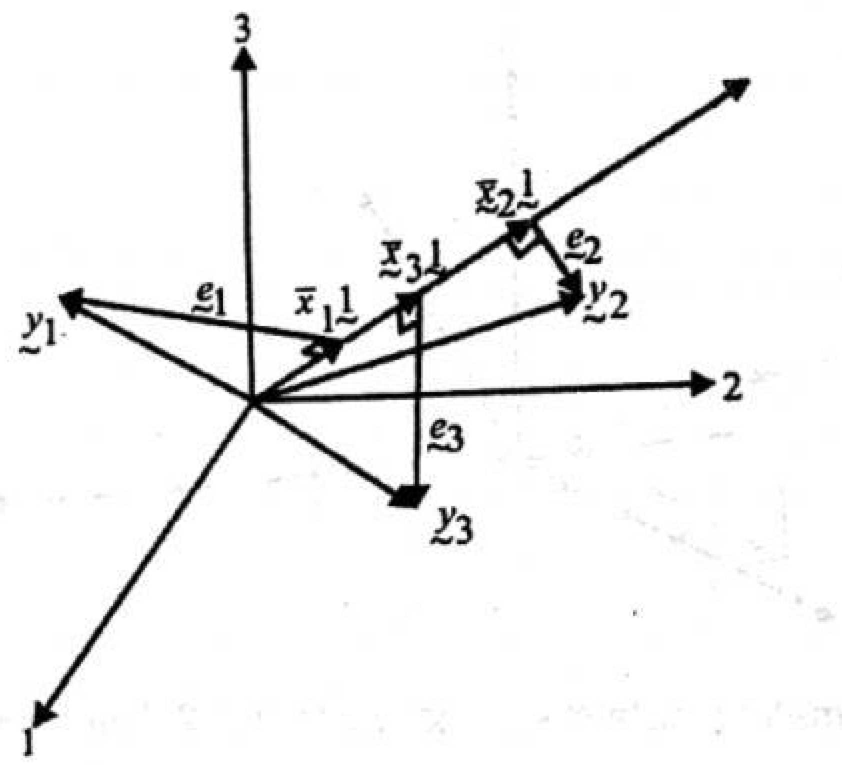


Dengan $\overline{x}_i\underset{\sim}{1}$ tegak lurus $\underset{\sim}{y}_i - \overline{x}_i\underset{\sim}{1}$. Vektor selisih $\underset{\sim}{e}_1 = \underset{\sim}{y}_i - \overline{x}_i\underset{\sim}{1}$ mempunyai elemen

$$\underset{\sim}{e}_1 = \underset{\sim}{y}_i - \bar{x}_i \underset{\sim}{1} = \begin{bmatrix} x_{i1} - \bar{x}_i \\ x_{i2} - \bar{x}_i \\ \vdots \\ x_{in} - \bar{x}_i \end{bmatrix}$$

yang merupakan deviasi pengukuran pada variabel ke-i dari rata-rata sampel.

Dekomposisi vektor $\underset{\sim}{y}_j$ menjadi komponen mean dan deviasi ditunjukkan dalam Gambar 3.3 untuk $p = 3$ dan $n = 3$.



Gambar 3.3
Dekomposisi $\underset{\sim}{y}_i$ menjadi $\bar{x}_i \underset{\sim}{1}$ dan komponen deviasi $\underset{\sim}{e}_1 = \underset{\sim}{y}_i - \bar{x}_i \underset{\sim}{1}, i = 1,2,3.$
1,2,3

Contoh 3.3

Carilah dekomposisi $\underset{\sim}{y}_i$ menjadi $\bar{x}_i \underset{\sim}{1}$ dan $\underset{\sim}{e}_1 = \underset{\sim}{y}_i - \bar{x}_i \underset{\sim}{1}, i = 1,2$ untuk data

$$\underset{\sim}{X} = \begin{bmatrix} 4 & -1 & 3 \\ 1 & 3 & 5 \end{bmatrix}$$

Dalam hal ini $\bar{x}_1 = \dfrac{4-1+3}{3} = 2$ dan $\bar{x}_2 = \dfrac{1+3+5}{3} = 3$, dengan demikian

$$\bar{x}_1\underset{\sim}{1} = 2\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}2\\2\\2\end{bmatrix}$$

$$\bar{x}_2\underset{\sim}{1} = 3\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}3\\3\\3\end{bmatrix}$$

Akibatnya

$$\underset{\sim}{e}_1 = \underset{\sim}{y}_1 - \bar{x}_1\underset{\sim}{1} = \begin{bmatrix}4\\-1\\3\end{bmatrix} - \begin{bmatrix}2\\2\\2\end{bmatrix} = \begin{bmatrix}2\\-3\\1\end{bmatrix}$$

$$\underset{\sim}{e}_2 = \underset{\sim}{y}_2 - \bar{x}_1\underset{\sim}{1} = \begin{bmatrix}1\\3\\5\end{bmatrix} - \begin{bmatrix}3\\3\\3\end{bmatrix} = \begin{bmatrix}-2\\0\\2\end{bmatrix}$$

Dapat ditunjukkan bahwa $\bar{x}_i\underset{\sim}{1}$ dan $\underset{\sim}{e}_1 = \underset{\sim}{y}_i - \bar{x}_i\underset{\sim}{1}$ tegak lurus, yaitu:

$$(\bar{x}_1\underset{\sim}{1})'\left(\underset{\sim}{y}_1 - \bar{x}_1\underset{\sim}{1}\right) = \begin{bmatrix}2 & 2 & 2\end{bmatrix}\begin{bmatrix}2\\-3\\1\end{bmatrix} = 4-6+2=0$$

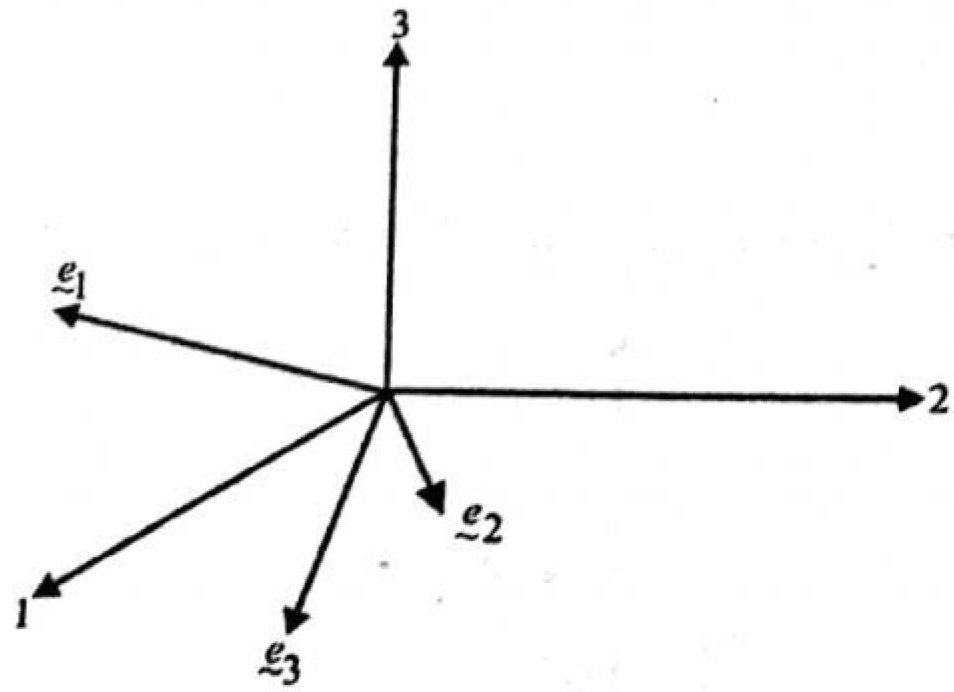
Hal yang sama juga berlaku untuk $\bar{x}_2\underset{\sim}{1}$ dan $\underset{\sim}{e}_2 = \underset{\sim}{y}_2 - \bar{x}_2\underset{\sim}{1}$ dengan komposisi

$$\underset{\sim}{y}_1 = \begin{bmatrix}4\\-1\\3\end{bmatrix} = \begin{bmatrix}2\\2\\2\end{bmatrix} + \begin{bmatrix}2\\-3\\1\end{bmatrix}$$

$$\underset{\sim}{y}_2 = \begin{bmatrix}1\\3\\5\end{bmatrix} = \begin{bmatrix}3\\3\\3\end{bmatrix} + \begin{bmatrix}-2\\0\\2\end{bmatrix}$$

Sampai saat ini Anda telah mempelajari vektor deviasi atau residual $\underset{\sim}{e}_1 = \underset{\sim}{y}_i - \bar{x}_i\underset{\sim}{1}$.

Plot vektor deviasi yang disajikan dalam Gambar 3.3 dengan translasi vektor deviaasi ke titik nol, tanpa merubah panjang maupun arahnya.



Gambar 3.4
Vektor Deviasi $\underset{\sim}{e}_i$ dari Gambar 3.3

Pandang kuadrat panjang vektor deviasi

$$Le^2 = \underset{\sim}{e}_1' \underset{\sim}{e}_i = \sum (x_j - \overline{x}_i)^2$$

(Panjang vektor deviasi)^2 = jumlah kuadrat deviasi

Nampak bahwa kuadrat panjang sebanding dengan variansi pengukuran variabel ke-i, secara ekuivalen panjang vektor sebanding dengan standar deviasi. Vektor yang lebih panjang menunjukkan variabilitas yang lebih besar.

Pandang 2 vektor deviasi $\underset{\sim}{e}_i$ dan $\underset{\sim}{e}_k$

$$\underset{\sim}{e}_i' \underset{\sim}{e}_k = \sum_{j=1}^{n} \left(x_{ij} - \overline{x}_i\right)\left(x_{kj} - \overline{x}_k\right)$$

Misal θ_{ik} adalah sudut antara vektor $\underset{\sim}{e}_i$ dan $\underset{\sim}{e}_k$. Didapat hubungan

$$\underset{\sim}{e}_i' \underset{\sim}{e}_k = L\underset{\sim}{e}_i L\underset{\sim}{e}_k (\theta_{ik})$$

yang berarti

$$\sum \left(x_{ij} - \overline{x}_i\right)\left(x_{kj} - \overline{x}_k\right) = \sqrt{\sum \left(x_{ij} - \overline{x}_i\right)^2} \sqrt{\sum \left(x_{kj} - \overline{x}_k\right)^2} \cos(\theta_{ik})$$

atau

$$r_{ik} = \frac{S_{ik}}{\sqrt{S_{ii}}\sqrt{S_{kk}}} = \cos(\theta_{ik})$$

Cosinus sudut merupakan koefisien koreksi sampel. Jadi, apabila kedua vektor deviasi mempunyai arah yang hampir sama, koefisien koreksi mendekati 1. Apabila dua vektor deviasi hampir tegak lurus, koefisien koreksi sampel mendekati nol.

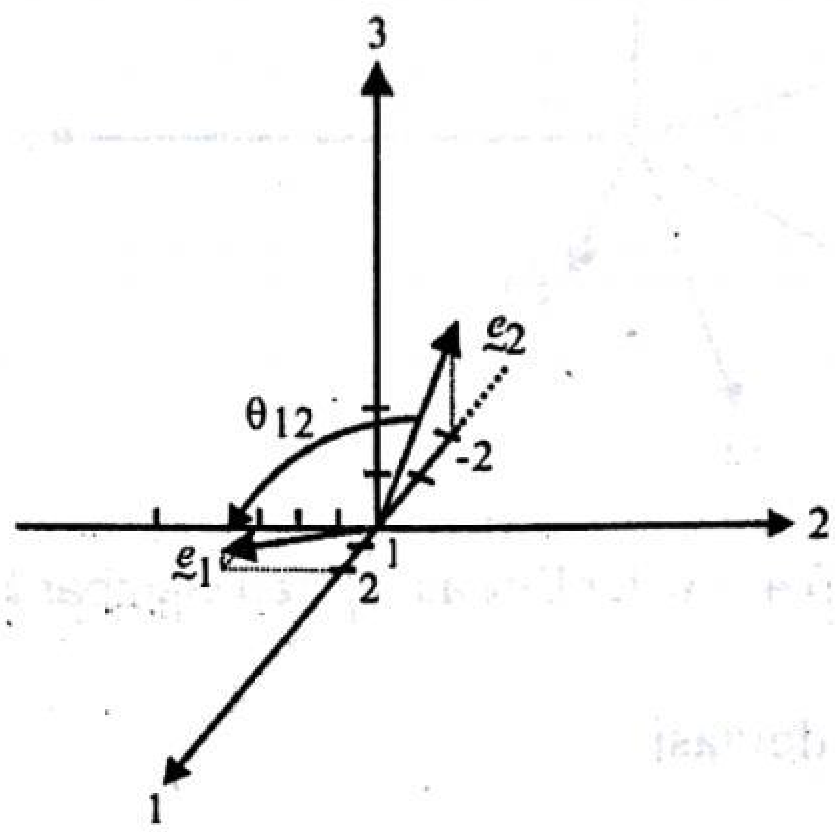
Bila kedua vektor mempunyai arah berlawanan, koefisien koreksi sampel mendekati -1.

Contoh 3.4

Diberikan vektor deviasi seperti dalam contoh 3.3. hitung matriks variansi-kovariansi $\underset{\sim}{S}_n$ dan matriks korelasi sampel $\underset{\sim}{R}$ dengan menggunakan konsep geometri yang baru saja Anda pelajari.

$$\underset{\sim}{e}_1 = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix} \qquad \underset{\sim}{e}_2 = \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix}$$

dengan translasi ke titik nol kedua vektor disajikan dalam Gambar 3.5.



Gambar 3.5
Vektor e_1 dan e_2

$$\underset{\sim}{e}_1'\underset{\sim}{e}_1 = \begin{bmatrix} 2 & -3 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix} = 14 = 3S_{11} \left(\text{bila } S_{ii} = \frac{\sum_{i=1}^{n} (x_{ij} - \overline{x}_i)^2}{n} \right)$$

atau $S_{11} = \dfrac{14}{3}$

$$\underset{\sim}{e}_2'\underset{\sim}{e}_2 = \begin{bmatrix} -2 & 0 & 2 \end{bmatrix} \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix} = 8 = 3S_{22}$$

atau $S_{22} = \dfrac{8}{3}$

$$\underset{\sim}{e}_1'\underset{\sim}{e}_2 = \begin{bmatrix} 2 & -3 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix} = -2 = 3S_{12}$$

atau $S_{12} = -\dfrac{2}{3}$

Jadi, $r_{12} = \dfrac{S_{12}}{\sqrt{S_{11}}\sqrt{S_{22}}} = \dfrac{-\frac{2}{3}}{\sqrt{\frac{14}{3}}\sqrt{\frac{8}{3}}} = -0{,}189$ dan

$$\underset{\sim}{S}_n = \begin{bmatrix} \frac{14}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{8}{3} \end{bmatrix}$$

$$\underset{\sim}{R} = \begin{bmatrix} 1 & -0{,}189 \\ -0{,}189 & 1 \end{bmatrix}$$

LATIHAN

Untuk memperdalam pemahaman Anda mengenai materi di atas, kerjakanlah latihan berikut!

Diberikan data matriks $\underset{\sim}{X} = \begin{bmatrix} 9 & 5 & 1 \\ 1 & 3 & 2 \end{bmatrix}$

1) Buatlah grafik diagram pencar dalam dimensi p = 2. Lokasikan mean sampel pada diagram yang sama!
2) Buatlah sketsa representasi dimensi n = 3 untuk data tersebut dan buat plot vektor deviasi $\underset{\sim}{y}_1 - \bar{x}_1\underset{\sim}{1}$ dan $\underset{\sim}{y}_2 - \bar{x}_2\underset{\sim}{1}$!
3) Buatlah sketsa vektor deviasi dengan mentranslasi vektor ini. Selanjutnya hitung panjang vektor ini dan cosinus sudut antara kedua vektor. Hubungkan kuantitas ini dengan $\underset{\sim}{S}_n$ dan $\underset{\sim}{R}$!

Petunjuk Jawaban Latihan

1) Setelah Anda membuat grafik pada ruang berdimensi 2, hitung mean dari *X,* selanjutnya buat plotnya pada sumbu yang sama.
2) Buat grafik pada dimensi tiga, jadi ada 2 titik yaitu $y_{12}(9,5,1)$ dan $y_{22}(1,3,2)$ pada ruang tersebut. Kemudian hitung $\bar{x}_1$ dan $\bar{x}_2$ dan dilanjutkan dengan menghitung $\underset{\sim}{y}_1 - \bar{x}_1\underset{\sim}{1}$ dan $\underset{\sim}{y}_2 - \bar{x}_2\underset{\sim}{1}$ yang disebut dengan vektor deviasi.
3) Geser vektor deviasi sehingga titik awalnya berada di titik nol, hitung panjang vektor $\underset{\sim}{e}_i'\underset{\sim}{e}_i$. Hitung cos (θ_{ik}) dan selanjutnya hitung $\underset{\sim}{S}_n$ dan $\underset{\sim}{R}$.

RANGKUMAN

Proyeksi baris y_i' dari matriks dara $\underset{\sim}{X}$ ke vektor satuan $\underset{\sim}{1}$ dengan sudut sama dengan setiap sumbu adalah $x_1'\underset{\sim}{1}$. Vektor ini mempunyai panjang $\sqrt{n}\bar{x}_i$ .Dengan demikian, mean sampel $\bar{x}_1$ mempunyai hubungan dengan panjang proyeksi $\underset{\sim}{y}_i$ pada vektor $\underset{\sim}{1}$.

Informasi untuk mendapatkan $\underset{\sim}{S}_n$ didapat vektor deviasi $\underset{\sim}{e}_i = \underset{\sim}{y}_i - \bar{x}_i \underset{\sim}{1} = \left[x_{i1} - \underset{\sim}{\bar{x}}_i, ..., x_{in} - \underset{\sim}{\bar{x}}_i\right]'$. Setelah translasi ke titik nol. Kuadrat panjangnya adalah nS_{ii} dan inner produk keduanya ($\underset{\sim}{e}_i$ dan $\underset{\sim}{e}_k$) adalah nS_{ik} .Koefisien korelasi sampel σ_{ik} adalah cosinus sudut antara $\underset{\sim}{e}_i$ dan $\underset{\sim}{e}_k$.

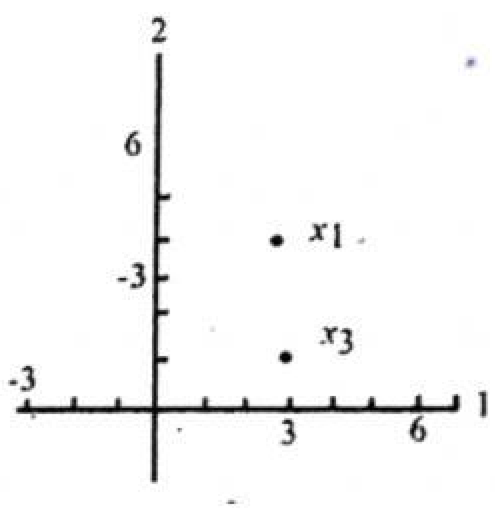
TES FORMATIF 1

Pilihlah satu jawaban yang paling tepat!

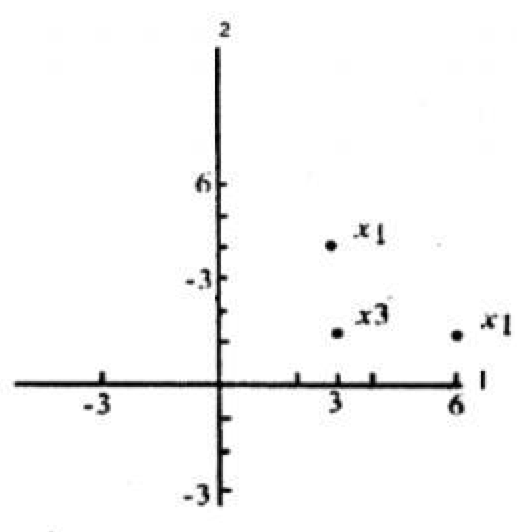
$$\underset{\sim}{X} = \begin{bmatrix} 3 & 6 & 3 \\ 4 & -2 & 1 \end{bmatrix}$$

1) Diagram pencar dari data dalam dimensi $p = 2$ adalah

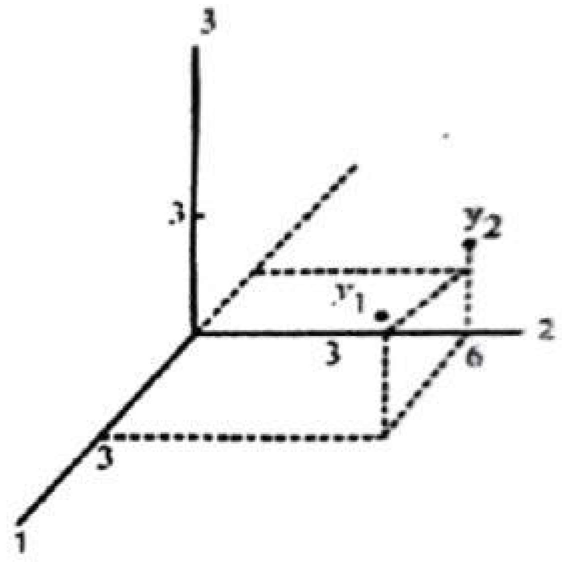
A.



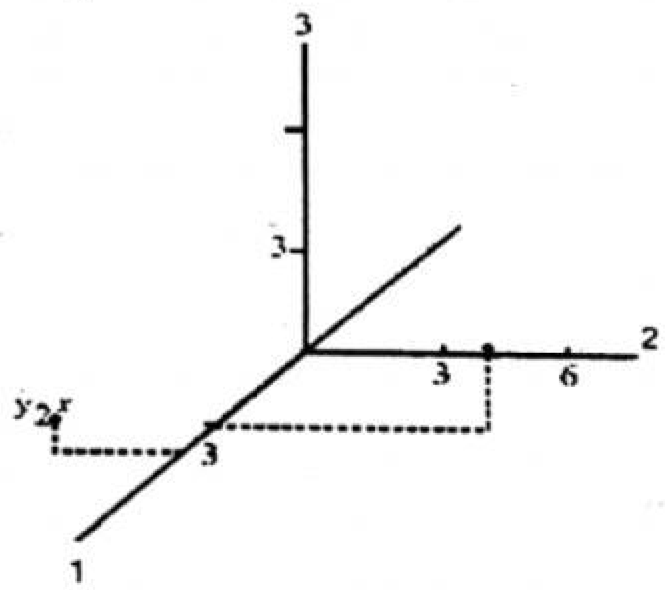
B.



C.



D.



2) Hitung mean sampel dari data tersebut

A. $\begin{bmatrix} 4 \\ 1 \end{bmatrix}$

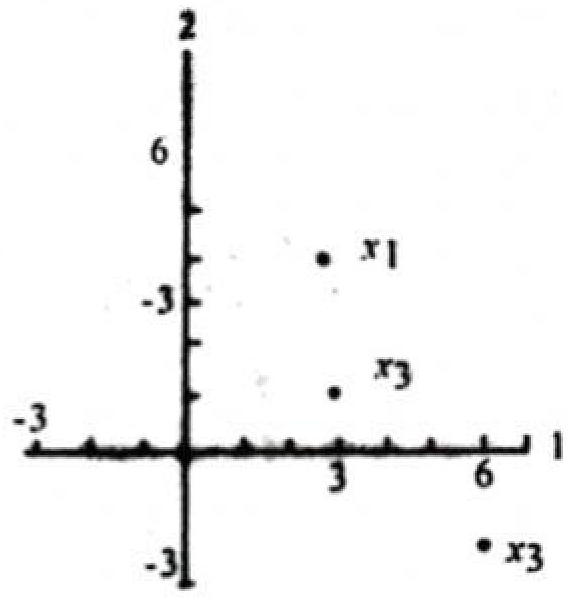
B. $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$

C. $\begin{bmatrix} 3{,}5 & 2 & 2 \end{bmatrix}$

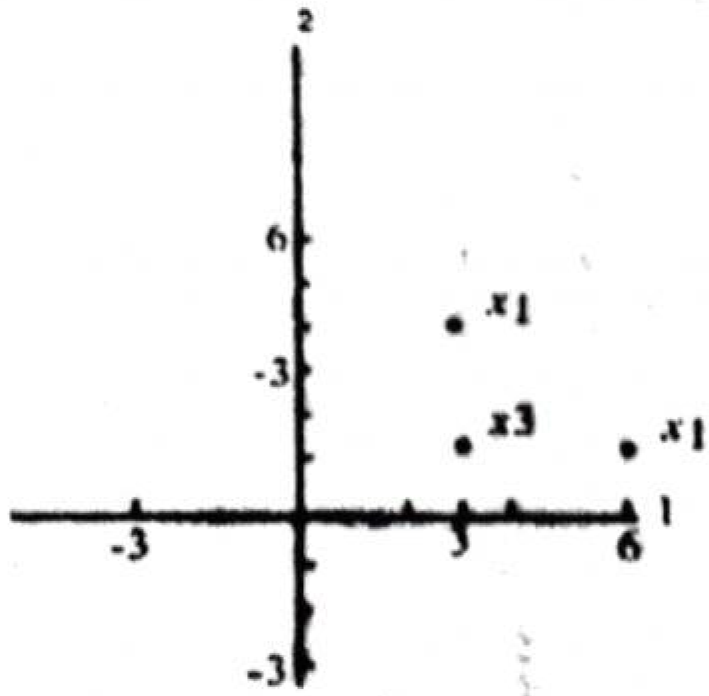
D. $\begin{bmatrix} 3 & 2 & 2 \end{bmatrix}$

3) Buat diagram pencar dari data dalam dimensi $n = 3$.

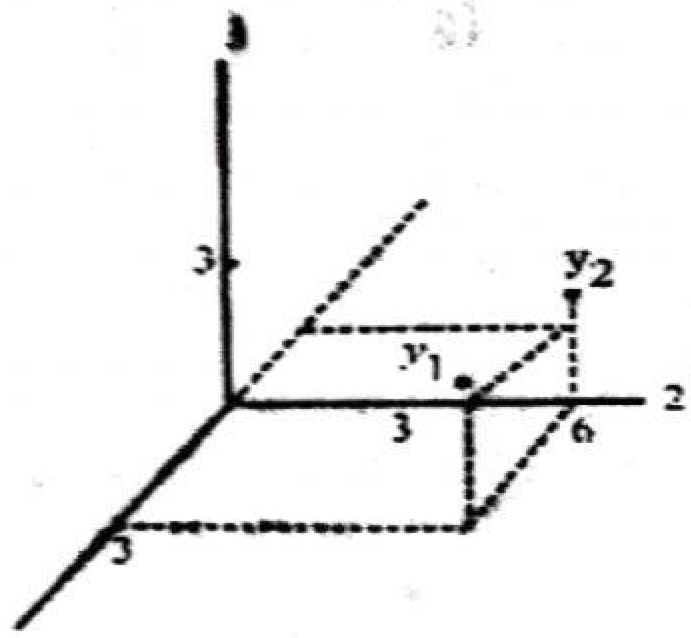
A.



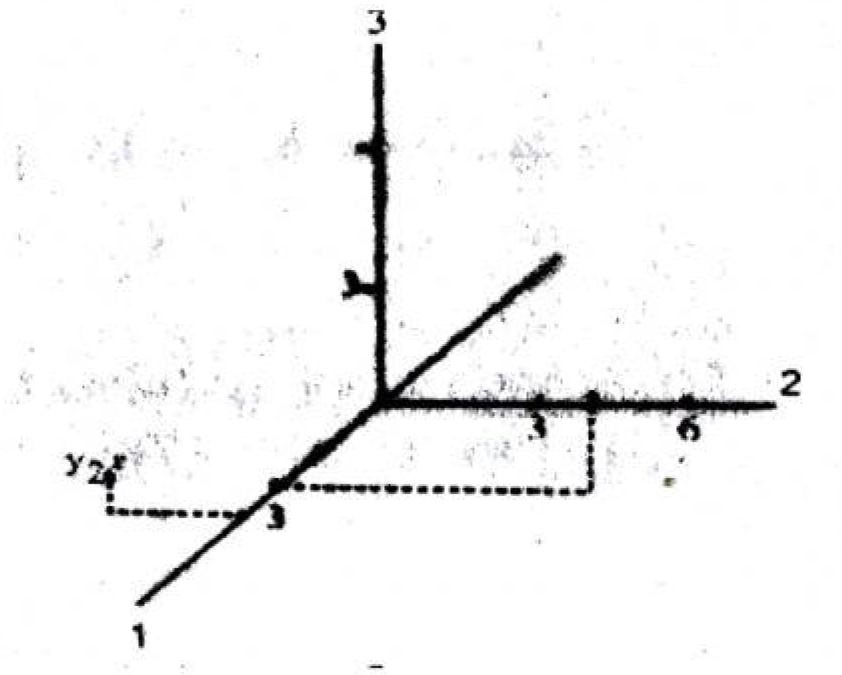
B.



C.



D.



4) Hitung vektor deviasi $\underset{\sim}{e}_1$ dan $\underset{\sim}{e}_2$

A. (1 -2 1) dan (0 3 -3)
B. (-1 2 -1) dan (3 -3 0)
C. (1 1 -2) dan (3 0 -3)
D. (-2 1 1) dan (0 3 -3)

5) Hitung $\underset{\sim}{S}_n$

A. $\begin{bmatrix} \frac{4}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{5}{3} \end{bmatrix}$

B. $\begin{bmatrix} 4 & -\frac{1}{3} \\ -\frac{1}{3} & 2 \end{bmatrix}$

C. $\begin{bmatrix} \frac{4}{3} & -1 \\ -1 & 5 \end{bmatrix}$

D. A, B, C tidak benar

Cocokkanlah jawaban Anda dengan Kunci Jawaban Tes Formatif 1 yang terdapat di bagian akhir modul ini. Hitunglah jawaban yang benar. Kemudian, gunakan rumus berikut untuk mengetahui tingkat penguasaan Anda terhadap materi Kegiatan Belajar 1.

$$\text{Tingkat penguasaan} = \frac{\text{Jumlah Jawaban yang Benar}}{\text{Jumlah Soal}} \times 100\%$$

Arti tingkat penguasaan: 90 - 100% = baik sekali
80 - 89% = baik
70 - 79% = cukup
< 70% = kurang

Apabila mencapai tingkat penguasaan 80% atau lebih, Anda dapat meneruskan dengan Kegiatan Belajar 2. **Bagus!** Jika masih di bawah 80%, Anda harus mengulangi materi Kegiatan Belajar 1, terutama bagian yang belum dikuasai.

KEGIATAN BELAJAR 2

Sifat-sifat Statistik dan Sampel Random

Untuk mempelajari sifat-sifat statistik dari sampel random yaitu $\bar{x}$ dan S_n yang akan digunakan dalam inferensi, diperlukan anggapan tentang variabel yang harga observasinya adalah data $\underset{\sim}{X}$.

Pada waktu data belum diobservasi kita menginginkan pengukuran pada *p* variabel. Sebelum pengukuran dibuat, harga tersebut tidak dapat diprediksi dengan tepat. Untuk itu kita memperlakukannya sebagai variabel random. Dalam konteks ini sel ke (*i,j*) dari matriks data disebut sebagai variabel X_{ij}, setiap himpunan pengukuran X_i pada *p* variabel adalah vektor random dan kita mempunyai matriks random.

$$\underset{p\times n}{\underset{\sim}{X}} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix} = \left[\underset{\sim}{x}_1, \underset{\sim}{x}_2, ..., \underset{\sim}{x}_n\right]$$

Berdasarkan matriks ini didefinisikan sampel random. Apabila vektor kolom $\underset{\sim}{X}_1, \underset{\sim}{X}_2, ..., \underset{\sim}{X}_n$ menunjukkan observasi independen dari distribusi bersama dengan fungsi densitas $f(x) = f(x_1, x_2, ..., x_n)$ maka $\underset{\sim}{X}_1, \underset{\sim}{X}_2, ..., \underset{\sim}{X}_n$ disebut sampel random dari $f(x)$. Secara matematis $\underset{\sim}{X}_1, \underset{\sim}{X}_2, ..., \underset{\sim}{X}_n$ membentuk sampel random apabila fungsi densitas bersamanya adalah hasil kali $f(x_1)f(x_2)...f(x_n)$ dengan $f(x_j) = f(x_{1j}, x_{2j}, ...x_{pj})$ merupakan fungsi densitas vektor kolom ke *j*.

Di bawah ini dituliskan dua hal yang berhubungan dengan sampel random.

1. Pengukuran pada *p* variabel dalam 1 trial seperti $\underset{\sim}{X}'_j = f\left[X_{1j}, X_{2j}, ...X_{pj}\right]$ biasanya berkorelasi. Pengukuran pada trial yang berbeda harus independen.
2. Independensi pengukuran dari trial ke trial tidak berlaku untuk data time series.

Contoh 3.5

X_1 adalah harga minyak mentah dalam juta dollar sedangkan X_2 adalah banyak impor minyak mentah.

Tabel 3.1.
Minyak mentah

Tahun	1970	1971	1972	1973	1974	1975
X_1 (harga)	54,8	58,4	58,4	67,0	116,2	137,9
X_2 (impor)	2716	2431	4541	6876	7360	8688

Dalam hal ini pengukuran $\underset{\sim}{X}' = \left[X_1 \quad X_2\right]$ tidak dapat dipandang sebagai sampel random berukuran n = 6 karena merupakan data runtun waktu (*time series*) yang independen satu dengan yang lain.

Telah dibahas dalam Modul 2 bahwa sifat independensi statistik mempunyai implikasi penting dalam pengukuran jarak. Jarak euclid sesuai digunakan dalam pengukuran vektor independen dan mempunyai variansi sama. Misal baris ke i $\underset{\sim}{Y}_i' = \left[X_{i1},\ldots,X_{1n}\right]$ dari $\underset{\sim}{X}$ dipandang sebagai titik dalam dimensi n. Lokasi titik ini ditentukan oleh distribusi probabilitas bersama $f(\underset{\sim}{y}_i) = f(x_{i1}, x_{i2}, \ldots, x_{in})$. Apabila pengukuran $x_{i1}, x_{i2}, \ldots, x_{in}$ merupakan sampelrandom, $f(\underset{\sim}{y}_i) = f(x_{i1}, x_{i2}, \ldots x_{in}) = f(x_{i1}) f(x_{i2}) \ldots f(x_{in})$ dan sebagai konsekuensinya setiap koordinat x_{ij} memberikan kontribusi terhadap lokasi melalui distribusi marginal identik $f_i(x_{ij})$.

Apabila setiap n komponen tidak independen atau distribusi marginal tidak identik, pengaruh-pengaruh pengukuran individu (koordinat) pada lokasi tidak simetris. Sehingga memerlukan konsep fungsi jarak yang memberikan bobot beda pada masing-masing koordinat yang disebut jarak statistik yang telah dibahas dalam modul sebelumnya.

Beberapa hal dapat disimpulkan mengenai distribusi sampling $\bar{\underset{\sim}{X}}$ dan $\underset{\sim}{S}$ tanpa memerlukan anggapan mengenai bentuk distribusi bersama dari variabel-variabel. Khususnya Anda akan melihat bagaimana sifat $\bar{\underset{\sim}{X}}$ dan $\underset{\sim}{S}_n$ sebagai estimator titik dari vektor mean $\underset{\sim}{\mu}$ dan matriks kovariansi $\underset{\sim}{\Sigma}$.

Teorema 3.1

Misal $\underset{\sim}{X}_1, \underset{\sim}{X}_2, ..., \underset{\sim}{X}_n$ adalah sampel random dari distribusi bersama dengan mean $\underset{\sim}{\mu}$ dan matriks kovariansi $\underset{\sim}{\Sigma}$ maka $\overline{\underset{\sim}{X}}$ adalah penduga tak bias untuk $\underset{\sim}{\mu}$ dan matriks kovariansinya adalah $\frac{1}{3}\underset{\sim}{\Sigma}$. Untuk matriks kovariansi $\underset{\sim}{S}_n$,

$\frac{n}{n-1}\underset{\sim}{S}_n$ merupakan penduga tak bias untuk $\underset{\sim}{\Sigma}$ dengan bias = $-\frac{1}{n}\underset{\sim}{\Sigma}$.

Bukti

$$\overline{\underset{\sim}{X}} = \frac{\underset{\sim}{X}_1, \underset{\sim}{X}_2, ..., \underset{\sim}{X}_n}{n}$$

Dengan menggunakan sifat-sifat ekspektasi

$$\begin{aligned}
E\left(\overline{\underset{\sim}{X}}\right) &= E\left(\frac{1}{n}\underset{\sim}{X}_1 + \frac{1}{n}\underset{\sim}{X}_2 + ... + \frac{1}{n}\underset{\sim}{X}_n\right) \\
&= E\left(\frac{1}{n}\underset{\sim}{X}_1\right) + E\left(\frac{1}{n}\underset{\sim}{X}_2\right) + ... + E\left(\frac{1}{n}\underset{\sim}{X}_n\right) \\
&= \frac{1}{n}E\left(\underset{\sim}{X}_1\right) + \frac{1}{n}E\left(\underset{\sim}{X}_2\right) + ... + \frac{1}{n}E\left(\underset{\sim}{X}_n\right) \\
&= \frac{1}{n}\underset{\sim}{\mu} + \frac{1}{n}\underset{\sim}{\mu} + ... + \frac{1}{n}\underset{\sim}{\mu} \\
&= \underset{\sim}{\mu}
\end{aligned}$$

Selanjutnya

$$\begin{aligned}
\underset{\sim}{S}_n &= \frac{1}{n}\left(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}}\right)\left(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}}\right)' \\
&= \frac{1}{n}\left[\sum_{j=1}^{n}\left(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}}\right)X_j' + \sum_{j=1}^{n}\left(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}}\right)\left(-\overline{\underset{\sim}{X}}\right)'\right] \\
&= \frac{1}{n}\left[\sum_{j=1}^{n}X_jX_j' - n\overline{\underset{\sim}{X}}\,\overline{\underset{\sim}{X}}'\right]
\end{aligned}$$

$$E(S_n) = \frac{1}{n} E\left[\sum_{j=1}^{n} \underset{\sim}{X}_j \underset{\sim}{X}_j' - n\underset{\sim}{\bar{X}}\underset{\sim}{\bar{X}}'\right]$$

$$= \frac{1}{n}\left[\sum_{j=1}^{n}\left(\underset{\sim}{X}_j \underset{\sim}{X}_j'\right) - nE\left(\underset{\sim}{\bar{X}}\underset{\sim}{\bar{X}}'\right)\right]$$

$$= \frac{1}{n}\left|\sum_{j=1}^{n}\left(\underset{\sim}{\Sigma} + \underset{\sim}{\mu}\underset{\sim}{\mu}'\right) - n\left(\frac{1}{n}\underset{\sim}{\Sigma}\underset{\sim}{\mu}\underset{\sim}{\mu}'\right)\right|$$

$$= \frac{1}{n}\left|n\underset{\sim}{\Sigma} + n\underset{\sim}{\mu}\underset{\sim}{\mu}' - \underset{\sim}{\Sigma} - n\underset{\sim}{\mu}\underset{\sim}{\mu}'\right|$$

$$= \frac{1}{n}\left|(n-1)\underset{\sim}{\Sigma}\right|$$

$$= \frac{n-1}{n}\underset{\sim}{\Sigma}$$

$\underset{\sim}{S}_n$ merupakan penduga untuk $\underset{\sim}{\Sigma}$ dengan bias $E(\underset{\sim}{S}_n) - \underset{\sim}{\Sigma} = -\frac{1}{n}\underset{\sim}{\Sigma}$

$$\left(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu}\right)\left(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu}\right)' = \left(\frac{1}{n}\sum_{j=1}^{n}\left(\underset{\sim}{X}_j - \underset{\sim}{\mu}\right)\right)\left(\frac{1}{n}\sum_{l=1}^{n}\left(\underset{\sim}{X}_l - \underset{\sim}{\mu}\right)\right)'$$

$$= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{l=1}^{n}\left(\underset{\sim}{X}_j - \underset{\sim}{\mu}\right)\left(\underset{\sim}{X}_l - \underset{\sim}{\mu}\right)'$$

$$Cov\left(\underset{\sim}{\bar{X}}\right) = E\left(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu}\right)\left(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu}\right)'$$

$$= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{l=1}^{n} E\left(\underset{\sim}{X}_j - \underset{\sim}{\mu}\right)\left(\underset{\sim}{X}_l - \underset{\sim}{\mu}\right)'$$

Untuk $j \neq l$ setiap komponen $\left(\bar{X}_j - \underset{\sim}{\mu}\right)\left(\bar{X}_l - \underset{\sim}{\mu}\right)'$ sama dengan nol karena $\underset{\sim}{x}_j$ dan $\underset{\sim}{x}_l$ independen sehingga

$$Cov\left(\underset{\sim}{\bar{X}}\right) = \frac{1}{n^2}\sum_{j=1}^{n}\left(\underset{\sim}{\bar{X}}_j - \underset{\sim}{\mu}\right)\left(\underset{\sim}{\bar{X}}_j - \underset{\sim}{\mu}\right)'$$

$$= \frac{1}{n^2}(\underset{\sim}{\Sigma} + \underset{\sim}{\Sigma} + \ldots + \underset{\sim}{\Sigma})$$

$$= \frac{1}{n^2} n\underset{\sim}{\Sigma} = \frac{1}{n}\underset{\sim}{\Sigma}$$

Catatan:

Penduga tak bias untuk kovarian matriks $\underset{\sim}{\Sigma}$ adalah:

$$\underset{\sim}{S} = \frac{n-1}{n-1}\underset{\sim}{S}_n = \frac{1}{n-1}\left(\overline{\underset{\sim}{X}}_j - \overline{\underset{\sim}{X}}\right)\left(\overline{\underset{\sim}{X}}_j - \overline{\underset{\sim}{X}}\right)'$$

Elemen baris ke *i* kolom ke *k* dari $\underset{\sim}{S}$ adalah

$$\frac{1}{n-1}\sum_{j=1}^{n}\left(\overline{X}_{ij} - \overline{\underset{\sim}{X}}_i\right)\left(\overline{X}_{kj} - \overline{\underset{\sim}{X}}_k\right)'$$

Definisi kovariansi sampel ini biasa digunakan dalam statistik uji multivariat. Kuantitas ini banyak menggantikan S_n sebagai matriks kovariansi sampel dalam sebagian besar pembahasan selanjutnya.

Dengan variabel tunggal, variansi sampel sering digunakan sebagai ukuran besarnya variansi pada variabel tersebut. Bila observasi menyangkut *p* variabel, ukuran variansi diberikan oleh matriks variansi kovariansi sampel.

$$\underset{\sim}{S} = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1p} \\ S_{21} & S_{22} & \cdots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \cdots & S_{pp} \end{bmatrix} = \left[S_{ik} = \frac{1}{n-1}\sum\left(x_{ij} - \overline{x}_i\right)\left(x_{ij} - \overline{x}_k\right) \right]$$

Matriks ini memuat p variansi dan $\frac{1}{n}p(p-1)$ kovariansi yang berbeda. Sering kali diperlukan harga numerik tunggal yang berkaitan dengan ukuran variansi di atas. Satu pikiran untuk harga ini adalah determinan S, yaitu $|S|$, yang harganya merupakan variansi sampel untuk $p = 1$.

$|\underset{\sim}{S}|$ selanjutnya disebut *Generalized sample variance* atau variansi sampel tergenarisasi. Jadi, $|\underset{\sim}{S}|$ merupakan ringkasan yang berupa 1 angka dari matriks variansi sampel.

Arti dari $|\underset{\sim}{S}|$

Misal dalam ruang berdimensi n

$$\underset{\sim}{e}_1 = y_1 - \overline{x}_1\underset{\sim}{1}$$

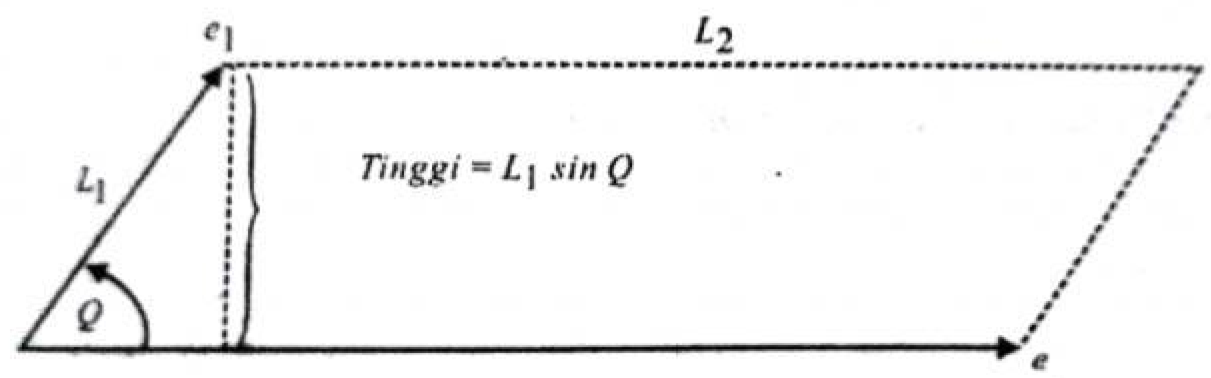
$$\underset{\sim}{e}_2 = y_2 - \overline{x}_2\underset{\sim}{1}$$

$\underset{\sim}{1}$ = matriks dengan semua elemen sama dengan 1. Dalam hal ini $\underset{\sim}{1}$ bertipe $n \times 1$.

L_1 panjang $\underset{\sim}{e}_1$

L_2 panjang $\underset{\sim}{e}_2$

$$L_1 = \sqrt{\sum_{j=1}^{n} \left(x_{1j} - \bar{x}_1 \right)^2} = \sqrt{(n-1)\, S_{11}}$$



$$L_2 = \sqrt{\sum_{j=1}^{n} \left(x_{2j} - \bar{x}_2 \right)^2} = \sqrt{(n-1)\, S_{22}}$$

$$\text{Cos}\,\theta = r_{12}, \text{ Luas : } L_1 L_2 \sqrt{1 - \cos^2 \theta}$$

$$\text{Luas} = (n-1)\sqrt{S_{11}}\sqrt{S_{22}}\sqrt{1 - r_{12}^2} = (n-1)\sqrt{S_{11} S_{22} \left(1 - r_{12}^2\right)}$$

Juga $|\underset{\sim}{S}| = \begin{vmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{vmatrix} = \begin{vmatrix} S_{11} & \sqrt{S_{11}}\sqrt{S_{22}}\, r_{12} \\ \sqrt{S_{11}}\sqrt{S_{22}}\, r_{12} & S_{22} \end{vmatrix}$

$$= S_{11}S_{22} - S_{11}S_{22}r_{12}^2 = S_{11}S_{22}\left(1 - r_{12}^2\right)$$

sehingga $|S| = (luas)^2 |(n-1)^2$.

Secara umum dengan matriks induksi *generalized sample variance* adalah $|\underset{\sim}{S}| = (n-1)^{(p-1)} (volume)^2$. Yang dimaksud volume adalah yang dihasilkan oleh $\underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots, \underset{\sim}{e}_{p-1}$.

Generalized sample variance berharga nol berarti paling sedikit 1 baris dari matriks deviasi

$$\begin{bmatrix} y_1' - \bar{x}_1 \underset{\sim}{1}' \\ y_1' - \bar{x}_2 \underset{\sim}{1}' \\ \\ y_p' - \bar{x}_p \underset{\sim}{1}' \end{bmatrix} = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_1 & \cdots & x_{1n} - \bar{x}_1 \\ x_{21} - \bar{x}_2 & x_{22} - \bar{x}_2 & \cdots & x_{2n} - \bar{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} - \bar{x}_p & x_{p2} - \bar{x}_p & \cdots & x_{pn} - \bar{x}_p \end{bmatrix}$$

dapat ditulis sebagai kombinasi linear dari baris-baris yang lain.

Secara geometris hal ini berarti bahwa $\underset{\sim}{e}_i' = (x_{i1} - \bar{x}_i, ..., x_{in} - \bar{x}_i)$ terletak pada *hiperplane* yang dihasilkan oleh $\underset{\sim}{e}_1 ... \underset{\sim}{e}_{i-1}, \underset{\sim}{e}_{i+1} ... \underset{\sim}{e}_p$.

Teorema 3.2

Bila $n \le p$, yaitu ukuran sampel $\le$ jumlah variabel, maka $|\underset{\sim}{S}| = 0$ untuk semua sampel.

Bukti:

Akan ditunjukkan bahwa rank dari $\underset{\sim}{S} \le P$

Untuk semua sampel $\sum_{j=1}^{n} \left(x_{ij} - \bar{x}_i \right) = 0$

Adanya kombinasi linear berarti rank dari $\underset{\sim}{X} - \underset{\sim}{\bar{x}} \underset{\sim}{1}' \le n - 1 \le p - 1$

$$(n-1) \underset{p\times p}{\underset{\sim}{S}} = \left(\underset{p\times p}{\underset{\sim}{X}} - \underset{p\times p}{\underset{\sim}{\bar{x}} \underset{\sim}{1}'} \right) (\underset{\sim}{X} - \underset{\sim}{\bar{x}} \underset{\sim}{1}')'$$

$$\begin{aligned} (n-1) kol_k(\underset{\sim}{S}) &= (\underset{\sim}{X} - \underset{\sim}{\bar{x}} \underset{\sim}{1})' kol_k(\underset{\sim}{X} - \underset{\sim}{\bar{x}} \underset{\sim}{1}') \\ &= (x_{ki} - \bar{x}_k) kol(\underset{\sim}{X} - \underset{\sim}{\bar{x}} \underset{\sim}{1}')' + ... + (x_{kn} - \bar{x}_k)' kol_n(\underset{\sim}{X} - \underset{\sim}{\bar{x}} \underset{\sim}{1}') \end{aligned}$$

Karena vektor kolom $\underset{\sim}{X} - \underset{\sim}{\bar{x}} \underset{\sim}{1}'$ mempunyai jumlah elemen nol maka Anda dapat menuliskan, misalnya $kol_1(\underset{\sim}{X} - \underset{\sim}{\bar{x}} \underset{\sim}{1}')$ sebagai negatif dari jumlah vektor kolom yang lain. Dengan demikian $kol_k(\underset{\sim}{S})$ merupakan kombinasi linear dari paling banyak $(n-1)$ vektor kolom yang linear independen, yaitu:

$$kol_2(\underset{\sim}{X} - \underset{\sim}{\bar{x}} \underset{\sim}{1}'), ..., kol_n(\underset{\sim}{X} - \underset{\sim}{\bar{x}} \underset{\sim}{1}')$$

sehingga rank $(\underset{\sim}{S}) \le p - 1$, atau S singular, dengan demikian $|\underset{\sim}{S}| = 0$.

Besarnya *Generalized sample variance* dipengaruhi oleh besarnya variabilitas masing-masing variabel $(S_{ii}, i = 1, ..., p)$. Misal S_{ii} besar, secara geometris vektor residual $\underset{\sim}{e}_i = \left(\underset{\sim}{y}_i - \overline{x}_i \underset{\sim}{1}'\right)$ panjang, hal ini menentukan besarnya volume. Akibatnya dari hal ini, kadang-kadang diperlukan membuat skala untuk semua vektor residual sedemikian sehingga mereka mempunyai panjang yang sama.

Membuat skala untuk vektor residual adalah mengganti observasi. x_{ij} dengan harga standarisasinya, yaitu

$$\frac{x_{ij} - \overline{x}_i}{\sqrt{S_{ii}}}$$

Matriks kovariansi sampel untuk variabel standarisasi ini adalah $\underset{\sim}{R}$ yang telah Anda kenal.

Generalized sample variance untuk variabel standarisasi $= \left|\underset{\sim}{R}\right|$.

Vektor $\left|\left(x_{i1} - \overline{x}_1\right) \big/ \sqrt{S_{11}}, \ldots, \left(x_{in} - \overline{x}_i\right) \Big/ \sqrt{S_{ii}}\right| = \dfrac{(y_i - x_i \underset{\sim}{1}')}{\sqrt{S_{ii}}}$ mempunyai panjang $\sqrt{n-1}$.

Generalized sample variance untuk variabel standarisasi akan besar bila vektor-vektor kolomnya hampir tegak lurus; berharga kecil bila 2 atau lebih dari vektor-vektor ini mempunyai hampir sama.

Bila θ_{ik} adalah sudut antara $(y_i - \overline{x}_i \underset{\sim}{1}')\sqrt{S_{ii}}$ dan $(y_k - \overline{x}_k \underset{\sim}{1}')\sqrt{S_{kk}}$ maka cos $\theta_{ik} = r_{ik}$

$\left|\underset{\sim}{R}\right|$ besar bila semua r_{ik} hampir sama dengan nol dan kecil bila r_{ik} hampir sama dengan. 1 atau – 1.

$$\left|\underset{\sim}{R}\right| = (n-1)^{-p}(\text{volume})^2$$

Hubungan antara $\left|\underset{\sim}{S}\right|$ dan $\left|\underset{\sim}{R}\right|$

$$\left|\underset{\sim}{S}\right| = \left(S_{11}S_{22}\ldots S_{pp}\right)\left|\underset{\sim}{R}\right|$$

$$(n-1)^p\left|\underset{\sim}{S}\right| = (n-1)^p\left(S_{11}S_{22}\ldots S_{pp}\right)\left|\underset{\sim}{R}\right|$$

$$= (n-1)S_{11}(n-1)S_{22}\ldots(n-1)S_{pp}\left|\underset{\sim}{R}\right|$$

Generalized variance yang lain adalah total variansi sampel total variansi sampel $= S_{11} + S_{22} + \ldots + S_{pp}$.

A. MEAN KOVARIANSI DAN KORELASI SEBAGAI HASIL OPERASI MATRIKS

$$\bar{\underset{\sim}{X}}_i = \left(x_{i1}.\underset{\sim}{1} + x_{i2}.\underset{\sim}{1} + \ldots + x_{in}.\underset{\sim}{1}\right)/n = y_i \underset{\sim}{1}/n$$

$$\bar{\underset{\sim}{X}} = \begin{bmatrix} \bar{\underset{\sim}{x}}_1 \\ \bar{\underset{\sim}{x}}_2 \\ \vdots \\ \bar{\underset{\sim}{x}}_p \end{bmatrix} = \begin{bmatrix} \dfrac{\underset{\sim}{y}_1{}'\underset{\sim}{1}}{n} \\ \dfrac{\underset{\sim}{y}_2{}'\underset{\sim}{1}}{n} \\ \vdots \\ \dfrac{\underset{\sim}{y}_p{}'\underset{\sim}{1}}{n} \end{bmatrix} = \frac{1}{n} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\bar{\underset{\sim}{x}} = \frac{1}{n} \underset{\sim}{X}'\underset{\sim}{1}$$

$$\underset{\sim}{X}\underset{\sim}{1}' = \frac{1}{n} \underset{\sim}{X}\underset{\sim}{1}\underset{\sim}{1}' = \begin{bmatrix} \bar{x}_1 & \bar{x}_1 & \cdots & \bar{x}_1 \\ \bar{x}_2 & \bar{x}_2 & \cdots & \bar{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_p & \bar{x}_p & \cdots & \bar{x}_p \end{bmatrix}$$

$$\underset{\sim}{X} - \frac{1}{n} \underset{\sim}{X}\underset{\sim}{1}\underset{\sim}{1}' = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_1 & \cdots & x_{1n} - \bar{x}_1 \\ x_{21} - \bar{x}_2 & x_{22} - \bar{x}_2 & \cdots & x_{2n} - \bar{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} - \bar{x}_p & x_{p2} - \bar{x}_p & \cdots & x_{pn} - \bar{x}_p \end{bmatrix}$$

$$(n-1)\underset{\sim}{S} = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_1 & \cdots & x_{1n} - \bar{x}_1 \\ x_{21} - \bar{x}_2 & x_{22} - \bar{x}_2 & \cdots & x_{2n} - \bar{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} - \bar{x}_p & x_{p2} - \bar{x}_p & \cdots & x_{pn} - \bar{x}_p \end{bmatrix} \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{pn} - \bar{x}_p \\ x_{12} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{p2} - \bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} - \bar{x}_1 & x_{2n} - \bar{x}_2 & \cdots & x_{pn} - \bar{x}_p \end{bmatrix}$$

$$= \left(\underset{\sim}{X} - \frac{1}{n} \underset{\sim}{X}\underset{\sim}{1}\underset{\sim}{1}'\right)\left(\underset{\sim}{X} - \frac{1}{n} \underset{\sim}{X}\underset{\sim}{1}\underset{\sim}{1}'\right)'$$

$$= \underset{\sim}{X}\left(\underset{\sim}{I} - \frac{1}{n} \underset{\sim}{X}\underset{\sim}{1}\underset{\sim}{1}'\right)\underset{\sim}{X}'$$

$$\text{Karena } \left(\underset{\sim}{I} - \frac{1}{n}\underset{\sim}{1}\,\underset{\sim}{1}'\right)\left(\underset{\sim}{I} - \frac{1}{n}\underset{\sim}{1}\,\underset{\sim}{1}'\right)' = I - \frac{1}{n}\underset{\sim}{1}\,\underset{\sim}{1}' - \frac{1}{n}\underset{\sim}{1}\,\underset{\sim}{1}' + \frac{1}{n^2}\underset{\sim}{1}\,\underset{\sim}{1}'\,\underset{\sim}{1}\,\underset{\sim}{1}'$$

$$= I - \frac{1}{n}\underset{\sim}{1}\,\underset{\sim}{1}'$$

Jadi, $\underset{\sim}{\bar{x}} = \frac{1}{n}\underset{\sim}{X}'\underset{\sim}{1}$

$$\underset{\sim}{S} = \frac{1}{n-1}\underset{\sim}{X}\left(\underset{\sim}{I} - \frac{1}{n}\underset{\sim}{1}\,\underset{\sim}{1}'\right)\underset{\sim}{X}'$$

Tampak bahwa $\underset{\sim}{\bar{x}}$ dan $\underset{\sim}{S}$ merupakan hasil operasi matriks $\underset{\sim}{X}$.

$$\underset{p\times p}{\underset{\sim}{D}^{1/2}} = \begin{bmatrix} \sqrt{S_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{S_{22}} & \cdots & 0 \\ 0 & \cdots & \cdots & \sqrt{S_{pp}} \end{bmatrix}$$

$$\underset{p\times p}{\underset{\sim}{D}^{-\frac{1}{2}}} = \begin{bmatrix} \frac{1}{\sqrt{S_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{S_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & \cdots & \cdots & \frac{1}{\sqrt{S_{pp}}} \end{bmatrix}$$

$$\underset{\sim}{R} = \begin{bmatrix} \frac{S_{11}}{\sqrt{S_{11}}\sqrt{S_{11}}} & \frac{S_{12}}{\sqrt{S_{11}}\sqrt{S_{22}}} & \cdots & \frac{S_{1p}}{\sqrt{S_{11}}\sqrt{S_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{S_{1p}}{\sqrt{S_{11}}\sqrt{S_{pp}}} & \frac{S_{2p}}{\sqrt{S_{22}}\sqrt{S_{pp}}} & \cdots & \frac{S_{pp}}{\sqrt{S_{pp}}\sqrt{S_{pp}}} \end{bmatrix} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{2p} & \cdots & 1 \end{bmatrix}$$

$$\underset{\sim}{R} = \underset{\sim}{D}^{-\frac{1}{2}}\underset{\sim}{S}\underset{\sim}{D}^{-\frac{1}{2}}$$

$$\underset{\sim}{S} = \underset{\sim}{D}^{\frac{1}{2}}\underset{\sim}{R}\underset{\sim}{D}^{\frac{1}{2}}$$

Mean dari kombinasi linera p variabel

$$\underset{\sim}{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

$$\underset{\sim}{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}$$

mean sampel dari $\underset{\sim}{c}'\underset{\sim}{x} = \dfrac{c'\underset{\sim}{x}_1 + c'\underset{\sim}{x}_2 + \ldots + c'\underset{\sim}{x}_n}{n}$

$$= c'\left(x_1 + \ldots + x_n\right)\frac{1}{n} = \underset{\sim}{c}'\underset{\sim}{\overline{x}}$$

Variansi sampel dari $\underset{\sim}{c}'\underset{\sim}{x}$

$$= \frac{\left(\underset{\sim}{c}'\underset{\sim}{x}_1 - \underset{\sim}{c}'\underset{\sim}{\overline{x}}\right)^2 + \ldots + \left(\underset{\sim}{c}'\underset{\sim}{x}_n - \underset{\sim}{c}'\underset{\sim}{\overline{x}}\right)^2}{n-1}$$

$$= \frac{\underset{\sim}{c}'\left(\underset{\sim}{x}_1 - \overline{x}\right)\left(\underset{\sim}{x}_1 - \overline{x}\right)'\underset{\sim}{c} + \ldots + c'\left(\underset{\sim}{x}_n - \underset{\sim}{\overline{x}}\right)\left(\underset{\sim}{x}_n - \underset{\sim}{\overline{x}}\right)'\underset{\sim}{c}}{n-1}$$

$$= \underset{\sim}{c}'\left|\frac{\left(\underset{\sim}{x}_1 - \underset{\sim}{\overline{x}}\right)\left(\underset{\sim}{x}_1 - \underset{\sim}{x}\right)' + \ldots + \left(\underset{\sim}{x}_n - \underset{\sim}{\overline{x}}\right)\left(\underset{\sim}{x}_n - \underset{\sim}{\overline{x}}\right)'}{n-1}\right|\underset{\sim}{c}$$

$$= \underset{\sim}{c}'\underset{\sim}{S}\underset{\sim}{c}$$

Bila $\underset{\sim}{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_p \end{bmatrix}$

Kovariansi sampel dari $\underset{\sim}{b}'\underset{\sim}{x}$ dan $\underset{\sim}{c}'\underset{\sim}{x}$

$$= \frac{\left(\underset{\sim}{b}'\underset{\sim}{x}_1 - \underset{\sim}{b}'\overline{\underset{\sim}{x}}\right)^2 \left(\underset{\sim}{c}'\underset{\sim}{x}_1 - \underset{\sim}{c}'\overline{\underset{\sim}{x}}\right) + ... + \left(\underset{\sim}{b}'\underset{\sim}{x}_n - \underset{\sim}{b}'\overline{\underset{\sim}{x}}\right)\left(\underset{\sim}{c}'\underset{\sim}{x}_n - \underset{\sim}{c}'\overline{\underset{\sim}{x}}\right)}{n-1}$$

$$= \frac{\underset{\sim}{b}'\left(\underset{\sim}{x}_1 - \overline{\underset{\sim}{x}}\right)\left(\underset{\sim}{x}_1 - \overline{\underset{\sim}{x}}\right)' c + ... + \underset{\sim}{b}'\left(\underset{\sim}{x}_n - \overline{\underset{\sim}{x}}\right)\left(\underset{\sim}{x}_n - \overline{\underset{\sim}{x}}\right)' \underset{\sim}{c}}{n-1}$$

$$= \underset{\sim}{b}' \left| \frac{\left(\underset{\sim}{x}_1 - \overline{\underset{\sim}{x}}\right)\left(\underset{\sim}{x}_1 - \overline{\underset{\sim}{x}}\right)' + ... + \left(\underset{\sim}{x}_n - \overline{\underset{\sim}{x}}\right)\left(\underset{\sim}{x}_n - \overline{\underset{\sim}{x}}\right)'}{n-1} \right| \underset{\sim}{c}$$

Kombinasi linear $\underset{\sim}{b}'\underset{\sim}{X} = b_1 X_1 + b_2 X_2 + ... + b_p X_p$

$\underset{\sim}{c}'\underset{\sim}{X} = c_1 X_1 + c_2 X_2 + ... + c_p X_p$

mean sampel dari $\underset{\sim}{b}'\underset{\sim}{X} = \underset{\sim}{b}'\overline{\underset{\sim}{x}}$

mean sampel dari $\underset{\sim}{c}'\underset{\sim}{X} = \underset{\sim}{c}'\overline{\underset{\sim}{x}}$

Variansi sampel dari $\underset{\sim}{b}'\underset{\sim}{X} = \underset{\sim}{b}'\underset{\sim}{S}\underset{\sim}{b}$

Variansi sampel dari $\underset{\sim}{c}'\underset{\sim}{X} = c'\underset{\sim}{S}c$

Kovariansi sampel dari $\underset{\sim}{b}'\underset{\sim}{X}$ dan $\underset{\sim}{c}'\underset{\sim}{X} = \underset{\sim}{b}'\underset{\sim}{S}\underset{\sim}{c}$

Bila $\underset{\sim}{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{q1} & \cdots & \cdots & a_{qp} \end{bmatrix}$

mean sampel dari $\underset{\sim}{A}\underset{\sim}{X}$ adalah $\underset{\sim}{A}\overline{\underset{\sim}{x}}$

Variansi sampel dari $\underset{\sim}{A}\underset{\sim}{X}$ adalah $\underset{\sim}{A}\underset{\sim}{S}\underset{\sim}{A}'$

Contoh 3.6

Dari 25 perusahaan diukur jumlah modal (x_1) dan keuntungan kotor (x_2) masing-masing dalam puluhan ribu rupiah.

Diketahui matriks kovariansi S.

$$\underset{\sim}{S} = \begin{bmatrix} 14.808 & 14.213 \\ 14.213 & 15.538 \end{bmatrix}$$

Hitung variansi sampel tergeneralisasi.

Penyelesaian:
$|\underset{\sim}{S}| = (14.808)(15.538) - (14.213)(14.213) = 28,08 \times 10^6$

Contoh 3.7

Diketahui matriks $\underset{\sim}{X} = \begin{bmatrix} 1 & 4 & 4 \\ 2 & 1 & 0 \\ 5 & 6 & 4 \end{bmatrix}$

Tunjukkan bahwa $|\underset{\sim}{S}| = 0$ dan tunjukkan adanya dependensi antara vektor-vektor residual.

Penyelesaian:
$\underset{\sim}{\bar{x}}' = \begin{bmatrix} 3 & 1 & 5 \end{bmatrix}$

$$\underset{\sim}{X} - \underset{\sim}{\bar{x}}\underset{\sim}{1}' = \begin{bmatrix} 1-3 & 4-3 & 4-3 \\ 2-1 & 1-1 & 0-1 \\ 5-5 & 6-5 & 4-5 \end{bmatrix} = \begin{bmatrix} -2 & 1 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}$$

$\underset{\sim}{e}_1' = \begin{bmatrix} -2 & 1 & 1 \end{bmatrix}$

$\underset{\sim}{e}_2' = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}$

$\underset{\sim}{e}_3' = \begin{bmatrix} 0 & 1 & -1 \end{bmatrix}$

$\underset{\sim}{e}_3 = \underset{\sim}{e}_1 + 2\underset{\sim}{e}_2$ berarti ada dependensi vektor-vektor baris (juga vektor kolom). Hal ini berarti bahwa $\underset{\sim}{e}_3$ terletak pada bidang yang dibentuk oleh $\underset{\sim}{e}_1$ dan $\underset{\sim}{e}_2$. Sehingga volume ruang yang dihasilkan oleh $\underset{\sim}{e}_1$, $\underset{\sim}{e}_2$, dan $\underset{\sim}{e}_3$ sama dengan nol atau $|\underset{\sim}{S}| = 0$.

Hal ini juga dapat ditunjukkan sebagai berikut:

$$\underset{\sim}{S} = \begin{bmatrix} 3 & -3/2 & 0 \\ -3/2 & 1 & 1/2 \\ 0 & 1/2 & 1 \end{bmatrix}$$

$$|\underset{\sim}{S}| = (3)\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}(-1)^2 + \left(-\frac{3}{2}\right)\begin{bmatrix} -\frac{3}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix}(-1)^2 + (0)\begin{bmatrix} -\frac{3}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix}(-1)^2$$

$$= 3\left(1-\frac{1}{4}\right)+\left(\frac{3}{2}\right)\left(-\frac{3}{2}-0\right)+0 = \frac{9}{4}-\frac{9}{4} = 0$$

Catatan:

Dalam analisa statistik $|\underset{\sim}{S}| = 0$ berarti bahwa pengukuran pada suatu atau beberapa variabel harus dihilangkan. Sehingga Data matriks baru mempunyai variansi sampel tergeneralisasi tidak sama dengan nol.

Contoh 3.8

$$\underset{\sim}{S} = \begin{bmatrix} 4 & 3 & 1 \\ 3 & 9 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

Tunjukkan bahwa $|S| = S_{11}S_{22}S_{33}|\underset{\sim}{R}|$

Penyelesaian:

$S_{11} = 4$

$S_{22} = 9$

$S_{33} = 1$

$$\underset{\sim}{R} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{2}{3} \\ \frac{1}{2} & \frac{2}{3} & 1 \end{bmatrix}$$

$$|\underset{\sim}{S}| = 4\begin{bmatrix} 9 & 2 \\ 2 & 1 \end{bmatrix}(-1)^2 + 3\begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix}(-1)^2 + 1\begin{bmatrix} 3 & 9 \\ 1 & 2 \end{bmatrix}(-1)^4$$

$$= 4(9-4) - 3(3-2) + 1(6-9) = 14$$

$$|\underset{\sim}{R}| = 1\begin{bmatrix} 1 & \frac{2}{3} \\ \frac{2}{3} & 1 \end{bmatrix}(-1)^2 + \frac{1}{2}\begin{bmatrix} \frac{1}{2} & \frac{2}{3} \\ \frac{1}{2} & 1 \end{bmatrix}(-1)^3 + \frac{1}{2}\begin{bmatrix} \frac{1}{2} & 1 \\ \frac{1}{2} & \frac{2}{3} \end{bmatrix}(-1)^4$$

$$= \left(1-\frac{4}{9}\right)-\left(\frac{1}{2}\right)\left(\frac{1}{2}-\frac{1}{3}\right)+\left(\frac{1}{2}\right)\left(\frac{1}{3}-\frac{1}{2}\right)=\frac{7}{18}$$

$$S_{11}S_{22}S_{33}\left|\underset{\sim}{R}\right| = (4)(9)(1)\left(\frac{7}{18}\right)=14$$

Jadi, $\left|\underset{\sim}{S}\right| = S_{11}S_{22}S_{33}\left|\underset{\sim}{R}\right|$

Contoh 3.9

Hitung total variansi sampel dari

1. $\underset{\sim}{S}_1 = \begin{bmatrix} 14.808 & 14.213 \\ 14.213 & 15.538 \end{bmatrix}$
2. $\underset{\sim}{S}_2 = \begin{bmatrix} 3 & -3/2 & 0 \\ -3/2 & 1 & 1/2 \\ 0 & 1/2 & 1 \end{bmatrix}$

Penyelesaian:

1. Total variansi sampel $= S_{11} + S_{22} = 14.808 + 15.538 = 30.346$
2. Total variansi sampel $= S_{11} + S_{22} + S_{33}$
   $= 3 + 1 + 1 = 5$

Contoh 3.10

$$\underset{\sim}{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} = \begin{bmatrix} 1 & 4 & 4 \\ 2 & 1 & 0 \\ 5 & 6 & 4 \end{bmatrix}$$

misalkan dua kombinasi linear berikut:

$$\underset{\sim}{b}'\underset{\sim}{X} = \begin{bmatrix} 2 & 2 & -1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = 2X_1 + 2X_2 - X_3$$

$$\underset{\sim}{c}'\underset{\sim}{X} = \begin{bmatrix} 1 & -1 & 3 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = X_1 - X_2 - 3X_3$$

Tunjukkan bahwa mean sampel dari $\underset{\sim}{b}'\underset{\sim}{X} = \underset{\sim}{b}'\overline{\underset{\sim}{x}}$

Penyelesaian:

$$\underset{\sim}{b}'\underset{\sim}{x}_1 = 2x_{11} + 2x_{21} - x_{31} = 2(1) + 2(2) - (5) = 1$$
$$\underset{\sim}{b}'\underset{\sim}{x}_2 = 2x_{12} + 2x_{22} - x_{32} = 2(4) + 2(1) - (6) = 4$$
$$\underset{\sim}{b}'\underset{\sim}{x}_3 = 2x_{13} + 2x_{23} - x_{33} = 2(4) + 2(0) - (4) = 4$$

mempunyai mean sampel $= \dfrac{1+4+4}{3} = 3$ dan

variansi sampel $\dfrac{1}{3-1}\left((1-3)^2 + (4-3)^2 + (4-3)^2\right) = 3$

Anda dapat memeriksa bahwa:

$$\underset{\sim}{\overline{X}} = \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix} \text{dan } \underset{\sim}{S} = \begin{bmatrix} 3 & -3/2 & 0 \\ -3/2 & 1 & 1/2 \\ 0 & 1/2 & 1 \end{bmatrix}$$

$$\underset{\sim}{b}'\underset{\sim}{\overline{x}} = \begin{bmatrix} 2 & 2 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix} = 3$$

Jadi, mean sampel dari $\underset{\sim}{b}'\underset{\sim}{X}$ adalah $\underset{\sim}{b}'\underset{\sim}{\overline{x}}$

$$\underset{\sim}{b}'\underset{\sim}{S}\underset{\sim}{b} = \begin{bmatrix} 2 & 2 & -1 \end{bmatrix} \begin{bmatrix} 3 & -3/2 & 0 \\ -3/2 & 1 & 1/2 \\ 0 & 1/2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}$$
$$= \begin{bmatrix} 2 & 2 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ -\frac{3}{2} \\ 0 \end{bmatrix} = 3$$

Jadi, variansi dari $\underset{\sim}{b}'\underset{\sim}{X}$ adalah $\underset{\sim}{b}'\underset{\sim}{S}\underset{\sim}{b}$

Kovariansi sampel dihitung dari pasangan
$(\underset{\sim}{b}'\underset{\sim}{x}_1, \underset{\sim}{c}'\underset{\sim}{x}_1), (\underset{\sim}{b}'\underset{\sim}{x}_2, \underset{\sim}{c}'\underset{\sim}{x}_2)$ dan $(\underset{\sim}{b}'\underset{\sim}{x}_3, \underset{\sim}{c}'\underset{\sim}{x}_3)$

Kovariansi sampel $= \dfrac{(1-3)(14-17) + (4-3)(21-17) + (4-3)(16-17)}{3-1} = \dfrac{9}{2}$

$$\underset{\sim}{b}'\underset{\sim}{S}\underset{\sim}{c} = \begin{bmatrix} 2 & 2 & -1 \end{bmatrix} \begin{bmatrix} 3 & -3/2 & 0 \\ -3/2 & 1 & 1/2 \\ 0 & 1/2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 2 & -1 \end{bmatrix} \begin{bmatrix} \frac{9}{2} \\ -1 \\ \frac{5}{2} \end{bmatrix} = \frac{9}{2}$$

Jadi, kovariansi sampel $= \underset{\sim}{b}'\underset{\sim}{S}\underset{\sim}{c}$.

LATIHAN

Untuk memperdalam pemahaman Anda mengenai materi di atas, kerjakanlah latihan berikut!

1) Diketahui data matriks

$$\underset{\sim}{X} = \begin{bmatrix} -1 & 2 & 5 \\ 3 & 4 & 2 \\ -2 & 2 & 3 \end{bmatrix}$$

a) Hitung matriks residual $\underset{\sim}{X} - \underset{\sim}{\bar{x}}\underset{\sim}{1}'$
b) Apakah matriks ini full rank?
c) Hitung $\underset{\sim}{S}$ dan hitung variansi sampel tergeneralisasi $|\underset{\sim}{S}|$!
d) Hitung total variansi sampel!

2) Diberikan

$$\underset{\sim}{S} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ dan } \underset{\sim}{S} = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & 1 \end{bmatrix}$$

a) Hitung total variansi sampel kedua S!
b) Hitung variansi sampel tergeneralisasi untuk kedua S!
c) Bandingkan hasil a dan b.

3) Buktikan bahwa $\left|\underset{\sim}{S}\right| = S_{11}S_{22}...S_{pp}\left|\underset{\sim}{R}\right|$!

Petunjuk Jawaban Latihan

1) a) Hitung $\underset{\sim}{X} - \underset{\sim}{\overline{x}}\underset{\sim}{1}'$. Apakah baris-barisnya independen?

   b) $\underset{\sim}{S} = \left(\underset{\sim}{X} - \underset{\sim}{\overline{x}}\underset{\sim}{1}'\right)\left(\underset{\sim}{X} - \underset{\sim}{\overline{x}}\underset{\sim}{1}'\right)'$

   c. Total variansi sampel $= S_{11} + S_{22} + S_{33}$.

2) a) Hitung total variansi sampel $= S_{11} + S_{22} + S_{33}$.

   b) Hitung variansi sampel tergeneralisasi.

3) $\underset{\sim}{S} = \underset{\sim}{D}^{\frac{1}{2}}\underset{\sim}{R}\underset{\sim}{D}^{\frac{1}{2}}$

   $\left|\underset{\sim}{S}\right| = \left|\underset{\sim}{D}^{\frac{1}{2}}\right|\left|\underset{\sim}{R}\right|\left|\underset{\sim}{D}^{\frac{1}{2}}\right|$

   Hitung $\left|\underset{\sim}{D}^{\frac{1}{2}}\right|$

RANGKUMAN

1. $\underset{\sim}{X} - N_p\left(\underset{\sim}{\mu}, \underset{\sim}{\Sigma}\right)$

   $\tilde{\mu} = \overline{\underset{\sim}{X}}$

   $\tilde{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}\left(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}}\right)\left(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}}\right)'$

   Merupakan penduga likelihood maksimum untuk $\underset{\sim}{\mu}$ dan $\underset{\sim}{\Sigma}$.

2. $E\left(\overline{\underset{\sim}{X}}\right) = \underset{\sim}{\mu}$

   $E\left(\tilde{\Sigma}\right) = \frac{n-1}{n}\underset{\sim}{\Sigma}$

   $Cov\left(\overline{\underset{\sim}{X}}\right) = \frac{1}{n}\underset{\sim}{\Sigma}$

3. Variansi sampel tergeneralisasi $= |\underset{\sim}{S}|$

   Di mana $\underset{\sim}{S} = \frac{1}{n-1}\sum_{j=1}^{n}\left(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}}\right)\left(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}}\right)'$

4. Bila $n \leq p$ $|\underset{\sim}{S}| = 0$ untuk semua sampel
5. Variansi sampel tergeneralisasi
   Untuk variabel standarisasi $= |\underset{\sim}{R}|$
6. Total variansi sampel $= S_{11} + ... + S_{pp}$
7. $\overline{\underset{\sim}{X}} = \frac{1}{n}\underset{\sim}{X}\underset{\sim}{1}$

$$(n-1)\underset{\sim}{S} = \left(\underset{\sim}{X} - \frac{1}{n}\underset{\sim}{X}\underset{\sim}{1}\underset{\sim}{1}'\right)\left(\underset{\sim}{X} - \frac{1}{n}\underset{\sim}{X}\underset{\sim}{1}\underset{\sim}{1}'\right)'$$

$$= \underset{\sim}{X}\left(\underset{\sim}{I} - \frac{1}{n}\underset{\sim}{X}\underset{\sim}{1}\underset{\sim}{1}'\right)\underset{\sim}{X}'$$

8. $\underset{p\times p}{\underset{\sim}{D}^{1/2}} = \begin{bmatrix} \sqrt{S_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{S_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \sqrt{S_{pp}} \end{bmatrix}$

$$\underset{\sim}{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{2p} & \cdots & 1 \end{bmatrix}$$

$$\underset{\sim}{R} = \underset{\sim}{D}^{-\frac{1}{2}}\underset{\sim}{S}\underset{\sim}{D}^{-\frac{1}{2}}$$

$$\underset{\sim}{S} = \underset{\sim}{D}^{\frac{1}{2}}\underset{\sim}{R}\underset{\sim}{D}^{\frac{1}{2}}$$

TES FORMATIF 2

Pilihlah satu jawaban yang paling tepat!

$$\underset{\sim}{X} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 11 & 12 & 13 & 14 & 15 \\ 21 & 22 & 23 & 24 & 25 \end{bmatrix}$$

1) $\underset{\sim}{S} = \ldots.$

   A. $\begin{bmatrix} 2,5 & 2,5 & 2,5 \\ & 2,5 & 2,5 \\ & & 2,5 \end{bmatrix}$

   B. $\begin{bmatrix} 1,58 & 2,5 & 2,5 \\ & 15,8 & 250 \\ & & 158 \end{bmatrix}$

   C. $\begin{bmatrix} 1,58 & 2,5 & 2,5 \\ & 1,58 & 2,5 \\ & & 1,58 \end{bmatrix}$

   D. $\begin{bmatrix} 1,58 & 0,25 & 2,5 \\ & 1,58 & 25 \\ & & 1,58 \end{bmatrix}$

2) $\left|\underset{\sim}{S}\right| = \ldots.$

   A. 0
   B. 13,5
   C. 57,6
   D. 59,8

3) Total variansi sampel = ….

   A. 175,380
   B. 47,400
   C. 7,500
   D. 0,474

4) Variansi sampel tergeneralisasi = ….
   A. 13,5
   B. 57,6
   C. 58,7
   D. A, B, C tak ada yang benar

5) $\left|\underset{\sim}{R}\right| = \ldots.$
   A. 0
   B. 1,412
   C. 2,019
   D. 2,375

Cocokkanlah jawaban Anda dengan Kunci Jawaban Tes Formatif 2 yang terdapat di bagian akhir modul ini. Hitunglah jawaban yang benar. Kemudian, gunakan rumus berikut untuk mengetahui tingkat penguasaan Anda terhadap materi Kegiatan Belajar 2.

$$\text{Tingkat penguasaan} = \frac{\text{Jumlah Jawaban yang Benar}}{\text{Jumlah Soal}} \times 100\%$$

Arti tingkat penguasaan: 90 - 100% = baik sekali
80 - 89% = baik
70 - 79% = cukup
< 70% = kurang

Apabila mencapai tingkat penguasaan 80% atau lebih, Anda dapat meneruskan dengan modul selanjutnya. **Bagus!** Jika masih di bawah 80%, Anda harus mengulangi materi Kegiatan Belajar 2, terutama bagian yang belum dikuasai.

Kunci Jawaban Tes Formatif

Tes Formatif 1

1) A
2) A
3) D
4) B
5) D

Tes Formatif 2

1) A
2) A
3) C
4) D
5) A

Daftar Pustaka

Johnson, R. A., Wichern, D. W. (1982). *Applied multivariate statistical analysis.* Prentice Hall Inc.

Rencher, A. C. & Christensen, W. F. (2012). *Methods of multivariate analysis*. Third Edition. New Jersey: John Wiley & Sons, Inc.

Tabachnick, B. G., and Fidell, L. S. (2007). *Using multivariate statistics* (5^{th} ed.). Boston: Pearson Education, Inc.