I. Interpretasi Geometri dan Sampel Acak

A. Interpretasi Geometri dari Sampel

Data sampel kami terlihat dalam susunan matriksseperti berikut ini :

$$\underset{(pxn)}{\mathbf{X}} = \begin{bmatrix} \mathbf{x}_{11} & \mathbf{x}_{12} & \cdots & \mathbf{x}_{1n} \\ \mathbf{x}_{21} & \mathbf{x}_{22} & \cdots & \mathbf{x}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{n1} & \mathbf{x}_{n2} & \cdots & \mathbf{x}_{pn} \end{bmatrix} = \left[\underset{\substack{\uparrow \\ \text{Pengamatan ke} \\ \text{1 (multivariate)}}}{\mathbf{x}_1}, \mathbf{x}_2, \cdots, \underset{\substack{\uparrow \\ \text{Pengamatan ke} \\ \text{n (multivariate)}}}{\mathbf{x}_n} \right]$$

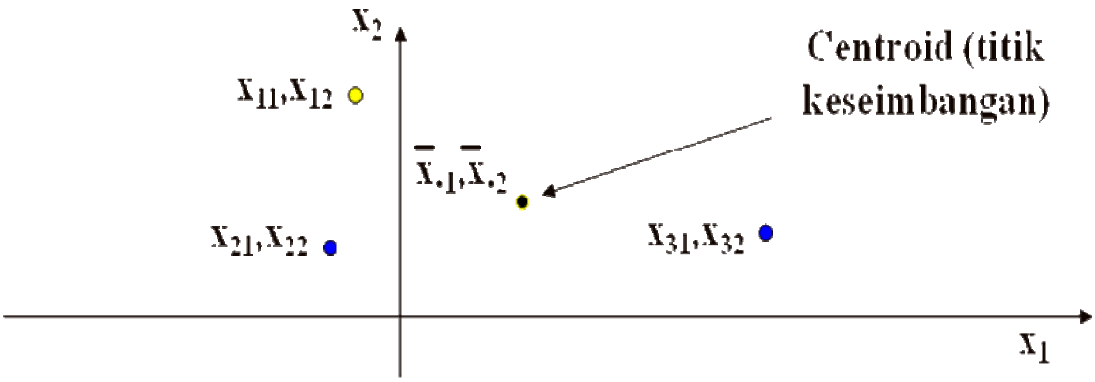
Dalam Beberapa buku juga menuliskan sebagai berikut:

$$\underset{(nxp)}{\mathbf{X}} = \begin{bmatrix} \mathbf{x}_{11} & \mathbf{x}_{12} & \cdots & \mathbf{x}_{1p} \\ \mathbf{x}_{21} & \mathbf{x}_{22} & \cdots & \mathbf{x}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{n1} & \mathbf{x}_{n2} & \cdots & \mathbf{x}_{np} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{\mathbf{x}}'_1 \\ \underset{\sim}{\mathbf{x}}'_2 \\ \vdots \\ \underset{\sim}{\mathbf{x}}'_n \end{bmatrix} \begin{matrix} \text{Separate} \\ \text{multivariate} \\ \text{observations} \end{matrix}$$

Penyebaran dari titik n dalam ruang dimensi p menunjukkan data pada lokasi dan variabelnya. Jika titik dianggap penuh, rataan sampel vektor $\bar{X}$ adalah pusat keseimbangan. Untuk contoh diberikan 2 variabel dan 3 pengamatan:

$$\underset{(pxn)}{\mathbf{X}} = \begin{bmatrix} \mathbf{x}_{11} & \mathbf{x}_{12} & \mathbf{x}_{13} \\ \mathbf{x}_{21} & \mathbf{x}_{22} & \mathbf{x}_{23} \end{bmatrix}$$

Kita



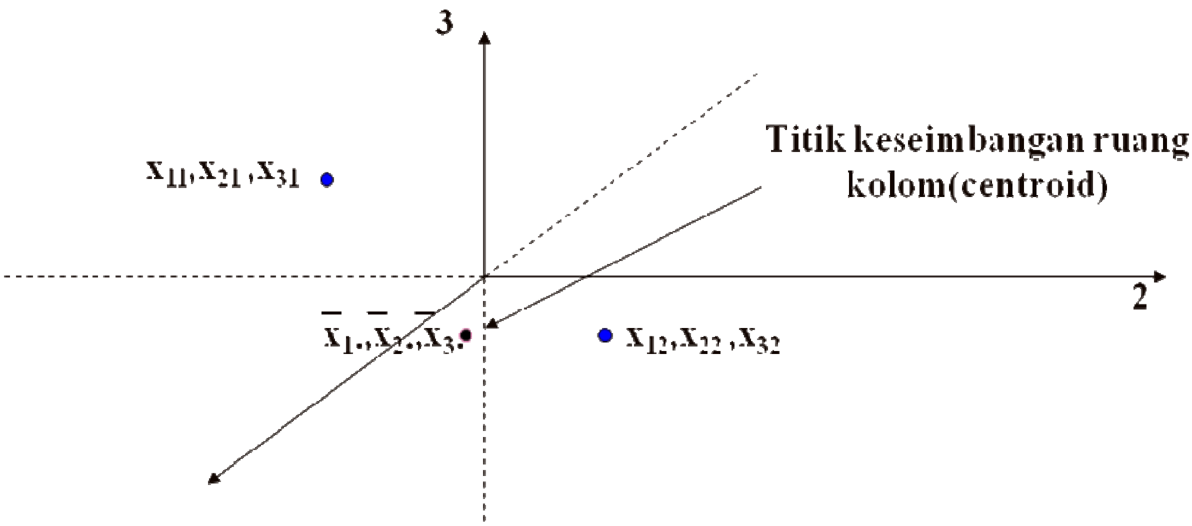
punya

p = 2 variabel atau ruang baris (karena baris dianggap sebagai koordinat vektor)

Dengan data yang sama yaitu:

$$\underset{(pxn)}{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix}$$

pemetaan pada item atau ruang kolom, akan terlihat seperti gambar dibawah ini:



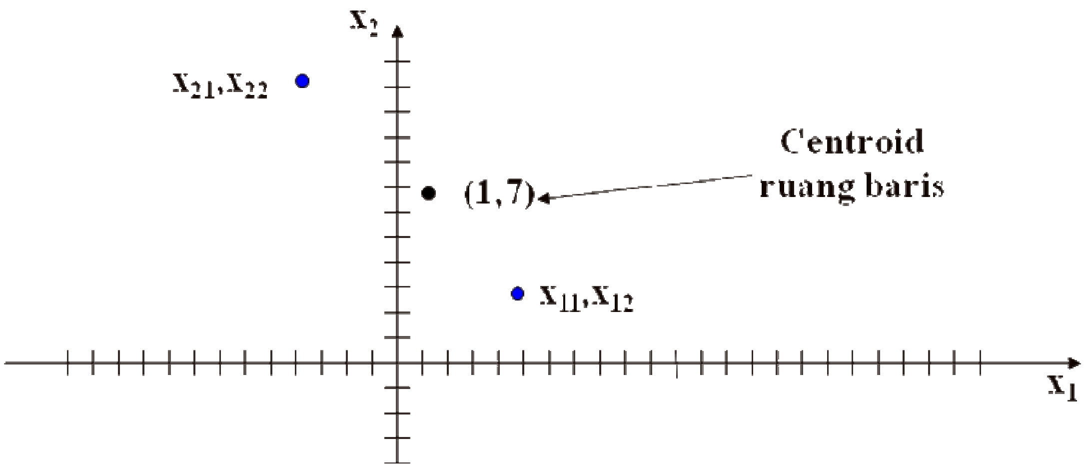
Ini dianggap sebagai ruang kolom karena kolomnya dianggap sebagai koordinat vektor

Contoh:

Hitunglah rataan vektor dari $\bar{x}$ data matriks X. Dengan n=2 dan p=2 dan lokasi x dapat dilihat dari gambar berikut:

$$\underset{\sim}{\mathbf{X}} = \begin{bmatrix} 5 & -3 \\ 3 & 11 \end{bmatrix}$$

ruang baris, dengan *p*=2, dapat digambarkan sebagai berikut:



rataan vektor (Centroid) $\bar{x}$ adalah (1,7)

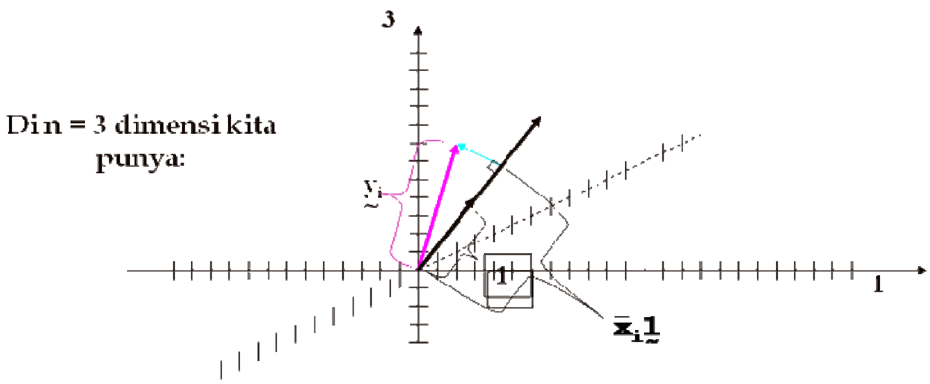
Cara menghitung titik keseimbangan atau centroid ruang baris:

$$\overline{X} = \begin{bmatrix} \frac{5-3}{2} \\ \frac{3+11}{2} \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \end{bmatrix}$$

Soal:

Dengan menggunakan data yang sama. Hitunglah rataan vektor dari $\bar{x}$ data matriks X dengan menggunakan ruang kolom!

ruang dimensi memperlihatkan suatu penarikan interpretasi geometri dari titik keseimbangan (centroid) kita mulai dengan pendefinisian n x 1 vektor 1:



Ini vektor bentuk nyata persamaan sudut dengan tiap koordinat axis – ini berarti normalisasi dari vektor hasil

$$\left(\frac{1}{\sqrt{n}}\right)\underset{\sim}{1}$$

Ini sering dianggap sebagai vektor *deviasi* dan didefinisikan dengan:

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

Contoh: Misalkan dengan menggunakan dekomposisi dari yi ke xi 1 dan ei= yi – xi1, i=1,2, untuk data X, dengan *p* = 2 vektor dan *n* = 3 dimensi :

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

Dengan: x1 = (4 – 1 + 3)/3 = 2 dan x2 = (1 + 3 + 5)/3 = 3, jadi

$$\bar{X}_1 = 2\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} \qquad \bar{X}_2 = 3\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$$

Akibatnya,

$$e_1 = y_1 - \bar{x}_1 1 = \begin{bmatrix} 4 \\ -1 \\ 3 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix} \qquad e_2 = y_2 - \bar{x}_2 1 = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} - \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix}$$

Kita catat bahwa x11 dan e1 = y1 - x11 tegak lurus, karena

$$(\bar{x}_1)'(y_1 - \bar{x}_1 1) = \begin{bmatrix} 2 & 2 & 2 \end{bmatrix}\begin{bmatrix} 2 \\ -3 \\ 3 \end{bmatrix} = 4 - 6 + 2 = 0$$

Dengan cara yang sama dapat dicari x21 dan e2 = y2 – x21. Sehingga dekomposisinya adalah

$$y_1 = \begin{bmatrix} 4 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix} \qquad y_2 = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} + \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix}$$

Kuadrat panjang vektor deviasi:

$$\underbrace{L^2_{e_i} = e_i' e_i}_{\text{Kuadrat panjang vektor deviasi}} = \underbrace{\sum_{j=1}^{n} \left(x_{ij} - \bar{x}_i \right)^2}_{\text{Jumlah kuadrat panjang}}$$

Mengingat kembali varians sampel:

$$s_i^2 = \frac{\sum_{j=1}^{n} \left(x_{ji} - \bar{x}_i \right)^2}{n - 1}$$

Kita dapat melihat kuadrat panjang dari suatu variabel vektor deviasi adalah proporsional dengan varians (sehingga panjang proporsional dengan standard deviasi)!

Untuk dua vektor deviasi ei dan ek:

$$e_i' e_k = \sum_{j=1}^{n} \left(x_{ij} - \bar{x}_i \right) \left(x_{kj} - \bar{x}_k \right) = x' y$$

Misalkan q_{ik} notasi untuk sudut antara dua vektor deviasi. Ingat bahwa:

$$\cos\left(\theta_{xy}\right) = \frac{x' y}{L_x L_y} = \frac{x' y}{\sqrt{x' x}\sqrt{y' y}}$$

Dengan subtitusi diperoleh

$$\mathbf{e}_i'\mathbf{e}_k = L_{\mathbf{e}_i} L_{\mathbf{e}_k} \frac{\mathbf{e}_i'\mathbf{e}_k}{L_{\mathbf{e}_i} L_{\mathbf{e}_k}} = L_{\mathbf{e}_i} L_{\mathbf{e}_k} \cos(\theta_{ik})$$

Subtitusi lainya didasarkan pada:

$$L_{\mathbf{e}_i}^2 = \mathbf{e}_i'\mathbf{e}_i = \sum_{j=1}^{n} \left(x_{ij} - \bar{x}_i\right)^2$$

Misalkan q_{ik} notasi untuk sudut antara dua vektor deviasi. Ingat bahwa:

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

Dengan subtitusi diperoleh

$$\sum_{j=1}^{n} \left(x_{ij} - \bar{x}_i\right)\left(x_{kj} - \bar{x}_k\right) = \sqrt{\sum_{j=1}^{n} \left(x_{ij} - \bar{x}_i\right)^2}\sqrt{\sum_{j=1}^{n} \left(x_{kj} - \bar{x}_k\right)^2}\cos(\theta_{ik})$$

Dengan demikian

$$\cos(\theta_{ik}) = \frac{\sum_{j=1}^{n} \left(x_{ij} - \bar{x}_i\right)\left(x_{kj} - \bar{x}_k\right)}{\sqrt{\sum_{j=1}^{n} \left(x_{ij} - \bar{x}_i\right)^2}\sqrt{\sum_{j=1}^{n} \left(x_{kj} - \bar{x}_k\right)^2}}$$

$$= \frac{\left(\sum_{j=1}^{n} \left(x_{ij} - \bar{x}_i\right)\left(x_{kj} - \bar{x}_k\right) \big/ n\right)}{\left(\sqrt{\sum_{j=1}^{n} \left(x_{ij} - \bar{x}_i\right)^2 \big/ n}\right)\left(\sqrt{\sum_{j=1}^{n} \left(x_{kj} - \bar{x}_k\right)^2 \big/ n}\right)}$$

$$= \frac{s_{ik}}{\sqrt{s_{ii}}\sqrt{s_{kk}}} = r_{ik}$$

Contoh:

Diberikan vektor deviasi pada contoh sebelumnya, misalkan kita hitung varians sampel dan kovarian matriks Sn, dan korelasi sampel matriks R dengan menggunakan konsep interpretasi geometri.

Dari contoh sebelumnya

$$\mathbf{e_1} = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix} \qquad \mathbf{e_2} = \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix}$$

$$\mathbf{e_1'e_1} = \begin{bmatrix} 2 & -3 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix} = 14 = 3s_{11}$$ atau $s_{11} = 14/3$

$$\mathbf{e_2'e_2} = \begin{bmatrix} -2 & 0 & 2 \end{bmatrix} \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix} = 8 = 3s_{22}$$ atau $s_{22} = 8/3$

$$\mathbf{e_1'e_2} = \begin{bmatrix} 2 & -3 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix} = -2 = 3s_{12}$$ atau $s_{12} = -2/3$

Akibatnya,

$$r_{12} = \frac{s_{12}}{\sqrt{s_{11}}\ \sqrt{s_{22}}} = \frac{-\frac{2}{3}}{\sqrt{\frac{14}{3}}\ \sqrt{\frac{8}{3}}} = -.189$$

dan

$$s_n = \begin{bmatrix} \frac{14}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{8}{3} \end{bmatrix} \qquad R = \begin{bmatrix} 1 & -.189 \\ -.189 & 1 \end{bmatrix}$$