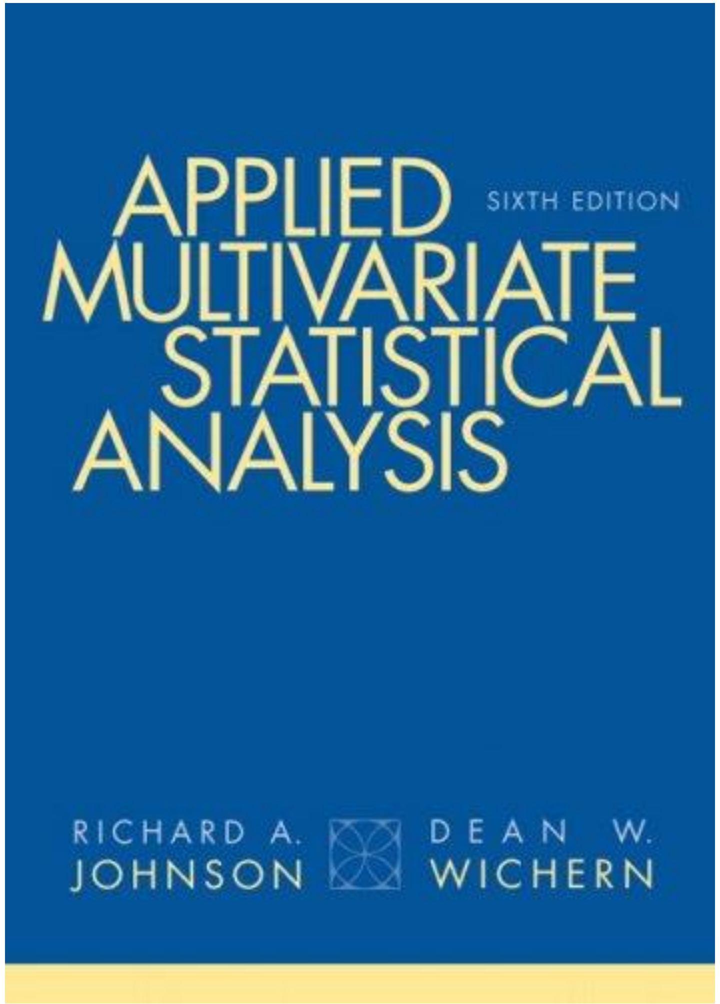
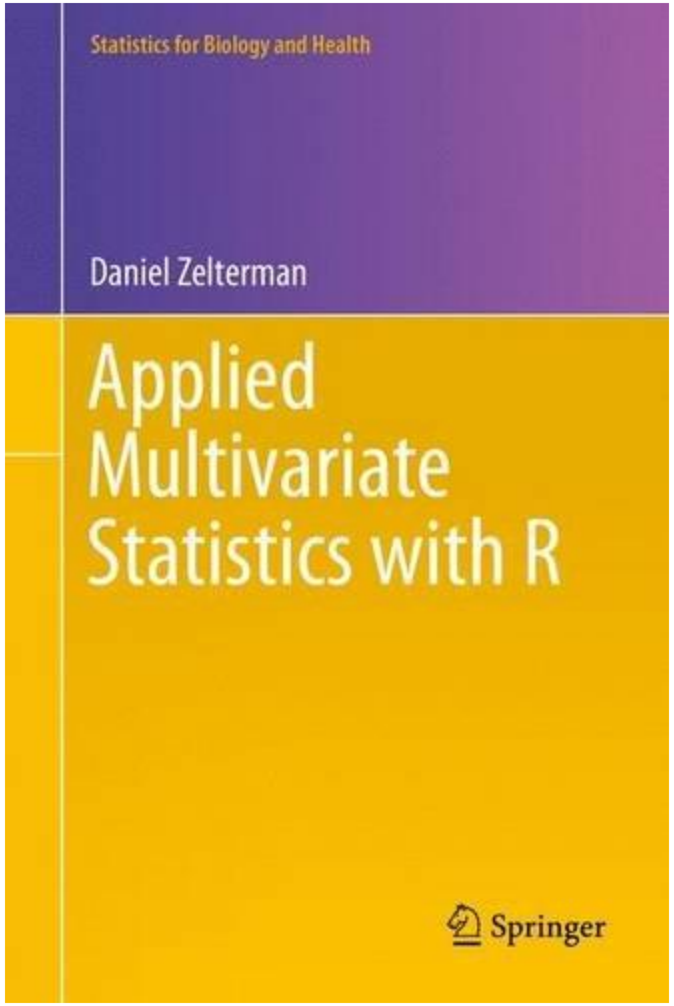
STK342
Teknik Peubah Ganda

Deskripsi MK

- Mata kuliah ini membahas tentang inferensia berdasarkan contoh acak dari sebaran normal ganda, analisis ragam peubah ganda, analisis profil, dan selang kepercayaan simultan. Berbagai teknik pereduksian dimensi akan diberikan seperti analisis komponen utama, analisis faktor, analisis biplot, dan analisis korespondensi. Mencakup juga analisis gerombol berhirarki dan tak berhirarki. Serta pembahasan mengenai analisis diskriminan linier, kuadratik, dan kanonik.

Buku referensi





SIlabus

1. Konsep Dasar Peubah Ganda
2. Sebaran Normal Ganda
3. Inferensia Peubah Ganda
4. Inferensia Peubah Ganda
5. MANOVA
6. Analisis profil
7. Analisis Komponen Utama

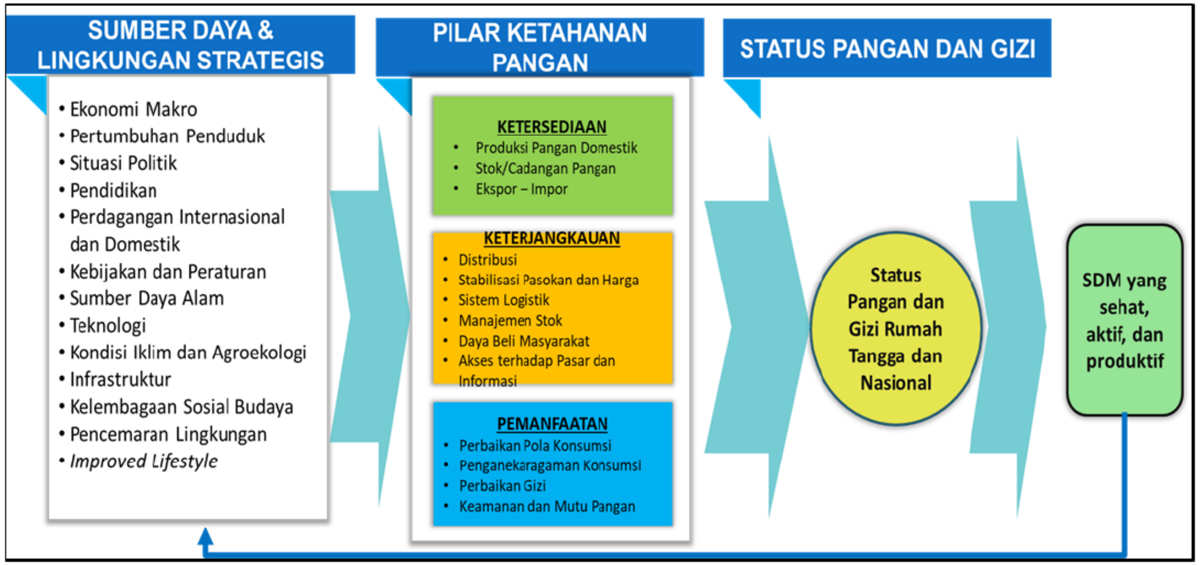
SIlabus

8. Pengenalan Analisis Faktor
9. Pengenalan Analisis Gerombol
10. Pengenalan Analisis Diskriminan
11. Pengenalan Analisis Biplot
12. Pengenalan Analisis Korespondensi
13. Pengenalan Analisis Korelasi Kononik
14. Pengenalan Analisis Penskalaan Dimensi Ganda (MDS)

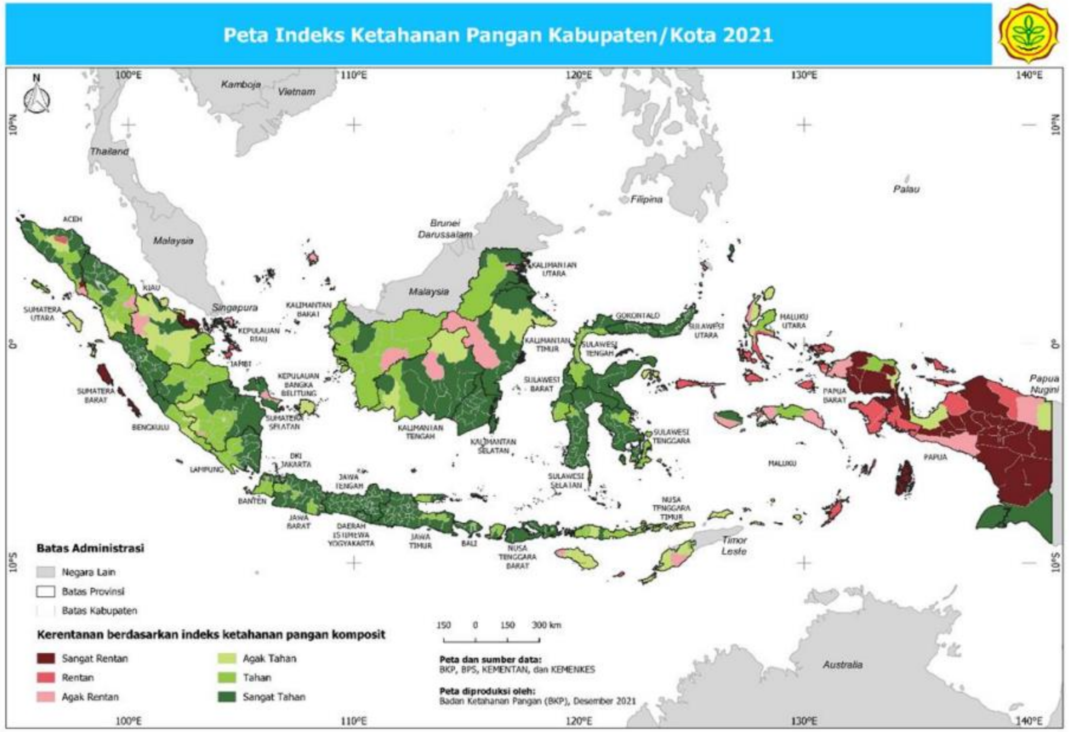
Pengamatan Peubah Ganda

- Menggambarkan suatu objek tidak cukup menggunakan satu peubah saja
- Kasus pengamatan peubah ganda dijumpai di seluruh bidang terapan
- Perlu analisis lebih canggih, jika antar peubah tidak saling bebas

Gambar 1.
Kerangka Konsep Ketahanan Pangan dan Gizi



Gambar 2.
Peta Indeks Ketahanan Pangan Kabupaten dan Kota 2021



Peranan analisis peubah ganda

- Data reduction or structural simplification
- Sorting and grouping
- Investigation of the dependence among variables
- Prediction
- Hypothesis construction and testing

Bagaimana menangani data peubah ganda?

Simple Linear Regression Equation

The simple linear regression equation provides an estimate of the population regression line

Estimated (or predicted) y value for observation i

Estimate of the regression intercept

Estimate of the regression slope

Value of x for observation i

$$\hat{y}_i = b_0 + b_1 x_i$$

The individual random error terms e_i have a mean of zero

$$e_i = (y_i - \hat{y}_i) = y_i - (b_0 + b_1 x_i)$$

Least Squares Estimators

- b_0 and b_1 are obtained by finding the values of b_0 and b_1 that minimize the sum of the squared differences between y and $\hat{y}$:

$$\min \ SSE = \min \sum e_i^2$$

$$= \min \sum (y_i - \hat{y}_i)^2$$

$$= \min \sum [y_i - (b_0 + b_1 x_i)]^2$$

Differential calculus is used to obtain the coefficient estimators b_0 and b_1 that minimize SSE

Least Squares Estimators

(continued)

- The slope coefficient estimator is

$$\boxed{b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = r_{xy}\frac{s_Y}{s_X}}$$

- And the constant or y-intercept is

$$\boxed{b_0 = \bar{y} - b_1\bar{x}}$$

- The regression line always goes through the mean x, y

– –

Regresi berganda

Notasi matriks model

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^\mathsf{T}\boldsymbol\beta + \varepsilon_i, \qquad i = 1, \ldots, n,$$

$$\mathbf{y} = X\boldsymbol\beta + \boldsymbol\varepsilon,$$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

$$X = \begin{pmatrix} \mathbf{x}_1^\mathsf{T} \\ \mathbf{x}_2^\mathsf{T} \\ \vdots \\ \mathbf{x}_n^\mathsf{T} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix},$$

$$\boldsymbol\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Proses pendugaan koefisien regresi

$$\hat{\vec\beta} = \arg\min_{\vec\beta} L\left(D, \vec\beta\right) = \arg\min_{\vec\beta} \sum_{i=1}^{n} \left(\vec\beta \cdot \vec{x_i} - y_i\right)^2$$

$$\begin{aligned} L\left(D, \vec\beta\right) &= \|X\vec\beta - Y\|^2 \\ &= \left(X\vec\beta - Y\right)^\mathsf{T}\left(X\vec\beta - Y\right) \\ &= Y^\mathsf{T}Y - Y^\mathsf{T}X\vec\beta - \vec\beta^\mathsf{T}X^\mathsf{T}Y + \vec\beta^\mathsf{T}X^\mathsf{T}X\vec\beta \end{aligned}$$

$$\begin{aligned} \frac{\partial L\left(D, \vec\beta\right)}{\partial\vec\beta} &= \frac{\partial\left(Y^\mathsf{T}Y - Y^\mathsf{T}X\vec\beta - \vec\beta^\mathsf{T}X^\mathsf{T}Y + \vec\beta^\mathsf{T}X^\mathsf{T}X\vec\beta\right)}{\partial\vec\beta} \\ &= -2X^\mathsf{T}Y + 2X^\mathsf{T}X\vec\beta \end{aligned}$$

$$\begin{aligned} -2X^\mathsf{T}Y + 2X^\mathsf{T}X\vec\beta &= 0 \\ \Rightarrow X^\mathsf{T}X\vec\beta &= X^\mathsf{T}Y \\ \Rightarrow \hat{\vec\beta} &= \left(X^\mathsf{T}X\right)^{-1}X^\mathsf{T}Y \end{aligned}$$

Beberapa Notasi

- Misal p peubah diamati dari n objek
- Skala pengukuran peubah bisa nominal, ordinal, interval, atau rasio
- Dapat berupa peubah bebas maupun peubah tak bebas

Obs	X1	X2	…	Xp
1	x_{11}	x_{21}	…	x_{p1}
2	x_{12}	x_{22}	…	x_{p2}
3	x_{13}	x_{23}	…	x_{p3}
4	x_{14}	x_{24}	…	x_{p4}
…	…	…	…	…
n	x_{1n}	x_{2n}	…	x_{pn}

Notasi berikut khusus untuk peubah-peubah berskala interval atau rasio

- Vektor peubah acak:

$$ {}_p\underline{X}_1 = \begin{bmatrix} X_1 \\ X_2 \\ \ldots \\ X_p \end{bmatrix} $$

- Nilai harapan vektor peubah acak

$$ E\left({}_p\underline{X}_1\right) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \ldots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \ldots \\ \mu_p \end{bmatrix} = \underline{\mu} $$

Matriks Ragam-peragam (Variance Covariance Matrix)

$$Cov(\underline{X}) = {}_p\Sigma_p = \begin{bmatrix} \mathrm{var}(x_1) & \mathrm{cov}(x_1, x_2) & \cdots & \mathrm{cov}(x_1, x_p) \\ \mathrm{cov}(x_2, x_1) & \mathrm{var}(x_2) & \cdots & \mathrm{cov}(x_2, x_p) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(x_p, x_1) & \mathrm{cov}(x_p, x_2) & \cdots & \mathrm{var}(x_p) \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

dimana:

$$Cov(X_i, X_i) = Var(X_i) = \sigma_{ii} = \sigma_i^2 = E\left(X_i - E(X_i)\right)^2$$

$$Cov(X_i, X_j) = \sigma_{ij} = E\left(X_i - E(X_i)\right)\left(X_j - E(X_j)\right)$$

- Matriks korelasi berukuran pxp

$$ {}_pR_p = \begin{bmatrix} 1 & \rho_{12} & & \rho_{1p} \\ \rho_{21} & 1 & & \rho_{2p} \\ & & \ldots & \\ \rho_{p1} & \rho_{p2} & & 1 \end{bmatrix} $$

- Hubungan matriks ragam peragam dengan matriks korelasi

$$ {}_pR_p = D\left(\frac{1}{\sigma_i}\right)\sum D\left(\frac{1}{\sigma_i}\right) $$

- Matriks D, matriks diagonal berukuruan pxp

$$ D\left(\frac{1}{\sigma_i}\right) = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \ldots & 0 \\ 0 & \frac{1}{\sigma_2} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \frac{1}{\sigma_p} \end{bmatrix} $$

Beberapa notasi untuk data sampel

- Vektor rataan berukuran px1, merupakan penduga bagi vektor $\underline{\mu}$

$$_p\underline{\bar{x}}_1 = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \cdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} \dfrac{\sum_{i=1}^{n} x_{1i}}{n} \\ \dfrac{\sum_{i=1}^{n} x_{2i}}{n} \\ \cdots \\ \dfrac{\sum_{i=1}^{n} x_{pi}}{n} \end{bmatrix} = \underline{\hat{\mu}}$$

- Matriks ragam peragam berukuran pxp, merupakan penduga bagi matriks Σ

$$_pS_p = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ & & & \\ s_{p1} & s_{p2} & & s_{pp} \end{bmatrix} = \hat{\Sigma}$$

$$s_{ii} = s_i^2 = \frac{\sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2}{n-1}$$

$$s_{ij} = \frac{\sum_{k=1}^{n} (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{n-1}$$

- Matriks korelasi berukuran pxp, merupakan penduga bagi matriks ρ

$$ {}_pR_p = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \dots & \dots & \dots & \dots \\ r_{p1} & r_{p2} & \dots & 1 \end{bmatrix} = \hat{\rho} $$

$$ r_{ii} = 1 $$

$$ r_{ij} = \frac{\sum_{k=1}^{n} (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{n} (x_{ik} - \bar{x}_i)^2 \sum_{k=1}^{n} (x_{jk} - \bar{x}_j)^2}} $$

Konsep Jarak

- Jarak pengamatan ke titik pusat

$$d_i = \sqrt{\left(\underline{x}_i - \underline{\bar{x}}\right)'\left(\underline{x}_i - \underline{\bar{x}}\right)}$$

- Jarak antar pengamatan (i,j)
  - Jarak Euclidean

$$d_{i.j} = \sqrt{\left(\underline{x}_i - \underline{x}_j\right)'\left(\underline{x}_i - \underline{x}_j\right)}$$

  - Jarak Mahalanobis

$$d_{i.j} = \sqrt{\left(\underline{x}_i - \underline{x}_j\right)' S^{-1}\left(\underline{x}_i - \underline{x}_j\right)}$$

  - Jarak Minkowski (City Block)

$$d_{i.j} = \sqrt[k]{\left(\underline{x}_i - \underline{x}_j\right)'\left(\underline{x}_i - \underline{x}_j\right)}, k = 2,3,4,....$$

Ringkasan Teknik Analisis Peubah Ganda

MANOVA

- MANOVA = Multivariate Analysis of Variance
- Analog dengan ANOVA (analysis of variance) pada analisis peubah tunggal
- Digunakan untuk membandingkan rata-rata populasi (misal membandingkan rata-rata perlakuan dalam percobaan)

MANOVA

- Mengasumsikan data berasal dari populasi multi-normal
- Model yang ada bervariasi dari Satu Arah (one-way) hingga model Banyak Arah (multi-way) termasuk adanya interaksi

MANOVA

Ilustrasi

Sebuah studi ingin membandingkan apakah ada perbedaan hasil panen tiga buah varietas padi. Respon yang diamati adalah tinggi tanaman ketika panen, bobot 100 butir padi, umur panen, panjang galur.

Analisis Komponen Utama

- Digunakan untuk melakukan reduksi peubah
- Menghasilkan peubah baru yang disebut KOMPONEN UTAMA dengan karakteristik : (1) informasi yang terkandung memuat hampir seluruh informasi (2) saling bebas
- Umumnya merupakan analisis antara, bukan analisis akhir

Analisis Faktor

- Digunakan juga untuk mereduksi peubah
- Menghasilkan peubah baru yang disebut FAKTOR atau PEUBAH LATEN
- Banyak digunakan di bidang terapan sosial karena sulit melakukan pengukuran secara langsung terhadap peubah yang diinginkan

Analisis Korelasi Kanonik

- Melakukan analisis keterkaitan antara dua gugus peubah
- Analisis korelasi tidak dilakukan antar pasangan peubah asal dari kedua gugus, namun antar peubah kanonik di kedua gugus

Analisis Korelasi Kanonik

Ilustrasi

Seorang konsultan olah gerak harus mampu melihat korelasi antara peubah anatomi tubuh manusia (lingkar lengan, berat badan, lingkar dada, panjang lengan, panjang kaki, tinggi badan, dsb) dengan berbagai aktifitas aerobik (sit-up, push-up, pull-up, scotch-jump, dsb)

Analisis Gerombol

- Digunakan untuk mengelompokkan objek-objek
- Kemiripan antar objek ditentukan oleh nilai-nilai pengamatan peubah ganda
- Dikenal teknik berhirarki dan tak-berhirarki

Analisis Gerombol

Ilustrasi

Pemberian program bantuan pengembangan sekolah tidak bisa disamakan di setiap sekolah. Berdasarkan karakteristik sekolah (fasilitas, input siswa, mutu pengajar, dukungan masyarakat), DEPDIKNAS melakukan pengelompokan sekolah sehingga diperoleh 4 kelompok SMU dan 4 jenis paket program pengembangan SMU.

Analisis Biplot

- Merupakan analisis eksplorasi untuk melihat (1) kedekatan antar objek (2) karakteristik atau peubah penciri setiap objek, dan (3) keterkaitan antar peubah
- Menggunakan konsep penguraian nilai singular
- Bermula di dunia pertanian, sekarang lebih banyak dipakai di riset pemasaran

Analisis Biplot

Ilustrasi

Perencanaan layanan perbankan memerlukan informasi mengenai posisi bank kita dibandingkan bank-bank pesaing. Untuk itulah dilakukan survei pasar mengenai persepsi nasabah mengenai berbagai atribut layanan perbankan, sehingga bisa kita ketahui atribut mana saja yang perlu ditingkatkan dan siapa pesaing terdekat yang harus diantisipasi.

Analisis Korespondensi

- Teknik eksplorasi yang diterapkan untuk melihat asosiasi kategori peubah kategorik
- Menggunakan konsep Generalized Singular Value Decomposition
- Banyak digunakan di dunia riset pemasaran

Analisis Korespondensi

Ilustrasi

Segmentasi produk bisa dilakukan berdasarkan peubah demografi konsumen. Melalui analisis ini bisa diketahui konsumen utama produk kita dari kelompok usia berapa, jenis kelamin apa, tinggal dimana, dan sebagainya.

Analisis Diskriminan

- Menghasilkan fungsi yang digunakan untuk mengklasifikasikan objek tertentu berasal dari populasi yang mana
- Linear Discriminant, Quadratic Discriminat, Canonical Discriminant, Logistic Regression, Non-Parametric Discriminant

Analisis Diskriminan

Ilustrasi

Perusahaan penyedia jasa layanan kartu kredit harus mampu membuat fungsi diskriminan yang mampu memisahkan calon pemegang kartu yang potensial melakukan transaksi dan yang tidak (idle) berdasarkan data dalam formulir aplikasi.

Penskalaan Dimensi Ganda – Multi Dimensional Scaling

- Menghasilkan peta atau gambar posisi objek berdasarkan matriks jarak yang diketahui
- Peta dihasilkan pada dimensi rendah (umumnya dimensi dua) sehingga mudah menginterpretasikan kedekatan antar objek
- Metric MDS vs Non-Metric MDS, tergantung tipe peubahnya