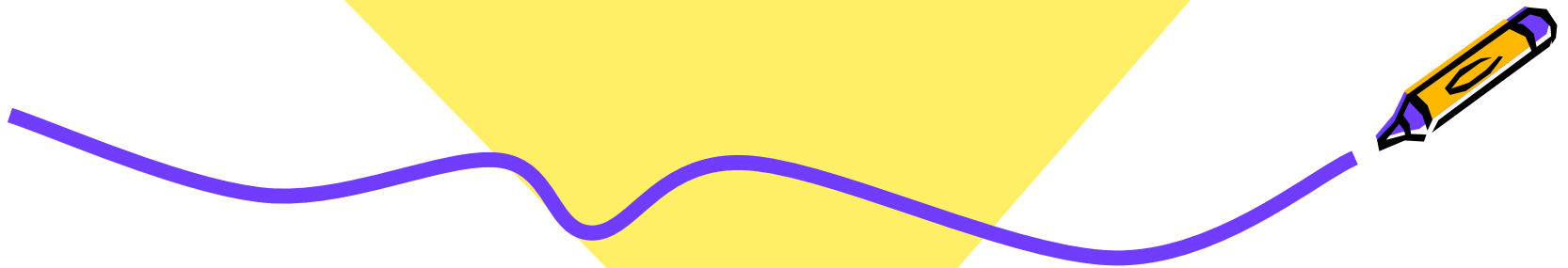




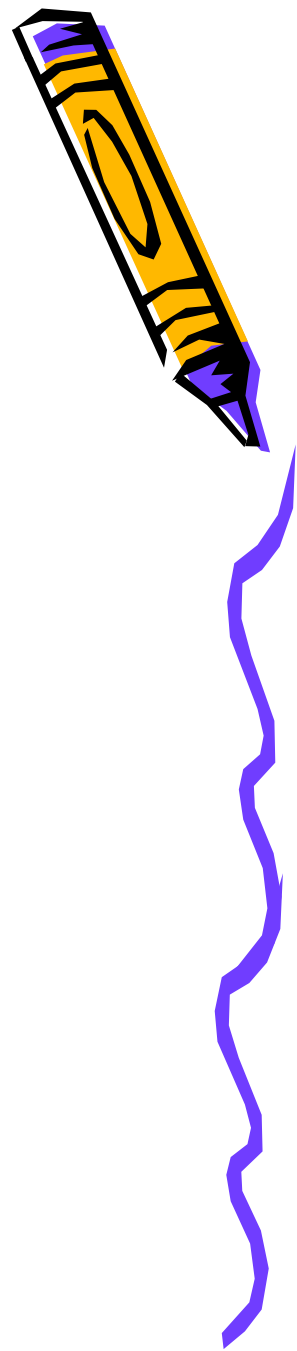
Canonical Correlation Analysis



Politeknik Statistika STIS
2021

Analisis Korelasi Kanonik (AKK)

**Bertujuan untuk mempelajari
kekuatan hubungan antara
dua himpunan/gugus peubah
ganda.**



Pengertian Awal



- Metode analisis pada AKK bersifat deskriptif ataupun dapat inferensia tergantung pada datanya.
- AKK merupakan pengembangan dari konsep korelasi linier dari Pearson.
- Analisis diawali dengan mengidentifikasi adanya hubungan liner antar sepasang peubah pada kedua himpunan peubah ganda.
- Pilihan basis AKK adalah matriks ragam-peragam atau matriks korelasi.

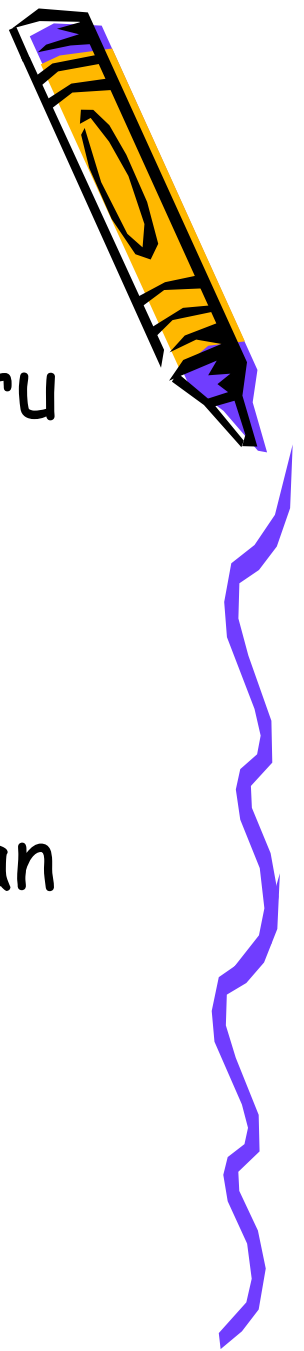


Metode Analisis

- Setiap himpunan peubah ganda akan diubah menjadi himpunan peubah baru (disebut peubah kanonik).

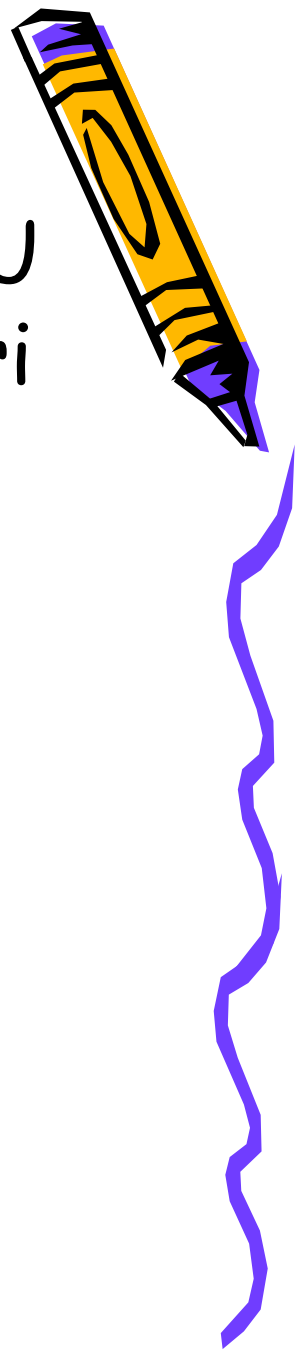
Misalkan:

Himpunan peubah ganda pertama (1)
sebanyak p peubah: X_1, X_2, \dots, X_p dan
Himpunan peubah ganda kedua (2)
sebanyak q peubah: X_1, X_2, \dots, X_q



AKK basis matriks ragam peragam

- Peubah kanonik (*canonical variable*): U dan V merupakan kombinasi linier dari himpunan peubah ganda.



$$U = a' X^{(1)} \text{ dan } V = b' X^{(2)}$$

$$\text{Var}(U) = a' \Sigma_{11} a \quad \text{Var}(V) = b' \Sigma_{22} b$$

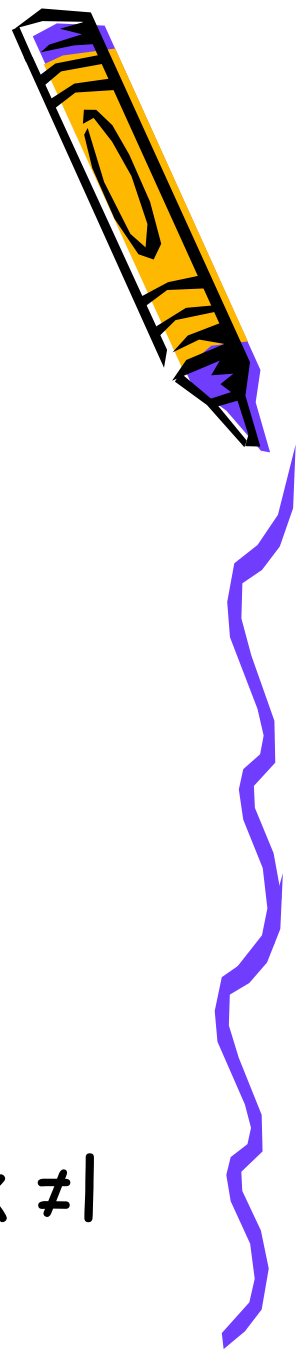
$$\text{Cov}(U, V) = a' \Sigma_{12} b$$



- Peubah kanonik dari masing-masing himpunan akan terurut berdasarkan pentingnya membentuk **korelasi pasangan kanonik** [**Corr(U,V)**]

Sifat variabel kanonik:

1. $\text{Var}(U_k) = \text{Var}(V_k) = 1$
2. $\text{Cov}(U_k, U_l) = \text{Corr}(U_k, U_l) = 0 \quad k \neq l$
3. $\text{Cov}(V_k, V_l) = \text{Corr}(V_k, V_l) = 0 \quad k \neq l$
4. $\text{Cov}(U_k, V_l) = \text{Corr}(U_k, V_l) = 0 \quad k \neq l$



Formulasi Korelasi Kanonik

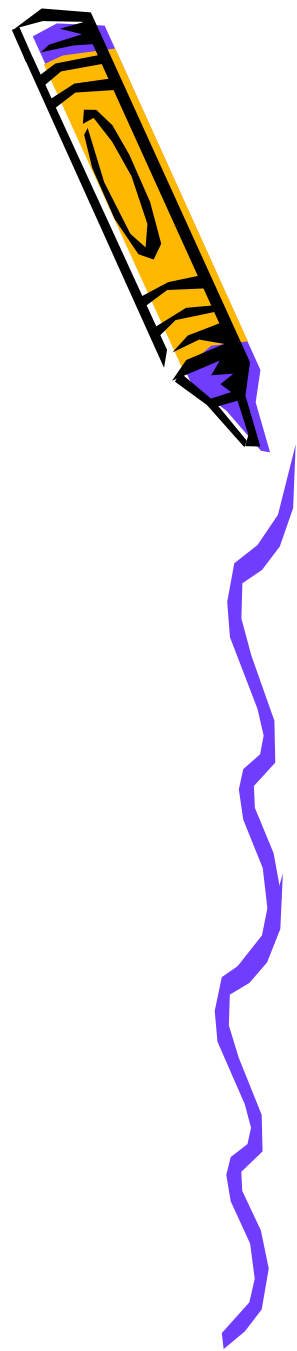
$$\text{Corr}(U,V) = \frac{\text{Cov}(U,V)}{[\text{Var}(U)]^{1/2} [\text{Var}(V)]^{1/2}}$$

$$= \frac{\mathbf{a}' \Sigma_{12} \mathbf{b}}{[\mathbf{a}' \Sigma_{11} \mathbf{a}]^{1/2} [\mathbf{b}' \Sigma_{22} \mathbf{b}]^{1/2}}$$

$\text{Corr}(U,V): \text{Corr}(U_k, V_k)$

untuk $k = 1, 2, \dots, r$

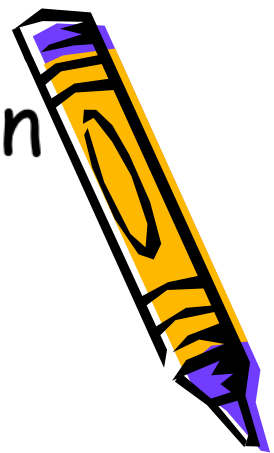
$r = \min(p,q)$



Matriks Ragam Peragam dari 2 himpunan
peubah ganda:

$$\Sigma_{(p+q) \times (p+q)} = \begin{bmatrix} \Sigma_{11(p \times p)} & \Sigma_{12(p \times q)} \\ \Sigma_{21(q \times p)} & \Sigma_{22(q \times q)} \end{bmatrix}$$

Catatan: notasi matriks disesuaikan basis AKK
dan amatan populasi atau sampel



Penghitungan vektor **a** dan **b**



- Vektor **a** diperoleh dari persamaan ciri:

$$(\Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}} - \rho^{*2} \mathbf{I}) \mathbf{e} = 0$$

$$\mathbf{a}_i = \mathbf{e}_i \Sigma_{11}^{-\frac{1}{2}}$$

- Vektor **b** diperoleh dari persamaan ciri:

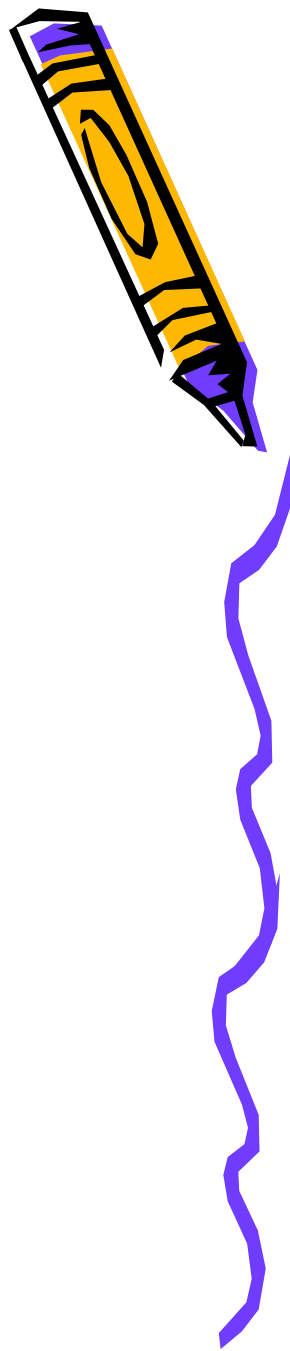
$$(\Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}} - \rho^{*2} \mathbf{I}) \mathbf{f} = 0$$

$$\mathbf{b}_i = \mathbf{f}_i \Sigma_{22}^{-\frac{1}{2}}$$



Karena $\text{Var}(U_k) = \text{Var}(V_k) = 1$

maka: $\text{Corr}(U_k, V_k) = \rho_k^*$

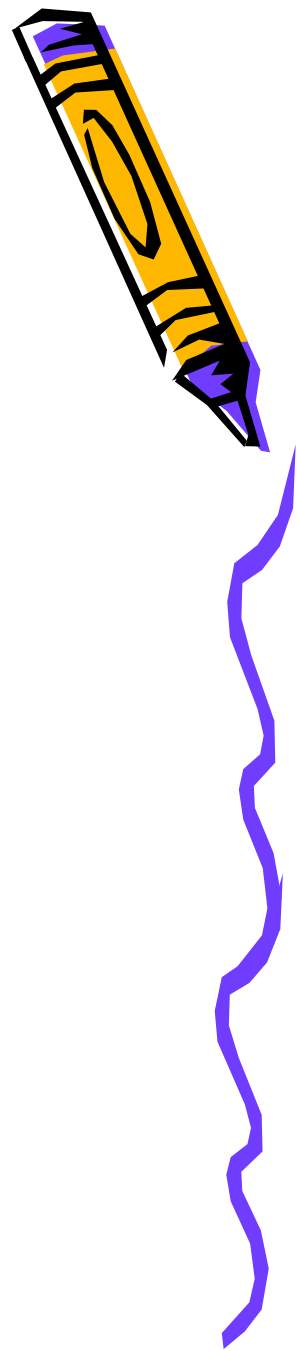


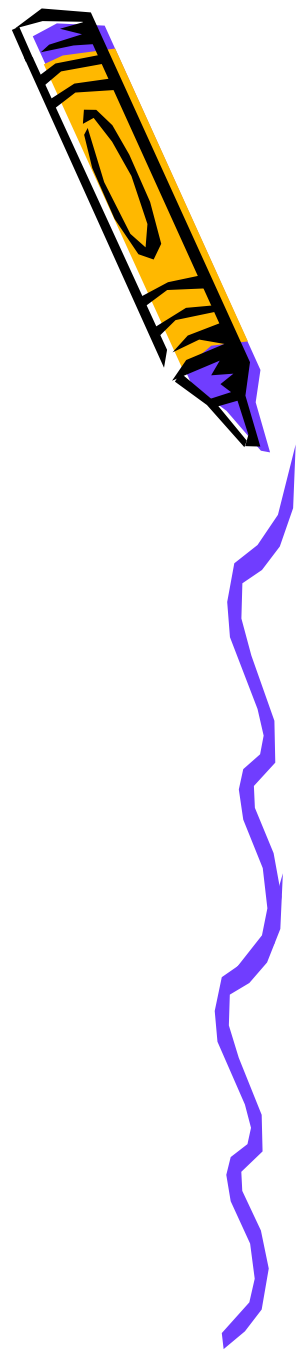
Kontribusi variabel kanonik

- Keragaman masing-masing himpunan peubah ganda dpt dijelaskan oleh suatu variabel kanonik:

Himpunan peubah ganda (1):

$$\bullet R^2_{Z(1)|U_i} = \frac{\sum_i \sum_k r^2_{Ui,Zk}^{(1)}}{p} \quad \begin{array}{l} i = 1, 2, \dots, r \\ k = 1, 2, \dots, p \end{array}$$





Himpunan peubah ganda (2):

$$R^2_{Z^{(2)}|V_i} = \frac{\sum_i \sum_k r^2_{Vi,Zk}^{(2)}}{q}$$

$$i = 1, 2, \dots, r$$

$$k = 1, 2, \dots, q$$

