# *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: $X_1$, $X_2$, ....... $X_p$ dan

Himpunan peubah ganda kedua (2) sebanyak ***q*** peubah: $X_1$, $X_2$, ....... $X_q$

# AKK basis matriks ragam peragam

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

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

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

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

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

**Sifat variabel kanonik**:

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

# Formulasi Korelasi Kanonik

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

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

Corr(U,V): Corr($U_k$,$V_k$)

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

$r = \min(p,q)$

# Matriks Ragam Peragam dari 2 himpunan peubah ganda:

$$\Sigma_{(p+q)x(p+q)} = \left[\begin{array}{c|c} \Sigma_{11_{(pxp)}} & \Sigma_{12_{(pxq)}} \\ \hline \Sigma_{21(qxp)} & \Sigma_{22_{(qxq)}} \end{array}\right]$$

Catatan: notasi matriks disesuaikan basis AKK

dan amatan populasi atau sampel

# Penghitungan vektor a dan b

- Vektor a diperoleh dari persamaan ciri:

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

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

- Vektor b diperoleh dari persamaan ciri:

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

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

Karena $Var(U_k) = Var(V_k) = 1$

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

# Kontribusi variabel kanonik

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

Himpunan peubah ganda (1):

- $R^2{}_{Z^{(1)}|U_i} = \dfrac{\sum_i \sum_k r^2{}_{U_i, Z_k^{(1)}}}{p}$

i= 1,2, …, r

k= 1,2, …, p

Himpunan peubah ganda (2):

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

i= 1,2, ..., r

k= 1,2, ..., q