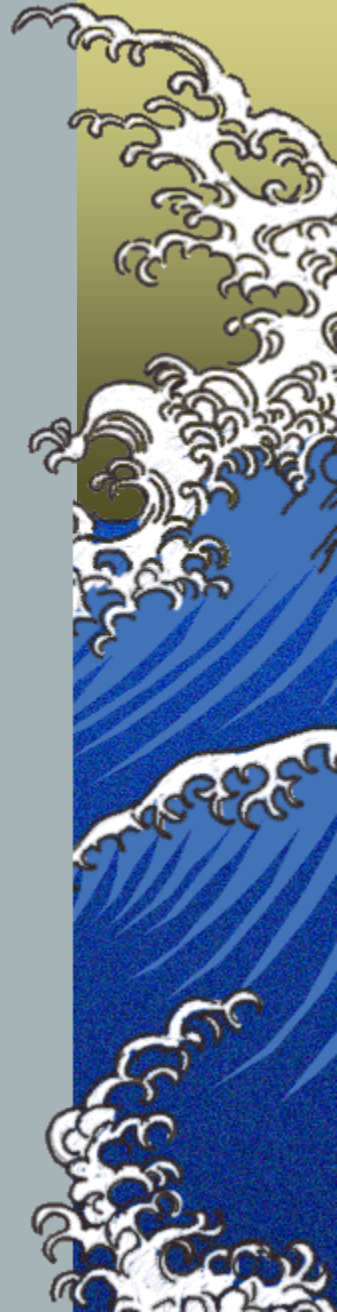


Canonical Correlation



I Made Sumertajaya



Pendahuluan

Hubungan antar variabel yang telah dikenal:

➤ Dua arah

- ❖ 1 var dependen vs 1 var independen → korelasi sederhana (simple correlation): pearson, spearman, tau kendall; tabel kontingensi (contingency table): uji khi-kuadrat, analisis korespondensi (correspondency analysis)
- ❖ 1 var dependen vs >1 var dependen → korelasi parsial (parsial correlation), korelasi ganda (multiple correlation), tabel multi arah: analisis korespondensi ganda (multiple correspondency analysis)

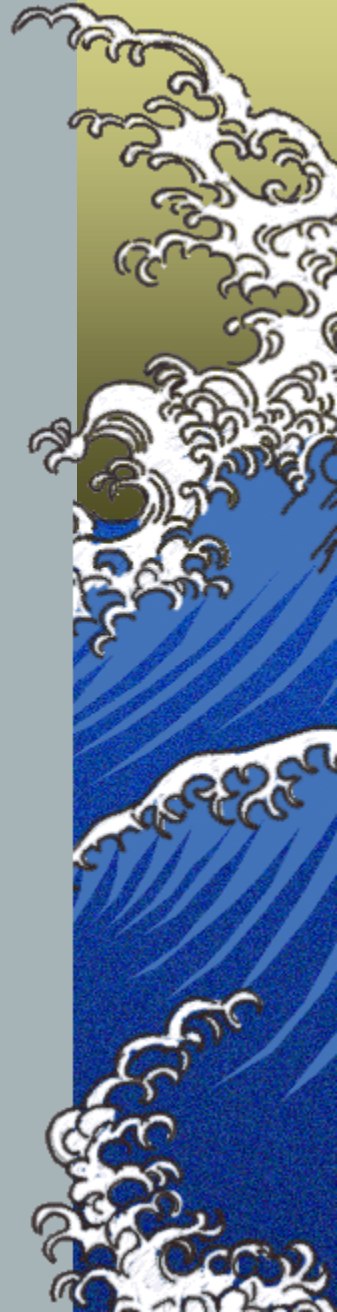


Pendahuluan: lanjutan

Hubungan antar variabel yang telah dikenal:

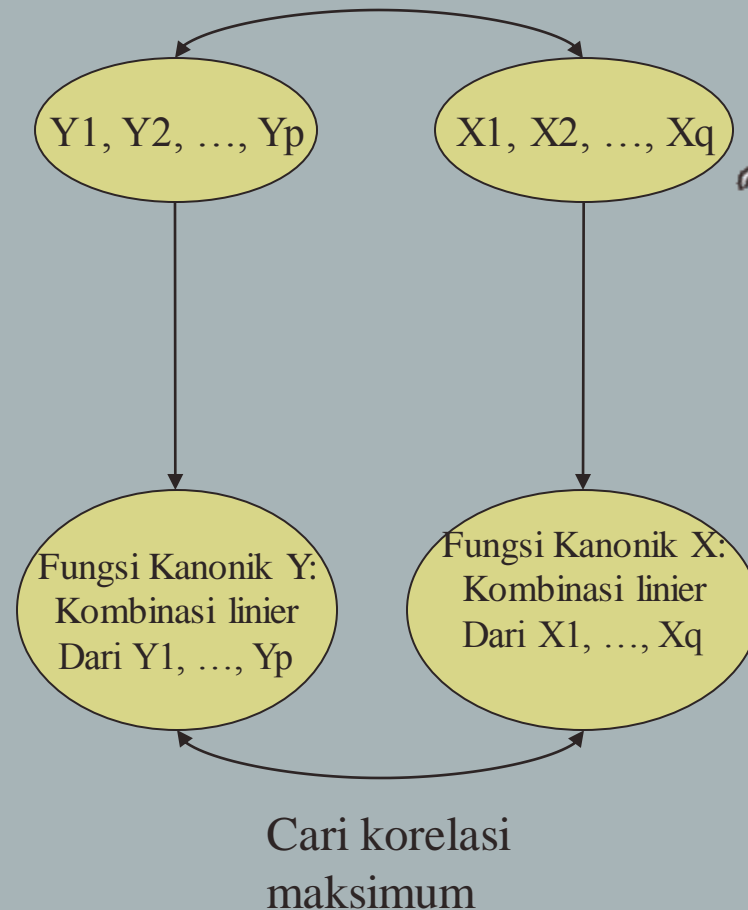
➤ Satu arah

- ❖ 1 var dependen vs 1 var independen → Analisis regresi sederhana (simple regression analysis); Analisis perancangan percobaan: faktor tunggal; dll
- ❖ 1 var dependen vs >1 var dependen → Analisis regresi berganda (multiple regression analysis); Analisis perancangan percobaan: faktorial, split-plot; dll

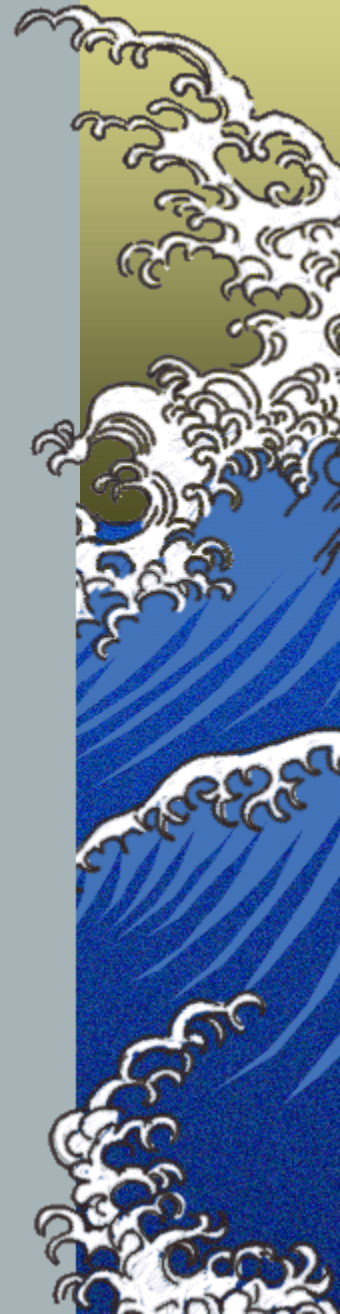
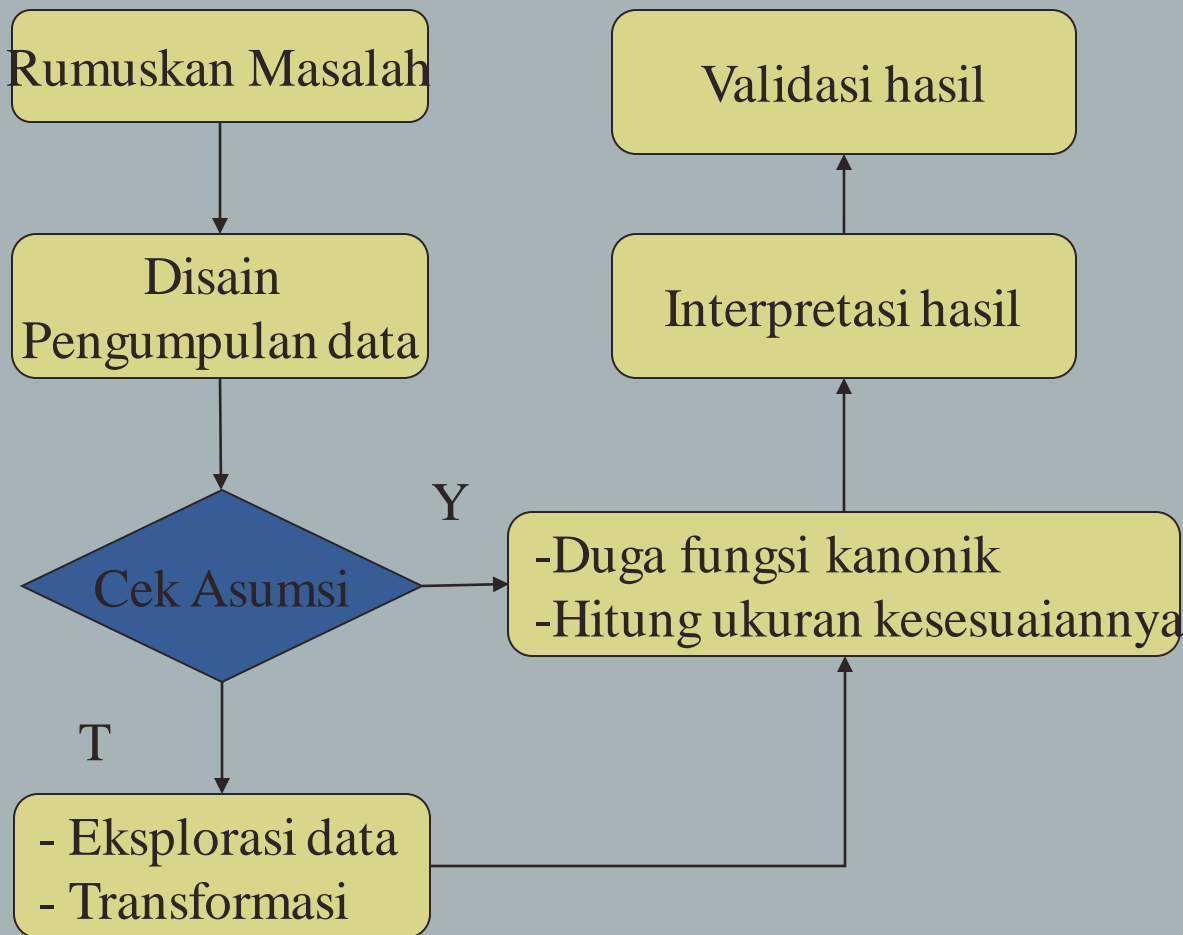


Apa itu korelasi kanonik?

- ▶ *Mengkaji hubungan antar gugus var dependen dengan gugus var independen*
- ▶ *Mengkaji struktur setiap gugus var baik indepen maupun dependen*



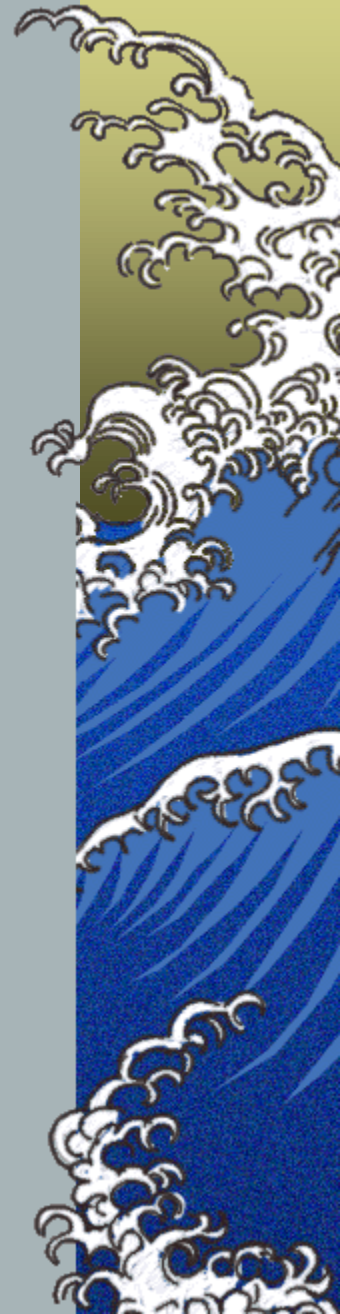
Tahapan Analisis Korelasi Kanonik



Asumsi

Beberapa asumsi yang harus diperhatikan dalam analisis korelasi kanonik yaitu:

- a. Korelasi antar peubah asal didasarkan pada hubungan linier
- b. Korelasi kanonik adalah hubungan linier antar variate
- c. Multivariate normal (normal ganda), asumsi ini diperlukan pada saat melakukan pengujian terhadap fungsi kanonik.



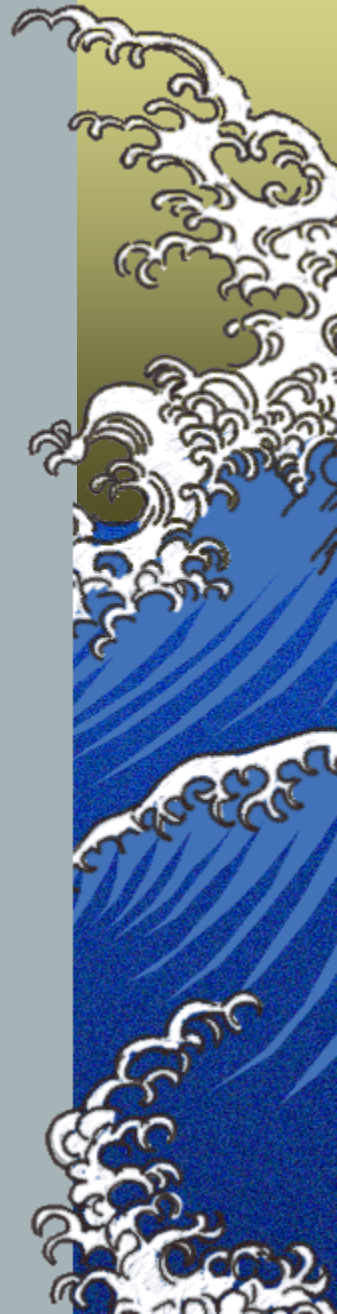
Pendugaan Fungsi Kanonik

- ▶ Misal, gugus peubah dependen Y_1, Y_2, \dots, Y_p dan gugus peubah independen X_1, X_2, \dots, X_q . Misalkan, karakteristik dari vektor peubah acak X dan Y adalah sebagai berikut:

$$E(\mathbf{Y}) = \boldsymbol{\mu}_Y \quad \text{Cov}(\mathbf{Y}) = \boldsymbol{\Sigma}_{Y.Y}$$

$$E(\mathbf{X}) = \boldsymbol{\mu}_X \quad \text{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}_{X.X}$$

$$\text{Cov}(\mathbf{X}, \mathbf{Y}) = \boldsymbol{\Sigma}_{X.Y} = \boldsymbol{\Sigma}_{Y.X}'$$



Pendugaan Fungsi Kanonik: lanjutin

- ▶ Kombinasi linier dari kedua gugus peubah tersebut dapat dituliskan sebagai berikut:

$$U = \mathbf{a}' \mathbf{X} = a_1 X_1 + a_2 X_2 + \dots + a_q X_q$$

$$V = \mathbf{b}' \mathbf{Y} = b_1 Y_1 + b_2 Y_2 + \dots + b_p Y_p$$

Sehingga,

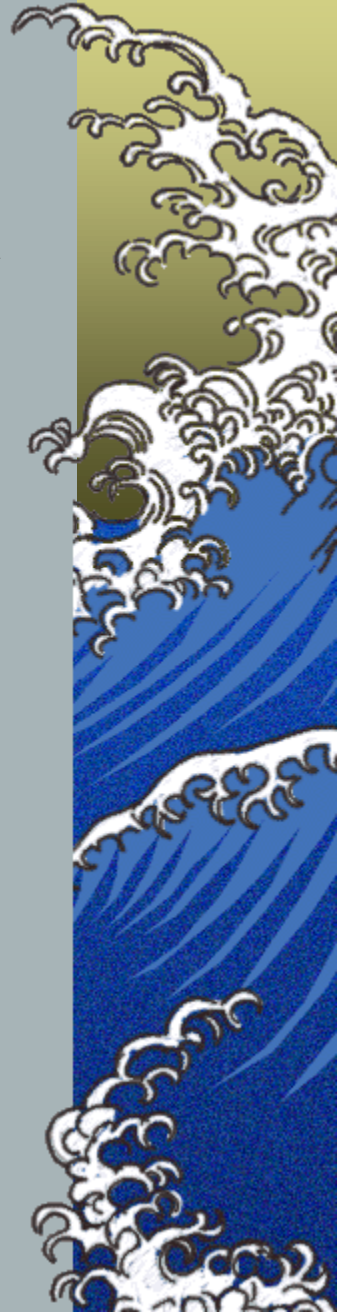
$$\text{Var}(U) = \mathbf{a}' \text{Cov}(\mathbf{X}) \mathbf{a} = \mathbf{a}' \Sigma_{X.X} \mathbf{a}$$

$$\text{Var}(V) = \mathbf{b}' \text{Cov}(\mathbf{Y}) \mathbf{b} = \mathbf{b}' \Sigma_{Y.Y} \mathbf{b}$$

$$\text{Cov}(U, V) = \mathbf{a}' \text{Cov}(\mathbf{X}, \mathbf{Y}) \mathbf{b} = \mathbf{a}' \Sigma_{X.Y} \mathbf{b}$$

- ▶ Dari sini kita mencari vektor koefesien \mathbf{a} dan \mathbf{b} sehingga korelasinya maksimum,

$$\text{Corr}(U, V) = \frac{\mathbf{a}' \Sigma_{X.Y} \mathbf{b}}{\sqrt{\mathbf{a}' \Sigma_{X.X} \mathbf{a}} \sqrt{\mathbf{b}' \Sigma_{Y.Y} \mathbf{b}}}$$



Pendugaan Fungsi Kanonik: lanjutan

Definisi:

▲ Peubah kanonik pertama: korelasi terbesar pertama

$$U_1 = \mathbf{a}_1' \mathbf{X} \quad \text{Var}(U_1) = 1$$

$$V_1 = \mathbf{b}_1' \mathbf{Y} \quad \text{Var}(V_1) = 1$$

$$\text{Maksimum Corr}(U_1, V_1) = \rho_1$$

▲ Peubah kanonik kedua: korelasi terbesar kedua

$$U_2 = \mathbf{a}_2' \mathbf{X} \quad \text{Var}(U_2) = 1 \quad \text{Cov}(U_1, U_2) = 0 \quad \text{Cov}(U_1, V_2) = \text{Cov}(U_2, V_1) = 0$$

$$V_2 = \mathbf{b}_2' \mathbf{Y} \quad \text{Var}(V_2) = 1 \quad \text{Cov}(V_1, V_2) = 0$$

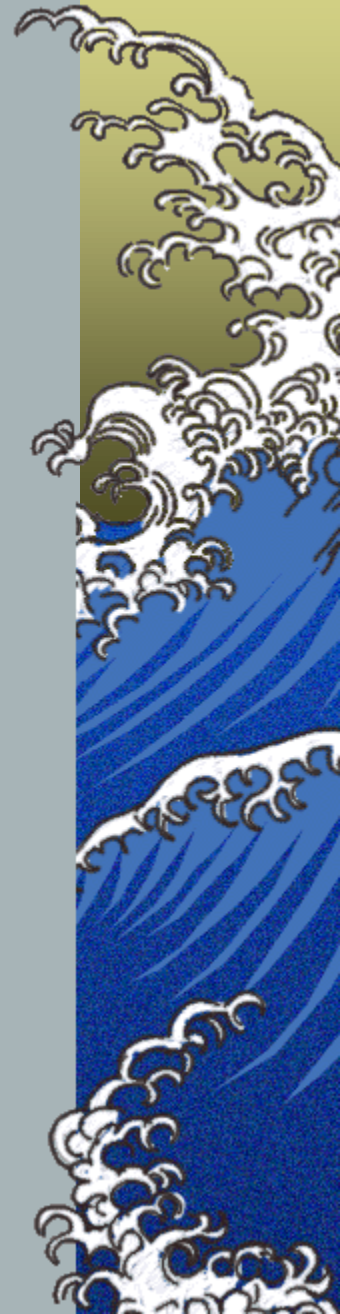
$$\text{Maksimum Corr}(U_2, V_2) = \rho_2$$

▲ Peubah kanonik ke-k:

$$U_k = \mathbf{a}_k' \mathbf{X} \quad \text{Var}(U_k) = 1 \quad \text{Cov}(U_k, U_l) = 0, \quad k \neq l \quad \text{Cov}(U_k, V_l) = 0, \quad k \neq l$$

$$V_k = \mathbf{b}_k' \mathbf{Y} \quad \text{Var}(V_k) = 1 \quad \text{Cov}(V_l, V_k) = 0, \quad k \neq l$$

$$\text{Maksimum Corr}(U_k, V_k) = \rho_k$$



Pendugaan Fungsi Kanonik: lanjutan

▲ Dengan menggunakan ketaksamaan Cauchy-Schwarz atau metode langrange maka diperoleh: (Pembuktiannya dapat dilihat Johnson, 1988 hal 441)

▲ $\rho_1^2, \rho_2^2, \dots, \rho_p^2$ adalah akar ciri-akar ciri (eigenvalues) dari matriks

$$\Sigma_{YY}^{-1/2} \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YY}^{-1/2}$$

yang berpadanan dengan vektor ciri f_1, f_2, \dots, f_p .

▲ $\rho_1^2, \rho_2^2, \dots, \rho_p^2$ juga merupakan akar ciri-akar ciri (eigenvalues) dari matriks

$$\Sigma_{XX}^{-1/2} \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX} \Sigma_{XX}^{-1/2}$$

yang berpadanan dengan vektor ciri e_1, e_2, \dots, e_p .

▲ Sehingga vektor koefisien a dan b diperoleh sebagai berikut:

$$a_1 = e_1 \Sigma_{XX}^{-1/2}$$

$$a_2 = e_2 \Sigma_{XX}^{-1/2}$$

....

$$a_p = e_p \Sigma_{XX}^{-1/2}$$

$$b_1 = f_1 \Sigma_{YY}^{-1/2}$$

$$b_2 = f_2 \Sigma_{YY}^{-1/2}$$

....

$$b_p = f_p \Sigma_{YY}^{-1/2}$$



Ukuran Kesesuaian Fungsi Kanonik

(1). Proporsi keragaman

$$R_{Z_X|U_1, \dots, U_r}^2 = \frac{\sum_{i=1}^r \sum_{k=1}^q r_{U_i Z_{X_k}}^r}{q} \quad R_{Z_Y|V_1, \dots, V_r}^2 = \frac{\sum_{i=1}^r \sum_{k=1}^p r_{V_i Z_{Y_k}}^r}{p}$$

(2). Inferensia

(i). Apakah secara keseluruhan peubah kanonik berhubungan ?

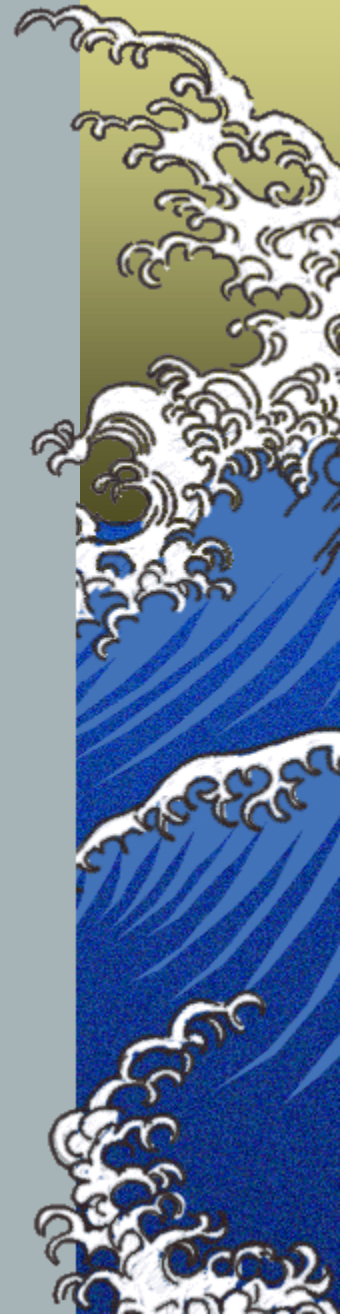
Bentuk hipotesisnya sebagai berikut:

$$H_0: \sum_{XY} = 0 \quad \text{vs} \quad H_1: \sum_{XY} \neq 0$$

Hipotesis nol ditolak jika nilai berikut besar,

$$-2 \ln \Lambda = n \ln \left(\frac{|S_{XX}| |S_{YY}|}{|S|} \right) = -n \ln \prod_{i=1}^p (1 - \hat{\rho}_i^2)$$

$$-(n-1-1/2(p+q+1)) \ln \prod_{i=1}^p (1 - \hat{\rho}_i^2) > \chi_{pq}^2(\alpha)$$



Ukuran Kesesuaian Fungsi Kanonik: lanjutan

(ii). Apakah ada sebagian peubah kanonik berhubungan ?

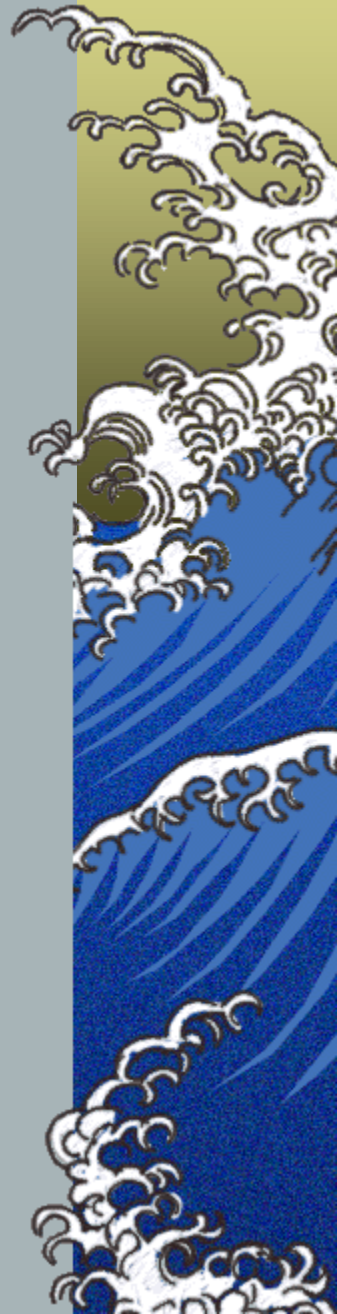
Hipotesisnya dapat dituliskan sebagai berikut:

$$H_0: \rho_1 \neq 0, \dots, \rho_k \neq 0, \rho_{k+1} = 0, \dots, \rho_p = 0$$

$$H_1: \rho_i \neq 0 \text{ untuk beberapa } i > k$$

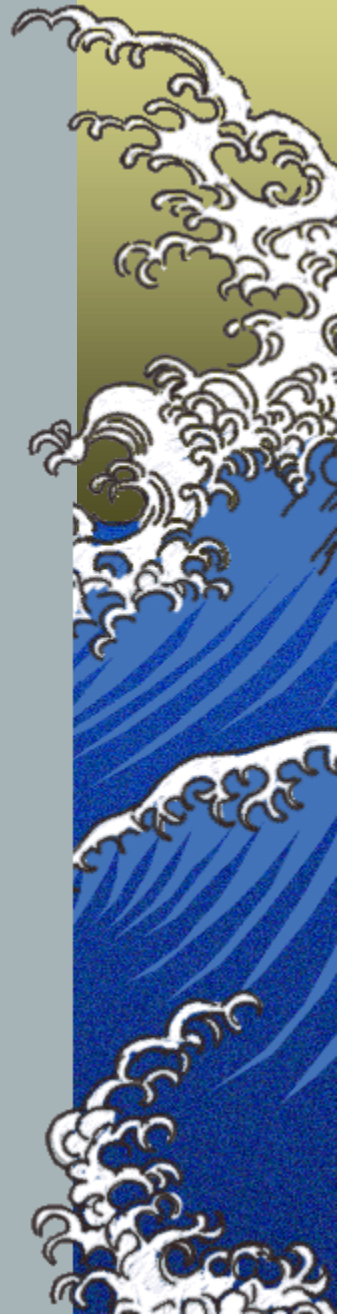
Tolak hipotesis nol pada taraf α , jika

$$-(n-1-1/2(p+q+1)) \ln \prod_{i=k+1}^p (1 - \hat{\rho}_i^2) > \chi_{(p-k)(q-k)}^2(\alpha)$$



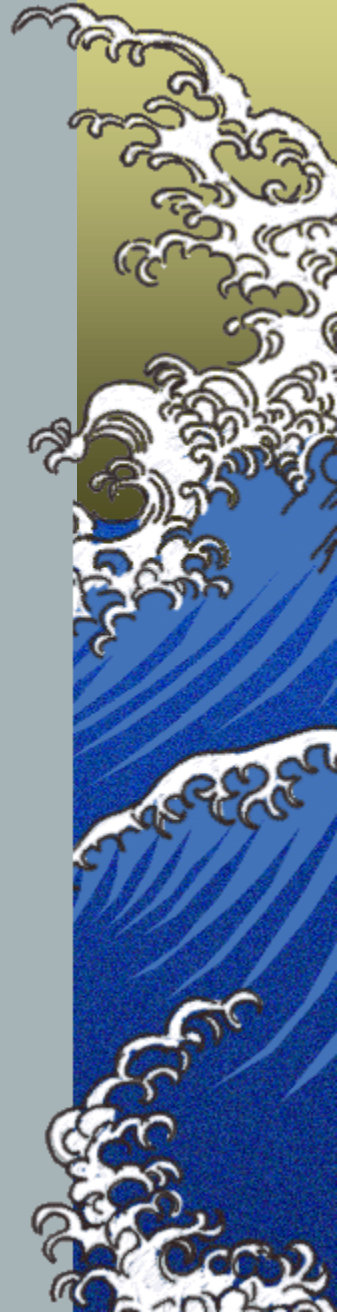
Interpretasi

- ▶ Koefesien kanonik yaitu a dan b yang telah dibakukan dapat diinterpretasikan sebagai besarnya kontribusi peubah asal terhadap variate kanonik.
- ▶ Loading kanonik dapat dihitung dari korelasi antara peubah asal dengan masing-masing fungsi kanonik.



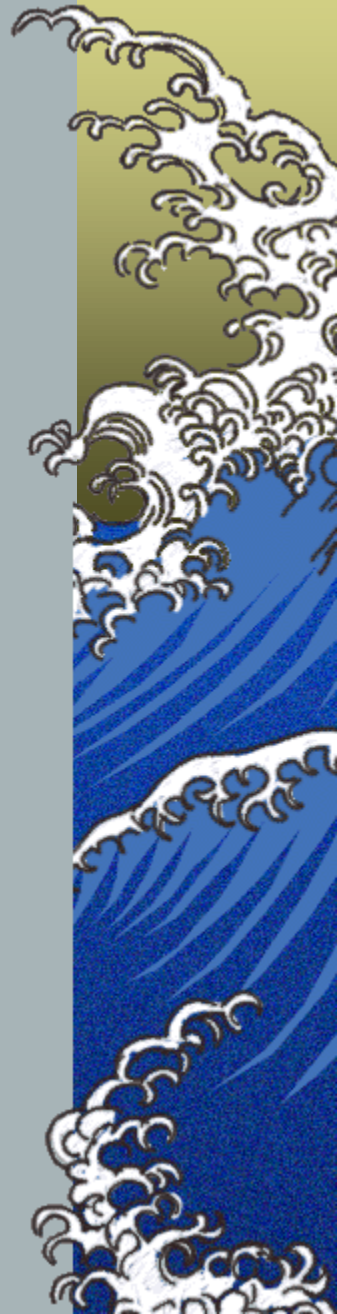
Validasi

- a. Membagi sampel menjadi dua bagian, bagian pertama digunakan untuk menduga fungsi kanonik dan bagian kedua digunakan sebagai validasi.*
- b. Analisis sensitivitas untuk peubah-peubah independen, yaitu dengan membandingkan loading kanonik apabila salah-satu dari peubah independen disisihkan dari analisis.*



Ilustrasi

Hubungan antara pengeluaran dengan karakteristik rumah tangga (penelitian mahasiswa pasca sarjana, 2000)



Program SAS

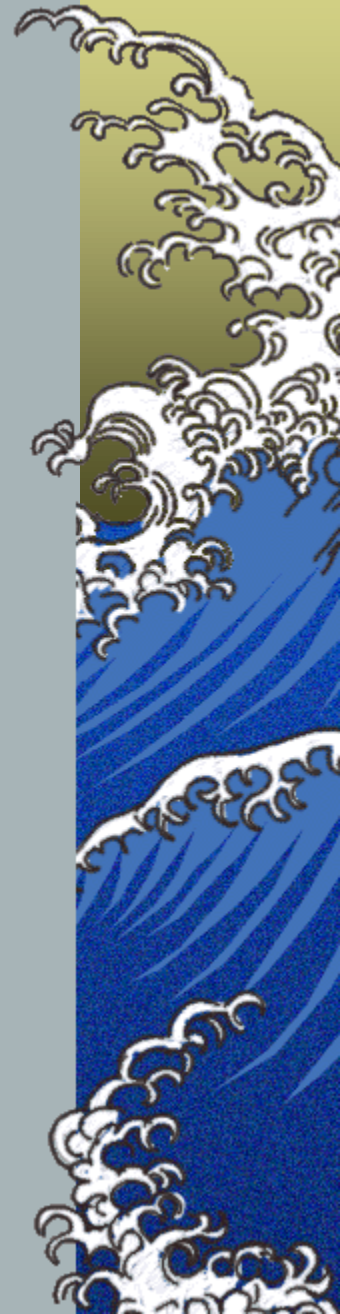
```
options ps=100 ls=76 nonumber nodate;  
title ' ';
```

```
data dt;  
input id$ y1 y2 y3 y4 x1 x2 x3 x4;  
label y1='Jumlah kartu kredit'  
      y2='Pengeluaran untuk konsumsi (Rp)'  
      y3='Pengeluaran untuk sandang (Rp)'  
      y4='Pengeluaran lain-lain'  
      x1='Jumlah anggota keluarga'  
      x2='Pendidikan (tahun)'  
      x3='Pendapatan per bulan (Rp)'  
      x4='Umur (tahun)';
```

```
cards;  
ld_001 4 1757500 169688 113125 2 13 2262500 32  
ld_002 1 1587500 140625 375000 5 18 1875000 40  
-- dst --  
;
```

```
Title1 'Hasil Analisis Korelasi Kanonik';
```

```
proc cancorr redundancy ncan=3 corr data=dt;  
var y1-y4;  
with x1-x4;  
run;
```



Hasil Analisis

- **Korelasi antar variabel**

Correlations Among the Original Variables

Correlations Among the VAR Variables

	y1	y2	y3	y4
y1	1.0000	0.7119	0.6570	-0.2788
y2	0.7119	1.0000	0.8446	-0.0050
y3	0.6570	0.8446	1.0000	0.2853
y4	-0.2788	-0.0050	0.2853	1.0000

Correlations Between the VAR Variables and the WITH Variables

	x1	x2	x3	x4
y1	-0.1241	-0.1554	0.5486	-0.1356
y2	0.4491	0.1106	0.8620	0.3914
y3	0.0704	0.2630	0.9236	0.3175
y4	0.1506	0.1631	0.1057	0.1871

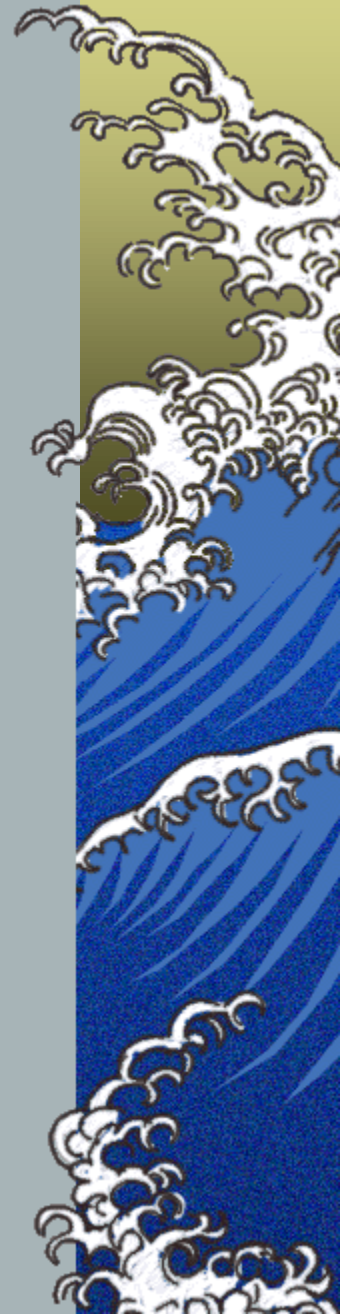


Hasil Analisis: lanjutan

✦ Korelasi Kanonik

Canonical Correlation Analysis

	<i>Canonical Correlation</i>	<i>Adjusted Canonical Correlation</i>	<i>Approximate Standard Error</i>	<i>Squared Canonical Correlation</i>
1	0.996511	0.996331	0.000700	0.993035
2	0.930033	0.927719	0.013572	0.864961
3	0.497259	0.483800	0.075653	0.247267
4	0.135327	.	0.098663	0.018313



Hasil Analisis: lanjutan

▲ Inferensia

*Eigenvalues of Inv(E)*H
= CanRsqu/(1-CanRsqu)*

	<i>Eigenvalue</i>	<i>Difference</i>	<i>Proportion</i>	<i>Cumulative</i>
1	142.5652	136.1600	0.9548	0.9548
2	6.4053	6.0768	0.0429	0.9977
3	0.3285	0.3098	0.0022	0.9999
4	0.0187		0.0001	1.0000

Test of H0: The canonical correlations in the current row and all that follow are zero

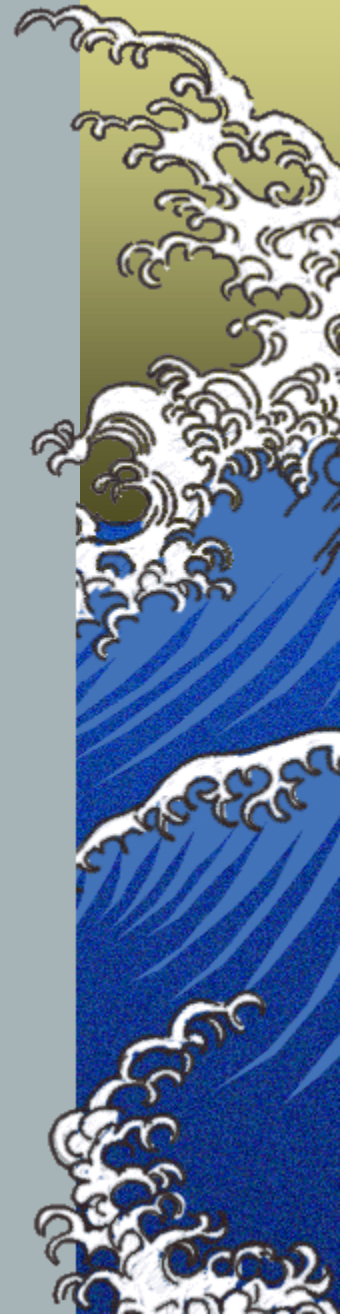
	<i>Likelihood Ratio</i>	<i>Approximate F Value</i>	<i>Num DF</i>	<i>Den DF</i>	<i>Pr > F</i>
1	0.00069506	172.66	16	281.7	<.0001
2	0.09978675	39.71	9	226.49	<.0001
3	0.73894810	7.68	4	188	<.0001
4	0.98168668	1.77	1	95	0.1863

Multivariate Statistics and F Approximations

S=4 M=-0.5 N=45

<i>Statistic</i>	<i>Value</i>	<i>F Value</i>	<i>Num DF</i>	<i>Den DF</i>	<i>Pr > F</i>
<i>Wilks' Lambda</i>	0.00069506	172.66	16	281.7	<.0001
<i>Pillai's Trace</i>	2.12357579	26.88	16	380	<.0001
<i>Hotelling-Lawley Trace</i>	149.31764851	849.45	16	178.09	<.0001
<i>Roy's Greatest Root</i>	142.56522889	3385.92	4	95	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.



Hasil Analisis: lanjutan

▲ Fungsi Kanonik

Standardized Canonical Coefficients for the VAR Variables

		V1	V2	V3
y1	Jumlah kartu kredit	-0.6172	0.1083	-1.5173
y2	Pengeluaran untuk konsumsi (Rp)	1.3406	-1.6030	-0.4402
y3	Pengeluaran untuk sandang (Rp)	-0.0004	2.0796	1.3661
y4	Pengeluaran lain-lain	0.0011	-0.6226	-0.5139

Standardized Canonical Coefficients for the WITH Variables

		W1	W2	W3
x1	Jumlah anggota keluarga	0.5995	-0.8084	0.1557
x2	Pendidikan (tahun)	0.0288	0.0508	1.1757
x3	Pendapatan per bulan (Rp)	0.7448	0.6827	-0.3808
x4	Umur (tahun)	-0.0424	-0.0349	-0.1344



Hasil Analisis: lanjutan

↳ Loading kanonik

Correlations Between the VAR Variables and Their Canonical Variables

		V1	V2	V3
y1	Jumlah kartu kredit	0.3366	0.5071	-0.7898
y2	Pengeluaran untuk konsumsi (Rp)	0.9010	0.2337	-0.3639
y3	Pengeluaran untuk sandang (Rp)	0.7268	0.6192	-0.1492
y4	Pengeluaran lain-lain	0.1664	-0.0514	0.3012

Correlations Between the WITH Variables and Their Canonical Variables

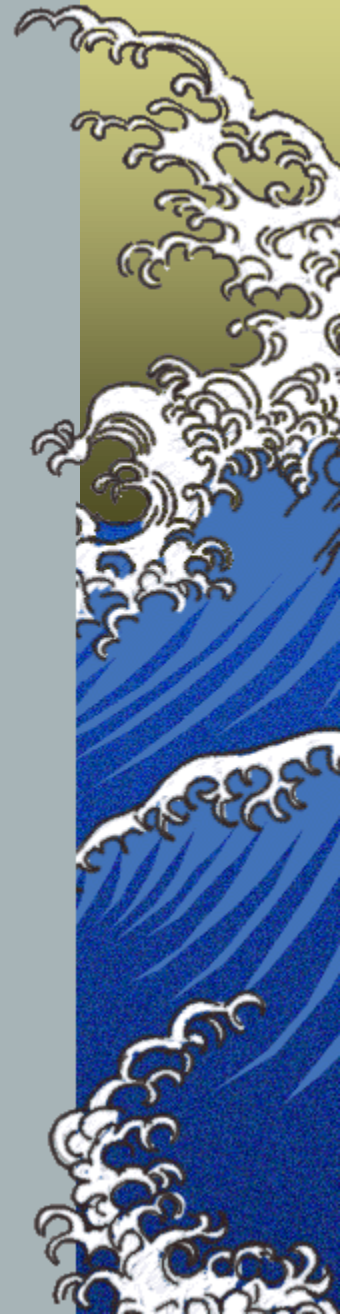
		W1	W2	W3
x1	Jumlah anggota keluarga	0.6812	-0.7319	0.0189
x2	Pendidikan (tahun)	0.2451	0.2703	0.9305
x3	Pendapatan per bulan (Rp)	0.8197	0.5727	-0.0087
x4	Umur (tahun)	0.6106	-0.1057	0.7461

Correlations Between the VAR Variables and the Canonical Variables of the WITH Variables

		W1	W2	W3
y1	Jumlah kartu kredit	0.3354	0.4716	-0.3927
y2	Pengeluaran untuk konsumsi (Rp)	0.8978	0.2173	-0.1809
y3	Pengeluaran untuk sandang (Rp)	0.7242	0.5759	-0.0742
y4	Pengeluaran lain-lain	0.1658	-0.0478	0.1498

Correlations Between the WITH Variables and the Canonical Variables of the VAR Variables

		V1	V2	V3
x1	Jumlah anggota keluarga	0.6788	-0.6807	0.0094
x2	Pendidikan (tahun)	0.2443	0.2514	0.4627
x3	Pendapatan per bulan (Rp)	0.8168	0.5326	-0.0043
x4	Umur (tahun)	0.6085	-0.0983	0.3710



Terima Kasih

