

MULTIKOLINEARITAS (Lanjutan)

Tjipto Juwono, Ph.D.

July 2018



SURYA
UNIVERSITY

Pengertian Multikolinearitas

- Secara historis multikolinearitas menunjukkan hubungan yang sempurna antara variabel-variabel independent

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k = 0 \quad (1)$$

$\lambda_1, \lambda_2, \dots, \lambda_k$ adalah konstanta-konstanta yang tidak semuanya sama dengan nol.

- Dewasa ini, istilah multikolinearitas digunakan untuk hubungan yang tidak sempurna juga

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k + \nu_i = 0 \quad (2)$$

Dengan ν_i adalah suku error random/stokastik.

Pengertian Multikolinieritas

Anggap bahwa $\lambda_2 \neq 0$

Sempurna

$$X_{2i} = -\frac{\lambda_1}{\lambda_2} X_{1i} - \frac{\lambda_3}{\lambda_2} X_{3i} - \dots - \frac{\lambda_k}{\lambda_2} X_{ki} \quad (3)$$

Persamaan [3] menunjukkan bahwa X_2 mempunyai relasi linear yang eksak linear dengan semua variabel independent lainnya.

Anggap bahwa $\lambda_2 \neq 0$

Tidak Sempurna

$$X_{2i} = -\frac{\lambda_1}{\lambda_2} X_{1i} - \frac{\lambda_3}{\lambda_2} X_{3i} - \dots - \frac{\lambda_k}{\lambda_2} X_{ki} - \frac{1}{\lambda_2} \nu_i \quad (4)$$

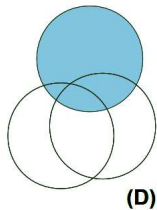
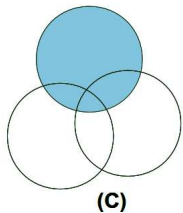
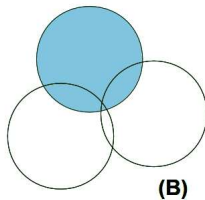
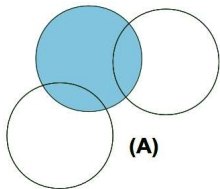
Persamaan [4] menunjukkan bahwa X_2 mempunyai relasi linear yang tidak eksak linear dengan semua variabel independent lainnya karena adanya suku stokastik $(-1/\lambda_2)\nu_i$.

Table 1: Tabel Kolinear

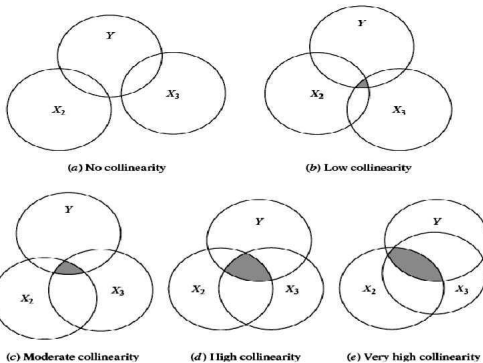
X_2	X_3	X_3^*
10	50	52
15	75	75
18	90	97
24	120	129
30	150	152

- Tabel 1 di atas memperlihatkan bahwa $X_{3i} = 5X_{2i}$. Berarti ada **kolinearitas sempurna** antara X_2 dan X_3 karena $r_{23} = 1$.
- X_3^* diperoleh dengan menambahkan angka random 2,0,7,9,2 pada X_3 . Sekarang kita mempunyai $X_{3i}^* = 5X_{2i} + \nu_i$ yang bukan merupakan kolinearitas sempurna.
- Namun koefisien korelasi antara X_3 dan X_3^* adalah 0.9959, yang merupakan korelasi yang kuat.

Multikolinieritas



Multikolinearitas



- Pembahasan tentang multikolinearitas hanya memperhitungkan hubungan linear antara variabel-variabel independen. Hubungan non-linear tidak diperhitungkan.
- Contoh hubungan non-linear

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i \quad (5)$$

- Variabel-variabel X_i^2 dan X_i^3 jelas terhubung dengan X_i namun hubungannya tidak bersifat linear.

Apa akibat multikolinieritas?

Mengapa Classical Linear Regression Model (CLRM) mengasumsikan ketiadaan multikolinieritas?

Jika **multikolinieritas sempurna**. Koefisien regresi dari variabel-variabel independent menjadi **indeterminate** dan standard errornya **tak berhingga**.

Jika **multikolinieritas tidak sempurna**. Walaupun koefisien regresi **determinate** namun standard errornya besar, sehingga koefisien regresi tersebut **tidak mempunyai akurasi yang baik**.

Alasan munculnya multikolinearitas?

- 1 **Metode pengambilan data.** Sampling tidak dilakukan dengan baik.
- 2 **Multikolinearitas yang muncul dari model itu sendiri.** Pada waktu menyusun model, orang memilih variabel-variabel yang secara natural berkaitan satu sama lain. Misalnya, regresi antara penggunaan listrik terhadap income dan ukuran rumah. Namun biasanya ada kaitan antara income dan ukuran rumah. Semakin besar incomenya, semakin besar pula ukuran rumah.
- 3 **Spesifikasi model.** Misalnya menambahkan suku-suku polinomial terutama jika range dari X kecil
- 4 **Model yang berlebihan.** Jumlah variabel independen lebih banyak dari jumlah observasi.
- 5 **Trend yang serempak pada variabel independen.** Biasanya muncul pada data-data time series di mana variabel-variabel independen-nya mempunya **common trend**. Yang satu naik, yang lain ikut naik. Atau sebaliknya.

MENGAPA INDETERMINATE?

Dua variabel independen:

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{u}_i \quad (6)$$

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2} \quad (7)$$

$$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2} \quad (8)$$

Apa yang terjadi jika $X_{3i} = \lambda X_{2i}$?

MENGAPA INDETERMINATE?

Misalkan $X_{3i} = \lambda X_{2i}$ maka dapat dibuktikan bahwa

$$\hat{\beta}_2 = \frac{0}{0} \quad (9)$$

$$\hat{\beta}_3 = \frac{0}{0} \quad (10)$$

MENGAPA INDETERMINATE?

Mengapa kita memperoleh hasil yang indeterminate?

- Ingat arti dari $\hat{\beta}_2$, yaitu merupakan laju perubahan Y ketika X_2 berubah sebanyak satu unit dengan menjaga X_3 tetap konstant.
- Tetapi jika X_2 dan X_3 kolinear secara sempurna, maka mustahil untuk menjaga X_3 tetap konstan apabila X_2 dirubah nilainya. Jika X_2 berubah, maka X_3 juga berubah dengan faktor λ
- Artinya? Tidak mungkin untuk memisahkan pengaruh X_2 dan X_3 pada sampel yang tersedia.

Bagaimana jika multikolinearitas tidak sempurna?

- $\hat{\beta}_2$ dan $\hat{\beta}_3$ dapat dihitung.
- Standard error besar, sehingga hasilnya tidak akurat.
- Confidence interval lebih lebar
- Harga- t cenderung tidak signifikan, tetapi R^2 dapat sangat tinggi
- Estimasi koefisien regresi dan standard error dapat menjadi sangat sensitif terhadap perubahan kecil pada data.
- Pertanda multikolinearitas yang paling kelihatan: R^2 tinggi, H_0 ditolak berdasarkan F-test, tetapi harga- t ternyata tidak signifikan.

Download data: cars.xlsx

- 1 Lakukan analisa regresi MPG vs CYL,ENG,WGT
- 2 Lakukan analisa regresi MPG vs CYL
- 3 Lakukan analisa regresi MPG vs ENG
- 4 Lakukan analisa regresi MPG vs WGT
- 5 Lakukan analisa regresi MPG vs CYL,ENG
- 6 Lakukan analisa regresi MPG vs ENG,WGT
- 7 Lakukan analisa regresi MPG vs CYL,WGT

Buat analisa, dan tentukan apakah ada di antara ketiga variabel (CYL,ENG,atau WGT) yang perlu dihilangkan dari model.