UNIVERSITAS SULTAN AGENG TIRTAYASA

UNTIRTA Jawara

Jujur Adil Wibawa Amanah Religius Akuntabel

Analisis Regresi

#11 Meeting

Multikolinearitas

Ferdian Bangkit Wijaya, S.Stat., M.Si
NIP. 199005202024061001

Regresi Linier - Multikolinearitas

Multikolinearitas adalah sebuah fenomena di mana terdapat hubungan linear yang kuat (atau sempurna) di antara dua atau lebih variabel independen (X) dalam sebuah model regresi.
Asumsi OLS yang ideal adalah Non-Multikolinearitas, yang berarti semua variabel X kita bersifat independen satu sama lain.

Jenis Multikolinearitas

1. Multikolinearitas Sempurna (Perfect Multicollinearity)
   - Terjadi ketika satu variabel X adalah kombinasi linear sempurna dari X lainnya.
   - Contoh: X3 = 2X1 + 5X2 (sempurna)
   - Contoh klasik: "Dummy Variable Trap" (Jebakan Variabel Dummy).
     - Misal, kita punya variabel kategori Jenis Kelamin (Pria, Wanita).
     - Kita buat dua dummy: D1 = 1 jika Pria (0 jika lain), dan D2 = 1 jika Wanita (0 jika lain).
     - Jika kita masukkan kedua dummy ini ke model (bersama intersep b0), kita akan mengalami multikolinearitas sempurna, karena D1 + D2 = 1 (dan b0 juga merupakan vektor angka 1).
2. Multikolinearitas Tinggi (High Multicollinearity)
   - Ini yang paling sering terjadi di dunia nyata.
   - Terjadi ketika ada korelasi yang sangat kuat (namun tidak sempurna) antar variabel X.
   - Contoh: X3 = 2X1 + 5X2 + error kecil
   - Contoh praktis: Memasukkan variabel Umur dan Pengalaman Kerja ke dalam model yang sama. Keduanya pasti berkorelasi sangat tinggi.

Dampak Multikolinearitas

Kenapa (X’X) menjadi singular???

$$\hat{\beta} = (X'X)^{-1}X'y$$

Skenario 1: Multikolinearitas Sempurna

1. Jika X3 = 2X1, maka kolom ke-3 dalam matriks X adalah kombinasi linear dari kolom ke-1.
2. Ini membuat matriks X tidak memiliki "full rank" (rank deficient).
3. Akibatnya, matriks X'X (matriks varians-kovarians dari X) juga menjadi tidak "full rank".
4. Matriks yang tidak "full rank" memiliki Determinan = 0.
5. Matriks yang determinannya 0 disebut MATRIKS SINGULAR.
6. Matriks singular tidak bisa di-inverse.
7. Hasil: inv(X'X) tidak dapat dihitung. Estimator b_hat tidak bisa ditemukan.
8. Software statistik (seperti R atau SPSS) akan memberikan error, NA, atau "singular matrix".

UNIVERSITAS SULTAN AGENG TIRTAYASA

Dampak Multikolinearitas

UNTIRTA Jawara

Skenario 2: Multikolinearitas Tinggi (Yang Umum Terjadi)

1. Jika X3 ~= 2X1, maka determinan dari X'X tidak benar-benar 0, tapi SANGAT DEKAT DENGAN 0 (misal, 0.0000001).
2. Matriks ini secara teknis bisa di-inverse, tapi hasilnya akan "meledak" secara numerik.
3. Inverse dari matriks X'X ini inv (X'X) berisi informasi Varians dari koefisien b_hat kita.
4. Jika det(X'X) sangat kecil, elemen-elemen di matriks inversenya Var(b_hat(j) akan menjadi SANGAT BESAR (Inflated / Menggembung).

Dampak Multikolinearitas

Konsekuensi Praktis dari Varians yang Meledak:

1. Standard Error (SE) Menjadi Sangat Besar: Ingat $SE(\hat{\beta}) = \sqrt{Var(\hat{\beta})}$
2. Uji-t Menjadi Sangat Kecil: Ingat $t_{hitung} = \frac{\hat{\beta}}{SE(\hat{\beta})}$

   Jika SE (penyebut) sangat besar, t_hitung akan menjadi sangat kecil (mendekati 0).
3. p-value Menjadi Sangat Besar: Akibatnya, p-value akan besar.
4. Gejala Aneh:
   - Variabel yang secara logika seharusnya sangat signifikan, di hasil output malah tidak signifikan (karena p-value > 0.05).
   - Nilai koefisien b_hat bisa memiliki tanda yang salah (misal, harusnya positif jadi negatif).
   - Model menjadi sangat tidak stabil. Perubahan kecil pada data (menambah/mengurangi 1-2 observasi) bisa mengubah hasil b_hat secara drastis.
   - Sering terjadi: Uji-F (Uji Simultan) signifikan, tetapi tidak ada Uji-t (Uji Parsial) yang signifikan. Ini adalah gejala klasik multikolinearitas.

Deteksi Multikolinearitas

1. Deteksi Informal (Indikasi Awal)

- Gejala Aneh: (Seperti yang disebut di Slide 3) Uji F sig, tapi Uji t tidak sig; tanda koefisien salah; SE sangat besar.
- Matriks Korelasi (Pairwise):
  - Hitung matriks korelasi antar semua variabel X.
  - Cari nilai korelasi (di luar diagonal) yang tinggi, misal $|r| > 0.8$ atau 0.9.
  - Kelemahan: Metode ini hanya bisa mendeteksi korelasi antar 2 variabel (pairwise). Metode ini akan GAGAL mendeteksi jika X4 adalah korelasi dari X1 + X2 + X3 (multikolinearitas multivariat).

Deteksi Multikolinearitas

2. Deteksi Formal: Variance Inflation Factor (VIF)

- **VIF (Variance Inflation Factor):** Mengukur seberapa besar varians dari $\hat{\beta}_j$ "menggembung" (inflated) akibat adanya korelasi X_j dengan variabel X lainnya.
- **Tahapan Menghitung VIF (untuk X_j):**

1. Buat **regresi auxilier (pembantu)**: Regresikan X_j pada *semua variabel X lainnya*.

$$X_j = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + ... + v$$

(Model ini *tidak* melibatkan Y)

2. Dapatkan nilai R_j^2 dari regresi auxilier tersebut. R_j^2 ini mengukur seberapa besar X_j bisa dijelaskan oleh X lainnya.
3. Hitung VIF untuk X_j:

$$VIF_j = \frac{1}{1 - R_j^2}$$

Deteksi Multikolinearitas

- **Interpretasi VIF:**
  - $R_j^2 = 0$ (Ideal, X_j tidak ada hubungan dgn X lain) $\rightarrow$ **VIF = 1**
  - $R_j^2 = 0.5$ (X_j 50% dijelaskan X lain) $\rightarrow$ **VIF = 2**
  - $R_j^2 = 0.9$ (X_j 90% dijelaskan X lain) $\rightarrow$ **VIF = 10**
  - $R_j^2 = 0.99$ (Sangat tinggi) $\rightarrow$ **VIF = 100**
  - $R_j^2 = 1$ (Sempurna/Singular) $\rightarrow$ **VIF = ∞ (Tak terhingga)**
- **Aturan Praktis (Rule of Thumb):**
  - **VIF = 1:** Tidak ada multikolinearitas (ideal).
  - $1 <$ **VIF** < 5**:** Multikolinearitas rendah (aman).
  - $5 \leq$ **VIF** ≤ 10**:** Multikolinearitas sedang (batas waspada).
  - **VIF** > 10**:** Multikolinearitas tinggi (masalah serius, perlu ditangani).

Penanganan Multikolinearitas

1. Biarkan Saja (Do Nothing)
   - Jika tujuan utama model hanya untuk prediksi (bukan interpretasi beta), dan VIF tidak terlalu parah (misal, < 10), multikolinearitas seringkali bisa diabaikan.
2. Mengeluarkan Variabel (Drop Variable)
   - Identifikasi variabel X dengan VIF tertinggi.
   - Keluarkan variabel tersebut dari model.
   - Kelemahan: Ini bisa menyebabkan Omitted Variable Bias (masalah baru), jika ternyata variabel yang dibuang itu secara teori penting.
3. Menggabungkan Variabel (Combine Variables)
   - Jika X1 (umur) dan X2 (pengalaman kerja) berkorelasi tinggi, mungkin kita bisa buat 1 variabel baru, misal "Senioritas" (X1 + X2) atau Indeks.
   - Metode formal untuk ini adalah Principal Component Analysis (PCA), di mana kita membuat komponen baru yang tidak saling berkorelasi.
4. Menambah Data (Increase Sample Size)
   - Menambah n (jumlah observasi) dapat membantu menstabilkan estimasi dan menurunkan Standard Error. Namun, ini seringkali tidak praktis (biaya, waktu).
5. Metode Estimasi Lanjutan
   - Gunakan metode regresi yang memang dirancang untuk tahan terhadap multikolinearitas, seperti Ridge Regression atau LASSO. (Ini di luar OLS) → Akan dipelajari pada **Peminatan Komputasi.**

SEE YOU NEXT WEEK !

Ferdian Bangkit Wijaya, S.Stat., M.Si
NIP. 199005202024061001
ferdian.bangkit@untirta.ac.id