

Nama = Rani Arum Puji Susanti
NPM = 210301040
UAS = Ekonomi Publik

1. Dalam analisis regresi linear, asumsi klasik sangat penting untuk memastikan bahwa hasil model yang dibangun valid dan akurat. Berikut adalah keuntungan dari asumsi klasik dan metode untuk mengujinya

a) Asumsi Linearitas

Manfaat : Menjamin bahwa hubungan antara variabel bebas (independen) dan variabel terikat (dependen) adalah linier, yang berarti bahwa variabel terikat (dependen) berubah sebanding dengan variabel independen (bebas).

Cara Menguji : Menggunakan plot scatter antara variabel bebas dan terikat, atau memeriksa plot residual untuk mengetahui apakah ada pola sistematis yang menunjukkan ketidaklinieran.

b) Asumsi Independen (tidak ada autokorelasi)

Manfaat : Menunjukkan bahwa residual atau kesalahan prediksi dari suatu peristiwa tidak berhubungan dengan residual dari peristiwa lain.

Ini penting untuk data time series dan panel.

Cara menguji : Menggunakan Durbin Watson test untuk mendeteksi autokorelasi pada residual. Nilai mendekati 2 menunjukkan tidak adanya autokorelasi.

c) Asumsi Homoskedastisitas

Manfaat : Mengasumsikan bahwa varians residual konstan untuk setiap level variabel independen. Ini sangat penting untuk estimasi koefisien regresi yang efektif dan tidak bias.

Metode pengujian : Plot residual dibandingkan dengan nilai prediksi atau variabel independen. Heteroskedastisitas ditandai dengan pola tertentu. Homoskedastisitas juga dapat diukur dengan uji white, Breusch Pagan.

d) Asumsi Normalitas Residual :

Manfaat : Menunjukkan bahwa residual model regresi terdistribusi normal. Ini penting untuk inferensi statistik, terutama untuk uji sampel.

Cara menguji : Untuk mengetahui distribusi residual, gunakan histogram atau plot Q-Q, juga dikenal sebagai plot kumulatif-kumulatif. Uji normalitas dapat menggunakan uji statistik seperti Shapiro Wilk atau Kolmogorov Smirnov.

e) Asumsi tidak ada multikolinearitas

Manfaat : Mengasumsikan bahwa variabel independen tidak berkorelasi tinggi satu sama lain. Multikolinearitas dapat membuat estimasi koefisien tidak stabil dan sulit diinterpretasikan.

Cara menguji: Menghitung variance inflation factor (VIF) untuk mendeteksi multikolinearitas. Nilai VIF lebih dari 10 menunjukkan masalah multikolinearitas yang serius.

f) Asosiasi tidak ada variabel yang ditentukan salah (model specification)

Maksud: Memastikan bahwa model spesifikasi benar, artinya semua variabel relevan termasuk dan variabel tidak relevan tidak termasuk dalam model.

Cara menguji: Uji Ramsey RESET dapat digunakan untuk mendeteksi kesalahan spesifikasi model.

2. Ketidakefektifan data dapat terjadi karena beberapa alasan seperti kesalahan input individu saat memasukkan data, adanya data yang hilang (missing data), data yang duplikat, inkonsistensi dalam format atau nilai data, dan adanya outliers atau anomali yang signifikan. Kesalahan input data dapat diidentifikasi dengan validasi otomatis atau pemeriksaan manual. Sementara data hilang dan duplikat dapat diidentifikasi dengan pemeriksaan untuk nilai kosong atau baris identik. Inkonsistensi dapat ditemukan dengan pemeriksaan format data, dan outliers dengan menggunakan statistik deskriptif atau visualisasi seperti box plot.

Ketidaknormalan data biasanya disebabkan oleh distribusi data yang tidak simetris (skewness atau kurtosis), kehadiran outliers atau anomali, serta distribusi data yang bimodal atau multimodal. Distribusi yang tidak simetris dapat di deteksi menggunakan histogram atau plot Q-Q, outliers dapat diidentifikasi dengan box atau plot atau analisis IQR, dan distribusi bimodal atau multimodal dapat diidentifikasi melalui histogram atau density plot. Uji statistik seperti Shapiro-Wilk atau Kolmogorov-Smirnov juga dapat digunakan untuk mendeteksi ketidaknormalan distribusi data.

3. a) Regresi linear sederhana

Definisi = Analisis statistik yang memodelkan hubungan antara satu variabel independen (x) menggunakan garis lurus.

$$\text{Persamaan} = Y = \beta_0 + \beta_1 X + \epsilon$$

Contoh = Menghubungkan tinggi badan (x) dengan berat badan (Y)

a) Regresi linear berganda

Definisi = Memodelkan hubungan antara satu variabel dependen (Y) dengan dua atau lebih variabel ($X_1, X_2, \dots, X_n$)

$$\text{Persamaan} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

Contoh = Menghubungkan berat badan (Y) dengan tinggi badan (X_1) dan umur



Tahapan mengolah data regresi:

1. Pengumpulan data • Kumpulkan data yang relevan untuk variabel bebas dan terikat.
2. Eksplorasi data • Analisis deskriptif untuk memahami karakteristik dan distribusi data.
3. Pembersihan data • Tangani missing data, outliers, dan inkonsistensi.
4. Pembagian data • Pecah data menjadi set pelatihan dan pengujian.
5. Pemodelan
 - Untuk regresi sederhana gunakan metode OLS untuk menemukan β_0 dan β_1 .
 - Untuk regresi berganda gunakan metode OLS untuk menemukan $\beta_0, \beta_1, \beta_2, \dots, \beta_n$.
6. Evaluasi Model • Gunakan metrik R-squared, adjusted, dan RMSE.
7. Interpretasi Hasil • Tafsirlah koefisien regresi dan uji signifikansi statistiknya.

Contoh analisis regresi

no Regresi Linier Sederhana

- Data : Menghubungkan tinggi badan (x) dengan berat badan (y)
- Langkah-langkah
 1. Plot data : Buat scatter plot untuk melihat hubungan antara tinggi & berat.
 2. Model : $Y = \beta_0 + \beta_1 X + \epsilon$
 3. Estimasi koefisien = Gunakan software statistik (misalnya Python, R) untuk mengestimasi β_0 dan β_1 .
 4. Evaluasi = Hitung R-squared untuk melihat seberapa baik model menjelaskan variabilitas data.
 5. Interpretasi = Jika β_1 signifikan, interpretasikan bahwa setiap penambahan satu satuan tinggi badan berkorelasi dengan perubahan β_1 satuan dalam berat badan.

no Regresi Linier Berganda

- Data • Menghubungkan berat badan (Y) & tinggi badan (x_1) & usia (x_2)
- Langkah-langkah
 1. Plot data = Buat scatter plot 3D untuk melihat hubungan antar variabel
 2. Model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$
 3. Estimasi koef = Gunakan software statistik untuk mengestimasi β_0, β_1 , dan β_2
 4. Evaluasi = Hitung adjusted R-squared dan pantau multikolinearitas menggunakan VIF
 5. Interpretasi = Menjelaskan makna koefisien β_1 dan β_2 dalam konteks hubungan mereka & berat badan.