

1. Menurut pendapat anda, apakah manfaat masing-masing asumsi klasik dan bagaimana cara mengujinya?
Jawab: Asumsi klasik dalam analisis regresi linear berganda memiliki beberapa manfaat penting untuk memastikan validitas dan reliabilitas model yang digunakan. Asumsi klasik meliputi:

a. Normalitas : menyatakan bahwa distribusi error (kesalahan) dalam model regresi harus mengikuti distribusi normal. Manfaatnya adalah memastikan bahwa estimasi parameter model regresi adalah BLUE (Best Linear Unbiased Estimator). Uji Normalitas dapat dilakukan dengan uji statistik seperti Kolmogorov-Smirnov, Shapiro-Wilk, atau melalui visualisasi dengan histogram dan Q-Q plot.

b. Heteroskedastisitas : Asumsi ini, ~~the~~ varians error harus konstan untuk semua nilai variabel independen. Manfaat dari heteroskedastisitas adalah untuk memastikan estimasi yang konsisten dan efisien. Jika ada heteroskedastisitas, maka varians error tidak konstan dan estimasi model bisa tidak efisien. Uji heteroskedastisitas dapat dilakukan dengan uji Breusch-Pagan atau Glejser.

c. Autokorelasi : Asumsi (Tidak ada korelasi diantara error). Manfaat dari asumsi ini adalah untuk memastikan bahwa error bersifat acak dan tidak memiliki pola tertentu. Autokorelasi sering terjadi dalam data deret waktu. Uji autokorelasi dapat dilakukan dengan uji Durbin-Watson.

d. Multikolinearitas : Guna menentukan apakah ada korelasi tinggi antar variabel apakah ada korelasi tinggi antar variabel independen dengan model regresi linear berganda, dan variabel dependen terganggu.
- Menguji: bisa dengan SPSS, metode toleransi, dan VIF (Variance Inflation Factor).

2. Jelaskan Penyebab terjadinya Ketidakevalidan data dan Ketidaknormalan data serta metode untuk mendeteksinya

Jawab: Ketidaknormalan dan ketidakevalidan data, yang sering terjadi dalam pengolahan data, dapat mempengaruhi kualitas analisis dan hasil yang dihasilkan. Berikut ini adalah penjelasan tentang penyebab dan cara menemukannya: Faktor penyebab Ketidakevalidan data: Ketidakevalidan data terjadi ketika data tidak memenuhi persyaratan atau standar yang telah ditetapkan. Beberapa faktor umum yang menyebabkan ketidakevalidan data meliputi:

- Kesalahan Penginputan:** Kesalahan seperti salah ketik atau kesalahan format saat memasukkan data.
 - Data Hilang (Missing Data):** Data yang seharusnya ada tetapi tidak tersedia.
 - Data Duplikat:** Data yang sama yang dicatat lebih dari satu kali.
 - Data tidak konsisten:** Data yang bertentangan dengan informasi lain dalam data set.
 - Kesalahan Pengukuran:** Kesalahan pengukuran: kesalahan yang disebabkan oleh alat ukur/pengukur yang tidak akurat atau proses pengukuran yang tidak tepat.
 - Format Data yang salah:** Terdapat kesalahan dalam format data, seperti tanggal yang tidak sesuai dengan format yang diharapkan.
- Penyebab Ketidaknormalan data:** Data yang sangat berbeda dari pola umum atau data lainnya dalam kumpulan data disebut ketidaknormalan data.

Beberapa penyebab umum Ketidaknormalan data meliputi:

- Outliers** adalah nilai yang sangat berbeda dari nilai-nilai lainnya dalam kumpulan data.

- Noise** adalah data acak atau kesalahan yang mengaburkan pola data asli.
- Perubahan mendadak** adalah perubahan tiba-tiba dalam data yang tidak terduga.
- Serangan atau penipuan** adalah data yang berasal dari aktivitas yang tidak sah atau berbahaya.

Metode untuk Mendeteksi Ketidakevalidan Data

- Validasi Input:** Menggunakan aturan dan format tertentu untuk memastikan data yang dimasukkan sesuai dengan yang diharapkan.
- Konsistensi:** Memastikan data konsisten dengan data set lain.

atau aturan bisnis tertentu.

3. **Pendeteksian Duplikasi**: Menggunakan algoritma atau fungsi tertentu untuk menemukan dan menghapus data duplikat.
4. **Pemeriksaan Missing Values**: Menggunakan teknik seperti imputasi untuk menangani missing data.
5. **Verifikasi Manual**: Meminta konfirmasi dari sumber data atau melalui inspeksi manual.

Metode untuk mendeteksi Ketidake normalan Data.

1. **Statistical Methods**: Menggunakan statistik deskriptif seperti mean, median, dan standar deviasi untuk mengidentifikasi outliers.
2. **Box Plot Analysis**: Menggunakan visualisasi seperti box plot untuk melihat data yang berada diluar jangkauan (Outliers).
3. **Z-Score**: Menghitung Z-score untuk mengidentifikasi nilai yang sangat jauh dari mean.
4. **Machine Learning Techniques**: Menggunakan algoritma clustering seperti K-means untuk mendeteksi Outliers.
5. **Anomaly Detection Algorithms**: Khusus seperti Isolation Forest, One-class untuk mendeteksi anomali dalam data.
6. **Time Series Analysis**: menggunakan analisis deret waktu untuk mendeteksi perubahan mendadak atau pola yang tidak biasa.

3. Regresi linear sederhana : Teknik analisis Statistik yang digunakan memahami hubungan antar 2 variabel (x dan y)

Bedanya : Regresi linear sederhana

- Melibatkan 1 variabel Independen

Regresi linear Berganda

- Melibatkan 2/ lebih variabel independen
- lebih kompleks

Tahapan mengolah data regresi

Contoh

⇒ Pengumpulan data

⇒ Ekspolarasi data

⇒ Pembagian data

⇒ Pembangunan model

⇒ Evaluasi model

⇒ Interpretasi Model

• Regresi linear sederhana

x vs y

• Regresi linear Berganda

• x, y, z