

Jawaban:

1. Asumsi linieritas memastikan bahwa model regresi memberikan estimasi yang akurat terhadap hubungan antara variabel independen dan dependen cara mengujinya plot residual vs. nilai prediksi atau menggunakan plot scatter antara variabel independen dan dependen.

Normalitas manfaatnya hipotesis dan membentuk interval kepercayaan yang akurat cara ujinya dengan kolmogorov smirnov / Jarque-Bera

Autokorelasi manfaat menghindari kesalahan yang tidak akurat dan parameter bias efisien Uji Durbin Watson/Breusch Godfrey

Homoskedastisitas Manfaat menghindari varians yang bervariasi cara ujinya dengan Breusch Pagan/White

Multikolinearitas, untuk memastikan variabel independen tidak berkorelasi cara Uji VIF/matriks korelasi

2. Penyebab ketidakvalidan data biasanya terdapat kesalahan pada input manual, kesalahan sistem bug atau kesalahan perangkat lunak, data redundansi atau data yg tercatat lebih dari sekali, data yang hilang dan data yang tidak konsisten.

Cara mendeteksinya adalah dengan menggunakan aturan saat memasukkan data seperti format yg benar, range nilai yg diterima, memastikan konsistensi data antara berbagai sumber dan sistem metode cross checking, menggunakan alat atau skrip untuk mendeteksi dan memperbaiki kesalahan data yang hilang.

Penyebab terjadinya ketidaknormalan data adalah noise atau data yang mengandung gangguan atau informasi tidak relevan yg tidak mencerminkan fenomena sebenarnya, Anomali atau pola data yg biasanya tidak terduga, Outlier data yang secara signifikan berbeda dari sebagian besar data lainnya. Untuk mendeteksi menggunakan statistik deskriptif, visualisasi data, algoritma deteksi anomali dan analisis multivariat

3. Regresi Linear Sederhana adalah metode statistik yang digunakan untuk memahami hubungan antara dua variabel: satu variabel independen (predictor) dan satu variabel dependen (response). Model ini merepresentasikan hubungan tersebut dengan persamaan garis lurus dalam bentuk sedangkan regresi linear berganda adalah perluasan dari regresi linear sederhana, di mana hubungan antara satu variabel dependen dan dua atau lebih variabel independen dipelajari. Model ini direpresentasikan dengan persamaan

tahapan pengolahan data regresi yaitu pengumpulan data, pemeriksaan data, eksplorasi data, pemilihan model, pembagian data, pelatihan model, pengujian model, evaluasi model

Contoh regresi linear sederhana: analisis hubungan antara jumlah jam belajar dengan nilai ujian.

regresi linear berganda: analisis hubungan antara variabel independen jam belajar, jumlah latihan dan nilai ujian