General Linear Model — 3 — Residual Analysis

Sandeep Sharma · Follow

3 min read · May 30, 2022

> *A residual is a measure of how far away a point is vertically from the regression line. In simple words, difference between actual and predicted value.*

Residual is a diagnostic measure used when assessing the quality of a model. They are also known as errors.

What is a Residual?

- Formula:

  Residual = Observed value - Predicted value

$$e = y - \hat{y}$$

- Tells us how far off our prediction is
- Tells us if our prediction was too high or too low
- Both the sum and the mean of the residuals are equal to zero

We can check the assumptions of regression by examining the residuals

*Examine for linearity assumption (***Linearity)**

*Evaluate independence assumption (***Independence of errors)**

*Evaluate normal distribution assumption (***Normality of errors)**

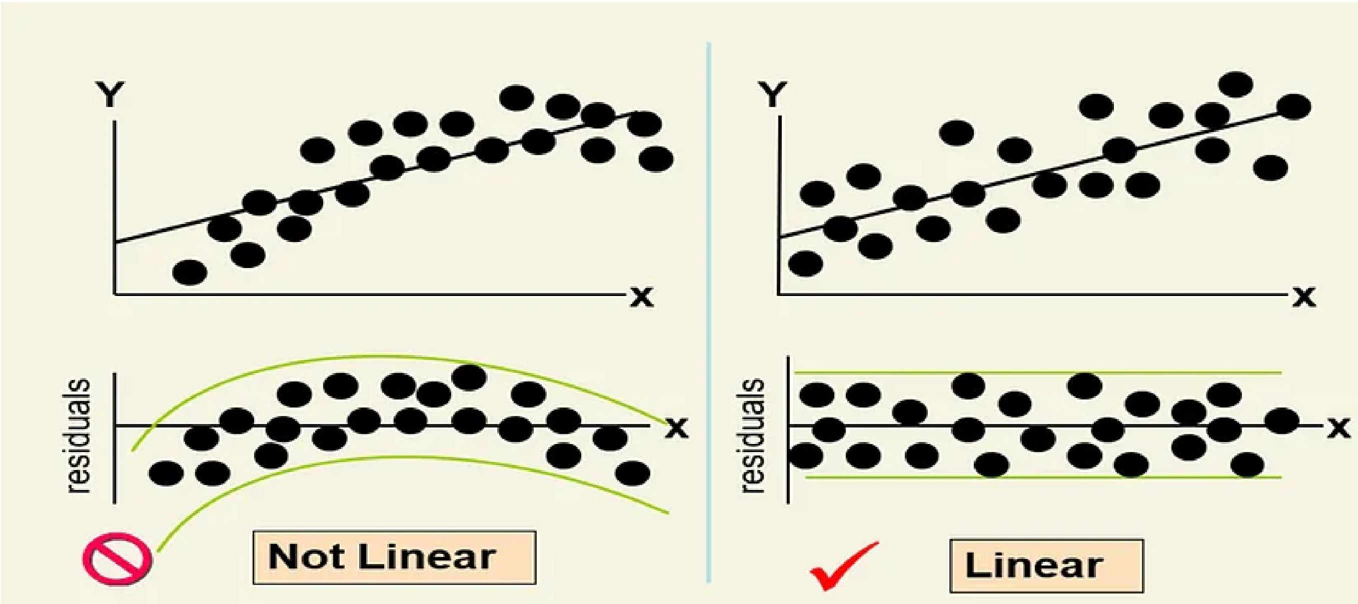
*Examine for constant variance for all levels of X (**homoscedasticity/* **Equal variance***)*

> ***Linearity:*** *The relationship between x and y must be linear. Check this assumption by examining a scatterplot of x and y.*
>
> ***Independence of errors:*** *There should not a relationship between the residuals and the predicted values. We can check this assumption by examining a scatterplot of "residuals versus fits". The correlation should be approximately 0.*
>
> ***Normality of errors:*** *The residuals must be approximately normally distributed. We can check this assumption by examining a normal probability plot.*
>
> ***Equal variances:*** *The variance of the residuals should be consistent across all predicted values. We can check this assumption by examining the scatterplot of "residuals versus fits".*

Residual Analysis for Linearity

> *A residual plot is a graph that shows the residuals on the vertical axis and the independent variable on the horizontal axis. If the points in a residual plot are*

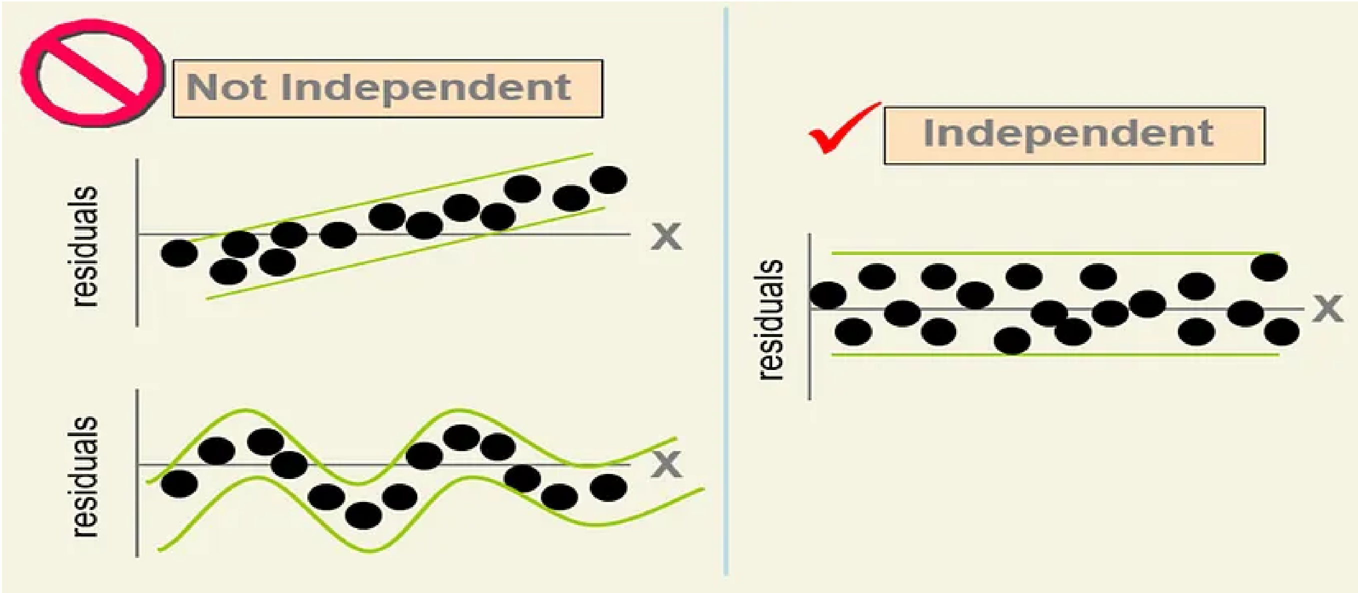
> *randomly dispersed around the horizontal axis, a linear regression model is appropriate for the data; otherwise, a nonlinear model is more appropriate.*



Linear and Non-Linear

Residual Analysis for Independence

> *we can plot the residuals vs. the sequential number of the data point. If we notice a pattern, we say that there is an autocorrelation effect among the residuals and the independence assumption is not valid.*



Residual Analysis for Independence — Another Method — Durbin Watson (D-W) Statistics

It's the sum of the squares of the differences between consecutive errors divided by the sum of the squares of all errors.

Another way to look at the Durbin-Watson Statistic is: **$D = 2(1-\rho)$**
where ρ = **the correlation between consecutive errors.**

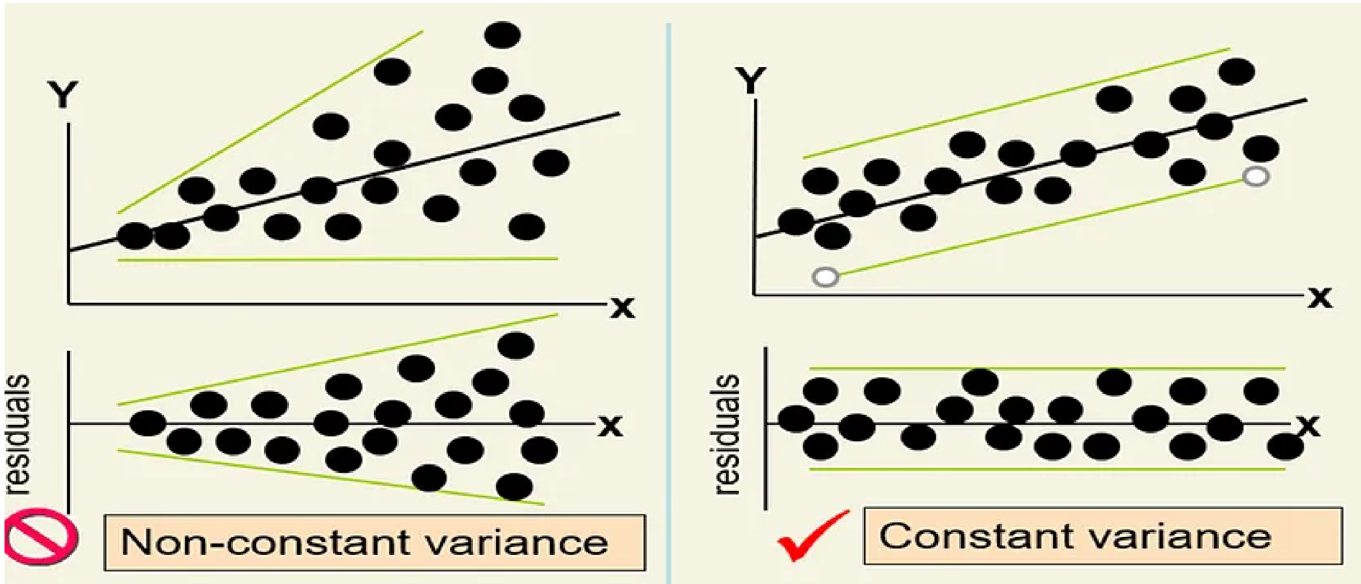
> *There are 3 important values for D:*
> *D=0: This means that $\rho=1$, indicating a positive correlation.*

> *D=2: In this case, ρ=0, indicating no correlation.*
>
> *D=4: ρ=-1, indicating a negative correlation*

Residual Analysis for Equal Variance

1. **White test statistics**

2. **Residual vs fitted value plot**



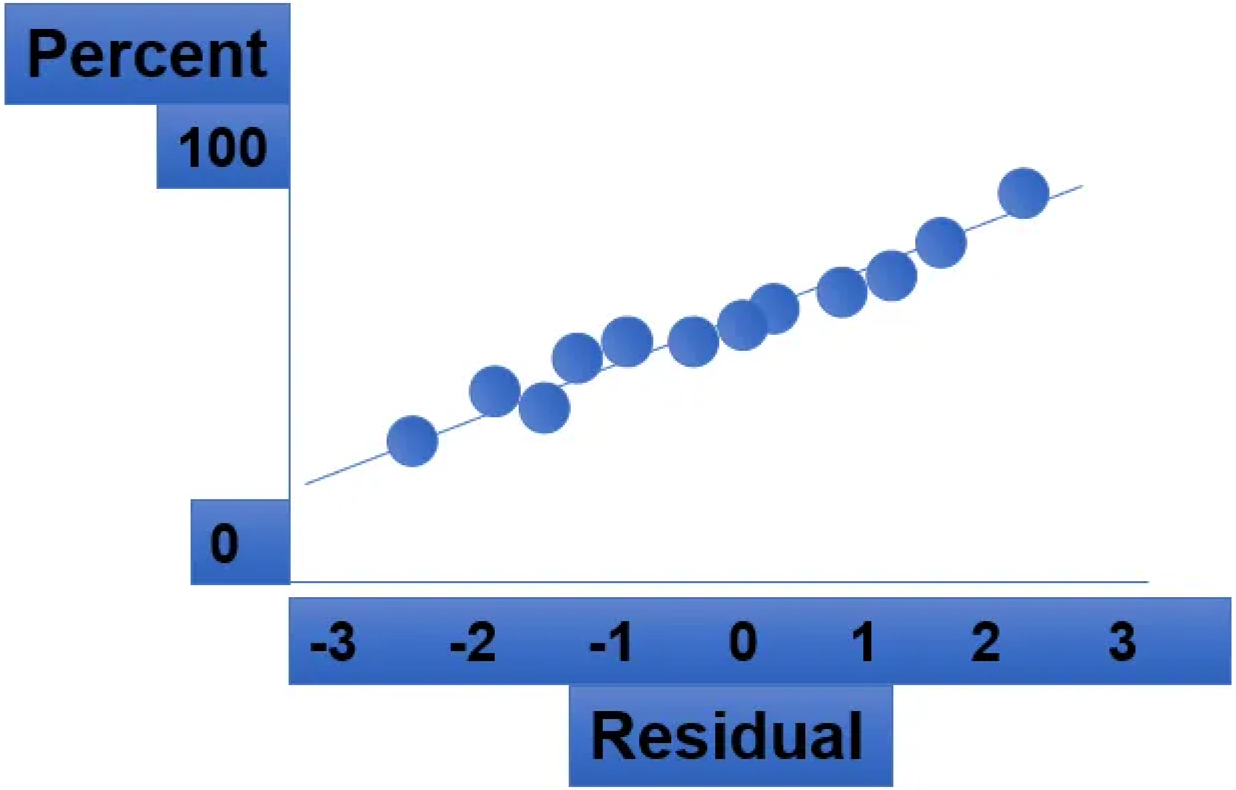
Checking for Normality

1. Examine the Stem-and-Leaf Display of the Residuals.

2. Examine the Boxplot of the Residuals
3. Examine the Histogram of the Residuals
4. Construct a Normal Probability Plot of the Residuals

Residual Analysis for Normality

When using a normal probability plot, normal errors will approximately display in a straight line.



Thank you for reading. Links to other blogs: —

General Linear Model — 2

General and Generalized Linear Models

The Poisson Distribution

Uniform Distribution

Normal Distribution

Binomial Distribution

10 alternatives for Cloud based Jupyter notebook!!

Generalized Linear Model | Linear Regression | Regression | Regression Analysis | Residual Analysis

Written by Sandeep Sharma

Follow

179 Followers

Manager Data Science — Coffee Lover — Machine Learning — Statistics — Management Consultant — Product Management — Business Analyst

More from Sandeep Sharma

Sandeep Sharma

Hessian Matrix

Wikipedia Definition — In mathematics, the Hessian matrix or Hessian is a square matrix...

4 min read · Aug 25, 2022

126 2

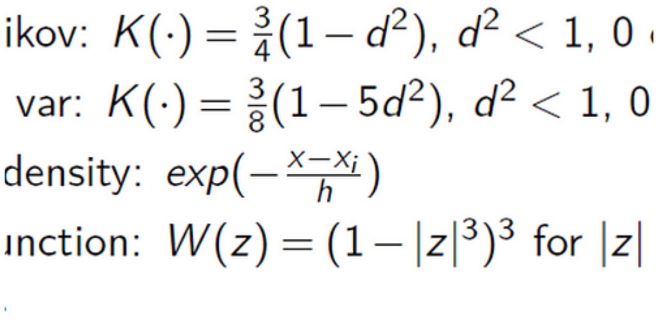
Sandeep Sharma

General and Generalized Linear Models

The General Linear Model is a framework of statistical methods to relate some number o...

3 min read · May 18, 2022

18



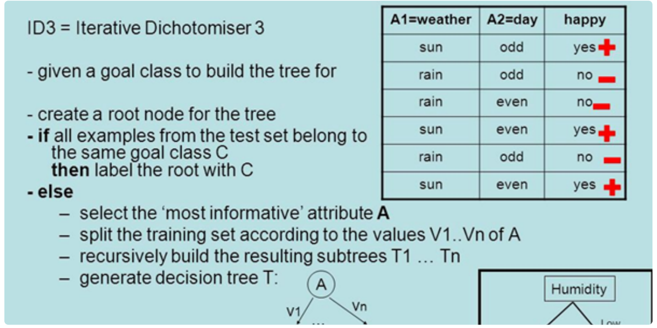
Sandeep Sharma

Non Parametric Regression

Non Parametric Regression is a form of regression analysis in which none of the...

4 min read · Jun 18, 2022

40



Sandeep Sharma

Decision Tree and its types

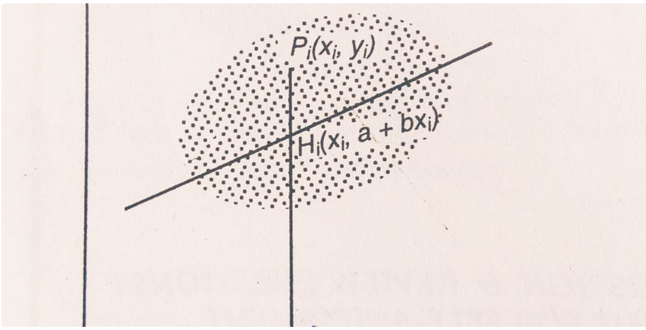
Q:—What is Decision Tree? A:—It is a set of rules used to classify data into categories....

4 min read · Nov 8, 2021

14

See all from Sandeep Sharma

Recommended from Medium



NANDINI VERMA

The Fundamentals of Linear Regression

Regression analysis is a powerful statistical technique that plays a pivotal role in data…

3 min read · Oct 5, 2023

2

Risdan Kristori in Python in Plain English

Simple Linear Regression. Clearly Explained.

Detail Explanation of Simple Linear Regression with examples in Python

6 min read · Oct 10, 2023

14

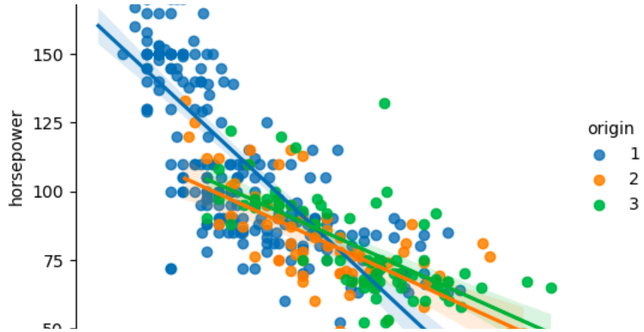
Lists

Practical Guides to Machine Learning

10 stories · 962 saves

Predictive Modeling w/ Python

20 stories · 830 saves



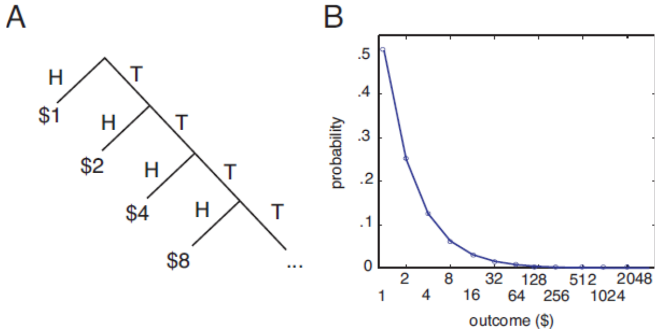
Jhon Vargas Reyes

5 Key points to train a Linear Regression model

Machine learning framework use two main ingredients, first one is the algorithms which...

8 min read · Sep 5, 2023

29 3



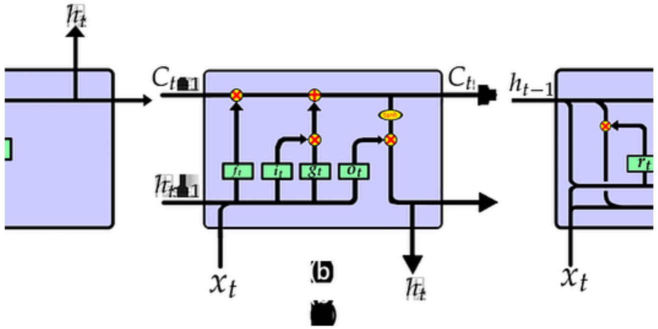
Hossam Hamdy

St. Petersburg Paradox — The Infinite Money Game

Consider a hypothetical game where a fair coin is flipped repeatedly until it comes up...

7 min read · Sep 20, 2023

103

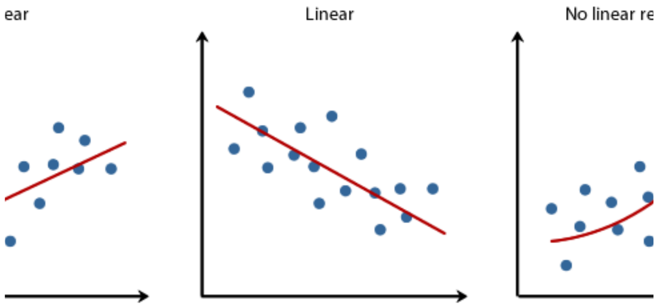


Ayesha Shabbir

Understanding RNN, LSTM, and GRU: Architectures and Challeng...

Strong deep learning models made for sequential data processing include Gated...

5 min read · Oct 27, 2023



Jaleel Adejumo

Regression Analysis from Scratch: Incorporating p-value Computation

In a previous article, I explained the concept of gradient descent and its application in a...

7 min read · Aug 13, 2023

127 2

See more recommendations