



SNS COLLEGE OF ENGINEERING



Kurumbapalayam(Po), Coimbatore – 641 107

Accredited by NAAC-UGC with 'A' Grade

Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of Information Technology

Course Name – 23ADT202 Fundamental of Data
science and Analytics

II Year / IV Semester

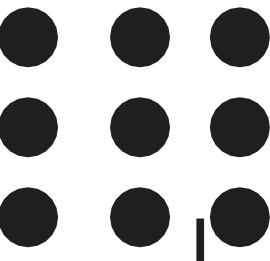
Unit 2 – Descriptive Analytics

Standard error of estimation





Introduction to SEE



- The Standard Error of Estimate (SEE) measures the accuracy of predictions made by a regression model.

It quantifies the average distance between observed and predicted values.

A smaller SEE indicates a better fit between the regression model and data.

SEE helps assess the model's prediction reliability.



Definition of SEE



SEE is the standard deviation of residuals in a regression model.

Residuals are the differences between observed values (Y) and predicted values (\hat{Y}).

Formula:

$$SEE = \sqrt{[\Sigma(Y_i - \hat{Y}_i)^2 / (n - 2)]}$$



Steps to Calculate SEE



1. Calculate the predicted values using the regression equation.
2. Compute the residuals (differences between observed and predicted).
3. Square the residuals and sum them.
4. Divide the sum by $(n - 2)$, then take the square root.



Example Dataset



$$X = [1, 2, 3, 4, 5]$$

$$Y = [2, 4, 5, 4, 5]$$

Use these values to calculate the SEE for a regression model.
SEE will help assess how well the model fits the data.



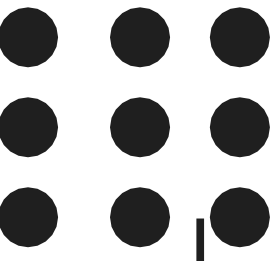
Interpreting SEE - Low SEE



A low SEE indicates that the data points are close to the regression line. This represents a good fit between the model and the data. It suggests that the model is making accurate predictions. A low SEE is desirable for reliable regression analysis.



Interpreting SEE - High SEE



A high SEE indicates significant deviations between observed and predicted values. This represents a poor fit between the model and the data. The model may not be accurately capturing the relationship between variables. High SEE suggests that predictions are unreliable.



Visualization of SEE



A scatter plot with the regression line and residual lines can help visualize SEE. The closer the residuals are to zero, the smaller the SEE. Large residuals indicate higher SEE and a poorer model fit. Residual plots help detect patterns in errors that the model cannot capture.



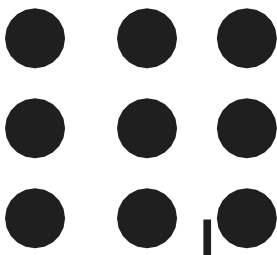
Applications of SEE



1. **Business:** Evaluate sales forecast accuracy based on advertising spend.
 2. **Healthcare:** Assess the reliability of models predicting patient recovery times.
 3. **Education:** Measure the accuracy of student performance predictions based on study hours.
- SEE helps in improving decision-making across industries.



Exercise & Conclusion



Exercise 1: Calculate the SEE for $X = [10, 20, 30, 40]$, $Y = [15, 25, 35, 45]$, Predicted $Y = [14, 24, 34, 44]$.

Exercise 2: Plot the residuals for $X = [1, 2, 3, 4]$, $Y = [5, 6, 7, 8]$, Predicted $Y = [5.2, 5.8, 7.1, 7.9]$.

SEE provides valuable insights into the quality of regression models.

Understanding SEE enhances model evaluation and prediction accuracy.



THANK YOU