Fits & Residuals

Become Future Fit

Canopus
Business Management Group

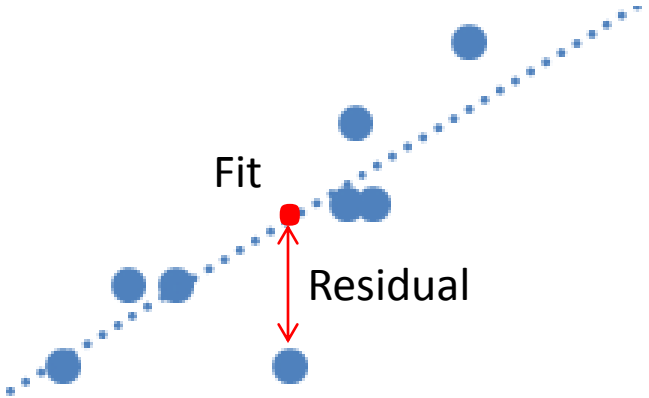
You will learn

What are fits, residuals & how it relates to Least Squares Method

Level of Difficulty

High

Fits & Residuals



Least Squares Method

The true “best fit” model minimizes the total deviation (residuals) of the observations from the model.

Fits & Residuals

Fit : It is the predicted value at some X

Residual : It is an Error. The difference between the actual value and the predicted value at some X

Properties of Residuals

1. The average and the sum value of the Residuals for any given data set = 0
2. Are distributed normally around zero
3. Have homogeneous variance
4. Are independent of each other

Fits & Residuals

Lets compute fits and residuals in Minitab

Fits & Residuals Analysis is mandatory for validation of Regression Model

Properties of Residuals

1. The average and the sum value of the Residuals for any given data set = 0
2. Are distributed normally around zero
3. Have homogeneous variance
4. Are independent of each other

Multiple Linear Regression

Become Future Fit

Canopus
Business Management Group

You will learn

How to construct a Multiple Linear Regression

Level of Difficulty

High

Improving Output per employee

A retail organization wants to find out what factors impact its Sales Revenue of channel partners. Data is collected different partners for the following influencing factors:

- Enquires
- Labor Deployment
- Avg Response Time
- No. of Stock outs

Identify the factors and build a regression model.

Steps Involved

1. Screening of factors using scatter plot & correlation coefficient
2. Study Partial Correlation
3. Run Regression
    1. Hypo Test
    2. Rsq Vs Rsq Adj
    3. Multi collinearity check
    4. Subset Analysis
    5. Re Run Regression

Regression Hypothesis

$$H_0 : b_1 = b_2 = \ldots = b_n = 0$$

(i.e. all slopes = 0)

$$H_a : b_j \neq 0 \text{ for at least one } j$$

(i.e. at least one X is significant)

Importance of R-Sq

Become Future Fit

Canopus®
Business Management Group

You will learn

What is R-sq & Adj R-sq?
How is it computed?
What is its significance?

Level of Difficulty

High

R-sq

Coefficient of Determination, represents the quantum of variation in Dependent Variable (Y) that can be expressed by the Independent Variables (Xs)

Coefficient of Determination

Canopus Business Management Group

R^2 value > 0.65 (65%)

Regression equation is complete

Xs included in study explain most of the variation in Y

R^2 value < 0.65 (65%)

- Regression equation is not complete
- Xs included in study can't explain most of the variation in Y
- Include additional Xs
- Re-run regression till R-sq >0.65

R-sq

What is Adjusted R-sq ?

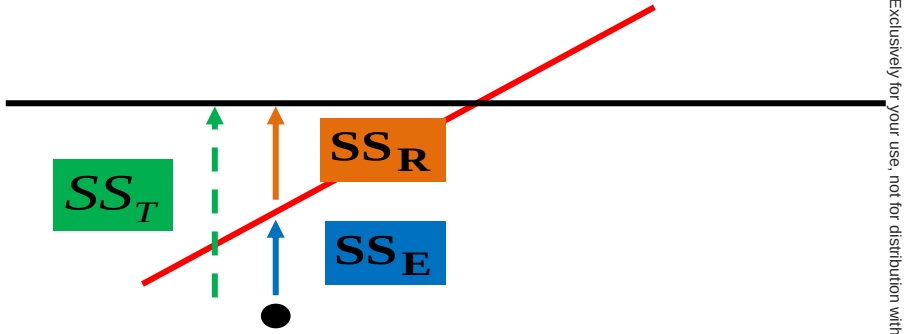
R-sq Adjusted

Canopus Business Management Group

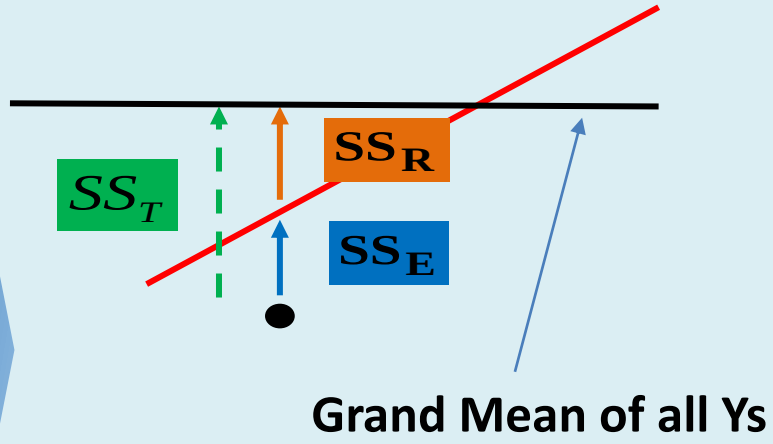
R-sq value almost always Increases but never decreases when a factors added to the model, regardless of whether the predictor was statistically significant.

Adjusted R-sq value does not increase with the addition of more factors. It will decrease when the predictor is insignificant.

Sum of Squares Computation



Sum of Squares Computation



$$SS_T = SS_R + SS_E$$

Where

SSR = Sums of Squares for the Regression

SSE = Sums of Squares for the Error

SST = Sums of Squares for the Total

R-sq Formula

$$SS_T = SS_R + SS_E$$

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}$$

Where…

SSR = Sums of Squares for the Regression

SSE = Sums of Squares for the Error

SST = Sums of Squares for the Total

R-sq = coefficient of determination

R-sq Formula

$$SS_T = SS_R + SS_E$$

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}$$

R-sq Formula

$$SS_T = SS_R + SS_E$$

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}$$

$$= 1 - \frac{\text{SS Error}}{\text{SS Total}} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Multi-Collinearity

Become Future Fit

Canopus
Business Management Group

You will learn

What is Multi-Collinearity & how to detect it?

Level of Difficulty

High

Multi-Collinearity

Lets check if any dependent variables are influencing other dependent variables…

Variance Inflation Factor

Variance Inflation Factor (VIF) quantifies the severity of multicollinearity

Guidelines for VIFs:

- 1.0: absolutely no collinearity.
- <5.0: little collinearity.
- >10.0: major collinearity.
- From 5 to 10: gray area.

Partial Correlation

Become Future Fit

You will learn

What is partial correlation and how to compute it?

Level of Difficulty

Low

Studying Relationships

Data of Productivity, Competency Score and Experience for employees is given. Both Competency Score and Experience impact Productivity. But you want to know what is the correlation between Productivity and Competency, if we exclude the impact of experience on both?

Partial-correlation

Is the correlation between two variables after removing the effect of one or more additional variables

Covariance - Measures the linear relationship between two variables. Covariance is not standardized, unlike the correlation coefficient. Signs matter and not value.

Autocorrelation – is the correlation between observations of a same time series.

Cross-correlation – is the correlation between observations of two different time series.