Quality Improvement through Exploratory Data Analysis

Lecture XVI

[Chapter 11 in textbook]

Exploratory Data Analysis

Correlation

Regression

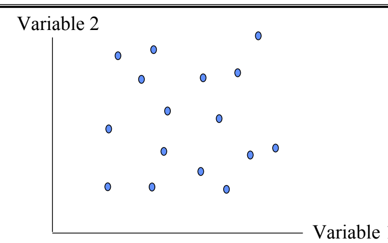
Exploratory Data Analysis

Correlation

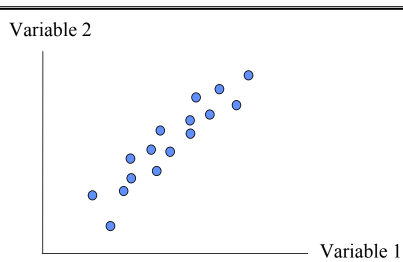
versus

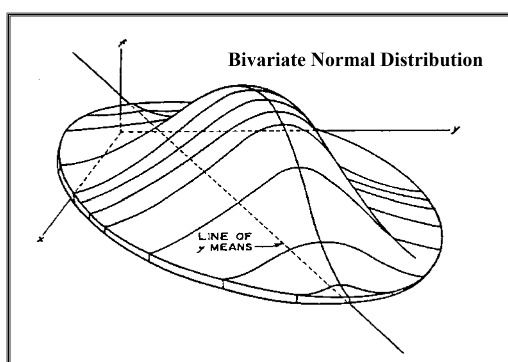
Causation

Low Correlation

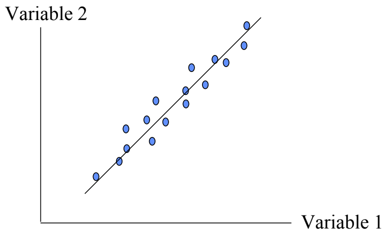


High Correlation

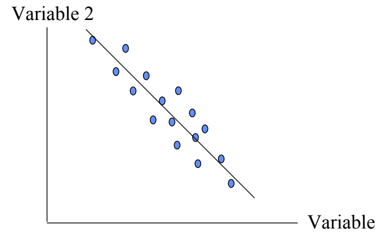




Positive Correlation



Negative Correlation



Example of Correlation Matrix

	V1	V2	V3	V4	V5
V1	1.0	.85	.24	.50	.70
V2	.85	1.0	.63	.57	.80
V3	.24	.63	1.0	.32	.44
V4	.50	.57	.32	1.0	.72
V5	.70	.80	.44	.72	1.0

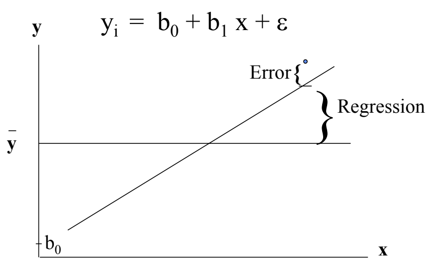
Regression

Is the Model (Regression Equation) Better at Prediction than the Grand Mean?

y_i = (intercept) + (slope)*(x value)

Model

$$y_i = b_0 + b_1 x + \varepsilon$$



Procedure for Regression

Partitioning Variation Into

Sum of Squares Regression (SSR)
Sum of Squares Error (SSE)

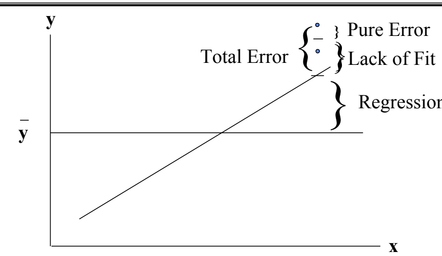
Regression Analysis Measures

α : *Reliability*

R^2 : Coefficient of Determination

[Percent of Variance Accounted For]

Lack of Fit



Procedure for Regression

Partitioning Variation Into

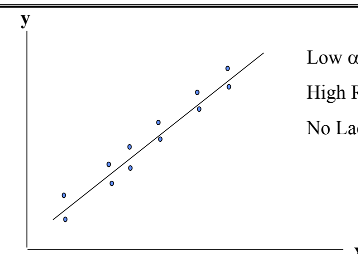
Sum of Squares Regression (SSR)

Sum of Squares Error (SSE)

SS Lack of Fit

SS Pure Error

Evaluation of Regression Model

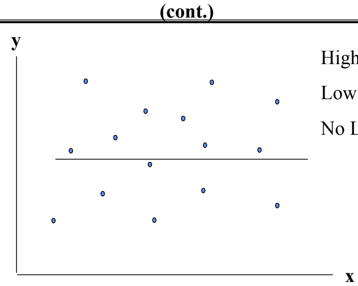


Low α

High R^2

No Lack of Fit

Evaluation of Regression Model (cont.)

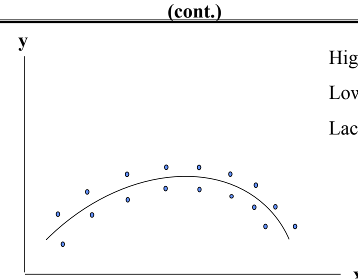


High α

Low R^2

No Lack of Fit

Evaluation of Regression Model (cont.)

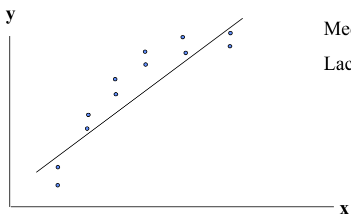


High α

Low R^2

Lack of Fit

Evaluation of Regression Model
(cont.)



Medium α

Medium R^2

Lack of Fit

Multiple Regression

More than 1 Dependent (Predictor) Variable

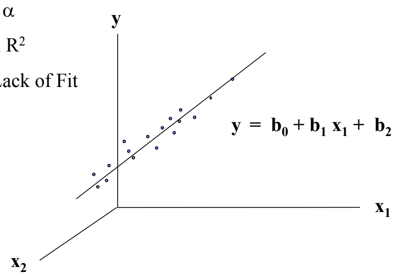
$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \varepsilon$$

Example of Multiple Regression

Low α

High R^2

No Lack of Fit

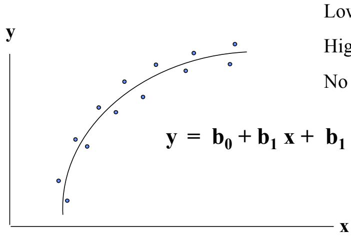


Polynomial Regression

Quadratic and Higher Order Terms

$$y = b_0 + b_1 x + b_1 x^2 + b_3 x^3 + \varepsilon$$

Example of Polynomial Regression



Low α

High R^2

No Lack of Fit

Stepwise Regression

Selects the Variable that Account for the Most Variation First. (Highest R^2)

Next, Selects the Combination of Two Variables That, Together, Account for the Most Variation

And So On.