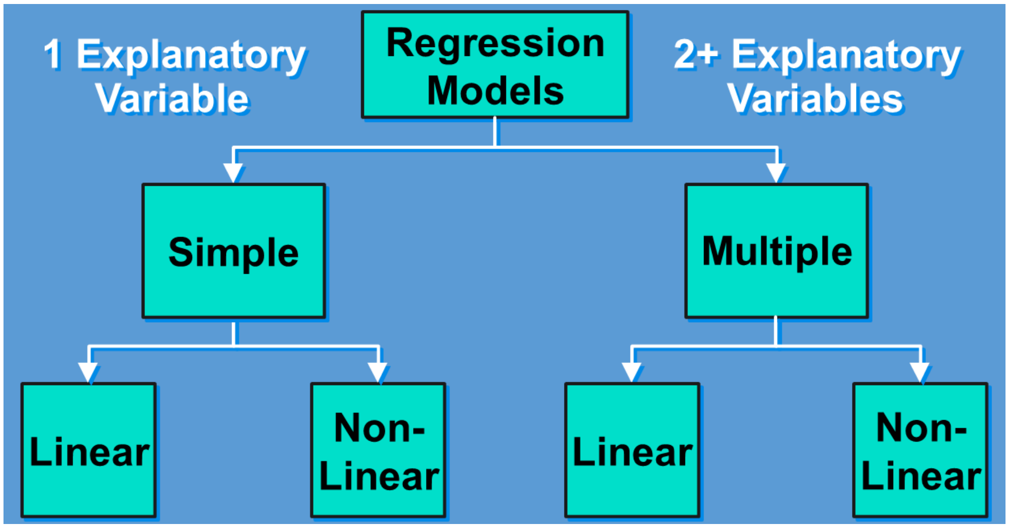
Types of Regression Models

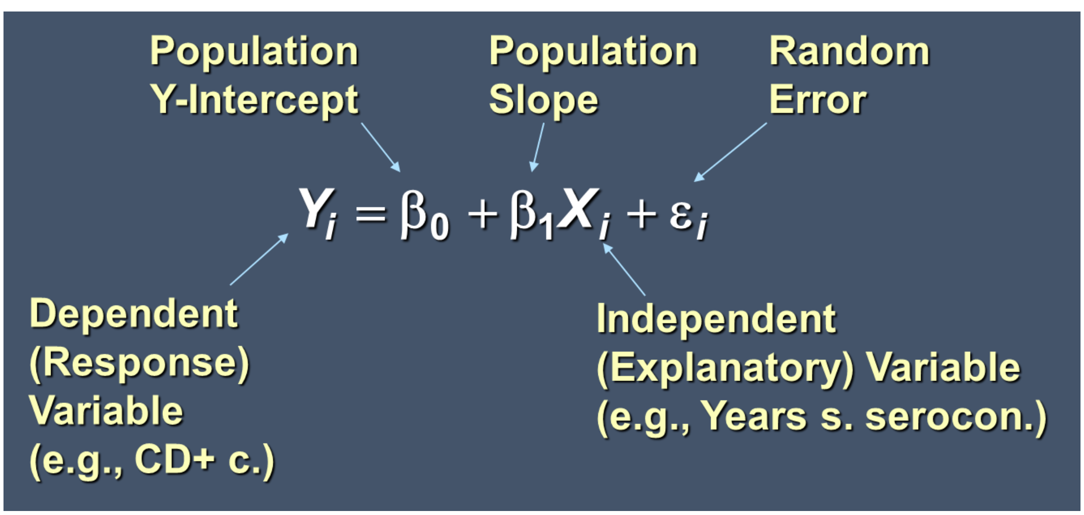


Simple Regression

- Simple regression gives us the ability to estimate the mathematical relationship between a dependent variable (usually called y) and an independent variable (usually called x).
- The dependent variable is the variable for which we want to make a prediction
- Regression Analysis was first developed by Sir Francis Galton, who studied the relation between heights of sons and fathers.

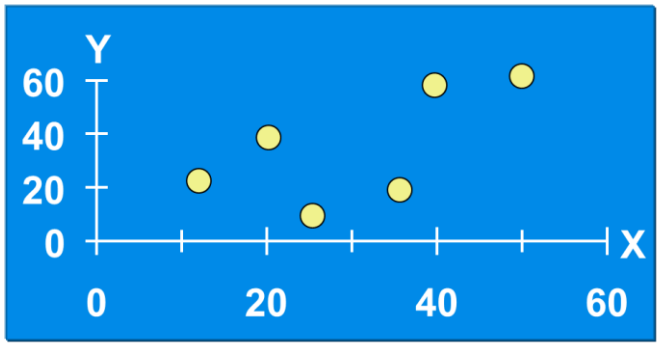
Simple Linear Regression

Relationship between Variables is a Linear Function

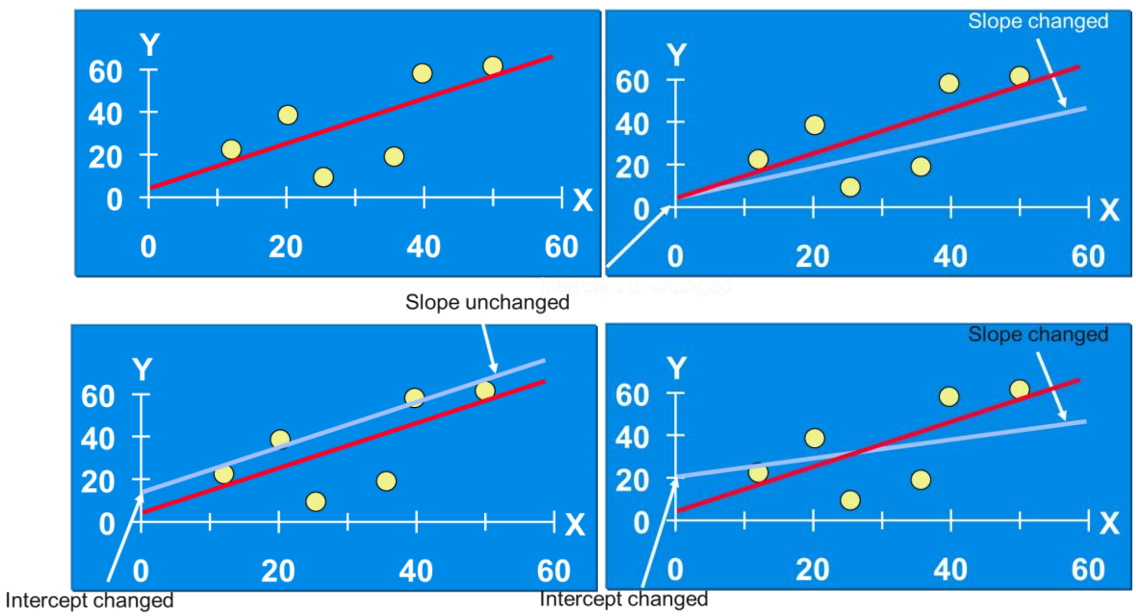


Parameter Estimation of Simple Linear Regression Model using Least Squares Method

Plot of All (X_i, Y_i) Pairs

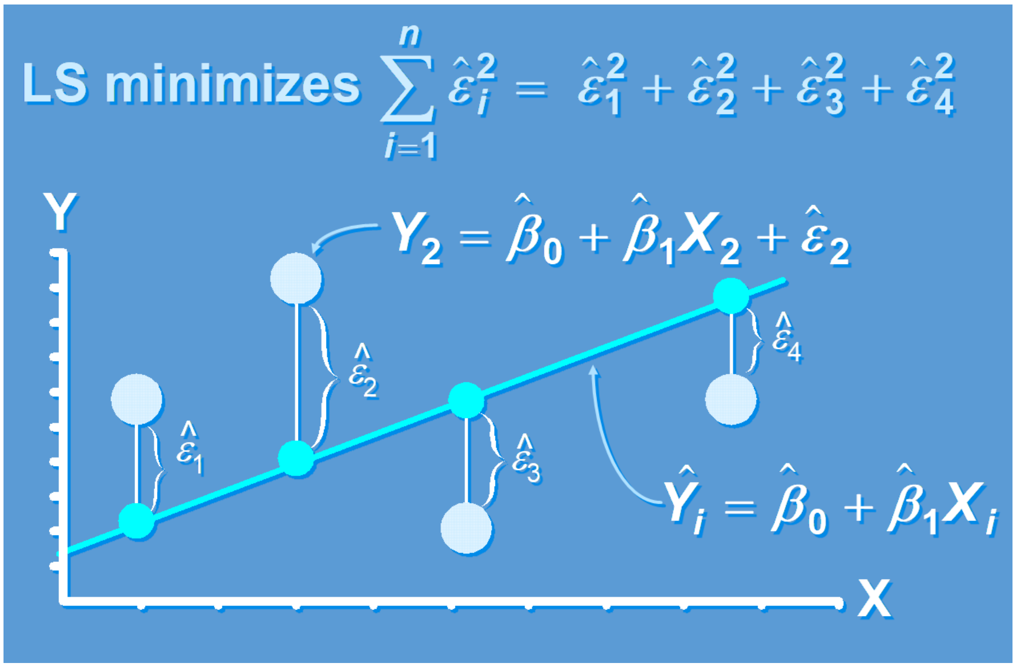


How would you draw a line through the points?   How do you determine which line 'fits best'?



‘Best Fit’ Means Difference Between Actual Y Values & Predicted Y Values Are a Minimum. *But* Positive Differences Off-Set Negative ones. Hence, LS Minimizes the Sum of the Squared Differences (errors) (SSE)

$$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

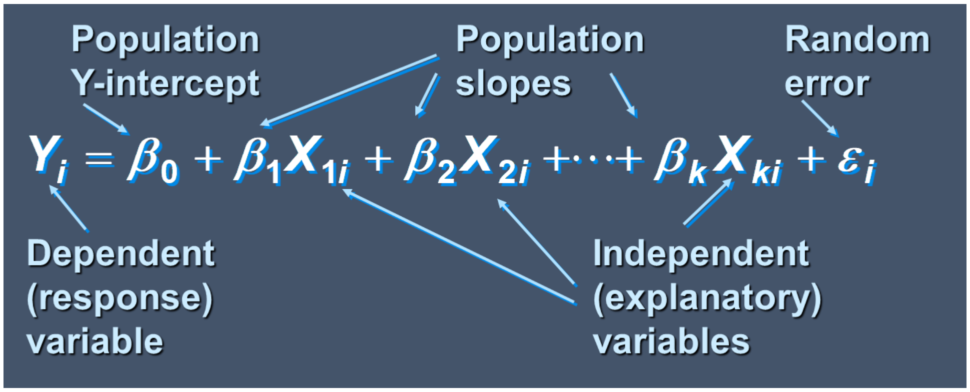


- 1. Slope ($\hat{\beta}_1$)
  - Estimated Y Changes by $\hat{\beta}_1$ for Each 1 Unit Increase in X
    - If $\hat{\beta}_1 = 2$, then Y Is Expected to Increase by 2 for Each 1 Unit Increase in X
- 2. Y-Intercept ($\hat{\beta}_0$)
  - Average Value of Y When $X = 0$
    - If $\hat{\beta}_0 = 4$, then Average Y Is Expected to Be 4 When X Is 0

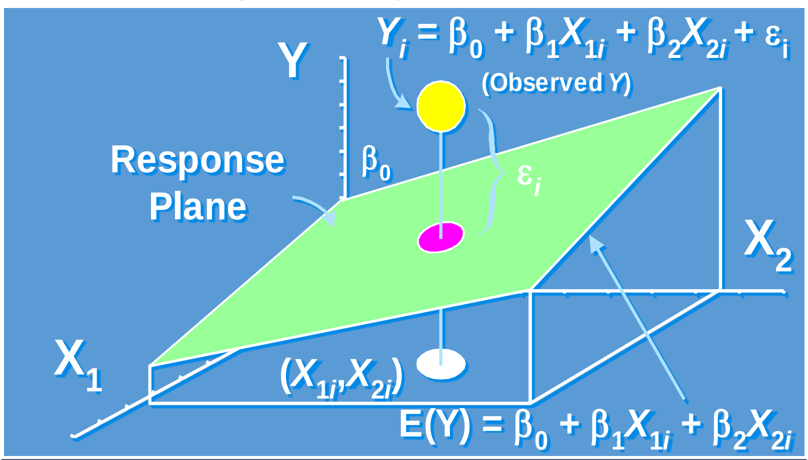
Multiple Linear Regression

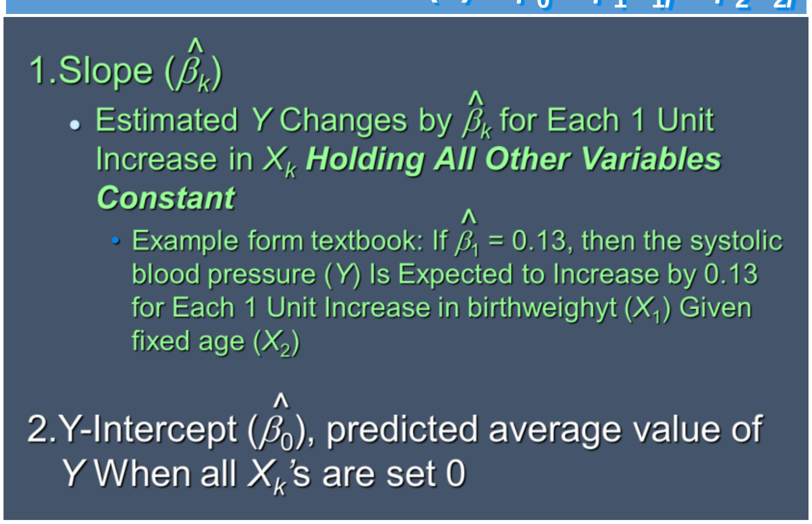
Relationship between 1 dependent & 2 or more independent variables is a linear function

- Decide what you want to do and select the dependent variable
- List all potential independent variables for your model



Parameter Estimation using Least Squares





Model Evaluation

- Proportion of Variation in *Y* 'Explained' by All *X* Variables **Taken Together**

$$R^2 = \frac{\text{Explained variation}}{\text{Total variation}} = \frac{SS_{yy} - SSE}{SS_{yy}} = 1 - \frac{SSE}{SS_{yy}}$$

SSE = Residual Error
SS_{yy} = Variance

- R^2 Never Decreases When New *X* Variable Is Added to Model (Disadvantage When Comparing Models)
- Solution: Adjusted R^2
  - Each additional variable reduces adjusted R^2, unless SSE reduces enough to compensate

$$R_a^2 = 1 - \left[\frac{n-1}{n-(k+1)}\right]\frac{SSE}{SS_{yy}} \le 1 - \frac{SSE}{SSyy} = R^2$$