5 assumptions of Simple Linear Regression (SLR)

- **Existence –** for any fixed value of X, Y is a Random Variable with finite mean & variance
  - this defines a set of conditional RVs: $Y|X=x$
- **Independence –** Y_i are independent of each other
  - the Y_i are ***I***ndependent & ***I***dentically ***D***istributed (***iid***) RVs
- **Linearity –** the mean value of Y is a straight-line function of X
  - The SLR equation describes $\mu_{Y|X=x}$ in addition to $Y|X=x$

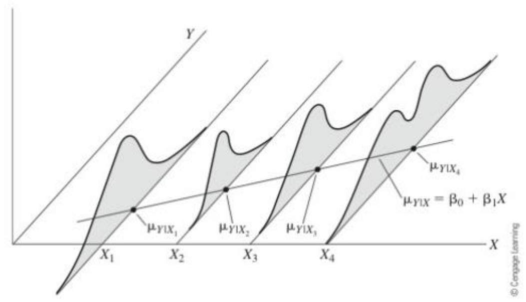
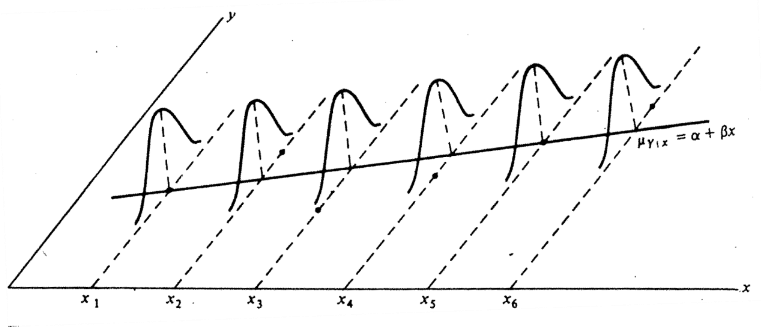


FIGURE 5.5 Straight-line assumption

- **Homoskedasticity –** the variance of Y is the same for any value of X
- **Normality –** for any fixed value of X, Y has a normal distribution
  - $Y_i \,|\, X=x \overset{iid}{\sim} N(\mu_{Y|X=x}, \sigma^2)$ with no subscript on σ^2



Statistical Model for Simple Linear Regression

- $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ $i = 1, 2, \ldots, n$
- $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ $e_i = Y_i - \hat{Y}_i$ are the <u>errors</u> or <u>residuals</u>

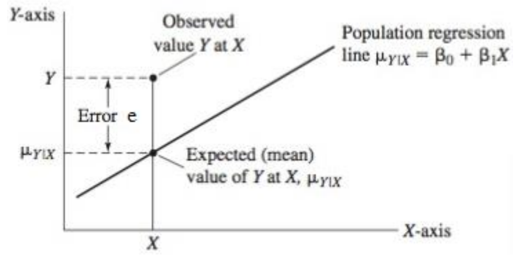


FIGURE 5.6 Error component e

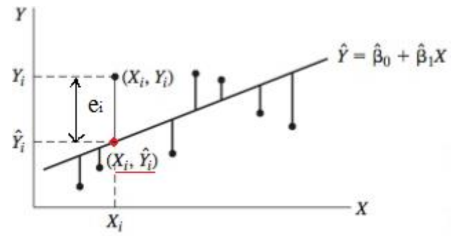


FIGURE 5.7 Deviations of observed points

- <u>Least Squares Estimates</u> (LSE) for β_0 and β_1
  - Minimize the Sum of Squared Errors (SSE)

$$S = SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \left[\sum x_i \sum y_i\right]/n}{\sum x_i^2 - \left(\sum x_i\right)^2/n} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{SSXY}{SSX}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$