

Simple Linear Regression: Inferential Methods

Pop (x, y)

$$y = \alpha + \beta x + e$$

sample (x, y)

random sample

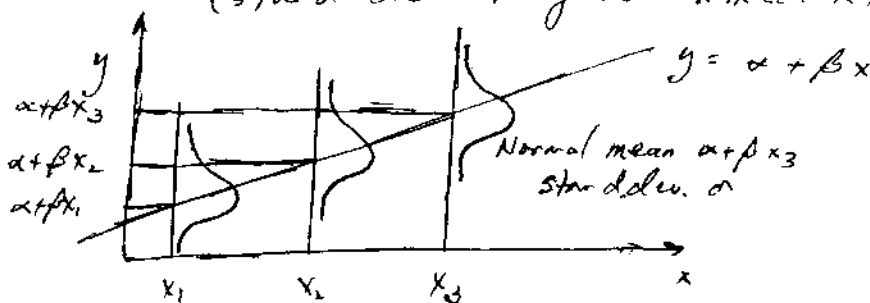
$$\hat{y} = a + b x$$

Assumptions about e : 1) has mean zero 2) has stand. dev. σ which doesn't depend on x , 3) is normally distributed 4) the random deviations e_1, \dots, e_n associated with different observations are independent of one another.

For any fixed x value, y itself has a normal distribution with

$$\left(\begin{array}{l} \text{mean of } y \text{ value} \\ \text{for fixed } x \end{array} \right) = \left(\begin{array}{l} \text{height of the pop.} \\ \text{regression line above } x \end{array} \right) = \alpha + \beta x$$

$$\left(\text{stand. dev of } y \text{ for fixed } x \right) = \sigma$$



Estimate the pop. regression line with the least squares line

$$b = \text{point est. of } \beta = \frac{S_{xy}}{S_{xx}} \quad S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$$

$$a = \text{point est. of } \alpha = \bar{y} - b\bar{x}$$

$$\text{so } \hat{y} = a + bx \text{ estimates } y = \alpha + \beta x$$

The statistic used for estimating the variance σ^2

$$s_e^2 = \frac{SS_{\text{Resid}}}{n-2}$$

$$SS_{\text{Resid}} = \sum (y - \hat{y})^2 = \sum y^2 - a \sum y - b \sum xy$$

the estimate of σ is $s_e = \sqrt{s_e^2}$

The coefficient of determination is $r^2 = 1 - \frac{SS_{\text{Resid}}}{SS_{\text{Total}}}$

$$\text{where } SS_{\text{Total}} = \sum (y - \bar{y})^2 = \sum y^2 - \frac{(\sum y)^2}{n} = S_{yy}$$

The value of r^2 can be interpreted as the proportion of the observed y variation that can be explained by the model relationship

Inference b has sampling distribution that is Normal mean β and stand dev $\sigma_b = \frac{\sigma}{\sqrt{S_{xx}}}$

$$\text{CI for } \beta \quad b \pm t^* s_b \quad \text{H.T for } H_0: \beta = \beta_0 \quad H_A: \beta \neq \beta_0 \quad t = \frac{b - \beta_0}{s_b}$$