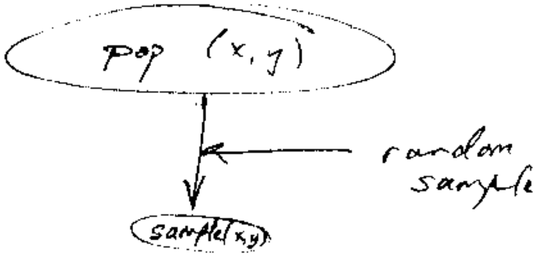
Simple Linear Regression : Inferential Methods



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

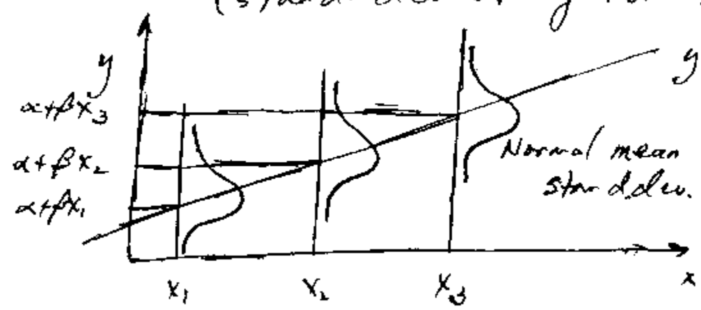
$y = a + b x$

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

$\begin{pmatrix} \text{mean } y \text{ value} \\ \text{for fixed } x \end{pmatrix} = \begin{pmatrix} \text{height of the pop.} \\ \text{regression line above } x \end{pmatrix} = \alpha + \beta x$

(stand. dev of y for fixed x) $= \sigma$



Estimate the pop. regression line with the least squares line

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

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

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

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

The statistic used for estimating the variance σ^2

$s_e^2 = \frac{SSResid}{n-2}$ $\quad SSResid = \Sigma(y-\hat{y})^2 = \Sigma y^2 - a\Sigma y - b\Sigma xy.$

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

The coefficient of determination is $r^2 = 1 - \frac{SSResid}{SSTotal}$

where $SSTotal = \Sigma(y-\bar{y})^2 = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = S_{yy}$

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 $\sigma_b = \frac{\sigma}{\sqrt{S_{xx}}}$

estimated stand dev $s_b = \frac{s_e}{\sqrt{S_{xx}}}$

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