Analyze

Straight Line Equation	$y = \beta_0 + \beta_1 x$ $\beta_0: y - intercept\ when\ x = 0$ $\beta_1: slope\ of\ line$ $y_{observed} = (mean\ value\ of\ y\ for\ a\ given\ value\ of\ x) + random\ error$ $y = \beta_0 + \beta_1 x + \varepsilon$ Random error: difference between $y_{observed}\ and\ y_{mean}$
The Method of Least Squares	Best-fitting line prediction equation (regression line): $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ $\hat{y}: predicted\ value\ of\ y\ obtained\ from\ the\ fitted\ line$ $\hat{\beta}_0\ \&\ \hat{\beta}_1: estimates\ of\ the\ true\ \beta_0\ \&\ \beta_1$ Principle of Least Squares: $Sum\ of\ Squared\ Errors = SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ *By substituting the value of* $\hat{y}_i$: $Sum\ of\ Squared\ Errors = SSE = \sum_{i=1}^{n} [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$ Least square estimates of $\hat{\beta}_0\ \&\ \hat{\beta}_1$ can be estimated as follows: $S_{x^2} = \sum_{i=1}^{n} {x_i}^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}$ $S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}$ $\hat{\beta}_1 = \frac{S_{xy}}{S_{x^2}} \qquad and \qquad \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$

Analyze

Calculating S_e^2, an estimator of σ_ε^2

σ_ε^2 : variability of random errors (ε)

$$\hat{\sigma}_\varepsilon^2 = \frac{SSE}{n-2}$$

($\hat{\sigma}_\varepsilon^2$ sometimes referred to as S_e^2).

SSE can also be written as :

$$SSE = S_{y^2} - \hat{\beta}_1 S_{xy}$$

$$= S_{y^2} - \frac{(S_{xy})^2}{S_{x^2}}$$

where:

$$S_{y^2} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

95 % of values fall in the region of $\pm 1.96\, \hat{\sigma}_\varepsilon$ (as a check for σ_ε).

Inferences concerning the slope β_1 of a line

Null hypothesis $H_0 : \beta_1 = 0$

Alternate hypothesis $H_1 : \beta_1 \neq 0$

test statistic is t-distribution with $(n-2)$ degrees of freedom.

$$t = \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}}, \quad \text{where } S_{\hat{\beta}_1} = \frac{\hat{\sigma}_\varepsilon}{\sqrt{S_{x^2}}}$$

if $\beta_1 \neq 0 \Rightarrow$ linear relationship

Standard deviation of the slope =

$$S_b = \frac{S}{\sqrt{S_{x^2}}}$$

use critical value of t at $\alpha = 0.025$ (two-tail). e.g. ($t_{0.025,(n-2)}$).

then, if calculated $t > t_{critical}$ for +ve slope

or " " $< -t_{critical}$ for -ve slope

then the null hypothesis indicating that calculated results are highly significant is rejected.

(see example pp. IX-16).