Simple Linear Regression

A first order (straight line) probabilistic model

$$y = \beta_0 + \beta x + \varepsilon,$$

where $y =$ dependent variable, $x =$ independent variable, $E(y) = \beta_0 + \beta_1 x =$ deterministic component, $\varepsilon =$ random error component, $\beta_0 =$ y–intercept of the line, $\beta_1 =$ slope of the line.

The least–square approach

Sum of Squared Errors (SSE) can be found as

$$SSE = \sum (y_i - \widehat{y}_i)^2 = \sum (y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i))^2,$$

where $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x$ is called the least squares lines.

Formulas for the least squares Estimates

Slope: $\widehat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$;
y– intercept: $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$, where
$SS_{xy} = \sum (x_i - \overline{x})(y_i - \overline{y}) = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$,
$SS_{xx} = \sum (x_i - \overline{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$,
$n =$Sample size.

Estimation of σ^2 for a straight-line model

$$s^2 = \frac{SSE}{n-2},$$

where $SSE = \sum (y_i - \widehat{y}_i)^2 = SS_{yy} - \widehat{\beta}_1 SS_{xy}$ and $SS_{yy} = \sum (y_i - \overline{y}_i)^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$.
Moreover,

$$s = \sqrt{s^2} = \sqrt{\frac{SSE}{n-2}}.$$

Making inferences about the slope β_1

One–tailed test	Two–tailed test
$H_0: \ \beta_1 = 0$	$H_0: \ \beta_1 = 0$
$H_a: \ \beta_1 < 0$ (or $H_a: \ \beta_1 > 0$)	$H_a: \ \beta_1 \neq 0$
Test statistic: $t = \frac{\widehat{\beta}_1}{s_{\widehat{\beta}_1}} = \frac{\widehat{\beta}_1}{s/\sqrt{SS_{xx}}}$	
Rejection region: $t < -t_\alpha$ (or $t > t_\alpha$)	$\lvert t \rvert > t_{\alpha/2}$,

where t_α and $t_{\alpha/2}$ are based on $(n-2)$ degrees of freedom.

The coefficient of Correlation

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$$

The coefficient of Determination

$$r^2 = \frac{SS_{yy} - SSE}{SS_{yy}} = 1 - \frac{SSE}{SS_{yy}}$$

Confidence Interval for the mean Value of y at $x = x_p$

$$\widehat{y} \pm t_{\alpha/2} \cdot s \cdot \sqrt{\frac{1}{n} + \frac{(x_p - \overline{x})^2}{SS_{xx}}},$$

where $t_{\alpha/2}$ is based on $(n-2)$ degrees of freedom.

Prediction Interval for an Individual New Value of y at $x = x_p$

$$\widehat{y} \pm t_{\alpha/2} \cdot s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \overline{x})^2}{SS_{xx}}},$$

where $t_{\alpha/2}$ is based on $(n-2)$ degrees of freedom.