Regression

$y_i = \beta_o + \beta_1 x_i + \varepsilon_i$ $\qquad E(y) = \mu_{y|x}$

Slope: $\hat{\beta}_1 = \frac{SS_{XY}}{SS_{XX}}$ $\qquad$ **y-intercept:** $\hat{\beta}_o = \bar{y} - \hat{\beta}_1 \bar{x}$

Line: $\hat{y}_i = \hat{\beta}_o + \hat{\beta}_1 x_i = \bar{y} + \hat{\beta}_1 (x_i - \bar{x}) = \hat{\mu}_y$

$$SS_{XY} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}$$

$$SS_{XX} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

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

$SSE = \Sigma(y_i - \hat{y}_i)^2 = \Sigma e_i^2 = SS_{YY} - \frac{SS_{XY}^2}{SS_{XX}} = SS_{YY} - \hat{\beta}_1 SS_{XY}$ $\qquad s^2 = \frac{SSE}{n-2}$

$\mu_{\hat{\beta}_1} = \beta_1$ $\qquad \sigma_{\hat{\beta}_1}^2 = \frac{\sigma^2}{SSE}$ $\qquad s_{\hat{\beta}_1} = \frac{s}{\sqrt{SS_{XX}}}$

H_o: $\beta_1 = 0$ $\qquad$ **Test Statistic:** $t = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}} = \frac{\hat{\beta}_1}{\frac{s}{\sqrt{SS_{XX}}}}$

Confidence Interval for β_1: $\hat{\beta}_1 \pm t_{\frac{\alpha}{2}, n-2} \frac{s}{\sqrt{SS_{XX}}}$

$\mu_{\hat{y}_i} = \beta_o + \beta_1 x_i$ $\qquad \sigma_{\hat{y}} = \sigma \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{SS_{XX}}}$ $\qquad s_{\hat{\mu}} = s \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{SS_{XX}}}$

Confidence Interval for $\mu_{y|x=x^*}$: $\hat{y}^* \pm t_{\frac{\alpha}{2}, n-2}\, s_{\hat{\mu}}$

Prediction Interval for $y_{new|x=x^*}$: $\hat{y}^* \pm t_{\frac{\alpha}{2}, n-2}\, s \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{SS_{XX}}}$

Residuals: $e_i = y_i - \hat{y}_i$ $\varepsilon_i = y_i - E(y_i)$

Standardized residuals: $\frac{e_i - \overline{e}}{\sqrt{MSE}} = \frac{e_i}{\sqrt{MSE}}$

CORRELATION

Pearson's Correlation Coefficient

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

$$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n}$$

$$SS_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n}$$

$$SS_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n}$$

Parameter: ρ

Test Statistic: $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$, $df = n-2$

Spearman's Rank Correlation Coefficient

$$r_s = \frac{SS_{uv}}{\sqrt{SS_{uu}\,SS_{vv}}}$$

$$SS_{uv} = \sum u_i v_i - \frac{\sum u_i \sum v_i}{n}$$

$$SS_{uu} = \sum u_i^2 - \frac{\left(\sum u_i\right)^2}{n}$$

$$SS_{vv} = \sum v_i^2 - \frac{\left(\sum v_i\right)^2}{n}$$

u_i = *Rank of the ith observation in sample* 1.

v_i = *Rank of the ith observation in sample* 2.

n = *Number of pairs of observations*.

Shortcut Formula for r_s: $r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$, *where* $d_i = u_i - v_i$