Linear Regression - Cheat Sheet

Jeremy Buss

2021-10-10

$y = \beta_0 + \beta_1 x$

$e_i = y_i - \hat{y}_i = y_i - (b_0 + b_i x_i)$

$SS_{res} = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - b_0 + b_1 x_i)^2$

$\frac{\partial SS_{res}}{\partial \beta_0}\Big|_{b_0,b_1} = -2\sum_{i=1}^n (y_i - b_0 - b_1 x_i)$

$\frac{\partial SS_{res}}{\partial \beta_1}\Big|_{b_0,b_1} = -2\sum_{i=1}^n x_i(y_i - b_0 - b_1 x_i)$

$b_0 = \overline{y} - b_1\overline{x}$

$b_1 = \frac{\sum_{i=1}^n y_i(x_i-\overline{x})}{\sum_{i=1}^n x_i(x_i-\overline{x})} = \frac{\sum_{i=1}^n (x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^n (x_i-\overline{x})(x_i-\overline{x})} = \frac{\sum_{i=1}^n (x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^n (x_i-\overline{x})^2} = \frac{S_{xy}}{S_{xx}}$

$Var(b_1) = \sigma^2 \sum_{i=1}^n c_i^2 = \frac{\sigma^2 \sum_{i=1}^n (x_i-\overline{x})^2}{S_{xx}^2} = \frac{\sigma^2}{S_{xx}}$

$Var(b_0) = \sigma^2\left(\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}\right)$

$E(b_0) = \beta_0; E(b_1) = \beta_1$

$\sum_{i=1}^n (y_i - \hat{y}_i) = \sum_{i=1}^n e_i = 0$

$\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}$

Least Squares Regression line passes through the centroid point $(\overline{x}, \overline{y})$

$\sum_{i=1}^n e_i x_i = 0$

$\sum_{i=1}^n e_i \hat{y}_i = 0$

$\hat{\sigma}^2 = \frac{SS_{Res}}{n-2} = MS_{Res}$

β_1: $b_1 \pm t_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2/S_{xx}}$

β_0: $b_0 \pm t_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2(1/n + \overline{x}^2/S_{xx})}$

$\hat{\sigma^2} = MS_{res} \sim \sigma^2\frac{\chi^2_{n-2}}{n-2}$

$\hat{\sigma^2}$range$(\frac{SS_{Res}}{\chi^2_{\alpha/2,n-2}}, \frac{SS_{Res}}{\chi^2_{1-\alpha/2,n-2}})$

$se(\hat{\beta}_1) = \sqrt{\frac{MS_{Res}}{S_{xx}}}$

$se(\hat{\beta}_0) = \sqrt{MS_{Res}(1/n + \overline{x}^2/S_{xx})}$

Reject H_0 in favor of H_1 if Critical Value Method: $|t_{test}| > t_{\alpha/2,n-2}$ P-Value Method: p-value $= 2P(t_{n-2} > |t_{test}|) < \alpha$

Test statistic: $t_{test} = \frac{b_1 - \beta_1^0}{se(b_1)} \sim t_{n-2}$ under H_0

Test statistic: $t_{test} = \frac{b_0 - \beta_0^0}{se(b_0)} \sim t_{n-2}$ under H_0

(SS_T) : $\sum_{i=1}^{n}(y_i - \overline{y})^2$

(SS_R): $\sum_{i=1}^{n}(\hat{y_i} - \overline{y})^2$ or $\hat{\beta}_1 S_{xy}$

(SS_{Res}): $\sum_{i=1}^{n}(y_i - \hat{y_i})^2$ or $SS_T - \hat{\beta}_1 S_{xy}$

$F_{test} = \frac{SS_R/df_R}{SS_{Res}/df_{Res}} = \frac{SS_R/1}{SS_{Res}/(n-2)} = \frac{MS_R}{MS_{Res}}$

reject H_0 if $F_{test} > F_{\alpha,1,n-2}$ or p-value $= P(F_{1,n-2} > F_{test}) > \alpha$

Table 1: ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_{test}
Regression	SS_R	1	MS_R	MS_R/MS_{Res}
Residual	SS_{Res}	n-2	MS_{Res}	
Total	SS_T	n-1		

$R^2 = \frac{SS_R}{SS_T} = \frac{SS_T - SS_{Res}}{SS_T} = 1 - \frac{SS_{Res}}{SS_T}$

$\hat{\mu_{y|x_0}} \pm t_{\alpha/2,n-2}\hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x_0-\overline{x})^2}{S_{xx}}}$

$\hat{y_0} \pm t_{\alpha/2,n-2}\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0-\overline{x})^2}{S_{xx}}}$

$f(y_i|\beta_0, \beta_1, \sigma^2) = \frac{e^{\frac{-(y_i-\beta_0-\beta_i x_i)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$

$L(\beta_0, \beta_1, \sigma^2|Y_i) = \prod_{i=1}^{n} e^{\frac{-(y_i-\beta_0-\beta_i x_i)^2}{2\sigma^2}} \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n$

$l(\beta_0, \beta_1, \sigma^2|Y_i) = \frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_i x_i)^2 - \frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma^2)$

$\left.\frac{\partial l}{\partial \beta_0}\right|_{\hat{\beta}_0,\hat{\beta}_1,\hat{\sigma}} = \frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$

$\left.\frac{\partial l}{\partial \beta_1}\right|_{\hat{\beta}_0,\hat{\beta}_1,\hat{\sigma}} = \frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i = 0$

$\left.\frac{\partial l}{\partial \sigma^2}\right|_{\hat{\beta}_0,\hat{\beta}_1,\hat{\sigma}} = -\frac{1}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 0$

$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x} = b_0$

$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = b_1$

$\hat{\sigma^2} = \frac{\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n}$