

Linear Regression - Cheat Sheet

Jeremy Buss

2021-10-10

$$y = \beta_0 + \beta_1 x$$

$$e_i = y_i - \hat{y}_i = y_i - (b_0 + b_1 x_i)$$

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

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

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

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

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

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

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

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

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

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

LRM contains centroid (\bar{x}, \bar{y})

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

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

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

$$\beta_1: b_1 \pm t_{\alpha/2, n-2} \sqrt{\hat{\sigma}^2 / S_{xx}}$$

$$\beta_0: b_0 \pm t_{\alpha/2, n-2} \sqrt{\hat{\sigma}^2 (1/n + \bar{x}^2 / S_{xx})}$$

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

$$\hat{\sigma}^2 \text{range} \left(\frac{SS_{Res}}{\chi_{\alpha/2, n-2}^2}, \frac{SS_{Res}}{\chi_{1-\alpha/2, n-2}^2} \right)$$

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

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

Reject H_0 in favor of H_1

if Critical Value Method: $|t_{test}| > t_{\alpha/2, n-2}$

P-Value Method: $p\text{-value} = 2P(t_{n-2} > |t_{test}|) < \alpha$

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

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

$(SS_R) : \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ 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 $p\text{-value} = P(F_{1,n-2} > F_{test}) > \alpha$

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

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

$\hat{y}_0 \pm t_{\alpha/2, n-2} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$

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

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

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

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

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

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

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

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

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

$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \epsilon_i$

$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \epsilon_i, \quad i = 1, 2, \dots, n$

$S(\beta_0, \beta_1, \dots, \beta_k) = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2$

$\left. \frac{\partial S}{\partial \beta_0} \right|_{b_0, b_1, \dots, b_k} = -2 \sum_{i=1}^n (y_i - b_0 - \sum_{j=1}^k b_j x_{ij}) = 0$

$\left. \frac{\partial S}{\partial \beta_j} \right|_{b_0, b_1, \dots, b_k} = -2 \sum_{i=1}^n \left(y_i - b_0 - \sum_{j=1}^k b_j x_{ij} \right) x_{ij} = 0$

$$\mathbf{y} = \mathbf{X}\beta + \epsilon, \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

\mathbf{X} : Design matrix $\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$

$$S(\beta) = \mathbf{y}^\top \mathbf{y} - 2\beta^\top \mathbf{X}^\top \mathbf{y} + \beta^\top \mathbf{X}^\top \mathbf{X} \beta$$

$$\frac{\partial \mathbf{t}^\top \mathbf{a}}{\partial \mathbf{t}} = \frac{\partial \mathbf{a}^\top \mathbf{t}}{\partial \mathbf{t}} = \mathbf{a}$$

$$\frac{\partial \mathbf{t}^\top \mathbf{A} \mathbf{t}}{\partial \mathbf{t}} = 2\mathbf{A} \mathbf{t}$$

$$\left. \frac{\partial S}{\partial \beta} \right|_{\mathbf{b}} = -2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X} \mathbf{b} = 0$$

$$\mathbf{b} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

$$\hat{\mathbf{y}} = \mathbf{X} \mathbf{b} = \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} = \mathbf{H} \mathbf{y}$$

$$\text{hat matrix } \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$$

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X} \mathbf{b} = \mathbf{y} - \mathbf{H} \mathbf{y} = (\mathbf{I} - \mathbf{H}) \mathbf{y}$$

$\mathbf{H}, \mathbf{I} - \mathbf{H}$ symmetric idempotent projection matrices

\mathbf{H} projects \mathbf{y} to $\hat{\mathbf{y}}$ on column space \mathbf{X} , $Col(\mathbf{X})$

$\mathbf{I} - \mathbf{H}$ projects \mathbf{y} to \mathbf{e} on space **perpendicular** to $Col(\mathbf{x})$

$$Col(\mathbf{X}) = \{\mathbf{X} \mathbf{b} : \mathbf{b} \in \mathbf{R}^p\}$$

$$\mathbf{y} \notin Col(\mathbf{X})$$

$$\hat{\mathbf{y}} = \mathbf{X} \mathbf{b} = \mathbf{H} \mathbf{y} \in Col(\mathbf{X})$$

Minimize distance of \mathbf{A} to $Col(\mathbf{X})$: Find the point in $Col(\mathbf{X})$ that is closest to \mathbf{A}

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X} \mathbf{b} = \mathbf{y} - \mathbf{H} \mathbf{y} = (\mathbf{I} - \mathbf{H}) \mathbf{y} \perp Col(\mathbf{X})$$

$$\mathbf{X}^\top (\mathbf{y} - \mathbf{X} \mathbf{b}) = 0$$

$$Var(\mathbf{b}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$$

$$SS_{Res} = \mathbf{y}^\top \mathbf{y} - \mathbf{b}^\top \mathbf{X}^\top \mathbf{y}$$

$$MS_{Res} = \frac{SS_{Res}}{n-p} \text{ with } p = k + 1$$

$$\hat{\sigma}^2 = MS_{Res} \text{ is an unbiased estimator for } \sigma^2, \text{ i.e. } E[MS_{Res}] = \sigma^2$$

$\hat{\sigma}^2$ measures **unexplained** var. prefer small residual mean square.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \beta_j \neq 0 \text{ for at least one } j$$

Table 1: ANOVA Table

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

- Note: k is # of coefficients for regressors. For SLR $k = 1$

$$SS_T = \mathbf{y}^\top \mathbf{y} - \frac{1}{n} \sum_{i=1}^n y_i^2$$

$$SS_{Res} = \mathbf{y}^\top \mathbf{y} - \mathbf{b}^\top \mathbf{X}^\top \mathbf{y}$$

$$SS_R = SS_T - SS_{Res} = \mathbf{b}^\top \mathbf{X}^\top \mathbf{y} - \frac{1}{n} \sum_{i=1}^n y_i^2$$

Reject H_0 if $F_{test} > F_{\alpha, k, n-k-1}$

$$E[MS_{Res}] = \sigma^2$$

$$E[MS_R] = \sigma^2 + \frac{\beta_{1:k}^\top \mathbf{X}_c^\top \mathbf{X}_c \beta_{1:k}}{k\sigma^2} \text{ where } \beta_{1:k} = (\beta_1, \dots, \beta_k)^\top$$

$$\mathbf{X}_c = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \dots & x_{1k} - \bar{x}_k \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \dots & x_{2k} - \bar{x}_k \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \dots & x_{nk} - \bar{x}_k \end{bmatrix}$$

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

$$R_{adj}^2 = 1 - \frac{SS_{Res}/(n-p)}{SS_T/(n-1)}$$

Penalty (through p) for the number of variables in model

Partial: Tests the contribution of X_j given all other regressors in the model

$$H_0 : \beta_j = 0 \text{ and } H_1 : \beta_j \neq 0$$

$$t_{test} = \frac{b_j}{\sqrt{\hat{\sigma}^2 C_{jj}}}, \text{ where } C_{jj} \text{ is the } j\text{-th diagonal element of } (\mathbf{X}^\top \mathbf{X})^{-1}$$

$$E(a + cY) = a + cE(Y)$$

$$Var(a + cY) = c^2 Var(Y)$$

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))]$$

$$Cov(a + cX, b + dY) = cd Cov(X, Y)$$

$$E(\sum_{i=1}^n a_i Y_i) = \sum_{i=1}^n a_i E(Y_i)$$

$$Var(\sum_{i=1}^n a_i Y_i) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j Cov(Y_i, Y_j)$$

$$\text{If } Y_1, Y_2, \dots \text{ inde. } Cov(Y_i, Y_j) = 0 \text{ for } i \neq j \text{ and } Var(\sum_{i=1}^n a_i Y_i) = \sum_{i=1}^n a_i^2 Var(Y_i)$$

$$Y \sim N(\mu, \sigma^2), Z = \frac{Y - \mu}{\sigma} \sim N(0, 1)$$

$$Y_i \stackrel{iid}{\sim} N(\mu_i, \sigma_i^2), \text{ then distr. } N(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2)$$

$$\text{If } Z \sim N(0, 1) \text{ then } Z^2 \sim \chi_1^2$$

$$\text{If } Z_i \stackrel{iid}{\sim} N(0, 1) \text{ then } \sum_{i=1}^n Z_i^2 \sim \chi_n^2$$

rank = # of Linearly Independent columns

Idempotent $AA = A$

Orthogonal: $A^{-1} = A^{\top}$ and $A^{\top}A = I$

Symmetric: $A = A^{\top}$

Inverse: $A^{-1}A = AA^{-1} = I$

Quadratic: $y^{\top}Ay = \sum_{i=1}^n \sum_{j=1}^n a_{ij}y_iy_j$

trace: sum of diags

$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$

$E(\mathbf{B}\mathbf{y}) = \mathbf{B}E(\mathbf{y})$

$Var(\mathbf{B}\mathbf{y}) = \mathbf{B}Var(\mathbf{y})\mathbf{B}^{\top} = \mathbf{B}\Sigma\mathbf{B}^{\top}$

$\mathbf{y} \sim N_n(\mu, \sigma^2\mathbf{I})$ and $\bar{y} = \frac{1}{n}\mathbf{1}^{\top}\mathbf{Y}$

$\bar{y} \sim N(\frac{1}{n}\mathbf{1}^{\top}\mathbf{Y}, \frac{1}{n}\mathbf{1}^{\top}(\sigma^2\mathbf{I})\frac{1}{n}\mathbf{1}) = N(\mu, \frac{\sigma^2}{n})$