Linear Regression - Cheat Sheet

Jeremy Buss

2021-10-10

$y = \beta_0 + \beta_1 x$

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

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

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

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

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

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

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

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

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

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

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

LRM contains centroid $(\overline{x}, \overline{y})$

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

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

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

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 $= P(F_{1,n-2} > F_{test}) > \alpha$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$\frac{\partial S}{\partial \beta}\Big|_{\mathbf{b}} = -2\mathbf{X}^\intercal\mathbf{y} + 2\mathbf{X}^\intercal\mathbf{X}\mathbf{b} = 0$

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

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

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

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

$\mathbf{H}$, $\mathbf{I} - \mathbf{H}$ **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 A to $Col(\mathbf{X})$: Find the point in $Col(\mathbf{X})$ that is closest to A

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

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

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

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

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

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

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

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

$H_1 : \beta_j \neq 0$ 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*

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

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

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

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

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

$E[MS_R] = \sigma^2 + \frac{\beta_{1:k}^\intercal\mathbf{X_c}^\intercal\mathbf{X_c}\boldsymbol{\beta_{1:k}}}{k\sigma^2}$ where $\beta_{1:k} = (\beta_1, \ldots, \beta_k)^\intercal$

$$\mathbf{X_c} = \begin{bmatrix} x_{11}-\overline{x_1} & x_{12}-\overline{x_2} & \ldots & x_{1k}-\overline{x_k} \\ x_{21}-\overline{x_1} & x_{22}-\overline{x_2} & \ldots & x_{2k}-\overline{x_k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1}-\overline{x_1} & x_{n2}-\overline{x_2} & \ldots & x_{nk}-\overline{x_k} \end{bmatrix}$$

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

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

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

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

$H_0 : \beta_j = 0$ and $H_1 : \beta_j \neq 0$

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

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

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

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

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

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

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

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

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

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

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

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

rank = # of Linearly Independent columns

Idempotent $AA = A$

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

Symmetric: $A = A^\intercal$

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

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

trace: sum of diags

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

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

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

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

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