

Least Squares Regression

$x_i \quad i=1, \dots, p-1$ predictors

$$X = [1 \quad x_1, \dots, x_{p-1}] \quad n \times p$$

$$\text{rank}(X) = p \leq n$$

$$\tilde{Y} = X\beta + \tilde{\epsilon}, \quad \tilde{\epsilon} \sim N_n[0, \sigma^2 I_n]$$

LS: minimizing $(\tilde{Y} - X\beta)'(\tilde{Y} - X\beta)$

$$\hat{\beta} = (X'X)^{-1}X'\tilde{Y}$$

$$\hat{\beta} \sim N_p\{\beta, \sigma^2(X'X)^{-1}\}$$

$$\text{var}(\hat{\beta}) = \sigma^2 (X'X)^{-1} \quad \begin{matrix} \text{Gauss-Markov} \\ \text{BLUE} \end{matrix}$$

$$\hat{y} = \hat{Y} = X\hat{\beta} = X(X'X)^{-1}X'\tilde{Y} \\ = HY$$

H "hat-matrix"

$$\hat{Y} \sim N_n[X\beta, \sigma^2 X(X'X)^{-1}X']$$

$$\hat{Y} = X\hat{\beta} = \underbrace{\frac{1}{n}\beta_0 + x_1\beta_1 + \dots + x_p\beta_p}_{H Y}$$

H projects \tilde{Y} into the cos. space of X

ANOVA Decomposition

$$\mathbf{I}_n = \bar{\mathbf{J}}_n + [\mathbf{H} - \bar{\mathbf{J}}_n] + [\mathbf{I}_n - \mathbf{H}] \\ = \mathbf{A}_1 + \mathbf{A}_2 + \mathbf{A}_3$$

$\mathbf{X} = [1, \mathbf{X}_c]$ where cols. of \mathbf{X}_c are centered, $\mathbf{1}' \mathbf{X}_c = 0$

$$\mathbf{H} = \bar{\mathbf{J}}_n + \mathbf{X}_c (\mathbf{X}_c' \mathbf{X}_c)^{-1} \mathbf{X}_c'$$

$$\text{So } \mathbf{A}_2 = \mathbf{X}_c (\mathbf{X}_c' \mathbf{X}_c)^{-1} \mathbf{X}_c'$$

$$\text{And } \mathbf{A}_i \subset \tilde{\mathbf{A}}_i \quad i=1, 2, 3$$

$$\mathbf{A}_i \mathbf{A}_j = 0 \quad i \neq j$$

$$\text{rank}(\mathbf{A}_1) = 1, \text{ rank}(\mathbf{A}_2) = p-1$$

$$\text{rank}(\mathbf{A}_3) = n-p$$

Source	df	SS
Overall mean	1	$\mathbf{Y}' \bar{\mathbf{J}} \mathbf{Y}$
Regression	p-1	$\mathbf{Y}' \mathbf{X}_c (\mathbf{X}_c' \mathbf{X}_c)^{-1} \mathbf{X}_c \mathbf{Y}$
Residual	n-p	$\mathbf{Y}' (\mathbf{I} - \mathbf{H}) \mathbf{Y}$
Total	n	$\mathbf{Y}' \mathbf{Y}$
Corrected total	n-1	$\mathbf{Y}' (\mathbf{I} - \bar{\mathbf{J}}) \mathbf{Y}$

Expected Mean Squares $\Sigma = \sigma^2 I$

Recall $E(Y' \Sigma Y) = \text{tr}(\Sigma A) + \mu' A \mu$

- $E(Y' \bar{J} Y) = \text{tr}(\sigma^2 \bar{J}) + \beta' X' \bar{J} X \beta$
 $= \sigma^2 + n \beta_0^2$

because $\bar{J} X = [1, 0]$

- $E(Y' X_c (X_c' X_c)^{-1} X_c' Y)$
 $= (p-1) \sigma^2 + \beta' X' [X_c (X_c' X_c)^{-1} X_c'] X \beta$
 $= (p-1) \sigma^2 + \beta_c' X_c' X_c \beta_c$

where $\beta_c = (\beta_1, \dots, \beta_{p-1})$

$$E(\text{Regression MS}) = \sigma^2 + \frac{\beta_c' X_c' X_c \beta_c}{p-1}$$

- $E(\text{Residual MS}) = \sigma^2$

because $\beta' X' (I - H) X \beta$

Note that $H X = X$ b/c $X(X' X)^{-1} X' X = X$

Assuming normality

$$A_i \Sigma = \sigma^2 A_i \xrightarrow{\text{Bhat's}} c_i = \sigma^2$$

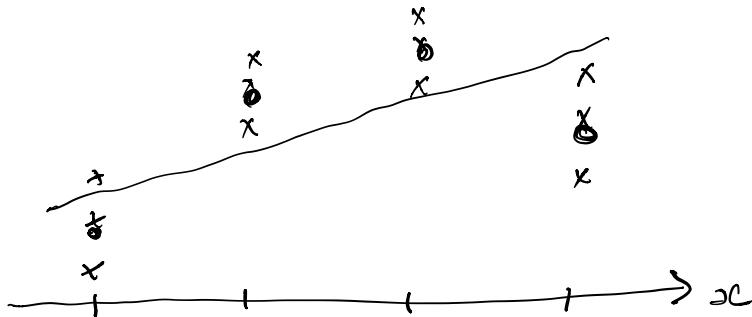
$$Y' \bar{J} Y \sim \sigma^2 \chi^2_{n-p}(\delta_1) \quad \delta_1 = n \beta_0^2 / \sigma^2$$

$$Y' A_2 Y \sim \sigma^2 \chi_{p-1}^2(\delta_2) , \quad \delta_2 = \beta_c' X_c' X_c \beta_c / \sigma^2$$

$$Y' A_3 Y \sim \sigma^2 \chi_{n-p}^2(0)$$

$$\frac{Y' A_2 Y / (p-1)}{Y' A_3 Y / (n-p)} \sim F_{p-1, n-p} (\delta_2)$$

Lack of fit



$$\tilde{Y} = \mu + \varepsilon = X\beta + (\mu - X\beta) + \varepsilon$$

$$\text{Suppose } n = \sum_{j=1}^k r_j$$

r_j is the number of responses at
jth combination of predictors

$$X = \left\{ \begin{array}{l} 1_{r_1} \otimes \tilde{x}_1' \\ 1_{r_2} \otimes \tilde{x}_2' \\ \vdots \\ 1_{r_k} \otimes \tilde{x}_k' \end{array} \right\}$$

$$X' = [I_{r_1}' \otimes \tilde{x}_1, \dots, I_{r_k}' \otimes \tilde{x}_k]$$

$$A_{pe} = \text{blockdiag} \left[I_{r_j} - \bar{J}_{r_j} \right]_{C_{r_j}}$$

$$X' A_{pe} = [(I_{r_1}' \otimes \tilde{x}_1) C_{r_1}, \dots, (I_{r_k}' \otimes \tilde{x}_k) C_{r_k}] \\ = 0$$

$$\text{Similarly } \bar{J}_n A_{pe} = 0, \quad X_c' A_{pe} = 0$$

Hence

$$I_n = \bar{J}_n + X_c (X_c' X_c)^{-1} X_c' \\ + [I - H - A_{pe}] + A_{pe} \\ = A_1 + A_2 + A_{lof} + A_{pe}$$

$$\text{rank}(A_{pe}) = \sum_{j=1}^k (r_j - 1) = n - k$$

$$\text{rank}(A_{lof}) = n - p - (n - k) = k - p$$

Note that $p \leq k$ because $\text{rank}(X) \leq k$
 X has k distinct rows

$$S_{lof} = \mu' A_{lof} \mu / \sigma^2 \\ = \mu' [I - H] \mu / \sigma^2 - \cancel{\mu' A_{pe} \mu / \sigma^2} \rightarrow 0$$

and so $S_{pe} = 0$

Test for lack of fit using

$$\frac{Y' A_{\text{lof}} Y / (k-p)}{Y' A_{\text{ref}} Y / (n-k)} \sim F_{k-p, n-k} (\delta_{\text{lof}})$$

$$\delta_{\text{lof}} = 0 \text{ if } \mu = X\beta.$$

Partitioning to Regression SS

Suppose X is partitioned into groups of columns

$$X = (X_1; X_2; \dots; X_m) \quad \begin{matrix} n \times p \\ n \times p_1 \\ n \times p_2 \\ \vdots \\ n \times p_m \end{matrix} \quad \sum_{j=1}^m p_j = p$$

$$X_1 = I_n, \quad p_1 = 1$$

$$R_i = X_i, \quad R_j = (X_1, \dots, X_j) \quad j=1, \dots, m$$

Consider this identity

$$X(X'X)^{-1}X' = R_1(R_1'R_1)^{-1}R_1' + \sum_{j=2}^m [R_j(R_j'R_j)^{-1}R_j' - R_{j-1}(R_{j-1}'R_{j-1})^{-1}R_{j-1}']$$

Hence the Regression SS decomposes as

$$Y' X (X'X)^{-1} X' Y = Y' \bar{Y} + \text{SS}(X_2 | X_1)$$

$$+ \\ \vdots \\ + SS(x_m | x_1, \dots, x_{m-1})$$

where

$$SS(x_j | x_1, \dots, x_{j-1}) \\ = Y' [R_j (R_j' R_j)^{-1} R_j' - R_{j-1} (R_{j-1}' R_{j-1})^{-1} R_{j-1}'] Y$$

Comments

$$R_j (R_j' R_j)^{-1} R_j' R_j = R_j$$

equivalently

$$\left\{ \begin{array}{l} x'_1 \\ x'_2 \\ \vdots \\ x'_j \end{array} \right\} R_j (R_j' R_j)^{-1} R_j' = \left\{ \begin{array}{l} x'_1 \\ x'_2 \\ \vdots \\ x'_j \end{array} \right\}$$

Hence for $1 \leq i \leq j \leq m$

$$(a) x'_i R_j (R_j' R_j)^{-1} R_j = x'_i$$

$$(b) R'_i R_j (R_j' R_j)^{-1} R_j = R'_i$$

So

$$R_j (R_j' R_j)^{-1} R_j' = (R_{j-1}, x_j) [- J^{-1} (R_{j-1}, x_j)']$$

$$= (R_{j-1} X_j) \begin{Bmatrix} R_j' R_{j-1} & R_j' X_j \\ X_j' R_{j-1} & X_j' X_j \end{Bmatrix}^{-1} \begin{pmatrix} R_j' \\ X_j \end{pmatrix}$$

If $X_i' X_j = 0$ for $i \neq j$ then

$$R_j (R_j' R_j)^{-1} R_j' = R_{j-1} (R_{j-1}' R_{j-1})^{-1} R_{j-1} + X_j (X_j' X_j)^{-1} X_j'$$

$$\text{and so } SS(X_j | X_{1, \dots, j-1}) = Y' X_j (X_j' X_j)^{-1} X_j' Y$$

In this case the sequential decomposition does not depend on the ordering of the groups.