Multiple linear regression model

Y dependent (responce) var. $\quad$ k indep. var. $X_1, \dots, X_k$ (predictor, explantory or regressor var.)

(i) X_i's are non random (fixed var.)

(ii) For each set of X_i values there is a subpopl^n of Y

$$Y_1 \equiv Y \mid X_1 = x_{11}, \dots, X_k = x_{1k} \quad \mu_{Y|x_{11}, \dots, x_{1k}} \quad \sigma^2$$

$$Y_2 \equiv Y \mid X_1 = x_{21}, \dots, X_k = x_{2k} \quad \mu_{Y|x_{21}, \dots, x_{2k}} \quad \sigma^2$$

$$\vdots \qquad\qquad \vdots$$

$$Y_n \equiv Y \mid X_1 = x_{n1}, \dots, X_k = x_{nk} \quad \mu_{Y|x_{n1}, \dots, x_{nk}} \quad \sigma^2$$

$$Y_i = \mu_{Y|x_{i1}, \dots, x_{ik}} + \epsilon_i$$

$$i = 1, \dots, n$$

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \epsilon_i \quad , i = 1, \dots, n$$

$$E(\epsilon_i \epsilon_j) = 0 \quad , i \neq j \quad E(\epsilon_i) = 0, V(\epsilon_i) = \sigma^2$$

$$\underset{\sim}{Y} = \underset{\sim}{X} \underset{\sim}{\beta} + \underset{\sim}{\epsilon} \qquad\qquad p = k+1$$

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}_{n \times 1} ; \quad \underset{\sim}{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & & \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}_{n \times p} ; \quad \underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}_{p \times 1} ; \quad \underline{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}_{n \times 1}$$

$$= \left(\underset{\sim}{1}, \underset{\sim}{x_1}, \dots, \underset{\sim}{x_k} \right)_{1 \times p}$$

minimize $L(\beta) = \sum_{i=1}^{n} \epsilon_i^2 = \epsilon'\epsilon = (Y - X\beta)(Y - X\beta)'$

$$= Y'Y - 2\beta' X' Y + \beta' X' X \beta$$

$$\left.\frac{\partial L(\beta)}{\partial \beta}\right|_{\hat{\beta}} = 0 \Rightarrow 0 - 2X'Y + 2X'X\hat{\beta} = 0$$

$$\boxed{X'X\hat{\beta} = X'Y}$$ normal equation

l.s. estimator for β is $\hat{\beta} = (X'X)^{-1} X'Y$ provided $(X'X)^{-1}$ exist

if the regressors are linearly independent.

fitted model

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

residuals $e_i = y_i - \hat{y}_i$

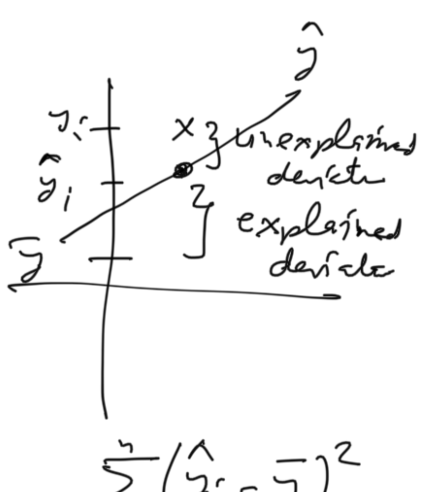
$$e = y - \hat{y} = y - X\hat{\beta} = y - Hy = (I - H)y$$

$$SS_E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = y'y - \hat{\beta}' X' y$$

$$SS_{Total} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = y'y - \frac{\left(\sum_i y_i\right)^2}{n}$$

$$SS_{Total} = SS_{Reg.} + SS_E$$

$$SS_{Reg} = \hat{\beta} X' y - \frac{\left(\sum_i y_i\right)^2}{n}$$



Coeff of multiple determination

$$R^2 = R^2_{y.12\ldots k} = \frac{SS_{Reg}}{\ldots} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

$$SS_{Total} \quad \sum_{i=1}^{n}(y_i - \bar{y})^2$$

$\underset{\sim}{Y}$ MVN with mean $\underset{\sim}{X}\underset{\sim}{\beta}$ and disp Σ

$$l(\beta, \Sigma) = \frac{k}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(Y - X\beta)' \Sigma^{-1} (Y - X\beta)\right\}$$

assumption $\quad \epsilon_i \sim NID(0, \sigma^2)$, $i = 1, \ldots, n$

$$E(\underset{\sim}{Y}) = \underset{\sim}{X}\underset{\sim}{\beta} \;;\; D(Y) = \Sigma = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix} = \sigma^2 I$$

$$E(\hat{\underset{\sim}{\beta}}) = E\left((X'X)^{-1} X' Y\right)$$

$$= (\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}' \underbrace{E(Y)}_{\underset{\sim}{X}\underset{\sim}{\beta}} = \underset{\sim}{\beta}.$$

$\hat{\underset{\sim}{\beta}}$ UE for $\underset{\sim}{\beta}$

$$D(\hat{\underset{\sim}{\beta}}) = \sigma^2 (X'X)^{-1} = \sigma^2 C \quad \text{, where } \underline{C = (X'X)^{-1}}$$

$$\underset{\sim}{Y} \sim N_n(\underset{\sim}{X}\underset{\sim}{\beta}, \sigma^2 I)$$

$$\hat{\underset{\sim}{\beta}} = (X'X)^{-1} X' Y = \underset{\sim}{u}\,\underset{\sim}{Y} \sim N_p(\underset{\sim}{\beta}, \sigma^2 C)$$

$$\frac{\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}}{\sigma\sqrt{C}} \sim N_p(0, I) \qquad \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \sim t_{n-p}$$

<u>For testing significance of regression</u>

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

$$H_1 : \beta_j \neq 0 \text{ for atleast one } j.$$

SoV	df	SS	MS	F
Reg.	$p-1=k$	SS_{Reg}	$MS_{Reg}=\frac{SS_{Reg}}{p-1}$	$\frac{MS_{Reg}}{MS_E}=F_0$
Error	$n-p$	SS_E	$MS_E = SS_E/_{n-p}$	
Total	$n-1$	SS_{Total}		

reject H_0 at los α if $F_0 > F_{\alpha, p-1, n-p}$.

—x—

$$\underset{n\times 1}{\underset{\sim}{y}} = \underset{n\times p}{\underset{\sim}{X}}\,\underset{p\times 1}{\underset{\sim}{\beta}} + \underset{n\times 1}{\underset{\sim}{\epsilon}} \qquad p = k+1$$

$$\underset{\sim}{\beta} = (\beta_1, \beta_2)'$$

$$= \underset{\sim}{X_1}\,\underset{\sim}{\beta_1} + \underset{\sim}{X_2}\,\underset{\sim}{\beta_2} + \epsilon. \qquad n\times 1 \quad (p-r)\times 1$$

test $\boxed{H_0 : \underset{\sim}{\beta_1} = 0 \quad vs \quad H_1 : \underset{\sim}{\beta} \neq 0}$

For full model $\hat{\beta} = (X'X)^{-1}X'y$

$$SS_R(\underset{\sim}{\beta}|\beta_0) = \hat{\beta}X'y - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$$

reduced model $\underset{\sim}{y} = \underset{\sim}{X_2}\,\underset{\sim}{\beta_2} + \epsilon$

$$\hat{\beta}_2 = (X_2'X_2)^{-1}X_2'\underset{\sim}{y}$$

$$\underbrace{SS_R(\beta_2|\beta_0)} = \hat{\beta}_2 X_2' y - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$$

$$\underbrace{SS_R(\beta_1|\beta_2, \beta_0)} = SS_R(\underset{\sim}{\beta}|\beta_0) - SS_R(\beta_2|\beta_0)$$

extra SS due to β_1

$$F_0 = \frac{SS_R(\beta|\beta_2, \beta_0)/r}{MS}$$

MSE

reject H_0 if $F_0 > F_{\alpha, r, n-p}$.

Polynomial regression

$$y = \beta_0 + \beta_1 x + \beta_{11} x^2 + \epsilon$$

$$H_0 : \beta_{11} = 0 \text{ vs } H_1 : \beta_{11} \neq 0$$

Rejecting H_0 means the quadratic term contributes significantly to the model.

ANOVA

SOV	SS	df
Regression	$SS_R(\beta_1, \beta_{11} \mid \beta_0)$	2
Linear	$SS_R(\beta_1 \mid \beta_0)$	1
Quadratic	$SS_R(\beta_{11} \mid \beta_0, \beta_1)$	1
Error	SS_E	$n-3$
Total	SS_{Tot}	$n-1$