$$\underset{\sim}{Y} = X \underset{\sim}{\beta} + \underset{\sim}{\epsilon}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ \vdots \\ Y_2 \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$Y_i = \alpha + \beta X_i + \epsilon_i \qquad i = 1, 2, \ldots, n$$

$$RSS = \underset{1\times n}{\underset{\sim}{\epsilon}^T} \underset{n\times 1}{\underset{\sim}{\epsilon}} = \sum \epsilon_i^2 = \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2$$

$$= \underset{\sim}{Y}^T \underset{\sim}{Y} - 2 \underset{1\times 2}{\underset{\sim}{\beta}^T} \underset{2\times n}{X^T} \underset{n\times 1}{\underset{\sim}{Y}} + \underset{1\times 2}{\underset{\sim}{\beta}^T} \underset{2\times n}{X^T} \underset{n\times 2}{X} \underset{2\times 1}{\underset{\sim}{\beta}}$$

minimise RSS by choice of α, β

i.e. by choice of vector $\underset{\sim}{\beta}$

Solve
$$\frac{\partial RSS}{\partial \underset{\sim}{\beta}} = \begin{pmatrix} \dfrac{\partial RSS}{\partial \alpha} \\ \dfrac{\partial RSS}{\partial \beta} \end{pmatrix} = \underset{\sim}{0}$$

$$\underset{\sim}{b}^T \underset{\sim}{a} = \begin{pmatrix} b_1 & b_2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = a_1 b_1 + a_2 b_2$$

$$\frac{\partial \underset{\sim}{b}^T \underset{\sim}{a}}{\partial \underset{\sim}{b}} = \begin{pmatrix} \partial/\partial b_1 \\ \partial/\partial b_2 \end{pmatrix} \underset{\sim}{b}^T \underset{\sim}{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \underset{\sim}{a}$$

$$\overset{1\times 2}{\begin{pmatrix} b_1 & b_2 \end{pmatrix}} \overset{2\times 2}{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}} \overset{2\times 1}{\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}} = \underset{\sim}{b}^T A \underset{\sim}{b}$$

$$= \overset{1\times 2}{\begin{pmatrix} b_1 a_{11} + b_2 a_{21} & b_1 a_{12} + b_2 a_{22} \end{pmatrix}} \overset{2\times 1}{\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}}$$

$$= \overset{1\times 1}{b_1^2 a_{11} + b_1 b_2 a_{21} + b_1 b_2 a_{12} + b_2^2 a_{22}}$$

$$\frac{\partial}{\partial \underset{\sim}{b}} \underset{\sim}{b}^T A \underset{\sim}{b} = \begin{pmatrix} \partial/\partial b_1 \\ \partial/\partial b_2 \end{pmatrix} \left(b_1^2 a_{11} + \underbrace{b_1 b_2 a_{21} + b_1 b_2 a_{12}}_{b_1 b_2 (a_{21} + a_{12})} + b_2^2 a_{22} \right)$$

$$= \begin{pmatrix} 2b_1 a_{11} + b_2 (a_{21} + a_{12}) \\ b_1 (a_{21} + a_{12}) + 2b_2 a_{22} \end{pmatrix}$$

i.e. A symmetric

NB: If $a_{12} = a_{21}$ i.e. $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} = A'$

$$\Rightarrow \frac{\partial}{\partial \underset{\sim}{b}} \underset{\sim}{b}^T A \underset{\sim}{b} = \begin{pmatrix} 2b_1 a_{11} + 2b_2 a_{12} \\ 2b_1 a_{12} + 2b_2 a_{22} \end{pmatrix} = 2 \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

quadratic form

$$= 2A\underset{\sim}{b}$$

$$RSS = \underset{\sim}{y}^T \underset{\sim}{y} - 2\underset{\sim}{\beta}^T \overbrace{X^T \underset{\sim}{y}}^{\underset{\sim}{b}^T \underset{\sim}{a}} + \underset{\sim}{\beta}^T X^T X \underset{\sim}{\beta}$$

NB: $X^T X$ is symmetric, since

$$(X^T X)^T = X^T (X^T)^T = X^T X$$

$$(AB)^T = B^T A^T$$

$$\Rightarrow \frac{\partial RSS}{\partial \underset{\sim}{\beta}} = \underset{\sim}{0} - 2X^T \underset{\sim}{y} + 2(X^T X)\underset{\sim}{\beta}\Big|_{\underset{\sim}{\beta} = \hat{\underset{\sim}{\beta}}} = \underset{\sim}{0}$$

$$\Rightarrow \quad -2X^T \underset{\sim}{y} + 2(X^T X)\hat{\underset{\sim}{\beta}} = \underset{\sim}{0}$$

$$\Rightarrow \quad X^T X \hat{\underset{\sim}{\beta}} = X^T \underset{\sim}{y} \qquad \text{(normal equations)}$$

$$\Rightarrow \quad \underbrace{(X^T X)^{-1} X^T X \hat{\underset{\sim}{\beta}}}_{I\, \hat{\underset{\sim}{\beta}}} = (X^T X)^{-1} X^T \underset{\sim}{y}$$

$$\Rightarrow \quad \hat{\underset{\sim}{\beta}} = (X^T X)^{-1} X^T \underset{\sim}{y}.$$

is OLS estimator of $\underset{\sim}{\beta}$.

How to evaluate?

$$X^T X = \overset{2\times n}{\begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{pmatrix}} \overset{n\times 2}{\begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}}$$

$$= \overset{2\times 2}{\begin{pmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{pmatrix}}$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$$\cancel{\Rightarrow (X^T X)^{-1} =}$$

$$\frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\Rightarrow (X^T X)^{-1} = \frac{1}{\underset{a}{n} \sum \underset{d}{X_i^2} - \underset{bc}{(\sum X_i)^2}} \begin{pmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{pmatrix}$$

$$= \frac{1}{n\sum (X_i - \bar{X})^2} \begin{pmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{pmatrix}$$

$$= \frac{1}{\sum (X_i - \bar{X})^2} \begin{pmatrix} \frac{1}{n}\sum X_i^2 & -\bar{X} \\ -\bar{X} & 1 \end{pmatrix}$$

$$\underset{\sim}{X^T y} = \overset{2\times n}{\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix}} \overset{n\times 1}{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}$$

$$= \overset{2\times 1}{\begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}}$$

$$\Rightarrow \hat{\beta} = (X^T X)^{-1} X^T \underset{\sim}{y}$$

$$= \frac{1}{n \sum (x_i - \bar{x})^2} \overset{2\times 2}{\begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix}} \overset{2\times 1}{\begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}}$$

$$= \frac{1}{n \sum (x_i - \bar{x})^2} \begin{pmatrix} \sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i \\ -\sum x_i \sum y_i + n \sum x_i y_i \end{pmatrix}$$

$$n \sum x_i y_i - \sum x_i \sum y_i$$

$$= n \left(\sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i \right)$$

$$\sum (x_i - \bar{x})(y_i - \bar{y}) = \sum (x_i y_i - \bar{y} x_i - \bar{x} y_i + \bar{x}\bar{y}$$

$$= \sum x_i y_i - \bar{y} \sum x_i - \bar{x} \sum y_i + n\bar{x}$$

$$= \sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i$$

$$\hat{\underset{\sim}{\beta}} = \frac{1}{n\sum(X_i-\bar{X})^2}\begin{pmatrix} \sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i \\ n\sum(X_i-\bar{X})(Y_i-\bar{Y}) \end{pmatrix}$$

$$= \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}$$

$$\hat{\beta} = \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sum(X_i-\bar{X})^2}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n\sum(X_i-\bar{X})^2}$$

$$RSS = \underset{\sim}{\epsilon}^T \underset{\sim}{\epsilon}$$

$$= \underset{\sim}{Y}^T \underset{\sim}{Y} - 2\underset{\sim}{\beta}^T X^T \underset{\sim}{Y} + \underset{\sim}{\beta}^T X^T X \underset{\sim}{\beta}$$

$$\frac{\partial RSS}{\partial \underset{\sim}{\beta}} = \underset{\sim}{0} \quad \Rightarrow \quad X^T X \hat{\underset{\sim}{\beta}} = X^T \underset{\sim}{Y}$$

$$\Rightarrow \quad \hat{\underset{\sim}{\beta}} = (X^T X)^{-1} X^T \underset{\sim}{Y}.$$

Extend to multiple regression.

ie $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \epsilon_i$

$$\Rightarrow \begin{array}{cccc} n\times 1 & n\times(p+1) & (p+1)\times 1 & n\times 1 \\ \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = & \begin{pmatrix} 1 & X_{11} & \cdots & X_{p1} \\ 1 & X_{12} & \cdots & X_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & \cdots & X_{pn} \end{pmatrix} & \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + & \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \end{array}$$

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\epsilon}$$

$$\Rightarrow \underset{\sim}{\epsilon}^T \underset{\sim}{\epsilon} = (\underline{y} - X\underset{\sim}{\beta})^T(\underline{y} - X\underset{\sim}{\beta})$$

$$= \underline{y}^T\underline{y} - 2\underset{\sim}{\beta}^T X^T \underline{y} + \underset{\sim}{\beta}^T X^T X \underset{\sim}{\beta}$$

as before

$$\Rightarrow \frac{\partial RSS}{\partial \underset{\sim}{\beta}} = -2\cancel{\beta^T} X^T \underline{y} + 2X^T X \underset{\sim}{\beta}$$

as before.

$$(b_1\ b_2\ \cdots\ b_p)\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = \sum a_i b_i$$

$$\frac{\partial}{\partial \underset{\sim}{b}}(\underset{\sim}{b}^T \underset{\sim}{a}) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}$$ extends

Also $\frac{\partial}{\partial \underset{\sim}{b}} \underset{\sim}{b}^T A \underset{\sim}{b} = 2A\underset{\sim}{b}$ provided $A = A'$

$$\Rightarrow \quad X^T X \underset{\sim}{\hat{\beta}} = X^T \underset{\sim}{y}$$

as before

$$\Rightarrow \quad \underset{\sim}{\hat{\beta}} = (X^T X)^{-1} \underbrace{X^T}_{(p+1)\times n} \underbrace{\underset{\sim}{y}}_{n\times 1} \quad (p+1)\times 1$$

as before.

$$\underbrace{\overset{(p+1)\times n}{X^T} \overset{n\times(p+1)}{X}}_{(p+1)\times(p+1)} \Rightarrow \underset{(p+1)\times(p+1)}{(X^T X)^{-1}}$$

$$\underset{\sim}{\hat{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} \quad (p+1)\times 1$$

$$\begin{pmatrix} n & \Sigma X_1 & \Sigma X_2 & \cdots & \Sigma X_p \\ \Sigma X_1 & \Sigma X_1^2 & \Sigma X_1 X_2 & & \\ & \Sigma X_1 X_2 & \Sigma X_2^2 & & \\ & & & \ddots & \end{pmatrix}$$

Stochastic properties of $\underset{\sim}{\hat{\beta}}$, $\underset{\sim}{\hat{y}}$ etc

$E(\hat{\beta})$, $\mathrm{cov}(\hat{\beta})$?

$$\underset{\sim}{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$$E(\underset{\sim}{X}) = \begin{pmatrix} E(X_1) \\ E(X_2) \end{pmatrix}$$

$$\underset{\substack{\uparrow \\ \text{matrix}}}{var(\underset{\sim}{X})} = \begin{pmatrix} var(X_1) & cov(X_1, X_2) \\ cov(X_1, X_2) & var(X_2) \end{pmatrix}$$

$$\Rightarrow E\left[\overset{n \times 1}{(\underset{\sim}{X} - E(\underset{\sim}{X}))} \, \overset{1 \times n}{(\underset{\sim}{X} - E(\underset{\sim}{X}))^T} \right]$$