2.1) $y_i = \mu + \varepsilon_i$

$$\hat{\mu} = (X^T X)^{-1} X^T y \quad \text{Least Squares estimate} \qquad \text{where} \quad \underline{X} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$X^T X = n \rightarrow (X^T X)^{-1} = \frac{1}{n}$$

$$\rightarrow (X^T X)^{-1} X^T = \left[\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right]$$

$$\rightarrow (X^T X)^{-1} X^T y = \sum_{i=1}^{n} \frac{y_i}{n} = \bar{y} \quad \blacksquare$$

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

$$\text{where} \quad \underline{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdots & \cdots \\ 1 & x_n \end{bmatrix} \qquad X^T X = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum_{i=1} x_i^2 \end{bmatrix}$$

$$\det(X^T X) = \frac{1}{n \sum x_i^2 - (\sum x_i)^2} = \frac{1}{n \sum x_i^2 - (n\bar{x})^2}$$

$$(X^T X)^{-1} = \frac{1}{n \sum x_i^2 - n^2 \bar{x}^2} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix}$$

$$X^T y = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix} \qquad \therefore \hat{\beta}_1 = \frac{n \sum x_i y_i - \overset{n\bar{y}\bar{x}}{\sum y_i \sum x_i}}{n \sum x_i^2 - n^2 \bar{x}^2} \rightarrow$$

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

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

$Var(\hat{\beta}_1) = \frac{1}{\sum x_i^2 - n\bar{x}} \cdot \sigma^2$ as it is the 2x2 element

of $(X^T X)^{-1} \sigma^2$ see pg 30 in Aslam

$$\frac{1}{\sum x_i^2 - n\bar{x}^2} = \frac{1}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{1}{Var(x_i)}$$

$\therefore$ $Var(\hat{\beta}_1) = \frac{\sigma^2}{Var(x_1)}$ $\therefore$ if $var(x_i)$ is big.

our fit of $\hat{\beta}_1$ will be more precise

2.10

$$\underset{\sim}{y} - P_x \underset{\sim}{y} = (I - P_x) y$$

$$y^T P_x (I - P_x) y = y^T P_x y - y^T P_x P_x y =$$

$$y^T P_x y - y^T P_x y = \underset{\sim}{0}$$

$\therefore$ Py and $y - Py$ are $\perp$

2.23

$$X = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & & & 1 \end{bmatrix} = I$$

$$P_x = X (X^T X)^{-1} X^T = I (I)^{-1} I = I$$

$$\therefore \quad I - P_x = 0$$

$$\therefore \quad \hat{\beta} = \underset{\sim}{y} \qquad \hat{u} = \underset{\sim}{y} \qquad s = 0$$

Model does not allow any standard error, # of parameters = # of data points