Consider iid data $(X_1, Z_1), \ldots, (X_n, Z_n)$, where $X_i \in \mathbb{R}^p$ and $Z_i \in \mathbb{R}$, coming from the linear regression model

$$E(Z_i | X_i = x) = \alpha + \beta^T x$$

where $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^p$ are unknown parameters.

1. The least squares estimator of α and β is found by minimising

$$\sum_{i=1}^{n} (Z_i - \alpha - \beta^T X_i)^2$$

wrt α and β. Show that the least squares estimators of α and β are equal to

$$\hat{\beta} = (\mathcal{X}_C^T \mathcal{X}_C)^{-1} \mathcal{X}_C^T Z_C, \quad \hat{\alpha} = \bar{Z} - \hat{\beta}^T \bar{X}$$

where $\mathcal{X}_C$ is the $n \times p$ matrix of centered data $X_i^C = X_i - \bar{X}$ and $Z_C = (Z_{C,1}, \ldots, Z_{C,n})^T$, with $Z_{C,i} = Z_i - \bar{Z}$.

$$\underset{(p\times 1)}{\bar{X}} = \begin{bmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_p \end{bmatrix}$$

Minimize : $(Z - \alpha \cdot 1_n - X\beta)^T (Z - \alpha \cdot 1_n - X\beta)$

$$= Z^T Z + \alpha^2 1_n^T 1_n + \beta^T X^T X \beta - 2 Z^T 1_n \cdot \alpha - 2\beta^T X^T Z + 2\alpha \beta^T X^T 1_n$$

$$\Rightarrow \frac{\partial}{\partial \alpha} = 0 \Rightarrow 2\alpha \underbrace{1_n^T 1_n}_{= n} - 2\underbrace{Z^T 1_n}_{= \bar{Z}\cdot n} + 2\beta^T \underbrace{X^T 1_n}_{= \bar{X}\cdot n} = 0$$

Scalar derivative $f(x) \rightarrow \frac{\mathrm{d}f}{\mathrm{d}x}$			Vector derivative $f(\mathbf{x}) \rightarrow \frac{\mathrm{d}f}{\mathrm{d}\mathbf{x}}$		
bx	$\rightarrow$	b	$\mathbf{x}^T\mathbf{B}$	$\rightarrow$	$\mathbf{B}$
bx	$\rightarrow$	b	$\mathbf{x}^T\mathbf{b}$	$\rightarrow$	$\mathbf{b}$
x^2	$\rightarrow$	$2x$	$\mathbf{x}^T\mathbf{x}$	$\rightarrow$	$2\mathbf{x}$
bx^2	$\rightarrow$	$2bx$	$\mathbf{x}^T\mathbf{B}\mathbf{x}$	$\rightarrow$	$2\mathbf{B}\mathbf{x}$

$$2\alpha n - 2n\bar{Z} + 2n\beta^T \bar{X} = 0$$

$$2\alpha n = 2n\bar{Z} - 2n\beta^T \bar{X}$$

$$\alpha = \bar{Z} - \beta^T \bar{X}$$

$$\Rightarrow \frac{\partial}{\partial \beta} = 0 \Rightarrow 2X^T X\beta - 2X^T Z + 2\alpha \underbrace{X^T 1_n}_{= \bar{X}\cdot n} = 0$$

(sub α) $\quad X^T X\beta - X^T Z + n(\bar{Z} - \beta^T \bar{X})\bar{X} = 0$

$$X^T X\beta - X^T Z + n\bar{Z}\bar{X} - n\underbrace{\beta^T \bar{X}}_{= \bar{X}^T\beta \text{ (scalar)}}\bar{X} = 0$$

$$X^T X\beta - X^T Z + n\bar{Z}\bar{X} - n\bar{X}(X^T\beta) = 0$$

$$[X^T X\beta - n\bar{X}X^T]\beta - X^T Z + n\bar{Z}\bar{X} = 0$$

Now, we have:

$$X_c = X - 1_n \cdot \bar{X}^T \qquad Z_c = Z - \bar{Z} \cdot 1_n$$

① $$\Rightarrow X_c^T X = (X - 1_n \bar{X}^T)^T (X - 1_n \bar{X}^T)$$

$$= X^T X - 2\bar{X} \underbrace{1_n^T X}_{\bar{X}^T \cdot n} + \bar{X} \underbrace{1_n^T 1_n}_{n} \bar{X}^T$$

$$= X^T X - 2n \bar{X} \bar{X}^T + n \bar{X} \bar{X}^T$$

$$= X^T X - n \bar{X} \bar{X}^T$$

② $$\Rightarrow X_c^T Z_c = (X - 1_n \bar{X}^T)^T (Z - \bar{Z} 1_n)$$

$$= X^T Z - X^T \bar{Z} 1_n - \bar{X} \underbrace{1_n^T Z}_{\bar{Z} \cdot n} + \bar{X} \underbrace{1_n^T \bar{Z} 1_n}_{\bar{Z} \cdot n}$$

$$= X^T Z - \underbrace{X^T 1_n}_{\bar{X} \cdot n} \bar{Z} - \cancel{n \bar{Z} \bar{X}} + \cancel{n \bar{Z} \bar{X}}$$

$$= X^T Z - n \bar{Z} \bar{X}$$

$$\underbrace{[X^T X - n \bar{X} \bar{X}^T]}_{X_c^T X_c} \underset{\sim}{\beta} \underbrace{- X^T Z + n \bar{Z} \bar{X}}_{- X_c^T Z_c} = 0$$

$$\Rightarrow (X_c^T X_c) \underset{\sim}{\beta} = X_c^T Z_c$$

$$\Rightarrow \underset{\sim}{\beta} = (X_c^T X_c)^{-1} X_c^T Z_c$$

2. To predict a new value Z_{new} from a new value X_{new}, we take

$$\hat{Z}_{new} = \hat{\alpha} + \hat{\beta}^T X_{new} = \bar{Z} + \hat{\beta}^T (X_{new} - \bar{X}). \tag{1}$$

Since

$$\hat{\beta} = (\mathcal{X}_C^T \mathcal{X}_C)^{-1} \mathcal{X}_C^T Z_C$$

is also the LS estimator of β in the model for centered data

$$Z_{C,i} = \beta^T X_{C,i} + \epsilon_i,$$

for prediction one can simply estimate β from centered data and then take $\hat{Z}_{new}$ at (1). In other words, we don't even need to estimate α.

$$\begin{aligned}\hat{Z}_{new} &= \hat{\alpha} + \hat{\beta}^T X_{new} \\ &= (\bar{Z} - \hat{\beta}^T \bar{X}) + \hat{\beta}^T X_{new} \\ &= \bar{Z} + \hat{\beta}^T (X_{new} - \bar{X})\end{aligned}$$