Q1: when $p=2$: $y_i = \beta_0 + \beta_1 x_i + 6 z_i$

using linear equations, we obtained:

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i^2 \frac{1}{n}\sum_{i=1}^{n} y_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^2 - \frac{1}{n^2}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i + \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i^2 \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \frac{1}{n}\sum_{i=1}^{n} x_i y_i}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2}$$

Now switch to matrix notation:

$$\begin{vmatrix} y_1 \\ \vdots \\ y_n \end{vmatrix} = \begin{vmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{vmatrix} \cdot \begin{vmatrix} \beta_0 \\ \beta_1 \end{vmatrix} + 6 z_i$$

$$\hat{\beta} = (X^T X)^{-1} X^T \cdot y$$

$$= \left(\begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{vmatrix} \cdot \begin{vmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{vmatrix} \right)^{-1} \cdot \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{vmatrix} \cdot \begin{vmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{vmatrix}$$

$$= \begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix}^{-1} \cdot \begin{vmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{vmatrix}$$

$$= \frac{1}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \cdot \begin{vmatrix} \sum_{i=1}^{n} x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{vmatrix} \begin{vmatrix} n\bar{y} \\ \sum_{i=1}^{n} x_i y_i \end{vmatrix}$$

$$= \begin{vmatrix} \dfrac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} \\ \\ \dfrac{-\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i + n\sum_{i=1}^{n} x_i y_i}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} \end{vmatrix}$$

$$= \begin{vmatrix} \dfrac{\frac{1}{n}\sum_{i=1}^{n} x_i^2 \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \frac{1}{n}\sum_{i=1}^{n} x_i y_i}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2} \\ \\ \dfrac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2} \end{vmatrix}$$

which corresponds exactly to our answer from solving linear equations

Q2:

Using straightforward calculations, we find

$$E(\hat\beta_0) = \beta_0$$
$$E(\hat\beta_1) = \beta_1$$
$$Var(\hat\beta_0) = \frac{\sigma^2}{n}\left(1 + \frac{\bar{x}^2}{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2}\right)$$
$$Var(\hat\beta_1) = \frac{\sigma^2}{n}\frac{1}{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2}$$
$$Cov(\hat\beta_0, \hat\beta_1) = -\frac{\sigma^2}{n}\frac{\bar{x}}{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2}$$

Now lets consider matrix notation

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

$$E(\hat\beta) = \beta \Rightarrow \quad E(\hat\beta_0) = \beta_0, \quad E(\hat\beta_1) = \beta_1$$

$$Var(\hat\beta) = \sigma^2 (X^T X)^{-1} = \sigma^2 \cdot \left(\begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{vmatrix} \cdot \begin{vmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{vmatrix} \right)^{-1}$$

$$= \sigma^2 \cdot \left(\begin{vmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{vmatrix} \right)^{-1}$$

$$= \frac{\sigma^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \cdot \begin{vmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{vmatrix}$$

$$= \frac{\sigma^2}{n} \frac{1}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2} \cdot \begin{vmatrix} \frac{1}{n}\sum_{i=1}^n x_i^2 & -\frac{1}{n}\sum_{i=1}^n x_i \\ -\frac{1}{n}\sum_{i=1}^n x_i & 1 \end{vmatrix}$$

So: $$Var(\hat\beta_0) = \frac{\sigma^2}{n} \cdot \frac{\frac{1}{n}\sum_{i=1}^n x_i^2}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2} = \frac{\sigma^2}{n}\left(1 + \frac{\bar{x}^2}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2}\right)$$

$$Var(\hat\beta_1) = \frac{\sigma^2}{n} \cdot \frac{1}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2}$$

$$Cov(\hat\beta_0, \hat\beta_1) = -\frac{\sigma^2}{n} \cdot \frac{\bar{x}}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2}$$

which corresponds exactly to our answer from direct calculation

Q_3: $y \sim N(X\beta, \sigma^2 I_n)$

$$\sigma^2_c = c\|y - X\hat{\beta}\|^2$$
$$= c\|y - (X^TX)^{-1}X^Ty\|^2$$
$$= c\|(I_n - (X^TX)^{-1}X^T)y\|^2$$
$$= c\|(I_n - (X^TX)^{-1}X^T)(X\beta + \sigma z)\|^2$$
$$= c\|(I_n - (X^TX)^{-1}X^T)\cdot \sigma z\|^2$$
$$= c\sigma^2\|(I_n - (X^TX)^{-1}X^T)\cdot z\|^2 \sim c\cdot\sigma^2 \chi^2_{n-p}$$

where $p = \text{rank}(X^TX)^{-1}X^TX$

① $E(\sigma^2_c) = \sigma^2 \Rightarrow c\cdot\sigma^2(n-p) = \sigma^2$

$$c = \frac{1}{n-p}$$

② $MLE = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\|y - X\beta\|^2\right)$

$$\log MLE = -n\log\sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}\|y - X\beta\|^2$$
$$= -n\log\sqrt{2\pi} - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\|y - X\beta\|^2$$

$$\hat{\beta} = LSE(\beta)$$

$$\frac{\partial \log MLE}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{\|y - X\beta\|^2}{2(\sigma^2)^2} = 0$$

$$\frac{n}{2}\sigma^2 = \frac{\|y - X\beta\|^2}{2}$$

$$\sigma_c^2 = \frac{1}{n}\|y - X\hat{\beta}\|^2$$

$$c = \frac{1}{n}$$

③ $\arg\min E(\hat{\sigma}_c^2 - \sigma^2)^2$

$$= \arg\min E(c\|y - X\hat{\beta}\|^2 - \sigma^2)^2$$
$$= \arg\min E(c\sigma^2\|(I_n - (X^TX)^{-1}X^T)z\| - \sigma^2)^2$$
$$= \arg\min (\sigma^2)^2 E(c\|I_n - (X^TX)^{-1}X^Tz\| - 1)^2$$
$$= \arg\min \left((E(c\|I_n - (X^TX)^{-1}X^Tz\|) - 1)^2 + Var(c\|I_n - (X^TX^{-1})^{-1}X^Tz\| - 1)\right)$$
$$= \arg\min \left((c(n-p) - 1)^2 + 2c^2(n-p)\right)$$
$$= \arg\min \left(c^2(n-p)^2 - 2c(n-p) + 1 + 2c^2(n-p)\right)$$

$$\frac{\partial\, c^2(n-p)^2 - 2c(n-p) + 1 + 2c^2(n-p)}{\partial c} = 2c(n-p)^2 - 2(n-p) + 4c(n-p) = 0$$

$$c = \frac{2(n-p)}{2(n-p)^2 + 4(n-p)} = \frac{1}{n-p+2}$$

(4)

$$\frac{\|y - X\hat{\beta}\|^2}{\sigma^2} \sim \chi^2_{n-p}$$

So $$P\left(\chi^2_{n-p,\frac{\alpha}{2}} < \frac{\|y - X\hat{\beta}\|}{\sigma^2} < \chi^2_{n-p,1-\frac{\alpha}{2}}\right) = \alpha$$

So $$P\left(\frac{\|y - X\hat{\beta}\|}{\chi^2_{n-p,1-\frac{\alpha}{2}}} < \sigma^2 < \frac{\|y - X\hat{\beta}\|}{\chi^2_{n-p,\frac{\alpha}{2}}}\right) = \alpha$$

α level confidence interval:

$$\left(\frac{\|y - X\hat{\beta}\|}{\chi^2_{n-p,1-\frac{\alpha}{2}}}, \frac{\|y - X\hat{\beta}\|}{\chi^2_{n-p,\frac{\alpha}{2}}}\right)$$

$$= \left(\frac{\hat{\sigma}^2_c}{c\chi^2_{n-p,1-\frac{\alpha}{2}}}, \frac{\hat{\sigma}^2_c}{c\chi^2_{n-p,\frac{\alpha}{2}}}\right)$$

Q4:

(a)

$$y \sim N(X\beta, \sigma^2 I_n)$$

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

$(x^*)^T \cdot \hat\beta$ is still <u>normal</u>. as it is a linear combination of $\hat\beta$

$$E((x^*)^T \cdot \hat\beta) = (x^*)^T \cdot E(\hat\beta) = (x^*)^T \cdot \beta$$

$$\begin{aligned} Var((x^*)^T \cdot \hat\beta) &= (x^*)^T \cdot Var(\hat\beta) ((x^*)^T)^T \\ &= (x^*)^T \cdot \sigma^2 (X^TX)^{-1} \cdot x^* \\ &= \sigma^2 (x^*)^T \cdot (X^TX)^{-1} \cdot x^* \end{aligned}$$

$$(x^*)^T \cdot \hat\beta \sim N((x^*)^T \cdot \beta, \sigma^2 (x^*)^T \cdot (X^TX)^{-1} \cdot x^*)$$

(b)

$$(x^*)^T \cdot \hat\beta - (x^*)^T \cdot \beta \sim N(0, \sigma^2 (x^*)^T (X^TX)^{-1} x^*)$$

$$\frac{(x^*)^T \cdot \hat\beta - (x^*)^T \cdot \beta}{\sqrt{\sigma^2 (x^*)^T (X^T \cdot X)^{-1} x^*}} \sim N(0,1)$$

$$\frac{\|y - X\hat\beta\|}{\sigma^2} \sim \chi^2_{n-p}$$

$$\hat\beta \perp\!\!\!\perp \frac{\|y - X\hat\beta\|^2}{\sigma^2}$$

So: $\dfrac{(x^*)^T (\hat\beta - \beta)}{\sigma^2 (x^*)^T (X^T X^{-1}) x^*} \perp\!\!\!\perp \dfrac{\|y - X\hat\beta\|^2}{\sigma^2}$

$$\frac{\dfrac{(x^*)^T (\hat\beta - \beta)}{\sqrt{\sigma^2} \cdot \sqrt{(x^*)^T (X^TX)^{-1} x^*}}}{\sqrt{\frac{1}{n-p}} \cdot \sqrt{\dfrac{\|y - X\hat\beta\|^2}{\sigma^2}}} \sim t_{n-p}$$

$$\frac{(x^*)^T \cdot (\hat\beta - \beta)}{\sqrt{\frac{1}{n-p}} \sqrt{(x^*)^T (X^TX)^{-1} x^*} \cdot \sqrt{\|y - X\hat\beta\|^2}} \sim t_{n-p}$$

α level confidence interval:

$$\left((x^*)^T\hat\beta - \sqrt{\tfrac{1}{n-p} (x^*)^T (X^TX)^{-1} x^* \|y - X\hat\beta\|^2} \cdot t_{n-p, 1-\frac{\alpha}{2}},\ (x^*)^T\hat\beta + \sqrt{\tfrac{1}{n-p} (x^*)^T (X^TX)^{-1} x^* \|y - X\hat\beta\|^2} \cdot t_{n-p, 1-\frac{\alpha}{2}}\right)$$

Q5: We showed in class that

$$Var(\hat{\beta}_1) = \sigma^2 \cdot \frac{1}{n} \cdot \frac{1}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

Here n is fixed, σ^2 is out of our control

Then: $\min Var(\hat{\beta}_1) \iff \max \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$

$\Rightarrow$ choose x_1, x_2 very close to 1

x_3, x_4 very close to -1

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

Q6: (a) $E(\hat{\theta}-\theta)^2 = E(\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2)$

$$= E(\hat{\theta}^2) - 2E(\hat{\theta})\theta + \theta^2$$

$$= (E(\hat{\theta}))^2 - 2E(\hat{\theta})\theta + \theta^2 + E(\hat{\theta}^2) - (E(\hat{\theta}))^2$$

$$= (E(\hat{\theta}) - \theta)^2 + Var(\hat{\theta})$$

(b) $\sum_{i=1}^{n}(x_i-\theta)^2 = \sum_{i=1}^{n}(x_i - \bar{x} + \bar{x} - \theta)^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2 + \sum_{i=1}^{n}(\bar{x}-\theta)^2 + \sum_{i=1}^{n} 2(x_i-\bar{x})(\bar{x}-\theta)$

$$\sum_{i=1}^{n}(x_i-\bar{x})\cdot(\bar{x}-\theta) = (\bar{x}-\theta)\cdot\sum_{i=1}^{n}(x_i-\bar{x}) = (\bar{x}-\theta)\cdot 0 = 0$$

$$= \sum_{i=1}^{n}(x_i-\bar{x})^2 + \sum_{i=1}^{n}(\bar{x}-\theta)^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\theta)^2$$

(c) $\sum_{i=1}^{n}\|x_i-\theta\|^2 = \sum_{i=1}^{n}\|x_i-\bar{x}+\bar{x}-\theta\|^2 = \sum_{i=1}^{n}\|x_i-\bar{x}\|^2 + \sum_{i=1}^{n}\|\bar{x}-\theta\|^2 + 2\sum_{i=1}^{n}(x_i-\bar{x})^T(\bar{x}-\theta)$

$$2\sum_{i=1}^{n}(x_i-\bar{x})^T(\bar{x}-\theta) = 2\left(\sum_{i=1}^{n}(x_i-\bar{x})^T\right)(\bar{x}-\theta) = 2\,(0\,\mathbb{1}_n)^T(\bar{x}-\theta) = 0$$

$$= \sum_{i=1}^{n}\|x_i-\bar{x}\|^2 + \sum_{i=1}^{n}\|\bar{x}-\theta\|^2 = \sum_{i=1}^{n}\|x_i-\bar{x}\|^2 + n\|\bar{x}-\theta\|^2$$

(d) $X \in \mathbb{R}^{n\times p}$, $y \in \mathbb{R}^{n\times 1}$, $\beta \in \mathbb{R}^{p\times 1}$

$$\|y-X\beta\|^2 = \|y - X\hat{\beta} + X\hat{\beta} - X\beta\|^2 = \|y-X\hat{\beta}\|^2 + \|X\hat{\beta}-X\beta\|^2 + 2(y-X\hat{\beta})^T(X\hat{\beta}-X\beta)$$

$$\hat{\beta} = (X^TX)^{-1}X^Ty$$

$$(y-X\hat{\beta})^T(X\hat{\beta}-X) = ((y - X(X^TX)^{-1}X^Ty))^T \cdot X\cdot(\hat{\beta}-\beta) = ((I_n - X(X^TX)^{-1}X^T)y)^T X(\hat{\beta}-\beta)$$

$$= y^T(I_n - X(X^TX)^{-1}X^T)X(\hat{\beta}-\beta) = y^T(X - X(X^TX)^{-1}X^TX)(\hat{\beta}-\beta)$$

$$= y^T(X-X)(\hat{\beta}-\beta) = 0$$

So $\|y-X\beta\|^2 = \|y-X\hat{\beta}\|^2 + \|X\hat{\beta}-X\beta\|^2$

(e) For each, we separate distance from the data to the real value.

into distance between data and prediction using estimated parameters and distance between prediction using real and estimated parameters

The commonality lies in that the estimation is chosen to minimize distance to data within the subspace of possible parameter predictions

It is thus orthogonal to all vectors in the parameter prediction subspace

The underlying parameter's prediction is also in this subspace

Then the cross term is 0.

Thus the distance can be separated as above