Problem 3

(a)

If $\beta=0$, $y_i=\alpha+e_i$. It means that there's no association between y and x.
The regression function will be plotted as a horizontal straight line which is $y_i=\alpha$

(b) $y_i=\alpha+e_i$

$\hat{y}_i=\hat{\alpha}$

$SSE=\sum_{i=1}^n (y_i-\hat{y}_i)^2=\sum_{i=1}^n (y_i-\hat{\alpha})^2=\sum_{i=1}^n y_i^2-2\hat{\alpha}y_i+\hat{\alpha}^2$

$\frac{\partial SSE}{\partial \alpha}=\sum_{i=1}^n -2y_i+2\hat{\alpha}=0$

$\sum_{i=1}^n(-y_i+\hat{\alpha})=0.$

$n\hat{\alpha}=\sum_{i=1}^n y_i$

$\hat{\alpha}=\frac{1}{n}\sum_{i=1}^n y_i$

$\hat{\alpha}=\bar{y}$

(c)

To prove $\hat{\alpha}$ is an unbiased estimator of α, I'll show $E(\hat{\alpha})=\alpha$.

$y_i=\alpha+e_i \quad e_i \sim NID(0,\sigma^2)$

$E(y_i)=E(\alpha+e_i)=\alpha+0=\alpha$

$Var(y_i)=Var(\alpha+e_i)=Var(e_i)=\sigma^2$

So y_i's have $N(\alpha,\sigma^2)$ distribution and are independent.

$y_i \sim NID(\alpha,\sigma^2)$

$E(\hat{\alpha})=E(\frac{1}{n}\sum_{i=1}^n y_i)$ y_i's are independent

$=\frac{1}{n}\sum_{i=1}^n E(y_i)$

$=\frac{1}{n}\cdot n\cdot\alpha$

$\therefore E(\hat{\alpha}) = \alpha$

$\therefore \hat{\alpha}$ is an unbiased estimator of α.

(d)

$\hat{\alpha} = \bar{y}$

$$Var(\hat{\alpha}) = Var(\tfrac{1}{n}\textstyle\sum_{i=1}^{n} y_i)$$

$$= \tfrac{1}{n^2} Var(y_1 + y_2 + \cdots + y_n)$$

$$= \tfrac{1}{n^2}[Var(y_1) + Var(y_2) + \cdots + Var(y_n)]$$

y_i's are independent.

$$= \tfrac{1}{n^2} \cdot n \cdot \sigma^2$$

$$= \frac{\sigma^2}{n}$$

$\therefore Var(\hat{\alpha}) = \frac{\sigma^2}{n}$.

(e)

$\hat{\alpha} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

In question (c), I showed $y_i \sim NID(\alpha, \sigma^2)$, i.e. y_i's are independent and have $N(\alpha, \sigma^2)$ distributions.

$\hat{\alpha}$ is the summation of n $y_i \sim NID(\alpha, \sigma^2)$ divided by n

Since the summation of independent random variables with normal distribution also has normal distribution, $\sum_{i=1}^{n} y_i \sim N(n\alpha, n\sigma^2)$.

The division of n doesn't influence its distribution.

So, $\hat{\alpha}$ has normal distribution.

(Note. even if y_i's don't have normal distribution, by Central Limit theorem, when n is large enough, $\hat{\alpha}$ will still have normal distribution.)

f).

Suppose $\hat{\alpha} = \sum_{i=1}^{n} c_i y_i$. To make $\hat{\alpha}$ unbiased, $E(\hat{\alpha}) = \alpha$

$$E(\hat{\alpha}) = E(\sum_{i=1}^{n} c_i y_i)$$ y_i's are independent.
$$= \sum_{i=1}^{n} c_i E(y_i)$$
$$= \alpha \sum_{i=1}^{n} c_i$$

This implies $\sum_{i=1}^{n} c_i = 1$

$$c_i = d_i + \frac{1}{n}$$
$$\sum_{i=1}^{n} c_i = \sum_{i=1}^{n} (d_i + \frac{1}{n})$$
$$= d_1 + \frac{1}{n} + d_2 + \frac{1}{n} + \cdots\cdots + d_n + \frac{1}{n}$$
$$= 1 + \sum_{i=1}^{n} d_i$$

$\therefore \sum_{i=1}^{n} c_i = 1 \rightarrow 1 + \sum_{i=1}^{n} d_i = 1 \rightarrow \sum_{i=1}^{n} d_i = 0 \rightarrow \sum_{i=1}^{n} d_i / n = 0.$

$$Var(\hat{\alpha}) = Var(\sum_{i=1}^{n} c_i y_i)$$ y_i's are independent
$$= \sum_{i=1}^{n} Var(c_i y_i)$$
$$= c_1^2 Var(y_1) + c_2^2 Var(y_2) + \cdots\cdots + c_n^2 Var(y_n)$$
$$= \sigma^2 (c_1^2 + \cdots\cdots + c_n^2)$$
$$= \sigma^2 \left[(d_1 + \frac{1}{n})^2 + (d_2 + \frac{1}{n})^2 + \cdots\cdots + (d_n + \frac{1}{n})^2 \right]$$

$$\min Var(\hat{\alpha}) \iff \min \sigma^2 \left[(d_1 + \frac{1}{n})^2 + (d_2 + \frac{1}{n})^2 + \cdots\cdots + (d_n + \frac{1}{n})^2 \right]$$
$$\iff \min d_1^2 + \frac{2d_1}{n} + \frac{1}{n^2} + d_2^2 + \frac{2d_2}{n} + \frac{1}{n^2} + \cdots\cdots + d_n^2 + \frac{2d_n}{n} + \frac{1}{n^2}$$
$$\iff \min d_1^2 + d_2^2 + \cdots\cdots + d_n^2 + \frac{2}{n}(d_1 + d_2 + \cdots\cdots + d_n) + n \cdot \frac{1}{n^2}$$

$\because \sum d_i = 0$

$$\iff \min d_1^2 + \cdots\cdots + d_n^2 + \frac{1}{n}$$

The final optimization problem is:

$$\min M = d_1^2 + \cdots\cdots + d_n^2 + \frac{1}{n}$$
$$\text{s.t. } \sum_{i=1}^{n} d_i = 0.$$

$$\frac{\partial M}{\partial d_i} = 2d_i = 0$$

$$d_i = 0. \quad \text{for } i=1,\ldots\ldots,n$$

Therefore, to fuufill 2 conditions: unbiased and lowest variance, $d_i = 0$ $\forall i=1,\ldots\ldots,n$. $\hat{\alpha} = \bar{y}$ is the "BLUE".