1. (2.4)

   (a) In order to use the least squares procedure, we must first define RSS under this model. In our case
$$RSS(b) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - x_i b)^2$$
   Our solution will be given by $\widehat{\beta} = \arg\min_b(RSS(b))$. Taking the derivative with respect to b we have
$$\frac{d}{db}RSS(b) = \frac{d}{db}\sum_{i=1}^{n}(y_i - x_i b)^2 = \sum_{i=1}^{n}\frac{d}{db}(y_i - x_i b)^2 = -2\sum_{i=1}^{n}(y_i - x_i b)x_i$$
$$= -2\left(\sum_{i=1}^{n} y_i x_i - b\sum_{i=1}^{n} x_i^2\right) \overset{set}{=} 0$$
   Solving for b we have
$$\widehat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

   (b) i.
$$\mathbb{E}(\widehat{\beta}|X) = \mathbb{E}\left(\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}|X = x_i\right) = \frac{\sum_{i=1}^n x_i \mathbb{E}(y_i|X = x_i)}{\sum_{i=1}^n x_i^2} = \frac{\sum_{i=1}^n x_i(\beta x_i)}{\sum_{i=1}^n x_i^2}$$
$$= \beta\frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i^2} = \beta$$

   ii.
$$\mathbb{V}(\widehat{\beta}|X) = \mathbb{V}\left(\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}|X = x_i\right) = \frac{\sum_{i=1}^n x_i^2 \mathbb{V}(y_i|X = x_i)}{\left(\sum_{i=1}^n x_i^2\right)^2} = \frac{\sigma^2\sum_{i=1}^n x_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}$$

   iii. Recall that a linear combination of normally distributed variables is normally distributed. Here,
$$\widehat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} = \frac{\sum_{i=1}^n x_i(\beta x_i + e_i)}{\sum_{i=1}^n x_i^2} = \beta + \frac{1}{\sum_{i=1}^n x_i^2}\sum_{i=1}^{n} x_i e_i$$
   Recall in our framework the $\{x_i\}_{i=1}^n$ are fixed constants. Moreover by assumption $e_i \sim N(0, \sigma^2)$. Using this, we have $\widehat{\beta}|X$ is a linear combination of normally distributed random variables (namely the e_i). Hence, using the calculations for part a and b, $\widehat{\beta} \sim N(\beta, \frac{\sigma^2}{\sum_{i=1}^n x_i^2})$

2. (2.6) Here we will try to decompose SST (normally written $SSY = \sum_{i=1}^n (y_i - \overline{y})^2$) into $RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ and $SSreg = \sum_{i=1}^n (\hat{y}_i - \overline{y})^2$. That is we will try to show $SST = RSS + SSreg$. Adding and subtraction $\hat{y}_i$ in the STT formula we have

$$SST = \sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}\left[(y_i - \hat{y}_i) + (\hat{y}_i - \overline{y})\right]^2$$

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \overline{y}) + \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 = RSS + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \overline{y}) + SSreg$$

Using this form of SST, our task now reduces to showing $\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \overline{y}) = 0$.

(a) $(y_i - \hat{y}_i) = (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) = (y_i - \overline{y} + \widehat{\beta}_1\overline{x} - \widehat{\beta}_1 x_i) = (y_i - \overline{y}) - \widehat{\beta}_1(x_i - \overline{x})$

(b) $(\hat{y}_i - \overline{y}) = (\widehat{\beta}_0 + \widehat{\beta}_1 x_i - \overline{y}) = (\overline{y} - \widehat{\beta}_1\overline{x} + \widehat{\beta}_1 x_i - \overline{y}) = \widehat{\beta}_1(x_i - \overline{x})$

(c) $\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \overline{y}) = \sum_{i=1}^{n}\left[(y_i - \overline{y}) - \widehat{\beta}_1(x_i - \overline{x})\right]\left[\widehat{\beta}_1(x_i - \overline{x})\right]$
$= \widehat{\beta}_1\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x}) - \widehat{\beta}_1^2\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{SXY}{SXX}SXY - \left(\frac{SXY}{SXX}\right)^2 SXX$
$= \frac{SXY^2}{SXX} - \frac{SXY^2}{SXX} = 0$

Having shown this term does in fact equal zero, we can decompose the total variation in Y into the variation of the data from the regression line (RSS) and the regression's variation from the mean ($SSreg$). That is $SST = RSS + SSreg$.

3. (a) The parameter $\alpha = \beta_0 + \beta_1\overline{x} = \mathbb{E}(Y|X = \overline{x})$ is the expected value of Y at $X = \overline{x}$. That is, α is the expected value of the response at the sample mean of the $\{x_i\}_{i=1}^{n}$.

(b) First we define the residuals sum of squares (RSS) in in terms of candidate parameters (a, b).

$$RSS(a,b) = \sum_{i=1}^{n}\hat{e}_i^2 = \sum_{i=1}^{n}\left(y_i - (a + b(x_i - \overline{x}))\right)^2$$

From here, we define our least squares estimates by $(\hat{\alpha}, \widehat{\beta}_1) = \underset{(a,b)}{\arg\min} RSS(a,b)$. Taking partial derivatives, we have

$$\frac{\partial}{\partial a}RSS(a,b) = \sum_{i=1}^{n}\frac{\partial}{\partial a}\left(y_i - (a + b(x_i - \overline{x}))\right)^2 = -2\sum_{i=1}^{n}\left(y_i - (a + b(x_i - \overline{x}))\right)$$

$$= -2\Big(\sum_{i=1}^{n}y_i - an - b\sum_{i=1}^{n}(x_i - \overline{x})\Big) = -2\Big(\sum_{i=1}^{n}y_i - an\Big) \overset{set}{=} 0$$

Where the second to last equality is due to the fact $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$. Solving for a yields our least squares estimate of α, $\widehat{\alpha} = \overline{y}$. Now taking the derivative of RSS with respect to b we have

$$\frac{\partial}{\partial b}RSS(a,b) = \sum_{i=1}^{n}\frac{\partial}{\partial b}\left(y_i - (a + b(x_i - \overline{x}))\right)^2 = -2\sum_{i=1}^{n}\left[y_i - (a + b(x_i - \overline{x}))\right](x_i - \overline{x})$$

$$-2\Big(\sum_{i=1}^{n} y_i(x_i-\overline{x}) - a\sum_{i=1}^{n}(x_i-\overline{x}) - b\sum_{i=1}^{n}(x_i-\overline{x})^2\Big) = -2\Big(\sum_{i=1}^{n} y_i(x_i-\overline{x}) - b\sum_{i=1}^{n}(x_i-\overline{x})^2\Big)$$

Now notice that $\sum_{i=1}^{n} y_i(x_i-\overline{x}) = \sum_{i=1}^{n} y_i(x_i-\overline{x}) - \sum_{i=1}^{n} \overline{y}(x_i-\overline{x}) = \sum_{i=1}^{n}(y_i-\overline{y})(x_i-\overline{x})$. Using this fact, we have

$$-2\Big(\sum_{i=1}^{n} y_i(x_i-\overline{x}) - b\sum_{i=1}^{n}(x_i-\overline{x})^2\Big) = -2\Big(\sum_{i=1}^{n}(y_i-\overline{y})(x_i-\overline{x}) - b\sum_{i=1}^{n}(x_i-\overline{x})^2\Big) \overset{set}{=} 0$$

Solving for b, we have our estimate of β_1, $\widehat{\beta}_1 = \frac{SXY}{SXX}$

(c) i. $\mathbb{V}(\hat{\alpha}|X) = \mathbb{V}(\overline{y}|X) = \mathbb{V}(\frac{1}{n}\sum_{i=1}^{n} y_i|X) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(y_i|X=x_i) = \frac{\sigma^2}{n}$

ii. Recall that we already derived that under the assumption that $e_i \sim N(0,\sigma^2)$, $\widehat{\beta}_1|X \sim N(\beta_1, \frac{\sigma^2}{SXX})$. Hence $\mathbb{V}(\widehat{\beta}_1|X) = \frac{\sigma^2}{SXX}$

iii.

$$\text{Cov}(\hat{\alpha},\hat{\beta}_1) = \text{Cov}(\hat{\beta}_0 + \overline{x}\hat{\beta}_1, \hat{\beta}) = \mathbb{E}(\hat{\beta}_1\hat{\beta}_0 + \hat{\beta}_1^2\overline{x}) - \mathbb{E}(\hat{\beta}_0 + \overline{x}\hat{\beta}_1)\mathbb{E}(\hat{\beta}_1)$$

$$= \mathbb{E}(\hat{\beta}_1\hat{\beta}_0) + \overline{x}\mathbb{E}(\hat{\beta}_1^2) - \mathbb{E}(\hat{\beta}_0)\mathbb{E}(\hat{\beta}_1) - \overline{x}\mathbb{E}(\hat{\beta}_1)^2$$

$$= \text{Cov}(\hat{\beta}_0,\hat{\beta}_1) + \mathbb{E}(\hat{\beta}_1)\mathbb{E}(\hat{\beta}_0) + \overline{x}\mathbb{V}(\hat{\beta}_1) + \overline{x}\mathbb{E}(\hat{\beta}_1)^2 - \mathbb{E}(\hat{\beta}_0)\mathbb{E}(\hat{\beta}_1) - \overline{x}\mathbb{E}(\hat{\beta}_1)^2$$

$$= \text{Cov}(\hat{\beta}_0,\hat{\beta}_1) + \overline{x}\mathbb{V}(\hat{\beta}_1) = -\sigma^2\frac{\overline{x}}{SXX} + \sigma^2\frac{\overline{x}}{SXX} = 0$$

Hence, under this parameterization of simple linear regression, the model parameters (α, β_1) are independent.