3.4.a. The simple regression model written in matrix form becomes

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = X\beta + \varepsilon = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

The normal equations $X'Xb = X'y$ for this case are

$$\begin{pmatrix} 1 \cdots 1 \\ x_1 \cdots x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \cdots 1 \\ x_1 \cdots x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

or equivalently

$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix},$$

so that

$$\begin{cases} an + b\sum x_i &= \sum y_i \\ a\sum x_i + b\sum x_i^2 &= \sum x_i y_i \end{cases}$$

The last equations coincide with the expressions (2.9) and (2.10).

The least squares estimates are obtained from

$$\begin{aligned} \begin{pmatrix} a \\ b \end{pmatrix} &= (X'X)^{-1}X'y = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix} \\ &= \frac{1}{n\sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix} \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix} \\ &= \frac{1}{n\sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} \sum x_i^2 \sum y_j - \sum x_i \sum x_j y_j \\ n\sum x_i y_i - \sum x_i \sum y_j \end{pmatrix}. \end{aligned}$$

(In products of summed terms we use different indices i and j to avoid confusion). We can simplify the expressions for a and b as follows.

$$\begin{aligned} b &= \frac{1}{n\sum x_i^2 - (\sum x_i)^2} \left(n\sum x_i y_i - \sum x_i \sum y_j \right) = \frac{1}{n\sum x_i^2 - n^2\bar{x}^2} \left(n\sum x_i y_i - n^2\bar{x}\bar{y} \right) \\ &= \frac{1}{n\sum (x_i - \bar{x})^2} \left(n\sum (x_i - \bar{x})(y_i - \bar{y}) \right) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}, \\ a &= \frac{1}{n\sum x_i^2 - n^2\bar{x}^2} \left(\bar{y}(n\sum x_i^2 - n^2\bar{x}^2) + n^2\bar{y}\bar{x}^2 - n\bar{x}\sum x_i y_i \right) \\ &= \bar{y} - \bar{x}\frac{\sum x_i y_i - n\bar{y}\bar{x}}{\sum x_i^2 - n\bar{x}^2} = \bar{y} - b\bar{x}. \end{aligned}$$

These results coincide with the expressions (2.8) and (2.6).

Finally, the 2×2 covariance matrix of the estimators a and b is given by

$$\sigma^2(X'X)^{-1} = \frac{\sigma^2}{n\sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix},$$

so that

$$\begin{aligned} \mathsf{var}(b) &= \frac{n\sigma^2}{n\sum (x_i - \bar{x})^2} = \frac{\sigma^2}{\sum (x_i - \bar{x})^2} \\ \mathsf{var}(a) &= \frac{\sigma^2 \sum x_i^2}{n\sum (x_i - \bar{x})^2} = \sigma^2 \frac{\frac{1}{n}\sum (x_i - \bar{x})^2 + \bar{x}^2}{\sum (x_i - \bar{x})^2} = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} \right). \end{aligned}$$

These results coincide with (2.27) and (2.28).

b. Let the $k \times 1$ vector μ and $k \times k$ matrix Σ be partitioned in parts according to (the single variable) y and the ($(k-1) \times 1$ vector) $Z = (x_2, \cdots, x_k)'$. Then the result in (1.22) implies that the (scalar) random variable y_c is normally distributed with

$$y_c \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(z - \mu_2), \sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}).$$

Define $x = (1, x_2, \cdots, x_k)'$, $\beta = \begin{pmatrix} \mu_1 - & \Sigma_{12}\Sigma_{22}^{-1}\mu_2 \\ & \Sigma_{22}^{-1}\Sigma_{21} \end{pmatrix}$ and $\sigma^2 = \sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, then this can be written as

$$y_c \sim N(x'\beta, \sigma^2).$$

We can now check Assumptions 1-7.

1. The i-th observation is $y_{ci} = y_i | \{x_{2i}, \cdots, x_{ki}\}$, and as we condition on the given values of $\{x_{2i}, \cdots, x_{ki}\}$ these values are fixed. The fact that Σ is nonsingular ensures that the $n \times k$ matrix X (with $n \geq k$) has rank k (with probability 1, that is, only in exceptional cases it may have smaller rank).

2. Conditional on x_i, it follows that $\varepsilon_i = y_{c,i} - x_i'\beta \sim N(0, \sigma^2)$ so that $E[\varepsilon_i] = 0$.

3. This follows as $\sigma^2 = \sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ is constant and does not depend on the values of x_i.

4. As the n observations are obtained by a random sample this means that the values of $\{y, x_2, \cdots, x_k\}$ are independent for different observations $i \neq j$, so that also $\varepsilon_i = y_{c,i} - x_i'\beta$ and $\varepsilon_j = y_{c,j} - x_j'\beta$ are independent for $i \neq j$.

5. The values of $\beta = \begin{pmatrix} \mu_1 - & \Sigma_{12}\Sigma_{22}^{-1}\mu_2 \\ & \Sigma_{22}^{-1}\Sigma_{21} \end{pmatrix}$ and $\sigma^2 = \sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ are the same for all n observations, as they do not depend on the values of x_i.

6. As $\varepsilon_i = y_{c,i} - x_i'\beta$ it follows that $y_{c,i} = x_i'\beta + \varepsilon_i$.

7. $\varepsilon_i \sim N(0, \sigma^2)$, see Assumption 2, and independence follows from the above comments on Assumption 4.