1. Suppose we have postulated the model

$$Y_i = \beta_1 x_i + \epsilon_i,$$

for $i = 1, 2, ..., n$, where $\epsilon_i \sim$ iid $\mathcal{N}(0, \sigma^2)$. This is our simple linear regression model where the intercept parameter $\beta_0 = 0$. This model is called the *no-intercept model*.
(a) The least squares estimator of β_1 in the no-intercept model is the value of β_1 that minimizes the objective function

$$Q(\beta_1) = \sum_{i=1}^{n} (Y_i - \beta_1 x_i)^2.$$

Show this least squares estimator is

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}.$$

(b) Derive the sampling distribution of $\widehat{\beta}_1$ in the no-intercept model. Identify which error assumptions are needed where in your derivations.

2. Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

for $i = 1, 2, ..., n$, where $\epsilon_i \sim$ iid $\mathcal{N}(0, \sigma^2)$.
(a) Recall the ith residual is $e_i = Y_i - \widehat{Y}_i$. When least squares is used to estimate the simple linear regression model, show

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i) = 0,$$

that is, the sum of the residuals is equal to zero. Interestingly, this "residuals sum to zero" property does not necessarily hold in the no-intercept model in Problem 1.
(b) Show

$$e_i \sim \mathcal{N}\left(0, \sigma^2(1 - h_{ii})\right),$$

where

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}.$$

This implies the residuals (unlike the errors) do not have constant variance; i.e., $V(e_i)$ changes as i does. In data analysis, h_{ii} is called the *leverage* associated with the ith observation. Leverages can be useful in classifying observations as outliers or not.

3. Suppose $Y_1, Y_2, ..., Y_n$ are independent normal random variables with $E(Y_i) = \beta_0 + \beta_1 x_i$ and $V(Y_i) = \sigma^2$, for $i = 1, 2, ..., n$.

(a) Show the maximum likelihood estimators of β_0, β_1, and σ^2 are

$$\begin{aligned}
\widehat{\beta}_0 &= \overline{Y} - \widehat{\beta}_1 \overline{x} \\
\widehat{\beta}_1 &= \frac{\sum_{i=1}^n (x_i - \overline{x})(Y_i - \overline{Y})}{\sum_{i=1}^n (x_i - \overline{x})^2} \\
\widehat{\sigma}^2 &= \frac{\text{SSE}}{n},
\end{aligned}$$

where $\text{SSE} = \sum_{i=1}^n (Y_i - \widehat{Y}_i)^2 = \sum_{i=1}^n (Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i)^2$. Under normality assumptions, note the MLEs of β_0 and β_1 are identical to the least squares estimators. The MLE of the error variance σ^2 is slightly different than the (unbiased) version presented in the notes (the MLE here is biased).

(b) Show the likelihood ratio test for

$$\begin{gathered}
H_0 : \beta_1 = 0 \\
\text{versus} \\
H_a : \beta_1 \neq 0
\end{gathered}$$

is equivalent to the t test given in the notes with $\beta_{1,0} = 0$ (see pp 109-110).

4. Suppose

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$$

is a trivariate random vector with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\mathbf{V}$, where

$$\boldsymbol{\mu} = \begin{pmatrix} 4 \\ 6 \\ 10 \end{pmatrix} \quad \text{and} \quad \mathbf{V} = \begin{pmatrix} 4 & -3 & 0 \\ -3 & 9 & 2 \\ 0 & 2 & 1 \end{pmatrix}.$$

(a) Find the mean and variance of $U = Y_1 - Y_2 + Y_3$. Find the covariance of U and V, where $V = 2Y_1 + 3Y_2 - Y_3$.

(b) Define

$$\mathbf{A} = \begin{pmatrix} 2 & 4 & -2 \\ 1 & 2 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{c} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}.$$

Calculate $E(\mathbf{c} + \mathbf{AY})$ and $\text{Cov}(\mathbf{c} + \mathbf{AY})$.

(c) Define

$$\mathbf{B} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 3 & -1 \\ -1 & 4 & -1 \end{pmatrix}.$$

Calculate $E(\mathbf{Y}'\mathbf{BY})$.

5. Suppose the random vector

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$$

has a trivariate normal distribution with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\mathbf{V}$, where

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{V} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 2 & -1 \\ 1 & -1 & 4 \end{pmatrix},$$

that is, $\mathbf{Y} \sim \mathcal{N}_3(\boldsymbol{\mu}, \mathbf{V})$.
(a) What is the distribution of $U = Y_1 + 3Y_2 - 2Y_3$?
(b) Define

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 0 & 3 \\ 2 & 3 & 1 \end{pmatrix}.$$

Calculate $E(\mathbf{Y}'\mathbf{A}\mathbf{Y})$.
(c) When $\mathbf{Y}$ is multivariate normal, then the variance of a quadratic form is

$$V(\mathbf{Y}'\mathbf{A}\mathbf{Y}) = 2[\mathrm{tr}(\mathbf{A}\mathbf{V})]^2 + 4\boldsymbol{\mu}'\mathbf{A}\mathbf{V}\mathbf{A}\boldsymbol{\mu}.$$

This formula is only correct when $\mathbf{Y}$ follows a multivariate normal distribution (and it also requires $\mathbf{A}$ to be symmetric). Calculate $V(\mathbf{Y}'\mathbf{A}\mathbf{Y})$ with the matrix $\mathbf{A}$ in part (b).