2.25 Consider the simple linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$, with $E(\varepsilon) = 0$, $\mathrm{Var}(\varepsilon) = \sigma^2$, and ε uncorrelated.

a. Show that $\mathrm{Cov}\left(\hat{\beta}_0, \hat{\beta}_1\right) = -\overline{x}\sigma^2 / S_{xx}$.

b. Show that $\mathrm{Cov}\left(\overline{y}, \hat{\beta}_1\right) = 0$.

2.26 Consider the simple linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$, with $E(\varepsilon) = 0$, $\mathrm{Var}(\varepsilon) = \sigma^2$, and ε uncorrelated.

a. Show that $E(MS_{\mathrm{R}}) = \sigma^2 + \beta_1^2 S_{xx}$.

b. Show that $E(MS_{\mathrm{Res}}) = \sigma^2$.

2.27 Suppose that we have fit the straight-line regression model $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1$ but the response is affected by a second variable x_2 such that the true regression function is

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

a. Is the least-squares estimator of the slope in the original simple linear regression model unbiased?

b. Show the bias in $\hat{\beta}_1$.

2.28 Consider the maximum-likelihood estimator $\tilde{\sigma}^2$ of σ^2 in the simple linear regression model. We know that $\tilde{\sigma}^2$ is a biased estimator for σ^2.

a. Show the amount of bias in $\tilde{\sigma}^2$.

b. What happens to the bias as the sample size n becomes large?

3.27 Show that $\mathrm{Var}(\hat{\mathbf{y}}) = \sigma^2 \mathbf{H}$.

3.28 Prove that the matrices $\mathbf{H}$ and $\mathbf{I} - \mathbf{H}$ are idempotent, that is, $\mathbf{HH} = \mathbf{H}$ and $(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H}) = \mathbf{I} - \mathbf{H}$.

3.29 For the simple linear regression model, show that the elements of the hat matrix are

$$h_{ij} = \frac{1}{n} + \frac{(x_i - \overline{x})(x_j - \overline{x})}{S_{xx}} \quad \text{and} \quad h_{ii} = \frac{1}{n} + \frac{(x_i - \overline{x})^2}{S_{xx}}$$

Discuss the behavior of these quantities as x_i moves farther from $\overline{x}$,

3.30 Consider the multiple linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Show that the least-squares estimator can be written as

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \mathbf{R}\boldsymbol{\varepsilon} \quad \text{where} \quad \mathbf{R} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

3.31 Show that the residuals from a linear regression model can be expressed as $\mathbf{e} = (\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon}$. [Hint: Refer to Eq. (3.15b).]

3.32 For the multiple linear regression model, show that $SS_{\mathrm{R}}(\boldsymbol{\beta}) = \mathbf{y}'\mathbf{H}\mathbf{y}$.

3.33 Prove that R^2 is the square of the correlation between $\mathbf{y}$ and $\hat{\mathbf{y}}$.

3.35 Let $\mathbf{x}_j$ be the jth row of $\mathbf{X}$, and $\mathbf{X}_{-j}$ be the $\mathbf{X}$ matrix with the jth row removed. Show that

$$\operatorname{Var}\left[\hat{\beta}_j\right]=\sigma^2\left[\mathbf{x}_j'\mathbf{x}_j-\mathbf{x}_j'\mathbf{X}_{-j}(\mathbf{X}_{-j}'\mathbf{X}_{-j})^{-1}\mathbf{X}_{-j}'\mathbf{x}_j\right]$$

3.36 Consider the following two models where $E(\boldsymbol{\varepsilon})=\mathbf{0}$ and $\operatorname{Var}(\boldsymbol{\varepsilon})=\sigma^2\mathbf{I}$:

Model A: $\mathbf{y}=\mathbf{X}_1\boldsymbol{\beta}_1+\boldsymbol{\varepsilon}$

Model B: $\mathbf{y}=\mathbf{X}_1'\boldsymbol{\beta}_1+\mathbf{X}_2\boldsymbol{\beta}_2+\boldsymbol{\varepsilon}$

Show that $R_A^2\le R_B^2$.

3.37 Suppose we fit the model $\mathbf{y}=\mathbf{X}_1\boldsymbol{\beta}_2+\boldsymbol{\varepsilon}$ when the true model is actually given by $\mathbf{y}=\mathbf{X}_1\boldsymbol{\beta}_2+\mathbf{X}_2\boldsymbol{\beta}_2+\boldsymbol{\varepsilon}$. For both models, assume $E(\boldsymbol{\varepsilon})=\mathbf{0}$ and $\operatorname{Var}(\boldsymbol{\varepsilon})=\sigma^2\mathbf{I}$. Find the expected value and variance of the ordinary least-squares estimate, $\hat{\boldsymbol{\beta}}_1$. Under what conditions is this estimate unbiased?

3.38 Consider a correctly specified regression model with p terms, including the intercept. Make the usual assumptions about $\boldsymbol{\varepsilon}$. Prove that

$$\sum_{i=1}^{n}\operatorname{Var}(\hat{y}_i)=p\sigma^2$$

N. Show that the square of the multiple correlation coefficient R^2 is equal to the square of the correlation between $\mathbf{Y}$ and $\hat{\mathbf{Y}}$

O. Consider the formal regression of the residuals e_i onto a quadratic function $\alpha_0+\alpha_1\hat{Y}_i+\alpha_2\hat{Y}_i^2$ of the fitted values $\hat{Y}_i$, by least squares. Show that all three estimated coefficients depend on $T_{12}=\Sigma e_i\hat{Y}_i^2$. What does this imply?

P. We fit a straight line model to a set of data using the formulas $\mathbf{b}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, $\hat{\mathbf{Y}}=\mathbf{X}\mathbf{b}$ with the usual definitions. We define $\mathbf{H}=\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Show that

$$\begin{aligned}\text{SS(due to regression)}&=\mathbf{Y}'\mathbf{H}\mathbf{Y}\\&=\hat{\mathbf{Y}}'\hat{\mathbf{Y}}\\&=\hat{\mathbf{Y}}'\mathbf{H}^3\mathbf{Y}.\end{aligned}$$

Q. Show that $\mathbf{X}'\mathbf{e}=\mathbf{0}$.

T. Suppose $\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}$ is a regression model containing a β_0 term in the first position, and $\mathbf{1}=(1,1,\ldots,1)'$ is an $n\times 1$ vector of ones. Show that $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{1}=(1,0,\ldots,0)'$ and hence that $\mathbf{1}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{1}=n$. (*Hint:* $\mathbf{X}'\mathbf{1}$ is the first column of $\mathbf{X}'\mathbf{X}$.) These results can be useful in regression matrix manipulations. For connected reading, see letters in The *American Statistician*, April 1972, 47–48.

U. By noting that $\mathbf{X}_0=(1,\overline{X}_1,\overline{X}_2,\ldots)'$ can be written as $\mathbf{X}'\mathbf{1}/n$, and applying the result in Exercise T above, show that $V(\hat{Y})$ at the point $(\overline{X}_1,\overline{X}_2,\ldots,\overline{X}_k)$ is σ^2/n.