Homework 14

Put your name and student ID here

2021-06-03

Q1: Let us consider fitting a straight line, $y = \beta_0 + \beta_1 x$, to points (x_i, y_i), where $i = 1, \ldots, n$.

1. Write down the normal equations for the simple linear model via the matrix formalism.
2. Solve the normal equations by the matrix approach and see whether the solutions agree with the earlier calculation derived in the simple linear models.

Q2: Consider fitting the curve $y = \beta_0 x + \beta_1 x^2$ to points (x_i, y_i), where $i = 1, \ldots, n$.

1. Use the matrix formalism to find expressions for the least squares estimates of β_0 and β_1.
2. Find an expression for the covariance matrix of the estimates.

Q3: Suppose that in the model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ i = 1, \ldots, n,$$

the errors ϵ_i have mean zero and are uncorrelated, but $\mathrm{Var}(\epsilon_i) = \rho_i^2 \sigma^2$, where the $\rho_i > 0$ are known constants, so the errors do not have equal variance. Because the variances are not equal, the theory developed in our class does not apply.

(a) Try to transform suitably the model such that the basic assumptions (i.e., the errors have zero mean and equal variance, and are uncorrelated) of the standard statistical model are satisfied.

(b) Find the least squares estimates of β_0 and β_1 for the transformed model.

(c) Find the variances of the estimates of Part (b).

Q4: Consider the multiple linear model $Y = X\beta + \epsilon$, where X is the $n \times p$ design matrix, $\beta = (\beta_0, \ldots, \beta_{p-1})^\top$ is a vector of p parameters, and the error $\epsilon \sim N(0, \sigma^2 I_n)$. Now consider the problem of estimating $\theta = \beta_0 + \beta_1 + \cdots + \beta_{p-1}$. Assume that $\mathrm{rank}(X) = p < n$. Let $\hat{\beta} = (\hat{\beta}_0, \ldots, \hat{\beta}_{p-1})^\top$ be the least squares estimate of β. Let $\hat{\theta} = \hat{\beta}_0 + \hat{\beta}_1 + \cdots + \hat{\beta}_{p-1}$.

(a) Show that $\hat{\theta}$ is an unbaised estimate of θ.

(b) Find the variance of the estimate $\hat{\theta}$.

(c) Let $\hat{\theta}_c = c^\top Y$ be an unbiased estimate of θ for any $\beta \in \mathbb{R}^{p\times 1}$, where $c \in \mathbb{R}^{n\times 1}$ is any fixed vector. Prove that $\mathrm{Var}(\hat{\theta}_c) \geq \mathrm{Var}(\hat{\theta})$. (Notice that $\hat{\theta}$ is also a linear combination of y_i. This result implies that $\hat{\theta}$ is the best linear unbiased estimator for θ.)

Q5: Consider the linear model in matrix formalism

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{Y} = (y_1, \ldots, y_n)^\top$, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{p-1})^\top$, $\boldsymbol{X}$ is the $n \times p$ design matrix, and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top \sim N(\boldsymbol{0}, \sigma^2 I_n)$ with unknown $\sigma > 0$. Assume that $\mathrm{rank}(\boldsymbol{X}) = r < p$.

(a) Show that the least squares estimator (LSE) for $\boldsymbol{\beta}$ is not unique.

(b) Show that there exists an $n \times r$ submatrix $\boldsymbol{X}^*$ of $\boldsymbol{X}$ with rank r such that $\boldsymbol{X} = \boldsymbol{X}^*\boldsymbol{Q}$, where $\boldsymbol{Q}$ is a $r \times p$ matrix.

(c) Let $\boldsymbol{\beta}^* = \boldsymbol{Q}\boldsymbol{\beta}$. Then the linear model becomes $\boldsymbol{Y} = \boldsymbol{X}^*\boldsymbol{\beta}^* + \boldsymbol{\epsilon}$. Find an LSE for $\boldsymbol{\beta}^*$ and show that the LSE is unique. Find an unbiased estimate of σ^2 and show its variance.