24500 HW5

Bin Yu

Feb 13, 2025

Question 1

Calculation of $\hat{\beta}_1$ and $\hat{\beta}_0$ and Their Variances

Consider the linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n,$$

with $\varepsilon_i \sim N(0, \sigma^2)$. The least squares estimators are obtained by minimizing

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2.$$

Taking partial derivatives with respect to β_0 and β_1 and setting them to zero gives the normal equations:

$$\frac{\partial Q}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i,$$

$$\frac{\partial Q}{\partial \beta_1} = -2 \sum_{i=1}^{n} x_i (y_i - \beta_0 - \beta_1 x_i) = 0.$$

From the first equation, it follows that

$$\beta_0 = \bar{y} - \beta_1 \bar{x}, \quad \text{with} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \quad \text{and} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Substituting $\beta_0 = \bar{y} - \beta_1 \bar{x}$ into the second normal equation:

$$\sum_{i=1}^{n} x_i \Big(y_i - (\bar{y} - \beta_1 \bar{x}) - \beta_1 x_i \Big) = 0.$$

$$\sum_{i=1}^{n} x_i \Big(y_i - \bar{y} + \beta_1 \bar{x} - \beta_1 x_i \Big) = 0.$$

$$\sum_{i=1}^{n} \Big(x_i y_i - x_i \bar{y} + \beta_1 \bar{x}\, x_i - \beta_1 x_i^2 \Big) = 0.$$

Separate the sum into individual terms:

$$\sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i + \beta_1 \bar{x} \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i^2 = 0.$$

Since:

$$\sum_{i=1}^{n} x_i = n\bar{x}.$$

$$\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} + \beta_1 \bar{x}\,(n\bar{x}) - \beta_1 \sum_{i=1}^{n} x_i^2 = 0.$$

$$\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} + \beta_1 n\bar{x}^2 - \beta_1 \sum_{i=1}^{n} x_i^2 = 0.$$

Rearrange the terms to isolate the terms involving β_1:

$$\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} = \beta_1 \left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \right).$$

Notice that

$$\sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \bar{x}^2 = \sum_{i=1}^{n}\Big(x_i^2 - 2x_i\bar{x} + \bar{x}^2 \Big) = \sum_{i=1}^{n}(x_i - \bar{x})^2,$$

and also observe that

$$\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} = \sum_{i=1}^{n} (x_i y_i - \bar{x}\bar{y}) = \sum_{i=1}^{n}\Big[(x_i - \bar{x})y_i\Big],$$

since

$$\sum_{i=1}^{n}(x_i - \bar{x})y_i = \sum_{i=1}^{n} x_i y_i - \bar{x}\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}.$$

Thus, we have:

$$\beta_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

Then, the estimator for β_0 is given by

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}.$$

Variance of $\hat{\beta}_1$: Since

$$\hat{\beta}_1 = \frac{1}{S_x}\sum_{i=1}^{n}(x_i - \bar{x})y_i, \quad \text{with} \quad S_x = \sum_{i=1}^{n}(x_i - \bar{x})^2,$$

and each y_i has variance $\mathrm{Var}(y_i) = \sigma^2$ (with the y_i's independent), it follows that

$$\mathrm{Var}(\hat{\beta}_1) = \frac{1}{S_x^2}\sum_{i=1}^{n}(x_i - \bar{x})^2\ \mathrm{Var}(y_i) = \frac{\sigma^2}{S_x}.$$

Variance of $\hat{\beta}_0$:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}.$$

Using the properties of variance for a linear combination, we have

$$\mathrm{Var}(\hat{\beta}_0) = \mathrm{Var}(\bar{y}) + \bar{x}^2\,\mathrm{Var}(\hat{\beta}_1) - 2\bar{x}\ \mathrm{Cov}(\bar{y}, \hat{\beta}_1).$$

Since

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\mathrm{Var}(\bar{y}) = \frac{\sigma^2}{n},$$

and

$$\hat{\beta}_1 = \frac{1}{S_x}\sum_{i=1}^{n}(x_i - \bar{x})y_i,$$

it can be shown (using $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$) that

$$\mathrm{Cov}(\bar{y}, \hat{\beta}_1) = 0.$$

Thus,

$$\mathrm{Var}(\hat{\beta}_0) = \frac{\sigma^2}{n} + \bar{x}^2\,\frac{\sigma^2}{S_x}.$$

Proof that $\mathrm{Cov}(\hat{\beta}_1, r_i) = 0$

Since

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\,\bar{x},$$

Rewrite the residual as

$$r_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i = \big(y_i - \bar{y}\big) - \hat{\beta}_1(x_i - \bar{x}).$$

Thus,

$$\mathrm{Cov}(\hat{\beta}_1, r_i) = \mathrm{Cov}\Big(\hat{\beta}_1,\ (y_i - \bar{y}) - \hat{\beta}_1(x_i - \bar{x})\Big).$$

By the linearity of covariance,

$$\mathrm{Cov}(\hat{\beta}_1, r_i) = \mathrm{Cov}(\hat{\beta}_1, y_i - \bar{y}) - (x_i - \bar{x})\ \mathrm{Var}(\hat{\beta}_1).$$

Since

$$\hat{\beta}_1 = \frac{1}{S_x}\sum_{j=1}^{n}(x_j - \bar{x})\,y_j, \quad \text{with} \quad S_x = \sum_{j=1}^{n}(x_j - \bar{x})^2.$$

Then, by the linearity of covariance,

$$\mathrm{Cov}(\hat{\beta}_1, y_i - \bar{y}) = \frac{1}{S_x}\sum_{j=1}^{n}(x_j - \bar{x})\ \mathrm{Cov}(y_j, y_i - \bar{y}).$$

For any j:

$$\mathrm{Cov}(y_j, y_i - \bar{y}) = \begin{cases} \mathrm{Var}(y_i) - \mathrm{Cov}(y_i, \bar{y}) = \sigma^2 - \frac{\sigma^2}{n}, & \text{if } j = i, \\ 0 - \mathrm{Cov}(y_j, \bar{y}) = -\frac{\sigma^2}{n}, & \text{if } j \neq i. \end{cases}$$

Thus,

$$\mathrm{Cov}(\hat{\beta}_1, y_i - \bar{y}) = \frac{1}{S_x}\left[(x_i - \bar{x})\Big(\sigma^2 - \frac{\sigma^2}{n}\Big) - \frac{\sigma^2}{n}\sum_{j\neq i}(x_j - \bar{x})\right].$$

Since

$$\sum_{j=1}^{n}(x_j - \bar{x}) = 0$$

$$\sum_{j\neq i}(x_j - \bar{x}) = -(x_i - \bar{x}),$$

$$\mathrm{Cov}(\hat{\beta}_1, y_i - \bar{y}) = \frac{1}{S_x}\left[(x_i - \bar{x})\left(\sigma^2 - \frac{\sigma^2}{n}\right) + \frac{\sigma^2}{n}(x_i - \bar{x})\right] = \frac{(x_i - \bar{x})\sigma^2}{S_x}.$$

and,

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{S_x}.$$

Substitute the above results into the expression for $\mathrm{Cov}(\hat{\beta}_1, r_i)$:

$$\mathrm{Cov}(\hat{\beta}_1, r_i) = \frac{(x_i - \bar{x})\sigma^2}{S_x} - (x_i - \bar{x})\,\frac{\sigma^2}{S_x} = 0.$$

Proof that $\mathrm{Cov}(\hat{\beta}_0, r_i) = 0$

Since

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\,\bar{x},$$

we have

$$\mathrm{Cov}(\hat{\beta}_0, r_i) = \mathrm{Cov}(\bar{y} - \hat{\beta}_1\,\bar{x}, r_i) = \mathrm{Cov}(\bar{y}, r_i) - \bar{x}\ \mathrm{Cov}(\hat{\beta}_1, r_i).$$

Since we have already shown that $\mathrm{Cov}(\hat{\beta}_1, r_i) = 0$, it suffices to show that

$$\mathrm{Cov}(\bar{y}, r_i) = 0.$$

$$r_i = (y_i - \bar{y}) - \hat{\beta}_1(x_i - \bar{x}),$$

so that

$$\mathrm{Cov}(\bar{y}, r_i) = \mathrm{Cov}(\bar{y}, y_i - \bar{y}) - (x_i - \bar{x})\ \mathrm{Cov}(\bar{y}, \hat{\beta}_1).$$

For $\mathrm{Cov}(\bar{y}, \hat{\beta}_1)$:

$$\bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j \quad \text{and} \quad \hat{\beta}_1 = \frac{1}{S_x}\sum_{j=1}^{n}(x_j - \bar{x})\,y_j.$$

Thus,

$$\mathrm{Cov}(\bar{y}, \hat{\beta}_1) = \frac{1}{nS_x}\sum_{j=1}^{n}(x_j - \bar{x})\ \mathrm{Var}(y_j) = \frac{\sigma^2}{nS_x}\sum_{j=1}^{n}(x_j - \bar{x}) = 0,$$

since $\sum_{j=1}^{n}(x_j - \bar{x}) = 0$.

For $\mathrm{Cov}(\bar{y}, y_i - \bar{y})$:

$$\mathrm{Cov}(\bar{y}, y_i - \bar{y}) = \mathrm{Cov}(\bar{y}, y_i) - Var(\bar{y}) = \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0$$

Thus:

$$\mathrm{Cov}(\bar{y}, r_i) = 0 - 0 = 0,$$

and hence,

$$\mathrm{Cov}(\hat{\beta}_0, r_i) = 0.$$

We have shown that for every i,

$$\mathrm{Cov}(\hat{\beta}_1,\, y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad \text{and} \quad \mathrm{Cov}(\hat{\beta}_0,\, y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0.$$

Since $\hat{\beta}_0$, $\hat{\beta}_1$, and r_i are all linear combinations of the normally distributed y_j's, zero covariance implies independence. Therefore, the residual sum of squares

$$\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2,$$

being a function of the residuals, is independent of $(\hat{\beta}_0, \hat{\beta}_1)$.

Question 2

(a)

Define the design matrix and response vector as

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_{n_1+n_2} \end{pmatrix}, \qquad y = \begin{pmatrix} y_1 \\ \vdots \\ y_{n_1+n_2} \end{pmatrix}.$$

Suppose the regression model is given by

$$y = X\beta + \varepsilon,$$

where the error vector ε follows a multivariate normal distribution with mean zero and covariance matrix

$$V = \operatorname{diag}(\underbrace{\sigma^2, \ldots, \sigma^2}_{n_1}, \underbrace{2\sigma^2, \ldots, 2\sigma^2}_{n_2}).$$

Thus,

$$\varepsilon \sim N(0, V).$$

Since $y = X\beta + \varepsilon$, it follows that

$$y \sim N(X\beta, V).$$

The probability density function for y is given by

$$f(y \mid \beta) = \frac{1}{(2\pi)^{\frac{n}{2}} |V|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2}(y - X\beta)^T V^{-1}(y - X\beta) \right\},$$

where $n = n_1 + n_2$.

Taking the natural log of the likelihood function:

$$\ell(\beta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|V| - \frac{1}{2}(y - X\beta)^T V^{-1}(y - X\beta).$$

The first two terms, $-\frac{n}{2}\log(2\pi)$ and $-\frac{1}{2}\log|V|$, do not depend on β. Therefore, maximizing $\ell(\beta)$ with respect to β is equivalent to minimizing

$$Q(\beta) = (y - X\beta)^T V^{-1}(y - X\beta).$$

Expanding $Q(\beta)$:

$$\begin{aligned} Q(\beta) &= (y - X\beta)^T V^{-1}(y - X\beta) \\ &= y^T V^{-1} y - y^T V^{-1} X\beta - \beta^T X^T V^{-1} y + \beta^T X^T V^{-1} X\beta. \end{aligned}$$

$y^T V^{-1} X\beta$ is a scalar and equals its transpose:

$$y^T V^{-1} X\beta = \beta^T X^T V^{-1} y.$$

Thus, the objective function simplifies to

$$Q(\beta) = y^T V^{-1} y - 2\beta^T X^T V^{-1} y + \beta^T X^T V^{-1} X\beta.$$

use the following matrix calculus results to differentiate with respect to β

- The derivative of a constant term $y^T V^{-1} y$ is 0.
- The derivative of the linear term $-2\beta^T X^T V^{-1} y$ is $-2X^T V^{-1} y$.
- For the quadratic term $\beta^T X^T V^{-1} X\beta$, since $X^T V^{-1} X$ is symmetric, its derivative is $2X^T V^{-1} X\,\beta$.

Thus, the derivative is given by

$$\frac{\partial Q(\beta)}{\partial \beta} = -2X^T V^{-1} y + 2X^T V^{-1} X\,\beta.$$

Set the derivative equal to zero:

$$-2X^T V^{-1} y + 2X^T V^{-1} X\,\beta = 0.$$
$$X^T V^{-1} X\,\beta = X^T V^{-1} y.$$

Assuming that $X^T V^{-1} X$ is invertible, the solution is

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} y.$$

Therefore, the MLE is

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} y.$$

(b)

Since the model is $y = X\beta + \varepsilon$ with $E(\varepsilon) = 0$, we have

$$E(\hat{\beta}) = (X^T V^{-1} X)^{-1} X^T V^{-1} E(y) = (X^T V^{-1} X)^{-1} X^T V^{-1} X\,\beta = \beta.$$

Thus, the MLE is unbiased.

(c)

Given that $\hat{\beta}$ is a linear estimator,

$$\hat{\beta} = A\,y, \quad \text{with} \quad A = (X^T V^{-1} X)^{-1} X^T V^{-1},$$

and using the formula for the variance of a linear transformation,

$$\mathrm{Var}(Ay) = A\,\mathrm{Var}(y)\,A^T,$$

and since $\mathrm{Var}(y) = V$, and V^{-1}, $X^T V^{-1} X$ is symmetric, we have

$$\mathrm{Var}(\hat{\beta}) = (X^T V^{-1} X)^{-1} X^T V^{-1}\,V\,V^{-1} X (X^T V^{-1} X)^{-1}.$$
$$\mathrm{Var}(\hat{\beta}) = (X^T V^{-1} X)^{-1} X^T V^{-1}\,\mathrm{Var}(y) V^{-1} X (X^T V^{-1} X)^{-1}$$
$$\mathrm{Var}(\hat{\beta}) = (X^T V^{-1} X)^{-1} X^T V^{-1} V V^{-1} X (X^T V^{-1} X)^{-1}$$

Noticing that

$$V^{-1} V\,V^{-1} = V^{-1},$$

this simplifies to

$$\mathrm{Var}(\hat{\beta}) = (X^T V^{-1} X)^{-1} X^T V^{-1} X (X^T V^{-1} X)^{-1} = (X^T V^{-1} X)^{-1}.$$

(d)

Since $y \sim N(X\beta, V)$ and $\hat{\beta}$ is a linear function of y, it follows that

$$\hat{\beta} \sim N\Big(\beta,\, (X^T V^{-1} X)^{-1}\Big).$$