Assignment 1

1. Question 1

Suppose the experimenter postulates a model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \varepsilon_i \sim i.i.d. N(0, \sigma^2),\ i = 1, 2, \cdots, n$$

where β_0 is known.

(a). What is the appropriate least squares estimator of β_1? Justify your answer.

(b). What is the variance of the estimator of the slope in (a)?

1.1 (a) Derivation of the Least Squares Estimator (LSE) for the Slope β_1

The model is $y = \beta_0 \mathbf{1}_n + \beta_1 \mathbf{x} + \varepsilon$ (where β_0 is known, $\varepsilon \sim i.i.d. N(0, \sigma^2 \mathbf{I}_n)$, $\mathbf{1}_n$ is an n-dimensional vector of ones, and $\mathbf{x}$ is the independent variable vector).

The goal of least squares is to minimize the sum of squared residuals:

$$M(\beta_1) = \|y - \beta_0 \mathbf{1}_n - \beta_1 \mathbf{x}\|_2^2$$

Since $M(\beta_1)$ is a convex function, take the derivative with respect to β_1 and set it to zero:

$$\frac{\partial M}{\partial \beta_1} = -2(y - \beta_0 \mathbf{1}_n - \beta_1 \mathbf{x})^T \mathbf{x} = 0$$

Expanding this equation gives:

$$y^T \mathbf{x} - \beta_0 \mathbf{1}_n^T \mathbf{x} - \beta_1 \mathbf{x}^T \mathbf{x} = 0$$

Solving for the LSE of β_1 yields:

$$\hat{\beta}_1 = \frac{y^T \mathbf{x} - \beta_0 \mathbf{1}_n^T \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \frac{y^T \mathbf{x} - n\beta_0 \bar{x}}{\mathbf{x}^T \mathbf{x}}$$

(where $\bar{x} = \frac{1}{n} \mathbf{1}_n^T \mathbf{x}$ is the sample mean of x_i)

1.2 (b) Calculation of the Variance of $\hat{\beta}_1$

Substitute $y = \beta_0 \mathbf{1}_n + \beta_1 \mathbf{x} + \varepsilon$ into the expression for $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\mathbf{x}^T(\beta_0 \mathbf{1}_n + \beta_1 \mathbf{x} + \varepsilon)}{\mathbf{x}^T \mathbf{x}} - n\frac{\beta_0 \bar{x}}{\mathbf{x}^T \mathbf{x}}$$

Expand and simplify the expression:

$$\begin{aligned} \hat{\beta}_1 &= \frac{n\beta_0 \bar{x} + \beta_1 \mathbf{x}^T \mathbf{x} + \mathbf{x}^T \varepsilon}{\mathbf{x}^T \mathbf{x}} - \frac{n\beta_0 \bar{x}}{\mathbf{x}^T \mathbf{x}} \\ &= \beta_1 + \frac{\mathbf{x}^T \varepsilon}{\mathbf{x}^T \mathbf{x}} \end{aligned}$$

Given that $\varepsilon \sim N(0, \sigma^2 \mathbf{I}_n)$, using the variance property $\mathrm{Var}(a^T \varepsilon) = a^T \mathrm{Var}(\varepsilon) a$, we obtain:

$$\mathrm{Var}(\hat{\beta}_1) = \mathrm{Var}\left(\frac{\mathbf{x}^T \varepsilon}{\mathbf{x}^T \mathbf{x}}\right) = \frac{\mathbf{x}^T \cdot \sigma^2 \mathbf{I}_n \cdot \mathbf{x}}{(\mathbf{x}^T \mathbf{x})^2} = \frac{\sigma^2 \mathbf{x}^T \mathbf{x}}{(\mathbf{x}^T \mathbf{x})^2} = \frac{\sigma^2}{\mathbf{x}^T \mathbf{x}}$$

---

2. Question 2

Consider a model $y_i = \beta + \varepsilon_i$ where β is a constant parameter and $\varepsilon_i \sim i.i.d. N(0, \sigma^2)$. Find LSE and MLE for β.

2.1 solution:

To solve for the LSE and MLE of β in the model $y_i = \beta + \varepsilon_i$ ($\varepsilon_i \sim i.i.d. N(0, \sigma^2)$), we can analyze as follows:

2.1.1 Least Squares Estimator (LSE)

Model in vector form: $\mathbf{y} = \beta \mathbf{1}_n + \varepsilon$, where $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$, $\mathbf{1}_n$ is an n-dimensional vector of ones, and $\varepsilon \sim N(0, \sigma^2 \mathbf{I}_n)$.

- **Objective**: Minimize the L_2-norm of residuals:

$$\|\mathbf{y} - \beta \mathbf{1}_n\|_2$$

- **Define residual function**: Let $M(\beta) = \|\mathbf{y} - \beta \mathbf{1}_n\|_2^2$. Since $M(\beta)$ is a convex function in β, we find its minimum by taking the derivative and setting it to zero.
- **Differentiate and solve**:

  $$\frac{\partial M}{\partial \beta} = -2(\mathbf{y} - \beta \mathbf{1}_n)^T \mathbf{1}_n = -2(n\bar{y} - n\beta) = 0$$

  (where $\bar{y} = \frac{1}{n}\mathbf{1}_n^T \mathbf{y}$ is the sample mean of y_i).

  Solving gives the LSE:

  $$\hat{\beta}_{\text{LSE}} = \bar{y}$$

2.1.2 Maximum Likelihood Estimator (MLE)

- **Likelihood function**: Since $\mathbf{y} \sim N(\beta \mathbf{1}_n, \sigma^2 \mathbf{I}_n)$, the likelihood function is:

  $$L(\beta) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \beta \mathbf{1}_n)^T(\mathbf{y} - \beta \mathbf{1}_n)\right)$$

- **Log-likelihood function**:

  $$\ell(\beta) = \ln L(\beta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\|\mathbf{y} - \beta \mathbf{1}_n\|_2^2$$

- **Maximize log-likelihood**: To maximize $\ell(\beta)$, we minimize the term $\|\mathbf{y} - \beta \mathbf{1}_n\|_2^2$ (since other terms are constant with respect to β).

  This is exactly the same optimization problem as for the LSE, so:

  $$\hat{\beta}_{\text{MLE}} = \arg\max \ell(\beta) = \arg\min \|\mathbf{y} - \beta \mathbf{1}_n\|_2^2 = \bar{y}$$

Thus, $\hat{\beta}_{\text{LSE}} = \hat{\beta}_{\text{MLE}} = \bar{y}$.

---

3. Question3

A sample of n boys and n girls is taken from a secondary school and their heights are measured.

Let $y_1, y_2, \cdots, y_n$ denote the heights of the n girls, and $y_{n+1}, y_{n+2}, \cdots, y_{2n}$ those of the n boys, respectively.

It is believed that the random quantities y_i satisfy $y_i = \alpha + \beta x_i + \varepsilon_i, \varepsilon_i \sim i.i.d.N(0, \sigma^2), i = 1, \cdots 2n$

where α is known, β is unknown and the covariates x_i are defined by [$x_i = -1$ for girls, $x_i = 1$ for boys]. Find the least squares estimators of β.

3.1 solution

To find the Least Squares Estimator (LSE) of β in the model $y_i = \alpha + \beta x_i + \varepsilon_i$ (where α is known, x_i takes values -1 for girls and 1 for boys, and there are $2n$ observations in total), we can proceed as follows:

The goal of least squares is to minimize the L_2-norm of residuals:

$$\|\mathbf{y} - \alpha \mathbf{1}_{2n} - \beta \mathbf{x}\|_2$$

Let $M(\beta) = \|\mathbf{y} - \alpha \mathbf{1}_{2n} - \beta \mathbf{x}\|_2^2$. Expanding this norm:

$$\begin{aligned} M(\beta) &= (\mathbf{y} - \alpha \mathbf{1}_{2n} - \beta \mathbf{x})^T (\mathbf{y} - \alpha \mathbf{1}_{2n} - \beta \mathbf{x}) \\ &= \mathbf{y}^T \mathbf{y} + \alpha^2 \|\mathbf{1}_{2n}\|_2^2 + \beta^2 \|\mathbf{x}\|_2^2 - 2\alpha \mathbf{1}_{2n}^T \mathbf{y} + 2\beta \mathbf{1}_{2n}^T \mathbf{x} - 2\beta \mathbf{x}^T \mathbf{y} \end{aligned}$$

Since $\|\mathbf{1}_{2n}\|_2^2 = 2n$ and $\|\mathbf{x}\|_2^2 = n(-1)^2 + n(1)^2 = 2n$, and $\mathbf{1}_{2n}^T \mathbf{x} = -n + n = 0$, the expression simplifies to:

$$M(\beta) = \beta^2 \cdot 2n - 2\beta \mathbf{x}^T \mathbf{y} + C$$

(where C is a constant term independent of β).

Treat $M(\beta)$ as a quadratic function in β (it opens upward, so the minimum exists at the vertex). Take the derivative and set it to zero:

$$\frac{dM(\beta)}{d\beta} = 4n\beta - 2\mathbf{x}^T \mathbf{y} = 0$$

Solve for β:

$$\hat{\beta} = \frac{\mathbf{x}^T \mathbf{y}}{2n}$$

Since $\mathbf{x}^T\mathbf{y} = -\sum_{i=1}^n y_i + \sum_{i=n+1}^{2n} y_i$, we get:

$$\hat{\beta} = \frac{-(y_1 + y_2 + \cdots + y_n) + (y_{n+1} + y_{n+2} + \cdots + y_{2n})}{2n}$$

Thus, the LSE of β is $\hat{\beta} = \frac{-\sum_{i=1}^n y_i + \sum_{i=n+1}^{2n} y_i}{2n}$.

---

4. Question 4

Consider a simple linear regression model where $\{x_i\}$ are constants, α and β are parameters, ε_i are iid $N(0, \sigma^2)$ and $i = 1, \cdots, n$. A weighted least square estimators for α and β are obtained by minimizing [weighted sum of squared residuals, with predefined known constant weights w_i satisfying $\sum_{i=1}^n w_i = 1$. Find the weighted least square estimators for α and β and their variance-covariance matrix.

4.1 solution:

To derive the weighted least squares estimators for α and β in the model $y_i = \alpha + \beta x_i + \varepsilon_i$ (with weights w_i and $\mathbf{D} = \text{diag}(w_1, w_2, \ldots, w_n)$), follow these steps:

We aim to minimize the **weighted sum of squared residuals**:

$$(\alpha, \beta) = \arg\min \sum_{i=1}^n w_i (y_i - \alpha - \beta x_i)^2$$

In vector form, this is equivalent to:

$$(\alpha, \beta) = \arg\min (\mathbf{y} - \alpha \mathbf{1}_n - \beta \mathbf{x})^T \mathbf{D} (\mathbf{y} - \alpha \mathbf{1}_n - \beta \mathbf{x})$$

where:

- $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$ (response vector),
- $\mathbf{1}_n$ is an n-dimensional vector of ones,
- $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ (covariate vector),
- $\mathbf{D}$ is a diagonal matrix with w_i on the diagonal.

Let $m(\alpha, \beta) = (\mathbf{y} - \alpha \mathbf{1}_n - \beta \mathbf{x})^T \mathbf{D} (\mathbf{y} - \alpha \mathbf{1}_n - \beta \mathbf{x})$ (weighted residual squared norm).

- **Derivative with respect to α:**

$$\frac{\partial m}{\partial \alpha} = -2\mathbf{1}_n^T\mathbf{D}(\mathbf{y} - \alpha\mathbf{1}_n - \beta\mathbf{x}) = 0 \tag{1}$$

- **Derivative with respect to β:**

$$\frac{\partial m}{\partial \beta} = -2\mathbf{x}^T\mathbf{D}(\mathbf{y} - \alpha\mathbf{1}_n - \beta\mathbf{x}) = 0 \tag{2}$$

Define $w^T = \mathbf{1}_n^T\mathbf{D}$ (weighted sum vector). From equation (1):

$$w^T\mathbf{y} - \alpha w^T\mathbf{1}_n - \beta w^T\mathbf{x} = 0 \implies \hat{\alpha} = w^T\mathbf{y} - \hat{\beta}w^T\mathbf{x} \tag{3}$$

Substitute (3) into equation (2). Note that $\mathbf{x}^T\mathbf{D}\mathbf{1}_n = w^T\mathbf{x}$, so:

$$\mathbf{x}^T\mathbf{D}\mathbf{y} - (w^T\mathbf{y} - \hat{\beta}w^T\mathbf{x})w^T\mathbf{x} - \hat{\beta}\mathbf{x}^T\mathbf{D}\mathbf{x} = 0$$

Simplify and solve for $\hat{\beta}$:

$$\hat{\beta} = \frac{\mathbf{x}^T\mathbf{D}\mathbf{y} - (w^T\mathbf{x})(w^T\mathbf{y})}{\mathbf{x}^T\mathbf{D}\mathbf{x} - (w^T\mathbf{x})^2}$$

Substitute $\hat{\beta}$ back into (3) to find $\hat{\alpha}$:

$$\hat{\alpha} = \frac{\mathbf{x}^T\mathbf{D}\mathbf{x} \cdot w^T\mathbf{y} - \mathbf{x}^T\mathbf{D}\mathbf{y} \cdot w^T\mathbf{x}}{\mathbf{x}^T\mathbf{D}\mathbf{x} - (w^T\mathbf{x})^2}$$

In matrix form, the estimators are:

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \frac{1}{\mathbf{x}^T\mathbf{D}\mathbf{x} - (w^T\mathbf{x})^2} \begin{bmatrix} \mathbf{x}^T\mathbf{D}\mathbf{x} \cdot w^T\mathbf{y} - \mathbf{x}^T\mathbf{D}\mathbf{y} \cdot w^T\mathbf{x} \\ \mathbf{x}^T\mathbf{D}\mathbf{y} - (w^T\mathbf{x})(w^T\mathbf{y}) \end{bmatrix}$$

Assume $\varepsilon \sim N(0, \sigma^2\mathbf{I}_n)$. Let $\begin{bmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \end{bmatrix} = \frac{1}{\mathbf{x}^T\mathbf{D}\mathbf{x} - (w^T\mathbf{x})^2} \begin{bmatrix} \mathbf{x}^T\mathbf{D}\mathbf{x}w^T - (w^T\mathbf{x})\mathbf{x}^T\mathbf{D} \\ -(w^T\mathbf{x})w^T \end{bmatrix} \varepsilon.$

The variance-covariance matrix is derived as:

$$\mathrm{Var}\left(\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}\right) = \sigma^2 \begin{bmatrix} \mathbf{x}^T\mathbf{D}\mathbf{x}w^T - (w^T\mathbf{x})\mathbf{x}^T\mathbf{D} \\ -(w^T\mathbf{x})w^T \end{bmatrix} \begin{bmatrix} \mathbf{x}^T\mathbf{D}\mathbf{x}w^T - (w^T\mathbf{x})\mathbf{x}^T\mathbf{D} \\ -(w^T\mathbf{x})w^T \end{bmatrix}^T \cdot \frac{1}{(\mathbf{x}^T\mathbf{D}\mathbf{x} - (w^T\mathbf{x})^2)^2}$$

5. Question 5

Consider a simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \varepsilon_i \sim i.i.d.N\left(0, \sigma^2\right), i = 1, 2, \cdots, n$$

and $\varepsilon_1, \cdots, \varepsilon_n$ are independent. We have three data points: $(-1, y_1)$, $(0, y_2)$, $(1, y_3)$.
(a) Express the least squares estimates of β_0 and β_1 in terms of y_1, y_2 and y_3.
(b) Find out the expression of the regression sum of squares in terms of y_1, y_2 and y_3. Suppose now we know that $y_3 = y_1 + 2$ and the total corrected sum of squares $S_{yy} = \sum_{i=1}^{3}(y_i - \bar{y})^2 = 2.5$.
(c) What is the value of the coefficient of determination r^2? Does the regression line fit the three data points well?
(d) Find the 95% confidence intervals for β_0 and β_1 respectively.

5.1 Part (a): Derive Least Squares Estimates

The model in vector form is $\mathbf{y} = \beta_0 \mathbf{1}_3 + \beta_1 \mathbf{x} + \varepsilon$, where:

- $\mathbf{y} = (y_1, y_2, y_3)^T$,
- $\mathbf{1}_3 = (1, 1, 1)^T$,
- $\mathbf{x} = (-1, 0, 1)^T$.

We minimize the L_2-norm of residuals:

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{\beta_0, \beta_1} \|\mathbf{y} - \beta_0 \mathbf{1}_3 - \beta_1 \mathbf{x}\|_2^2$$

Let $m(\beta_0, \beta_1) = \|\mathbf{y} - \beta_0 \mathbf{1}_3 - \beta_1 \mathbf{x}\|_2^2$.

- **Derivative with respect to β_0:**

$$\frac{\partial m}{\partial \beta_0} = -2(\mathbf{y} - \beta_0 \mathbf{1}_3 - \beta_1 \mathbf{x})^T \mathbf{1}_3 = 0$$

Let $\bar{y} = \frac{1}{3}\sum_{i=1}^{3} y_i$ and $\bar{x} = 0$ (since $x = [-1, 0, 1]$), simplifying gives:

$$\bar{y} - \beta_0 - \beta_1 \bar{x} = 0 \implies \beta_0 = \bar{y} \tag{1}$$

- **Derivative with respect to β_1:**

$$\frac{\partial m}{\partial \beta_1} = -2(\mathbf{y} - \beta_0 \mathbf{1}_3 - \beta_1 \mathbf{x})^T \mathbf{x} = 0$$

  Compute $\mathbf{x}^T \mathbf{y} = -y_1 + 0 \cdot y_2 + y_3 = y_3 - y_1$ and $\mathbf{x}^T \mathbf{x} = (-1)^2 + 0^2 + 1^2 = 2$, simplifying gives:

$$\mathbf{x}^T \mathbf{y} - \beta_0 \mathbf{x}^T \mathbf{1}_3 - \beta_1 \mathbf{x}^T \mathbf{x} = 0 \tag{2}$$

Substitute $\beta_0 = \bar{y}$ (from (1)) into (2):

$$(y_3 - y_1) - \bar{y} \cdot 0 - 2\beta_1 = 0 \implies \hat{\beta}_1 = \frac{y_3 - y_1}{2}$$

Then substitute $\hat{\beta}_1$ back into (1):

$$\hat{\beta}_0 = \bar{y} = \frac{y_1 + y_2 + y_3}{3}$$

- $\hat{\beta}_1 = \frac{y_3 - y_1}{2}$
- $\hat{\beta}_0 = \frac{y_1 + y_2 + y_3}{3}$

Thus, the least squares estimates are $\hat{\beta}_0 = \frac{y_1 + y_2 + y_3}{3}$ and $\hat{\beta}_1 = \frac{y_3 - y_1}{2}$.

5.2 Part (b): Regression Sum of Squares (RSS)

The regression sum of squares is defined as $RSS = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$, where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ are the fitted values.

- Compute fitted values:
  - $\hat{y}_1 = \hat{\beta}_0 - \hat{\beta}_1 = \frac{y_1 + y_2 + y_3}{3} - \frac{y_3 - y_1}{2} = \frac{5y_1 + 2y_2 - y_3}{6}$
  - $\hat{y}_2 = \hat{\beta}_0 = \frac{y_1 + y_2 + y_3}{3}$
  - $\hat{y}_3 = \hat{\beta}_0 + \hat{\beta}_1 = \frac{y_1 + y_2 + y_3}{3} + \frac{y_3 - y_1}{2} = \frac{-y_1 + 2y_2 + 5y_3}{6}$
- Calculate RSS:

$$
\begin{aligned}
RSS &= \left(\frac{5y_1+2y_2-y_3}{6}-\frac{y_1+y_2+y_3}{3}\right)^2+\left(\frac{y_1+y_2+y_3}{3}-\frac{y_1+y_2+y_3}{3}\right)^2+\left(\frac{-y_1+2y_2+5y_3}{6}-\frac{y_1+y_2+y_3}{3}\right)^2 \\
&= \left(\frac{3y_1+y_2-3y_3}{6}\right)^2+0+\left(\frac{-3y_1+y_2+3y_3}{6}\right)^2 \\
&= \frac{(y_1-y_3)^2}{2}
\end{aligned}
$$

5.3 Part (c): Coefficient of Determination r^2

The coefficient of determination is $r^2 = \frac{RSS}{S_{yy}}$, where $S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = 2.5$ (given).

- Substitute $RSS = \frac{(y_1-y_3)^2}{2}$ and $S_{yy} = 2.5$:

$$
r^2 = \frac{\frac{(y_1-y_3)^2}{2}}{2.5}
$$

Assuming $y_3 = y_1 + 2$ (implied by RSS structure), then $(y_1 - y_3)^2 = 4$:

$$
r^2 = \frac{\frac{4}{2}}{2.5} = \frac{2}{2.5} = 0.8
$$

5.4 Part (d):CI calculating

5.4.1 Expression and Expansion of $\hat{\beta}_0$

The least squares estimator of β_0 is:

$$
\hat{\beta}_0 = \bar{y} - \frac{\mathbf{x}^T\mathbf{y} - n\bar{x}\bar{y}}{\mathbf{x}^T\mathbf{x} - n\bar{x}^2}\bar{x}
$$

By substituting the model $\mathbf{y} = \beta_0 \mathbf{1}_n + \beta_1 \mathbf{x} + \varepsilon$, we derive:

$$
\begin{aligned}
\hat{\beta}_0 &= \frac{\mathbf{1}_n^T \mathbf{y}}{n} - \frac{\mathbf{x}^T \mathbf{y} - \bar{x}\mathbf{1}_n^T \mathbf{y}}{\mathbf{x}^T \mathbf{x} - n\bar{x}^2}\bar{x} \\
&= \frac{n\beta_0 + \beta_1 n\bar{x} + \mathbf{1}_n^T \varepsilon}{n} - \frac{n\beta_0\bar{x} + \beta_1 \mathbf{x}^T \mathbf{x} + \mathbf{x}^T \varepsilon - \bar{x}(n\beta_0 + \beta_1 n\bar{x} + \mathbf{1}_n^T \varepsilon)}{\mathbf{x}^T \mathbf{x} - n\bar{x}^2}\bar{x} \\
&= \beta_0 + \beta_1 \bar{x} + \frac{\mathbf{1}_n^T \varepsilon}{n} - \frac{\beta_1(\mathbf{x}^T \mathbf{x} - n\bar{x}^2) + (\mathbf{x}^T \varepsilon - \bar{x}\mathbf{1}_n^T \varepsilon)}{\mathbf{x}^T \mathbf{x} - n\bar{x}^2}\bar{x} \\
&= \beta_0 + \frac{\mathbf{1}_n^T \varepsilon}{n} - \frac{\bar{x}(\mathbf{x}^T \varepsilon - \bar{x}\mathbf{1}_n^T \varepsilon)}{\mathbf{x}^T \mathbf{x} - n\bar{x}^2} \\
&= \beta_0 + \frac{\mathbf{x}^T \mathbf{x}\mathbf{1}_n^T \varepsilon - n\bar{x}^2 \mathbf{1}_n^T \varepsilon}{n(\mathbf{x}^T \mathbf{x} - n\bar{x}^2)} \\
&= \beta_0 + \frac{\mathbf{x}^T \mathbf{x}\mathbf{1}_n^T - n\bar{x}^2 \mathbf{1}_n^T}{n(\mathbf{x}^T \mathbf{x} - n\bar{x}^2)}\varepsilon
\end{aligned}
$$

5.4.2 Variance and Confidence Interval of $\hat{\beta}_0$

- **Variance**:

$$
\mathrm{Var}(\hat{\beta}_0) = \sigma^2 \frac{\mathbf{x}^T \mathbf{x}}{n(\mathbf{x}^T \mathbf{x} - n\bar{x}^2)}
$$

- **Confidence Interval Derivation**:
  Since $\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\frac{s^2 \mathbf{x}^T \mathbf{x}}{n(\mathbf{x}^T \mathbf{x} - n\bar{x}^2)}}} \sim t_{n-2},\ s^2 = \frac{\mathrm{MSE}}{n-2} = 0.5, \mathbf{x}^T \mathbf{x} = 2, n = 3$, and $\bar{x} = 0$, the standard error is:

$$
\mathrm{SE}(\hat{\beta}_0) = \sqrt{\frac{s^2 \mathbf{x}^T \mathbf{x}}{n(\mathbf{x}^T \mathbf{x} - n\bar{x}^2)}} = \sqrt{\frac{0.5 \times 2}{3 \times 2}} = \frac{1}{\sqrt{6}} \approx 0.408
$$

From the t-distribution table, $t_{0.025}(1) = 12.706$. Thus, the 95% confidence interval for β_0 is:

$$
\left[\frac{y_1 + y_2 + y_3}{3} - 12.706 \times 0.408, \frac{y_1 + y_2 + y_3}{3} + 12.706 \times 0.408\right]
$$

5.4.3 Expression and Expansion of $\hat{\beta}_1$

The least squares estimator of β_1 is:

$$
\hat{\beta}_1 = \frac{\mathbf{x}^T \mathbf{y} - n\bar{x}\bar{y}}{\mathbf{x}^T \mathbf{x} - n\bar{x}^2}
$$

By substituting the model $\mathbf{y} = \beta_0 \mathbf{1}_n + \beta_1 \mathbf{x} + \varepsilon$, we derive:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\mathbf{x}^T(\beta_0 \mathbf{1}_n + \beta_1 \mathbf{x} + \varepsilon) - \bar{x}\mathbf{1}_n^T(\beta_0 \mathbf{1}_n + \beta_1 \mathbf{x} + \varepsilon)}{\mathbf{x}^T\mathbf{x} - n\bar{x}^2} \\
&= \frac{n\beta_0\bar{x} + \beta_1 \mathbf{x}^T\mathbf{x} + \mathbf{x}^T\varepsilon - n\beta_0\bar{x} - \beta_1 n\bar{x}^2 - \bar{x}\mathbf{1}_n^T\varepsilon}{\mathbf{x}^T\mathbf{x} - n\bar{x}^2} \\
&= \frac{\beta_1(\mathbf{x}^T\mathbf{x} - n\bar{x}^2) + (\mathbf{x}^T\varepsilon - \bar{x}\mathbf{1}_n^T\varepsilon)}{\mathbf{x}^T\mathbf{x} - n\bar{x}^2} \\
&= \beta_1 + \frac{\mathbf{x}^T\varepsilon - \bar{x}\mathbf{1}_n^T\varepsilon}{\mathbf{x}^T\mathbf{x} - n\bar{x}^2}
\end{aligned}$$

5.4.4 Variance and Confidence Interval of $\hat{\beta}_1$

- **Variance**:

$$\mathrm{Var}(\hat{\beta}_1) = \sigma^2 \frac{1}{\mathbf{x}^T\mathbf{x} - n\bar{x}^2}$$

- **Confidence Interval Derivation**:
  Since $\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{s^2}{\mathbf{x}^T\mathbf{x} - n\bar{x}^2}}} \sim t_{n-2}$, $s^2 = 0.5$, $\mathbf{x}^T\mathbf{x} - n\bar{x}^2 = 2$, and $t_{0.025}(1) = 12.706$, the standard error is:

$$\mathrm{SE}(\hat{\beta}_1) = \sqrt{\frac{s^2}{\mathbf{x}^T\mathbf{x} - n\bar{x}^2}} = \sqrt{\frac{0.5}{2}} = 0.5$$

  Thus, the 95% confidence interval for β_1 is:

$$\left[\hat{\beta}_1 - 12.706 \times 0.5, \hat{\beta}_1 + 12.706 \times 0.5\right] = \left[\frac{y_3 - y_1}{2} - 6.353, \frac{y_3 - y_1}{2} + 6.353\right]$$

  Assuming $\hat{\beta}_1 = 1$, the interval becomes $[-5.353, 7.353]$.