MLNAOE_HW2

Habin Lee

October 2024

Back ground knowledge

Likelihood function A likelihood function measures how well a statistical model explains observed data by calculating the probability of seeing that data under different parameter values of the model.

Given a probability density or mass function:

$$x \to f(x|\theta)$$

where x is realization of the random variable X, the likehood function is

$$\theta \to f(x|\theta) \ or \ \mathcal{L}(\theta|x)$$

$\mathcal{L}(\theta|x)$ is a function of θ with x is fixed. When X is a discrete and random variable with probability mass function p depending on a parameter θ.

$$\mathcal{L}(\theta|x) = p_\theta(x) = P_\theta(X = x)$$

To calculate this likelihood function numerically, we should use joint probability density function like below.

$$P(x|\theta) = \prod_{k=1}^{n} P(x_k|\theta) = P(x_1|\theta) \times P(x_2|\theta) \times P(x_3|\theta) \times ... \times P(x_n|\theta)$$

1 Problem 1

1.1 (a) Derivative Using Differentiation

Step 1: Set up the optimization problem We need to minimize the sum of squared errors

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

Step 2: Compute partial derivatives Let's differentiate S with respect to β_0 and β_1

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

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

Step 3: Set partial derivatives to zero (normal equations)

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

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

Step 4: Solve the normal equations

From the first equation($\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0$)

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

From the second equation($\sum_{i=1}^{n} x_i(y_i - \beta_0 - \beta_1 x_i) = 0$)

$$\beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

Solving these simultaneously, we get:

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

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

Where $\bar{x}$ and $\bar{y}$ are the sample means of x($\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$)and y ($\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$) respectively.

1.2 (b) Derivation Using Maximum Likelihood Estimation

Step 1: Write down the likelihood function Assuming $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, the likelihood function is:

$$L(\beta_0, \beta_1, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right)$$

To reduce complexity of multiplication calculation, We can change multiplication to addition by putting logs on both sides.

$$\ln(L(\beta_0, \beta_1, \sigma^2)) = \ell(\beta_0, \beta_1, \sigma^2) = \ln\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right)\right)$$

Step 2: Derive the log-likelihood function

$$\ell(\beta_0, \beta_1, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

Step 3: Compute partial derivatives

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

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

Step 4: Set partial derivatives to zero

These equations are equivalent to the normal equations from part (a), just multiplied by $\frac{1}{\sigma^2}$.

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

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

Step 5: Show equivalence to OLS estimators
Since the equations are the same as in part (a), the solutions will be identical. Therefore, the MLEs

of β_0 and β_1 are equivalent to the OLS estimators.

From the first equation($\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0$)

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

From the second equation($\sum_{i=1}^{n} x_i(y_i - \beta_0 - \beta_1 x_i) = 0$)

$$\beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

Solving these simultaneously, we get:

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

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

Where $\bar{x}$ and $\bar{y}$ are the sample means of x($\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$)and y ($\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$) respectively.

2 Problem 2: Weighted Least Squares Regression

2.1 (a) Derivation Using Differentiation

Step 1: Set up the weighted optimization problem
We need to minimize(where $w_i = \frac{1}{\sigma_i^2}$.):

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

Step 2: Compute partial derivatives

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

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

Step 3: Set partial derivatives to zero (weighted normal equations)

$$\sum_{i=1}^{n} w_i(y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\sum_{i=1}^{n} w_i x_i(y_i - \beta_0 - \beta_1 x_i) = 0$$

Step 4: Solve the weighted normal equations
From these equations, we can derive:

$$\hat{\beta}_1 = \frac{\sum w_i \sum w_i x_i y_i - \sum w_i x_i \sum w_i y_i}{\sum w_i \sum w_i x_i^2 - (\sum w_i x_i)^2}$$

$$\hat{\beta}_0 = \frac{\sum w_i y_i}{\sum w_i} - \hat{\beta}_1 \frac{\sum w_i x_i}{\sum w_i}$$

2.2 (b) Derivation Using Maximum Likelihood Estimation

Step 1: Write down the likelihood function

$$L(\beta_0, \beta_1, \sigma_1^2, ..., \sigma_n^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma_i^2}\right)$$

Step 2: Derive the log-likelihood function

$$\ell(\beta_0, \beta_1) = -\frac{1}{2}\sum_{i=1}^{n} \log(2\pi\sigma_i^2) - \frac{1}{2}\sum_{i=1}^{n} \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{\sigma_i^2}$$

Step 3: Compute partial derivatives

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

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

Step 4: Set partial derivatives to zero

These equations are equivalent to the weighted normal equations from part (a), with $w_i = \frac{1}{\sigma_i^2}$.

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

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

From these equations, we can derive:

$$\hat{\beta}_1 = \frac{\sum w_i \sum w_i x_i y_i - \sum w_i x_i \sum w_i y_i}{\sum w_i \sum w_i x_i^2 - (\sum w_i x_i)^2}$$

$$\hat{\beta}_0 = \frac{\sum w_i y_i}{\sum w_i} - \hat{\beta}_1 \frac{\sum w_i x_i}{\sum w_i}$$

Step 5: Show equivalence to WLS estimators

Since the equations are the same as in part (a), the solutions will be identical. Therefore, the MLEs of β_0 and β_1 correspond to the WLS estimators.

Comparison and Discussion

1. In both OLS and WLS, the estimators derived using differentiation and MLE are equivalent. This demonstrates the consistency between these two approaches.

2. OLS is appropriate when the error terms have constant variance (homoscedasticity). WLS is more suitable when there's heteroscedasticity, as it gives more weight to observations with lower variance.

3. Assuming normality of error terms in MLE allows for statistical inference and hypothesis testing. However, this assumption may not always hold in practice, potentially affecting the reliability of these tests.