Lecture 11 Multiple Linear Regression

Lecturer:Xiangyu Chang *Scribe: Lingtao Ouyang, Yuxi Wu*

Edited by: Zhihong Liu

1 Recall of SLR

For the observed values $\{(X_i, Y_i)\}_{i=1}^n$, we can have **the simple linear regression model** among parameters.

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i. \tag{1}$$

where ϵ_i satisfies $\mathbb{E}[\epsilon_i] = 0$ and $\mathrm{Var}(\epsilon_i) = \sigma^2$.

The least squares estimates (LS) of the regression coefficients are:

$$\begin{cases} \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}, \\ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \end{cases}$$

where $S_{xy} = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ is the sample covariance and $S_{xx} = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$.

1.1 Expectations and variance properties

$$\mathbb{E}[\hat{\beta}_0] = \beta_0. \tag{2}$$

$$\mathbb{E}[\hat{\beta}_1] = \beta_1 \Rightarrow \mathbb{E}[S_{xy}] = \mathbb{E}[\hat{\beta}_1 S_{xx}] = \beta_1 S_{xx}. \tag{3}$$

$$\mathrm{Var}(\hat{\beta}_1) = \frac{1}{n^2 S_{xx}^2} \sum_i (x_i - \bar{x})^2 \mathrm{Var}(y_i) = \frac{\sigma^2}{n S_{xx}}. \tag{4}$$

$$\mathrm{Var}(\hat{\beta}_0) = \left(\frac{1}{n} + \frac{\bar{x}^2}{n S_{xx}}\right)\sigma^2 = \frac{S_{xx} + \bar{x}^2}{n S_{xx}}\sigma^2 = \frac{\sum x_i^2}{n^2 S_{xx}}\sigma^2. \tag{5}$$

where $S_{xx} + \bar{x}^2 = \frac{1}{n}\sum x_i^2$.

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{n S_{xx}} = \frac{\mathrm{Var}(S_{xy})}{\sigma^2 S_{xx}} \Rightarrow \mathrm{Var}(S_{xy}) = \frac{\sigma^2 S_{xx}}{n}. \tag{6}$$

1.2 For noise ϵ in SLR

Theorem 1. *Let $\hat{\epsilon}_i = y_i - \hat{y}_i$. The estimator*

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

is an unbiased estimator of σ^2, i.e., $\mathbb{E}(\hat{\sigma}^2) = \sigma^2$.

Proof. The fitted values in the regression model are given by:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i = \bar{y}_n - \hat{\beta}_1 \bar{x}_n + \hat{\beta}_1 x_i.$$

The residual sum of squares can be expanded as:

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}\left[(y_i - \bar{y}_n) - \hat{\beta}_1(x_i - \bar{x}_n)\right]^2.$$

Expanding this expression, we have:

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \underbrace{\sum_{i=1}^{n}(y_i - \bar{y}_n)^2}_{\mathrm{T}_1} + \underbrace{\hat{\beta}_1^2\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}_{\mathrm{T}_2} - \underbrace{2\hat{\beta}_1\sum_{i=1}^{n}(y_i - \bar{y}_n)(x_i - \bar{x}_n)}_{\mathrm{T}_3}.$$

First, we compute $\mathbb{E}(\mathrm{T}_2)$ and $\mathbb{E}(\mathrm{T}_3)$:

$$\mathbb{E}(\mathrm{T}_2) = nS_{xx}\mathbb{E}(\hat{\beta}_1^2) = nS_{xx}\cdot\frac{\mathbb{E}(S_{xy}^2)}{S_{xx}^2} = \frac{n\mathbb{E}(S_{xy}^2)}{S_{xx}}.$$

$$\mathbb{E}(\mathrm{T}_3) = 2n\mathbb{E}\left[\hat{\beta}_1 S_{xy}\right] = 2n\mathbb{E}\left[\frac{S_{xy}}{S_{xx}}\right] = \frac{2n}{S_{xx}}\mathbb{E}(S_{xy}^2).$$

Therefore:

$$\mathbb{E}(\mathrm{T}_2) - \mathbb{E}(\mathrm{T}_3) = -\frac{n}{S_{xx}}\left[\mathrm{Var}(S_{xy}) + \mathbb{E}(S_{xy})^2\right].$$

Substituting $\mathrm{Var}(S_{xy}) = \frac{\sigma^2 S_{xx}}{n}$ and $\mathbb{E}(S_{xy}) = \beta_1 S_{xx}$, we obtain:

$$\mathbb{E}(\mathrm{T}_2) - \mathbb{E}(\mathrm{T}_3) = -\sigma^2 - n\beta_1^2 S_{xx}.$$

Next, we compute $\mathbb{E}(\mathrm{T}_1)$. The regression model can be expressed as:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \bar{y}_n = \beta_0 + \beta_1\bar{x}_n + \bar{\epsilon}_n.$$

Therefore:

$$\mathbb{E}(y_i - \bar{y}_n)^2 = \mathbb{E}\left[\beta_1(x_i - \bar{x}_n) + (\epsilon_i - \bar{\epsilon}_n)\right]^2.$$

Expanding this expression, we get:

$$\mathbb{E}(y_i - \bar{y}_n)^2 = \beta_1^2(x_i - \bar{x}_n)^2 + \mathbb{E}(\epsilon_i - \bar{\epsilon}_n)^2 + 2\beta_1(x_i - \bar{x}_n)\mathbb{E}(\epsilon_i - \bar{\epsilon}_n).$$

where:

$$\mathbb{E}(\epsilon_i - \bar{\epsilon}_n)^2 = \mathbb{E}\left[\epsilon_i^2 + \bar{\epsilon}_n^2 - 2\epsilon_i\bar{\epsilon}_n\right].$$

Since $\mathrm{Var}(\bar{\epsilon}_n) = \frac{\sigma^2}{n}$ and $\mathbb{E}(\epsilon_i\bar{\epsilon}_n) = \frac{\sigma^2}{n}$, it follows that:

$$\mathbb{E}(\epsilon_i - \bar{\epsilon}_n)^2 = \frac{n-1}{n}\sigma^2.$$

Hence:

$$\mathbb{E}(\mathrm{T}_1) = \sum_{i=1}^{n}\mathbb{E}(y_i - \bar{y}_n)^2 = n\beta_1^2 S_{xx} + (n-1)\sigma^2.$$

$$\mathbb{E}\left(\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right) = \mathbb{E}(\mathrm{T}_1) + \mathbb{E}(\mathrm{T}_2) - \mathbb{E}(\mathrm{T}_3)$$

$$= n\beta_1^2 S_{xx} + (n-1)\sigma^2 - \sigma^2 - n\beta_1^2 S_{xx} = (n-2)\sigma^2.$$

Therefore:

$$\mathbb{E}(\hat{\sigma}^2) = \frac{1}{n-2}\mathbb{E}\left(\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right) = \sigma^2.$$

□

1.3 Maximum Likelihood Estimation and Least Squares Equivalence in Simple Linear Regression

Consider the simple linear regression model:

$$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

where y_i is the response variable, X_i is the predictor variable, β_0 and β_1 are unknown parameters, and ε_i are independent error terms following a normal distribution $\mathcal{N}(0, \sigma^2)$. Under this model, the conditional distribution of y_i given X_i is:

$$y_i \mid X_i \sim \mathcal{N}(\beta_0 + \beta_1 X_i, \sigma^2).$$

The pdf of y_i given X_i is:

$$f_{\beta_0,\beta_1}(y_i \mid X_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 X_i)^2}{2\sigma^2}\right).$$

The log-likelihood function for the parameters β_0 and β_1 is obtained as follows:

$$\ell(\beta_0, \beta_1) = \sum_{i=1}^{n} \log f_{\beta_0,\beta_1}(y_i \mid X_i).$$

Substituting the pdf into the log-likelihood function, we have:

$$\ell(\beta_0, \beta_1) = \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 X_i)^2}{2\sigma^2}\right)\right).$$

Simplifying the expression:

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

1.3.1 Equivalence of MLE and Least Squares

To find the maximum likelihood estimates (MLEs) of β_0 and β_1, we need to maximize the log-likelihood function with respect to these parameters. Since the term $-\frac{n}{2}\log(2\pi\sigma^2)$ is a constant with respect to β_0 and β_1, maximizing $\ell(\beta_0, \beta_1)$ is equivalent to minimizing the residual sum of squares:

$$\max_{\beta_0,\beta_1} \ell(\beta_0, \beta_1) \Leftrightarrow \min_{\beta_0,\beta_1} \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 X_i)^2.$$

This minimization problem is precisely the objective of the least squares (LS) method. Therefore, under the normality assumption of the error terms, the MLEs of β_0 and β_1 coincide with the LS estimates:

$$\hat{\beta}_0^{\mathrm{MLE}} = \hat{\beta}_0^{\mathrm{LS}}, \quad \hat{\beta}_1^{\mathrm{MLE}} = \hat{\beta}_1^{\mathrm{LS}}. \tag{7}$$

1.3.2 Distributional Properties of $\hat{\beta}_1$

The LS estimate of the slope parameter β_1 is given by:

$$\hat{\beta}_1 = \frac{1}{n}\sum_i (x_i - \hat{x})y_i,$$

Under the normality assumption of the error terms, $\hat{\beta}_1$ follows a normal distribution with mean β_1 and variance $\frac{\sigma^2}{S_{XX}}$:

$$\hat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2}{nS_{XX}}\right). \tag{8}$$

1.3.3 Standardization of $\hat{\beta}_1$

To conduct hypothesis testing for β_1, we standardize $\hat{\beta}_1$ by subtracting its expected value, multiplying by $\sqrt{S_{XX}}$, and dividing by σ. The standardized variable follows a standard normal distribution:

$$\frac{\sqrt{nS_{XX}}(\hat{\beta}_1 - \beta_1)}{\sigma} \sim \mathcal{N}(0, 1). \tag{9}$$

This result is fundamental for constructing confidence intervals and performing significance tests for the slope parameter β_1.

1.4 Regression Model Fit Evaluation

In statistics and econometrics, TSS (Total Sum of Squares), ESS (Explained Sum of Squares), and RSS (Residual Sum of Squares) are important metrics for measuring the fit of a regression model. Their relationship can be expressed by the formula:

$$TSS = ESS + RSS.$$

1.5 Definitions and Explanations

- **TSS** : The total sum of squares, representing the total variation between the observed values and their mean.
$$TSS = \sum (y_i - \bar{y})^2.$$
- **ESS** : The explained sum of squares, representing the part of the variation that is explained by the regression model.
$$ESS = \sum (\hat{y}_i - \bar{y})^2.$$
- **RSS** : The residual sum of squares, representing the part of the variation that is not explained by the regression model.
$$RSS = \sum (y_i - \hat{y}_i)^2.$$

Specifically, TSS can be decomposed into two parts: one part is ESS, indicating the variation in y caused by changes in the independent variable x; the other part is RSS, indicating the variation in y caused by other factors besides the linear influence of x on y.

Table 1: Degree of Freedom and Distribution

Sum of Square	**Degree of Freedom**	**Distribution**
RSS	1	$\sigma^2\chi^2(1)$
ESS	$n-2$	$\sigma^2\chi^2(n-2)$
TSS	$n-1$	$\sigma^2\chi^2(n-1)$

Coefficient of Determination R^2

The coefficient of determination is defined as:

$$0 \le R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \le 1. \tag{10}$$

where $TSS = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$ is the total sum of squares. R^2 represents the proportion of variance explained by the model, ranging between $[0, 1]$.

$$\frac{\frac{\sqrt{nS_{xx}}(\hat{\beta}_1 - \beta_1)}{\sigma}}{\sqrt{\frac{ESS}{\sigma^2}}} \Rightarrow \frac{\sqrt{nS_{xx}}(\hat{\beta}_1 - \beta_1)}{\sqrt{ESS}} \sim t(n-2). \tag{11}$$

2 Multiple Linear Regression Model

Consider the multiple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i.$$

or in matrix form:

$$\mathbf{Y_{n\times 1}} = \mathbf{X_{n\times p}}\beta_{\mathbf{p\times 1}} + \epsilon_{\mathbf{n\times 1}}. \tag{12}$$

where $\mathbf{X}$ is an $n \times p$ design matrix, β is a $p \times 1$ vector of regression coefficients, and ϵ is an $n \times 1$ error vector satisfying $\mathbb{E}[\epsilon] = \mathbf{0}$ and $\mathrm{Var}(\epsilon) = \sigma^2\mathbf{I}$.

To find the least squares (LS) estimator $\hat{\beta}$, we start by defining the objective function:

$$L(\beta) = \frac{1}{2}\|\mathbf{Y} - \mathbf{X}\beta\|^2.$$

Taking the gradient of $L(\beta)$ with respect to β and setting it to zero yields the normal equations:

$$\nabla(L(\beta)) = -\mathbf{X^T}(\mathbf{Y} - \mathbf{X}\beta) = 0.$$

Assuming that $\mathbf{X^T X}$ is invertible (i.e., the columns of $\mathbf{X}$ are linearly independent), we can solve for β:

$$\hat{\beta} = (\mathbf{X^T X})^{-1}\mathbf{X^T Y}. \tag{13}$$

This is the LS estimator for β.

2.1 Statistical Properties of the Estimator

2.1.1 Unbiasedness

The LS estimator $\hat{\beta}$ is unbiased. This can be shown by taking the expectation of $\hat{\beta}$:

$$\mathbb{E}[\hat{\beta}] = \mathbb{E}[(\mathbf{X^T X})^{-1}\mathbf{X^T Y}].$$

Substituting $\mathbf{Y} = \mathbf{X}\beta + \epsilon$, we get:

$$\mathbb{E}[\hat{\beta}] = (\mathbf{X^T X})^{-1}\mathbf{X^T}\mathbb{E}[\mathbf{X}\beta + \epsilon] = (\mathbf{X^T X})^{-1}\mathbf{X^T X}\beta = \beta.$$

Thus, $\hat{\beta}$ is an unbiased estimator of β.

2.1.2 Variance

The variance of $\hat{\beta}$ is derived as follows:

$$\mathrm{Var}(\hat{\beta}) = \mathrm{Var}[(\mathbf{X^T X})^{-1}\mathbf{X^T Y}].$$

Substituting $\mathbf{Y} = \mathbf{X}\beta + \epsilon$, we have:

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

Since $\mathrm{Var}(\mathbf{Y}) = \mathrm{Var}(\mathbf{X}\beta + \epsilon) = \sigma^2\mathbf{I}$, this simplifies to:

$$\mathrm{Var}(\hat{\beta}) = (\mathbf{X^T X})^{-1}\mathbf{X^T}(\sigma^2\mathbf{I})\mathbf{X}(\mathbf{X^T X})^{-1} = \sigma^2(\mathbf{X^T X})^{-1}.$$

The least squares (LS) estimator $\hat{\beta}$ has the following properties:

$$\mathbb{E}[\hat{\beta}] = \beta. \tag{14}$$

$$\mathrm{Var}(\hat{\beta}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}. \tag{15}$$

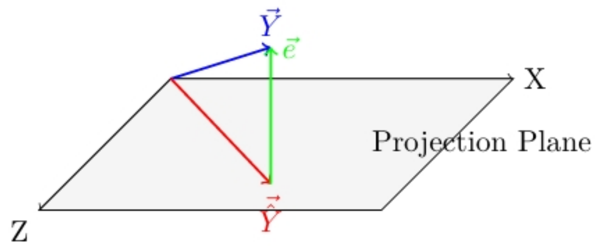

Figure 1: Projection Illustration

2.2 Projection and Error Vectors

2.2.1 Projection Matrix

The predicted values $\hat{\mathbf{Y}}$ and residuals $\mathbf{e}$ are:

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X^T X})^{-1}\mathbf{X^T Y} = \mathbf{PY} \quad \text{(P is called projection matrix)}$$

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} \quad \text{where} \quad \mathbf{e} \cdot \mathbf{Y} = 0. \tag{16}$$

2.2.2 Eigenvalues of Projection Matrix

Theorem 2. *A projection matrix* $\mathbf{P}$ *satisfies:*

$$\mathbf{P^T} = \mathbf{P} \quad \textit{(Symmetric)}, \quad \mathbf{P^2} = \mathbf{P} \quad \textit{(Idempotent)}.$$

The eigenvalues of $\mathbf{P}$ *are either 0 or 1.*

Proof. Let $\mathbf{Px} = \lambda\mathbf{x}$. Using the idempotent property:

$$\mathbf{P}^2\mathbf{x} = \mathbf{Px} \implies \lambda^2\mathbf{x} = \lambda\mathbf{x} \implies \lambda(\lambda - 1) = 0 \implies \lambda = 0 \text{ or } \lambda = 1.$$

□

2.2.3 Unbiased Estimation of σ^2

Theorem 3. *The residual sum of squares* $\mathbf{e}^T\mathbf{e}$ *divided by* $n - p$ *is an unbiased estimator of* σ^2*:*

$$\mathbb{E}\left[\frac{\mathbf{e}^T\mathbf{e}}{n-p}\right] = \sigma^2. \tag{17}$$

Proof. Let $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{P})\mathbf{Y}$, where $\mathbf{P} = \mathbf{X}(\mathbf{X^T X})^{-1}\mathbf{X^T}$ is the projection matrix. The residual sum of squares is:

$$\mathbf{e}^T\mathbf{e} = \mathbf{Y}^T(\mathbf{I} - \mathbf{P})^T(\mathbf{I} - \mathbf{P})\mathbf{Y} = \mathbf{Y}^T(\mathbf{I} - \mathbf{P})\mathbf{Y}.$$

Substituting $\mathbf{Y} = \mathbf{X}\beta + \epsilon$ into the expression:

$$\mathbf{e}^T\mathbf{e} = \epsilon^T(\mathbf{I} - \mathbf{P})\epsilon.$$

Taking the expectation:

$$\mathbb{E}[\mathbf{e}^T\mathbf{e}] = \sigma^2\text{tr}(\mathbf{I} - \mathbf{P}) = \sigma^2(n - p).$$

Thus:

$$\mathbb{E}\left[\frac{\mathbf{e}^T\mathbf{e}}{n-p}\right] = \sigma^2.$$

□