Statistical Inference and Multivariate Analysis (MA324)

Lecture Slides
Lecture 26

Simple Linear Regression

Indian Institute of Technology Guwahati

Jan-May 2023

Simple Linear Regression

- Just one predictor x, i.e. $p = 1$, where p indicates the number of predictors in a multiple linear regression model.

- The model for the **simple linear regression** is given by

$$y = \beta_0 + \beta_1 x + \epsilon,$$

  where y is the **outcome** variable (random), x is the **independent/predictor** variable (non-random) and ϵ is the **random error term**. β_0 (intercept) and β_1 (slope) are model parameters (unknown constants).

- Equivalently, the **model can be written** for $i = 1, 2, \cdots, n$ number of observations $(x_1, y_1), \cdots, (x_n, y_n)$ as

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

- How do you **interpret** β_0 (intercept) and β_1 (slope)?

Assumptions:

1. The **errors** (ϵ_i) are **uncorrelated** with each other.

2. The **expected value of the errors is zero** ($E(\epsilon_i) = 0$ for all i).

3. The assumption of ***homoscedasticity (constant variance)***:

   $Var(Y|X = x) = \sigma^2$

4. Errors (ϵ_i) are **normally distributed**

- The first three assumptions are required to estimate the model parameters. They mean:
  The ***errors*** ϵ_i**'s are uncorrelated** with

$$\begin{aligned} E(\epsilon_i|x_i) &= E(\epsilon_i) = 0, i = 1, \ldots, n \\ Var(\epsilon_i|x_i) &= Var(\epsilon_i) = \sigma^2, i = 1, \ldots, n \end{aligned}$$

Assumptions: What they means?

1. The **errors** (ϵ_i) are **uncorrelated** with each other. This violation most often occurs in data that are ordered in time (time series data), where errors that are near each other in time are often similar to each other (such time-related correlation is called autocorrelation). **Violation of this assumption** can lead to very **misleading assessments** of the strength of the regression.

2. The **expected value of the errors is zero** ($E(\epsilon_i) = 0$ for all i). That is, it cannot be true that for certain observations the model is systematically too low, while for others it is systematically too high.

3. The assumption of ***homoscedasticity (constant variance)***: each outcome observation (Y) has exact same variance (also for error). If Y follows Poisson distribution, this assumption is violated. Why?

4. The **assumption of *normality* of errors is NOT necessary for estimation**; however it is **necessary for inference** (e.g. testing of hypothesis, confidence intervals).

Least Squares Estimation

- Goal: To **estimate** β_0, β_1 **by minimizing error** in some sense (e.g. squared error)
- One reasonable way is to use the ***principle of Least Squares*, i.e. minimize the objective function**

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

  with respect to β_0, β_1.
- **Differentiate** $Q(\beta_0, \beta_1)$ with respect to β_0, β_1 and **equate the partial derivatives to zero** to get the estimates $\hat{\beta}_0, \hat{\beta}_1$.
- The resulting equations are called ***normal equations***:

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

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

Least Squares Estimation

- The **solution is given by**

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$
$$\hat{\beta}_1 = S_{xy}/S_{xx}$$

where

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \text{ and } S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}).$$

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are called the **least squares estimator (LSE)** of β_0 and β_1 respectively.

Properties of $\hat{\beta}_0$ and $\hat{\beta}_1$

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are linear combination of y_i's.
- $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased for β_0 and β_1, respectively.
- Variances of $\hat{\beta}_0$ and $\hat{\beta}_1$:

$$\begin{aligned} Var(\hat{\beta}_0) &= \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right) \\ Var(\hat{\beta}_1) &= \frac{\sigma^2}{S_{xx}} \end{aligned}$$

- Do the least square based estimation problem different from the likelihood based estimation procedure?

- **Geometrically** speaking, **minimizing** $\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$ this is **equivalent to minimizing the sum of squared vertical distances** between the actual y-values and any regression line to be fitted.

Linear Estimator

Def: [**Linear Estimator**] An estimator $\hat{\theta}$ is called linear estimator of θ if $\hat{\theta}$ is a **linear combination** of random observations.

Def: [**Linear Estimator**] An estimator $\hat{\theta}$ is called the best linear unbiased estimator (**BLUE**) of θ if $\hat{\theta}$ is **linear and unbiased estimator** of θ and $\hat{\theta}$ has **minimum variance among** all linear unbiased estimator of θ.

Gauss-Markov Theorem

Theorem

Under the first three assumptions of linear regression, the LSEs $\hat{\beta}_0$ and $\hat{\beta}_1$ are BLUE of β_0 and β_1, respectively.

A Few Definitions

- ***Fitted values***: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i,\ i = 1, \ldots, n$
- ***Residuals*: difference between observed and fitted values** of y, i.e.

$$e_i = y_i - \hat{y}_i,\ i = 1, \ldots, n$$

- The objective function evaluated at the LSE, i.e. $Q(\hat{\beta}_0, \hat{\beta}_1)$ is called the ***residual sum of squares* (SS_{Res}).**

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

Properties of Least Squares Fit:

1. $\sum_{i=1}^{n} e_i = 0$. (from the first normal equation)
2. $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$. (using (1))
3. $\sum_{i=1}^{n} x_i e_i = 0$.
4. $\sum_{i=1}^{n} \hat{y}_i e_i = 0$. (using (1) and (3))

Estimation of Error Variance

- It can be shown that
$$E(SS_{Res}) = (n-2)\sigma^2$$

- Hence, $\hat{\sigma}^2 = \frac{SS_{Res}}{n-2} = MS_{Res}$ is an unbiased estimator of σ^2. Here MS_{Res} is residual mean square.

- Observed value of $\hat{\sigma}^2 = \frac{SS_{Res}}{n-2}$ is called ***Residual variance***. It's **square root is called *residual standard error*.**

- Note that SS_{Res} depends on the model. Therefore, any violation of model assumptions has serious effect on $\hat{\sigma}^2$.

- A convenient computing formula for SS_{Res} is
$$SS_{Res} = SS_T - \hat{\beta}_1 S_{xy},$$
where
$$SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2.$$