Linear Regression Models

Segment 1 – Simple Linear Regression Model

Topic 3 – Ordinary Least Squares (OLS) Estimators and Estimates

Sudarsan N.S. Acharya (sudarsan.acharya@manipal.edu)

Topics

1. Estimator and Estimate
2. Interpreting OLS Estimators for a Continuous Predictor
3. Interpreting OLS Estimators for a Categorical Predictor
4. Properties of OLS estimators
5. Prediction Problem
6. Prediction Error
7. Summary

Estimator and Estimate

Estimator and Estimate

- Suppose that we want to build an SLRM for response Y and a single predictor predictor X_1:

Estimator and Estimate

- Suppose that we want to build an SLRM for response Y and a single predictor predictor X_1: $\hat{Y} = \beta_0 + \beta_1 X_1$.

Estimator and Estimate

- Suppose that we want to build an SLRM for response Y and a single predictor predictor X_1: $\hat{Y} = \beta_0 + \beta_1 X_1$.
- To that end, we want to use a randomly sampled dataset of size n

Estimator and Estimate

- Suppose that we want to build an SLRM for response Y and a single predictor predictor X_1: $\hat{Y} = \beta_0 + \beta_1 X_1$.
- To that end, we want to use a randomly sampled dataset of size n *but the samples are not identified yet.*

Estimator and Estimate

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- Suppose that we want to build an SLRM for response Y and a single predictor predictor X_1: $\hat{Y} = \beta_0 + \beta_1 X_1$.
- To that end, we want to use a randomly sampled dataset of size n *but the samples are not identified yet.*
- Recall the OLS estimators of β_0 and β_1:

Estimator and Estimate

- Suppose that we want to build an SLRM for response Y and a single predictor predictor X_1: $\hat{Y} = \beta_0 + \beta_1 X_1$.
- To that end, we want to use a randomly sampled dataset of size n *but the samples are not identified yet.*
- Recall the OLS estimators of β_0 and β_1: $\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{X}_n$,

Estimator and Estimate

- Suppose that we want to build an SLRM for response Y and a single predictor predictor X_1: $\hat{Y} = \beta_0 + \beta_1 X_1$.
- To that end, we want to use a randomly sampled dataset of size n *but the samples are not identified yet.*
- Recall the OLS estimators of β_0 and β_1: $\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{X}_n$, and $\hat{\beta}_1 = \dfrac{S_{XY}}{S_{XX}}$.

Estimator and Estimate

- Suppose that we want to build an SLRM for response Y and a single predictor predictor X_1: $\hat{Y} = \beta_0 + \beta_1 X_1$.
- To that end, we want to use a randomly sampled dataset of size n *but the samples are not identified yet.*
- Recall the OLS estimators of β_0 and β_1: $\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{X}_n$, and $\hat{\beta}_1 = \dfrac{S_{XY}}{S_{XX}}$.
- The OLS estimators are random variables because they depend on the dataset.

Estimator and Estimate

- Suppose that we want to build an SLRM for response Y and a single predictor predictor X_1: $\hat{Y} = \beta_0 + \beta_1 X_1$.
- To that end, we want to use a randomly sampled dataset of size n *but the samples are not identified yet.*
- Recall the OLS estimators of β_0 and β_1: $\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{X}_n$, and $\hat{\beta}_1 = \dfrac{S_{XY}}{S_{XX}}$.
- The OLS estimators are random variables because they depend on the dataset.
- Suppose we identify, say, $n = 32$ samples;

Estimator and Estimate

- Suppose that we want to build an SLRM for response Y and a single predictor predictor X_1: $\hat{Y} = \beta_0 + \beta_1 X_1$.
- To that end, we want to use a randomly sampled dataset of size n *but the samples are not identified yet.*
- Recall the OLS estimators of β_0 and β_1: $\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{X}_n$, and $\hat{\beta}_1 = \dfrac{S_{XY}}{S_{XX}}$.
- The OLS estimators are random variables because they depend on the dataset.
- Suppose we identify, say, $n = 32$ samples; then we have that dataset-specific estimates:

Estimator and Estimate

- Suppose that we want to build an SLRM for response Y and a single predictor predictor X_1: $\hat{Y} = \beta_0 + \beta_1 X_1$.
- To that end, we want to use a randomly sampled dataset of size n *but the samples are not identified yet.*
- Recall the OLS estimators of β_0 and β_1: $\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{X}_n$, and $\hat{\beta}_1 = \dfrac{S_{XY}}{S_{XX}}$.
- The OLS estimators are random variables because they depend on the dataset.
- Suppose we identify, say, $n = 32$ samples; then we have that dataset-specific estimates: $\hat{\beta}_0 = \bar{y}_n - \hat{\beta}_1 \bar{x}_n$,

Estimator and Estimate

- Suppose that we want to build an SLRM for response Y and a single predictor predictor X_1: $\hat{Y} = \beta_0 + \beta_1 X_1$.
- To that end, we want to use a randomly sampled dataset of size n *but the samples are not identified yet.*
- Recall the OLS estimators of β_0 and β_1: $\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{X}_n$, and $\hat{\beta}_1 = \dfrac{S_{XY}}{S_{XX}}$.
- The OLS estimators are random variables because they depend on the dataset.
- Suppose we identify, say, $n = 32$ samples; then we have that dataset-specific estimates: $\hat{\beta}_0 = \bar{y}_n - \hat{\beta}_1 \bar{x}_n$, and $\hat{\beta}_1 = \dfrac{s_{xy}}{s_{xx}}$.

Interpreting OLS Estimators for a Continuous Predictor

Interpreting OLS Estimators for a Continuous Predictor

- Suppose that we consider the *mtcars* dataset with mpg as the response and hp as the predictor.

Interpreting OLS Estimators for a Continuous Predictor

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- Suppose that we consider the *mtcars* dataset with mpg as the response and hp as the predictor.
- SLRM predicts $\widehat{mpg} = \hat{\beta}_0 + \hat{\beta}_1 hp$.

Interpreting OLS Estimators for a Continuous Predictor

- Suppose that we consider the *mtcars* dataset with mpg as the response and hp as the predictor.
- SLRM predicts $\widehat{mpg} = \hat{\beta}_0 + \hat{\beta}_1 hp$.
- $\hat{\beta}_0$ is the predicted mpg when hp is 0.

Interpreting OLS Estimators for a Continuous Predictor

- Suppose that we consider the *mtcars* dataset with mpg as the response and hp as the predictor.
- SLRM predicts $\widehat{mpg} = \hat{\beta}_0 + \hat{\beta}_1 hp$.
- $\hat{\beta}_0$ is the predicted mpg when hp is 0.
- What about $\hat{\beta}_1$?

Interpreting OLS Estimators for a Continuous Predictor

- Suppose that we consider the *mtcars* dataset with mpg as the response and hp as the predictor.
- SLRM predicts $\widehat{mpg} = \hat{\beta}_0 + \hat{\beta}_1 hp$.
- $\hat{\beta}_0$ is the predicted mpg when hp is 0.
- What about $\hat{\beta}_1$? It is the change in the predicted mpg for a 1 unit increase in hp:

Interpreting OLS Estimators for a Continuous Predictor

- Suppose that we consider the *mtcars* dataset with mpg as the response and hp as the predictor.
- SLRM predicts $\widehat{mpg} = \hat{\beta}_0 + \hat{\beta}_1 hp$.
- $\hat{\beta}_0$ is the predicted mpg when hp is 0.
- What about $\hat{\beta}_1$? It is the change in the predicted mpg for a 1 unit increase in hp:

$$\begin{cases} \widehat{mpg}_{\text{old}} & = \hat{\beta}_0 + \hat{\beta}_1 hp \\ \widehat{mpg}_{\text{new}} & = \hat{\beta}_0 + \hat{\beta}_1 (hp + 1) \end{cases} \Rightarrow \widehat{mpg}_{\text{new}} - \widehat{mpg}_{\text{old}} = \hat{\beta}_1.$$

Interpreting OLS Estimators for a Categorical Predictor

Interpreting OLS Estimators for a Categorical Predictor

- Suppose that in the *mtcars* dataset we add a new column called $heavy$ with *yes* or *no* entries indicating whether a car is heavy or not.

Interpreting OLS Estimators for a Categorical Predictor

- Suppose that in the *mtcars* dataset we add a new column called $heavy$ with *yes* or *no* entries indicating whether a car is heavy or not.
- Suppose that mpg is the response and that $heavy$ is the predictor.

Interpreting OLS Estimators for a Categorical Predictor

- Suppose that in the *mtcars* dataset we add a new column called $heavy$ with *yes* or *no* entries indicating whether a car is heavy or not.
- Suppose that mpg is the response and that $heavy$ is the predictor.
- As there are 2 levels (*yes*, *no*) for the categorical variable $heavy$,

Interpreting OLS Estimators for a Categorical Predictor

- Suppose that in the *mtcars* dataset we add a new column called $heavy$ with *yes* or *no* entries indicating whether a car is heavy or not.
- Suppose that mpg is the response and that $heavy$ is the predictor.
- As there are 2 levels (*yes*, *no*) for the categorical variable $heavy$, 1 new *dummy variable* called $heavyyes$ is created.

Interpreting OLS Estimators for a Categorical Predictor

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- Suppose that in the *mtcars* dataset we add a new column called $heavy$ with *yes* or *no* entries indicating whether a car is heavy or not.
- Suppose that mpg is the response and that $heavy$ is the predictor.
- As there are 2 levels (*yes*, *no*) for the categorical variable $heavy$, 1 new *dummy variable* called $heavyyes$ is created.
- Note that the level *no* is the reference level per alphabetical order;

Interpreting OLS Estimators for a Categorical Predictor

- Suppose that in the *mtcars* dataset we add a new column called $heavy$ with *yes* or *no* entries indicating whether a car is heavy or not.
- Suppose that mpg is the response and that $heavy$ is the predictor.
- As there are 2 levels (*yes*, *no*) for the categorical variable $heavy$, 1 new *dummy variable* called $heavyyes$ is created.
- Note that the level *no* is the reference level per alphabetical order; that is, $heavyyes$ is equal to 0 if car is not heavy and 1 if it is.

Interpreting OLS Estimators for a Categorical Predictor

- Suppose that in the *mtcars* dataset we add a new column called $heavy$ with *yes* or *no* entries indicating whether a car is heavy or not.
- Suppose that mpg is the response and that $heavy$ is the predictor.
- As there are 2 levels (*yes*, *no*) for the categorical variable $heavy$, 1 new *dummy variable* called $heavyyes$ is created.
- Note that the level *no* is the reference level per alphabetical order; that is, $heavyyes$ is equal to 0 if car is <u>not</u> heavy and 1 if it is.
- The SLRM is $\widehat{mpg} = \hat{\beta}_0 + \hat{\beta}_1 heavyyes.$

Interpreting OLS Estimators for a Categorical Predictor

MANIPAL ACADEMY of HIGHER EDUCATION (Institution of Eminence Deemed to be University)

- Suppose that in the *mtcars* dataset we add a new column called $heavy$ with *yes* or *no* entries indicating whether a car is heavy or not.
- Suppose that mpg is the response and that $heavy$ is the predictor.
- As there are 2 levels (*yes*, *no*) for the categorical variable $heavy$, 1 new *dummy variable* called $heavyyes$ is created.
- Note that the level *no* is the reference level per alphabetical order; that is, $heavyyes$ is equal to 0 if car is not heavy and 1 if it is.
- The SLRM is $\widehat{mpg} = \hat{\beta}_0 + \hat{\beta}_1 heavyyes$.
- $\hat{\beta}_0$ is the average mpg of the not heavy cars (reference level).

Interpreting OLS Estimators for a Categorical Predictor

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- Suppose that in the *mtcars* dataset we add a new column called $heavy$ with *yes* or *no* entries indicating whether a car is heavy or not.
- Suppose that mpg is the response and that $heavy$ is the predictor.
- As there are 2 levels (*yes*, *no*) for the categorical variable $heavy$, 1 new *dummy variable* called $heavyyes$ is created.
- Note that the level *no* is the reference level per alphabetical order; that is, $heavyyes$ is equal to 0 if car is <u>not</u> heavy and 1 if it is.
- The SLRM is $\widehat{mpg} = \hat{\beta}_0 + \hat{\beta}_1 heavyyes$.
- $\hat{\beta}_0$ is the average mpg of the not heavy cars (reference level).
- $\hat{\beta}_1$ is the difference between the average mpg of the heavy cars and the average mpg of the not heavy cars (reference level).

Properties of OLS estimators

Properties of OLS estimators

MANIPAL ACADEMY of HIGHER EDUCATION (Institution of Eminence Deemed to be University)

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of β_0 and β_1, respectively:

Properties of OLS estimators

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of β_0 and β_1, respectively: $E[\hat{\beta}_0] = \beta_0$ and $E[\hat{\beta}_1] = \beta_1$.

Properties of OLS estimators

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of β_0 and β_1, respectively: $E[\hat{\beta}_0] = \beta_0$ and $E[\hat{\beta}_1] = \beta_1$.
- In the linear approximation $Y^{(i)} = \beta_0 + \beta_1 X_1^{(i)} + \varepsilon^{(i)}$, recall the assumptions about the random error term $\varepsilon^{(i)}$:

Properties of OLS estimators

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of β_0 and β_1, respectively: $E[\hat{\beta}_0] = \beta_0$ and $E[\hat{\beta}_1] = \beta_1$.
- In the linear approximation $Y^{(i)} = \beta_0 + \beta_1 X_1^{(i)} + \varepsilon^{(i)}$, recall the assumptions about the random error term $\varepsilon^{(i)}$: zero mean, constant variance = σ^2, and uncorrelated across the samples.

Properties of OLS estimators

MANIPAL ACADEMY of HIGHER EDUCATION (Institution of Eminence Deemed to be University)

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of β_0 and β_1, respectively: $E[\hat{\beta}_0] = \beta_0$ and $E[\hat{\beta}_1] = \beta_1$.
- In the linear approximation $Y^{(i)} = \beta_0 + \beta_1 X_1^{(i)} + \varepsilon^{(i)}$, recall the assumptions about the random error term $\varepsilon^{(i)}$: zero mean, constant variance = σ^2, and uncorrelated across the samples.
- Variance of $\hat{\beta}_0$ and $\hat{\beta}_1$:

Properties of OLS estimators

MANIPAL
ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of β_0 and β_1, respectively: $E[\hat{\beta}_0] = \beta_0$ and $E[\hat{\beta}_1] = \beta_1$.
- In the linear approximation $Y^{(i)} = \beta_0 + \beta_1 X_1^{(i)} + \varepsilon^{(i)}$, recall the assumptions about the random error term $\varepsilon^{(i)}$: zero mean, constant variance = σ^2, and uncorrelated across the samples.
- Variance of $\hat{\beta}_0$ and $\hat{\beta}_1$: $Var[\hat{\beta}_0] = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}_n^2}{s_{xx}}\right)$,

Properties of OLS estimators

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of β_0 and β_1, respectively: $E[\hat{\beta}_0] = \beta_0$ and $E[\hat{\beta}_1] = \beta_1$.
- In the linear approximation $Y^{(i)} = \beta_0 + \beta_1 X_1^{(i)} + \varepsilon^{(i)}$, recall the assumptions about the random error term $\varepsilon^{(i)}$: zero mean, constant variance = σ^2, and uncorrelated across the samples.
- Variance of $\hat{\beta}_0$ and $\hat{\beta}_1$: $Var[\hat{\beta}_0] = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}_n^2}{s_{xx}}\right)$, and $Var[\hat{\beta}_1] = \frac{\sigma^2}{s_{xx}}$.

Properties of OLS estimators

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of β_0 and β_1, respectively: $E[\hat{\beta}_0] = \beta_0$ and $E[\hat{\beta}_1] = \beta_1$.
- In the linear approximation $Y^{(i)} = \beta_0 + \beta_1 X_1^{(i)} + \varepsilon^{(i)}$, recall the assumptions about the random error term $\varepsilon^{(i)}$: zero mean, constant variance = σ^2, and uncorrelated across the samples.
- Variance of $\hat{\beta}_0$ and $\hat{\beta}_1$: $Var[\hat{\beta}_0] = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}_n^2}{s_{xx}}\right)$, and $Var[\hat{\beta}_1] = \frac{\sigma^2}{s_{xx}}$.
- σ^2 is typically unknown:

Properties of OLS estimators

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of β_0 and β_1, respectively: $E[\hat{\beta}_0] = \beta_0$ and $E[\hat{\beta}_1] = \beta_1$.
- In the linear approximation $Y^{(i)} = \beta_0 + \beta_1 X_1^{(i)} + \varepsilon^{(i)}$, recall the assumptions about the random error term $\varepsilon^{(i)}$: zero mean, constant variance = σ^2, and uncorrelated across the samples.
- Variance of $\hat{\beta}_0$ and $\hat{\beta}_1$: $Var[\hat{\beta}_0] = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}_n^2}{s_{xx}}\right)$, and $Var[\hat{\beta}_1] = \frac{\sigma^2}{s_{xx}}$.
- σ^2 is typically unknown: use the in-sample approximation $\sigma^2 \approx \frac{1}{n-2}\sum_{i=1}^{n} \left(y^{(i)} - \hat{y}^{(i)}\right)^2$.

Properties of OLS estimators

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of β_0 and β_1, respectively: $E[\hat{\beta}_0] = \beta_0$ and $E[\hat{\beta}_1] = \beta_1$.
- In the linear approximation $Y^{(i)} = \beta_0 + \beta_1 X_1^{(i)} + \varepsilon^{(i)}$, recall the assumptions about the random error term $\varepsilon^{(i)}$: zero mean, constant variance = σ^2, and uncorrelated across the samples.
- Variance of $\hat{\beta}_0$ and $\hat{\beta}_1$: $Var[\hat{\beta}_0] = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}_n^2}{s_{xx}} \right)$, and $Var[\hat{\beta}_1] = \frac{\sigma^2}{s_{xx}}$.
- σ^2 is typically unknown: use the in-sample approximation $\sigma^2 \approx \frac{1}{n-2} \sum_{i=1}^{n} \left(y^{(i)} - \hat{y}^{(i)} \right)^2$.
- Recall that the summation term is the RSS.

Prediction Problem

Prediction Problem

- Let $x_1^{(i)}$ and $y^{(i)}$ represent the ith sample's predictor and response values in a dataset with, say, n samples.

Prediction Problem

- Let $x_1^{(i)}$ and $y^{(i)}$ represent the ith sample's predictor and response values in a dataset with, say, n samples.
- The true population relationship is given by $Y = f(X) + \varepsilon$ for an unknown function f.

Prediction Problem

- Let $x_1^{(i)}$ and $y^{(i)}$ represent the ith sample's predictor and response values in a dataset with, say, n samples.
- The true population relationship is given by $Y = f(X) + \varepsilon$ for an unknown function f.
- The prediction problem is to build an approximation $\hat{f}$ of f using the dataset such that

Prediction Problem

- Let $x_1^{(i)}$ and $y^{(i)}$ represent the ith sample's predictor and response values in a dataset with, say, n samples.
- The true population relationship is given by $Y = f(X) + \varepsilon$ for an unknown function f.
- The prediction problem is to build an approximation $\hat{f}$ of f using the dataset such that for a new predictor value X from the population, the *prediction error* between $\hat{f}(X)$ and the corresponding response value Y is minimized.

Prediction Error

Prediction Error

- How do we measure the prediction error?

Prediction Error

- How do we measure the prediction error?
- For regression, we have, for example,

Prediction Error

- How do we measure the prediction error?
- For regression, we have, for example, squared error $\left(Y - \hat{f}(X)\right)^2$,

Prediction Error

- How do we measure the prediction error?
- For regression, we have, for example, squared error $\left(Y - \hat{f}(X)\right)^2$, and absolute deviation $\left|Y - \hat{f}(X)\right|$.

Prediction Error

- How do we measure the prediction error?
- For regression, we have, for example, squared error $\left(Y - \hat{f}(X)\right)^2$, and absolute deviation $\left|Y - \hat{f}(X)\right|$.
- We will use the squared error as the measure of prediction error.

Prediction Error

- How do we measure the prediction error?
- For regression, we have, for example, squared error $\left(Y - \hat{f}(X)\right)^2$, and absolute deviation $\left|Y - \hat{f}(X)\right|$.
- We will use the squared error as the measure of prediction error.
- How do we build a good “fitted” model?

Prediction Error

- How do we measure the prediction error?
- For regression, we have, for example, squared error $\left(Y - \hat{f}(X)\right)^2$, and absolute deviation $\left|Y - \hat{f}(X)\right|$.
- We will use the squared error as the measure of prediction error.
- How do we build a good "fitted" model? Minimizes the prediction error on *unseen* data.

Prediction Error

- How do we measure the prediction error?
- For regression, we have, for example, squared error $\left(Y - \hat{f}(X)\right)^2$, and absolute deviation $\left|Y - \hat{f}(X)\right|$.
- We will use the squared error as the measure of prediction error.
- How do we build a good "fitted" model? Minimizes the prediction error on *unseen* data.
- Train-validation-test split of data.

Summary

Summary

- Interpreted OLS estimators for continuous and categorical predictors.

Summary

- Interpreted OLS estimators for continuous and categorical predictors.
- Described the properties of OLS estimators.

Summary

- Interpreted OLS estimators for continuous and categorical predictors.
- Described the properties of OLS estimators.
- Described the prediction problem and interpreted the prediction error.