Multiple Linear Regression

Dr. Supaporn Erjongmanee

Department of Computer Engineering
Kasetsart University
fengspe@ku.ac.th

1

Outline

- Introduction
- Estimating Parameters
- Residual, SSE, SST, R^2
- Example

2

Multiple Regression Model

Multiple Regression

$$x_1, x_2, \ldots, x_n \rightarrow \boxed{f = ?} \rightarrow y$$

- General model

  $Y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k + \varepsilon$

  - where
    - X = independent/predictor variables
    - Y = dependent variable
    - k = number of predictors
    - ε is normally distributed with $\mu = 0$, var = σ^2
      - When σ^2 -> 0, ε is close to zero, Y is closer to true regression line
      - When σ^2 is large, ε is close to zero, Y is closer to true regression line
    - b_j = rate how y increases according to x_j increases

3

Observation Set

- Let x_{ij} = observed j^{th} predictor in i^{th} data set
  - i = 1,2, ..., n, j = 1,2, ..., k
- The data set is composed of n sets:
  - $(x_{11}, x_{12}, x_{13}, \ldots, x_{1k}, y_1)$ $\quad \boxed{y_1 = b_0 + b_1x_{11} + b_2x_{12} + \ldots + b_kx_{1k} + \varepsilon}$
  - $(x_{21}, x_{22}, x_{23}, \ldots, x_{2k}, y_2)$ $\quad \boxed{y_2 = b_0 + b_1x_{21} + b_2x_{22} + \ldots + b_kx_{2k} + \varepsilon}$
  - …
  - $(x_{n1}, x_{n2}, x_{n3}, \ldots, x_{nk}, y_n)$ $\quad \boxed{y_n = b_0 + b_1x_{n1} + b_2x_{n2} + \ldots + b_kx_{nk} + \varepsilon}$

4

Outline

- Introduction
- Estimating Parameters
- Residual, SSE, SST, R^2
- Example

5

Estimating Parameters

$$Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k + \varepsilon$$

- Parameters to estimate are
  - $b_1, b_2, \dots, b_k$
- Also, apply principle of least squares to minimize sum of errors [g(·)] to find estimated parameters

$$g(b_0, b_1, b_2, b_k) = \sum_{i=1}^{n} [\, y_i - (b_0 + b_1\, x_{i1} + \dots + b_k\, x_{ik})]^2$$

6

Estimating Parameters (cont.)

$$g(b_0, b_1, b_2, b_k) = \sum_{i=1}^{n} [\, y_i - (b_0 + b_1 x_{i1} + \dots + b_k x_{ik})]^2$$

- Take partial derivatives of g(·) with respect to each b_j and set to zero.

$$
\begin{aligned}
b_0 \sum_{i=1}^{n} 1 + b_1 \sum_{i=1}^{n} x_{i1} + \dots + b_k \sum_{i=1}^{n} x_{ik} &= \sum_{i=1}^{n} y_i \\
b_0 \sum_{i=1}^{n} x_{i1} + b_1 \sum_{i=1}^{n} x_{i1} x_{i1} + \dots + b_k \sum_{i=1}^{n} x_{i1} x_{ik} &= \sum_{i=1}^{n} x_{i1} y_i \\
\dots \\
b_0 \sum_{i=1}^{n} x_{ik} + b_1 \sum_{i=1}^{n} x_{ik} x_{i1} + \dots + b_k \sum_{i=1}^{n} x_{ik} x_{ik} &= \sum_{i=1}^{n} x_{ik} y_i
\end{aligned}
$$

- Help for solving b_j's
  - Use software for help finding values of b_j's
  - Use matrix to help out calculation

7

Regression with Matrices

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k + \varepsilon$$

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ & & \vdots & \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} b_0 \\ \vdots \\ b_k \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ & & \vdots & \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} b_0 \\ \vdots \\ b_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

8

Regression with Matrices (cont.)

$g(b_0, b_1, b_2, b_k)$
$=\sum_{i=1}^{n}[\, y_i \; - (b_0 \; + b_1 \; x_{i1} + \; \ldots + b_k \; x_{ik})]^2$
$= (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})$
$= \|\, \boldsymbol{y} - \boldsymbol{X\beta} \,\|^2$

- Apply principle of least squares to minimize sum of errors (·) to find $\boldsymbol{\beta}$
- Take partial derivatives of g(·) with respect to each b_j and set to zero.

$$
\begin{aligned}
&b_0 \sum_{i=1}^{n} 1 \; + b_1 \sum_{i=1}^{n} x_{i1} \; + \ldots + b_k \sum_{i=1}^{n} x_{ik} = \sum_{i=1}^{n} y_i \\
&b_0 \sum_{i=1}^{n} x_{i1} + b_1 \sum_{i=1}^{n} x_{i1}\, x_{i1} + \ldots + b_k \sum_{i=1}^{n} x_{i1}\, x_{ik} = \sum_{i=1}^{n} x_{i1} y_i \\
&\ldots \\
&b_0 \sum_{i=1}^{n} x_{ik} + b_1 \sum_{i=1}^{n} x_{ik}\, x_{i1} + \ldots + b_k \sum_{i=1}^{n} x_{ik}\, x_{ik} = \sum_{i=1}^{n} x_{ik} y_i
\end{aligned}
$$

Regression with Matrices (cont.)

- Take partial derivatives of g(·) with respect to each b_j and set to zero.

$$
\begin{aligned}
&b_0 \sum_{i=1}^{n} 1 \; + b_1 \sum_{i=1}^{n} x_{i1} \; + \ldots + b_k \sum_{i=1}^{n} x_{ik} = \sum_{i=1}^{n} y_i \\
&b_0 \sum_{i=1}^{n} x_{i1} + b_1 \sum_{i=1}^{n} x_{i1}\, x_{i1} + \ldots + b_k \sum_{i=1}^{n} x_{i1}\, x_{ik} = \sum_{i=1}^{n} x_{i1} y_i \\
&\ldots \\
&b_0 \sum_{i=1}^{n} x_{ik} + b_1 \sum_{i=1}^{n} x_{ik}\, x_{i1} + \ldots + b_k \sum_{i=1}^{n} x_{ik}\, x_{ik} = \sum_{i=1}^{n} x_{ik} y_i
\end{aligned}
$$

- In a matrix, we can write:

$$
\begin{bmatrix}
\sum_{i=1}^{n} 1 & \sum_{i=1}^{n} x_{i1} & \ldots & \sum_{i=1}^{n} x_{ik} \\
\sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}\, x_{i1} & \ldots & \sum_{i=1}^{n} x_{i1}\, x_{ik} \\
 & & \ldots & \\
\sum_{i=1}^{n} x_{ik} & \sum_{i=1}^{n} x_{ik}\, x_{i1} & \ldots & \sum_{i=1}^{n} x_{ik}\, x_{ik}
\end{bmatrix}
\begin{bmatrix} b_0 \\ \vdots \\ b_k \end{bmatrix}
=
\begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1} y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ik} y_i \end{bmatrix}
$$

Regression with Matrices (cont.)

- In a matrix, we can write:

$$\begin{bmatrix} \sum_{i=1}^{n} 1 & \sum_{i=1}^{n} x_{i1} & \dots & \sum_{i=1}^{n} x_{ik} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1} x_{i1} & \dots & \sum_{i=1}^{n} x_{i1} x_{ik} \\ & \dots & & \\ \sum_{i=1}^{n} x_{ik} & \sum_{i=1}^{n} x_{ik} x_{i1} & \dots & \sum_{i=1}^{n} x_{ik} x_{ik} \end{bmatrix} \begin{bmatrix} b_0 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1} y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ik} y_i \end{bmatrix}$$

$$\boldsymbol{X'X\beta} = \boldsymbol{X'y}$$

Note: $\boldsymbol{X'}$ = transpose of X

$$\boxed{\hat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y}}$$

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$$

After finding b_0, b_1, b_2, …, b_k, we can compute SSE, SST, R^2,

11

Example

- Use data of 6 cars, build a model to predict horsepower (hp) using two inputs: engine size (liters) and fuel type

Make	hp	Engine size	Fuel
Ford	132	2.0	Regular
Mazda	167	2.0	Premium
Subaru	170	2.5	Regular
Lexus	204	2.5	Premium
Mitsubishi	230	3.0	Regular
BMW	260	3.0	Premium

- Let
  - x_1 = Engine size
  - x_2 = Fuel (replace 0 for regular, 1 for premium)
  - Y = hp

Source: [1]

12

Example (cont.)

- n = 6
- k = 2

Make	hp	Engine size	Fuel
Ford	132	2.0	Regular
Mazda	167	2.0	Premium
Subaru	170	2.5	Regular
Lexus	204	2.5	Premium
Mitsubishi	230	3.0	Regular
BMW	260	3.0	Premium

$$\boxed{\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}}$$

$$\boldsymbol{X}=\begin{bmatrix}1 & 2.0 & 0\\1 & 2.0 & 1\\1 & 2.5 & 0\\1 & 2.5 & 1\\1 & 3.0 & 0\\1 & 3.0 & 1\end{bmatrix} \quad \boldsymbol{y}=\begin{bmatrix}132\\167\\170\\204\\230\\260\end{bmatrix} \quad \boldsymbol{X}'\boldsymbol{X}=\begin{bmatrix}6 & 15 & 3\\15 & 38.5 & 7.5\\3 & 7.5 & 3\end{bmatrix}$$

$$\boldsymbol{X}'\boldsymbol{y}=\begin{bmatrix}1163\\3003\\631\end{bmatrix}$$

Source: [1]

13

Example (cont.)

$$\boxed{\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}}$$

$$\boldsymbol{X}=\begin{bmatrix}1 & 2.0 & 0\\1 & 2.0 & 1\\1 & 2.5 & 0\\1 & 2.5 & 1\\1 & 3.0 & 0\\1 & 3.0 & 1\end{bmatrix} \quad \boldsymbol{y}=\begin{bmatrix}132\\167\\170\\204\\230\\260\end{bmatrix} \quad \boldsymbol{X}'\boldsymbol{X}=\begin{bmatrix}6 & 15 & 3\\15 & 38.5 & 7.5\\3 & 7.5 & 3\end{bmatrix}$$

$$\boldsymbol{X}'\boldsymbol{y}=\begin{bmatrix}1163\\3003\\631\end{bmatrix}$$

$$\boxed{\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}} = \begin{bmatrix}79/12 & -5/2 & -1/3\\-5/2 & 1 & 0\\-1/3 & 0 & 2/3\end{bmatrix}\begin{bmatrix}1163\\3003\\631\end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix}\hat{b}_0\\\hat{b}_1\\\hat{b}_2\end{bmatrix} = \begin{bmatrix}-61.42\\95.50\\33.00\end{bmatrix}$$

$$\boxed{y = -61.42 + 95.5x_1 + 33.0x_2}$$

b1, b2 are not zero. Both factors (engine size, fuel) affect horsepower (hp).

Source: [1]

14

Example (cont.)

$$y = -61.42 + 95.5x_1 + 33.0x_2$$

What is SSE?
SST?
R^2?

- Let
  - x_1 = Engine size
  - x_2 = Fuel (replace 0 for regular, 1 for premium)
  - Y = hp
- If fuel has no effect, we increase engine size by one, horsepower is increased by 95.5
- If engine size has no effect, we increase fuel by one (change from regular to premium), horsepower is increased by 33.

Source: [1]

Outline

- Introduction
- Estimating Parameters
- Residual, SSE, SST, R^2
- Example

Residuals, Sum of Errors (SSE), σ^2

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

$$\sigma^2 = \frac{SSE}{n-2}$$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$$

$$\boxed{\text{Residual} = \boldsymbol{y} - \hat{\boldsymbol{y}}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$$

$$SSE = (\boldsymbol{y} - \hat{\boldsymbol{y}})(\boldsymbol{y} - \hat{\boldsymbol{y}}) = ||\, \boldsymbol{y} - \hat{\boldsymbol{y}}\, ||^2$$

$$\sigma^2 = \frac{SSE}{[n-(k+1)]} = MSE$$

How to compute residuals ?

How to compute SSE ?

How to compute σ^2 ?

What is σ^2 ?

17

Example (cont.)

$$SSE = ||\, \boldsymbol{y} - \hat{\boldsymbol{y}}\, ||^2$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & 2.0 & 0 \\ 1 & 2.0 & 1 \\ 1 & 2.5 & 0 \\ 1 & 2.5 & 1 \\ 1 & 3.0 & 0 \\ 1 & 3.0 & 1 \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} = \begin{bmatrix} -61.42 \\ 95.50 \\ 33.00 \end{bmatrix}$$

$$\boldsymbol{y} = \begin{bmatrix} 132 \\ 167 \\ 170 \\ 204 \\ 230 \\ 260 \end{bmatrix}$$

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1 & 2.0 & 0 \\ 1 & 2.0 & 1 \\ 1 & 2.5 & 0 \\ 1 & 2.5 & 1 \\ 1 & 3.0 & 0 \\ 1 & 3.0 & 1 \end{bmatrix} \begin{bmatrix} -61.42 \\ 95.50 \\ 33.00 \end{bmatrix} = \begin{bmatrix} 129.58 \\ 162.58 \\ 177.33 \\ 210.33 \\ 225.08 \\ 258.08 \end{bmatrix}$$

$$\text{Residual} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \begin{bmatrix} 132 \\ 167 \\ 170 \\ 204 \\ 230 \\ 260 \end{bmatrix} - \begin{bmatrix} 129.58 \\ 162.58 \\ 177.33 \\ 210.33 \\ 225.08 \\ 258.08 \end{bmatrix} = \begin{bmatrix} 2.42 \\ 4.42 \\ -7.33 \\ -6.33 \\ 4.92 \\ 1.92 \end{bmatrix}$$

$$SSE = ||\, \boldsymbol{y} - \hat{\boldsymbol{y}}\, ||^2 = 2.42^2 + \ldots + 1.92^2 = 147.08$$

18

Example (cont.)

$\sigma^2 = \frac{SSE}{[n-(k+1)]} = MSE$

$SSE = ||\boldsymbol{y} - \hat{\boldsymbol{y}}||^2 = 2.42^2 + \ldots + 1.92^2 = 147.08$

$\sigma^2 = \frac{SSE}{[n-(k+1)]} = \frac{147.08}{[6-(2+1)]} = 49.03$ ➡ $\boxed{\sigma = \sqrt{49.03} = 7.002}$

data = 6

2 factors: x_1, x_2

Make	hp	Engine size	Fuel
Ford	132	2.0	Regular
Mazda	167	2.0	Premium
Subaru	170	2.5	Regular
Lexus	204	2.5	Premium
Mitsubishi	230	3.0	Regular
BMW	260	3.0	Premium

Total Sum of Error (*SST*) and Regression Sum of Squares (*SSR*)

$SST = (\boldsymbol{y} - \bar{\boldsymbol{y}})(\boldsymbol{y} - \bar{\boldsymbol{y}}) = ||\boldsymbol{y} - \bar{\boldsymbol{y}}||^2$

$SST = ||\boldsymbol{y} - \bar{\boldsymbol{y}}||^2$

$= [(\boldsymbol{y} - \hat{\boldsymbol{y}}) + (\hat{\boldsymbol{y}} - \bar{\boldsymbol{y}})]'[(\boldsymbol{y} - \hat{\boldsymbol{y}}) + (\hat{\boldsymbol{y}} - \bar{\boldsymbol{y}})]$

$= ||\boldsymbol{y} - \hat{\boldsymbol{y}}||^2 + ||\hat{\boldsymbol{y}} - \bar{\boldsymbol{y}}||^2$ Other terms are zeros

Note that $SST = SSE + SSR$

$SSR = ||\hat{\boldsymbol{y}} - \bar{\boldsymbol{y}}||^2$ = Regression sum of squares
= Sum of squared fitted values from $\bar{\boldsymbol{y}}$

Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{SSE}{SST}$$

Proportion of fitted values that can be explained by multiple linear model

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Example (cont.)

$SST = ||\, \boldsymbol{y} - \bar{\boldsymbol{y}}||^2$ $\quad R^2 = 1 - \frac{SSE}{SST}$

$SST = ||\, \boldsymbol{y} - \bar{\boldsymbol{y}}||^2 = \sum_{i=1}^{n}(y_i - 193.83)^2 = 10{,}900.83$

$SSE = ||\, \boldsymbol{y} - \hat{\boldsymbol{y}}\,||^2 = 2.42^2 + \ldots + 1.92^2 = 147.08$

$SSR = SST - SSE = 10{,}573.75$

Make	hp
Ford	132
Mazda	167
Subaru	170
Lexus	204
Mitsubishi	230
BMW	260

$\bar{\boldsymbol{y}} = 193.83$

$$\boxed{R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{147.08}{10900.83} = 0.9865}$$

98.65% of fitted values that can be explained by multiple linear model

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Outline

- Introduction
- Estimating Parameters
- Residual, SSE, SST, R^2
- Example

Example 2

- What affects human wingspan? Height? Foot length?
- Use the given data to find the followings:
  - σ, σ^2
  - SSE
  - R^2
  - SST, SSR
  - MSR, MSE

$$\sigma^2 = \frac{SSE}{[n-(k+1)]}$$

$$R^2 = 1 - \frac{SSE}{SST}$$

$$SSR = SST - SSE$$

$$n = 16, k = 2$$

Example 2 (cont.)

- What affects human wingspan? Height? Foot length?
- Use the given data to find the followings:
  - At α = 0.05 , does height affect wingspan?
  - At α = 0.05 , does foot length affect wingspan?

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Example 2 (cont.)

- Use the given data to find the followings:
  - At α = 0.05 , does <u>height</u> affect wingspan (not consider foot length)?

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Example 2 (cont.)

$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- Compute correlation between
  - Wingspan and height (r_1)
  - Wingspan and foot length (r_2)
  - Height and foot length (r_3)

References

1. J.L. Devore and K.N.Berk, Modern Mathematical Statistics with Applications, Springer, 2012.