Linear Regression

Haixia Liu

September 13, 2020

Linear regression

- Data:
  - Quantitative response y
  - p different prdictors $x_1, x_2, \cdots, x_p$
- There is some relationship between y and $\mathbf{x} = (x_1, \cdots, x_p)$

$$y = f(\mathbf{x}) + \epsilon. \tag{1}$$

Linear regression

- Data:
    - Quantitative response y
    - p different prdictors $x_1, x_2, \cdots, x_p$
- There is some relationship between y and $\mathbf{x} = (x_1, \cdots, x_p)$

$$y = f(\mathbf{x}) + \epsilon. \tag{1}$$

- What is *linear regression*?
    - Equation (1) is a regression problem,
    - function f is linear.

Linear regression

- Data:
    - Quantitative response y
    - p different prdictors $x_1, x_2, \cdots, x_p$
- There is some relationship between y and $\mathbf{x} = (x_1, \cdots, x_p)$

$$y = f(\mathbf{x}) + \epsilon. \tag{1}$$

- What is *linear regression*?
    - Equation (1) is a regression problem,
    - function f is linear.
- Let y_i, $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})$ are the i-th observation, $i = 1, \cdots, n$. Consider linear regression model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i, i = 1, \cdots, n.$$

Simple linear regression $p = 1$

The linear regression model is

$$y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i,\ i = 1, \cdots, n.$$

Simple linear regression $p = 1$

The linear regression model is

$$y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i,\ i = 1, \cdots, n.$$

Minimizing the sum of squares of error:

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1})^2.$$

Simple linear regression $p = 1$

The linear regression model is

$$y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i,\ i = 1, \cdots, n.$$

Minimizing the sum of squares of error:

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1})^2.$$

The estimator is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_{i1} - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_{i1} - \bar{x})^2},$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_{i1}$ and $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$.

Multiple linear regression $p > 1$

Consider linear regression model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i, i = 1, \cdots, n. \tag{2}$$

Set $\tilde{\mathbf{x}}_i = [1, \mathbf{x}_i^T], i = 1, \cdots, n$

$$X = \begin{bmatrix} \tilde{\mathbf{x}}_1 \\ \vdots \\ \tilde{\mathbf{x}}_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}, \ \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix} \ \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \ \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix},$$

Equation (2) can be rewritten as

$$\mathbf{y} = X\beta + \epsilon.$$

The least squares estimation

Minimizing the sum of squares of error:

$$\min_{\beta_0,\cdots,\beta_p} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}))^2 = \min_{\beta} \sum_{i=1}^{n} (y_i - \tilde{\mathbf{x}}_i \beta)^2 = \|X\beta - \mathbf{y}\|_2^2.$$

By some linear algebra calcuation, the least squares estimator of is then

$$\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y}$$

Then

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

is called the fitted values, viewed as the predicted values of the reponses based on the linear model.

Statistical Analysis of the least squares estimation

Two problems we will consider:

- How about the linear model we choose?
  – R^2-statistic.
- If yes, how about relationship between predictors and responser?
  hypothesis test
  – t-statistic.
  – p-value.

Set $\hat{\epsilon} = \mathbf{y} - \hat{\mathbf{y}}$, which is residuals. The sum of squares of these residuals

$$\sum_{i=1}^{n} \hat{\epsilon}_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2.$$

Set $\hat{\epsilon} = \mathbf{y} - \hat{\mathbf{y}}$, which is residuals. The sum of squares of these residuals

$$\sum_{i=1}^{n} \hat{\epsilon}_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2.$$

$$X^T \hat{\epsilon} = X^T(\mathbf{y} - X\hat{\beta}) = X^T\mathbf{y} - X^T X (X^T X)^{-1} X^T \mathbf{y} = \mathbf{0}.$$

Set $\hat{\epsilon} = \mathbf{y} - \hat{\mathbf{y}}$, which is residuals. The sum of squares of these residuals

$$\sum_{i=1}^{n} \hat{\epsilon}_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2.$$

$$X^T \hat{\epsilon} = X^T(\mathbf{y} - X\hat{\beta}) = X^T\mathbf{y} - X^T X (X^T X)^{-1} X^T \mathbf{y} = \mathbf{0}.$$

- The residual $\hat{\epsilon}$ is orthogonal to all columns of X,
- The residual vector $\hat{\epsilon}$ is orthogonal to the hyperplane formed by the all columns in n dimensional real space.

$$X = \begin{bmatrix} \tilde{\mathbf{x}}_1 \\ \vdots \\ \tilde{\mathbf{x}}_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}, \ \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix}$$

- $\hat{\mathbf{y}}$ and $[\bar{y}, \cdots, \bar{y}]^T = \bar{y} * [1, \cdots, 1]^T$ are orthogonal to $\hat{\epsilon} = \mathbf{y} - \hat{\mathbf{y}}$ or

$$\hat{\mathbf{y}}^T(\mathbf{y} - \hat{\mathbf{y}}) = 0, \ [\bar{y}, \cdots, \bar{y}](\mathbf{y} - \hat{\mathbf{y}}) = 0,$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

$$X = \begin{bmatrix} \tilde{\mathbf{x}}_1 \\ \vdots \\ \tilde{\mathbf{x}}_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}, \ \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix}$$

- $\hat{\mathbf{y}}$ and $[\bar{y}, \cdots, \bar{y}]^T = \bar{y} * [1, \cdots, 1]^T$ are orthogonal to $\hat{\epsilon} = \mathbf{y} - \hat{\mathbf{y}}$ or

  $$\hat{\mathbf{y}}^T(\mathbf{y} - \hat{\mathbf{y}}) = 0, \ [\bar{y}, \cdots, \bar{y}](\mathbf{y} - \hat{\mathbf{y}}) = 0,$$

  where $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$.
- This implies

  $$\|\mathbf{y} - [\bar{y}, \cdots, \bar{y}]^T\|^2 = \|\hat{\mathbf{y}} - [\bar{y}, \cdots, \bar{y}]^T\|^2 + \|\mathbf{y} - \hat{\mathbf{y}}\|^2.$$

Recall

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \ \hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{bmatrix},$$

so

$$\|\mathbf{y} - [\bar{y}, \cdots, \bar{y}]^T\|^2 = \|\hat{\mathbf{y}} - [\bar{y}, \cdots, \bar{y}]^T\|^2 + \|\mathbf{y} - \hat{\mathbf{y}}\|^2$$

can be rewritten as

$$\underbrace{\sum_{i=1}^{n}(y_i - \bar{y})^2}_{\text{TSS}} = \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}_{\text{ESS}} + \underbrace{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}_{\text{RSS}}.$$

- **TSS**: Total Sum of Squares
- **ESS**: Explained Sum of Squares
- **RSS**: Residual Sum of Squares

R^2-statistic

$$R^2 = \frac{\text{TSS} - \text{RSS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}.$$

- Range of R^2: $[0, 1]$.
- R^2 value that is extremely close to 1 means the data truly comes from a linear model with a small residual error.
- smaller R^2 value might indicate a serious problem with the experiment in which the data were generated.

Model assumptions

The linear model is given by

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}, \ i = 1, \cdots, n.$$

- ϵ_i are random (need some assumptions),
- $x_{i1}, \cdots, x_{ip}, i = 1, \cdots, n$ are fixed (independent/predictor variable),
- y_i are random (dependent/response variable).

How about relationship between predictors and responser?

The most common *hypothesis test* involves testing the *null test hypothesis* of

$$H_0 : \text{There is no relationship between } x \text{ and } y,$$

versus the *alternative hypothesis*

$$H_a : \text{There is some relationship between } x \text{ and } y.$$

How about relationship between predictors and responser?

The most common *hypothesis test* involves testing the *null test hypothesis* of

$$H_0 : \text{There is no relationship between } x \text{ and } y,$$

versus the *alternative hypothesis*

$$H_a : \text{There is some relationship between } x \text{ and } y.$$

Mathematically,

$$p = 1 : \; H_0 : \beta_1 = 0 \text{ versus } H_a : \beta_1 \neq 0.$$

$$p \neq 1 : H_0 : \beta_1 = \cdots = \beta_p = 0 \text{ versus } H_a : \beta_{i^*} \neq 0 \text{ for some } i^*.$$

Review: Simple linear regression $p = 1$

The linear regression model is

$$y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i, \; i = 1, \cdots, n.$$

Minimizing the sum of squares of error:

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1})^2.$$

The estimator is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_{i1} - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_{i1} - \bar{x})^2},$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_{i1}$ and $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$.

Theorem

Define $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_{i1}$, $l_{xx} = \sum_{i=1}^{n}(x_{i1} - \bar{x})^2$, $S_e^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1})^2$ *and* $\hat{\sigma}^2 = \frac{S_e^2}{n-2}$. *Let* $\epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2), i = 1, \cdots, n$. *We have*

1. $\hat{\beta}_0 \sim \mathcal{N}(\beta_0, (\frac{1}{n} + \frac{\bar{x}^2}{l_{xx}})\sigma^2)$,
2. $\hat{\beta}_1 \sim \mathcal{N}(\beta_1, \frac{\sigma^2}{l_{xx}})$,
3. $\frac{(n-2)\hat{\sigma}^2}{\sigma^2} = \frac{S_e^2}{\sigma^2} \sim \chi(n-2)$,
4. $\hat{\sigma}^2$ *is independent of* $(\hat{\beta}_0, \hat{\beta}_1)$.

Theorem

Define $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_{i1}$, $l_{xx} = \sum_{i=1}^{n}(x_{i1} - \bar{x})^2$, $S_e^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1})^2$ *and* $\hat{\sigma}^2 = \frac{S_e^2}{n-2}$. *Let* $\epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2), i = 1, \cdots, n$. *We have*

1. $\hat{\beta}_0 \sim \mathcal{N}(\beta_0, (\frac{1}{n} + \frac{\bar{x}^2}{l_{xx}})\sigma^2)$,
2. $\hat{\beta}_1 \sim \mathcal{N}(\beta_1, \frac{\sigma^2}{l_{xx}})$,
3. $\frac{(n-2)\hat{\sigma}^2}{\sigma^2} = \frac{S_e^2}{\sigma^2} \sim \chi(n-2)$,
4. $\hat{\sigma}^2$ *is independent of* $(\hat{\beta}_0, \hat{\beta}_1)$.

So we have

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}\sqrt{1/\sum_{i=1}^{n}(x_i - \bar{x})^2}} \sim t_{n-2},$$

$$\frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}\sqrt{1/n + \bar{x}^2/\sum_{i=1}^{n}(x_i - \bar{x})^2}} \sim t_{n-2}.$$

To test the null hypothesis, we need to determine

- whether $\hat{\beta}_1$, our estimate for β_1, is sufficiently far from zero that we can be confident that β_1 is non-zero.
- How far is far enough?

To test the null hypothesis, we need to determine

- whether $\hat{\beta}_1$, our estimate for β_1, is sufficiently far from zero that we can be confident that β_1 is non-zero.
- How far is far enough?

In practice, we compute a *t-statistic*

$$t = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}\sqrt{1/\sum_{i=1}^{n}(x_i - \bar{x})^2}}.$$

Theorem

Define $S_e^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \cdots - \hat{\beta}_p x_{ip})^2$ *and* $\hat{\sigma}^2 = \frac{S_e^2}{n-p-1}$*. Let* $\epsilon \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2 I_n)$ *and* $rank(X) = p + 1 < n$*. We have*

1. $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2(X^T X)^{-1})$,
2. $\frac{(n-p-1)\hat{\sigma}^2}{\sigma^2} = \frac{S_e^2}{\sigma^2} \sim \chi(n - p - 1)$,
3. $\hat{\epsilon}$ *is independent of* $\hat{\mathbf{y}}$,
4. $\hat{\sigma}^2$ *is independent of* $\hat{\beta}$.

Theorem

Define $S_e^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \cdots - \hat{\beta}_p x_{ip})^2$ *and* $\hat{\sigma}^2 = \frac{S_e^2}{n-p-1}$. *Let* $\epsilon \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2 I_n)$ *and* $rank(X) = p+1 < n$. *We have*

1. $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2(X^T X)^{-1})$,
2. $\frac{(n-p-1)\hat{\sigma}^2}{\sigma^2} = \frac{S_e^2}{\sigma^2} \sim \chi(n-p-1)$,
3. $\hat{\epsilon}$ *is independent of* $\hat{\mathbf{y}}$,
4. $\hat{\sigma}^2$ *is independent of* $\hat{\beta}$.

So we have

$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}\sqrt{c_{jj}}} \sim t_{n-p-1}$$

$$\frac{(\hat{\beta}-\beta)^T(\mathbf{X}^T\mathbf{X})(\hat{\beta}-\beta)/p}{\hat{\sigma}^2} \sim F_{p+1,n-p-1}$$

where $c_{00}, c_{11}, ..., c_{pp}$ are the diagonal elements of $(\mathbf{X}^T\mathbf{X})^{-1}$.