Chapter 2. Linear and Nonlinear Regression Models

Overview

- Regression models capture linear or nonlinear relations of one or more attribute variables with how one or more target variables
- 2.1 Linear Regression Models
- 2.2 Least-Squares Method and Maximum Likelihood Method of Parameter Estimation
- 2.3 Nonlinear Regression Models and Parameter Estimation
- 2.4 Software

2.1 Linear Regression Models

- A simple linear regression model

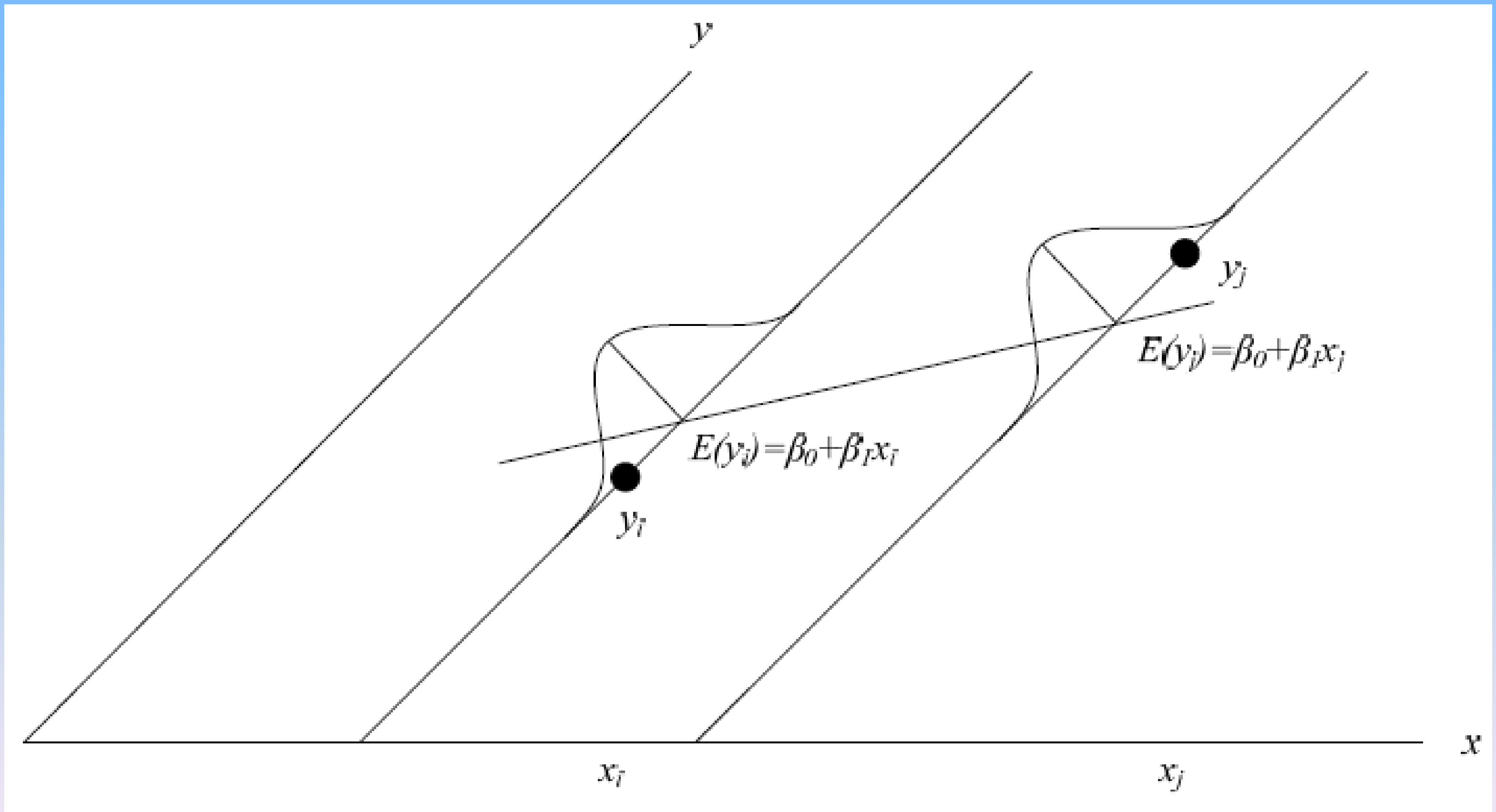
$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

  - Assumptions

1) $E(\varepsilon_i) = 0$, that is, the mean of ε_i is zero;
2) $\mathrm{var}(\varepsilon_i) = \sigma^2$, that is, and the variance of ε_i is σ^2;
3) $\mathrm{cov}(\varepsilon_i, \varepsilon_j)=0$ for $i \neq j$, that is, the covariance of ε_i and ε_j for any two different data observations, the i^{th} observation and the j^{th} observation, is zero.

2.1 Linear Regression Models

- A simple linear regression model



2.1 Linear Regression Models

- Examples of other linear regression models

$$y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_1 x_{i,p} + \varepsilon_i$$

$$y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,1}^k + \varepsilon_i$$

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 \log x_{i,1} x_{i,2} + \varepsilon_i$$

- A general form of linear regression models

$$y_i = \beta_0 + \beta_1 \Phi_1(x_{i,1}, \dots, x_{i,p}) + \cdots + \beta_k \Phi_k(x_{i,1}, \dots, x_{i,p}) + \varepsilon_i$$

2.2 Least-Squares Method and Maximum Likelihood Method of Parameter Estimation

- Parameter estimation
  - Estimate parameters β's to fit the regression model to a set of training data $(\boldsymbol{x}_i, y_i)$, $\boldsymbol{x}_i = (x_{i,1}, \ldots, x_{i,k})$, $i = 1, \ldots, n$

2.2 Least-Squares Method and Maximum Likelihood Method of Parameter Estimation

- Parameter estimation
  - Least-squares method
    - For the simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Look for $\hat{\beta}_0$ and $\hat{\beta}_1$ to minimize the sum of squared errors (*SSE*):

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

2.2 Least-Squares Method and Maximum Likelihood Method of Parameter Estimation

- Parameter estimation
  - Least-squares method
    - The partial derivatives of *SSE* with respective to $\hat{\beta}_0$ and $\hat{\beta}_1$ should be zero at the point where *SSE* is minimized

$$\frac{\partial SSE}{\partial \hat{\beta}_0} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

$$\frac{\partial SSE}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n} x_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

2.2 Least-Squares Method and Maximum Likelihood Method of Parameter Estimation

- Parameter estimation
  - Least-squares method
    - Solution

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

$$\hat{\beta}_0 = \frac{1}{n}\left(\sum_{i=1}^{n} y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i\right) = \bar{y} - \hat{\beta}_1 \bar{x}.$$

2.2 Least-Squares Method and Maximum Likelihood Method of Parameter Estimation

- Parameter estimation
  - Least-squares method
    - Example 2.1: fit a linear regression model to the space shuttle O-rings data set in Table 2.1, and determine the predicted target value for each observation using the linear regression model

TABLE 2.1

The Data set of O-rings with Stress Along with the Predicted Target Value from the Linear Regression

Instance	Launch Temperature	Number of O-rings with Stress
1	66	0
2	70	1
3	69	0
4	68	0
5	67	0
6	72	0
7	73	0
8	70	0
9	57	1
10	63	1
11	70	1
12	78	0
13	67	0
14	53	2
15	67	0
16	75	0
17	70	0
18	81	0
19	76	0
20	79	0
21	75	0
22	76	0
23	58	1

2.2 Least-Squares Method and Maximum Likelihood Method of Parameter Estimation

- Parameter estimation
  - Least-squares method
    - Example 2.1

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{-65.91}{1382.82} = -0.05$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 0.30 - (-.05)(69.57) = 3.78$$

$$y_i = 3.78 - 0.05x_i + \varepsilon_i$$

2.2 Least-Squares Method and Maximum Likelihood Method of Parameter Estimation

- Parameter estimation
  - Maximum likelihood method
    - Assumption: ε_i is normally distributed with the mean of zero and the constant, unknown variance of σ^2, denoted by $N(0, \sigma^2)$

2.2 Least-Squares Method and Maximum Likelihood Method of Parameter Estimation

- Parameter estimation
  - Maximum likelihood method
    - The assumption that ε_i's are independent $N(0, \sigma^2)$ gives the normal distribution of y_i

$$E(y_i) = \beta_0 + \beta_1 x_i$$

$$var(y_i) = \sigma^2$$

$$f(y_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y_i - E(y_i)}{\sigma}\right)^2} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y_i - \beta_0 - \beta_1 x_i}{\sigma}\right)^2}$$

2.2 Least-Squares Method and Maximum Likelihood Method of Parameter Estimation

- Parameter estimation
  - Maximum likelihood method
    - y_i's are independent, and the likelihood of observing y_1, …, y_n:

$$L(\beta_0, \beta_1, \sigma) = \prod_{i=1}^{n} \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{-\frac{1}{2}\left(\frac{y_i - \beta_0 - \beta_1 x_i}{\sigma}\right)^2}$$

2.2 Least-Squares Method and Maximum Likelihood Method of Parameter Estimation

- Parameter estimation
  - Maximum likelihood method
    - $\hat{\beta}_0, \hat{\beta}_1$ *and* $\hat{\sigma}^2$, which maximize the likelihood function, are obtained:

$$\frac{\partial lnL(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2)}{\partial \hat{\beta}_0} = \frac{1}{\hat{\sigma}^2} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\frac{\partial lnL(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2)}{\partial \hat{\beta}_1} = \frac{1}{\hat{\sigma}^2} \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\frac{\partial lnL(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2)}{\partial \hat{\sigma}^2} = -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 0$$

2.2 Least-Squares Method and Maximum Likelihood Method of Parameter Estimation

- Parameter estimation
  - Maximum likelihood method
    - $\hat{\beta}_0, \hat{\beta}_1$ *and* $\hat{\sigma}^2$, which maximize the likelihood function, are obtained by solving the following:

$$\frac{\partial lnL(\hat{\beta}_0,\hat{\beta}_1,\hat{\sigma}^2)}{\partial \hat{\beta}_0} = \frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\frac{\partial lnL(\hat{\beta}_0,\hat{\beta}_1,\hat{\sigma}^2)}{\partial \hat{\beta}_1} = \frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\frac{\partial lnL(\hat{\beta}_0,\hat{\beta}_1,\hat{\sigma}^2)}{\partial \hat{\sigma}^2} = -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 0$$

2.2 Least-Squares Method and Maximum Likelihood Method of Parameter Estimation

- Parameter estimation
  - Maximum likelihood method

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

$$\hat{\beta}_0 = \frac{1}{n}\left(\sum_{i=1}^{n} y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i\right) = \bar{y} - \hat{\beta}_1 \bar{x}.$$

2.2 Nonlinear Regression Models and Parameter Estimation

- Nonlinear regression models are nonlinear in model parameters

$$y_i = f(\boldsymbol{x}_i, \boldsymbol{\beta}) + \varepsilon_i$$

$$\boldsymbol{x}_i = \begin{bmatrix} 1 \\ x_{i,1} \\ \vdots \\ x_{i,p} \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

2.2 Nonlinear Regression Models and Parameter Estimation

- Examples of nonlinear regression models
  - An exponential regression model

$$y_i = \beta_0 + \beta_1 e^{\beta_2 x_i} + \varepsilon_i$$

  - A logistic regression model

$$y_i = \frac{\beta_0}{1+\beta_1 e^{\beta_2 x_i}} + \varepsilon_i$$

2.2 Nonlinear Regression Models and Parameter Estimation

- Parameter estimation
  - Equations derived from least-squares method and maximum likelihood method to estimate the parameters of a nonlinear regression model do not have analytical solutions
  - Numerical search methods, using the Gauss-Newton method and the gradient decent search, are used to search for the solution to the equations

2.4 Software

- Statistica (http://www.statsoft.com),
- SAS (http://www.sas.com),
- SPSS (http://www.ibm/com/software/analytics/spss/).