ML
LECTURE-3

BY
Dr. Ramesh Kumar Thakur
Assistant Professor (II)
School Of Computer Engineering

Criterion for minimization of error

- In regression, we would like to write the numeric output y, called the dependent variable, as a function of the input x, called the independent variable.
- We **assume that the output is the sum of a function f(x) of the input and some random error denoted by ϵ :**

$$y = f(x) + \epsilon.$$

- Here the **function f(x) is unknown and we would like to approximate it by some estimator g(x, θ) containing a set of parameters θ.**
- We **assume that the random error ϵ follows normal distribution with mean 0.**
- Let $x_1, \ldots, x_n$ **be a random sample of observations of the input variable x and** $y_1, \ldots, y_n$ **the corresponding observed values of the output variable y.**
- Using the assumption that the error ϵ follows normal distribution, we can **apply the method of maximum likelihood estimation to estimate the values of the parameter θ.**
- It can be shown that the values of θ which maximizes the likelihood function are the values of θ that minimizes the following sum of squares:

$$E(\theta) = (y_1 - g(x_1, \theta))^2 + \cdots + (y_n - g(x_n, \theta))^2$$

- The **method of finding the value of θ as that value of θ that minimizes E(θ) is known as the ordinary least squares method.**

Criterion for minimization of error

x	x_1	x_2	$\cdots$	x_n
y	y_1	y_2	$\cdots$	y_n

Table 7.1: Data set for simple linear regression

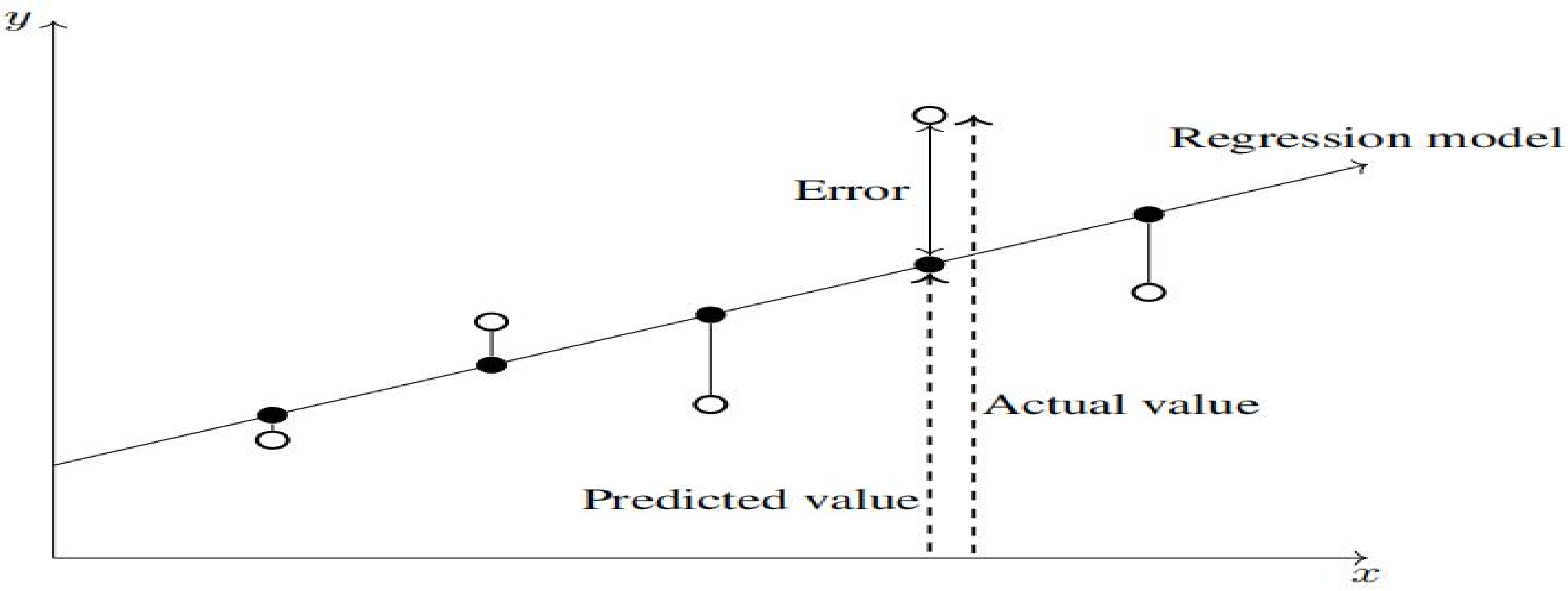


Figure 7.1: Errors in observed values

Simple linear regression

- Let x be the independent predictor variable and y the dependent variable.
- Assume that we have a set of observed values of x and y. A simple linear regression model defines the relationship between x and y using a line defined by an equation in the following form:

$$y = a + bx$$

- To determine the optimal estimates of α and β, an estimation method known as Ordinary Least Squares (OLS).

- **The OLS method**
- In the OLS method, the values of y-intercept and slope are chosen such that they minimize the sum of the squared errors; that is, the sum of the squares of the vertical distance between the predicted y-value and the actual y-value (see Figure 7.1). Let $\hat{y}_i$ be the predicted value of y_i

- Then the sum of squares of errors is given by
- $\mathrm{E} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

  $= \sum_{i=1}^{n} [y_i - (a + bx_i)]^2$

- So we are required to find the values of **a** and **b** such that E is minimum.

Solution of Simple linear regression using OLS

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

  $= \sum_{i=1}^{n} [y_i - (a + bx_i)]^2$
- To solve the above equation we have to take two partial derivations as below:
- $\frac{\partial E}{\partial a} = 0$ ------(i) and $\frac{\partial E}{\partial b} = 0$ ------(ii)
- By solving eq(i)

=> $2\sum_{i=1}^{n} [y_i - a - bx_i](-1) = 0$

=> $-2\sum_{i=1}^{n} y_i + 2a\sum_{i=1}^{n} 1 + 2b\sum_{i=1}^{n} x_i = 0$

=> $-\sum_{i=1}^{n} y_i + an + b\sum_{i=1}^{n} x_i = 0$

=> $an = \sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i$

=> $a = \frac{1}{n}\sum_{i=1}^{n} y_i - b\frac{1}{n}\sum_{i=1}^{n} x_i$

=> $\boldsymbol{a = \overline{y} - b\overline{x}}$

where $\overline{\boldsymbol{y}} = \frac{1}{n}\sum_{i=1}^{n} y_i$ (mean of values of y), $\overline{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (mean of values of x)

Solution of Simple linear regression using OLS

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

  $= \sum_{i=1}^{n} [y_i - (a + bx_i)]^2$

- To solve the above equation we have to take two partial derivations as below:

- $\frac{\partial E}{\partial a} = 0$ -------(i) and $\frac{\partial E}{\partial b} = 0$ -------(ii)

- By solving eq(ii)

$\Rightarrow 2\sum_{i=1}^{n} [y_i - a - bx_i](-x_i) = 0$

$\Rightarrow -2\sum_{i=1}^{n} x_i y_i + 2a\sum_{i=1}^{n} x_i + 2b\sum_{i=1}^{n} x_i{}^2 = 0$

$\Rightarrow -\sum_{i=1}^{n} x_i y_i + a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i{}^2 = 0$

$\Rightarrow -\sum_{i=1}^{n} x_i y_i + (\bar{y} - b\bar{x})\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i{}^2 = 0$

$\Rightarrow b(\sum_{i=1}^{n} x_i{}^2 - \bar{x}\sum_{i=1}^{n} x_i) = \sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i$

$\Rightarrow b[\sum_{i=1}^{n} x_i\{\sum_{i=1}^{n} (x_i - \bar{x})\}] = \sum_{i=1}^{n} x_i\{\sum_{i=1}^{n} (y_i - \bar{y})\}$

$\Rightarrow \boldsymbol{b = \frac{\sum_{i=1}^{n} (y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})}}$

By multiplying $\frac{1}{n-1}\{\sum_{i=1}^{n} (x_i - \bar{x})\}$ in the numerator and denominator of RHS

$\Rightarrow b = \frac{\frac{1}{n-1}\{\sum_{i=1}^{n} (x_i - \bar{x})\}\sum_{i=1}^{n} (y_i - \bar{y})}{\frac{1}{n-1}\{\sum_{i=1}^{n} (x_i - \bar{x})\}\sum_{i=1}^{n} (x_i - \bar{x})} = \boldsymbol{\frac{Cov(x,y)}{Var(x)}}$

Solution of Simple linear regression using OLS

Formulas to find a and b

- Recall that the means of x and y are given by

$$\bar{x} = \frac{1}{n}\sum x_i$$
$$\bar{y} = \frac{1}{n}\sum y_i$$

- and also that the variance of x is given by

$$\mathrm{Var}(x) = \frac{1}{n-1}\sum (x_i - \bar{x})^2$$

- The covariance of x and y, denoted by Cov(x, y) is defined as

$$\mathrm{Cov}(x, y) = \frac{1}{n-1}\sum (x_i - \bar{x})(y_i - \bar{y})$$

- It can be shown that the values of a and b can be computed using the following formulas:

$$b = \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)}$$
$$a = \bar{y} - b\bar{x}$$

Variations of Least Square for Solving Simple linear regression

- **Remarks:**
- It is interesting to note why the least squares method discussed above is christened as “ordinary” least squares method.
- Several different variants of the least squares method have been developed over the years. For example, in the weighted least squares method, the coefficients **a** and **b** are estimated such that the weighted sum of squares of errors.

$$E = \sum_{i=1}^{n} w_i [y_i - (a + bx_i)]^2$$

- for some positive constants $w_1, \ldots, w_n$, is minimum. There are also methods known by the names **generalised least squares method, partial least squares method, total least squares method**, etc.
- The OLS method has a long history. The method is usually credited to **Carl Friedrich Gauss (1795),** but it was first published by **Adrien-Marie Legendre (1805).**

Simple linear regression Example

- **Example:**
- Obtain a linear regression for the data in Table 7.2 assuming that y is the independent variable.

x	1.0	2.0	3.0	4.0	5.0
y	1.00	2.00	1.30	3.75	2.25

Table 7.2: Example data for simple linear regression

- **Solution:**
- In the usual notations of simple linear regression, we have

$$
\begin{aligned}
n &= 5 \\
\bar{x} &= \frac{1}{5}(1.0 + 2.0 + 3.0 + 4.0 + 5.0) \\
&= 3.0 \\
\bar{y} &= \frac{1}{5}(1.00 + 2.00 + 1.30 + 3.75 + 2.25) \\
&= 2.06 \\
\mathrm{Cov}\,(x, y) &= \frac{1}{4}[(1.0 - 3.0)(1.00 - 2.06) + \cdots + (5.0 - 3.0)(2.25 - 2.06)] \\
&= 1.0625 \\
\mathrm{Var}\,(x) &= \frac{1}{4}[(1.0 - 3.0)^2 + \cdots + (5.0 - 3.0)^2] \\
&= 2.5 \\
b &= \frac{1.0625}{2.5} \\
&= 0.425 \\
a &= 2.06 - 0.425 \times 3.0 \\
&= 0.785
\end{aligned}
$$

Simple linear regression Example

- Therefore, the linear regression model for the data is

$$y = 0.785 + 0.425x.$$

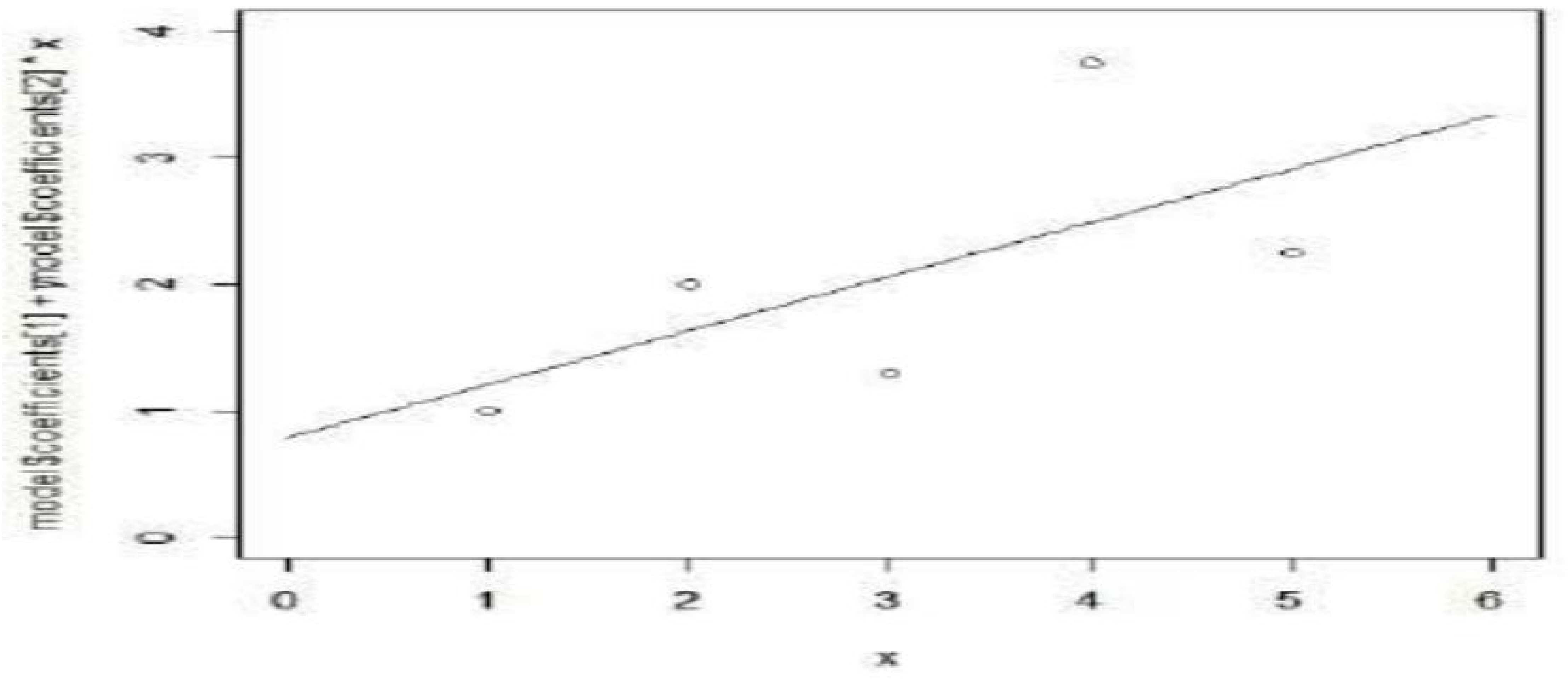
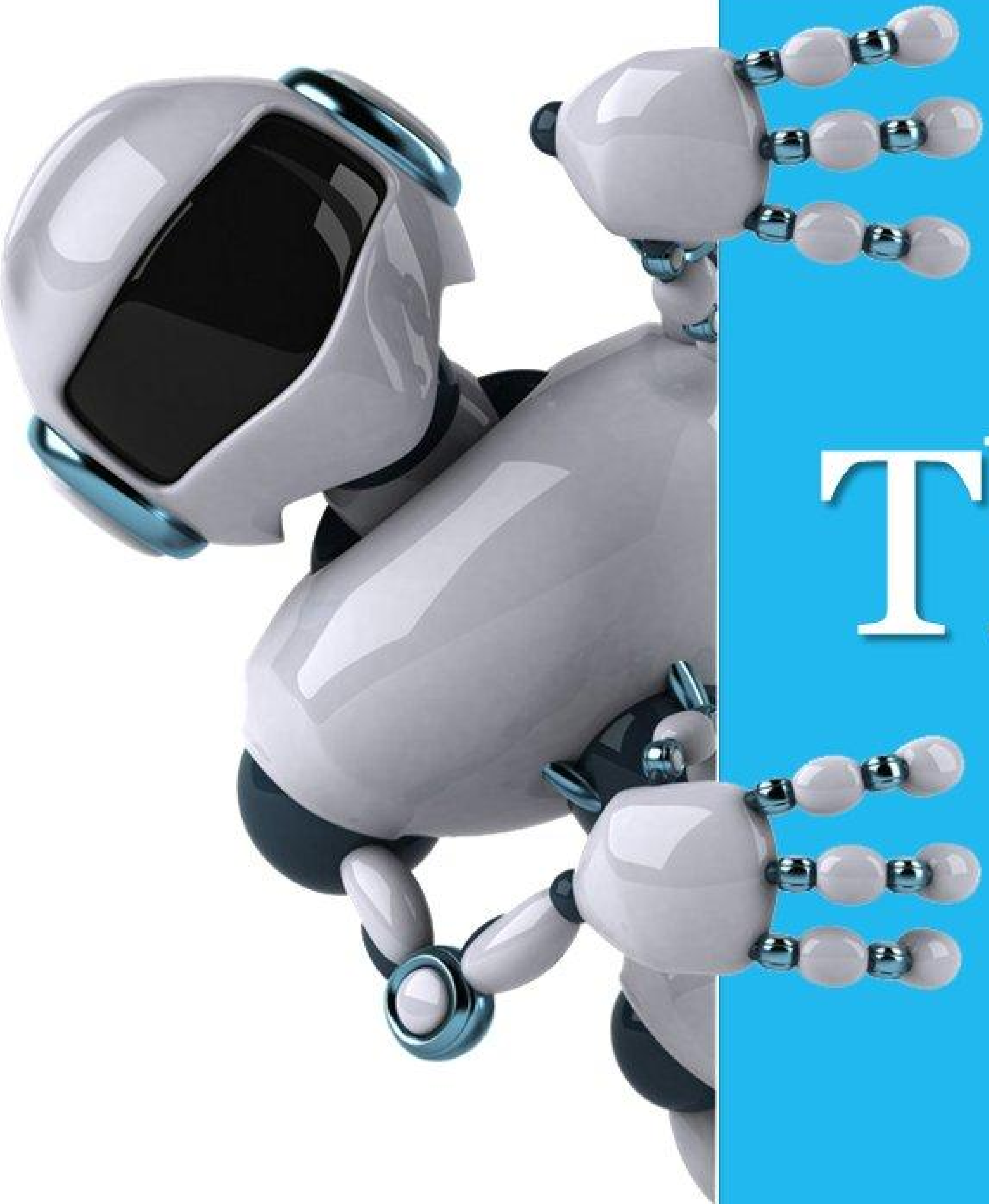


Figure 7.2: Regression model for Table 7.2



Thank you