Math 4543: Numerical Methods

Lecture 10 — Linear Regression

Syed Rifat Raiyan
Lecturer
Department of Computer Science & Engineering
Islamic University of Technology, Dhaka, Bangladesh
Email: rifatraiyan@iut-dhaka.edu

Lecture Plan

The agenda for today

- Understand the concept of Regression Analysis
- What is Linear Regression?
- Explore the appropriateness of different optimization criteria for the regression model
- Uniqueness of the Least-Square-Errors criterion
- Derive the slope of a Zero-intercept Linear Regression model

Regression Analysis

What is a regression model?

In statistical modeling, regression analysis is a set of statistical processes for ***estimating*** the relationships between a dependent variable and one or more independent variables.

The problem statement for a regression model is as follows. Given n data pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, best fit $y = f(x)$ to the data (Figure 1).

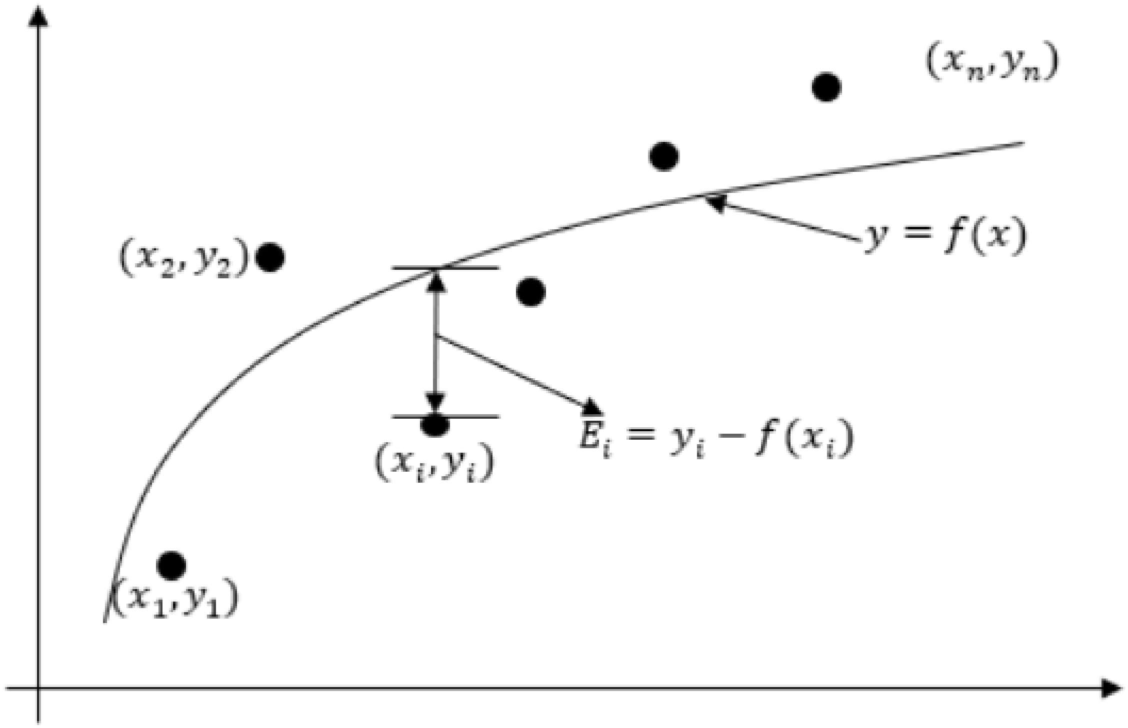


Figure 1. A general regression model for discrete y vs. x data

Linear Regression

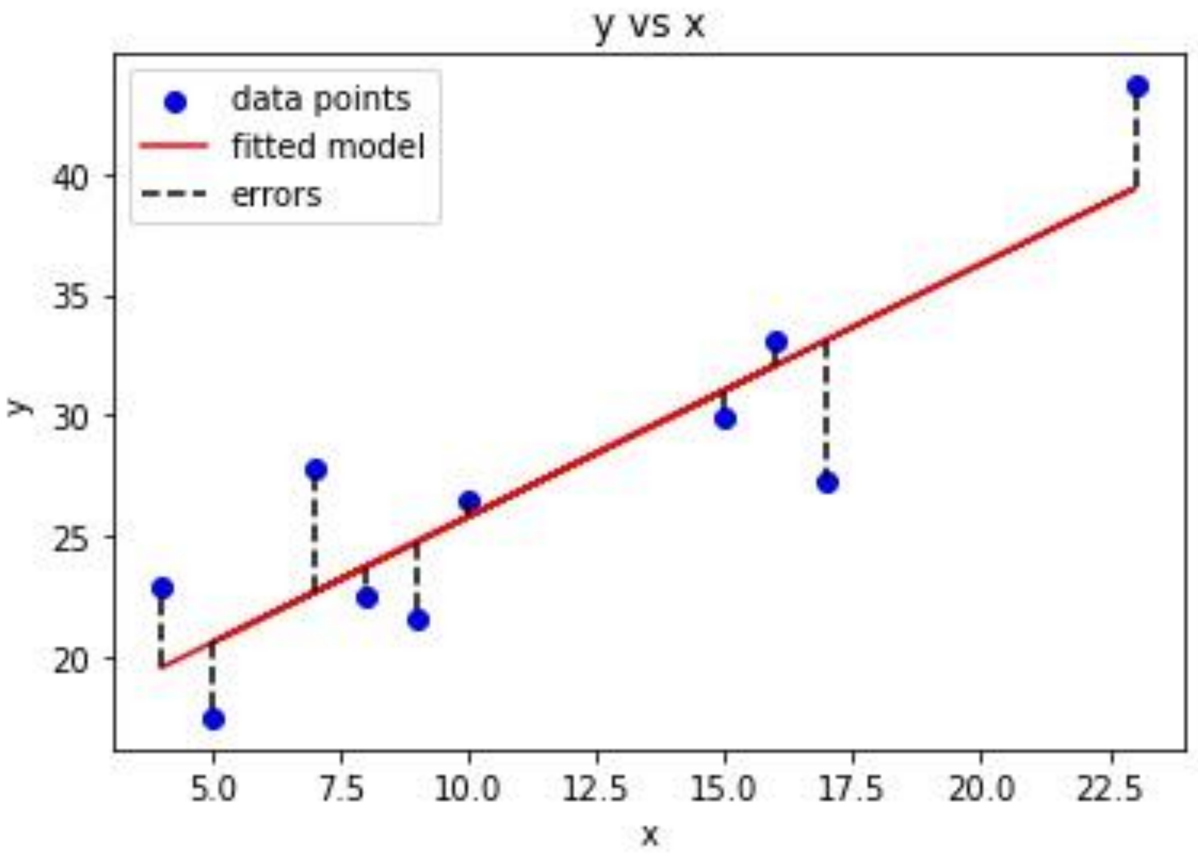
What is it?

In linear regression, the relationships are modeled using ***linear predictor functions*** whose unknown model parameters (slope and *y*-intercept) are estimated from the data.

Linear regression is the most popular regression model. In this model, we wish to predict response to n data points $(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_n, y_n)$ by a regression model given by

$$y = a_0 + a_1 x \qquad (1)$$

where a_0 and a_1 are the constants of the regression model.



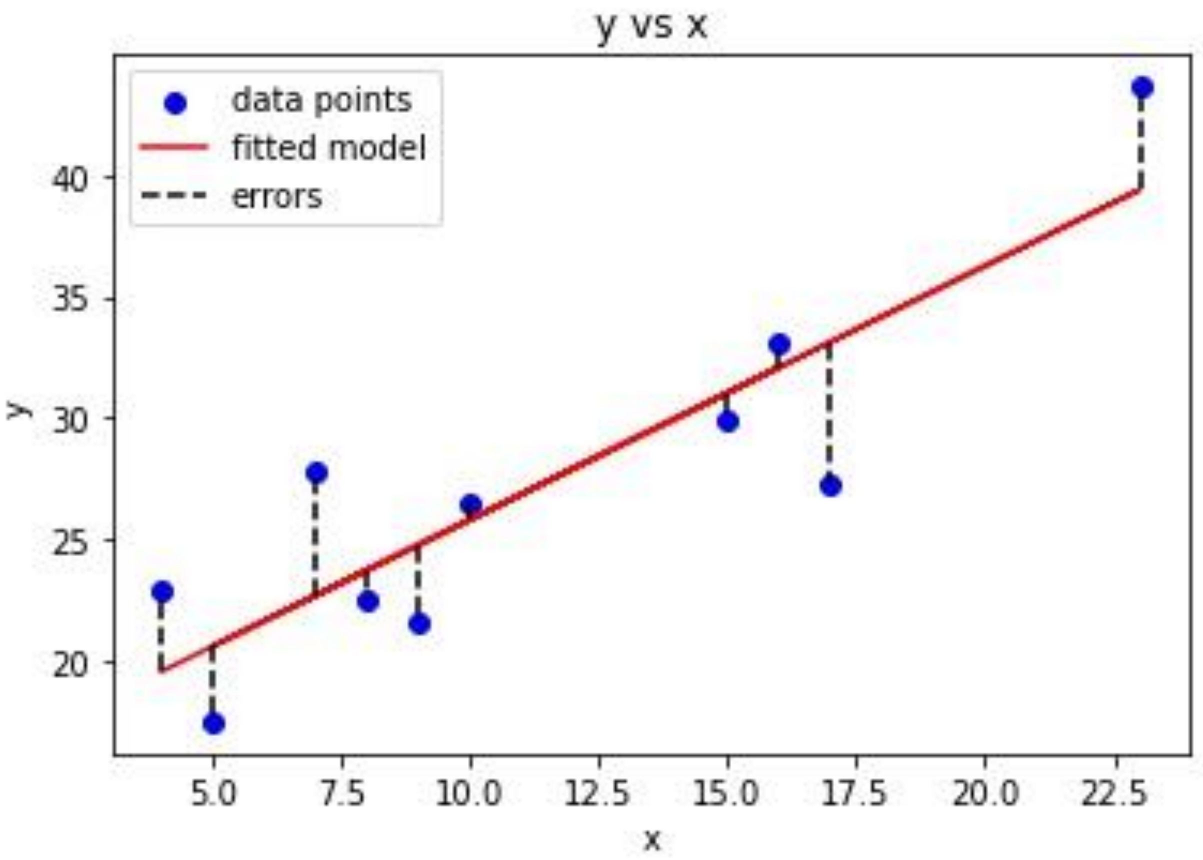
Linear Regression

How to quantify the *goodness* of fit?

A measure of goodness of fit, that is, how well $a_0 + a_1 x$ predicts the response variable y is the magnitude of the residual E_i at each of the n data points.

$$E_i = y_i - (a_0 + a_1 x_i) \qquad (2)$$

Ideally, if all the residuals E_i are zero, one has found an equation in which all the points lie on the model. Thus, minimization of the residuals is an objective of obtaining regression coefficients.



Linear Regression

Choosing the right criterion for optimization

$$E_i = y_i - (a_0 + a_1 x_i) \qquad (2)$$

The most popular method to minimize the residual is the least-squares method, where the estimates of the constants of the models are chosen such that the sum of the squared residuals is minimized, that is, minimize

$$S_r = \sum_{i=1}^{n} {E_i}^2 \qquad (3)$$

Why minimize the sum of the square of the residuals, S_r?

$$\boxed{\text{Why not } \sum_{i=1}^{n} E_i \text{ or } \sum_{i=1}^{n} |E_i| \text{ ?}}$$

Linear Regression

The case *against* the $\sum_{i=1}^{n} E_i$ criterion

Table 1 Data points.

x	y
2.0	4.0
3.0	6.0
2.0	6.0
3.0	8.0

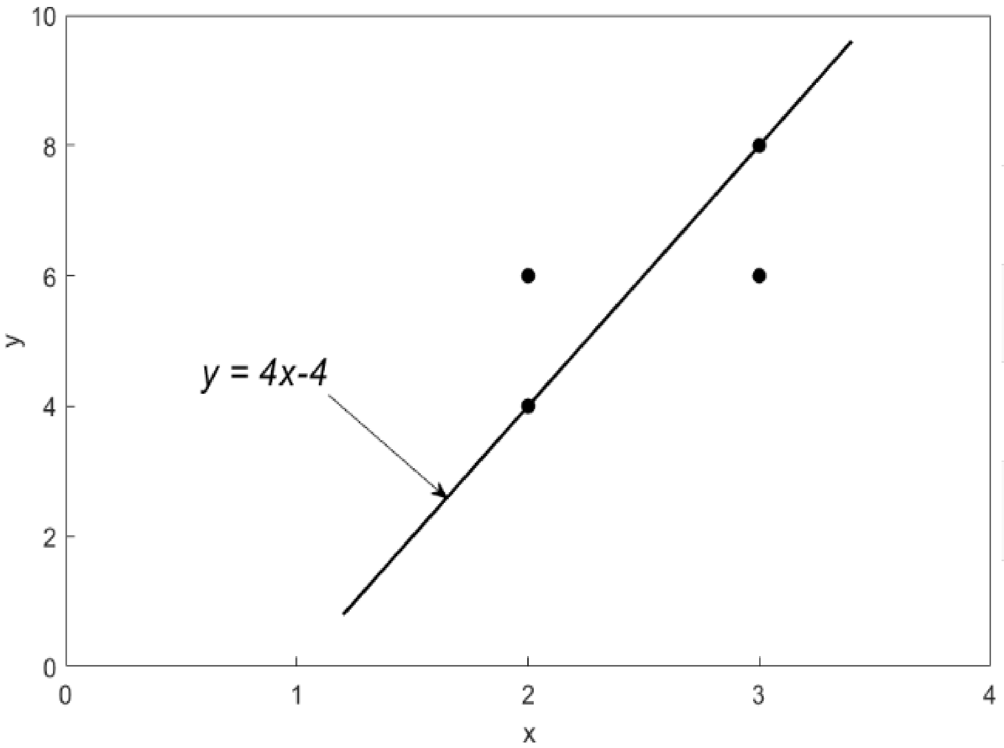
To explain this data by a straight line regression model,

$$y = a_0 + a_1 x \qquad (4)$$

Let us use minimizing $\sum_{i=1} E_i$ as a criterion to find a_0 and a_1. Assume randomly that

$$y = 4x - 4 \qquad (5)$$

as the resulting regression model (Figure 2).



x	y	$y_{predicted}$	$E = y - y_{pred}$
2.0	4.0	4.0	0.0
3.0	6.0	8.0	−2.0
2.0	6.0	4.0	2.0
3.0	8.0	8.0	0.0

$$\sum_{i=1}^{4} E_i = 0$$

Linear Regression

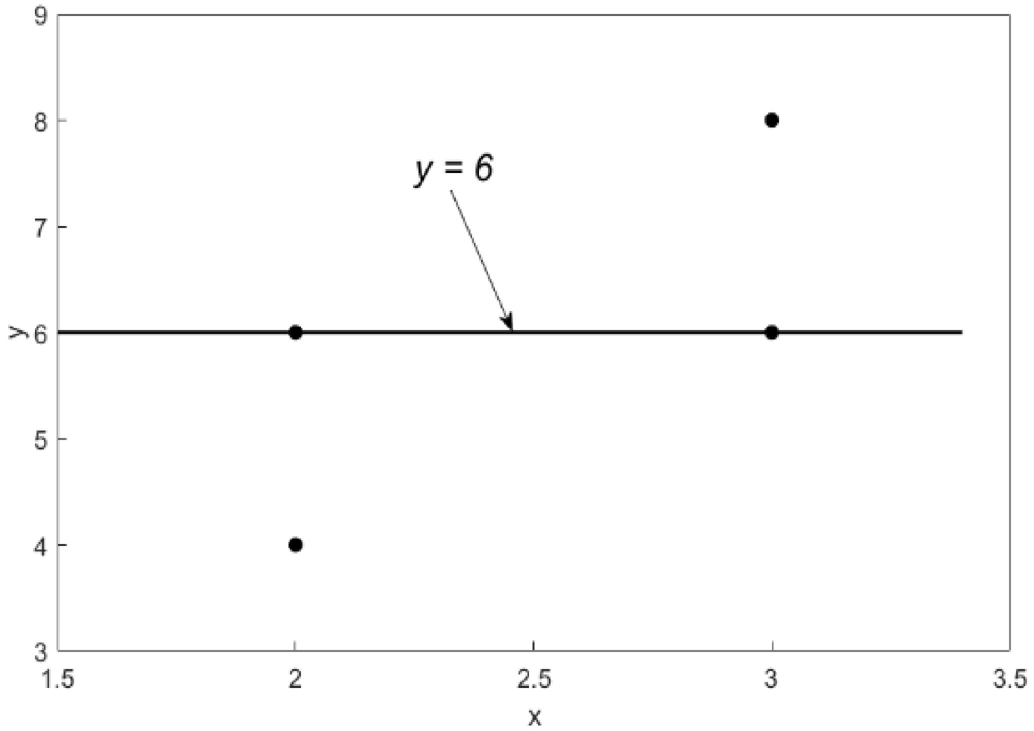
The case *against* the $\sum_{i=1}^{n} E_i$ criterion

So does this give us the smallest possible sum of residuals? For this data, it does as

$$\sum_{i=1}^{4} E_i = 0,$$ and it cannot be made any smaller. But does it give unique values for the parameters of the regression model? No, because, for example, a straight-line model (Figure 3)

$$y = 6 \qquad (6)$$

also gives $\sum_{i=1}^{4} E_i = 0$ as shown in Table 3.



x	y	$y_{\text{predicted}}$	$E = y - y_{pred}$
2.0	4.0	6.0	-2.0
3.0	6.0	6.0	0.0
2.0	6.0	6.0	0.0
3.0	8.0	6.0	2.0

$$\sum_{i=1}^{4} E_i = 0$$

Non-unique regression models. So, bad criterion!

Linear Regression

The case *against* the $\sum_{i=1}^{n} |E_i|$ criterion

Non-unique regression models. Also bad criterion!

Table 1 Data points.

x	y
2.0	4.0
3.0	6.0
2.0	6.0
3.0	8.0

To explain this data by a straight line regression model,

$$y = a_0 + a_1 x \tag{4}$$

You may think that the reason the criterion of minimizing $\sum_{i=1}^{n} E_i$ does not work is because negative residuals cancel with positive residuals. So, is minimizing the sum of absolute values of the residuals, that is, $\sum_{i=1}^{n} |E_i|$ better? Let us look at the same example data given in Table 1. For the regression

$$y = 4x - 4$$

x	y	$y_{predicted}$	$E = y - y_{pred}$
2.0	4.0	4.0	0.0
3.0	6.0	8.0	2.0
2.0	6.0	4.0	2.0
3.0	8.0	8.0	0.0

$$\sum_{i=1}^{4} |E_i| = 4$$

$$y = 6$$

x	y	$y_{predicted}$	$E = y - y_{pred}$
2.0	4.0	6.0	2.0
3.0	6.0	6.0	0.0
2.0	6.0	6.0	0.0
3.0	8.0	6.0	2.0

$$\sum_{i=1}^{4} |E_i| = 4$$

No other straight-line model that you may choose for this data has $\sum_{i=1}^{4} |E_i| < 4$.

Linear Regression

The case *for* the $\sum_{i=1}^{n} E_i^2$ criterion

To get a unique regression model, the least-squares criterion where we minimize the sum of the square of the residuals

$$S_r = \sum_{i=1}^{n} E_i^{2}$$
$$= \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 \qquad (7)$$

is recommended.

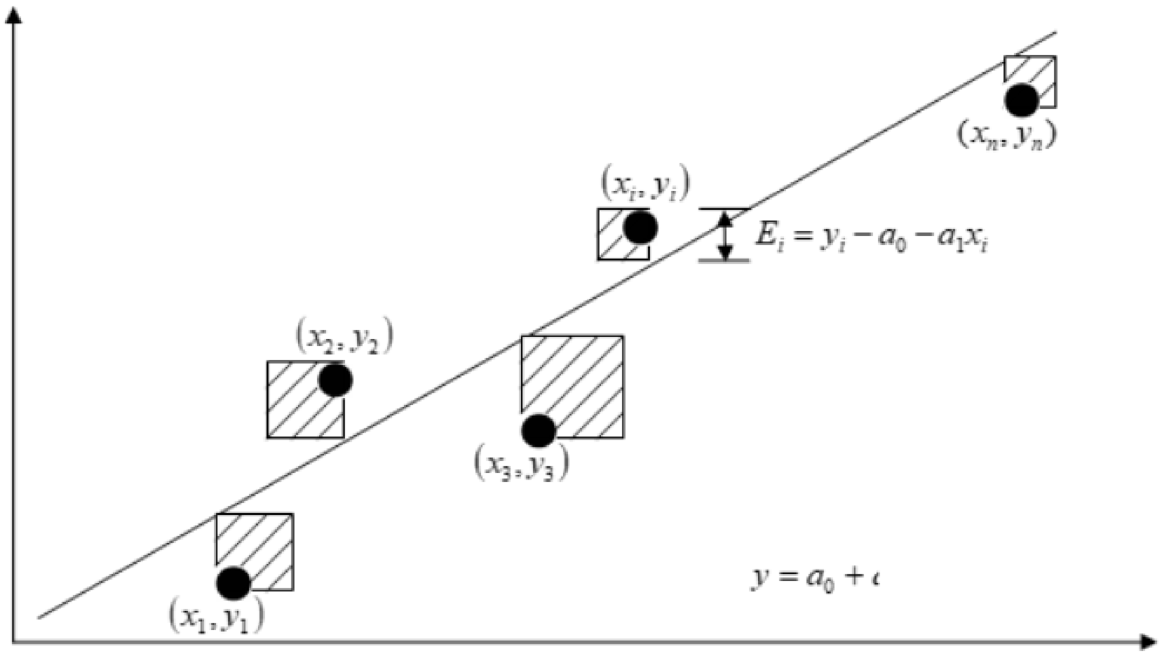


Figure 1. Linear regression of y vs. x data showing residuals and square of residual at a typical point, x_i

To find a_0 and a_1, we need to calculate where the sum of the square of the residuals, S_r is the absolute minimum. We start this process of finding the absolute minimum first by

1) taking the partial derivative of S_r with respect to a_0 and a_1 and set them equal to zero, and

2) conducting the second derivative test.

Linear Regression

The case *for* the $\sum_{i=1}^{n} E_i^2$ criterion

Taking the partial derivative of S_r with respect to a_0 and a_1 and set them equal to zero

$$\frac{\partial S_r}{\partial a_0} = 2\sum_{i=1}^{n}(y_i - a_0 - a_1 x_i)(-1) = 0 \qquad (2)$$

$$\frac{\partial S_r}{\partial a_1} = 2\sum_{i=1}^{n}(y_i - a_0 - a_1 x_i)(-x_i) = 0 \qquad (3)$$

Dividing both sides by 2 and expanding the summations in Equations (2) and (3) gives,

$$-\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} a_0 + \sum_{i=1}^{n} a_1 x_i = 0$$

$$-\sum_{i=1}^{n} y_i x_i + \sum_{i=1}^{n} a_0 x_i + \sum_{i=1}^{n} a_1 x_i^2 = 0$$

Noting that

$$\sum_{i=1}^{n} a_0 = a_0 + a_0 + \ldots + a_0 = n a_0$$

$$n a_0 + a_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \qquad (4)$$

$$a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \qquad (5)$$

Solving the above two simultaneous linear equations (4) and (5) gives

$$\boxed{a_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}} \qquad (6)$$

$$\boxed{a_0 = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}} \qquad (7)$$

Linear Regression

The case *for* the $\sum_{i=1}^{n} E_i^2$ criterion

Redefining

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} \qquad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

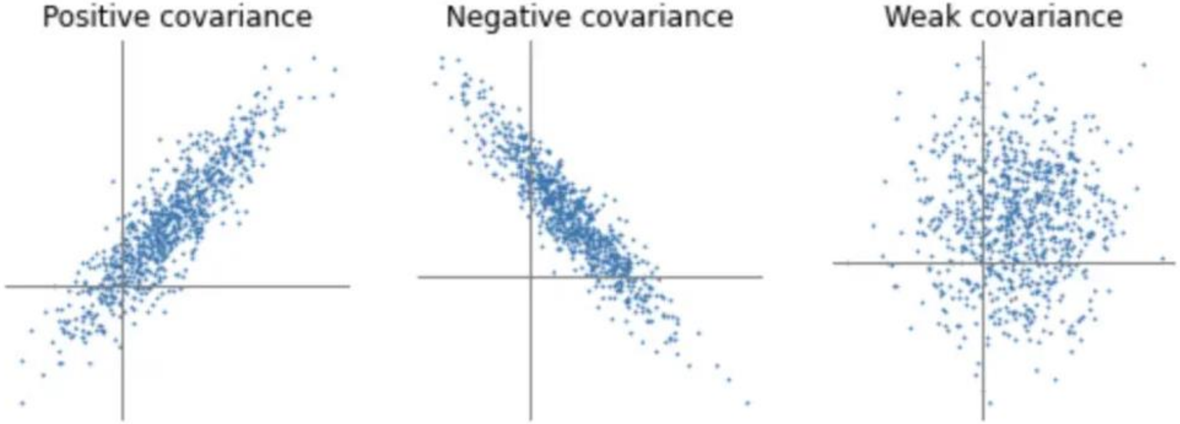
$$S_{xx} = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \qquad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

we can also rewrite the constants a_0 and a_1 from (6) and (7) as

$$a_1 = \frac{S_{xy}}{S_{xx}} \tag{8}$$

$$a_0 = \bar{y} - a_1\bar{x} \tag{9}$$

A small detour on *Variance* and *Covariance*



$Cov(X, Y) > 0$ $\quad$ $Cov(X, Y) < 0$ $\quad$ $Cov(X, Y) \approx 0$

Variance is the measure of dispersion of data points from the mean.

Covariance is the measure of how 2 random variables vary with respect to each other.

Linear Regression

The case *for* the $\sum_{i=1}^{n} E_i^2$ criterion

Putting the first derivative equations equal to zero only gives us a critical point. For a general function, it could be a local minimum, a local maximum, a saddle point, or none of the previous. The second derivative test, though, given in the shows that it is a local minimum.

If you have a function $f(x, y)$ and we found a critical point (a, b) from the first derivative test, then (a, b) is a minimum point if

$$\frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 > 0, \text{ and} \qquad (A.6)$$

$$\frac{\partial^2 f}{\partial x^2} > 0 \text{ or } \frac{\partial^2 f}{\partial y^2} > 0 \qquad (A.7)$$

$$\frac{\partial^2 S_r}{\partial a_0^2} = -2\sum_{i=1}^{n} -1$$
$$= 2n \qquad (A.10)$$

$$\frac{\partial^2 S_r}{\partial a_1^2} = 2\sum_{i=1}^{n} x_i^2 \qquad (A.11)$$

$$\frac{\partial^2 S_r}{\partial a_0 \partial a_1} = 2\sum_{i=1}^{n} x_i \qquad (A.12)$$

$$\frac{\partial^2 S_r}{\partial a_0^2}\frac{\partial^2 S_r}{\partial a_1^2} - \left(\frac{\partial^2 S_r}{\partial a_0 \partial a_1}\right)^2 = (2n)\left(2\sum_{i=1}^{n} x_i^2\right) - \left(2\sum_{i=1}^{n} x_i\right)^2$$
$$= 4\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]$$
$$= 4\sum_{\substack{i=1\\ i<j}}^{n} (x_i - x_j)^2 > 0 \qquad (A.13)$$

Unique regression model. Good criterion!

Linear Regression

An example

Tip: Use the *"Statistics"* mode of your calculator. (**Mode 6** for **Casio fx-991EX**)

The torque T needed to turn the torsional spring of a mousetrap through an angle, θ is given below

Angle, θ	*Torque*, T
Radians	$\mathrm{N \cdot m}$
0.698132	0.188224
0.959931	0.209138
1.134464	0.230052
1.570796	0.250965
1.919862	0.313707

Find the constants k_1 and k_2 of the regression model

$$T = k_1 + k_2\theta \qquad (E1.1)$$

Solution

For the linear regression model,

$$T = k_1 + k_2\theta$$

the constants of the regression model are given by

$$k_2 = \frac{n\sum_{i=1}^{5}\theta_i T_i - \sum_{i=1}^{5}\theta_i \sum_{i=1}^{5} T_i}{n\sum_{i=1}^{5}\theta_i^2 - \left(\sum_{i=1}^{5}\theta_i\right)^2} \qquad (E1.2)$$

$$k_1 = \bar{T} - k_2\bar{\theta} \qquad (E1.3)$$

Linear Regression

An example

Table 2 shows the summations needed for the calculation of the above two constants k_1 and k_2 of the regression model.

Table 2. Tabulation of data for calculation of needed summations.

i	θ	T	θ^2	$T\theta$
	$Radians$	$N \cdot m$	$Radians^2$	$N \cdot m$
1	0.698132	0.188224	4.87388×10^{-1}	1.31405×10^{-1}
2	0.959931	0.209138	9.21468×10^{-1}	2.00758×10^{-1}
3	1.134464	0.230052	1.2870	2.60986×10^{-1}
4	1.570796	0.250965	2.4674	3.94215×10^{-1}
5	1.919862	0.313707	3.6859	6.02274×10^{-1}
$\sum_{i=1}^{5}$	6.2831	1.1921	8.8491	1.5896

$$k_2 = \frac{n\sum_{i=1}^{5}\theta_i T_i - \sum_{i=1}^{5}\theta_i \sum_{i=1}^{5} T_i}{n\sum_{i=1}^{5}\theta_i^2 - \left(\sum_{i=1}^{5}\theta_i\right)^2}$$

$$= \frac{5(1.5896) - (6.2831)(1.1921)}{5(8.8491) - (6.2831)^2}$$

$$= 9.6091 \times 10^{-2}\text{N-m/rad}$$

$$\bar{T} = \frac{\sum_{i=1}^{5} T_i}{n} = \frac{1.1921}{5} = 2.3842 \times 10^{-1} N - m$$

$$\bar{\theta} = \frac{\sum_{i=1}^{5}\theta_i}{n} = \frac{6.2831}{5} = 1.2566\ radians$$

$$k_1 = \bar{T} - k_2\bar{\theta}$$

$$= 2.3842 \times 10^{-1} - (9.6091 \times 10^{-2})(1.2566)$$

$$= 1.1767 \times 10^{-1}\text{N-m}$$

Linear Regression

An example

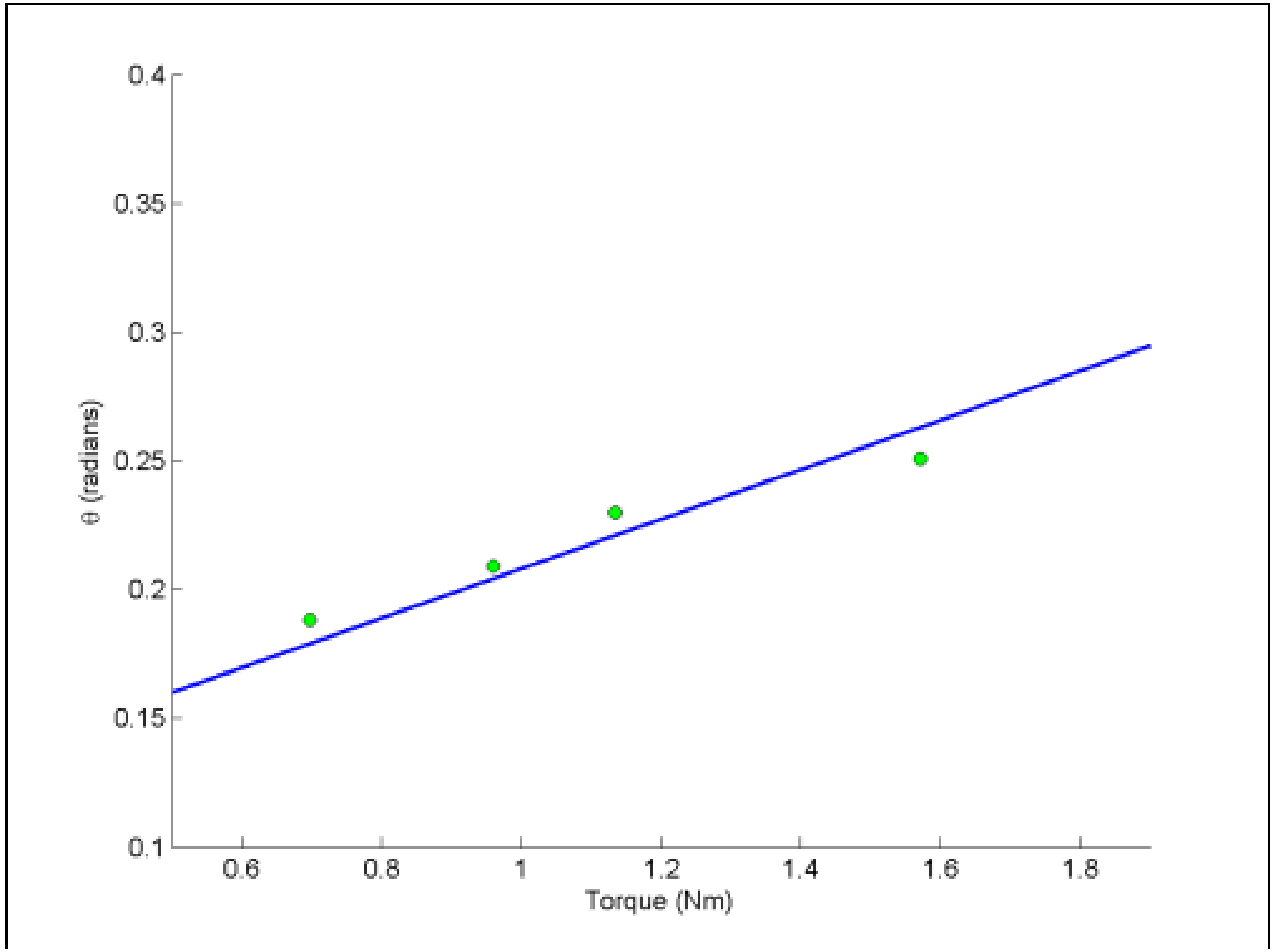


Figure 4 Linear regression of torque vs. angle data

$$k_2 = \frac{n\sum_{i=1}^{5}\theta_i T_i - \sum_{i=1}^{5}\theta_i \sum_{i=1}^{5} T_i}{n\sum_{i=1}^{5}\theta_i^2 - \left(\sum_{i=1}^{5}\theta_i\right)^2}$$

$$= \frac{5(1.5896) - (6.2831)(1.1921)}{5(8.8491) - (6.2831)^2}$$

$$= 9.6091 \times 10^{-2}\text{N-m/rad}$$

$$k_1 = \bar{T} - k_2\bar{\theta}$$

$$= 2.3842 \times 10^{-1} - (9.6091 \times 10^{-2})(1.2566)$$

$$= 1.1767 \times 10^{-1}\text{N-m}$$

Mini Quiz

Forcing the regression line to pass through the origin

If we use $y = a_1 x$ instead of $y = a_0 + a_1 x$ will the same formula **doesn't** seem to work. Why not?

$$a_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2}$$

$$a_1 = \frac{S_{xy}}{S_{xx}}$$

$$a_0 = \bar{y} - a_1 \bar{x}$$

Linear Regression

The zero y-intercept variant of the model

In this model, we wish to predict response to n data points $(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_n, y_n)$ by a regression model

$$y = a_1 x \qquad (1)$$

where a_1 is the only constant of the regression model.

A measure of goodness of fit, that is, how well $a_1 x$ predicts variable y is the sum of the square of the residuals, S_r

$$S_r = \sum_{i=1}^{n} E_i^{\,2} = \sum_{i=1}^{n} (y_i - a_1 x_i)^2 \qquad (2)$$

To find a_1, we look for the value of a_1 for which S_r is the absolute minimum.

We will begin by conducting the first derivative test. Take the derivative of Equation (2)

$$\frac{dS_r}{da_1} = 2\sum_{i=1}^{n} (y_i - a_1 x_i)(-x_i) = 0 \qquad (3)$$

Now putting $\frac{dS_r}{da_1} = 0$ gives $2\sum_{i=1}^{n} (y_i - a_1 x_i)(-x_i) = 0$

giving

$$-2\sum_{i=1}^{n} y_i x_i + 2\sum_{i=1}^{n} a_1 x_i^2 = 0$$

$$-2\sum_{i=1}^{n} y_i x_i + 2a_1\sum_{i=1}^{n} x_i^2 = 0$$

Solving the above equation for a_1 gives

$$\boxed{a_1 = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}} \qquad (4)$$

Linear Regression

Is the slope a_1 unique?

Let's conduct the second derivative test.

$$\frac{d^2 S_r}{d{a_1}^2} = \frac{d}{da_1}\left(2\sum_{i=1}^{n}(y_i - a_1 x_i)(-x_i)\right)$$

$$= \frac{d}{da_1}\sum_{i=1}^{n}(-2x_i y_i + 2a_1 {x_i}^2)$$

$$= \sum_{i=1}^{n} 2{x_i}^2 > 0 \qquad (5)$$

for at most one $x_i \neq 0$, which is a pragmatic assumption that all the x-values are not zero.

This inequality shows that the Equation (2) value of a_1 corresponds to a location of local minimum. Since the sum of the squares of the residuals, S_r is a continuous function of a_1, that S_r has only one point where $\frac{dS_r}{da_1} = 0$, and at that point, we have $\frac{d^2 S_r}{d{a_1}^2} > 0$, it corresponds not only to a local minimum but an absolute minimum as well. Hence, Equation (4) gives us the value of the constant, a_1 of the regression model, $y = a_1 x$.

Linear Regression

A zero y-intercept example

To find the longitudinal modulus of a composite material, the following data, as given in Table 1, is collected.

Table 1 Stress vs. strain data for a composite material.

Strain (%)	Stress (MPa)
0	0
0.183	306
0.36	612
0.5324	917
0.702	1223
0.867	1529
1.0244	1835
1.1774	2140
1.329	2446
1.479	2752
1.5	2767
1.56	2896

Find the longitudinal modulus E using the regression model.

$$\sigma = E\varepsilon$$

Solution

Rewriting data from Table 1 in the base SI system of units is given in Table 2.

Table 2 Stress vs strain data for a composite in SI system of units

Strain (m/m)	Stress (Pa)
0.0000	0.0000
1.8300×10^{-3}	3.0600×10^{8}
3.6000×10^{-3}	6.1200×10^{8}
5.3240×10^{-3}	9.1700×10^{8}
7.0200×10^{-3}	1.2230×10^{9}
8.6700×10^{-3}	1.5290×10^{9}
1.0244×10^{-2}	1.8350×10^{9}
1.1774×10^{-2}	2.1400×10^{9}
1.3290×10^{-2}	2.4460×10^{9}
1.4790×10^{-2}	2.7520×10^{9}
1.5000×10^{-2}	2.7670×10^{9}
1.5600×10^{-2}	2.8960×10^{9}

Linear Regression

A zero y-intercept example

Using Equation (4) gives

$$E = \frac{\sum_{i=1}^{n} \sigma_i \varepsilon_i}{\sum_{i=1}^{n} {\varepsilon_i}^2} \qquad (E1.1)$$

The summations used in Equation (E1.1) are given in Table 3.

Table 3 Tabulation for Example 2 for needed summations

i	ε	σ	ε^2	$\varepsilon\sigma$
1	0.0000	0.0000	0.0000	0.0000
2	1.8300×10^{-3}	3.0600×10^{8}	3.3489×10^{-6}	5.5998×10^{5}
3	3.6000×10^{-3}	6.1200×10^{8}	1.2960×10^{-5}	2.2032×10^{6}
4	5.3240×10^{-3}	9.1700×10^{8}	2.8345×10^{-5}	4.8821×10^{6}
5	7.0200×10^{-3}	1.2230×10^{9}	4.9280×10^{-5}	8.5855×10^{6}
6	8.6700×10^{-3}	1.5290×10^{9}	7.5169×10^{-5}	1.3256×10^{7}
7	1.0244×10^{-2}	1.8350×10^{9}	1.0494×10^{-4}	1.8798×10^{7}
8	1.1774×10^{-2}	2.1400×10^{9}	1.3863×10^{-4}	2.5196×10^{7}
9	1.3290×10^{-2}	2.4460×10^{9}	1.7662×10^{-4}	3.2507×10^{7}
10	1.4790×10^{-2}	2.7520×10^{9}	2.1874×10^{-4}	4.0702×10^{7}
11	1.5000×10^{-2}	2.7670×10^{9}	2.2500×10^{-4}	4.1505×10^{7}
12	1.5600×10^{-2}	2.8960×10^{9}	2.4336×10^{-4}	4.5178×10^{7}
$\sum_{i=1}^{12}$			1.2764×10^{-3}	2.3337×10^{8}

$$n = 12$$

$$\sum_{i=1}^{12} \varepsilon_i^2 = 1.2764 \times 10^{-3}$$

$$\sum_{i=1}^{12} \sigma_i \varepsilon_i = 2.3337 \times 10^{8}$$

$$E = \frac{\sum_{i=1}^{12} \sigma_i \varepsilon_i}{\sum_{i=1}^{12} {\varepsilon_i}^2} = \frac{2.3337 \times 10^{8}}{1.2764 \times 10^{-3}} = 1.8284 \times 10^{11}\ \mathrm{Pa} = 182.84\ \mathrm{GPa}$$

Linear Regression

A zero y-intercept example

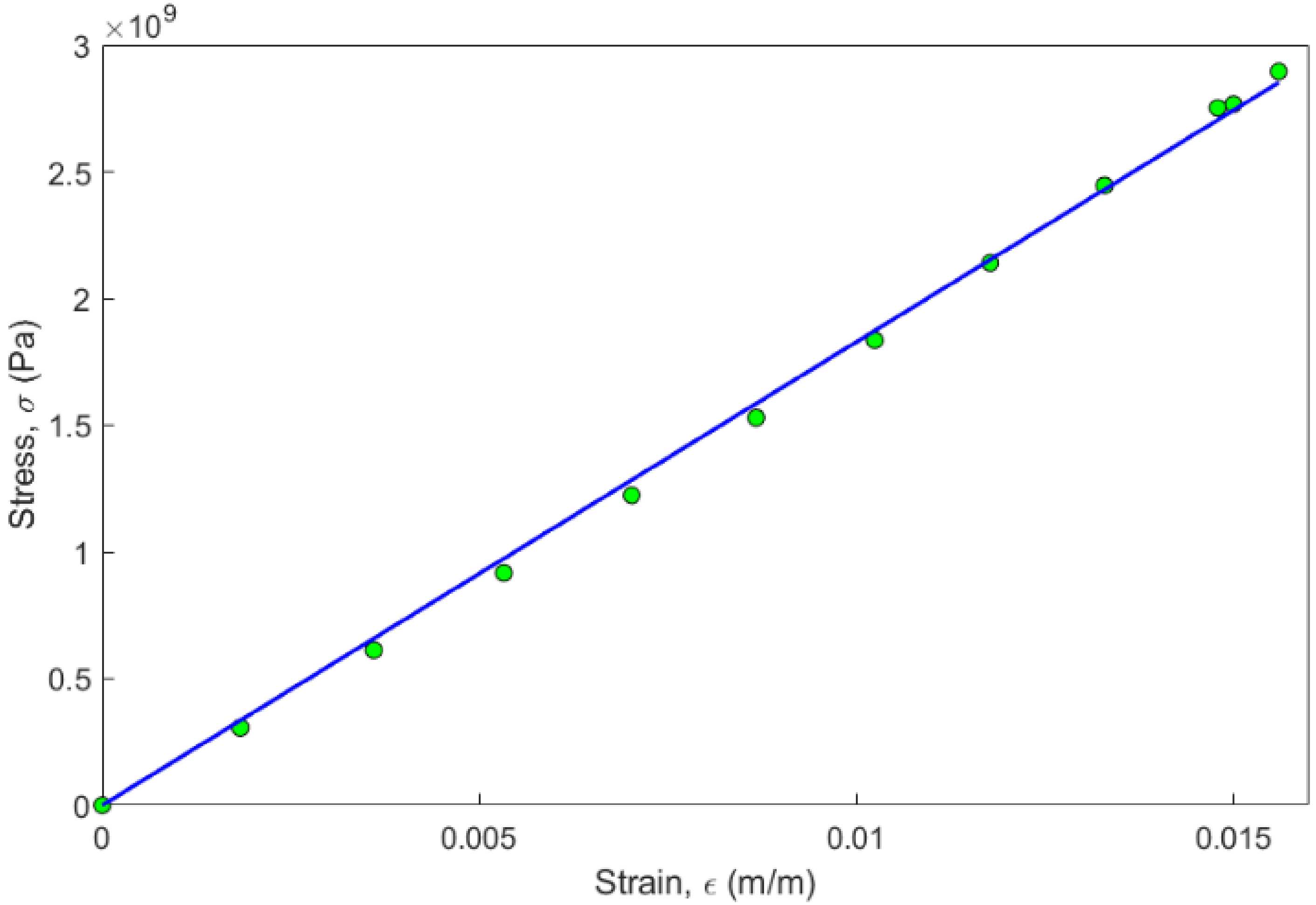


Figure 1. Stress vs strain data and regression model for a composite material uniaxial test

$$
\begin{aligned}
E &= \frac{\sum_{i=1}^{12} \sigma_i \varepsilon_i}{\sum_{i=1}^{12} {\varepsilon_i}^2} \\
&= \frac{2.3337 \times 10^8}{1.2764 \times 10^{-3}} \\
&= 1.8284 \times 10^{11} \text{ Pa} \\
&= 182.84 \text{ GPa}
\end{aligned}
$$