Deriving the General Straight-Line Regression Model

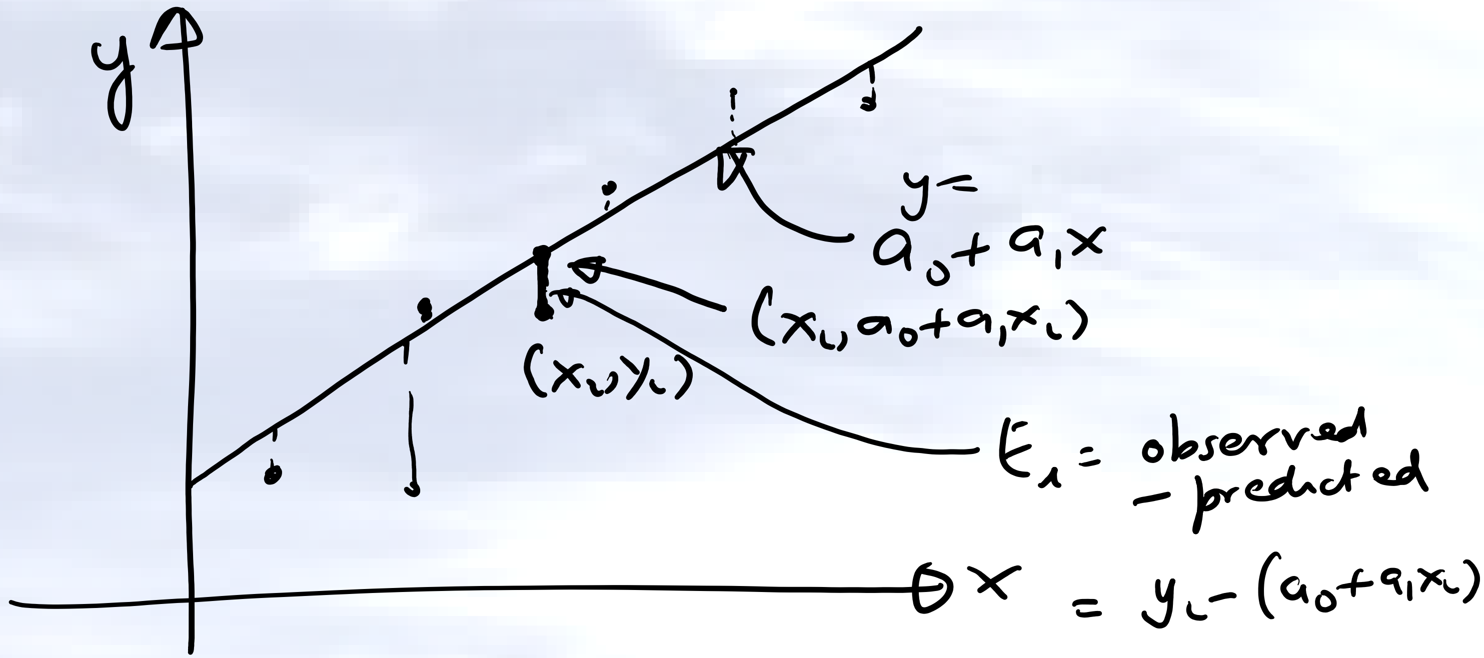
MathForCollege.com
Open Education Resources

http://nm.MathForCollege.com
Transforming Numerical Methods Education for STEM Undergraduates

For more details on this topic

- Go to http://nm.MathForCollege.com
- Click on Regression

Given $(x_1, y_1), \ldots, (x_n, y_n)$, best fit $y = a_0 + a_1 x$ to the data.



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

$$= \boxed{\sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2}$$

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

make it as small as possible

First derivative part

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

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

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

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

Divide both eqns by 2

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

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

or

$$n a_0 + a_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \quad \text{--- (1)}$$

$$a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \quad \text{--- (2)}$$

Multiply eqn (1) by $\sum_{i=1}^{n} x_i$, and eqn (2) by n

and subtract,

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

Eqn (1)

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

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

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

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

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

$$\boxed{y = a_0 + a_1 x}$$

Second derivative test tell us that a_0, a_1 values found correspond to a local minimum.

✓ S_r is a continuous function of a_0 and a_1; $\underline{\frac{\partial S_r}{\partial a_0} = 0; \frac{\partial S_r}{\partial a_1} = 0}$ only yields one solution.

That implies that a_0, a_1 values found correspond to an absolute minimum.

$\boxed{\text{END}}$

You are free

- to **Share** – to copy, distribute, display and perform the work
- to **Remix** – to make derivative works

under the following conditions

- **Attribution** — You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).
- **Noncommercial** — You may not use this work for commercial purposes.
- **Share Alike** — If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar license to this one.

Acknowledgement

This instructional resource is brought to you by Numerical Methods for STEM undergraduate
http://nm.MathForCollege.com
Committed to bringing numerical methods to the undergraduate

This material is based upon work supported by the National Science Foundation under Grant #2013271 (Transforming Undergraduate Engineering Education through Adaptive Learning and Student Data Analytics). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

MathForCollege.com

Open Education Resources

Like what you see - tell your friends.

Subscribe to the NumericalMethodsGuy Channel – help us reach our goal of 100,000 subscribers.