

# BUDAPESTI UNIVERSITY OF TECHNOLOGY AND ECONOMICS

INSTITUTE OF MATHEMATICS

BACHELOR THESIS

---

## Linear Regression through Origin

---

*Author:*

Dyussenov Nuraly

*Supervisor:*

Dr. Jozsef Mala

Associate Professor, BME Fac. of Nat. Sci.

Budapest, November 29, 2023



M Ű E G Y E T E M 1 7 8 2

# Contents

<b>1</b>	<b>Preliminaries</b>	<b>1</b>
1.1	Introduction . . . . .	1
<b>2</b>	<b>Linear Regression Regression with no intercept term</b>	<b>5</b>
2.1	Simple Linear Regression with no intercept term . . . . .	5
2.1.1	Motivational example . . . . .	8
2.1.2	Relevant Literature . . . . .	10
2.2	Comparative Analysis . . . . .	10
<b>3</b>	<b>Applications to Linear Regression through Origin</b>	<b>11</b>
3.1	Something to add 1 . . . . .	12
3.2	Something to add 1 . . . . .	12
<b>4</b>	<b>Theoretical results</b>	<b>13</b>
4.1	A theoretical resilt . . . . .	13
4.2	Towards some advanced topic . . . . .	13
<b>5</b>	<b>Programming simulations</b>	<b>14</b>
<b>6</b>	<b>Summary and closing words</b>	<b>15</b>
<b>A</b>	<b>Program Codes</b>	<b>17</b>

# 1. Preliminaries

*"A place for future inspirational quote."*

– Name of the author

## 1.1 Introduction

In the world of regression analysis, choosing the right model is a constant challenge, balancing simplicity and accuracy. This thesis focuses on a specific aspect—linear regression through the origin (RTO) —examining its statistical properties when dealing with just one explanatory variable. Our goal is to identify situations where this approach might be more suitable than the commonly used simple linear regression. Through this study, we aim to shed light on the conditions that make regression through the origin a preferable choice, offering insights that bridge mathematical rigor with real-world applicability. Join us on this journey as we navigate the complexities of statistical modeling, striving to understand when and why regression through the origin might outperform its more conventional counterpart.

### Simple Linear Regression

Before we start delving into RTO, it's best to get familiar with a more general case - Simple Linear Regression.

In Simple Linear Regression, we are given a random sample of data points  $(x_1, y_1), \dots, (x_n, y_n)$  from a population, and our goal is to find a linear function

$$y = \beta_1 x + \beta_o \tag{1.1}$$

That describes the relationship between two variables  $x$  and  $y$  as good as possible.

## 1. PRELIMINARIES

---

Since, sample is random, the equation (1.1) is not true in general, so we take into account the error term  $\varepsilon$ :

$$y = \beta_1 x + \beta_0 + \varepsilon \quad (1.2)$$

The objective of simple linear regression is, under some conditions\*, to estimate the parameters  $\beta_0$  and  $\beta_1$ , so that they will provide best fit.

List out the conditions

Likelihood function

$$L(y_1, \dots, y_n | \beta_0, \beta_1, \sigma) = \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \right)$$

Log-likelihood function

$$l(y_1, \dots, y_n | \beta_0, \beta_1, \sigma) = c - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2$$

$$\frac{\partial l}{\partial \beta_0} = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i)) = 0 \quad (1.3)$$

$$\frac{\partial l}{\partial \beta_1} = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i)) x_i = 0 \quad (1.4)$$

$$\frac{\partial l}{\partial \sigma} = \frac{n}{\sigma} - \frac{1}{\sigma^3} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 = 0 \quad (1.5)$$

Let the solutions of the above equations be denoted as  $\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2$  for  $\beta_0, \beta_1, \sigma^2$ . If  $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ , then...

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x};$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}};$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{1}{n} SSE.$$

So  $\hat{\beta}_0, \hat{\beta}_1$  are Maximum Likelihood Estimators of the model.

## 1. PRELIMINARIES

---

**Proposition 1.1.1.** *Finding values of  $\beta_0, \beta_1$  that minimize MSE is same as finding MLE of  $\beta_0, \beta_1$*

*Proof.* □ todo

**Proposition**

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

*Proof.* The vector  $\mathbf{y} - \hat{\mathbf{y}}$  is perpendicular to  $\hat{\mathbf{y}} - \mathbf{1} \bar{y}$ , thus the proposition is true by the Pythagorean theorem.

Alternatively, it is enough to show that

$$\sum (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) = 0$$

,

since then:

$$\begin{aligned} \sum_{i=1}^n (y_i - \bar{y})^2 &= \left( \sum_{i=1}^n (\hat{y}_i - \bar{y}) + \sum_{i=1}^n (y_i - \hat{y}_i) \right)^2 = \\ &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + 2 \sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) + \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \\ &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \end{aligned}$$

From 1.3 we know that  $\sum (y_i - \bar{y}) = 0$ . From 1.4 we know that  $\sum (y_i - \hat{y}_i)x_i = 0$ ,  $\hat{y}_i = \beta_0 + \beta_1 x_i \Rightarrow x_i = \frac{1}{\beta_1}(\hat{y}_i - \beta_0) \Rightarrow \sum \hat{y}_i(y_i - \hat{y}_i) = 0$

Finally,

$$\sum (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) = \sum \hat{y}_i(y_i - \hat{y}_i) - \bar{y} \sum (y_i - \hat{y}_i) = 0$$

□

**Proposition 1.4.** The estimators  $\hat{\beta}_0, \hat{\beta}_1, \frac{SSE}{n-2}$  are unbiased estimators of  $\beta_0, \beta_1, \sigma^2$  respectively.

**Proof:**

1. **Unbiasedness of  $\hat{\beta}_1$ :**

$$\begin{aligned}
 \mathbb{E}[\hat{\beta}_1] &= \mathbb{E}\left[\frac{S_{xy}}{S_{xx}}\right] = \mathbb{E}\left[\frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}\right] = \\
 &= \mathbb{E}\left[\frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2}\right] = \frac{\sum(x_i - \bar{x})\mathbb{E}[y_i]}{\sum(x_i - \bar{x})^2} = \\
 &= \frac{\sum(x_i - \bar{x})(\beta_0 + \beta_1 x_i)}{\sum(x_i - \bar{x})^2} = \frac{\sum(x_i \beta_0 - \bar{x} \beta_0 + \beta_1 x_i^2 - \beta_1 x_i \bar{x})}{\sum x_i^2 - n \bar{x}^2} = \\
 &= \frac{\cancel{n \bar{x} \beta_0} - \cancel{n \bar{x} \beta_0} + \sum \beta_1 x_i^2 - n \beta_1 \bar{x}^2}{\sum x_i^2 - n \bar{x}^2} = \frac{(\sum x_i^2 - n \bar{x}^2) \beta_1}{\sum x_i^2 - n \bar{x}^2} = \beta_1
 \end{aligned}$$

2. **Unbiasedness of  $\hat{\beta}_0$ :**

$$\begin{aligned}
 \mathbb{E}(\hat{\beta}_0) &= \mathbb{E}(\bar{y} - \hat{\beta}_1 \bar{x}) = \bar{y} - \bar{x} \mathbb{E}(\hat{\beta}_1) = \frac{1}{n} \mathbb{E}[\sum y_i] - \beta_1 \bar{x} = \\
 &= \frac{1}{n} \mathbb{E}[\sum (\beta_0 + \beta_1 x_i)] - \beta_1 \bar{x} = \frac{1}{n} n \beta_0 + \frac{1}{n} n \beta_1 \bar{x} - \bar{x} \beta_1 = \beta_0
 \end{aligned}$$

3. **Unbiasedness of  $\frac{SSE}{n-2}$  as an estimator of  $\sigma^2$ :**

$$\mathbb{E}\left(\frac{SSE}{n-2}\right) = \frac{1}{n-2} (n-2) \sigma^2 = \sigma^2$$

**Proposition 1.1.2.**  $\mathbb{V}ar[\hat{\beta}_1] = \frac{\sigma^2}{S_{xx}}$

*Proof.* Assume,  $Y_i \sim N(0, \sigma^2)$

$$\mathbb{V}ar[\hat{\beta}_1] = \mathbb{V}ar\left(\frac{1}{S_{xx}} \sum (x_i - \bar{x}) Y_i\right) = \frac{1}{S_{xx}^2} \sum \mathbb{V}ar Y_i = \frac{\sigma}{S_{xx}^2}$$

□

## 2. Linear Regression Regression with no intercept term

### 2.1 Simple Linear Regression with no intercept term

In certain statistical applications, the conventional assumption of a non-zero intercept term ( $\beta_0$ ) in a simple linear regression model may not align with the nature of the data. For example, in economics the cost of production be assumed to be zero, when there is no production, or in physics, when we are describing the relationship between force and the displacement, forced is assumed to be zero, when there is no displacement.

Likelihood function  $L$  is:

$$\begin{aligned} L(y_1, \dots, y_n | \beta, \sigma) &= \prod \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \beta x_i)^2\right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum (y_i - \beta x_i)^2\right) \end{aligned}$$

Log-likelihood  $l$  is:

$$l(y_1, \dots, y_n | \beta, \sigma) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum (y_i - \beta x_i)^2$$

$$\begin{aligned}\frac{\partial l}{\partial \beta} &= -\frac{1}{2\sigma^2} \sum 2(y_i - \beta x_i)(-x_i) = 0 \\ \frac{1}{\sigma^2} \sum (y_i x_i - \beta x_i^2) &= 0 \\ \sum x_i y_i &= \sum x_i^2 \beta \\ \hat{\beta} &= \frac{\sum x_i y_i}{\sum x_i^2}\end{aligned}$$

**Proposition 2.1.1.**  $\hat{\beta}$  is unbiased

*Proof.*

$$\begin{aligned}\mathbb{E}[\hat{\beta}] &= \mathbb{E}\left[\frac{\sum x_i y_i}{\sum x_i^2}\right] = \sum \frac{1}{x_i^2} \mathbb{E}[\sum x_i y_i] = \\ &= \frac{\sum x_i \mathbb{E}[y_i]}{\sum x_i^2} = \frac{\sum x_i^2 \beta}{\sum x_i^2} = \beta\end{aligned}$$

□

**Proposition 2.1.2.**  $\mathbb{V}ar[\hat{\beta}_1^0] = \frac{\sigma^2}{\sum x_i^2}$

*Proof.*

$$\mathbb{V}ar[\hat{\beta}_1^0] = \mathbb{V}ar\left[\frac{\sum x_i y_i}{\sum x_i^2}\right] = \frac{1}{(\sum x_i^2)^2} \sum x_i^2 \mathbb{V}ar[Y_i] = \frac{\sigma^2}{\sum x_i^2}$$

□

**Proposition 2.1.3.**

$$\mathbb{V}ar[\hat{\beta}_1^0] < \mathbb{V}ar[\hat{\beta}_1]$$

*Proof.*

$$\begin{aligned}\sum x_i^2 &> \sum (x_i - \bar{x})^2 \\ \frac{1}{\sum x_i^2} &< \frac{1}{\sum (x_i - \bar{x})^2} \\ \frac{\sigma^2}{\sum x_i^2} &< \frac{\sigma^2}{\sum (x_i - \bar{x})^2} \\ \mathbb{V}ar[\hat{\beta}_1^0] &< \mathbb{V}ar[\hat{\beta}_1]\end{aligned}$$



□

This might suggest that  $\hat{\beta}_1^0$  might be more accurate estimator than  $\hat{\beta}_1$  for the slope term. This gives us some motivation to compare the two estimators more closely.

It would be convenient for us to find confidence interval for  $\beta_1^0 - \beta_1$ , since if 0 lies in the CI, then we can statistically infer that two estimators are very close.

**Proposition 2.1.4.** *The difference of the two estimators is normally distributed as follows:*

$$\hat{\beta}_1^0 - \hat{\beta}_1 \sim N(\beta_1^0 - \beta_1, \sigma^2(\frac{1}{S_{xx}} - \frac{1}{S_{xx}^0}))$$

Before proving proposition 2.1.4, let's first understand some properties of  $\hat{\beta}_1^0 - \hat{\beta}_1$ :

**Proposition 2.1.5.**  $\text{Cov}(\hat{\beta}_1^0, \hat{\beta}_1^0 - \hat{\beta}_1) = 0$

*Proof.*

$$\begin{aligned} \text{Cov}(\hat{\beta}_1^0, \hat{\beta}_1^0 - \hat{\beta}_1) &= \text{Cov}\left(\frac{\sum xy}{\sum x^2}, \frac{\sum xy}{\sum x^2} - \frac{\sum(x - \bar{x})y}{\sum(x - \bar{x})^2}\right) \\ &= \frac{\sum x^2}{(\sum x^2)^2} \text{Var } y - \frac{\sum x(x - \bar{x})}{\sum x^2 \sum (x - \bar{x})^2} \text{Var } y \\ &= \sigma^2 \left( \frac{1}{\sum x^2} - \frac{\sum x^2 - \bar{x} \sum x}{\sum x^2 \sum (x^2 - 2x\bar{x} + \bar{x}^2)} \right) \\ &= \sigma^2 \left( \frac{1}{\sum x^2} - \frac{\sum x^2 - n\bar{x}^2}{\sum x^2 (\sum x^2 - 2n\bar{x}^2 + n\bar{x}^2)} \right) \\ &= \sigma^2 \left( \frac{1}{\sum x^2} - \frac{\sum x^2 - n\bar{x}^2}{\sum x^2 (\sum x^2 - n\bar{x}^2)} \right) \\ &= \sigma^2 \left( \frac{1}{\sum x^2} - \frac{1}{\sum x^2} \right) = 0 \end{aligned}$$

□

**Proposition 2.1.6.**  $\mathbb{E}[\hat{\beta}_1^0 - \hat{\beta}_1] = \beta_1^0 - \beta_1$

*Proof.* By linearity of expected value,  $\mathbb{E}[\hat{\beta}_1^0 - \hat{\beta}_1] = \mathbb{E}[\hat{\beta}_1^0] - \mathbb{E}[\hat{\beta}_1] = \beta_1^0 - \beta_1$

□

Now we are ready to prove proposition 2.1.4

## 2. LINEAR REGRESSION REGRESSION WITH NO INTERCEPT TERM

*Proof of prop. 2.1.4.*  $\hat{\beta}_1^0 - \hat{\beta}_1$  is a linear combination of mutually independent normally distributed r.v.-s  $\Rightarrow$  it is normally distributed.

you can  
show that

$$\begin{aligned}\text{Var}(\hat{\beta}_1^0 - \hat{\beta}_1) &= \text{Var}(\hat{\beta}_1^0) + \text{Var}(\hat{\beta}_1) - 2\text{Cov}(\hat{\beta}_1^0, \hat{\beta}_1) = \frac{\sigma^2}{S_{xx}^0} + \frac{\sigma^2}{S_{xx}} - 2\text{Cov}(\hat{\beta}_1^0, \hat{\beta}_1^0 - \hat{\beta}_1) - 2\text{Cov}(\hat{\beta}_1^0, \hat{\beta}_1^0) \\ &= \frac{\sigma^2}{S_{xx}^0} + \frac{\sigma^2}{S_{xx}} - 0 - 2\text{Var} \hat{\beta}_1^0 = \frac{\sigma^2}{S_{xx}} - \frac{\sigma^2}{S_{xx}^0}\end{aligned}$$

□

Using proposition 2.1.4, we can now construct confidence interval for  $\beta_1^0 - \beta_1$ :

$$\begin{aligned}\hat{\beta}_1^0 - \hat{\beta}_1 &\sim N(\beta_1^0 - \beta_1, \sigma^2(\frac{1}{S_{xx}} - \frac{1}{S_{xx}^0})) \\ Z &:= \frac{\hat{\beta}_1^0 - \hat{\beta}_1 - \mathbb{E}[\hat{\beta}_1^0 - \hat{\beta}_1]}{\text{Var}(\hat{\beta}_1^0 - \hat{\beta}_1)} = \frac{\hat{\beta}_1^0 - \hat{\beta}_1 - (\beta_1^0 - \beta_1)}{\sigma^2(\frac{1}{S_{xx}} - \frac{1}{S_{xx}^0})} \\ \beta_1^0 - \beta_1 &= \hat{\beta}_1^0 - \hat{\beta}_1 - \sigma^2 Z(\frac{1}{S_{xx}} - \frac{1}{S_{xx}^0})\end{aligned}$$

CI is:

$$\hat{\beta}_1^0 - \hat{\beta}_1 - \sigma^2 Z_\alpha(\frac{1}{S_{xx}} - \frac{1}{S_{xx}^0}) \leq \beta_1^0 - \beta_1 \leq \hat{\beta}_1^0 - \hat{\beta}_1 + \sigma^2 Z_\alpha(\frac{1}{S_{xx}} - \frac{1}{S_{xx}^0}) \quad (2.1)$$

### 2.1.1 Motivational example

Even though RTO might be worse than performing full linear regression model in terms of SSE (since full model always gives unbiased estimators, they are also MLE estimators, so SSE is minimized), there are other statistical parameters that are better in RTO when intercept term is small enough.

Given that the true model is known,  $y_i = \beta_1 x_i + \alpha$ , one might simulate the data points by adding normally-distributed error terms:

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

## 2. LINEAR REGRESSION REGRESSION WITH NO INTERCEPT TERM

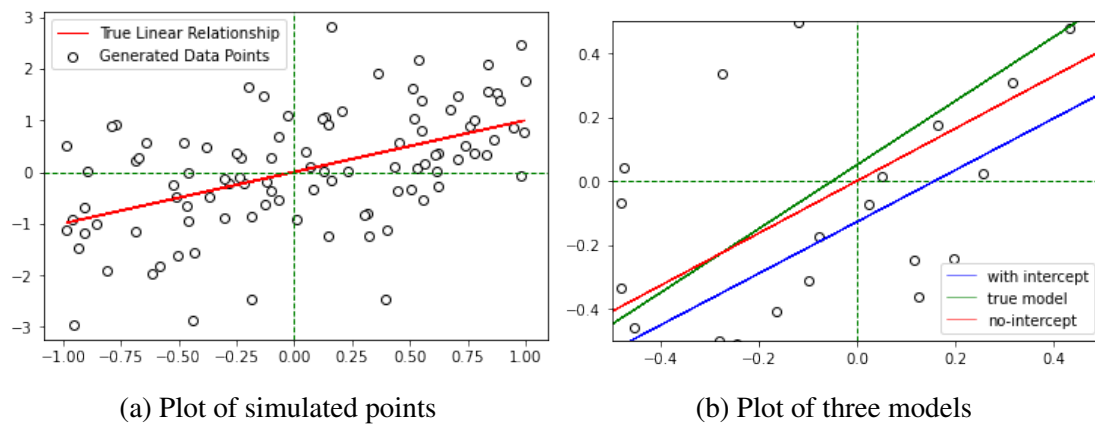


Figure 2.1: Overall caption for the figure

One might be interested now in comparing sum of squared deviations (SSD) of fitted points  $\sum(\hat{y}_i - \hat{y}_i)^2$ .

Let's start by generating a sample of 100 points, and assume that parameters are known as  $\beta = 1, \alpha = 0.05$  (see fig. 2.1a):

Now, let's fit two models: no-intercept model, and full model. For that purposes we will use `LinearRegression` function from `sklearn.linear_model` package. On fig. 2.1b you can see the two models as well as the true model plotted up close.

Now let's compare two statistics for these two models: BIC (Bayesian Information Criterion) and SSD.

The results are as follows:

$$BIC_{no} = 386.1685814527955$$

$$BIC = 391.09782858386046$$

$$SSD_{no} = 1.2267420929010404$$

$$SSD = 4.150248687009578$$

It is clearly seen that no-intercept model provides better fit in terms of these parameters.

### **2.1.2 Relevant Literature**

Refer to seminal works and research studies that have explored or utilized the simple linear regression model without an intercept term. A brief review of the literature provides additional context and allows for a synthesis of existing knowledge in this specialized domain.

## **2.2 Comparative Analysis**

### **3. Applications to Linear Regression through Origin**

### **3.1 Something to add 1**

### **3.2 Something to add 1**

## **4. Theoretical results**

### **4.1 A theoretical resilt**

### **4.2 Towards some advanced topic**

## **5. Programming simulations**



## **6. Summary and closing words**

# **Bibliography**

## **A. Program Codes**