

# BUDAPESTI UNIVERSITY OF TECHNOLOGY AND ECONOMICS

INSTITUTE OF MATHEMATICS

FACULTY OF MATHEMATICS

---

## Linear Regression through Origin

---

*Author:*

Dyussenov Nuraly

*Supervisor:*

Dr. Jozsef Mala

Associate Professor, BME Fac. of Nat. Sci.

Budapest, October 24, 2023



M Ű E G Y E T E M 1 7 8 2

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Theoretical background</b>	<b>2</b>
2.1	Statistics Basics . . . . .	2
2.2	Simple Linear Regression . . . . .	4
2.3	Simple Linear Regression with no intercept term . . . . .	5
2.4	Comparative Analysis . . . . .	6
<b>3</b>	<b>Applications to Linear Regression through Origin</b>	<b>7</b>
3.1	Something to add 1 . . . . .	8
3.2	Something to add 1 . . . . .	8
<b>4</b>	<b>Theoretical results</b>	<b>9</b>
4.1	A theoretical resilt . . . . .	9
4.2	Towards some advanced topic . . . . .	9
<b>5</b>	<b>Programming simulations</b>	<b>10</b>
<b>6</b>	<b>Summary and closing words</b>	<b>11</b>
<b>A</b>	<b>Program Codes</b>	<b>13</b>

# List of Tables

# List of Figures

# 1. Introduction

*"Bla-bla-bla"*

– XY

## 2. Theoretical background

### 2.1 Statistics Basics

**Definition (Data)** Let  $(x_1, \dots, x_n)$ , where  $x_i \in S$  for  $i = 1, \dots, n$ . The set  $S$  is typically  $\mathbb{R}$ ,  $\mathbb{R}^d$ , or it can be any abstract set. However, for our purposes,  $S$  (the sample space) will usually be  $\mathbb{R}$ .

**Definition (Sample)** In statistics, our data are often modeled by a vector  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  of i.i.d. (independent, identically distributed) random variables, called the sample (of which size is  $n$ ), where the random variables  $X_i$  take values in  $\mathbb{Z}$  or  $\mathbb{R}$ . The common distribution of the  $X_i$  is called the parent distribution, and we say that the sample is from that parent distribution.

**Definition (Model)** A statistical model is a family  $\{P_\theta \mid \theta \in \Theta\}$  of distributions on the sample space. When  $\Theta \subset \mathbb{R}^d$ , we say that we have a parametric model, and we call  $\Theta$  the parameter set (space).

**Definition (p-th Quantile of Data)** If  $p \in (0, 1)$ , then a  $p$ -th quantile (or a  $p$ -th percentile) of the data  $(x_1, \dots, x_n)$  is a  $p$ -th quantile of the corresponding empirical distribution function  $\hat{F}_n$ .

**Definition (Sample mean)** Let  $(X_1, \dots, X_n)$  be a sample. Then the random variable

$$\bar{X} = X = \frac{1}{n} \sum_{i=1}^n X_i$$

is called the sample mean.

**Definition (Estimator)** An estimator is a statistic (a function of the sample data) used to estimate an unknown parameter in a statistical model. An estimator for the parameter  $\theta$ , denoted as  $\hat{\theta}$ , is any measurable function of the random variables  $X_1, X_2, \dots, X_n$ .

**Definition (Biased)** If  $\hat{\theta}$  is an estimator of  $\theta$ , then we can define the quantity  $Bias(\hat{\theta}) = \mathbb{E}_\theta[\hat{\theta}] - \theta$ . The estimator  $\hat{\theta}$  is called unbiased if its bias is 0.

**Definition (MSE of an Estimator)** Let us have the model  $\{P_\theta \mid \theta \in \Theta\}$  and let us have the sample  $(X_1, \dots, X_n)$  from it. The mean square error (or the quadratic risk) of an estimator  $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$  for the parameter  $\theta$  is defined by

$$\text{MSE}_\theta(\hat{\theta}) = \mathbb{E}_\theta((\hat{\theta} - \theta)^2)$$

when  $\theta$  is the true parameter.

**Steiner's identity:**  $\mathbb{E}((X - a)^2) = \text{Var}(X) + (a - \mathbb{E}(X))^2$

**Interpretation in the context of mean square error (MSE):**

$$\text{MSE}_\theta(\hat{\theta}) = \text{Var}_\theta(\hat{\theta}) + (\text{Bias}_\theta(\hat{\theta}))^2$$

**Definition (Sufficiency)** Let the model be  $\{P_\theta \mid \theta \in \Theta\}$  and  $\mathbf{X} = (X_1, \dots, X_n)$  be a sample from it. The statistic  $T$  is called *sufficient* for the parameter  $\theta$  (or, for the model  $\{P_\theta \mid \theta \in \Theta\}$ ) if the conditional distribution  $P_\theta(\mathbf{X} \in \cdot \mid T = t)$  does not depend on  $\theta$ .

**Theorem (Neyman-Fisher Factorization Theorem)** If the model is  $\{p(x|\theta) \mid \theta \in \Theta\}$  where  $p(x|\theta)$  is a probability mass/density function and  $\mathbf{X} = (X_1, \dots, X_n)$  is a sample from it, then the statistic  $T$  is *sufficient* for the parameter  $\theta$  if and only if we can find nonnegative functions  $g$  and  $h$  such that

$$p_{\mathbf{X}}(x|\theta) = g(T(x), \theta)h(x).$$

**Definition (Likelihood)** Let  $\{p(x, \theta), \theta \in \Theta\}$  be a model. If the observed value of  $X$  is  $x$ , we say that  $p(x|\theta)$  is the *likelihood* of  $\theta$ :  $L(\theta) = p(x|\theta)$ . Thus, we are considering the mass/density as a function of  $\theta$ , for a fixed  $x$ . If  $x = (x_1, \dots, x_n)$  is a realization of the sample  $\mathbf{X} = (X_1, \dots, X_n)$ , then  $p(x|\theta)$  is the product of the marginals,

$$L(\theta) = p(x|\theta) = \prod_{i=1}^n p(x_i|\theta).$$

**Theorem (Rao-Blackwell)** Let  $\{P_\theta \mid \theta \in \Theta\}$  be a model and  $(X_1, \dots, X_n)$  be a sample. Let  $\hat{\theta}$  be an estimator of  $\theta$  with  $\text{Var}_\theta(\hat{\theta})$  finite for each  $\theta$ . If  $T$  is a sufficient statistic for  $\theta$ , then  $\theta^* = \mathbb{E}_\theta(\hat{\theta}|T)$  is a statistic, and we have for all  $\theta$  that

$$\text{MSE}_\theta(\theta^*) \leq \text{MSE}_\theta(\hat{\theta}) \quad (1)$$

and the inequality is strict unless  $\hat{\theta}$  is a function of  $T$  with probability 1.

## 2.2 Simple Linear Regression

Consider the model function:

$$y = \alpha + \beta x,$$

which describes a line with slope  $\beta$  and y-intercept  $\alpha$ . In general, such a relationship may not hold exactly for the largely unobserved population of values of the independent and dependent variables; we call the unobserved deviations from the above equation the errors. Suppose we observe  $n$  data pairs and call them  $\{(x_i, y_i), i = 1, \dots, n\}$ . We can describe the underlying relationship between  $y_i$  and  $x_i$  involving this error term  $\varepsilon_i$  by:

$$y_i = \alpha + \beta x_i + \varepsilon_i.$$

This relationship between the true (but unobserved) underlying parameters  $\alpha$  and  $\beta$  and the data points is called a linear regression model.

The goal is to find estimated values  $\hat{\alpha}$  and  $\hat{\beta}$  for the parameters  $\alpha$  and  $\beta$  which would provide the "best" fit in some sense for the data points. As mentioned in the introduction, in this article the "best" fit will be understood as in the least-squares approach: a line that minimizes the sum of squared residuals (see also Errors and residuals)  $\hat{\varepsilon}_i$  (differences between actual and predicted values of the dependent variable  $y$ ), each of which is given by, for any candidate parameter values  $\alpha$  and  $\beta$ :

$$\hat{\varepsilon}_i = y_i - \alpha - \beta x_i.$$

In other words,  $\hat{\alpha}$  and  $\hat{\beta}$  solve the following minimization problem:

$$\text{Find } \min_{\alpha, \beta} Q(\alpha, \beta), \quad \text{for } Q(\alpha, \beta) = \sum_{i=1}^n \hat{\varepsilon}_i^2 = \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2.$$

By expanding to get a quadratic expression in  $\alpha$  and  $\beta$ , we can derive values of  $\alpha$  and  $\beta$  that minimize the objective function  $Q$  (these minimizing values are denoted  $\hat{\alpha}$  and  $\hat{\beta}$ ):[6]



## **2.3 Simple Linear Regression with no intercept term**

## **2.4 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

?<ch:closing>?

# **Bibliography**



## A. Program Codes

? $\langle$ ap:codes $\rangle$ ?