Time Series Analysis

Victor Hugo F. Francheto
Institute of Mathematical and Computer Sciences
University of São Paulo
São Carlos, Brazil
victor.francheto@usp.br

March 1, 2025

Abstract: This material is a guide for students who wish to learn more about time series analysis. Our focus here is on presenting the fundamental concepts involved in time series.

Keywords: *write*

1 Introduction

2 Linear Regression

Before diving into time series analysis, it's important to revisit the concepts of linear regression. Linear regression provides a fundamental basis for understanding how variables relate to each other over time, serving as a stepping stone to more complex models.

Many time series techniques rely on estimating trends and relationships between variables—an idea already explored in linear regression. Moreover, time series models like ARIMA and dynamic regressions often incorporate linear components to capture patterns in the data.

Reviewing the fundamentals of linear regression helps reinforce key concepts such as minimizing the mean squared error, interpreting coefficients, and understanding residuals—all of which are crucial in time series analysis. With this foundation in place, transitioning to specific time series models becomes more natural and intuitive.

Regression analysis is a statistical tool that utilizes the relationship between two or more quantitative variables to make predictions. For example, if one knows the relationship between advertising expenditures and sales, one can predict sales using regression analysis once the level of advertising expenditures has been set.

2.1 Relation Between Variables

A functional relation between two variables is expressed by a mathematical formula. If X is the *independent* variable and Y is the *dependent* variable, a functional relation between X and Y is of the form:

$$Y = f(X).$$

Give a particular value of Xm the function f indicates the corresponding value of Y.

Example: Consider the relationship between dollar sales Y of a product sold at a fixed price and the number of units sold X. If the selling price is \$2 per unit, then the relationship is given by:

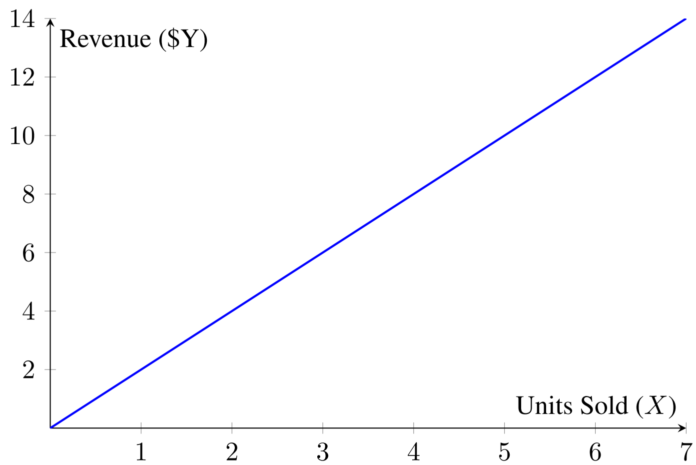
$$Y = 2X.$$

This means that for every unit sold, the total revenue increases by \$2. The table below shows the number of units sold (X) and the corresponding revenue (\$Y):

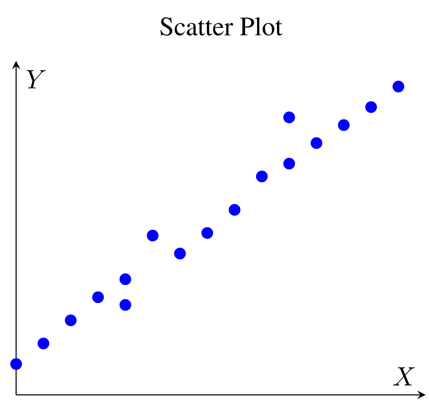
Units Sold (X)	Revenue (\$Y)
0	0
1	2
2	4
3	6
4	8
5	10
6	12

Table 1: Revenue (\$Y) as a function of units sold (X) for a price of \$2 per unit.

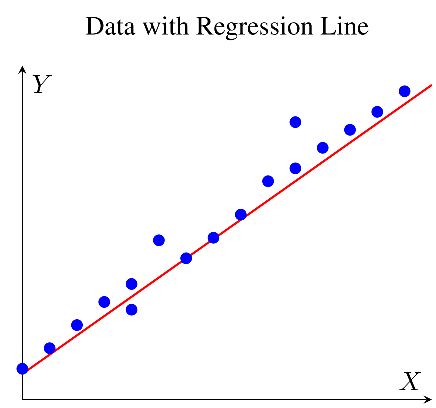
The graph below represents the function $Y = 2X$, showing how revenue (\$Y) increases linearly with the number of units sold (X).



A statistical relation, unlike a functional relation, is not perfect. In general, observations in a statistical relation do not fall directly on the curve of the relationship.



(a) Only data points.



(b) Data points with regression line.

2.2 Regression Models and Their Uses

A regression model is a formal approach to representing the fundamental relationship between two statistical quantities.

1. A tendency of the dependent variable Y to vary with the independent variable or variables in a systematic fashion;

2. A scattering of observations around the curve of the statistical relationship.

These two characteristics are embodied in a regression model by postulating that:

1. In the population of observations associated with the sampled process, there is a probability distribution of Y for each level of X.
2. The means of these probability distributions vary systematically with X.

The expressions *independent variable* or **predictor variable** for X, and *response variable* for Y in a regression model are merely conventional labels. There is no implication that Y causally depends on X in a given case. No matter how strong the statistical relationship, a regression model does not necessarily imply a cause-and-effect pattern.

In some applications, an independent variable may actually be causally dependent on the response variable. For example, when estimating temperature (the response) from the height of mercury (the independent variable) in a thermometer.

2.3 Regression Models with Multiple Independent Variables

Regression models may contain more than one independent variable. For example:

1. **Predicting Used Car Prices:** A dealership wants to predict the prices of used cars to adjust its sales strategy. To achieve this, it considers factors such as mileage, model year, engine power, and the number of previous owners as independent variables. Each of these variables helps understand how different characteristics influence the perceived value of vehicles;
2. **Predicting Urban Real Estate Prices:** A real estate company aims to estimate the prices of residential properties in a specific city. The independent variables considered include total built area, location (neighborhood), proximity to schools or green areas, property age, and the presence of a garage. These variables provide insights into which factors have the greatest impact on property valuation;
3. **Predicting Stock Prices in the Financial Market:** A team of financial analysts wants to forecast stock price fluctuations for technology sector companies. They consider independent variables such as quarterly profit, company debt, annual growth rate, brand reputation, and the overall economic outlook. These variables help understand the factors that may influence investor behavior and stock prices.

2.4 Regression Model With Unspecified Error Term Distribution

Previously, we considered a basic regression model with only one independent variable and a linear regression function. The model can be stated as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{1}$$

where:

- Y_i is the value of the response variable in the ith trial, where β_0 and β_1 are parameters;
- X_i is a known constant, that is, the value of the independent variable in the ith trial;
- ϵ_i is a random error term with $\mathbb{E}(\epsilon_i) = 0$ and variance $Var(\epsilon_i) = \sigma^2$. Moreover, ϵ_i and ϵ_j are uncorrelated, meaning that $Cov(\epsilon_i, \epsilon_j) = 0$ for all $i \neq j$, where $i, j = 1, 2, \ldots, n$.

The model in (1) is referred to as a *simple linear model*, as it is linear in both the parameters and the independent variable. It is termed "simple" because it includes only one independent variable. The model is "linear in the independent variable" since this variable appears only to the first power. A model that is linear in both the parameters and the independent variable is also called a *first-order model*.

The observed value of Y in the ith trial consists of two components: the deterministic term $\beta_0 + \beta_1 X_i$ and the random term ϵ_i. Consequently, Y_i is a random variable.

Since $\mathbb{E}(\epsilon_i) = 0$, it follows that

$$\mathbb{E}(Y_i) = \mathbb{E}(\beta_0 + \beta_1 X_i + \epsilon_i) = \beta_0 + \beta_1 X_i + \mathbb{E}(\epsilon_i) = \beta_0 + \beta_1 X_i. \tag{2}$$

Note that, $\beta_0 + \beta_1 X_i$ plays the role of the constant.

Thus, the response Y_i, when the level of X existing in the ith trial is X_i, comes from a probability distribution whose mean is:

$$\mathbb{E}(Y_i) = \beta_0 + \beta_1 X_i. \tag{3}$$

We therefore know that regression function for model (1) is:

$$\mathbb{E}(Y) = \beta_0 + \beta_1 X, \tag{4}$$

since the regression function relates the means of the probability distributions of Y for any given X to level of X.

The observed value of Y in the ith trial exceeds or falls short of the value of the regression function by the error term amount ϵ_i.

The error terms ϵ_i are assumed to have constant variance σ^2. It therefore, follows that the variance of the response Y_i is:

$$\sigma^2(Y_i) = \sigma^2 \tag{5}$$

since,

$$\sigma^2(\beta_0 + \beta_1 X_i + \epsilon_i) = \sigma^2(\epsilon_i) = \sigma^2. \tag{6}$$

Thus, the model in (1) assumes that the probability distribution of Y have the same variance σ^2, regardless of the level of the independent variable X.

References