

A Brief Overview of Regression Analysis

This article is supposed to be a quick refresher on regression analysis, and more precisely - on one of its types - the logistic regression model, as it will be the tool we will implement in our analysis while solving our business task.

Therefore, if you are confident with this material, just skip the remaining paragraphs of this article, and proceed with the next video directly. If you feel this is not the case yet, then please read on, as we will provide enough information and intuition on the analytical approach to be used from now on.

A popular tool in data analytics, machine learning, advanced statistics, and econometrics, is *regression analysis*.

Roughly speaking, this is an equation which on one side has a variable, called a *dependent* variable, because its value will *depend* on the values of all variables you see on the other side.

The variables on the right side are all *independent*, or *explanatory*. Their role is to *explain* the value of the *dependent* variable.

There are more terms that can be used for these variables in the same context. The dependent variable can also be called a *target*, while the independent variables can be called *predictors*.

Furthermore, as mentioned in the previous video, to avoid confusion with the term 'variable', which has a different meaning when used by programmers, we, as BI analysts or data scientists, will call the explanatory variables 'features'. Remember that if you prefer, you could call them 'attributes' or 'inputs' as well.

Having said that, it is easy to describe what *logistic regression* is about.

It is a type of a regression model whose dependent variable is binary. That is, the latter can assume one of two values – 0 or 1, True or False, Yes or No.

Therefore, considering the values of all our features, we want to be able to predict whether the dependent variable will take the value of 0 or 1.

Apart from logistic regression, there are many other types of equations that allow you to calculate the dependent variable in a different way. Logistic regression is just one of them – and it is one that has been used massively.

Anyway, you would most often hear professionals say that they are trying to find a regression model, or, simply, find a *regression*, that has a high predictive power. In other words, what they are trying to do is settle upon an equation that could be used to estimate expected values for the *dependent* variable with great precision.

For the moment, this introduction to regression analysis and logistic regression gives us the necessary grounds to proceed with the pre-processing part of our task.

See you in the next video!