PRADIPTA NANDI

Serial Number: 250

Batch: 7

Domain: Data Science with Python

COMPLETE
OVERVIEW
OF
REGRESSION

WHAT IS REGRESSION?

Regression is a statistical method used in finance, investing, and other disciplines that attempt to determine the strength and character of the relationship between one dependent variable (usually denoted by Y) and a series of other variables (known as independent variables).

IMPORTANCE OF REGRESSION

The importance of regression analysis for a small business is that it helps determine which factors matter most, which it can ignore, and how those factors interact with each other. The importance of regression analysis lies in the fact that it provides a powerful statistical method that allows a business to examine the relationship between two or more variables of interest.

Another importance is that it is all about data: data means numbers and figures that define your business. The advantage of regression analysis is that it can allow you to essentially crunch the numbers to help you make better decisions for your business now and in the future.

The regression method of forecasting means studying the relationships between data points, which can help you to:

- **Predict sales in the near and long term.**
- **Understand inventory levels.**
- **Understand supply and demand.**
- **Review and understand how different variables impact all of these things.**

Companies might use regression analysis to understand, for example:

- **Why customer service calls dropped in the past year or even the past month.**
- **Predict what sales will look like in the next six months.**
- **Whether to choose one marketing promotion over another.**
- **Whether to expand the business or create and market a new product.**

USING REGRESSION ANALYSIS TO FORMULATE STRATEGIES

It's important to understand that regression analysis is, essentially, a statistical problem. Businesses have adopted many concepts from statistics because they can prove valuable in helping a company determine any number of important things and then make informed, well-studied decisions based on various aspects of data. And data, according to Merriam-Webster, is merely factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation.

Regression analysis uses data, specifically two or more variables, to provide some idea of where future data points will be. The benefit of regression analysis is that this type of statistical calculation gives businesses a way to see the future. The regression method of forecasting allows businesses to use specific strategies so that those predictions, such as future sales, future needs for labor or supplies, or even future challenges, will yield meaningful information.

APPLICATIONS OF REGRESSION ANALYSIS

Predictive analytics:

This application, which involves forecasting future opportunities and risks, is the most widely used application of regression analysis in business. For example, predictive analytics might involve demand analysis, which seeks to predict the number of items that consumers will purchase in the future.

Operation efficiency:

Companies use this application to optimize the business process. This kind of data-driven decision-making can eliminate guesswork and make the process of creating optimum efficiency less about gut instinct and more about using well-crafted predictions based on real data.

Supporting decisions:

Many companies and their top managers today are using regression analysis (and other kinds of data analytics) to make an informed business decision and eliminate guesswork and gut intuition. Regression helps businesses adopt a scientific angle in their management strategies.

Correcting errors:

Even the most informed and careful managers do make mistakes in judgment. Regression analysis helps managers, and businesses in general, recognize and correct errors. Using regression analysis could help a manager determine that an increase in hours would not lead to an increase in profits. This could help the manager avoid making a costly mistake.

New Insights:

Looking at the data can provide new and fresh insights. Many businesses gather lots of data about their customers. But that data is meaningless without proper regression analysis, which can help find the relationship between different variables to uncover patterns.

REGRESSION ANALYSIS EXAMPLE

Though this sounds complicated, it's fairly simple. You could simply look back at the activity of the GDP in the last quarter or the last three-month period, and compare it to your sales figure. In 2018, the government reported that the GDP grew 2.6 percent in the fourth quarter. If your sales rose 5.2 percent during that same period, you'd have a pretty good idea that your sales generally rise at twice the rate of GDP growth because:

5.2 percent (your sales) / 2.6 percent = 2

The "2" means that your sales are rising at twice the rate of the GDP. You might want to go back a couple of more quarters to be sure this trend continues, say for an entire year. Suppose you sell car parts, wheat, or forklifts. It would be the same regardless of the products or services you sell. Since you know that your sales are increasing at twice the rate of GDP growth, then if the GDP increases 4 percent the next quarter, your sales will likely rise 8 percent. If the GDP goes up 3 percent, your sales would likely rise 6 percent, and so on.

In this way, regression analysis can be a valuable tool for forecasting sales and help you determine whether you need to increase supplies, labor, production hours, and any number of other factors.

HOW TO INTERPRET REGRESSION OUTPUT

To answer questions using regression analysis, you first need to fit and verify that you have a good model. Then, you look through the regression coefficients and p-values. When you have a low p-value (typically < 0.05), the independent variable is statistically significant. The coefficients represent the average change in the dependent variable given a one-unit change in the independent variable (IV) while controlling the other IVs.

For instance, if your dependent variable is income and your IVs include IQ and education (among other relevant variables), you might see output like this:

Coefficients

Term	Coef	SE Coef	T	P
Constant	483.670	39.5671	12.2241	0.000
IQ	4.796	0.9511	5.0429	0.000
Education	24.215	1.9405	12.4785	0.000

The low p-values indicate that both education and IQ are statistically significant. The coefficient for IQ indicates that each additional IQ point increases your income by an average of approximately $4.80 while controlling everything else in the model. Furthermore, an additional unit of education increases average earnings by $24.22 while holding the other variables constant.

Regression analysis is a form of inferential statistics. The p-values help determine whether the relationships that you observe in your sample also exist in the larger population.

TYPES OF REGRESSION

The two basic types of regression are simple linear regression and multiple linear regression, although there are non-linear regression methods for more complicated data and analysis. Simple linear regression uses one independent variable to explain or predict the outcome of the dependent variable Y, while multiple linear regression uses two or more independent variables to predict the outcome.

The general form of each type of regression is:

Simple linear regression: $Y = a + bX + u$

Multiple linear regression: $Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_tX_t + u$

Where:

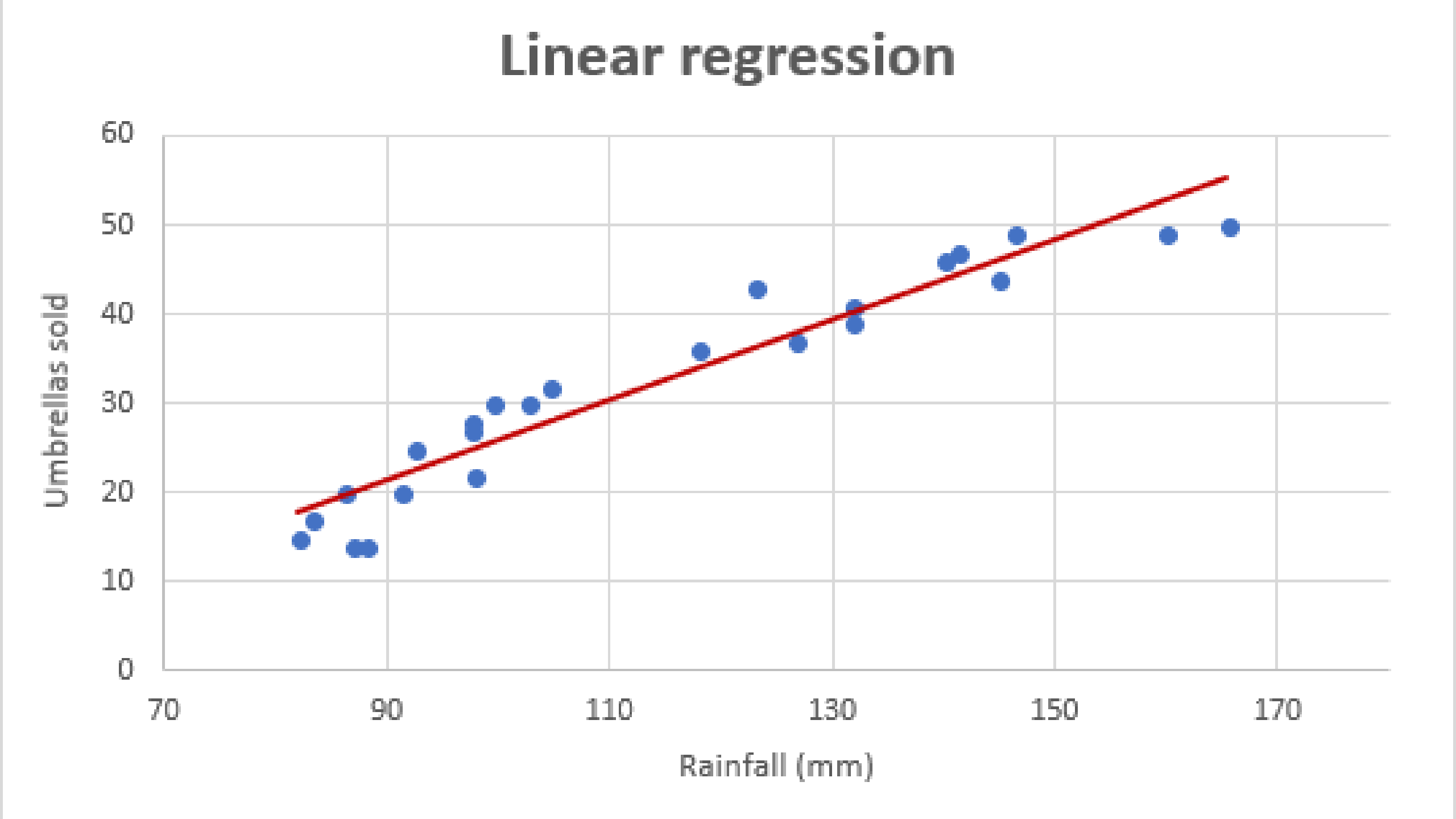
Y = the variable that you are trying to predict (dependent variable).

X = the variable that you are using to predict Y (independent variable).

a = the intercept.

b = the slope.

u = the regression residual.


Linear regression
Umbrellas sold
Rainfall (mm)
60
50
40
30
20
10
0
70
90
110
130
150
170

Regression takes a group of random variables, thought to be predicting Y, and tries to find a mathematical relationship between them. This relationship is typically in the form of a straight line (linear regression) that best approximates all the individual data points. In multiple regression, the separate variables are differentiated by using subscripts.

LET'S CODE NOW:

OBTAINING TRUSTWORTHY REGRESSION RESULTS

With the vast power of using regression comes great responsibility. Sorry, but that's the way it must be. To obtain regression results that you can trust, you need to do the following:

- **Specify the correct model. As we saw, if you fail to include all the important variables in your model, the results can be biased.**
- **Check your residual plots. Be sure that your model fits the data adequately.**
- **Correlation between the independent variables is called multicollinearity. Some multicollinearity is OK. However, excessive multicollinearity can be a problem.**
- **Using regression analysis gives you the ability to separate the effects of complicated research questions. You can disentangle the spaghetti noodles by modeling and controlling all relevant variables, and then assess the role that each one plays.**

CONCLUSION

Regression is a parametric technique used to predict continuous (dependent) variables given a set of independent variables. It is parametric in nature because it makes certain assumptions (discussed next) based on the data set. If the data set follows those assumptions, regression gives incredible results.

THANK YOU