You probably know by now that whenever possible you should be making [data-driven decisions at work](https://hbr.org/2014/05/an-introduction-to-data-driven-decisions-for-managers-who-dont-like-math). But do you know how to parse through all the data available to you? The good news is that you probably don’t need to do the number crunching yourself (hallelujah!) but you do need to correctly understand and interpret the analysis created by your colleagues. One of the most important types of data analysis is called regression analysis.

To better understand this method and how companies use it, I talked with Thomas Redman, author of [*Data Driven: Profiting from Your Most Important Business Asset*](https://www.amazon.com/Data-Driven-Profiting-Important-Business/dp/1422119122). He also advises organizations on their data and data-quality programs.

**What is Regression Analysis?**

Redman offers this example scenario: Suppose you’re a sales manager trying to predict next month’s numbers. You know that dozens, perhaps even hundreds of factors — from the weather to a competitor’s promotion to the rumor of a new and improved model — can impact the numbers. Perhaps people in your organization even have a theory about what will have the biggest effect on sales. “Trust me. The more rain we have, the more we sell.” “Six weeks after the competitor’s promotion, sales jump.”

Regression analysis is a way of mathematically sorting out which of those variables does indeed have an impact. It answers the questions: Which factors matter most? Which can we ignore? How do those factors interact with one another? And, perhaps most important, how certain

are we about all these factors?

In regression analysis, those factors are called “variables.” You have your dependent variable — the main factor that you’re trying to understand or predict. In Redman’s example above, the dependent variable is monthly sales. And then you have your independent variables— the factors you suspect have an impact on your dependent variable.

## How Does it Work?

To conduct a regression analysis, you gather the data on the variables in question. (Reminder: You likely don’t have to do this yourself, but it’s helpful for you to understand the process your data analyst colleague uses.) You take all your monthly sales numbers for, say, the past three years and any data on the independent variables you’re interested in. So, in this case, let’s say you find out the average monthly rainfall for the past three years as well. Then you plot all that information on a chart that looks like this:

The y-axis is the amount of sales (the dependent variable, the thing you’re interested in, is always on the y-axis), and the x-axis is the total rainfall. Each blue dot represents one month’s data—how much it rained that month and how many sales you made that same month.

Glancing at this data, you probably notice that sales are higher on days when it rains a lot. That’s interesting to know, but by how much? If it rains three inches, do you know how much you’ll sell? What about if it rains four inches?

Now imagine drawing a line through the chart above, one that runs roughly through the middle of all the data points. This line will help you answer, with some degree of certainty, how much you typically sell when it rains a certain amount.

## How Do Companies Use It?

Regression analysis is the “go-to method in analytics,” says Redman. And smart companies use it to make decisions about all sorts of business issues. “As managers, we want to figure out how we can affect sales, retain employees, or recruit the best people. It helps us figure out what we can do.”

Most companies use regression analysis to explain a phenomenon they want to understand (for example, Why did customer service calls drop last month?); predict things about the future (for example, What will sales look like over the next six months?); or to decide what to do (for example, Should we go with this promotion or a different one?).

A note about “correlation is not causation”: Whenever you work with regression analysis or any other analysis that tries to explain the impact of one factor on another, you need to remember the important adage: Correlation is not causation. This is critical. Here’s why: It’s easy to say that there is a correlation between rain and monthly sales. The regression shows that they are indeed related. But it’s an entirely different thing to say that rain caused the sales. Unless you’re selling umbrellas, it might be difficult to prove that there is cause and effect.

Sometimes [factors that are so obviously not connected by cause and effect are correlated](https://tylervigen.com/spurious-correlations), but more often in business, it’s not so obvious. When you see a correlation from a regression analysis, you can’t make assumptions, says Redman. Instead, “you must go out and see what’s happening in the real world. What’s the physical mechanism that’s causing the relationship?” Observe consumers buying your product in the rain, talk to them, and find out what is actually causing them to make the purchase. “A lot of people skip this step, and I think it’s because they’re lazy. The goal is not to figure out what is going on in the data but to figure out is what is going on in the world. You must pound the pavement,” he says. Redman wrote about [his own experiment and analysis in trying to lose weight](https://hbr.org/2014/02/how-to-explore-cause-and-effect-like-a-data-scientist) and the connection between his travel and weight gain. He noticed that when he traveled, he ate more and exercised less. Was his weight gain caused by travel? Not necessarily. “It was nice to quantify what was happening, but travel wasn’t the cause. It may be related,” he says, but it’s not like his being on the road put those extra pounds on. He had to understand more about what was happening during his trips. “I’m often in new environments, so maybe I’m eating more because I’m nervous.” He needed to look more closely at the correlation. And this is his advice to managers: Use the data to guide more experiments, not to make conclusions about cause and effect.

egression Analysis, a statistical technique, is used to evaluate the relationship between two or more variables. Regression analysis helps an organisation to understand what their data points represent and use them accordingly with the help of business analytical techniques in order to do better decision-making. In this analysis, you will understand how the typical value of the dependent variable changes when one of the independent variables is varied, while the other independent variables are held fixed. Business analysts and data professionals use this powerful statistical tool for removing the unwanted variables and select the important ones. There are several ways that an organisation use

#### **| Decision-making**

Organisations, in order to run smoothly as well as efficiently, need better decisions and must understand the effects of the decision taken. Organisations collect data about sales, investments, expenditures and other parameters and analyse it for improvement. The regression analysis helps the organisations to make sense of the data which is then used for gaining insights into an organisation. Business analysts and data professionals use the [regression analysis](https://analyticsindiamag.com/how-to-code-linear-regression-models-with-r/) to make strategic business decisions.

#### **Optimisation Of Business**

The motive of regression analysis is to turn the collected data into actionable insights. The organisations are adopting data-driven decision making which eliminates the old-school techniques like guesswork or assuming a hypothesis which eventually improves the performance of work in an organisation. This analysis provides practical assistance to the management unit of the organisation. With lots of data available, the data can be analysed as well as understood to gain efficient insights and work smartly.

#### **3| Predictive Analysis**

Organisations use regression analysis in order to predict future events. In this process, the business analysts predict the man of the dependent variables for given specific values of the dependent variables. The multivariate linear regression is used for various important purposes such as forecasting sale volumes or create growth plans, etc. According to this [article](https://statisticsbyjim.com/regression/predictions-regression/), the general procedure for using regression in order to make good predictions are mentioned below:

#### **4| Risk Analysis**

Most of the time while analysis data, the analyst make mistakes and made a confusion between correlation and causation. It is important to note and understand that correlation is not causation. The regression analysis is used by the organisations to assess the risk in the financial domain and other such domains and thus guide to make crucial business decisions. The linear regression analysis developed the capital asset pricing model which helps in taking financial decisions.

#### **5| Understand The Failures**

Besides the [analysis](https://analyticsindiamag.com/how-ixigo-increased-its-user-base-by-110-using-ai-and-analytics/) of data, regression analysis also helps an organisation to understand any failure and learns from it to correct in the future analysis. Thee regression analysis basically provides quantitative support for the decision-making process. Predicting success can be said as one of the main assets of regression analysis such as analysing the data points of previous sales data as well as current sales data in an organisation to understand and predict the future success.

### **Use Cases**

**Pharmaceutical Companies**

In pharmaceutical companies, regression analysis is used to analyse the quantitative stability data for the retest period or estimation of shelf life. In this approach, the nature of the relationship between an attribute and time determines whether the data should be transformed for linear regression analysis or non-linear regression analysis.

**Finance**

Simple linear regression is also known as Ordinary Least Squares (OLS) provides an overall rationale for the placing of the line of the best fit among the data points which are being studied. This tool is commonly used in forecasting and financial analysis. Another application such as the statistical method is fundamental to the [Capital Asset Pricing Model](https://corporatefinanceinstitute.com/resources/knowledge/finance/what-is-capm-formula/) (CAPM) which describes the relationship between the expected return and risk of investing in a security.

**Credit Card**

In the credit card company, regression analysis helps in understanding various factors like customer’s risk of credit default, expected consumer behaviour, prediction of credit balance, etc. and based on these results the company implements specific EMI options while minimising the default among risky customers.