Top of Form

**Chapter 1.3**

**What is Simple Linear Regression?**

**Simple linear regression** is a statistical method that allows us to summarize and study relationships between two continuous (quantitative) variables:

* One variable, denoted *x*, is regarded as the **predictor**, **explanatory**, or **independent** variable.
* The other variable, denoted *y*, is regarded as the **response**, **outcome**, or **dependent** variable.

Because the other terms are used less frequently today, we'll use the "**predictor**" and "**response**" terms to refer to the variables encountered in this course. The other terms are mentioned only to make you aware of them should you encounter them. Simple linear regression gets its adjective "simple," because it concerns the study of only one predictor variable. In contrast, multiple linear regression, which we study later in this course, gets its adjective "multiple," because it concerns the study of two or more predictor variables.

**Types of relationships**

Before proceeding, we must clarify what types of relationships we won't study in this course, namely, **deterministic** (or **functional**) **relationships**. Here is an example of a deterministic relationship.

Note that the observed (*x*, *y*) data points fall directly on a line. As you may remember, the relationship between degrees Fahrenheit and degrees Celsius is known to be:

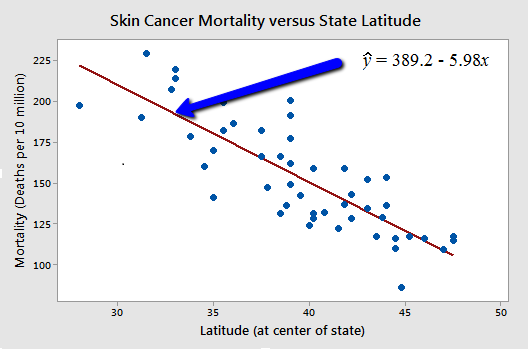
That is, if you know the temperature in degrees Celsius, you can use this equation to determine the temperature in degrees Fahrenheit *exactly*.

Here are some examples of other deterministic relationships that students from previous semesters have shared:

* Circumference = π × diameter
* Hooke's Law: *Y* = α +*βX*, where *Y* = amount of stretch in a spring, and *X* = applied weight.
* Ohm's Law: *I* = *V*/*r*, where *V* = voltage applied, *r* = resistance, and *I* = current.
* Boyle's Law: For a constant temperature, *P* = α/*V*, where *P* = pressure, α = constant for each gas, and *V* = volume of gas.

For each of these deterministic relationships, the equation *exactly* describes the relationship between the two variables. This course does not examine deterministic relationships. Instead, we are interested in **statistical relationships**, in which the relationship between the variables is not perfect.

Here is an example of a statistical relationship. The response variable *y* is the mortality due to skin cancer (number of deaths per 10 million people) and the predictor variable *x* is the latitude (degrees North) at the center of each of 49 states in the U.S. ([skincancer.txt](https://online.stat.psu.edu/stat462/sites/onlinecourses.science.psu.edu.stat462/files/data/skincancer/index.txt)) (The data were compiled in the 1950s, so Alaska and Hawaii were not yet states, and Washington, D.C. is included in the data set even though it is not technically a state.)



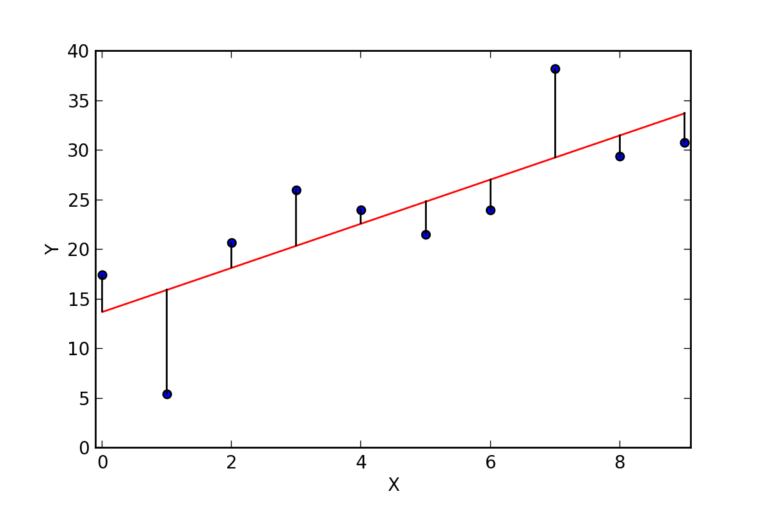
You might anticipate that if you lived in the higher latitudes of the northern U.S., the less exposed you'd be to the harmful rays of the sun, and therefore, the less risk you'd have of death due to skin cancer. The scatter plot supports such a hypothesis. There appears to be a negative linear relationship between latitude and mortality due to skin cancer, but the relationship is not perfect. Indeed, the plot exhibits some "**trend**," but it also exhibits some "**scatter**." Therefore, it is a statistical relationship, not a deterministic one.

**Regression lines**

**Why Regression lines are important?**

Regression lines are useful in forecasting procedures. Its purpose is to describe the interrelation of the dependent variable(y variable) with one or many independent variables(x variable).

Using the equation obtained from the regression line acts as an analyst who can forecast future behaviors of the dependent variables by inputting different values for the independent ones.



Regression Line Formula: y = a + bx + u

Multiple Regression Line Formula: y= a + b1x1 +b2x2 + b3x3 +…+ btxt + u

**Where linear regression is used?**

Regression lines are used in the financial sector and in business. Various financial analyst employs linear regressions to forecast stock prices, commodity prices and to perform valuations for many different securities. Various companies employ linear regressions for the purpose of forecasting sales, inventories, and many other variables.

**Book Reading and Video Material**

·        Understanding Machine Learning: From Theory to Algorithms by Shai Shalev-Shwartz and Shai Ben-David-Cambridge University Press 2014 [Download](https://www.cse.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf) Buy at Amazon

·        Introduction to Machine Learning – the Wikipedia guide [Download](http://datascienceassn.org/sites/default/files/Introduction%20to%20Machine%20Learning.pdf)

https://www.toppr.com/guides/fundamentals-of-business-mathematics-and-statistics/correlation-and-regression/regression-lines/

Bottom of Form