

8.1: Introduction to Linear Regression and Correlation



Figure 13.1 Linear regression and correlation can help you determine if an auto mechanic's salary is related to his work experience.
(credit: Joshua Rothhaas)

Professionals often want to know how two or more numeric variables are related. For example, is there a relationship between the grade on the second math exam a student takes and the grade on the final exam? If there is a relationship, what is the relationship and how strong is it?

In another example, your income may be determined by your education, your profession, your years of experience, and your ability, or your gender or color. The amount you pay a repair person for labor is often determined by an initial amount plus an hourly fee.

These examples may or may not be tied to a model, meaning that some theory suggested that a relationship exists. This link between a cause and an effect, often referred to as a model, is the foundation of the scientific method and is the core of how we determine what we believe about how the world works. Beginning with a theory and developing a model of the theoretical relationship should result in a prediction, what we have called a hypothesis earlier. Now the hypothesis concerns a full set of relationships. As an example, in Economics the model of consumer choice is based upon assumptions concerning human behavior: a desire to maximize something called utility, knowledge about the benefits of one product over another, likes and dislikes, referred to generally as preferences, and so on. These combined to give us the demand curve. From that we have the prediction that as prices rise the quantity demanded will fall. Economics has models concerning the relationship between what prices are charged for goods and the market structure in which the firm operates, monopoly verse competition, for example. Models for who would be most likely to be chosen for an on-the-job training position, the impacts of Federal Reserve policy changes and the growth of the economy and on and on.

Models are not unique to Economics, even within the social sciences. In political science, for example, there are models that predict behavior of bureaucrats to various changes in circumstances based upon assumptions of the goals of the bureaucrats. There are models of political behavior dealing with strategic decision making both for international relations and domestic politics.

The so-called hard sciences are, of course, the source of the scientific method as they tried through the centuries to explain the confusing world around us. Some early models today make us laugh; spontaneous generation of life for example. These early models are seen today as not much more than the foundational myths we developed to help us bring some sense of order to what seemed chaos.

The foundation of all model building is the perhaps the arrogant statement that we know what caused the result we see. This is embodied in the simple mathematical statement of the functional form that $y = f(x)$. The response, Y , is caused by the stimulus, X . Every model will eventually come to this final place and it will be here that the theory will live or die. Will the data support this hypothesis? If so then fine, we shall believe this version of the world until a better theory comes to replace it. This is the process by which we moved from flat earth to round earth, from earth-center solar system to sun-center solar system, and on and on.

The scientific method does not confirm a theory for all time: it does not prove "truth". All theories are subject to review and may be overturned. These are lessons we learned as we first developed the concept of the hypothesis test earlier in this book. Here, as we begin this section, these concepts deserve review because the tool we will develop here is the cornerstone of the scientific method and the stakes are higher. Full theories will rise or fall because of this statistical tool; regression and the more advanced versions call econometrics.

In this chapter we will begin with correlation, the investigation of relationships among variables that may or may not be founded on a cause and effect model. The variables simply move in the same, or opposite, direction. That is to say, they do not move randomly. Correlation provides a measure of the degree to which this is true. From there we develop a tool to measure cause and effect relationships; regression analysis. We will be able to formulate models and tests to determine if they are statistically sound. If they

are found to be so, then we can use them to make predictions: if as a matter of policy we changed the value of this variable what would happen to this other variable? If we imposed a gasoline tax of 50 cents per gallon how would that effect the carbon emissions, sales of Hummers/Hybrids, use of mass transit, etc.? The ability to provide answers to these types of questions is the value of regression as both a tool to help us understand our world and to make thoughtful policy decisions.

This page titled [8.1: Introduction to Linear Regression and Correlation](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- **13.0: Introduction to Linear Regression and Correlation** by [OpenStax](#) is licensed [CC BY 4.0](#). Original source: <https://openstax.org/details/books/introductory-business-statistics>.