

MITOCW | MIT15_071S17_Session_2.2.01_300k

In this lecture, we introduce linear regression a simple but very powerful method to analyze data and make predictions and apply it in a very unexpected context-- predicting the quality of wines.

Bordeaux is a region in France popular for producing wine.

While this wine has been produced in much the same way for hundreds of years, there are differences in price and quality from year to year that are sometimes very significant.

Bordeaux wines are widely believed to taste better when they are older, so there's an incentive to store young wines until they are mature.

The main problem is that it is hard to determine the quality of the wine when it is so young just by tasting it, since the taste will change so significantly by the time it will actually be consumed.

This is why wine tasters and experts are helpful.

They taste the wines and then predict which ones will be the best wines later.

The question we'll address in this lecture-- can analytics model this process better and make stronger predictions?

On March 4, 1990, the New York Times announced that Princeton economics professor Orley Ashenfelter can predict the quality of Bordeaux wine without tasting a single drop.

Ashenfelter's predictions have nothing to do with assessing the aroma of the wine, looking at the legs, or declaring that the wine tastes citrusy, oaky, or nutty.

They are the results of a mathematical model.

Ashenfelter used a method called linear regression.

The method predicts an outcome variable or dependent variable.

And in doing so, it uses a set of what is called independent variables.

For the dependent variable, Ashenfelter chose a typical price in 1990-1991 for Bordeaux wine in an auction.

This approximates quality.

As independent variables, he used age of the wine-- so the older wines are more expensive-- and weather-related information, specifically the average growing season temperature, the harvest rain, and winter rain.

In these figures, we depict the data during the period from 1952 to 1978.

There are four independent variables-- the age of the wine, the average growing season temperature, the harvest rain, and winter rain.

And on the vertical axis, you observe the logarithm of the price, the realization in an auction.

So these are the primitive data that Ashenfelter used.

So Ashenfelter believed that his predictions are more accurate than those of the world's most influential wine critic.

His name is Robert Parker.

In response, Parker called Ashenfelter to be "an absolute total sham," and he adds that, "rather like a movie critic who never goes to see the movie but tells you how good it is based on the actors and the director."