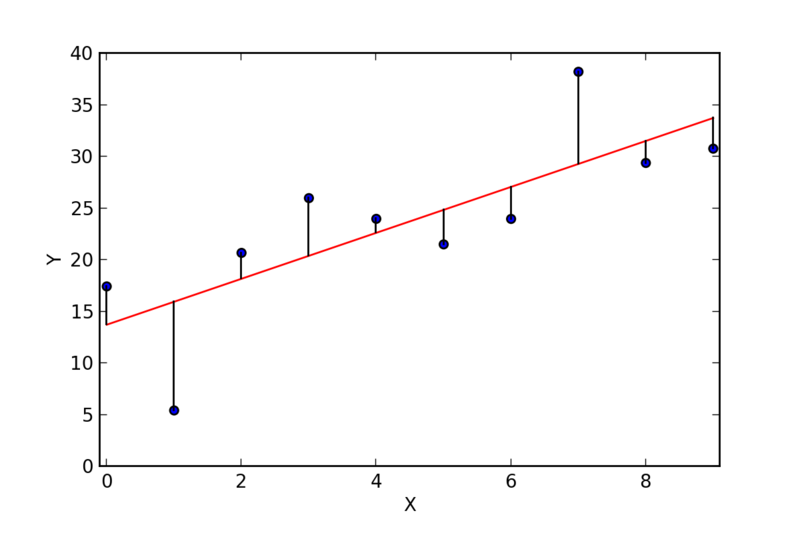
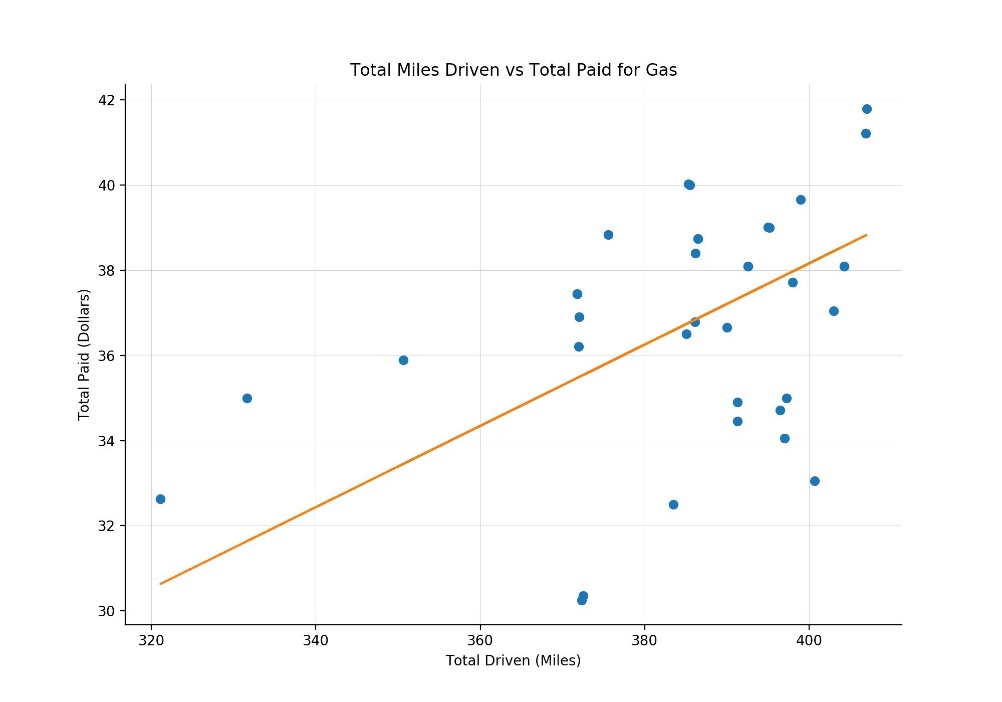
In case one hasn’t noticed, the modern world practically runs on predictions. Weather predictions, stock predictions, market trend predictions, and oil price predictions are just some of the predictions that are prevalent in everyday life. One might be asking how all these predictions are accomplished. Fortunately, the answer is a lot simpler than having a trio of predictors a la Minority Report. These predictions are accomplished via the use of linear regression, which make use of a least-squares regression line to predict future values of dependent variables. I will be discussing what a least-squares fit entails, as well as some things it could be used for and some current real-life use cases.

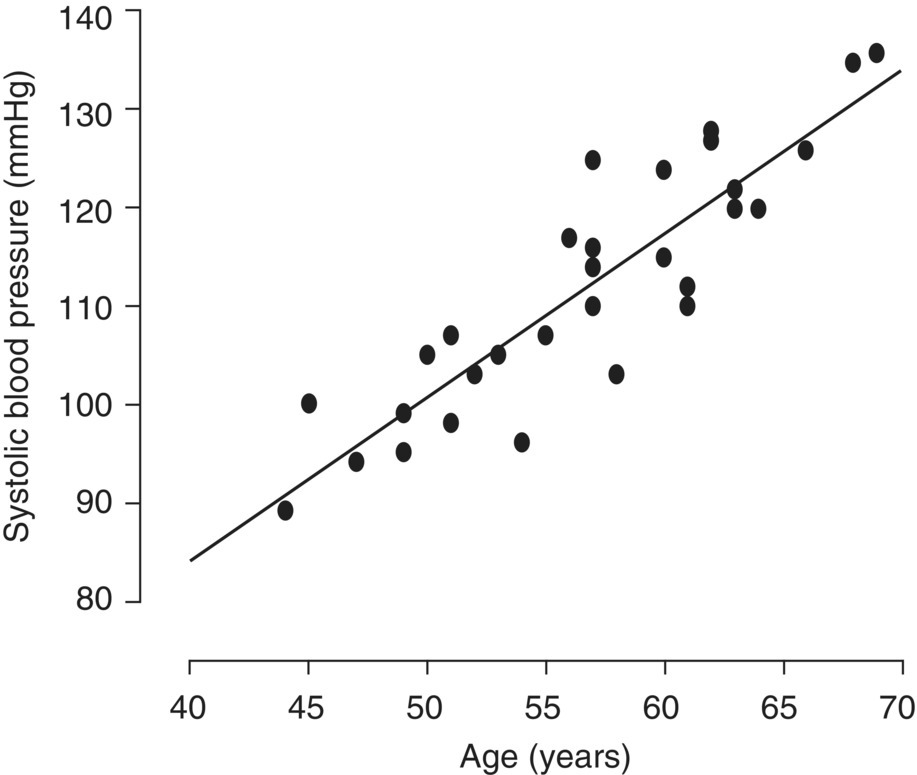
A set of collected data points that are plotted on a scatter plot may display a linear relationship, and the Pearson correlation coefficient can be calculated to determine the magnitude of said correlation. How, then, can we determine future values of the dependent variable with only knowledge of the current data points? We determine the slope of the linear relationship by adding a least-squares fit line to the data points. This is a line placed on the scatter plot that minimizes the sum of squares of residuals. Residuals are the vertical difference between the actual data points and the least-squares fit line, so minimizing the sum of their squares results in a line whose equation can most accurately predict the dependent variable for future values of the independent variable. Below is an example of a scatter plot with a fitted least-squares regression line, showing how the line minimizes the residuals, and thus, the sum of their squares. The line’s equation would give us an accurate prediction of the y-values.



What could you use the least-squares method for? Let’s say you’re going on a road trip and you want to know how much you’d expect to spend on gas for the whole trip. If you’ve measured how many miles you can drive on one tank and the cost of a full tank of gas for those miles, then you’d have your answer (assume gas price remains constant in an area). A scatter plot of price per tank vs miles driven per tank can be plotted and then a least-squares line can be fitted to the plot. The equation of this line will take our desired miles as input and return the predicted value of the gas price for the trip! Below is a scatter plot illustrating the data and the least-squares line. Note the outliers and how they skew the least-squares line.



In reality, least-squares fits are used for much greater applications than just road trip planning. Since revenue generated in a business commonly follows a linear trend, businesses frequently use linear least-square fits to predict quarterly revenue based on historical marketing data. For example, by graphing a least-squares fit line on a plot of revenue vs advertising expenses, a business can know how much they can expect to spend on marketing if they desire a certain revenue. Meanwhile, in the health sciences, medical researchers use these fits to predict the blood pressure of patients from different conditions such as age, height, weight, blood solute levels, or drug presence. Below is an example scatter plot of blood pressure vs age with an overlaid least-squares fit line. The equation of the line will tell us what blood pressure we can expect for a person of a certain age.



I use linear regression at my own work as well. I ferment bacteria in bio-reactors (closed mixing vats with sensors attached) and measure their sensor output such as acidity, temperature, and oxygen use. If I construct a scatter plot of one of these parameters vs time and then overlay a least-squares fit line, I can know what levels of acid, oxygen, metabolites, etc. to expect at a certain time point. This allows me to catch anomalies if said levels become unusual.

While linear regression helps to save time and resources, one must remember that the equation of the line is still a prediction that only applies to linear relationships, hence the name. Despite the line minimizing the sum of squares of residuals, the residuals remain, meaning some error in the model will inevitably remain (as was discussed previously, a model is never perfect).

References:

https://basicmedicalkey.com/regression-analyses/

https://www.ehdp.com/methods/least-squares-regression-1.htm

https://towardsdatascience.com/linear-regression-in-real-life-4a78d7159f16

http://www.intellspot.com/linear-regression-examples/