# Initial Posts

|  |
| --- |
| **What is Regression Analysis?**  The basic point of regression analysis is to try and find out if variables are related and if so, the strength and significance of the relationship. "The goal of regression analysis is to describe the relationship between one set of variables, called the dependent variables, and another set of variables, called independent or explanatory variables." (Downey, 2014). The ultimate goal would be to find independent (aka explanatory) variables whose variance explains the variance in the dependent (aka outcome) variables. |
| **Multiple vs. Simple Regression**  Simple Regression: When there is only 1 explanatory variable.  Multiple Regression: When there are more than 1 explanatory variables.  Regression formula:    *Formula obtained from Google.com* |
| **Linear Regression**  Linear regression describes the relationship between variables via a straight line. The best fitting line would have an intercept and slope which minimizes the error.  For simple linear regression this would be:    For multiple regression, this would be:    Although a 'best' line can be found between the variables, it does not necessarily mean that it explains the relationship well. Also, what is interesting is that for simple regression, we are fitting a line and if there are 2 explanatory variables, we are fitting a plane. As we keep adding more and more explanatory variables the object we are fitting grows in dimensions which becomes visually uninterpretable for us 3 dimensional beings (*nerd alert!*).  Also, as mentioned in Think Stats, data can be transformed into a non-linear version of itself and then linear regression ran but the analysis is still considered linear. "Using computed variables like agepreg2 [squared variable] is a common way to fit polynomials and other functions to data. This process is still considered linear regression, because the dependent variable is a linear function of the explanatory variables, regardless of whether some variables are non-linear functions of others." (Downey, 2014).  Reference:  Downey, Allen B.. Think Stats: Exploratory Data Analysis |
| **Predicting Results**  When a predictive model is created, there are parameters that are used to give weight to the estimators that are used to calculate an outcome. In a simple linear regression model, let's say that that we've created a model which has the below parameters:   * Intercept (aka beta 0 parameter) = 1 * Slope (aka beta 1 parameter) = 2     This would give the model:  prediction = 2\*x + 1  So if we knew 'x', let's say x=2, we can predict that the outcome will equal 11. In other words, the slope (aka beta) parameter tells us about the relationship between the x<-->y variables. In this case, it is telling us that for each unit increase in x, the y value increases by 200%. |
|  |

# Replies

|  |
| --- |
|  |
|  |
|  |
|  |
|  |