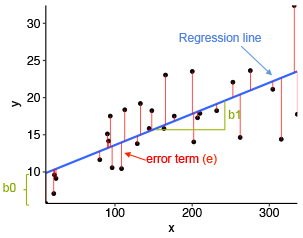
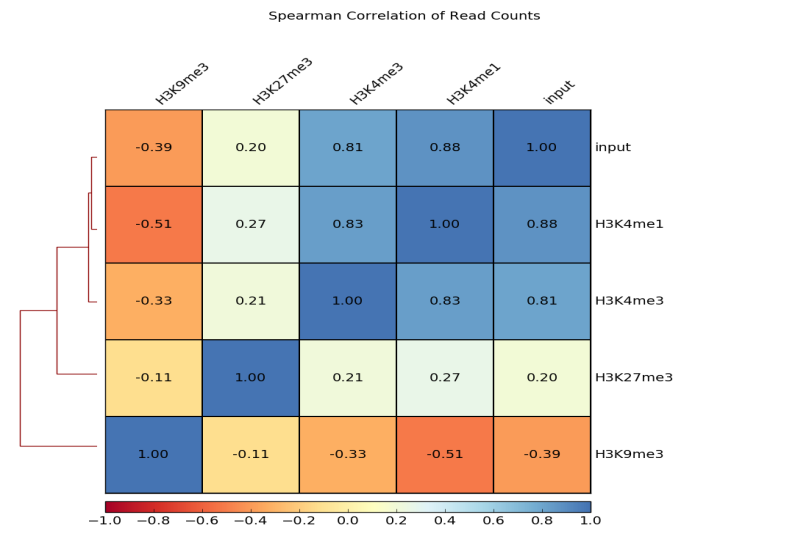
# Linear Regression

Linear Regression is a supervised machine learning algorithm where the predicted output is continuous and has a constant slope. It’s used to predict values within a continuous range, (e.g. sales, price) rather than trying to classify them into categories (e.g. cat, dog).

Linear regression is the next step up after correlation. It is used when we want to predict the value of a variable based on the value of another variable. The variable we want to predict is called the dependent variable (or sometimes, the outcome variable).



Correlation is used to describe the linear relationship between two continuous variables (e.g., height and weight). In general, correlation tends to be used when there is no identified response variable. It measures the strength (qualitatively) and direction of the linear relationship between two or more variables.



As per the given IPL dataset runs scored by each player in the previous year is given and we have to predict how many runs will the players score in the coming season of IPL. As the runs scored is numerical value and to bring ease in maintaining, understanding and running the program we are using regression.

As we have to maintain balance between the accuracy and time complexity, we did the correlation, which gives the correlation matrix. As the datapoints are not large in number, we cannot reduce more no of features. Keeping all these points in mind we have cleaned the dataset based on the requirements of the given problem. We trained our model on the given training data and predicting the required independent feature i.e., the runs players may score in the coming season.