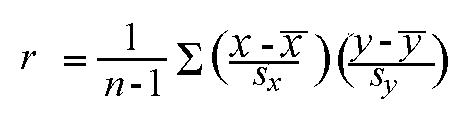
# Correlation

The strength of the linear association between two variables is quantified by the ***correlation coefficient***.  
Given a set of observations (*x1, y1*), (*x2,y2*),...(*xn,yn*), the formula for computing the correlation coefficient is given by  


The correlation coefficient always takes a value between -1 and 1, with 1 or -1 indicating perfect correlation (all points would lie along a straight line in this case). A positive correlation indicates a positive association between the variables (increasing values in one variable correspond to increasing values in the other variable), while a negative correlation indicates a negative association between the variables (increasing values is one variable correspond to decreasing values in the other variable). A correlation value close to 0 indicates no association between the variables.

Since the formula for calculating the correlation coefficient standardizes the variables, changes in scale or units of measurement will not affect its value. For this reason, the correlation coefficient is often more useful than a graphical depiction in determining the strength of the association between two variables.

## Correlation in Linear Regression

The square of the correlation coefficient, *r²*, is a useful value in [linear regression](http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm). This value represents the fraction of the variation in one variable that may be explained by the other variable. Thus, if a correlation of 0.8 is observed between two variables (say, height and weight, for example), then a linear regression model attempting to explain *either* variable in terms of the other variable will account for 64% of the variability in the data.

The correlation coefficient also relates directly to the regression line ***Y = a + bX*** for any two variables, where IMG_257.  
Because the least-squares regression line will always pass through the means of x and y, the regression line may be entirely described by the means, standard deviations, and correlation of the two variables under investigation.

For some good examples of studies using correlation, see the [correlation index](http://lib.stat.cmu.edu/cgi-bin/isearchb?DASL+METHODS/Correlation) in the Statlib Data and Story Library (DASL).