**MATH221, Week 7**

Linear Correlation Coefficient and the Coefficient of Determination

Introduction

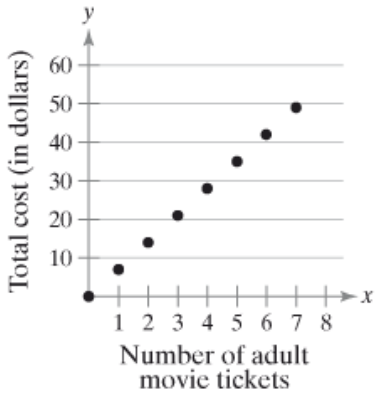
The linear correlation coefficient and the coefficient of determination are values that are used to measure the strength or reliability of the relationship between the independent (or explanatory) variable x and the dependent (or response) variable y.

Linear Correlation Coefficient

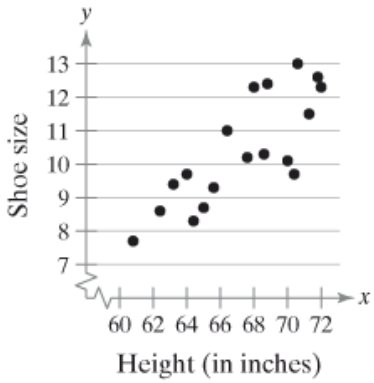
The linear correlation coefficient is the measure of the strength and direction of a linear relationship between two variables. It is represented by the letter **r.** If r is positive, the direction of the correlation is upward from left to right. If r is negative, the direction of the correlation is downward from left to right.

**-1 ≤ r ≤ 1** r must be between -1 and 1 inclusive.

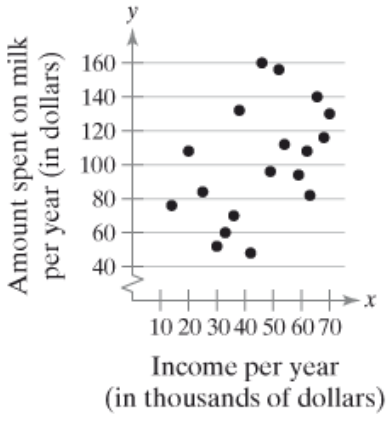
The closer r is to -1 or to 1 the stronger the relationship between the two variables will be. If r = 0, there is no correlation.



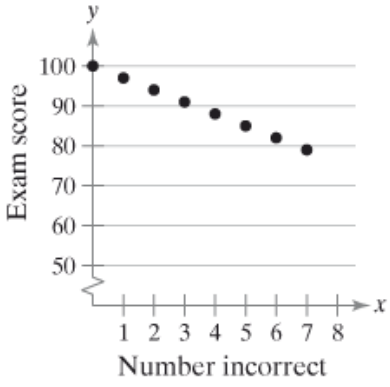

**Perfect positive correlation** $r = 1$



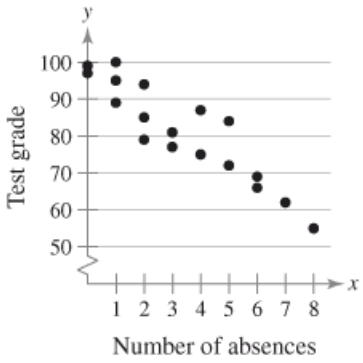

**Strong positive correlation** $r = 0.81$



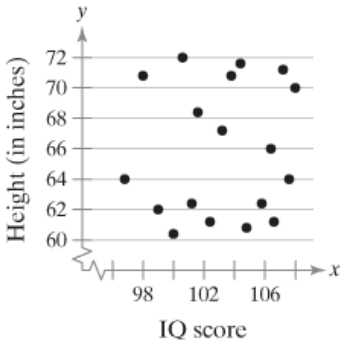

**Weak positive correlation** $r = 0.45$




**Perfect negative correlation** $r = -1$



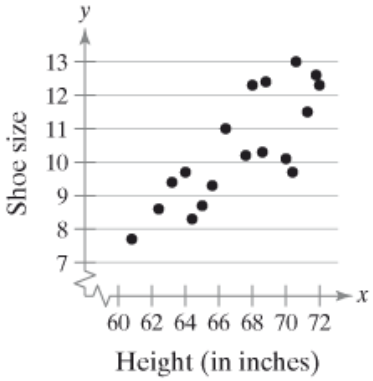

**Strong negative correlation** $r = -0.92$




**No correlation** $r = 0.04$

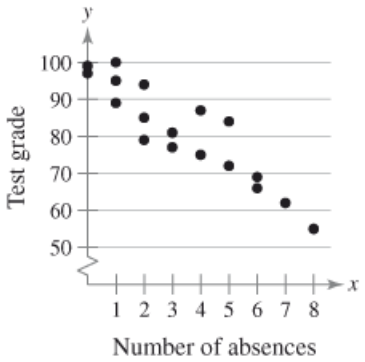
Coefficient of Determination

The coefficient of determination is the percentage of the variation in the dependent variable y that can be explained by the behavior of the independent variable x. The larger this value is the better the independent variable x can be used to explain (or estimate) the value of the dependent variable y. It is designated as **$r^2$**. It is the square of the linear correlation coefficient. It will always be positive.




**Strong positive correlation** $r = 0.81$

The coefficient of determination for height and shoe size is $r^2 = (0.81)^2 = 0.6561 \approx 0.66$
66% of variation in shoe size is explained by ones height in inches.
100% - 66% = 34% of the variation in shoe size is explained by variables other than height or due to random chance.




**Strong negative correlation** $r = -0.92$

$r^2 = (-0.92)^2 = 0.8464 \approx 0.85$

85% of the variation in test grades can be explained by number of absences.

100% - 85% = 15% of the variation in test grades can be explained by variables other than number of absences or due to random chance.