What goes on the x-axis and the y-axis ?

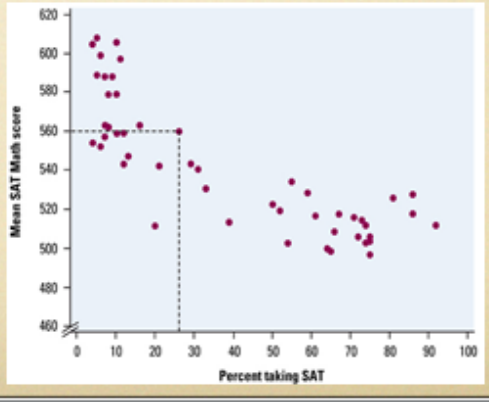
Scatterplots

- Scatterplot shows the relationship between two quantitative variables measured on the same individuals.

* Explanatory variables along X axis, Response variables along Y.

- Each individual in data appears as the point in the plot fixed by the values of both variables for that individual.

- Example:



Association

Positive Association, Negative Association

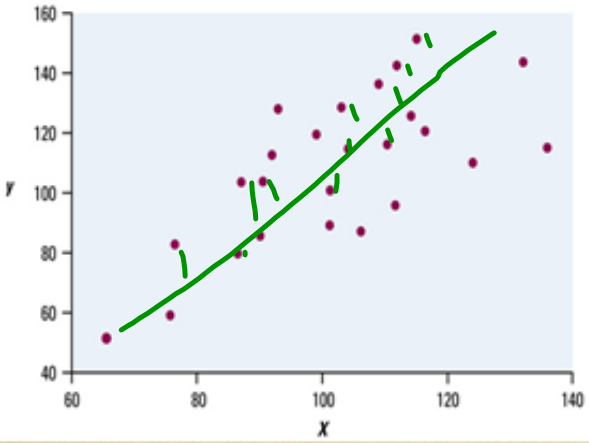
Two variables are **positively associated** when above-average values of one tend to accompany above-average values of the other, and below-average values also tend to occur together.

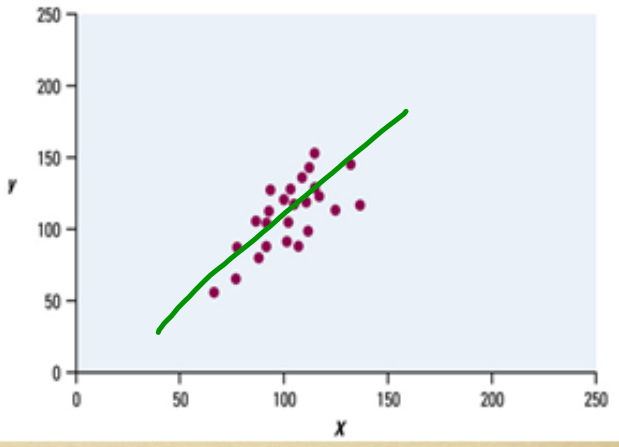
Two variables are **negatively associated** when above-average values of one tend to accompany below-average values of the other, and vice versa.

What happens when you re-scale the data... be careful

Correlation

- Caution- our eyes can be fooled! Our eyes are not good judges of how strong a linear relationship is. The 2 scatterplots depict the same data but drawn with a different scale. Because of this we need a numerical measure to supplement the graph.





Interpreting r

- The absolute value of r tells you the strength of the association (0 means no association, 1 is a strong association)
- The sign tells you whether it's a positive or a negative association. So r ranges from -1 to +1
  - Note- it makes no difference which variable you call x and which you call y when calculating correlation, but stay consistent!
  - Because r uses standardized values of the observations, r does not change when we change the units of measurement of x, y, or both. (Ex: Measuring height in inches vs. ft. won't change correlation with weight)
  - values of -1 and +1 occur ONLY in the case of a perfect linear relationship , when the variables lie exactly along a straight line.

What is Association?

The statistical term association is defined as a relationship between two random variables which makes them statistically dependent. It refers to rather a general relationship without specifics of the relationship being mentioned, and it is not necessary to be a causal relationship.

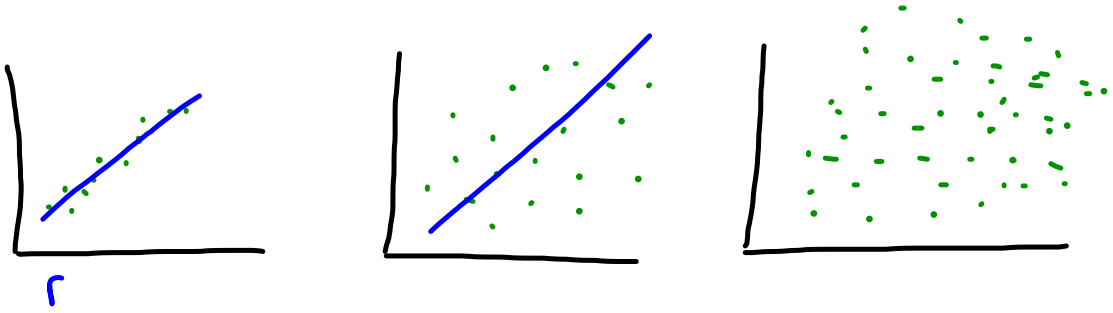
Many statistical methods are used to establish the association between two variables. Pearson's correlation coefficient, odds ratio, distance correlation, Goodman's and Kruskal's Lambda and Spearman's rho (ρ) are a few examples.

What is Correlation?

Correlation is a measure of the strength of the relationship between two variables. The correlation coefficient quantifies the degree of change of one variable based on the change of the other variable. In statistics, correlation is connected to the concept of dependence, which is the statistical relationship between two variables.

The Pearson's correlation coefficient or just the correlation coefficient r is a value between -1 and 1 ($-1 \leq r \leq +1$). It is the most commonly used correlation coefficient and valid only for a linear relationship between the variables. If $r=0$, no relationship exist, and if $r \geq 0$, the relation is directly proportional; the value of one variable increases with the increase in the other. If $r \leq 0$, the relationship is inversely proportional; one variable decreases as the other increases. Because of the linearity condition, correlation coefficient r can also be used to establish the presence of a linear relationship between the variables.

Read more: http://www.differencebetween.com/difference-between-association-and-vs-correlation/#ixzz2hHN3cgnP


r