Statistical Correlation

Explorable.com, Lyndsay T Wilson433K reads

Statistical correlation is a statistical technique which tells us if two variables are related.

For example, consider the variables of family income and family expenditure. It's well known that income and expenditure increase or decrease together. Thus they are related in the sense that change in any one variable is accompanied by change in the other variable.

Likewise, the price and the demand of a commodity are related variables; when price increases, demand will tend to decreases and vice versa.

If the change in one variable is accompanied by a change in the other, then the variables are said to be correlated. We can therefore say that family income and family expenditure are correlated, as are commodity price and demand.

The Relationship Between Variables

Correlation is about the relationship between variables [1]. Correlations tell us:

1. whether this relationship is positive or negative
2. the strength of the relationship.

In the case of family income and family expenditure, it is easy to see that they both rise or fall together in the same direction. This is called a positive correlation.

In case of price and demand, change occurs in opposing directions so that increase in one is accompanied by decrease in the other. This is called a negative correlation.

Coefficient of Correlation

Statistical correlation [2] is measured by what is called the coefficient of correlation (r). Its numerical value ranges from +1.0 to -1.0. It gives us an indication of both the strength and direction of the relationship between variables.

In general, r > 0 indicates a positive relationship, r < 0 indicates a negative relationship and r = 0 indicates no relationship (or that the variables are independent of each other and not related). Here r = +1.0 describes a perfect positive correlation and r = -1.0 describes a perfect negative correlation.

The closer the coefficients are to +1.0 and -1.0, the greater the strength of the relationship between the variables.

As a rule of thumb, the following guidelines on strength of relationship are often useful (though many experts would somewhat disagree on the choice of boundaries).

Value of r	Strength of relationship
-1.0 to -0.5 or 1.0 to 0.5	Strong
-0.5 to -0.3 or 0.3 to 0.5	Moderate
-0.3 to -0.1 or 0.1 to 0.3	Weak
-0.1 to 0.1	None or very weak

Correlation is only appropriate for examining the relationship between meaningful quantifiable data (e.g. air pressure, temperature) rather than categorical data such as gender, color etc.

Disadvantages

While 'r' (the correlation coefficient) is a powerful tool, it has to be handled with care.

1. The most used correlation coefficients only measure linear relationship [3]. It is therefore perfectly possible that while there is strong non linear relationship [4] between the variables [5], r is close to 0 or even 0. In such a case, a scatter diagram can roughly indicate the existence or otherwise of a non linear relationship.
2. One has to be careful in interpreting the value of 'r'. For example, it has been shown [6] that the number of people who have fallen into swimming pools each year since 1999 correlates with the number of films Nicolas Cage has appeared in. Obviously, irrespective of the value of 'r', this is what's called a non-sense correlation - and for good reason!
3. 'r' should never be used to say anything about a cause and effect relationship [7]. Put

differently, by examining the value of 'r', we could only conclude that variables X and Y are related. However the same value of 'r' does not tell us if X influences Y or the other way round. Statistical correlation should not be the primary tool used to study causation [8], because of the problem with third variables [9].

Source URL: https://explorable.com/statistical-correlation

Links
[1] https://explorable.com/relationship-between-variables
[2] http://www.surveysystem.com/correlation.htm
[3] https://explorable.com/linear-relationship
[4] https://explorable.com/non-linear-relationship
[5] https://explorable.com/research-variables
[6] http://www.tylervigen.com/view_correlation?id=359
[7] https://explorable.com/cause-and-effect
[8] https://explorable.com/correlation-and-causation
[9] https://explorable.com/confounding-variables