

PROPERTIES OF CORRELATION

1. Correlation requires that both variables be quantitative (numerical).

You can't calculate a correlation between "income" and "city of residence" because "city of residence" is a qualitative (non-numerical) variable.

2. Positive r indicates positive association between the variables, and negative r indicates negative association.

A positive r indicates that above average values of x tend to be matched with above average values of y and below average values of x tend to be matched with below average values of y .

POSITIVE r high with high, low with low

A negative r indicates that above average values of x tend to be matched with below average values of y and below average values of x tend to be matched with above average values of y .

NEGATIVE r high with low, low with high

3. The correlation coefficient (r) is always a number between -1 and +1.

Values of r near 0 indicate a very weak linear relationship. The extreme values of -1 and +1 indicate the points in a scatterplot lie exactly along a straight line.

4. The correlation coefficient (r) is a pure number without units.

r is not affected by:

--interchanging the two variables
(it makes no difference which variable is called x and which is called y)

--adding the same number to all the values of one variable

--multiplying all the values of one variable by the same positive number

Because r uses the standardized values of the observations, r does not change when we change units of measurement (inches vs. centimeters, pounds vs. kilograms, miles vs. meters). r is "scale invariant".

- 5. The correlation coefficient measures clustering about a line, but only relative to the SD's.**

Pictures can be misleading.

- 6. The correlation can be misleading in the presence of outliers or nonlinear association.**

r does not describe curved relationships. r is affected by outliers. When possible, check the scatterplot.

- 7. Ecological correlations based on rates or averages tend to overstate the strength of associations.**

(See demo problem on worksheet #6)

- 8. Correlation measures association. But association does not necessarily show causation.**

Both variables may be influenced simultaneously by some third variable.