1 Correlation

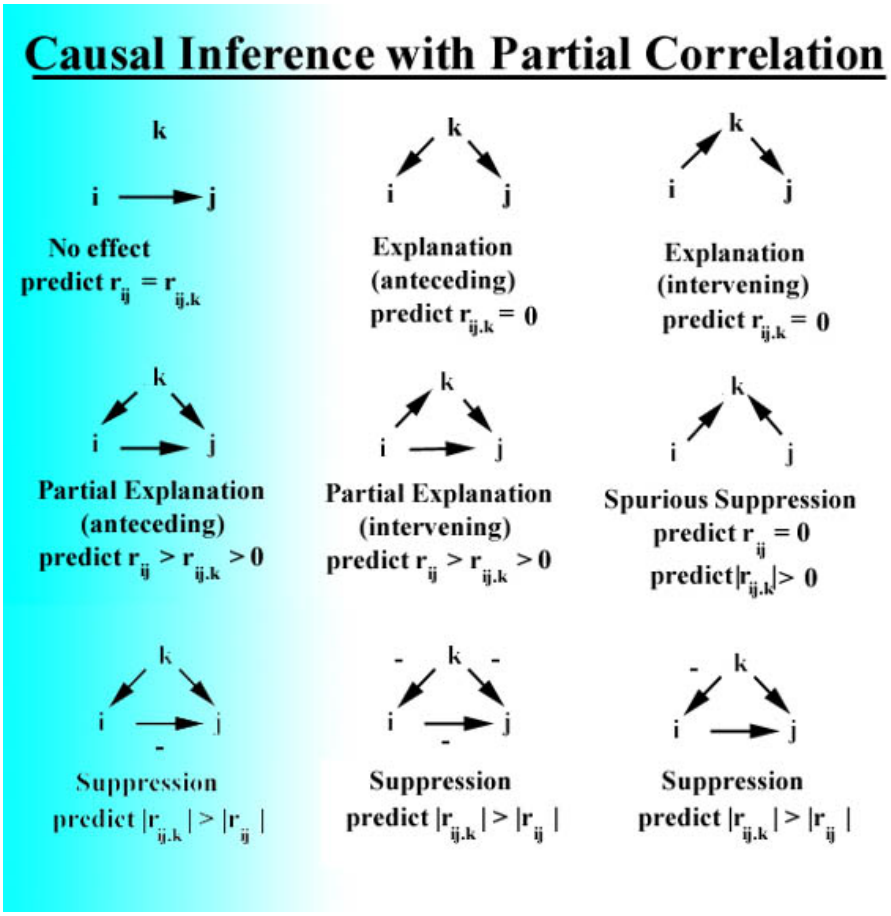
Pearson's correlation coefficient, r_{xy}, for two numerical variables measures the strength of the linear relationship between x and y.
We saw

$$r_{xy} = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$$

It is conceivable that the relationship between the variables is mostly due to a shared relationship with a third variable, z.
For example, if measuring the average number of TVs per household (x_1), and mean life expectancy (y) for different countries, we would find a high correlation between the two variables, but the relationship can probably explained through the mean per capita income (x_2) for the different countries, or other measures of economic wealth.
In order to study this effect we find the partial correlation between y and x_1, correcting for x_2, where both (y and x_1 are corrected for x_2)

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{x_1x_2}r_{yx_2}}{\sqrt{1 - r_{x_1x_2}^2}\sqrt{1 - r_{yx_2}^2}}$$

Which would be close to zero for the example, indicating that in fact the correlation between y and x_1 is only due to the relationship of both with x_2.



Example 1
The Wechsler Adult Intelligence Scale (WAIS) is a device often used to measure "intelligence" beyond the years of childhood. Among its several sub-scales are three labeled as C, A, and V. The "C" stands

for "comprehension," which chiefly reflects the test-taker's ability to comprehend the meanings and implications of written passages.
The "A" refers to the test-taker's ability to perform tasks that require arithmetic ability.
And the "V" stands for "vocabulary," which is a measure that increases or decreases in accordance with the breadth of the test-taker's vocabulary within the domain of the language in which the test is constructed.
The following table shows the correlations typically found among these three sub-scales.

C versus A: $r_{CA} = +.49$ $r^2_{CA} = .24$
C versus V: $r_{CV} = +.73$ $r^2_{CV} = .53$ then
A versus V: $r_{AV} = +.59$ $r^2_{AV} = .35$

$$r_{AC \cdot V} = \frac{r_{AC} - r_{AV} r_{CV}}{\sqrt{1 - r^2_{AV}} \sqrt{1 - r^2_{CV}}} = 0.11$$

Hence $r^2_{AC \cdot V} = 0.01$.
With the effects of vocabulary removed, the correlation between comprehension and arithmetic ability collapses down to hardly anything at all.
The practical inference is that if we were to administer the WAIS to a sample of subjects who were homogeneous with respect to breadth of vocabulary, the correlation between their scores on the comprehension and arithmetic sub-scales would prove fairly scant, on the order of r=+.11 and r2=.01.

The partial correlation is when squared interpreted as the percent of unique variance in y uniquely accounted for by x_1, after both y and x_1 are controlled by x_2. (it is the correlation of the residuals of y (for the model $y = \beta_0 + \beta_1 x_2 + e$)with the residuals of x_1 (for the model $x_1 = \beta_0 + \beta_1 x_2 + e$)).

Semi partial correlation or part correlation is the correlation between response variable y and predictor x_1, when correcting only x_1 for x_2. This is the appropriate correlation coefficient to use for multiple regression, in order to evaluate, if it would be meaningful to add x_1 to a model relating already y with x_2. A small part correlation would indicate that all the information that x_1 bears on y is already included by x_2.
The squared part correlation of y and x_1, correcting for x_2 is

$$r^2_{yx_1(x_2)} = R^2_{y,x_1x_2} - R^2_{y,x_2}$$

(R^2_{y,x_1x_2}=coefficient of determination for the multiple linear regression model for response y and predictors x_1 and x_2.)
The squared part correlation is the increase in the coefficient of determination, when adding x_1 to the model, including already the variable we are correcting for, or is interpreted as the percent of total (unique plus joint) variance in y uniquely accounted for by x_1 and not by x_2. (It is the correlation of the residuals (for the model $x_1 = \beta_0 + \beta_1 x_2 + e$) and y).