Partial Correlation

Reading an association after linear adjustment for a third variable

Ratiomera Statistics

Table of contents

Course-Page Summary

Partial correlation asks whether two quantitative variables still move together linearly after each has been adjusted for the same third variable. It is an adjusted association, not an experimental effect and not a prediction equation.

Two Equivalent Routes

Let X and Y be the focal variables and Z the variable being adjusted for.

1. Regress X on Z and keep the residuals e_X.
2. Regress Y on Z and keep the residuals e_Y.
3. Calculate the correlation between e_X and e_Y.

Each residual is the part of its focal variable not linearly predicted by Z in that fitted model. Their correlation is $r_{XY \cdot Z}$.

With one adjustment variable, the same result follows from the three pairwise correlations:

$$r_{XY \cdot Z} = \frac{r_{XY} - r_{XZ} r_{YZ}}{\sqrt{(1 - r_{XZ}^2)(1 - r_{YZ}^2)}}.$$

Compare the Raw and Adjusted Questions

Table 1: Raw and partial correlations answer related but different questions.

Coefficient	Question
r_{XY}	How are X and Y linearly associated before adjustment?
$r_{XY \cdot Z}$	How are the residual parts of X and Y linearly associated after their separate relationships with Z are represented?

An adjusted coefficient may become smaller, stay similar, change sign, or become larger. A smaller value suggests that Z represented part of the raw pattern. A larger value can reflect suppression, where adjustment reveals a relationship that the raw coefficient partly hid. None of these changes proves the causal role of Z.

Checks and Limits

- Use the same complete set of cases for every correlation entering the calculation.
- Inspect X with Z, Y with Z, and the residual-residual plot. Linear adjustment cannot remove an unmodeled curved relationship.
- State precisely which variable was adjusted for and why it belongs in the question.
- Adjustment does not establish time order, remove unmeasured variables, correct measurement error, or reproduce random assignment.
- Partial correlation is symmetric: exchanging X and Y leaves its value unchanged.

Topic 7 keeps one named outcome and places several predictors into one regression equation. Its conditional coefficients carry the same central idea of adjustment, but express changes in outcome units and can include categorical predictors and interactions.

Expanded Reference

Purpose and foundations

Partial correlation describes the linear association between two variables after removing the fitted linear association that each has with one measured third variable. With focal variables X and Y and a control variable Z, the first-order partial correlation is written $r_{XY \cdot Z}$. Read the dot as "controlling for." The coefficient asks whether cases that are higher than expected on X, given Z, also tend to be higher than expected on Y, given the same Z.

The phrase **holding Z constant** describes a model-based comparison. It does not mean that the observed cases literally share one identical Z value. Linear regression is used to predict X from Z and Y from Z. Each case's residual records how far its observed value lies above or below that prediction. Partial correlation is the ordinary Pearson correlation between those two sets of residuals.

This method answers a conditional association question. It can show that a raw correlation shrinks, grows, or changes sign after adjustment. It cannot decide by itself whether Z is a confounder, mediator, collider, or suitable control. Those roles concern the data-generating process and require subject-matter reasoning, time order, and research design.

Quantity	What it correlates	Question answered
Raw correlation r_{XY}	Observed X with observed Y	How are the two measured variables linearly associated?
Partial correlation $r_{XY \cdot Z}$	Residualized X with residualized Y	How are their remaining linear components associated after adjustment for Z?

Core ideas

Residualization happens in three transparent steps. First, regress X on Z and save each residual e_{Xi}. Second, regress Y on Z and save each residual e_{Yi}. Third, calculate Pearson correlation between e_X and e_Y. Values above zero on a residualized variable mean "higher than the linear prediction based on Z," and values below zero mean "lower than predicted."

Standardizing the original variables does not perform this adjustment. Standardization subtracts a mean and divides by a standard deviation. It aligns units and reference points while preserving Pearson correlation. Residualization removes the fitted linear component associated with the control variable. The distinction matters because two plots can have matching standardized axes yet different correlations after residualization.

Observed coefficient change	Possible reading	What must still be checked
Partial coefficient is smaller	Some raw overlap was shared with Z	Whether Z is a defensible control and models are adequate
Partial coefficient is similar	Linear adjustment for Z changed little	Nonlinear effects, measurement, range, and sampling uncertainty
Partial coefficient is larger	Adjustment revealed previously obscured association	Whether suppression is substantively plausible rather than an accidental pattern

The adjustment is linear. If Z has a curved relationship with X or Y, a straight-line residualization can leave systematic structure behind. The method also inherits sensitivity to influential observations and range restriction from Pearson correlation and regression. Inspect the raw relations X with Y, X with Z, and Y with Z, then inspect the residualized relationship.

Partial correlation has a close conceptual connection with multiple regression. Both ask conditional linear questions after other measured information is considered. Their numerical coefficients have different scales: partial correlation is standardized to the interval from -1 to 1, while a later regression slope expresses a fitted outcome difference per predictor unit. Topic 7 develops that broader multiple-predictor framework.

Formula guide

Residualizing X against Z begins with a fitted line and keeps what the line does not explain:

$$e_{Xi} = x_i - (a_X + b_X z_i)$$

Do the corresponding operation for Y:

$$e_{Yi} = y_i - (a_Y + b_Y z_i)$$

The partial correlation is then the Pearson correlation of the two residual variables:

$$r_{XY \cdot Z} = r(e_X, e_Y)$$

When exactly one variable Z is controlled, the coefficient can also be calculated from the three pairwise correlations:

$$r_{XY \cdot Z} = \frac{r_{XY} - r_{XZ} r_{YZ}}{\sqrt{(1 - r_{XZ}^2)(1 - r_{YZ}^2)}}$$

This formula is compact, but the residual method is often easier to understand and diagnose because it shows the two adjustment models directly. The two routes agree when they use the same complete cases, ordinary linear regressions with intercepts, and the same three measured variables. That agreement checks the calculation; it does not establish that the control was causally appropriate.

The denominator also states when the direct formula is defined. If $|r_{XZ}| = 1$ or $|r_{YZ}| = 1$, one residualized focal variable has no remaining variation and the denominator is zero. A correlation cannot be calculated for a variable with zero spread. Near-perfect relations with Z can likewise make the adjusted result highly sensitive to small changes or rounding.

Kind of control	What changes	What conclusion it can support
Experimental control	The study design assigns or holds conditions before outcomes are observed	Can strengthen a causal comparison when the design and assumptions justify it
Statistical control	The analysis represents fitted linear relations with measured Z	Produces an adjusted association, not random assignment
Unmeasured third variable	Nothing in the calculation represents it	Its possible contribution remains unresolved

Residualization removes only the fitted linear component associated with the measured version of Z. It does not produce error-free variables, remove nonlinear relations automatically, or guarantee that Z was an appropriate control.

Reading the explanatory figure

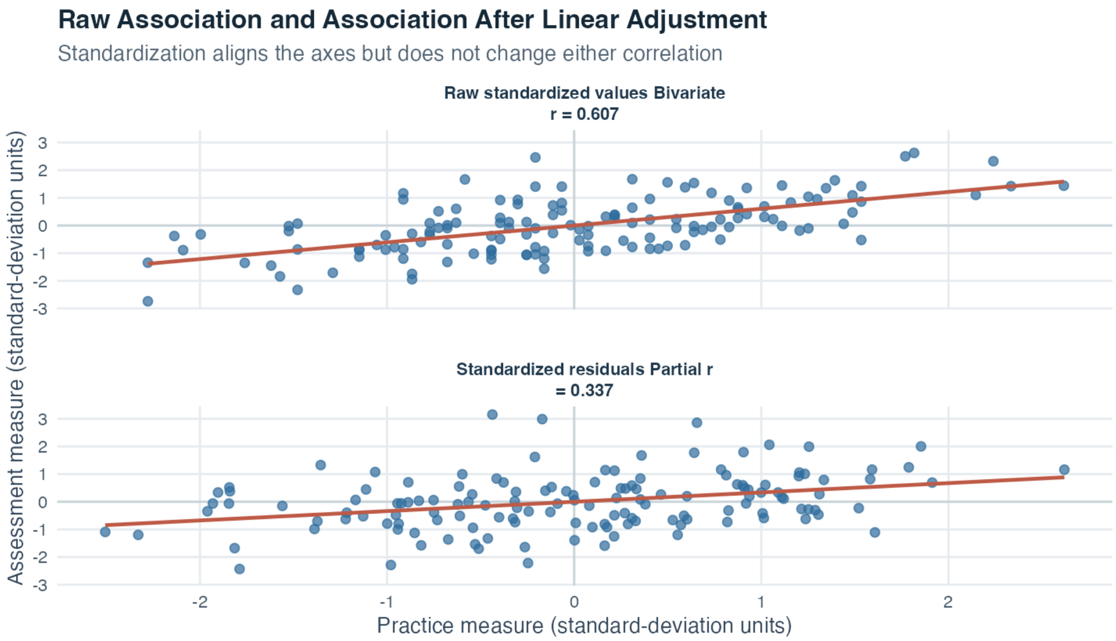

Figure 1: Two scatterplots compare the standardized raw association with the association between standardized residuals after linear adjustment for a third measure.

The left panel displays standardized practice and assessment measures. Standard-deviation units place zero at each variable's mean and make one unit equal to one sample standard deviation. The visible upward line and reported bivariate correlation of 0.607 describe a moderately positive raw linear association in these simulated values.

The right panel displays standardized residuals after both variables have been predicted from the control measure. Each horizontal coordinate now means how far a case's practice value lies above or below the value predicted from the control. Each vertical coordinate means the corresponding departure for assessment. The line still rises, but less steeply in standardized terms, and the partial correlation is 0.337. The adjustment therefore reduced the measured association while leaving a positive residual relationship.

The figure's subtitle emphasizes that standardization alone does not alter either correlation. The coefficient changes because the second panel uses residuals, not because its axes are standardized. The difference between 0.607 and 0.337 is a descriptive clue that some raw co-variation was aligned with the control variable. It is not a diagnosis of causation, and it should be interpreted with the three raw pairwise plots and the substantive role of the control.

Interpretation checklist

Name X, Y, and every control variable and explain why each control belongs in the analysis. Draw the assumed temporal or causal ordering in words before adjusting. Inspect all raw pairwise plots, missing-data patterns, range, and influential observations. Check that the adjustment relations are reasonably linear. Report the raw and partial correlations together so readers can see what changed.

Describe the coefficient as an association after linear adjustment for the named third variable. Report sample size, the three bivariate correlations, and the resulting partial correlation so the calculation can be followed. Do not interpret coefficient shrinkage as automatic proof of confounding, growth as automatic proof of suppression, or a near-zero value as proof of no relationship. Consider measurement quality and sampling uncertainty throughout.

How this topic connects

Covariance and Pearson correlation introduced paired linear co-variation. Simple regression then split each outcome into a fitted value and residual. Partial correlation combines those ideas: it uses two regressions to remove the components linearly associated with Z, then correlates the remaining components.

Multiple regression is the natural next step. It estimates the conditional contribution of several predictors in a single outcome model and provides slopes in their original units. The partial-correlation view remains useful because it explains what “holding other predictors constant” means: compare cases through the portions of a predictor and outcome that remain after linear adjustment. This connection turns a technical phrase into a concrete residual comparison.

Document ID: `topic-06-partial-correlation-summary-en`

Notes 1 of 3

Partial Correlation | Topic 6

Notes 2 of 3

Partial Correlation | Topic 6

Notes 3 of 3

Partial Correlation | Topic 6