*Tutorials and worked examples for simulation, curve fitting, statistical analysis, and plotting.*
*http://www.simfit.org.uk*

Partial correlation analysis is used to evaluate the extent to which the correlations between two or more columns (called $Y$-variables) of a $n$ by $m$ data matrix with $m > 2$ depend on correlations between these columns and other columns in the matrix (called $X$-variables). Either a data set or a correlation matrix together with sample size can be input, and it is most often used to study the way that the correlations between two columns depend on a third column.

## Example 1

From the main SIMFIT menu select [Statistics], [Multivariate], [Partial correlation] and then read in the test file `g02byf.tf1` provided. In the special case when $n = m$ you have to specify whether a data file or correlation matrix is being input, but this is a data matrix with fifteen rows and three columns as follows.

Column 1: number of deaths
Column 2: smoke($mg/m^3$)
Column 3: sulphur dioxide(parts/million)

| | | |
|---|---|---|
| 112 | 0.30 | 0.09 |
| 140 | 0.49 | 0.16 |
| 143 | 0.61 | 0.22 |
| 120 | 0.49 | 0.14 |
| 196 | 2.64 | 0.75 |
| 294 | 3.45 | 0.86 |
| 513 | 4.46 | 1.34 |
| 518 | 4.46 | 1.34 |
| 430 | 1.22 | 0.47 |
| 274 | 1.22 | 0.47 |
| 255 | 0.32 | 0.22 |
| 236 | 0.29 | 0.23 |
| 256 | 0.50 | 0.26 |
| 222 | 0.32 | 0.16 |
| 213 | 0.32 | 0.16 |

However the following important trailer section has been added to the data.

```
begin{indicators}
-1   -1   1
end{indicators}
```

Negative indicator values denote $Y$-variables, zero values indicate suppression, while positive indictor values identify $X$ variables. In other words, the default partial correlation between deaths and smoke is required when sulphur dioxide is considered as fixed. However, it should be noted that the assigning of columns to $Y$ or $X$ groups can also be done interactively.

First the overall Pearson product-moment correlation matrix is calculated and displayed along with the two-tail $p$-values.

Pearson product moment correlation results:
Strict upper triangle: $r$
Strict lower triangle: corresponding two-tail $p$ values

| | | |
|---|---|---|
| ..... | 0.7560 | 0.8309 |
| 0.0011 | ..... | 0.9876 |
| 0.0001 | 0.0000 | ..... |

This is then followed by a likelihood ratio test

Test for absence of any significant correlations
$H_0$: correlation matrix is the identity matrix

| | | |
|---|---|---|
| Determinant | 0.003484 | |
| Test statistic ($TS$) | 68.86 | |
| Degrees of freedom | 3 | |
| $P(\chi^2 \geq TS)$ | 0.0000 | *Reject $H_0$ at 1% significance level* |

but, in addition, the partial correlation matrix is displayed as in the next table for variables indicated as $YYX$. That is, correlation for columns 1 and 2, regarding column 3 as fixed.

Partial correlation results for variables: $YYX$
Strict upper triangle: partial $r$
Strict lower triangle: corresponding 2-tail $p$ values

| | |
|---|---|
| ... | -0.7381 |
| 0.0026 | ... |

## Example 2

This is the test file `pacorr.tf1` which contains a correlation matrix.

Correlation matrix: sample size = 30

| | | |
|---|---|---|
| 3 | 3 | |
| 1.0000 | 0.6162 | 0.8267 |
| 0.6162 | 1.0000 | 0.7321 |
| 0.8267 | 0.7321 | 1.0000 |
| 3 | | |

variable 1: Intelligence
variable 2: Weight
variable 3: Age

By systematically altering the definition for $Y$ variables and $X$ variables SIMFIT can calculate all the correlations and partial correlations as follows.

$r(1,2) = 0.6162$
$r(1,3) = 0.8267$
$r(2,3) = 0.7321$
...
$r(1,2|3) = 0.0286$ (95% confidence limits $= -0.3422, 0.3918$)
$t = 0.1488, ndof = 27, p = 0.8828$
...
$r(1,3|2) = 0.7001$ (95% confidence limits $= 0.4479, 0.8490$)
$t = 5.094, ndof = 27, p = 0.0000$ *Reject $H_0$ at 1% significance level*
...
$r(2,3|1) = 0.5025$ (95% confidence limits $= 0.1659, 0.7343$)
$t = 3.020, ndof = 27, p = 0.0055$ *Reject $H_0$ at 1% significance level*

From this table it is clear that when variable 3 is regarded as fixed, the correlation between variables 1 and 2 is not significant but, when either variable 1 or variable 2 are regarded as fixed, there is evidence for significant correlation between the other variables. Exactly what commonsense would predict.

## Theory

Assuming a multivariate normal distribution and linear correlations, the partial correlations between any two variables from the set $i$, $j$, $k$ conditional upon the third can be calculated using the usual correlation coefficients as

$$r_{i,j|k} = \frac{r_{ij} - r_{ik}r_{jk}}{\sqrt{(1 - r_{ik}^2)(1 - r_{jk}^2)}}.$$

If there are $p$ variables in all but $p - q$ are fixed then the sample size $n$ can be replaced by $n - (p - q)$ in the usual significance tests and estimation of confidence limits, e.g. $n - (p - q) - 2$ for a $t$ test.

The situation is more involved when there are more than three variables, say $n_x$ $X$ variables which can be regarded as fixed, and the remaining $n_y$ $Y$ variables for which partial correlations are required conditional on the fixed variables.

Then the variance-covariance matrix $\Sigma$ can be partitioned as in

$$\Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$$

when the variance-covariance of $Y$ conditional upon $X$ is given by

$$\Sigma_{y|x} = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy},$$

while the partial correlation matrix $R$ is calculated by normalizing as

$$R = \mathrm{diag}(\Sigma_{y|x})^{-\frac{1}{2}}\,\Sigma_{y|x}\,\mathrm{diag}(\Sigma_{y|x})^{-\frac{1}{2}}.$$

Exactly as for the full correlation matrix, the strict upper triangle of the output from the partial correlation analysis contains the partial correlation coefficients $r_{ij}$, while the strict lower triangle holds the corresponding two tail probabilities $p_{ij}$ where

$$p_{ij} = P\left(t_{n-n_x-2} \le -|r_{ij}|\sqrt{\frac{n - n_x - 2}{1 - r_{ij}^2}}\right) + P\left(t_{n-n_x-2} \ge |r_{ij}|\sqrt{\frac{n - n_x - 2}{1 - r_{ij}^2}}\right).$$

However, for convenience, the output table may display the subscripted partial correlation coefficients with indicated conditional variables together with confidence limits as in Example 2.