

1 Subject

Computation of Partial Correlations with Tanagra. Partial correlation measures the degree of association between two random variables, with the effect of a set of controlling variables removed (http://en.wikipedia.org/wiki/Partial_correlation).

In this tutorial, we show how to use the PARTIAL CORREALTION component of Tanagra. We duplicate the example described on a website. Thus, in addition to the presentation of the theoretical method, we can trace the detail of all the calculations that we will achieve¹.

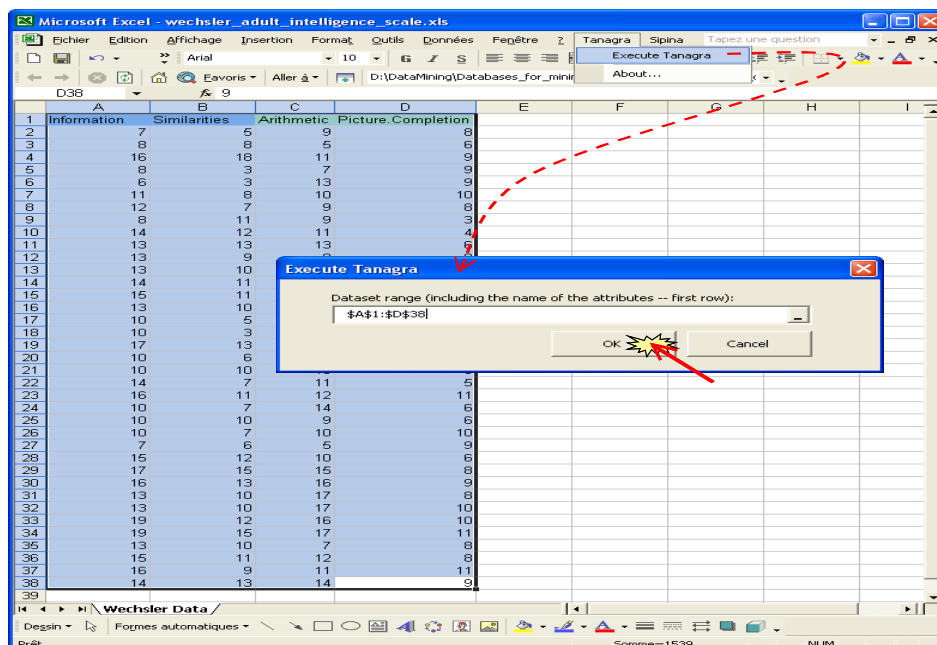
2 Dataset

The dataset comes from the WAIS general test of intelligence (Wechsler Adult Intelligence Scale -- http://en.wikipedia.org/wiki/Wechsler_Adult_Intelligence_Scale). We have 37 instances described with 4 dimensions: INFORMATION, SIMILARITIES, ARITHMETIC and PICTURE.COMPLETION. We intended to characterize the association between INFORMATION and SIMILARITIES by controlling the effect of ARITHMETIC and PICTURE.COMPLETION.

3 Computing the partial correlation

3.1 Dataset importation

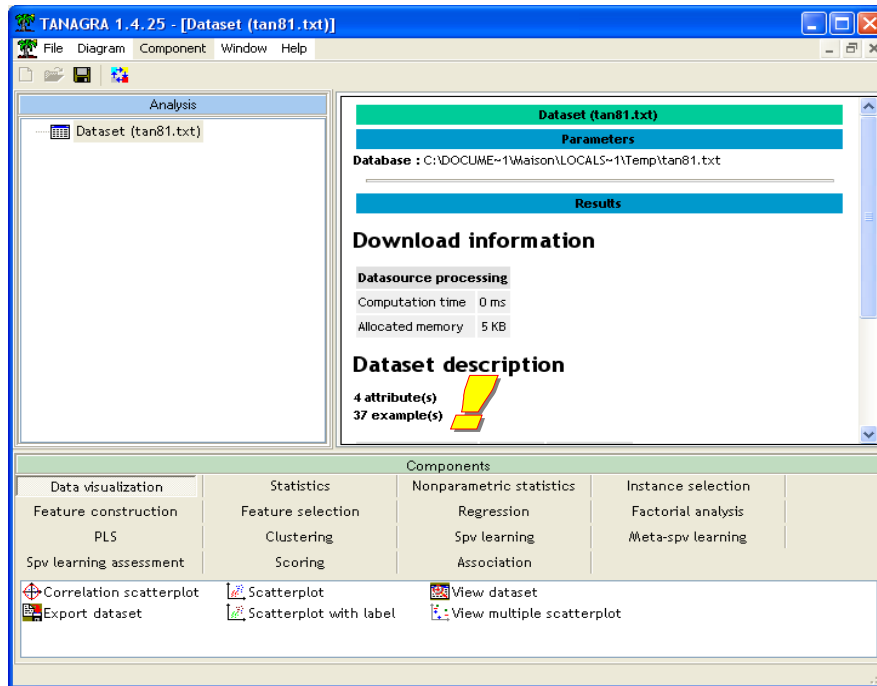
The simplest way in order to create a diagram is to load the dataset in the EXCEL spreadsheet (http://eric.univ-lyon2.fr/~ricco/tanagra/fichiers/wechsler_adult_intelligence_scale.xls). We select the data range and we click on the menu TANAGRA/EXECUTE TANAGRA².



¹ See http://www.stat.psu.edu/online/development/stat505/07_partcor/01_partcor_intro.html and http://www.stat.psu.edu/online/development/stat505/07_partcor/06_partcor_partial.html

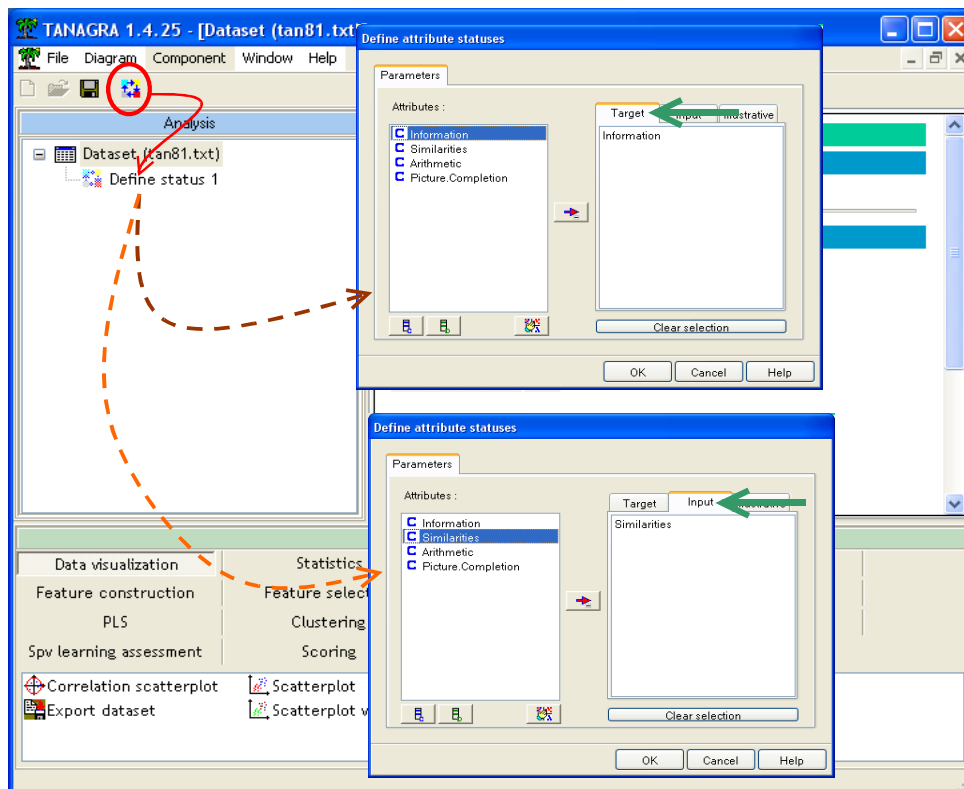
² The EXCEL add-in TANAGRA.XLA is available since the version 1.4.11. See the tutorial on the web site for the installation and the utilization of this add-in in your spreadsheet.

After checking the range selection, we click on OK. Tanagra is automatically launched and the dataset transferred. We have 37 individuals and 4 variables.

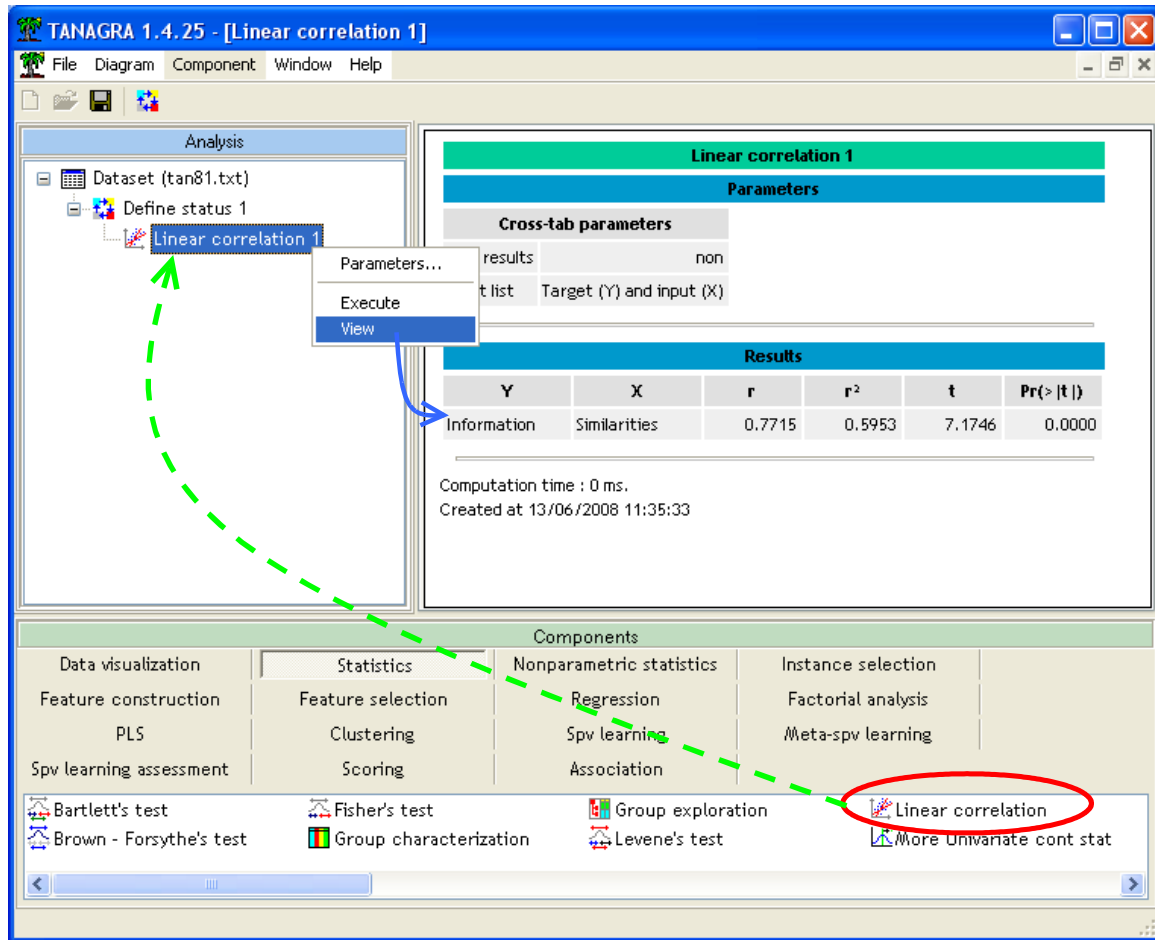


3.2 Zero order correlation

In a first step, we want to measure the correlation between SIMILARITIES and INFORMATION. We insert the DEFINES STATUS in the diagram. We use the shortcut in the toolbar. We set INFORMATION as TARGET, SIMILARITIES as INPUT. Because the analysis is symmetrical, we could have done the opposite.



We add the LINEAR CORRELATION COMPONENT (STATISTICS tab) into the diagram. We click on the VIEW menu in order to obtain the results.



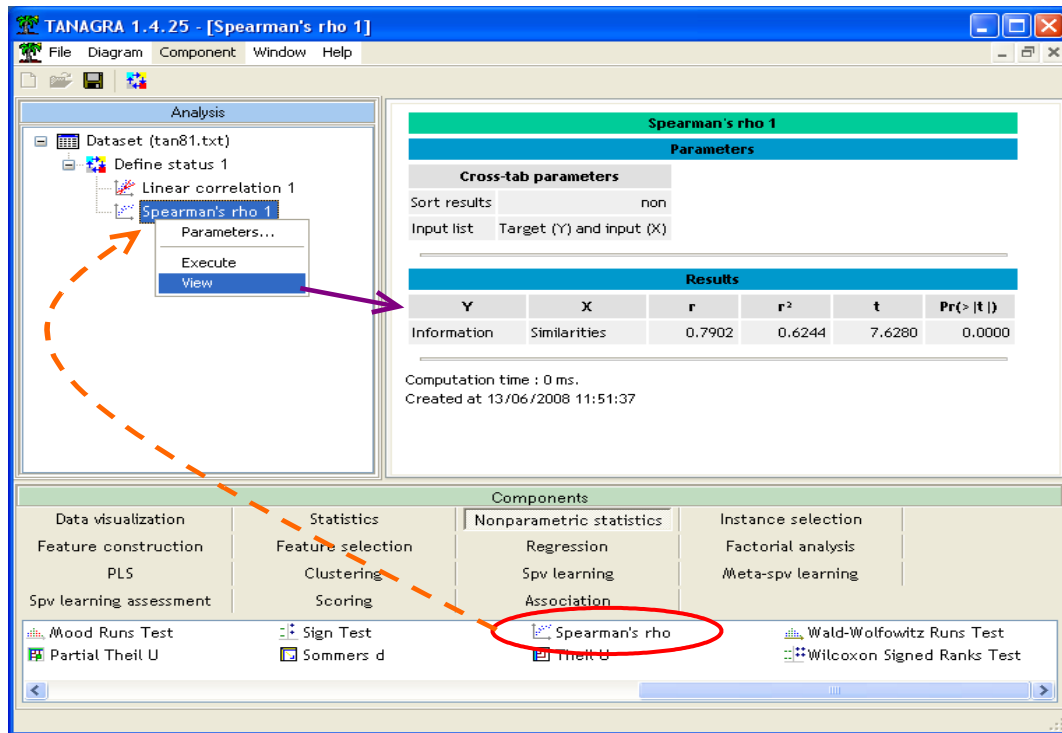
The zero order correlation between INFORMATION and SIMILARITIES is 0.7715. The t-statistic for the significance testing is 7.1746. Under the null hypothesis assumption (the correlation is 0), its sampling distribution is the t distribution with $(n - 2 = 37 - 2 = 35)$ degrees of freedom. The software computes directly the p-value of the test. We compare this with your significance level. In our example, the p-value is lower than 0.0001. It seems there is a significant correlation between INFORMATION and SIMILARITIES. The individuals with high INFROMATION scores have also high SIMILARITIES scores.

Note: The details of the computation are available on line. http://www.stat.psu.edu/online/development/stat505/06_propmean/03_propmean_infercorr.html.

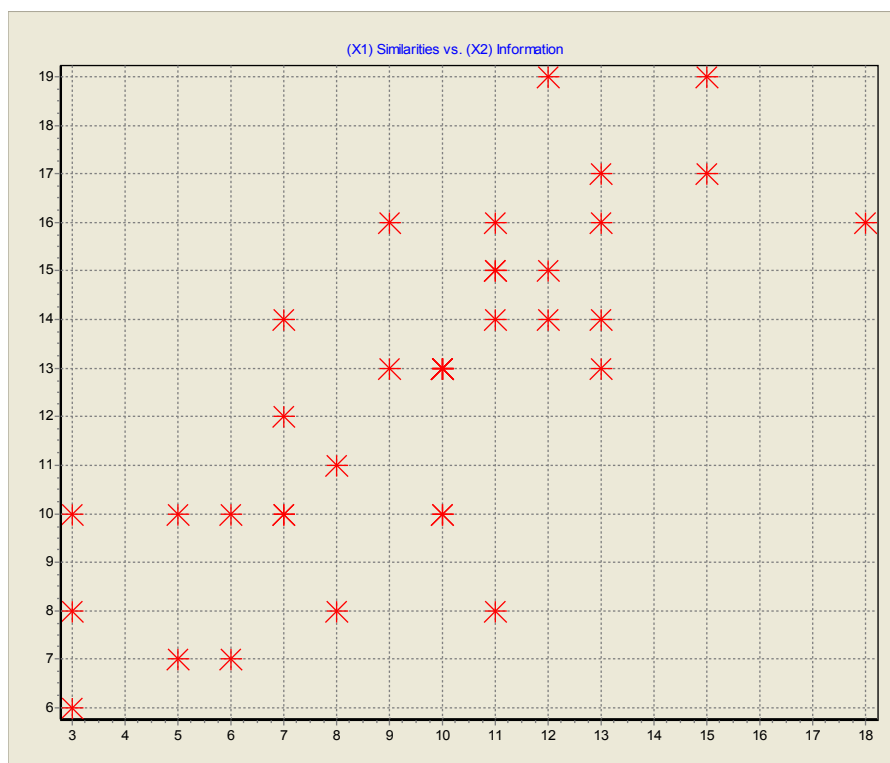
3.3 Rank correlation: Spearman’s correlation

The Pearson’s correlation coefficient characterizes only linear association. If we have doubt about this assumption, we can use the Spearman correlation. Basically, it turns the dataset in ranks, and the compute the correlation coefficient on these new variables. This new indicator can handle non linear (but monotonic) association. Furthermore, it is not disturbed by outliers.

We insert the SPEARMAN’S RHO component (NONPARAMETRIC STATISTICS tab) into the diagram. We click on the VIEW menu, we obtain the following results.



The rho value ($\rho = 0.7902$) is similar to the r value. It is also highly significant. It seems that the association is rather linear. If we build a graphical display of the data, a scatter plot, we note that the linear association hypothesis is credible³.



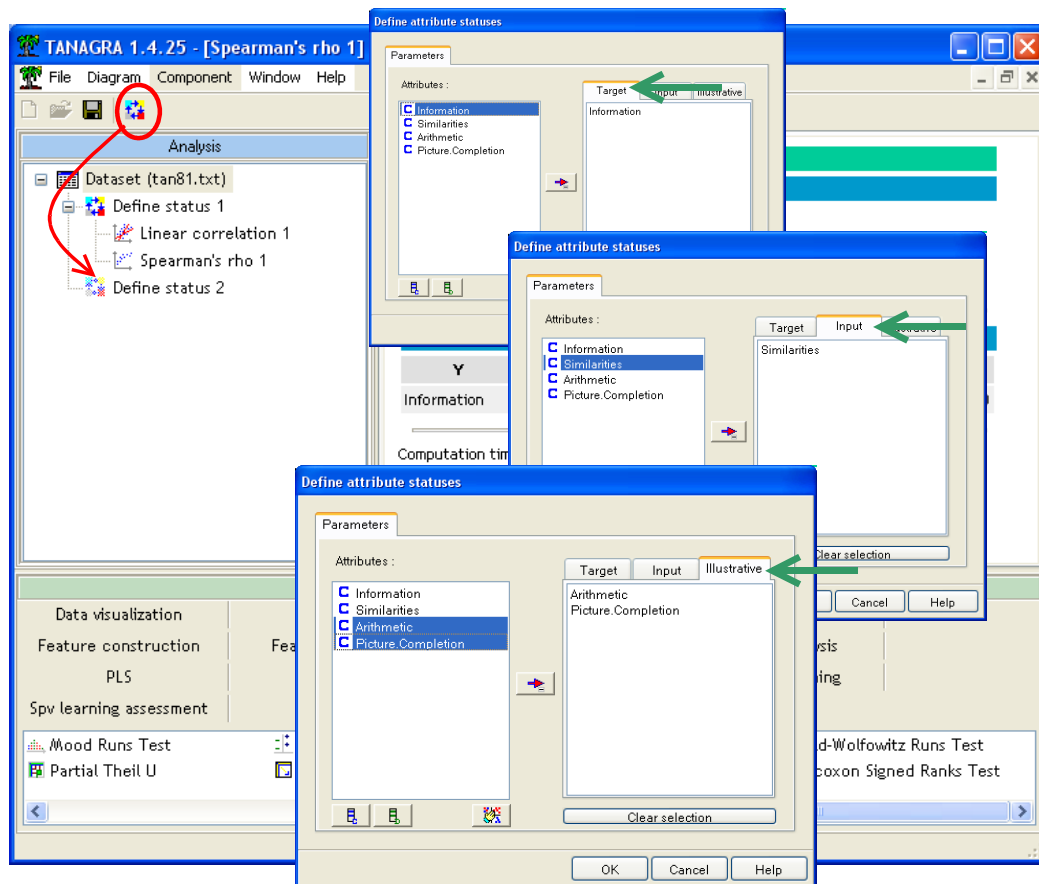
³ See <http://www.tufts.edu/~gdallal/corr.htm> about the relation between the graphical display of the data and the correlation coefficient.

3.4 Partial correlation

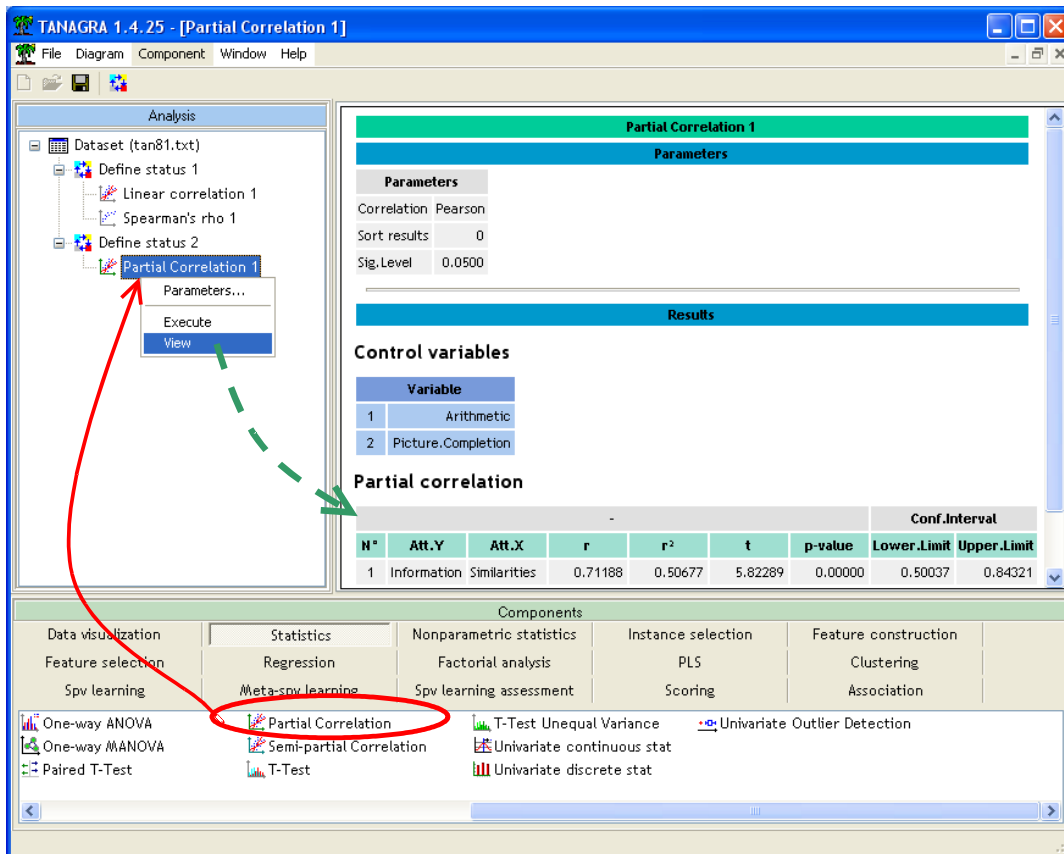
We want now to measure the association between INFORMATION and SIMILARITIES by controlling the (eventually) effect of ARITHMETIC and PICTURE.COMPLETION.

We insert again the DEFINE STATUS component into the diagram. We set INFORMATION as TARGET, SIMILARITIES as INPUT, **ARITHMETIC and PICTURE.COMPLETION as ILLUSTRATIVE**.

Note: The ILLUSTRATIVE term is not really appropriate in our study. But the DEFINE STATUS component is generic, here, we point that a third kind of variables are used in the analysis.



We add the PARTIAL CORRELATION (STATISTICS tab) into the diagram. We click on the VIEW menu in order to obtain the results.



The partial correlation is 0.71188. It is also highly significant. We note that the degrees of freedom is now $(37 - 2 - 2 = 33)$.

By using the Fisher transformation, we can compute the confidence interval of the correlation. We obtain for the $[0.50037; 0.84321]$ interval for the 95% confidence level.

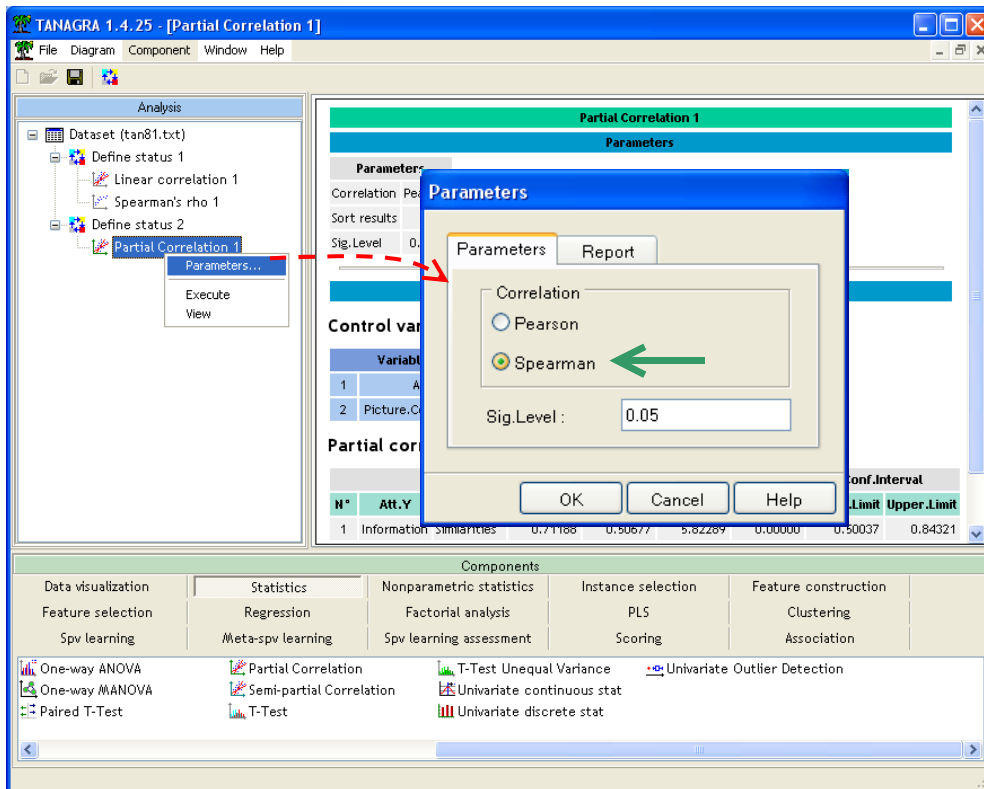
The zero order and the partial correlations are very similar in our example. It seems that the association between INFORMATION and SIMILARITIES cannot be explained by the control variables ARITHMETIC and PICTURE.COMPLETION.

Note: See http://www.stat.psu.edu/online/development/stat505/07_partcor/06_partcor_partial.html for the details of computation.

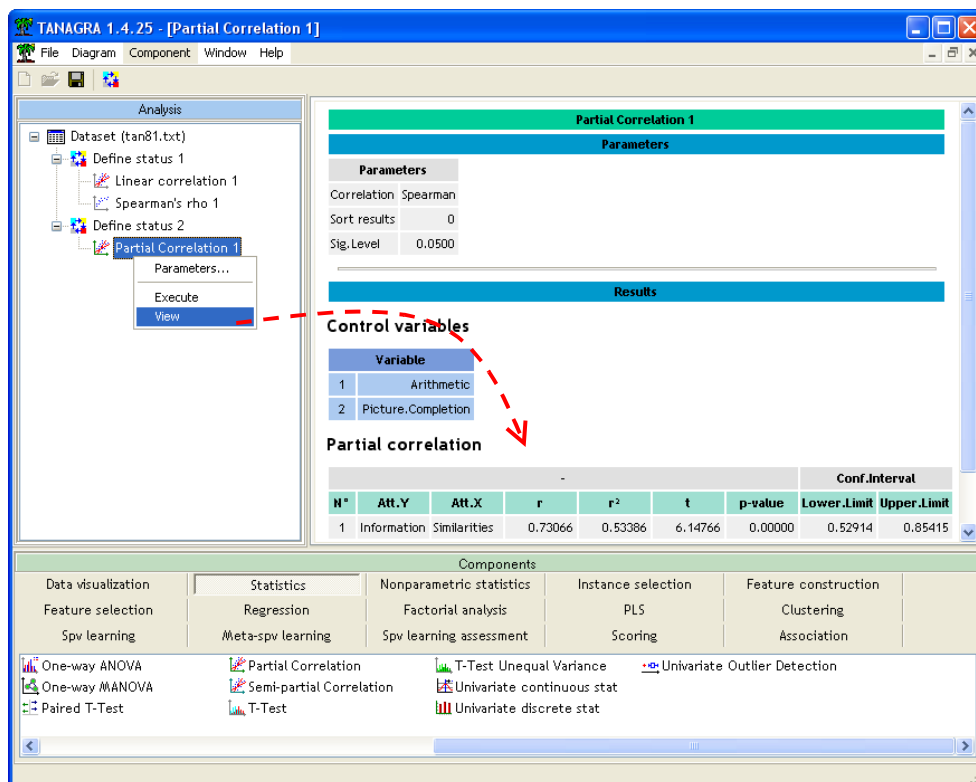
3.5 Spearman's partial correlation

As previous, we can calculate the partial rank correlation. The computation is exactly the same one, but we work on the ranks⁴. With TANAGRA, we modify the parameters of the component. We click on the PARAMETERS menu, we select the option SPEARMAN and then we click on OK. We activate the VIEW menu to get the results.

⁴ http://support.sas.com/documentation/cdl/en/procstat/59629/HTML/default/procstat_corr_sect017.htm



We obtain very similar results to the Pearson's partial correlation. The partial Spearman's rho is 0.73066.



4 Conclusion

The analysis of partial correlations is complicated. The numerical approaches are used mainly to validate intuitions. It is obvious that the search on the causalities must rely on domain knowledge.

An excellent website reference details the various kind of interaction between variables when we compute the partial correlations. See <http://www2.chass.ncsu.edu/garson/PA765/partialr.htm>

