INTERCITERIA AND CORRELATION ANALYSES:
SIMILARITIES, DIFFERENCES AND SIMULTANEOUS USE

Pencho G. Marinov and Stefka Fidanova

Department of Parallel Algorithms
Institute of Information and Communication Technologies
Bulgarian Academy of Sciences
Acad. G. Bonchev Str., Bl. 25-A, 1113 Sofia, Bulgaria,
e-mails: pencho@parallel.bas.bg, stefka@parallel.bas.bg

Abstract: Short remarks on intercriteria and correlation analyses are given. An example that their results do not coincide, is discussed. Results with simultaneous use and reuse of InterCriteria Analysis and a few types Correlation Analyses are presented.

Keywords: Intercriteria analysis, Intuitionistic fuzziness; Correlation analysis – Pearson, Spearman, Kendall; Confidence Interval for Correlation

1 Introduction

In this work we compare generally accepted Pearson, Spearman rank and Kendall rank correlations with ICA (InterCriteria Analysis) and some numerical experiments are provided for discussion. For more detailed considerations, some other statistical indicators are also calculated: average, confidence intervals, etc.

\( CL \) – 95% Confidence Interval for correlation, Left and Right bounds of \( CI \).

\( P_v \) – P-value indicates the risk of concluding that a correlation exists – when actually, no correlation exists – is 5%.

\( \text{Corr}_X \) – \( X \)-correlation, \( X \in \{ P, S, K, K0 \}; \)

\( P \) – for Pearson correlation,

\( S \) – for Spearman rank correlation,

\( K \) – for Kendall rank correlation, taking into account the emergence of ties.

\( K0 \) – for Kendall rank correlation w/o taking into account the emergence of ties
$Av_n$ – Average (mean) of Objects for $n$-Criteria ($n \in \{k, l\}$);  
$k, l, Np$ – $k, l$ - Criteria serial numbers and $Np$ - Number of this pair $(k, l)$;  
m$(k, l)$ – $\mu_{k,l}$ - is a measure of concordant for ICA  
n$(k, l)$ – $\nu_{k,l}$ - is a measure of discordant (negative concordant) for ICA  
p$(k, l)$ – $\pi_{k,l} \; -$ is a measure of uncertainty for ICA

2 Main results

2.1 Input/Output text files

Input file has header with info for number of criteria $nC$ – first line, number of column for each of these criteria – second line; number of rows $NnR$, number of columns $NnC$ and ’XXXXX’-name of experiment in 3-th line. Next $NnR$ rows are with $NnC$ comma separated values (CSV-format) are INPut data. This ’XXXXX’-name is important for next reuse of results or for comparisons of experiments.  
For Input data with $nC$ criteria the Output file has 16 columns and $\frac{nC(nC-1)}{2}$ rows. One column more has info about data file XXXXX-corr. The description of these columns is:

1. $k$ – the number of first criterion in the pair $(k, l)$,  
2. $l$ – the number of second criterion in the pair $(k, l)$,  
3. $Np$ – the number in order of this pair $(k, l)$ (from 1 to $\frac{nC(nC-1)}{2}$)  
4. $\mu(k, l)$ – from ICA  
5. $\nu(k, l)$ – from ICA  
6. $\pi(k, l)$ – from ICA  
7. $Av_k$ – Average (mean) for k-th criterion  
8. $Av_l$ – Average (mean) for l-th criterion  
9. $Pv_P$ – $P$-value for Pearson Correlation  
10. $Pv_S$ – $P$-value for Spearman Correlation  
11. $CI_{Left}$ – Confidence Interval for $P$-corr, Left boundary  
12. $CI_{right}$ – Confidence Interval for $P$-corr, Right boundary  
13. $corr$ – $P$ – Pearson Correlation  
14. $corr$ – $S$ – Spearman rank Correlation  
15. $corr$ – $K$ – Kendall rank Correlation taking into account the emergence of ties  
16. $corr$ – $K0$ – Kendall rank Correlation w/o taking into account ...
**Fig. 1-a:** Visualization for a part of Input data (smooth case).

**Fig. 1-b:** Visualization for zoomed part of Input data (noised case).

**Fig. 2-a:** Visualization for a part of Output data from Table 2-a.

**Fig. 2-b:** Visualization for a part of Output data from Table 2-b.
Table 1: Input data for example art0b. (13 columns, 101 rows)
Table 2-a: Output data for example art0a. (78 rows = 13*12/2)
Table 2-b: Output data for example art0b
3 Reuse OUTput files as INPut files

Now we reuse the output data (results) as new input. In two versions - alone and in combination of two such sets.

First variant. Let we look at just seven indicators (from all 16): $\mu, \nu, \pi, \text{corr}\_P, \text{corr}\_S, \text{corr}\_K, \text{corr}\_K0$ in output files (like these from Table 2-a and 2-b) as at 7 criteria, we can use 78 rows (pairs criteria) as objects, and reuse calculations as described above. We’ll get a new table with 16 columns and 21 rows ($21 = \frac{7\times16}{2}$). (see Tables 3-a and 3-b). Before that let see the description of these 21 rows (see Table 3-o).

<table>
<thead>
<tr>
<th>XX</th>
<th>col-1</th>
<th>col-2</th>
<th>XX</th>
<th>col-1</th>
<th>col-2</th>
<th>XX</th>
<th>col-1</th>
<th>col-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>$\mu$</td>
<td>$\nu$</td>
<td>08</td>
<td>$\nu$</td>
<td>Corr_P</td>
<td>15</td>
<td>$\pi$</td>
<td>Corr_K0</td>
</tr>
<tr>
<td>02</td>
<td>$\mu$</td>
<td>$\pi$</td>
<td>09</td>
<td>$\nu$</td>
<td>Corr_S</td>
<td>16</td>
<td>Corr_P</td>
<td>Corr_S</td>
</tr>
<tr>
<td>03</td>
<td>$\mu$</td>
<td>Corr_P</td>
<td>10</td>
<td>$\nu$</td>
<td>Corr_K</td>
<td>17</td>
<td>Corr_P</td>
<td>Corr_K</td>
</tr>
<tr>
<td>04</td>
<td>$\mu$</td>
<td>Corr_S</td>
<td>11</td>
<td>$\nu$</td>
<td>Corr_K0</td>
<td>18</td>
<td>Corr_P</td>
<td>Corr_K0</td>
</tr>
<tr>
<td>05</td>
<td>$\mu$</td>
<td>Corr_K</td>
<td>12</td>
<td>$\pi$</td>
<td>Corr_P</td>
<td>19</td>
<td>Corr_S</td>
<td>Corr_K</td>
</tr>
<tr>
<td>06</td>
<td>$\mu$</td>
<td>Corr_K0</td>
<td>13</td>
<td>$\pi$</td>
<td>Corr_S</td>
<td>20</td>
<td>Corr_S</td>
<td>Corr_K0</td>
</tr>
<tr>
<td>07</td>
<td>$\nu$</td>
<td>$\pi$</td>
<td>14</td>
<td>$\pi$</td>
<td>Corr_K</td>
<td>21</td>
<td>Corr_K</td>
<td>Corr_K0</td>
</tr>
</tbody>
</table>

Table 3-o: Description of rows in Tables 3-a and 3-b

Second variant. Let we merge two output files (like these from Tables 2-a and 2-b) but only these 7 columns as above. The new INPut files are with 14 columns and 78 rows. All pairs with criteria are $91 = 14 \times 13/2$. Interesting for us are the same 7 indicators $\mu, \nu, \pi, \text{corr}\_P, \text{corr}\_S, \text{corr}\_K, \text{corr}\_K0$ and these pairs columns (1,8), (2,9), (3,10), ..., (7,14) which are numbered in sequence 7, 20, 32, 43, 53, 62, 70 (see column 2 in Table 4).
Table 3-b: Output data after reuse Table 2-b. (experiment art0b_2)

<table>
<thead>
<tr>
<th>xx</th>
<th>N</th>
<th>mu</th>
<th>nu</th>
<th>pi</th>
<th>Cor-P</th>
<th>Cor-S</th>
<th>Cor-K</th>
<th>Cor-K0</th>
<th>AAAA-cor2</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu</td>
<td>7</td>
<td>0.824</td>
<td>0.090</td>
<td>0.086</td>
<td>0.914</td>
<td>0.893</td>
<td>0.770</td>
<td>0.734</td>
<td>0.900</td>
</tr>
<tr>
<td>nu</td>
<td>20</td>
<td>0.842</td>
<td>0.084</td>
<td>0.074</td>
<td>0.925</td>
<td>0.798</td>
<td>0.758</td>
<td>0.900</td>
<td>0.900</td>
</tr>
<tr>
<td>pi</td>
<td>32</td>
<td>0.092</td>
<td>0.007</td>
<td>0.902</td>
<td>0.392</td>
<td>0.261</td>
<td>0.263</td>
<td>0.085</td>
<td>0.860</td>
</tr>
<tr>
<td>P</td>
<td>43</td>
<td>0.943</td>
<td>0.000</td>
<td>0.057</td>
<td>1.000</td>
<td>0.998</td>
<td>0.994</td>
<td>0.943</td>
<td>0.900</td>
</tr>
<tr>
<td>S</td>
<td>53</td>
<td>0.902</td>
<td>0.049</td>
<td>0.049</td>
<td>0.998</td>
<td>0.863</td>
<td>0.853</td>
<td>0.900</td>
<td>0.900</td>
</tr>
<tr>
<td>K</td>
<td>62</td>
<td>0.896</td>
<td>0.061</td>
<td>0.043</td>
<td>0.986</td>
<td>0.959</td>
<td>0.856</td>
<td>0.836</td>
<td>0.900</td>
</tr>
<tr>
<td>K0</td>
<td>70</td>
<td>0.854</td>
<td>0.014</td>
<td>0.132</td>
<td>1.000</td>
<td>0.954</td>
<td>0.904</td>
<td>0.840</td>
<td>0.900</td>
</tr>
</tbody>
</table>

Table 4: Output data after merge Tables 2-a and 2-b. (experiments art0a, art0b)

4 Conclusion

About some cases of dependences: (see Tables 2-a, 2-b, 3, 3-a, 3-b and 4) We look at the seven indicators: \( \mu, \nu, \pi, \text{corr}_P, \text{corr}_S, \text{corr}_K, \text{corr}_K0 \).

- about \( \mu \) \_ICA and Correlations \( P, S, K, K0 \): (cases \( Np \in \{3, 4, 5, 6\} \) from Table 3-a and 3-b) \( \nu \) and \( \pi \) are relatively small (< 16%), the rest indicators are relatively high. Especially for \( \text{corr}_P \) the following exceptions are interesting: \( Np \in \{33, 49, 68\} \) from Tables 2-a and 2-b.

- about \( \nu \) \_ICA and Correlations \( P, S, K, K0 \): (cases \( Np \in \{8, 9, 10, 11\} \) from Table 3-a and 3-b) \( \mu \) and \( \pi \) are relatively small (< 16%), the rest indicators are relatively high. Especially for \( \text{corr}_P \) the following exceptions are interesting: \( Np \in \{42, 56, 72\} \) from Tables 2-a and 2-b.

- about \( \pi \) \_ICA and Correlations \( P, S, K, K0 \): (cases \( Np \in \{12, 13, 14, 15\} \) from Table 3-a and 3-b) All Correlations are small (< 10%), the uncertainty is very high (> 87%) for the case with silent data.
about $\mu$, $\nu$, $\pi$ from ICA: (cases $Np \in \{1, 2, 7\}$ from Table 3-a and 3-b) Between $\mu$ and $\nu$ anti-correlation for silent case is more pronounced than in smooth case of data. Between $(\mu, \pi)$ and $(\nu, \pi)$ uncertainty for silent case is more pronounced than in smooth case of data.

- about $corr_P$, $corr_S$: (case $Np = 16$ from Table 3-a and 3-b) $\nu$ and $\pi$ are small ($< 10\%$), the rest indicators are high and in most cases $|corr_P| > |corr_S|$. Interesting exceptions are $Np \in \{33, 42, 49, 56, 68, 72, 76\}$ from Tables 2-a and 2-b.

- about the type of data - smooth, silent: /see Table 4/ For $\pi$: indicator $\pi$ is very high ($> 90\%$) but all rest indicators are small. For all others indicators (the rest six) $\nu$ and $\rho$ are small ($< 10\%$) but the others are high.

Acknowledgements
The authors are thankful for the support provided by the Bulgarian National Science Fund under Grant Ref. No. DFNI-I-02-5 “InterCriteria Analysis: A New Approach to Decision Making”.

References