Distance correlation: Discovering meta-analytic relationships between variables when other correlation coefficients fail

Research Synthesis, Dubrovnik: Methods in meta-analysis (29.05.2019)

Lukasz Stasielowicz & Reinhard Suck

• Osnabrück University
• lukasz.stasielowicz@uos.de
Foosball (table soccer)
Correlations in meta-analyses

- Usual main goal of a meta-analysis: Computing the mean correlation across studies (i.e. \( r \))
  - Example: Is there some kind of dependence between personality constructs?

- Issue 1: Correlation ≠ causation

- Issue 2: \( r = 0 \) ≠ Lack of dependence
  - Crux of this presentation
  - \( r = 0 \) only means that there is no linear relationship
  - Risk of failing to identify nonlinear relationships, e.g. inverted-U
Nonlinear relationship: Example 1

- **Yerkes–Dodson law**
  - Relationship between arousal and performance
  - Nonlinear relationship (inverted-U relationship)

Diamond et al. (2007, p. 3)
Nonlinear relationship: Example 2

- Relationship between Age and cognitive abilities
  - Non-monotonic relationship
    - Increase + decrease of cognitive abilities

(Li et al., 2004, p. 158)
Nonlinear relationships in a meta-analysis

- $r$ can fail in meta-analyses when dealing with nonlinear relationships

- What about other well-known effect sizes?
  - Spearman’s rho, Kendall’s tau etc. cannot detect non-monotonic relationships

- Distance correlation ($\mathcal{R}$) as a potential solution (Rizzo & Székely, 2016)
  - Different types of dependence can be assessed simultaneously
  - $\mathcal{R}_{\text{Min}} = 0$, $\mathcal{R}_{\text{Max}} = 1$
    - 0 means that there is no dependence
Assessing nonlinear relationships

- Comparison of four different coefficients
  - Distance correlation, Pearson’s $\rho$, Kendall’s $\tau$, and Spearman’s $\rho$

**Linear relationship**
- Distance correlation = .560
- Pearson's $\rho$ = .601
- Kendall's $\tau$ = .415
- Spearman's $\rho$ = .591

**Inverted-U relationship**
- Distance correlation = .430
- Pearson's $\rho$ = -.014
- Kendall's $\tau$ = .003
- Spearman's $\rho$ = .005
Many applications of distance correlation

- Exploratory data analysis (Székely & Rizzo, 2009)
- Variable selection in regression models (Kong et al., 2015; Li et al., 2012; Yenigün & Rizzo, 2015)
- Principal component analysis (Mishra, 2014)
- Modelling autocorrelation in longitudinal studies (Edelmann et al., 2018; Zhou, 2012)
- Measuring dependence between networks in brain imaging studies (Chen et al., 2019)

Potentially relevant in the meta-analytic context (Székely et al., 2007)

- „Distance correlation can also be applied as an index of dependence; for example, in meta-analysis distance correlation would be a more generally applicable index than product-moment correlation” (p. 2770)
Goals of the present study

- Testing the feasibility of using distance correlation in a meta-analysis
- Comparing distance correlation to standard effect sizes

Computing distance correlation

- R package *energy*
- Conceptual similarity to Pearson correlation:
  \[
  \mathcal{R}(X, Y) = \frac{\nu(x,y)}{\sqrt{\nu(x) \cdot \nu(y)}}
  \]

  Distance Correlation = \[
  \frac{\text{Distance Covariance}}{\sqrt{\text{Distance Variance}_X \cdot \text{Distance Variance}_Y}}
  \]

- It is based on distances between individual values
  - i.e. X = cognitive abilities: Person 1 vs Person 2; Person 1 vs Person 3 etc.
  - i.e. Y = age: Person 1 vs Person 2; Person 1 vs Person 3 etc.
Computing distance correlation

- **Distances for the X variable:**

  \[ a_{km} = |X_k - X_m| \]

  \[ \bar{a}_{.k} = \frac{1}{n} \sum_{m=1}^{n} a_{km} \]

  \[ \bar{a}_{.m} = \frac{1}{n} \sum_{k=1}^{n} a_{km} \]

  \[ \bar{a}_{..} = \frac{1}{n^2} \sum_{k,m=1}^{n} a_{km} \]

  \[ A_{km} = a_{km} - \bar{a}_{.k} - \bar{a}_{.m} + \bar{a}_{..} \]

- **For the Y variable b values are computed**

\[ \psi^2 = \frac{1}{N} \sum_{i=1}^{n} \frac{\hat{b}_i^2}{\hat{n}_i} - 2 \frac{\hat{b}_i}{\hat{n}_i} \]

\[ = \frac{1}{N} \sum_{i=1}^{n} \hat{b}_i^2 - 2 \frac{1}{N} \sum_{i=1}^{n} \hat{b}_i \]

\[ = \frac{1}{N} \sum_{i=1}^{n} \hat{b}_i^2 - \frac{2}{N} \sum_{i=1}^{n} \hat{n}_i \]

\[ = \frac{1}{N} \sum_{i=1}^{n} \hat{b}_i^2 - \frac{2}{N} \sum_{i=1}^{n} \hat{n}_i \]

\[ D_{km} = A_{km} - \bar{a}_{.k} - \bar{a}_{.m} + \bar{a}_{..} \]

**Distance Variance**

\[ \nu^2_n(X) = \nu^2_n(X,X) = \frac{1}{n^2} \cdot \sum_{k,m=1}^{n} A_{km}^2 \]

**Distance Covariance**

\[ \nu^2_n(X, Y) = \frac{1}{n^2} \cdot \sum_{k,m=1}^{n} A_{km} B_{km} \]
Current study

- 36 scenarios (4 x 3 x 3)
  - 4 different kinds of dependence (see figure)
  - Number of samples in the meta-analysis (k: 20, 50, 100)
  - Size of each sample (N: 50, 200, 1000)

- For each sample the following effect sizes were computed: Kendall’s tau (τ), Spearman’s rho (ρ), Pearson correlation (r), distance correlation (R), and unbiased distance correlation (RU) were computed

- Next the mean effect sizes were computed (180 in total)
Current study

- R packages: energy, bootstrap, metafor

- Meta-analytic model: Random-effects model

- Heterogeneity estimator: Restricted maximum likelihood (REML)
  - Good performance in simulation studies
    (Langan et al., 2017; Veroniki et al., 2016)
Distance correlation in meta-analysis

- Usually effect sizes are weighted \((w_i)\) in a meta-analysis
  - They depend on the sampling variance \((v_i)\)
  - Small samples \(\rightarrow\) large variance \(\rightarrow\) small weight

- Sampling variance for distance correlation
  - Jackknife method has been recommended (Székely & Rizzo, 2009)
    - Leave-one-out procedure
    - Compute distance correlation after „deleting“ one pair of observation (i.e. data for one person)
    - Compare mean correlation across leave-one-out subsets to the correlation of each subset

Image by HOerwin56 from Pixabay
Results (pattern A)

- Data sets were simulated based on a true Pearson correlation ($r$) of .60

- $r$ performs best

- $\tau$ underestimates the dependence

- Spearman’s rho and distance correlations ($R$) perform similarly (slight underestimation)
  
  - Interestingly distance correlations perform worse in large samples

<table>
<thead>
<tr>
<th>$k$</th>
<th>$N$</th>
<th>$\tau$</th>
<th>$\rho$</th>
<th>$r$</th>
<th>$R$</th>
<th>$R_U$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>50</td>
<td>.397</td>
<td>.563</td>
<td>.592</td>
<td>.574</td>
<td>.539</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>.410</td>
<td>.579</td>
<td>.605</td>
<td>.561</td>
<td>.551</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>.408</td>
<td>.579</td>
<td>.596</td>
<td>.550</td>
<td>.548</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>.402</td>
<td>.565</td>
<td>.612</td>
<td>.588</td>
<td>.556</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>.411</td>
<td>.581</td>
<td>.606</td>
<td>.563</td>
<td>.553</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>.409</td>
<td>.581</td>
<td>.601</td>
<td>.552</td>
<td>.550</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>.408</td>
<td>.571</td>
<td>.621</td>
<td>.595</td>
<td>.563</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>.407</td>
<td>.577</td>
<td>.601</td>
<td>.560</td>
<td>.550</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>.412</td>
<td>.585</td>
<td>.604</td>
<td>.556</td>
<td>.554</td>
</tr>
</tbody>
</table>


Results (pattern B)

- $\tau$, $r$, and $\rho$ fail to identify an inverted-U relationship
  - Values close to 0

- Only distance correlations ($R$) yield large values

<table>
<thead>
<tr>
<th>$k$</th>
<th>$N$</th>
<th>$\tau$</th>
<th>$\rho$</th>
<th>$r$</th>
<th>$R$</th>
<th>$R_U$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>50</td>
<td>-.040</td>
<td>-.039</td>
<td>-.066</td>
<td>.586</td>
<td>.522</td>
</tr>
<tr>
<td>20</td>
<td>200</td>
<td>-.017</td>
<td>-.024</td>
<td>-.056</td>
<td>.550</td>
<td>.533</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>-.006</td>
<td>-.005</td>
<td>.019</td>
<td>.540</td>
<td>.537</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>-.048</td>
<td>-.056</td>
<td>-.079</td>
<td>.590</td>
<td>.524</td>
</tr>
<tr>
<td>50</td>
<td>200</td>
<td>-.011</td>
<td>-.015</td>
<td>-.014</td>
<td>.552</td>
<td>.535</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>.008</td>
<td>.010</td>
<td>.008</td>
<td>.540</td>
<td>.536</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>-.020</td>
<td>-.026</td>
<td>-.046</td>
<td>.584</td>
<td>.520</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>-.006</td>
<td>-.005</td>
<td>.021</td>
<td>.548</td>
<td>.531</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0</td>
<td>0</td>
<td>-.007</td>
<td>.541</td>
<td>.538</td>
</tr>
</tbody>
</table>
Results (pattern C)

- τ, r, and ρ fail to identify the non-monotonic relationship
  - Values close to 0

- Only distance correlations (ℛ) yield values greater than 0
  - Unbiased estimator yielded negative values for some small samples (N = 50)

<table>
<thead>
<tr>
<th>k</th>
<th>N</th>
<th>τ</th>
<th>ρ</th>
<th>r</th>
<th>ℛ</th>
<th>ℛ_u</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>50</td>
<td>.030</td>
<td>.053</td>
<td>.053</td>
<td>.248</td>
<td>.170^a</td>
</tr>
<tr>
<td>20</td>
<td>200</td>
<td>.005</td>
<td>.010</td>
<td>.004</td>
<td>.180</td>
<td>.147</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>-.004</td>
<td>-.006</td>
<td>-.005</td>
<td>.154</td>
<td>.146</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>.012</td>
<td>.018</td>
<td>.016</td>
<td>.242</td>
<td>.171^a</td>
</tr>
<tr>
<td>50</td>
<td>200</td>
<td>.007</td>
<td>.013</td>
<td>.010</td>
<td>.178</td>
<td>.145</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>.002</td>
<td>.003</td>
<td>.003</td>
<td>.152</td>
<td>.144</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>.011</td>
<td>.018</td>
<td>.016</td>
<td>.243</td>
<td>.172^a</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>-.004</td>
<td>-.006</td>
<td>-.005</td>
<td>.177</td>
<td>.143</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0</td>
<td>-.001</td>
<td>-.001</td>
<td>.151</td>
<td>.143</td>
</tr>
</tbody>
</table>
Results (pattern D)

- \( \tau, r, \) and \( \rho \) fail to identify the non-monotonic relationship
  - Values close to 0

- Only distance correlations (\( R \)) yield values greater than 0
  - Unbiased estimator yielded negative values for some small samples (\( N = 50 \))

\[
\begin{array}{cccccc}
 k & N & \tau & \rho & r & R & R_U \\
 50 & 50 & -.003 & -.005 & -.002 & .239 & .157^a \\
 20 & 200 & .006 & .012 & .013 & .185 & .157 \\
 & 1000 & -.002 & -.006 & -.007 & .174 & .168 \\
 50 & 50 & .002 & .008 & .017 & .241 & .161^a \\
 & 200 & -.005 & -.008 & -.008 & .189 & .163 \\
 & 1000 & -.001 & -.002 & -.003 & .172 & .166 \\
 50 & 50 & .004 & .007 & .009 & .241 & .168^a \\
 & 200 & -.001 & -.003 & -.003 & .190 & .163 \\
 & 1000 & 0 & -.002 & -.002 & .173 & .167 \\
\end{array}
\]
Summary

- Only distance correlation was able to identify dependence across all 36 scenarios
- Use of distance correlation in a meta-analysis can be fruitful

• Recommendation: Preliminary check
  - No dependence? Use r (software available: metafor etc.)
  - Dependence: Check scatter plots for each sample

• If the relationship is linear – use r
• Nonlinear or nonmonotonic relationships – use distance correlation
Issues

- **Interpretation**: Does a value of .01 imply dependence?
  - Statistical tests exist (Székely & Rizzo, 2009; Székely et al., 2007)
    - Pitfalls of $p$-value (Amrhein, Greenland, & McShane, 2019)

- **Unbiased estimator**: Problems in small samples (negative values)
  - Common when dealing with unbiased statistics, i.e. $adjR^2$ in multiple regression etc. (Rizzo & Székely, 2016; Székely & Rizzo, 2013)
  - How to deal with this issue in a meta-analysis?
    - Set negative values to zero?
      - Requires adjusting the jackknife technique – setting distance correlations to 0
    - Delete them from the meta-analysis?
Issues

- Full data sets needed to compute distance correlation
  - It cannot be derived from summary statistics ($M, SD, t, p$ etc.)
  - It cannot be derived from standard effect sizes ($r, d, OR$ etc.)
  - Open Science to the rescue!
    - Willingness to share data is increasing
    - Many platforms available (osf, PsychArchives etc.)
    - Multi-lab studies (replications)
    - Peer Reviewers' Openness Initiative
Issues

- The same distance correlation value can correspond to different patterns across samples (i.e. linear, quadratic)

- Dealing with heterogeneity
  - Common heterogeneity statistics ($I^2$, $Q$, $\tau$) may fail
    - Different patterns but the same distance correlation value
  - Failure of identifying moderators may lead to bad consequences, i.e.
    - Approval of interventions with side effects in certain groups
    - Rejection of promising interventions
  - Visual inspection of the data necessary
  - Changing the sign of distance correlation if plausible (i.e. U-relationship vs inverted-U relationship)
  - Subgroup analysis: Analyzing data sets with different patterns separately
Future research questions

- Conducting meta-analyses based on real data
- Benchmarks for interpreting $R$ values
- Applying distance correlation to three-level meta-analytic models
- Bayesian distance correlation
- Comparing distance correlation to other new dependence measures
  - Maximal Information Coefficient (MIC), Total Information Coefficient (TIC), Heller Heller Gorfine measure (HHG) or Hoeffding’s D (de Siqueira Santos et al., 2014; Kinney & Atwal, 2014; Reshef et al. 2018; Speed, 2011)
  - MIC and TIC seem to perform worse when dealing with linear patterns but are better when dealing with nonlinear patterns (Reshef et al., 2018).
References


References


References


References


Distance correlation in meta-analysis

Lukasz Stasielowicz & Reinhard Suck

Research Synthesis, Dubrovnik
Methods in meta-analysis (29.05.2019)
Appendix: Unbiased vs standard estimator

Unbiased estimator (Rizzo & Székely, 2016)

\[ \tilde{A}_{i,j} = \begin{cases} 
   a_{i,j} - \frac{1}{n-2} \sum_{i=1}^{n} a_{i,j} - \frac{1}{n-2} \sum_{j=1}^{n} a_{i,j} + \frac{1}{(n-1)(n-2)} \sum_{i,j=1}^{n} a_{i,j}, & i \neq j; \\
   0, & i = j. 
\end{cases} \] p. 33

Standard estimator

\[ A_{km} = a_{km} - \bar{a}_k - \bar{a}_m + \bar{a} . \]
## Appendix: All results

<table>
<thead>
<tr>
<th>k</th>
<th>N</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>τ</td>
<td>ρ</td>
<td>r</td>
<td>R</td>
</tr>
<tr>
<td>50</td>
<td>20</td>
<td>.397</td>
<td>.563</td>
<td>.592</td>
<td>.574</td>
</tr>
<tr>
<td>50</td>
<td>200</td>
<td>.410</td>
<td>.579</td>
<td>.605</td>
<td>.561</td>
</tr>
<tr>
<td>50</td>
<td>1000</td>
<td>.408</td>
<td>.579</td>
<td>.596</td>
<td>.550</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>.402</td>
<td>.565</td>
<td>.612</td>
<td>.588</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>.411</td>
<td>.581</td>
<td>.606</td>
<td>.563</td>
</tr>
<tr>
<td>200</td>
<td>1000</td>
<td>.409</td>
<td>.581</td>
<td>.601</td>
<td>.552</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>.408</td>
<td>.571</td>
<td>.621</td>
<td>.595</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>.407</td>
<td>.577</td>
<td>.601</td>
<td>.560</td>
</tr>
<tr>
<td>100</td>
<td>1000</td>
<td>.412</td>
<td>.585</td>
<td>.604</td>
<td>.556</td>
</tr>
<tr>
<td>τ</td>
<td>ρ</td>
<td>r</td>
<td>R</td>
<td>R_U</td>
<td></td>
</tr>
<tr>
<td>-.040</td>
<td>-.039</td>
<td>-.066</td>
<td>.586</td>
<td>.522</td>
<td></td>
</tr>
<tr>
<td>-.017</td>
<td>-.024</td>
<td>-.056</td>
<td>.550</td>
<td>.533</td>
<td></td>
</tr>
<tr>
<td>-.006</td>
<td>-.005</td>
<td>.019</td>
<td>.540</td>
<td>.537</td>
<td></td>
</tr>
<tr>
<td>-.048</td>
<td>-.056</td>
<td>-.079</td>
<td>.590</td>
<td>.524</td>
<td></td>
</tr>
<tr>
<td>-.011</td>
<td>-.015</td>
<td>-.014</td>
<td>.552</td>
<td>.535</td>
<td></td>
</tr>
<tr>
<td>.008</td>
<td>.010</td>
<td>.008</td>
<td>.540</td>
<td>.536</td>
<td></td>
</tr>
<tr>
<td>-.020</td>
<td>-.026</td>
<td>-.046</td>
<td>.584</td>
<td>.520</td>
<td></td>
</tr>
<tr>
<td>-.006</td>
<td>-.005</td>
<td>.021</td>
<td>.548</td>
<td>.531</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>-.007</td>
<td>.541</td>
<td>.538</td>
<td></td>
</tr>
<tr>
<td>τ</td>
<td>ρ</td>
<td>r</td>
<td>R</td>
<td>R_U</td>
<td></td>
</tr>
<tr>
<td>.030</td>
<td>.053</td>
<td>.047</td>
<td>.248</td>
<td>.170</td>
<td></td>
</tr>
<tr>
<td>.005</td>
<td>.010</td>
<td>.004</td>
<td>.180</td>
<td>.147</td>
<td></td>
</tr>
<tr>
<td>-.004</td>
<td>-.006</td>
<td>-.005</td>
<td>.154</td>
<td>.146</td>
<td></td>
</tr>
<tr>
<td>.012</td>
<td>.018</td>
<td>.016</td>
<td>.242</td>
<td>.171</td>
<td></td>
</tr>
<tr>
<td>.007</td>
<td>.013</td>
<td>.010</td>
<td>.178</td>
<td>.145</td>
<td></td>
</tr>
<tr>
<td>.002</td>
<td>.003</td>
<td>.003</td>
<td>.152</td>
<td>.144</td>
<td></td>
</tr>
<tr>
<td>.011</td>
<td>.018</td>
<td>.016</td>
<td>.243</td>
<td>.172</td>
<td></td>
</tr>
<tr>
<td>.004</td>
<td>.007</td>
<td>.009</td>
<td>.241</td>
<td>.168</td>
<td></td>
</tr>
<tr>
<td>.004</td>
<td>.007</td>
<td>.009</td>
<td>.241</td>
<td>.168</td>
<td></td>
</tr>
<tr>
<td>τ</td>
<td>ρ</td>
<td>r</td>
<td>R</td>
<td>R_U</td>
<td></td>
</tr>
<tr>
<td>-.003</td>
<td>-.005</td>
<td>-.002</td>
<td>.239</td>
<td>.157</td>
<td></td>
</tr>
<tr>
<td>.006</td>
<td>.012</td>
<td>.013</td>
<td>.185</td>
<td>.157</td>
<td></td>
</tr>
<tr>
<td>-.002</td>
<td>-.006</td>
<td>-.007</td>
<td>.174</td>
<td>.168</td>
<td></td>
</tr>
</tbody>
</table>