Hartung, Joachim

Working Paper

A note on combining dependent tests of significance

Technical Report, SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund, No. 1998,47

Provided in Cooperation with:
Collaborative Research Center 'Reduction of Complexity in Multivariate Data Structures' (SFB 475), University of Dortmund

Suggested Citation: Hartung, Joachim (1998) : A note on combining dependent tests of significance, Technical Report, SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund, No. 1998,47, SFB 475, Universität Dortmund, Dortmund

This Version is available at:
http://hdl.handle.net/10419/77121

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
A Note on Combining Dependent Tests of Significance

Joachim Hartung

Department of Statistics
University of Dortmund
Vogelpothsweg 87
D–44221 Dortmund
Germany

Abstract

In combining several tests of significance the individual test statistics are allowed to be dependent. By choosing the weighted inverse normal method for the combination, the dependency of the original test statistics is then characterized by a correlation of the transformed statistics. For this correlation a confidence region, an unbiased estimator and an unbiased estimate of its variance are derived. The combined test statistic is extended to include the case of possibly dependent original test statistics. A simulation study shows the performance of the actual significance level.

Key Words: Combining dependent test statistics, Combining p-values, Non-parametric meta-analysis, Inverse normal method, Multiple endpoints.

1 Introduction

In many situations we are lead to combine several test statistics. For instance in a clinical trial there may be several aspects of the same underlying drug influence, so-called multiple endpoints, and one might be interested in the question, whether there is any drug effect at all. Or in the meta-analysis of a series of similar studies resp. experiments a common overall effect is of interest. Often the various test statistics are so different, that a direct combination is not possible, but the corresponding p-values of the test statistics under the individual null hypothesis have to be combined.
Now the crucial point in the assumptions of the known methods is the claim that the various test statistics have to be independent, cf. for instance Birnbaum (1954), Liptak (1958), van Zwet and Oosterhoff (1967), Hedges and Olkin (1985). Particularly in the first case of application mentioned above this condition is seldomly fulfilled.

Allowing for dependency in the original test statistics, in the class of quasi means of the p-values characterized by Liptak (1958) his proposal to work with the nowadays generally called inverse normal method is chosen, because then dependency becomes equivalent to correlation. Indeed a single parametric formulation of this correlation does have to be assumed, but of course it can also be regarded as a mean correlation approximating the case of possibly different correlations between the transformed statistics. Further, the parameter sets belonging to the individual hypotheses might not be different, cf. Hedges and Olkin (1985) for a general discussion of the inverse normal method, a slight modification of it is now given below to include the case of dependent test statistics, at large following the notations of van Zwet and Oosterhoff (1967).

2 Main Results

For \( i = 1, \ldots, n \), let \( T_i \) be one-sided test statistics for testing the null-hypotheses

\[
H_{i,0} : \vartheta_i = \vartheta_{i0}
\]

for the real-valued parameters \( \vartheta_i \), against the one-sided alternatives

\[
H_{i,1} : \vartheta_i > \vartheta_{i0}
\]

where large values of \( T_i \) may lead to a rejection of \( H_{i,0} \).

It is desired to combine the results of these tests, i.e. to construct a function of \( T_1, \ldots, T_n \) that can be used to test the combined null-hypothesis

\[
H_0 : \vartheta_i = \vartheta_{i0}, \text{ for all } i = 1, \ldots, n
\]
against the alternative

\[ H_1 : \beta_i \geq \beta_{i0} \quad \text{for all } i = 1, \ldots, n \]

with strict inequality in at least one case.

If \( T_i \) has a continuous distribution function \( F_{i0} \) under the null-hypothesis \( H_{i0} \), then \( F_{i0}(T_i) \), i.e. the 1- \( p \)-value(\( T_i \)), is uniformly distributed on \((0, 1)\), and denoting \( \Phi^{-1} \) the inverse of the standard normal distribution function, the 'probit'

\[(2.1) \quad t_i = \Phi^{-1}(F_{i0}(T_i))\]

has a standard normal distribution under \( H_{i0} \), \( i = 1, \ldots, n \).

Dependency in the original test statistics \( T_1, \ldots, T_n \) leads to a dependency in the probits \( t_1, \ldots, t_n \), which now is equivalent to some correlation of the \( t_i \)'s, and we assume

\[ \text{Cov}(t_i, t_j) = \rho, \quad \text{for } i \neq j, \ i, j = 1, \ldots, n \]

with the real–valued parameter \( \rho \) in the natural parameter set, i.e. \(-\frac{1}{n-1} \leq \rho \leq 1\).

Let now \( \lambda_1, \ldots, \lambda_n \) be a set of real valued weights, with \( \sum_{i=1}^{n} \lambda_i \neq 0 \), so according to the 'weighted inverse normal method' the combined test statistic, which under \( H_0 \) is standard normally distributed, is given as follows:

\[ (2.2) \quad t(\rho) = \frac{\sum_{i=1}^{n} \lambda_i t_i}{\sqrt{\left(1 - \rho\right) \sum_{i=1}^{n} \lambda_i^2 + \rho \left(\sum_{i=1}^{n} \lambda_i\right)^2}} \sim_{H_0} N(0, 1), \]

and we see, for \( \rho = 0 \) we get the usual 'weighted inverse normal method' test statistic \( t(0) \) for independent test statistics \( T_i \), c.f. for instance Hedges and Olkin (1985), and for \( \rho = 1 \) we have just the weighted mean of the probits \( t_1, \ldots, t_n \), with weights \( \lambda_i / \sum_{j=1}^{n} \lambda_j \).

If there is no further information available about \( \rho \), the \( t_i \)'s have to be used themselves for drawing some inference about \( \rho \). So let be defined the quadratic
form

\[ q = \frac{1}{n-1} \sum_{i=1}^{n} \left( t_i - \frac{1}{n} \sum_{i=1}^{n} t_i \right)^2, \]

then, under \( H_0 \), a confidence interval for \( \rho \) at level \( 1 - \gamma, 0 < \gamma < 1 \), is given by

\[ 1 - \frac{(n-1) \cdot q}{\chi_{n-1; \gamma_1}^2} \leq \rho \leq 1 - \frac{(n-1) \cdot q}{\chi_{n-1; \gamma_2}^2}, \quad \gamma_1 + \gamma_2 = \gamma, \quad \gamma_1 \geq 0, \quad \gamma_2 \geq 0, \]

where \( \chi_{n-1; \gamma}^2 \) denotes the \( \gamma \)-quantile of the central \( \chi^2 \) distribution with \( (n-1) \) degrees of freedom.

Further, under \( H_0 \), an unbiased estimator of \( \rho \) is given by

\[ \hat{\rho} = 1 - q \]

with the variance

\[ \text{Var}(\hat{\rho}) = \frac{2}{n-1} (1 - \rho)^2, \]

which can be estimated unbiasedly by

\[ \hat{\text{Var}}(\hat{\rho}) = \frac{2}{n+1} \cdot q^2. \]

Extending the usual ‘weighted inverse normal method’ in order to include the case of possibly dependent test statistics, an estimated form of (2.2), that under \( H_0 \) is approximately standard normally distributed, is now given by

\[ t(\hat{\rho}^*, \kappa) = \frac{\sum_{i=1}^{n} \lambda_i t_i}{\sqrt{\sum_{i=1}^{n} \lambda_i^2 + \left( \sum_{i=1}^{n} \lambda_i \right)^2 - \sum_{i=1}^{n} \lambda_i^2} \left\{ \hat{\rho}^* + \kappa \cdot \sqrt{\frac{2}{n+1} (1 - \hat{\rho}^*)} \right\}^{\text{approx.}} \]

\[ \sim_{H_0} N(0, 1), \]
where

\[ \hat{\rho}^* = \max \left\{ -\frac{1}{n-1}, \hat{\rho} \right\} \]  

and \( \kappa \geq 0 \) is a parameter regulating the actual significance level, chosen for instance as \( \kappa = \kappa_1 = 0.2 \) or as \( \kappa = \kappa_2 = (1 + \frac{1}{n-1} - \hat{\rho}^*) \cdot 0.1 \), as in Table 1.

For demonstrating the influence of correlation on the significance level given by the usual test statistic \( t(0) \), cf. (2.2), and its corrections by the statistic \( t(\hat{\rho}^*, \kappa) \), cf. (2.8), with \( \kappa = \kappa_1 \) and \( \kappa = \kappa_2 \), in table 1 for some constellations of \( n, \lambda_1, \ldots, \lambda_n \) and \( \rho \) the realizations of significance levels \( \hat{\alpha} \) are simulated (10,000 runs each) for a prescribed nominal significance level of \( \alpha = 0.05 \).

For the most practical case of a nonnegative correlation the correction statistics (2.8) have a good performance, whereas for negative correlations — being more of a theoretical interest — all considered statistics are too conservative. For improvements in that case the factor \( \kappa \) in (2.8) could be weakened, allowing then, of course, the significance level to increase for nonnegative correlations.
Table 1: (*) Simulated actual significance levels \( \hat{\alpha} \) to the nominal significance level \( \alpha = 0.05 \) in several constellations of \( n, \lambda_1, \ldots, \lambda_n \) and \( \rho \) for the combined test statistics \( t(0) \), cf. (2.2), and \( t(\hat{\rho}^*, \kappa) \), c.f. (2.8), with \( \kappa_1 = 0.2 \) and \( \kappa_2 = [1 + 1/(n - 1) - \hat{\rho}^*] \cdot 0.1 \).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \lambda_i )</th>
<th>test stat.</th>
<th>( -\frac{1}{n} )</th>
<th>( -\frac{1}{2(n-1)} )</th>
<th>( \rho \geq 0 )</th>
<th>( \rho &lt; 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( t(0) )</td>
<td>0.002</td>
<td>0.009</td>
<td>0.05</td>
<td>0.06</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>( t(\hat{\rho}^*, \kappa_1) )</td>
<td>0.011</td>
<td>0.019</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>( t(\hat{\rho}^*, \kappa_2) )</td>
<td>0.013</td>
<td>0.023</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>\sqrt{3}</td>
<td>( t(0) )</td>
<td>0.003</td>
<td>0.013</td>
<td>0.05</td>
<td>0.06</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>( t(\hat{\rho}^*, \kappa_1) )</td>
<td>0.014</td>
<td>0.022</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>( t(\hat{\rho}^*, \kappa_2) )</td>
<td>0.016</td>
<td>0.025</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>( i^2 )</td>
<td>( t(0) )</td>
<td>0.024</td>
<td>0.026</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>( t(\hat{\rho}^*, \kappa_1) )</td>
<td>0.040</td>
<td>0.040</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>( t(\hat{\rho}^*, \kappa_2) )</td>
<td>0.040</td>
<td>0.040</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>( t(0) )</td>
<td>0.001</td>
<td>0.009</td>
<td>0.05</td>
<td>0.07</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>( t(\hat{\rho}^*, \kappa_1) )</td>
<td>0.001</td>
<td>0.014</td>
<td>0.04</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>( t(\hat{\rho}^*, \kappa_2) )</td>
<td>0.003</td>
<td>0.026</td>
<td>0.05</td>
<td>0.06</td>
<td>0.07</td>
</tr>
<tr>
<td>\sqrt{3}</td>
<td>( t(0) )</td>
<td>0.001</td>
<td>0.012</td>
<td>0.05</td>
<td>0.06</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>( t(\hat{\rho}^*, \kappa_1) )</td>
<td>0.003</td>
<td>0.018</td>
<td>0.04</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>( t(\hat{\rho}^*, \kappa_2) )</td>
<td>0.007</td>
<td>0.030</td>
<td>0.05</td>
<td>0.06</td>
<td>0.07</td>
</tr>
<tr>
<td>( i^2 )</td>
<td>( t(0) )</td>
<td>0.015</td>
<td>0.027</td>
<td>0.05</td>
<td>0.06</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>( t(\hat{\rho}^*, \kappa_1) )</td>
<td>0.023</td>
<td>0.030</td>
<td>0.04</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>( t(\hat{\rho}^*, \kappa_2) )</td>
<td>0.028</td>
<td>0.037</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
</tr>
</tbody>
</table>

(*) The notation 0.001 for \( \hat{\alpha} \) should be understood as \( \hat{\alpha} \leq 0.001 \).
Table 1: continued.

<table>
<thead>
<tr>
<th>$\alpha = 0.05$ nominal significance level</th>
<th>realized significance level $\hat{\alpha}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>$\lambda_i$</td>
</tr>
<tr>
<td>1</td>
<td>$t(0)$</td>
</tr>
<tr>
<td></td>
<td>$t(\hat{\rho}^*, \kappa_1)$</td>
</tr>
<tr>
<td></td>
<td>$t(\hat{\rho}^*, \kappa_2)$</td>
</tr>
<tr>
<td>10</td>
<td>$\sqrt{t}$</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>$i^2$</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>$\sqrt{t}$</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>$i^2$</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$i^2$</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3 Proof of the Results

Denote $A^T$ the transpose of a real matrix $A$, $I$ the $(n \times n)$-identity matrix and $1$ the $(n \times 1)$ vector of ones, i.e. $1^T = (1, \ldots, 1)^T_{1\times n}$. With respect to distributional properties of linear and quadratic forms let us in general refer for instance to Mathai and Provost (1992).

Putting now, c.f. (2.1),

\begin{equation}
(3.1) \quad x = (t_1, \ldots, t_n)^T
\end{equation}

we get under $H_0$ the covariance matrix of $x$

\begin{equation}
(3.2) \quad \text{Cov}(x) = (1 - \rho)I + \rho 11^T,
\end{equation}

and with $\lambda = (\lambda_1, \ldots, \lambda_n)^T$ for the variance of $\lambda^T x$:

\begin{equation}
(3.3) \quad \text{Var}(\lambda^T x) = \lambda^T \text{Cov}(x) \lambda
\end{equation}

\begin{equation*}
= (1 - \rho)\lambda^T I \lambda + \rho \cdot \lambda^T 1 1^T \lambda
\end{equation*}

\begin{equation*}
= (1 - \rho) \sum_{i=1}^{n} \lambda_i^2 + \rho \cdot \left( \sum_{i=1}^{n} \lambda_i \right)^2,
\end{equation*}

which yields (2.2). Let us define now the projection matrix

\begin{equation}
(3.4) \quad K = I - \frac{1}{n}11^T, \quad K = K^T, \quad K = K^2,
\end{equation}

then for the quadratic form $q$ of (2.3) we have the equivalent representation

\begin{equation}
(3.5) \quad q = \frac{1}{n-1}x^T K x,
\end{equation}
such that by the following property
\[
K \cdot \text{Cov}(x) = (1 - \rho)K + \rho \cdot K \cdot 11^T
\]
\[
= (1 - \rho)K + \rho \cdot (I - \frac{1}{n}11^T)11^T
\]
\[
= (1 - \rho)K + \rho \cdot (1 - \frac{1}{n}1 \cdot n)1^T
\]
\[
= (1 - \rho)K ,
\]
with \( \text{trace}(K) = n - 1 \), we get under \( H_0 \)
\[
\frac{1}{1 - \rho} \cdot (n - 1)q \overset{H_0}{\sim} \chi^2_{n-1} ,
\]
which, with \( \gamma_1 + \gamma_2 = \gamma \), gives the \((1 - \gamma)\)-confidence interval for \( 1/(1 - \rho) \):
\[
\frac{1}{(n - 1)q} \cdot \chi^2_{n-1;\gamma_1} \leq \frac{1}{1 - \rho} \leq \frac{1}{(n - 1)q} \cdot \chi^2_{n-1;1-\gamma_2} ,
\]
respectively for \((1 - \rho)\):
\[
\frac{(n - 1)q}{\chi^2_{n-1;1-\gamma_2}} \leq 1 - \rho \leq \frac{(n - 1)q}{\chi^2_{n-1;\gamma_1}} ,
\]
yielding now directly the \((1 - \gamma)\)-confidence interval for \( \rho \) given in (2.4).

Now by (3.6), under \( H_0 \), we get for the expectation
\[
E \left( \frac{1}{1 - \rho}(n - 1)q \right) = n - 1 , \text{ respectively}
\]
\[
E q = 1 - \rho ,
\]
and for the variance
\[
2(n - 1) = \text{Var} \left( \frac{1}{1 - \rho}(n - 1)q \right)
\]
\[
= \frac{1}{(1 - \rho)^2} \cdot (n - 1)^2 \cdot \text{Var}(q) , \text{ respectively}
\]
\[
\text{Var}(q) = 2 \cdot \frac{1}{n - 1} \cdot (1 - \rho)^2 ,
\]
an unbiased estimator which is given by, cf. Hartung and Voet (1986),

\begin{equation}
\hat{\text{Var}}(q) = 2 \cdot \frac{1}{n+1} q^2.
\end{equation}

Now \( \text{Var}(q) = \text{Var}(1 - q) \), so that (2.5), (2.6) and (2.7) are also shown.

Finally, although under \( H_0 \) in (2.2) we may use the unbiased estimator (2.5) for \( \rho \), we have to recognize the square root function to be concave, so that in expectation the denominator would be underestimated. Therefore in (2.8) the estimate of \( \rho \) is corrected by adding a 'small amount' of its estimated standard deviation under \( H_0 \).

\section{Final Remark}

In this paper we have shown how to modify the 'inverse normal method' of non-parametric meta-analysis in order to include the case of possibly dependent test statistics resp. dependent \( p \)-values. In the worst case one can take at least a weighted mean of the probits.

So we recommend to take our procedure in consideration if one is not absolutely sure whether the statistics to be combined are really independent.

\textbf{Acknowledgement:} Thanks are due to Boris Weimann, University of Dortmund, for carrying out the simulations.
References


