Multiple Rank-based Testing for Ordered Alternatives with Incomplete Data

Paul Cabilio and Jianan Peng

Acadia University

Outline

- Introduction
- Example with Incomplete Blocks
- Rank Correlations and the Page and Jonckheere Tests
- Spearman and Kendall Correlations for Incomplete Rankings
- Extended Page and Jonckheere Statistics
- Step Down Multiple Test Procedures
- Application to an Example
- Discussion

INTRODUCTION

Consider an experiment

$$x_{ij} = \beta_i + \tau_j + e_{ij} \quad i = 1, .., n \quad j = 1, ..., t,$$

t = number of treatments; n = number of blocks;

τ_j are treatment effects, and β_i are block effects.

$$H_0 : \tau_1 = \tau_2 = ... = \tau_t \qquad H_1 : \tau_1 \leq \tau_2 \leq ... \leq \tau_t$$

with at least one inequality in the alternative strict.

Rank based procedures include the tests of Page (1963) and Jonckheere (1954) . If the null hypothesis is rejected, it is of interest to conduct tests so as to identify which inequalities in the alternative are strict, and in so doing maintain the experimentwise error rate at a pre-assigned level.

Example: Experimental results given in Jones, Wentzell, and Toews (1992)

	Treatment			
Block	$T1$	$T2$	$T3$	$T4$
1	11.865	9.832	7.567	10.168
2	5.601	4.892	4.032	3.126
3		14.415	14.185	7.800
4	13.267			9.953
5	8.006	7.793		7.582
6	17.692	16.644	15.327	11.573
7	9.027	7.973	11.855	6.820
8	9.789	7.967	7.758	7.849

Lymph Heart Pressure in mm Hg taken over a 24 hour period at 6-hour intervals on 8 toads during dehydration.

$$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 \qquad H_1 : \tau_1 \geq \tau_2 \geq \tau_3 \geq \tau_4$$

with at least one inequality strict.

Both *Page* and *Jonckheere* statistics are of the form

$$\sum_{i=1}^{n} \mathcal{A}(i)$$

where $\mathcal{A}(i)$ is an (unstandardized) rank correlation between the ranking in each block and the criterion ranking chosen to be $(1, 2, ..., t)$. For Page it is the *Spearman* correlation, and for Jonckheere the *Kendall* correlation:

$$\begin{aligned} \mathcal{A}_S(i) &= \sum_{j=1}^{t} \left(j - \frac{t+1}{2}\right)\left(\mu_i(j) - \frac{t+1}{2}\right) = \sum_{j=1}^{t} j\mu_i(j) - \frac{t(t+1)^2}{4} \\ \mathcal{A}_K(i) &= \sum_{l<m}^{t} sgn\left(\mu_i(m) - \mu_i(l)\right) \end{aligned}$$

where $\mu_i(j)$ is the rank of treatment j in block i

INCOMPLETE RANKINGS

Alvo and Cabilio (1995a), developed measures of similarity for incomplete rankings. With one ranking the complete criterion ranking, and with k_i observed treatments in the other, $2 \le k_i \le t$, these become:

$$\textit{Spearman: } \mathbf{A}_S(i) = \frac{t+1}{k_i+1} \sum_{j=1}^{t} j u_{ij} - \frac{t(t+1)^2}{4}$$

$$u_{ij} = \mu_i^*(j)\,\delta_i(j) + \frac{k_i+1}{2}\left(1 - \delta_i(j)\right)$$

$\mu_i^*(j)$ the (incomplete) ranking of treatment j in block i,

$\delta_i(j) = 1$ or 0, indicating whether treatment j is or is not present in block i.

Criterion	1	2	3	4	5	6	7	8	9
Ranking	2	1	3	-	5	-	6	4	-
u_{ij}	2	1	3	3.5	5	3.5	6	4	3.5
$\frac{t+1}{k_i+1}u_{ij}$	$\frac{20}{7}$	$\frac{10}{7}$	$\frac{30}{7}$	5	$\frac{50}{7}$	5	$\frac{60}{7}$	$\frac{40}{7}$	5

$(t = 9, k_i = 6)$

$$\textit{Kendall} \quad : \quad \mathbf{A}_K(i) = \sum_{l<m}^{i} a_i(l,m)$$

$$a_i(l,m) = \begin{cases} sgn\left(\mu_i^*(m) - \mu_i^*(l)\right) & \text{if } \delta_i(l) = \delta_i(m) = 1 \\ 2\mu_i^*(m)/(k_i+1) - 1 & \text{if } \delta_i(l) = 0, \delta_i(m) = 1 \\ 1 - 2\mu_i^*(l)/(k_i+1) & \text{if } \delta_i(l) = 1, \delta_i(m) = 0 \\ 0 & \text{otherwise} \end{cases}$$

Example :

Treatment	1	2	3	4	5	6	7	8	9
Ranking	2	1	3	-	5	-	6	4	-

.

$$\begin{aligned} a_i(1,2) &= sgn\left(\mu_i^*(2) - \mu_i^*(1)\right) = -1 \\ a_i(2,3) &= sgn\left(\mu_i^*(3) - \mu_i^*(2)\right) = 1 \\ a_i(4,5) &= 2\mu_i^*(5)/7 - 1 = 3/7 \\ a_i(5,6) &= 1 - 2\mu_i^*(5)/7 = -3/7 \\ a_i(4,6) &= 0 \end{aligned}$$

TEST STATISTICS: COMPLETE CASE

In the complete case the *Page* and *Jonckheere* statistics may be expressed as

$$L = \sum_{i=1}^{n} \mathcal{A}_S(i) + \frac{nt(t+1)^2}{4} = \sum_{j=1}^{t} jR_j$$

$$JK = \frac{1}{2}\sum_{i=1}^{n} \mathcal{A}_K(i) + \frac{nt(t-1)}{4} = \sum_{i=1}^{n}\sum_{l<m} U_i(l,m)$$

$R_j = \sum_{i=1}^{n} \mu_i(j)$, and $U_i(l,m)$ is the *Mann-Whitney* score,

$$U_i(l,m) = \frac{1}{2}\left(1 + sgn\left(\mu_i(m) - \mu_i(l)\right)\right) = \begin{cases} 1 \text{ if } \mu_i(l) < \mu_i(m) \\ 0 \text{ if } \mu_i(l) > \mu_i(m) \end{cases}$$

TEST STATISTICS: INCOMPLETE CASE

In the incomplete situation the *Extended Page* and the *Extended Jonckheere* statistics may be written as

$$L^* = \sum_{i=1}^{n} \mathbf{A}_S(i) + \frac{nt(t+1)^2}{4} = \sum_{j=1}^{t} jR_j^*$$

$$JK^* = \frac{1}{2}\sum_{i=1}^{n} \mathbf{A}_K(i) + \frac{nt(t-1)}{4} = \sum_{i=1}^{n}\sum_{l<m} U_i^*(l,m)$$

$R_j^* = \sum_{i=1}^{n} \frac{(t+1)}{(k_i+1)} u_{ij}$, and $U_i^*(l,m) = (1 + a_i(l,m))/2$ is an *Extended Mann-Whitney* score.

It can be shown that under H_0, as $n \to \infty$, and conditional on the pattern of missing observations, both statistics L^* and JK^* are asymptotically normally distributed.

STEP DOWN TEST PROCEDURE

The form of both the extended Page and Jonckheere statistics makes it possible to apply a general step-down testing procedure for multiple comparisons such as that proposed in Marcus, Peritz, and Gabriel (1976). In terms of our example with $t = 4$, the procedure is as follows, with all individual tests at level α.

Step 1 Test $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 \quad H_1 : \tau_1 \geq \tau_2 \geq \tau_3 \geq \tau_4$. If H_0 is rejected proceed to Step 2. Otherwise stop.

Step 2 Test

[1] $H_0 : \tau_1 = \tau_2 = \tau_3;\ H_1 : \tau_1 \geq \tau_2 \geq \tau_3$

[2] $(H_0 : \tau_1 = \tau_2;\ H_1 : \tau_1 > \tau_2) \wedge (H_0 : \tau_3 = \tau_4;\ H_1 : \tau_3 > \tau_4)$

[3] $H_0 : \tau_2 = \tau_3 = \tau_4;\ H_1 : \tau_2 \geq \tau_3 \geq \tau_4$

If no H_0 is rejected, stop. If one or more H_0 is rejected, go to Step 3.

Step 3 The candidate tests at this step are

[1] $H_0 : \tau_1 = \tau_2;\ H_1 : \tau_1 > \tau_2$

[2] $H_0 : \tau_2 = \tau_3;\ H_1 : \tau_2 > \tau_3$

[3] $H_0 : \tau_3 = \tau_4;\ H_1 : \tau_3 > \tau_4$

A test in Step 3 is conducted **if and only if all the null hypotheses in Step 2 which imply it** were rejected.

Example: **Step 1** ($\alpha = 0.05$).

First test $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4$; $H_1 : \tau_1 \geq \tau_2 \geq \tau_3 \geq \tau_4$

			Treatment			
Block	$T1$	$T2$	$T3$	$T4$	k_i	$\frac{t+1}{k_i+1}$
1	1	3	4	2	4	1
2	1	2	3	4	4	1
3	2	1	2	3	3	5/4
4	1	1.5	1.5	2	2	5/3
5	1	2	2	3	4	5/4
6	1	2	3	4	4	1
7	2	3	1	4	4	1
8	1	2	4	3	4	1
R_j^*	11.416	18.250	22.500	27.833		

Scores u_{ij} for the example.

$L^* = 226.75$. The Normal approximation gives a $p - value \approx 0.0002$. Exact tail probabilities are given in Alvo and Cabilio (1995b) , giving $P(L^* \geq 226.75) = 0.00006$. Conclusion: $\tau_1 > \tau_4$

Test all the hypotheses in **Step 2**

$H_0 : \tau_1 = \tau_2 = \tau_3$; $H_1 : \tau_1 \geq \tau_2 \geq \tau_3$. Now t is taken to be 3.

	Treatment				
Block	$T1$	$T2$	$T3$	k_i	$\frac{t+1}{k_i+1}$
1	1	2	3	3	1
2	1	2	3	3	1
3	1.5	1	2	2	4/3
4	$*$	$*$	$*$	1	$*$
5	1	2	1.5	2	4/3
6	1	2	3	3	1
7	2	3	1	3	1
8	1	2	3	3	1
R_j^*	9.333	15	17.667		

Scores u_{ij} for data in partition (123).

$L^* = 92.333$,and using the Normal approximation gives a $p - value \approx 0.006$.
Conclusion: $\tau_1 > \tau_3$

Step 2 (continued)$(H_0 : \tau_1 = \tau_2;\ H_1 : \tau_1 > \tau_2) \wedge (H_0 : \tau_3 = \tau_4;\ H_1 : \tau_3 > \tau_4)$. Note that with $t = 2$, the block is either complete, in which case $(t+1)\,/\,(k_i+1) = 1$, or the block is unusable. With only two treatments, the test is the usual Page test, which is equivalent to a one sided Sign test.

	Treatment		*Treatment*	
Block	$T1$	$T2$	$T3$	$T4$
1	1	2	2	1
2	1	2	1	2
3	*	*	1	2
4	*	*	*	*
5	1	2	*	*
6	1	2	1	2
7	1	2	1	2
8	1	2	2	1
R_j^*	6	12	8	10

Scores u_{ij} *for data in partition* $((12) \wedge (34))$.

For the two subsets in this case $L_1^* = 30$, and $L_2^* = 28$ and both statistics have large sample Normal distributions.

With $L^* = L_1^* + L_2^*$ as the test statistic, use of a Normal approximation gives a $p - value \approx 0.011$.

Using binomial probabilities we have an exact $p - value = 0.019$.

Conclusion $\tau_1 > \tau_2$ or $\tau_3 > \tau_4$.

Other partitions are handled in a similar fashion.

RECAP OF CONCLUSIONS USING EXTENDED PAGE TEST

To recap the results we have the following.

Stage	Partition	L^*	Stand. L^*	$p-value$	Conclusion
1	(1234)	226.75	3.502	0.00006	$\tau_1 > \tau_4$
2	(123)	92.333	2.525	0.006	$\tau_1 > \tau_3$
	(12) $\wedge$ (34)	58	2.309	0.019 (Exact)	$\tau_1 > \tau_2$ or $\tau_3 > \tau_4$
	(234)	92.333	2.245	0.012	$\tau_2 > \tau_4$
3	(12)	30		0.016 (Exact)	$\tau_1 > \tau_2$
	(23)	29		0.109 (Exact)	–
	(34)	28		0.344 (Exact)	–

Conclusions using extended Page statistic ($\alpha = 0.05$).

Thus we conclude that $\tau_1 > \tau_2 > \tau_4$ and $\tau_1 > \tau_3$.

CONCLUSIONS USING EXTENDED JONCKHEERE TEST

All the calculations at each stage were carried out using an implementation of a *Minitab* macro. Another macro was written in order to carry out the testing using the extended Jonckheere test. The results are very similar to the ones given above, and are recapped in the table below. The cases in Stage 3, with $t = 2$, are identical, since in this case also the test is equivalent to the Sign test.

Stage	Partition	JK^*	Stand. JK^*	$p - value$	Conclusion
1	(1234)	38.667	3.805	0.00006	$\tau_1 > \tau_4$
2	(123)	17	2.833	0.002	$\tau_1 > \tau_3$
	(12) $\wedge$ (34)		2.309	0.011	$\tau_1 > \tau_2$ or $\tau_3 > \tau_4$
	(234)	16.333	2.344	0.010	$\tau_2 > \tau_4$
3	(12)	30		0.016 (Exact)	$\tau_1 > \tau_2$
	(23)	29		0.109 (Exact)	–
	(34)	28		0.344 (Exact)	–

DISCUSSION

Efficiency

The extensions to the Page and Jonckheere tests were developed in order to make use of the information contained in blocks with missing observations. Such an approach is preferable to the situation where blocks with missing observations are deleted so as to be able to use the original statistics. Alvo and Cabilio (1995b) show that in the special case where the missing observations follow a balanced incomplete block design (BIBD), with k out of t observations per block, the asymptotic relative efficiency in the Pitman sense of the Extended Page test to the Page test is $(k-1)(t+1) / [(k+1)(t-1)]$. This value is close to $(k-1)t / [k(t-1)]$, the efficiency factor of the BIBD (Scheffé, 1959).

DISCUSSION

Power

Cabilio and Tilley (1999), conducted simulations to calculate the power of tests of trend using the Spearman and Kendall correlations with missing observations developed in Alvo and Cabilio (1995a). It was shown that for various types of trend and patterns of missing observations, in most cases both of these statistics were more powerful than the corresponding statistics in which missing observations were deleted. Since the extended Page and Jonckheere tests are sums of such independent statistics, this gives an indication that power may be lost if blocks with missing observations are deleted.

DISCUSSION

Bonferroni

An alternative to the step down test procedure was introduced by Budde and Bauer (1989). This is a Bonferroni-type of procedure in which only $t-1$ pairwise comparisons of adjacent treatments are considered. In the case of missing observations, such a procedure would be equivalent to the last step given in the example above, in which all pairs with missing observations are deleted. The levels would be Bonferroni adjusted. Such a procedure does not take advantage of the additional information used in the extended tests at previous steps, and as such may be expected to have smaller power.

REFERENCES

Alvo, M. and Cabilio, P. (1995a). "Rank correlation methods for missing data," Canadian Journal of Statistics, 23, 345-358.

Alvo, M. and Cabilio, P. (1995b). "Testing ordered alternatives in the presence of incomplete data,", Journal of the American Statistical Association, 90, 1015-1024.

Budde, M., and Bauer, P. (1989). "Multiple test procedures in clinical dose finding trials," Journal of the American Statistical Association, 84, 792-796.

Cabilio, P. and Tilley, J. (1999). "Power calculations for tests of trend with missing observations,". Environmetrics,10, 803-816.

Jonckheere, A. R. (1954). "A test of significance for the relation between m rankings and k ranked categories," The British Journal of Statistical Psychology, 7, 93-100.

Jones, J. M., Wentzell, L. A., and Toews, D. P. (1992). "Posterior lymph heart pressure and rate and lymph flow in the toad Bufo marinus in response to hydrated and dehydrated conditions," The Journal of Experimental Biology, 169, 207-220.

Marcus, R., Peritz, E., and Gabriel, K. R. (1976). "On closed testing procedures with special reference to ordered analysis of variance

Page, E. B., (1963). "Ordered hypotheses for multiple treatments: a significance test for linear ranks," Journal of the American Statistical Association, 58, 216-230.

Scheffé, H. (1959). The Analysis of Variance. John Wiley & Sons, New York.