Studying norms in small groups by means of multi-group concordance analysis[#]

Gaj Vidmar[1] and *Matej Černigoj[2]*
[1]University of Ljubljana, Faculty of Medicine, Institute of Biomedical Informatics, Ljubljana, Slovenia
[2]University of Ljubljana, Faculty of Arts, Department of Psychology, Ljubljana, Slovenia

Abstract: An experiment in the field of social psychology is presented, emphasising the application of some less renowned statistical methods for analysing ranked data. The experiment addresses two principal issues: the association between norm congruency and formation of peer groups and the internal (personal) vs. the external (behavioural) aspect of norms. This calls for analysing agreement within and between several groups of judges ranking several objects. After an overview of the interrelations between different forms and measures of rank-correlation and concordance is provided, the problem is tackled by means of ANACONDA (a procedure analogous to ANOVA, developed by Schucany and associates in the nineteen-seventies), whereby a solution is proposed for the previously unaddressed issue of testing overall (factor) effects. Finally, estimation of confidence intervals for concordance coefficient in the non-null case is illustrated.

Key words: norms, concordance coefficients, ANACONDA, ranks

Preučevanje norm v majhnih skupinah z medskupinsko analizo konkordance

Gaj Vidmar[1] in *Matej Černigoj[2]*
[1]Univerza v Ljubljani, Medicinska fakulteta, Institut za biomedicinsko informatiko, Ljubljana
[2]Univerza v Ljubljani, Oddelek za psihologijo, Ljubljana

Povzetek: V članku je predstavljen eksperiment s področja socialne psihologije, pri čemer je izpostavljena uporaba manj znanih statističnih metod za analizo podatkov v obliki rangov. Eksperiment je namenjen preučevanju dveh vprašanj: povezave skladnosti norm z oblikovanjem vrstniških skupin ter razlike med notranjim (osebnim) in zunanjim (vedenjskim) vidikom norm. V ta namen je potrebna analiza skladnosti znotraj skupin in med skupinami ocenjevalcev, ki ocenjujejo več objektov. Po teoretičnem pregledu obstoječih koeficientov rang-korelacije in konkordance je za analizo zbranih podatkov uporabljena ANACONDA (večskupinska analiza konkordance, analogna analizi variance, ki so jo v 1970-ih letih razvili Schucany in sodelavci). V okviru te metode je predlagana rešitev za doslej neobravnavani problem testiranja skupnega učinka posameznega dejavnika. Na koncu je prikazan postopek ocenjevanja intervala

---

zaupanja za koeficient konkordance v primeru, da ničelna hipoteza ne drži.

Ključne besede: norme, koeficienti konkordance, ANACONDA, rangi

Background

Norms are regarded a paramount social phenomenon by many social psychologists. Turner (1991), for example, considers social norm to be the key concept in understanding what researches mean by social influence, which is a topic central to social psychology. Despite this general acknowledgement, the concept of norm remains very poorly defined up to this day. There is still no generally accepted theory of norms and the term itself is also quite elusive. According to Bečaj (1995), most researches use the term *social norm* to address a behavioural rule appertaining to the macrosocial level (e.g., culture), whereas for small groups the terms *group norm* and *group standard* are preferred. Nevertheless, as such distinctions are far beyond the scope of this paper, the term norm will be used hereinafter regardless of the size of the group in question.

Smith (1988) provides a general working definition of what norms might be. When interacting with strangers, people must base their behaviour on assumptions about how the others would react to it. These assumptions may be internalised in the form of feelings as to what does or does not feel like an appropriate behaviour in a given setting, or they may be externalised as feelings of wishing to please (or, in some cases, to displease) particular others one encounters. When a group of people continues to meet for some time, opportunities arise to test these general assumptions and to adapt them to the others' actual reactions. In time, any group that continues to meet develops an increasingly unique set of norms about how the members should behave in each other's presence. Although in order to be considered proper norms, these behavioural rules have to be shared by a great majority of members (ranging from 2/3 to 3/4 according to different authors), they are usually not spelled out very clearly. In fact, they are rarely accessible to conscious inspection. However, when somebody starts acting in discordance with them, other members of the group react more or less vigorously in order to put the deviant back "on track".

Our investigation is a modification of Smith's (1988) exercise, which is aimed to investigate the process of norm formation and maintenance in an actual group of students for the purpose of demonstration of basic concepts in social psychology and its research methodology. The nature of that investigation is purely correlational and there is no explicit hypothesis testing. The data are supposed to be used to determine some important features of the group of students, such as the level of group members' consensus regarding particular behavioural rules and the level of concordance

regarding their relative importance. Only very basic statistics is required for that. The setting, however, allows for the use of more advanced methods for analysing ranked data, including some less renowned and more complex ones. The application of these methods could provide some insight into certain questions of broader theoretical importance.

For example, the association between norm congruency and formation of peer groups could be analysed. It is a well-known fact in social psychology that people seek the company of similar individuals for various reasons (often to evaluate themselves, as suggested by Festinger's theory, according to Turner, 1991). Therefore, it was expected that peer groups of four to six students, formed on the basis of free choice before the onset of the investigation (and not for the purpose of it), would exhibit a greater degree of concordance in judging the relative importance of behavioural rules as compared to the entire class. Moreover, the two-way ANACONDA procedure, which is the central subject of the paper, allows one to assess interaction between the judgements of behavioural rules in peer groups and two different judging conditions: one concerned with the external aspect of norms (i.e., how does one react to other's deviance from the norm) and the other concerned with the internal aspects of norms (i.e., how does one feel when he/she happens to be the deviant himself/herself).

To introduce concordance analysis, the interrelations between different forms and measures of rank-correlation and concordance are briefly reviewed next. Computation of confidence intervals for concordance in the non-null case is briefly illustrated at the end of the paper. As the emphasis of the whole study is on methodology rather than on interpretation, only statistical aspects of the various procedures and parameters are presented and discussed henceforth, instead of their psychological justifications and/or implications.

Overview of concordance coefficients

While numerous references can provide the necessary pieces of information for such a review, ours is based upon Palachek & Schucany (1984) and Siegel & Castellan (1988). There are several coefficients available for assessing inter-group concordance, i.e., for testing agreement among m rankings of k objects. All such measures are based upon bivariate ordinal correlation – either Spearman's ρ or Kendall's τ. The best-known and most widely used solution was proposed by Kendall and Babington-Smith in 1939. Their coefficient of concordance W, which is basically the average ρ of all possible pairs of rankings [1], is related to Friedman's χ^2 for two-way analysis of variance using ranks [2], introduced in 1937, which also provides the means for testing significance of W in terms of rejection of the null-hypothesis of no concordance:

$$W = [(m\text{-}1) \text{ average}(\rho_{\text{ranking-ranking}}) + 1]/m \qquad (1)$$

$$\chi^2 = W m (k\text{-}1), df = k\text{-}1 \qquad (2)$$

Kendall's coefficient of agreement is obtained by substituting ρ with τ in [1] ($u_K = W_\tau$), while in 1952 Ehrenberg proposed the simple average of τ's as a related coefficient of agreement ($u_E = \text{average}(\tau_{\text{ranking-ranking}})$). Another measure, proposed by Quade in 1972, is the expected value of ρ's ($u_Q = E(\rho_{\text{ranking-ranking}})$), which is the basis for estimating confidence intervals for coefficient of concordance in the non-null case. Last, but not least relevant to two-group concordance analysis, there is the correlation between a group of judges and a criterion, defined as the average of individual judge-criterion correlations ($T_C = \text{average}(\tau_{\text{ranking-criterion}})$).

Method

Participants

47 undergraduate psychology students, attending introductory course in social psychology, participated in the experiment. They were previously (not for the purpose of the experiment) divided into 9 groups of 4 to 6 according to their own preferences.

Procedure

*S*s first discussed which implicit behavioural rules (meaning those not being imposed by some authority) are respected in their class. The discussions went on in the small groups and at the end each group rendered a list of 10 such rules to the experimenter, who later chose 12 most frequently stated rules. On the next encounter *S*s were presented with the list of 12 selected rules and given instructions as described below.

Instruments and research design

A list of 12 behavioural rules (formed on the basis of previous relevance assessment described below), expressed in terms of rule violation, was applied with the following instructions on how to rank the norm violations in two successive conditions:

- *Condition 1: S*s had to rank the rules' violations according to how they would react themselves if a fellow student acted that way, rank 12 meaning "I couldn't bear such behaviour" and rank 1 meaning "I wouldn't condemn such behaviour or would even approve of it".
- *Condition 2: S*s had to rank the rules' violations according to how they would feel if it happened that they had acted that way themselves. Here, rank 12

meant one would have felt very bad about it or ashamed of it, while rank 1 meant one would not have felt troubled or would even be proud of it.

Results and discussion

Description of the data

The behaviours listed in table 1 were chosen for further study. To provide a first impression of the data, descriptive statistics (mean, minimum and maximum rank) for ranks pooled across all groups are reported in table 2 for both conditions. The purpose of this is purely descriptive, since concordance rather than central tendency is the subject of investigation, and groups of students are of primary interest rather than the class as a whole.

Kendall's *W* values for the two experimental conditions are reported in table 3 for all groups and for the whole class, together with group sizes. All concordance coefficients are significantly different from zero: *p* values obtained by the standard χ^2 test are smaler than 0.001 for all the peer groups under both conditions, except for

Table 1. Behaviour rules studied.

A	Someone does not lend his/her course notes when asked to by a colleague.
B	Someone doesn't even notice a colleague when passing him/her in town.
C	Someone has obsequious attitude towards the professors.
D	A group always sticks together and doesn't allow others to join their company.
E	Someone copies a report (assignment, essay) without permission by the author.
F	Someone does not congratulate a colleague for having passed an exam.
G	When discussing exam schedule, someone always tries to impose his/her will.
H	Someone behaves in a self-conscious and contrived manner.
I	Someone is very active in class, asking a lot of questions.
J	Someone does not spontaneously offer a chewing gum (candy).
K	Someone is "stoned" in the class.
L	Someone chats with the neighbour during a lecture.

Table 2. Descriptive statistics for rule ranks.

Rule		A	B	C	D	E	F	G	H	I	J	K	L
	Mean	10.2	7.9	7.9	7.0	9.9	4.4	9.3	6.0	2.5	2.6	5.0	5.3
Condition 1	*Min*	3	1	2	3	3	1	5	1	1	1	1	1
	Max	12	12	11	12	12	9	12	12	7	7	12	11
	Mean	10.3	7.7	8.7	6.5	9.9	4.6	8.2	7.0	2.1	2.9	6.5	3.7
Condition 2	*Min*	5	2	3	2	3	1	4	2	1	1	1	1
	Max	12	12	12	11	12	9	11	12	6	6	12	9

Table 3. Kendall's W values for norms' rankings.

Group	1	2	3	4	5	6	7	8	9	Class
Condition 1	0.65	0.83	0.62	0.56	0.52	0.61	0.62	0.82	0.61	0.54
Condition 2	0.55	0.79	0.60	0.60	0.59	0.62	0.59	0.81	0.68	0.55
N	5	5	4	6	5	5	6	5	6	47

group 1 under condition 2 (where $p=0.01$). For the class, p is less than 10^{-50} under both conditions. This indicates agreement within each group, as well as within the class as a whole, regarding both the external and the internal aspect of violating given behaviour rules.

Two-group and multi-group concordance analysis

The basis of multi-group concordance analysis is the two-group $\mathscr{L}$ statistic, introduced by Schucany & Frawley (1973). They proposed a generalisation of the logic of the Page test for ordered alternatives (i.e., of the L statistics) as a means for testing concordance both within and between two groups of judges (of size n_1 and n_2, respectively) ranking k objects. If **s** and **t** denote vectors of column sums whereby the judges (rankings) are represented by n_1 and n_2 rows, respectively, and the objects are represented by k columns of a table, the statistic is

$$\mathscr{L} = \Sigma_k\, S_i\, T_i = \mathbf{s'}\, \mathbf{t} \tag{3}$$

The appropriate test statistics is $\mathscr{L}^*$, which is based upon the expected value and variance of $\mathscr{L}$, which, in turn, are derived from $E(S_i)$, $var(S_i)$ and $cov(S_i, S_j)$:

$$\mathscr{L}^* = [\mathscr{L} - E(\mathscr{D})] / var(\mathscr{D}) \tag{4}$$

$$E(\mathscr{D}) = n_1\, n_2\, k\, (k+1)^2 / 4 \tag{5}$$

$$var(\mathscr{D}) = n_1\, n_2\, (k-1)\, k^2\, (k+1)^2 / 144 \tag{6}$$

The normal approximation of $\mathscr{L}^*$ has been appraised by Schucany & Frawley (1973). Due to the finite range of the $\mathscr{L}$ statistic (its minimum and maximum are given by [8] and [9], respectively), the authors proposed another standardisation in the same article, producing a statistics within the usual interval for a correlation measure ($\mathscr{W} \in [-1, 1]$; $\mathscr{W} = 1$ when there is complete agreement within each group on the same rank ordering, $\mathscr{W} = -1$ when there is perfect agreement within each group but the two groups manifest opposite ordering):

$$\mathscr{W} = [\mathscr{L} - E(\mathscr{D})] / [\max(\mathscr{D}) - E(\mathscr{D})] \tag{7}$$

$$\min(\mathscr{D}) = n_1\, n_2\, k\, (k+1)\, (k+2)\, /\, 6 \qquad (8)$$

$$\max(\mathscr{D}) = n_1\, n_2\, k\, (k+1)\, (2k+1)\, /\, 6 \qquad (9)$$

The next step was to extend the two-group test to multiple groups. Zero correlation and asymptotic independence between any two $\mathscr{L}$'s and between one of the $\mathscr{L}$ statistics and a related Friedman χ^2 were established by Beckett (1975, cited in Beckett & Schucany, 1975), extending the 2-group proof by Li & Schucany (1975), and Beckett & Schucany (1975) introduced ANACONDA (multi-group Analysis of Concordance, named on the basis of the similarity of the concept to ANOVA). The procedure they proposed is exemplified by table 4. It should be emphasised that whenever a discordant pair of groups (i.e., nonsignificant $\mathscr{L}^*$) is found in ANACONDA, one should proceed with the Friedman's χ^2-analysis within the two discordant groups in order to find out whether there is disagreement within one and/or two groups or the rankings are concordant within each group but the groups differ in their overall ranking of the objects.

A problem arises, though, when addressing the peer-groups related issues of our experiment, as the test statistics for factor effects cannot be found in the literature. Namely, the article introducing $\mathscr{L}$ (Schucany & Frawley, 1973) even suggests a different approach ($\mathscr{L}_3 = S_k\, R_i\, S_i\, T_i$ for the 3-group case) when addressing multi-group concordance in its final part, while the articles presenting ANACONDA (Beckett & Schucany, 1975, and Beckett & Schucany, 1979) do not present such an "overall"

Table 4. One-way ANACONDA table for 3x2 layout (first factor levels A,B,C; second factor levels x,y; adapted from Beckett & Schucany, 1975).

Source of Variation	Sum of Products	Test Statistic	Interpretation
Within Group Ax	$s_{Ax}'\, s_{Ax}$	χ^2_{Ax}	
Within Group Ay	$s_{Ay}'\, s_{Ay}$	χ^2_{Ay}	
Within Group Bx	$s_{Bx}'\, s_{Bx}$	χ^2_{Bx}	Agreement Within
Within Group By	$s_{By}'\, s_{By}$	χ^2_{By}	Each Group
Within Group Cx	$s_{Cx}'\, s_{Cx}$	χ^2_{Cx}	
Within Group Cy	$s_{Cy}'\, s_{Cy}$	χ^2_{Cy}	
Group A vs. Group B	$(s_{Ax} + s_{Ay})'\, (s_{Bx} + s_{By})$	$\mathscr{L}_{AB}$	
Group A vs. Group C	$(s_{Ax} + s_{Ay})'\, (s_{Cx} + s_{Cy})$	$\mathscr{L}_{AC}$	Factor 1 Effect
Group B vs. Group C	$(s_{Bx} + s_{By})'\, (s_{Cx} + s_{Cy})$	$\mathscr{L}_{BC}$	
Group x vs. Group y	$(s_{Ax} + s_{Bx} + s_{Cx})'\, (s_{Ay} + s_{By} + s_{Cy})$	$\mathscr{L}_{xy}$	Factor 2 Effect
Group Ax vs. Group By	$s_{Ax}'\, s_{By}$	$\mathscr{L}_{AxBy}$	
Group Ax vs. Group Cy	$s_{Ax}'\, s_{Cy}$	$\mathscr{L}_{AxCy}$	
Group Bx vs. Group Ay	$s_{Bx}'\, s_{Ay}$	$\mathscr{L}_{BxAy}$	Factor 1 $\times$ Factor 2
Group Bx vs. Group Cy	$s_{Bx}'\, s_{Cy}$	$\mathscr{L}_{BxCy}$	Interactions
Group Cx vs. Group Ay	$s_{Cx}'\, s_{Ay}$	$\mathscr{L}_{CxAy}$	
Group Cx vs. Group By	$s_{Cx}'\, s_{By}$	$\mathscr{L}_{CxBy}$	
Total	$(s_{Ax} +...+ s_{Cy})'\, (s_{Ax} +...+ s_{Cy})$	χ^2_{Class}	Overall Agreement

test. While a single $\mathscr{L}^*$ suffices for the now classic problem of French and American expert wine connoisseurs ranking wine vintages, presented in all the three articles, in the 3-schools-by-3-methods example with data on evaluation of reading-teaching presented by Beckett & Schucany (1979) concordance is found only within and between two schools and the analysis hence proceeds in a different direction (the data from the two concordant schools are combined for the purpose of multiple comparisons between methods, while the third group is discarded from further analysis as there is no unanimous judgement within this internally discordant group regarding teaching methods).

The solution we propose for testing the analogue of an ANOVA factor effect is an intuitive appraisal of the expected value and variance of the sum of i $\mathscr{L}$ values of a factor, which provides the necessary elements for a *z*-test (i.e., a factor $\mathscr{L}^*$):

$$\mathrm{E}(\mathscr{D}) = \Sigma_i \ \mathrm{E}(\mathscr{L}_i) \tag{10}$$

$$\mathrm{var}(\mathscr{D}) = [\Sigma_i \ \mathrm{var}(\mathscr{L}_i) \ / \ N] + \mathrm{var}(\mathrm{E}(\mathscr{L}_i)) \tag{11}$$

This seems a justifiable procedure given the mutual independence of $\mathscr{L}$'s and the approximately normal distribution of every $\mathscr{L}$. Another (probably excessively conservative) possibility might be applying the Bonferroni inequality within each factor to correct the alpha-level requested for significance of individual $\mathscr{L}^*$ values. Yet another possibility would be to test factor effects on the basis of $\mathscr{L}$ values obtained from independent comparisons from $\mathscr{L}$ values, whereby there would be *j*-1 such comparisons for a factor with *j* levels (e.g., for 9 levels of the group factor in our experiment, group 1 vs. groups 2 to 9 combined, group 2 vs. groups 3 to 9 combined, etc.). Then, expected value and variance would be appraised as above, but to us this approach seems contradictory to the reasoning of Schucany and associates, plus there are different possibilities to form the comparisons and the choice between them would be arbitrary.

The results of ANACONDA applied to our experimental data are summarised in table 5, whereby the analysis of variation within groups has been omitted, since it has already presented in table 3 in the form of *W*'s. There is agreement between and within each pair of groups, which is not inconsistent with the notion of peer-groups formation on the basis of norms. Agreement can also be observed when testing all the groups "at once" (factor effect of group). No difference has been found between the externally and internally recognised norms, while the absence of interaction between group membership and ranking condition indicates that none of the groups appears to practice "double standards". The overall agreement (*W*=0.53) is self-evident given the other results.

It should be noted that our data might have been validly analysed without resorting to ANACONDA. A possible approach could be to treat rule rankings as interval scale and perform a mixed-model ANOVA with group (between-subjects),

Table 5. Summary of two-way ANACONDA results.

Source of Variation	Sum of Products or E($\mathscr{D}$) and var($\mathscr{D}$)	Test Statistic	Interpretation
Group 1 vs. Group 2	$(s_{11} + s_{12})' (s_{21} + s_{22})$		Agreement Within and Between Groups 1 and 2
… (34 comparisons)	…	all $\mathscr{L}^* > 50$	…
Group 9 vs. Group 8	$(s_{91} + s_{92})' (s_{81} + s_{82})$		Agreement Within and Between Groups 9 and 8
All Groups 1 … 9	E($\mathscr{D}$) = 1096134 var($\mathscr{D}$) = 4019158	$\mathscr{L}^* = 79.2$	Agreement Within and Between Groups
Condition 1 vs. *Condition 2*	$(s_{11} +...+ s_{91})' (s_{12} +...+ s_{92})$	$\mathscr{L}^* = 667.6$	Agreement Within and Between Conditions
Group 1 *Condition 1* vs. Group 2 *Condition 2*	$s_{11}' s_{22}$		No Interaction Between Group 1 *Cond. 1* and Group 2 *Cond. 2*
… (70 comparisons)	…	all $\mathscr{L}^* > 5$	…
Group 9 *Condition 1* vs. Group 8 *Condition 2*	$s_{91}' s_{82}$		No Interaction Between Group 9 *Cond. 1* and Group 8 *Cond. 2*
All Interactions	E($\mathscr{D}$) = 1111133 var($\mathscr{D}$) = 13069835	$\mathscr{L}^* = 38.4$	No Interactions
Total	$(s_{11} +...+ s_{92})' (s_{11} +...+ s_{92})$	$\chi^2 = 541.8$	Overall agreement

condition and rule (both within-subjects) as factors. Subsequently, group might be dropped from the model (as it would have most probably turned out not to have notable effect) to obtain simpler and hence more powerful tests of the possible condition effect and rule-condition interactions. Focusing on the condition effect, an even simpler analysis can be based solely on mean rule ranks, which, as expected, results in high correlation (r=0.764, p=0.004; ρ=0.772, p=0.003). Nevertheless, we believe that analysis of agreement is the most natural approach to the data from experiments of the type we performed, and it might eventually provide scientific insight in addition to pedagogical benefit.

Distribution of concordance coefficient in the non-null case

Usually, the significance of inter-group concordance is considered only in terms of departure from the null hypothesis of no agreement (equal likelihood of all permutations of ranks for every judge). While the usual χ^2-approximation is only useful for testing W under the null hypothesis of random rankings, the distribution of concordance statistics in the non-null case is a much less explored issue. The first thorough study of this subject was published by Wood (1970), who presented a variance stabilising transformation for W, ρ, W_τ and τ. Next, Palachek and Schucany (1984) used studentised U-statistic estimator of u_Q in the non-null case to obtain approximate confidence intervals for the strength of concordance. They also refined the

Table 6. Approximate confidence intervals for two selected concordance coefficients.

Group	Average	95 % c.i.	Estimated *df*	N
Class (*Condition 1*)	0.528	[0.262, 0.795]	35.762	47
Group 2 (*Condition 1*)	0.786	[0.667, 0.905]	3.356	5

τ-approximation by estimating the degrees of freedom using jack-knife.

We applied their procedure, but as precise strength of concordance within different significantly concordant groups was not central to our investigation, we present only two examples of confidence intervals obtained in this way in table 6. It is worth noting that in spite of the seemingly small difference in the point-estimate of population concordance and the fact that the *p*-value of *W* for the class is about 10^{-50} (as opposed to about 10^{-5} for the chosen group), comparison of 95% confidence intervals between the class and the chosen group (chosen having the fourth largest *W* and the largest average ρ) under *Condition 1* provides a nice illustration of the *similis simili gaudet* principle.

Conclusion

An attempt has been made to demonstrate the usefulness of the presented experimental procedure and associated data-analysis methods. It has been shown how the association between norm congruency and formation of peer groups, as well as the difference between the private and the public aspect of norms, can be studied by analysing agreement within and between several groups of judges ranking several behaviour rules. Multi-group analysis of concordance was applied, so an overview of the interrelations between different concordance measures had to be presented first, and the fundamentals of the little known ANACONDA procedure had to be explained. An overall (factor) test of between-and-within-groups agreement in ANACONDA has been proposed and its possible drawbacks and alternatives have been discussed. Lastly, an illustration of computation of confidence intervals for concordance coefficient in the non-null case has been presented.

A problem preventing wider usage of two-group concordance analysis and ANACONDA, and especially the usage of procedure for estimating confidence intervals for u_Q, lies in the fact that these procedures are neither included in any statistical software package, nor can a specialised program for performing them be found on the Internet. However, the computations for two-group concordance analysis and ANACONDA are fairly simple, and they can be easily programmed in a general-purpose spreadsheet or in any major statistical software package. On the other hand, the procedure for estimating confidence intervals for u_Q (Palachek and Schucany, 1984) is rather complex, and its detailed description or software implementation for

public use is beyond the scope of our paper, so we refer the interested researchers to the original paper.

Within a broader historical framework of statistics based on ranked data, one can observe a first nonparametric boom in the late fifties, represented by Siegel's (1956) classic text, for which a major incentive was the desire to reduce the extensive and time-consuming calculations required by the advanced parametric methods of the époque. A second wave of prosperity of ordinal methods came some two decades later: in addition to the "nonmetric revolution" (multidimensional scaling and subsequent nonlinear multivariate methods), it is exemplified by Conover's (1980) and Meddis' (1984) textbooks and the application of ANOVA to ranked data (Hora & Conover, 1984). Finally, the last decade might be called the era of resampling-based statistics (especially bootstrap), developing together with ever growing computer power. In this sense, our exposition fits into the framework of revived interest in ordinal statistical methods, in spite of the fact that a most prominent example of this renaissance (Cliff, 1996) does not mention multi-way concordance analysis.

We believe that adequate attention must be paid to analysis of ordinal and/or ranked data in the statistics and methodology curricula in the social sciences and elsewhere. Just as certain pitfalls must be identified (e.g., ranking interval or pseudo-interval variables with few distinct values), the wealth of procedures available must be presented, ranging from explicitly confirmatory to strictly exploratory ones. On the one hand the methodology of structural equation modelling has fully mastered ordinal manifest variables (Jöreskog, 2002, 2004), and on the other hand the developments within the data mining and knowledge discovery field have further accelerated the growth of the body of visualisation methods appropriate for ordinal data (cf. Friendly, 2000). While the multitude of available approaches, together with the plenitude of associated software and divergent literature, can only confuse a researcher less proficient in statistics, one should remember that the analysis of "awkward" or "exotic" data often contributes crucially to the progress of science.

References

Beckett, J. & Schucany, W.R. (1975). ANACONDA: Analysis of concordance of g groups of judges. In *Proceedings of Social Statistics Section of the American Statistical Association* (pp. 311-313).

Beckett, J. & Schucany, W.R. (1979). Concordance among categorized groups of judges. *Journal of Educational Statistics*, *4* (2), 125-137.

Bečaj, J. (1995). *Temelji socialnega vplivanja [Fundamentals of social influence]*. Ljubljana: Univerza v Ljubljani, Filozofska fakulteta.

Cliff, N. (1996). *Ordinal methods for behavioral data analysis*. Mahwah, NJ: Lawrence Erlbaum.

Conover, W.J. (1980). *Practical nonparametric statistics* (2nd ed.). New York: Wiley.

Coenders, G., Satorra, A. & Saris, W.E. (1997). Alternative approaches to structural modeling of ordinal data: a Monte Carlo study. *Structural Equation Modelling*, *4* (4), 261-282.

Friendly, M. (2000). *Visualizing categorical data*. Cary: SAS Institute.

Hora, S.C. & Conover, W.J. (1984). The F statistic in the two-way layout with rank-score transformed data. *Journal of the American Statistical Association*, *79* (387), 668-673.

Jöreskog, K.G. (2002, revised 2004). *Structural equation modelling with ordinal variables using LISREL* [WWW document]. URL www.ssicentral.com/lisrel/ordinal.pdf

Li, L. & Schucany, W.R. (1975). Some properties of a test for concordance of two groups of rankings. *Biometrika*, *62* (2), 417-423.

Meddis, R. (1984). *Statistics using ranks*. Oxford: Basil Blackwell.

Palachek, A.D. & Schucany, W.R. (1984). On approximate confidence intervals for measures of concordance. *Psychometrika*, *49* (1), 133-141.

Schucany, W.R. & Frawley, W.H. (1973). A rank test for two group concordance. *Psychometrika*, *38* (2), 249-258.

Siegel, S. (1956). *Nonparametric statistics*. New York: McGraw-Hill.

Siegel, S. & Castellan, J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

Smith, P. B. (1988). Norms and roles in the small group. In G. M. Breakwell, H. Foot & R. Gilmour (Eds.), *Doing Social Psychology* (pp. 291-308). Leicester: British Psychological Society.

Turner, J. C. (1991). *Social Influence.* Buckingham: Open University Press.

Wood, J.T. (1970). A variance stabilizing transformation for coefficients of concordance and for Spearman's rho and Kendalls's tau. *Biometrika*, *57* (3), 619-627.