A Comparison of Different Approaches to Hierarchical Clustering of Ordinal Data

Aleš Žiberna, Nataša Kejžar, and Petra Golob[1]

Abstract

The paper tries to answer the following question: "How should we treat ordinal data in hierarchical clustering?" The question is strongly connected to the use of questionnaires with ordinal scales in the social sciences. The results could help to differentiate among answers to the questions from questionnaires that could be considered as scale variables, those it would be better to convert to ranks and those that should be treated as nominal variables. To make the results general several two-dimensional combinations of group sizes, shapes and differences between their centers were used as well as one three-dimensional combination. Each combination was simulated both with and without unessential variables.

All datasets consisted of 3 groups, each with its own multivariate distribution (2 or 3 variables) with known means and covariances.

From each design several datasets were simulated. Each variable was cut and recoded to achieve an ordinal scale. Different cutting schemes were used (the intervals were of equal size, either increasing/decreasing from the lowest to the highest value or decreasing from the mean to both extremes). These new variables were then treated as interval, converted to ranks and treated as nominal. Then hierarchical clustering algorithms were used. Ward's algorithm with Squared Euclidean distance was used when data were considered interval or converted to ranks, and Ward's algorithm with matching coefficient as dissimilarity measure was used when they were considered nominal.

The quality of the results was assessed by comparing the gained partitions with the three original groups. We also compared results from clustering the original (uncut) data with the three original groups for comparison. The comparison was made using Corrected Rand Index. The results indicate that in most cases treating the data as interval or converting them to ranks yields better results than treating them as nominal, but the differences are sometimes diminished when cutting into a smaller number of intervals.

---

[1] University of Ljubljana, Slovenia

1   Introduction

The paper tries to answer the following question: "How should we treat ordinal data in hierarchical clustering?" Measurements in the social sciences are often done with an ordinal scale, usually using questionnaires with categorical scales of 3 to 9 categories. The results of these measurements are ordinal data. However, many statistical methods are not suitable to handle this kind of data.

One of them is hierarchical clustering. The problem when dealing with ordinal data in hierarchical clustering lies in how to compute the dissimilarity matrix (distances among units). Dissimilarity matrix is the input of several different clustering algorithms. We used Ward's algorithm (Ward, 1963), one of the most commonly used clustering algorithms. The different procedures we used thus differ only in the way the dissimilarity matrix is computed, which mainly depends on the assumptions about the nature of the data (interval, ordinal, nominal).

The purpose of this paper is to determine which of the different procedures used outperforms the others and under which circumstances. To achieve this, all procedures were tested on a number of two-dimensional data designs (a combination of group covariance matrices and sizes of three groups), distances among population means and number of unessential variables.

50 repetitions of each data design combination (which was cut with respect to different parameters to produce ordinal data) were made in order to obtain more reliable results.

The simulation was also done on a specific three-dimensional data design. In a similar manner to the two-dimensional designs we simulated three groups, each from a different three-dimensional normal distribution.

2   Simulation of the data

In order to test different clustering procedures, we first had to simulate the data on which we could apply these procedures. As mentioned, we started by simulating three groups, each simulated from a different two or three-dimensional normal distribution. These groups were simulated 50 times – that is, the whole process of simulation, clustering and assessment was repeated 50 times for each setting (data design combination). The three different distances among population means for data design 1 are presented in Figure 1. The same three sets of distances among population means were used in combination with all two-dimensional data designs. The remaining two-dimensional data designs (2 to 7) are shown in Figure 2. Only the first set of distances among population means is shown. The black dots in the graphs represent an example of an actual simulation of the data setting, while gray

dots represent a result of a simulation where the number of cases in each group was multiplied by 100. All of these combinations were simulated with zero, one and two unessential variables.
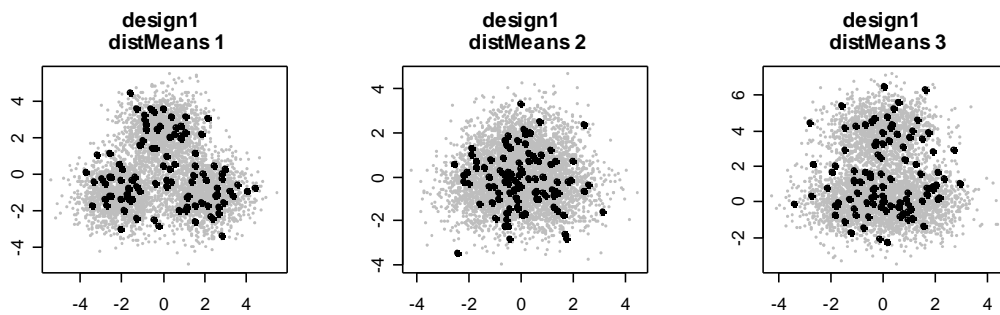
design1
distMeans 1

design1
distMeans 2

design1
distMeans 3

Figure 1: Data design 1 with different sets of distances among population means.

design2
distMeans 1

design3
distMeans 1

design4
distMeans 1

design5
distMeans 1

design6
distMeans 1
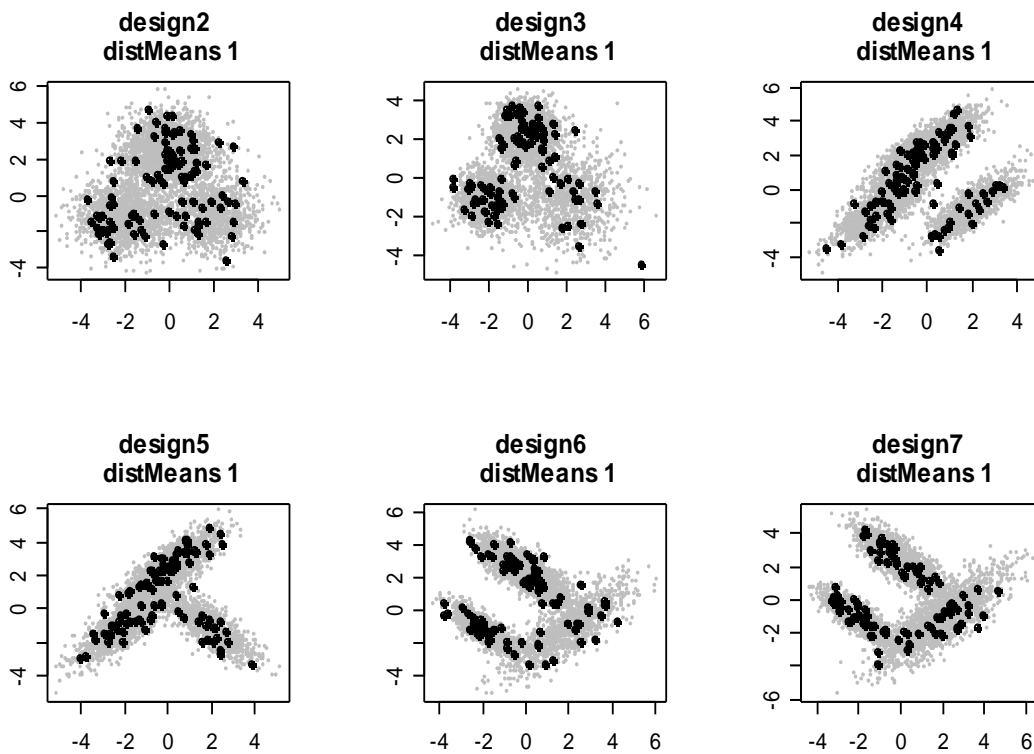
design7
distMeans 1

Figure 2: Data designs 2 to 7 with distances between population means 1.

The design and distances among group means for the three-dimensional case is shown in Table 1. In Figure 3 we can see an example of the simulated data, one of the 50 cases. As can be seen, the groups differ in their centers or mean vectors, in

their covariance matrices and in size (number of units). They also slightly overlap. These groups were simulated with unessential variables – zero, one and three.[2]

Table 1: Design for the simulated groups.

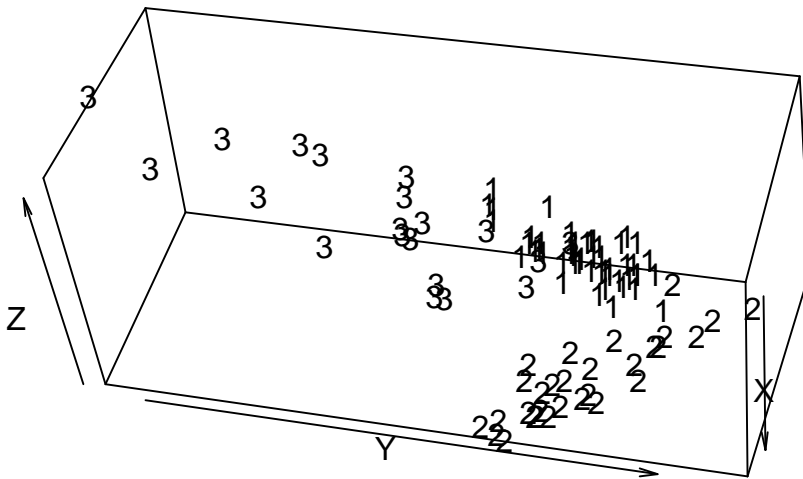| Group number	μ (vector of population means)	Σ (population covariance matrix)	Number of units
1	(-5,1,2)	$\begin{bmatrix} 6,25 & 9 & 0 \\ 9 & 20,25 & 0 \\ 0 & 0 & 4 \end{bmatrix}$	50
2	(8,2.1,-2)	$\begin{bmatrix} 16 & -17 & -5,04 \\ -17 & 25 & 5,25 \\ -5,04 & 5,25 & 4,41 \end{bmatrix}$	30
3	(1,-20,12)	$\begin{bmatrix} 9 & -10,8 & 1,8 \\ -10,8 & 144 & -57,6 \\ 1,8 & -57,6 & 36 \end{bmatrix}$	20

Figure 3: An example of the three-dimensional simulated data. The numbers indicate group membership.

---

[2] Here the largest number of unessential variables is 3. It equals the number of essential variables in the dataset.

3 Cutting the interval data into categories (ordinal data)

Though we described the process of simulating interval data, the purpose of this paper is to test several procedures of hierarchical clustering on ordinal data. With this in mind, we transformed interval data into ordinal data by cutting the variables into categories in several different ways. In this process we varied the following parameters:

- Number of categories (noClass)

- Type of cutting (typeCut)

- Extension factor (f)

Let us first describe these parameters.

3.1 Number of categories

The first parameter provides the number of categories into which values of the original variables were recoded. The values of the original variables were cut into intervals (their number matches the number of categories). All values in the same interval were then recoded into the same value, namely the sequential number of the interval (e.g. values within the interval with the lowest values were recoded into 1). In our simulation, we recoded the interval variable into 3, 4 and 5 categories. In all cases, values of both or all three interval variables used were recoded into the same number of categories, the values of each variable entered separately. Cutoff points differed from variable to variable, but the parameters were the same.

3.2 Type of cutting

This parameter shows how the direction in which the widths of the interval increase was determined. We used the following types of cutting:

- *no-sayers*: The values of the original interval variables are cut in a way such that the first interval is the widest and each subsequent interval is narrower – this corresponds to asymmetrically formed categories (more categories at the end of the scale) or the answers of subjects who tend to disagree.

- *yes-sayers*: Opposite to the first type (the first interval is the narrowest and each subsequent interval is wider) – this corresponds to asymmetrically formed categories (more categories at the beginning of the scale) or the answers of subjects who tend to agree.

- *average*: The widest interval is at the center and intervals narrow in both directions from the center – this corresponds to a scale with wide categories at the center (more categories at the extremes of the scale) or the answers of subjects who tend to try to blend in.
- *simple*: All intervals are of equal width – this corresponds to symmetrically formed categories or subjects who answer "correctly".
- *mix*: We randomly chose one of the types of cutting above for each unit based on the probabilities that we assumed[3] to represent the portion of subjects from a population that would answer in such a way – this would correspond to a real sample from a population answering questions with symmetrically formed categories. We assumed the following probabilities:
    - P(*yes-sayers*) = P(*no-sayers*) = 0.14
    - P(*average*) = 0.29
    - P(*simple*) = 0.43.

The effects of type of cutting on the borders of the intervals are seen in Figure 4. The figure presents the case with extension factor 2 and 5 categories. There are four different lines split into intervals. The value above each interval is that which would be assigned to all cases within that interval for the variable recoded at a given time. The type of cutting is also indicated.
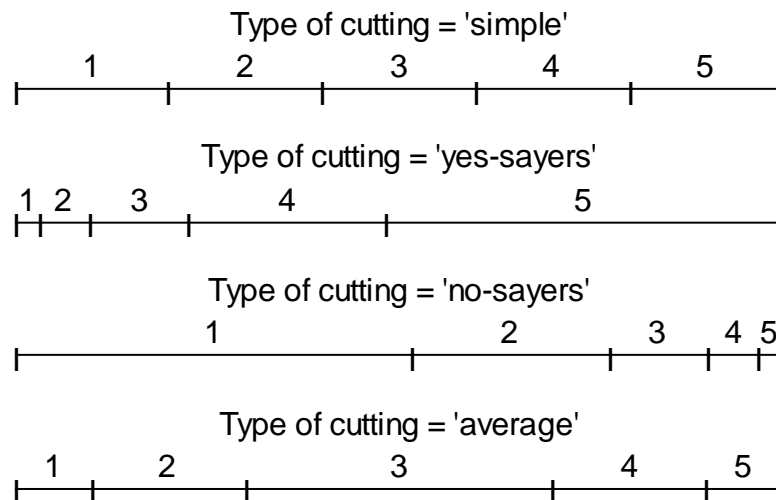

Figure 4: The effect of the type of cutting.

---

[3] The assumed probabilities were based on the results of the research "Drobno gospodarstvo v Sloveniji" by Professor Dr. Janez Prašnikar of University of Ljubljana (Faculty of Economics) and review of the "Slovensko javno mnenje" (Toš et. Al., 1999).

3.3 Extension factor

This parameter determines how much wider each subsequent interval is than the previous one, looking in the direction defined by the type of cutting. f is actually the parameter of the geometric sequence (e.g. if the starting interval is of width i, then the next is of width $i*f$). If $f = 1$, then all intervals are of equal width irrespective of type of cutting.
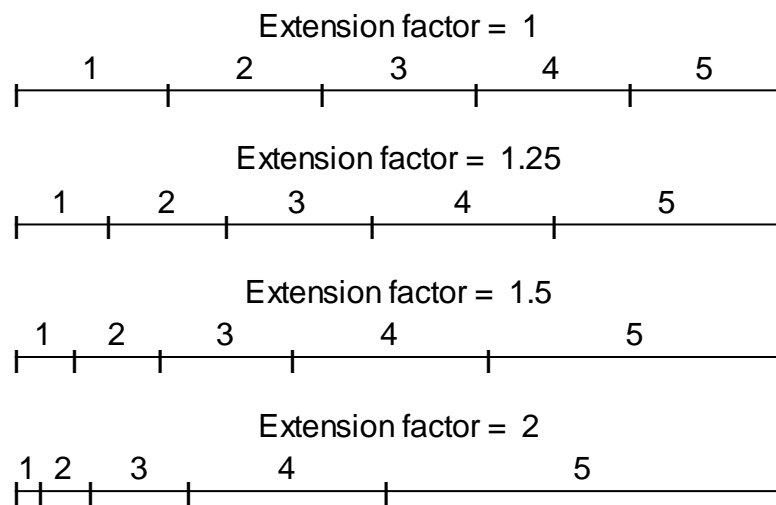

Figure 5: The effect of the extension factor.

The effects of the extension factor on the interval borders are seen in Figure 5. In the figure the case with type of cutting *yes-sayers* and 5 categories is presented. Again there are four different lines split into intervals. Above each interval we can see the value that would be assigned to all cases within that interval for the variable recoded at a given time. The extension factor is indicated above the values.

4 Computation of the dissimilarity matrices

The usual input of the hierarchical clustering algorithm is the dissimilarity matrix. In this paper we treated data in three different ways (as interval, as ordinal and as nominal), although we know they are ordinal. These different treatments influenced the procedures for computing dissimilarity matrices. Actually this is the only part of the computations that differs from procedure to procedure.

4.1   Interval data

When treating data (ordinal variables, which were the result of the recoding described in the previous section) as interval, we used these transformed (recoded) data and computed the dissimilarity matrix by Squared Euclidean distance:

$d(X,Y) = \Sigma(x_i - y_i)^2$, where X and Y represent two different units and x_i and y_i their values at the ith variable.

4.2   Ordinal data

When treating data as ordinal[4], we first computed ranks for values of each (transformed) variable respectively and then computed the dissimilarity matrix on these ranks by Squared Euclidean distance (as before):

$d(X,Y) = \Sigma(x_i - y_i)^2$, where X and Y represent two different units and x_i and y_i their values at the ith variable.

4.3   Nominal data

When treating data as nominal, our dissimilarity measure was simply the number of variables with different values, a matching coefficient:

$$d(X,Y) = \sum_i \delta_i; \quad \delta_i = \begin{cases} 1, x_i = y_i \\ 0, x_i \neq y_i \end{cases}$$

X and Y again represent two different units and x_i and y_i their values at the ith variable.

5   Hierarchical clustering procedure

The dissimilarity matrices were used as an input in the Ward hierarchical clustering algorithm (Ward, 1963). Numerous empirical comparisons have shown that Ward's method is the most suitable method for spherical data (e.g. Everitt, 1974; Mojena, 1978; Ferligoj and Batagelj, 1980). Like most hierarchical methods it is based on the consecutive merger of two groups into a new group (the other hierarchical method work in the opposite direction). By Ward's algorithm, dissimilarity measure between the new group, which was formed by a merger of groups C_i and C_j and group C_k, is computed by the following formula:

$$d(C_i \cup C_j, C_k) = \frac{(n_i + n_j)n_k}{(n_i + n_j + n_k)} d^2(T_{i,j}, T_k),$$

---

[4] We are not stating that this is the only or the best way for treating ordinal data, but this technique is widely used when computing correlation for ordinal variables.

- C_i, C_j, C_k – groups or cases
- $T_{i,j}$ – the center of the new group $C_i \cup C_j$
- T_k – the center of the group C_k.
- n_i, n_j and n_k – the number of cases in groups C_i, C_j and C_k.

(Ferligoj, 1989).

However, since we performed our analysis on the basis of dissimilarity matrices, this formula can not be used. The appropriate formula is based on the Lance Williams recurrence formula and here the dissimilarity between a newly joined group and previous groups can be computed solely on the basis of the dissimilarity matrix. The Lance Williams recurrence formula for Ward method is:

$$d(C_i \cup C_j, C_k) = \frac{(n_i + n_k)}{(n_i + n_j + n_k)} d(C_k, C_i) + \frac{(n_j + n_k)}{(n_i + n_j + n_k)} d(C_k, C_j) - \frac{n_k}{(n_i + n_j + n_k)} d(C_i, C_j)$$

(Everitt et. al., 2001, p. 61 and 63)

The method is also monotonous according to the Lance Williams formula, which means it always logically reveals the structure of the data.

6 Measure of quality of the results of clustering procedures

We decided to measure the quality of the results derived by different procedures of hierarchical clustering (the procedures differ in the method of dissimilarity matrix computation) by comparing the results of these procedures to the original group membership. These groups were already introduced in section 2 (Simulation of the data) and are defined by the multivariate distribution from which they were simulated.

However, the result of clustering procedure is a dendrogram. A tree structure shows which cases, groups or clusters were joined in each step of the clustering algorithm. To get the result which would be comparable with the original group membership we extracted the partition, a part of the dendrogram with the same number of clusters or groups as the original number of groups.

For the measure of similarity of gained partitions with the original groups we chose Corrected Rand Index (Hubert, 1977). Rand Index (Rand, 1971) is one of most widely used measures for comparing partitions (Hubert in Arabie, 1985). It takes values on the interval [0,1], where 1 means total agreement and 0 no agreement. It is based on the comparison of pairs of units.

To calculate the Rand Index, we formed a R×C contingency table. There are $\binom{n}{2}$ pairs, each belonging to one of the following types:

1. Objects in the pair are placed in the same group in both partitions.
2. Objects in the pair are placed in different groups in both partitions.

3. Objects in the pair are placed in different groups in the first partition and in the same group in the second partition.
4. Objects in the pair are placed in the same group in the first partition and in different groups in the second partition.

Pairs of type 1. and 2. can be described as agreements (A) and pairs of type 3. and 4. as disagreements (D). A represents the total number of agreements and D the total number of disagreements:

$$A + D = \binom{n}{2}.$$

The Rand Index can be written as:

$$Rand = \frac{A}{\frac{n(n-1)}{2}}.$$

However, we used the Corrected Rand Index (CRand) instead. As the name indicates, it is in fact the Rand index, corrected for chance. The Corrected Rand Index is computed using the following formula:

$$CRand = \frac{Rand - E(Rand)}{Max(Rand) - E(Rand)}.$$

This index takes values on the interval [-1,1], where 1 again means total agreement and 0 means that agreement between the two partitions is as good as can be expected by chance (in cases where the group membership of each unit was chosen randomly).

7   Results

Up to this point, we have described the simulation of the data, the transformation performed on them, computation of the dissimilarity matrices, clustering algorithm and, in the previous section, the measure used to assess the quality of results of clustering procedure. The values of this measure represent the main results of this research.

For the analysis of the data, analysis of variance (Rice, 1995; Faraway, 2002) and graphical representation were used. We first analyzed two-dimensional data settings and then compared the findings with three-dimensional design.

Analysis of variance was used to compare the means of the different settings and treatments of the data.

The main interest is to determine the effect of data treatment (with respect to other factors) on the quality of classification. Data designs (designs 1 to 7), distances among population means and number of unessential variables can not be influenced or even known prior to the data collection and analysis. Since determining them is often the aim of the analysis, they are considered only separately (as additive factors).

Our basic model was:

CRand = μ + a_i + b_j + interaction effect (a, b) + c_k + d_l + e_m + f_n + g_o + interaction effects (d, e, f, g) + ε

where *a, b,...,g* are factors occurring at levels:

$i = 1,...,3$ levels for distance of population means (distMeans)
$j = 1,...,7$ levels for data design – covariance matrices (design)
$k = 1,...,3$ levels for number of unessential variables added
$l = 1,...,3$ levels for number of categories (noClass)
$m = 1,...,4$ levels for type of cutting (typeCut)
$n = 1,...,4$ levels for extension factor (f)
$o = 1,...,3$ levels for data treatment (varType)

We assumed that the errors (ε) are independent and identically normally distributed with mean 0 and variance σ^2. We used a least squares approach (we fitted the linear model) to estimate the effects of factors. We fitted the model on only 10 repetitions (due to restrictions in computer memory).

Table 2: Analysis of variance table for the basic model.

	Df	Sum of Squares	Mean Square	F
DistMeans	2	1022.0	511.0	26331.3
Design	6	95.1	15.9	816.8
unessential variables	2	186.2	143.1	7375.1
NoClass	2	23.6	11.8	606.7
TypeCut	3	74.3	24.4	1258.5
F	3	58.8	19.6	1009.4
VarType	2	517.6	258.8	13335.0
distMeans:design	12	11.8	1.0	50.7
noClass:typeCut	6	24.4	4.1	209.2
noClass:f	6	9.1	1.5	78.0
typeCut:f	9	50.8	5.7	291.0
noClass:varType	4	36.3	9.1	467.1
typeCut:varType	6	5.9	1.0	50.7
f:varType	6	52.7	8.8	453.0
noClass:typeCut:f	18	15.6	0.9	44.7
noClass:typeCut:varType	12	4.8	0.4	20.7
noClass:f:varType	12	12.9	1.1	55.3
typeCut:f:vartype	18	5.8	0.3	16.7
noClass:typeCut:f:varType	36	4.5	0.1	6.5
Residuals	90554	1747.3	0.02	

Table 2 shows an analysis of variance table for the basic model. Adjusted R^2 for this model was 0.57. It can be seen that the interaction effects are present but are smaller than the main effects. The two largest main effects explain about 39% of total variation. All other main effects and interactions in the model explain only about 17% of variance.

The largest effects in the model are based on difference of population means and data treatment. From the coefficients of the linear model it can be seen that the larger the population means' distance, the lower the quality of classification. The overall effect of data treatment suggests that we treat ordinal data as interval or at least as ranks. There is also a large effect due to added unessential variables. The more unessential variables added, the lower the quality of classification.

Graphical presentation of results can be seen in Figure 4, with the four graphs presenting the results of four types of cutting respectively.
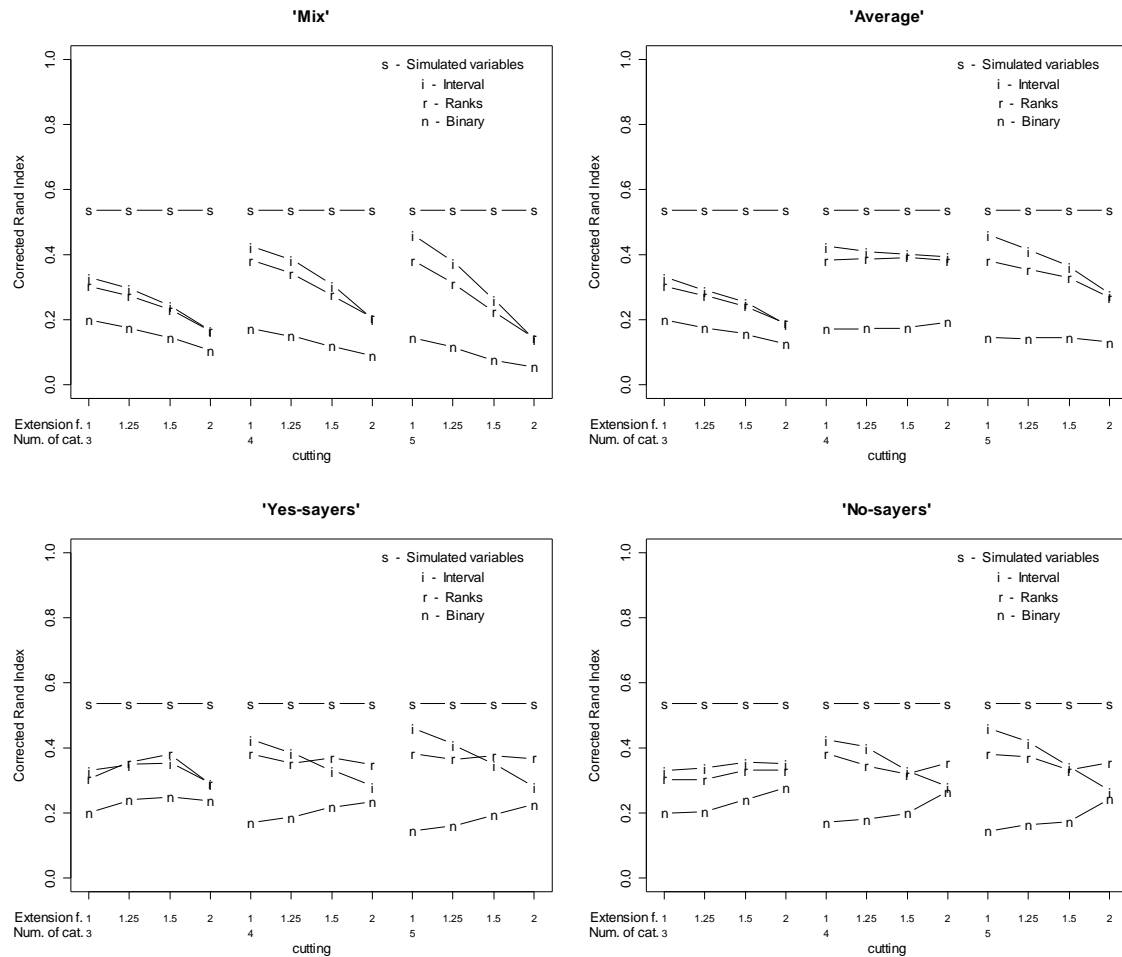

Figure 6: The quality of results of different procedures.

All graphs show the average Corrected Rand Index (on the y axis) for all 50 repetitions for all two-dimensional data settings. The results are averaged over all data designs, sets of distances and number of unessential variables. The results for

individual settings are not shown due to the large number of different combinations, but they were examined and are referred to where applicable.

Corrected Rand Index measures the similarity of a partition, which is gained after a specific transformation of the data and after a specific procedure for clustering the data was used. The parameters of the transformations of the original, simulated variables into ordinal variables can be seen in two places. The title of the graph denotes the type of cutting used. The other two parameters (extension factor and number of categories) that influenced this transformation are indicated at the x axis.

The characters indicate the treatment of the data according to the legend in the top right corner of the graph:

s – Simulated variables: The original simulated variables (before the transformation into ordinal variables) were put through the same procedure – dissimilarity matrix computed using Squared Euclidean distance and clustering done by Ward's method. Since the variables in this procedure were not subjected to the transformations, different cutting options (or parameters) had no effect and therefore the Corrected Rand Index is always the same for them. These results are shown only for the comparison, in order to simplify the estimation of the effect of the transformation. In principle[5], better results could not be expected after any transformation, since this procedure retained the most data.

i – Interval: The results are based on having treated the ordinal data as interval (as described in subsection 4.1)

r – Ranks: The results are based on having treated the ordinal data as ordinal (as described in subsection 4.2), which was done by computing ranks and then performing on them the same analysis as on interval data.

n – Nominal: The results are based on having treated the ordinal data as nominal (as described in subsection 4.3).

It can be seen from all graphs that all treatments of the data as well as all cutting options have an effect on the quality of the result. Furthermore, the treatment of the ordinal data as interval and the treatment where we compute ranks on these ordinal data yield similar results, which are in all cases better than those gained by treating the data as nominal. Since treating the data as nominal is obviously an inappropriate approach, we further focused exploration and interpretation of the results on the remaining two treatments (as interval and ordinal data).

---

[5] It could happen by "chance" if the borders of intervals used in the cutting (or recoding) procedure were to line up with the borders of our groups.

7.1   Focus on the results of treating data as interval and transformation to ranks

Observing these two data treatments from the graphs, we noticed that the quality of the results decreases with the increase of the extension factor. When the extension factor is small, an increase in the number of categories improves the quality of the results. However, when the extension factor increases, these effects are more or less diminished.

Since treating data as nominal yields much poorer results compared to treating them as ranks or interval, we decided to eliminate them. We fitted the basic model on the new set of data. Adjusted R^2 for this data was 0.55. The results can be seen in Table 3.

Table 3: Analysis of variance table for the basic model
(data treatments: interval and rank).

	Df	Sum of Squares	Mean Square	F
distMeans	2	963.2	481.6	22204.2
design	6	69.1	11.5	531.3
unessential variables	2	252.4	126.2	5817.6
noClass	2	48.9	24.4	1126.3
typeCut	3	44.6	14.9	686.0
f	3	98.6	32.9	1514.5
varType	1	6.7	6.7	307.3
distMeans:design	12	11.8	1.0	45.2
noClass:typeCut	6	25.8	4.3	198.3
noClass:f	6	15.5	2.6	119.4
typeCut:f	9	30.9	3.4	158.6
noClass:varType	2	0.4	0.2	10.0
typeCut:varType	3	3.1	1.0	47.3
f:varType	3	12.31	4.1	189.2
noClass:typeCut:f	18	16.1	0.9	41.1
noClass:typeCut:varType	6	1.3	0.2	9.9
noClass:f:varType	6	4.9	0.8	37.6
typeCut:f:vartype	18	3.4	0.4	17.5
noClass:typeCut:f:varType	36	2.2	0.1	5.7
Residuals	90554	1309.2	0.02	

It can again be seen that most of main effects are larger than interaction effects. The largest effect in the model is based on difference of population means. Data treatment effect is not so strong anymore. It explains about 0.2% of variance compared to 13% when nominal data were included. Its interaction with extension

factors accounts for more Sum of Squares (variance) than its main effect. All other main effects are stronger. Unessential variables (as in the previous model) are the second strongest effect. Extension factor and different number of classes then follow. The more distorted the groups are, the lesser the quality in classification, and the larger the number of classes, the better the quality of the classification.

For the special three-dimensional data design we fitted a similar anova model (excluding distance of population means and data designs, since we had just one):

$$CRand = \mu + c_k + d_l + e_m + f_n + g_o + \text{interaction effects (d, e, f, g)} + \varepsilon$$

where $c,...,g$ are factors occurring at levels:

$k = 1,...,3$ levels for number of unessential variables added
$l = 1,...,3$ levels for number of categories (noClass)
$m = 1,...,4$ levels for type of cutting (typeCut)
$n = 1,...,4$ levels for extension factor (f)
$o = 1,...,3$ levels for data treatment (varType)

The results were very similar to the ones described above. The main difference was that the effect of number of unessential variables was much smaller. Possible reasons for that could be a larger number of variables describing a group and a special design of population means and covariance matrices. This special data design could also explain the smaller effect of the number of categories.

7.2 Comparison of the results of treating data as interval and of transforming them into ranks

After reviewing these results, the following question arises: "Is it better to treat ordinal variables as interval or to transform them into ranks?" Unfortunately, the answer is too complicated to be answered in a brief manner. It was observed that transforming data ranks produced better results in most scenarios when the transformed variables are quite far away from interval scale. By this we mean that the extension factor is large or the number of categories is 3 and the type of cutting is set to *yes-sayers* or *no-sayers*, which is the usual result of asymmetrical scales (wider categories at one of the extremes). Of course, this differs from one individual setting to another, but as a general rule it can be said that on average the result of treating data as ranks compared to treating them as interval improves when we move away from the interval scale.

On the other hand, treating data as interval produces better results in the remaining scenarios: when the type of cutting is *average* or *mix* and number of categories is 5 or 4 or extension factor is not very large (approximately below 1.5). These settings are very important, since we can assume that the *mix* type of cutting

is the most realistic assumed type for a general social survey (with well designed questions, which do not impose asymmetry). It is likely that the respondents of a survey will be a mixture of people with different "answering habits" (*yes-sayers*, *no-sayers*, *average*, *simple*). Furthermore, 5 is one of the most common numbers of categories in scales found in social surveys. We also believe that in well-designed surveys extreme extension factors are quite rare (e.g. extension factor 2 and *no-sayers* type of cutting would mean that more than one half of the units has the extreme value).

Finally, we should point out that although we estimated the effects of different data treatments on a number of different data designs, there is no guarantee that these results hold in general or even more for a specific dataset. Even in our datasets the effects of treatments were sometimes quite different, mainly due to "chance" cutting.

8   Conclusion

In this research we dealt with the following question: "How should we treat ordinal data in hierarchical clustering?". We can state that ordinal data in hierarchical clustering should be either treated as interval or converted to ranks, not as nominal or converted to a set of binary variables, as long as the original ordinary variables have at least 3 categories. Technically our results can be applied only to the data designs used in our simulations. However, since we assessed the treatment effects on a large variety of different data designs, we believe that our results can be used to select an appropriate treatment for most datasets. The results for our specific clustering procedures (Ward's method, Squared Euclidean distance) have clearly shown this (large difference in quality of the results); we believe that the suggestion is also useful for similar procedures.

Unfortunately, we can not point to only one of the two treatments suggested in the previous paragraph. There are some differences in the quality of the results gained by the two treatments, which were described in more detail in subsection 7.2, but no general conclusion can be reached.

To sum up, ordinal data in hierarchical clustering should not be treated as nominal, no matter how far from interval data they are. However, the question whether it is better to treat them as interval or to convert them to ranks remains unanswered.

References

[1] Everitt, B.S. (1974): *Cluster Analysis*. London: Heinemann Educational Books.

[2] Everitt, B.S. et. al. (2001): *Cluster Analysis*. Fourth Edition. London: Arnold.

[3] Faraway, J.J (2002): *Practical Regression and Anova Using R*. URL: www.stat.lsa.umich.edu/~faraway/book.

[4] Ferligoj, A. (1989): Razvrščanje v skupine, Teorija in uporaba v družboslovju. *Metodološki zvezki,* **4**, Ljubljana.

[5] Ferligoj, A. and Batagelj, V. (1980): Taksonomske metode v družboslovnem raziskovan`ju. *Poročilo RSS*, Ljubljana: RIFSPN.

[6] Hubert, L.J. (1977): Nominal Scale Response Agreement as a Generalized Correlation. *British Journal of Mathematical and Statistical Psychology*, **30**, 98-103.

[7] Hubert, L. in Arabie, P. (1985): Comparing Partitions. *Journal of Classification*, **2**, 193-218.

[8] Leenen, I. in van Mechelen, I. (2001): An Evaluation of Two Algorithms for Hierarchical Classes Analysis. *Journal of Classification*, **18**, 57-80.

[9] Milligan, G.W. (1981): A Review of Monte Carlo Tests of Cluster Analysis. *Multivaritate Behavioral Research*, **16**, 379-407.

[10] Mojena, R. (1977): Hierarchical Grouping Methods and Stopping Rules, An Evaluation. *Computer Journal*, **20**, 359-363.

[11] Rand, W.M. (1971): Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association*, **66**, 846-850.

[12] Rice, J. A. (1995): Mathematical Statistics and Data Analysis, Second Edition, Belmond: Duxbury Press.

[13] Toš, N. et. Al. (1999): Vrednote v prehodu I in II, Slovensko javno mnenje 1990-1998. Ljubljana.

[14] Ward, J. H. (1963): Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, **58**, 236-244.