Categorising Count Data into Ordinal Responses with Application to Ecological Communities

D. Fernández [1*], S. Pledgera [1]

[1]School of Mathematics, Statistics and Operations Research, Victoria University of Wellington, Wellington, New Zealand

Rec: 29 September 2015, Accp: 22 November 2015,

Abstract

Count data sets may involve overdispersion from a set of species and underdispersion from another set which would require fitting different models (e.g. a negative binomial model for the overdispersed set and a binomial model for the underdispersed one). Additionally, many count data sets have very high counts and very low counts. Categorising these counts into ordinal categories makes the actual counts less influential in the model fitting, giving broad categories which enable us to detect major ~~broadly-based~~ broadly based patterns of turnover or nestedness shown by groups of species. In this paper, a strategy of categorising count data into ordinal data was carried out and also we implemented measures to compare different cluster structures. The application of this categorising strategy and a comparison of clustering results between count and categorised ordinal data in two ecological community data sets are shown. A major advantage of using our ordinal approach is that it allows for the inclusion of all different levels of dispersion in the data in one methodology, without treating the data differently. This reduction of the parameters on modelling different levels of dispersion does not substantially change the results in clustering structure. In the two data sets used in this paper, we observed ordinal clustering structure up to ~~93.1%~~ 93.1 % similar to those from the count data approaches. This has the important implication of supporting simpler, faster data collection using ordinal scales only.

Supplementary materials accompanying this paper appear on-line.

Keywords: Cluster analysis, Clustering measures, ~~EM-algorithm~~ EM algorithm, Finite mixture model, Ordinal data, Stereotype model

ArticleNote

Electronic Supplementary Material
Supplementary materials for this article are available at 10.1007/s13253-015-0240-3.

Introduction

Count Data

One of the most common types of data recorded is a count of the number of times an event occurs, for example, the number of a particular species at a certain site. The values are non-negative integers, with the zero value being included or not depending on whether it is ecologically important. The counts may have no upper bound, or may have a known maximum (as in a binomial or multinomial distribution of *n* objects over different categories). In this article, unbounded count data are considered.

Rogers (**1974**, Chapter 1) describes a stochastic scheme for classifying count data in relation to its variance-mean ratio. When this ratio is equal to unity, i.e. the variance is equal to the mean, the dispersion of the data relative to a predefined study region follows a *random* point pattern (a Poisson process). On the other hand, if the data have a variance-mean ratio greater than unity, i.e. \(\hbox {variance}>\hbox {mean}\) (overdispersion), this indicates a more *clustered* (e.g. spatial or temporal clustering) than random point pattern. Finally, if the data ~~has~~ have a variance-mean ratio less than unity, i.e. \(\hbox {variance}<\hbox

{mean}\) (underdispersion), the point pattern is more likely to result from a more *regular* than random or clustered process. In the case of count data distributed as a *random* point pattern, the dispersion is expected to follow a Poisson distribution as the variance of this distribution is equal to its mean. Rogers (**1974**, Chapter 2) derives the densities under linearity assumptions detailed below when the dispersion follows a clustered or regular pattern. This is determined from a random point pattern resulting in a negative binomial distribution when the dispersion is clustered and in a binomial distribution when the dispersion follows a regular pattern. In the case of a clustered point pattern (\(\hbox {variance}>\hbox {mean}\), negative binomial distribution), the probability of an object settling in a quadrat is positively linearly related to the number of objects already there (e.g. shoal of sardines). If this probability is completely independent of the number of objects already in a quadrat (e.g. plants with well-dispersed seeds) then the point pattern is random (\(\hbox {variance}=\hbox {mean}\), Poisson distribution). In a regular point pattern (\(\hbox {variance}<\hbox {mean}\), binomial distribution), the probability of an object settling in a quadrat decreases linearly with the number of objects already there (e.g. gannet nests in a colony or songbirds establishing territories). The mean-variance relationship is a critical property of count data. When not properly controlled for, trends in location (mean abundance) may be confounded with changes in dispersion (variance), leading to misleading results (Warton et al. **2012**). This is a major problem, and one way to deal with the variance-mean ratio problem is to turn the count data into ordinal data.

The detection of patterns in matrices of count data was considered by Pledger and Arnold (**2014**) using the Poisson distribution. Both single-mode clustering (row clustering only or column clustering only) and biclustering (simultaneous clustering of rows and columns) were done using finite mixtures. This gave rise to model-based analogues of correspondence analysis, multidimensional scaling, association analysis, ordination, biplots and other methods in multivariate analysis. However, for some count data ~~sets~~ sets, the assumption of a Poisson distribution is unrealistic, and in Hui et al. (**2014**) the negative binomial distribution was found to be more appropriate for multidimensional scaling of a matrix of 28 sites by the species composition of 12 spider species (the same data set used below in the example in ~~Section~~ Sect. **3.1**). Each species was given its own dispersion parameter, to allow for different degrees of spatial clustering. The question arises of whether a species-specific grouping into ordinal data (e.g. zero or a low, medium or high count for that species) would provide essentially the same information about overall clustering and association patterns.

There are several possible problems which could arise from using count data. Firstly, one of the causes of overdispersion in count data is the presence of outliers. Secondly, count data ~~is~~ are often supplied from data sets that structurally exclude zero counts (e.g. hospital length of stay data set). Thirdly, a more frequent situation is count data having an excess of zero counts which are far more that the expected zero counts under NB or Poisson distributional assumptions (e.g. number captured from spatially rare or hard to detect species). Lastly, as we described in ~~Section~~ Sect. **1**, the binomial distribution is a useful model to use when count data ~~has~~ have underdispersion . The difficulty in this scenario is the estimation of the number of trials parameter. Although different models for count data may be fitted depending on the data features, a good alternative is to recode the data into ordinal scale to fit our ordinal model approach. For instance, compared with the count data distributions, an ordinal variable is less sensitive to the presence of outliers and is not affected by the omission of zeros or large number of count outcomes in the data.

The main aims of this paper are to show the advantages of categorising count data into ordinal data in different situations such as when the count data set has extreme values or involves overdispersion, and to compare clustering results between count and categorised ordinal data. The plan of the article is as follows. The methodology of our strategy for determining the optimal number of ordinal categories using likelihood-based models for matrices of ordinal data is ~~i s~~ presented in ~~Section~~ Sect. **2**. Additionally, this Section presents three measures to compare clusterings from count and ordinal data methods over the same data set: the adjusted Rand index, the ~~normalized~~ normalised variation of information and the ~~normalized~~ normalised information distance. The results of clustering comparison in two real data ~~s e t~~ sets from community ecology are given in ~~Section~~ Sect. **3**, and we conclude with a discussion in ~~Section~~ Sect. **4**.

Methodology

In this section, we introduce the mixture likelihood-based models built on the ordinal stereotype model to define our clustering approach ~~(Section~~ (Sect. **2.1**). Furthermore, the strategy to categorise the count data

into ordinal outcomes is developed in ~~Section~~ Sects. 2.2, and ~~Section~~ 2.3 introduces the measures to compare different clustering structures.

Ordinal Stereotype Model Approach

The extension of the likelihood-based models proposed in Pledger and Arnold (2014) for an $n \times m$ data matrix with ordinal data was considered by Fernández et al. (2014a). This approach also considered finite mixtures to define a fuzzy clustering and used the ordered stereotype model introduced by Anderson (1984) in order to formulate the ordinal approach. The ordered stereotype model including row clustering, column clustering or biclustering for the probability that y_{ij} takes the category *k* that is ~~characterized~~ characterised by the following log ~~odds~~odds:

- Row clustering

MediaObjects/13253_2015_240_Equ5.gif

$$\begin{aligned} \begin{aligned}&{\mathrm {log}} \left(\frac{P\left[y_{ij}=k \mid i\in r \right]}{P\left[y_{ij}=1 \mid i\in r \right] }\right) =\mu_{k}+ \phi_{k}(\alpha_{r} + \beta_{j} + \gamma_{rj}),\\&k=2,\ldots ,\ell , \qquad r=1,\ldots ,R, \qquad j=1,\ldots ,m. \end{aligned} \end{aligned}$$

- Column clustering

MediaObjects/13253_2015_240_Equ6.gif

$$\begin{aligned} \begin{aligned}&{\mathrm {log}} \left(\frac{P\left[y_{ij}=k \mid j\in c \right]}{P\left[y_{ij}=1 \mid j\in c \right] }\right) =\mu_{k}+ \phi_{k}(\alpha_{i} + \beta_{c} + \gamma_{ic}),\\&k=2,\ldots ,\ell , \qquad i=1,\ldots ,n, \qquad c=1,\ldots ,C. \end{aligned} \end{aligned}$$

- Biclustering

MediaObjects/13253_2015_240_Equ7.gif

$$\begin{aligned} \begin{aligned}&{\mathrm {log}} \left(\frac{P\left[y_{ij}=k \mid i\in r, j\in c \right]}{P\left[y_{ij}=1 \mid i\in r, j\in c \right] }\right) =\mu_{k}+ \phi_{k}(\alpha_{r} + \beta_{c} + \gamma_{rc}),\\&k=2,\ldots ,\ell , \qquad r=1,\ldots ,R, \qquad c=1,\ldots ,C, \end{aligned} \end{aligned}$$

where the inclusion of the monotone increasing constraint $0=\phi_{1} \le \phi_{2} \le \dots \le \phi_{\ell }=1$ ensures that the variable response $\varvec{Y}$ is ordinal ~~(see~~ [see Anderson (1984~~)).~~)]. The parameters $\{\mu_{2},\ldots ,\mu_{\ell }\}$ are the *cut points*, and $\{\phi_{2},\ldots ,\phi_{\ell }\}$ are the parameters which can be interpreted as the "scores" for the categories of the response variable y_{ij}. The sets of parameters $\{\alpha_{1},\ldots ,\alpha_{n}\}$ and $\{\beta_{1},\ldots ,\beta_{m}\}$ quantify the main effects of the *n* rows and *m* ~~columns~~ columns, respectively, and the set $\{\gamma_{11},\ldots ,\gamma_{nm}\}$ ~~are~~ is the associations between the different rows and columns. We restrict $\mu_{1}=\phi_{1}=0$, $\phi_{\ell }=1$, $\sum_{i=1}^{n}\alpha_{i}=\sum_{j=1}^{m}\beta_{j}=0$ and we impose sum-to-zero constraints on each row and column of the association (or pattern detection) matrix γ to ensure identifiability. $R\le n$ is the number of row groups, $C\le m$ the number of column groups, $i\in r$ means row *i* is classified in the row cluster *r* and $j\in c$ means column *j* is classified in the column cluster *c*. It is important to note that the actual membership of the rows among the *R* row-clusters and the columns among the *C* column-clusters is unknown and, therefore, it is considered as missing information. Further, we define $\{\pi_{1},\ldots ,\pi_{R}\}$ and $\{\kappa_{1},\ldots ,\kappa_{C}\}$ as the (unknown) proportions of rows and columns in each row and column ~~g r o u p~~ g r o u p , respectively, with $\sum_{r=1}^{R}{\pi_{r}}=\sum_{c=1}^{C}{\kappa_{c}}=1$. We can view π_{r} and κ_{c} as the a priori row and column membership probabilities.

One of the most common uses of the EM algorithm (Dempster et al. 1977; McLachlan and Krishnan 1997) is in the case of the estimation of the parameters for a finite mixture-density model with incomplete data which in this case is the actual unknown cluster membership of each row and/or column. This method performs a fuzzy assignment of rows/columns to clusters based on the posterior probabilities after likelihood ~~maximization.~~ maximisation. In this paper, we have fitted a suite of clustering models using the EM algorithm, and the information criterion AIC was computed for each model to select the best models.

An advantage of using the stereotype model in our ordinal approach is the interpretation of the score parameters \(\{\phi_k\}\). If two ordinal categories have the same (or very similar) score parameter values, this provides evidence that those ordinal categories are not distinguishable and we can collapse them into a single category in our data (Fernández et al. **2014a**, Section 1.2.2). It is useful to know into how many cuts (i.e. into how many ordinal categories) the data must be divided. The following subsection introduces a strategy to categorise count data based on the ordinal stereotype model.

How Many Ordinal Categories?

One of the questions arising from recoding count data into an ordinal scale is related to determining how many ordinal categories into which the data should be optimally categorised. We implemented the option of replacing the count data by their ranks, and then cutting the ranks into groups based on percentiles because percentiles are not strongly influenced by extreme values in the count data, and can be calculated even if the counts are skewed. Therefore, percentiles do not depend on the variance-mean ratio scheme of the count data. When recoding a matrix \(\varvec{Y}=\{y_{ij}\}\), one option is to recode across the whole count data set with the chosen criterion. However, it may be more appropriate to analyse count data sets where the columns (or rows) have a dramatically different count pattern. For instance in an ecological community, a data set of abundance of species (columns) by sites (rows) might have a set of species with a ~~high~~ high-count pattern because they are either numerous or easily detectable species, whereas other species (more difficult to observe) have a ~~low count~~ low-count pattern. In order to standardise the species, a recoding strategy where the columns are recoded separately should be taken.

Given a \(n\times m\) matrix \(\varvec{Y}\) of count data, our strategy to categorise \(\varvec{Y}\) is as follows:

1. Start by setting a large number of ordinal categories \(\ell \) (e.g. \(\ell =10\)).
2. Rescale each observation \(y_{ij}\) \((i=1,\ldots ,n\), \(j=1,\ldots ,m)\)~~as:~~ as
MediaObjects/13253_2015_240_Equ8.gif
$$\begin{aligned} y^{\mathrm{st}}_{ij}=\frac{y_{ij}-{\mathrm{min}}(\varvec{Y}_{j})}{\mathrm{max}(\varvec{Y}_{j}) -{\mathrm{min}}(\varvec{Y}_{j})}, \end{aligned}$$
where \(\varvec{Y}_{j}\) (\(j=1,\ldots ,m)\) is the column vector. After this step, we have a new ~~standardized~~ standardised \(n \times m\) data matrix \(\varvec{Y}^{\mathrm{st}}=\{y^{\mathrm{st}}_{ij}\}\) which lies ~~on~~ in the range [0, 1].
3. Divide each new column vector \(\varvec{Y}^{\mathrm{st}}_{j}\) into \(\ell +1\) quantiles: \(Q^{(0)},\ldots ,Q^{(\ell)}\).
There is a number of equivalent ways of defining the sample quantiles. However, the sample quantiles used in statistical packages in common use such as **R** are all based ~~in one or two order~~ on one- or two-order statistics, and can be written ~~as:~~ as
MediaObjects/13253_2015_240_Equ1.gif
$$\begin{aligned} Q^{(k)}= \left\{ \begin{array}{ll} 0 &{} \qquad {\mathrm{if}}\,\,k=0,\\ \\ (1-\varphi)y^{\mathrm{st}}_{(i)j}+\varphi y^{\mathrm{st}}_{(i+1)j} &{} \qquad {\mathrm{if}}\,\,k=1,\ldots ,\ell -1,\\ \\ 1 &{}\qquad {\mathrm{if}}\,\,k=\ell ,\\ \end{array} \right. \end{aligned} \quad (1)$$
where \(\varphi =nk+s-j\), \(s=\frac{1}{3}(k+1)\), \(j=\lfloor kn+s \rfloor \) is the floor function for \(kn+s\) (i.e. the largest integer not greater than \(kn+s\)), and \(y^{\mathrm{st}}_{(i)j}\) denotes the \(i\hbox {th}\) order statistics of the column vector \(\varvec{Y}^{\mathrm{st}}_{j}\)~~(see~~ [see Hyndman and Fan (**1996**) for more ~~details).~~details].
4. Recode each observation \(y^{\mathrm{st}}_{ij}\) \((i=1,\ldots ,n\), \(j=1,\ldots ,m)\)~~as:~~ as
MediaObjects/13253_2015_240_Equ2.gif
$$\begin{aligned} y'_{ij}= \left\{ \begin{array}{ll} 0 &{} \qquad {\mathrm{if}}\,\, y^{\mathrm{st}}_{ij}=0,\\ \\ k &{} \qquad {\mathrm{if}}\,\, y^{\mathrm{st}}_{ij}>0 \,\, {\mathrm{and}}\,\,y^{\mathrm{st}}_{ij} \in (Q^{(k-1)},Q^{(k)}], \\ \end{array} \right. \end{aligned} \quad (2)$$
where \((Q^{(k-1)},Q^{(k)}]\) is the interval of values from vector \(\varvec{Y}^{\mathrm{st}}_{j}\) between the ~~\((k-1)^{\mathrm{th}}\)~~ \((k-1){\mathrm{th}}\) and ~~\(k^{\mathrm{th}}\)~~ \(k{\mathrm{th}}\) quantiles, for \(k=1,\ldots ,\ell \). Each interval contains \(\frac{100}{\ell ~~)\%~~}\,\%\) of the non-zero data.
As a result of this step, we obtain an ordinal view \(\varvec{Y'}\) of the original data set \(\varvec{Y}\). A graphical illustration of the recoding from count data \(y_{ij}\) into ordinal responses \(y'_{ij}\) based on the quantiles is given in Figure 3 in Web Appendix A.
5. Fit our ordinal mixture methodology to \(\varvec{Y'}\)~~(Section~~ (Sect. **2.1**).
6. If two or more adjacent categories have the same score parameter value, collapse them, set the new number of ordinal categories \(\ell \) and return to step 2. Otherwise, the categorisation is appropriate

and returns the results of model fitting.

Note that we ~~standardize~~ standardise the original count data with the aim of reducing the number of quantiles to calculate in the step 3. Thus, we need to calculate only $((\ell +1)$ quantiles for the whole data set $\varvec{Y}^{\mathrm{st}}$, instead of $m\times (\ell +1)$ quantiles (i.e. $\ell +1$ quantiles for each column in $\varvec{Y}$). However, this ~~standardization~~ standardisation might not work suitably for some data sets (e.g. when there is no variation in a column and so the maximum and minimum values in that column are the same) and other strategies can be used. For instance, computing the $\ell +1$ quantiles for groups of columns. Additionally, we may wish to directly assign zero values from the original count data into a particular category in the ordinal scale ~~(see eq.~~ [see Eq. (2~~)).~~)]. The reason for this procedure is related to the particular meaning of the zero value in some data sets such as ecological community data regarding species abundance, where it is important to keep absences separated from presences. However, this category could be removed and ~~equation~~ Eq. (2) would simply turn into

MediaObjects/13253_2015_240_Equ9.gif

$$\begin{aligned} y'_{ij}= \left\{ \begin{array}{ll} 0&{}\qquad {\mathrm{if}} \,\, y^{\mathrm{st}}_{ij} \in [Q^{(0)},Q^{(1)}],\\ \\ k &{}\qquad {\mathrm{if}} \,\, y^{\mathrm{st}}_{ij} \in (Q^{(k)},Q^{(k+1)}], \\ \end{array} \right. \end{aligned}$$

for $k=1,\ldots ,\ell -1$. Finally, this strategy was presented on categorising throughout columns but the same idea might be applied over the rows just exchanging columns for rows above.

Comparing Clusterings. Definition of Measures

Let *Y* be a data set of *N* observations, then $\varvec{U}=\{U_1,\ldots ,U_K\}$ is a partition, or clustering, of *Y* into *K* groups, where $\bigcup_{k=1}^K U_k=Y$ and $U_i \cap U_j=\varnothing$ for $i\ne j$ (non-overlapping). Equivalently, $\varvec{V}=\{V_1,\ldots ,V_{K'}\}$ on *Y* is an alternative of clustering *Y* into K' groups. The information on the overlap between these two clusterings $\varvec{U}$ and $\varvec{V}$ can be summarised in the form of a $K\times K'$ contingency table as illustrated in Table 1.

Given two clusterings $\mathbf{U}$ and $\mathbf{V}$, the following quantities are defined via the marginal and the joint distributions of data items in $\mathbf{U}$ and $\mathbf{V}$ ~~respectively~~, respectively, as Vinh et al. (2010):

MediaObjects/13253_2015_240_Equ3.gif

$$\begin{aligned} \begin{aligned} {\mathrm{H}}(\mathbf U)&= -\sum_{i=1}^{K} \frac{a_i}{N}\log \left(\frac{a_i}{N}\right) , \qquad \qquad \,\, \text{(Entropy } \text{ for } \mathbf U)\\ {\mathrm{H}}(\mathbf V)&= -\sum_{j=1}^{K'} \frac{b_j}{N}\log \left(\frac{b_j}{N}\right) , \qquad \qquad \text{(Entropy } \text{ for } \mathbf V)\\ {\mathrm{H}}(\mathbf U ,\mathbf V)&= -\sum_{i=1}^{K}\sum_{j=1}^{K'} \frac{n_{ij}}{N}\log \left(\frac{n_{ij}}{N}\right) , \qquad \text{(Joint } \text{ entropy } \text{ for } {} \mathbf U \hbox { and }{} \mathbf V) \\ {\mathrm{I}}(\mathbf U ,\mathbf V)&= \sum_{i=1}^{K}\sum_{j=1}^{K'} \frac{n_{ij}}{N}\log \left(\frac{n_{ij}/N}{a_i b_j/N^2}\right) \\&={\mathrm{H}}(\mathbf U)+{\mathrm{H}}(\mathbf V)-{\mathrm{H}}(\mathbf U ,\mathbf V). \qquad \text{(Mutual } \text{ information } \text{ for } {} \mathbf U \hbox { and }{} \mathbf V) \end{aligned} \end{aligned} \quad (3)$$

We use three measures in common use to compare clusterings: the adjusted Rand Index (ARI, Hubert and Arabie ~~(~~1985~~)),~~) , the variation of information (VI, Meila ~~(~~2005~~)),~~) and the ~~normalized~~ normalised information distance (NID, Kraskov et al. ~~(~~2005~~)).~~). The ARI is a pair counting-based measure developed from the Rand index (Rand 1971) and corrected for chance as suggested by Hubert and Arabie (1985). The ARI remains the most well-known and widely used measure to compare clusterings. For instance, Žiberna et al. (2004) used this measure to compare clusterings for ordinal data. The formulation of the ARI from Table 1 is as ~~follows,~~follows:

MediaObjects/13253_2015_240_Equ10.gif

$$\begin{aligned} {\mathrm{ARI}}(\mathbf U ,\mathbf V)= \frac{\sum_{i=1}^{K}\sum_{j=1}^{K'} \left({\begin{array}{c}n_{ij}\\ 2\end{array}}\right) - \frac{\left[\sum_{i=1}^{K} \left({\begin{array}{c}a_i\\ 2\end{array}}\right) \right] \left[\sum_{j=1}^{K'} \left({\begin{array}{c}b_j\\ 2\end{array}}\right) \right] }{\left({\begin{array}{c}N\\ 2\end{array}}\right) }}{\frac{1}{2}\left[\sum_{i=1}^{K} \left({\begin{array}{c}a_i\\ 2\end{array}}\right) + \sum_{k=1}^{K'} \left({\begin{array}{c}b_j\\ 2\end{array}}\right) \right] -\frac{\left[\sum_{i=1}^{K} \left({\begin{array}{c}a_i\\ 2\end{array}}\right) \right] \left[\sum_{j=1}^{K'} \left({\begin{array}{c}b_j\\ 2\end{array}}\right) \right] }{\left({\begin{array}{c}N\\ 2\end{array}}\right) }}. \end{aligned}$$

This measure is bounded above by 1, and a 0 value indicates independent clusterings.

An alternative to pair counting-based measures ~~are~~ is information theoretic-based distance measures. They

are based on the relationship between an observation from *Y* and its cluster in each of the two clusterings that are compared. Based on the quantities defined in (3), the VI for clustering $\mathbf U$ and $\mathbf V$ is formulated as

MediaObjects/13253_2015_240_Equ11.gif

$$\begin{aligned} {\mathrm{VI}}(\mathbf U ,\mathbf V)={\mathrm{H}}(\mathbf U ,\mathbf V)-{\mathrm{I}}(\mathbf U ,\mathbf V) ={\mathrm{2H}}(\mathbf U ,\mathbf V)-{\mathrm{H}}(\mathbf U)-{\mathrm{H}}(\mathbf V). \end{aligned}$$

This measure is bounded between 0 and $(\log (N))$. In order to bound it between 0 and 1, the ~~normalized~~ normalised VI (NVI, Kraskov et al. ~~(~~2005~~))~~) is defined, which consists of dividing $({\mathrm{VI}}(\mathbf U ,\mathbf V))$ by $({\mathrm{H}}(\mathbf U ,\mathbf V))$:

MediaObjects/13253_2015_240_Equ12.gif

$$\begin{aligned} {\mathrm{NVI}}(\mathbf U ,\mathbf V) =1-\frac{{\mathrm{I}}(\mathbf U ,\mathbf V)}{{\mathrm{H}}(\mathbf U ,\mathbf V)}. \end{aligned}$$

Another distance measure is the NID which is bounded between 0 and 1 and formulated as

MediaObjects/13253_2015_240_Equ13.gif

$$\begin{aligned} {\mathrm{NID}}(\mathbf U ,\mathbf V)=1-\frac{{\mathrm{I}}(\mathbf U ,\mathbf V)}{{\mathrm{max\{H}}(\mathbf U),H(\mathbf V)\}}. \end{aligned}$$

A zero value indicates that $\varvec{U}$ and $\varvec{V}$ are exactly the same clusterings and a value of one is interpreted as independent clusterings for both NVI and NID. Thus, we use the unit-complements of these measures (i.e. 1-NVI and 1-NID) in our comparisons in order to have the same scale interpretation between ARI, NVI and NID.

Application

In this section, clusterings from approaches for count and ordinal data are compared. The count data-based clusterings are obtained by applying the likelihood-based methodology described ~~in~~ by Pledger and Arnold (2014) for basic Poisson and NB building blocks, and the ordinal data-based clustering by applying Fernández et al. (2014a~~)~~)'s approach ~~(Section~~ (Sect. 2.1). Two real-life data sets are used to illustrate the comparison among clusterings. These data sets have small dimension. We have not found drawbacks in the application of our approach to larger data sets (e.g. thousands of sites and species) apart from the higher computational speed required. For the sake of increasing this speed, we compiled certain functions in **C** code and called them from **R**. Regarding model choice, we have used Akaike's Information Criterion (AIC, Akaike ~~(~~1973~~)).~~). Fernández et al. (2014a) set up a comprehensive simulation study and conclude that the AIC is the best information criteria when dealing with ordinal data and the likelihood-based finite mixture model with the stereotype model as the components in the mixture ~~is~~ are fitted. In general terms, the lower is the AIC value of a ~~model~~ model, the best is the fitting of this model for a data set.

Example 1: Spider Data Set

The spider abundance data set (Van der Aart and Smeenk-Enserink 1974) shows the distribution of 12 different hunting spider species across 28 different sites. We obtained the original count data set from the **R** package *mvabund* using the data set called "spider". The original count data ~~is~~ are given in Table 5 in Web Appendix B.1. Note the large number of zeros and also the high counts, suggesting the NB model is preferable to the Poisson model (Hui et al. 2014). Additionally, Figure 4 in Web Appendix B.1 depicts the mean and variance for all of the species of spiders throughout the sites. The variance is greater than the mean in all the species indicating possible overdispersion (the variance-mean ratio ranges from 6.8 to 65.4). We categorised the original data into four ordinal responses by following the strategy described above ~~(Section~~ (Sect. 2.2), setting:

MediaObjects/13253_2015_240_Equ4.gif

$$\begin{aligned} y_{ij}= \left\{ \begin{array}{ll} ~~\text{(0)}~~ \text {(0) }\mathbf{None } ~~\mathbf{None }~~ &{} \qquad ~~\text{ No } \text{ data } \text{ recorded }\\ ~~\text{(1) } \mathbf{Low } &{} \qquad ~~\text{ Species } \text{ coverage } \text{ is } \text{ below } 25\%\\ ~~\text{(2) } \mathbf{Medium } &{}\qquad ~~\text{ Species } \text{ coverage } \text{ is } \text{ between } 25\%-65\% \\ ~~\text{(3) } \mathbf{High } &{} \qquad ~~\text{ Species } \text{ coverage } \text{ is } \text{ higher } \text{ than } 65\%.\\ \text {No data recorded}\\ \text {(1) }\mathbf{Low } &{} \qquad \text {Species coverage is below }25\,\%\\ \text {(2) }\mathbf{Medium } &{}\qquad \text {Species coverage is between }25\text{-}65\,\% \\ \text {(3) }\mathbf{High } &{} \qquad \text {Species coverage is higher than }65\,\%.\\ \end{array} \right. \end{aligned} \quad (4)$$

The whole ordinal data set and a summary of the frequencies of spider abundance data for this new ordinal scale are shown in Table 6 in Web Appendix ~~B.1~~ B.1, respectively. All the categories have similar frequency (between 56 and 66 observations) apart from the first category, which is for sites and spider species without the presence recorded.

A suite of models was ~~fitted~~ fitted, and the information criteria measures were computed. The results are summarised in Table 7 in Web Appendix B.1. AIC indicates that the best model is the stereotype model version including row (sites) clustering with $R=3$ row groups (i.e. $\mu_{k}+\phi_{k}(\alpha_{r}+\beta_{j})$) with $\hbox{AIC}=397.28$. Each row is allocated to the group to which the site belongs with the highest posterior probability. The resultant row clustering setting is ${\mathrm{R}}1=\{1-7,13,14\}$, ${\mathrm{R}}2=\{8,21-24,27,28\}$, ~~a~~ n d ${\mathrm{R}}3=\{9-12,15-20,25,26\}$.

The scatter plot and histogram ~~(Figure~~ (Fig. 1) display the average fitted scores $\{\overline{\phi}_{(.j)}\}$ over the 28 sites, using a weighted average which accounts for the fitted spacings (Fernández et al. **2014a**).

Different colour and shape points and colour bars represent the resultant $R=3$ row (site) clustering settings. Three groups are distinguished in the scatter ~~plot~~ plot, and the histogram presents three clear modes. Since each ordinal response category k ~~$(k=0,\dots$~~ $(k=0,\ldots,3)$ is associated with a score parameter ϕ_{k}, the spacing between adjacent ϕ_{k} values shows us how similar or different categories are in terms of the effect of rows and columns ~~(see~~ [see Agresti (**2010**, Section 4.3.5.) and Fernández et al. (**2014a**, Section ~~5.2.2)).~~ 5.2.2)]. For this data set, the fitted score parameters were $\widehat{\phi}_{0}=0$, $\widehat{\phi}_{1}=0.39$, $\widehat{\phi}_{2}=0.89$ and $\widehat{\phi}_{3}=1$ (the end points being fixed at 0 and 1). Therefore, the distance between ordinal categories "Low" and "Medium" (0.50) is greater than that between categories "None" and "Low" (0.11) or categories "Medium" and "High" (0.39). This spacing is illustrated in the spaced mosaic plot (Fernández et al. **2014b**) in ~~Figure~~ Fig. 2.

Table 2 summarises all the ~~3~~ three clusterings. They are also ~~also~~ shown in Figure 5 in Web Appendix B.1.

For all the sites, the highest posterior probability stands out from the other two probabilities except for the sites 16, 17 and 19 (e.g. $\kappa_1=0.52$ and $\kappa_3=0.42$ for site 17). The clustering which allocates the sites 16, 17 and 19 to their ~~highest~~ highest, a posteriori probability cluster is thus not the only reasonable crisp clustering. For this ~~reason~~ reason, we make an alternative allocation ("Stereotype 2") which allocates site 17 to cluster R1 and sites {16, 19} to cluster R2 (whereas they had all been originally allocated to cluster R3). This ~~enable~~ enables us to test for the effect of the fuzziness when comparing clusterings. Furthermore, we obtained the count data-based clustering for 3 site groups for Poisson and NB building blocks, using the highest probability-based allocation criteria. Taking into account the "Stereotype 2" clustering, the results show that sites ~~$\{1-7, 13-20, 22-24, 27, 28\}$~~ 1–7, 13–20, 22–24, 27, 28 are classified into the same cluster for all three probability models. Sites 8 and 21 are allocated to group R2 according to the ordinal model and in group R3 according to the other two models. The opposite happens in site 26. The rest of the sites ~~($\{9-12, 25\}$)~~ (9–12, 25) are classified into a different cluster depending on the fitted model. These clustering structures show that the ordinal stereotype approach is closer to the NB approach than the Poisson approach, which is as expected given the overdispersion shown for the data. However, we want to compare the clustering not only graphically but also using the measures described in ~~Section~~ Sect. 2.3. The measures ARI, ~~NVI,~~ NVI and NID were computed for the three clusterings (Poisson, NB and ~~Stereotype)~~ Stereotype), and the results are summarised in the Table 3.

Large values of these measures indicate similarity of clustering. Furthermore, we computed an index which indicates the percentage of the Poisson vs. NB clustering explained by the clustering with ordinal data for each measure. For example, the ARI value for the Poisson vs. NB comparison is 0.555 and the ARI value for the NB vs. Stereotype comparison is 0.409. Therefore, the clustering structure with the stereotype approach explains a $\left(1-\frac{0.555-0.409}{0.555}\right) \times$ ~~100=73.7\%~~100=73.7\,\%\) of the count data clustering structure. For the three comparison measures, the Poisson and NB clusterings are the closest as it is expected. Between count and ordinal data-based models, the "Stereotype 2" clustering is closer to the NB clustering than the Poisson one. The clustering from the other ordinal data-based model ("Stereotype") is also closer to NB than ~~Poisson~~ Poisson, although less similar than the "Stereotype 2". We observed that the

"Stereotype 2" clustering structure is up to ~~84.2%~~ 84.2 % similar to those from the count data approach. The observed similarity between NB and stereotype clusterings is a satisfactory result because the data ~~is~~ are overdispersed suggesting that NB is preferred over Poisson.

Example 2: Urban Bird Data Set

The urban bird data set (Dolédec et al. **1996**) is a list of information about 40 bird species (columns) across 51 sites (rows). We obtained the original count data set from the **R** package *ade4* using the data set called "aviurba". As in the previous example, there is a large number of zeros ~~(76%~~ (76 % of the whole data set) but the difference with the spider data set is that the range of count data values is small, from 0 to 4, i.e. it is a data set with low counts. Additionally, Figure 6 in Web Appendix B.2 shows that the variance and the mean have similar values for almost all the species (the variance-mean ratio only ranges from 0.5 to 2.1) indicating that the point pattern is ~~random~~ random, and therefore a Poisson model is preferable. As this data set only ~~consist~~ consists of 5 possible count data values and with the aim of obtaining as more similar ordinal data set to the count data set as possible, we categorised the count values (0, 1, 2, 3, 4) into their corresponding five ordinal categories with labels $\{0,1,2,3,4\}$. However, models fitted to these data indicated that the corresponding estimated score parameters for the adjacent categories 3 and 4, $\widehat{\phi }_{3}$ and $\widehat{\phi }_{4}$, were very close to each other. This spacing is illustrated in the spaced mosaic plot in Figure 9 in Web Appendix B.2. This implies that the relative frequencies in these two categories are independent of the clustering structure. ~~Therefore~~ Therefore, retaining the distinction between 3 and 4 is not informative about the clustering structure ~~(see~~ [see Fernández et al. (**2014a**, Section 5.2.2) for more ~~detail).~~ details]. In that case, the model still holds with the same scores if the ordinal scale is collapsed by combining those two adjacent categories into one single response category. Table 8 in Web Appendix B.2 summarises the frequencies of urban bird data in the final ~~4-level~~ four-level ordinal scale.

Furthermore, a suite of models was fitted and a summary of the AIC results are in the bar plot depicted in Figure 7 in Web Appendix B.2. This bar plot is sorted by AIC and the model version is distinguished by different bar colours.

AIC indicates that the best model is the column effect model (i.e. $\mu_{k}+\phi_{k}\beta_{j}$). However, as we want to compare clustering structures, we select the stereotype model version including species clustering with $C=3$ column groups (i.e. $\mu_{k}+\phi_{k}\beta_{c}$) which is ~~labeled~~ labelled as $\{rR1,cC3\}$ in the bar plot and is the second best model for the data, according to AIC. The three species groups are distinguished in the scatter plot in Figure 8 in Web Appendix B.2.

In order to compare clusterings, we fitted the same model $\{rR1,cC3\}$ for $C=3$ species groups for Poisson and NB building blocks (i.e. $\mu +\beta_{c}$). Table 9 and Figure 10 in Web Appendix B.2 ~~summarize~~ summarise the clusterings structures for Poisson, NB and ordinal stereotype building blocks based on the highest posterior probability allocation criterion. We observe that our ordinal approach is closer to the Poisson approach. This is also confirmed in Table 10 in Web Appendix B.2 which shows the ARI, ~~1-NVI,~~ 1-NVI and 1-NID calculations for the three clusterings. The three measures indicate that the stereotype approach is closer to Poisson. For instance, according to the 1-NID measure, the ordinal clustering structure in comparison with the Poisson clustering structure is ~~93.1%~~ 93.1 % similar to the count data approaches. We observe that all three comparison measures between Poisson and stereotype clusterings result in low values. This is because each model defines the clustering structure differently. Table **4** shows the average of the species mean and the variance-mean for each column cluster in both models. We note that the three groups in the Poisson model have different species mean values, but that difference is not equally reflected in the variance-mean ratio ~~(i.e.~~ [i.e. 1.100 (Cluster 2) vs. 1.021 (Cluster ~~3)).~~ 3)]. However, the means are not as well separated in the stereotype model as in the Poisson model (particularly between groups C2 and C3), but the three clusters are very well distinguished across the variance-mean ratio. Therefore, the Poisson model is clustering based on the species mean across the sites, whereas the stereotype model is doing it based more on the species variance-mean ratio because this model can cluster data sets of all different levels of dispersion.

Discussion

We have shown some features of categorising count data into ordinal data. In our view, a major advantage is that by using our approach for ordinal data, we do not have to decide among different parametric models for

the data. It enables the inclusion of all of the different levels of dispersion in one methodology. For example, if a count data set involves overdispersion from one set of species and underdispersion from another set, probably the optimal strategy using the original data would be to fit a NB model for the overdispersed set and a binomial model for the underdispersed one. However, we may fit our ordinal stereotype methodology to both of these without treating the data differently. Additionally, many count data sets have extreme values, for ~~example~~ example, very high counts and very low counts in ecological community data. Replacing these counts with "high" and ~~"low" respectively~~ "low", respectively, ordinal categories makes the actual counts less influential in the model fitting, giving broad categories which enable us to detect major ~~broadly based~~ broadly based patterns of turnover or nestedness shown by group of species. These features in count data were illustrated in the two examples presented. The spider data set has large number of zeros, high counts, and also overdispersion is shown examining the variance-mean ratio. It suggests that the NB model is preferable. However, similar values for almost all the species in the urban bird data set suggest that the Poisson would be the best model. In both ~~examples~~ examples, we can fit our ordinal approach regardless the level of dispersion in the data. Thus, our ordinal data approach has similarities to non-parametric tests, being based on ranks, which are less susceptible to outliers. Our approach is an alternative analysis for researchers to consider.

The problem of changes in abundance being confounded with changes in dispersion (Warton et al. **2012**) was addressed in Hui et al. (**2014**) using complicated negative binomial models with latent variables or finite mixtures. Our approach has been to simplify the count data into ordered categories, allowing the data to dictate the spacings, which is a more non-parametric method. The statistical methodology is now available to evaluate the option of using ordinal rather than count data to identify broad overall patterns of species abundance. The saving in cost of sampling time in collecting only ordinal data (such as the Braun-Blanquet scale) may be justified by the benefits of being able to sample more sites and identify overall patterns more accurately.

Two future research directions may be set in order to investigate the differences between recoded and original count data. Firstly, setting up an empirical comprehensive study through numerical experiments across a wide range of scenarios. Secondly, developing a measure to quantify the loss of information due to use of the ordinal categorisation instead of the original count data.

Although our examples have been drawn from ecological communities, the methods in the article are widely applicable over many disciplines.

13253_2015_240_OnlinePDF.pdf

Electronic supplementary material

Below is the link to the electronic supplementary material. Supplementary material 1 (pdf ~~436~~ 529 KB)

References

Agresti, A. *Analysis of Ordinal Categorical Data*. Wiley Series in Probability and Statistics. Wiley, 2nd edition, 2010.

Akaike, H. Information theory and an extension of the maximum likelihood principle. In Petrov, B. N. and Csaki, F., editors, *2nd International Symposium on Information Theory*, pages 267–281, 1973.

Anderson, J. A. Regression and ordered categorical variables. *Journal of the Royal Statistical Society Series B*, 46(1):1–30, 1984.

Dempster, A. P., Laird, N. M., and Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

Dolédec, S., Chessel, D., Ter Braak, C. S. J., and Champely, S. Matching species traits to environmental variables: a new three-table ordination method. *Environmental and Ecological Statistics*, 3(2):143–166, 1996.

Fernández, D., Arnold, R., and Pledger, S. Mixture-based clustering for the ordered stereotype model. *Computational Statistics and Data Analysis*, 2014a. URL http://www.sciencedirect.com/science/article/pii/S016794731400317X.

Fernández, D., Pledger, S., and Arnold, R. Introducing spaced mosaic plots. Research Report Series. ISSN: 1174-2011. 14-3, School of Mathematics, Statistics and Operations Research, VUW, 2014b. URL http://msor.victoria.ac.nz/foswiki/pub/Main/ResearchReportSeries/TechReport_Spaced_Mosaic_Plots.

Hubert, L. and Arabie, P. Comparing partitions. *Journal of Classification*, 2(1):193–218, 1985.

Hui, F. K. C., Taskinen, S., Pledger, S., Foster, S. D., and Warton, D. I. Model-based approaches to unconstrained ordination. *Methods in Ecology and Evolution*, 2014.

Hyndman, R. J. and Fan, Y. Sample quantiles in statistical packages. *Statistical Computing*, 50(4):361–365, 1996.

Kraskov, A., Stögbauer, H., Andrzejak, R. G., and Grassberger, P. Hierarchical clustering using mutual information. *EPL (Europhysics Letters)*, 70(2):278–284, 2005.

McLachlan, G. J. and Krishnan, T. *The EM Algorithm and Extensions*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. John Wiley, 1997.

Meila, M. Comparing clusterings: an axiomatic view. In *ICML 2005: Proceedings of the 22nd International Conference on Machine Learning*, pages 577–584. ACM Press, 2005.

Pledger, S. and Arnold, R. Multivariate methods using mixtures: Correspondence analysis, scaling and pattern-detection. *Computational Statistics and Data Analysis*, 71:241–261, 2014.

Rand, W. M. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850, 1971.

Rogers, A. *Statistical Analysis of Spatial Dispersion: The Quadrat Method*. Monographs in Spatial and Environmental Systems Analysis. Pion, 1974.

Van der Aart, P. and Smeenk-Enserink, N. Correlations between distributions of hunting spiders (Lycosidae, Ctenidae) and environmental characteristics in a dune area. *Netherlands Journal of Zoology*, 25(1):1–45, 1974.

Vinh, N. X., Epps, J., and Bailey, J. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research*, 11(1):2837–2854, 2010.

Warton, D. I., Wright, S. T., and Wang, Y. Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution*, 3(1):89–101, 2012.

Žiberna, A., Kejžar, N., and Golob, P. A comparison of different approaches to hierarchical clustering of ordinal data. *Metodološki Zvezki - Advances in Methodology and Statistics*, 1(1):57–73, 2004.

13253_2015_240_Fig1_print.png

Fig. 1 Spider abundance data ~~set:~~ set: *Scatter plot* and histogram of the \(R=3\) fitted sites clusters \(\{\overline{\phi }_{(i.)}\}\) from the row clustering version of the stereotype model (\(\mu_{k}+\phi_{k}(\alpha_{r}+\beta_{j})\)).

13253_2015_240_Fig2_print.png

Fig. 2 Spaced mosaic plot with spacing for the \(R=3\) fitted spider site clusters from the row clustering version of the stereotype model (\(\mu_{k}+\phi_{k}(\alpha_{r}+\beta_{j})\)). The plot is divided into three horizontal bands over the ~~y-axis,~~ *y*-axis, one for each group, and four vertical bands over the ~~x-axis,~~ *x*-axis, one for each ordinal category. The fitted spacing \(\{\phi_{k}\}\) is depicted with different ~~color~~ colour blocks ~~(yellow,~~ (*yellow*, *red* and ~~blue)~~ *blue*) showing differences between two adjacent categories. We observe that ordinal categories 2 (“Medium”) and 3 (“High”) are close to each other ~~(blue band)~~(*blue band*).

Table 1 The contingency table for clusterings \(\mathbf U \) and \(\mathbf V \) on *Y* where \(n_{ij}\) is interpreted as the number of observations from *Y* that are common to clusters \(U_i\) and \(V_j\) (i.e. \(n_{ij}=|U_i \cap V_j|\)), \(a_i\) is the sum of row *i* (i.e. \(a_{i}=|U_i |\)), and \(b_j\) is the sum of column *j* (i.e. \(b_{j}=|V_j |\)).

\(\mathbf{U \backslash V}\)	\(V_1\)	\(V_2\)	\(\ldots \)	\(V_{K'}\)	Total
\(U_1\)	\(n_{11}\)	\(n_{12}\)	\(\ldots \)	\(n_{1K'}\)	\(a_1\)
\(U_2\)	\(n_{21}\)	\(n_{22}\)	\(\ldots \)	\(n_{2K'}\)	\(a_2\)
\(\vdots \)	\(\vdots \)	\(\vdots \)	\(\ddots \)	\(\vdots \)	\(\vdots \)
\(U_K\)	\(n_{K1}\)	\(n_{K2}\)	\(\ldots \)	\(n_{KK'}\)	\(a_K\)
Total	\(b_1\)	\(b_2\)	\(\ldots \)	\(b_{K'}\)	\(N=\sum_{i=1}^K\sum_{j=1}^{K'} n_{ij}\)

Table 2 Spider data ~~set:~~ set: Site clustering results for Poisson, NB and ordered stereotype model. ~~The number of fitted clusters is \(R=3\). All the allocations are based on highest posterior probabilities except for the "Stereotype 2" clustering which has a fuzzy allocation in the sites shown in boldface~~

Groups	Clustering (highest probability)			Stereotype 2	
	Poisson	NB	Stereotype		
R1~~{1-7,9-14,25}~~	{1–7,9—14,25}	{1-7,13,14}~~{1-7,13,14}~~	{1–7,13,14}	~~{1-7,13,14,~~{1–7,13,14,**17**}	
R2~~{22-24,26-28}~~	{22–24,26–28}~~{9-12,22-28}~~	{9–12,22–28}~~{8,21-24,27,28}~~	{8,21–24,27,28}	{8,**16**,**19**~~,21-24,27,28}~~,21–24,27,28}	
R3~~{8,15-21}~~	{8,15–21}~~{8,15-21}~~	{8,15–21}~~{9-12,15-20,25-26}~~	{9–12,15–20,25–26}	~~{9-12,15,18,20,25-26}~~	{9–12,15,18,20,25–26}

The number of fitted clusters is \(R=3\). All the allocations are based on the highest posterior probabilities except for the "Stereotype 2" clustering which has a fuzzy allocation in the sites shown in boldface.

Table 3 Spider data ~~set:~~ set: Clustering results for Poisson, ~~NB,~~ NB and two classifications based on the ordered stereotype model ("Stereotype" and "Stereotype 2").

Clustering comparison	ARI	1-NVI	1-NID
Poisson versus NB	0.555	0.562	0.701
Poisson versus Stereotype	0.280 (50.5 %)	0.229 (40.7 %)	0.361 (51.5 %)
NB versus Stereotype	0.409 (73.7 %)	0.335 (59.6 %)	0.500 (71.3 %)
Poisson versus Stereotype 2	0.334 (60.2 %)	0.304 (54.1 %)	0.457 (65.2 %)
NB versus Stereotype 2	**0.465 (83.8 %)**	**0.423 (75.3 %)**	**0.590 (84.2 %)**

The number of fitted clusters is \(R=3\). Large values indicate similarity of clustering. The percentage of the Poisson ~~vs.~~ versus NB clustering explained by the clustering with ordinal data is indicated in parenthesis. The closest clusterings are the two count data-based models (Poisson and NB) over the three measures. Between count and ordinal data-based models, "Stereotype 2" is closer to NB than Poisson and is shown in ~~boldface~~

~~Clustering Comparison~~	~~ARI~~	~~1-NVI~~	~~1-NID~~
~~Poisson vs. NB~~	~~0.555~~	~~0.562~~	~~0.701~~
~~Poisson vs. Stereotype~~	~~0.280 (50.5%)~~	~~0.229 (40.7%)~~	~~0.361 (51.5%)~~
~~NB vs. Stereotype~~	~~0.409 (73.7%)~~	~~0.335 (59.6%)~~	~~0.500 (71.3%)~~
~~Poisson vs. Stereotype 2~~	~~0.334 (60.2%)~~	~~0.304 (54.1%)~~	~~0.457 (65.2%)~~
~~NB vs. Stereotype 2~~	~~**0.465 (83.8%)**~~	~~**0.423 (75.3%)**~~	~~**0.590 (84.2%)**~~

~~**Table 4 Urban bird abundance data set**: Comparison between Poisson and ordered stereotype model cluster structure.~~ boldface.

Table 4 Urban bird abundance data set: Comparison between Poisson and ordered stereotype model cluster structure.

Groups	Poisson		Stereotype	
	Mean	Var-mean ratio	Mean	Var-mean ratio
C1	2.206	0.659	2.206	0.659
C2	0.528	1.100	0.501	1.548

C3	0.114	1.021	0.213	0.937

The number of fitted clusters is \(C=3\). Cells in the columns labelled as “Mean” and ~~“Var-Mean Ratio”~~ “Var-mean ratio” are the average of the mean of the species and variance-mean ration across the 51 ~~sites~~ sites, respectively, for each species ~~cluster~~

~~Groups~~	~~Poisson~~		~~Stereotype~~	
	~~Mean~~	~~Var-Mean Ratio~~	~~Mean~~	~~Var-Mean Ratio~~
~~C1~~	~~2.206~~	~~0.659~~	~~2.206~~	~~0.659~~
~~C2~~	~~0.528~~	~~1.100~~	~~0.501~~	~~1.548~~
~~C3~~	~~0.114~~	~~1.021~~	~~0.213~~	~~0.937~~

cluster.