`clustord` Ordinal Models

Louise McMillan

2025-05-27

Contents

TL:DR

The proportional-odds model is named `"POM"` in the `clustord()` model argument. It is the most commonly-used model for ordinal data analysis, and it is the simplest, in that it has the fewest parameters and a consistent pattern for all categories of the ordinal response variables.

The ordered stereotype model is named `"OSM"` in the `clustord()` model argument. Compared with the proportional-odds model, it has one additional parameter for every category/level of the ordinal response variable. These additional parameters, plus its non-cumulative structure, make it much more flexible. Use the OSM if you think that your ordinal data may be very heterogeneous in terms of the patterns of different categories of the response variables.

The $\texttt{phi}_k$ score parameters for the response categories k indicate how much information is available in each response category. If $\texttt{phi}_k$ and $\texttt{phi}_{k+1}$ are very similar (within 0.1 of each other) then that indicates that response categories k and $k+1$ do not provide much information about any clustering structure, which means you could simplify your data by combining those two response categories without having much effect on the results of the analysis.

You can also use the $\texttt{phi}_k$ scores from the OSM as a data-driven numerical encoding of your ordinal response categories that is better than simply numbering the categories $1, 2, \ldots, q$ and then carry out analysis using methods for numerical data, such as k-means (Lloyd, 1982 and MacQueen, 1967).

Introduction

For this package, we assume that you have a dataset of ordinal data. The most common form of this is survey data, such as you might get by asking participants to ask a series of questions with Likert-scale answers (for example, ranking from 1 = “Strongly Disagree” to 5 = “Strongly Agree”).

Name	Q1	Q2	Q3	Q4	Q5	Q6
Wen	3	3	2	3	3	3
Mirai	1	2	3	1	3	2
An	2	2	2	1	3	2
Max	2	1	1	1	2	1

We will refer to the data matrix as $\boldsymbol{Y}$. We index the rows of the data matrix with i and the columns of the data matrix with j, so an individual response value is defined as Y_{ij}. The q categories of each response variable Y_{ij} are indexed with k, with $k = 1, \ldots, q$.

There are three broad types of clustering: row clustering, column clustering and biclustering. Within each of these, there are multiple possible clustering structures. These are discussed in detail in the `clustord` *Tutorial* vignette, and summarised in the *Clustering Structure Summary* vignette.

If there are row clusters, they are indexed with r and if there are column clusters they are indexed with c.

This vignette discusses the two types of ordinal models that are available in `clustord`: the **proportional-odds model (POM)** and the **ordered stereotype model (OSM)**.

Proportional-odds model (POM)

The first model is the proportional-odds model (Agresti, 2000). This is the most widely-used ordinal model, and the simplest. The model is more easily recognisable as a regression model:

$$\text{logit}P(Y \leq k \,|\, \boldsymbol{x}) = \log\left(\frac{P(Y \leq k \,|\, \boldsymbol{x})}{P(Y > k \,|\, \boldsymbol{x})}\right) = \mu_k - \boldsymbol{\beta}^T\boldsymbol{x} \;\; \text{for } k = 1, \ldots, q-1$$

where μ_k is the intercept parameter for response category k and $\boldsymbol{\beta}$ are the coefficients controlling the effect of $\boldsymbol{x}$ on the response, Y. The model is named "proportional-odds" because the coefficients do not depend on the response category, k. That is, the effect of $\boldsymbol{x}$ is the same for every single category of the response. Thus, the number of coefficients is only one more than the number of covariates.

This is also a **cumulative** ordinal model in that it is expressed as the probability of obtaining a given response category or lower, relative to the probability of getting any of the higher response categories. Along with the coefficients staying the same for all categories, this is the other part of how this model enforces similar patterns of effects for every response category.

The proportional-odds clustering forms in `clustord` were introduced in Matechou et al. (2016). Considering cell (i, j) in the data matrix of responses, where if row i is in row cluster r and/or column j is in column c then the model has this general shape:

$$\log\left(\frac{P(Y_{ij} \leq k \,|\, i \in r,\ j \in c)}{P(Y_{ij} > k \,|\, i \in r,\ j \in c)}\right) = \mu_k - \eta_{rcij} \;\; \text{for } k = 1, \ldots, q-1$$

where μ_k is a parameter that controls the default probabilities of the different response categories in the absence of clustering, and η_{rcij} is the remaining part of the linear predictor. η_{rcij} is the part of the model that determines the clustering structure, and it is the same in both the proportional-odds model and the ordered stereotype model within `clustord`.

η_{rcij} is subtracted from μ_k, rather than added, so that if a parameter within η_{rcij} is positive then that corresponds to a higher probability of obtaining higher response categories (whereas if it were added, positive effects would lead to a higher probability of obtaining lower response categories).

The logit notation above is the most compact way of expressing this model, but alternatively we can express it in terms of the probability, θ_{ijk}, of getting a single response category, k, in cell Y_{ij}:

$$\theta_{ijk} \,|\, i \in r,\ j \in c = \begin{cases} \dfrac{\exp(\mu_k - \eta_{rcij})}{1 + \exp(\mu_k - \eta_{rcij})} & k = 1 \\ \dfrac{\exp(\mu_k - \eta_{rcij})}{1 + \exp(\mu_k - \eta_{rcij})} - \dfrac{\exp(\mu_{k-1} - \eta_{rcij})}{1 + \exp(\mu_{k-1} - \eta_{rcij})} & 1 < k < q \\ 1 - \sum_{k=1}^{q-1} \theta_{ijk} & k = q \end{cases}$$

These are some of the potential clustering structures within `clustord`, expressed in POM form:

- Row clustering only:

$$\log\left(\frac{P(Y_{ij} \le k \,|\, i \in r)}{P(Y_{ij} > k \,|\, i \in r)}\right) = \mu_k - \texttt{rowc}_r$$

  In this case, $\eta_{rcij} = \texttt{rowc}_r$ i.e. the only structure is the presence of row clusters.

- Row clustering with individual column effects:

$$\log\left(\frac{P(Y_{ij} \le k \,|\, i \in r)}{P(Y_{ij} > k \,|\, i \in r)}\right) = \mu_k - \texttt{rowc}_r - \texttt{col}_j$$

  In this case, $\eta_{rcij} = \texttt{rowc}_r + \texttt{col}_j$, which gives row clusters whilst allowing for each individual column to have a different pattern.

- Row clustering with individual column effects and interactions:

$$\log\left(\frac{P(Y_{ij} \le k \,|\, i \in r)}{P(Y_{ij} > k \,|\, i \in r)}\right) = \mu_k - \texttt{rowc}_r - \texttt{col}_j - \texttt{rowc\_col}_{rj}$$

  In this case, $\eta_{rcij} = \texttt{rowc}_r + \texttt{col}_j + \texttt{rowc\_col}_{rj}$, which gives row clusters whilst allowing for each individual column to have a different pattern, and includes the interaction between those individual patterns and the row clusters so that each row cluster has a different set of individual column patterns. $\texttt{rowc\_col}_{rj}$ is a matrix of parameters.

- Row clustering with row covariates $\boldsymbol{x}_i$:

$$\log\left(\frac{P(Y_{ij} \le k \,|\, i \in r, \boldsymbol{X})}{P(Y_{ij} > k \,|\, i \in r, \boldsymbol{X})}\right) = \mu_k - \texttt{rowc}_r - \texttt{cov}^T \boldsymbol{x}_i$$

  In this case, $\eta_{rcij} = \texttt{rowc}_r + \texttt{cov}^T \boldsymbol{x}_i$, so there is row clustering structure and also covariates, but the effect of the covariates is the same for all the clusters. `cov` is a **vector** of covariate coefficients that is the same length as the number of covariates. $\boldsymbol{x}_i$ is a **vector** of covariates for the response row $\boldsymbol{Y}_i$.

- Row clustering with row covariates $\boldsymbol{x}_i$ and coefficients dependent on row cluster:

$$\log\left(\frac{P(Y_{ij} \le k \,|\, i \in r, \boldsymbol{X})}{P(Y_{ij} > k \,|\, i \in r, \boldsymbol{X})}\right) = \mu_k - \texttt{rowc}_r - \texttt{cov}_r^T \boldsymbol{x}_i$$

  In this case, $\eta_{rcij} = \texttt{rowc}_r + \texttt{cov}_r^T \boldsymbol{x}_i$, so there is row clustering structure and also covariates, and the effect of the covariates is different for each row cluster. $\texttt{cov}_r$ is a **vector** of covariate coefficients that is the same length as the number of covariates, and there is a different set of coefficients for each row cluster r. $\boldsymbol{x}_i$ is a **vector** of covariates for the response row $\boldsymbol{Y}_i$.

- Row clustering with column covariates $\boldsymbol{w}_j$:

$$\log\left(\frac{P(Y_{ij} \le k \,|\, i \in r, \boldsymbol{X})}{P(Y_{ij} > k \,|\, i \in r, \boldsymbol{X})}\right) = \mu_k - \texttt{rowc}_r - \texttt{cov}^T \boldsymbol{w}_j$$

In this case, $\eta_{rcij} = \texttt{rowc}_r + \texttt{cov}^T \boldsymbol{w}_j$, so there is row clustering structure and also covariates, but the effect of the covariates is the same for all the clusters. `cov` is a **vector** of covariate coefficients that is the same length as the number of covariates. $\boldsymbol{w}_j$ is a **vector** of covariates for the response column $\boldsymbol{Y}_j$.

- Row clustering with column covariates $\boldsymbol{w}_j$ and coefficients dependent on row cluster:

$$\log\left(\frac{P(Y_{ij} \le k \,|\, i \in r, \boldsymbol{X})}{P(Y_{ij} > k \,|\, i \in r, \boldsymbol{X})}\right) = \mu_k - \texttt{rowc}_r - \texttt{cov}_r^T \boldsymbol{w}_j$$

In this case, $\eta_{rcij} = \texttt{rowc}_r + \texttt{cov}_r^T \boldsymbol{w}_j$, so there is row clustering structure and also covariates, and the effect of the covariates is different for each row cluster. $\texttt{cov}_r$ is a **vector** of covariate coefficients that is the same length as the number of covariates, and there is a different set of coefficients for each row cluster r. $\boldsymbol{w}_j$ is a **vector** of covariates for the response column $\boldsymbol{Y}_j$.

- Column clustering only:

$$\log\left(\frac{P(Y_{ij} \le k \,|\, j \in c)}{P(Y_{ij} > k \,|\, j \in c)}\right) = \mu_k - \texttt{colc}_c$$

In this case, $\eta_{rcij} = \texttt{colc}_c$ i.e. the only structure is the presence of column clusters.

- Column clustering with individual row effects:

$$\log\left(\frac{P(Y_{ij} \le k \,|\, j \in c)}{P(Y_{ij} > k \,|\, j \in c)}\right) = \mu_k - \texttt{rowc}_r - \texttt{col}_j$$

In this case, $\eta_{rcij} = \texttt{rowc}_r + \texttt{row}_i$, which gives column clusters whilst allowing for each individual row to have a different pattern.

- Column clustering with individual row effects and interactions:

$$\log\left(\frac{P(Y_{ij} \le k \,|\, j \in c)}{P(Y_{ij} > k \,|\, j \in c)}\right) = \mu_k - \texttt{rowc}_r - \texttt{col}_j - \texttt{colc\_row}_{ci}$$

In this case, $\eta_{rcij} = \texttt{rowc}_r + \texttt{col}_j + \texttt{colc\_row}_{ci}$, which gives column clusters whilst allowing for each individual row to have a different pattern, and includes the interaction between those individual patterns and the column clusters so that each column cluster has a different set of individual row patterns. $\texttt{colc\_row}_{ci}$ is a matrix of parameters.

- Biclustering:

$$\log\left(\frac{P(Y_{ij} \le k \,|\, i \in r,\ j \in c)}{P(Y_{ij} > k \,|\, i \in r,\ j \in c)}\right) = \mu_k - \texttt{rowc}_r - \texttt{colc}_c$$

In this case, $\eta_{rcij} = \texttt{rowc}_r + \texttt{colc}_c$ so there are both row clusters and column clusters.

- Biclustering with interaction between row clusters and column clusters:

$$\log\left(\frac{P(Y_{ij} \le k \,|\, i \in r,\ j \in c)}{P(Y_{ij} > k \,|\, i \in r,\ j \in c)}\right) = \mu_k - \texttt{rowc}_r - \texttt{colc}_c - \texttt{rowc\_colc}_{rc}$$

In this case, $\eta_{rcij} = \texttt{rowc}_r + \texttt{colc}_c + \texttt{rowc\_colc}_{rc}$. $\texttt{rowc\_colc}_{rc}$ is a matrix of parameters.

Column clustering and biclustering can similarly include covariates, in the same way as row clustering can. Note that because the coefficient/covariate structure of including covariates is the same in the clustering models regardless of whether the covariates are attached to the rows ($\boldsymbol{x}_i$) or the columns ($\boldsymbol{w}_j$), the `clustord` package combines their coefficients into one single `cov` parameter vector, with the order of the coefficients corresponding to the order in which they're included in the formula provided to `clustord()`.

Ordered stereotype model

The second model is the ordered stereotype model (introduced by Anderson, 1984 and described in Agresti, 2000). This is a more flexible model than the proportional-odds model. It has one additional set of parameters, $\{\phi_k\}$, and a non-cumulative logit structure. This is the regression model form:

$$\log\left(\frac{P(Y=k\,|\,\boldsymbol{x})}{P(Y=1\,|\,\boldsymbol{x})}\right)=\mu_k+\phi_k\boldsymbol{\beta}^T\boldsymbol{x}\ \text{ for } k=2,\ldots,q$$

where $\boldsymbol{\beta}$ are the coefficients controlling the effect of $\boldsymbol{x}$ on the response, Y.

μ_1 is set to 0 to ensure identifiability and the ϕ_k parameters are constrained to be ordered: $0=\phi_1\leq\phi_2\leq\cdots\leq\phi_q=1$. (The non-ordered stereotype model lacks this constraint, and can be used to model nominal data.)

In the ordered stereotype model, the ϕ_k parameters modify the effect of the covariate on the response so that the effect varies between response categories. Moreover, the model is non-cumulative, so the pattern of response category 3, relative to category 1, can be different than the pattern of response category 2, relative to category 1.

The ordered stereotype clustering forms in `clustord` were defined in Fernández et al. (2016) and Fernández et al. (2019). Again for cell (i,j) in the data matrix of responses, if row i is in row cluster r and/or column j is in column c then the model has this general shape:

$$\log\left(\frac{P(Y_{ij}=k\,|\,i\in r,\ j\in c)}{P(Y_{ij}=1\,|\,i\in r,\ j\in c)}\right)=\mu_k+\phi_k\eta_{rcij}\ \text{ for } k=2,\ldots,q$$

where μ_k is a parameter that controls the default probabilities of the different response categories in the absence of clustering, ϕ_k is the score parameter for category k and η_{rcij} is the remaining part of the linear predictor. η_{rcij} is the same as in the proportional-odds model.

As for POM, the notation above is the most compact way of expressing the OSM, but alternatively we can express it in terms of the probability, θ_{ijk}, of getting a single response category, k, in cell Y_{ij}:

$$\theta_{ijk}\,|\,i\in r,\ j\in c=\frac{\exp(\mu_k+\phi_k\eta_{rcij})}{\sum_{l=1}^{q}\exp(\mu_l+\phi_l\eta_{rcij})}k=1,\ldots,q$$

These are some of the potential clustering structures within `clustord`, expressed in OSM form:

- Row clustering only:

  $$\log\left(\frac{P(Y_{ij}=k\,|\,i\in r)}{P(Y_{ij}=1\,|\,i\in r)}\right)=\mu_k+\phi_k(\texttt{rowc}_r)$$

  In this case, $\eta_{rcij}=\texttt{rowc}_r$ i.e. the only structure is the presence of row clusters.

- Row clustering with individual column effects:

  $$\log\left(\frac{P(Y_{ij}=k\,|\,i\in r)}{P(Y_{ij}=1\,|\,i\in r)}\right)=\mu_k+\phi_k(\texttt{rowc}_r+\texttt{col}_j)$$

  In this case, $\eta_{rcij}=\texttt{rowc}_r+\texttt{col}_j$, which gives row clusters whilst allowing for each individual column to have a different pattern.

- Row clustering with individual column effects and interactions:

  $$\log\left(\frac{P(Y_{ij}=k\,|\,i\in r)}{P(Y_{ij}=1\,|\,i\in r)}\right)=\mu_k+\phi_k(\texttt{rowc}_r+\texttt{col}_j+\texttt{rowc\_col}_{rj})$$

  In this case, $\eta_{rcij}=\texttt{rowc}_r+\texttt{col}_j+\texttt{rowc\_col}_{rj}$, which gives row clusters whilst allowing for each individual column to have a different pattern, and includes the interaction between those individual patterns and the row clusters so that each row cluster has a different set of individual column patterns. $\texttt{rowc\_col}_{rj}$ is a matrix of parameters.

- Row clustering with row covariates $\boldsymbol{x}_i$:

$$\log\left(\frac{P(Y_{ij}=k\,|\,i\in r,\boldsymbol{X})}{P(Y_{ij}=1\,|\,i\in r,\boldsymbol{X})}\right)=\mu_k+\phi_k(\texttt{rowc}_r+\texttt{cov}^T\boldsymbol{x}_i)$$

  In this case, $\eta_{rcij}=\texttt{rowc}_r+\texttt{cov}^T\boldsymbol{x}_i$, so there is row clustering structure and also covariates, but the effect of the covariates is the same for all the clusters. `cov` is a **vector** of covariate coefficients that is the same length as the number of covariates. $\boldsymbol{x}_i$ is a **vector** of covariates for the response row $\boldsymbol{Y}_i$.

- Row clustering with row covariates $\boldsymbol{x}_i$ and coefficients dependent on row cluster:

$$\log\left(\frac{P(Y_{ij}=k\,|\,i\in r,\boldsymbol{X})}{P(Y_{ij}=1\,|\,i\in r,\boldsymbol{X})}\right)=\mu_k+\phi_k(\texttt{rowc}_r+\texttt{cov}_r^T\boldsymbol{x}_i)$$

  In this case, $\eta_{rcij}=\texttt{rowc}_r+\texttt{cov}_r^T\boldsymbol{x}_i$, so there is row clustering structure and also covariates, and the effect of the covariates is different for each row cluster. $\texttt{cov}_r$ is a **vector** of covariate coefficients that is the same length as the number of covariates, and there is a different set of coefficients for each row cluster r. $\boldsymbol{x}_i$ is a **vector** of covariates for the response row $\boldsymbol{Y}_i$.

- Row clustering with column covariates $\boldsymbol{w}_j$:

$$\log\left(\frac{P(Y_{ij}=k\,|\,i\in r,\boldsymbol{X})}{P(Y_{ij}=1\,|\,i\in r,\boldsymbol{X})}\right)=\mu_k+\phi_k(\texttt{rowc}_r+\texttt{cov}^T\boldsymbol{w}_j)$$

  In this case, $\eta_{rcij}=\texttt{rowc}_r+\texttt{cov}^T\boldsymbol{w}_j$, so there is row clustering structure and also covariates, but the effect of the covariates is the same for all the clusters. `cov` is a **vector** of covariate coefficients that is the same length as the number of covariates. $\boldsymbol{w}_j$ is a **vector** of covariates for the response column $\boldsymbol{Y}_j$.

- Row clustering with column covariates $\boldsymbol{w}_j$ and coefficients dependent on row cluster:

$$\log\left(\frac{P(Y_{ij}=k\,|\,i\in r,\boldsymbol{X})}{P(Y_{ij}=1\,|\,i\in r,\boldsymbol{X})}\right)=\mu_k+\phi_k(\texttt{rowc}_r+\texttt{cov}_r^T\boldsymbol{w}_j)$$

  In this case, $\eta_{rcij}=\texttt{rowc}_r+\texttt{cov}_r^T\boldsymbol{w}_j$, so there is row clustering structure and also covariates, and the effect of the covariates is different for each row cluster. $\texttt{cov}_r$ is a **vector** of covariate coefficients that is the same length as the number of covariates, and there is a different set of coefficients for each row cluster r. $\boldsymbol{w}_j$ is a **vector** of covariates for the response column $\boldsymbol{Y}_j$.

- Column clustering only:

$$\log\left(\frac{P(Y_{ij}=k\,|\,j\in c)}{P(Y_{ij}=1\,|\,j\in c)}\right)=\mu_k+\phi_k(\texttt{colc}_c)$$

  In this case, $\eta_{rcij}=\texttt{colc}_c$ i.e. the only structure is the presence of column clusters.

- Column clustering with individual row effects:

$$\log\left(\frac{P(Y_{ij}=k\,|\,j\in c)}{P(Y_{ij}=1\,|\,j\in c)}\right)=\mu_k+\phi_k(\texttt{rowc}_r+\texttt{col}_j)$$

  In this case, $\eta_{rcij}=\texttt{rowc}_r+\texttt{row}_i$, which gives column clusters whilst allowing for each individual row to have a different pattern.

- Column clustering with individual row effects and interactions:

$$\log\left(\frac{P(Y_{ij}=k\,|\,j\in c)}{P(Y_{ij}=1\,|\,j\in c)}\right)=\mu_k+\phi_k(\texttt{rowc}_r+\texttt{col}_j+\texttt{colc\_row}_{ci})$$

  In this case, $\eta_{rcij}=\texttt{rowc}_r+\texttt{col}_j+\texttt{colc\_row}_{ci}$, which gives column clusters whilst allowing for each individual row to have a different pattern, and includes the interaction between those individual patterns and the column clusters so that each column cluster has a different set of individual row patterns. $\texttt{colc\_row}_{ci}$ is a matrix of parameters.

- Biclustering:

$$\log\left(\frac{P(Y_{ij}=k\,|\,i\in r,\ j\in c)}{P(Y_{ij}=1\,|\,i\in r,\ j\in c)}\right)=\mu_k+\phi_k(\texttt{rowc}_r+\texttt{colc}_c)$$

  In this case, $\eta_{rcij}=\texttt{rowc}_r+\texttt{colc}_c$ so there are both row clusters and column clusters.

- Biclustering with interaction between row clusters and column clusters:

$$\log\left(\frac{P(Y_{ij}=k\,|\,i\in r,\ j\in c)}{P(Y_{ij}=1\,|\,i\in r,\ j\in c)}\right)=\mu_k+\phi_k(\texttt{rowc}_r+\texttt{colc}_c+\texttt{rowc\_colc}_{rc})$$

  In this case, $\eta_{rcij}=\texttt{rowc}_r+\texttt{colc}_c+\texttt{rowc\_colc}_{rc}$. $\texttt{rowc\_colc}_{rc}$ is a matrix of parameters.

As for POM, because the coefficient/covariate structure of including covariates is the same in the clustering models regardless of whether the covariates are attached to the rows ($\boldsymbol{x}_i$) or the columns ($\boldsymbol{w}_j$), the `clustord` package combines their coefficients into one single `cov` parameter vector, with the order of the coefficients corresponding to the order in which they're included in the formula provided to `clustord()`.

Note that η_{rcij} takes the same forms for POM and OSM even though the overall distribution shapes differ.

Table 1: Comparison of `clustord` notation with the notation used in the original journal articles.

α_r	`rowc`
β_j	`col`
γ_{rj}	`rowc_col`
β_c	`colc`
α_i	`row`
γ_{ic}	`colc_row`
γ_{rc}	`rowc_colc`

A note about notation

If you are looking at the cited journal articles by Pledger and Arnold (2014), Matechou et al. (2016), and Fernández et al. (2016 and 2019), the notation in those is slightly different than the notation used in this tutorial. The package and tutorial notation was changed to reduce confusion between the parameters in the row clustering and column clustering models.

Table 1 is a glossary of the notation used in `clustord` and the corresponding notation used in the articles.

The rest of the parameters retain the same names in this tutorial and the cited references.

Note also that, although it is theoretically possible in this model structure to add α_r and α_i to the same model, ie. row cluster effects **and** individual row effects, `clustord` does not allow this, and will warn you if you try to use `Y ~ ROWCLUST + ROW` or similar formulae. And the biclustering model, which has α_r and β_c, does not allow either individual row or individual column effects, partly because this would introduce too many parameters and be too difficult to fit correctly.

References

Agresti, A. (2010). *Analysis of ordinal categorical data*. Vol. 656, John Wiley & Sons.

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

Fernández, D., Arnold., R. and Pledger, S. (2016). Mixture-based clustering for the ordered stereotype model. *Computational Statistics & Data Analysis*, **93**, pp. 46–75.

Fernández, D., Arnold, R., Pledger, S., Liu, I., & Costilla, R. (2019). Finite mixture biclustering of discrete type multivariate data. *Advances in Data Analysis and Classification*, **13**, pp. 117–143.

Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, **28**(2), pp. 129–137.

MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press*, **1**(14), pp. 281–297.

Matechou, E., Liu, I., Fernández, D. Farias, M., and Gjelsvik, B. (2016). Biclustering models for two-mode ordinal data. *Psychometrika*, **81**, pp. 611–624.

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