



# Modelling scale effects in rating data: a Bayesian approach

Maria Iannario<sup>1</sup> · Maria Kateri<sup>2</sup> · Claudia Tarantola<sup>3</sup>

Accepted: 26 December 2023  
© The Author(s) 2024

## Abstract

We present a Bayesian approach for the analysis of rating data when a scaling component is taken into account, thus incorporating a specific form of heteroskedasticity. Model-based probability effect measures for comparing distributions of several groups, adjusted for explanatory variables affecting both location and scale components, are proposed. Markov Chain Monte Carlo techniques are implemented to obtain parameter estimates of the fitted model and the associated effect measures. An analysis on students' evaluation of a university curriculum counselling service is carried out to assess the performance of the method and demonstrate its valuable support for the decision-making process.

**Keywords** Bayesian ordinal superiority measures · Heterogeneity of variances · MCMC · Ordinal responses · Scale effects

## 1 Introduction

Rating surveys have gained significant prominence as crucial tools for evaluating service quality, having been established as pivotal instruments for the quantitative assessment of respondents' satisfaction regarding the received services. This practice is not only widespread, but has also become an integral part of performance appraisal in various public sectors (see Poister and Gary 1994, among others).

The higher education sector is noteworthy, where evaluation surveys serve as powerful tool for shaping the narrative; see e.g., Kanwar (2022) and reference herein. Surveys enable a multifaceted approach to comprehending the dynamics within educational institutions. By actively involving students, faculties, staff, and other stakeholders, these surveys open up valuable channels of feedback on crucial aspects such as teaching quality, available resources, campus facilities, and the overall environment; see e.g., Bejan et al. (2015). This feedback, in turn, provides an overview of an institution's strengths and areas in need of improvement.

---

✉ Maria Iannario  
maria.iannario@unina.it

<sup>1</sup> Department of Political Sciences, University of Naples Federico II, Via L. Rodinó, 22, Naples, Italy

<sup>2</sup> Institute of Statistics, RWTH Aachen University, Aachen, Germany

<sup>3</sup> Department of Economics and Management, University of Pavia, Pavia, Italy

Structured and standardized evaluation surveys facilitate the systematic collection of data on various courses, programs, and departments. This methodology empowers institutions to discern trends and patterns over time, laying the foundation for informed decision-making. Additionally, the feedback from these surveys spotlights specific areas requiring attention and improvement, ranging from teaching techniques and course content to the quality of facilities, availability of support services, and the general climate of the surveyed institution.

Using evaluation survey data, institutions gain the opportunity to assess their own performance in relation to similar facilities or industry benchmarks. This comparative analysis yields valuable insights into the institution's position and serves as a compass for identifying areas of excellence and those in need of further development. Moreover, this data-driven approach supports evidence-based decision-making, providing administrators with empirical evidence to guide choices regarding resource allocation, program development, and policy implementation.

By actively seeking feedback through evaluation surveys, institutions demonstrate a commitment to accountability and transparency. This practice underscores respect for the perspectives and experiences of stakeholders and signals that these contributions are valued and will be utilized to bring about positive changes within the institutions.

Understanding students' opinions through evaluation surveys leads to improvements in various facets of education. This encompasses refining teaching methods, adapting curriculum design, and enhancing student support services, ultimately resulting in greater student engagement, higher levels of satisfaction, and increased retention rates.

As noted in QAA (2018), engaging students in quality assurance and improvement practices reflects the fundamental principles of higher education. Students become not simply recipients of university services; they also serve as instrumental participants in evaluating their effectiveness.

The regular administration of assessment surveys fosters a culture of continuous improvement within higher education institutions. This practice signifies a dedication to continuous self-evaluation and a willingness to adapt and evolve in response to feedback. It serves as an example of an institution's commitment to the constant pursuit of excellence.

Furthermore, evaluation survey data are of great importance in accreditation processes and external assessments; see, e.g., Brennan and Shah (2000). They function as an additional source of evidence, demonstrating an institution's compliance with established quality standards and its proficiency in achieving desired educational outcomes. In this manner, evaluation surveys play a pivotal role not only in understanding the current state of an institution but also in charting its future trajectory in higher education.

In the higher education sector, university curriculum counselling has emerged as a notable domain. In this context, the evaluation of students' feedback has become a cornerstone for enhancing service quality. This reflects a broader trend towards prioritizing user-centric approaches in service delivery, highlighting the significance placed on the perspectives and experiences of the recipients of these services. Related surveys aim to establish a virtuous cycle of student-centered education support by identifying areas with low satisfaction and improving them. Moreover, they play a crucial role in future government financial support projects and the establishment of plans for universities. These surveys measure the level of satisfaction or awareness of current students, and provide a wide range of information about the students' educational experience, see Shin et al. (2020).

We focus on this kind of surveys analysing a set of data resulting from a study fulfilled at University of Naples Federico II from 2002 to 2004, and replicated in 2007 and 2008. The purpose of the related surveys was to produce basic data for establishing development

strategies such as improving service quality and satisfaction by identifying educational or infrastructural problems and points for replanning or further development.

The early years of the 21st century were transformative for the Italian higher education system, marked by substantial reforms designed to align it with European standards, elevate quality, and promote innovation and internationalization. This context underscores the enduring relevance of the insights gleaned from our data analysis in the contemporary academic landscape. The investigation aimed to improve the quality of the service according to the needs of students for the necessity of strengthening university competitiveness. In Sect. 3 we provide a brief description of the data; further details can be found in Capecchi and Piccolo (2010), and reference herein.

To assess the quality of the service, students reported scores to items on ordinal rating scales. Frequently it is assumed that an ordinal response is the discretization of an underlying unobserved (latent) continuous variable, with every possible value (score) for the ordinal response corresponding to an interval of the latent variable. The response to an item is selected on a discrete ordered scale ranging from one qualitative end point to another one (as in the mentioned survey). It does not have metric information, although numerically labeled options may be considered. When a single item question is considered, the main model to analyse the obtained scores is the cumulative model with the proportional assumption for the odds, scaled by the model's link function; see Agresti (2010). Under this model, the effect of the explanatory variables is the same across all cumulative odds defined over the response categories.

The cumulative model represents an unconditional framework, unlike other ordinal regression models such as the adjacent categories model and the whole class of hierarchically structured models (see Tutz 2022 for further details). The assumption that effects of covariates are not category-specific makes it a simply structured model that allows to interpret parameters in terms of cumulative odds (alternative structure are in Boes and Winkelmann 2006, among others). However, if unobserved heterogeneity related to the latent variable is present, scale effects in the regression structure with ordinal responses are needed.

The unobserved heterogeneity can be determined by several causes. It can be triggered by the sampling system; in this case, observed outcomes differ between studies even though they estimate a common underlying parameter. In other cases, observed outcomes are more variable than expected on the basis of sampling variability alone. This is typically interpreted as evidence of the presence of variability in the underlying outcomes. When heterogeneity is present, one can try to examine whether certain predictor variables (also known as moderators or effect modifiers) are able to explain at least part of the heterogeneity in the outcomes. This is the objective of the present research work that aims to detect category specific covariates effects (see Bürkner and Vuorre 2019). In particular, here we rely on the location-scale models (McCullagh 1980) and we present simple ways for effects interpretation for these types of models.

There is a wide literature regarding ordinal data analysis both from a frequentist and Bayesian perspective, see e.g., Johnson and Albert (1999) for an insightful comparisons of the two approaches. While frequentist methods for the analysis of location-scale models are available (see e.g., McCullagh 1980; Tutz 2022), here, following Bürkner and Vuorre (2019); Bürkner (2017), we adopt a Bayesian framework, gaining flexibility in specifying the model and enhancing richness and accuracy in providing parameter estimates. In a Bayesian setup it is possible to combine in a natural way prior information and new data. Once the posterior distribution of the parameter of interest is obtained, one can easily derive the distribution of functions of interest of the examined parameters. Another advantage is the possibility to use the same framework and approach both in case the sample

size is small or large. For a discussion of the advantages of the use of a Bayesian approach for ordinal data see, e.g., Liddell and Kruschke (2018) and references therein. For some milestones regarding Bayesian approach in the context of ordinal data see Albert and Chib (1993), Albert and Chib (1997), Dellaportas and Smith (1993), Johnson (1996), Johnson and Albert (1999).

The focus of the contribution lies on the implementation of the proposed model, the use of marginal effects to address the interpretation of the results on the extreme categories of the rating scales, and the introduction of Bayesian probability-based measures for comparing clusters on ratings. The obtained results favour the proposed model compared to the location baseline one and show significant subject heterogeneity.

The paper is organized as follows. The next section provides a brief description of the parametric ordinal models in their location-scale version, presenting also marginal effects and ordinal superiority measures for this class of models. Section 3 introduces university curriculum counselling data and reports the main results of the model's implementation. The paper ends with some concluding remarks.

## 2 Model description

We consider the following setting. Let  $Y = (Y_1, Y_2, \dots, Y_n)'$  be a sample of size  $n$ , generated by an ordinal random variable  $Y \sim G(y)$  on the support  $\{1, \dots, k\}$ , where  $k$  is a known integer. We indicate with  $Y_i^*$  the underlying (continuous) latent variable such that when  $\alpha_{j-1} < Y_i^* \leq \alpha_j$ , then  $Y_i = j$ ,  $j = 1, 2, \dots, k$ . Here  $-\infty = \alpha_0 < \alpha_1 < \dots < \alpha_k = +\infty$  are the thresholds of  $Y^*$ . Let also  $X$  be a  $n \times p$  real matrix that includes all covariates ( $p \geq 1$ ) that are relevant for explaining  $Y^*$  and characterize the students' profile. The  $i$ -th row of  $X$  is the row vector  $\mathbf{x}_i$  with the covariates values for the  $i$ -th student,  $i = 1, \dots, n$ .

In our context,  $Y_i$  is the rating expressed by the  $i$ -th student on a specific question concerning the evaluation of the university curriculum counsellor service. For each student, we collect information  $\mathcal{I}_i = (y_i, \mathbf{x}_i)$ ,  $i = 1, 2, \dots, n$ , where  $y_i$  is the observed value of the rating. The latent regression model behind the process of response is  $Y_i^* = \mathbf{x}_i\boldsymbol{\beta} + \sigma_i\epsilon_i$ ,  $i = 1, 2, \dots, n$ , where  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$  are the covariates coefficients. This structure yields the location scale model from which we can obtain the standard cumulative link model (see Agresti 2010) by setting  $\sigma = 1$ . In the latent regression  $\sigma_i$  is the standard deviation of the noise variable  $\epsilon \sim F_\epsilon(\cdot)$ , and it may depend on covariates, yielding  $\sigma_i = \exp(\mathbf{z}_i\boldsymbol{\gamma})$ . Here,  $\mathbf{z}_i$  is a row vector of the matrix  $\mathbf{Z}$  which includes all the  $q \geq 1$  relevant covariates and  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)'$  are the related covariates coefficients. Note that, the vectors of covariates  $\mathbf{x}_i$  and  $\mathbf{z}_i$  may have a non-empty intersection.

The location scale model can be seen as a collection of binary response models postulating that intercepts are ordered and the effect of explanatory variables captured in the predictor is the same in all of the models-proportional assumption-see Tutz (2022).

The probability mass function of  $Y_i$ , for  $j = 1, 2, \dots, k$ , is then

$$\begin{aligned} \Pr(Y_i = j \mid \boldsymbol{\theta}, \mathbf{x}) &= \Pr(\alpha_{j-1} < Y_i^* \leq \alpha_j) \\ &= F_\epsilon \left[ (\alpha_j - \mathbf{x}_i\boldsymbol{\beta}) / \sigma_i \right] - F_\epsilon \left[ (\alpha_{j-1} - \mathbf{x}_i\boldsymbol{\beta}) / \sigma_i \right] \\ &= F_\epsilon \left[ (\alpha_j - \mathbf{x}_i\boldsymbol{\beta}) / \exp(\mathbf{z}_i\boldsymbol{\gamma}) \right] \\ &\quad - F_\epsilon \left[ (\alpha_{j-1} - \mathbf{x}_i\boldsymbol{\beta}) / \exp(\mathbf{z}_i\boldsymbol{\gamma}) \right]. \end{aligned}$$

Among the alternatives for  $F_{\epsilon}(\cdot)$ , which can be any strictly increasing distribution function, we focus on the logit link function  $F(\eta) = \exp(\eta)/(1 + \exp(\eta))$  (i.e., the canonical link) for reasons of robustness, see Iannario et al. (2017). More specifically the various link functions, which provide the relationship between the linear predictor and the probabilities of the response categories, differ in terms of sensitivity to anomalous data. Extreme design points and/or anomalous responses can negatively affect likelihood based inference in ordinal response models. For this reason, recent studies emphasise the need for a robust method or, alternatively, for link functions that ensure that maximum likelihood estimators have a limited influence function to eliminate problems associated with outlier data, such as the logistic link (see Scalera et al. 2021, for details).

The threshold values  $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_k)$  are assumed to be controlled by the anchor labels of the response prompt (e.g., ‘very satisfied’, ‘mildly satisfied’, and so on), which in our context are the same for all faculties. Because of algebraic trade-offs, the outer thresholds of the ordered data models are fixed, and only the interior thresholds are estimated from the data.

Furthermore, as already mentioned in the introduction, a proportional assumption yields the regression parameters to be invariant to the outcome category. As alternatives, parallel (non-proportional) and semi-parallel (partial proportional) odds models has been proposed (see Peterson and Harrell 1990, among others) generating an increasing level of complexity and difficulty in interpretation. The strength of the selected model for our study is the possibility of interpreting parameters intuitively, even when referring to a complex structure.

Since we do not have relevant prior information, following the approach prosed by Bürkner and Vuorre (2019), Bürkner (2017), we use non-informative priors on all parameters of interest, letting the data guide the behaviour of the posterior distributions. More precisely on the covariate coefficients we assign improper uniform priors,  $unif(-\infty, +\infty)$ , while on the intercepts we consider Student- $t$  priors with 3 degrees of freedom. This ensures that the tails are very wide, while the distribution is still proper with finite mean and variance (here set—without loss of generality—equal to 0 and 2.5 respectively). In order to obtain posterior samples we rely on Markov Chain Monte Carlo (MCMC) method. In particular we use the R package `brm` (Bürkner 2017). It implements a Hamiltonian MCMC using Stan; see e.g., Betancourt and Girolami (2013), Kreuzer et al. (2023, 2022), Neal (2011). The ordering of the intercepts is ensured via the `order` class in Stan. More precisely, the joint prior distribution is truncated to support over points satisfying the ordering constraints.

We then consider the Watanabe–Akaike information criterion (WAIC), proposed by Watanabe (2010), for model selection among different possible models for the data under consideration, see also Gelman et al. (2014), Stander et al. (2019), Watanabe (2013).

WAIC is a method for estimating pointwise out-of-sample prediction accuracy from a fitted Bayesian model using the log-likelihood evaluated at the posterior simulations of the parameter values. WAIC is obtained by adding to the log pointwise posterior predictive density a correction for the effective number of parameters to adjust for overfitting. In its original formulation WAIC (see Eq. 6 in Watanabe 2010) was defined as

$$WAIC = TL + \mathbb{V}/n, \quad (2.1)$$

where  $TL$ , the training loss, is given by

$$TL = -\frac{1}{n} \sum_{i=1}^n \log \mathbb{E}_{\theta}[\Pr(Y_i | \theta)],$$

with  $\mathbb{E}_{\theta}[\Pr(Y_i | \theta)]$  being the expectation value with respect to the posterior distribution of  $\theta$  and  $\mathbb{V}$  being the functional variance

$$\mathbb{V} = \sum_{i=1}^n \text{var} \log(\Pr(Y_i | \boldsymbol{\theta})) = \sum_{i=1}^n \{ \mathbb{E}_{\boldsymbol{\theta}} [\log \Pr(Y_i | \boldsymbol{\theta})^2] - \mathbb{E}_{\boldsymbol{\theta}} [\log \Pr(Y_i | \boldsymbol{\theta})]^2 \}.$$

Both the training loss and the functional variance can be approximated by simulations, replacing the expectations by averages over the posterior draws of  $\boldsymbol{\theta}$ ; see Gelman et al. (2014) for further details. For alternative Bayesian model assessment procedures we refer to Burnham and Anderson (2002) and Piironen and Vehtari (2017), and references herein. Their presentation and comparison to WAIC are beyond our scope.

## 2.1 Effect measures for covariates in location-scale models

For the examined models, as a consequence of the nonlinearity of the link function, model parameters are not as simple to interpret as slopes and correlations for the standard metric models (i.e., ordinary linear regression). This enhances the need to introduce simpler ways to interpret the effects of the covariates as advocated in Agresti and Tarantola (2018) for the cumulative models. Extending the mentioned approach, we evaluate the effect of each explanatory variable, on the location and on the scale components, using the so-called Marginal Effect (ME) measures, see e.g., Greene (2008). The latter measure how a change in a specific covariate  $x_{il}$  ( $z_{il}$ ) affects the response variable when all other covariates are fixed at certain values  $\mathbf{x}_{i \setminus l}^*$  ( $\mathbf{z}_{i \setminus l}^*$ ). For further details and interpretation of ME in ordered response models, see e.g., Greene (2008) and Greene and Hensher (2010).

In this section we report the ME measure of a continuous variable  $x_{il}$  involved only in the location component of the model. The ME on  $\Pr(Y_i = j)$  is given by the partial derivative of  $\Pr(Y_i = j)$  with respect to  $x_{il}$ .

$$ME_{\{Y_i=j, x_{il}\}} = \frac{\partial \Pr(Y_i = j | \mathbf{x}_i = (x_{il}, \mathbf{x}_{i \setminus l}^*), \mathbf{z}_i)}{\partial x_{il}}. \quad (2.2)$$

In Eq. (2.2) the partial derivative of  $\Pr(Y_i = j)$  with respect to  $x_{il}$  indicates the rate of change in  $\Pr(Y_i = j)$  with respect to  $x_{il}$  when other covariates are fixed at value  $\mathbf{x}_{i \setminus l}^*$  and  $\mathbf{z}_i$ . It is obtained by

$$\begin{aligned} & \frac{\partial \Pr[Y_i = j | \mathbf{x}_i = (x_{il}, \mathbf{x}_{i \setminus l}^*), \mathbf{z}_i]}{\partial x_{il}} \\ &= \begin{cases} -\beta_l \times \frac{f_j}{\exp(\mathbf{z}_i \boldsymbol{\gamma})} & j = 1 \\ -\beta_l \times \frac{f_j}{\exp(\mathbf{z}_i \boldsymbol{\gamma})} + \beta_l \times \frac{f_{j-1}}{\exp(\mathbf{z}_i \boldsymbol{\gamma})} & j = 2, \dots, k-1 \\ \beta_l \times \frac{f_{j-1}}{\exp(\mathbf{z}_i \boldsymbol{\gamma})} & j = k \end{cases} \quad (2.3) \end{aligned}$$

where  $f_j = f\left[\frac{(\alpha_j - \mathbf{x}_i \boldsymbol{\beta})}{\exp(\mathbf{z}_i \boldsymbol{\gamma})}\right]$  and  $f_{j-1} = f\left[\frac{(\alpha_{j-1} - \mathbf{x}_i \boldsymbol{\beta})}{\exp(\mathbf{z}_i \boldsymbol{\gamma})}\right]$  are the density function corresponding to the examined cumulative model.

If  $x_{il}$  is a categorical variable, we need to calculate the discrete change. For a dichotomous variable, it is given by

$$ME_{\{Y_i=j, x_{il}\}} = \Pr(Y_i = j | \mathbf{x} = (1, \mathbf{x}_{i \setminus l}^*), \mathbf{z}_i) - \Pr(Y_i = j | \mathbf{x} = (0, \mathbf{x}_{i \setminus l}^*), \mathbf{z}_i).$$

If the number of possible categories is greater than two, the discrete change indicates how the probability of assuming a particular level adjusts when we move from the examined level to the reference one.

When the same covariate influences also the scale component ( $x_{il} = z_{im}$ ), ME measures become

$$\frac{\partial \Pr[Y_i = j \mid \mathbf{x}_i = (x_{il}, \mathbf{x}_{i \setminus l}^*), z_{i \setminus m}^*]}{\partial x_{il}} = \begin{cases} -\frac{[\beta_l + \gamma_m \times (\alpha_j - x_i \beta)]}{\exp(z_i \gamma)} \times f_j & j = 1 \\ -\frac{[\beta_l + \gamma_m \times (\alpha_j - x_i \beta)]}{\exp(z_i \gamma)} \times (f_j - f_{j-1}) & j = 2, \dots, k-1 \\ \frac{[\beta_l + \gamma_m \times (\alpha_j - x_i \beta)]}{\exp(z_i \gamma)} \times f_{j-1} & j = k. \end{cases} \quad (2.4)$$

According to the way we fix the values of the covariates, we can obtain three different types of ME measures. Among the alternatives, extensively reported in Iannario and Tarantola (2021), Long (1997), Long and Freese (2014), Sun (2015), the Average Marginal Effect (AME) are obtained calculating the mean of the marginal effects evaluated on the  $n$  sample values; Marginal Effects at Representative values (MER) are obtained calculating the marginal effect for specific values of interest in the examined study; Marginal Effects at the Mean (MEM) are obtained by computing the marginal effects setting all covariates equal to their mean value. In our analysis, we analyse MER (ME hereafter) measures. Specifically, the mean is used for continuous variables and the reference category is considered for factor ones. Marginal effects and the corresponding posterior distribution can be obtained via the MCMC sampler following Bürkner (2017). MCMC standard errors are also computed following the main content in Flegal and Jones (2011).

## 2.2 Bayesian ordinal superiority measures

Ordinal superiority measures for group comparison were introduced by Agresti and Kateri (2017), and further discussed in Iannario and Tarantola (2021). We will now present an extension of these measures in a Bayesian context. An application to our location-scale model will be presented in Sect. 3.

Given a categorical variable  $d$  playing a role in the scale parameter, we use these measures to compare the probability that an observation from one group  $g_0$  is scored above an independent observation from the alternative group  $g_1$ . Let indicate with  $\mathbf{w}_{\setminus d}$  the set of all covariates except for  $d$ . The following two ordinal superiority measures  $\Delta$  and  $\gamma$  were proposed in Agresti and Kateri (2017)

$$\Delta(\mathbf{w}_{\setminus d}^*) = \Pr(Y_{g_0} > Y_{g_1}) - \Pr(Y_{g_1} > Y_{g_0});$$

$$\gamma(\mathbf{w}_{\setminus d}^*) = \Pr(Y_{g_0} > Y_{g_1}) + \frac{1}{2} \Pr(Y_{g_0} = Y_{g_1}).$$

These two measures are functionally related and they differ by their reference value. A positive value of  $\Delta(\mathbf{w}_{\setminus d}^*)$  indicates that it is more likely to obtain a higher rating in  $g_0$  than in  $g_1$ . The same consideration holds for  $\gamma$ , if it is greater than 0.5.

Their Bayesian estimates can be obtained generating a set of  $S$  samples for the posterior predictive probability distribution. At a generic value  $s$  ( $s = 1, \dots, S$ ) and for a specific value  $\mathbf{w}_{\setminus d}^*$  the ordinal superiority measure  $\Delta$  with respect to the variable  $d$  is given by

$$\begin{aligned}\Delta^s(\mathbf{w}_{\setminus d}^*) &= \Pr^s(Y_{g_0} > Y_{g_1}) - \Pr^s(Y_{g_1} > Y_{g_0}) \\ &= \sum_{l>k} \pi_{0l}^s(\mathbf{w}_{\setminus d}^*) \pi_{1k}^s(\mathbf{w}_{\setminus d}^*) - \sum_{k>l} \pi_{0l}^s(\mathbf{w}_{\setminus d}^*) \pi_{1k}^s(\mathbf{w}_{\setminus d}^*)\end{aligned}$$

where  $\pi_{0j}^s(\mathbf{w}_{\setminus d}^*) = \Pr^s(Y = j \mid d = 0, \mathbf{w}_{\setminus d}^*)$  is the predictive probability distribution obtained from the examined model for  $g_0$ ;  $\pi_{1j}^s(\mathbf{w}_{\setminus d}^*)$  is obtained in a similar way for  $g_1$ , that is  $\pi_{1j}^s(\mathbf{w}_{\setminus d}^*) = \Pr^s(Y = j \mid d = 1, \mathbf{w}_{\setminus d}^*)$ .

The Bayesian estimates of  $\Delta$  is given by

$$\widehat{\Delta(\mathbf{w}_{\setminus d}^*)} = \frac{1}{S} \sum_s \Delta^s(\mathbf{w}_{\setminus d}^*).$$

A value of  $\widehat{\Delta(\mathbf{w}_{\setminus d}^*)}$  greater than zero indicates that it is more likely to obtain a higher rating in  $g_0$  than in  $g_1$ . Alternatively, one can calculate the  $\gamma$  measure having null value equal to 0.5. Its Bayesian estimate under the squared error loss, is given by

$$\widehat{\gamma(\mathbf{w}_{\setminus d}^*)} = \frac{1}{S} \sum_s \gamma^s(\mathbf{w}_{\setminus d}^*) \quad \text{with} \quad \gamma^s(\mathbf{w}_{\setminus d}^*) = (\Delta^s(\mathbf{w}_{\setminus d}^*) + 1)/2.$$

MCMC standard errors can be easily obtained from the estimated values, providing an indication of the variability of the considered measures.

### 3 University curriculum counsellor data analysis

The model discussed above and its implementation is illustrated next on a real case study. The associated data refer to University of Naples Federico II curriculum counsellor and are the result of a well thought-out experimental design. They consist of sample information stratified with respect to Faculty, gender, day of the week. The survey has been replicated in different years even if the research design does not involve repeated observations of the same variables over periods of time; a posteriori check confirmed that proportions of selected students for interviews in different subgroups have been substantially consistent with expected proportions in the strata. The interviewed students were asked to reply to questions regarding the orientation services provided by the 13 faculties. The purpose of these questions was to discover improvement tasks such as in education service and curriculum composition and explore strategies for improving efficiency in increasing the students' satisfaction. The sampled students considered in the analysis were only users of the evaluated service. Respondents were asked to express on a seven-point scale, ranging from 1 = 'Completely Unsatisfied' to 7 = 'Completely Satisfied', opinions with regard to satisfaction. Questionnaires have been filled anonymously and are composed by two main sections. The first one concerns some demographics and students related topics (i.e., *Gender* of the respondent, a dichotomous variable (1 = Female, 0 = Male); *Age*, on continuous scale; *Freq serv*, a factor with levels: 0 = Not-Frequent users, 1 = Frequent users; *Diploma* of admission, a factor with levels: 1 = Classic Studies, 2 = Scientific studies, 3 = Linguistic, 4 = Professional, 5 = Technical/Accountancy, 6 = Others; *Full-time* position, a binary

variable with levels: 1 = Full-time 0 = Otherwise; *Area* of study, a factor variable related to the different Faculties: 0 = Scientific, 1 = Health Science and 2 = Humanistic). The second section is about Likert-type item questions (ratings) related to respondents' satisfaction towards 5 different items: Information, Willingness, Opening Hours, Competence and Global satisfaction. The reported analysis refers to 2002, 2003 and 2004 waves because of homogeneous sampling. For waves 2007 and 2008, three more items have been added with reference to Usefulness, Structure and Advertising of the Service: for an indepth study analyses for data concerning the last waves, see Corduas et al. (2010), Iannario (2008).

The analysed data consist of 2179, 2535, 3183 complete responses for the three waves, respectively. The main information is reported in Table 1 whereas Fig. 1 shows the distribution of the *Area* of study in the three waves.

Among the ratings, we point out results on the evaluation of the Office hours (*Office*)—the duration of office hours (see Table 2). The selection is related to the comparatively larger heterogeneity for this item than the others (see Capeccchi and Piccolo 2010 for further details on this issue). Table 2 also reports summary statistics concerning the global satisfaction expressed in the three waves.

From a preliminary analysis of the data, we identified the following relevant covariates: *Gender* of the respondent; *Age*; *Freq serv*, and *Area*. The dependence between the ratings concerning *Office* hours satisfaction and *Area* of study in the three waves is clearly evident in the mosaic plots displayed in the three panels of Fig. 2. *Area* represents a covariate which affects both location and scale components in the estimated models.

The Bayesian estimates of the location and scale parameters are reported in Table 3 (posterior mean, MCMC Standard Error and 95% credible intervals). These results are obtained via the R package *brms* (Bayesian regression model using “Stan”); see Bürkner (2017). Standard convergence diagnostics have been considered. Possible interactions among covariates were tested, but they yield posterior support to zero.

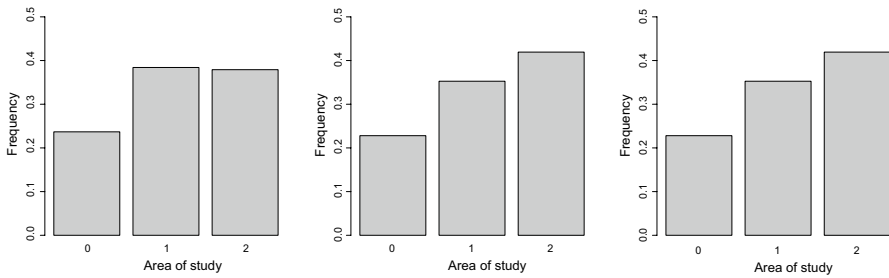
We run in parallel 4 chains of 2000 iteration with a burnin period of 1000 iteration each. The Bayesian estimate of the standard deviation is obtained from the posterior samples of  $\log\text{-disc}$  (log-discrimination) with  $\text{disc}$  corresponding to the inverse of the standard deviation. More precisely, for every iteration  $t$ ,  $t = 1, 2, \dots, T$ , we transformed  $\log(\text{disc})^t$  to  $sd^t$  with  $sd^t = 1/\exp(\log(\text{disc}))$ ; the Bayesian estimates of  $sd$  is then obtained as the average of all  $sd^t$ .

To assess the better performance among models with or without scale effect, we compute the WAIC index reported in (2.1); see Table 4 for the results. Here, with exception of the third wave (2004), where WAIC indexes are closely related, the mentioned indexes support the presence of heterogeneity in the data.

In order to further evaluate the effect of heterogeneity we rely on conditional effects on specific covariates with respect to the response variable. First of all we concentrate upon the variable *Freq serv* affecting only the location component. In Fig. 3 we provide

**Table 1** Summary statistics concerning the characteristics of the respondents of the surveys

| Year | Proportion of female | Average age | Proportion of residence in Naples | Frequency of service | Proportion of scientific and classic high-school (diploma) |
|------|----------------------|-------------|-----------------------------------|----------------------|------------------------------------------------------------|
| 2002 | 0.481                | 22.60       | 0.337                             | 0.350                | 0.616                                                      |
| 2003 | 0.505                | 22.15       | 0.377                             | 0.324                | 0.529                                                      |
| 2004 | 0.519                | 22.51       | 0.400                             | 0.388                | 0.638                                                      |



**Fig. 1** Distribution of the *Area* of study (0 = Scientific, 1 = Health Science and 2 = Humanistic) in the three waves: left (2002), middle (2003), right (2004)

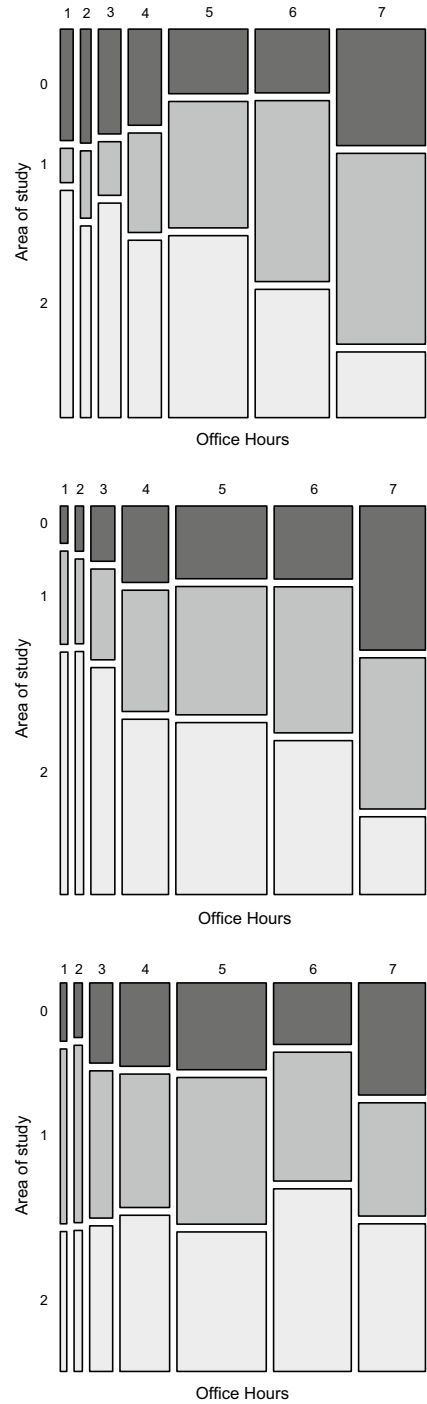
**Table 2** Summary statistics concerning *Office* hours and Global satisfaction

| Year | <i>Office</i> hours satisfaction<br>average–median (SD) | Global satisfaction<br>average–median<br>(SD) |
|------|---------------------------------------------------------|-----------------------------------------------|
| 2002 | 5.281–6 (1.606)                                         | 5.718–6 (1.303)                               |
| 2003 | 5.186–5 (1.443)                                         | 5.620–6 (1.241)                               |
| 2004 | 5.202–5 (1.424)                                         | 5.668–6 (1.211)                               |

a visual representation of the estimated relationship between *Freq serv* and *Office* hours. Each panel displays the estimated probabilities of the seven response categories for the two groups in the three waves. We notice that a higher evaluation is provided by students using more the service with an increasing distance in the three waves, that is students who frequently used the service were more satisfied of *Office* hours than the others. In Fig. 4 we report the evaluation for the different *Area* of study, a covariate affecting both the location and the scale component. First of all, it is possible to observe a reduced evaluation of the service in the three waves checking for the extreme values (pink colour = 7). Health Science students provide the highest reduction of the maximum evaluation in the three waves. Focusing on each wave, students of Scientific area tend to provide a higher evaluation than students of Health Science and Humanistic area. The latter, by inspection of scale effects in Table 3, present for the first wave a lower level of heterogeneity with respect to the cluster of students attending Scientific area. More precisely, the standard deviation of the latent variable for wave 2002 was lower both for Health science ( $Sd = 0.68$ ) and Humanistic area than Scientific one ( $Sd = 0.71$ ) for whom the standard deviation was fixed to 1. For waves 2003 and 2004 the difference with respect to Scientific area was not substantial. Generally a reduction of scale effect is assessed during the 3 years. Fig. 5 summarizes the main findings. The size of the arrows displays the reduced heterogeneity in the three waves, whereas the comparison among *Area* in the three waves supports the comprehensive best performance of the Scientific Area.

All these results are underlined by means of the main findings of the Bayesian ordinal superiority measures in Table 5. Here, the different evaluation of *Office* hours depending on the *Area* of study is further stressed signaling the need of a changing timing for the counsellor services in area different from Scientific one in order to improve the evaluation. The only exception is reserved to Health Science students in the first wave (as also remarked in Fig. 5 and in Table 5). More precisely, for 2002 the order superiority measures,  $\widehat{\Delta}_{01} = -0.1380$  and  $\widehat{\gamma}_{01} = 0.4310$ , indicate a higher probability to observe an elevate

**Fig. 2** Mosaic plot for *Office hours satisfaction* and *Area of study* in the three waves: upper row (2002), middle (2003), lower (2004). Each mosaic plot reports information on the seven-point satisfaction scale versus the area of study (0 = Scientific, 1 = Health Science and 2 = Humanistic)



**Table 3** Bayesian estimation for the location-scale model; i.e., posterior mean estimates, standard deviations and 95% Credible Intervals (CI) for its parameters

|                         | Year 2002     |          |          | Year 2003     |          |          | Year 2004     |          |          |
|-------------------------|---------------|----------|----------|---------------|----------|----------|---------------|----------|----------|
|                         | Estimate (SD) | l-95% CI | u-95% CI | Estimate (SD) | l-95% CI | u-95% CI | Estimate (SD) | l-95% CI | u-95% CI |
| $\hat{\alpha}_1$        | -1.87 (0.22)  | -2.30    | -        | -3.39 (0.34)  | -4.08    | -2.75    | -2.65 (0.27)  | -3.18    | -2.13    |
| $\hat{\alpha}_2$        | -1.36 (0.21)  | -1.79    | -0.94    | -2.67 (0.32)  | -3.32    | -2.05    | -1.79 (0.25)  | -2.26    | -1.30    |
| $\hat{\alpha}_3$        | -0.76 (0.21)  | -1.15    | -0.35    | -1.73 (0.30)  | -2.32    | -1.15    | -0.76 (0.23)  | -1.19    | -0.31    |
| $\hat{\alpha}_4$        | -0.22 (0.21)  | -0.60    | 0.19     | -0.81 (0.29)  | -1.38    | -0.25    | 0.30 (0.23)   | -0.13    | 0.76     |
| $\hat{\alpha}_5$        | 0.68 (0.21)   | 0.27     | 1.10     | 0.39 (0.29)   | -0.17    | 0.96     | 1.55 (0.24)   | 1.09     | 2.01     |
| $\hat{\alpha}_6$        | 1.50 (0.22)   | 1.07     | 1.94     | 1.56 (0.29)   | 1.00     | 2.15     | 2.76 (0.25)   | 2.26     | 3.25     |
| Age                     | 0.03 (0.01)   | 0.01     | 0.04     | 0.02 (0.01)   | -0.01    | 0.04     | 0.05 (0.01)   | 0.03     | 0.07     |
| Gender                  | 0.12 (0.06)   | 0.00     | 0.24     | 0.09 (0.07)   | -0.04    | 0.22     | -0.08 (0.06)  | -0.21    | 0.04     |
| Freq serv               | 0.38 (0.07)   | 0.25     | 0.51     | 0.80 (0.08)   | 0.63     | 0.96     | 1.00 (0.08)   | 0.85     | 1.16     |
| Health Science          | 0.32 (0.10)   | 0.13     | 0.52     | -0.28 (0.09)  | -0.47    | -0.10    | -0.29 (0.09)  | -0.46    | -0.13    |
| Humanistic              | -0.52 (0.10)  | -0.71    | -0.33    | -0.88 (0.10)  | -1.08    | -0.69    | -0.03 (0.08)  | -0.19    | 0.13     |
| log_disc_health science | 0.39 (0.06)   | 0.26     | 0.52     | 0.06 (0.06)   | -0.05    | 0.18     | 0.00 (0.05)   | -0.09    | 0.09     |
| sd_health science       | 0.68 (0.04)   | 0.60     | 0.77     | 0.94 (0.05)   | 0.84     | 1.05     | 1.00 (0.05)   | 0.91     | 1.09     |
| log_disc_humanistic     | 0.35 (0.06)   | 0.23     | 0.46     | 0.11 (0.05)   | 0.00     | 0.21     | 0.02 (0.05)   | -0.07    | 0.11     |
| sd_humanistic           | 0.71 (0.04)   | 0.63     | 0.79     | 0.90 (0.05)   | 0.81     | 1.00     | 0.98 (0.05)   | 0.90     | 1.07     |

**Table 4** Watanabe–Akaike information criterion (WAIC) for the location-scale and cumulative models in the three waves

| Year | Model                | WAIC     |
|------|----------------------|----------|
| 2002 | Location scale model | 7130.2   |
|      | Cumulative model     | 7180.6   |
| 2003 | Location scale model | 8276.9   |
|      | Cumulative model     | 8278.0   |
| 2004 | Location scale model | 10,402.5 |
|      | Cumulative model     | 10,398.9 |

evaluation of the response variable in group 1 (Health Science) respect group 0 (Scientific area).

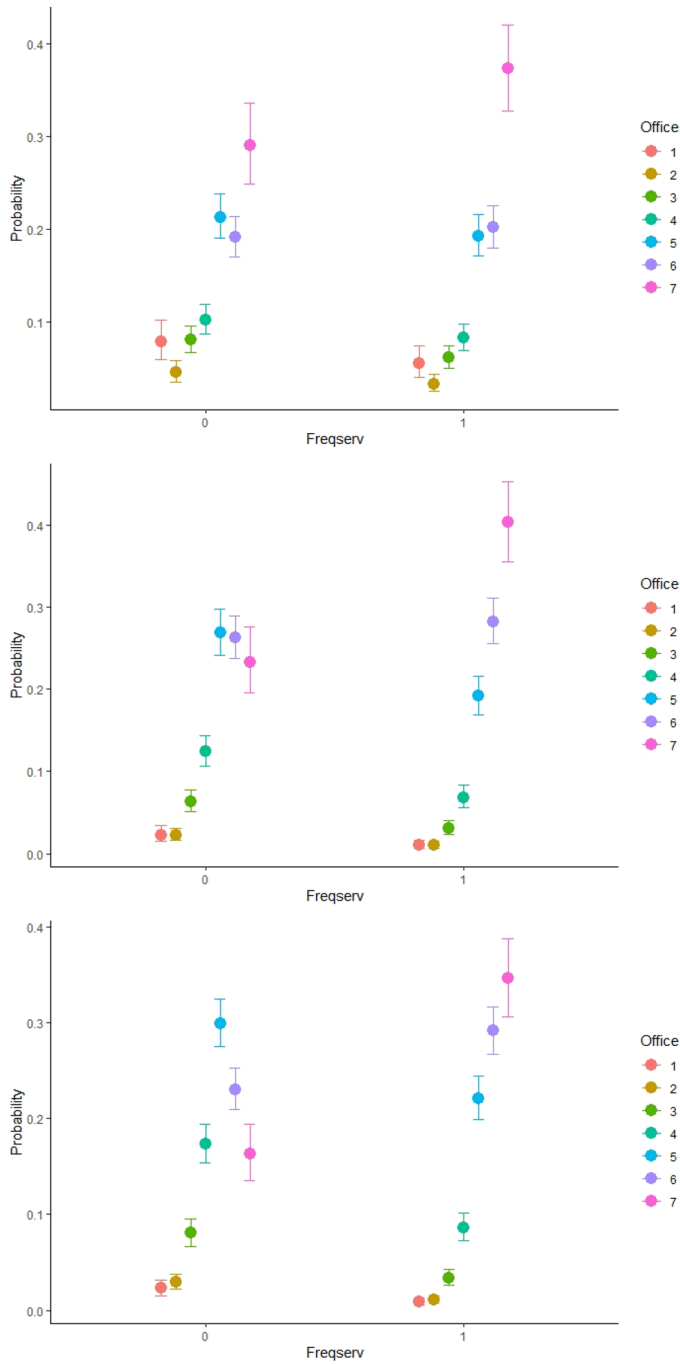
Looking for other covariates, taking into consideration the variable *Age*, in Table 3, we observe that a higher evaluation of the service is provided by older people, whereas for *Gender*, male seems more critical with the service.

## 4 Discussion and conclusion

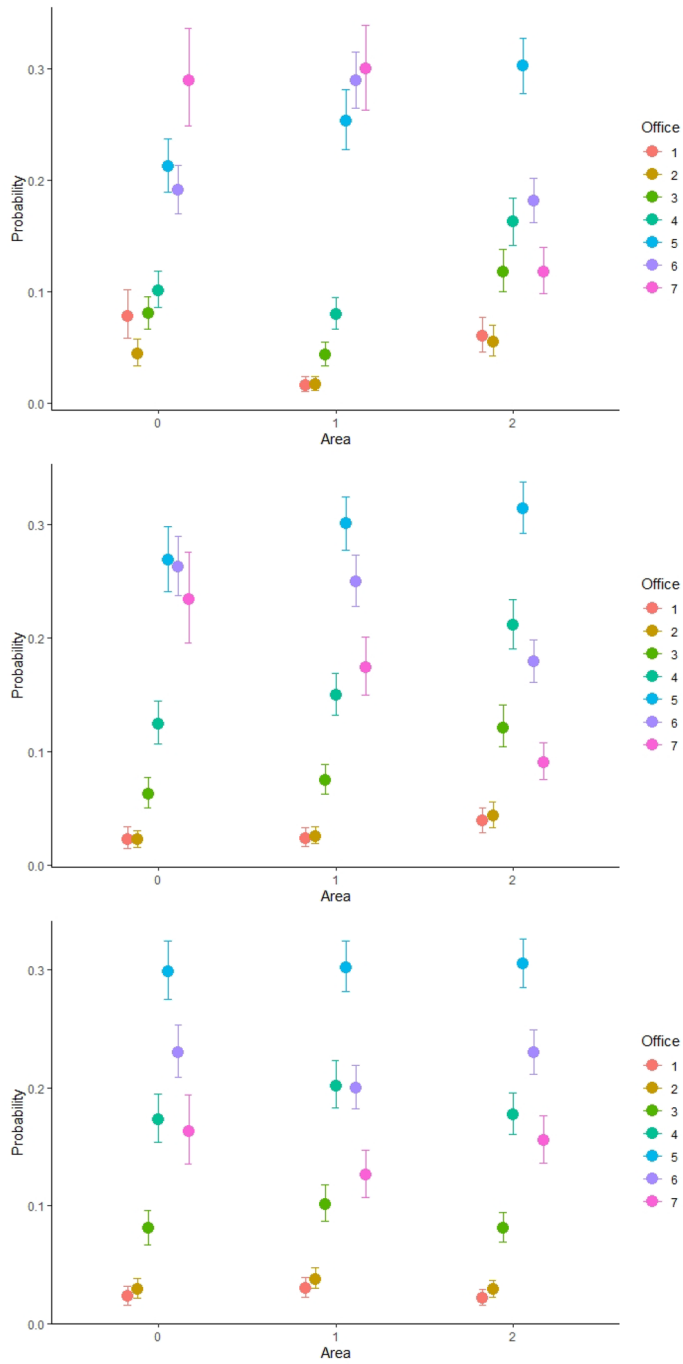
The overall level of satisfaction with university curriculum counsellor focused on students interpreted as ‘consumers’ can accurately diagnose the current status of university services and contribute to improving the quality of them. Our contribution fits this area, analysing data collected in 2002, 2003 and 2004 at University of Naples Federico II. The analysis allowed to detect some trends in the respondents’ behaviour and to relate observed patterns to the subjects’ characteristics. The study has been pursued employing of location-scale models in a Bayesian framework taking into account the presence of heterogeneity and improving the accuracy in providing parameter estimates.

Main results show different evaluation of *Office* hours depending on the *Area* of study signalling the need of a changing timing for the counselling services in area different from the Scientific one. The only exception is reserved to Health Science students in the first wave. Furthermore, male and young students seem more critical with the service. The scale effect remarked on the *Area* for the first two waves reduced in the third ones (2004) where the WAIC index do not discriminate in favour of location-scale framework. The Bayesian analysis allowed improving the flexibility in specifying the model and providing model summaries. Estimates of the proposed order superiority measure, together with their credible intervals, can be easily obtained via the MCMC output. The model can also be easily extended to include prior information deriving from other studies if it becomes available.

Further analyses concerning other aspects of the service may improve the institution overall service strategy. Furthermore, an analysis in which we scrutinize the complex relationships between the latent variable on different levels (faculties), exploiting a multilevel framework, may allow us to study how group membership is expected to influence data analysis results. This represents a future research hypothesis following the framework proposed in Doroshenko and Liseo (2023), Hedeker et al. (2008). Furthermore, the adoption of a Bayesian nonparametric approach (such as for example Barone and Dalla Valle 2023; Turek et al. 2021) for rating data and its comparison to the current results in terms of this case study is a possible follow-up work.



**Fig. 3** Marginal effect for *Freqserv* in the three waves: upper (2002), middle (2003), lower (2004). Points indicate the posterior mean estimates and error bars corresponds to the 95% Credible Intervals



**Fig. 4** Marginal effect for *Area* in the three waves: upper (2002) middle (2003), lower (2004). Points indicate the posterior mean estimates and error bars corresponds to the 95% Credible Intervals

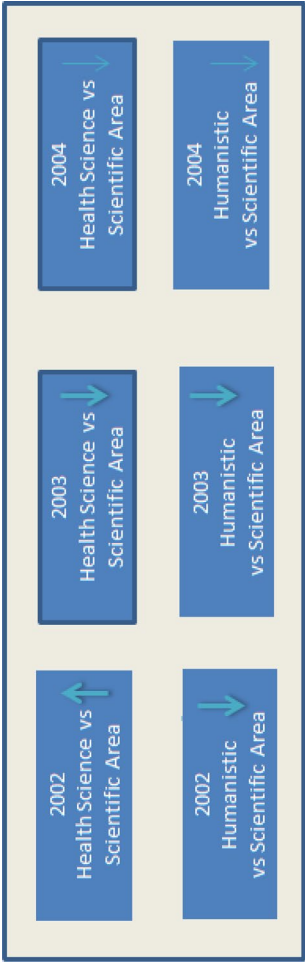


Fig. 5 Summary results concerning *Office* hours satisfaction and *Area* of study in the three waves

**Table 5** Bayesian estimation of ordinal superiority measures; i.e., posterior mean estimates, standard deviations and 95% Credible Intervals (CI) for its parameters

|                | Group          | Year | Estimate | SD     | l-95% CI | u-95% CI |
|----------------|----------------|------|----------|--------|----------|----------|
| $\hat{\Delta}$ | Health science | 2002 | - 0.1380 | 0.0119 | - 0.1456 | - 0.1281 |
| $\hat{\gamma}$ | Health science | 2002 | 0.4310   | 0.0059 | 0.4272   | 0.4359   |
| $\hat{\Delta}$ | Humanistic     | 2002 | 0.1861   | 0.0125 | 0.1789   | 0.1954   |
| $\hat{\gamma}$ | Humanistic     | 2002 | 0.5931   | 0.0063 | 0.5894   | 0.5977   |
| $\hat{\Delta}$ | Health science | 2003 | 0.0937   | 0.0117 | 0.0850   | 0.1007   |
| $\hat{\gamma}$ | Health science | 2003 | 0.5469   | 0.0059 | 0.5425   | 0.5504   |
| $\hat{\Delta}$ | Humanistic     | 2003 | 0.2847   | 0.0117 | 0.2782   | 0.2906   |
| $\hat{\gamma}$ | Humanistic     | 2003 | 0.6423   | 0.0058 | 0.6391   | 0.6453   |
| $\hat{\Delta}$ | Health science | 2004 | 0.0894   | 0.0120 | 0.0799   | 0.0974   |
| $\hat{\gamma}$ | Health science | 2004 | 0.5447   | 0.0060 | 0.5400   | 0.5487   |
| $\hat{\Delta}$ | Humanistic     | 2004 | 0.0115   | 0.0136 | 0.0024   | 0.0215   |
| $\hat{\gamma}$ | Humanistic     | 2004 | 0.5058   | 0.0068 | 0.5012   | 0.5108   |

**Acknowledgement** The authors would like to thank the constructive suggestions of the referees. They would also like to thank Stefania Capecchi and Domenico Piccolo, from the University of Naples Federico II, coordinators and project managers for data collection and preliminary analysis.

**Funding** Open access funding provided by Università degli Studi di Napoli Federico II within the CRUI-CARE Agreement. There is no funding to be reported.

## Declarations

**Conflict of interest** Authors declare that they have no conflict of interest.

**Ethical approval** This article does not contain any studies with human participants or animals performed by any of the authors. Informed consent was obtained from all individual involved in the examined surveys.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Agresti, A.: Analysis of Ordinal Categorical Data, 2nd edn. Wiley, Hoboken (2010)
- Agresti, A., Kateri, M.: Ordinal probability effect measures for group comparisons in multinomial cumulative link models. *Biometrics* **73**, 214–219 (2017)
- Agresti, A., Tarantola, C.: Simple ways to interpret effects in modeling ordinal categorical data. *Stat. Neerl.* **72**, 210–223 (2018)
- Albert, J., Chib, S.: Bayesian methods for cumulative, sequential and two step ordinal data regression models. Technical Report (1997)

- Albert, J.H., Chib, S.: Bayesian analysis of binary and polychotomous response data. *J. Am. Stat. Assoc.* **88**(422), 669–679 (1993)
- Barone, R., Dalla Valle, L.: Bayesian nonparametric modelling of conditional multidimensional dependence structures. *J. Comput. Graph. Stat.* (2023) (in press)
- Bejan, S.A., Janatuinen, T., Jurvelin, J., Klöpping, S., Malinen, H., Minke, B., Vacareanu, R.: Quality assurance and its impact from higher education institutions' perspectives: methodological approaches, experiences and expectations. *Qual. High. Educ.* **21**, 343–371 (2015)
- Betancourt, M., Girolami, M.: Hamiltonian Monte Carlo for hierarchical models. [arXiv:1312.0906](https://arxiv.org/abs/1312.0906) (2013)
- Boes, S., Winkelmann, R.: Ordered response models. *Allgemeines Statistisches Arch.* **90**, 165–179 (2006)
- Brennan, J., Shah, T.: Quality assessment and institutional change: experiences from 14 countries. *High. Educ.* **40**, 331–349 (2000)
- Bürkner, P.: brms: An R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**(1), 1–28 (2017)
- Bürkner, P., Vuorre, M.: Ordinal regression models in psychology: a tutorial. *Adv. Methods Pract. Psychol. Sci.* **2**(1), 77–101 (2019)
- Burnham, K.P., Anderson, D.R.: *Model Selection and Multi-model Inference: A Practical Information-Theoretic Approach*. Springer, New York (2002)
- Capecchi, S., Piccolo, D.: Modelling approaches for ordinal data: the case of orientation service evaluation. *Quaderni di Statistica* **12**, 1–26 (2010)
- Corduas, M., Iannario, M., Piccolo, D.: A class of statistical models for evaluating services and performances. In: Bini, M., Monari, P., Piccolo, D., Salmaso, L. (eds.) *Statistical Methods for the Evaluation of Educational Services and Quality of Products, Contribution to Statistics Series*, pp. 99–117. Springer, Berlin (2010)
- Dellaportas, P., Smith, A.F.M.: Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. *J. R. Stat. Soc. Ser. C* **42**, 443–459 (1993)
- Doroshenko, L., Liseo, B.: Generalized linear mixed model with Bayesian rank likelihood. *Stat. Methods Appl.* **32**, 425–446 (2023)
- Flegal, J.M., Jones, G.L.: Implementing Markov chain Monte Carlo: estimating with confidence. In: Brooks, S., Gelman, A., Jones, G.L., Meng, X. (eds.) *Handbook of Markov Chain Monte Carlo*, pp. 175–197. Chapman & Hall/CRC Press, Boca Raton (2011)
- Gelman, A., Hwang, J., Vehtari, A.: Understanding predictive information criteria for Bayesian models. *Stat. Comput.* **24**, 997–1016 (2014)
- Greene, W.: *Econometric Analysis*, 6th edn. Pearson Prentice Hall, New York (2008)
- Greene, W.H., Hensher, D.A.: *Modeling Ordered Choices: A Primer*. Cambridge University Press, Cambridge (2010)
- Hedeker, D., Mermelstein, R.J., Demirtas, H.: An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics* **64**(2), 627–34 (2008)
- Iannario, M.: A class of models for ordinal variables with covariates effects. *Quaderni di Statistica* **10**, 53–72 (2008)
- Iannario, M., Tarantola, C.: Effect measures for group comparisons in a two-component mixture model: a cyber risk analysis. In: Balzano, S., Porzio, G.C., Salvatore, R., Vistocco, D., Vichi, M. (eds.) *Studies in Classification, Data Analysis and Knowledge Organization*, pp. 97–104. Springer Nature, Switzerland (2021)
- Iannario, M., Monti, A.C., Piccolo, D., Ronchetti, E.: Robust inference for ordinal response models. *Electron. J. Stat.* **11**, 3407–3445 (2017)
- Johnson, V.E.: On Bayesian analysis of multirater ordinal data: an application to automated essay grading. *J. Am. Stat. Assoc.* **91**, 42–51 (1996)
- Johnson, V.E., Albert, J.H.: *Ordinal Data Modeling*. Springer, New York (1999)
- Kanwar, A., Sanjeeva, M.: Student satisfaction survey: a key for quality improvement in the higher education institution. *J. Innov. Entrep.* **11**, 1–10 (2022)
- Kreuzer, A., Dalla Valle, L., Czado, C.: A Bayesian non-linear state space copula model to predict air pollution in Beijing. *J. R. Stat. Soc. Ser. C* **71**, 613–638 (2022)
- Kreuzer, A., Dalla Valle, L., Czado, C.: Bayesian multivariate nonlinear state space copula models. *Comput. Stat. Data Anal.* **188**, 107820 (2023)
- Liddell, T.M., Kruschke, J.K.: Analyzing ordinal data with metric models: What could possibly go wrong? *J. Exp. Soc. Psychol.* **79**, 328–348 (2018)
- Long, J.S.: *Regression Models for Categorical and Limited Dependent Variables*. Sage, California (1997)
- Long, J.S., Freese, J.: *Regression Models for Categorical Dependent Variables Using Stata*. College Station. Stata Press, TX (2014)
- McCullagh, P.: Regression models for ordinal data. *J. R. Stat. Soc. B* **42**, 109–142 (1980)

- Neal, R.M.: MCMC Using Hamiltonian Dynamics. In: Brooks, S., Gelman, A., Jones, G.L., Meng, X. (eds.) *Handbook of Markov Chain Monte Carlo*, pp. 113–162. Chapman & Hall/CRC Press, Boca Raton (2011)
- Peterson, B., Harrell, F.E., Jr.: Partial proportional odds models for ordinal response variables. *J. R. Stat. Soc. Ser. C* **39**, 205–217 (1990)
- Piironen, J., Vehtari, A.: Comparison of Bayesian predictive methods for model selection. *Stat. Comput.* **27**, 711–735 (2017)
- Poister, T.H., Gary, T.H.: Citizen ratings of public and private service quality: a comparative perspective. *Public Adm. Rev.* **54**, 155–160 (1994)
- QAA: QAA Briefing: Student engagement in quality assurance and enhancement (Issue July) (2018). <https://www.qaa.ac.uk/docs/qaa/about-us/qaa-briefing-student-engagement-in-quality-assurance-and-enhancement.pdf>
- Scalera, V., Iannario, M., Monti, A.C.: Robust link functions. *Statistics* **55**, 963–977 (2021)
- Shin, H.J., You, Y.E., Kim, M.Y.: A study on developing and validating a measurement tool for the collegiate education satisfaction. *J. Learner-Centered Curric. Instr.* **20**(21), 139–161 (2020)
- Stander, J., Dalla Valle, L., Taglioni, C., Liseo, B., Cortina-Borja, A.: Analysis of paediatric visual acuity using Bayesian copula model with Sinh–Arcsinh marginal densities. *Stat. Med.* **38**, 3421–3443 (2019)
- Sun, C.: *Empirical Research in Economics: Growing up with R*. Pine Square LLC, Starkville, MS (2015)
- Turek, D., Wehrhahn, C., Gimenez, O.: Bayesian non-parametric detection heterogeneity in ecological models. *Environ. Ecol. Stat.* **28**, 355–381 (2021)
- Tutz, G.: Ordinal regression: a review and a taxonomy of models. *WIREs Comput. Stat.* **14**, e1545 (2022)
- Watanabe, S.: Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* **11**, 3571–3594 (2010)
- Watanabe, S.: A widely applicable Bayesian information criterion. *J. Mach. Learn. Res.* **14**, 867–897 (2013)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.