

A Bivariate Multilevel Ordinal Response Model for the Analysis of University Course Evaluations

Un Modello Multilivello Bivariato a Risposta Ordinale per l'Analisi delle Valutazioni dei Corsi Universitari

Leonardo Grilli

Dipartimento di Statistica “G. Parenti”
Universita' degli Studi di Firenze
grilli@ds.unifi.it

Carla Rampichini¹

Dipartimento di Statistica “G. Parenti”
Universita' degli Studi di Firenze
carla@ds.unifi.it

Riassunto: I giudizi espressi dagli studenti costituiscono un utile strumento per la valutazione dei corsi universitari. L'articolo propone un modello per l'analisi di tali giudizi che consente di tener conto degli aspetti peculiari del fenomeno: la natura ordinale delle valutazioni espresse; la struttura gerarchica dei dati (valutazioni in corsi); la natura multidimensionale del costrutto che si intende misurare; la dipendenza delle valutazioni da caratteristiche e aspettative degli studenti. Il modello proposto è impiegato per l'analisi congiunta di due quesiti del questionario per la valutazione della didattica dell'Universita' di Firenze.

Keywords: multivariate multilevel models, ordinal response, course evaluation.

1. The model

Let $Y_{ij}^{(h)}$ be the h -th observed ordinal variable ($h = 1, 2, \dots, H$) for the i -th subject ($i = 1, 2, \dots, n_j$) of the j -th cluster ($j = 1, 2, \dots, J$). In our application the clusters are the courses, the subjects are the questionnaires and the ordinal variables are the ratings on two items of the questionnaire (i.e. $H = 2$). Now assume that each of the observed ratings $Y_{ij}^{(h)}$, with $c = 1, 2, \dots, C$ categories, is generated by a latent variable $\tilde{Y}_{ij}^{(h)}$ through the following relationship:

$$\{Y_{ij}^{(h)} = c^{(h)}\} \Leftrightarrow \{\gamma_{c^{(h)}-1}^{(h)} < \tilde{Y}_{ij}^{(h)} \leq \gamma_{c^{(h)}}^{(h)}\}, \quad (1)$$

where the thresholds satisfy $-\infty = \gamma_0^{(h)} \leq \gamma_1^{(h)} \leq \dots \leq \gamma_{C-1}^{(h)} \leq \gamma_C^{(h)} = +\infty$. The joint probability $P(Y_{ij}^{(1)} = c^{(1)}, Y_{ij}^{(2)} = c^{(2)})$ is equal to $P(\gamma_{c^{(1)}-1}^{(1)} < \tilde{Y}_{ij}^{(1)} \leq \gamma_{c^{(1)}}^{(1)}, \gamma_{c^{(2)}-1}^{(2)} < \tilde{Y}_{ij}^{(2)} \leq \gamma_{c^{(2)}}^{(2)})$.

Now let us consider the following two-level model for the latent variables: $\tilde{Y}_{ij}^{(h)} = \alpha^{(h)} + \beta^{(h)} \mathbf{x}_{ij}^{(h)} + u_j^{(h)} + \varepsilon_{ij}^{(h)}$, where, for each h , $\mathbf{x}_{ij}^{(h)}$ is a p_h -vector of subject and cluster level covariates (in general, the covariates are partly or totally the same for the two response variables), $\beta^{(h)}$ is the corresponding p_h -vector of fixed parameters, $\alpha^{(h)}$ is the intercept, $u_j^{(h)}$ is the cluster's random effect (level two error) and $\varepsilon_{ij}^{(h)}$ is the subject's disturbance (level one error). From now on, all the distributions and moments are considered conditionally on the $\mathbf{x}$'s, without explicitly indicate it. The errors are assumed to be

¹Financial support from MIUR research project “Statistical education and evaluation: instruments, methods and new technologies” is gratefully acknowledged.

distributed as: $(\varepsilon_{ij}^{(1)}, \varepsilon_{ij}^{(2)})' \stackrel{iid}{\sim} N(\mathbf{0}, \Sigma)$ and $(u_j^{(1)}, u_j^{(2)})' \stackrel{iid}{\sim} N(\mathbf{0}, \Omega)$, with

$$\Sigma = \begin{pmatrix} \sigma_{\varepsilon_1}^2 & \sigma_{\varepsilon_1 \varepsilon_2} \\ \sigma_{\varepsilon_1 \varepsilon_2} & \sigma_{\varepsilon_2}^2 \end{pmatrix}, \quad \Omega = \begin{pmatrix} \tau_1^2 & \tau_{12} \\ \tau_{12} & \tau_2^2 \end{pmatrix}; \quad (2)$$

moreover the first and second level errors are assumed to be independent, so $\text{Cov}_{ij}(\varepsilon_{ij}^{(h)}, u_j^{(k)}) = 0, \forall i, j, h, k$. Therefore Σ is the covariance matrix for a given subject of a given cluster (conditional covariance matrix, where conditional is referred to the cluster's random effects). Then the ratio $\sigma_{\varepsilon_1 \varepsilon_2} / \sigma_{\varepsilon_1} \sigma_{\varepsilon_2}$ can be interpreted as the conditional polychoric correlation (Drasgow (1981)).

The unconditional covariance structure is: $\text{Cov}(\tilde{Y}_{ij}^{(h)}, \tilde{Y}_{i'j'}^{(k)}) = E(\varepsilon_{ij}^{(h)} \varepsilon_{i'j'}^{(k)}) + E(u_j^{(h)} u_j^{(k)})$, with $\text{Cov}(\tilde{Y}_{ij}^{(h)}, \tilde{Y}_{i'j'}^{(k)}) = 0$ if $j \neq j'$, while if $j = j'$ three types of correlation can be defined: (a) the correlation between the same variable for two distinct subjects of the same cluster (intraclass correlation coefficient, ICC): $\text{Corr}(\tilde{Y}_{ij}^{(h)}, \tilde{Y}_{i'j}^{(h)}) = \tau_h^2 / (\sigma_{\varepsilon_h}^2 + \tau_h^2)$, $h = 1, 2$; (b) the correlation between the two variables for the same subject (marginal polychoric correlation): $\text{Corr}(\tilde{Y}_{ij}^{(1)}, \tilde{Y}_{ij}^{(2)}) = (\sigma_{\varepsilon_1 \varepsilon_2} + \tau_{12}) / \sqrt{(\sigma_{\varepsilon_1}^2 + \tau_1^2)(\sigma_{\varepsilon_2}^2 + \tau_2^2)}$; and (c) the correlation between the two variables for two distinct subjects of the same cluster: $\text{Corr}(\tilde{Y}_{ij}^{(1)}, \tilde{Y}_{i'j}^{(2)}) = \tau_{12} / \sqrt{(\sigma_{\varepsilon_1}^2 + \tau_1^2)(\sigma_{\varepsilon_2}^2 + \tau_2^2)}$.

2. Application

The model presented in the previous Section has been used to analyze some of the data gathered in the survey of course quality carried out by the University of Florence, in all the schools of the University, for classes in the 2000-2001 academic year. The questionnaire contains 26 questions which require the same type of ordinal response: 1. decidedly no; 2. more no than yes; 3. more yes than no; 4. decidedly yes. In our application we jointly analyzed two of the questions posed in the questionnaire, that is: *Is the course workload acceptable?* (Q3) *Is the teacher clear?* (Q13). Specifically, we considered the ratings relative to the courses held in the School of Pharmacy, excluding the courses with less than five respondents. We also excluded the questionnaires with a missing response in either of the two items which enter the analysis (3.2% of the total). Altogether, 2888 questionnaires have been considered, corresponding to 87 courses. The number of respondents per course goes from 6 to 136 (median=32, mean=34); the average proportion of respondents is 71%, with a minimum of 29% and a maximum of 100%.

In order to make an initial assessment of the proportion of variance in the ratings which is linked to the course, we fitted, for each of the two considered items, a two-level probit model without covariates (null model).

All the models presented in the paper are fitted with the SAS NLMIXED procedure (SAS (1999)), which maximizes an arbitrary conditional likelihood through a dual quasi-Newton algorithm, integrating out the random effects by means of Gaussian quadrature.

The univariate null models provide an estimate of the intraclass correlation coefficient (ICC). This estimate is very different for the two considered items: the proportion of variability in the evaluations which is attributable to the courses is 0.24 for the course workload and 0.41 for the teacher's clarity.

Now let us consider the bivariate two-level model defined in Section 1, which can be used to jointly analyze the items Q3 and Q13 (Rabe-Hesketh et al. (2001)). The bivariate

model includes the correlation structure between the two items, which is interesting in itself and might also influence the other parameters. In order to identify the model, for each item we fix to zero the first threshold $\gamma(h)$ and to one the first level standard deviation $\sigma_{\varepsilon h}$, i.e. at the first level we estimate only the correlation.

We start the model selection with a bivariate probit model without random effects from which the estimate of the polychoric correlation is 0.4483 (Table 1).

Interpreting the random effects as factors, two alternative hypotheses can be set up for the second level model:

- *two-factor model*: there are two random effects at course level (one for each item), whose variances and covariance can be estimated;
- *one-factor model*: there is a single random effect at course level, entering the two linear predictors with different factor loadings (which correspond to the standard deviations of the random effects).

The one-factor model is a special case of the two-factor model in which the factors are perfectly correlated. Having a single factor is useful in that the courses can be easily ranked on the basis of the predicted values of that factor; however such a model should be used only if supported by the data.

Table 1 reports the estimation results for the one-factor and two-factor null models and for the two-factor model with covariates. In terms of deviances ($-2\log L$) the two-factor null model is clearly preferable over the one-factor null model. Even if the two models lead to a similar estimate of the marginal polychoric correlation (0.477 versus 0.450), in the two-factor model the correlation between the factors (0.62) is farther from unity, which is the value assumed by the one-factor model. The consequences of this incorrect restriction seem to concern mainly the item Q3: in fact in the one-factor model the ICC goes down to 0.056, causing an attenuation in the intercept and thresholds. The better fit of the model with two second level factors agrees with our expectations, since Q3 and Q13 really refer to different aspects of the course. However, the correlation between the factors is quite high and greater than the marginal correlation in the ratings: this could be interpreted in the sense that, at the course level, a good teacher causes the course workload to be perceived as lighter or, viceversa, a light course workload improves the rating of the teacher.

The evaluations expressed by the students are conditioned by individual expectations and traits of the students themselves, so to make fair comparisons among courses it is necessary to net out the influence of each of these components. The last model of Table 1 includes some binary individual covariates derived from other items of the questionnaire, where the code one means a positive evaluation (rating 3 or 4). Moreover, two contextual variables, computed as the course average of some of the previous individual variables, are found to be particularly relevant: the class average grounding in the subject influences the workload (Q3) rating, while the class average interest in the subject influences the teacher's clarity (Q13) evaluation. The introduction of the covariates has reduced both the ICC's and the correlation between the two factors.

With a two-factor model, the ranking of the courses on the basis of the considered items cannot rely upon a single measure. Instead, the predicted values of both factors (second level residuals) should be computed and plotted: the best courses are then the ones lying in the I quadrant, while the worst courses are the ones lying in the III quadrant. Extreme cases should be selected for further investigation.

Table 1: Bivariate models results. The University of Florence, School of Pharmacy, academic year 2000-2001.

	no random eff.		one u factor*		two u factors*		2 u fact. + covar.*	
<i>Parameter</i>	<i>Estim.</i>	<i>s.e.</i>	<i>Estim.</i>	<i>s.e.</i>	<i>Estim.</i>	<i>s.e.</i>	<i>Estim.</i>	<i>s.e.</i>
<i>fixed</i>								
$\alpha^{(1)}$	1.7965	0.0440	1.8094	0.0463	1.9879	0.0632	0.6467	0.1257
$\gamma_2^{(1)}$	0.9460	0.0413	0.9735	0.0424	1.0806	0.0471	1.1268	0.0496
$\gamma_3^{(1)}$	2.4575	0.0479	2.5255	0.0503	2.7848	0.0563	2.8935	0.0594
$\beta_{individual\ homework}^{(1)}$	.	.	.	.	.	.	0.3564	0.0480
$\beta_{grounding}^{(1)}$	.	.	.	.	.	.	0.3765	0.0518
$\beta_{grounding(course\ mean)}^{(1)}$	.	.	.	.	.	.	0.9934	0.1690
$\beta_{interest}^{(1)}$	.	.	.	.	.	.	0.3184	0.0610
$\alpha^{(2)}$	1.3552	0.0329	1.6771	0.0578	1.5710	0.0530	-0.3673	0.1783
$\gamma_2^{(2)}$	0.6931	0.0288	0.9409	0.0383	0.9398	0.0382	0.9978	0.0409
$\gamma_3^{(2)}$	1.7285	0.0357	2.3271	0.0494	2.3318	0.0495	2.4364	0.0525
$\beta_{individual\ homework}^{(2)}$	.	.	.	.	.	.	0.2970	0.0486
$\beta_{grounding}^{(2)}$	.	.	.	.	.	.	0.3717	0.0508
$\beta_{interest}^{(2)}$	.	.	.	.	.	.	0.4862	0.0621
$\beta_{interest(course\ mean)}^{(2)}$	.	.	.	.	.	.	1.7667	0.2255
$\beta_{first.year}^{(2)}$	.	.	.	.	.	.	-0.2132	0.0621
<i>covariance</i>								
$\text{Corr}(\tilde{Y}^{(1)}, \tilde{Y}^{(2)} \mathbf{u})$	.	.	0.4014	0.0227	0.4208	0.0205	0.3758	0.0220
$\text{Corr}(\tilde{Y}^{(1)}, \tilde{Y}^{(2)})$	0.4483	0.0185	0.4500	0.0188	0.4768	0.0245	0.3774	0.0205
$\text{Corr}(u^{(1)}, u^{(2)})$	.	.	1	.	0.6237	0.0503	0.4029	0.0524
ICC ⁽¹⁾	0	.	0.0561	0.0111	0.2448	0.0289	0.1754	0.0197
ICC ⁽²⁾	0	.	0.4258	0.0229	0.5008	0.0248	0.3103	0.0213
-2logL	13058		12029		11738		10980	
n. of param.	7		9		10		19	

obs=2888, *non-adaptive gaussian quadrature with 21 points. The superscripts are: 1=Q3, 2=Q13.

Future work is aimed at extending the model to include all the items in order to identify the total number of latent dimensions which define the perceived quality of the course.

References

- Dragow, F. (1981). Polychoric and polyserial correlations, in: *Encyclopaedia of Statistics*, Johnson and Kotz eds. pp. 68–73.
- Rabe-Hesketh, S., Pickles, A. and Skrondal, A. (2001). *GLLAMM Manual, Technical report*, number 2001/01, Department of Biostatistics and Computing, Institute of Psychiatry, King's College, University of London.
- SAS (1999). *SAS/STAT User's Guide*, version 8. edn, Cary: SAS Institute Inc.