Two logistic random effects models to analyze bounded outcome scores

Inmaculada Arostegui[1] and Vicente Núñez-Antón[2]

[1] Departamento de Matemática Aplicada y Estadística e Investigación Operativa. Universidad del País Vasco UPV/EHU. Leioa, Spain
[2] Departamento de Econometría y Estadística (E.A. III). Universidad del País Vasco UPV/EHU. Bilbao, Spain

Responses that are restricted to finite intervals are called bounded outcome scores. Those can be found in many research areas, as for instance quality of life studies. Health-related quality of life (HRQoL) measures often are restricted to the [0,100] closed interval, in an ordinal scale and with non standard distributions, e.g., J- or U-shaped. We propose a recoding method for the original scores from 0 to n, where n is the number of possible values. The recoded response variable can be considered as grouped data for a dichotomous outcome (0 vs. 1), where correlation within individuals exists, and it represents the number of points obtained by an individual in a HRQoL domain. Therefore, we propose an analysis of such data based on the logistic regression with random effects model. In general, suppose that the ith individual has a set of k regression covariates $\mathbf{x}_i = (x_{i1}, \ldots, x_{ik})^T$ for each i = 1, ..., N. These covariates are linked to the response probability θ_i in a way that depends on the selected methodological approach. Two different models are considered and compared: the binomial-logit-normal regression (BLNR) and the beta-binomial regression (BBR) models.
The BLNR model is given by:

$$logit(\theta_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \sigma u_i, \qquad i = 1, \ldots, N, \tag{1}$$

where, the constant σ is the unknown positive parameter for the random effect and the u_i's are independent and identically distributed random variables with mean zero and variance one.
The BBR model is given by:

$$logit(\theta_i) = \mathbf{x}_i^T \boldsymbol{\beta} + logit(\theta_{0i}), \qquad i = 1, \ldots, N, \tag{2}$$

where the θ_{0i}'s are independent beta random variables, with shape parameters p/ϕ and $(1-p)/\phi$, ($p \in (0,1)$ and $\phi > 0$), which represent the baseline response probability for each subject i.
The two different methods are illustrated on a dataset with the Short Form – 36 (SF–36) HRQoL instrument. The same covariates are selected as significant for both methods and the eight HRQoL domains. Model fit is tested with the AIC and the BIC criteria and both approaches featured very similar results. Distributional hypotheses are tested numerically and graphically. Results for the BBR approach are shown to be better than for the BLNR approach, specially for very skewed HRQoL domains and specifically in the tails of the distribution. The interpretation of the exponentiated regression coefficients in terms of the *odds ratio* shows the usefulness of these two analytical approaches for HRQoL data in epidemiological research.