Thailand Statistician
2015; 13(1): 67-78
http://statassoc.or.th
Contributed paper

The Covariate-Adjusted Frequency Plot for the Rasch Poisson Counts Model

Heinz Holling $[a]$, **Walailuck Böhning** $[a]$ and **Dankmar Böhning** $*[b]$,
$[a]$ Statistics and Quantitative Methods,
Faculty of Psychology and Sports Science
University of Münster, Münster, Germany
$[b]$ Southampton Statistical Sciences Research Institute
Mathematics & Medicine, Southampton, SO17 1BJ, UK
$*$ Corresponding author; e-mail: d.a.bohning@soton.ac.uk

Received: 31 March 2014
Accepted: 7 October 2014

Abstract

The Rasch Poisson Counts model is an appropriate item response theory (IRT) model for analyzing many kinds of count data in educational and psychological testing. The evaluation of a fitted Rasch Poisson model by means of a graphical display or graphical device is difficult and, hence, very much an open problem, since the observations come from different distributions. Hence methods, potentially straightforward in the univariate case, cannot be applied for this model. However, it is possible to use a method, called the *covariate–adjusted frequency plot*, which incorporates covariate information into a marginal frequency plot. We utilize this idea here to construct a covariate-adjusted frequency plot for the Rasch Poisson Counts model. This graphical method is useful in illustrating goodness-of-fit of the model as well as identifying potential areas (items) with problematic fit. A case study using typical data from a frequently used intelligence test illustrates the method which is easy to use.

Keywords: Frequency plot, Goodness-of-fit, Model Diagnostics, Graphical Display, Residual Analysis.

1. Introduction

The Rasch model is the standard model for analyzing count data from speed tests. Compared to other models the Rasch model has the best statistical attributes, e. g. it is the only model where the scores are the sufficient statistics. Hence if the Rasch model fits the data it should be used. The Rasch Poisson Counts model (RPCM), first published in the classical monograph by Rasch [1, 2], may be considered as the first item response theory (IRT) model. This model allows for the analysis of count data which are assumed to be distributed according to a Poisson distribution. Nevertheless the RPCM has by far not gained as much attention as other IRT models such as the logistic 1PL- or 2PL-model. The minor attraction of the RPCM may be explained by the field of previous applications which are mainly confined to very elementary cognitive tasks such as errors in reading tasks (see e. g. Jansen [3]). However, Jansen [3, 4] showed that data from many common intelligence tests can be well modeled by the RPCM.

Statistical inference for the RPCM, as for any other statistical models, requires checking the distributional assumptions. This can be easily accomplished by comparing the estimated, theoretical distribution with the observed distribution, the latter usually displayed in a frequency plot in which the observed frequency f_y of count value y is plotted against y. This is fairly easy when only a few parameters are involved but becomes more complex when covariate information needs to be taken into account. For this reason, Holling *et al.* [5] recently suggested a marginal plot which incorporates covariate information to any available refined degree. For example, it is possible to construct an observed frequency plot that incorporates covariate patterns which are unique for each study participant. This graphical device was called *covariate–adjusted frequency plot (CAFP)*.

More precisely, let Y be random variable with values in $\{0, 1, 2, ...\}$ with the distributional model $P(Y = y) = p_y = p_y(\psi)$ involving a parameter ψ and f_y as the frequency distribution of n observations $y_1, y_2, \cdots, y_n$ (more precisely, f_y is the count of sample values in $y_1, y_2, \cdots, y_n$ equal to y). Given a consistent estimate $\hat{\psi}$ of ψ, usually $p_y(\hat{\psi})n$ is compared with f_y. This is done on the basis that $p_y(\hat{\psi}) \to p_y(\psi)$ and $f_y/n \to p_y(\psi)$. Here $\to$ means convergence in probability.

This principle is generalized in the covariate-adjusted frequency plot for a distributional model incorporating covariates and thus involving potentially many different parameters. Let the count variable Y_i follow a distributional model $p(\mu(\psi, \eta_i))$ with $i = 1, ..., n$ where η_i is a known vector or scalar (typically the values of the covariate(s)), and $\mu(\cdot, \cdot)$ is a known function. A covariate-adjusted frequency is now defined as $\hat{f}_y(\hat{\psi}_n) = \sum_{i=1}^{n} p_y(\hat{\mu}_i)$ and a covariate-adjusted probability as $\hat{p}_y(\hat{\psi}_n) = \frac{1}{n} \sum_{i=1}^{n} p_y(\hat{\mu}_i)$ where $\hat{\mu}_i = \mu(\hat{\psi}_n, \eta_i)$ for $i = 1, ..., n$ and $\hat{\psi}_n$ is a consistent estimate of ψ. The associated plots are constructed by plotting $\hat{f}_y(\hat{\psi}_n)$ and $\hat{p}_y(\hat{\psi}_n)$ against y, respectively. $\hat{f}_y(\hat{\psi}_n)$ is constructed as a marginal operation over the distribution of the η_i. The rationale for this construction is given in Holling *et al.* [5] and provides as the essential result that, given the model is correct, *the covariate adjusted frequency* $\hat{f}_y(\hat{\psi}_n)$ *and* f_y *converge to the same object, hence, they are comparable*.

The purpose of this paper is to apply the ideas of the covariate-adjusted frequency plot to frequency data underlying the RPCM with the ultimate goal of illustrating goodness-of-fit of the model. In the following section the Rasch Poisson Counts model is introduced, followed by the description of the data which are then used to apply the covariate-adjusted frequency plot to the RPCM. The paper ends with a discussion comparing alternative approaches to the covariate-adjusted frequency plot for the RPCM.

2. Rasch Poisson Counts model and the covariate-adjusted frequency plot

Let Y_{ij} denote the count (number of correct solutions, number of errors, number of marks, etc.) for person i and item j. Assume that Y_{ij} follows a Poisson distribution

$$P(Y_{ij} = y) = Po(y|\mu_{ij}) = \exp(-\mu_{ij})\mu_{ij}^y/y!. \tag{1}$$

where μ_{ij} is the expected count for person i and item j, $i = 1, \cdots, n$ and $j = 1, \cdots, k$. The core assumption of the RPCM is the existence of a person-specific ability parameter θ_i and an item-specific easiness parameter λ_j such that the expected value

$$\mu_{ij} = \theta_i \lambda_j \tag{2}$$

is a product of the two parameters. The major problem with this model consists in the number of person parameters which grows with the number of study participants. Therefore, the person parameters have been considered as random effects in extended models. The item parameters λ_j are usually considered as fixed effects as they are relatively few in number and do not increase with the sample size. As a prior for the person parameters Jansen and van Duijn [6] introduced the Gamma distribution which is conjugate to the Poisson distribution. See also Verhelst and Kamphuis [7]. Thus, the arising negative binomial distribution is used for the item scores. In this case, common goodness-of-fit displays may be applied.

Instead of a Gamma distribution Jansen [3] used a log-normal distribution for the person parameters. Under this assumption which seems to be quite natural the distribution of the items scores is not analytically known as in the case of the fixed-effects-model. We will only consider the log-normal distribution for the person parameters since a covariate-adjusted plot can be applied here as well.

The RPCM may be regarded as special log-linear model where

$$\log \mu_{ij} = \beta_0 + \beta_i^P + \beta_j^I + \beta_{ij}^{PI} \tag{3}$$

with the usual constraints $\sum_i \beta_i^P = \sum_j \beta_j^I = 0$ and $\sum_i \beta_{ij}^{PI} = \sum_j \beta_{ij}^{PI} = 0$ in the fixed effects case. Here β_i^P is the main effect of the $i-$th person whereas β_j^I is the main effect of the $j-$th item. Furthermore, β_{ij}^{PI} is the person-item interaction. Note that (3) corresponds to the full model and the RPCM is achieved if $\beta_{ij}^{PI} = 0$ for all i and j. In other words, θ_i corresponds to $\exp(\beta_i^P)$ and λ_j to $\exp(\beta_j^I)$. The model can be fitted using the Poisson likelihood based upon (1)

$$L(\beta_0, \beta_j^I, \beta_i^P) = \prod_i \prod_j Po(y_{ij}|\mu_{ij}). \tag{4}$$

For the RPCM the likelihood becomes in particular

$$\prod_{i=1}^{n} \prod_{j=1}^{k} Po(y_{ij}|\exp(\beta_0 + \beta_j^I + \beta_i^P)). \tag{5}$$

We are now considering the more appropriate random effects model, i. e. considering the person parameters as *random* effects. The likelihood for the RPCM with a log-normally distributed person parameters is then provided as

$$L(\mu, \beta_j^I) = \prod_i \int \prod_j Po(y_{ij}|\exp(\mu + \beta_j^I + \beta_i^P))\phi(\beta_i^P|0, \sigma_P^2)d\beta_i^P, \tag{6}$$

where $\phi(\beta_i^P|0, \sigma_P^2)$ is the normal density with mean 0 and variance σ_P^2.

Note that in the random effects approach the fixed person effects are replaced by normal random effects $\beta_i^P \sim N(0, \sigma_P^2)$. Parameter estimates were found using maximum likelihood choosing the Laplace approximation for solving the marginal integral in the case of the random effects model. All computation were done with the software STATA, version12. For more details on parameter estimation see e. g. Fischer and Molenaar [2].

To apply the covariate-adjusted frequency plot to the RPCM we will use some typical intelligence data from the *Berlin Structure of Intelligence Test for Youth: Assessment of Talent and Giftedness* (BIS-HB see Jäger et al. [8]). This test, based on the Berlin model of Intelligence Structure (BIS; e. g. Jäger [9]), is one of the most comprehensive intelligence tests. It comprises four major intelligence facets *reasoning*, *processing speed*, *memory* and *creativity*. Many of the subtests in the BIS-HB measuring the three last mentioned facets yield count data which could follow the RPCM. Since the BIS-HB consists of a representative sample of intelligence items many subtests of other common intelligence tests or ability tests should be analyzed by the RPCM as well.

As an illustrative example we will use six typical subtests of the BIS-HB measuring processing speed. These tests require simple cognitive operations which have to be accomplished as fast as possible. A typical example is to tick those numbers in a series which are by three greater than the previous number. In an another task missing letters have to be added to incomplete words to achieve correct words. The data are based on a representative sample of $N = 1,327$ high school students from Germany which has been tested to establish the norms of the BIS-HB. The data for the $1,327$ participants on six items result in total $7,962$ observations. Y_{ij} is the count for participant i and item j. Figure 1 shows the empirical frequency distribution f_y for the counts Y_{ij}. This is a simply frequency plot on the basis of all $7,962$ observations and no account is taken for the special structure of the data. It is clear from the bi-modal character that a simple, one-parameter model Poisson model is not appropriate for the data.

One of the *core assumptions* of the RPCM is the additivity (no interaction term). Hence there is considerable interest in a graphical device illustrating the fit of the RPCM. The question is how a fitted frequency $\hat{f}_y$ can be constructed since every observation y_{ij} has it own fitted value $\hat{\theta}_i\hat{\lambda}_j$. This can be now easily accomplished using the concept of the *covariate-adjusted frequency plot.*

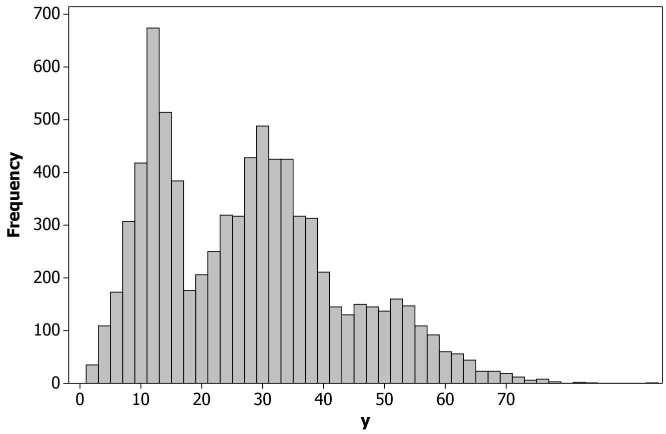


Figure 1: Empirical frequency distribution f_y for counts Y_{ij} of the processing speed data.

To illustrate the covariate-adjusted frequency plot for the RPCM, using $\mu_{ij} = E(Y_{ij}) = \exp(\beta_0 + \beta_i^P + \beta_j^I)$ we have that

$$p_y(\mu(\psi, \eta_{ij})) = Po(y|\exp[\beta_0 + \beta_i^P + \beta_j^I])$$

with $p_y(\mu) = Po(y|\mu)$, $\mu(\psi, \eta_{ij}) = \exp(\beta_0 + \beta_i^P + \beta_j^I)$ and $\boldsymbol{\psi} = (\beta_0, \beta_1^P, \cdots, \beta_n^P, \beta_1^I, \cdots, \beta_k^I)^T$, potentially with the usual constraints $\sum_i \beta_i^P = \sum_j \beta_j^I = 0$ in the fixed effect case. Furthermore,

$$\hat{f}_y = \sum_{i=1}^{n}\sum_{j=1}^{k} Po(y|\hat{\mu}_{ij}) = \sum_{i=1}^{n}\sum_{j=1}^{k} Po(y|\exp[\hat{\beta}_0 + \hat{\beta}_i^{(P)} + \hat{\beta}_j^I]), \tag{7}$$

where k is the number of items and n the number of study participants; $\hat{\mu}_{ij}$, $\hat{\beta}_0$, $\hat{\beta}_i^{(P)}$, $\hat{\beta}_j^I$ are the fitted values under this model. In Figure 2 we see the CAFP for the RPCM using model (3) with a random effects for the person parameters i.e. $\beta_i^P \sim N(0, \sigma_p^2)$. The goodness-of-fit appears reasonably well. It is also possible to the method to investigate the effect of removing certain items of groups of items. In Figure 3 we have removed the item effect entirely and the lack-of-fit is very evident. Hence the item effect is required for this data set. Another option is to investigate individual items separately. This is done by constructed

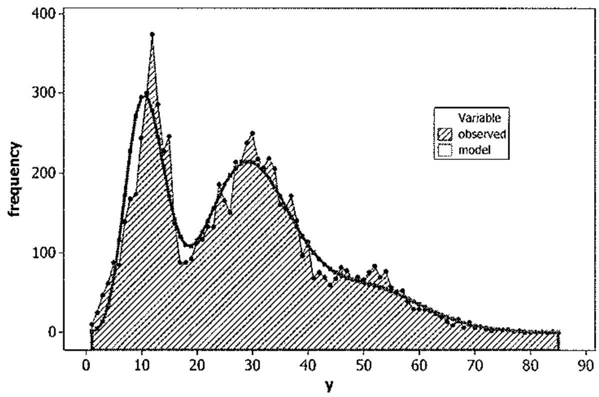


Figure 2: Empirical frequency distribution f_y and covariate adjusted frequency plot $\hat{f}_y$ using random person parameter effects for the processing speed data.

a CAFP for the data consisting only out of the item of interest (observed at the 1,327 participants). This has been done for each of the six items involved in the study and the associated CAFPs for item 1 and item 3 are provided in Figure 4. We see that item 3 experience some problems in reaching a good goodness-of-fit whereas this seems better for item 1.

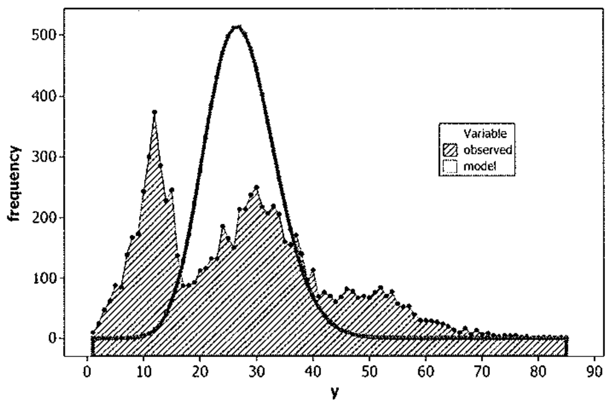
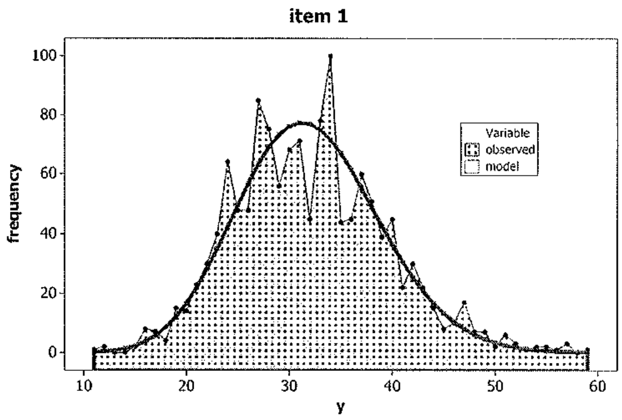


Figure 3: Empirical frequency distribution f_y and covariate adjusted frequency plot $\hat{f}_y$ using the Rasch model without item effects.



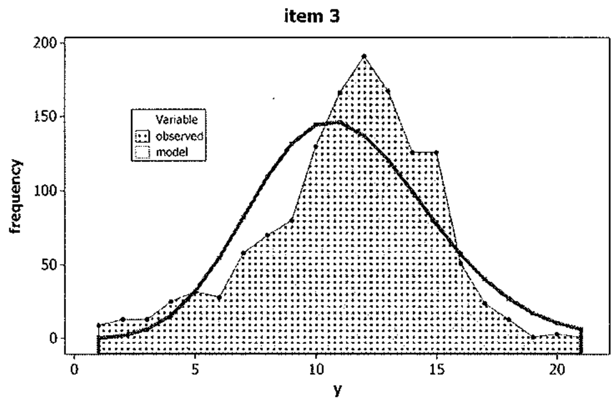


Figure 4: Covariate adjusted frequency plots for the RPCM for items 1 and 3 measuring processing speed.

3. Discussion and extensions

Cameron and Trivedi [10] consider covariate modelling for count outcomes in detail though model evaluation is focusing on residual analysis. Pearson residuals have been discussed by many authors including Lindsey [11], Zelterman [12] or Winkelmann [13]. However, index-plots or Q-Q-plots on the basis of Pearson residuals can be misleading, since even if the model is correct in terms of covariates and distributional assumption the graph might still indicate some deficiencies. In Figure 5 we see an index plot of the full Rasch model fitted for the processing speed data (upper left panel), a standardized residual plot of the Pearson residuals against fitted values (upper right panel), a Q-Q plot of the Pearson residuals (lower left panel) and a Q-Q plot of the Anscombe residual (lower right panel). The latter is suggested in Cameron and Trivedi [10] to adjust for the Poisson nature of the data. The major drawback of these residual diagnostic plots is that they all focus on the normal distribution as the comparison model whereas this is not the model of interest for count data. This is partly acknowledged when using the Anscombe residuals although the difference to the Pearson residuals are minor in the Q-Q plot as Figure 6 (lower panel) shows. See also Augustin, Sauleau and Wood [14] or Ben and Yohai [15] for further discussion.

Finally, it should be noted, that the covariate-adjusted plot can be applied to many other IRT-models based on count data, e.g. to the 1PL-, 2PL- and 3PL-model or linear logistic test models. These models are based instead of a Poisson distribution on a binomial distribution. As can be worked out fairly easy, a covariate-adjusted plot can also be used here to support decision making, e. g. whether to use a 1PL- or 2PL-model.

To widen the scope, let us now leave the area of RPCM and look beyond IRT models and discuss how the CAFP can be applied in other contexts. An issue of illustrating fit (or lack of fit) arises when dealing with *big data* e. g. within large scale assessments. Plotting and, ultimately, looking at individual values becomes problematic due to the very many points that cover the plotting area. We illustrate this with the following data constellation. Suppose a set of values e_i are available, where $i = 1, \cdots, 500,000$. These have been generated from the two-component mixture $e_i \sim 0.5N(10,1) + 0.5N(15,4)$ followed by Poisson counts $Y_i \sim Po(e_i)$. In Figure 6 (upper panels), we see index plots of the standardized residuals $(Y_i - e_i)/\sqrt{e_i}$. These plots are very difficult to interpret due the mass of plots in the graphical area which persists even if a scaling factor of 0.1 is used for the size of the symbols. In the lower panels of Figure 6 we see the frequency plots, on the left panel with the standard Poisson distribution fitted $\hat{f}_y = \exp(-\hat{\mu})\hat{\mu}^y/y!$, $\hat{\mu} = \sum_i y_i/n$, which shows a clear lack of fit. In the right lower panel of Figure 6 we see the covariate-adjusted frequency plot, in this case defined as

$$\hat{f}_y = \sum_{i=1}^{500,000} \exp(-e_i)e_i^y/y!, \tag{8}$$

which gives the correct view, namely that $Y_i \sim Po(e_i)$ is the appropriate model

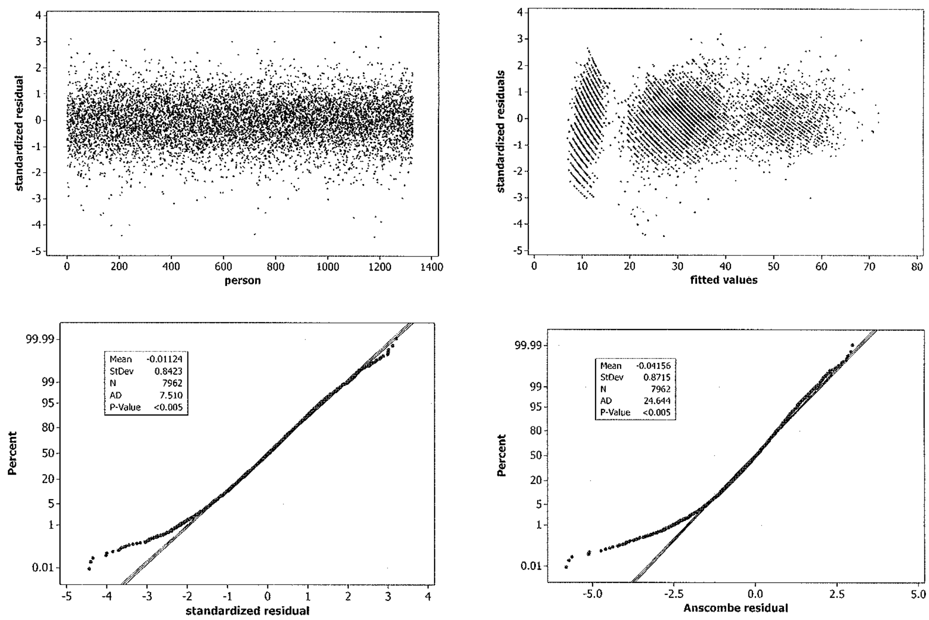


Figure 5: Residual analysis for the processing speed data: index plot of Pearson residual (upper left panel), Pearson residual against fitted value (upper right panel), Q-Q plot of Pearson residuals (lower left panel) and Q-Q plot of Anscombe residuals (lower right panel)

for this data set. Hence, it appears that the covariate-adjusted frequency plot, as a marginal graphical instrument, is also a valuable graphical tool for big or huge data sets since its appearance – in contrasts to residual diagnostic plots – is not affected by large amount of data.

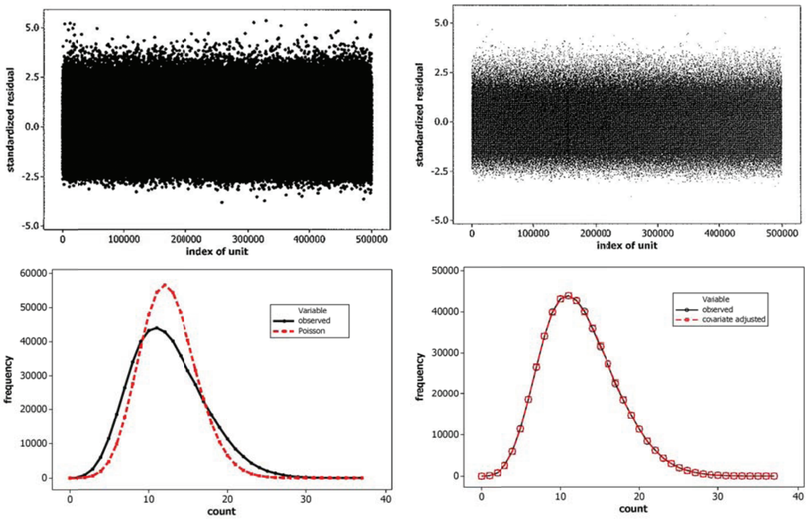


Figure 6: Residual plotting and covariate-adjusted frequency plots for a big data set: index plot of Pearson residual in standard size (upper left panel), with symbols decreased in size with factor 10 (upper right panel), frequency plot with standard Poisson distribution fitted (lower left panel) and frequency plot with covariate-adjusted frequency plot (lower right panel)

Acknowledgements: This work was funded by the *German Research Foundation* (GZ: Ho1286/6-2). The authors are grateful to the Editor, Associate Editor and a referee for their very helpful comments.

References

[1] Rasch, G., *Probabilistic models for some intelligence and attainment tests.* Copenhagen: The Danish Institute of Educational Research, 1960.

[2] Fischer, G.H., & Molenaar, I.W., *Rasch models. Foundations, recent developments, and applications.* New York: Springer, 1995.

[3] Jansen, M.G.H., Parameters of the latent distribution in Raschs Poisson counts model. In G. H. Fischer & D. Laming (Eds.), *Contributions to mathe-*

matical psychology, psychometrics and methodology, New York: Springer, 1994: 319-326.

[4] Jansen, M.G.H., Applications of Raschs Poisson counts model to longitudinal count data, In J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences*, Münster: Waxmann, 1997: 380-389.

[5] Holling, H., Böhning, W., Böhning, D., & Formann, A.K., The covariate-adjusted frequency plot, *Statistical Methods in Medical Research*, 2013; DOI: 10.1177/0962280212473386.

[6] Jansen, M. G. H., & van Duijn, M. A. J., Extensions of Raschs multiplicative Poisson model, *Psychometrika*, 1992; 57: 405-414.

[7] Verhelst, N. D.& Kamphuis, F.H., Poisson-Gamma model for speed tests, *Technical Report 2009-2*, Cito: Arnhem, 2009.

[8] Jäger, A. O., Holling, H., Preckel, F., Schulze, R., Vock, M., Süß, H.-M., & Beauducel, A., *BIS-HB - Berliner Intelligenzstrukturtest für Jugendliche: Begabungs- und Hochbegabungsdiagnostik*, Göttingen: Hogrefe, 2006.

[9] Jäger, A. O., Mehrmodale Klassifikation von Intelligenzleistungen: Experimentell kontrollierte Weiterenwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica*, 1982; 28: 195-225.

[10] Cameron, A.C. and Trivedi P.K., *Regression analysis of count data.* Cambridge: Cambridge University Press, 1996.

[11] Lindsey, J.K., *Modelling frequency and count data*, Oxford: Clarendon Press, 1995.

[12] Zelterman, D., *Models for discrete data*, Oxford: Oxford University Press, 2006.

[13] Winkelmann, R., *Econometric analysis of count data*, New York: Springer, 2003.

[14] Augustin, N.H., Sauleau, E.A., & Wood, S.N., On quantile-quantile plots for generalized linear models, *Computational Statistics and Data Analysis*, 2012; 56: 2404-2409.

[15] Ben, M.G., & Yohai, V., Quantile-quantile plot for deviance residuals in the generalized linear model, *Journal of Computational and Graphical Statistics*, 2004; 13: 36-47.