

Patrícia Martinková; Karel Zvára
Reliability in the Rasch model

Kybernetika, Vol. 43 (2007), No. 3, 315--326

Persistent URL: <http://dml.cz/dmlcz/135776>

Terms of use:

© Institute of Information Theory and Automation AS CR, 2007

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* <http://project.dml.cz>

RELIABILITY IN THE RASCH MODEL

PATRÍCIA MARTINKOVÁ AND KAREL ZVÁRA

This paper deals with the reliability of composite measurement consisting of true-false items obeying the Rasch model. A definition of reliability in the Rasch model is proposed and the connection to the classical definition of reliability is shown. As a modification of the classical estimator *Cronbach's alpha*, a new estimator *logistic alpha* is proposed. Finally, the properties of the new estimator are studied via simulations in the Rasch model.

Keywords: Cronbach's alpha, Rasch model, reliability

AMS Subject Classification: 62F10, 62P25

1. INTRODUCTION

Let us consider the problem of measuring the reliability of a composite measurement such as an educational test. Consider a set of items

$$Y_j = T_j + e_j \quad \text{for } j = 1, \dots, m, \quad (1)$$

where T_j are the unobservable true scores and e_j are the error terms with zero mean and a positive variance, independent from the true scores. The observed overall score is given by $Y = Y_1 + \dots + Y_m$ and the overall unobservable true score is $T = T_1 + \dots + T_m$.

The reliability of such a measurement is defined as the ratio of the variability of the true score to the observed variability, that is

$$R_m = \text{var}(T)/\text{var}(Y). \quad (2)$$

Also, when having two independent measurements $Y_1 = T + e_1$, $Y_2 = T + e_2$ of the same property T , where $\text{var}(e_1) = \text{var}(e_2)$, the reliability can be expressed as the correlation between these two measurements

$$\text{corr}(T + e_1, T + e_2) = \frac{\text{cov}(T + e_1, T + e_2)}{\sqrt{\text{var}^2(Y)}} = \frac{\text{var}(T)}{\text{var}(Y)} = R_m. \quad (3)$$

Since we cannot estimate $\text{var}(T)$, $\text{var}(e)$, nor measure the knowledge by the same test twice and independently, measures to estimate the reliability have been developed.

A widely used characteristic of reliability is called Cronbach’s alpha. It was proposed by Cronbach in [6] as a generalization of Kuder–Richardson formula 20 for binary data (see [9]). Cronbach’s alpha is defined as

$$\alpha_{CR} = \frac{m}{m-1} \frac{\text{var}(Y) - \sum_j \text{var}(Y_j)}{\text{var}(Y)} = \frac{m}{m-1} \frac{\sum \sum_{j \neq k} \sigma_{jk}}{\sum \sum_{j,k} \sigma_{jk}}, \tag{4}$$

where σ_{jk} is the covariance of the pair (Y_j, Y_k) . A pleasant property of Cronbach’s alpha is the fact that this characteristic is easy to estimate from the data simply by using sample variances and sample covariances instead of their population counterparts in (4).

Novick and Lewis have shown in [11] that Cronbach’s alpha is always a lower bound of the reliability and is equal to reliability if, and only if, the test is composed of items that are *essentially tau-equivalent*, that is if for the items’ true scores it holds simultaneously

$$\begin{aligned} \text{var}(T_1) &= \dots = \text{var}(T_m) = \sigma_T^2 \\ \text{corr}(T_j, T_k) &= 1, \quad j, k = 1, \dots, m. \end{aligned} \tag{5}$$

In [13] ten Berge and Zegers came with a series $\mu_0 \leq \mu_1 \leq \dots \leq R_m$ of lower bounds to the reliability, where $\mu_0 = \alpha_{CR}$ is Cronbach’s alpha, and where

$$\mu_1 = \frac{1}{\sum \sum_{j \neq k} \sigma_{jk}} \left(\sum \sum_{j \neq k} \sigma_{jk} + \left(\frac{m}{m-1} \sum \sum_{j \neq k} \sigma_{jk}^2 \right)^{1/2} \right)$$

was proposed by Guttman in [8].

Connection between Cronbach’s alpha and the intraclass correlation coefficient (ICC) in terms of the 2-way ANOVA model was investigated in [3]. ICC itself was deeply studied in [5], from where this work also takes inspiration.

The nonrobustness of sample estimate $\hat{\alpha}_{CR}$ is discussed and a robust estimator of reliability proposed in [14] and more recently in [4].

In this note, we concentrate on the case of educational tests with *dichotomously scored items*. In such a case, the assumptions of the classical model (1) are violated. Therefore, in the next section we propose a new estimate of reliability which should be more appropriate for binary data.

2. ESTIMATION OF RELIABILITY

Interesting findings about Cronbach’s alpha can be made when its sample estimate

$$\hat{\alpha}_{CR} = \frac{m}{m-1} \frac{\sum \sum_{j \neq k} \hat{\sigma}_{jk}}{\sum \sum_{j,k} \hat{\sigma}_{jk}} \tag{6}$$

is further rewritten in terms of the two-way ANOVA mixed-effects model: Let us suppose that the score reached by the i th student in the j th item can be expressed as

$$Y_{ij} = A_i + b_j + e_{ij}, \tag{7}$$

where ability of i th person $A_i \sim N(0, \sigma_A^2)$ is a random variable obeying the normal distribution, b_j is an unknown parameter describing the difficulty of j th item and $e_{ij} \sim N(0, \sigma_e^2)$ is a normally distributed error term, independent from abilities A_p for $p = 1, \dots, n$. In this situation the true score can be expressed as $T_{ij} = A_i + b_j$, and one can easily see, that conditions (5) of essential tau-equivalence are satisfied, and therefore $\alpha_{CR} = R_m$. When considering model (7), the sample estimate $\hat{\alpha}_{CR}$ can be rewritten as

$$\hat{\alpha}_{CR} = \frac{MS_A - MS_e}{MS_A} = 1 - \frac{1}{F_A}, \tag{8}$$

where MS_A and MS_e are the mean squares and F_A is statistic widely used for testing the hypothesis $\text{var}(A) = 0$, either in a fixed effect model (where also student abilities are understood as fixed) or in mixed effect model (7), see [10] p.947. As an interpretation of (8) we can say, that the greater the estimate of reliability $\hat{\alpha}_{CR}$ is, the better the educational test can distinguish between the students. Besides, formula (8) can be used for construction of the confidence interval for Cronbach's alpha (see also [7]).

Nevertheless, Feldt in [7] warns, that for a test with dichotomously scored items, the assumptions of analysis of variance are violated. The distribution of error terms may be far from the normal distribution, and moreover the error term and the true score cannot be considered independent anymore. Therefore, it is a matter of question as to what extent at all the classical estimate Cronbach's alpha (or better said Kuder–Richardson formula 20) is appropriate for tests with dichotomously scored items.

The idea of the present contribution (first mentioned in [15]) is to replace the F-statistic in (8) best suited for normally distributed variables by the analogous statistic appropriate for dichotomous data.

Testing the hypothesis $H_0 : \text{var}(T) = 0$ is equal to testing the submodel B where the score Y_{ij} depends only on the test item (and does not depend on the student's ability) against the model $A + B$ where the score Y_{ij} depends on the student and on the test item. In the fixed-effect model of logistic regression, the appropriate statistic is the difference of deviances in the submodel and in the model

$$X^2 = D(B) - D(A + B), \tag{9}$$

where deviance D is defined as a function of the difference of the log-likelihood for the model and for the saturated model (for details see, e.g. [1], p.139). Statistic (9) has under the null hypothesis asymptotically (for n fixed and m approaching infinity) the $\chi^2(n - 1)$ distribution. Therefore, the proposed estimate is

$$\hat{\alpha}_{\log} = 1 - \frac{n - 1}{X^2}. \tag{10}$$

In the next sections, we study the properties of the proposed estimate (10), which we call *logistic alpha*, in the Rasch model.

3. RELIABILITY IN THE RASCH MODEL

The model used most often for describing dichotomously scored items (in particular in the context of Item Response Theory) is the logit-normal model, called the Rasch model (see [12]). In the Rasch model, the probability of correct response $y_{ij} = 1$ or false response $y_{ij} = 0$ of person i on item j is given by

$$P(Y_{ij} = y_{ij}|A_i) = \frac{\exp[y_{ij}(A_i + b_j)]}{1 + \exp(A_i + b_j)}, \tag{11}$$

where $A_i \sim N(0, \sigma_A^2)$ describes the level of ability of person i , and b_j is an unknown parameter describing the difficulty of item j . The conditional distributions are assumed to be independent. Since no error term is assumed in model (11), the classical definition of reliability (2) is not applicable here.

Inspired by formula (2.3) in [5] we propose to define the reliability of measurement composed of binary data obeying the mixed effect model by the ratio

$$R_m = \frac{\text{var}[E(Y_i|A_i)]}{\text{var}(Y_i)}. \tag{12}$$

Similarly to the classical definition, there is the total observed variability in the denominator, and there is the part of the $\text{var}(Y_i)$ due to variability of A_i in the numerator.

For the classical (mixed-effect two-way ANOVA) model, where

$$E(Y|A) = \text{var}(T)$$

the new definition merges with the classical definition of reliability (2).

Formula (12) can be used for defining reliability for binary data obeying any type of distribution. The formula for the reliability in the Rasch model (11) is following (see the Appendix for detailed derivation):

$$R_m = \frac{\sum_{j=1}^m \sum_{t=1}^m (C_{jt} - D_j D_t)}{\sum_{j=1}^m \sum_{t=1}^m (C_{jt} - D_j D_t) + \sum_{j=1}^m B_j}, \tag{13}$$

where

$$B_j = E_A \frac{e^{A+b_j}}{(1 + e^{A+b_j})^2} = \int_{-\infty}^{\infty} \frac{e^{A+b_j}}{(1 + e^{A+b_j})^2} \frac{1}{\sqrt{2\pi\sigma_A^2}} e^{-\frac{A^2}{2\sigma_A^2}} dA,$$

$$D_j = E_A \frac{e^{A+b_j}}{1 + e^{A+b_j}} = \int_{-\infty}^{\infty} \frac{e^{A+b_j}}{1 + e^{A+b_j}} \frac{1}{\sqrt{2\pi\sigma_A^2}} e^{-\frac{A^2}{2\sigma_A^2}} dA$$

and

$$C_{jt} = E_A \frac{e^{A+b_j}}{1 + e^{A+b_j}} \frac{e^{A+b_t}}{1 + e^{A+b_t}} = \int_{-\infty}^{\infty} \frac{e^{A+b_j}}{1 + e^{A+b_j}} \frac{e^{A+b_t}}{1 + e^{A+b_t}} \frac{1}{\sqrt{2\pi\sigma_A^2}} e^{-\frac{A^2}{2\sigma_A^2}} dA.$$

These integrals cannot be evaluated explicitly, but can be evaluated numerically. Table 1 shows the values of the reliability for some numbers of items L and some

Table 1. Reliability in the Rasch model for different number of items.

Number of items	Variability of abilities σ_A						
	0.01	0.1	0.2	0.5	0.9	2.5	10
L=3	0.00008	0.00741	0.02881	0.15047	0.34335	0.73121	0.94152
<i>SB R₃</i>	<i>0.00008</i>	<i>0.00742</i>	<i>0.02882</i>	<i>0.15054</i>	<i>0.34345</i>	<i>0.73125</i>	<i>0.94153</i>
L=11	0.00028	0.02667	0.09814	0.39386	0.65731	0.90890	0.98335
L=20	0.00050	0.04747	0.16519	0.54160	0.77717	0.94775	0.99077
<i>SB R₂₀</i>	<i>0.00050</i>	<i>0.04746</i>	<i>0.16518</i>	<i>0.54159</i>	<i>0.77716</i>	<i>0.94775</i>	<i>0.99077</i>
L=50	0.00125	0.11078	0.33098	0.74709	0.89711	0.97843	0.99629
<i>SB R₅₀</i>	<i>0.00125</i>	<i>0.11077</i>	<i>0.33095</i>	<i>0.74707</i>	<i>0.89710</i>	<i>0.97843</i>	<i>0.99629</i>
L=100	0.00249	0.19947	0.49735	0.85524	0.94577	0.98910	0.99814
<i>SB R₁₀₀</i>	<i>0.00249</i>	<i>0.19944</i>	<i>0.49731</i>	<i>0.85522</i>	<i>0.94576</i>	<i>0.98910</i>	<i>0.99814</i>

variabilities of student abilities σ_A , when the equidistantly distributed item difficulties between -0.1 and 0.1 of length L are chosen. The values were calculated using the function `integrate` in software `R`, using multiple of ± 25 of the variability σ_A as the limits of integration. The maximum absolute error reached in integrations for $L = 3$, $L = 11$, and $L = 20$ was less than 0.000025 , for $L = 50$ and $L = 100$ it was less than 0.00013 . Table 1 gives an impression that the relationship between reliability and number of items follows the Spearman–Brown formula. This formula for two tests consisting of different numbers (m_1 and m_2) of tau-equivalent items has been proved in the ANOVA model (7) (see [2]) and it says

$$R_{m_2} = \frac{\frac{m_2}{m_1} R_{m_1}}{1 + (\frac{m_2}{m_1} - 1) R_{m_1}}. \tag{14}$$

Emphasized lines in Table 1 named *SB R_{m₂}* are the values R_{m_2} we would get via the Spearman–Brown formula when setting $m_1 = 11$ and taking the values of the bold line **L=11** as R_{m_1} . The question is, whether the differences are due to integration error or not. A theoretical proof for Spearman–Brown formula in the Rasch model would be needed to answer the question.

In Table 2 the true reliabilities are displayed for the case of 11 items, when the item difficulties are unequidistantly distributed with different variability. One can see that the variability of item difficulties has only a slight impact on test reliability when compared with impact of the number of items.

4. SIMULATIONS AND A PRACTICAL EXAMPLE

This study was inspired by the data describing 11 dichotomously scored items in biology. We made, first of all, a simulation study for this case. We studied the

Table 2. Reliability in the Rasch model for different item difficulties.

Item difficulties	Variability of abilities σ_A						
	0.01	0.1	0.2	0.5	0.9	2.5	10
B1	0.00028	0.02667	0.09814	0.39386	0.65731	0.90890	0.98335
B2	0.00027	0.02665	0.09806	0.39368	0.65719	0.90889	0.98334
B3	0.00027	0.02665	0.09806	0.39368	0.65719	0.90889	0.98334
B4	0.00028	0.02667	0.09813	0.39385	0.65730	0.90890	0.98335
B5	0.00027	0.02664	0.09803	0.39360	0.65711	0.90886	0.98334
B6	0.00027	0.02664	0.09802	0.39359	0.65715	0.90889	0.98335
B7	0.00025	0.02439	0.09044	0.37481	0.64241	0.90608	0.98318

B1	=	{-0.1, -0.08, -0.06, -0.04, -0.02, 0, 0.02, 0.04, 0.06, 0.08, 0.1}
B2	=	{-0.1, -0.099, -0.098, -0.097, -0.096, -0.095, -0.094, -0.093, 0, 0.05, 0.1}
B3	=	{-0.1, -0.05, 0, 0.093, 0.094, 0.095, 0.096, 0.097, 0.098, 0.099, 0.1}
B4	=	{-0.1, -0.05, -0.03, 0, 0.02, 0.04, 0.06, 0.07, 0.08, 0.09, 0.1}
B5	=	{-0.1, -0.1, -0.1, -0.1, -0.1, 0, 0.1, 0.1, 0.1, 0.1, 0.1}
B6	=	{-0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1}
B7	=	{-1, -0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8, 1}

number of students of $n = 20$, $n = 30$ and of $n = 50$. Besides, the number of items of $m = 20$ and $m = 50$ was studied. The item difficulties were always taken equidistant between -0.1 and 0.1 . In each case, the number of 55 values of σ_A were chosen so that the resulting 55 reliabilities would cover the interval $\langle 0, 1 \rangle$.

For each of five combinations of number of students and number of items (five figures) and for each of 55 values of σ_A (55 points in the figure) the true reliability was computed via formula (13). Further, the following procedure was repeated 500-times for each point:

1. The set of n students abilities A_i was generated from the $N(0, \sigma_A)$ distribution
2. For each of n abilities A_i , the m scores on the test items were generated from the Rasch model (11)
3. The sample estimate of the Cronbach's alpha (8) and the logistic alpha (10) was computed from the data

For each of 500 sample estimates of Cronbach's alpha and logistic alpha, their average value and sample variance was computed, and finally the bias and mean squared error (MSE) were displayed.

As shown in the enclosed figures, the new estimate gives better results (smaller bias and mean squared error), except for the case of the true reliability value close to 1. The new estimate tends to give very good results for the case when the number

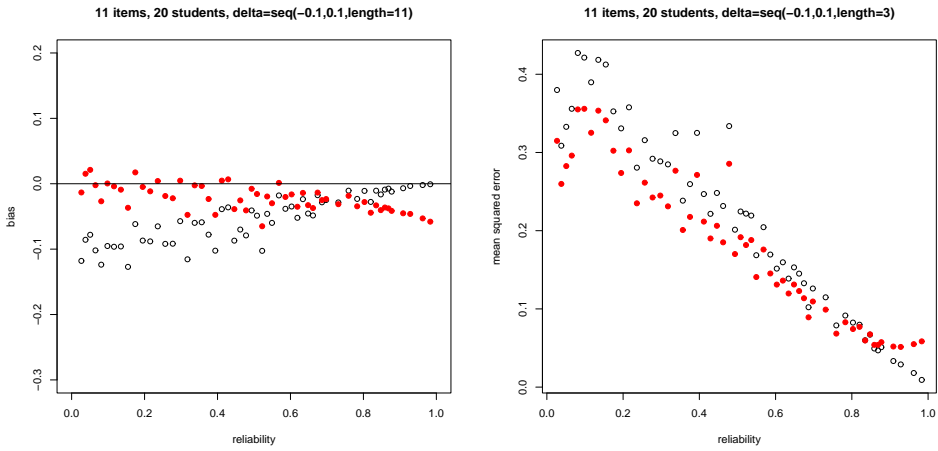


Fig. 1. Bias and MSE for classical (empty circles) and logistic (solid circles) estimator of reliability. Number of students 20, number of items 11.

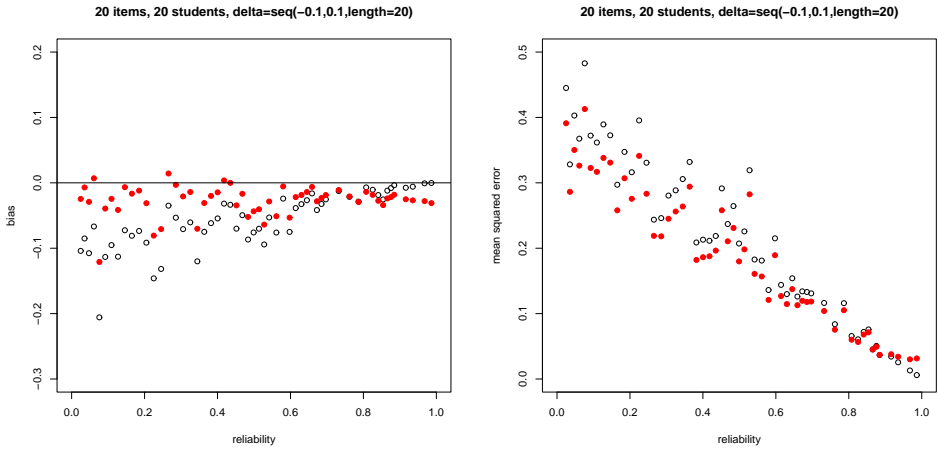


Fig. 2. Bias and MSE for classical (empty circles) and logistic (solid circles) estimator of reliability. Number of students 20, number of items 20.

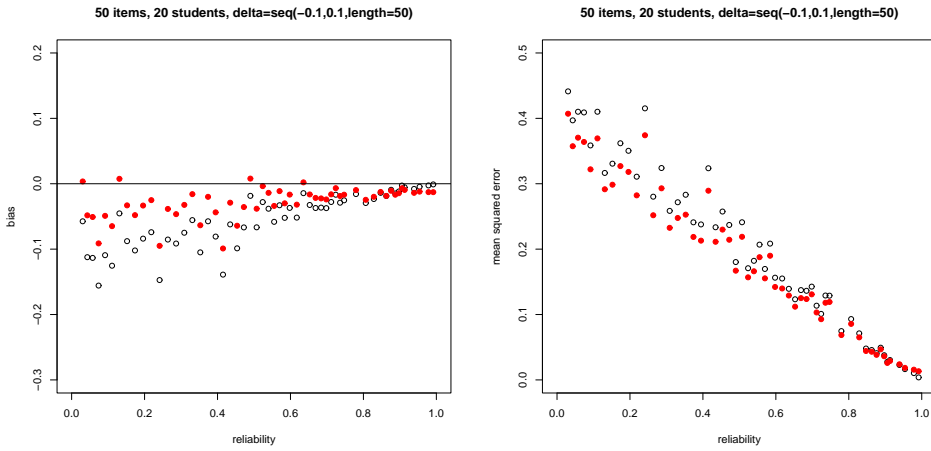


Fig. 3. Bias and MSE for classical (empty circles) and logistic (solid circles) estimator of reliability. Number of students 20, number of items 50.

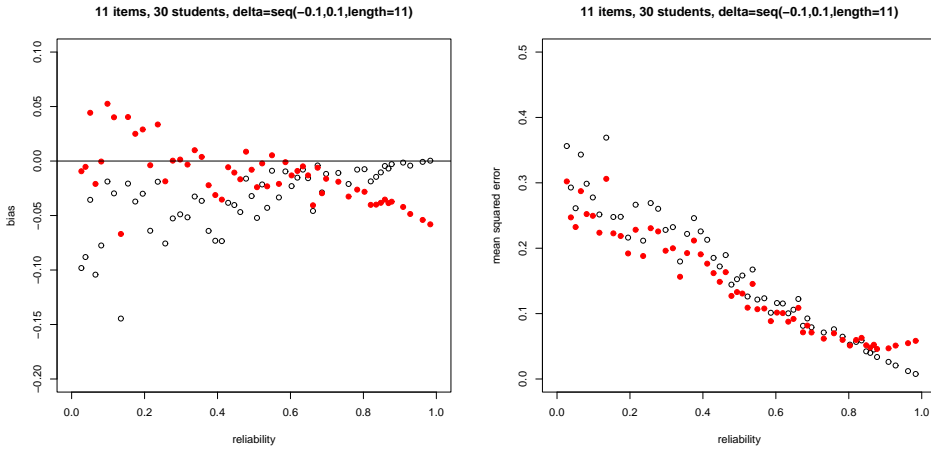


Fig. 4. Bias and MSE for classical (empty circles) and logistic (solid circles) estimator of reliability. Number of students 30, number of items 11.

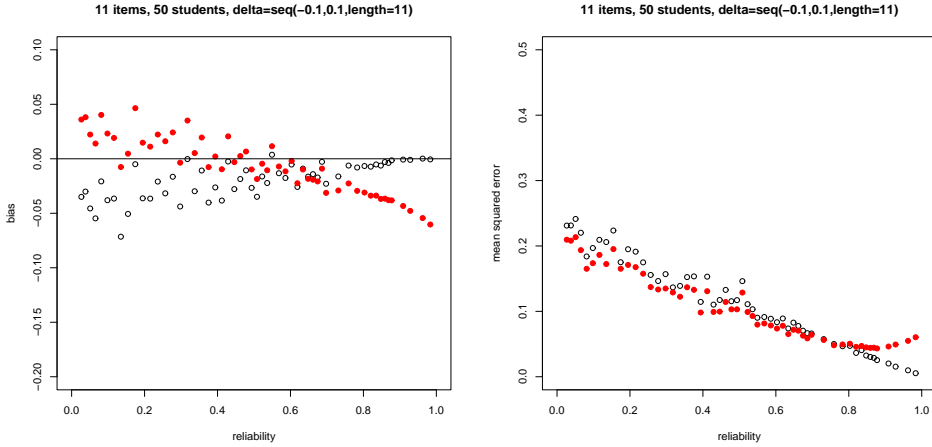


Fig. 5. Bias and MSE for classical (empty circles) and logistic (solid circles) estimator of reliability. Number of students 50, number of items 11.

of items exceeds the number of students (Figure 3). In the case of high number of students in proportion to the number of items (Figure 5), the results of the new estimate are a bit worse.

Let us now look at an example of a real data analysis. We analysed responses of total number of 224 students to a biology test composed of 11 dichotomous items. The students were divided into nine groups (nine classes). In Table 3, we can see the number of students in each group, the estimate of reliability based on Cronbach’s alpha and the estimate of reliability via logistic alpha.

Table 3. Estimation of reliability via Cronbach’s and logistic alpha.

Group	# of students	$\hat{\alpha}_{CR}$	$\hat{\alpha}_{log}$
1	21	0.2492918	0.30021227
2	24	-0.1319623	-0.02784881
3	28	0.2562637	0.29611042
4	21	0.4248927	0.47305085
5	31	0.3694124	0.40444782
6	25	0.6944165	0.68331857
7	24	0.2213452	0.26716854
8	23	0.4833022	0.53131265
9	27	0.3570429	0.44172929

In group number 2, both logistic and Cronbach’s alpha gave a negative estimate of reliability. This was caused by small variability of total scores reached by students

in this group. Except one group, the logistic alpha gave always higher estimate of reliability than the Cronbach's alpha. This might be an example of underestimation of reliability by Cronbach's alpha. Nevertheless, the real data examples do not tell us much about which of the two estimates is better, since we do not know the true value of the reliability.

5. CONCLUSIONS AND DISCUSSION

While the classical definition of reliability (2) is not appropriate for mixed effect models of binary data, we proposed a new definition of reliability (12), which is shown to have the same properties as the classical definition. The new definition merges with the classical definition for the classical model (1).

As a counterpart to the classical estimator of reliability Cronbach's alpha (4), which is based on F-statistics appropriate for continuous data, a new estimate named *logistic alpha* (10), appropriate for binary data, is proposed.

In simulations in the Rasch model, the new estimate gave better results (smaller bias and mean squared error), except for the case of real reliability values close to 1. In particular, the new estimate gave better results for the case of a high number of items compared to the number of students. The results of the new estimate tend to be worse for the case of high number of students in proportion to the number of items.

Further work should contain a study of the theoretical properties of the new estimate in the case of null hypothesis $H_0 : R_m = 0$ and also in the case when the alternative $H_1 : R_m > 0$ holds. This could lead to improvement of the proposed estimate *logistic alpha* for true values of reliability close to 1.

APPENDIX: DERIVATION OF THE RELIABILITY IN THE RASCH MODEL

The Rasch model is defined by

$$P(Y_{ij} = 1|A_i) = E(Y_{ij}|A_i) = \frac{e^{A_i+b_j}}{1 + e^{A_i+b_j}} = p_{ij},$$

where $A_i \sim N(0, \sigma_A^2)$ describes the ability of the i th student, $i = 1, \dots, n$ and b_j are fixed unknown parameters describing difficulty of the j th item $j = 1, \dots, m$. Therefore the conditional variance is

$$\text{var}(Y_{ij}|A_i) = p_{ij}(1 - p_{ij}) = \frac{e^{A_i+b_j}}{(1 + e^{A_i+b_j})^2},$$

and its mean value is

$$E \text{var}(Y_{ij}|A_i) = \int_{-\infty}^{\infty} \frac{e^{A+b_j}}{(1 + e^{A+b_j})^2} \frac{1}{\sqrt{2\pi\sigma_A^2}} e^{-\frac{A^2}{2\sigma_A^2}} dA = B_j.$$

The unconditional mean value can be written as

$$E Y_{ij} = E E(Y_{ij}|A_i) = E \frac{e^{A+b_j}}{1 + e^{A+b_j}} = \int_{-\infty}^{\infty} \frac{e^{A+b_j}}{1 + e^{A+b_j}} \frac{1}{\sqrt{2\pi\sigma_A^2}} e^{-\frac{A^2}{2\sigma_A^2}} dA = D_j.$$

Similarly, for the total score of the i th student $Y_i = \sum_{j=1}^m Y_{ij}$ it holds that

$$E Y_i = E \sum_{j=1}^m Y_{ij} = \sum_{j=1}^m E Y_{ij} = \sum_{j=1}^m D_j$$

and the unconditional variance is

$$\begin{aligned} \text{var } Y_i &= \text{var } E(Y_i|A_i) + E(\text{var}(Y_i|A_i)) \\ &= \text{var } E\left(\sum_{j=1}^m Y_{ij}|A_i\right) + E \text{var}\left(\sum_{j=1}^m Y_{ij}|A_i\right) \\ &= \text{var} \sum_{j=1}^m E(Y_{ij}|A_i) + E \sum_{j=1}^m \text{var}(Y_{ij}|A_i) \\ &= \sum_{j=1}^m \sum_{t=1}^m E(E(Y_{ij}|A_i)E(Y_{it}|A_i)) \\ &\quad - \sum_{j=1}^m \sum_{t=1}^m E(E(Y_{ij}|A_i))E(E(Y_{it}|A_i)) + \sum_{j=1}^m E \frac{e^{A_i+b_j}}{(1 + e^{A_i+b_j})^2} \\ &= \sum_{j=1}^m \sum_{t=1}^m (C_{jt} - D_j D_t) + \sum_{j=1}^m B_j, \end{aligned}$$

where the third equation holds because of assumption of independence of conditional distributions and

$$C_{jt} = E(E(Y_{ij}|A_i)E(Y_{it}|A_i)) = \int_{-\infty}^{\infty} \frac{e^{A+b_j}}{1 + e^{A+b_j}} \frac{e^{A+b_t}}{1 + e^{A+b_t}} \frac{1}{\sqrt{2\pi\sigma_A^2}} e^{-\frac{A^2}{2\sigma_A^2}} dA.$$

Therefore the reliability in the Rasch model can be written as

$$R_m = \frac{\text{var } E(Y_i|A_i)}{\text{var } Y_i} = \frac{\sum_{j=1}^m \sum_{t=1}^m (C_{jt} - D_j D_t)}{\sum_{j=1}^m \sum_{t=1}^m (C_{jt} - D_j D_t) + \sum_{j=1}^m B_j}. \tag{15}$$

ACKNOWLEDGEMENT

This work was supported by projects 1M06014 and MSM 0021620839 of the Ministry of Education, Youth and Sports of the Czech Republic.

(Received December 7, 2006.)

REFERENCES

-
- [1] A. Agresti: *Categorical Data Analysis*. Wiley, New York 2002.
 - [2] P. D. Allison: A simple proof of the Spearman–Brown formula for continuous test lengths. *Psychometrika* 40 (1975), 135–136.
 - [3] G. Bravo and L. Potvin: Estimating the reliability of continuous measures with Cronbach’s alpha or the intraclass correlation coefficient: Toward the integration of two traditions. *J. Clin. Epidemiol.* 44 (1991), 381–390.
 - [4] A. Christmann and S. Van Aelst: Robust estimation of Cronbach’s alpha. *J. Multivariate Anal.* 97 (2006), 1660–1674.
 - [5] D. Commenges and H. Jacqmin: The intraclass correlation coefficient distribution-free definition and test. *Biometrics* 50 (1994), 517–526.
 - [6] L. J. Cronbach: Coefficient alpha and the internal structure of tests. *Psychometrika* 16 (1951), 297–334.
 - [7] L. S. Feldt: The approximate sampling distribution of Kuder–Richardson reliability coefficient twenty. *Psychometrika* 30 (1965), 357–370.
 - [8] L. A. Guttman: A basis for analyzing test-retest reliability. *Psychometrika* 30 (1945), 357–370.
 - [9] G. Kuder and M. Richardson: The theory of estimation of test reliability. *Psychometrika* 2 (1937), 151–160.
 - [10] J. Neter, W. Wasserman, and M. H. Kutner: *Applied Linear Statistical Models*. Richard D. Irwin, Homewood, Il. 1985.
 - [11] M. R. Novick and C. Lewis: Coefficient alpha and the reliability of composite measurement. *Psychometrika* 32 (1967), 1–13.
 - [12] G. Rasch: *Probabilistic Models for Some Intelligence and Attainment Tests*. The Danish Institute of Educational Research, Copenhagen 1960.
 - [13] J. M. F. ten Berge and F. E. Zegers: A series of lower bounds to the reliability of a test. *Psychometrika* 43 (1978), 575–579.
 - [14] R. R. Wilcox: Robust generalizations of classical test reliability and Cronbach’s alpha. *British J. Math. Statist. Psych.* 45 (1992), 239–254.
 - [15] K. Zvára: Measuring of reliability: Beware of Cronbach. (Měření reliability aneb bacha na Cronbacha, in Czech.) *Inform. Bull. Czech Statist. Soc.* 12 (2002), 13–20.

Patřicia Martinková, Center of Biomedical Informatics, Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 2, 182 07 Praha 8. Czech Republic.

e-mail: martinkova@euromise.cz

Karel Zvára, Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University in Prague, Sokolovská 83, 186 75 Praha 8. Czech Republic.

e-mail: karel.zvara@mff.cuni.cz