ACTA UNIVERSITATIS PALACKIANAE OLOMUCENSIS FACULTAS RERUM NATURALIUM

1993 Mathematica XXXII Vol. 110

VARIANT II OF PROBABILISTIC MODEL OF SCHOOL–ACHIEVEMENT TEST WITH DOUBLE CHOICE RESPONSE

EVA TESAŘÍKOVÁ

(Received October 10, 1992)

Abstract

In connection with the analysis of achievement tests with double choice response it is possible to construct two variants of mathematical models describing the probability structure of such tests. In contrast to the Variant I presented in [1], Variant II proceeds from the assumption that in the case of partial knowledge the person under examination can be really familiar only with one correct answer among the offered alternatives. The proportion of the topic to be examined, which the tested person really does not know, can be expressed in the form of a weighted sum $\tau = \delta/2 + \nu$, where ν is the proportion of total unfamiliarity and δ is the proportion of partial unfamiliarity.

Key words: school-achievement test, double-choice response, probabilistic model.

MS Classification: 62P10, 62P15

1 Introduction

The construction of mathematical models, describing the probability structure of school-achievement tests ,should take into account a certain difference between the real knowledge of the examined person and the result of the test evaluated by the examiner.

Variant I of a probabilistic model of the school-achievement test with double choice response, published by the author of this paper in 1991 proceeds from the assumption that when the person under examination is really familiar with the topic of the question, he or she will select both the correct responses from the offered alternatives.

The aim of the present paper is to construct variant II of the probabilistic model of such a test, proceeding from the assumption, that in the case of a partial knowledge the examinee can be really familiar only with one correct response among offered alternatives.

The present paper is the third from a series of articles dealing with this problem.

2 Assumption of the model

Let us consider the school-achievement test with compulsory choice of two correct responses from $q > 3$ response alternatives of which two are correct. A missing answer is evaluated as an incorrect one. The examined persons are informed about the nature of the test before testing. It is also presumed that the test consists of n independent questions of the same difficulty, the number of offered alternatives is the same in all questions and that the function of the alternatives is equivalent . Alternative responses to each question should be chosen in order to avoid similarity and discrepancy.

The whole proportion τ of the topic to be examined which the tested person really does not know can be expressed in the form of a weighted sum $\tau = \delta/2+\nu$, where ν is the proportion of total unfamiliarity and δ is the proportion of partial unfamiliarity.

3 Model of probabilistic structure of the test

The description of the probabilistic structure of the test comprises the following notation for random events relative to the i-th question of the test:

Z_i tested person is total familiar with the topic of the i-th question

D_i tested person is partial familiar with the topic of the i-th question

N_i tested person is unfamiliar with the topic of the i-th question

According to the assumptions mentioned above the random events Z_i, D_i, N_i have probabilities

$$P(Z_i) = 1 - \nu - \delta, \quad P(D_i) = \delta, \quad P(N_i) = \nu$$

In relation to the registered results of the test the following three random events are considered

S_{i0} no correct answer was given to i-th question

S_{i1} only one correct answer was given to i-th question

S_{i2} both correct answers were given to i-th question

In the case of total familiarity with the given topic, the examined person will give two correct responses, so that

$$P(S_{i0}|Z_i) = 0, \quad P(S_{i1}|Z_i) = 0, \quad P(S_{i2}|Z_i) = 1,$$

In the case of partial familiarity with the topic of the i-th question, the examinee actually knows only one correct response, while his second response is effectively chosen at random by him from among the $q-1$ remaining alternatives. The following conditional probabilities discribe this situation

$$P(S_{i0}|D_i) = 0, \quad P(S_{i1}|D_i) = \frac{q-2}{q-1}, \quad P(S_{i2}|D_i) = \frac{1}{q-1}$$

In the case of unfamiliarity with the topic tested by the i-th question the examined person can use only the random choice. With regard to the presence of two correct responses among $q > 3$ offered alternatives the following relations hold:

$$P(S_{i0}|N_i) = \frac{\binom{q-2}{2}}{\binom{q}{2}} = \frac{(q-2)(q-3)}{q(q-1)}$$

$$P(S_{i1}|N_i) = \frac{\binom{q-2}{1}\binom{2}{1}}{\binom{q}{2}} = \frac{4(q-2)}{q(q-1)}$$

$$P(S_{i2}|N_i) = \frac{\binom{2}{2}}{\binom{q}{2}} = \frac{2}{q(q-1)}$$

These conditinal probabilities are defined by the hypergeometric distribution which is applicable in this situation.

Unconditional probabilities of evets S_{i0}, S_{i1}, S_{i2} can be calculated according to the theorem of total probability:

$$\begin{aligned}
p_{i0} &= P(S_{i0}) \;=\; P(S_{i0}\cap Z_i)+P(S_{i0}\cap D_i)+P(S_{i0}\cap N_i) \;= \\
&= P(Z_i)P(S_{i0}|Z_i)+P(D_i)P(S_{i0}|D_i)+P(N_i)P(S_{i0}|N_i) \;= \qquad (1)\\
&= \nu\frac{(q-2)(q-3)}{q(q-1)}
\end{aligned}$$

$$\begin{aligned}
p_{i1} &= P(S_{i1}) \;=\; P(S_{i1}\cap Z_i)+P(S_{i1}\cap D_i)+P(S_{i1}\cap N_i) \;= \\
&= P(Z_i)P(S_{i1}|Z_i)+P(D_i)P(S_{i1}|D_i)+P(N_i)P(S_{i1}|N_i) \;= \qquad (2)\\
&= \delta\frac{q-2}{q-1}+\nu\frac{4(q-2)}{q(q-1)}
\end{aligned}$$

$$\begin{aligned}
p_{i2} &= P(S_{i2}) \;=\; P(S_{i2}\cap Z_i)+P(S_{i2}\cap D_i)+P(S_{i2}\cap N_i) \;= \\
&= P(Z_i)P(S_{i2}|Z_i)+P(D_i)P(S_{i2}|D_i)+P(N_i)P(S_{i2}|N_i) \;= \qquad (3)\\
&= (1-\nu-\delta)+\delta\frac{1}{q-1}+\nu\frac{2}{q(q-1)} \;=\; 1-\delta\frac{q-2}{q-1}-\nu\frac{q(q-1)-2}{q(q-1)}
\end{aligned}$$

The assumption of the same difficulty of questions and the same schema of offered answers allows us to omit the index of the question and to introduce the following notation

$$p_{i0}=p_0,\quad p_{i1}=p_1,\quad p_{i2}=p_2,\quad S_{i0}=S_0,\quad S_{i1}=S_1,\quad S_{i2}=S_2$$

for all considered indices.

To the whole test, i.e. the whole series of n independent questions we can associate three random variables M_0, M_1, M_2.

Variable M_0 represents the number of questions in the whole test to which no correct answer was given. This variable has a binomial probability distribution defined by the relation

$$P(M_0=m_0)=\binom{n}{m_0}\left(\nu\frac{(q-2)(q-3)}{q(q-1)}\right)^{m_0}\left(1-\nu\frac{(q-2)(q-3)}{q(q-1)}\right)^{n-m_0} \qquad (4)$$

with the expectation

$$E(M_0)=n\nu\frac{(q-2)(q-3)}{q(q-1)} \qquad (5)$$

and with the variance

$$D(M_0)=n\nu\frac{(q-2)(q-3)}{q(q-1)}\left(1-\nu\frac{(q-2)(q-3)}{q(q-1)}\right) \qquad (6)$$

Variable M_1 represents the number of questions to which only one correct answer was given. This variable has a binomial probability distribution defined by relation

$$P(M_1 = m_1) =$$
$$= \binom{n}{m_1} \left(\delta \frac{q-2}{q-1} + \nu \frac{4(q-2)}{q(q-1)} \right)^{m_1} \left(1 - \delta \frac{q-2}{q-1} - \nu \frac{4(q-2)}{q(q-1)} \right)^{n-m_1} \quad (7)$$

with the expectation

$$E(M_1) = n \left(\delta \frac{q-2}{q-1} + \nu \frac{4(q-2)}{q(q-1)} \right) \quad (8)$$

and with the variance

$$D(M_1) = n \left(\delta \frac{q-2}{q-1} + \nu \frac{4(q-2)}{q(q-1)} \right) \left(1 - \delta \frac{q-2}{q-1} - \nu \frac{4(q-2)}{q(q-1)} \right) \quad (9)$$

The random variable M_2 representing the number of questions to which both correct responses were given has a binomial probability distribution in the form

$$P(M_2 = m_2) =$$
$$= \binom{n}{m_2} \left(1 - \delta \frac{q-2}{q-1} - \nu \frac{q(q-1)-2}{q(q-1)} \right)^{m_2} \left(\delta \frac{q-2}{q-1} + \nu \frac{q(q-1)-2}{q(q-1)} \right)^{n-m_2} \quad (10)$$

with the expectation

$$E(M_2) = n \left(1 - \delta \frac{q-2}{q-1} - \nu \frac{q(q-1)-2}{q(q-1)} \right) \quad (11)$$

and variance

$$D(M_2) = n \left(1 - \delta \frac{q-2}{q-1} - \nu \frac{q(q-1)-2}{q(q-1)} \right) \left(\delta \frac{q-2}{q-1} + \nu \frac{q(q-1)-2}{q(q-1)} \right) \quad (12)$$

With regard to the fact

$$m_0 + m_1 + m_2 = n$$

the random vector $M = (M_0, M_1, M_2)$ has a multinomial probability distribution defined by the following relation

$$P(M_0 = m_0, M_1 = m_1, M_2 = m_2) =$$
$$= \frac{n!}{m_0!\, m_1!\, m_2!} \left(\nu \frac{(q-2)(q-3)}{q(q-1)} \right)^{m_0} \left(\delta \frac{q-2}{q-1} + \nu \frac{4(q-2)}{q(q-1)} \right)^{m_1} \quad (13)$$
$$\left(1 - \delta \frac{q-2}{q-1} - \nu \frac{q(q-1)-2}{q(q-1)} \right)^{m_2}$$

This probabilistic model makes possible to estimate the parameters ν and δ as parameters of the mentioned above probability distributions. In this way it is also possible to estimate the parameter $\tau = \nu + \delta/2$ representing the whole proportion of the tested topic with which the examinee is unfamiliar.

References

[1] Tesaříková, E.: *Probabilistic model of school-achievement test with double-choice response — variant I.* Acta UPO, Fac. Rer. Nat., 105(1992), 119–126

[2] Tesaříková, E.: *Statistical analysis of probabilistic model—variant I of the school-achievement test with double choice response.* Acta UPO, Fac. Rer. Nat., 105(1992), 127–146

[3] Tesaříková E.: *Probabilistic models of school-achievement test with double choice response and their statistical analysis.* The Statistician. London 1993 (in press)

Author's address: Department of Algebra and Geometry
Faculty of Science
Palacký University
Tomkova 38, Hejčín
779 00 Olomouc
Czech Republic