Acta Universitatis Palackianae Olomucensis. Facultas Rerum Naturalium. Mathematica

Eva Tesaříková

Probabilistic model of school-achievement test with triple choice response

Acta Universitatis Palackianae Olomucensis. Facultas Rerum Naturalium. Mathematica, Vol. 33 (1994), No. 1, 133--140

Persistent URL: http://dml.cz/dmlcz/120307

ACTA UNIVERSITATIS PALACKIANAE OLOMUCENSIS FACULTAS RERUM NATURALIUM
1994 Mathematica XXXIII Vol. 114

PROBABILISTIC MODEL OF SCHOOL–ACHIEVEMENT TEST WITH TRIPLE CHOICE RESPONSE

EVA TESAŘÍKOVÁ

(Received January 17, 1994)

Abstract

In connection with the analysis of achievement tests with triple choice response it is possible to construct various mathematical models describing the probability structure of such tests. The present variant proceeds from the assumption of the existence of four possible levels of the real familiarity with the one question tested topic. The proportion of the whole topic to be examined which the tested person really does not know then can be expressed in the form of a weighted sum $\tau = \nu + \delta/3 + 2\omega/3$, where ω is the proportion of "$one - third - familiarity$", δ is the proportion of "two-thirds-familiarity", and ν is the proportion of "total unfamiliarity" with the tested topic.

Key words: school-achievement test, triple-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.

In papers [1], [2], [3], [4] published by the author in years 1990–1993 were constructed two different variants of probabilistic models of the school-achievement tests with double choice response.

In connection with the analysis of achievement tests with triple choice response it is possible to construct various mathematical models describing the probability structure of such tests. The present variant proceeds from the assumption of the existence of four possible levels of the real familiarity with the one question tested topic: in case of "one-third-knowledge" the person under examination is really familiar only with one of the three correct answers,in case of "two-thirds-knowledge" the person under examination is really familiar with two of the three correct answers, in case of "total knowledge" the person under examination is really familiar with everyone of the three correct answers, in case of "total unknowledge" the tested person is really unfamiliar with everyone of the three correct answers among the offered alternatives. The proportion of the whole topic to be examined which the tested person really does not know then can be expressed in the form of a weighted sum $\tau = \nu + \delta/3 + 2\omega/3$, where ω is the proportion of "one-third-familiarity", δ is the proportion of "two-thirds-familiarity" , and ν is the proportion of "total unfamiliarity" with the tested topic.

2 Assumption of the model

Let us consider the school-achievement test with compulsory choice of three correct responses from q response alternatives, $q > 5$. 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.

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 "two-thirds-familiar" with the topic of the i-th question

O_i tested person is "one-third-familiar" with the topic of the i-th question

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

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

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

In relation to the registered results of the test the following four 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} two correct answers were given to i-th question

S_{i3} three correct answers were given to i-th question

In the case of "total familiarity" with the given topic the person under examination will give all the three correct responses, so that

$$P(S_{i0}|Z_i) = 0, \quad P(S_{i1}|Z_i) = 0, \quad P(S_{i2}|Z_i) = 0, \quad P(S_{i3}|Z_i) = 1.$$

In the case of "two-thirds-familiarity" with the topic of the i-th question, the examinee actually knows only two correct responses, while his third response is effectively chosen at random by him from among the $q-2$ remaining alternatives. The following conditional probabilities describe this situation

$$P(S_{i0}|D_i) = 0, \quad P(S_{i1}|D_i) = 0, \quad P(S_{i2}|D_i) = \frac{q-3}{q-2}, \quad P(S_{i3}|D_i) = \frac{1}{q-2},$$

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

$$P(S_{i0}|O_i) = 0,$$

$$P(S_{i1}|O_i) = \frac{\binom{q-3}{2}}{\binom{q-1}{2}} = \frac{(q-3)(q-4)}{(q-1)(q-2)}$$

$$P(S_{i2}|O_i) = \frac{\binom{q-3}{1}\binom{2}{1}}{\binom{q-1}{2}} = \frac{4(q-3)}{(q-1)(q-2)},$$

$$P(S_{i3}|O_i) = \frac{\binom{2}{2}}{\binom{q-1}{2}} = \frac{2}{(q-1)(q-2)}.$$

In case of "total unfamiliarity" with the topic tested by the i-th question the examined person can only use the random choice. With regard to the presence of three correct responses among $q > 5$ offered alternatives the following relations hold:

$$P(S_{i0}|N_i) = \frac{\binom{q-3}{3}}{\binom{q}{3}} = \frac{(q-3)(q-4)(q-5)}{q(q-1)(q-2)},$$

$$P(S_{i1}|N_i) = \frac{\binom{q-3}{2}\binom{3}{1}}{\binom{q}{3}} = \frac{9(q-3)(q-4)}{q(q-1)(q-2)},$$

$$P(S_{i2}|N_i) = \frac{\binom{q-3}{1}\binom{3}{2}}{\binom{q}{3}} = \frac{18(q-3)}{q(q-1)(q-2)},$$

$$P(S_{i3}|N_i) = \frac{\binom{3}{3}}{\binom{q}{3}} = \frac{6}{q(q-1)(q-2)}.$$

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

Unconditional probabilities of events $S_{i0}, S_{i1}, S_{i2}, S_{i3}$ can be calculated according to the theorem of total probability:

$$\begin{aligned}
p_{i0} = & \\
&= P(S_{i0}) = P(S_{i0} \cap Z_i) + P(S_{i0} \cap D_i) + P(S_{i0} \cap O_i) + P(S_{i0} \cap N_i) \\
&= P(Z_i)P(S_{i0}|Z_i) + P(D_i)P(S_{i0}|D_i) + P(O_i)P(S_{i0}|O_i) + P(N_i)P(S_{i0}|N_i) \\
&= (1-\nu-\omega-\delta).0 + \delta.0 + \omega.0 + \nu\frac{(q-3)(q-4)(q-5)}{q(q-1)(q-2)} \\
&= \nu\frac{(q-3)(q-4)(q-5)}{q(q-1)(q-2)} \qquad (1)
\end{aligned}$$

$$\begin{aligned}
p_{i1} = & \\
&= P(S_{i1}) = P(S_{i1} \cap Z_i) + P(S_{i1} \cap D_i) + P(S_{i1} \cap O_i) + P(S_{i1} \cap N_i) \\
&= P(Z_i)P(S_{i1}|Z_i) + P(D_i)P(S_{i1}|D_i) + P(O_i)P(S_{i1}|O_i) + P(N_i)P(S_{i1}|N_i) \\
&= (1-\nu-\omega-\delta).0 + \delta.0 + \omega\frac{(q-3)(q-4)}{(q-1)(q-2)} + \nu\frac{9(q-3)(q-4)}{q(q-1)(q-2)} \\
&= \omega\frac{(q-3)(q-4)}{(q-1)(q-2)} + \nu\frac{9(q-3)(q-4)}{q(q-1)(q-2)} \qquad (2)
\end{aligned}$$

$$
\begin{aligned}
p_{i2} &= \\
&= P(S_{i2}) = P(S_{i2} \cap Z_i) + P(S_{i2} \cap D_i) + P(S_{i2} \cap O_i) + P(S_{i2} \cap N_i) \\
&= P(Z_i)P(S_{i2}|Z_i) + P(D_i)P(S_{i2}|D_i) + P(O_i)P(S_{i2}|O_i) + P(N_i)P(S_{i2}|N_i) \\
&= (1-\nu-\omega-\delta).0 + \delta\frac{q-3}{q-2} + \omega\frac{4(q-3)}{(q-1)(q-2)} + \nu\frac{18(q-3)}{q(q-1)(q-2)} \\
&= \delta\frac{q-3}{q-2} + \omega\frac{4(q-3)}{(q-1)(q-2)} + \nu\frac{18(q-3)}{q(q-1)(q-2)}
\end{aligned} \tag{3}
$$

$$
\begin{aligned}
p_{i3} &= \\
&= P(S_{i3}) = P(S_{i3} \cap Z_i) + P(S_{i3} \cap D_i) + P(S_{i3} \cap O_i) + P(S_{i3} \cap N_i) \\
&= P(Z_i)P(S_{i3}|Z_i) + P(D_i)P(S_{i3}|D_i) + P(O_i)P(S_{i3}|O_i) + P(N_i)P(S_{i3}|N_i) \\
&= (1-\nu-\omega-\delta).1 + \delta\frac{1}{q-2} + \omega\frac{2}{(q-1)(q-2)} + \nu\frac{6}{q(q-1)(q-2)} \\
&= 1 - \delta\frac{q-3}{q-2} - \omega\frac{(q-1)(q-2)-2}{(q-1)(q-2)} + \nu\frac{q(q-1)(q-2)-6}{q(q-1)(q-2)}.
\end{aligned} \tag{4}
$$

The assumption of equal 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

$$
\begin{aligned}
p_{i3} &= p_3, \quad p_{i2} = p_2, \quad p_{i1} = p_1, \quad p_{i0} = p_0, \\
S_{i0} &= S_0, \quad S_{i1} = S_1, \quad S_{i2} = S_2, \quad S_{i3} = S_3
\end{aligned}
$$

for all the considered i.

To the whole test, i.e. the whole series of n independent question we can associate four random variables M_0, M_1, M_2, M_3.

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

$$
\begin{aligned}
P(M_0 = m_0) &= \\
&= \binom{n}{m_0}\left(\nu\frac{(q-3)(q-4)(q-5)}{q(q-1)(q-2)}\right)^{m_0}\left(1-\nu\frac{(q-3)(q-4)(q-5)}{q(q-1)(q-2)}\right)^{n-m_0}
\end{aligned} \tag{5}
$$

with the expectation

$$
E(M_0) = n\nu\frac{(q-3)(q-4)(q-5)}{q(q-1)(q-2)} \tag{6}
$$

and with the variance

$$
D(M_0) = n\nu\frac{(q-3)(q-4)(q-5)}{q(q-1)(q-2)}\left(1-\nu\frac{(q-3)(q-4)(q-5)}{q(q-1)(q-2)}\right) \tag{7}
$$

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 the relation

$$P(M_1 = m_1) = \\ = \binom{n}{m_1}\left[\left(\omega+\nu\frac{9}{q}\right)\frac{(q-3)(q-4)}{(q-1)(q-2)}\right]^{m_1}\left[1-\left(\omega+\nu\frac{9}{q}\right)\frac{(q-3)(q-4)}{(q-1)(q-2)}\right]^{n-m_1} \tag{8}$$

with the expectation

$$E(M_1) = n\left(\omega+\nu\frac{9}{q}\right)\frac{(q-3)(q-4)}{(q-1)(q-2)} \tag{9}$$

and with the variance

$$D(M_1) = n\left(\omega+\nu\frac{9}{q}\right)\frac{(q-3)(q-4)}{(q-1)(q-2)}\left[1-\left(\omega+\nu\frac{9}{q}\right)\frac{(q-3)(q-4)}{(q-1)(q-2)}\right] \tag{10}$$

Variable M_2 represents the number of questions to which only two correct answers were given. This variable has a binomial probability distribution defined by relation

$$\begin{aligned} P(M_2 = m_2) &= \binom{n}{m_2}\left[\left(\delta+\frac{4\omega}{q-1}+\frac{18\nu}{q(q-1)}\right)\left(\frac{q-3}{q-2}\right)\right]^{m_2}\cdot \\ &\cdot\left[1-\left(\delta+\frac{4\omega}{q-1}+\frac{18\nu}{q(q-1)}\right)\left(\frac{q-3}{q-2}\right)\right]^{n-m_2} \end{aligned} \tag{11}$$

with the expectation

$$E(M_2) = n\left(\delta+\frac{4\omega}{q-1}+\frac{18\nu}{q(q-1)}\right)\left(\frac{q-3}{q-2}\right) \tag{12}$$

and with the variance

$$\begin{aligned} D(M_2) &= n\left(\delta+\frac{4\omega}{q-1}+\frac{18\nu}{q(q-1)}\right)\cdot \\ &\cdot\left(\frac{q-3}{q-2}\right)\left[1-\left(\delta+\frac{4\omega}{q-1}+\frac{18\nu}{q(q-1)}\right)\left(\frac{q-3}{q-2}\right)\right]. \end{aligned} \tag{13}$$

The random variable M_3 representing the number of questions to which all the three correct responses were given has a binomial probability distribution again in the form

$$\begin{aligned} &P(M_3 = m_3) = \\ &\binom{n}{m_3}\left[1-\delta\frac{q-3}{q-2}-\omega\frac{(q-1)(q-2)-2}{(q-1)(q-2)}-\nu\frac{q(q-1)(q-2)-6}{q(q-1)(q-2)}\right]^{m_3}\cdot \\ &\cdot\left[\delta\frac{q-3}{q-2}+\omega\frac{(q-1)(q-2)-2}{(q-1)(q-2)}+\nu\frac{q(q-1)(q-2)-6}{q(q-1)(q-2)}\right]^{n-m_3}. \end{aligned} \tag{14}$$

The expectation is defined by

$$E(M_3) = n\left(1 - \delta\frac{q-3}{q-2} - \omega\frac{(q-1)(q-2)-2}{(q-1)(q-2)} - \nu\frac{q(q-1)(q-2)-6}{q(q-1)(q-2)}\right) \qquad (15)$$

and the variance by

$$\begin{aligned} D(M_3) \;=\; & n\left(1 - \delta\frac{q-3}{q-2} - \omega\frac{(q-1)(q-2)-2}{(q-1)(q-2)} - \nu\frac{q(q-1)(q-2)-6}{q(q-1)(q-2)}\right)\cdot \\ & \cdot\left(\delta\frac{q-3}{q-2} + \omega\frac{(q-1)(q-2)-2}{(q-1)(q-2)} + \nu\frac{q(q-1)(q-2)-6}{q(q-1)(q-2)}\right). \end{aligned} \qquad (16)$$

With regard to the fact $m_0 + m_1 + m_2 + m_3 = n$ the random vector $M = (M_0, M_1, M_2, M_3)$ has a multinomial probability distribution defined by the following relation

$$\begin{aligned} & P(M_0 = m_0, M_1 = m_1, M_2 = m_2, M_3 = m_3) = \\ & = \frac{n!}{m_0!\,m_1!\,m_2!\,m_3!}\left[\nu\frac{(q-3)(q-4)(q-5)}{q(q-1)(q-2)}\right]^{m_0}\left[\left(\omega + \nu\frac{9}{q}\right)\frac{(q-3)(q-4)}{(q-1)(q-2)}\right]^{m_1}\cdot \\ & \cdot\left[\left(\delta + \frac{4\omega}{q-1} + \frac{18\nu}{q(q-1)}\right)\frac{q-3}{q-2}\right]^{m_2}\left[1 - \delta\frac{q-3}{q-2} - \omega\frac{(q-1)(q-2)-2}{(q-1)(q-2)} - \right. \\ & \left. - \nu\frac{q(q-1)(q-2)-6}{q(q-1)(q-2)}\right]^{m_3}. \end{aligned} \qquad (17)$$

This probabilistic model makes possible to estimate the parameters ν, ω and δ as parameters of the above mentioned probability distributions. In this way is also possible to estimate the parameter $\tau = \nu + \delta/3 + 2\omega/3$ 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 UP Olomucensis, 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 UP Olomucensis, Fac. rer. nat., 105 (1992), 127–146.

[3] Tesaříková, E.: *Variant II of probabilistic model of the school-achievement test with double-choice response,* Acta UP Olomucensis, Fac. rer. nat., 110 (1993), 159–164.

[4] Tesaříková, E.: *Statistical analysis of variant II of the probabilistic model of the school-achievement test with double choice response,* Acta UP Olomucensis, Fac. rer. nat., (1993), 165–175.

Author's address: Department of Algebra and Geometry
Faculty of Science
Palacký University
Tomkova 40, Hejčín
779 00 Olomouc
Czech Republic