Testový výkon a úspěšnost

ZDENĚK PŮLPÁN

Anotace: V práci se diskutují možnosti hodnocení testových výsledků bez použití klasických statistických metod. Vychází se z myšlenky teorie fuzzy množin a k posouzení úspěšnosti respondenta se využívá odhadu věrohodným hodnotitelem.

Klíčová slova: didaktický test, testový výkon, obtížnost položky, fuzzy odhad úspěšnosti.

Každou testovou položku se obvykle snažíme konstruovat tak, aby měla jednoznačně formulovatelnou správnou odpověď. Je-li pak možné z výsledků odpovědí na testovou položku (nebo homogenní skupinu položek) odhadnout pravděpodobnost správné odpovědi (nebo správných odpovědí) p, $p \in \langle 0, 1 \rangle$, můžeme zavést vztah pro tzv. informační obtížnost $v_1 = v(p)$ podle [1] resp. [2] ve tvaru

$$v_1 = v_1(p) = \frac{1 - I(S; p)}{2}, \qquad v_1 \in \langle 0; 1 \rangle \tag{1}$$

kde

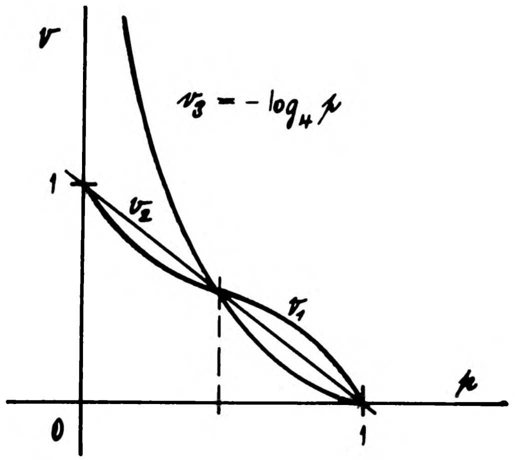
$$\begin{aligned} I(S; p) &= I(p) \quad \text{pro} \quad p \geqq 0{,}5 \\ &= -I(p) \quad \text{pro} \quad 0 \leqq p < 0{,}5 \\ I(p) &= 1 - (-p \log_2 p - (1 - p) \log_2 (1 - p)) \\ &\quad (\text{zde definujeme } 0 \,.\, \log 0 = 0). \end{aligned}$$

Mírou obtížnosti testové položky může být také např.

$$v_2 = v_2(p) = 1 - p; \qquad v_2 \in \langle 0, 1 \rangle \tag{2}$$

$$\text{nebo} \quad v_3 = v_3(p) = -\log_z p; \qquad v_3 \in \langle 0, \infty \rangle\,; \; z > 1. \tag{3}$$

Je-li $v_i \to 0$ $(i = 1, 2, 3)$, je testová položka velmi snadná, je-li $v_i \to 1$ $(i = 1, 2)$ nebo $v_3 \to \infty$, je testovaná položka velmi obtížná. (Obr. 1)



Obtížnost testové položky může být také stanovena na základě subjektivního odhadu věrohodným hodnotitelem. Používá se při tom T. L. Saatyho metody ([3]).

Má-li test k položek, hodnotitel porovnává vzájemné obtížnosti testových položek např. na základě devítibodové stupnice:

1... označuje obě položky za stejně obtížné,

3... označuje první položku mírně obtížnější než druhou,

5... označuje první položku významně obtížnější,

7... označuje první položku podstatně obtížnější,

9... označuje první položku absolutně obtížnější,

hodnoty 2, 4, 6, 8 vyjadřují kompromis.

Předpokládejme, že známe odhadované obtížnosti jednotlivých položek $v^1, v^2, \dots, v^k$; $v^i \in (0, 1\rangle$ *pro* $i = 1, 2, \dots, k$. Pak můžeme matici **Q** poměrů obtížností jednotlivých položek zapsat ve tvaru (4):

$$\mathbf{Q} = \begin{bmatrix} \frac{v^1}{v^1}, & \frac{v^1}{v^2}, & \dots & \frac{v^1}{v^k} \\ \frac{v^2}{v^1}, & \frac{v^2}{v^2}, & \dots & \frac{v^2}{v^k} \\ . & & & \\ . & & & \\ . & & & \\ \frac{v^k}{v^l}, & \frac{v^k}{v^2}, & \dots & \frac{v^k}{v^k} \end{bmatrix} = \left[\frac{v^i}{v^j}\right] \qquad i, j = 1, 2, \dots, k \tag{4}$$

Pak platí

$$\mathbf{Q}.\begin{bmatrix} v^1 \\ v^2 \\ \cdot \\ \cdot \\ \cdot \\ v^k \end{bmatrix} = \mathbf{k}.\begin{bmatrix} v^1 \\ v^2 \\ \cdot \\ \cdot \\ \cdot \\ v^k \end{bmatrix} \tag{5}$$

Označíme-li

$$\boldsymbol{v}_4 \, 1 \begin{bmatrix} v^1 \\ v^2 \\ \cdot \\ \cdot \\ \cdot \\ v^k \end{bmatrix}, \tag{6}$$

můžeme (5) psát ve tvaru (7):

$$\mathbf{Q} \,.\, \boldsymbol{v}_4 \, 1 \, \mathrm{k} \,.\, \boldsymbol{v}_4 \tag{7}$$

$$(\mathbf{Q} - \mathrm{kJ}) \,.\, \mathbf{v}_4 \, 1 \, 0. \tag{8}$$

Pak platí

$$(\mathbf{Q} - \lambda J) = 0, \tag{9}$$

když λ je vlastní číslo matice **Q**, a kde J resp. 0 označujeme jednotkovou resp. nulovou matici řádu $k \times \mathrm{k}$.

Rovnice (9) má nenulové řešení, právě když k je vlastní číslo matice **Q**. Existuje právě jedno vlastní číslo $\lambda \neq 0$, $\lambda = k$.

Řešením rovnice (8) je kterýkoliv sloupec matice **Q**. Jednoznačnosti řešení pro $\boldsymbol{v}_4$ dosáhneme zavedením dodatečné podmínky, např.

$$\sum_{i=1}^{k} v^i = 1.$$

(Pro $\boldsymbol{v}_4$ dále označíme každou složku ještě indexem 4.)

Obtížnosti položek mohou sloužit jako podklad k vícerozměrnému škálování (celkový skór testu s k položkami si můžeme představit jako k-rozměrný vektor) nebo k vyvažování položek testu. Vyvážený skór dostaneme tak, že za každou správně zodpovězenou testovou položku obdrží respondent skór, rovný obtížnosti příslušné položky. Tak se zvýhodní oproti obvyklému skórování (kdy se za každou správně vyřešenou testovou položku přiřazuje vždy stejný skór) ze dvou respondentů ten, který vyřeší při stejném počtu správné zodpovězených položek položky obtížnější (ve smyslu příslušné obtížnosti (1), (2), (3) nebo (6)).

Má-li test k položek s obtížnostmi v_j^i ($i = 1, 2, \ldots, k$; $j = 1, 2, 3, 4$), je celkový vyvážený skór respondenta

$$S^j = \sum_{i=1}^{k} s_i v_j^i \,; \qquad j = 1, 2, 3, 4 \tag{10}$$

nebo po normování

$$S^j_n = \frac{S^j}{\sum_i v^i_j} \; ; \qquad S^j_n \in \langle 0; 1\rangle \; j = 1, 2, 4 \qquad (10')$$

kde $s_i = 1$ resp. $s_i = 0$, když respondent i-tou položku vyřešil resp. nevyřešil. Pro $j = 3$ platí, že když pro některé i je $p_i \to 0$, je pak $S^3 \to \infty$. Součiny $s_i \, . \, v^i_j$ můžeme chápat jako úspěšnosti respondentů v i-té položce a skóry S^1, S^2, S^3, S^4 jako úspěšnosti respondentů v celém testu.

Nelze-li buď přesně formulovat nebo jednoznačně hodnotit odpověď respondenta, musí být testový výkon odhadnout jistým hodnotitelem, např. jako číslo μ z intervalu $\langle 0; 1\rangle$. Takové číslo μ pak vyjadřuje subjektivní odhad úrovně správnosti odpovědi.

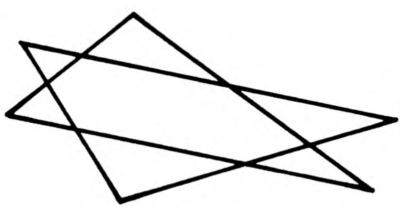
Příklad 1: Může být průnikem dvou trojúhelníků šestiúhelník?

Odpovědi respondentů na tuto testovou položku mohou být následujících typů

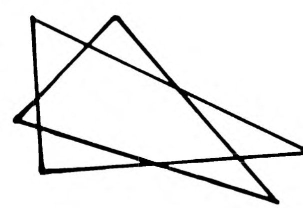
a) diskutují se možné vzájemné polohy polorovin, určující oba trojhúhelníky; úvahy jsou zcela obecné, správné;

b) průnikem dvou trojúhelníků může být šestiúhelník; jako důkaz je uveden některý z následujích obrázků 2 resp. 3 resp. 4:

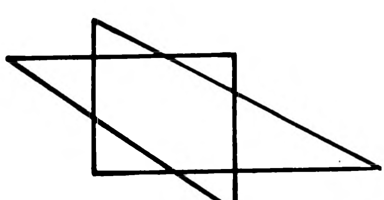
α)



β)



γ)



Z hlediska matematického jsou někdy cennější řešení, která vycházejí z co nejobecnějších představ. Proto zde hodnotitel např. určí pro jednotlivé odpovědi postupně $\mu_a = 1$; $\mu_\alpha = 0{,}9$; $\mu_\beta = 0{,}8$; $\mu_\gamma = 0{,}7$. Řešení, kde respondent nevyznačí v žádném obrázku průnik dvou trojúhelníků (ať je tím průnikem cokoliv), hodnotitel označí $\mu = 0$ a pod.

Je-li subjektivní odhad úrovně správnosti řešení testové položky jisté $\mu \in [0,1]$ je vhodné stanovit míru úspěšnosti respondenta $u(\mu)$ vztahem (11)

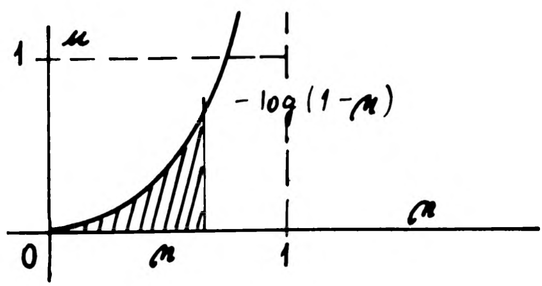
$$u(\mu) = \frac{1}{log_z e} \int_0^{\mu} -\log_z (1 - t)\, dt ; \tag{11}$$

$$\mu \in \langle 0; 1 \rangle ; \qquad z \geqq 2$$

Pak je

$$0 \leqq u(\mu) \leqq 1, \tag{12}$$

kde mezní hodnoty u získá ten respondent, který byl hodnocen $\mu = 0$ nebo $\mu = 1$ (viz obrázek 5):



Úspěšnost respondenta v celém testu s k položkami se pak může hodonotit mírou U^j, $j = 1, 2, 4$, vztahem

$$U^j = \sum_{i=1}^{k} u(\mu_i)\, v_j^i \tag{13}$$

$$0 \leqq U_j \leqq \sum_i v_j^i$$

nebo její normovanou hodnotou

$$U_n^j = \frac{1}{\sum_i v_j^i} U^j ; \qquad 0 \leqq U_n^j \leqq 1 ; \qquad j = 1, 2, 4 \tag{14}$$

Pro posouzení úspěšnosti určité skupiny respondentů můžeme využít experimentálně odhadnutého rozložení relativních četností na μ, které označíme $m(\mu)$. Úspěšnost skupiny respondentů v jedné testové položce pak určíme následujícím postupem:

a) stanovíme „přísnost“ volbou čísla $z \in \langle 2, \infty)$; čím je z větší, tím je i „přísnost“ vyšší;

b) je-li experimentální rozložení relativních četností spojitě aproximováno rozložením $m(\mu)$, stanoví se úspěšnost $W(z)$ skupiny respondentů pomocí (15):

$$W(z) = \int_0^1 \min(m(\mu) ; -\log_z (1 - \mu))\, d\mu \tag{15}$$

c) je-li experimentální rozložení $m(\mu)$ diskrétní s body $\mu_0 = 0, \mu_1, \mu_2, \ldots, \mu_n = 1$, kde $\mu_i < \mu_j$ pro $i < j$, $i, j = 0, 1, \ldots, n$, na kterých jsou dány hodnoty relativní četnosti $m(\mu_i)$, stanoví se úspěšnost $W(z)$ skupiny respondentů pomocí (16)

$$W(z) = \sum_{i=0}^{n} \min\,(m(\mu_i);\ -\log_z(1-\mu_i)) \tag{16}$$

V obou případech platí, že

$$0 \leqq W(z) \leqq 1$$

pro libovolné $z \in \langle 2, \infty)$.

Příklad 2: Nechť je dáno diskrétní rozložení $m(\mu)$ následující tabulkou:

μ	0,3	0,4	0,5	0,6
m	0,1	0,6	0,2	0,1

Pro $z = 2$ dostaneme úspěšnost $W(2)$ následujícím výpočtem

$$\begin{aligned} W(2) &= \min\,(0{,}1;\ -\log_2(1-0{,}3)) + \min(0{,}6;\ -\log_2(1-0{,}4)) + \\ &+ \min\,(0{,}2;\ -\log_2(1-0{,}5)) + \min\,(0{,}1;\ -\log_2(1-0{,}6)) = \\ &= 0{,}1 + 0{,}6 + 0{,}2 + 0{,}1 = 1{,}0. \end{aligned}$$

Pro $z = 10$ určíme úspěšnost $W(10)$ z výpočtu

$$\begin{aligned} W(10) &= \min\,(0{,}1;\ -\log 0{,}7) + \min\,(0{,}6;\ -\log 0{,}6) + \min\,(0{,}2;\ -\log 0{,}5) + \\ &+ \min\,(0{,}1;\ -\log 0{,}4) = 0{,}1 + 0{,}2 + 0{,}2 + 0{,}1 = 0{,}6. \end{aligned}$$

Volíme-li z přísněji, dostaneme úspěšnost menší.

Celkovou úspěšnost pro k položek pak stanovíme jako vážený součet úspěšností jednotlivých testových položek $W_i(z)$:

$$W^j(z) = \sum_{i=1}^{k} v_j^i\, W_i(z); \qquad W^j(z) \in \langle 0; \textstyle\sum_i v_j^i \rangle; \qquad j = 1, 2, 4 \tag{17}$$

Po normování dostaneme úspěšnost ve tvaru

$$W_n^j(z) = \frac{W^j(z)}{\sum_i v_j^i}; \quad W_n^j(z) \in \langle 0, 1 \rangle; \quad j = 1, 2, 4 \tag{18}$$

ZÁVĚR

Testové položky je možné rozlišovat nejen podle obsahu, ale také na základě koeficientu obtížnosti.

Pro stanovení koeficientu obtížnosti je možné použít buď objektivního odhadu pravděpodobnosti vyřešení položky (např. z relativní četnosti vyřešení v homogenní populaci anebo z relativní četnosti vyřešení více homogenních položek respondentem) nebo odhadu věrohodným hodnotitelem. Tak dostaneme obtížnosti v_1, v_2, v_3, v_4.

Pomocí některé z obtížností můžeme stanovit odpovídající testový výsledek (který zde nazýváme úspěšnost) ze vztahu pro S_n^1, S_n^2, S_n^3 nebo S^4 (viz (10), (10′). Nejsou-li odpovědi respondentů zcela přesně hodnotitelné, lze použít odhadu věrohodným hodnotitelem tak, že se úspěšnost v testu dá posoudit pomocí $W_n^1(z)$, $W_n^2(z)$, $W_n^4(z)$ (*viz* (18)).

LITERATURA

Půlpán, Zd.: Některá informační kritéria pro posouzení homogenních skupin respondentů. Československá psychologie č. 5, roč. XXV, 1981, s. 440–451

Půlpán, Zd.: Zur Bestimmung des Schwierigkeitsgrades von Aufgaben mit Hilfe der Semantischen Information, grkg/Humankybernetik, Band 29, Heft 4 (1988).

Půlpán, Zd., Kuřina, Fr., Kebza, Vl.: O představivosti. Praha, Academia 1992.

Saaty, Th.: Exploring the Interface between Hierarchies, Multiple Objectives and Fuzzy Sets. In: Fuzzy Sets and Systems 1, 1978, č. 1

ZDENĚK PŮLPÁN
THE TEST PERFORMANCE AND SUCCESSFULNESS

In the work the tests results evalutions possibilities are being discussed without applying the classical statistiacal methods. The autor proceeds from the idea of the sets fuzzy, and estimations of competent evalutors are exploited for the evalutions of the respondent's successfulness.

the ballance of the total tests score the coefficient of the information difficulty is offered which was introduced by author (1988) or by Saaty (1978). Further on the so called measure of successfulness limiting the subjective estimation by the mathematical estimation is being discussed.

The offered characterization are interpretted in a suitable way.

Došlo do redakce: 23. 1. 1991

Autor: Doc. RNDr. PhDr. ZDENĚK PŮLPÁN, CSc., pedagogická fakulta, nám. Svobody 301, 500 00 Hradec Králové