INFORMAČNÍ ÚČINNOST DIDAKTICKÉHO TESTU S VOLBOU Z NABÍDNUTÝCH ODPOVĚDÍ

STANISLAV KOMENDA, JANA ZAPLETALOVÁ

Test typu multiple-choice

Didaktický test s volbou správné odpovědi z předložených nabídek zprostředkovává zpětnou vazbu mezi učitelem a žákem v procesu učení. K hlavním přednostem dobře sestaveného testu zmíněného typu patří jeho objektivnost a také ekonomičnost při měření, jaké úrovně znalostí o daném tématu bylo opravdu dosaženo.

Struktura testové položky, jejichž soubor pak tvoří test, zahrnuje tzv. kmen, což je tvrzení nebo otázka či úloha, týkající se zkoušené znalosti. Spolu s otázkou se pak zkoušenému subjektu předkládá několik nabídek odpovědi, z nichž zpravidla právě jediná bývá správná a ostatní hrají roli distraktorů.

Úkolem zkoušeného subjektu je vybrat z nabídek odpovědi tu, kterou považuje za správnou. Pokud se jeho volba shoduje se záměrem examinátora, je položka skórována hodnotou 1 (jedním bodem), což je její příspěvek k testovému skóre správných odpovědí v testu. V případě nesprávného řešení položky je její příspěvek k testovému skóre nulový. Existují však i určité modifikace těchto zásad skórování v testu.

Konstrukce testových položek vyžaduje, vedle dobré znalosti zkoušeného tématu, i určitý důvtip a znalost zásad tvorby testu. Způsob formulace problému rozhoduje nejen o zkoušeném tématu, ale i o tom, jak hluboké mají být znalosti, na něž je dotazováno. To souvisí s tzv. taxonomií výukových cílů, např. v pojetí Bloomově.

Hlavním problémem bývá při konstrukci testové položky formulace distraktorů. Požaduje se, aby tyto plně fungovaly - v tom smyslu, že žádná není pro subjekt triviální a subjekt bez znalosti zkoušeného tématu volí mezi nabídkami odpovědi podle zásad náhodného výběru. Protože vyšší počet nabídek snižuje, při zachování jejich kvality - možnost dosažení správného řešení náhodným uhádnutím, bývá na straně examinátora snaha formulovat nabídek (tj. vlastně distraktorů) s danou položkou co nejvíce. To však bývá velmi obtížné - což examinátor někdy překonává tím, že jako distraktory nabízí odpovědi z různých „dimenzí" zkoušeného tématu. Tímto způsobem lze, samozřejmě, zvyšovat počet distraktorů prakticky bez omezení - dá se však pochybovat, že by přitom zůstal respektován požadavek rovnocennosti všech nabídnutých odpovědí.

Tomuto problému se budeme později věnovat hlouběji a ukážeme, že z hlediska jistého rozumně zvoleného kritéria informační účinnosti jsou rozsáhlejší testy s jednoduššími položkami ekvivalentní testům méně rozsáhlým, užívajícím položky s vyšším počtem distraktorů.

Pravděpodobnostní model didaktického testu

Uvažujme situaci, kdy je test aplikován na subjekt se záměrem zjišťovat, jakou část sledovaného tématu tento opravdu zná resp. nezná.

Test je složen z n položek považovaných za vzájemně nezávislé. Tím se rozumí, že znalost správného řešení dané položky subjektem nijak neovlivňuje schopnost subjektu správně řešit kteroukoliv z dalších n-1 testových položek.

Spolu s položkou se subjektu předkládá q nabídek odpovědi, z nichž jedna je správná a q-1 představuje distraktory. O tom je zkoušený subjekt informován, spolu se žádostí, aby jednu z q nabídek - totiž tu, kterou považuje za řešení položky - označil za správnou (a tím současně prohlásil zbylých q-1 položek za nesprávné). Podle instrukce má subjekt jednu z nabídek označit za správnou i tehdy, není-li si správným řešením úlohy jist; tím využije možnosti dosáhnout správné odpovědi náhodným uhádnutím.
Test má tedy dva parametry: rozsah n (počet položek) a q (počet nabídek odpovědi s každou položkou).

Vedle toho uvažuje model testu další parametr π ($0 \leq \pi \leq 1$), totiž podíl látky zkoušeného tématu, který subjekt neovládá. Přitom platí, že zatímco parametry n a q jsou kontrolovány examinátorem, je parametr π pod kontrolou zkoušeného subjektu. Veličina $1-\pi$ tak představuje podíl látky, kterou zkoušený subjekt ovládá.

Parametr π bude objektem statistického odhadování, využívajícího naměřených údajů o chování subjektu.

Odhad parametru π modelu testu

Za předpokladu, že naší evidencí je dostupná informace o počtu chyb k, kterých se subjekt dopustil v testu (k je skóre nesprávných odpovědí subjektu), lze odhadovat číselnou hodnotu parametru π metodou maximální věrohodnosti. Tímto způsobem nás výsledek testu dokáže informovat o tom, jaký podíl zkoušené látky subjekt ve skutečnosti neovládá či ovládá.

Odhady získané metodou maximální věrohodnosti mají dobré vlastnosti, žádoucí ve smyslu požadavků teorie statistického odhadování. Jsou asymptoticky nevychýlené, asymptoticky eficientní a jejich rozdělení se blíží k rozdělení normálnímu – s rostoucím rozsahem testu n.

Maximálně věrohodný odhad $\hat{\pi}$ parametru π tedy je dán výrazem

$$\hat{\pi} = \frac{k}{n}\frac{q}{q-1}\,,\ \text{jestliže}\quad k \le n\frac{q-1}{q}$$

$$\hat{\pi} = 1\,,\quad \text{jestliže}\quad k > n\frac{q-1}{q}.$$

Střední hodnota odhadu

Za předpokladu, že π je skutečná hodnota parametru, lze odvodit střední hodnotu odhadu. Platí přitom

$$E(\hat{\pi}|\pi) = B\left(\left[n\frac{q-1}{q}\right]-1 \middle| n-1, \pi\frac{q-1}{q}\right) +$$

$$+\,1 - B\left(\left[n\frac{q-1}{q}\right] \middle| n, \pi\frac{q-1}{q}\right),$$

kde $[x]$ označuje celou část čísla x a

$$B(x|n,p) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}$$

je distribuční funkce binomického rozdělení s parametry n a p, v bodě x.

Odhad (3) je vychýlený; jeho střední hodnota se liší od odhadované skutečné hodnoty parametru, a to tak, že očekávaná (střední) hodnota $E(\hat{\pi}|\pi)$ odhadu parametru π je nižší nežli skutečná hodnota parametru π. Odhadem $\hat{\pi}$ je tedy odhadovaná hodnota parametru π v průměru podceňována.

Odchylka činí

$$\pi - E(\hat{\pi}|\pi) =$$

$$= \pi\left\{1 - B\left(\left[n\frac{q-1}{q}\right] - 1 \mid n-1, \pi\frac{q-1}{q}\right)\right\} -$$

$$- \left\{1 - B\left(\left[n\frac{q-1}{q}\right] \mid n, \pi\frac{q-1}{q}\right)\right\} .$$

Rozdíl postupně mizí k nule s rostoucím q, tak, že v případě (hypotetického) testu s volně formulovanou odpovědí vymizí docela.

Rozptyl odhadu

Pro rozptyl $D(\hat{\pi}|\pi)$ odhadu tak dostáváme

$$D(\hat{\pi}|\pi) =$$

$$= \frac{n-1}{n}\pi^2 B\left(\left[n\frac{q-1}{q}\right] - 2 \mid n-2, \pi\frac{q-1}{q}\right) +$$

$$+ \frac{1}{n}\frac{q}{q-1}\pi B\left(\left[n\frac{q-1}{q}\right] - 1 \mid n-1, \pi\frac{q-1}{q}\right) +$$

$$+1-B\left(\left[n\frac{q-1}{q}\right]\mid n,\pi\frac{q-1}{q}\right)-$$

$$-\left\{\pi B\left(\left[n\frac{q-1}{q}\right]-1\mid n-1,\pi\frac{q-1}{q}\right)+\right.$$

$$\left.+1-B\left(\left[n\frac{q-1}{q}\right]\mid n,\pi\frac{q-1}{q}\right)\right\}^2.$$

Ve speciálním případě testu s volně tvořenou odpovědí ($q = \infty$) se rozptyl odhadu redukuje do podstatně jednoduššího tvaru

$$D(\hat{\pi}|\pi)=\frac{1}{n}\pi(1-\pi).$$

Testy stejné informační účinnosti

Rozptyl $D(\hat{\pi}|\pi)$ odhadu $\hat{\pi}$ je mírou informační účinnosti (vlastně spíše neúčinnosti) testu. Při daném π (skutečné, neznámé hodnotě parametru) závisí rozptyl dále na vnějších parametrech n a q. To nabízí možnost hledat takové kombinace hodnot (n, q), které vedou ke shodným hodnotám rozptylu – a tedy kvantifikovat představu, která se zcela přirozeně nabízí: rozsáhlý test (s velkým počtem položek n) s jednoduchými položkami (o malém počtu nabídek odpovědi q) může být stejně účinný jako poměrně malý test (s nízkým n) se složitějšími položkami (o větším počtu nabídek odpovědi q).

Hodnotou rozptylu $D(\hat{\pi}|\pi)$ jsou tedy zaváděny určité skupiny testů (n, q) dané úrovně ekvivalence informační účinnosti. Vyhledávání ekvivalentních testů je poměrně jednoduché. Pro zvolenou hodnotu parametru π a danou hodnotu rozptylu $D(\hat{\pi}|\pi)$ se najdou pro dané hodnoty q příslušné rozsahy testů n. Hodnotu π je třeba volit na základě předběžné zkušenosti s úrovní znalostí subjektů ve zkoušené tematické oblasti.

Tak například, pro $\pi = 0{,}4$ (předpokládá se, že subjekty neovládají asi 40% zkoušené látky) zjišťujeme, že téže úrovně rozptylu odhadu ve výši 0,010 dosahují testy ($n = 64$, $q = 2$),

$(n = 44,\ q = 3)$, $(n = 37,\ q = 4)$, $(n = 34,\ q = 5)$ a $(n = 24,\ q = \infty)$. Podobně můžeme postupovat také při jiných volbách π a $D(\$|\pi)$.

Grafickou modifikaci postupu představují soustavy křivek na grafech. Pro zvolenou úroveň neznalostí a zvolenou kvalitu odhadu (na svislé ose) se odečítají souřadnice (na vodorovné ose) průsečíků této kvality odhadu s křivkami pro jednotlivé počty distraktorů q. Uvažována je úroveň neznalosti $\pi = 0{,}25$ a $\pi = 0{,}50$.