DIDAKTICKÝ TEST Z HLEDISKA STATISTICKÉ INDUKCE

Stanislav Komenda

Jana Zapletalová

Lékařská fakulta UP Olomouc

Uvažujme situaci, kdy je předložen test s volbou odpovědi subjektu, s cílem změřit, jaký díl látky, která je předmětem zkoušky, subjekt skutečně ovládá.

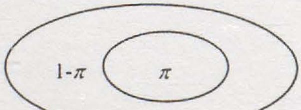
Test je tvořen n položkami (otázkami, úlohami). Spolu s každou testovou položkou se předkládá subjektu q alternativ odpovědi (q=2,3,4,...); právě jediná z nich je správným řešením, zatímco zbylých q-1 alternativ působí jako distraktory (falešné nabídky).

V rámci administrace testu je subjekt instruován, že má právě jednu z nabízených alternativ zvolit (označit, zaškrtnout, zakroužkovat) v souladu se svou představou o řešení úlohy. Je poučen, že správná je právě jediná alternativa a že zbylých q-1 alternativ jsou řešení nesprávná. Zároveň je upozorněn na možnost dosáhnout správného řešení náhodným uhádnutím i v případě, kdy si není správností své volby jist. Tím jsou vymezeny tři veličiny, jejichž vztah bude předmětem zájmu:

1.*Rozsah testu* (počet testových položek) n a *počet alternativních odpovědí* q, předkládaných subjektu s každou testovou položkou. Obě tyto veličiny jsou kontrolovány (řízeny) autorem testu, případně examinátorem v rámci projektu, přípravy, sestavování testu.

2. *Relativní podíl látky, kterou subjekt ovládá (či neovládá),* z tématu, jež je předmětem zkoušení (ověřování, zjišťování, měření). Tuto veličinu budeme formalizovat symbolem π ($0<\pi<1$), pro podíl látky, který subjekt neovládá. Specifikuje ji svými znalostmi zkoušený subjekt. Veličina π není dostupná přímému měření; lze však předpokládat, že ovlivňuje některé další veličiny, které už přímému měření dostupné jsou - a může se proto stát objektem indukce, usuzování z toho, co lze měřit, na to, co měřit nelze. A protože jsou tu dobré důvody chápat vztah mezi přímo měřitelným a přímo neměřitelným jako *vztah statistický*, je přirozené sáhnout po metodách *statistické indukce, odhadování parametrů vhodného modelu a testování hypotéz o parametrech takového modelu.*

Elementárním způsobem se dá výše naznačená představa zachytit množinovým diagramem: prvky universa znalostí se rozkládají do dvou podmnožin. Jedna obsahuje poznatky subjektu neznámé - míra její mohutnosti je π. Do druhé patří poznatky subjektu známé - její míra je rovna 1-π (viz Obr.a).



Obr. a

Stav vědomostí, které subjekt o předmětu má, se promítá do jeho chování, reakcí. V situaci testování je elementární reakcí subjektu na otázku odpověď, řešení dílčí úlohy, položky testu.

Označme symbolem p pravděpodobnost, že odpověď subjektu na uvažovanou otázku testu bude nesprávná; $1-p$ je pak doplňková pravděpodobnost, že subjekt vyřeší úlohu správně.

Za předpokladu, že

(a) testové položky jsou náhodnými prvky universa znalostí daného tématu a

(b) pro subjekt bez znalosti tématu položky je q nabídnutých alternativních odpovědí souborem rovnocenných možností volby (bez preference), lze závislost dimenze odpovědí na dimenzi znalostí popsat pomocí elementárního vzorce pro úplnou pravděpodobnost náhodných jevů

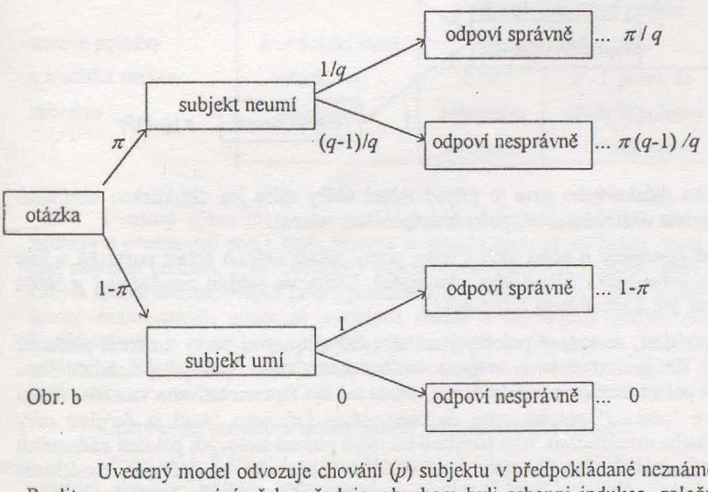
$$1 - p = (1 - \pi) + \frac{\pi}{q}, \qquad (1)$$

v němž je formulována skutečnost, že subjekt ovládající látku zkoušené otázky odpoví správně s jistotou, zatímco subjekt bez znalosti může dosáhnout správné odpovědi mechanismem náhodného uhádnutí s pravděpodobností $1/q$. Trajektorie tohoto mechanismu spolu s jejich pravděpodobnostmi zachycuje Obr. b.

Vzorec (1) lze upravit na tvar

$$p = \pi (q - 1) / q < \pi .$$

Pravděpodobnost p nesprávného řešení testové položky náhodně zvolené z množiny zkoušených poznatků je tedy menší než podíl zkoušené látky, který subjekt neovládá. Zároveň je zřejmé, že pravděpodobnost p by neměla - ani v případě úplné neznalosti ($\pi = 1$) - klesnout pod hranici $(q-1) / q$, představující tak jakousi "statistickou nulu" této stupnice.



Obr. b

Uvedený model odvozuje chování (p) subjektu v předpokládané neznámé skutečnosti (π). Realita procesu poznání však vyžaduje, abychom byli schopni indukce, založené na inverzním postupu: ze známého chování (p) usuzovat na neznámou skutečnost (π). S využitím předchozího vztahu a Bayesova vzorce pro inverzní pravděpodobnost se odvodí pro testovou položku vztah

$$\text{P(subjekt umí | subjekt odpověděl správně)} = \frac{q(1-\pi)}{\pi + q(1-\pi)} \qquad (2)$$

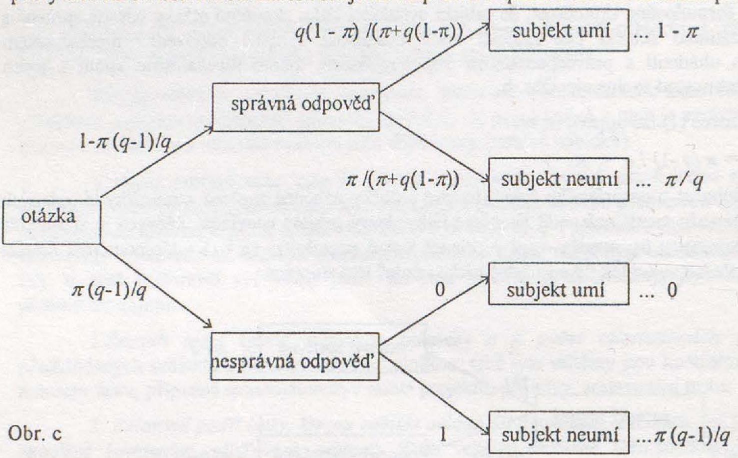
a

$$\text{P(subjekt umí | subjekt odpověděl nesprávně)} = 0\,.$$

Možné trajektorie této úvahy demonstruje Obr.c.

Nerovnosti $$0 \leq 1 - \pi \leq \frac{1-\pi}{1-\pi\frac{q-1}{q}}$$

poskytují kvantitativní představu o tom, jak evidence examinátora o chování subjektu mění jeho "názor" na znalosti subjektu o tématu pokrytém testovou položkou. Zatímco prostřední člen $1\text{-}\pi$ představuje očekávanou míru znalosti v případě, kdy má examinátor k dispozici jenom všeobecnou informaci o úrovni znalostí v referenční populaci, k níž subjekt náleží, jsou levý a pravý člen nerovnosti mírou znalostí subjektu nesprávně a správně řešícího aktuální problém.



Obr. c

Z hlediska didaktického testu je případ jediné úlohy spíše jen didaktickou abstrakcí. Obvykle zahrnuje test větší nebo menší počet úloh (položek, otázek).

Označme symbolem n počet úloh v testu, pokrývajícím určitou oblast poznatků v jisté hloubce, která je pedagogicky zdůvodněna jako účelná. Uvažujme subjekt neovládající π 100% látky z této oblasti, jež je předmětem zkoušení.

Předpokládejme, že testové položky jsou náhodně vybranými prvky universa poznatků zkoušené oblasti. To je v souladu se zřejmou snahou autora testu "pokrýt" položkami testu zkoušenou oblast pokud možno rovnoměrně; test tématu má být reprezentativním vzorkem vědění o tématu. Relace "padnutí" položek testu do podmnožiny "π" nebo "$1\text{-}\pi$" je do jisté míry výsledkem náhodného mechanismu. Jeho působení lze měřit počtem testových položek padnuvších do podmnožiny poznatků, které subjekt neovládá a modelovat binomickým rozdělením pravděpodobností

$$P(l|\pi,n) = \binom{n}{l}\pi^{l}(1-\pi)^{n-l}, \qquad (l = 0, 1, ..., n) \qquad (3)$$

Tímto pravděpodobnostním rozdělením je modelován náhodný mechanismus, kontrolující, které z dílčích poznatků uvažované látky zahrne examinátor do svého testu formou testových položek.

Označme jako k, $k = 0, 1, \ldots, l$ počet testových položek, které zkoušený subjekt nedokázal vyřešit, tj., na které odpověděl nesprávně; zřejmě je potom $l\text{-}k$ počet položek, které se podařilo subjektu vyřešit, aniž by příslušnou úlohu znal (tzn., že správnou úlohu uhádl).

Vzájemné relace veličin n, l a k přehledně znázorňuje schéma (4)

O tom, které z l položek (jejichž předmět zkoušený subjekt neovládá) budou řešeny správně a které nesprávně, rozhoduje mechanismus náhodného uhadování daný volbou alternativy odpovědi z q předložených nabídek. Za předpokladu (a) plného fungování distraktorů u všech testových položek a (b) vzájemné nezávislosti odpovědí na položky - lze tento mechanismus modelovat binomickým rozdělením pravděpodobností

$$P(k|l,q) = \binom{l}{k}\left(\frac{q-1}{q}\right)^{k}\left(\frac{1}{q}\right)^{l-k} \qquad (5)$$

$(k = 0,1,\ldots,l)$, $(l = 0,1,\ldots,n)$, $(q = 2,3,4,\ldots)$, l a q daná

<table>
<tr><td></td><td colspan="3">n položek testu s volbou odpovědi</td></tr>
<tr><td>testové položky
z hlediska znalosti
subjektu</td><td>l položek, jejichž
předmět subjekt
neovládá (neumí)</td><td colspan="2">n - l položek, jejichž
předmět subjekt
ovládá (umí)</td></tr>
<tr><td rowspan="2">testové položky
z hlediska reakce
subjektu</td><td rowspan="2">k položek, které
subjekt řeší
nesprávně</td><td colspan="2">n - k položek, které subjekt
řeší správně, z toho</td></tr>
<tr><td>l - k
náhodným
uhádnutím</td><td>n - l proto, že
předmět položky
ovládá (umí)</td></tr>
</table>

(4)

Uvedený výraz (5) je podmíněné rozdělení pravděpodobností náhodné veličiny, jejíž hodnoty k představují počet chyb, kterých se subjekt dopustil při řešení testu, za podmínky, že jiná náhodná veličina, jejíž hodnoty l udávají počet testových položek, které subjekt neovládá, nabyla známé úrovně. Případ testu s položkami, předpokládajícími volně konstruovanou odpověď (essay examination), odpovídá modelové situaci s neomezeně velkým počtem nabídek q. Rozdělení (5) je v takovém případě jednobodové - veličiny l a k jsou totožné. Dosáhnout správného řešení úlohy náhodným uhadováním nelze, tento mechanismus je pak z procesu vyloučen.

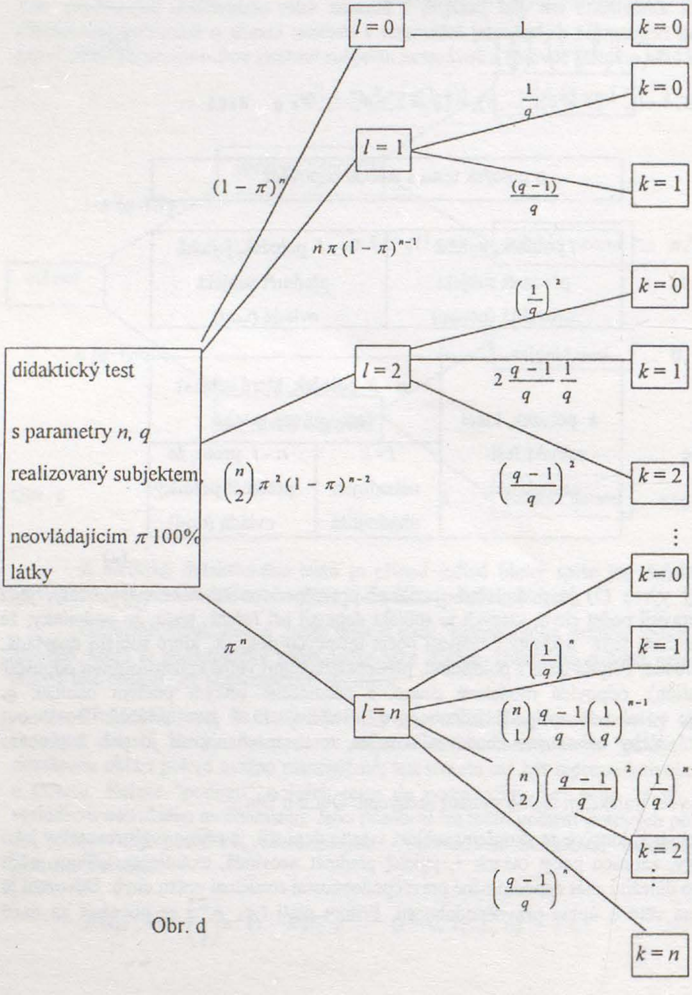
Úplný výčet trajektorií těchto úsudků zachycují Obr.d a Obr.e.

Počet chyb k, kterých se zkoušený subjekt v testu dopustil, je přímo registrovatelný jako výsledek zkoušky, zatímco počet otázek l, jejichž předmět neovládá, examinátor přímo měřit nemůže. Je proto důležité znát nepodmíněné pravděpodobnostní rozdělení počtu chyb. Odvození je možno založit na větě o úplné pravděpodobnosti. Přitom platí (π, n, q se považují za dané konstanty)

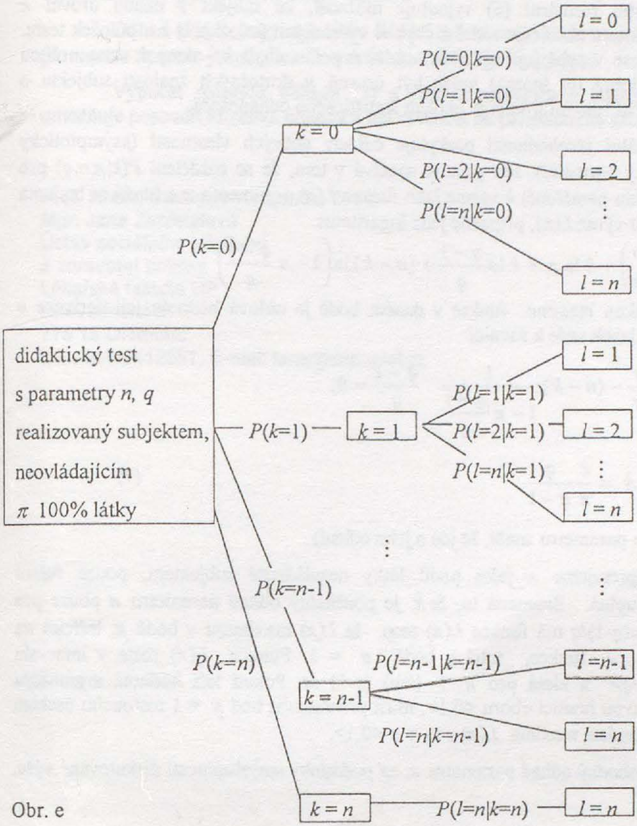
$$P(k|\pi,n,q) = \sum_{l=k}^{n} P(k|l,q)P(l|\pi,n) = \binom{n}{k}\left(\pi - \frac{q-1}{q}\right)^{k}\left(1 - \pi\frac{q-1}{q}\right)^{n-k} \qquad (6)$$

Odvozené rozdělení je opět binomické. Jeho empirickým protějškem je četnostní rozdělení v souboru subjektů podle počtu chyb (resp. správných řešení), kterých se dopustili (resp. kterých dosáhli) při řešení testu. Tím jsou zároveň vytvořeny podmínky pro statistické odhadování parametrů modelu a rovněž pro ověření oprávněnosti předpokladů, za nichž bylo rozdělení odvozeno.

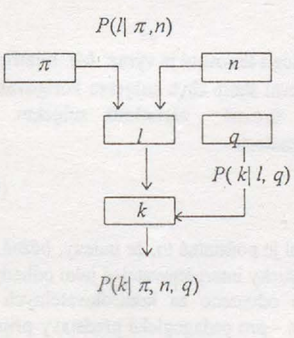
Vztahy mezi elementy didaktického testu, jak je vymezují příslušná pravděpodobnostní rozdělení, zachycuje blokové schéma následujícího grafu (Obr. f). Tato rozdělení jsou specifikována vzorci (3), (5) a (6) uvedenými výše.



Obr. d


l = 0
P(l=0|k=0)
P(l=1|k=0)
l = 1
k = 0
P(l=2|k=0)
l = 2
P(l=n|k=0)
P(k=0)
l = n
didaktický test
s parametry n, q
realizovaný subjektem,
neovládajícím
π 100% látky
l = 1
P(l=1|k=1)
P(k=1)
k = 1
P(l=2|k=1)
l = 2
P(l=n|k=1)
l = n
P(k=n-1)
P(l=n-1|k=n-1)
l=n-1
P(k=n)
k = n-1
P(l=n|k=n-1)
l = n
k = n
P(l=n|k=n)
l = n

Obr. e


P(l| π,n)
π
n
l
q
P(k| l, q)
k
P(k| π, n, q)

Obr. f

Odhad podílu látky nezvládnuté subjektem

Pravděpodobnostní rozdělení (6) vyjadřuje možnost, že subjekt s danou úrovní π neznalosti zkoušeného tématu absolvuje test s k chybně vyřešenými položkami z n položek testu. Po skončení zkoušky nese v sobě její výsledek - údaj o počtu chyb k - naopak zase určitou informaci o tom, jak dobrá (či špatná) může být úroveň π skutečných znalostí subjektu o zkoušeném tématu. Vytěžit tuto informaci je úkolem statistického odhadování.

Metoda maximální věrohodnosti poskytuje odhady dobrých vlastností (asymptoticky eficientní a asymptoticky normální). Její princip spočívá v tom, že se rozdělení $P(k|\pi,n,q)$ pro dané (tj. u našeho subjektu naměřené) k vezme jako funkce $L(\pi)$ argumentu π a hledá se hodnota argumentu, maximizující výraz $L(\pi)$, případně jeho logaritmus

$$\ln L(\pi) = \ln\binom{n}{k} + k\ln\pi + k\ln\frac{q-1}{q} + (n-k)\ln\left(1-\pi\frac{q-1}{q}\right)$$

Nutnou podmínkou maxima funkce v daném bodě je nulová hodnota její derivace v tomto bodě. Tento požadavek vede k rovnici

$$\frac{\partial \ln L(\pi)}{\partial \pi} = \frac{k}{\pi} - (n-k)\frac{1}{1-\pi\frac{q-1}{q}}\,\frac{q-1}{q} = 0,$$

Řešením je výraz $$\hat{\pi} = \frac{k}{n}\frac{q}{q-1} \qquad (7)$$

(symbol"^" nad znakem parametru značí, že jde o jeho odhad).

Poněvadž interpretujeme π jako podíl látky nezvládnuté subjektem, pouze řešení $0 \leq \hat{\pi} \leq 1$ jsou smysluplná. Znamená to, že $\hat{\pi}$ je použitelný odhad parametru π pouze pro $k \leq n(q-1)/q$. Pro $k > n(q-1)/q$ má funkce $L(\pi)$ resp. $\ln L(\pi)$ maximum v bodě π, ležícím na hranici definičního oboru funkce, totiž v bodě $\pi = 1$. Funkce $L(\pi)$ roste v intervalu $\pi \in \langle 0, (k/n)(q-1)/q\rangle$ a klesá pro $\pi > (k/n)(q-1)/q$. Pokud leží hodnota argumentu $\pi = (k/n)(q-1)/q$ za pravou hranicí oboru $\langle 0,1\rangle$, musí překračovat bod $\pi = 1$ rostoucím úsekem funkce $L(\pi)$ a $\pi = 1$ je bodem maxima $L(\pi)$ pro $\pi \in \langle 0,1\rangle$.

Maximálně věrohodný odhad parametru π, za podmínky smysluplnosti diskutované výše, je tedy

$$\begin{aligned} \hat{\pi} &= \frac{k}{n}\frac{q}{q-1} && \text{pro } k \leq n\frac{q}{q-1} \\ \hat{\pi} &= 1 && \text{pro } k > n\frac{q-1}{q} \end{aligned} \qquad (8)$$

Vraťme se ještě k odhadu (7). V terminologii testování je výraz k/n relativní hrubé skóre nesprávných odpovědí subjektu a odhad $\hat{\pi}$ relativní skóre chyb subjektu korigované na možnost náhodného uhádnutí, nebo také index úrovně neznalosti subjektu. Vedle tvaru $\hat{\pi} = (k/n)\,q/(q-1)$ se používá také index úrovně znalosti

$$1-\hat{\pi} = \frac{n-k}{n} - \frac{k}{n}\frac{1}{q-1}. \qquad (9)$$

Z hlediska teorie statistického odhadování je podstatné to, že indexy, běžně používané při hodnocení výsledků didaktických testů, jsou statisticky interpretovatelné jako odhady parametru π pravděpodobnostního rozdělení (6), které bylo odvozeno za kontrolovatelných předpokladů. Parametr π se dá přitom interpretovat způsobem - pro pedagogické představy přijatelným - jako podíl látky nezvládnuté zkoušeným subjektem. Odhad $\hat{\pi}$ byl získán metodou maximální věrohodnosti, jež je zárukou kvality odhadu.

Odhad (7) je vychýlený - jeho střední (očekávaná) hodnota je nižší než skutečná hodnota parametru π, takže je odhadem neznámá skutečnost podceňována.

Výpočet rozptylu odhadu (7) a jeho konfrontace s Raovou-Cramerovou nerovností umožňuje posoudit eficienci odhadu a její závislost na parametrech didaktického testu.

Prof. RNDr. Stanislav Komenda, DrSc.
Mgr. Jana Zapletalová
Ústav sociálního lékařství
a zdravotní politiky
Lékařská fakulta UP
Hněvotínská 3
775 15 Olomouc
tel. : 086/5412551, E-mail jana@risc.upol.cz