EDUKOMETRIE: DIDAKTICKÝ TEST A JEHO MODELOVÁNÍ

*Komenda S., Zapletalová J.**

1. 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 odpovědí 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 žádný 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.

2. 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

---

* *odd. biometrie, LF Univerzity Palackého, Hněvotínská 3, 775 15 Olomouc*
komendas@risc.upol.cz, jana@risc.upol.cz

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.

Symbolem p budeme označovat pravděpodobnost, že zkoušený subjekt vyřeší správně položku náhodně vybranou v souboru n položek testu. Pravděpodobnost p souvisí s parametrem π vztahem

$$1-p = (1-\pi).1 + \pi(1/q)\,, \tag{1}$$

využívajícím vzorce pro úplnou pravděpodobnost. Vyjádřeno slovně:

Pravděpodobnost správné odpovědi subjektu na položku testu =
Pravděpodobnost, že položka patří do oblasti látky, kterou subjekt ovládá ×
Pravděpodobnost, že subjekt správně vyřeší položku ze známé části látky +
Pravděpodobnost, že položka patří do oblasti látky, kterou subjekt neovládá ×
Pravděpodobnost, že subjekt správně vyřeší položku z neznámé části látky.

Ze vztahu (1) plyne

$$p = (q-1)/q\,.\,\pi < \pi\,. \tag{2}$$

Ze (2) je zřejmé, že v důsledku možnosti dosáhnout správného řešení položky náhodným uhádnutím bez znalosti zkoušené látky bude pravděpodobnost nesprávné odpovědi p nižší nežli podíl π zkoušené látky, který subjekt neovládá. Protože π musí ležet v intervalu od 0 do 1, bude se pravděpodobnost nesprávné odpovědi p nacházet v intervalu od 0 do $(q-1)/q$. Doplňková hodnota $1-p = 1/q$ tak představuje jakousi „statistickou nulu" na škále možnosti správné odpovědi.

Pomocí (1) odvodíme inverzní pravděpodobnost, že subjekt, který vyřešil položku správně, opravdu látku pokrytou položkou ovládá. Ta je dána podílem, v jehož čitateli je výraz

Pravděpodobnost, že položka patří do oblasti látky, kterou subjekt ovládá ×
Pravděpodobnost, že takovou položku subjekt vyřeší správně $= (1-\pi).1$

a ve jmenovateli pravděpodobnost správného vyřešení náhodně vzaté položky, tj. $1-p$ z (1), tedy $1-\pi+\pi/q$. Podíl je tedy roven

$$\frac{1-\pi}{1-\pi+\dfrac{\pi}{q}} = \frac{q(1-\pi)}{\pi+q(1-\pi)}\,. \tag{3}$$

Komplementární pravděpodobnost udává možnost, že subjekt, který správně vyřešil úlohu, téma ve skutečnosti neovládá, totiž

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

Podobně se dá zjistit, že subjekt, který úlohu správně nevyřešil, ji s jistotou, tj. s pravděpodobností 1, neovládá.

3. Případ několika položek

Nezná-li zkoušený subjekt $\pi.100\%$ látky, lze předpokládat, že z n položek testu padne l $(0 \le l \le n)$ do této části látky s binomickou pravděpodobností

$$P(l|\pi,n)=\binom{n}{l}\pi^{l}(1-\pi)^{n-l}. \qquad (4)$$

Výraz (4) je zároveň pravděpodobnost, že z n položek testu jich padne $n-l$ do oblasti látky, kterou subjekt ovládá. Úvaha je správná za předpokladu, že testové položky pokrývají oblast zkoušené látky v jistém smyslu „rovnoměrně".

Položky patřící do oblasti látky, kterou subjekt ovládá, budou vyřešeny správně. Přitom však budou správně vyřešeny i některé z l položek, jejichž téma subjekt neovládá, a to díky možnosti náhodného uhádnutí.

Označme symbolem k počet položek řešených subjektem nesprávně; $l-k$ bude tedy počet položek řešených správně mechanismem náhodného uhádnutí, $(0 \le k \le l)$. Situaci znázorňuje schéma

<table>
<tr><td rowspan="2">Chování Subjektu</td><td rowspan="2">k položek řešených nesprávně</td><td colspan="2">$n-k$ položek řešených správně, z toho</td></tr>
<tr><td>$l-k$ díky uhádnutí</td><td>$n-l$ díky znalosti tématu</td></tr>
<tr><td>Znalosti subjektu</td><td colspan="2">l položek, jejichž téma subjekt neovládá</td><td>$n-l$ položek, jejichž téma subjekt ovládá</td></tr>
<tr><td></td><td colspan="3">n položek testu</td></tr>
</table>

Pravděpodobnost, že z l položek, jejichž téma subjekt neovládá, bude k položek řešeno chybně (a tedy $l-k$ řešeno správně náhodou), je opět binomická:

$$P(k|l,q)=\binom{l}{k}\left(\frac{q-1}{q}\right)^{k}\left(\frac{1}{q}\right)^{l-k}, \ (k=0,1,\ldots,l) \qquad (5)$$

Binomické rozdělení (5) je podmíněným rozdělením náhodné veličiny X nabývající hodnot k, tj. počtu chyb v testu, za podmínky, že jiná náhodná veličina Y (počet položek testu, které subjekt neovládá) nabyla dané hodnoty l. Tato hodnota l vymezuje zároveň definiční obor veličiny X.

Speciální případ, kdy zkoušenému subjektu není předkládána žádná odpověď (což odpovídá položkám s volnou odpovědí, tzv. essay–test), může být považován za případ s $q=\infty$ (s nekonečně mnoha nabídnutými odpověďmi). Je v tom jistá dialektika: subjekt, jemuž bylo nabídnuto nekonečně mnoho odpovědí, je ve stejné situaci, jako by mu nebyla nabídnuta žádná odpověď. Rozdělení (5) pak pro každé l degeneruje do jednobodového rozdělení

$$P(k=l\,|\,l)=1\,. \qquad (6)$$

(Všech l položek, které subjekt neovládá, vyřeší chybně.)

Veličiny X a Y se v rámci budovaného modelu testu zásadně liší. Zatímco hodnoty veličiny X (počet chyb v testu) jsou dostupné přímé evidenci jako skóre chybných odpovědí v testu, neexistuje žádná možnost empirického zjišťování stavu veličiny Y. Znamená to, že jako výsledek realizace testu na souboru subjektů můžeme konstruovat empirickou distribuci hodnot náhodné veličiny X – jako jedině dostupnou informaci o znalostech. Je proto třeba, abychom odvodili teoretický protějšek tohoto empirického rozdělení – což nám umožní odhadovat hodnotu parametru π, resp. l. Vyjdeme přitom z rozdělení pravděpodobnosti náhodného vektoru (X, Y), k němuž najdeme hledané rozdělení veličiny X jako rozdělení marginální.

V prvním kroku se vlastně jedná o aplikaci věty o úplné pravděpodobnosti

$$P(X=k|\pi,n,q)=\sum_{l=k}^{n}P(k|l,q)P(l|\pi,n) \ , \tag{7}$$

z čehož po dosazení ze (4) a (5) odvodíme

$$\begin{aligned} P(X=k\mid\pi,n,q) &= \sum_{l=k}^{n}\binom{l}{k}\left(\frac{q-1}{q}\right)^{k}\left(\frac{1}{q}\right)^{l-k}\binom{n}{l}\pi^{l}(1-\pi)^{n-l} \\ &= \binom{n}{k}\pi^{k}\left(\frac{q-1}{q}\right)^{k}\sum_{v=0}^{n-k}\binom{n-k}{v}\left(\frac{\pi}{q}\right)^{v}(1-\pi)^{n-k-v} \\ &= \binom{n}{k}\pi^{k}\left(\frac{q-1}{q}\right)^{k}\left[\frac{\pi}{q}+(1-\pi)\right]^{n-k}. \end{aligned}$$

S ohledem na vztah

$$\frac{\pi}{q}+1-\pi = 1-\pi\,\frac{q-1}{q}$$

dostáváme konečně

$$P(X=k|\pi,n,q)=\binom{n}{k}\left(\pi\frac{q-1}{q}\right)^{k}\left(1-\pi\frac{q-1}{q}\right)^{n-k}. \tag{8}$$

Jde opět o binomické rozdělení. Přitom zřejmě bude (8) identické se (4) v případě $q\to\infty$ (což je případ testu–eseje).

4. Odhad parametru π modelu testu

Za předpokladu, že naší evidenci 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.

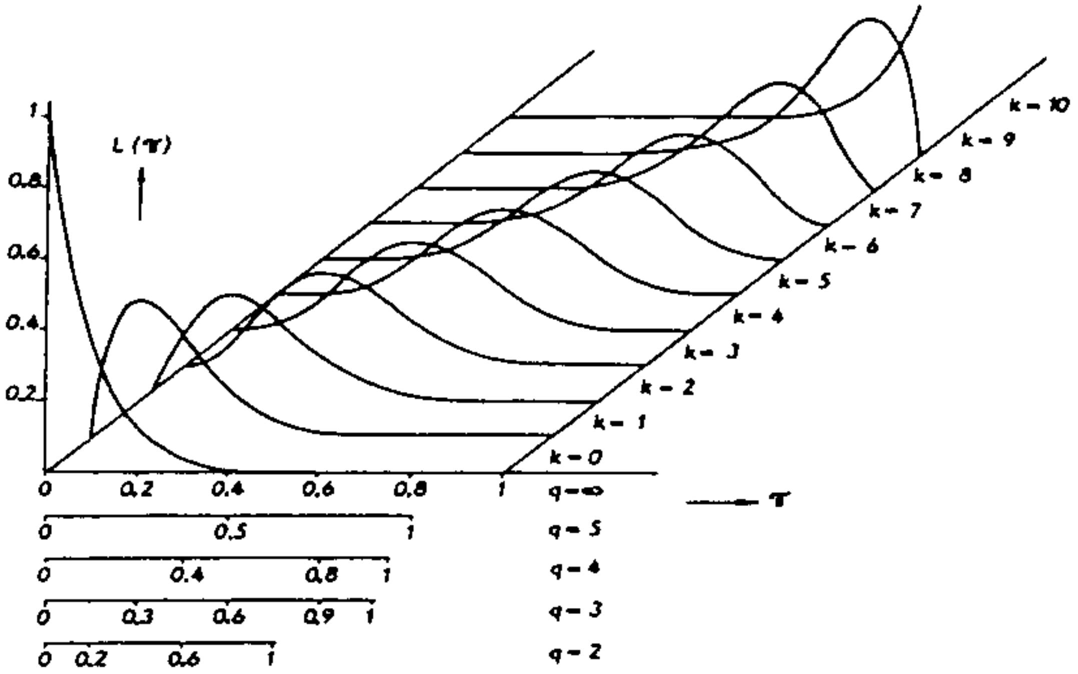
Princip metody maximální věrohodnosti spočívá v tom, že se hledá taková hodnota parametru π, která činí pozorovanou hodnotu k co nejpravděpodobnější. Jde tedy o to najít tu hodnotu π, která maximizuje pravděpodobnost (8). Princip naznačuje Obr.1 pro n = 10.

Úvaha se zjednodušuje, uvažuje-li se místo (8) její logaritmus. Funkce (8) se ovšem nyní chápe jako funkce argumentu π (při dané hodnotě k), a jako taková se označuje názvem věrohodnost.

Logaritmus (přirozený) věrohodnostní funkce má tvar

$$\ln L(\pi)=\ln\binom{n}{k}+k\ln\pi+k\ln\frac{q-1}{q}+(n-k)\ln\left(1-\pi\frac{q-1}{q}\right) \tag{9}$$

a hledané řešení musí splňovat podmínku $0\le\pi\le 1$.



Obr. 1

Nutnou podmínkou existence extrému funkce $\ln L(\pi)$ je nulová hodnota její první derivace v bodě, ve kterém má $\ln L(\pi)$ dosahovat extrému. Dostáváme tak

$$\frac{\mathrm{d}}{\mathrm{d}\pi} \ln L(\pi) = \frac{k}{\pi} - \frac{n-k}{1-\pi\dfrac{q-1}{q}} \frac{q-1}{q} = 0$$

$$\tilde{\pi} = \frac{k}{n}\frac{q}{q-1}.$$

Tento výraz je hledaným řešením za podmínky, že

$$\frac{k}{n}\frac{q}{q-1} \leq 1\,, \quad \text{tj., že} \quad k \leq n\,\frac{q-1}{q}.$$

Pro $k > n(q-1)/q$ má funkce $L(\pi)$ a také $\ln L(\pi)$ svůj extrém v bodě π ležícím na hranici definičního oboru, tj. v bodě $\pi = 1$. To vyplývá ze skutečnosti, že funkce $L(\pi)$ roste s rostoucím π v intervalu od $\pi = 0$ do hodnoty $\pi = \frac{k}{n}\frac{q}{q-1}$ a klesá s π rostoucím od hodnoty $\pi = \frac{k}{n}\frac{q}{q-1}$ výše. Je-li argument $\pi = \frac{k}{n}\frac{q}{q-1}$ maxima funkce $L(\pi)$ mimo oblast (0, 1), musí být hranice $\pi = 1$ překročena rostoucí větví funkce $L(\pi)$ a (takto podmíněné) maximum musí mít funkce $L(\pi)$ v bodě $\pi = 1$. Maximálně věrohodný odhad $\tilde{\pi}$ parametru π je tedy dán výrazem

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

(10)

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

Postačující podmínkou existence maxima funkce $L(\pi)$ v bodě $\pi = \frac{k}{n}\frac{q}{q-1}$ je záporná hodnota její druhé derivace v tomto bodě. Platí

$$\frac{d^2 \ln L(\pi)}{d\pi^2} = -\frac{k}{\pi^2} - \left(\frac{q-1}{q}\right)^2 \frac{n-k}{\left(1-\pi\frac{q-1}{q}\right)^2} .$$

Tento výraz je záporný pro všechna π, tedy i pro $\pi = \frac{k}{n}\frac{q}{q-1}$, což dokazuje existenci maxima funkce $L(\pi)$ resp. $\ln L(\pi)$ v tomto bodě.

Pro případ $n = 10$ je věrohodnostní funkce $L(\pi)$ znázorněna na Obr.1, kde je patrna poloha maxima (v závislosti na počtu nabídek odpovědi q).

5. Střední hodnota odhadu (10)

Za předpokladu, že π je skutečná hodnota parametru, lze odvodit střední hodnotu odhadu (10). Platí přitom

$$E(\tilde{\pi}|\pi) = \pi\, Bi\left(\left[n\frac{q-1}{q}\right] - 1/n-1, \pi\frac{q-1}{q}\right) + 1 - Bi\left(\left[n\frac{q-1}{q}\right]/n, \pi\frac{q-1}{q}\right) , \qquad (11)$$

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

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

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

Odhad (10) je vychýlený; jeho střední hodnota se liší od odhadované skutečné hodnoty parametru, a to tak, že očekávaná (střední) hodnota $E(\tilde{\pi}|\pi)$ odhadu parametru π je nižší nežli skutečná hodnota parametru π. Odhadem $\tilde{\pi}$ je tedy odhadovaná hodnota parametru π v průměru podceňována.

Odchylka činí

$$\pi - E\left(\tilde{\pi}|\pi\right) = \pi\left\{1 - Bi\left(\left[n\frac{q-1}{q}\right] - 1/n-1, \pi\frac{q-1}{q}\right)\right\} - \left\{1 - Bi\left(\left[n\frac{q-1}{q}\right]/n, \pi\frac{q-1}{q}\right)\right\} \qquad (12)$$

Výraz (12) postupně mizí k nule s rostoucím q, tak, že v případě (hypotetického) testu s volně formulovanou odpovědí vymizí docela.

6. Rozptyl odhadu

Při odvozování vzorce pro rozptyl $D(\tilde{\pi}|\pi)$ odhadu (10) se využívá vztah

$$D(\tilde{\pi}|\pi) = E(\tilde{\pi}^2|\pi) - E^2(\tilde{\pi}|\pi) . \qquad (13)$$

Pro rozptyl $D(\tilde{\pi}|\pi)$ odhadu (10) tak dostáváme (viz (11) a (13))

$$D(\tilde{\pi}|\pi) = \frac{n-1}{n}\pi^2 Bi\left(\left[n\frac{q-1}{q}\right] - 2/n-2, \pi\frac{q-1}{q}\right) + \frac{1}{n}\frac{q}{q-1}\pi Bi\left(\left[n\frac{q-1}{q}\right] - 1/n-1, \pi\frac{q-1}{q}\right) +$$

$$1 - Bi\left(\left[n\frac{q-1}{q}\right]/n, \pi\frac{q-1}{q}\right) - -\left\{\pi Bi\left(\left[n\frac{q-1}{q}\right] - 1/n-1, \pi\frac{q-1}{q}\right) + 1 - Bi\left(\left[n\frac{q-1}{q}\right]/n, \pi\frac{q-1}{q}\right)\right\}^2 . \qquad (14)$$

Ve speciálním případě testu s volně tvořenou odpovědí ($q = \infty$) se rozptyl (14) odhadu (10) redukuje do podstatně jednoduššího tvaru

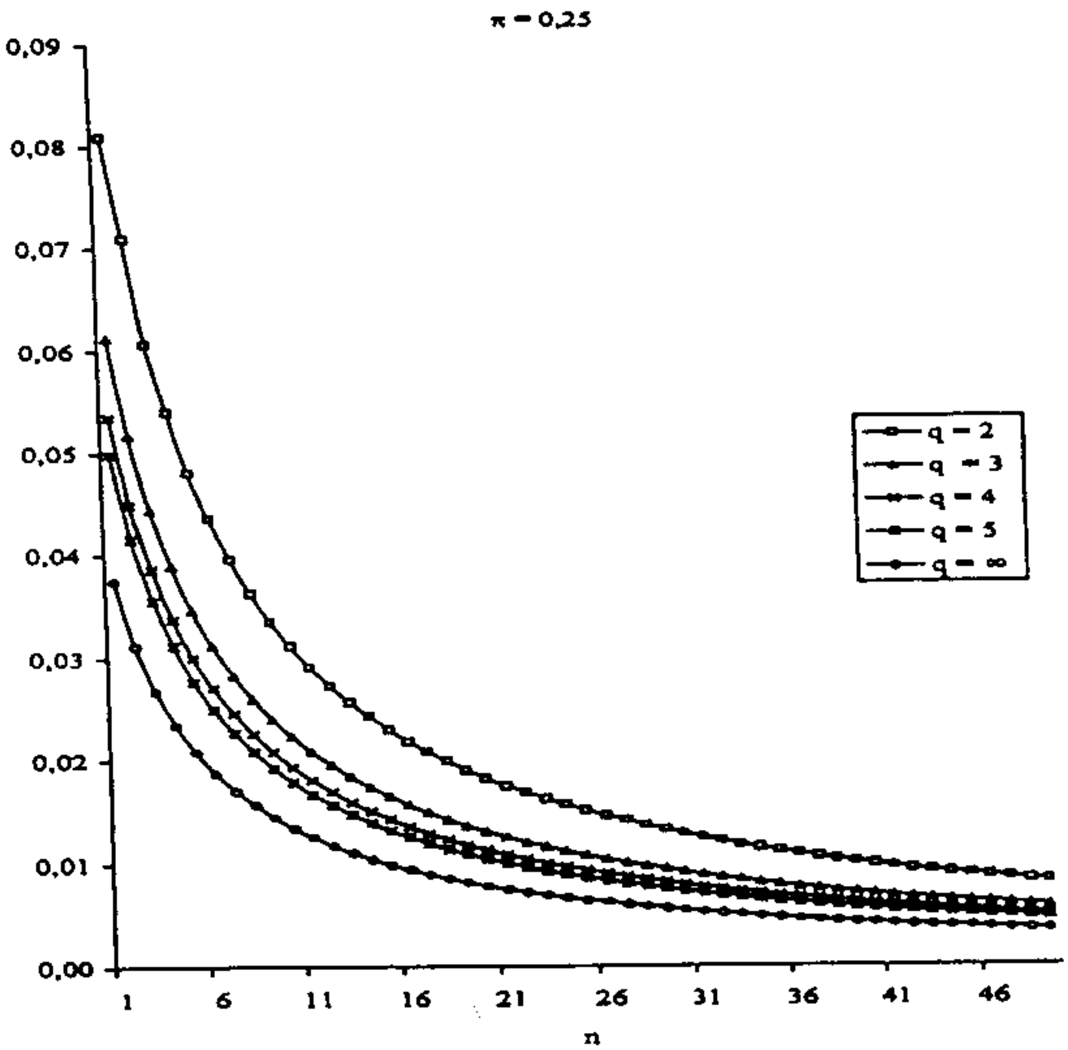
$$D(\tilde{\pi}|\pi) = \frac{1}{n}\pi(1-\pi) . \qquad (15)$$

7. Testy stejné informační účinnosti

Rozptyl $D(\tilde{\pi}|\pi)$ odhadu (10) 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 $D(\tilde{\pi}|\pi)$ jsou tedy zaváděny určité skupiny testů (n, q) dané úrovně ekvivalence informační účinnosti

Alternativní způsob stanovení testů stejné informační účinnosti umožňují grafy. Na Obr.2 konstruovaném pro π = 0,25 stačí vést v dané úrovni účinnosti (na svislé ose) rovnoběžku s vodorovnou osou a na ní (z průsečíků s jednotlivými křivkami pro dané q) odečítat příslušné rozsahy testů n.



Obr. 2