

ENTROPIE JAKO NÁSTROJ OVĚŘOVÁNÍ FUNKCE DISTRAKTORŮ

Stanislav Komenda, Jana Mazuchová, Josef Petřek

***Klíčová slova:** Didaktický test, multiple-choice, testová položka, distraktory, entropie.*

***Anotace:** Entropie - pojem původně fyzikální, zavedený v termodynamice - je dnes běžně využíván počtem pravděpodobnosti a matematickou statistikou jako ukazatel variability četnostního rozdělení nominálního znaku. Článek demonstruje možnou aplikaci entropie při hodnocení funkce distraktorů v testu znalostí, založeném na principu „multiple-choice“.*

Měření znalostí s využitím testu vícenásobné volby (multiple-choice test) se stalo v četných situacích standardní metodou. V prostředí vysoké školy je tato metoda využívána jednak jako test orientovaný na populační normu (což je typické pro přijímací zkoušky uchazečů o studium na fakultě), ale také jako test orientovaný problémově na kritérium dosažené dílčí znalosti (což je typické pro zkoušku z určitého předmětu). V každém z obou těchto případů je test sestavován z položek podle poněkud odlišných zásad [1]. Zatímco úkolem testu orientovaného na zjištění pozice subjektu vzhledem k populační normě je maximálně diferencovat v referenční populaci a k tomu účelu využít položek širokého spektra obtížnosti, aby tak byla stupnice výsledného testového skóre naplno uplatněna - je záměrem kritériálního testu co nejcitlivěji rozpoznat subjekty, které danou látku ovládají, od těch, které ji nezvládnou.

Testům obou kategorií je přitom vlastní společný požadavek účinného fungování distraktorů, tj. alternativních, nesprávných řešení, které test s každou svou položkou zkoušenému subjektu nabízí.

1. Struktura didaktického testu

Didaktický test jako nástroj měření znalostí subjektu je zvažován v širším kontextu pedagogické reality, v níž jsou identifikovány tyto strukturální úrovně:

- a) skutečné znalosti, které subjekt má o látce zkoušeného předmětu,
- b) skutečné znalosti, které subjekt má o látce „pokryvané“ položkami testu,
- c) manifestované znalosti, tj. odpověď subjektu na podnět příslušných testových položek.

Vycházíme přitom z představy, že „skutečné znalosti subjektu“ jsou teoretický koncept, nepřístupný přímému měření. Evidována jsou řešení či odpovědi subjektu na testové položky. Předpokládá se, že na odpovědích subjektu se jeho skutečné znalosti zkoušeného tématu podílejí a že skutečné a naměřené (manifestované) znalosti spolu pozitivně korelují.

Závislost odpovědi subjektu, tj. jeho řešení dílčích testových úloh, která se promítají do skóre správných, nebo nesprávných odpovědí testu, může být matematicky specifikována. Tato specifikace se stává základem modelu odpovědi.

Snad vůbec nejjednodušším možným modelem postulujícím závislost odpovědi subjektu na jeho skutečných znalostech zkoušeného tématu je - chápat počet otázek testu, jejichž předmět subjekt neovládá, jako binomicky rozdělenou náhodnou veličinu Y , přičemž podíl nezvládnuté látky ($\approx$ úroveň skutečné neznalosti subjektu) vystupuje jako parametr π ($0 \leq \pi \leq 1$) tohoto rozdělení. Touto představou je v uvažované situaci formalizována závislost úrovně (b) na úrovni (a). Dá se ukázat, že odpověď subjektu ($\approx$ jeho testové skóre X) závisí na znalosti testu Y opět binomicky; parametrem pravděpodobnostního rozdělení je přitom znalost testových položek. Tím je vyjádřena závislost úrovně (c) na úrovni (b) v naší výchozí koncepci.

Celou tuto strukturu shrnuje schéma na obr. 1. Představa silně simplifikuje; předpokládá totiž, že pro zkoušený subjekt jsou všechny položky testu stejně obtížné a že subjekt řeší položky vzájemně nezávisle (to, že vyřeší správně položku předcházející, nijak neovlivní možné vyřešení položky následující). Matematicky pak lze dokázat, že i ve vztahu ke skutečné znalosti celého předmětu je pravděpodobnostní rozdělení měřitelné odpovědi subjektu binomické [3]. Speciální analýza dat ukazuje, že příslušné empirické rozdělení má vyšší rozptyl nežli rozdělení binomické s danými parametry.

Obr. 1

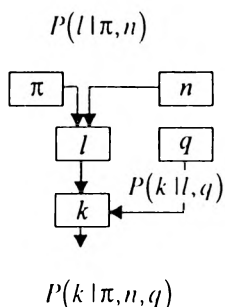


Schéma na obr. 1 lze interpretovat takto: Neovládá-li testovaný subjekt π 100 % zkoušené látky, lze očekávat, že mezi n položek didaktického testu se dostane $Y=l$ otázek z jím nezvládnuté oblasti ($0 \leq l \leq n$) s pravděpodobností

$$P(Y = l | \pi, n) = \binom{n}{l} \pi^l (1 - \pi)^{n-l} \quad (1)$$

Tato binomická pravděpodobnost udává tedy zároveň možnost, že subjekt dostane mezi n položkami testu ($n-l$) položek z oblasti, kterou ovládá. Položky, které „padly“ do oblasti jemu známých poznatků, vyřeší subjekt správně. Správně však může vyřešit rovněž některé z l položek, patřících do oblasti poznatků, které neovládá - díky mechanismu umožňujícímu náhodné uhádnutí správné odpovědi.

Označíme-li jako k počet testových položek, které subjekt vyřeší nesprávně, potom je $l-k$ počet položek správně vyřešených díky náhodě ($0 \leq k \leq l$). Situaci zachycuje schematicky obr. 2 na další straně.

Pravděpodobnost, že z l položek testu, které subjekt řešit neumí, se mu podaří náhodně uspět v $l-k$ případech (k položek bude řešeno chybně), je rovna

$$P(X = k | l, q) = \binom{l}{k} \left(\frac{q-1}{q} \right)^k \left(\frac{1}{q} \right)^{l-k} \quad (2)$$

$$k = 0, 1, \dots, l$$

Binomické rozdělení (2) je vlastně podměněným rozdělením náhodné veličiny X , udávající počet chyb k v testu, za podmínky, že l otázek testu (náhodná veličina Y) padlo do oblasti poznatků subjektu neznámých. Při daném l může veličina X zřejmě nabývat hodnot $k = 0, 1, \dots, l$.

Jak už bylo naznačeno výše, je mezi veličinami X a Y podstatný rozdíl: zatímco X je veličina, jejíž hodnoty jsou přímo



pozorovatelné a registrovatelné jako výsledek testu, neposkytuje test možnost přímo registrovat stav veličiny Y . Jako výsledek testu může být pozorován počet chybně řešených položek, ne však počet položek, jejichž řešení subjekt nezná.

Pravděpodobnostní metody umožňují najít i nepodmíněné rozdělení počtu chyb X - což je mimořádně důležité, protože právě toto četnostní rozdělení je teoretickým, modelovým protějškem četností získávaných empiricky na souboru realizací testu. Dá se odvodit, že hledané rozdělení má tvar

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

$$k = 0, 1, \dots, n$$

(3)

V takto zavedeném modelu odpovědi subjektu jsou parametry n (rozsah testu) a q (počet alternativ odpovědi nabízených s každou testovou položkou) kontrolovány examina-

torem (autorem testu), zatímco π je parametr kontrolovaný zkoušeným subjektem.

Jak je patrné, jedním z parametrů zmiňovaného modelu je počet alternativních řešení q , nabízených subjektu spolu s každou testovou položkou. Je-li právě jediná z nabídek správným řešením, je ve hře $q-1$ distraktorů.

Mechanismus náhody vstupuje do uvažované situace dvakrát - a je proto také dvakrát modelem zachycován.

Poprvé intervenuje náhoda v okamžiku, kdy autor testu (examinátor) formuluje testové položky. V této fázi procesu se totiž rozhoduje o tom, kolik položek „pokryje“ problémy subjektem zvládnuté a kolik položek naopak „pokryje“ problémy subjektem nezvládnuté. Náhoda tak kontroluje, respektující hodnotu π subjektu, chování veličiny Y , tj. rozhoduje o tom, kolik se do testu dostane položek subjektem nezvládnutých.

Podruhé ovlivní náhoda chování subjektu, tj. jeho testové skóre X , ve chvíli, kdy zkoušený subjekt náhodně uhaduje správné řešení v případě položek, které ve skutečnosti „neumí“. V této fázi realizace testu sehrává

Obr. 2

n položek testu		
l položek, které subjekt řešit neumí		$n-l$ položek, které subjekt umí
k položek, které subjekt řešil chybně	$n-k$ položek řešených správně:	
	$l-k$ náhodným uhádnutím	$n-l$ díky svým znalostem

zásadní úlohu počet distraktorů $q-1$, a je proto důležité ujistit se, že všechny distraktory použité u dané testové položky jsou plně rovnocenné.

Praktická zkušenost „polního“ testování ukazuje, že ve hře není jenom počet nabídnutých alternativ odpovědi q . Ve své jednoduché, základní verzi model odpovědi předpokládá, že všechny distraktory jsou opravdu rovnocenné a odvádějí pozornost subjektu, který téma „neumí“, stejně intenzivně. To by se pak mělo empiricky projevit ve výsledcích testu tím, že u dané testové položky se soubor jedinců, kteří řešili položku chybně, rozdělí mezi distraktory co nejvíce rovnoměrně.

2. Položková analýza

Didaktický test je zpravidla aplikován jako komplex n testových položek na soubor N jedinců. Informace obsažená v záznamu o tom, jak každý z N jedinců vyřešil n položek testu, představuje matici ohodnocených odpovědí, v jejíchž $N \cdot n$ polích se umísťují nuly a jedničky jako skóre odpovědi (0 = nesprávně, 1 = správně řešení). V našem případě je položková analýza prováděna v situaci, kdy test, tj. soubor testových položek, každému subjektu sestavoval výpočetní program metodou Monte Carlo. Výsledky umožňují zkoumat obtížnost položek i funkce distraktorů – ne však počítat koeficienty diskriminační schopnosti položek.

Matice odpovědí poskytuje materiál pro analýzu dvojího druhu:

- analýzu znalostí subjektů, dovolávající se teoretického konceptu popsaného v předchozím odstavci,

- položkovou analýzu, představující jakousi „kalibraci“ testových položek jako nástroje měření znalostí. Statisticky vzato – zatímco v prvním případě se znalosti subjektu (event. některé jeho jiné vlastnosti) odhadují na

množině realizací n položek testu, odhaduje se ve druhém případě obtížnost položky (event. i jiné její vlastnosti) na množině řešení poskytovaných N subjekty.

Položková analýza je ve srovnání s analýzou znalostí mnohostrannější. Obtížnost položky je konceptuálním protějškem úrovně znalosti subjektu; totéž platí i o koeficientech diskriminační schopnosti položky. Analýza fungování distraktorů se už však váže výhradně na položku testu.

Podobně jako analýza znalostí také položková analýza vychází ze tří strukturních úrovní studované situace:

- a) podíl jedinců obecně referenční populace (do níž patří jedinci podstupující test), kteří ovládají téma formulované uvažovanou položkou;

- b) počet jedinců mezi těmi, kteří jsou testováni, již ovládají „předmět“ položky;

- c) počet jedinců, kteří vyřešili (resp. nevyřešili) úlohu zadanou testovou položkou.

Pouze úroveň c) je dostupná empirické evidenci, může být bezprostředně měřena jako skóre správných, či nesprávných řešení dané položky. Úroveň a) je podmiňována úrovní b), která pak dále ovlivňuje výsledně, testové skóre položky na úrovni c).

Specifikace modelu je analogií úvahy předložené v předchozím odstavci:

Za předpokladu, že jedinci řeší uvažovanou testovou položku vzájemně nezávisle (to se dá zajistit) a že znalosti každého z nich dávají stejnou šanci správnému řešení (což je výrazné zjednodušení pedagogické skutečnosti) – je počet jedinců, kteří neovládají „téma“ položky, veličinou s binomickým rozdělením analogickým (1), kde π lze chápat jako podíl jedinců v referenční populaci, kteří téma položky neovládají. Při daném počtu jedinců neovládajících uvažovanou úlohu je rozdělení počtu chybujících jedinců analogické formuli (2). Přímý vztah mezi počtem chybných řešení dané položky

N jedinců, kteří podstoupili test a řešili v něm uvažovanou položku	Jedinci, kteří úlohu neznají	Jedinci, kteří řeší úlohu chybně	
		Jedinci, kteří řeší úlohu správně	náhodným uhádnutím
	Jedinci, kteří úlohu "pokrytou" položkou znají		díky své znalosti

a podílem „nepoučených“ subjektů je statisticky zachytitelný binomickou formulí analogickou (3). Transponovat lze i obr. 1. Příslušná analogie k obr. 2 by měla podobu, jakou naznačuje obr. 3.

3. Ověření funkčnosti distraktorů

Jestliže se funkčnost použitých $q-1$ distraktorů projeví rovnoměrným nebo téměř rovnoměrným rozdělením souboru chybujících subjektů, pak by měla výrazně nerovnoměrná distribuce chybujících subjektů na množině distraktorů indikovat jejich dysfunkčnost. To je důvod, proč je rozumné a žádoucí sledovat četnostní rozdělení subjektů mezi alternativy odpovědi pro každou položku testu, jako součást položkové analýzy.

Hodnocení rovnoměrnosti (uniformity) rozdělení četností lze zjednodušit, a tedy usnadnit použitím entropie.

Statistická charakteristika

$$H_a = - \left\{ \frac{v_1}{N - N_j} \log \frac{v_1}{N - N_j} + \dots + \dots + \frac{v_{q-1}}{N - N_j} \log \frac{v_{q-1}}{N - N_j} \right\} \quad (4)$$

(j je index testové položky), pomocí níž se entropie H zavádí, nabývá svých hodnot na stupnici od minima nula do maxima $\log(q-1)$; podle volby základu logaritmu se pak entropie měří v jednotkách bit (zvolen dvojkový, binární logaritmus), nit (zvolen přirozený logaritmus o základu e), případně dit (zvolen desítkový, dekadický logaritmus). Minimální hodnoty 0 se přitom dosahuje v případě, kdy se všechny chybné odpovědi soustředí na volbu jediného distraktoru (lhostejno kterého). Maxima $\log(q-1)$ dosahuje entropie tehdy, je-li příslušné četnostní rozdělení zcela rovnoměrné a každý distraktor byl zvolen týmž počtem chybujících subjektů.

Ve vzorci (4) označuje přitom
 N ... počet jedinců, kteří podstoupili test
 N_j ... počet jedinců, kteří položku j vyřešili správně
 $v_1, \dots, v_{q-1}$ je četnostní rozdělení chybných odpovědí mezi $q-1$ distraktorů,
 $v_1 + \dots + v_{q-1} = N - N_j$ [4]

Pro četnost správné odpovědi je vyhrazeno značení v_q .

Zmíněné vlastnosti entropie nabízejí možnost jejího využití jako kritéria fungování distraktorů: čím víc se liší hodnota entropie (4) skutečného rozdělení chybujících subjektů od hodnoty maximálně dosažitelné, tím výrazněji svědčí tato evidence pro jejich funkční neúplnost. Přitom je třeba poznamenat, že onoho maxima $\log(q-1)$ může být dosaženo jen tehdy, je-li celkový počet chybujících subjektů $N - N_j$ celočíselným násobkem počtu distraktorů $q-1$. Pro několik desítek typických, běžně se vyskytujících situací shrnují - pro daný počet alternativních odpovědí q a daný počet chybujících subjektů $N - N_j$ - rozdělení hodnot entropie speciální tabulky (viz [4]).

4. Obtížnost položky a funkčnost distraktorů

Součástí zkoušky z fyziologie na Lékařské fakultě Univerzity Palackého v Olomouci je i test typu multiple-choice. Vzhledem k celkovému počtu použitých testových položek, který dosahuje několika stovek, počtu studentů (kolem 200) a počtu položek, které každý student jako součást testu absolvoval (těch bylo 100), bylo možno evidovat pro každou testovou položku asi 15-35 studentů, kteří ji jako součást své zkoušky řešili. Rozdílné počty studentů připadajících na jednotlivé položky jsou důsledkem jejich náhodného vybírání z celkového balíku položek (postup vybírání je zajišťo-

ván výpočetním programem, který rovněž garantuje vlastní zkoušení a jednotnou evidenci výsledků zkoušky formou databáze. Z této databáze vychází i naše položková analýza. Autorem programu je Ing. Josef Chytil.

Protože studenti neabsolvovali test jako předem sestavený soubor položek, omezuje se položková analýza na hodnocení obtížnosti (snadnosti) položky a hodnocení funkčnosti distraktorů. Koefficienty diskriminační schopnosti citlivosti položek hodnoceny být nemohly.

Jak je patrné z připojené tabulky 1, jsou uváděny s každou testovou položkou následující údaje:

- identifikační číslo testové položky j
- počet N subjektů, které řešily položku
- počet alternativních nabídek odpovědi q
- kód správné odpovědi (z q nabídek)
- četnostní rozdělení subjektů na množině odpovědí $v_1, \dots, v_q$
- relativní podíl správných odpovědí X_j/N
- aktuální entropie H_j počítaná podle (4)
- maximální dosažitelná entropie při daném q a $N - N_j$
- index F_j snadnosti položky j , odvozený z relativního podílu správných odpovědí s použitím korekce na náhodné uhádnutí, podle vztahu

$$F_j = \frac{X_j}{N} - \frac{N - X_j}{N} \frac{1}{q-1} \quad (5)$$

5. Účinnost distraktorů

Počet nabídek alternativní odpovědi q a tedy i počet distraktorů $q-1$ ovlivňují míru intervence náhody mechanismem náhodného uhádnutí v případě zcela nepoučeného subjektu neznalého tématu, jak je patrné ve vzorci (2). Platí přitom, že vysoký počet nabídek q snižuje možnost dospět



náhodně ke správnému řešení úlohy. Parametr q ovlivňuje také index snadnosti F_j položky j , jak je patrné ze vzorce (5). Index snadnosti se přitom dá interpretovat jako podíl látky zkoušeného tématu, který subjekt opravdu ovládá. Navíc hraje parametr q a zejména to, zdali všechny distraktory plně fungují, zásadní roli při určování četnostního rozdělení $v_1, \dots, v_q$.

Lze konstatovat, že soubor $q-1$ distraktorů funguje naplno jenom výjimečně. Znamená to, že by se vlastně mělo uvažovat o jakémusi „efektivním“ počtu distraktorů, což je číslo, které by se dalo odvodit právě na základě aktuální hodnoty entropie. Jedním z takových návrhů je postup daný formulí

$$q_{ef} = 1 + \exp(H) \quad (6)$$

Odvození a zdůvodnění diskutuje práce [4]. Prakticky je tedy vždy hodnota F_j nadhodnocována - látku „pokrytou“ položkou j ovládá nižší podíl populace, než naznačuje vypočítaná hodnota F_j .

Rozhodnutí o tom, zdali fungují u dané položky distraktory ještě přijatelným způsobem, lze opírat o stejný princip, na němž spočívá testování statistických hypotéz. K rozhodnutí o event. dysfunkci distraktorů by se mělo dospívat následujícím postupem:

V tab. 1 v práci [4] se vyhledá část odpovídající použitému počtu distraktorů q a počtu $N-N_j$ subjektů, jejichž řešení byla chybná. Dílčí tabulka uvádí hodnoty entropie H_j , k nimž může volba chybujících subjektů pro tuto situaci vést, spolu s pravděpodobnostním rozdělením na množině těchto hodnot. Pokud by všech $q-1$ distraktorů rovnocenně fungovalo, měla by entropie dosahovat nejvyšších možných hodnot. Na opačném pólu stupnice jsou minimální hodnoty entropie (s extrémní hodnotou nula) a lze stanovit obor hodnot, které nejvíce svědčí v neprospěch hypotézy o fungování distraktorů a jejichž výskyt je - za platnosti hypotézy - jen

málo pravděpodobný (téměř nemožný, tj. očekávaný s pravděpodobností nejvýše 0,05). Tento obor může být považován za tzv. kritický obor pro test zmíněné hypotézy: padne-li vypočítaná hodnota entropie H_j do kritického oboru, usoudí se, že distraktory plnohodnotně nefungují. Jak se lze z dílčích tabulek přesvědčit, test na hladině signifikance 0,05 nelze aplikovat, je-li počet chybujících subjektů $N-N_j$ menší než 6 (v případě $q=3$), eventuálně menší než 4 (v případě $q=4$ nebo $q=5$).

Příklady

V tab. 1 je zachycena část výsledků, jak tyto vystoupily v případě testových položek určených pro hodnocení znalostí z tématu „dýchání“. Všimněme si některých z nich, abychom tak stručně demonstrovali diagnostické možnosti námi navrženého postupu.

Položka č. 2 je extrémně snadná - všech $N=21$ subjektů ji vyřešilo správně. Entropii počítat nelze (není z čeho), což signalizuje program údajem „-9. 999“.

Položka č. 3 je přiměřeně obtížná: $F_j=0,613$ vypovídá, že 61 % subjektů její téma ovládá. Všechny $N-N_j=31-22=9$ chybujících subjektů se rozdělilo mezi $q-1=3$ distraktory v poměru $(v_1, v_2, v_3)=(8,0,1)$, čemuž odpovídá entropie $H=0,349$. V dílčí tabulce (2 ve [4]) zjišťujeme, že hodnoty entropie 0,349 a nižší jsou při plné funkci distraktorů očekávány pouze s pravděpodobností 0,0033. To nám dává právo usuzovat, že distraktory položky $j=3$ téměř jistě plnohodnotně nefungují.

Stejný závěr lze učinit pro testovou položku $j=5$ či $j=7$.

Speciálním případem jsou testové položky, u nichž je některý z distraktorů volen subjekty častěji než správné řešení. I v takovém případě je doporučeníhodné, aby byla úloha, testovou položkou formulovaná, pečlivě revidována.

Závěr

Užitečnou součástí položkové analýzy testu typu multiple-choice je prověření funkce distraktorů každé testové položky. Je diskutován postup řešení, využívající entropii četnostního rozdělení souboru jedinců, kteří řešili uvažovanou položku chybně.

Literatura:

- [1] BYČKOVSKÝ, P. : Základy měření výsledků výuky. Tvorba didaktického testu. Praha, ČVUT 1982.
- [2] KALBFLEISCH, J. G. : Probability and Statistical Inference. Volume 1: Probability. Springer - Verlag 1985.
- [3] KOMENDA, S. - KLEMENTA, J. : Analýza náhodného v pedagogickém experimentu a praxi. Praha, SPN 1981.
- [4] KOMENDA, S. - TESAŘIKOVÁ, E. : Number of effective distractors of the school - achievement test item. Acta Univ. Palacki. Olomuc. (Olomouc), Fac. Med. 107, 417-433 (1984).

ANNEX:

PRÁVĚPODOBNOSTNÍ ROZDĚLENÍ ENTROPIE

Tab. 1. (str. 233) [4], o níž se testování (nulové) hypotézy plně účinnosti distraktorů testové položky opírá, byla vypočítána s využitím úvahy o tzv. *problému obsazení* (occupancy problem - [2]). Totéž platí i pro rozšířený výpočet pomocí výkonnější výpočetní techniky, který se aktuálně připravuje.

„Obsazením“ se rozumí rozmístění $N-N_j$ chybných řešení položky j mezi jejími $q-1$ distraktory. Při rozlišování identity chybně řešivších subjektů je takových možných rozmístění $(q-1)^{N-N_j}$ - a za platnosti nulové hypotézy je každé z těchto rozmístění stejně pravděpodobné.

Jednoduchý kombinatorický model této situace se dostane tak, že se do čepice vloží

$q-1$ lístků s čísly $1, \dots, q-1$ a $(N-N_j)$ krát se vždy po pečlivém zamíchání - táhne lístek, zaznamenaná na něm zapsaná čísla a tažený „distraktor“ se vrací zpět do čepice. Výsledkem $N-N_j$ tažení je uspořádaná posloupnost četností $v_1, \dots, v_{q-1}$, $v_1 + \dots + v_{q-1} = N-N_j$.

Situaci, při níž $N-N_j$ chybujících subjektů „obsadí“ $q-1$ distraktorů s četnostmi $v_1, \dots, v_{q-1}$, očekáváme s multinomickou pravděpodobností

$$\frac{(N-N_j)!}{v_1! \dots v_{q-1}!} \left(\frac{1}{q-1} \right)^{N-N_j} \quad (7)$$

$$v_1 + \dots + v_{q-1} = N-N_j \\ 0 \leq v_1, \dots, v_{q-1} \leq N-N_j$$

Pro četnostní rozdělení $(v_1, \dots, v_{q-1})$ se zavádí entropie H_a vzorcem (4). Multinomické rozdělení pravděpodobností (7) není ještě rozdělením na množině hodnot entropie H_a . Hodnota entropie je totiž stejná pro všechna obsazení $(v_1, \dots, v_{q-1})$, která se liší pouze permutací svých četností. Takových permutací je $(q-1)!/I$, kde

$$I = w_1! \dots w_k!, \quad 1 \leq w_1, \dots, w_k \leq q-1, \\ w_1 + \dots + w_k = q-1, \quad 1 \leq k \leq q-1;$$

„ w “ přitom označují, kolikrát se ta která z $q-1$ četností $v_1, \dots, v_{q-1}$ opakuje.

Entropie H_a nabývá tedy své hodnoty (4) s pravděpodobností

$$\frac{(N-N_j)!}{v_1! \dots v_{q-1}!} \left(\frac{1}{q-1} \right)^{N-N_j} \cdot \frac{(q-1)!}{I} \quad (8)$$

Příklad:

Uvažujme testovou položku s $q-1=3$ distraktory, kterou řešilo chybně $N-N_j=6$ subjektů. Počet různých, stejně pravděpodobných obsazení je tedy roven

$(q-1)^{N-N_j}=3^6=729$. Protože naše statistika nerozlišuje, který subjekt „obsadil“ ten který distraktor, a bere v úvahu jenom počet subjektů, redukuje se množina různých



obsazení na podstatně menší počet typů, totiž ty, které uvádí tab. 2.

Praktické využití tabulky probíhá takto:

Kdyby všechny $q=1=3$ distraktory položky j účinně fungovaly, měly by být obsazeny $N-N_j=6$ chybujících subjekty co nejvíce rovnoměrně (a tomu odpovídající entropie by měla být co nejvyšší). Jestliže naopak bude mít vypočítaná entropie hodnotu nepravděpodobně signifikantně nízkou, tj. nižší než 0,05, usoudíme, že distraktory účinně

nefungují. Podíváme-li se na tabulku, zjišťujeme, že k signifikantnímu výsledku vedou hodnoty entropie rovné nule, případně 0,4506. Řečeno statisticky, kritickým oborem testu hypotézy o účinnosti fungování distraktorů je množina $\{H=0, H=0,4506\}$. Test nemá přitom signifikanci 0,05, ale $0,0041+0,0494=0,0535$, tj. nepatrně vyšší.

Z tabulky je zřejmé, že v případě malého počtu chybujících subjektů ($N-N_j \leq 5$) je informace k zamítnutí testované hypotézy nedostatečná.

Tab. 2

v_1, v_2, v_3	(1) $\frac{(N-N_j)!}{v_1! v_2! v_3!}$	(2) $\frac{(q-1)!}{f}$	(3) $(3)=(1)(2)$	H_j
6 0 0				
0 6 0	$\frac{6!}{6! 0! 0!} = 1$	$\frac{3!}{1! 2!} = 3$	$\frac{3}{729} = 0,0041$	0
0 0 6				
5 1 0				
5 0 1				
0 5 1	$\frac{6!}{5! 1! 0!} = 6$	$\frac{3!}{1! 1! 1!} = 6$	$\frac{36}{729} = 0,0494$	0,4506
1 5 0				
1 0 5				
0 1 5				
4 1 1				
1 4 1	$\frac{6!}{4! 1! 1!} = 30$	$\frac{3!}{1! 2!} = 3$	$\frac{90}{729} = 0,1235$	0,6365
1 1 4				
3 3 0				
3 0 3	$\frac{6!}{3! 3! 0!} = 20$	$\frac{3!}{2! 1!} = 3$	$\frac{60}{729} = 0,0823$	0,6931
0 3 3				
4 2 0				
4 0 2				
2 4 0	$\frac{6!}{4! 2! 0!} = 15$	$\frac{3!}{1! 1! 1!} = 6$	$\frac{90}{729} = 0,1235$	0,8676
0 4 2				
2 0 4				
0 2 4				
3 2 1				
3 1 2				
2 3 1	$\frac{6!}{3! 2! 1!} = 60$	$\frac{3!}{1! 1! 1!} = 6$	$\frac{360}{729} = 0,4938$	1,0114
1 3 2				
2 1 3				
1 2 3				
2 2 2	$\frac{6!}{2! 2! 2!} = 90$	$\frac{3!}{3!} = 1$	$\frac{90}{729} = 0,1235$	1,0986
Σ			1,0000	

Tabulka I

pol. j	počet subj.	poč. alt.	správ. alt.	četnosti alternativ					$\frac{X_{\cdot j}}{N}$	entrop. IIa	max. II	index Fj
1	27	4	3	3	1	16	7		0.593	0.860	1.090	0.457
2	21	4	1	21	0	0	0		1.000	-9.999	0.000	1.000
3	31	4	4	8	0	1	22		0.710	0.349	1.099	0.613
4	27	4	4	1	6	1	19		0.704	0.736	1.082	0.605
5	18	4	3	0	1	11	6		0.611	0.410	1.079	0.481
6	24	4	2	3	20	1	0		0.833	0.562	1.040	0.778
7	27	4	4	4	0	0	23		0.852	0.000	1.040	0.802
8	27	4	2	1	26	0	0		0.963	0.000	0.000	0.951
9	32	4	2	1	26	0	5		0.813	0.451	1.099	0.750
10	27	4	2	2	13	8	4		0.481	0.956	1.093	0.309
11	20	4	3	3	2	12	3		0.600	1.082	1.082	0.467
12	21	4	3	2	3	13	3		0.619	1.082	1.082	0.492
13	24	4	3	3	1	19	1		0.792	0.950	1.055	0.722
14	20	4	4	0	4	1	15		0.750	0.500	1.055	0.667
15	19	4	1	17	0	0	2		0.895	0.000	0.693	0.860
16	19	4	4	0	1	0	18		0.947	0.000	0.000	0.930
17	30	4	3	1	0	29	0		0.967	0.000	0.000	0.956
18	32	4	4	3	0	1	28		0.875	0.562	1.040	0.833
19	23	4	1	14	3	3	3		0.609	1.099	1.099	0.478
20	19	4	2	2	16	1	0		0.842	0.637	1.099	0.789
21	26	4	4	1	6	0	19		0.731	0.410	1.079	0.641
22	22	4	3	0	4	18	0		0.818	0.000	1.040	0.758
23	23	4	3	2	0	16	5		0.696	0.598	1.079	0.594
24	18	4	4	0	0	1	17		0.944	0.000	0.000	0.926
25	26	4	1	17	2	6	1		0.654	0.849	1.099	0.538
26	30	4	2	0	30	0	0		1.000	-9.999	0.000	1.000
27	31	4	3	5	7	12	7		0.387	1.087	1.096	0.183
28	33	4	1	32	0	1	0		0.970	0.000	0.000	0.960
29	29	4	4	4	1	0	24		0.828	0.500	1.055	0.770
30	23	4	3	3	0	13	7		0.565	0.611	1.089	0.420