Praktická ukázka využití testování hypotéz

ONDŘEJ VENCÁLEK

Přírodovědecká fakulta UP, Olomouc

Tradiční kurzy pravděpodobnosti a statistiky na všech úrovních vzdělávání začínají příklady z oblasti hazardních her – hody mincí, jednou či vícero kostkami, tahy čísel z osudí. Je to snad dáno tím, že právě oblast hazardních her vedla k zájmu o nahodilost a potřebě jejího popisu jazykem matematiky. Pokud však při výuce u těchto příkladů zůstaneme a nedojde na příklady poněkud závažnější, nabude velká část studentů dojmu, že „celá pravděpodobnost a statistika se vlastně zabývají jen malichernými problémy". Cílem tohoto příspěvku je přesvědčit čtenáře, že navzdory výše uvedenému přesvědčení je statistika velmi užitečným nástrojem, který napomáhá k rozvoji nejrůznějších oblastí lidského vědění.

Úskalí zobecňování

Předmětem zkoumání statistiky jsou nejrůznější vztahy. Představme si například prodejce, který prodává nějaké přístroje (např. fotoaparáty) dvou různých typů. Vede si záznamy o tom, zda u prodaných kusů došlo v záruční době k poruše a následné reklamaci či nikoliv. U typu A, kterého se prodalo 200 kusů, zaznamenal čtyři reklamace, u typu B, kterého se prodalo 100 kusů, jen jednu. Reklamována byla tedy 2 % přístrojů typu A a 1 % přístrojů typu B. Lze tedy říci, že typ A je *obecně* poruchovější? Slůvkem *obecně* zde míníme skutečnost, že pro další (zatím neprodané) výrobky bude opět poruchovost typu A vyšší než typu B.

Výrobek typu A se jeví jako více poruchový, a to dvojnásobně oproti typu B. Když by však bylo byť jen o jednu reklamaci typu B více, byly by u obou typů výrobku 2 % reklamací. Nezdá se tedy fér tvrdit něco obecného na základě tak malého rozdílu v pozorované poruchovosti.

Představme si ještě několik dalších situací. V tabulce 2 je popsána situace, kdy jsou počty reklamací vyšší než v tabulce 1, ale stále platí, že poruchovost typu A je dvojnásobná oproti typu B (20 % oproti 10 %). Tomuto rozdílu již budeme přikládat větší význam, neboť změna jednoho či dvou pozorování nebude znamenat zásadní změnu našich závěrů o poruchovosti. V situaci prezentované v tabulce 3 si pak již můžeme být rozdílností v poruchovosti jednotlivých typů dost jisti. Procentuální zastoupení reklamovaných výrobků je sice stejné jako v předešlé situaci, ale máme podstatně více pozorování.

Tabulka 1 Četnost reklamací u jednotlivých typů přístroje

	reklamace	bez závad	celkem
typ A	4	196	200
typ B	1	99	100
celkem	5	295	300

Tabulka 2 Četnost reklamací u jednotlivých typů přístroje (situace s vyšší poruchovostí)

	reklamace	bez závad	celkem
typ A	40	160	200
typ B	10	90	100
celkem	50	250	300

Tabulka 3 Četnost reklamací u jednotlivých typů přístroje (situace s vyšší poruchovostí a vyšším počtem pozorování)

	reklamace	bez závad	celkem
typ A	400	1 600	2 000
typ B	100	900	1 000
celkem	500	2 500	3 000

Otázka, kterou si ve všech třech výše uvedených situacích klademe, je stejná: „Může být pozorovaná rozdílnost v poruchovosti dvou typů přístroje náhodná, nebo jde o obecné pravidlo?“ Statistika dává „návod“, jak tuto otázku zodpovědět.

Od označení k hypotéze

U žádného z přístrojů nevíme předem, zda bude reklamován či nikoliv. Výskyt závady tedy lze považovat za *náhodný jev.* Míru poruchovosti přístrojů můžeme vyjádřit pomocí pravděpodobnosti výskytu závady. Budeme ji značit P(Z). Je-li například P(Z) = 0,1, znamená to, že k závadě dojde u přibližně 10 % výrobků. Poruchovost může záviset na typu výrobku. Označme proto P(Z|A) pravděpodobnost výskytu závady u výrobků typu A a P(Z|B) pravděpodobnost výskytu závady u výrobků typu B.

Zajímá nás, zda je *obecně* pravděpodobnost výskytu závady u výrobků typu A stejná jako u výrobků typu B, tedy zda platí

$$\mathrm{P(Z|A)} = \mathrm{P(Z|B)}. \tag{1}$$

V takovém případě pravděpodobnost výskytu závady *nezávisí* na typu výrobku.

Naprosto zásadní je uvědomit si, že pravděpodobnosti P(Z|A) a P(Z|B) *neznáme*. Máme k dispozici pouze jejich odhady učiněné na základě našich pozorování, a to

P(Z|A) odhadneme hodnotou 4/200 = 0,02,

P(Z|B) odhadneme hodnotou 1/100 = 0,01.

To, že se mezi 200 již prodanými výrobky typu A vyskytly 4 závady totiž neznamená, že se v dalších 200 prodaných výrobcích typu A opět vyskytnou přesně 4 závady. Očekáváme sice, že by jich mohlo být kolem čtyř, ale mírně větší či menší počet je jistě možný také.

Otázka, zda je pozorovaná rozdílnost v poruchovosti dvou typů přístroje náhodná, nebo jde o obecné pravidlo, je vlastně otázkou, zda platí rovnost (1). Skutečnost, že platí rovnost (1), je nějaká *hypotéza*, konkrétně jde o *hypotézu nezávislosti* poruchovosti a typu výrobku. Ta může a nemusí být pravdivá. O její pravdivosti chceme rozhodnout na základě námi učiněných pozorování. Konfrontace hypotézy (rovnost (1)) s daty (tabulka 1) je příkladem jednoho z hlavních úkolů statistiky – *testování hypotéz*.

Trocha počítání (od otázek k odpovědím)

Uvažujme situaci, uvedenou na začátku, tedy celkem 5 reklamací u celkem 300 přístrojů, z nichž 2/3 jsou typu A a 1/3 typu B. Pokud by hypotéza platila, očekávali bychom, že stejné zastoupení (tj. 2/3 a 1/3) jednotlivých typů bude i mezi reklamovanými, avšak 2/3 z 5 nejsou celé číslo. Tři reklamace u typu A a dvě u typu B je tedy možnost blízká našim očekáváním vycházejícím z předpokladu nezávislosti. Drobná kolísání jsou však jistě díky náhodě možná.

Označme symbolem X počet reklamací u typu A. Při celkovém počtu pět reklamací může nastat jen jedna ze šesti následujících možností: $X = 0$ (pokud všech pět reklamací je u typu B), $X = 1$, $X = 2$, $X = 3$, $X = 4$, nebo $X = 5$ (pokud všech pět reklamací je u typu A). Tyto možnosti však *nejsou stejně pravděpodobné*. Již jsme si ukázali, že pokud platí hypotéza nezávislosti, bude s velkou pravděpodobností hodnota X blízká 3.

Výpočet pravděpodobností jednotlivých možností je dán vzorcem

$$\mathrm{P}(X = k) = \frac{\dbinom{200}{k}\dbinom{100}{5-k}}{\dbinom{300}{5}}, \quad k = 0, 1, 2, \ldots, 5. \tag{2}$$

Ve jmenovateli je počet všech možností, které z 300 přístrojů mohou být těmi pěti reklamovanými. Tyto možnosti jsou za platnosti hypotézy stejně pravděpodobné. V čitateli jsou započteny už jen možnosti, kdy nastane jev $X = k$. Je celkem $\binom{200}{k}$ možných k-tic pokažených přístrojů typu A a ke každé k-tici existuje $\binom{100}{5-k}$ možných $(5-k)$-tic pokažených přístrojů typu B.

Po dosazení hodnot $k = 0, \ldots, 5$ do vzorce (2) dostaneme pravděpodobnosti shrnuté v tabulce 4.

Tabulka 4 Pravděpodobnosti vypočtené podle vzorce (2)

k	0	1	2	3	4	5
$\mathrm{P}(X = k)$	0,004	0,040	0,164	0,332	0,330	0,129

Je-li alternativou k naší hypotéze možnost, že přístroje typu A jsou poruchovější, tj. $\mathrm{P}(Z|A) > \mathrm{P}(Z|B)$, svědčí ve prospěch této alternativy vysoké hodnoty X.

Stále předpokládáme, že platí hypotéza nezávislosti. Předpokládanou hypotézu nazýváme nulovou. Položme si nyní otázku, nakolik „extrémním“ výsledkem (svědčícím ve prospěch alternativy) je za tohoto předpokladu naše pozorování $X = 4$. Tedy, s jakou pravděpodobností dojde k tomu, že z pěti porouchaných přístrojů budou čtyři nebo více typu A. Vypočteme proto

$$\mathrm{P}(X \geq 4) = \mathrm{P}(X = 4) + \mathrm{P}(X = 5) \doteq 0{,}330 + 0{,}129 = 0{,}459.$$

Čtyři nebo více poruch je tedy poměrně očekávaný výsledek (cca 46 %), i když hypotéza platí. Poznamenejme, že jsme se právě seznámili s tzv. *Fisherovým faktoriálovým testem.* Hodnotě pravděpodobnosti, kterou jsme vyčíslili na 0,459 se říká *p-hodnota* (anglicky *p-value*).

Kdyby byla p-hodnota příliš nízká, např. 0,001, odmítli bychom hypotéze nadále věřit, tj. řekli bychom, že *hypotézu nezávislosti zamítáme.* Uvědomme si, že se může stát, že zamítneme i platnou hypotézu. Například výše uvedená pravděpodobnost 0,001 je sice malá, za platnosti nulové hypotézy však asi v jednom z tisíce případů takový výsledek nastane a my hypotézu zamítneme nesprávně.

Statistici tedy vědí, že se při zamítnutí (ale i případném nezamítnutí) hypotézy mohou dopustit chyby. Vědí ale, že k tomu dojde jen s malou pravděpodobností. Dokonce mohou tuto pravděpodobnost kontrolovat, a to volbou parametru α v následujícím rozhodovacím pravidle: Je-li

p-hodnota $\leq \alpha$, pak zamítáme hypotézu,

p-hodnota $> \alpha$, pak nezamítáme hypotézu.

Pravděpodobnost chybného zamítnutí hypotézy, která platí, je pak nejvýše α. Hodnotu tohoto parametru sami dopředu zvolíme. Obvykle se volí $\alpha = 0{,}05$, tj. připouštíme 5% pravděpodobnost zamítnutí platné hypotézy. Parametru α se říká *hladina testu.*

Vraťme se ještě k tabulkám 2 a 3. Pro tabulku 2 je p-hodnota 0,019. Pro tabulku 3 pak vyjde p-hodnota řádově 10^{-13}. V obou případech je tedy p-hodnota menší než obvyklá hodnota $\alpha = 0{,}05$. Znamená to, že v situacích popsaných v tabulkách 2 a 3 zamítáme hypotézu nezávislosti na hladině 5 %. Zatímco v situaci popsané v úvodu (tabulka 1) rozdíl v poruchovosti není statisticky významný, v obou dalších situacích už naše pozorování poskytují dost silný argument k tomu, abychom mohli tvrdit, že typ A je obecně poruchovější než typ B.

Příklady z oblasti psychologie a medicíny

V předešlé části jsme se seznámili s jednou statistickou technikou – Fisherovým faktoriálovým testem nezávislosti. Nyní si ukážeme dvě situace, v nichž je možné tento test využít.

Příklad z oblasti psychologie

Daniel Kahneman představuje v knize *Myšlení rychlé a pomalé* [2] svůj pohled na to, jak se generují intuitivní názory na složité záležitosti. Tvrdí, že pokud naše intuice nenajde na nějakou obtížnější otázku rychle uspokojivou odpověď, vyhledá související otázku, která je snazší, a na tu pak odpoví. Tento fenomén nazývá *substitucí*.

Kahneman cituje německou studii, jejímiž účastníky byli studenti, kteří měli odpovídat na dvě otázky:

- Jak jste v poslední době spokojeni?
- Kolikrát jste byli za poslední měsíc na rande?

Realizátoři studie chtěli zjistit, zda odpovědi na tyto otázky spolu souvisejí. Zjistili však (poměrně překvapivě), že nikoliv. Evidentně studentům při otázce na jejich spokojenost nepřišly schůzky na mysl. Jiné skupině byly položeny tytéž otázky, ale v opačném pořadí:

- Kolikrát jste byli za poslední měsíc na rande?
- Jak jste v poslední době spokojeni?

Výsledky byly zcela odlišné. Zjistila se úzká vazba mezi mírou spokojenosti a počtem schůzek.

Kahneman tyto výsledky vysvětluje tím, že zde šlo o substituci:

> *Randění zjevně nepředstavovalo ústřední prvek života těchto studentů . . . , ale když byli nejprve dotázáni na milostný život, vyvolalo to v nich určitou emoční reakci. Studenti, kteří měli hodně schůzek, si připomněli tento šťastný aspekt svého života, zatímco těm, kteří schůzky neměli, to připomenulo jejich osamělost nebo nepříjemné zkušenosti s odmítnutím. Emoce vzbuzené otázkou na schůzky ještě měli všichni v mysli, když odpovídali na další otázku o své celkové spokojenosti.*

Pojďme se nyní podívat, jak mohla tato studie vypadat. Zjednodušíme odpovědi studentů ohledně spokojenosti na *spokojen* a *nespokojen*, a ohledně schůzek na *měl* a *neměl*.

Tabulka s četnostmi jednotlivých kombinací pro skupinu, která nejprve odpovídala na otázku týkající se štěstí, by mohla vypadat jako tabulka 5. Tabulka 6 pak udává četnosti jednotlivých kombinací odpovědí pro skupinu, která nejprve odpovídala na otázku týkající se schůzek.

Jestliže použijeme dříve uvedený Fisherův test nezávislosti, dospějeme k p-hodnotě 0,253 pro skupinu 1 a 0,001 pro skupinu 2. Zatímco v prvním případě tedy hypotézu nezávislosti nemůžeme zamítnout, u druhé skupiny je významně vyšší zastoupení šťastných ve skupině těch, kteří měli schůzku, oproti skupině těch, kteří schůzku neměli.

Tabulka 5 Počty studentů podle jejich odpovědí týkajících se štěstí a schůzek – skupina 1

	šťastný	nešťastný	celkem
měl schůzku	20	40	60
neměl schůzku	10	30	40
celkem	30	70	100

Tabulka 6 Počty studentů podle jejich odpovědí týkajících se štěstí a schůzek – skupina 2

	šťastný	nešťastný	celkem
měl schůzku	25	35	60
neměl schůzku	5	35	40
celkem	30	70	100

Příklad z oblasti medicíny

Přišel vám předešlý výsledek zajímavý? Následující příklad bude nejen zajímavý, ale výsledek zde zjištěný je i velice důležitý.

Výzkumníci z Ústavu hematologie a krevní transfuze v Praze ve spolupráci s několika dalšími pracovišti sbírají a analyzují data týkající se zhoubných nádorů hlavy a krku. Podařilo se jim ukázat, že pacienti, kteří mají virový původ zhoubného nádoru (jde o tzv. papilomaviry) mají významně lepší prognózu přežití. V tabulce 7 jsou pacienti rozděleni jednak podle HPV pozitivity[1] (tj. pozitivní pacienti mají papilomavirový původ nádoru) a jednak podle toho, zda přežili tři roky od diagnózy.

Tabulka 7 Počty pacientů podle přežití doby 3 let od diagnózy a podle HPV pozitivity.

	přežil	zemřel	celkem
HPV pozit.	66	17	83
HPV negat.	45	63	108
celkem	111	80	181

Test nezávislosti hovoří jasně: p-hodnota $= 9 \cdot 10^{-8} < 0{,}05$. Pacienti s virovým původem nádoru mají významně vyšší pravděpodobnost přežití

[1] Zkratka HPV pochází z anglického označení human papilloma virus.

doby tří let od diagnózy. V analýze této datové sady samozřejmě musí statistik zohlednit i další důležité faktory, které mají či mohou mít vliv na délku přežití pacienta, jako jsou například věk, pohlaví nebo, zda jde o kuřáka či nekuřáka.

Výzkumnou zprávu shrnující zjištěné výsledky se podařilo publikovat v prestižním časopise *International Journal of Cancer* [3]. Předmětem dalšího zkoumání je, zda pacienti HPV pozitivní reagují jinak na různé typy léčby než pacienti HPV negativní. Taková skutečnost by mohla výrazně napomoci v rozhodování o způsobu terapie.

Literatura

[1] *Calda, E. – Dupač, V.*: Matematika pro gymnázia – Kombinatorika, pravděpodobnost a statistika. Prometheus, Praha, 2006.

[2] *Kahneman, D.*: Myšlení rychlé a pomalé. Jan Melvil Publishing, Brno, 2012.

[3] *Vojtěchová, Z. a kol.*: Analysis of the integration of human papillomaviruses in head and neck tumours in relation to patients' prognosis. International Journal of Cancer, roč. 138 (2016), č. 2, s. 386–395.

Zajímavé matematické úlohy

Uveřejňujeme další část pravidelné rubriky Zajímavé matematické úlohy a uvádíme zadání další dvojice úloh. Řešení nových úloh 227 a 228 můžete zaslat nejpozději do 20. 11. 2016 na adresu: Redakce časopisu MFI, 17. listopadu 12, 771 46 Olomouc nebo také elektronickou cestou (pouze však v TEXovských verzích, příp. v MS Wordu) na emailovou adresu: *mfi@upol.cz*.

Úloha 227

Je dán pravoúhlý trojúhelník ABC, pro jehož odvěsny platí nerovnost $|AC| > |BC|$. Nechť D je průsečík osy vnitřního úhlu při vrcholu C s přeponou AB a E, F nechť značí po řadě středy kružnic opsaných trojúhelníkům BDC, ADC. Dokažte, že průsečík P osy vnitřního úhlu při vrcholu A a přímky EF je středem kružnice vepsané trojúhelníku CDE.

Jaroslav Švrček & Waldemar Pompe (*Warszawa*)