Petr Emanovský

Jak pivovarský sládek způsobil revoluci ve statistice

Učitel matematiky, Vol. 29 (2021), No. 2, 96–110

Persistent URL: http://dml.cz/dmlcz/149001

JAK PIVOVARSKÝ SLÁDEK ZPŮSOBIL REVOLUCI VE STATISTICE

PETR EMANOVSKÝ[1]

Úvod

Statistické testování hypotéz neboli testy statistické významnosti představují jednu ze základních metod matematické statistiky, kterou dnes běžně používají nejen profesionální statistici, ale i vědečtí pracovníci a studenti při svých výzkumech. Myšlenky, na nichž je tato metoda založena, se objevují již počátkem 18. století v práci skotského matematika Johna Arbuthnota (Arbuthnot, 1710; Emanovský, 2021). Do dnešní podoby byla tato metoda rozvinuta až ve 20. a 30. letech 20. století a je spojována zejména se jmény matematiků Ronalda Fishera (1890–1962), Jerzyho Neymana (1894–1981) a Egona Pearsona (1895–1980). Významnou měrou k rozvoji této metody však přispěl rovněž méně známý William Sealy Gosset (1876–1937), autor nových statistických metod, které vyvinul během svého působení v pivovaru Guinness v irském Dublinu. Gosset spolupracoval s významným statistikem Karlem Pearsonem, který mu pomáhal s matematickým obsahem jeho článků, včetně významného článku *The probable error of a mean* z roku 1908.

Statistická inference jako základní princip matematické statistiky

Statistickou inferencí neboli indukcí rozumíme induktivní myšlenkový postup, kdy na základě určitých vlastností náhodně vybraného vzorku usuzujeme o vlastnostech základního souboru, z něhož byl vzorek vybrán. Toto zobecnění je založeno na výpočtu tzv.

[1]Článek vznikl za podpory projektu Univerzity Palackého Olomouc „Matematické struktury“ IGA PrF 2020 014.

výběrových charakteristik (statistik), které si můžeme představit jako vzorce, do kterých dosazujeme hodnoty zjištěné u vybraného vzorku. Výběrové charakteristiky jsou náhodné veličiny, jejichž hodnoty jsou pro různé vzorky různé, takže závisí na náhodě. Má tedy smysl hovořit o pravděpodobnostním rozdělení takto získaných hodnot neboli o výběrovém rozdělení. Výběrovým rozdělením náhodné veličiny (statistiky) máme na mysli pravděpodobnostní rozdělení hodnot, které náhodná veličina nabývá ve všech možných výběrech o daném rozsahu ze specifikovaného základního souboru (Hendl, 2004). Odhadujeme-li některý neznámý parametr základního souboru, např. střední hodnotu μ, provedeme to tak, že určíme tzv. výběrový průměr $\bar{x}_1$ pro náhodně vybraný vzorek. Ten se s největší pravděpodobností od skutečné hodnoty μ liší. Pokud náhodně vybereme jiný stejně rozsáhlý vzorek, dostaneme pravděpodobně jinou hodnotu výběrového průměru, řekněme $\bar{x}_2$. Opakováním tohoto postupu bychom mohli dostat výběrové průměry $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$ pro k pozorování. Vznikla by tak nová náhodná veličina, která každému vzorku přiřazuje jeho výběrový průměr. Pokud každému z možných výběrových průměrů přiřadíme pravděpodobnost, s jakou jej můžeme dostat, vznikne tzv. výběrové rozdělení průměrů. Takto vzniklá náhodná veličina má svou teoretickou střední hodnotu $\mu_{\bar{x}}$ a teoretickou směrodatnou odchylku $\sigma_{\bar{x}}$. Parametr $\sigma_{\bar{x}}$ se také nazývá směrodatná chyba odhadu (průměru) nebo střední chyba průměru. Za předpokladu, že základní soubor má normální rozdělení, je známo, že $\mu_{\bar{x}} = \mu$ a $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. Zjednodušeně řečeno, průměr výběrových průměrů odpovídá skutečnému průměru původní náhodné veličiny a rozptýlenost výběrových průměrů kolem této střední hodnoty se s rostoucím rozsahem vzorků zmenšuje. Vzhledem k tomu, že přesnou hodnotu σ zpravidla neznáme, odhadujeme hodnotu směrodatné chyby $\sigma_{\bar{x}}$ hodnotou $s_{\bar{x}}$ výběrové směrodatné chyby průměru, kterou můžeme vypočítat pomocí zjištěných výběrových průměrů a vzorce

$$s_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{k} (\bar{x}_i - \bar{\bar{x}})^2}{k-1}},$$

kde $\bar{\bar{x}}$ je aritmetický průměr ze zjištěných výběrových průměrů.

Výběrovou směrodatnou chybu lze rovněž určit ze vztahu $s_{\bar{x}} = \frac{s}{\sqrt{n}}$, kde s je výběrová směrodatná odchylka náhodně vybraného vzorku (Hendl, 2004). Díky Gossetovi víme, že testovací statistika tvaru $\frac{\bar{x}-\mu}{s_{\bar{x}}}$ má známé pravděpodobnostní rozdělení, tzv. t-rozdělení, pro jehož hodnoty sestavil tabulku. Máme-li odhadnout hodnotu parametru μ, můžeme tuto neznámou hodnotu porovnat s konkrétní hodnotou μ_0, o které se domníváme, že by mohla být blízká skutečné hodnotě μ. Provedeme to v rámci statistického testování nulové hypotézy $H_0\colon \mu = \mu_0$. Právě statistické testování hypotéz neboli určování statistické významnosti představuje jednu z typických forem statistického usuzování, při němž se uplatňuje princip statistické inference. Tato metoda je podrobně popsána v každé učebnici základů matematické statistiky, viz např. (Hendl, 2004; Chráska, 2007; Klementa et al., 1984; Komenda, 1994). Předmětem tohoto testování může být pouze statistická hypotéza týkající se tzv. hromadného jevu, jehož výskyt či nepřítomnost lze sledovat opakovaně při velkém počtu pozorování. Testování je založeno na porovnání pozorovaného výsledku, který jsme zjistili u náhodně vybraného vzorku s teoretickým matematickým modelem, který předpokládá platnost tzv. nulové hypotézy. Na základě tohoto srovnání nulovou hypotézu buď zamítáme, nebo nezamítáme. Rozhodujícím kritériem tohoto rozhodovacího procesu je přitom výpočet vhodné výběrové charakteristiky pro náhodně vybraný vzorek. Tuto vypočtenou hodnotu srovnáváme s tabulkovou (kritickou) hodnotou. Např. je-li pro alternativu $H_A\colon \mu > \mu_0$ vypočtená hodnota větší než hodnota kritická, zamítáme nulovou hypotézu na zvolené hladině významnosti. V tomto případě existuje totiž pravděpodobnost, že nesprávně zamítneme nulovou hypotézu neboli že se dopustíme chyby 1. druhu, menší než riziko, které jsme ochotni akceptovat (hladina významnosti). Poznamenejme, že v dnešní době použijeme při statistickém testování hypotéz místo tabulky spíše počítač s vhodným softwarem. Pak bude při rozhodování klíčová tzv. p-hodnota, tj. pravděpodobnost chyby 1. druhu. Vypočtenou p-hodnotu srovnáváme s tzv. hladinou významnosti, tedy s rizikem chyby 1. druhu, kterou jsme ještě ochotni připustit. Teoretickým srovnávacím modelem bývá vzorec (testové kritérium, statistika), do kterého dosazujeme naměřené hodnoty a který určuje náhodnou veličinu, jejíž pravděpodobnostní rozdělení je při platnosti nulové hypotézy známo. V dnešní době existuje řada testů statistické významnosti využívajících různé matematické modely podle typu dat a konkrétní situace statistického usuzování (Hendl, 2004). K nejčastěji používaným testům patří Studentův t-test, chí kvadrát χ^2-test a F-test. Všimněme si blíže Studentova t-testu.

Studentův t-test

Studentův t-test je jedním z nejpoužívanějších statistických testů významnosti, který patří mezi tzv. parametrické testy. Dnešní statistika zná tento test v několika variantách užívaných podle typu statistického usuzování (jednovýběrový, dvouvýběrový, párový) (Hendl, 2004). Typickým příkladem použití tohoto testu je odhad aritmetického průměru μ základního souboru při neznámém rozptylu σ. Hlavní myšlenka, na které je toto testování založeno, spočívá v tom, že výběrové rozdělení aritmetického průměru $\bar{x}$ odpovídá po standardizaci střední hodnotou μ a výběrovou střední chybou $s_{\bar{x}}$ tzv. t-rozdělení, jehož tvar závisí na počtu $n-1$ stupňů volnosti (Hendl, 2004). Symbolicky můžeme tento fakt zapsat takto:

$$\frac{\bar{X}-\mu}{s_{\bar{x}}} = \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}.$$

Se vzrůstajícím počtem stupňů volnosti se t-rozdělení blíží normovanému normálnímu rozdělení a shoda nastává teoreticky při nekonečném počtu stupňů volnosti. V praxi však můžeme t-rozdělení považovat za normované normální, pokud je n alespoň 30 (Hendl, 2004).

Příklad. Studentům 1. ročníku VŠ byl na začátku studia zadán test připravenosti na vysokoškolské studium. Test obsahoval 20 otázek, přičemž studenti vždy vybírali jednu správnou odpověď ze dvou nabízených možností. Tabulka 1 obsahuje počty správných odpovědí čtrnácti náhodně vybraných studentů. Rozhodněte na základě těchto údajů, zda jsou studenti 1. ročníku opravdu připraveni na studium, nebo zda jsou správné odpovědi výsledkem náhodného hádání.

Tab. 1: Výsledky testu vybraného vzorku studentů

i (student)	x_i (počet správných odpovědí)	$\bar{x} - x_i$	$(\bar{x} - x_i)^2$
1	12	−0,36	0,13
2	10	1,64	2,69
3	9	2,64	6,97
4	13	−1,36	1,85
5	13	−1,36	1,85
6	8	3,64	13,25
7	11	0,64	0,41
8	7	4,64	21,53
9	14	−2,36	5,57
10	11	0,64	0,41
11	15	−3,36	11,29
12	17	−5,36	28,73
13	11	0,64	0,41
14	12	−0,36	0,13
	$\bar{x} = 11{,}64$		$\sum 95{,}22$

Řešení. Nulovou hypotézu H_0 stanovíme tak, aby odpovídala náhodnému hádání studentů, tj. průměrné hodnotě správných odpovědí $\mu_0 = 10$. Alternativní hypotéza H_A pak předpokládá, že studenti jsou na studium připraveni, tedy že $\mu > 10$. Jedná se tedy o jednostranný (a to pravostranný) t-test hypotézy $H_0\colon \mu = 10$ proti alternativě $H_A\colon \mu > 10$. Testujeme-li na obvyklé 1%, resp. 5% hladině významnosti, kritickou hodnotou testu bude 99%, resp. 95% kvantil t-rozdělení.

Vypočítáme výběrový průměr $\bar{x}$ pro náš čtrnáctiprvkový vzorek:

$$\bar{x} = \frac{\sum_1^{14} x_i}{14} \approx 11{,}64.$$

Pro výběrovou směrodatnou odchylku platí:

$$s = \sqrt{\frac{\sum_1^{14} (\bar{x} - x_i)^2}{14 - 1}} \approx \sqrt{\frac{95{,}22}{13}} \approx 2{,}71.$$

Vypočítáme hodnotu statistiky t pro náš vzorek:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{11{,}64 - 10}{\frac{2{,}71}{\sqrt{14}}} \approx 2{,}28.$$

Nyní porovnáme naši vypočtenou hodnotu s tabulkovými kritickými hodnotami pro Studentovo t-rozdělení s 13 stupni volnosti. Pro naši vypočtenou hodnotu platí: $2{,}28 < 2{,}650 = t_{0{,}99;13}$, $2{,}28 > 1{,}771 = t_{0{,}95;13}$. Hypotézu H_0 tedy nezamítáme na hladině významnosti $\alpha = 0{,}01$, ale zamítáme na hladině významnosti $\alpha = 0{,}05$.

W. S. Gosset alias Student

William Sealy Gosset se narodil v roce 1876 v anglickém Canterbury. Vystudoval chemii a matematiku ve Winchesteru a na New College v Oxfordu. Po ukončení studií v roce 1899 začal pracovat jako chemik a experimentální statistik v pivovaru Guinness v irském Dublinu. Během svého působení v pivovaru řešil Gosset zejména problém statistického testování kvality piva pomocí malých vzorků. Rozvoj této novátorské metody vyžadoval řadu experimentů a poměrně náročných matematických výpočtů. Proto začal Gosset od roku 1905 spolupracovat s uznávaným statistikem Karlem Pearsonem (1857–1936). Pearson v té době pracoval v Biometrics Laboratory, University College London a zabýval se, tak jako většina tehdejších statistiků, statistickou analýzou využívající rozsáhlých vzorků při známém pravděpodobnostním rozdělení základního souboru. Karl Pearson je mimo jiné autorem dnes hojně používaného χ^2 testu dobré shody. V letech 1906–1907 byl Gosset vyslán na speciální studium do Pearsonovy biometrické laboratoře, kde pracoval na řešení problému statistické inference z malých vzorků. Během roční práce v laboratoři publikoval Gosset několik článků, z nichž nejvýznamnější je článek *The probable error of a mean* z roku 1908. V článku je popsána nová metoda vhodná pro statistické testování málo početných vzorků, kterou dnes známe pod názvem Studentův t-test. Článek byl publikován v Pearsonově časopisu *Biometrika* pod pseudonymem Student. Důvodem utajení skutečné identity autora a jeho vazby na Guinessův pivovar byla pravděpodobně obava jeho zaměstnavatele z využití publikovaných výsledků konkurencí. Článek zaujal zejména statistika Ronalda A. Fishera, který tuto metodu dále rozvinul. Fisher byl v roce 1912 geniálním studentem matematiky na University of Cambridge. Objevil chybu v Gossetových výpočtech, když dokázal, že místo čísla n, které udává velikost vzorku, má být ve vzorci pro statistiku t číslo $n - 1$, kterému dnes říkáme počet stupňů volnosti. Uvádí se, že si Fisher s Gossetem v letech 1912–1934 vyměnili na 150 dopisů, v nichž řešili hlavně problematiku t-testu. Postupně mezi nimi vzniklo přátelství, i když osobně se setkali až v roce 1922 (Ziliak, 2008). Fisher znovu uvedl významné Gossetovy výsledky ve své práci *Statistical Methods for Research Workers* z roku 1934 a postupně se zasloužil o široké uplatnění Studentova t-testu v biometrice, ekonomii, medicíně a dalších vědních oborech. V dublinském pivovaru pracoval Gosset až do roku 1935, kdy se stal hlavním sládkem v nově otevřeném Guinnessově pivovaru Park Royal Brewery v Londýně. Bohužel, již v roce 1937 náhle zemřel na infarkt. Poznamenejme, že Gosset rovněž prováděl experimenty v oblasti agronomie. Kromě toho byl mužem mnoha zálib. Mimo jiné byl vášnivým zahrádkářem, rybářem, lovcem, stavěl lodě a hrál golf (Boland, 1984).

W. S. Gosset experimentátor

Většina statistických analýz byla v Gossetově době založena především na „teorii rozsáhlých vzorků“ předpokládající normální rozdělení sledovaného znaku v populaci. Při velkém počtu experimentů, resp. při dostatečně rozsáhlém vzorku dokážeme určit relativně přesně skutečnou hodnotu průměrné hodnoty sledovaného znaku v populaci. Gossetův přístup zaměřený na málo početné vzorky byl v tehdejší době opravdu revoluční. Pokud je vzorek malý, vznikají dva zdroje nejistoty: 1) kvůli chybě průměru náhodného vzorku se průměr vzorku může významně odchylovat od skutečné hodnoty průměru populace a 2) malý rozsah vzorku neumožňuje určit pravděpodobnostní rozdělení výběrového průměru. Obvyklá metoda určení pravděpodobnosti, s jakou leží hodnota populačního průměru v dané vzdálenosti od výběrového průměru, je založena na předpokladu, že výběrový průměr má normální rozdělení se směrodatnou odchylkou $\frac{\sigma}{\sqrt{n}}$, kde σ je směrodatná odchylka základního souboru a n je rozsah vzorku (Student, 1908). Hlavním cílem Gossetovy práce bylo zjistit, jak reprezentativní mohou být malé vzorky (2 až 10 pozorování) v porovnání s velkými vzorky s tisíci pozorováními. Přesněji řečeno, chtěl zjistit, jak závisí chyba odhadu skutečného průměru základního souboru na velikosti vzorku (počtu náhodných odběrů z jednotlivých barelů). V pivovarském provozu to konkrétně znamenalo řešit problém, kolik náhodných odběrů z jednotlivých barelů sladového extraktu je nutno provést, abychom si byli „téměř jisti“, že cukernatost extraktu se liší maximálně o 0,5 stupňů od požadované hodnoty 133 stupňů na barel. Gosset problém řešil v roce 1904 simulací dat. Použil sladový extrakt s relativně přesně známou hodnotou cukernatosti. Z tohoto základního souboru nejprve realizoval velké množství dvouprvkových pozorování (náhodných odběrů ze dvou barelů). Zjistil, že v 80 % případů se zjištěná cukernatost vzorku shoduje s přesností 0,5 stupně se skutečnou cukernatostí základního souboru. Neboli při náhodném výběru vzorků ze dvou barelů máme šanci 4 : 1, že cukernatost vzorku souhlasí s požadovanou přesností se skutečnou hodnotou. Gosset provedl podobný experiment pro $n = 3, 4, 5, \ldots, 82$. Výsledky jeho experimentů jsou uvedeny v tabulce 2 (Ziliak, 2008).

Gosset považoval za přijatelnou šanci pro získání relativně přesné hodnoty průměrné cukernatosti 10 : 1. Vzhledem k údajům uvedeným v tabulce 1 tedy Gosset učinil závěr, že je třeba provádět alespoň čtyřprvková pozorování.

Tab. 2: Šance chyby odhadu průměrné cukernatosti menší než 0,5 stupňů (zdroj: Ziliak, 2008)

Počet pozorování	Šance	Pravděpodobnost
2	4 : 1	80 %
3	7 : 1	87,5 %
4	12 : 1	92 %
5	19 : 1	95 %
82	Prakticky nekonečno	Téměř 100 %

Své matematické výpočty, zejména chování náhodné veličiny $z = \frac{\bar{x}-\mu}{s}$, se Gosset snažil již v letech 1906–1907 ověřit i empiricky pomocí simulace dat, postupem, který bychom dnes nazvali metodou Monte Carlo. Využil k tomu korelační tabulku obsahující údaje o výšce a délce levého prostředníčku 3 000 vězňů. Naměřené hodnoty byly zapsány na 3 000 kartiček, které byly řádně zamíchány. Následně byly postupně všechny kartičky náhodně vybírány a příslušné údaje zapisovány do knihy. Tím vznikl náhodně uspořádaný základní soubor dat o rozsahu $N = 3\,000$. Z tohoto souboru bylo náhodně vybráno 750 čtyřprvkových vzorků. Pro každý vzorek byl určen výběrový průměr $\bar{x}$ a výběrová směrodatná odchylka s. Nyní mohl Gosset pro každý vzorek určit rozdíl $\bar{x}-\mu$ jeho výběrového průměru $\bar{x}$ a průměru základního souboru μ. Dělením takto získaných hodnot hodnotami s příslušných výběrových směrodatných odchylek získal přehled o hodnotách náhodné veličiny $z = \frac{\bar{x}-\mu}{s}$ (Ziliak, 2011). Gosset pochopil, že by bylo výhodnější místo opakování mnoha pozorování pracovat s jedním vzorkem a jeho hodnotu porovnávat s matematickým modelem, který by udával pravděpodobnost, s kterou můžeme tento výsledek obdržet. To ho postupně dovedlo až k vytvoření náhodné veličiny z a sestavení tabulky jejího pravdivostního rozdělení.

W. S. Gosset statistik

Gosset se však nespokojil s výsledky svých experimentů a snažil se je potvrdit přesnými matematickými výpočty. Proto začal spolupracovat s uznávaným statistikem Karlem Pearsonem. V letech 1906–1907 pracoval v jeho biometrické laboratoři, kde se snažil zejména určit výběrové rozdělení statistiky $z = \frac{\bar{x}-\mu}{s}$ a zpracovat pravděpodobnostní tabulku tohoto rozdělení pro hodnoty $n = 2$ až $n = 10$. Později ve spolupráci s R. A. Fisherem ukázal, že výběrové rozdělení aritmetického průměru může být po standardizaci střední hodnotou μ a výběrovou střední chybou $s_{\bar{x}}$ reprezentováno tzv. t-statistikou tvaru

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}},$$

která má tzv. Studentovo t-rozdělení (Hendl, 2004). Tvar tohoto pravděpodobnostního rozdělení závisí pouze na jediném parametru, a sice na tzv. počtu stupňů volnosti. Počet stupňů volnosti je roven $n - 1$, závisí tedy na rozsahu vzorku. S rostoucím rozsahem vzorku roste počet stupňů volnosti a křivka hustoty pravděpodobnosti t-rozdělení se aproximativně blíží křivce standardizovaného normálního rozdělení.

Gosset definuje ve svém článku z roku 1908 náhodnou veličinu z ve tvaru $z = \frac{\bar{x}-\mu}{s}$, kde $\bar{x}$ je výběrový průměr, μ je skutečný průměr základního souboru (o kterém se předpokládá, že má normální rozdělení $N(\mu, \sigma^2)$ s neznámým σ) a

$$s^2 = \frac{\sum_i^n (x_i - \bar{x})^2}{n}$$

je výběrový rozptyl. Statistiku z je tedy možné vyjádřit ve tvaru

$$z = \frac{\bar{x} - \mu}{\sqrt{\frac{\sum_i^n (x_i - \bar{x})^2}{n}}} = \frac{\bar{x} - \mu}{\sqrt{\sum_i^n (x_i - \bar{x})^2}} \sqrt{n}.$$

Článek obsahuje rovněž pravděpodobnostní tabulku z rozdělení pro n od 4 do 10 (Tab. 3). V roce 1917 Gosset zveřejnil rozšířenou tabulku pro n od 2 do 30. Později, v roce 1925, R. A. Fisher transformoval náhodnou veličinu z na $t = z\sqrt{n-1}$ (Eisenhart, 1979).

Tab. 3: Pravděpodobnostní tabulka z rozdělení pro n od 4 do 10
(Zdroj: Student, 1908)

$z\left(=\frac{x}{s}\right)$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$	For comparison $\left(\frac{\sqrt{7}}{\sqrt{(2\pi)}}\int_{-\infty}^{x} e^{-\frac{7x^2}{2}}\,dx\right)$
0.1	0.5633	0.5745	0.5841	0.5928	0.6006	0.60787	0.61462	0.60411
0.2	0.6241	0.6458	0.6634	0.6798	0.6936	0.70705	0.71846	0.70159
0.3	0.6804	0.7096	0.7340	0.7549	0.7733	0.78961	0.80423	0.78641
0.4	0.7309	0.7657	0.7939	0.8175	0.8376	0.85465	0.86970	0.85520
0.5	0.7749	0.8131	0.8428	0.8667	0.8863	0.90251	0.91609	0.90691
0.6	0.8125	0.8518	0.8813	0.9040	0.9218	0.93600	0.94732	0.94375
0.7	0.8440	0.8830	0.9109	0.9314	0.9468	0.95851	0.96747	0.96799
0.8	0.8701	0.9076	0.9332	0.9512	0.9640	0.97328	0.98007	0.98253
0.9	0.8915	0.9269	0.9498	0.9652	0.9756	0.98279	0.98780	0.99137
1.0	0.9092	0.9419	0.9622	0.9751	0.9834	0.98890	0.99252	0.99820
1.1	0.9236	0.9537	0.9714	0.9821	0.9887	0.99280	0.99539	0.99926
1.2	0.9354	0.9628	0.9782	0.9870	0.9922	0.99528	0.99713	0.99971
1.3	0.9451	0.9700	0.9832	0.9905	0.9946	0.99688	0.99819	0.99986
1.4	0.9451	0.9756	0.9870	0.9930	0.9962	0.99791	0.99885	0.99989
1.5	0.9598	0.9800	0.9899	0.9948	0.9973	0.99859	0.99926	0.99999
1.6	0.9653	0.9836	0.9920	0.9961	0.9981	0.99903	0.99951	
1.7	0.9699	0.9864	0.9937	0.9970	0.9986	0.99933	0.99968	
1.8	0.9737	0.9886	0.9950	0.9977	0.9990	0.99953	0.99978	
1.9	0.9970	0.9904	0.9959	0.9983	0.9992	0.99967	0.99985	
2.0	0.9797	0.9919	0.9967	0.9986	0.9994	0.99976	0.99990	
2.1	0.9821	0.9931	0.9973	0.9989	0.9996	0.99983	0.99993	
2.2	0.9841	0.9941	0.9978	0.9992	0.9997	0.99987	0.99995	
2.3	0.9858	0.9950	0.9982	0.9993	0.9998	0.99991	0.99996	
2.4	0.9873	0.9957	0.9985	0.9995	0.9998	0.99993	0.99997	
2.5	0.9886	0.9963	0.9987	0.9996	0.9998	0.99995	0.99998	
2.6	0.9898	0.9967	0.9989	0.9996	0.9999	0.99996	0.99999	
2.7	0.9908	0.9972	0.9989	0.9997	0.9999	0.99997	0.99999	
2.8	0.9916	0.9975	0.9989	0.9998	0.9999	0.99998	0.99999	
2.9	0.9924	0.9978	0.9989	0.9998	0.9999	0.99998	0.99999	
3.0	0.9931	0.9981	0.9989	0.9998	—	0.99999	—	

Ve svém nejvýznamnějším článku Gosset uvádí následující příklad ilustrující jeho novou metodu. S její pomocí ukazuje rozdílný efekt dvou různých medikamentů na spánek deseti pacientů. Tabulka 4 zachycuje průměrný počet hodin spánku vyvolaných u těchto pacientů pomocí obou preparátů.

Z tabulky 4 je zřejmé, že pro preparát 1 platí $z = \frac{0{,}75}{1{,}70} = 0{,}44$. Podle tabulky 3 je pravděpodobnost této hodnoty mezi 0,8697 a 0,9161. Interpolací lze zjistit, že tato pravděpodobnost je 0,8873. To odpovídá šanci $\frac{0{,}887}{0{,}113}$, tj. přibližně 8 : 1. V případě preparátu 2 máme $z = \frac{2{,}33}{1{,}90} = 1{,}23$. To odpovídá pravděpodobnosti 0,9974, tedy máme šanci téměř 400 : 1 pro tuto hodnotu. Abychom ukázali, že účinek preparátu 2 je významně lepší než u preparátu 1, vypočítáme hodnotu $z = \frac{1{,}58}{1{,}17} = 1{,}35$ pro rozdíly mezi oběma medikamenty (poslední sloupec tabulky 4). Tato hodnota odpovídá tabulkové pravděpodobnosti 0,9985. Šance, že preparát 2 je lepší než preparát 1, je tedy přibližně 666 : 1.

Tab. 4: První publikovaný příklad na užití t-testu
(Student, 1908)

Additional hours' sleep gained by the use of hyoscyamine hydrobromide.

Patient	1 (Dextro-)	2 (Laevo-)	Difference (2–1)
1.	+ ·7	+ 1·9	+ 1·2
2.	– 1·6	+ ·8	+ 2·4
3.	– ·2	+ 1·1	+ 1·3
4.	– 1·2	+ ·1	+ 1·3
5.	– 1	– ·1	0
6.	+ 3·4	+ 4·4	+ 1·0
7.	+ 3·7	+ 5·5	+ 1·8
8.	+ ·8	+ 1·6	+ ·8
9.	0	+ 4·6	+ 4·6
10.	+ 2·0	+ 3·4	+ 1·4
	Mean + ·75	Mean + 2·33	Mean + 1·58
	S. D. 1·70	S. D. 1·90	S. D. 1·17

Z hlediska dnešní statistiky odpovídá tento příklad použití tzv. párového t-testu. Testové kritérium má v tomto případě tvar

$$t = \frac{\bar{D}}{s_D}\sqrt{n},$$

kde $\bar{D}$ je výběrový průměr rozdílů jednotlivých párů, s_D je výběrová směrodatná odchylka rozdílů, n je počet párů (Škaloudová, 1998). Při platnosti nulové hypotézy má tato statistika t-rozdělení s $n-1$ stupni volnosti. Nulovou hypotézou zde rozumíme tvrzení, že mezi účinky obou preparátů není žádný statisticky významný rozdíl. Proti ní stojí alternativní hypotéza o statisticky významném rozdílu ve prospěch medikamentu 2. Pro náš příklad tedy máme $t = \frac{1{,}58}{1{,}17}\sqrt{10} \approx 4{,}269$. Nahlédneme-li do tabulky kvantilů t-rozdělení, zjistíme, že námi vypočtená hodnota je větší, než 95%

kvantil pro 9 stupňů volnosti $t_{0,95;9} = 1{,}833$ i 99% kvantil pro 9 stupňů volnosti $t_{0,99;9} = 2{,}821$. Nulovou hypotézu tedy zamítáme na hladině významnosti 5 % i 1 %. Udělali bychom tedy stejný závěr jako Gosset, tj. preparát 2 je účinnější než preparát 1. (Přesněji: můžeme s 1% rizikem omylu tvrdit, že preparát 2 je účinnější než preparát 1.)

Závěr

Výsledky Gossetovy práce a zejména jeho nejvýznamnější článek z roku 1908 představují pozoruhodné obohacení tehdejší statistiky. Přináší novou problematiku statistické inference pomocí malých vzorků a ukazuje metodu simulace dat, která umožňuje určovat pravděpodobnostní rozdělení náhodných veličin (metoda Monte Carlo). Přesto, že hlavním Gossetovým objevem ve statistice je jeho t-rozdělení, neměli bychom zapomenout uvést i další výsledky, jimiž přispěl k rozvoji této vědní disciplíny. Gosset se zasloužil mimo jiné o rozvoj efektivního experimentálního designu a o ekonomický přístup k logice nejistoty (Ziliak, 2008). Gossetův význam spočívá rovněž v tom, že jako první zavedl statistickou kontrolu do výrobní praxe.

Literatura

[1] Arbuthnot, J. (1710). An argument for divine providence, taken from the constant regularity observed in the births of both sexes. *Philosophical Transactions of the Royal Society, 27*, 186–190. Dostupné z `http://www.jstor.org/stable/103111`.

[2] Boland, P. J. (1984). A biographical glimpse of William Sealy Gosset. *American Statistician, 38*, 179–183.

[3] Box, J. F. (1981). Gosset, Fisher, and t distribution. *The American Statistician, 35*(2), 61–66.

[4] Box, J. F. (1987). Guinness, Gosset, Fisher, and small samples. *Statistical Science, 2*(1), 45–52.

[5] Eisenhart, C. (1979). On the transition from “Student’s” z to “Student’s” t. *American Statistician, 33*, 6–10.

[6] Emanovský, P. (2021). První statistické testování hypotézy podle Johna Arbuthnota. *Učitel matematiky, 29*(1), 26–36.

[7] Fienberg, S. E. (1992). A brief history of statistics in three and one-half chapters: a review essay. *Statistical Science, 7*(2), 208–225.

[8] Fisher, R. A. (1934). *Statistical methods for research workers.* Paternoster Row.

[9] Hendl, J. (2004). *Přehled statistických metod – zpracování dat.* Portál.

[10] Chráska, M. (2007). *Metody pedagogického výzkumu.* Portál.

[11] Klementa, J, Komenda, S., & Kunert, E. (1984). *Statistické metody v pedagogickém výzkumu.* VUP.

[12] Komenda, S. (1994). *Biometrie.* VUP.

[13] Plackett, R. I., & Barnard, G. A. (1990). *Student – A statistical biography of William Sealy Gosset.* Clarendon Press.

[14] Student. (1908). The probable error of a mean. *Biometrika, 6*, 1–25.

[15] Škaloudová, A. (1998). *Statistika v pedagogickém a psychologickém výzkumu.* Pedagogická fakulta UK.

[16] Ziliak, S. T. (2008). Guinnessometrics: The economic foundation of „Student’s t. *Journal of Economic Perspectives, 22*(4), 199–216.

[17] Ziliak, S. T. (2011). W.S. Gosset and some neglected concepts in experimental statistics: Guinnessometrics II. *Journal of Wine Economics, 6*(2), 252–277.

[18] Ziliak, S. T. (2014). Balanced versus randomized field experiments in economics: Why W. S. Gosset aka “Student” matters. *Review of Behavioral Economics, 1*, 167–208.

Abstract

One of the most used tests of statistical significance is the so called Student's t-test. The author of the mathematical model on which this test is based is the British statistician William Sealy Gosset (1876–1937). The article describes the essence and origin of this test in a mathematical and historical context.

Petr Emanovský
Přírodovědecká fakulta Univerzity Palackého v Olomouci
17. listopadu 1192/12
771 46 Olomouc
e-mail: petr.emanovsky@upol.cz