Petr Emanovský

Jak Frank Wilcoxon pomohl statistikům objevit neparametrické testy

Učitel matematiky, Vol. 31 (2023), No. 3, 164–177

Persistent URL: http://dml.cz/dmlcz/152010

JAK FRANK WILCOXON POMOHL STATISTIKŮM OBJEVIT NEPARAMETRICKÉ TESTY

Petr Emanovský[1]

Úvod

Jedním z významných úkolů moderní statistiky je poskytovat informace o rozsáhlém základním souboru na základě dat získaných na menším vzorku neboli provádět tzv. statistickou inferenci. Tato procedura je popsána v každé učebnici matematické statistiky (např. Hendl, 2004; Anděl, 2007) a představuje jeden ze základních principů této vědecké disciplíny. Vznik a historický vývoj tohoto myšlenkového postupu je stručně popsán například v (Arbuthnot, 1710; Fienberg, 1992; Emanovský, 2021a; Emanovský, 2021b). Princip statistické inference se významně uplatňuje při statistickém testování hypotéz neboli při testech statistické významnosti. Toto testování představuje jednu ze základních metod moderní statistiky, kterou dnes používají nejen profesionální statistici, ale i vědečtí pracovníci a studenti při svých výzkumech. 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 testové statistiky pro náhodně

[1]Článek vznikl za podpory projektu Univerzity Palackého Olomouc „Algebraické a geometrické struktury“ IGA PrF 2022 017.

vybraný vzorek. Tuto vypočtenou hodnotu srovnáváme s tabulkovou (kritickou) hodnotou. Padne-li vypočtená hodnota do tzv. kritického oboru, zamítáme nulovou hypotézu na zvolené hladině významnosti. V tomto případě je totiž pravděpodobnost toho, 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 spíše než tabulky 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).

Parametrické versus neparametrické testy

Testy statistické významnosti zpravidla dělíme do dvou skupin. Rozlišujeme tzv. testy parametrické a testy neparametrické. Definice obou typů testu nejsou ve statistické literatuře zcela jednoznačné, ale můžeme se pokusit o jejich stručnou charakteristiku. Parametrickými statistickými metodami většinou rozumíme takové postupy, jejichž použití předpokládá některé vlastnosti měřených proměnných. Častým předpokladem bývá normalita měřených dat. K nejznámějším parametrickým testům patří Studentův t-test (jednovýběrový, dvouvýběrový) a F-test (Anděl, 2007; Hendl, 2004).

V praxi se však často setkáváme se situací, kdy o rozdělení pravděpodobnosti zkoumané veličiny nic nevíme, nebo víme, že má tvar, který není vhodný pro použití některého z běžných parametrických testů. V tomto případě lze použít postupů, které nevyžadují znalost nebo některou specifickou vlastnost rozdělení zkoumané veličiny. Těmto statistickým metodám říkáme neparametrické metody neboli metody nezávislé na tvaru rozdělení (distribution free). Jejich výhodou je univerzálnější použití, což je na druhé straně vykoupeno menší silou těchto testů. Tedy pravděpodobnost, že zamítneme neplatnou nulovou hypotézu, je u neparametrických testů menší než u testů parametrických. Je-li však nulová hypotéza zamítnuta pomocí neparametrického testu, je již zbytečné ji testovat silnějším parametrickým testem (Blatná, 1996; Papica, 1992; Škaloudová, 1998). Nejběžnější neparametrické testy jsou znaménkový test, Wilcoxonovy testy (jednovýběrový, párový, dvouvýběrový) a χ^2-test (Anděl, 2007; Hendl, 2004). Všimněme si blíže Wilcoxonova testu pro dva závislé výběry (párového testu).

Wilcoxonův test pro dva závislé výběry

Wilcoxonův test pro dva závislé výběry, nazývaný též párový test, pracuje s hodnotami rozdílů $d_i = x_i - y_i$ mezi pozorovanými párovými hodnotami x_i, y_i. Na rozdíl od jednoduššího znaménkového testu však nebere v úvahu pouze směr odchylek uvnitř párů, ale rovněž relativní velikost těchto odchylek. Díky tomu má Wilcoxonův test větší sílu zamítnout nulovou hypotézu než znaménkový test (Emanovský, 2022a).

Postup při Wilcoxonově párovém testu je následující:

1. Určíme rozdíly $d_i = x_i - y_i$ pro všechny páry výběru. Případné nulové rozdíly ignorujeme.
2. Nenulové rozdíly uspořádáme vzestupně podle velikosti bez ohledu na znaménko.
3. Každému nenulovému rozdílu přiřadíme pořadí, přičemž stejným hodnotám přiřadíme průměrné pořadí.
4. Určíme součet pořadí kladných rozdílů W^+ a součet pořadí záporných rozdílů W^-. Snadno lze ukázat, že $W^+ + W^- = 1 + 2 + \cdots + n = \frac{n(n+1)}{2}$.
5. Nulovou hypotézou je zpravidla tvrzení, že mezi párovými hodnotami x_i, y_i není statisticky významný rozdíl, a alternativní hypotéza naopak předpokládá, že tento rozdíl statisticky významný je. V tomto případě se jedná o dvoustranný test a jako testovací kritérium použijeme statistiku

$W = \min(W^+, W^-)$. Můžeme samozřejmě uvažovat i jednostranný test, a sice levostranný s alternativní hypotézou předpokládající, že hodnoty x_i jsou statisticky významně menší než hodnoty y_i nebo pravostranný s alternativní hypotézou předpokládající, že hodnoty x_i jsou statisticky významně větší než hodnoty y_i. V prvním případě bereme za testové kritérium W^+, ve druhém W^-.

6. Pro malé hodnoty n jsou kritické hodnoty $w(\alpha, n)$ tabelovány (např. Hendl, 2004; Anděl, 2007). Je-li vypočtená hodnota W menší nebo rovna tabulkové kritické hodnotě $w(\alpha, n)$, zamítáme nulovou hypotézu na hladině významnosti α.

Právě popsaný test je někdy nazýván přesným testem. Pro větší počet párových pozorování ($n \geq 30$) lze provést tzv. asymptotické testování, které využívá toho, že rozdělení statistiky W se asymptoticky blíží normálnímu rozdělení se střední hodnotou $E(W) = \frac{n(n+1)}{4}$ a rozptylem $D(W) = \frac{n(n+1)(2n+1)}{24}$. Jako testové kritérium použijeme testovou statistiku $U = \frac{W - E(W)}{\sqrt{D(W)}}$. Kritické hodnoty pak najdeme v tabulce standardizovaného normálního rozdělení (Hendl, 2004; Blatná, 1996; Papica, 1992).

Příklad. Deset náhodně vybraných žáků 8. ročníku školy vykazovalo výsledky vědomostních testů z matematiky a fyziky uvedené v tabulce 1. Oba testy byly přibližně stejně náročné a maximální počet bodů v každém testu byl 100. Zajímá nás, jestli je rozdíl ve výsledcích obou testů pro žáky 8. ročníku této školy statisticky významný.

Výsledky obou testů představují dva závislé spárované datové soubory, které můžeme rovněž chápat jako jeden soubor obsahující párové rozdíly. Lze tedy použít Wilcoxonův párový test, který bude probíhat v následujících krocích:

1. Určíme rozdíly $d_i = x_i - y_i$ (4. sloupec tabulky 1).
2. Nenulové rozdíly uspořádáme vzestupně podle velikosti bez ohledu na znaménko.
3. Každému nenulovému rozdílu přiřadíme pořadí, přičemž stejným hodnotám přiřadíme průměrné pořadí. Pořadí klad-

Tab. 1: Výsledky testů z matematiky a fyziky

i (žák)	x_i (Test M)	y_i (Test F)	$d_i = x_i - y_i$	Pořadí d_i^+	Pořadí d_i^-
1	82	63	19	9	
2	69	42	27	10	
3	73	74	-1		1,5
4	43	37	6	7	
5	58	51	7	8	
6	93	91	2	4	
7	96	94	2	4	
8	58	54	4	6	
9	76	77	-1		1,5
10	88	86	2	4	
				$W^+ = 52$	$W^- = 3$

ných rozdílů jsou v 5. sloupci tabulky 1 a pořadí záporných rozdílů jsou v 6. sloupci tabulky 1. První v pořadí jsou dvě stejné hodnoty 1 (znaménko neuvažujeme). Každé z těchto hodnot přiřadíme průměrné pořadí $\frac{1+2}{2} = 1{,}5$. Následují tři stejné hodnoty 2, které zaujímají 3., 4. a 5. pořadí. Těmto hodnotám přiřadíme průměrné pořadí $\frac{3+4+5}{2} = 4$. Následuje hodnota 4, která je 6. v pořadí atd.

4. Určíme součet pořadí kladných rozdílů $W^+ = 52$ a součet pořadí záporných rozdílů $W^- = 3$.
5. Nulovou hypotézou je zde tvrzení, že mezi výsledky obou testů není statisticky významný rozdíl, a alternativní hypotéza naopak předpokládá, že tento rozdíl statisticky významný je. Jedná se tedy o dvoustranný test. Hodnota testovacího kritéria je v tomto případě $W = \min(W^+, W^-) = 3$. Intuitivně cítíme, že velký rozdíl mezi hodnotami W^+ a W^- je nepříznivý nulové hypotéze.
6. Vzhledem k tomu, že je vypočtená hodnota menší než kritická tabulková hodnota pro Wilcoxonův test $w(0{,}05, 10) = {}$ $= 8$ (tab. 4), zamítáme nulovou hypotézu na hladině vý-

znamnosti 5 %. Vzhledem k tomu, že $w(0{,}01, 10) = 3$, zamítli bychom nulovou hypotézu i na hladině významnosti 1 %.

Z tabulky 1 vidíme, že počet kladných rozdílů je větší než počet záporných rozdílů. Nabízí se tedy alternativní hypotéza ve tvaru: „Bodové hodnocení žáků v testu z matematiky je statisticky významně vyšší než v testu z fyziky.“ V tomto případě se jedná o jednostranný (přesněji pravostranný) test, pro nějž je testovým kritériem $W^- = 3$. Tedy i v případě pravostranného testu nulovou hypotézu zamítáme na hladině významnosti 5 %. Pro levostranný test by ovšem testovým kritériem byla hodnota $W^+ = 52$, která je větší než kritická, takže bychom nulovou hypotézu nezamítli.

Testování lze provést i pomocí znaménkového testu, který však nemá dostatečnou sílu na to, abychom mohli zamítnout nulovou hypotézu na hladině významnosti 0,05 (Emanovský, 2021). Wilcoxonův párový test je silnější než znaménkový test, neboť kromě počtu kladných (záporných) rozdílů zohledňuje také velikosti těchto rozdílů.

Frank Wilcoxon

Frank Wilcoxon a jeho sestra-dvojče se narodili 2. září 1892 zámožným americkým rodičům v irském County Cork. Část svého vzdělání získal v Anglii od soukromých učitelů, ale převážně vyrůstal v New Yorku. Prožil značně bouřlivé mládí, neboť toužil poznávat stále nové věci. V 16 letech se dal k námořnictvu, později pracoval na ropném vrtu a také jako lesní ošetřovatel stromů. V roce 1917 ukončil přírodovědné bakalářské studium na Pennsylvania Military College a po první světové válce získal v roce 1921 magisterský titul v oboru chemie na Rutgers University. Ve studiu dále pokračoval na Cornell University, kde v roce 1924 obhájil doktorát z anorganické chemie. V následujících letech se Frank Wilcoxon intenzivně zabýval vědeckým výzkumem v oblasti chemie. Předmětem jeho vědecké práce bylo zejména zkoumání vlivu některých chemických látek na růst rostlin. Významně se podílel na vývoji insekticidů a fungicidů. V letech 1925–1957 postupně působil v několika výzkumných ústavech v USA zabývajících se aplikacemi chemických látek v botanice (Boyce Thomson Institute for Plant Research, Atlas Powder Company, American Cyanamid). Během této práce se postupně začal intenzivně zabývat metodami statistické analýzy empirických dat. V roce 1957 ve věku 65 let odešel do důchodu, dále však působil jako konzultant pro Boyce Thomson Institute a přednášel na Katedře statistiky Floridské státní university. Zemřel po krátké nemoci v roce 1965. (Borja & Stander, 2016)

Frank Wilcoxon statistik

Frank Wilcoxon se proslavil především svou průkopnickou prací ve statistice, přestože byl v podstatě statistik samouk. Známý je zejména jako autor Wilcoxonových testů (jednovýběrového a dvouvýběrového), které představují neparametrické alternativy příslušných Studentových t-testů. Wilcoxonův zájem o statistiku lze zaznamenat již v roce 1925, kdy se snažil statisticky testovat efektivitu účinku nově vyvinutých insekticidů (Borja & Stander, 2016). Později byl silně ovlivněn Fisherovým významným článkem *Statistical Methods for Research Workers* z roku 1934 (Borja & Stander, 2016). Přestože oficiálně pracoval jako statistik pouze posledních sedm let svého produktivního života, publikoval řadu originálních prací z oblasti statistiky. K jeho nejvýznamnějším publikacím patří bezesporu článek *Individual Comparisons by Ranking Methods*, který vyšel v časopisu *Biometrics Bulletin* v roce 1945. V tomto článku Wilcoxon popisuje dvě nové metody neparametrického testování dnes známé jako dvouvýběrový Wilcoxonův test (Wilcoxon rank sum test) a párový Wilcoxonův test (Wilcoxon signed-rank test) (Siegel, 1957). Hlavní myšlenka Wilcoxonova neparametrického testování spočívá v tom, že se netestují přímo pozorovaná data, ale jejich pořadí v uspořádaném souboru dat.

Wilcoxonovy výpočty

Wilcoxon se snažil určit pravděpodobnost toho, že hodnota testového kritéria W bude rovna nebo ještě extrémnější než pozorovaná hodnota. Zkusme naznačit alespoň základní úvahy směřující k tomuto cíli. Předpokládejme, že $W = W^+$ a pro jednoduchost uvažujme všechna pořadí rozdílů navzájem různá. Označme $F_W(w/n)$ počet všech možností, kterými lze vybrat z množiny $\{1, 2, \ldots, n\}$ takovou podmnožinu, že součet všech čísel této podmnožiny je právě w. Je zřejmé, že $w \in \left\{0, 1, 2, \ldots, \frac{n(n+1)}{2}\right\}$.

Například pro 3 pozorované páry ($n = 3$) máme celkem $2^3 =$ $= 8$ možných pořadí rozdílů (uvažujeme pořadí $(\pm1, \pm2, \pm3)$). Potom $F_W(6/3) = 1$, neboť máme právě jedno pořadí $(+1, +2, +3)$, pro něž $w = 6$. Podobně $F_W(5/3) = 1$ (pořadí $(-1, +2, +3)$), $F_W(4/3) = 1$ (pořadí $(+1, -2, +3)$), $F_W(3/3) = 2$ (pořadí $(+1, +2, -3)$ a $(-1, -2, +3)$), $F_W(2/3) = 1$ (pořadí $(-1, +2, -3)$), $F_W(1/3) = 1$ (pořadí $(+1, -2, -3)$) a $F_W(0/3) = 1$ (pořadí $(-1, -2, -3)$) (tab. 2).

Tab. 2: Hodnoty $F_W(w/n)$ pro $n = 3$

Součet w všech kladných členů pořadí	Všechna pořadí se součtem kladných členů w	Počet $F_W(w/3)$ všech pořadí se součtem kladných členů w
6	$(+1, +2, +3)$	1
5	$(-1, +2, +3)$	1
4	$(+1, -2, +3)$	1
3	$(+1, +2, -3)$, $(-1, -2, +3)$	2
2	$(-1, +2, -3)$	1
1	$(+1, -2, -3)$	1
0	$(-1, -2, -3)$	1

Statistika W^+ má tedy na množině hodnot $\{0, 1, 2, \ldots, 6\}$ (tzv. *kritických hodnot*) rozdělení pravděpodobností $\left\{\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{2}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right\}$. Pokud by nás např. zajímala pravděpodobnost $P(W^+ \leq 4)$ toho, že pozorovaná hodnota statistiky W^+ bude menší nebo rovna 4 (tzv. *pravděpodobnostní hladina* pro $w = 4$), vypočteme $P(W^+ \leq 4) = \frac{1}{8} + \frac{2}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{6}{8} = 0{,}75$ (Carine et al., 2010).

Každé kritické hodnotě w lze tedy přiřadit pravděpodobnostní hladinu $P(W^+ \leq w)$.

Tabulka 3 udává kritické hodnoty a pravděpodobnostní hladiny pro $n = 3$. Pro větší hodnoty n lze při výpočtu hodnot $F_W(w/n)$ využít rekurentního vzorce

$$F_W(w/n) = F_W\big(w/(n-1)\big) + F_W\big((w-n)/(n-1)\big)$$

(McCornack, 1965).

Tab. 3: Kritické hodnoty a pravděpodobnostní hladiny pro $n = 3$

Pořadí	Součet kladných pořadí w (kritické hodnoty)	Pravděpodobnostní hladiny $(P(W^+ \leq w))$
$(-1, -2, -3)$	0	$\frac{1}{8} = 0{,}125$
$(+1, -2, -3)$	1	$\frac{1}{8} + \frac{1}{8} = \frac{2}{8} = 0{,}25$
$(-1, +2, -3)$	2	$\frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8} = 0{,}375$
$(+1, +2, -3)$ $(-1, -2, +3)$	3	$\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{2}{8} = \frac{5}{8} = 0{,}625$
$(+1, -2, +3)$	4	$\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{2}{8} + \frac{1}{8} = \frac{6}{8} = 0{,}75$
$(-1, +2, +3)$	5	$\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{2}{8} + \frac{1}{8} + \frac{1}{8} = \frac{7}{8} = 0{,}875$
$(+1, +2, +3)$	6	$\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{2}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{8}{8} = 1$

Pro výpočet pravděpodobnostních hladin pak platí obecný vzorec

$$P(W^+ \leq w) = \frac{\sum_{k=0}^{n} F_W(k/n)}{2^n}.$$

Pomocí tohoto vzorce můžeme pro dané n vypočítat pravděpodobnostní hladiny příslušné jednotlivým kritickým hodnotám. Pokud bychom uvažovali i případy, kdy mohou být některá pořadí stejná, výpočet by byl poněkud složitější (Lehman, 1998). Podobným způsobem postupoval i Wilcoxon při sestavování tabulky publikované v jeho článku (Wilcoxon, 1945) (tab. 4). První sloupec tabulky udává počet párů pozorovaných dat (tedy n), ve druhém sloupci jsou některé vybrané kritické hodnoty a ve třetím sloupci jim příslušející pravděpodobnostní hladiny. Wilcoxona zřejmě zajímaly zejména kritické hodnoty odpovídající konvenční hladině významnosti 0,05, případně 0,01. Tabulka 4 například

Tab. 4: Wilcoxonova tabulka kritických hodnot a pravděpodobnostních hladin (Wilcoxon, 1945, s. 82)

TABLE II
For Determining the Significance of Differences in Paired Experiments

Number of Paired Comparisons	Sum of rank numbers, + or —, whichever is less	Probability of this total or less
7	0	0.016
7	2	0.047
8	0	0.0078
8	2	0.024
8	4	0.055
9	2	0.0092
9	3	0.019
9	6	0.054
10	3	0.0098
10	5	0.019
10	8	0.049
11	5	0.0093
11	7	0.018
11	11	0.053
12	7	0.0093
12	10	0.021
12	14	0.054
13	10	0.0105
13	13	0.021
13	17	0.050
14	13	0.0107
14	16	0.021
14	21	0.054
15	16	0.0103
15	19	0.019
15	25	0.054
16	19	0.0094
16	23	0.020
16	29	0.053

uvádí důležité kritické hodnoty pro $n = 10$ (kritická hodnota 8 s pravděpodobnostní hladinou 0,049 ≈ 0,05 a kritická hodnota 3 s pravděpodobnostní hladinou 0,0098 ≈ 0,01). Na základě těchto úvah Wilcoxon (1945) a Wilcoxon et al. (1963) sestavili tabulky kritických hodnot a pravděpodobnostních úrovní do 50 párů pozorování. Rozšířené tabulky vytvořil McCornack (1965). Je třeba si uvědomit, že tabulky vznikaly v době, kdy sice již existovaly počítače, ale nebyly zdaleka tak výkonné jako ty dnešní. S dnešní podobou tabulky kritických hodnot Wilcoxonova párového testu, kterou používáme při rozhodování o zamítnutí či nezamítnutí nulové hypotézy na příslušné hladině významnosti, se můžeme seznámit téměř v každé učebnici matematické statistiky (tab. 5). Samozřejmě, že v dnešní době výkonných počítačů a sofistikovaného softwaru se obejdeme i bez tabulek. Vzhledem k tomu, že pravděpodobnostní hladiny klesají s rostoucím n, dostaneme nenulové

Tab. 5: Dnešní podoba tabulky kritických hodnot Wilcoxonova párového testu (Anděl, 2007, s. 260)

T4. Kritické hodnoty $w_n(\alpha)$ jednovýběrového Wilcoxonova testu

$P[\min(S^+, S^-) \leq w_n(\alpha)] \leq \alpha$

n	$w_n(0{,}05)$	$w_n(0{,}01)$	n	$w_n(0{,}05)$	$w_n(0{,}01)$	n	$w_n(0{,}05)$	$w_n(0{,}01)$
6	0	–	26	98	75	46	361	307
7	2	–	27	107	83	47	378	322
8	3	0	28	116	91	48	396	339
9	5	1	29	126	100	49	415	355
10	8	3	30	137	109	50	434	373
11	10	5	31	147	118	51	453	390
12	13	7	32	159	128	52	473	408
13	17	9	33	170	138	53	494	427
14	21	12	34	182	148	54	514	445
15	25	15	35	195	159	55	536	465
16	29	19	36	208	171	56	557	484
17	34	23	37	221	182	57	579	504
18	40	27	38	235	194	58	602	525
19	46	32	39	249	207	59	625	546
20	52	37	40	264	220	60	648	567
21	58	42	41	279	233	61	672	589
22	65	48	42	294	247	62	697	611
23	73	54	43	310	261	63	721	634
24	81	61	44	327	276	64	747	657
25	89	68	45	343	291	65	772	681

kritické hodnoty s pravděpodobnostními hladinami blízkými hodnotě 0,05 až při $n \geq 7$. V našem příkladu jsme použili kritickou hodnotu $w(0{,}05, 10) = 8$, kterou lze přečíst v tabulce 5. Znamená to, že při $n = 10$ je pravděpodobnost $P(W^+ \leq 8) \leq 0{,}05$, přičemž 8 je největší ze všech kritických hodnot, které tuto nerovnost splňují, neboli hodnota 8 je největší hodnota, která nás ještě opravňuje zamítnout nulovou hypotézu na hladině významnosti 0,05. Pro každou větší kritickou hodnotu je již příslušná pravděpodobnostní hladina (a tedy riziko chyby 1. druhu) větší než 0,05. To vede k zamítnutí nulové hypotézy na hladině významnosti 0,05. Podobně pro hladinu významnosti 0,01 máme kritickou hodnotu $w(0{,}01, 10) = 3$, což je největší hodnota ze všech kritických hodnot w, pro něž je pravděpodobnost $P(W^+ \leq w) \leq 0{,}01$.

Nakonec Wilcoxon vyšetřoval asymptotické vlastnosti rozdělení statistiky W^+ a ukázal, že se toto rozdělení blíží normálnímu se střední hodnotou $E(W) = \frac{n(n+1)}{4}$ a rozptylem $D(W) = \frac{n(n+1)(2n+1)}{24}$ (Wilcoxon, 1947).

Závěr

Výsledky Wilcoxonovy průkopnické práce ve statistice a zejména jeho nejvýznamnější článek z roku 1945 představují pozoruhodné obohacení statistických metod. Objev Wilcoxonových testů přináší novou široce použitelnou alternativu jednovýběrového a dvouvýběrového t-testu. O originalitě a využitelnosti tohoto konceptu ve statistice svědčí mimo jiné citovanost článku (Wilcoxon, 1945). Uvádí se, že od roku 1946 do roku 2015 byl počet citací tohoto článku 4 229 (Borja & Stander, 2016). Neparametrické metody se uplatňují zejména v pedagogickém, sociologickém a psychologickém výzkumu, kde se často setkáváme se statistickými znaky, na které nelze parametrické testy aplikovat.

Literatura

[1] Anděl, J. (2007). *Statistické metody.* Matfyzpress.

[2] 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. http://www.jstor.org/stable/103111.

[3] Blatná, D. (1996). *Neparametrické metody.* VŠE.

[4] Borja, M. C., & Stander, J. (2016). How Frank Wilcoxon helped statisticians walk the non-parametric path. *Significance, 13*(6).

[5] Bellera, C. A., Julien, M., & Hanley, J. A. (2010). Normal approximations to the distributions of the Wilcoxon Statistics: Accurate to what N? Graphical Insights. *Journal of Statistics Education, 18*(2), https://doi.org/10.1080/10691898.2010.11889486

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

[7] Emanovský, P. (2022a). Jak pivovarský sládek způsobil revoluci ve statistice. *Učitel matematiky, 29*(2), 96–110.

[8] Emanovský, P. (2022b). Statistické testování hypotéz třikrát jinak. *Učitel matematiky, 30*(1), 3–14.

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

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

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

[12] Lehmann, E. (1998). *Nonparametrics–statistical methods based on ranks* (Revised First Edition). Holden-Day Inc.

[13] McCornack, R. L. (1965). Extended Tables of the Wilcoxon Matched Pair Signed Rank Statistic. *Journal of the American Statistical Association, 60*, 864–871.

[14] Papica, J. (1992). *Neparametrická statistika.* VUP.

[15] Siegel, S. (1957). Nonparametric statistics. *The American*

Statistician, 11(3), 13–19.

[16] Škaloudová, A. (1998). *Statistika v pedagogickém a psychologickém výzkumu.* PedF UK.

[17] Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin, 1*(6), 80–83.

[18] Wilcoxon, F. (1947). Probability Tables for Individual Comparisons by Ranking Methods. *Biometrics, 3*(3), 119–122.

[19] Wilcoxon, F., Katti, S. K., & Wilcox R. (1963). *Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test.* Pearl River, American Cyanamid Co. and Florida State University.

Abstract

Wilcoxon tests belong among the frequently used nonparametric tests of statistical significance. The author of the mathematical model on which these tests are based is the American statistician Frank Wilcoxon (1892–1965). The article describes the essence and origin of these tests, especially Wilcoxon signed-rank test, in mathematical and historical contexts.

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