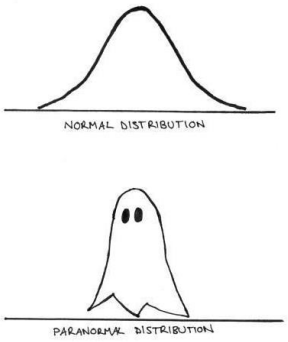
ANALYTICKÁ CHEMOMETRIE

neparametrické testy



Patrik Kania
Ústav analytické chemie, VŠCHT Praha

04. 11. 2025

Testování neparametrických hypotéz (angl. *nonparametric statistics*)

- nejsou většinou založeny na parametrech a nezávisí na typu rozdělení ZS
- použití v případě nemožnosti použít test parametrický
- mají nižší sílu testu $1-\beta$ než parametrické
- ideální pro malé výběry pocházející z nenormálních či neznámých rozdělení
- mají jednoduchý výpočet, ale těžkou teorii
- rozhodovací pravidla často nejsou stavěna přímo na hodnotách, ale na jejich pořadí v náhodném výběru
- lze použít i na případy, kdy neznáme hodnoty náhodného výběru

Znaménkový test pro jeden výběr (pro výběr s $n \leq 20$) (angl. *sign test*)

- alternativa k testu správnosti
- lze užít i jako párový test

testování hypotézy $\tilde{x} = \mu_0$

- neznáme rozdělení → medián

n - počet všech prvků výběru
P - počet kladných výchylek prvků výběru od mediánu
Z - počet nulových výchylek prvků výběru od mediánu
$n_{red} = n - Z$

pokud $P \leq k_{\alpha 1}(n_{red})$ nebo $P \geq k_{\alpha 2}(n_{red})$ hypotézu H_0 zamítáme

$k_{\alpha 1}(n_{red})$ a $k_{\alpha 2}(n_{red})$ jsou kritické hodnoty pro znaménkový test

<u>Příklad:</u> Při sledování rozpadu radioaktivního izotopu byly počítány částice emitované izotopem během 1 minuty. Pro celkem 7 pozorování byly získány tyto výsledky:

číslo měření	1	2	3	4	5	6	7
počet částic	205	194	189	211	206	208	202

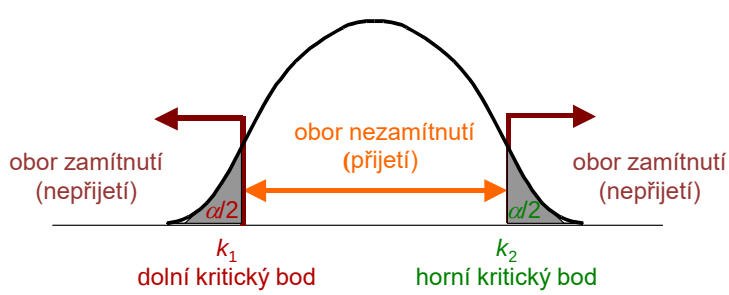
Na hladině významnosti 5 % určete, zda medián ZS je roven hodnotě 200.

$n = 7$
$P = 5$
$Z = 0$
$n_{red} = 7 - 0 = 7$
$k_{1;0,025}(7) = 0$
$k_{2;0,025}(7) = 7$

$0 < 5 < 7$ hypotézu H_0 nelze zamítnout

Kritické hodnoty pro znaménkový test

	$\alpha = 0{,}05$		$\alpha = 0{,}01$			$\alpha = 0{,}05$		$\alpha = 0{,}01$	
n	k_1	k_2	k_1	k_2	n	k_1	k_2	k_1	k_2
6	0	6	-	-	24	6	18	5	19
7	0	7	-	-	25	7	18	5	20
8	0	8	0	8	26	7	19	6	20
9	1	8	0	9	27	7	20	6	21
10	1	9	0	10	28	8	20	6	22
11	1	10	0	11	29	8	21	7	22
12	2	10	1	11	30	9	21	7	23
13	2	11	1	12	31	9	22	7	24
14	2	12	1	13	32	9	23	8	24
15	3	12	2	13	33	10	23	8	25
16	3	13	2	14	34	10	24	9	25
17	4	13	2	15	35	11	24	9	26
18	4	14	3	15	36	11	25	9	27
19	4	15	3	16	37	12	25	10	27
20	5	15	3	17	38	12	26	10	28
21	5	16	4	17	39	12	27	11	28
22	5	17	4	18	40	13	27	11	29
23	6	17	4	19					



pro oboustranný test platí:

zamítáme H_0, pokud počet kladných výchylek $P \leq k_1$ nebo $\geq k_2$

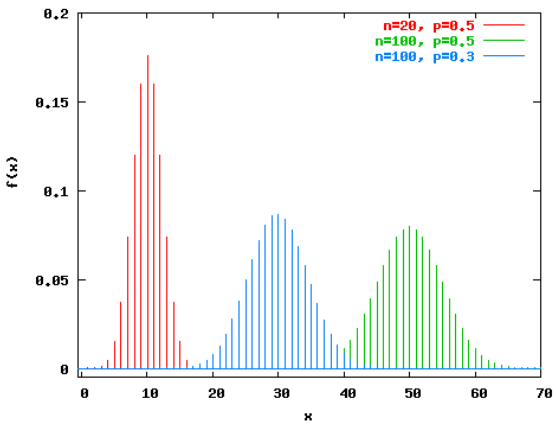
pro k_1 platí, že P(počet kladných výchylek $\leq k_1$) $\leq \alpha/2$
k_2 platí, že P(počet kladných výchylek $\geq k_2$) $\leq \alpha/2$

Binomické rozdělení (angl. *binomial distribution*)

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

$$f(x,n,p) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad \text{pro } x = 0, 1, 2, \dots, n$$

značí se Bi(n, p) p – prst jevu, n – počet opakování, x – počet výskytů jevu



$$\mathrm{E}(\xi) = np$$

$$\mathrm{D}(\xi) = np(1-p)$$

pro $n = 100$
$p = 0{,}3$

je $\mathrm{E}(\xi) = 30$
$\mathrm{D}(\xi) = 21$

$$f(x,n,p) = \binom{n}{x} p^x (1-p)^{n-x},$$

x počet kladných výchylek

n počet měření

p pravděpodobnost kladné výchylky

$$f(0;7;0{,}5) = \binom{7}{0} 0{,}5^0 (1-0{,}5)^{7-0} = 0{,}0078125 \quad < \alpha = 0{,}025$$

$$f(1;7;0{,}5) = \binom{7}{1} 0{,}5^1 (1-0{,}5)^{7-1} = 0{,}0546875 \quad > \alpha = 0{,}025$$

$$f(2;7;0{,}5) = \binom{7}{2} 0{,}5^2 (1-0{,}5)^{7-2} = 0{,}1640625$$

$$f(3;7;0{,}5) = \binom{7}{3} 0{,}5^3 (1-0{,}5)^{7-3} = 0{,}2734375$$

$$f(4;7;0{,}5) = \binom{7}{4} 0{,}5^4 (1-0{,}5)^{7-4} = 0{,}2734375$$

$$f(5;7;0{,}5) = \binom{7}{5} 0{,}5^5 (1-0{,}5)^{7-5} = 0{,}1640625$$

$$f(6;7;0{,}5) = \binom{7}{6} 0{,}5^6 (1-0{,}5)^{7-6} = 0{,}0546875$$

$$f(7;7;0{,}5) = \binom{7}{7} 0{,}5^7 (1-0{,}5)^{7-7} = 0{,}0078125$$

vzhledem k symetrii rozdělení platí: $k_2 = n - k_1$

Znaménkový test pro jeden výběr (pro výběr s $n \geq 20$) (angl. *sign test*)

- alternativa k testu správnosti
- lze užít i jako párový test

testování hypotézy $\tilde{x} = \mu_0$

- neznáme rozdělení → medián

$$TK = \frac{2 \cdot P - n_{red}}{\sqrt{n_{red}}}$$

n - počet všech prvků výběru
P - počet kladných výchylek od mediánu
Z - počet nulových výchylek od mediánu
$n_{red} = n - Z$

pokud $|TK| > z_{(\alpha/2)}$ hypotézu H_0 zamítáme

$z_{\alpha/2}$ je oboustranná kritická hodnota normovaného normálního rozdělení pro hladinu významnosti α

Pořadí hraje prim!

- u většiny neparametrických testů nejsou rozhodovací pravidla stavěna přímo na hodnotách, ale na jejich pořadí v náhodném výběru

x_i	1,2	2,1	1,8	2,7	2,3
R_i	1	3	2	5	4

v případě stejných hodnot ve výběru → spojená pořadí

x_i	1,2	2,1	2,1	2,7	2,3
R_i	1	2,5	2,5	5	4

Jednovýběrový Wilcoxonův test (angl. *one-sample Wilcoxon test*)

- použití ve stejných případech jako test znaménkový, ale je citlivější
- kromě mediánu používá i pořadí pozorovaných hodnot
- nejčastější použití jako párový test (jeden základní soubor)

testování hypotézy $X_1 - X_2$ má rozdělení symetrické kolem 0

1) určíme rozdíly $D_1 = X_{11} - X_{21}, \dots, D_n = X_{1n} - X_{2n}$
2) určíme počet nulových rozdílů D_i pro výpočet n_{red}
3) určíme pořadí nenulových rozdílů D_i dle jejich absolutních hodnot
4) provedeme zvlášť součet pořadí kladných S^+ a záporných S^- rozdílů

možná kontrola správnosti výpočtu: $S^+ + S^- = \frac{n_{red}(n_{red}+1)}{2}$

5) ze dvou získaných hodnot S^+ a S^- vybereme tu nižší a porovnáme s kritickou hodnotou

pokud $min(S^+, S^-) \leq \omega_\alpha(n_{red})$ hypotézu H_0 zamítáme

Příklad: Při sledování rozpadu radioaktivního izotopu byly počítány částice emitované izotopem během 1 minuty. Při studii byly použity dva různé detektory. Provedeno bylo celkem 7 párových experimentů a byly získány výsledky uvedené níže. Na hladině významnosti 5 % rozhodněte, zda použité detektory se liší svými citlivostmi.

číslo měření	1	2	3	4	5	6	7
detektor č. 1	205	194	189	211	206	208	202
detektor č. 2	200	194	186	215	200	198	205
rozdíl	5	0	3	-4	6	10	-3
pořadí hodnot	4	-	1,5	3	5	6	1,5

$n = 7$
$Z = 1$
$n_{red} = 7 - 1 = 6$
$S^+ = 4 + 1{,}5 + 5 + 6 = 16{,}5$
$S^- = 3 + 1{,}5 = 4{,}5$
$min(16{,}5;\ 4{,}5) = 4{,}5$
$\omega_{0,05}(6) = 0{,}6$

$4{,}5 > 0{,}6$

hypotézu H_0 nelze zamítnout

Kritické hodnoty pro jednovýběrový Wilcoxonův test

n	$\alpha = 0{,}05$	$\alpha = 0{,}01$	$\alpha = 0{,}001$
6	0,6	-	-
7	2,1	-	-
8	3,7	0,3	-
9	5,5	1,6	-
10	8,1	3,1	-
11	10,7	5,1	-
12	13,8	7,2	1,0
13	17,2	9,8	2,5
14	21,1	12,7	4,4
15	25,3	15,9	6,5
16	29,9	19,5	9,0
17	34,9	23,4	11,7
18	40,3	27,7	14,8
19	46,1	32,4	18,2
20	52,3	37,4	21,9
21	59,0	42,7	26,0
22	66,0	48,6	30,4
23	73,4	54,8	35,2
24	81,3	61,3	40,3
25	89,5	68,3	45,8
30	137,1	109,0	78,5
35	195,3	159,7	120,5
40	264,2	220,4	172,1
45	343,9	291,5	233,3
50	434,5	373,0	304,5

Příklad: Při sledování rozpadu radioaktivního izotopu byly počítány částice emitované izotopem během 1 minuty. Při studii byly použity dva různé detektory. Provedeno bylo celkem 7 párových experimentů a byly získány výsledky uvedené níže. Na hladině významnosti 5 % rozhodněte, zda použité detektory se liší svými citlivostmi.

číslo měření	1	2	3	4	5	6	7
detektor č. 1	204	194	189	211	206	208	202
detektor č. 2	200	194	186	215	216	198	205
rozdíl	4	0	3	-4	-10	10	-3
pořadí hodnot	3,5	-	1,5	3,5	5,5	5,5	1,5

$n = 7$

$Z = 1$

$n_{red} = 7 - 1 = 6$

$S^+ = 3{,}5 + 1{,}5 + 5{,}5 = 10{,}5$

$S^- = 3{,}5 + 1{,}5 + 5{,}5 = 10{,}5$

$min(10{,}5;\ 10{,}5) = 10{,}5$

$\omega_{0,05}(6) = 0{,}6$

1. modifikace příkladu

$10{,}5 > 0{,}6$

hypotézu H_0 nelze zamítnout

Příklad: Při sledování rozpadu radioaktivního izotopu byly počítány částice emitované izotopem během 1 minuty. Při studii byly použity dva různé detektory. Provedeno bylo celkem 7 párových experimentů a byly získány výsledky uvedené níže. Na hladině významnosti 5 % rozhodněte, zda použité detektory se liší svými citlivostmi.

číslo měření	1	2	3	4	5	6	7
detektor č. 1	204	194	189	215	216	208	205
detektor č. 2	200	194	186	211	206	198	202
rozdíl	+4	0	+3	+4	+10	+10	+3
pořadí hodnot	3,5	-	1,5	3,5	5,5	5,5	1,5

$n = 7$

$Z = 1$

$n_{red} = 7 - 1 = 6$

2. modifikace příkladu

$S^+ = 3{,}5 + 1{,}5 + 3{,}5 + 5{,}5 + 5{,}5 + 1{,}5 = 21$

$S^- = 0$

$min(21; 0) = 0$

$\omega_{0,05}(6) = 0{,}6$

$0 < 0{,}6$

hypotézu H_0 lze zamítnout

Dvouvýběrový Wilcoxonův test (*two-sample Wilcoxon/Mann-Whitney U test*)

- alternativa k parametrickému dvouvýběrovému testu rovnosti středních hodnot, kdy nelze předpokládat normalitu obou základních souborů
- i v případech, kdy známe pouze pořadí hodnot výběru nikoliv hodnoty
- není třeba provádět test rovnosti rozptylů
- výběry musí být nezávislé

H_0: $\tilde{x}_1 = \tilde{x}_2$ H_1: $\tilde{x}_1 \neq \tilde{x}_2$

1) určíme tzv. sdružené pořadí hodnot X_{11}, ….., X_{1n}, X_{21}, ….., X_{2n}
2) vypočteme součty T_1 a T_2 sdruženého pořadí hodnot 1. výběru a 2. výběru
3) vypočteme testovací kritéria U_1 a U_2
4) porovnáme hodnoty U_1 a U_2 a vybereme tu nižší
5) vybranou nižší U porovnáme s kritickou hodnotou

pokud $min(U_1, U_2) \leq W_\alpha(n_1, n_1)$ hypotézu H_0 zamítáme

Příklad: Porovnáním výsledků čistoty epoxidu byl testován vliv nového způsobu jeho výroby. Pro analýzu bylo vybráno 8 vzorků získaných starou metodou výroby a 5 vzorků z nové výrobny.

starý způsob:	5,7	5,5	4,3	5,9	5,2	5,6	5,8	5,1
nový způsob:	5,0	4,5	4,2	5,4	4,4			

H_0: způsob výroby nemá vliv na čistotu epoxidu H_1: neplatí H_0

hodnota	4,2	4,3	4,4	4,5	5,0	5,1	5,2	5,4	5,5	5,6	5,7	5,8	5,9
pořadí	1	2	3	4	5	6	7	8	9	10	11	12	13

$$T_1 = \sum_{1.výběr} R_j = 2 + 6 + 7 + 9 + 10 + 11 + 12 + 13 = 70$$

$$T_2 = \sum_{2.výběr} R_j = 1 + 3 + 4 + 5 + 8 = 21$$

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - T_1 = 8 \cdot 5 + \frac{8(8+1)}{2} - 70 = 6$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - T_2 = 8 \cdot 5 + \frac{5(5+1)}{2} - 21 = 34$$

$$min(U_1, U_2) = 6$$

$$W_{0,05}(8,5) = 6$$

Jelikož $min(U_1, U_2) = W_\alpha(n_1, n_1)$ hypotézu H_0 zamítáme

způsob výroby má vliv na čistotu epoxidu

odvození vzorce pro Mann-Whitney test

Kritické hodnoty pro Wilcoxonův dvouvýběrový test; $W_\alpha(n_1, n_2)$. $\alpha = 0.05$

n_1 \ n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
4	-	-	-	0																
5	-	-	0	1	2															
6	-	-	1	2	3	5														
7	-	-	1	3	5	6	8													
8	-	0	2	4	6	8	10	13												
9	-	0	2	4	7	10	12	15	17											
10	-	0	3	5	8	11	14	17	20	23										
11	-	0	3	6	9	13	16	19	23	26	30									
12	-	1	4	7	11	14	18	22	26	29	33	37								
13	-	1	4	8	12	16	20	24	28	33	37	41	45							
14	-	1	5	9	13	17	22	26	31	36	40	45	50	55						
15	-	1	5	10	14	19	24	29	34	39	44	49	54	59	64					
16	-	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75				
17	-	2	6	11	17	22	28	34	39	45	51	57	63	69	75	81	87			
18	-	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99		
19	-	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	
20	-	2	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127
21	-	3	8	15	22	29	36	43	50	58	65	73	80	88	96	103	111	119	126	134
22	-	3	9	16	23	30	38	45	53	61	69	77	85	93	101	109	117	125	133	141
23	-	3	9	17	24	32	40	48	56	64	73	81	89	98	106	115	123	132	140	149
24	-	3	10	17	25	33	42	50	59	67	76	85	94	102	111	120	129	138	147	156
25	-	3	10	18	27	35	44	53	62	71	80	89	98	107	117	126	135	145	154	163
26	-	4	11	19	28	37	46	55	64	74	83	93	102	112	122	132	141	151	161	171
27	-	4	11	20	29	38	48	57	67	77	87	97	107	117	127	137	147	158	168	178
28	-	4	12	21	30	40	50	60	70	80	90	101	111	122	132	143	154	164	175	186
29	-	4	13	22	32	42	52	62	73	83	94	105	116	127	138	149	160	171	182	193
30	-	5	13	23	33	43	54	65	76	87	98	109	120	131	143	154	166	177	189	200

Konec