

10. Neparametrické testy o mediánech

10.1. Motivace

10.2. Pojem pořadí a průměrného pořadí

10.3. Jednovýběrový znaménkový test a jeho asymptotická varianta

10.4. Příklad: U 10 náhodně vybraných vzorků benzínu byly zjištěny následující hodnoty oktanového čísla: 98,2 96,8 96,3 99,8 96,9 98,6 95,6 97,1 97,7 98,0. Na hladině významnosti 0,05 testujte hypotézu, že medián oktanového čísla je 98 proti oboustranné alternativě.

Řešení:

rozdíly $x_i - 98$: 0,2 -1,2 -1,7 1,8 -1,1 0,6 -2,4 -0,9 -0,3 0,0

$S_Z^+ = 3$, nenulových rozdílů je 9. Ve statistických tabulkách najdeme pro $n = 9$ a $\alpha = 0,05$ kritické hodnoty $k_1 = 1$, $k_2 = 8$. Protože kritický obor $W = \langle 0,1 \rangle \cup \langle 8,9 \rangle$ neobsahuje hodnotu 3, nemůžeme H_0 zamítnout na hladině významnosti 0,05.

10.5. Párový znaménkový test

10.6. Příklad: U osmi osob byl změřen systolický krevní tlak před pokusem a po něm.

č. osoby	1	2	3	4	5	6	7	8
tlak před	130	185	162	136	147	181	138	139
tlak po	139	190	175	135	155	175	158	149

Na hladině významnosti 0,05 testujte hypotézu, že pokus neovlivní systolický krevní tlak

Řešení:

rozdíly $x_i - y_i$: -9 -5 -13 1 -8 6 -30 -10

Testová statistika $S_Z^+ = 2$. Ve statistických tabulkách najdeme pro $n = 8$ a $\alpha = 0,05$ kritické hodnoty $k_1 = 0$, $k_2 = 8$. Protože kritický obor $W = 0 \cup 8$ neobsahuje hodnotu 2, nemůžeme H_0 zamítnout na hladině významnosti 0,05. Znamená to, že s rizikem omylu nejvýše 0,05 je zvýšení krevního tlaku stejně pravděpodobné jako jeho pokles.

10.7. Jednovýběrový Wilcoxonův test a jeho asymptotická varianta

10.8. Příklad: Pro zadání příkladu 10.4. proveďte jednovýběrový Wilcoxonův test.

Řešení:

Absolutní hodnoty rozdílů $x_i - 98$ setřídíme vzestupně podle velikosti (přitom vynecháme nulový rozdíl a kladné rozdílů značíme tučně):

abs ($x_i - 98$) **0,2** 0,3 **0,6** 0,9 1,1 1,2 1,7 **1,8** 2,4

pořadí R_i **1** 2 **3** 4 5 6 7 **8** 9

Součet pořadí přes kladné hodnoty rozdílů: $S_W^+ = 12$

Součet pořadí přes záporné hodnoty rozdílů: $S_W^- = 33$

Testová statistika = $\min(12,33) = 12$, tabelovaná kritická hodnota pro $\alpha = 0,05$ a $a = 9$ je 5.

Protože $12 > 5$, H_0 nezamítáme na hladině významnosti 0,05.

10.9. Párový Wilcoxonův test

10.10. Příklad: Pro data z příkladu 10.6. proveďte párový Wilcoxonův test.

Řešení:

Absolutní hodnoty rozdílů $x_i - y_i$ seřídíme vzestupně podle velikosti (kladné rozdíly značíme tučně):

abs ($x_i - y_i$) **1** 5 **6** 8 9 10 13 20

pořadí R_i **1** 2 **3** 4 5 6 7 8

Součet pořadí přes kladné hodnoty rozdílů: $S_W^+ = 4$

Součet pořadí přes záporné hodnoty rozdílů: $S_W^- = 32$

Testová statistika = $\min(4,3) = 4$, tabelovaná kritická hodnota pro $\alpha = 0,05$ a $n = 8$ je 3.

Protože $4 > 3$, H_0 nezamítáme na hladině významnosti 0,05.

10.11. Příklad (na asymptotickou variantu Wilcoxonova testu):

30 náhodně vybraných osob mělo nezávisle na sobě bez předchozího nácviku odhadnout, kdy od daného signálu uplyne 1 minuta. Byly získány následující výsledky (v sekundách): 53 48 45 55 63 51 66 56 50 58 61 51 64 63 59 47 46 58 52 56 61 57 48 62 54 49 51 46 53 58.

Na asymptotické hladině významnosti 0,05 testujte hypotézu, že medián rozložení, z něhož daný náhodný výběr pochází, je 60 sekund proti oboustranné alternativě (nulová hypotéza vlastně tvrdí, že polovina osob délku jedné minuty podhodnotí a druhá nadhodnotí).

Řešení:

Testujeme $H_0: x_{0,50} = 60$ proti oboustranné alternativě $H_1: x_{0,50} \neq 60$.

Obvyklým způsobem stanovíme statistiku $S_W^+ = 55$.

Asymptotická testová statistika:

$$U_0 = \frac{S_W^+ - E(S_W^+)}{\sqrt{D(S_W^+)}} = \frac{S_W^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} = \frac{55 - \frac{30(30+1)}{4}}{\sqrt{\frac{30(30+1)(2 \cdot 30+1)}{24}}} = -3,65$$

Kritický obor:

$$W = (-\infty, -u_{1-\alpha/2}) \cup (u_{1-\alpha/2}, \infty) = (-\infty, -u_{0,975}) \cup (u_{0,975}, \infty) = (-\infty, -1,96) \cup (1,96, \infty).$$

Testová statistika se realizuje v kritickém oboru, tedy H_0 zamítáme na asymptotické hladině významnosti 0,05.

10.12. Dvouvýběrový Wilcoxonův test a a jeho asymptotická varianta

10.13. Příklad: Výrobce určitého výrobku se má rozhodnout mezi dvěma dodavateli polotovarů vyrábějících je různými technologiemi. Rozhodující je procentní obsah určité látky.

1. technologie: 1,52 1,57 1,71 1,34 1,68

2. technologie: 1,75 1,67 1,56 1,66 1,72 1,79 1,64 1,55

Na hladině významnosti 0,05 posuďte pomocí dvouvýběrového Wilcoxonova testu, zda je oprávněný předpoklad, že obě technologie poskytují stejné procento účinné látky.

Řešení:

usp.h. **1,34** **1,52** 1,55 1,56 **1,57** 1,64 1,66 1,67 **1,68** **1,71** 1,72 1,75 1,79

pořadí **1** **2** 3 4 **5** 6 7 8 **9** **10** 11 12 13

$T_1 = 1 + 2 + 5 + 9 + 10 = 27$, $T_2 = 3 + 4 + 6 + 7 + 8 + 11 + 12 + 13 = 64$

$U_1 = 5 \cdot 8 + 5 \cdot 6 / 2 - 27 = 28$, $U_2 = 5 \cdot 8 + 8 \cdot 9 / 2 - 64 = 12$

Kritická hodnota pro $\alpha = 0,05$, $\min(5,8) = 5$, $\max(5,8) = 8$ je 6. Protože $\min(28,12) > 2$, nemůžeme na hladině významnosti 0,05 zamítnout hypotézu, že obě technologie poskytují stejné procento účinné látky.

10.14. Kruskalův - Wallisův test

10.15. Příklad: V roce 1980 byly získány tři nezávislé výběry obsahující údaje o průměrných ročních příjmech (v tisících dolarů) čtyř sociálních skupin ve třech různých oblastech USA.

jižní oblast: 6 10 15 29

pacifická oblast: 11 13 17 131

severovýchodní oblast: 7 14 28 25

Na hladině významnosti 0,05 testujte hypotézu, že příjmy v těchto oblastech se neliší.

Zamítnete-li nulovou hypotézu, vyšetřete, které dvojice výběrů se od sebe liší na hladině významnosti 0,05.

Řešení:

Usp.hodnoty	6	7	10	11	13	14	15	17	25	28	29	131
Pořadí 1.výběru	1		3				7				11	
Pořadí 2.výběru				4	5			8				12
Pořadí 3.výběru		2				6			9	10		

$$T_1 = 22, T_2 = 29, T_3 = 27, Q = \frac{12}{12 \cdot 13} \left(\frac{22^2}{4} + \frac{29^2}{4} + \frac{27^2}{4} \right) - 3 \cdot 13 = 0,5, \chi_{0,95}^2(2) = 5,991.$$

Protože $Q < 5,991$, H_0 nezamítáme na asymptotické hladině významnosti 0,05. Rozdíly mezi průměrnými ročními příjmy v uvedených třech oblastech se neprokázaly.

10.16. Mediánový test

10.17. Příklad: Pro data z příkladu 10.15. proveďte mediánový test.

Řešení:

Medián všech 12 hodnot je 14,5. V 1. výběru jsou dvě hodnoty větší nebo rovny mediánu, ve

2. výběru 2 hodnoty, ve 3. výběru 2 hodnoty. $Q_M = 4 \left[\frac{1}{4} (2^2 + 2^2 + 2^2) \right] - 12 = 0, \chi_{0,95}^2(2) =$

5,991. Protože $Q_M < 5,991$, H_0 nezamítáme na asymptotické hladině významnosti 0,05.

10.18. Metody mnohonásobného porovnávání

a) Neményiho metoda

b) Obecná metoda mnohonásobného porovnávání