

Jednoduché třídění s nezávislými výběry

Kruskalův - Wallisův test

je zobecněním dvouvýběrového Wilcoxonova testu

hypotéza $H_0: F_1(x) = \dots = F_I(x)$ pro vš. x

alternativa: $F_i \neq F_j$ pro nějaké i, j

postup: všechny výběry se uspořádají ^{jedné} dohromady ^{podle} rostoucí ^{postupnosti} (vznikne uspořádaný sdružený výběr)

pro každé y_{ij} se určí pořadí R_{ij} ve sdruženém výběru
máme:

<u>výběr</u>	<u>pořadí ve sdruženém výběru</u>	<u>součet pořadí</u>
1	$R_{11} \ R_{12} \ \dots \ R_{1n_1}$	T_1
2	$R_{21} \ R_{22} \ \dots \ R_{2n_2}$	T_2
$\vdots$	$\vdots$	$\vdots$
I	$R_{I1} \ R_{I2} \ \dots \ R_{In_I}$	T_I

platí $T = T_1 + \dots + T_I = \frac{1}{2} n(n+1)$, kde $n = \sum_{i=1}^I n_i$

za platnosti H_0 je $E T_i = \frac{1}{2} n_i (n+1)$

Testová statistika je

$$Q = \frac{12(n+1)}{n(n+1)(n-1)} \sum_{i=1}^I \frac{1}{n_i} \left(T_i - \frac{1}{2} n_i (n+1) \right)^2 =$$

$$= \frac{12}{n(n+1)} \sum_{i=1}^I \frac{T_i^2}{n_i} - 3(n+1), \quad EQ = I - 1$$

Q má asymptoticky rozdělení χ^2_{I-1}

H_0 se zamítá když $Q \geq \chi^2_{I-1}(\alpha)$

krit. hodnota

%

(případně se v tabulkách dají najít i přesné krit. hodnoty)

> Pokud se H_0 zamítá, pokračuje se testování rozdílu dvojic:

$$\text{když } \left| \frac{T_i}{n_i} - \frac{T_j}{n_j} \right| > \sqrt{\frac{1}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right) n(n+1) h_{I-1}(\alpha)}, \text{ pak prohlásíme,}$$

že se distribuční fee i -tého a j -tého výběru významně liší.

$h_{I-1}(\alpha)$ je přesná krit. hodnota

$h_{I-1}(\alpha) \doteq \chi^2_{I-1}(\alpha)$ pro velké výběry

Friedmanův test

- 1) Naměřené hodnoty se rozdělí do skupin podle úrovně náhodného efektu
V každé skupině se jim přiřadí pořadí (podle uspořádání v rostoucím pořadí)
- 2) Data se přerozdělí do výběrů podle perného efektu
V rámci každého výběru se sečtou pořadí přiřazená v bodě 1)

Pro platnosti H_0 budou součty pořadí ve všech výběrech z bodu 2) přibližně stejné

R_{ij} = pořadí j -tého prvku z i -tého výběru v rámci skupiny dané náhodným efektem

$$Q = \frac{12}{rI(I+1)} \sum_{i=1}^I \left(\sum_{j=1}^r R_{ij} \right)^2 - 3r(I+1) \quad \text{je testová statistika}$$

kritické hodnoty bývají tabelované

pro velké hodnoty I a r se dají použít krit. hodnoty χ^2_{I-1} rozdělení

Dá se použít i pro nezávislé výběry.

data file: quicksorttime.STA [25 cases with 9 variables]

MN *M1* *M3* *M3-5* *M3-7* *M3-11* *M3-17*
~~*M3-11*~~ ~~*M3-17*~~ ~~*M3-5-11*~~ *M3-5-11* *M3-5-17*

	1 VAR1	2 VAR2	3 VAR3	4 VAR4	5 VAR5	6 VAR6	7 VAR7	8 VAR8	9 VAR9
1	7.950	8.010	7.710	7.530	7.580	7.520	7.500	7.510	7.530
2	8.060	8.120	7.580	7.520	7.420	7.460	7.480	7.450	7.470
3	8.290	7.960	7.630	7.520	7.530	7.520	7.470	7.520	7.520
4	8.310	8.210	7.680	7.670	7.510	7.470	7.460	7.480	7.430
5	8.060	8.070	7.740	7.540	7.500	7.540	7.470	7.500	7.330
6	8.050	8.280	7.670	7.640	7.480	7.520	7.480	7.520	7.490
7	7.980	8.010	7.630	7.560	7.520	7.530	7.460	7.480	7.480
8	8.060	7.950	7.670	7.530	7.580	7.490	7.470	7.520	7.430
9	8.130	8.050	7.640	7.530	7.540	7.500	7.440	7.460	7.450
10	8.080	7.950	7.600	7.440	7.500	7.530	7.480	7.440	7.420
11	8.440	8.000	7.610	7.460	7.470	7.600	7.420	7.490	7.560
12	8.060	8.260	7.610	7.610	7.520	7.490	7.470	7.480	7.440
13	8.110	7.960	7.630	7.680	7.650	7.470	7.470	7.520	7.420
14	8.080	7.970	7.610	7.660	7.550	7.480	7.440	7.510	7.470
15	8.200	7.970	7.680	7.560	7.420	7.470	7.430	7.460	7.480
16	8.050	7.940	7.630	7.520	7.500	7.500	7.460	7.530	7.540
17	8.270	7.960	7.670	7.520	7.440	7.460	7.410	7.440	7.470
18	8.140	7.980	7.610	7.500	7.500	7.500	7.470	7.470	7.470
19	8.110	8.180	7.670	7.540	7.550	7.550	7.490	7.450	7.480
20	8.050	8.100	7.710	7.520	7.500	7.510	7.480	7.490	7.490
21	8.020	8.020	7.550	7.600	7.550	7.460	7.480	7.530	7.500
22	8.110	8.110	7.600	7.460	7.490	7.510	7.510	7.500	7.440
23	8.050	8.030	7.620	7.630	7.560	7.490	7.500	7.520	7.420
24	8.300	7.800	7.580	7.530	7.510	7.380	7.470	7.490	7.460
25	8.170	8.210	7.660	7.510	7.520	7.410	7.480	7.520	7.400

Quicksort - postupyprasti delky 1000 000

meri se oas

Skupiny:	pevný efekt, M3	M3-5	M3-7	M3-11
na h. efekt				
1	4.	2.	3.	1.
2	4.	3.	1.	2.
3	4.	1.	3.	1.
4	4.	3.	2.	1.
5	4.	2.	1.	2.
6	4.	3.	1.	2.
7	4.	3.	1.	2.
8	4.	2.	3.	1.
9	4.	2.	3.	1.
10	4.	1.	2.	3.
11	4.	1.	2.	3.
12	3.	3.	2.	1.
13	2.	4.	3.	1.
14	3.	4.	2.	1.
15	4.	3.	1.	2.
16	4.	3.	1.	1.
17	4.	3.	1.	2.
18	4.	1.	1.	1.
19	4.	1.	2.	2.
20	4.	3.	1.	2.
21	2.	4.	2.	1.
22	4.	1.	2.	3.
23	3.	4.	2.	1.
24	4.	3.	2.	1.
25	4.	2.	3.	1.

součty pořadí: 93 62 47 39

$$Q = \frac{12}{25 \cdot 4 \cdot 5} \sum_{i=1}^4 \left(\sum_{j=1}^{25} R_{ij} \right)^2 - 3 \cdot 25 \cdot 5 = \frac{12}{500} (93^2 + 62^2 + 47^2 + 39^2) - 375 =$$

$$= \frac{12}{500} (8649 + 3844 + 2209 + 1521) - 375 =$$

$$= \frac{12}{500} \cdot 16223 - 375 = 389,352 - 375 = 14,352$$

krit. hodnota Friedmannova testu na hladině 0,05 je 7,82

-11- $\chi^2_3(0,05) = 7,81$

Hypotéza, že použité algoritmy nemá vliv na dobu výpočtu, se zamítá!

Zjištění, která dvojice hypotézu porušuje:

vyšetříme rozdíly $|\sum_{j=1}^n R_{ij} - \sum_{j=1}^n R_{tj}|$ pro všechny dvojice i, t

a porovnáme s tabulovanými kritickými hodnotami

$$|93 - 62| = 31 \quad |93 - 47| = 46 \quad |93 - 39| = 54$$

$$|62 - 47| = 15 \quad |62 - 39| = 23$$

$$|47 - 39| = 8$$

krit. hodnota na hladině 0,05 je asi 19 (možná o něco větší - tabulky jsou neúplné)

podle toho jsou jediné nevýznamné rozdíly mezi algoritmy

M3-5 a M3-7

M3-7 a M3-11

(možná ještě M3-5 a M3-11)

Pro srovnání - v klasickém jednoduchém třídění vyšly všechny rozdíly nevýznamné

dvouvýběrovým t-testem vyšel nevýznamný rozdíl mezi M3-7 a M3-11