26/04/2017

k πληθυσμοί - k δείγματα ανεξάρτητα : Το τεστ των Kruskal-Wallis

Έστω k πληθυσμοί, με μέσες τιμές $\mu_1, \dots, \mu_k$ (ή $\omega_1, \dots, \omega_k$) και σ.κ. $F_1, \dots, F_k$.

Έστω X_{ij}, $i=1,\dots,k$, $j=1,\dots,n_i$ οι παρατηρήσεις, όπου X_{ij} η j μέτρηση στο δείγμα i, $n = \sum_{i=1}^{k} n_i$.

Ενδιαφέρει: $H_0: \mu_1 = \dots = \mu_k$ (ή $\omega_1 = \dots = \omega_k$ ή $F_1 = \dots = F_k$) v

H_a: όχι όλα τα μ_i ίσα μεταξύ τους

Αναμειγνύουμε τα k δείγματα σε ένα ενιαίο δείγμα και ταξινομούμε με την αύξηση τάξη μεγέθους, συμβολίζοντας με R_{ij} την αντίστοιχη τάξη της X_{ij} μέτρησης (Υποθέτουμε ότι δεν υπάρχουν ισοτιμίες).

Έστω, R_i το άθροισμα των τάξεων για τις παρατηρήσεις του $i^{\text{οστού}}$ δείγματος, δηλαδή $R_i = \sum_{j=1}^{n_i} R_{ij}$, με $\frac{R_i}{n_i}$ τη μέση τιμή.

Υπό την H_0, $\frac{R_1}{n_1} = \dots = \frac{R_k}{n_k} = \frac{R_1 + \dots + R_k}{n_1 + \dots + n_k} = \frac{\frac{n(n+1)}{2}}{n} = \frac{n+1}{2}$,

όπου $\sum_{i=1}^{k} R_i = R_1 + \dots + R_k = \frac{n(n+1)}{2}$ και $\sum_{i=1}^{k} \left(\frac{R_i}{n_i} - \frac{n+1}{2} \right)^2 \cong 0$.

Άρα, έχουμε το στατιστικό:

$K = \frac{12}{n(n+1)} \sum_{i=1}^{k} n_i \left(\bar{R}_i - \bar{R} \right)^2$, $\bar{R}_i = \frac{R_i}{n_i}$, $\bar{R} = \frac{n+1}{2}$ $\underline{\underline{n}}$

$$\boxed{K = \frac{12}{n(n+1)} \sum_{i=1}^{k} n_i \left(R_i/n_i - {}^{n+1}/_2 \right)^2} \quad \left(\leftarrow SS_{tre} = \sum_i n_i (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 \right)$$

$$\underline{\underline{n}} \quad \boxed{K = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1)}$$

- $\frac{R_i - E(R_i)}{\sqrt{Var R_i}} \overset{\text{ασυμπτ.}}{\sim} N(0,1) \rightarrow \sum_{i=1}^{k} \left(\frac{R_i - E(R_i)}{\sqrt{Var R_i}} \right)^2 \overset{\text{ασυμπτ.}}{\sim} X^2_{k-1}$ με

$$E(R_i) = n_i \frac{n+1}{2}, \quad Var R_i = \frac{n_i (n-n_i)(n+1)}{12}$$

Άρα είναι: $K \sim X^2_{k-1}$, γιατί $\Sigma R_i = \frac{n(n+1)}{2}$ και κρίσιμη περιοχή $K \geq X^2_{a,k-1}$

Για r σετ από ισοτιμίες, $c = 1 - \sum_{i=1}^{r} f_i(f_i^2 - 1) / [n(n^2-1)]$ και $K^* = \frac{K}{c}$

$K = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1)$, $K = \frac{12}{n(n+1)} \sum_{i=1}^{k} n_i \left(\frac{R_i}{n_i} - \frac{n+1}{2} \right)^2$, $p = P(K \geq x)$ εα.

$K^* = \frac{K}{c}$, $K \geq X^2_{k-1,a}$, $c = 1 - \sum_{1}^{r} f_i(f_i^2-1)/n(n^2-1)$

Παράδειγμα: (5.1, σελ. 89 - Μπατσίδης)

$K \sim ;$ $n_1 = 2$, $n_2 = 1$, $n_3 = 1$ και $P(K \geq 2.7) = 0.5$

ΛΥΣΗ

Έστω $\overset{x}{X_{11}}, \overset{x}{X_{12}}, \overset{y}{X_{21}}, \overset{z}{X_{31}}$ οι παρατηρήσεις. Οι $12 = \binom{4}{2,1,1}$ δυνατές περιπτώσεις είναι:

Δείγματα: xxyz, xyxz, xyzx, yxzx, yzxx, yxxz, xxzy, xzxy, xzyx, zxyx, zyx[illegible]

(R_1, R_2, R_3): (3,3,4), (4,2,4), (5,2,3), (6,1,3), (7,1,2), (5,1,4), (3,4,3), (4,4,2), (5,3,2), (6,3,1), (7,2,[illegible]

($n=4$) K: 2.7 1.8 0.3 1.8 2.7 2.7 2.7 1.8 0.3 1.8 2.7

$$P(X=x) = \begin{cases} 1/2, & x = 2.7 \\ 1/3, & x = 1.8 \\ 1/6, & x = 0.3 \end{cases}$$

Παράδειγμα: (5.2, σελ. 90 - Μπατσίδης)

Για $k=2$, W-M-W και K-W ταυτίζονται ως X^2_1

$K = \frac{12}{n(n+1)} \sum_{1}^{2} n_i \left(\frac{R_i}{n_i} - \frac{n+1}{2} \right)^2$ με $R_1 + R_2 = \frac{n(n+1)}{2}$, $n = n_1 + n_2$,

$ER_1 = \frac{n_1(n+1)}{2}$, $VarR_1 = \frac{n_1 n_2 (n+1)}{2}$

Οπότε $X^2_1 \sim Z^2 = \left(\frac{R_1 - ER_1}{\sqrt{VarR_1}} \right)^2$ των W-M-W

Παράδειγμα 7.12 : $R_3 = 5+7+8 = 20$

$K = \left(\ldots + \frac{20^2}{3} + \frac{19^2}{2} \right) = 6.93$

$K^* = 6.93 / 0.994 = 6.9728$ και $\chi^2_{3,0.05} = 7.815$

Παράδειγμα:

H_1 | 340, 345, 330, 342

H_2 | 339, 333, 344, 338

H_3 | 347, 343, 349, 355

$H_0: \mu_1 = \mu_2 = \mu_3$ v H_a: όχι όλα τα μ_i ίσα μεταξύ τους

$n_1 = n_2 = n_3 = 4$, $n = 12$, $a = 0.05$

ΛΥΣΗ

Δείγματα:	1	2	2	2	1	1	3	2	1	3	3	3
Ενιαίο Διαταγμένο δείγμα:	330	333	338	339	340	342	343	344	345	347	349	355
Τάξεις:	1	2	3	4	5	6	7	8	9	10	11	12

$R_1 = 1+5+6+9 = 21$

$R_2 = 2+3+4+8 = 17$

$R_3 = 7+10+11+12 = 40$

$K = \frac{12}{12(12+1)} \left(\frac{21^2}{4} + \frac{17^2}{4} + \frac{40^2}{4} \right) - 3(12+1) = 5.8077$

$p = P(K \geq \overset{5.6923}{5.8077}) < 0.049$ $(< 0.05 = a)$, άρα απορρίπτεται η H_0

	Δείγματα			
	1°	2°	3°	4°
	φωτοπεριοδικές συνθήκες (ώρες φωτός/μέρα)			
	6h	12h	18h	24h
χρόνος (μέρες) που χρειάζεται να τραφεί κάθε τσιμπούρι	7.1	8.6	12.0	9.1
	14.3	11.0	13.9	14.5
	14.3	9.0	14.1	11.5
	13.4	12.6	8.7	12.7
	10.7	14.8	13.2	11.7
	11.1	14.3	11.1	11.1
			12.0	12.0

H_0: τα τέσσερα δείγματα προέρχονται από τον ίδιο πληθυσμό (δηλαδή δεν επηρεάζουν οι φωτοπεριοδικές συνθήκες)

$n_1 = 6$, $n_2 = 6$, $n_3 = 7$, $n_4 = 7$, $n = 26$, $a = 0.05$

Δείγματα	1	2	3	2	4	1	2	1	3	4	4	4	3	3	4	2	4
Ενιαίο	7.1	8.6	8.7	9	9.1	10.7	11	11.1	11.1	11.1	11.5	11.7	12	12	12	12.6	12.7
Τάξεις	1	2	3	4	5	6	7	9	9	9	11	12	14	14	14	16	17

(...) Δείγματα	3	1	3	3	1	1	2	4	2
Ενιαίο	13.2	13.4	13.9	14.1	14.3	14.3	14.3	14.5	14.8
Τάξεις	18	19	20	21	23	23		25	26

$$
\left.\begin{aligned}
R_1 &= 1+6+9+19+23+23 = 81 \\
R_2 &= 2+4+7+16+23+26 = 93 \\
R_3 &= 3+9+14+14+18+20+21 = 99 \\
R_4 &= 5+9+11+12+14+17+25 = 93
\end{aligned}\right\} \Rightarrow
$$

$$\Rightarrow K = \frac{12}{26(26+1)} \left(\frac{81^2}{6} + \frac{93^2}{6} + \frac{99^2}{7} + \frac{93^2}{7} \right) - 3(26+1) = 0.081$$

$f_1 = 3 = f_2 = f_3, \quad C = 1 - \frac{3(3^2-1)+3(3^2-1)+3(3^2-1)}{26(26^2-1)} = 0.996$

$K^* = \frac{0.081}{0.996} = 0.0813 < \chi^2_{c-1,a} \left(= \chi^2_{3,0.05} = 7.815 \right)$

Άρα, δεν απορρίπτεται η H_0