Μαθημα 8:

17/4/19

Κ πληθυσμοί - κ δειγματα ανεξ/τα | Το test Kruskal-Wallis

Εστω κ πληθυσμοί με μέσες τιμές $\mu_1, \dots, \mu_k$ (ή διαμεσους $m_1, \dots, m_k$) και αθροιστικές σ.κ $F_1, \dots, F_k$.
Εστω x_{ij}, $i=1, \dots, k$, $j=1, \dots, n_i$ οι παρατηρήσεις όπου x_{ij} η j μέτρηση στο δειγμα i. $n = \sum_{i=1}^{k} n_i$ (συνολικές παρατηρ.)
Μας ενδιαφέρει ο έλεγχος $H_0: \mu_1 = \dots = \mu_k$ $(m_1 = \dots = m_k)$
ή $H_0: F_1 = \dots = F_k$
v H_a: οχι όλα τα μ_i ισα μεταξύ τους.

Αναμειγνύουμε τα κ-δειγματα σε ενα ενιαιο δειγμα κατα αύξουσα τάξη μεγέθους.
Εστω R_{ij} οι τάξεις των x_{ij} μετρήσεων και $R_i = \sum_{j=1}^{n_i} R_{ij}$ το άθροισμα των τάξεων για το δειγμα i (στο ενιαιο δειγμα).

Υπο την Ho, περιμένουμε $\frac{R_1}{n_1} = \dots = \frac{R_k}{n_k}$ (μεσοι όροι των τάξεων)

$$\left(= \frac{R_1 + \dots + R_k}{n_1 + \dots + n_k} = \frac{\frac{n(n+1)}{2}}{n} = \frac{n+1}{2} \right).$$

Αρα ενδείξεις οτι η Ho αληθευει εχουμε αν $\frac{R_i}{n_i}$ ειναι περίπου ισο με $\frac{n+1}{2}$ ή εναλλακτικά το

$$\sum_{i=1}^{k} \left(\frac{R_i}{n_i} - \frac{n+1}{2} \right)^2$$ να ειναι κοντά στο μηδέν.

Το τεστ των Kruskal-Wallis (K-W) για τον έλεγχο της Ho γινεται με το εναλλακτικο

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

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

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

Άρα απορρ. την Ho για μεγάλες τιμές του K

$$K \overset{\text{προσ.}}{\sim} \chi^2_{k-1}, \quad K \geq \chi^2_{\alpha, k-1}$$

όπου $\sum R_i = \frac{n(n+1)}{2}$

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

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

Ισοτιμίες : $K^* = K / c$, $c = 1 - \sum_{i=1}^{r} \frac{t_i (t_i^2 - 1)}{n(n^2-1)}$

r σετ ισοτιμιών, t_i αριθμ. ισοτ. του σετ. i.

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

$p = P(K \geq x \mid Ho)$; $p \leq a$

$K \geq \chi^2_{\alpha, n-1}$ (αυτό το χρησ. όταν ο δεν μας καλύπτει ο πίνακας).

Παράδειγμα 3 (5.1, σελ. 89 - Μπατσίδης)

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

$$K = \frac{12}{4.5} \sum_{1}^{3} \frac{R_i^2}{n_i} - 3.5$$

Δλδ. $\binom{4}{2\,1\,1} = 12$ δυνατές περιπτώσεις με $x_{11}, x_{12}, x_{21}, x_{31}$
$x \quad x \quad y \quad z$

Δείγματα: xxyz, xyxz, xyzx, yxzx, yzxx, yxxz, xxzy, xzxy, xzyx
Τάξεις: 1234, 1234, 1234, 1234
(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)
K : 2,7 1,8 0,3 1,8 2,7 2,7 2,7 1,8 0,3

zxyx, zyxx, zxxy
(6, 3, 1), (7, 2, 1), (5, 4, 1)
1,8 2,7 2,7

Άρα
$$P(K = x) = \begin{cases} 6/12 = 0.5 & , x = 2.7 \\ 4/12 & , x = 1.8 \\ 2/12 & , x = 0.3 \end{cases}$$

και
$P(K \geq 2.7) = 0.5$.

Παράδειγμα 4 (7.12)

Έχει λάθη τα σωστά είναι:
$R_3 = 5 + 7 + 8 = 20$

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

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

Παράδειγμα 5 (5.2 σελ. 90 - Μπατσίδης)

$k=2$, W-M-W και K-W ταυτίζονται (ασυμπτωτικά ως χ^2_1)

από K-W: $$k = \frac{12}{n(n+1)} \sum_{1}^{2} n_i \left(\frac{R_i}{n_i} - \frac{n+1}{2} \right)^2$$

και $$Z^2 = \left(\frac{R_1 - E R_1}{Var R_1} \right)^2$$ των W-M-W

$$R_1 + R_2 = \frac{n(n+1)}{2}, \quad n = n_1 + n_2$$

$$E R_1 = \frac{n_1(n+1)}{2}$$

$$Var R_1 = \frac{n_1 n_2 (n+1)}{12}$$

Παράδειγμα 1

3 μηχανές χρησιμοποιούνται για την παραγωγή δοχείων αναψυκτικών. Κατά την διάρκεια μιας βδομάδας καστ. ο αριθμός δοχ. που ακολουθεί για τη M_1, M_2, M_3. Θέλουμε να ελέγξουμε αν υπάρχ. στατ. σημ. διαφ. ως προς τον αριθμό δοχείων των μηχανών.

M_1	340	345	330	342
M_2	339	333	344	338
M_3	347	343	349	355

αρ. δοχείων σε μια εβδομάδα (4 μέρες) από 3 μηχανές

Λύση

$H_0: \mu_1 = \mu_2 = \mu_3$ · H_a: όχι όλα τα μ_i ίσα μεταξύ των

$\alpha = 0.05$

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

Δείγμα: 1 2 2 2 1 1 3 2 1 3 3 3

Ενιαίο διατεταγμένο 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

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

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

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

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

$$- 3(12+1)$$

$\rightsquigarrow K = 5.8077$

$p = P(K \geq 5.8077) < 0.049 < \alpha (= 0.05)$

άρα απορρ. H_0

Παράδειγμα 2

Θέλουμε να ελέγξουμε αν οι φωτοπεριοδικά συνθήκα επηρεάζουν τον χρόνο που χρειάζεται να τραφεί κάθε τσιμπούρι.

Χρόνος σε μέρες που χρειάζεται να τραφεί κάθε τσιμπούρι	Δείγματα 1	2	3	4
	Φωτοπεριοδικά συνθήκα (ώρες φωτός/μέρα)			
	6h	12h	18h	24h
	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

1-

H_0: Τα 4 δείγματα προέρχονται από τον ίδιο πληθυσμό. (δηλ. ότι οι φωτοπεριοδικά συνθήκα δεν επηρεάζουν).

Δείγμα:	1	2	3	2	4	1	2	1	3	4	4	4
Σύνολο Δεδομένων:	7.1,	8.6,	8.7,	9,	9.1,	10.7,	11	11.1 (8),	11.1 (9),	11.1 (10),	11.5,	11.7
Τάξη:	1	2	3	4	5	6	7	9	9	9	11	12

(*) →

3	3	4	2	4	3	1	3	3	1	1	2
12 (13),	12 (14),	12 (15),	12.6,	12.7,	13.2,	13.4,	13.9,	14.1,	14.3 (22),	14.3 (23),	14.3 (24)
14	14	14	16	17	18	19	20	21	23	23	23

4 2

14.5, 14.8.

$K_2 = \ldots$

$C = \ldots$

$n_1 = 6, \; n_2 = 6$

$n_3 = 7, \; n_4 = 7$

$n = 26$

$$k^* = \frac{k}{c} \qquad k^* \sim \chi^2_3$$

$$k^* > \chi^2_{0.05,3} = 7.815$$

$R_1 = 1 + 6 + 9 + 19 + 23 + 23 = 81$

$R_2 = 78$

$R_3 = 99$

$R_4 = 93$

$$K = \frac{12}{26(26+1)} \left(\frac{81^2}{6} + \frac{78^2}{6} + \frac{99^2}{7} + \frac{93^2}{7} \right) - 3(26+1) =$$

$$= 0.081.$$

$$C = 1 - \frac{1}{26(26^2-1)} \left(3(3^2-1) + 3(3^2-1) + 3(3^2-1) \right) = 0.996$$

όπου 3 ότι, $f_1 = 3, \; f_2 = 3, \; f_3 = 3$ (*)

→ διόρθωση με ισοτιμίες

και $k^* = \frac{k}{c} = \frac{0.081}{0.996} = 0.0813$

Επειδή $0.0813 < 7.815$ δεν απορρ. η H_0