Μαθημα 4:

13/3/19

Το προσημικό τεστ των τάξεων του Wilcoxon

Εστω $x_1, \dots, x_n$ τ.δ. απο συνεχή και συμμετρικό πληθυσμό με α.σ.κ $F(x)$. Ενδιαφέρει ο έλεγχος $H_0: \mu = \mu_0$ (ή $H_0: m = m_0$ αν δεν ειναι συμμετρικό).

Το τεστ του Wilcoxon στηρίζεται στις τάξεις των διαφορών $D_i = x_i - \mu_0$ (ή $D_i = x_i - m_0$), όπου η μέτρηση $x_i = \mu_0$ διαγράφεται

Εστω $R_1, \dots, R_n$ οι τάξεις των δειγματικών τιμών $x_1, \dots, x_n$

R_i = αριθμός των $x_j \leq x_i$, $i = 1, \dots, n$

$$= \sum_{j=1}^{n} I(x_j \leq x_i)$$

Στα δεδομένα μας δεν υπάρχουν ισοτιμίες - δεσμοί (ties)

Πρόταση: Αν $R_1, \dots, R_n$ είναι οι τάξεις των $x_1, \dots, x_n$ από συνεχή ασκ. Τότε:

(i) $R_1, \dots, R_n$ είναι δείγμα από την διακριτή κατανομή

(ii) $E(R_i) = \frac{n+1}{2}$, $Var(R_i) = \frac{n^2-1}{12}$, $Cov(R_i, R_j) = -\frac{n+1}{12}$, $i \neq j$

απόδειξη: (δεν μας ενδιαφέρει να τα θυμόμαστε)

(i) Οι δυνατές τιμές των R_i είναι από το σύνολο $\{1, 2, \dots, n\}$ με $p = P(R_i = j) = \frac{1}{n}$, $j = 1, \dots, n$

(ii) $E(R_i) = \sum_{j=1}^{n} j P(R_i = j) = \sum_{j=1}^{n} j \frac{1}{n} = \frac{1}{n} \sum_{j=1}^{n} j = \frac{1}{n} \frac{n(n+1)}{2} =$

$= \frac{n+1}{2}$

$E(R_i^2) = \sum_{j=1}^{n} j^2 P(R_i = j) = \sum_{j=1}^{n} j^2 \frac{1}{n} = \frac{(n+1)(2n+1)}{6}$

Άρα $Var(R_i) = \frac{n^2-1}{12}$ ■

Μεθοδολογία:

$D_i = X_i - \mu_0$ (ή $D_i = X_i - m_0$) κ $H_0 : \mu = \mu_0$ ($H_0 : m = m_0$)

Διατάσσουμε τις απόλυτες τιμές κατά αύξουσα τάξη μεγέθους $|D_1|, \dots, |D_n|$ και έστω $R(|D_i|)$ οι τάξεις των απόλυτων τιμών των διαφορών.

T^+ = αθροισμα των θετικών διαφορών =

$= \sum_{i=1}^{n} I(D_i > 0) R(|D_i|)$

αθροίσμε των $D_i > 0$ όπου $I = \begin{cases} 1, & \text{όταν ισχύει} \\ 0, & \text{όταν δεν ισχύει} \end{cases}$

$T^- =$ -//- αρνητικών διαφορών $= \sum_{i=1}^{n} I(D_i < 0) R(|D_i|)$

$$T^+ + T^- = 1 + 2 + \dots + n = \frac{n(n+1)}{2}$$

Ο έλεγχος γίνεται με το στατιστικό:

$$T = \min\{T^+, T^-\} = \min\left\{T^+, \frac{n(n+1)}{2} - T^+\right\}$$

με κρ. περιοχή $T \le T_{n,a}$ (μονόπλευρος έλεγχος)

$T \le T_{n,\frac{a}{2}}$ (δίπλευρος —"—)

Παράδειγμα 1 (3.3, σ. 56 Μπατσίδης)

$n = 4$, $T^+ \overset{H_0}{\sim}$; $D_i = X_i - \mu_0$ (D_1, D_2, D_3, D_4).

$2^4 = 16$ δυνατές περιπτώσεις

τάξεις D_i: 1 2 3 4

άθροισμα θετικών τάξεων

$\underset{0}{----}$, $\underset{2}{-+--}$, $\underset{3}{--+-}$, $\underset{4}{---+}$, $\underset{5}{-++-}$, $\underset{6}{-+-+}$, $\underset{7}{--++}$,

$\underset{9}{-+++}$, $\underset{1}{+---}$, $\underset{3}{++--}$, $\underset{4}{+-+-}$, $\underset{5}{+--+}$, $\underset{6}{+++-}$, $\underset{7}{++-+}$,

$\underset{8}{+-++}$, $\underset{10}{++++}$.

$$P(T^+ = t) = \begin{cases} 1/16, & t = 0, 1, 2, 8, 9, 10 \\ 2/16, & t = 3, 4, 5, 6, 7 \end{cases}$$

$$\underbrace{0 \quad 1 \quad 2}_{1/16} \quad \underbrace{3 \quad 4 \quad 5 \quad 6 \quad 7}_{2/16} \quad \underbrace{8 \quad 9 \quad 10}_{1/16}$$

Παράδειγμα 2 (3.2, σ. 55 Μπατσίδης)

126, 142, 156, 228, 245, 246, 370, 419, 433, 454, 478, 503

$n = 12$

$H_0: m = 220$ σελίδες vs $H_a: m \neq 220$ $(a = 5\%)$

$D_i = X_i - m_i = X_i - 220$:

$-94, -78, -64, 8, 25, 26, 150, 199, 213, 234, 258, 283$

προέρχεται από αρνητικό

$|D_i|$: $8, 25, 26, 64, 78, 94, 150, 199, 213, 234, 258, 283$

$R(|D_i|)$: $1, 2, 3, \underbrace{4, 5, 6}, 7, 8, 9, 10, 11, 12$

$T^- = 4 + 5 + 6 = 15$

$$T = \min\left\{T^-, \overbrace{\frac{n(n+1)}{2} - T^-}^{T^+}\right\} = \min\left\{15, \overbrace{\frac{12 \times 13}{2} - 15}^{63}\right\} = 15$$

$$T = \min\{T^-, T^+\} = 15 \not< T_{12, 0.025} = 14$$

δεν απορρ. η H_0.

$$T^+ \text{ προσ. } N\left(E(T^+) = \frac{n(n+1)}{4}, Var(T^+) = \frac{n(n+1)(2n+1)}{24} - \frac{\sum_{i=1}^{k}(t_i^3 - t_i)}{48}\right)$$ (k σετ ισοτιμιών)

$n \geq 20$

$$W = \frac{T^+ - E(T^+) \pm 0.5}{\sqrt{Var(T^+)}}$$, κρ. περιοχή $|w| \geq z_{\frac{a}{2}}$

Παράδειγμα 3 (7.4) Παλμοί καρδιάς.

$H_0: \mu = 160$ v $H_a: \mu > 160$

$D_i = x_i - \mu_0$: 3, 5, 29, 1, 11, -2, 9, 2, 3, -21, 12, 5, -12, 6, 12, 3, 27, 10 / $n = 18$

$|D_i|$: 1, 2, 2, 3, 3, 3, 5, 5, 6, 9, 10, 11, 12, 12, 12, 21, 27, 29.

$R(|D_i|)$: 1, 2.5, 2.5, 5, 5, 5, 7.5, 7.5, 9, 10, 11, 12, 14, 14, 14, 16, 17, 18.

$\frac{2+3}{2} = 2.5$

(κ=4)

$t_1 = 2$, $t_2 = 3$, $t_3 = 2$, $t_4 = 3$

$T^- = 2.5 + 14 + 16 = 32,5$

$T = \min\left\{T^-, \frac{n(n+1)}{2} - T^-\right\} = \min\{32.5, 138.5\} = 32,5$

$T = 32,5 < T_{18, 0.05} = 47$ άρα απορρ. η H_0.

Το τεστ των ροών

α, β ← ονομαστικά δεδομένα
+, - ← διατάξιμα δεδομένα

Έστω n_1 της κατηγορίας + } συνολικά $n = n_1 + n_2$ δεδομένα
n_2 της κατηγορίας -

Η H_0: Η διαδικασία που προκαλεσε την ακολουθία (των α, β ή +, -) είναι τυχαία ελέγχεται με το στατιστικό

R = αριθμός των ακολουθιών ομοίων συμβόλων στην ακολουθία των n το πλήθος + κ −

$= 1 + \sum_{i=2}^{n} I_i$, I δεικτρια συνάρτηση που είναι 1 αν το i-οστό σύμβολο είναι διαφορετικό απο το $(i-1)$-οστό (Bernoulli, $p = P(ab \text{ ή } ba) =$

$$= \frac{n_1}{n}\frac{n_2}{n-1} + \frac{n_2}{n}\frac{n_1}{n-1} = \frac{2n_1 n_2}{n(n-1)}\Big).$$

(σελ. 39 - Μπατσίδη)

Προφανώς, πολύ μικρές ή πολύ μεγάλες τιμές του R υποδεικνύουν απόκλιση απο την τυχαιότητα

π.χ.

οι ακολουθίες: $\underbrace{+++++++}_{n_1=7}\ \underbrace{-----}_{n_2=5}$

$n=12,\ n_1=7,\ n_2=5,\ R=2$

: $+-+-+-+-+-+-+$:

$n=13,\ n_1=7,\ n_2=6,\ R=13$

Παράδειγμα 1

Να ελεγχθεί κατα πόσο η ακολουθία της σειράς αποτελείται απο 8 αγόρ. κ 12 κορίτσια είναι τυχαία όπου:

AAAKKKKAAKKKKKAAAKKK

H_0: η σειρά είναι τυχαία

H_a: παιδιά του ίδιου φύλου τείνουν να κάθονται μαζί.

$n=20,\ n_1=8,\ n_2=12,\ R=6.$

απορρ. των H_0 για μικρές τιμές του R

δλδ. $R \leq R_{n_1, n_2, a} = R_{8,12,0.05} = 7$

απορ. H_0.

παράδειγμα 2: (2.1, σ.41 - Μπατσίδη)

Το τεστ. των ροών εκτός ύλης