15/3/2017



# Το προσημικό Τεστ

Έστω $X_1, X_2, \ldots, X_n$ τυχαίο δείγμα από κάποιο πληθυσμό. μας ενδιαφέρει η συμπερασματολογία για το $\mu$.

Αν ο πληθυσμός είναι συνεχής και συμμετρικός τότε

$P(X>\mu) = P(X<\mu) = 1/2$, $P(X=\mu)=0$ (*)

Ελέγχουμε την $H_0: \mu=\mu_0$ vs $H_a: \mu\neq\mu_0$
$\mu<\mu_0$
$\mu>\mu_0$

Αντικαθιστούμε κάθε τιμή μεγαλύτερη του $\mu_0$ με + και κάθε μικρότερη με −. Έστω $X$ = πλήθος των + στο δείγμα μεγέθους $n$ (τιμή ίση με $\mu_0$ δε λαμβάνονται υπόψιν προσαρμόζοντας και το $n$ ανάλογα)

Αν αληθεύει η $H_0$, $X \sim B(n, p=\frac{1}{2})$. Άρα, ο έλεγχος μεταφέρεται ισοδύναμα στον $H_0: p=\frac{1}{2}$ vs $H_a$: $p\neq\frac{1}{2}$
$p<1/2$
$p>1/2$

και κρίσιμες περιοχές μεγέθους $a$ → $X\leq k_{a/2}'$ ή $X\geq k_{a/2}$
$X\leq k_{a/2}'$
$X\geq k_a$

όπου $k_a'$ και $k_a$ οι μεγαλύτερος και μικρότερος ακέραιος αντίστοιχα με $\sum^{k_a'} \binom{n}{x}\left(\frac{1}{2}\right)^n \leq a$ και $\sum^{n} \binom{n}{x}\left(\frac{1}{2}\right)^n \leq a$

Εναλλακτικά, $X \sim B(n, \frac{1}{2}) \Rightarrow X \overset{\text{προσ.}}{\sim} N(\frac{n}{2}, \frac{n}{4})$

Έτσι χρησιμοποιούμε το στατιστικό: $Z = \begin{cases} \dfrac{X - \frac{n}{2} - \frac{1}{2}}{\sqrt{n}/2}, & X > \frac{n}{2} \\ \dfrac{X - \frac{n}{2} + \frac{1}{2}}{\sqrt{n}/2}, & X < \frac{n}{2} \end{cases}$

και κρίσιμη περιοχή μεγέθους α.

Παράδειγμα 1 (7.4)

163 (+), 165 (+), 160, 189 (+), 161 (+), 171 (+), 158 (−), 169 (+), 162 (+), 163 (+), 139 (−), 172 (+), 165 (+), 148 (−),
166 (+), 172 (+), 163 (+), 187 (+), 170 (+)

$H_0: \mu = 160$ vs $H_a: \mu > 160$ $(\alpha = 0{,}05)$

$X = 15$

$n = 18$ (διαγράψαμε μια τιμή)

ΚΠ: $Z \geq z_\alpha (= z_{0,05} = 1{,}645)$

ΣΣΤ $Z = \dfrac{X - \frac{n}{2} - \frac{1}{2}}{\sqrt{n}/2} = \dfrac{15 - \frac{18}{2} - \frac{1}{2}}{\sqrt{18}/2} = 2{,}59$

Άρα $2{,}59 > 1{,}645 \Rightarrow$ Απορρ. η $H_0$.

## Το προσημικό τεστ των τάξεων του Wilcoxon

Για το τυχαίο δείγμα έστω $R_1, \dots, R_n$ οι τάξεις των δειγματικών τιμών. Δηλαδή, $R_i =$ αριθμός των $X_j$ με $X_j \le X_i$, $i = 1, 2, \dots, n$ $= \sum_{j=1}^{n} I(X_j \le X_i)$, $i = 1, 2, \dots, n$

Υποθέτουμε ότι στο τ.δ. $X_1, X_2, \dots, X_n$ δεν υπάρχουν δεσμοί, δηλαδή $X_i \neq X_j$ για $i \neq j$

Πρόταση: Αν $R_1, \dots, R_n$ είναι οι τάξεις των $X_1, X_2, \dots, X_n$ από συνεχή αθροιστική συνάρτηση κατανομής, τότε:

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

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

(i) Αφού δεν υπάρχουν δεσμοί προκύπτει ότι οι δυνατές τιμές των τάξεων $R_i$ ανήκουν στο σύνολο $\{1, 2, \dots, n\}$ με $P(R_i = j) = \frac{1}{n}$, $i = 1, 2, \dots, n$

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

$$E(R_i^2) = \sum_{j=1}^{n} j^2 P(R_i = j) = \sum_{j=1}^{n} j^2 \cdot \frac{1}{n} = \frac{n(n+1)\cdot(2n+1)}{6} \cdot \frac{1}{n} = \dots = \frac{n^2-1}{12}$$

• Προσημικό Τεστ Wilcoxon:

Έστω $X_1, X_2, \dots, X_n$ τ.δ από συνεχή (κ' συμμετρικό) πληθυσμό. Μας ενδιαφέρει ο έλεγχος $H_0: \mu = \mu_0$ $(m = m_0)$

Υπολογίζουμε τις διαφορές: $D_i = X_i - \mu_0$ (διαγράφουμε τις $X_i = \mu_0$). Διατάσσουμε τις απόλυτες τιμές $|D_1|, \dots, |D_n|$ κατά αύξουσα τάξη και υπολογίζουμε τις τάξεις, έστω $R(|D_i|)$, $i = 1, 2, \dots, n$. Αν δυο ή περισσότερες διαφορές είναι ίσες μεταξύ τους, τότε η τάξη καθεμιας είναι ο μέσος όρος των τάξεων που θα είχαν αν ήταν διαφορετικές μεταξύ τους. Αν αληθεύει η $H_0$ τότε το άθροισμα των τάξεων των θετικών διαφορών, έστω $T^+$ αναμένεται να είναι περίπου ίσο με το άθροισμα των τάξεων των αρνητικών διαφορών, έστω $T^-$.

Είναι $T^+ = \sum_{i=1}^{n} I(D_i > 0) \cdot R(|D_i|)$ και $T^- = \sum_{i=1}^{n} I(D_i < 0) \cdot R(|D_i|)$

όπου $I(D_i < 0) = 1 - I(D_i > 0)$

$$T^+ + T^- = \sum_{i=1}^{n} R(|D_i|) = \frac{n(n+1)}{2}$$

Για τον έλεγχο της μηδενικής χρησιμοποιείται το στατιστικό $T = \min(T^+, T^-) = \min\left(T^+, \frac{n(n+1)}{2} - T^+\right)$

και κρίσιμη περιοχή $T \le T_{n,a}$ για μονόπλευρο και $T \le T_{n,a/2}$ για δίπλευρο έλεγχο

$$E(T^+) = \frac{n(n+1)}{4}$$

$$Var(T^+) = \frac{n(n+1)\cdot(2n+1)}{24}$$

Παράδειγμα 1 (3.3 Μπατσίδης)

Αν $n=4$, $T^+ \sim$ ; (χωρίς δεσμούς)

Έστω $X_1, X_2, X_3, X_4$

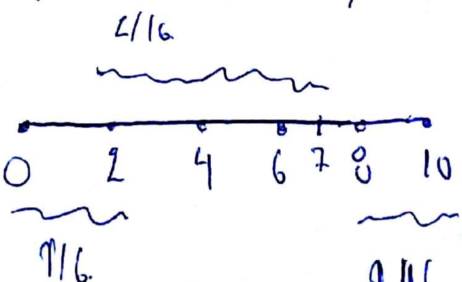
$$\left.\begin{aligned} D_1 &= X_1 - m_0 \\ D_2 &= X_2 - m_0 \\ D_3 &= X_3 - m_0 \\ D_4 &= X_4 - m_0 \end{aligned}\right\} \quad |D_1|, |D_2|, |D_3|, |D_4|$$

Διακρίνουμε περιπτώσεις: $2^4 = 16$ περιπτώσεις

Αν: $----$, $-+--$, $--+-$, $---+$, $-++-$, $-+-+$,
$--++$, $-+++$, $+---$, $++--$, $+-+-$, $+--+$,
$+++-$, $++-+$, $+-++$, $++++$

$t^+$: 0, 2, 3, 4, 5, 6, 7, | 1, 3, 4, 5, 6, 7, 8, 10

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



Παραδ. 2 (3.2 Μπατσίδης)

$n=12$: 126 142 156 228 245 246 370 419 433 454 478 503

$H_0: m = 220$ vs $H_a: m \neq 220$

$D_i: x_i - m_0 = x_i - 220$: $-94, -78, -64, 8, 25, 26, 150, 199, 213, 234, 258, 283$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | + | + | + | − | − | − | + | + | + | + | + | + |
| $\lvert D_i \rvert =$ | 8 | 25 | 26 | 64 | 78 | 94 | 150 | 199 | 213 | 234 | 258 | 283 |
| $R(\lvert D_i \rvert) =$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |

$$\left.\begin{array}{l} T^- = 4+5+6 = 15 \\ T = \min(T^+, T^-) = 15 \end{array}\right\} \quad T = \min\left(\frac{n(n+1)}{2} - T^-, T^-\right) = 15$$

① κ.π. $T \leq T_{n,\frac{a}{2}} = (T_{12,0{,}025})$

Όμως $T_{12,0{,}025} = 14$, $T = 15 \Rightarrow 15 > 14$

Άρα στατιστικά εκτός κρίσιμης περιοχής, δεν απορρίπτω την $H_0$.

Scanned by CamScanner

Παράδειγμα (3,74)

$H_0: \mu = 160$ vs $H_a: \mu > 160$  $n = 18$

$D_i: x_i - 160$ : 3, 5, 29, 1, 11, −2, 9, 2, 3, −21, 12, 5, −12, 6, 12, 3, 27, 10

$|D_i| =$ 1, $\bar{2}$, 2, 3, 3, 3, 5, 5, 6, 9, 10, 11, $\bar{12}$, 12, 12, $\bar{21}$, 27, 29

2, 3, 4, 5, 6, 7, 8, 13, 14, 15

$R(|D_i|) =$ 1, 9, 10, 11, 12, 16, 17, 18

= 1, 2.5, 2.5, 5, 5, 5, 7.5, 7.5, 9, 10, 11, 12, 14, 14, 14, 16, 17, 18

$T^- = 2{,}5 + 14 + 16 = 32{,}5$

$T^+ = \frac{n(n+1)}{2} - T^- = 138{,}5$

$T = 32{,}5$ ($\min(T^+, T^-) = T = 32{,}5$

$T_{0{,}05,18} = 47$

$T < T_{0{,}05,18}$

Άρα, απορρ. την $H_0$.