## Μάθημα 5ο 20/3/19

## Δύο Πληθυσμοί - Δύο δείγματα (συσχετισμένα): n-ζεύγη $(x_i, y_i)$

Συχνά τα δεδομένα αποτελούν ένα τυχαίο δείγμα μεγέθους n από συσχετισμένα ζεύγη μετρήσεων $(x_i, y_i)$. Δεδομένα αυτής της μορφής εμφανίζονται συνήθως όταν το ίδιο το αντικείμενο χρησιμοποιείται ως πειραματική μονάδα και ως μονάδα ελέγχου (control).

π.χ.
$x_i$: βάρος πριν από κάποια δίαιτα
$y_i$: μετά από κάποιο χρονικό διάστημα.

Σκοπός ο έλεγχος της αποτελεσματικότητας της δοκιμασίας δηλ. ο έλεγχος της υπόθεσης ότι τα x και y προέρχονται από τον ίδιο πληθυσμό (ίδια μέση τιμή - διάμεσο)

Χρησιμοποιούνται τις $d_i = x_i - y_i$

# 1. Το Προσημικό τεστ για συγκρίσεις κατά ζεύγη

$(x_1, y_1), \dots, (x_n, y_n)$ $\quad |d_i = x_i - y_i|$ διατάξιμα εντός των ζευγών.

Αν υποτεθεί ότι οι διαφορές προέρχονται από συνεχή κ συμμετρικό πληθυσμό, η $H_0 : \mu_D = \mu_x - \mu_x = 0$ είναι ισοδύναμη με την

$$H_0 : P(d_i > 0) = P(d_i < 0) = \frac{1}{2} \; (= p)$$

ελέγχεται με το στατιστικό $X =$ πλήθος των θετικών διαφορών, γιατί $X \sim Bin(n, p = \frac{1}{2})$

κρ. περιοχή $\quad X \geq k_{\frac{a}{2}}$ ή $X \leq k'_{a/2}$

$X \geq k_a$

$X \leq k'_a$

με $k_a, k'_a$ ο μικρότερος κ μεγαλύτερος ακέραιος αντίστοιχα:

$$\sum_{x=k_a}^{n} \binom{n}{x} \left(\frac{1}{2}\right)^n \leq a \quad \text{και} \quad \sum_{x=0}^{k'_a} \binom{n}{x} \left(\frac{1}{2}\right)^n \leq a.$$

Προσεγγιστικά, $\quad Z = \dfrac{X - \frac{n}{2} \pm \frac{1}{2}}{\sqrt{n}/2}$

κρ. περιοχή: $\quad |Z| \geq Z_{a/2}$

$Z \geq Z_a$

$Z \leq -Z_a$

Παράδειγμα 1 (Ι.5)

$d_i = x_i - y_i$, $H_0: \mu_D = 0$ v $H_a: \mu_D > 0$.

$n = 10$, $d_i$: 0.03, 0.21, 1.17, 0.15, -0.02, -0.04, 0.52, 0.22, -0.01, 0.13.

$X = 7$ (πλήθος των +).

$B(n = 10, p = \frac{1}{2})$ (από τον πίνακα της διωνυμικής)

$P(x=0)$

$P(x=8) = 0.0439$

$P(x=9) = 0.0098$

$P(x=10) = 0.001$

$P(x \geq 9) = 0.0108$ (μικρότερο του 0.05)

$P(x \geq 8) = 0.0547$

$K_{0.005} = 9$

Άρα $X \geq K_{0.05} (= 9)$ δεν απορρ. η $H_0$.

$$Z = \frac{7 - \frac{10}{2} - \frac{1}{2}}{\sqrt{10}/2} = 0.95 < Z_{0.05} = 1,645$$ δεν απορ η $H_0$

## Το προσημικό τεστ του Wilcoxon

$(x_1, y_1), \ldots, (x_n, y_n)$

$|d_i = x_i - y_i|$ διατάξιμα εντός των ζευγών

A. υποτεθεί ότι η διαφορά προέρχεται από συνεχή, συμμ. πληθυσμό η $H_0: \mu_D = \mu_x - \mu_y = 0$ είναι ισοδύναμη με την $H_0: P(d_i > 0) = P(d_i < 0) = \frac{1}{2}$

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

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

$T \le T_{a/2, n}$

$T \le T_{a, n}$

$T = \min\{T^-, T^+\}$

### Παράδειγμα 2 (7.5)

$d_i = x_i - y_i$ , $H_0: \mu_D = 0$ , $H_a: \mu_D > 0$

$d_i$: 0.03, 0.21, 1.17, 0.15, −0.02, −0.04, 0.52, 0.22, −0.01, 0.13

$|d_i|$: 0.01, 0.02, 0.03, 0.04, 0.13, 0.15, 0.21, 0.22, 0.52, 1.17.

$R(|d_i|)$: 1 2 3 4 5 6 7 8 9 10.

$T^- = 1+2+4 = 7$

$T = \min\{T^-, \frac{n(n+1)}{2} - T^-\} = \min\{7, \underbrace{\frac{10 \cdot 11}{2} - 7}_{48}\} = 7$

κρ. περιοχή $T \leq T_{0.05, 10}$ $(=11)$

$7 < 11$

απορρ. $H_0$

## Παράδειγμα 1 (3.5, σελ. 63, Μπατσίδης)

$d_i$: 7, 5, 12, −3, −5, 2, 14, 18, 19, 21, −1 [$d_i$ = τιμή λ − τιμή μ]

$\alpha = 0.05$, $H_0: \mu_D = 0$ ν $H_a: \mu_D \neq 0$.

με το τεστ του Wilcoxon

$|d_i|$: 1, 2, 3, 5, 5, 7, 12, 14, 18, 19, 21

$R(|d_i|)$: 1, 2, 3, 4.5, 4.5, 6, 7, 8, 9, 10, 11.

$T^- = 1 + 3 + 4.5 = 8,5$

$T = \min\{T^-, T^+ = \frac{n(n+1)}{2} - T^- = \frac{11(11+1)}{2} - 8.5 = 57.5\}$

$= 8.5$.

κρ. περιοχή $T \leq T_{a/2, n}$ $(= T_{0.025, 11} = 11)$

άρα η $H_0$ απορρ. αφού $8.5 < 11$

ή

$$W = \frac{T^{+} - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum (t_i^3 - t_i)}{48}}} \quad \left(\overset{\text{προσ.}}{\underset{H_0}{\sim}} N(0,1)\right)$$

$t_1 = 2$

$$W = \frac{57.5 - \frac{11 \cdot 12}{4}}{\sqrt{\frac{11(11+1)(22+1)}{24} - \frac{(2^3-2)}{48}}} = 2.18$$

κρ. περιοχή $W \geq Z_{\frac{a}{2}} \left(= Z_{0.025} = 1.96\right)$

Επειδή $W = 2.18 > 1.96 \left(= Z_{0.025}\right)$ απορ. $H_0$