Μάθημα 6ο: 27/3/19

Δύο πληθυσμοί - Δύο Δείγματα Ανεξάρτητα

Το τεστ των Kolmogorov-Smirnov [redacted] (K-S)

Εστω τα ανεξάρτητα τυχαία δείγματα $x_1, .., x_m, y_1, .., y_n$ από δύο πληθυσμούς $F_X(x)$ κ $F_Y(y)$ αντίστοιχα.
Για τον έλεγχο της $H_0: F_X(x) = F_Y(y)$ v $H_a: F_X(x) \neq F_Y(y)$ το τεστ. K-S χρησιμοποιεί το στατιστικό

$$D_{m,n} = \sup_t |F_m(t) - F_n(t)| = \max_i |F_m(t_i) - F_n(t_i)|$$

όπου $F_m(t) = (\text{ο αριθμός των } x_i \le t)/m$

και $F_n(t) = (\text{ο αριθμός των } y_i \le t)/n$

οι ε.α.δ.κ των $F_X(x)$ και $F_Y(y)$ αντίστοιχα.

Οι ε.α.δ.κ. υπολογίζονται είτε με ομαδοποίηση των x και y σε κοινές κατηγορίες είτε με ανάμειξη των x_i και y_i σε ένα δείγμα και διάταξη του δείγματος κατά πραγματικό μέγεθος (βλέπε παράδειγμα).
Απορρίπτουμε την H_0, για επίπεδο σημαντικότητας α όταν $D_{m,n} \ge D_{m,n,\alpha}$ (πίνακες).

Παράδειγμα 1

Τυχαίο δείγμα $x_1, ..., x_9$ από έναν πληθ. και άλλο τ.δ. $y_1, ..., y_{12}$ από άλλο $\overset{\text{ανεξ.}}{\text{πληθ.}}$

Να ελεγχθεί $H_0: F_X(x) = F_Y(y)$

x_i : 7.6 8.4 8.6 8.7 9.3 9.9 10.1 10.6 11.2

y_i : 5.2 5.7 5.9 6.5 6.8 8.2 9.1 9.8 10.8 12.3 13.4 14.6

$F_m(t_i) = F_9(t)$: 0 0 0 0 0 1/9 1/9 2/9 3/9 4/9 4/9 5/9 5/9 6/9 7/9 8/9 8/9 1 1 1 1

$F_n(t_i) = F_{12}(t)$:

: 1/12 2/12 3/12 4/12 5/12 5/12 6/12 6/12 6/12 6/12 7/12 7/12 8/12 8/12 8/12 8/12 9/12 9/12 10/12 11/12 1

$F_9(t_i) - F_{12}(t_i)$:

: $-3/36$ $-6/36$ $-9/36$ $-12/36$ $\boxed{-15/36}$ $-11/$ $-14/$ $-10/$ $-4/$ $-2/$ $-5/$ $-1/$ $-2/$ 0 $4/36$ $2/$ $5/$ $9/$ $6/$ $3/$ 0

$$D_{9,12} = \max_i |F_9(t_i) - F_{12}(t_i)| = \frac{15}{36} = 0.4166$$

$$D_{9,12,\alpha} = \frac{20}{36} = 0.5556$$

$D_{9,12} < D_{9,12,\alpha}$ δεν απορρ. H_0

ή

προσεγγιστικά;

$$D_{9,12,0.05} \simeq 1.36 \sqrt{\frac{m+n}{mn}} = 1.36 \sqrt{\frac{9+12}{9 \cdot 12}} = 0.5997$$

Παράδειγμα 2 (7.8)

x_i : $-2, 1, 3, 4, 8, 9, 10$ $\quad n=7$

y_i : $-10, -7, -8, -4, -3, -1, 7, -3, 28, -3$ $\quad n=10$

H_0 : $F_X(x) = F_Y(Y)$

y	y	y	y	y	y	y	x	y	x	x	x	y	x	x	x	y
-10	-8	-7	-4	-3	-3	-3	-2	-1	1	3	4	7	8	9	10	28

$x \leftarrow F_7(t)$:

0 0 0 0 0 0 0 1/7 1/7 2/7 3/7 4/7 4/7 5/7 6/7 1 1

$y \leftarrow F_{10}(t)$:

1/10 2/10 3/10 4/10 5/10 6/10 7/10 7/10 8/10 8/10 8/10 8/10 9/10 9/10 9/10 9/10 1.

$F_7(t) - F_{10}(t)$:

$-7/10$

$$D_{7,10} = \max_i |F_7(t) - F_{10}(t)| = 0.7 > D_{7,10,\,0.05} = \frac{43}{70} = 0.61$$

άρα απορρ. η H_0.

$$D_{7,10,0.05} \approx 1.36\sqrt{\frac{7+10}{7\cdot 10}} = 0.67$$

Τον προσεγγιστικό τον χρησιμοποιούμε όταν δεν μας καλύπτει πίνακας

(μπορούμε να τα γράψουμε και τα δύο).

Παράδειγμα 3 (7.4)

Βαθμός: 1 2 3 4 5 6 7 8 9 10 Σύνολο

Μ1(συχνότητα): 0 0 4 8 13 10 4 0 1 0 40

Μ2(συχνότητα): 1 2 3 10 20 10 3 0 0 1 50

$H_0 : F_X(x) = F_Y(y)$ v $H_a : F_X(x) \neq F_Y(y)$

τιμές t : 1 2 3 4 5 6 7 8 9 10

$F_{40}(t)$: 0 0 $\overset{\overset{8+4}{}}{}$ 4/40 12/40 25/40 35/40 39/40 39/40 1 1

0.1 0.3 0.685 0.875 0.975 0.975 1 1

$F_{50}(t)$: 1/50 3/50

0.02 0.06 0.12 0.32 0.72 0.92 0.98 0.98 0.98 1

$|F_{40}(t) - F_{50}(t)|$: 0.02 0.06 0.02 0.02 $\boxed{0.095}$ 0.045 0.005 0.005 0.02 0.

$$D_{40,50} = \max_i |F_{40}(t) - F_{50}(t)| = 0.095$$

$$D_{40,50,0.05} \cong 1.36\sqrt{\frac{40+50}{40 \cdot 50}} = 0.29$$

$D_{40,50} < D_{40,50,0.05}$ άρα δεν απορρ. η H_0

Άρα η Μ1, Μ2 δεν διαφέρουν
έχουν την ίδια αποτελεσματικότητα