23/3/2017

Δυο πληθυσμοί - Δυο δείγματα (ανεξάρτητα)
- → χ^2 test (ομογενής, ανεξαρ. πίνακας συναφείας)
- → Kolmogorov-Smirnov
- → Wilcoxon-Mamm-Whitney.

Το Test των W-M-W:

Έστω τα ανεξάρτητα τ.δ. $x_{11},\dots x_{1n_1}$ & $x_{21},\dots,x_{2n_2}$ από τους πληθυσμούς x_1 & x_2 αντίστοιχα.

H_0: Οι τ.μ. X_1 & X_2 έχουν την ίδια κατανομή ν → $H_0: P(x_1>x_2)=1/2$

H_a: -II- -II- έχουν διαφορετική κατανομή → $\boxed{\begin{array}{l} H_a: P(x_1>x_2)\neq 1/2 \\ H_a: P(x_1>x_2)< 1/2 \\ H_a: P(x_1>x_2)> 1/2 \end{array}}$

- Έστω $R(x_{ij})$, $j=1,\dots,n$, $i=1,2,\dots,n$ οι τάξεις των μετρήσεων στο ενιαίο δείγμα $n=n_1+n_2$ παρατηρήσεων, και $R_i=\sum_{j=1}^{n_i}R(x_{ij})$, $i=1,2,\dots$ το άθροισμα των αντίστοιχων τάξεων.

Αν H_0 είναι αληθής, τότε θα περιμέναμε $(P(x_1>x_2)=1/2)$ οι μέσες τιμές να είναι ίσες δηλαδή $R_1/n_1=R_2/n_2$ & είναι $\sum_{i=1}^{n}R_i(=1+2+\dots+n)=\frac{n(n+1)}{2}$ άρα,

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

Πρόταση (σελ. 66 Μπατσίδης): Υπό την H_0, αν δεν υπάρχουν δεσμοί ισχύει:

(i) $R_1\xrightarrow[\text{προσ.}]{H_0}N(ER_1,VarR_1)$, αφού $R_1=\sum_{j=1}^{n_1}R(x_{ij})$

(ii) $ER_i=\frac{n_i(n+1)}{2}$ $\left(ER_1=\frac{n_1(n+1)}{2}\right)$

$VarR_i=\frac{n_i(n+1)(n-n_i)}{12}$ $\left(VarR_1=\frac{n_1n_2(n+1)}{12}\right)$

Απόδειξη:

$E(R_i)=E(\sum R(x_{ij}))=\sum E(R(x_{ij}))=\sum\sum e\frac{1}{n}=\frac{1}{\cancel{n}}\sum\frac{\cancel{n}(n+1)}{2}=\frac{n_i(n+1)}{2}$

τιμές $R(x_{ij})$: $1,2,\dots,n$.

πιθανότητες: $1/n$ $1/n,\dots$ $1/n$

Το test των W-M-W χρησιμοποιεί για τον έλεγχο H_0 το στατιστικό:

u_1 = αριθμός των φορών που μια x_1 παρατήρηση ακολουθεί μια x_2 παρατήρηση.

(ή άθροισμα αριθμού παρατηρήσεων x_1 που ακολουθούν-είναι μεγαλύτερες-από κάθε μέτρηση x_2 για το ενσωματωμένο διατεταγμένο δείγμα.)

Πχ:

	1	2	3	4	5	6	7	8	9	10	11
	x_2	x_2	x_1	x_1	x_2	x_1	x_2	x_1	x_2	x_2	x_1
από πόσες x_2 είναι μεγαλύτερη η x_1:			2	2		3		4			6
$R(x_{ij})$	1	2	3	4	5	6	7	8	9	10	11

$n = 11$, $n_1 = 5$, $n_2 = 6$

$u_1 = 2+2+3+4+6 = 17$

Ισοδύναμα: $U_1 = R_1 - \frac{n_1(n_1+1)}{2}$

$R_1 = 3+4+6+8+11 = 32$

$U_1 = 32 - \frac{5(5+1)}{2} = 17$

Ανάλογα, $U_2 = R_2 - \frac{n_2(n_2+1)}{2}$ & τελικά, χρησιμοποιούμε το $U = \min(U_1, U_2) =$

$= U_1 + U_2 = \underset{\sim}{R_1} - \frac{n_1(n_1+1)}{2} - \underset{\sim}{R_2} + \frac{n_2(n_2+1)}{2} =$

$= \frac{n(n+1)}{2} - \frac{n_1(n_1+1)}{2} - \frac{n_2(n_2+1)}{2} = n_1 n_2$. Άρα $U = \min(U_1, n_1 n_2 - U_1)$

και κρίσιμη περιοχή α, αν η $P(U \leq u) \leq \alpha$ (μονόπλευρο) ή $P(U \leq u) \leq \alpha/2$ για δ' έλεγχο

η τιμή που παρατηρήθηκε

Προσεγγιστικά: $z = \frac{R_1 - \frac{n_1(n+1)}{2}}{\sqrt{n_1 n_2 (n+1)/12}}$

Παράδειγμα 3 (4.1 Μπατσίδης): Να βρεθεί η κατανομή του R_1 υπό την H_0 όταν $n_1 = 2$, $n_2 = 3$ και δεν υπάρχουν δεσμοί.

Λύση

$x_{11}, x_{12} \to x$

$x_{21}, x_{22}, x_{23} \to y$

$\binom{5}{2\ 3} = 10$ πιθανές θέσεις που θα μπορούσαν να προκύψουν από μια δειγματοληψία

			R_1			R_1			
1. (x x y y y)	(1,2)	(3,4,5)	3	4. (y, x x y y)	(2,3)	5	7. (y x y y x)	(2,5)	7
2. (x y x y y)	(1,3)	(2,4,5)	4	5. (x y y y x)	(1,5)	6	8. (y y x x y)	(3,4)	7
3. (x y y x y)	(1,4)	(2,3,5)	5	6. (y x y x y)	(2,4)	6	9. (y y x y x)	(3,5)	8
							10. (y y y x x)	(4,5)	=9

Τιμες R_1: $\overset{1/10}{3}, \overset{1/10}{4}, \overbrace{5, 6, 7}^{3/10}, \overset{1/10}{8}, \overset{1/10}{9}$.

$$P(R_1 = r) = \begin{cases} 2/10, & r = 5,6,7 \\ 1/10, & r = 3,4,8,9 \end{cases}$$

Παραδ. 4 (4.3 Μπατσίδης)

Παραδειγμα 1 (7.9):

x_1: 9 11 15 $\qquad H_0: F(x_1) = F(x_2)$ v $H_a: F(x_1) \neq F(x_2)$.

x_2: 6 8 10 13

Λύση

Αναμειγμενο διατετ. δειγμα:	6	8	9	10	11	13	15	
	x_2	x_2	x_1	x_2	x_1	x_2	x_1	
Ταξη $R(x_i)$	1	2	3	4	5	6	7	

$n_1 = 3 \quad n_2 = 4 \quad n = 7$

$R_1 = \cancel{1} + 3 + 5 + 7 = 15$

$U_1 = R_1 - \frac{n_1(n_1+1)}{2} = 25 - 6 = 9$

$U = \min(U_1, n_1 n_2 - U_1) = \min(9, 3) = 3$

κρισιμη περιοχη: $P(U \leq 3) \leq \frac{\alpha}{2} \left(= \frac{0.05}{2} = 0.025 \right)$

$\hookrightarrow = 0.200 = p$

Επειδή $2 \cdot p = 0.4 > 0.05$ (για διπλευρο ελεγχο) δεν απορ. H_0.

Παραδειγμα 2: $H_0: F(x_1) = F(x_2)$ v $H_a: F(x_1) \neq F(x_2)$

x_1: 26, 27, 25, 31

x_2: 28, 35, 32, 29

Λύση

x_1	x_1	x_1	x_2	x_2	x_1	x_2	x_2
25	26	27	28	29	31	32	35
1	2	3	4	5	6	7	8

$R_1 = 1 + 2 + 3 + 6 = 12$, $U_1 = 12 - \frac{4 * 5}{2} = 2$

$U = \min(2, n_1 n_2 - 2) = \min(2, 14) = 2$

$P(U \leq 2) = 0.057 = p$

Είναι (επειδή διπλευρος ελεγχος) $2p = 2 \cdot 0.057 > 0.05$

δεν απορ. H_0.