10/9/2017

ΤΕΣΤ ΙΣΟΤΗΤΑΣ ΔΥΟ ΠΟΛΥΩΝΙΜΙΚΩΝ ΚΑΤΑΝΟΜΩΝ - ΤΕΣΤ ΟΜΟΓΕΝΕΙΑΣ

Θεωρούμε τα δείγματα από δύο ανεξάρτητους πληθυσμούς:

$(n_{11},\dots,n_{1k}) \sim M(n_{1\cdot}, p_{11},\dots,p_{1k})$, $n_{1\cdot} = \sum n_{1j}$, $\sum p_{1j} = 1$

$(n_{21},\dots,n_{2k}) \sim M(n_{2\cdot}, p_{21},\dots,p_{2k})$, $n_{2\cdot} = \sum n_{2j}$, $\sum p_{2j} = 1$

Ενδιαφέρει $H_0: p_{1j} = p_{2j} = p_j$, $j = 1,\dots,k$ (θεωρ. σελ. 15 Ζωγράφος).

Αν τα p_{ij} είναι γνωστά, $\displaystyle\sum_{j=1}^{k} \frac{(n_{1j} - \overbrace{n_{1\cdot} p_{1j}}^{e_{1j}})^2}{n_{1\cdot} p_{1j}} \overset{ασυμ.}{\sim} \chi^2_{k-1}$

και $\displaystyle\sum_{j=1}^{k} \frac{(n_{2j} - \overbrace{n_{2\cdot} p_{2j}}^{e_{2j}})^2}{n_{2\cdot} p_{2j}} \overset{ασυμ.}{\sim} \chi^2_{k-1}$

Τότε:

$$\sum_{i=1}^{2}\sum_{j=1}^{k} \frac{(n_{ij} - n_{i\cdot} p_{ij})^2}{n_{i\cdot} p_{ij}} \sim \chi^2_{2(k-1)} \text{ με κρίσιμη περιοχή } \chi^2 \geq \chi^2_{\alpha, 2(k-1)}.$$

Όταν τα p_{ij} άγνωστα και $H_0: p_{1j} = p_{2j} = p_j$, $j = 1,\dots,k$ οι εκτιμ. μεγ. πιθ. (ΕΜΠ)

για $\hat{p}_j = \dfrac{n_{1j} + n_{2j}}{n_{1\cdot} + n_{2\cdot}} = \dfrac{n_{\cdot j}}{n}$, $n = n_{1\cdot} + n_{2\cdot}$ και $e_{ij} = \dfrac{n_{i\cdot} * n_{\cdot j}}{n}$.

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{k} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \sim \chi^2_{2(k-1) - \underset{\text{εκτιμ. παρ.}}{s}} = \chi^2_{2(k-1)} \text{ με κρ.π. } \chi^2 \geq \chi^2_{\alpha, k-1}$$

$s = k-1$, γιατί $\sum p_j = 1$

Παρ. 1 (παρ. 1.6.1 σελ. 32 Ζωγρ.)

Παρ. 2:

	Ντουιμαντερ (I=1)	Σειρές (I=2)	Μόδα (I=3)	Σύνολο
Αγόρια (i=1)	$n_{11} = 50 / e_{11} = 33.3$	$n_{12} = 30 / e_{12} = 36.7$	$e_{13} = 30$ $n_{13} = 20$	$n_{1\cdot} = 100$
Κορίτσια (j=1)	$n_{21} = 50 / e_{21} = 66.7$	$n_{22} = 80 / e_{22} = 73.3$	$e_{23} = 60$ $n_{23} = 70$	$n_{2\cdot} = 200$
	$n_{\cdot 1} = 100$	$n_{\cdot 2} = 110$	$n_{\cdot 3} = 90$	$n = 300$

$$e_{ij} = \frac{n_{i\cdot} * n_{\cdot j}}{n}$$

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{3} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = \ldots = 19.39$$

$$\chi^2 \geq \chi^2_{0.05,2} = 5.991$$

Απορ. H_0, δηλαδή υπαρχει διαφορά αναμεσα σε αγ. και κοριτσια.

Παραδειγμα 3:

Οστα ανα φυλο και περιοδο

	Πρωτη(j=1)	Δευτερη(j=2)	Τριτη(j=3)	Συνολο
Ανδρας (i=1)	$n_{11}=162$	$n_{12}=180$	$n_{13}=210$	$n_{1\cdot}=552$
	$e_{11}=152.12$	$e_{12}=170.58$	$e_{13}=229.3$	
Γυναικα (i=2)	$n_{21}=110$	$n_{22}=125$	$n_{23}=200$	$n_{2\cdot}=435$
	$e_{21}=119.88$	$e_{22}=134.42$	$e_{23}=180.70$	
Συνολο	$n_{\cdot 1}=272$	$n_{\cdot 2}=305$	$n_{\cdot 3}=410$	$n=987$

Λυση

$H_0: p_{1j} = p_{2j} = p_{ij} = p_j$ με $\chi^2 \geq \chi^2_{0.05,2} = 5.991$

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{3} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

Απορ. H_0, αρα υπαρχει διαφορά.

X^2 TEST ΟΜΟΓΕΝΕΙΑΣ Γ ΑΝΕΞΑΡΤΗΤΩΝ ΠΟΛΥΩΝΥΜΙΚΩΝ ΚΑΤΑΝΟΜΩΝ.

Κατ./Δειγμα (j) (i)	1	2	...	j	...	k	
1	n_{11}	n_{12}	...	n_{1j}	...	n_{1k}	$n_{1\cdot}$
2	n_{21}	n_{22}	...	n_{2j}	...	n_{2k}	
⋮							
i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ik}	$n_{i\cdot}$
⋮							
r	n_{r1}	n_{r2}	...	n_{rj}	...	n_{rk}	$n_{r\cdot}$
	$n_{\cdot 1}$			$n_{\cdot j}$		$n_{\cdot k}$	n

n_{ij} = αριθμος παρατηρησεων του i-οστου δειγματος που ανηκει στην κατηγορια j.

p_{ij} = P(ενα μελος του δειγμ. να προερχ. απο τον $i^{οστο}$ πληθυσμό και να εχει το χαρακτ. j.

$n_{i\cdot}$ = πληθος παρατ. του $i^{οστού}$ δειγματ.

$n_{\cdot j}$ = αρ. παρατ. που εχουν το χαρακ. j.

n = Το συνολο των παρατηρησεων

Αν τα p_{ij} γνωστα, εστω $e_{ij}=n_{i\cdot} * p_{ij}$ οι αναμενομενες συχνοτητες. Άρα για το i-οστό δειγμα $i=1,\dots,r$:

$$\chi^2=\sum_{j=1}^{k}\frac{(n_{ij}-e_{ij})^2}{e_{ij}}=\sum_{j=1}^{k}\frac{(n_{ij}-n_i p_{ij})^2}{n_{i\cdot}p_{ij}}\sim\chi^2_{k-1}$$

και συνολικα για τα r ανεξαρτητα δειγματα:

$$\chi^2=\sum_{i=1}^{r}\sum_{j=1}^{k}\frac{(n_{ij}-e_{ij})^2}{e_{ij}}\sim\chi^2_{r(k-1)},\ e_{ij}=n_{i\cdot} * p_{ij} \text{ και } \chi^2\geq\chi^2_{a,r(k-1)} \text{ απορ.}$$

Για p_{ij} αγνωστα, οταν Ho αληθευει $p_{ij}=\dots=p_{rj}=p_j$ & ΕΜΠ για p_j

$$\hat{p}_j=\frac{n_{1j}+\dots n_{rj}}{n}=\frac{n_{\cdot j}}{n}\rightsquigarrow e_{ij}=n_{i\cdot} * \frac{n_{\cdot j}}{n}=\frac{n_{i\cdot} * n_{\cdot j}}{n}.$$

$$\chi^2=\sum_{i=1}^{r}\sum_{j=1}^{k}\frac{(n_{ij}-e_{ij})^2}{e_{ij}}\sim\chi^2_{r(k-1)},\ \chi^2\geq\chi^2_{a,(r-1)(k-1)}.$$

Παραδειγμα 1:

Κατηγ.(j)/Δειγμα(i)	Αποτελεσμα: Ευνοϊκό (j=1)	Μη Ευνοϊκό (j=2)	Συνολο
Θεραπεια I (i=1)	$n_{11}=70$ / 79.6	$n_{12}=30$ / 20.4	$n_{1\cdot}=100$
θεραπεια II (i=2)	$n_{21}=160$ / 159.2	$n_{22}=40$ / 40.8	$n_{2\cdot}=200$
θεραπεια III (i=3)	$n_{31}=168$ / 159.2	$n_{32}=32$ / 40.8	$n_{3\cdot}=200$
Συνολο	$n_{\cdot 1}=398$	$n_{\cdot 2}=102$	$n=500$

Λύση

$r=3,\ k=2.$

$H_0: p_{1j}=p_{2j}=p_j\quad j=1,2$ ⟶ Τα ποσοστά των ευν. αποτελεσματων είναι ισα και στις 3 θεραπειες.

$$e_{ij}=\frac{n_{i\cdot} * n_{\cdot j}}{n}$$

$$\chi^2=\sum_{i=1}^{3}\sum_{j=1}^{2}\frac{(n_{ij}-e_{ij})^2}{e_{ij}}=8.08.$$

$\chi^2>\chi^2_{a(3-1)(2-1)}\ (=5.991)$

Απορ. Ho, δηλ. οι θεραπειες διαφεραν ως προς την ποιότητά τους.