Μάθημα 10:

15/5/19

r×κ Πίνακας Συνάφειας

Μεταβλ. Χ (Α) \ Μεταβλ. Υ (Β)	B_1 ...	B_j ...	B_κ	
A_1	n_{11} ...	n_{1j} ...	$n_{1\kappa}$	$n_{1\cdot}$
⋮				
A_i	n_{i1} ...	n_{ij} ...	$n_{i\kappa}$	$n_{i\cdot}$
⋮				
A_r	n_{r1} ...	n_{rj} ...	$n_{r\kappa}$	$n_{r\cdot}$
	$n_{\cdot 1}$	$n_{\cdot j}$	$n_{\cdot \kappa}$	$n = n_{\cdot\cdot}$

όπου

$$n = n_{\cdot\cdot} = \sum_{i=1}^{r}\sum_{j=1}^{\kappa} n_{ij}$$

$$n_{i\cdot} = \sum_{j=1}^{\kappa} n_{ij}$$

$$n_{\cdot j} = \sum_{i=1}^{r} n_{ij}$$

$P_{ij} = P(A_i \cap B_j) = P($ταυτόχρονα κ τα 2 χαρακτηριστικά$)$

$P_{i\cdot} = P(A_i)$, $P_{\cdot j} = P(B_j)$

χ^2 τεστ Ανεξαρτησίας

H_0: Τα χαρ. Α και Β είναι ανεξ/τα v.
H_a: Τα χαρ. Α και Β δεν είναι ανεξ/τα

ή

$H_0: P(A_i \cap B_j) = P(A_i) \cdot P(B_j)$

ή

$H_0: P_{ij} = P_{i\cdot} \cdot P_{\cdot j}$ v. $H_a: P_{ij} \neq P_{i\cdot} \cdot P_{\cdot j}$ τουλ. για ένα (i, j)

Η από κοινού κατανομή των n_{ij} είναι πολυωνυμική με παράμετρο n και P_{ij}, $i=1,\dots,r$, $j=1,\dots,k$

με $\sum_{i=1}^{r}\sum_{j=1}^{k} P_{ij} = 1 = \sum_{i=1}^{r} P_{i\cdot} = \sum_{j=1}^{k} P_{\cdot j}$

Για τον έλεγχο της Ho, χρησιμοποιούμαι το στατιστικό

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

όπου e_{ij} = αναμενόμενες συχνότητες $= n\cdot P_{ij}$

και

$\sum_i\sum_j (n_{ij}-e_{ij}) = 0$ γιατί $\sum_i\sum_j n_{ij} - n\sum_i\sum_j P_{ij} = n - n\cdot 1 = 0$

δεν μας ενδιαφέρει

Όταν οι παράμετροι είναι άγνωστες:

$$\hat{P}_{i\cdot} = \frac{n_{i\cdot}}{n} \quad \text{και} \quad \hat{P}_{\cdot j} = \frac{n_{\cdot j}}{n}.$$

Όταν αληθεύει η Ho: $\hat{P}_{ij} = \hat{P}_{i\cdot}\cdot\hat{P}_{\cdot j} = \frac{n_{i\cdot}}{n}\cdot\frac{n_{\cdot j}}{n}$

και $e_{ij} = \cancel{n}\,\frac{n_{i\cdot}\cdot n_{\cdot j}}{n^{\cancel{2}}} = \frac{n_{i\cdot}\cdot n_{\cdot j}}{n}$

και $\chi^2 \sim \chi^2_{rk-1-(r-1)-(k-1)} \equiv \chi^2_{(r-1)(k-1)}$

γιατί $\sum_{i=1}^{r} P_{i\cdot} = 1$ και $\sum_{j=1}^{k} P_{\cdot j} = 1$

Κρίσιμη περιοχή: $\chi^2 \geq \chi^2_{a,(r-1)(k-1)}$

$$\chi^2 = \frac{n(|n_{11}n_{22} - n_{12}n_{21}| - n/2)^2}{n_1 . n_2 . n_{.1}\, n_{.2}} \sim \chi^2_1$$ 2×2 πίνακ

Παράδειγμα 1 (r.r)

Σχέση Εργασίας \ Σχέδιο Συνταξιοδότησης	1	2	3	
Μισθωτοί	160 (136)	n_{12} = 140 e_{12} = (136)	40 (68)	340 = $n_{1.}$
Ωρομίσθιοι	40 (64)	60 (64)	n_{23} = 60 (32)	160 = $n_{2.}$
	200 = $n_{.1}$	200 = $n_{.2}$	100 = $n_{.3}$	500 = n

$H_0: P_{ij} = P_{i\cdot} \cdot P_{\cdot j}$ (ανεξ.)

$\hat{P}_{i\cdot} = \frac{n_{i\cdot}}{n}$, $\hat{P}_{\cdot j} = \frac{n_{\cdot j}}{n}$, $e_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$

λύση

$\frac{340 \cdot 200}{500} = 136$ πόσοι αναμένουμε στο 1ο κελί

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{3} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \; (= 49{,}632)$$

κρ. περιοχή $\chi^2 \geq \chi^2_{0.05,(r-1)(k-1)} \; (= \chi^2_{0.05,2} = 5{,}991)$

απορρ. η H_0.

Παράδειγμα 2 (Παραδ. 2.2.1, Ζωγράφος, σ.56)

Παράδειγμα 3 (Μογγολισμού κ. Λοιμώδους Ηπατίτιδα)

Μητέρα (μεταβλ. 1)	Παιδί (μεταβλ. 2) (Ασθένεια) (+) (j=1) Μογγολοειδή	(−) Μη Μογγολοειδές	Σύνολο
Με λοιμώδη (+) Ηπατίτιδα (i=1)	$n_{11}=65$ P_{11} ($e_{11}=5.15$)	$n_{12}=76$ P_{12} ($e_{12}=135,85$)	$n_{1.}=141$
Χωρίς λοιμώδη (−) Ηπατίτιδα (i=2)	$n_{21}=8$ P_{21} ($e_{21}=67.85$)	$n_{22}=1851$ P_{22} ($e_{22}=1791,15$)	$n_{2.}=1859$
Σύνολο	$n_{.1}=73$	$n_{.2}=1927$	$n=2000$

Παράγοντες

Μια ομάδα γιατρών μιας τεαν/κής κλινικής ισχυρίζεται ότι ο μογγολισμός στα παιδιά έχει σχέση με λοιμώδη Ηπατίτιδα κατά των

Για τον έλεγχο του ισχυρισμού αυτού, 2000 περιπτώσεις εκλέγονται τυχαία από μια μεγάλη μαιευτική κλινική και εξετάζονται ως προς την ηπατίτιδα της μητέρας. Μετά, και για κάθε περίπτωση, εξετάζεται αφού ευτομοτά το παιδί αν είναι μογγολικό ή όχι.
Τα αποτελέσματα δίνονται στον παραπάνω πίνακα 2×2.

Λύση

Ho: Οι δύο μεταβλητές ανεξ/τες (λοιμώδη ηπατίτιδα της μητέρας και μογγολισμός ανεξ/τες μεταβλ.)

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(n_{ij}-e_{ij})^2}{e_{ij}} = 776,7 \quad (\sim \chi^2_1)$$

διόρθωση 764,33

, $\chi^2 > \chi^2_{0.05, 1} = 3.841$, απορρ. H_0

Μέτρα Συνάφειας για 2×2 πίνακες (Σωμο άφου §2.2.5 σ. 41)

Παράγοντες: Κάπνισμα , Ασθένεια: Καρκ. πνεύμονα.

Ο λόγος σχετικών πιθανοτήτων (odds) ή λόγος διαγώνιων γινομένων είναι το πιο γνωστό απο τα μέτρα συσχέτισης για 2×2 πίνακα συνάφειας. Ορίζεται ως $\theta = \dfrac{p_{11} p_{22}}{p_{12} p_{21}}$

και εκτιμάται απο την $\hat{\theta} = \dfrac{n_{11} n_{22}}{n_{12} n_{21}}$

(Τα odds ενός ενδεχ. Ε είναι το πηλίκο $\dfrac{P(E)}{1-P(E)}$)

σχετική πιθανότητα.

$\hat{\theta} = 197,88$ για το παράδειγμα 3.

$\theta = 1$: ανεξ/τες

$\theta > 1$: θετικά συσχ.

$\theta < 1$: αρνητικά συσχ.