



ΑΡΙΣΤΟΤΕΛΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΟΝΙΚΗΣ

Συσχέτιση μεταξύ κατηγορικών (ποιοτικών) μεταβλητών

Κωνσταντίνος Ι. Μπουγιούκας, PhD



ΘΕΣΣΑΛΟΝΙΚΗ

Η δοκιμασία χ^2 (chi-square test)

- Σύγκριση κατηγορικών (ποιοτικών) μεταβλητών.

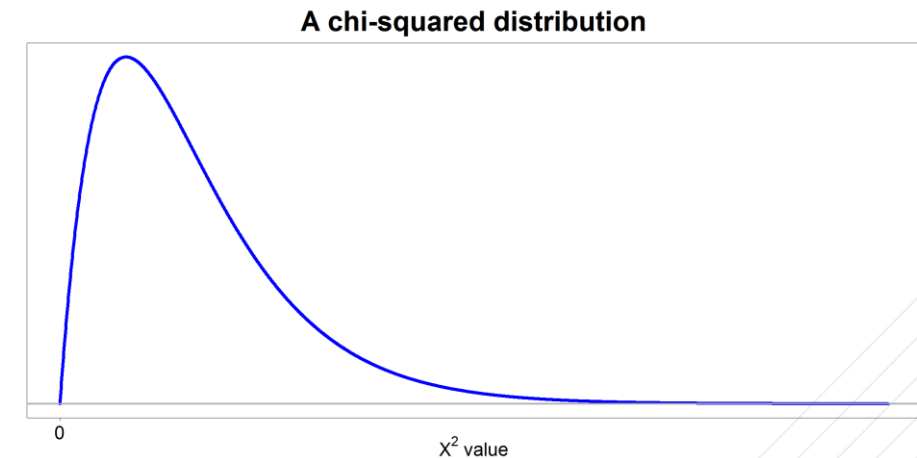
H0: Δεν υπάρχει συσχέτιση μεταξύ των δύο μεταβλητών.

H1: Υπάρχει συσχέτιση μεταξύ των δύο μεταβλητών.

- Στατιστικός τύπος:
$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Μεγάλη τιμή στο $\chi^2 \Rightarrow$ στατιστικά σημαντική συσχέτιση.

Δηλαδή, εάν $p < \alpha \Rightarrow$ **απορρίπτεται** η H0.



- **Προϋπόθεση:** Αναμενόμενες συχνότητες > 5 σε αντίθετη περίπτωση χρησιμοποιείται η ακριβής δοκιμασία κατά Fisher (Fisher's exact test)

Πίνακας συνάφειας (πίνακας 2x2)

Συσχέτιση της μεταβλητής 1 (Επίπεδο 1, Επίπεδο 2) και της μεταβλητής 2 (Επίπεδο 1, Επίπεδο 2)

Μεταβλητή 1	Μεταβλητή 2		ΣΥΝΟΛΟ
	Επίπεδο 1	Επίπεδο 2	
Επίπεδο 1	a	b	$a+b$
Επίπεδο 2	c	d	$c+d$
ΣΥΝΟΛΟ	$a+c$	$b+d$	n

Παράδειγμα: Σχέση πρωρότητας και φύλου (πίνακας 2x2)

- Παρατηρούμενες συχνότητες (Observed frequencies)

ΦΥΛΟ	ΠΡΩΡΟΤΗΤΑ		ΣΥΝΟΛΟ
	ΝΑΙ	ΟΧΙ	
ΑΓΟΡΙΑ	33	49	82
ΚΟΡΙΤΣΙΑ	12	47	59
ΣΥΝΟΛΟ	45	96	$n = 141$

- Αναμενόμενες συχνότητες (Expected frequencies)

ΦΥΛΟ	ΠΡΩΡΟΤΗΤΑ		ΣΥΝΟΛΟ
	ΝΑΙ	ΟΧΙ	
ΑΓΟΡΙΑ	26.2	55.8	82
ΚΟΡΙΤΣΙΑ	18.8	40.2	59
ΣΥΝΟΛΟ	45	96	$n = 141$

$$E = \frac{R \times C}{n}$$

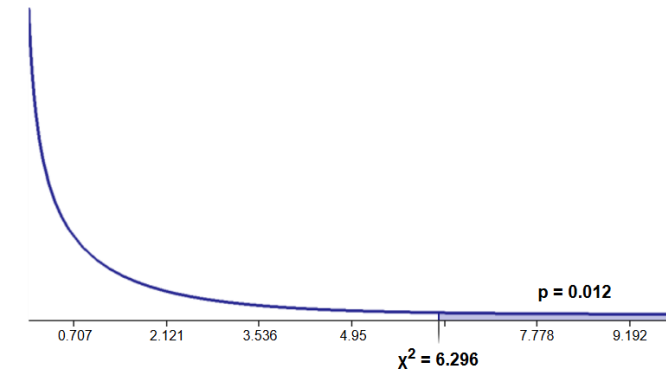
Παράδειγμα (πίνακας 2x2)

- Σχέση προωρότητας και φύλου

ΦΥΛΟ	ΠΡΟΩΡΟΤΗΤΑ		ΣΥΝΟΛΟ
	ΝΑΙ	ΟΧΙ	
ΑΓΟΡΙΑ	$a = 33$	$b = 49$	$a+b = 82$
ΚΟΡΙΤΣΙΑ	$c = 12$	$d = 47$	$c+d = 59$
ΣΥΝΟΛΟ	$a+c = 45$	$b+d = 96$	$n = 141$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$df=(r-1)(c-1)=(2-1)(2-1)=1$$



H_0 : Δεν υπάρχει σχέση μεταξύ φύλου και προωρότητας

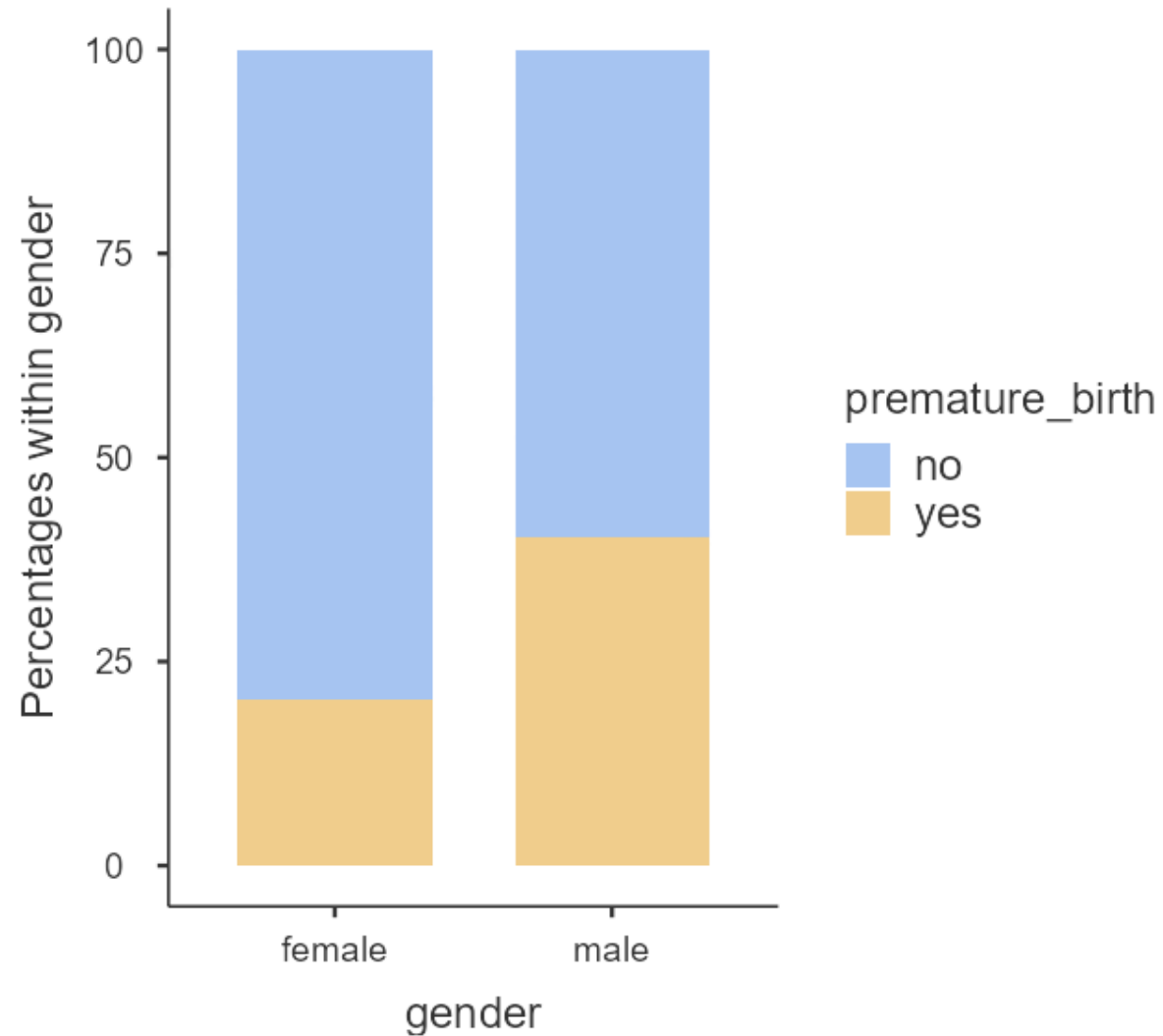
H_1 : Υπάρχει σχέση μεταξύ φύλου και προωρότητας

$$\chi^2 = \frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)} = \frac{141(33 * 47 - 49 * 12)^2}{82 * 59 * 45 * 96} = 6.256 \Rightarrow P = 0.012$$

⇒ Η προωρότητα σχετίζεται σημαντικά με το φύλο ($p=0.012<0.05$)

Παράδειγμα: Ερμηνεία με σύγκριση ποσοστών (πίνακας 2x2)

Plots



Παράδειγμα: Ερμηνεία με σύγκριση ποσοστών (πίνακας 2x2)

- Σχέση πρωρότητας και φύλου

ΦΥΛΟ	ΠΡΩΡΟΤΗΤΑ		ΣΥΝΟΛΟ
	ΝΑΙ	ΟΧΙ	
ΑΓΟΡΙΑ	$a = 33$	$b = 49$	$a+b = 82$
ΚΟΡΙΤΣΙΑ	$c = 12$	$d = 47$	$c+d = 59$
ΣΥΝΟΛΟ	$a+c = 45$	$b+d = 96$	$n = 141$

$$H_0: p_A = p_K$$

$$H_1: p_A \neq p_K$$

Ποσοστό πρόωρων αγοριών $p_A = 33/82 = 0.40$ (40%)

Ποσοστό πρόωρων κοριτσιών $p_K = 12/59 = 0.20$ (20%)

⇒ Το ποσοστό πρόωρων αγοριών (40%) είναι σημαντικά υψηλότερο από το ποσοστό πρόωρων κοριτσιών (20%) ($p=0.012 < 0.05$)

Λόγος σχετικών πιθανοτήτων (odds ratio)

- Σχέση προωρότητας και φύλου

ΦΥΛΟ	ΠΡΟΩΡΟΤΗΤΑ		ΣΥΝΟΛΟ
	ΝΑΙ	ΟΧΙ	
ΑΓΟΡΙΑ	$a = 33$	$b = 49$	$a+b = 82$
ΚΟΡΙΤΣΙΑ	$c = 12$	$d = 47$	$c+d = 59$
ΣΥΝΟΛΟ	$a+c = 45$	$b+d = 96$	$N = 141$

- Σχετική πιθανότητα πρόωρων αγοριών $= a/b = 33/49$
- Σχετική πιθανότητα πρόωρων κοριτσιών $= c/d = 12/47$
- $OR = (a/b)/(c/d) = a*d/b*c = 33*47/49*12 = \mathbf{2.638}$ (95% CI: 1.2 έως 5.7, δεν περιλαμβάνεται το 1 και έχουμε στατιστικά σημαντικό αποτέλεσμα, δηλαδή απόρριψη H_0).
- Το OR λέγεται και λόγος διασταυρούμενων (διαγώνιων) γινομένων ή λόγος αναλογιών

Λόγος σχετικών πιθανοτήτων (odds ratio)

- $OR = 1$ (η τιμή της H_0) δεν υπάρχει διαφορά στα ποσοστά (ίση πιθανότητα: $p_A = p_K$)
- $OR < 1$ μειωμένη πιθανότητα
- $OR > 1$ αυξημένη πιθανότητα

$OR = 2.6 \Rightarrow$ Η πιθανότητα προωρότητας ήταν **2.6 φορές μεγαλύτερη** στα αγόρια απ' ότι στα κορίτσια

- αντιστροφή λόγου (κορίτσια/αγόρια)
 $OR' = 1 / OR$

π.χ. $OR' = 1/2.6 = 0.38 \Rightarrow$ **Η πιθανότητα προωρότητας ήταν κατά $(0.38-1 = -0.62)$ 62% μικρότερη στα κορίτσια απ' ότι στα αγόρια**