



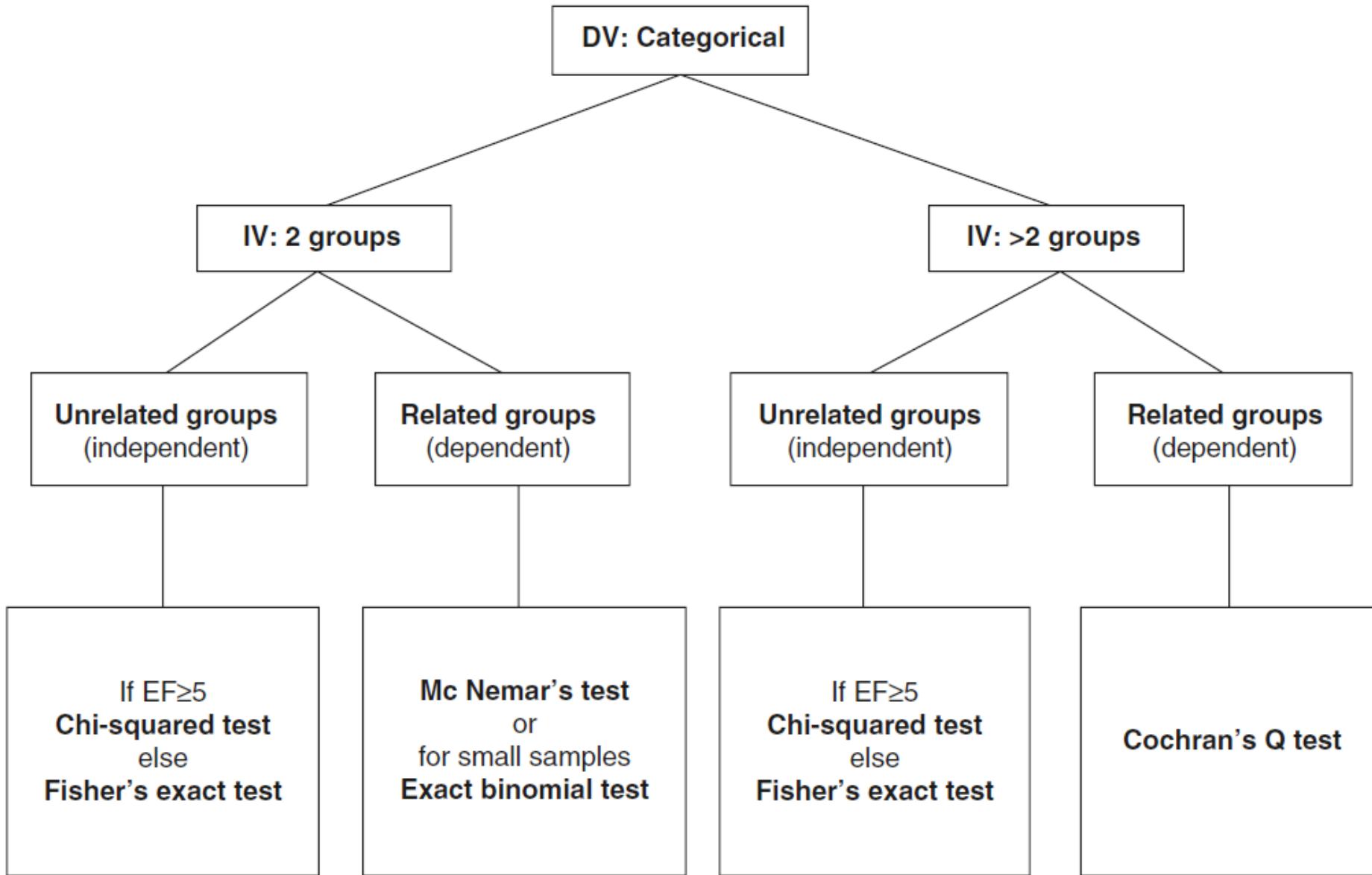
ΑΡΙΣΤΟΤΕΛΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΟΝΙΚΗΣ

Συσχέτιση μεταξύ κατηγορικών (ποιοτικών) μεταβλητών

Κωνσταντίνος Ι. Μπουγιούκας, PhD



ΘΕΣΣΑΛΟΝΙΚΗ



Η δοκιμασία χ^2 (chi-square test)

- Σύγκριση κατηγορικών (ποιοτικών) μεταβλητών.

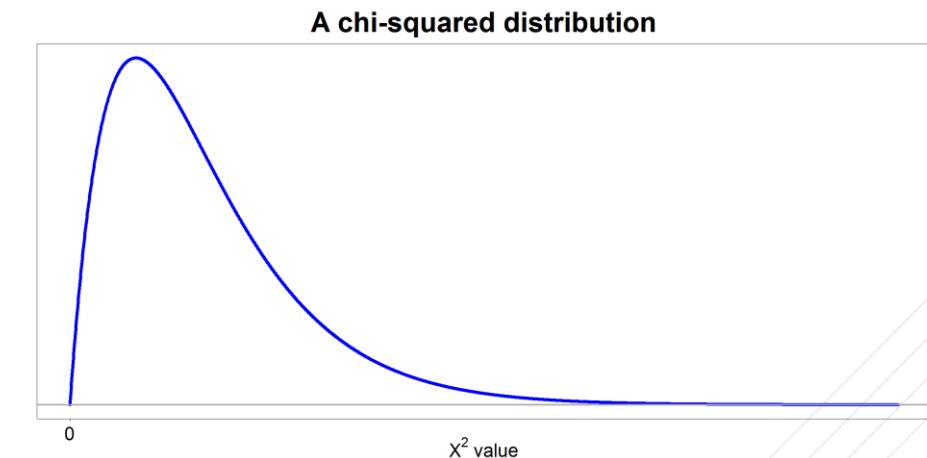
H_0 : Δεν υπάρχει συσχέτιση μεταξύ των δύο μεταβλητών.

H_1 : Υπάρχει συσχέτιση μεταξύ των δύο μεταβλητών.

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

Μεγάλη τιμή στο χ^2 \Rightarrow στατιστικά σημαντική συσχέτιση.

Δηλαδή, εάν $p < \alpha$ \rightarrow απορρίπτεται η H_0 .



- 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 προωρότητας και φύλου (πίνακας 2x2)

- Παρατηρούμενες συχνότητες (Observed frequencies)

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

- Αναμενόμενες συχνότητες (Expected frequencies)

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

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



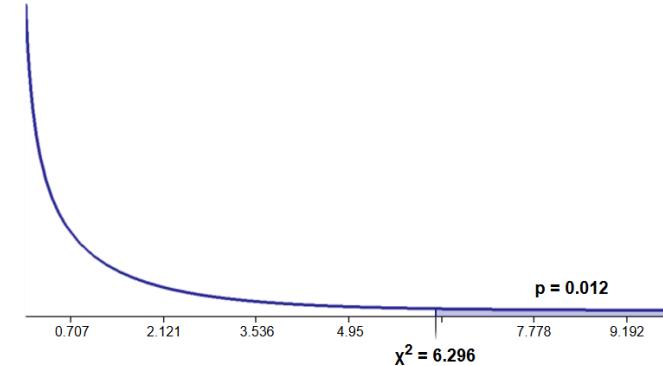
Παράδειγμα (πίνακας 2x2)

- Σχέση προωρότητας και φύλου

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

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

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



H_0 : Δεν υπάρχει σχέση μεταξύ φύλου και προωρότητας

H_1 : Υπάρχει σχέση μεταξύ φύλου και προωρότητας

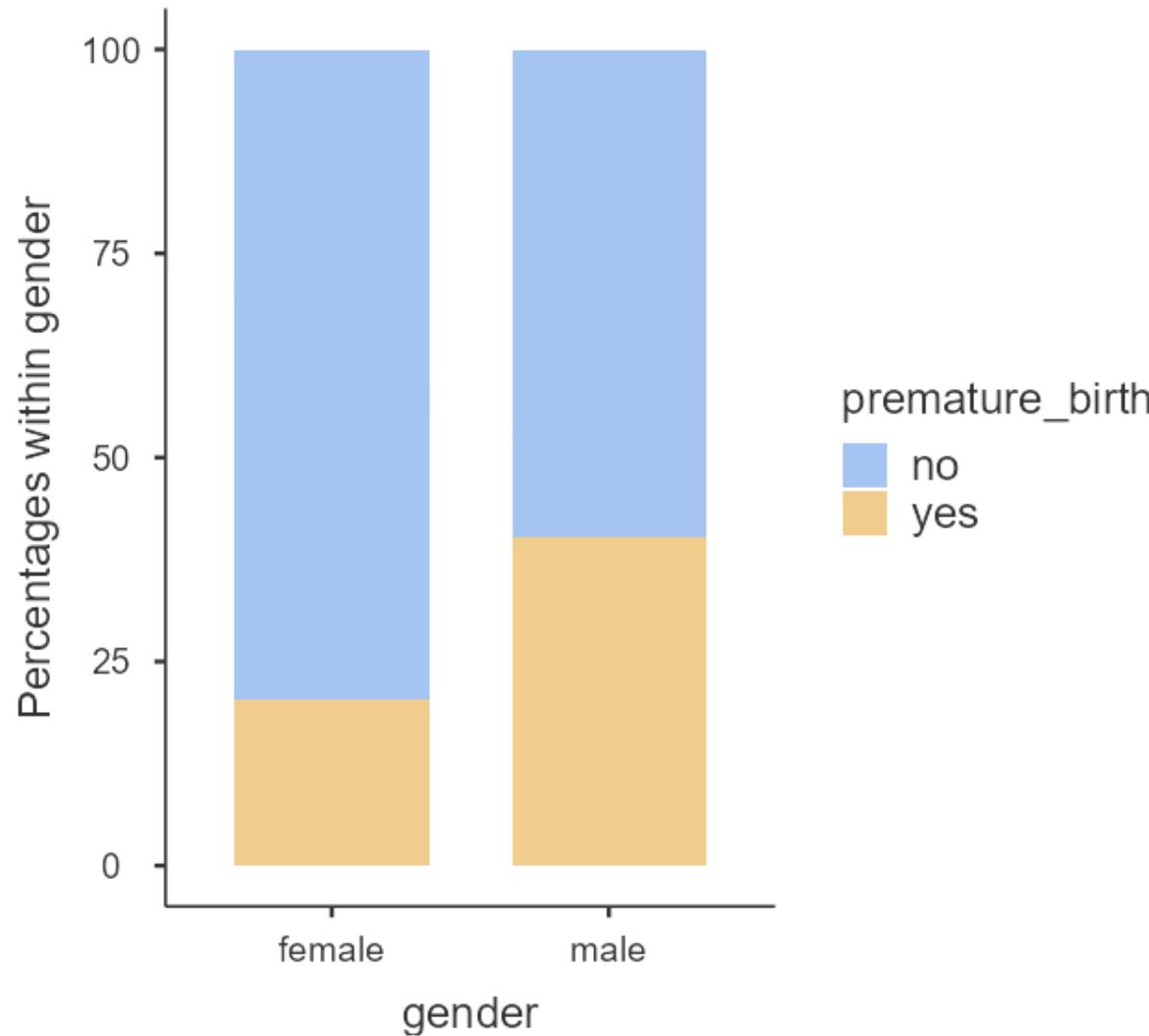
$$\chi^2 = \frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)} = \frac{141(33 * 47 - 49 * 12)^2}{82 * 59 * 45 * 96} = 6.256 \quad \Rightarrow P = 0.012$$

\Rightarrow 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 πρωτότητας και φύλου

ΦΥΛΟ	ΠΡΩΤΟΤΗΤΑ		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 πρώτων αγοριών $p_A = 33/82 = 0.402$ (40.2%)

Ποσοστό πρώτων κοριτσιών $p_K = 12/59 = 0.203$ (20.3%)

⇒ **Το ποσοστό πρώτων αγοριών (40.2%) είναι σημαντικά υψηλότερο από το ποσοστό πρώτων κοριτσιών (20.3%) ($p=0.012 < 0.05$)**



Σχετικός κίνδυνος (relative ratio ή risk ratio), RR

- Σχέση προωρότητας και φύλου

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

- Κίνδυνος πρόωρων αγοριών, $\text{Risk}_{\text{Αγόρια}} = a/(a+b) = 33/82=0.402$
- Κίνδυνος πρόωρων κοριτσιών, $\text{Risk}_{\text{Κορίτσια}} = c/(c+d) = 12/59=0.203$
- $\text{RR} = \text{Risk}_{\text{Αγόρια}} / \text{Risk}_{\text{Κορίτσια}} = 0.402/0.203 = 1.98$ (**95% CI: 1.12 έως 3.5, δεν περιλαμβάνεται το 1 και έχουμε στατιστικά σημαντικό αποτέλεσμα, δηλαδή απόρριψη } H_0**).



Σχετικός κίνδυνος (relative ratio ή risk ratio), RR

- $RR = 1$ (η τιμή της H_0) δεν υπάρχει διαφορά στον κίνδυνο
- $RR < 1$ μειωμένος κίνδυνος πρωρότητας
- $RR > 1$ αυξημένος κίνδυνος πρωρότητας

$RR = 1.98 \Rightarrow$ Ο κίνδυνος πρωρότητας είναι **1.98 φορές μεγαλύτερος** στα αγόρια απ' ότι στα κορίτσια.

Αυτό σημαίνει ότι:

Τα αγόρια έχουν $(1.98-1) \times 100\% = 98\%$ αυξημένο κίνδυνο να είναι πρόωρα σε σχέση με τα κορίτσια.



Λόγος σχετικών πιθανοτήτων (odds ratio), OR

- Σχέση προωρότητας και φύλου

ΠΡΟΩΡΟΤΗΤΑ			
ΦΥΛΟ	ΝΑΙ	ΟΧΙ	ΣΥΝΟΛΟ
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 $=c/d = 12/47$
- $OR = (a/b)/(c/d) = a*d/b*c = 33*47/49*12 = 2.638$ (**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 στη σχετική πιθανότητα
- $OR < 1$ μειωμένη σχετική πιθανότητα
- $OR > 1$ αυξημένη σχετική πιθανότητα

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

Επίσης, μπορεί να υπολογιστεί η σχετική πιθανότητα προωρότητας στα κορίτσια σε σχέση με τα αγόρια, με την αντιστροφή του λόγου OR. Σε αυτή την περίπτωση το νέο odds ratio είναι: $OR' = 1/ OR$.

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

