

Στατιστικοί έλεγχοι για διχότομα δεδομένα

Γεωργία Σαλαντή
Λέκτορας επιδημιολογίας

Διακριτές μεταβλητές

- Διχότομα δεδομένα (δύο πιθανές εκβάσεις)
 - Φύλο, ναι/οχι
- Διακριτά δεδομένα (πάνω από δύο εκβάσεις)
 - Ναι/όχι/ίσως, βελτίωση/επιδείνωση/σταθερός

Μεταβλητή 1/ Μεταβλητή 2	Συνεχής	Διακριτή
Συνεχής	Συντελεστής συσχέτισης	t-test Z-test Wilcoxon
Διακριτή	t-test Z-test Wilcoxon	X ² test Fisher's exact test McNeamar test

Αναιμία και φύλο: 2x2 πίνακας

		Αναιμία		Σύνολο
		Όχι	Ναι	
Φύλο	Γυναίκες	47 (82%)	10 (18%)	57
	Άνδρες	59 (91%)	6 (9%)	65
		106	16	

Είναι πιο συχνή η αναιμία στους άντρες
ή στις γυναίκες;

Μηδενική υπόθεση

- Η κατανομή των παρατηρήσεων στα κελιά του 2×2 πίνακα είναι τυχαία
- (η πιθανότητα αναιμίας στους άνδρες = πιθανότητα αναιμίας στις γυναίκες)

χ^2 ΤΕΣΤ

- Υπολογίζουμε τον αναμενόμενο αριθμό σε κάθε κελί **εάν** η πιθανότητα αναιμίας στους άνδρες και τις γυναίκες ήταν ίδιος

		Αναιμία		Σύνολο
		Όχι	Ναι	
Φύλο	Γυναίκες	47 $57 \cdot 106 / 122$ $= 49.5$	10	57
	Άνδρες	59	6	65
		106	16	122

χ^2 ΤΕΣΤ

- Υπολογίζουμε τον αναμενόμενο αριθμό σε κάθε κελί **εάν** η πιθανότητα αναιμίας στους άνδρες και τις γυναίκες ήταν ίδιος – ταιριάζουν με τις παρατηρήσεις;

		Αναιμία		Σύνολο
		Όχι	Ναι	
Φύλο	Γυναίκες	O=47 E=49.5	O=10 E=7.5	57
	Άνδρες	O=59 E=56.5	O=6 E=8.5	65
		106	16	122

χ^2 ΤΕΣΤ

- Υπολογίζουμε τον αναμενόμενο αριθμό σε κάθε κελί **εάν** η πιθανότητα αναιμίας στους άνδρες και τις γυναίκες ήταν ίδιος – ταιριάζουν με τις παρατηρήσεις;

$$\chi^2 = \sum_{i=1,2,3,4} \frac{(O_i - E_i)^2}{E_i}$$

O_i οι παρατηρήσεις, E_i οι προσδοκώμενες τιμές

χ^2 ΤΕΣΤ

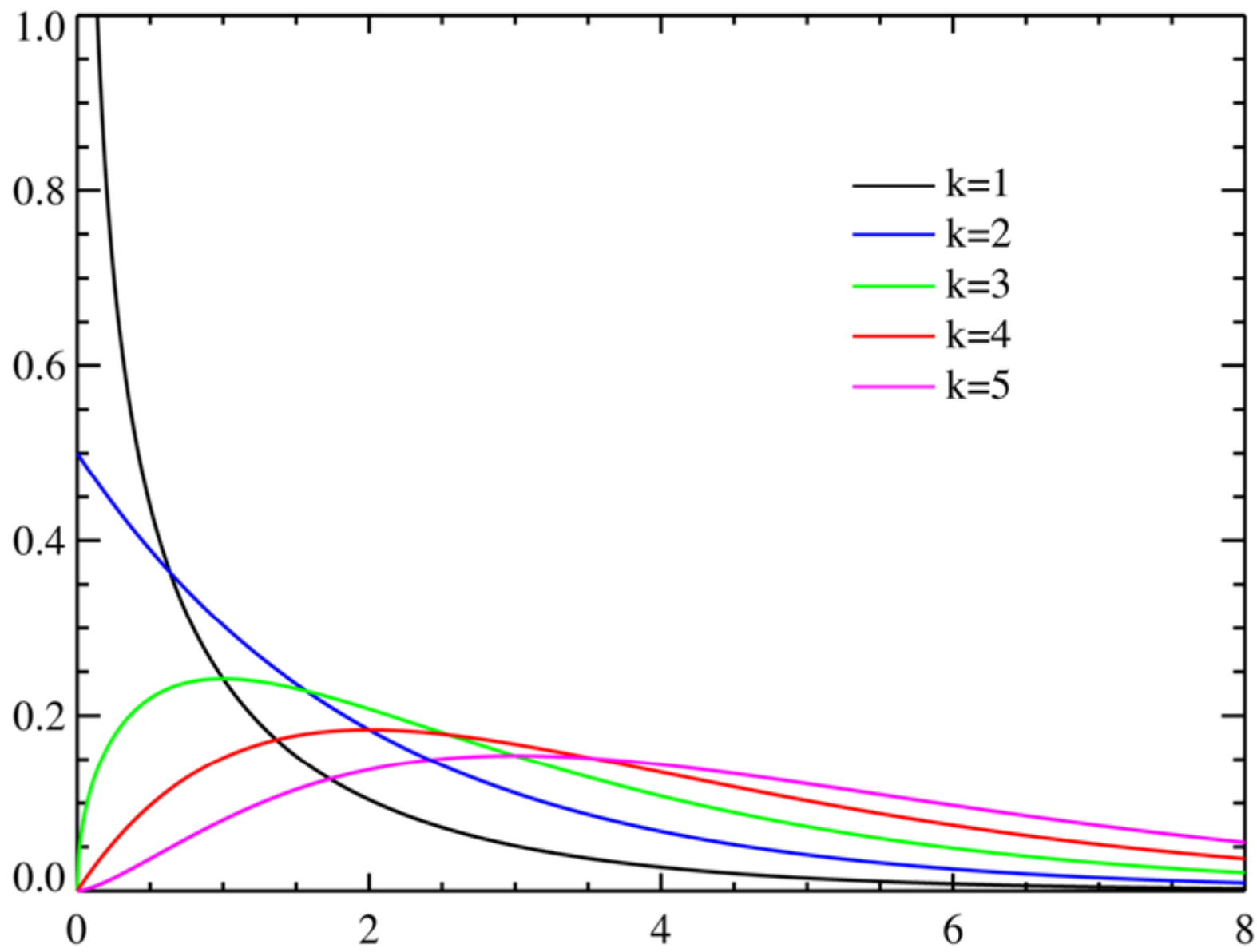
- Υπολογίζουμε τον αναμενόμενο αριθμό σε κάθε κελί *εάν* η πιθανότητα αναιμίας στους άνδρες και τις γυναίκες ήταν ίδιος – ταιριάζουν με τις παρατηρήσεις;

$$\chi^2 = 1.81$$

Ακολουθεί χ^2 κατανομή με 1 βαθμό ελευθερίας

Κατανομή χ^2

- χ^2 με k βαθμούς ελευθερίας (degrees of freedom)
- Όσο πιο πολλοί βαθμοί ελευθερίας, τόσο πιο «κοντή» η κατανομή



P-value

- Η πιθανότητα να βρούμε μια τόσο ακραία τιμή όσο αυτή του τέστ (=1.81) εάν το πείραμα επαναληφθεί με το ίδιο μέγεθος δείγματος
- P-value=0.17
- Τι σημαίνει αυτό;
 - Ότι η H_0 δεν απορρίπτεται
 - Ότι η συσχέτιση δεν είναι στατιστικά σημαντική

χ^2 ΤΕΣΤ

- Πάνω από δύο κατηγορίες

		Αναιμία		Σύνολο
		Όχι	Ναι	
Γονότυπος	AA	53	10	63
	AC	45	5	50
	CC	8	1	9
		106	16	122

χ^2 ΤΕΣΤ

- Υπολογίζω τις προσδοκώμενες τιμές όπως πριν
 - $\text{Σύνολο κολώνας} \cdot \text{Σύνολο γραμμής} / \text{Σύνολο}$
- Χρησιμοποιώ τον ίδιο τύπο για το χ^2 όπως πριν
- Κατανομή χ^2 με 2 βαθμούς ελευθερίας
- P-value=0.64
- Τι σημαίνει αυτό;

Βαθμοί ελευθερίας

- Αν ξέρω τα συνολικά (σε γραμμές και στήλες), πόσα κελιά χρειάζομαι για να συμπληρώσω όλον τον πίνακα;

		Αναιμία		Σύνολο
		Όχι	Ναι	
Φύλο	Γυναίκες	●		57
	Άνδρες			65
		106	16	122

Βαθμοί ελευθερίας

- Αν ξέρω τα συνολικά (σε γραμμές και στήλες), πόσα κελιά χρειαζόμαστε για να συμπληρώσω όλον τον πίνακα;

		Αναιμία		Σύνολο
		Όχι	Ναι	
Γονότυπος	AA	●		63
	AC	●		50
	CC			9
		106	16	122

χ^2 τεστ

Προϋποθέσεις για την εφαρμογή του: Όλες οι προσδοκώμενες τιμές να είναι πάνω από 5

χ^2 ΤΕΣΤ

- Το τεστ δεν είναι αξιόπιστο...

		Αναιμία		Σύνολο
		Όχι	Ναι	
Γονότυπος	AA	53	10	63
	AC	45	5	50
	CC	8	1 1.18	9
		106	16	122

Fisher test

- Μπορούμε να το εφαρμόσουμε όταν έχουμε λίγες ή σπάνιες παρατηρήσεις με κάποια κελιά να έχουν λίγες ή και καθόλου παρατηρήσεις
- $p\text{-value} = 0.77$

McNemar test

- Για ζευγαρωτές παρατηρήσεις και διχότομο αποτέλεσμα
- Πιθανότητα έκβασης πριν = Πιθανότητα έκβασης μετά