

# ΙΑΤΡΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

## Ακαδ. Έτος 2024-2025

Ανάλυση ποιοτικών χαρακτηριστικών - 1

# ΠΟΙΟΤΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ

Διαξονική κατανομή

$\chi^2$  για τον έλεγχο συσχέτισης ποιοτικών μεταβλητών

Υπολογισμός αναμενόμενων συχνοτήτων

Έλεγχος προϋποθέσεων

Υπολογισμός του  $\chi^2$

Βαθμοί ελευθερίας

Αξιολόγηση, συμπέρασμα, ερμηνεία

$\chi^2$  ετερογένειας

$\chi^2$  σε τετράπτυχους πίνακες

$\chi^2$  κατά ζεύγη

$\chi^2$  καλής εφαρμογής

Διαστρωμάτωση

$\chi^2$  κατά Mantel-Haenszel

Εκτίμηση βαθμού συσχέτισης ποιοτικών δεδομένων:

Σχετικός κίνδυνος – Σχετικός λόγος

Διάστημα εμπιστοσύνης- όρια αξιοπιστίας αναλογιών

Έλεγχος υποθέσεων για αναλογίες

# Κατανομή συχνοτήτων κατά ένα ποιοτικό μέγεθος

Κατανομή δείγματος 100 ατόμων κατά οικογενειακή κατάσταση			
Παντρεμένοι	Ανύπαντροι	Άλλα	Σύνολο
63	27	10	100

Κατανομή δείγματος 100 ατόμων σύμφωνα με την παρουσία ή απουσία μιας νόσου		
Ασθενείς	«Υγιείς»	Σύνολο
30	70	100



## Διαξονική κατανομή συχνοτήτων

Ταυτόχρονη κατανομή συχνοτήτων κατά δύο ποιοτικά χαρακτηριστικά

		Άσκηση			Σύνολο
		Έντονη	Μέτρια	Σπάνια/Καθόλου	
Έμφραγμα μυοκαρδίου	Ναι	10	30	60	100
	Όχι	20	40	40	100
Σύνολο		30	70	100	200

Υπόθεση για έλεγχο: Υπάρχει σχέση ανάμεσα στο βαθμό άσκησης και στον κίνδυνο εμφάνισης εμφράγματος μυοκαρδίου?

Εναλλακτική διατύπωση: Διαφέρει η αναλογία υγιών (ή ασθενών) ανά κατηγορία άσκησης ?

Εναλλακτική διατύπωση: Διαφέρει η κατανομή της άσκησης σε ασθενείς/υγιείς ?

## Παράδειγμα 2: Διερεύνηση της ύπαρξης σχέσης ανάμεσα στις καπνισματικές συνήθειες και στον κίνδυνο εμφάνισης καρκίνου του στομάχου

		Κάπνισμα		
		Ναι	Όχι	Σύνολο
Καρκίνος στομάχου	Ναι	60	40	100
	Όχι	120	80	200
Σύνολο		180	120	300

Υπόθεση για έλεγχο: Υπάρχει σχέση ανάμεσα στο κάπνισμα και στο Ca στομάχου?

ή (ισοδύναμα)

- Η αναλογία των καπνιστών διαφέρει ανάμεσα σε καρκινοπαθείς και μη καρκινοπαθείς?
- Η αναλογία των μη καπνιστών διαφέρει ανάμεσα σε καρκινοπαθείς και μη καρκινοπαθείς?
- Η αναλογία των καρκινοπαθών διαφέρει ανάμεσα σε καπνιστές και μη καπνιστές?
- Η αναλογία των μη καρκινοπαθών διαφέρει ανάμεσα σε καπνιστές και μη καπνιστές?

## Αντι-παράδειγμα.

**Πρόβλημα: Διερεύνηση της σχέσης ανάμεσα στο κάπνισμα και στον καρκίνο του στομάχου**

Ένας γιατρός ρωτάει 100 ασθενείς με καρκίνο του στομάχου αν καπνίζουν και βρίσκει ότι:

	Καπνίζουν	Δεν καπνίζουν	Σύνολο
Ασθενείς	80	20	100

Μπορεί να συμπεράνει από τα δεδομένα αυτά ότι συσχετίζεται η εμφάνιση καρκίνου του στομάχου με το κάπνισμα;

Τι άλλο χρειάζεται να γνωρίζει;

## $\chi^2$ ως κριτήριο συσχέτισης ποιοτικών χαρακτηριστικών

Παράδειγμα: Κατανομή 890 ατόμων κατά επάγγελμα και ασθένεια (ύπαρξη ή όχι καρκίνου της ουροδόχου κύστεως)

	Επάγγελμα			Σύνολο
	Γεωργοί	Υπάλληλοι	Εργάτες	
Καρκινοπαθείς	140 (37%)	125 (32%)	55 (46%)	320 (36%)
«Υγιείς»	240	265	65	570
Σύνολο	380	390	120	890

Φαίνεται να υπάρχουν διαφορές στις αναλογίες των ασθενών/“υγιών” μεταξύ ατόμων διαφορετικών επαγγελμαμάτων.

Οφείλονται οι διαφορές στην “τύχη”, την τυχαία δειγματοληπτική διακύμανση, ή υπάρχουν πραγματικά διαφορές;

Οι στήλες του Πίνακα  $K=3$ , οι γραμμές του Πίνακα  $L=2$ , κελιά=6

## ΕΡΕΥΝΗΤΙΚΗ ΥΠΟΘΕΣΗ – ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΕΩΝ

- Υπάρχει σχέση μεταξύ του επαγγέλματος (ποιοτική μεταβλητή) και της πιθανότητας εμφάνισης καρκίνου της ουροδόχου κύστης (ποιοτική μεταβλητή);
- $H_0$  : Δεν υπάρχει σχέση μεταξύ των 2 μεταβλητών
- $H_1$  : Υπάρχει σχέση μεταξύ των δύο μεταβλητών



## Στατιστική αξιολόγηση

- Για να απαντήσουμε σε αυτό το ερώτημα πρέπει να κάνουμε στατιστική αξιολόγηση.
- Η συνηθισμένη μέθοδος με την οποία αξιολογούμε στατιστικά την συσχέτιση μεταξύ ποιοτικών χαρακτηριστικών είναι το  $\chi^2$  **“ως κριτήριο συσχέτισης ποιοτικών χαρακτηριστικών”**.

## Στατιστική αξιολόγηση (2)

- 1) υπολογισμός των **αναμενόμενων** συχνοτήτων:  
συχνότητες που αναμένουμε να παρατηρήσουμε κάτω από την μηδενική υπόθεση, δηλαδή κάτω από την υπόθεση ότι δεν υπάρχει σχέση μεταξύ των δύο ποιοτικών χαρακτηριστικών
- 2) Υπολογισμός **στατιστικής δοκιμασίας**  $\chi^2$
- 3) Σύγκριση του αποτελέσματος της στατιστικής δοκιμασίας με την **οριακή τιμή της  $\chi^2$  κατανομής** σε  $(K-1) \times (L-1)$  βαθμούς ελευθερίας ( $K$ =αριθμός στηλών,  $L$ =αριθμός γραμμών) και σε επίπεδο σημαντικότητας  $\alpha\%$  (συνήθως  $\alpha=5$ , ή  $\alpha=1$ )
- 4) Αξιολόγηση με τον γνωστό τρόπο...

# 1) Αναμενόμενες συχνότητες

Παράδειγμα: Κατανομή 890 ατόμων κατά επάγγελμα και ασθένεια (ύπαρξη ή όχι καρκίνου της ουροδόχου κύστεως)

	Επάγγελμα			Σύνολο
	Γεωργοί	Υπάλληλοι	Εργάτες	
Καρκινοπαθείς	140	125	55	320
«Υγιείς»	240	265	65	570
Σύνολο	380	390	120	890

Τρόπος σκέψης και υπολογισμού.

Αν δεν υπάρχει σχέση των 2 μεταβλητών η κατανομή σε κάθε κελί θα ήταν ίδια με την συνολική κατανομή

Π.χ. Για το κελί των καρκινοπαθών-γεωργών:

Σε σύνολο 890 ατόμων έχουμε 380 γεωργούς

Σε σύνολο 320 καρκινοπαθών πόσους (X) γεωργούς αναμένουμε να έχουμε **αν δεν συσχετίζεται το επάγγελμα με τον καρκίνο ουροδόχου κύστεως?**

$$X = (320 * 380) / 890 = 136,6$$

# 1) Αναμενόμενες συχνότητες (συν)

Παράδειγμα: Κατανομή 890 ατόμων κατά επάγγελμα και ασθένεια (ύπαρξη ή όχι καρκίνου της ουροδόχου κύστεως)

	Επάγγελμα			Σύνολο
	Γεωργοί	Υπάλληλοι	Εργάτες	
Καρκινοπαθείς	140	125	55	320
«Υγιείς»	240	265	65	570
Σύνολο	380	390	120	890

Τρόπος σκέψης και υπολογισμού.

- Π.χ. Για το κελί των «υγιών» υπαλλήλων:

Σε 890 άτομα υπάρχουν 390 υπάλληλοι.

Σε 570 «υγιείς» πόσοι (X) θα είναι υπάλληλοι (εάν ισχύει η  $H_0$ )?

$$X = (390 * 570) / 890 = 249,8$$

# 1) Αναμενόμενες συχνότητες (συν)

Ανάλογα σκεφτόμαστε και για τα 6 κελιά και προκύπτουν 6 αναμενόμενες συχνότητες.

Πάντα καταλήγουμε τον τύπο:

**ΑΝΑΜΕΝ. ΣΥΧΝ = (αντ. οριζόντιο συν. Χ αντ. κάθετο συν.)/γενικό συν.**

# 1) Αναμενόμενες συχνότητες (συν)

Παράδειγμα: Κατανομή 890 ατόμων κατά επάγγελμα και ασθένεια (ύπαρξη ή όχι καρκίνου της ουροδόχου κύστεως)

	Επάγγελμα			
	Γεωργοί	Υπάλληλοι	Εργάτες	Σύνολο
Καρκινοπαθείς	140 <b>136,6</b>	125 <b>140,2</b>	55 <b>43,2</b>	320
«Υγιείς»	240 <b>243,4</b>	265 <b>249,8</b>	65 <b>76,8</b>	570
Σύνολο	380	390	120	890

## 2) Υπολογισμός στατιστικής δοκιμασίας $\chi^2$

$$\chi^2 = \sum_{l=1}^m \frac{(O_l - E_l)^2}{E_l}$$

Όπου  $O$  = παρατηρηθείσες συχνότητες (Observed)

$E$  = αναμενόμενες συχνότητες (Expected)

$m$  = αριθμός κελιών του πίνακα

## 2) Υπολογισμός στατιστικής δοκιμασίας $\chi^2$ (συν)

$$\begin{aligned}\chi^2 &= [(140-136,6)^2/136,6] + [(125-140,2)^2/140,2] + \\ & [(55-43,2)^2/43,2] + [(240-243,4)^2/243,4] + \\ & [(265-249,8)^2/249,8] + [(65-76,8)^2/76,8] = \\ & = 0,08 + 1,65 + 3,22 + 0,05 + 0,92 + 1,81 = 7,73\end{aligned}$$



### 3) Στατιστική αξιολόγηση

Σύγκριση του αποτελέσματος με την οριακή τιμή της  $\chi^2$  κατανομής που ακολουθεί η συνάρτηση της στατιστικής δοκιμασίας υπό την  $H_0$

Χρησιμοποιούμε τον Πίνακα Π3, που έχει την μορφή:

ΒΕ	Στατιστικά σημαντικές διαφορές στο επίπεδο			
	10%	5%	1%	0,1%
1	2,71	3,84	6,64	10,83
2	4,61	5,99	9,21	13,82
3	6,25	7,82	11,35	16,27
.....	.....	.....	.....	.....
40	51,81	55,76	63,69	73,40

### 3) Στατιστική αξιολόγηση (συν)

Υπολογισμός των βαθμών ελευθερίας:

$$(K-1)*(L-1) = (3-1)*(2-1) = 2$$

(K= αριθμός στηλών, L= αριθμός γραμμών)

Αξιολόγηση

BE	10%	5%	1%	0,1%
2	4,61	5,99	9,21	13,82

$$5,99 < \mathbf{7,73} < 9,21$$

$$5\% > P > 1\%$$

ή

$$0,05 > P > 0,01$$

Τι μας λέει αυτή η αξιολόγηση;

## Συμπέρασμα- Ερμηνεία

- Συμπέρασμα: η σχέση ανάμεσα στο επάγγελμα και στον κίνδυνο εμφάνισης καρκίνου της ουροδόχου κύστεως είναι στατιστικά σημαντική στο επίπεδο 5%
- Ερμηνεία: Οι εργάτες έχουν μεγαλύτερη πιθανότητα από αυτή που αναμένεται να εμφανίσουν τη νόσο, οι γεωργοί περίπου ίδια, ενώ οι υπάλληλοι έχουν μικρότερη πιθανότητα από αυτή που αναμένεται να εμφανίσουν την νόσο.

## Προϋποθέσεις εφαρμογής του $\chi^2$ (αναφέρονται στις αναμενόμενες συχνότητες)

- 1. Όλες οι αναμενόμενες συχνότητες  $>1$
- 2. Οι περισσότερες αναμενόμενες συχνότητες (τα  $4/5$ )  $>5$ .

## Τι είναι οι Βαθμοί Ελευθερίας?

- Πόσες συχνότητες μπορούμε να τοποθετήσουμε αυθαίρετα στα κελιά, αν τα μερικά (κάθετα και οριζόντια) σύνολα θεωρηθούν δεδομένα;

## ΒΑΘΜΟΙ ΕΛΕΥΘΕΡΙΑΣ (συν)

	Επάγγελμα			
	Γεωργοί	Υπάλληλοι	Εργάτες	Σύνολο
Καρκινοπαθείς	<b><math>\alpha</math></b>	$320 - \alpha - (120 - \beta)$	$120 - \beta$	<b>320</b>
“Υγιείς”	$380 - \alpha$	$390 - (320 - \alpha - (120 - \beta))$	<b><math>\beta</math></b>	<b>570</b>
Σύνολο	<b>380</b>	<b>390</b>	<b>120</b>	<b>890</b>

# Δοκιμασία $\chi^2$ ως κριτήριο ετερογενείας

- Διερεύνηση του αν η αναλογία κάποιων χαρακτηριστικών διαφέρει ανάμεσα σε διάφορες ομάδες.
- Η διαξονική κατανομή γίνεται με άξονες αφενός την παρουσία ή απουσία αυτού του χαρακτηριστικού και αφετέρου τις ομάδες.
- Στην περίπτωση αυτή το  $\chi^2$  που εφαρμόζεται (ίδιο ακριβώς με το  $\chi^2$  για τη συσχέτιση ποιοτικών χαρακτηριστικών) λέγεται  **$\chi^2$  ως κριτήριο ετερογένειας**.

# Δοκιμασία $\chi^2$ ως κριτήριο ετερογενείας

Κατανομή 620 καρκινοπαθών που νοσηλεύτηκαν σε 5 νοσοκομεία κατά στάδιο της νόσου

	Νοσοκομείο					
<b>Στάδιο νόσου</b>	<b>1ο</b>	<b>2ο</b>	<b>3ο</b>	<b>4ο</b>	<b>5ο</b>	<b>Σύνολο</b>
Εντοπι- σμένος	135	62	85	80	153	515
Μεταστα- τικός	40	10	15	10	30	105
Σύνολο	175	72	100	90	183	620



# Έλεγχος υπόθεσης

- Ερευνητική υπόθεση: Υπάρχει διαφορά (ετερογένεια) στην κατανομή της νόσου ανά στάδιο στα 5 νοσοκομεία?
- $H_0$ : Δεν υπάρχει διαφορά (ετερογένεια) στην κατανομή της νόσου ανά στάδιο στα 5 νοσοκομεία
- $H_1$ : Υπάρχει διαφορά (ετερογένεια) στην κατανομή της νόσου ανά στάδιο στα 5 νοσοκομεία

## Δοκιμασία $\chi^2$ ως κριτήριο ετερογενείας

Κατανομή 620 καρκινοπαθών που νοσηλεύτηκαν σε 5 νοσοκομεία κατά στάδιο της νόσου

Στάδιο νόσου	Νοσοκομείο					Σύνολο
	1ο	2ο	3ο	4ο	5ο	
Εντοπι- σμένος	135	62	85	80	153	515
Μεταστα- τικός	40	10	15	10	30	105
Σύνολο	175	72	100	90	183	620

$$\chi^2 = \sum_{i=1}^{10} \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2 = 8,21$$

$$BE = (5-1) * (2-1) = 4$$

## Αξιολόγηση- συμπέρασμα -ερμηνεία

BE	10%	5%	1%	0,1%
4	7,78	9,49	13,28	18,47

$$7,78 < 8,21 < 9,49$$

$$10\% > P > 5\%$$

- Συμπέρασμα: Η ετερογένεια των περιστατικών ανά στάδιο νόσου, δεν είναι στατιστικά σημαντική, είναι όμως ενδεικτική, γιατί  $P < 10\%$
- Ερμηνεία: Υπάρχει ένδειξη ετερογένειας κατά στάδιο της νόσου στα διάφορα νοσοκομεία η οποία δεν φτάνει στο συμβατικό όριο της στατιστικής σημαντικότητας

# Δοκιμασία $\chi^2$ στους τετράπτυχους πίνακες

Κατανομή 2239 γυναικών με καρκίνο του μαστού και 1370 «υγειών» (μη-καρκινοπαθών) κατά αναμνηστικό προκλητής ή φυσιολογικής εμμηνόπαυσης

Τρόπος εμμηνόπαυσης	Καρκινοπαθείς	Μη-καρκινοπαθείς	Σύνολο
Προκλητή εμμ.	a= 469 <b>(21%)</b>	b= 473 <b>(35%)</b>	a+b= 942
Φυσιολογική εμμ.	c= 1770 <b>(79%)</b>	d= 897 <b>(65%)</b>	c+d= 2667
Σύνολο	a+c=2239	b+d= 1370	n= 3609

Ή με διόρθωση κατά Yates

$$\chi^2 = \frac{(a \cdot d - b \cdot c)^2 \cdot n}{(a + c)(b + d)(a + b)(c + d)}$$

$$\chi^2 = \frac{(|a \cdot d - b \cdot c| - n / 2)^2 \cdot n}{(a + c)(b + d)(a + b)(c + d)}$$

Εφαρμογή του τύπου, ΒΕ, αξιολόγηση

$$X^2 = \frac{(|a \cdot d - b \cdot c| - n / 2)^2 \cdot n}{(a + c)(b + d)(a + b)(c + d)} =$$
$$= \frac{(|1469 \cdot 897 - 473 \cdot 1770| - 3609 / 2)^2 \cdot 3609}{942 \cdot 2667 \cdot 2239 \cdot 1370} = 80,54$$

ΒΕ= (2-1) \* (2-1)=1 πάντα στους τετράπτυχους

ΒΕ	10%	5%	1%	0,1%
1	2,71	3,84	6,64	10,83

10,83 < 80,54

0,1% > P

Τι συμπεραίνουμε;

## Συμπέρασμα- ερμηνεία

- Συμπέρασμα: υπάρχει στατιστικά πολύ σημαντική σχέση ανάμεσα στον τρόπο εμμηνόπαυσης και στον κίνδυνο εμφάνισης καρκίνου του μαστού.
- Ερμηνεία: οι γυναίκες με προκλητή εμμηνόπαυση έχουν μικρότερο κίνδυνο εμφάνισης καρκίνου του μαστού