ΙΑΤΡΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Επαναληπτικό μάθημα στην ανάλυση ποιοτικών χαρακτηριστικών

X^2-test

Αν θέλουμε να εξετάσουμε τη συσχέτιση μεταξύ

2 ποιοτικών μεταβλητών

(π.χ. κάπνισμα και φύλο)

X^2-test

(μία από τις εφαρμογές του)

Εφαρμογές του X^2-test

1. **Δοκιμασία X^2 ως κριτήριο συσχέτισης ποιοτικών χαρακτηριστικών (ή ως κριτήριο ετερογένειας)**
   Στατιστικά σημαντικό αποτέλεσμα: Ενδεικτικό συσχέτισης του ενός ποιοτικού χαρακτηριστικού με το άλλο

2. **Δοκιμασία X^2 ως κριτήριο διαφοράς δύο αναλογιών (ποσοστών)**
   Στατιστικά σημαντικό αποτέλεσμα: Τα ποσοστά διαφέρουν

3. **Δοκιμασία X^2 ως κριτήριο καλής εφαρμογής**
   Όταν θέλουμε να ελέγξουμε αν η κατανομή ενός ποιοτικού χαρακτηριστικού είναι συμβατή με μία θεωρητική κατανομή (π.χ σε 60 ρίψεις ζαριού, αν η συχνότητα εμφάνισης των αριθμών 1 έως 6 ακολουθεί την ομοιόμορφη κατανομή)
   Στατιστικά σημαντικό αποτέλεσμα: Ενδεικτικό ότι δεν υπάρχει καλή εφαρμογή

4. **Δοκιμασία X^2 για σύγκριση ποιοτικών παρατηρήσεων κατά ζεύγη** (paired X^2, McNemar’s test)

X^2-test

1. **Γενικός τύπος**

$$X^2 = \sum_i \frac{(O - E)^2}{E}$$

2. **Ειδικά για πίνακα 2x2**
   (όταν δηλαδή και οι 2 μεταβλητές είναι ποιοτικές <u>με 2 κατηγορίες</u>)

a	b
c	d

$$X^2 = \frac{\left(|ad - bc| - \frac{n}{2}\right)^2 n}{(a+b)(c+d)(a+c)(b+d)}$$

X^2-test

Την υπολογισθείσα τιμή του X^2 την συγκρίνουμε στη συνέχεια με τις οριακές τιμές στον αντίστοιχο πίνακα

- σε επίπεδο σημαντικότητας 5%
- σε βαθμούς ελευθερίας = (K-1)*(L-1) όπου K: αριθμός στηλών, L: αριθμός γραμμών του πίνακα

Αν X^2 ≥οριακή τιμή → απορρίπτω H_0 και συμπεραίνω ότι η σχέση είναι στατιστικά σημαντική στο συγκεκριμένο επίπεδο σημαντικότητας (5% ή και μικρότερο, π.χ. 1%)

Αν X^2 <οριακή τιμή → η σχέση δεν είναι στατιστικά σημαντική

Παράδειγμα 4 (α)

Σε 100 άτομα με καρκίνο του πνεύμονα (cases) οι 24 είναι μη καπνιστές, οι 15 πρώην καπνιστές και οι 61 καπνιστές. Σε δείγμα 105 υγιών ατόμων (controls) οι 82 είναι μη καπνιστές και οι 13 πρώην καπνιστές.

(α) Υπάρχει σχέση ανάμεσα στον καρκίνο του πνεύμονα και τις καπνιστικές συνήθειες;

Κάπνισμα (καπνιστής/ πρώην καπνιστής/ μη καπνιστής) → **Ca πνεύμονα (ασθενείς, υγιείς)**

Ποιοτική (με 3 κατηγορίες)

X^2-test

Ποιοτική (με 2 κατηγορίες)

Ca Πνεύμονα	Κάπνισμα			
	Μη	Πρώην	Νυν	Σύνολο
Ασθενείς	24	15	61	100
Υγείς	82	13	10	105
Σύνολο	106	28	71	205

Ca Πνεύμονα	Κάπνισμα			
	Μη	Πρώην	Νυν	Σύνολο
Ασθενείς	24 (23%)	15 (54%)	61 (86%)	100
Υγείς	82	13	10	105
Σύνολο	106 (100%)	28 (100%)	71 (100%)	205

Είναι αυτή η διαφορά πραγματική

ή

είναι αποτέλεσμα τυχαίας διακύμανσης?

Δοκιμασία X^2 για συσχέτιση ποιοτικών χαρακτηριστικών

- Τρόπος υπολογισμού:
  - Υπολογίζω τον αριθμό ατόμων σε κάθε κελί του πίνακα που θα περίμενα αν ΔΕΝ υπήρχε σχέση μεταξύ καπνίσματος και Ca πνεύμονα (→ **Αναμενόμενες συχνότητες**)
  - Αν οι **παρατηρηθείσες** (**O**bserved) συχνότητες διαφέρουν πολύ από τις **αναμενόμενες** (**E**xpected), συμπεραίνω ότι υπάρχει σχέση μεταξύ καπνίσματος και Ca πνεύμονα

$$X^2 = \sum_i \frac{(O-E)^2}{E}$$

Αναμενόμενες συχνότητες (Ε)

Π.χ. Για το κελί καρκινοπαθών-μη καπνιστών

Ca Πνεύμονα	**Κάπνισμα**			
	Μη	**Πρώην**	**Νυν**	Σύνολο
Ασθενείς	24	15	61	100
Υγείς	82	13	10	105
Σύνολο	106	28	71	205

Στους 205 υπάρχουν 106 μη καπνιστές

Στους 100 καρκινοπαθείς Χ;

$$X = \frac{100 * 106}{205} = 51,7$$

$$X = \frac{\alpha\nu\tau.o\rho\iota\zeta\acute{o}\nu\tau\iota o\ \sigma\acute{\upsilon}\nu o\lambda o * \alpha\nu\tau.\kappa\acute{\alpha}\theta\varepsilon\tau o\ \sigma\acute{\upsilon}\nu o\lambda o}{\gamma\varepsilon\nu\iota\kappa\acute{o}\ \sigma\acute{\upsilon}\nu o\lambda o}$$

Πίνακας παρατηρηθεισών και αναμενόμενων συχνοτήτων

Ca Πνεύμονα	**Κάπνισμα**			
	Μη	**Πρώην**	**Νυν**	Σύνολο
Ασθενείς	**24/51.7**	**15/13,7**	**61/ 34,6**	100
Υγείς	**82 /54,3**	**13 / 14,3**	**10 / 36,4**	105
Σύνολο	106	28	71	205

Είναι αυτή η διαφορά πραγματική ή
είναι αποτέλεσμα τυχαίας διακύμανσης?

→ **Χ²-test**: θα υπολογίσω την πιθανότητα να βρω μία τέτοια τάξης διαφορά στην τύχη - όταν δηλαδή δεν υπάρχει σχέση μεταξύ καπνίσματος και Ca (σφάλμα τύπου Ι)

→ αν η πιθανότητα να βρω μία τέτοια τάξης διαφορά **κατά τύχη** είναι πολύ μικρή (<0.05) τότε η διαφορά είναι στατιστικά σημαντική

Προϋποθέσεις για X^2

- Όλες οι αναμενόμενες συχνότητες > 1

- Τα 4/5 των αναμενόμενων συχνοτήτων > 5

Υπολογισμός του X^2

$$X^2 = \sum_i \frac{(O - E)^2}{E}$$

O : παρατηρηθείσες συχνότητες

E : αναμενόμενες συχνότητες

Ca Πνεύμονα	Κάπνισμα		
	Μη	Πρώην	Νυν
Ασθενείς	24/ 51.7	15/ 13,7	61/ 34,6
Υγείς	82/ 54,3	13/ 14,3	10/ 36,4

$$X^2 = \frac{(24-51,7)^2}{51,7} + \frac{(15-13,7)^2}{13,7} + \frac{(61-34,6)^2}{34,6} +$$

$$\frac{(82-54,3)^2}{54,3} + \frac{(13-14,3)^2}{14,3} + \frac{(10-36,4)^2}{36,4}$$

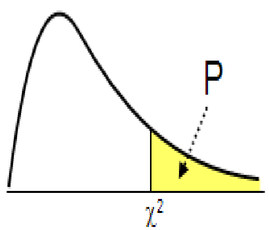
X^2= 14,8 + 0,1 +20,1+14,1 + 0,1 + 19,2 = 68,4

Πως κρίνω αν η διαφορά είναι «αρκετά μεγάλη»

- Ανατρέχω στον αντίστοιχο πίνακα
  - Βαθμοί ελευθερίας = (K-1)(L-1) = (3-1)(2-1) = 2

    K: αριθμός στηλών, L: αριθμός γραμμών

Chi-square distribution table



	P										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124

BE	10%	5%	1%	0.1%
2	4.61	5.99	9.21	13.82

X^2=68,4
68,4>13.82
p< 0.1%

Συμπέρασμα: Υπάρχει στατιστικά σημαντική σχέση μεταξύ καρκίνου του πνεύμονα και ιστορικού καπνίσματος

Ερμηνεία: Οι καπνιστές έχουν αυξημένο κίνδυνο ανάπτυξης καρκίνου του πνεύμονα συγκριτικά με τους μη καπνιστές

Παράδειγμα 4 (β)

(β) Ανάμεσα στα άτομα με καρκίνο του πνεύμονα (cases) οι 61 είναι καπνιστές (νυν και πρώην) και πίνουν καφέ, 7 είναι μη καπνιστές και πίνουν καφέ, ενώ 17 δεν πίνουν καφέ και δεν καπνίζουν.

Στο δείγμα 105 υγιών ατόμων (controls) οι 15 είναι καπνιστές (νυν και πρώην) και πίνουν καφέ, 18 είναι μη καπνιστές και πίνουν καφέ, ενώ 64 δεν πίνουν καφέ και δεν καπνίζουν.

Υπάρχει σχέση ανάμεσα στον καρκίνο του πνεύμονα και στην κατανάλωση καφέ;

Ανάμεσα στα άτομα με καρκίνο του πνεύμονα (cases) οι 61 είναι καπνιστές (νυν και πρώην) και πίνουν καφέ, 7 είναι μη καπνιστές και πίνουν καφέ, ενώ 17 δεν πίνουν καφέ και δεν καπνίζουν.

	Καφές		Σύνολο
	Ναι	Όχι	
Ασθενείς	61+7=68	100-68=32	100
Υγιείς			105

Στο δείγμα 105 υγιών ατόμων (controls) οι 15 είναι καπνιστές (νυν και πρώην) και πίνουν καφέ, 18 είναι μη καπνιστές και πίνουν καφέ, ενώ 64 δεν πίνουν καφέ και δεν καπνίζουν.

	Καφές		Σύνολο
	Ναι	Όχι	
Ασθενείς	61+7=68	100-68=32	100
Υγιείς	15+18=33	105-33=72	105

Κατανάλωση καφέ και καρκίνος πνεύμονα

	Καφές		Σύνολο
	Ναι	Όχι	
Ασθενείς	68 (68%)	32	100 (100%)
Υγιείς	33 (31%)	72	105 (100%)

- 68% των ασθενών πίνει καφέ
- 31% των υγιών πίνει καφέ
- Είναι αυτή η διαφορά τυχαία ή στατιστικά σημαντική;

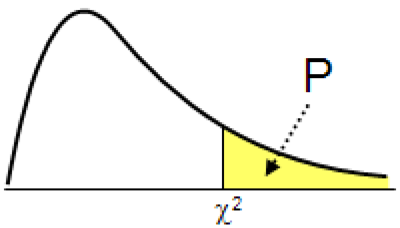
Κατανάλωση καφέ και καρκίνος πνεύμονα

	Καφές		Σύνολο
	Ναι	Όχι	
Ασθενείς	68	32	100
Υγιείς	33	72	105

$$X^2 = \frac{\left(\left|ad - bc\right| - \frac{n}{2}\right)^2 n}{(a+b)(c+d)(a+c)(b+d)} = 26$$

B.E.=(2-1)x(2-1)=1

Chi-square distribution table



	P										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124

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

26> 10,83

p<0.1%

Συμπέρασμα: Υπάρχει στατιστικά σημαντική σχέση μεταξύ καρκίνου του πνεύμονα και καθημερινής κατανάλωσης καφέ

Παράγοντες που συντελούν σε παρατηρούμενη σχέση μεταξύ κάποιων μεταβλητών

- **Τυχαίο σφάλμα**
  - Ψευδείς συσχετίσεις μπορεί να βρεθούν κατά τύχη

  → Το αντιμετωπίζουμε με χρήση στατιστικών δοκιμασιών

- **Συγχυτικοί παράγοντες**
  - Η σχέση μεταξύ του παράγοντα και της έκβασης συγχέεται από την παρουσία μίας τρίτης μεταβλητής

Συγχυτικός Παράγοντας

Παράγοντας Α

Ασθένεια

Πιθανός συγχυτικός παράγοντας

Παράγοντας Β

Κατανάλωση καφέ

Καρκίνος πνεύμονα

Κάπνισμα

Το κάπνισμα είναι πιθανός συγχυτικός παράγοντας στη σχέση καφέ-καρκίνου

Πως μπορώ να ελέγξω τη σχέση καφέ-καρκίνου λαμβάνοντας υπόψη το κάπνισμα;

- Θα ελέγξω τη σχέση καφέ-καρκίνου ΧΩΡΙΣΤΑ για καπνιστές και μη καπνιστές → έτσι θα δω την «καθαρή» επίδραση του καφέ στον καρκίνο
- Διαστρωμάτωση ή στρωματοποίηση (stratification)

Στρωματοποίηση σύμφωνα με καπνιστικές συνήθειες

Ανάμεσα στα άτομα με καρκίνο του πνεύμονα οι 61 είναι καπνιστές (νυν και πρώην) και πίνουν καφέ, 7 είναι μη καπνιστές και πίνουν καφέ, ενώ 17 δεν πίνουν καφέ και δεν καπνίζουν.

Στο δείγμα 105 υγιών ατόμων (controls) οι 15 είναι καπνιστές (νυν και πρώην) και πίνουν καφέ, 18 είναι μη καπνιστές και πίνουν καφέ, ενώ 64 δεν πίνουν καφέ και δεν καπνίζουν.

Μη Καπνιστές

	Καφές	
	Ναι	Όχι
Ασθενείς	7	17
Υγιείς	18	64

X^2=0,2 (Β.Ε.=1)

Καπνιστές

	Καφές	
	Ναι	Όχι
Ασθενείς	61	100-(7+17+61)= 15
Υγιείς	15	105-(18+64+15)= 8

X^2=1,5 (Β.Ε.=1)

Αξιολόγηση X^2

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

0,2 <2,71 & 1,5 <2,71

p >0,10

Η σχέση καφέ-καρκίνου παύει να υπάρχει όταν κοιτάζω χωριστά καπνιστές και μη-καπνιστές

Συμπέρασμα: α) Δεν υπάρχει σχέση μεταξύ κατανάλωσης καφέ και καρκίνου του πνεύμονα

β) το κάπνισμα δρα ως συγχυτικός παράγοντας της σχέσης καφέ-καρκίνου πνεύμονα

Παράδειγμα 5 (α)

100 άτομα ρωτήθηκαν αν αρρώστησαν από κοινό κρυολόγημα το Σεπτέμβριο και το Νοέμβριο. Η πιθανότητα νόσησης στους 2 μήνες διαφέρει σε βαθμό στατιστικά σημαντικό;

Πιθανότητα νόσησης = % ατόμων που νοσούν

↓

Σύγκριση αναλογιών (για τους 2 μήνες)

↓

X^2 τεστ

ΟΜΩΣ, ΣΤΑ ΙΔΙΑ ΑΤΟΜΑ → Κατά ζεύγη X^2

Σεπτέμβριος	Νοέμβριος		
+	+	20	
-	-	60	
+	-	5	ε
-	+	15	ζ

ΣΤΑ ΙΔΙΑ ΑΤΟΜΑ → Κατά ζεύγη X^2

$$X^2 = \frac{(\varepsilon - \zeta)^2}{\varepsilon + \zeta}$$

Πιθανότητα νόσησης = % ατόμων που νοσούν

P(νόσος Σεπτέμβριο)=(20+5)/100=25%

P(νόσος Νοέμβριο)=(20+15)/100=35%

Τυχαία ή στατιστικά σημαντική διαφορά;

Σεπτέμβριος	Νοέμβριος		
+	+	20	
-	-	60	
+	-	5	ε
-	+	15	ζ

Με διόρθωση κατά Yates:

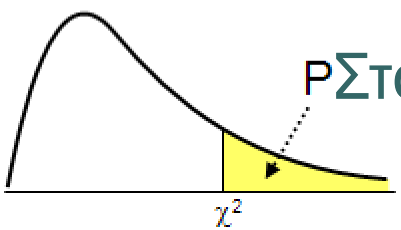
$$X^2 = \frac{(|\varepsilon - \zeta| - 1)^2}{\varepsilon + \zeta} = \frac{(|5 - 15| - 1)^2}{20} = \frac{81}{20} = 4{,}05$$

Β.Ε.=1

Πως κρίνω αν η διαφορά είναι στατιστικά σημαντική;

- Ανατρέχω στον αντίστοιχο πίνακα
  - Βαθμοί ελευθερίας = 1

Chi-square distribution table



4,05>3.84

Στατιστικά σημαντικό στο 5% (p<0.05)

	P										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124

Συμπέρασμα: Η πιθανότητα νόσησης στους 2 μήνες

διαφέρει σε βαθμό στατιστικά σημαντικό.

P(νόσος Σεπτέμβρη)=(20+5)/100=25%

P(νόσος Νοέμβρη)=(20+15)100=35%

Ερμηνεία: Η πιθανότητα νόσησης το Νοέμβριο είναι μεγαλύτερη (σε βαθμό στατιστικά σημαντικό) από την πιθανότητα νόσησης το Σεπτέμβριο.

Αν δεν υπήρχε αντιστοιχία κατά ζεύγη, τι θα έβρισκα;

Αναλογίες που συγκρίνονται:

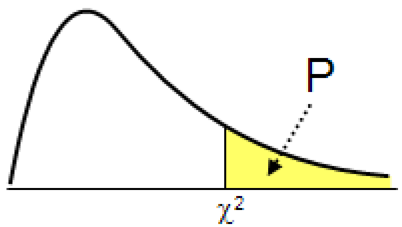
$\frac{25}{100} =$ αναλογία ατόμων που αρρώστησαν τον Σεπτέμβριο

$\frac{35}{100} =$ αναλογία ατόμων που αρρώστησαν τον Νοέμβριο

	Σεπτέμβριος	Νοέμβριος	Σύνολο
+	25	35	60
-	75	65	140
Σύνολο	100	100	200

$$X^2 = \frac{\left(|ad-bc|-\frac{n}{2}\right)^2 n}{(a+b)(c+d)(a+c)b+d)} = \frac{(|25\cdot 65-35\cdot 75|-100)^2\cdot 200}{60\cdot 140\cdot 100\cdot 100} = 1,93$$

Chi-square distribution table



1.93<3.84

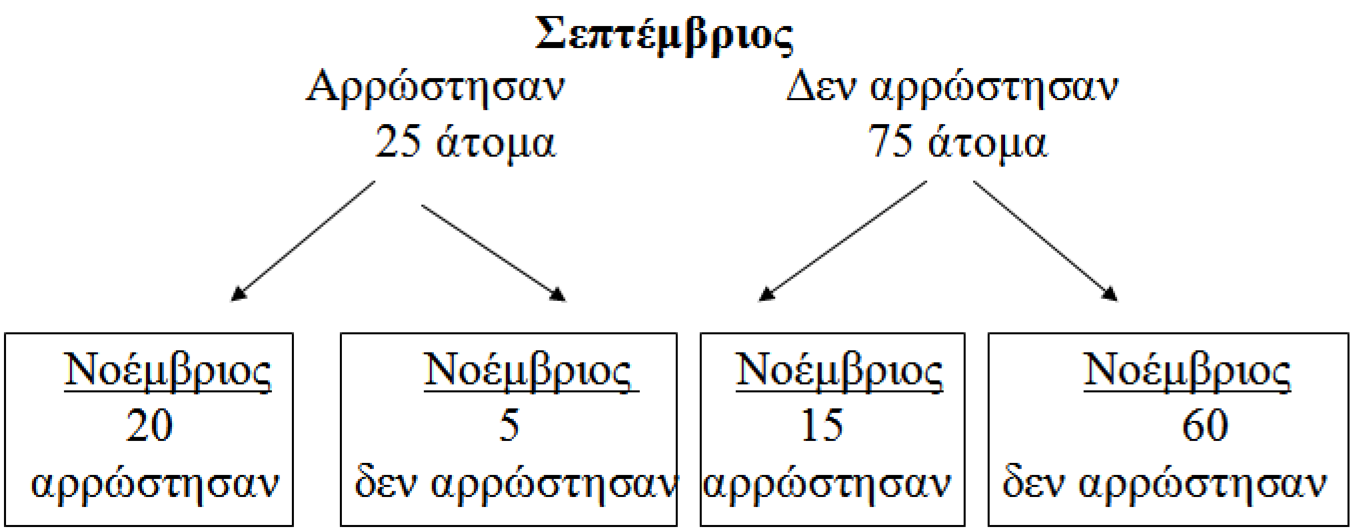
p>0.05 (μη στατιστικά σημαντικό)

	P										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124

→ Κατά ζεύγη X^2 πιο ισχυρό

Παράδειγμα 5 (β)

Τα άτομα που αρρώστησαν το Σεπτέμβριο έχουν διαφορετική πιθανότητα να αρρωστήσουν και το Νοέμβριο συγκριτικά με τα άτομα που δεν αρρώστησαν το Σεπτέμβριο της ίδιας χρονιάς;

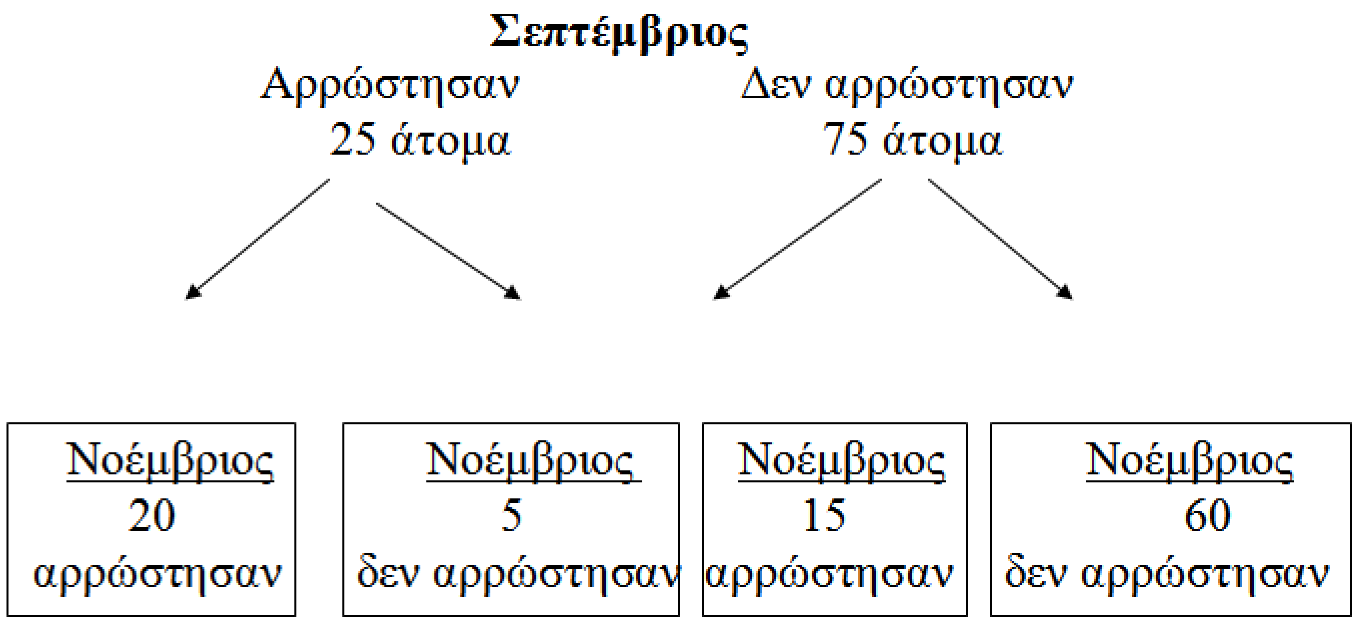


- αναλογία αυτών που αρρώστησαν το Νοέμβριο στο σύνολο όσων αρρώστησαν το Σεπτέμβριο

<u>20</u> <u>άτομα που αρρώστησαν και το Νοέμβριο</u>
25 άτομα που αρρώστησαν το Σεπτέμβριο

-αναλογία αυτών που αρρώστησαν το Νοέμβριο στο σύνολο όσων δεν αρρώστησαν το Σεπτέμβριο

<u>15</u> <u>άτομα που αρρώστησαν μόνο Νοέμβριο</u>
75 άτομα που δεν αρρώστησαν το Σεπτέμβριο



Σεπτέμβριος	Νοέμβριος		Σύνολο
	Ναι	Όχι	
Ναι	20	5	25
Όχι	15	60	75
Σύνολο	35	65	100

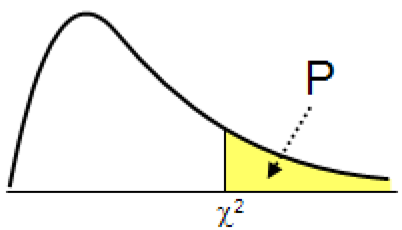
Δοκιμασία X^2 ως κριτήριο διαφοράς δύο αναλογιών (ποσοστών)

Σεπτέμβριος	Νοέμβριος		Σύνολο
	Ναι	Όχι	
Ναι	20	5	25
Όχι	15	60	75
Σύνολο	35	65	100

$$X^2 = \frac{\left(\left|20 \cdot 60 - 15 \cdot 5\right| - 50\right)^2 100}{35 \cdot 65 \cdot 25 \cdot 75} = 27,09$$

Βαθμοί ελευθερίας: (2-1)x(2-1)=1

Chi-square distribution table



DF	P 0.995	0.975	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.002	0.001
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124

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

27,09>10,83

p<0.1%

Συμπέρασμα: Τα άτομα που αρρώστησαν το Σεπτέμβριο έχουν σε βαθμό στατιστικά σημαντικό διαφορετική πιθανότητα να αρρωστήσουν και το Νοέμβριο συγκριτικά με τα άτομα που δεν αρρώστησαν το Σεπτέμβριο της ίδιας χρονιάς.

Ερμηνεία: Τα άτομα που αρρώστησαν το Σεπτέμβριο έχουν σημαντικά μεγαλύτερη πιθανότητα να αρρωστήσουν και το Νοέμβριο συγκριτικά με τα άτομα που δεν αρρώστησαν το Σεπτέμβριο.

Παράδειγμα 6

- Θνητότητα
  - Συνήθης αγωγή =50%
  - Νέα αγωγή: 76 θάνατοι σε 200 άτομα =38%

- Διαφέρουν στατιστικά σημαντικά;

Αγωγή (Νέα/Συνήθης) → Θάνατος (ναι/όχι)

Ποιοτική

Ποιοτική

X^2-test

Ποιο είναι το πρόβλημα που παρατηρείτε εδώ?

Αγωγή	Θάνατος		Σύνολο
	Ναι	Όχι	
Νέα	76	124	200
Συνήθης	?	?	?

Εναλλακτικός τρόπος

Συνήθης αγωγή: 50% → Δοθείσα αναλογία p_0

Νέα αγωγή: 76 θάνατοι σε 200 =38% → Αναλογία p_1

- Σύγκριση της αναλογίας p_1 με τη δοθείσα αναλογία p_0:

$$\frac{|p_1 - p_0|}{SE_p}$$

Η ποσότητα αυτή συγκρίνεται με οριακές τιμές από πίνακα t-κατανομής και αν είναι μεγαλύτερη από την οριακή τιμή → η αναλογία p_1 διαφέρει από τη δοθείσα αναλογία p_0 στατιστικά σημαντικά

Προϋποθέσεις: $0,1 < p < 0,9$ και $n > 40$

ή $n * p > 5$ και $n * q > 5$

Σύγκριση της p_1 με τη δοθείσα αναλογία p_0

$$\frac{|p_1 - p_0|}{SE_p}$$

- Όταν συγκρίνουμε με μία δεδομένη αναλογία, τότε χρησιμοποιούμε αυτή για να προσδιορίσουμε το πιθανό σφάλμα, δηλαδή:

$$SE_p = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{0{,}50 * 0{,}50}{200}} = 0{,}0354$$

$$\frac{|\mathrm{p}_1 - \mathrm{p}_0|}{\mathrm{SE_P}} = \frac{|0{,}38 - 0{,}50|}{0{,}0354} = \frac{0{,}12}{0{,}0354} = 3{,}3898$$

Σύγκριση της p_1 με τη δοθείσα αναλογία p_0

Αξιολόγηση

Πίνακας t-test, στους ∞ Β.Ε.

ΒΕ	10%	5%	1%	0.1%
∞	1.65	1.96	2.58	3.29

3,39 > 3.29

$p < 0.1\%$

Συμπέρασμα:

Υπάρχει πάρα πολύ ισχυρά στατιστικά σημαντική διαφορά ανάμεσα στη θνητότητα της συνηθισμένης και της νέας θεραπευτικής αγωγής

Ερμηνεία:

Η νέα θεραπευτική αγωγή υπερέχει σε βαθμό πολύ στατιστικά σημαντικό της συνηθισμένης αγωγής.

95% όρια αξιοπιστίας της αναλογίας

$$p_1 \pm 1.96 * SE_{P_1}$$

Αφού η p_1 βρέθηκε να διαφέρει στατιστικά σημαντικά :

$$SE_{P_1} = \sqrt{\frac{p_1 * q_1}{n_1}} = \sqrt{\frac{0,38 * 0,62}{200}} = 0,0343$$

95% CI: 0,38 ± 1,96 * 0,0343

0,3128 0,4472

Αναμενόμενος αριθμός θανάτων σε ένα δείγμα 150 ατόμων

- Θνητότητα με νέα θεραπεία: 38% (31.3%, 44.7%)
- Σε 150 άτομα:

38% (31.3%, 44.7%)

57 47 67

Ερμηνεία: Αν η νέα θεραπευτική αγωγή εφαρμοστεί σε 150 άτομα ο αναμενόμενος αριθμός θανάτων είναι 57 άτομα με 95% όρια αξιοπιστίας 47-67 άτομα.

Μήνες	Ι	Φ	Μ	Α	Μ	Ι	Ι	Α	Σ	Ο	Ν	Δ	Σύνολο
Γεννήσεις	5	7	13	8	12	9	13	7	6	10	12	18	120

X^2 ΚΑΛΗΣ ΕΦΑΡΜΟΓΗΣ

H_0: Η κατανομή των γεννήσεων ακολουθεί την ομοιόμορφη κατανομή

H_α: Η κατανομή των γεννήσεων ΔΕΝ ακολουθεί την ομοιόμορφη κατανομή

$$X^2 = \sum_{i=1}^{12} \frac{(O-E)^2}{E}$$

Ποιες οι αναμενόμενες συχνότητες (Ε) αν οι γεννήσεις κατανέμονταν ομοιόμορφα στους 12 μήνες?

Μήνες	Ι	Φ	Μ	Α	Μ	Ι	Ι	Α	Σ	Ο	Ν	Δ	Σύνολο
παρατ.	5	7	13	8	12	9	13	7	6	10	12	18	120
αναμ.	10	10	10	10	10	10	10	10	10	10	10	10	120

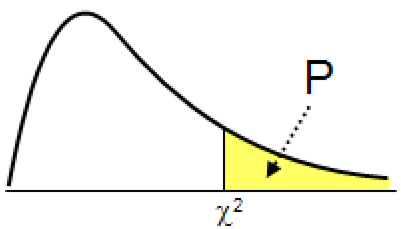
$$X^2 = \sum_{i=1}^{12} \frac{(O-E)^2}{E}$$

$$X^2 = \frac{25+9+9+4+4+1+9+9+16+0+4+64}{10} = \frac{154}{10} = 15,4$$

Β.Ε.: Αριθμός κελιών - αριθμός δεδομένων που χρησιμοποιήθηκαν για υπολογισμό των αναμενόμενων συχνοτήτων.

ΒΕ = 12-1=11 (ομοιόμορφη κατανομή)

Chi-square distribution table



15.4 <19.68

p>0.05 (όχι στατιστικά σημαντικό)

	P										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124
9	1.735	2.700	12.242	14.684	16.919	19.023	19.679	21.666	23.589	26.056	27.877
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588
11	2.603	3.816	14.631	17.275	19.675	21.920	22.618	24.725	26.757	29.354	31.264
12	3.074	4.404	15.812	18.549	21.026	23.337	24.054	26.217	28.300	30.957	32.909
13	3.565	5.009	16.985	19.812	22.362	24.736	25.472	27.688	29.819	32.535	34.528

Συμπέρασμα: Δεν υπάρχει στατιστικά σημαντική διαφορά ανάμεσα στην παρατηρηθείσα και την αναμενόμενη, με βάση την ομοιόμορφη κατανομή, κατανομή των γεννήσεων.

Ερμηνεία: Η κατανομή των γεννήσεων ακολουθεί την ομοιόμορφη κατανομή. Δεν φαίνεται να υπάρχει σημαντική διαφοροποίηση στις γεννήσεις ανάλογα με τον μήνα.