

Στατιστικές Υποθέσεις III

*(Ερευνητικά Ερωτήματα / Υποθέσεις
προς επιβεβαίωση)*

Περίπτωση 3: Ποιοτική με Ποιοτική

- Έστω ότι η μία ερώτηση Q1 είναι **ποιοτική** και η Q2 είναι **ποιοτική** (τα χαρακτηριστικά μπορεί να είναι και τα δύο ονομαστικά ή και τα δύο διατάξιμα ή ένα ονομαστικό κι ένα διατάξιμο)
- Τότε **ΣΧΕΣΗ** μεταξύ των ερωτήσεων, σημαίνει τα ποσοστά των διαφόρων κατηγοριών της Q1, ΔΕΝ είναι τα ίδια σε όλες τις κατηγορίες της Q2
- Δηλαδή??

Παράδειγμα 1

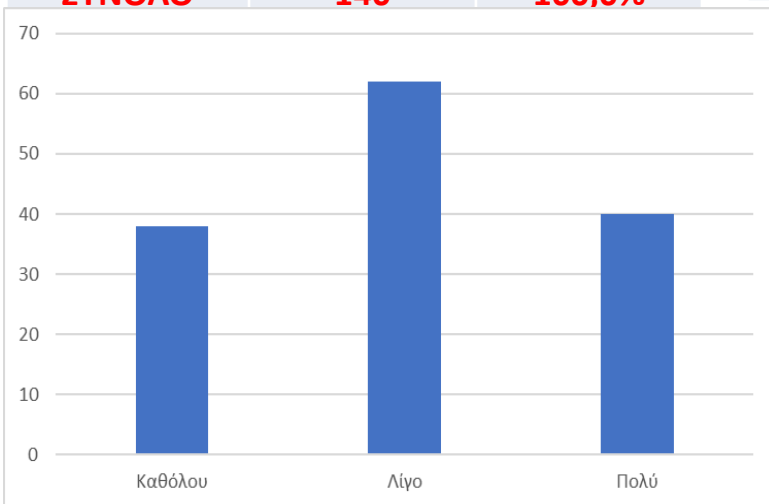
- Σε ένα ερωτηματολόγιο που δόθηκε σε 140 κοινωνικούς λειτουργούς, καταγράφηκε η δομή που εργαζόταν και η ικανοποίηση τους από τις συνθήκες εργασίας

<u>ΙΚΑΝΟΠΟΙΗΣΗ</u>	Μονάδες απεξάρτησης	Νοσοκομεία / Κ.Υ.	Τοπ. Αυτοδιοίκηση	Σχολεία
Καθόλου	7	14	12	5
Λίγο	22	14	14	12
Πολύ	6	9	15	10

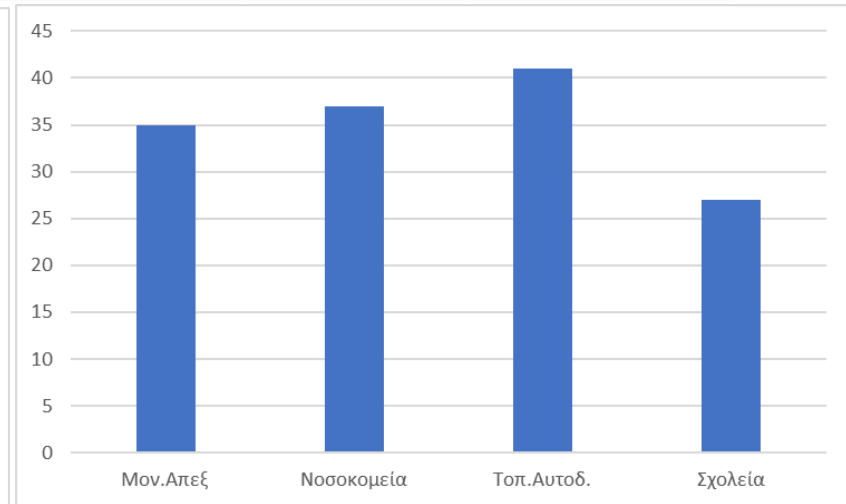
Πίνακας Συνάφειας (CROSSTAB)

Αν είχαμε μελετήσει χωριστά τις ερωτήσεις:

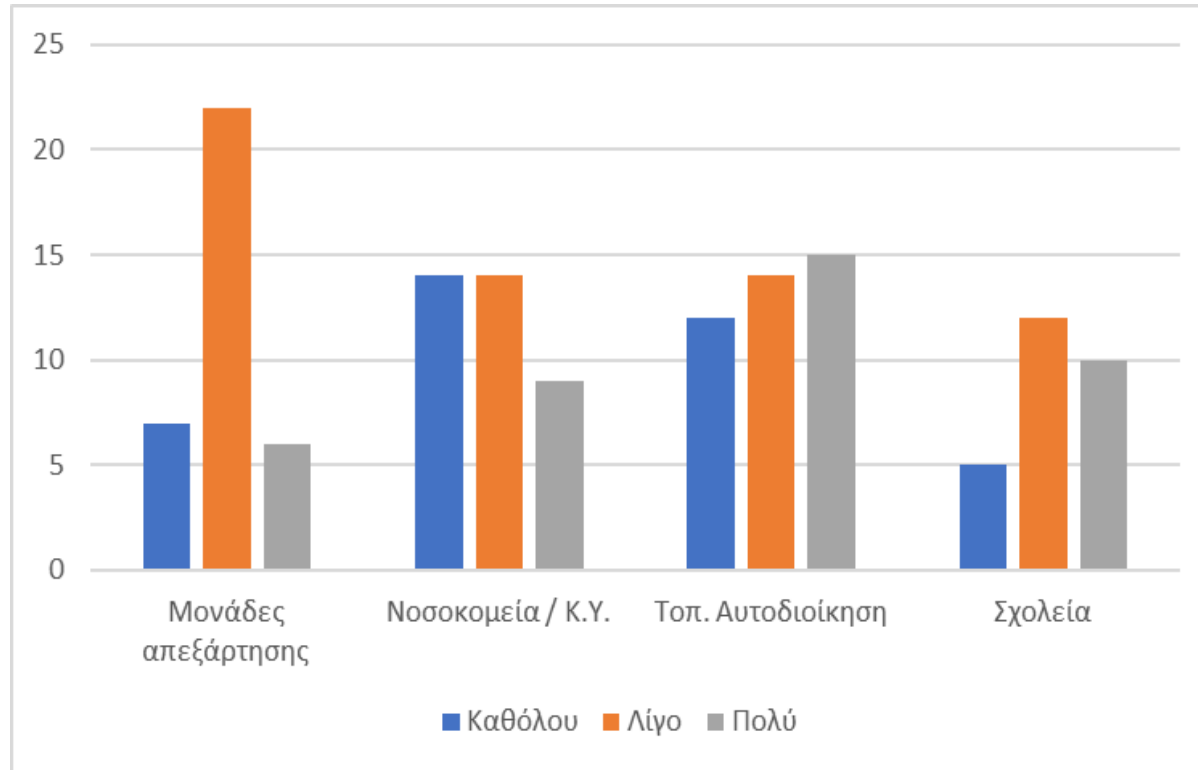
Βαθμός ικανοπ.	Συχνότητα	Σχ.Συχνότητα
Καθόλου	38	27,1%
Λίγο	62	44,3%
Πολύ	40	28,6%
ΣΥΝΟΛΟ	140	100,0%



Δομή	Συχνότητα	Σχ.Συχνότητα
Μον.Απεξ	35	25,0%
Νοσοκομεία	37	26,4%
Τοπ.Αυτοδ.	41	29,3%
Σχολεία	27	19,3%
ΣΥΝΟΛΟ	140	100,0%



Το Σύνθετο Ραβδόγραμμα



Τι συμπεράσματα βγάζουμε από αυτό το ραβδόγραμμα;;

- **Ερώτημα:**

Έχει σχέση ο βαθμός ικανοποίησης με την δομή που εργάζονται;

Τι σημαίνει «σχέση»; -

Πρέπει να παρατηρήσουμε το ραβδόγραμμα ή να ανάγουμε τις αρχικές μας μετρήσεις του πίνακα σε ΠΟΣΟΣΤΑ.

Πως βρίσκουμε ποσοστά σε ένα πίνακα διπλής εισόδου;

1. Χρειαζόμαστε τα ΣΥΝΟΛΑ, γραμμών και στηλών

-	Μονάδες απεξάρτησης	Νοσοκομεία / Κ.Υ.	Τοπ. Αυτοδιοίκηση	Σχολεία	ΣΥΝΟΛΟ
Καθόλου	7	14	12	5	38
Λίγο	22	14	14	12	62
Πολύ	6	9	15	10	40
ΣΥΝΟΛΟ	35	37	41	27	140

2. Στη συνέχεια αποφασίζουμε αν θα βγάλουμε ποσοστά ως προς τις **γραμμές (rows)** ή ως προς τις **στήλες (columns)**

3. Αν αποφασίσουμε να βρούμε ποσοστά ως προς τις στήλες, τότε ΔΙΑΙΡΟΥΜΕ τις συχνότητες που έχουμε στον πίνακα με τα αντίστοιχα σύνολα των στηλών (και πολλαπλασιάζουμε με το 100) – π.χ. $7/35 \cdot 100$, $22/35 \cdot 100$ κ.λ.π.

Ποσοστά ως προς στήλες

<u>ΙΚΑΝΟΠΟΙΗΣΗ</u>	Μονάδες απεξάρτησης	Νοσοκομεία / Κ.Υ.	Τοπ. Αυτοδιοίκηση	Σχολεία
Καθόλου	20,0%	37,8%	29,3%	18,5%
Λίγο	62,9%	37,8%	34,1%	44,4%
Πολύ	17,1%	24,3%	36,6%	37,0%
ΣΥΝΟΛΟ	100%	100%	100%	100%

- Όταν τα ποσοστά κατανέμονται με τον ίδιο τρόπο στις τέσσερις δομές, ΤΟΤΕ ΔΕΝ υπάρχει σχέση ανάμεσα στο βαθμό ικανοποίησης και την δομή
- Στο παράδειγμα μας όμως έχουμε

Μ.απεξάρτησης	20% - 63% - 17%
Νοσοκομεία	38% - 38% - 24%
Τοπ.Αυτοδιοίκηση	29% - 34% - 37%
Σχολεία	19% - 44% - 37%

Φαίνεται ότι η κατανομή των ποσοστών δεν είναι η ίδια, άρα υπάρχει μια σχέση (ΣΤΟ ΔΕΙΓΜΑ)

Δείγμα $\rightarrow$ Πληθυσμός

- Να θυμηθούμε πάντα πως παίρνουμε δείγματα και αυτό που βλέπουμε στα δείγματα δεν ισχύει πάντα στον πληθυσμό
- Βλέπουμε λοιπόν μια σχέση στα δείγματα, πως θα εξασφαλίσουμε πως αυτή η σχέση θα υπάρχει και στον πληθυσμό;

πως ελέγχεται η σημαντικότητα της σχέσης;

Υπάρχει πάλι το μαγικό νούμερο p (p -value), που μας δίνει την απάντηση (κατάλληλη διαδικασία: **έλεγχος χ^2 – Chi Square Test**):

Αν $p < 0,05$, τότε τα ποσοστά είναι διαφορετικά και στον πληθυσμό, άρα τα χαρακτηριστικά που μελετήσαμε έχουν κάποια σχέση! (τα αποτελέσματα μας ήταν στατιστικά σημαντικά)

Αν $p > 0,05$, τότε τα ποσοστά στον πληθυσμό είναι πιθανόν να είναι τα ίδια, άρα τα χαρακτηριστικά που μελετήσαμε δεν έχουν κάποια σχέση! (τα αποτελέσματα μας ήταν τυχαία)

Αποτελέσματα Απόφαση

$$\chi^2 = 10,48$$

$$P = 0,11$$

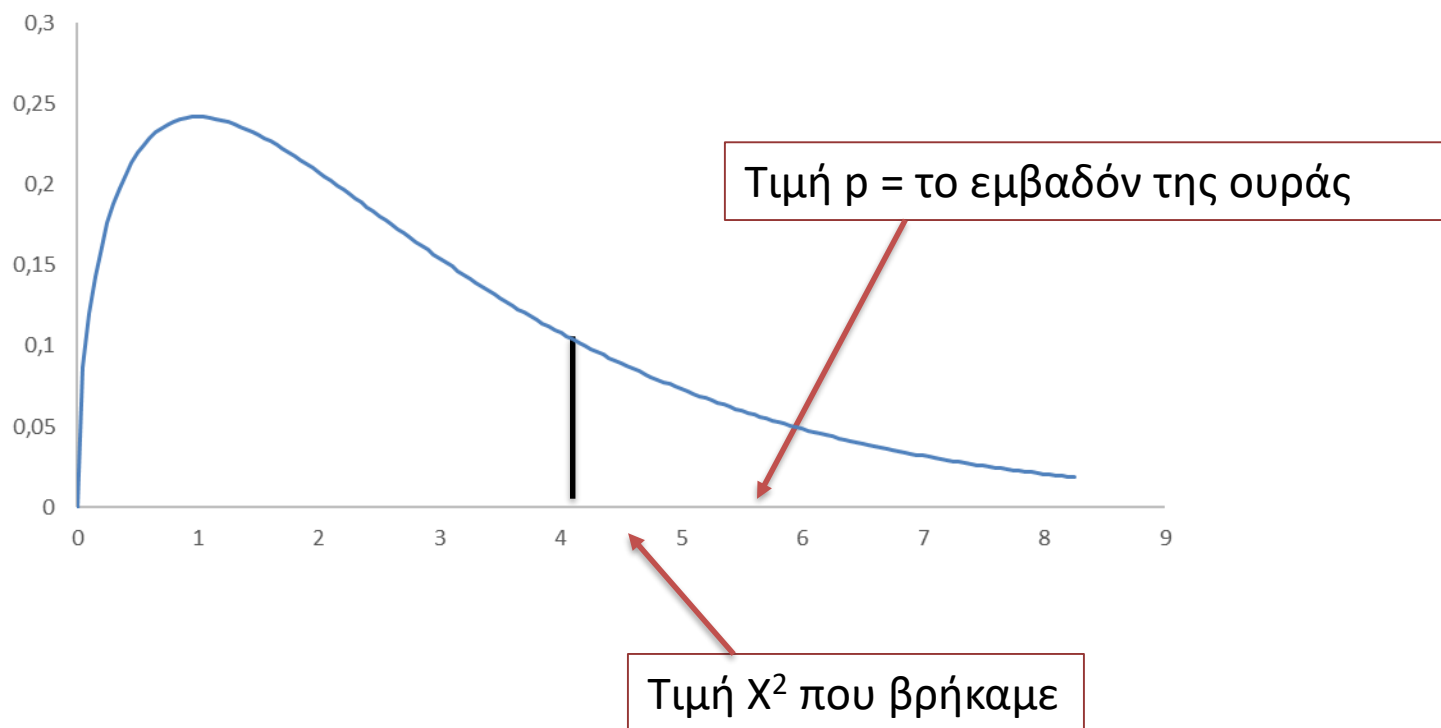
- Αφού $p > 0,05$ τότε δεν φαίνεται να υπάρχει κάποια σχέση ανάμεσα στο βαθμό ικανοποίησης και στην δομή που εργάζονται οι κοινωνικοί λειτουργοί
- Οι διαφοροποιήσεις σε ποσοστά που βρήκαμε στο δείγμα είναι μάλλον τυχαία

Πως ο υπολογιστής βρίσκει το p ;

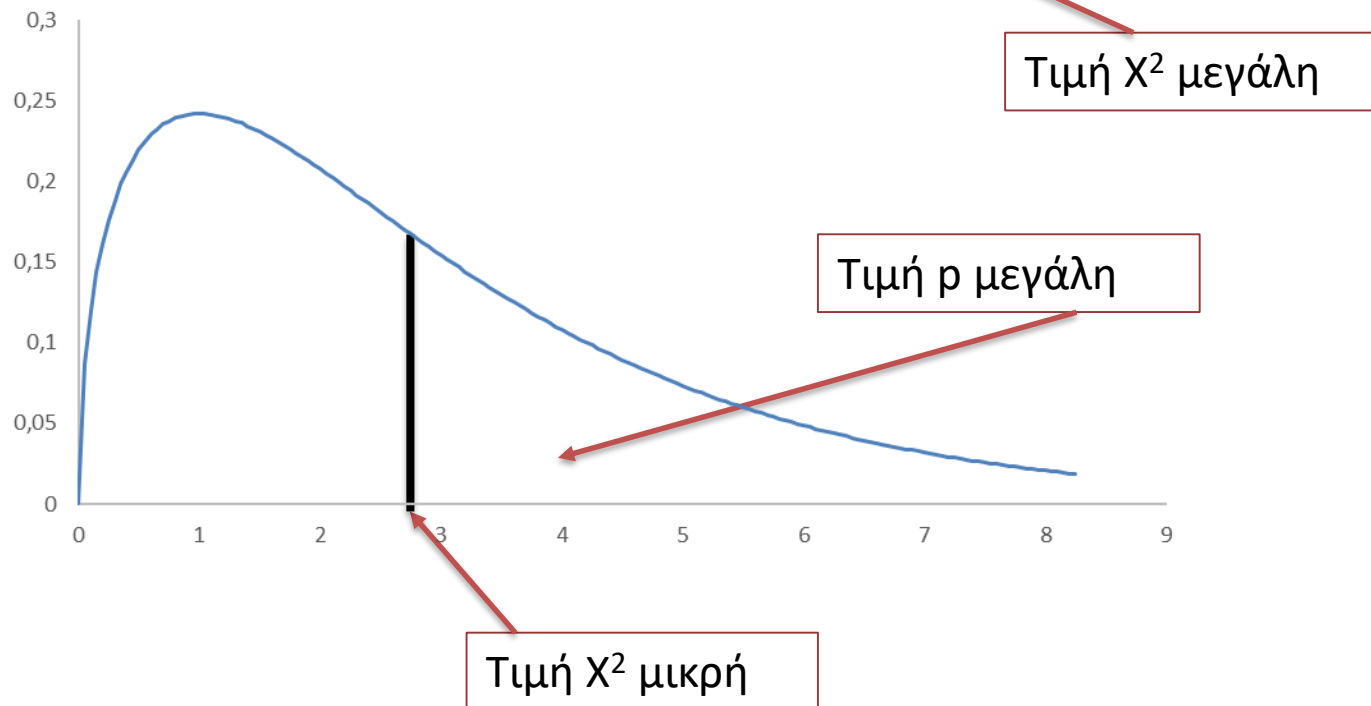
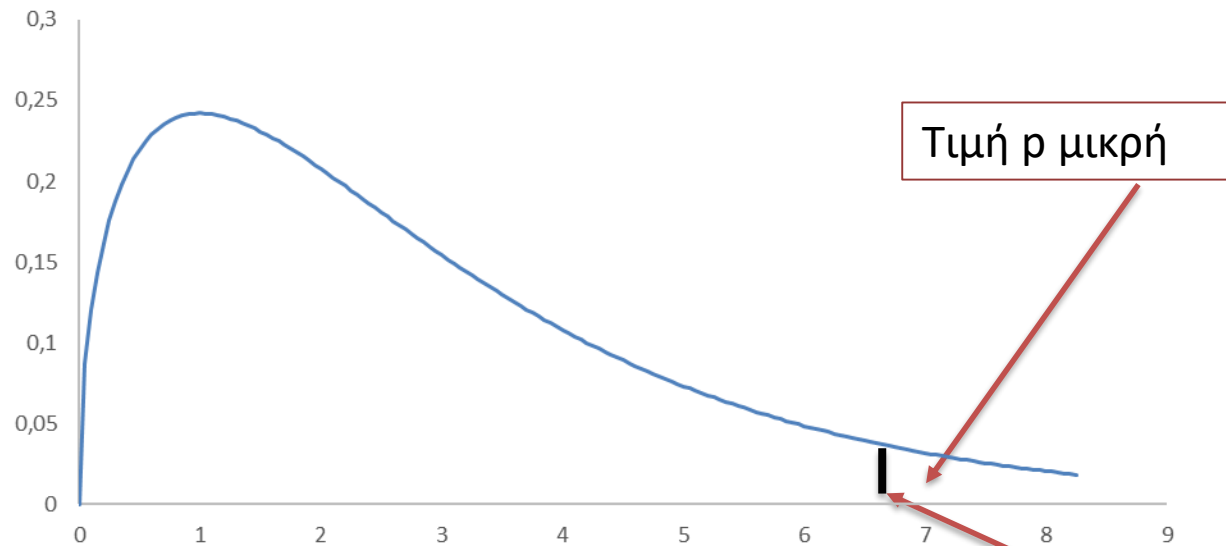
- Όπως στο t -test χρησιμοποιείται η ποσότητα t , στην ANOVA χρησιμοποιείται η F , έτσι σε αυτή τη διαδικασία χρησιμοποιείται μια ποσότητα που ονομάζεται χ^2
- Αφού υπολογιστεί το χ^2 , βρίσκεται ένα εμβαδό στην κατανομή, που ονομάζεται χ^2

Το χ^2 και το p

χ^2 κατανομή



χ^2 και p είναι
αντιστρόφως
ανάλογα (όπως
συμβαίνει και με
το t και το F)



Από τι εξαρτάται το χ^2

Το πόσο μεγάλο θα βγεί το χ^2 και συνεπώς πόσο μικρό θα είναι το p , εξαρτάται

1. Από το πόσο διαφορετικές είναι οι κατανομές των ποσοστών
2. Από το πόσο μεγάλο είναι το δείγμα μας

Παράδειγμα 2

Σε μια μελέτη με ερωτηματολόγια, υπήρχαν δημογραφικά στοιχεία (π.χ. φύλο) και ερωτήσεις όπως η παρακάτω: Πόσο συχνά επισκέπτεστε χώρους πρασίνου στην πόλη σας;

Συνδυάζοντας την ερώτηση αυτή με το φύλο είχαμε:

επισκέψεις σε χώρους πρασίνου	Φύλο		ΣΥΝΟΛΟ
	Γυναίκες	Άνδρες	
Καθημερινά	56	30	86
Μια φορά τη βδομάδα	130	72	202
Μια φορά το μήνα	43	17	60
Λιγότερο από μία φορά το μήνα	33	14	47
ΣΥΝΟΛΟ	262	133	395

Ποσοστά για το Παράδειγμα 2

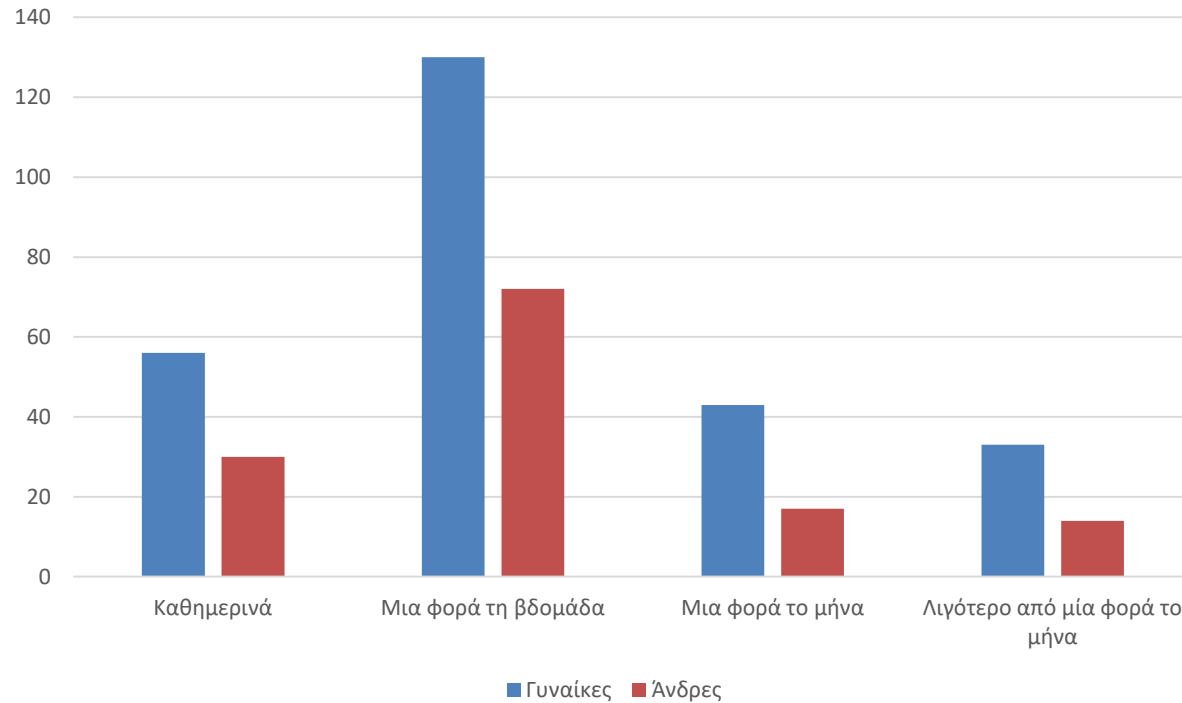
επισκέψεις σε χώρους πρασίνου	Φύλο		ΣΥΝΟΛΟ
	Γυναίκες	Άνδρες	
Καθημερινά	21,4%	22,6%	21,8%
Μια φορά τη βδομάδα	49,6%	54,1%	51,1%
Μια φορά το μήνα	16,4%	12,8%	15,2%
Λιγότερο από μία φορά το μήνα	12,6%	10,6%	11,9%
ΣΥΝΟΛΟ	100,0%	100,0%	100,0%

Παρατηρούμε ότι οι κατανομές των ποσοστών σε κάθε στήλη δεν είναι ίδιες στο Δείγμα μας. Οι διαφορές βέβαια δεν είναι μεγάλες, π.χ.

49,6% των γυναικών επισκέπτεται χώρους πρασίνου μια φορά τη βδομάδα, ενώ στους άνδρες το ποσοστό αυτό είναι 54,1%

Κλπ...

Σύνθετο Ραβδόγραμμα



Αποτελέσματα

Chi-Square Tests

	Value	df	Asymp. Sig. (2- sided)
Pearson Chi-Square	2,418 ^a	3	,490
Likelihood Ratio	2,482	3	,479
Linear-by-Linear Association	1,745	1	,187
N of Valid Cases	395		

Τιμή p

Τιμή χ^2

Συμπέρασμα

- Αφού το $p=0,490 > 0,05$, άρα οι ερωτήσεις δεν έχουν σχέση μεταξύ τους.
- Αυτό σημαίνει πως αυτό που βρήκαμε στο δείγμα μας ήταν τυχαίο.
- Δηλαδή, οι (μικρές) διαφορές που είδαμε ανάμεσα σε άντρες και γυναίκες αναφορικά με τη συχνότητα επίσκεψης σε χώρους πρασίνου δεν είναι στατιστικά σημαντικές και δεν αξιολογούνται.

Ποσοστά ως προς στήλες ή ως προς γραμμές;

επισκέψεις σε χώρους πρασίνου	Φύλο		ΣΥΝΟΛΟ
	Γυναίκες	Άνδρες	
Καθημερινά	21,4%	22,6%	21,8%
Μια φορά τη βδομάδα	49,6%	54,1%	51,1%
Μια φορά το μήνα	16,4%	12,8%	15,2%
Λιγότερο από μία φορά το μήνα	12,6%	10,6%	11,9%
ΣΥΝΟΛΟ	100,0%	100,0%	100,0%

επισκέψεις σε χώρους πρασίνου	Φύλο		ΣΥΝΟΛΟ
	Γυναίκες	Άνδρες	
Καθημερινά	65,1%	34,9%	100%
Μια φορά τη βδομάδα	64,4%	35,6%	100%
Μια φορά το μήνα	71,7%	28,3%	100%
Λιγότερο από μία φορά το μήνα	74,4%	25,6%	100%
ΣΥΝΟΛΟ	66,8%	33,2%	100,0%

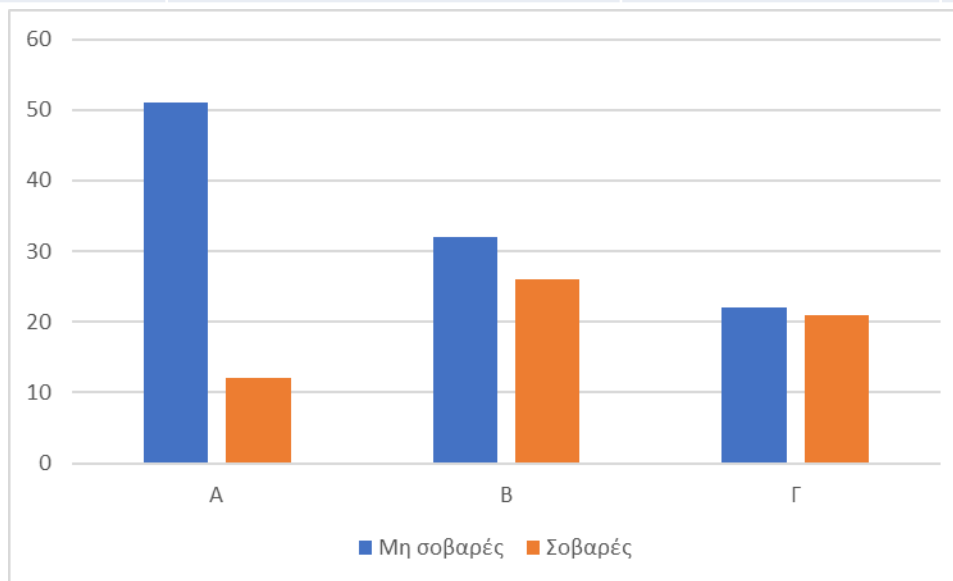
Παράδειγμα 3

Σε μια έρευνα που έγινε σε σχέση με τις παρενέργειες τριών φαρμάκων για την υπέρταση είχαμε τα αποτελέσματα

	Παρενέργειες	
Φάρμακα	Μη σοβαρές	Σοβαρές
A	51	12
B	32	26
Γ	22	21

Πίνακας Ποσοστών

	Παρενέργειες		
Φάρμακα	Μη σοβαρές	Σοβαρές	ΣΥΝΟΛΟ
A	81,0%	19,0%	100%
B	55,2%	44,8%	100%
Γ	51,2%	48,8%	100%



Αποτελέσματα - Απόφαση

$$\chi^2 = 12,90$$

$$P = 0,0016$$

- Αφού $p < 0,05$ τότε φαίνεται ότι υπάρχει σχέση ανάμεσα στις παρενέργειες και το είδος φαρμάκου που χορηγήθηκε
- από τον πίνακα ποσοστών φαίνεται ότι το φάρμακο Α, έχει τις λιγότερο σοβαρές παρενέργειες από τα άλλα δύο.