

# Ανάλυση δεδομένων στο περιβάλλον του SPSS

Λαβίδας Κωνσταντίνος  
Μαθηματικός

[lavidas@upatras.gr](mailto:lavidas@upatras.gr)

# Ο έλεγχος $\chi^2$ – chi square

- Αναπτύχθηκε από τον Karl Pearson το 1900.
- Είναι μη παραμετρικό τεστ
- Χρησιμοποιούνται συχνότητες και όχι μετρήσεις
- Γενικά χρησιμοποιείται για να μετρήσει τη διαφορά μεταξύ των παρατηρούμενων και των αναμενόμενων συχνοτήτων.
- Τρεις μορφές ελέγχου:
  1. **Ομοιογένειας (Homogeneity test)**
    - Έλεγχος της ύπαρξης διαφοράς ανάμεσα στις ομάδες που συγκροτεί η μια μεταβλητή ως προς την άλλη.
  2. **Ανεξαρτησίας (Συνάφειας - Contingency test)**
    - Έλεγχος της σχέσης μεταξύ δύο ποιοτικών μεταβλητών
  3. **Καλής προσαρμογής (Goodness of Fit test)**
    - Έλεγχος του πόσο καλά προσαρμόζεται «ταιριάζει» η κατανομή των παρατηρήσεων μας με μια γνωστή θεωρητική κατανομή
- Οι περιπτώσεις 1 και 2 βασίζονται στην ίδια υπολογιστική τεχνική και έτσι εξετάζονται με τον ίδιο τρόπο η μόνη διαφορά έγκειται στην διατύπωση της υπόθεσης.

# Έλεγχος της ανεξαρτησίας δύο ποιοτικών μεταβλητών

# Έλεγχος χ<sup>2</sup> ανεξαρτησίας (1/2)

- Βασίζεται στον υπολογισμό ενός μέτρου (χ<sup>2</sup>) από τις συχνότητες του πίνακα συνάφειας
  - Ελέγχει πόσο απέχουν οι παρατηρούμενες από τις αναμενόμενες συχνότητες
  - Από τις παρατηρούμενες (**O**bserved) συχνότητες (count ή frequencies) υπολογίζονται οι αναμενόμενες (**E**xpected)

$$E = \frac{(\text{Περιθώρια συχνότητα γραμμής}) \times (\text{Περιθώρια συχνότητα στήλης})}{\text{Σύνολο τιμών πίνακα}}$$

# έλεγχος $\chi^2$ ανεξαρτησίας (2/2)

- Υποθέσεις:
  - Μηδενική: οι δύο μεταβλητές είναι ανεξάρτητες (**δεν σχετίζονται**)
  - Εναλλακτική: οι δύο μεταβλητές **σχετίζονται**

# Η ελεγχοσυνάρτηση $\chi^2$

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

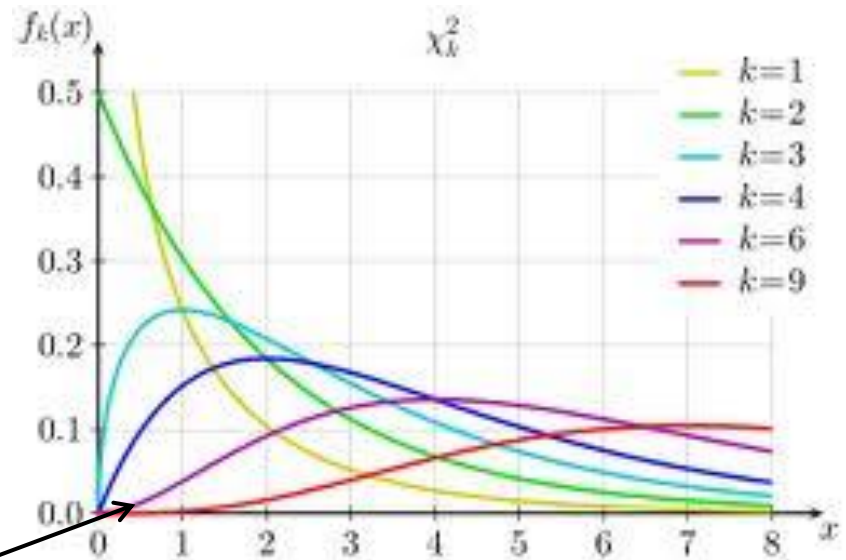
$r$  = αριθμός γραμμών

$c$  = αριθμός στηλών

$O$  = Παρατηρούμενες συχνότητες

$E$  = Αναμενόμενες συχνότητες

$n = rc$



- Ακολουθεί την  $\chi^2$  κατανομή με  $k=(r-1)(c-1)$  βαθμούς ελευθερίας

# Παράδειγμα υπολογισμού της ελεγχοσυνάρτησης $\chi^2$ με το χέρι (1/3)

- Έστω ο πίνακας συνάφειας

αναγνωσιμότητα \* κοινωνικο-οικονομικό επίπεδο  
Crosstabulation

Count		κοινωνικο-οικονομικό επίπεδο			Total
		1 χαμηλό	2 μεσαίο	3 υψηλό	
αναγνωσιμότητα	1 τακτικός	20	45	40	105
	2 περιστασιακός	60	25	30	115
Total		80	70	70	220

Περιθώριες  
συχνότητες  
στήλης

- Από τις παρατηρούμενες θα υπολογίσουμε της αναμενόμενες συχνότητες

Παράδειγμα:

$$E = \frac{(\text{Περιθώρια συχνότητα γραμμής}) \times (\text{Περιθώρια συχνότητα στήλης})}{\text{Σύνολο τιμών πίνακα}}$$

$$E_{1.1} = \frac{80 \cdot 105}{220} \approx 38,2$$

Περιθώριες  
συχνότητες  
γραμμής

# Παράδειγμα υπολογισμού της ελεγκοσυνάρτησης $\chi^2$ με το χέρι(2/3)

- Ο ίδιος πίνακας με της παρατηρούμενες και τις αναμενόμενες συχνότητες

αναγνωσιμότητα \* κοινωνικο-οικονομικό επίπεδο Crosstabulation

			κοινωνικο-οικονομικό επίπεδο			
			1 χαμηλό	2 μεσαίο	3 υψηλό	Total
αναγνωσιμότητα	1 τακτικός	Count	20	45	40	105
		Expected Count	38,2	33,4	33,4	105,0
	2 περιστασιακός	Count	60	25	30	115
		Expected Count	41,8	36,6	36,6	115,0
Total		Count	80	70	70	220
		Expected Count	80,0	70,0	70,0	220,0

Στον πίνακα count και Expected count είναι οι παρατηρούμενες και οι αναμενόμενες συχνότητες αντίστοιχα



# Παράδειγμα υπολογισμού της ελεγκοσυνάρτησης $\chi^2$ με το χέρι(3/3)

## Εφαρμογή του τύπου

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

$r$ = αριθμός γραμμών

$c$ =αριθμός στηλών

$O$ =Παρατηρούμενες συχνότητες

$E$ =Αναμενόμενες συχνότητες

$n=rc$

αναγνωσιμότητα \* κοινωνικο-οικονομικό επίπεδο Crosstabulation

		κοινωνικο-οικονομικό επίπεδο			Total	
		1 χαμηλό	2 μεσαίο	3 υψηλό		
αναγνωσιμότητα	1 τακτικός	Count	20	45	40	105
		Expected Count	38,2	33,4	33,4	105,0
	2 περιστασιακός	Count	60	25	30	115
		Expected Count	41,8	36,6	36,6	115,0
Total		Count	80	70	70	220
		Expected Count	80,0	70,0	70,0	220,0

## Τελικά

$\chi^2=26,75$  μεγαλύτερο από την κρίσιμη τιμή σε επίπεδο  $\alpha=0,05$ . Η Κρίσιμη Τιμή για 2 βαθμούς ελευθερίας και  $\alpha=0,05$  είναι 5,99.

# Ο έλεγχος $\chi^2$ ανεξαρτησίας χρησιμοποιείται για να ελέγξει την ανεξαρτησία ή όχι των ποιοτικών μεταβλητών.

- Δεν δίνει πληροφορίες, για το αν η έλλειψη ανεξαρτησίας αφορά όλο τον πίνακα ή ένα μόνο τμήμα του.
  - Αυτό το προσδιορίζουμε από τα υπόλοιπα (residuals) κυρίως τα προσαρμοσμένα (adjusted) τα οποία ακολουθούν κατά προσέγγιση την τυπική κανονική κατανομή:
    - Τιμές μεγαλύτερες από 1,96 κατά απόλυτη τιμή οδηγούν με μεγαλύτερη ισχύ στην έλλειψη ανεξαρτησίας.
- **Δίνει πληροφορίες, για την ένταση της σχέσης, με μέτρα που βασίζονται στο  $\chi^2$  στατιστικό**
- **Δεν δίνει πληροφορίες, για την φύση της σχέσης, στην περίπτωση δύο κατηγορικών μεταβλητών.**
  - Στην περίπτωση **δύο μεταβλητών διάταξης έχει νόημα η φύση της σχέσης θετική ή αρνητική** και προσδιορίζεται με μέτρα όπως: συντελεστής gamma, Tau-b, Tau-c του Kendall και συντελεστής d του Somers.

# Προϋποθέσεις χρήσης του ελέγχου $\chi^2$ ανεξαρτησίας

- Ο ερευνητής πρέπει να εξασφαλίσει ότι ένας ικανός αριθμός συμμετεχόντων της έρευνας θα βρεθεί σε κάθε κατηγορία. **Συνήθως σε κάθε κατηγορία θα πρέπει να έχουμε περίπου 20 συμμετέχοντες.**
- Οι αναμενόμενες συχνότητες στα κελιά δεν πρέπει να είναι μικρότερες του 1 και
- **Μόνο το πολύ 20% των κελιών θα μπορεί να έχει αναμενόμενη συχνότητα κάτω από 5.**

# Διαδικασία: έλεγχου $\chi^2$ Ανεξαρτησίας

## Έλεγχος της σχέσης μεταξύ δύο ποιοτικών μεταβλητών

### 1<sup>ο</sup> βήμα: Εκτελούμε την διαδικασία $\chi^2$ :

- Η ύπαρξη συστηματικής σχέσης ανιχνεύεται με τον έλεγχο  $\chi^2$  (Chi-Square)
- Αν ο έλεγχος  $\chi^2$  είναι **στατιστικά σημαντικός**
  - προσδιορίζεται μέσω του πίνακα συνάφειας η μορφή της σχέσης

### 2<sup>ο</sup> βήμα: Πίνακας συνάφειας ή διπλής εισόδου (**contingency table or cross tabulation table**):

- Διαβάζοντας τον πίνακα συνάφειας «μάλλον» είμαστε σε θέση να καταλάβουμε τη **μορφή της σχέσης**
  - Στην περίπτωση που ο πίνακας συνάφειας είναι πολύ μεγάλος (πολλές γραμμές και στήλες) αυτό επιτυγχάνεται με επιπλέον διερεύνηση (π.χ. Correspondence analysis)

### 3<sup>ο</sup> βήμα: Προσδιορίζουμε την ένταση της σχέσης

# Πίνακας συνάφειας ή πίνακας σύμπτωσης

## -Contingency table or Cross-tabulation:

- Παρουσιάζει τη **μορφή** της σχέσης (relation or association) μεταξύ δύο ποιοτικών μεταβλητών.
- Έχει τόσες γραμμές (Rows) όσες οι κατηγορίες της μιας μεταβλητής και τόσες στήλες (Columns) όσες οι κατηγορίες της άλλης μεταβλητής.
  - Στα κελιά που δημιουργούνται από το συνδυασμό γραμμών και στηλών γράφουμε τις απόλυτες συχνότητες που αντιστοιχούν στον συνδυασμό των κατηγοριών των δύο μεταβλητών.

# Σύνθετη παρουσίαση δύο ποιοτικών μεταβλητών: Πίνακας συνάφειας

Εξαρτημένη μεταβλητή

(Φύλο) \* (Σειρά επιλογής του Φυσικού τμήματος) Crosstabulation

			Σειρά επιλογής του Φυσικού τμήματος			Σύνολα
			1η θέση	2η θέση	άλλη θέση	
Φύλο	Αγόρι	Count	22	17	20	59
		% within (Φύλο)	37,3%	28,8%	33,9%	100,0%
		% within (Σειρά επιλογής του Φυσικού τμήματος)	53,7%	53,1%	62,5%	56,2%
		% of Total	21,0%	16,2%	19,0%	56,2%
	Κορίτσι	Count	19	15	12	46
		% within (Φύλο)	41,3%	32,6%	26,1%	100,0%
		% within (Σειρά επιλογής του Φυσικού τμήματος)	46,3%	46,9%	37,5%	43,8%
		% of Total	18,1%	14,3%	11,4%	43,8%
	Σύνολα	Count	41	32	32	105
		% within (Φύλο)	39,0%	30,5%	30,5%	100,0%
		% within (Σειρά επιλογής του Φυσικού τμήματος)	100,0%	100,0%	100,0%	100,0%
		% of Total	39,0%	30,5%	30,5%	100,0%

Ανεξάρτητη μεταβλητή

Τιμές στην γραμμή σύνολα (Total) και στην στήλη σύνολα ονομάζονται **περιθώριες** απόλυτες και σχετικές συχνότητες (marginal frequencies) ή περιθώρια κατανομή.

Οι τιμές στο κύριο μέρος του πίνακα αντιστοιχούν σε **συνδυασμό τιμών** της μιας μεταβλητής με την άλλη και αναπαρίστανται με απόλυτες και σχετικές συχνότητες (**joint frequencies**).

## Σύνθετη παρουσίαση δύο ποιοτικών μεταβλητών: Πίνακας συνάφειας

- Απαραίτητη η παρουσίαση των διμεταβλητών **σχετικών συχνοτήτων - % (percent)**.
- Οι διμεταβλητές σχετικές συχνότητες μπορούν να υπολογιστούν:
  - διαιρώντας την απόλυτη συχνότητα του κελιού δια του συνόλου της γραμμής (% within row variable)
  - διαιρώντας την απόλυτη συχνότητα του κελιού δια του συνόλου της στήλης (% within column variable)
  - διαιρώντας την απόλυτη συχνότητα του κελιού δια του γενικού συνόλου (% of total) .

# Σύνθετη παρουσίαση δύο ποιοτικών μεταβλητών: πίνακας συνάφειας

Παράδειγμα: Η σχολική ηλικία των μαθητών και η σημασία που δίνουν στο να παίζουν παιχνίδια.

- Πίνακας συνάφειας

*Πόσο σημαντικό είναι να παίζεις παιχνίδια \* Σχολική Τάξη Crosstabulation*

Count		Σχολική Τάξη			Total
		Ε΄ Δημοτικού	Β΄ Γυμνασίου	Β΄ Λυκείου	
Πόσο σημαντικό είναι να παίζεις παιχνίδια	Καθόλου	1	3	3	7
	Λίγο	0	4	7	11
	Μέτρια	3	12	8	23
	Πολύ	16	15	7	38
	Πάρα πολύ	20	11	6	37
Total		40	45	31	116

*Πόσο σημαντικό είναι να παίζεις παιχνίδια \* Σχολική Τάξη Crosstabulation*

		Σχολική Τάξη			Total	
		Ε΄ Δημοτικού	Β΄ Γυμνασίου	Β΄ Λυκείου		
Πόσο σημαντικό είναι να παίζεις παιχνίδια	Καθόλου	Count	1	3	3	7
		Expected Count	2,4	2,7	1,9	7,0
	Λίγο	Count	0	4	7	11
		Expected Count	3,8	4,3	2,9	11,0
	Μέτρια	Count	3	12	8	23
		Expected Count	7,9	8,9	6,1	23,0
	Πολύ	Count	16	15	7	38
		Expected Count	13,1	14,7	10,2	38,0
	Πάρα πολύ	Count	20	11	6	37
		Expected Count	12,8	14,4	9,9	37,0
Total	Count	40	45	31	116	
	Expected Count	40,0	45,0	31,0	116,0	

Μόνο απόλυτες συχνότητες

Απόλυτες και σχετικές συχνότητες (διά του συνόλου των τιμών κάθε γραμμής)



# Σύνθετη παρουσίαση δύο ποιοτικών μεταβλητών: πίνακας συνάφειας

- Για να προσδιορίσουμε τη **μορφή** της όποιας σχέσης μεταξύ δύο ποιοτικών μεταβλητών:
  - Θα πρέπει να ορίσουμε, έστω και καταχρηστικά, τη μια μεταβλητή ως ανεξάρτητη και την άλλη ως εξαρτημένη.
    - Συνήθως ανεξάρτητη είναι αυτή που προηγείται χρονικά.
    - **Τις τιμές της ανεξάρτητης τις τοποθετούμε στις στήλες**
  - Ελέγχουμε τα ποσοστά που αντιστοιχούν σε κάθε τιμή της «εξαρτημένης» μεταβλητής τα οποία υπολογίστηκαν για κάθε τιμή της εξαρτημένης μεταβλητής επί του συνόλου κάθε κατηγορίας της «ανεξάρτητης» μεταβλητής.
  - Στην περίπτωσή μας οριζόντιο έλεγχο «% within Σχολική Τάξη»

# Σύνθετη παρουσίαση δύο ποιοτικών μεταβλητών: πίνακας συνάφειας

Πόσο σημαντικό είναι να παίζεις παιχνίδια \* Σχολική Τάξη Crosstabulation

			Σχολική Τάξη			Total
			Ε΄ Δημοτικού	Β΄ Γυμνασίου	Β΄ Λυκείου	
Πόσο σημαντικό είναι να παίζεις παιχνίδια	Καθόλου	Count	1	3	3	7
		% within Σχολική Τάξη	2,5%	6,7%	9,7%	6,0%
	Λίγο	Count	0	4	7	11
		% within Σχολική Τάξη	0,0%	8,9%	22,6%	9,5%
	Μέτρια	Count	3	12	8	23
		% within Σχολική Τάξη	7,5%	26,7%	25,8%	19,8%
	Πολύ	Count	16	15	7	38
		% within Σχολική Τάξη	40,0%	33,3%	22,6%	32,8%
	Πάρα πολύ	Count	20	11	6	37
		% within Σχολική Τάξη	50,0%	24,4%	19,4%	31,9%
Total		Count	40	45	31	116
		% within Σχολική Τάξη	100,0%	100,0%	100,0%	100,0%

Μέτρα  
προσδιορισμού της  
έντασης της σχέσης  
μεταξύ δύο  
**ΠΟΙΟΤΙΚΩΝ**  
μεταβλητών

# Ένταση της σχέσης δύο ποιοτικών μεταβλητών

- Μετά την όποια στατιστικά σημαντική σχέση (μορφή της σχέσης) φαίνεται να συνδέει τις δύο ποιοτικές μεταβλητές
- Ενδιαφέρον παρουσιάζει η ένταση της σχέσης (Συμμετρική ή Ασύμμετρη) αυτής
  - Αναφερόμαστε στο μέγεθος της σχέσης
  - Μέτρα συνάφειας

# Συμμετρική και Ασύμμετρη σχέση

- **Συμμετρική (symmetric) σχέση**
  - Αυτό που εξετάζουμε είναι η **σχέση** μεταξύ των μεταβλητών χωρίς να μας ενδιαφέρει η κατεύθυνση της σχέσης
    - Ερωτήματα του τύπου: υπάρχει **σχέση** μεταξύ ...
- **Ασύμμετρη (asymmetric) σχέση**
  - Αυτό που εξετάζουμε είναι η **επίδραση** μιας μεταβλητής πάνω στην άλλη. Μας ενδιαφέρει η κατεύθυνση της σχέσης
    - Ερωτήματα του τύπου: **επιδρά** η μεταβλητή ... πάνω στην μεταβλητή ...

# Μέτρα προσδιορισμού της έντασης της σχέσης ή μέτρα συνάφειας

## Συμμετρικές σχέσεις

- **Symmetric Measures**

- **Contingency coefficient:**  $C = \sqrt{\frac{x^2}{x^2 + n}}$  ,  $n$  μέγεθος δείγματος

- Με τιμές [0 και 1)
- Ο συντελεστής  $C$  μπορεί να χρησιμοποιηθεί για τη σύγκριση δυο διαφορετικών πινάκων συνάφειας. Συγκρίνοντας τους αντίστοιχους συντελεστές  $C$ , είναι δυνατόν να αποφανθούμε για τη διαφορά στο βαθμό συνάφειας.

- **Phi Coefficient :**  $\Phi = \sqrt{\frac{x^2}{n}}$  (αναφέρεται και ως συντελεστής του Pearson)

- Χρησιμοποιείται μόνο για  $2 \times 2$  πίνακες συνάφειας

- **Cramer's :**  $V = \sqrt{\frac{x^2}{(k-1)N}}$  ,  $k = \min\{r, c\}$

- Χρησιμοποιείται κυρίως για πίνακες συνάφειας διαφόρων διαστάσεων
- $V = 0.10$  small effect
- $V = 0.30$  medium effect
- $V = 0.50$  large effect

# Μέτρα προσδιορισμού της έντασης της σχέσης ή μέτρα συνάφειας

## Ασύμμετρες σχέσεις

- **Directional Measures**
- **Μέτρα αναλογικής μείωσης του λάθους (πρόβλεψης) (Proportional Reduction Error):**
- **Lambda:**
  - Ο συντελεστής  $\lambda$  παίρνει τιμές από 0 μέχρι και 1.
  - Η τιμή 1 σημαίνει πλήρη συσχέτιση μεταξύ των δύο μεταβλητών, οπότε θεωρητικά γνωρίζοντας τις τιμές της μιας, αυτόματα και επακριβώς μπορούμε να προβλέψουμε τις τιμές της άλλης. Έχουμε δηλαδή 100% μείωση του λάθους στην πρόβλεψη της μιας μεταβλητής από την άλλη.
  - Παράδειγμα: αν το μέτρο έχει τιμή 0,234, σημαίνει ότι έχουμε 23,4% μείωση του σφάλματος πρόβλεψης της μιας μεταβλητής από την άλλη σε σχέση με το σφάλμα που θα κάναμε αν προβλέπαμε στην τύχη.

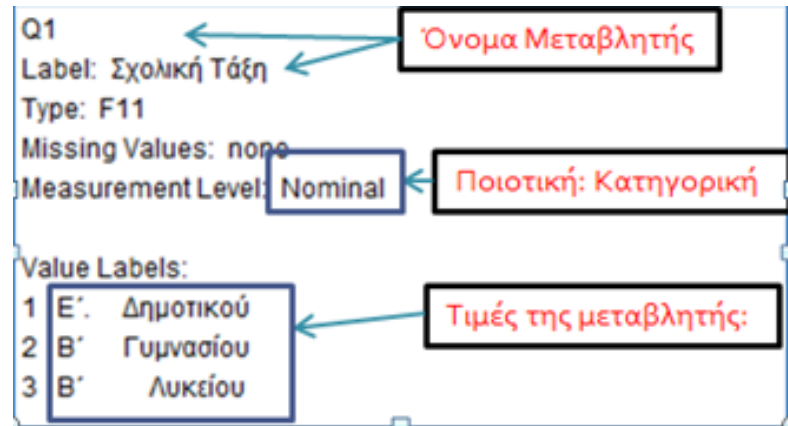
Μέτρα που προσδιορίζουν και την κατεύθυνση της σχέσης.

- **Μόνο για διατάξιμες ή τακτικές (ordinal) ποιοτικές μεταβλητές**
- Παίρνουν τιμές στο διάστημα  $[-1,1]$  με την τιμή  $-1$  να αντιστοιχεί σε τέλεια αρνητική σχέση, η τιμή  $0$  σε μη ύπαρξη σχέσης και η τιμή  $1$  σε τέλεια θετική σχέση. Αυτά κυρίως είναι:
  - Gamma, (Συμμετρική σχέση)
  - Somer's d. **Ασύμμετρη** επέκταση του συντελεστή Gamma.
  - Kendall's tau-b :κατάλληλος για συμμετρικούς πίνακες, (**Ασύμμετρη σχέση**)
  - Kendall's tau-c :κατάλληλος για μη συμμετρικούς (**Ασύμμετρη σχέση**)

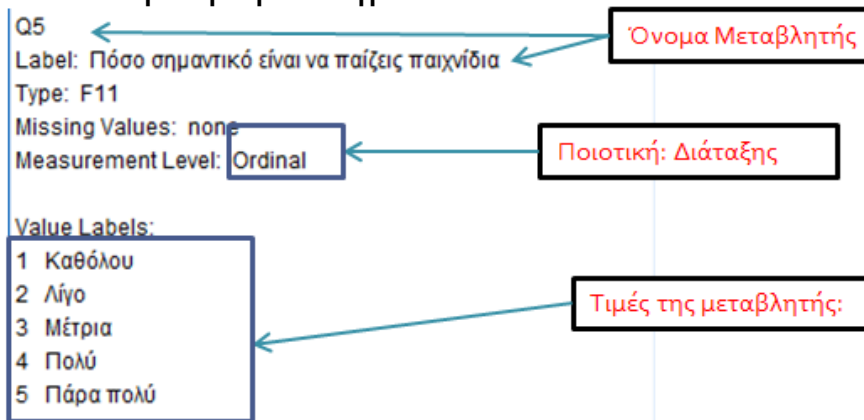


# Παράδειγμα 1<sup>ο</sup>

- Μας ενδιαφέρει να διερευνήσουμε την σχέση μεταξύ της σχολικής ηλικίας και της σημασίας που δίνουν οι μαθητές στο παιχνίδι.
- 1<sup>η</sup> Μεταβλητή: «Σχολική τάξη»:



- 2<sup>η</sup> Μεταβλητή: «Σημασία που δίνουν στο παιχνίδι»



# Πρέπει να παρουσιαστεί ο επόμενος πίνακας

Η παρουσίαση της «συνάφειας» των δύο μεταβλητών

Πόσο σημαντικό είναι να παίζεις παιχνίδια \* Σχολική Τάξη Crosstabulation

		Σχολική Τάξη			Total
		Ε' Δημοτικού	Β' Γυμνασίου	Β' Λυκείου	
Πόσο σημαντικό είναι να παίζεις παιχνίδια	Καθόλου	Count 1	Count 3	Count 3	Count 7
		% within Σχολική Τάξη 2,5%	% within Σχολική Τάξη 6,7%	% within Σχολική Τάξη 9,7%	% within Σχολική Τάξη 6,0%
	Λίγο	Count 0	Count 4	Count 7	Count 11
		% within Σχολική Τάξη 0,0%	% within Σχολική Τάξη 8,9%	% within Σχολική Τάξη 22,6%	% within Σχολική Τάξη 9,5%
	Μέτρια	Count 3	Count 12	Count 8	Count 23
	% within Σχολική Τάξη 7,5%	% within Σχολική Τάξη 26,7%	% within Σχολική Τάξη 25,8%	% within Σχολική Τάξη 19,8%	
	Πολύ	Count 16	Count 15	Count 7	Count 38
		% within Σχολική Τάξη 40,0%	% within Σχολική Τάξη 33,3%	% within Σχολική Τάξη 22,6%	% within Σχολική Τάξη 32,8%
	Πάρα πολύ	Count 20	Count 11	Count 6	Count 37
		% within Σχολική Τάξη 50,0%	% within Σχολική Τάξη 24,4%	% within Σχολική Τάξη 19,4%	% within Σχολική Τάξη 31,9%
Total	Count	40	45	31	116
	% within Σχολική Τάξη	100,0%	100,0%	100,0%	100,0%

Q1	Q5
Β' Γυμνασίου	Πολύ
Β' Γυμνασίου	Πάρα πολύ
Β' Γυμνασίου	Πολύ
Β' Γυμνασίου	Μέτρια
Ε' Δημοτικού	Καθόλου
Ε' Δημοτικού	Πάρα πολύ
Β' Γυμνασίου	Λίγο
Ε' Δημοτικού	Πολύ
Β' Λυκείου	Λίγο
Β' Λυκείου	Πολύ
Β' Λυκείου	Πάρα πολύ
Β' Λυκείου	Πάρα πολύ
Β' Λυκείου	Λίγο
Β' Λυκείου	Μέτρια
Β' Λυκείου	Πολύ
Β' Λυκείου	Μέτρια
Β' Λυκείου	Καθόλου
Β' Λυκείου	Πολύ
Β' Γυμνασίου	Πολύ
Ε' Δημοτικού	Πολύ
Β' Γυμνασίου	Πολύ
Β' Γυμνασίου	Πολύ
Β' Γυμνασίου	Πάρα πολύ
Β' Γυμνασίου	Πάρα πολύ
Β' Γυμνασίου	Μέτρια
Β' Γυμνασίου	Πολύ
Ε' Δημοτικού	Πάρα πολύ
Β' Λυκείου	Πάρα πολύ
Ε' Δημοτικού	Πολύ
Β' Λυκείου	Πολύ

- Έχουμε την βάση δεδομένων με τις τιμές των μεταβλητών

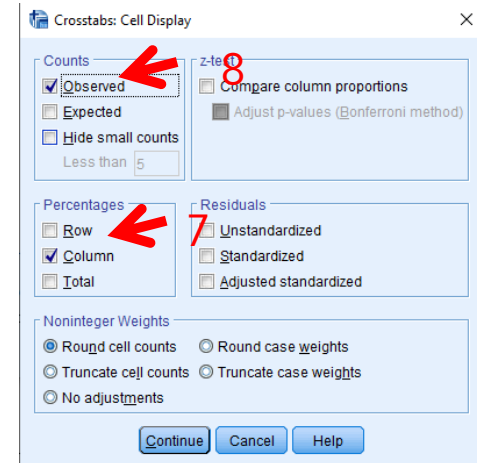
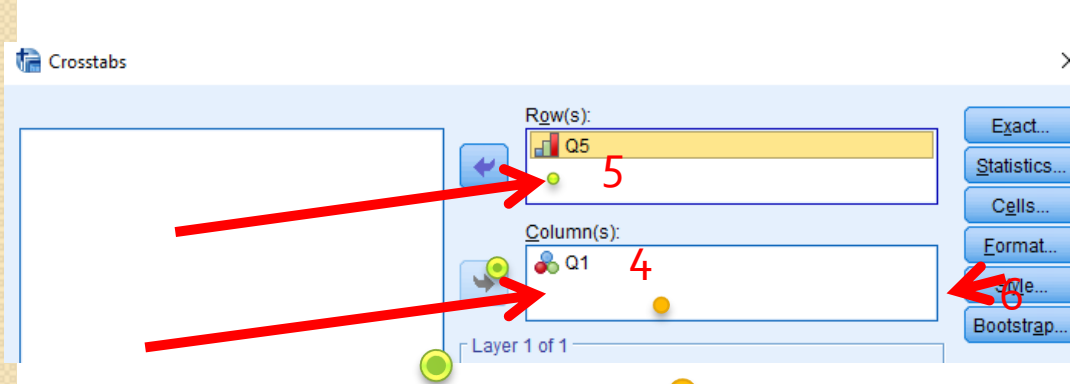
Αρχείο: «children game.sav»

## Σύνθετη παρουσίαση δύο ποιοτικών μεταβλητών, με το SPSS

- Για να μπορέσουμε να καταλάβουμε την όποια σχέση θα πρέπει να ορίσουμε, έστω και καταχρηστικά, τη μια μεταβλητή ως ανεξάρτητη και την άλλη ως εξαρτημένη.
- Συνήθως ανεξάρτητη είναι αυτή που προηγείται χρονικά
- Στο παράδειγμα αυτό:
  - Ανεξάρτητη : «Σχολική τάξη»
  - Εξαρτημένη: «Σημασία που δίνουν στο παιχνίδι»

# Σύνθετη παρουσίαση δύο ποιοτικών μεταβλητών, με το SPSS

- Εντολές: Analyze<sup>1</sup> → descriptive statistics<sup>2</sup> → crosstabs<sup>3</sup>



«Εξαρτημένη»

«Ανεξάρτητη»

# Πίνακας συνάφειας μεταξύ: του πόσο σημαντικό είναι να παίζουν τα παιδιά παιχνίδια ανά ηλικιακή ομάδα

- Προσπαθούμε να καταλάβουμε τη μορφή της σχέσης (μεγάλη προσοχή)
- Θεωρούμε ανεξάρτητη μεταβλητή αυτή που τοποθετήθηκε στη στήλη.
- Ελέγχουμε οριζόντια τα ποσοστά «% within μεταβλητή στήλης»:
- Συγκρίνοντας τα ποσοστά των τιμών κάθε τιμής της εξαρτημένης για κάθε τιμή της ανεξάρτητης

Πόσο σημαντικό είναι να παίζεις παιχνίδια \* Σχολική Τάξη Crosstabulation

		Σχολική Τάξη			Total
		Ε' Δημοτικού	Β' Γυμνασίου	Β' Λυκείου	
Πόσο σημαντικό είναι να παίζεις παιχνίδια	Καθόλου	Count 1	3	3	7
	% within Σχολική Τάξη	2,5%	6,7%	9,7%	6,0%
Λίγο	Count	0	4	7	11
	% within Σχολική Τάξη	0,0%	8,9%	22,6%	9,5%
Μέτρια	Count	3	12	8	23
	% within Σχολική Τάξη	7,5%	26,7%	25,8%	19,8%
Πολύ	Count	16	15	7	38
	% within Σχολική Τάξη	40,0%	33,3%	22,6%	32,8%
Πάρα πολύ	Count	20	11	6	37
	% within Σχολική Τάξη	50,0%	24,4%	19,4%	31,9%
Total	Count	40	45	31	116
	% within Σχολική Τάξη	100,0%	100,0%	100,0%	100,0%

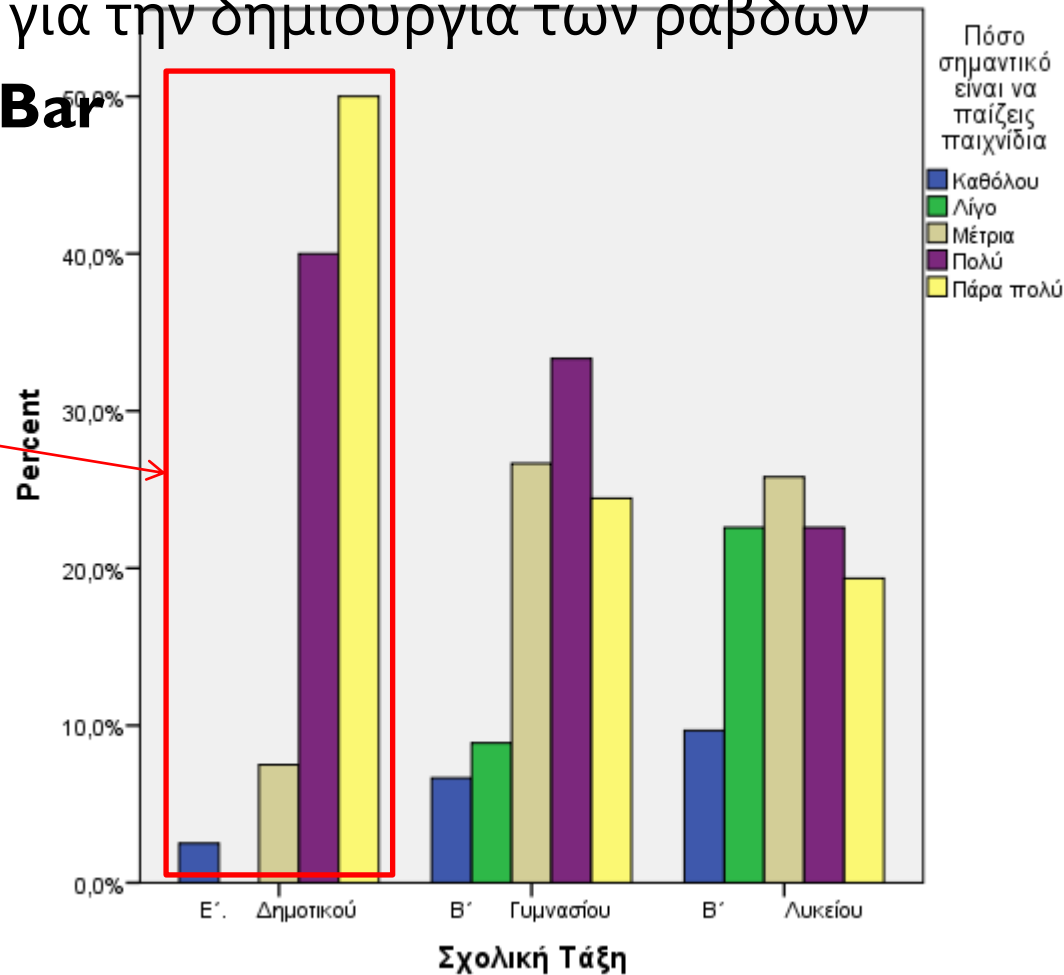
Η σύγκριση δηλαδή γίνεται μεταξύ των υποκατηγοριών της κάθετης μεταβλητής αφού διεξάγουμε κάθετη ερμηνεία των αποτελεσμάτων χρησιμοποιώντας τις στήλες των σχετικών συχνοτήτων «% within», στο παράδειγμα μας «σχολική τάξη».

# Κάθετη παρουσίαση του πίνακα

- Από τους μαθητές της Ε΄ δημοτικού
  - Το 50% δήλωσε ότι είναι πάρα πολύ σημαντικό να παίζεις παιχνίδια
  - Το 40% δήλωσε ότι είναι πολύ σημαντικό
  - Το 7,5% μέτρια
  - Το 2,5% λίγο
- Από τους μαθητές της Β΄ Γυμνασίου
  - Το 24,4% δήλωσε ότι είναι πάρα πολύ σημαντικό να παίζεις παιχνίδια
  - .....
- Από τους μαθητές της Β΄ Λυκείου
  - Το 19,4% δήλωσε ότι είναι πάρα πολύ σημαντικό να παίζεις παιχνίδια
  - .....

# Γράφημα παρουσίασης της σχέσης μεταξύ δύο ποιοτικών μεταβλητών

- Τα δεδομένα (ποσοστά) από τον πίνακα συνάφειας θα χρειαστούν για την δημιουργία των ράβδων
- **Clustered Bar**



Κάθε ομάδα ράβδων αντιστοιχεί σε κάθε τιμή της ανεξάρτητης μεταβλητής (γραμμή στον πίνακα συνάφειας) και έχει άθροισμα ποσοστών 100%

# Γράφημα παρουσίασης της σχέσης μεταξύ δύο ποιοτικών μεταβλητών, με το SPSS

- Εντολές: Graphs <sup>1</sup> → Chart Builder <sup>2</sup> →.

The image shows the SPSS Chart Builder interface with several dialog boxes open. Red arrows and numbers 1 through 8 indicate the steps for creating a clustered bar chart. A yellow cloud at the bottom left contains the text «Ανεξάρτητη» (Independent), and a green cloud at the bottom right contains «εξαρτημένη» (Dependent).

**Chart Builder:**

- Variables:** 4. Σχολική Τάξη (Categorical), Πόσο σημαντικό είναι να παίζεις... (Quantitative)
- Chart preview:** 5. Clustered bar chart. X-axis: Σχολική Τάξη (Ε' Δημοτικού, Β' Γυμνασίου, Β' Λυκείου). Y-axis: Πόσο σημαντικό είναι να παίζεις... (Percentage). 3. Cluster on X: set color.
- Gallery:** 3. Basic Elements tab. Bar chart selected.

**Element Properties:**

- Edit Properties of:** Bar1. Variable: Πόσο σημαντικό είναι να παίζεις... 6. Statistic: Percentage ().
- Set Parameters:** 7. Denominator for Computing Percentage: Total for Each X-Axis Category.
- Buttons:** 8. Continue.

**Buttons:** 9. Apply.

21/5/2024

«Ανεξάρτητη»

«εξαρτημένη»

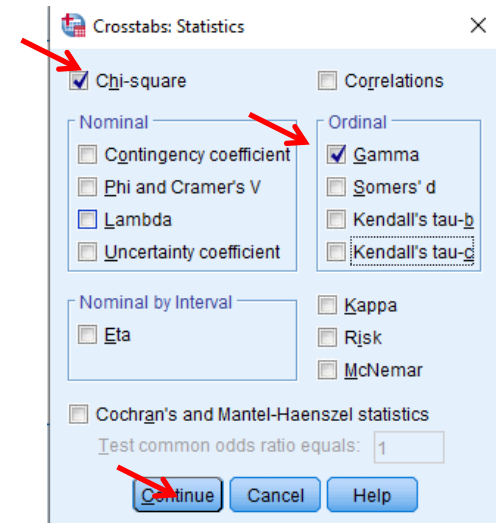
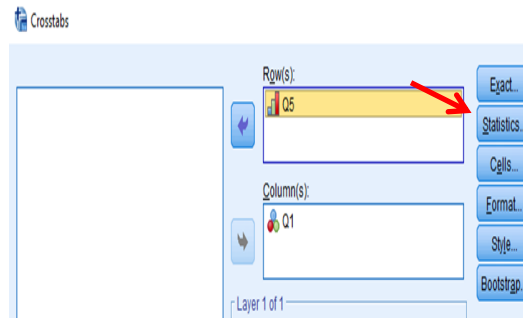
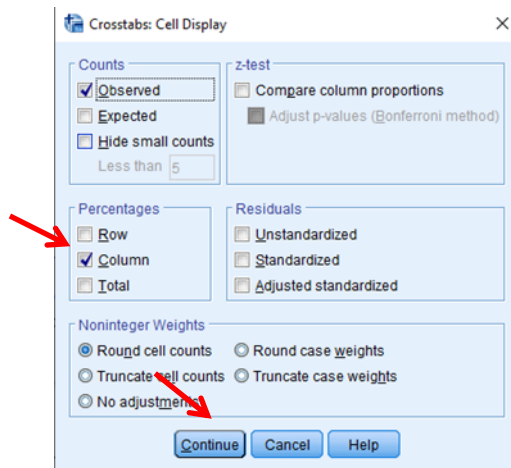
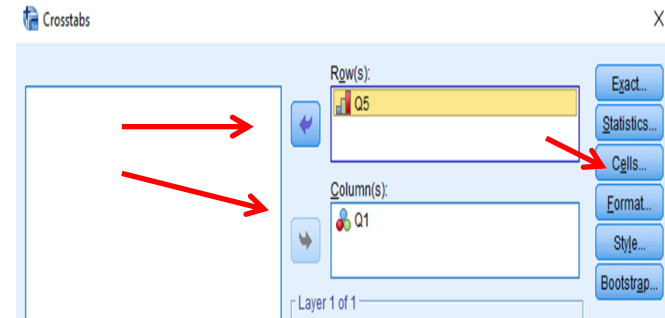
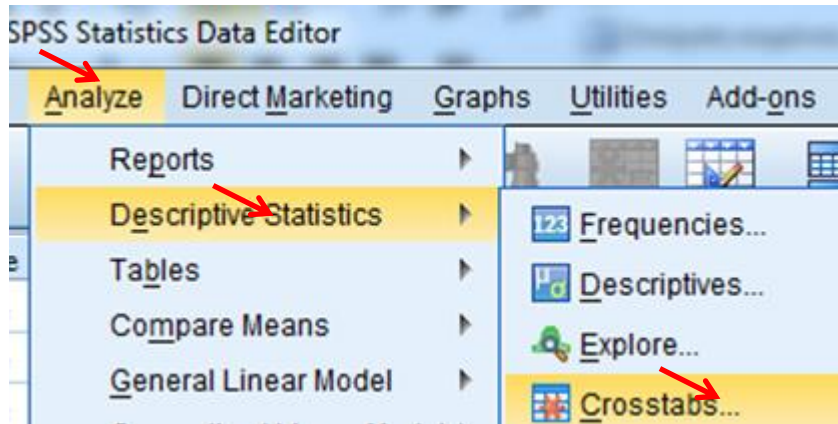


# Στατιστική συμπερασματολογία

- Από τα παραπάνω φαίνεται ότι οι η σημασία που δίνουν οι μαθητές στο παιχνίδι αποδυναμώνεται όσο οι μαθητές μεγαλώνουν.
- Μένει να ελέγξουμε κατά πόσο η σχέση αυτή που βρήκαμε δεν είναι τυχαία, αλλά **στατιστικά σημαντική**.
- Για έλεγχο της σημαντικότητας της σχέσης δύο ποιοτικών μεταβλητών, θα χρησιμοποιήσουμε τεστ:
  - $\chi^2$  ανεξαρτησίας.

# Έλεγχος της σημαντικότητας της σχέσης με το spss

21/5/2024



# Ο Πίνακας του ελέγχου $\chi^2$

## Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	23,691 <sup>a</sup>	8	,003
Likelihood Ratio	26,630	8	,001
Linear-by-Linear Association	17,424	1	,000
N of Valid Cases	116		

a. 6 cells (40,0%) have expected count less than 5. The minimum expected count is 1,87.

Προσοχή?

«Υπάρχει πρόβλημα»

Θα πρέπει τα κελιά με αναμενόμενες  
συχνοτήτες κάτω από 5 να είναι το πολύ το 20%  
των κελιών

Τι θα κάνουμε?

«αν οι προϋποθέσεις  
δεν ισχύουν!»

# Αν οι προϋποθέσεις δεν ισχύουν:

- 1<sup>ος</sup> τρόπος

- Διόρθωση Fisher:

- στις περιπτώσεις που έχουμε πίνακα συνάφειας 2x2 (δύο ποιοτικές μεταβλητές με δύο κατηγορίες η κάθε μία)

- Test προσομοίωσης: Exact test

- 2<sup>ος</sup> τρόπος

- Συγχώνευση κατηγοριών και δημιουργία ευρύτερων κατηγοριών.

- *Με ότι αυτό μπορεί να σημαίνει για την αξιοπιστία των αποτελεσμάτων.*

# 1<sup>ος</sup> τρόπος

## Test προσομοίωσης Exact test

### Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	23,691 <sup>a</sup>	8	,003	,002		
Likelihood Ratio	26,630	8	,001	,001		
Fisher's Exact Test	23,527			,001		
Linear-by-Linear Association	17,424 <sup>b</sup>	1	,000	,000	,000	,000
N of Valid Cases	116					

a. 6 cells (40,0%) have expected count less than 5. The minimum expected count is 1,87.

b. The standardized statistic is -4,174.

# 2<sup>ος</sup> τρόπος

Ας συγχωνεύσουμε κατηγορίες της εξαρτημένης μεταβλητής

Πόσο σημαντικό είναι να παίζεις παιχνίδια			Πόσο σημαντικό είναι να παίζεις παιχνίδια	
Καθόλου	Λίγο	Μέτρια	Πολύ	Πάρα πολύ

1<sup>η</sup>: καθόλου – μέτρια

2<sup>η</sup>: πολύ και πάρα πολύ

Εντολές: Transform - into different variable

1,2, 3 → 1

4,5 → 2

Προέκυψε η μεταβλητή q5\_trans

και

- Τρέχουμε ξανά τον έλεγχο

# Πίνακες του ελέγχου $\chi^2$

Πίνακας του  $\chi^2$ :

Αν  $sign < 0,05$  στατιστικά σημαντική σχέση

Γράφουμε σε κάθε περίπτωση:

$(\chi^2(2, N=4)=3, p=1)$  δηλαδή  $\chi^2(2, N=116)=19,17, p=0,001$

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	19,177 <sup>a</sup>	2	,000
Likelihood Ratio	21,235	2	,000
Linear-by-Linear Association	18,216	1	,000
N of Valid Cases	116		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 10,96.



Symmetric Measures

		Value	Asymptotic Standard Error <sup>a</sup>	Approximate T <sup>b</sup>	Approximate Significance	Exact Significance
Ordinal by Ordinal	Gamma	-,482	,094	-4,785	,000	,000
N of Valid Cases		116				

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Πίνακας προσδιορισμού της Έντασης της σχέσης: Συντελεστής Gamma

21/5/2024



# Παρουσίαση του αποτελέσματος

- Οι μαθητές της Ε' δημοτικού φαίνεται να δίνουν πολύ μεγάλη σημασία για το παιχνίδι, αφού η συντριπτική πλειονότητα αυτών (90%) δηλώνει ότι είναι πολύ (40%) και πάρα πολύ (50%) σημαντικό να παίζεις παιχνίδια. Η εικόνα αυτή με μικρότερη όμως ένταση (57,7%) παραμένει και για τους μαθητές της Β γυμνασίου. Τέλος για τους μαθητές της Β λυκείου παρατηρούμε ότι μόνο το 42% αυτών δηλώνει ότι είναι σημαντικό από πολύ έως πάρα πολύ να παίζεις παιχνίδια.
- Από τα παραπάνω φαίνεται ότι οι η σημασία που δίνουν οι μαθητές στο παιχνίδι αποδυναμώνεται όσο οι μαθητές μεγαλώνουν. Επομένως, τα δεδομένα του δείγματος μας υποδεικνύουν μια αρνητική σχέση μεταξύ της σχολικής ηλικίας και της σημασίας που δίνουν για το παιχνίδι.
- Από τον έλεγχο  $\chi^2$  που πραγματοποιήσαμε παρατηρούμε ότι η σχέση αυτή είναι και στατιστικά σημαντική ( $\chi^2(2, N=116)=19,17, p=0,001$ ). Επιπλέον η σχέση αυτή είναι ισχυρή αρνητική, αφού ο αντίστοιχος συντελεστής gamma είναι  $\gamma = -0,634$ .

# Παράδειγμα 2<sup>ο</sup>

- Διερευνούμε την σχέση αναγνωσιμότητας ενός εντύπου ανάλογα με το κοινωνικο-οικονομικό προφίλ του αναγνωστικού κοινού
  - Μεταβλητές:
    - Αναγνωσιμότητα (συχνότητα ανάγνωσης): **1** - τακτικός και ο - περιστασιακός αναγνώστης
    - Κοινωνικο-οικονομικό προφίλ: 1-χαμηλό, 2 - μεσαίο και 3 - υψηλό.

# Ερευνητικά ερωτήματα και υποθέσεις

- Υποθέσεις:
  - $H_0$ : Η αναγνωσιμότητα του εντύπου Α είναι ανεξάρτητη του κοινωνικό-οικονομικού προφίλ των πολιτών
  - $H_1$ : Η αναγνωσιμότητα εξαρτάται από το κοινωνικό-οικονομικό προφίλ των πολιτών
- Εναλλακτικά με τη μορφή ερωτήματος:
  - Ερευνητικό ερώτημα
    - Σχετίζεται η αναγνωσιμότητα του εντύπου Α και πως με το κοινωνικό –οικονομικό προφίλ των πολιτών

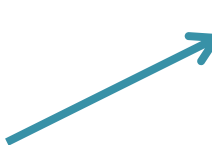
# Έστω ότι έχουμε στην διάθεσή μας τον πίνακα συχνοτήτων

		αναγνωσιμότητα	
		τακτικός	περιστασιακός
κοινωνικο-οικονομικό επίπεδο	χαμηλό	20	60
	μεσαίο	45	25
	υψηλό	40	30

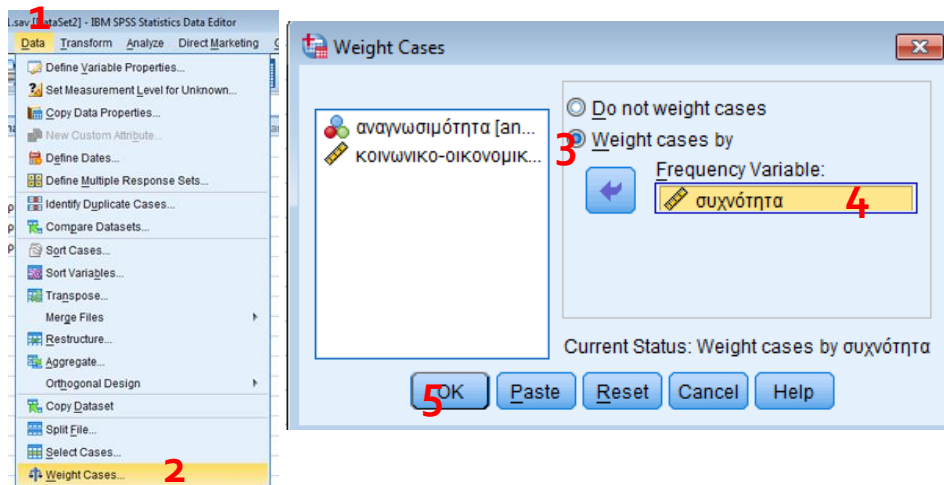
Άραγε με ποιο τρόπο θα περάσω τις μεταβλητές στο SPSS ;

# Ο προφανής τρόπος είναι χρονοβόρος!

- Εναλλακτικά
- Φτιάχνω τρεις μεταβλητές
  - (η τρίτη είναι η συχνότητα)



anagnwsim	koinwnioikon	συχνότητα
τακτικός	χαμηλό	20,00
τακτικός	μεσαίο	45,00
τακτικός	υψηλό	40,00
περιστασιακός	χαμηλό	60,00
περιστασιακός	μεσαίο	25,00
περιστασιακός	υψηλό	30,00



- Η διαδικασία καλείται στάθμιση περιπτώσεων (weight cases)

# Εργασία για το σπίτι

- Να ελέγξετε την προηγούμενη σχέση και να γράψετε τα συμπεράσματα σας
- Εναλλακτικά
  - Αρχείο «[anagnvsimothta.sav](#)»

# Παράδειγμα 3<sup>ο</sup>

- Από το αρχείο δεδομένων «x2 2x2.sav» μας ενδιαφέρει η αναζήτηση της πιθανής σχέσης μεταξύ των μεταβλητών τρόπος επιμόρφωσης και αποτέλεσμα εξετάσεων. Αν υπάρχει σχέση, ποια είναι η φύση και η ένταση της σχέσης των δύο αυτών μεταβλητών;
- Γράψτε τα συμπεράσματά σας...