

## 10ο ΕΡΓΑΣΤΗΡΙΟ

### Συσχέτιση Ποιοτικών (Κατηγορικών) δεδομένων - ordinal/nominal

Για δύο ποιοτικές μεταβλητές χρησιμοποιείται ο έλεγχος  $\chi^2$  (chi-square test)

Analyze → Descriptive Statistics → Crosstabs

Στην επιλογή “rows” τοποθετούμε την μία μεταβλητή και στην επιλογή “columns” την άλλη. Για να έχουμε τα αποτελέσματα του ελέγχου  $\chi^2$ , πρέπει να επιλέξουμε Statistics και Chi-square. Επίσης είναι χρήσιμο να έχουμε και τα ποσοστά ανα γραμμές (rows) ή ανα στήλες (columns), οπότε ενεργοποιούμε επίσης την επιλογή Cells, οπότε διαλέγουμε Percentages (ποσοστά), ανά γραμμή ή ανα στήλη.

#### Παράδειγμα 1

Συσχετίζοντας τις μεταβλητές ΠΑΧΥΣΑΡΚΙΑ και ΚΑΡΔΙΑΓΓΕΙΑΚΟΣ ΚΙΝΔΥΝΟΣ έχουμε στο output

		Καρδιαγγειακός κίνδυνος				
			ΟΧΙ	ΥΨΗΛΟΣ	ΠΟΛΥ ΥΨΗΛΟΣ	Total
Παχυσαρκία	Φυσιολογικός	Count	133	110	9	252
		% within Παχυσαρκία	52,80%	43,70%	3,60%	100,00%
	Υπέρβαρος	Count	81	49	2	132
		% within Παχυσαρκία	61,40%	37,10%	1,50%	100,00%
	Παχύσαρκος	Count	56	40	1	97
		% within Παχυσαρκία	57,70%	41,20%	1,00%	100,00%
Total		Count	270	199	12	481
		% within Παχυσαρκία	56,10%	41,40%	2,50%	100,00%

Και

#### Chi-Square Tests

	Value	df	Asymp. Sig. (2- sided)
Pearson Chi-Square	4,598 <sup>a</sup>	4	,331
Likelihood Ratio	4,788	4	,310
Linear-by-Linear Association	2,234	1	,135
N of Valid Cases	481		

a. 2 cells (22,2%) have expected count less than 5. The minimum expected count is 2,42.

Από τον τελευταίο πίνακα, στην πρώτη σειρά, βλέπουμε το  $\chi^2$  και το p.

Το  $p$  είναι στην περίπτωση αυτή μεγάλο ( $p=0,331 > 0,05$ ), οπότε τα δύο χαρακτηριστικά δεν έχουν καμία σχέση.

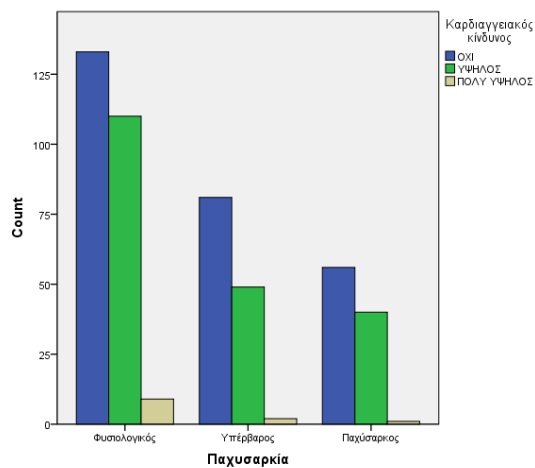
Ο πρώτος πίνακας με τα ποσοστά, μπορεί να διαμορφωθεί πιο απλά, ως εξής

	Καρδιαγγειακός κίνδυνος			
Παχυσαρκία	ΟΧΙ	ΥΨΗΛΟΣ	ΠΟΛΥ ΥΨΗΛΟΣ	
Φυσιολογικός	52,80%	43,70%	3,60%	100%
Υπέρβαρος	61,40%	37,10%	1,50%	100%
Παχύσαρκος	57,70%	41,20%	1,00%	100%

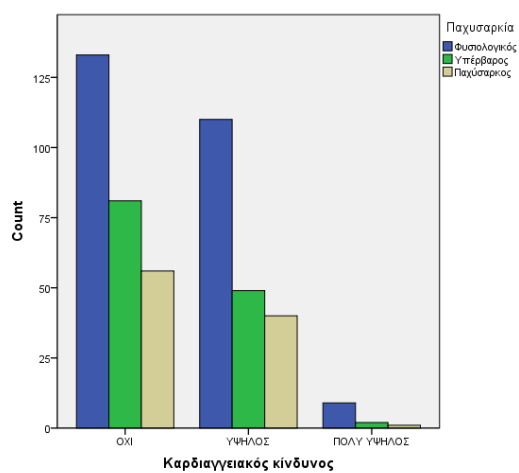
Για να παρουσιάσουμε γραφικά την συσχέτιση 2 ποιοτικών μεταβλητών χρησιμοποιούμε ένα σύνθετο Ραβδόγραμμα ως εξής

[Graphs](#) → [Legacy Dialogs](#) → [Bar](#) → [Clustered](#)

Η μία μεταβλητή τοποθετείται στο “category axis” και η άλλη στο “define clusters by”, οπότε έχουμε για τις παραπάνω μεταβλητές



Αν τοποθετήσουμε τις μεταβλητές ανάποδα, έχουμε



## Παράδειγμα 2

Για τη σχέση ανάμεσα στο ΦΥΛΟ και στο ΚΑΠΝΙΣΜΑ, έχουμε

**Φύλο \* Κάπνισμα Crosstabulation**

			Κάπνισμα		Total
			ΟΧΙ	ΝΑΙ	
Φύλο	Ανδρας	Count	129	59	188
		% within Φύλο	68,6%	31,4%	100,0%
	Γυναίκα	Count	172	121	293
		% within Φύλο	58,7%	41,3%	100,0%
Total		Count	301	180	481
		% within Φύλο	62,6%	37,4%	100,0%

**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4,806 <sup>a</sup>	1	,028		
Continuity Correction <sup>b</sup>	4,392	1	,036		
Likelihood Ratio	4,857	1	,028		
Fisher's Exact Test				,034	,018
Linear-by-Linear Association	4,796	1	,029		
N of Valid Cases	481				

Εδώ έχουμε  $p=0,028 < 0,05$  είναι στατιστικά σημαντικό, που σημαίνει ότι οι μεταβλητές έχουν κάποια σχέση. Παρατηρώντας τον πίνακα με τα ποσοστά, βλέπουμε ότι οι άντρες καπνίζουν σε ποσοστό 31,4% ενώ για τις γυναίκες το ποσοστό αυτό είναι αυξημένο και είναι 41,3%