



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών

Media Monitoring

Ενότητα 7: Εισαγωγή & Ανάλυση δεδομένων με το
SPSS

Σταμάτης Πουλακιδάκος

Σχολή ΟΠΕ

Τμήμα ΕΜΜΕ

Output

Είναι ο όρος που χρησιμοποιείται για να περιγράψει τα αποτελέσματα από αναλύσεις που έχουν γίνει και είναι ο αντικειμενικός στόχος του προγράμματος. Στα αριστερά του output view είναι η χρήση του περιγράμματος (outline view) που περιγράφει περιληπτικά τα αποτελέσματα που έχουμε βγάλει στο output

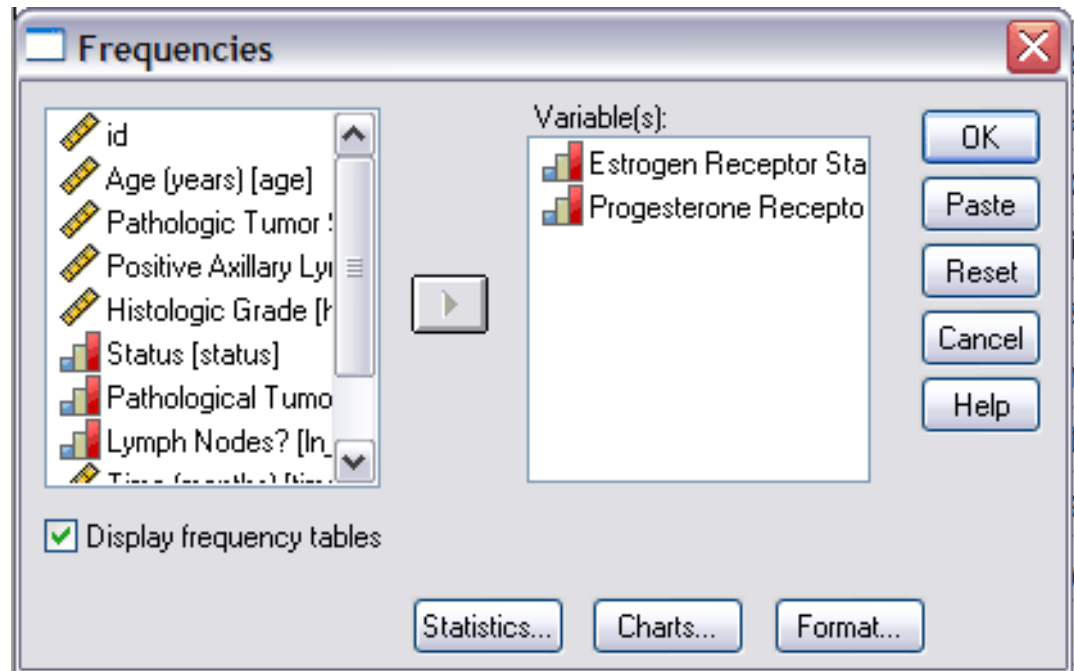
The screenshot shows the SPSS Output Viewer window. The left pane displays the 'Outline View' with a tree structure of the output, including 'Frequencies', 'Statistics', and 'Frequency Table'. The right pane shows the 'Frequencies' output for the dataset 'C:\Program Files\SPSS\Breast cancer survival.sav'. It includes a 'Statistics' table, a 'Frequency Table' for 'Estrogen Receptor Status', and the beginning of a 'Progesterone Receptor Status' table.

	Estrogen Receptor Status	Progesterone Receptor Status
N	869	851
Valid		
Missing	338	356
Median	1,00	1,00

		Estrogen Receptor Status			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Negative	338	28,0	38,9	38,9
	Positive	531	44,0	61,1	61,1
	Total	869	72,0	100,0	100,0
Missing	Unknown	338	28,0		
Total		1207	100,0		

Παράθυρο διαλόγου

Στις περισσότερες αναλύσεις που θα θελήσουμε να κάνουμε, επικοινωνούμε με τον υπολογιστή μέσω ενός παραθύρου διαλόγου



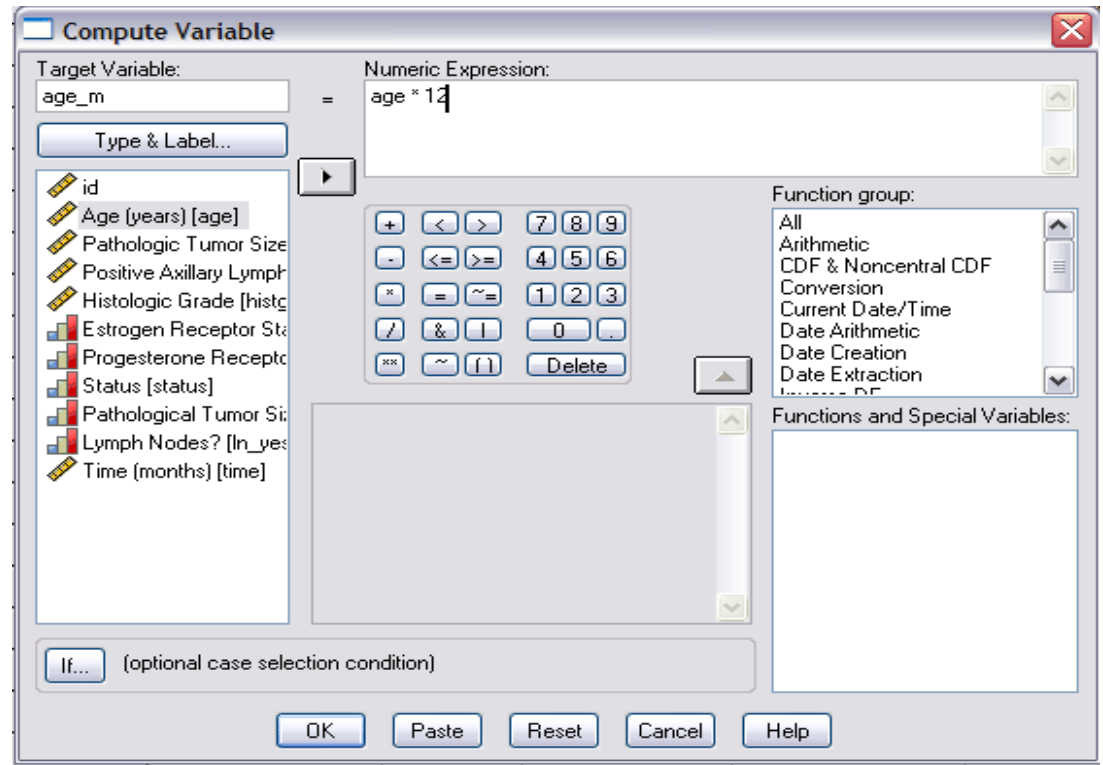
- Χωρίζεται σε 4 μέρη:
- 1. Αριστερά είναι το πλαίσιο με τις διαθέσιμες μεταβλητές (αυτές που έχουμε ορίσει στο variable view)
- 2. Το δεύτερο πλαίσιο ακριβώς δίπλα αφορά τις μεταβλητές που θα επιλέξουμε να συμμετάσχουν
- 3. Στο κάτω μέρος υπάρχουν επιλογές που σχετίζονται με την ανάλυση
- 4. Δεξιά υπάρχουν τα πλήκτρα OK, paste, cancel, help που επιλέγονται προκειμένου να ολοκληρωθούν οι εργασίες στο συγκεκριμένο παράθυρο



Το μενού transform

Η εντολή compute χρησιμοποιείται για να υπολογίσετε μια νέα μεταβλητή συναρτήσει κάποιας ήδη υπάρχουσας. Αν π.χ. Έχετε την ηλικία του ανθρώπου σε έτη και θέλετε να την έχετε σε μήνες θα χρησιμοποιήσετε αυτή την εντολή

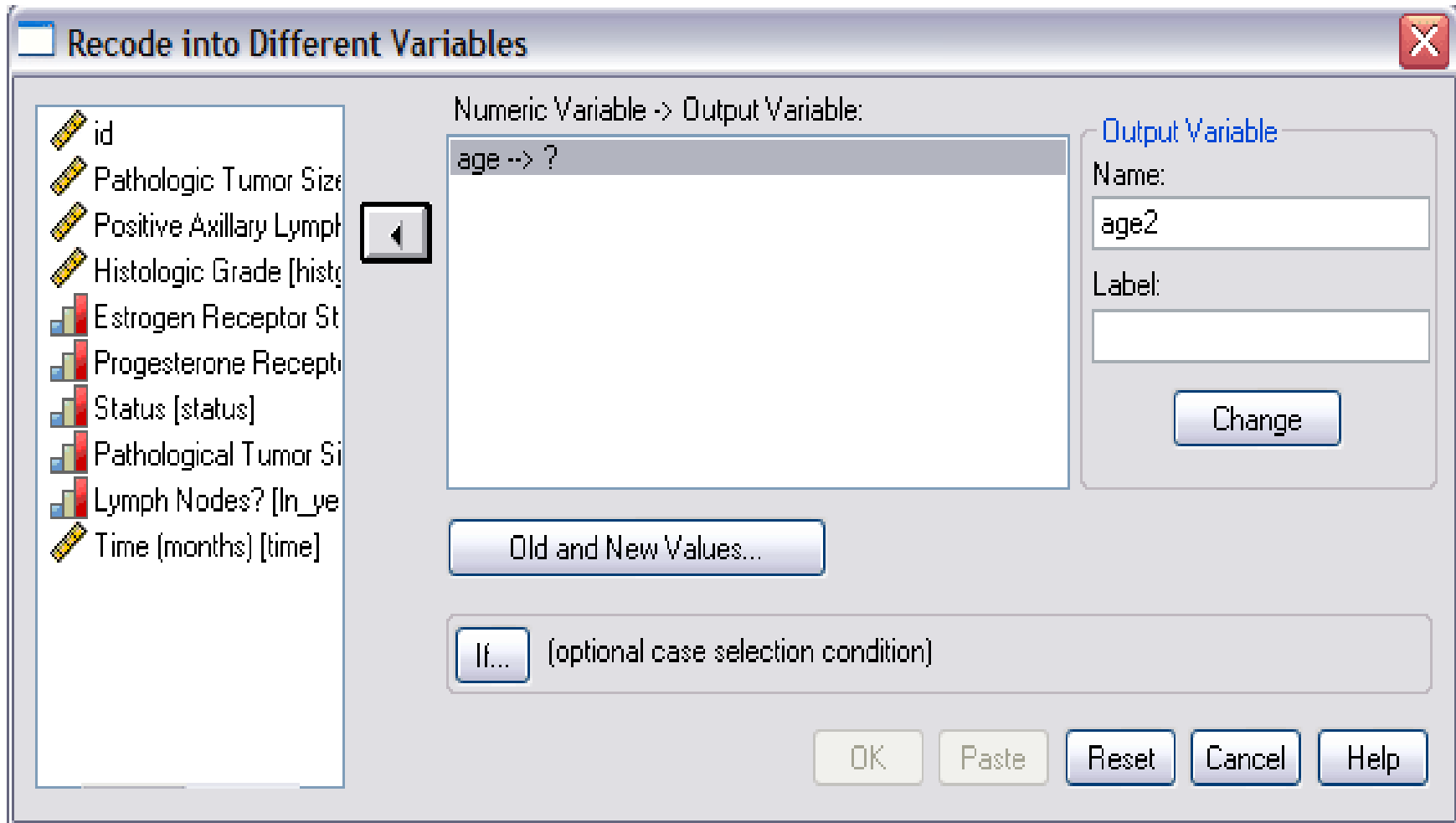
Επιλέγουμε
transform→compute→



- Με την εντολή recode αλλάζουμε την κωδικοποίηση των τιμών μιας μεταβλητής, έτσι ώστε να πάρει μορφή κατάλληλη για την ανάλυσή μας
- Χρησιμοποιείται σε μεταβλητές των οποίων οι τιμές αντιπροσωπεύουν κατηγορίες. Π.χ. Αν επιθυμούμε να ομαδοποιήσουμε τη μεταβλ. Age, μπορούμε να δώσουμε εντολή έτσι ώστε οι ηλικίες μέχρι 18 να αντιπροσωπεύονται πλέον με τον αριθμό 1, οι 19-24 με το 2 κ.ο.κ.
- Έχουμε τη δυνατότητα είτε να μετατρέψουμε την ήδη υπάρχουσα μεταβλητή, είτε να δημιουργήσουμε καινούρια, διατηρώντας την παλιά ανέπαφη



Επιλέγοντας transform-recode into different variable εμφανίζεται το εξής παράθυρο



Από τη λίστα με τις μεταβλητές αριστερά επιλέγουμε εκείνη που θέλουμε να επανακωδικοποιήσουμε και τη μεταφέρουμε στο διπλανό πλαίσιο (numeric variable-output variable), ορίζουμε όνομα και ετικέτα της νέας μεταβλητής & επιλέγουμε old and new values:

Recode into Different Variables

Numeric Variable -> Output Variable: _____ Output Variable: _____

Recode into Different Variables: Old and New Values

Old Value

Value: _____

System-missing

System- or user-missing

Range:

through

Range, LOWEST through value:

Range, value through HIGHEST:

All other values

New Value

Value: _____

System-missing

Copy old value(s)

Old --> New:

Add

Change

Remove

Lowest thru 18 --> 1

19 thru 24 --> 2

25 thru 40 --> 3

41 thru 60 --> 4

61 thru Highest --> 5

Output variables are strings Width: 8

Convert numeric strings to numbers ('5'->5)

Continue Cancel Help



Αντίστοιχη είναι η διαδικασία και για την επανακωδικοποίηση της ίδιας μεταβλητής.

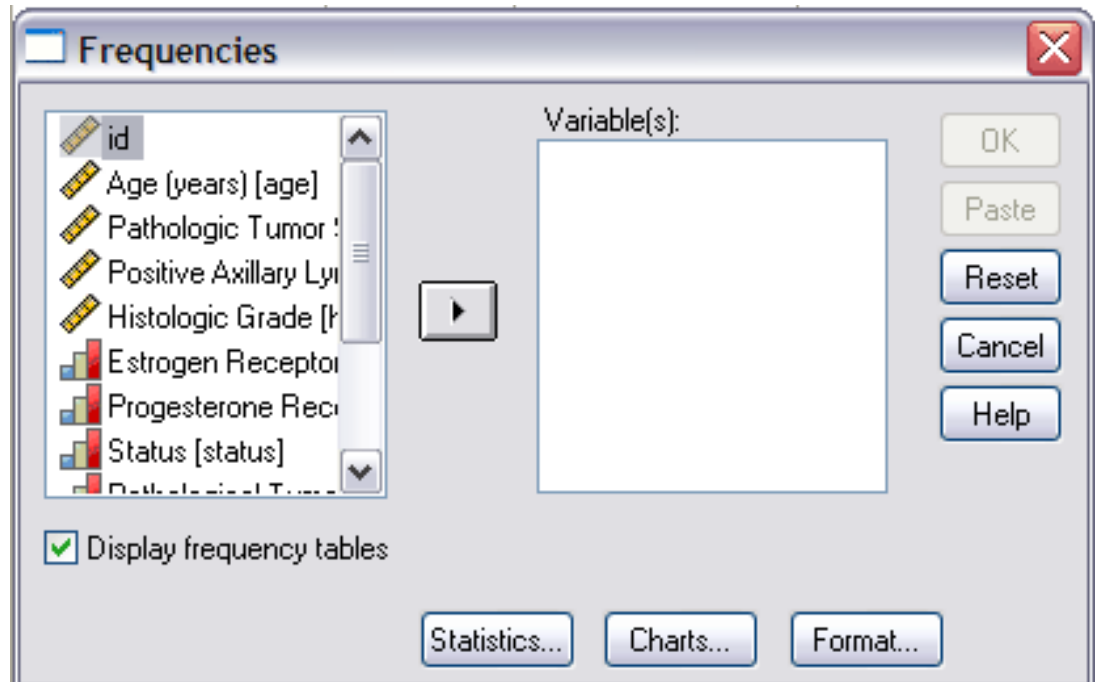
Το recode μας χρησιμεύει:

1. Όταν έχουμε διαφορετικά αρχεία με τις ίδιες μεταβλητές αλλά με διαφορετική κωδικοποίηση σε κάποια ή κάποιες
2. Έχουμε έναν αριθμό κατηγοριών σε μεταβλητή που θέλουμε να μειώσουμε
3. Όταν κάνουμε χρήση της κλίμακας Likert και θέλουμε να προχωρήσουμε σε αναλύσεις παραγόντων ή ανάλυση αξιοπιστίας, όπου οι επιλογές πρέπει να έχουν την ίδια φορά

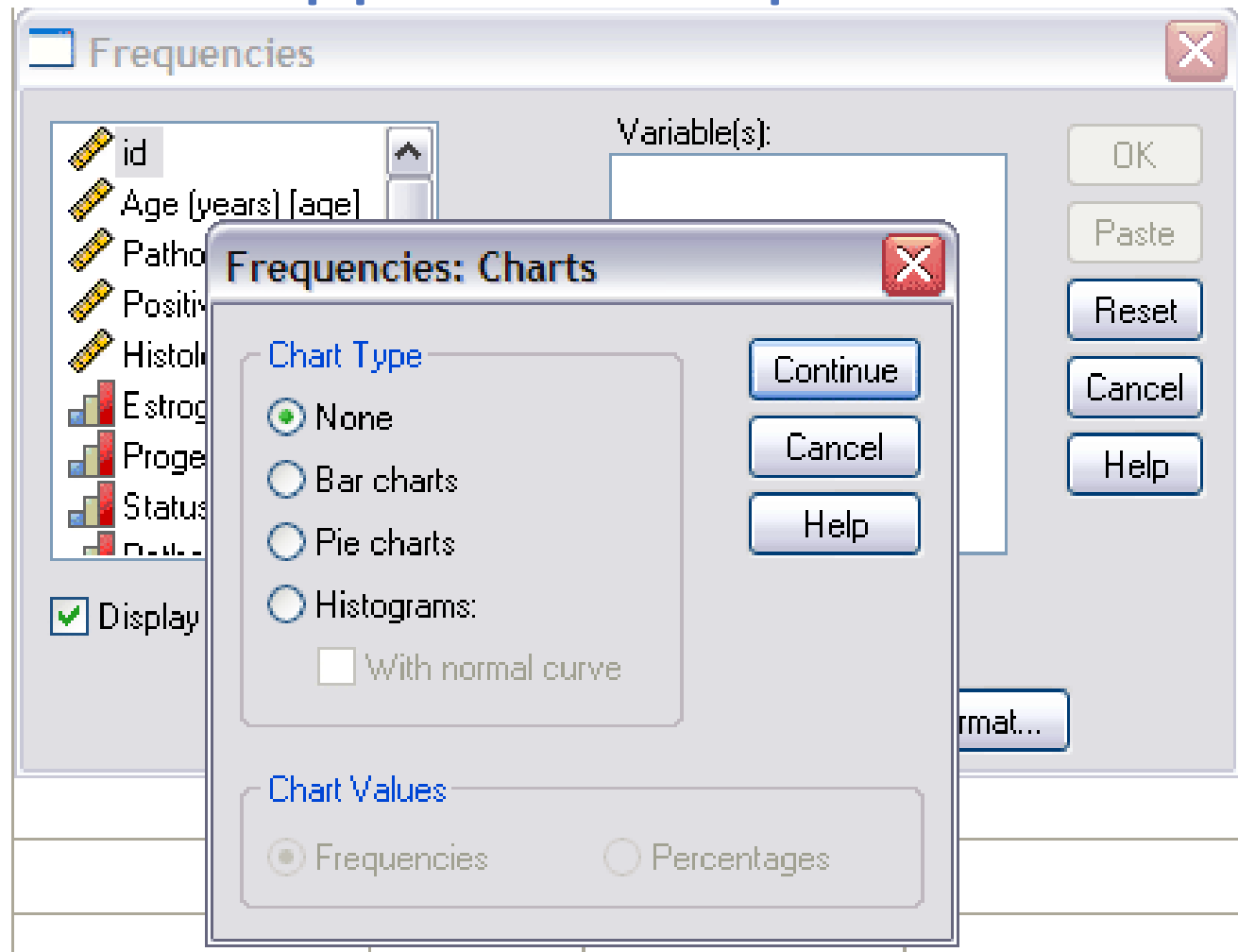


Το μενού analyze

Η εντολή frequencies μας δίνει τη συχνότητα εμφάνισης κάθε τιμής των μεταβλητών που έχουμε επιλέξει να αναλύσουμε, καθώς και κάποια βασικά μέτρα περιγραφικής στατιστικής (analyze-descriptive statistics-frequencies). Κοιτάμε να είναι οπωσδήποτε ενεργοποιημένη η επιλογή display frequency tables



Μπορούμε να επιλέξουμε και γραφική απεικόνιση μέσω του μενού charts



Πατώντας OK παράγεται πίνακας σαν τον ακόλουθο:

QUIZ1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	2	1,9	1,9	1,9
	2	1	1,0	1,0	2,9
	3	6	5,7	5,7	8,6
	4	6	5,7	5,7	14,3
	5	6	5,7	5,7	20,0
	6	12	11,4	11,4	31,4
	7	14	13,3	13,3	44,8
	8	16	15,2	15,2	60,0
	9	8	7,6	7,6	67,6
	10	34	32,4	32,4	100,0
	Total	105	100,0	100,0	



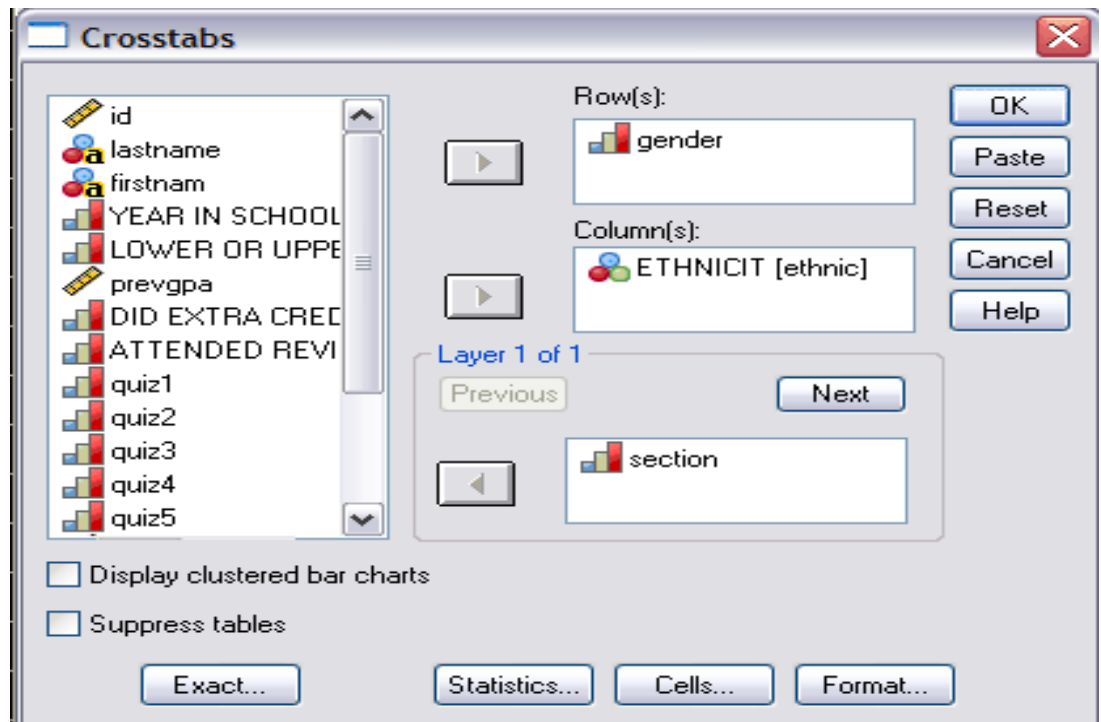
Η εντολή crosstabs

- Με την εντολή crosstabs παρουσιάζεται σε μορφή πίνακα η σχέση δύο **ποιοτικών** μεταβλητών. Τέτοιες είναι οι μεταβλητές που έχουν διακριτές κατηγορίες, χωρίς καμιά σειρά κατάταξης/ιεράρχησης, όπως το φύλο (γυναίκα, άνδρας), ο τόπος κατοικίας (πόλη, προάστια, εξοχή), απαντήσεις (ναι, όχι) κ.α.
- Μπορεί να χρησιμοποιηθεί και για συνεχή δεδομένα μόνο όταν αυτά είναι ομαδοποιημένα σε επιμέρους κατηγορίες, όπως η ηλικία (0-19, 20-39, 40-59, 60-79, 80-99)
- Μπορούμε να παρουσιάσουμε τη σχέση 2 μεταβλητών με βάση κάποια τρίτη



Crosstabs

Π.χ. υπάρχουν 3 τμήματα (sections) που παρακολουθούν ένα μάθημα, στα οποία υπάρχουν μαθήτριες και μαθητές (gender) διαφορετικών εθνοτήτων (ethnicit). Θέλουμε να δούμε πως σχετίζονται οι δύο μεταβλητές gender & ethnicit ανά τμήμα (section). Επιλέγουμε analyze-descriptive statistics-crosstabs



Επιλέγουμε τις μεταβλητές που θέλουμε να συσχετίσουμε και τις τοποθετούμε στα πλαίσια row & column (δεν έχει ιδιαίτερη σημασία σε ποιό από τα δύο). Στο τρίτο πλαίσιο (layer 1 of 1) τοποθετούμε τη μτβλ. Κριτήριο (π.χ. section)

GENDER ^ ETHNICIT ^ SECTION Crosstabulation

Count

SECTION			ETHNICIT					Total
			AMERICAN INDIAN	ASIAN	AFRO-AMERICAN	CAUCASIAN	HISPANIC	
1	GENDER	FEMALE		6	2	9	3	20
		MALE		1	5	7		13
		Total		7	7	16	3	33
2	GENDER	FEMALE	3	5	6	11	1	26
		MALE	1	4	1	7		13
		Total	4	9	7	18	1	39
3	GENDER	FEMALE	1	2	6	6	3	18
		MALE		2	4	5	4	15
		Total	1	4	10	11	7	33



Crosstabs & ανάλυση χ^2

Η εντολή `frequencies` μπορεί να μας δώσει τις συχνότητες συγκεκριμένων μεμονωμένων μεταβλητών, αλλά δεν μπορεί να μας φανερώσει συχνότητες περιπτώσεων για τις οποίες χρειαζόμαστε περισσότερες από μία μεταβλητές, προκειμένου να τις προσδιορίσουμε



Έλεγχος ανεξαρτησίας χ^2

- Μαζί με τις συχνότητες μέσα σε κάθε κελί του πίνακα διπλής εισόδου (crosstab), το SPSS μπορεί να υπολογίσει την αναμενόμενη τιμή για κάθε κελί
- Η αναμενόμενη τιμή βασίζεται στην υπόθεση ότι οι δύο μεταβλητές είναι ανεξάρτητες μεταξύ τους. Έστω ότι σε μια ομάδα 100 ανθρώπων οι 30 είναι άνδρες και οι 70 γυναίκες. Αν υπάρχουν 10 Ασιάτες ανάμεσα στους 100, οι 3 θα είναι άνδρες και οι 7 γυναίκες (η ίδια αναλογία που παρατηρείται στο σύνολο της ομάδας)



Έλεγχος ανεξαρτησίας χ^2

- Ο σκοπός του ελέγχου ανεξαρτησίας chi-square είναι να καθορίσει αν οι πραγματικές τιμές για τα κελιά αποκλίνουν σημαντικά από τις αντίστοιχες αναμενόμενες τιμές
- Αν υπάρχει μεγάλη ασυμφωνία μεταξύ των πραγματικών τιμών και των αναμενόμενων η τιμή της χ^2 θα είναι υψηλή, υποδηλώνοντας σημαντική διαφορά μεταξύ των πραγματικών και των αναμενόμενων τιμών
- Ο τύπος του ελέγχου chi-square: $\chi^2 = \sum [(f_0 - f_e)^2 / f_e]$
- Μαζί με το μέτρο αυτό υπολογίζεται και μια τιμή πιθανότητας. Με $p < 0.05$, είναι κοινώς αποδεκτό ότι οι τιμές που παρατηρήθηκαν διαφέρουν σημαντικά από τις αναμενόμενες τιμές και ότι οι δύο μεταβλητές ΔΕΝ είναι ανεξάρτητες μεταξύ τους.



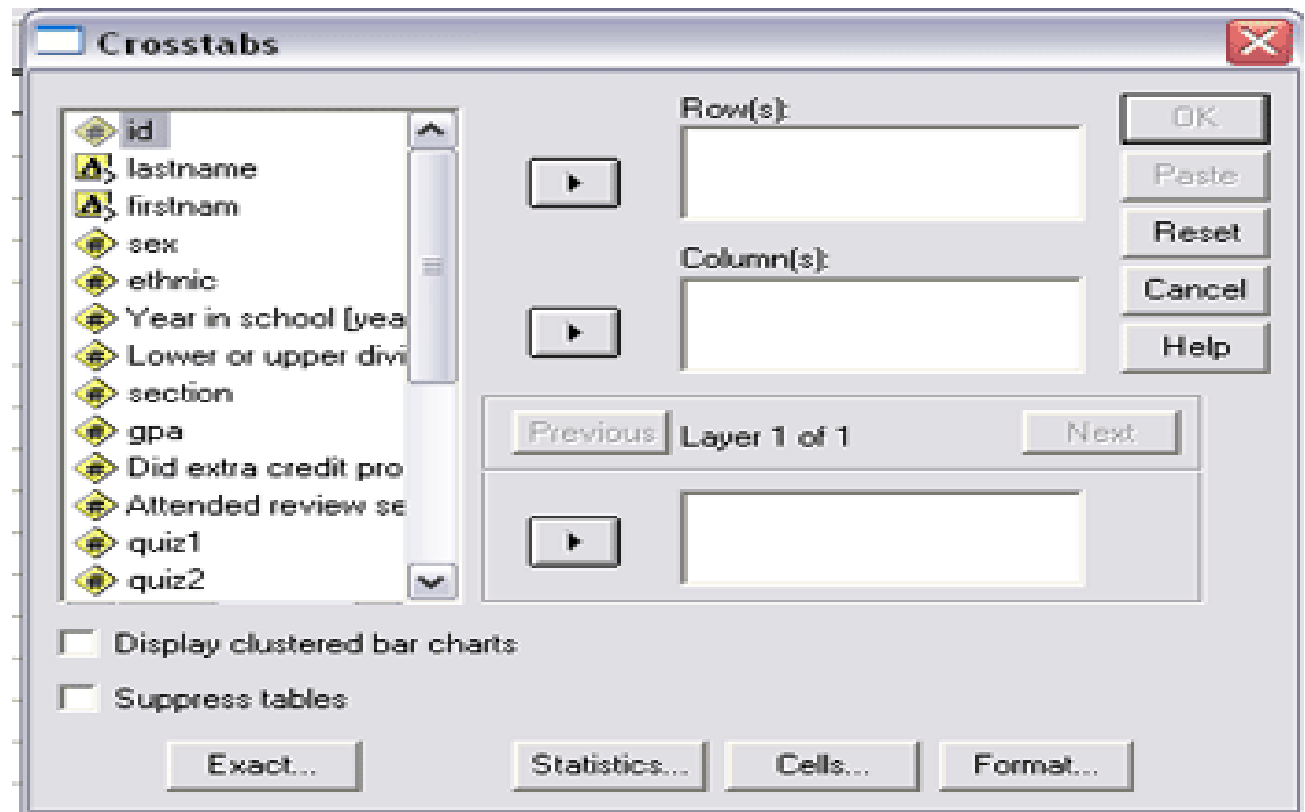
Έλεγχος ανεξαρτησίας χ^2

- Επειδή η χ^2 θεωρείται συχνά έλεγχος σχέσης (το αντίθετο της ανεξαρτησίας) & έχει ορισμένους περιορισμούς, ο Pearson πρότεινε το μέτρο (ϕ), που διαιρεί την τιμή της χ^2 με το πλήθος του δείγματος (N) και στη συνέχεια λαμβάνει τη θετική τετραγωνική ρίζα του αποτελέσματος.
- Ο σκοπός ήταν να υπάρξει ένα τυποποιημένο μέτρο σχέσης που θα λαμβάνει τιμές μεταξύ 0 & 1, όπου το 0 υποδηλώνει εντελώς ανεξάρτητες μεταβλητές, ενώ το 1 ισχυρή σχέση μεταξύ μεταβλητών.
- Με ακόμα λιγότερα προβλήματα είναι το τεστ Cramer's V $\chi^2/[N(k-1)]$ (k είναι ο μικρότερος από τον αριθμό των στηλών και των γραμμών), του οποίου οι τιμές ποικίλουν μεταξύ 0 & 1 και χρησιμοποιείται ευρέως για τη δύναμη της σχέσης μεταξύ των μεταβλητών σε έναν έλεγχο chi-square



Ο έλεγχος chi-square βήμα προς βήμα

Πάμε στο μενού analyze-
descriptive statistics-
crosstabs



Αφού επιλέξουμε τις μεταβλητές που θέλουμε να αναλύσουμε πατάμε το κουμπί cells



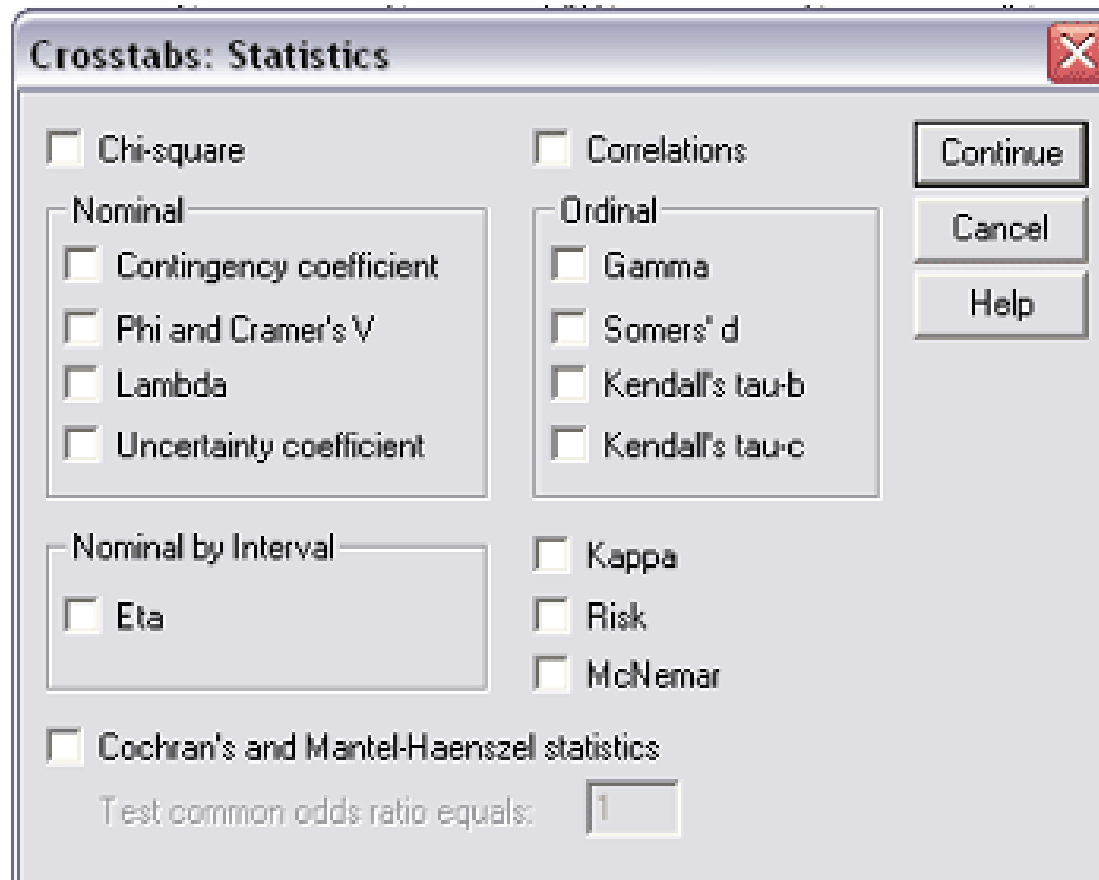
Cell Display

- Observed count: ο πραγματικός αριθμός παρατηρήσεων σε κάθε κελί
- Expected count: Η αναμενόμενη τιμή για κάθε κελί
- Row percentages: το ποσοστό των τιμών σε κάθε κελί για κάθε γραμμή
- Column percentages: το ποσοστό των τιμών σε κάθε κελί για κάθε στήλη
- Total percentages: το ποσοστό των τιμών σε κάθε κελί για ολόκληρο τον πίνακα
- Unstandardized residuals: πραγματική τιμή μείον αναμενόμενη τιμή



Συνήθως μαζί με την ανάλυση crosstabs γίνεται και ο έλεγχος chi-square

Πατάμε το πλήκτρο statistics και επιλέγουμε το chi-square



Crosstab

			Smoking						Total
			None	Light	Medium	Heavy	No Alcohol	Alcohol	None
Staff Group	Sr Managers	Count	4	2	3	2	0	11	22
		Expected Count	4,7	3,3	3,7	1,5	1,0	7,7	22,0
		Residual	-,7	-1,3	-,7	,5	-1,0	3,3	
	Jr Managers	Count	4	3	7	4	1	17	36
		Expected Count	7,6	5,5	6,1	2,5	1,7	12,6	36,0
		Residual	-3,6	-2,5	,9	1,5	-,7	4,4	
	Sr Employees	Count	25	10	12	4	5	46	102
		Expected Count	21,6	15,5	17,2	7,1	4,8	35,7	102,0
		Residual	3,4	-5,5	-5,2	-3,1	,2	10,3	
	Jr Employees	Count	18	24	33	13	10	78	176
		Expected Count	37,3	26,8	29,7	12,3	8,3	61,6	176,0
		Residual	-19,3	-2,8	3,3	,7	1,7	16,4	
	Secretaries	Count	10	6	7	2	7	18	50
		Expected Count	10,6	7,6	8,4	3,5	2,4	17,5	50,0
		Residual	-,6	-1,6	-1,4	-1,5	4,6	,5	
	National Average	Count	42	29	20	9	0	0	100
		Expected Count	21,2	15,2	16,9	7,0	4,7	35,0	100,0
		Residual	20,8	13,8	3,1	2,0	-4,7	-35,0	
	Total	Count	103	74	82	34	23	170	486
		Expected Count	103,0	74,0	82,0	34,0	23,0	170,0	486,0



Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	117,025(a)	25	,000
Likelihood Ratio	150,226	25	,000
Linear-by-Linear Association	51,455	1	,000
N of Valid Cases	486		

a 11 cells (30,6%) have expected count less than 5. The minimum expected count is 1,04.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,491	,000
	Cramer's V	,219	,000
N of Valid Cases		486	

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.



- Το τελευταίο βήμα ανάλυσης είναι η παρατήρηση των πραγματικών και των αναμενόμενων τιμών σε κάθε κελί
- Με την πρώτη ματιά βλέπουμε ότι διαφέρουν αισθητά
- Η τιμή chi-square είναι υψηλή (117,025) και η πιθανότητα σφάλματος μικρότερη από 0.05
- ΠΡΟΣΟΧΗ!!! **Πριν** σχολιάσουμε το chi-square πρέπει να δούμε στο κάτω μέρος του πίνακα το ποσοστό των κελιών με αναμενόμενες τιμές κάτω από 5.
- Αν το ποσοστό αυτό είναι μέχρι 20%, τότε το chi-square είναι έγκυρο και μπορούμε να πάμε να δούμε αν υποδηλώνει συσχέτιση μεταξύ των δύο μεταβλητών ή όχι
- Αν το ποσοστό είναι από 20% και πάνω, τότε το chi-square είναι άκυρο και ΔΕΝ εξετάζουμε τις τιμές του



- Count: ο αριθμός των παρατηρήσεων σε κάθε κελί
- Exp. Count: ο αριθμός που θα εμφανιζόταν αν οι δύο μτβλ. Ήταν εντελώς ανεξάρτητες μεταξύ τους
- Residual: η πραγματική μείον την αναμενόμενη τιμή
- Row total: ο αριθμός των παρατηρήσεων σε κάθε γραμμή
- Column total: ο αριθμός των παρατηρήσεων σε κάθε στήλη
- Chi-square: pearson & likelihood ratio: Δύο διαφορετικές μέθοδοι για υπολογισμό των στατιστικών της chi-square. Όταν το N είναι μεγάλο οι τιμές αυτές θα είναι σχεδόν ίσες



- Value: για τις μεθόδους pearson & maximum likelihood, καθώς η τιμή ελέγχου αυξάνει, η πιθανότητα οι δύο μεταβλητές να μην είναι ανεξάρτητες, επίσης αυξάνει
- Degrees of freedom: βαθμοί ελευθερίας είναι ο αριθμός των επιπέδων στην πρώτη μεταβλητή μείον 1 ($6-1=5$), επί τον αριθμό των επιπέδων στη δεύτερη μεταβλητή μείον 1 ($6-1=5$) ($5 \times 5=25$)
- Significance: η πιθανότητα τα αποτελέσματα του ελέγχου να προέκυψαν τυχαία. Η μικρή τιμή p υποδηλώνει ότι οι πραγματικές τιμές διαφέρουν σημαντικά από τις αναμενόμενες τιμές



Τέλος Ενότητας

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αθηνών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Σημειώματα

Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση 1.0.

Έχουν προηγηθεί οι κάτωθι εκδόσεις:

- Έκδοση διαθέσιμη [εδώ](#).



Σημείωμα Αναφοράς

Copyright Εθνικών και Καποδιστριακών Πανεπιστημίων Αθηνών, Δρ. Σταμάτης Πουλακιδάκος 2015. Σταμάτης Πουλακιδάκος. «Media Monitoring. Εισαγωγή & Ανάλυση δεδομένων με το SPSS». Έκδοση: 1.0. Αθήνα 2015. Διαθέσιμο από τη δικτυακή διεύθυνση: <http://opencourses.uoa.gr/courses/MEDIA102/>.



Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



[1] <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Ως **Μη Εμπορική** ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο

Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.



Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.

