

2^ο ΕΡΓΑΣΤΗΡΙΟ

Στο 2^ο κατά σειρά εργαστήριο θα παρουσιάσουμε πώς εξάγουμε **περιγραφικούς στατιστικούς δείκτες** για καθεμία από τις μεταβλητές που υπάρχουν σε μια βάση δεδομένων, χρησιμοποιώντας το πρόγραμμα SPSS. Επίσης, θα παρουσιάσουμε τη **διαδικασία επιλογής περιπτώσεων** (select cases), καθώς και την **προσθήκη ή διαγραφή περιπτώσεων και μεταβλητών** σε ήδη υπάρχουσα βάση δεδομένων (insert cases, insert variables και clear).

Για να επιλέξουμε τον σωστό τρόπο περιγραφής των δεδομένων κάθε μεταβλητής (δηλ. τους σωστούς περιγραφικούς στατιστικούς δείκτες) θα πρέπει να είμαστε σε θέση να αναγνωρίζουμε το είδος κάθε μεταβλητής και τη φύση των δεδομένων που μας δίνει, δηλαδή εάν πρόκειται για κατηγορική-ονομαστική μεταβλητή (nominal), ιεραρχική μεταβλητή¹ (ordinal), μεταβλητή ίσων διαστημάτων (interval) ή αναλογική μεταβλητή (ratio).

Χρειάζεται, επίσης, να εξηγήσουμε ότι τα τέσσερα παραπάνω είδη μεταβλητών ταξινομούνται και σε δύο ευρύτερες κατηγορίες: (α) τις ποιοτικές και (β) τις ποσοτικές μεταβλητές. **Ποιοτικές μεταβλητές** είναι εκείνες των οποίων οι τιμές αναφέρονται σε κατηγορίες, ονόματα ή θέσεις κατάταξης, όπως το φύλο (άντρας/γυναίκα), η εθνικότητα (Έλληνας/Γερμανός/Ιταλός) ή η σειρά τερματισμού (π.χ. 1^{ος}, 2^{ος}, 3^{ος} κ.λπ.) και δεν έχουν ποσοτικές ιδιότητες, δηλαδή δεν μπορούμε να τις προσθέσουμε, να τις αφαιρέσουμε, να τις πολλαπλασιάσουμε, κ.λπ. **Ποσοτικές μεταβλητές** είναι εκείνες των οποίων οι τιμές έχουν ποσοτικές ιδιότητες (δηλ. μπορούμε να κάνουμε με αυτές αριθμητικές πράξεις) και γνωρίζουμε ακριβώς πόσο απέχουν μεταξύ τους. Μπορούμε λοιπόν να καταλάβουμε ότι ποιοτικές μεταβλητές είναι οι κατηγορικές (nominal) και οι ιεραρχικές (ordinal), ενώ ποσοτικές μεταβλητές είναι οι μεταβλητές ίσων διαστημάτων (interval) και οι αναλογικές (ratio). Η διάκριση των μεταβλητών παρουσιάζεται συνοπτικά στο ακόλουθο διάγραμμα:



¹ Οι ιεραρχικές μεταβλητές ονομάζονται και 'διατάξιμες' ή 'τακτικές' μεταβλητές.

Οι μεταβλητές διακρίνονται, επίσης, σε συνεχείς (continuous variables) και σε ασυνεχείς ή διακριτές (discrete variables). **Συνεχείς** ονομάζονται οι μεταβλητές που μπορούν να λάβουν (έστω και θεωρητικά) μια οποιαδήποτε τιμή εντός ενός διαστήματος, συμπεριλαμβανομένων και δεκαδικών αριθμών (π.χ. ύψος, βάρος, εισόδημα). Από την άλλη μεριά, **ασυνεχείς ή διακριτές** ορίζονται οι μεταβλητές που μπορούν να λάβουν μόνο συγκεκριμένες τιμές εντός ενός διαστήματος (συνήθως τιμές ισοδύναμες με ακέραιους αριθμούς, όπως ο αριθμός παιδιών σε μία οικογένεια).

Τέλος, οι μεταβλητές που παίρνουν μόνο δύο τιμές (π.χ. Ναι/Όχι, Αρσενικό/Θηλυκό) ονομάζονται **διχοτομικές μεταβλητές** (dichotomous), καθώς διχοτομούν το δείγμα μας σε δύο υποομάδες ανάλογα με την τιμή που παίρνει το κάθε άτομο στην αντίστοιχη μεταβλητή. Οι διχοτομικές μεταβλητές είναι συνήθως κατηγορικές-ονομαστικές μεταβλητές.

Έχοντας υπόψη όλα τα παραπάνω, οι περιγραφικοί στατιστικοί δείκτες και τα διαγράμματα που ενδείκνυνται για την περιγραφή κάθε είδους μεταβλητών παρουσιάζονται συνοπτικά στον παρακάτω πίνακα:

ΕΙΔΟΣ ΜΕΤΑΒΛΗΤΩΝ	ΠΕΡΙΓΡΑΦΙΚΟΙ ΣΤΑΤΙΣΤΙΚΟΙ ΔΕΙΚΤΕΣ ΚΑΙ ΔΙΑΓΡΑΜΜΑΤΑ
ΠΟΙΟΤΙΚΕΣ <ul style="list-style-type: none"> - Κατηγορικές - Ιεραρχικές 	Κατανομές συχνοτήτων, δεσπόζουσα τιμή, ραβδόγραμμα, κυκλικό διάγραμμα.
ΠΟΣΟΤΙΚΕΣ (ΑΣΥΝΕΧΕΙΣ): <ul style="list-style-type: none"> - Οιονεί (ή πλαστικές) ίσων διαστημάτων - Ίσων διαστημάτων - Αναλογικές 	Κατανομές συχνοτήτων, δείκτες κεντρικής τάσης (Μ.Ο., δεσπόζουσα, διάμεσος), δείκτες διασποράς, ραβδόγραμμα, διάγραμμα κουτιού.
ΠΟΣΟΤΙΚΕΣ (ΣΥΝΕΧΕΙΣ): <ul style="list-style-type: none"> - Ίσων διαστημάτων - Αναλογικές 	Δείκτες κεντρικής τάσης (Μ.Ο., δεσπόζουσα, διάμεσος), δείκτες διασποράς, δείκτες λοξότητας και κυρτότητας (σχήματος κατανομής), ιστόγραμμα, διάγραμμα κουτιού, διάγραμμα γραμμής, διάγραμμα διασποράς.

Σε αυτό το εργαστήριο θα εστιάσουμε στους περιγραφικούς δείκτες που εφαρμόζονται στις ποιοτικές μεταβλητές (κατηγορικές και ιεραρχικές), καθώς και στις ποσοτικές ασυνεχείς μεταβλητές. Πιο συγκεκριμένα, θα εξηγήσουμε πώς εξάγουμε στο SPSS κατανομές συχνοτήτων, ραβδογράμματα, κυκλικά διαγράμματα και δείκτες κεντρικής τάσης.

1. Περιγραφικοί στατιστικοί δείκτες και διαγράμματα για ποιοτικές (κατηγορικές και ιεραρχικές) μεταβλητές.

ΣΗΜΕΙΩΣΗ: Οι περιγραφικοί στατιστικοί δείκτες που παρουσιάζονται σε αυτή την ενότητα μπορούν να εφαρμοστούν και σε ποσοτικές μεταβλητές που είναι διακριτές-ασυνεχείς. Ωστόσο, επειδή οι ποσοτικές μεταβλητές μπορούν να περιγραφούν και με άλλους στατιστικούς δείκτες, οι μεταβλητές αυτές παρουσιάζονται ξεχωριστά στην επόμενη ενότητα.

Η περιγραφική στατιστική για ποιοτικές μεταβλητές (nominal και ordinal) περιλαμβάνει τον υπολογισμό συχνοτήτων. Συχνότητα είναι ο αριθμός των ατόμων (συμμετεχόντων μιας έρευνας) που εμπίπτουν σε μια κατηγορία. Οι συχνότητες παρουσιάζονται με δύο τρόπους: α) τις απόλυτες συχνότητες και β) τις σχετικές συχνότητες (ποσοστά). **Απόλυτη συχνότητα** είναι ο απόλυτος αριθμός των ατόμων που εμπίπτουν σε μια κατηγορία, ενώ **σχετική συχνότητα** είναι το ποσοστό των ατόμων που εμπίπτουν σε μια κατηγορία επί του συνόλου των συμμετεχόντων της έρευνας.

Για την περιγραφή ποιοτικών μεταβλητών ενδείκνυται, επίσης, η δημιουργία κυκλικών διαγραμμάτων (pie charts) ή ραβδογραμμάτων (bar charts).

ΠΑΡΑΔΕΙΓΜΑ 1: Ανοίγουμε το αρχείο Store_spss_lab.sav που περιλαμβάνει δεδομένα από μια έρευνα ικανοποίησης πελατών τεσσάρων καταστημάτων τηλεπικοινωνίας. Ας υποθέσουμε ότι θέλουμε να περιγράψουμε το δείγμα της συγκεκριμένης έρευνας ως προς τα εξής δημογραφικά χαρακτηριστικά: φύλο, ηλικία και μορφωτικό επίπεδο. Επίσης, ας υποθέσουμε ότι θέλουμε να απεικονίσουμε και γραφικά, με τη χρήση κυκλικών διαγραμμάτων (pie charts), τα ποσοστά των ερωτώμενων που εμπίπτουν σε κάθε δημογραφική κατηγορία.

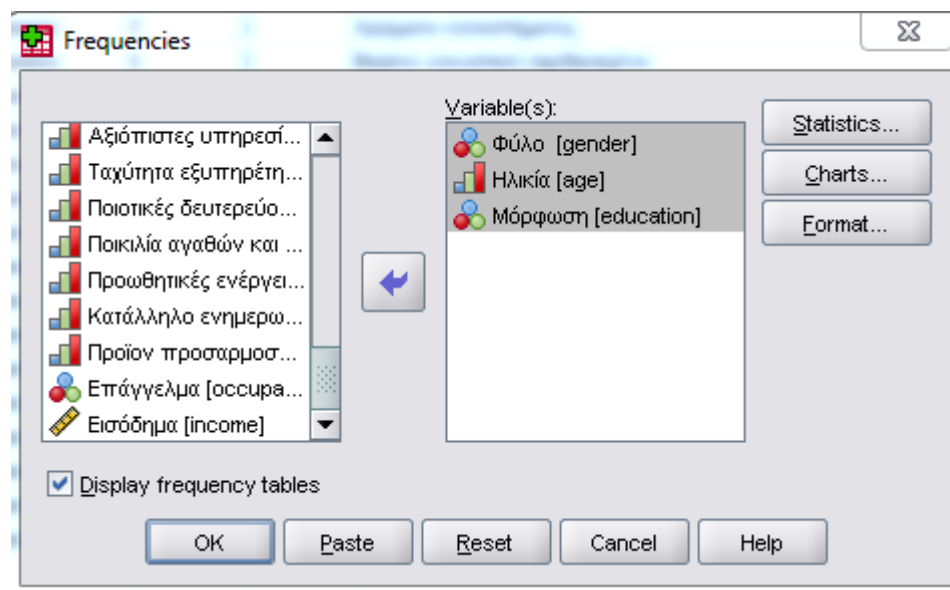
Για το σκοπό αυτό ακολουθούμε τα εξής βήματα:
Analyze → Descriptive Statistics → Frequencies

Στο πλαίσιο Variable(s) μεταφέρουμε τις μεταβλητές Φύλο, Ηλικία, και Μόρφωση.

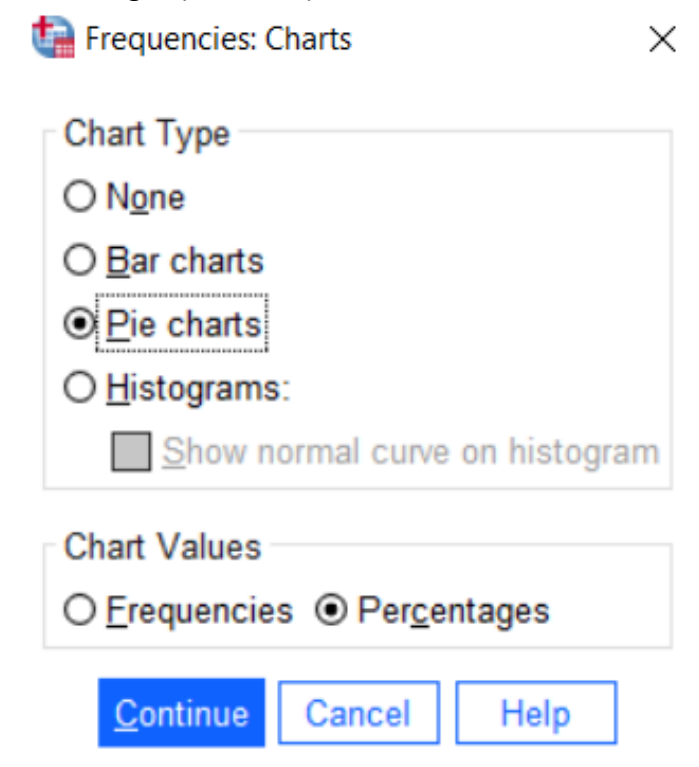
Η μεταβλητή «φύλο» είναι κατηγορική μεταβλητή, ενώ οι μεταβλητές «ηλικία» και «μόρφωση» είναι ιεραρχικές μεταβλητές καθώς οι τιμές τους μπορούν να μπουν σε αύξουσα ή φθίνουσα σειρά. Όλοι οι περιγραφικοί στατιστικοί δείκτες που εφαρμόζονται σε κατηγορικές μεταβλητές μπορούν να εφαρμοστούν και σε ιεραρχικές μεταβλητές, οι οποίες ωστόσο σε κάποιες περιπτώσεις (όπως π.χ. στις περιπτώσεις των «πλαστικών» κλιμάκων τύπου Likert²) περιγράφονται και με άλλους

²Στις κοινωνικές έρευνες, αρκετές από τις ιεραρχικές κλίμακες που χρησιμοποιούνται, όπως οι κλίμακες μέτρησης της ακαδημαϊκής επίδοσης (π.χ. 0-20 με βάση το 10) ή οι διαβαθμιστικές κλίμακες αξιολόγησης (1=καθόλου αντιπροσωπευτικό.....5=πολύ αντιπροσωπευτικό) ή οι κλίμακες μέτρησης των στάσεων (5=συμφωνώ απόλυτα.....1=διαφωνώ απόλυτα), παρά την αντικειμενική αδυναμία να αποδοθούν οι τιμές τους

στατιστικούς δείκτες (όπως ο Μέσος Όρος) που θα παρουσιαστούν στην επόμενη ενότητα.



Στη συνέχεια, επιλέγουμε το κουμπί charts. Εκεί επιλέγουμε Pie Charts και Percentages (Ποσοστά).



σε κλίμακα ίσων διαστημάτων, στην πράξη θεωρούνται ότι μετριοούνται σε κλίμακα ίσων διαστημάτων καθώς μόνο τότε μπορούμε να τις χειριστούμε και να τις επεξεργαστούμε πιο εύκολα στατιστικά (π.χ. βρίσκοντας τον Μέσο Όρο). Σε αυτές τις περιπτώσεις μιλάμε για «πλαστικές κλίμακες ίσων διαστημάτων» και για «πλαστικά διαστήματα» (Ρούσσοσ & Τσαούσης, 2002).

Έτσι, προκύπτουν οι παρακάτω πίνακες για κάθε δημογραφικό χαρακτηριστικό. Η στήλη Frequency περιγράφει τις απόλυτες συχνότητες για κάθε τιμή (κατηγορία) της μεταβλητής. Η στήλη Percent δείχνει τις αντίστοιχες σχετικές συχνότητες (ποσοστά) για κάθε τιμή (κατηγορία) της μεταβλητής. Η στήλη Valid Percent παρουσιάζει τα ποσοστά αφού έχουν αφαιρεθεί περιπτώσεις με missing values (στη δική μας περίπτωση επειδή δεν έχουμε missing values το valid percent είναι ίδιο με το percent). Η στήλη cumulative percent δείχνει το αθροιστικό ποσοστό που προκύπτει αφού σε κάθε τιμή (κατηγορία) έχει προστεθεί το ποσοστό όλων των προηγούμενων τιμών. Στην περίπτωση του φύλου παρατηρούμε πως το 47,5% (95 άτομα) των ερωτώμενων είναι άνδρες ενώ το 52,5% (105 άτομα) είναι γυναίκες.

Φύλο

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Άνδρας	95	47,5	47,5	47,5
Γυναίκα	105	52,5	52,5	100,0
Total	200	100,0	100,0	

Ως προς την ηλικία, βάσει του παρακάτω πίνακα, διαπιστώνεται πως το 40,5% (81 άτομα) του δείγματος ανήκουν στην ηλικιακή ομάδα 18-30 ετών, το 38% (76 άτομα) στην ομάδα 31-40 ετών, το 14,5% (29 άτομα) στην ομάδα 41-50 ετών, ενώ το 7% (14 άτομα) είναι ηλικίας 51 ετών και άνω. Επίσης, κοιτώντας το αθροιστικό ποσοστό αυτής της μεταβλητής συμπεραίνουμε, μεταξύ άλλων, ότι το 78,5% των συμμετεχόντων στην έρευνα είναι ηλικίας έως 40 ετών.

Ηλικία

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 18-30	81	40,5	40,5	40,5
31-40	76	38,0	38,0	78,5
41-50	29	14,5	14,5	93,0
51-60	7	3,5	3,5	96,5
61 και άνω	7	3,5	3,5	100,0
Total	200	100,0	100,0	

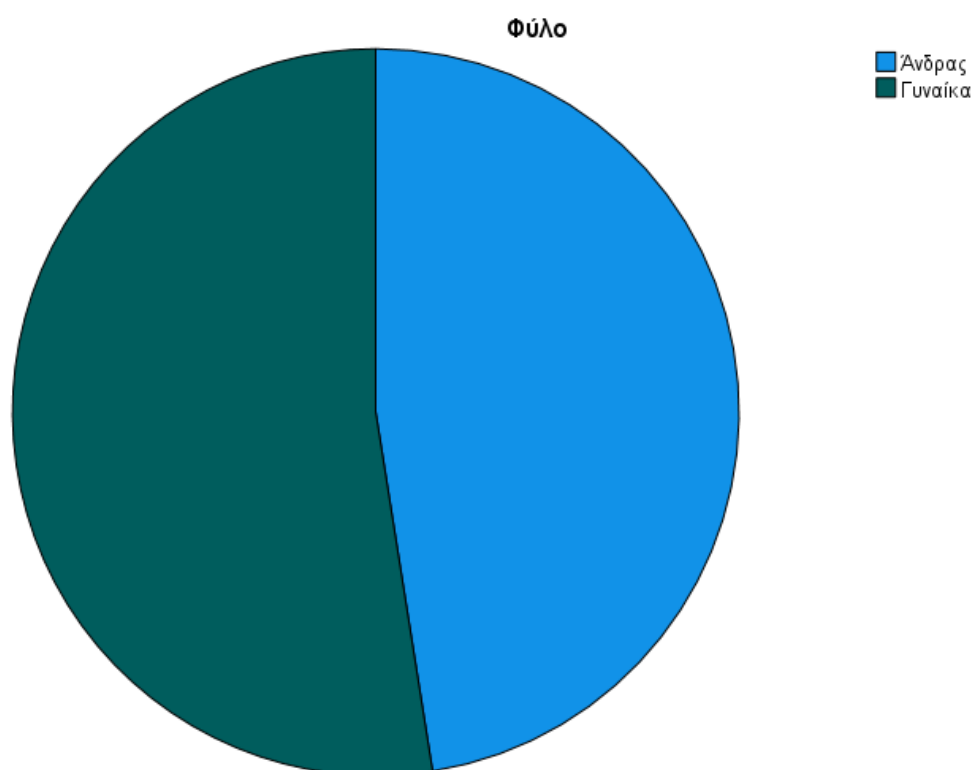
Σχετικά με το μορφωτικό επίπεδο των ερωτώμενων, το 26% (52 άτομα) έχουν τελειώσει το Γυμνάσιο-Λύκειο, το 23% (46 άτομα) είναι φοιτητές, το 42,5% (85 άτομα) είναι απόφοιτοι ΤΕΙ-Πανεπιστημίου, ενώ το 8,5% (17 άτομα) είναι κάτοχοι μεταπτυχιακού διπλώματος. Κοιτώντας το αθροιστικό ποσοστό αυτής της

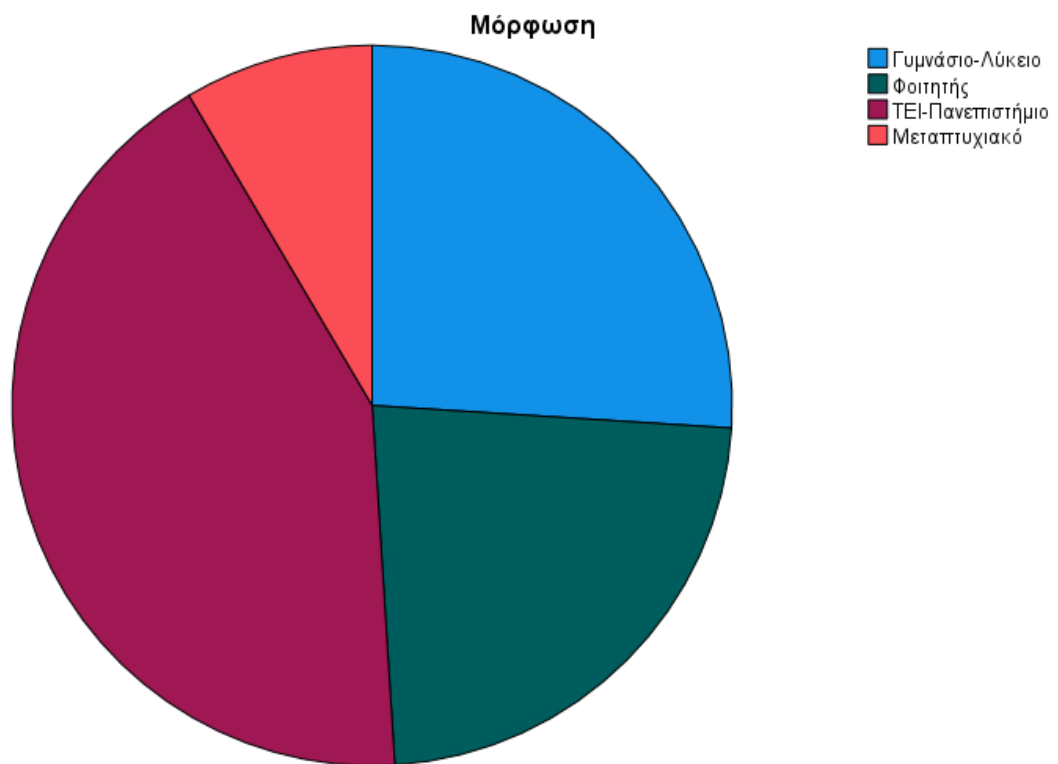
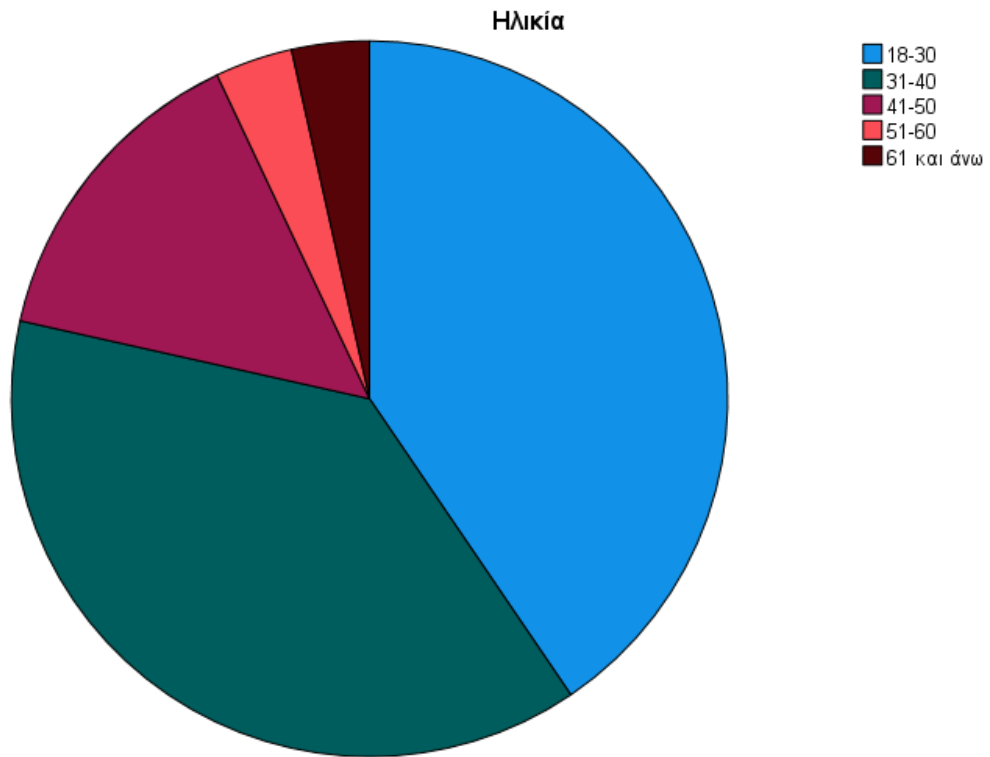
μεταβλητής συμπεραίνουμε ότι η συντριπτική πλειοψηφία των συμμετεχόντων (91,5%) έχουν ανώτατο επίπεδο εκπαίδευσης που δεν υπερβαίνει το πτυχίο ΑΕΙ-ΤΕΙ.

Μόρφωση

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Γυμνάσιο-Λύκειο	52	26,0	26,0	26,0
	Φοιτητής	46	23,0	23,0	49,0
	ΤΕΙ-Πανεπιστήμιο	85	42,5	42,5	91,5
	Μεταπτυχιακό	17	8,5	8,5	100,0
	Total	200	100,0	100,0	

Αμέσως μετά τους πίνακες, στο output εμφανίζονται τα κυκλικά διαγράμματα.

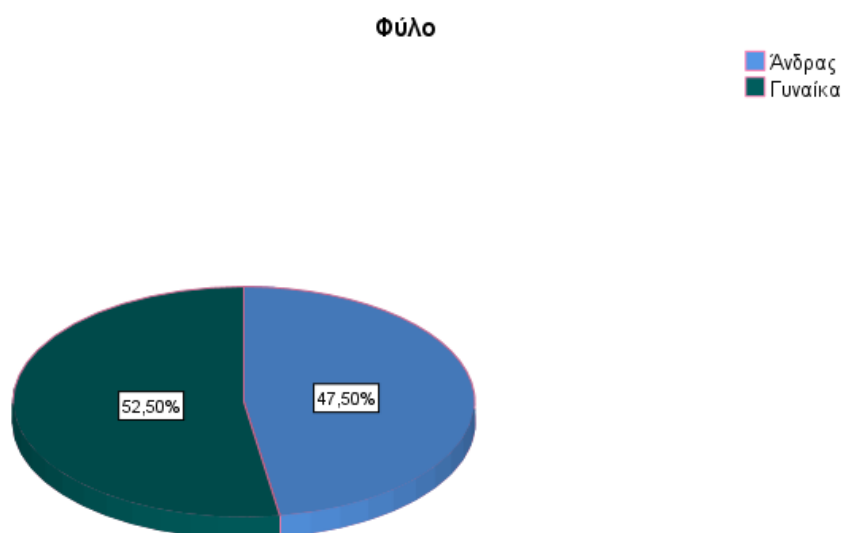




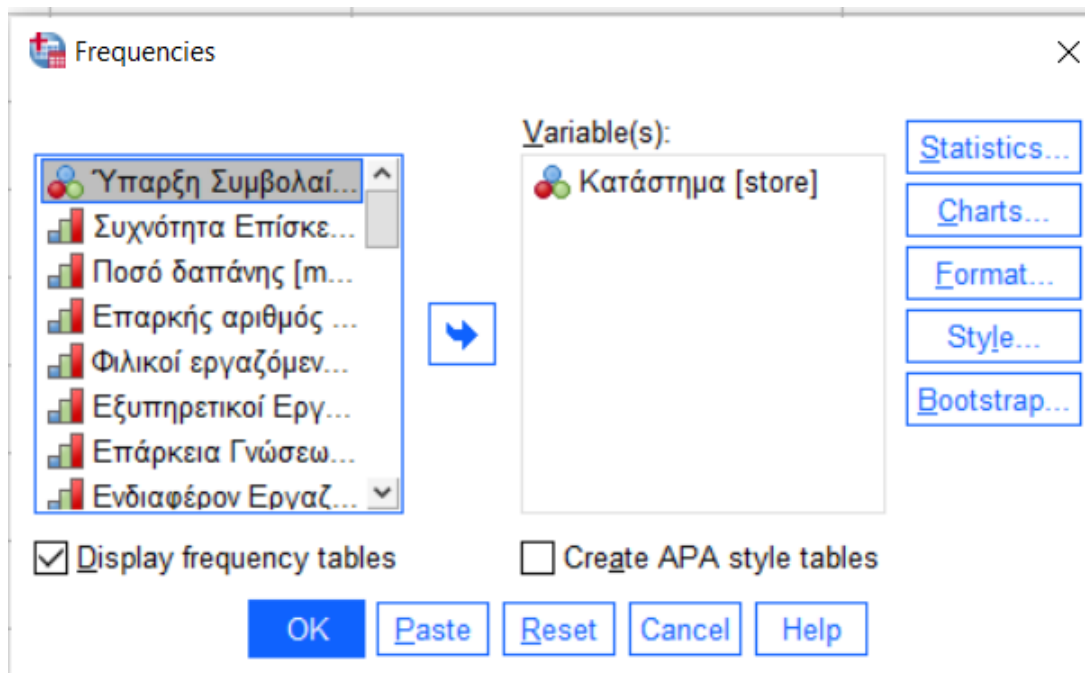
Βέβαια, η μορφή στην οποία εμφανίζονται είναι αρκετά απλή και θα πρέπει να τα επεξεργαστούμε. Το SPSS δίνει τη δυνατότητα επεξεργασίας των γραφημάτων. Για

παράδειγμα, αν θέλουμε να επεξεργαστούμε τη μορφή του γραφήματος για το φύλο, πηγαίνουμε στο output πάνω στο γράφημα και κάνουμε διπλό κλικ. Αμέσως ανοίγει το πλαίσιο chart editor. Εδώ, μπορούμε να αλλάξουμε χρώματα, να προσθέσουμε ετικέτες, να διορθώσουμε τους τίτλους και να αλλάξουμε γενικότερα τη μορφή του γραφήματος.

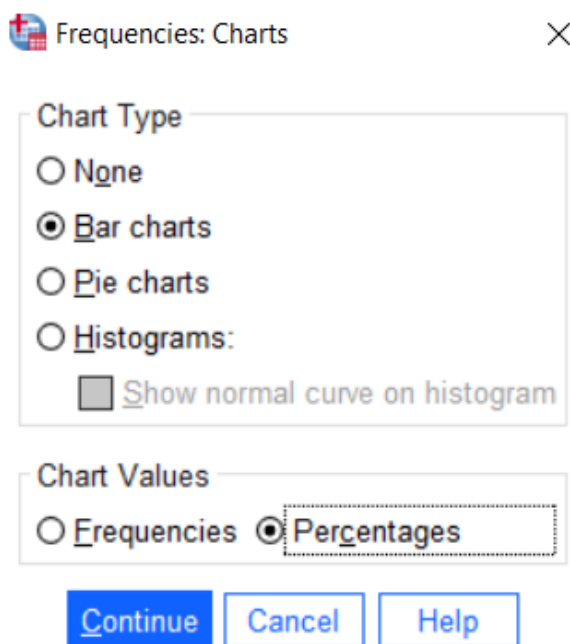
Για να προσθέσουμε ετικέτες, στο μενού Elements επιλέγουμε Show Data Labels. Για να αλλάξουμε χρώμα στα διάφορα κομμάτια του διαγράμματος απλά κάνουμε κλικ στα αντίστοιχα τμήματα και στο παράθυρο διαλόγου Properties που εμφανίζεται στην καρτέλα Fill & Border επιλέγουμε το επιθυμητό χρώμα και στη συνέχεια Apply. Αν θέλουμε να αλλάξουμε το διάγραμμα σε τρισδιάστατη μορφή τότε στην καρτέλα Depth & Angle επιλέγουμε 3-D και μετά Apply. Έτσι, το γράφημα αλλάζει μορφή.



ΠΑΡΑΔΕΙΓΜΑ 2: Ένας ερευνητής θέλει να βρει το ποσοστό των ερωτώμενων που επισκέπτονται καθένα από τα 4 διαφορετικά καταστήματα τηλεπικοινωνίας (Γερμανος, Cosmote, Vodafone, Wind). Επίσης, επιθυμεί να απεικονίσει τα ποσοστά αυτά με τη μορφή ραβδογράμματος. Έτσι, με τον ίδιο τρόπο Analyze → Descriptive Statistics → Frequencies μεταφέρω τη μεταβλητή Κατάστημα στο πλαίσιο Variable(s).



Από την επιλογή Charts επιλέγω Bar Charts και Percentages.



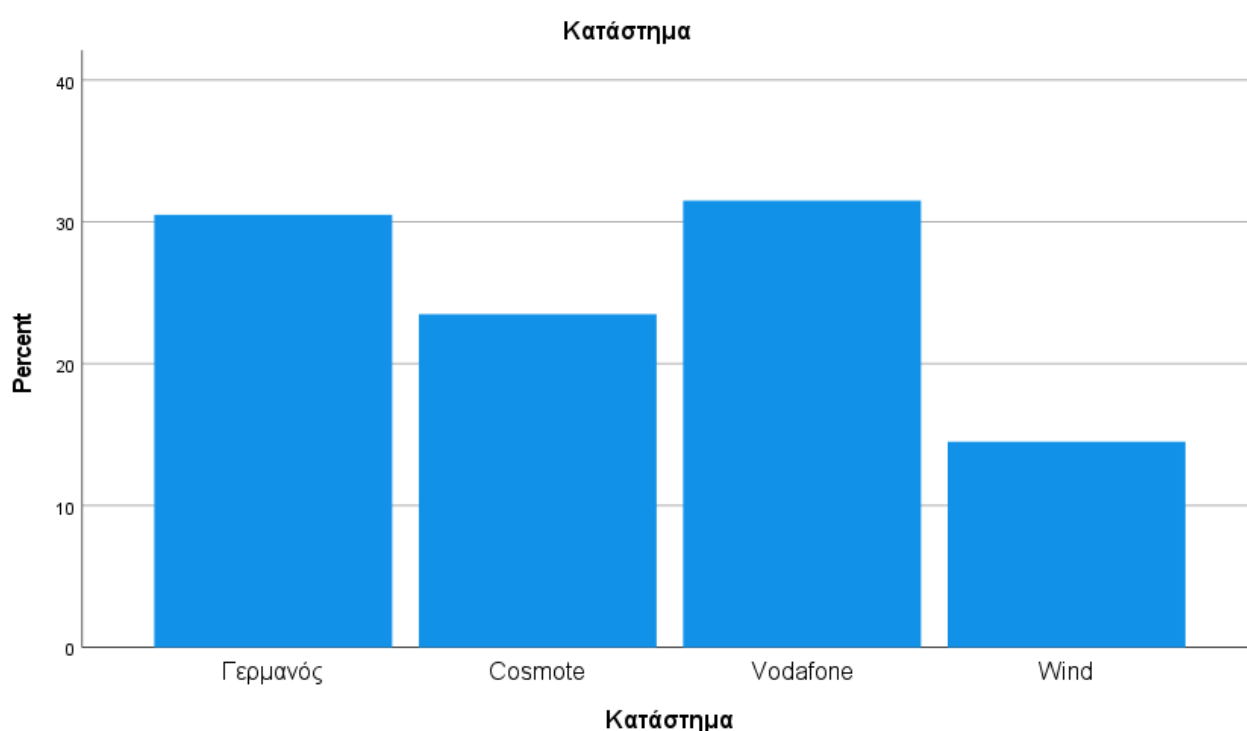
Έπειτα Continue και OK.

Στη συνέχεια στο Output εμφανίζεται ο πίνακας συχνοτήτων. Απαντώντας στο ερώτημα που θέσαμε, το 30,5% του δείγματος (61 άτομα) επισκέφτηκαν κατάστημα Γερμανός, το 23,5% (47 άτομα) επισκέφτηκαν κατάστημα της Cosmote, το 31,5% (63 άτομα) επισκέφτηκαν κατάστημα της Vodafone και το 14,5% του δείγματος (29

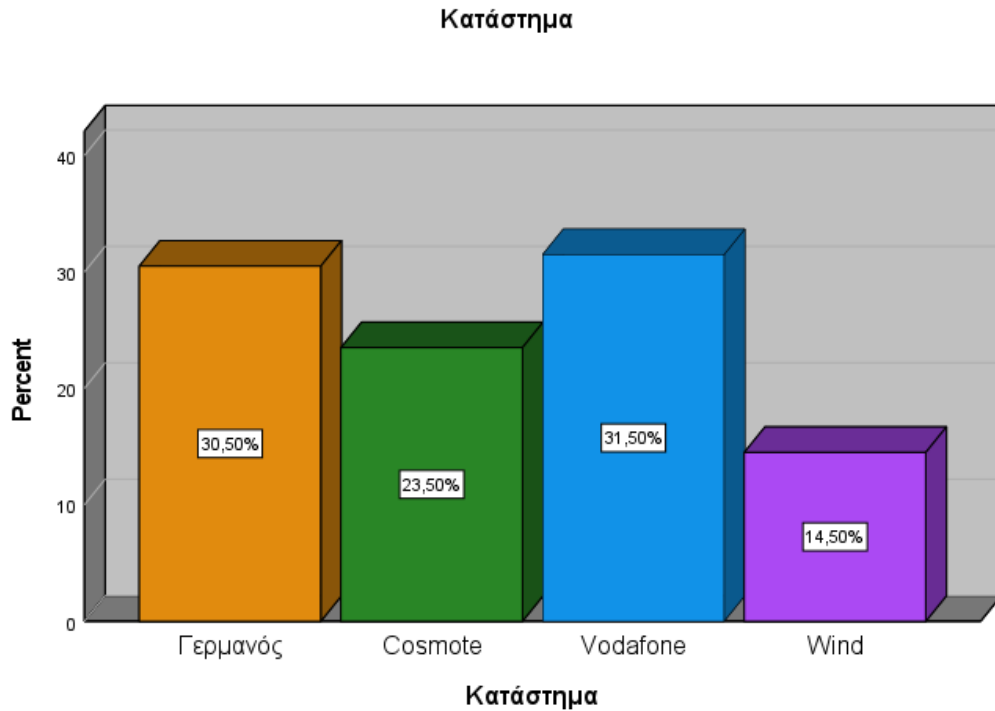
άτομα) επισκέφτηκαν κατάστημα της Wind. Συνεπώς, τα περισσότερα άτομα είχαν επισκεφτεί κατάστημα της εταιρίας Vodafone, ενώ η εταιρεία Wind είχε τις λιγότερες επισκέψεις.

		Κατάστημα			Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	Γερμανός	61	30,5	30,5	30,5
	Cosmote	47	23,5	23,5	54,0
	Vodafone	63	31,5	31,5	85,5
	Wind	29	14,5	14,5	100,0
	Total	200	100,0	100,0	

Αμέσως πιο κάτω εμφανίζεται το ραβδόγραμμα.



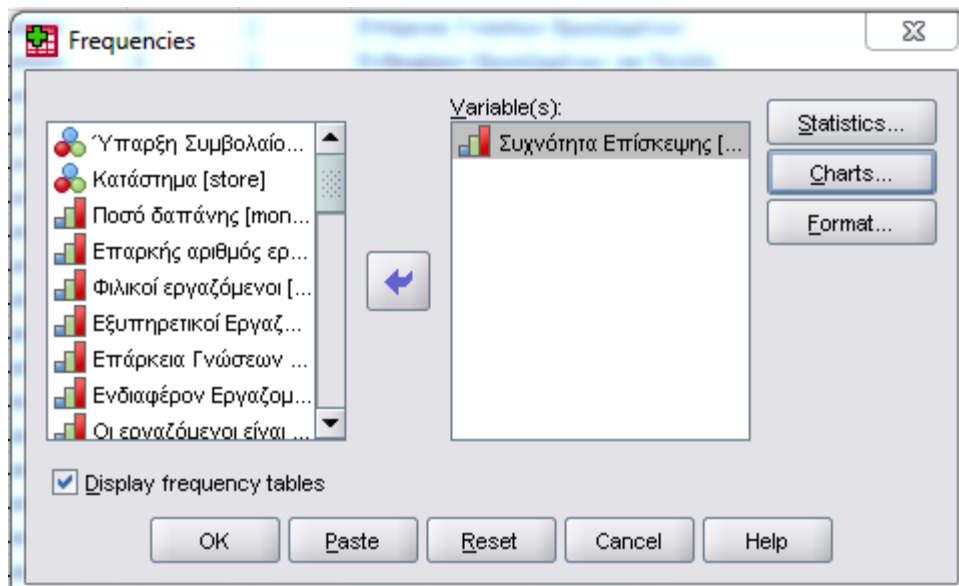
Πάλι με τη χρήση του Chart Editor (κάνοντας διπλό κλικ πάνω στο γράφημα) μπορούμε να μορφοποιήσουμε κατάλληλα το ραβδόγραμμα.



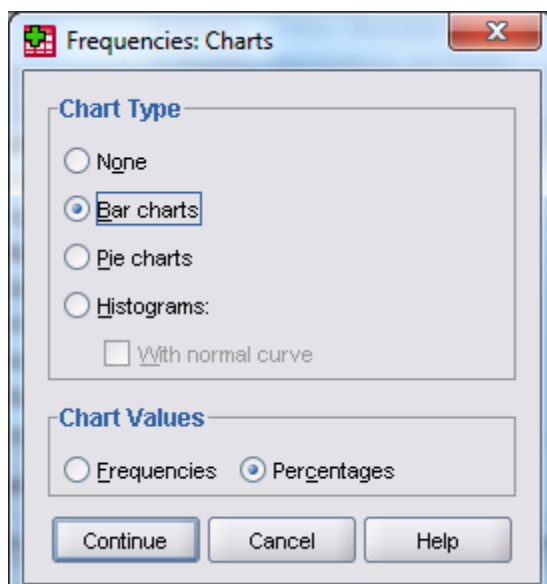
ΠΑΡΑΔΕΙΓΜΑ 3: Ας υποθέσουμε ότι ένας ερευνητής θέλει να βρει το ποσοστό των ερωτώμενων που επισκέπτονται το κατάστημα 2 έως 3 φορές το μήνα. Επίσης, επιθυμεί να απεικονίσει τα ποσοστά της συχνότητας επίσκεψης με τη μορφή ραβδογράμματος.

Σημείωση: Η μεταβλητή 'Συχνότητα Επίσκεψης' είναι ιεραρχική και μπορούμε να εφαρμόσουμε τους ίδιους περιγραφικούς στατιστικούς δείκτες που ισχύουν για τις κατηγορικές μεταβλητές.

Έτσι, με τον ίδιο τρόπο επιλέγουμε Analyze → Descriptive Statistics → Frequencies. Μεταφέρω τη μεταβλητή Συχνότητα επίσκεψης στο πλαίσιο Variable(s)



Από την επιλογή Charts επιλέγω Bar Charts και Percentages.



Έπειτα Continue και OK.

Στη συνέχεια στο Output εμφανίζεται ο πίνακας συχνοτήτων.

Απαντώντας στο ερώτημα που θέσαμε, το 17% (34 άτομα) του δείγματος επισκέπτεται το κατάστημα 2-3 φορές το μήνα.

Συχνότητα Επίσκεψης

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1 φορά	156	78,0	78,0	78,0
2-3 φορές	34	17,0	17,0	95,0
περισσότερες απο 3 φορές	10	5,0	5,0	100,0
Total	200	100,0	100,0	

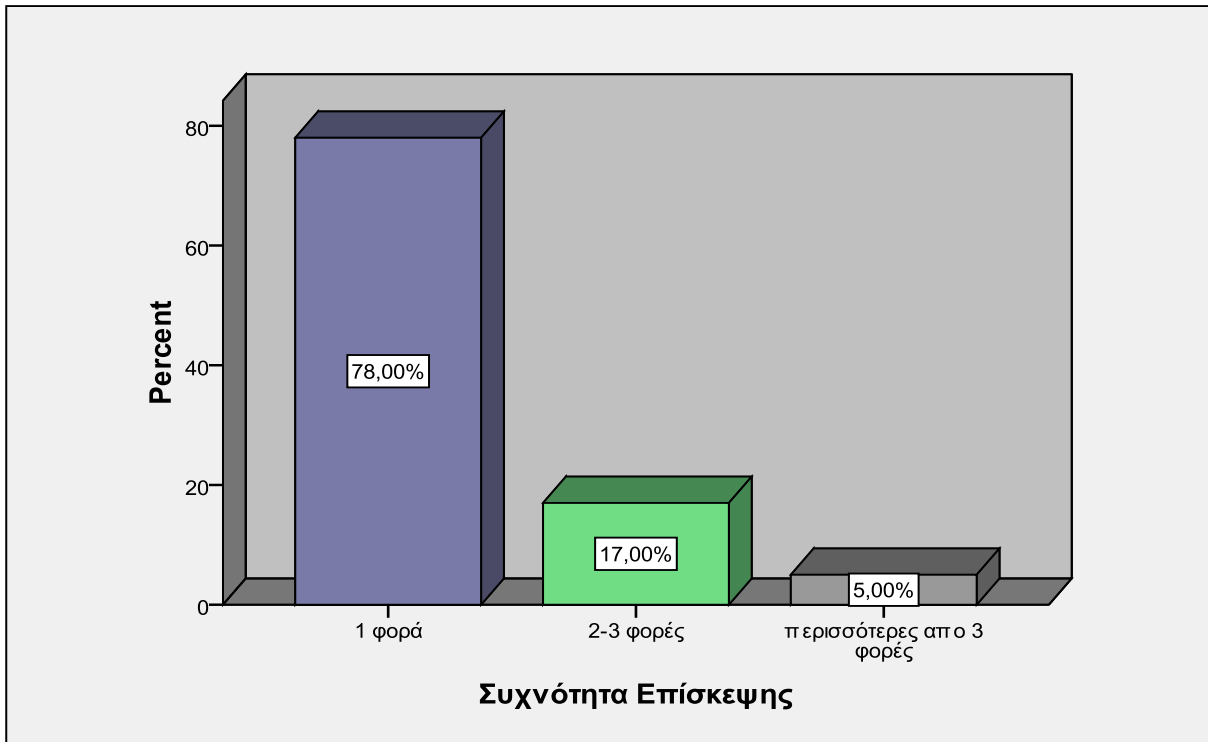
Αμέσως πιο κάτω εμφανίζεται το ραβδόγραμμα.

Συχνότητα Επίσκεψης



Πάλι με τη χρήση του Chart Editor (κάνοντας διπλό κλικ πάνω στο γράφημα) μπορούμε να μορφοποιήσουμε κατάλληλα το ραβδόγραμμα.

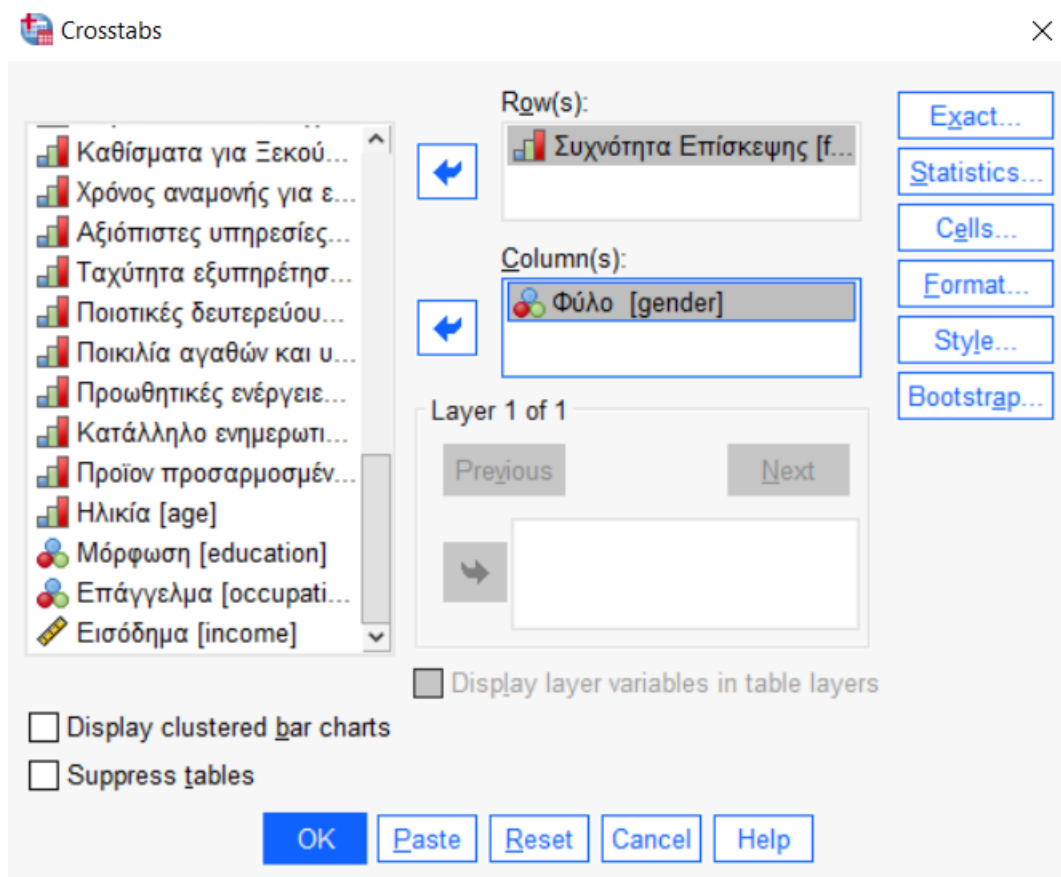
Συχνότητα Επίσκεψης



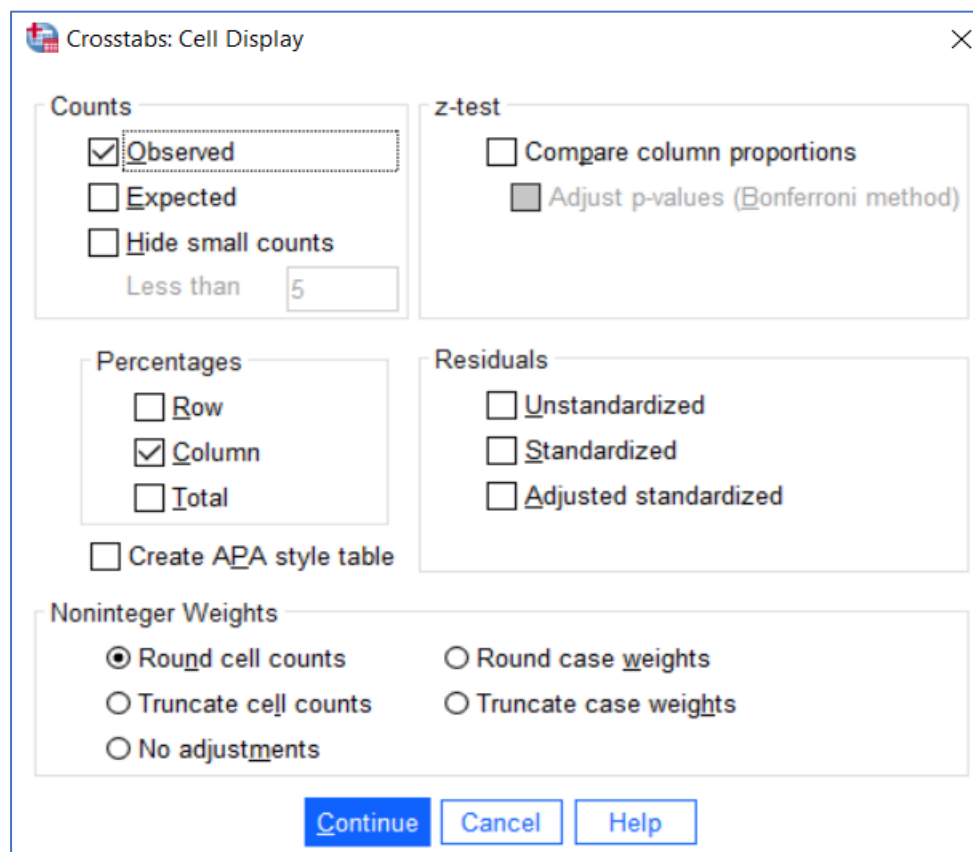
Ένας ερευνητής μπορεί να θέλει να παρουσιάσει τα ποσοστά των ερωτώμενων που εμπíπτουν σε κάθε κατηγορία μιας μεταβλητής ομαδοποιώντας τα ανάλογα με τις τιμές μιας άλλης nominal - ordinal μεταβλητής. Για παράδειγμα, ένας ερευνητής μπορεί να θέλει να παρουσιάσει τα ποσοστά της συχνότητας επίσκεψης στο κατάστημα ανά φύλο. Για το σκοπό αυτό ακολουθούμε τα εξής βήματα:

Analyze → Descriptive Statistics → Crosstabs

Στο παράθυρο που ανοίγει στο πλαίσιο Row(s) μεταφέρω τη μεταβλητή Συχνότητα Επίσκεψης ενώ στο πλαίσιο Column(s) μεταφέρω τη μεταβλητή Φύλο.



Στο κουτάκι Cells στο πλαίσιο Percentages επιλέγω Column.



Έπειτα, πατάω Continue και OK. Στο output εμφανίζεται ο παρακάτω πίνακας συνάφειας:

Συχνότητα Επίσκεψης * Φύλο Crosstabulation

		Φύλο		Total	
		Ανδρας	Γυναίκα		
Συχνότητα Επίσκεψης	1 φορά	Count	67	89	156
		% within Φύλο	70,5%	84,8%	78,0%
	2-3 φορές	Count	20	14	34
		% within Φύλο	21,1%	13,3%	17,0%
	περισσότερες από 3 φορές	Count	8	2	10
		% within Φύλο	8,4%	1,9%	5,0%
Total		Count	95	105	200
		% within Φύλο	100,0%	100,0%	100,0%

Με βάση τον πίνακα παρατηρούμε πως από το σύνολο των αντρών που συμμετείχαν στην έρευνα (95 άντρες) το 70,5% (67 άντρες) επισκέπτονται το κατάστημα 1 φορά το μήνα, το 21,1% (20 άντρες) το επισκέπτονται 2-3 φορές το μήνα και το 8,4% (8 άντρες) το επισκέπτονται περισσότερες από 3 φορές το μήνα. Αντίστοιχα, από το σύνολο των γυναικών που συμμετείχαν στην έρευνα (105 γυναίκες) το 84,8% (89 γυναίκες) επισκέπτονται το κατάστημα 1 φορά το μήνα, το 13,3% (14 γυναίκες) το επισκέπτονται 2 έως 3 φορές το μήνα και το 1,9% (2 γυναίκες) το επισκέπτονται περισσότερες από 3 φορές το μήνα. **Στο συγκεκριμένο δείγμα, δηλαδή, παρατηρούμε μεγαλύτερες συχνότητες επίσκεψης στην ομάδα των αντρών σε σύγκριση με εκείνη των γυναικών.**

2. Περιγραφικοί στατιστικοί δείκτες και διαγράμματα για ποσοτικές μεταβλητές τύπου scale.

Όπως προαναφέρθηκε, οι περιγραφικοί στατιστικοί δείκτες που εφαρμόζονται στις ποιοτικές μεταβλητές (δηλ. οι απόλυτες και οι σχετικές συχνότητες) δύνανται να εφαρμοστούν και σε ποσοτικές μεταβλητές που είναι διακριτές-ασυνεχείς. Σε αυτή την ενότητα, θα παρουσιάσουμε κάποιους άλλους περιγραφικούς δείκτες που εφαρμόζονται σε ποσοτικές μεταβλητές ή σε μεταβλητές που είναι *οιονεί* ποσοτικές (δηλ. «πλαστικές»). Συγκεκριμένα, θα μιλήσουμε για τους δείκτες κεντρικής τάσης, όπως τον αριθμητικό μέσο όρο, τη διάμεσο και τη δεσπόζουσα τιμή³. Επίσης, θα δούμε πώς μπορούμε να κατασκευάσουμε ιστογράμματα.

³Η δεσπόζουσα τιμή μπορεί να εφαρμοστεί και σε ποιοτικές μεταβλητές (κατηγορικές ή ιεραρχικές).

Τι είναι οι δείκτες κεντρικής τάσης;

Οι δείκτες κεντρικής τάσης είναι αριθμητικές τιμές που μας δείχνουν **το «κέντρο» των τιμών** μιας μεταβλητής. Οι δείκτες αυτοί μας δείχνουν την αριθμητική θέση ή το αριθμητικό σημείο γύρω από το οποίο συγκεντρώνονται οι τιμές μιας μεταβλητής. Οι δείκτες κεντρικής τάσης περιλαμβάνουν:

α) Τον αριθμητικό Μέσο Όρο ή Μέση Τιμή (Mean, \bar{x}).

Αυτός είναι ο πιο συχνά χρησιμοποιούμενος δείκτης κεντρικής τάσης και ορίζεται ως το άθροισμα των τιμών μιας μεταβλητής διά του πλήθους των τιμών αυτής (εξαιρουμένων των missing values):

$$\bar{x} = \Sigma x / N$$

Σx = το άθροισμα των τιμών μιας μεταβλητής.

N = το πλήθος των τιμών μιας μεταβλητής (εκτός των missing values).

β) Τη Διάμεσο (Median)

Διάμεσος (Δμ) μιας μεταβλητής είναι η τιμή εκείνη που χωρίζει τα δεδομένα σε δύο ίσα μέρη, έτσι ώστε το πλήθος των μετρήσεων που βρίσκονται αριστερά της να είναι ίσο με το πλήθος των μετρήσεων που βρίσκονται δεξιά της.

Για να βρούμε τη διάμεσο χρειάζεται να διατάξουμε τις τιμές μιας μεταβλητής σε αύξουσα σειρά (από τη μικρότερη στη μεγαλύτερη) και στη συνέχεια, να βρούμε την μεσαία τιμή που βρίσκεται ακριβώς στο κέντρο αυτής της διάταξης χωρίζοντας την σε δύο ίσα μέρη (όταν το πλήθος των τιμών N είναι περιττός αριθμός) ή να βρούμε το ημίάθροισμα των δύο μεσαίων τιμών σε περίπτωση που το N είναι ζυγός αριθμός.

γ) Τη Δεσπόζουσα (επικρατούσα) τιμή μιας μεταβλητής (Mode)

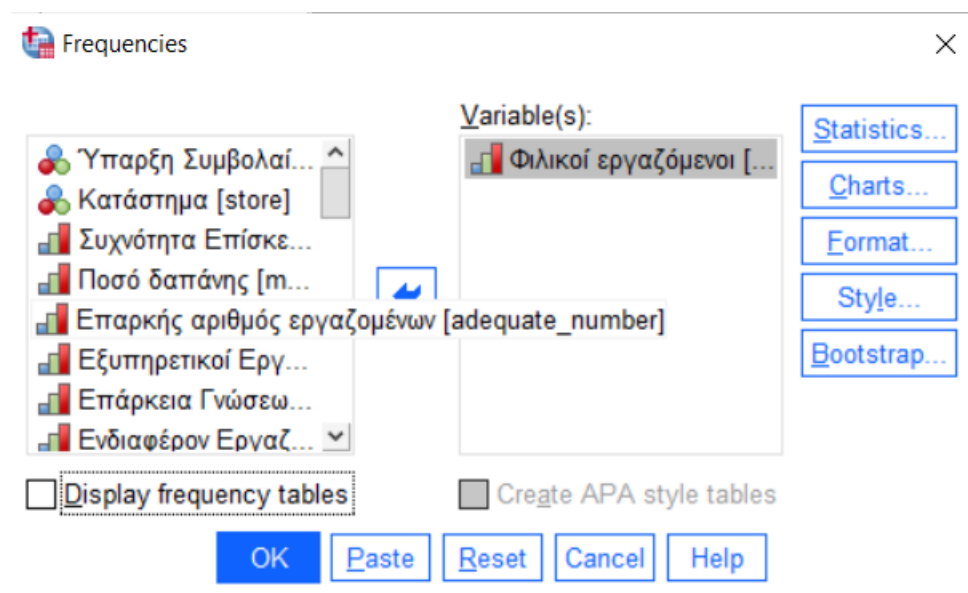
Δεσπόζουσα (Δσπ) είναι η τιμή της μεταβλητής με τη μεγαλύτερη συχνότητα, δηλαδή η τιμή που εμφανίζεται περισσότερες φορές στο δείγμα για τη συγκεκριμένη μεταβλητή. Μπορεί σε μία μεταβλητή να υπάρχουν δύο δεσπόζουσες τιμές με την ίδια μέγιστη συχνότητα (δικόρυφη κατανομή μεταβλητής) ή ακόμα και περισσότερες από δύο δεσπόζουσες τιμές με την ίδια μέγιστη συχνότητα (πολυκόρυφη κατανομή μεταβλητής). **Η Δεσπόζουσα είναι ο μόνος δείκτης κεντρικής τάσης που μπορεί να εφαρμοστεί και σε ποιοτικές μεταβλητές τύπου nominal ή ordinal.**

Όλοι οι παραπάνω δείκτες κεντρικής τάσης μπορούν να υπολογιστούν μέσω της διαδικασίας Frequencies, η οποία μας δίνει τη δυνατότητα να σχηματίσουμε και το ιστόγραμμα για τη μεταβλητή που εξετάζουμε. Για το σκοπό αυτό ακολουθούμε τα εξής βήματα:

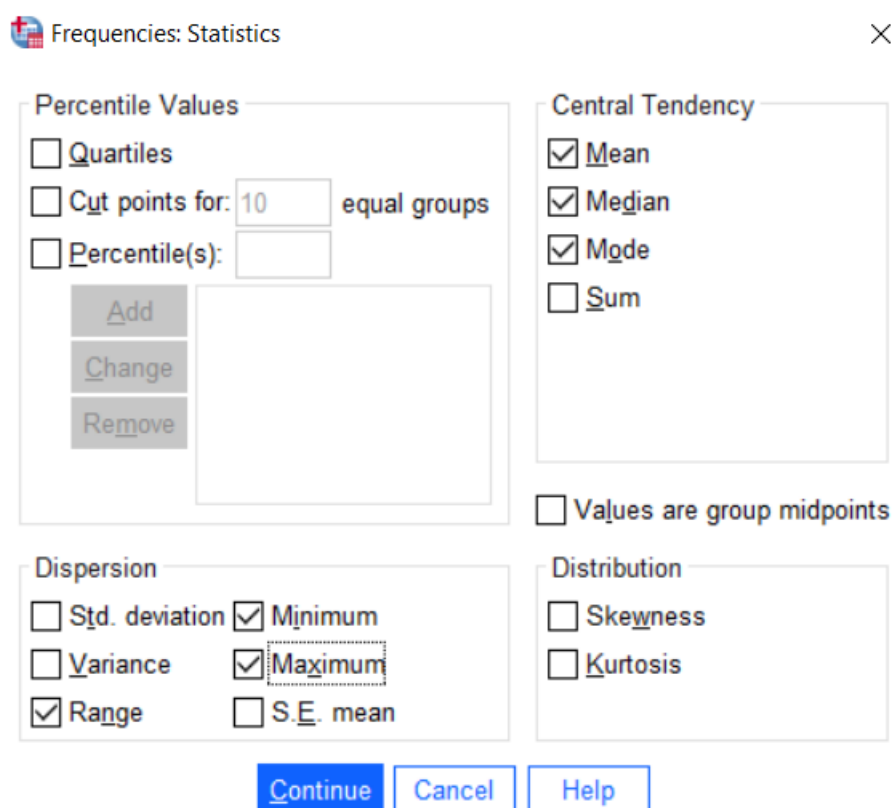
Analyze → Descriptive Statistics → Frequencies

ΠΑΡΑΔΕΙΓΜΑ 1: Ας υποθέσω ότι θέλω να δω κατά πόσο οι εργαζόμενοι στα διάφορα καταστήματα φάνηκαν φιλικόι στους συμμετέχοντες της έρευνας. Επιλέγω Analyze → Descriptive Statistics → Frequencies. Στο παράθυρο διαλόγου που ανοίγει μεταφέρω

τη μεταβλητή 'Φιλικό εργαζόμενοι' στο πλαίσιο Variable(s) και απενεργοποιώ το Display frequency tables.



Έπειτα πατώντας στο κουμπί Statistics στο παράθυρο διαλόγου που ανοίγει επιλέγω τα εξής: Mean (μέση τιμή), Median (διάμεσος), Mode (δεσπόζουσα τιμή), Range (εύρος), Minimum (ελάχιστο), Maximum (μέγιστο). Έπειτα Continue.



Στο output εμφανίζεται ο παρακάτω πίνακας με τους αριθμητικούς περιγραφικούς δείκτες που ζητήσαμε.

Statistics

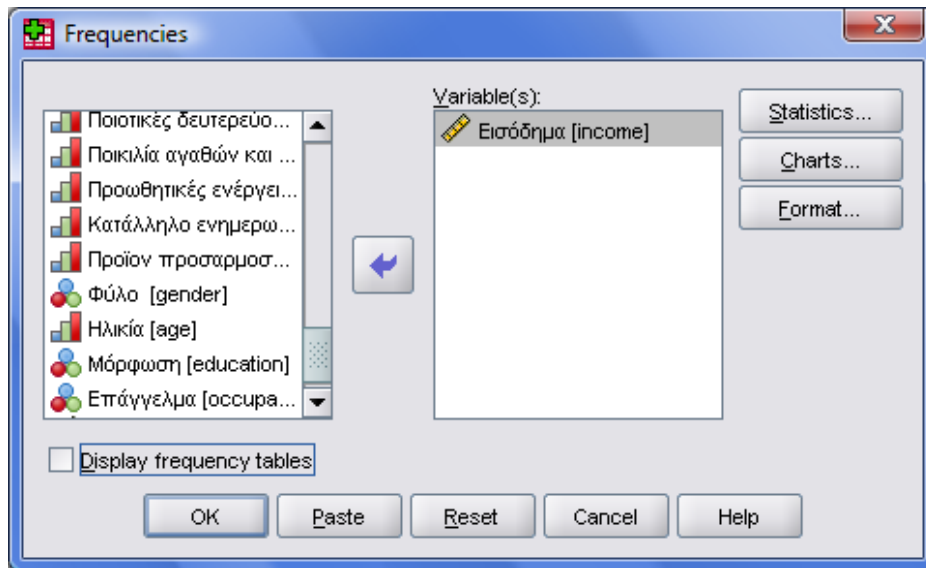
Φιλικοί εργαζόμενοι		
N	Valid	200
	Missing	0
Mean		4,4750
Median		5,0000
Mode		5,00
Range		3,00
Minimum		2,00
Maximum		5,00

Παρατηρούμε ότι ο μέσος όρος των απαντήσεων στη ερώτηση αναφορικά με τη φιλικότητα των εργαζομένων (βαθμός συμφωνίας/διαφωνίας) είναι 4,47 σε πεντάβαθμη κλίμακα που κυμαίνεται από 1=Διαφωνώ απόλυτα έως 5=Συμφωνώ απόλυτα. Δηλαδή ο μέσος όρος πλησιάζει (βρίσκεται κοντά) στο δεξί άκρο της κλίμακας απαντήσεων που δηλώνει πως ο συμμετέχων 'συμφωνεί απόλυτα'. **Επομένως, οι συμμετέχοντες στην έρευνα τείνουν να συμφωνούν ότι οι εργαζόμενοι στα καταστήματα τηλεπικοινωνίας είναι φιλικοί.** Η Διάμεσος των απαντήσεων είναι ίση με το 5, ενώ Δεσπόζουσα τιμή είναι και πάλι το 5. Άρα συμπεραίνουμε ότι οι περισσότεροι συμμετέχοντες απάντησαν ότι 'Συμφωνούν απόλυτα' με τη φιλικότητα των εργαζομένων. Επιπλέον, παρατηρούμε ότι η μέγιστη τιμή που καταγράφηκε είναι το 5 και η ελάχιστη το 2. Τέλος, το εύρος των τιμών μεταξύ ελάχιστης και μέγιστης τιμής είναι 3.

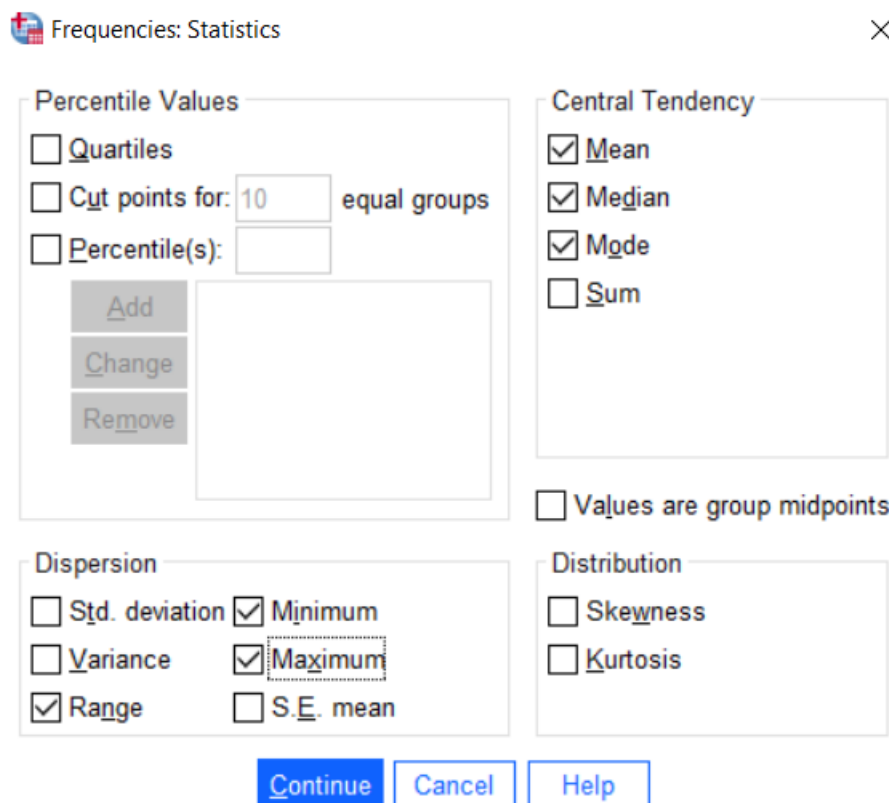
ΣΗΜΕΙΩΣΗ: Η ελάχιστη και η μέγιστη τιμή (minimum-maximum value), καθώς και το εύρος τιμών (range), ανήκουν στους δείκτες διασποράς μιας μεταβλητής και όχι στους δείκτες κεντρικής τάσης. Εδώ παρουσιάζονται μόνο ενδεικτικά, καθώς θα παρουσιαστούν εκ νέου στο 3^ο Εργαστήριο.

ΠΑΡΑΔΕΙΓΜΑ 2: Ας υποθέσουμε τώρα ότι θέλω να περιγράψω το μηνιαίο εισόδημα των συμμετεχόντων στην έρευνα. Το εισόδημα είναι μια συνεχής ποσοτική μεταβλητή, ωστόσο μπορούμε να εφαρμόσουμε σε αυτήν τους ίδιους δείκτες κεντρικής τάσης που χρησιμοποιούνται και για την περιγραφή διακριτών ποσοτικών μεταβλητών.

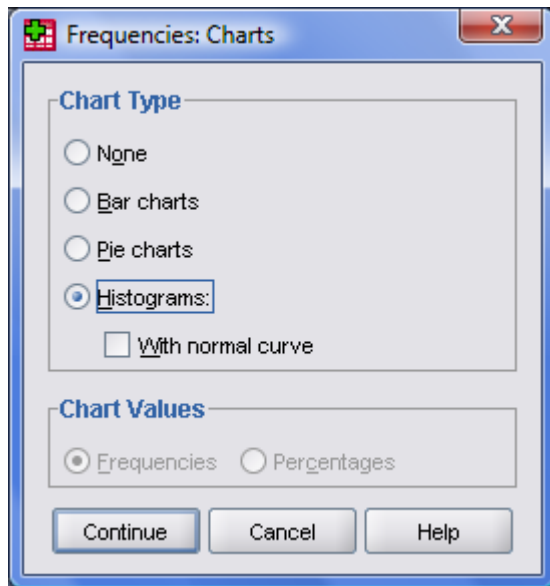
Επιλέγω Analyze → Descriptive Statistics → Frequencies. Στο παράθυρο διαλόγου που ανοίγει μεταφέρω τη μεταβλητή income στο πλαίσιο Variable(s) και απενεργοποιώ το Display frequency tables.



Έπειτα πατώντας στο κουμπί Statistics στο παράθυρο διαλόγου που ανοίγει επιλέγω τα εξής: Mean (μέση τιμή), Median (διάμεσος), Mode (δεσπόζουσα τιμή), Range (εύρος), Minimum (ελάχιστο), Maximum (μέγιστο). Έπειτα Continue.



Στη συνέχεια πατώντας στο κουμπί Charts, στο παράθυρο διαλόγου που ανοίγει επιλέγω Histograms (ιστόγραμμα). Έπειτα Continue και OK.



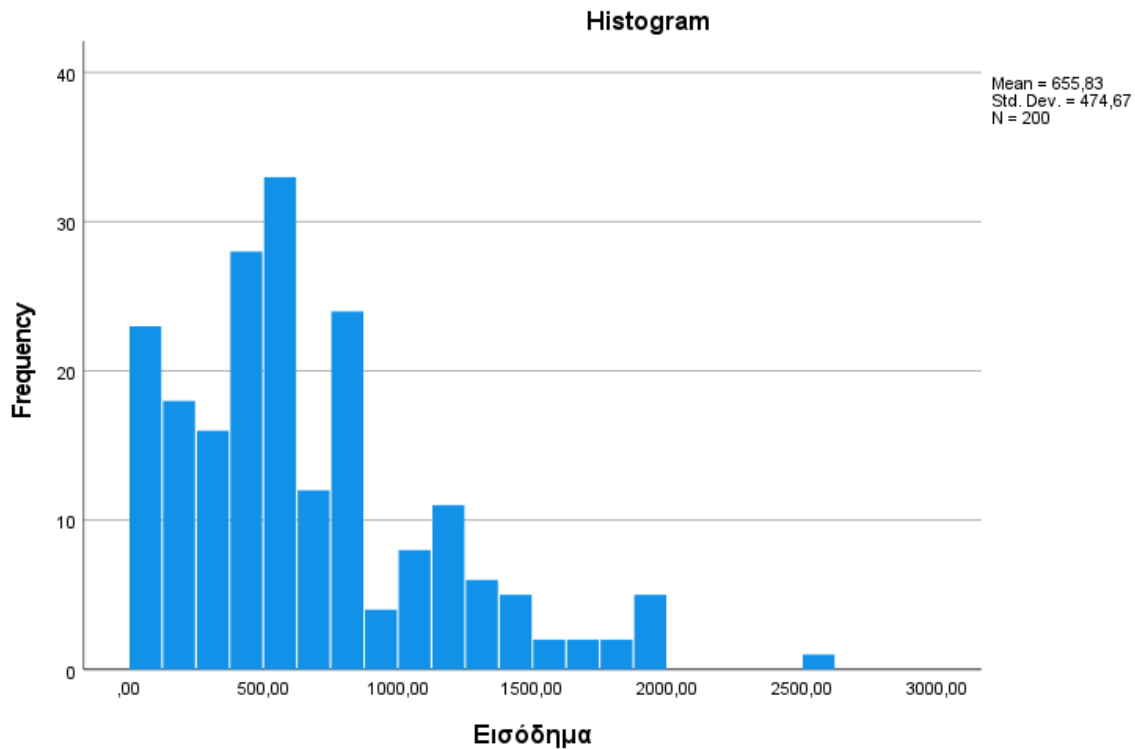
Στο output εμφανίζεται ο παρακάτω πίνακας με τους αριθμητικούς περιγραφικούς δείκτες που ζητήσαμε, καθώς επίσης και το ιστόγραμμα.

Statistics

Εισόδημα

N	Valid	200
	Missing	0
Mean		655,8250
Median		550,0000
Mode		77,00 ^a
Range		2500,00
Minimum		,00
Maximum		2500,00

a. Multiple modes exist. The smallest value is shown



Παρατηρούμε ότι το μέσο εισόδημα των συμμετεχόντων στην έρευνα είναι 655,82 ευρώ το μήνα. Η Διάμεσος του μηνιαίου εισοδήματος ισούται με 550 ευρώ, δηλαδή οι μισοί συμμετέχοντες παίρνουν κάτω από 550 ευρώ το μήνα και οι άλλοι μισοί πάνω από 550 ευρώ. Η δεσπόμενη τιμή που δίνεται είναι τα 77 ευρώ το μήνα, ωστόσο η υποσημείωση εξηγεί ότι υπάρχουν και άλλες δεσπόμενες τιμές με τη συγκεκριμένη να είναι η μικρότερη. Επιπλέον, βλέπουμε ότι το μέγιστο εισόδημα που καταγράφηκε είναι ίσο με 2500 ευρώ το μήνα, ενώ το ελάχιστο ισούται με 0 ευρώ (δηλ. απουσία εισοδήματος). Τέλος το εύρος τιμών (range) από το ελάχιστο έως το μέγιστο εισόδημα είναι 2500 ευρώ.

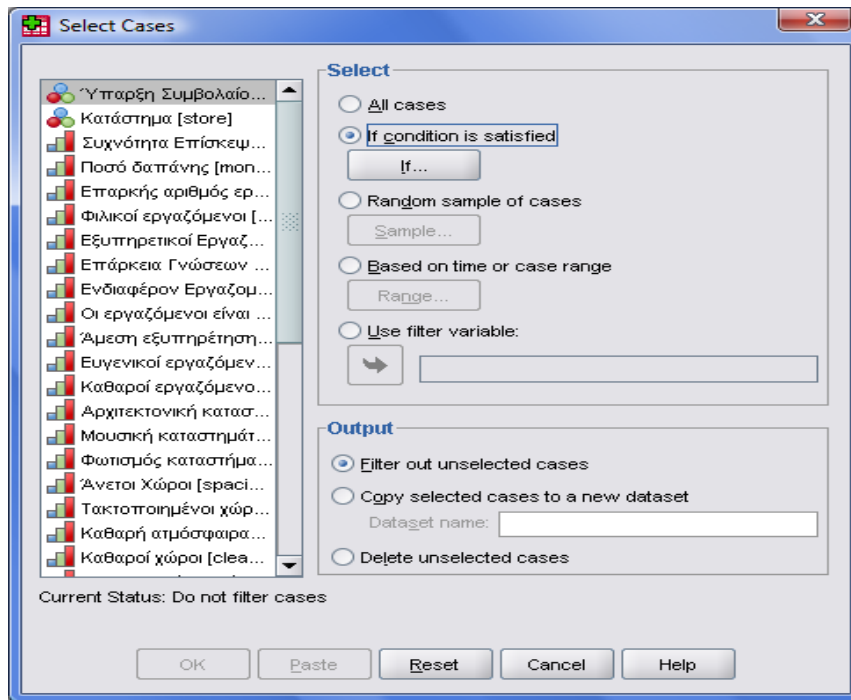
3. Η διαδικασία επιλογής περιπτώσεων (select cases).

Σε ορισμένες περιπτώσεις ένας ερευνητής μπορεί να θέλει να εξετάσει ένα μέρος των δεδομένων που έχει στη διάθεση του. Για παράδειγμα, χρησιμοποιώντας τα δεδομένα του αρχείου `Store_spss_lab.sav` ας υποθέσουμε ότι θέλουμε να εξετάσουμε πώς συμπεριφέρονται οι άνδρες του δείγματός μας.

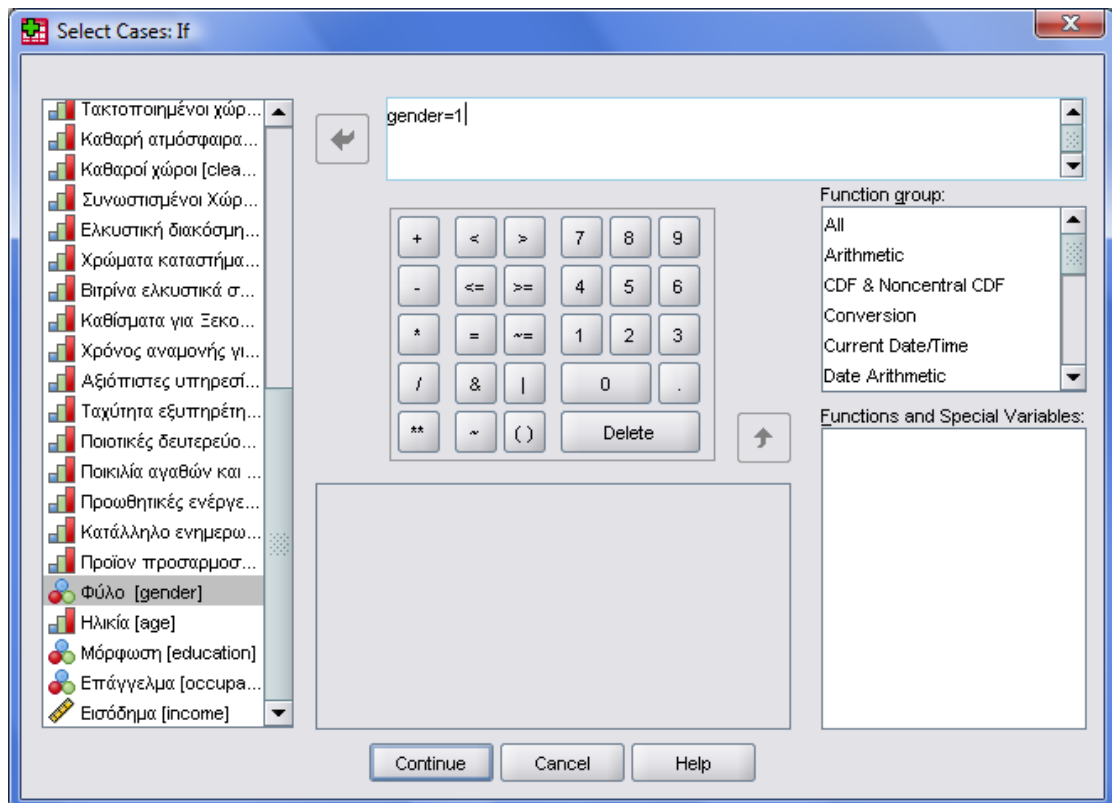
Η διαδικασία της επιλογής περιπτώσεων στο SPSS γίνεται ως εξής:

- Data
 - Select Cases

- Επιλέγουμε `if condition is satisfied` και κάνουμε κλικ στο `if`.



- Επιλέγουμε την μεταβλητή Φύλο, την εισάγουμε στο πλαίσιο δεξιά και σημειώνουμε $gender=1$ (η τιμή 1 αντιστοιχεί στους άνδρες).



Πατάμε continue και μετά OK.

	contract	store	frequency_visits	money_spended	adequate_number	friendly_employees	service_employees	knowledge_employees	Interest_employees	Busy_employees	service_2_employees	polite_employees	clean_employees
1	1,00	2,00	1,00	2,00	2,00	4,00	4,00	3,00	3,00	3,00	3,00	4,00	
2	1,00	2,00	1,00	2,00	4,00	3,00	3,00	3,00	3,00	3,00	3,00	3,00	
3	1,00	2,00	1,00	2,00	4,00	4,00	4,00	4,00	3,00	3,00	3,00	4,00	
4	1,00	1,00	1,00	2,00	5,00	5,00	5,00	4,00	4,00	2,00	4,00	5,00	
5	1,00	1,00	1,00	2,00	5,00	5,00	5,00	4,00	4,00	3,00	4,00	5,00	
6	1,00	1,00	2,00	3,00	5,00	5,00	5,00	5,00	5,00	1,00	5,00	5,00	
7	1,00	1,00	1,00	2,00	4,00	4,00	4,00	3,00	4,00	3,00	4,00	4,00	
8	1,00	2,00	1,00	1,00	4,00	4,00	4,00	3,00	3,00	3,00	3,00	4,00	
9	1,00	2,00	1,00	2,00	4,00	4,00	4,00	4,00	4,00	1,00	3,00	4,00	
10	1,00	2,00	1,00	2,00	5,00	5,00	5,00	4,00	4,00	4,00	4,00	5,00	
11	1,00	2,00	1,00	2,00	3,00	3,00	3,00	3,00	3,00	2,00	3,00	4,00	
12	1,00	2,00	1,00	3,00	4,00	4,00	4,00	4,00	4,00	3,00	4,00	5,00	
13	1,00	2,00	1,00	2,00	5,00	5,00	5,00	5,00	5,00	3,00	5,00	5,00	
14	1,00	2,00	1,00	2,00	5,00	4,00	5,00	4,00	4,00	1,00	4,00	5,00	
15	1,00	2,00	1,00	1,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	
16	1,00	2,00	1,00	2,00	5,00	5,00	5,00	4,00	4,00	2,00	4,00	4,00	
17	1,00	2,00	2,00	3,00	4,00	4,00	4,00	4,00	4,00	3,00	3,00	4,00	
18	1,00	2,00	1,00	1,00	2,00	4,00	4,00	3,00	3,00	3,00	3,00	4,00	
19	1,00	2,00	3,00	3,00	4,00	4,00	4,00	3,00	4,00	2,00	4,00	4,00	
20	1,00	2,00	2,00	2,00	5,00	5,00	5,00	4,00	4,00	1,00	4,00	5,00	
21	1,00	2,00	1,00	1,00	4,00	4,00	4,00	4,00	4,00	1,00	3,00	4,00	
22	1,00	2,00	1,00	4,00	4,00	4,00	4,00	4,00	4,00	1,00	4,00	5,00	
23	1,00	2,00	1,00	1,00	4,00	3,00	4,00	3,00	4,00	3,00	3,00	4,00	
24	1,00	2,00	1,00	1,00	5,00	5,00	4,00	4,00	3,00	1,00	3,00	3,00	
25	1,00	2,00	1,00	1,00	4,00	3,00	3,00	2,00	3,00	2,00	3,00	3,00	
26	1,00	2,00	1,00	1,00	5,00	3,00	4,00	3,00	3,00	2,00	3,00	3,00	

Η όποια στατιστική ανάλυση γίνει στη συνέχεια αναφέρεται μόνο στους άνδρες του δείγματος.

Η επιλογή περιπτώσεων διατηρείται μέχρι εμείς πάλι να την απομακρύνουμε με τον ίδιο τρόπο, μόνο που τώρα επιλέγουμε All cases.

4. Προσθήκη και διαγραφή περιπτώσεων ή μεταβλητών σε υπάρχουσα βάση δεδομένων.

Για να εισάγουμε μια νέα περίπτωση (γραμμή) μεταξύ περιπτώσεων που ήδη υπάρχουν, στο Φύλλο Εργασίας «Data View» επιλέγουμε ένα οποιοδήποτε κελί της γραμμής που βρίσκεται κάτω από τη θέση όπου θέλουμε να εισάγουμε τη νέα γραμμή και από το Edit επιλέγουμε Insert Case. Εναλλακτικά, κάνουμε κλικ στον αριθμό της γραμμής που βρίσκεται κάτω από τη θέση όπου θα εισάγουμε τη νέα γραμμή και με δεξί κλικ επιλέγουμε Insert Case.

Για να εισάγουμε νέα μεταβλητή (στήλη) μεταξύ μεταβλητών που ήδη υπάρχουν, κάνουμε κλικ σε ένα οποιοδήποτε κελί της στήλης που βρίσκεται δεξιά από τη θέση που θέλουμε να εισάγουμε τη νέα στήλη και από το Edit επιλέγουμε Insert Variable. Εναλλακτικά, επιλέγουμε τη στήλη που βρίσκεται δεξιά από τη θέση όπου θα εισάγουμε τη νέα στήλη και με δεξί κλικ επιλέγουμε Insert Variable.

Νέες μεταβλητές μπορούμε να εισάγουμε και από το Φύλλο Εργασίας «Variable View» με παρόμοιο τρόπο.

Αν η εισαγωγή της γραμμής ή στήλης έχει γίνει σε λάθος θέση, αναιρούμε την ενέργεια με Edit → Undo.

Για να διαγράψουμε μια περίπτωση ή μια μεταβλητή, επιλέγουμε τη γραμμή ή τη στήλη στην οποία αντιστοιχεί και με δεξί κλικ επιλέγουμε Clear.