

ΤΟΠΟΣ  
Επιστημονικές Εκδόσεις

Επιμέλεια-Διόρθωση: Μαρία Αποστολοπούλου  
Ηλ. επεξεργασία: Μαρία Παπαδάκη  
Εξώφυλλο: ΜΟΤΙΒΟ Α.Ε.

© 2010 Εκδόσεις Τόπος και Αθανάσιος Κατσής,  
Γεώργιος Σιδερίδης, Αναστάσιος Εμβάλωτης  
[Οι Εκδόσεις Τόπος είναι εμπορικό σήμα της ΜΟΤΙΒΟ Α.Ε.]

ISBN 978-960-6863-80-6

Παραγωγή - κεντρική διάθεση:  
Πλαπούτα 2 και Καλλιδρομίου, 11473, Αθήνα  
Τηλ.: 210 8222835-856  
Fax: 210 8222684

Βιβλιοπωλείο  
Γενναδίου 6  
10678, Αθήνα  
Τηλ.: 210 3221580  
Fax: 210 3211246

Η πνευματική ιδιοκτησία αποκτάται χωρίς καμία διατύπωση και χωρίς την ανάγκη ρήτρας απαγορευτικής των προσβολών της. Κατά το Ν. 2387/20 (όπως έχει τροποποιηθεί με το Ν. 2121/93 και ισχύει σήμερα) και κατά τη Διεθνή Σύμβαση της Βέρνης (που έχει κυρωθεί με το Ν. 100/1975) απαγορεύεται η αναδημοσίευση, η αποθήκευση σε κάποιο σύστημα διάσωσης και γενικά η αναπαραγωγή του παρόντος έργου με οποιονδήποτε τρόπο ή μορφή, τμηματικά ή περιληπτικά, στο πρωτότυπο ή σε μετάφραση ή άλλη διασκευή, χωρίς γραπτή άδεια του εκδότη.

[www.toposbooks.gr](http://www.toposbooks.gr)

## Πρόλογος

Το βιβλίο αυτό φιλοδοξεί να παρουσιάσει τις κυριότερες τεχνικές της στατιστικής ανάλυσης στο χώρο των κοινωνικών επιστημών και των επιστημών της Αγωγής, αφού η Στατιστική έχει καταστεί πλέον ένα αναπόσπαστο κομμάτι της ποσοτικής ανάλυσης στα συγκεκριμένα ερευνητικά πεδία. Απευθύνεται σε όσους ξεκινούν την ενασχόλησή τους με το χώρο της στατιστικής ανάλυσης των δεδομένων, αλλά και στους πιο έμπειρους ερευνητές οι οποίοι επιθυμούν να ενημερωθούν για εξελιγμένες μεθοδολογίες στο χώρο της εφαρμοσμένης Στατιστικής.

Το κυρίαρχο πνεύμα κατά τη διάρκεια της συγγραφής του βιβλίου ήταν η παρουσίαση των διαφόρων στατιστικών τεχνικών χωρίς τη χρήση «δύσκολων» μαθηματικών εννοιών, οι οποίες πολλές φορές λειτουργούν αποτρεπτικά για το ενδιαφέρον του αναγνώστη. Για το λόγο αυτό επικεντρωθήκαμε στη «χρηστική» διάσταση της Στατιστικής, με εφαρμογές κυρίως (αλλά όχι μόνο) μέσω του λογισμικού Statistical Package for the Social Sciences (SPSS) και της μετεξέλιξής του σε Predictive Analysis SoftWare (PASW). Από την άλλη πλευρά όμως, δεν επιθυμούσαμε το βιβλίο να θεωρηθεί αποκλειστικά ως μια παράθεση «έτοιμων συνταγών» για τη Στατιστική. Έτσι, για το μνημένο μαθηματικά αναγνώστη και για τη σε βάθος κατανόηση των εννοιών υπάρχουν πολλές αναφορές στη μαθηματική μεθοδολογία, οι οποίες όμως δεν είναι απαραίτητες για την πρακτική κατανόηση των στατιστικών τεχνικών από όσους δεν έχουν το κατάλληλο υπόβαθρο.

Το βιβλίο μπορεί να «χωρισθεί» σε δύο μέρη. Το πρώτο μέρος αποτελείται από τα Κεφάλαια 1-9 και καλύπτει τις πιο γνωστές στατιστικές προσεγγίσεις στο χώρο των κοινωνικών επιστημών. Στο μεγαλύτερο μέρος αυτών των κεφαλαίων δεν απαιτείται ισχυρό μαθηματικό υπόβαθρο και η ύλη είναι ανάλογη ενός προπτυχιακού προγράμματος σπουδών ή είναι επαρκής για μια αρχική προσέγγιση σε μεταπτυχιακό επίπεδο. Προτείνουμε στον αναγνώστη να μην επιμείνει στην (πλήρη) κατανόηση ενός μαθη-

ματικού τύπου, αν αυτό δημιουργεί πρόβλημα, αλλά στην ποιοτική διάσταση της μεθόδου και των αποτελεσμάτων της. Από την άλλη πλευρά όμως, αυτή η προσπάθεια «απο-μαθηματικοποίησης» δε σημαίνει ότι δεν απαιτείται σοβαρή και συστηματική προσπάθεια για την κατανόηση του υλικού.

Στο δεύτερο μέρος (Κεφάλαια 10-14) παρουσιάζονται οι πιο σύγχρονες στατιστικές τεχνικές. Εδώ είναι απαραίτητη η καλή γνώση των προηγούμενων κεφαλαίων, ενώ υπάρχει πιο συχνή αναφορά στις μαθηματικές ιδιότητες των στατιστικών τεχνικών. Τα κεφάλαια αυτά είναι πιο χρήσιμα για ερευνητές, κυρίως σε διδακτορικό επίπεδο, με καλό μαθηματικό επίπεδο, ενώ γίνεται χρήση και άλλων στατιστικών λογισμικών. Στο τέλος του βιβλίου υπάρχουν παραρτήματα με αναφορές και σε ηλεκτρονικές σελίδες, για καλύτερη επεξήγηση ορισμένων σημείων του κειμένου, και μια εκτενής βιβλιογραφία για πληρέστερη μελέτη.

Είναι προφανές ότι υπάρχουν αρκετές ακόμα στατιστικές τεχνικές οι οποίες δεν περιλαμβάνονται στο βιβλίο. Επιλέξαμε τις, κατά τη γνώμη μας, πιο «χρήσιμες» στους ερευνητές του χώρου των κοινωνικών επιστημών, ενώ είμαστε πάντα ανοικτοί σε υποδείξεις και σχόλια για τυχόν παραλείψεις.

Κλείνοντας αυτό το εισαγωγικό σημείωμα επιθυμούμε να εκφράσουμε τις ευχαριστίες μας στις οικογένειές μας για την υπομονή και την υποστήριξή τους κατά τη διάρκεια της συγγραφής. Το βιβλίο αφιερώνεται σε αυτές!

## Εισαγωγή

### Η έννοια της μεταβλητής

Η έρευνα στις κοινωνικές επιστήμες βασίζεται στη μελέτη των σχετικών εμπειρικών δεδομένων. Στην ποσοτική ανάλυση, το σύνολο των δεδομένων αποτελείται από τις επιμέρους «μεταβλητές» (variables). Γενικώς, μεταβλητή είναι κάθε χαρακτηριστικό ή ιδιότητα το οποίο επιδέχεται περισσότερες από μια τιμές (Argyrous, 2005). Ως σχετικό παράδειγμα, ας θεωρήσουμε μια έρευνα για τη σχολική επίδοση όπου ο ερευνητής<sup>1</sup> συλλέγει δεδομένα. Τα δεδομένα αυτά αποτελούνται από τις βαθμολογίες των μαθητών σε ορισμένα μαθήματα. Η βαθμολογία κάθε μαθήματος είναι μια μεταβλητή που χρησιμοποιείται συχνά στην εκπαιδευτική έρευνα. Δέχεται περισσότερες από μια τιμές, οι οποίες μάλιστα δε χρειάζεται να είναι κατ' ανάγκη αριθμητικές. Άλλες μεταβλητές μπορεί να αναφέρονται σε κάποια δημογραφική ιδιότητα των ερευνητικών υποκειμένων, όπως στην περίπτωση του φύλου (αγόρι, κορίτσι). Η επιλογή των μεταβλητών που εξετάζονται σε κάθε ερευνητικό πρόβλημα προκύπτει μέσα από μια σειρά μεθοδολογικών κριτηρίων, όπως το θεωρητικό πλαίσιο της έρευνας, τα ερευνητικά ερωτήματα, η διαθεσιμότητα των δεδομένων, ζητήματα δεοντολογίας κ.λπ.

Στις κοινωνικές επιστήμες έχουν προταθεί και καθιερωθεί διάφορες ταξινομήσεις αναφορικά με τους τύπους των μεταβλητών. Η συνηθέστερη ταξινόμηση (τουλάχιστον στην ελληνική βιβλιογραφία) ορίζει τις **ποιοτικές** (qualitative) και τις **ποσοτικές** (quantitative) μεταβλητές. Οι ποσοτικές μεταβλητές εκφράζονται σε αριθμητική μορφή (π.χ., το εισόδημα μετρημένο σε ευρώ), ενώ οι ποιοτικές σε μη αριθμητική μορφή (π.χ., το φύλο). Ας σημειωθεί ότι τα ποιοτικά δεδομένα (δηλαδή το σύνολο των ποιοτικών μεταβλητών) δεν είναι αναγκαστικά λέξεις ή κείμενο. Ως ποιοτικά δεδομένα μπορεί, επίσης, να θεωρηθούν εικόνες, καταγραφές ήχων κ.λπ.

Η διάκριση μεταξύ ποσοτικών και ποιοτικών δεδομένων έχει αποτελέσει αντικείμενο πολλών συζητήσεων σχετικά με το ποια ερευνητικά δεδομένα θεωρούνται «καλύτερα». Οι υπέρμαχοι της ποιοτικής προσέγγισης ισχυρίζονται ότι τα ποιοτικά δε-

1. Σε όλο το βιβλίο όπου αναφέρεται το αρσενικό (ερευνητής, φοιτητής, μαθητής, μαθητές) θα εννοείται και η αντίστοιχη χρήση του θηλυκού (π.χ., ερευνήτρια, φοιτήτρια, μαθήτρια, μαθήτριες), καθώς και το αντίστροφο, εκτός και αν υπάρχει σαφής διαφοροποίηση.



δομένα είναι πιο «λεπτομερή» και συνήθως αποκαλύπτουν πολλές «κρυφές» πλευρές του ερευνητικού προβλήματος. Από την άλλη πλευρά, οι της ποσοτικής ανάλυσης θεωρούν ότι τα δεδομένα τους είναι «αντικειμενικά» και «αξιόπιστα». Για μια καλή και ευρύτερη ανασκόπηση του σχετικού θέματος, ο ενδιαφερόμενος αναγνώστης μπορεί να διαβάσει την 1η ενότητα του Εγχειριδίου Μελέτης της Θεματικής Ενότητας: «Εκπαιδευτική Έρευνα στην Πράξη» του Μεταπτυχιακού Προγράμματος Σπουδών: Σπουδές στην Εκπαίδευση του Ελληνικού Ανοικτού Πανεπιστημίου (Bird, Hammersley, Gomm, Woods, 1999).

Κατά την προσωπική μας άποψη, η παραπάνω διαμάχη είναι υπερβολική, αντιπαραγωγική και εν πολλοίς λανθασμένη, αφού αποκρύπτει το γεγονός ότι **τα ποιοτικά και τα ποσοτικά δεδομένα συνδέονται μεταξύ τους**. Πιο συγκεκριμένα, πολλές ποσοτικές μεταβλητές κατασκευάζονται με τη βοήθεια ποιοτικών κρίσεων και αναλύσεων. Προς επίρρωση αυτού του ισχυρισμού, ας επικεντρωθούμε στη μεταβλητή της αυτοεκτίμησης (self esteem), η οποία χρησιμοποιείται σε αρκετές ψυχολογικές μελέτες. Η συγκεκριμένη μεταβλητή μετρείται ποσοτικά, βάσει των αριθμητικών απαντήσεων σε μια σειρά ερωτήσεων. Η κατασκευή όμως αυτών των ερωτήσεων βασίζεται σε πολλές ποιοτικές αναλύσεις που έχουν να κάνουν, μεταξύ άλλων, με τον ορισμό της αυτοεκτίμησης, τη διαφοροποίησή της από άλλες παρόμοιες έννοιες, τον τρόπο βαθμολόγησης κάθε απάντησης, ώστε να καταγράφονται οι ποιοτικές διαφορές μεταξύ των ερωτηθέντων κ.λπ.

Αντιστοίχως, οι ποιοτικές μεταβλητές μπορούν να εκφραστούν και να αναλυθούν αριθμητικά. Για παράδειγμα, έστω ότι στη μεταβλητή με την ενδεικτική περιγραφή “πολιτική τοποθέτηση” ανήκουν οι γενικές διατυπώσεις “προοδευτικός”, “κεντρώος” και “συντηρητικός”. Ένας ερευνητής που επιθυμεί την ποσοτική ανάλυση της παραπάνω μεταβλητής μπορεί να θέσει συγκεκριμένες αριθμητικές τιμές σε κάθε απάντηση (π.χ., 1=“προοδευτικός”, 2=“κεντρώος” και 3=“συντηρητικός”, ή αντίστροφα). Για όλους τους παραπάνω λόγους δε θα επιμένουμε ιδιαίτερος στη διάκριση μεταξύ ποιοτικών και ποσοτικών δεδομένων. Απλώς επισημαίνουμε ότι στο βιβλίο αυτό η έμφαση θα δοθεί αποκλειστικά στην ανάλυση και ερμηνεία ποσοτικών δεδομένων.

Μια πολύ ενδιαφέρουσα ταξινόμηση αφορά στις ποσοτικές μεταβλητές. Αυτές διακρίνονται συνήθως σε **συνεχείς** (continuous) και **διακριτές** (discrete) ή **κατηγορικές** (categorical) μεταβλητές.

Η συνεχής (ποσοτική) μεταβλητή επιδέχεται ένα πολύ μεγάλο πλήθος τιμών στο σύνολο όλων των πραγματικών αριθμών. Παράδειγμα συνεχούς ποσοτικής μεταβλητής είναι το βάρος ενός υποκειμένου, το οποίο δέχεται τιμές από το εύρος τιμών μέτρησης βάρους. Έτσι, οι τιμές μπορεί να είναι 67.220 (Kgr), 67.221 (Kgr), 67.222 (Kgr) κ.λπ. Για τους μαθηματικά μνημόνους αναγνώστες ο όρος «ένα πολύ μεγάλο πλήθος τιμών» μπορεί να αντικατασταθεί με τον όρο «άπειρες τιμές».

Οι διακριτές (ποσοτικές) μεταβλητές επιδέχονται επιλεγμένες τιμές, οι οποίες είναι συνήθως ακέραιοι αριθμοί. Παράδειγμα διακριτής μεταβλητής είναι η σύνθεση της οικογένειας, όταν η αποτύπωση αναφέρεται στον αριθμό των προσώπων που την αποτελούν. Έτσι, τιμές της παραπάνω μεταβλητής μπορεί να είναι 2 άτομα, 3 άτομα, 4 άτομα, 5 άτομα, αλλά ποτέ 2,5 άτομα.

Ενδιαφέρον παρουσιάζει η μετάβαση από τον έναν τύπο ποσοτικής μεταβλητής στον άλλο. Με βάση μια συνεχή μεταβλητή μπορούμε να δημιουργήσουμε την αντίστοιχη διακριτή μεταβλητή. Για παράδειγμα, στην περίπτωση της συνεχούς μεταβλητής του ετήσιου εισοδήματος (το οποίο μετριέται σε ευρώ), μπορούμε να θέσουμε όρια που διαχωρίζουν το εισόδημα σε υψηλό, μεσαίο και χαμηλό. Εντελώς ενδεικτικά ορίζουμε τη διακριτή μεταβλητή του εισοδήματος με τιμές 1=μέχρι 15 χιλιάδες ευρώ (χαμηλά εισοδήματα), 2=15-25 χιλιάδες ευρώ (μεσαία εισοδήματα) και 3=άνω των 25 χιλιάδων ευρώ (υψηλά εισοδήματα). **Προσοχή όμως: Η αντίστροφη μετάβαση είναι αδύνατη!** Μια διακριτή μεταβλητή δεν μπορεί να μετασχηματισθεί σε συνεχής. Για το λόγο αυτό, όταν υπάρχει η δυνατότητα επιλογής, συνιστάται η αρχική καταγραφή των δεδομένων να γίνεται σε συνεχείς μεταβλητές.

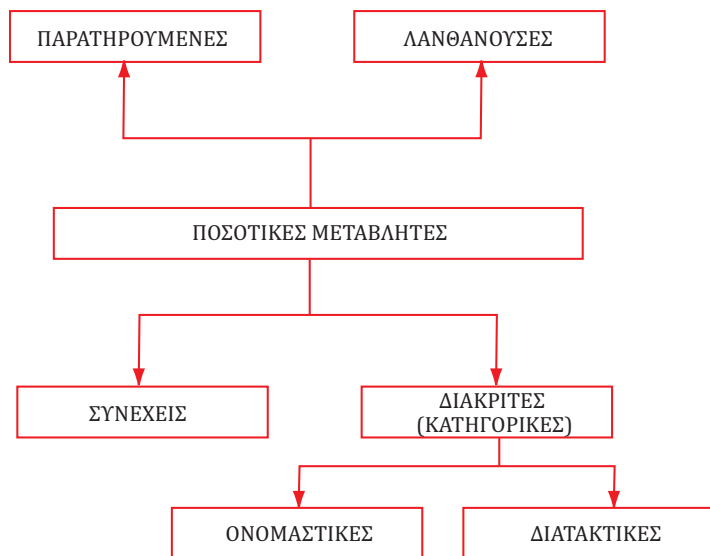
Οι διακριτές μεταβλητές παρουσιάζουν μια επιπλέον σημαντική ταξινόμηση. Χωρίζονται σε **ονομαστικές** (nominal) και **διατακτικές** (ordinal). Ως ονομαστικές μεταβλητές θεωρούνται οι διακριτές μεταβλητές στις οποίες εφαρμόζεται απλή ταξινόμηση χαρακτηριστικών. Δεν υπάρχει, δηλαδή, ποιοτική σύγκριση μεταξύ των τιμών της μεταβλητής. Παραδείγματα ονομαστικής μεταβλητές είναι το φύλο και η οικογενειακή κατάσταση ενός υποκειμένου. Στις δύο αυτές περιπτώσεις δεν μπορούμε να πούμε ότι κάποια τιμή της μεταβλητής «υπερτερεί» των άλλων τιμών της μεταβλητής.

Ως διατακτικές μεταβλητές θεωρούμε τις διακριτές μεταβλητές των οποίων οι τιμές τους αναπτύσσονται κατά τέτοιο τρόπο, ώστε να είναι δυνατή η ανίχνευση κάποιας σχέσης διάταξης. Για παράδειγμα, η διακριτή μεταβλητή του εισοδήματος που δημιουργήσαμε προηγουμένως ή η αξιολόγηση ενός διδακτικού εγχειριδίου (1=άσχημο, 2=μέτριο, 3=καλό) είναι περιπτώσεις διατακτικής μεταβλητής. Στις παραπάνω μεταβλητές υπάρχει συγκριτική σχέση μεταξύ των τιμών των μεταβλητών, αφού, λ.χ., η τιμή 1 υποδηλώνει χειρότερη αξιολόγηση του εγχειριδίου από την τιμή 2. Αξίζει να σημειωθεί ότι, στις διατακτικές μεταβλητές, ίσες διαφορές μεταξύ των τιμών της μεταβλητής δε συνεπάγονται αναγκαστικά και ίσες διαφορές για το χαρακτηριστικό που συνοδεύει τη μεταβλητή. Για παράδειγμα, κάποιος με τιμή 2 στη μεταβλητή του εισοδήματος (μεσαίο εισόδημα) δε σημαίνει αναγκαστικά ότι έχει διπλάσιο εισόδημα από κάποιον με τιμή 1. Σημαίνει απλώς ότι το εισόδημα του πρώτου υποκειμένου είναι υψηλότερο από το αντίστοιχο εισόδημα του δεύτερου υποκειμένου.

Εκτός από τη διάκριση μεταξύ ποσοτικών-ποιοτικών, συνεχών-διακριτών και ονομαστικών-διατακτικών μεταβλητών, μια σημαντική διαφοροποίηση υπάρχει

επίσης μεταξύ **παρατηρούμενων** (observable) και **λανθανουσών** (latent) ή **μη παρατηρούμενων** μεταβλητών. Οι παρατηρούμενες μεταβλητές είναι αυτές που μπορεί εύκολα και άμεσα να μετρηθούν από τον ερευνητή. Πολλές δημογραφικές μεταβλητές αποτελούν παραδείγματα τέτοιων μεταβλητών (π.χ., ύψος, βάρος, φύλο κ.λπ.). Αντίθετα, ως λανθάνουσες μεταβλητές θεωρούνται αυτές των οποίων η μέτρηση δεν είναι εύκολη και άμεση, αλλά συνήθως γίνεται με τη χρήση άλλων μεταβλητών των οποίων η καταγραφή τους γίνεται με ευκολότερο τρόπο. Παραδείγματα λανθανουσών μεταβλητών είναι οι στάσεις των ερωτωμένων για ένα συγκεκριμένο ζήτημα, προτιμήσεις, ψυχολογικά χαρακτηριστικά κ.λπ. Είναι προφανές ότι η μέτρηση των λανθανουσών μεταβλητών δεν είναι κάτι απλό ποσοτικά (βλ. Bartholomew & Knott, 1999). Επίσης, προφανές είναι ότι οι λανθάνουσες μεταβλητές αποτελούν την πιο ενδιαφέρουσα ερευνητικά κατηγορία μεταβλητών.

Το Σχήμα 1 παρουσιάζει με διαγραμματικό τρόπο τις κυριότερες ταξινομήσεις των ποσοτικών μεταβλητών. Εκτός όμως από αυτές υπάρχουν και άλλες κατηγοριοποιήσεις ποσοτικών μεταβλητών, στις σημαντικότερες εκ των οποίων θα αναφερθούμε στη συνέχεια του βιβλίου.



**Σχήμα 1:** Οι κυριότερες κατηγορίες των ποσοτικών μεταβλητών

## Κεφάλαιο 1

### Περιγραφική Στατιστική

#### 1.1. Εισαγωγή

Μετά τη συλλογή των ποσοτικών δεδομένων, ακολουθεί το στάδιο της στατιστικής επεξεργασίας για την εξαγωγή των πρώτων ερευνητικών ευρημάτων. Το πρώτο βήμα της στατιστικής ανάλυσης περιλαμβάνει συνήθως την Περιγραφική Στατιστική (Descriptive Statistics), όπου ο ερευνητής αποκτά μια προκαταρκτική εικόνα-περιγραφή των δεδομένων. Μέσα από αυτή την εισαγωγική διαδικασία είναι δυνατόν να εξαχθούν τα πρώτα ουσιαστικά ευρήματα. Έστω, για παράδειγμα, ότι ένας ερευνητής επιθυμεί να εξετάσει τις επιδόσεις ενός συγκεκριμένου μαθητικού πληθυσμού ως προς το φύλο. Χωρίς να προβεί σε εξειδικευμένες στατιστικές τεχνικές, ένας απλός υπολογισμός των μέσων όρων της επίδοσης μεταξύ των δύο φύλων, σε συνδυασμό με τα κατάλληλα διαγράμματα, παρέχει στον ερευνητή μια αρχική ένδειξη-απάντηση για το συγκεκριμένο ερευνητικό ερώτημα.

Υπάρχουν δύο κύριοι άξονες περιγραφικής στατιστικής ανάλυσης. Ο πρώτος περιλαμβάνει τον υπολογισμό ορισμένων συγκεκριμένων στατιστικών δεικτών (μέσος όρος, διάμεσος κ.λπ.), βάσει των οποίων είναι δυνατή μια αρχική αποτύπωση των περισσότερων ποσοτικών μεταβλητών. Στη στατιστική ορολογία οι συγκεκριμένοι δείκτες ονομάζονται συνήθως «μέτρα περιγραφικής στατιστικής». Εκτός όμως από αυτά τα μέτρα (δείκτες), υπάρχει επίσης η δυνατότητα δημιουργίας κατάλληλων διαγραμμάτων για την οπτική απεικόνιση των μεταβλητών.

Στη συνέχεια θα περιγράψουμε τους βασικούς άξονες περιγραφικής στατιστικής ανάλυσης, έχοντας πάντα υπόψη ότι αποτελούν μια αρχική στατιστική προσέγγιση. Για μια εκτενέστερη μελέτη μεθόδων περιγραφικής αλλά και πιο εξειδικευμένης στατιστικής ανάλυσης στις κοινωνικές επιστήμες παραπέμπουμε, μεταξύ άλλων, στους Γναρδέλλη (2009), Martin (2008), Κασίλλη (2006), Δαφέρμο (2005), Ρούσσο και Τσαούση (2002), Τσάντα, Μωυσιάδη, Μπαγιάτη και Χατζηπαντελή (1999), Μπεχράκη (1999), Ρίτσαρντσον και Βασιλίαινα (1999), Καλαματιανού (1998), Μπόρα-Σέντα και Κολυβά-Μαχαίρα (1998).

#### 1.2. Μέτρα Περιγραφικής Στατιστικής

Τα μέτρα της περιγραφικής στατιστικής αναφέρονται σε τρία κυρίως χαρακτηριστικά των δεδομένων:

1. Τη *θέση του κέντρου* (ή του μεγαλύτερου όγκου) των δεδομένων κάθε μεταβλητής. Αυτά ονομάζονται **μέτρα κεντρικής θέσης** ή **τάσης**.

2. Τη *διασπορά* των δεδομένων κάθε μεταβλητής. Τα συγκεκριμένα μέτρα ονομάζονται **μέτρα διασποράς**.

3. Την *ασυμμετρία* της κατανομής των δεδομένων. Έτσι έχουμε τα **μέτρα ασυμμετρίας**.

Στη συνέχεια θα αναλύσουμε κάθε κατηγορία μέτρων περιγραφικής στατιστικής ξεχωριστά. Επισημαίνουμε ότι θα γίνει μια παρουσίαση των κυριότερων χαρακτηριστικών κάθε μέτρου, χωρίς την παράθεση των μαθηματικών τύπων υπολογισμού τους. Στο τέλος του κεφαλαίου υπάρχει ένα παράδειγμα εύρεσής τους μέσω του SPSS.

Επίσης, σε κάποιες περιπτώσεις έχουμε **ομαδοποιημένα** δεδομένα όπου, αντί για κάθε δειγματική μονάδα, τα δεδομένα παρουσιάζονται με βάση τη συχνότητα εμφάνισης σε συγκεκριμένες κατηγορίες τιμών, π.χ. πόσοι μαθητές στην εικοσαβάθμια κλίμακα έχουν βαθμολογηθεί κάτω από το 10, πόσοι έχουν βαθμό μεταξύ 11-15 κ.λπ. Σε αυτές τις περιπτώσεις ο αναγνώστης πρέπει να συμβουλευθεί τους μαθηματικούς τύπους για την εύρεση των μέτρων περιγραφικής στατιστικής για ομαδοποιημένα δεδομένα. Οι περιπτώσεις αυτές είναι περισσότερο θεωρητικές και δεν εφαρμόζονται συχνά στην πράξη (βλ., ενδεικτικά, Δαμιανού και Κούτρας, 2000).

### 1.2.1. Μέτρα κεντρικής θέσης ή τάσης

Τα μέτρα κεντρικής θέσης υπολογίζουν, με διάφορους τρόπους, την κεντρική θέση κάθε μεταβλητής. Ανάμεσά τους περιλαμβάνεται ο μέσος όρος, που χρησιμοποιείται ευρύτατα στην ποσοτική ανάλυση. Οφείλουμε όμως να τονίσουμε ότι τα μέτρα κεντρικής θέσης αφορούν σχεδόν αποκλειστικά **συνεχείς μεταβλητές** (π.χ., εισόδημα, επίδοση σε εξετάσεις και στις αρχικές μονάδες μέτρησης).

Όταν έχουμε διακριτές ονομαστικές μεταβλητές δεν έχει νόημα η χρήση των μέτρων κεντρικής θέσης. Για παράδειγμα, αν υποθέσουμε ότι η κωδικοποίηση της μεταβλητής φύλο είναι 0 και 1 (για άνδρες και γυναίκες αντίστοιχα), τι νόημα έχει η τιμή του μέσου όρου; Σε ορισμένες έρευνες, τα μέτρα κεντρικής θέσης χρησιμοποιούνται σε διακριτές διατακτικές μεταβλητές. Αυτό καλό είναι να γίνεται με φειδώ και κατά περίπτωση, γιατί η ερμηνεία τους δεν είναι πάντα ιδιαίτερος διαφωτιστική. Για παράδειγμα, ίσως να μην έχει άμεση χρησιμότητα η έννοια του μέσου όρου (π.χ. 2,3) σε μια πενταβάθμια κλίμακα (1-5) για τη μέτρηση του βαθμού επαγγελματικής ικανοποίησης.

Για συνεχείς μεταβλητές, τα κυριότερα μέτρα κεντρικής θέσης είναι τα ακόλουθα:

■ **Μέσος όρος (mean, average, expected value):** Είναι το πιο γνωστό μέτρο θέσης και υπολογίζεται από το λόγο του αθροίσματος όλων των τιμών μίας μεταβλητής προς το πλήθος των τιμών. Το χαρακτηριστικό γνώρισμα του μέσου όρου είναι ότι επηρεάζεται πολύ έντονα από τις ακραίες τιμές (outliers) της μεταβλητής.

Οι ακραίες τιμές μπορεί, σε ορισμένες περιπτώσεις, να είναι είτε πολύ μεγάλες είτε πολύ μικρές. Αξίζει να τονίσουμε ότι ο ερευνητής **δεν** πρέπει να θεωρεί ότι οι ακραίες τιμές έχουν κατ' ανάγκη καταχωρισθεί λάθος και να προχωρεί στη διόρθωση-αντικατάστασή τους. Αν και το ενδεχόμενο λανθασμένης καταχώρισης των δεδομένων είναι υπαρκτό, εντούτοις δεν αποκλείεται κάποιο ερευνητικό υποκείμενο να παρουσιάζει αυτή την ακραία τιμή (π.χ., ένα πολύ πλούσιο άτομο για τη μεταβλητή του εισοδήματος). Είναι, λοιπόν, αναγκαία η διερεύνηση των ακραίων τιμών, ώστε να διαπιστωθεί η ορθότητά τους.

Ο μέσος όρος συμβολίζεται με  $\mu$  ή  $\bar{x}$  αν αναφερόμαστε σε πληθυσμό ή σε δείγμα, αντιστοίχως. Προφανώς, συνηθέστερος είναι ο συμβολισμός  $\bar{x}$ .

■ **Διάμεσος (median):** Ουσιαστικά είναι η μεσαία παρατήρηση. Είναι, δηλαδή, η παρατήρηση που χωρίζει το δείγμα σε δύο ίσα μέρη. Αν για παράδειγμα έχουμε 3 παρατηρήσεις, οι οποίες σε αύξουσα σειρά είναι οι αριθμοί 6, 11, 19, τότε ως διάμεσος είναι η 2η παρατήρηση, δηλαδή το 11. Σε περίπτωση ζυγού αριθμού παρατηρήσεων, ως διάμεσος συνήθως λαμβάνεται ο μέσος όρος των δύο μεσαίων παρατηρήσεων ή οποιαδήποτε άλλη τιμή μεταξύ των δύο μεσαίων τιμών. Έτσι, αν έχουμε τις παρατηρήσεις 7, 9, 15, 25, τότε ως διάμεσος λαμβάνεται το 12 (ο μέσος όρος των παρατηρήσεων 9 και 15). Σε αντίθεση με το μέσο όρο, η διάμεσος **δεν** επηρεάζεται τόσο πολύ από ακραίες παρατηρήσεις. Αν, για παράδειγμα, στην προηγούμενη περίπτωση αντί της τιμής 25 είχαμε την τιμή 2500, η διάμεσος δε θα άλλαζε τιμή. Αντίθετα, ο μέσος όρος θα παρουσίαζε μεγάλη διαφοροποίηση (από 14 που ήταν η αρχική τιμή, θα γινόταν 632,75). Καλό είναι, λοιπόν, να αποφεύγεται η χρήση του μέσου όρου και να υιοθετείται η διάμεσος, όταν οι μεταβλητές έχουν ακραίες τιμές. Αυτό συμβαίνει χαρακτηριστικά στην περίπτωση της μεταβλητής του εισοδήματος και για το λόγο αυτό στη διεθνή βιβλιογραφία είναι συχνή η αναφορά στη «διάμεσο του εισοδήματος» (median income).

■ **Τα τεταρτημόρια (quartiles):** Είναι οι τιμές που χωρίζουν τα δεδομένα σε τέσσερα ίσα μέρη. Έστω για παράδειγμα ότι έχουμε δεδομένα, που αφορούν στην επίδοση των μαθητών σε μια εξέταση, αναπτυγμένα στην εικοσαβάθμια κλίμακα. Θεωρούμε ότι το πρώτο τεταρτημόριο έχει τιμή 11. Τότε μπορούμε να πούμε ότι το 25% των μαθητών έχει βαθμολογική επίδοση μέχρι και 11. Ομοίως, ορίζονται τα δεύτερο και το τρίτο τεταρτημόριο για το 50% και 75% των παρατηρήσεων, αντιστοίχως. Είναι προφανές ότι στο δεύτερο τεταρτημόριο εντοπίζεται η διάμεσος.

■ **Επέκταση της έννοιας των τεταρτημορίων είναι τα εκατοστημόρια ή ποσοστημόρια ή ποσοστιαία σημεία (percentiles):** Έστω ότι το 40ό εκατοστημό-

ριο της βαθμολογικής επίδοσης είναι η τιμή 13. Αυτό σημαίνει ότι το 40% των μαθητών έχουν επίδοση μέχρι και 13, ενώ το υπόλοιπο 60% πάνω από 13.

■ **Επικρατούσα τιμή ή κορυφή (mode):** Πρόκειται για τη συχνότερη παρατήρηση. Σε αντίθεση με το μέσο όρο, η κορυφή μπορεί να μην είναι μοναδική. Για παράδειγμα, όταν τα δεδομένα είναι οι αριθμοί 5, 6, 9, 9, 12, 12, 17, 21, 23 τότε έχουμε 2 κορυφές (9 και 12). Θεωρείται το λιγότερο χρήσιμο μέτρο κεντρικής θέσης.

### 1.2.2. Μέτρα διασποράς

Τα μέτρα διασποράς εξετάζουν το πόσο «απλωμένα» ή «διεσπαρμένα» είναι τα δεδομένα. Όπως και τα μέτρα θέσης αναφέρονται σε συνεχή δεδομένα. Ας δούμε ένα παράδειγμα:

Τα δεδομένα του Πίνακα 1 αναφέρονται σε 3 δείγματα βαθμολογίας μαθητών στις πανελλαδικές εξετάσεις (παρατίθενται τα μόρια εισαγωγής).

**Πίνακας 1:** Δεδομένα βαθμολογίας από 3 δείγματα

<b>Δείγμα I</b>	<b>Δείγμα II</b>	<b>Δείγμα III</b>
12.546	14.623	9.785
14.128	14.789	12.675
14.082	14.341	14.014
15.431	14.567	17.098
16.672	14.513	19.012

Είναι εύκολο να εξακριβώσει κανείς ότι και στα τρία δείγματα οι μέσοι όροι είναι κοντά μεταξύ τους (γύρω στα 14.560 μόρια). Όμως είναι φανερό ότι το «άπλωμα» των δεδομένων διαφέρει ανάλογα με το δείγμα. Οι παρατηρήσεις στο δείγμα II είναι πολύ κοντά στο μέσο όρο ενώ στο δείγμα III απέχουν πάρα πολύ. Μια ενδιάμεση κατάσταση παρατηρείται στο δείγμα I.

Τα κυριότερα μέτρα διασποράς είναι τα ακόλουθα:

■ **Διακύμανση ή διασπορά (variance) και τυπική απόκλιση (standard deviation):** Η διακύμανση και η τυπική απόκλιση είναι ουσιαστικά μαθηματικά ισοδύναμες ποσότητες, αφού η δεύτερη είναι η τετραγωνική ρίζα της πρώτης. Ειδικά η τυπική απόκλιση αποτελεί το κυριότερο μέτρο διασποράς και δηλώνει τη μέση απόσταση κάθε παρατήρησης από το μέσο όρο των δεδομένων. Όταν οι τιμές των δεδομένων απέχουν πολύ (λίγο) από το μέσο όρο, τότε η τυπική απόκλιση λαμβάνει μεγάλες



(μικρές) τιμές. Αξίζει να προσέξουμε ότι η διακύμανση και η τυπική απόκλιση έχουν πάντα **μη αρνητικές τιμές**. Δε νοείται υπολογισμός αυτών των δεικτών να καταλήγει σε αρνητικό αποτέλεσμα.

Στην πράξη χρησιμοποιείται σχεδόν αποκλειστικά η τυπική απόκλιση, γιατί εκφράζεται στις ίδιες μονάδες με αυτές των δεδομένων. Έτσι, στο προηγούμενο παράδειγμα οι μονάδες της τυπικής απόκλισης για το κάθε δείγμα είναι τα μόρια εισαγωγής των μαθητών. Αντίθετα, οι μονάδες της διακύμανσης είναι οι αρχικές μονάδες των δεδομένων υψωμένες στο τετράγωνο (στο προηγούμενο παράδειγμα αυτές θα ήταν τα μόρια εισαγωγής στο τετράγωνο). Είναι προφανές ότι η πρακτική κατανόηση της διακύμανσης είναι πολύ δύσκολη (γι' αυτό άλλωστε δε χρησιμοποιείται συχνά).

Η διασπορά συμβολίζεται με  $\sigma^2$  ή  $s^2$  αν αναφερόμαστε σε πληθυσμό ή σε δείγμα, αντιστοίχως. Ομοίως, η τυπική απόκλιση συμβολίζεται με  $\sigma$  ή  $s$  για πληθυσμό ή δείγμα, αντιστοίχως. Προφανώς, στις περισσότερες περιπτώσεις έχουμε στοιχεία του δείγματος και, καθώς μας ενδιαφέρει (κυρίως) η τυπική απόκλιση, χρησιμοποιείται ο συμβολισμός  $s$ .

■ **Εύρος (range):** Είναι η διαφορά μεταξύ της μικρότερης και της μεγαλύτερης παρατήρησης. Είναι εύκολο μέτρο όσον αφορά τον υπολογισμό, αλλά δε θεωρείται ιδιαίτερα αξιόπιστο, αφού βασίζεται σε δύο μόνο παρατηρήσεις. Επηρεάζεται και αυτό έντονα από ακραίες τιμές.

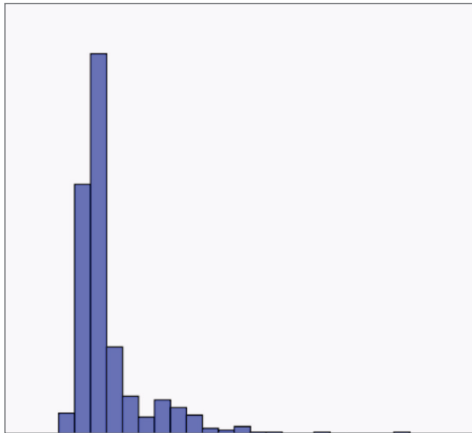
■ **Ενδοτεταρτημοριακό εύρος (interquartile range):** Είναι η διαφορά του πρώτου από το τρίτο τεταρτημόριο. Στο μεταξύ τους διάστημα συγκεντρώνεται το μεσαίο 50% των τιμών του δείγματος. Μια μικρή τιμή αυτού του δείκτη δηλώνει μικρή διασπορά του «μεσαίου» 50% των δεδομένων.

### 1.2.3. Άλλα μέτρα Περιγραφικής Στατιστικής

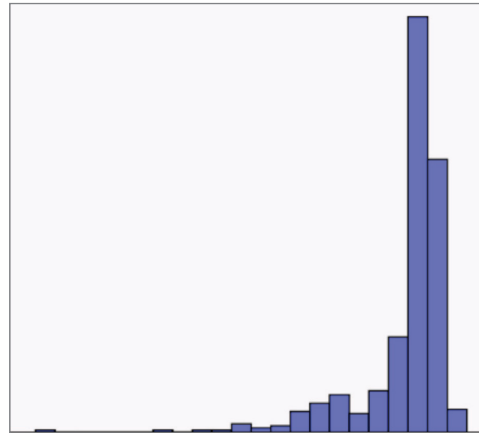
Εκτός από τα μέτρα θέσης και διασποράς υπάρχουν και άλλα χρήσιμα μέτρα περιγραφικής στατιστικής που αφορούν στην ασυμμετρία, στη σχετική μεταβλητότητα αλλά και στην κυρτότητα των δεδομένων. Ας δούμε τους σημαντικότερους από αυτούς τους δείκτες.

■ **Συντελεστές ασυμμετρίας (skewness coefficients):** Καταγράφουν τη συμμετρία ή μη των δεδομένων σχετικά με την κορυφή τους. Υπάρχουν πολλοί σχετικοί συντελεστές. Χωρίς να υπεισέλθουμε σε μαθηματικές προσεγγίσεις σημειώνουμε ότι όταν ο συντελεστής που υπολογίζεται από το SPSS είναι θετικός, τότε τα δεδομένα είναι «τραβηγμένα» προς τα δεξιά, όπως στο παρακάτω διάγραμμα (Σχήμα 2).





**Σχήμα 2:** Ιστόγραμμα με θετικό συντελεστή ασυμμετρίας



**Σχήμα 3:** Ιστόγραμμα με αρνητικό συντελεστή ασυμμετρίας

Αντίθετα, η αρνητική τιμή του συντελεστή ασυμμετρίας δηλώνει ότι τα δεδομένα σχηματίζουν «ουρά» προς τα αριστερά, όπως φαίνεται και στο Σχήμα 3. Προφανώς τα παραπάνω ισχύουν για έντονα θετικές ή αρνητικές τιμές του συντελεστή. Τιμές πολύ κοντά στο 0 (είτε θετικές είτε αρνητικές) δηλώνουν ότι έχουμε ουσιαστικά μια συμμετρική κατανομή.

Μια ενδιαφέρουσα σχέση προκύπτει μεταξύ του μέσου όρου και της διαμέσου σε περίπτωση ασυμμετρίας. Σε περίπτωση θετικής ασυμμετρίας, τα δεδομένα έχουν «τραβηχτεί» προς τα δεξιά. Άρα υπάρχουν ορισμένες υψηλές θετικές (ακραίες) τιμές. Αφού ο μέσος όρος επηρεάζεται περισσότερο από αυτές τις τιμές σε σχέση με τη διάμεσο, αναμένουμε ότι στη θετική ασυμμετρία ο μέσος όρος θα είναι μεγαλύτερος αριθμητικά από τη διάμεσο. Αντιστρόφως, στην αρνητική ασυμμετρία αναμένουμε ότι ο μέσος όρος θα είναι μικρότερος αριθμητικά από τη διάμεσο.

■ **Συντελεστής μεταβλητότητας (CV-coefficient of variation):** Χρησιμοποιείται για τη σύγκριση της μεταβλητότητας-ανομοιογένειας μεταξύ δύο ή περισσότερων δειγμάτων, όταν διαφέρει η τυπική απόκλιση αλλά και ο μέσος όρος τους. Είναι επίσης χρήσιμος δείκτης αν τα δείγματα εκφράζονται σε διαφορετικές μονάδες μέτρησης (π.χ., το εισόδημα δύο πληθυσμών σε ευρώ και σε δολάρια). Αυτό συμβαίνει γιατί ο συγκεκριμένος συντελεστής είναι ανεξάρτητος των μονάδων μέτρησης των τιμών του δείγματος.

Για κάθε δείγμα, ως συντελεστής μεταβλητότητας ορίζεται ο λόγος της τυπικής απόκλισης προς το μέσο όρο. Ως ένα γενικό κανόνα, μπορούμε να πούμε ότι όταν ο

συντελεστής μεταβλητότητας δεν ξεπερνά το 10%, το κάθε δείγμα μπορεί να θεωρηθεί αρκετά ομοιογενές. Έχουμε, λοιπόν, τον παρακάτω τύπο:

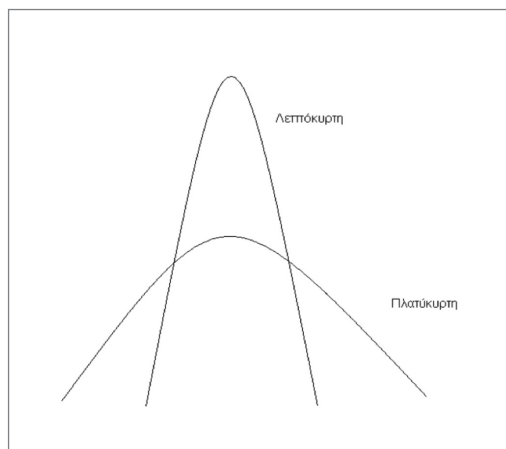
$$CV = \frac{\text{Τυπική απόκλιση}}{\text{Μέσος όρος}} \times 100\%$$

### Ας δούμε ένα σχετικό παράδειγμα:

Ένας ερευνητής καταγράφει το μέσο όρο και την τυπική απόκλιση της βαθμολογικής επίδοσης στις πανελλαδικές εξετάσεις από δύο σχολικές μονάδες Α και Β. Για τη σχολική μονάδα Α, ο μέσος όρος είναι 14.752 μόρια εισαγωγής και η τυπική απόκλιση είναι 1.215 μόρια. Αντίστοιχα, για τη σχολική μονάδα Β, ο μέσος όρος είναι 10.985 μόρια εισαγωγής και η τυπική απόκλιση είναι 1.120 μόρια. Σύμφωνα με τον παραπάνω τύπο, ο συντελεστής μεταβλητότητας στη σχολική μονάδα Α είναι 8,25% ενώ στη μονάδα Β είναι 10,2%. Κατά συνέπεια υπάρχει μεγαλύτερη ανομοιογένεια στη μονάδα Β, έστω και αν έχει μικρότερη τυπική απόκλιση.

Παρόμοιας χρήσης αλλά λιγότερο διαδεδομένος είναι ο συντελεστής Gini.

■ **Συντελεστής κύρτωσης (kurtosis coefficient):** Εξετάζει την κύρτωση της καμπύλης των δεδομένων. Στην περίπτωση που οι τιμές είναι συγκεντρωμένες γύρω από μια κεντρική τιμή, τότε η κατανομή ονομάζεται *λεπτόκυρτη* και ο συντελεστής κύρτωσης είναι θετικός. Στην αντίθετη περίπτωση (δηλαδή όταν ο συντελεστής είναι αρνητικός και οι τιμές είναι πιο διάσπαρτες), η κατανομή των δεδομένων ονομάζεται *πλατύκυρτη*. Η διαφορά στην κύρτωση των δεδομένων φαίνεται στη διαγραμματική απεικόνιση δύο διαφορετικών κατανομών (Σχήμα 4).



Σχήμα 4: Παραδείγματα κύρτωσης

### 1.3. Πίνακες και Διαγράμματα Περιγραφικής Στατιστικής

Εκτός από τον υπολογισμό των προαναφερόμενων μέτρων, στην οπτικοποίηση των πρώτων ερευνητικών ευρημάτων είναι απαραίτητη η δημιουργία ορισμένων πινάκων ή/και διαγραμμάτων. Με τον τρόπο αυτό ο ερευνητής αποκτά μια ολοκληρωμένη εικόνα των ευρημάτων του δείγματος πριν από την πραγματοποίηση πιο σύνθετων στατιστικών αναλύσεων. Ας σημειωθεί ότι σε μια ερευνητική εργασία όπου γίνεται χρήση πινάκων ή/και διαγραμμάτων, αυτά θα πρέπει να συνοδεύονται από αρίθμηση, τίτλο και πιθανώς πηγή (όταν τα δεδομένα προέρχονται από άλλη ερευνητική μελέτη). Στη συνέχεια θα αναφερθούμε στα πιο σημαντικά είδη πινάκων και διαγραμμάτων.

#### 1.3.1. Πίνακες

Πίνακες χρησιμοποιούνται κυρίως για την απεικόνιση διακριτών μεταβλητών. Μπορούν να καταγράφουν την κατανομή μίας ή δύο μεταβλητών. Ας δούμε ένα σχετικό παράδειγμα:

##### Παράδειγμα:

Μια ερευνήτρια καταγράφει, μεταξύ άλλων, το οικονομικό και μορφωτικό επίπεδο ενός δείγματος 300 ερωτωμένων. Οι τιμές που λαμβάνει η κάθε μεταβλητή είναι οι ακόλουθες:

Για το οικονομικό επίπεδο (ετήσιο εισόδημα σε ευρώ):

- 1= έως 12 χιλιάδες ευρώ
- 2= από 12-24 χιλιάδες ευρώ
- 3= από 24-36 χιλιάδες ευρώ
- 4=άνω των 36 χιλιάδων ευρώ

Για το μορφωτικό επίπεδο (ανώτερος τίτλος σπουδών):

- 1= απόφοιτος Δημοτικού-Γυμνασίου
- 2= απόφοιτος Λυκείου
- 3= απόφοιτος ΑΕΙ-ΑΤΕΙ
- 4= μεταπτυχιακός τίτλος

Οι παραπάνω μεταβλητές είναι διατακτικές αφού η τιμή 1 αντιστοιχεί σε χαμηλότερο οικονομικό και αντίστοιχα μορφωτικό επίπεδο από την τιμή 2 κ.ο.κ. Η δημιουργία των κατάλληλων πινάκων για κάθε μεταβλητή χωριστά προϋποθέτει τον υπολογισμό των συχνοτήτων για την κάθε τιμή της μεταβλητής. Για τη μεταβλητή του μορφωτικού επιπέδου, ένας ενδεικτικός πίνακας είναι ο ακόλουθος (Πίνακας 2):

Μορφωτικό επίπεδο	Συχνότητα	Ποσοστό (%)	Αθροιστική συχνότητα	Αθροιστικό ποσοστό (%)
1	56	18,7	56	18,7
2	127	42,3	183	61
3	85	28,3	268	89,3
4	32	10,7	300	100
ΣΥΝΟΛΟ	300	100		

**Πίνακας 2:** Συχνότητες τιμών για το μορφωτικό επίπεδο

Οι στήλες της αθροιστικής συχνότητας και του αθροιστικού ποσοστού αθροίζουν τις συχνότητες και τα ποσοστά μέχρι και τη συγκεκριμένη τιμή της μεταβλητής. Για παράδειγμα, για το μορφωτικό επίπεδο με τιμή 2, η αθροιστική συχνότητα υπολογίζεται αθροίζοντας τις συχνότητες των τιμών 1 και 2 ( $127+56=183$ ).

Μια άλλη επιλογή στην απεικόνιση πινάκων είναι οι **πίνακες διπλής εισόδου (two-way tables)**. Οι πίνακες αυτοί καταγράφουν τις συχνότητες των τιμών **δύο** μεταβλητών σε γραμμές και στήλες. Για το παράδειγμά μας ένας ενδεικτικός πίνακας διπλής εισόδου είναι ο ακόλουθος (Πίνακας 3):

Μορφωτικό επίπεδο	Οικονομικό επίπεδο				ΣΥΝΟΛΟ
	1	2	3	4	
1	26	20	6	4	56
2	18	53	51	5	127
3	14	43	22	6	85
4	7	5	8	12	32
ΣΥΝΟΛΟ	65	121	87	27	300

**Πίνακας 3:** Πίνακας διπλής εισόδου μορφωτικού και οικονομικού επιπέδου

Στην ανάλυση των πινάκων διπλής εισόδου θα αναφερθούμε εκτενέστερα στο Κεφάλαιο 7. Αξίζει όμως στο σημείο αυτό να σημειώσουμε ότι ο ερευνητής πρέπει να είναι ιδιαίτερα προσεκτικός στην ομαδοποίηση στοιχείων από πολλούς πίνακες σε ένα γενικό πίνακα, καθώς μπορεί να αντιστραφεί η αρχική τάση. Το φαινόμενο είναι γνωστό ως το παράδοξο του Simpson (Simpson's paradox) και αναλύεται στο Παράρτημα 1.

### 1.3.2. Διαγράμματα<sup>1</sup>

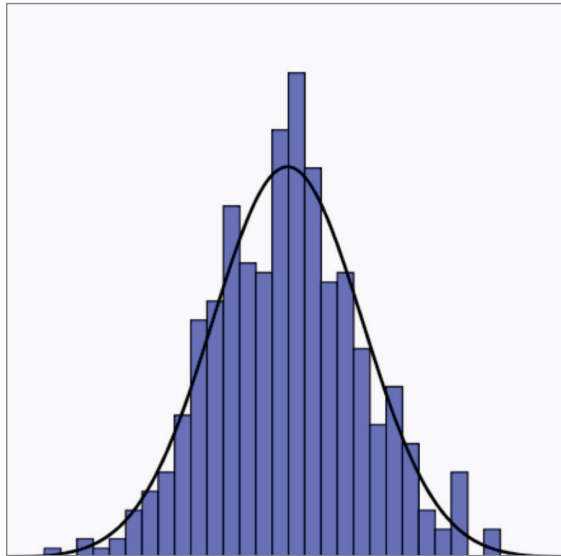
Μια πολύ χρήσιμη επιλογή της περιγραφικής στατιστικής είναι η παρουσίαση των αποτελεσμάτων με τη χρήση κατάλληλων διαγραμμάτων. Τονίζουμε ότι ο ερευνητής θα πρέπει να είναι προσεκτικός, ώστε το διάγραμμα να βοηθά στην κατανόηση των ερευνητικών δεδομένων, χωρίς να χρειάζεται να δοθούν πολλές σχετικές διευκρινίσεις. Συχνά λάθη που γίνονται στην κατασκευή των διαγραμμάτων είναι η παρουσία πολλών και επικαλυπτόμενων τιμών για τις διακριτές μεταβλητές και η πυκνή ή αραιή αρίθμηση αξόνων για τις συνεχείς μεταβλητές.

Στη συνέχεια θα αναφερθούμε στα κυριότερα είδη διαγραμμάτων:

■ **Το Ιστόγραμμα (Histogram):** Πρόκειται για τη συνηθέστερη επιλογή γραφικής παράστασης συνεχών (κυρίως) μεταβλητών<sup>2</sup>. Στα συνεχή δεδομένα, οι τιμές της μεταβλητής ομαδοποιούνται και οι ομάδες διατάσσονται στον οριζόντιο άξονα κατ' αύξουσα σειρά. Στη συνέχεια από κάθε ομάδα υψώνουμε ορθογώνια, το ύψος των οποίων αντιστοιχεί στη συχνότητα κάθε ομάδας. Το ιστόγραμμα μας δίνει μια πολύ καλή οπτική απεικόνιση για την κατανομή των δεδομένων. Αρχικά, όπως σημειώσαμε και στα μέτρα ασυμμετρίας, φαίνεται όχι μόνο αν υπάρχει αλλά και το είδος της ασυμμετρίας. Επίσης, όπως θα δούμε και στο Κεφάλαιο 2, το ιστόγραμμα χρησιμεύει ιδιαίτερα προκειμένου να διαπιστώσουμε αν τα δεδομένα ακολουθούν κανονική κατανομή. Στο Σχήμα 5 παρουσιάζεται ένα συμμετρικό ιστόγραμμα μαζί με την καμπύλη η οποία αποτυπώνει την κατανομή των δεδομένων. Σε πολλές περιπτώσεις ο ερευνητής, αντί της πλήρους αποτύπωσης του ιστογράμματος, προτιμά να παρουσιάσει μόνο την καμπύλη.

■ **Το Θηκόγραμμα (box-plot):** Πρόκειται για ένα χρήσιμο διάγραμμα, ειδικά όσον αφορά τη σύγκριση συνεχών κατανομών σε διαφορετικούς πληθυσμούς (π.χ., το εισόδημα μεταξύ ανδρών-γυναικών). Για την κατασκευή του δημιουργείται ένα ορθογώνιο πλαίσιο με βάσεις το πρώτο και το τρίτο τεταρτημόριο. Ενδιάμεσα τοποθετείται η διάμεσος (το 2ο τεταρτημόριο). Από τα μέσα των βάσεων αναπτύσσονται γραμμές οι οποίες συνδέουν τις οριακές τιμές της μεταβλητής. Ο υπολογισμός των οριακών τιμών γίνεται βάσει ενός μαθηματικού τύπου που δε θα αναφερθεί. Απλώς επισημαίνουμε ότι «καλύπτει» αρκετή απόσταση προς τις δύο πλευρές των

1. Αρκετές φορές, αντί του όρου «διάγραμμα» χρησιμοποιείται ο όρος «γράφημα». Αν και υπάρχουν κάποιες διαφορές μεταξύ των δύο εννοιών, θα χρησιμοποιούνται ως ταυτόσημες.  
2. Στην περίπτωση διακριτών δεδομένων, οι τιμές της μεταβλητής αντιστοιχούν στα μεσαία σημεία των βάσεων των ορθογώνιων.



Σχήμα 5: Συμμετρικό ιστόγραμμα

δεδομένων. Αν υπάρχουν και άλλες τιμές εκτός των οριακών τιμών, αυτές δηλώνονται επιπλέον, συνήθως με αστερίσκους ή κουκκίδες.

■ **Το κυκλικό διάγραμμα (pie chart):** Πρόκειται για μια εναλλακτική γραφική απεικόνιση διακριτών δεδομένων. Κάθε «κομμάτι» του κύκλου αντιπροσωπεύει τη συχνότητα ή το ποσοστό κάθε τιμής της διακριτής μεταβλητής. Η χρήση αυτού του διαγράμματος συνιστάται όταν έχουμε μικρό αριθμό τιμών της διακριτής μεταβλητής.

■ **Το διάγραμμα σκεδασμού (scatter plot):** Αναφέρεται στην απεικόνιση δύο συνεχών μεταβλητών σε έναν οριζόντιο και σε έναν κάθετο άξονα (άξονες X-Y, αντιστοίχως). Ουσιαστικά, αποτυπώνει την ταυτόχρονη μεταβολή των δύο μεταβλητών. Μπορεί να θεωρηθεί ότι είναι κάτι ανάλογο του πίνακα διπλής εισόδου, με τη διαφορά ότι έχουμε συνεχείς και όχι διακριτές μεταβλητές. Οφείλουμε να επισημάνουμε την κατάλληλη (και ορθή) βαθμονόμηση των αξόνων, προκειμένου να είναι σωστή η απεικόνιση. Για παράδειγμα, αν η συνεχής μεταβλητή έχει ελάχιστη και μέγιστη τιμή τους αριθμούς 23 και 92 αντιστοίχως, τότε προφανώς δεν έχει νόημα ο σχετικός άξονας να εκτείνεται λ.χ. ως το 1000. Είναι πιθανό η πυκνή ή αραιή αρίθμηση των αξόνων να «αποκρύψει» την πραγματική τάση μεταβολής των δύο μεταβλητών.

■ **Το ραβδόγραμμα (bar chart):** Το ραβδόγραμμα επιτρέπει τη γραφική απεικόνιση διακριτών μεταβλητών. Η συνηθέστερη πρακτική είναι η τοποθέτηση στον ορι-

ζόντιο άξονα των τιμών της μεταβλητής και η ανάπτυξη ράβδων ύψους ανάλογου της συχνότητας για κάθε κατηγορία. Μια διασκευή του ραβδογράμματος είναι το διάγραμμα Pareto, το οποίο εμφανίζει τις κατηγορίες της διακριτής μεταβλητής σε φθίνουσα σειρά.

Για την καλύτερη κατανόηση των παραπάνω μέτρων, πινάκων και διαγραμμάτων θα παρουσιάσουμε στη συνέχεια αναλυτικές εφαρμογές τους στο SPSS.

## 1.4. Εφαρμογές στο SPSS

(α) Τα δεδομένα αυτής της εφαρμογής αναφέρονται σε 80 περιοχές της χώρας και περιέχουν τις εξής μεταβλητές:

*area*: Το είδος της περιοχής με τιμές 1 και 2 για Αστική και Αγροτική-Ημιαστική αντιστοίχως

*income*: Το μέσο ετήσιο εισόδημα των κατοίκων της περιοχής σε ευρώ

*crime*: Ένας δείκτης εγκληματικότητας (αριθμός εγκληματικών πράξεων ανά 1000 κατοίκους)

Η μεταβλητή *area* είναι διακριτή ονομαστική μεταβλητή ενώ οι άλλες δύο είναι συνεχείς μεταβλητές. Τα δεδομένα παρουσιάζονται στην Εικόνα 1:

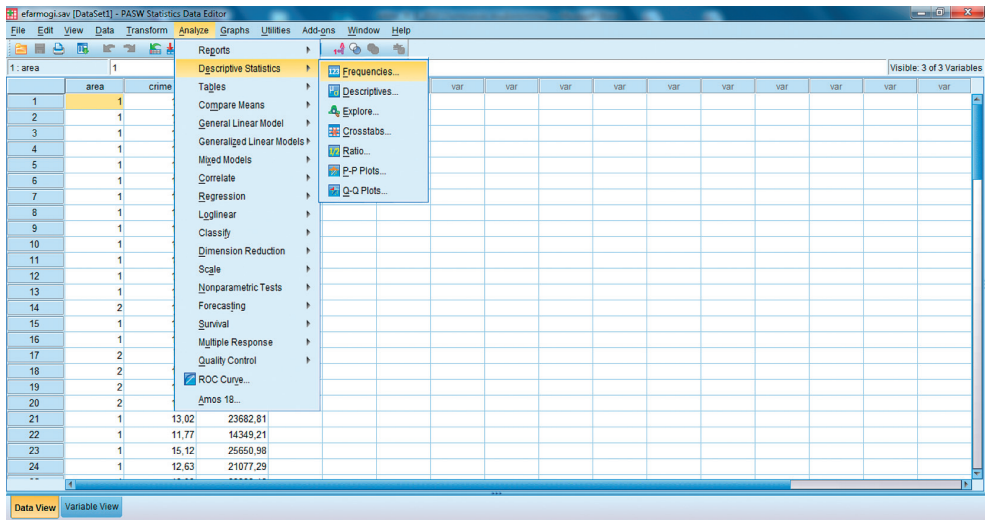
	area	crime	income	var	var	var	var	var	var	var	var	var	var
1	1	11,73	19476,48										
2	1	12,95	26010,19										
3	1	12,61	25427,70										
4	1	12,50	24968,08										
5	1	10,40	14689,79										
6	1	10,53	13394,00										
7	1	12,56	26009,92										
8	1	15,29	27677,87										
9	1	12,74	22374,85										
10	1	12,56	21893,28										
11	1	12,52	20687,93										
12	1	11,38	16810,36										
13	1	15,00	24982,73										
14	2	10,78	19972,12										
15	1	12,37	20476,26										
16	1	10,50	14198,71										
17	2	9,87	10777,43										
18	2	13,42	27730,50										
19	2	13,03	27095,63										
20	2	12,82	22888,56										
21	1	13,02	23682,81										
22	1	11,77	14349,21										
23	1	15,12	25650,98										
24	1	12,63	21077,29										

Εικόνα 1: Δεδομένα του παραδείγματος

Αρχικά επιθυμούμε την εύρεση των μέτρων θέσης, διασποράς και ασυμμετρίας των συνεχών μεταβλητών *crime* και *income*. Υπάρχουν διάφοροι τρόποι για τον υπολογισμό των συγκεκριμένων δεικτών. Όλες οι διαθέσιμες επιλογές περιέχονται στο μενού

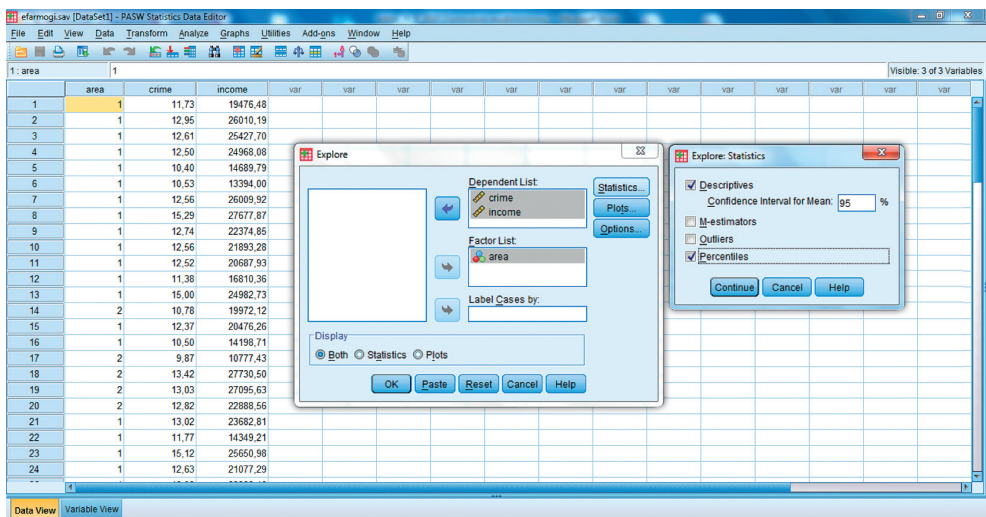
**Analyze→Descriptive Statistics**

Αξιοποιήσιμες είναι οι εντολές Explore, Descriptives και Frequencies (Εικόνα 2).



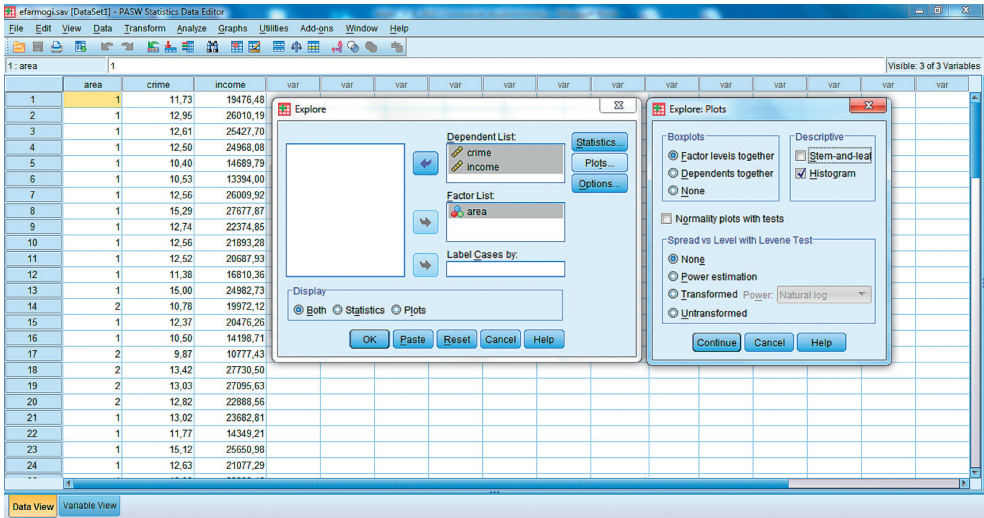
**Εικόνα 2:** Εντολές του SPSS για την εύρεση των μέτρων περιγραφικής στατιστικής

Για την εντολή Explore αρκεί να βάλουμε το όνομα κάθε μεταβλητής στη θέση Dependent List. Αν ο ερευνητής επιθυμεί η ανάλυση να γίνει χωριστά σε κάθε υποσύνολο των δεδομένων, εισάγει τη σχετική μεταβλητή στη θέση Factor List. Εισάγουμε τα ονόματα των δύο (συνεχών) μεταβλητών στο Dependent List και τη μεταβλητή area στη θέση Factor List. Η επιλογή Statistics δίνει τη δυνατότητα να πάρουμε τα κυριότερα μέτρα θέσης και διασποράς, ενώ η επιλογή Plots παρέχει την ευκαιρία κατασκευής ιστογράμματος και θηκογράμματος (Εικόνες 3-4).



**Εικόνα 3:** Επιλογές της εντολής Explore





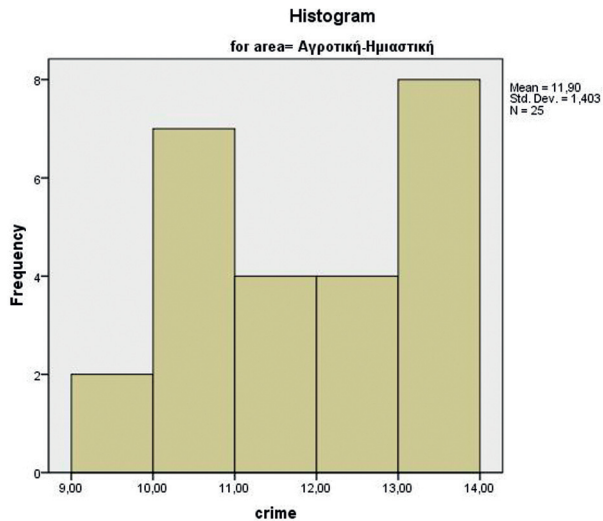
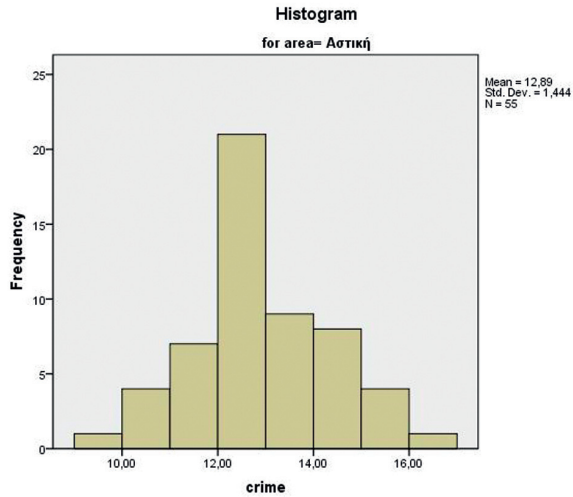
Εικόνα 4: Επιλογές της εντολής Explore (συνέχεια)

Πατώντας OK, η εφαρμογή επιστρέφει τα αποτελέσματα της ανάλυσης, τα κυριότερα μέρη της οποίας παρουσιάζονται παρακάτω (Εικόνα 5).

		Case Processing Summary					
		Cases					
		Valid		Missing		Total	
	area	N	Percent	N	Percent	N	Percent
crime	Αστική	55	100,0%	0	,0%	55	100,0%
	Αγροτική-Ημιαστική	25	100,0%	0	,0%	25	100,0%
income	Αστική	55	100,0%	0	,0%	55	100,0%
	Αγροτική-Ημιαστική	25	100,0%	0	,0%	25	100,0%

Descriptives				Statistic	Std. Error	
crime	Αστική	Mean		12,8908	,19476	
		95% Confidence Interval for	Lower Bound	12,5003		
		Mean	Upper Bound	13,2812		
		5% Trimmed Mean		12,8933		
		Median		12,6271		
		Variance		2,086		
		Std. Deviation		1,44436		
		Minimum		9,65		
		Maximum		16,29		
		Range		6,63		
		Interquartile Range		1,60		
		Skewness		,090		,322
		Kurtosis		-,069		,634
		Αγροτική-Ημιαστική		Mean		
95% Confidence Interval for	Lower Bound			11,3171		
Mean	Upper Bound			12,4754		
5% Trimmed Mean				11,9296		
Median				11,9952		
Variance				1,968		
Std. Deviation				1,40296		
Minimum				9,22		
Maximum				13,91		
Range				4,69		
Interquartile Range				2,61		
Skewness				-,279	,464	
Kurtosis				-1,291	,902	
income	Αστική			Mean		19521,6666
		95% Confidence Interval for	Lower Bound	18471,7641		
		Mean	Upper Bound	20571,5692		
		5% Trimmed Mean		19412,9665		
		Median		19221,7457		
		Variance		15082876,727		
		Std. Deviation		3883,66795		
		Minimum		12030,21		
		Maximum		29491,07		
		Range		17460,86		
		Interquartile Range		5082,93		
		Skewness		,484	,322	
		Kurtosis		-,264	,634	
		Αγροτική-Ημιαστική		Mean		19535,5905
95% Confidence Interval for	Lower Bound			17850,5880		
Mean	Upper Bound			21220,5931		
5% Trimmed Mean				19547,6803		
Median				19620,6562		
Variance				16663431,474		
Std. Deviation				4082,08666		
Minimum				10777,43		
Maximum				27730,50		
Range				16953,07		
Interquartile Range				5507,78		
Skewness				,003	,464	
Kurtosis				,042	,902	

		Percentiles							
		5	10	25	50	75	90	95	
Weighted Average(Definition 1)	crime	Αστική	10,3261	10,8307	12,0699	12,6271	13,6724	15,0483	15,4081
		Αγροτική-Ημιαστική	9,4111	9,9832	10,5333	11,9952	13,1433	13,5277	13,8142
	income	Αστική	14037,7648	14648,8000	16810,3566	19221,7457	21893,2840	25517,0150	26343,7296
		Αγροτική-Ημιαστική	11486,3382	13953,1639	16976,1789	19620,6562	22483,9563	25321,7984	27540,0422
Tukey's Hinges	crime	Αστική			12,1952	12,6271	13,6228		
		Αγροτική-Ημιαστική			10,6269	11,9952	13,0683		
	income	Αστική			16831,7888	19221,7457	21820,2432		
		Αγροτική-Ημιαστική			17856,2369	19620,6562	22079,3508		

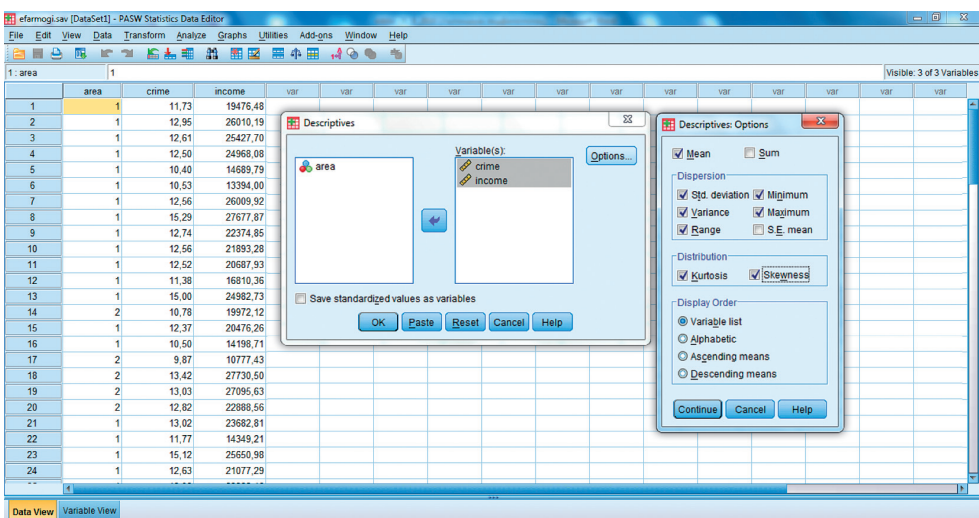


**Εικόνα 5:** Αποτελέσματα της εντολής Explore

Στον πίνακα με τίτλο Descriptives και στη στήλη Statistic εμφανίζονται τα εξής μέτρα: Μέσος όρος (mean), διάμεσος (median), διασπορά (variance), τυπική απόκλιση

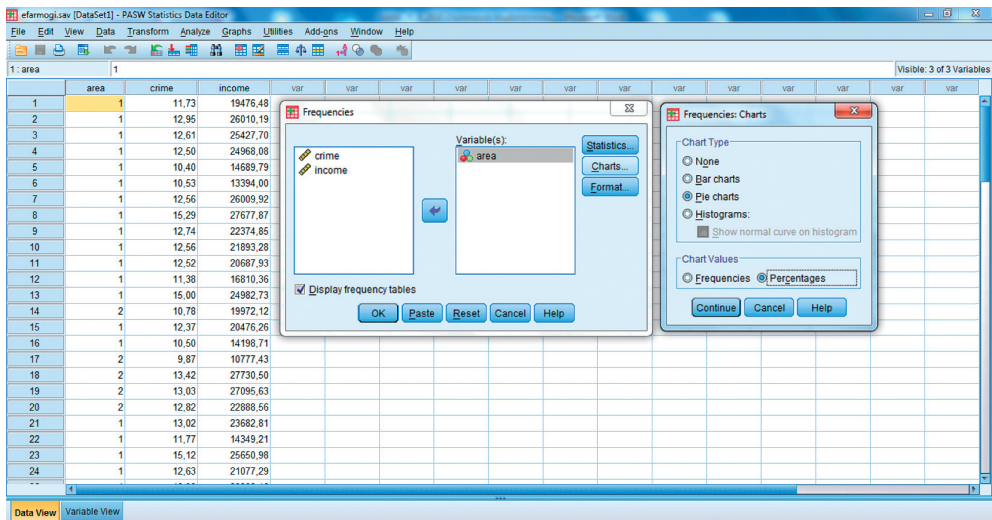
(standard deviation), ελάχιστη και μέγιστη τιμή (minimum, maximum), εύρος (range), ενδοτεταρτημοριακό εύρος (interquartile range) καθώς και οι συντελεστές ασυμμετρίας και κύρτωσης (skewness, kurtosis). Για καθεμία από τις μεταβλητές income και crime, τα αποτελέσματα παρουσιάζονται ξεχωριστά για κάθε επίπεδο τιμών της μεταβλητής area. Στον Πίνακα με τίτλο Percentiles και ειδικά στη γραμμή Weighted Average εμφανίζονται το 5ο, 10ο, 25ο (1ο τεταρτημόριο), 50ό (2ο τεταρτημόριο=διάμεσος), 75ο (3ο τεταρτημόριο), 90ό και 95ο εκατοστιαίο σημείο. Τέλος, ακολουθούν το ιστόγραμμα για τη μεταβλητή crime σε αστικές περιοχές και τα θηκογράμματα για τη μεταβλητή της εγκληματικότητας, ξεχωριστά για αστικές και ημιαστικές-αγροτικές περιοχές. Παρατηρούμε ότι το ιστόγραμμα της Εικόνας 5 είναι μάλλον συμμετρικό, κάτι το οποίο είναι αναμενόμενο αφού ο σχετικός συντελεστής είναι πολύ κοντά στο 0 (0.09). Από τα θηκογράμματα παρατηρούμε ότι η εγκληματικότητα είναι μεγαλύτερη στα αστικά κέντρα, σε σχέση με τις ημιαστικές-αγροτικές περιοχές, κάτι το οποίο επιβεβαιώνεται από τις υψηλότερες τιμές του μέσου και της διαμέσου στον πίνακα Descriptives. Επίσης, υπάρχει μεγαλύτερο εύρος τιμών για την εγκληματικότητα στις αστικές περιοχές. Αυτό φαίνεται τόσο από τις αριθμητικές τιμές (εύρος 6,63 έναντι 4,69) όσο και από το «άνοιγμα» των τιμών στα θηκογράμματα. Ειδικά για τις αστικές περιοχές υπάρχει μια περιοχή πολύ υψηλής εγκληματικότητας (58η παρατήρηση), αλλά και 3 περιοχές χαμηλότερης εγκληματικότητας. Στις περιοχές αυτές η εγκληματικότητα είναι παρόμοια με τις περιοχές χαμηλής εγκληματικότητας των ημιαστικών-αγροτικών περιοχών.

Εκτός της εντολής Explore, μέτρα θέσης παρέχει και η εντολή Descriptives της επιλογής Options. Η συγκεκριμένη εντολή όμως δεν παράγει διαγράμματα (Εικόνα 6).



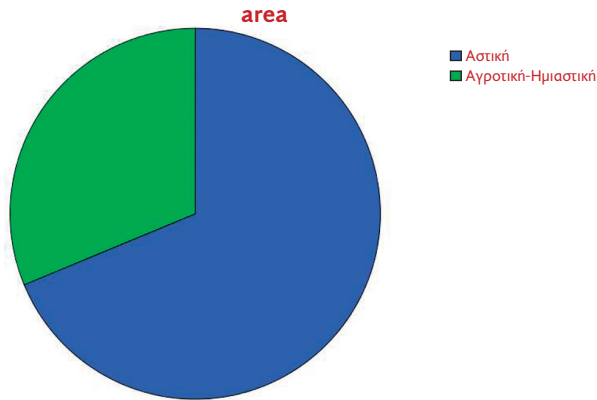
Εικόνα 6: Επιλογές της εντολής Descriptives

Τέλος, η εντολή Frequencies χρησιμοποιείται για την κατασκευή πινάκων συχνότητων σε **διακριτές** μεταβλητές. Με την επιλογή Charts της εντολής, κατασκευάζονται ραβδογράμματα και κυκλικά διαγράμματα. (Εικόνες 7-8). Σε περίπτωση όμως συνεχούς μεταβλητής, η εντολή Frequencies δίνει σειρά μέτρων και (κυρίως) ιστόγραμμα (με τη σχετική καμπύλη), με τη βοήθεια των επιλογών Statistics και Charts, αντίστοιχως (Εικόνες 9-10). Ένα σημαντικό πλεονέκτημα της συγκεκριμένης εντολής είναι ότι παρέχει την κορυφή (mode) με την επιλογή Statistics. Επίσης, στην επιλογή Percentiles ο ερευνητής μπορεί να πληκτρολογήσει το εκατοστημόριο που επιθυμεί και να το εισαγάγει στο σχετικό κουτί επιλέγοντας Add. Τέλος, με την επιλογή Cut points for μπορεί να ζητήσει τα εκατοστημόρια που «κόβουν» το δείγμα σε έναν ορισμένο αριθμό τμημάτων (π.χ., τα δεκατημόρια χωρίζουν το δείγμα σε 10 κομμάτια). Μειονέκτημα για τη χρήση της εντολής στην περίπτωση συνεχούς μεταβλητής είναι ο πολύ μεγάλος πίνακας συχνότητων στο παράθυρο αποτελεσμάτων. Για το λόγο αυτό, στις συνεχείς μεταβλητές καλό είναι ο ερευνητής να απενεργοποιήσει την προεπιλεγμένη εντολή Display frequency tables.



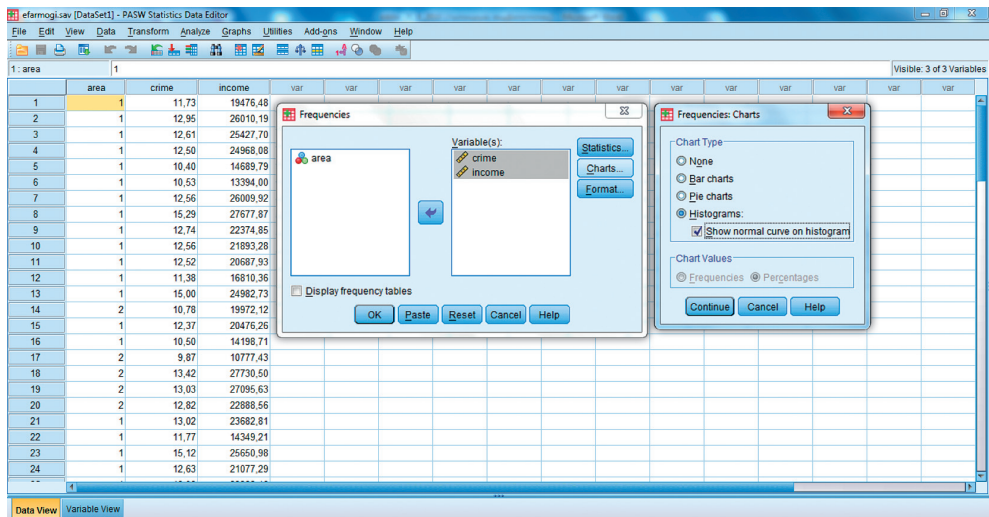
Εικόνα 7: Επιλογές της εντολής Frequencies

		area			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Αστική	55	68,8	68,8	68,8
	Αγροτική-Ημιαστική	25	31,3	31,3	100,0
	Total	80	100,0	100,0	

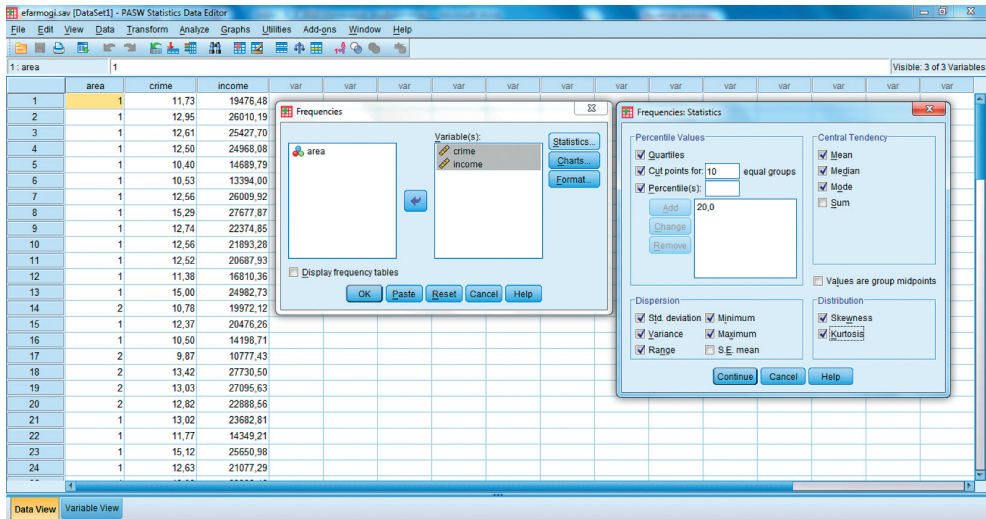


Εικόνα 8: Αποτελέσματα της εντολής Frequencies

Στην πρώτη στήλη του πίνακα area εμφανίζονται οι περιγραφές των μεταβλητών, όπως καταχωρίσθηκαν στο Variable View του SPSS. Στη δεύτερη στήλη παρουσιάζονται οι συχνότητες (στήλη Frequency), τα ποσοστά (στήλη Percent), τα έγκυρα ποσοστά (στήλη Valid Percent) επί του συνόλου αυτών που απάντησαν με έγκυρο τρόπο, καθώς και τα αθροιστικά ποσοστά (στήλη Cumulative Percent). Αλλαγές στο κυκλικό διάγραμμα επιτυγχάνονται με τη χρήση του Chart Editor, ο οποίος ενεργοποιείται με διπλό κλικ στην περιοχή του διαγράμματος. Περισσότερες πληροφορίες για τη δημιουργία διαγραμμάτων στο SPSS παρουσιάζονται στην επόμενη εφαρμογή.



Εικόνα 9: Η επιλογή Charts της εντολής Frequencies



Εικόνα 10: Η επιλογή Statistics της εντολής Frequencies

### (β) Η χρήση του μενού Graph στο SPSS

Από την έκδοση 15 και μετά, το SPSS διαθέτει (εκτός των επιλογών που ενδεικτικά παρουσιάστηκαν μέσω της εντολής Descriptive Statistics), έναν εξελιγμένο δημιουργό διαγραμμάτων-γραφημάτων. Στη συνέχεια θα προσπαθήσουμε να σας εξοικειώσουμε με τη δημιουργία και την τροποποίηση γραφημάτων. Ασφαλώς, προκειμένου να δημιουργήσετε κάποιο γράφημα, θα πρέπει να έχετε ολοκληρώσει την καταχώριση των δεδομένων σας στον Data Editor.

#### Ας δούμε το σχετικό παράδειγμα:

Έστω ότι θέλουμε να κατασκευάσουμε ένα ιστόγραμμα για τις τιμές της μεταβλητής Ηλικία από ένα σύνολο δεδομένων που αφορούν στο βαθμό πολιτικής ενημέρωσης μαθητών Δευτεροβάθμιας Εκπαίδευσης αλλά και κάποιων τάξεων του Δημοτικού. Η μεταβλητή ηλικία έχει καταχωρισθεί ως V2.

Προκειμένου να δημιουργήσετε ένα γράφημα, ακολουθήστε τις παρακάτω οδηγίες:

1. Από το menu File επιλέξτε Open και στη συνέχεια Data. Ανοίξτε το αρχείο των δεδομένων (Εικόνα 11).



	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10a	v10b	v10c	v10d	v11	var
1	2	12	1	1	35	35	6	6	6	1	1	1	1	3	
2	1	12	1	1	10	21	6	6	3	1	1	1	1	3	
3	2	12	1	1	40	50	2	2	1	1	0	1	0	4	
4	2	11	1	1	10	21	6	2	3	1	1	1	0	2	
5	1	12	1	1	35	15	6	6	6	1	0	1	1	4	
6	1	12	1	1	20	10	6	6	6	1	0	1	1	1	
7	2	12	1	1	21	21	5	5	3	0	0	1	1	3	
8	1	12	1	1	21	22	6	2	3	1	0	1	0	5	
9	1	12	1	1	20	35	6	5	3	1	1	1	1	3	
10	1	11	12	1	0	60	2	1	1	1	1	0	0	2	
11	2	12	1	1	31	50	5	3	2	1	1	1	1	4	
12	2	12	1	1	10	50	6	6	3	1	0	1	1	4	
13	1	12	1	1	31	31	4	4	4	1	1	1	1	3	
14	1	12	1	1	10	10	6	6	6	1	1	1	1	1	
15	2	13	1	1	35	35	6	6	6	1	1	1	1	0	
16	1	12	1	1	21	21	6	6	3	1	1	1	1	3	
17	2	12	1	1	26	26	6	4	5	1	1	1	1	5	
18	2	11	1	1	21	22	4	3	3	2	2	2	2	3	
19	1	12	1	1	0	0	2	2	1	1	2	2	2	2	
20	1	12	1	1	0	0	6	6	6	1	1	1	1	3	
21	2	12	1	1	10	50	6	3	3	1	1	0	1	4	
22	2	12	1	1	15	22	4	4	3	2	2	2	1	2	
23	2	11	1	1	31	31	3	3	4	2	2	1	1	2	
24	1	12	1	1	35	10	6	6	2	1	1	1	1	4	

Εικόνα 11: Τα δεδομένα του παραδείγματος

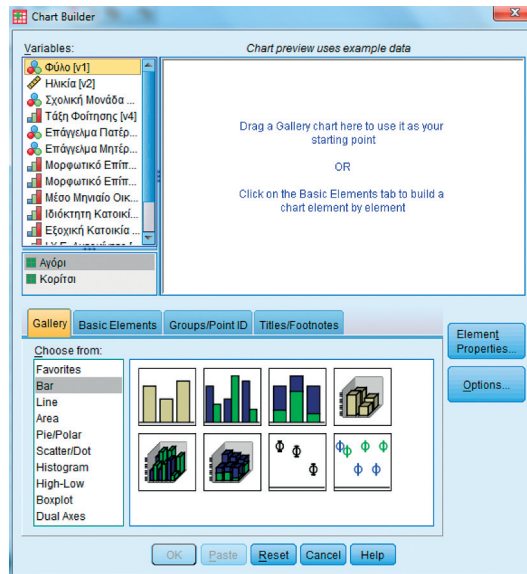
2. Από το menu Graphs επιλέξτε Chart Builder (Εικόνα 12).

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10a	v10b	v10c	v10d	v11	var
1	2	12	1	1	35	35	6	6	6	1	1	1	1	3	
2	1	12	1	1	10	21	6	6	3	1	1	1	1	3	
3	2	12	1	1	40	50	2	2	1	1	0	1	0	4	
4	2	11	1	1	10	21	6	2	3	1	1	1	0	2	
5	1	12	1	1	35	15	6	6	6	1	0	1	1	4	
6	1	12	1	1	20	10	6	6	6	1	0	1	1	1	
7	2	12	1	1	21	21	5	5	3	0	0	1	1	3	
8	1	12	1	1	21	22	6	2	3	1	0	1	0	5	
9	1	12	1	1	20	35	6	5	3	1	1	1	1	3	
10	1	11	12	1	0	60	2	1	1	1	1	0	0	2	
11	2	12	1	1	31	50	5	3	2	1	1	1	1	4	
12	2	12	1	1	10	50	6	6	3	1	0	1	1	4	
13	1	12	1	1	31	31	4	4	4	1	1	1	1	3	
14	1	12	1	1	10	10	6	6	6	1	1	1	1	1	
15	2	13	1	1	35	35	6	6	6	1	1	1	1	0	
16	1	12	1	1	21	21	6	6	3	1	1	1	1	3	
17	2	12	1	1	26	26	6	4	5	1	1	1	1	5	
18	2	11	1	1	21	22	4	3	3	2	2	2	2	3	
19	1	12	1	1	0	0	2	2	1	1	2	2	2	2	
20	1	12	1	1	0	0	6	6	6	1	1	1	1	3	
21	2	12	1	1	10	50	6	3	3	1	1	0	1	4	
22	2	12	1	1	15	22	4	4	3	2	2	2	1	2	
23	2	11	1	1	31	31	3	3	4	2	2	1	1	2	
24	1	12	1	1	35	10	6	6	2	1	1	1	1	4	

Εικόνα 12: Το μενού εντολών Graphs

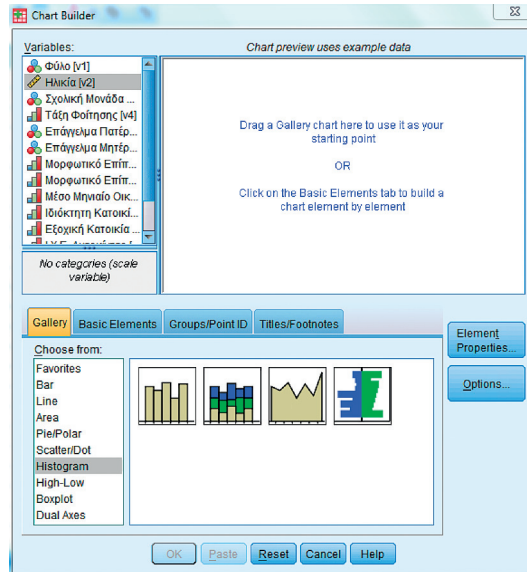


3. Θα διαπιστώσετε ότι το πλαίσιο διαλόγου που αναπτύσσεται (Εικόνα 13) είναι αλληλεπιδραστικό.



Εικόνα 13: Η εντολή Chart Builder

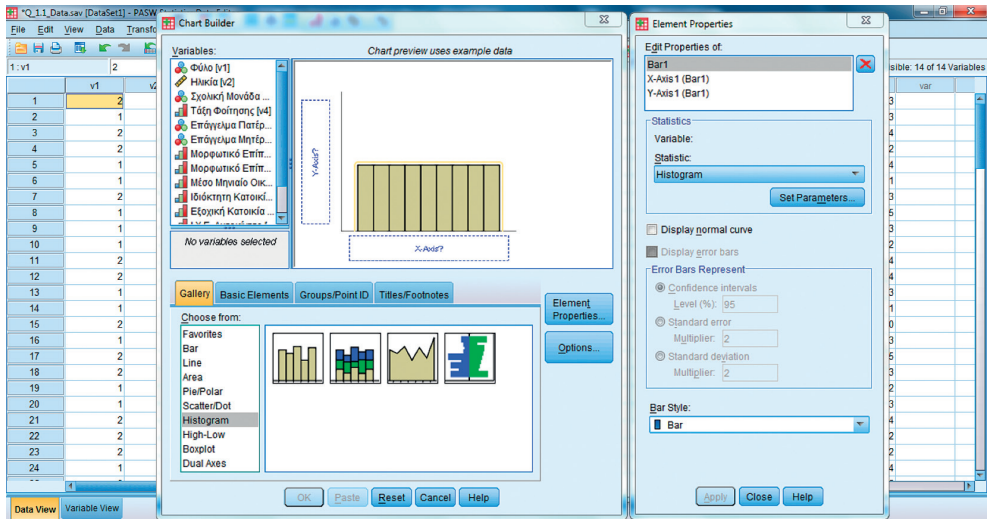
4. Θα πρέπει στο προσκήνιο να είναι ενεργή η επιλογή Gallery, σε διαφορετική περίπτωση επιλέξτε την και φέρτε την στο προσκήνιο, κάνοντας κλικ στη λέξη Gallery.
5. Από τη λίστα των προτεινόμενων επιλογών επιλέξτε τον τύπο γραφήματος (εδώ Histogram). Διαπιστώνετε ότι υπάρχουν πολλοί και διαφορετικοί τύποι γραφημάτων (Εικόνα 14).



Εικόνα 14: Η εντολή Chart Builder

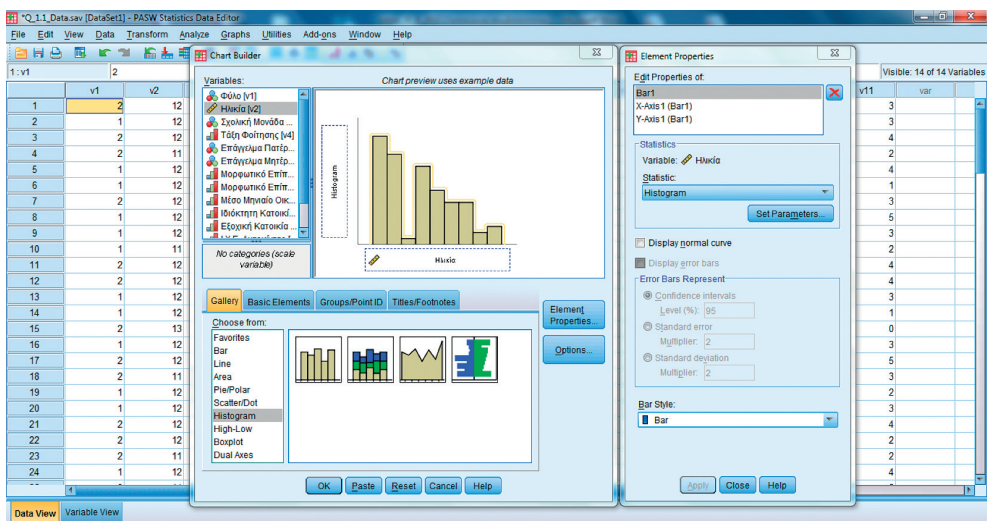
6. Επιλέξτε τον ειδικότερο τύπο γραφήματος<sup>3</sup> από την ποικιλία των προτεινόμενων γραφημάτων (ιστογραμμάτων) και οδηγήστε τον ειδικό τύπο στη λευκή περιοχή (Εικόνα 15). Ένα νέο πλαίσιο διαλόγου ενεργοποιείται και σας προτρέπει να ορίσετε τις βασικές ιδιότητες του γραφήματος. Στην παρούσα φάση δεν είναι απαραίτητο να κάνετε καμιά σχετική επιλογή και σας προτείνεται να κλείσετε το αναδυόμενο πλαίσιο. Άλλωστε, μπορείτε οποιαδήποτε στιγμή να ανακαλέσετε τις ιδιότητες του γραφήματος.

3. Μετακινώντας το δείκτη στους ειδικότερους τύπους ραβδογράμματος, εμφανίζονται οι περιγραφές του κάθε τύπου [απλό ραβδόγραμμα (simple), ραβδόγραμμα συστάδας (clustered), ραβδόγραμμα στοιβάδας (stacked), τρισδιάστατες απεικονίσεις τους κ.ά.].



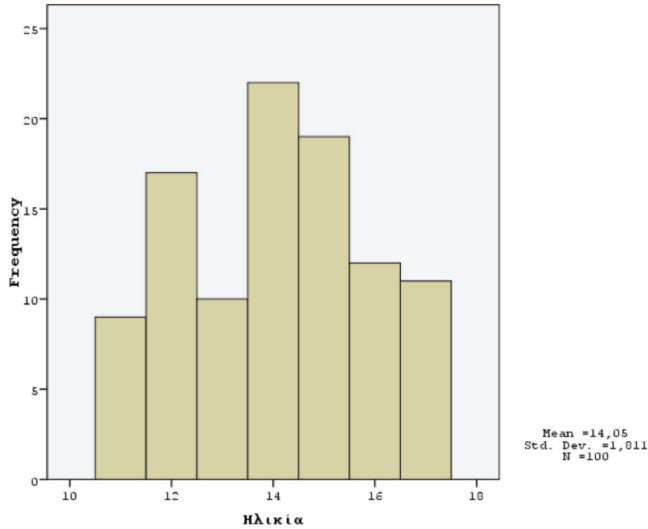
Εικόνα 15: Η εντολή Chart Builder

7. Από τη λίστα μεταβλητών στα αριστερά του ενδεικτικού γραφήματος επιλέξετε τη μεταβλητή για την οποία ζητάτε τη γραφική παράσταση των τιμών της (στη συγκεκριμένη περίπτωση τη μεταβλητή Ηλικία [v2]) και οδηγήστε τη στη θέση X-Axis. Η τοποθέτηση της μεταβλητής στη θέση X-Axis μετασχηματίζει τη μορφή του ενδεικτικού γραφήματος και έχετε ήδη μια πρώτη (ενδεικτική) ένδειξη για το παραγόμενο ιστόγραμμα (Εικόνα 16).



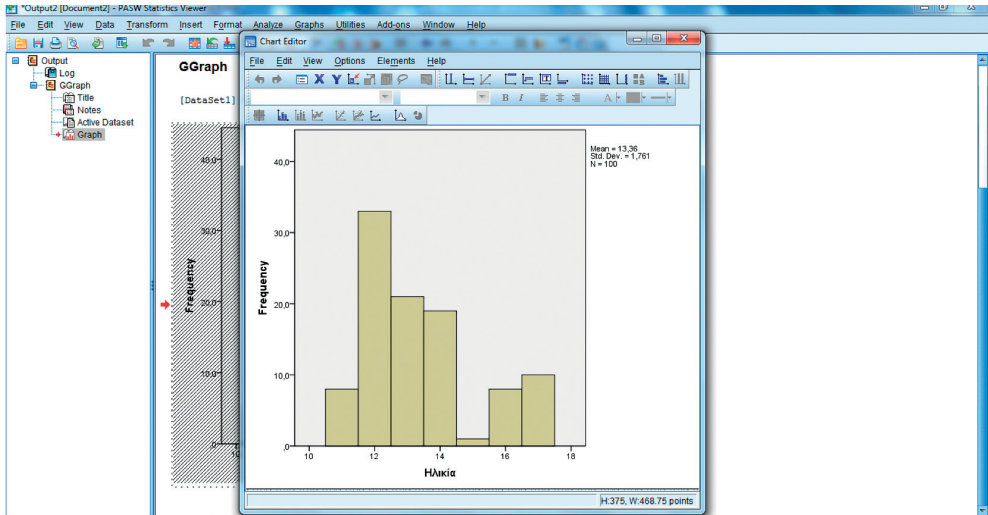
Εικόνα 16: Η εντολή Chart Builder

8. Μπορείτε στη συνέχεια να προσθέσετε τίτλους και υποσημειώσεις με επιλογή της κατάλληλης κάρτας (Titles/Footnotes).
9. Προκειμένου να έχετε μια πρώτη μορφή του ιστογράμματος για την κατανομή των τιμών στη μεταβλητή της Ηλικίας, επιλέξτε το OK. Το αποτέλεσμα εμφανίζεται στο Σχήμα 6.

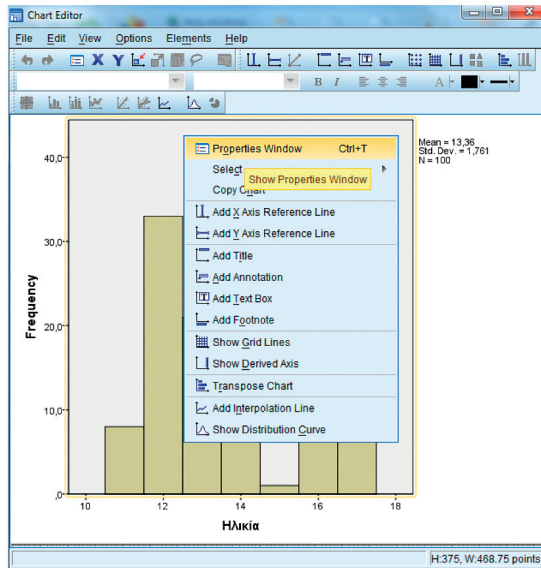


**Σχήμα 6:** Το παραγόμενο ιστόγραμμα

Κάθε παρέμβαση και τροποποίηση του γραφήματος είναι εφικτή με την προϋπόθεση ότι θα ενεργοποιηθεί ο συντάκτης γραφημάτων (Chart Editor) (Εικόνα 17). Η ενεργοποίηση επιτυγχάνεται είτε με δύο κλικ στην επιφάνεια του γραφήματος, είτε με δεξιό κλικ στο γράφημα, επιλογή από το διαθέσιμο menu της επιλογής SPSS Chart Object και στη συνέχεια Open (Εικόνα 18).



Εικόνα 17: Η επιλογή Chart Editor

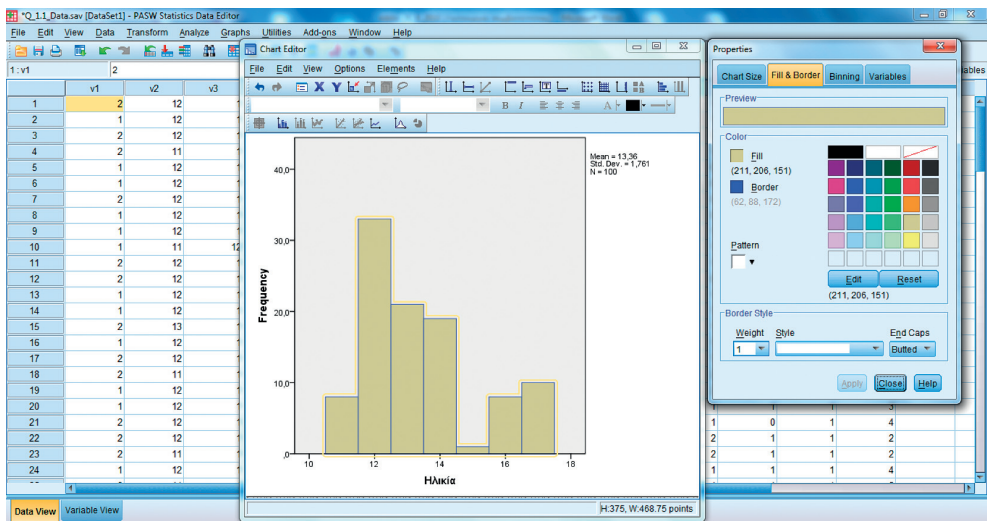


Εικόνα 18: Η επιλογή Chart Editor

Σας παρέχετε η δυνατότητα να:

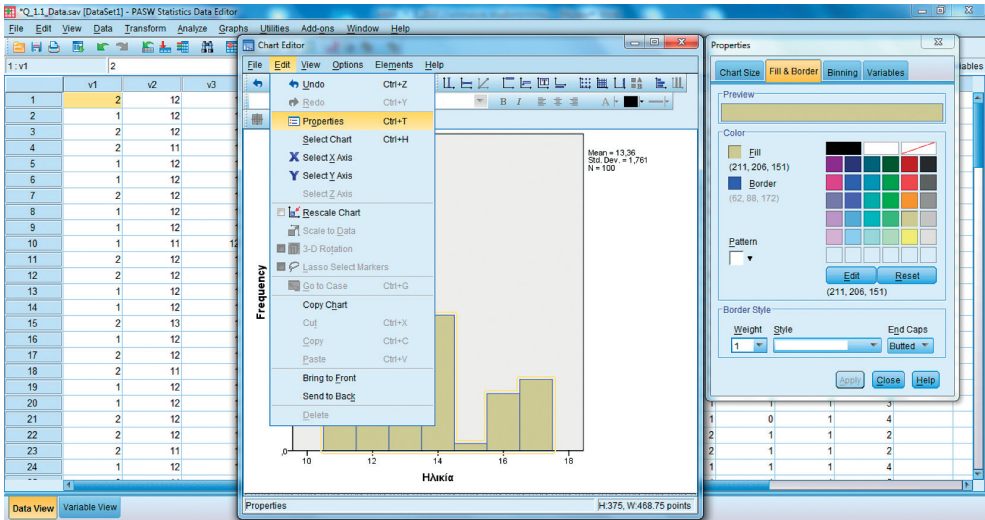
1. αλλάξετε τις χρωματικές (προ)επιλογές
2. μορφοποιήσετε αριθμούς και περιγραφές
3. συντάξετε εκ νέου ή να διορθώσετε κείμενα (τίτλους, υπότιτλους, σημειώσεις κ.λπ.)
4. εμφανίσετε τα αριθμητικά δεδομένα για τις επιλεγείσες τιμές
5. χρησιμοποιήσετε προκαθορισμένους τύπους γραφήματος (templates).

Για να επηρεάσετε οποιοδήποτε στοιχείο, θα πρέπει πρώτα να το επιλέξετε. Έτσι, και προκειμένου να τροποποιήσουμε τη μορφή των στηλών στο ιστόγραμμα που δημιουργήσαμε, κάνουμε κλικ σε οποιαδήποτε στήλη του. Με τον τρόπο αυτό επιλέγονται όλες οι στήλες του γραφήματος και γύρω από την κάθε στήλη εμφανίζεται ένα περίγραμμα (μπλε) (Εικόνα 19).



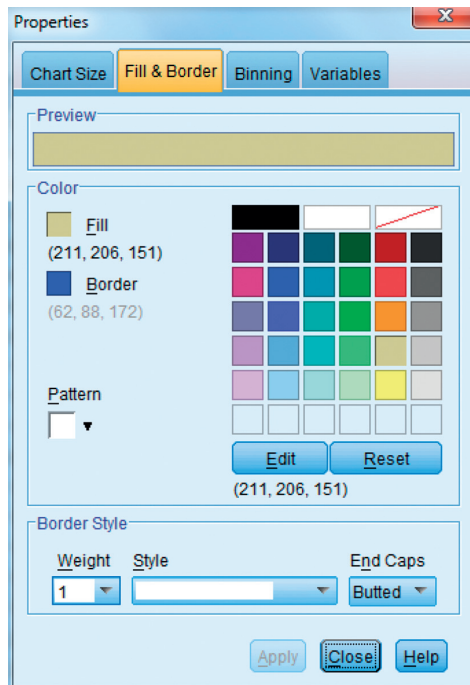
Εικόνα 19: Άλλες επιλογές του Chart Editor

Για την αναίρεση της επιλογής, πατήστε απλώς το πλήκτρο Esc. Για την επιλογή μίας μόνο στήλης, κάντε κλικ στη συγκεκριμένη στήλη. Για την επιλογή περισσότερων από μίας, αλλά όχι για το σύνολο των στηλών, κάντε κλικ στις συγκεκριμένες στήλες, κρατώντας πατημένο το πλήκτρο Ctrl. Ολοκληρώνοντας τις επιλογές ως προς το πλήθος των στηλών, επιλέξτε από το menu Edit του Chart Editor την εντολή Properties (Εικόνα 20).



Εικόνα 20: Η εντολή Properties από το σχετικό μενού του Chart Editor

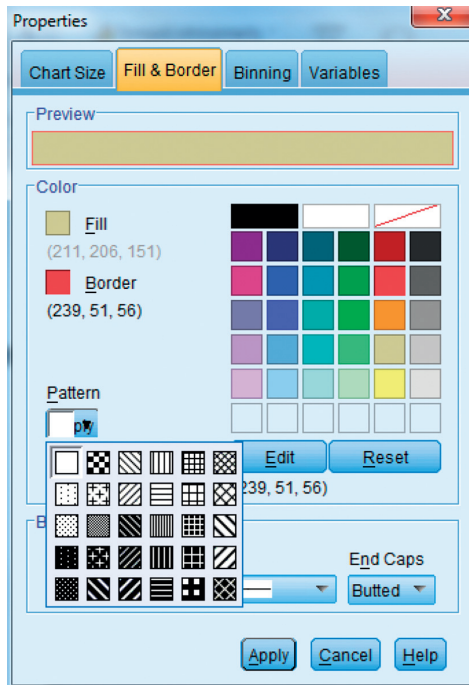
Ένα νέο πλαίσιο διαλόγου αναδύεται, το οποίο περιλαμβάνει μια σειρά από κάρτες (tabs), παρέχοντας δυνατότητες τροποποίησης επιλεγμένων στοιχείων του γραφήματος (Εικόνα 21).



Εικόνα 21: Αλλαγές χρωμάτων με την εντολή Properties

Έτσι, η επιλογή Fill & Border επηρεάζει τις ιδιότητες των στοιχείων του γραφήματος ως προς το χρώμα και τα περιγράμματα, η επιλογή Chart Size επηρεάζει το μέγεθος του γραφήματος κ.τ.ό.

Σημειώστε ότι, προκειμένου να αλλάξετε τη μορφοποίηση της κλίμακας βαθμολογίας σε κάποιον άξονα, θα πρέπει να επιλέξετε τους αριθμούς στον άξονα αναφοράς και από το Menu Edit να επιλέξετε την εντολή Properties (Εικόνα 22).

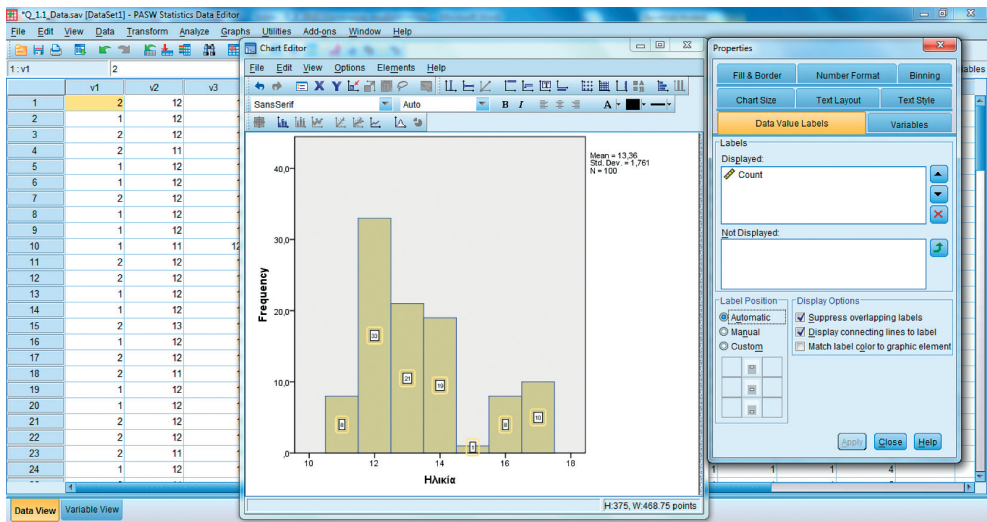
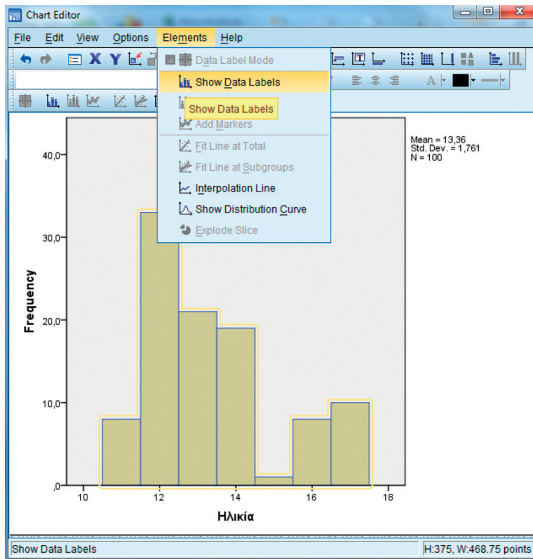


**Εικόνα 22:** Αλλαγές στα περιγράμματα με την εντολή Properties

Στην καρτέλα Number Format σάς παρέχεται η δυνατότητα τροποποίησης επιμέρους στοιχείων των αξόνων.



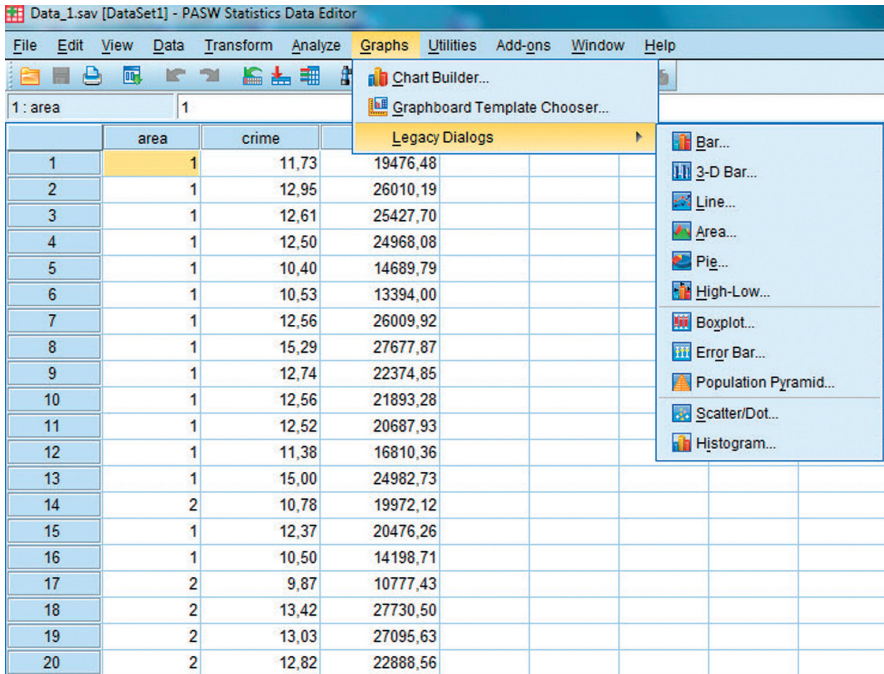
Προκειμένου να εμφανίσετε τις αριθμητικές τιμές στην επιφάνεια ενός γραφήματος, επιλέξτε (έχοντας φυσικά ενεργό τον Chart Editor) την εντολή Elements και Show Data Labels (Εικόνες 23-24).



Εικόνες 23-24: Η επιλογή Data Value Labels

Η τροποποίηση της αρχικής μορφής των τιμών επιτυγχάνεται μέσω της επιλογής Data Value Labels της επιλογής Properties (Εικόνα 24).

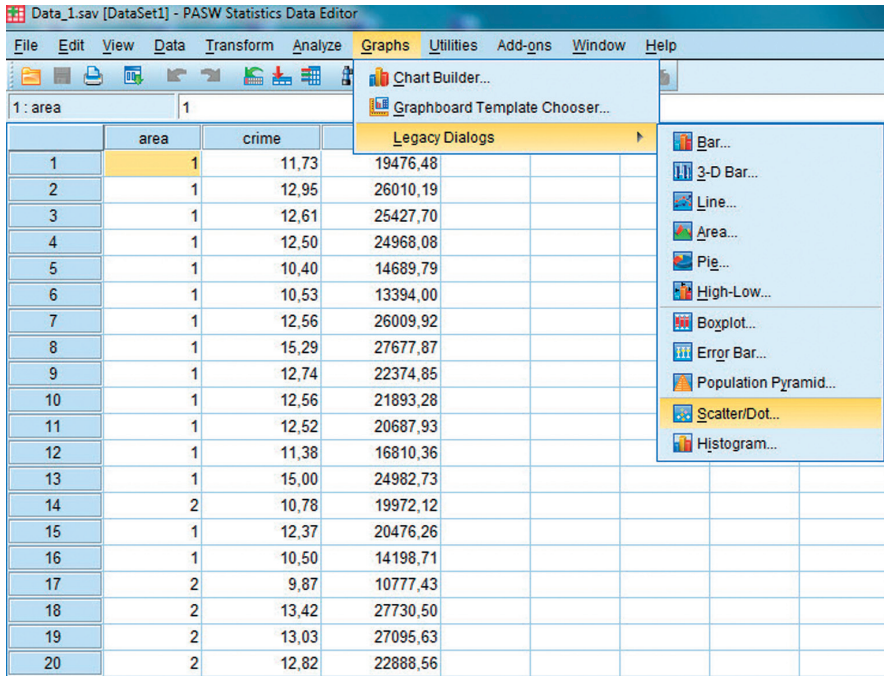
Εκτός από όσα μέχρι στιγμής προτάθηκαν, το PASW παρέχει τη δυνατότητα δημιουργίας αλληλεπιδραστικών γραφημάτων. Από το μενού Graphs επιλέξτε την εντολή Legacy Dialogs (Εικόνα 25).



Εικόνα 25: Η επιλογή Legacy Dialogs

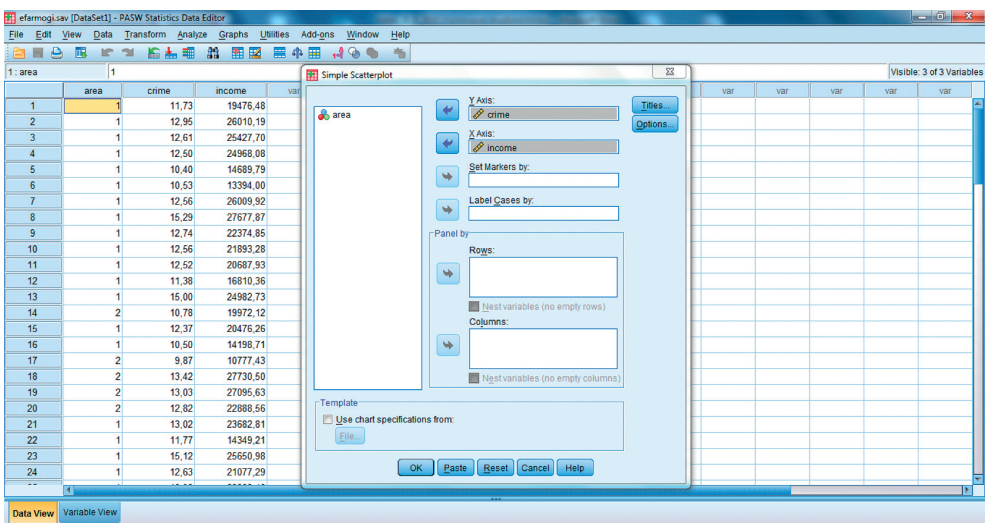
Από εκεί μπορείτε να επιλέξετε τον τύπο του γραφήματος που επιθυμείτε. Στα γραφήματα (ραβδόγραμμα, ιστόγραμμα) το όνομα της μεταβλητής (διακριτής ή συνεχούς, αντίστοιχα) εισάγεται στον οριζόντιο άξονα. Στον κάθετο άξονα τοποθετείτε την επιλογή \$count ή % για συχνότητα ή ποσοστά, αντίστοιχα. Αντίθετα για την παραγωγή θηκογραμμάτων, τοποθετήστε τη μεταβλητή του θηκογράμματος στον κάθετο άξονα. Προφανώς για το διάγραμμα σκεδασμού οι δύο μεταβλητές εισάγονται στους αντίστοιχους άξονες. Ας δούμε ένα σχετικό παράδειγμα:

Έστω ότι μας ενδιαφέρει η δημιουργία ενός διαγράμματος σκεδασμού μεταξύ των μεταβλητών crime και income της προηγούμενης εφαρμογής. Για το σκοπό αυτό επιλέγουμε Scatterplot από το μενού Graphs και την επιλογή Legacy Dialogs (Εικόνα 26).

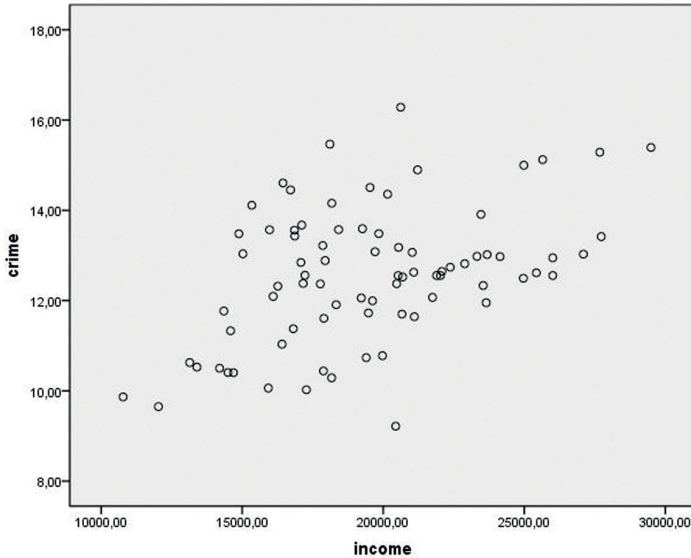


Εικόνα 26: Η δημιουργία Scatterplot

Στη συνέχεια εισάγουμε τις μεταβλητές στους δύο άξονες και πατάμε OK (Εικόνα 27). Τα αποτελέσματα παρουσιάζονται στο Σχήμα 7. Αναλυτικότερη ερμηνεία των διαγραμμάτων σκεδασμού θα παρουσιασθεί στα Κεφάλαια 4-5.

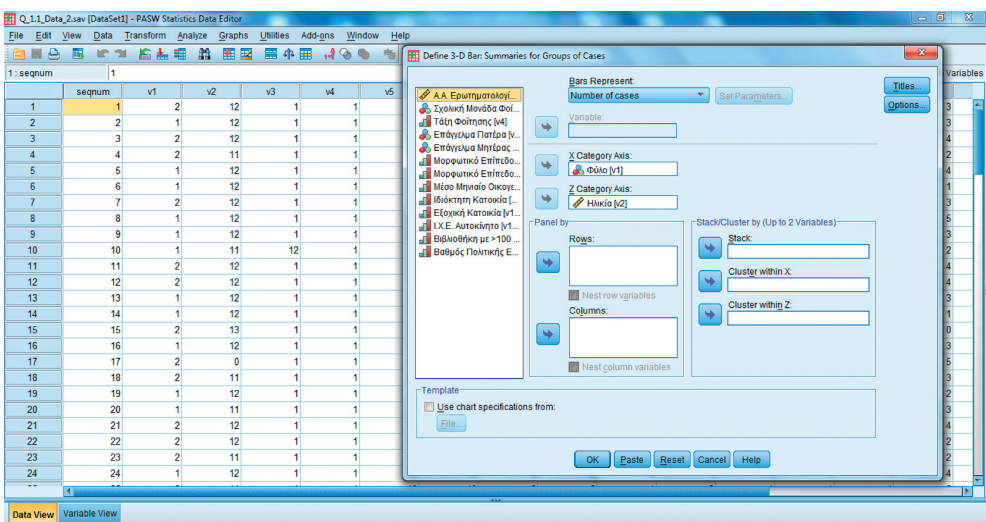


Εικόνα 27: Επιλογές για Scatterplot

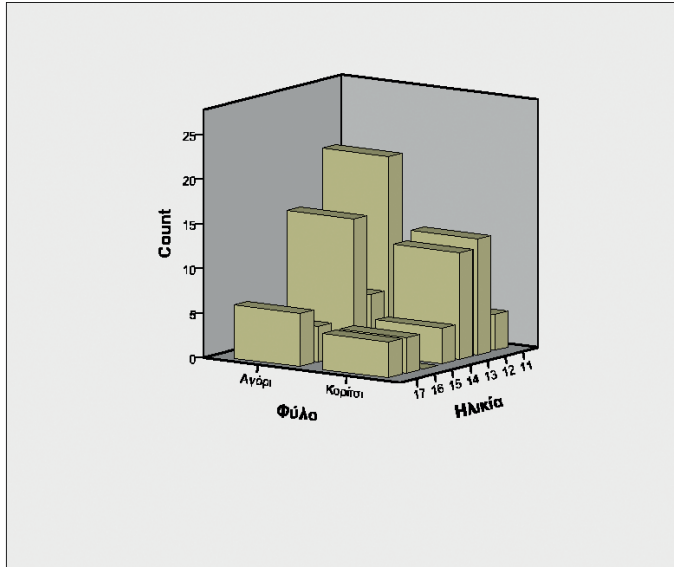


Σχήμα 7: Το παραγόμενο Scatterplot

Με τη βοήθεια των διαθέσιμων επιλογών είναι επίσης δυνατή η δημιουργία τρισδιάστατων γραφημάτων. Στην Εικόνα 28 βλέπετε τις αναγκαίες επιλογές για τη δημιουργία ενός τρισδιάστατου γραφήματος συνάρτησης των μεταβλητών Ηλικία [V2] και Βαθμός Πολιτικής Ενημέρωσης [V10] ως προς τις συχνότητες των τιμών τους, και στην Εικόνα 29 το αποτέλεσμα των επιλογών ως γράφημα στον Chart Editor.



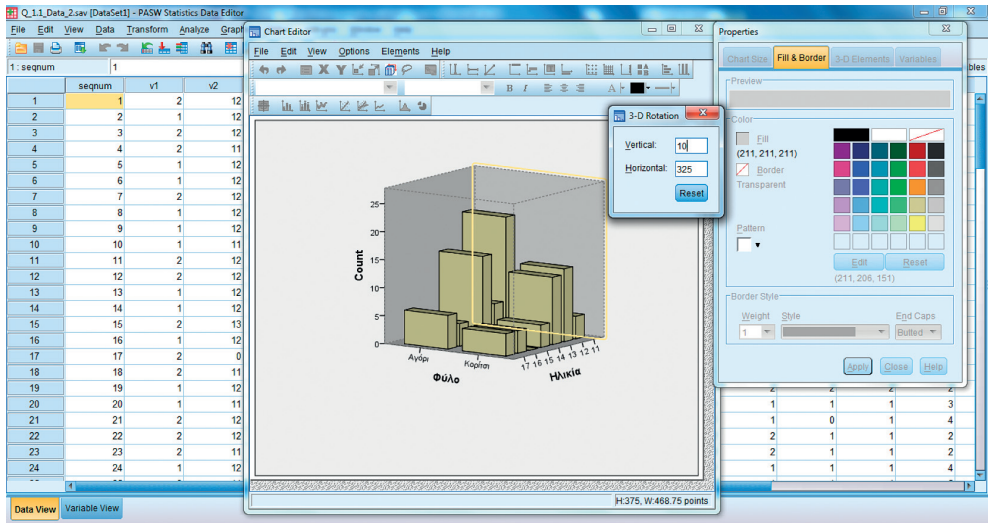
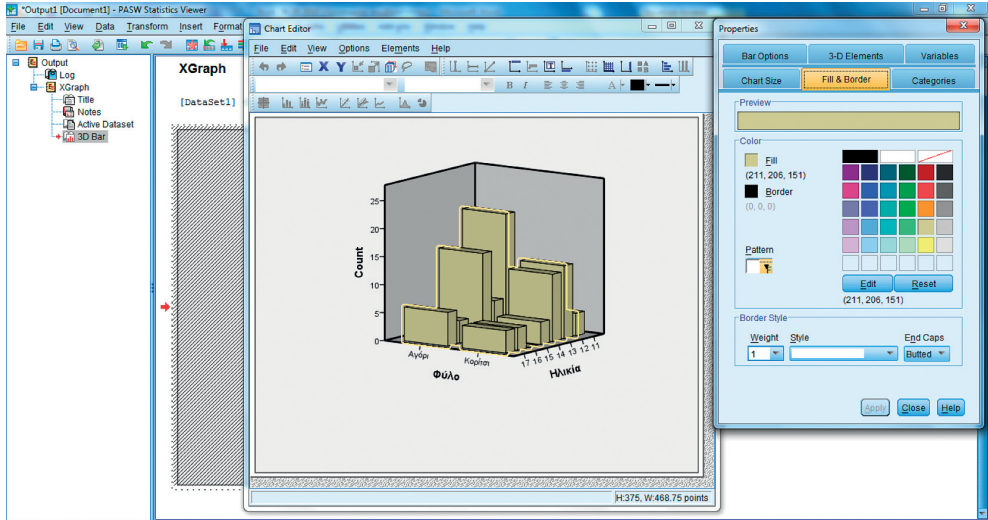
Εικόνα 28: Δημιουργία τρισδιάστατων γραφημάτων



**Εικόνα 29:** Το παραγόμενο τρισδιάστατο γράφημα

Μια ακόμη δυνατότητα δημιουργίας γραφημάτων παρέχεται μέσω της επιλογής Legacy Dialogs, διαθέσιμη επίσης στο Menu Graphs (Εικόνες 30-31). Ωστόσο, οι επιλογές και οι δυνατότητες ανάπτυξης του γραφήματος είναι περιορισμένες σε σχέση με τις δύο προηγούμενες διαδικασίες (μέσω του Chart Builder και της εντολής Interactive).

Η διαμόρφωση του γραφήματος είναι εφικτή μέσω του Chart Editor (Εικόνες 30-31).



Εικόνες 30-31: Επιλογές του Chart Editor για τη διαμόρφωση τρισδιάστατου γραφήματος

