ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΟΣ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

«ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ»

ΘΕΟΔΟΣΑΚΗΣ ΚΑΡΣΤΕΝ ΜΗΝΑΣ ΓΚΕΡΧΑΡΝΤ

ΜΑΓΚΑΦΑΣ ΑΝΑΣΤΑΣΙΟΣ

ΡΥΣΣΑΚΗΣ ΦΑΝΟΥΡΙΟΣ

Εποπτεύουσα καθηγήτρια: Χρυσάνθη Παπαθανασοπούλου

ΠΑΤΡΑ 2014

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΟΣ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

«ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ»

ΘΕΟΔΟΣΑΚΗΣ ΚΑΡΣΤΕΝ ΜΗΝΑΣ ΓΚΕΡΧΑΡΝΤ Α.Μ 11009

ΜΑΓΚΑΦΑΣ ΑΝΑΣΤΑΣΙΟΣ Α.Μ 981

ΡΥΣΣΑΚΗΣ ΦΑΝΟΥΡΙΟΣ Α.Μ. 11137

Εποπτεύουσα καθηγήτρια: Χρυσάνθη Παπαθανασοπούλου

ΠΑΤΡΑ 2014

ΠΡΟΛΟΓΟΣ

Η συγγραφή της πτυχιακής εργασίας πραγματοποιείται στο πλαίσιο του Προπτυχιακού Προγράμματος Σπουδών του τμήματος Διοίκησης Επιχειρήσεων Πάτρας του Τεχνολογικού Ιδρύματος Δυτικής Ελλάδας.

Το θέμα που πραγματεύεται η παρούσα εργασία εστιάζεται στην *Περιγραφική Στατιστική και την ανάλυση των δεδομένων*. Αυτός είναι και ο *αντικειμενικός σκοπός της πτυχιακής εργασίας*, η παρουσίαση και η περιγραφή δηλαδή των «εργαλείων», είτε αυτά είναι πίνακες είτε είναι διαγράμματα του συγκεκριμένου τομέα της Στατιστικής.

Συμπερασματικά, θα πραγματοποιηθεί μια *θεωρητική επισκόπηση* της Περιγραφικής Στατιστικής και συμπληρωματικά σε κάθε θεματική ενότητα, όπου αυτό είναι εφικτό, θα υλοποιείται και μια *πρακτική και υπολογιστική εφαρμογή* στο στατιστικό λογισμικό SPSS με παραδείγματα ανάλυσης δεδομένων.

ΠΕΡΙΛΗΨΗ

Ως **Στατιστική** ορίζεται η *επιστήμη που ασχολείται με την συλλογή, την ανάλυση και την ερμηνεία δεδομένων*. Παρόλο που ο συγκεκριμένος ορισμός αντιπροσωπεύει ένα σημαντικό μέρος των δραστηριοτήτων της Στατιστικής, εντούτοις δεν αποτελεί το αποκλειστικό αντικείμενο αυτής της επιστήμης. Μπορούμε να πούμε πιο σωστά πως ο παραπάνω ορισμός αναφέρεται στο κομμάτι αυτό που ονομάζεται **Περιγραφική Στατιστική**. Υπάρχει και μια άλλη διάσταση της Στατιστικής που έχει ως κύρια ενασχόληση την **συμπερασματολογία**. Για αυτήν την πλευρά της Στατιστικής ένας ορισμός που θα μπορούσε να δοθεί είναι ο ακόλουθος: «*Στατιστική είναι η προσπάθεια εξαγωγής συμπερασμάτων κάτω από συνθήκες αβεβαιότητας*».

Κάλλιστα κάποιος θα μπορούσε να προβάλλει τον ισχυρισμό ότι η Στατιστική Συμπερασματολογία είναι πιο σημαντική συγκριτικά με την Περιγραφική Στατιστική και αυτό εξαιτίας του ότι χρειάζεται πιο πολύπλοκα εργαλεία για την ανάπτυξη της. Είναι κοινός τόπος τα τελευταία έτη, ότι η Περιγραφική Στατιστική βρίσκει ολοένα και περισσότερες εφαρμογές και μια αιτία για αυτήν την εξέλιξη αποτελεί το γεγονός πως σε όλες τις επιστήμες υπάρχει η ανάγκη για ποσοτική προσέγγιση των διαφόρων εννοιών και μεθόδων, κάτι το οποίο υλοποιείται με την συγκέντρωση, την ανάλυση και την παρουσίαση των δεδομένων που υπάρχουν.

Κατά συνέπεια, η Περιγραφική Στατιστική έχει βαρύνουσα σημασία καθώς είναι ουσιαστικά η πρώτη προσπάθεια επεξεργασίας των υπό εξέταση δεδομένων-στοιχείων και ως εκ τούτου λάθη και παραλείψεις θα επηρεάσουν αρνητικά το τελικό αποτέλεσμα και θα οδηγήσουν σε λανθασμένα συμπεράσματα. Άρα η περιγραφή και η ανάλυση των δεδομένων είναι καθοριστικής σημασίας για την εξαγωγή ορθών αποτελεσμάτων.

Σκοπός της παρούσας εργασίας είναι η θεωρητική και πρακτική επισκόπηση της Περιγραφικής Στατιστικής καθώς και η υλοποίηση παραδειγμάτων ανάλυσης δεδομένων στο στατιστικό πακέτο SPSS.

Πίνακας Περιεχομένων

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

1.1 Ορισμός Στατιστικής

Όταν ένας απλός άνθρωπος ακούσει στην καθημερινότητα του την λέξη «Στατιστική» συνήθως το μυαλό του πηγαίνει σε εκλογές, δημοσκοπήσεις, exit polls και γενικά διάφορες μετρήσεις που συσχετίζονται με συγκεκριμένες εκφάνσεις του δημόσιου ή ιδιωτικού βίου. Γενικότερα, η Στατιστική είναι μια έννοια που παραπέμπει σε μια συστηματική απαρίθμηση και παρουσίαση αριθμητικών δεδομένων και στοιχείων, τα οποία έχουν προκύψει από διάφορες μετρήσεις ή παρατηρήσεις, οι οποίες με την σειρά τους αναφέρονται σε ορισμένο αντικείμενο ή γεγονός. Έχουμε επομένως την Γεωργική Στατιστική που προφανώς αναφέρεται σε στοιχεία που αφορούν πλευρές (παραγωγή, εισαγωγές, εξαγωγές κ.τ.λ.) της γεωργίας μιας χώρας, την Στατιστική Επιχειρήσεων ή Στατιστική Εργατικού Δυναμικού που εξετάζει τις πτυχές που αναπτύσσονται γύρω από το εργατικό δυναμικό και την επιχείρηση γενικότερα και άλλες στατιστικές κατηγορίες (Δημογραφική Στατιστική, Εγκληματολογική Στατιστική κ.α.).

Αναφορικά με την επιστημονική ορολογία, η Στατιστική δεν είναι απλώς ένα σύνολο μετρήσεων αλλά μια ευρύτερη έννοια που ασχολείται με τις επιστημονικές μεθόδους συλλογής, οργάνωσης, παρουσίασης και ανάλυσης των αριθμητικών εκείνων στοιχείων που αναφέρονται σε χαρακτηριστικές ιδιότητες διαφόρων οικονομικών, κοινωνικών, δημογραφικών, φυσικών κ.λ.π. φαινομένων και έχει ως σκοπό την συστηματική μελέτη αυτών των στοιχείων για την κατάληξη σε γενικά συμπεράσματα, τα οποία είναι χρήσιμα στην διαδικασία της λήψης ορθών αποφάσεων (Κιόχος Π., 1993).

Ακολούθως αναφέρουμε τα βασικά στάδια που έπονται από τον ορισμό της Στατιστικής:

i. Η συγκέντρωση των στατιστικών στοιχείων που είναι απαραίτητη για την μελέτη του προβλήματος που επιθυμούμε να εξετάσουμε.

ii. Η μεθοδική επεξεργασία και παρουσίαση των στατιστικών στοιχείων σε μορφή αριθμητικών πινάκων και γραφικών παραστάσεων.
iii. Η ανάλυση των στοιχείων αυτών και κατόπιν η εξαγωγή χρήσιμων συμπερασμάτων προκειμένου να ληφθούν σωστές αποφάσεις.

Ενώ αρχικά η Στατιστική ασχολείτο μόνο με την παράθεση τεραστίων αριθμητικών πινάκων, σήμερα μια στατιστική έρευνα χωρίζεται στα εξής στάδια:

• Συλλογή δεδομένων.
• Έλεγχος δεδομένων (καταμέτρηση, διάταξη, διόρθωση λαθών, συμπλήρωση ελλιπών στοιχείων).
• Παρουσίαση δεδομένων (πίνακες, διαγράμματα).
• Επεξεργασία των δεδομένων με σκοπό την εξαγωγή χρήσιμων συμπερασμάτων.

Ο κλάδος της Στατιστικής που ασχολείται με τα δύο πρώτα στάδια λέγεται ***σχεδιασμός πειραμάτων (experimental design***). Ο κλάδος που ασχολείται με το τρίτο στάδιο είναι η ***περιγραφική στατιστική (descriptive statistics***). Πρέπει να σημειώσουμε πως η Περιγραφική Στατιστική χρησιμοποιεί αριθμητικά μέτρα όπως η μέση τιμή, η διάμεσος, η διακύμανση καθώς και διαγραμματική απεικόνιση για να περιγράψει κάποια χαρακτηριστικά των δεδομένων, χωρίς όμως να μπορούν να γενικευτούν αυτά για ολόκληρο τον πληθυσμό. Τέλος, η ***επαγωγική στατιστική*** περιλαμβάνει τις μεθόδους με τις οποίες τα χαρακτηριστικά ενός μεγάλου συνόλου δεδομένων (πληθυσμός) προσεγγίζονται από την μελέτη των χαρακτηριστικών ενός μικρού υποσυνόλου των δεδομένων (δείγμα).

1.2 Ιστορική αναδρομή

Η προέλευση της λέξης «**Στατιστική**» ίσως να πηγάζει από την αρχαία ελληνική λέξη *στατίζω* που σημαίνει (τοποθετώ, ταξινομώ, συμπεραίνω) ή από τη λατινική λέξη «*status*» που σημαίνει (πολιτεία, κράτος).

Η πιο αρχαία πιθανώς συλλογή στοιχείων θεωρείται η απογραφή πληθυσμού που έγινε το 2238 π.Χ. στην Κίνα από τον αυτοκράτορα Yao. Πιο μετά στοιχειώδεις απογραφές φαίνεται να έχουν πραγματοποιηθεί τόσο από τους Αιγυπτίους όσο και από τους Πέρσες. Στην αρχαιότητα, η συγκέντρωση στατιστικών στοιχείων είχε ως στόχο τον εντοπισμό των πολιτών που είχαν υποχρέωση να υπηρετήσουν στο στρατό ή να πληρώσουν φόρο.

Μάλιστα, η συστηματική συλλογή στοιχείων απασχόλησε και τους κατοίκους διαφόρων χωρών της Ευρώπης. Ο μεγάλος αριθμός θανάτων που οφειλόταν σε πολέμους, λιμοκτονίες, επιδημικές ασθένειες και διάφορες άλλες αιτίες είχε επιπτώσεις στον πληθυσμό και στην οικονομία. Για παράδειγμα, το 1620 ο Άγγλος Graunt σε μια δειγματοληπτική έρευνα που έκανε σε οικογένειες του Λονδίνου διαπίστωσε ότι σε κάθε 88 άτομα υπήρχαν 3 θάνατοι. Χρησιμοποιώντας τους επίσημους καταλόγους, οι οποίοι έδιναν 13.200 θανάτους το 1620, έκανε την εκτίμηση ότι ο πληθυσμός του Λονδίνου το έτος αυτό κυμαινόταν περίπου στα 387.200 άτομα.

Μεταξύ 16ου και 19ου αιώνα, σημειώθηκε μια αλματώδης ανάπτυξη του εμπορίου που ώθησε τις ηγεσίες των διαφόρων κρατών στη μελέτη οικονομικών δεδομένων, όπως την παραγωγικότητα των βιομηχανιών, το εξαγωγικό εμπόριο, κ.τ.λ.

Εν γένει, μέχρι τα τέλη του 18^{ου} αιώνα, η Στατιστική έχει περιγραφικό χαρακτήρα και το βασικό θέμα ενασχόλησης της είναι η ***Δημογραφία***. Αργότερα, θα ξεφύγει από τον περιγραφικό χαρακτήρα και θα αναπτυχθεί ένας νέος κλάδος, αυτός του ***Λογισμού των Πιθανοτήτων***, με κύριους εκφραστές του Γάλλους Pascal και Fermat.

Σήμερα η Στατιστική γνωρίζει μια ταχύτατη εξέλιξη, απόρροια κυρίως της τεχνολογικής εξέλιξης και εισαγωγής των ηλεκτρονικών υπολογιστών και επιπρόσθετα διαθέτει μια ευρεία ποικιλία πεδίων εφαρμογής.

1.3 Κράτος και Στατιστική

Σε κάθε χώρα έχουν δημιουργηθεί αυτοτελείς στατιστικοί οργανισμοί με σκοπό τον αποτελεσματικό συντονισμό όλων των στατιστικών εργασιών. Ο αντίστοιχος οργανισμός στην Ελλάδα είναι η ***Ε.Σ.Υ.Ε. (Εθνική Στατιστική Υπηρεσία Ελλάδο*ς**) (http://www.statistics.gr/). Οι στατιστικές που πραγματοποιεί η Ε.Σ.Υ.Ε. είναι μηνιαίες, τριμηνιαίες, ετήσιες, ανά 5ετία και ανά 10ετία, και καλύπτουν όλους σχεδόν τους τομείς δραστηριότητας. Πληθυσμιακά στοιχεία, στοιχεία απασχόλησης και ανεργίας, στοιχεία που αφορούν την υγεία, την κοινωνική ασφάλιση, την παιδεία, την παραγωγική διαδικασία, τις τιμές, το εθνικό εισόδημα κ.τ.λ.

ΚΕΦΑΛΑΙΟ 2: ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΣΤΑΤΙΣΤΙΚΗΣ

2.1 Στατιστική Μονάδα- Στατιστικός Πληθυσμός

Η στατιστική μονάδα είναι δυνατόν να είναι ένα αντικείμενο, ένα άτομο, μια εταιρεία, ένα ίδρυμα ή κάποιο γεγονός (εκλογική αναμέτρηση) και γενικά είναι αυτό από το οποίο λαμβάνουμε τις πληροφορίες που επιθυμούμε να επεξεργαστούμε και να αναλύσουμε στατιστικά.

Οι στατιστικές μονάδες μπορεί να είναι *απλές* (ένα άτομο, ένα αντικείμενο, μια μέρα), είναι όμως δυνατό να είναι και *σύνθετες* και να αποτελούνται από περισσότερα αντικείμενα ή πρόσωπα, όπως για παράδειγμα η οικογένεια, η μηνιαία ή ετήσια παραγωγή μιας βιομηχανίας κ.λ.π.

Το σύνολο των στατιστικών μονάδων, των οποίων επιθυμούμε τη μελέτη ενός ή περισσοτέρων συγκεκριμένων χαρακτηριστικών, ονομάζεται ***πληθυσμός ή στατιστικός πληθυσμός***.

Ένα από τα βασικά στοιχεία που πρέπει να οριστούν για τον πληθυσμό είναι τα *όριά του*, δηλαδή ποιες είναι ακριβώς οι στατιστικές μονάδες του πληθυσμού που θα μελετηθούν. Για παράδειγμα, αν θελήσουμε να μελετήσουμε ορισμένα χαρακτηριστικά των κατοίκων της Ραφήνας, δεν είναι συνετό να διεξάγουμε την έρευνα στο λιμάνι μια Παρασκευή απόγευμα ή Σάββατο πρωί, ή ένα Σαββατοκύριακο τους θερινούς μήνες. Δεν αρκεί, λοιπόν μόνο ο προσδιορισμός των γεωγραφικών ορίων μιας πόλης.

Ο στατιστικός πληθυσμός μπορεί να είναι *άπειρος*, όπως η παραγωγή ενός προϊόντος, οι γεννήσεις βρεφών σε μία πόλη ή *πεπερασμένος*, όπως οι αφίξεις και αναχωρήσεις των αεροσκαφών μια συγκεκριμένη μέρα στο αεροδρόμιο Ελ. Βενιζέλος.

2.2 Στατιστική Μεταβλητή

Μια μεταβλητή είναι ένας μαθηματικός και φυσικός όρος, που βρίσκει εφαρμογή σε κοινωνικούς, ψυχολογικούς και άλλους τομείς. Πιο αναλυτικά, ως ***μεταβλητή*** ορίζεται κάτι το οποίο μεταβάλλεται και παίρνει διάφορες τιμές. Όμως πιο συγκεκριμένα, ο όρος μεταβλητή στον ερευνητικό τομέα, ορίζεται ως το κάθε φυσικό μέγεθος που μπορεί να μετρηθεί ή να σημανθεί. Αυτό το γεγονός συμβαίνει και στις λεγόμενες συμπεριφορικές επιστήμες, όπου μια μεταβλητή εμπεριέχει όλους εκείνους τους παράγοντες, οι οποίοι ορίζουν τις εκάστοτε συμπεριφορές και οι οποίες διαφοροποιούνται από άνθρωπο σε άνθρωπο, καθώς και σε ομάδες υπολογίζοντας κυρίως τις διάφορες τιμές έντασης.

Εναλλακτικά μπορούμε να ορίσουμε τις μεταβλητές ως τις χαρακτηριστικές ιδιότητες των στατιστικών μονάδων ενός πληθυσμού με την μελέτη των οποίων ασχολείται η Στατιστική. Οι αριθμοί που αντιπροσωπεύουν τις διάφορες καταστάσεις μιας μεταβλητής καλούνται *τιμές της μεταβλητής*. Όσον αφορά τον συμβολισμό μιας μεταβλητής, αυτές συμβολίζονται συνήθως με ένα από τα κεφαλαία γράμματα X, Y, Z ενώ οι τιμές της με μικρά γράμματα όπως παραδείγματος χάριν $x_1, x_2, \ldots, x_k$.

2.3 Είδη Στατιστικών Μεταβλητών

1. ***Ποσοτικές (quantitative***) είναι οι μεταβλητές που δύναται να επιδέχονται αριθμητική μέτρηση.

Οι ποσοτικές μεταβλητές παίρνουν αριθμητικές «τιμές» και εκφράζονται με μια μονάδα μέτρησης. Διακρίνονται σε μεταβλητές ***διαστήματος (interval***) και σε μεταβλητές ***αναλογίας (ratio***).

Οι ποσοτικές μεταβλητές διακρίνονται σε δυο ακόμα κατηγορίες:

Α. ***Ασυνεχείς ή Διακριτές (Discrete Variables***) είναι εκείνες που λαμβάνουν ακέραιες τιμές (αριθμός λευκών ή ερυθρών αιμοσφαιρίων, αριθμός υπαλλήλων ενός λογιστηρίου, αριθμός παιδιών μιας οικογένειας, αριθμός ραδιενεργών κρούσεων, αριθμός ελαττωματικών προϊόντων) ή διαφορετικά πεπερασμένο πλήθος τιμών.

Β. ***Συνεχείς (Continuous Variables***) είναι εκείνες που μπορούν να λάβουν όλες τις τιμές ενός διαστήματος πραγματικών αριθμών (βάρος, ύψος, εισόδημα, ηλικία, ταχύτητα, θερμοκρασία κ.α.).

2. ***Ποιοτικές (qualitative***) είναι οι μεταβλητές που δεν επιδέχονται μέτρηση και οι τιμές τους εκφράζονται με λέξεις. Παραδείγματα ποιοτικών μεταβλητών είναι η οικογενειακή κατάσταση, το φύλο, το μορφωτικό επίπεδο κ.α.

2.4 Κλίμακες Μέτρησης

Ευρέως χρησιμοποιούνται οι εξής τέσσερις κλίμακες: ***κατηγορίας, διάταξης, διαστήματος και αναλογίας***. *Οι δυο πρώτες κλίμακες μέτρησης αφορούν τις ποιοτικές μεταβλητές ενώ οι δυο τελευταίες τις ποσοτικές*.

2.4.1 Κλίμακα Κατηγορίας ή Ονομαστική (Nominal)

Στην Ονομαστική κλίμακα ή Κατηγορίας (nominal) υπάγονται οι μεταβλητές των οποίων το σύνολο των τιμών δεν έχει καμία ιδιότητα. Για τη μεταβλητή αυτή, μοναδική σημασία έχουν οι διαφορετικές τιμές (το πλήθος των κατηγοριών της) που μπορεί να πάρει. Η μοναδική σχέση που μπορεί να προσδιοριστεί μεταξύ των κατηγοριών αυτών είναι απλά η ύπαρξη διαφοράς.

2.4.2 Κλίμακα Διάταξης (Ordinal)

Πρόκειται για τις μεταβλητές που για το σύνολο τιμών τους μπορούμε να ορίσουμε μια σχέση **διάταξης**, δηλαδή δύνανται να τοποθετηθούν στη σειρά. Η διάταξη μπορεί να είναι αύξουσα ή φθίνουσα. Οι ίσες διαφορές μεταξύ των τιμών μιας τέτοιας μεταβλητής δεν συνεπάγονται και ίσες διαφορές για το χαρακτηριστικό που μετράει η μεταβλητή. Δεν υπάρχει δηλαδή αντιστοίχιση σε υποδιαιρέσεις ή πολλαπλάσια κάποιας μονάδας. Η διάταξη επομένως, το μόνο που εξασφαλίζει είναι τον προσδιορισμό της μεγαλύτερης, καλύτερης, προτιμότερης κατηγορίας αλλά όχι πόσο μεγαλύτερη, καλύτερη, προτιμότερη είναι σε σχέση με κάποια από τις υπόλοιπες.

2.4.3 Κλίμακα Διαστήματος (Interval)

Περιλαμβάνονται οι μεταβλητές των οποίων οι ίσες διαφορές μεταξύ των τιμών τους συνεπάγονται και ίσες διαφορές για το χαρακτηριστικό που μετράει η μεταβλητή (π.χ. ηλικία, θερμοκρασία). *Η κλίμακα αυτή δεν επιτρέπει μόνο την ιεράρχηση των υποκειμένων αλλά προσδιορίζει επίσης και την ακριβή διαφορά τους*. Η απόσταση μεταξύ δυο οποιονδήποτε διαδοχικών τιμών της μεταβλητής αυτής, είναι ίση με την απόσταση δυο άλλων τυχαίων διαδοχικών τιμών της. Παράλληλα, δεν έχει νόημα ο υπολογισμός αναλογιών. Αυτό που χαρακτηρίζει βασικά τις μεταβλητές αυτής της διάταξης είναι ο αυθαίρετος ορισμός του μηδενός, που δεν υποδηλώνει παντελή έλλειψη του μετρήσιμου χαρακτηριστικού.

2.4.4 Κλίμακα Αναλογίας (Ratio)

Αφορά τις μεταβλητές των οποίων οι τιμές αντιστοιχούν αναλογικά στην ποσότητα του χαρακτηριστικού που μετρούν. Εδώ το μηδέν ανήκει στο διάστημα τιμών της μεταβλητής και δηλώνει την πλήρη απουσία. Επιπλέον για τις τιμές των μεταβλητών αυτών έχει έννοια ο υπολογισμός των αναλογιών. Η ταχύτητα, το ύψος, ο ημερήσιος τζίρος μιας εταιρείας είναι μερικά παραδείγματα μεταβλητών με κλίμακα αναλογίας.

ΚΕΦΑΛΑΙΟ 3: ΣΤΑΤΙΣΤΙΚΗ ΜΟΝΑΔΑ – ΣΤΑΤΙΣΤΙΚΟΣ ΠΛΗΘΥΣΜΟΣ

3.1 ΜΕΘΟΔΟΙ ΣΥΛΛΟΓΗΣ ΣΤΟΙΧΕΙΩΝ- ΑΠΟΓΡΑΦΗ

Οι διάφορες μέθοδοι συλλογής στατιστικών στοιχείων διακρίνονται σε δύο μεγάλες κατηγορίες: τις ***απογραφές*** και τις ***δειγματοληπτικές έρευνες***.

Απογραφή (census) καλείται η διαδικασία με την οποία συλλέγονται οι παρατηρήσεις όλων των μονάδων ενός πληθυσμού σε μια συγκεκριμένη χρονική στιγμή.

Είναι δυνατόν να έχουμε τις:

• *Δημογραφικές απογραφές* στις οποίες συλλέγονται στοιχεία σχετικά με το φύλο, την ηλικία, το επάγγελμα, κ.τ.λ.

• *Οικονομικές απογραφές* στις οποίες συγκεντρώνονται στοιχεία σχετικά με την οικονομική κατάσταση.

• *Βιομηχανικές απογραφές*, στις οποίες συλλέγουμε πληροφορίες σχετικά με την οικονομική δραστηριότητα των βιομηχανιών, τον αριθμό των απασχολούμενων, το επίπεδο μηχανοργάνωσης, κ.τ.λ.

• *Γεωργικές απογραφές*, στις οποίες συγκεντρώνουμε στοιχεία που αφορούν τις εκτάσεις που καλλιεργούνται, το είδος της γεωργικής παραγωγής, τον αριθμό των γεωργικών μηχανημάτων, κ.τ.λ.

Τα ***μειονεκτήματα*** των απογραφών είναι:

• Το *μεγάλο κόστος*, καθώς χρειάζεται ειδική προεργασία και μεγάλο αριθμό απογραφέων.

• Την *μη ύπαρξη πολλών εξειδικευμένων ατόμων*, που έχει ως συνέπεια την συγκέντρωση εσφαλμένων στοιχείων τα οποία μπορεί να δώσουν λανθασμένη εικόνα της σύνθεσης του πληθυσμού.

• Την *μη επίκαιρη έκδοση των αποτελεσμάτων*, λόγω του μεγάλου όγκου των πληροφοριών.

3.2 ΜΕΘΟΔΟΙ ΣΥΛΛΟΓΗΣ ΣΤΟΙΧΕΙΩΝ- ΔΕΙΓΜΑΤΟΛΗΨΙΑ

Δειγματοληψία ή **Δειγματοληπτική Μέθοδος** καλείται η απογραφή ορισμένων συγκεκριμένων χαρακτηριστικών ενός τμήματος του πληθυσμού.

Το τμήμα του πληθυσμού που απογράφεται ονομάζεται **δείγμα**. Η δειγματοληψία πραγματοποιείται σε συγκεκριμένα στάδια ανάλογα με τη μέθοδο που επιλέγουμε.

Οι κανόνες και οι μέθοδοι συλλογής και ανάλυσης δεδομένων από πεπερασμένους πληθυσμούς συνιστούν την περιοχή της Στατιστικής που είναι γνωστή ως **Μέθοδοι Δειγματοληπτικών Ερευνών (Sample Survey Methods)**. Η θεωρητική τους βάση ονομάζεται **Θεωρία Δειγματοληψίας (Sampling Theory)**. Μια «καλή» δειγματοληπτική μέθοδος παρουσιάζει τα εξής πλεονεκτήματα:

- *Χαμηλότερο κόστος*: Οι πληροφορίες στα δεδομένα προέρχονται από ένα τμήμα του πληθυσμού.
- *Μεγαλύτερη ταχύτητα*: Η συλλογή και η επεξεργασία των δεδομένων είναι ταχύτερη.
- *Μεγαλύτερη ακρίβεια*: Πιο προσεκτική εποπτεία της διεξαγωγής της έρευνας και πιο προσεκτική επεξεργασία των αποτελεσμάτων είναι εφικτή.

Ο κύριος στόχος είναι ***η λήψη ενός δείγματος το οποίο να είναι αντιπροσωπευτικό του πληθυσμού*** *και το οποίο να οδηγεί σε εκτιμήσεις των χαρακτηριστικών του πληθυσμού με όσο το δυνατόν μεγαλύτερη ακρίβεια μπορούμε να επιτύχουμε για το κόστος ή για την προσπάθεια που είμαστε έτοιμοι να καταβάλλουμε.*

Κατά κανόνα, η έρευνα θα μπορούσε να βασισθεί στην επισκόπηση όλων των μελών ενός πεπερασμένου πληθυσμού. Μια τέτοια διαδικασία είναι η **απογραφή** που

αναφέραμε προηγουμένως. Η απογραφή δηλαδή, είναι μια δειγματοληπτική έρευνα με κάλυψη 100%. Το ενδιαφέρον μας όμως εστιάζεται σε πολύ χαμηλότερα επίπεδα κάλυψης, συχνά του ύψους του 1% ή 5%.

Η στοιχειωδέστερη μορφή δειγματοληψίας κατά πιθανότητα είναι η *απλή τυχαία δειγματοληψία (Simple Random Sampling).* Το σχήμα αυτό χρησιμοποιείται ευρύτατα, κυρίως λόγω της απλότητας του από την άποψη της στατιστικής συμπερασματολογίας. Στην *απλή τυχαία δειγματοληψία* κάθε μία από τις μονάδες του πληθυσμού έχει ίση πιθανότητα να επιλεγεί. Διευκρινίζεται ότι πρόκειται για *απλή τυχαία δειγματοληψία χωρίς επανάθεση (επανατοποθέτηση),* δηλαδή κάθε μονάδα απομακρύνεται από τον πληθυσμό μετά την επιλογή της στο τυχαίο δείγμα. Βασική προϋπόθεση για το σχηματισμό ενός δείγματος από έναν πληθυσμό είναι ο σαφής καθορισμός του πληθυσμού.

3.3 ΠΛΕΟΝΕΚΤΗΜΑΤΑ & ΜΕΙΟΝΕΚΤΗΜΑΤΑ ΤΗΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Αναφέρουμε επιγραμματικά μερικά από τα *πλεονεκτήματα* της Δειγματοληψίας:

- Μεγαλύτερη ταχύτητα πληροφοριών.
- Μεγαλύτερη ακρίβεια.
- Μεγαλύτερη ευχέρεια εφαρμογής.
- Χαμηλό κόστος.
- Ολοκληρωτική δύναμη εφαρμογής της γενικής απογραφής.

Συνοπτικά μερικά από τα *μειονεκτήματα* της Δειγματοληπτικής Μεθόδου αποτελούν:

- Αν οι μονάδες του πληθυσμού που εξετάζουμε εμφανίζονται σποραδικά, πρέπει να μελετήσουμε ένα σημαντικά μεγάλο δείγμα.
- Ο σχεδιασμός και η εκτέλεση της Δειγματοληψίας απαιτούν μεγάλη προσοχή και είναι επιτακτική η αυστηρή και πιστή εφαρμογή του θεωρητικής διαδικασίας που επιβάλλεται για την επιλογή του δείγματος και η στατιστική ανάλυση των αποτελεσμάτων.
- Η κακή σχεδίαση και εκτέλεση της Δειγματοληψίας.
- Η μη αντιπροσωπευτικότητα του δείγματος.
- Η μη κατάλληλη μέθοδος διεξαγωγής Δειγματοληψίας.
- Τα ανεπαρκή δεδομένα.
- Τα Δειγματοληπτικά Σφάλματα.

ΚΕΦΑΛΑΙΟ 4: ΠΑΡΟΥΣΙΑΣΗ ΣΤΑΤΙΣΤΙΚΩΝ ΣΤΟΙΧΕΙΩΝ

4.1 ΓΕΝΙΚΑ

Τα στατιστικά δεδομένα πρέπει να παρουσιάζονται με τρόπο απλό και σαφή, έτσι ώστε να είναι εύκολη η κατανόησή τους από τον κάθε ενδιαφερόμενο. Η παρουσίαση μπορεί να γίνει υπό μορφή:

Α. ***Πινάκων*** *που διακρίνονται στις εξής κατηγορίες:*

I. Στατιστικοί Πίνακες

II. Πίνακες Κατανομής Συχνοτήτων

Β. ***Γραφικών Παραστάσεων***

Γ. ***Εκθέσεων ή Αναφορών***

4.2 ΣΤΑΤΙΣΤΙΚΟΙ ΠΙΝΑΚΕΣ

Σε κάθε πίνακα, που έχει συνταχθεί σωστά, εκτός από το κύριο σώμα, που περιέχει διαχωρισμένα μέσα στις γραμμές και στήλες τα στατιστικά δεδομένα, παρατηρούνται και τα εξής ειδικότερα στοιχεία:

Α. Ο **τίτλος**, που γράφεται στο πάνω μέρος και πρέπει να δηλώνει με σαφήνεια και με περιληπτικό τρόπο το περιεχόμενο του πίνακα.

Β. Οι **επικεφαλίδες των στηλών (και γραμμών)**, που δείχνουν συνοπτικά τη φύση και τη μονάδα μετρήσεως των δεδομένων.

Γ. Η **πηγή** που γράφεται στο κάτω μέρος του πίνακα και δείχνει την προέλευση των δεδομένων.

Δ. Οι **υποσημειώσεις** που γράφονται στο κάτω μέρος του πίνακα και πριν από την πηγή, αν θεωρηθεί απαραίτητο να δοθούν κάποιες επεξηγήσεις.

Δύο είναι οι βασικοί τύποι των στατιστικών πινάκων:

- Πίνακες απλής εισόδου.
- Πίνακες διπλής εισόδου.

Οι ***πίνακες απλής εισόδου*** αναφέρονται στην μελέτη ενός φαινομένου υπό το πρίσμα ενός μόνο χαρακτηριστικού και χρησιμοποιούνται συχνά για συγκρίσεις και εξαγωγή συμπερασμάτων (παραδείγματος χάριν ο πληθυσμός της Ελλάδας το 1981). Ακολούθως παρατηρούμε έναν πίνακα απλής εισόδου:

Πίνακας 3.1

Εξέλιξη του ελληνικού πληθυσμού 1920-1981

Χρόνια	*Πληθυσμός*
1920	5.016.886
1928	6.204.684
1940	7.344.860
1951	7.637.801
1961	8.388.553
1971	8.768.641
1981	9.739.500
1991	10.269.907

Πηγή: Ε.Σ.Υ.Ε

Βλέπουμε πως αναγράφεται ο τίτλος (Εξέλιξη του ελληνικού πληθυσμού 1920-1981), οι επικεφαλίδες των γραμμών και των στηλών οι οποίες αντίστοιχα είναι «Χρόνια» και «Πληθυσμός», η πηγή άντλησης των δεδομένων που είναι η Ε.Σ.Υ.Ε. (αναγράφεται κάτω από τον πίνακα) και τέλος δεν υπάρχουν υποσημειώσεις στον συγκεκριμένο πίνακα.

Οι ***πίνακες διπλής εισόδου*** μας πληροφορούν σχετικά με έναν πληθυσμό όπου εξετάζουμε δύο χαρακτηριστικά του, ποσοτικά ή ποιοτικά (παραδείγματος χάριν βάρος και ύψος των μαθητών μιας τάξης ενός Λυκείου). Στη συνέχεια μπορούμε να δούμε έναν πίνακα διπλής εισόδου:

Πίνακας 3.11

Κατανομή της βαθμολογίας 4.961 μαθητών λυκείου θεωρητικής κατεύθυνσης στα Μαθηματικά και στην Έκθεση

Βαθμοί στα Μαθηματικά	*Βαθμοί στην Έκθεση*						
	≤ 3	4	5	6	7	≥ 8	Σύνολο
≤ 3	20	51	53	20	2	—	146
4	33	134	246	115	15	—	543
5	24	205	551	489	84	8	1.361
6	3	114	551	746	230	34	1.678
7	1	17	176	453	247	47	941
≥ 8	—	1	22	104	124	41	292
Σύνολο	81	522	1.599	1.927	702	130	4.961

Πηγή: T. Salvemini, Lezioni Di Statistica, Roma 1974

Ο τίτλος του πίνακα βρίσκεται πάνω από τον πίνακα και είναι «Κατανομή της βαθμολογίας 4.961 μαθητών λυκείου θεωρητικής κατεύθυνσης στα Μαθηματικά και στην Έκθεση». Έχουμε δύο μετρήσιμες ποσοτικές μεταβλητές, τους βαθμούς στα Μαθηματικά και τους βαθμούς στην Έκθεση. Η πρώτη μεταβλητή (Μαθηματικά) αντιστοιχεί στις γραμμές του πίνακα διπλής εισόδου και η δεύτερη μεταβλητή (Εκθεση) αντιστοιχείς τις στήλες. Η πηγή αναγράφεται κάτω από τον πίνακα και είναι «T. Salverimini, Lezioni Di Statistica, Roma 1974». Δεν υπάρχουν υποσημειώσεις.

4.3 ΠΙΝΑΚΕΣ ΚΑΤΑΝΟΜΗΣ ΣΥΧΝΟΤΗΤΩΝ

Ένας απλός αλλά αποτελεσματικός τρόπος για να περιγράψουμε την κατανομή των κατηγοριών (ή τιμών) μιας μεταβλητής, είναι να παρουσιάσουμε πόσο συχνά (πόσες φορές) εμφανίζεται η κάθε κατηγορία (ή τιμή) στα δεδομένα. Οι κατανομές που παρουσιάζονται κατά αυτόν τον τρόπο είναι γνωστές ως **κατανομές συχνοτήτων**. Κατανομές συχνοτήτων μπορούν να χρησιμοποιηθούν για την περιγραφή οποιασδήποτε μεταβλητής, που μετριέται σε οποιαδήποτε κλίμακα μέτρησης.

Στην περίπτωση που η κλίμακα μέτρησης είναι *ονομαστική*, η σειρά με την οποία κατατάσσονται και παρουσιάζονται οι κατηγορίες της μεταβλητής δεν είναι σημαντική. Όταν όμως, είναι τακτική ή υψηλότερη, οι κατηγορίες πρέπει να παρουσιάζονται με τη σειρά τους (αύξουσα ή φθίνουσα). Αυτό συνιστάται για τις συχνότητες όλων των ειδών και είναι υποχρεωτικό για τις αθροιστικές.

Οι κατανομές συχνοτήτων μπορούν να παρουσιαστούν υπό μορφή πινάκων ή διαγραμμάτων. Οι **πίνακες συχνοτήτων,** όπως είναι γνωστοί, μπορεί να περιλαμβάνουν μέχρι και έξι στήλες, αλλά η παρουσίαση πινάκων με τρεις ή τέσσερις στήλες είναι πολύ πιο συνηθισμένη.

Οι συγκεκριμένοι πίνακες συντάσσονται με κατάλληλη κατάταξη και συστηματική ομαδοποίηση των τιμών της μεταβλητής που εξετάζεται. Ο τρόπος κατασκευής τους εξαρτάται από το είδος των χαρακτηριστικών, δηλαδή το είδος της μεταβλητής.

Α. Ασυνεχής ή Διακριτή Μεταβλητή

Αν τα χαρακτηριστικά είναι *διακριτά* και τα δυνατά αποτελέσματα της μέτρησης σχετικά λίγα τότε ο πίνακας παίρνει την ακόλουθη μορφή:

Δυνατές τιμές της μεταβλητής	**Αριθμός φορών που παρατηρήθηκε η κάθε τιμή (Συχνότητα)**
x_1	f_1
x_2	f_2
.	.
.	.
.	.
x_k	f_k
Σύνολο	$\sum_{i=1}^{k} f_i = n$

Τα $x_1,\ldots, x_k$ είναι οι τιμές της διακριτής μεταβλητής Χ οι οποίες τοποθετούνται κατά αύξουσα σειρά, από τη μικρότερη στη μεγαλύτερη. Τα $f_1, \ldots, f_k$ εκφράζουν πόσες φορές εμφανίζεται στο συνολικό πληθυσμό κάθε τιμή της μεταβλητής, είναι επομένως οι συχνότητες της μεταβλητής Χ.

Όταν η τιμή x_i εμφανίζεται f_i φορές τότε λέμε ότι f_i είναι η **απόλυτη συχνότητα** ή απλά **συχνότητα** της x_i. Η **σχετική συχνότητά της** ορίζεται ως η ποσότητα f_i/n.. Ο συμβολισμός $\sum_{i=1}^{k} f_i$ ταυτίζεται με το άθροισμα όλων των f_i.

Β. Συνεχής Μεταβλητή

Αν τα χαρακτηριστικά είναι *συνεχή ή διακριτά με μεγάλο πλήθος δυνατών τιμών*, τότε δυσχεραίνεται η μορφή του πίνακα, οπότε κρίνεται απαραίτητη η *ομαδοποίηση των παρατηρήσεων*. Η ομαδοποίηση αυτή πραγματοποιείται με το χωρισμό του διαστήματος μεταβολής (α_0 , α_1) της μεταβλητής Χ σε υποδιαστήματα της μορφής

[α_{i-1} , α_i), που ονομάζονται **τάξεις** ή **ομάδες** ή **κλάσεις.**

Τα άκρα των τάξεων καλούνται αντίστοιχα, το μεν **α_{i-1} κατώτερο όριο**, το δε **α_i ανώτερο όριο.** Η διαφορά των δυο ορίων καλείται *πλάτος της τάξεως* και συμβολίζεται με δ. Το ημιάθροισμα των ορίων της κάθε τάξης καλείται **κεντρική τιμή της τάξεως**, δηλαδή:

$$x_i = \frac{a_{i-1} + a_i}{2}$$

Οι συχνότητες εδώ δίνουν τον αριθμό των παρατηρήσεων που περιέχονται στις αντίστοιχες τάξεις. Ακόμα με ***M και m συμβολίζονται η μέγιστη και η ελάχιστη τιμή αντίστοιχα της μεταβλητής.***

Μια προφανής δυσκολία που υπάρχει στην ομαδοποίηση των παρατηρήσεων είναι *ο προσδιορισμός του αριθμού των τάξεων k που θα χρησιμοποιηθούν*.

Στην πράξη, συνήθως ο αριθμός των τάξεων κυμαίνεται κατά μέσο όρο στις 8 με 10. Επίσης, συχνά έχουμε *τάξεις ίσου πλάτους*. Φυσικά, υπάρχουν και οι περιπτώσεις *άνισου πλάτους*, όπως για παράδειγμα στις κατανομές δαπανών, ημερών ανεργίας κ.ο.κ.

Εν γένει, είτε έχουμε κατανομή ίσου πλάτους, είτε άνισου πλάτους, ο πίνακας συχνοτήτων συνεχούς μεταβλητής θα έχει την ακόλουθη μορφή:

Πίνακας 4.3

Τάξεις	*Συχνότητες* (f_i)	*Κεντρικές τιμές των τάξεων* (x_i)
$\alpha_0 - \alpha_1$	f_1	x_1
$\alpha_1 - \alpha_2$	f_2	x_2
$\alpha_2 - \alpha_3$	f_3	x_3
.	.	.
.	.	.
.	.	.
$\alpha_{i-1} - \alpha_i$	f_i	x_i
.	.	.
.	.	.
.	.	.
$\alpha_{n-1} - \alpha_n$	f_n	x_n
Σύνολο	$\sum f_i$	

4.4 ΒΑΣΙΚΑ ΣΤΟΙΧΕΙΑ ΓΙΑ ΤΗΝ ΚΑΤΑΣΚΕΥΗ ΠΙΝΑΚΑ ΚΑΤΑΝΟΜΗΣ ΣΥΧΝΟΤΗΤΑΣ

1. Οι ***κλάσεις*** που επιλέγουμε πρέπει να είναι τέτοιες ώστε να δίνουν τη σωστή εικόνα της κατανομής των δεδομένων. Ο καθορισμός του αριθμού των κλάσεων είναι εν γένει αυθαίρετος, ωστόσο ως επί το πλείστον γίνεται χρήση του ***κανόνα του Sturges.*** Σύμφωνα με αυτό τον κανόνα, ο αριθμός των τάξεων κατά προσέγγιση θα δίνεται από τον εμπειρικό τύπο: $k = 1 + 3,322 \log_{10} N$, όπου k είναι ο αριθμός των κλάσεων και N ο αριθμός των δεδομένων- παρατηρήσεων. Συνήθως χρησιμοποιούνται από 5 ως 20 κλάσεις. Όσο περισσότερα δεδομένα έχουμε, τόσο περισσότερες κλάσεις πρέπει να χρησιμοποιούνται. Αυτό διότι αν ο αριθμός των

κλάσεων που θα χρησιμοποιηθούν είναι πολύ μικρός είναι ενδεχόμενο να αποκρύβονται σημαντικά χαρακτηριστικά των δεδομένων με την ομαδοποίησή του. Από την άλλη πλευρά, αν ο αριθμός των κλάσεων είναι μεγάλος σε σχέση με τα δεδομένα θα έχουμε πολλές κλάσεις που θα είναι κενές ή με μικρό αριθμό παρατηρήσεων και η κατανομή που θα εμφανίζουν θα δίνει μια όχι ικανοποιητική περιγραφή των δεδομένων.

2. Στη συνέχεια πρέπει να καθορίσουμε *το **εύρος κάθε κλάσης***. Γενικά το εύρος προκύπτει διαιρώντας τη διαφορά της μικρότερης από τη μεγαλύτερη μέτρηση με τον επιθυμητό αριθμό των κλάσεων που θέλουμε να χρησιμοποιήσουμε. Κατά κανόνα προσπαθούμε να έχουμε ***κλάσεις ίσου εύρους***.

3. Τα όρια των κλάσεων πρέπει να καθορίζονται με τρόπο ώστε οι μετρήσεις να κατανέμονται σε μία μόνο από τις δυνατές κατηγορίες.

4.5 ΥΠΟΛΟΓΙΣΜΟΣ ΣΥΧΝΟΤΗΤΩΝ ΜΕ ΜΙΚΡΟΫΠΟΛΟΓΙΣΤΕΣ ΚΑΙ SPSS

Σε μερικές περιπτώσεις οι κατανομές συχνοτήτων μπορούν να υπολογιστούν χωρίς μικροϋπολογιστές όταν ο αριθμός των δεδομένων είναι μικρός. Η κατάσταση αμβλύνεται όσο αυξάνεται ο αριθμός τους και γίνεται απελπιστικά δύσκολη, όταν ο αριθμός των κατηγοριών και των δεδομένων είναι πολύ μεγάλος. Η χρήση ηλεκτρονικών υπολογιστών και στατιστικών προγραμμάτων κάνει τον υπολογισμό συχνοτήτων, ομαδικών δεδομένων και οποιουδήποτε άλλου στατιστικού ευκολότατη, ανεξάρτητα από τον αριθμό δεδομένων και κατηγοριών. Με τους μικροϋπολογιστές η όλη δυσκολία είναι να κατασκευασθούν τα δεδομένα και να αποθηκευτούν σε αρχεία, έτσι ώστε να μπορούν να χρησιμοποιηθούν από το πρόγραμμα στατιστικής που θα χρησιμοποιήσουμε.

4.6 ΓΡΑΦΙΚΕΣ ΠΑΡΑΣΤΑΣΕΙΣ

Οι στατιστικοί πίνακες, παρά την πληρότητα την οποία παρουσιάζουν και την ακρίβεια των πληροφοριών που περιέχουν, είναι σχεδόν πάντοτε χρήσιμο οι πληροφορίες που περιέχουν να παρασταθούν με μορφή *διαγραμμάτων ή γραφικών παραστάσεων*. Επιπρόσθετα τα στατιστικά δεδομένα που συγκεντρώνονται σε ένα πίνακα συχνοτήτων μπορούν να παρουσιαστούν με τη μορφή *γραφικών παραστάσεων ή διαγραμμάτων*.

Με αυτό τον τρόπο επιτυγχάνεται μια εποπτική αντίληψη του φαινομένου και επιτρέπεται ο τονισμός των κύριων χαρακτηριστικών του, αδιαφορώντας για τις λεπτομέρειες, που τις περισσότερες φορές δεν έχουν και μεγάλη σημασία.

Γράφημα αποκαλείται μια γραφική αναπαράσταση μίας ή περισσοτέρων μεταβλητών. Τα γραφήματα είναι χρήσιμα για να βλέπουμε και να καταλαβαίνουμε το σχήμα της κατανομής μίας μεταβλητής.

Η σημασία των διαγραμμάτων έγκειται στο ότι όταν είναι ορθά σχεδιασμένα, είναι πιο εύκολο στον αναγνώστη να το συγκρατήσει στην μνήμη του από ότι έναν στατιστικό πίνακα. Επιπλέον, μετατρέπουν τους αριθμούς που περιέχονται μέσα σε έναν πίνακα και που είναι μια αφηρημένη έννοια, σε μια συγκεκριμένη απεικόνιση με γεωμετρικό σχήμα και έτσι αποκτούμε μια αντίληψη του φαινομένου που εξετάζουμε.

Τα διαγράμματα πρέπει να έχουν *τίτλο* που να είναι σύντομος και σαφής και αναγράφεται συνήθως στο κάτω μέρος τους. Κατά μήκος των αξόνων των διαγραμμάτων πρέπει να σημειώνονται οι *κλίμακες* των τιμών των μεγεθών που απεικονίζονται. Όταν είναι αναγκαίο, θα πρέπει κάτω από το διάγραμμα να αναγράφονται οι τυχόν *υποσημειώσεις* για διευκρινήσεις ή συμπληρωματικές επεξηγήσεις των μεγεθών που απεικονίζονται.

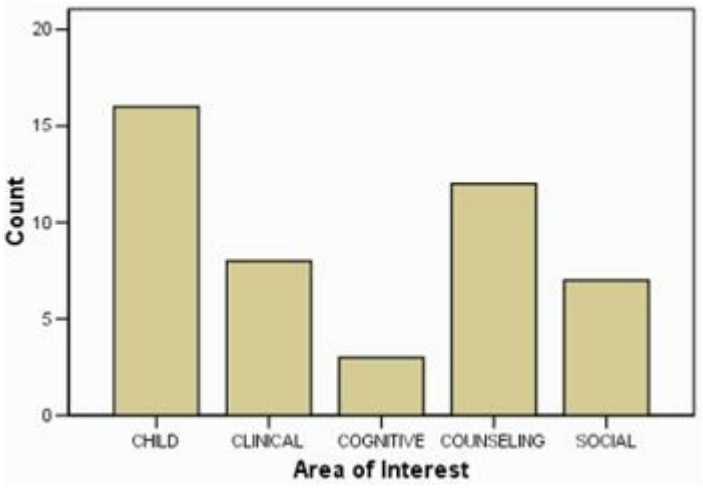
Τέλος, πρέπει να αναφέρεται και η *πηγή* από την οποία πήραμε τα αριθμητικά δεδομένα των μεγεθών. Από τις γραφικές παραστάσεις ή τα διαγράμματα μπορούμε

να αναλύσουμε συνοπτικά διάφορες χρήσιμες πληροφορίες. Με τα διαγράμματα διευκολύνεται η σύγκριση μεταξύ ομοειδών στοιχείων για τα ίδια ή διαφορετικά χαρακτηριστικά.

Τα κυριότερα είδη ***γραφικών παραστάσεων*** είναι τα εξής:

- **Ραβδόγραμμα**
- **Διάγραμμα συχνοτήτων ή Ιστόγραμμα**
- **Κυκλικό διάγραμμα**
- **Χρονόγραμμα**

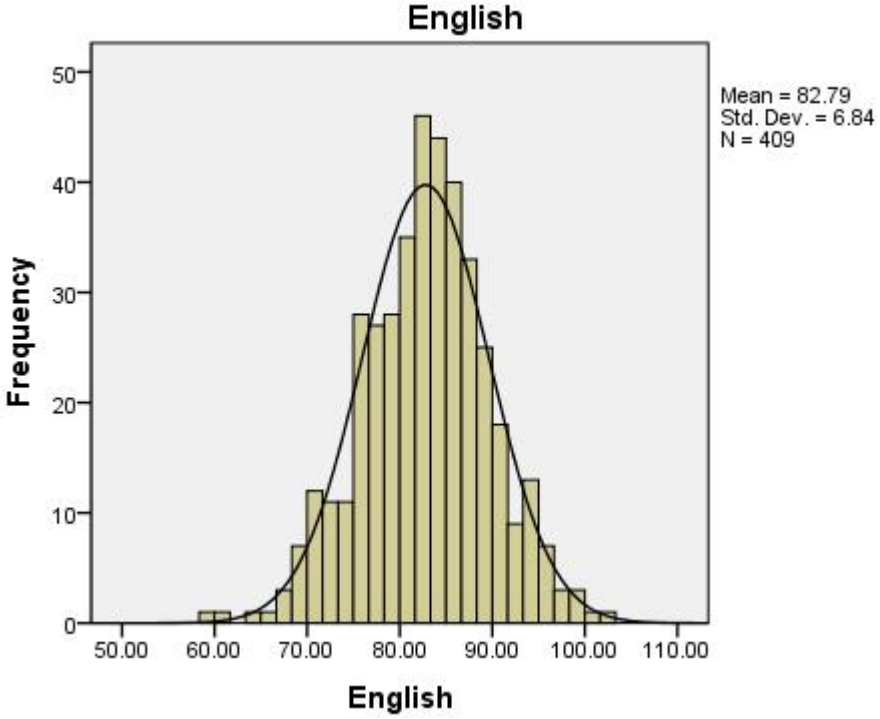
4.6.1 Ραβδόγραμμα (Bar Chart)



Το **ραβδόγραμμα** χρησιμοποιείται συνήθως όταν η μεταβλητή Χ είναι ποιοτική. Αποτελείται από ορθογώνιες στήλες, που οι βάσεις τους έχουν κοινό μήκος που επιλέγεται αυθαίρετα, είναι όμως τέτοιο ώστε *να εξασφαλίζονται κενά μεταξύ δύο διαδοχικών ορθογωνίων*. Οι στήλες υψώνονται πάνω σε ορθογώνιο ή κατακόρυφο άξονα. ***Το ύψος κάθε στήλης είναι ίσο με την συχνότητα (*** ν_i***) ή τη σχετική***

συχνότητα fi της αντίστοιχης τιμής της μεταβλητής. Όταν θέλουμε να κάνουμε σύγκριση των αντίστοιχων τιμών της ίδιας μεταβλητής Χ για δύο διαφορετικά δείγματα, τότε κατασκευάζουμε διπλά ορθογώνια για την ίδια τιμή της μεταβλητής Χ, ένα για το κάθε δείγμα.

4.6.2 Διάγραμμα συχνοτήτων ή Ιστόγραμμα (Histogram)



Στην περίπτωση ποσοτικών μεταβλητών αντί του ραβδογράμματος χρησιμοποιείται το **διάγραμμα συχνοτήτων και σχετικών συχνοτήτων ή Ιστόγραμμα** όπως αλλιώς ονομάζεται. Το Ιστόγραμμα κατά κανόνα, χρησιμοποιείται όταν θέλουμε να αναπαραστήσουμε κατανομές συχνοτήτων με διαστήματα τάξεων.

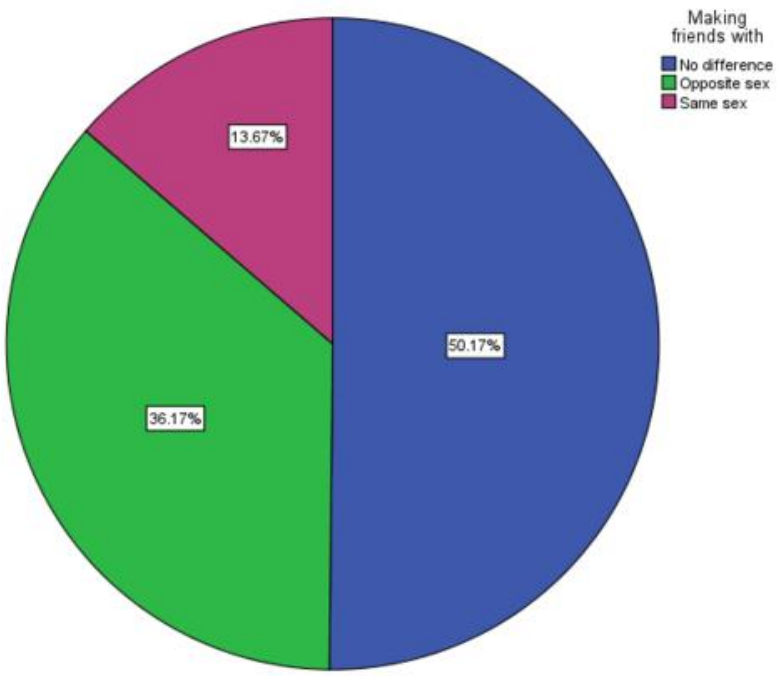
Αποτελείται από ευθύγραμμα τμήματα κάθετα στον άξονα της μεταβλητής Χ, ένα για κάθε τιμή x_i. Οι τιμές της μεταβλητής τοποθετούνται στο διάγραμμα σε αύξουσα

σειρά στον οριζόντιο άξονα xx΄. Το ύψος κάθε ευθύγραμμου τμήματος που αντιστοιχεί στον κάθετο άξονα yy΄ είναι ίσο με την αντίστοιχη συχνότητα v_i ή με την αντίστοιχη σχετική συχνότητα f_i. Αν ενώσουμε τα σημεία (x_i, v_i), i = 1, 2, …, κ με διαδοχικά ευθύγραμμα τμήματα δημιουργείται μια πολυγωνική γραμμή που ονομάζεται ***πολύγωνο συχνοτήτων***. Ομοίως αν ενώσουμε τα σημεία (x_i, f_i) δημιουργείται το ***πολύγωνο σχετικών συχνοτήτων***.

4.6.3 Πολύγωνα Συχνότητας (Frequency Polygon)

Το ***Πολύγωνο Συχνότητας (Frequency Polygon***) είναι ένας έτερος τρόπος παρουσίασης κατανομής συχνότητας. Κατασκευάζεται ως εξής: Σημειώνουμε τη συχνότητα κάθε κλάσης πάνω από το μέσο σημείο της κλάσης και στη συνέχεια ενώνουμε τα διαδοχικά σημεία που προκύπτουν με ευθύγραμμα τμήματα. Το πολύγωνο «κλείνει» με τη θεώρηση μιας πρόσθετης θεωρητικής κλάσης (με συχνότητα 0) σε κάθε ένα από τα άκρα του διαστήματος των τιμών της κατανομής και την προσθήκη, στη συνέχεια στην τεθλασμένη γραμμή δύο ευθυγράμμων τμημάτων που συνδέουν τα άκρα της τεθλασμένης με τα ενδιάμεσα σημεία των κλάσεων αυτών.

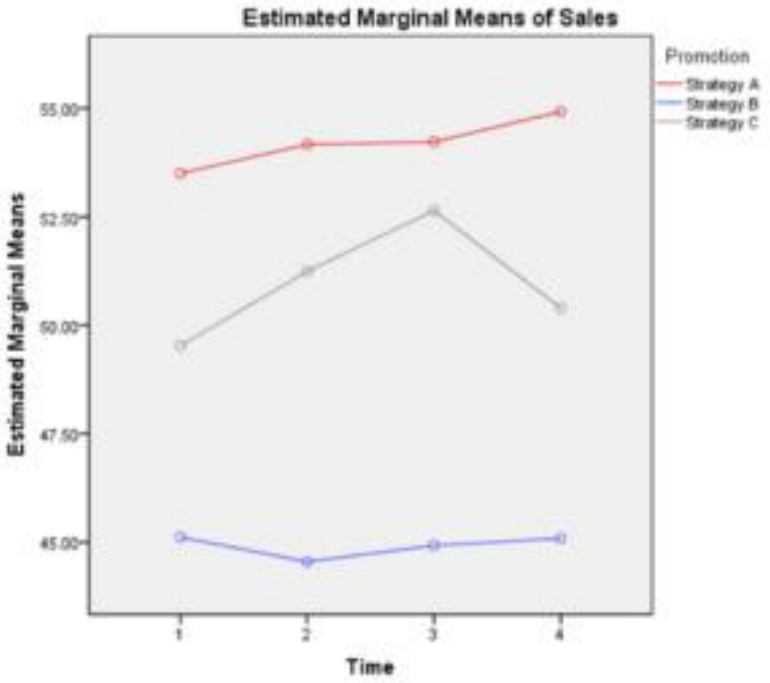
4.6.4 Κυκλικό διάγραμμα (Pie Chart)



Το **κυκλικό διάγραμμα** χρησιμοποιείται για την γραφική παράσταση ποιοτικών ή ποσοτικών μεταβλητών όταν οι τιμές της μεταβλητής X είναι σχετικά λίγες. Τα διαγράμματα αυτά είναι αποτελεσματικά στις περιπτώσεις που αντικειμενικός σκοπός είναι η παρουσίαση των συνιστωσών μιας ολότητας με τέτοιο τρόπο που να αναδεικνύονται τα σχετικά τους μεγέθη.

Το κυκλικό διάγραμμα είναι ένας κυκλικός δίσκος χωρισμένος σε k κυκλικούς τομείς (όσες και οι τιμές της μεταβλητής X). Το εμβαδόν E_i κάθε κυκλικού τομέα είναι ανάλογο προς τις αντίστοιχες συχνότητες v_i ή τις σχετικές συχνότητες f_i.

4.6.5 Χρονόγραμμα (Time Chart)

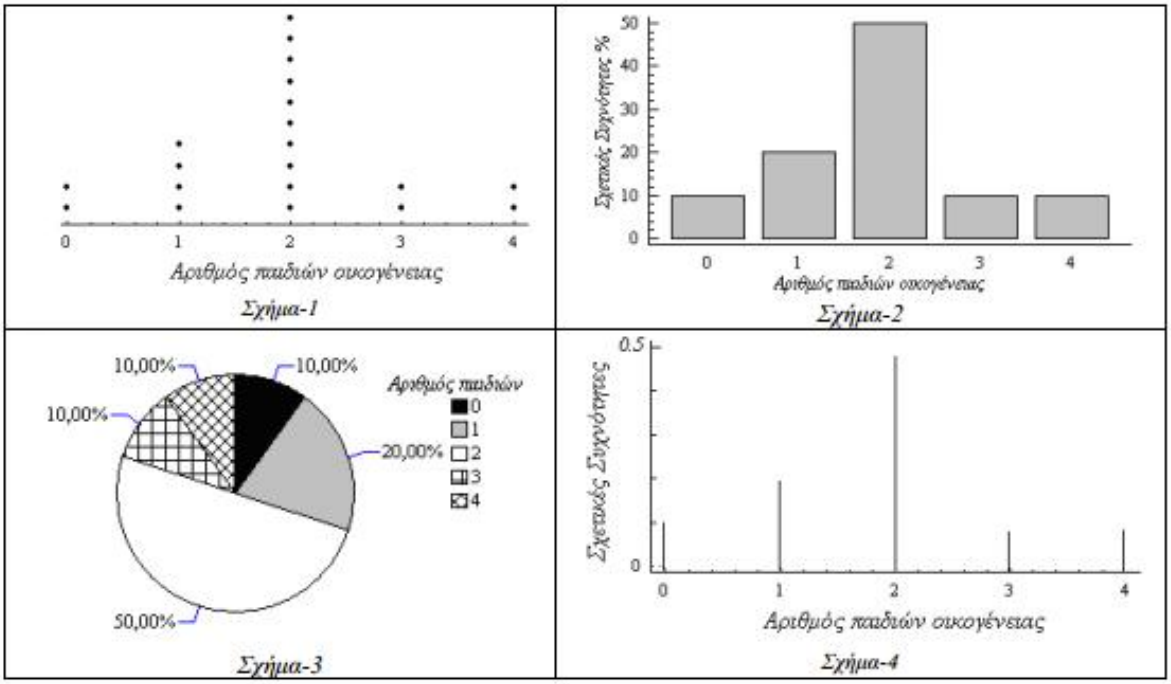


Όταν θέλουμε να παρακολουθήσουμε την διαχρονική εξέλιξη διαφόρων μεγεθών, τότε κατασκευάζουμε τη γραφική παράσταση στην οποία στον οριζόντιο άξονα xx΄ λαμβάνουμε ισομήκη διαδοχικά τμήματα, καθένα από τα οποία αντιστοιχεί στη μονάδα του χρησιμοποιούμενου χρόνου και στον κατακόρυφο άξονα yy΄ παίρνουμε κλίμακα η οποία πρέπει να καλύπτει τις τιμές της μεταβλητής.

4.7 ΠΑΡΑΔΕΙΓΜΑ ΑΠΕΙΚΟΝΙΣΗΣ ΜΕΤΑΒΛΗΤΗΣ

Όσον αφορά την περιγραφή της *κατανομής δεδομένων από ποσοτικές μεταβλητές*, η Περιγραφική Στατιστική μας προσφέρει πολλές δυνατότητες γραφικής παρουσίασής τους.

Στα ακόλουθα τέσσερα σχήματα φαίνονται ***τέσσερις διαφορετικές γραφικές αναπαραστάσεις/απεικονίσεις της κατανομής του δείγματος από τη μεταβλητή U***:



Από κοινού και οι τέσσερις γραφικές απεικονίσεις, δίνουν μια πιο παραστατική και πιο ευκρινή εικόνα της κατανομής του δείγματος από αυτήν του αντίστοιχου πίνακα συχνοτήτων. Δε μας τροφοδοτούν με περισσότερη ή διαφορετική πληροφορία από αυτήν που μας δίνει ο πίνακας συχνοτήτων, γιατί κατασκευάζονται με βάση την πληροφορία που παίρνουμε από αυτόν και αυτή η πληροφορία απεικονίζεται γραφικά. Όμως δίνουν αυτήν την πληροφορία με πιο παραστατικό τρόπο και την κάνουν πιο εύκολα κατανοητή. Ιδιαίτερα, συμπεράσματα που αφορούν τη μορφή και

τη θέση της κατανομής, προκύπτουν με πιο άμεσο και προφανή τρόπο και χωρίς να απαιτείται ιδιαίτερη εμπειρία.

Ο τρόπος που κατασκευάζονται τα διαγράμματα αυτά είναι προφανής και απλός.

• Στο **Σημειόγραμμα**, απεικονίζουμε τα δεδομένα ως *κουκίδες* στις αντίστοιχες θέσεις ενός οριζόντιου άξονα.

• Στο **Ραβδόγραμμα**, απεικονίζουμε τις *συχνότητες* ή τις *σχετικές συχνότητες* των διαφορετικών τιμών, i =1,2,...,k ως ύψη ορθογωνίων που σχεδιάζουμε στις αντίστοιχες θέσεις του οριζόντιου άξονα. Τα ορθογώνια έχουν ίδιο μήκος βάσης που επιλέγουμε αυθαίρετα. Επίσης, το ραβδόγραμμα μπορεί να σχεδιασθεί με οριζόντιο, αντί κατακόρυφο προσανατολισμό.

• Στο **Διάγραμμα Συχνοτήτων** (ή σχετικών συχνοτήτων), απεικονίζουμε τις συχνότητες (αντίστοιχα τις σχετικές συχνότητες) των διαφορετικών τιμών y_i, όπου i=1,2,...,k όπως και στο ραβδόγραμμα, με την διαφορά ότι στις θέσεις των y_i χαράσσουμε κάθετα ευθύγραμμα τμήματα αντί ορθογωνίων.

• Στο **Κυκλικό Διάγραμμα**, απεικονίζουμε τις *συχνότητες* ή τις *σχετικές συχνότητες* των διαφορετικών τιμών y_i, i=1,2,...,k με ένα διαφορετικό τρόπο.

4.8 ΣΤΑΤΙΣΤΙΚΕΣ ΕΚΘΕΣΕΙΣ- ΑΝΑΦΟΡΕΣ

Πλην των πινάκων και των γραφικών παραστάσεων υπάρχει και εναλλακτικός τρόπος παρουσίασης των στατιστικών στοιχείων και αυτός είναι οι **εκθέσεις ή αναφορές.** Στο κείμενο που συνοδεύει αυτές τις εκθέσεις αναφέρονται τα πιο κύρια σημεία των αποτελεσμάτων, γίνεται σχολιασμός της σημασίας τους και επίσης αναγράφονται και ορισμένες παρατηρήσεις του προσώπου που συνέταξε την αναφορά, όπως επίσης μπορεί να αναφέρεται και με μια σύντομη σημείωση η στατιστική τεχνική που χρησιμοποιήθηκε στην έρευνα. Προαιρετικά, εντός της έκθεσης περιλαμβάνονται και περιληπτικοί πίνακες.

ΚΕΦΑΛΑΙΟ 5: ΑΡΙΘΜΗΤΙΚΗ ΠΑΡΟΥΣΙΑΣΗ ΣΤΑΤΙΣΤΙΚΩΝ ΣΤΟΙΧΕΙΩΝ-

ΠΕΡΙΓΡΑΦΙΚΟΙ ΠΑΡΑΜΕΤΡΟΙ Η΄ ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ ΚΑΙ ΜΕΤΑΒΛΗΤΕΣ

5.1 ΑΡΙΘΜΗΤΙΚΗ ΠΕΡΙΓΡΑΦΗ ΔΕΔΟΜΕΝΩΝ

Οι γραφικές μέθοδοι αποτελούν μια χρήσιμη και γρήγορη παρουσίαση των δεδομένων ωστόσο δεν είναι εύκολο να χρησιμοποιηθούν για *στατιστική συμπερασματολογία*. Ως εκ τούτου, για την αποφυγή αυτού του προβλήματος χρησιμοποιούμε ***αριθμητικά περιγραφικά μέτρα (numerical descriptive measures***). Εκείνο που μας ενδιαφέρει κυρίως είναι να χρησιμοποιήσουμε τα δειγματικά δεδομένα για να υπολογίσουμε ένα σύνολο τιμών στατιστικών συναρτήσεων, που θα μας δώσουν μια καλή θεωρητική εικόνα της δειγματικής κατανομής σχετικής συχνότητας και τα οποία θα είναι χρήσιμα για να βοηθηθούμε στη στατιστική συμπερασματολογία που αναφέρεται στην κατανομή της σχετικής συχνότητας του *πληθυσμού*.

Δύο είναι τα χαρακτηριστικά ενός δείγματος που μπορούν να μας αποφέρουν μια καλή συνοπτική εικόνα για το δείγμα. Το πρώτο χαρακτηριστικό είναι κάποια τιμή γύρω από την οποία τα δεδομένα τείνουν να συσσωρεύονται. Τέτοια μέτρα ονομάζονται ***μέτρα θέσης (measures of location) ή μέτρα κεντρικής τάσης (measures of central tendency***). Το άλλο χαρακτηριστικό αναφέρεται στη ***μεταβλητότητα των δεδομένων***. Στον καθορισμό δηλαδή της διασποράς των δεδομένων γύρω από κάποιο μέτρο αριθμητικής θέσης.

5.2 ΜΕΤΡΑ ΘΕΣΗΣ ΤΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ Η΄ ΠΑΡΑΜΕΤΡΟΙ ΚΕΝΤΡΙΚΗΣ ΤΑΣΗΣ

Τα ***μέτρα θέσης ή κεντρικής τάσης*** των στατιστικών δεδομένων μετρούν τη θέση στην οποία τα δεδομένα έχουν την τάση να συγκεντρώνονται και μπορούν να χρησιμοποιηθούν σαν μια αντιπροσωπευτική τιμή των δεδομένων.

5.2.1 ΜΕΤΡΑ ΘΕΣΗΣ ΓΙΑ ΜΗ ΟΜΑΔΟΠΟΙΗΜΕΝΕΣ ΠΑΡΑΤΗΡΗΣΕΙΣ

Ο Αριθμητικός Μέσος ή Μέση Τιμή (Mean) ορίζεται ως το πηλίκο του αθροίσματος των τιμών της μεταβλητής δια το πλήθος των τιμών της. Η τιμή του επηρεάζεται από την τιμή όλων των όρων της μεταβλητής.

Συμβολισμός: **μ** αν αναφερόμαστε σε πληθυσμό και $\overline{x}$ όταν αναφερόμαστε σε δείγμα.

Τύποι: $\mu = \frac{\sum_{i=1}^{N} x_i}{N}$, $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

Διάμεσος (M) (Median) είναι η τιμή εκείνη της μεταβλητής που χωρίζει το σύνολο των τιμών σε δυο ίσα μέρη, ώστε ο αριθμός των παρατηρήσεων που είναι μικρότερες από το Μ, να είναι ίσος με τον αριθμό αυτών που είναι μεγαλύτερες από το Μ. Είναι το σημείο της κατανομής που αφήνει 50% των παρατηρήσεων προς τα πάνω και 50% προς τα κάτω.

Για να υπολογίσουμε την διάμεσο, οι παρατηρήσεις κατατάσσονται κατά τη φυσική τους διάταξη. Στην περίπτωση που οι τιμές της μεταβλητής δεν περιέχονται σε πίνακα συχνοτήτων, η διάμεσος δίνεται από τον όρο (N+1)/2, όπου N το πλήθος των παρατηρήσεων. Εάν το N είναι περιττός αριθμός η διάμεσος είναι η παρατήρηση που βρίσκεται στη (N+1)/2 θέση, γιατί αυτή η παρατήρηση αφήνει (N-1)/2 παρατηρήσεις προς τα κάτω και (N-1)/2 παρατηρήσεις προς τα πάνω.

Αν το N είναι άρτιος, τότε στο μέσο των τιμών υπάρχουν δυο τιμές, οπότε η διάμεσος είναι ο μέσος όρος των δυο αυτών μεσαίων τιμών.

Κορυφή (Mode) ή Επικρατούσα Τιμή (Mo) είναι η τιμή της μεταβλητής με τη μεγαλύτερη συχνότητα. Η κορυφή δεν καθορίζεται πάντοτε μονοσήμαντα.

Το **κ-Ποσοστημόριο (Pκ) (Percentiles**) ενός συνόλου τιμών είναι η τιμή εκείνη για την οποία το κ% των παρατηρήσεων είναι μικρότερες από αυτή την τιμή. Π.χ., αν κ=90, τότε το P_{90} είναι η τιμή που αφήνει προς τα κάτω το 90 % των παρατηρήσεων.

5.2.2 ΜΕΤΡΑ ΘΕΣΗΣ ΓΙΑ ΟΜΑΔΟΠΟΙΗΜΕΝΕΣ ΠΑΡΑΤΗΡΗΣΕΙΣ

Τα κυριότερα μέτρα θέσης είναι:

Ø Ο ***Αριθμητικός Μέσος*** ή ***Μέσος Όρος***: Στην περίπτωση που τα στατιστικά δεδομένα δίνονται σε μορφή κατανομής συχνοτήτων και η μεταβλητή είναι διακριτή, τότε ο Αριθμητικός Μέσος δίνεται από την σχέση: $m=\frac{\sum f_i x_i}{\sum f_i}$ όπου f_i είναι οι συχνότητες και x_i είναι οι τιμές της μεταβλητής.

Αν τα δεδομένα εμφανίζονται με μορφή κατανομής συχνοτήτων κατά τάξεις ο τύπος υπολογισμού είναι και πάλι ο ίδιος, μόνο που εδώ x_i είναι οι κεντρικές τιμές των τάξεων.

Ø Ο ***Αρμονικός Μέσος***: Χρησιμοποιείται κυρίως για τον υπολογισμό μέσων ποσοστών χρόνου και προβλημάτων αποστάσεων. Ο τύπος υπολογισμού τους είναι:

$$H=\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+...+\frac{1}{x_n}}=\frac{n}{\sum \frac{1}{x_i}}$$

Ø Ο ***Γεωμετρικός Μέσος***: Επηρεάζεται πολύ λίγο από μεμονωμένες υψηλές τιμές και υπολογίζεται ως εξής: $G=\sqrt[N]{x_1 \times x_2 \times x_3 ... \times x_N}$

Ø Ο ***Σταθμισμένος ή Σταθμικός Μέσος***: Χρησιμοποιείται στην περίπτωση που θέλουμε να δώσουμε μεγαλύτερη βαρύτητα σε ορισμένες παρατηρήσεις. Όταν η μεταβλητή είναι ασυνεχής δίνεται από τον τύπο: $m=\frac{\sum f_i x_i}{\sum f_i}$

Ø Η Δ***ιάμεσος ή Διχοτόμος***: Είναι η τιμή της μεσαίας παρατήρησης όταν όλες οι παρατηρήσεις είναι ταξινομημένες σε αύξουσα ή φθίνουσα σειρά. Εναλλακτικά, ορίζεται ως η στατιστική παράμετρος που διαχωρίζει τις τιμές της μεταβλητής σε δύο ίσες ομάδες όπου το 50% εξ αυτών είναι μικρότερο από την τιμή της Διαμέσου και το υπόλοιπο 50% είναι μεγαλύτερο από την Διάμεσο.

Αν έχουμε ταξινομημένα δεδομένα ο τύπος υπολογισμού της διαμέσου είναι:

$$M = a_{i-1} + \frac{d}{f_i}\left(\frac{N}{2} - F_{i-1}\right)$$

ø Τα ***Εκατοστημόρια ή Ποσοστιαία Σημεία***: Τα ***εκατοστημόρια (percentiles***) ή ***ποσοστιαία σημεία (quartiles***) υπολογίζονται με τρόπο ανάλογο προς αυτό της διαμέσου. Το Εκατοστημόριο ενός συνόλου είναι εκείνη η τιμή η οποία, όταν οι τιμές διαταχθούν σε αύξουσα σειρά, έχει από αριστερά της το p% των δεδομένων και από δεξιά της το υπόλοιπο (100-p)%.

Ø Η ***Επικρατούσα Τιμή ή Τύπος ή Σημείο Μέγιστης Συχνότητας (Mode)***: Ορίζεται ως η τιμή με τη μεγαλύτερη συχνότητα.

5.3 ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ

Τα μέτρα διασποράς είναι τα αριθμητικά μεγέθη που μας δίνουν την διασπορά των παρατηρήσεων γύρω από τις κεντρικές τιμές της κατανομής.

Τα μέτρα διασποράς είναι το εύρος, η διακύμανση, η τυπική απόκλιση και το ενδοτεταρτημοριακό εύρος. Ένα μέτρο διασποράς μας δίνει με τρόπο περιληπτικό και αντικειμενικό την μεταβλητότητα ή ανομοιογένεια των παρατηρήσεων.

Τα κυριότερα μέτρα διασποράς είναι:

- Το **Εύρος** των τιμών: Το απλούστερο από τα μέτρα διασποράς είναι το **Εύρος ή Κύμανση (*Range*)** (*R*), που ορίζεται ως η διαφορά της ελάχιστης παρατήρησης από τη μέγιστη παρατήρηση. Υπολογίζεται ως ακολούθως:

$R=X_{max} - X_{min}$

- Το **Ενδοτεταρτημοριακό Εύρος**: ***Το ενδοτεταρτημοριακό εύρος (Interquartile Range)*** είναι η διαφορά του πρώτου τεταρτημορίου Q_1 από το τρίτο τεταρτημόριο Q_3.. Στο μεταξύ τους διάστημα περιλαμβάνεται το 50% των παρατηρήσεων. Επομένως όσο μικρότερο είναι αυτό το διάστημα, τόσο μεγαλύτερη θα είναι η συγκέντρωση των τιμών και άρα μικρότερη η διασπορά των τιμών της μεταβλητής. Ο τύπος υπολογισμού είναι: $Q = \frac{Q_3 - Q_1}{2}$

- Η **Διακύμανση**: Ένας άλλος τρόπος για να υπολογίσουμε τη διασπορά των παρατηρήσεων $t_1,t_2,...,t_v$ μιας μεταβλητής X θα ήταν να αφαιρέσουμε τη μέση τιμή $\bar{x}$ από κάθε παρατήρηση και να βρούμε τον αριθμητικό μέσο των διαφορών αυτών. Κατά συνέπεια η Διακύμανση είναι ουσιαστικά ο αριθμητικός μέσος των τετραγώνων των αποκλίσεων των τιμών των παρατηρήσεων από τον αριθμητικό μέσο.

Όταν έχουμε αταξινόμητες παρατηρήσεις ο τύπος υπολογισμού της είναι:

$$s^2 = \frac{\sum (x_i - m)^2}{N}.$$

Όταν οι παρατηρήσεις είναι ομαδοποιημένες δίνεται από την σχέση:

$$s^2 = \frac{\sum f_i (x_i - m)^2}{\sum f_i}$$

- **Τυπική Απόκλιση**: Η διακύμανση είναι μια αξιόπιστη παράμετρος διασποράς, αλλά έχει ένα *μειονέκτημα*. Δεν εκφράζεται με τις μονάδες με τις οποίες εκφράζονται οι παρατηρήσεις. Για παράδειγμα, αν οι παρατηρήσεις εκφράζονται σε cm, η διακύμανση εκφράζεται σε cm^2. Αν όμως πάρουμε τη *θετική τετραγωνική ρίζα της διακύμανσης*, θα έχουμε ένα μέτρο διασποράς που θα εκφράζεται με την ίδια μονάδα μέτρησης του χαρακτηριστικού, όπως ακριβώς είναι και όλα τα άλλα μέτρα θέσης, που εξετάσαμε έως τώρα. Η ποσότητα αυτή λέγεται **Τυπική Απόκλιση** και υπολογίζεται ως εξής: $s = \sqrt{s^2} = \sqrt{\frac{\sum (x_i - m)^2}{N}}$

- **Συντελεστής Μεταβλητότητας**: Ο συντελεστής μεταβολής εκφράζεται σε ποσοστό επί τοις εκατό, είναι συνεπώς ανεξάρτητος από τις μονάδες μέτρησης και παριστάνει ένα μέτρο σχετικής διασποράς των τιμών και όχι της απόλυτης διασποράς. Ο συντελεστής μεταβλητότητας της μεταβλητής X δίνεται από τον τύπο:

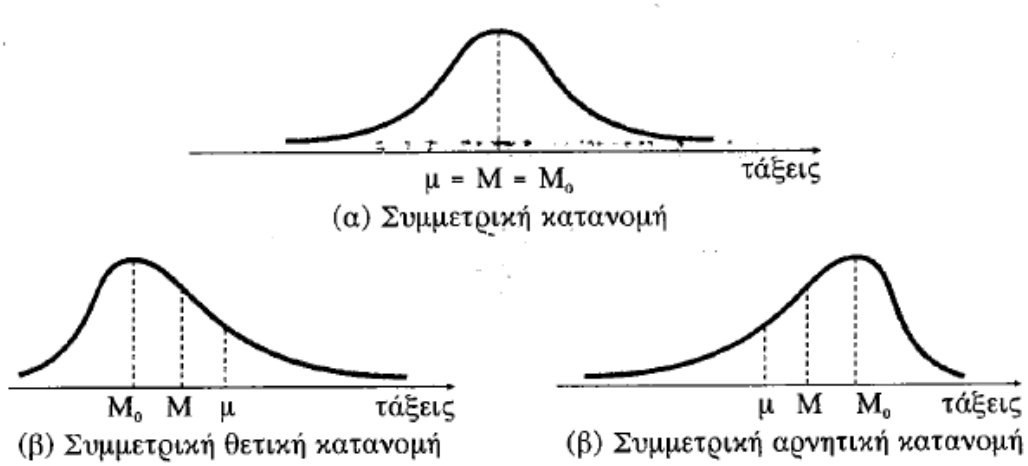
$$CV(X) = \frac{s}{m} \times 100\%$$

ΚΕΦΑΛΑΙΟ 6: ΑΣΥΜΜΕΤΡΙΑ & ΚΥΡΤΩΣΗ

6.1 ΜΕΤΡΑ Η΄ ΣΥΝΤΕΛΕΣΤΕΣ ΑΣΥΜΜΕΤΡΙΑΣ

Τα *μέτρα κεντρικής τάσης και διασποράς* μιας κατανομής δίνουν μία πρώτη εικόνα της μορφής της. Η εικόνα αυτή βελτιώνεται αν προσδιορίσουμε και ένα μέτρο ασυμμετρίας της. Με άλλα λόγια αν προσδιορίσουμε πόσο και προς ποια κατεύθυνση αποκλίνει η κατανομή μας από την πλήρως συμμετρική κατανομή. Η ασυμμετρία δύναται να είναι θετική ή αρνητική.

Συμμετρική είναι μια κατανομή όταν οι τιμές της τοποθετούνται συμμετρικά γύρω από την μέση τιμή. ***Θετικά Ασυμμετρική*** είναι μια κατανομή όταν παρουσιάζει εξόγκωση προς τα αριστερά και επιμήκυνση του άκρου της που αντιστοιχεί στις μεγαλύτερες τιμές του χαρακτηριστικού. Διαφοροτρόπως, η μεγάλη συγκέντρωση των παρατηρήσεων βρίσκεται στις μικρές τιμές της μεταβλητής. Αντίθετα, ***Αρνητικά Ασυμμετρική*** είναι μία κατανομή όταν παρουσιάζει εξόγκωση προς τα δεξιά και επιμήκυνση του άκρου της που αντιστοιχεί στις μικρότερες τιμές του χαρακτηριστικού. Παρακάτω παρατηρούμε μια συμμετρική και δύο ασυμμετρικές κατανομές:



(α) Συμμετρική κατανομή

(β) Συμμετρική θετική κατανομή

(β) Συμμετρική αρνητική κατανομή

Ο **Συντελεστής Ασυμμετρίας** μετρά το βαθμό ασυμμετρίας στην κατανομή συχνοτήτων μιας τυχαίας μεταβλητής. Υπάρχουν στην βιβλιογραφία διάφοροι συντελεστές που μετρούν την ασυμμετρία μιας κατανομής. Ενδεικτικά παραθέτουμε τους εξής:

- Συντελεστής Ασυμμετρίας S_k του Pearson: $S_k = \frac{\text{m-}\ M_0}{\text{s}}$
- Συντελεστής Ασυμμετρίας S_k του Bowley: $S_k = \frac{(Q_3 - M) - (M - Q_1)}{(Q_3 - M) + (M - Q_1)}$

Η τιμή του συντελεστή S_k μπορεί να πάρει τιμή από το -1 μέχρι και το +1. Δηλαδή $-1 \leq S_k \leq +1$.

- Αν S_k =0 τότε έχουμε συμμετρική κατανομή.
- Αν S_k>0 έχουμε θετική ασυμμετρία.
- Αν S_k<0 έχουμε αρνητική ασυμμετρία.

6.2 ΜΕΤΡΑ Η΄ ΣΥΝΤΕΛΕΣΤΕΣ ΚΥΡΤΩΣΗΣ

Δύο μονοκόρυφες κατανομές είναι δυνατόν να έχουν τον ίδιο αριθμητικό μέσο, την ίδια τυπική απόκλιση, να είναι συμμετρικές αλλά παρόλα αυτά να διαφέρουν ως προς την οξύτητα της κορυφής τους. Η κύρτωση μιας κατανομής μετράει τον βαθμό συγκέντρωσης των τιμών της μεταβλητής στην περιοχή του αριθμητικού μέσου και προς τα άκρα του μέσου. Αλλιώτικα, η κύρτωση κρίνει το πόσο λεπτή ή πλατιά είναι η κατανομή.

Με βάση το χαρακτηριστικό αυτό μια κατανομή μπορεί να θεωρηθεί:

- ***Λεπτόκυρτη***
- ***Πλατύκυρτη***
- ***Μεσόκυρτη***

Ως μέτρο κύρτωσης της κατανομής n παρατηρήσεων ο K. Pearson όρισε το συντελεστή: $b_4 = \dfrac{\sum_{i=1}^{n} (X_i - \bar{X})^4}{n S^4}$

Το μέτρο αυτό είναι γνωστό ως ***συντελεστής κύρτωσης β_4 του Pearson***.

Αποδεικνύεται ότι :

$$\beta_4 = \begin{cases} > 3 & \lambda\epsilon\pi\tau o\kappa\upsilon\rho\tau\eta \\ = 3 & \mu\epsilon\sigma o\kappa\upsilon\rho\tau\eta \\ < 3 & \pi\lambda\alpha\tau\upsilon\kappa\upsilon\rho\tau\eta \end{cases}$$

Πρέπει να σημειώσουμε πως η Κύρτωση αναφέρεται σε Συμμετρικές κατανομές.

ΚΕΦΑΛΑΙΟ 7: ΤΥΠΙΚΗ ΚΑΝΟΝΙΚΗ ΚΑΤΑΝΟΜΗ

7.1 ΤΥΠΙΚΗ ΚΑΝΟΝΙΚΗ ΚΑΤΑΝΟΜΗ

Η χρήση των ***τιμών z*** γίνεται ιδιαίτερα αποτελεσματική, όταν συνδυαστούν με τις κανονικές κατανομές. Αυτό που χρειάζεται, για να καθοριστεί η κατανομή μιας οποιασδήποτε κανονικά κατανεμημένης μεταβλητής, είναι ο αριθμητικός μέσος και η τυπική απόκλιση της μεταβλητής. Μπορούμε παράλληλα, αν γνωρίζουμε τον αριθμητικό μέσο και την τυπική απόκλιση της μεταβλητής, να υπολογίσουμε το εμβαδόν που πέφτει κάτω από ένα ορισμένο μέρος της κανονικής καμπύλης (από μια τιμή και πάνω, από μια τιμή και κάτω ή μεταξύ δύο τιμών της μεταβλητής). Βέβαια, αν ξέρουμε το εμβαδόν ενός μέρους της κανονικής καμπύλης, μπορούμε να υπολογίσουμε το ποσοστό των τιμών που περιλαμβάνονται σε αυτό διαιρώντας το διά του ολικού εμβαδού.

Αντίθετα, αν γνωρίζουμε το εμβαδόν (ποσοστό) κάτω από ένα μέρος της κανονικής καμπύλης, είναι δυνατόν να υπολογίσουμε τις τιμές της μεταβλητής που αντιστοιχούν στο ποσοστό αυτό.

Αυτά ισχύουν για όλες τις κανονικές κατανομές. Οι *τυπικές* (z*) κατανομές*, από την άλλη μεριά, έχουν πάντοτε αριθμητικό μέσο 0 και τυπική απόκλιση +1. Αυτό σημαίνει ότι τα ποσοστά που αντιστοιχούν σε μια τιμή z είναι τα ίδια για όλες τις κανονικές κατανομές, αφού ο αριθμητικός μέσος και η τυπική απόκλιση της τυπικής κατανομής είναι πάντοτε 0 και 1. Με άλλα λόγια, αν έχουμε μια τυπική κανονική κατανομή, τα ποσοστά αυτά χρειάζεται να υπολογιστούν μόνο μια φορά. Πράγματι, αυτοί οι υπολογισμοί έχουν γίνει και τα ποσοστά που πέφτουν μεταξύ δύο τιμών z παρέχονται σε ειδικούς πίνακες.

Τα ποσοστά αυτά αντικατοπτρίζουν τις πιθανότητες να πάρει μια μεταβλητή τιμές πάνω ή κάτω από ένα δεδομένο όριο ή μεταξύ δύο ορίων. Αν λοιπόν, έχουμε μια κανονική κατανομή και χρειαζόμαστε τα ποσοστά ή τις πιθανότητες που πέφτουν μεταξύ δυο τιμών της, αυτό που έχουμε να κάνουμε είναι να τις μετατρέψουμε σε

τιμές z και να βρούμε τα αντίστοιχα ποσοστά στον πίνακα κανονικών τυπικών κατανομών.

Τα βασικά χαρακτηριστικά της κατανομής είναι τα εξής:

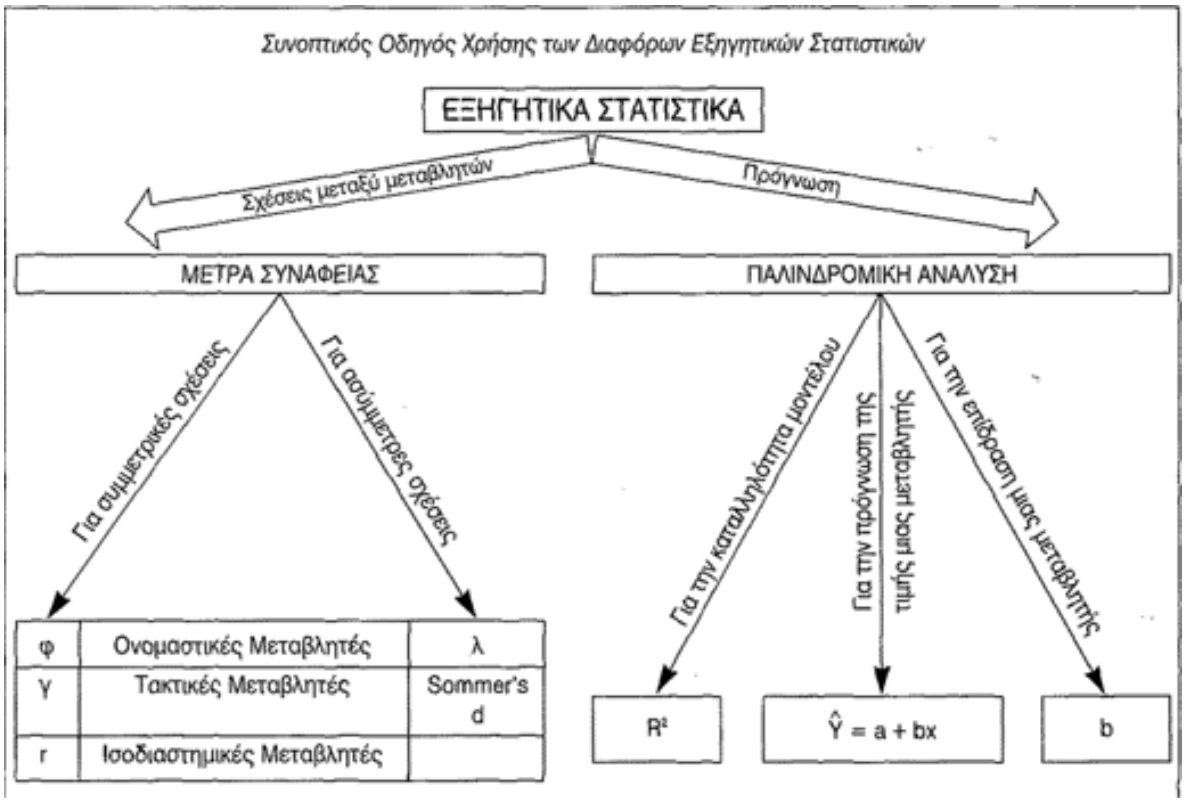
1. Η κανονική τυπική κατανομή έχει αριθμητικό μέσο 0 και τυπική απόκλιση 1.
2. Η κανονική τυπική κατανομή είναι συμμετρική, δηλαδή ο χώρος (ή το ποσοστό ή η πιθανότητα) από το 0 μέχρι το +1 είναι ακριβώς ο ίδιος με αυτόν από το 0 μέχρι το –1.
3. Η πιθανότητα ότι μια z τιμή πέφτει μεταξύ 0 και 1 είναι περί που 0,34, μεταξύ 0 και 2 είναι 0,475 και μεταξύ 0 και 3 περίπου 0,499.

Αν τα συνδυάσουμε όλα αυτά οδηγούμαστε στην διατύπωση του ***Εμπειρικού Κανόνα*** σύμφωνα με τον οποίο, το *68% όλων των τιμών μιας κανονικής κατανομής περιέχονται μεταξύ μιας τυπικής απόκλισης πάνω και κάτω από τον αριθμητικό μέσο, περίπου 95% περιέχονται μεταξύ δύο τυπικών αποκλίσεων και σχεδόν όλες οι τιμές (99%) περιέχονται μεταξύ τριών τυπικών αποκλίσεων πάνω και κάτω από τον αριθμητικό μέσο.*

ΚΕΦΑΛΑΙΟ 8: ΕΞΗΓΗΤΙΚΑ ΣΤΑΤΙΣΤΙΚΑ

8.1 ΕΞΗΓΗΤΙΚΑ ΣΤΑΤΙΣΤΙΚΑ

Η γνώση των σχέσεων μεταξύ μεταβλητών μας επιτρέπει να εξηγήσουμε ή να προβλέψουμε τη μία από την άλλη, γεγονός που είναι και ο κύριος στόχος του μεγαλύτερου μέρους της επιστημονικής έρευνας. Τα στατιστικά που περιγράφουν συλλογικά τις σχέσεις μεταξύ μεταβλητών ονομάζονται «**εξηγητικά**».



8.2 ΜΕΤΡΑ ΣΥΝΑΦΕΙΑΣ

Οι *σχέσεις μεταξύ δύο ή περισσότερων μεταβλητών* μπορεί να μετρηθούν με διάφορα στατιστικά. Τα κυριότερα από αυτά παρουσιάζονται παρακάτω. Γενικά, τα στατιστικά αυτά, που είναι γνωστά και ως **μέτρα *ή συντελεστές συνάφειας***, λαμβάνουν τιμές από -1 μέχρι +1. Το 0 δείχνει παντελή έλλειψη σχέσης, το +1 δείχνει τέλεια θετική σχέση και το -1 τέλεια αρνητική σχέση. Όλοι οι συντελεστές συνάφειας δείχνουν την ισχύ της σχέσης μεταξύ δυο μεταβλητών.

Η επιλογή του κατάλληλου μέτρου εξαρτάται από διαφόρους παράγοντες, εκ των οποίων δύο είναι πολύ σημαντικοί: *το* επίπεδο μέτρησης της μεταβλητής και το *αν η σχέση είναι* συμμετρική *ή* ασύμμετρη**.** Μία σχέση είναι *συμμετρική*, αν αυτό που εξετάζουμε είναι απλά η σχέση μεταξύ μεταβλητών ανεξάρτητα από την κατεύθυνση της σχέσης αυτής. Αντίθετα, μια σχέση είναι *ασύμμετρη*, αν εξετάζουμε την επίδραση μιας μεταβλητής σε μια άλλη.

Περιληπτικά, η διαδικασία επιλογής και υπολογισμού του κατάλληλου μέτρου σχέσης μεταξύ μεταβλητών έχει ως εξής: Βασισμένοι στις ερωτήσεις της έρευνας και την μέτρηση των μεταβλητών αποφασίζουμε, αν η σχέση είναι συμμετρική ή ασύμμετρη και ποιο είναι το επίπεδο μέτρησης.

8.3 ΘΕΤΙΚΕΣ ΚΑΙ ΑΡΝΗΤΙΚΕΣ ΣΧΕΣΕΙΣ ΜΕΤΑΞΥ ΜΕΤΑΒΛΗΤΩΝ

Η σχέση μεταξύ δύο μεταβλητών δύναται να είναι θετική ή αρνητική. Όταν οι σειρές ή οι τιμές *και των δύο μεταβλητών* αυξάνονται ή μειώνονται ταυτόχρονα, τότε έχουμε μια **θετική** σχέση. Όταν η σειρά ή η τιμή *της μιας αυξάνεται, ενώ της άλλης μειώνεται,* τότε έχουμε **αρνητική** σχέση. Για παράδειγμα, αν εξετάσουμε τη σχέση μεταξύ της εκπαίδευσης (δημοτικό, γυμνάσιο, λύκειο**,** ανώτερη ή ανώτατη σχολή) και εισοδήματος (χαμηλό, μεσαίο ή υψηλό) θα δούμε ότι τα υψηλότερα εισοδήματα τείνουν να συνυπάρχουν με τα υψηλότερα επίπεδα εκπαίδευσης και αντιστρόφως. Σε αυτή την περίπτωση έχουμε μια θετική σχέση. Αν, όμως, εξετάσουμε τη σχέση μεταξύ του χρόνου που ξοδεύουν οι μαθητές παίζοντας και της σχολικής επίδοσης μπορεί να βρούμε ότι όσο πιο πολύ χρόνο ξοδεύουν με παιχνίδι, τόσο χαμηλότερη θα είναι η επίδοση τους. Μια τέτοια σχέση θα είναι αρνητική. *Συνοψίζοντας, όταν δύο μεταβλητές αλλάζουν προς την ίδια κατεύθυνση, η σχέση τους είναι θετική και όταν αλλάζουν προς αντίθετη κατεύθυνση, η σχέση τους είναι αρνητική.*

Αρνητικές και θετικές σχέσεις υπάρχουν για όλες τις μεταβλητές. Για τις ονομαστικές μεταβλητές, όμως, το πρόσημο μιας σχέσης δεν έχει καμιά αξία γιατί το τί καθιστά "άνω" και "κάτω" είναι εντελώς αυθαίρετα. Για παράδειγμα, αν εξετάσουμε τη σχέση μεταξύ φύλου και εισοδήματος, θα βρούμε ότι οι άνδρες έχουν υψηλότερο εισόδημα. Επειδή, όμως, το φύλο είναι ονομαστική μεταβλητή (δηλαδή η μια κατηγορία δεν είναι ανώτερη ή κατώτερη από την άλλη - απλώς διαφορετική) δεν μπορούμε να πούμε ότι η σχέση είναι αρνητική ή θετική.

Είναι απαραίτητο να τονιστεί ότι αυτό ισχύει για την ερμηνεία της σχέσης ονομαστικών μεταβλητών και όχι για τον υπολογισμό των στατιστικών. Όταν υπολογίζονται οι σχέσεις μεταξύ ονομαστικών μεταβλητών, το αποτέλεσμα θα είναι μια θετική ή μια αρνητική σχέση, ανάλογα με το πώς κωδικοποιήθηκε η μεταβλητή. Έτσι, αν στην κωδικοποίηση είχαμε δώσει στις γυναίκες τον αριθμό 1 και στους άνδρες 0, η παραπάνω σχέση θα ήταν αρνητική. Αν είχαμε κωδικοποιήσει τη

μεταβλητή αντίστροφα, η σχέση θα ήταν θετική. Επειδή όμως, αυτό είναι καθαρά αυθαίρετο, ο χαρακτηρισμός τέτοιων σχέσεων ως θετικών ή αρνητικών δεν έχει νόημα.

8.4 ΓΡΑΜΜΙΚΕΣ ΚΑΙ ΜΗ ΓΡΑΜΜΙΚΕΣ ΣΧΕΣΕΙΣ ΜΕΤΑΞΥ ΜΕΤΑΒΛΗΤΩΝ- ΑΠΛΗ ΓΡΑΜΜΙΚΉ ΠΑΛΙΝΔΡΟΜΗΣΗ (SIMPLE LINEAR REGRESSION)

Στην στατιστική, ως **Γραμμική Παλινδρόμηση** ορίζεται η μοντελοποίηση της σχέσης μιας απλής εξαρτημένης μεταβλητής Y με μια ή περισσότερες ανεξάρτητες μη ερμηνευτικές μεταβλητές X_1, $X_{2,}$…, X_n. Η μεταβλητή X_i δεν θεωρείται πως είναι τυχαία ενώ η Y θεωρείται τυχαία μεταβλητή. Όταν υπάρχει μόνο μια ανεξάρτητη μεταβλητή X τότε η μοντελοποίηση ονομάζεται ***απλή γραμμική παλινδρόμηση (simple linear regression)***.

Απαραίτητη προϋπόθεση για τη χρήση του συντελεστή συσχέτισης είναι η σχέση μεταξύ των μεταβλητών να είναι γραμμική (ευθύγραμμη). **Γραμμική σχέση** υπάρχει όταν η αλλαγή της μιας μεταβλητής (Y) είναι σταθερή για κάθε μονάδα αλλαγής της άλλης μεταβλητής (X), ανεξαρτήτως του επιπέδου της X. Πράγμα που σημαίνει ότι σε μια γραμμική σχέση, η μεταβλητή Y θα αλλάξει κατά β (μία σταθερά), άσχετα αν η μεταβλητή X πάει από 2 σε 3, από 10 σε 11 ή από 333 σε 334. Η σχέση αυτή ονομάζεται γραμμική γιατί, αν την παραστήσουμε γραφικά, όλα τα σημεία της Ψ που αντιστοιχούν στις διάφορες τιμές της X σχηματίζουν μια *ευθεία γραμμή*. Αλγεβρικά, ο ορισμός της γραμμής αυτής είναι **$\alpha+\beta X_i$.** Κατά συνέπεια, η γραμμική σχέση μεταξύ των μεταβλητών X και Y παριστάνεται με την εξίσωση $Y_i = \alpha+\beta X_i$.

8.5 ΘΕΤΙΚΕΣ ΚΑΙ ΑΡΝΗΤΙΚΕΣ ΓΡΑΜΜΙΚΕΣ ΣΧΕΣΕΙΣ

Μια γραμμική σχέση μπορεί να είναι θετική ή αρνητική. Μια σχέση, όπως είπαμε πιο πάνω είναι θετική, αν η μεταβολή και των δύο μεταβλητών είναι προς την ίδια κατεύθυνση, και αρνητική αν είναι προς διαφορετική κατεύθυνση. Δηλαδή, αν η τιμή της Y_i αυξάνεται όπως αυξάνεται και η τιμή της X_i (ή ελαττώνεται όπως ελαττώνεται και η τιμή της X_i) τότε η σχέση είναι θετική. Αν η τιμή Y_i αυξάνεται όπως ελαττώνεται η τιμή της X_i (ή και αντίστροφα) τότε η σχέση είναι αρνητική. Οι μη γραμμικές σχέσεις είναι συνήθως και θετικές και αρνητικές, ή καλύτερα, αρχίζουν ως θετικές ή αρνητικές σχέσεις και στη συνέχεια αλλάζουν πρόσημο.

Η σχέση μεταξύ προσπάθειας και σχολικής επίδοσης είναι καθαρότατα μια θετική σχέση, αφού όσο πιο πολύ προσπαθήσει κάποιος (εντός λογικών ορίων) τόσο καλύτερα θα αποδώσει στο σχολείο. Αντίθετα, ο αριθμός των απουσιών και η σχολική επίδοση έχουν αρνητική σχέση, αφού όσο μεγαλύτερος είναι ο αριθμός απουσιών τόσο χαμηλότερη θα είναι η επίδοση.

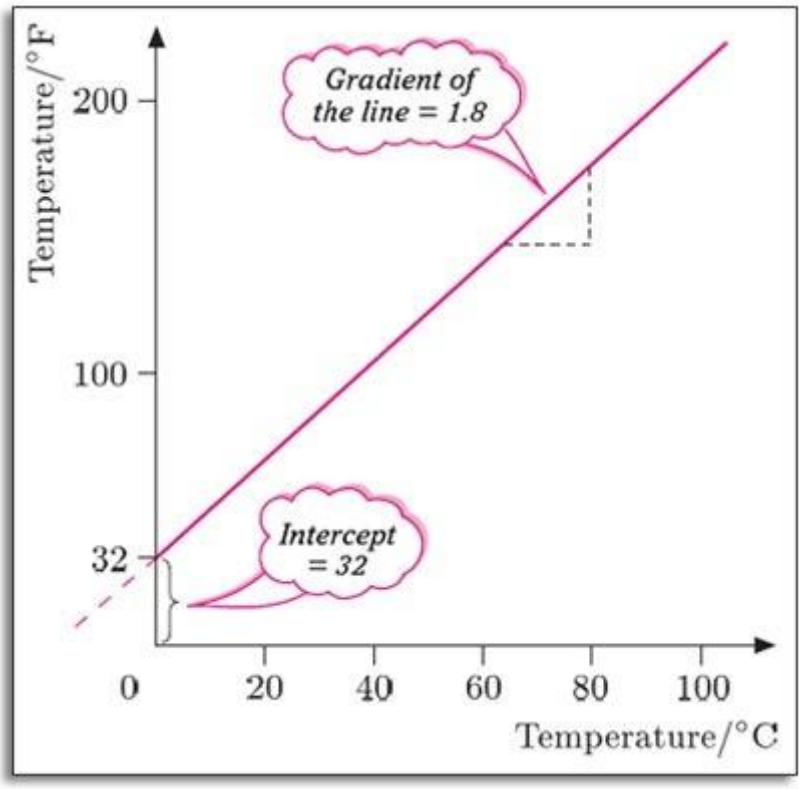
8.6 ΤΕΛΕΙΕΣ ΚΑΙ ΜΗ ΤΕΛΕΙΕΣ ΓΡΑΜΜΙΚΕΣ ΣΧΕΣΕΙΣ

Η εξίσωση $Y_i = a + \mathrm{b}\, X_i$ παριστάνει μια τέλεια γραμμική σχέση, γιατί για κάθε μεταβολή της X_i γνωρίζουμε ακριβώς την μεταβολή στην Y_i. Χρησιμοποιώντας γραφικές παραστάσεις, μια γραμμική σχέση είναι τέλεια, αν οι τιμές της μεταβλητής Y_i για κάθε τιμή της X_i πέφτουν ακριβώς πάνω σε μια ευθεία. Ένα παράδειγμα τέλειας γραμμικής σχέσης είναι η σχέση μεταξύ βαθμών Φαρενάιτ (F) και Κελσίου (C). Η σχέση των δύο καθορίζεται από την εξίσωση: $F_i = 32 + \frac{9}{5} C_i$

όπου α = 32 και β = 9/5=1,8, F_i = βαθμοί Φαρενάιτ και C_i = βαθμοί Κελσίου.

Ερμηνεύοντας την παραπάνω εξίσωση συμπεραίνουμε ότι όταν το C_i είναι 0 βαθμοί κελσίου, το F_i είναι 32 βαθμοί Φαρενάιτ και ότι για κάθε βαθμό Κελσίου έχουμε ακριβώς 9/5=1,8 επιπλέον βαθμών Φαρενάιτ.

Η τέλεια θετική γραμμική σχέση απεικονίζεται παρακάτω:

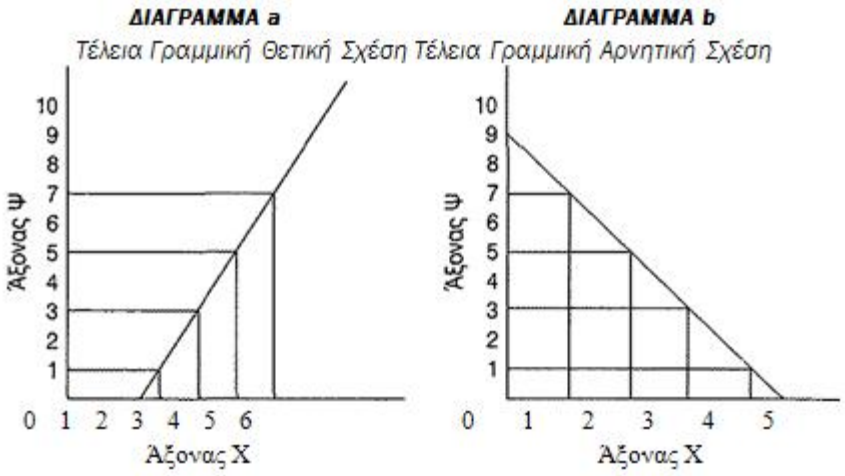


Στις κοινωνικές επιστήμες οι σχέσεις μεταξύ μεταβλητών δεν είναι σχεδόν ποτέ τέλειες, δηλαδή δεν υφίσταται απόλυτη ανταπόκριση μεταξύ των τιμών τους. Αλγεβρικά αυτές οι σχέσεις παριστάνονται με εξισώσεις του τύπου:

$$Y_i = a + \mathrm{b}\, X_i + e_i$$

όπου το e_i παριστάνει την απόσταση της τιμής της μεταβλητής από την ευθεία. Αν παρασταθούν γραφικά τέτοιες σχέσεις, τα σημεία τους παρουσιάζουν (περίπου) το σχήμα μιας ευθείας (ή καμπύλης σε περίπτωση μη γραμμικής σχέσης), αλλά δεν

πέφτουν όλα πάνω σε μία ευθεία γραμμή. Όσο πιο καλά διακρίνεται η ευθεία γραμμή, δηλαδή όσο πιο κοντά στη γραμμή βρίσκονται τα σημεία, τόσο ισχυρότερη είναι η σχέση. Ακολούθως παρατηρούμε δύο διαγράμματα Τέλειας Γραμμικής Θετικής και Αρνητικής Σχέσης:



ΚΕΦΑΛΑΙΟ 9: ΜΕΛΕΤΕΣ ΠΕΡΙΠΤΩΣΗΣ- ΠΑΡΑΔΕΙΓΜΑΤΑ ΜΕ ΤΗΝ ΧΡΗΣΗ ΤΟΥ SPSS

Η επεξεργασία των δεδομένων και η παράθεση των παρακάτω παραδειγμάτων διεξήχθη με τη χρήση του ***στατιστικού πακέτου για κοινωνικές επιστήμες SPSS 19***. Τα ακόλουθα παραδείγματα παρουσιάζουν την εφαρμογή ορισμένων ελέγχων στην στατιστική επιστήμη, μέσω του SPSS. Πιο συγκεκριμένα δίνονται παραδείγματα ***ελέγχου κανονικότητας*** των μεταβλητών μιας έρευνας (απόρριψη ή αποδοχή μηδενικής υπόθεσης H_0) και παραδείγματα για τον ***έλεγχο υπόθεσης μέσων τιμών*** π.χ. του πληθυσμού (απόρριψη ή αποδοχή μηδενικής υπόθεσης H_0). Παράλληλα στο *Παράρτημα* μπορούμε να δούμε πώς καταχωρούμε τα δεδομένα στο φύλλο εργασίας[1] (*Dataset*) του SPSS.

9.1 ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ ΤΩΝ ΜΕΤΑΒΛΗΤΩΝ

Αρχικά εισάγουμε τις μεταβλητές και τα δεδομένα που έχουμε συλλέξει στο SPSS. Στην συνέχεια απαιτείται *έλεγχος κανονικότητας των μεταβλητών* αυτών και ακολουθεί η επεξεργασία και η ανάλυση των αποτελεσμάτων, Για να ελέγξουμε αν η κατανομή μιας μεταβλητής είναι συμβατή με την κανονική εφαρμόζουμε το **Test Kolmogorov-Smirnov (1- Sample K-S).**

Πριν ξεκινήσουμε καλό είναι να δούμε τι σημαίνει μηδενική υπόθεση H_0. ***Μηδενική υπόθεση*** είναι ότι η υπό έλεγχο κατανομή δεν διαφέρει από την κανονική κατανομή έναντι της ***Εναλλακτικής υπόθεσης*** H_1 η οποία αποδέχεται ότι η υπό έλεγχο κατανομή διαφέρει από την κανονική κατανομή.

Ο έλεγχος κανονικότητας των μεταβλητών μιας έρευνας και η απόρριψη ή αποδοχή της μηδενικής υπόθεσης, είναι απαραίτητα πριν την επεξεργασία των

[1] Βλ. Παράρτημα

δεδομένων. Ο έλεγχος αυτός μπορεί να γίνει μέσω του SPSS δίνοντας του κάποιες εντολές και εφαρμόζοντας τα κατάλληλα κάθε φορά τεστ. Μπορεί να γίνει στατιστικά με το τεστ Kolmogorov – Smirnov αλλά και γραφικά. Επίσης μπορεί να γίνει και έλεγχος υπόθεσης για τη Μέση Τιμή ενός πληθυσμού (*1-Sample T-test*). Όλα αυτά παρουσιάζονται αναλυτικά στη συνέχεια.

§ *Έλεγχος κανονικότητας με SPSS*

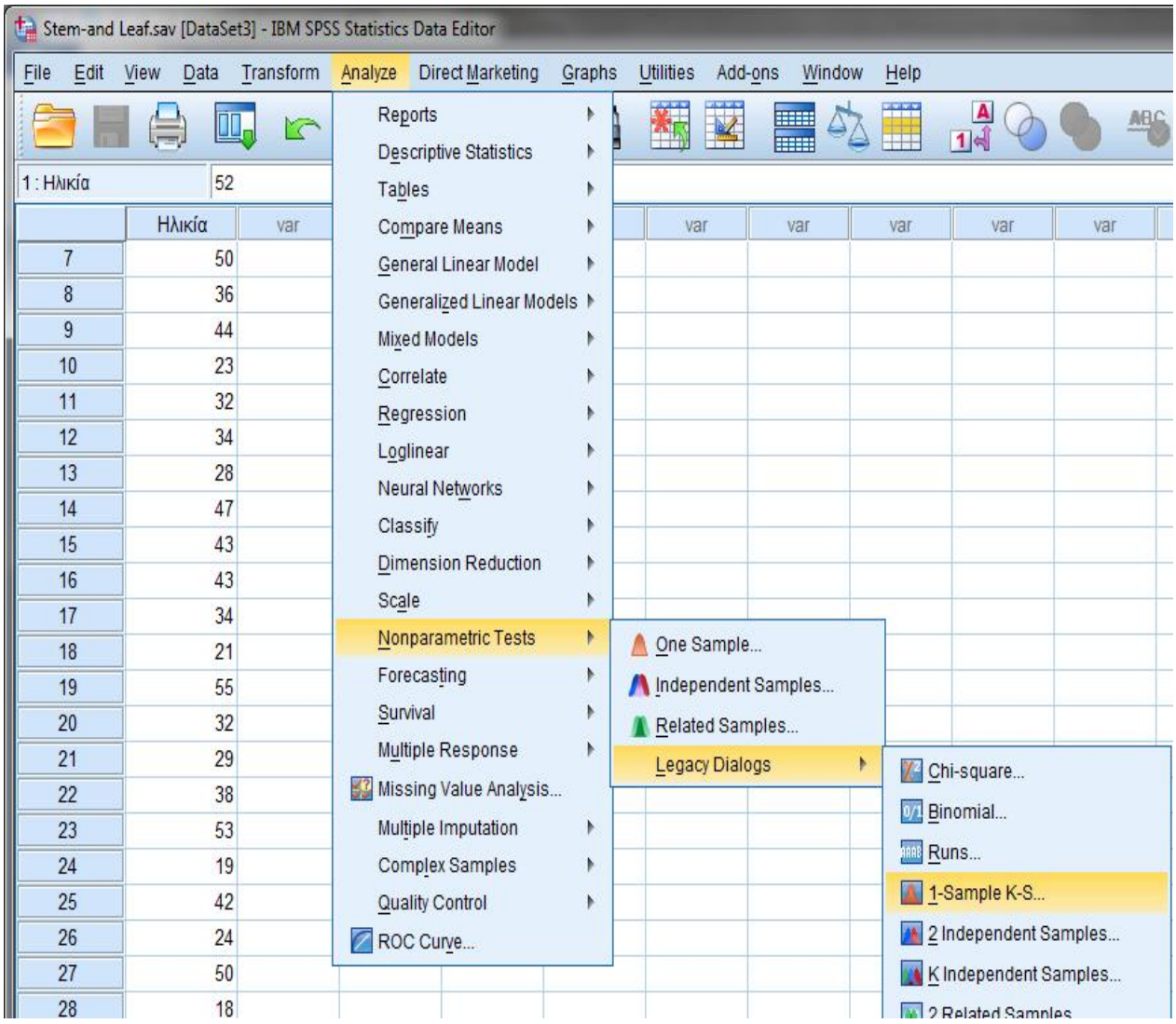
Θα ελέγξουμε αν ισχύει η υπόθεση της κανονικότητας των δεδομένων στο SPSS μέσω της εντολής **Analyze/ Nonparametric Tests/ 1- Sample K-S** (*Μη-Παραμετρικός Έλεγχος*). Τα δεδομένα τα αντλήσαμε από το βιβλίο της ***Χριστίνας Νόβα- Καλτσούνη «Μεθοδολογία Εμπειρικής Έρευνας στις Κοινωνικές Επιστήμες- Ανάλυση δεδομένων με την χρήση του SPSS 13***». Πιο συγκεκριμένα για τον έλεγχο Κανονικότητας Kolmogorov- Smirnov χρησιμοποιήσαμε το αρχείο *Stem-and Leaf.sav* που περιέχει μόνο μια μεταβλητή, την Ηλικία.

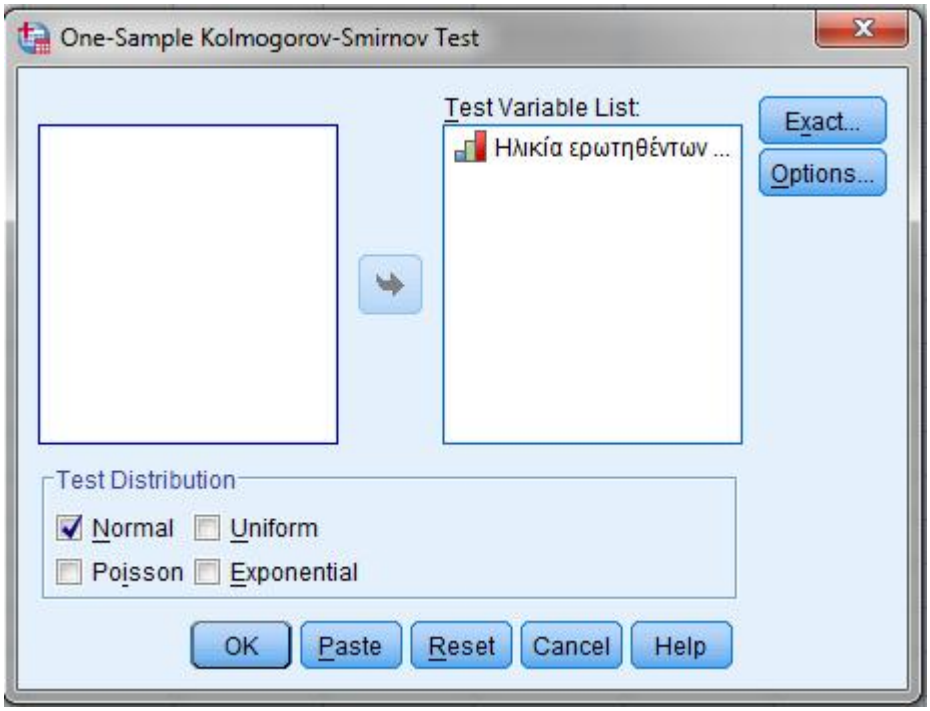
Η μηδενική υπόθεση του ελέγχου υπόθεσης της κανονικότητας Kolmogorov-Smirnov είναι ότι η κατανομή των δεδομένων δε διαφέρει από την Κανονική κατανομή ενώ η εναλλακτική υπόθεση είναι ότι η κατανομή των δεδομένων διαφέρει από την Κανονική κατανομή. Το επίπεδο στατιστικής σημαντικότητας συνήθως το ορίζουμε σε a=0,05 ή 5%. Το ***παρατηρηθέν επίπεδο στατιστικής σημαντικότητας p- value*** ορίζεται ως η πιθανότητα η τιμή του ελέγχου (ελεγχοσυνάρτησης) να πάρει μία τιμή τόσο ακραία ή περισσότερο ακραία από αυτή που πήρε στο συγκεκριμένο δείγμα κάτω από τη μηδενική υπόθεση.

Αν η p-value είναι μικρότερη του 0,05, τότε λέμε ότι η μηδενική υπόθεση απορρίπτεται. Αν η p-value είναι μεγαλύτερη ή ίση του 0,05, τότε λέμε ότι η μηδενική υπόθεση δεν απορρίπτεται. Το SPSS εμφανίζει τις τιμές των παρατηρηθέντων επιπέδων στατιστικής σημαντικότητας και τις ονομάζει (***Asymptotic) Significances***.

Ο λόγος που χρειαζόμαστε την κανονικότητα των δεδομένων, είναι για να έχουν ισχύ κάποιες στατιστικές τεχνικές που χρησιμοποιούμε, όπως οι έλεγχοι υποθέσεων για τους Μέσους, η Γραμμική Παλινδρόμηση κ.ά.

Στη συνέχεια βλέπουμε τα *μονοπάτια εντολών* που δίνουμε στο SPSS για να εκτελέσει τον έλεγχο κανονικότητας K-S:





Τα αποτελέσματα (***Output***) του SPSS είναι το ακόλουθα:

One-Sample Kolmogorov-Smirnov Test

		Ηλικία ερωτηθέντων
N		185
Normal Parameters[a,b]	Mean	37,05
	Std. Deviation	10,695
Most Extreme Differences	Absolute	,072
	Positive	,072
	Negative	-,063
Kolmogorov-Smirnov Z		,974
Asymp. Sig. (2-tailed)		**,299**

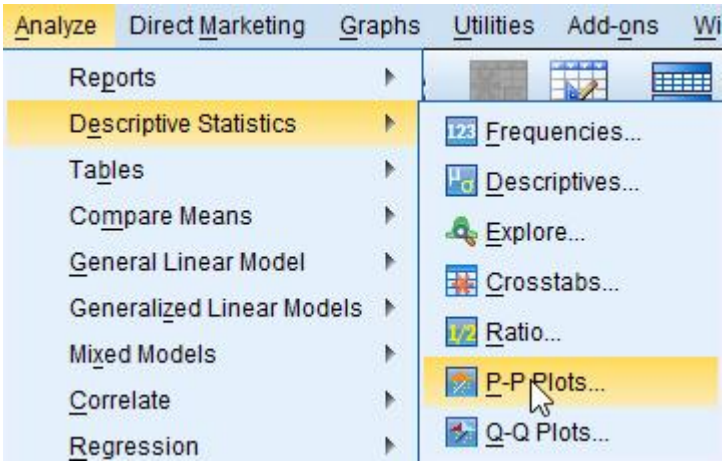
a. Test distribution is Normal.

b. Calculated from data.

Παρατηρούμε ότι η έντονη τιμή της τελευταίας γραμμής [Asymp.Sig.(2-tailed)] που αντιστοιχεί στο παρατηρούμενο επίπεδο σημαντικότητας (p-value) για τον αμφίπλευρο έλεγχο είναι ίση με 0,299, άρα μεγαλύτερη από το επίπεδο σημαντικότητας α=0,05 οπότε ***δεν απορρίπτουμε την υπόθεση της κανονικότητας για αυτήν την μεταβλητή.*** Αυτό το αποτέλεσμα είναι επιθυμητό και αν θέλουμε να εφαρμόσουμε περαιτέρω ελέγχους θα χαρακτηρίζονται από εγκυρότητα αποτελεσμάτων. *Εναλλακτικά, στην περίπτωση που δεν ισχύει η υπόθεση της Κανονικότητας μπορούμε να καταφύγουμε στη χρήση* **μη-παραμετρικών μεθόδων**[2].

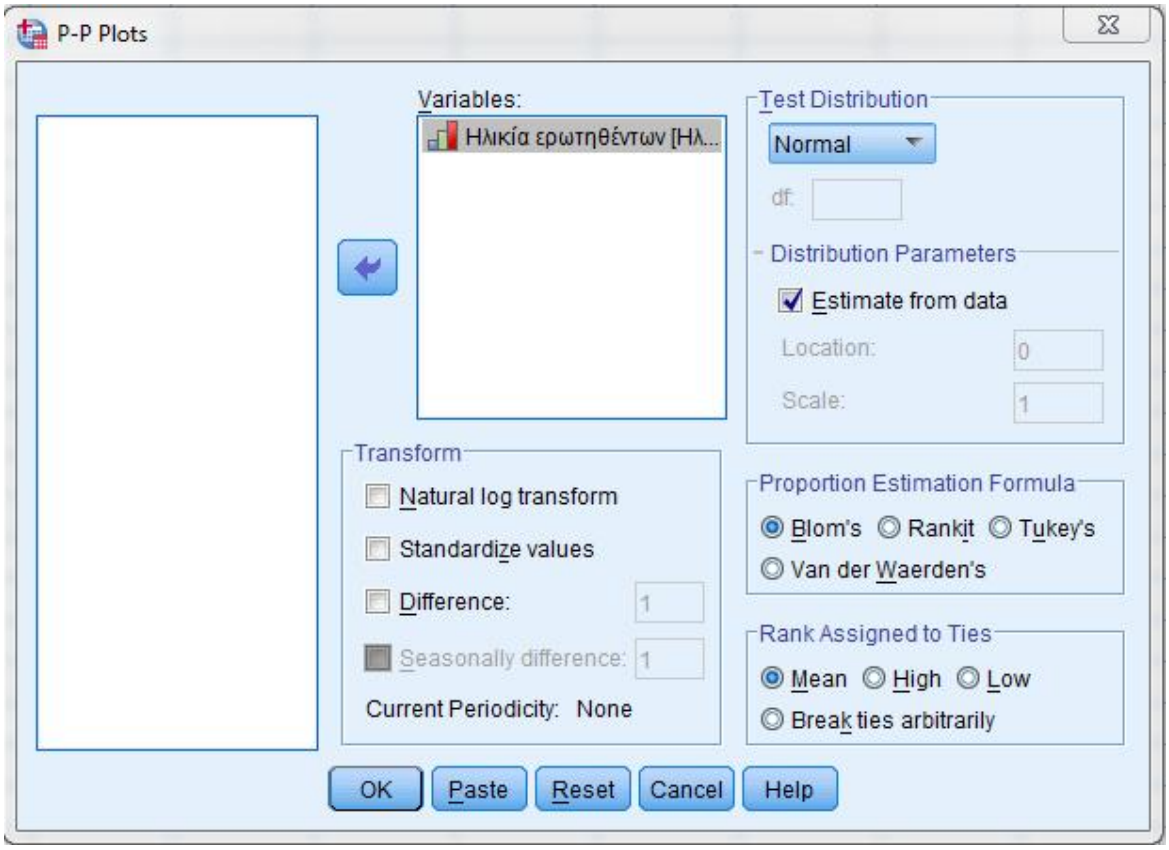
§ *Γραφικός έλεγχος κανονικότητας*

Ο έλεγχος γραφικά γίνεται μέσω της εντολής **Analyze/ Descriptive Statistics/ P-P Plots.**

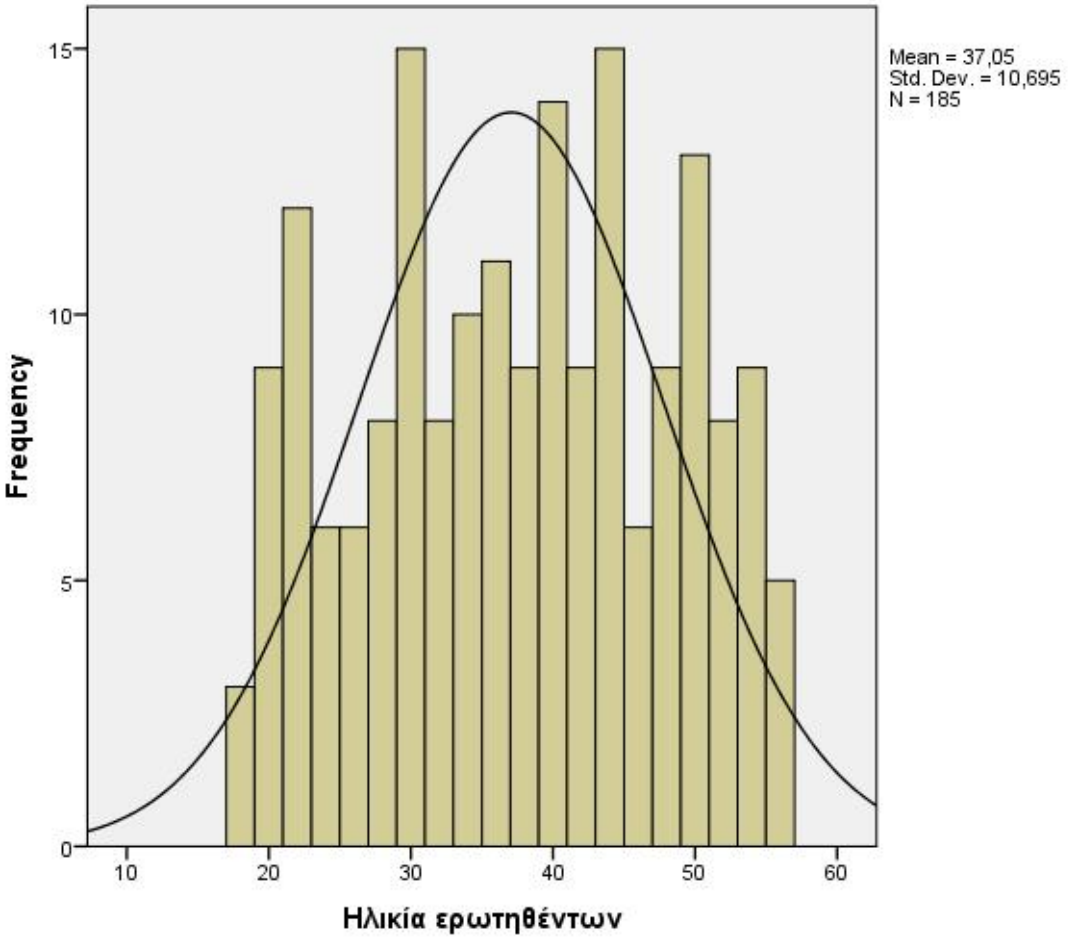


Αρχικά τοποθετούμε τη μεταβλητή που θέλουμε να ελέγξουμε αν είναι κανονική στο πλαίσιο *Variables* και στο πεδίο *Test distribution* επιλέγουμε *Normal* και κατόπιν πατάμε OK.

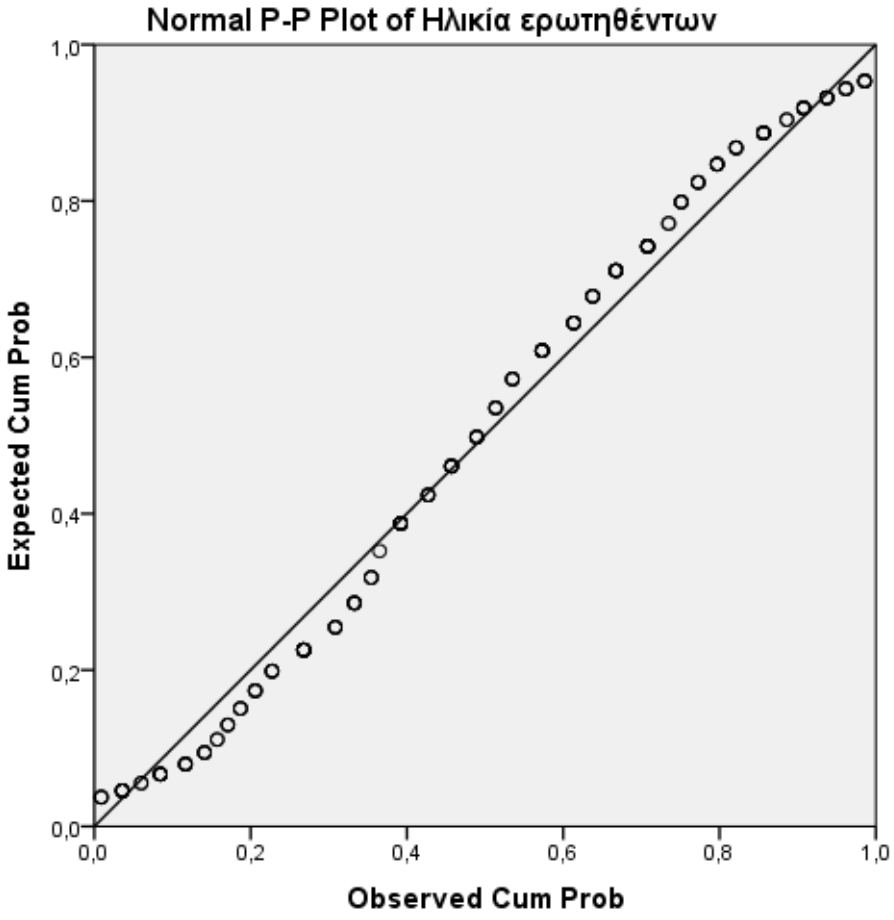
[2] Βλ. Παράρτημα



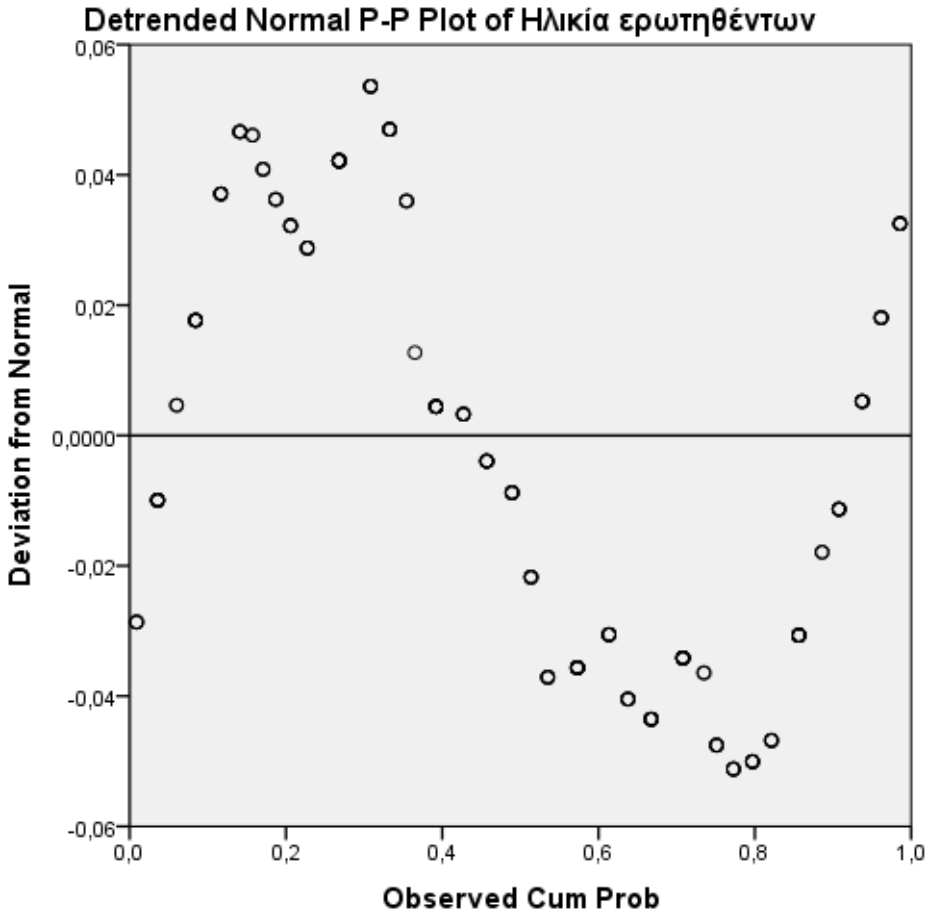
Παρακάτω βλέπουμε το *Ιστόγραμμα* της μεταβλητής με κάποια περιγραφικά μέτρα όπως την μέση τιμή (Mean), την τυπική απόκλιση (Std. Dev.) και το πλήθος των παρατηρήσεων (N). Η μαύρη γραμμή παριστάνει την καμπύλη της κανονικότητας ή αλλιώς την «καμπάνα».



Το συμπέρασμα που προκύπτει από το Ιστόγραμμα, όπως και από τα επόμενα γραφήματα P-P Plot και Detrended P-P Plot είναι <u>ότι η Κανονικότητα της μεταβλητής «Ηλικία ερωτηθέντων» είναι σε ισχύ</u>.



Παρατηρούμε πως τα σημεία βρίσκονται πάνω ή πολύ κοντά στην ευθεία γραμμή, και πιο ειδικά στο μέσο της πρέπει τα σημεία (κουκκίδες) να είναι πάνω στην ευθεία. Η υπόθεση της Κανονικότητας ενισχύεται και από το ακόλουθο γράφημα ***Detrended Normal P-P Plot:***


Detrended Normal P-P Plot of Ηλικία ερωτηθέντων
Deviation from Normal
0,06
0,04
0,02
0,0000
-0,02
-0,04
-0,06
0,0
0,2
0,4
0,6
0,8
1,0
Observed Cum Prob

9.2 ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΗΣ ΓΙΑ ΤΗΝ ΜΕΣΗ ΤΙΜΗ ΕΝΟΣ ΔΕΙΓΜΑΤΟΣ (1- SAMPLE- T-TEST)- ΠΑΡΑΔΕΙΓΜΑ 1

Το μονοπάτι εντολών που ακολουθούμε στο στατιστικό πακέτο SPSS είναι: **Analyze → Compare Means → One-Sample T-Test**.

Με το *One-sample T-test* ελέγχουμε αν η μέση τιμή μ ενός πληθυσμού διαφέρει σημαντικά από κάποια συγκεκριμένη τιμή μ_0. Ελέγχουμε δηλαδή τη μηδενική υπόθεση H_0: $\mu = \mu_0$ έναντι της εναλλακτικής H_1: $\mu \neq \mu_0$ (δίπλευρος έλεγχος). Η τιμή μ_0 με την οποία θέλουμε να γίνει η σύγκριση, μπορεί να είναι γνωστή εμπειρικά ή από άλλες έρευνες. Μπορεί λόγου χάρη να προέρχεται από ένα άλλο δείγμα που λήφθηκε από τον ίδιο πληθυσμό ή από μια άλλη δειγματοληπτική έρευνα που καλύπτει ένα μεγαλύτερο τμήμα του πληθυσμού.

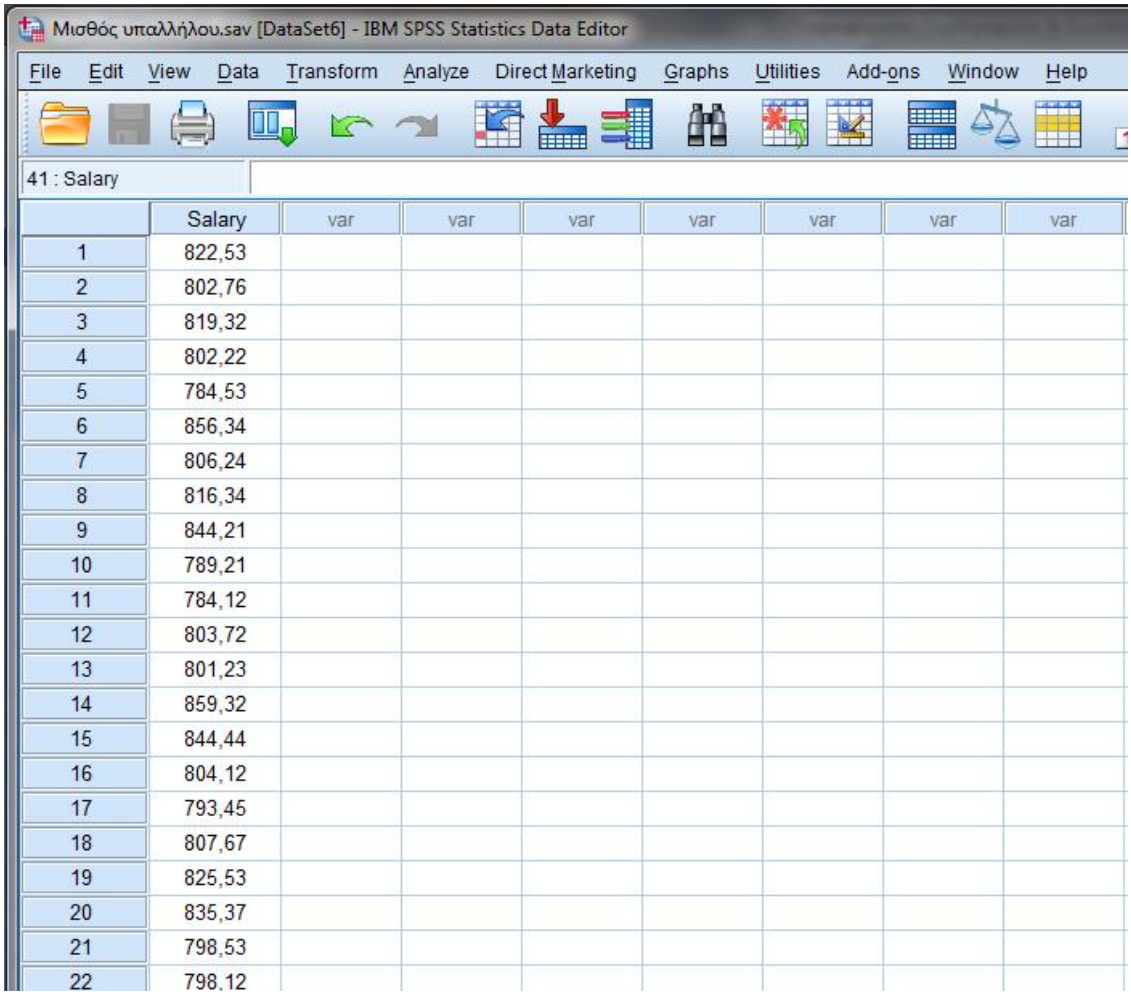
Στον στατιστικό έλεγχο υποθέσεων μας ενδιαφέρει να εξετάσουμε αν μια παράμετρος του πληθυσμού (για παράδειγμα μέση τιμή ή διακύμανση) ικανοποιεί μια υπόθεση (*Ho* → μηδενική) έναντι μια εναλλακτικής υπόθεσης (H_1 → εναλλακτική).

Στον παρακάτω πίνακα δίνονται οι μισθοί (σε χρηματικές μονάδες) 40 τυχαία επιλεγμένων υπαλλήλων που εργάζονται στον ιδιωτικό τομέα. Ζητείται να ελεγχθεί, σε επίπεδο σημαντικότητας a=5%, αν ο μέσος μισθός μ του πληθυσμού από τον οποίο προέρχεται το παραπάνω δείγμα είναι ίσος με μ_0 = 817 χρηματικές μονάδες.

Τα δεδομένα φαίνονται στη συνέχεια:

822,53	784,53	844,21	801,23	793,45	798,53	803,46	823,88	837,13	809,43
802,76	856,34	789,21	859,32	807,67	798,12	800,56	840,43	787,13	815,43
819,32	806,24	784,12	844,44	825,53	779,23	824,12	813,56	817,43	825,51
802,22	816,34	803,72	804,12	835,37	849,43	834,29	826,21	827,10	828,78

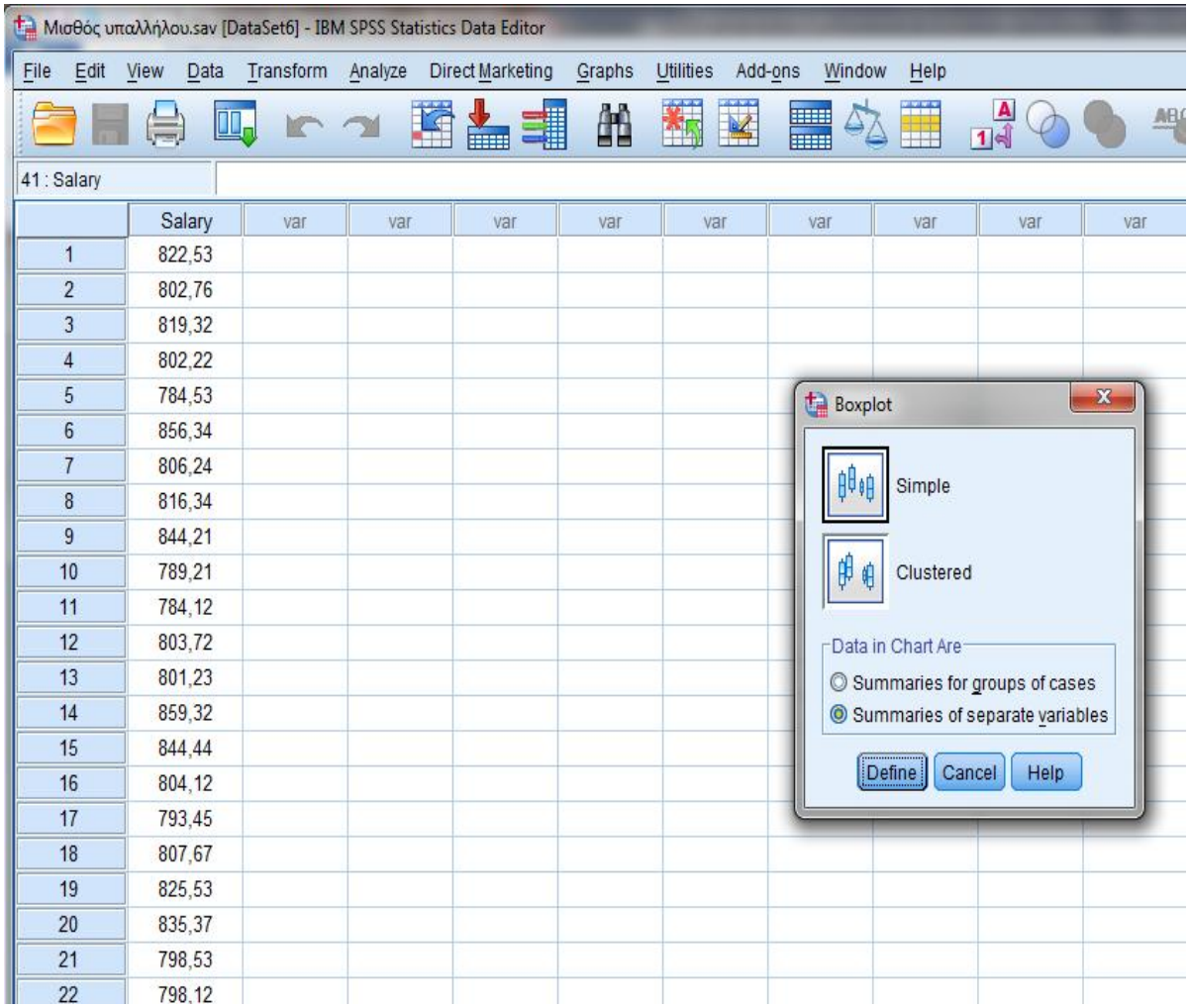
Ξεκινάμε δημιουργώντας την μεταβλητή *salary* στο *Variable View* στο SPSS και στη συνέχεια καταχωρούμε στο Data View τα δεδομένα του παραπάνω πίνακα. Το αρχείο που αποθηκεύουμε το ονομάζουμε «Μισθός υπαλλήλου.sav». Στο παρακάτω σχήμα διαφαίνεται το Data View όπως προκύπτει από την εισαγωγή των δεδομένων (40 περιπτώσεις).



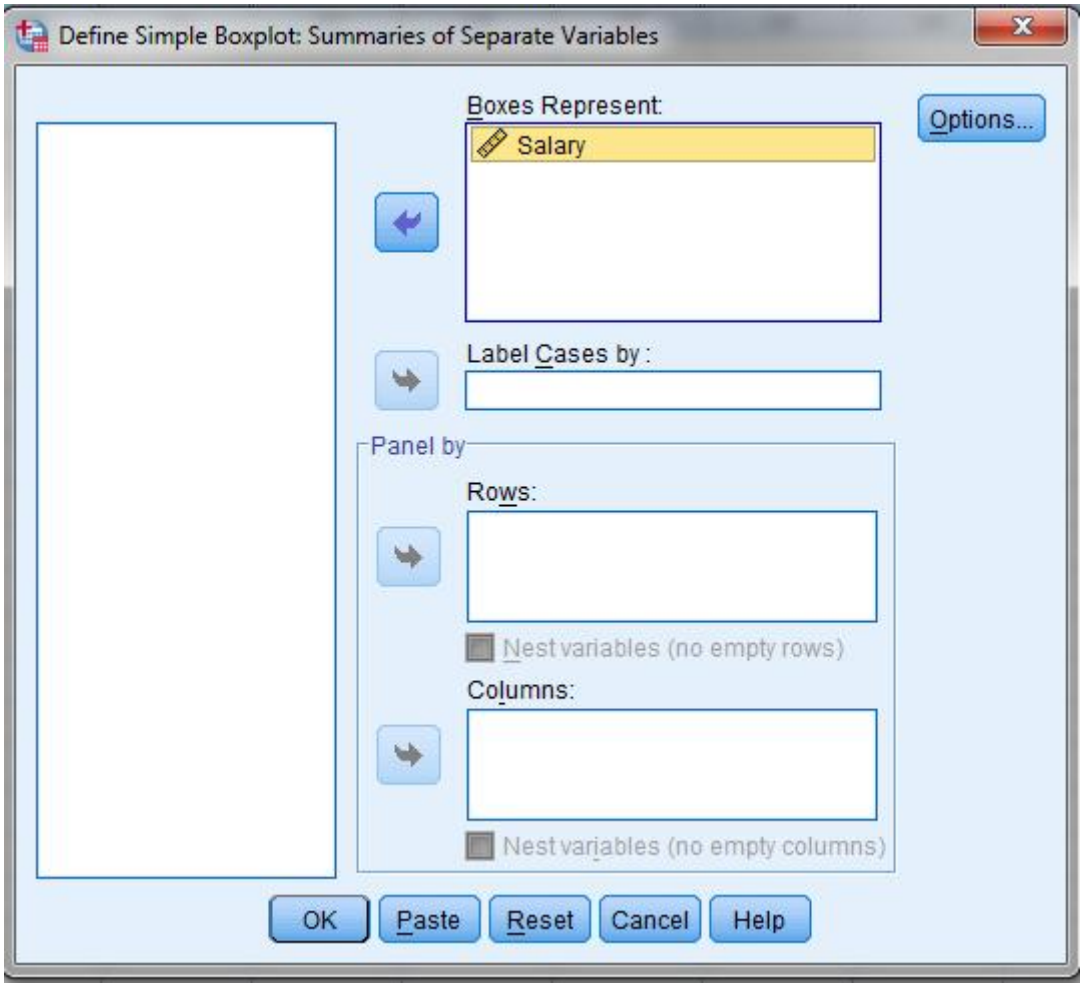
	Salary	var	var	var	var	var	var	var
1	822,53							
2	802,76							
3	819,32							
4	802,22							
5	784,53							
6	856,34							
7	806,24							
8	816,34							
9	844,21							
10	789,21							
11	784,12							
12	803,72							
13	801,23							
14	859,32							
15	844,44							
16	804,12							
17	793,45							
18	807,67							
19	825,53							
20	835,37							
21	798,53							
22	798,12							

Επίσης ελέγχουμε αν υπάρχουν *ακραίες παρατηρήσεις (outliers*). Ο έλεγχος θα γίνει γραφικά με τη βοήθεια του ***διαγράμματος πλαισίου- απολήξεων ή Θηκογράμματος (Boxplot***). Για να κατασκευαστεί το γράφημα αυτό δίνονται οι παρακάτω εντολές στο SPSS: **Graphs → Legacy Dialogs → Boxplot.**

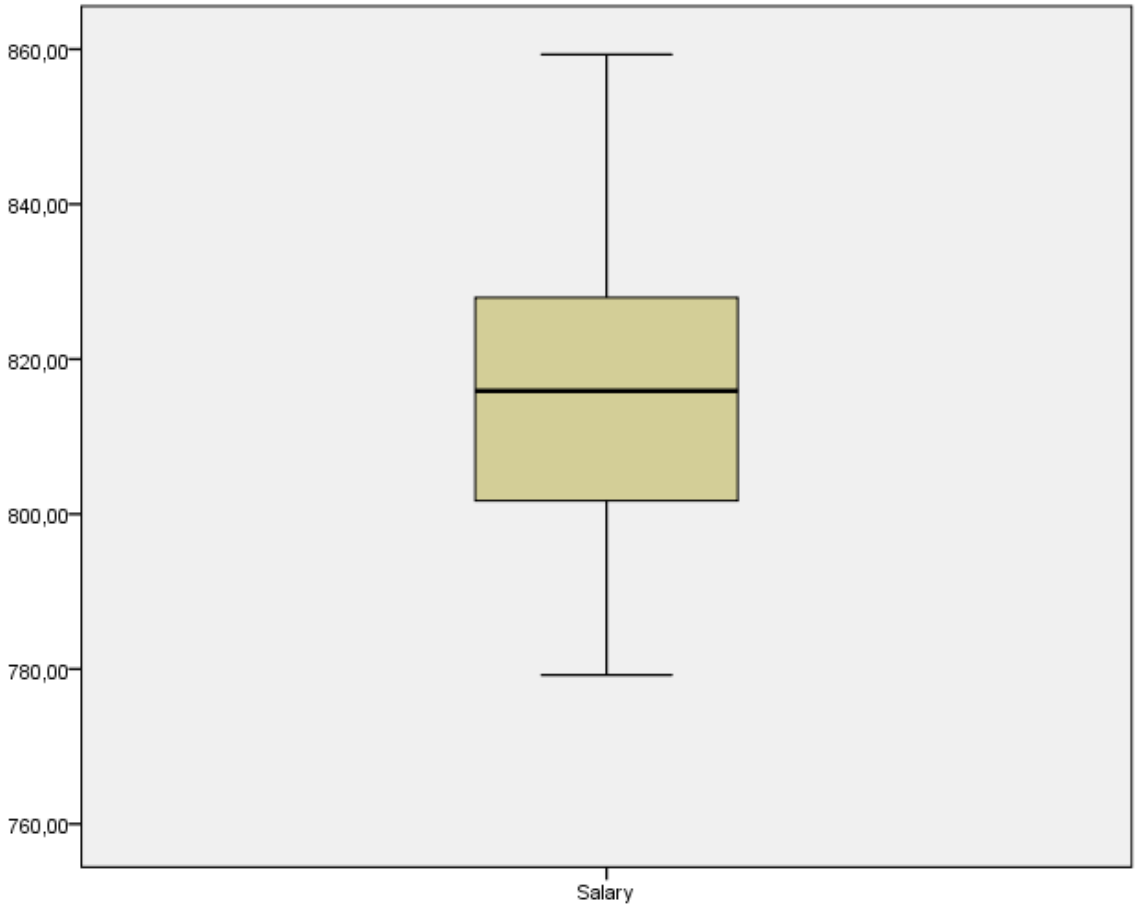
Στην συνέχεια στο παράθυρο που ακολουθεί επιλέγουμε **Simple** και **Summaries of seperate variables**, όπως φαίνεται στο σχήμα που ακολουθεί:



Τοποθετούμε την μεταβλητή **Salary** στο πλαίσιο **Boxes Represent**, όπως διαφαίνεται παρακάτω:

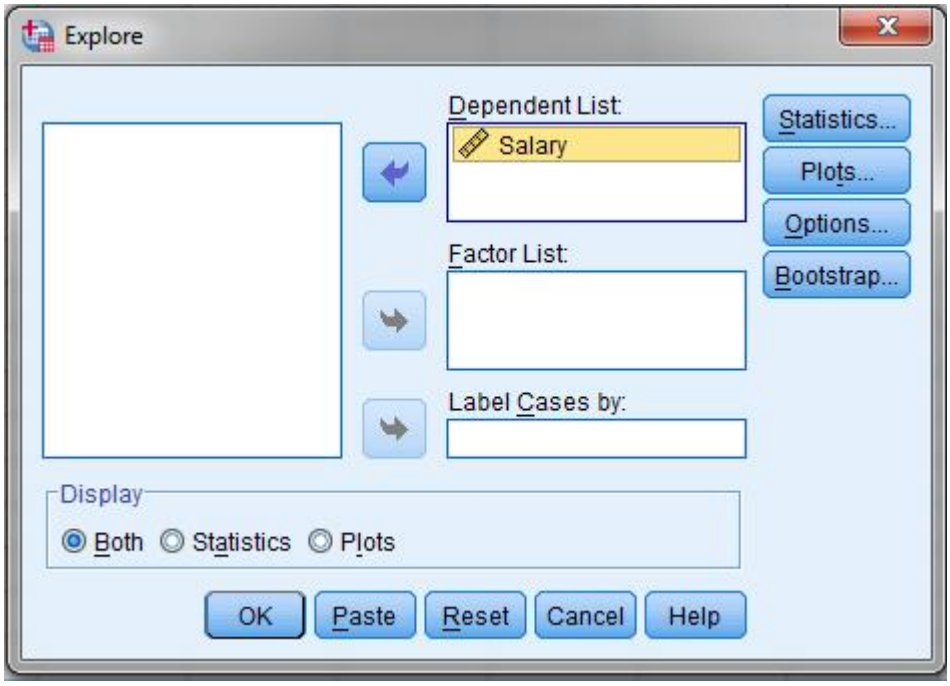


Στη συνέχεια επιλέγουμε OK και προκύπτει το Boxplot:

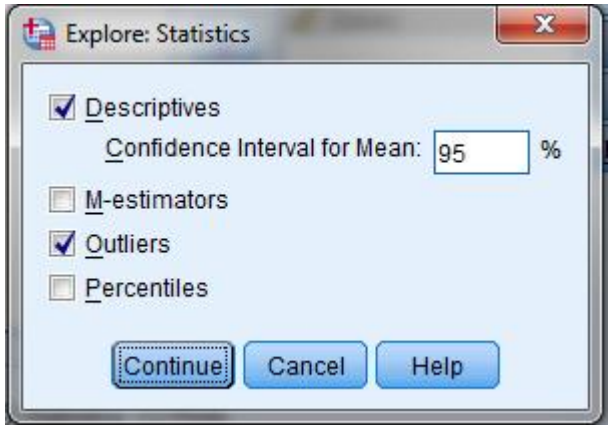


Παρατηρούμε ότι δεν υπάρχουν ακραίες τιμές. Σε περίπτωση που υπήρχαν, πάνω στο γράφημα θα σημειώνονταν με λευκή κουκκίδα και επίσης θα αναγραφόταν και ο αύξοντας αριθμός της τιμής αυτής δίπλα στην κουκκίδα.

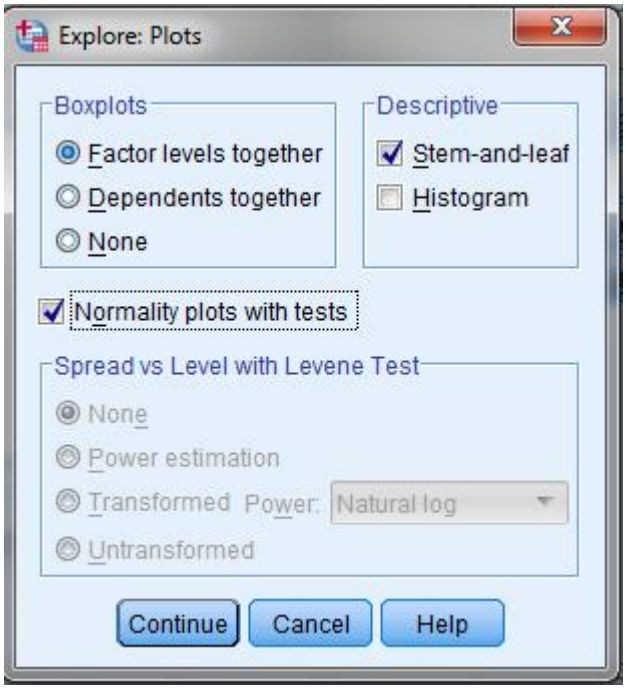
Όπως προαναφέρθηκε, ο έλεγχος κανονικότητας του δείγματος γίνεται τόσο γραφικά, όσο και στατιστικά με την χρήση κάποιων κριτηρίων. Ο έλεγχος με τα στατιστικά τεστ των Kolmogorov-Smirnov και Shapiro-Wilk γίνεται με τον παρακάτω τρόπο: Δίνονται στο SPSS οι εξής εντολές: **Analyze → Descriptive Statistics → Explore.**



Σε αυτό το παράθυρο ελέγχουμε αν έχει επιλεχθεί η ένδειξη ***Both***, κάτω αριστερά, που σημαίνει ότι το SPSS θα παραθέσει και στατιστικούς δείκτες και γραφήματα που μας βοηθούν στον έλεγχο της κανονικότητας. Στη συνέχεια μετακινούμε την μεταβλητή ***Salary*** δεξιά στο πλαίσιο ***Dependent List***, και επιλέγουμε ***Statistics, Plots, Options***. Στο πλαίσιο Statistics επιλέγουμε την ένδειξη ***Descriptives***, αν θέλουμε διαφορετικό διάστημα εμπιστοσύνης για την μέση τιμή και την ένδειξη ***Outliers*** αν θέλουμε να εμφανιστούν οι ακραίες τιμές.



Μετά επιλέγουμε *Continue* και επιστρέφουμε στο προηγούμενο παράθυρο. Επιλέγοντας το πλαίσιο *Plots* προκύπτει το παρακάτω παράθυρο.



Στο παραπάνω παράθυρο επιλέγουμε οπωσδήποτε την ένδειξη *Normality plots with tests* για να προκύψουν οι πίνακες και τα γραφήματα που σχετίζονται με τα στατιστικά κριτήρια για τον έλεγχο της κανονικότητας.

Στο πλαίσιο *Options* το SPSS μας δίνει τη δυνατότητα να καθορίσουμε την πολιτική διαχείρισης των **ελλιπών ή απουσών τιμών (missing values**). Στην περίπτωση μας δεν έχει σημασία τί θα επιλέξουμε καθώς δεν έχουμε ελλιπείς τιμές.

Στην συνέχεια επιλέγουμε OK και προκύπτουν στο SPSS τα γραφήματα και οι πίνακες που μας χρειάζονται για τον έλεγχο της κανονικότητας:

Tests of Normality

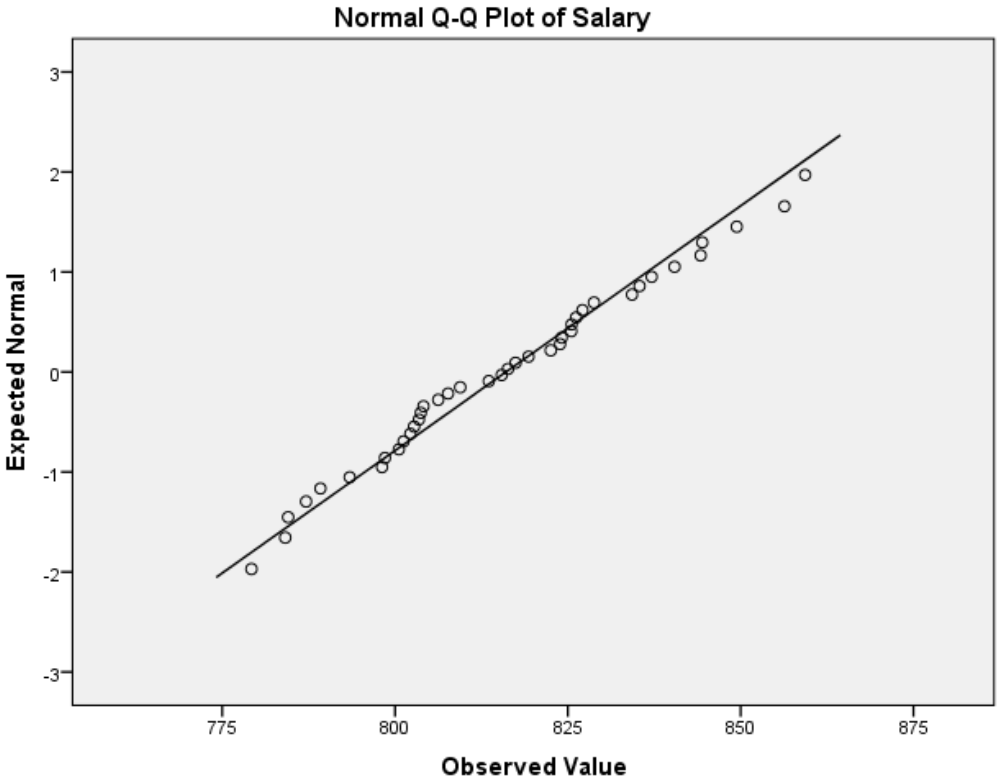
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Salary	,096	40	,200*	,978	40	,623

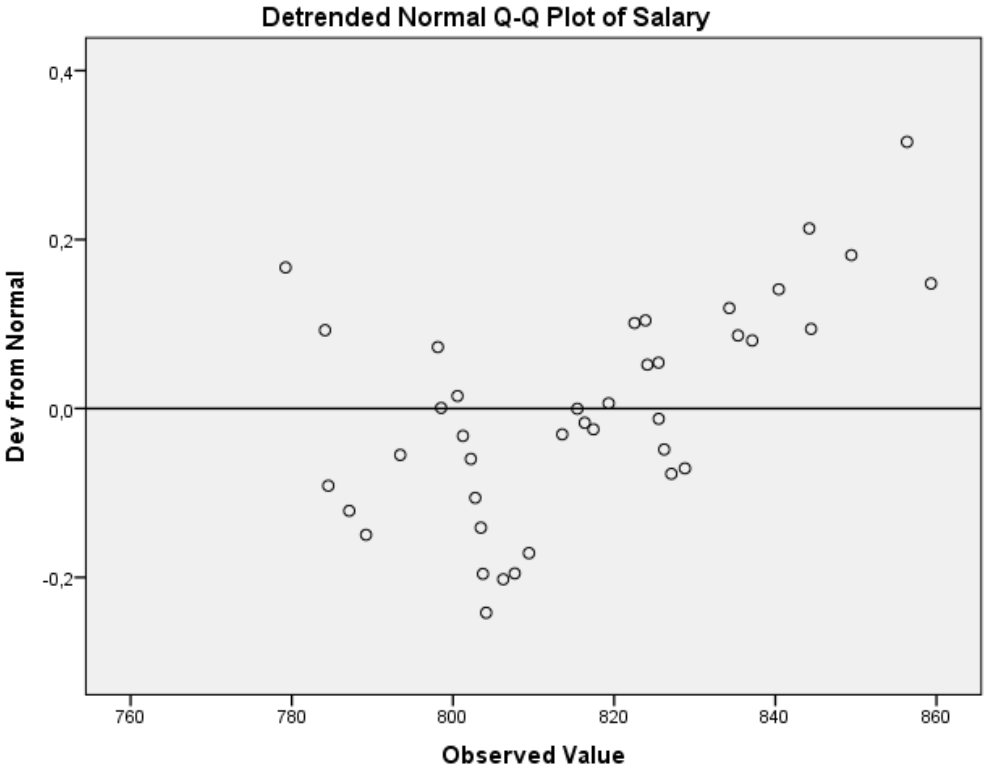
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Παρατηρούμε ότι Sig.(p-value)=0,20 δηλαδή 20% για το στατιστικό κριτήριο Kolmogorov- Smirnov και το Sig.(p-value)=0,623 δηλαδή 62,3% για το στατιστικό κριτήριο των Shapiro- Wilk. Για το στατιστικό τεστ των Shapiro-Wilk ισχύει Sig.=62,3%> a=5% (το όριο που θέσαμε για να κρίνουμε την μηδενική μας υπόθεση), και έτσι συμπεραίνουμε ότι **δεν μπορούμε να απορρίψουμε την μηδενική μας υπόθεση**. Δηλαδή, ***η κατανομή του πληθυσμού από τον οποίο προέρχεται το δείγμα μας είναι προσεγγιστικά κανονική***. Ο έλεγχος της κανονικότητας του δείγματος μπορεί να γίνει και γραφικά με τα διαγράμματα *Normal Q-Q plot* και *Detrended Normal Q-Q Plot*.

Παρακάτω απεικονίζονται τα διαγράμματα ελέγχου της Κανονικότητας:


Normal Q-Q Plot of Salary
Expected Normal
Observed Value
3
2
1
0
-1
-2
-3
775
800
825
850
875


Detrended Normal Q-Q Plot of Salary
Dev from Normal
Observed Value
0,4
0,2
0,0
-0,2
760
780
800
820
840
860

Εφόσον η υπόθεση της Κανονικότητας είναι σε ισχύ, θα εξετάσουμε τον έλεγχο υπόθεσης που αναφέραμε στην αρχή μέσω του ελέγχου ισότητας μέσων τιμών T-test.

Στο Output του SPSS εξάγονται τα ακόλουθα αποτελέσματα:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Salary	40	816,0608	20,42072	3,22880

Ο πίνακας One-Sample Statistics μας δίνει:

- Το πλήθος των παρατηρήσεων του δείγματος (**N** = 40).
- Τον αριθμητικό μέσο των παρατηρήσεων του δείγματος (**Mean** μ = 816,0608).
- Την τυπική απόκλιση των παρατηρήσεων του δείγματος (**Std. Deviation**=20,42072).
- Το τυπικό σφάλμα του αριθμητικού μέσου του δείγματος (**St. Error Mean**=3,22880).

One-Sample Test

	Test Value = 817					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Salary	-,291	39	,773	-,93925	-7,4701	5,5916

Ο πίνακας **One-Sample Test** μας δίνει :

- Την τιμή του T– test (**t** = -0,291).
- Τους βαθμούς ελευθερίας (df)=39.
- Το sig. του T – test (**Sig.** = 0,773).
- Την διαφορά της μέσης τιμής της μεταβλητής που ελέγχεται και της αριθμητικής τιμής που έχουμε ορίσει (**Mean Difference**= -0,93925).
- Το 95% διάστημα εμπιστοσύνης της διαφοράς της μέσης τιμής της μεταβλητής που ελέγχεται και της αριθμητικής τιμής που έχουμε ορίσει (95% (**Confidence Interval of the Difference**= -7,4701 , 5,5916).

Να σημειώσουμε πως για την διενέργεια αμφίπλευρων ελέγχων η διαδικασία υλοποίησης στο SPSS είναι ακριβώς η ίδια με αυτήν των μονόπλευρων ελέγχων. Η διαφορά έγκειται στην ερμηνεία των αποτελεσμάτων.

Επομένως, στην περίπτωσή μας έχουμε αριθμητικό μέσο ***Mean*** = 816,0608 και ***Sig.=0,773> 0,05****. Άρα αποδεχόμαστε την μηδενική υπόθεση* $H_0: \mu=817$.

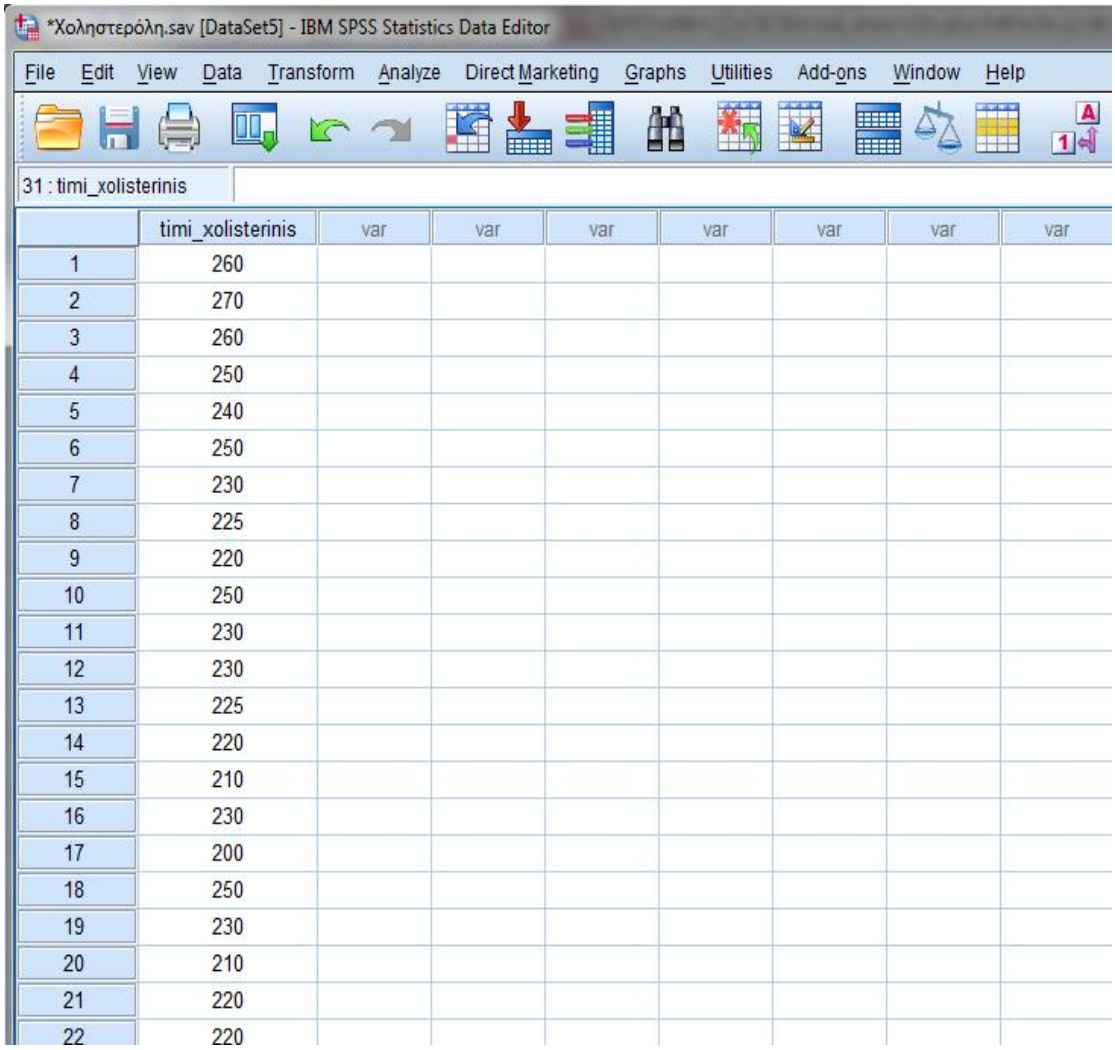
Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5% αληθεύει ο ισχυρισμός ότι ο μέσος μισθός των υπαλλήλων είναι ίσος με 817 Ευρώ.

9.3 ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΗΣ ΓΙΑ ΤΗΝ ΜΕΣΗ ΤΙΜΗ ενός ΔΕΙΓΜΑΤΟΣ (1- SAMPLE- T-TEST)- <u>ΠΑΡΑΔΕΙΓΜΑ 2</u>

Μια φαρμακευτική εταιρεία υποστηρίζει ότι η μέση χοληστερόλη αίματος σε καρδιοπαθείς μετά από χορήγηση ενός νέου φαρμάκου της για 15 ημέρες είναι 230mg/100ml. Ο Εθνικός Οργανισμός Φαρμάκων προκειμένου να διαπιστώσει εάν αληθεύει ο ισχυρισμός της εταιρείας, πήρε ένα τυχαίο δείγμα από 30 καρδιοπαθείς και μετά από χορήγηση 15 ημερών του νέου φαρμάκου μέτρησε τη χοληστερόλη του αίματος. Οι μετρήσεις του Ε.Ο.Φ. δίνονται στον παρακάτω πίνακα:

260	270	260	250	240	250	230	225	220	250
230	230	225	220	210	230	200	250	230	210
220	220	230	235	240	245	230	240	220	210

Δημιουργούμε τη μεταβλητή **timi_xolisterinis** και περνάμε τα δεδομένα (30 περιπτώσεις) του πίνακα στο SPSS ώστε να γίνουν οι απαραίτητες στατιστικές αναλύσεις. Στη συνέχεια βλέπουμε το Data View του αρχείου «Χοληστερόλη.sav»:



	timi_xolisterinis	var	var	var	var	var	var	var
1	260							
2	270							
3	260							
4	250							
5	240							
6	250							
7	230							
8	225							
9	220							
10	250							
11	230							
12	230							
13	225							
14	220							
15	210							
16	230							
17	200							
18	250							
19	230							
20	210							
21	220							
22	220							

Πριν αναφερθούμε στον έλεγχο υπόθεσης που θέλουμε να εξετάσουμε, είναι αναγκαίο να ελέγξουμε επίσης αν ισχύει η Κανονικότητα της υπό εξέταση μεταβλητής. Ακολουθώντας τα βήματα που περιγράψαμε προηγουμένως προκύπτει ο ακόλουθος πίνακας:

One-Sample Kolmogorov-Smirnov Test

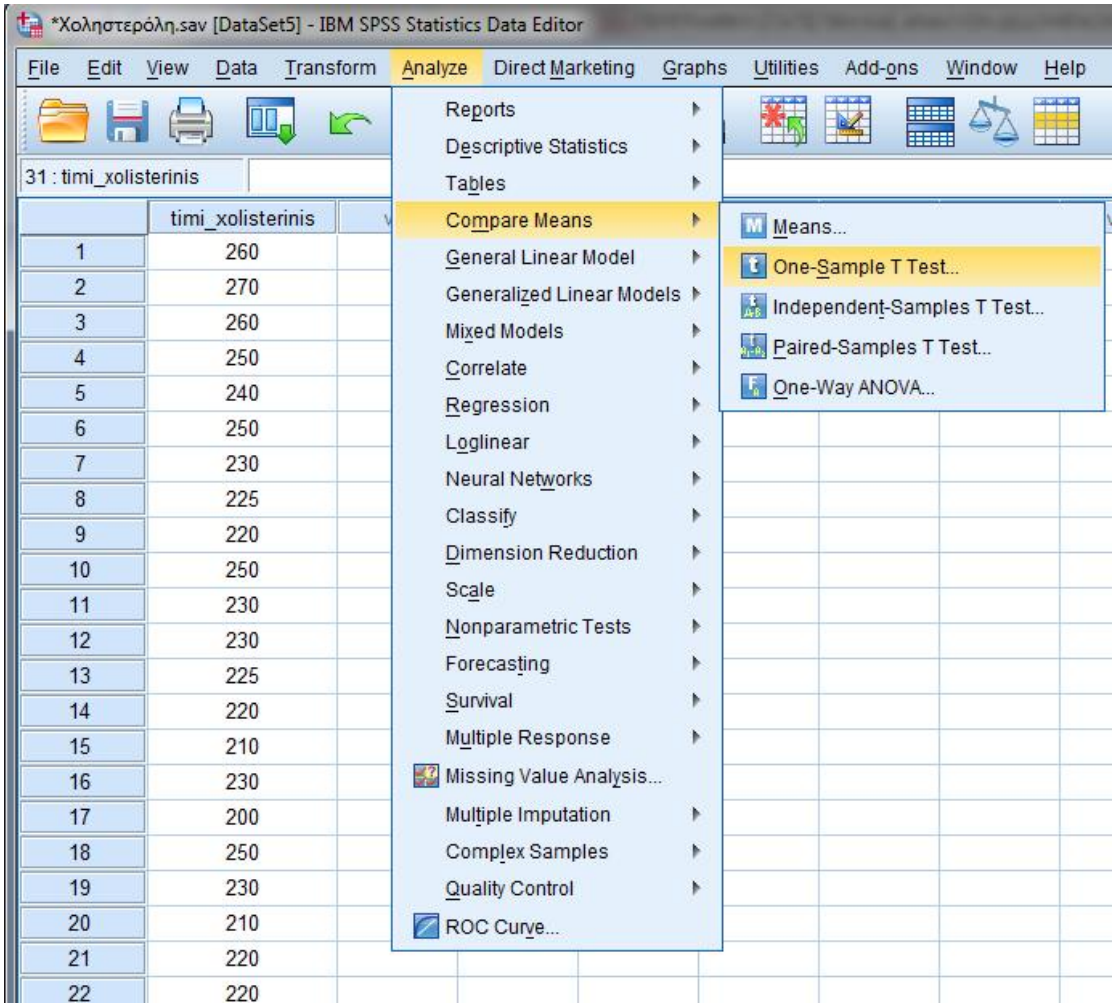
		timi_xolisterinis
N		30
Normal Parameters[a,b]	Mean	232,67
	Std. Deviation	16,595
Most Extreme Differences	Absolute	,164
	Positive	,164
	Negative	-,089
Kolmogorov-Smirnov Z		,897
Asymp. Sig. (2-tailed)		**,396**

a. Test distribution is Normal.

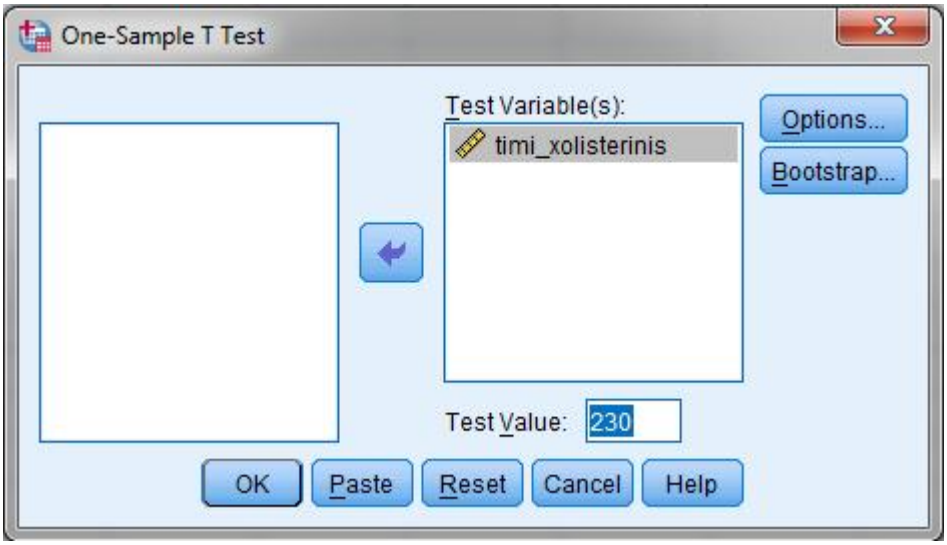
b. Calculated from data.

Πράγματι, συμπεραίνουμε πως ***η μεταβλητή timi_xolisterinis ακολουθεί την Κανονική κατανομή διότι p-value (asymp. Sig)=0,396>0,05. Συνεπώς αφού ισχύει η κανονικότητα είμαστε σε θέση να προχωρήσουμε σε έλεγχο ισότητας μέσων τιμών T-test.***

Όσον αφορά τον εξεταζόμενο έλεγχο υπόθεσης, μπορούμε να ισχυριστούμε με κίνδυνο σφάλματος a=5% (βαθμός εμπιστοσύνης 95%) ότι η φαρμακευτική εταιρεία έχει δίκιο σε αυτά που υποστηρίζει για το νέο της φάρμακο; Για να απαντήσουμε το παραπάνω ερώτημα θα πρέπει να κάνουμε έναν δίπλευρο έλεγχο της H_0: μ=230 έναντι της H_1: μ≠230. Αφού καταχωρήσουμε τα δεδομένα μας στο Data View, δίνουμε τις εντολές **Analyze → Compare Means → One-Sample T-Test,** όπως βλέπουμε στο ακόλουθο παράθυρο:



Εν συνεχεία, όπως βλέπουμε στο παρακάτω πλαίσιο διαλόγου, στη θέση Test Variable εισάγουμε τη μεταβλητή (timi_xolisterinis) της οποίας τη μέση τιμή θέλουμε να ελέγξουμε. Στη θέση Test Value ορίζουμε την τιμή μ_0 με την οποία θα γίνει η σύγκριση (εδώ η τιμή είναι ίση με 230). Από την επιλογή Options μπορούμε να ζητήσουμε και ένα διάστημα εμπιστοσύνης για τη διαφορά των μέσων τιμών.



Πατώντας OK παίρνουμε τα ακόλουθα αποτελέσματα:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
timi_xolisterinis	30	232,67	16,595	3,030

One-Sample Test

	Test Value = 230					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
timi_xolisterinis	,880	29	,386	2,667	-3,53	8,86

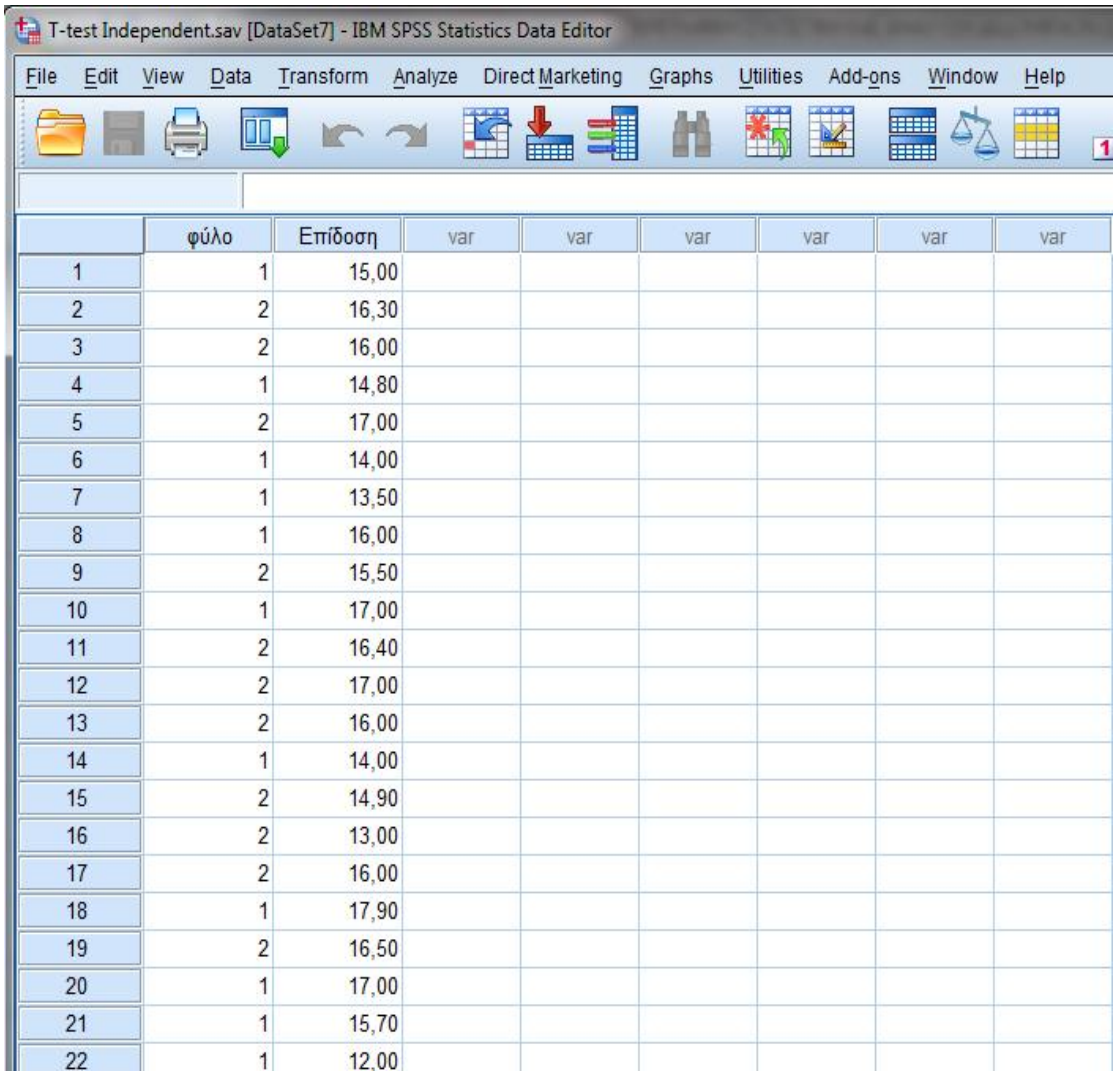
Ο πρώτος πίνακας μας πληροφορεί ότι το δείγμα μας περιλαμβάνει 30 άτομα (τιμές χοληστερόλης), με μέση τιμή χοληστερόλης 232,67 (που αποτελεί και μια εκτίμηση για τη μέση τιμή του πληθυσμού). Μας δίνει ακόμη την τυπική απόκλιση (Std. Deviation) των τιμών χοληστερόλης και το τυπικό σφάλμα της μέσης τιμής (Std. Error Mean).

Ο δεύτερος πίνακας δίνει τα *αποτελέσματα του ελέγχου υπόθεσης*. Ειδικότερα, δίνει την τιμή της ελεγχοσυνάρτησης t (t = 0,88), τους βαθμούς ελευθερίας df για την κατανομή t-Student (df = 29), και το επίπεδο σημαντικότητας (p-value – τιμή) του δίπλευρου ελέγχου (sig. 2-tailed = 0,386). Περιλαμβάνεται και ένα 95% ΔΕ για τη διαφορά των μέσων τιμών χοληστερόλης. Το σημαντικότερο σημείο αυτού του πίνακα είναι το ***sig. του t-test*** γιατί με βάση αυτό απορρίπτουμε ή αποδεχόμαστε τη μηδενική υπόθεση H_0. Αν το sig. είναι μικρότερο του 0,05 τότε απορρίπτουμε την H_0, ενώ σε αντίθετη περίπτωση την αποδεχόμαστε.

Αφού sig. = p-value = 0,386 είναι μεγαλύτερο από το επίπεδο σημαντικότητας a=0,05 που έχουμε θέσει, η υπόθεση H_0: μ = 230 δεν απορρίπτεται. **Συνεπώς, θα πρέπει να αποδεχτούμε, με κίνδυνο σφάλματος 5%, ότι ο ισχυρισμός της εταιρείας αληθεύει.**

9.4 ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΗΣ ΓΙΑ ΤΗΝ ΔΙΑΦΟΡΑ ΜΕΣΩΝ ΤΙΜΩΝ ΔΥΟ ΔΕΙΓΜΑΤΩΝ (INDEPENDENT SAMPLES T-TEST)- ΠΑΡΑΔΕΙΓΜΑ 3

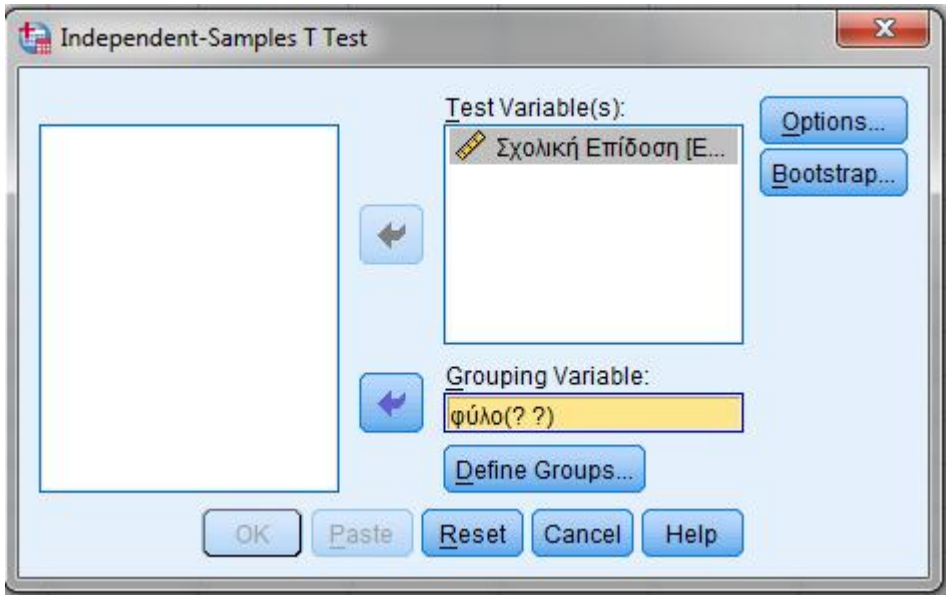
Αντλούμε και πάλι τα δεδομένα μας από το CD του βιβλίου της Νόβα-Καλτσούνη και πιο συγκεκριμένα από το αρχείο «T-test Independent.sav». Έχουμε στη διάθεση μας δύο μεταβλητές το «Φύλο» και την «Επίδοση», οι οποίες υποθέτουμε ότι ακολουθούν Κανονική κατανομή. Παρακάτω βλέπουμε το Data View:



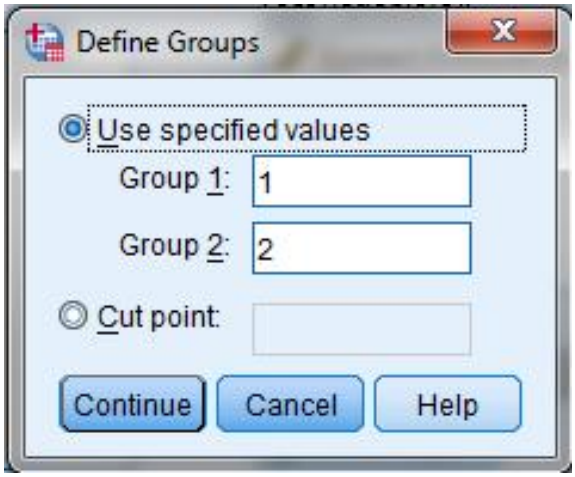
	φύλο	Επίδοση	var	var	var	var	var	var
1	1	15,00						
2	2	16,30						
3	2	16,00						
4	1	14,80						
5	2	17,00						
6	1	14,00						
7	1	13,50						
8	1	16,00						
9	2	15,50						
10	1	17,00						
11	2	16,40						
12	2	17,00						
13	2	16,00						
14	1	14,00						
15	2	14,90						
16	2	13,00						
17	2	16,00						
18	1	17,90						
19	2	16,50						
20	1	17,00						
21	1	15,70						
22	1	12,00						

Με βάση τα δεδομένα θα ελέγξουμε αν, σε επίπεδο σημαντικότητας a=5% αν οι μέσες σχολικές επιδόσεις διαφοροποιούνται σημαντικά ανάμεσα στα αγόρια και τα κορίτσια. Θα πραγματοποιηθεί σύγκριση μέσων (t-test) για τον έλεγχο υπόθεσης που μας ενδιαφέρει.

Στη συνέχεια δίνουμε στο SPSS τις εντολές **Analyze → Compare Means → Independent-Samples T-Test** και προκύπτει το παρακάτω σχήμα.



Στο *Test Variable* τοποθετούμε την μεταβλητή ως προς την οποία θα γίνει η σύγκριση (Σχολική επίδοση μαθητών) και στο *Grouping Variable* την μεταβλητή που ορίζει τις ομάδες που συγκρίνονται (φύλο μαθητών). Η επιλογή *Define Groups*, ορίζει τις δύο αυτές ομάδες. Όπως φαίνεται και στο παρακάτω σχήμα, όπως προκύπτει από το SPSS, ο κωδικός των αγοριών είναι 1 και των κοριτσιών 2. Στη συνέχεια εισάγουμε τους δύο αυτούς κωδικούς στις αντίστοιχες θέσεις, πατάμε *Continue* και *OK*.



Τα αποτελέσματα που προκύπτουν δίνονται αναλυτικά στους ακόλουθους πίνακες όπως προκύπτουν από το SPSS.

Group Statistics

	φύλο μαθητή	N	Mean	Std. Deviation	Std. Error Mean
Σχολική Επίδοση	Αγόρι	722	15,5939	1,36914	,05095
	Κορίτσι	740	15,5035	1,36848	,05031

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Σχολική Επίδοση	Equal variances assumed	,034	**,855**	1,262	1460	**,207**	,09039	,07160	-,05006	,23085
	Equal variances not assumed			1,262	1459,079	,207	,09039	,07160	-,05006	,23085

Ο πρώτος πίνακας (*Group Statistic*) δίνει τα βασικά στατιστικά στοιχεία για τις δύο ομάδες που συγκρίνονται, δηλαδή αγόρια και κορίτσια. Άρα, για το δείγμα των κοριτσιών έχουμε πληθυσμό N=740 κορίτσια, με μέση σχολική επίδοση ίση με 15,5035 και με τυπική απόκλιση 1,36848 μονάδες. Το δείγμα των αγοριών έχει πληθυσμό N=722 αγόρια, με μέση σχολική επίδοση ίση με 15,5939 και με τυπική απόκλιση 1,36914 μονάδες. Τα αποτελέσματα δείχνουν ότι η μέση σχολική επίδοση των αγοριών είναι λίγο μεγαλύτερη από των κοριτσιών. Αυτή η διαφορά παρατηρείται για τα δύο συγκεκριμένα δείγματα. Ακολουθεί έλεγχος στατιστικής σημαντικότητας αυτής της διαφοράς ο οποίος δίνεται στον δεύτερο πίνακα που αναγράφει τα αποτελέσματα του στατιστικού ελέγχου T-test για ανεξάρτητα δείγματα.

Υπάρχουν δύο είδη t - test ανεξάρτητων δειγμάτων και φαίνονται στον πίνακα με την παρουσία δύο γραμμών. Η πρώτη γραμμή αποτελεσμάτων ονομάζεται ***Equal variances assumed*** και η δεύτερη ***Equal variances not assumed***. Πρέπει να

ακολουθήσουμε δύο βήματα προκειμένου να ερμηνεύσουμε τα αποτελέσματα του δεύτερου πίνακα. Σύγκριση των μέσων τιμών των δύο ομάδων (έλεγχος της H_0: $\mu_1 = \mu_2$ έναντι της H_1: $\mu_1 \neq \mu_2$).

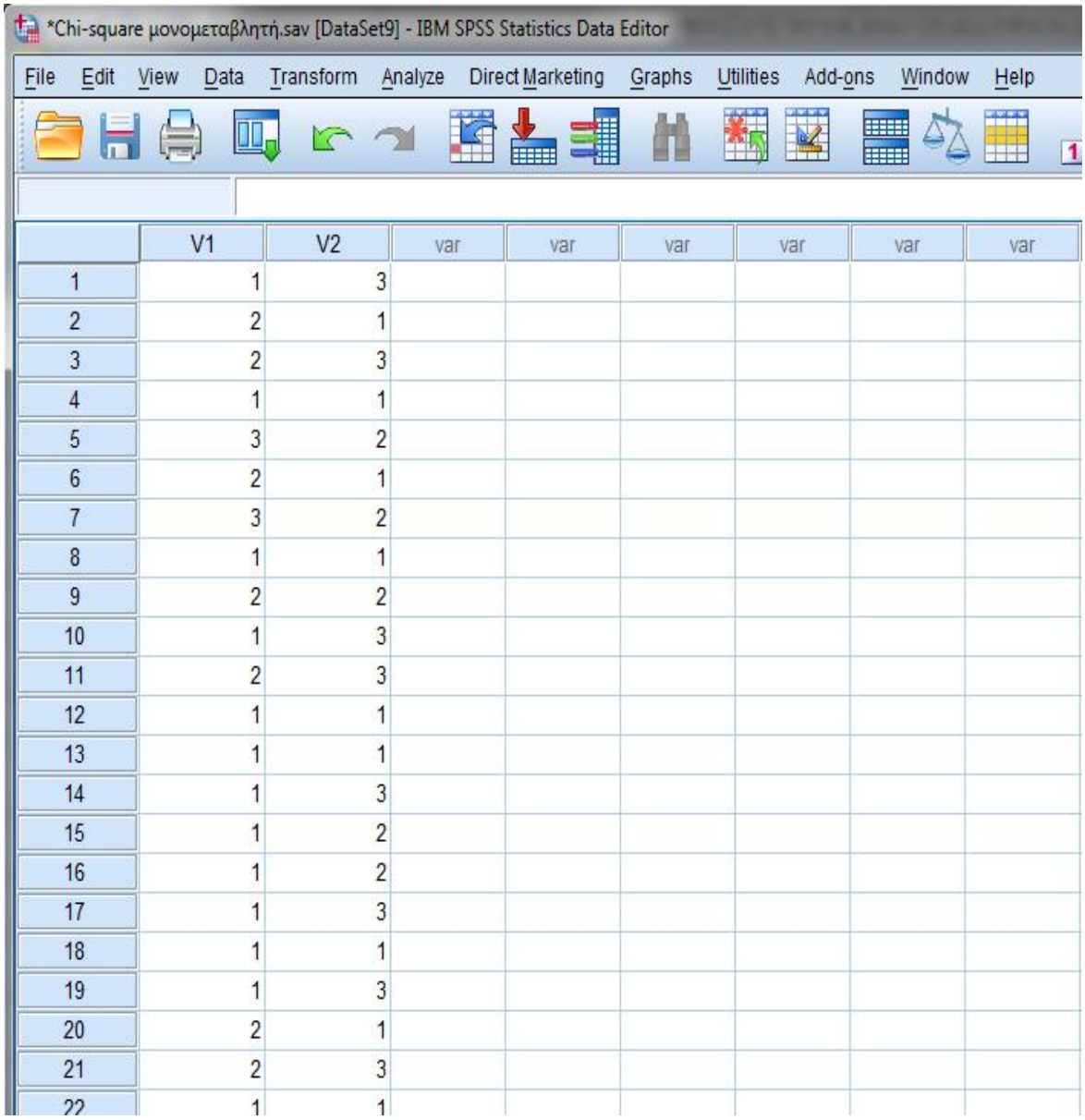
Ανάλογα με τα αποτελέσματα του ***ελέγχου Levene***, επιλέγουμε ποιο από τα δύο είδη του t-test είναι το κατάλληλο, δηλαδή ποια από τις δύο γραμμές είναι η κατάλληλη. Εάν το sig. (2-tailed) είναι μικρότερο του 0,05 τότε η υπόθεση H_0 της ισότητας των δύο μέσων απορρίπτεται (στατιστικά σημαντική διαφορά). Στην αντίθετη περίπτωση αποδεχόμαστε ότι οι δυο πληθυσμοί δεν διαφέρουν σημαντικά ως προς τη μέση τιμή τους.

Στο συγκεκριμένο παράδειγμα ο έλεγχος Levene δίνει sig. = 0,855 μεγαλύτερο του 0,05 και αποδεχόμαστε την υπόθεση των ίσων διακυμάνσεων. Για το t-test έλεγχο των μέσων τιμών τα αποτελέσματα της πρώτης γραμμής είναι κατάλληλα. Επειδή sig. (2-tailed) = 0,207 είναι μεγαλύτερο του 0,05 αποδεχόμαστε την υπόθεση των ίσων μέσων τιμών. ***Συνεπώς ο βαθμός στην σχολική επίδοση των κοριτσιών (κατά μέσο όρο) δεν παρουσιάζει στατιστικά σημαντική διαφορά (στο επίπεδο 0,05) από την σχολική επίδοση των αγοριών.***

9.5 ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΗΣ ΓΙΑ ΤΗΝ ΑΝΕΞΑΡΤΗΣΙΑ ΔΥΟ ΜΕΤΑΒΛΗΤΩΝ (PEARSON'S X^2 CHI-SQUARE)- **ΠΑΡΑΔΕΙΓΜΑ 4**

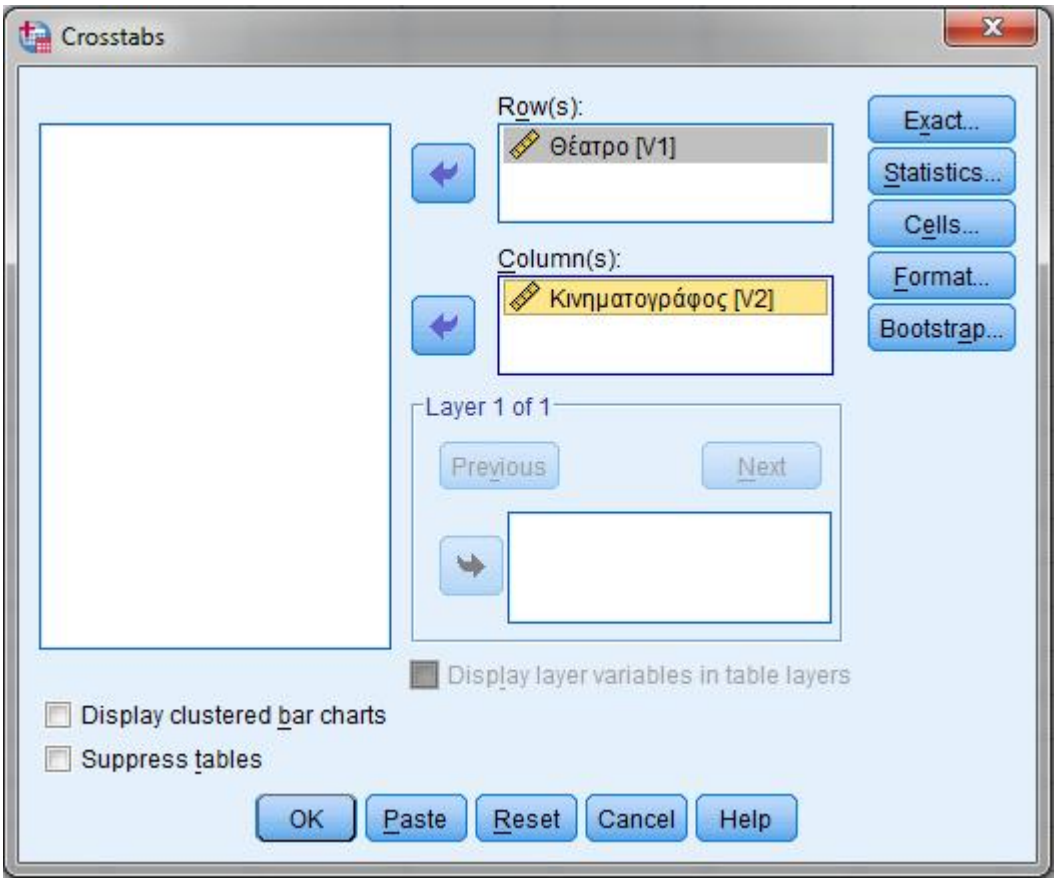
Ο έλεγχος X^2 του Pearson εφαρμόζεται όταν θέλουμε να διαπιστώσουμε αν δύο ποιοτικές μεταβλητές ενός πληθυσμού σχετίζονται μεταξύ τους ή είναι ανεξάρτητες. Πιο συγκεκριμένα, ελέγχουμε τη «μηδενική» υπόθεση H_0: οι *δυο ποιοτικές μεταβλητές είναι ανεξάρτητες μεταξύ τους*, έναντι της εναλλακτικής H_1: *οι δυο ποιοτικές μεταβλητές είναι εξαρτημένες*.

Θα χρησιμοποιήσουμε δεδομένα για άλλη μια φορά από το CD του βιβλίου της Νόβα- Καλτσούνη και πιο συγκεκριμένα από το αρχείο «Chi-square μονομεταβλητή.sav». Έχουμε στη διάθεση μας δύο μεταβλητές το «Θέατρο-V1» και τον «Κινηματογράφο-V2». Παρακάτω βλέπουμε το Data View:

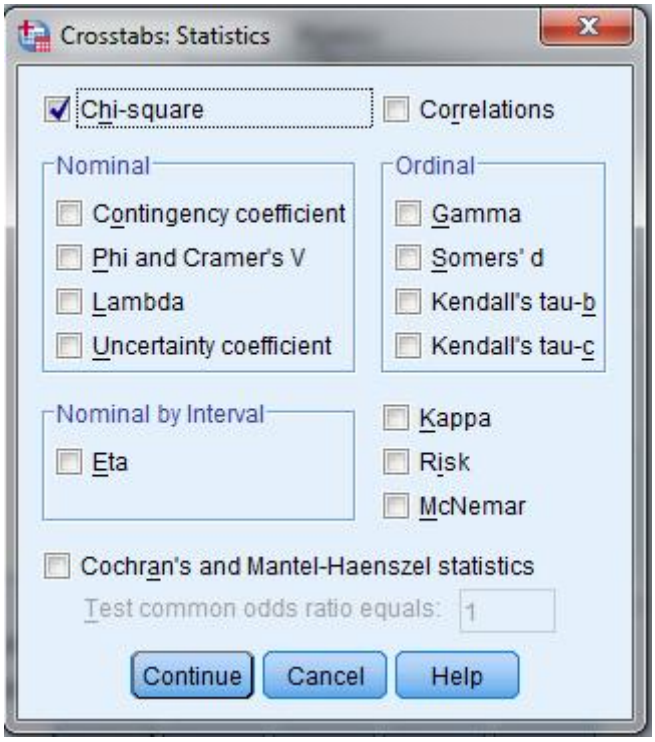


	V1	V2	var	var	var	var	var	var
1	1	3						
2	2	1						
3	2	3						
4	1	1						
5	3	2						
6	2	1						
7	3	2						
8	1	1						
9	2	2						
10	1	3						
11	2	3						
12	1	1						
13	1	1						
14	1	3						
15	1	2						
16	1	2						
17	1	3						
18	1	1						
19	1	3						
20	2	1						
21	2	3						
22	1	1						

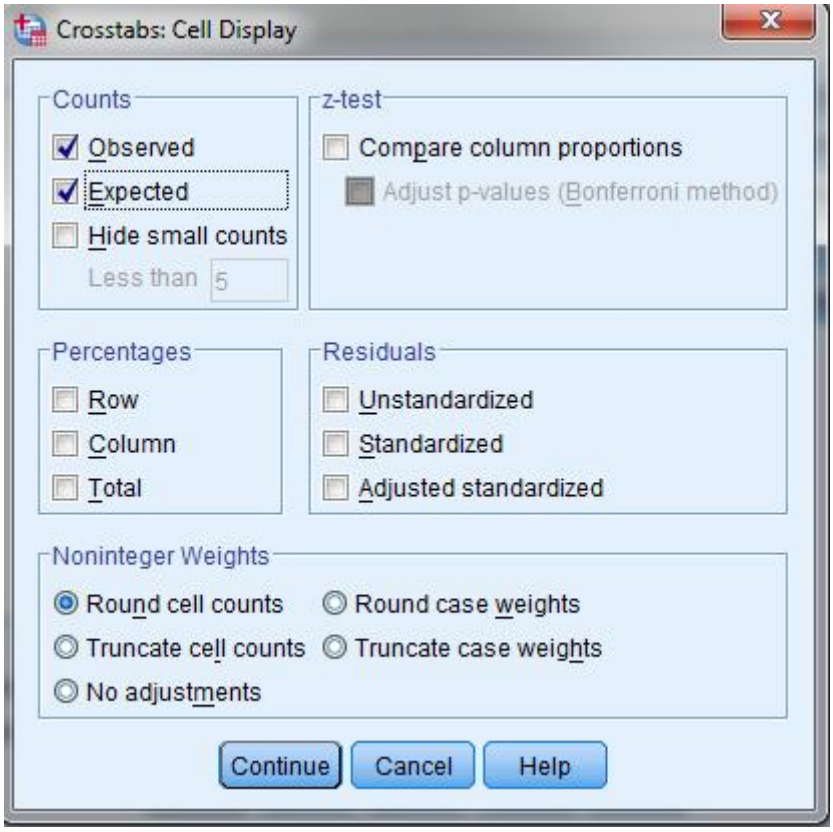
Θα ελέγξουμε αν, σε επίπεδο σημαντικότητας a=5%, η συχνότητα επισκεψιμότητας του θεάτρου συσχετίζεται με την συχνότητα επισκεψιμότητας του κινηματογράφου. Επιλέγουμε τις παρακάτω εντολές: **Analyze → Descriptive Statistics → Crosstabs.** Στις θέσεις Row και Column βάζουμε τις δύο μεταβλητές που θα ελεγχθούν, δηλαδή το Θέατρο-V1 και τον Κινηματογράφο-V2 αντίστοιχα και εμφανίζεται το παράθυρο που ακολουθεί:



Ακολούθως, επιλέγοντας Statistics ανοίγει το εξής πλαίσιο διαλόγου όπου επιλέγουμε το πλαίσιο Chi-square:



Πατάμε το Continue. Επιλέγονας το Cells ανοίγει ένα νέο παράθυρο διαλόγου στο οποίο επιλέγουμε να εμφανίζονται οι παρατηρούμενες συχνότητες (observed counts) και οι αναμενόμενες συχνότητες (expected counts). Πατάμε Continue και *OK*.



Επιλέγοντας όλα τα παραπάνω στο SPSS προκύπτουν τα ακόλουθα αποτελέσματα:

Θέατρο * Κινηματογράφος Crosstabulation

			Κινηματογράφος			Total
			πολύ συχνά	συχνά	σπάνια	
Θέατρο	πολύ συχνά	Count	97	38	78	213
		Expected Count	84,2	57,4	71,4	213,0
	συχνά	Count	41	18	39	98
		Expected Count	38,8	26,4	32,9	98,0
	σπάνια	Count	0	38	0	38
		Expected Count	15,0	10,2	12,7	38,0
Total		Count	138	94	117	349
		Expected Count	138,0	94,0	117,0	349,0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	116,122[a]	4	**,000**
Likelihood Ratio	113,787	4	,000
Linear-by-Linear Association	,603	1	,438
N of Valid Cases	349		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 10,23.

Ο πρώτος πίνακας μας δίνει τη συνάφεια ανάμεσα στο Θέατρο- V1 και τον Κινηματογράφο- V2. Ο πίνακας αυτός δείχνει τις συχνότητες (counts) των διασταυρούμενων κατηγοριών του δείγματος, καθώς και τις συχνότητες (expected counts) που αναμένονταν σε περίπτωση ανεξαρτησίας των δύο μεταβλητών. Για παράδειγμα, έστω ότι στο δείγμα μας παρατηρήθηκαν 97 άτομα που επισκέπτονται «Πολύ Συχνά» τόσο το θέατρο όσο και τον κινηματογράφο. Αν η συχνότητα επισκεψιμότητας του θεάτρου ήταν ανεξάρτητη της συχνότητα επισκεψιμότητας του κινηματογράφου θα αναμέναμε 84,2 άτομα που επισκέπτονται Πολύ Συχνά τόσο το θέατρο όσο και τον κινηματογράφο (κατά μέσο όρο). Ο βαθμός διαφοροποίησης των παρατηρούμενων συχνοτήτων από τις συχνότητες που θα αναμένονταν υπό την υπόθεση της ανεξαρτησίας καθορίζει και το αποτέλεσμα του ελέγχου της ανεξαρτησίας.

Ο δεύτερος πίνακας παρουσιάζει τα αποτελέσματα τριών ελέγχων ανεξαρτησίας. Μας ενδιαφέρει η πρώτη γραμμή που δείχνει τα αποτελέσματα του ελέγχου X^2 του Pearson. Η τιμή της ελεγχοσυνάρτησης είναι X^2 = 116,122 με 4 βαθμούς ελευθερίας και το επίπεδο σημαντικότητας (p – value) του ελέγχου είναι 0,000 (πρακτικά 0%). **Επειδή έχουμε επίπεδο σημαντικότητας μικρότερο από 0,05 η υπόθεση της ανεξαρτησίας ανάμεσα στις δύο μεταβλητές απορρίπτεται**.

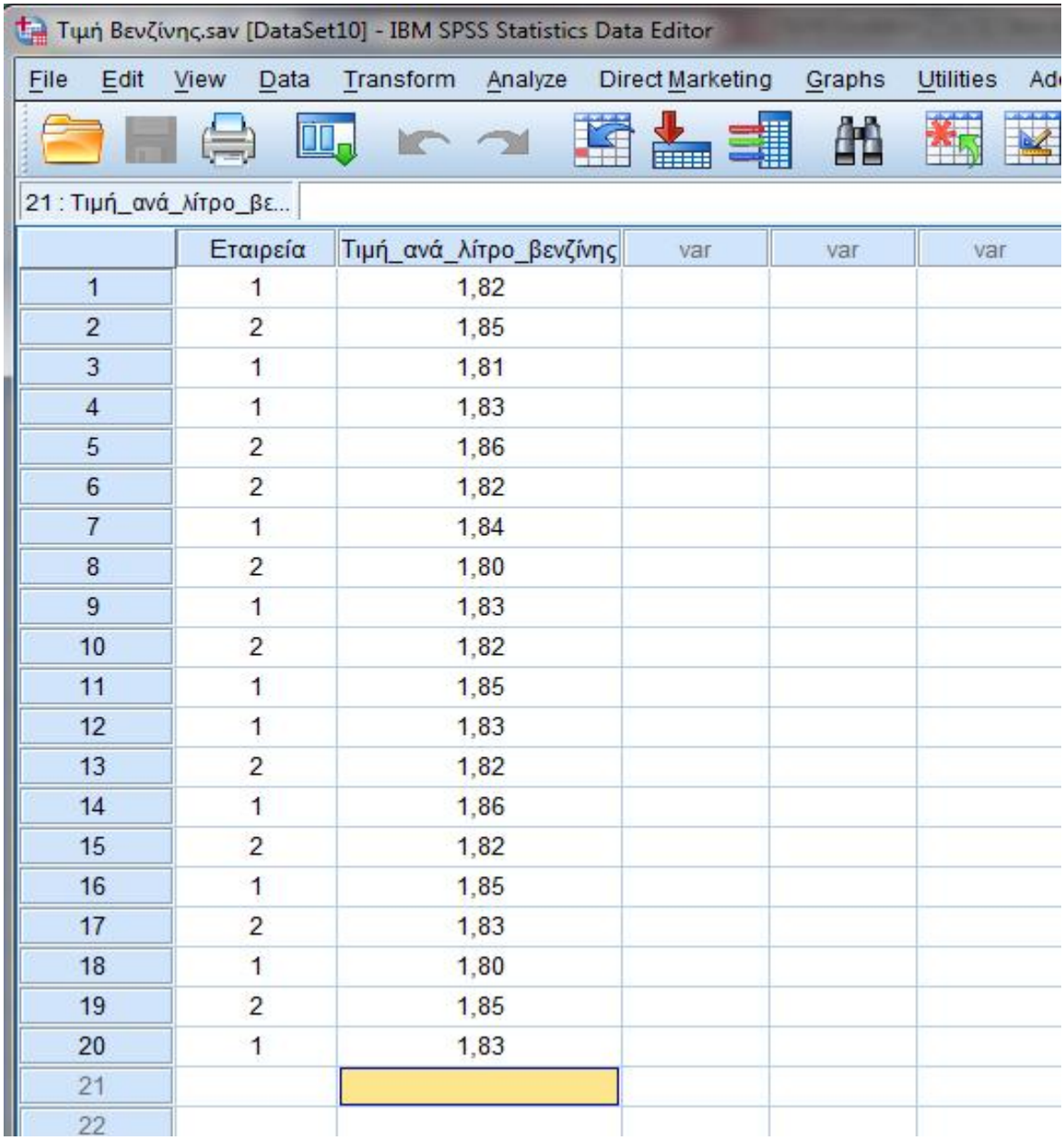
Συμπερασματικά, παρουσιάζεται σημαντική διαφοροποίηση ανάμεσα στις παρατηρούμενες και τις αναμενόμενες συχνότητες της συχνότητας επισκεψιμότητας του θεάτρου και του κινηματογράφου και οι δύο μεταβλητές είναι εξαρτημένες.

9.6 ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΗΣ ΓΙΑ ΤΗΝ ΔΙΑΦΟΡΑ ΜΕΣΩΝ ΤΙΜΩΝ ΔΥΟ ΔΕΙΓΜΑΤΩΝ (INDEPENDENT SAMPLES T-TEST)- <u>ΠΑΡΑΔΕΙΓΜΑ 5</u>

Στον πίνακα που ακολουθεί δίνονται τιμές της βενζίνης, σε χρηματικές μονάδες, είκοσι πρατηρίων υγρών καυσίμων δύο διαφορετικών εταιρειών (τυχαία επιλογή). Θέλουμε να ελέγξουμε, σε επίπεδο σημαντικότητας 5%, αν η μέση τιμή μ_1 του πληθυσμού των πρατηρίων υγρών καυσίμων της πρώτης εταιρείας είναι ίση με τη μέση τιμή μ_2 του πληθυσμού των πρατηρίων υγρών καυσίμων της δεύτερης εταιρείας.

Εταιρεία	Τιμή/ανά λίτρο βενζίνης
1	1,82
2	1,85
1	1,81
1	1,83
2	1,86
2	1,82
1	1,84
2	1,80
1	1,83
2	1,82
1	1,85
1	1,83
2	1,82
1	1,86
2	1,82
1	1,85
2	1,83
1	1,80
2	1,85
1	1,83

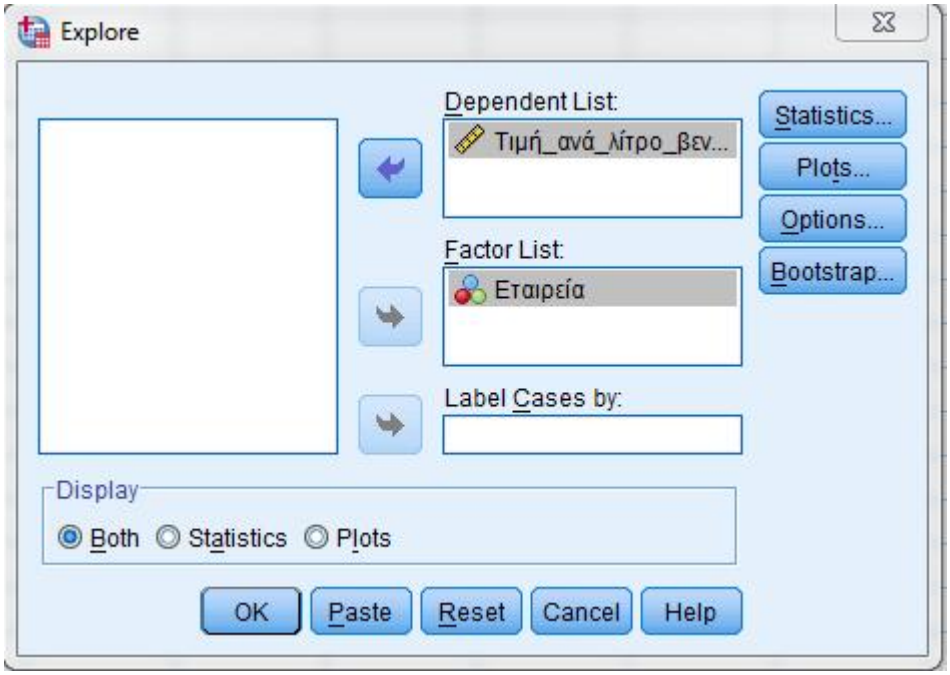
Δημιουργούμε το αρχείο «Τιμή Βενζίνης.sav» και παρακάτω βλέπουμε το Data View:



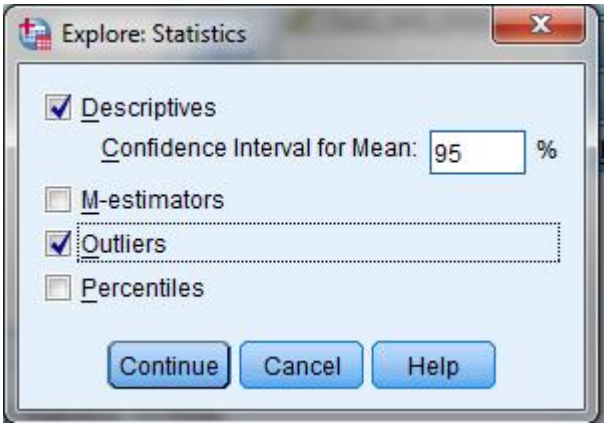
	Εταιρεία	Τιμή_ανά_λίτρο_βενζίνης	var	var	var
1	1	1,82			
2	2	1,85			
3	1	1,81			
4	1	1,83			
5	2	1,86			
6	2	1,82			
7	1	1,84			
8	2	1,80			
9	1	1,83			
10	2	1,82			
11	1	1,85			
12	1	1,83			
13	2	1,82			
14	1	1,86			
15	2	1,82			
16	1	1,85			
17	2	1,83			
18	1	1,80			
19	2	1,85			
20	1	1,83			
21					
22					

9.6.1 ΥΠΟΘΕΣΗ ΠΡΩΤΗ: ΚΑΝΟΝΙΚΗ ΚΑΤΑΝΟΜΗ ΤΩΝ ΜΕΤΑΒΛΗΤΩΝ

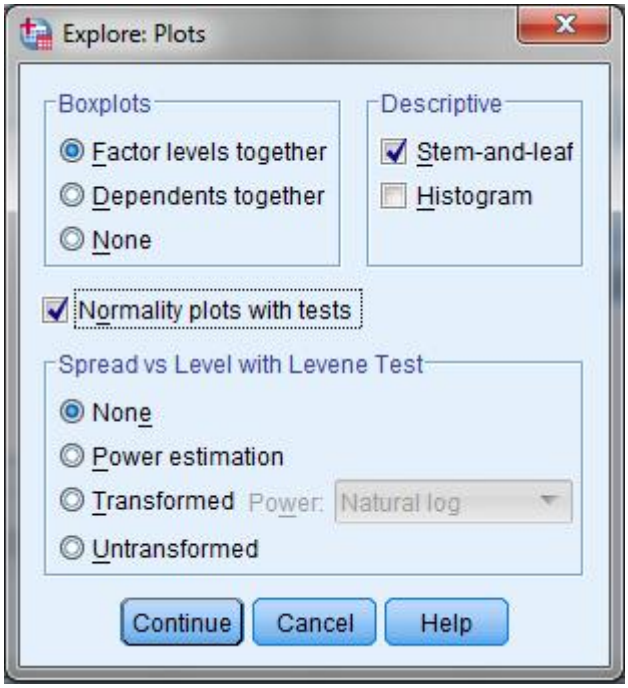
Αφού έχουμε δημιουργήσει τις δυο μεταβλητές μας ξεκινάμε τις αναλύσεις στο SPSS. Αρχικά θα ελέγξουμε αν η κατανομή των μεταβλητών μας είναι κανονική. Για να ελέγξουμε την κανονικότητα των δύο δειγμάτων δίνουμε τις εντολές: Analyze → Descriptive Statistics → Explore. Στο παράθυρο διαλόγου που προκύπτει τοποθετούμε την μεταβλητή Τιμή ανά λίτρο βενζίνης στη θέση Dependent List και την μεταβλητή Εταιρεία στη θέση Factor List.



Στο Statistics επιλέγουμε τις ενέργειες που θέλουμε να γίνουν, όπως φαίνονται στο παρακάτω σχήμα.



Στη συνέχεια στο Plots επιλέγουμε τη ρύθμιση Normality plots with tests.



Μετά επιλέγουμε Continue, OK και προκύπτει ο παρακάτω πίνακας:

Tests of Normality

	Εταιρεία	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Τιμή_ανά_λίτρο_βενζίνης	1	,187	11	,200*	,959	11	**,756**
	2	,253	9	,101	,899	9	**,249**

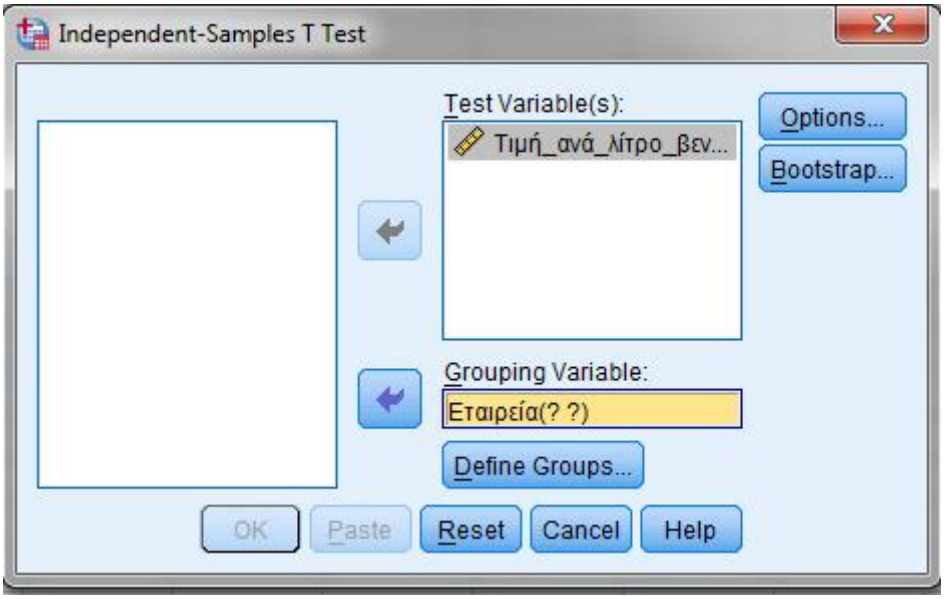
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

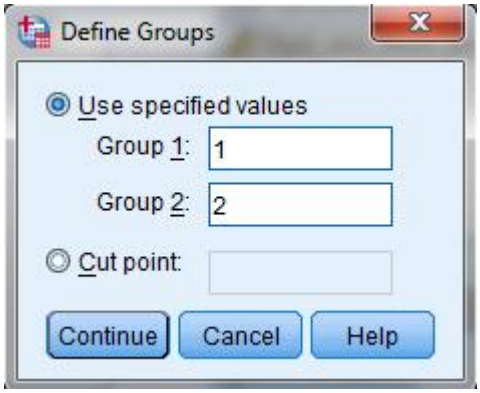
Επειδή τα δείγματα μας είναι μεγέθους < 50 μπορούμε να ασχοληθούμε μόνο με το στατιστικό τεστ των Shapiro-Wilk. Παρατηρούμε ότι για τα πρατήρια υγρών καυσίμων της πρώτης εταιρείας το p−value$_1$ = sig. = 0,756 δηλαδή 75,6% > 5% και για τα πρατήρια υγρών καυσίμων της δεύτερης εταιρείας το p – value$_2$ = sig.= 0,249, δηλαδή 24,9%>5%. **Άρα και οι δυο εταιρείες παρουσιάζουν κανονικές κατανομές.**

9.6.2 Η ΕΚΤΕΛΕΣΗ ΤΟΥ T-TEST ΓΙΑ ΔΥΟ ΑΝΕΞΑΡΤΗΤΑ ΔΕΙΓΜΑΤΑ.

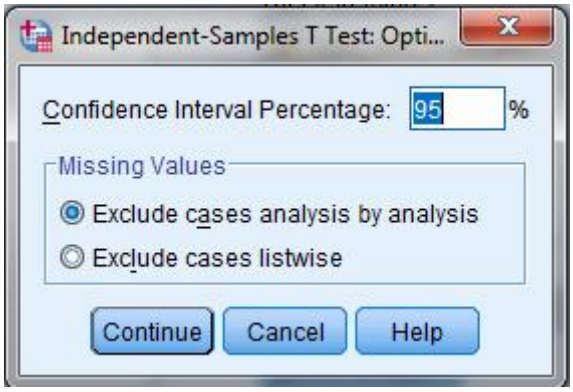
Οι προϋποθέσεις χρήσης του t-test ικανοποιούνται μιας και το δείγμα μας είναι τυχαίο και οι κατανομές των μεταβλητών κανονική. Για την έναρξη του ελέγχου t-test ακολουθούνται οι εξής εντολές: **Analyze → Compare Means → Independent-Samples T-test**. Στο παράθυρο διαλόγου που προκύπτει τοποθετούμε την μεταβλητή «Τιμή ανά λίτρο βενζίνης» στο Test Variables και την μεταβλητή «Εταιρεία» στο Grouping Variable για να προσδιορίσουμε σε ποια μεταβλητή θα κάνουμε έλεγχο μέσων τιμών και ως προς ποια μεταβλητή θα κάνουμε διαχωρισμό περιπτώσεων.



Στη συνεχεία επιλεγούμε Define Groups και θέτουμε τις τιμές για να διαχωρίσουμε δυο ομάδες της μεταβλητής price.



Μετά επιλέγουμε Continue και στο Options καθορίζουμε το επίπεδο στατιστικής σημαντικότητας και διαχειριζόμαστε τις απούσες τιμές (missing values).



Στη συνέχεια επιλέγουμε Continue και OK και εμφανίζονται οι παρακάτω πίνακες.

Group Statistics

	Εταιρεία	N	Mean	Std. Deviation	Std. Error Mean
Τιμή_ανά_λίτρο_βενζίνης	1	11	1,8318	,01779	,00536
	2	9	1,8300	,01936	,00645

Στον παραπάνω πίνακα δίνονται στατιστικά στοιχεία των δυο δειγμάτων. Το μέγεθος του πρώτου δείγματος είναι 11 και του δεύτερου 9. Επίσης, προκύπτει ότι η μέση τιμή της βενζίνης για τα πρατήρια της πρώτης εταιρείας είναι 1,8318 και για τα πρατήρια της δεύτερης εταιρείας είναι 1,83.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Τιμή_ανά_λίτρο_βενζίνης	Equal variances assumed	,235	**,634**	,219	18	**,829**	,00182	,00832	-,01566	,01929
	Equal variances not assumed			,217	16,548	,831	,00182	,00839	-,01592	,01956

Στον παραπάνω πίνακα βλέπουμε δύο t-test και θα επιλέξουμε το κατάλληλο αφού εξετάσουμε το τεστ του Levene, που βρίσκεται στο αριστερό μέρος του πίνακα και μας πληροφορεί για την ισότητα των πληθυσμιακών διασπορών. ***Παρατηρούμε ότι για το τεστ Levene ισχύει 63,4% > 5% (ή p-value= 0.634 > 0.05) και επομένως δεχόμαστε ότι οι διακυμάνσεις είναι ίσες. Επομένως, θα κοιτάξουμε στην πρώτη γραμμή του πίνακα (Equal variances assumed) όπου έχουμε p − value = 0,829 > 0,05 και επομένως δεν μπορούμε να απορρίψουμε την μηδενική υπόθεση ότι οι μέσες τιμές στα πρατήρια υγρών καύσιμων των δύο εταιρειών είναι ίσες.*** Αν στο τεστ Levene βρίσκαμε p − value < 0,05 (όταν έχουμε ε.σ. a=5%) τότε θα κοιτάζαμε τη δεύτερη γραμμή του παραπάνω πίνακα.

ΒΙΒΛΙΟΓΡΑΦΙΑ

ΕΛΛΗΝΙΚΕΣ ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ (REFERENCES)

Αγιακλόγλου Χρήστος, Μπένος Θεοφάνης, (2007). «Εισαγωγή στην οικονομετρική ανάλυση». Εκδόσεις Μπένου.

Ατζουλάκος Δ., (2008). «Στατιστικός Έλεγχος Ποιότητας – Β' Έκδοση». Σημειώσεις παραδόσεων. Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς.

Βάμβουκας Γεώργιος, (2007). «Σύγχρονη Οικονομετρία». Εκδόσεις Οικονομικού Πανεπιστημίου Αθηνών.

Βενέτης Ιωάννης, (2009). «Εισαγωγικές διαλέξεις στην Οικονομετρία». Εκδόσεις Γκιούρδας, Αθήνα.

Κιντής Α., (2010). «Σύγχρονη οικονομετρική ανάλυση - Τόμος Α'». Εκδόσεις Gutenberg, Αθήνα.

Κιόχος Π. (1993). «Περιγραφική Στατιστική». Εκδόσεις Interbooks, Αθήνα.

Κολύβα-Μαχαίρα Φ., Μπόρα-Σέντα Ε., (1998). «Στατιστική θεωρία και εφαρμογές». Εκδόσεις Ζήτη, Θεσσαλονίκη.

Νόβα- Καλτσούνη Χ., (2006). «Μεθοδολογία Εμπειρικής Έρευνας στις Κοινωνικές Επιστήμες- Ανάλυση Δεδομένων με τη χρήση του SPSS 13», Αθήνα.

Ξεκαλάκη Ευδ. (1995). «Τεχνικές Δειγματοληψίας». Αθήνα.

Ξεκαλάκη Ευδ. (2001). «Μη-Παραμετρική Στατιστική», Αθήνα.

Πανάρετος Ιωαν. & Ξεκαλάκη Ευδ. (1993). «Εισαγωγή στη Στατιστική Σκέψη- Τόμος 1 (Περιγραφική Στατιστική)». Αθήνα.

Πανάρετος Ιωαν. (1997). «Γραμμικά Μοντέλα με έμφαση στις εφαρμογές». Γ' έκδοση, Αθήνα.

Ρούσσας Γ. (1994). «Στατιστική συμπερασματολογία Τ. ΙΙ, Έλεγχος υποθέσεων». Εκδόσεις Ζήτη, Αθήνα.

Τζαβαλής Ηλίας, (2008). «Οικονομετρία». Εκδόσεις Οικονομικού Πανεπιστημίου Αθηνών.

Τριχόπουλος Δ., Τζώνου Α., Κατσουγιάννη κ., (2001). «Βιοστατιστική». Εκδόσεις Παρισιάνου, Αθήνα.

Τσιώνας Βασίλειος, (2009). «Στατιστική με εφαρμογές στην Οικονομετρία». Εκδόσεις Οικονομικού Πανεπιστημίου Αθηνών.

Χατζηνικολάου Δ. (2002). «Στατιστική για οικονομολόγους. Β' Έκδοση». Εκδόσεις Printshop, Θεσσαλονίκη.

Ψώνιος Δ., (1999). «Στατιστική». Εκδόσεις Ζήτη, Θεσσαλονίκη.

ΞΕΝΕΣ ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ (REFERENCES)

Begg David, Fischer Stanley, Dornbusch Rudiger (2006). «Εισαγωγή στην οικονομική». Εκδόσεις Κριτική, Αθήνα.

Howitt, D. Cramer «Στατιστική με το SPSS 13», Γ' έκδοση.

ΗΛΕΚΤΡΟΝΙΚΕΣ ΠΗΓΕΣ (ELECTRONIC SOURCES)

http://www.aua.gr/gpapadopoulos/files/hypoth-tests-4.pdf

http://users.auth.gr/dkugiu/Teach/CivilEngineer/hypothesis.pdf

http://www.mednet.gr/archives/2010-4/pdf/691.pdf

http://www.aua.gr/gpapadopoulos/files/anova12-13a.pdf

http://www.aua.gr/gpapadopoulos/files/descrstat12a.pdf

http://statistics.scientist.gr/23.pdf

http://ebooks.edu.gr/modules/ebook/show.php/DSGL-C100/493/3203,13012/)

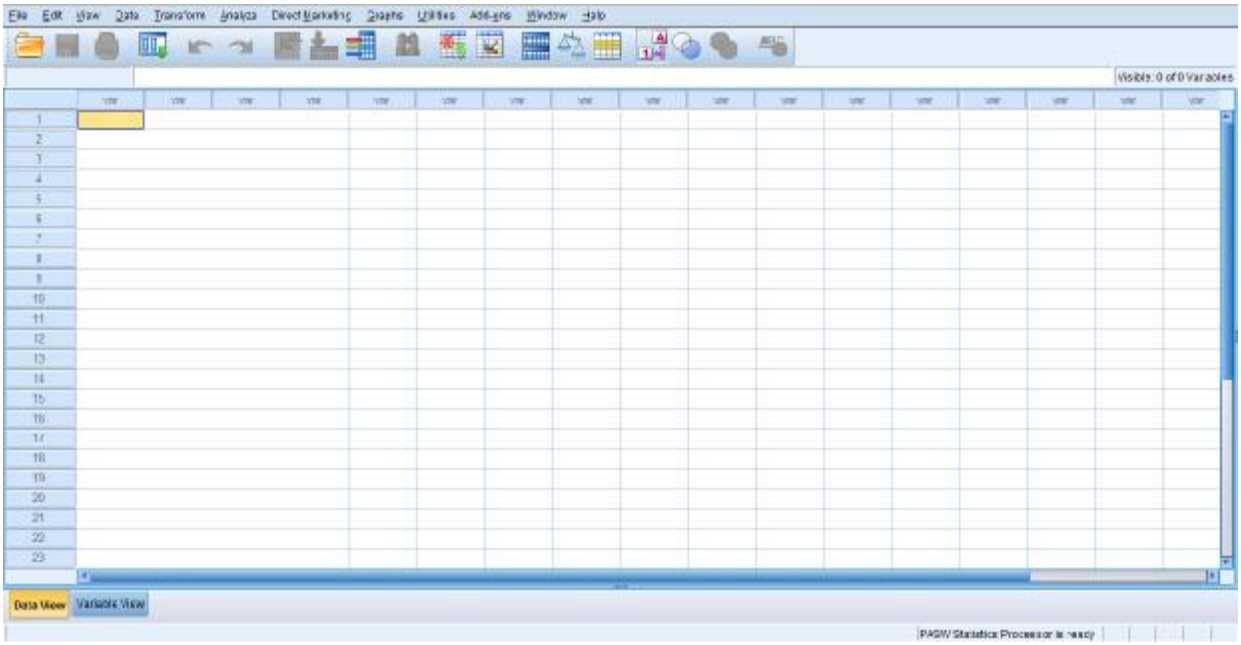
http://www.toposbooks.gr/behavioralstats/samplechapter.pdf

http://www.arnos.gr/system/files/1_6.pdf

http://www.hjn.gr/actions/get_pdf.php?id=262

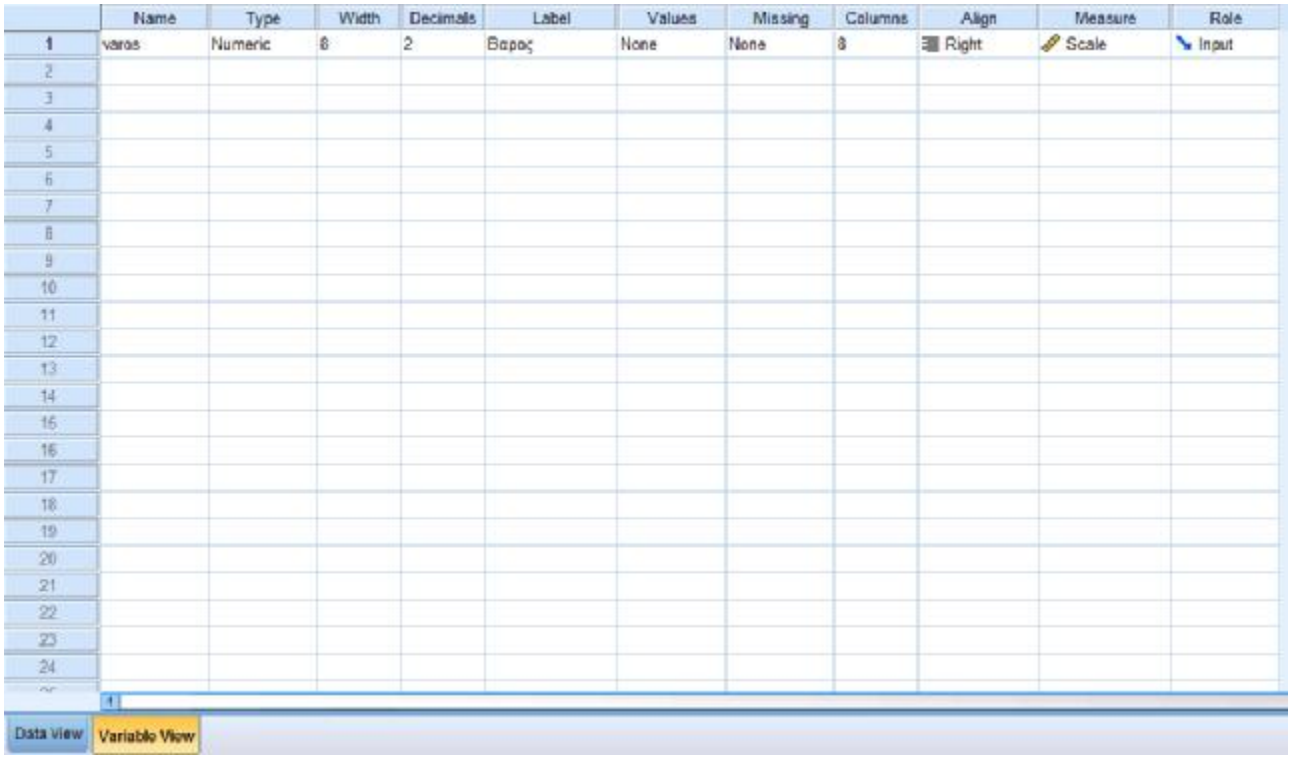
ΠΑΡΑΡΤΗΜΑ

1. Η επιφάνεια όπου ο χρήστης καλείται να εργαστεί είναι η ακόλουθη:

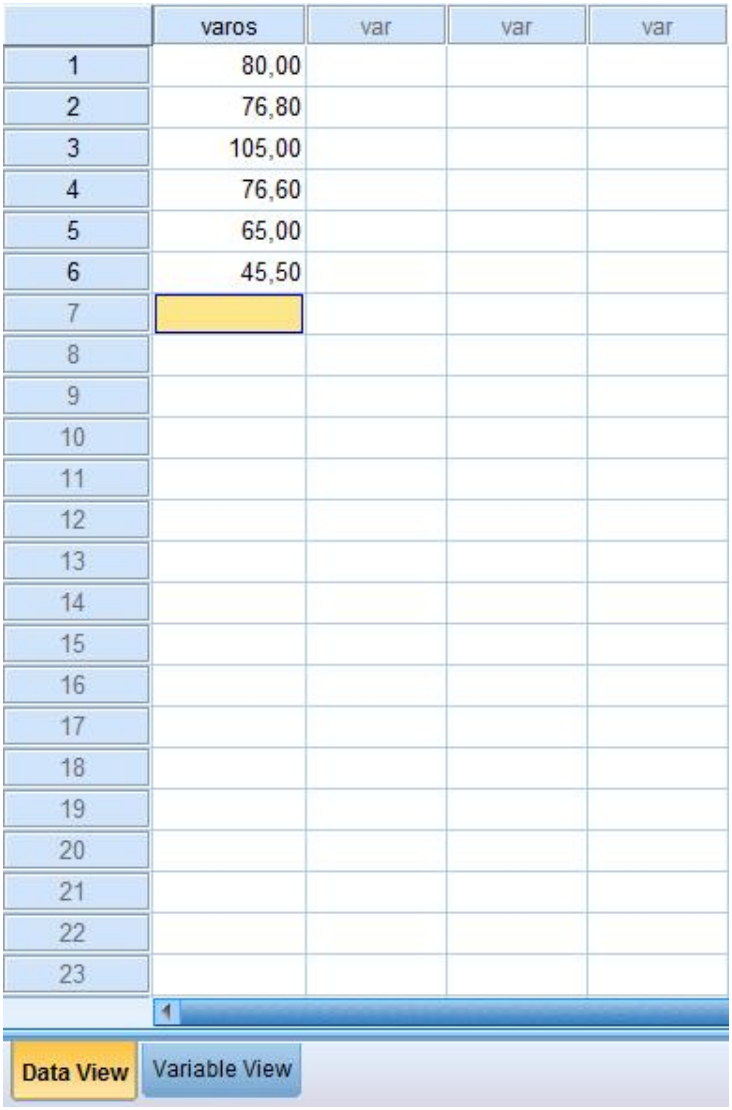


Αυτά είναι τα Φύλλα Εργασίας του SPSS που εμπεριέχουν 2 σελίδες: το *Data View* όπου εμφανίζονται τα δεδομένα τα οποία εισάγει ο χρήστης και το *Variable View* όπου εμφανίζονται οι μεταβλητές που εισαγάγει ο χρήστης, δηλαδή ο τύπος των δεδομένων.

Στην ακόλουθη εικόνα φαίνεται πως εμφανίζεται το Variable View ενώ έχει δημιουργηθεί η Μεταβλητή Βάρος (varos). Στο SPSS δεν γίνεται το όνομα της μεταβλητής να ξεκινά με νούμερο ή σύμβολο αλλά μπορεί να δοθεί ως *Label* κάτι τέτοιο.



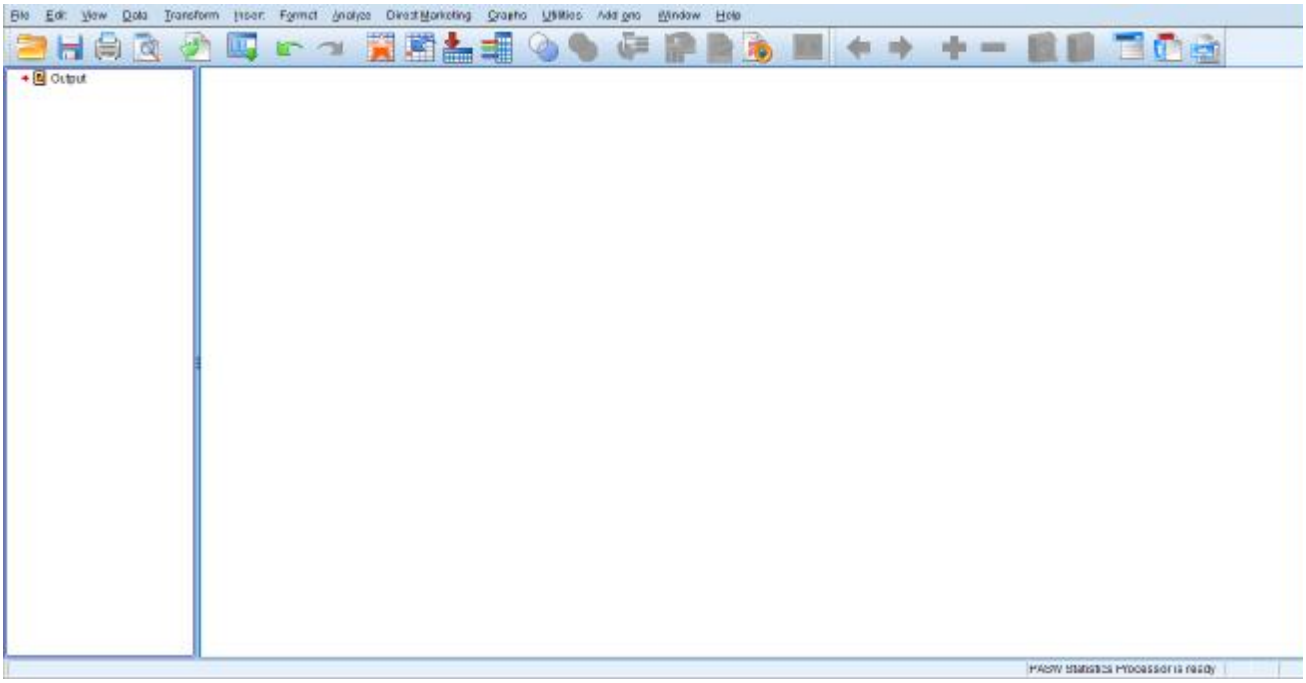
Στην ακόλουθη εικόνα φαίνεται πως εμφανίζεται το Data View ενώ έχει εισαχθεί η Μεταβλητή Βάρος (varos) και στη συνέχεια έχουμε καταχωρήσει έξι (6) τιμές για αυτήν.



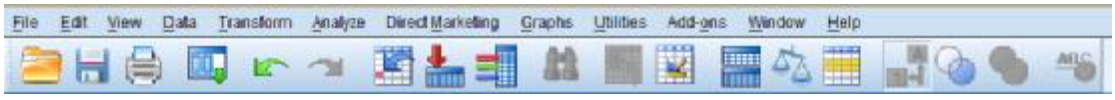
	varos	var	var	var
1	80,00			
2	76,80			
3	105,00			
4	76,60			
5	65,00			
6	45,50			
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				

Data View | Variable View

Το Output, δηλαδή το αρχείο όπου εμφανίζονται τα διάφορα αποτελέσματα του SPSS έχει την ακόλουθη μορφή:



Η μπάρα Εντολών του SPSS είναι η ακόλουθη (από κάτω εμφανίζονται εντολές σε συντόμευση):



και όπως βλέπουμε περιέχει εντολές σχετικά με:

- Το αρχείο.
- Την Διαχείριση του αρχείου.
- Την προβολή του.
- Τα δεδομένα του αρχείου.
- Τις αλλαγές που μπορούν να γίνουν στο αρχείο.
- Την ανάλυση των δεδομένων του αρχείου.
- Το Direct Marketing.
- Τα Γραφήματα που μπορούν να δημιουργηθούν.
- Την Χρησιμότητα.

- Τα Πρόσθετα.
- Τις επιλογές του Παράθυρου.
- Την βοήθεια.

2. «Οι **Μη-Παραμετρικές μέθοδοι** α) αποβλέπουν σε ευρύτερα πεδία εφαρμογής λόγω του ότι οι κατανομές στις οποίες αναφέρονται είναι λιγότερο περιορισμένες από ό,τι στα αντίστοιχα παραμετρικά προβλήματα, β) δεν είναι εξίσου ισχυρές με τις αντίστοιχες παραμετρικές μεθόδους και γ) είναι περισσότερο ευσταθείς επειδή ακριβώς δεν επηρεάζονται από την μορφή της κατανομής των δεδομένων. Παρ' όλα αυτά, οι Μη Παραμετρικές μέθοδοι συχνά είναι σχεδόν το ίδιο αποτελεσματικές όπως οι παραμετρικές μέθοδοι οι οποίες κάνουν αυστηρές υποθέσεις για τον πληθυσμό από τον οποίο προέρχονται τα δεδομένα.

Ένα έτερο σημαντικό πλεονέκτημα των μη παραμετρικών μεθόδων είναι ότι μπορούν να εφαρμοσθούν σε δεδομένα που είναι ταξινομημένα σε κατηγορίες (κατηγορικά δεδομένα) και τα οποία είναι σε κλίμακα διάταξης ή ακόμα και απλώς σε ονομαστική κλίμακα, ενώ οι παραμετρικές μέθοδοι προϋποθέτουν ακριβείς μετρήσεις. Τέλος, θα πρέπει να σημειωθεί ότι οι μη παραμετρικές μέθοδοι μπορούν να θεωρηθούν ως προπαρασκευαστικές για τις παραμετρικές μεθόδους, με την έννοια ότι, η χρησιμοποίηση μιας παραμετρικής μεθόδου, η οποία βασίζεται στην υπόθεση της κανονικότητας, θα πρέπει να έπεται ενός ελέγχου, με μία μη παραμετρική μέθοδο, της υπόθεσης ότι τα δεδομένα έχουν προέλθει από μια κανονική κατανομή».