

**ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ
ΕΛΛΑΔΑΣ**

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

Τίτλος Εργασίας

«ΣΤΑΤΙΣΤΙΚΗ ΕΛΕΓΧΟΥ ΜΕ ΤΗΝ ΧΡΗΣΗ ΤΟΥ SPSS»

Πτυχιακή Εργασία των:

Ασημακόπουλος Αλέξης

Παυλόπουλος Πάτροκλος

Επιβλέπων : Γεώργιος Κακαρελίδης

ΠΑΤΡΑ, 2015

ΠΕΡΙΛΗΨΗ

Η παρούσα εργασία υλοποιείται στα πλαίσια των προπτυχιακών μας σπουδών και έχει τίτλο «Στατιστική ελέγχου με την χρήση του spss». Στην αρχή παρουσιάζουμε βασικές έννοιες σχετικές με την στατιστική, όπως είναι ο πληθυσμός, το δείγμα, τα δεδομένα, οι μεταβλητές κλπ. Στη συνέχεια μέσα από τη βιβλιογραφική ανασκόπηση προσπαθούμε να περιγράψουμε τους κλάδους της στατιστικής, περιγραφικής και επαγωγική (ή στατιστική συμπερασματολογία). Στο δεύτερο κεφάλαιο εκτός των ανωτέρω περιγράφουμε επίσης τη έννοια του στατιστού ελέγχου, ποια η χρησιμότητα αυτού και πως σχετίζεται με την οικονομία και τη διοίκηση. Στο επόμενο (τρίτο κεφάλαιο), περιγράφουμε την εξέλιξη και χρήση του στατιστικού πακέτου SPSS, ύστερα από την περιγραφή ορισμένων διδασκαστικών τεχνικών θεμάτων στη συνέχεια εστιάζουμε στον τρόπο με τον οποίο πραγματοποιούνται οι έλεγχοι υποθέσεων με τη χρήση του συγκεκριμένου εργαλείου. Το τέταρτο κεφάλαιο παρουσιάζει ένα παράδειγμα στο οποίο πραγματοποιούνται παραμετρικοί και μη παραμετρικοί ελεγχοι υποθέσεων. Η εργασία ολοκληρώνεται με την εξαγωγή συμπερασμάτων και τη βιβλιογραφία από όπου αντλήθηκαν τα βασικά στοιχεία για τη συγγραφή αυτής. Σε όλο το μήκος αυτής είναι έκδηλη η σπουδαιότητα αλλά και η χρησιμότητα της στατιστικής σε πληθώρα τομέων, κοινωνιολογία, οικονομία, διοίκηση, κ.α..

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ

Στατιστική, SPSS, Έλεγχοι Υποθέσεων, Οικονομία, Διοίκηση, t-test, F-statistic, Chi-square

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΕΡΙΛΗΨΗ.....	1
ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ	1
ΠΡΟΛΟΓΟΣ	5
ΚΕΦΑΛΑΙΟ 1 ^ο : ΕΙΣΑΓΩΓΙΚΕΣ ΕΝΝΟΙΕΣ.....	7
1.1. ΕΙΣΑΓΩΓΗ	7
1.2. ΕΡΕΥΝΗΤΙΚΑ ΕΡΩΤΗΜΑΤΑ.....	8
1.3. ΒΑΣΙΚΕΣ ΣΤΑΤΙΣΤΙΚΕΣ ΕΝΝΟΙΕΣ - ΟΡΟΙ.....	8
1.3.1. ΠΛΗΘΥΣΜΟΣ ΚΑΙ ΔΕΙΓΜΑ.....	8
1.3.2. ΚΑΤΗΓΟΡΙΕΣ ΣΤΑΤΙΣΤΙΚΩΝ ΜΕΘΟΔΩΝ	9
1.3.3. ΚΑΤΗΓΟΡΙΕΣ ΜΕΤΑΒΛΗΤΩΝ ΚΑΙ ΔΕΔΟΜΕΝΩΝ	10
1.4. ΣΤΑΤΙΣΤΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ – ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ	12
ΚΕΦΑΛΑΙΟ 2 ^ο : ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΕΠΙΣΚΟΠΗΣΗ: Η ΣΤΑΤΙΣΤΙΚΗ ΩΣ ΕΡΓΑΛΕΙΟ	15
2.1. ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ	16
2.1.1. ΠΑΡΑΔΕΙΓΜΑΤΑ ΠΕΡΙΓΡΑΦΙΚΗΣ ΣΤΑΤΙΣΤΙΚΗΣ.....	18
2.2. ΕΠΑΓΩΓΙΚΗ ΣΤΑΤΙΣΤΙΚΗ.....	20
2.2.1 Ο ΣΤΑΤΙΣΤΙΚΟΣ ΕΛΕΓΧΟΣ ΩΣ ΕΠΑΓΩΓΙΚΟ ΜΕΤΡΟ.....	21
2.2.2. Η ΧΡΗΣΙΜΟΤΗΤΑ ΤΟΥ ΣΤΑΤΙΣΤΙΚΟΥ ΕΛΕΓΧΟΥ	29
2.2.3. ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ	30
2.2.4. ΤΟ Χ ² ΣΤΑΤΙΣΤΙΚΟ ΤΕΣΤ (Chi –Square Test).....	32
2.2.5. ΣΥΣΧΕΤΙΣΗ.....	33
2.3. ΜΟΝΤΕΛΟΠΟΙΗΣΗ.....	36
2.4. ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΟΙΚΟΝΟΜΙΑ.....	39
2.5. ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΔΙΟΙΚΗΣΗ ΕΠΙΧΕΙΡΗΣΕΩΝ	41

2.6. ΣΧΕΤΙΚΑ ΠΑΡΑΔΕΙΓΜΑΤΑ	42
ΚΕΦΑΛΑΙΟ 3 ^ο : SPSS: ΕΞΕΛΙΞΗ ΚΑΙ ΧΡΗΣΗ.....	45
3.1. ΣΤΑΤΙΣΤΙΚΗ ΜΕ ΤΗ ΧΡΗΣΗ SPSS	45
3.2. ΤΙ ΕΙΝΑΙ SPSS - ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ – ΕΞΕΛΙΞΗ.....	45
3.3. ΔΥΝΑΤΟΤΗΤΕΣ ΤΟΥ SPSS.....	50
3.3.1. ΣΤΑΤΙΣΤΙΚΟΙ ΕΛΕΓΧΟΙ ΜΕ ΤΟ SPSS	53
ΚΕΦΑΛΑΙΟ 4 ^ο : ΕΦΑΡΜΟΓΗ	57
ΣΥΜΠΕΡΑΣΜΑΤΑ	105
ΒΙΒΛΙΟΓΡΑΦΙΑ	107

ΜΕΡΟΣ Α΄: ΘΕΩΡΗΤΙΚΗ ΠΡΟΣΕΓΓΙΣΗ

ΠΡΟΛΟΓΟΣ

Στην σημερινή εποχή, κάθε φαινόμενο της φύσης και οποιαδήποτε ενέργεια του ανθρώπου αποτελεί πεδίο εφαρμογής της Στατιστικής. Ιδιαίτερα χρήσιμη είναι στις Κοινωνικές, Οικονομικές και Διοικητικές επιστήμες. Οι στατιστικές τεχνικές είναι απαραίτητες για την επιτυχή λειτουργία κάθε επιχείρησης γιατί παρέχουν στην Διοίκηση όλες τις απαραίτητες πληροφορίες που αφορούν την ίδια την επιχείρηση (εσωτερικό περιβάλλον) αλλά και τους ανταγωνιστές (εξωτερικό περιβάλλον). Οι επιχειρήσεις χρησιμοποιούν ολοένα και περισσότερο στατιστικές μεθόδους για να μετατρέπουν δεδομένα σε πληροφορίες.

Η ραγδαία ανάπτυξη των νέων τεχνολογιών και η ενσωμάτωσή τους στην παραγωγή και τις υπηρεσίες έχει δημιουργήσει πρωτόγνωρες συνθήκες για τους χειριστές της οικονομικής πολιτικής. Η πρόκληση που αντιμετωπίζουν είναι: 1) να μπορέσουν να παρακολουθήσουν όσα διαδραματίζονται μέσω της συλλογής κατάλληλων στατιστικών στοιχείων, και 2) μέσω οικονομικής ανάλυσης να προτείνουν και να εφαρμόσουν κατάλληλες πολιτικές που θα οδηγήσουν στη μέγιστη δυνατή ανάπτυξη και ευημερία.

Η Στατιστική χρησιμοποιείται σε όλες σχεδόν τις επιστήμες και στους περισσότερους τομείς της ανθρώπινης δραστηριότητας (οικονομία, βιομηχανία, εμπόριο, δημογραφία, μετεωρολογία, πολιτική, ιατρική, κ.ά.). Η επίδρασή της στη ζωή μας είναι πολύ μεγάλη. Ως εφαρμοσμένη επιστήμη ασχολείται με τις επιστημονικές μεθόδους σχεδιασμού μιας μελέτης, συλλογής, επεξεργασίας, παρουσίασης και ανάλυσης αριθμητικών δεδομένων, με σκοπό την εξαγωγή συμπερασμάτων για τη λήψη σωστών αποφάσεων. Επίσης, αρκετά συχνά, τόσο στην καθημερινή ζωή όσο και στις επιστήμες εμφανίζεται η ανάγκη μελέτης φαινομένων, η έκβαση των οποίων δεν είναι γνωστή εκ των προτέρων και, ως εκ τούτου, δεν μπορεί να προβλεφθεί με βεβαιότητα.

Στατιστική είναι Επιστήμη της Συλλογής, Παρουσίασης, Ανάλυσης και Ερμηνείας των προς έρευνα αριθμητικών δεδομένων που αναφέρονται σε φυσικά ή κοινωνικά φαινόμενα. Ο κλάδος της Στατιστικής συμβαδίζει στον επιστημονικό χώρο με κάθε άλλη Επιστήμη που απαιτεί επεξεργασία δεδομένων, τεκμηρίωση και αξιοπιστία αποτελεσμάτων. Η Στατιστική επίσης, κατά την εκτίμηση και αξιολόγηση των προς έρευνα δεδομένων, διαθέτει μεθόδους Πρόγνωσης,

και κατά συνέπεια παρέχει αξιοπιστία Πρόβλεψης αποτελεσμάτων, προς όλες τις επιστήμες που υιοθετούν τις στατιστικές μεθόδους, ελαχιστοποιώντας και ελέγχοντας το πιθανόν περιθώριο λάθους.

Ο κλάδος της Στατιστικής θεωρείται αναγκαίος και σημαντικός σε όλους τους τομείς, τόσο της Επιστήμης, όσο και της Αγοράς Εργασίας, παρέχοντας γνώση και εφόδια κατάρτισης απαραίτητα στις σύγχρονες κοινωνίες. Η γνώση και η πρακτική εφαρμογή στατιστικών μεθόδων και εργαλείων, εξασφαλίζει εύκολη αντιμετώπιση και εφαρμογή των αντίστοιχων διαδικασιών. Αξίζει επίσης να αναφερθεί ότι η αναγκαιότητα της χρήσης της Στατιστικής επιβάλλει άμεση και ευρεία κατανόηση των εννοιών και εργαλείων αυτής. Κατά συνέπεια, η βαρύτητα του μαθήματος πρέπει να εστιάζεται σε σημεία ενδιαφέροντος άμεσης πρακτικής εφαρμογής.

Στην εποχή της τεχνολογικής ανάπτυξης και πληροφορικής έκρηξης, ως Στατιστική ορίζεται η επιστήμη που απασχολείται με την συγκέντρωση, παρουσίαση, αξιολόγηση και επεξεργασία συμπερασμάτων. Η ιδιαιτερότητα της ως επιστήμη είναι ότι ενώ σε θεωρητικό επίπεδο η χρήση μαθηματικών μοντέλων είναι αναγκαία για την επίλυση προβλημάτων, στο επίπεδο εφαρμογών χρησιμοποιεί το πλαίσιο όλων σχεδόν των άλλων γνωστικών περιοχών (ιατρική, χρηματοοικονομικά, μάρκετινγκ, αστρονομία, αρχαιολογία, ψυχολογία κ.α.). Τα άτομα που ασχολούνται με την στατιστική ονομάζονται στατιστικοί αναλυτές.

Δουλειά του στατιστικού αναλυτή είναι να αναπτύσσει μοντέλα κατάλληλα τόσο ως προς τις ερωτήσεις όσο προς τα δεδομένα. Επιπλέον πρέπει να μπορεί να αντλήσει από τα μοντέλα αυτά σχετικές πληροφορίες που πηγάζουν από τα δεδομένα.

ΚΕΦΑΛΑΙΟ 1^ο: ΕΙΣΑΓΩΓΙΚΕΣ ΕΝΝΟΙΕΣ

1.1. ΕΙΣΑΓΩΓΗ

Είναι γνωστό ότι η επιστημονική έρευνα βασίζεται στην παρατήρηση φαινομένων για να συλλέξει και να αναλύσει πληροφορίες και μετρήσεις που προέρχονται από αυτές τις παρατηρήσεις, και στη συνέχεια να προχωρήσει στην εξαγωγή χρήσιμων και γενικεύσιμων συμπερασμάτων. Η αναφορά σε αυτό το σύνολο των πληροφοριών και μετρήσεων συνήθως γίνεται με τον επιστημονικό όρο δεδομένα (data). Ένας ερευνητής διεξάγει επιστημονικές μελέτες και έρευνες, με σκοπό να δώσει απαντήσεις σε συγκεκριμένα ερωτήματα και γενικούς προβληματισμούς που σχετίζονται με μία συγκεκριμένη ομάδα ή ομάδες ατόμων ή αντικειμένων (ή άλλων οντοτήτων).

Η επιστημονική έρευνα επιτρέπει στους ερευνητές να προβαίνουν σε παρατηρήσεις των φαινομένων και των προβλημάτων που εξετάζουν. Η Στατιστική παρέχει τα κατάλληλα εργαλεία για το σχεδιασμό της διαδικασίας συλλογής, περιγραφής, ανάλυσης και επεξεργασίας αυτών των παρατηρήσεων. Δηλαδή, ένα σύνολο από μεθόδους και κανόνες που επιτρέπουν την οργάνωση μιας συλλογής δεδομένων, τη συνοπτική και αποτελεσματική παρουσίαση και ανάλυση τους, και τέλος την εξαγωγή χρήσιμων συμπερασμάτων.

Εύκολα γίνεται αντιληπτό λοιπόν, ότι η επιστήμη της στατιστικής παρέχει εργαλεία για οποιαδήποτε επιστημονική έρευνα. Το παρόν κεφάλαιο αποσκοπεί στην εξοικείωση του αναγνώστη με τις βασικές στατιστικές έννοιες και μεθόδους που χρησιμοποιούνται στα πλαίσια των επιστημονικών ερευνών, και πιο συγκεκριμένα των ερευνών που μελετούν την επίδραση της χρήσης συνεργατικών τεχνολογιών στην εργασία και στη μάθηση. Είναι εμφανές ότι στα πλαίσια ενός κεφαλαίου είναι αδύνατο να γίνει μία αναλυτική παρουσίαση όλων των τεχνικών που περιλαμβάνει η στατιστική ανάλυση δεδομένων. Για τους αναγνώστες που ενδιαφέρονται να ασχοληθούν περισσότερο με τον επιστημονικό κλάδο της στατιστικής υπάρχει πληθώρα

διαθέσιμων συγγραμμάτων (Gravetter & Wallnau 1996, Howell, 1989, Klugh, 1986, Keppel and Saufley, 1980).

1.2. ΕΡΕΥΝΗΤΙΚΑ ΕΡΩΤΗΜΑΤΑ

Στη συγκεκριμένη εργασία σκοπός μας είναι να διαπιστώσουμε σε ποιο βαθμό η χρήση SPSS διευκολύνει τη λειτουργία της διοίκησης των επιχειρήσεων και συμβάλλει στην καλύτερη δυνατή λειτουργία αυτών. Αντλώντας δεδομένα από προηγούμενες έρευνες ή από παραδείγματα σχετικά με τη χρήση του εν λόγω πακέτου στην οικονομία και τη διοίκηση θα πραγματοποιήσουμε ορισμένους ελέγχους προκειμένου να προσδιορίσουμε τη χρησιμότητά αυτού κατά τη διάρκεια λήψης αποφάσεων κατά την επιχειρηματική λειτουργία και δραστηριότητα.

1.3. ΒΑΣΙΚΕΣ ΣΤΑΤΙΣΤΙΚΕΣ ΕΝΝΟΙΕΣ - ΟΡΟΙ

Σε αυτό το σημείο, και πριν προχωρήσουμε σε μία σύντομη περιγραφή των μεθόδων που περιλαμβάνει η στατιστική για την οργάνωση και ανάλυση των δεδομένων, κρίνεται σκόπιμο να παρουσιαστούν οι βασικές έννοιες και οι όροι που χρησιμοποιούνται στη στατιστική.

1.3.1. ΠΛΗΘΥΣΜΟΣ ΚΑΙ ΔΕΙΓΜΑ

Όπως αναφέρεται και παραπάνω, η επιστημονική έρευνα σχετίζεται με ένα γενικό προβληματισμό ή μια συγκεκριμένη ερώτηση που αφορά ομάδες ατόμων, αντικειμένων ή άλλων οντοτήτων. Οι ομάδες αυτές αποτελούν το επίκεντρο του ενδιαφέροντος σε μία μελέτη, και είναι γνωστές με τον όρο πληθυσμός (population) στην επιστήμη της στατιστικής. Ο πληθυσμός μπορεί να αποτελείται από έμψυχα αντικείμενα (ομάδες ανθρώπων), από άψυχα αντικείμενα (εξαρτήματα αυτοκινήτων), ή από οτιδήποτε άλλο ο ερευνητής θέλει να εξετάσει (αφηρημένες έννοιες). Είναι εμφανές ότι ένας πληθυσμός μπορεί να είναι από πολύ μικρός (π.χ. οι έφηβοι σε μία τάξη σε ένα σχολείο), έως πολύ μεγάλος (π.χ. το σύνολο όλων των εφήβων του

πλανήτη). Επειδή η επιστημονική έρευνα οδηγεί σε ένα σύνολο από παρατηρήσεις για τα υποκείμενα (subjects) του πληθυσμού, τις περισσότερες φορές ο όρος πληθυσμός αναφέρεται σε ένα σύνολο παρατηρήσεων που περιγράφουν τα υπό εξέταση υποκείμενα.

Όμως συχνά η μελέτη ολόκληρου του πληθυσμού δεν είναι πρακτικά δυνατή, γιατί το μέγεθος του είναι πολύ μεγάλο. Για αυτό το λόγο, τις περισσότερες φορές ένα υποσύνολο του πληθυσμού, που ονομάζεται δείγμα (sample), αποτελεί το αντικείμενο της μελέτης. Το δείγμα λοιπόν, είναι ένα σύνολο από υποκείμενα που έχει επιλεγεί κατάλληλα ώστε να αντιπροσωπεύει έναν ολόκληρο πληθυσμό. Προφανώς, η διαδικασία επιλογής του δείγματος αποτελεί μία εξαιρετικά σημαντική διαδικασία, καθώς καθορίζει την εγκυρότητα των αποτελεσμάτων της μελέτης. Αν το δείγμα δεν είναι αντιπροσωπευτικό, υπάρχει μεγάλη πιθανότητα η μελέτη να οδηγήσει σε εσφαλμένα συμπεράσματα.

Στην πράξη, ένα δείγμα είναι αντιπροσωπευτικό του πληθυσμού όταν έχει χρησιμοποιηθεί η διαδικασία της τυχαίας δειγματοληψίας (random sampling) για την απόκτηση του. Η τυχαία δειγματοληψία απαιτεί κάθε υποκείμενο του πληθυσμού να έχει την ίδια πιθανότητα να επιλεγεί. Όπως και οι πληθυσμοί, ένα δείγμα μπορεί να είναι αντίστοιχα από πολύ μικρό έως και πολύ μεγάλο.

Εύκολα γίνεται αντιληπτό ότι όσο μεγαλύτερο είναι το δείγμα, τόσο πιο αντιπροσωπευτικό θα είναι, καθώς αυξάνεται ο αριθμός των υποκειμένων που επιλέγονται από τον πληθυσμό.

Στην στατιστική όταν χρησιμοποιούμε δεδομένα κρίνεται αναγκαίο να προσδιορίζουμε εάν τα δεδομένα προέρχονται από ένα πληθυσμό ή από ένα δείγμα. Για να εξυπηρετηθεί αυτός ο διαχωρισμός, η στατιστική χρησιμοποιεί τον όρο παράμετρος (parameter) για να περιγράψει δεδομένα που αναφέρονται στον πληθυσμό, και τον όρο στατιστικός δείκτης (statistic) για τα δεδομένα που σχετίζονται με ένα δείγμα

1.3.2. ΚΑΤΗΓΟΡΙΕΣ ΣΤΑΤΙΣΤΙΚΩΝ ΜΕΘΟΔΩΝ

Αν και η επιστήμη της στατιστικής προτείνει ένα μεγάλο σύνολο από τεχνικές για την οργάνωση και την ανάλυση των δεδομένων, αυτές οι τεχνικές μπορούν να κατηγοριοποιηθούν σε δύο γενικές κατηγορίες.

Η πρώτη κατηγορία, η περιγραφική στατιστική (descriptive statistics), περιλαμβάνει μεθόδους για την οργάνωση, απλοποίηση και συνοπτική παρουσίαση των δεδομένων. Αν και υπάρχουν πολλές τεχνικές που ανήκουν σε αυτήν την κατηγορία, η πιο διαδεδομένη είναι ο υπολογισμός της μέσης τιμής (mean) και της τυπικής απόκλισης (standard deviation). Άλλοι περιγραφικοί στατιστικοί δείκτες θα παρουσιαστούν συνοπτικά στη συνέχεια του κεφαλαίου.

Όπως ήδη αναφέρθηκε, τις περισσότερες φορές είναι αδύνατο να συλλέξουμε μετρήσεις για κάθε υποκείμενο του πληθυσμού, και γι αυτό χρησιμοποιούμε ένα αντιπροσωπευτικό υποσύνολο του, τα δείγματα. Η δεύτερη κατηγορία, η επαγωγική στατιστική (inferential statistics), περιλαμβάνει τεχνικές που επιτρέπουν την ανάλυση τέτοιων δεδομένων ώστε να γίνει εφικτή η εξαγωγή χρήσιμων και γενικεύσιμων συμπερασμάτων για τους πληθυσμούς, με βάση τις πληροφορίες που συλλέγονται από τα δείγματα.

Ωστόσο, είναι προφανές ότι όσο αντιπροσωπευτικό και αν θεωρείται ένα δείγμα, δε μπορεί να παρέχει μια πιστή αναπαράσταση του πληθυσμού. Έτσι, πάντα θα υπάρχει μία ασυμφωνία, ή με άλλα λόγια ένα ποσοστό λάθους, ανάμεσα στο στατιστικό δείκτη που προκύπτει από το δείγμα, και την αντίστοιχη τιμή της παραμέτρου του πληθυσμού. Αυτό το ποσοστό λάθους ονομάζεται σφάλμα δειγματοληψίας (sampling error), και αποτελεί ένα από τα κύρια προβλήματα που καλείται να αντιμετωπίσει ο ερευνητής όταν προσπαθεί να εξάγει γενικά συμπεράσματα για πληθυσμούς έχοντας στη διάθεση του ένα αριθμό από δείγματα. Το σφάλμα δειγματοληψίας τονίζει το γεγονός ότι ένας στατιστικός δείκτης αποτελεί μόνο μία εκτίμηση της αντίστοιχης παραμέτρου του πληθυσμού.

1.3.3. ΚΑΤΗΓΟΡΙΕΣ ΜΕΤΑΒΛΗΤΩΝ ΚΑΙ ΔΕΔΟΜΕΝΩΝ

Τυπικά, αυτό που ενδιαφέρει έναν ερευνητή είναι να απαντήσει σε ερωτήματα που σχετίζονται με έναν ή περισσότερους πληθυσμούς, έχοντας στη διάθεση του κάποιο αριθμό δειγμάτων.

Είναι φανερό ότι τις περισσότερες φορές ο στόχος της έρευνας είναι να δημιουργήσει μία σχέση αιτίου-αποτελέσματος ανάμεσα σε δύο μεταβλητές (variables)¹. Με άλλα λόγια, ο στόχος σε μία έρευνα είναι να αποδειχθεί ότι οι αλλαγές που εμφανίζονται στην τιμή μιας μεταβλητής οφείλονται στην αλλαγή της τιμής μιας άλλης μεταβλητής. Για να γίνει αυτό εφικτό, η επιστημονική έρευνα χρησιμοποιεί δείγματα και προσεκτικά σχεδιασμένα πειράματα, όπου ο ερευνητής μπορεί να ελέγχει την τιμή μιας μεταβλητής και ταυτόχρονα να παρατηρεί τις τιμές της άλλης μεταβλητής, χωρίς η διαδικασία του πειράματος να επηρεάζεται από εξωτερικούς παράγοντες ή μεταβλητές. Η μεταβλητή που ελέγχει ο ερευνητής ονομάζεται ανεξάρτητη μεταβλητή (independent variable), ενώ αυτή που αποτελεί το αντικείμενο παρατήρησης ονομάζεται εξαρτημένη μεταβλητή (dependent variable). Το ζητούμενο για τον ερευνητή είναι να εξετάσει την επίδραση των διαφορετικών τιμών της ανεξάρτητης μεταβλητής στις τιμές της εξαρτημένης.

Όπως έχει ήδη αναφερθεί, μία επιστημονική μελέτη απαιτεί τη συλλογή ενός συνόλου δεδομένων. Τα δεδομένα που συλλέγονται μπορεί να είναι ποιοτικά ή ποσοτικά. Ποιοτικά ονομάζονται τα δεδομένα που δεν επιδέχονται μέτρηση, όπως για παράδειγμα το επίπεδο της μόρφωσης, το φύλο, το χρώμα των ματιών, κλπ, ενώ ποσοτικά αυτά που μπορούν να μετρηθούν με τη βοήθεια κάποιας κλίμακας, όπως για παράδειγμα το βάρος, το ύψος, ο βαθμός επίδοσης σε κάποιες εξετάσεις, κλπ. Τα ποιοτικά δεδομένα μπορεί να είναι οργανωμένα απλά σαν ονομαστικές κατηγορίες (nominal data), όπως για παράδειγμα το φύλο, ή σαν ταξινομημένες κατηγορίες (ordinal data), όπως για παράδειγμα η κατάταξη των μεταλλίων στους Ολυμπιακούς Αγώνες. Τα ποσοτικά δεδομένα διακρίνονται σε συνεχή (continuous), όπως για παράδειγμα ο χρόνος επίλυσης ενός μαθηματικού προβλήματος, και διακριτά (discrete), όπως για παράδειγμα ο αριθμός των παιδιών που έχει μία οικογένεια.

¹ Μεταβλητή ονομάζεται οποιοδήποτε χαρακτηριστικό ή κατάσταση παρουσιάζει αλλαγή, ή έχει διαφορετική τιμή για διαφορετικά υποκείμενα τα οποία ανήκουν σε ένα πληθυσμό ή ένα δείγμα. Δεν είναι ασυνήθιστο ο ερευνητής να ελέγχει και να παρατηρεί πολλές μεταβλητές ταυτόχρονα, αλλά για ακαδημαϊκούς λόγους χρησιμοποιούμε την απλούστερη περίπτωση όπου ελέγχει μία μεταβλητή και παρατηρεί μία άλλη.



1.4. ΣΤΑΤΙΣΤΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ – ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ

Η Στατιστική είναι μία μεθοδική μαθηματική, παλαιότερα τεχνική και σήμερα επιστήμη που επιχειρεί να εξαγάγει έγκυρη γνώση χρησιμοποιώντας εμπειρικά δεδομένα παρατήρησης, ή πειράματος. Κύριο αντικείμενο έρευνας και μελέτης της Στατιστικής είναι η συλλογή, ταξινόμηση, επεξεργασία, παρουσίαση, ανάλυση και ερμηνεία διαφόρων δεδομένων με απώτερο στόχο την εξαγωγή ασφαλών συμπερασμάτων για λήψη ορθών αποφάσεων. Πρόκειται για σημαντική επιστήμη της οποίας οι εφαρμογές έχουν ευρύτατο πεδίο στη διοικητική, τις επιχειρήσεις, καθώς και στις θετικές και συμπεριφορικές ή Κοινωνικές επιστήμες.

Η Στατιστική αποτελεί σήμερα κλάδο των εφαρμοσμένων μαθηματικών, οι δε ασχολούμενοι στο στατιστικό πεδίο έρευνας και ανάλυσης καλούνται γενικά στατιστικοί ή στατιστικολόγοι.

Γενικά ο όρος Στατιστική φέρεται με διττή σημασία², αφενός υποδηλώνοντας μαθηματικές μεθόδους χειρισμού δεδομένων που λήφθηκαν με απαρίθμηση ή μέτρηση και αφετέρου αυτά τα ίδια τα δεδομένα που έχουν υποστεί αυτούς τους χειρισμούς. Σύμφωνα με τον Α. Αλεξόπουλο εξετάζοντας τον ορισμό όπως αυτός καθορίστηκε στη δεκαετία του 1950 "Στατιστική είναι σύνολο μεθόδων που καθοδηγούν στη λήψη ορθών αποφάσεων σε περιπτώσεις αβεβαιότητας" τονίζει την εννοιολογική διάκριση του συνόλου των στοιχείων ενός φαινομένου και το σύνολο των μεθόδων που εξετάζουν αυτά προς τον κοινό σκοπό.

² UNESCO "Λεξικό Κοινωνικών Όρων" τομ.3ος, σελ. 865

Σύμφωνα με το Λεξικό Οικονομικοτεχνικών Όρων του Ελληνικού Κέντρου Παραγωγικότητας³ "Στατιστική είναι α) τα αριθμητικά δεδομένα που αναφέρονται σε σύνολο ατόμων, (έμψυχων, άψυχων, φαινόμενα κ,λπ.) και β) επιστήμη συλλογής, ανάλυσης και ερμηνείας τούτων των δεδομένων".

Η Στατιστική έρευνα βασίζεται στη χρήση της στατιστικής θεωρίας, ενός κλάδου των εφαρμοσμένων μαθηματικών. Στη στατιστική, η τυχαιότητα και η απροσδιοριστία ορίζονται στα πλαίσια της θεωρίας πιθανοτήτων. Η πρακτική της στατιστικής περιλαμβάνει την σχεδίαση, συλλογή και ερμηνεία δεδομένων που προκύπτουν από αβέβαιες παρατηρήσεις. Επειδή η στατιστική αποσκοπεί στην εξαγωγή των «καλύτερων» πληροφοριών από τα διαθέσιμα δεδομένα, κατατάσσεται από μερικούς σαν κλάδος της θεωρίας των αποφάσεων.

Η Στατιστική ως έννοια εμφανίζεται από τους μυθικούς χρόνους από της πρώτης δημιουργίας οργανωμένων κοινωνιών. Μια πρώτη γραφή στατιστικής μορφής με αριθμητικά δεδομένα είναι ο νεών κατάλογος (κατάλογος των πλοίων) των Αχαιών στον Τρωικό πόλεμο από τον Όμηρο⁴. Από τον κατάλογο αυτό οι ιστορικοί απέσπασαν σημαντικές εκτιμήσεις της οικονομικής ευρωστίας και του πληθυσμού των πόλεων-κρατών που σημετείχαν καθώς και σημαντικά στοιχεία για την τότε ναυπηγική, ναυτιλία και ναυτική τέχνη. Πρώτη ιστορική συλλογή καθαρά στατιστικών στοιχείων θεωρείται η απογραφή πληθυσμού από τον Αυτοκράτορα της Κίνας Γιάο (Υαο) το 2238 π.Χ. Στοιχειώδεις τέτοιες απογραφές είχαν πραγματοποιήσει και άλλοι αρχαίοι λαοί όπως οι αρχαίοι Αιγύπτιοι, οι Βαβυλώνιοι, οι Πέρσες οι αρχαίοι Έλληνες και οι Ρωμαίοι με χαρακτηριστικότερη την απογραφή του Οκταβιανού του Αυγούστου.

Στην αρχαιότητα πρώτος στόχος συλλογής στατιστικών στοιχείων ήταν η στράτευση και η φορολόγηση τόσο των πολιτών όσο και ολόκληρων πόλεων, π.χ. των σατραπειών της Περσίας, πόλεων της "Αθηναϊκής Συμμαχίας" κ.λπ. Χαρακτηριστική ήταν και η κοινωνική διάρθρωση της αρχαίας Αθήνας στην Αρχαϊκή εποχή λαμβάνοντας υπόψη ως στατιστικά στοιχεία τον μέδιμνο και τον ίππο. Αλλά και η εκπροσώπηση των φυλών και Δήμων της Αθήνας στην Εκκλησία του Δήμου, οι ψηφοφορίες ακόμα και ο οστρακισμός στηρίζονταν σε στατιστικά δεδομένα. Αργότερα με βάση στατιστικών στοιχείων προχώρησαν οι Ρωμαίοι στη διοικητική διαίρεση της

³ Λεξικό Οικονομικοτεχνικών Όρων του Ελληνικού Κέντρου Παραγωγικότητας, 2η έκδοση - Αθήνα 1965

⁴ Ιλιάδα β' (στ. 494-759).

Αυτοκρατορίας τους και ακολούθως η Βυζαντινή Αυτοκρατορία δημιουργώντας τα βυζαντινά θέματα.

Η συστηματική όμως συλλογή δεδομένων για πληθυσμό και οικονομία (δημογραφική στατιστική) άρχισε στη διάρκεια της Αναγέννησης και ειδικότερα στη Βενετία και την Φλωρεντία όπου και γρήγορα επεκτάθηκε σ' όλα τα τότε Βασίλεια της Ευρώπης. Περί το τέλος του 11ου αιώνα, επί εποχής Γουλιέλμου του Κατακτητή, πραγματοποιήθηκε μια σπουδαία στατιστική απογραφή που αφορούσε μονάδες παραγωγής της Αγγλίας, όπως μεταλλεία, ιχθυοτροφεία κ.λπ. Οι μεγάλοι όμως ρυθμοί θνησιμότητας που άρχισαν να παρατηρούνται λίγο αργότερα, από επιδημικές ασθένειες, πολέμους και λιμοκτονίες έδωσαν ιδιαίτερη ώθηση στη στατιστική έρευνα καταγράφοντας αιτίες και απώλειες. Έτσι το 1348 ξεκίνησαν οι καταγραφές θανάτων από την πανώλη, την φοβερή ασθένεια που κράτησε τέσσερις αιώνες. Στις καταγραφές αυτές προστέθηκαν και θάνατοι από άλλες αιτίες. Το 1620 ο Άγγλος έμπορος Τζον Γκράουντ ξεκίνησε πρώτος τη δειγματοληπτική έρευνα σε οικογένειες του Λονδίνου όπου και διαπίστωσε ότι σε κάθε 88 άτομα υπήρχαν τρεις θάνατοι. Από το στοιχείο αυτό και χρησιμοποιώντας τους εν λόγω καταλόγους που έδιναν 13.200 θανάτους εκτιμήθηκε ότι ο πληθυσμός του Λονδίνου το 1620 αριθμούσε 387.000 κατοίκους.

Έτσι πολλοί επιστήμονες θέτουν αφετηρία της Στατιστικής το έτος 1663, με την έκδοση του βιβλίου Φυσικές και Πολιτικές παρατηρήσεις της Θνησιμότητας του John Graunt⁵.

Η ραγδαία ανάπτυξη του εμπορίου που σημειώθηκε από τον 16ο μέχρι τον 19ο αιώνα εξανάγκασε τις αρχές των κρατών στη μελέτη των νέων οικονομικών δεδομένων του εμπορίου των μεταφορών και των βιομηχανιών καθώς και του εργατικού δυναμικού. Σήμερα η στατιστική έρευνα από μαθηματική τεχνική έχει αναχθεί σε σπουδαία αυτοτελή επιστήμη ακολουθώντας ιδιαίτερες μεθόδους ανάλυσης.

⁵ Willcox, Walter (1938) The Founder of Statistics. Review of the International Statistical Institute 5(4):321–328.

ΚΕΦΑΛΑΙΟ 2^ο: ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΕΠΙΣΚΟΠΗΣΗ: Η ΣΤΑΤΙΣΤΙΚΗ ως ΕΡΓΑΛΕΙΟ

Η επιστήμη της στατιστικής αποτελεί ένα αναντικατάστατο εργαλείο για κάθε ερευνητή, καθώς επιτρέπει την οργάνωση μιας συλλογής δεδομένων, τη συνοπτική και αποτελεσματική παρουσίαση τους, την ανάλυση τους, και τέλος την εξαγωγή χρήσιμων και γενικεύσιμων συμπερασμάτων.

Η στατιστική είναι η επιστήμη που ασχολείται με τη συγκέντρωση, παρουσίαση, αξιολόγηση και επεξεργασία της πληροφορίας, με στόχο την εξαγωγή συμπερασμάτων και χρησιμεύει σε όλες της επιστήμες, όπως στην οικονομία, κοινωνιολογία, ιατρική, βιολογία, ψυχολογία, κτλ. Η στατιστική απαιτεί πολύ καλή γνώση του φαινομένου που μελετά, σωστή οργάνωση των παρατηρήσεων και σωστή καταγραφή των δεδομένων. Οι πληροφορίες που χρησιμοποιεί πρέπει να είναι πλήρεις, ποιοτικά ελεγμένες και χωρίς λάθη. Για την επεξεργασία των δεδομένων απαιτείται θεωρητική γνώση ώστε να επιλεγούν οι σωστές τεχνικές και μέθοδοι για την στατιστική ανάλυση που επιδιώκουμε αλλά και να πληρούνται οι κατάλληλες προϋποθέσεις για την εφαρμογή των στατιστικών τεχνικών.

Οι μετρήσεις ή αλλιώς τα δεδομένα μιας επιστημονικής έρευνας είναι συγκεκριμένες πληροφορίες, στοιχεία ή γεγονότα, οι οποίες μπορούν να χρησιμοποιηθούν για περαιτέρω αποφάσεις και συμπεράσματα. Γεγονός είναι ότι όλες οι ερευνητικές προσπάθειες οι οποίες παράγουν δεδομένα επιδέχονται στατιστική επεξεργασία. Η επεξεργασία αυτή διακρίνεται σε δύο φάσεις. Οι μέθοδοι της περιγραφικής στατιστικής αποτελούν το επιστημονικό εργαλείο για την συγκέντρωση, ταξινόμηση και παρουσίαση των πρωτογενών δεδομένων, ενώ τα εργαλεία της στατιστικής συμπερασματολογίας επιτρέπουν την εκτίμηση χαρακτηριστικών του συνόλου των δεδομένων από την μελέτη ενός υποσυνόλου τους.

Η ανάγκη για καταγραφή και επεξεργασία δεδομένων οδήγησε στην δημιουργία αρχικά στατιστικών υπηρεσιών κατάλληλων για την αποθήκευση και αρχειοθέτηση δεδομένων καθώς και την αποδοτικότερη επεξεργασία και συλλογή αποτελεσμάτων. Η διαδικασία της επεξεργασίας δεδομένων και συλλογής αποτελεσμάτων περιλαμβάνει ένα προστάδιο, το στάδιο της ανάλυσης των δεδομένων.

Με τον όρο αυτό εννοούνται όχι μόνο οι τεχνικές και οι μέθοδοι επεξεργασίας πληροφοριών που προέκυψαν από πραγματικά ή εικονικά πειράματα ή παρακολούθηση φαινομένων αλλά και η θεσμοθέτηση κοινά παραδεκτών τεχνικών με σκοπό την εκτίμηση των χαρακτηριστικών του πληθυσμού.

Παράλληλα η χρήση των υπαρχόντων στατιστικών – μαθηματικών εργαλείων ανάλυσης δεδομένων οδήγησε στην αποδοχή κοινών μεθοδολογιών για την αποδοτικότερη επεξεργασία δεδομένων. Ο διαχωρισμός των υπάρχων μεθοδολογιών περιλαμβάνει:

- Περιγραφική στατιστική,
- Στατιστική συμπερασματολογία – Επαγωγική στατιστική,
- Ανάλυση παλινδρόμησης και διακύμανσης,
- Στοχαστική ανάλυση,
- Μπεϋζιανή ανάλυση,
- Πολυμεταβλητή ανάλυση – Ανάλυση κατηγορικών δεδομένων,
- Μη- παραμετρική στατιστική.

2.1. ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Τα αποτελέσματα ενός πειράματος συνήθως δημιουργούν ένα μεγάλο αριθμό δεδομένων. Κρίνεται λοιπόν αναγκαία η εύρεση διαδικασιών, με τις οποίες τα αποτελέσματα αυτά μπορούν να οργανωθούν και να παρουσιαστούν με απλό και εύληπτο τρόπο. Αυτός ακριβώς είναι και ο στόχος των περιγραφικών στατιστικών δεικτών, δηλαδή να παρέχουν μεθόδους που απλοποιούν και διευκολύνουν την οργάνωση και παρουσίαση των αποτελεσμάτων. Στην ενότητα αυτή παρουσιάζονται σύντομα ορισμένοι από τους πιο γνωστούς περιγραφικούς στατιστικούς δείκτες.

Οι κατανομές συχνότητας (frequency distributions) αποτελούν μία από τις πιο γνωστές μεθόδους για την παρουσίαση ενός συνόλου δεδομένων. Οι κατανομές συχνότητας μπορούν να δομηθούν είτε σαν πίνακες είτε σαν γραφικές παραστάσεις. Και στις δύο περιπτώσεις, ο στόχος είναι να δοθεί μία πλήρης εικόνα για την κατανομή των δεδομένων στην κλίμακα μέτρησης.

Ανάλογα με την κατηγορία των δεδομένων (ποσοτικά ή ποιοτικά), χρησιμοποιούνται διαφορετικές μορφές αυτών των κατανομών (ιστογράμματα ή ραβδογράμματα, αντίστοιχα).

Ο σκοπός των μέτρων κεντρικής τάσης (measures of central tendency) είναι να προσδιοριστεί ένα στατιστικό μέγεθος το οποίο μπορεί να χρησιμοποιηθεί για να αντιπροσωπεύσει ένα σύνολο δεδομένων. Για την επίτευξη αυτού του στόχου χρησιμοποιούνται συνήθως τρία μέτρα κεντρικής τάσης: η μέση τιμή (mean), η διάμεσος (median) και η επικρατούσα τιμή (mode). Τα μέτρα κεντρικής τάσης επιτρέπουν τη συνοπτική περιγραφή και τη σύγκριση των διαθέσιμων δεδομένων.

Στις περισσότερες περιπτώσεις χρησιμοποιείται η μέση τιμή σαν μέτρο κεντρικής τάσης, καθώς δίνει την πιο αντιπροσωπευτική εικόνα για ένα σύνολο δεδομένων. Ωστόσο, σε ορισμένες περιπτώσεις ο υπολογισμός της μέσης τιμής δεν είναι δυνατός (π.χ. υπάρχουν απροσδιόριστες τιμές στα δεδομένα, ή έχουμε διαθέσιμα ποιοτικά δεδομένα), ή δεν παρέχει την πιο αντιπροσωπευτική εικόνα (π.χ. η μέση τιμή μπορεί να επηρεαστεί σημαντικά από μία πολύ μεγάλη ή πολύ μικρή παρατήρηση). Σε αυτές τις περιπτώσεις χρησιμοποιείται σαν μέτρο κεντρικής τάσης η διάμεσος, που αντιστοιχεί σε εκείνη την τιμή για την οποία το 50% των παρατηρήσεων έχει τιμή ίση ή μικρότερη.

Τέλος, σαν μέτρο κεντρικής τάσης μπορεί να χρησιμοποιηθεί και η επικρατούσα τιμή. Η επικρατούσα τιμή ουσιαστικά αντιστοιχεί στην παρατήρηση με τη μεγαλύτερη συχνότητα, και χρησιμοποιείται συνήθως συμπληρωματικά με τα δύο προηγούμενα μέτρα κεντρικής τάσης, ή όταν είναι διαθέσιμα ποιοτικά δεδομένα με τη μορφή ονομαστικών κατηγοριών, όπου δεν έχει νόημα ο υπολογισμός της μέσης τιμής ή της διαμέσου.

Τα μέτρα διασποράς (measures of variability) δίνουν μία εικόνα σχετικά με το πόσο συγκεντρωμένες είναι οι παρατηρήσεις σε ένα σύνολο δεδομένων. Τα κυριότερα μέτρα διασποράς είναι το εύρος (range), το ήμι-ενδοτεταρτομοριακό εύρος (semi-interquartile range), η τυπική απόκλιση (standard deviation) και η διακύμανση (variance).

Η διακύμανση και η τυπική απόκλιση αποτελούν το πιο αξιόπιστο και το πιο συνηθισμένο μέτρο διασποράς. Αυτά τα μέτρα διασποράς χρησιμοποιούν ως σημείο αναφοράς τη μέση τιμή, και λαμβάνουν υπόψη την απόσταση όλων των παρατηρήσεων από αυτήν.

Συμπερασματικά μπορούμε να αναφέρουμε πως σκοπός της Περιγραφικής Στατιστικής είναι γενικά η άθροιση και σύνοψη δεδομένων. Ειδικότερα αποτελεί ένα στατιστικό εργαλείο με σκοπό την συγκέντρωση ταξινόμηση και παρουσίαση πρωτογενών δεδομένων σε κατανοητή μορφή. Γίνεται με την χρήση πινάκων (συχνοτήτων, διπλής εισόδου), γραφημάτων (ραβδογράμματα, θηκογράμματα, διασποράς), και στατιστικών μέτρων (μέτρα κεντρικής τάσης, μέτρα κύμανσης, και μεταβλητότητας).

2.1.1. ΠΑΡΑΔΕΙΓΜΑΤΑ ΠΕΡΙΓΡΑΦΙΚΗΣ ΣΤΑΤΙΣΤΙΚΗΣ

Σε αυτήν την ενότητα περιγράφονται ορισμένα τυπικά σενάρια μελετών και προβληματισμών που μπορεί να συναντήσει ένας ερευνητής ο οποίος μελετά την επίδραση της χρήσης συνεργατικών τεχνολογιών σε εργασιακά και μαθησιακά περιβάλλοντα. Τα σενάρια αυτά θα μας βοηθήσουν να κατανοήσουμε το ρόλο της στατιστικής στην επιστημονική έρευνα.

- **Σενάριο 1^ο:** ένας καθηγητής Πανεπιστημίου διδάσκει Μαθηματικά για 20 χρόνια. Χρησιμοποιώντας παλαιότερες βαθμολογίες φοιτητών έχει υπολογίσει ότι η βαθμολογία τους ακολουθεί κανονική κατανομή με μέση τιμή 6 και διακύμανση 22. Ο καθηγητής θέλει τώρα να εξετάσει αν η χρήση συστημάτων ασύγχρονης τηλεκπαίδευσης θα επηρεάσει θετικά τη βαθμολογία των φετινών φοιτητών του. Ο καθηγητής χρησιμοποιεί τυχαίο δείγμα 15 μαθητών
- **Σενάριο 2^ο:** το σενάριο αυτό αποτελεί μία παραλλαγή του πρώτου. Τώρα ο καθηγητής θέλει να εξετάσει την επίδραση της χρήσης ασύγχρονης τηλεκπαίδευσης στο μάθημα των Μαθηματικών για όλα τα τμήματα του Πανεπιστημίου. Επειδή ο πληθυσμός τώρα είναι πολύ μεγαλύτερος επιλέγει ένα δείγμα 100 φοιτητών. Από τις βαθμολογίες όλων των καθηγητών που παρακολουθούν Μαθηματικά γνωρίζει μόνο ότι ο μέσος όρος της βαθμολογίας των φοιτητών είναι 6.
- **Σενάριο 3^ο:** ένας ερευνητής θέλει να εξετάσει εάν μια νέα διδακτική μέθοδος για μαθητές Γυμνασίου, η οποία χρησιμοποιεί συνεργατική τεχνολογία, επιδρά στη μαθησιακή διαδικασία. Ο ερευνητής επιλέγει τυχαία δύο δείγματα 30 μαθητών και εφαρμόζει στο πρώτο δείγμα την κλασική μέθοδο διδασκαλίας, ενώ στο δεύτερο δείγμα

τη νέα μέθοδο διδασκαλίας. Για να διαπιστώσει την επίδραση της αλλαγής διδακτικής μεθόδου (ανεξάρτητη μεταβλητή) στη μαθησιακή διαδικασία, υποβάλλει τους μαθητές κάθε δείγματος σε ένα βαθμολογημένο τεστ με άριστα το 100, και καταγράφει τις βαθμολογίες τους (εξαρτημένη μεταβλητή). Επιπρόσθετα, από τα δεδομένα που έχει συλλέξει υπολογίζει ότι η διακύμανση του πρώτου δείγματος είναι 36,3, ενώ του δεύτερου είναι 41,8.

- **Σενάριο 4^ο:** το σενάριο αυτό αποτελεί μία παραλλαγή του προηγούμενου. Τώρα, επειδή ο ερευνητής πιστεύει ότι οι νοητικές ικανότητες κάθε μαθητή μπορεί να αλλοιώσουν τα αποτελέσματα, δεν επιλέγει με τελείως τυχαίο τρόπο τα δύο δείγματα των 30 μαθητών, αλλά χρησιμοποιεί ένα τεστ IQ, και για κάθε μαθητή με ένα συγκεκριμένο δείκτη IQ που τοποθετεί στο πρώτο δείγμα, τοποθετεί και ένα ακόμη με τον ίδιο δείκτη IQ στο δεύτερο δείγμα, δηλαδή "ζευγαρώνει" τα υποκείμενα που ανήκουν στα δείγματα, χρησιμοποιώντας ως κριτήριο τον δείκτη IQ τους.
- **Σενάριο 5^ο:** ένας μάνατζερ θέλει να εξετάσει αν η χρήση χώρων σύσκεψης με υποστήριξη από υπολογιστικά συστήματα επιδρά στο χρόνο λήψης της τελικής απόφασης. Για το σκοπό αυτό επιλέγει ένα αντιπροσωπευτικό δείγμα 30 εργαζομένων, και τους τοποθετεί τη μία μέρα σε ένα "κλασικό" χώρο σύσκεψης, και την επόμενη μέρα σε ένα χώρο σύσκεψης που υποστηρίζεται από υπολογιστικά συστήματα, χρονομετρώντας κάθε φορά τη διαδικασία λήψης της τελικής απόφασης.
- **Σενάριο 6^ο:** ένας ερευνητής θέλει να μελετήσει την επίδραση τριών διαφορετικών μεθόδων διδασκαλίας Μαθηματικών στην ικανότητα επίλυσης προβλημάτων από φοιτητές. Η πρώτη μέθοδος διδασκαλίας είναι η "κλασική" μέθοδος διδασκαλίας, όπου ο καθηγητής διδάσκει ένα σύνολο φοιτητών. Στη δεύτερη μέθοδο οι φοιτητές αναλαμβάνουν με μία προκαθορισμένη σειρά να παρουσιάσουν την ύλη κάθε διάλεξης, και ο καθηγητής παρακολουθεί και επεμβαίνει όποτε κρίνει ότι είναι απαραίτητο. Στην τρίτη μέθοδο ο καθηγητής παρουσιάζει τη διάλεξη χρησιμοποιώντας υποστηρικτικά ασύρματες τεχνολογίες συνεργασίας. Ο ερευνητής επιλέγει τρία δείγματα 50 φοιτητών, και εφαρμόζει σε καθένα μία από τις τρεις διαφορετικές μεθόδους διδασκαλίας. Στο τέλος του εξαμήνου χρησιμοποιεί για κάθε δείγμα ένα τεστ μαθηματικών και συγκεντρώνει τις βαθμολογίες όλων των φοιτητών. Με τα δεδομένα αυτά υπολογίζει ότι οι διακυμάνσεις των τριών δειγμάτων είναι 30, 45 και 40 αντίστοιχα.

- **Σενάριο 7^ο:** ένας ερευνητής θέλει να κάνει μία έρευνα για να διαπιστώσει ποια από τις τέσσερις κατηγορίες συνεργατικής αλληλεπίδρασης (ασύγχρονη - από απόσταση, σύγχρονη - από απόσταση, σύγχρονη - τοπική, ασύγχρονη - τοπική) είναι η πιο δημοφιλής. Για το σκοπό αυτό επιλέγει ένα τυχαίο δείγμα 1000 ατόμων και χρησιμοποιεί ένα ερωτηματολόγιο στο οποίο τα υποκείμενα δηλώνουν την προτίμησή τους.

2.2. ΕΠΑΓΩΓΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Συνήθως όταν ένας ερευνητής σχεδιάζει κάποιο πείραμα έχει κάποια θεωρία που θέλει να επαληθεύσει, και επομένως κάποια ένδειξη για τα τελικά αποτελέσματα. Αυτή η πρόβλεψη του ερευνητή για το τελικό αποτέλεσμα ονομάζεται στην στατιστική υπόθεση (hypothesis). Ουσιαστικά μία υπόθεση αποτελεί την πρόβλεψη του ερευνητή για την επίδραση της αλλαγής της ανεξάρτητης μεταβλητής στην εξαρτημένη. Ο έλεγχος υποθέσεων (hypothesis testing) αποτελεί ένα σημαντικό κομμάτι της επαγωγικής στατιστικής και ένα πολύ χρήσιμο εργαλείο για τον ερευνητή, καθώς του επιτρέπει να ελέγξει την εγκυρότητα της θεωρίας του.

Σκοπός της Στατιστικής Συμπερασματολογίας είναι η διεξαγωγή από τα δεδομένα νόμων, κανόνων και συμπερασμάτων των οποίων η ισχύς ξεπερνά το επίπεδο των παρατηρήσεων. Οι προτεινόμενοι κανόνες καθορίζουν ένα μαθηματικό μοντέλο με σκοπό την καλύτερη και απλούστερη ερμηνεία των δεδομένων.

Η Επαγωγική Στατιστική είναι εκείνος ο τομέας της εφαρμοσμένης στατιστικής, που ασχολείται με τη γενίκευση των συμπερασμάτων της περιγραφικής στατιστικής, που προέκυψαν από τμήμα του συνόλου που ερευνάται, σε ολόκληρο το σύνολο αυτό. Η επαγωγική στατιστική με τη σειρά της, μπορεί να διακριθεί στην εκτιμητική στατιστική και στον έλεγχο υποθέσεων.

Η εκτιμητική στατιστική αντιμετωπίζει το πρόβλημα πως από τα δεδομένα ενός δείγματος είναι δυνατόν να υπολογιστεί (εκτιμηθεί) η τιμή μιας ή περισσότερων αγνώστων παραμέτρων του πληθυσμού και μπορεί να είναι σημείου και διαστήματος. Στις σημειακές εκτιμήσεις, χρησιμοποιούνται οι κατάλληλες στατιστικές συναρτήσεις και ένα δείγμα από τον υπό διερεύνηση πληθυσμό, προκειμένου να πραγματοποιηθεί η εκτίμηση μιας άγνωστης

πληθυσμιακής παραμέτρου και για κάθε παράμετρο παρέχεται μία μόνο τιμή. Το μειονέκτημα των σημειακών εκτιμήσεων είναι ότι διαφορετικά δείγματα, δίνουν διαφορετικές τιμές για την αντίστοιχη πληθυσμιακή παράμετρο. Αντίθετα, στις εκτιμήσεις διαστήματος χρησιμοποιείται ένα δείγμα από τον πληθυσμό, προκειμένου να εκτιμηθεί ένα διάστημα πιθανών τιμών, εντός του οποίου μπορεί να βρίσκεται η άγνωστη πληθυσμιακή παράμετρος.

Κατά τον έλεγχο υποθέσεων όπως και στην εκτιμητική επαγωγική στατιστική, γίνεται προσπάθεια εκτίμησης άγνωστων πληθυσμιακών παραμέτρων μέσα από συγκεκριμένα δείγματα και με τη χρήση κατάλληλων στατιστικών συναρτήσεων, πραγματοποιώντας όμως τον έλεγχο συγκεκριμένων υποθέσεων, που αφορούν τις υπό διερεύνηση πληθυσμιακές παραμέτρους. Αναλυτικότερα, αντιμετωπίζεται το πρόβλημα πως από τα δεδομένα ενός δείγματος, μπορούμε να συμπεράνουμε αν γίνεται αποδεκτή ή ορθότερα αν πρέπει να απορριφθεί ή όχι μια υπόθεση, που διατυπώνεται για την τιμή κάποιας πληθυσμιακής παραμέτρου.

2.2.1 Ο ΣΤΑΤΙΣΤΙΚΟΣ ΕΛΕΓΧΟΣ ΩΣ ΕΠΑΓΩΓΙΚΟ ΜΕΤΡΟ

Υπάρχουν πολλές παραλλαγές της μεθόδου ελέγχου υποθέσεων για κάθε ερευνητική περίπτωση, η βασική λογική που ακολουθείται είναι πάντα η ίδια. Αξίζει να σημειωθεί ότι συνήθως ένας ερευνητής μελετά πληθυσμούς οι οποίοι προσεγγίζουν μία κανονική κατανομή (normal distribution), δηλαδή οι περισσότερες παρατηρήσεις συγκεντρώνονται γύρω από τη μέση τιμή, και ο αριθμός τους μειώνεται συμμετρικά και προς τις δύο μεριές.

Ο έλεγχος υποθέσεων έχει ως αποτέλεσμα μια απόφαση για την τιμή μιας παραμέτρου ενός στατιστικού πληθυσμού (μέσος, διακύμανση, αναλογία). Άλλοι έλεγχοι έχουν ως σκοπό την σύγκριση πληθυσμών ή τον βαθμό τυχαιότητας των δειγμάτων ή το είδος συσχέτισης δυο μεταβλητών.

Υπόθεση είναι μια δήλωση για έναν ή περισσότερους πληθυσμό (ους) ή τις παραμέτρους τους. Σε ένα έλεγχο υποθέσεων έχουμε την αντικαταβολή δυο υποθέσεων, της μηδενικής και της εναλλακτικής. Η μηδενική είναι η υπόθεση που ελέγχεται για την ορθότητα της. Η εναλλακτική είναι η υπόθεση η οποία είναι στην διάθεση μας όταν απορρίπτεται η μηδενική υπόθεση.

Στους στατιστικούς ελέγχους υποθέσεων:

Επιθυμούμε να ελέγξουμε αν μία ή περισσότερες παράμετροι (π.χ. μέση τιμή μ) ενός πληθυσμού ικανοποιούν μια βασική υπόθεση (π.χ. $\mu = 100$) έναντι μιας εναλλακτικής υπόθεσης (π.χ. $\mu > 100$).

Σχεδόν πάντοτε δεν είμαστε σε θέση να καταγράψουμε όλον τον πληθυσμό (ώστε να αποφανθούμε με βεβαιότητα για το αν ισχύει η βασική υπόθεση) οπότε αρκούμαστε σε ένα τυχαίο δείγμα από αυτόν.

Με βάση αυτό το τυχαίο δείγμα θέλουμε να πάρουμε μια απόφαση: να απορρίψουμε ή όχι ότι ισχύει η βασική υπόθεση.

Συνήθως εργαζόμαστε ως εξής:

Κατασκευάζουμε κάποια συνάρτηση του δείγματος (στατιστική συνάρτηση) η οποία, όταν ισχύει η βασική υπόθεση, ακολουθεί μια συγκεκριμένη (γνωστή) κατανομή. Ενώ όταν ισχύει η εναλλακτική υπόθεση, λαμβάνει «ακραίες» τιμές (πολύ «μεγάλες» ή πολύ «μικρές»).

Αν, με βάση το τ.δ. που πήραμε, αυτή η στατιστική συνάρτηση λάβει κάποια «ακραία» τιμή τότε απορρίπτουμε την βασική υπόθεση. Το πότε μια τιμή θεωρείται «ακραία» ώστε να απορρίψουμε την βασική υπόθεση εξαρτάται από το «επίπεδο σημαντικότητας» που έχουμε προαποφασίσει.

Αφού το ο ερευνητής επιλέξει και προσδιορίσει πλήρως και με σαφήνεια το πρόβλημα προς διερεύνηση, ένα σημαντικό θέμα που ακολουθεί είναι η διατύπωση της υπόθεσης που πρόκειται να ελεγχθεί. Η λέξη υπόθεση στην έρευνα έχει περιορισμένη έννοια και αναφέρεται σε εικασίες που πιθανόν επεξηγούν παρατηρήσεις. Οι ερευνητικές υποθέσεις δεν αποδεικνύονται με μαρτυρίες ή γεγονότα που είναι σύμφωνα με αυτές, απλώς δεν διαψεύδονται. Στην αντίθετη περίπτωση όμως, όταν δηλαδή οι μαρτυρίες ή τα γεγονότα δεν είναι σύμφωνα με τη υπόθεση, τότε η υπόθεση απορρίπτεται συμπερασματικά. Αυτή είναι και η διαφορά των ερευνητικών υποθέσεων με τους καθημερινούς συλλογισμούς.

Η διατύπωση μιας ερευνητικής υπόθεσης προκύπτει λογικά από την βιβλιογραφική ανασκόπηση του προβλήματος. Επίσης μπορεί να προκύψει και από τη θεωρία, αφού η θεωρία καθοδηγεί την

έρευνα και δημιουργεί προβλέψεις που πρέπει να ελεγχθούν. Το πιο σημαντικό στοιχείο ωστόσο, δεν είναι από πού προκύπτει η υπόθεση αλλά ότι οι υποθέσεις πρέπει να εκφράζονται με τέτοιο τρόπο ώστε να είναι δυνατή είτε η επιβεβαίωσή τους είτε η απόρριψή τους. Η υπόθεση που δεν μπορεί να ελεγχθεί, δεν ανήκει στη σφαίρα της επιστήμης.

Πολλοί ερευνητές υποστηρίζουν ότι μπορεί να υπάρξει κάποιου είδους ιεραρχία στις υποθέσεις που μπορεί να προκύψουν σε μια ερευνητική μελέτη. Κάθε υπόθεση που ακολουθεί, θα είναι πιο συμπαγής στις έννοιες και με μεγαλύτερη δυνατότητα ελέγχου από την προγενέστερή της. Συχνά, για την αρχική υπόθεση χρησιμοποιείται ο όρος Ερευνητική, ενώ για εκείνη που προκύπτει στο τέλος, (τη σαφέστερα διατυπωμένη), χρησιμοποιείται ο όρος Λειτουργική.

Οι όροι που χρησιμοποιούνται σε κάθε ερευνητική υπόθεση πρέπει να εκφράζονται με ακρίβεια, με σωστή ορολογία, να είναι λειτουργικές και να έχουν ξεκάθαρο νόημα. Γενικοί όροι του τύπου προσωπικότητα, αυτοεκτίμηση, επίτευγμα, πρέπει να αποφεύγονται. Η υπόθεση απαιτεί, λιτή και περιεκτική γλώσσα, καθώς και ορισμό της ορολογίας.

Επίσης, όπως αναφέρθηκε, την κάθε υπόθεση πρέπει να τη διακρίνει η δυνατότητα για έλεγχο. Αφού οι υποθέσεις είναι εκείνες που προλέγουν το αποτέλεσμα μιας έρευνας, είναι προφανές ότι συγκεκριμένα εργαλεία πρέπει να υπάρχουν που να μπορούν να παρέχουν αξιόπιστες και έγκυρες μετρήσεις των μεταβλητών που εμπλέκονται.

Οι υποθέσεις θα πρέπει να εκφράζουν διαφορές ή σχέσεις μεταξύ των μεταβλητών αυτών. Μια ικανοποιητική υπόθεση είναι εκείνη μέσω της οποίας οι αναμενόμενες συσχετίσεις μεταξύ των μεταβλητών γίνονται αναμφίβολες. Η υπόθεση πρέπει να είναι λιτή, συγκεκριμένη και σχετικά απλή στον έλεγχο. Οι γενικές υποθέσεις δεν προτιμούνται. Οι υποθέσεις δε θα πρέπει να είναι ασύμβατες με γεγονότα που είναι γνωστά και θα πρέπει να στηρίζονται στην υπάρχουσα γνώση. Φυσικά υπάρχουν και οι μελέτες που δίνουν αντίθετα αποτελέσματα από εκείνα που υποστηρίζουν τα γνωστά γεγονότα. Σε περιπτώσεις σαν κι αυτές η υπόθεση πρέπει να διατυπώνεται προκειμένου να επιλυθεί η διαφωνία.

Συνήθως, οι τρόποι διατύπωσης υπόθεσης που υπάρχουν, οι οποίοι ταυτίζονται με τις κύριες προσεγγίσεις του ερευνητικού σχεδιασμού είναι:

- Η διερεύνηση της ισότητας παραμέτρου με τιμή,

- Η διερεύνηση διαφορών ανάμεσα σε ομάδες και
- Η διερεύνηση σχέσεων ανάμεσα σε ομάδες.

Όταν λέμε ερευνητικός σχεδιασμός εννοούμε το πλάνο ή τη στρατηγική που ακολουθούμε προκειμένου να πάρουμε απαντήσεις σε ερευνητικές ερωτήσεις.

Από τους τρόπους διατύπωσης μιας υπόθεσης οδηγούμαστε στο σημαντικό σημείο ότι στη στατιστική ανάλυση, στην πραγματικότητα, δεν ελέγχουμε την λειτουργική υπόθεση, αλλά τη λογική της αντίθετη, ή αλλιώς μηδενική υπόθεση.

Αναλυτικότερα, δεν ελέγχουμε την υπόθεση ότι υπάρχει στατιστικά σημαντική πληθυσμιακή διαφορά ή συσχέτιση, αντίθετα ελέγχουμε την υπόθεση ότι δεν υπάρχει στατιστικά σημαντική διαφορά ή συσχέτιση. Όταν βέβαια έχουμε να κάνουμε τέτοιου είδους ελέγχους, συνήθως εργαζόμαστε με δειγματικά δεδομένα, που έχουν προκύψει από τον υπό μελέτη πληθυσμό, τα οποία χρησιμοποιούμε για να εκτιμήσουμε τις πληθυσμιακές τιμές των μεταβλητών μας. Οποιαδήποτε εκτίμηση μιας πληθυσμιακής παραμέτρου εμπεριέχει δειγματοληπτικό σφάλμα. Έτσι, αυτό το οποίο στην ουσία γίνεται με τη διατύπωση της μηδενικής υπόθεσης, είναι να αποδεικνύεται αν υπάρχει στατιστικά ασήμαντη διαφορά (ή συσχέτιση), η οποία προέρχεται από το γεγονός ότι διερευνάται μόνο τμήμα του πληθυσμού και για το λόγο αυτό μπορεί να αποδοθεί στις αυξομειώσεις-κυμάνσεις της τυχαίας δειγματοληψίας με τη μορφή δειγματοληπτικού σφάλματος, ή ότι είναι διαφορά στατιστικά σημαντική, δηλαδή πραγματική, που δεν μπορεί να εξηγηθεί μόνο από τις κυμάνσεις της τυχαίας δειγματοληψίας.

Με τον τόπο αυτό, είναι δυνατόν να προσδιοριστούν τα ακριβή όρια (επίπεδο σημαντικότητας) για την απόρριψη της μηδενικής υπόθεσης. Αναλυτικότερα, αν μια διαφορά (μεταξύ εκτίμησης και παραμέτρου) τόσο μεγάλη όσο αυτή που παρατηρείται ή και ακόμη μεγαλύτερη, μπορεί να συμβεί από τύχη (δηλαδή λόγω δειγματοληπτικού σφάλματος) λιγότερο από μία φορά στις εκατό, τότε λέμε ότι η διαφορά αυτή είναι στατιστικά σημαντική σε επίπεδο σημαντικότητας 0,01 ή 1%. Αν η διαφορά αυτή μπορεί να συμβεί τυχαία λιγότερες από πέντε φορές αλλά περισσότερες από μία στις εκατό, τότε λέμε ότι η διαφορά αυτή είναι στατιστικά σημαντική σε επίπεδο σημαντικότητας 0,05 ή 5%.

Στην περίπτωση που απορριφθεί η μηδενική υπόθεση, τότε γίνεται αποδεκτή η εναλλακτική υπόθεση, δηλαδή, ότι η διαφορά (ή η συσχέτιση) πράγματι υπάρχει. Στην πραγματικότητα αυτή

είναι και η υπόθεση για την οποία αρχικά ενδιαφερόμασταν, αλλά τώρα ονομάζεται Εναλλακτική Υπόθεση.

Τη μηδενική υπόθεση συνήθως τη συμβολίζουμε H_0 και την εναλλακτική της H_1 . Εδώ πρέπει να τονίσουμε ότι αν δεν απορριφθεί η μηδενική υπόθεση αυτό δε σημαίνει ότι δεν υπάρχει διαφορά ή συσχέτιση, αλλά ότι δεν υπάρχει διαφορά ή συσχέτιση μεγαλύτερη από εκείνη που οφείλεται στη τύχη ή σε τυχαίες αυξομειώσεις.

Ο τρόπος με τον οποίο μπορούμε να γνωρίζουμε ότι η συσχέτιση ή η διαφορά είναι αρκετά μεγάλες ώστε να ξεπερνούν εκείνες που οφείλονται στη τύχη, είναι η χρησιμοποίηση των επιπέδων σημαντικότητας (0,01, 0,05, 0,001) και πινάκων κατάλληλων στατιστικών δοκιμασιών.

Η λογική του ελέγχου υποθέσεων μπορεί να συνοψιστεί στα εξής τέσσερα βήματα: διατύπωση υποθέσεων, διαμόρφωση των κριτηρίων για την λήψη μιας απόφασης, συλλογή δεδομένων από δείγματα και αξιολόγηση της μηδενικής υπόθεσης.

Βήμα 1ο: Διατύπωση των Υποθέσεων

Το πρώτο βήμα σε έναν έλεγχο υποθέσεων είναι να διατυπωθεί η μηδενική υπόθεση H_0 (null hypothesis). Η μηδενική υπόθεση είναι μία πρόταση που προβλέπει ότι η αλλαγή της ανεξάρτητης μεταβλητής (μεταβλητή που ελέγχω) δεν έχει καμία επίδραση στην εξαρτημένη μεταβλητή (μεταβλητή που παρατηρώ). Με αυτόν τον τρόπο μπορούμε να θεωρήσουμε γνωστές τις παραμέτρους του νέου πληθυσμού που δημιουργείται από την αλλαγή της τιμής της ανεξάρτητης μεταβλητής, καθώς υποθέτουμε ότι αυτή η αλλαγή δεν έχει καμία επίδραση. Επομένως, οι τιμές των παραμέτρων του πληθυσμού μετά την αλλαγή της ανεξάρτητης μεταβλητής είναι ίδιες με αυτές πριν την αλλαγή.

Η δεύτερη υπόθεση που διατυπώνει ο ερευνητής ονομάζεται εναλλακτική υπόθεση H_1 (alternative hypothesis), και είναι η ακριβώς αντίθετη από τη μηδενική υπόθεση, δηλαδή δηλώνει ότι η αλλαγή της ανεξάρτητης μεταβλητής επιδρά στην εξαρτημένη μεταβλητή. Ο στόχος της μεθόδου ελέγχου υποθέσεων είναι να αποδείξει ότι η μηδενική υπόθεση μπορεί να απορριφθεί με κάποιο βαθμό βεβαιότητας.

Βήμα 2ο: Διαμόρφωση των Κριτηρίων για τη Λήψη μιας Απόφασης

Ο ερευνητής τελικά θα συλλέξει κάποια δεδομένα από δείγματα για να αξιολογήσει την αξιοπιστία της μηδενικής υπόθεσης. Με άλλα λόγια, θα χρησιμοποιήσει στατιστικούς δείκτες για να βγάλει συμπεράσματα για την τιμή μιας παραμέτρου του πληθυσμού. Ωστόσο, όπως προαναφέρθηκε, όσο αντιπροσωπευτικό και αν είναι το δείγμα, πάντα θα υπάρχει μια ασυμφωνία (σφάλμα δειγματοληψίας) ανάμεσα στην πραγματική τιμή της παραμέτρου και του στατιστικού δείκτη.

Το ερώτημα λοιπόν που προκύπτει είναι αν η διαφορά ανάμεσα στην τιμή του στατιστικού δείκτη και στην τιμή που προβλέπει η μηδενική υπόθεση για την παράμετρο του πληθυσμού οφείλεται πραγματικά στην επίδραση της ανεξάρτητης μεταβλητής, ή οφείλεται απλά σε σφάλματα δειγματοληψίας. Για να απαντήσει σε αυτό το ερώτημα ο ερευνητής χρειάζεται να θέσει κάποια κριτήρια που καθορίζουν επακριβώς πόση διαφορά χρειάζεται να υφίσταται ανάμεσα σε αυτές τις δύο τιμές, ώστε να δικαιολογείται η απόφαση να απορριφθεί η μηδενική υπόθεση. Αν τελικά αποφασίσει να απορρίψει τη μηδενική υπόθεση, θα πρέπει να καθορίσει το βαθμό βεβαιότητας με τον οποίο προβαίνει σε μία τέτοια απόφαση, ή με άλλα λόγια την πιθανότητα τα αποτελέσματα της ανάλυσής του να είναι εσφαλμένα. Η πιθανότητα αυτή ονομάζεται επίπεδο σημαντικότητας ή *επίπεδο α (level of significance or alpha level)*, και καθορίζει τη μέγιστη πιθανότητα το αποτέλεσμα μιας στατιστικής ανάλυσης να οφείλεται σε σφάλματα ή τυχαίους παράγοντες.

Βήμα 3ο: Συλλογή Δεδομένων από Δείγματα

Το επόμενο βήμα στη μέθοδο ελέγχου υποθέσεων είναι η συλλογή των δεδομένων από το δείγμα και ο υπολογισμός των κατάλληλων περιγραφικών στατιστικών δεικτών. Είναι εξαιρετικά σημαντικό να επιλεγεί ένα αντιπροσωπευτικό δείγμα, ώστε η μέθοδος να παρέχει αξιόπιστα αποτελέσματα. Όπως ήδη αναφέρθηκε, η μέθοδος της τυχαίας δειγματοληψίας εξασφαλίζει αυτή την απαίτηση. Σε αυτό το σημείο αξίζει να σημειωθεί ότι η διαδικασία της συλλογής των δεδομένων από το δείγμα γίνεται αφού ο ερευνητής διαμορφώσει τα κριτήρια για την λήψη μιας απόφασης, ώστε να είναι αμερόληπτος και ανεπηρέαστος από τα δεδομένα αυτά.

Βήμα 4ο: Αξιολόγηση της Μηδενικής Υπόθεσης

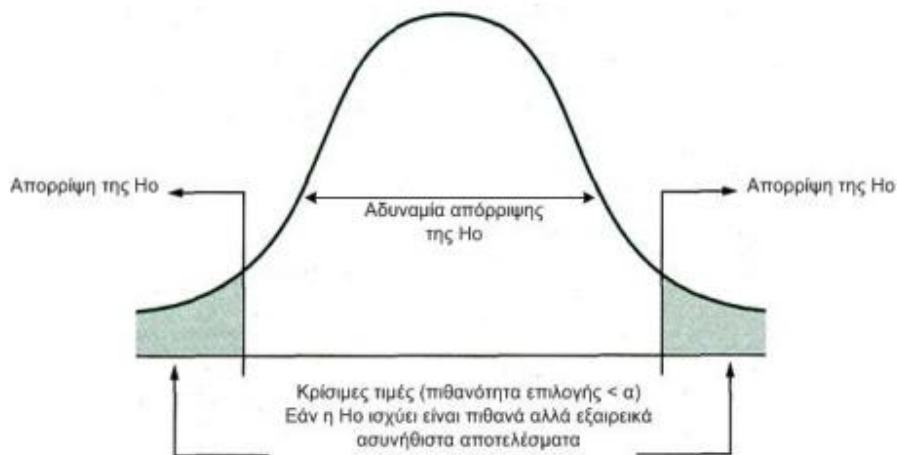
Στο τελευταίο βήμα της μεθόδου ελέγχου υποθέσεων ο ερευνητής μελετά την εγκυρότητα της μηδενικής υπόθεσης με τη βοήθεια στατιστικών δεικτών, ώστε να λάβει μια απόφαση σύμφωνα

με τα κριτήρια που έχει θέσει στο δεύτερο βήμα. Υπάρχουν λοιπόν δύο δυνατότητες, είτε αποφασίζει να απορρίψει τη μηδενική υπόθεση (reject the null hypothesis) όταν τα αποτελέσματα που προκύπτουν από το δείγμα είναι σημαντικά (significantly) διαφορετικά από αυτά που προβλέπει η μηδενική υπόθεση, είτε οδηγείται στην απόφαση ότι απέτυχε να απορρίψει τη μηδενική υπόθεση (fail to reject the null hypothesis) όταν τα αποτελέσματα του πειράματος δεν παρέχουν επαρκείς αποδείξεις ότι η μηδενική υπόθεση δεν ισχύει.

Ο ερευνητής για να οδηγηθεί σε μία από τις δύο παραπάνω αποφάσεις χρησιμοποιεί κάποιο στατιστικό δείκτη ελέγχου (test statistic).

Για παράδειγμα, αν ο ερευνητής έχει καθορίσει επίπεδο σημαντικότητας $\alpha=0.05$, αυτό σημαίνει ότι θέλει να είναι σίγουρος κατά 95% ότι τα συμπεράσματα του δεν είναι εσφαλμένα. Αν και η ακριβής τιμή κάθε στατιστικού δείκτη ελέγχου διαφέρει, σε γενικές γραμμές για $\alpha=0.05$ θέλουμε τα αποτελέσματα μας να εμφανίζονται με διπλάσια πιθανότητα από την πιθανότητα να οφείλονται απλά στην τύχη ή σε σφάλματα (για $\alpha=0.001$ η πιθανότητα γίνεται περίπου τριπλάσια). Στις επιστημονικές μελέτες η μεγαλύτερη αποδεκτή τιμή για το α είναι 0.05, δηλαδή η μέγιστη αποδεκτή πιθανότητα τα αποτελέσματα του ερευνητή να είναι εσφαλμένα είναι 5%.

Είναι φανερό ότι οι περιοχές στις οποίες μπορεί να απορριφθεί η μηδενική υπόθεση (ονομάζονται και κρίσιμες περιοχές) αντιστοιχούν σε μεγάλες τιμές του στατιστικού δείκτη ελέγχου, οι οποίες εξασφαλίζουν ότι τα αποτελέσματα που έλαβε ο ερευνητής είναι αρκετά ασυνήθιστα (έχουν μεγάλη διαφορά από τη μέση τιμή του αρχικού πληθυσμού) για το επίπεδο σημαντικότητας που έχει θέσει. Για την εύρεση της τιμής που ορίζει τις κρίσιμες περιοχές της κατανομής, ο ερευνητής χρειάζεται να ανατρέξει στον κατάλληλο στατιστικό πίνακα σύμφωνα με το στατιστικό δείκτη ελέγχου που χρησιμοποιεί.



Οι έλεγχοι υποθέσεων στηρίζονται σε στατιστικές πρακτικές. Είναι επομένως δυνατό να γίνει δεκτή (ή να απορριφτεί) μια υπόθεση που στην πραγματικότητα είναι λανθασμένη (ή σωστή). Για παράδειγμα, είναι δυνατό να απορριφθεί η μηδενική υπόθεση αν και είναι σωστή, γιατί έτυχε ο έλεγχος να στηριχθεί σε ένα μη αντιπροσωπευτικό δείγμα. Γενικά τα σφάλματα αποφάσεως που μπορούν να γίνουν είναι δύο τύπων.

- **Σφάλμα τύπου I (ή πρώτου είδους).** Αντιστοιχεί στην περίπτωση που απορρίπτεται η μηδενική υπόθεση, η οποία στην πραγματικότητα αληθεύει και έπρεπε να γίνει αποδεκτή.
- **Σφάλμα τύπου II (ή δευτέρου είδους).** Αντιστοιχεί στην περίπτωση που γίνεται αποδεκτή η μηδενική υπόθεση, η οποία στην πραγματικότητα είναι λανθασμένη.

Αν και τα δύο είδη σφαλμάτων μπορεί να φαίνονται το ίδιο σοβαρά, συνήθως τα σφάλματα τύπου I είναι σοβαρότερα, επειδή η αποδοχή της εναλλακτικής υπόθεσης συνήθως οδηγεί σε αποφάσεις αλλαγής της υπάρχουσας κατάστασης, οι οποίες αντιστοιχούν σε σημαντικό κόστος.

Στο σημείο αυτό αξίζει να αναφέρουμε ότι, θέτοντας μικρότερο επίπεδο σημαντικότητας, απαιτούμε πιο «σημαντικές αποδείξεις» για την απόρριψη της H_0 και τον χαρακτηρισμό των ευρημάτων μας στο δείγμα ως στατιστικά σημαντικών. Έτσι, μπορεί, σε κάποιο επίπεδο σημαντικότητας α , π.χ. $\alpha = 0.05$, να απορρίπτουμε την H_0 και σε κάποιο μικρότερο, π.χ. $\alpha = 0.01$, να μην την απορρίπτουμε γιατί απαιτούμε σημαντικότερες αποδείξεις.

Όσο πιο μικρό είναι το επίπεδο σημαντικότητας στο οποίο μπορούμε να απορρίψουμε την H_0 , τόσο πιο σημαντική είναι η τιμή της στατιστικής συνάρτησης ελέγχου που παρατηρείται στο δείγμα, με την έννοια ότι δίνει πιο ισχυρές αποδείξεις εναντίον της H_0 . Άρα, όσο πιο μικρό είναι το επίπεδο σημαντικότητας στο οποίο μπορούμε να απορρίψουμε την H_0 , τόσο πιο σημαντικό, στατιστικά, είναι το αποτέλεσμα του ελέγχου. Τέλος, είναι προφανές, ότι αν η H_0 απορρίπτεται σε κάποιο επίπεδο σημαντικότητας α , τότε επίσης απορρίπτεται σε οποιοδήποτε μεγαλύτερο, ενώ αν δεν απορρίπτεται σε κάποιο επίπεδο σημαντικότητας α , τότε επίσης δεν απορρίπτεται σε οποιοδήποτε μικρότερο.

2.2.2. Η ΧΡΗΣΙΜΟΤΗΤΑ ΤΟΥ ΣΤΑΤΙΣΤΙΚΟΥ ΕΛΕΓΧΟΥ

Οι έλεγχοι υποθέσεων μαζί με τα διαστήματα εμπιστοσύνης είναι τα δύο σημαντικότερα εργαλεία της στατιστικής συμπερασματολογίας, αποτελούν το συνδυαστικό κρίκο ανάμεσα στην περιγραφική και την επαγωγική στατιστική γενικεύοντας τα όσα προκύπτουν μελετώντας ένα δείγμα, με σκοπό να περιγράφουν επαρκώς και τα χαρακτηριστικά του αντίστοιχου πληθυσμού. Αντικείμενο του κλάδου αυτού της στατιστικής είναι η εξαγωγή συμπερασμάτων σχετικά με τα χαρακτηριστικά ενός πληθυσμού από πληροφορίες που αναφέρονται σε ένα μόνο δείγμα. Σύμφωνα με την πρακτική των διαστημάτων εμπιστοσύνης, υπολογίζεται μια σημειακή εκτίμηση μιας παραμέτρου του πληθυσμού (π.χ. μέση τιμή) και στη συνέχεια σχηματίζεται ένα διάστημα εμπιστοσύνης γύρω από την εκτίμηση αυτή, για το οποίο υπάρχει βεβαιότητα (εμπιστοσύνη) κατά ένα ποσοστό ότι βρίσκεται η ζητούμενη παράμετρος του πληθυσμού. Αυτή η ανάλυση δε λαμβάνει υπόψη οποιαδήποτε πεποίθηση υπάρχει σχετικά με τον πληθυσμό.

Στην άλλη πλευρά βρίσκονται οι έλεγχοι υποθέσεων (ή έλεγχοι σημαντικότητας ή κανόνες αποφάσεων). Η πρακτική αυτή λαμβάνει υπόψη την πεποίθηση που υπάρχει σχετικά με την παράμετρο ενός πληθυσμού, η οποία οδηγεί στην κατάσχεση μιας υπόθεσης. Σκοπός της ανάλυσης είναι η αποδοχή ή η απόρριψη της υπόθεσης χρησιμοποιώντας τις πληροφορίες που παρέχει το δείγμα του πληθυσμού. Μέσα από τον έλεγχο υπόθεσης μπορούν να προκύψουν συμπεράσματα ιδιαίτερα σημαντικά για τη διαδικασία της λήψης αποφάσεων.

Είναι φανερό ότι ο έλεγχος υποθέσεων αποτελεί ένα απαραίτητο εργαλείο για τον ερευνητή, καθώς του επιτρέπει να βγάλει γενικά συμπεράσματα για πληθυσμούς χρησιμοποιώντας πληροφορίες που προέρχονται από δείγματα. Η επαγωγική στατιστική περιλαμβάνει μία πληθώρα στατιστικών μεθόδων που βασίζονται στον έλεγχο υποθέσεων και χρησιμοποιούνται σε διαφορετικές ερευνητικές περιστάσεις. Σε κάθε ερευνητική μελέτη οι πληροφορίες που χρειάζεται να λάβουμε υπόψη για την επιλογή της κατάλληλης μεθόδου ελέγχου υποθέσεων είναι ο στόχος του πειράματος, το πλήθος των ανεξάρτητων μεταβλητών και των διαφορετικών τιμών τους, το πλήθος των εξαρτημένων μεταβλητών, ο αριθμός των διαθέσιμων δειγμάτων, η κατηγορία των δεδομένων, και τέλος το είδος του στατιστικού ελέγχου σε συνδυασμό με τις προϋποθέσεις που πρέπει να ικανοποιούνται για τη χρήση του.

2.2.3. ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ

Σε συνέχεια των παραπάνω να αναφέρουμε μια ακόμη περίπτωση ελέγχου υποθέσεων, όπου εξετάζεται η ύπαρξη επιδράσεων από μια ή περισσότερες εξωγενείς μεταβλητές στην εξαρτημένη μεταβλητή που είναι αυτή η οποία μας ενδιαφέρει κατά κύριο λόγο. Υπάρχουν και περιπτώσεις στις οποίες ο ερευνητής θέλει να εξετάσει την επίδραση δύο ή περισσότερων εξαρτημένων μεταβλητών με δύο ή περισσότερες δυνατές τιμές στην παρατηρούμενη μεταβλητή. Η ανάλυση διακύμανσης (ANalysis Of VAriance-ANOVA) εξυπηρετεί αυτόν ακριβώς το σκοπό.

Σε αυτό το σημείο μπορεί να αναρωτηθεί κανείς γιατί χρειαζόμαστε μία καινούρια στατιστική μέθοδο για να ελέγξουμε την επίδραση δύο ή περισσότερων τιμών της ανεξάρτητης μεταβλητής, αφού θα μπορούσαμε κάλλιστα να εξετάσουμε την επίδραση όλων αυτών των δυνατών τιμών ανά δύο, χρησιμοποιώντας όσες φορές χρειάζεται μία από τις στατιστικές μεθόδους που παρουσιάστηκαν προηγουμένως. Ωστόσο, εύκολα γίνεται αντιληπτό ότι η χρήση πολλαπλών στατιστικών τεστ αυξάνει τη συνολική πιθανότητα να οδηγηθούμε σε λανθασμένα

συμπεράσματα, και για αυτό χρησιμοποιείται η ανάλυση διακύμανσης σε αυτές τις περιπτώσεις⁶.

Η ανάλυση διακύμανσης μπορεί να χρησιμοποιηθεί είτε με ανεξάρτητα δείγματα είτε με εξαρτημένα δείγματα. Επίσης, η ανάλυση διακύμανσης μπορεί να χρησιμοποιηθεί και σε πειράματα που ο ερευνητής θέλει να ελέγχει δύο ή περισσότερες ανεξάρτητες μεταβλητές, οι οποίες στην ορολογία της ανάλυσης διακύμανσης ονομάζονται παράγοντες (factors). Επειδή αυτές οι ερευνητικές περιπτώσεις αυξάνουν σημαντικά το βαθμό πολυπλοκότητας της στατιστικής ανάλυσης και ξεφεύγουν από τους στόχους αυτού του κεφαλαίου, θα περιγράψουμε στη συνέχεια μόνο την ανάλυση διακύμανσης ενός παράγοντα (single-factor ANOVA). Ο αναγνώστης που ενδιαφέρεται και για τις υπόλοιπες μεθόδους μπορεί να ανατρέξει στη βιβλιογραφία που αναφέρθηκε στην εισαγωγή.

Η ανάλυση διακύμανσης ενός παράγοντα χρησιμοποιείται για τον έλεγχο υποθέσεων σε μελέτες ανεξάρτητων δειγμάτων. Ο ερευνητής επιλέγει ένα δείγμα για κάθε διαφορετική τιμή της ανεξάρτητης μεταβλητής, και ελέγχει υποθέσεις που συγκρίνουν τις μέσες τιμές των δειγμάτων αυτών. Η μηδενική υπόθεση δηλώνει ότι δεν υφίσταται καμία διαφορά ανάμεσα σε όλες τις μέσες τιμές των δειγμάτων. Είναι εμφανές λοιπόν, ότι στην περίπτωση που ο ερευνητής απορρίψει τελικά τη μηδενική υπόθεση, το μόνο που μπορεί να ισχυριστεί είναι ότι τα δείγματα διαφέρουν μεταξύ τους, αλλά δεν είναι σε θέση να γνωρίζει ποια συγκεκριμένα δείγματα διαφέρουν. Αν θέλει να εξακριβώσει τέτοιου είδους πληροφορίες μπορεί να χρησιμοποιήσει μετά την ανάλυση διακύμανσης ειδικά στατιστικά τεστ, τα οποία ονομάζονται post hoc tests.

Η ανάλυση διακύμανσης χρησιμοποιεί την κατανομή F για την εύρεση των κρίσιμων τιμών.

Οι προϋποθέσεις που πρέπει να ισχύουν για να χρησιμοποιηθεί η ανάλυση διακύμανσης ενός παράγοντα συνοψίζονται στις εξής:

- τα δείγματα είναι αντιπροσωπευτικά και οι τιμές που τα απαρτίζουν οφείλονται σε ανεξάρτητες παρατηρήσεις,

⁶ Σε κάθε τεστ υπάρχει μία πιθανότητα τα συμπεράσματα να είναι εσφαλμένα, η οποία εκφράζεται από το συντελεστή σημαντικότητας α . Αν χρησιμοποιήσουμε διαδοχικά τεστ, τότε ο συνολικός συντελεστής σημαντικότητας ισούται με το άθροισμα των επιμέρους συντελεστών του κάθε τεστ, και επομένως αυξάνεται η συνολική πιθανότητα να οδηγηθούμε σε εσφαλμένα συμπεράσματα

Η ανάλυση διακύμανσης χρησιμοποιεί την κατανομή F για την εύρεση των κρίσιμων τιμών.

Οι προϋποθέσεις που πρέπει να ισχύουν για να χρησιμοποιηθεί η ανάλυση διακύμανσης ενός παράγοντα συνοψίζονται στις εξής:

τα δείγματα είναι αντιπροσωπευτικά και οι τιμές που τα απαρτίζουν οφείλονται σε ανεξάρτητες παρατηρήσεις,

2.2.4. ΤΟ Χ² ΣΤΑΤΙΣΤΙΚΟ ΤΕΣΤ (Chi –Square Test)

Το στατιστικό τεστ X^2 (Chi-Square test) είναι ίσως το πιο δημοφιλές μη-παραμετρικό τεστ. Αν και υπάρχουν διάφορες παραλλαγές του, ανάλογα με το διαθέσιμο αριθμό δειγμάτων και το στόχο της ανάλυσης, εμείς θα εξετάσουμε την πιο απλή μορφή του. Σε αυτές τις περιπτώσεις, ο ερευνητής έχει στη διάθεση του ένα δείγμα ποιοτικών δεδομένων οργανωμένο σε ονομαστικές κατηγορίες, και ο στόχος είναι να χρησιμοποιηθούν τα δεδομένα αυτά ώστε να προσδιοριστεί η αναλογία (ή το ποσοστό) του πληθυσμού που ανήκει στην κάθε κατηγορία. Για την επίτευξη αυτού του στόχου διατυπώνεται μια μηδενική υπόθεση, που είτε δηλώνει ότι δεν υπάρχει κάποια συγκεκριμένη προτίμηση στις διαθέσιμες ονομαστικές κατηγορίες, είτε δηλώνει ότι τα ποσοστά που προτιμώνται από τα υποκείμενα δε διαφέρουν από τα ποσοστά άλλων πληθυσμών οι οποίοι αποτελούν το σημείο αναφοράς.

Και στις δύο περιπτώσεις, αυτό που προσδιορίζει η μηδενική υπόθεση είναι ο αναμενόμενος αριθμός (expected frequency – f_e) των υποκειμένων που ανήκει σε κάθε ονομαστική κατηγορία. Ο έλεγχος υποθέσεων που ακολουθεί αξιολογεί αυτή τη μηδενική υπόθεση, συγκρίνοντας τον αριθμό των υποκειμένων που αναμένεται σε κάθε ονομαστική κατηγορία με τον αριθμό των υποκειμένων που παρατηρείται ότι ανήκει σε κάθε ονομαστική κατηγορία (observed frequency - f_o), με βάση τις μετρήσεις του δείγματος. Ο στατιστικός δείκτης ελέγχου που χρησιμοποιείται για αυτή την αξιολόγηση είναι το X^2 .

Αν και, τα μη-παραμετρικά τεστ δε χρειάζεται να ικανοποιούν συγκεκριμένες προϋποθέσεις, τα συμπεράσματα που προκύπτουν από την εφαρμογή του στατιστικού τεστ X^2 είναι πιο αξιόπιστα

όταν χρησιμοποιείται αντιπροσωπευτικό δείγμα, και η αναμενόμενη συχνότητα όλων των κατηγοριών είναι μεγαλύτερη από πέντε.

2.2.5. ΣΥΣΧΕΤΙΣΗ

Γενικά ο όρος συσχέτιση αναφέρεται στο βαθμό με τον οποίο σχετίζονται (συμμεταβάλλονται) δύο ή περισσότερες μεταβλητές. Αφορά την κατεύθυνση της σχέσης των δυο μεταβλητών που μελετώνται. Ο βαθμός της σχέσης μεταξύ δύο μεταβλητών ονομάζεται απλή συσχέτιση και μεταξύ περισσότερων από δύο ονομάζεται πολλαπλή συσχέτιση, συχνά η συσχέτιση αυτή (αν είναι γραμμική) εκτιμάται με τη μέθοδο της απλής ή της πολλαπλής παλινδρόμησης αντίστοιχα.

Πιο συγκεκριμένα, η συσχέτιση μεταξύ δύο μεταβλητών ονομάζεται γραμμική όταν στο διάγραμμα διασκορπισμού τα σημεία τείνουν να συγκεντρώνονται γύρω από μια ευθεία, και μη γραμμική όταν τα σημεία τείνουν να συγκεντρώνονται γύρω από μια καμπύλη. Η απλή συσχέτιση ασχολείται με το βαθμό με τον οποίο τα σημεία συγκεντρώνονται γύρω από την ευθεία χωρίς να προσδιορίζεται αυτή η ευθεία που διέρχεται από το νέφος των σημείων. Η κατεύθυνση του νέφους των σημείων σηματοδοτεί το είδος της σχέσης, ενώ η συγκέντρωση των σημείων είναι ενδεικτική του βαθμού συμμεταβολής των δεδομένων. Η γραμμική συσχέτιση μπορεί να είναι:

- **Θετική συσχέτιση** όταν δύο μεταβλητές τείνουν να μεταβάλλονται προς την ίδια κατεύθυνση. Στην περίπτωση αυτή οι τιμές τείνουν να αυξάνονται ή να μειώνονται.
- **Αρνητική συσχέτιση** όταν δύο μεταβλητές τείνουν να μεταβάλλονται προς αντίθετη κατεύθυνση. Στην περίπτωση αυτή οι τιμές της μίας μεταβλητής τείνουν να αυξάνονται και της άλλης να μειώνονται
- **Μηδενική συσχέτιση** όταν οι μεταβολές των τιμών της μίας μεταβλητής δεν συνδέονται με τις μεταβολές της άλλης. Τα σημεία του νέφους είναι διασκορπισμένα σε όλο το μήκος του διαγράμματος.

Αναλυτικότερα, η συσχέτιση (correlation) είναι μία στατιστική τεχνική που εξυπηρετεί τον ερευνητή σε αυτές τις περιπτώσεις, καθώς του επιτρέπει να παρατηρεί δύο μεταβλητές στο

"φυσικό τους περιβάλλον" (μη-ελεγχόμενο περιβάλλον), και να προσδιορίζει και να μετρήσει επακριβώς τη σχέση που τις συνδέει. Σε αυτήν την εισαγωγική ενότητα θα επικεντρώσουμε το ενδιαφέρον μας μόνο στην αναζήτηση γραμμικών σχέσεων ανάμεσα σε μεταβλητές, και επομένως στον υπολογισμό γραμμικών συντελεστών συσχέτισης.

Η κατεύθυνση του νέφους των σημείων δηλώνει το είδος της σχέσης μεταξύ των δύο μεταβλητών ενώ η παραστατική συγκέντρωση των σημείων είναι ενδεικτική του βαθμού συμμεταβολής των δεδομένων. Η ομόρροπη ή αντίρροπη συμμεταβολή των δεδομένων υποδηλώνεται από το πρόσημο της συνδιακύμανσης.

Ο υπολογισμός του συντελεστή συσχέτισης απαιτεί τη μέτρηση δύο μεταβλητών για κάθε υποκείμενο (τα οποία συνήθως συμβολίζονται με X, Y), σε ένα μη-ελεγχόμενο περιβάλλον. Ο στόχος της τεχνικής είναι να προσδιορίσει τη σχέση που συνδέει τις δύο μεταβλητές, υπολογίζοντας την τιμή του συντελεστή συσχέτισης. Για παράδειγμα, ένας ερευνητής μπορεί να θέλει να προσδιορίσει και να ποσοτικοποιήσει τη σχέση που υπάρχει (αν υπάρχει) ανάμεσα στο ύψος και το βάρος των ανθρώπων.

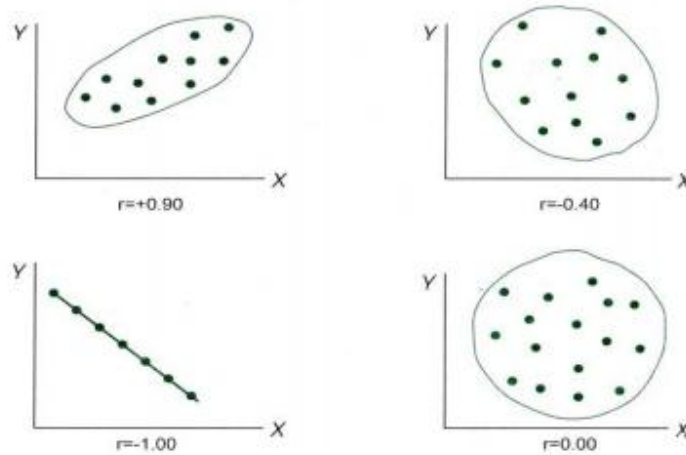
Οι τιμές που μπορεί να πάρει ο συντελεστής συσχέτισης κυμαίνονται από -1 έως $+1$. Αν η τιμή του είναι θετική, τότε οι δύο μεταβλητές τείνουν να μεταβάλλονται προς την ίδια κατεύθυνση, δηλαδή όταν αυξάνεται η μία αυξάνεται και η άλλη. Αν αντίθετα η τιμή του είναι αρνητική, τότε οι μεταβλητές τείνουν να κινούνται προς την αντίθετη κατεύθυνση, δηλαδή όταν η μία αυξάνεται η άλλη μειώνεται. Όσο μεγαλύτερη είναι η τιμή του συντελεστή συσχέτισης (ανεξαρτήτως προσήμου), τόσο μεγαλύτερη είναι η σχέση που συνδέει τις δύο μεταβλητές. Στην ακραία περίπτωση που οι δύο μεταβλητές έχουν συντελεστή συσχέτισης 0 , ο ερευνητής μπορεί να συμπεράνει ότι δεν υπάρχει κάποια σχέση ανάμεσα τους.

Αν και υπάρχουν διάφορες τεχνικές υπολογισμού του γραμμικού συντελεστή συσχέτισης ανάλογα με τα διαθέσιμα δεδομένα (ποσοτικά και ποιοτικά), η πιο συχνά χρησιμοποιούμενη μέθοδος είναι ο υπολογισμός του συντελεστή συσχέτισης Pearson (Pearson correlation), Ο οποίος συμβολίζεται με το γράμμα r .

Αξίζει να σημειωθεί ότι η τιμή του συντελεστή συσχέτισης υψωμένη στο τετράγωνο (r^2) εκφράζει το ποσοστό της διακύμανσης (αλλαγές στην τιμή) της Y μεταβλητής που μπορεί να προβλεφθεί από τη διακύμανση της X μεταβλητής. Για παράδειγμα, αν ο συντελεστής

συσχέτισης έχει υπολογιστεί σε $r=0.80$, τότε το $r^2=0.64=64\%$ των αλλαγών της Y μεταβλητής μπορεί να προβλεφθεί από τις αλλαγές στην X μεταβλητή.

Τέλος, αξίζει να αναφέρουμε ότι όταν ο ερευνητής θέλει να βγάλει συμπεράσματα και να ποσοτικοποιήσει τις σχέσεις ανάμεσα σε παραμέτρους του πληθυσμού, μπορεί να χρησιμοποιήσει το συντελεστή συσχέτισης ως στατιστικό δείκτη ελέγχου σε μία μέθοδο ελέγχου υποθέσεων. Σε αυτήν την περίπτωση, η μηδενική υπόθεση δηλώνει ότι δεν υπάρχει καμία συσχέτιση ανάμεσα στις δύο υπό εξέταση παραμέτρους, και οι κρίσιμες περιοχές ορίζονται από τιμές που βρίσκονται σε στατιστικούς πίνακες με κρίσιμες τιμές για το συντελεστή συσχέτισης Pearson.



Σημειώνεται ότι

- Θετικές τιμές – ομόρροπη συμμεταβολή – θετική συσχέτιση. Όσο πλησιέστερα στο 1 τόσο ισχυρότερη η θετική συσχέτιση - $\rho=1$ πλήρη θετική γραμμική συσχέτιση.
- Αρνητικές τιμές – αντίρροπη συμμεταβολή - αρνητική συσχέτιση. Όσο πλησιέστερα στο 1 τόσο ισχυρότερη η αρνητική συσχέτιση - $\rho=-1$ πλήρη αρνητική γραμμική συσχέτιση.
- Πλησιέστερα στο μηδέν – ασθενής θετική ή αρνητική συσχέτιση, γραμμική σχέση των μεταβλητών.
- Μηδενική - $\rho=0$, μηδενική συνδιακύμανση – ανυπαρξία γραμμικής συσχέτισης.

Με βάση το βαθμό συσχέτισης το πόσο ισχυρή είναι η συσχέτιση μεταξύ δύο μεταβλητών έχουμε τις παρακάτω διαβαθμίσεις

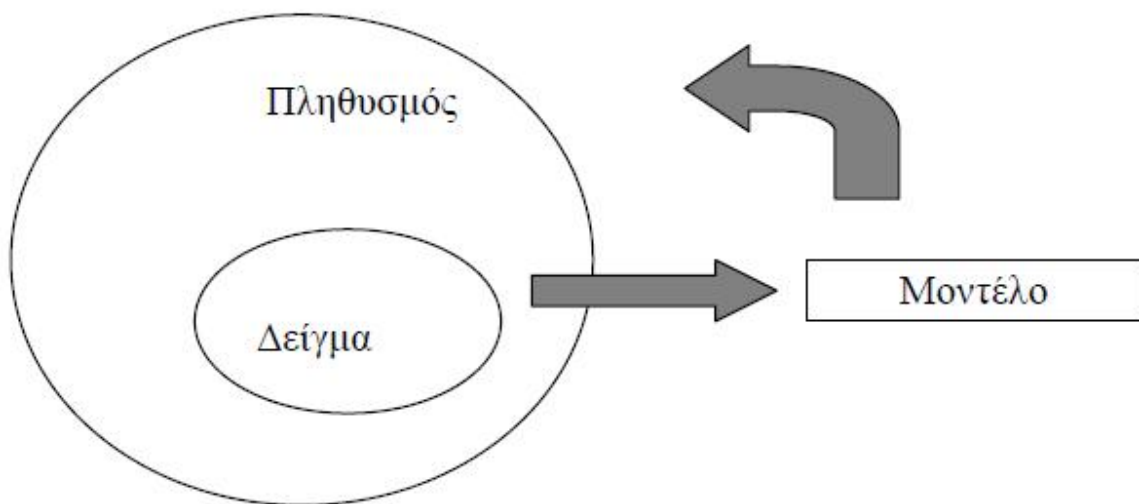
- Όταν $0.8 < r < 1$ ή $-1 < r < -0.8$ – πολλή σημαντική ή πολλή ισχυρή συσχέτιση
- Όταν $0.7 < r < 0.8$ ή $-0.8 < r < -0.7$ – σημαντική ή ισχυρή συσχέτιση.
- Όταν $0.5 < r < 0.7$ ή $-0.7 < r < -0.5$ – μέση συσχέτιση.
- Όταν $0.3 < r < 0.5$ ή $-0.5 < r < -0.3$ – ασθενής συσχέτιση.
- Όταν $-0.3 < r < 0.3$ – ανύπαρκτη συσχέτιση.
- Όταν $r = \pm 1$ – τέλεια συσχέτιση.

2.3. ΜΟΝΤΕΛΟΠΟΙΗΣΗ

Τα μαθηματικά μοντέλα αποτελούν σήμερα την πιο διαδεδομένη μέθοδο μελέτης φυσικών, κοινωνικών, οικονομικών, ιατρικών φαινομένων. Σε γενικό πλαίσιο, χρησιμοποιούνται για την ανάλυση και μελέτη τέτοιου είδους φαινομένων καθώς και την παράλληλη διεξαγωγή αποτελεσμάτων. Μαθηματικό μοντέλο μπορεί να θεωρηθεί προσομοίωση των πραγματικών φαινομένων τα οποία ακολουθούν συγκεκριμένους κανόνες. Οι κανόνες αυτοί και γενικότερα τα χαρακτηριστικά γνωρίσματα προσπαθούν να αντικατασταθούν από ανάλογους μαθηματικούς συσχετισμούς. Συχνά η πολυπλοκότητα των υπό μελέτη φαινομένων μας αναγκάζει να προβούμε σε απλοποιήσεις και παραδοχές. Η βασική απαίτηση είναι το μαθηματικό μοντέλο να εξηγεί με τον απλούστερο και καταλληλότερο τρόπο το συγκεκριμένο πρόβλημα που εξετάζεται.

Η μοντελοποίηση ή υποδειγματοποίηση των υπό εξέταση δεδομένων συχνά εξυπηρετούν τους μελετητές κυρίως όταν πρόκειται να βγάλουν συμπεράσματα ή να λάβουν μια απόφαση βασισμένοι σε γεγονότα που πιθανότατα επαναλαμβάνονται. Η επιτυχία στην κατασκευή ενός συγκεκριμένου μαθηματικού μοντέλου, έγκειται στην κατάλληλη χρήση της απλοποίησης και της παραδοχής, στο σωστό χρόνο, και στο σωστό στάδιο έτσι ώστε να περιέχει όσο το δυνατό περισσότερη πραγματικότητα. Η κατασκευή του μοντέλου στηρίζεται αρχικά στην παρατήρηση, την εμπειρία και την διαίσθηση που οδηγούν στην εξήγηση και στην διατύπωση θεωριών οι οποίες περιγράφουν με αντιπροσωπευτικό τρόπο το συγκεκριμένο φαινόμενο (προβλήματος). Μετά τον έλεγχο τους ακολουθεί η αναπροσαρμογή τους, η ανατροφοδότησή τους με καινούργια στοιχεία και η επαναδιατύπωσή τους με σκοπό τον έλεγχο (και σύγκριση) του προτεινόμενου μοντέλου με τα ήδη υπάρχοντα.

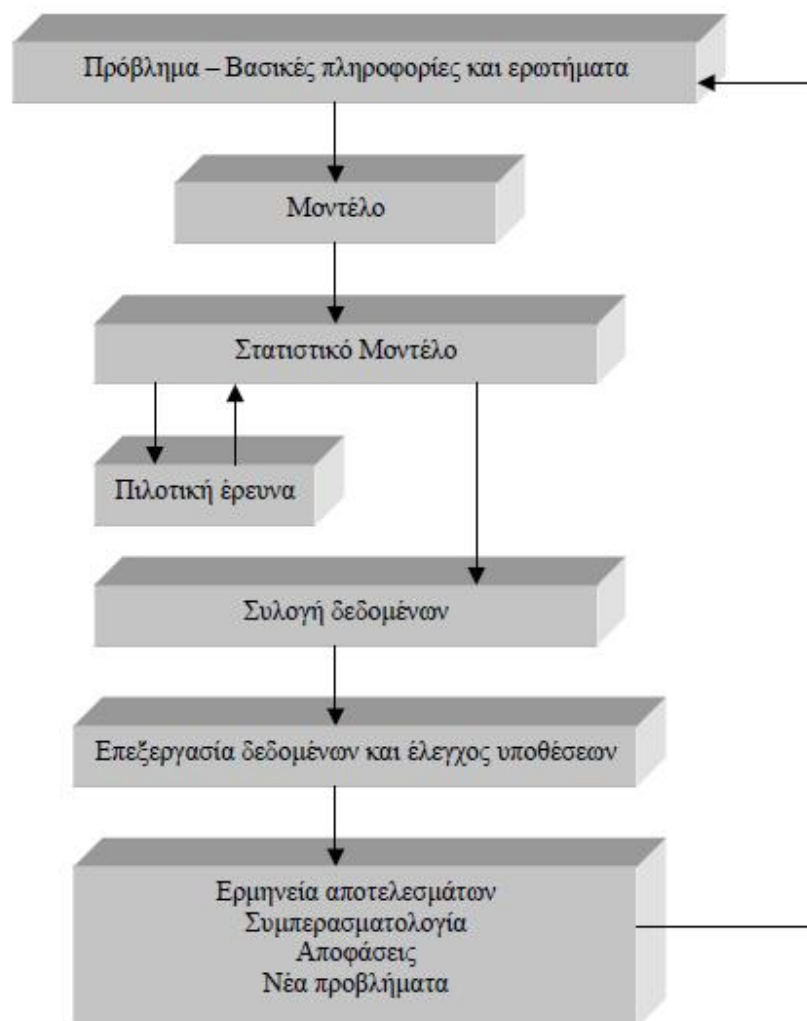
Κάθε μοντέλο (θεωρητικά ή πρακτικά) εξηγεί ένα συγκεκριμένο (ή ομάδα συγκεκριμένων) φαινομένων. Κάθε φαινόμενο εξελίσσεται όχι αφηρημένα αλλά υποκείμενα, αυτόνομες μονάδες παρατήρησης, το σύνολο των οποίων ορίζει τον πληθυσμό. Το φαινόμενο αναλύεται σε επιμέρους μετρήσιμα χαρακτηριστικά, τις μεταβλητές, στις οποίες αντιστοιχούμε τιμές. Η αντιστοίχιση αυτή ονομάζεται μέτρηση και γίνεται με την χρήση εργαλείων γενικού χαρακτήρα. Πρακτικά, τις περισσότερες φορές, είναι αδύνατη η μελέτη του πληθυσμού; στην περίπτωση αυτή ένα υπο-σύνολο του πληθυσμού λαμβάνεται με σκοπό την ανάλυση και διεξαγωγή συμπερασμάτων. Το υπό-σύνολο αυτό ονομάζεται δείγμα του πληθυσμού.



Μεταφέροντας ένα πρόβλημα σε υποθέσεις θα πρέπει να επιλέξουμε από τα χαρακτηριστικά του, εκείνα τα οποία αφενός είναι σημαντικά αφετέρου μετρήσιμα. Η χρήση της στατιστικής συμπερασματολογίας καθώς και μεθόδων ελέγχου μοντέλων αποτελεί σημαντικό βήμα για την καταλληλότερη επιλογή του συγκεκριμένου μοντέλου όπου αφενός θα εξηγεί με απλό και κατανοητό το συγκεκριμένο πρόβλημα αφετέρου θα είναι απλό στην χρήση και τον χρόνο λήψης αποφάσεων. Τα στάδια επεξεργασίας και ανάλυσης των δεδομένων δίνονται στο σχήμα που ακολουθεί.

Μεταφέροντας ένα πρόβλημα σε υποθέσεις, θα πρέπει να επιλέξουμε από τα χαρακτηριστικά του (μεταβλητές) εκείνα που είναι σημαντικά και μετρήσιμα. Εστιάζοντας την προσοχή σε τμήματα που είναι σημαντικά προσπαθούμε να κατασκευάσουμε μοντέλο / μοντέλα διαφορετικά

από τα υπάρχοντα διακρινόμενα για την απλότητα, σαφήνεια και εφαρμογή των υπό μελέτη στοιχείων του προβλήματος. Ένα μοντέλο είναι καλύτερο όταν μας πληροφορεί πληρέστερα για το πρόβλημα προς ανάλυση ενώ μας επιτρέπει να κάνουμε ακριβέστερες προβλέψεις. Είναι ιδιαίτερα σημαντικό να αναφερθεί ότι μετά τον έλεγχο προσαρμογής των δεδομένων που συγκεντρώθηκαν με το στατιστικό μοντέλο που δημιουργήσαμε, μπορούμε να υπολογίσουμε ένα διάστημα εμπιστοσύνης για τις παραμέτρους του πληθυσμού και να προχωρήσουμε στον έλεγχο υποθέσεων γι' αυτές.



2.4. ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΟΙΚΟΝΟΜΙΑ

Η οικονομία και η διοίκηση επιχειρήσεων αποτελούν τομείς ευρέως διαδεδομένους και ιδιαίτερα σημαντικούς για την εξέλιξη των κοινωνιών. Είναι πλέον γεγονός πως και οι δυο αυτές επιστήμες προκειμένου να λειτουργήσουν ορθά και ομαλά απαιτούν τη χρήση μεγάλου όγκου δεδομένων προκειμένου να προσδιορίσουν για παράδειγμα την καταναλωτική συμπεριφορά, τις ανάγκες της αγοράς κ.ά. Επομένως και σε αυτή την περίπτωση απαιτείται ο κατάλληλος τρόπος διαχείρισης μεγάλου όγκου δεδομένων.

Όπως ήδη αναφέρθηκε η μοντελοποίηση συχνά εξυπηρετεί πολλούς από τους στόχους των οικονομολόγων και των υπεύθυνων για τη λήψη αποφάσεων σε επιχειρήσεις και οργανισμούς. Από τα παραπάνω προκύπτει πως η εξόρυξη δεδομένων εφαρμόζεται και στον τομέα της οικονομίας. Τα οικονομικά δεδομένα κυρίως συλλέγονται από τράπεζες και από άλλους οικονομικούς οργανισμούς. Τα δεδομένα αυτά συνήθως είναι αξιόπιστα, ολοκληρωμένα και έχουν υψηλή ποιότητα και απαιτούν συστηματική μέθοδο για την ανάλυση αυτών. Η συνεισφορά της εξόρυξης δεδομένων στην επιστήμη της οικονομίας συναντάται στην συλλογή και κατανόηση των δεδομένων, στην βελτίωση δεδομένων (data refinement), στην δημιουργία και εκτίμηση ενός μοντέλου και στην ανάπτυξη αυτού. Η σωστή ανάλυση των οικονομικών δεδομένων μας διευκολύνει στο να παίρνουμε καλύτερες αποφάσεις ενεργώντας σύμφωνα με την ανάλυση της αγοράς. Τα εργαλεία και οι τεχνικές της εξόρυξης δεδομένων βοηθούν στο να αναλύσουμε τα οικονομικά δεδομένα με τους παρακάτω τρόπους:

Τα δεδομένα που συλλέγονται από διάφορα οικονομικά ιδρυτήματα, όπως οι τράπεζες, συγκεντρώνονται αρχικά στην αποθήκη δεδομένων (data warehouse). Οι τεχνικές της πολυδιάστατης ανάλυσης δεδομένων χρησιμοποιούνται για την ανάλυση τέτοιων δεδομένων που συλλέγονται στην αποθήκη δεδομένων για τις γενικές ιδιότητές του.

Μία άλλη εφαρμογή της εξόρυξης δεδομένων σχετίζεται με την πρόβλεψη αποπληρωμής δανείου και πολιτικές πίστωσης του πελάτη. Μέθοδοι της εξόρυξης όπως η επιλογή χαρακτηριστικών (feature selection) βοηθάει στην ταυτοποίηση ποικίλων χαρακτηριστικών όπως το επίπεδο εισοδήματος του πελάτη, την εξόφληση ανάλογα με τα έσοδα, την πιστωτική του ιστορία κτλ. Με την επεξεργασία αυτών των χαρακτηριστικών, η τράπεζα μπορεί να

αποφασίζει για τις πολιτικές δανειοδότησης βάσει των σχετικά χαμηλών κινδύνων. Οι τεχνικές της συσταδοποίησης και της ταξινόμησης βοηθούν τα οικονομικά ινστιτούτα να ομαδοποιούν διάφορους πελάτες που έχουν κοινά χαρακτηριστικά. Η αποτελεσματική συσταδοποίηση και οι μέθοδοι φιλτραρίσματος βοηθούν τις τράπεζες να ταυτοποιούν μία ομάδα πελατών, να συσχετίζουν ένα νέο πελάτη με την παρούσα ομάδα και να τους παρέχουν κοινά οφέλη.

Τα εργαλεία της εξόρυξης δεδομένων βοηθούν τα οικονομικά ινστιτούτα να αναγνωρίζουν τις ανάγκες και τα εγκλήματα από παραποιημένα δεδομένα από τις διάφορες βάσεις δεδομένων και από το ιστορικό συναλλαγών που έγιναν από τους πελάτες. Οι τεχνικές οπτικοποίησης βοηθούν στην παρουσίαση δεδομένων με διαφορετικές μορφές, όπως γράφοι που βασίζονται σε συγκεκριμένα γνωρίσματα. Προβάλλοντας τα δεδομένα από διάφορες οπτικές γωνίες, η τράπεζα δύναται να διακρίνει τους πελάτες που έχουν επιχειρήσει παράνομες πράξεις και μετά μια λεπτομερή έρευνα αυτών των ύποπτων περιπτώσεων βοηθάει στην εξιχνίαση των απατών και των εγκλημάτων.

Οι επιχειρήσεις και οι οργανισμοί που αξιοποιούν το σύνολο της διαθέσιμης πληροφορίας συνδυάζοντας τις τεχνολογίες συλλογής (Data Collection & Market Research) και ανάλυσης δεδομένων (Data Mining), επιτυγχάνουν σημαντικά ακριβέστερες προβλέψεις για την μελλοντική συμπεριφορά του πελάτη.

Όλες οι διαθέσιμες πηγές δεδομένων των επιχειρήσεων (δεδομένα Marketing, συναλλαγών, Web, Text, Survey data κλπ) συνδέονται με την αναλυτική πλατφόρμα πρόβλεψης της SPSS που οδηγεί στην κατανόηση των επιχειρηματικών προβλημάτων.

Η ραγδαία ανάπτυξη των νέων τεχνολογιών και η ενσωμάτωσή τους στην παραγωγή και τις υπηρεσίες έχει δημιουργήσει πρωτόγνωρες συνθήκες για τους χειριστές της οικονομικής πολιτικής. Η πρόκληση που αντιμετωπίζουν είναι: 1) να μπορέσουν να παρακολουθήσουν όσα διαδραματίζονται μέσω της συλλογής κατάλληλων στατιστικών στοιχείων, και 2) μέσω οικονομικής ανάλυσης να προτείνουν και να εφαρμόσουν κατάλληλες πολιτικές που θα οδηγήσουν στη μέγιστη δυνατή ανάπτυξη και ευημερία.

Τα επίσημα στατιστικά στοιχεία διαδραματίζουν πολύ σημαντικό ρόλο τόσο στη διαμόρφωση πολιτικής, όσο και στην παρακολούθηση της υλοποίησης στόχων, παρέχοντας στους χρήστες μια ολοκληρωμένη και επιστημονικά τεκμηριωμένη εικόνα για καταστάσεις που τυγχάνουν

μελέτης και ελέγχου. Ο ρόλος των επίσημων στατιστικών γίνεται ακόμα πιο ζωτικός σε περιόδους οικονομικής κρίσης, όπως και αυτή που διανύουμε τώρα, καθώς αποτελούν το βασικό εργαλείο για την παρακολούθηση της πορείας της οικονομίας αλλά και για τον έλεγχο επίτευξης προκαθορισμένων στόχων, με την ποσοτικοποίηση των διαφόρων μεγεθών και καταστάσεων που παρατηρούνται στο κοινωνικό και οικονομικό περιβάλλον.

2.5. ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΔΙΟΙΚΗΣΗ ΕΠΙΧΕΙΡΗΣΕΩΝ

Όπως ήδη αναφέρθηκε η Στατιστική είναι επιστήμη που επιχειρεί να εξαγάγει γνώση χρησιμοποιώντας εμπειρικά δεδομένα.

Η λέξη στατιστική: προέρχεται από τη λατινική φράση *statisticum collegium* (διάλεξη για υποθέσεις της πολιτείας), από την οποία προήρθε η Ιταλική λέξη *statista*, που σημαίνει πολιτικός, και η Γερμανική λέξη *Statistik*, η οποία αρχικά αναφερόταν στην ανάλυση των δεδομένων για την πολιτεία. Πήρε την έννοια της συλλογής και ταξινόμησης δεδομένων γενικά στις αρχές του δεκάτου ένατου αιώνα.

Η Στατιστική είναι ένα χρήσιμο εργαλείο για την κατανόηση και τη μελέτη των πραγματικών καταστάσεων των οποίων δεν γνωρίζουμε την πραγματική τους εσωτερική δομή.

Η χρήση τους στη βιομηχανία, μηχανική, οικονομία, διοίκηση, βιολογία, ιατρική, χημεία είναι σήμερα τόσο διαδεδομένη και τόσο καθοριστική που θεωρείται πλέον ένα από τα πιο χρήσιμα και δημοφιλή εργαλεία της μαθηματικής επιστήμης.

Οι επιχειρήσεις χρησιμοποιούν ολοένα και περισσότερο στατιστικές μεθόδους για να μετατρέπουν δεδομένα σε πληροφορίες μιας και οι στατιστικές μέθοδοι αποτελούν ένα κρίσιμο εργαλείο για τη σύγχρονη διοίκηση και οικονομία.

Η λύση κάθε στατιστικού προβλήματος χωρίζεται σε τρία στάδια, που σημειώνονται σε κάθε παράδειγμα: (1) επιλογή της μεθόδου, (2) υπολογισμός των στατιστικών μεγεθών, και (3) ερμηνεία των αποτελεσμάτων. Ουσιαστικά, η στατιστική είναι ένα εργαλείο που μετατρέπει δεδομένα σε πληροφορίες.

2.6. ΣΧΕΤΙΚΑ ΠΑΡΑΔΕΙΓΜΑΤΑ

ΠΑΡΑΔΕΙΓΜΑ 1^ο:

Ένας νέος τύπος τσιγάρων βρίσκεται στο στάδιο ποιοτικού ελέγχου. Αν το τμήμα ποιοτικού ελέγχου της καπνοβιομηχανίας παραγωγής, ενδιαφέρεται να γνωρίζει τη μέση ποσότητα νικοτίνης που περιέχεται στα νέου τύπου τσιγάρα, μπορεί να υπολογίσει ένα διάστημα εμπιστοσύνης και να πάρει έτσι μια εκτίμηση για την άγνωστη μέση ποσότητα νικοτίνης. Στην περίπτωση όμως, που ενδιαφέρεται να γνωρίζει μόνο αν στα νέου τύπου τσιγάρα η μέση ποσότητα νικοτίνης δεν υπερβαίνει ένα μέγιστο επιτρεπτό όριο, τότε πρέπει να κάνει κατάλληλο στατιστικό έλεγχο υποθέσεων ώστε να μπορεί να αποφασίσει μεταξύ των υποθέσεων:

Η μέση ποσότητα νικοτίνης δεν υπερβαίνει το μέγιστο επιτρεπτό όριο.

Η μέση ποσότητα νικοτίνης υπερβαίνει το μέγιστο επιτρεπτό όριο.

Ο στατιστικός έλεγχος υποθέσεων (hypothesis testing) είναι μια συμπερασματική διαδικασία/ μέθοδος που προσφέρει η Στατιστική Συμπερασματολογία και βρίσκει εφαρμογή σε στοχαστικά προβλήματα απόφασης μεταξύ δύο εναλλακτικών υποθέσεων. Η μία υπόθεση έχει επικρατήσει να συμβολίζεται με H_0 και ονομάζεται μηδενική υπόθεση (null hypothesis), και η άλλη με H_1 και ονομάζεται εναλλακτική υπόθεση (alternative hypothesis).

Αναγκαία προϋπόθεση για τη σωστή εφαρμογή των στατιστικών ελέγχων και κυρίως για τη σωστή ερμηνεία των αποτελεσμάτων τους, είναι η κατανόηση της λογικής και του νοήματός τους.

ΜΕΡΟΣ Β': ΠΕΡΙ SPSS Ο ΛΟΓΟΣ

ΚΕΦΑΛΑΙΟ 3^ο: SPSS: ΕΞΕΛΙΞΗ ΚΑΙ ΧΡΗΣΗ

3.1. ΣΤΑΤΙΣΤΙΚΗ ΜΕ ΤΗ ΧΡΗΣΗ SPSS

Υπάρχει πληθώρα διαθέσιμων στατιστικών εργαλείων και πακέτων λογισμικού, τα οποία καλύπτουν όλες τις ανάγκες μιας στατιστικής ανάλυσης, από τις πιο απλές (όπως οι περιγραφικές και επαγωγικές τεχνικές που αναφέραμε σε αυτό το κεφάλαιο), μέχρι και τις πιο σύνθετες (π.χ. μη γραμμικές συσχετίσεις, παλινδρόμηση, εξόρυξη γνώσης από δεδομένα, κλπ).

Τα πιο γνωστά στατιστικά εργαλεία είναι τα εμπορικά πακέτα λογισμικού (π.χ. SPSS, SAS, Statistica, κλπ) τα οποία προσφέρουν μία μεγάλη ποικιλία στατιστικών αναλύσεων.

Η ανάγκη για άμεση εφαρμογή της στατιστικής σε ποικίλες επιστήμες οδήγησε στην δημιουργία στατιστικών πακέτων τα οποία έχουν την δυνατότητα εισαγωγής, επεξεργασίας, ανάλυσης και παρουσίασης δεδομένων σε σύντομο χρονικό διάστημα. Φυσικά χάρη στους υπολογιστές υπάρχουν απεριόριστες πλέον δυνατότητες όσον αφορά τον όγκο των δεδομένων.

Το S.P.S.S. (Statistical Package for Social Sciences) είναι ένα στατιστικό πακέτο που έχει πολλές δυνατότητες όσον αφορά την επεξεργασία και παρουσίαση των δεδομένων μιας επιστημονικής έρευνας αλλά και μεγάλη αξιοπιστία. Οι τελευταίες εκδόσεις του S.P.S.S. έχουν γραφικό περιβάλλον, πράγμα που το καθιστά πολύ εύκολο για την χειριστή του.

3.2. ΤΙ ΕΙΝΑΙ SPSS - ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ – ΕΞΕΛΙΞΗ

Το SPSS (Statistical Package for Social Sciences) είναι ένα ολοκληρωμένο στατιστικό πακέτο και μπορεί κανείς να το κατατάξει στην ίδια κατηγορία με τα SAS, GENSTAT, STATISTICA, NAG και STATGRAPHICS. Οι σύγχρονες εξελίξεις στους υπολογιστές έχουν επιτρέψει τη χρήση του σε οικιακούς υπολογιστές PCs. Αρχικά αναπτύχθηκε στο Πανεπιστήμιο Stanford της Καλιφόρνιας το 1965 για την επίλυση στατιστικών προβλημάτων, που αφορούν τις κοινωνικές επιστήμες. Στη συνέχεια συμπληρώθηκε με στατιστικά εργαλεία που χρησιμοποιούνται σε

διάφορους επιστημονικούς και επαγγελματικούς τομείς με αποτέλεσμα σήμερα να χρησιμοποιείται σχεδόν σε κάθε είδους στατιστική ανάλυση και να έχει καταξιωθεί ως ένα από τα φιλικότερα και αποτελεσματικότερα στατιστικά προγράμματα.

Η πρώτη έκδοση του πακέτου κυκλοφόρησε το 1965 και έτρεχε σε μεγάλα υπολογιστικά συστήματα. Στη συνέχεια στις αρχές τις δεκαετίας του 80 έγινε προσαρμογή του πακέτου για μικροϋπολογιστές. Η πρώτη έκδοση του πακέτου για μικροϋπολογιστές ήταν η SPSS/PC+ ver. 1.0. και έτρεχε σε περιβάλλον λειτουργικού συστήματος MS-DOS.

Στη συνέχεια ακολούθησαν τέσσερις εκδόσεις σε περιβάλλον λειτουργικού συστήματος MS-DOS, οι οποίες περιελάμβαναν νέες στατιστικές επεξεργασίες, νέες δυνατότητες ελέγχου του περιβάλλοντος εργασίας και των βημάτων, που ακολουθούνται σε μία στατιστική ανάλυση και νέες δυνατότητες βοήθειας στην εξήγηση και στην επιλογή των εντολών του SPSS/PC+.

Στις αρχές της δεκαετίας του 90 με την έλευση του γραφικού περιβάλλοντος των παραθύρων (Windows 3.0, Windows 3.1) και στη συνέχεια με την εμφάνιση του νέου λειτουργικού συστήματος Windows 95TM το SPSS εμφανίζει εκδόσεις, οι οποίες μπορούν και τρέχουν θαυμάσια κάτω από το παραθυρικό περιβάλλον. Το SPSS για Windows γίνεται πακέτο πολύ φιλικό προς τον χρήστη, επειδή όπως όλες οι παραθυρικές εφαρμογές λειτουργεί με ένα σύστημα μενού και πλαισίων διαλόγου (dialogs boxes). Επίσης είναι πολύ πιο συμπληρωμένο σε σχέση με τις εκδόσεις του πακέτου σε περιβάλλον MS-DOS. Έτσι δημιουργώντας την έκδοση για τα Windows το SPSS κάνει την ανάλυση δεδομένων διαθέσιμη στις μάζες δηλ ότι έκανε ο Bill Gates για τους υπολογιστές.

Στο παραθυρικό περιβάλλον, η πρώτη έκδοση που κυκλοφόρησε ήταν το SPSS για τα Windows ver. 5.0 και στην συνέχεια ακολούθησαν μέχρι σήμερα 8 εκδόσεις. Η τελευταία έκδοση είναι η 22.0. Η εταιρεία κατασκευής του πακέτου φροντίζει, ώστε σε κάθε νέα έκδοση ο βασικός κορμός του πακέτου να παραμένει σταθερός και απλά να συμπληρώνονται καινούργιες στατιστικές διαδικασίες.

Στην Ελλάδα το στατιστικό πακέτο SPSS για Windows υποστηρίζεται από την SPSS Hellas, η οποία είναι αποκλειστική αντιπρόσωπος των προϊόντων SPSS εκτός από την Ελλάδα στην Κύπρο και στη Μάλτα.

Όπως ήδη έχει αναφερθεί, το SPSS για Windows καλύπτει ένα ευρύ πεδίο στατιστικών εφαρμογών και επεξεργασίας δεδομένων όλων των κατηγοριών και επιπέδων, τα οποία χωρίζονται σε επιμέρους τμήματα (modules). Τα τρία κύρια τμήματα του SPSS, που αναφέρονται πιο κάτω είναι τα εξής :

1. Βασικό Τμήμα (Base Module). Αυτό περιλαμβάνει τεχνικές διαχείρισης και επεξεργασίας δεδομένων. Μία σειρά απλών στατιστικών αναλύσεων, όπως κατανομές συχνοτήτων, πίνακες διπλής, τριπλής εισόδου και παρουσίασης αποτελεσμάτων.

2. Επαγγελματικό Τμήμα (Professional Module). Αυτό περιλαμβάνει στατιστικές τεχνικές παραμετρικής (π.χ. t-test, ανάλυση της διακύμανσης (Analysis of Variance), συσχέτιση (Correlation) και παλινδρόμηση (Regression)) και μη παραμετρικής ανάλυσης (π.χ. Wilcoxon, X^2 , Friedman, Mann-Whitney και Kruskal –Wallis) και

3. Προχωρημένο Τμήμα (Advanced Statistics). Αυτό περιλαμβάνει πολύπλοκες στατιστικές αναλύσεις (εφαρμογές), όπως, παραγοντική ανάλυση (Factor analysis), ανάλυση κατά συστάδες (Cluster analysis), μη γραμμική παλινδρόμηση, διακριτική ανάλυση (Discriminant analysis) κλπ.

Επίσης εκτός από τα τρία κύρια τμήματα το SPSS για Windows περιλαμβάνει και άλλα. Μερικά από αυτά είναι:

1. Τμήμα Χρονολογικών Σειρών (Trend Module). Αυτό περιλαμβάνει τεχνικές επεξεργασίας και ανάλυσης δεδομένων σε χρονολογικές σειρές.

2. Τμήμα Παραγωγής Πινάκων (Table Module). Με αυτό μπορεί κανείς να δημιουργήσει πίνακες απλούς ή περίπλοκους, μίας, δύο ή τριών διαστάσεων έτοιμους προς παρουσίαση ή δημοσίευση.

3. Τμήμα των Γραφικών (Graphic Module). Αυτό περιλαμβάνει περισσότερα από 50 είδη διαγραμμάτων δισδιάστατα ή τρισδιάστατα σε μορφή ραβδογράμματος (bar chart), ιστογράμματος, γραμμικών διαγραμμάτων (line chart), διαγραμμάτων διασποράς (scatterplots) κλπ. Για την δημιουργία και επεξεργασία αυτών των διαγραμμάτων το SPSS για Windows διαθέτει ειδικό ενσωματωμένο συντάκτη διαγραμμάτων και υποστηρίζει όλα τα γνωστά πρότυπα της Microsoft για το περιβάλλον των Windows, όπως TDynamic Data Exchange (TDDEx), TObject Linking and Embedding (TOLE), TObject Database Connectivity (TODBCT), τα

οποία επιτρέπουν να μοιράζεται κανείς τα δεδομένα και τα διαγράμματα με άλλα πακέτα, που τρέχουν σε παραθυρικό περιβάλλον.

4. Τμήμα Κατηγορικών Δεδομένων (Categories Module). Αυτό περιλαμβάνει στατιστικές τεχνικές για την ανάλυση δεδομένων έρευνας αγοράς.

5. Τμήμα για Εισαγωγή δεδομένων (Data Entry) . Με αυτό μπορεί κάποιος να δημιουργήσει φόρμες για την εισαγωγή των δεδομένων σε αρχεία μορφής δεδομένων SPSS για Windows, δηλαδή σε αρχεία SAV.

6. Τμήμα Επεξεργασίας Δεδομένων Ποιοτικού Ελέγχου (QI Analyst Module). Αυτό το τμήμα περιλαμβάνει στατιστικές τεχνικές για δεδομένα προερχόμενα από το χώρο του ποιοτικού ελέγχου.

7. Τμήμα δημιουργίας εξελιγμένων Αναφορών (Report Writer). Με αυτό μπορεί κανείς να δημιουργήσει εύκολα και γρήγορα επαγγελματικές αναφορές.

Τα πολλά χαρακτηριστικά του SPSS είναι προσβάσιμα μέσω pull-down μενού ή μπορεί να χρησιμοποιηθούν μέσω μιας προγραμματιστικής γλώσσας 4ης γενιάς για σύνταξη εντολών. Το παραπάνω έχει τα πλεονεκτήματα της αναπαραγωγής, την απλούστευση των επαναλαμβανόμενων εργασιών καθώς και τη διαχείριση πολύπλοκων χειρισμών και στοιχείων ανάλυσης. Επιπλέον, κάποιες σύνθετες εφαρμογές μπορεί να προγραμματιστούν μόνο σε σύνταξη και δεν είναι προσβάσιμα μέσω της δομής του μενού. Το pull-down μενού δημιουργεί επίσης σύνταξη των εντολών. Αυτό μπορεί να εμφανίζεται στην έξοδο, αν και οι προεπιλεγμένες ρυθμίσεις πρέπει να αλλάξουν για να γίνει η σύνταξη ορατή στο χρήστη. Μπορούν επίσης να επικολληθούν σε ένα αρχείο χρησιμοποιώντας την επιλογή "paste" που υπάρχει σε κάθε μενού. Τα προγράμματα μπορούν να τρέχουν με αλληλεπίδραση ή χωρίς επιτήρηση, χρησιμοποιώντας τη παρεχόμενη επιλογή Production Job Facility.

Επιπλέον, μία «μακρο» γλώσσα μπορεί να χρησιμοποιηθεί για να γράψει υπορουτίνες εντολών και επέκταση προγραμματισμού Python που μπορεί να έχει πρόσβαση στις πληροφορίες, στο λεξικό δεδομένων και δυναμικά στην οικοδόμηση των προγραμμάτων σύνταξη της εντολής. Η επέκταση προγραμματισμού Python, που εισήχθη στο SPSS 14, αντικατέστησε το λιγότερο λειτουργικό SAX Basic "σενάρια" για τους περισσότερους σκοπούς, αν και SAX Basic

παραμένει διαθέσιμο. Επιπλέον, η επέκταση Python επιτρέπει στο SPSS να ασκεί οποιαδήποτε από τις στατιστικές στο ελεύθερο λογισμικό πακέτο R. Μετά έκδοση SPSS 14 μπορεί να οδηγηθεί έξω από Python ένα πρόγραμμα ή VB.NET με το παρεχόμενο "plug-ins".

Το πρόγραμμα SPSS θέτει περιορισμούς στην εσωτερική δομή του αρχείου, τους τύπους δεδομένων, στην επεξεργασία δεδομένων και στα αρχεία που ταιριάζουν, και έτσι απλουστεύεται σημαντικά. Στο SPSS τα σύνολα δεδομένων έχουν μια δύο-διαστάσεων δομή του πίνακα, όπου οι σειρές αντιπροσωπεύουν συνήθως περιπτώσεις (όπως τα άτομα ή νοικοκυριά) και οι στήλες αντιπροσωπεύουν μετρήσεις (όπως η ηλικία, το φύλο ή το εισόδημα των νοικοκυριών). Μόνο δύο τύποι δεδομένων ορίζονται: αριθμητική και κείμενο. Όλη η επεξεργασία των δεδομένων γίνεται διαδοχικά κατά περίπτωση μέσω του αρχείου. Τα αρχεία μπορούν να συνδυαστούν ένα προς ένα και ένα προς πολλά, αλλά όχι πολλά προς πολλά.

Η γραφική επαφή χρήστη έχει δύο όψεις που μπορεί να εναλλάσσεται κάνοντας κλικ σε μία από τις δύο καρτέλες στο κάτω αριστερό μέρος του παραθύρου του SPSS.

Το «Data view» δείχνει μια άποψη φύλλου των περιπτώσεων (σειρές) και μεταβλητών (στήλες). Σε αντίθεση με τα φύλλα, τα κελιά δεδομένων μπορεί να περιέχουν μόνο αριθμούς ή κείμενο και ενώ τύποι δεν μπορούν να αποθηκευτούν σε αυτά τα κελιά. Το «Variable View» εμφανίζει το λεξικό μετα-δεδομένων, όπου κάθε σειρά αντιπροσωπεύει μια μεταβλητή και δείχνει το όνομα της μεταβλητής και μια ποικιλία από άλλα χαρακτηριστικά (variable label, value label(s), print width, measurement type). Τα κελιά στις δύο προβολές (data view και variable view) μπορεί να είναι χειροκίνητα, καθορίζοντας τη δομή των αρχείων και επιτρέποντας την εισαγωγή δεδομένων χωρίς τη χρήση εντολής σύνταξης. Αυτό μπορεί να είναι επαρκές για τις μικρές σειρές δεδομένων. Τα μεγαλύτερα σύνολα δεδομένων, όπως στατιστικών ερευνών, πιο συχνά δημιουργούνται στο λογισμικό εισαγωγής δεδομένων, ή εισάγονται κατά τη διάρκεια με τη βοήθεια ηλεκτρονικού υπολογιστή με προσωπική συνέντευξη, με σάρωση και τη χρήση οπτικής αναγνώρισης χαρακτήρων και λογισμικό οπτικής αναγνώρισης σήμα, ή με άμεση σύλληψη από ερωτηματολόγια σε απευθείας σύνδεση. Αυτά τα σύνολα δεδομένων τότε διαβάζονται σε SPSS.

Το SPSS μπορεί να διαβάσει και να γράψει δεδομένα από αρχεία κειμένου ASCII (συμπεριλαμβανομένων των ιεραρχικών αρχείων), άλλα πακέτα στατιστικά στοιχεία, λογιστικά φύλλα και βάσεις δεδομένων. Επίσης το SPSS μπορεί να διαβάσει και να γράψει σε εξωτερικούς

σχετικούς πίνακες της βάσης δεδομένων μέσω ODBC και SQL. Η στατιστική έξοδος είναι ένα αρχείου (*. Αρχείο SPV, υποστηρίζοντας ρινοτ πίνακες) το οποίο μπορεί να μεταφορτωθεί. Το αρχείο μπορούν να εξαχθούν σε κείμενο ή το Microsoft Word, PDF, Excel, και άλλες μορφές. Εναλλακτικά, η παραγωγή μπορεί να συλληφθεί ως δεδομένα (χρησιμοποιώντας την εντολή OMS), όπως κείμενο, κείμενο οριοθετημένο με στηλοθέτες, PDF, XLS, HTML, XML, SPSS σύνολο δεδομένων ή μια ποικιλία από μορφές γραφικών εικόνας (JPEG, PNG, BMP και EMF).

3.3. ΔΥΝΑΤΟΤΗΤΕΣ ΤΟΥ SPSS

Το SPSS για Windows μπορεί να δεχθεί και να διαχειρισθεί πολλά είδη εξωτερικών αρχείων δεδομένων. Υπάρχουν όμως και κάποια συγκεκριμένα αρχεία, που τα φτιάχνει κατά την εφαρμογή του. Αυτά είναι :

Αρχεία δεδομένων (Data files). Παίρνουν αυτόματα την προέκταση *.SAV. Περιέχουν τα δεδομένα που καταχωρίζουμε στο παράθυρο επεξεργασίας δεδομένων (Data editor).

Αρχεία αποτελεσμάτων (Output Files). Παίρνουν αυτόματα την προέκταση *.SPO. Περιέχουν τα αποτελέσματα των στατιστικών αναλύσεων. Μπορούμε να τα επεξεργαστούμε μέσα από το παράθυρο προβολής αποτελεσμάτων (Viewer). Από εκεί μπορούμε επίσης, αν θέλουμε, να τα αντιγράψουμε και να τα επικολλήσουμε απευθείας στο κείμενο των ευρημάτων της έρευνας.

Αρχεία εντολών (Syntax files). Παίρνουν αυτόματα την προέκταση *.SPS. Περιέχουν εντολές του SPSS.

Αρχεία γραφημάτων (Chart files). Παίρνουν αυτόματα την προέκταση *.CHT. Περιέχουν γραφήματα που δημιουργήθηκαν με εντολές του SPSS.

Μια μεταβλητή μπορεί να μετρηθεί σε διάφορες κλίμακες. Η πιο ατελής είναι η ονομαστική (nominal), η οποία χρησιμοποιείται αποκλειστικά για όλες σχεδόν τις ποιοτικές μεταβλητές, εκείνες δηλαδή που δεν μπορούν να μετρηθούν αλλά μόνο να απαριθμηθούν. Τέτοιες είναι το φύλο, το επάγγελμα, η οικογενειακή κατάσταση, η καταγωγή. Στην ονομαστική κλίμακα η ταξινόμηση των ατόμων μιας ομάδας γίνεται σε κατηγορίες που προσδιορίζονται με βάση

ορισμένα κριτήρια. Οι κατηγορίες είναι ανεξάρτητης σειράς και αμοιβαία αποκλειόμενες, έτσι ώστε μία εγγραφή που εμφανίζεται σε μία κατηγορία αποκλείεται να εμφανίζεται σε άλλη. Η θέση της εγγραφής στην κατηγορία καθορίζει την ομοιότητα ή την διαφορά της έναντι των άλλων και όχι τη σειρά της, το βαθμό υπεροχής ή υστέρησής της.

Η τακτική κλίμακα (ordinal ή κλίμακα τάξης) μοιάζει με την ονομαστική, έχει όμως το επιπλέον χαρακτηριστικό ότι οι εγγραφές ταξινομούνται σε κατηγορίες που ακολουθούν σειρά φυσική ή λογικά παραδεκτή, αύξουσα η φθίνουσα. Η ιεράρχηση των εγγραφών γίνεται εδώ με βάση τη σχετική υπεροχή ή την υστέρηση σε σχέση με τις άλλες ως προς ορισμένο κριτήριο. Για παράδειγμα, στην ερώτηση «Πόσο συχνά παρακολουθείτε τηλεπαιχνίδια;» οι απαντήσεις θα μπορούσαν να διακριθούν στις εξής κατηγορίες: «ποτέ», «σπάνια», «μερικές φορές», «συχνά», «κάθε μέρα».

Τέλος, η κλίμακα ισο-διαστημάτων (scale) έχει τις ίδιες ιδιότητες με την τακτική και επιπλέον την ιδιότητα του προκαθορισμού ίσων διαστημάτων ή διαφορών μεταξύ των κατηγοριών σε οποιοδήποτε τμήμα της κλίμακας. Χαρακτηρίζεται από κοινή και σταθερή μονάδα μέτρησης. Μπορούμε να προσδιορίσουμε λοιπόν την απόσταση μεταξύ δύο εγγραφών στην κλίμακα αυτή. Το σημείο μηδέν όμως ορίζεται αυθαίρετα και για το λόγο αυτό οι σχέσεις μεταξύ δύο εγγραφών δεν μπορούν να εκφραστούν με βάση κάποια αναλογία. Για παράδειγμα, αν συγκρίνουμε την ευφυΐα δύο ατόμων, μπορούμε να πούμε ότι κάποιος που έχει IQ 120 έχει 10 μονάδες μεγαλύτερο IQ από κάποιον με δείκτη 110. Δεν μπορούμε όμως να πούμε ότι κάποιος με δείκτη 150 έχει διπλάσια ευφυΐα από κάποιον με 75.

Χρησιμοποιώντας τα μενού ή τις εντολές του SPSS ζητάει κανείς να εκτελεσθούν κυρίως τέσσερις κατηγορίες εργασιών, συνήθως με τη παρακάτω σειρά:

- Αναγνώριση, Εισαγωγή Δεδομένων
- Μετασχηματισμός Δεδομένων
- Γραφική Παρουσίαση Δεδομένων
- Ανάλυση Δεδομένων

Για την καταχώρηση και επεξεργασία των δεδομένων μιας πειραματικής έρευνας, το πρώτο στάδιο είναι η καταχώρηση των δεδομένων στον Data Editor του S.P.S.S., δηλαδή στο περιβάλλον του S.P.S.S.

Δηλαδή σε κάθε στήλη καταχωρούμε τις απαντήσεις της ίδιας ερώτησης, ενώ κάθε γραμμή αντιστοιχεί στις απαντήσεις κάθε υποκειμένου.

Τα βασικά μενού του Data Editor, η κύρια γραμμή επιλογής συνοψίζεται στα ακόλουθα:

File με βασικές εντολές:

- New : δημιουργία νέου αρχείων δεδομένων
- Open : άνοιγμα αρχείων δεδομένων
- Save & Save as : σώζουμε το αρχείο δεδομένων που δημιουργήσαμε ή ανοίξαμε
- Print : εκτυπώνουμε
- Exit : βγαίνουμε από το πρόγραμμα

Edit με βασικές εντολές:

- Undo : αναίρεση της τελευταίας ενέργειας
- Cut : αποκοπή
- Copy : αντιγραφή
- Paste : επικόλληση
- Find : αναζήτηση εγγραφής
- Options : Γενικές ρυθμίσεις

View με βασικές εντολές:

- Status bar : γραμμή κατάστασης
- Toolbars : διαμόρφωση γραμμών εντολών
- Fonts : διαμόρφωση γραμματοσειρών
- Grid lines : εμφάνιση - απόκρυψη περιγραμμάτων κελιών
- Value labels : εμφάνιση – απόκρυψη labels

Data με βασικές εντολές:

- Insert case : εισαγωγή νέας γραμμής
- Insert variable : εισαγωγή νέας στήλης
- Got to case : γρήγορη μετάβαση στον αριθμό της γραμμής που θέλουμε

- Sort case : ταξινόμηση δεδομένων κατά αύξουσα ή φθίνουσα σειρά

Transform αλλαγές σε μεταβλητές π.χ. επανακωδικοποίηση τιμών, πράξεις μεταξύ μεταβλητών

Analyze Στατιστική ανάλυση π.χ. υπολογισμός στατιστικών μέτρων

Graphs Γραφικές παραστάσεις

Utilities δημιουργία ενός πίνακα περιεχομένων για τις μεταβλητές

Windows πληροφορίες για αρχεία σχετικά με το πρόγραμμα π.χ. αρχεία δεδομένων, αποτελεσμάτων, γραφημάτων και εντολών που ενεργοποιήθηκαν κατά τη διάρκεια της στατιστικής ανάλυσης

Help βοήθεια.

3.3.1. ΣΤΑΤΙΣΤΙΚΟΙ ΕΛΕΓΧΟΙ ΜΕ ΤΟ SPSS

ΤΕΣΤ ΑΝΕΞΑΡΤΗΣΙΑΣ – ΣΥΣΧΕΤΙΣΗ ΜΕΤΑΒΛΗΤΩΝ

Δυο ποιοτικές μεταβλητές ή μια ποιοτική και μια ποσοτική μεταβλητή

Ελέγχουμε την υπόθεση: οι δύο μεταβλητές δεν συσχετίζονται.

Analyze

Descriptive Statistics

Crosstabs

Output

- Πίνακας: Chi-Square Tests

Pearson Chi –Square (γραμμικός συντελεστής συσχέτισης Pearson)

Αν η σημαντικότητα (significance) του συντελεστή συσχέτισης Pearson < 0.05 , τότε η υπόθεση H_0 απορρίπτεται, δηλαδή οι δύο μεταβλητές είναι συσχετισμένες.

Δυο ποσοτικές μεταβλητές: ΠΑΛΙΝΔΡΟΜΗΣΗ

Ο γενικός όρος ‘Ανάλυση Παλινδρόμησης’ περιγράφει μία σειρά αναλύσεων που είναι πλέον πολύ δημοφιλείς στην επιστημονική κοινότητα. Γενικά, η ανάλυση παλινδρόμησης μας βοηθά να εξετάσουμε σχέσεις αιτιότητας (causal relationships), να ερμηνεύσουμε δηλαδή τη διακύμανση των τιμών της εξαρτημένης μεταβλητής με βάση μία ή περισσότερες ανεξάρτητες μεταβλητές.

Χρησιμοποιείται λοιπόν προκειμένου να:

- Προβλέψουμε τις τιμές μιας μεταβλητής με βάση τις τιμές μίας ή περισσότερων άλλων μεταβλητών.
- Αποφασίσουμε αν κάποια μεταβλητή είναι ‘καλή’ για την πρόβλεψη κάποιας άλλης μεταβλητής.
- Να βρούμε ποιό είναι το ποσοστό της διακύμανσης των τιμών μιας μεταβλητής που μπορεί να εξηγηθεί από τις τιμές μίας μεμονωμένης ή ενός συνόλου μεταβλητών.
- Να φτιάξουμε καινούρια μοντέλα και να ελέγξουμε υπάρχουσες θεωρίες.

Απλή γραμμική παλινδρόμηση:

Τα ζεύγη των τιμών των δυο μεταβλητών (x , y) προσαρμόζονται σε μια ευθεία. Ψάχνουμε τους συντελεστές της ευθείας και αν γίνεται καλή προσαρμογή. Εξετάζουμε την υπόθεση

H_0 : δεν υπάρχει γραμμική σχέση μεταξύ των μεταβλητών X , Y .

Analyze

Regression

Curve Estimation ή Linear Regression

Πολλαπλή γραμμική παλινδρόμηση:

Επέκταση της απλής γραμμικής παλινδρόμησης για δύο ή περισσότερες ανεξάρτητες μεταβλητές. Εξετάζουμε την υπόθεση

H_0 : δεν υπάρχει γραμμική σχέση μεταξύ των μεταβλητών.

Από το μενού επιλέγουμε πάλι:

Analyze

Regression

Linear

Μία απαραίτητη προϋπόθεση η οποία είναι απαραίτητη γενικά σε όλα τα μοντέλα με περισσότερες εκ της μίας ανεξάρτητων μεταβλητών είναι η έλλειψη συγγραμμικότητας. Η συγγραμμικότητα είναι ένα σοβαρό πρόβλημα για την πολλαπλή γραμμική παλινδρόμηση. Όταν μία ανεξάρτητη μεταβλητή συσχετίζεται με μία άλλη ανεξάρτητη, δηλαδή μέσω της μίας μπορούμε να εκτιμήσουμε τις τιμές της άλλης τότε μιλάμε για πρόβλημα συγγραμμικότητας. Επομένως η ύπαρξη και των δύο μεταβλητών στο μοντέλο δεν είναι δυνατή.

ΣΥΓΚΡΙΣΗ ΜΕΣΩΝ ΟΡΩΝ ΔΕΙΓΜΑΤΩΝ/ ΠΛΗΘΥΣΜΟΥ

Το δείγματα πρέπει να κατανέμονται κανονικά ή να είναι > 30 .

(1) Σύγκριση μέσου όρου ενός δείγματος ή ενός πληθυσμού με μια σταθερή τιμή.

Μηδενική υπόθεση H_0 : η μέση τιμή ενός δείγματος είναι ίση με μια δοσμένη τιμή.

Analyze

Compare Means

One Sample T-Test

Μεταφέρουμε στο πλαίσιο Test Variable(s) τη μεταβλητή της οποίας τη μέση τιμή θέλουμε να συγκρίνουμε με δεδομένη τιμή. Την τιμή για την οποία θα γίνει η σύγκριση εισάγουμε στη συνέχεια στη θέση Test Value.

(2) Σύγκριση των μέσων όρων δύο ανεξάρτητων δειγμάτων – Κατηγοριοποίηση των τιμών μιας ποσοτικής μεταβλητής με βάση μια ποιοτική μεταβλητή.

Μηδενική υπόθεση H_0 : οι μέσες τιμές δύο δειγμάτων είναι ίσες.

Analyze

Compare Means

Independent Samples t-Test

(3) Σύγκριση των μέσων όρων δύο εξαρτημένων δειγμάτων

Πολλές φορές για την πραγματοποίηση μιας μελέτης, είναι απαραίτητο να συλλέξουμε παρατηρήσεις πάνω στα ίδια αντικείμενα σε δυο διαφορετικές χρονικές στιγμές μετρώντας την ίδια παράμετρο/ μεταβλητή.

Μηδενική υπόθεση H_0 : οι μέσες τιμές δύο δειγμάτων είναι ίσες.

Analyze

Compare Means

Paired Samples t-Test

Αν η σημαντικότητα (significance) κάθε ελέγχου < 0.05 τότε η μηδενική υπόθεση της ισότητας των μέσων τιμών απορρίπτεται και άρα συμπεραίνουμε ότι τα δείγματα δεν έχουν ίσες μέσες τιμές.

ΚΕΦΑΛΑΙΟ 4^ο: ΕΦΑΡΜΟΓΗ

Για τη ολοκλήρωση της παρούσας ενότητας χρησιμοποιήσαμε δεδομένα τα οποία αφορούσαν στοιχεία 100 ατόμων για τις ακόλουθες μεταβλητές: Φύλο (Sex 1: Άνδρας, 2: Γυναίκα), Ηλικία (Age), Βάρος (Weight), Ύψος (Height), Τύπος εργασίας (Job_type 1: Αυτοαπασχολούμενος, 2: Υπάλληλος γραφείου, 3: Διευθυντικό στέλεχος), Επίπεδο εβδομαδιαίας σωματικής άσκησης, όχι απαραίτητως αθλητικής μορφής (Exercise_index 1: καθόλου, 2: ελαφρά μορφή, 3: μέτρια μορφή 4: βαριά άσκηση).

Με βάση τα συγκεκριμένα δεδομένα πραγματοποιήσαμε μια σειρά ελέγχων, παραμετρικών και μη παραμετρικών. Αναλυτικότερα:

Προκειμένου να απαντήσουμε στα εξής ερωτήματα:

α) άτομα διαφορετικού φύλου έχουν παρόμοιο βάρος.

β) άτομα διαφορετικού φύλου έχουν παρόμοιο δείκτη μάζας σώματος.

γ) άτομα διαφορετικής ηλικιακής ομάδας έχουν παρόμοιο βάρος.

δ) άτομα διαφορετικού ηλικιακής ομάδας έχουν παρόμοιο δείκτη μάζας σώματος.

Πραγματοποιούμε παραμετρικούς και μη παραμετρικούς ελέγχους.

Αναλυτικότερα:

ΠΑΡΑΜΕΤΡΙΚΟΙ ΕΛΕΓΧΟΙ:

Προκειμένου να εφαρμόσουμε έναν παραμετρικό έλεγχο πριν ελέγχουμε τις υποθέσεις εφαρμογής αυτού: κανονικότητα, ιστόγραμμα και Q-Q plot.

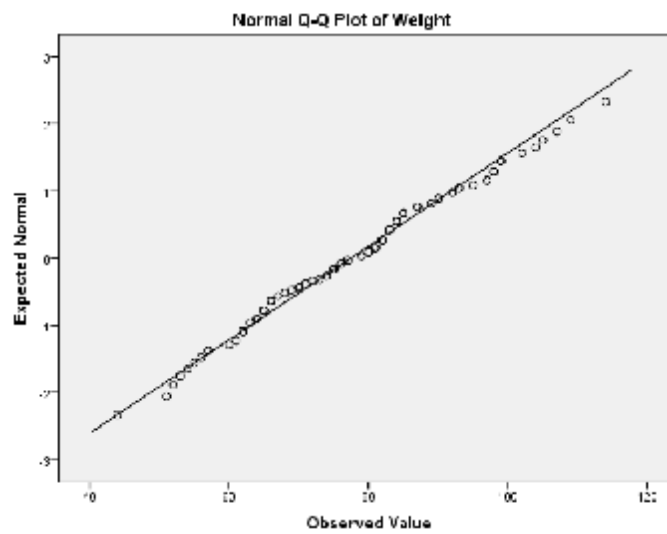
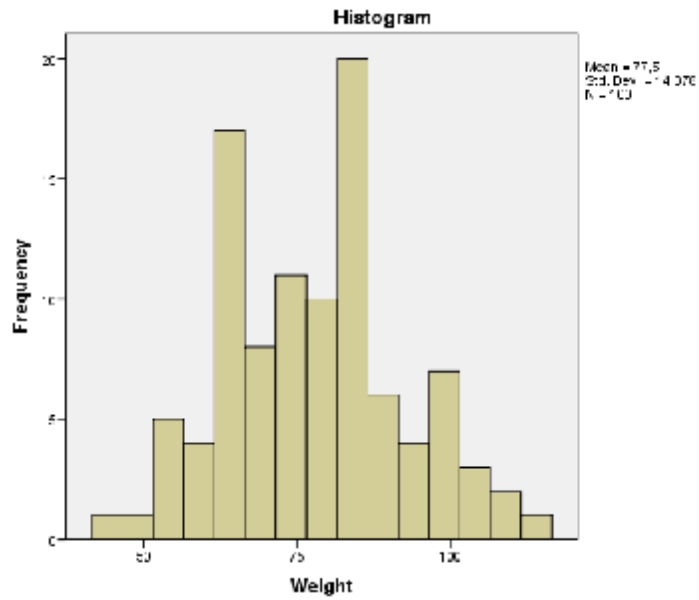
Για κάθε ένα από τα ερωτήματα που ακολουθούν εφαρμόζουμε το κριτήριο το Sig. όπου όταν αυτό είναι μεγαλύτερο από 0,05 έχουμε κανονικά κατανεμημένα δεδομένα οπότε μπορούμε να εφαρμόσουμε τον αντίστοιχο έλεγχο.

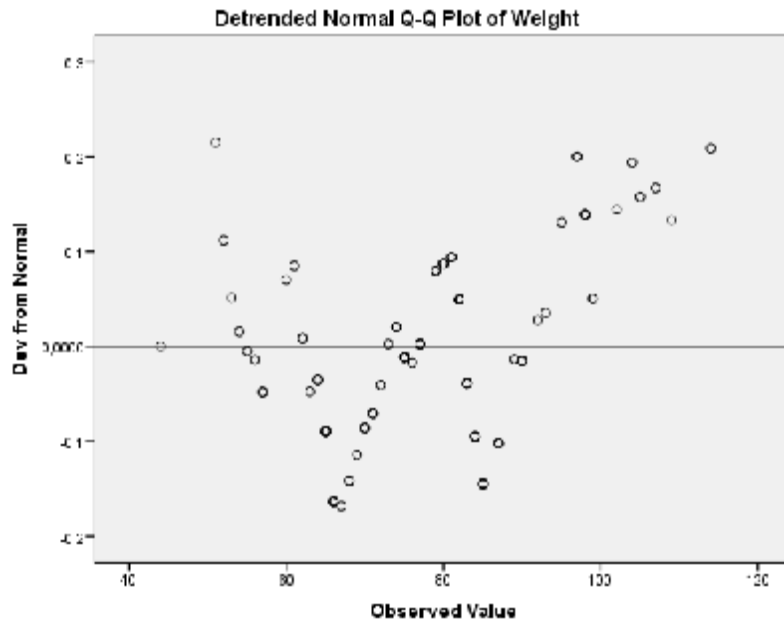
Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Weight	,071	100	,200*	,989	100	,562

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

a. Lilliefors Significance Correction

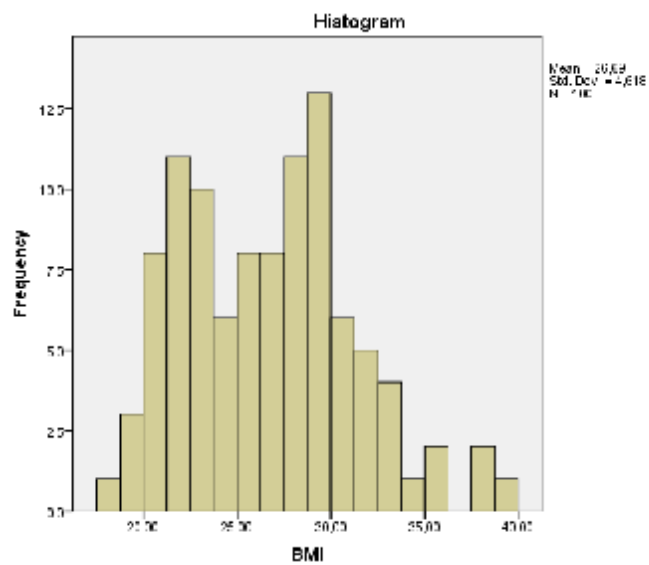


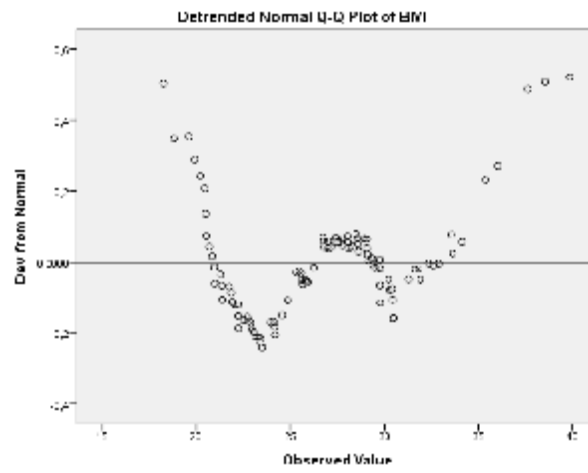
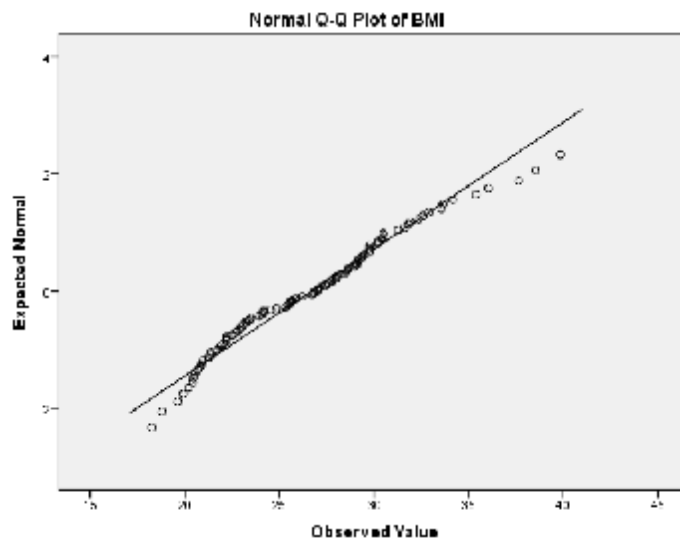


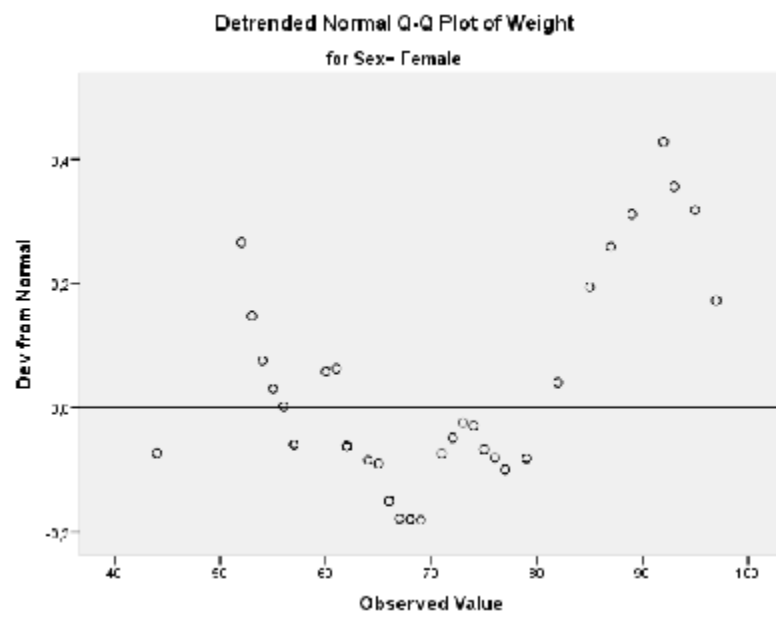
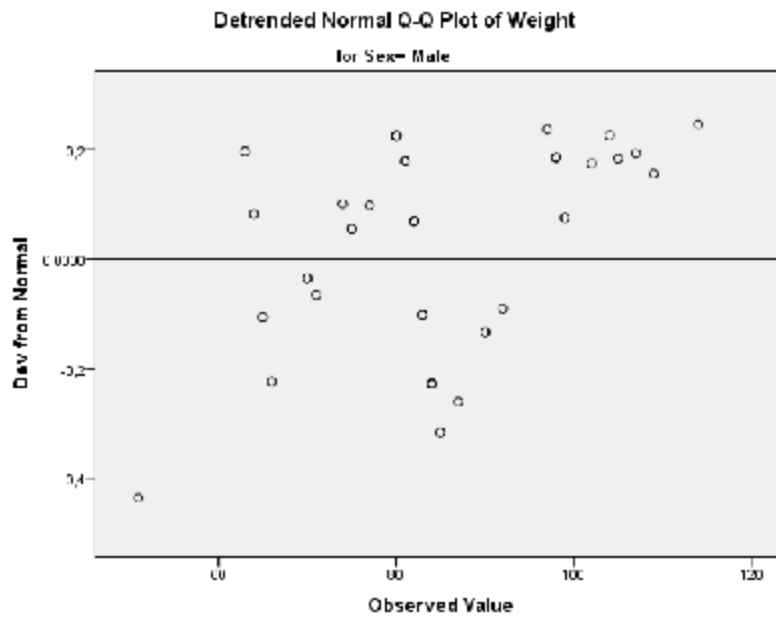
Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
BMI	,085	100	,074	,972	100	,031

a. Lilliefors Significance Correction





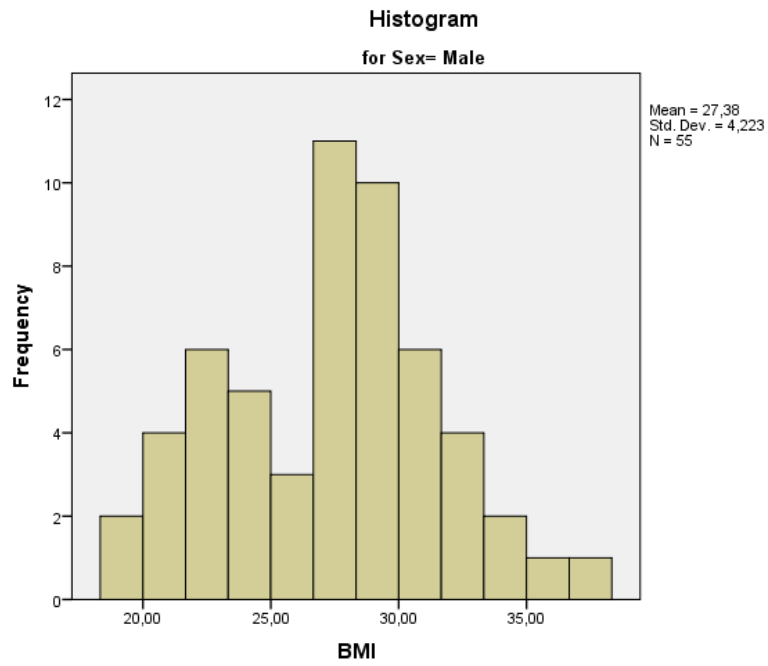


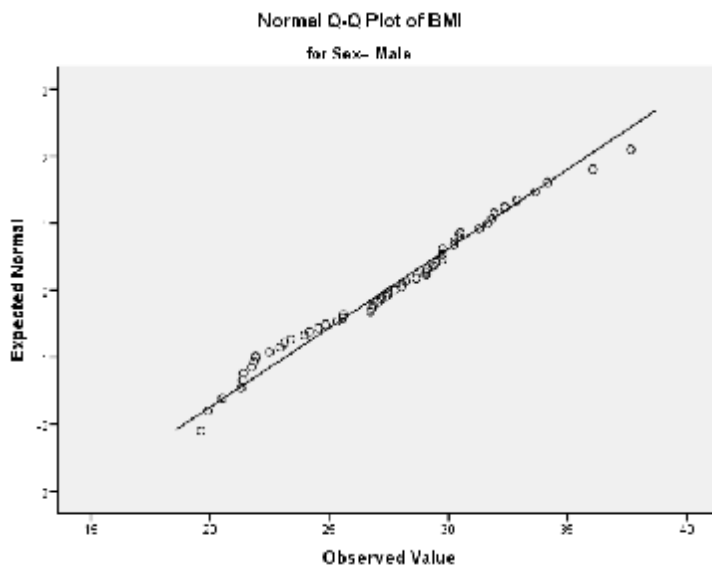
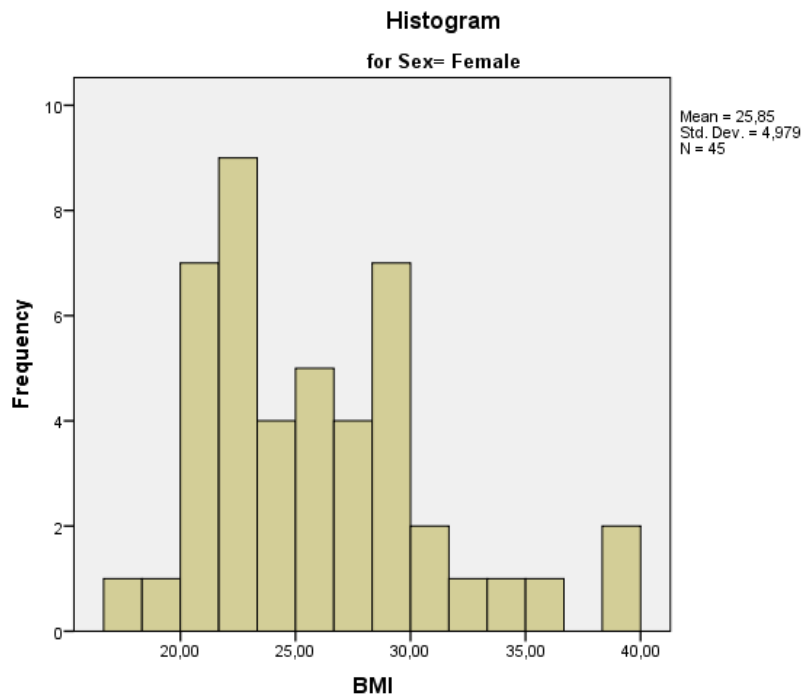
Tests of Normality

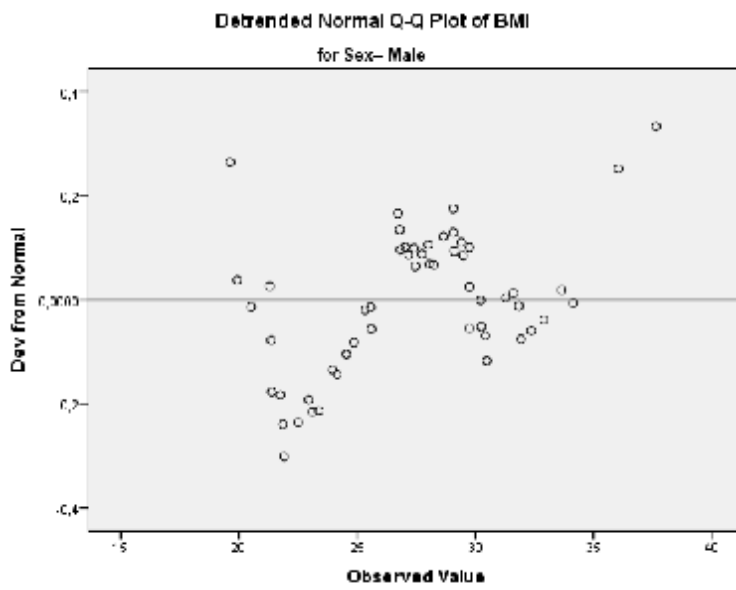
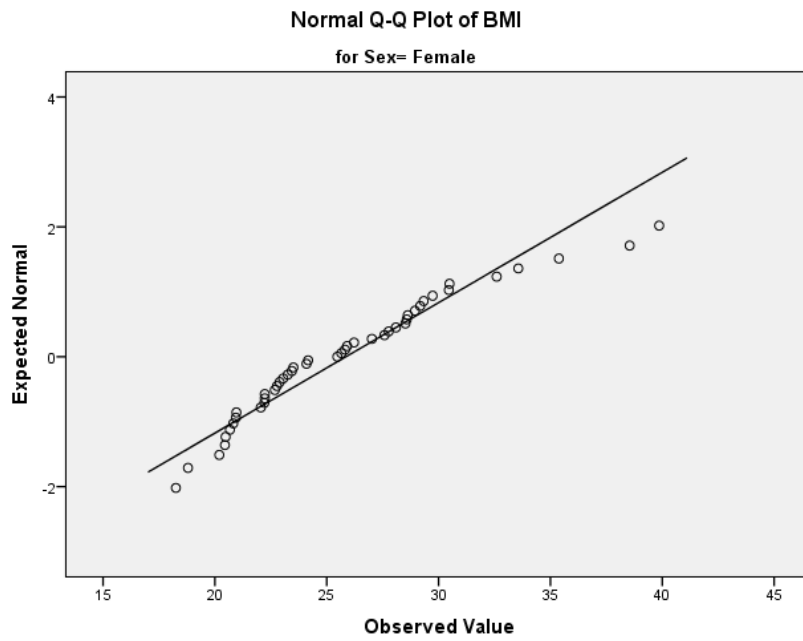
	Sex	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
BMI	Male	,076	55	,200*	,979	55	,435
	Female	,126	45	,072	,935	45	,015

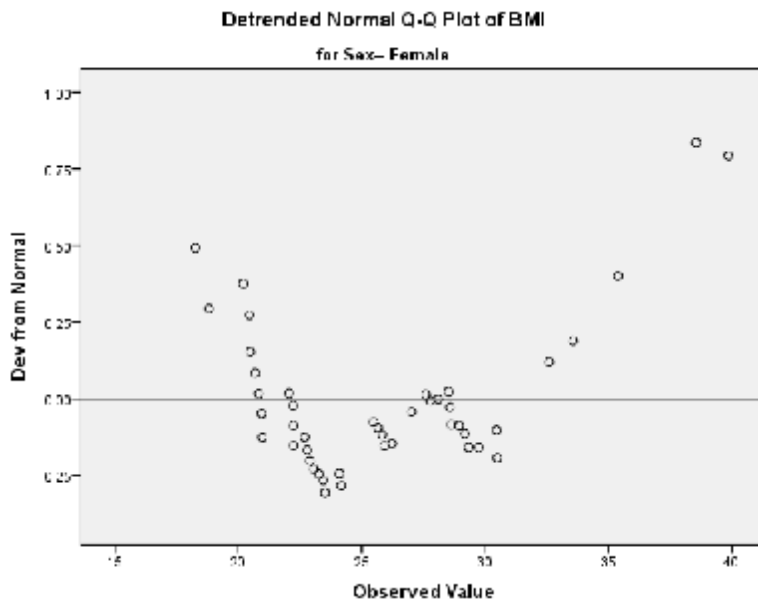
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction







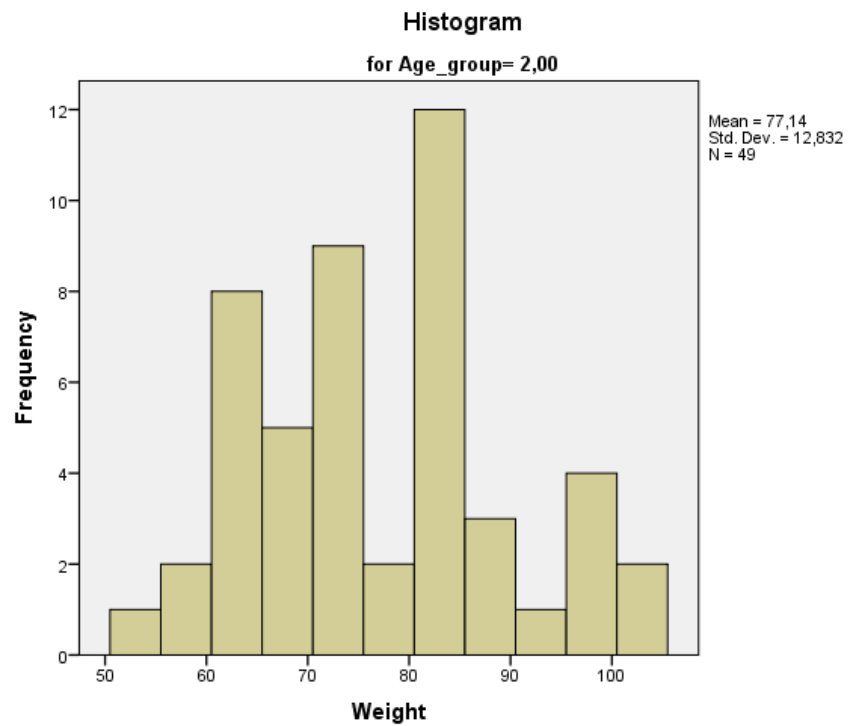
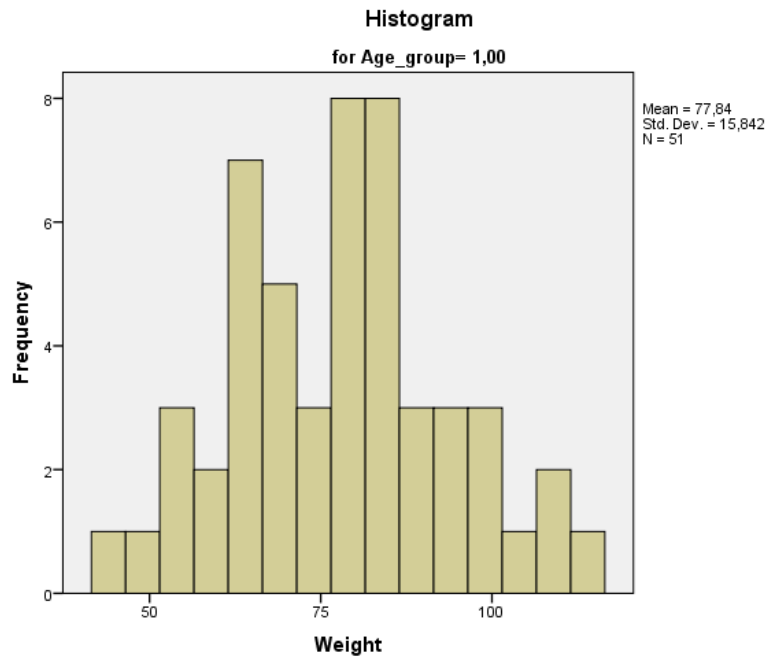


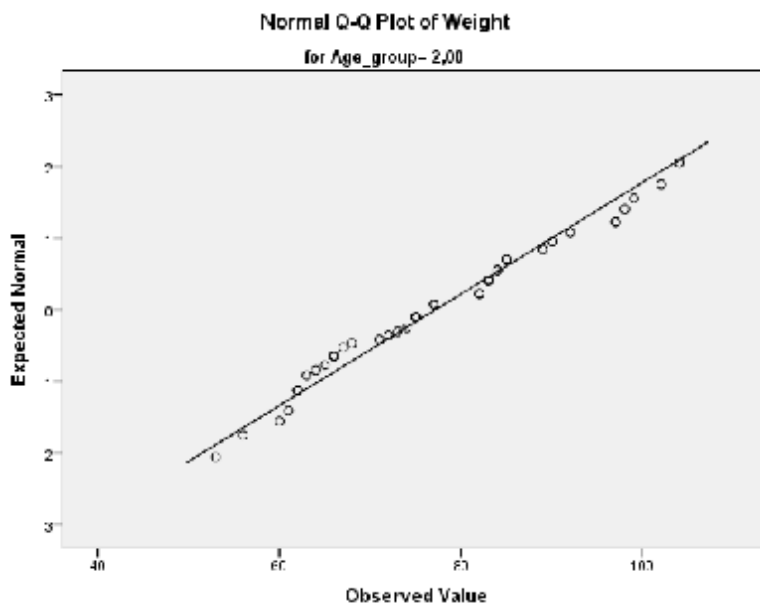
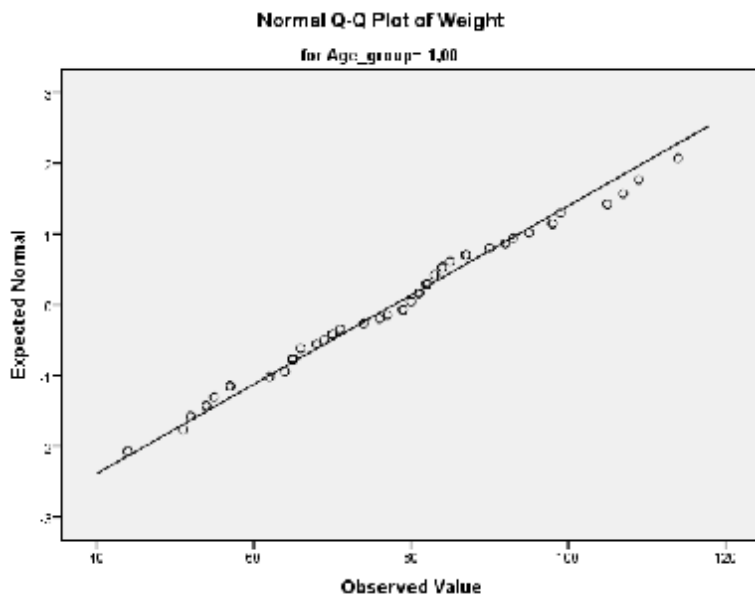
Tests of Normality

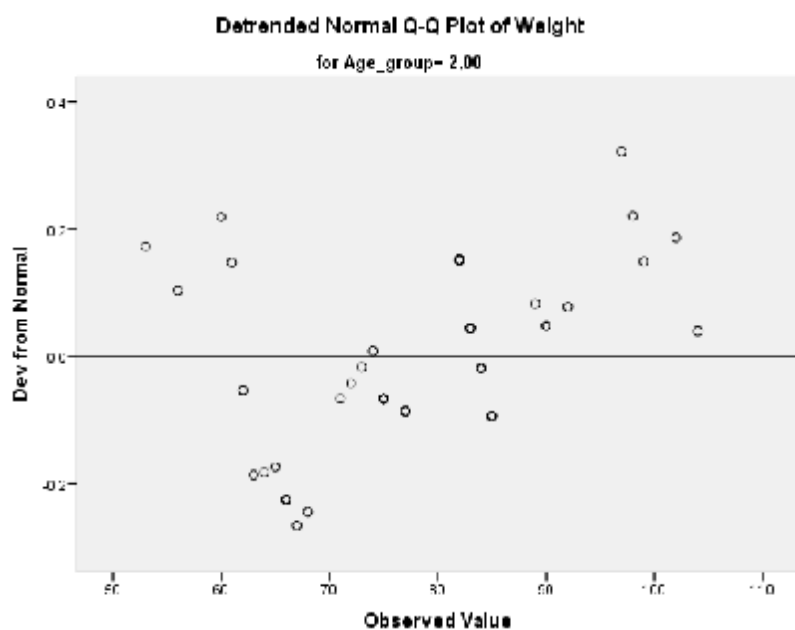
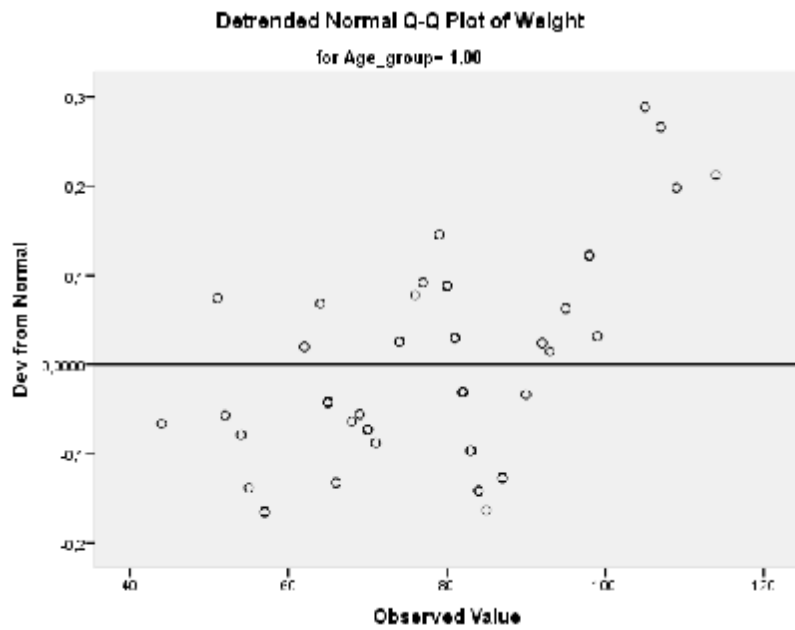
	Age_group	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Weight	<=45	,078	51	,200*	,986	51	,819
	>45	,096	49	,200*	,970	49	,233

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

a. Lilliefors Significance Correction







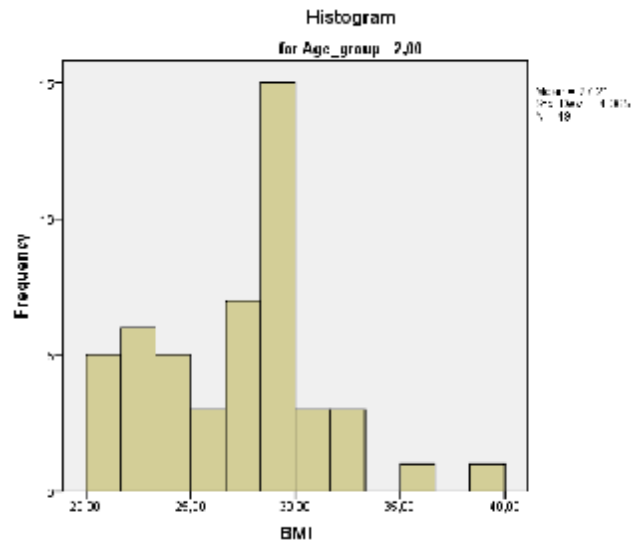
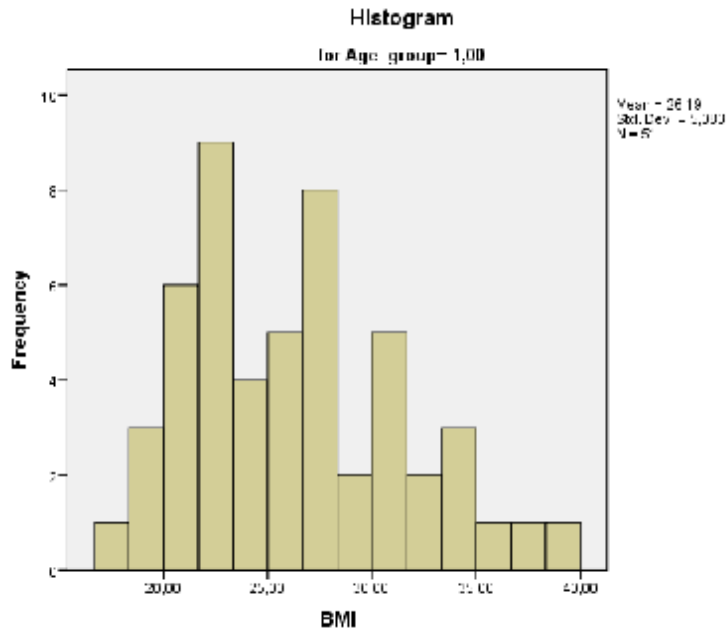
Tests of Normality

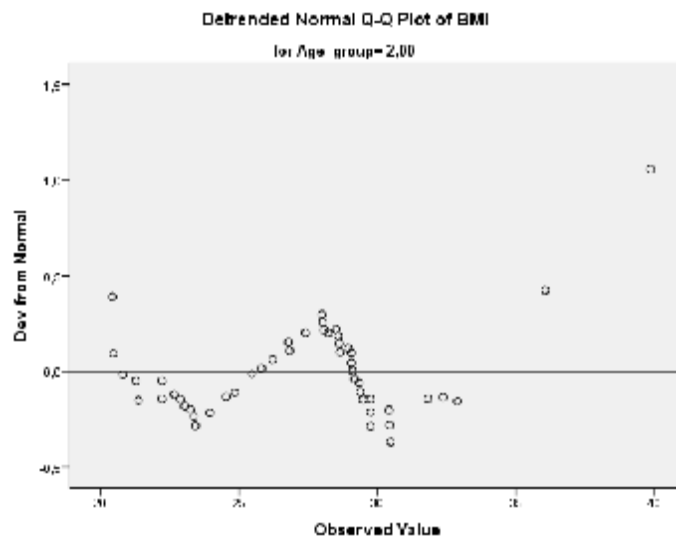
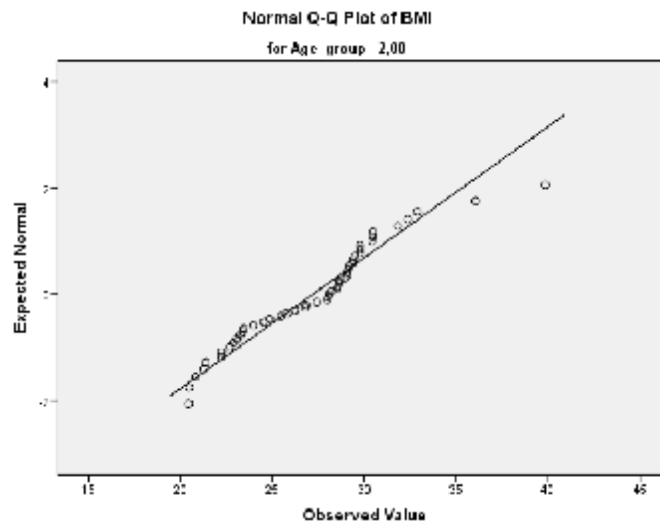
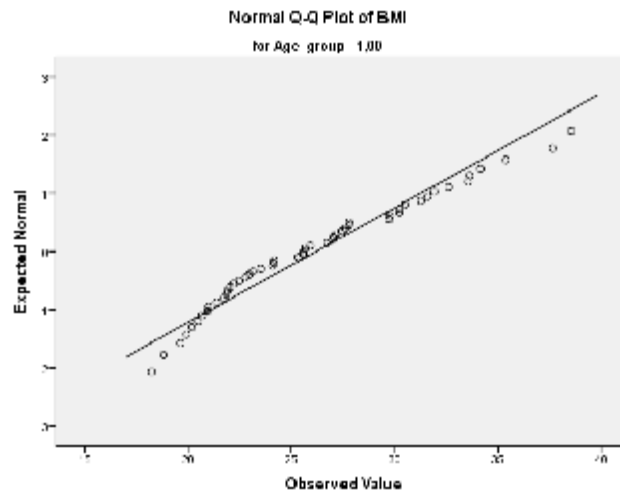
	Age_group	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.

BMI	1,00	,106	51	,200*	,955	51	,053
BMI	2,00	,129	49	,041	,947	49	,028

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

a. Lilliefors Significance Correction





ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΕΛΕΓΧΟΙ:

Ranks

	Sex	N	Mean Rank	Sum of Ranks
Weight	Male	55	63,35	3484,50
	Female	45	34,79	1565,50
	Total	100		
BMI	Male	55	55,68	3062,50
	Female	45	44,17	1987,50
	Total	100		

Test Statistics^a

	Weight	BMI
Mann-Whitney U	530,500	952,500
Wilcoxon W	1565,500	1987,500
Z	-4,901	-1,975
Asymp. Sig. (2-tailed)	,000	,048

a. Grouping Variable: Sex

Από τον παραπάνω πίνακα παρατηρούμε πως και για το βάρος αλλά και για το δείκτη μάζας το Sig. είναι μικρότερο από 0,05, οπότε σε επίπεδο σημαντικότητας 5%, οι τιμές των μεταβλητών αυτών διαφέρουν ανάμεσα στα δυο φύλα.

Ranks

	Age_grou p	N	Mean Rank	Sum of Ranks
Weight	<45	51	51,01	2601,50
	>45	49	49,97	2448,50

	Total	100		
	<45	51	46,57	2375,00
BMI	>45	49	54,59	2675,00
	Total	100		

Test Statistics^a

	Weight	BMI
Mann-Whitney U	1223,500	1049,000
Wilcoxon W	2448,500	2375,000
Z	-,179	-1,382
Asymp. Sig. (2-tailed)	,858	,167

a. Grouping Variable: Age_group

Αντίθετα με το φύλο η ηλικιακή κατηγορία δεν φαίνεται να επηρεάζει στατιστικά σημαντικά τις τιμές του βάρους και του δείκτη σώματος δηλαδή δεν υπάρχει διαφοροποίηση αυτών των μεταβλητών για άτομα μέχρι 45 ετών ή πάνω από αυτά.

Όπως και στο προηγούμενο ερώτημα έτσι και στο συγκεκριμένο προκειμένου να ελέγξουμε τα παρακάτω:

α) άτομα διαφορετικού επιπέδου εβδομαδιαίας σωματικής άσκησης έχουν παρόμοιο βάρος.

β) άτομα διαφορετικού επιπέδου εβδομαδιαίας σωματικής άσκησης έχουν παρόμοιο δείκτη μάζας σώματος.

γ) άτομα διαφορετικού τύπου εργασίας έχουν παρόμοιο βάρος.

δ) άτομα διαφορετικού τύπου εργασίας έχουν παρόμοιο δείκτη μάζας σώματος.

Πραγματοποιούμε παραμετρικούς και μη παραμετρικούς ελέγχους.

Στην πρώτη περίπτωση πριν από την εφαρμογή αυτών ελέγχουμε την ικανοποίηση υποθέσεων όπως η κανονικότητα, προκειμένου να εφαρμοστούν:

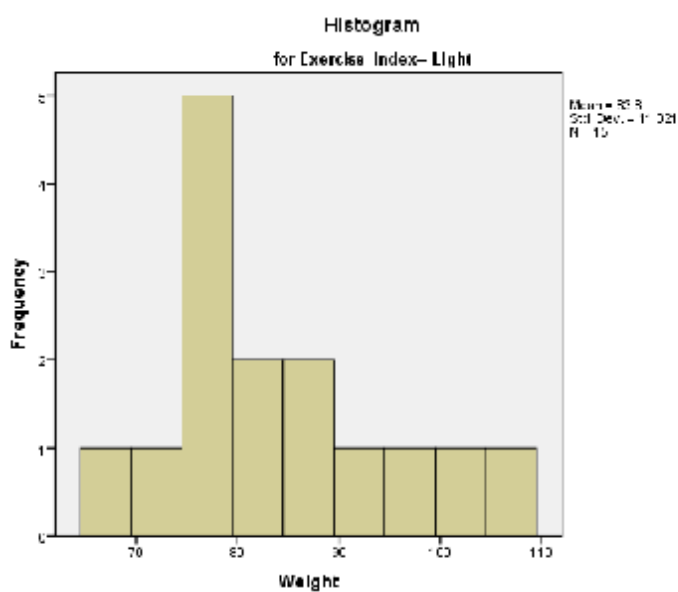
ΠΑΡΑΜΕΤΡΙΚΟΙ ΕΛΕΓΧΟΙ

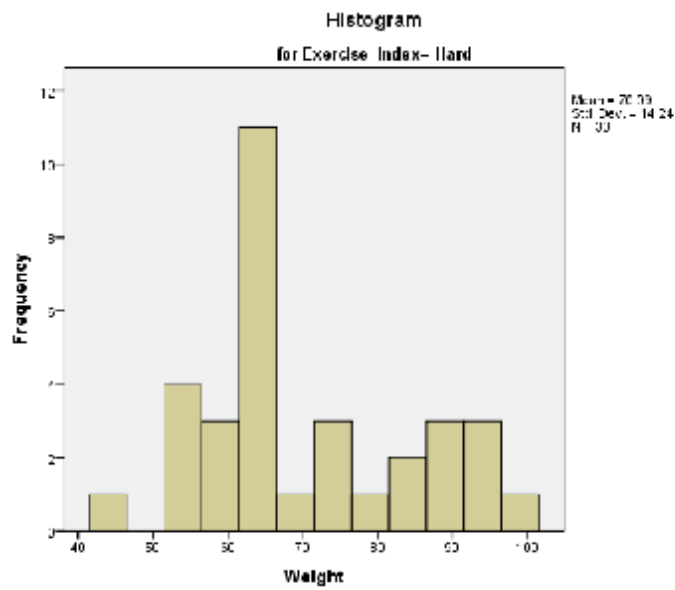
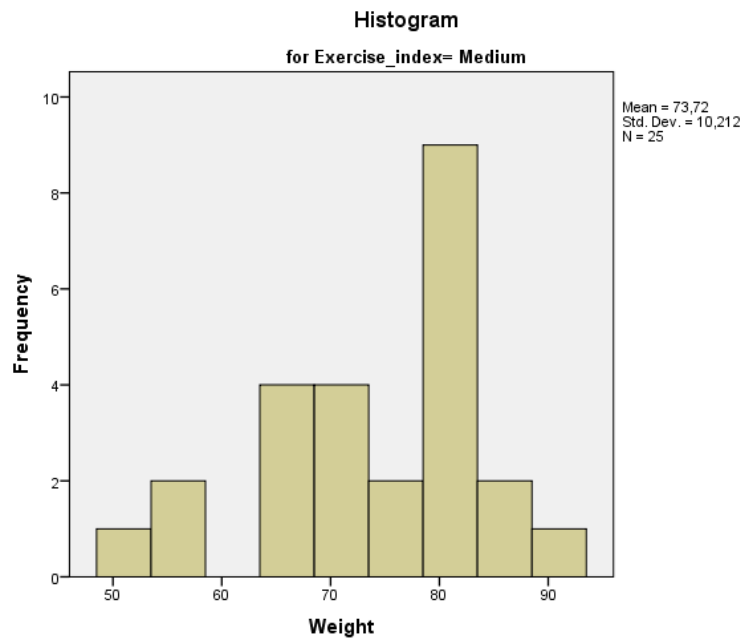
Tests of Normality

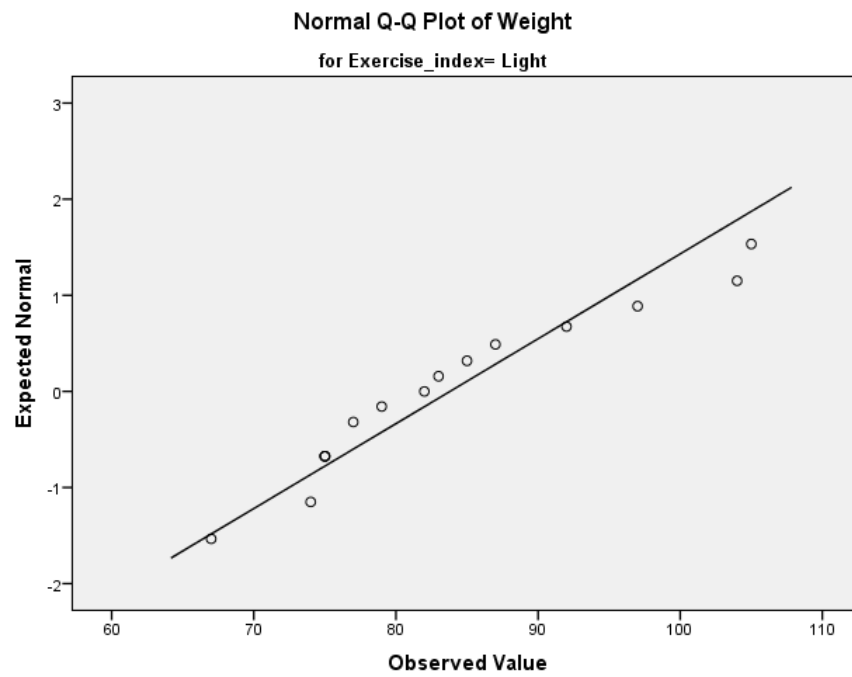
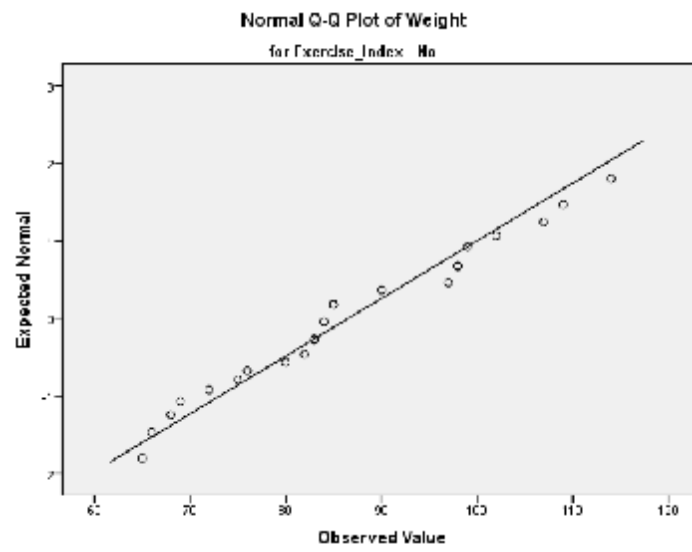
	Exercise_index	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	Df	Sig.	Statistic	df	Sig.
Weight	No	,176	27	,032	,958	27	,327
	Light	,131	15	,200*	,922	15	,206
	Medium	,177	25	,041	,928	25	,077
	Hard	,189	33	,004	,938	33	,060

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

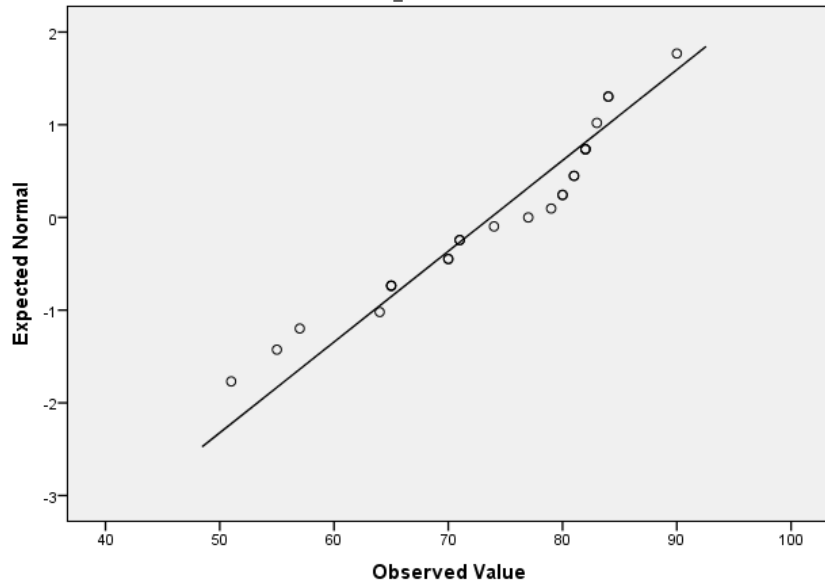
a. Lilliefors Significance Correction



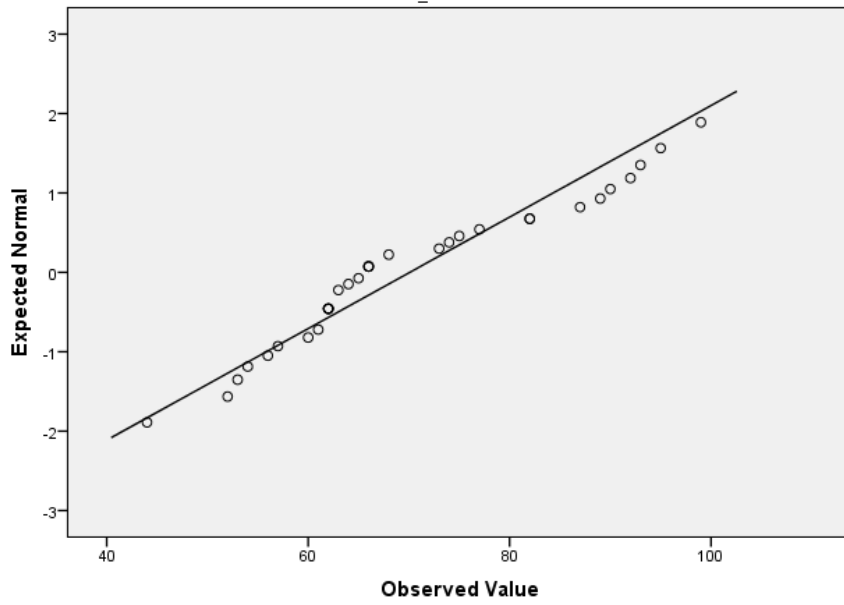


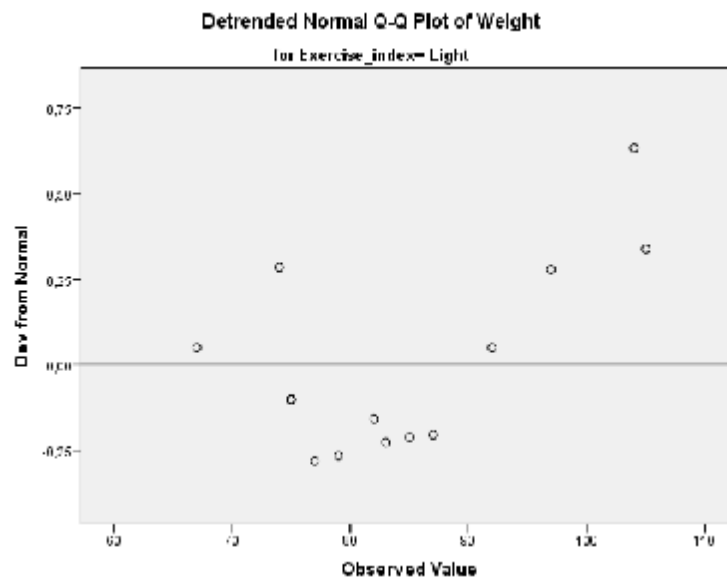
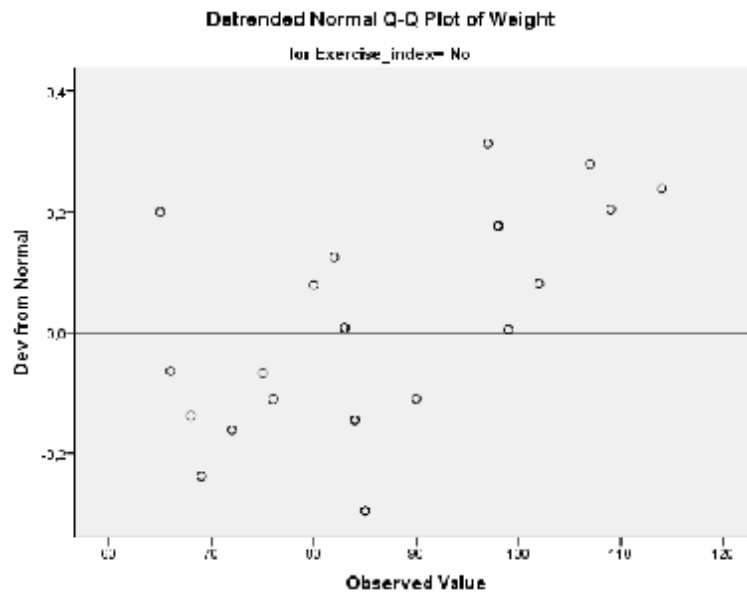


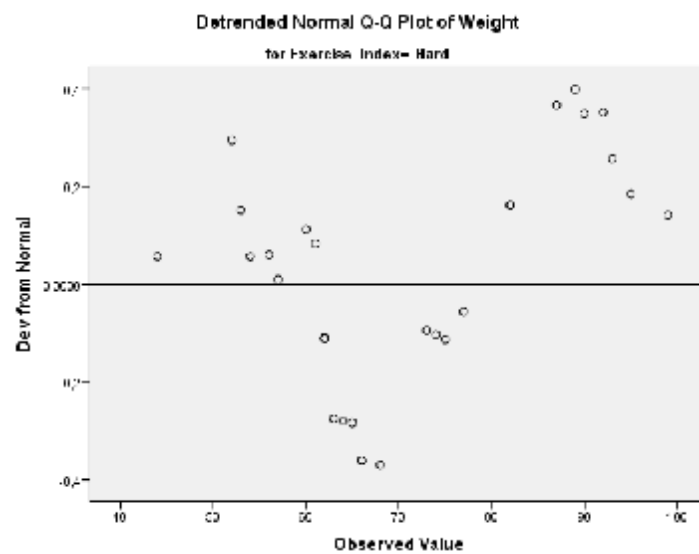
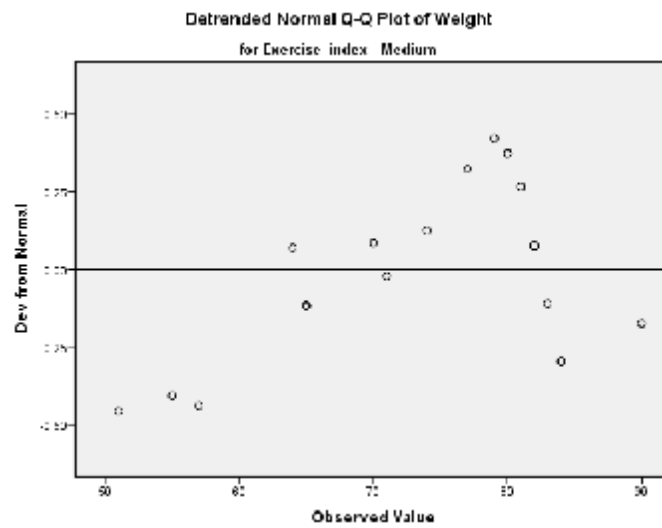
Normal Q-Q Plot of Weight
for Exercise_index= Medium



Normal Q-Q Plot of Weight
for Exercise_index= Hard





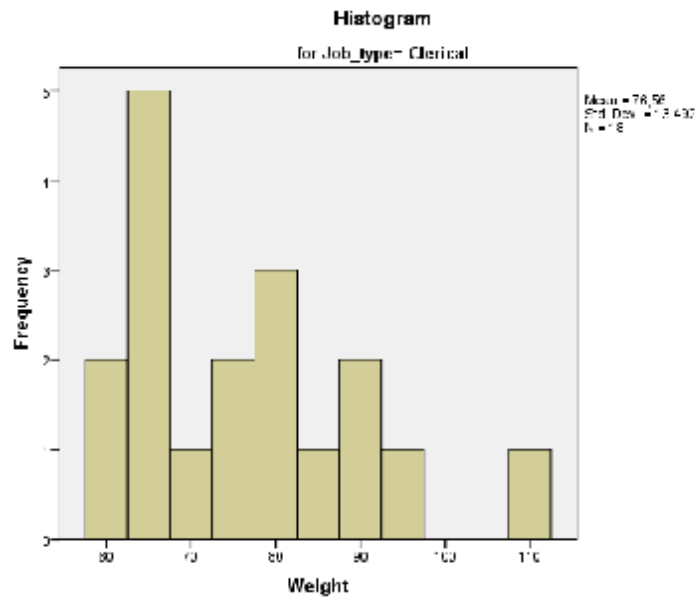
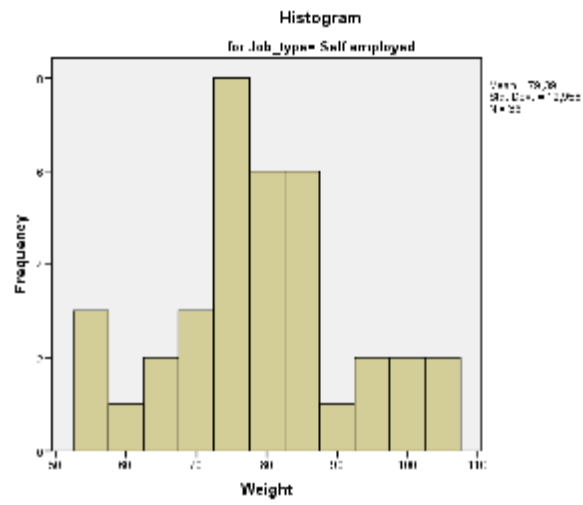


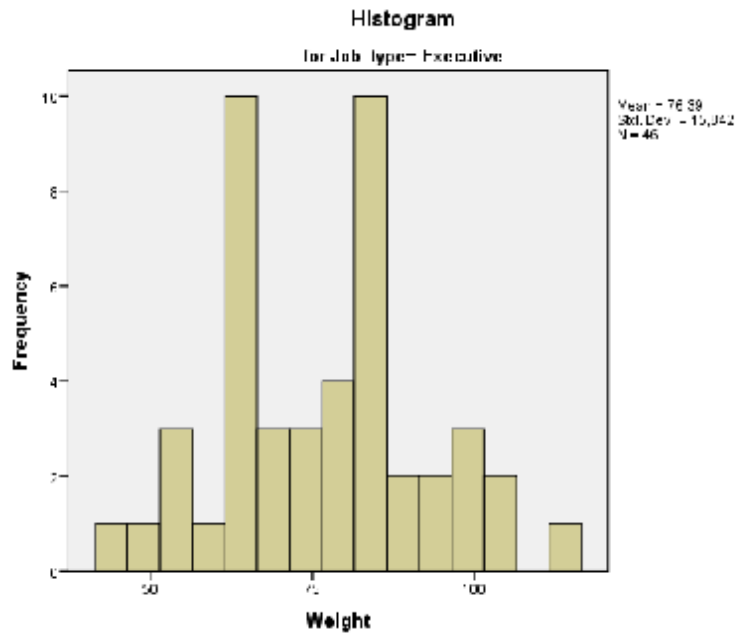
Tests of Normality

	Job_type	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Weight	Self-employed	,111	36	,200*	,975	36	,561
	Clerical	,172	18	,169	,926	18	,164
	Executive	,092	46	,200*	,984	46	,772

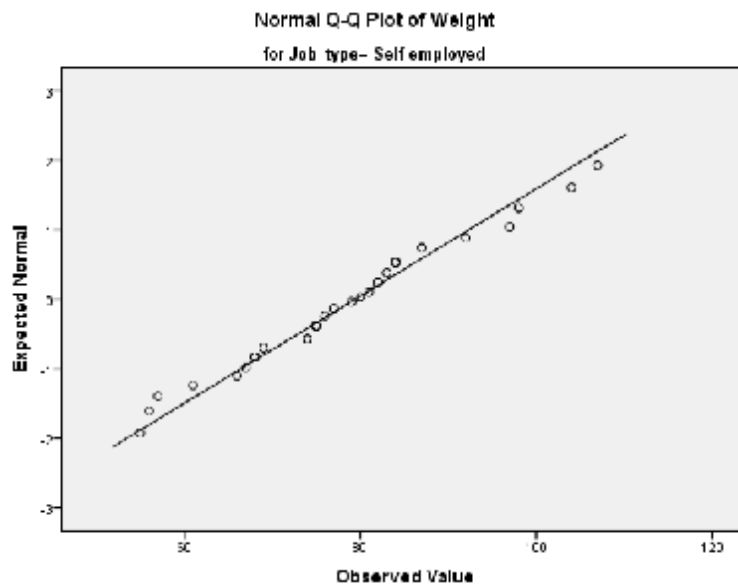
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction



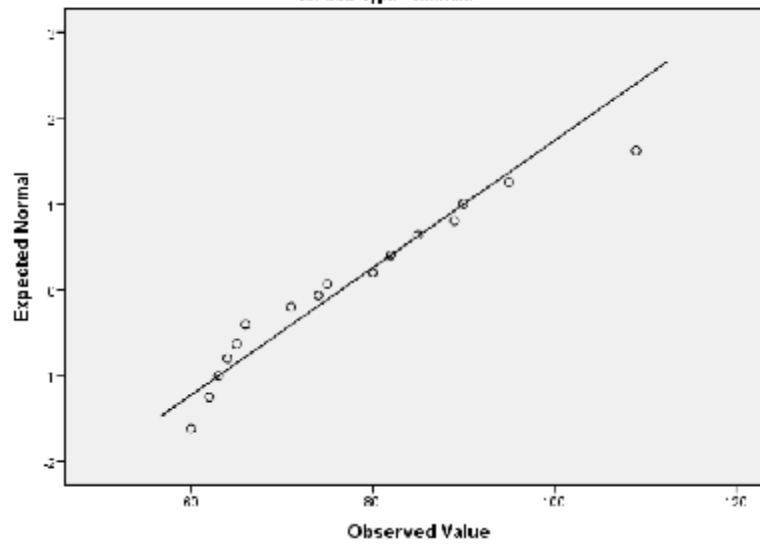


Normal Q-Q Plots



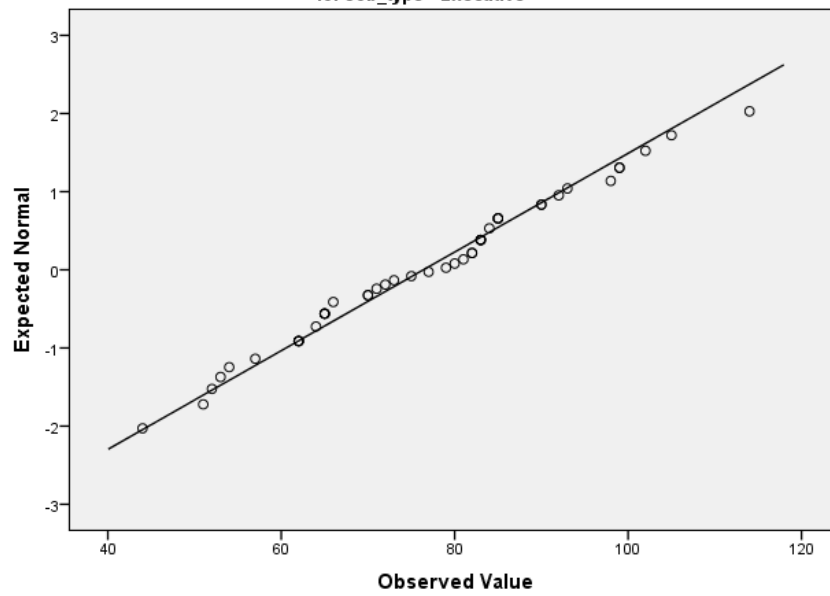
Normal Q-Q Plot of Weight

for Job_type= Clerical



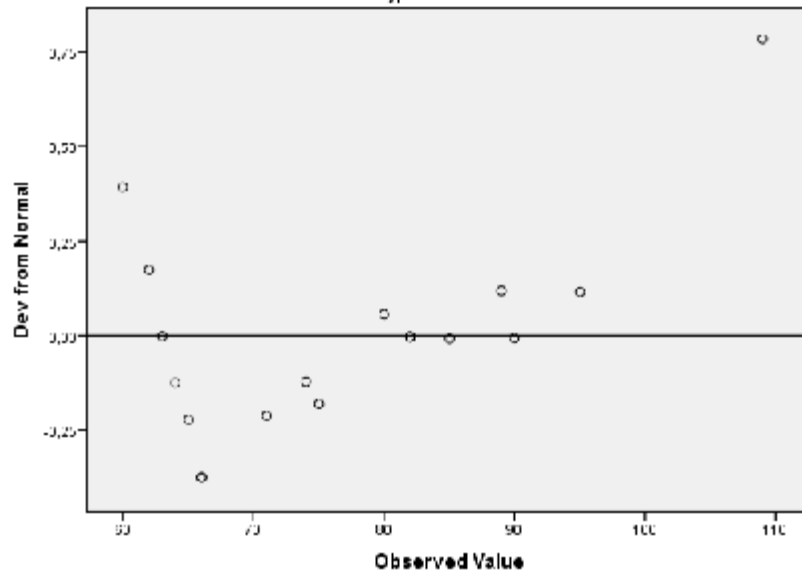
Normal Q-Q Plot of Weight

for Job_type= Executive



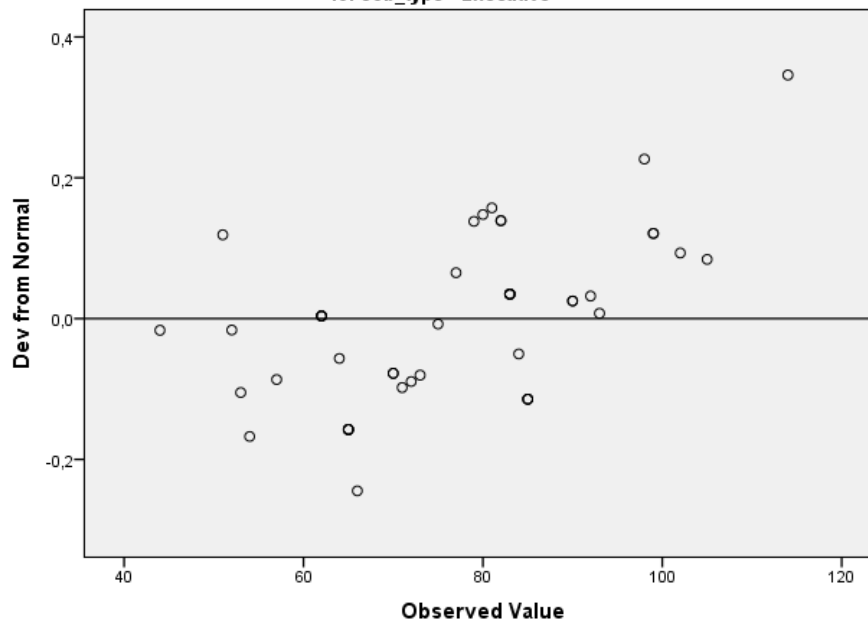
Detrended Normal Q-Q Plot of Weight

for Job_type= Clerical



Detrended Normal Q-Q Plot of Weight

for Job_type= Executive

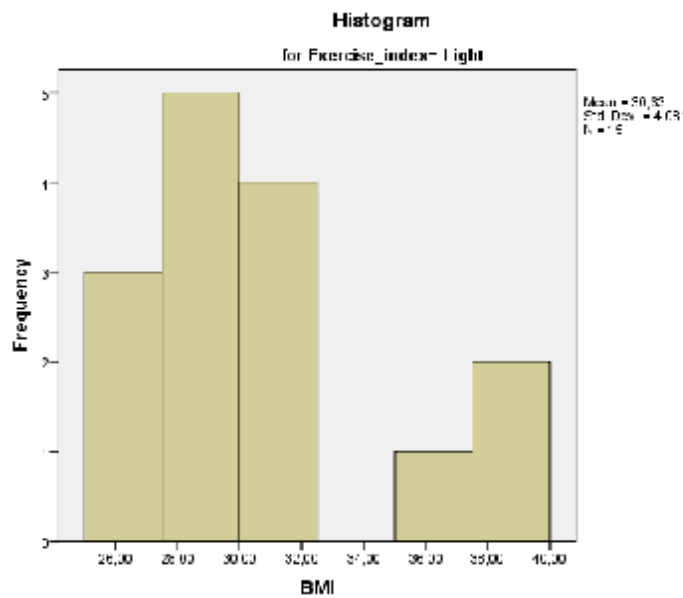


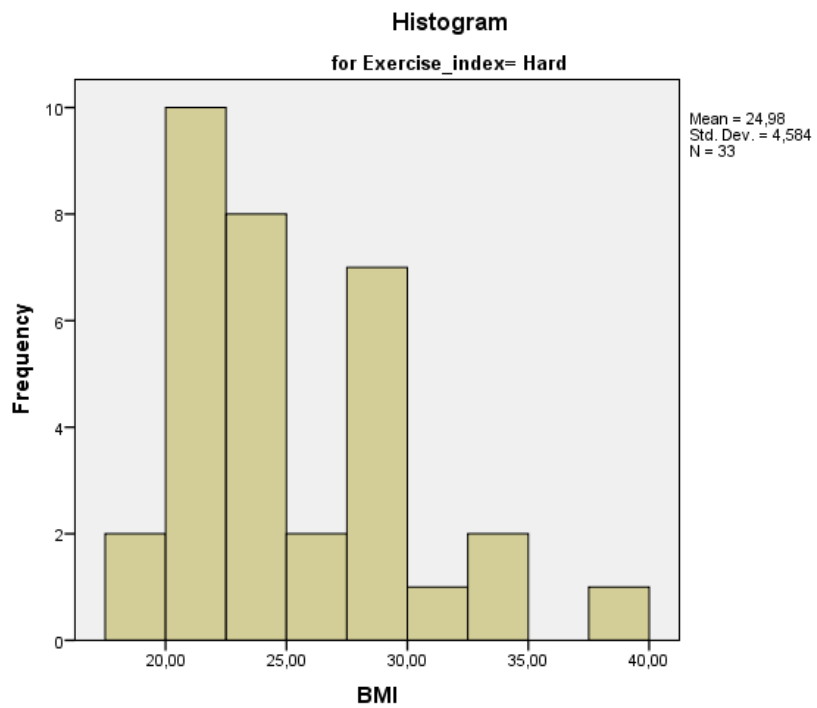
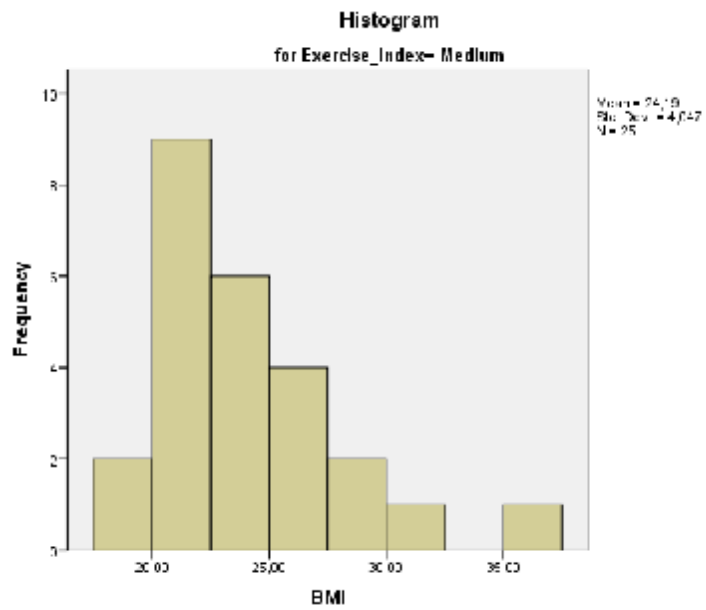
Tests of Normality

	Exercise_index	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
BMI	No	,090	27	,200*	,984	27	,934
	Light	,247	15	,014	,888	15	,062
	Medium	,145	25	,184	,882	25	,008
	Hard	,177	33	,010	,922	33	,020

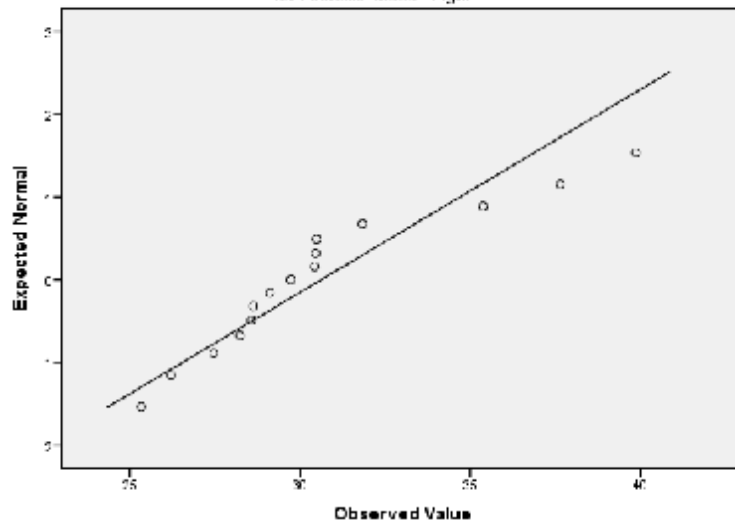
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

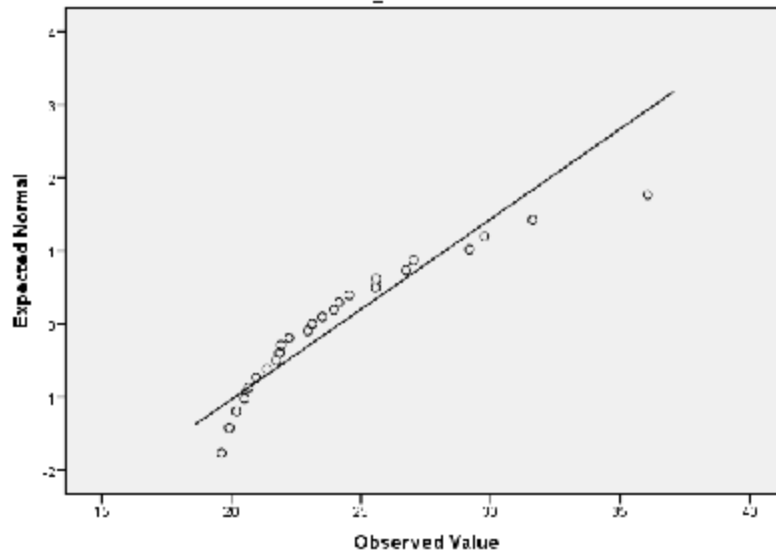


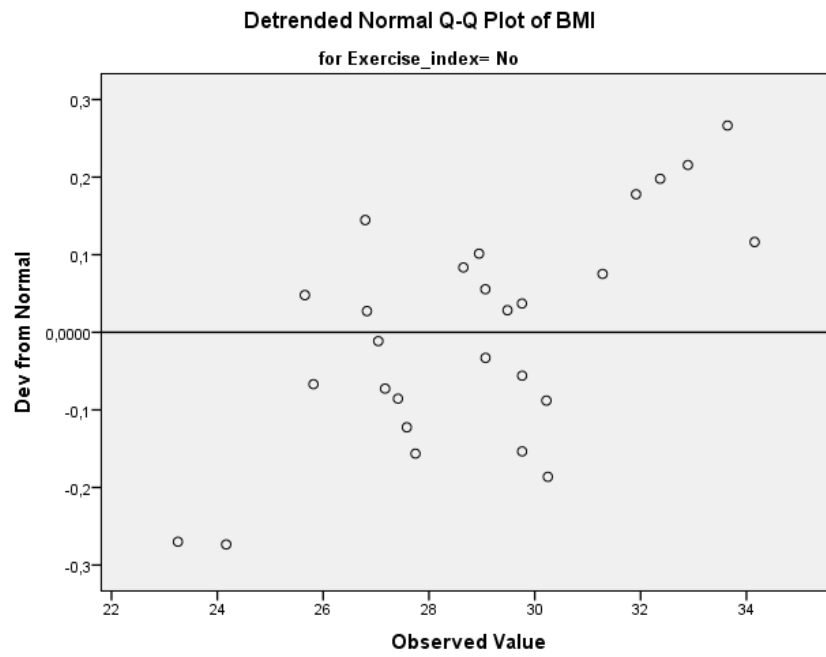
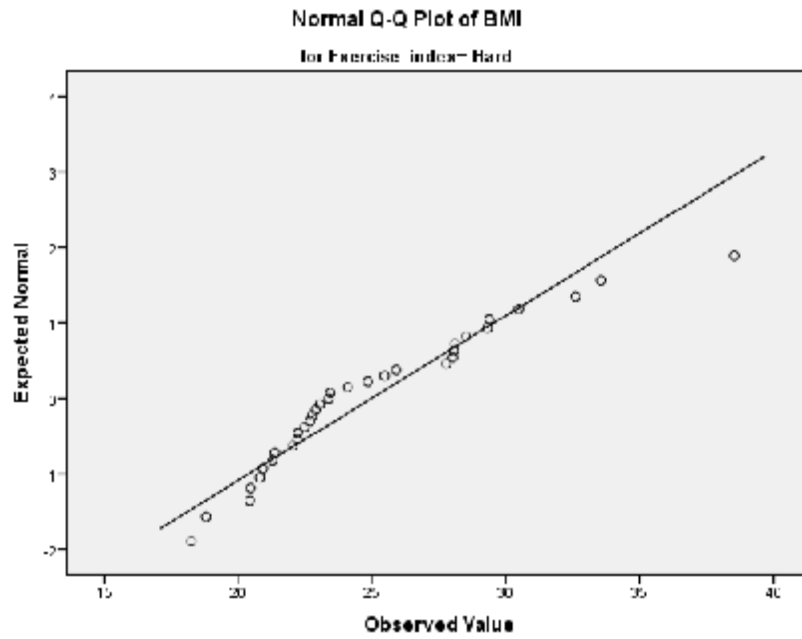


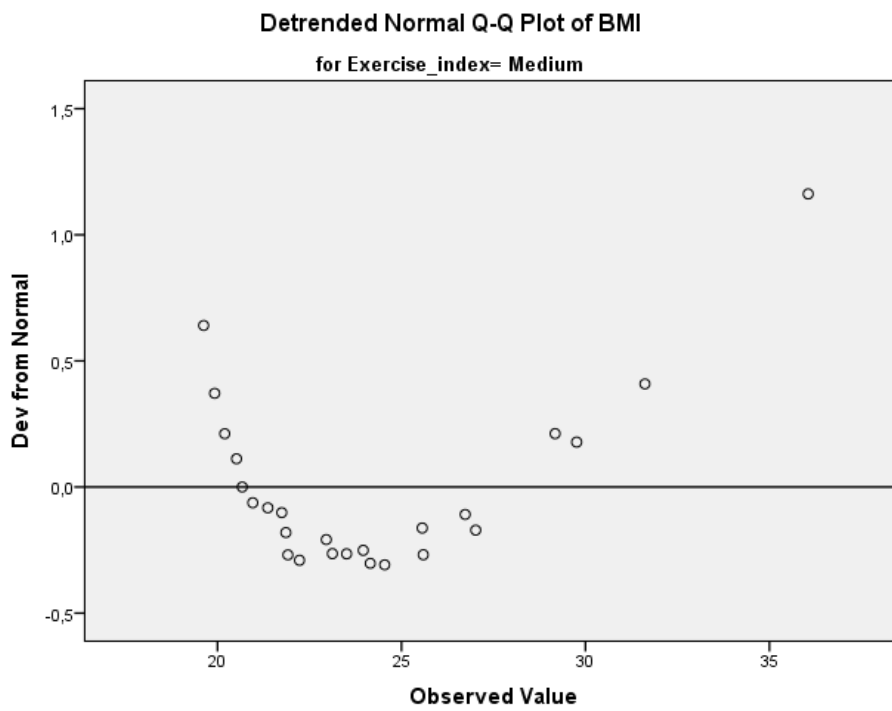
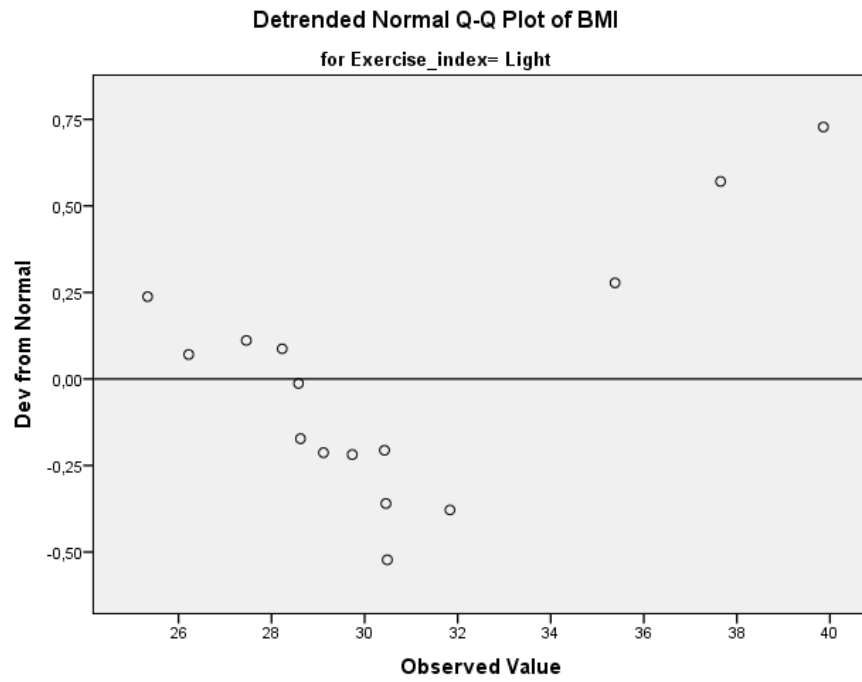
Normal Q-Q Plot of BMI
for Exercise_index = Light

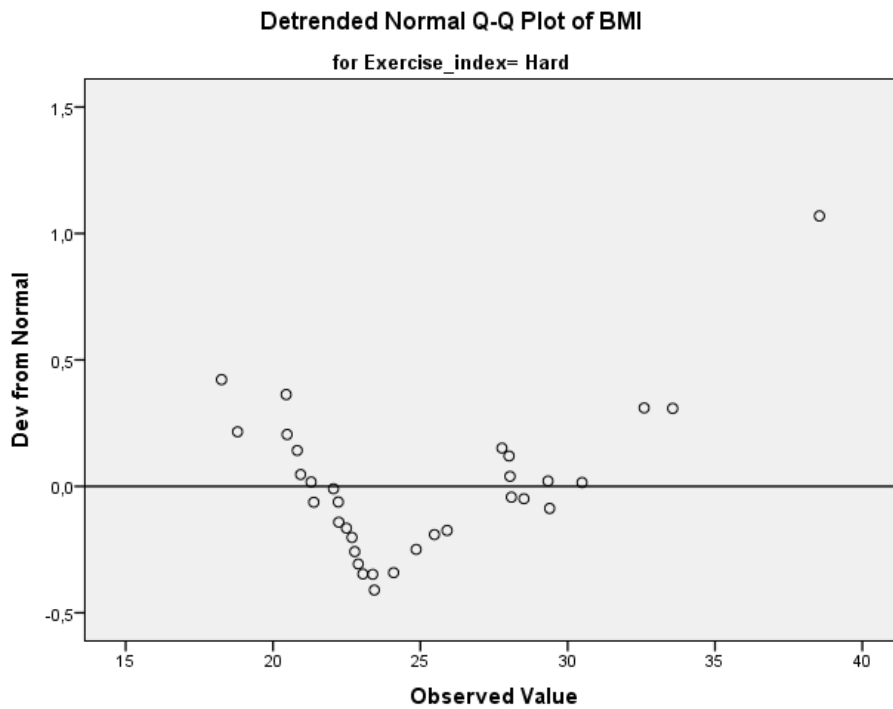


Normal Q-Q Plot of BMI
for Exercise_index = Medium







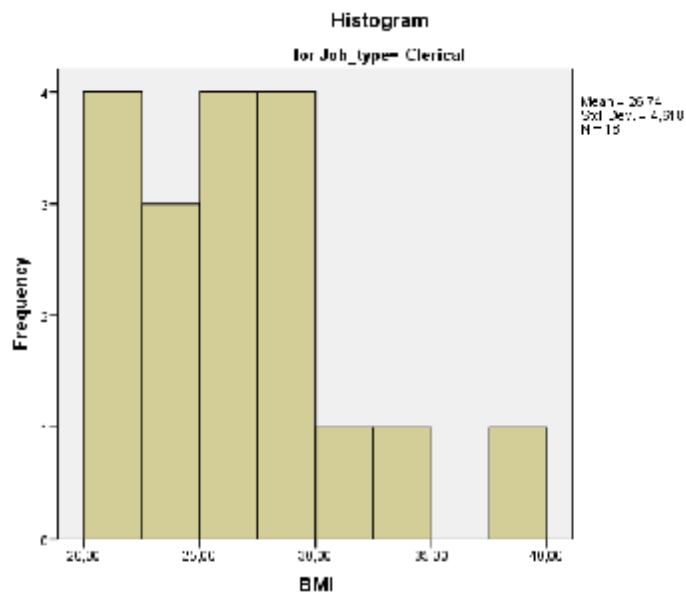
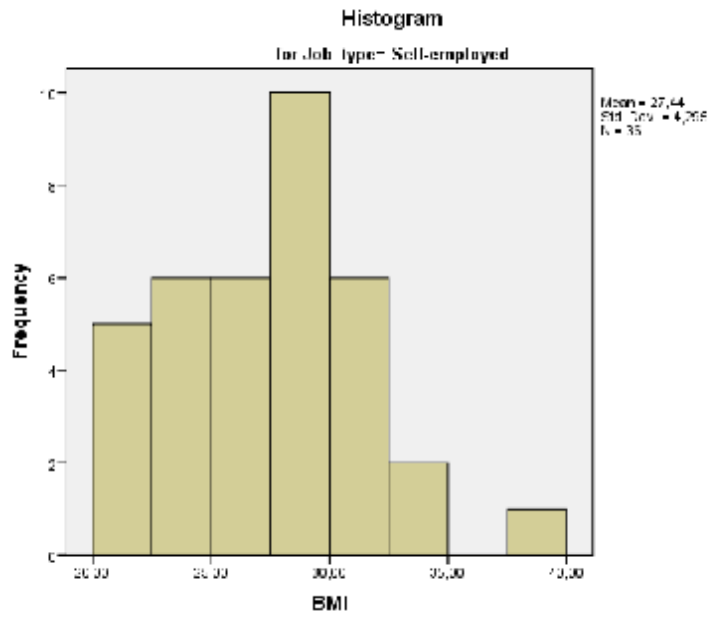


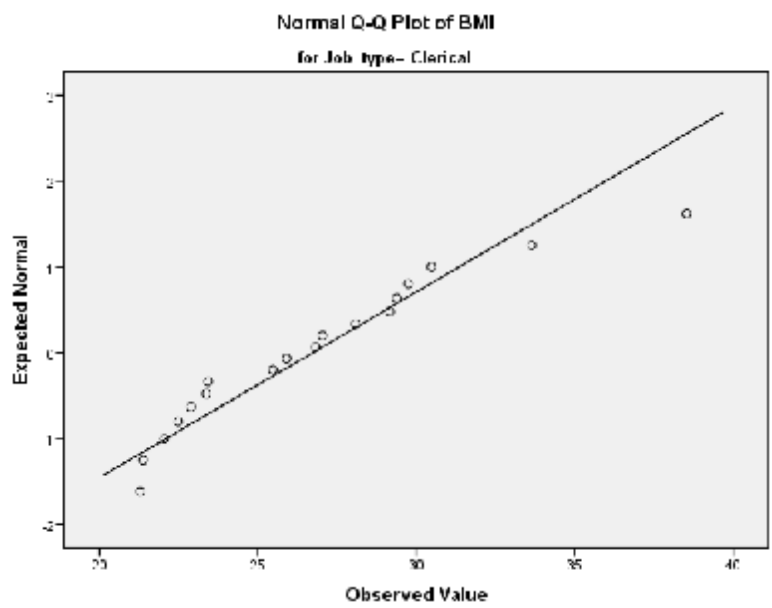
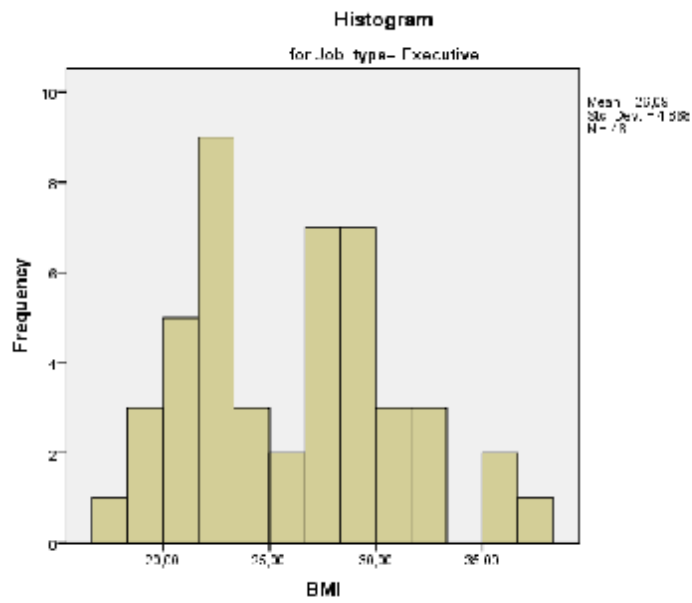
Tests of Normality

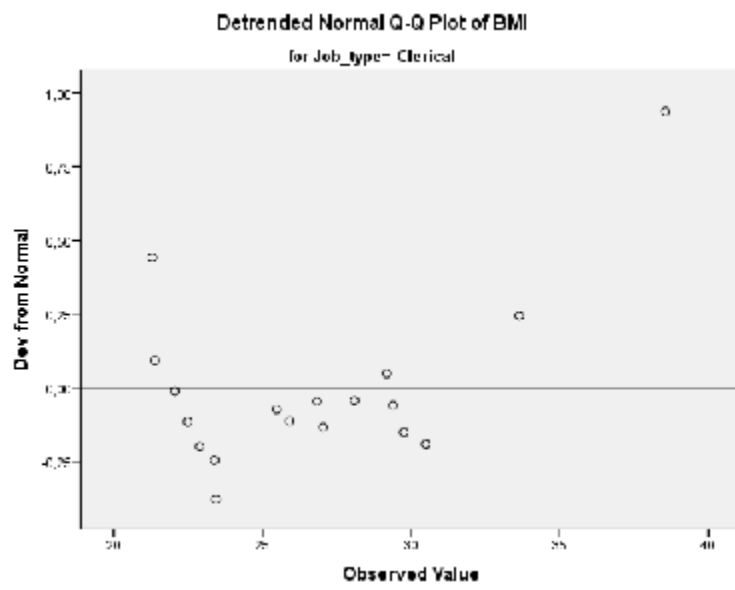
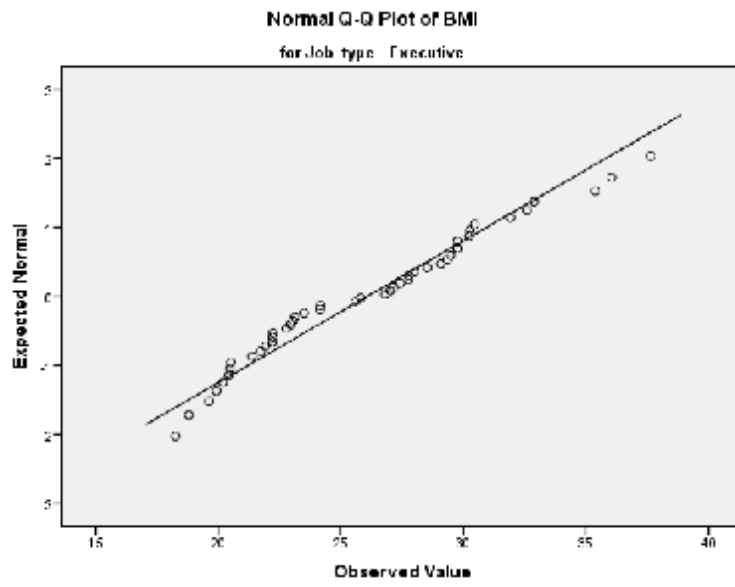
	Job_type	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
BMI	Self-employed	,070	36	,200*	,967	36	,338
	Clerical	,151	18	,200*	,918	18	,121
	Executive	,120	46	,093	,960	46	,115

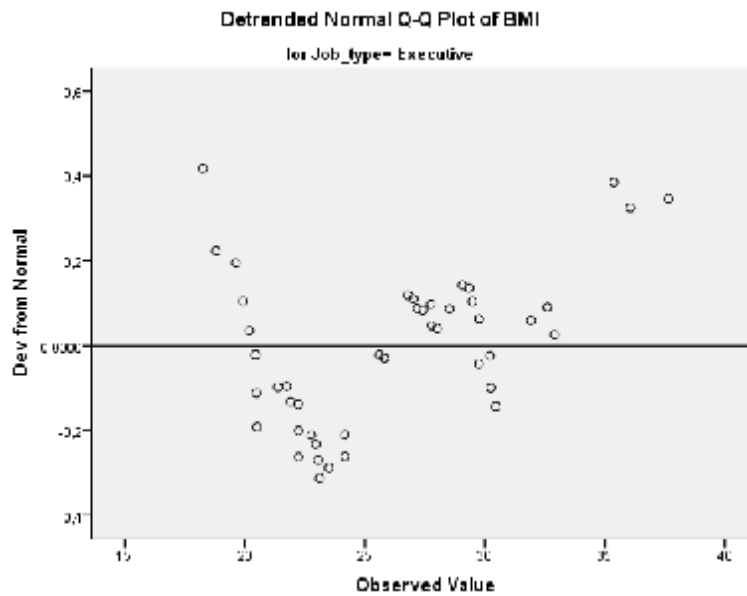
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction









ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΕΛΕΓΧΟΙ

Σε αντιστοιχία με τους παραπάνω παραμετρικούς ελέγχους τα ερωτήματα ελέγχονται και με τους ακόλουθους μη παραμετρικούς ελέγχους:

Ranks

	Exercise_inde x	N	Mean Rank
	No	27	68,65
	Light	15	63,30
Weight	Medium	25	43,02
	Hard	33	35,50
	Total	100	

Test Statistics^{a,b}

	Weight
Chi-Square	23,997

df	3
Asymp. Sig.	,000

a. Kruskal Wallis

Test

b. Grouping Variable:

Exercise_index

Το επίπεδο της σωματικής άσκησης διαφοροποιεί το βάρος των ατόμων ($0,000 < 0,05$)

Ranks

	Exercise_index	N	Mean Rank
BMI	No	27	67,06
	Light	15	73,97
	Medium	25	33,62
	Hard	33	39,08
	Total	100	

Test Statistics^{a,b}

	BMI
Chi-Square	32,188
df	3
Asymp. Sig.	,000

a. Kruskal Wallis

Test

b. Grouping Variable:

Exercise_index

Σε αντιστοιχία με το βάρος και ο δείκτης BMI διαφοροποιείται με κριτήριο την σωματική άσκηση ($0,000 < 0,05$)

Ranks

	Age_grou p	N	Mean Rank
Weight	<45	51	51,01
	>45	49	49,97
Total		100	

Test Statistics^{a,b}

	Weight
Chi-Square	,032
df	1
Asymp. Sig.	,858

a. Kruskal Wallis

Test

b. Grouping Variable:

Age_group

Αντίθετα με την σωματική άσκηση η ηλικιακή κατηγορία δεν διαφοροποιεί το βάρος του ατόμου στατιστικά σημαντικά ($0,858 > 0,05$)

Ranks

	Age_grou p	N	Mean Rank
BMI	<45	51	46,57
	>45	49	54,59

Total	100	
-------	-----	--

Test Statistics^{a,b}

	BMI
Chi-Square	1,911
df	1
Asymp. Sig.	,167

a. Kruskal Wallis Test

b. Grouping Variable:
Age_group

Όπως και στο βάρος έτσι και με το δείκτη BMI δεν παρατηρείται στατιστικά σημαντική διαφοροποίηση αυτού ανάλογα με την ηλικιακή κατηγορία στην οποία ανήκει το άτομο.

Προκειμένου να ελέγξουμε τα παρακάτω:

- α) «Obesity_index» και «Sex».
- β) «Obesity_index» και «Exercise index».
- γ) «Obesity_index» και «Job_type».
- δ) «Job_type» και «Exercise index».
- ε) «Job_type» και «Sex».
- στ) «Job_type» και «Age_group».
- ζ) «Exercise_index» και «Age_group».

Πραγματοποιούμε Crosstab ανάλυση χρησιμοποιώντας το κριτήριο Chi-Square, πιο συγκεκριμένα όταν η τιμή του Sig. αυτού είναι μικρότερη από 0,05 υπάρχει σχέση μεταξύ των υπό εξέταση μεταβλητών σε αντίθετη περίπτωση δεν υπάρχει.

Πιο αναλυτικά:

Obesity_index * Sex Crosstabulation

Count

		Sex		Total
		Male	Female	
Obesity_index	<25	17	22	39
	>=25=<30	24	16	40
	>30	14	7	21
Total		55	45	100

Chi-Square Tests

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	3,610 ^a	2	,164
Likelihood Ratio	3,631	2	,163
Linear-by-Linear Association	3,355	1	,067
N of Valid Cases	100		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 9,45.

Δεν υπάρχει σχέση μεταξύ «Obesity_index» και «Sex».

Crosstab

Count

		Exercise_index				Total
		No	Light	Medium	Hard	
Obesity_index	<25	2	0	17	20	39
	>=25=<30	17	8	6	9	40

>30	8	7	2	4	21
Total	27	15	25	33	100

Chi-Square Tests

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	38,177 ^a	6	,000
Likelihood Ratio	45,325	6	,000
Linear-by-Linear Association	20,980	1	,000
N of Valid Cases	100		

a. 1 cells (8,3%) have expected count less than 5. The minimum expected count is 3,15.

Υπάρχει στατιστικά σημαντική σχέση μεταξύ: «Obesity_index» και «Exercise index».

Crosstab

Count

	Job_type			Total
	Self-employed	Clerical	Executive	
<25	11	7	21	39
Obesity_index >=25=<30	16	8	16	40
>30	9	3	9	21
Total	36	18	46	100

Chi-Square Tests

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2,240 ^a	4	,692
Likelihood Ratio	2,269	4	,686

Linear-by-Linear Association	1,448	1	,229
N of Valid Cases	100		

a. 1 cells (11,1%) have expected count less than 5. The minimum expected count is 3,78.

Δεν υπάρχει στατιστικά σημαντική σχέση μεταξύ «Obesity_index» και «Job_type».

Crosstab

Count

		Exercise_index				Total
		No	Light	Medium	Hard	
Job_type	Self-employed	9	12	8	7	36
	Clerical	4	0	1	13	18
	Executive	14	3	16	13	46
Total		27	15	25	33	100

Chi-Square Tests

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	28,681 ^a	6	,000
Likelihood Ratio	29,180	6	,000
Linear-by-Linear Association	,632	1	,427
N of Valid Cases	100		

a. 3 cells (25,0%) have expected count less than 5. The minimum expected count is 2,70.

Υπάρχει στατιστικά σημαντική σχέση μεταξύ «Job_type» και «Exercise index».

Crosstab

Count

		Sex		Total
		Male	Female	
Job_type	Self-employed	19	17	36
	Clerical	10	8	18
	Executive	26	20	46
Total		55	45	100

Chi-Square Tests

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	,117 ^a	2	,943
Likelihood Ratio	,117	2	,943
Linear-by-Linear Association	,111	1	,739
N of Valid Cases	100		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 8,10.

Δεν υπάρχει στατιστικά σημαντική σχέση μεταξύ «Job_type» και «Sex».

Crosstab

Count

		Age_group		Total
		<45	>45	
Job_type	Self-employed	16	20	36
	Clerical	7	11	18

	Executive	28	18	46
Total		51	49	100

Chi-Square Tests

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	3,469 ^a	2	,177
Likelihood Ratio	3,493	2	,174
Linear-by-Linear Association	2,329	1	,127
N of Valid Cases	100		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 8,82.

Δεν υπάρχει στατιστικά σημαντική σχέση μεταξύ «Job_type» και «Age_group».

Exercise_index * Age_group Crosstabulation

Count

		Age_group		Total
		<45	>45	
Exercise_index	No	12	15	27
	Light	5	10	15
	Medium	20	5	25
	Hard	14	19	33
Total		51	49	100

Chi-Square Tests

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	11,722 ^a	3	,008
Likelihood Ratio	12,391	3	,006
Linear-by-Linear Association	,313	1	,576
N of Valid Cases	100		

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 7,35.

Υπάρχει στατιστικά σημαντική σχέση μεταξύ:
«Exercise_index» και «Age_group».

Model Summary^{b,c}

Model	R		R Square	Adjusted R Square	Std. Error of the Estimate
	Sex = Male (Selected)	Sex ~= Male (Unselected)			
1	,766 ^a	,838	,587	,580	2,73815

a. Predictors: (Constant), Weight

b. Unless noted otherwise, statistics are based only on cases for which Sex = Male.

c. Dependent Variable: BMI

ANOVA^{a,b}

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	565,769	1	565,769	75,461	,000 ^c
	Residual	397,367	53	7,497		

Total	963,136	54			
-------	---------	----	--	--	--

a. Dependent Variable: BMI

b. Selecting only cases for which Sex = Male

c. Predictors: (Constant), Weight

Coefficients^{a,b}

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
1 (Constant)	6,372	2,446		2,605	,012	1,466	11,278
Weight	,251	,029	,766	8,687	,000	,193	,309

a. Dependent Variable: BMI

b. Selecting only cases for which Sex = Male

Το γραμμικό υπόδειγμα το οποίο να δίνει δυνατότητα πρόβλεψης του Δείκτη μάζας σώματος ενός Άνδρα γνωρίζοντας το βάρος του, με βάση τα διαθέσιμα δεδομένα είναι το εξής:

$$\text{BMI} = 6,372 + 0,251 * \text{Weight}$$

Το μοντέλο είναι στατιστικά σημαντικό, γεγονός που προκύπτει είτε από έλεγχο t για τη στατιστική σημαντικότητα του β1 είτε από έλεγχο F για τη στατιστική σημαντικότητα του μοντέλου.

Το βάρος ενός άνδρα ερμηνεύει το 58,6% της συνολικής μεταβλητότητας του δείκτη BMI

Model Summary^{b,c}

Model	R		R Square	Adjusted R Square	Std. Error of the Estimate
	Sex = Female (Selected)	Sex ~ = Female (Unselected)			
1	,838 ^a	,766	,703	,696	2,74641

a. Predictors: (Constant), Weight

b. Unless noted otherwise, statistics are based only on cases for which Sex = Female.

c. Dependent Variable: BMI

ANOVA^{a,b}

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	766,583	1	766,583	101,631	,000 ^c
	Residual	324,340	43	7,543		
	Total	1090,923	44			

a. Dependent Variable: BMI

b. Selecting only cases for which Sex = Female

c. Predictors: (Constant), Weight

Coefficients^{a,b}

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	2,275	2,375		,958	,343	-2,514	7,064
	Weight	,337	,033	,838	10,081	,000	,270	,405

a. Dependent Variable: BMI

b. Selecting only cases for which Sex = Female

Το γραμμικό υπόδειγμα το οποίο να δίνει δυνατότητα πρόβλεψης του Δείκτη μάζας σώματος μιας Γυναίκας γνωρίζοντας το βάρος της, με βάση τα διαθέσιμα δεδομένα είναι το εξής:

$$\text{BMI} = 2,275 + 0,337 * \text{Weight}$$

Το μοντέλο είναι στατιστικά σημαντικό, γεγονός που προκύπτει είτε από έλεγχο t για τη στατιστική σημαντικότητα του β1 είτε από έλεγχο F για τη στατιστική σημαντικότητα του μοντέλου.

Το βάρος μιας γυναίκας ερμηνεύει το 70,3% της συνολικής μεταβλητότητας του δείκτη BMI.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Από όλα όσα προηγήθηκαν προκύπτει ότι η στατιστική είναι η επιστήμη, η «τέχνη» του να μαθαίνουμε από τα δεδομένα. Η στατιστική συνίσταται στη συλλογή δεδομένων, στην περιγραφή τους και κυρίως στην ανάλυση τους που οδηγεί και στην απόκτηση συμπερασμάτων. Η Στατιστική ανάλυση- μέσα από την εξιδεικευμένη καταγραφή και ανάλυση των δεδομένων αποτελούν χρησιμότητα εργαλείο για την εξαγωγή ασφαλών συμπερασμάτων σχετικά με διάφορα οικονομικά μεγέθη και μεγέθη διοίκησης επιχειρήσεων. Η στατιστική ανάλυση δεδομένων αποτελεί εργαλείο χρήσιμο για πληθώρα επιστημών (κοινωνιολογία, οικονομία, διοίκηση, μάρκετινγκ, θεωρία καταναλωτή κ.ά.).

Η εμπειρική έρευνα ανήκει σαφώς σε μια από αυτές τις δύο γενικές κατηγορίες. Στη συσχετιστική έρευνα προσπαθούμε να μην επηρεάζουμε τις μεταβλητές αλλά μόνο τις μετράμε και ψάχνουμε τις σχέσεις (συσχετισμούς) μεταξύ κάποιου συνόλου μεταβλητών, όπως η πίεση αίματος και το επίπεδο χοληστερόλης.

Στην πειραματική έρευνα, χειριζόμαστε μερικές μεταβλητές και μετράμε έπειτα τα αποτελέσματα αυτού του χειρισμού σε άλλες μεταβλητές, παραδείγματος χάριν, ένας ερευνητής τεχνητά να αυξήσει την πίεση αίματος και να καταγράψει έπειτα το επίπεδο χοληστερόλης. Η ανάλυση στοιχείων στην πειραματική έρευνα οδηγεί στον υπολογισμό των συσχετισμών μεταξύ των μεταβλητών, συγκεκριμένα, εκείνων που χειρίζονται και εκείνων που επηρεάζονται από το χειρισμό. Εντούτοις, τα πειραματικά στοιχεία μπορούν ενδεχομένως να παρέχουν τις ποιοτικά καλύτερες πληροφορίες: Μόνο τα πειραματικά στοιχεία μπορούν αποφασιστικά να καταδείξουν τις αιτιώδεις σχέσεις μεταξύ των μεταβλητών. Παραδείγματος χάριν, εάν διαπιστώσαμε ότι όποτε αλλάζουμε τις μεταβλητές A οι μεταβλητές B αλλάζουν, κατόπιν μπορούμε να καταλήξουμε στο συμπέρασμα ότι το A επηρεάζει το B. Τα στοιχεία από τη συσχετιστική έρευνα μπορούν μόνο να ερμηνεύσουν τους αιτιώδεις όρους βασισμένους σε μερικές θεωρίες που έχουμε, αλλά τα συσχετιστικά στοιχεία δεν μπορούν αποφασιστικά να αποδείξουν την αιτιότητα.

Οι ανεξάρτητες μεταβλητές είναι εκείνες που επεξεργαζόμαστε ενώ οι εξαρτημένες μεταβλητές μετριούνται μόνο ή καταχωρούνται. Αυτή η διάκριση εμφανίζεται ως ιδιόζουσα αφού

δημιουργεί σύγχυση σε πολλούς επειδή, όπως μερικοί φοιτητές λένε, όλες οι μεταβλητές εξαρτώνται από κάτι. Εντούτοις, μόλις κατακτήσετε αυτήν την διάκριση θα τη θεωρείτε αυτονόητη. Η ορολογία εξαρτημένη και ανεξάρτητη μεταβλητή ισχύει συνήθως για την πειραματική έρευνα όπου μερικές μεταβλητές επιδέχονται επεξεργασία, και από αυτή την άποψη είναι ανεξάρτητες κατά τα αρχικά σχέδια αντίδρασης, τα χαρακτηριστικά γνωρίσματα, τις προθέσεις, κ.λπ. των θεμάτων. Μερικές άλλες μεταβλητές αναμένονται να θεωρηθούν εξαρτημένες από την επεξεργασία τους χειρισμό ή τις πειραματικές συνθήκες. Δηλαδή, εξαρτώνται από την απάντηση στην ερώτηση "τι θα κάνει το υποκείμενο".

Κάπως αντίθετα προς τη φύση αυτής της διάκρισης, αυτοί οι όροι χρησιμοποιούνται επίσης στις μελέτες όπου δεν χειριζόμαστε κυριολεκτικά τις ανεξάρτητες μεταβλητές, αλλά μόνο ορίζουν ιδιότητες στις πειραματικές ομάδες που ορίζονται από μερικές προϋπάρχουσες ιδιότητες των θεμάτων. Παραδείγματος χάριν, εάν σε ένα πείραμα, τα αρσενικά συγκρίνονται με τα θηλυκά σχετικά με τα χρωμοσώματα τους το φύλος θα μπορούσε να κληθεί ως ανεξάρτητα μεταβλητή και το χρωμόσωμα ως εξαρτημένη μεταβλητή.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ανδρικόπουλος, Α. (2000). Οικονομετρία. Θεωρία και Εμπειρικές Εφαρμογές. Αθήνα, Εκδόσεις Μπένου.

Δημητριάδης, Ε. (2012). ΣΤΑΤΙΣΤΙΚΗ ΕΠΙΧΕΙΡΗΣΕΩΝ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΕ SPSS ΚΑΙ LISREL, ΚΡΙΤΙΚΗ, ΑΘΗΝΑ.

Δημητριάδης Ε., (2007) “ Στατιστικές Εφαρμογές με S.P.S.S ” , Εκδόσεις Κριτική Α.Ε.

Ζαΐρης, Π. Ε. (2010). Στατιστική Μεθοδολογία, ΚΡΙΤΙΚΗ, ΑΘΗΝΑ.

Ζαΐρης Ποσειδών, (2005) “ Στατιστική Μεθοδολογία ” , Τόμος Α, Εκδόσεις Κριτική

Ιωαννίδης, Δ. (2011). Στατιστική μεθοδολογία , Εκδόσεις ΖΗΤΗ, Θεσσαλονίκη.

Κάτος Α. Β. (2004). Οικονομετρία: Θεωρία και Εφαρμογές. Θεσσαλονίκη, Ζυγός.

Μουσιάδης Π., Μπόρα Ε. (2002). Εφαρμοσμένη Στατιστική. Θεσσαλονίκη, ΑΠΘ.

Οικονόμου Γ., & Γεωργίου Α., (2000) “ Ποσοτική Ανάλυση για Λήψη Διοικητικών Αποφάσεων”, Τόμοι Α' & Β', Εκδόσεις Μπένου

Παπαϊωάννου Τ. & Λουκάς Σ. (2002) “ Εισαγωγή στη Στατιστική ” , Εκδόσεις Σταμούλη

Σφακιανάκης Μ., (2002) “ Προσομοίωση και Εφαρμογές ” , Εκδόσεις Πατάκη

Χαλκιάς, Ι. (2009). Στατιστική: Μέθοδοι Ανάλυσης για Επιχειρηματικές Αποφάσεις, Εκδόσεις Rosili

Χατζηνικολάου, Δ. (2002). Στατιστική για Οικονομολόγους. Ιωάννινα.

Brooks C. (2002). Introductory Econometrics for Finance, Cambridge University Press.

Doane, D., Seward, L. (2012). Applied Statistics in Business & Economics, Fourth Edition, McGraw- Hill International Edition

Gujarati D. N. (2003). Basic Econometrics, New York, Mc Grow-Hill.

Krzanowski W. (1998). *An Introduction to Statistical Modelling*, New York, Oxford University Press.

Maddala G.S. (1992). *Introductory Econometrics*, New Jersey, Prentice-Hall.

Myers R. (1990). *Classical and Modern Regression with Applications*, Belmont California, Duxbury Press.

Myers R., Montgomery D., Vining G. (2002). *Generalized Linear Models*, New York, John Wiley.

Thomas R.L. (1997). *Modern Econometrics: An Introduction*. Harlow, Addison-Wesley.

Αρθρογραφία

Bera, A.K. and Jarque, C.M. (1981). An efficient large-sample test for normality of observations and regression residuals, *Working Papers in Econometrics*, 40. Canberra, Australian National University.

Chow G.C. (1960). Test of equality between sets of coefficients in two linear regressions, *Econometrica*, Vol. 28, No 3, pp. 591 – 605.

Dritsakis N. and S. Athanasiadis (2000). An econometric model of tourist demand, The case of Greece. *Journal of Hospitality and Leisure Marketing*, Vol. 7, No 2, pp. 39 - 49.

Dritsakis N. (2002). Productivity of the health care sector in Greece, *Social Science Tribune*, Vol. 33, pp. 35 - 45.

Dritsakis N., Kaparis A., Mastriotis Ch. and Tzililis J. (2003). An index of efficiency for the Athens Stock Exchange, *Spoudai*, Vol. 53, No3, pp. 37 - 56.

Dritsakis N. (2003). A theoretical model for the optimal allocation of health resources in Greece, *Health Services Management Research*, Vol. 16, pp. 39 - 44.

Dourbin J. and Watson G. (1951). Testing for serial correlation in least squares regression. *Biometrika*, Vol. 37, pp. 409 - 428.

Koenker R. and Bassett G. (1982). Robust tests for heteroscedasticity based on regression quantiles, *Econometrica*, Vol. 50, pp. 43 - 61.

Ramsey J. B. (1969). Test for the specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society*, pp 350 - 371.

Wald A. (1943). Test of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Society*, Vol. 54.

ΠΑΡΑΡΤΗΜΑ - ΔΕΔΟΜΕΝΑ

Τα δεδομένα που χρησιμοποιήθηκαν για την εκπόνηση της εν λόγω εργασίας αντλήθηκαν από το τμήμα «Στατιστικής & Ασφαλιστικής Επιστήμης» του Πανεπιστημίου Πειραιά, δόθηκαν στα πλαίσια του μαθήματος «Στατιστικά Προγράμματα».

Τα δεδομένα παρουσιάζονται στον ακόλουθο πίνακα:

SEX	AGE	WEIGHT	HEIGHT	JOB_TYPE	EXERSICE_INDEX
Female	41	54	1,72	Executive	Hard
Female	28	44	1,53	Executive	Hard
Male	34	65	1,82	Executive	Medium
Male	41	51	1,60	Executive	Medium
Female	43	57	1,68	Executive	Medium
Female	46	53	1,61	Executive	Hard
Female	64	62	1,74	Executive	Hard
Male	24	65	1,78	Executive	Medium
Female	31	77	1,93	Self-employed	Medium
Female	53	56	1,64	Self-employed	Hard
Female	27	57	1,65	Self-employed	Hard
Female	29	55	1,62	Self-employed	Medium
Male	78	63	1,72	Clerical	Hard
Male	28	70	1,81	Executive	Medium
Male	53	64	1,73	Clerical	Hard
Male	37	81	1,93	Executive	Medium
Male	28	74	1,84	Self-employed	Medium
Male	24	71	1,80	Executive	Medium
Female	26	66	1,73	Clerical	Hard
Female	44	52	1,53	Executive	Hard

Female	59	65	1,71	Executive	Medium
Female	80	62	1,67	Executive	Hard
Male	39	65	1,70	Clerical	Hard
Female	46	61	1,64	Self-employed	Hard
Female	33	62	1,65	Executive	Hard
Female	46	75	1,81	Clerical	Hard
Male	30	82	1,89	Executive	Medium
Female	59	62	1,64	Executive	Hard
Male	34	70	1,74	Executive	Medium
Female	53	68	1,71	Self-employed	No
Male	51	66	1,68	Clerical	Hard
Female	58	60	1,60	Clerical	Hard
Female	26	64	1,65	Executive	Medium
Male	74	82	1,85	Self-employed	Medium
Female	37	68	1,68	Self-employed	Hard
Male	25	80	1,82	Executive	Medium
Female	30	65	1,64	Executive	No
Male	50	84	1,85	Self-employed	Medium
Male	58	77	1,76	Self-employed	Hard
Male	39	83	1,81	Self-employed	Light
Female	48	62	1,56	Clerical	Hard
Male	36	81	1,78	Self-employed	Medium
Male	35	82	1,79	Executive	Medium
Female	28	69	1,64	Self-employed	No
Female	72	72	1,67	Executive	No
Female	32	74	1,69	Clerical	Hard
Female	49	74	1,68	Self-employed	Light
Male	37	80	1,73	Self-employed	Medium
Male	50	83	1,76	Executive	No
Male	49	85	1,78	Clerical	No
Female	32	79	1,71	Executive	Medium
Male	37	80	1,72	Clerical	No
Male	42	90	1,82	Executive	No
Male	53	83	1,74	Executive	No
Male	43	87	1,78	Self-employed	Light
Female	44	76	1,66	Self-employed	No
Male	26	84	1,74	Executive	No
Female	34	93	1,83	Executive	Hard
Male	53	99	1,88	Executive	Hard
Male	48	82	1,71	Self-employed	Hard
Female	48	89	1,78	Clerical	Hard
Male	52	75	1,63	Self-employed	Light
Female	66	73	1,60	Executive	Hard
Female	70	75	1,62	Self-employed	Light
Female	61	67	1,53	Self-employed	Light
Male	50	97	1,84	Self-employed	No

Female	48	66	1,51	Self-employed	No
Male	56	84	1,70	Self-employed	No
Male	48	85	1,71	Executive	No
Male	57	104	1,89	Self-employed	Light
Female	71	71	1,56	Clerical	Medium
Female	49	66	1,50	Executive	Hard
Male	47	90	1,75	Clerical	Hard
Male	48	102	1,86	Executive	No
Female	30	79	1,63	Self-employed	Light
Male	47	82	1,66	Clerical	No
Male	60	83	1,67	Executive	No
Male	43	83	1,67	Executive	Medium
Male	52	85	1,69	Executive	No
Male	39	99	1,81	Executive	No
Male	37	98	1,80	Executive	No
Male	66	75	1,57	Self-employed	Light
Female	58	77	1,59	Executive	Light
Male	28	82	1,64	Clerical	Hard
Female	46	82	1,64	Self-employed	Light
Male	40	98	1,77	Self-employed	No
Male	45	84	1,63	Self-employed	Medium
Male	57	92	1,70	Self-employed	Light
Male	42	114	1,89	Executive	No
Male	56	98	1,74	Self-employed	No
Female	35	92	1,68	Executive	Hard
Male	54	75	1,51	Executive	No
Female	44	87	1,61	Self-employed	Hard
Male	45	109	1,80	Clerical	No
Male	41	107	1,77	Self-employed	No
Female	38	85	1,55	Executive	Light
Male	46	90	1,58	Executive	Medium
Male	32	105	1,67	Executive	Light
Female	40	95	1,57	Clerical	Hard
Female	49	97	1,56	Self-employed	Light

Παράρτημα με τα δεδομένα- διαφάνειες στο τεταρτο κεφαλαιο power-point

Βήματα ελέγχου chi-square