

ΑΛΕΞΑΝΔΡΕΙΟ ΤΕΙ ΘΕΣΣΑΛΟΝΙΚΗΣ

ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Στατιστική επεξεργασία ιατρικών δεδομένων

Ο ρόλος της Στατιστικής στην Ιατρική και ειδικότερα του κλάδου της Βιοστατιστικής, του οποίου και αναπτύσσονται η βασική ορολογία και οι βασικές έννοιες. Περιλαμβάνονται βασικές τεχνικές της περιγραφικής Στατιστικής με εφαρμογές σε δοσμένα ιατρικά δεδομένα, καθώς και εισαγωγή στην επαγωγική στατιστική με εξόρυξη συμπερασμάτων για έναν πληθυσμό με τη χρήση ενός αντιπροσωπευτικού δείγματος.



Του φοιτητή

Ιωάννη Κυριάκου

Αρ.Μητρώου:02/1891

Επιβλέπων καθηγήτρια

Μαρία Μπαργούλη

ΘΕΣΣΑΛΟΝΙΚΗ 2010

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να εκφράσω τις ευχαριστίες μου προς την κα. Μαρία Μπάργουλη, επιβλέποντα καθηγήτρια της παρούσας πτυχιακής εργασίας, για την πολύτιμη βοήθεια της και στήριξή της που μου παρείχε για όσο χρονικό διάστημα χρειάστηκε για την ολοκλήρωση της. Παράλληλα θα ήθελα να ευχαριστήσω τους γονείς μου και γενικότερα τους πολύ κοντινούς μου ανθρώπους οι οποίοι μου συμπαραστάθηκαν και με βοήθησαν να ολοκληρώσω την συγκεκριμένη πτυχιακή εργασία. Ένα μεγάλο ευχαριστώ στο αφεντικό μου το οποίο αν και χωρίς πτυχίο μου έδωσε την ευκαιρία να εργαστώ κοντά του και να ξεκινήσω την σταδιοδρομία μου.

ΠΡΟΛΟΓΟΣ

Η παρούσα πτυχιακή εργασία περιέχει τις γενικές αρχές και τη βασική θεωρία της στατιστικής καθώς και της ιατρικής στατιστικής.

Η στατιστική μεθοδολογία αποτελεί αναπόσπαστο μέρος της βιοιατρικής έρευνας. Είναι σχεδόν αδύνατο να μην χρησιμοποιηθεί η στατιστική ως βοηθητική επιστημονική μέθοδος για τη συλλογή, επεξεργασία, παρουσίαση, ανάλυση και αξιολόγηση των ευρημάτων της ιατρικής και βιολογικής έρευνας. Επομένως η στατιστική αποβλέπει και στην διευκόλυνση των ιατρών και βιολόγων στην αναζήτηση και εφαρμογή των καταλληλότερων στατιστικών τεχνικών, στην επεξεργασία των ποιοτικών και ποσοτικών δεδομένων της ερευνητικής τους εργασίας.

Η στατιστική είναι επιστήμη η οποία απαιτεί συνεχή αναδρομή σε τύπους και πίνακες. Η σύνοψη αυτών των πινάκων μας βοηθά στο να συγκρίνουμε τις διάφορες οριακές τιμές οι οποίες εμφανίζονται στις στατιστικές δοκιμασίες.

ΠΕΡΙΛΗΨΗ

Στην παρούσα πτυχιακή εργασία έγινε προσπάθεια να γίνει μια ανάλυση του ρόλου της στατιστικής στην ιατρική. Η πτυχιακή αυτή ξεκινά με κάποια γενικά στοιχεία για την ιστορία της στατιστικής από την αρχαιότητα μέχρι την σύγχρονη περίοδο της στατιστικής και αναλύει τους διάφορους τρόπους με τους οποίους γίνεται η συλλογή των στατιστικών δεδομένων.

Στην συνέχεια βλέπουμε με ποιο τρόπο η στατιστική παίζει ενεργό ρόλο στον τομέα της ιατρικής καθώς και στις βιολογικές επιστήμες. Γίνεται αναφορά στον ρόλο που διαδραματίζουν οι ηλεκτρονικοί υπολογιστές στον τομέα της ιατρικής, στην ασφάλεια των ιατρικών δεδομένων και στον ιατρικό φάκελο του ασθενή.

Θα δούμε ότι η στατιστική χωρίζεται σε τρεις μεγάλες ενότητες οι οποίες είναι η θεωρία και οι τεχνικές δειγματοληψίας, η περιγραφική στατιστική και η επαγωγική στατιστική. Στόχος της περιγραφικής στατιστικής είναι η ανάπτυξη μεθόδων για τη συνοπτική και την αποτελεσματική παρουσίαση των δεδομένων (Διαγράμματα, πίνακες συχνοτήτων). Στην επαγωγική στατιστική εξάγουμε συμπεράσματα σχετικά με τον πληθυσμό από δείγματα. Έχουμε τον έλεγχο υποθέσεων ο οποίος προσπαθεί να απορρίψει την μηδενική υπόθεση και τα διαστήματα εμπιστοσύνης.

Στην συνέχεια γίνεται αναφορά στις κύριες λειτουργίες του στατιστικού πακέτου SPSS οι οποίες θα χρησιμοποιηθούν ώστε να εξαχθούν τα αποτελέσματα από τα δεδομένα τα οποία έχουμε και αφορούν τιμές διάφορων καρκινικών δεικτών από ασθενείς, με βασικό στόχο την μελέτη της σχέσης των δεικτών καθώς και τη δημιουργία ενός προγνωστικού μοντέλου του βασικού καρκινικού δείκτη στους ασθενείς.

ΚΕΦΑΛΑΙΟ 1

1.1 Η επιστήμη της Στατιστικής

Στατιστική είναι η επιστήμη που ασχολείται με τη συστηματική συγκέντρωση και ταξινόμηση κοινωνικών ή φυσικών φαινομένων και δεδομένων, που προσφέρονται για αριθμητική επεξεργασία και εξαγωγή συμπερασμάτων και νόμων με γενική ισχύ. Αποτελεί ένα από τους πυρήνες της σύγχρονης επιστημονικής μεθοδολογίας. Η γνώση της είναι απαραίτητη για τον σχεδιασμό πειραμάτων και ερευνών, τη μεθοδολογία σχεδιασμού, ανάλυσης και παρουσίασης αυτών και την αξιοποίηση τους στην εξαγωγή συμπερασμάτων. Επίσης είναι απαραίτητη στους γιατρούς για να είναι σε θέση να αξιολογήσουν κριτικά την ιατρική βιβλιογραφία.

Η στατιστική σαν επιστήμη έκανε την εμφάνιση της τον 16^ο αιώνα για την καταγραφή και επεξεργασία των διάφορων χαρακτηριστικών του κράτους (=status) και στο γεγονός αυτό οφείλει και την ονομασία της. Στη συνέχεια οι μέθοδοι της επεκτάθηκαν, με την πρόοδο που παρουσίασαν οι επιστήμες σε πάρα πολλούς τομείς της ανθρώπινης δραστηριότητας.

Η στατιστική μέθοδος χρησιμοποιείται σήμερα σε τομείς πολύ διαφορετικούς μεταξύ τους όπως η φυσική, η οικονομία, και οι ασφάλειες ζωής. Μια καλά οργανωμένη στατιστική υπηρεσία θεωρείται απαραίτητη για κάθε εκσυγχρονισμένο κράτος και έχει αποστολή να καταγράψει και να κατατάσσει τα δεδομένα όλων σχεδόν των κοινωνικών τομέων (π.χ αύξηση πληθυσμού κάθε χρόνο, αριθμός μαθητών που μπαίνουν σε κάθε τάξη κάθε χρόνο, μεταβολές τιμαρίθμου, κατανομή πολιτών στην παραγωγική δραστηριότητα κ.ο.κ).

Ο σωστός τρόπος χρησιμοποίησεως των στατιστικών δεδομένων είναι απαραίτητος για την εξαγωγή σωστών συμπερασμάτων. Το σύνολο των αντικειμένων (έμψυχων ή απύχων) με τα οποία ασχολείται η στατιστική λέγεται στατιστικός πληθυσμός. Οι ιδιότητες που την ενδιαφέρουν είναι ποιοτικές και ποσοτικές. Οι αριθμοί που εκφράζουν τις ιδιότητες των στοιχείων του στατιστικού πληθυσμού αποτελούν τα στατιστικά δεδομένα τα οποία συγκεντρώνονται με απογραφή ή με δειγματοληψία ή με συνεχή εγγραφή.

1.1.1 Η στατιστική ως απαρίθμηση, ως μέθοδος έρευνας και ως επιστήμη

Η αρχική έννοια της στατιστικής ήταν η απαρίθμηση και καταγραφή των μετρήσεων. Με την πάροδο του χρόνου η έννοια διευρύνθηκε και σήμερα ο όρος στατιστική αναφέρεται σε τρία ξεχωριστά αντικείμενα, την στατιστική ως απαρίθμηση, ως μέθοδο έρευνας και ως επιστήμη. Τα αντικείμενα αυτά πολλές φορές συγχέονται και το περιεχόμενο τους αλληλοεπικαλύπτεται.

1.1.2 Η στατιστική ως απαρίθμηση

Η διατήρηση συνεχούς ροής στατιστικών στοιχείων της κοινωνικής και οικονομικής δραστηριότητας της πολιτείας, απαραίτητων για την μελέτη των δημογραφικών μεταβολών, των οικονομικών μεγεθών και της παραγωγής αποτέλεσε την βάση για την ανάπτυξη της στατιστικής ως απαριθμήσεως και καταγραφής παρατηρήσεων. Αυτή άλλωστε η δραστηριότητα της στατιστικής ήταν και ο σπουδαιότερος παράγοντας της ιστορικής της εξέλιξης.

Η μορφή αυτής της στατιστικής καθαρά περιγραφική δεν είναι υπόθεση αποκλειστικά της κρατικής λειτουργίας. Συναντάται σε κάθε βήμα της οργανωμένης κοινωνίας και είναι περισσότερο γνωστή μορφή στο κοινό. Η παρουσίαση π.χ υπό μορφή αριθμών, πινάκων ή γραφικών απεικονίσεων της δραστηριότητας ενός νοσηλευτικού ιδρύματος ή ενός νοσοκομειακού εργαστηρίου ή μιας κλινικής χαρακτηρίζεται ως στατιστική. Γίνεται μια προσπάθεια, ανεπιτυχής ως τώρα, να διαχωριστεί η έννοια αυτής της στατιστικής με την χρησιμοποίηση των όρων «τα στατιστικά» ή «οι στατιστικές». Με την πάροδο του χρόνου πιθανόν οι όροι να καθιερωθούν.

1.1.3 Η στατιστική ως μέθοδος έρευνας

Η στατιστική, όπως και κάθε επιστημονική μέθοδος έρευνας, χρησιμοποιεί τον αριστοτελικό επαγωγικό συλλογισμό για την εδραίωση των ευρημάτων και την συναγωγή των τελικών συμπερασμάτων. Η θεμελιώδης αυτή μεθοδολογία συνίσταται από:

1. την παρατήρηση
2. την κατάστρωση και διαμόρφωση του προβλήματος ή την διατύπωση της υποθέσεως
3. τον πειραματισμό
4. την συλλογή, ταξινόμηση και ανάλυση των στοιχείων της έρευνας ή του πειράματος
5. την εξαγωγή και έκθεση των συμπερασμάτων

Η ακολουθία αυτή των συλλογισμών, πράξεων και ενεργειών βρίσκει ιδεώδη εφαρμογή στην στατιστική. Η αλληλουχία των ανωτέρω οδηγεί είτε στην επαλήθευση της υποθέσεως είτε στην απόρριψη της και στην διατύπωση νέας υποθέσεως ή στην τροποποίηση της αρχικής, οπότε αρχίζει νέα σειρά πειραματισμών και παρατηρήσεων προς επαλήθευση της νέας υποθέσεως.

Με αυτό το περιεχόμενο συναντούμε την στατιστική μεθοδολογία ολοένα και περισσότερο στην ιατρική, ιδίως στην επιδημιολογία και την βιολογία και σε κάθε σχεδόν τομέα των εφαρμοσμένων και κοινωνικών επιστημών, διότι προσφέρει στον ερευνητή την δυνατότητα μελέτης των ιατρικών, δημογραφικών και κοινωνικών φαινομένων κατά συστηματικό τρόπο.

1.1.4 Η στατιστική ως επιστήμη

Το αντικείμενο της στατιστικής ως επιστήμης είναι οι ομαδικές εκδηλώσεις τις οποίες περιγράφει αριθμητικά. Οι ομαδικές εκδηλώσεις, οι οποίες εκφράζονται αριθμητικά, δηλαδή οι παρατηρήσεις και οι μετρήσεις, είναι συνήθως πολυάριθμες. Η στατιστική επιχειρεί τον περιορισμό αυτού του πλήθους με την συστηματική τους άθροιση και ταξινόμηση. Έπειτα προβαίνει στην ακριβή εκτίμηση των ποσοτικών σχέσεων μεταξύ των ποικίλων μεταβλητών και στην κριτική τους ανάλυση, ώστε να παύουν να αποτελούν οι παρατηρήσεις και τα στοιχεία των πειραμάτων αόριστες και ασύνδετες μεταξύ

τους εντυπώσεις. Ο αντικειμενικός αυτός σκοπός συμπληρώνεται από τον καθορισμό μεγεθών που περιγράφουν τις τάσεις οι οποίες ενδεχομένως υφίστανται μεταξύ των μεταβλητών.

Η στατιστική όμως ως επιστήμη δεν περιορίζει την έρευνά της μόνο στα ιατρικά και βιολογικά φαινόμενα και τις ομαδικές εκδηλώσεις της ανθρώπινης κοινωνικής ζωής. Διευρύνει το πεδίο και μελετά τις τακτικότητες, οι οποίες σταθερά και περιοδικά εκδηλώνονται κατά την διαμόρφωση και ανάπτυξη τους ή προηγούνται και ενδεχομένως προκαλούν αυτά τα φαινόμενα. Μελετά επομένως και τις αιτίες οι οποίες προκαλούν αυτές τις τυπικότητες. Οι αιτίες αυτές πάντως δεν εξομοιώνονται με νόμους. Ο πραγματικός άλλωστε σκοπός και προορισμός της στατιστικής δεν είναι η θέσπιση νόμων, αλλά η αποκάλυψη αιτιωδών συσχετίσεων και αλληλεξαρτήσεων. Η πλευρά αυτή της ιατρικής στατιστικής είναι αρκετά συνδεδεμένη με την επιδημιολογία.

Η αιτιολογική επεξήγηση των διαφόρων βιολογικών νόμων, όπως γενικότερα και κάθε παρατηρήσεως ή πειραματικού ευρήματος ανήκει στην ειδική επιστήμη, δηλαδή στην ιατρική ή την βιολογία και όχι στην βοηθητική δηλαδή την στατιστική.

Κατά την εξέλιξη της ως επιστήμη η στατιστική αντιμετώπισε και ορισμένες σθεναρές αντιδράσεις οι οποίες έφθασαν ως το σημείο διατυπώσεως σοβαρών κατηγοριών, ότι πχ «υπάρχουν τριών ειδών ψεύδη τα κατ' ανάγκην, τα από σκοπού και η στατιστική». Σύντομα βέβαια αποδείχθηκε ότι η στατιστική δεν είναι κάποιο επιστημονικό ψεύδος, αλλά αληθινή και γόνιμη επιστήμη. Η αρχική αρνητική αντίδραση φαίνεται ότι πρέπει να αποδοθεί στην έλλειψη καταρτίσεως και στην φυσική, κατά κάποιο τρόπο αποστροφή των ανθρώπων προς τα μαθηματικά. Οι παραπάνω παράγοντες, κατά παράδοση περισσότερο έκδηλοι στους ιατρούς δυσκόλεψαν την προσέγγιση της ιατρικής επιστήμης με την στατιστική.

Ένα πρόβλημα το οποίο απασχόλησε τους ειδικούς και πάνω στο οποίο ακούστηκαν οι πιο αντιφατικές εισηγήσεις ήταν η χρησιμοποίηση των ανωτέρων μαθηματικών στην στατιστική. Πολλοί διαπρεπείς στατιστικολόγοι αντιτάχθηκαν στη χρήση των ανωτέρων μαθηματικών στην στατιστική

ανάλυση, πρεσβεύοντας ότι μόνο τα στοιχειώδη μαθηματικά πρέπει να χρησιμοποιούνται στην στατιστική. Πράγματι για τις συνηθισμένες στατιστικές εργασίες, οι οποίες και αποτελούν την τεράστια πλειονότητα της καθημερινής ιατρικής πράξεως, η γνώση των απλών μαθηματικών, γυμνασιακού επιπέδου θεωρείται επαρκής ώστε να μη χρειάζεται προσφυγή στα ανώτερα μαθηματικά. Ως επιστήμη όμως η στατιστική χρησιμοποιεί συχνά τον ανώτερο μαθηματικό λογισμό. Η ευρύτερη θεώρηση της στατιστικής, ως κλάδου της μαθηματικής επιστήμης, απαιτεί την χρήση του ανώτερου μαθηματικού λογισμού, ιδιαίτερα στην διατύπωση αφηρημένων υποθέσεων και στη δημιουργία στατιστικών υποδειγμάτων ή προτύπων. Με αυτήν την μορφή όμως η στατιστική γίνεται πολλές φορές αυτοσκοπός.

Στην ιατρική και τις βιολογικές επιστήμες τα προβλήματα δεν είναι τόσο σύνθετα ώστε να μην λύνονται με τα κατώτερα, γυμνασιακά μαθηματικά. Αντιθέτως, τελείως απαραίτητες είναι η λογική και η στατιστική σκέψη για την προσπέλαση των ιατρικών και βιολογικών προβλημάτων.

1.2 Ιστορία της ιατρικής στατιστικής

Η λέξη στατιστική προέρχεται από τη λατινική λέξη status, δηλαδή κράτος, πολιτεία. Η αρχική σημασία της λέξεως στατιστική αναφερόταν στην απαρίθμηση και καταγραφή αριθμητικών ενδείξεων, οι οποίες ήταν απαραίτητες για την κρατική λειτουργία. Με την έννοια αυτή μπορούμε να υποστηρίξουμε ότι η στατιστική υφίσταται ήδη στην αρχαιότητα, ενώ ως επιστήμη και ως μέθοδος έρευνας είναι πρόσφατη.

1.2.1 Η στατιστική στην αρχαιότητα

Η αρχική μορφή της στατιστικής ήταν οι αριθμητικές ενδείξεις για τον πληθυσμό και την παραγωγή, που ενδιέφεραν την κρατική λειτουργία. Τα στοιχεία αυτά, απαραίτητα για διοικητικούς, φορολογικούς και στρατιωτικούς σκοπούς, συλλέγονταν ακόμη και στην απώτατη αρχαιότητα.

Οι πρώτες απαριθμήσεις πληθυσμών φαίνεται ότι έγιναν στη Βαβυλώνα. Στην Κίνα επίσης κατεβλήθη προσπάθεια διακριβώσεως του πληθυσμού της χώρας. Στην Παλαιά Διαθήκη υπάρχουν σημαντικές στατιστικές πληροφορίες. Γνωστή είναι η απαρίθμηση του εβραϊκού λαού στο όρος Σινά και η κατανομή του εδάφους από το Μωϋσή.

Στην αρχαία Ελλάδα η απαρίθμηση των κατοίκων της Αττικής, η οποία έγινε από τον Κέκροπα, η κατανομή των γαιών από τον Λυκούργο στη Σπάρτη και η διαίρεση των Αθηναίων σε τάξεις από τον Σόλωνα, αποδεικνύουν ότι υπήρχε παράδοση συγκεντρώσεως στατιστικών πληροφοριών και ίσως διενέργειας ερευνών επί του πληθυσμού, της ιδιοκτησίας και του εισοδήματος.

1.2.2 Η στατιστική στον Μεσαίωνα

Τα μοναδικά στατιστικά στοιχεία ήταν τα εκκλησιαστικά δίπτυχα, θα λέγαμε η απαρχή των ληξιαρχικών καταγραφών. Μέχρι και τον 15^ο αιώνα το ενδιαφέρον για την απαρίθμηση των πληθυσμών ήταν ελάχιστο. Εξαιρεση αποτελούν ορισμένες μετρήσεις οι οποίες έγιναν στη Γαλλία για τον πληθυσμό και τα εκκλησιαστικά κτήματα επί Καρλομάγνου και στην Αγγλία τον 11^ο αιώνα.

1.2.3 Η στατιστική στους νεώτερους χρόνους

Από τον 16^ο αιώνα και ύστερα διαπιστώνεται αξιόλογη πρόοδος στο πεδίο καταγραφής βιοτικών γεγονότων, κοινωνικών παρατηρήσεων και στις απογραφές του πληθυσμού, ιδιαίτερα στη Γαλλία, Αγγλία και Γερμανία. Στις μεγάλες πόλεις που αναπτύχθηκαν οικονομικά και πολιτιστικά κατά την περίοδο του Εμποροκρατισμού και αργότερα του Διαφωτισμού η στατιστική των απαριθμήσεων έλαβε μεγάλη ώθηση. Προέκυψε ένα ευρύ και δύσχρηστο σύστημα απογραφών του πληθυσμού για φορολογικούς, στρατιωτικούς και διοικητικούς σκοπούς. Πρέπει να σημειωθεί ότι τα αποτελέσματα των πληθυσμιακών μετρήσεων μέχρι και τα μέσα του 19^{ου} αιώνα κατά κανόνα δεν

δημοσιεύονταν. Χρησίμευαν ως πληροφορίες για τις ενδιαφερόμενες αρχές και αποτελούσαν κατά κάποιο τρόπο μυστικά του κράτους. Σε εξαιρετικές περιπτώσεις δίνονταν ορισμένα στοιχεία σε επιστήμονες.

Τον 17^ο αιώνα είδε τη γέννησή της η Δημογραφία, ο κλάδος της στατιστικής που ασχολείται αποκλειστικά με τον πληθυσμό σαν συνέχεια της Πολιτικής αριθμητικής.

Κατά τον 18^ο αιώνα σε πολλά πανεπιστήμια της Δύσεως καθιερώθηκε η διδασκαλία ενός μαθήματος, της Πολιτειογραφίας. Στην πραγματικότητα το μάθημα συνίστατο από ενότητα τριών αντικειμένων, της γεωγραφίας, της ιστορίας και της πολιτικής. Το 1748 την Πολιτειογραφία δίδαξε ο Gottfried Achenwall ο οποίος στις παραδόσεις του χρησιμοποίησε για πρώτη φορά τον όρο στατιστική. Ο όρος έγινε γενικά αποδεκτός και καθιερώθηκε. Συνέπεια των ανωτέρω ήταν να ονομασθεί από πολλούς ο Achenwall πατέρας της στατιστικής.

Παράλληλα με την ανάπτυξη της στατιστικής των απαριθμήσεων των κοινωνικών φαινομένων και της μετρήσεως των δημογραφικών παραμέτρων σημειώνεται κατά την ίδια περίοδο και σοβαρή εξέλιξη της στατιστικής ως επιστήμης και ως μεθόδου έρευνας. Η ανάπτυξη οφείλεται κατά κύριο λόγο στην πρόοδο των σχετικών κλάδων των μαθηματικών και ιδιαίτερα της θεωρίας των πιθανοτήτων.

Κατά τον 19^ο αιώνα οι στατιστικές έρευνες στράφηκαν κυρίως στη σύλληψη των απροσώπων ομαδικών εκδηλώσεων του πολιτικού και κοινωνικοοικονομικού βίου. Η στατιστική των απαριθμήσεων έλαβε συστηματικότερη μορφή αρχικά για την συγκέντρωση δημογραφικών στοιχείων και κατόπιν για την παρακολούθηση της παραγωγής. Η ίδρυση εξάλλου εθνικών στατιστικών υπηρεσιών υποβοήθησε πολύ στην παρακολούθηση παγκοσμίων στατιστικών σειρών κοινού ενδιαφέροντος από τους πολλούς δε στατιστικούς οργανισμούς που κατά καιρούς ιδρύθηκαν βασικής σημασίας για την πρόοδο της στατιστικής ήταν το Διεθνές Συμβούλιο υγιεινής και δημογραφίας, το διεθνές ινστιτούτο στατιστικής της Χάγης, ενώ στις ημέρες μας πολλά οφείλονται στις στατιστικές υπηρεσίες του οργανισμού ηνωμένων εθνών και της παγκόσμιας οργανώσεως υγείας.

Η άνθιση της στατιστικής ως ανεξάρτητου κλάδου επιστημονικής έρευνας κατά τον 19^ο αιώνα οφείλεται κατά πολύ στο Βέλγο επιστήμονα Adolphe Quetelet, η ευρύτητα και η σπουδαιότητα του έργου του οποίου παραμένουν ανεκτίμητες. Ο Quetelet βασίστηκε στις θεωρίες του Laplace και επιχείρησε την συστηματική καταμέτρηση του ανθρωπίνου σώματος και των πνευματικών ικανοτήτων των ανθρώπων.

1.2.4 Η σύγχρονη περίοδος της στατιστικής

Μετά τη γόνιμη σπορά του παρελθόντος η εξέλιξη της στατιστικής στις αρχές του 20^{ου} αιώνα έλαβε τη μορφή αληθινής επιστημονικής επανάστασης. Η ριζική αλλαγή στην στατιστική μεθοδολογία και στον τρόπο σκέψης γεννήθηκε στην Αγγλία χάρη στις εργασίες του Karl Pearson και του Ronald Fisher αρχικά σε προβλήματα γεωπονίας και αγρονομίας και αργότερα ιατρικής, βιολογίας, κοινωνιολογίας και οικονομικής. Το 1900 ο Pearson επινόησε τη μέθοδο διατυπώσεως της συμφωνίας μεταξύ παρατηρήσεως και υποθέσεως, το γνωστό χ^2 . Ταυτόχρονα άρχισαν να διαμορφώνονται σπουδαίες στατιστικές μέθοδοι δειγματοληψίας και μέθοδοι ελέγχου της στατιστικής σημαντικότητας και αξιοπιστίας. Ένα σπουδαίο βήμα συντελέστηκε το 1908 όταν ο Gosset δημοσίευσε με το ψευδώνυμο student, πραγματεία για τον έλεγχο της στατιστικής σημαντικότητας μεταξύ των μέσων όρων δειγμάτων μικρού αριθμού παρατηρήσεων, του γνωστού κριτηρίου t , του t -test.

1.3 Βασική στατιστική ορολογία

Η συγκέντρωση, η ταξινόμηση και ανάλυση των παρατηρήσεων και των μετρήσεων είναι το αρχικό βήμα της στατιστικής. Το υλικό αυτό, δηλαδή τα στατιστικά στοιχεία ή δεδομένα προκύπτουν είτε από την απαρίθμηση βιολογικών φαινομένων και ομοειδών αντικειμένων είτε από την μέτρηση των διαφόρων βιολογικών τιμών ή παραμέτρων και των λοιπών γνωρισμάτων των οργανισμών, διαφόρων ουσιών και αντικειμένων. Το στατιστικό υλικό που παρέχεται από απαριθμήσεις εκφράζεται πάντοτε σε ακέραιους αριθμούς, ενώ

οι μετρήσεις οδηγούν τόσο σε ακέραιους όσο και σε δεκαδικούς αριθμούς. Η χρήση των δεκαδικών στις μετρήσεις είναι θέμα ακριβείας των οργάνων και επιθυμητής προσεγγίσεως τιμής.

1.3.1 Ποσοτικό και ποιοτικό στατιστικό υλικό, πρωτογενές και δευτερογενές

Διακρίνονται τέσσερις μορφές στατιστικού υλικού, το ποσοτικό, το ποιοτικό, το πρωτογενές και το δευτερογενές. Το ποσοτικό εκφράζεται με αριθμούς οι οποίοι περιγράφουν ποσοτικές ιδιότητες και το ποιοτικό αναφέρεται σε ποιοτικές ιδιότητες ενός χαρακτηριστικού. Το πρωτογενές ή πρωτότυπο στατιστικό υλικό ονομάζεται έτσι όταν ο ερευνητής που το συλλέγει είναι ο ίδιος ο οποίος το χρησιμοποιεί. Δευτερογενές στατιστικό υλικό είναι αντίθετα το στατιστικό υλικό που συγκεντρώθηκε για γενικούς σκοπούς και χρησιμοποιείται από τους ερευνητές για ειδικές έρευνες.

Το υλικό είναι δυνατό να αναφέρεται σε κάποια συγκεκριμένη χρονική στιγμή ή διάρκεια. Τότε είναι στατικό στατιστικό υλικό. Η θεώρηση του κάτω από το πρίσμα της ροής του χρόνου δίνει το δυναμικό στατιστικό υλικό.

1.3.2 Στατιστικός πληθυσμός και στατιστικό δείγμα

Η έννοια του στατιστικού πληθυσμού ή πλήθους και του δείγματος είναι βασικής σημασίας στην στατιστική. Το σύνολο των ομοειδών στατιστικών μονάδων, των παρατηρήσεων ή των μετρήσεων συνιστά τον στατιστικό πληθυσμό ή πλήθος. Το στατιστικό πλήθος αναφέρεται σε οποιοδήποτε σύνολο δεδομένων ατόμων, βιολογικών γεγονότων, φαινομένων ή βιολογικών τιμών, το οποίο προέκυψε από την απαρίθμηση ομοειδών στατιστικών μονάδων ή από τη μέτρηση των διαφόρων χαρακτηριστικών αυτών. Το μέγεθος του πλήθους μπορεί να είναι άπειρο ή πεπερασμένο.

Κατά τις διάφορες επιστημονικές ή βιοστατιστικές εργασίες είναι σχεδόν αδύνατη ή πολύ δύσκολα πραγματοποιήσιμη η μελέτη του συνόλου των στατιστικών μονάδων, όταν μάλιστα το σύνολο τους είναι τεράστιο. Σε

αυτές τις περιπτώσεις η έρευνα στρέφεται στην μελέτη ενός μικρού μέρους του στατιστικού πλήθους το οποίο λέγεται στατιστικό δείγμα ή δείγμα. Αν το δείγμα είναι αντιπροσωπευτικό του πληθυσμού από τον οποίο προέρχεται τότε τα ευρήματα από την μελέτη μπορούν να αναχθούν στο σύνολο του πληθυσμού.

1.3.3 Μεταβλητές

Μεταβλητή ονομάζεται στη στατιστική κάθε χαρακτηριστικό το οποίο είναι δυνατό να καταμετρηθεί με τη βοήθεια των συνήθων μονάδων μετρήσεως και να λάβει ολόκληρη σειρά τιμών με μορφή αριθμών. Ένα μεταβλητό ποσό ή μεταβλητή τιμή που μπορεί να λάβει όλες τις δυνατές τιμές στην κλίμακα των μετρήσεων μεταξύ δυο ορίων και να μεταβάλλεται κατά μικρές ποσότητες καλείται συνεχής μεταβλητή. Ασυνεχής ή διακεκομμένη μεταβλητή είναι εκείνη της οποίας οι τιμές μεταβάλλονται με καθορισμένη ποσότητα στην κλίμακα των μετρήσεων. Από την μέτρηση των χαρακτηριστικών των στατιστικών μονάδων προκύπτουν γενικώς οι συνεχείς μεταβλητές ενώ οι ασυνεχείς είναι προϊόν καταμετρήσεως ή απαριθμήσεως των στατιστικών μονάδων. Όταν κατά την μέτρηση δύο μεταβλητών οι διάφορες τιμές τις οποίες λαμβάνει η μια αποτελούν συνάρτηση των τιμών που παίρνει η άλλη τότε η πρώτη λέγεται εξαρτημένη ενώ αν οι τιμές της παραμένουν ανεπηρέαστες αποτελεί ανεξάρτητη μεταβλητή.

1.3.4 Συλλογή και επεξεργασία του στατιστικού υλικού

1.3.4.1 Ο στατιστικός πληθυσμός

Ο όρος στατιστικός πληθυσμός χρησιμοποιείται για να περιγράψει ένα μεγάλο αριθμό ζώντων οργανισμών, αναφέρεται όμως και σε σειρές μετρήσεων και παρατηρήσεων. Οι στατιστικοί πληθυσμοί μπορούν να διακριθούν σε άπειρους ή τόσο πολυπληθείς ώστε να θεωρούνται πρακτικά άπειροι ή αόριστοι και σε πεπερασμένους.

Οι στατιστικές μονάδες που περιλαμβάνονται σε ένα πεπερασμένο πληθυσμό μπορούν να απαριθμηθούν όλες. Σε αυτή τη περίπτωση μιλούμε για πλήρη καταμέτρηση ή απογραφή. Όταν ο πληθυσμός είναι άπειρος η καταμέτρηση των στατιστικών μονάδων οι οποίες τον απαρτίζουν είναι ανέφικτη αλλά και όταν ο πληθυσμός είναι πεπερασμένος η απαρίθμηση καθίσταται δυσχερής ή και αδύνατη.

1.3.4.2 Το στατιστικό δείγμα

Το στατιστικό δείγμα παριστάνει οποιαδήποτε ειδική συλλογή παρατηρήσεων, η οποία λαμβάνεται από ένα αντίστοιχο στατιστικό πληθυσμό και αποτελεί επομένως μέρος αυτού. Το δείγμα μπορεί κατά περίπτωση να είναι μικρό και να περιλαμβάνει λίγες μονάδες ή μερικές μόνο δεκάδες στατιστικών μονάδων του πληθυσμού από τον οποίο πηγάζει. Αντιθέτως μπορεί να αποτελείται από ένα σεβαστό ποσοστό του αριθμού των στατιστικών μονάδων.

Κατά τον πειραματισμό στην ιατρική και την βιολογία ασχολούμαστε σχεδόν πάντοτε με δείγματα πάνω στα οποία μελετούμε διάφορα χαρακτηριστικά και από αυτά στη συνέχεια επεκτείνουμε τα συμπεράσματα της έρευνας στον γενικό πληθυσμό. Για να είναι εφικτή η γενίκευση των συμπερασμάτων της έρευνας από το δείγμα στον πληθυσμό απαιτείται το δείγμα να είναι αντιπροσωπευτικό του πληθυσμού από τον οποίο προέρχεται. Το δείγμα αντιπροσωπεύει τον πληθυσμό, όταν τα προς διερεύνηση χαρακτηριστικά των στατιστικών μονάδων του δείγματος είναι όμοια με τα χαρακτηριστικά των στατιστικών μονάδων του πληθυσμού και αντιστρόφως.

Η αντιπροσωπευτικότητα αυτή μπορεί να εξασφαλιστεί μόνο εφ' όσον παρέχεται η ίδια ευκαιρία σε όλες τις στατιστικές μονάδες ή παρατηρήσεις του πληθυσμού να αποτελέσουν μέρος του δείγματος. Σε αυτή την περίπτωση το δείγμα είναι αντιπροσωπευτικό ή όπως λέγεται στην στατιστική ορολογία αδιάβλητο. Σε όλες τις άλλες περιπτώσεις στις οποίες το δείγμα σχηματίζεται χωρίς να τηρείται η βασική αρχή της παροχής της ίδιας ευκαιρίας σε όλες τις στατιστικές μονάδες να αποτελέσουν και μονάδες του δείγματος μιλούμε για

μεροληπτικό δείγμα. Η αποφυγή μεροληψίας επιτυγχάνεται με μια καθορισμένη μεθοδολογία η οποία οδηγεί στο σχηματισμό του απλού τυχαίου δείγματος.

1.3.4.3 Τυχαία δειγματοληψία

Η επίτευξη της ίδιας πιθανότητας, ώστε κάθε στατιστική μονάδα να αποτελέσει μέρος του δείγματος η οποία αποτελεί θεμελιώδη αρχή της στατιστικής δειγματοληψίας, επιτυγχάνεται πολλές φορές με την κατάταξη προηγουμένως των μονάδων ή των χαρακτηριστικών του πληθυσμού σε στρώματα, δηλαδή σε ομοιογενείς ομάδες, τάξεις ή κλάσεις. Ο διαχωρισμός του πληθυσμού σε στρώματα γίνεται με βάση καθορισμένα από πριν κριτήρια. Αν δεν προηγηθεί κατάταξη των μονάδων του στατιστικού πλήθους σε στρώματα αλλά ολόκληρος ο πληθυσμός θεωρείται μια ομάδα, ένα στρώμα, τότε η δειγματοληψία καλείται απλή τυχαία δειγματοληψία. Στην περίπτωση που το στατιστικό πλήθος διαιρείται με βάση ορισμένα κριτήρια σε ομοιογενείς μονάδες, σε στρώματα, έχουμε την ενστρωματωμένη τυχαία δειγματοληψία. Στην συγκεκριμένη περίπτωση τα μέλη του δείγματος λαμβάνονται από κάθε στρώμα τυχαία και με βάση τις αναλογίες με τις οποίες τα στρώματα συναντώνται στον στατιστικό πληθυσμό. Με τη διεργασία αυτή το τελικό τυχαίο δείγμα σχηματίζεται από τόσα επί μέρους στρώματα όσα και εκείνα στα οποία διαιρέθηκε το αρχικό πλήθος. Έτσι τα διάφορα στοιχεία του πληθυσμού αντικατοπτρίζονται αναλογικά στο δείγμα. Το δείγμα αυτό καλείται ενσωματωμένο.

Ο αριθμός των στρωμάτων ποικίλλει ανάλογα με τη φύση και σύνθεση του πληθυσμού. Εύλογο είναι ότι όσο αυξάνεται ο αριθμός τους τόσο περισσότερο αυξάνεται η αντιπροσωπευτικότητα του δείγματος. Λόγοι όμως συνήθως τεχνικοί δεν επιτρέπουν τον διαμερισμό του πληθυσμού σε περισσότερα των 12 στρωμάτων πράγμα που δεν επιτυγχάνεται ή παραβλέπεται. Η ενστρωματωμένη δειγματοληψία είναι συνηθισμένη στη διερεύνηση κοινωνικών φαινομένων. Αλλά και στην ιατρική και στην βιολογία εφαρμόζεται συχνά.

Η επιλογή των μελών του δείγματος κατά τη δειγματοληψία επιτυγχάνεται με την χρήση ειδικών απλών μεθόδων και κυρίως με τη χρήση τυχαίων αριθμών. Με τη μέθοδο της τυχαίας δειγματοληψίας δίνεται η ίδια πιθανότητα σε όλα τα μέλη του στατιστικού πληθυσμού να περιληφθούν στο δείγμα, αποκλείεται η υποκειμενικότητα και εξουδετερώνεται η ενδεχόμενη μεροληπτική επίδραση του ερευνητή στην επιλογή των μονάδων του δείγματος.

Ο αριθμός των πηγών συλλογής ιατρικών και βιολογικών στοιχείων είναι προφανώς τεράστιος. Γενικότερα το στατιστικό υλικό προέρχεται από τρεις μεγάλες κατηγορίες.

1. Τον πειραματισμό στην κλινική
2. Από τις ιατροκοινωνικοοικονομικές και επιδημιολογικές έρευνες
3. Από τον συνδυασμό αυτών των δύο

Στην ιατρική και τις βιολογικές επιστήμες η στατιστική ύλη είναι στο μεγαλύτερο της ποσοστό αποτέλεσμα πειραματικής εργασίας και έρευνας. Μικρότερο ποσοστό προέρχεται και από ιατροκοινωνικές έρευνες, επιδημιολογικές, δημόσιας υγείας, κοινωνικής ιατρικής και ιατρικής δημογραφίας. Με την ευρύτερη πάντως έννοια στο υλικό πειραματισμού περιλαμβάνονται όλα τα στατιστικά στοιχεία, τα οποία προκύπτουν από την καθημερινή ιατρική πράξη στην κλινική και στο εργαστήριο και όχι μόνο τα στατιστικά δεδομένα από τις ειδικές ερευνητικές εργασίες.

Ο αριθμός των στατιστικών μονάδων είναι συνήθως μικρός στο υλικό πειραματισμών και η απαρίθμησή τους δεν παρουσιάζει δυσκολίες. Η μέτρηση επίσης των χαρακτηριστικών των στατιστικών μονάδων είναι ακριβής, ιδίως όταν οι μέθοδοι πραγματοποιήσεως των μετρήσεων και τα όργανα τα οποία χρησιμοποιούνται είναι τα κατάλληλα. Το στατιστικό υλικό που προκύπτει από τα πειράματα είναι κατά κανόνα ακριβές, εκτός εάν εξαιρεθούν οι περιπτώσεις τεχνικών λαθών και λανθασμένων μεθόδων.

Αντίθετα το στατιστικό υλικό των ιατροκοινωνικών ερευνών στερείται μεγάλης ακριβείας, διότι και ο αριθμός των στατιστικών μονάδων είναι

συνήθως μεγάλος και η μέτρηση των χαρακτηριστικών των ιατροκοινωνικών παρατηρήσεων δυσχερής. Ο τελευταίος παράγοντας επιτείνεται πολλές φορές από τον ανθρώπινο παρεμβατισμό. Τούτο οφείλεται στο γεγονός ότι οι παρατηρήσεις στις ιατροκοινωνικές, επιδημιολογικές και ιατροδημογραφικές έρευνες προέρχονται συνήθως από πληροφορίες, τις οποίες παρέχουν τα άτομα μετά από σχετικές ερωτήσεις.

Σε γενικές γραμμές η στατιστική ύλη από πειραματισμούς παριστάνει υλικό αντικειμενικών μετρήσεων, ενώ το υλικό των ιατροκοινωνικών ερευνών αντιστοιχεί συνήθως σε εκτιμήσεις του μεγέθους των χαρακτηριστικών που μελετώνται.

1.3.4.4 Το στατιστικό δελτίο. Το ερωτηματολόγιο

Υπάρχουν πολλοί τρόποι συγκεντρώσεως του στατιστικού υλικού. Ο αριθμός τους και η ποικιλία είναι αρκετά μεγάλος και η εκλογή της κατάλληλης μεθόδου συλλογής στατιστικού υλικού κατά περίπτωση εξαρτάται βασικά από τη φύση της ειδικής εργασίας,, τη οποία πραγματοποιεί ο ιατρός ή ο βιολόγος.

Ανεξάρτητα όμως από τη μέθοδο, η οποία θα χρησιμοποιηθεί για τη συλλογή των στατιστικών δεδομένων, εκείνο που έχει μεγαλύτερη σημασία είναι ο τρόπος σκέψεως για την διεξαγωγή της έρευνας ή του πειράματος. Η λογική και η τιμιότητα είναι δύο από τις θεμελιωδέστερες αρετές για τη συλλογή, την επεξεργασία και γενικά τη διαχείριση της στατιστικής ύλης. Γι 'αυτό και σε κάθε δημοσίευμα στατιστικού υλικού πρέπει να περιγράφονται με κάθε ακρίβεια ο τρόπος συλλογής του υλικού και η μεθοδολογία της επεξεργασίας του.

Ο χειρότερος τρόπος με τον οποίο μπορεί να συλλεγεί στατιστικό υλικό, και ο οποίος ατυχώς χρησιμοποιείται επιπόλαια αρκετές φορές είναι η αναγραφή των διαφόρων ευρημάτων και των μετρήσεων σε ένα φύλλο χαρτί ή σε ένα τετράδιο. Το αποτέλεσμα είναι συνήθως η χωρίς τάξη καταχώρηση των δεδομένων, οι παραλήψεις και παρείσφρηση λαθών. Η κατάληξη είναι φυσικά η δυσχέρεια ή ακόμη και η αδυναμία διαλογής των λανθασμένων στοιχείων. Γι 'αυτό η κατάστρωση και χρησιμοποίηση ενός στατιστικού

δελτίου, προσαρμοσμένου στον αντικειμενικό σκοπό της ειδικής έρευνας, είναι βασικής σημασίας για την συλλογή και κατάταξη του στατιστικού υλικού.

Στις ιατροκοινωνικές μελέτες το δελτίο παίρνει τη μορφή δελτίου απογραφής ή ερωτηματολογίου. Στο εργαστήριο και στην κλινική μπορεί να περιορισθεί στην απλή μορφή ενός ατομικού δελτίου μεταφοράς των κλινικών ευρημάτων ή των εργαστηριακών και παρακλινικών εξετάσεων για κάθε άρρωστο που αποτελεί αντικείμενο της επιστημονικής εργασίας. Η αναγραφή των στατιστικών δεδομένων στο δελτίο κάθε στατιστικής μονάδας μπορεί να γίνει απ' ευθείας ή κατόπιν μεταφοράς από τα πρωτόκολλα της έρευνας, τα ιστορικά και τα φύλλα νοσηλείας ή από τους πίνακες μηχανογραφήσεως των στοιχείων των ασθενών.

Πρέπει πάντως να τονιστεί ότι δεν υπάρχει κάποιο κοινό πρότυπο στατιστικού δελτίου. Τα χαρακτηριστικά του δελτίου ποικίλλουν ανάλογα με την εργασία. Ο ερευνητής σχεδιάζει μόνος του το στατιστικό δελτίο και το προσαρμόζει στις πραγματικές ανάγκες της δουλειάς του.

1.3.4.5 Η διαλογή του στατιστικού υλικού

Μετά την αποπεράτωση της εργασίας, την διεξαγωγή των πειραμάτων ή των κλινικών παρατηρήσεων και την συλλογή όλων των στοιχείων ακολουθεί η διαλογή της στατιστικής ύλης. Διαχωρίζονται τα στοιχεία των μετρήσεων και των παρατηρήσεων και σχηματίζονται ομάδες ομοειδών στατιστικών μονάδων και κοινών χαρακτηριστικών.

Η προσπάθεια κατά την διαλογή του στατιστικού υλικού αποβλέπει στην εξάλειψη της ανομοιογένειας, στην οποία βρίσκεται η μάζα των στοιχείων και στην κατάταξη της ύλης σε ευσύνοπτο αριθμό μονάδων κοινών χαρακτηριστικών. Για τον σκοπό αυτό τα στοιχεία μεταφέρονται αρχικά σε πρόχειρους πίνακες. Από αυτούς επακολουθεί σε δεύτερο στάδιο η κατάστρωση του βασικού ή των βασικών πινάκων στους οποίους απεικονίζεται πιστά και ταυτόχρονα μόνιμα η τελική εικόνα του συλλεγμένου στατιστικού υλικού.

Η κατάστρωση των προχείρων πινάκων του ιατρικού και βιολογικού στατιστικού υλικού γίνεται εύκολα με το χέρι, όταν ο αριθμός των στατιστικών μονάδων είναι μικρός. Από τα στατιστικά δελτία μεταφέρονται τα διάφορα χαρακτηριστικά στους πρόχειρους πίνακες.

Όταν ο αριθμός των στατιστικών μονάδων είναι μεγάλος και τα χαρακτηριστικά τα οποία μελετούμε είναι πολλά, η πινακοποίηση προφανώς συναντά ορισμένες δυσχέρειες. Η εργασία τότε διευκολύνεται με τον σχηματισμό κατανομών συχνοτήτων. Στις εξαιρετικά εκτεταμένες ιατροκοινωνικές έρευνες των χρονίων νοσημάτων στο γενικό πληθυσμό η πινακοποίηση μπορεί να γίνει με την βοήθεια των ηλεκτρονικών διερευνητών μέσα στα γενικότερα πλαίσια της επεξεργασίας του στατιστικού υλικού με τις μηχανικές μεθόδους.

Η διαλογή των δελτίων είναι συνήθως η πιο ανιαρή και μονότονη εργασία από την όλη μελέτη. Μπορεί να γίνει είτε με το χέρι είτε μηχανικά ή ημιμηχανικά. Η διαλογή με το χέρι είναι αναπόφευκτη και γι ' αυτό αποτελεί τον συνηθέστερο τρόπο διαχωρισμού του στατιστικού υλικού στα εργαστήρια και στις κλινικές, όταν μάλιστα ο αριθμός των στατιστικών μονάδων είναι μικρός. Καθίσταται όμως αδύνατη σε εκτεταμένες έρευνες με μεγάλο αριθμό στατιστικών δελτίων και όταν οι διάφοροι συνδυασμοί των στοιχείων που μελετούμε είναι πολυάριθμοι. Την εργασία αυτή την επιτυγχάνουμε σήμερα με τους ηλεκτρονικούς υπολογιστές.

Η μηχανική επεξεργασία του στατιστικού υλικού ήταν απόρροια της αναπτύξεως του προγραμματισμού και της ευρείας χρησιμοποίησεως των ηλεκτρονικών υπολογιστών πρώτης και δεύτερης γενιάς. Σήμερα χρησιμοποιείται ελάχιστα και σύντομα και αυτή η μέθοδος θα αποτελέσει ιστορική ανάμνηση. Επιτυγχάνεται με τη χρήση ειδικών δελτίων στα οποία η μεταφορά των πληροφοριών και των αριθμητικών δεδομένων γίνεται με διατρήσεις, τις οποίες πραγματοποιούν ειδικά πολύπλοκα μηχανήματα. Αρχικά τα στατιστικά στοιχεία κωδικοποιούνται από τον ειδικό προγραμματιστή και με τη βοήθεια μιας διατρητικής μηχανής μεταφέρονται με διατρήσεις σε κατάλληλες θέσεις του δελτίου. Τα διάτρητα δελτία (punch cards) επεξεργάζεται κατόπιν ο ηλεκτρονικός διερευνητής διά μέσου μιας

σειράς πολύπλοκων μονάδων του, όπως είναι η μονάδα επαλήθευσης και η μονάδα πινακοποίησης. Επιτυγχάνονται έτσι σύνθετες εργασίες διαχωρισμού, συσχέτισεως και αποθηκεύσεως σε μνήμες των κωδικοποιημένων στατιστικών δεδομένων.

Η ημιμηχανική διαλογή των δελτίων είναι μέθοδος απλή. Δεν απαιτεί πολύπλοκα και δαπανηρά μηχανήματα και στηρίζεται στα δελτία με περιφερικές ή κεντρικές διατρήσεις. Η πρόοδος όμως της τεχνολογίας παραγκώνισε και αυτή τη μέθοδο και γι' αυτό απλώς μνημονεύεται εδώ για ιστορικούς λόγους.

ΚΕΦΑΛΑΙΟ 2

2.1 Η χρήση των ηλεκτρονικών υπολογιστών ή ηλεκτρονικών διερευνητών στην ιατρική στατιστική

Οι σύγχρονοι μικροϋπολογιστές ή μικροδιερευνητές, προσωπικοί ή οικογενειακοί υπολογιστές, έχουν μεταβάλει εξαιρετικά τον τρόπο επεξεργασίας των στατιστικών δεδομένων. Αυτό επιτυγχάνεται με την άμεση επικοινωνία του χρήστη με τον ηλεκτρονικό υπολογιστή ή τον ηλεκτρονικό διερευνητή χάρη στην δημιουργία ειδικών ανωτέρων γλωσσών προγραμματισμού, όπως η BASIC, FORTRAN, PASCAL, COBOL κλπ. Είναι τέτοια η πρόοδος της μικροηλεκτρονικής και της επιστήμης των ηλεκτρονικών υπολογιστών σε συνδυασμό και με την ανάπτυξη της τεχνολογίας ώστε σε 10-15 χρόνια να έχουμε αλληπάλληλες γενιές μικροϋπολογιστών τεραστίων δυνατοτήτων και απεριόριστης μνήμης. Το σχετικά μικρό κόστος αποκτήσεως ενός μικροϋπολογιστή με μεγάλες δυνατότητες προγραμματισμού και μνήμης και ακόμη η ευχέρεια επικοινωνίας μεταξύ των περισσότερων μικρών μονάδων οδήγησαν στην δημιουργία δικτύων μικροϋπολογιστών ή μικροδιερευνητών στη θέση των μεγάλων κεντρικών μονάδων ηλεκτρονικών υπολογιστών. Η επιστήμη από την άλλη πλευρά της Πληροφορικής βοήθησε σημαντικά στη διάδοση της στατιστικής πληροφορίας και στην δυνατότητα επεξεργασίας ογκώδους στατιστικής ύλης σε εξαιρετικά σύντομο χρόνο και χωρίς λάθη χάρη στην ύπαρξη κατάλληλου λογισμικού.

Εκτός από την επεξεργασία του στατιστικού υλικού, της ταξινομήσεως του δηλαδή, και το υπολογιστικό μέρος της στατιστικής αναλύσεως έχει διευκολυνθεί αφάνταστα από τη χρήση των ηλεκτρονικών διερευνητών. Μπορούμε αβίαστα να υποστηρίξουμε ότι σήμερα δεν υπάρχει πλέον θέμα υπολογισμών και μαθηματικών πράξεων στην επίλυση των στατιστικών προβλημάτων. Έτοιμα προγράμματα, κατάλληλα προσαρμοσμένα, παρέχουν σε ελάχιστο χρόνο τη λύση στατιστικών προβλημάτων της περιγραφικής ή συμπερασματολογικής στατιστικής.

Εκείνο βεβαίως που δεν είναι δυνατόν να υποκατασταθεί, είναι η γνώση της στατιστικής μεθοδολογίας για την χρήση του κατάλληλου προγράμματος και φυσικά η ερμηνεία των αποτελεσμάτων, τα οποία ο διερευνητής παρουσιάζει στην οθόνη ή στον εκτυπωτή.

Από τα εκτεταμένα, έτοιμα και οργανωμένα προγράμματα διαχείρισης της στατιστικής ύλης, τα οποία επικράτησε να λέγονται στατιστικά πακέτα (statistical packages) είναι τα SPSS (Statistical Package for Social Sciences), το BMDP (Biomedical Computer Programs), το MINITAB και το EGRET (Epidemiological Graphics Estimation and Testing Package).

2.2 Η στατιστική στην ιατρική και στις βιολογικές επιστήμες (Βιοστατιστική)

Πάρα πολλοί επιστημονικοί κλάδοι χρησιμοποιούν εκτεταμένα την στατιστική. Η εφαρμογή των στατιστικών μεθόδων στην ιατρική και τις βιολογικές επιστήμες είναι σχετικά πρόσφατη και ονομάζεται βιοστατιστική ή και βιομετρία. Μόλις κατά τα τέλη της προηγούμενης εκατονταετίας άρχισαν να χρησιμοποιούνται συστηματικά οι στατιστικές μέθοδοι στην αξιολόγηση της ιατρικής έρευνας. Αυτό οφείλεται στο γεγονός ότι επί αιώνες η ιατρική και η βιολογία είχαν κατά κάποιο τρόπο στρέψει τα νώτα στην μέτρηση. Η σπουδή των ιατρικών προβλημάτων στηριζόταν στην παρατήρηση και η γνώση των βιομετρικών μεθόδων φαινόταν ότι δεν ήταν απαραίτητη. Οι ιατροί, οι φαρμακοποιοί, οδοντίατροι, κτηνίατροι αλλά και οι βιολόγοι μέχρι μιας όχι απομακρυσμένης εποχής δυσπιστούσαν στην ιδέα της χρησιμοποίησεως των μαθηματικών στις επιστήμες τους. Στην αρχή φαινόταν δυσχερής η παραδοχή του στατιστικού υπολογισμού ως μέσου έρευνας με τον ίδιο ακριβώς τρόπο με τον οποίο είχε χρησιμοποιηθεί ως τότε το μικροσκόπιο, ο ζυγός και η κλινική παρατήρηση. Όπως όμως συμβαίνει με όλες τις εφαρμοσμένες επιστήμες, έτσι και η ιατρική και η βιολογία στηρίζονται σήμερα κυρίως στις μετρήσεις, με τις οποίες οι γνώσεις αποβαίνουν ακριβέστερες, συστηματοποιούνται αλλά και σταθεροποιούνται στην ανθρώπινη μνήμη.

Από τις μετρήσεις, ύστερα από κατάλληλη κριτική ανάλυση, συνάγονται τα συμπεράσματα και καθορίζονται οι τακτικότητες, οι οποίες διέπουν τις επιστήμες. Η βιοστατιστική είναι, στην πραγματικότητα, δύο λέξεις - και δύο τομείς της μελέτης - που συνδυάζεται. Το βιο (bio) μέρος περιλαμβάνει τη βιολογία, τη μελέτη των πραγμάτων διαβίωσης. Το μέρος στατιστικών (statistics) περιλαμβάνει τη συσσώρευση, την καταδίωξη, την ανάλυση, και

την εφαρμογή των στοιχείων. Επίσης είναι η χρήση των διαδικασιών και της ανάλυσης στατιστικών στη μελέτη και πρακτική της βιολογίας. Υπό αυτήν τη μορφή, έχει πολλές πραγματικές και επιστημονικές εφαρμογές.

Η Βιοστατιστική χρησιμοποιείται συνήθως στα πειράματα της βιολογίας. Το στοιχείο συγκεντρώνεται και αναλύεται πριν από, κατά τη διάρκεια, και μετά από ένα πείραμα της βιολογίας, με την πρόθεση να έρθει σε κάποια μορφή λογικής κατάληξης. Αφ' ενός, ένα πείραμα βιοστατιστικής μπορεί να είναι εξ ολοκλήρου μαθηματικό παραδείγματος χάριν, η μέτρηση της θερμοκρασίας ενός ζώου στους διάφορους χρόνους της ημέρας.

Επειδή οι ερευνητικές υποθέσεις και σενάρια στις επιστήμες της βιολογίας και της ιατρικής είναι ποικίλες, η βιοστατιστική συμπεριλαμβάνει γενικότερα κάθε ποσοτική, και όχι μόνο στατιστική, προσέγγιση που μπορεί να χρησιμοποιηθεί για να απαντήσει σε ερευνητικά ερωτήματα ή να ελέγξει την ορθότητα επιστημονικών θεωριών, υποθέσεων και σεναρίων. Ο Σχεδιασμός και η ανάλυση των κλινικών δοκιμών είναι ίσως η περισσότερο γνωστή εφαρμογή των στατιστικών μεθόδων στη ιατρική. Τα εκπαιδευτικά προγράμματα Βιοστατιστικής είναι σχεδόν αποκλειστικά μεταπτυχιακά. Συχνά βρίσκονται στα σχολεία της δημόσιας υγείας, που συνδέονται με τις σχολές της ιατρικής, της δασονομίας, ή της γεωργίας ή ως ειδικές κατευθύνσεις εφαρμογής στα τμήματα Στατιστικής επιστήμης.

Επιπλέον, πολλά πανεπιστημιακά τμήματα σχετικά με τις οικολογικές επιστήμες διοργανώνουν μια σειρά μαθημάτων βιοστατιστικής που εισάγει έννοιες όπως αυτή των ελέγχων υποθέσεων για μία μεταβλητή ή για πολυμεταβλητά δεδομένα ενός ή περισσότερων δειγμάτων ή ομάδων. Συχνά αυτά τα μαθήματα συνδιάζονται με μαθήματα Πειραματικού Σχεδιασμού. Πολλοί επιστήμονες της στατιστικής δυσανασχετούν για το διαχωρισμό της Βιοστατιστικής από τη Στατιστική με την έννοια ότι ένας στατιστικός μπορεί να χειριστεί και Ιατρικά προβλήματα. Όμως η πράξη στις χώρες της Ευρώπης και τις Η.Π.Α έδειξε ότι χρειάζονται εξειδικευμένες γνώσεις (τόσο στην ορολογία όσο και στην ανάλυση) που ένας στατιστικός δεν γνωρίζει. Αυτό, σε συνδυασμό με αυξημένη ζήτηση Βιοστατιστικών επιστημόνων από Ιατρικές και φαρμακευτικές εταιρείες, οδήγησε στη δημιουργία εντατικών

μεταπτυχιακών προγραμμάτων βιοστατιστικής ή ιατρικής στατιστικής σε πολλά πανεπιστήμια της Ευρώπης και των ΗΠΑ εκ των οποίων αρκετά είναι άμεσα συνδεδεμένα με τη διαδικασία της παραγωγής και της Ιατρικής έρευνας.

Οι μετρήσεις και άλλες χρήσεις της βιοστατιστικής, όχι πάντα, περιλαμβάνουν τέτοια εγκόσμια πράγματα. Η βιοστατιστική χρησιμοποιείται συνήθως στην μεγάλη κλίμακα προσπάθειες εξεύρεσης λύσεων, όπως η δοκιμή φαρμάκων. Σε αυτή την περίπτωση των δοκιμών για τα νέα φαρμακευτικά είδη, η βιοστατιστική στηρίζεται σε μεγάλο ποσοστό επάνω στη διαδρομή και ερμηνεύει τα στοιχεία και για να καταστήσει τις συστάσεις βασισμένες σε εκείνες τις ερμηνείες.

Μια άλλη ζωτικής σημασίας χρήση της βιοστατιστικής είναι στην αξιολόγηση της εξάπλωσης μιας ασθένειας. Οι επιστήμονες διεξάγουν δοκιμές στους ανθρώπους που έχουν συμβληθεί μια ασθένεια – το δείγμα – και συγκρίνουν το DNA, την ιστορία ζωής, και τους κοινωνικούς όρους τους με άλλους που ζουν στην ίδια περιοχή – το υπόλοιπο του πληθυσμού – προκειμένου να δουν γιατί μερικοί άνθρωποι πήραν την ασθένεια και άλλοι όχι. Κατά αυτόν τον τρόπο, η βιοστατιστική μπορεί να λύσει ορισμένα περιβαλλοντικά ή βιολογικά μυστήρια. Μια σχετική χρήση της βιοστατιστικής είναι στη γενετική έρευνα. Δείγματα, πληθυσμοί, πειράματα, έρευνα - όλα γίνονται στο όνομα της συνέχισης των θεραπειών για τις θανάσιμες ασθένειες και τις προβλέψεις των πιθανοτήτων.

Η βιοστατιστική χρησιμοποιείται επίσης στη διαμόρφωση και την υπόθεση. Λαμβάνοντας υπόψη ένα σύνολο στοιχείων, οι επιστήμονες συνδυάζουν τη θεωρία βιοστατιστικής και πιθανότητας προκειμένου να αποφασιστεί η πιθανότητα των ασθενειών για να χτυπήσουν τους πληθυσμούς, τα φάρμακα για να θεραπεύσουν εκείνες τις ασθένειες, και την αντίδραση των ανθρώπων σε εκείνα τα φάρμακα. Κατά αυτόν τον τρόπο, η βιοστατιστική υπόσχεται να είναι τόσο καλή στην πρόβλεψη του μέλλοντος όσο είναι και στην ανάλυση του παρελθόντος.

Η συλλογή, η κατάταξη και η επεξεργασία μιας σειράς παρατηρήσεων των βιολογικών φαινομένων, του φυσικού και κοινωνικού περιβάλλοντος ήταν

ανέκαθεν το βασικότερο γνώρισμα της επιστήμης, ακόμη και στην στοιχειοδέστερή της μορφή. Αρχικά η επιστημονική έρευνα στηρίχθηκε, όπως ήταν φυσικό, στην ποιοτική ανάλυση των παρατηρήσεων. Με την πάροδο όμως του χρόνου η παραδοσιακή μέθοδος της ποιοτικής αναλύσεως των παρατηρήσεων συμπληρώθηκε ή και αντικαταστήθηκε βαθμιαία και οι εφαρμοσμένες επιστήμες στράφηκαν προς τις περισσότερο ακριβείς και συγκεκριμένες ποσοτικές μετρήσεις και στην στατιστική μεθοδολογία.

Η κατάταξη των μετρήσεων σε λειτουργικά σχήματα οδήγησε στην ανάπτυξη της στατιστικής ως μεθόδου συλλογής και εκθέσεως του στατιστικού υλικού. Επειδή όμως η επιστημονική μέτρηση είναι στην πραγματικότητα σύγκριση, η στατιστική, ως επιστήμη και ως μεθοδολογία έρευνας, διαμορφώθηκε και εξελίχθηκε τελικά ως επιστήμη της αιτιολογικής αξιολογήσεως των μετρήσεων και των συγκρίσεων. Στο πεδίο της περιλαμβάνεται σήμερα πλέον όλη η επιστημονική μεθοδολογία συλλογής, οργανώσεως και αναλύσεως των ποσοτικών και των ποιοτικών δεδομένων του πειραματισμού και της παρατηρήσεως. Με αυτό το εννοιολογικό περιεχόμενο η στατιστική ανάλυση αποτέλεσε και την απαραίτητη βοηθητική μέθοδο σε κάθε σχεδόν μορφή έρευνας της σημερινής περιόδου. Στην ιατρική και τις βιολογικές επιστήμες ειδικότερα, η στατιστική προσφέρει τις κατάλληλες βιομετρικές μεθόδους για την εκτίμηση και αξιολόγηση της βιολογικής μεταβλητότητας.

Το χαρακτηριστικότερο γνώρισμα της εποχής την οποία διατρέχουμε είναι οι αριθμοί. Ο ερευνητής συναντά τους αριθμούς σε κάθε βήμα της καθημερινής επιστημονικής εμπειρίας του και δεν μπορεί να παρατηρεί και να πειραματίζεται χωρίς προκαθορισμένο σχεδιασμό και οργάνωση της έρευνας και του πειράματος. Ο ιατρός, κλινικός και εργαστηριακός, και ο βιολόγος δεν μπορούν να ανοίξουν ένα επιστημονικό σύγγραμμα ή ιατρικό περιοδικό χωρίς να συναντήσουν μετρήσεις, συγκρίσεις, στατιστική ορολογία και γενικά χωρίς να έρθουν σε επαφή με την στατιστική μεθοδολογία.

Η στατιστική είναι σήμερα μια αυτόνομη επιστήμη με την οποία ασχολούνται χιλιάδες ερευνητές. Ταυτόχρονα είναι ένας βοηθητικός κλάδος όλων των επιστημών. Κάθε χρόνο βλέπει το φως της δημοσιότητας τεράστιος

αριθμός αμιγών στατιστικών εκδόσεων, αλλά και σε κάθε σχεδόν επιστημονική εργασία ή διατριβή είναι εμφανής η συμβολή της στατιστικής επεξεργασίας για την εμπέδωσή της. Ως μεθοδολογία αποτελεί απαραίτητο εφόδιο του ιατρού και του βιολόγου στην οργάνωση της δικής του εργασίας και την τεκμηρίωση των ευρημάτων. Αλλά και η επιστημονική δραστηριότητα των άλλων επιβάλλει την στοιχειώδη στατιστική μόρφωση, αφού μόνο έτσι είναι εφικτή η παρακολούθηση και κατανόηση, αλλά και η αξιολόγηση του έργου των υπόλοιπων εργατών της επιστήμης.

Γενικά η στατιστική στην ιατρική πρακτική περιλαμβάνει:

1. τον σχεδιασμό πειραμάτων και μελετών
2. τη συλλογή δεδομένων
3. την επεξεργασία δεδομένων
4. την ανάλυση δεδομένων
5. την παρουσίαση δεδομένων και αποτελεσμάτων
6. την ερμηνεία αποτελεσμάτων

Οι λόγοι ύπαρξης της είναι:

1. διεξάγουμε επιστημονική έρευνα
2. κατανοούμε και εκτιμούμε σωστά τα πορίσματα άλλων μελετών
3. καταλαβαίνουμε την αιτιολόγηση σε απόψεις άλλων επιστημόνων
4. κρίνουμε πόση αξιοπιστία υπάρχει σε κάθε πόρισμα
5. κάνουμε κριτική στα επιχειρήματα των άλλων

2.2.1 Η βιοστατιστική, ως πηγή ποιοτικών και ποσοτικών δεδομένων που επηρεάζουν την γεωγραφία της υγειονομικής φροντίδας

Η ιατρική δημογραφία (βιοστατιστική), είναι μια κατ' εξοχήν εφαρμοσμένη ανθρωπο – βιολογική επιστήμη, η οποία θεωρείται σήμερα ένας

κλάδος της υγιεινής. Χρησιμοποιεί ποιοτικού και ποσοτικού χαρακτήρα στοιχεία για την μέτρηση, περιγραφή και ανάλυση των διαφόρων χαρακτηριστικών του πληθυσμού. Η μελέτη των πληθυσμιακών γνωρισμάτων κατευθύνεται πρωτίστως στη διαπίστωση του όγκου και της γεωγραφικής διασποράς του πληθυσμού, της κατανομής του κατά φύλο και της δομής του κατά ηλικία, της αναπαραγωγικότητας, της νοσηρότητας και της θνησιμότητας του.

Η ιατροδημογραφική έρευνα στηρίζεται στην ύπαρξη αξιόπιστων αριθμητικών στοιχείων. Η συλλογή του πρωτογενούς στατιστικού υλικού είναι βασική προϋπόθεση για τη διεξαγωγή κάθε ιατροδημογραφικής μελέτης. Οι βασικές πηγές συλλογής των ιατροδημογραφικών στοιχείων είναι οι απογραφές του πληθυσμού, οι ληξιαρχικές πράξεις, τα βιοστατιστικά στοιχεία των υγειονομικών κέντρων, των νοσηλευτηρίων και των ασφαλιστικών οργανισμών, καθώς επίσης και οι ειδικές έρευνες.

2.3 Ασφάλεια ιατρικών δεδομένων

2.3.1 Ιατρικός φάκελος

Ο όρος «ιατρικός φάκελος» είναι κατανοητός σε όλους. Το περιεχόμενο ενός ιατρικού φακέλου αφορά σε έγγραφα σχετικά με την κατάσταση της υγείας ενός ασθενούς: παραπεμπτικά εξετάσεων, καταγραφή στοιχείων νοσηλείας, αποτελέσματα απεικονιστικών και εργαστηριακών εξετάσεων, διαγνώσεις, χορήγηση αγωγών κ.λπ. Πρακτικά, ο ιατρικός φάκελος αποτελεί το μέσο επικοινωνίας ανάμεσα στο ιατρικό και παραϊατρικό προσωπικό που ασχολείται με έναν συγκεκριμένο ασθενή. Οδηγίες θεραπείας, παραπεμπτικά με ειδικές οδηγίες, καταγραφή της πορείας μίας νόσου κ.λπ. δρομολογούνται στους διαφόρους εμπλεκόμενους, που δεν έχουν την δυνατότητα της μεταξύ τους άμεσης επικοινωνίας, μέσω του ιατρικού φακέλου. Κατά την περίοδο αντιμετώπισης του προβλήματος, ο ιατρικός φάκελος αποτελεί το σημείο αναφοράς στο οποίο ανατρέχει κάποιος για να έχει μια εικόνα της κατάστασης του ασθενούς. Οι εμπλεκόμενοι σε ένα ιατρικό περιστατικό γνωρίζουν ότι για να δουν το αποτέλεσμα μίας εξέτασης πρέπει να ανατρέξουν στον ιατρικό

φάκελο του ασθενή. Ανεπίσημα, ο ιατρικός φάκελος χρησιμεύει ως «χώρος εργασίας», όπου καταγράφονται ιδέες και εντυπώσεις για το πρόβλημα ενός ασθενή, καθώς και για την πορεία της αντιμετώπισης του προβλήματος.

Με την ολοκλήρωση ενός περιστατικού, ο ιατρικός φάκελος είναι το μέρος όπου φυλάσσονται όλα τα εργαστηριακά και κλινικά δεδομένα για μελλοντική χρήση. Τα δεδομένα, μάλιστα, που έχουν καταγραφεί στον ιατρικό φάκελο μπορούν να χρησιμεύσουν για τον έλεγχο των διαδικασιών που ακολουθήθηκαν κατά τη διάρκεια της θεραπείας του ασθενή.

2.3.1.1 Ιατρικός φάκελος βασισμένος σε χαρτί

Δεν υπάρχει αμφιβολία ότι για δεκαετίες, αν όχι για αιώνες, η λύση του ιατρικού φακέλου βασισμένου στο χαρτί (paper-based) έχει χρησιμοποιηθεί με σχετική επιτυχία. Η μέθοδος που χρησιμοποιεί ως βάση το χαρτί, υλικό με το οποίο είναι εξοικειωμένοι οι περισσότεροι άνθρωποι και σίγουρα οι ασχολούμενοι με τα ιατρικά πράγματα, παρουσιάζει ως βασικό πλεονέκτημα αφενός την πιο πάνω εξοικείωση και αφετέρου το αυτόνομο της μεθόδου (π.χ. δεν είναι απαραίτητη η ύπαρξη υπολογιστή, ρεύματος, πρίζας, για να ανακτήσει κανείς το περιεχόμενο ενός φακέλου).

Το χαρτί, ωστόσο, από μόνο του ως υλικό παρουσιάζει κάποια σημαντικότερα μειονεκτήματα:

1. Μπορεί να καταστραφεί εύκολα, ενώ είναι αρκετά επίπονη και χρονοβόρα η διαδικασία
2. Δεν δημιουργούνται εύκολα αντίγραφα ασφαλείας.
3. Φθείρεται με τη χρήση ή το χρόνο, οπότε έχει περιορισμένη διάρκεια ζωής.
4. Είναι διαθέσιμο σε ένα μόνο μέρος την ίδια στιγμή.

Παρόλα αυτά, τα σημαντικότερα προβλήματα που εμφανίζονται από την χρήση ενός paper-based ιατρικού φακέλου δεν οφείλονται στο βασικό χρησιμοποιούμενο υλικό, αλλά στο τι αυτό συνεπάγεται. Οι ιατρικοί φάκελοι απαιτούν ιδιαίτερα πολύ χρόνο για να βρεθούν, κυρίως σε μεγάλους

οργανισμούς (νοσοκομεία κ.λπ.), ενώ συνήθως το περιεχόμενό τους είναι διασκορπισμένο σε διαφορετικά σημεία (γραφεία ιατρών, διαγνωστικά και εξεταστικά κέντρα κ.λπ.).

Ένα άλλο σημαντικό μειονέκτημα είναι ότι η πληθώρα ιατρικών φακέλων, τόσο σε ένα ιατρείο όσο και σε έναν οργανισμό, λειτουργεί εις βάρος του σε ό,τι αφορά το κόστος, καθώς απαιτείται μεγάλος αριθμός σε ανθρώπινο δυναμικό για τη σωστή αποθήκευση και ταξινόμηση χιλιάδων πιθανά φακέλων. Αυτός είναι και ο σημαντικότερος λόγος για τον οποίο πάρα πολλά νοσηλευτικά ιδρύματα και οργανισμοί διατηρούν αρχείο για ένα μικρό σχετικά χρονικό διάστημα (σε πολλά νοσοκομεία τα αρχεία φυλάσσονται για περίπου μία πενταετία).

Έχει αποδειχθεί ότι το κλινικό προσωπικό κατ' εξακολούθηση αποτυγχάνει στην ανεύρεση πληροφοριών από ένα paper-based ιατρικό φάκελο, κατά τη διάρκεια μιας συνεδρίας με ασθενή: σε μελέτη 168 περιπτώσεων αποδείχθηκε ότι αναζητήθηκαν και δε βρέθηκαν πληροφορίες σε ποσοστό 81%. Στο 95% αυτών των περιπτώσεων ο ιατρικός φάκελος δεν ήταν διαθέσιμος κατά τη διάρκεια της συνεδρίας.

Τα ποσοστά ανά κατηγορία μη διαθέσιμης πληροφορίας ήταν 36% για πληροφορίες που αφορούσαν σε εργαστηριακές εξετάσεις και πράξεις, 23% για φαρμακευτική και θεραπευτική αγωγή, 31% για ιατρικό ιστορικό και 10% για άλλες πληροφορίες. Στην παραπάνω ανάλυση δεν αναφέρονται οι όποιες πιθανές συνέπειες δημιουργήθηκαν σε ό,τι αφορά την εξέλιξη και το αποτέλεσμα της θεραπείας που ακολουθήθηκε για τους συγκεκριμένους ασθενείς.

Είναι σημαντικό να τονίσουμε ότι στην περίπτωση του paper-based ιατρικού φακέλου δεν είναι δυνατό να μιλάμε για την άμεση χρησιμοποίηση του περιεχομένου πληθώρας ιατρικών φακέλων για έρευνα, μιας και κάτι τέτοιο απαιτεί όχι μόνο ιδιαίτερη προσπάθεια (ανάγνωση των φακέλων, κωδικοποιημένη καταγραφή των στοιχείων τους κ.λπ.), αλλά κυρίως χρόνο. Τέλος, θα πρέπει να αναφερθούμε στις ανάγκες που προκύπτουν από τη σύγχρονη αντίληψη γύρω από τη διοίκηση και τη διαχείριση, οι οποίες απαιτούν τη χρήση όσο το δυνατόν περισσότερων πληροφοριών που θα

λειτουργήσουν υποστηρικτικά σε αποφάσεις που αφορούν στη διαχείριση ενός οργανισμού παροχής ιατρικών υπηρεσιών (νοσοκομειακό ίδρυμα κ.λπ.), αλλά και ολόκληρου του συστήματος υγείας μιας περιοχής ή ακόμα και ενός κράτους.

2.3.1.2 Ηλεκτρονικός ιατρικός φάκελος

Ο τομέας της Υγείας θεωρείται ο μεγαλύτερης εντάσεως πληροφορίας τομέας του κόσμου. Υπολογίζεται ότι κάθε χρόνο δαπανούνται στις ΗΠΑ πάνω από 450 δισ. δολάρια για τη διαχείριση ιατρικών πληροφοριών, σχεδόν το ένα τρίτο του συνολικού ετήσιου προϋπολογισμού της βιομηχανίας της Υγείας. Η διείσδυση των τεχνολογιών αιχμής στον ιατρικό κόσμο καταρρίπτει τα αντικειμενικά εμπόδια του παρελθόντος, που δημιουργούν τόσο η απόσταση όσο και ο χρόνος, ενώ παράλληλα προσφέρει τα απαραίτητα εργαλεία και τις μεθόδους για την παροχή αναβαθμισμένων υπηρεσιών πρωτοβάθμιας υγείας. Η Κοινωνία των Πληροφοριών αλλάζει ριζικά τον τομέα της Υγείας, εισάγοντας αλλαγές στη διαχείριση των συστημάτων αρχειοθέτησης ιατρικών δεδομένων, οι οποίες στοχεύουν στη σωστή διαχείριση των ιατρικών πληροφοριών ενός ασθενούς.

Παράλληλα, στην ιατρική επιστήμη, όλο και περισσότερο διαπιστώνεται η ανάγκη για τη συγκέντρωση όσο το δυνατόν μεγαλύτερου όγκου πληροφοριών για την πληρέστερη εικόνα της υγείας ενός ασθενή. Έχει δημιουργηθεί, δηλαδή, η ανάγκη για εύκολη πρόσβαση στο σύνολο των δεδομένων ενός ιατρικού φακέλου, η επεξεργασία των οποίων θα έχει σαν αποτέλεσμα τη διεξαγωγή σωστών συμπερασμάτων σε ό,τι αφορά την εικόνα των παρελθόντων ιατρικών εξετάσεων και των μελλοντικών ενεργειών που αφορούν τον ασθενή.

Η πληροφόρηση παρέχει ένα μέσο επικοινωνίας μεταξύ των ιατρών του ίδιου ή ακόμα και διαφορετικών νοσοκομείων ή ακόμα και ιατρείων, με απώτερο σκοπό την καλύτερη διάγνωση, καθώς και την άμεση περίθαλψη του ασθενούς, όπου και όποτε αυτό κρίνεται απαραίτητο. Ο εκάστοτε θεράπων ιατρός μπορεί πλέον να έχει πρόσβαση κάθε στιγμή στα αρχεία των ασθενών

του, είτε από το γραφείο του είτε από τον χώρο που του παρέχει το νοσοκομείο με το οποίο συνεργάζεται.

Στην ουσία, ο ηλεκτρονικός ιατρικός φάκελος δύναται να καταστεί ιδιαίτερα χρήσιμος οδηγός τόσο σε μία πιθανή διάγνωση από τον ίδιο ή από διαφορετικό θεράποντα ιατρό, όσο και σε κάποια μελλοντική περίθαλψη του ασθενούς.

Τα ιατρικά αρχεία ενός ασθενούς αποτελούν ιδιαίτερα ευαίσθητα προσωπικά δεδομένα. Αυτό σημαίνει ότι όποιος τα αναλύει ή έχει πρόσβαση σε αυτά, πρέπει να είναι άτομο το οποίο δε θα τα χρησιμοποιήσει προς όφελος του. Η ασφάλεια των ιατρικών δεδομένων είναι ένα σημαντικότερο θέμα για το οποίο, ωστόσο, η τεχνολογία έχει δώσει ουσιαστικές λύσεις, οι οποίες μάλιστα μπορεί να θεωρηθούν αποτελεσματικότερες από αυτές που μέχρι σήμερα εφαρμόζονται για την τήρηση και φύλαξη των ιατρικών φακέλων των ασθενών. Στον ηλεκτρονικό ιατρικό φάκελο πρέπει να δίδεται ιδιαίτερη έμφαση στην Προστασία των Προσωπικών Δεδομένων τα οποία θα αρχειοθετούνται. Λόγω της ευαισθησίας των προσωπικών στοιχείων, είναι επιτακτική η ανάγκη να πληρούνται όλες εκείνες οι προϋποθέσεις ασφαλείας που θα εξασφαλίζουν το αδιάβλητο των δεδομένων. Οι βασικές απαιτήσεις ασφαλείας αφορούν στα εξής πέντε διαφορετικά επίπεδα:

1. Οργανωτική Ασφάλεια
2. Φυσική Ασφάλεια
3. Ασφάλεια Υλικού
4. Ασφάλεια Λειτουργικού Συστήματος
5. Ασφάλεια Εφαρμογής.

2.3.1.3 Ανάγκη για το παρόν – Βάση για το μέλλον

Ο σημερινός πολίτης έχει γίνει αρκετά πιο απαιτητικός σε σχέση με το παρελθόν σε ότι αφορά τις υπηρεσίες που του προσφέρονται, ιδιαίτερα σε έναν χώρο τόσο ευαίσθητο όσο αυτός της υγείας. Έχει την απαίτηση από τον

ιατρό του να είναι ενήμερος για την κατάσταση της υγείας του, καθώς και να ενημερώνει και τον ίδιο αποτελεσματικά, μεταφέροντας του κατ' αυτόν τον τρόπο το συναίσθημα της ασφάλειας.

Ουσιαστικά, ο ηλεκτρονικός ιατρικός φάκελος, ακριβώς λόγω της πληρότητας που μπορεί να τον χαρακτηρίζει, δεν αφορά μόνο στον ασθενή πληθυσμό, αλλά σε όλους όσους επιθυμούν να έχουν τα ιατρικά δεδομένα τους αρχειοθετημένα. Αφορά σε ανθρώπους οι οποίοι ταξιδεύουν πολύ για επαγγελματικούς λόγους ή για λόγους αναψυχής και οι οποίοι δεν είναι δυνατό να έχουν μαζί τους τον ιατρικό τους φάκελο. Απευθύνεται ακόμα σε ανθρώπους που ζουν σε απομακρυσμένες περιοχές, παρέχοντάς τους εύκολη πρόσβαση και μεγαλύτερη ευελιξία στις υπηρεσίες υγείας. Απευθύνεται, επίσης, σε γονείς οι οποίοι επιθυμούν να έχουν τη δυνατότητα παρακολούθησης – διαχείρισης των ιατρικών φακέλων των παιδιών τους, δίνοντάς τους παράλληλα πολύτιμες συμβουλές και εν κατακλείδι παρέχοντάς τους μεγαλύτερη ασφάλεια. Ένας πλήρης ιατρικός φάκελος αποτελεί από μόνος του στοιχείο ασφάλειας προς το πρόσωπο το οποίο αφορά, αφού του παρέχει τη δυνατότητα να παρακολουθεί με σωστό τρόπο την πορεία της υγείας του.

Ο ηλεκτρονικός ιατρικός φάκελος ασθενούς, εφόσον είναι σωστά δομημένος και συμπληρωμένος, δε μπορεί παρά να αποτελεί ανάγκη για το παρόν, ανάγκη η οποία θα γίνει βάση για το μέλλον.

ΚΕΦΑΛΑΙΟ 3

3. Στατιστική και Ιατρική

3.1 Οι υπολογιστές στην Ιατρική Επιστήμη

Με την βοήθεια των υπολογιστών η Ιατρική έφθασε εδώ που είναι σήμερα. Γενικά οι υπολογιστές έχουν βοηθήσει όλες τις επιστήμες. Δεν θα μπορούσαν λοιπόν να αφήσουν ανεπηρέαστη και την ιατρική επιστήμη. Ίσως να φανούμε υπερβολικοί αν πούμε ότι στο χώρο της υγείας η βοήθεια που μας παρέχουν, είναι μεγαλύτερη από κάθε άλλη.

Ας σκεφτούμε τα τόσα ιατρικά μηχανήματα, τα μηχανήματα Laser και την μαγνητική τομογραφία, που μπορεί να μελετήσει ο γιατρός, ένα τμήμα του σώματος σαν να το έχει κόψει φέτες, μπορεί να το περιστρέψει και να το κάνει zoom, για να το δει από κοντά. Μέχρι τώρα το μόνο που τα μηχανήματα αυτά δεν μπορούσαν να αποκαλύψουν ήταν η βιοχημεία του τμήματος που μελετούσαν. Τώρα ο τομογράφος εκπομπής ποζιτρονίου, το έκανε και αυτό πραγματικότητα. Έτσι μπορεί να διαγνώσει την επιληψία, την χορεία Huntingtons και διάφορα σύνδρομα. Βοηθάει να μελετάμε τον εγκέφαλο σαν φυσικό όργανο, όταν ακούμε, θυμόμαστε, σκεπτόμαστε, μαθαίνουμε, κινούμαστε. Με κατάλληλα όργανα δε, που λέγονται γωνιόμετρα, μελετάμε την κινητική δραστηριότητα του εγκεφάλου. Βέβαια τα μηχανήματα αυτά είναι ακόμα σε αρχικά στάδια, όμως η εξέλιξή τους υπόσχεται πολλά.

Ας σκεφτούμε ακόμα την ρομποτική ιατρική, που θα μπορούσε ένας χειρουργός να χειρουργήσει από απόσταση χιλιομέτρων (αν υπάρχει ο κατάλληλος εξοπλισμός). Η τηλεϊατρική για την παρακολούθηση από απόσταση ενός ασθενούς. Το ηλεκτρονικό μικροσκόπιο, που ψάχνει μόνο του τα νεοπλασματικά κύτταρα και να τα εντοπίζει. Το ίδιο και τα ηλεκτρονικά διασκόπια. Στους υπερηχογράφους, που έφτασαν να βγάζουν σχεδόν φωτογραφία του εμβρύου. Έχουμε φτάσει στην ευχάριστη θέση να μπορούμε να διαγνώσουμε, περίπου το 80% των παθήσεων του εμβρύου. Στην αρθροπλαστική, ειδικά μηχανήματα αντιγράφουν με ακρίβεια την άρθρωση, ώστε το μόσχευμα να ταιριάζει απόλυτα. Την αποκωδικοποίηση του DNA, δεν θα μπορούσε να τη φαντασθεί κάποιος, αν δεν υπήρχαν οι ηλεκτρονικοί υπολογιστές. Και γενικά μηχανήματα για την διάγνωση και για την θεραπεία των ασθενών, θα μπορούσαμε να μιλάμε για ώρες.

Ο κόσμος που αποκαλύπτει ο υπολογιστής είναι απέραντος, αγγίζει τα όρια της φαντασίας. Μπορεί τυφλοί να δουν το φως τους, με κατάλληλα μηχανήματα (αν η πάθησή τους το επιτρέπει), άτομα βαρήκοα, να μπορούν να ακούν και να χειρίζονται μηχανήματα με οπτικά σήματα και πολλά άλλα.

Αλλά η πιο ελπιδοφόρα ιατρική εφαρμογή μπορεί να είναι η προληπτική ιατρική. Χρησιμοποιώντας δεδομένα της κρυσταλλογραφίας με ακτίνες Χ, ειδικοί προγραμματιστές μπορούν να δημιουργήσουν μοντέλα στον υπολογιστή, που περιγράφουν τις δομές των ιών. Αυτές οι εικόνες μπορούν κάποια μέρα να οδηγήσουν στην προστασία του ανθρώπου, από πάρα πολλές ασθένειες. Ας ελπίσουμε ότι οι ηλεκτρονικοί υπολογιστές θα βοηθήσουν στην θεραπεία του καρκίνου, του AIDS και των άλλων ανιάτων ασθενειών.

Όπως έχει αναφερθεί και στο προηγούμενο κεφάλαιο, οι σύγχρονοι μικροϋπολογιστές ή μικροδιερευνητές, προσωπικοί ή οικογενειακοί υπολογιστές, έχουν μεταβάλει εξαιρετικά τον τρόπο επεξεργασίας των στατιστικών δεδομένων. Φαίνεται λίγο περίεργο, αλλά με την βοήθεια των υπολογιστών, η ιατρική έφθασε εδώ που είναι σήμερα. Γενικά οι υπολογιστές έχουν βοηθήσει όλες τις επιστήμες. Δεν θα μπορούσαν λοιπόν να αφήσουν ανεπηρέαστη και την ιατρική επιστήμη. Ίσως να φανούμε υπερβολικοί αν πούμε ότι στο χώρο της υγείας η βοήθεια που μας παρέχουν, είναι μεγαλύτερη από κάθε άλλη.

3.2 Φύση και δομή της ιατρικής πληροφορίας

Η ραγδαία ανάπτυξη της τεχνολογίας των υπολογιστών και της δικτύωσής τους σε τοπικό αλλά και διεθνές επίπεδο έδωσε τη δυνατότητα άμεσης επικοινωνίας μεταξύ ιατρών. Την επικοινωνία ακολούθησε η αμοιβαιότητα στο μοίρασμα της πληροφορίας, και έτσι δημιουργήθηκαν διεθνώς «κατανεμημένες» βάσεις ιατρικών δεδομένων. Οι εξελίξεις αυτές είχαν ως αποτέλεσμα να διατίθεται σήμερα στην ιατρική κοινότητα ένας τεράστιος όγκος πληροφοριών, στον οποίο η πρόσβαση είναι άμεση.

Ο τεράστιος όγκος πληροφορίας και η αμεσότητα στην πρόσβασή της είναι αναγκαίες συνθήκες για την επίλυση σύνθετων ιατρικών προβλημάτων, δεν είναι όμως ικανές. Πράγματι, όσο μεγαλύτερος είναι ο όγκος της

διατιθέμενης πληροφορίας, τόσο πιο δύσκολη γίνεται η ανεύρεση μιας συγκεκριμένης πληροφορίας. Είναι σαφές ότι για να βρεθεί η συγκεκριμένη πληροφορία θα πρέπει να υπάρχει ο κατάλληλος αλγόριθμος διερεύνησης, μέσω του οποίου θα γίνει ο εντοπισμός της. Εκτός όμως από τη διάσταση που σχετίζεται με την εντόπιση χρήσιμων ιατρικών πληροφοριών, υπάρχει και η διάσταση του συνδυασμού τους για τη λήψη μιας ιατρικής πληροφορίας που αφορά τη διάγνωση, την πρόγνωση ή τη θεραπεία. Η διαχείριση των ιατρικών πληροφοριών κάνει χρήση των Η/Υ, αλλά δεν μένει σ' αυτούς, απαιτεί νέες μεθόδους κωδικοποίησης και ανάλυσης, που συνιστούν τη βάση της «Ιατρικής Πληροφορικής».

Η Ιατρική Πληροφορική παρουσιάζει σοβαρές διαφορές σε σχέση με τις εφαρμογές της Πληροφορικής στις βασικές επιστήμες. Τα φυσικά ή χημικά φαινόμενα περιγράφονται με νόμους που δίνονται συνήθως από αναλυτικές μαθηματικές εκφράσεις (συναρτήσεις). Το ίδιο δεν ισχύει για τα ιατρικά φαινόμενα, που συνήθως αναφέρονται σε παθολογικές λειτουργίες σύνθετων οργάνων για τις οποίες δεν υπάρχει ένας κοινός κώδικας (λέγεται ότι δεν υπάρχουν ασθένειες, υπάρχουν ασθενείς) και, επομένως, κάθε πρόβλεψη ή απόφαση γι' αυτές απορρέει από σύνθετες λογικές διαδικασίες που δεν μπορούν να δοθούν με συστηματικό τρόπο. Συνήθως ο ιατρός, αντίθετα από το βασικό επιστήμονα, λαμβάνει αποφάσεις ακολουθώντας μια μη αναλυτική προσέγγιση, η οποία καλείται «ευρετική» (heuristic) και είναι αντικείμενο μελέτης μιας νέας επιστήμης, που ασχολείται με την Τεχνητή Νοημοσύνη (Artificial Intelligence).

Κατά την ευρετική διαδικασία οι διάφορες πληροφορίες εξετάζονται και έχουν μια σύνθετη μμεταξύ τους αλληλεπίδραση, η οποία καθορίζει την τελική απόφαση. Η ικανότητα των ιατρών στην άσκηση της ευρετικής προσέγγισης ποικίλλει και εξαρτάται μεν από την εμπειρία και την αρτιότητα της εκπαίδευσης, αλλά όχι μόνο από αυτά (συχνά λέμε ότι αυτός ο ιατρός έχει ιατρική διαίσθηση). Σχεδόν πάντα, η ιατρική απόφαση λαμβάνεται σε συνθήκες αβεβαιότητας (μεγάλης ή μικρής). Οι υπολογιστές και η Ιατρική Πληροφορική έρχονται να υποστηρίξουν τη λήψη ιατρικών αποφάσεων, πρώτον, μειώνοντας την αβεβαιότητα και την υποκειμενικότητα και, δεύτερον,

χρησιμοποιώντας με πιο αποδοτικό τρόπο τα υπάρχοντα δεδομένα. Οι εφαρμογές της Ιατρικής Πληροφορικής απαιτούν:

- Πλήθος μαθηματικών εργαλείων ώστε να κωδικοποιηθούν όσο το δυνατόν περισσότερο υπάρχοντα δεδομένα.
- Μεθόδους στατιστικής ανάλυσης, αφού όλες οι ιατρικές μετρήσεις και παρατηρήσεις υπόκεινται σε τυχαία σφάλματα.
- Δημιουργική εφαρμογή της αναλυτικής θεωρίας των αποφάσεων.
- Ανάλυση των γνωστικών μηχανισμών και γνωστική ψυχολογία.

Η ιατρική πληροφορία είναι ένας συνδυασμός σημάτων, το καθένα από τα οποία συνοδεύεται από τυχαίο θόρυβο. Η πρώτη προσπάθεια της Πληροφορικής είναι η μείωση αυτού του θορύβου. Όταν πρόκειται για σήματα που αντιστοιχούν σε εργαστηριακές μεταβλητές ή εικόνες που πρόκειται να υποστούν μια επεξεργασία μέσω υπολογιστών, εφαρμόζονται αλγοριθμικά φίλτρα, που βασίζόμενα στη γνώση του μετρητικού πρωτοκόλλου και της απόκρισης του μετρητικού συστήματος στοχεύουν στο ξεκαθάρισμα του θορύβου από το σήμα. Θόρυβο έχουν όλα τα ιατρικά σήματα, ακόμα και αυτά που προέρχονται από τη φυσική εξέταση, γιατί η περιγραφή, π.χ., ενός συμπτώματος από τον ασθενή εξαρτάται από το μορφωτικό του επίπεδο, την ψυχική του κατάσταση, την ηλικία κτλ. Σ' αυτή την περίπτωση το φιλτράρισμα γίνεται με τη χρήση εναλλακτικών ερωτήσεων (η διαμόρφωση των ερωτηματολογίων είναι μέρος της Ιατρικής Πληροφορικής).

Ο ιατρός στη λήψη μιας ιατρικής απόφασης (διάγνωση, πρόγνωση ή θεραπεία) λαμβάνει υπόψη κλινικές και εργαστηριακές μεταβλητές, οι οποίες υπόκεινται σε στατιστικά σφάλματα και δεν δίνουν σε κάθε περίπτωση (η καθεμία ξεχωριστά) απόλυτη βεβαιότητα στις αποφάσεις του. Μερικές από τις προαναφερόμενες μεταβλητές έχουν μεγαλύτερη και άλλες μικρότερη προβλεπτική αξία, αλλά ο συνδυασμός τους είναι εκείνος που αυξάνει την πεποίθηση του ιατρού προς τη μία ή την άλλη απόφαση. Η συνδυαστική διαδικασία γίνεται συνήθως με έναν τρόπο «ευρετικό» και ακωδικοποίητο (αυτό που καλούμε ιατρική εμπειρία). Η Ιατρική Πληροφορική δίνει τη

δυνατότητα της στατιστικής ταξινόμησης προτύπων (φυσιολογικών ή παθολογικών καταστάσεων) και επιτρέπει τη χρησιμοποίηση εκτεταμένων βάσεων ιατρικών δεδομένων. Έτσι, σε πρώτη φάση ενσωματώνει τις δυνατότητες της «ευρετικής» υπό την μορφή των έμπειρων συστημάτων (expert systems), που τρόπον τινά συγκεντρώνουν και ταξινομούν την υπάρχουσα εμπειρία δίνοντας επιπλέον κανόνες (if-then), που συνήθως ακολουθούνται σε συγκεκριμένες ιατρικές «ρουτίνες». Σήμερα, αναπτύσσονται νέες προσεγγίσεις στη διαχείριση των ιατρικών πληροφοριών, με τη χρήση νευρωνικών δικτύων και δικτύων πεποίθησης κατά Bayes (Bayesian Belief Networks).

3.2.1 Φύση της ιατρικής πληροφορίας

Η διαφορετικότητα της ιατρικής πληροφορίας σε σχέση με τις πληροφορίες των βασικών θετικών επιστημών είναι προϊόν πολλών αιτίων και έχει να κάνει, αφενός ,με τα αντικείμενα της ιατρικής και, αφετέρου, με την τελείως διαφορετική δομή των συσχετίσεών τους. Για να γίνουν πιο συγκεκριμένα τα παραπάνω, ας εξετάσουμε αυτές που θεωρούνται επιστήμες βασικού επιπέδου. Στην ιεραρχική σχέση μεταξύ των επιστημών η φυσική βρίσκεται στη βάση. Η φυσική χαρακτηρίζεται από ένα είδος απλότητας αλλά και γενίκευσης.

Οι έννοιες και οι περιγραφές των αντικειμένων και των μηχανισμών της φυσικής χρησιμοποιούνται απαραίτητα σε όλες τις εφαρμοσμένες επιστήμες, συμπεριλαμβανομένης και της ιατρικής. Οι φυσικοί νόμοι και οι περιγραφές ορισμένων φυσικών διαδικασιών είναι ουσιαστικοί παράγοντες στην ανάλυση και εξήγηση των ιατρικών λειτουργιών. Είναι, για παράδειγμα, απαραίτητο να γνωρίζουμε ορισμένες βασικές έννοιες της Μοριακής Φυσικής για να κατανοήσουμε γιατί το νερό είναι τόσο καλός διαλύτης ή πώς μεταβολίζονται τα θρεπτικά μόρια.

Η εφαρμογή των υπολογιστών για τη λύση κάποιου προβλήματος φυσικής μέσα στα ιατρικά πλαίσια δεν παρουσιάζει διαφορές σε σχέση με τις υπολογιστικές εφαρμογές που συναντώνται στα εργαστήρια φυσικής και

μηχανολογίας. Η χρήση των υπολογιστών στις διάφορες διαδικασίες βασικού επιπέδου (όπως είναι αυτές της φυσικής ή χημείας) είναι παρόμοια και ανεξάρτητη της συγκεκριμένης εφαρμογής. Εάν εξετάζουμε τις διαλυτικές ιδιότητες του νερού, δεν έχει σημασία αν αυτές έχουν να κάνουν με εφαρμογές στη γεωλογία, στη χημεία ή στην ιατρική. Οι διαδικασίες βασικού επιπέδου της φυσικής είναι ιδιαίτερα προσιτές στη μαθηματική κωδικοποίηση, έτσι η χρήση των υπολογιστών σε αυτές τις εφαρμογές απαιτεί μόνο συμβατικό αριθμητικό προγραμματισμό. Στην ιατρική, όμως, υπάρχουν και άλλες διαδικασίες υψηλής πολυπλοκότητας, οι οποίες αναφέρονται σε σύνθετα αντικείμενα, όπως οι οργανισμοί (φυσιολογικοί ή παθολογικοί).

Όταν αναλύονται, περιγράφονται ή καταγράφονται ιδιότητες ή λειτουργίες ανθρώπων, χρησιμοποιούνται περιγραφές αντικειμένων πολύ υψηλής πολυπλοκότητας, η συμπεριφορά των οποίων δεν έχει αντίστοιχο στο χώρο της φυσικής ή της μηχανολογίας. Αυτές οι περιγραφές είναι πολύ δύσκολο να κωδικοποιηθούν χρησιμοποιώντας μαθηματικούς αλγόριθμους και λογισμικά πακέτα, τα οποία εφαρμόζονται τόσο καλά στα βασικά επίπεδα. Από τα προηγούμενα συνάγεται ότι η Ιατρική Πληροφορική περιέχει εφαρμογές που κινούνται μεταξύ της ανάλυσης μηχανισμών χαμηλού επιπέδου και της επεξεργασίας εξαιρετικά σύνθετων φαινομένων. Όταν μελετώνται ολιστικά ανθρώπινοι οργανισμοί (συμπεριλαμβανομένων των στοιχείων της αντίληψης, της αυτοσυνείδησης και της συμπεριφοράς), αναδεικνύονται πολλά και σύνθετα προβλήματα, για τα οποία η συμβατική λογική και τα συμβατικά μαθηματικά είναι δύσκολο να εφαρμοστούν. Γενικά, τα χαρακτηριστικά των αντικειμένων βασικού επιπέδου είναι καθαρά ορισμένα και σαφώς διακριτά, ενώ εκείνα των αντικειμένων υψηλού επιπέδου είναι ασαφώς ορισμένα και όχι με ακρίβεια διακριτά.

3.2.2 Η δομή των ιατρικών δεδομένων

Είναι γνωστό ότι η θεμελίωση μιας επιστήμης απαιτεί αυστηρότητα ορισμών, ώστε να μην υπάρχει σύγχυση εννοιών, και κατόπιν τη δόμηση θεωρημάτων και την ανάπτυξη μιας συμβολογίας που να έχει συνέπεια και να μην αφήνει περιθώρια ασάφειας. Από την άλλη πλευρά, η ιατρική είναι

γνωστή για την έλλειψη ενός τυποποιημένου λεξικού και τυποποιημένης ορολογίας. Είναι αρκετοί αυτοί που πιστεύουν ότι η πραγματική επιστημονική θεμελίωση της ιατρικής είναι αδύνατη εάν δεν λυθεί αυτό το μείζον πρόβλημα. Υπάρχουν, βέβαια, ορισμένοι οι οποίοι αναρωτιούνται αν είναι επιθυμητή μια τέτοια τυποποίηση για την ιατρική, και μάλιστα πιστεύουν ότι κάτι τέτοιο ίσως μειώσει τις δυνατότητες, μιας και τη θεωρούν περισσότερο «τέχνη» παρά επιστήμη.

Η συζήτηση πάνω σε αυτό το θέμα έχει γίνει ιδιαίτερα πιο έντονη τελευταία, με την εισαγωγή των υπολογιστών στη διαχείριση των ιατρικών δεδομένων. Πράγματι, γι' αυτού του είδους τη διαχείριση είναι αναγκαία η ομοιομορφία στην τυποποίηση των δεδομένων και ορισμών. Είναι προφανές ότι χωρίς αυτή την τυποποίηση – ενδεχομένως – να υπάρχει διάσταση απόψεων μεταξύ του παρατηρητή ή του καταγραφέα μιας μεταβλητής και εκείνου που την αναλύει. Η έλλειψη ακρίβειας και τυποποίησης δημιουργεί ακόμα μεγαλύτερα προβλήματα όταν πρέπει να αναλυθούν συνδυαστικά οι εμπειρίες και παρατηρήσεις πολλών ιατρών μαζί σχετικά με την εξέλιξη μιας ασθένειας στον πληθυσμό ή την επίδραση κάποιου φαρμάκου στη θεραπεία. Γενικά, χωρίς μια προκαθορισμένη ορολογία είναι αδύνατον να γίνει μια αυτοματοποιημένη σύνθεση των δεδομένων.

Για παράδειγμα, ένας ιατρός μπορεί να σημειώσει ότι ένας ασθενής έχει «βραχύτητα αναπνοής». Αργότερα, άλλος ιατρός μπορεί να καταγράψει για τον ίδιο ασθενή ότι έχει «δύσπνοια». Αν αυτοί οι δύο όροι δεν χαρακτηριστούν ως συνώνυμοι, τότε ένα αυτοματοποιημένο πρόγραμμα θα αποτύχει στο να υποδείξει ότι ο ασθενής έχει το ίδιο πρόβλημα στις δύο περιπτώσεις. Με δεδομένη αυτή την έλλειψη, οι επιστήμονες στο χώρο της υγείας επικοινωνούν όσο καλύτερα μπορούν χρησιμοποιώντας εντονότατα το στοιχείο της «ευρετικής» αλληλεπίδρασης. Εκ των πραγμάτων, αποδεικνύεται ότι μόνο σε ελάχιστες περιπτώσεις μπορεί να τεθεί σε κίνδυνο η υγεία του ασθενούς λόγω επικοινωνιακών παρεξηγήσεων. Εάν, όμως, πρόκειται να χρησιμοποιηθούν ηλεκτρονικές βάσεις δεδομένων για τη διαχείριση των ασθενών, τότε είναι, αφενός, απαραίτητη η κωδικοποίηση των πληροφοριών και, αφετέρου, ο σαφέστατος ορισμός των δεδομένων.

ΚΕΦΑΛΑΙΟ 4

4.1 Ενότητες Στατιστικής

Η Στατιστική χωρίζεται σε 3 μεγάλες ενότητες

1. Την θεωρία και τις τεχνικές δειγματοληψίας (Sampling Theory) η οποία αντιμετωπίζει το πρόβλημα επιλέγοντας το κατάλληλο δείγμα.
2. Την Περιγραφική Στατιστική (Descriptive Statistics) η οποία χρησιμοποιείται για την ταξινόμηση και επεξεργασία των δεδομένων του δείγματος μας. Ο σκοπός μας είναι ο σχηματισμός μιας συνοπτικής παρουσίασης του δείγματος μας με πίνακες και γραφήματα από την οποία μπορούν να προκύψουν εύκολα διάφορα συμπεράσματα πάνω στο δείγμα μας. Στόχος της περιγραφικής στατιστικής είναι η ανάπτυξη μεθόδων για τη συνοπτική και την αποτελεσματική παρουσίαση των δεδομένων. Για το σκοπό αυτό έχουν αναπτυχθεί μέθοδοι πινακοποίησης δεδομένων, μέθοδοι γραφικής παρουσίασης δεδομένων και αριθμητικά περιγραφικά μέτρα τα οποία βοηθούν στην περιγραφή της κατανομής των δεδομένων.

Για τις ποσοτικές μεταβλητές η περιγραφική στατιστική προσφέρει τις ακόλουθες δυνατότητες:

- Κατασκευή πίνακα συχνοτήτων

Ο πίνακας συχνοτήτων μιας ποσοτικής μεταβλητής περιλαμβάνει τις συχνότητες, τις σχετικές συχνότητες, τις αθροιστικές συχνότητες, και τις σχετικές αθροιστικές συχνότητες των τιμών της.

- Κατασκευή διαγραμμάτων

Διάγραμμα και πολύγωνο συχνοτήτων, σχετικών συχνοτήτων, αθροιστικών συχνοτήτων και σχετικών αθροιστικών συχνοτήτων, ιστόγραμμα, φυλλογράφημα, θηκόγραμμα.

- Υπολογισμός μέτρων θέσης – κεντρικής τάσης

Τα μέτρα θέσης κεντρικής τάσης μας δίνουν πληροφορίες για τη θέση της κατανομής των παρατηρήσεων. Τα πλέον χρησιμοποιούμενα είναι η μέση τιμή και η διάμεσος.

4.1.1 Μέση τιμή

Η μέση τιμή ενός πληθυσμού συμβολίζεται με μ και η μέση τιμή ενός δείγματος με $\bar{\chi}$

$$\bar{\chi} = \frac{1}{\nu} \sum_{i=1}^{\nu} \chi_i = \frac{\sum_{i=1}^k \nu_i y_i}{\nu} = \sum_{i=1}^k f_i y_i \quad (4.1)$$

Από τον ορισμό της μέσης τιμής είναι φανερό ότι αν οι τιμές $\chi_1, \chi_2, \dots, \chi_n$ είναι όλες μεταξύ τους ίσες, θα είναι ίσες με τη μέση τιμή τους. Με τη μέση τιμή επιδιώκεται να ορισθεί ένας «τυπικός εκπρόσωπος των παρατηρήσεων». Το γεγονός όμως ότι στον υπολογισμό συμμετέχει το άθροισμα όλων των παρατηρήσεων την καθιστά ευαίσθητη σε ακραίες-έκτροπες παρατηρήσεις. Κατά συνέπεια η μέση τιμή αποκρύπτει τις έκτροπες παρατηρήσεις. Δηλαδή όταν υπάρχουν έκτροπες παρατηρήσεις, η μέση τιμή δίνει παραπλανητική εικόνα αν θεωρηθεί «τυπικός εκπρόσωπος» των παρατηρήσεων.

Η μέση τιμή έχει τα ακόλουθα πλεονεκτήματα και μειονεκτήματα:

Πλεονεκτήματα της μέσης τιμής:

1. Για τον υπολογισμό της χρησιμοποιούνται όλες οι τιμές.
2. Είναι μοναδική για κάθε σύνολο δεδομένων
3. Είναι εύκολα κατανοητή
4. Ο υπολογισμός της είναι σχετικά εύκολος
5. Αξιοποιείται στη στατιστική συμπερασματολογία

Μειονεκτήματα της μέσης τιμής:

1. Επηρεάζεται πολύ από ακραίες τιμές
2. Μπορεί να μην αντιστοιχεί σε δυνατή τιμή της μεταβλητής
3. Δεν υπολογίζεται για ποιοτικά δεδομένα

4. Είναι δύσκολος ο υπολογισμός της σε ομαδοποιημένα δεδομένα με ανοικτές τις ακραίες κλάσεις

4.1.2 Διάμεσος

Η διάμεσος του δείγματος συμβολίζεται με δ . Είναι το σημείο της κατανομής των παρατηρήσεων κάτω από το οποίο βρίσκεται το 50% των παρατηρήσεων και πάνω από αυτό το υπόλοιπο 50% των παρατηρήσεων. Εκφράζει τη κεντρική θέση της κατανομής των παρατηρήσεων γι 'αυτό συναντάται και ως μέσος θέσης. Αν το πλήθος n των παρατηρήσεων είναι περιττός αριθμός τότε η διάμεσος είναι η ενδιάμεση παρατήρηση, ενώ αν είναι άρτιος αριθμός, τότε η διάμεσος υπολογίζεται από τη μέση τιμή των δύο ενδιάμεσων παρατηρήσεων.

Πλεονεκτήματα της διαμέσου:

1. Είναι εύκολα κατανοητή
2. Δεν επηρεάζεται από ακραίες τιμές
3. Υπολογίζεται και στην περίπτωση που οι ακραίες κλάσεις είναι ανοικτές
4. Ο υπολογισμός της είναι απλός
5. Είναι μοναδική σε κάθε σύνολο δεδομένων

Μειονεκτήματα της διαμέσου:

1. Δεν χρησιμοποιούνται όλες οι τιμές για τον υπολογισμό της
2. Είναι δύσκολη η αξιοποίηση της στη στατιστική συμπερασματολογία
3. Δεν υπολογίζεται για κατηγορικά δεδομένα
4. Για τον υπολογισμό της μπορεί να χρειαστεί παρεμβολή

4.1.3 Διασπορά

Πολύ συχνά δυο ή περισσότεροι μέσοι όροι έχουν την ίδια τιμή όμως προέρχονται από τελείως διαφορετικές κατανομές συχνοτήτων. Η μέση τιμή δεν αρκεί για να αντιπροσωπεύσει και να περιγράψει πλήρως την κατανομή συχνοτήτων διότι αυτή είναι η τιμή του κέντρου μόνο, ενώ η κατανομή συχνοτήτων έχει και διασπορά στο χώρο. Ο καθορισμός του μεγέθους της διασποράς επιτυγχάνεται με διάφορα μέτρα από τα οποία τα σημαντικότερα είναι η τυπική απόκλιση και η διακύμανση.

Το εύρος της διασποράς (range) παριστάνει τη διαφορά μεταξύ της μέγιστης και της ελάχιστης τιμής μιας στατιστικής σειράς. Η κατανόηση του είναι εύκολη αλλά αδυνατεί να περιγράψει ουσιαστικά τη διασπορά της μεταβλητής και για το λόγο αυτό χρησιμοποιείται εμπειρικά στη στατιστική ανάλυση. Τα βασικά του μειονεκτήματα βρίσκονται στον επηρεασμό του από τις δύο ακραίες τιμές οι οποίες μπορεί να διασπείρονται τελείως ανώμαλα και στο ότι δεν λαμβάνονται υπόψη οι υπόλοιπες τιμές της μεταβλητής.

4.1.4 Διακύμανση

Είναι το μέτρο της μεταβλητότητας το οποίο παριστάνει το μέσο όρο των αποκλίσεων όλων των τιμών των παρατηρήσεων ή μετρήσεων από τον αριθμητικό μέσο όρο. Η τιμή η οποία λαμβάνεται με αυτή τη μεθοδολογία ονομάζεται διακύμανση (variance) και συμβολίζεται με s^2

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (4.2)$$

4.1.5 Τυπική απόκλιση

Η τυπική απόκλιση του πληθυσμού συμβολίζεται με σ και του δείγματος με s και μας λέει πόσο μακριά από τη μέση τιμή βρίσκονται οι παρατηρήσεις.

$$s = \sqrt{\frac{1}{v-1} \sum_{i=1}^v (\chi_i - \bar{\chi})^2} = \sqrt{\frac{1}{v-1} \left(\sum_{i=1}^v \chi_i^2 - v \cdot \bar{\chi}^2 \right)} \quad (4.3)$$

Έτσι όταν οι παρατηρήσεις δε διαφέρουν πολύ από τη μέση τιμή τους η τυπική απόκλιση είναι μικρή ενώ μεγαλώνει όσο περισσότερο απομακρύνονται οι παρατηρήσεις από τη μέση τιμή τους. Δηλαδή η τυπική απόκλιση μας δίνει ένα μέτρο της μέσης απόστασης – απόκλισης των παρατηρήσεων από τη μέση τιμή τους. Συνεπώς έχει νόημα να χρησιμοποιείται σε συνδυασμό με τη μέση τιμή.

Πλεονεκτήματα της τυπικής απόκλισης:

1. Για τον υπολογισμό της λαμβάνονται υπόψη όλες οι παρατηρήσεις
2. Έχουν μεγάλη εφαρμογή στη στατιστική συμπερασματολογία
3. Με βάση τη τυπική απόκλιση και τη μέση τιμή, μπορούν να ορισθούν διαστήματα στα οποία βρίσκεται γνωστό ποσοστό παρατηρήσεων

Μειονεκτήματα της τυπικής απόκλισης:

1. Το κυριότερο της μειονέκτημα είναι ότι δεν εκφράζεται στις ίδιες μονάδες με τη μεταβλητή. Το μειονέκτημα αυτό παύει να υπάρχει με τη χρησιμοποίηση της τυπικής απόκλισης
2. Απαιτούνται περισσότερες αλγεβρικές πράξεις για τον υπολογισμό τους απ' ότι στα άλλα μέτρα

4.2 Πίνακες

Το στατιστικό υλικό το οποίο συλλέγεται εκτίθεται σε δυο μορφές, τους στατιστικούς πίνακες και τις γραφικές παραστάσεις ή γραφικές απεικονίσεις ή διαγράμματα.

Η χρησιμότητα των πινάκων και των γραφικών παραστάσεων συνίσταται στην απάλειψη της ανομοιογένειας στην οποία βρίσκονται οι στατιστικές μονάδες και τα χαρακτηριστικά τους, στην ελάττωση του πλήθους των παρατηρήσεων σε λίγους αριθμούς οι οποίοι εύκολα γίνονται αντιληπτοί και στη δημιουργία ευκολονόητων εποπτικών εικόνων οι οποίες έχουν την ιδιότητα να παραμένουν για πολύ χρόνο στη μνήμη.

Η παρουσίαση με πίνακες των διαφόρων στατιστικών δεδομένων είναι εξαιρετικά διαδεδομένη στις επιστημονικές εργασίες. Ο αντικειμενικός της σκοπός είναι η εμφάνιση των στοιχείων σε περιληπτική και κατανοητή μορφή. Η ποικιλία των στατιστικών πινάκων είναι τεράστια παρόλα αυτά η κατασκευή τους διέπεται από ορισμένες γενικές αρχές και κανόνες.

Χαρακτηριστικά πινάκων

Το βασικό κριτήριο της κατασκευής κάθε πίνακα είναι η σαφής και σύντομη έκθεση του υλικού που συλλέχθηκε. Αρχικό χαρακτηριστικό του πίνακα είναι ο τίτλος του ο οποίος πρέπει να είναι όσο γίνεται συντομότερος, σαφέστερος και πλήρης. Αποτελεί κατά κάποιο τρόπο το στοιχείο της ταυτότητας του.

Ο ωφέλιμος χώρος του πίνακα ο οποίος περικλείεται μέσα σε περιθώρια, υποδιαιρείται σε κάθετες στήλες. Στην επικεφαλίδα των στηλών αναγράφεται το χαρακτηριστικό το οποίο προσδιορίζει το περιεχόμενο τους.

Ο κορμός του πίνακα διαιρείται και οριζοντίως σε περισσότερες γραμμές και σε κελιά ανάλογα με τη βάση με την οποία ταξινομούνται οι στατιστικές μονάδες ή τα χαρακτηριστικά τους.

Οι αριθμοί οι οποίοι αναγράφονται στο σώμα του πίνακα παριστάνουν συχνότητες, τον αριθμό δηλαδή των παρατηρήσεων σε κάθε ομάδα τάξη ή κλάση. Το άθροισμα των συχνοτήτων σε κάθε στήλη δίνει το σύνολο των συχνοτήτων ή το συνολικό αριθμό των παρατηρήσεων.

i. Πίνακας Συχνοτήτων

Shoe Size					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	6.0	1	6.7	6.7	6.7
	6.5	2	13.3	13.3	20.0
	7.0	3	20.0	20.0	40.0
	7.5	2	13.3	13.3	53.3
	8.5	2	13.3	13.3	66.7
	9.0	1	6.7	6.7	73.3
	9.5	2	13.3	13.3	86.7
	10.5	2	13.3	13.3	100.0
	Total	15	100.0	100.0	
Total		15	100.0		

4.3 Διαγράμματα – Η αναγκαιότητα των διαγραμμάτων

Ο άλλος τρόπος με τον οποίο εκτίθεται το στατιστικό υλικό είναι οι γραφικές παραστάσεις ή διαγράμματα. Η έκθεση του υλικού με γραφικές παραστάσεις αποτελεί μια εξαιρετικά εύκαμπτη και ωφέλιμη οδό αναλύσεως και ερμηνείας των πολυαρίθμων στατιστικών στοιχείων με οπτική μορφή. Τα διαγράμματα χρησιμοποιούνται ευρύτατα στη καθαρά επιστημονική έρευνα. Ο σκοπός τους έγκειται στο να βοηθήσουν τους αναγνώστες να αποκτήσουν ταχύτερη, σαφέστερη και γενικά πληρέστερη αντίληψη των δεδομένων τα οποία είχαν προηγουμένως αποδοθεί αριθμητικά.

Η παράσταση του στατιστικού υλικού με μορφή πίνακα παρέχει τη μεγαλύτερη δυνατή ακρίβεια εντούτοις δεν είναι πάντοτε επαρκώς εποπτική και δεν δίνει μια συνοπτική εικόνα των ευρημάτων. Αυτό οφείλεται στο μειονέκτημα του πίνακα να περιλαμβάνει συνήθως μεγάλο πλήθος αριθμών. Είναι γνωστό ότι η παρακολούθηση και ανάγνωση πολλών αριθμών αποβαίνει δυσάρεστη για όλους εμάς. Εξάλλου η άμεση και σαφής αντίληψη του περιεχομένου των αριθμών λείπει από πολλά άτομα πολύ περισσότερο μάλιστα όταν οι αριθμοί εκφράζουν τις σχέσεις οι οποίες υπάρχουν μεταξύ τους. Στην ίδια αυτή δυσχέρεια δηλαδή στην έλλειψη ταχείας αντίληψης των μεγεθών θα πρέπει να αποδοθεί και η γρήγορη επισκόπηση των πινάκων οι

οποίοι περιλαμβάνονται μεταξύ του κειμένου από πολλούς αναγνώστες. Ευνόητη είναι η στροφή της στατιστικής σε αναζήτηση μεθόδου ώστε να γίνει δυνατή η άμεση κατανόηση της έννοιας των αριθμών και των μεταξύ τους σχέσεων.

4.3.1 Η χρησιμότητα των διαγραμμάτων

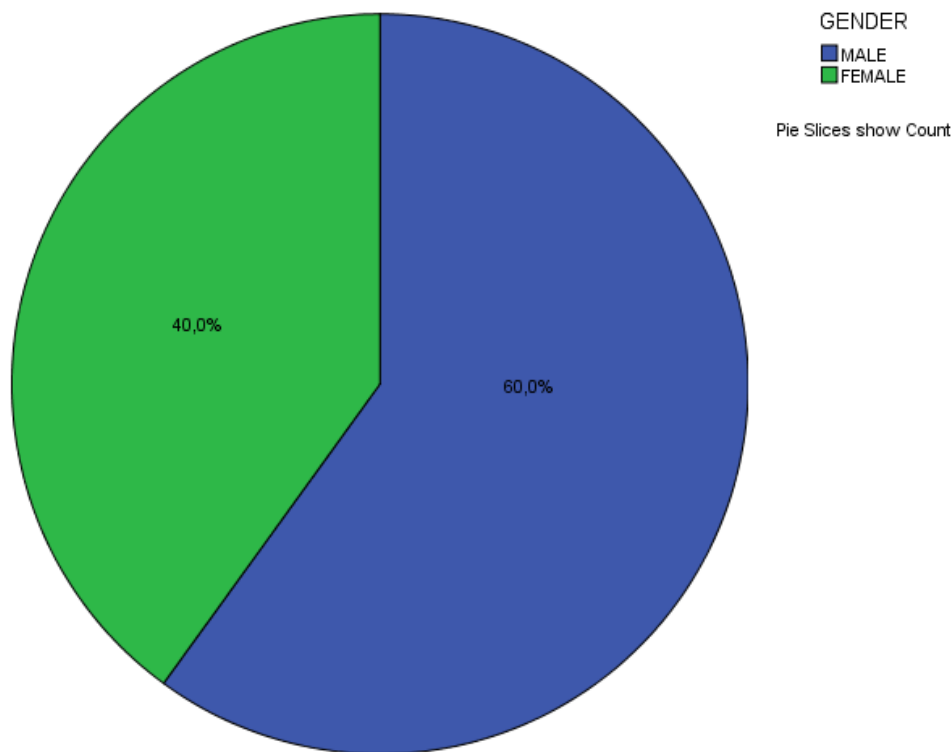
Με τη χρησιμοποίηση των γραφικών απεικονίσεων αποβλέπουμε στην εξυπηρέτηση δυο βασικών αντικειμενικών σκοπών. Ο πρώτος είναι η εποπτική παράσταση των στατιστικών στοιχείων. Αυτή η χρησιμότητα των διαγραμμάτων προέκυψε από την ανάγκη να γίνονται γνωστά τα συμπεράσματα των διαφόρων ερευνών στο ευρύτερο κοινό και όχι μόνο στον περιορισμένο κύκλο των ειδικών. Η έκθεση των ευρυμάτων με πίνακες απευθύνεται βασικά σε μικρό κύκλο αναγνωστών οι οποίοι είναι εξειδικευμένοι με αυτούς.

Αντίθετα οι γραφικές παραστάσεις συμβάλλουν πάρα πολύ στη διάδοση των στατιστικών γνώσεων και σε κύκλους οι οποίοι δεν θα μπορούσαν αλλιώς να τις εννοήσουν. Αυτό οφείλεται στο γεγονός ότι ο στατιστικός πίνακας απαιτεί αρκετή κρίση στην ερμηνεία του περιεχομένου του. Το διάγραμμα προσφέρει την ανάγλυφη εικόνα των αριθμών, η οποία είναι ευκολότερα κατανοητή. Οι γραφικές παραστάσεις αποτελούν τρόπο στατιστικής και τοπογραφικής μελέτης ορισμένων ιατροκοινωνικών φαινομένων και για το λόγο αυτό χρησιμοποιούνται στην Επιδημιολογία, Ιατρική δημογραφία και την κοινωνική ιατρική.

Παρόλα αυτά δεν πρέπει να υπερεκτιμηθεί η σημασία και η αξία των γραφικών παραστάσεων, διότι αυτές είναι επιβοηθητικές μορφές της στατιστικής αναλύσεως και δεν μπορούν να αντικαταστήσουν την ακρίβεια των στατιστικών πινάκων. Η χρησιμότητα τους παραμένει σε αυτό το οποίο αναφέρθηκε. Αποτελούν κατά κάποιον τρόπο τις υλικές εικόνες των αριθμών της στατιστικής έρευνας.

4.3.2 Κυκλικά διαγράμματα

Ένας συνηθισμένος τρόπος αναπαράστασης μιας κατανομής ποιοτικής μεταβλητής είναι τα κυκλικά διαγράμματα χωρισμένα σε κυκλικούς τομείς, το εμβαδόν των οποίων είναι ανάλογο του μεγέθους των μεταβλητών. Τα κυκλικά διαγράμματα προσφέρονται ασφαλώς και για την αναπαράσταση κατανομών ποσοτικών μεταβλητών. Η κατασκευή αυτών των διαγραμμάτων στηρίζεται στο γεγονός ότι όλη η επιφάνεια του κύκλου είναι 360° και αντιπροσωπεύει το 100% των παρατηρήσεων. Οι διάφοροι τομείς αντιπροσωπεύουν αναλογικά μέρη του συνόλου.

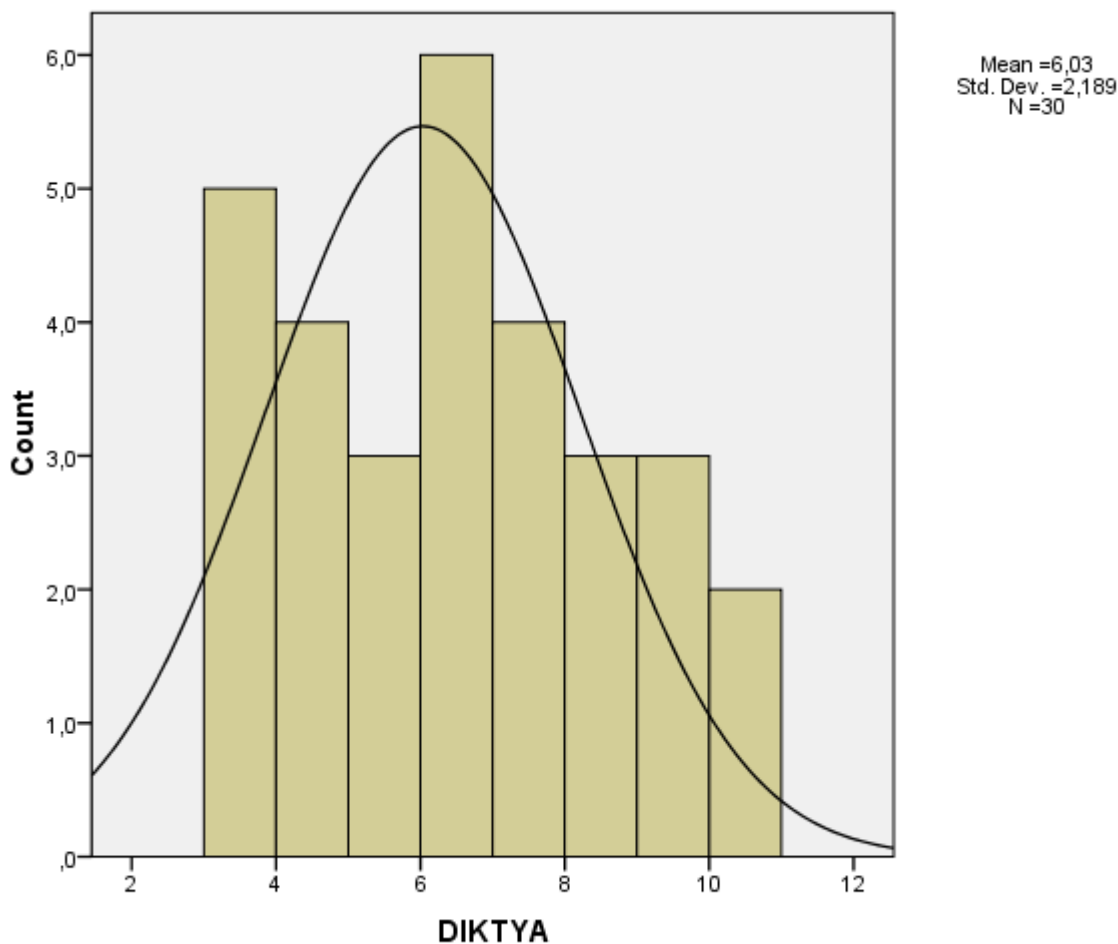


i. Παράδειγμα κυκλικού διαγράμματος

4.3.3 Ιστόγραμμα

Είναι μια μορφή διαγράμματος με στήλες. Υποστηρίζεται ότι τα ιστογράμματα ονομάστηκαν έτσι διότι προσομοιάζουν με τους ιστούς των ιστοφόρων σκαφών. Είναι η πλέον δόκιμη μορφή παρουσιάσεως κατανομών συχνότητων ποσοτικών συνεχών μεταβλητών γι' αυτό και χρησιμοποιούνται ευρύτατα στην στατιστική ανάλυση. Το ιστόγραμμα δίνει μια εξαιρετικά επαγωγική εικόνα του σχήματος, της μορφής της κατανομής. Ο οριζόντιος

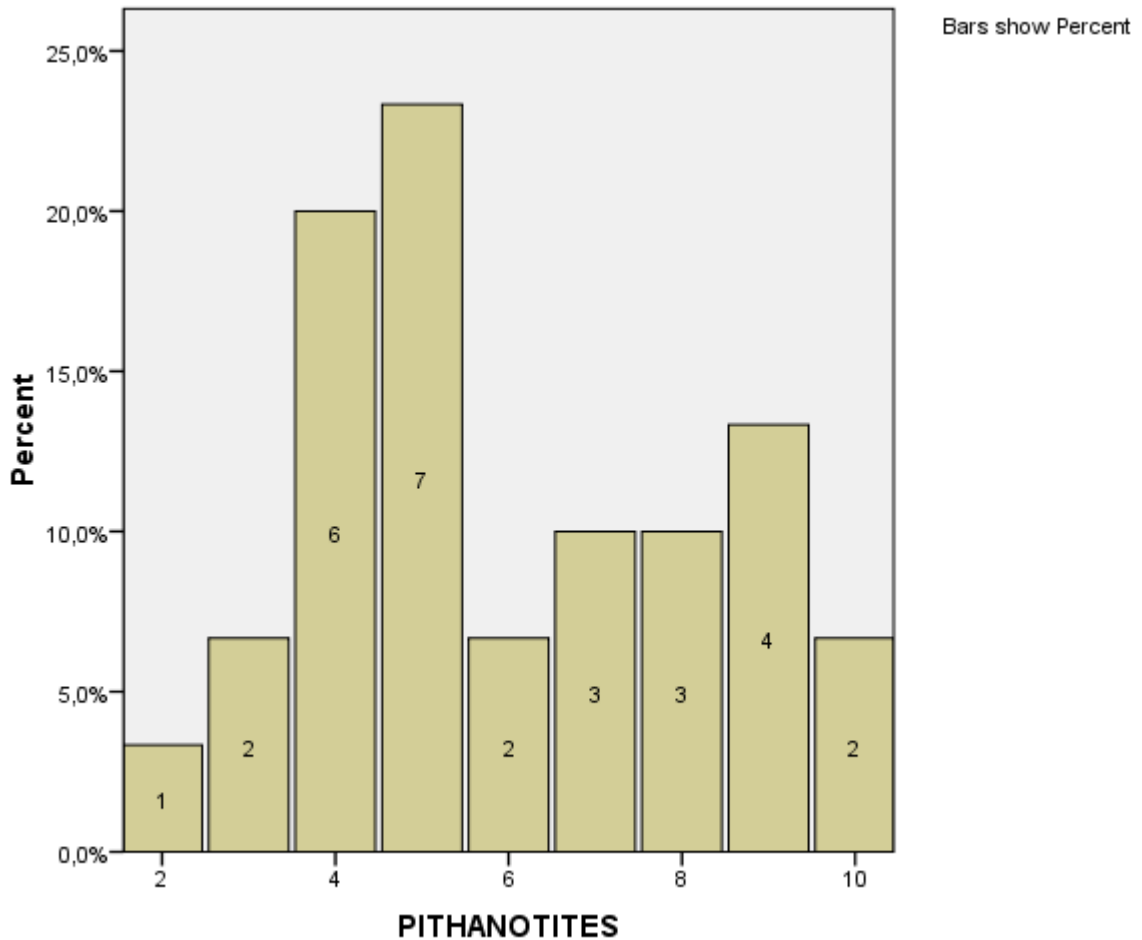
άξονας παρουσιάζει τα αληθινά όρια των διαφόρων διαστημάτων. Τα αληθινά όρια ενός διαστήματος είναι τα σημεία που το διαχωρίζουν από τα διαστήματα αμφοτέρων πλευρών. Ο κάθετος άξονας παρουσιάζει είτε τη συχνότητα ή τη σχετική συχνότητα των παρατηρήσεων εντός κάθε διαστήματος.



ii. Παράδειγμα ιστογράμματος

4.3.4 Ραβδογράμματα

Χρησιμοποιούνται συχνά για την παρουσίαση μιας κατανομής συχνοτήτων ποσοτικών δεδομένων. Σε ένα ραβδόγραμμα οι διάφορες κατηγορίες στις οποίες πέφτουν οι παρατηρήσεις παρουσιάζονται στον οριζόντιο άξονα. Μια κάθετη στήλη (ράβδος) πάνω από κάθε κατηγορία σχεδιάζεται ούτως ώστε το ύψος της να αντιπροσωπεύει είτε τη συχνότητα ή τη σχετική συχνότητα των παρατηρήσεων μέσα σε αυτή την τάξη.



iii. Παράδειγμα ραβδογράμματος

4.4 Κατανομές συχνοτήτων

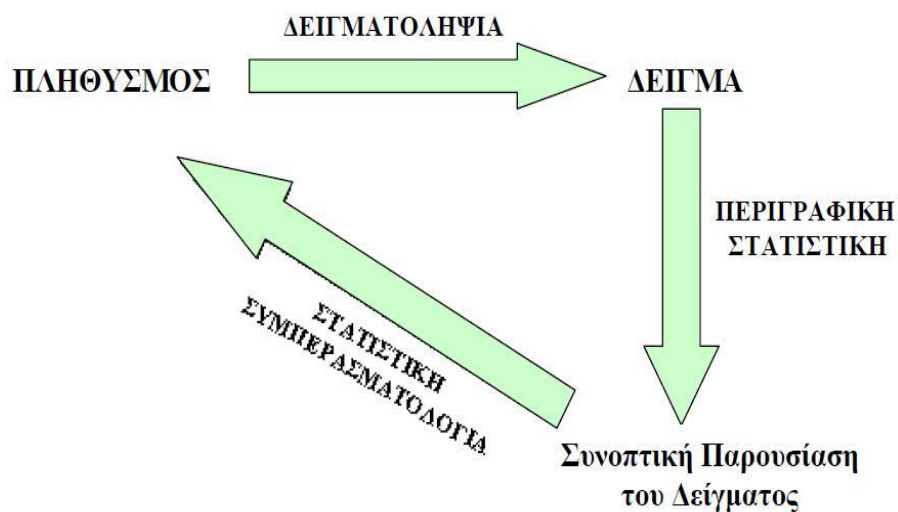
Όταν ο αριθμός των μετρήσεων είναι αρκετά μεγάλος υπάρχει πιθανότητα πολλές φορές να βρίσκουμε την ίδια τιμή. Τότε η κατάταξη των μετρήσεων κατά αύξουσα ή φθίνουσα σειρά γίνεται αρκετά δύσκολη και μια κατανομή συχνοτήτων η οποία λαμβάνει χώρα σε μια τόσο μεγάλη βάση είναι δύσχρηστη. Η δυσχέρεια αυτή παρακάμπτεται εύκολα με την οργάνωση των μετρήσεων σε αλληπάλληλες ομάδες διαδοχικών αριθμητικών τιμών, δηλαδή με το σχηματισμό τάξεων ή κλάσεων και την κατάστρωση μιας κατανομής συχνοτήτων.

Με τη διάταξη των τιμών μιας σειράς μετρήσεων κατά κλάσεις επιτυγχάνεται η ταξινόμηση της μεταβλητής κατά τρόπο ώστε το υλικό να

λάβει συγκεκριμένη μορφή, απαλείφεται η ανομοιογένεια και η ασάφεια και το κυριότερο διευκολύνεται ο περαιτέρω χειρισμός του υλικού για την ανεύρεση των στατιστικών παραμέτρων. Κατά τον σχηματισμό των κατανομών συχνοτήτων πρέπει να τηρούνται ορισμένες αρχές από τις οποίες η κυριότερη είναι η διατήρηση σταθερού του εύρους των κλάσεων. Είναι προτιμότερο να σχηματίσουμε αρχικά μια ποιο εκτεταμένη κατανομή συχνοτήτων σε τάξεις μικρού εύρους και εάν από την μελέτη της προσωρινής αυτής κατανομής προκύψει ότι η ομαδοποίηση δυο ή περισσότερων κλάσεων είναι εφικτή μας παρέχεται η ευχέρεια να το πράξουμε.

3. Την επαγωγική στατιστική στην οποία σκοπός μας είναι όχι μόνο να περιγράψουμε ότι παρατηρούμε και να εξάγουμε συμπεράσματα πάνω στο δείγμα μας αλλά να επεκτείνουμε αυτά τα συμπεράσματα σε ολόκληρο τον πληθυσμό μας. Θέλουμε δηλαδή από το δείγμα να επιστρέψουμε πίσω στον πληθυσμό, γενικεύοντας με το μικρότερο δυνατό κίνδυνο λάθους τα αποτελέσματα που προέκυψαν από το δείγμα μας έτσι ώστε να εξάγουμε αποτελέσματα για ολόκληρο το πληθυσμό

Το παρακάτω σχήμα μας βοηθάει στην κατανόηση των στατιστικών μεθοδολογιών:



iv. Στατιστικές μεθοδολογίες

Μια στατιστική μέθοδος χωρίζεται σε τρία στάδια:

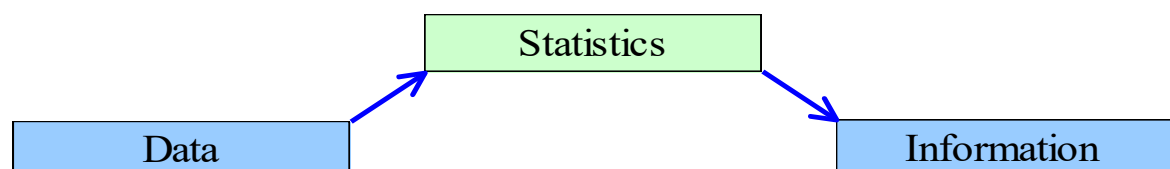
Συλλογή δεδομένων – Στη φάση αυτή, με απογραφή ή με οποιοδήποτε άλλο τρόπο, συλλέγονται οι σχετικές πληροφορίες για το πρόβλημα και ταξινομούνται κατάλληλα.

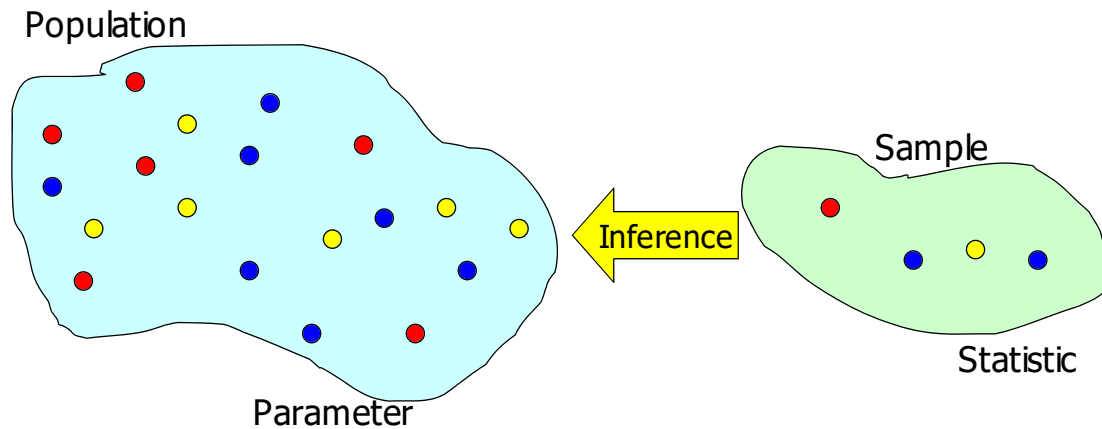
Μαθηματική επεξεργασία των δεδομένων που έχουν συλλεγεί. Είναι και το πιο πολύπλοκο αλλά και ενδιαφέρον στις εφαρμογές του στάδιο. Για μια τέτοια επεξεργασία χρησιμοποιούνται όλες οι μαθηματικές μέθοδοι που απαιτούνται. Κυρίως χρησιμοποιούνται στοιχεία από τον ολοκληρωτικό λογισμό και τη θεωρία των πιθανοτήτων. Κύριος σκοπός στη φάση αυτή είναι να βρεθεί κάποιος φυσικοκοινωνικός νόμος που διέπει το φαινόμενο που ενδιαφέρει.

Εξαγωγή συμπερασμάτων – Είναι το τελευταίο στάδιο, αλλά και αυτό που ενδιαφέρει στην όλη υπόθεση. Τα συμπεράσματα που βγαίνουν όμως με τη στατιστική μέθοδο περιέχουν ένα ποσοστό αβεβαιότητας που όμως η ίδια στατιστική μπορεί να υπολογίσει. Βασική προϋπόθεση για την εξαγωγή συμπερασμάτων είναι ο σωστός τρόπος στη μαθηματική επεξεργασία των δεδομένων.

4.5 Επαγωγική Στατιστική

Επαγωγική Στατιστική είναι η διαδικασία με την οποία αποκτούμε πληροφορίες και εξάγουμε συμπεράσματα σχετικά με τον πληθυσμό από δείγματα





v. Επαγωγική στατιστική

Για να εκτελέσουμε επαγωγική στατιστική, απαιτούνται γνώσεις περιγραφικής στατιστικής, κατανομές πιθανοτήτων, και δειγματοληπτικές κατανομές.

Στην επαγωγική στατιστική έχουμε:

A) Τον έλεγχο υποθέσεων

Ο έλεγχος υποθέσεων ξεκινάει από τη μηδενική υπόθεση (H_0) και προσπαθεί να την απορρίψει, υπολογίζοντας πιθανότητες και αντλώντας πληροφορία από τη δειγματοληπτική κατανομή του στατιστικού που ελέγχεται.

B) Δημιουργία διαστημάτων εμπιστοσύνης

Τα διαστήματα εμπιστοσύνης υπολογίζονται για τις παραμέτρους του πληθυσμού, χρησιμοποιώντας πληροφορία από το αντίστοιχο στατιστικό του δείγματος. Υπολογίζοντάς το περιορίζουμε την αβεβαιότητά μας για το πόσο πιθανό είναι να ισχύει για τον πληθυσμό η τιμή που βρήκαμε στο δείγμα.

ΚΕΦΑΛΑΙΟ 5

5.1 SPSS

Το SPSS είναι ένα στατιστικό πακέτο ανάλυσης δεδομένων, το οποίο προσφέρει στο χρήστη δυνατότητες για δημιουργία αναφορών, ανάλυση και μοντελοποίηση δεδομένων καθώς και για γραφική αναπαράσταση τους. Διαθέτει πολλές στατιστικές συναρτήσεις για ανάλυση δεδομένων μέσα από ένα εύχρηστο γραφικό περιβάλλον. Τα γράμματα SPSS (Statistical Package for the Social Sciences) σημαίνουν στατιστικό πακέτο για τις κοινωνικές επιστήμες και κυκλοφόρησε σε πρώτη έκδοση το 1968 από τους Norman Nie και Hadlai Hull. Είναι ένα από τα πιο ευρέως χρησιμοποιούμενα προγράμματα για τη στατιστική ανάλυση των κοινωνικών επιστημών.

Με την βοήθεια του SPSS όλα τα στάδια της αναλυτικής διαδικασίας ολοκληρώνονται κάτω από ένα ενοποιημένο περιβάλλον εργασίας καλύπτοντας την ανάλυση από άκρο σε άκρο.



vi. Αναλυτική διαδικασία

Ο χρηματοοικονομικός τομέας, ο τομέας των τηλεπικοινωνιών, ο τομέας της υγείας, της εκπαίδευσης και της έρευνας, ο ευρύτερος δημόσιος τομέας και ο τομέας της βιομηχανίας αποτελούν τα βασικότερα πεδία εφαρμογών της τεχνολογίας του SPSS.

5.1.1 χ^2 test

Ο δείκτης χ^2 προτάθηκε αρχικά από τον Γερμανό Helmer το 1875 αλλά παρέμεινε άγνωστος μέχρι ο Pearson τον εισήγαγε στην εφαρμοσμένη στατιστική. Βασικής σημασίας είναι το γεγονός ότι το χ^2 κριτήριο ακολουθεί μια κατανομή τη χ^2 - κατανομή η μορφή της οποίας εξαρτάται από τον αριθμό των βαθμών ελευθερίας. Εάν κατά το σχηματισμό της κατανομής των αποκλίσεων των πραγματικών από τις θεωρητικές τιμές υπολογίζουμε κάθε φορά την τιμή του χ^2 είναι ευνόητο ότι θα σχηματιστεί μια δειγματοληπτική κατανομή των τιμών του κριτηρίου χ^2 .

Η τυχαία μεταβλητή χ^2 προέρχεται από τον μετασχηματισμό μιας κανονικής τυχαίας μεταβλητής όπως παρακάτω. Αν Y είναι μία τυχαία μεταβλητή από την κανονική κατανομή $N(\mu, \sigma^2)$ (μέση τιμή μ και διακύμανση σ^2) τότε η τυποποιημένη μεταβλητή

$$Z = \frac{Y - \mu}{\sigma} \quad (5.1)$$

Ακολουθεί την κανονική κατανομή $N(0, 1)$, ενώ η τυχαία μεταβλητή

$$\chi^2(1) = \left(\frac{Y - \mu}{\sigma} \right)^2 = Z^2 \quad (5.2)$$

Ακολουθεί μια νέα κατανομή που λέγεται χ^2 με 1 βαθμό ελευθερίας.

Αν Y_1 και Y_2 είναι δυο ανεξάρτητες τυχαίες μεταβλητές από την κανονική κατανομή $N(\mu, \sigma^2)$ τότε η τυχαία μεταβλητή

$$\chi^2(2) = \left(\frac{Y_1 - \mu}{\sigma} \right)^2 + \left(\frac{Y_2 - \mu}{\sigma} \right)^2 = Z_1^2 + Z_2^2 \quad (5.3)$$

Ακολουθεί την κατανομή χ^2 με 2 βαθμούς ελευθερίας.

Τέλος αν Y_1, Y_2, \dots, Y_k είναι ανεξάρτητες τυχαίες μεταβλητές από την κανονική κατανομή $N(\mu, \sigma^2)$ τότε η τυχαία μεταβλητή

$$\chi^2(k) = \left(\frac{Y_1 - \mu}{\sigma}\right)^2 + \left(\frac{Y_2 - \mu}{\sigma}\right)^2 + \dots + \left(\frac{Y_k - \mu}{\sigma}\right)^2 = Z_1^2 + Z_2^2 + \dots + Z_k^2 \quad (5.4)$$

Ακολουθεί την κατανομή χ^2 με k βαθμούς ελευθερίας

5.2 Ορισμός της t-κατανομής

Όταν έχουμε μεγάλα δείγματα μπορούμε να καθορίσουμε με βάση τις στατιστικές τιμές του δείγματος όρια εμπιστοσύνης για τον άγνωστο πληθυσμιακό μέσο όρο καθώς επίσης και να προβούμε σε συγκρίσεις μεταξύ των μέσων όρων των δειγμάτων χρησιμοποιώντας το κριτήριο z .

Όταν ο αριθμός των στατιστικών μονάδων των δειγμάτων είναι μικρός η δειγματοληπτική κατανομή των μέσων όρων των δειγμάτων ακολουθεί την κατανομή t-κατανομή του student η οποία προσομοιάζει με την κανονική και εξαρτάται κάθε φορά από το μέγεθος του δείγματος. Από μερικούς η κατανομή αυτή ονομάζεται t-κατανομή του student-Fisher από τα ονόματα των δύο ερευνητών που την επεξεργάστηκαν. Η t-κατανομή διατυπώθηκε το 1908 από τον Gosset ο οποίος στη δημοσίευση του χρησιμοποίησε το ψευδώνυμο student με το οποίο έγινε γνωστή πολύ αργότερα χάρη στον Fisher.

5.2.1 Ιδιότητες t-κατανομής

Στην κατανομή αυτή η επιφάνεια που περικλείεται κάτω από την t-καμπύλη εξαρτάται από το μέγεθος του δείγματος. Στην ουσία η επιφάνεια εξαρτάται κάθε φορά από το μέγεθος του δείγματος ελαττωμένο κατά μονάδα, δηλαδή από το $n-1$. Στην t-κατανομή η απόσταση από τον αριθμητικό μέσο όρο εκφράζεται σε τιμές t . Υπάρχουν ειδικοί πίνακες που παρέχουν τις τιμές t για δεδομένο επίπεδο πιθανότητας.

5.3 Η δοκιμασία t του student

Για τον υπολογισμό των ορίων εμπιστοσύνης και της ελάχιστης σημαντικής διαφοράς χρησιμοποιούνται οι τιμές t για τον συγκεκριμένο αριθμό βαθμών ελευθερίας. Ο έλεγχος της διαφοράς προς το τυπικό σφάλμα της διαφοράς στα προβλήματα της t -κατανομής βασίζεται στον υπολογισμό της τιμής του t για τους δεδομένους βαθμούς ελευθερίας. Η δοκιμασία στην περίπτωση αυτή καλείται δοκιμασία t του student και το t ονομάζεται t -κριτήριο. Η δοκιμασία εφαρμόζεται αφού προηγουμένως ληφθεί πρόνοια για τον υπολογισμό μιας κοινής διακυμάνσεως για τα δείγματα.

Στην καθημερινή ιατρική στατιστική το t -κριτήριο βρίσκει πολυπληθείς εφαρμογές διότι ο αριθμός των πειραμάτων είναι περιορισμένος και οι συγκρίσεις μεταξύ μέσων όρων πάρα πολλές.

5.4 Το t -test για ανεξάρτητα δείγματα

Συγκρίνει τους μέσους όρους μίας αριθμητικής μεταβλητής μεταξύ δύο ομάδων. Η εντολή στο SPSS είναι: «ANALYZE» - «COMPARE MEANS» - «INDEPENDENT-SAMPLES T TEST».

Το αποτέλεσμα της ανάλυσης καταλήγει σε δύο πίνακες. Στον πρώτο, «Group statistics», εμφανίζονται περιγραφικοί στατιστικοί δείκτες: το μέγεθος («N») κάθε ομάδας, ο μέσος όρος («Mean»), η τυπική απόκλιση («Std. Deviation») και το τυπικό σφάλμα του μέσου όρου («Std. Error Mean»). Στο δεύτερο πίνακα παρουσιάζεται το t -test, ως εξής: υπολογίζονται δύο τιμές, μία t -τιμή για την περίπτωση που τα δείγματα είναι ομοιογενή («Equal variances assumed») και μία t -τιμή για την περίπτωση που τα δείγματα είναι ανομοιογενή («Equal variances not assumed»).

Ο έλεγχος της ομοιογένειας της διασποράς μεταξύ των συγκρινόμενων δειγμάτων γίνεται με το Levene test. Αν η F -τιμή του Levene test είναι στατιστικώς σημαντική («Sig.» < ,050), αυτό σημαίνει ότι τα δείγματα είναι ανομοιογενή.

Στη συνέχεια, εκτός από την t -τιμή, δίνονται οι βαθμοί ελευθερίας («df»), το επίπεδο στατιστικής σημαντικότητας («Sig.»), η διαφορά μεταξύ των μέσων όρων («Mean Dif-ference»), το τυπικό σφάλμα της διαφοράς των μέσων όρων («Std. Error Diffe-rence») και τα όρια του διαστήματος εμπιστοσύνης της διαφοράς των μέσων όρων («95% Confidence Interval of the Difference»).

5.5 To t -test για ζευγαρωτά δείγματα

Χρησιμοποιείται για να συγκρίνουμε αν ο μέσος όρος μιας αριθμητικής μεταβλητής διαφέρει από τον μέσο όρο μιας άλλης αριθμητικής μεταβλητής (η οποία έχει μετρηθεί με την ίδια μονάδα μέτρησης) σε μία, την ίδια ομάδα ατόμων. Πρόκειται δηλ. για αριθμητική διμεταβλητή σε εξαρτημένα δείγματα. Η σχετική εντολή στο SPSS είναι: «ANALYZE» - «COMPARE MEANS» - «PAIRED-SAMPLES T TEST»

Το αποτέλεσμα δίνει τρεις πίνακες. Στον πρώτο πίνακα δίνονται περιγραφικοί στατιστικοί δείκτες των δύο μεταβλητών: μέσοι όροι («Mean»), μέγεθος δείγματος («N»), τυπικές αποκλίσεις («Std. Deviation»), τυπικό σφάλμα των μέσων όρων («Std. Error Mean»). Στο δεύτερο πίνακα δίνεται ο δείκτης συνάφειας Pearson r («Correlation») μεταξύ των δύο μεταβλητών και το επίπεδο στατιστικής σημαντικότητάς του («Sig.»). Στον τρίτο πίνακα δίνονται τα στοιχεία από τη σύγκριση των μέσων όρων («Paired differences»), η τιμή t , οι βαθμοί ελευθερίας («df») και το επίπεδο στατιστικής σημαντικότητας («Sig.»).

5.6 ANOVA

Η ανάλυση διακύμανσης (Analysis of Variance, ANOVA) χρησιμοποιείται για τον έλεγχο της στατιστικής σημαντικότητας των διαφορών των μέσων όρων περισσότερων από δύο ομάδων. Επιπλέον, με την ANOVA μπορούμε να ελέγξουμε την αλληλεπίδραση δύο ή περισσότερων ανεξάρτητων (κατηγορικών) μεταβλητών πάνω στην εξαρτημένη (αριθμητική)

μεταβλητή. Δηλαδή, μπορεί να έχουμε μία μικτή διμεταβλητή, τριμεταβλητή ή ακόμα και πολυμεταβλητή.

Η ανάλυση της διακύμανσης (ANalysis Of Variance – ANOVA) είναι μία στατιστική μέθοδος με την οποία η μεταβλητότητα που υπάρχει σ' ένα σύνολο δεδομένων διασπάται στις επιμέρους συνιστώσες της με στόχο την κατανόηση της σημαντικότητας των διαφορετικών πηγών προέλευσής της. Η ανάπτυξη της μεθοδολογίας οφείλεται στον θεμελιωτή της σύγχρονης στατιστικής επιστήμης, άγγλο στατιστικό Sir Ronald Aylmer Fisher (1890-1962). Στην πραγματικότητα η ANOVA περιλαμβάνει μία ομάδα στατιστικών μεθόδων καταλλήλων για την ανάλυση δεδομένων που προκύπτουν από πειραματικούς σχεδιασμούς.

Τα δεδομένα ενός δείγματος ανάλογα με την προέλευσή τους διακρίνονται σε παρατηρήσεις (observational sampling) ή σε πειραματικά (designed sampling). Στην πρώτη κατηγορία ο στατιστικός ερευνητής απλά παρατηρεί τις τιμές που εμφανίζονται χωρίς να έχει δυνατότητα επέμβασης στις αντίστοιχες μεταβλητές. Αντίθετα στη δεύτερη κατηγορία ο στατιστικός ερευνητής προσπαθεί να ελέγξει τα επίπεδα μιας ή περισσότερων ανεξάρτητων (independent) μεταβλητών προκειμένου να προσδιορίσει την επίδραση που έχουν πάνω στην υπό μελέτη μεταβλητή που καλείται εξαρτημένη (dependent) ή απόκριση (response). Για παράδειγμα, απόκριση μπορεί να είναι η βαθμολογία στην εξέταση του μαθήματος της στατιστικής, ο όγκος των πωλήσεων μιας επιχείρησης ή το συνολικό εισόδημα μιας οικογένειας κατά τη διάρκεια του έτους.

5.7 Τεχνικές στο SPSS

5.7.1 Frequencies

Η διαδικασία Frequencies εμφανίζει πίνακες κατανομής συχνοτήτων και γραφικές παραστάσεις (ραβδογράμματα και ιστογράμματα) και υπολογίζει στατιστικά μέτρα. Η διαδικασία εκτελείται εφόσον δηλωθεί μία τουλάχιστον μεταβλητή. Τα ελάχιστα αποτελέσματα (default output) που παράγει είναι ένας πίνακας κατανομής συχνοτήτων για κάθε μεταβλητή. Χρησιμοποιείται κυρίως

για περιγραφική στατιστική κατηγοριοποιημένων μεταβλητών δηλαδή είτε ποιοτικών είτε ποσοτικών διακριτών.

5.7.2 Descriptives

Η διαδικασία Descriptives υπολογίζει και εμφανίζει στατιστικά μέτρα για ποσοτικές, συνεχείς κυρίως, μεταβλητές για τις οποίες δεν χρειάζεται να κατασκευαστεί πίνακας κατανομής συχνοτήτων.

5.7.3 Compute

Χρησιμοποιείται για τη δημιουργία τιμών νέων μεταβλητών με την εκτέλεση υπολογισμών πάνω σε τιμές μεταβλητών που ήδη υπάρχουν.

5.7.4 Recode

Με την εντολή RECODE είναι δυνατή η επανακωδικοποίηση μεταβλητής που ήδη υπάρχει, με δύο διαφορετικές δυνατότητες: Είτε τη μετατροπή της ίδιας μεταβλητής με διαγραφή των παλαιών τιμών και αντικατάστασή τους από τις νέες είτε τη δημιουργία νέας μεταβλητής με τιμές τις κωδικοποιημένες τιμές. Με αυτόν τον τρόπο, συνήθως δημιουργούμε ποιοτικές μεταβλητές με τη μορφή διαστημάτων, από ποσοτικές μεταβλητές.

5.7.5 Crosstabs

Η διαδικασία Crosstabs εμφανίζει πίνακες συνάφειας για ζεύγη ποιοτικών μεταβλητών και υπολογίζει διάφορα στατιστικά μέτρα και ελέγχους ανεξαρτησίας τους.

5.8 Γραμμική παλινδρόμηση

Η Παλινδρόμηση είναι μεθοδολογία της στατιστικής επιστήμης η οποία βρίσκει, χρησιμοποιεί και τελικά αξιοποιεί τις σχέσεις μεταξύ δύο ή περισσότερων ποσοτικών μεταβλητών, έτσι ώστε οι τιμές της μιας μεταβλητής η οποία ονομάζεται εξαρτημένη (Y), να μπορούν να εκτιμηθούν ή να προβλεφθούν από τις τιμές της άλλης (X) ή των άλλων (X_1, X_2, \dots, X_n) μεταβλητών που ονομάζονται ανεξάρτητες.

Η θεωρία της παλινδρόμησης βρίσκει πολύ μεγάλη εφαρμογή στις εφαρμοσμένες επιστήμες. Μέσω αυτής έχουν υπολογιστεί πολλές σχέσεις για διάφορες μεταβλητές όπως για ύψος, βάρος κλπ. Με την μέθοδο αυτή θα κάνουμε το μοντέλο πρόβλεψης των ιατρικών δεικτών.

5.9 Συσχέτιση

Είναι η διαδικασία εκείνη μέσω της οποίας προσπαθούμε να βρούμε το βαθμό ενδοεξάρτησης μεταξύ τυχαίων μεταβλητών. Για να εφαρμοστεί η διαδικασία της συσχέτισης, πρέπει όλες οι μεταβλητές που θα χρησιμοποιηθούν είτε ως εξαρτημένες είτε ως ανεξάρτητες, να είναι τυχαίες μεταβλητές.

Μπορούμε να διακρίνουμε τη συσχέτιση σε γραμμική όταν έχουμε δυο τυχαίες μεταβλητές οι οποίες θα πρέπει να προέρχονται από πληθυσμό ο οποίος ακολουθεί κανονική κατανομή και πολλαπλή όταν έχουμε περισσότερες από δυο τυχαίες μεταβλητές οι οποίες και αυτές όλες πρέπει να προέρχονται από πληθυσμό ο οποίος ακολουθεί κανονική κατανομή.

ΚΕΦΑΛΑΙΟ 6

ΠΡΑΚΤΙΚΗ ΕΦΑΡΜΟΓΗ

Η εφαρμογή μας έχει ως βασικό σκοπό την μελέτη συγκεκριμένων ιατρικών δεδομένων και τη δημιουργία ενός μοντέλου πρόβλεψης ενός καρκινικού δείκτη με τη βοήθεια άλλων καρκινικών δεικτών. Το δείγμα μας αποτελείται από 93 άτομα, από τα οποία τα 50 άτομα είναι άντρες (σε ποσοστό 53,8%) και τα υπόλοιπα 43 είναι γυναίκες (σε ποσοστό 46,2%).

Πίνακας κατανομής συχνοτήτων για το φύλο

	Συχνότητα	Ποσοστό	Ποσοστό που υπολογίζεται βάση όσων απάντησαν σε μία ερώτηση	Αθροιστικό ποσοστό
ΑΝΤΡΑΣ	50	53,8	53,8	53,8
ΓΥΝΑΙΚΑ	43	46,2	46,2	100,0
ΣΥΝΟΛΟ	93	100,0	100,0	

Τα στοιχεία των ασθενών τα οποία μελετάμε είναι οι καρκινικοί δείκτες και συγκεκριμένα οι δείκτες AFP, CA15.3, CA125, CA19.9 και CEA.

Οι καρκινικοί δείκτες αποτελούν ουσίες που κυκλοφορούν στο φυσιολογικό αίμα σε μικρές ποσότητες. Αυξημένη συγκέντρωση και αύξηση των τιμών τους υποδηλώνει την παρουσία νεοπλασίας. Οι καρκινικοί δείκτες βοηθούν:

- Στην διάγνωση της νεοπλασίας
- Μερικές φορές έχουν προγνωστική αξία
- Χρησιμεύουν για παρακολούθηση και εκτίμηση της θεραπείας
- Για την έγκαιρη διάγνωση υποτροπών

AFP (Άλφα-Φετοπρωτεΐνη ή AFP (Alpha-FetoProtein)

Παράγεται στο ήπαρ του εμβρύου, στο γαστρεντερικό σύστημα και στο λεκιθικό ασκό (το σακουλάκι του μωρού στην κοιλιά). Η πρωτεΐνη αυτή κυκλοφορεί στο αίμα του εμβρύου και εξαφανίζεται σταδιακά (ποτέ εντελώς, μένει κάτι ελάχιστο) και αντικαθίσταται από τη λευκωματίνη. Η κύρια

χρησιμότητα της είναι α) σε συνδυασμό με την HCG στη διάγνωση / παρακολούθηση / πρόγνωση ασθενών με σεμινωματώδεις όγκους (αυτό σημαίνει όγκους στα γεννητικά όργανα και στα δυο φύλα) και β) στη διάγνωση του ηπατοκυτταρικού καρκινώματος (αυτό είναι ο καρκίνος στο συκώτι). Φυσιολογικές τιμές: 0-10 kU/L.

CA15.3

Είναι προϊόν του γονιδίου MUC-1 και σχετίζεται με την κυτταρική προσκόλληση. Αυξημένα επίπεδα εμφανίζονται στον καρκίνο του μαστού, αλλά και σε άλλα αδenoκαρκινώματα. Αυξημένες τιμές μπορεί να εμφανιστούν και σε άλλες παθήσεις (καλοήθειες) του μαστού, αλλά και σε παθήσεις του ήπατος. Φυσιολογικές τιμές: 0 - 35 U/L.

CA125

Υπάρχουν πολλών ειδών καρκινικά αντιγόνα CA (cancer antigens), αλλά μόνο μερικά αποδείχτηκαν χρήσιμα. Βρίσκονται αυξημένο στον επιθηλιακό καρκίνο των ωοθηκών, αλλά και σε άλλα ενδοκοιλιακά καρκινώματα (καρκίνος στομάχου, ενδομητρίου, παχέος εντέρου, ήπατος). Μη νεοπλασματικές ασθένειες που σχετίζονται με αύξηση του CA 125 είναι η ενδομητρίωση, η παγκρεατίτιδα, η ηπατική κίρρωση, η περιτονίτιδα, ο ασκίτης και η φλεγμονώδης νόσος της πυέλου. Αυξάνεται και φυσιολογικά στην περίοδο και στην εγκυμοσύνη (όχι όμως περισσότερο από 100 kU/L). Φυσιολογικές τιμές: 0 - 35 kU/L

CA19.9

Ανιχνεύεται αυξημένη στον καρκίνο του παγκρέατος, στο 50% των ασθενών με καρκίνο στομάχου και στο 1/3 των ασθενών με καρκίνο παχέος εντέρου. Η παγκρεατίτιδα, ο ίκτερος, η κίρρωση και η κυστική ίνωση δίνουν ψευδώς θετικά αποτελέσματα. Φυσιολογικές τιμές: 0 - 37 U/L (για άλλους ως 100).

CEA

Είναι μια πρωτεΐνη που σχετίζεται με την προσκόλληση των κυτάρων. Μπορεί να είναι αυξημένο σχεδόν σε κάθε αδenoκαρκίνωμα προχωρημένου

σταδίου, ενώ σπάνια ανιχνεύεται σε αρχικά στάδια. Αυξημένα επίπεδα CEA μπορεί να εμφανιστούν και σε μη νεοπλασματικά νοσήματα (ηπατίτιδα, ηπατική κίρρωση, αποφρακτικός ίκτερος - πέτρα στη χολή, ελκώδη κολίτιδα, νόσο Crohn, παγκρεατίτιδα, βρογχίτιδα, εμφύσημα και νεφροπάθεια) ή ακόμα και σε κανονικούς ανθρώπους που απλά καπνίζουν πολύ. Χρησιμεύει για την παρακολούθηση ασθενών με καρκίνο παχέος εντέρου. Φυσιολογικές τιμές: 0 - 3,5 μg/L (0 - 5 για καπνιστές).

ΣΤΑΤΙΣΤΙΚΑ ΜΕΤΡΑ ΠΟΣΟΤΙΚΩΝ ΜΕΤΑΒΛΗΤΩΝ

Στην συνέχεια ακολουθούν η μέση τιμή, η διακύμανση, η μέγιστη και η ελάχιστη τιμή της κάθε μεταβλητής με βάση το δείγμα μας.

Μικρότερη, Μεγαλύτερη, Μέση τιμή και διακύμανση του δείκτη AFP

Μικρότερη τιμή:	0,00
Μεγαλύτερη τιμή:	12,50
Μέση τιμή:	2,1756
Διακύμανση:	2,03898

Παρατηρούμε ότι η μέση τιμή του δείκτη AFP ανήκει στα φυσιολογικά όρια τιμών.

Μικρότερη, Μεγαλύτερη, Μέση τιμή και διακύμανση του δείκτη CA15.3

Μικρότερη τιμή:	4,80
Μεγαλύτερη τιμή:	139,00
Μέση τιμή:	50,4146
Διακύμανση:	31,45115

Παρατηρούμε ότι η μέση τιμή του δείκτη CA15.3 δεν ανήκει στα φυσιολογικά όρια τιμών.

Μικρότερη, Μεγαλύτερη, Μέση τιμή και διακύμανση του δείκτη CA125

Μικρότερη τιμή:	2,60
Μεγαλύτερη τιμή:	91,60
Μέση τιμή:	11,1148
Διακύμανση:	10,13674

Παρατηρούμε ότι η μέση τιμή του δείκτη CA125 ανήκει στα φυσιολογικά όρια τιμών.

Μικρότερη, Μεγαλύτερη, Μέση τιμή και διακύμανση του δείκτη CA19.9

Μικρότερη τιμή:	1,30
Μεγαλύτερη τιμή:	51,79
Μέση τιμή:	15,1262
Διακύμανση:	7,23049

Παρατηρούμε ότι η μέση τιμή του δείκτη CA19.9 ανήκει στα φυσιολογικά όρια τιμών.

Μικρότερη, Μεγαλύτερη, Μέση τιμή και διακύμανση του δείκτη CEA

Μικρότερη τιμή:	0,37
Μεγαλύτερη τιμή:	5,71
Μέση τιμή:	1,4113
Διακύμανση:	0,85172

Παρατηρούμε ότι η μέση τιμή του δείκτη CEA ανήκει στα φυσιολογικά όρια τιμών.

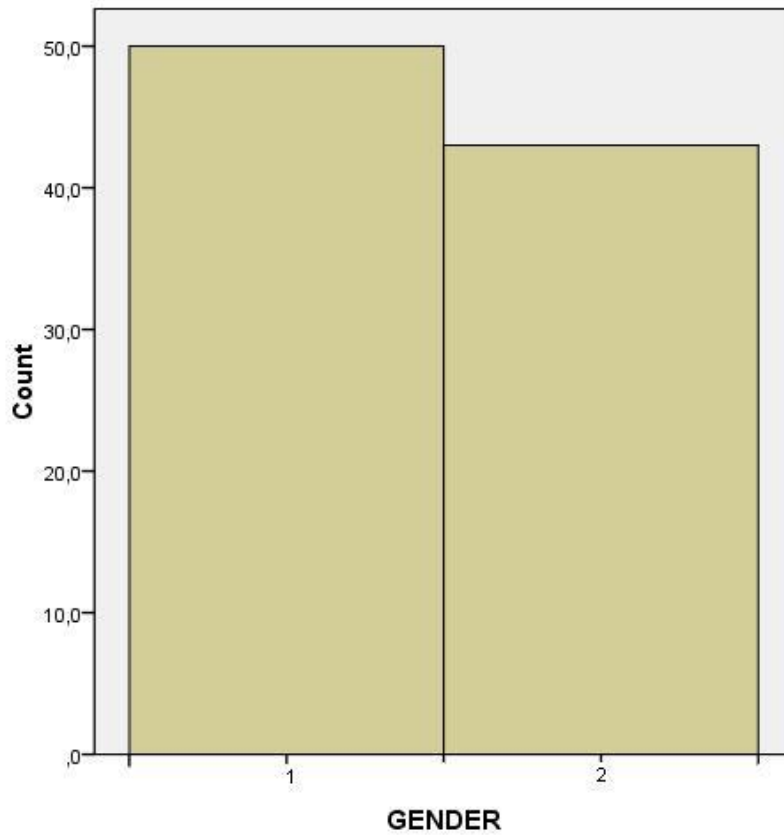
Οι τιμές αναφοράς όσον αφορά τον κάθε ένα καρκινικό δείκτη είναι:

Για τον AFP 0-15, για τον CA15.3 0-32,4, για τον CA125 0-35, για τον CA19.9 0-37 και για τον CEA όσον αφορά τους καπνιστές είναι μικρότερο του 10 και όσον αφορά τους μη καπνιστές είναι μικρότερο του 5.

Συμπέρασμα:

Σύμφωνα με τα στοιχεία τα οποία έχουμε βλέπουμε ότι η μέση τιμή του δείκτη AFP είναι 2.1756, η μέση τιμή του δείκτη CA125 είναι 11.1148, η μέση τιμή του δείκτη CA19.9 είναι 15.1262 και η μέση τιμή του δείκτη CEA είναι 1.4113. και παρατηρούμε ότι ανήκουν μέσα στο όριο το οποίο έχουμε σαν τιμή αναφοράς, σε αντίθεση με τον καρκινικό δείκτη CA15.3 του οποίου η μέση τιμή είναι 50.4146 η οποία είναι πολύ ψηλότερη από την τιμή αναφοράς (0-32,4)

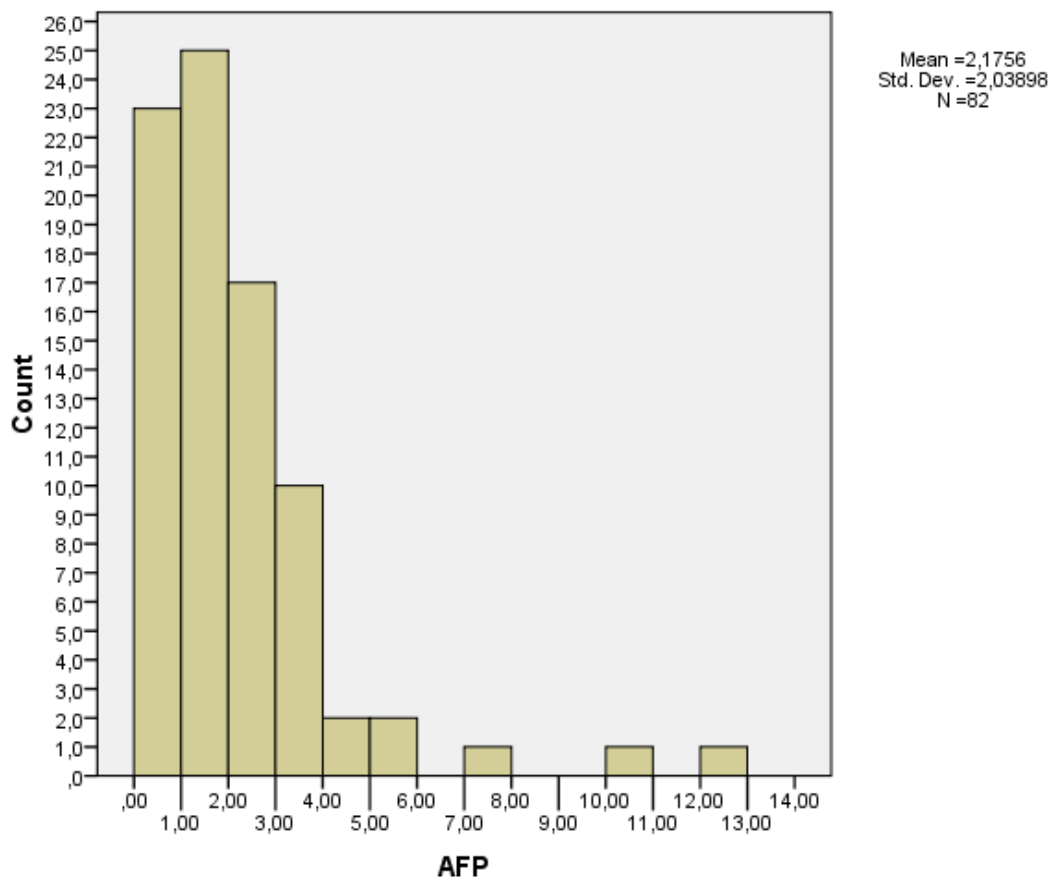
ΡΑΒΔΟΓΡΑΜΜΑΤΑ ΠΟΙΟΤΙΚΩΝ ΜΕΤΑΒΛΗΤΩΝ



vii. *Ραβδόγραμμα που παριστάνει το φύλο*

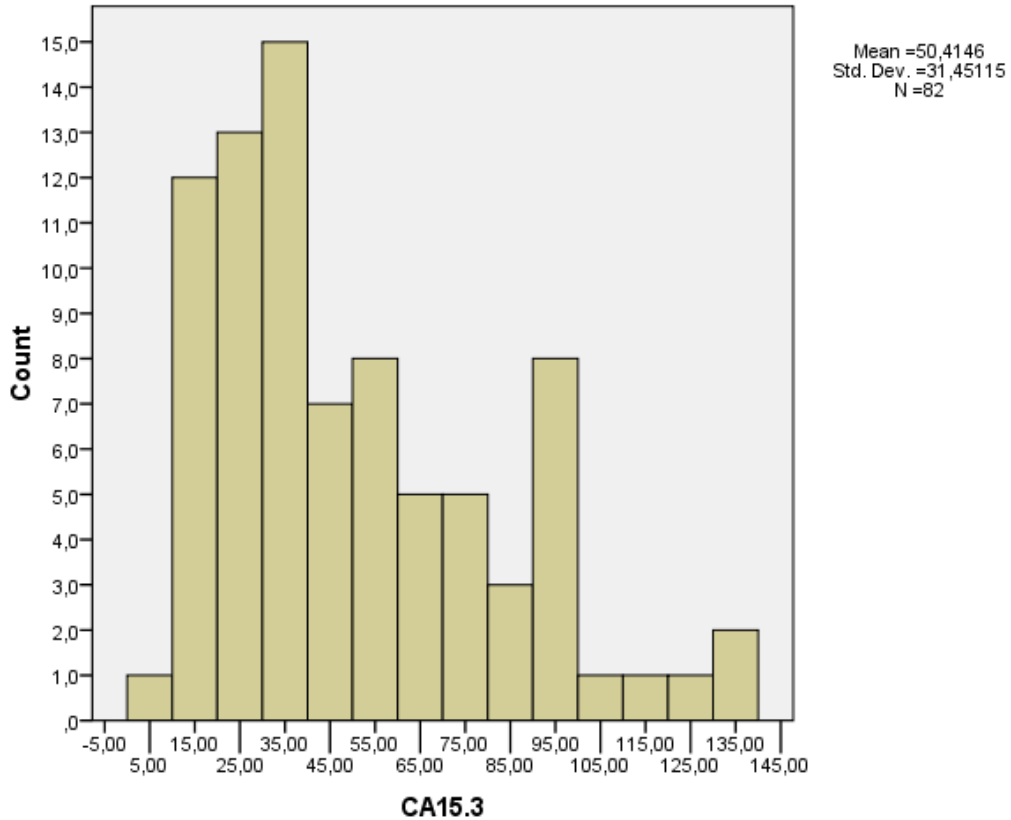
Το συγκεκριμένο ραβδόγραμμα μας δείχνει τον αριθμό των ανδρών και γυναικών οι οποίοι πήραν μέρος στις μετρήσεις μας. Παρατηρούμε ότι στις μετρήσεις έχουν λάβει μέρος 50 άντρες και 43 γυναίκες.

ΙΣΤΟΓΡΑΜΜΑΤΑ ΠΟΣΟΤΙΚΩΝ ΜΕΤΑΒΛΗΤΩΝ



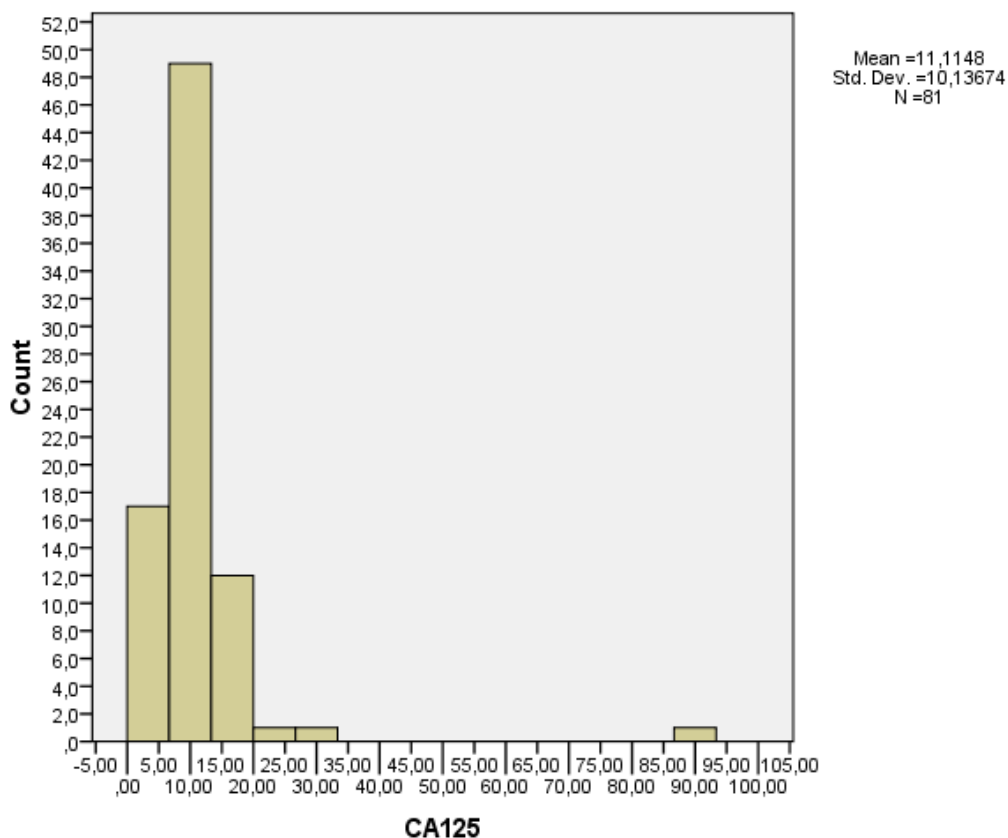
viii. *Ιστόγραμμα που αφορά τις τιμές του δείκτη AFP*

Βλέπουμε ότι το δείγμα μας αποτελείται από 82 τιμές (οι υπόλοιπες τιμές είναι χαμένες τιμές). Η μέση τιμή είναι 2,1756 η οποία πέφτει μέσα στο διάστημα των τιμών αναφοράς για τον συγκεκριμένο δείκτη, όπως επίσης και η μικρότερη τιμή η οποία είναι η τιμή 0 και η μεγαλύτερη τιμή η οποία είναι η τιμή 12,50. Οι τιμές ανάμεσα στο 1 και 2 έχουν την μεγαλύτερη συχνότητα στους ασθενείς από τους οποίους προέρχονται τα δεδομένα.



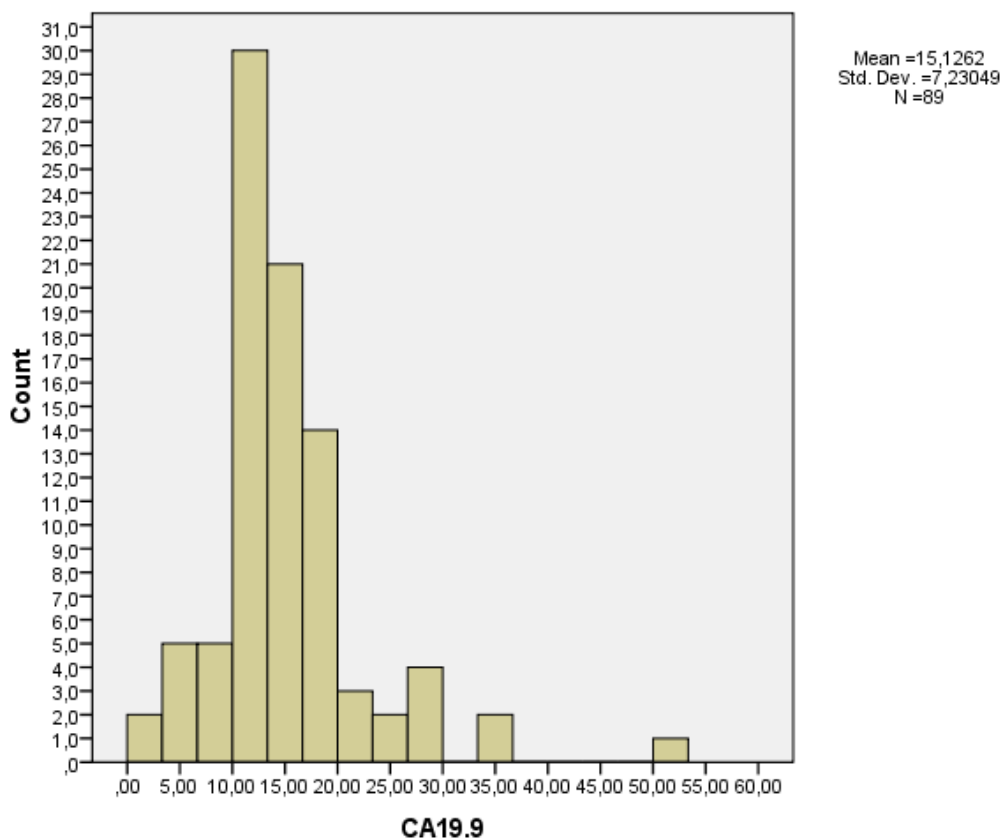
ix. Ιστογράμμα που αφορά τις τιμές του δείκτη CA15.3

Βλέπουμε ότι το δείγμα μας αποτελείται από 82 τιμές (οι υπόλοιπες τιμές είναι χαμένες τιμές). Η μέση τιμή είναι 50,4146 η οποία είναι μεγαλύτερη από τις τιμές του διαστήματος των τιμών αναφοράς για τον συγκεκριμένο δείκτη. Η μικρότερη τιμή είναι η τιμή 4,80 και μεγαλύτερη η τιμή 139,00.



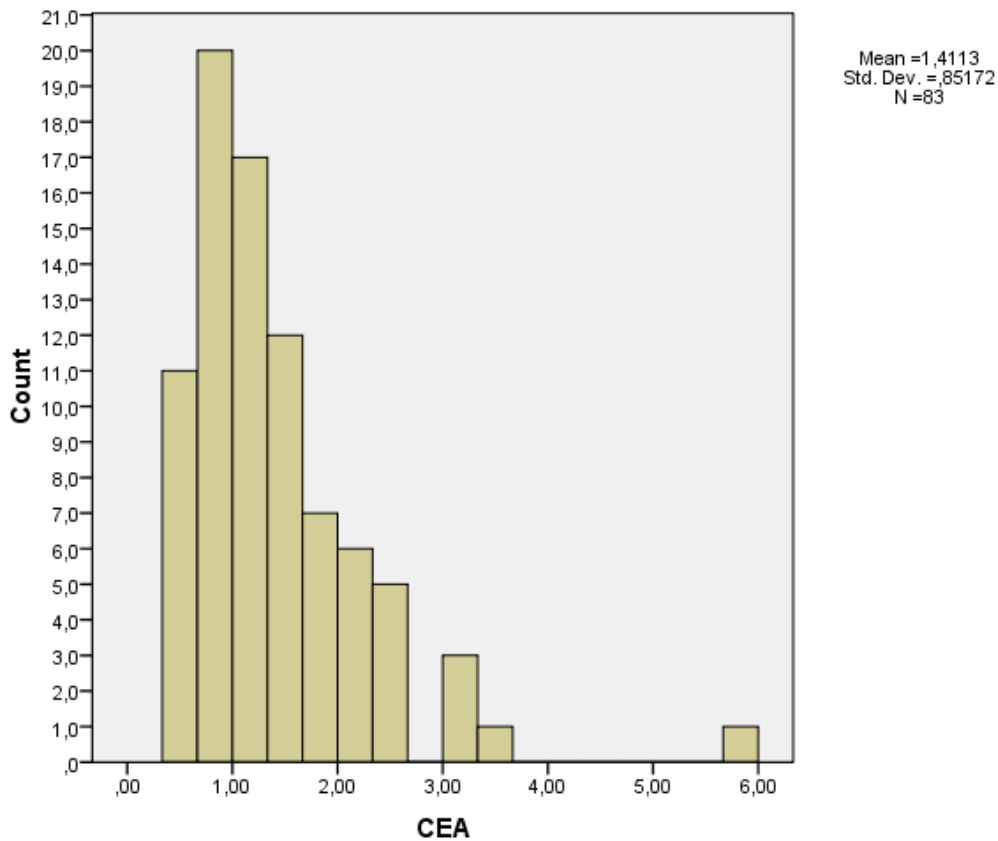
χ.Ιστογράμμο που αφορά τις τιμές του δείκτη CA125

Βλέπουμε ότι το δείγμα μας αποτελείται από 81 τιμές (οι υπόλοιπες τιμές είναι χαμένες τιμές). Η μέση τιμή είναι 11,1148 η οποία ανήκει στο διάστημα των τιμών αναφοράς του συγκεκριμένου δείκτη. Η μικρότερη τιμή είναι η τιμή 2,60 και μεγαλύτερη η τιμή 91,60. Βλέπουμε ότι οι τιμές που ανήκουν στο διάστημα 7,5 μέχρι 10 εμφανίζουν την μεγαλύτερη συχνότητα.



xi. Ιστογράμμα που αφορά τις τιμές του δείκτη CA19.9

Βλέπουμε ότι το δείγμα μας αποτελείται από 89 τιμές (οι υπόλοιπες τιμές είναι χαμένες τιμές). Η μέση τιμή είναι 15,1262 η οποία ανήκει στο διάστημα των τιμών αναφοράς του συγκεκριμένου δείκτη. Η μικρότερη τιμή είναι η τιμή 1,30 και μεγαλύτερη η τιμή 51,79. Βλέπουμε ότι οι τιμές που ανήκουν στο διάστημα 10 μέχρι 12,5 εμφανίζουν την μεγαλύτερη συχνότητα.



xii. Ιστογράμμα που αφορά τις τιμές του δείκτη AFP

Βλέπουμε ότι το δείγμα μας αποτελείται από 83 τιμές (οι υπόλοιπες τιμές είναι χαμένες τιμές). Η μέση τιμή είναι 1,4113 η οποία ανήκει στο διάστημα των τιμών αναφοράς του συγκεκριμένου δείκτη. Η μικρότερη τιμή είναι η τιμή 0,37 και μεγαλύτερη η τιμή 5,71. Βλέπουμε ότι οι τιμές που ανήκουν στο διάστημα 0,6 μέχρι 1 εμφανίζουν την μεγαλύτερη συχνότητα.

ΕΛΕΓΧΟΣ ΤΩΝ ΜΕΣΩΝ ΤΙΜΩΝ ΤΩΝ ΙΑΤΡΙΚΩΝ ΔΕΙΚΤΩΝ ΣΕ ΑΝΤΡΕΣ ΚΑΙ ΓΥΝΑΙΚΕΣ – ΕΦΑΡΜΟΓΗ ΤΟΥ Independent samples t-test

Το t τεστ έχει δύο “κατευθύνσεις”. Η μία κατεύθυνση είναι αυτή που δεν μπορούμε να υποθέσουμε ότι οι διακυμάνσεις των δύο δειγμάτων είναι περίπου ίσες και αυτή που μπορούμε να υποθέσουμε ότι είναι ίσες. Ο παραπάνω πίνακας έχει δύο γραμμές αποτελεσμάτων, η πρώτη αναφέρεται στην περίπτωση που μπορούμε να υποθέσουμε ισότητα των δύο διακυμάνσεων και η δεύτερη στην περίπτωση που δεν μπορούμε να υποθέσουμε ισότητα των δύο διακυμάνσεων. Ο πίνακας είναι χωρισμένος σε δύο κατηγορίες αποτελεσμάτων, η μία αφορά το Levene για την ισότητα των διακυμάνσεων και η άλλη περιέχει τα αποτελέσματα του t τεστ που επιλέξαμε να κάνουμε. Όπως αναφέραμε, ο πίνακας έχει δύο γραμμές αποτελεσμάτων, το αν θα κοιτάξουμε την πρώτη ή τη δεύτερη γραμμή αποτελεσμάτων του t τεστ θα μας το “πει” το τεστ του Levene. Το τεστ του Levene ελέγχει την υπόθεση της ισότητας των δύο διακυμάνσεων και υπολογίζει μία p-value. Αν η p-value είναι μικρότερη του 0,05, απορρίπτεται η υπόθεση της ισότητας των διακυμάνσεων. Στην αντίθετη περίπτωση δεν απορρίπτεται. Επομένως, ανάλογα με την p-value (Sig.) του τεστ του Levene, κοιτάζουμε την πρώτη ή τη δεύτερη γραμμή αποτελεσμάτων.

AFP

Στην προκειμένη περίπτωση η p-value είναι μεγαλύτερη του 0.05, άρα μπορούμε να υποθέσουμε ισότητα των δύο διακυμάνσεων. Επομένως θα κοιτάξω τη πρώτη γραμμή αποτελεσμάτων του πίνακα. Η p-value για τον έλεγχο της ισότητας των δύο μέσων είναι ίση με 0.615 (Sig. (2-tailed)). Άρα η μηδενική υπόθεση δεν απορρίπτεται, δηλαδή σε επίπεδο στατιστικής σημαντικότητας $\alpha=5\%$, οι μέσες τιμές των δύο πληθυσμών (αντρών – γυναικών) από τα οποία προήλθαν τα δύο δείγματα δεν διαφέρουν στατιστικά σημαντικά.

CA125

Η p-value είναι μεγαλύτερη του 0.05, άρα μπορούμε να υποθέσουμε ισότητα των δύο διακυμάνσεων. Επομένως θα κοιτάξω τη πρώτη γραμμή

αποτελεσμάτων του πίνακα. Η p-value για τον έλεγχο της ισότητας των δύο μέσων είναι ίση με 0.382 (Sig. (2-tailed)). Άρα η μηδενική υπόθεση δεν απορρίπτεται, δηλαδή οι μέσοι των δύο πληθυσμών από τα οποία προήλθαν τα δύο δείγματα δεν διαφέρουν στατιστικά σημαντικά.

CA15.3

Η p-value είναι μεγαλύτερη του 0.05, άρα μπορούμε να υποθέσουμε ισότητα των δύο διακυμάνσεων. Επομένως θα κοιτάξω τη πρώτη γραμμή αποτελεσμάτων του πίνακα. Η p-value για τον έλεγχο της ισότητας των δύο μέσων είναι ίση με 0.812 (Sig. (2-tailed)). Άρα η μηδενική υπόθεση δεν απορρίπτεται, δηλαδή οι μέσοι των δύο πληθυσμών από τα οποία προήλθαν τα δύο δείγματα δεν διαφέρουν στατιστικά σημαντικά.

CA19.9

Η p-value είναι μεγαλύτερη του 0.05, άρα μπορούμε να υποθέσουμε ισότητα των δύο διακυμάνσεων. Επομένως θα κοιτάξω τη πρώτη γραμμή αποτελεσμάτων του πίνακα. Η p-value για τον έλεγχο της ισότητας των δύο μέσων είναι ίση με 0.732 (Sig. (2-tailed)). Άρα η μηδενική υπόθεση δεν απορρίπτεται, δηλαδή οι μέσοι των δύο πληθυσμών από τα οποία προήλθαν τα δύο δείγματα δεν διαφέρουν στατιστικά σημαντικά.

CEA

Η p-value είναι μεγαλύτερη του 0.05, άρα μπορούμε να υποθέσουμε ισότητα των δύο διακυμάνσεων. Επομένως θα κοιτάξω τη πρώτη γραμμή αποτελεσμάτων του πίνακα. Η p-value για τον έλεγχο της ισότητας των δύο μέσων είναι ίση με 0.572 (Sig. (2-tailed)). Άρα η μηδενική υπόθεση δεν απορρίπτεται, δηλαδή οι μέσοι των δύο πληθυσμών από τα οποία προήλθαν τα δύο δείγματα δεν διαφέρουν στατιστικά σημαντικά.

Συμπέρασμα

Σε όλους τους ιατρικούς δείκτες, οι μέσες τιμές των δεικτών δεν διαφέρουν στατιστικά σημαντικά ανάμεσα στους άντρες και στις γυναίκες.

ΕΛΕΓΧΟΣ ΤΩΝ ΜΕΣΩΝ ΤΙΜΩΝ ΤΩΝ ΙΑΤΡΙΚΩΝ ΔΕΙΚΤΩΝ ΑΝΑ ΔΥΟ – ΕΦΑΡΜΟΓΗ ΤΟΥ Paired samples t-test

CA 15.3 – CA 125

Τα στατιστικά του ελέγχου υποθέσεως είναι: $t = 11,107$, $df = 79$, $sig = 0,000$. Παρατηρούμε ότι η σημαντικότητα είναι ίση με 0,000. Αφού είναι μικρότερη του 0,05 συμπεραίνουμε ότι οι μέσοι των δεικτών για τους 2 δείκτες (CA 15.3 και CA 125) διαφέρουν στατιστικά σημαντικά.

CA 15.3 – CA 19.9

Τα στατιστικά του ελέγχου υποθέσεως είναι: $t = 10,061$, $df = 80$, $sig = 0,000$. Παρατηρούμε ότι η σημαντικότητα είναι ίση με 0,000. Αφού είναι μικρότερη του 0,05 συμπεραίνουμε ότι οι μέσοι των δεικτών για τους 2 δείκτες (CA 15.3 και CA 19.9) διαφέρουν στατιστικά σημαντικά.

CA 15.3 – CEA

Τα στατιστικά του ελέγχου υποθέσεως είναι: $t = 14,130$, $df = 79$, $sig = 0,000$. Παρατηρούμε ότι η σημαντικότητα είναι ίση με 0,000. Αφού είναι μικρότερη του 0,05 συμπεραίνουμε ότι οι μέσοι των δεικτών για τους 2 δείκτες (CA 15.3 και CEA) διαφέρουν στατιστικά σημαντικά.

CA 125 – CA19.9

Τα στατιστικά του ελέγχου υποθέσεως είναι: $t = -3,098$, $df = 80$, $sig = 0,003$. Παρατηρούμε ότι η σημαντικότητα είναι ίση με 0,003. Αφού είναι μικρότερη του 0,05 συμπεραίνουμε ότι οι μέσοι των δεικτών για τους 2 δείκτες (CA 125 και CA19.9) διαφέρουν στατιστικά σημαντικά.

CA 125 – CEA

Τα στατιστικά του ελέγχου υποθέσεως είναι: $t = 8,505$, $df = 80$, $sig = 0,000$. Παρατηρούμε ότι η σημαντικότητα είναι ίση με 0,000. Αφού είναι μικρότερη του 0,05 συμπεραίνουμε ότι οι μέσοι των δεικτών για τους 2 δείκτες (CA 125 και CEA) διαφέρουν στατιστικά σημαντικά.

CA 19.9 – CEA

Τα στατιστικά του ελέγχου υποθέσεως είναι: $t = 19,111$, $df = 80$, $sig = 0,000$. Παρατηρούμε ότι η σημαντικότητα είναι ίση με 0,000. Αφού είναι μικρότερη του 0,05 συμπεραίνουμε ότι οι μέσοι των δεικτών για τους 2 δείκτες (CA 19.9 και CEA) διαφέρουν στατιστικά σημαντικά.

ΣΥΣΧΕΤΙΣΗ ΤΩΝ ΜΕΤΑΒΛΗΤΩΝ ΑΝΑ ΔΥΟ

Συντελεστής συσχέτισης AFP,CA15.3

Ο συντελεστής συσχέτισης όσον αφορά τους δείκτες AFP και CA15.3 είναι 0.136 το οποίο σημαίνει ότι μεταξύ τους υπάρχει ασθενής γραμμική συσχέτιση.

Συντελεστής συσχέτισης AFP,CA125

Ο συντελεστής συσχέτισης όσον αφορά τους δείκτες AFP και CA125 είναι 0.062 το οποίο σημαίνει ότι μεταξύ τους υπάρχει μηδενική γραμμική συσχέτιση.

Συντελεστής συσχέτισης AFP,CA19.9

Ο συντελεστής συσχέτισης όσον αφορά τους δείκτες AFP και CA19.9 είναι -0.014 το οποίο σημαίνει ότι μεταξύ τους υπάρχει μηδενική γραμμική συσχέτιση.

Συντελεστής συσχέτισης AFP,CEA

Ο συντελεστής συσχέτισης όσον αφορά τους δείκτες AFP και CA19.9 είναι 0.050 το οποίο σημαίνει ότι μεταξύ τους υπάρχει μηδενική γραμμική συσχέτιση.

Συντελεστής συσχέτισης CA15.3,CA125

Ο συντελεστής συσχέτισης όσον αφορά τους δείκτες CA15.3 και CA125 είναι 0.114 το οποίο σημαίνει ότι μεταξύ τους υπάρχει ασθενής γραμμική συσχέτιση.

Συντελεστής συσχέτισης CA15.3,CA19.9

Ο συντελεστής συσχέτισης όσον αφορά τους δείκτες CA15.3 και CA19.9 είναι -0.022 το οποίο σημαίνει ότι μεταξύ τους υπάρχει μηδενική γραμμική συσχέτιση.

Συντελεστής συσχέτισης CA15.3,CEA

Ο συντελεστής συσχέτισης όσον αφορά τους δείκτες CA15.3 και CEA είναι -0.218 πράγμα το οποίο σημαίνει ότι μεταξύ τους υπάρχει ασθενής γραμμική συσχέτιση.

Συντελεστής συσχέτισης CA125,CA19.9

Ο συντελεστής συσχέτισης όσον αφορά τους δείκτες CA125 και CA19.9 είναι 0.196 πράγμα το οποίο σημαίνει ότι μεταξύ τους υπάρχει ασθενής γραμμική συσχέτιση.

Συντελεστής συσχέτισης CA125,CEA

Ο συντελεστής συσχέτισης όσον αφορά τους δείκτες CA125 και CEA είναι -0.139 πράγμα το οποίο σημαίνει ότι μεταξύ τους υπάρχει ασθενής αρνητική γραμμική συσχέτιση.

Συντελεστής συσχέτισης CA19.9,CEA

Ο συντελεστής συσχέτισης όσον αφορά τους δείκτες CA19.9 και CEA είναι 0.010 πράγμα το οποίο σημαίνει ότι μεταξύ τους υπάρχει μηδενική γραμμική συσχέτιση.

Συμπέρασμα

Βλέπουμε ότι όταν συγκρίνουμε τις μεταβλητές ανά δύο ότι ο συντελεστής συσχέτισης παίρνει τιμές οι οποίες βρίσκονται πάρα πολύ κοντά στο μηδέν πράγμα το οποίο σημαίνει ότι οι ανά δυο μεταβλητές μεταξύ τους είναι ασυσχέτιστες.

ΤΟ ΜΟΝΤΕΛΟ ΠΡΟΒΛΕΨΗΣ ΤΟΥ ΙΑΤΡΙΚΟΥ ΔΕΙΚΤΗ AFP ΣΕ ΣΧΕΣΗ ΜΕ ΤΟΥΣ ΔΕΙΚΤΕΣ CA19.9, CA15.3, CA125

Ο δείκτης R^2 εκφράζει το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής το οποίο ερμηνεύεται από τη διακύμανση των τιμών της ανεξάρτητης μεταβλητής ή των ανεξάρτητων μεταβλητών. Ο συντελεστής αυτός ονομάζεται συντελεστής προσδιορισμού και υποδεικνύει την ποιότητα προσαρμογής της εξίσωσης παλινδρόμησης στα δεδομένα.

Συντελεστής προσδιορισμού του δείκτη AFP σε σχέση με τον δείκτη CA19.9.

Σε αυτήν την εφαρμογή μας, είναι $R^2 = 0,065$, δηλαδή το 6,5% της διακύμανσης των τιμών του καρκινικού δείκτη AFP ερμηνεύεται από την διακύμανση των τιμών του καρκινικού δείκτη CA19.9.

Συντελεστής προσδιορισμού του δείκτη AFP σε σχέση με τον δείκτη CA15.3.

Σε αυτήν την εφαρμογή μας, είναι $R^2 = 0,079$, δηλαδή το 7,9% της διακύμανσης των τιμών του καρκινικού δείκτη AFP ερμηνεύεται από την διακύμανση των τιμών του καρκινικού δείκτη CA15.3.

Συντελεστής προσδιορισμού του δείκτη AFP σε σχέση με τον δείκτη CA125.

Σε αυτήν την εφαρμογή μας, είναι $R^2 = 0,062$, δηλαδή το 6,2% της διακύμανσης των τιμών του καρκινικού δείκτη AFP ερμηνεύεται από την διακύμανση των τιμών του καρκινικού δείκτη CA125.

Συντελεστής προσδιορισμού του δείκτη AFP σε σχέση με τους δείκτες CA19.9, CA15.3.

Σε αυτήν την εφαρμογή μας, είναι $R^2 = 0,25$, δηλαδή το 25% της διακύμανσης των τιμών του καρκινικού δείκτη AFP ερμηνεύεται από την διακύμανση των τιμών των καρκινικών δεικτών CA19.9 και CA15.3.

Συντελεστής προσδιορισμού του δείκτη AFP σε σχέση με τους δείκτες CA19.9, CA125.

Σε αυτήν την εφαρμογή μας, είναι $R^2 = 0,40$, δηλαδή το 40% της διακύμανσης των τιμών του καρκινικού δείκτη AFP ερμηνεύεται από την διακύμανση των τιμών των καρκινικών δεικτών CA19.9 και CA125.

Συντελεστής προσδιορισμού του δείκτη AFP σε σχέση με τους δείκτες CA15.3, CA125.

Σε αυτήν την εφαρμογή μας, είναι $R^2 = 0,65$, δηλαδή το 65% της διακύμανσης των τιμών του καρκινικού δείκτη AFP ερμηνεύεται από την διακύμανση των τιμών των καρκινικών δεικτών CA15.3 και CA125.

Συντελεστής προσδιορισμού του δείκτη AFP σε σχέση με τους δείκτες CA19.9, CA15.3, CA125

Σε αυτήν την εφαρμογή μας, είναι $R^2 = 0,85$, δηλαδή το 85% της διακύμανσης των τιμών του καρκινικού δείκτη AFP ερμηνεύεται από την διακύμανση των τιμών των καρκινικών δεικτών CA19.9, CA15.3, CA125.

Συμπέρασμα:

Το μοντέλο με τον μεγαλύτερο συντελεστή προσδιορισμού είναι το μοντέλο με τους καρκινικούς δείκτες CA19.9, CA15.3, CA125 και μάλιστα το ποσοστό ερμηνεύσης του δείκτη AFP από τους υπόλοιπους τρεις δείκτες είναι αρκετά ικανοποιητικό ώστε να μπορούν οι γιατροί να το χρησιμοποιούν ως μοντέλο μελλοντικής πρόβλεψης του δείκτη αυτού όταν είναι γνωστοί οι υπολοίποι τρεις καρκινικοί δείκτες.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Βλέπουμε ότι η στατιστική από ανέκαθεν έπαιζε και θα συνεχίσει να παίζει σημαντικό ρόλο στην ζωή μας. Από τα αρχαία χρόνια μέχρι και σήμερα χρησιμοποιείται σε διάφορους τομείς της καθημερινότητας μας βοηθώντας μας να εξάγουμε χρήσιμα συμπεράσματα. Αυτά τα συμπεράσματα όμως για να είναι χρήσιμα πρέπει να τύχουν κατάλληλης επεξεργασίας. Με την βοήθεια της περιγραφικής στατιστικής μπορούμε να σχηματίσουμε μια συνοπτική διαδικασία του δείγματος μας με πίνακες και γραφήματα και με την επαγωγική στατιστική να επεκτείνουμε τα συμπεράσματα μας σε ολόκληρο τον πληθυσμό με την προϋπόθεση να έχουμε όσο το δυνατόν μικρότερο κίνδυνο λάθους.

Η εφαρμογή της στατιστικής επιστήμης στην επίλυση βιολογικών προβλημάτων μας οδήγησε στη δημιουργία του κλάδου της Βιοστατιστικής ο οποίος κλάδος αυτός ασχολείται με την επεξεργασία και ανάλυση στοιχείων που αφορούν βιολογικούς οργανισμούς.

ΒΙΒΛΙΟΓΡΑΦΙΑ

4. Shortliffe E.H., Perreault, L.E. Medical Informatics: Computer Applications in Health Care, Addison–Welsey, 1990.
5. Komaroff A.L., The variability and inaccuracy of Medical data, Proceedings of the IEEE, 67: 1196, 1979.
6. Blum B.I. Clinical Information Systems, Springer Verlag, 1996.
7. Ιατρικό Βήμα
8. Βιβλίο Βασική Ιατρική Στατιστική (Τρίτη Έκδοση), Βασίλειος Χρ. Κατσουγιαννόπουλος
9. Αλιβιζάτος, Γ.Π (1953): Στατιστική Μεθοδολογία
10. Altman, D.G (1991): Practical Statistics for Medical Research. Chapman, London
11. Βαλαώρας, Β.Γ (1943) Στοιχεία Βιομετρίας και Στατιστικής, Βαφειαδάκης, Αθήναι.
12. Daniel, W.W (1991): Biostatistics, J. Wiley and sons Inc. New York
13. Πληροφορίες από τον παγκόσμιο ιστό
<http://www.bestrong.org.gr/el/learncancer/cancerdetection/cancerbloodtests/>
<http://www.scribd.com/doc/900538/->
<http://gynaikologos-manolis.pblogs.gr/2009/06/pws-mporesan-ta-computers-na-bohthhsoyn-thn-iatrikh.html>
<http://www.aua.gr/gpapadopoulos/files/sisxetisi091.pdf>
<http://7.nsa-virtualeducation.com/images/corri.pdf>

ΠΕΡΙΕΧΟΜΕΝΑ

Ευχαριστίες.....	2
Πρόλογος.....	3
Περίληψη.....	4
Κεφάλαιο 1	
1.1 Η επιστήμη της Στατιστικής.....	6
1.2 Ιστορία της ιατρικής στατιστικής.....	10
1.3 Βασική στατιστική ορολογία.....	13
Κεφάλαιο 2	
2.1 Η χρήση των ηλεκτρονικών υπολογιστών στην ιατρική στατιστική.....	24
2.2 Η στατιστική στην ιατρική και στις βιολογικές επιστήμες.....	25
2.3 Ασφάλεια ιατρικών δεδομένων.....	30
Κεφάλαιο 3	
3.1 Οι υπολογιστές στην Ιατρική Επιστήμη.....	37
3.2 Φύση και δομή της ιατρικής πληροφορίας.....	38
Κεφάλαιο 4	
4.1 Ενόητες Στατιστικής.....	45

4.2 Πίνακες.....	49
4.3 Διαγράμματα – Η αναγκαιότητα των διαγραμμάτων.....	51
4.4 Κατανομές συχνοτήτων.....	55
4.5 Επαγωγική Στατιστική.....	57
Κεφάλαιο 5	
5.1 SPSS.....	60
5.2 Ορισμός της t-κατανομής.....	62
5.3 Η δοκιμασία t του student.....	63
5.4 Το t-test για ανεξάρτητα δείγματα.....	63
5.5 Το t-test για ζευγαρωτά δείγματα.....	64
5.6 ANOVA.....	64
5.7 Τεχνικές στο SPSS.....	65
5.8 Γραμμική παλινδρόμηση.....	66
5.9 Συσχέτιση.....	67
Κεφάλαιο 6	
Πρακτική εφαρμογή.....	69
Βιβλιογραφία.....	89