ΚΕΦΑΛΑΙΟ 1
ΕΙΣΑΓΩΓΗ ΣΤΑ ΒΑΣΙΚΑ ΠΡΟΒΛΗΜΑΤΑ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ

1.1 Αντικείμενο της Στατιστικής και Στατιστική Σκέψη

Πολλές απόψεις έχουν διατυπωθεί για το ποιό ακριβώς είναι το αντικείμενο της Στατιστικής. Σε πολλά εισαγωγικά βιβλία **ως Στατιστική ορίζεται η επιστήμη εκείνη που ασχολείται με τη συλλογή, την ανάλυση και την ερμηνεία δεδομένων**. Παρότι ο ορισμός αυτός αντιπροσωπεύει πράγματι ένα μεγάλο μέρος των δραστηριοτήτων της Στατιστικής δεν αποτελεί το αποκλειστικό αντικείμενο της επιστήμης αυτής. Αυτή η πλευρά των δραστηριοτήτων της Στατιστικής αποτελεί αυτό που ονομάζουμε **Περιγραφική Στατιστική**. Η άλλη διάσταση της Στατιστικής είναι εκείνη η οποία ασχολείται με την **συμπερασματολογία**. Για αυτή την πλευρά της Στατιστικής θα μπορούσαμε να δίναμε τον ορισμό ότι **Στατιστική είναι η προσπάθεια εξαγωγής συμπερασμάτων κάτω από συνθήκες αβεβαιότητας.**

Θα μπορούσε να ισχυρισθεί κανείς ότι η Στατιστική συμπερασματολογία είναι περισσότερο σημαντική. Ο λόγος, ίσως, είναι ότι χρειάζεται περισσότερο πολύπλοκα "εργαλεία" προκειμένου να αναπτυχθεί. Τα τελευταία χρόνια όμως και η περιγραφική Στατιστική βρίσκει όλο και περισσότερες εφαρμογές. Αυτό οφείλεται στο ότι σε όλες σχεδόν τις επιστήμες υπάρχει ανάγκη ποσοτικής προσέγγισης των διαφόρων εννοιών και μεθόδων. Αυτό γίνεται με συγκέντρωση, ανάλυση και παρουσίαση των υπαρχόντων στοιχείων.

Και οι δύο βασικές δραστηριότητες της Στατιστικής που αναπτύχθηκαν προηγουμένως χρειάζονται ικανότητα *στατιστικής σκέψης*. Η ικανότητα στατιστικής σκέψης πρέπει βέβαια να συνδυάζεται και με *γνώση του αντικειμένου* από το οποίο προέρχονται τα προς ανάλυση στοιχεία.

Είναι γνωστό ότι πολλοί ισχυρίζονται πως η Στατιστική είναι

ένας τρόπος για να λέει κανείς ψέματα. Στην πραγματικότητα βέβαια δεν είναι η Στατιστική υπεύθυνη για αυτό αλλά οι μη κατάλληλες στατιστικές μέθοδοι που χρησιμοποιούνται για την ανάλυση συγκεκριμένων δεδομένων ή η παραπλανητική ερμηνεία δεδομένων που γίνεται από μη στατιστικούς για εξυπηρέτηση συγκεκριμένων σκοπιμοτήτων. Είναι δηλαδή αποτέλεσμα λανθασμένης στατιστικής σκέψης. Γενικά θα μπορούσε να πει κανείς ότι η στατιστική σκέψη μας παρέχει τη δυνατότητα βασικής κατανόησης των στατιστικών μεθόδων και μας βοηθά να ανακαλύψουμε επαναλαμβανόμενες διαδικασίες (patterns) με την αξιοποίηση διαθέσιμων δεδομένων.

Ενα από τα σημαντικότερα βήματα για να μεγιστοποιήσει κανείς τη χρησιμότητα των στατιστικών εννοιών και μεθόδων βρίσκεται στην επιλογή των δεδομένων. Οσο πιο σχετικά είναι τα δεδομένα με το πρόβλημα που μας ενδιαφέρει να αντιμετωπίσουμε τόσο χρησιμότερη είναι η ανάλυσή τους. Η ενασχόληση με τον τρόπο συλλογής των υπό ανάλυση δεδομένων μπορεί να αποδώσει ιδιαίτερα χρήσιμα αποτελέσματα στο στάδιο της αξιοποίησης τους. Η συλλογή των δεδομένων και η ανάλυση που θα επακολουθήσει επηρεάζεται από τη γνώση του αντικειμένου από εκείνον, ο οποίος ενδιαφέρεται να πάρει κάποια απόφαση και από τις πληροφορίες που είναι αναγκαίες προκειμένου να ληφθεί απόφαση για ένα συγκεκριμένο πρόβλημα.

Καθοριστικό ρόλο στην ανάλυση δεδομένων παίζει η αντίληψη της έννοιας της **διακύμανσης (variation)** ή **μεταβλητότητας**. Η μεταβλητότητα είναι αναπόφευκτη σε όλες τις πλευρές της ανθρώπινης δραστηριότητας. Για οποιοδήποτε θέμα που αναφέρεται στον άνθρωπο και στο περιβάλλον του είναι προφανές ότι υπάρχει μεταβλητότητα. Οι άνθρωποι γύρω μας δεν έχουν το ίδιο βάρος, το περιεχόμενο σε μια συγκεκριμένη συσκευασία ενός προϊόντος δεν είναι ποτέ *ακριβώς* το ίδιο, το εμπόρευμα που πωλούν δύο υπάλληλοι με τα ίδια προσόντα στην ίδια χρονική περίοδο δεν είναι το ίδιο. Ετσι, ανεξάρτητα από το φαινόμενο από το οποίο προέρχεται ένα σύνολο δεδομένων, υπάρχει μεταβλητότητα στις τιμές των δεδομένων αυτών. Κατανόηση της

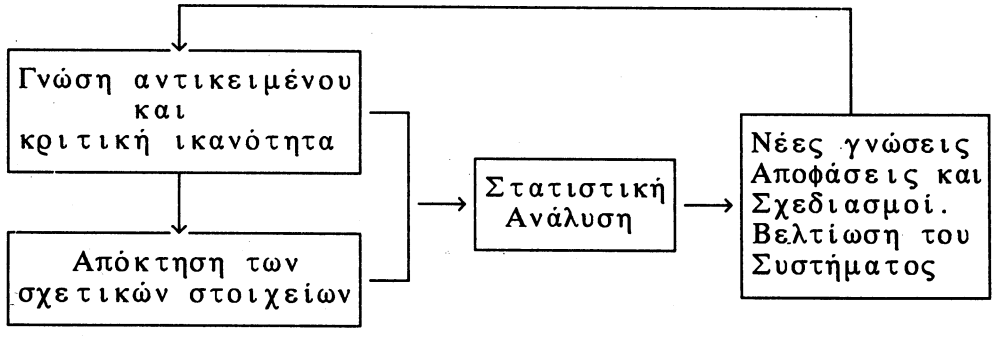
μεταβλητότητας και των λόγων που την προκαλούν είναι απαραίτητα για την ερμηνεία των δεδομένων. Θα μπορούσε κανείς να ισχυρισθεί ότι η κατανόηση και ερμηνεία της μεταβλητότητας σε ένα σύνολο δεδομένων είναι ακριβώς αυτό με το οποίο ασχολείται η Στατιστική. Η κατανόηση της μεταβλητότητας των δεδομένων οδηγεί στην ανακάλυψη, περιγραφή και κατανόηση συσχετίσεων και επαναλαμβανόμενων διαδικασιών (patterns). Η γνώση αυτή αποτελεί συχνό τη βάση για αποφάσεις που παίρνουμε σε σχέση με φαινόμενα που προκάλεσαν τα υπό ανάλυση δεδομένα.

Η έννοια της μεταβλητότητας είναι ίσως αυτή ακριβώς που οδήγησε στη σημαντική ανάπτυξη και αξιοποίηση των μεθόδων των Πιθανοτήτων και της Στατιστικής τα τελευταία χρόνια σε κατεύθυνση διαφορετική από εκείνη των Μαθηματικών. Τα Μαθηματικά, όπως είναι γνωστό, ασχολούνται με συγκεκριμένες και σαφώς καθορισμένες διαδικασίες όπου ένα σύνολο συγκεκριμένων υποθέσεων μπορεί να οδηγήσει σε ένα μονοσήμαντο αποτέλεσμα. Αντίθετα, η Στατιστική δημιουργήθηκε από την ανάγκη μελέτης φαινομένων όπου κάτω από τις ίδιες ακριβώς συνθήκες είναι δυνατόν να καταλήξουν σε διαφορετικά αποτελέσματα λόγω της ύπαρξης της μεταβλητότητας.

Θα μπορούσε επομένως κανείς να ισχυρισθεί ότι **στατιστική σκέψη είναι μια διαδικασία συλλογισμών που αναγνωρίζει ότι υπάρχει μεταβλητότητα σε όλα τα φαινόμενα και ότι η μελέτη της μεταβλητότητας οδηγεί σε νέες γνώσεις και καλύτερες αποφάσεις.**

Το σχήμα που ακολουθεί εξηγεί την χρησιμοποίηση της Στατιστικής ανάλυσης σε συνδυασμό με την γνώση του αντικειμένου και την κρίση αυτού που λαμβάνει τις αποφάσεις όπως αυτά συνδέονται με το σχεδιασμό, τη λήψη αποφάσεων και τη βελτίωση συστημάτων. Οπως φαίνεται από το σχήμα αυτό πρόκειται για μία διαδικασία συνεχούς βελτίωσης. Οσο περισσότερο κατανοούμε ένα φαινόμενο με τη χρήση της Στατιστικής ανάλυσης, τόσο περισσότερο βελτιώνουμε τη γνώση μας για το αντικείμενο και έτσι καθίσταται

εφικτή η καλύτερη κατανόηση της διαδικασίας την επόμενη φορά.



Σχήμα 1.1.1

1.2 Ενδεικτικές περιπτώσεις εφαρμογής της Στατιστικής

Ας δούμε τώρα μερικές περιπτώσεις προβλημάτων που θα απαιτήσουν ενδεχομένως την συμβολή ενός Στατιστικού. Τα προβλήματα αυτά θα μας δώσουν και ενδείξεις των κυρίων στοιχείων ενός στατιστικού προβλήματος.

1. Προκειμένου να προβλεφθεί το αποτέλεσμα των εκλογών, οι εταιρείες σφυγμομετρήσεων της κοινής γνώμης ερωτούν, με βάση ένα ερωτηματολόγιο, έναν προκαθορισμένο αριθμό ψηφοφόρων από όλη τη χώρα και καταγράφουν τις προτιμήσεις τους. Με βάση τις πληροφορίες που συλλέγονται οι εταιρείες κάνουν κάποια πρόβλεψη. Παρόμοια προβλήματα συναντώνται σε έρευνες αγοράς (ποιό είναι το ποσοστό των μελλοντικών αγοραστών που θα προτιμήσουν κάποια συγκεκριμένη μάρκα αυτοκινήτου;), στην Κοινωνιολογία (ποιό είναι το ποσοστό των σπιτιών στην επαρχία που έχουν τηλέφωνο;), στη Βιομηχανία (ποιό είναι το ποσοστό προιόντων που έχουν αγορασθεί ή παραχθεί που είναι ελαττωματικά;).

2. Σε έναν ελεγκτή του Υπουργείου Υγείας ανατίθεται να καθορίσει το αποθεματικό ενός μεγάλου νοσοκομείου. Οπως είναι φυσικό, το να μετρήσει ο ελεγκτής αυτός τον αριθμό όλων των αναλωσίμων και μη αναλωσίμων υλικών στο απόθεμα και να προσδιορίσει την αξία του είναι πολυδάπανο και χρονοβόρο και είναι πολύ πιθανό να υπόκειται σε λάθη εξαιτίας του μεγέθους του αποθέματος. Προκειμένου να ελαττωθεί το κόστος και να αποκτηθεί μία αξιόπιστη εκτίμηση της αξίας, ο ελεγκτής επιλέγει ένα δείγμα ειδών από τον κατάλογο των των αναλωσίμων και των εργαλείων του νοσοκομείου και με προσοχή καταμετρά τον αριθμό των μονάδων κάθε είδους που είναι διαθέσιμο και καταγράφει την συνολική αξία του είδους. Ο λόγος της συνολικής αξίας του δείγματος αυτού των ειδών προς τη συνολική αξία που προκύπτει από τα στοιχεία του νοσοκομείου οδηγεί σε μια εκτίμηση του ελλείμματος που οφείλεται σε κλοπές, στη μη καταγραφή χρησιμοποιηθέντων στοιχείων κ.λ.π.. Αυτό το ποσοστό ελλείμματος μπορεί στη συνέχεια να εφαρμοσθεί στην συνολική αξία του αποθεματικού που εμφανίζεται στα βιβλία του νοσοκομείου έτσι ώστε να αποκτηθεί μία εκτίμηση της πραγματικής αξίας του τρέχοντος αποθεματικού. Πόσο ακριβής είναι η εκτίμηση αυτή; Σε τι έκταση περιμένουμε η εκτίμηση αυτή να αποκλίνει από την πραγματική τιμή του αποθεματικού του νοσοκομείου;

3. Η παραγωγή ενός χημικού εργαστασίου εξαρτάται από πολλούς παράγοντες. Παρατηρώντας τους παράγοντες αυτούς και την παραγωγή για μια χρονική περίοδο είμαστε σε θέση να κατασκευάσουμε μια εξίσωση πρόβλεψης που να συσχετίζει την παραγωγή με τους παρατηρηθέντες παράγοντες. Στην ίδια κατεύθυνση μπορούμε να θεωρήσουμε έναν οικονομολόγο ο οποίος επιθυμεί να κατασκευάσει μία εξίσωση πρόβλεψης που θα είναι χρήσιμη για να προβλέπει την ανάπτυξη ή κάποιο άλλο μέτρο της οικονομίας ως συνάρτηση άλλων μεταβλητών. Με όμοιο τρόπο ένας μάνατζερ, ενδεχομένως, ενδιαφέρεται να προβλέψει τις πωλήσεις ενός προιόντος ως συνάρτηση

του προυπολογισμού διαφήμισης, του αριθμού των πωλητών που εργάζονται στην επιχείρηση ή διαφόρων άλλων μεταβλητών που ενδεχομένως σχετίζονται με τις πωλήσεις της συγκεκριμένης εταιρείας.
Πώς μπορούμε να βρούμε μία "καλή" εξίσωση πρόβλεψης ; Εαν η εξίσωση αυτή χρησιμοποιηθεί για να προβλέψει την παραγωγή είναι φυσικό ότι η πρόβλεψη σπάνια θα συμπίπτει με την πραγματική παραγωγή. Πάντοτε, δηλαδή, η πρόβλεψη θα υπόκειται σε κάποια μορφή λάθους (απόκλισης). Είναι δυνατόν να θέσουμε κάποιο όριο στο λάθος πρόβλεψης; Ποιοί είναι οι πιο σημαντικοί παράγοντες στην πρόβλεψη της παραγωγής;

4. Εκτός από το πρόβλημα της πρόβλεψης, όπως είπαμε, η Στατιστική ασχολείται με αποφάσεις που λαμβάνονται με βάση παρατηρηθέντα στοιχεία. Ας θεωρήσουμε το πρόβλημα του καθορισμού της αποτελεσματικότητας ενός νέου εμβολίου. Ας υποθέσουμε, για παράδειγμα, ότι το νέο αυτό εμβόλιο γρίππης γίνεται σε δέκα ανθρώπους τους οποίους παρατηρούμε κατά τη διάρκεια του χειμώνα. Εστω ότι οκτώ από τους ανθρώπους αυτούς περνούν το χειμώνα χωρίς να περάσουν γρίππη. Μπορούμε να θεωρήσουμε το εμβόλιο αποτελεσματικό;

5. Δύο διαφορετικές τεχνικές διδασκαλίας χρησιμοποιούνται προκειμένου να παρουσιαστεί σε δύο ομάδες φοιτητών παρόμοιας ικανότητας ένα θέμα. Στο τέλος της περιόδου των μαθημάτων κατασκευάζεται ένα μέτρο επίδοσης για κάθε μια από τις δύο ομάδες. Με βάση αυτή την πληροφορία μπορούμε να αναρωτηθούμε: Παρέχουν τα στοιχεία αυτά επαρκείς ενδείξεις ότι η μία μέθοδος είναι περισσότερο αποτελεσματική από την άλλη όσο αφορά την απόδοση των φοιτητών;

6. Ας θεωρήσουμε τον έλεγχο των προιόντων που αγοράζονται από

κάποια κατασκευαστική εταιρεία. Με βάση τον έλεγχο αυτό κάθε παρτίδα παραχθέντων αγαθών θα πρέπει ή να γίνει δεκτή ή να απορριφθεί και να επιστραφεί στον κατασκευαστή. Ο έλεγχος συνήθως γίνεται με την επιλογή ενός δείγματος δέκα στοιχείων από κάθε παρτίδα και την καταγραφή του αριθμού των ελαττωματικών. Η απόφαση για το κατά πόσον θα πρέπει να αποδεχθούμε ή να απορρίψουμε την παρτίδα μπορεί να στηριχθεί στον αριθμό των ελαττωματικών στοιχείων που παρατηρήθηκαν στο δείγμα.

7. Μια εταιρεία που κατασκευάζει σύνθετα ηλεκτρονικά αντικείμενα παράγει κάποια συστήματα που λειτουργούν κανονικά αλλά επίσης και μερικά που για κάποιους λόγους δεν λειτουργούν κανονικά. Τι είναι εκείνο που κάνει κάποιο σύστημα να λειτουργεί κανονικά ή όχι; Προκειμένου να απαντηθεί αυτή η ερώτηση ίσως αποφασίσουμε να κάνουμε κάποιες εσωτερικές μετρήσεις σε ένα σύστημα ώστε να καθορίσουμε σημαντικούς παράγοντες που διαχωρίζουν ένα αποδεκτό από ένα μη αποδεκτό σύστημα. Από ένα δείγμα αποδεκτών και απορριπτέων συστημάτων μπορούν στη συνέχεια να συναχθούν στοιχεία τα οποία θα βοηθήσουν στην κατανόηση του αρχικού σχεδιασμού ή των μεταβλητών παραγωγής που επηρεάζουν την ποιότητα του συστήματος.

1.3 Μορφές και είδη δεδομένων

Ο όρος **δεδομένα (data)** αναφέρεται σε μετρήσεις ή παρατηρήσεις που προέρχονται από ένα πείραμα ή μια δειγματοληπτική έρευνα. Τα δεδομένα μπορεί να είναι ή **ποσοτικά** (αριθμητικά) ή **ποιοτικά** (κατηγορικά). Συνηθως τον όρο *μέτρηση* τον αντιλαμβανόμαστε με την ποσοτική του έννοια σαν να πρόκειται για μετρήσεις σε μεταβλητές όπως ύψος, βάρος, ηλικία, απόσταση κ.λ.π.. Αλλα παραδείγματα ποσοτικών στοιχείων αναφέρονται στους μισθούς, στις τιμές του χρηματιστηρίου, στον αριθμό των βιβλίων που πουλάει ένα βιβλιοπωλείο κ.λ.π.. Τα δεδομένα στα παραδείγματα αυτά είναι

πραγματικοί αριθμοί και αναφέρονται στην τιμή κάποιου ποσοτικού χαρακτηριστικού που μας ενδιαφέρει.

Επομένως, γενικά, μπορούμε να πούμε ότι τα **ποσοτικά δεδομένα** (quantitative data) είναι αριθμητικές παρατηρήσεις.

Από το άλλο μέρος αν σε μια ερώτηση μιας έρευνας μας ενδιαφέρει να καταγράψουμε την οικογενειακή κατάσταση του ερωτώμενου (ελεύθερος, παντρεμένος, χωρισμένος, χήρος) οι απαντήσεις δεν θα είναι αριθμητικές. Παρ'όλα αυτά κάθε μία απάντηση μπορεί και πάλι να τοποθετηθεί σε μια από τέσσερις κατηγορίες. Παρατηρήσεις οι οποίες μπορούν να χωριστούν σε κατηγορίες με βάση ποιοτικά χαρακτηριστικά, όπως π.χ. οικογενειακή κατάσταση, φύλο, είδος απασχόλησης, είδος κατοικίας αποτελούν ποιοτικά δεδομένα. Τα δεδομένα σε τέτοια παραδείγματα είναι απλώς τα ονόματα των δυνατών ταξινομήσεων των παρατηρήσεων.

Επομένως τα **ποιοτικά δεδομένα** (qualitative data) είναι κατηγορικές παρατηρήσεις.

Η μόνη επεξεργασία ποιοτικών δεδομένων που μπορούμε να κάνουμε είναι να μετρήσουμε τον αριθμό των παρατηρήσεων σε κάθε κατηγορία και στη συνέχεια να μετρήσουμε την αναλογία ή το ποσοστό (proportion ή percentage) όλων των παρατηρήσεων που υπάγονται σε κάθε κατηγορία. Αυτό είναι σημαντικό να το κατανοήσουμε έστω και αν χρησιμοποιούνται αριθμοί ως ονόματα για κάθε κατηγορία. Για παράδειγμα, ας αναφερθούμε πάλι στο ερωτηματολόγιο που περιέχει την ερώτηση για την οικογενειακή κατάσταση των ερωτωμένων με πιθανές απαντήσεις "ελεύθερος", "παντρεμένος", "διαζευγμένος", "χήρος". Προκειμένου να απλοποιήσουμε τη διαδικασία καταγραφής των απαντήσεων (ιδιαίτερα όταν η διαδικασία αυτή αφορά την εισαγωγή τους σε υπολογιστή)

μετατρέπουμε συνήθως τα δεδομένα σε αριθμούς. Είναι όμως σημαντικό να αντιληφθούμε ότι παρ'όλα αυτά τα δεδομένα του είδους αυτού εξακολουθούν να είναι ποιοτικά διότι οι αριθμοί εδώ απλώς αντιπροσωπεύουν το "όνομα" της απάντησης. Δεν έχουν αριθμητικό νόημα. Κατά συνέπεια, αριθμητικοί υπολογισμοί με τα δεδομένα αυτά δεν έχουν έννοια.

Για παράδειγμα, ας υποθέσουμε ότι προκειμένου να καταγράψουμε τις απαντήσεις στο ερώτημα για την οικογενειακή κατάσταση χρησιμοποιήσαμε τις εξής κατηγορίες:

Ελεύθερος	1	Παντρεμένος	2
Διαζευγμένος	3	Χήρος	4

Εστω ότι στα πρώτα δέκα ερωτηματολόγια καταγράφηκαν οι απαντήσεις 1,1,3,4,1,1,2,3,1,3. Ο μέσος των αριθμητικών αυτών απαντήσεων είναι 2. Είναι προφανές ότι δεν μπορεί κανείς να θεωρήσει ότι η μέση οικογενειακή κατάσταση των ερωτηθέντων είναι "παντρεμένος". Ας υποθέσουμε ακόμα ότι συγκεντρώθηκαν τέσσερα ακόμα ερωτηματολόγια στα οποία τρείς από τους ερωτηθέντες ήταν χήροι και ένας διαζευγμένος. Στην περίπτωση αυτή ο μέσος γίνεται 2.5. Ούτε αυτό βέβαια σημαίνει ότι η μέση οικογενειακή κατάσταση ήταν "παντρεμένος" αλλά καθ'οδόν προς "διαζευγμένος"! Είναι δηλαδή προφανές ότι τέτοιοι αριθμητικοί υπολογισμοί που οδηγούν σε αριθμητικά αποτελέσματα, όπως ο υπολογισμός του μέσου, για ποιοτικά δεδομένα δίνει απαντήσεις χωρίς νόημα. Το μόνο που έχει έννοια να κάνουμε με ποιοτικά δεδομένα είναι να καταγράψουμε τον αριθμό των φορών που κάθε κατηγορία εμφανίζεται και στη συνέχεια να υπολογίσουμε το ποσοστό των ερωτηθέντων που, στο παράδειγμά μας π.χ., ανήκουν σε κάθε μια από τις οικογενειακές καταστάσεις.

Δεδομένα όπως αυτά του παραδείγματός μας, τα οποία αποτελούν απλώς τα ονόματα των δυνατών ταξινομήσεων των παρατηρήσεων λέμε ότι βρίσκονται σε **ονομαστική κλίμακα (nominal scale).**

Η κατανόηση του είδους των δεδομένων που μετράμε είναι σημαντική γιατί είναι ένας από τους παράγοντες που καθορίζουν ποιά στατιστική τεχνική θα χρησιμοποιήσουμε για την ανάλυσή τους. Συνήθως ο καθορισμός του κατά πόσον τα δεδομένα είναι ποιοτικά ή ποσοτικά είναι αρκετός. Υπάρχουν όμως μερικές περιπτώσεις (κυρίως στη μη Παραμετρική Στατιστική) που είναι σημαντικό να αντιληφθούμε αν μη ποσοτικά δεδομένα μπορούν να διαταχθούν (να έχουν διάταξη). Να διαπιστώσουμε δηλαδή αν η φύση των μετρήσεων ενός συνόλου μη ποσοτικών δεδομένων είναι τέτοια που να επιτρέπει να θεωρήσει κανείς και διάταξη των προκυπτουσών κατηγοριών.

Αν οι κατηγορίες για ένα σύνολο μη ποσοτικών δεδομένων υπαινίσσονται και κάποια διάταξη, τα δεδομένα ονομάζονται **διατεταγμένα δεδομένα (ranked data)**. Στην περίπτωση αυτή λέμε ότι τα δεδομένα βρίσκονται σε **διατεταγμένη κλίμακα** ή **κλίμακα διάταξης (ordinal scale)**.

Ως παράδειγμα για την περίπτωση αυτή μπορούμε να αναφέρουμε την αξιολόγηση ενός ξενοδοχείου ως εξαίρετου, καλού, ικανοποιητικού, ή μέτριου με βάση την ποιότητα διαμονής που προσφέρει. Από τη φύση των απαντήσεων και επειδή τα δεδομένα είναι μη ποσοτικά και κατηγορικά θα έλεγε κανείς ότι ανήκουν σε ονομαστική κλίμακα (nominal scale). Παρατηρούμε όμως, ότι οι απαντήσεις διατάσσονται κατά σειρά προτίμησης με βάση την ποιότητα διαμονής. Ετσι οποιαδήποτε αριθμητική παρουσίαση των τεσσάρων απαντήσεων θα πρέπει να διατηρεί τη διάταξη των απαντήσεων. Τέτοιας μορφής δεδομένα, δηλαδή, βρίσκονται σε διατεταγμένη κλίμακα . Ο μόνος περιορισμός στην επιλογή των αριθμών στην κλίμακα αυτή είναι ότι θα πρέπει να αντιπροσωπεύουν τη διάταξη των απαντήσεων. Οι τιμές που θα χρησιμοποιηθούν καθορίζονται αυθαίρετα. Για παράδειγμα θα μπορούσαμε να χρησιμοποιήσουμε τους εξής αριθμούς για να χαρακτηρίσουμε τις τέσσερις κατηγορίες:

Εξαίρετο	4	Καλό	3

Ικανοποιητικό	2	Μέτριο	1

Δεν θα υπήρχε όμως πρόβλημα αν χρησιμοποιούσαμε την εξής αξιολόγηση:

Εξαίρετο	9	Καλό	5
Ικανοποιητικό	3	Μέτριο	2

Θα μπορούσαμε επίσης να αντιστρέψουμε τη βαθμολογία και να χρησιμοποιήσουμε 1 για εξαίρετο και 4 για μέτριο χωρίς αυτό να επηρεάσει τη στατιστική τεχνική που θα χρησιμοποιήσουμε.

Η μόνη πληροφορία που παρέχουν τα διατεταγμένα δεδομένα και η οποία δεν υπάρχει στα ποιοτικά δεδομένα είναι η διάταξη των απαντήσεων. Ούτε στην περίπτωση των διατεταγμένων δεδομένων μπορούμε να ερμηνεύσουμε τη διαφορά μεταξύ τιμών γιατί οι τιμές που χρησιμοποιούνται είναι αυθαίρετες. Για παράδειγμα, στη διάταξη 4,3,2,1 η διαφορά μεταξύ εξαίρετου και καλού (4-3 = 1) είναι ίση με τη διαφορά μεταξύ ικανοποιητικού και μετρίου (2-1 = 1), ενώ στη διάταξη 9-5-3-2 η διαφορά μεταξύ εξαίρετου και καλού (9-5 = 4) είναι τέσσερις φορές μεγαλύτερη από τη διαφορά μεταξύ ικανοποιητικού και μετρίου (3-2 = 1). Δεδομένου ότι και τα δύο συστήματα αρίθμησης είναι αποδεκτά συστήματα (παρ'ότι είναι αυθαίρετα) δεν μπορεί κανείς να εξαγάγει συμπεράσματα για διαφορές τιμών μιας διατεταγμένης κλίμακας. Επομένως και εδώ περιγραφικά μέτρα που βασίζονται σε αριθμητικούς υπολογισμούς δεν είναι αποδεκτά. Οι μόνες στατιστικές συναρτήσεις που έχουν έννοια και μπορούν να υπολογισθούν με βάση διατεταγμένα δεδομένα είναι περιγραφικά μέτρα που βασίζονται στη διαδικασία διάταξης. Για παράδειγμα, όπως θα δούμε στη συνέχεια, ένα κατάλληλο μέτρο θέσης στην περίπτωση διατεταγμένων δεδομένων είναι η **διάμεσος**. Οπως επίσης θα δούμε αργότερα, οι μόνες στατιστικές τεχνικές που είναι κατάλληλες για ανάλυση διατεταγμένων στοιχείων είναι αυτές που βασίζονται στη διαδικασία διάταξης.

Θα πρέπει να παρατηρήσουμε ότι μπορούμε να αντιμετωπίζουμε ποσοτικά δεδομένα ως ποιοτικά δεδομένα εάν αυτό μας εξυπηρετεί. Για παράδειγμα, έστω ότι μας ενδιαφέρει να μετρήσουμε το βάρος μιας συσκευασίας καφέ που υποτίθεται ότι θα πρέπει να είναι 200gr. (Είναι προφανές ότι εδώ έχουμε ποσοτικά δεδομένα). Ας υποθέσουμε ότι ενδιαφερόμαστε στο πείραμα αυτό να εντοπίσουμε συσκευασίες με βάρος μικρότερο από 200gr οι οποίες θα θεωρηθούν ως μη αποδεκτές. Επομένως το αποτέλεσμα του πειράματός μας στην περίπτωση αυτή είναι ή "αποδεκτό" ή "μη αποδεκτό", διαχωρισμός που μπορεί να θεωρηθεί ότι καθιστά τα δεδομένα ποιοτικά. Βέβαια, ενώ υπάρχουν, εν γένει, περιπτώσεις που εξυπηρετεί τη μελέτη να θεωρηθούν ποσοτικά δεδομένα ως ποιοτικά, είναι προφανές ότι το αντίθετο δεν είναι δυνατό να συμβεί.

Τα ποσοτικά δεδομένα μπορούν να θεωρηθούν ότι χωρίζονται στις εξής κατηγορίες: **δεδομένα σε κλίμακα διαστήματος (interval scale)** και **δεδομένα σε κλίμακα λόγου (ratio scale)**. Και οι δυο κλίμακες είναι διατεταγμένες κλίμακες στις οποίες όμως έχει έννοια η σύγκριση του μεγέθους δυο τιμών είτε με θεώρηση της διαφοράς τους είτε με θεώρηση του λόγου τους.

Ετσι, η **κλίμακα διαστήματος** είναι μια διατεταγμένη κλίμακα στην οποία η διαφορά μεταξύ δύο τιμών έχει έννοια.

Για παράδειγμα, αν ο βαθμός ενός φοιτητή σε ένα συγκεκριμένο μάθημα είναι 8 και ενός άλλου φοιτητή είναι 6, τότε έχει έννοια να μιλήσουμε για διαφορά δύο μονάδων μεταξύ των βαθμών των δύο αυτών φοιτητών. Η τιμή μηδέν για δεδομένα σε κλίμακα διαστήματος δεν αποτελεί ένδειξη πλήρους ανυπαρξίας της ποσότητας αυτής, Για παράδειγμα, το 0 της κλίμακας Κελσίου για τη μέτρηση της θερμοκρασίας, δε σημαίνει ότι δεν υπάρχει θερμοκρασία. Σε μια άλλη κλίμακα η θερμοκρασία αυτή θα είχε διαφορετική τιμή, π.χ. 32 στην κλίμακα Farenheit. Δηλαδή στην κλίμακα διαστήματος δεν ορίζεται το απόλυτο μηδέν.

Η **κλίμακα λόγου** είναι μια διατεταγμένη κλίμακα στην οποία

πέρα από το ότι οι διαφορές τιμών έχουν έννοια, υπάρχει απόλυτο μηδέν έτσι ώστε και οι λόγοι των δύο τιμών να έχουν έννοια.

Δηλαδή, σε αντίθεση με την κλίμακα διαστήματος, η τιμή μηδέν για δεδομένα σε κλίμακα λόγου, αποτελεί ένδειξη πλήρους απουσίας της ποσότητας αυτής (ορίζεται το απόλυτο μηδέν) και ο λόγος δύο τιμών αποτελεί ένδειξη του σχετικού τους μεγέθους. Για παράδειγμα, αν οι εισπράξεις μιας επιχείρησης είναι μηδέν, αυτό έχει την έννοια ότι η επιχείρηση δεν έχει καθόλου έσοδα. Επίσης εισπράξεις 100 χιλ. δρχ. είναι διπλάσιες από εισπράξεις 50 χιλ. δρχ.. Τέτοια στοιχεία μπορούν να θεωρηθούν ως στοιχεία κλίμακας λόγου.

Είναι προφανές ότι δεδομένα που ενώ είναι δυνατό να θεωρηθούν ότι ανήκουν σε κλίμακα διαστήματος, δεν μπορούν να θεωρηθούν ότι ανήκουν σε κλίμακα λόγου. Για παράδειγμα, οι μετρήσεις της καθημερινής θερμοκρασίας σε βαθμούς Κελσίου αποτελούν δεδομένα σε κλίμακα διαστήματος αλλά όχι σε κλίμακα λόγου. Οπως τονίσθηκε προηγουμένως, η τιμή μηδέν δεν αποτελεί ένδειξη έλλειψης θερμοκρασίας. Με αυτή την έννοια δεν είναι δυνατόν να πούμε ότι αν σε μια συγκεκριμένη μέρα η θερμοκρασία είναι 30°C, η μέρα αυτή είναι δυο φορές πιο ζεστή από μία μέρα που είχε θερμοκρασία 15°C. Στις περισσότερες όμως περιπτώσεις δεδομένων που αντιμετωπίζουμε δεν υπάρχει διάκριση μεταξύ δεδομένων σε κλίμακα διαστήματος και δεδομένων σε κλίμακα λόγου.

Συνοπτικά μπορούμε να πούμε τα εξής:

Οι τέσσερις χρησιμοποιούμενες κλίμακες μέτρησης είναι - από την "ασθενέστερη" στην "ισχυρότερη" - η ονομαστική κλίμακα (nominal scale), η διατεταγμένη κλίμακα (ordinal scale), η κλίμακα διαστήματος (interval scale) και η κλίμακα λόγου (ratio scale).

- Η **ονομαστική κλίμακα** κάνει χρήση αριθμητικών τιμών μόνο ως μέσων διαχωρισμού των ιδιοτήτων ή των στοιχείων σε διάφορες κατηγορίες. Οι τιμές δηλαδή είναι τα αυθαίρετα ονόματα των δυνατών κατηγοριών.

- Η **κλίμακα διάταξης** είναι μια ονομαστική κλίμακα στην οποία συγκρίσεις της μορφής "μεγαλύτερη", "μικρότερη" ή "ίση" μεταξύ των τιμών έχουν νόημα.

- Η **κλίμακα διαστήματος** είναι μια διατεταγμένη κλίμακα στην οποία, πέρα από τη σχετική διάταξη των μετρήσεων, έχει έννοια και το μέγεθος του διαστήματος των δύο μετρήσεων (δηλαδή το μέγεθος της διαφοράς με την έννοια της αφαίρεσης των σχετικών τιμών). Το μηδέν ορίζεται αυθαίρετα στην κλίμακα αυτή, όπως και η μονάδα (μοναδιαία απόσταση).

- Η **κλίμακα λόγου** είναι μια κλίμακα διαστήματος στην οποία, πέρα από την διάταξη και το μέγεθος του διαστήματος μεταξύ δύο τιμών, έχει έννοια και ο λόγος δύο τιμών. Δηλαδή στην κλίμακα αυτή έχει έννοια η σύγκριση δύο τιμών μέσω του λόγου τους. Το μηδέν ορίζεται στην κλίμακα αυτή, δηλαδή υπάρχει φυσική μέτρηση που ονομάζεται μηδέν. Η μονάδα ορίζεται αυθαίρετα.

Δεν είναι δυνατό να καθορίσουμε ποιά είναι η κατάλληλη κλίμακα μόνο με τη γνώση των μετρήσεων. Ο καθορισμός της κατάλληλης κλίμακας γίνεται δυνατός μόνο μετά από προσεκτική εξέταση των μετρούμενων ποσοτήτων και της μεθόδου μέτρησης. Η μελέτη αυτή θα οδηγήσει στην ερμηνεία που μπορεί να δοθεί στη μέτρηση.

Ο πίνακας που ακολουθεί αποτελεί μια συνοπτική παρουσίαση των επιτρεπτών χειρισμών των δεδομένων ανάλογα με το είδος τους και την κλίμακα στην οποία ανήκουν.

Πίνακας 1.3.1

Επιτρεπτές συγκρίσεις και υπολογισμοί δεδομένων στις διάφορες κλίμακες μέτρησης

Κλίμακα

	Ονομαστική	**Διατεταγμένη**
Ποιοτικά δεδομένα	- Τιμές είναι αυθαίρετα ονόματα δυνατών κατηγοριών - Αποδεκτοί υπολογισμοί: ποσοστά και αναλογίες σε κάθε κατηγορία	- Τιμές είναι ονόματα δυνατών κατηγοριών που ταυτόχρονα αντιπροσωπεύουν την διάταξή τους - Αποδεκτοί υπολογισμοί: μόνο αυτοί που βασίζονται σε διαδικασία διάταξης

Κλίμακα

	Διαστήματος	**Λόγου**
Ποσοτικά δεδομένα	- Διάταξη - Εχουν έννοια οι διαφορές των τιμών	- Διάταξη - Εχουν έννοια οι διαφορές των τιμών - Εχουν έννοια οι λόγοι των τιμών

1.4 Οι βασικές έννοιες της Εφαρμοσμένης Στατιστικής

Παράδειγμα: Ο διευθυντής μιας πολυεθνικής ασφαλιστικής εταιρείας ενδιαφέρεται να κατανοήσει καλύτερα τον τρόπο με τον οποίο οι πελάτες του αποφασίζουν για τις ασφάλειες που κάνουν. Ιδιαίτερα ενδιαφέρεται να μελετήσει αν το φύλο και το εισόδημα είναι παράγοντες που επηρεάζουν το ποσό για το οποίο ασφαλίζεται ένας πελάτης και το είδος της ασφάλειας που κάνει. (Η συγκεκριμένη εταιρεία ενδαφέρεται να μελετήσει δύο είδη ασφαλειών. Τη γενική ασφάλεια (universal policy) η οποία δεν αποτελεί μόνο ασφάλεια αλλά και επένδυση. Ο πελάτης ασφαλίζεται για όλη του τη ζωή και έχει το δικαίωμα να ακυρώσει την ασφάλεια και να εισπράξει τα χρήματα σε μετρητά σε οποιαδήποτε χρονική στιγμή. Αντίθετα η ασφάλεια περιορισμένου χρόνου (term policy) καλύπτει μία καθορισμένη χρονική περίοδο. Οταν αγοραστεί η ασφάλεια ο πελάτης δεν έχει το δικαίωμα να ζητήσει επιστροφή των ασφαλίστρων (premium)). Ο διευθυντής της ασφαλιστικής εταιρείας πιστεύει ότι γνώση μιας τέτοιας σχέσης που τυχόν υφίσταται μπορεί να βοηθήσει την εταιρεία ώστε να διαμορφώσει καλύτερα τις προτάσεις που κάνει σε πιθανούς μελλοντικούς πελάτες. Για το σκοπό αυτό καταγράφηκαν δεδομένα από τους φακέλους 52 πελατών της εταιρείας οι οποίοι επελέγησαν με τυχαίο τρόπο. Για κάθε πελάτη καταγράφηκε (1) το είδος της ασφάλειας (1 = ασφάλεια περιορισμένου χρόνου, 0 = γενική ασφάλεια), (2) το ποσό ασφάλισης, (3) το φύλο του πελάτη (1 = γυναίκα , 0 = άνδρας) και (4) το ετήσιο εισόδημα του πελάτη (τα ποσά σε εκατοντάδες χιλιάδες δραχμές). Τα στοιχεία αυτά δίνονται στον πίνακα που ακολουθεί.

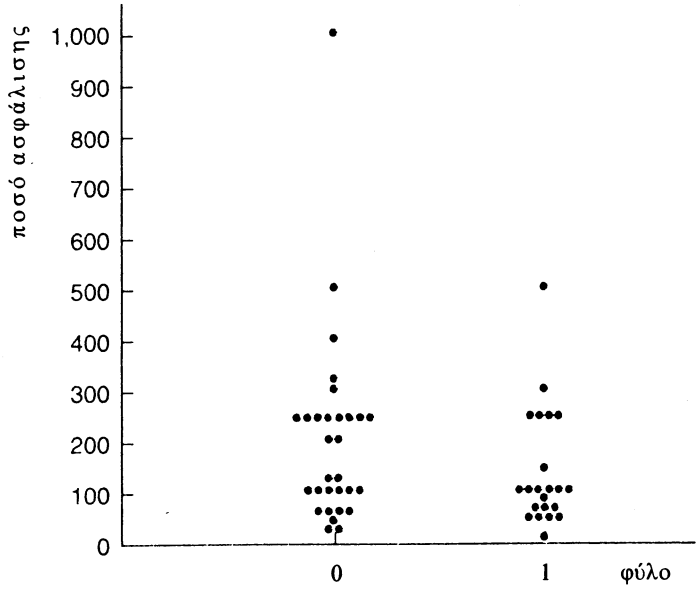
Πίνακας 1.4.1

Πελάτης	Φύλο	Ποσό ασφάλισης (εκ. χιλ. δρχ.)	Είδος ασφάλειας	Ετήσιο εισόδημα (εκ. χιλ. δρχ.)
1	1	75	1	46.0
2	0	250	1	52.0
3	0	250	1	42.5
4	1	100	1	31.0
5	1	100	0	40.5
6	1	50	0	20.0
7	0	100	1	27.5
8	0	25	0	30.0
9	1	50	1	21.0
10	1	80	0	18.0
11	0	250	1	43.5
12	1	50	1	17.0
13	0	50	1	19.0
14	1	250	1	30.0
15	0	500	1	85.0
16	0	200	0	62.0
17	0	250	1	26.0
18	1	250	1	29.0
19	0	40	0	26.0
20	1	15	0	17.0
21	0	25	1	44.5
22	0	250	1	36.0
23	0	50	1	21.0
24	0	50	1	29.0
25	1	50	1	23.0

Πίνακας 1.4.1 (συνέχεια)

Πελάτης	Φύλο	Ποσό ασφάλισης (εκ. χιλ. δρχ.)	Είδος ασφάλειας	Ετήσιο εισόδημα (εκ. χιλ. δρχ.)
26	0	1000	1	126.0
27	0	400	1	78.5
28	1	100	0	92.7
29	1	150	0	33.5
30	0	100	1	22.0
31	1	250	1	48.8
32	0	300	1	29.0
33	0	50	1	18.0
34	1	100	0	34.4
35	1	75	0	22.8
36	0	250	0	31.4
37	1	500	1	41.7
38	0	100	1	20.8
39	1	250	0	54.7
40	0	125	0	27.3
41	0	250	1	40.2
42	1	100	1	27.0
43	0	320	0	76.5
44	1	300	0	57.0
45	0	200	1	42.1
46	0	100	1	37.5
47	0	100	1	26.0
48	1	100	1	23.0
49	0	100	1	29.5
50	0	125	1	31.0
51	0	250	1	30.5

Ας ρίξουμε μια ματιά στα δεδομένα. Το πρώτο πράγμα που παρατηρούμε είναι η μεταβλητότητα. Το ποσό ασφάλισης κυμαίνεται όπως επίσης και το ετήσιο εισόδημα. Αν προσέξουμε περισσότερο, βλέπουμε ότι οι πελάτες με υψηλότερο ετήσιο εισόδημα τείνουν να ασφαλίζονται για μεγαλύτερα ποσά. Στο σχήμα 1.4.1 απεικονίζονται τα ποσά ασφάλισης χωριστά για άνδρες και γυναίκες πελάτες.



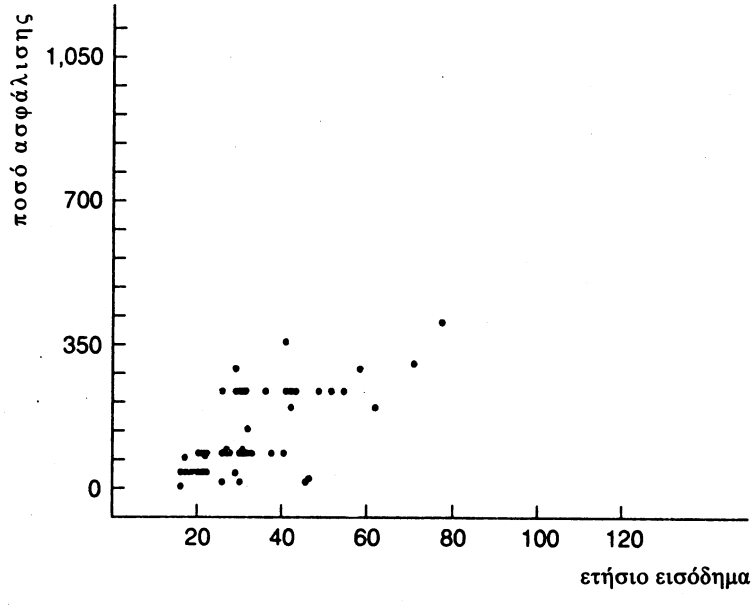
Σχήμα 1.4.1

Ποσά ασφάλισης για άνδρες και γυναίκες πελάτες

Οπως παρατηρούμε στο σχήμα υπάρχει σημαντική διακύμανση μεταξύ των ποσών ασφάλισης των ανδρών όπως επίσης και μεταξύ των αντιστοίχων ποσών ασφάλισης των γυναικών. (Αυτό φαίνεται από το κατακόρυφο άπλωμα των δύο συνόλων των τιμών των δεδομένων). Παρατηρούμε επίσης ότι τα ποσά ασφάλισης για τους άνδρες τείνουν να είναι, κατά κάποιο τρόπο, μεγαλύτερα από τα αντίστοιχα για τις γυναίκες παρ'ότι υπάρχει σημαντική επικάλυψη στο εύρος των τιμών των δύο ομάδων.

Το σχήμα 1.4.2 δίνει ένα διάγραμμα για τα ποσά ασφάλισης

(στον κατακόρυφο άξονα) σε σχέση με το ετήσιο εισόδημα (στον οριζόντιο άξονα).



Σχήμα 1.4.2

Είναι φανερό ότι πελάτες με υψηλότερο ετήσιο εισόδημα τείνουν να αγοράζουν ακριβότερες ασφάλειες. Παρ'όλα αυτά μπορούμε να παρατηρήσουμε ότι υπάρχει μία σημαντική επικάλυψη μεταξύ του εύρους των τιμών για πελάτες με χαμηλότερα εισοδήματα και του εύρους των τιμών για πελάτες με μεσαία εισοδήματα. (Στο σχήμα 1.4.2 υπάρχουν λιγότερα σημεία από ότι δεδομένα γιατί οι ίδιες τιμές εκφράζονται με το ίδιο σημείο στη γραφική παράσταση).

Τα σχήματα 1.4.1 και 1.4.2 αποτελούν ένδειξη ότι το ποσό ασφάλισης των πελατών σχετίζεται με το εισόδημά τους και σε μικρότερη έκταση με το φύλο. Τα στοιχεία αυτά θα έχουμε την ευκαιρία να τα αναλύσουμε περισσότερο αργότερα. Στο σημείο αυτό όμως μπορούμε να κάνουμε μερικές παρατηρήσεις.

1. 80% των ανδρών στο δείγμα έχει επιλέξει ασφάλεια περιορισμένου χρόνου ενώ μόνο 52% των γυναικών έχει κάνει το ίδιο. Αυτό ίσως αποτελεί ένδειξη ότι το είδος της ασφάλειας που επιλέγει ο πελάτης εξαρτάται κατά κάποιο τρόπο από το φύλο.

2. Το μέσο ποσό ασφάλισης για τους ανδρες είναι 20200000 δρχ. ενώ το αντίστοιχο για τις γυναίκες είναι 14260000 δρχ. Αυτό συμβαδίζει με τα προηγούμενα συμπεράσματα που βασίστηκαν στο σχήμα 1.4.1.

3. Κατά μέσο όρο το ποσό ασφάλισης τείνει να αυξάνεται κατά 710000 δρχ. για κάθε πρόσθετες 100000 δρχ ετησίου εισοδήματος. (Η διαπίστωση αυτή επιβεβαιώνεται με τη χρησιμοποίηση της ανάλυσης παλινδρόμησης που θα δούμε αργότερα).

Στο υπόλοιπο της ενότητας αυτής θα αναφερθούμε στα βασικά στοιχεία οποιασδήποτε στατιστικής μελέτης.

Τις περισσότερες φορές τα δεδομένα μπορούν να θεωρηθούν ως το αποτέλεσμα μιας διαδικασίας. Εν γένει ως **διαδικασία ή διεργασία (process)** μπορούμε να ορίσουμε *ένα σύνολο συνθηκών που χρησιμοποιούνται κατ'επανάληψη για να μετατρέψουν εισρέοντα στοιχεία (inputs) σε αποτελέσματα (outcomes).* Στο παράδειγμα της ασφαλιστικής εταρείας ο υπεύθυνος μελέτησε ένα δείγμα φακέλων από 51 πελάτες που επελέγη με τυχαίο τρόπο από το σύνολο των φακέλων όλων των τρεχόντων πελατών του. Ο φάκελος παριστάνει τα αποτέλεσματα (outcomes) πολλών διαδικασιών, κυρίως της διαφημιστικής διαδικασίας της εταιρείας, που οδηγεί σε πωλήσεις, της διαδικασίας εξυπηρέτησης των πελατών (που βοηθά στο να διατηρηθούν οι πελάτες και να δημιουργηθούν νέοι πελάτες από τις συστάσεις των ήδη πελατών) και της διοικητικής διαδικασίας (administrative process) (με την οποία γίνονται, διατηρούνται και

ενημερώνονται εγγραφές στοιχείων πελατών στους φακέλους. Κάθε μια από τις διαδικασίες αυτές περιλαμβάνει έναν αριθμό περιορισμών που δρούν από κοινού για να δημιουργήσουν αποτελέσματα.

Οι στατιστικές μελέτες αναφέρονται σε δεδομένα που προκύπτουν από κάποια ομάδα που ενδιαφέρει και, συνήθως, αποτελείται από ανθρώπους , αντικείμενα ή φυσικές μετρήσεις για κάποιο σημαντικό μέγεθος.

Για να δηλωθεί ένα σύνολο τέτοιων στοιχείων για τα οποία μας ενδιαφέρει να έχουμε μια καλύτερη αντίληψη χρησιμοποιούμε τον όρο **πληθυσμός (population)**. Πολλές φορές χρησιμοποιείται και ο όρος **ολότητα (universe)**.

Σε μερικές στατιστικές μελέτες ενδιαφερόμαστε να μελετήσουμε μια ομάδα, όπως αυτή είναι σε κάποια συγκεκριμένη χρονική στιγμή. Σε τέτοιες μελέτες ο πληθυσμός είναι ολόκληρη η ομάδα των στοιχείων τη συγκεκριμένη αυτή στιγμή. Για παράδειγμα, η απόφαση για το αν θα πρέπει να δοθεί ένα επίδομα στους εργαζόμενους μιας επιχείρησης ίσως εξαρτάται από τις τρέχουσες διαθέσεις των εργαζομένων όπως αυτές εκφράζονται σε μία καταγραφή. Στην περίπτωση αυτή το σύνολο των στοιχείων για τα οποία χρειάζεται γνώση αποτελείται από τις τρέχουσες γνώμες των εργαζομένων στη συγκεκριμένη χρονική στιγμή.

Σε άλλες περιπτώσεις αυτό που επιθυμούμε να αντιληφθούμε σχετίζεται με την δυνατότητα της διαδικασίας που παράγει αποτελέσματα (outcomes) από την οποία μπορούμε να μετρήσουμε την ποσότητα που μας ενδιαφέρει. Σε μια τέτοια περίπτωση ένα σύνολο ενδεχομένων μπορεί να θεωρηθεί σαν ένα δείγμα των πιθανών ενδεχομένων που θα μπορούσαν να παραχθούν από τη συγκεκριμένη διαδικασία εάν αυτή η διαδικασία συνεχιζόταν χωρίς μεταβολή. Για παράδειγμα, ας υποθέσουμε ότι μελετάμε το χρόνο αναμονής ασθενών σε ένα νοσοκομείο κατά τους μήνες Ιανουάριο και Φεβρουάριο. Οι παρατηρήσεις αυτές αντιπροσωπεύουν ένα δείγμα των δυνατοτήτων της διαδικασίας λειτουργίας του γραφείου καταγραφής ασθενών κατά την

περίοδο μελέτης. Θα μας ενδιέφερε να μάθουμε για το χρόνο αναμονής όλων των ασθενών με την προυπόθεση ότι οι συνθήκες της διαδικασίας που επικρατούν τον Ιανουάριο και Φεβρουάριο συνεχίζουν να υφίστανται χωρίς μεταβολή.

Η γνώση του αντικειμένου είναι εκείνη που καθορίζει τι ακριβώς θέλουμε να μάθουμε γύρω από τον πληθυσμό ή τη διαδικασία που μελετάμε. Το ερώτημα αυτό δεν είναι ένα στατιστικό ερώτημα παρ'ότι η στατιστική σκέψη επηρεάζει την επιλογή αυτή. Αυτό οδηγεί στον ορισμό στατιστικών μεταβλητών. Οι μετρήσεις που θα γίνουν στη συνέχεια πάνω σε κάθε μια από τις μεταβλητές απαιτούν έναν **λειτουργικό ορισμό (operational definition)**, ένα ορισμό ο οποίος είναι σαφής και ακριβής ώστε να επιτρέπει συνεπείς και αξιόπιστες μετρήσεις από όλους όσους εμπλέκονται στη διαδικασία μέτρησης. Ετσι η **στατιστική μεταβλητή** είναι ένας κατάλληλος (λειτουργικός) ορισμός του χαρακτηριστικού που μας ενδιαφέρει να μελετήσουμε.

Εστω ότι μας ενδιαφέρει να παρατηρήσουμε κάποια στατιστική μεταβλητή για ένα δείγμα από έναν πληθυσμό. Ενά "καλό" δείγμα θα πρέπει να αντικατοπτρίζει τα ουσιαστικά χαρακτηριστικά του πληθυσμού από τον οποίο επιλέγεται. Ενα σημαντικό βήμα για την επιλογή του δείγματος είναι η απαρίθμηση των στοιχείων του πληθυσμού από τον οποίο θα επιλεγεί το δείγμα.

Το **πλαίσιο (frame)** είναι το σύνολο των μελών του πληθυσμού τα οποία έχουν στην πραγματικότητα τη δυνατότητα να περιληφθούν στο δείγμα (η πηγή του δείγματος). Μερικές φορές οι μονάδες του πλαισίου δεν είναι τα μέλη του υπό μελέτη πληθυσμού αλλά μιας ευρύτερης υποδιαίρεσής του όπου κάθε μονάδα περιέχει ένα σύνολο μελών (π.χ. οικογένειες).

Προκειμένου να επιτευχθεί ικανοποιητική εκπροσώπηση είναι σημαντικό να επιλεγούν τα στοιχεία του δείγματος από ένα καλά ορισμένο και σχεδιασμένο πλαίσιο. Επομένως είναι σημαντικό να καθορίσουμε ένα πλαίσιο που να προσεγγίζει τον πραγματικό πληθυσμό όσο το δυνατόν καλύτερα (να "περιγράφει" τον πληθυσμό όσο το

δυνατόν πιστότερα). Με την έννοια αυτή ένα πλαίσιο μπορεί να θεωρηθεί ως μία **λειτουργική έκφραση (operational version)** της έννοιας ενός πληθυσμού. Για το λόγο αυτό πολλές φορές το πλαίσιο ονομάζεται **υπό δειγματοληψία ή δειγματοληπτικός πληθυσμός (sampled population)**. Η οποιαδήποτε στατιστική συμπερασματολογία μπορεί να αναφέρεται απευθείας μόνο στο συγκεκριμένο πλαίσιο. Η εφαρμογή της οποιασδήποτε συμπερασματολογίας στον πληθυσμό αυτόν καθεαυτό εξαρτάται από την εγγύτητα του πλαισίου στον πληθυσμό (από την πιστότητα του πλαισίου). Μία επιτυχημένη ως προς την πιστότητα επιλογή πλαισίου είναι σημαντική δεδομένου ότι οι οποιεσδήποτε αποφάσεις θα εφαρμοσθούν στον πληθυσμό και όχι στο πλαίσιο. Θα πρέπει βέβαια να τονισθεί ότι είναι ελάχιστες φορές δυνατό να κατασκευάσουμε ένα τέλειο πλαίσιο. Εκείνο που μπορούμε να προσπαθήσουμε να κάνουμε είναι το πλαίσιο αυτό να αποτελεί μία όσο το δυνατόν πιστότερη εικόνα (ένα όσο το δυνατόν πιστότερο αντίγραφο) του πληθυσμού. Στο παράδειγμα με την ασφαλιστική εταιρεία, το πλαίσιο είναι το σύνολο των φακέλων των πελατών που ήταν διαθέσιμοι για μελέτη στο διευθυντή της εταιρείας. Οι φάκελοι αυτοί, προφανώς, δεν περιλαμβάνουν νέες ασφάλειες που δεν έχουν ακόμα αρχειοθετηθεί, αντίθετα είναι ενδεχόμενο να περιλαμβάνουν ασφάλειες που έχουν σταματήσει αλλά οι φάκελοί τους δεν έχουν ακόμα ενημερωθεί;

Εστω τώρα ότι ενδιαφερόμαστε να μελετήσουμε την ικανότητα μιας διαδικασίας και όχι ένα πληθυσμό. Στην περίπτωση αυτή το πλαίσιο αποτελείται από όλα τα δυνατά αποτελέσματα της διαδικασίας στην υπό μελέτη χρονική περίοδο. Στο παράδειγμα με τους χρόνους αναμονής στο νοσοκομείο, το πλαίσιο αποτελείται από τους χρόνους αναμονής όλων των ασθενών που επισκέφθηκαν το γραφείο αυτό τον Ιανουάριο και το Φεβρουάριο. Εκείνο που μας ενδιαφέρει να καταλάβουμε είναι η δυνατότητα αυτής καθεαυτής της διαδικασίας. Επομένως οποιεσδήποτε αποφάσεις θα αναφέρονται στις διαδικασίες καταγραφής του συγκεκριμένου γραφείου και όχι στους ασθενείς που

συνέβη να εξετασθούν και να υποβληθούν σε κάποια θεραπεία τον Ιανουάριο και το Φεβρουάριο.

Σε όλες σχεδόν τις περιπτώσεις δεν είναι δυνατόν να αναμένουμε ότι θα πάρουμε πληροφορίες για κάθε μέλος ενός πλαισίου. (Εάν κάτι τέτοιο συμβεί τότε έχουμε **απογραφή (census)**. Αυτό συμβαίνει γιατί μια τέτοια διαδικασία είναι εξαιρετικά πολυδάπανη και απαιτεί πολύ χρόνο. Συνήθως στηριζόμαστε σε ένα δείγμα.

Ενα **δείγμα (sample)** είναι ένα υποσύνολο του πληθυσμού.

Για μελέτες που αναφέρονται σε διαδικασίες, ένα δείγμα είναι ένα υποσύνολο των πιθανών ενδεχομένων της διαδικασίας σε ένα συγκεκριμένο χρονικό διάστημα. Τα στοιχεία ενός δείγματος ονομάζονται **δειγματικές ή δειγματοληπτικές μονάδες (sampling units)**. Σε μερικές περιπτώσεις ονομάζονται **πειραματικές μονάδες (experimental units)**. Στη συνέχεια του βιβλίου θα θεωρήσουμε τις δύο αυτές έννοιες ως ισοδύναμες.

Είναι γεγονός ότι πολλοί εκφράζουν αμφιβολίες και επιφυλάξεις για το κατά πόσο τα συμπεράσματα που βασίζονται σε ένα δείγμα μπορούν να επεκταθούν σε ολόκληρο τον πληθυσμό. Ο λόγος που καταφεύγουμε σε δείγματα είναι ότι σε πολλές περιπτώσεις δεν υπάρχει εναλλακτική λύση αν θέλουμε αντικειμενικές πληροφορίες. Ενας άλλος λόγος είναι ότι η δειγματοληψία μας επιτρέπει να ελέγξουμε συστηματικότερα τη διαδικασία συλλογής των δεδομένων. Εάν δεν υπάρχει επαρκής έλεγχος και επίβλεψη είναι ενδεχόμενο να καταλήξουμε με πολύ μεγαλύτερο λάθος αν καταφύγουμε σε πλήρη απογραφή από ότι αν χρησιμοποιήσουμε μια διαδικασία δειγματοληψίας που είναι ευκολότερο να εποπτευθεί με σχολαστική προσοχή. Το γεγονός βέβαια ότι βασιζόμαστε σε ένα δείγμα έχει κάποιο κόστος. Σε οποιαδήποτε περίπτωση, επειδή ακριβώς το δείγμα είναι μόνο ένα υποσύνολο του πληθυσμού, κανένα δείγμα δεν μπορεί να αντικατοπτρίζει τέλεια το υπό μελέτη πλαίσιο.

Ας υποθέσουμε ότι μπορούμε να πραγματοποιήσουμε μία συνολική

απογραφή με την ίδια ακριβώς μέθοδο που θα χρησιμοποιούσαμε στην περίπτωση δειγματοληψίας - το ίδιο ερωτηματολόγιο, τα ίδια μέσα, την ίδια εκπαίδευση κ.λ.π. **Η απογραφή ονομάζεται και 100% δείγμα.** Η διαφορά των αποτελεσμάτων μιας δειγματοληψίας και των αποτελεσμάτων ενός 100% δείγματος (μιας απογραφής) ονομάζεται **δειγματικό ή δειγματοληπτικό σφάλμα (sampling error).** Το δειγματοληπτικό σφάλμα είναι ένα αναπόσπαστο στοιχείο της Στατιστικής Συμπερασματολογίας. Αυτό είναι το λάθος που συνήθως αναφέρεται σε πολιτικές δημοσκοπήσεις όταν μελετάται το ποσοστό εκείνων που υποστηρίζουν κάποιο συγκεκριμένο κόμμα. Η μελέτη αυτή συμπληρώνεται με την πρόταση "το περιθώριο λάθους είναι συν ή πλην 3 ποσοστιαίες μονάδες". Αυτό σημαίνει ότι το εκτιμώμενο ποσοστό εκείνων που υποστηρίζουν ένα κόμμα μπορεί να απέχει 3 ποσοστιαίες μονάδες από το αποτέλεσμα που θα παίρναμε από ένα 100% δείγμα. Αν, για παράδειγμα, το ποσοστό που καταγράφεται στη σφυγμομέτρηση ότι προτιμά ένα συγκεκριμένο κόμμα είναι 42%, το συγκεκριμένο περιθώριο λάθους οδηγεί στο συμπέρασμα ότι το πραγματικό ποσοστό σε ένα 100% δείγμα μπορεί να είναι μέχρι 39% στο χαμηλότερο επίπεδο και μέχρι 45% στο υψηλότερο επίπεδο εξ αιτίας του δειγματοληπτικού σφάλματος. Αυτό το είδος του σφάλματος δεν είναι δυνατό να το αποφύγει κανείς και επομένως είναι σημαντικό να εκτιμάται το μέγεθος του δειγματοληπτικού σφάλματος που σχετίζεται με οποιαδήποτε στατιστική μελέτη. Πολλές στατιστικές μέθοδοι περιλαμβάνουν τεχνικές που εκτιμούν την ποσότητα του δειγματοληπτικού σφάλματος με την προυπόθεση ότι η επιλογή του δείγματος έγινε με πρέποντα τρόπο.

Δύο άλλες βασικές έννοιες της Στατιστικής είναι η **παράμετρος (parameter)** και η **στατιστική συνάρτηση (statistic).** Η παράμετρος εκφράζει ένα συγκεκριμένο χαρακτηριστικό της συνάρτησης κατανομής του πληθυσμού (είναι δηλαδή μια παράμετρος της συνάρτησης αυτής) ή μιας διαδικασίας. Στατιστική συνάρτηση είναι μια συνάρτηση των παρατηρήσεων του δείγματος.

Ας υποθέσουμε ότι θα ήταν εφικτό να παρατηρήσουμε ένα 100% δείγμα. Στην περίπτωση αυτή, αν και θα είχαμε στη διάθεσή μας όλες τις πληροφορίες για κάποιο χαρακτηριστικό του πλαισίου, δεν θα είμαστε σε θέση να τις επεξεργασθούμε με αποτελεσματικότητα εκτός εάν τις συνοψίζαμε και ενδεχομένως προσδίδαμε στις συνοπτικές τιμές φυσική ερμηνεία. Το ίδιο θα χρειαζόταν να κάνουμε με τις πληροφοριές που θα είχαμε σε ένα δείγμα. Με άλλα λόγια και στις δύο περιπτώσεις χρειάζεται να συνοψίσουμε τις πληροφορίες για ένα χαρακτηριστικό σε μια αριθμητική ποσότητα. Η ποσότητα αυτή ονομάζεται **παράμετρος** στην περίπτωση του πληθυσμού. Στην περίπτωση που έχουμε ένα δείγμα, η ποσότητα που συνοψίζει το χαρακτηριστικό που μας ενδιαφέρει προκύπτει ως τιμή μιας συνάρτησης των παρατηρήσεων του δείγματος. Η συνάρτηση αυτή ονομάζεται **στατιστική συνάρτηση**. Ετσι, για παράδειγμα, ο δειγματικός μέσος είναι μια στατιστική συνάρτηση. Οι στατιστικές συναρτήσεις, συνήθως, χρησιμοποιούνται προκειμένου να γίνει κάποια συμπερασματολογία για τις παραμέτρους του πληθυσμού ή της διαδικασίας που μας ενδιαφέρει.

Γενικότερα, η χρησιμοποίηση πληροφοριών από ένα δείγμα προκειμένου να εξάχθούν συμπεράσματα για τον πληθυσμού από τον οποίο προέρχεται ονομάζεται **Στατιστική Συμπερασματολογία (Statistical Inference)**. Χαρακτηριστικά παραδείγματα Στατιστικής Συμπερασματολογίας είναι η **Εκτιμητική (Estimation)** και ο **Ελεγχος Υποθέσεων (Hypotheses Testing)**. Στην Εκτιμητική, συνήθως, χρησιμοποιούμε μια στατιστική συνάρτηση για να εκτιμήσουμε την τιμή μιας παραμέτρου του υπό μελέτη πληθυσμού ή διαδικασίας. Στον Ελεγχο Υποθέσεων εξετάζουμε το κατά πόσο μια υπόθεση που κάνουμε για μια παράμετρο είναι ισχυρή ή όχι.

Τέλος, η **εμπιστοσύνη (confidence)** που αναφέρεται στη Στατιστική Συμπερασματολογία είναι η πιθανοφάνεια της αλήθειας της συμπερασματολογίας αυτής.

Επομένως, συνοπτικά, μπορούμε να πούμε τα εξής για τα βασικά στοιχεία της Στατιστικής Ανάλυσης.

Τα βασικά στοιχεία της Στατιστικής Ανάλυσης

Πληθυσμός (population) είναι ένα σύνολο στοιχείων που μας ενδιαφέρει να μελετήσουμε. Πολλές φορές χρησιμοποιείται ο όρος **ολότητα (universe)**.

Διαδικασία (process) είναι ένα σύνολο περιορισμών που εμφανίζονται κατ'επανάληψη ώστε να μετατρέψουν πληροφορίες σε αποτελέσματα.

Πλαίσιο (frame) είναι το σύνολο των στοιχείων του πληθυσμού, ή των δυνατών αποτελεσμάτων μιας διαδικασίας, που είναι δυνατό να περιληφθούν στο δείγμα.

Στατιστική μεταβλητή (Statistical variable) είναι μια καλά ορισμένη μετρήσιμη έκφραση ενός χαρακτηριστικού που μας ενδιαφέρει.

Δείγμα (sample) είναι ένα υποσύνολο ενός πληθυσμού ή παρατηρηθέντων αποτελεσμάτων μιας διαδικασίας για μια χρονική περίοδο.

Παράμετρος (parameter) είναι μία αριθμητική ποσότητα που συνοψίζει κάποιο χαρακτηριστικό του πληθυσμού ή της ικανότητας μιας διαδικασίας.

Στατιστική συνάρτηση (statistic) είναι μια συνάρτηση των στοιχείων του δείγματος.

Στατιστική Συμπερασματολογία (Statistical Inference) είναι η διαδικασία χρησιμοποίησης πληροφοριών από το δείγμα με σκοπό την εξαγωγή συμπερασμάτων για τον πληθυσμό ή την ικανότητα μιας διαδικασίας.

Εμπιστοσύνη (confidence) είναι η πιθανοφάνεια ότι η Στατιστική Συμπερασματολογία στην οποία καταλήξαμε είναι σωστή ή ότι έχει κάποιο λάθος, το οποίο όμως δεν υπερβαίνει κάποια προκαθορισμένη ποσότητα.

Παράδειγμα: Μια εταιρεία που παράγει μπύρα ενδαφέρεται να διερευνήσει αν οι καταναλωτές προτιμούν μια καινούργια γεύση μπύρας από αυτήν που ήδη κυκλοφορεί στην αγορά. Προκειμένου να εξεταστεί αυτό δίνεται σε ένα τυχαίο δείγμα 125 καταναλωτών ένα δείγμα και από τις δύο γεύσεις. Οι καταναλωτές ερωτώνται ποιά γεύση προτιμούν. 70 από τα 125 άτομα που ρωτήθηκαν δήλωσαν ότι προτιμούν την νέα γεύση, ενώ 55 δήλωσαν ότι προτιμούν αυτή που ήδη κυκλοφορεί. Να καθορίσετε:

α) τον πληθυσμό ή τη διαδικασία (όποια από τις δύο έννοιες είναι κατάλληλη εδώ),
β) τη στατιστική μεταβλητή,
γ) τις παραμέτρους που μας ενδιαφέρουν,
δ) το δείγμα,
ε) την σχετική στατιστική συνάρτηση.

Λύση :

α) Ο πληθυσμός που μας ενδιαφέρει είναι οι προτιμήσεις σε γεύση των καταναλωτών της συγκεκριμένης περιοχής από την οποία επελέγη το δείγμα κατά τη συγκεκριμένη περίοδο που έγινε η μελέτη.
β) Η στατιστική μεταβλητή είναι η προτίμηση καθενός από τους καταναλωτές όπως αυτή εκφράσθηκε στο δείγμα.

γ) Οι παράμετροι που μας ενδιαφέρουν είναι τα ποσοστά προτίμησης για κάθε μια από τις δύο γεύσεις της μπύρας που θα προέκυπταν αν είχαν ερωτηθεί όλοι οι καταναλωτές της περιοχής στην οποία έγινε η δειγματοληψία (τα ποσοστά αυτά για όλη την περιοχή είναι βέβαια άγνωστα).

δ) Το δείγμα αποτελείται από τους 125 καταναλωτές που έκαναν το τεστ της γεύσης.

ε) Οι σχετικές στατιστικές συναρτήσεις είναι τα ποσοστά των προτιμήσεων για κάθε μια από τις δύο γεύσεις στο δείγμα.

1.5 Αξιολόγηση των Στατιστικών Μελετών

Οπως ήδη εξηγήσαμε, η Στατιστική είναι ένα απαραίτητο εργαλείο για όλες σχεδόν τις επιστήμες. Αυτό βέβαια δε σημαίνει ότι όλες οι απαντήσεις που δίνονται με τη χρήση της Στατιστικής είναι σωστές. Υπάρχουν πολλές περιπτώσεις όπου η χρήση της Στατιστικής οδηγεί σε λανθασμένα συμπεράσματα. Αυτό γιατί υπάρχουν πολλές πηγές λαθών που σχετίζονται με τις στατιστικές μελέτες. Αναφερθήκαμε ήδη σε μια από αυτές, το δειγματοληπτικό σφάλμα. Για να είναι μια στατιστική μελέτη πλήρης θα πρέπει να περιλαμβάνει ένα καμμάτι (μια συζήτηση) για όλες τις πιθανές πηγές λάθους σε οποιαδήποτε παρουσίαση. Δυστυχώς σε πολλές περιπτώσεις αυτό δεν συμβαίνει ή γίνεται μόνο αποσπασματικά. Αυτό είναι λυπηρό γιατί μια τέτοια παρουσίαση των πηγών ενδεχομένων λαθών θα έδινε μεγαλύτερη αξιοπιστία στις μελέτες παρά θα τους αφαιρούσε κύρος. Μερικές πηγές λάθους σε στατιστικές μελέτες είναι οι εξής:

- Αναντιστοιχία μεταξύ του πληθυσμού που μελετάται και του πληθυσμού στον οποίο θα εφαρμοσθούν οι όποιες αποφάσεις.

Λάθη αυτής της μορφής δεν είναι δυνατόν να αποφευχθούν ολοκληρωτικά δεδομένου ότι οι οποιεσδήποτε αποφάσεις ή σχέδια θα

εφαρμοσθούν σε μελλοντικές μορφές του πληθυσμού ή της διαδικασίας. Τα στατιστικά δεδομένα με βάση τα οποία έγινε η ανάλυση προέρχονται από τον παρόντα ή από προηγούμενους πληθυσμούς και διαδικασίες. Θα πρέπει, επομένως, να εξετάζεται κατά πόσο ο πληθυσμός ο οποίος έχει μελετηθεί αντιστοιχεί με τον μελλοντικό πληθυσμό στον οποίο θα εφαρμοσθούν οι όποιες αποφάσεις. Αυτό βέβαια ανήκει στην αρμοδιότητα εκείνου ο οποίος θα πάρει την απόφαση και όχι στο Στατιστικό. Για παράδειγμα, ο υπεύθυνος της ασφαλιστικής εταιρείας στο παράδειγμά μας κατέληξε στο συμπέρασμα ότι οι άνδρες κάνουν μεγαλύτερες ασφάλειες από τις γυναίκες με βάση τους εν ενεργεία ασφαλισμένους. Οποιαδήποτε απόφαση ληφθεί με βάση το συμπέρασμα αυτό υποθέτει ότι το χαρακτηριστικό αυτό θα εξακολουθεί να ισχύει για τους πελάτες μελλοντικά.

- Ακατάλληλη επεξεργασία των στοιχείων μιας διαδικασίας

Ενα από τα πιο συχνά λάθη των στατιστικών αναλύσεων είναι ότι αγνοείται το γεγονός πως τα δεδομένα έχουν συλλεγεί κατά τη διάρκεια μιας χρονικής περιόδου. Εάν αυτό δεν γίνει αντιληπτό, αν, δηλαδή, ο παράγων χρόνος δεν ληφθεί υπόψη , ίσως δεν αντιληφθούμε πιθανές αλλαγές που έχουν συμβεί στη διαδικασία. Επομένως, θα πρέπει πάντοτε να εξετάζουμε αν υπάρχουν συστηματικές αλλαγές στα δεδομένα με την πάροδο του χρόνου. Μόνο αν είναι σαφές ότι δεν έχουν συμβεί σημαντικές αλλαγές θα μπορεί κανείς να υποθέσει ότι τα δεδομένα προέρχονται από ένα πληθυσμό που δεν έχει μεταβληθεί με την πάροδο του χρόνου.

Στο παράδειγμα της ασφαλιστικής εταιρείας θεωρήσαμε το μέσο ποσό ασφάλισης για κάθε φύλο και με τον τρόπο αυτό χωρίσαμε ουσιαστικά τον πληθυσμό μας σε δύο πληθυσμούς. Τα στοιχεία από τα οποία επελέγη το δείγμα περιλαμβάνουν άνδρες και γυναίκες που ασφαλίστηκαν σε διαφορετικές χρονικές στιγμές στο παρελθόν που θα μπορούσαν να αναφέρονται και σε 20 χρόνια στο παρελθόν. Υπάρχει

πιθανότητα να έχει υπάρξει κάποια μεταβολή στον τρόπο συμπεριφοράς των ανδρών και των γυναικών στη χρονική αυτή περίοδο; Ενδελεχής μελέτη θα μπορούσε, ενδεχομένως, να δείξει ότι άνδρες που ασφαλίστηκαν δέκα ή περισσότερα χρόνια προηγουμένως ασφαλίστηκαν κατά μέσο όρο για ποσά μικρότερα από άνδρες οι οποίοι ασφαλίστηκαν τα τελευταία πέντε χρόνια. Σε μια τέτοια περίπτωση μπορεί ο μέσος των 20200000 δρχ. που υπολογίσαμε από τα δειγματικά δεδομένα να θεωρηθεί αξιόπιστος; Αγνοώντας την εξέλιξη του χρόνου μέσα στον οποίο ασφαλίστηκαν οι ασφαλισμένοι του δείγματος είναι ενδεχόμενο να χάσει κανείς σημαντικές πληροφορίες για τάσεις που υπάρχουν στους ασφαλισμένους από τους οποίους προήλθε το δείγμα και να οδηγηθεί σε λανθασμένα συμπεράσματα για το σύνολο των μελλοντικών ασφαλισμένων στην εταιρεία αυτή.

- Αναντιστοιχία πλαισίου και πληθυσμού

Ενα άλλο είδος λάθους της Στατιστικής Ανάλυσης προέρχεται από το ενδεχόμενο το πλαίσιο από το οποίο προήλθε το δείγμα να μην αντιστοιχεί πλήρως προς τον πληθυσμό. Στο παράδειγμα της ασφαλιστικής εταιρείας πλαίσιο ήταν όλοι οι πελάτες της εταιρείας κατά το χρόνο της μελέτης. Καινούργιοι πελάτες για τους οποίους δεν υπήρχε ακόμη φάκελος δεν περιλαμβάνονταν στο πλαίσιο, ενώ, αντίθετα, άλλοι πελάτες οι οποίοι είχαν ακυρώσει την ασφάλειά τους αλλά οι φάκελοί τους δεν είχαν ακόμα κλείσει, περιλαμβάνονταν στο πλαίσιο. Το πόσο σοβαρό είναι αυτό το πρόβλημα είναι κάτι που πρέπει να εκτιμηθεί από τον υπεύθυνο της εταιρείας.

Το πρόβλημα αυτό μπορεί να αποδειχθεί, σε μερικές περιπτώσεις, πολύ σοβαρό. Στο παράδειγμα για τον έλεγχο της γεύσης της μπύρας, το πλαίσιο αποτελείται από όλους τους καταναλωτές που είχαν την δυνατότητα να πάρουν μέρος στο συγκεκριμένο τεστ. Επομένως, το πλαίσιο περιορίζεται στους καταναλωτές που κατοικούν στη συγκεκριμένη πόλη όπου έγινε το

συγκεκριμένο τεστ. Εάν εξάλλου, όπως γίνεται συνήθως, το τεστ έγινε σε κάποιο συγκεκριμένο σημείο (π.χ. αγορά) κάποιο Σαββατοκύριακο, τότε το πλαίσιο περιορίζεται ακόμα περισσότερο στους καταναλωτές εκείνους οι οποίοι πηγαίνουν στη συγκεκριμένη αγορά τα Σαββατοκύριακα. Αν, επομένως, οι καταναλωτές στη συγκεκριμένη πόλη που πηγαίνουν στη συγκεκριμένη αγορά τα Σαββατοκύριακα δεν είναι αντιπροσωπευτικοί όλων των καταναλωτών μπύρας στη συγκεκριμένη χώρα, τα συμπεράσματα μιας τέτοιας μελέτης ενδέχεται να είναι λανθασμένα.

- Επιλογή ακατάλληλων στατιστικών μεταβλητών

Χαρακτηριστικό παράδειγμα του τρόπου με τον οποίο η επιλογή των μεταβλητών επηρεάζει την εξαγωγή στατιστικών συμπερασμάτων είναι η κατάσταση της Οικονομίας. Ανάλογα με το ποιά στατιστική μεταβλητή εξετάζει κανείς είναι δυνατό να ισχυρισθεί ότι μία Οικονομία είναι υγιής ή ότι έχει προβλήματα, μιλώντας πάντα για την ίδια Οικονομία. Μπορεί, για παράδειγμα, κανείς ως χαρακτηριστική στατιστική μεταβλητή για την Οικονομία να θεωρήσει το Ακαθάριστο Εθνικό Προιόν ή εναλλακτικά το Εμπορικό Ισοζύγιο ή το έλλειμμα του προυπολογισμού. Είναι χαρακτηριστικό ότι πολιτικοί αντιπάλων κομμάτων χρησιμοποιούν το ένα ή το άλλο χαρακτηριστικό προκειμένου να υποστηρίξουν ότι η Οικονομία βρίσκεται σε καλή κατάσταση ή έχει προβλήματα ανάλογα με το αν βρίσκονται στην κυβέρνηση ή στην αντιπολίτευση.

- Πρόβλημα μετρήσεων

Μία από τις πηγές σφαλμάτων που δεν της δίνεται αρκετή σημασία είναι η αποτυχία στο να γίνονται ακριβείς μετρήσεις οι οποίες να έχουν συνέπεια. Κάτι τέτοιο μπορεί να συμβαίνει για πολλούς λόγους. Τα όργανα που χρησιμοιούνται για μετρήσεις ίσως να

μην έχουν ελεγχθεί κατάλληλα. Οι άνθρωποι οι οποίοι κάνουν τις μετρήσεις ίσως να μην έχουν την κατάλληλη εκπαίδευση. Μπορεί επίσης ο καθορισμός του υπό μέτρηση αντικειμένου να μην είναι σαφής.

Το τελευταίο πρόβλημα εμφανίζεται πολύ συχνά σε δειγματοληπτικές έρευνες. Συχνά οι ερωτώμενοι δεν καταλαβαίνουν την ερώτηση με τον τρόπο που οι ερωτώντες θα επιθυμούσαν διότι η διατύπωση της ερώτησης δεν είναι σαφής. Αλλη παρόμοια περίπτωση είναι όταν η διατύπωση της ερώτησης "οδηγεί" στην επιθυμητή απάντηση. Τέλος μερικοί από τους ερωτώμενους δεν απαντούν με ειλικρίνεια, ειδικά σε ερωτήσεις που η ειλικρινής απάντηση δεν τους κάνει να αισθάνονται άνετα. Ενα τέτοιο παράδειγμα αποτελούν οι σφυγμομετρήσεις για τις επαναληπτικές εκλογές στην Β' Αθήνας το 1992. Στις εκλογές αυτές μόνο το κόμμα της τότε αξιωματικής αντιπολίτευσης, από τα μεγάλα κόμματα, δήλωσε ότι θα έχει υποψήφιο. Οι σφυγμομετρήσεις απέτυχαν να προσδιορίσουν το ποσοστό που ο υποψήφιος του ΠΑΣΟΚ έλαβε τελικά, όπως επίσης απέτυχαν έστω και να πλησιάσουν το ποσοστό που έλαβε το κόμμα του κ.Λεβέντη. Αυτό γιατί οι ψηφοφόροι της Ν.Δ. (που επίσημα απείχε από τις εκλογές), οι οποίοι τελικά ψήφισαν το κόμμα Λεβέντη, απέφευγαν να το δηλώσουν στις διάφορες σφυγμομετρήσεις.

- Ακατάλληλη επιλογή στατιστικής τεχνικής ή στατιστικού μοντέλου

Πέρα από το να μάθει κανείς μία σειρά από στατιστικές τεχνικές, είναι σημαντικό να έχει την ικανότητα να αντιλαμβάνεται τις περιπτώσεις εκείνες που η επιλογή και χρησιμοποίηση μιας συγκεκριμένης στατιστικής τεχνικής είναι δυνατό να οδηγήσει σε λανθασμένα συμπεράσματα. Το ίδιο συμβαίνει και με την επιλογή στατιστικών μοντέλων. Είναι σημαντικό να ελέγχει κανείς την καταλληλότητα οποιουδήποτε στατιστικού μοντέλου πριν το χρησιμοποιήσει για να πάρει σημαντικές αποφάσεις.

* Συνοπτικά θα μπορούσε να πει κανείς ότι ένα στατιστικό πρόβλημα περιλαμβάνει τα εξής στάδια:

1. Εναν σαφή ορισμό του αντικειμένου του πειράματος και του πληθυσμού στον οποίο το πείραμα αναφέρεται.
2. Τον σχεδιασμό του πειράματος ή της δειγματοληπτικής διαδικασίας.
3. Την συλλογή και την ανάλυση των δεδομένων.
4. Την διαδικασία στατιστικής συμπερασματολογίας για τον υπό μελέτη πληθυσμό με βάση τις πληροφορίες του δείγματος.
5. Την παροχή ενός μέτρου καταλληλότητας (αξιοπιστίας) της συμπερασματολογίας.

1.6 Ο ρόλος των Στατιστικών για την εξαγωγή Στατιστικών Συμπερασμάτων

Οπως είναι γνωστό, οι άνθρωποι καταγράφουν παρατηρήσεις και συλλέγουν δεδομένα επί πολλούς αιώνες. Επιπλέον χρησιμοποίησαν και χρησιμοποιούν τα δεδομένα ως βάση για προβλέψεις και για τη λήψη αποφάσεων χωρίς καμμιά βοήθεια της Στατιστικής. Τι είναι λοιπόν εκείνο που μπορούν οι Στατιστικοί και η Στατιστική να συνεισφέρουν;

Οπως ήδη είπαμε, η Στατιστική είναι η επιστημονική περιοχή η οποία ασχολείται με την εξαγωγή πληροφοριών από αριθμητικά δεδομένα και την χρησιμοποίηση των πληροφοριών αυτών για την εξαγωγή συμπερασμάτων σε σχέση με τον πληθυσμό από τον οποίο προήλθαν τα δεδομένα, κάτω πάντοτε από συνθήκες αβεβαιότητας. Ενας Στατιστικός ποσοτικοποιεί τις πληροφορίες, μελετά διαδικασίες σχεδιασμού και δειγματοληψίας και αναζητά την κατάλληλη μεθοδολογία η οποία θα δώσει τις περισσότερες δυνατές πληροφορίες, κάτω από τις συγκεκριμένες συνθήκες με ελάχιστο κόστος. Επομένως,

μια σημαντική συνεισφορά ενός Στατιστικού είναι ο σχεδιασμός των πειραμάτων και των δειγματοληπτικών ερευνών με τρόπο ώστε να ελαττωθεί το κόστος μελέτης μεγάλων πληθυσμών. Η δεύτερη σημαντική συνεισφορά είναι αυτή καθεαυτή η συμπερασματολογία. Ο Στατιστικός μελετά τις διάφορες μεθόδους συμπερασματολογίας και αναζητά την καλύτερη προκειμένου να κάνει προβλέψεις ή να λάβει αποφάσεις κάτω από ορισμένες συνθήκες. Το πιο σημαντικό όμως είναι ότι ο Στατιστικός δίνει πληροφορίες που αναφέρονται στην ποιότητα κάθε συγκεκριμένης συμπερασματολογίας. Οπως είναι φυσικό, όταν κάνουμε προβλέψεις μας ενδιαφέρει να ξέρουμε κάτι σε σχέση με το λάθος που είναι δυνατόν να κάνουμε στην πρόβλεψή μας. Επίσης, εάν πρόκειται να πάρουμε κάποια απόφαση, θέλουμε να ξέρουμε την πιθανότητα που υπάρχει να έχουμε πάρει μια λανθασμένη απόφαση. Η έμφυτη ικανότητα του ανθρώπου να προβλέπει και να παίρνει αποφάσεις δεν παρέχει άμεσες απαντήσεις στα σημαντικά αυτά ερωτήματα. Η έμφυτη αυτή ανθρώπινη ικανότητα μπορεί να εκτιμηθεί μόνο με μακροχρόνιες παρατηρήσεις. Σε αντίθεση, οι στατιστικές διαδικασίες παρέχουν άμεσες απαντήσεις στα ερωτήματα αυτά. Επομένως, η Στατιστική μας δίνει την δυνατότητα να εξαγάγουμε συμπεράσματα από δειγματικά δεδομένα και να αποτιμήσουμε την αξιοπιστία των συμπερασμάτων αυτών. Αυτής της μορφής η πληροφορία είναι χρήσιμη σε οποιασδήποτε μορφής απόφαση ή πρόβλεψη.

1.7 Τρόποι συλλογής στοιχείων

Οπως ήδη αναφέραμε, ο στόχος της Στατιστικής Ανάλυσης είναι να μάθουμε όσο το δυνατόν περισσότερα γύρω από έναν πληθυσμό ή μια διαδικασία που μελετάμε. Είναι, επομένως, επιθυμητό να συλλέξουμε στοιχεία (δεδομένα) που χαρακτηρίζουν τον πληθυσμό ή τη διαδικασία όσο το δυνατόν καλύτερα, ελαχιστοποιώντας το λάθος επιλογής. Υπάρχουν τέσσερις κύριες μέθοδοι συλλογής στοιχείων:

1. Η επιλογή **τυχαίου δείγματος (random sample)**,
2. Η διεξαγωγή ενός **τυχαιοποιημένου πειράματος (randomized experiment)**,
3. Η χρησιμοποίηση **διαθέσιμων στοιχείων (encountered data** ή **convenience data)** και
4. Η επιλογή **λογικών υποομάδων (rational subgroups)**, δηλαδή σχεδιασμένα δείγματα που επιλέγονται στην πορεία του χρόνου.

Τυχαία δείγματα (random samples)

Η πιο αποτελεσματική μέθοδος επιλογής δείγματος ώστε να ελέγχεται το δειγματικό λάθος είναι η **Τυχαία Δειγματοληψία (Random Sampling)**. Υπάρχουν πολλές εξειδικεύσεις της τυχαίας δειγματοληψίας. Η απλούστερη από αυτές είναι η **Απλή Τυχαία Δειγματοληψία (Simple Random Sampling)**, σύμφωνα με την οποία αν θέλουμε να επιλέξουμε ένα δείγμα μεγέθους n, κάθε υποομάδα n στοιχείων του πληθυσμού θα πρέπει να έχει την ίδια πιθανότητα να επιλεγεί. Η τυχαία δειγματοληψία έχει πολλά πλεονεκτήματα σε σχέση με μη τυχαίες μεθόδους επιλογής.

1. *Αποκλεισμός της μεροληψίας.* Αν επιλέξουμε τα στοιχεία ενός δείγματος αυθαίρετα υπάρχει πάντοτε το ενδεχόμενο της

μεροληψίας, έστω και αν αυτό δε γίνει ηθελημένα. Παρ'ότι η τυχαία δειγματοληψία δεν εξασφαλίζει ότι ένα δείγμα είναι αντιπροσωπευτικό, αποκλείει το ενδεχόμενο μεροληψίας στην επιλογή.

2. *Καθορισμός εμπιστοσύνης.* Η τυχαία δειγματοληψία παρέχει μια στατιστική βάση για τον καθορισμό της εμπιστοσύνης που συνδέεται με τη συγκεκριμένη συμπερασματολογία. Αυτό δεν μπορεί να συμβεί αν τα στοιχεία του δείγματος επιλέγονται με οποιοδήποτε άλλο τρόπο.

3. *Ελεγχος δειγματικού λάθους.* Η μεθοδολογία αυτή επιτρέπει τον έλεγχο του δειγματικού λάθους με αντίστοιχο έλεγχο του μεγέθους του δείγματος. Παρέχει, επομένως, τη δυνατότητα καθορισμού του δειγματικού λάθους σε επιθυμητό επίπεδο. Με μια μη τυχαία μέθοδο δειγματοληψίας είναι ενδεχόμενο να καταλήξουμε σε ένα μη αποδεκτό επίπεδο δειγματικού λάθους.

Υπάρχουν διάφοροι τρόποι επιλογής ενός δείγματος με τη μέθοδο της τυχαίας δειγματοληψίας. Οι μέθοδοι αυτές περιγράφονται σε σχετικά εγχειρίδια δειγματοληψίας.

Τυχαιοποιημένα πειράματα (randomized experiments)

Ενας από τους κυριότερους τρόπους από εκείνους με τους οποίους είναι δυνατό να μελετήσουμε τις αιτίες της μεταβλητότητας είναι να σχεδιάσουμε ένα πείραμα. Κλασσικό παράδειγμα αποτελεί ο έλεγχος της αποτελεσματικότητας ενός λιπάσματος στην παραγωγή σταριού. Εκείνο που συνήθως γίνεται είναι ότι μια συγκεκριμένη ποικιλία σταριού φυτεύεται σε ένα κομμάτι ενός αγρού και λιπαίνεται με το συγκεκριμένο λίπασμα. Η ίδια ποικιλία σταριού φυτεύεται σε μια άλλη περιοχή με παρόμοια χαρακτηριστικά και λιπαίνεται με ένα συνηθισμένο λίπασμα. Στη συνέχεια συγκρίνονται

οι παραγωγές που προέρχονται από τις δύο περιοχές του αγρού ώστε να εκτιμηθεί η αποτελεσματικότητα του νέου λιπάσματος σε σχέση με το συνήθως χρησιμοποιούμενο λίπασμα. Οι σπόροι του σταριού που χρησιμοποιούνται στο τέστ (πειραματικές μονάδες (experimantal units)) θα πρέπει να τοποθετηθούν τυχαία στα δύο κομμάτια του αγρού. Αυτό μπορεί να επιτευχθεί ως εξής: έστω ότι πρόκειται να φυτευθούν 100 σπόροι σταριού και 50 από αυτούς θα λιπανθούν με το συνηθισμένο λίπασμα, ενώ οι άλλοι 50 με το υπό μελέτη λίπασμα. Μπορούμε να τοποθετήσουμε και τους 100 σπόρους σε ένα σάκο, να ανακινήσουμε πολύ καλά το σάκο και να επιλέξουμε στην τύχη τους σπόρους. Η τυχαιοποιημένη αυτή κατανομή των σπόρων βεβαιώνει οτι η μελέτη έχει τα πλεονεκτήματα της τυχαίας δειγματοληψίας που αναλύθηκαν στην προηγούμενη ενότητα. Στην περίπτωση αυτή ο πληθυσμός δεν είναι βέβαια οι 100 σπόροι που χρησιμοποιήθηκαν στη μελέτη. Αντίθετα ο πληθυσμός είναι ένας υποθετικός πληθυσμός. Είναι η παραγωγή όλου του σταριού του συγκεκριμένου αυτού είδους που θα παίρναμε εάν χρησιμοποιούσαμε τις συνθήκες της μελέτης (παρόμοιο έδαφος, λίπασμα, νερό κ.λ.π.). Στην περίπτωσή μας, μας ενδιαφέρει κυρίως το αποτέλεσμα του λιπάσματος στη διαδικασία ανάπτυξης του σταριού. Χρησιμοποιούμε, λοιπόν, το τυχαιοποιημένο πείραμα για να αντιληφθούμε καλύτερα αυτή τη διαδικασία.

Ο προσεκτικός σχεδιασμός των τυχαιοποιημένων πειραμάτων μας επιτρέπει να πάρουμε τη μέγιστη δυνατή πληροφορία για το φαινόμενο που μελετάμε με ελάχιστο κόστος. Οι αρχές που χρησιμοποιούνται στο σχεδιασμό ενός τυχαιοποιημένου πειράματος μελετώνται περισσότερο στο σχεδιασμό και την ανάλυση πειραμάτων.

Διαθέσιμα δεδομένα (encountered data ή convenience data)

Σε πολλά προβλήματα δεν είναι δυνατό να χρησιμοποιήσουμε τυχαία δειγματοληψία. Αντίθετα, θα πρέπει να στηριχθούμε σε δεδομένα που απλώς είναι διαθέσιμα στον αναλυτή. Τέτοια δεδομένα έχουμε κυρίως στην Ιατρική όταν τα δεδομένα προέρχονται από

ανθρώπους που έχουν επισκεφθεί το νοσοκομείο ή το γιατρό για ένα συγκεκριμένο λόγο. Παρόμοια είναι η περίπτωση όταν αναλύουμε δεδομένα για σεισμούς.

Πολλές φορές η ανάλυση τέτοιων δεδομένων γίνεται με την υπόθεση ότι μπορούν να εξομοιωθούν με δεδομένα που έχουν προκύψει από τυχαία δειγματοληψία. Είναι όμως προφανές ότι μια τέτοια ανάλυση είναι παρακινδυνευμένη ως προς τα αποτελέσματά της. Είναι επίσης φυσικό ότι τέτοια δεδομένα, τις περισσότερες φορές, έχουν κάποιας μορφής μεροληπτικότητα. Εν γένει, μπορούμε να πούμε ότι δύο είναι τα σημαντικότερα μειονεκτήματα των διαθέσιμων δεδομένων:

1. Δεν παρέχουν τις δυνατότητες που παρέχουν τα δεδομένα που προέρχονται από τυχαία δειγματοληψία και

2. Δεν μας δίνουν τη δυνατότητα να σχεδιάσουμε τη διαδικασία επιλογής ώστε να επωφεληθούμε από τα πλεονεκτήματα που προσφέρουν τα τυχαιοποιημένα πειράματα.

Λογικές υποομάδες (rational subgroups)

Οταν μελετάμε τη συμπεριφορά μιας διαδικασίας προσπαθούμε να διερευνήσουμε τη μεταβλητότητα που παρατηρείται στα αποτελέσματά της και να καθορίσουμε τις αιτίες που την προκαλούν. Πολλές φορές δε μας ενδιαφέρει απλά να μελετήσουμε τη μεταβλητότητα των αποτελεσμάτων μιας διαδικασίας που προκύπτουν σε μια συγκεκριμένη χρονική στιγμή, αλλά ενδιαφερόμαστε επίσης για τη μεταβλητότητα που παρουσιάζεται στα αποτελέσματα της διαδικασίας με την πάροδο του χρόνου.

Οταν μελετάμε τη μεταβλητότητα μιας διαδικασίας σε μια συγκεκριμένη χρονική στιγμή παρατηρούμε τα αποτελέσματα που δίνει η διαδικασία όσο το δυνατόν πλησιέστερα στο συγκεκριμένο χρόνο και κάτω από τις ίδιες συνθήκες. Αυτό ελαχιστοποιεί το ενδεχόμενο ότι η διαδικασία μεταβάλλεται κατά τη διάρκεια της μικρής χρονικής περιόδου στην οποία αναφέρονται τα συγκεκριμένα δεδομένα. Αν όμως ενδιαφερόμαστε να μελετήσουμε τις μεταβολές μιας διαδικασίας σε

μια πιο εκτεταμένη χρονική περίοδο επιλέγουμε δείγματα σε τακτά χρονικά διαστήματα. Η μεταβλητότητα μεταξύ των δειγμάτων μας επιτρέπει να επισημάνουμε μεταβολές στη διαδικασία.

Τα διαστήματα στα οποία λαμβάνονται τα δείγματα επιλέγονται με υποκειμενικό τρόπο που βασίζεται στη γνώση της διαδικασίας. Για παράδειγμα, ένα δείγμα σε μια διαδικασία παραγωγής μπορεί να επιλέγεται κάθε μισή ώρα μιας δεδομένης ημέρας. Εάν, αντίθετα, επιλέξουμε ένα τυχαίο δείγμα από τη συνολική παραγωγή μιας ημέρας σε μια χρονική στιγμή δε θα έχουμε τη δυνατότητα να ξεχωρίσουμε τη μεταβλητότητα των προιόντων που παρήχθησαν την ίδια στιγμή από τη μεταβλητότητα που υπάρχει λόγω μεταβολής του χρόνου.

Η μέθοδος επιλογής μικρών δειγμάτων σε τακτά χρονικά διαστήματα όταν τα στοιχεία του δείγματος, για ένα δεδομένο δείγμα, λαμβάνονται κατά το δυνατόν περισσότερο κάτω από τις ίδιες συνθήκες και τον ίδιο περίπου χρόνο, ονομάζεται **μέθοδος των λογικών υποομάδων (rational subgroups)**.

1.8 Στατιστικός Τρόπος Σκέψης για Διαδικασίες Λήψης Αποφάσεων

Μια από τις κυριότερες εφαρμογές της Στατιστικής είναι η χρήση των μεθόδων της για την υποβοήθηση της λήψης αποφάσεων. Στο πλαίσιο αυτό **στατιστική σκέψη** είναι ένας τρόπος σκέψης που μας επιτρέπει να καταλάβουμε και, τελικά, να βελτιώσουμε κάποιες διεργασίες μέσω ενδελεχούς μελέτης της μεταβλητότητας των δεδομένων.

Οπως έχουμε ήδη πει, η κατανόηση και ερμηνεία της μεταβλητότητας και των λόγων που την προκαλούν αποτελεί το κύριο συστατικό στη διαδικασία λήψης αποφάσεων. Για παράδειγμα, έστω ότι ο διευθυντής πωλήσεων μιας εταιρείας διαπιστώνει ότι οι πωλήσεις σε μια συγκεκριμένη περιοχή μεταβάλλονται από μήνα σε μήνα. Εάν παρατηρήσει πτώση των πωλήσεων επί τρεις διαδοχικούς μήνες θα

πρέπει να θεωρήσει ότι αυτό αποτελεί ένδειξη μιας σημαντικής τάσης; Αν είναι έτσι ποιοι είναι οι λόγοι που προκαλούν την τάση αυτή και τι ενέργειες θα πρέπει να κάνει;

Ενα από τα κύρια στοιχεία για την αποτελεσματική ερμηνεία της μεταβλητότητας είναι η αντίληψη και κατανόηση της διαδικασίας (διεργασίας) από την οποία η μεταβλητότητα αυτή προήλθε. Το ουσιαστικό στοιχείο της στατιστικής σκέψης είναι να αναγνωρίσει ότι όλες οι διαδικασίες παράγουν μεταβλητότητα και ότι η ελάττωση της μεταβλητότητας είναι το κλειδί για τη βελτίωση της διαδικασίας. Η συστηματική χρησιμοποίηση της Στατιστικής για τη σωστή ερμηνεία της μεταβλητότητας των δεδομένων σε συνάρτηση με την γνώση του αντικειμένου παρέχει τα μέσα με τα οποία αυξάνεται η αντίληψη του τρόπου που εξελίσσεται η διαδικασία και, επομένως, παρέχει τη δυνατότητα βελτίωσης της διαδικασίας αυτής.

Στο σημείο αυτό θα πρέπει να υπενθυμίσουμε ότι *διεργασία (διαδικασία) είναι ένα σύνολο συνθηκών, μέσω της από κοινού κατ'επανάληψη λειτουργίας των οποίων, "διαθέσιμο υλικό" μετατρέπεται σε "προιόντα"*. Παραδείγματα "διαθέσιμου υλικού" είναι οι πρώτες ύλες σε μια κατασκευαστική διαδικασία, εκπαιδευτικά μαθήματα τα οποία παρακολουθούν τεχνικοί που ασχολούνται με την συντήρηση μηχανών, αλλά και οι σημειώσεις των φοιτητών σε ένα μάθημα που θα χρησιμοποιηθούν στη διαδικασία μελέτης για τις εξετάσεις. Ως "προιόντα" μιας διαδικασίας μπορούν να θεωρηθούν οι γραπτές αναφορές, τιμολόγια, υπηρεσίες, προιόντα και οι βαθμοί φοιτητών σε ένα διαγώνισμα. Τα "προιόντα" καταλήγουν στους "πελάτες" της διεργασίας. Επομένως μια διεργασία (διαδικασία) μπορεί να θεωρηθεί από αυτόν που τη συντηρεί ως ένας πελάτης. Την ίδια στιγμή τα προιόντα μιας διεργασίας μπορούν να θεωρηθούν ως "διαθέσιμο υλικό" της διεργασίας του πελάτη. Γενικά, επομένως, μια διεργασία λαμβάνει το διαθέσιμο υλικό της (input) από μια διεργασία ενός προμηθευτή και στέλνει το "προιόν" σε μια διαδικασία του πελάτη.

Τα βήματα που συνοψίζουν τη χρήση της στατιστικής σκέψης για την καλύτερη κατανόηση και τελική βελτίωση μιας διεργασίας είναι τα εξής:

1. Κατ'αρχήν θα πρέπει να αντιληφθούμε πώς η υπό μελέτη διεργασία λειτουργεί τη συγκεκριμένη στιγμή. Ποια είναι τα "διαθέσιμα υλικά" (inputs) στη διεργασία αυτή; Ποιοι είναι εκείνοι που προμηθεύουν τα "υλικά" αυτά; Ποιοι είναι οι σημαντικοί εσωτερικοί παράγοντες της διεργασίας; Ποιες είναι οι ενέργειες που γίνονται; Με ποια σειρά; Τι αποφάσεις πρέπει να ληφθούν στο εσωτερικό της διεργασίας; Ποια είναι τα "προιόντα"; Ποιοι είναι οι "πελάτες" της διεργασίας;

2. Στη συνέχεια αποτιμάμε την απόδοση της διεργασίας κατά τη στιγμή της μελέτης. Ποια είναι τα υψηλότερα επίπεδα απόδοσης που μπορούμε ρεαλιστικά να περιμένουμε από τη διεργασία αυτή με τον τρόπο που τη στιγμή αυτή λειτουργεί;

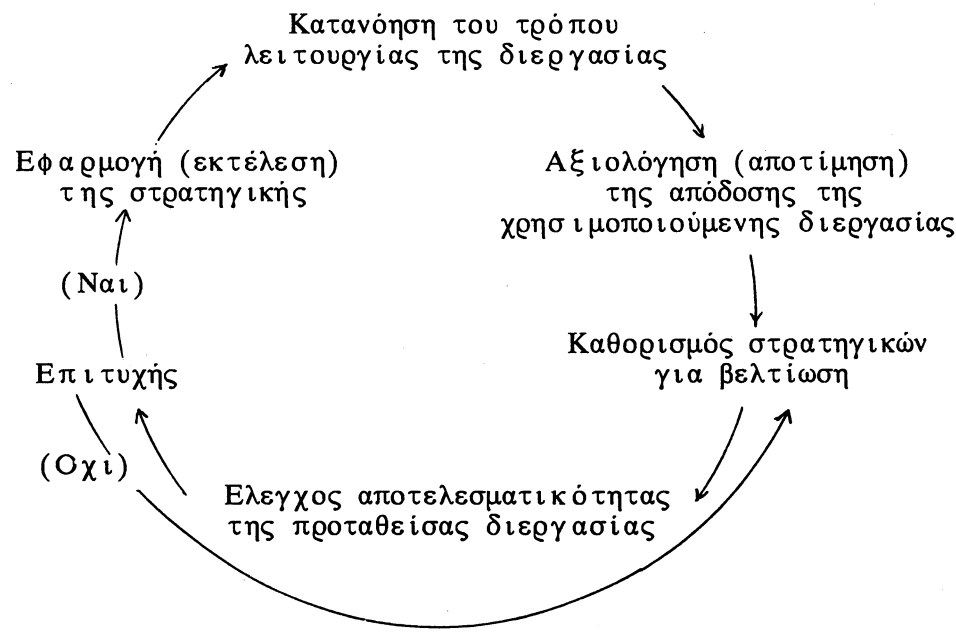
3. Καθορίζουμε πιθανές στρατηγικές που θα βοηθήσουν τη βελτίωση της διεργασίας.

4. Ελέγχουμε την αποτελεσματικότητα των στρατηγικών που επιλέξαμε.

5. Εάν οι έλεγχοι καταλήγουν σε κάποια ένδειξη επιτυχίας, εφαρμόζουμε τις μεταβολές αυτές στη διεργασία. Εάν τα αποτελέσματα των ελέγχων δεν είναι ενθαρρυντικά επιστρέφουμε στο βήμα 3.

6. Επιστρέφουμε στο βήμα 1 μιας κυκλικής διαδικασίας βελτίωσης χωρίς τέλος.

Ο ατελεύτητος αυτός κύκλος της βελτίωσης απεικονίζεται στο σχήμα 1.8.1.



Σχήμα 1.8.1

Εχουμε ήδη εξηγήσει ότι ο στατιστικός τρόπος σκέψης εμπεριέχει τη χρήση στατιστικών μεθόδων σε συνδυασμό με την γνώση του αντικειμένου. Ας δούμε πώς συνεισφέρουν ο στατιστικός τρόπος σκέψης και η γνώση του αντικειμένου στον κύκλο βελτίωσης που αναπτύξαμε. Το πρώτο στάδιο βασίζεται αποκλειστικά στις πρόσφατες γνώσεις του αντικειμένου. Η κατανόηση της αλληλεπίδρασης των στοιχείων μέσα σε μια διεργασία είναι απαραίτητη για την αποτελεσματικότητα της ανάλυσης που θα υιοθετηθεί. Τα βασικά εργαλεία για τη δουλειά αυτή είναι το *διάγραμμα ροής (flow diagram)* και το *διάγραμμα αιτίου και αιτιατού (cause and effect diagram).*

Το διάγραμμα ροής (flow diagram) είναι απλά ένα διάγραμμα που αποτυπώνει τη διεργασία. Αποτυπώνει, δηλαδή, με ένα συστηματικό τρόπο πώς οι διάφοροι εσωτερικοί παράγοντες της διεργασίας μετατρέπουν το "διαθέσιμο υλικό" (εισρέοντα στοιχεία) σε

"προιόντα". Δεδομένου ότι το πρώτο βήμα για τη βελτίωση μιας διεργασίας είναι η καλύτερη κατανόησή της, πολλοί θεωρούν το διάγραμμα ροής ως το πιο σημαντικό εργαλείο για τη βελτίωση της διεργασίας.

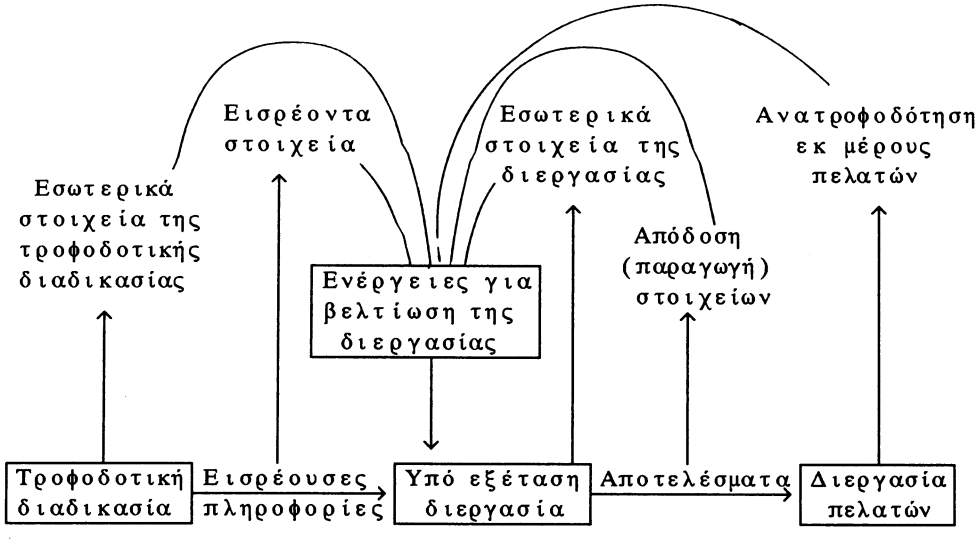
Το διάγραμμα **αιτίου και αιτιατού (cause and effect diagram)** εστιάζεται σε ένα "αποτέλεσμα" (effect) - ένα πρόβλημα ή ένα επιθυμητό αποτέλεσμα - και επιχειρεί να καθορίσει τις αιτίες (causes) που το προκαλούν. Είναι μια μορφή οργανωμένης δημιουργικής σκέψης (brain-storming). Σ'αυτό καθορίζονται πολλές βασικές κατηγορίες ενδεχομένων αιτίων. (Ενα σύνηθες σύνολο κατηγοριών που χρησιμοποιούνται είναι άνθρωποι, μηχανές, το περιβάλλον, υλικά, μέθοδοι και μετρήσεις.) Διαγράμματα αιτίου και αιτιατού αναπτύσσονται περισσότερο αποτελεσματικά από ομάδες που χρησιμοποιούν την τεχνική της οργανωμένης δημιουργικής σκέψης (brain-storming) για να προσδιορίσουν ενδεχόμενες αιτίες και να τις κατατάξουν στην κατάλληλη κατηγορία. Με το να δίνεται η δυνατότητα και ο τρόπος ώστε να εκφράζονται και να συλλέγονται πολλές απόψεις για κάποιο πρόβλημα, τα διαγράμματα αιτίου και αιτιατού αποκαλύπτουν συχνά μια λύση που εξυπηρετεί με τον καλύτερο τρόπο τον οργανισμό στον οποίο γίνεται η διεργασία ως σύνολο.

Για να χαρακτηρίσουμε την τρέχουσα απόδοση της διεργασίας (στάδιο 2) χρησιμοποιούμε τις γνώσεις για το αντικείμενο ώστε να καθορίσουμε ποιες μεταβλητές θα μελετήσουμε και το είδος της στατιστικής μεθόδου που θα χρησιμοποιήσουμε. Μεταξύ των κυριότερων στατιστικών γραφικών μεθόδων που χρησιμοποιούνται είναι *χαρτογραφήσεις (χάρτες) ροής (run charts), χαρτογραφήσεις ελέγχου (control charts), ιστογράμματα (histograms), διαγράμματα Pareto (Pareto diagrams)* και *διαγράμματα διάχυσης (scatter diagrams).* Προκειμένου να προτείνουμε δυνατές στρατηγικές για βελτίωση (στάδιο 3), βασιζόμαστε στη γνώση του αντικειμένου . Ο έλεγχος της αποτελεσματικότητας των προτεινομένων στρατηγικών (στάδιο 4)

στηρίζεται κυρίως σε στατιστικές μεθόδους, η πιο αποτελεσματική από τις οποίες είναι το *σχεδιασμένο πείραμα (designed experiment).* Η εφαρμογή της διαδικασίας (στάδιο 5) είναι κυρίως μια μη στατιστική εργασία παρ'ότι τα αποτελέσματα οποιασδήποτε σχεδιαζόμενης μεταβολής θα πρέπει να ελεγχθούν στατιστικά. Οπως περνάμε τα διάφορα στάδια για να μελετήσουμε, να καταλάβουμε και να βελτιώσουμε μια διεργασία είναι αναπόφευκτο να δημιουργούμε τόσα καινούργια ερωτήματα όσες είναι και οι απαντήσεις που δίνουμε. Είναι φυσικό ότι κάθε μελέτη οδηγεί σε μια άλλη μελέτη και στη διαρκή διαδικασία της βελτίωσης της διεργασίας (στάδιο 6). Επομένως, αυτή καθεαυτή η στατιστική σκέψη είναι μια συνεχής διαδικασία χωρίς τέλος που θα πρέπει να θεωρείται αναπόσπαστο τμήμα της πρακτικής της λήψης αποφάσεων.

Μια νέα θεώρηση που προέκυψε από τη συνεργασία και τις εφαρμογές της Στατιστικής στο management είναι ότι όλες οι πλευρές μιας διεργασίας θα πρέπει να μελετώνται προκειμένου να επισημαίνονται οι υπάρχουσες δυνατότητες για βελτίωση. Αυτό είναι το ουσιαστικό σημείο του **management ολικής ποιότητας (total quality management).** Το management ολικής ποιότητας είναι ένας τρόπος διοίκησης που εφαρμόζεται επί πολλά χρόνια στην Ιαπωνία, πρόσφατα μεταφέρθηκε και εξαπλώνεται με ταχύτητα στις Η.Π.Α. και στη συνέχεια στην Ευρώπη και στον υπόλοιπο κόσμο. Πολλοί οργανισμοί, ιδιωτικοί και δημόσιοι, συγκέντρωναν αποκλειστικά την προσοχή τους στο "προιόν" μιας διαδικασίας. Αυτό έχει αποδειχθεί πλέον ότι αποτελεί μια λανθασμένη πρακτική. Είναι πια φανερό ότι είναι δυνατόν να βελτιωθούν πολλά πράγματα από τη μελέτη όχι μόνο των "προιόντων" αλλά και του "διαθέσιμου υλικού" (των εισρεουσών πληροφοριών) των εσωτερικών παραγόντων της διεργασίας και από την ανατροφοδότητση πληροφοριών (feed back) εκ μέρους των πελατών για τα "προιόντα" της διεργασίας. (Πολλές φορές μάλιστα θεωρείται ότι η εικόνα του καταναλωτή (πελάτη) για ένα προιόν θα πρέπει να θεωρείται ως "προιόν").

Οι ροές των εισρεόντων στοιχείων και αποτελεσμάτων από την διαδικασία τροφοδότησης προς την διεργασία και τον πελάτη καθώς και των σχετικών δυνατοτήτων μελέτης απεικονίζονται στο στο διάγραμμα 1.8.2. (Το διάγραμμα αυτό εφαρμόζεται σε διεργασίες οποιασδήποτε φύσης, όπως πωλήσεις, εξυπηρέτηση, διοικητικές λειτουργίες κ.λ.π.).



Σχήμα 1.8.2

Προκειμένου να μελετήσουμε της απόδοση μιας διεργασίας (στάδιο 2) θα πρέπει να καθορίσουμε κάποιους δείκτες απόδοσης. Ας θεωρήσουμε, για παράδειγμα, τη διαδικασία οδήγησης από το σπίτι στον τόπο δουλειάς καθημερινά. Για τη διαδικασία αυτή ένας σημαντικός δείκτης απόδοσης (performance indicator) θα μπορούσε να θεωρηθεί ο χρόνος που απαιτείται.

Μια στατιστική μεταβλητή που χρησιμοποιείται για να χαρακτηρίσει την ποιότητα των προιόντων μιας διεργασίας ονομάζεται **ποιοτικό χαρακτηριστικό (quality characteristic)**. Ενα μεγάλο μέρος του έργου βελτίωσης μιας διεργασίας αναφέρεται στην παρατήρηση της μεταβλητότητας ενός ποιοτικού χαρακτηριστικού, στον προσδιορισμό

του αιτίου (των αιτίων) που προκαλεί τη μεταβλητότητα αυτή και στη λήψη απόφασης επί της διεργασίας ώστε να ελαττωθεί η μεταβλητότητα αυτή. Σε πολλές περιπτώσεις η μεταβλητότητα μεταξύ των "προιόντων" μιας διεργασίας μπορεί να αποδοθεί σε άλλες εσωτερικές μεταβλητές της διεργασίας. Για παράδειγμα, ο χρόνος που απαιτείται για να οδηγήσει κανείς από το σπίτι του στη δουλειά του ίσως εξαρτάται από τη διαδρομή που ακολουθεί, τη χρονική στιγμή που ξεκινά, τον αριθμό των φορών που υποχρεώνεται να σταματήσει σε σηματοδότες ή τις καιρικές συνθήκες.

Υπάρχουν δύο είδη αιτίων διακύμανσης. Οι **κοινές αιτίες (common causes)** και οι **ειδικές αιτίες** ή **προσδιορίσιμες αιτίες (special causes** ή **assignable causes).**

Οι κοινές αιτίες (common causes) μεταβλητότητας είναι συνηθισμένοι ή κανονικοί παράγοντες εσωτερικοί της διεργασίας (που περιλαμβάνουν και το διαθέσιμο υλικό για τη διεργασία) που μεταβάλλονται φυσιολογικά με την πάροδο του χρόνου, ώρα με ώρα, μέρα με μέρα. Οι κοινές αιτίες επηρεάζουν όλα τα "προιόντα". Για παράδειγμα, κοινές αιτίας διακύμανσης στο πρόβλημα του χρόνου που απαιτείται για να μεταβεί κανείς στη δουλειά του περιλαμβάνουν τη διαδρομή που ακολουθεί κανείς, την ταχύτητα με την οποία οδηγεί (μέσα στα συνήθη όρια), τον αριθμό των φορών που συναντά κανείς κόκκινο και την ένταση της κυκλοφορίας. Ο χρόνος μετάβασης επηρεάζεται από όλες αυτές τις μεταβλητές κάθε φορά που γίνεται η μετακίνηση αυτή.

Από το άλλο μέρος, οι ειδικοί παράγοντες είναι ασυνήθιστα περιστατικά που δεν αποτελούν συνηθισμένο μέρος της διεργασίας. Για παράδειγμα, μπορεί κανείς να πάθει λάστιχο ή για κάποιο λόγο να ξεκινήσει το πρωί από το σπίτι του αργότερα και να πέσει σε άλλο ρυθμό κυκλοφορίας. Ειδικές αιτίες, επίσης, είναι συχνά βλάβες σε κάποιο εξάρτημα ή μέρος της διεργασίας όπως, για παράδειγμα, η επιδιόρθωση μιας μηχανής ή το λάστιχο στο παράδειγμά μας. Πολλές φορές είναι δυνατόν να προκληθούν ανέλπιστες βελτιώσεις από

τυχαίες ανακαλύψεις. Εν γένει οι ειδικές αιτίες επηρεάζουν ένα, ή μερικά μόνο, από τα "προιόντα". Για παράδειγμα, τις περισσότερες φορές ο χρόνος που απαιτείται για να πάει κανείς στη δουλειά του δεν μπορεί να θεωρηθεί ότι επηρεάζεται από μηχανικές βλάβες, όμως, εάν κάποια βλάβη παρουσιασθεί, ο χρόνος μετάβασης θα είναι κατά τεκμήριο πολύ περισσότερος απ'ότι ο συνηθισμένος.

Ως **ευσταθή διεργασία (stable process)** χαρακτηρίζουμε τη διεργασία εκείνη στην οποία υπάρχουν μόνο κοινές αιτίες διακύμανσης. Σε τέτοιες διεργασίες, δηλαδή, δεν υπάρχουν ειδικά γεγονότα που προκαλούν ασυνήθιστη μεταβλητότητα. Δοθέντος ότι όλες οι κοινές αιτίες έχουν επίπτωση σε όλα τα "προιόντα", η μεταβλητότητα μεταξύ των "προιόντων" αντανακλά τη συνολική επίδραση της μεταβλητότητας σε όλα τα στοιχεία της διεργασίας. Αν κάποιο συγκεκριμένο "προιόν" έχει μεγαλύτερη τιμή από ότι τα περισσότερα άλλα, αυτό δεν μπορούμε να το αποδώσουμε σε μια μοναδική αιτία. Μάλλον θα πρέπει να θεωρήσουμε ότι η συνολική επίδραση όλων των κοινών αιτίων στο προιόν αυτό είχε ως αποτέλεσμα να οδηγήσει σε μια μεγαλύτερη τιμή για το προιόν. Σε μια ευσταθή διεργασία το *σύστημα των αιτιών που προκαλούν τη μεταβλητότητα παραμένει σταθερό στην πάροδο του χρόνου.* Τα "προιόντα" διαφέρουν μεταξύ τους αλλά το εύρος της μεταβλητότητας παραμένει κατ'ουσίαν σταθερό και, επομένως, είναι δυνατόν να προβλεφθεί.

Είναι σημαντικό να γίνει κατανοητό ότι ο όρος *ευστάθεια* στην περίπτωση αυτή δεν έχει την έννοια της "απόδοσης κατά το επιθυμητό". Αυτό που σημαίνει είναι ότι το φάσμα των προιόντων μπορεί να προβλεφθεί μέσα σε κάποια συγκεκριμένα όρια. Εάν συμβεί να παρατηρήσουμε ένα "προιόν" έξω από τα όρια αυτά, αυτό θα αποτελέσει ένα λόγο για να υποπτευθούμε ότι μια ειδική αιτία έχει επηρεάσει το συγκεκριμένο "προιόν".

Ασταθής διεργασία (unstable process) είναι η διεργασία εκείνη η οποία επηρεάζεται όχι μόνο από κοινές αιτίες αλλά και από ειδικές αιτίες. Δοθέντος ότι οι ειδικές αιτίες δεν αποτελούν

σύνηθες μέρος της διεργασίας δεν μπορούν να προβλεφθούν. Επομένως, το εύρος της μεταβλητότητας των ασταθών διεργασιών δεν είναι προβλέψιμο. Αν ένα "προιόν" έχει μια τιμή έξω από το σύνηθες εύρος μεταβλητότητας ή αν υπάρχει μια μη τυχαία επαναλαμβανόμενη απόκλιση μεταξύ των "προιόντων", τότε έχουμε λόγους να υποπτευθούμε ότι υφίστανται ειδικές αιτίες που επιδρούν στη διεργασία.

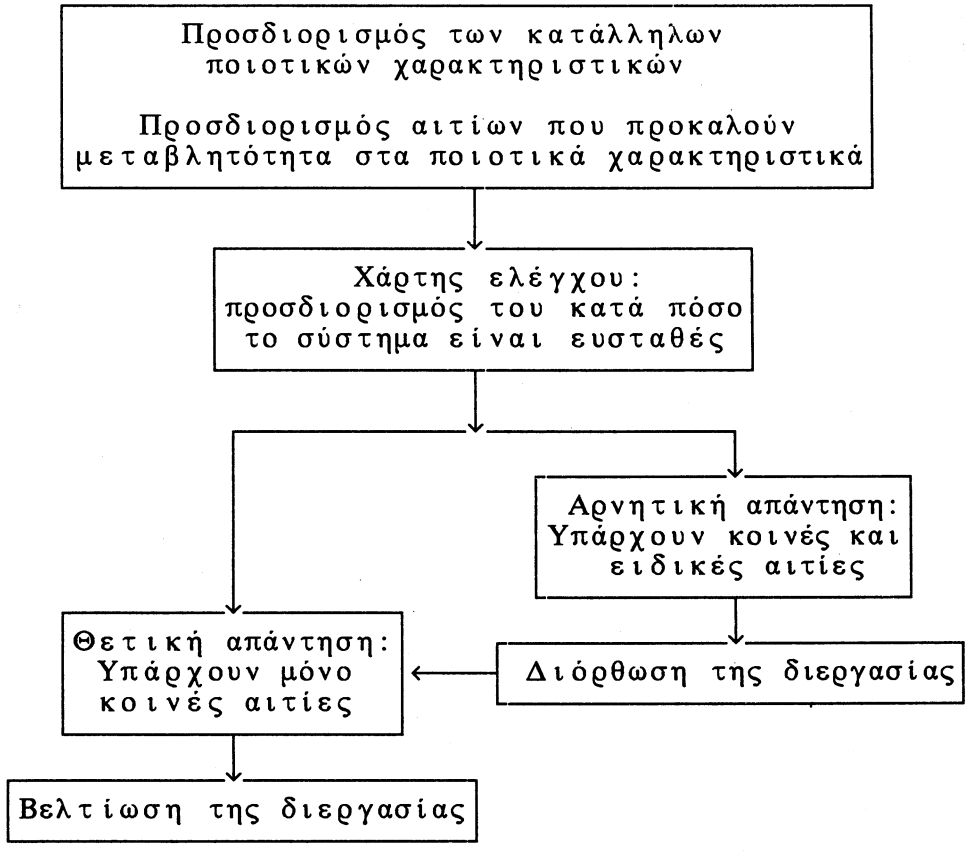
Η διάκριση μεταξύ ευσταθούς και ασταθούς διεργασίας είναι σημαντική δοθέντος ότι, προκειμένου να υπάρξει βελτίωση απαιτείται διαφορετικός τρόπος δράσης σε κάθε μια από αυτές. Η καλύτερη ευκαιρία για να βελτιωθεί ένα ασταθές σύστημα είναι να προσδιοριστούν οι ειδικές αιτίες μεταβλητότητας και να απομακρυνθούν. Να μετατραπεί δηλαδή μια ασταθής διεργασία σε ευσταθή διεργασία. Μπορούμε να θεωρήσουμε την ενέργεια αυτή ως *διόρθωση της διεργασίας*. Κάτι τέτοιο αποτελεί το πρώτο βήμα σε μια μακροπρόθεσμη διαδικασία βελτίωσης. Αν μια διεργασία είναι ασταθής (επηρεάζεται, δηλαδή, από πολλές ειδικές αιτίες μεταβλητότητας) είναι δύσκολο να προσδιοριστεί η επίδραση των όποιων μεταβολών επιφέρουμε στη διεργασία. Εάν αποφασίζαμε να μεταβάλλουμε τη διεργασία και στη συνέχεια παρατηρούσαμε μια μεταβολή στο εύρος των "προιόντων" πώς θα ήταν δυνατόν να ξέρουμε εάν οι μεταβολές στα "προιόντα" οφείλονται στις μεταβολές της διεργασίας που επιφέραμε ή σε κάποια άλλη ειδική αιτία; Είναι, επομένως, ιδιαίτερα χρήσιμο να μεταβάλλουμε μια διεργασία σε ευσταθή πριν επιχειρήσουμε οποιαδήποτε βελτίωσή της. Το πιο χρήσιμο και σύνηθες εργαλείο για να εκτιμηθεί κατά πόσο μια διεργασία είναι ευσταθής ή ασταθής (όσο αφορά ένα συγκεκριμένο χαρακτηριστικό ποιότητας είναι ο **χάρτης ελέγχου (control chart)**.

Είναι φανερό, από όσα προαναφέρθηκαν, ότι για να βελτιώσει κανείς ένα σύστημα θα πρέπει πρώτα να προσδιορίσει τις ειδικές αιτίες μεταβλητότητας. Αυτοί που συμμετέχουν στη διεργασία είναι ίσως οι καλύτεροι για να κάνουν κάτι τέτοιο. Από το άλλο μέρος

όμως, οι συμμετέχοντες σε μια διεργασία είναι λιγότερο πιθανό να έχουν τη δυνατότητα να προσδιορίσουν τρόπους βελτίωσης μιας ευσταθούς διεργασίας, εν μέρει διότι η εμφάνιση ενός "προιόντος" που πλησιάζει τα όρια της κανονικής μεταβλητότητας δεν μπορεί να εξηγηθεί από μια μόνο αιτία. Οταν η διεργασία είναι ευσταθής, βελτιώσεις μπορούν να επιτευχθούν μόνο με μεταβολές σ'αυτή καθεαυτή τη διεργασία. Μηχανικοί και managers είναι εκείνοι που μπορούν να κάνουν καλύτερα κάτι τέτοιο. Αυτό γιατί είναι σε καλύτερη θέση να αντιληφθούν τη διεργασία στην ολότητά της και είναι περισσότερο πιθανό να έχουν εκπαιδευτεί σε περισσότερο πολύπλοκες στατιστικές μεθόδους που απαιτούνται για κάτι τέτοιο. Με το να ελέγξουν τη διεργασία κάτω από διαφορετικές δοκιμαστικές συνθήκες, οι managers μπορεί να καταλήξουν σε αξιόπιστες ενδείξεις που μπορούν να οδηγήσουν σε βελτιώσεις της διεργασίας που είναι δυνατόν να διατηρηθούν.

Σε πολλές επιχειρήσεις οι βασικές αρχές του management διεργασιών που προαναφέρθηκαν δεν τηρούνται. Ενα σύνηθες λάθος στο management μιας διεργασίας είναι η **άκαιρη παρέμβαση (tampering)**. Το να παρεμβαίνει, δηλαδή, κανείς αντιδρώντας σε κοινές αιτίες μεταβλητότητας, ως εάν έχουν προέλθει από μια ειδική αιτία. Εάν προσαρμόσουμε μια διεργασία αντιδρώντας σε μια κοινή αιτία μεταβλητότητας, το μόνο που μπορεί να επιτύχουμε είναι να αυξήσουμε τη μεταβλητότητα.

Συμπερασματικά μπορεί να λεχθεί ότι η απομάκρυνση των ειδικών αιτιών συχνά οδηγεί σε ταχείες βελτιώσεις. Μεγαλύτερη όμως βελτίωση μακροπρόθεσμα επιτυγχάνεται με την μείωση της μεταβλητότητας που οφείλεται σε κοινές αιτίες μέσω μεταβολών που επιφέρονται σ'αυτή καθεαυτή τη διεργασία. Στο σχήμα 1.8.3 που ακολουθεί απεικονίζονται οι βελτιώσεις μιας διεργασίας όπως συζητήθηκαν στην ενότητα αυτή.



Σχήμα 1.8.3

Παράδειγμα: Κάθε ένα από τα παραδείγματα που ακολουθούν περιγράφει καταστάσεις στις οποίες διαπιστώνεται απουσία στατιστικής σκέψης. Να συζητηθεί η πρακτική που ακολουθεί κάθε manager με βάση τις αρχές του management διεργασιών που συζητήθηκαν.

α) Κάθε πρωί ο Διευθυντής Παραγωγής και οι συνεργάτες του σε μια εταιρεία εξετάζουν τα ελαττωματικά αντικείμενα που παρήχθησαν την προηγούμενη μέρα. Στόχος τους είναι να προσδιορίσουν και να

απομακρύνουν τις αιτίες του προβλήματος.

β) Ενας οικονομικός αναλυτής ισχυρίσθηκε τον Δεκέμβριο του 1973 ότι η κατάσταση του εξωτερικού εμπορίου βελτιώθηκε. Την άποψη αυτή την στήριξε στο γεγονός ότι, από τα στοιχεία της ΕΣΥΕ προέκυψε ότι το έλλειμμα του εμπορικού ισοζυγίου μειώθηκε τον Νοεμβρίου του 1973 κατά 1612080 χιλ. δρχ. σε σχέση με το έλλειμμα του Οκτωβρίου του ίδιου έτους. (Τον Οκτώβριο του 1973 οι εισαγωγές ήταν 10315710 χιλ. δρχ. και οι εξαγωγές 3751562 χιλ. δρχ. Αντίστοιχα, τον Νοέμβριο του ίδιου έτους οι εισαγωγές ήταν 8676203 ενώ οι εξαγωγές ήταν 3724135 χιλ. δρχ.).

γ) Ο Διευθυντής Πωλήσεων μιας εταιρείας αποφασίζει να προχωρήσει σε ειδικές εκπτώσεις βλέποντας ότι οι πωλήσεις του βρίσκονται σε χαμηλότερο επίπεδο από αυτό που είχε προγραμματιστεί.

Λύση :

α) Ο manager υποθέτει ότι κάθε ελαττωματικό αντικείμενο αντικατοπτρίζει μια ειδική αιτία μεταβλητότητας. Παρ'όλα αυτά είναι φυσιολογικό ότι μια διαδικασία παραγωγής, όταν είναι ευσταθής, παράγει μερικά ελαττωματικά αντικείμενα. Προσπάθεια προσδιορισμού μιας ειδικής αιτίας για κάθε ελαττωματικό αντικείμενο θα ήταν μάταιη και θα αποτελούσε σπατάλη χρόνου για το διευθυντή, εκτός αν είχε προηγουμένως διαπιστωθεί ότι είχε παραχθεί ένας ασυνήθιστα υψηλός αριθμός ελαττωματικών αντικειμένων. Η επιλογή αυτής της διαδικασίας είναι δυνατόν να οδηγήσει σε αύξηση της μεταβλητότητας της διαδικασίας.

β) Το έλλειμμα του Εμπορικού Ισοζυγίου είναι φυσικό να μεταβάλλεται από μήνα σε μήνα. Μόνο εάν η μεταβολή του τελευταίου μήνα από τον προηγούμενο ήταν πολύ μεγαλύτερη από αντίστοιχες

μεταβολές του παρελθόντος θα μπορούσε κανείς να οδηγηθεί στο συμπέρασμα αυτό. Κυρίως όμως αν είχε υπάρξει ουσιαστική βελτίωση σε ετήσια βάση.

γ) Ο Διευθυντής Πωλήσεων θα πρέπει πρώτα να καθορίσει αν η διεργασία είναι ευσταθής. Διαφορετικά οι ειδικές εκπτώσεις θα μπορούσαν να αποτελέσουν άκαιρη παρέμβαση και ίσως κάνουν περισσότερο κακό παρά καλό.

1.9 Χάρτες ή διαγράμματα ροής (run charts)

Ο χάρτης ροής (run chart) είναι μια γραφική παράσταση (plot) των τιμών των δεδομένων με τη σειρά με την οποία έχουν συλλεγεί.

Οταν έχουμε μια χρονολογική σειρά στοιχείων το διάγραμμα ροής είναι ένα απλό και πολύ χρήσιμο εργαλείο γαι τον προσδιορισμό της μεταβλητότητας.

1.10 Εισαγωγή στο Σχεδιασμό και Ανάλυση Πειραμάτων

Μια από τις κυριότερες εφαρμογές της Στατιστικής είναι η παροχή μεθόδων και πληροφοριών που μπορούν να χρησιμοποιηθούν για τη βελτίωση καταστάσεων. Ο **Στατιστικός Σχεδιασμός και Ανάλυση Πειραμάτων** είναι μια ισχυρότατη μεθοδολογία που επιτρέπει να αντιληφθούμε με ποιο τρόπο διαφορές μεταξύ διαφόρων παραγόντων είναι ενδεχόμενο να επηρεάσουν μια ποσότητα που μας ενδιαφέρει. Η κατανόηση αυτής της μορφής οδηγεί σε βελτίωση. Για παράδειγμα, αν ο Διευθυντής πωλήσεων μιας εταιρείας διαπιστώσει σημαντικές διαφορές στην ποσότητα πωλήσεων μεταξύ των πωλητών, ίσως αποφασίσει να χρησιμοποιήσει μια πρόσθετη εκπαίδευση ώστε να μειώσει τις διαφορές.

Στις μελέτες διεργασιών ο Στατιστικός Σχεδιασμός Πειραμάτων είναι εξαιρετικά χρήσιμος στον προσδιορισμό παραγόντων που συνεισφέρουν στη μεταβλητότητα μιας ευσταθούς διεργασίας και στο κατά πόσο υπάρχουν σημαντικές διαφορές μεταξύ τους. Για παράδειγμα, εάν υπάρχουν διαφορές μεταξύ μηχανών που χρησιμοποιούνται για την εμφιάλωση ενός προϊόντος, μια βελτίωση της πολιτικής συντήρησης είναι ενδεχόμενο να ελαττώσει τις διαφορές και, επομένως, να οδηγήσει σε βελτίωση της διαδικασίας. Εκείνο που, κατ'αρχήν, κάνει κανείς χρησιμοποιώντας διαγράμματα ροής και ελέγχου είναι να διαπιστώσει εάν η διεργασία είναι ευσταθής. Παρ'όλα αυτά το γεγονός ότι μια διαδικασία είναι ευσταθής από μόνο του δεν συνεπάγεται ότι είναι και ικανοποιητική. Μια από τις πιο ενδιαφέρουσες πλευρές βελτίωσης μιας διαδικασίας είναι να βρει κανείς τρόπους να μεταβάλει μια ευσταθή διαδικασία έτσι ώστε να ελαττώσει τη μεταβλητότητα των "προιόντων". Ενα απαραίτητο πρώτο βήμα για τη διοίκηση είναι να αντιληφθεί ότι έχει την ευθύνη για συνεχή βελτίωση. Το δεύτερο βήμα είναι να χρησιμοποιηθούν οι γνώσεις για το αντικείμενο - εμπειρία και

θεωρητική γνώση που αναφέρεται στην υπό μελέτη διεργασία - ώστε να προταθούν μεταβολές που ίσως οδηγήσουν σε βελτίωση της διεργασίας ή στον προσδιορισμό πιθανών αιτίων μεταβλητότητας στο εσωτερικό της διεργασίας. Το τρίτο βήμα στη βελτίωση μιας ευσταθούς διεργασίας είναι ο σχεδιασμός ενός πειράματος ώστε να ελεγχθεί το αποτέλεσμα μιας προτεινόμενης μεταβολής ή να ελεγχθεί κάποια ενδεχόμενη αιτία μεταβλητότητας.

Στο παρελθόν η δραστηριότητα αυτή, που έχει αποδειχθεί σήμερα ιδιαίτερα αποτελεσματική, αγνοείτο στις δυτικές βιομηχανίες (σε αντίθεση με την Ιαπωνία). Πρόσφατα όμως διευθυντικά στελέχη εταιρειών έχουν αρχίσει να χρησιμοποιούν σχεδιασμένα πειράματα πιο συστηματικά.

Στη συνέχεια θα δώσουμε τις βασικές αρχές που προσδιορίζουν και εξηγούν τον Στατιστικό Σχεδιασμό Πειραμάτων. Το θέμα αυτό αποτελεί ένα ολόκληρο αντικείμενο της Στατιστικής. Προκειμένου να δώσουμε τις βασικές αρχές θα χρησιμοποιήσουμε ένα παράδειγμα και στη συνέχεια θα διατυπώσουμε τις βασικές αρχές του πειραματικού σχεδιασμού.

Παράδειγμα : Εστω ότι ένας καθηγητής του τμήματος Στατιστικής ενδιαφέρεται να βρει τρόπους ώστε να βελτιώσει την απόδοση των φοιτητών του στο μάθημα το οποίο διδάσκει. Η εμπειρία του τον έχει οδηγήσει στο συμπέρασμα ότι ο τελικός βαθμός των φοιτητών επηρεάζεται από τον αριθμό των ασκήσεων που οι φοιτητές λύνουν κατά τη διάρκεια του εξαμήνου. Η εμπειρία του, από τους φοιτητές του, έχει δείξει ότι ένας μέσος φοιτητής προσπαθεί να λύσει περίπου τις μισές από τις ασκήσεις που δίνονται στα μαθήματα. Ο καθηγητής αυτός πιστεύει ότι αν οι φοιτητές προσπαθούσαν να λύσουν περισσότερες ασκήσεις θα απέδιδαν καλύτερα και στο τελικό διαγώνισμα. Προκειμένου να ελέγξει τη διαίσθησή του αυτή, ο καθηγητής οργανώνει το εξής πείραμα στην τάξη στην οποία διδάσκει. Οι 24 φοιτητές και φοιτήτριες που παρακολουθούν το μάθημα

συμφωνούν να πάρουν μέρος στο πείραμα αυτό ως ένα τρόπο κατανόησης της μεθόδου του Πειραματικού Σχεδιασμού. Το πείραμα αυτό γίνεται ως εξής:

Η τάξη διαιρείται σε τρεις "ομάδες μελέτης", κάθε μια από τις οποίες αποτελείται από οκτώ φοιτητές. Η μια από τις ομάδες συμφωνεί να προσπαθήσει να λύσει τις μισές από τις ασκήσεις που θα δοθούν από τον καθηγητή κατά τη διάρκεια του εξαμήνου. (Μια άσκηση θεωρείται ότι έχει λυθεί σωστά όταν ο φοιτητής έχει βρει το σωστό αποτέλεσμα και πείθει με την επιχειρηματολογία του ότι έχει καταλάβει καλά τη λύση). Η δεύτερη ομάδα δέχεται να προσπαθήσει να λύσει το 75% των ασκήσεων, ενώ η τρίτη συμφωνεί να προσπαθήσει να λύσει το 100% των ασκήσεων. Το ποιοτικό χαρακτηριστικό του προβλήματος αυτού είναι ο τελικός βαθμός που θα δοθεί από τον καθηγητή. Δοθέντος ότι οι καλοί φοιτητές τείνουν να πάρουν μεγαλύτερους βαθμούς στο τελικό διαγώνισμα από ότι οι λιγότερο καλοί φοιτητές, ο καθηγητής αποφασίζει να χωρίσει τους φοιτητές, στο συγκεκριμένο μάθημα, σε δύο ομάδες σύμφωνα με τον μέχρι τη στιγμή εκείνη μέσο όρο των βαθμών τους στα μαθήματα στα οποία έχουν εξετασθεί με επιτυχία. Οι 12 φοιτητές που έχουν μέσο όρο μικρότερο από 6 αποτελούν τη μια ομάδα, ενώ οι υπόλοιποι 12 που έχουν μέσο όρο μεγαλύτερο από 6 αποτελούν μια άλλη ομάδα. Οι 12 φοιτητές από κάθε μία από τις δύο αυτές ομάδες τοποθετούνται με τυχαίο τρόπο στις τρεις ομάδες μελέτης που προαναφέραμε. Επομένως, κάθε ομάδα μελέτης έχει μια ισόρροπη εκπροσώπιση των φοιτητών ανάλογα με το μέσο όρο της βαθμολογίας τους. (Δηλαδή φοιτητές από κάθε μια από τις δύο ομάδες που έχουν σχηματιστεί με βάση το μέσο όρο βαθμολογίας).

Ο πίνακας 1.10.1 δείχνει τον τρόπο με τον οποίο οι φοιτητές τοποθετήθηκαν στις διάφορες ομάδες. (Ο αριθμός αντιπροσωπεύει τον αύξοντα αριθμό του φοιτητή στην κατάσταση).

Πίνακας 1.10.1

Μέσος όρος ≤ 6

Ονομα	Ομάδα εργασίας που τοποθετήθηκε
1	50%
2	75%
3	100%
4	50%
5	100%
6	50%
7	75%
8	100%
9	75%
10	75%
11	50%
12	100%

Μέσος όρος > 6

Ονομα	Ομάδα εργασίας που τοποθετήθηκε
13	100%
14	50%
15	50%
16	75%
17	100%
18	75%
19	50%
20	75%
21	50%
22	75%
23	100%
24	100%

Ποσοστό ασκήσεων που λύθηκαν

50 %

1, 4, 6, 11
14, 15, 19, 21

75 %

2, 7, 9, 10
16, 18, 20, 22

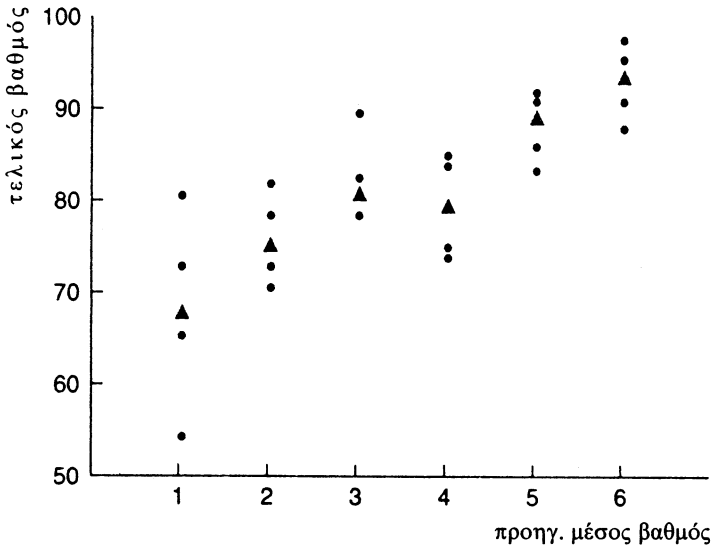
100 %

3, 5, 8, 12
13, 17, 23, 24

Η βαθμολογία στο τελικό διαγώνισμα του μαθήματος αυτού δίνεται στον πίνακα 1.10.2. Η γραφική παράσταση δίνεται στο σχήμα 1.10.1. Οι απλές κουκίδες αποτελούν ένδειξη του βαθμού, ενώ οι κουκίδες με τον κύκλο (ή τρίγωνο) αντιπροσωπεύουν μέσους όρους. Το στατιστικό ερώτημα που τίθεται με βάση τα στοιχεία αυτά είναι αν είμαστε σε θέση να συμπεράνουμε ότι ο αριθμός των ασκήσεων που λύνουν οι φοιτητές επηρεάζει τον τελικό βαθμό στις εξετάσεις.

Πίνακας 1.10.2

Προηγούμενος μέσος βαθμός	Ποσοστό ασκήσεων	Τελικός βαθμός στο μάθημα				Μέσος βαθμός ομάδας
≤ 6	50%	54	66	73	81	68.5
	75%	71	73	78	82	76.0
	100%	78	78	84	89	82.85
> 6	50%	74	75	84	85	79.5
	75%	83	87	96	94	89.25
	100%	88	93	96	98	93.75



Σχήμα 1.10.1
Γραφική παράσταση των αποτελεσμάτων

Προκειμένου να αναλύσουμε τα αποτελέσματα του πειράματος αυτού, ας παρατηρήσουμε για λίγο τα δεδομένα. Για την ομάδα των φοιτητών με μέσο όρο μικρότερο ή ίσο από 6 παρατηρούμε ότι καθώς το ποσοστό των ασκήσεων που λύθηκαν αυξάνεται από 50% σε 75% και σε 100%, η μέση τελική βαθμολογία μεταβάλλεται από 68.5 σε 76.0 και 82.25. Για την άλλη ομάδα των φοιτητών με μέση βαθμολογία μεγαλύτερη του 6 η μέση τελική βαθμολογία κλιμακώνεται από 79.5 σε 89.25 και 93.75 καθώς αυξάνεται το ποσοστό των ασκήσεων που έλυσαν. Επομένως, και για τις δύο ομάδες των φοιτητών, όπως αυτές χωρίστηκαν ανάλογα με το μέσο όρο της βαθμολογίας των μαθημάτων που έχουν ήδη εξετασθεί επιτυχώς, παρατηρούμε ότι η μέση τελική βαθμολογία αυξάνεται σημαντικά όσο αυξάνεται ο αριθμός των ασκήσεων που έλυσαν. Τα στοιχεία αυτά φαίνεται να οδηγούν στο συμπέρασμα ότι η τελική βαθμολογία των εξετάσεων βελτιώνεται ανάλογα με τον αριθμό των ασκήσεων που έχουν λυθεί.

Θα επιθυμούσαμε ακόμα, ενδεχομένως, να εξετάσουμε εάν ο μέσος όρος της βαθμολογίας στα ήδη εξετασθέντα μαθήματα αποτελεί ένα καλό στοιχείο για να χρησιμοποιηθεί ως πρόβλεψη για τον τελικό

βαθμό του διαγωνίσματος. Παρατηρούμε ότι εάν εξετάσουμε οποιεσδήποτε δύο υποομάδες που συμπλήρωσαν τον ίδιο αριθμό ασκήσεων, θα διαπιστώσουμε ότι η ομάδα με μέση βαθμολογία μεγαλύτερη από 6 έχει σταθερά τον μεγαλύτερο μέσο όρο τελικής βαθμολογίας. Εχουμε 68.5 έναντι 79.5 για την ομάδα του 50%, 76.0 έναντι 89.25 για την ομάδα του 75% και, τέλος, 82.25 έναντι 93.75 για την ομάδα του 100% των ασκήσεων. Και στις τρεις αυτές περιπτώσεις το μέσο τελικό σκορ για αυτούς που είχαν μέση προηγούμενη βαθμολογία μεγαλύτερη από 6 ήταν περίπου 12 μονάδες υψηλότερο για αυτούς που είχαν μέση προηγούμενη βαθμολογία μικρότερη ή ίση από το 6.

Εχοντας εξετάσει το απλό αυτό παράδειγμα πειραματικού σχεδιασμού, ας δούμε τώρα την ορολογία που χρησιμοποιείται και τους σημαντικότερους όρους της περιοχής αυτής της Στατιστικής.

Πειραματικές μονάδες (experimental units) είναι στοιχεία του πειράματος για τα οποία θα ορισθούν μεταβλητές πάνω στις οποίες θα έχουμε παρατηρήσεις. Στο παράδειγμά μας πειραματικές μονάδες είναι οι φοιτητές στην τάξη.

Απαντητική μεταβλητή (response variable) είναι μια στατιστική μεταβλητή που εκφράζει το αποτέλεσμα του πειράματος για κάποια πειραματική μονάδα. Συχνά η μεταβλητή αυτή είναι ένα ποιοτικό χαρακτηριστικό της διαδικασίας. Στο παράδειγμά μας απαντητική μεταβλητή είναι η βαθμολογία στο τελικό διαγώνισμα.

Παράγοντας (factor) είναι μια μεταβλητή την οποία εκ προθέσεως ελέγχουμε ώστε να παρατηρήσουμε την επίδρασή της στην απαντητική μεταβλητή. Ενας παράγοντας μπορεί να είναι μια ταξινόμηση (classification), π.χ. το φύλο, που ίσως επηρεάζει κατά κάποιο τρόπο την απαντητική μεταβλητή.

Επίπεδα (levels) είναι οι τιμές οι οποίες προκαθορίζονται για κάποιον παράγοντα.

Στο παράδειγμά μας ο παράγοντας που μας ενδιέφερε ήταν το ποσοστό των ασκήσεων που συμπλήρωσαν οι φοιτητές. Τα επίπεδα στα οποία αυτός ελέγχθηκε ήταν τρία: 50%, 75% και 100%.

Σε πολλά προβλήματα υπάρχουν πολλές ακόμα μεταβλητές εκτός από τους παράγοντες που μας ενδιαφέρουν που είναι ενδεχόμενο να προκαλέσουν μεταβλητότητα στην απαντητική μεταβλητή. Οι μεταβλητές αυτές ονομάζονται **λανθάνουσες** ή **αφανείς** ή **περιθωριακές μεταβλητές (background variables** ή **blocking variables)**. Κάθε επίπεδο μιας περιθωριακής μεταβλητής ονομάζεται **block**. Οταν σε ένα πρόβλημα υπάρχουν περιθωριακές μεταβλητές οι παράγοντες που μας ενδιαφέρουν συνήθως μελετώνται για κάθε block της περιθωριακής μεταβλητής.

Οταν σχεδιάζει κανείς ένα πείραμα θα πρέπει να εξετάζει όλες τις περιθωριακές μεταβλητές, τόσο τις σημαντικές όσο και αυτές που θεωρεί τετριμμένες. Στο παράδειγμά μας περιθωριακές μεταβλητές θα μπορούσαν να θεωρηθούν η ακαδημαική ικανότητα των φοιτητών (όπως αυτή εκφράζεται με το μέσο όρο βαθμολογίας στα προηγούμενα μαθήματα), η ψυχολογική κατάσταση στην οποία ο φοιτητής βρίσκεται σε μια συγκεκριμένη μέρα, η δυσκολία των θεμάτων των εξετάσεων, ο καθηγητής που εξετάζει το μάθημα και η προσπάθεια που κάνουν οι φοιτητές για να προετοιμαστούν για το τελικό διαγώνισμα (πέρα από τις ασκήσεις που έλυσαν κατά τη διάρκεια του χρόνου).

Κατά τη διάρκεια της εκτέλεσης ενός πειράματος προσπαθούμε κατά το δυνατόν να ελέγξουμε τη μεταβλητότητα των λανθανουσών (αφανών) μεταβλητών. Κάνοντας κάτι τέτοιο έχουμε τη δυνατότητα να συγκρίνουμε τα αποτελέσματα των παραγόντων που μας ενδιαφέρουν κάτω από τις ίδιες, όσο το δυνατόν, συνθήκες.

Υπάρχουν τρεις, κυρίως, τρόποι ελέγχου μιας λανθάνουσας μεταβλητής.

1. *Απομάκρυνσή της ως μεταβλητής.* Ενας τρόπος για να επιτευχθεί αυτό είναι να διατηρηθεί η λανθάνουσα μεταβλητή σε μία συγκεκριμένη τιμή καθ'όλη τη διάρκεια του πειράματος έτσι ώστε να μην είναι δυνατόν να προκαλέσει μεταβλητότητα στην απαντητική μεταβλητή. Στο παράδειγμά μας επιτύχαμε να ελέγξουμε τη δυσκολία του τελικού διαγωνίσματος και τον καθηγητή. (Είχαμε μόνο ένα τελικό διαγώνισμα για όλους και έναν καθηγητή για όλους). Ενα μειονέκτημα της προσέγγισης αυτής είναι ότι η συμπερασματολογία που προκύπτει από το πείραμα περιορίζεται για τις συνθήκες κάτω από τις οποίες έγινε το πείραμα. Ετσι τα συμπεράσματά μας για το αποτέλεσμα που είχε ο αριθμός των ασκήσεων που λύθηκαν στον τελικό βαθμό έχουν στατιστική αξία μόνο για τό συγκεκριμένο μάθημα και για τον συγκεκριμένο καθηγητή. Προκειμένου να ισχυρισθούμε στατιστικά ότι τα συμπεράσματα ισχύουν και για άλλα μαθήματα και για άλλους καθηγητές, το πείραμα θα πρέπει να επεκταθεί ώστε να καλύπτει και αυτές τις συνθήκες.

2. *Ελεγχος των επιπέδων της λανθάνουσας μεταβλητής.* Τα επίπεδα μιας λανθάνουσας μεταβλητής θα πρέπει να καλύπτουν όλο το εύρος των τιμών που είναι δυνατό να εμφανιστούν κατά την διεργασία. Στο παράδειγμά μας ο μέσος όρος βαθμολογίας των φοιτητών στα προηγούμενα μαθήματα καλύπτει όλους τους φοιτητές. Ο μέσος αυτός όρος ελέγχθηκε σε δύο επίπεδα (από 0 έως 6 και από 6 και πάνω).

3. *Καταγραφή των τιμών της λανθάνουσας μεταβλητής.* Υπάρχουν περιπτώσεις που δεν είναι δυνατόν να ελέγξουμε τα επίπεδα μιας μεταβλητής σε ένα πείραμα. Στην περίπτωση αυτή θα πρέπει να προσπαθήσουμε να καταγράψουμε τις τιμές της. Κάτι τέτοιο ίσως μας βοηθήσει να καταλάβουμε την επίδρασή της στην απαντητική μεταβλητή όταν αναλύουμε τα πειραματικά αποτελέσματα.

Εάν δεν μπορέσουμε να ελέγξουμε μια αφανή μεταβλητή με ένα από τους τρεις τρόπους που προαναφέρθηκαν, υπάρχει κίνδυνος η μεταβλητή αυτή να επηρεάσει το αποτέλεσμα του πειράματος με συστηματικό τρόπο. Αυτό μπορεί να έχει ως συνέπεια να αποδώσουμε την επίδρασή της, λανθασμένα, σε έναν από τους παράγοντες του πειράματος. Προκειμένου να αποφύγουμε ένα τέτοιο ενδεχόμενο τοποθετούμε τις πειραματικές μονάδες στα υποσύνολα με τυχαίο τρόπο αφού ήδη έχουν καθορισθεί τα επίπεδα των παραγόντων και οι αφανείς μεταβλητές.

Αυτή η πλευρά του Πειραματικού Σχεδιασμού ονομάζεται **τυχαιοποίηση (randomization)**. Στο παράδειγμά μας οι φοιτητές, που ανήκαν σε μια από τις δύο κατηγορίες με βάση τη μέση βαθμολογία σε μαθήματα στα οποία είχαν εξετασθεί προηγουμένως, τοποθετήθηκαν τυχαία στις τρεις ομάδες μελέτης (50%, 75%, 100%). Αυτό έχει ως αποτέλεσμα το ότι οι επιδράσεις όλων των μη ελεγχόμενων μεταβλητών κατανέμονται τυχαία και χωρίς μεροληπτικότητα στις υποομάδες του πειράματος. Η χρησιμοποίηση της τυχαιοποίησης στο παράδειγμα αυτό έχει ως αποτέλεσμα ότι η από μέρα σε μέρα μεταβλητότητα της απόδοσης των φοιτητών που οφείλεται στην προσωπικότητά τους όπως και η μεταβλητότητα που οφείλεται στο διαφορετικό χρόνο προετοιμασίας για το τελικό διαγώνισμα κατανέμονται τυχαία στις έξι υποομάδες. Ετσι, παρά το γεγονός ότι οι μεταβλητές αυτές δεν είναι δυνατό να ελεγχθούν, η τυχαιοποίηση τις εμποδίζει από το να επηρεάσουν *συστηματικά* μια υποομάδα σε βάρος μιας άλλης.

Μια λανθάνουσα μεταβλητή η οποία συχνά αγνοείται είναι η διαδικασία μέτρησης. Είναι σημαντικό να καθορισθεί, με τη μεγαλύτερη δυνατή ακρίβεια ο τρόπος με τον οποίο κάθε μια από τις μεταβλητές θα μετρηθεί και έτσι να ελαχιστοποιηθεί η μεταβλητότητα που θα οφείλεται σε ασυνέπειες του τρόπου μέτρησης. Οι κύριες δραστηριότητες για το σχεδιασμό ενός πειράματος είναι οι εξής:

Κύρια σημεία για το σχεδιασμό ενός πειράματος

1. Περιγραφή της διεργασίας που ενδιαφερόμαστε να μελετήσουμε με τον τρόπο που λειτουργεί τη στιγμή κατά την οποία γίνεται η μελέτη.
2. Καθορισμός των κυρίων ερωτήσεων που θα πρέπει να απαντηθούν με το πείραμα.
3. Καθορισμός των πειραματικών μονάδων.
4. Προσδιορισμός την απαντητικής μεταβλητής.
5. Προσδιορισμός των παραγόντων και των επιπέδων τους.
6. Προσδιορισμός των αφανών μεταβλητών.

(∘) Προσδιορισμός των μεταβλητών αυτών οι οποίες είναι δυνατόν να ελεγχθούν και της μεθόδου με την οποία θα ελεγχθούν.

(∘) Προσδιορισμός των μη ελεγχόμενων αφανών μεταβλητών των οποίων θα καταγράψουμε τις τιμές.

7. Καθορισμός της τεχνικής με την οποία θα γίνει η μέτρηση για την απαντητική μεταβλητή κάθε παράγοντα και κάθε λανθάνουσας μεταβλητής.
8. Χρήση τυχαιοποίησης ώστε να τοποθετηθούν οι πειραματικές μονάδες στις πειραματικές υποομάδες.

1.11 Χρήση των Υπολογιστών στη Στατιστική Ανάλυση

Η χρήση της Στατιστικής Ανάλυσης έχει επεκταθεί εκπληκτικά και έχει γίνει ευρύτερα διαθέσιμη κυρίως λόγω της ανάπτυξης των στατιστικών πακέτων για υπολογιστές. Πριν οι υπολογιστές διαδοθούν τόσο πολύ οι στατιστικές μέθοδοι απαιτούσαν επίπονη εργασία για πολύπλοκους υπολογισμούς. Αυτό καθιστούσε δύσκολη την χρησιμοποίηση στατιστικών μεθόδων πέρα από τις απλούστερες από αυτές που ήταν διαθέσιμες. Ηταν, δηλαδή, περισσότερο σημαντικό για μια μέθοδο, προκειμένου να χρησιμοποιηθεί, να έχει εύκολους υπολογισμούς παρά να είναι η πιο κατάλληλη. Εστω και αν ακόμα η

μέθοδος που εχρησιμοποιείτο βασιζόταν σε υποθέσεις που δεν ανταποκρίνονταν στην υπό μελέτη κατάσταση, οι μελετητές έτειναν να αγνοήσουν το γεγονός αυτό. Οι υπολογισμοί γίνονταν από υπαλλήλους και έτσι οι υπεύθυνοι (managers) δεν ασχολούνταν με τη Στατιστική Ανάλυση και δεν είχαν, για το λόγο αυτό, την ευκαιρία να αποκτήσουν οικειότητα με τα σχετικά δεδομένα.

Σήμερα η κατάσταση αυτή έχει αλλάξει. Υπάρχουν πάρα πολλά εύχρηστα στατιστικά πακέτα, τόσο για μεγάλους υπολογιστές όσο και για μικρουπολογιστές. Τέτοια πακέτα είναι το SAS, το SPSS, το BMDP, το MINITAB, το STATGRAPHICS κ.λ.π. Τα πακέτα αυτά δίνουν στο χρήστη πρόσβαση σε οποιαδήποτε στατιστική μέθοδο που ένας επαγγελματίας στατιστικός θα ήθελε να εφαρμόσει.

Παρ'όλα αυτά θα πρέπει να τονισθεί ότι η μεγάλη προσφορά και η χρήση στατιστικών πακέτων δεν συνεπάγεται υποχρεωτικά σωστή Στατιστική. Είναι, βέβαια, γεγονός ότι οι υπολογιστές κάνουν εύκολα διαθέσιμες τις στατιστικές μεθόδους, αλλά δεν παύει να υπάρχει ανάγκη επιλογής της κατάλληλης στατιστικής μεθόδου για κάθε συγκεκριμένο πρόβλημα, για τον έλεγχο των υποθέσεων που πρέπει κανείς να κάνει, όπως επίσης και για τις πιθανές τροποποιήσεις της μεθόδου που απαιτούνται για ένα συγκεκριμένο πρόβλημα. Πολλοί ισχυρίζονται ότι με το να κάνουν οι υπολογιστές τη Στατιστική τόσο εύκολα διαθέσιμη στο ευρύ κοινό την έκαναν επίσης και επικίνδυνη. Επομένως, περισσότερο από ποτέ άλλοτε, είναι σημαντικό για αυτούς που παίρνουν αποφάσεις χρησιμοποιώντας ποσοτικά δεδομένα να έχουν την ικανότητα της στατιστικής σκέψης. Ακόμα και ένας υπεύθυνος που δε θα κάνει ποτέ στατιστικούς υπολογισμούς και αναλύσεις είναι πολύ πιθανό να δεχθεί υποδείξεις από άλλους που βασίζονται σε στατιστική συμπερασματολογία. Μόνο εάν κανείς καταλαβαίνει και έχει αναπτύξει την βασική στατιστική σκέψη μπορεί να προφυλάσσεται από συνηθισμένες κακές χρήσεις των στατιστικών μεθόδων.

Ο βασικός σκοπός μας στο βιβλίο αυτό είναι η εξοικείωση με τη

στατιστική σκέψη και τις βασικές στατιστικές μεθόδους. Από το άλλο μέρος, είναι σημαντικό να χειρίζεται κανείς τα στατιστικά πακέτα που είναι διαθέσιμα και να ερμηνεύει τα αποτελέσματα που αυτά δίνουν. Για το λόγο αυτό, στην ανάπτυξη των θεμάτων χρησιμοποιούμε το στατιστικό πακέτο MINITAB (την έκδοση για μικρουπολογιστές).

Οπως θα μάθουμε να χρησιμοποιούμε τις εφαρμογές των στατιστικών πακέτων θα δούμε ότι τα πακέτα αυτά δίνουν πολύ περισσότερες πληροφορίες από αυτές που χρειαζόμαστε σε ένα συγκεκριμένο πρόβλημα ή που έχουμε την ικανότητα να ερμηνεύσουμε. Αυτό γιατί αυτοί που σχεδιάζουν τα στατιστικά πακέτα περιλαμβάνουν συνήθως κάθε πληροφορία ή δυνατότητα που είναι δυνατόν να χρειαστεί κάποιος χρήστης. Θα πρέπει, επομένως, να συνηθίσουμε να παίρνουμε τις πληροφορίες εκείνες που χρειαζόμαστε, χωρίς να ανησυχούμε για τις υπόλοιπες.

Το πακέτο MINITAB

Το πακέτο Minitab θεωρείται ως ένα από τα ευκολότερα στη χρήση τους στατιστικά πακέτα με πολλές δυνατότητες. Το Minitab μπορεί να χρησιμοποιηθεί τόσο σε μεγάλα συστήματα όσο και σε συστήματα μικρουπολογιστών. Μπορεί να χρησιμοποιηθεί σε IBM προσωπικούς υπολογιστές (PCs) (ή συμβατούς) αλλά και σε Apple Mackintosh. Εκτός από το κλασσικό πακέτο, το Minitab υπάρχει σήμερα διαθέσιμο και σε έκδοση για Windows. Η φιλοσοφία του Minitab είναι στο ότι αποκτά κανείς διέξοδο στη δομή των οδηγιών (command structure) μέσω του καταλόγου επιλογών (menu) με τον ίδιο τρόπο όπως μέσω πληκτρολογούμενων εντολών. Θα παρουσιάσουμε το πακέτο αυτό με τη χρήση ορισμένων από τις βασικές εντολές.

Πριν να έχουμε τη δυνατότητα να κάνουμε στατιστικούς υπολογισμούς με το Minitab θα πρέπει να εισάγουμε τα δεδομένα μας στον υπολογιστή. Οταν μπούμε στο subdirectory του Minitab σε ένα IBM συμβατό υπολογιστή και ξεκινήσουμε το πρόγραμμα, θα δούμε την *οθόνη εργασίας (session window)* με τη *λωρίδα βασικών επιλογών*

(main menu bar) στην κορυφή της οθόνης και την ένδειξη του υπολογιστή MTB> . Αυτή η οθόνη εργασίας (session window) χρησιμοποιείται για τη χρήση πληκτρολογούμενων εντολών και δεδομένων. Τα δεδομένα μπορούν επίσης να εισαχθούν απευθείας στην *οθόνη εργασίας δεδομένων (data worksheet)* στην οποία μπορούμε να έχουμε πρόσβαση με το να επιλέξουμε την *οθόνη δεδομένων (data screen)* από τις επιλογές εργασίας (edit menu). Η ένδειξη (prompt) MTB> αποτελεί την επίκληση για μια εντολή. Η επίκληση για δεδομένα (prompt for data) είναι DATA> . Μετά την ένδειξη MTB> το Minitab περιμένει μια εντολή, ενώ μετά την ένδειξη DATA> περιμένει την εισαγωγή στοιχείων.

Προκειμένου να εισαχθούν στοιχεία, δύο εντολές είναι δυνατόν να χρησιμοποιηθούν. Οι εντολές READ ή SET. Με οποιαδήποτε από τις εντολές αυτές το Minitab δημιουργεί δεδομένα σε τόσες στήλες (columns) που έχουν επικεφαλίδα C1, C2, ... (column1, column2,...) όσες διαφορετικές ομάδες παρατηρήσεων υπάρχουν στα στοιχεία μας.

Αφού πληκτρολογήσουμε μια εντολή ή μια σειρά δεδομένων πατάμε το πλήκτο με την ένδειξη ENTER. Είναι καλό πριν πατήσουμε το ENTER να ελέγξουμε τη γραμμή που πληκτρολογήσαμε για ενδεχόμενα λάθη. Εάν βρούμε κάποιο λάθος γυρίζουμε πίσω με το Backspace και κάνουμε τις όποιες διορθώσεις. Από τη στιγμή που θα πατήσουμε το πλήκτο με την ένδειξη ENTER, η πληροφορία που είχαμε στη γραμμή που πληκτρολογήσαμε θα εισαχθεί στη μνήμη του υπολογιστή και η οποιαδήποτε διόρθωση θα χρειασθεί κάποια άλλη διαδικασία. (Παρ'όλα αυτά, διορθώσεις σε δεδομένα γίνονται πιο εύκολα με την μεταφορά στην οθόνη των δεδομένων (data worksheet)).

Σημείωση: Οταν πληκτρολογούμε γράμματα και λέξεις δεν έχει σημασία στο Minitab εάν αυτό γίνεται με κεφαλαία ή μικρά γράμματα.

Η εντολή READ

Ας υποθέσουμε ότι θέλουμε να εισάγουμε τις παρατηρήσεις 71, 72, 68, 66 και 72 που αντιπροσωπεύουν το βάρος σε κιλά πέντε

ανθρώπων. Θα πρέπει να δώσουμε τις εξής εντολές:

```
MTB  > READ C1
DATA > 71
DATA > 72
DATA > 68
DATA > 66
DATA > 72
DATA > END
```

Με τον τρόπο αυτό οι πέντε παρατηρήσεις αρχειοθετούνται στη στήλη C1 (column C1) του Minitab.

Εστω ότι τα δεδομένα μας αποτελούνται από τα βάρη και τα ύψη πέντε ατόμων. Ας υποθέσουμε ότι εκτός από τα βάρη και τα ύψη επιθυμούμε να καθορίσουμε και την ταυτότητα κάθε ατόμου χρησιμοποιώντας τους αριθμούς 1,2,...,5. Στην περίπτωση αυτή τα δεδομένα θα έχουν την εξής μορφή:

Ατομο	Βάρος	Υψος
1	71	185
2	72	188
3	68	172
4	66	175
5	72	170

Προκειμένου να εισάγουμε στον υπολογιστή τα δεδομένα αυτά, θα χρησιμοποιήσουμε τις εξής εντολές:

```
MTB  > READ  C1  C2  C3
DATA >       1   71  185
DATA >       2   72  188
DATA >       3   68  172
```

```
DATA >        4   66   175
DATA >        5   72   170
DATA > END
```

Οι στήλες C1, C2 και C3 του Minitab περιέχουν τώρα την ταυτότητα των πέντε ατόμων, τα βάρη τους και τα ύψη τους αντίστοιχα. Οπως και στη συνήθη δακτυλογράφηση θα πρέπει να προσέχουμε να αφήνουμε ένα κενό μεταξύ των παρατηρήσεων ή μεταξύ των λέξεων που χρησιμοποιούμε.

Η εντολή SET

Η εντολή SET (τοποθέτησε) είναι ένας άλλος τρόπος εισαγωγής στοιχείων στον υπολογιστή. Η εντολή αυτή διαφέρει από την εντολή READ στο ότι μας επιτρέπει να τοποθετήσουμε τις παρατηρήσεις σε μια ή περισσότερες γραμμές. Για τις πέντε παρατηρήσεις που αναφέρονται στο βάρος των πέντε ατόμων, η κατάλληλη εντολή θα είναι:

```
MTB  > SET C1
DATA > 71  72  68  66  72
DATA > END
```

Προκειμένου να εισαχθούν τα στοιχεία για το βάρος και το ύψος των πέντε ατόμων, οι αντίστοιχες εντολές θα είναι:

```
MTB  > SET C1
DATA > 1  2  3  4  5
DATA > END
MTB  > SET C2
DATA > 71  72  68  66  72
DATA > END
MTB  > SET C3
```

```
DATA > 185  188  172  175  170
DATA > END
```

Υπενθυμίζουμε ότι η ένδειξη (prompt) MTB>, όπως επίσης και η ένδειξη DATA> εμφανίζεται στην οθόνη. Εκείνο που εμείς έχουμε να κάνουμε είναι να πληκτρολογήσουμε τις λέξεις για τις εντολές ή τους αριθμούς και να πατήσουμε το πλήκτρο ENTER μετά από κάθε γραμμή.

Η εντολή PRINT

Μπορούμε πάντοτε να τυπώσουμε τα δεδομένα που έχουμε εισαγάγει στον υπολογιστή χρησιμοποιώντας την εντολή PRINT. Για παράδειγμα, αν τα ύψη των πέντε ατόμων έχουν τοποθετηθεί στη στήλη C3, όταν γράψουμε την εντολή

```
MTB > PRINT C3
```

ο computer τυπώνει τις πέντε παρατηρήσεις για τα ύψη που περιέχονται στη στήλη C3. Εξάλλου εάν έχουμε τοποθετήσει την ταυτότητα του ατόμου στη στήλη C1, το βάρος του στην C2 και το ύψος του στην C3, τότε η εντολή

```
MTB > PRINT C1 C2 C3
```

ή, εναλλακτικά, η εντολή

```
MTB > PRINT C1 - C3
```

τυπώνει τις πληροφορίες που έχουν τοποθετηθεί στις στήλες C1, C2 και C3.

Διορθώσεις: Οι εντολές LET, DELETE, INSERT

Πολλές φορές θα συμβεί να ανακαλύψουμε ένα λάθος αφού έχουμε καταγράψει τα στοιχεία στον υπολογιστή. Ο ευκολότερος τρόπος να κάνουμε διορθώσεις είναι να μεταβούμε από την οθόνη εργασίας (session window) στο φύλλο των δεδομένων (data worksheet). Οταν βρισκόμαστε στο φύλλο των δεδομένων οι αλλαγές μπορούν να γίνουν

κατευθείαν χωρίς να χρειάζεται κάποια συγκεκριμένη εντολή. Για την πρόσβαση στο φύλλο εργασίας, οι υπολογιστές IBM επιλέγουν την οθόνη δεδομένων (data screen) από το edit menu.

Μεταβολές στα δεδομένα μπορούν επίσης να γίνουν από την οθόνη εργασίας (session window) χρησιμοποιώντας την εντολή LET. Για παράδειγμα, η εντολή

MTB > LET C3(4) = 10.2

αντικαθιστά την τέταρτη τιμή της στήλης C3 του Minitab με την τιμή 10.2.

Η εντολή

MTB > LET C2(1) = 6

αντικαθιστά την πρώτη τιμή στην στήλη C2 με τον αριθμό 6.

Εάν θέλουμε να διαγράψουμε (delete) μια ολόκληρη γραμμή στοιχείων χρησιμοποιούμε την εντολή DELETE. Για παράδειγμα, η εντολή

MTB > DELETE 2 C3

διαγράφει τη δεύτερη γραμμή της στήλης C3.

Η εντολή

MTB > DELETE 3:6 C1 C2

διαγράφει τις γραμμές από 3 μεχρι 6 από τις στήλες C1 και C2.

Εάν θέλουμε να παρεμβάλουμε ή να προσθέσουμε νέες γραμμές στα δεδομένα στηλών που υπάρχουν στο Minitb, χρησιμοποιούμε την εντολή INSERT. Για παράδειγμα, η εντολή

```
MTB  > INSERT 2 3 C1
DATA > 45
DATA > END
```

παρεμβάλλει την τιμή 45 μεταξύ των γραμμών 2 και 3 της στήλης C1. Δηλαδή η τρίτη γραμμή της C1 γίνεται τώρα 45.

Η εντολή

```
MTB  > INSERT 3 4 C1 C2
DATA > 98 11.8
```

DATA > END

παρεμβάλλει τις τιμές 98 και 11.8 μεταξύ των γραμμών 3 και 4 των δεδομένων των στηλών C1 και C2 αντίστοιχα.

Αριθμητικές πράξεις: Η εντολή LET

Εστω ότι θέλουμε να δημιουργήσουμε μια καινούργια στήλη (διάνυσμα) στο Minitab κάνοντας κάποια αριθμητική πράξη όπως, για παράδειγμα, πρόσθεση (addition (+)), αφαίρεση (substruction (-)), πολλαπλασιασμό (multiplication (*)), διαίρεση (division (/)) ή ύψωση σε δύναμη (exponentiation (**)) σε μια ή περισσότερες υπάρχουσες στήλες (διανύσματα) του Minitab. Αυτό μπορεί να γίνει με τη χρησιμοποίηση της εντολής LET. Για παράδειγμα, η εντολή

MTB > LET C4 = C2**2

δημιουργεί ένα καινούριο διάνυσμα - στήλη C4 στο Minitab που περιέχει τα τετράγωνα των τιμών των γραμμών που είναι τοποθετημένα στη στήλη C2.

Εξάλλου η εντολή

MTB > LET C5 = C2*C3

δημιουργεί στο Minitab τη στήλη C5 που περιέχει τα γινόμενα των τιμών των αντιστοίχων γραμμών των διανυσμάτων C2 και C3.

Η εντολή NAME

Στα διανύσματα δεδομένων του Minitab μπορούμε να δώσουμε ένα κανονικό όνομα (διαφορετικό από το C1,C2,... κ.λ.π.) και στη συνέχεια να αναφερόμαστε στο διάνυσμα αυτό με το όνομά του. Για να το κάνουμε αυτό χρησιμοποιούμε την εντολή NAME (ονομασία). Προκειμένου να χρησιμοποιήσουμε ένα όνομα για ένα διάνυσμα στοιχείων του Minitab θα πρέπει να είμαστε βέβαιοι ότι το όνομα που επιλέγουμε δεν περιέχει περισσότερους από οκτώ χαρακτήρες. Για παράδειγμα, σε σχέση με τα πέντε άτομα, τα βάρη τους και τα ύψη τους μπορούμε να χρησιμοποιήσουμε την εντολή

MTB > NAME C1='person' C2='weight' C3='height'.

Με τον τρόπο αυτό τα τρία διανύσματα στοιχείων του Minitab C1,C2 και C3 έχουν ονομασθεί "person", "weight", "height" αντίστοιχα. Παρατηρούμε ότι το σήμα της αποστρόφου (') είναι εκείνο με το οποίο αρχίζει και τελειώνει ένα δεδομένο όνομα. Αυτές οι αρχικές και τελικές απόστροφοι θα πρέπει να τοποθετούνται κάθε φορά που χρησιμοποιούμε το δοθέν όνομα ή ονόματα. Για παράδειγμα, εάν θέλουμε να αρχειοθετήσουμε δεδομένα στις μεταβλητές 'person', 'weight', 'height' θα πρέπει να χρησιμοποιήσουμε την εντολή

MTB > READ 'person' 'weight' 'height'.

Οι εντολές SAVE και RETRIEVE

Είναι δυνατόν νά αποθηκεύσουμε ένα φύλο εργασίας (worksheet) του Minitab στο σκληρό δίσκο του υπολογιστή για να το χρησιμοποιήσουμε αργότερα. Αυτό γίνεται με τη χρήση της εντολής SAVE, η οποία ακολουθείται από το όνομα που θέλουμε να δώσουμε στο file στο οποίο θα αποθηκευθεί η συγκεκριμένη εργασία. Για παράδειγμα, η εντολή

MTB > SAVE 'EXAMPLE'

αποθηκεύει όλα τα δεδομένα, όλες τις σταθερές, τα ονόματα των διανυσμάτων κ.λ.π. που υπήρχαν στο φύλο εργασίας στο file με όνομα 'EXAMPLE'. Μπορούμε να ανακαλέσουμε αργότερα ένα file που έχει ήδη αρχειοθετηθεί χρησιμοποιώντας την εντολή RETRIEVE ακολουθούμενη από το όνομα του file. Για παράδειγμα, η εντολή

MTB > RETRIEVE 'EXAMPLE'

επαναφέρει όλους τους αριθμούς, τα ονόματα των διανυσμάτων και τις σταθερές που είχαν αποθηκευθεί στο file με το όνομα 'EXAMPLE'.

Αποθήκευση και ανάσυρση δεδομένων σε δισκέττα

Για να αποθηκεύσουμε ένα φύλο εργασίας σε δισκέττα θα πρέπει να δώσουμε το όνομα του drive στο οποίο περιέχεται η δισκέττα. Για

παράδειγμα, η εντολή

MTB > SAVE 'A:EXAMPLE'

αποθηκεύει το file 'EXAMPLE' στη δισκέττα που βρίσκεται στο drive A του υπολογιστή IBM.

Ανάσυρση ενός φύλου εργασίας που έχει προηγουμένως αποθηκευθεί, από μια δισκέττα γίνεται με τον ίδιο τρόπο. Ετσι

MTB > RETRIEVE 'A:EXAMPLE'

ανασύρει το φύλο εργασίας από τη δισκέττα που βρίσκεται στο drive A του υπολογιστή.

Υποεντολές (Subcommands)

Προκειμένου να δοθεί η δυνατότητα επεξεργασίας περισσοτέρων πληροφοριών, οι περισσότερες από τις εντολές του Minitab έχουν υποεντολές. Προκειμένου να χρησιμοποιήσουμε μια υποεντολή, πληκτρολογούμε ένα ελληνικό ερωτηματικό (;) (το αγγλικό semicolon) στο τέλος της συγκεκριμένης εντολής. Στη συνέχεια, στην επόμενη γραμμή εμφανίζεται η ένδειξη (prompt) SUBC>. Θα πρέπει πάντα να θυμόμαστε **να τελειώνουμε την τελευταία σειρά που περιεχει υποεντολή με τελεία.**

Οι εντολές HELP και STOP

Εάν δεν θυμόμαστε πώς να χρησιμοποιήσουμε κάποια εντολή του Minitab, μπορούμε να χρησιμοποιήσουμε την εντολή HELP για να πάρουμε πληροφορίες για τη συγκεκριμένη εντολή. Για παράδειγμα, η εντολή

MTB > HELP SET

μας δίνει μια περιληπτική εξήγηση για την εντολή SET.

Μπορούμε να τερματίσουμε μια περίοδο εργασίας (session) του Minitab χρησιμοποιώντας την εντολή

MTB > STOP .

Το στατιστικό πακέτο SAS

Το στατιστικό πακέτο SAS (Statistical Analysis System) είναι, ίσως, το ισχυρότερο στατιστικό πακέτο που κυκλοφορεί στην αγορά με ικανότητα να αναλύσει μεγάλο όγκο δεδομένων που χρειάζονται πολύπλοκους στατιστικούς υπολογισμούς. Παρ'ότι σήμερα κυκλοφορεί έκδοση του SAS ειδική για μικρουπολογιστές συνιστάται η χρησιμοποίησή του σε μεγάλους υπολογιστές (mainframes). Αυτό γιατί, ακριβώς λόγω της πολυπλοκότητας του πακέτου η έκδοση για μικρουπολογιστές δεν έχει διαθέσιμες πολλές από τις δυνατότητες που έχει η έκδοση για μεγάλους υπολογιστές.

Οταν χρησιμοποιούμε το SAS θα πρέπει πάντα να θυμόμαστε ότι **όλες οι εντολές θα πρέπει να τελειώνουν με ένα ελληνικό ερωτηματικό (;) (αγγλικό semicolon).**

Εισαγωγή στοιχείων

Πριν κάνουμε εισαγωγή στοιχείων στον υπολογιστή θα πρέπει να χρησιμοποιήσουμε τις τρεις εντολές DATA, INPUT και CARDS με αυτή τη σειρά. Ενας εύκολος τρόπος για να γίνει αυτό είναι να χρησιμοποιήσουμε την εντολή INPUT για να χαρακτηρίσουμε (ονομάσουμε) τις ποσότητες για τις οποίες θα εισάγουμε στοιχεία στον υπολογιστή, ώστε να έχουμε τη δυνατότητα να τα χρησιμοποιήσουμε και αργότερα. Πριν από αυτό θα πρέπει να έχουμε πληκτρολογήσει την εντολή DATA; . Μετά από το INPUT θα πρέπει να πληκτρολογήσουμε την εντολή CARDS; σε διαφορετική βέβαια γραμμή. Στη συνέχεια θα ακολουθήσουν τα δεδομένα.

Ας δούμε, για παράδειγμα, πως θα γίνει η εισαγωγή των στοιχείων που αναφέρονται στα πέντε άτομα με τα βάρη και τα ύψη τους αντίστοιχα που χρησιμοποιήσαμε στην ενότητα για το Minitab. Για να γίνει η εισαγωγή των στοιχείων αυτών χρησιμοποιούμε τις παρακάτω εντολές:

```
DATA;
```

```
INPUT PERSON WEIGHT HEIGHT;
CARDS;
1  71  185
2  72  188
3  68  172
4  66  175
5  72  170
```

Παρατηρούμε ότι τα ονόματα που δώσαμε στα τρία διανύσματα -στήλες των δεδομένων είναι PERSON, WEIGHT και HEIGHT αντίστοιχα. Επομένως, σε οποιαδήποτε φάση της ανάλυσης των στοιχείων χρειάζεται να χρησιμοποιήσουμε τα δεδομένα για το ύψος στους υπολογισμούς μας θα καλέσουμε τη μεταβλητή HEIGHT όπως αυτή δόθηκε στην εντολή INPUT.

Σημείωση: Οπως και στο Minitab, το όνομα μιας μεταβλητής δεν μπορεί να έχει περισσότερους από οκτώ χαρακτήρες.

Προκειμένου να δώσουμε ένα ακόμα παράδειγμα, ας δούμε τα δεδομένα που αναφέρονται στον όγκο πωλήσεων Y ενός προϊόντος (σε εκατοντάδες χιλιάδες δραχμές) σε σχέση με τα έξοδα διαφήμισης X του προϊόντος αυτού (σε δεκάδες χιλιάδες δραχμές).

X	Y
1.2	4.4
2.4	4.9
3.1	5.8
4.0	5.7
4.8	6.4
5.6	7.1

Για τα δεδομένα αυτά χρειαζόμαστε τις ακόλουθες εντολές:

```
DATA;
```

```
INPUT X Y;
CARDS;
1.2  4.4
2.4  4.9
3.1  5.8
4.0  5.7
4.8  6.4
5.6  7.1
```

Εστω ότι έχουμε στοιχεία που αναφέρονται στην κατανάλωση βενζίνης τριων διαφορετικών αυτοκινήτων που δοκιμάσθηκαν από δύο διαφορετικούς οδηγούς. (Η κατανάλωση αναφέρεται σε χιλιόμετρα ανά λίτρο βενζίνης).

	Αυτοκίνητο		
Οδηγός	1	2	3
1	13.6	12.8	11.9
2	16.9	16.1	12.1

Από τα στοιχεία αυτά βλέπουμε ότι ο οδηγός 1 πέτυχε 13.6 χιλιόμετρα ανά λίτρο όταν οδηγούσε το αυτοκίνητο 1, 12.8 οδηγώντας το αυτοκίνητο 2 και 11.9 οδηγώντας το αυτοκίνητο 3 κ.λ.π. Για τα δεδομένα αυτά θα δώσουμε τις εξής εντολές:

```
DATA;
INPUT DRIVER AUTO DISTANCE;
CARDS;
1  1  13.6
1  2  12.8
1  3  11.9
2  1  16.9
2  2  16.1
2  3  12.1
```

Η ονομασία DRIVER αναφέρεται στον οδηγό, AUTO αναφέρεται στο αυτοκίνητο και DISTANCE αναφέρεται στα χιλιόμετρα ανά λίτρο που διανύθηκαν από κάθε οδηγό με κάθε αυτοκίνητο.

Παρατηρούμε ότι τα δεδομένα πληκτρολογήθηκαν με ένα συγκεκριμένο συνδυασμό ανά όνομα χρησιμοποιώντας τη διάταξη που καθορίσθηκε με τα ονόματα στην εντολή INPUT.

Εκτύπωση δεδομένων

Μπορούμε να τυπώσουμε δεδομένα που έχουν ήδη εισαχθεί στον υπολογιστή χρησιμοποιώντας την εντολή PROC PRINT. Για παράδειγμα, η εντολή

```
PROC PRINT;
```

τυπώνει όλα τα δεδομένα που έχουν εισαχθεί στον υπολογιστή.

Εάν θέλουμε να τυπώσουμε μόνο συγκεκριμένες ποσότητες μπορούμε να το κάνουμε περιλαμβάνοντας την εντολή VAR (variable) και τις συγκεκριμένες μεταβλητές, τα στοιχεία των οποίων θέλουμε να τυπώσουμε. (Τα ονόματα, βέβαια, των μεταβλητών μπορούν να είναι μόνο αυτά που δόθηκαν στην εντολή INPUT).

Για παράδειγμα, αν θέλουμε να τυπώσουμε μόνο τα έξοδα διαφήμισης (που τα είχαμε ονομάσει X στην εντολή INPUT) θα χρησιμοποιήσουμε τις ακόλουθες εντολές:

```
PROC PRINT;
VAR X;
```

Αριθμητικές πράξεις στο SAS

Μπορούμε να κάνουμε μαθηματικούς υπολογισμούς στο SAS σε μια ή περισσότερες από τις μεταβλητές που έχουμε ήδη χρησιμοποιήσει στην εντολή INPUT, χρησιμοποιώντας τα ίδια σύμβολα που χρησιμοποιούμε στο Minitab. Δηλαδή (+) για άθροιση, (-) για αφαίρεση, (*) για πολλαπλασιασμό, (/) για διαίρεση και (**) για

ύψωση σε δύναμη.

Για παράδειγμα, οι εντολές

```
RATIO = X/Y;
XSQUARE = X**2;
DIFF = X-Y;
SUM = X+Y;
PROD = X*Y;
```

δημιουργούν τις ποσότητες RATIO (λόγος), XSQUARE (X^2), DIFF (διαφορά), SUM (άθροισμα) και PROD (γινόμενο) για τις ποσότητες X και Y που είχαν ορισθεί στην εντολή INPUT.

ΑΣΚΗΣΕΙΣ

1.1 Δώστε δύο παραδείγματα για ποσοτικά δεδομένα και δύο για ποιοτικά δεδομένα.

1.2 Για κάθε ένα από τα παρακάτω παραδείγματα δεδομένων να καθορίσετε αν τα δεδομένα αυτά είναι ποσοτικά ή ποιοτικά:

α) οι αρχικοί μισθοί πτυχιούχων Στατιστικών

β) ο μήνας κατά τον οποίο ένας υπάλληλος του Πανεπιστημίου παίρνει την άδειά του

γ) ο βαθμός ενός φοιτητή στις εξετάσεις ενός μαθήματος Στατιστικής.

1.3 Ενα εβδομαδιαίο περιοδικό ενδιαφέρεται να μελετήσει ορισμένα από τα χαρακτηριστικά των αναγνωστών του. Μια δειγματοληπτική έρευνα σε 200 αναγνώστες του περιοδικού περιλάμβανε μια σειρά από ερωτήσεις. Για κάθε μια από τις ερωτήσεις που ακολουθούν να καθορισθεί εάν οι δυνατές απαντήσεις είναι ποιοτικές ή ποσοτικές:

α) Ποια είναι η ηλικία σας;

β) Ποια είναι η οικογενειακή σας κατάσταση;

γ) Είναι το μηνιαίο εισόδημά σας περισσότερο από 150.000 δραχμές ή λιγότερο απο 150.000 δραχμές;

δ) Πόσα άλλα περιοδικά αγοράζετε κάθε εβδομάδα;

1.4 Για κάθε ένα από τα παραδείγματα δεδομένων που ακολουθούν να καθορίσετε αν τα δεδομένα είναι ποιοτικά ή ποσοτικά:

α) ο μήνας με τον υψηλότερο αριθμό πωλήσεων για κάθε εταιρεία σε ένα δείγμα

β) το τμήμα στο οποίο κάθε καθηγητής από ένα δείγμα καθηγητών Πανεπιστημίου διδάσκει

γ) το μέγεθος ενός αεριούχου ποτού που παραγγέλλουν οι πελάτες ενός ξενοδοχείου (μικρό, μεσαίο ή μεγάλο).

Για τις καταστάσεις που αναφέρονται παρακάτω να καθορίσετε τα εξής στατιστικά μεγέθη:

1. τον πληθυσμό ή τη διεργασία
2. το πλαίσιο
3. τη στατιστική μεταβλητή (μεταβλητές)
4. την παράμετρο (παραμέτρους) που ενδιαφέρουν
5. τη στατιστική συνάρτηση (συναρτήσεις) που ενδιαφέρουν.

α) Η αντικαρκινική εταιρεία ενδιαφέρεται να ανανεώσει τις πληροφορίες που έχει για τις συνήθειες καπνίσματος.

β) Η διεύθυνση Στατιστικής μιας τράπεζας ενδιαφέρεται να εκτιμήσει το χρόνο αναμονής των πελατών πριν εξυπηρετηθούν.

γ) Ο κατασκευαστής ενός συγκεκριμένου μετάλλου ενδιαφέρεται να εκτιμήσει την αντοχή του μετάλλου αυτού.

δ) Το Εμπορικό Επιμελητήριο μιας πόλης ενδιαφέρεται να συγκεντρώσει πληροφορίες για το ποσό των χρημάτων που ξοδεύουν όσοι παρακολουθούν συνέδρια στην πόλη αυτή.

1.5 Ο προπονητής μιας ομάδας basketball, βλέποντας ότι η ομάδα του είναι σχετικά αδύναμη στη άμυνα, πιστεύει ότι θα πρέπει να σκοράρει 85 πόντους κατά μέσο όρο ανά παιχνίδι ώστε να έχει επιτυχία στο πρωτάθλημα. Μετά από δέκα παιχνίδια όμως, προκύπτει ότι η ομάδα του είχε ένα μέσο όρο 75 πόντων ανά παιχνίδι. Ο πρόεδρος της ομάδας του συνιστά να κάνει σημαντικές αλλαγές είτε στη στρατηγική είτε στους παίκτες που χρησιμοποιεί προκειμένου να αυξήσει την επιθετική απόδοση της ομάδας. Οι πόντοι που σημείωσε η ομάδα αυτή στα δέκα πρώτα παιχνίδια είναι ως εξής:

Παιχνίδι :	1	2	3	4	5	6	7	8	9	10
Πόντοι :	61	53	61	70	71	78	81	86	96	93

Πιστεύετε ότι οι αλλαγές που προτείνονται στον προπονητή είναι απαραίτητες; Εξηγήστε την απάντησή σας.

1.6 Μια φοιτήτρια ενδιαφέρεται να ελέγξει το βάρος της. Ως ένα πρώτο βήμα για κάτι τέτοιο αποφασίζει να καταγράφει καθημερινά τον αριθμό των θερμίδων που καταναλώνει (αριθμός θερμίδων ανά ημέρα) για ένα μήνα. Τα αποτελέσματα είναι τα εξής:

Εβδομάδα 1 :	1295	1720	1215	1210	1260	1075	1100
Εβδομάδα 2 :	1200	1435	1255	1300	1385	1515	1105
Εβδομάδα 3 :	1270	1200	1215	1225	995	1270	1350
Εβδομάδα 4 :	1285	1110	1430	1180	1385	1300	1175
Εβδομάδα 5 :	1475	1225					

α) Με βάση τις γνώσεις σας για τις συνήθειες διατροφής των νέων, προσδιορίστε κάποιες πιθανές πηγές μεταβλητότητας στα δεδομένα αυτά.

β) Προσδιορίστε κάθε μια πηγή μεταβλητότητας της ερώτησης 1 ως προερχόμενη από κοινές ή από ειδικές αιτίες μεταβλητότητας.

γ) Κατασκευάστε ένα διάγραμμα ροής για την καθημερινή κατανάλωση θερμίδων της φοιτήτριας αυτής.

δ) Το διάγραμμα ροής της προηγούμενης ερώτησης παρέχει κάποιες εμφανείς ενδείξεις αυτάρκειας της διεργασίας;

1.7 Οι πωλήσεις βιβλίων με σκληρό εξώφυλλο (αριθμός βιβλίων που πωλήθηκαν) σε κάποιο βιβλιοπωλείο για 30 συνεχείς ημέρες ήταν ως εξής:

Εβδομάδα 1 :	38	35	76	58	48	59
Εβδομάδα 2 :	67	63	33	69	53	51
Εβδομάδα 3 :	28	25	36	32	61	57
Εβδομάδα 4 :	49	78	48	42	72	52
Εβδομάδα 5 :	47	66	58	44	44	56

α) Κατά την κρίση σας η διαδικασία πωλήσεων είναι ευσταθής ή όχι;

β) Εάν σας έλεγαν να μετρήσετε τον αριθμό των βιβλίων που πωλούνται σε μια δεδομένη ημέρα πιστεύετε ότι θα είχατε δυσκολίες στο να αποφασίσετε τι είναι βιβλίο; Εξηγήστε.

γ) Απαριθμήσατε όσο το δυνατόν περισσότερες αιτίες μεταβλητότητας στα δεδομένα αυτά.

δ) Χαρακτηρίστε κάθε μια από τις αιτίες μεταβλητότητας της προηγούμενης ερώτησης ως προερχόμενη από κοινή αιτία ή από ειδική αιτία.

1.8 Ο αναλυτής εργασιών μιας τράπεζας παρατήρησε και κατέγραψε τον αριθμό των συναλλαγών (καταθέσεις και αποσύρσεις πελατών) κάθε ημέρας για την περίοδο επτά εβδομάδων. Τα δεδομένα (από Δευτέρα μέχρι Παρασκευή για κάθε εβδομάδα) ακολουθούν πιο κάτω. Με βάση ένα διάγραμμα ροής θα λέγατε ότι η διαδικασία των συναλλαγών εμφανίζεται να είναι ευσταθής για κάθε μέρα της εβδομάδας; Εξηγείστε την απάντησή σας στο πλαίσιο των αιτίων μεταβλητότητας για όποιες ημέρες πιστεύετε ότι δεν είναι σταθερή.

	Δευτέρα	Τρίτη	Τετάρτη	Πέμπτη	Παρασκευή
Εβδομάδα 1 :	64	96	75	105	169
Εβδομάδα 2 :	67	104	74	73	202
Εβδομάδα 3 :	70	116	89	112	230
Εβδομάδα 4 :	68	95	121	83	168
Εβδομάδα 5 :	55	109	99	94	157
Εβδομάδα 6 :	52	102	72	82	123
Εβδομάδα 7 :	68	90	105	78	179

1.9 Τα δεδομένα που ακολουθούν αναφέρονται σε αριθμό συναλλαγών ανά ημέρα για δύο διαδοχικούς μήνες στο υποκατάστημα μιας μεγάλης τράπεζας. Οι αριθμοί αντιπροσωπεύουν το συνολικό αριθμό ιδιωτικών και εμπορικών συναλλαγών που διεκπεραιώθηκαν στη συγκεκριμένη εργάσιμη ημέρα. (Δίνονται οι ημέρες της εβδομάδας).

α) Να εξηγήσετε γιατί η καταγραφή τέτοιων στοιχείων είναι σημαντική για τη Διεύθυνση της τράπεζας.

β) Χρησιμοποιώντας ένα διάγραμμα ροής να καθορίσετε αν οι δραστηριότητες του υποκαταστήματος αυτού ήταν ευσταθείς κατά τη διάρκεια των δύο αυτών μηνών.

γ) Σε μια γραφική παράσταση να απεικονίσετε τον αριθμό των συναλλαγών στον κάθετο άξονα που αντιστοιχεί σε κάθε μέρα της εβδομάδας (εμφανιζόμενη στον οριζόντιο άξονα). Εξηγήσατε τις διαπιστώσεις σας.

Μήνας 1		Μήνας 2	
Ημερομηνία	Αριθμός Συναλλαγών	Ημερομηνία	Αριθμός Συναλλαγών
2(Δε)	792	2(Πε)	821
3(Τρ)	791	3(Πα)	917
4(Τε)	781	6(Δε)	772
5(Πε)	818	7(Τρ)	724
6(Πα)	912	8(Τε)	701
9(Δε)	812	9(Πε)	776
10(Τρ)	782	10(Πα)	891
11(Τε)	911	13(Δε)	804
12(Πε)	811	14(Τρ)	762
13(Πα)	889	15(Τε)	711
16(Δε)	879	16(Πε)	890
17(Τρ)	801	17(Πα)	904
18(Τε)	768	21(Τρ)	836
19(Πε)	821	22(Τε)	762
20(Πα)	991	23(Πε)	803
23(Δε)	798	24(Πα)	961
24(Τρ)	891	27(Δε)	792
26(Πε)	801	28(Τρ)	781
27(Πα)	981	29(Τε)	741
30(Δε)	802	30(Πε)	817
31(Τρ)	888	31(Πα)	1011