ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ
ΜΕΣΟΛΟΓΓΙΟΥ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ & ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ

Π Τ Υ Χ Ι Α Κ Η Ε Ρ Γ Α Σ Ι Α

Στατιστική και Μέτρα Διασποράς και η εφαρμογή τους στις επιχειρήσεις

ΜΙΑΟΥΛΗΣ ΚΩΝΣΤΑΝΤΙΝΟΣ
ΤΣΙΑΛΑΜΑΝΗΣ ΔΗΜΗΤΡΙΟΣ

ΕΙΣΗΓΗΤΗΣ
ΑΘΑΝΑΣΙΟΣ ΜΕΓΑΡΙΤΗΣ Επιστημονικός συνεργάτης

ΜΕΣΟΛΟΓΓΙ 2012

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ
ΜΕΣΟΛΟΓΓΙΟΥ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ & ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ

Π Τ Υ Χ Ι Α Κ Η Ε Ρ Γ Α Σ Ι Α

Στατιστική και Μέτρα Διασποράς και η εφαρμογή τους στις επιχειρήσεις

ΜΙΑΟΥΛΗΣ ΚΩΝΣΤΑΝΤΙΝΟΣ (Α.Μ. 14178)

ΤΣΙΑΛΑΜΑΝΗΣ ΔΗΜΗΤΡΙΟΣ (Α.Μ. 14351)

ΕΙΣΗΓΗΤΗΣ
ΑΘΑΝΑΣΙΟΣ ΜΕΓΑΡΙΤΗΣ Επιστημονικός συνεργάτης

ΜΕΣΟΛΟΓΓΙ 2012

ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ
Τ.Ε.Ι. ΜΕΣΟΛΟΓΓΙΟΥ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ & ΟΙΚΟΝΟΜΙΑΣ
Τ Μ Η Μ Α   Λ Ο Γ Ι Σ Τ Ι Κ Η Σ

ΕΙΣΗΓΗΤΙΚΗ ΕΚΘΕΣΗ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ υποβάλλεται ενσωματωμένη σε κάθε αντίτυπο της πτ. εργασίας

ΤΙΤΛΟΣ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ *καταχώρηση θέματος όπως στον πίνακα θεμάτων*

Στατιστική και μέτρα διασποράς και η εφαρμογή τους στις επιχειρήσεις

ΦΟΙΤΗΤΕΣ

ΕΠΩΝΥΜΟ	ΟΝΟΜΑ	ΑΡ. ΜΗΤΡ.	E-mail
Μιαούλης	Κωνσταντίνος	14178	
Τσιαλαμάνης	Δημήτριος	14351	

ΕΚΘΕΣΗ ΕΓΚΡΙΣΗΣ ΕΙΣΗΓΗΤΗ ΕΚΠΑΙΔΕΥΤΙΚΟΥ *καταχώρηση από τον εισηγητή*

Η παρούσα πτυχιακή εργασία πραγματεύεται τις εφαρμογές της στατιστικής και ειδικότερα των μέτρων διασποράς στις επιχειρήσεις δίνοντας έμφαση στον έλεγχο ποιότητας. Κατά τη γνώμη μου, το θέμα αντιμετωπίστηκε επαρκώς, η δομή είναι ικανοποιητική και συνεπώς εισηγούμαι την έγκρισή της.

ΗΜΕΡΟΜΗΝΙΑ

(Σε περίπτωση ηλεκτρονικής υποβολής δεν χρειάζεται υπογραφή)

ΥΠΟΓΡΑΦΗ ……………………………………………

Πρόλογος

Η παρούσα Πτυχιακή Εργασία πραγματεύεται τις εφαρμογές της στατιστικής στις επιχειρήσεις. Συγκεκριμένα γίνεται παρουσίαση των μέτρων διασποράς και η εφαρμογή αυτών στην επίλυση διάφορων προβλημάτων που έχουν να κάνουν τόσο με την καθημερινή λειτουργία των επιχειρήσεων, όσο και με τη βελτίωση του παραγόμενου προϊόντος ή της προσφερόμενης υπηρεσίας. Για τη συγγραφή της εργασίας χρησιμοποιήθηκε σχετική Ελληνική αλλά και ξένη βιβλιογραφία, καθώς και διάφορες διαδικτυακές πηγές.

Ακόμα το περιεχόμενο της Πτυχιακής Εργασίας χωρίζεται σε τρία κεφάλαια. Στο πρώτο κεφάλαιο γίνεται μια εισαγωγή στην στατιστική, μέσω μιας σύντομης ιστορικής αναδρομής και με την παρουσίαση των βασικών στατιστικών εννοιών.

Στο δεύτερο κεφάλαιο παρουσιάζονται τα μέτρα διασποράς καθώς και ο συντελεστής μεταβολής. Για την καλύτερη κατανόηση της χρήσης των μέτρων αυτών, δίνονται αναλυτικά παραδείγματα εφαρμογής τους αλλά και μια σύγκριση ανάμεσά τους.

Στο τρίτο κεφάλαιο αναπτύσσονται με παραδείγματα οι εφαρμογές των μέτρων διασποράς στις επιχειρήσεις δίνοντας έμφαση στον έλεγχο ποιότητας.

Εισαγωγή

Η λέξη στατιστική, κατά μία άποψη, προέρχεται από τη λατινική λέξη status, που σημαίνει κράτος-κατάσταση, και δήλωνε, από τότε που πρωτοχρησιμοποιήθηκε, συλλογή στοιχείων για τις κρατικές ανάγκες. Οι ασχολούμενοι με τη συλλογή και ανάλυση αυτών των στοιχείων ονομάζονταν «κρατικοί» ή «στατιστικοί». Άλλη μία εκδοχή της προέλευσης της λέξης στατιστική, σύμφωνα με τις απόψεις που διατυπώνονται στα περισσότερα ετυμολογικά λεξικά, είναι ότι αυτή έλκει την καταγωγή της από το ελληνικό ρήμα στατίζω, που σημαίνει τοποθετώ, διαπιστώνω, προσδιορίζω [στατιστικός –ή –ό (γαλλ. statistique<(στατίζω)].

Για την εμφάνιση της στατιστικής αξίζει να αναφέρουμε ότι οι πρώτες προσπάθειες συλλογής στατιστικών στοιχείων ανάγονται στους αρχαιότατους χρόνους. Από την ιστορία πληροφορούμαστε ότι σε απογραφές πληθυσμού, γης, κ.τ.λ. προέβαιναν και λαοί της αρχαιότητας (Έλληνες, Κινέζοι, Αιγύπτιοι, Ρωμαίοι, κ.λ.π).

Αξίζει μάλιστα να αναφερθεί ότι ο πρώτος Κυβερνήτης της Ελλάδας, Ι. Καποδίστριας, είχε ενδιαφερθεί σοβαρότατα για τη δημιουργία στατιστικής υπηρεσίας στην Ελλάδα διαβλέποντας τον κρίσιμο ρόλο που θα έπαιζε στη δημιουργία του νέου κράτους.

Στις αρχές του 20ου αιώνα, χρησιμοποιήθηκε η στατιστική μέθοδος ελέγχου και ποιότητας των βιομηχανικών προϊόντων και θεωρείται ως η πρώτη σε ευρεία εφαρμογή χρήση στατιστικών μεθόδων στην παραγωγή.

Στις ημέρες μας, ακούμε από τα μέσα μαζικής ενημέρωσης ή διαβάζουμε στα διάφορα έντυπα για «στατιστικές έρευνες». Οι έρευνες αυτές είναι στην ουσία, πληροφορίες που αναφέρονται σε συγκεκριμένο αντικείμενο έρευνας ή γεγονός και δίνονται με μορφή πινάκων ή διαγραμμάτων. Ανάλογα με το αντικείμενο έρευνας ή το γεγονός, η στατιστική παίρνει και ιδιαίτερη ονομασία. Για παράδειγμα, μιλάμε για «Στατιστική Επιχειρήσεων», για «Γεωργική Στατιστική», για «Στατιστική Τουρισμού» κ.τ.λ., όταν τα αριθμητικά δεδομένα αναφέρονται αντίστοιχα στις επιχειρήσεις, στη γεωργία, ή στον τουρισμό κ.ο.κ.

Εκτός από τη σημασία αυτή, η «Στατιστική» ως κλάδος των εφαρμοσμένων επιστημών όχι μόνο συγκεντρώνει και παρουσιάζει πληροφορίες, αλλά ταυτόχρονα μελετά και αναλύει τις παρατηρήσεις ή μετρήσεις που αναφέρονται σε ένα συγκεκριμένο πεδίο έρευνας. Έτσι η Στατιστική περιλαμβάνει τόσο τις μεθόδους συλλογής και επεξεργασίας στοιχείων όσο και τις μεθόδους ανάλυσης και μελέτης τους. Από τη μελέτη αυτή προκύπτουν οι σχέσεις που υπάρχουν στα διάφορα φαινόμενα και διατυπώνονται συμπεράσματα που είναι χρήσιμα για τη λήψη ορθών αποφάσεων. Μπορούμε λοιπόν να πούμε ότι:

Στατιστική είναι η εφαρμοσμένη επιστήμη που ασχολείται με τις επιστημονικές μεθόδους σχεδιασμού μιας μελέτης, συλλογής, επεξεργασίας, παρουσίασης και ανάλυσης στοιχείων με σκοπό την εξαγωγή συμπερασμάτων για την λήψη ορθών αποφάσεων.

Σύμφωνα με τον ορισμό αυτό της Στατιστικής μπορούμε να απαριθμήσουμε τα βασικά στάδια που ακολουθούμε για τη μελέτη ενός φαινομένου.
Αυτά είναι:

1. Ο σχεδιασμός της μελέτης του φαινομένου.
2. Η συγκέντρωση των απαραίτητων πληροφοριών (στατιστικών στοιχείων)
3. Η επεξεργασία και παρουσίαση των στοιχείων αυτών.
4. Η ανάλυση των στοιχείων με επιστημονικές μεθόδους.
5. Η εξαγωγή χρήσιμων συμπερασμάτων.

Η Στατιστική ως επιστήμη έχει τους δικούς της συμβολισμούς, ορολογία, θεωρήματα και τεχνικές. Χρησιμοποιείται σε όλες σχεδόν τις άλλες επιστήμες και στους περισσότερους τομείς της ανθρώπινης δραστηριότητας (διοίκηση, δημογραφία, βιομηχανία, εμπόριο, πολιτική, ιατρική κ.τ.λ.).

Η επίδραση της Στατιστικής στη ζωή μας είναι σήμερα πολύ μεγάλη. Εκτός από την εφαρμογή της στην απογραφή, στατιστικές μέθοδοι χρησιμοποιούνται για παράδειγμα στη μελέτη της γεννητικότητας, της θνησιμότητας, της μετανάστευσης κ.τ.λ. προκειμένου να ληφθούν ορθές αποφάσεις. Χρησιμοποιούνται επίσης στον τομέα της εκπαίδευσης για την εξαγωγή χρήσιμων συμπερασμάτων, όσον αφορά το εκπαιδευτικό σύστημα μιας χώρας και στη μετεωρολογία, η οποία δεν μπορεί να κάνει καμία πρόβλεψη χωρίς τη στατιστική ανάλυση των διαφόρων στοιχείων που επηρεάζουν το φαινόμενο που μελετάει κάθε φορά.

Η μεθοδολογία της Στατιστικής χρησιμοποιείται επίσης σε γεωργικές μελέτες, στην ιατρική και φαρμακευτική έρευνα κ.τ.λ. Η Στατιστική και η θεωρία των πιθανοτήτων χρησιμοποιούνται για να ληφθούν σοβαρές αποφάσεις. Για παράδειγμα, με στατιστικές μεθόδους ελέγχεται η ποιότητα των προϊόντων που παράγονται από μια βιομηχανία και στη συνέχεια αποφασίζεται η διάθεσή τους στην αγορά.

Όλα τα αναπτυγμένα κράτη χρησιμοποιούν τη Στατιστική για τη μελέτη των οικονομικών δραστηριοτήτων (παραγωγή, εισαγωγές, εξαγωγές, τιμές αγαθών κ.τ.λ.) και των διοικητικών και κοινωνικών θεμάτων (διοίκηση, κοινωνικές ασφαλίσεις κ.τ.λ.).

Ως εδώ αναφέρθηκαν μερικοί μόνο τομείς στους οποίους προσφέρει τις υπηρεσίες της η Στατιστική, αλλά είναι σκόπιμο να γίνει ιδιαίτερη αναφορά στο ρόλο της Στατιστικής στο χώρο των επιχειρήσεων.

ΠΕΡΙΕΧΟΜΕΝΑ

Κεφάλαιο 1^{ο} : Εισαγωγή στη Στατιστική

1.1 Ιστορική Αναδρομή της Στατιστικής Αντίληψης

Η Στατιστική εμφανίζεται, με μορφή αναγνωρίσιμη σύμφωνα με το σημερινό τρόπο θεώρησης των πραγμάτων, τον 17^{ο} αιώνα. Επί τρεισήμισι αιώνες καταλαμβάνει σημαντική θέση στον κοινωνικό χώρο, τροφοδοτούμενη από αυτόν και τροφοδοτώντας τον με στοιχεία επεξεργασμένης πληροφορίας και εργαλεία ανάλυσης. Τον τελευταίο ενάμιση αιώνα δεσπόζει σε χώρους όπου οι μετρήσεις παίζουν κυρίαρχο ρόλο και διαμορφώνει μεθόδους με πολλαπλή εφαρμογή, ενώ αυτονομείται ως προς την καθαρά μαθηματική της διάσταση, διασταυρούμενη με τις πιθανότητες που κινούνται αυτόνομα και παράλληλα με αυτήν από τον 17^{ο} αιώνα.

Οι ποικίλες χρήσεις των λέξεων «στατιστική» και «στατιστικός» αντανακλούν τη διάσταση θεωρήσεων για την πραγματικότητα και τη μέθοδο. Για ορισμένους συνιστά μια διοικητική δραστηριότητα καταγραφής ποικίλων δεδομένων, που οδηγεί σε μη αμφισβητήσιμους αριθμούς, οι οποίοι τροφοδοτούν την κοινωνική συζήτηση και προσανατολίζουν τη δράση. Για άλλους είναι ένας κλάδος των Μαθηματικών που διδάσκεται στο πανεπιστήμιο και χρησιμοποιείται από επιστήμονες βιολόγους, γιατρούς, οικονομολόγους και ψυχολόγους. Η αυτονόμηση των δύο αυτών σημασιών ξεκινά από τις αρχές του 20^{ου} αιώνα, όταν καθιερώθηκαν και διαδόθηκαν οι τεχνικές της παλινδρόμησης και της συσχέτισης από το κέντρο Βιομετρίας του Karl Pearson και στη συνέχεια αυτές της Επαγωγικής Στατιστικής που αναπτύχθηκαν στο γεωργικό πειραματικό εργαστήριο του Ronald Fisher. Από τότε, η Στατιστική εμφανίζεται ως κλάδος των Εφαρμοσμένων Μαθηματικών.

Περιγράφοντας τις απαρχές της Στατιστικής στην Αγγλία και τη Γερμανία κατά τον 17^{ο} αιώνα, ο Desrosières επισημαίνει ότι στην Αριθμητική Πολιτική (στην Αγγλία) οι πίνακες θνησιμότητας χρησιμεύουν, για να τεκμηριώσουν τα ισόβια εισοδήματα ή τα ασφάλιστρα στις ασφάλειες ζωής. Οι εκτιμήσεις πληθυσμού κατά επαρχία είναι απαραίτητες για τη φορολόγηση ή τη στρατολόγηση. Αντίθετα, η γερμανική στατιστική προτείνει στον Ηγεμόνα ή στον υπεύθυνο υπάλληλο ένα πλαίσιο οργάνωσης των διαθέσιμων ποικιλόμορφων γνώσεων για ένα κράτος. Ο

Conring (1606-1681) αναζητούσε, σύμφωνα με τον Lazarsfeld, ένα σύστημα που θα καθιστούσε τα γεγονότα ευκολότερα στην απομνημόνευση, πιο εύληπτα στη διδασκαλία και πιο εύχρηστα από τους κυβερνώντες. Η ανάπτυξη και οργάνωση της Στατιστικής ακολουθεί τις ιδιαιτερότητες διοικητικά ισχυρών χωρών, όπως για παράδειγμα στη Γαλλία όπου το κράτος είναι συγκεντρωτικό και το ίδιο συμβαίνει και με τη Στατιστική από σκοπιά τόσο διοικητική, όσο και γεωγραφική. Η Διοίκηση έχει ενσωματώσει σε μεγάλο βαθμό την πραγματογνωμοσύνη, ενώ η συμβολή των Πανεπιστημίων είναι σχετικά μικρή.

Η Δημόσια Στατιστική οργανώνεται γύρω από τις απογραφές και παρεμβαίνει κυρίως στο αντικείμενο της Δημογραφίας και της Οικονομίας. Στη Μεγάλη Βρετανία οι διοικητικές υπηρεσίες είναι περισσότερο αυτόνομες, ενώ οι αρχές των νομών και των δήμων έχουν περισσότερες εξουσίες από ότι στη Γαλλία. Η Στατιστική δεν υπήρξε ποτέ συγκεντρωμένη σε μία μόνο υπηρεσία, ενώ τα εθνικά γραφεία οφείλουν να σχετίζονται με τα τοπικά. Οι δύο κύριοι τομείς εφαρμογής της στατιστικής οφείλουν να είναι αφενός το εξωτερικό εμπόριο και οι επιχειρήσεις και αφετέρου ο πληθυσμός, η φτώχεια, η υγιεινή και η δημόσια υγεία. Στη Γερμανία, η επίσημη Στατιστική ενοποιείται μετά το 1871, οργανώνοντας εκτεταμένες έρευνες και απογραφές της παραγωγικής μηχανής.

Οι πανεπιστημιακοί στατιστικοί και οικονομολόγοι παράγουν πολλές ογκώδεις περιγραφικές και ιστορικές μονογραφίες με έντονα στατιστικό περιεχόμενο. Στις Ηνωμένες Πολιτείες, τέλος, οι δραστηριότητες της Δημόσιας Στατιστικής ενεργοποιούνται από τις ανά δεκαετία απογραφές που προβλέπονται από το Σύνταγμα του 1787, προκειμένου να κατανεμηθούν μεταξύ των Πολιτειών τα οικονομικά βάρη και οι έδρες στη Βουλή των Αντιπροσώπων. Μετά το 1933, ο πολιτικός, ο οικονομικός και ο σχετιζόμενος με τον προϋπολογισμό ρόλος της Ομοσπονδιακής Διοίκησης αλλάζει ριζικά. Η επίσημη Στατιστική οργανώνεται ξανά και συντονίζεται κάτω από την ώθηση μιας ομάδας πανεπιστημιακών στατιστικών και οικονομολόγων.

Η εισαγωγή, η εφαρμογή και η αξιοποίηση των εννοιών της μέσης τιμής, της διακύμανσης, της παλινδρόμησης και της συσχέτισης αποτέλεσαν κομβικά σημεία στην εξέλιξη της Στατιστικής. Η μέση τιμή είχε αρχικά σχέση με αστρονομικές παρατηρήσεις. Ο Tycho Brahe προσφεύγει, τον $16^{ο}$ αιώνα, σε ένα μέσο αριθμητικό, για να απαλείψει σφάλματα παρατηρήσεων. Το 1632 ο Γαλιλαίος εισάγει μια προβληματική που εντάσσεται στο πεδίο της θεωρίας σφαλμάτων. Ο De Moivre είναι αυτός που χρησιμοποιεί πρώτος την κωδωνοειδή καμπύλη και καταστρώνει τον τύπο που περιγράφει το νόμο της κανονικής κατανομής ή νόμο των Gauss-Laplace, ο οποίος κατά τον 19^{ο} αιώνα θα ονομαστεί και νόμος των σφαλμάτων. Από τη θεωρία σφαλμάτων προκύπτει ότι όσο πιο μεγάλος είναι ο αριθμός των παρατηρήσεων, τόσο λιγότερο κυμαίνεται η μέση τιμή τους.

Η κανονικότητα των φυσικών μετρήσεων σε συνδυασμό με την κανονικότητα των κοινωνικών εκδηλώσεων, που αποκαλύπτεται από τη διαδιδόμενη συσσώρευση στατιστικών στοιχείων στις αρχές του 19^{ου} αιώνα, οδηγεί τον αστρονόμο και στατιστικό Quetelet στη θεμελίωση της Κοινωνικής Φυσικής. Κινούμενος στο πνεύμα αυτό προχωράει στη διατύπωση της έννοιας του μέσου ανθρώπου. Η έννοια αυτή θα υιοθετηθεί σε πάρα πολύ μεγάλο βαθμό, αλλά στα τέλη του 19^{ου} αιώνα ορθώνονται οι πρώτες ενστάσεις. Ο Durkheim, ο πρώτος κοινωνιολόγος που βασίζεται σε εκτεταμένες στατιστικές, υποστηρίζει ότι είναι θεμελιώδες σφάλμα να συγχέεται ο συλλογικός τύπος μιας κοινωνίας με το μέσο τύπο ανθρώπων που την αποτελούν.

Ο Galton δεν εμπιστευόταν τις μέσες τιμές και στα πλαίσια των ενασχολήσεων του ενδιαφερόταν για τις διαφορές μεταξύ των όντων, άρα και τις αποκλίσεις από τη μέση τιμή. Σε αντιδιαστολή με το νόμο της κανονικής κατανομής χρησιμοποιεί αυτό που ονομάζει «νόμο των παρεκκλίσεων από τη μέση τιμή». Είναι αυτός που μετατόπισε το ενδιαφέρον από τη μέση τιμή. Οι εργασίες του Galton καταλήγουν σε μια ανακάλυψη που καταστρέφει το σχήμα του Quetelet, δηλαδή ότι μια κανονική κατανομή αντί να είναι ένδειξη ομοιογένειας, μπορεί να προκύπτει ως συνιστώσα πολλών διαφορετικών κατανομών που αντιστοιχούν σε επιμέρους πληθυσμούς με μέσες τιμές πολύ διαφορετικές μεταξύ τους. Έτσι, η μέση τιμή μιας κανονικής κατανομής γίνεται ένα κινούμενο κέντρο βάρους που εξαρτάται από τη σύνθεση των εσωτερικών δυνάμεων των κοινωνιών. Η διαπίστωση αυτή οδηγεί τον Galton στην ανακάλυψη της παλινδρόμησης.

Σε σχέση με τις δειγματοληπτικές έρευνες, ο Kiaer οργανώνει το 1894 στη Νορβηγία την πρώτη παρόμοια έρευνα με άτομα που ερωτήθηκαν για τα επαγγέλματα, τα εισοδήματα, τις δαπάνες, τις ημέρες αργίας, το γάμο και τον αριθμό παιδιών τους. Μέχρι το 1930 χρησιμοποιείται στην Ευρώπη η αντιπροσωπευτική μέθοδος για σκοπούς διαχείρισης των κοινωνικών προβλημάτων από το κράτος. Μετά το 1930 διενεργούνται στις Η.Π.Α. δειγματοληπτικές έρευνες με θέμα μελέτες αγοράς καταναλωτικών αγαθών και εκλογικές προβλέψεις.

Όσον αφορά, τέλος, την εμπλοκή των στατιστικών στο θέμα της ταξινόμησης, αυτοί,

από τη μια μεριά, ως εμπλεκόμενοι στις διοικητικές καταγραφές, εργάζονται για τη συγκρότηση και τον ορισμό των κατηγοριών και την κωδικογράφηση των ειδικών περιπτώσεων, σε μια προοπτική που παρουσιάζει συγγένεια με αυτή των ειδικών του δικαίου και της διοίκησης. Από την άλλη μεριά, ως διερμηνείς των ίδιων τους των μορφοποιήσεων, επιχειρούν να συναγάγουν, μέσω όλο και πιο πολύπλοκων μαθηματικών κατασκευών, την ύπαρξη υπολανθανουσών κατηγοριών, οι οποίες αποκαλύπτονται από στατιστικές κανονικότητες ή από ειδικές μορφές κατανομής. Αργότερα, η Παραγοντική Ανάλυση των ψυχομετρών, που καταλήγει σε μια γενική νοημοσύνη, ή οι μέθοδοι της Ταξινόμησης, που προκύπτουν από την Ανάλυση Δεδομένων, έχουν τον ίδιο προσανατολισμό.

1.2 Βασικές Έννοιες Στατιστικής

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

Πληθυσμοί και Δείγματα

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

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

Στην πράξη, ένα δείγμα είναι αντιπροσωπευτικό του πληθυσμού όταν έχει χρησιμοποιηθεί η διαδικασία της τυχαίας δειγματοληψίας (random sampling) για την απόκτηση του. Η τυχαία δειγματοληψία απαιτεί κάθε υποκείμενο του πληθυσμού να έχει την ίδια πιθανότητα να επιλεγεί. Όπως και οι πληθυσμοί, ένα δείγμα μπορεί να είναι αντίστοιχα από πολύ μικρό έως και πολύ μεγάλο. Εύκολα γίνεται αντιληπτό ότι όσο μεγαλύτερο είναι το δείγμα, τόσο πιο αντιπροσωπευτικό θα είναι, καθώς αυξάνεται ο αριθμός των υποκειμένων που επιλέγονται από τον πληθυσμό.

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

Μεταβλητές

Η έννοια της μεταβλητής είναι μια έννοια των Μαθηματικών, που τη δεχόμαστε αξιωματικά χωρίς απόδειξη. Είναι ένα γράμμα που συμβολίζει ένα τυχαίο στοιχείο ενός συνόλου και χρησιμεύει για να δηλωθεί μια κοινή ιδιότητα των στοιχείων του. Χαρακτηρίζεται δηλαδή μια ποσότητα η οποία μπορεί να αλλάζει, κινούμενη εντός ενός συνόλου, το οποίο ονομάζεται πεδίο ορισμού της μεταβλητής. Με τις μεταβλητές κοινώς εννοούμε τα χαρακτηριστικά εκείνα, ως προς τα οποία εξετάζουμε έναν πληθυσμό. Οι δυνατές τιμές που μπορεί να πάρει μια μεταβλητή λέγονται τιμές της μεταβλητής.

1.3 Δειγματοληψία

Δειγματοληψία στη στατιστική είναι η τεχνική της επιλογής ενός μέρους του πληθυσμού (το οποίο ονομάζεται δείγμα), τέτοιο ώστε τα στατιστικά στοιχεία του δείγματος να ισχύουν, με ανεκτά σφάλματα, σε ολόκληρο τον πληθυσμό. Θεωρητικά ένα τέτοιο δείγμα είναι το τυχαίο δείγμα. Η δειγματοληψία προτιμάται από την απογραφή (η παρατήρηση όλου του πληθυσμού), γιατί ο όγκος των παρατηρήσεων, άρα και των υπολογισμών περιορίζεται δραστικά ή είναι δύσκολο οικονομικά ή πρακτικά ή αδύνατο να γίνει απογραφή. Για να είναι αξιόπιστα τα αποτελέσματα, το δείγμα πρέπει να είναι αντιπροσωπευτικό, με απλά λόγια πρέπει να είναι μια μικρογραφία του πληθυσμού.

Η αξιοπιστία των αποτελεσμάτων αυξάνεται όταν η σχεδίαση προβλέπει αντιπροσωπευτικό δείγμα πληθυσμού και βασίζεται σε επιστημονικά τεκμηριωμένες μεθόδους. Η επιχείρηση καλείται να καταστρώσει ένα σχέδιο δειγματοληψίας, το οποίο προϋποθέτει τρεις αποφάσεις:

- <u>Δείγμα</u>. Η έρευνα marketing πρέπει να ορίσει τον πληθυσμό - στόχο από τον οποίο θα ληφθεί το δείγμα.
- <u>Μέγεθος δείγματος</u>, δηλαδή πόσα άτομα πρέπει να περιληφθούν στο δείγμα. Τα μεγάλα δείγματα δίνουν πιο αξιόπιστα αποτελέσματα.
- <u>Διαδικασία δειγματοληψίας</u>. Για την επιλογή ενός αντιπροσωπευτικού δείγματος, πρέπει να ληφθεί ένα δείγμα πληθυσμού που τα στοιχεία θα βοηθήσουν στην επίλυση του προβλήματος.

Το μέγεθος ενός δείγματος μπορεί να έχει σημασία στην αξιοπιστία του αλλά το πόσο
αντιπροσωπευτικό είναι το δείγμα είναι ένα θέμα μεγαλύτερης σημασίας. Πολλές φορές είναι στην κρίση του ερευνητή να χρησιμοποιεί δείγμα που δεν είναι στατιστικά αντιπροσωπευτικό αλλά δείγμα που να του δίνει τις σχετικές πληροφορίες

μάρκετινγκ. Το μέγεθος ενός προϋπολογισμού μίας επιχείρησης μπορεί ακόμη να επηρεάσει το μέγεθος του δείγματος.

1.3.1 Μέθοδοι Δειγματοληψίας

Δειγματοληψία με πιθανότητα (Probability sampling)

Στη δειγματοληψία με πιθανότητα το δείγμα είναι ελεγχόμενο, εξετάζεται σύμφωνα με τους νόμους των πιθανοτήτων και δίνεται η δυνατότητα να υπολογίσουμε το σφάλμα εκτίμησης (table).

Μέθοδοι δειγματοληψίας με πιθανότητα

1. Απλή τυχαία δειγματοληψία (simple random sampling)
2. Συστηματική δειγματοληψία (Systematic sampling)
3. Κατά συστάδες (cluster sampling)
4. Στρωματοποιημένη δειγματοληψία (stratified sampling)

Δειγματοληψία χωρίς πιθανότητα (Non-Probability Sampling)

Σε αυτή την μέθοδο δειγματοληψίας, δεν είναι δυνατή η δειγματοληψία με πιθανότητα. Η δειγματοληψία δεν είναι γενικεύσιμη και δεν μπορούμε να υπολογίσουμε το σφάλμα εκτίμησης. Χρησιμοποιείται για εφαρμογές έρευνας που θέλουν να εξάγουν άμεσα τα συμπεράσματά τους.

1.4 Κατηγορίες Μεταβλητών και Δεδομένων

Ουσιαστικά, αυτό που ενδιαφέρει έναν ερευνητή είναι να απαντήσει σε ερωτήματα που σχετίζονται με έναν ή περισσότερους πληθυσμούς, έχοντας στη διάθεση του κάποιο αριθμό δειγμάτων. Είναι φανερό ότι τις περισσότερες φορές ο στόχος της έρευνας είναι να δημιουργήσει μία σχέση αιτίου-αποτελέσματος ανάμεσα σε δύο μεταβλητές (variables). Με άλλα λόγια, ο στόχος σε μία έρευνα είναι να αποδειχθεί ότι οι αλλαγές που εμφανίζονται στην τιμή μιας μεταβλητής οφείλονται στην αλλαγή της τιμής μιας άλλης μεταβλητής.

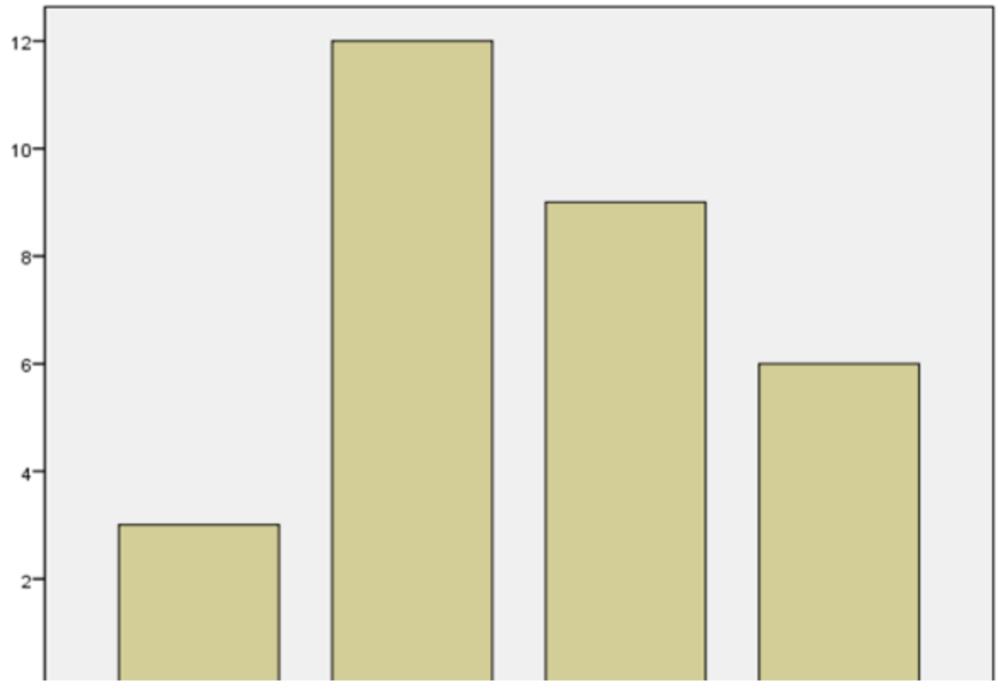
Για να γίνει αυτό εφικτό, η επιστημονική έρευνα χρησιμοποιεί δείγματα και προσεκτικά σχεδιασμένα πειράματα, όπου ο ερευνητής μπορεί να ελέγχει την τιμή μιας μεταβλητής και ταυτόχρονα να παρατηρεί τις τιμές της άλλης μεταβλητής, χωρίς η διαδικασία του πειράματος να επηρεάζεται από εξωτερικούς παράγοντες ή μεταβλητές. Η μεταβλητή που ελέγχει ο ερευνητής ονομάζεται ανεξάρτητη μεταβλητή (independent variable), ενώ αυτή που αποτελεί το αντικείμενο παρατήρησης ονομάζεται εξαρτημένη μεταβλητή (dependent variable). Το ζητούμενο για τον ερευνητή είναι να εξετάσει την επίδραση των διαφορετικών τιμών της ανεξάρτητης μεταβλητής στις τιμές της εξαρτημένης.

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

1.4.1 Παρουσίαση Στατιστικών Δεδομένων

Ραβδογράμματα

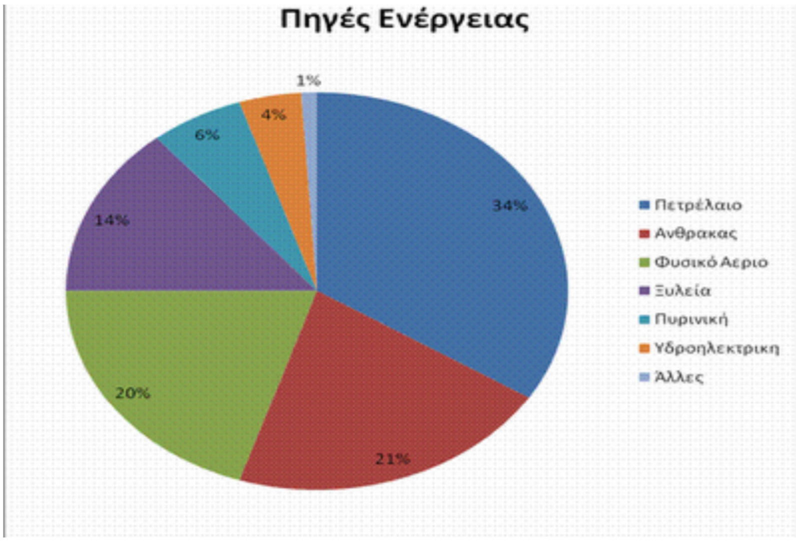
Τα δεδομένα ενός πίνακα συχνοτήτων δηλαδή τα δεδομένα ενός δείγματος για μια τ.μ. Χ που παίρνει τιμές σ΄ ένα σχετικά μικρό σύνολο διακεκριμένων τιμών (κατηγορίες ή αριθμητικές τιμές) μπορούν εύκολα να παρουσιαστούν σ΄ ένα πίνακα συχνοτήτων (frequency table). Ο πίνακας συχνοτήτων παρουσιάζει για κάθε τιμή xi της Χ τη συχνότητα εμφάνισής της fi, δηλαδή πόσες φορές εμφανίζεται η κάθε διακεκριμένη τιμή στο δείγμα. Μπορούν να παρασταθούν γραφικά με ένα ραβδόγραμμα, όπου το ύψος κάθε ράβδου παρουσιάζει τη συχνότητα (ή αθροιστική συχνότητα) ή σχετική συχνότητα (ή αθροιστική σχετική συχνότητα) για κάθε τιμή της μεταβλητής. Ένα παράδειγμα ραβδογράμματος συχνοτήτων φαίνεται στο επόμενο σχήμα.



Εικόνα : Παράδειγμα Ραβδογράμματος

Κυκλικό Διάγραμμα

Χρησιμοποιείται για τη γραφική παράσταση τόσο των ποιοτικών όσο και των ποσοτικών μεταβλητών όταν οι διάφορες τιμές της μεταβλητής είναι λίγες. Το κυκλικό διάγραμμα περιγράφει το ποσοστό του συνολικού αριθμού παρατηρήσεων που περιέχει κάθε κατηγορία, διαιρώντας ένα κύκλο σε κυκλικούς τομείς έτσι ώστε τα εμβαδά τους να είναι ανάλογα προς τις συχνότητες των αντίστοιχων κατηγοριών.



Φυλλόγραμμα

Το διάγραμμα αυτό δίνει την δυνατότητα ανασύστασης και ανάκλησης των μετρήσεων των αρχικών δεδομένων του δείγματος με ακρίβεια, πράγμα το οποίο δεν επιτυγχάνεται με το πρόγραμμα ή τους πίνακες συχνοτήτων.

Χρησιμοποιείται για την επεξεργασία μέτριου αριθμού παρατηρήσεων (περίπου 150). Η παρουσίαση του σχήματος μοιάζει με εκείνου του ιστογράμματος αλλά η τεχνική κατάρτισης δεν είναι η ίδια. Το φυλλογράφημα εμφανίζει τα δεδομένα σε όλο το εύρος των μετρήσεων που έχουν παρατηρηθεί, παρουσιάζει την συγκέντρωση των παρατηρήσεων (συχνότητες), δείχνει την μορφή της κατανομής, εμφανίζει τυχόν ακραίες παρατηρήσεις και επιτρέπει την επισήμανση της απουσίας συγκεκριμένων τιμών ή μετρήσεων.

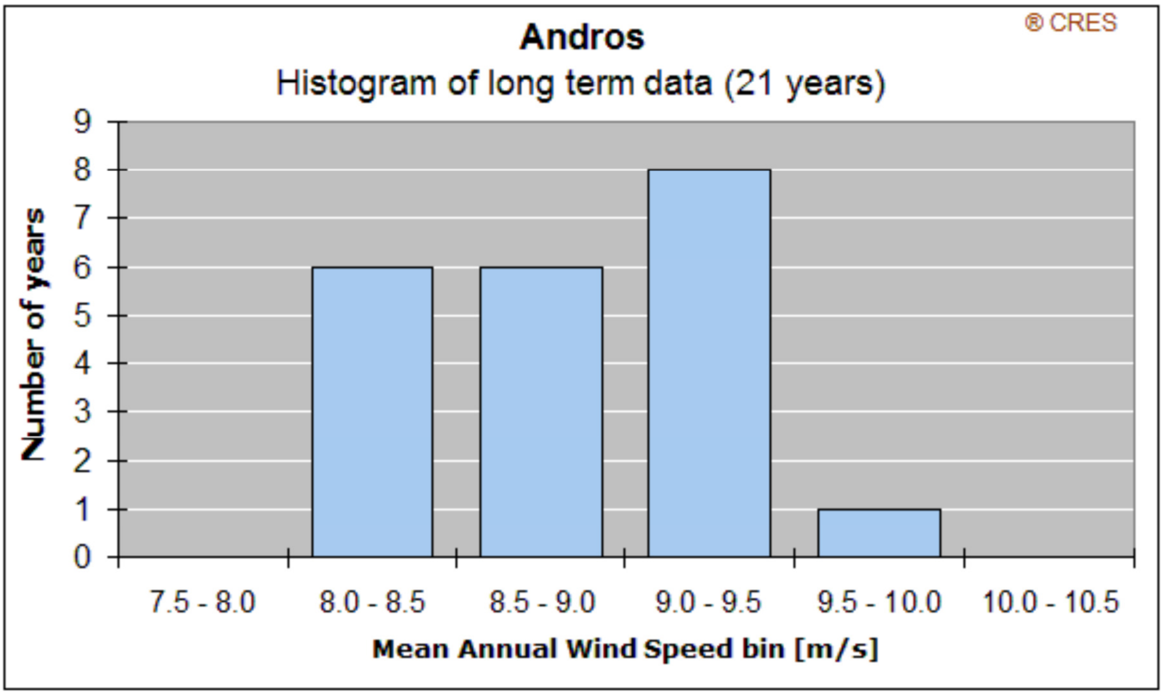
Παράδειγμα φυλλογράμματος

```
Frequency    Stem &  Leaf

     2,00        0 .  99
    24,00        1 .  000000011111122223333444
     9,00        1 .  555677789
    15,00        2 .  000011112233444
     8,00        2 .  56666889
     9,00        3 .  001122234
    11,00        3 .  55567788888
    10,00        4 .  0012223333
     2,00        4 .  78
     4,00        5 .  0022
     4,00        5 .  6688
     2,00        6 .  23
     1,00        6 .  5
     3,00        7 .  233
     1,00        7 .  8
     1,00        8 .  0
     2,00 Extremes    (>=9,3)
```

Ιστόγραμμα

Η γραφική παράσταση ενός δείγματος με ομαδοποιημένα δεδομένα γίνεται με το ιστόγραμμα. Στον οριζόντιο άξονα του συστήματος ορθογωνίων αξόνων σημειώνουμε τα όρια των κλάσεων. Στην συνέχεια, κατασκευάζουμε διαδοχικά ορθογώνια, το καθένα από τα οποία έχει βάση ίση με το πλάτος της κλάσης και ύψος ίσο με την συχνότητα (ή σχετική συχνότητα) της κλάσης αυτής.



1.5 Κατηγορίες Στατιστικών Μεθόδων

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

Η πρώτη κατηγορία, η περιγραφική στατιστική (descriptive statistics), περιλαμβάνει μεθόδους για την οργάνωση, απλοποίηση και συνοπτική παρουσίαση των δεδομένων. Αν και υπάρχουν πολλές τεχνικές που ανήκουν σε αυτήν την κατηγορία, η πιο διαδεδομένη είναι ο υπολογισμός της μέσης τιμής (mean) και της τυπικής απόκλισης (standard deviation). Τις περισσότερες φορές είναι αδύνατο να συλλέξουμε μετρήσεις για κάθε υποκείμενο του πληθυσμού, και για αυτό χρησιμοποιούμε ένα αντιπροσωπευτικό υποσύνολο του, τα δείγματα.

Η δεύτερη κατηγορία, η επαγωγική στατιστική (inferential statistics), περιλαμβάνει τεχνικές που επιτρέπουν την ανάλυση τέτοιων δεδομένων ώστε να γίνει εφικτή η εξαγωγή χρήσιμων και γενικεύσιμων συμπερασμάτων για τους πληθυσμούς, με βάση τις πληροφορίες που συλλέγονται από τα δείγματα.

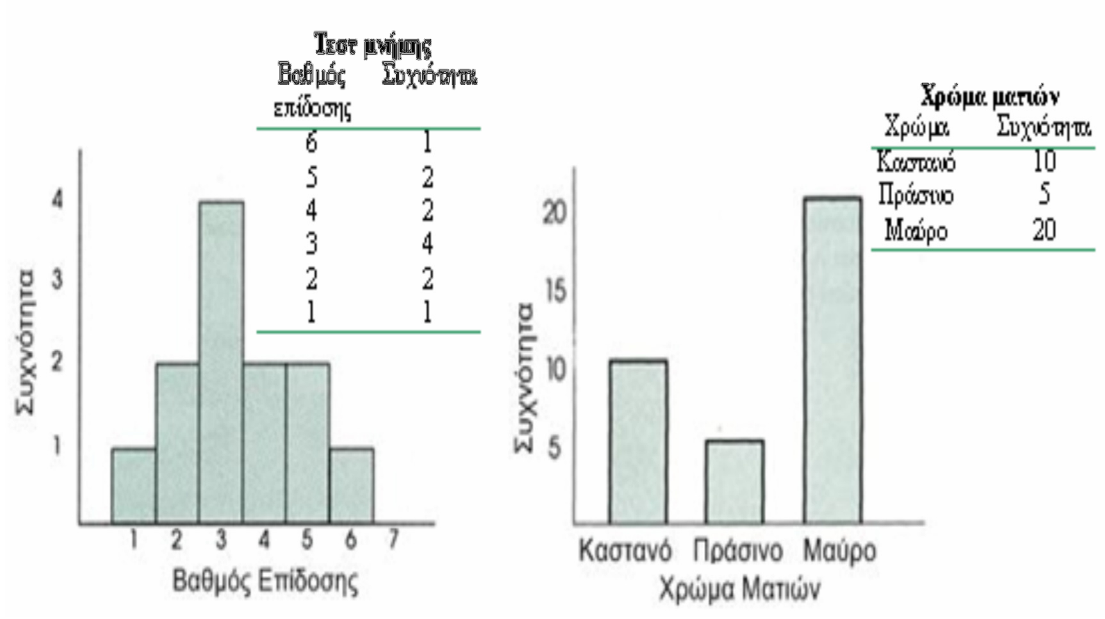
Ωστόσο, είναι προφανές ότι όσο αντιπροσωπευτικό και αν θεωρείται ένα δείγμα, δε μπορεί να παρέχει μια πιστή αναπαράσταση του πληθυσμού. Έτσι, πάντα θα υπάρχει μία ασυμφωνία, ή με άλλα λόγια ένα ποσοστό λάθους, ανάμεσα στο στατιστικό δείκτη που προκύπτει από το δείγμα, και την αντίστοιχη τιμή της παραμέτρου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

1.5.1 Περιγραφική Στατιστική

Όπως αναφέρθηκε και προηγουμένως, τα αποτελέσματα ενός πειράματος συνήθως δημιουργούν ένα μεγάλο αριθμό δεδομένων. Κρίνεται λοιπόν αναγκαία η εύρεση διαδικασιών, με τις οποίες τα αποτελέσματα αυτά μπορούν να οργανωθούν και να παρουσιαστούν με απλό και εύληπτο τρόπο. Αυτός ακριβώς είναι και ο στόχος των περιγραφικών στατιστικών δεικτών, δηλαδή να παρέχουν μεθόδους που απλοποιούν και διευκολύνουν την οργάνωση και παρουσίαση των αποτελεσμάτων.

Κατανομές Συχνότητας

Οι κατανομές συχνότητας (frequency distributions) αποτελούν μία από τις πιο γνωστές μεθόδους για την παρουσίαση ενός συνόλου δεδομένων. Οι κατανομές συχνότητας μπορούν να δομηθούν είτε σαν πίνακες είτε σαν γραφικές παραστάσεις (εικόνα 2). Και στις δύο περιπτώσεις, ο στόχος είναι να δοθεί μία πλήρης εικόνα για την κατανομή των δεδομένων στην κλίμακα μέτρησης. Ανάλογα με την κατηγορία των δεδομένων (ποσοτικά ή ποιοτικά), χρησιμοποιούνται διαφορετικές μορφές αυτών των κατανομών (ιστογράμματα ή ραβδογράμματα, αντίστοιχα).



Εικόνα : Κατανομές Συχνότητας υπό μορφή Πίνακα και Γραφικής Παράστασης για Ποσοτικά (ιστόγραμμα στα αριστερά) και Ποιοτικά (ραβδόγραμμα στα δεξιά) Δεδομένα

1.5.2 Στατιστικά Μέτρα

Μέτρα Κεντρικής Τάσης

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

Στις περισσότερες περιπτώσεις χρησιμοποιείται η μέση τιμή σαν μέτρο κεντρικής τάσης, καθώς δίνει την πιο αντιπροσωπευτική εικόνα για ένα σύνολο δεδομένων. Η μέση τιμή ορίζεται ως «το άθροισμα των παρατηρήσεων διά του πλήθους τους», δηλαδή ουσιαστικά πρόκειται για τον υπολογισμό του αριθμητικού μέσου όρου.

Ωστόσο, σε ορισμένες περιπτώσεις ο υπολογισμός της μέσης τιμής δεν είναι δυνατός (π.χ. υπάρχουν απροσδιόριστες τιμές στα δεδομένα, ή έχουμε διαθέσιμα ποιοτικά δεδομένα), ή δεν παρέχει την πιο αντιπροσωπευτική εικόνα (π.χ. η μέση τιμή μπορεί να επηρεαστεί σημαντικά από μία πολύ μεγάλη ή πολύ μικρή παρατήρηση). Σε αυτές τις περιπτώσεις χρησιμοποιείται σαν μέτρο κεντρικής τάσης η διάμεσος, που αντιστοιχεί σε εκείνη «την τιμή για την οποία το 50% των παρατηρήσεων έχει τιμή ίση ή μικρότερη». Για τον υπολογισμό της διαμέσου διατάσσουμε τις παρατηρήσεις σε αύξουσα σειρά, και αν το πλήθος τους είναι περιττό επιλέγουμε τη μεσαία παρατήρηση, ενώ αν είναι άρτιο η τιμή της διαμέσου δίνεται από το ημιάθροισμα των μεσαίων παρατηρήσεων.

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

Μέτρα Διασποράς

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

Το εύρος (range) είναι το πιο απλό μέτρο διασποράς και υπολογίζεται με τον ίδιο τρόπο τόσο για διακριτά ποσοτικά δείγματα, όσο και για τα συνεχή ποσοτικά δείγματα. Η τιμή του εύρους προκύπτει από τη διαφορά ανάμεσα στη μεγαλύτερη και τη μικρότερη τιμή των παρατηρήσεων στο δείγμα. Το ενδοτεταρτομοριακό εύρος (interquartile range) δίνει το εύρος που καλύπτουν τα μισά από τα δεδομένα που είναι πιο κοντά στην κεντρική τιμή (διάμεσο). Η διακύμανση και η τυπική απόκλιση αποτελούν το πιο αξιόπιστο και το πιο συνηθισμένο μέτρο διασποράς. Αυτά τα μέτρα διασποράς χρησιμοποιούν ως σημείο αναφοράς τη μέση τιμή, και λαμβάνουν υπόψη την απόσταση όλων των παρατηρήσεων από αυτήν. Η διάμεσος χωρίζει τα δεδομένα στα δύο. Μπορούμε να ορίσουμε άλλα σημεία χωρισμού του διατεταγμένου συνόλου τιμών που παίρνουμε από το δείγμα. Τέτοια σημεία είναι τα εκατοστιαία σημεία. Μια παρατήρηση καλείται το p-εκατοστιαίο σημείο (p–percentile) όταν ποσοστό παρατηρήσεων το πολύ p% είναι μικρότερες απ΄ αυτήν την παρατήρηση ($0 \leq p < 1$).

Η διάμεσος είναι το 50-εκατοστιαίο σημείο. Αλλά χαρακτηριστικά εκατοστιαία σημεία είναι αυτά που ορίζουν τέταρτα ή τεταρτομόρια (quartiles).

Κεφάλαιο 2ο: Μέτρα Διασποράς

2.1 Η έννοια της Διασποράς

Για να γίνει κατανοητός ο λόγος χρήσης των μέτρων διασποράς, παραθέτουμε το ακόλουθο παράδειγμα.

Παράδειγμα 1

Θεωρούμε μια επιχείρηση που παράγει ζυμαρικά. Για τη συσκευασία των προϊόντων της σε πακέτα των 500 gr θέλει να αγοράσει μια κατάλληλη μηχανή. Η επιλογή θα γίνει ανάμεσα σε δυο μηχανές Α και Β, οι οποίες πληρούν τις προδιαγραφές που έχει θέσει η επιχείρηση. Μια δειγματοληπτική χρήση των δυο μηχανών για 10 πακέτα ζυμαρικών έδωσε τα ακόλουθα αποτελέσματα σε γραμμάρια:

Πίνακας 1

Μηχανή Α	502	500	496	503	499	500	504	498	501	497
Μηχανή Β	500	498	510	495	500	492	501	497	503	504

Για τα δυο δείγματα υπολογίστηκαν οι μέσες τιμές τους, ώστε να εξεταστεί αν οι συσκευασίες των ζυμαρικών έχουν μέσο βάρος τα 500 gr.:

$$\bar{x}_A = \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{502+500+496+503+499+500+504+498+501+497}{10}$$

$$= \frac{5000}{10} = 500$$

και

$$\bar{x}_B = \frac{1}{10}\sum_{i=1}^{10} x_i$$

$$= \frac{500+498+510+495+500+492+501+497+503+504}{10}$$

$$= \frac{5000}{10} = 500$$

Όπως προκύπτει από τις τιμές των δειγμάτων, και οι δυο μηχανές έχουν μέση τιμή τα 500 γραμμάρια. Όμως οι συσκευασίες της μηχανής Α είναι όλες από 496 έως 504 γραμμάρια, ενώ της μηχανής Β από 492 μέχρι 510. Από τα διαστήματα αυτά προκύπτει ότι τα βάρη των πακέτων που προέρχονται από τη μηχανή Α βρίσκονται κοντά στη μέση τιμή, ενώ το αντίστοιχο διάστημα για τη μηχανή Β απλώνεται σε τιμές που βρίσκονται πιο μακριά από τη μέση τιμή. Γεννάται επομένως το ερώτημα, αν είναι αρκετή η γνώση της μέσης τιμής για την επιλογή του κατάλληλου μηχανήματος από την εταιρεία.

Όπως φαίνεται στο προηγούμενο παράδειγμα, τα μέτρα κεντρικής τάσης, όπως είναι η μέση τιμή, χρησιμοποιούνται ως αντιπρόσωποι μιας ομάδας δεδομένων. Η τιμή του μέτρου κεντρικής τάσης θεωρείται ότι είναι ο αριθμός γύρω από τον οποίο τείνουν να συγκεντρώνονται οι παρατηρήσεις του δείγματος. Όμως σε κάθε στατιστική έρευνα υπάρχει ένα διάστημα, μέσα στο οποίο διασπείρονται οι τιμές της τυχαίας μεταβλητής. Τα μέτρα εκείνα που δίνουν το βαθμό συγκέντρωσης των δεδομένων γύρω από κάποιο μέτρο θέσης, λέγονται μέτρα διασποράς ή διακύμανσης. Τα κυριότερα μέτρα διασποράς που θα αναπτύξουμε παρακάτω, είναι:

- Το εύρος του δείγματος
- Το ενδοτεταρτημοριακό εύρος
- Η διακύμανση και
- Η τυπική απόκλιση

2.2 Εύρος

Το εύρος (range) είναι το πιο απλό μέτρο διασποράς και υπολογίζεται με τον ίδιο τρόπο τόσο για διακριτά ποσοτικά δείγματα, όσο και για τα συνεχή ποσοτικά δείγματα. Η τιμή του εύρους προκύπτει από τη διαφορά ανάμεσα στη μεγαλύτερη και τη μικρότερη τιμή των παρατηρήσεων στο δείγμα, δηλαδή:

$$\boxed{R = \text{μεγαλύτερη τιμή της μεταβλητής } x - \text{μικρότερη τιμή της μεταβλητής } x}$$

Για τον ευκολότερο υπολογισμό του εύρους, οι τιμές του δείγματος διατάσσονται με αύξουσα σειρά, από τη μικρότερη προς τη μεγαλύτερη.

Παράδειγμα 2

Στο παράδειγμα της παραγράφου (2.1) παρατηρούμε ότι ισχύει:

$$R_A = 504 - 496 = 8$$

και

$$R_B = 510 - 492 = 18$$

Το στατιστικό μέγεθος του εύρους του δείγματος έχει το μειονέκτημα, ότι για τον υπολογισμό του λαμβάνονται υπόψη μόνο δυο τιμές του δείγματος, οι ακραίες. Στην περίπτωση όπου οι δυο ακραίες τιμές βρίσκονται σε μεγάλη απόσταση από τις υπόλοιπες, τότε η εικόνα της διασποράς είναι ψευδής. Παρά το σημαντικό μειονέκτημά του, το εύρος του δείγματος χρησιμοποιείται συχνά, κυρίως σε περιπτώσεις που οι ακραίες τιμές δεν απέχουν σημαντικά από τις υπόλοιπες τιμές του δείγματος.

2.3 Ενδοτεταρτημοριακό Εύρος

Προτού δώσουμε τον ορισμό του ενδοτεταρτημοριακού εύρους, θα πρέπει να ορίσουμε την έννοια των τεταρτημορίων. Τεταρτημόρια (quartiles) καλούνται οι τιμές της μεταβλητής οι οποίες χωρίζουν το σύνολο των τιμών του δείγματος σε τέσσερις ίσου πλήθους ομάδες. Για να βρούμε τις τιμές του πρώτου και του τρίτου τεταρτημορίου, διατάσουμε τις παρατηρήσεις του δείγματος κατά αύξουσα σειρά, και στη συνέχεια:

- Το πρώτο τεταρτημόριο Q_1 είναι η τιμή της μεταβλητής που είναι μεγαλύτερη από το ένα τέταρτο (25%) των παρατηρήσεων, και μικρότερη από τα τρία τέταρτα (75%) αυτών.

- Το τρίτο τεταρτημόριο Q_3 είναι η τιμή της μεταβλητής που είναι μεγαλύτερη από τα τρία τέταρτα (75%) των παρατηρήσεων, και μικρότερη από το ένα τέταρτο (25%) αυτών.

Μπορούμε πλέον να ορίσουμε το μέτρο διασποράς του ενδοτεταρτημοριακού εύρους, ως τη διαφορά μεταξύ του τρίτου και του πρώτου τεταρτημορίου:

$$\boxed{Q=Q3-Q1}$$

Το ενδοτεταρτημοριακό εύρος, μας δίνει το εύρος του 50% των κεντρικών παρατηρήσεων του δείγματος, απαλλαγμένο από το 25% των μικρότερων και το 25% των μεγαλύτερων παρατηρήσεων. Όσο μικρότερη είναι η τιμή του Q, τόσο μεγαλύτερη είναι η συγκέντρωση των τιμών της μεταβλητής. Το πλεονέκτημα του ενδοτεταρτημοριακού εύρους είναι το ότι δεν επηρεάζεται από τις ακραίες τιμές του δείγματος, αφού αυτές δεν λαμβάνονται υπ' όψιν στον υπολογισμό του, πράγμα που συμβαίνει με το εύρος μεταβολής.

2.3.1 Υπολογισμός του ενδοτεταρημοριακού εύρους σε ποσοτικά, διακριτά δείγματα

Για το υπολογισμό του πρώτου και του τρίτου τεταρτημορίου σε διακριτά ποσοτικά δείγματα, ισχύει:

Όταν οι n τιμές της τυχαίας μεταβλητής X, είναι διατεταγμένες κατά αύξουσα σειρά, δηλαδή ισχύει $x_1 \leq x_2 \leq \cdots \leq x_n$,τότε το πρώτο τεταρτημόριο Q_1, αντιστοιχεί στην τιμή που βρίσκεται στη θέση $\frac{n+1}{4}$,ενώ το τρίτο τεταρτημόριο Q_3, αντιστοιχεί στην τιμή που βρίσκεται στη θέση $\frac{3(n+1)}{4}$.

Παράδειγμα 3

Για το παράδειγμα της παραγράφου (2.1), διατάσουμε κατά αύξουσα σειρά τις τιμές του δείγματος της μηχανής A:

Πίνακας 2

Μηχανή A	496	497	498	499	500	500	501	502	503	504

Το πλήθος των παρατηρήσεων είναι $n_A = 10$. Επομένως για το πρώτο τεταρτημόριο έχουμε την τιμή που βρίσκεται στη θέση $\frac{10+1}{4} = \frac{11}{4} = 2{,}75$. Στην περίπτωση αυτή, όπου η θέση της τιμής του δείγματος δεν είναι ακέραιος αριθμός, για τον υπολογισμό του τεταρτημορίου παίρνουμε το ημιάθροισμα της αμέσως προηγούμενης και της αμέσως επόμενης τιμής. Δηλαδή έχουμε:

$$Q_1 = \frac{x_2+x_3}{2} = \frac{497+498}{2} = \frac{995}{2} = 497{,}5$$

Αντίστοιχα η τιμή του τρίτου τεταρτημορίου Q_3, είναι η τιμή του δείγματος που βρίσκεται στη θέση $\frac{3(10+1)}{4} = \frac{33}{4} = 7{,}75$. Όμοια με τον υπολογισμό του πρώτου τεταρτημορίου, έχουμε:

$$Q_3 = \frac{x_7 + x_8}{2} = \frac{501 + 502}{2} = \frac{1003}{2} = 501{,}5$$

Επομένως, το ενδοτεταρτημοριακό εύρος του δείγματος που προέρχεται από τη μηχανή Α, είναι:

$$Q = Q_3 - Q_1 = 501{,}5 - 497{,}5 = 4$$

2.3.2 Υπολογισμός ενδοτεταρτημοριακού εύρους σε συνεχή ποσοτικά δείγματα

Ο υπολογισμός του πρώτου και του τρίτου τεταρτημορίου σε ποσοτικά και συνεχή δεδομένα, γίνεται με τη βοήθεια των συχνοτήτων και των αθροιστικών συχνοτήτων. Σε ένα δείγμα n πλήθους, τα τεταρτημόρια Q_1 και Q_2 σε ομαδοποιημένα δεδομένα, δεχόμαστε ότι βρίσκονται στις θέσεις n4 και 3n4 αντίστοιχα. Έτσι μπορούμε να χρησιμοποιήσουμε για την εύρεσή τους, τους ακόλουθους τύπους:

$$Q_1 = a_{i-1} + \left(\frac{n}{4} - N_{i-1}\right) \cdot \frac{c}{n_i}$$

και

$$Q_3 = a_{i-1} + \left(\frac{3n}{4} - N_{i-1}\right) \cdot \frac{c}{n_i}$$

όπου:

a_{i-1}το κατώτερο όριο της κλάσης που περιέχεται το αντίστοιχο τεταρτημόριο

N_{i-1}η αθροιστική συχνότητα της κλάσης που βρίσκεται αμέσως πριν την κλάση που περιέχει το τεταρτημόριο

c το εύρος της κλάσης που περιέχει το τεταρτημόριο

n_i η συχνότητα της κλάσης που περιέχει το τεταρτημόριο

n το πλήθος των παρατηρήσεων

Για την καλύτερη κατανόηση του υπολογισμού του ενδοτεταρτημοριακού εύρους σε συνεχή ποσοτικά δεδομένα, παραθέτουμε το ακόλουθο παράδειγμα:

Παράδειγμα 4

Σε μια εταιρεία κινητής τηλεφωνίας, έγινε καταμέτρηση του χρόνου διάρκειας των τηλεφωνικών κλήσεων στο κέντρο εξυπηρέτησης πελατών, για μια εβδομάδα. Καταγράφηκαν συνολικά 1200 τηλεφωνικές κλήσεις που ομαδοποιήθηκαν σε κλάσεις των 2 λεπτών. Οι τιμές του δείγματος παρουσιάζονται στον ακόλουθο πίνακα συχνοτήτων:

Πίνακας 3

Χρόνος διάρκειας κλήσεων x_i	**Αριθμός κλήσεων n_i**	**Αθροιστική Συχνότητα N_i**
[0, 2)	100	100
[2, 4)	200	300
[4, 6)	250	550
[6, 8)	250	800
[8, 10)	200	1.000
[10, 12)	100	1.100
[12, 14)	60	1.160
[14, 16)	20	1.180
[16, 18)	10	1.190

[18, 20)	10	1.200
Σύνολο	**1.200**	

Η θέση του πρώτου τεταρτημορίου δίνεται από τη σχέση:

$$\frac{n}{4} = \frac{1.200}{4} = 300$$

Δηλαδή η τιμή του Q_1 θα πρέπει να είναι τέτοια ώστε 300 παρατηρήσεις του δείγματος να είναι μικρότερες και 900 παρατηρήσεις μεγαλύτερες από αυτή. Η κλάση που περιέχει το πρώτο τεταρτημόριο είναι η δεύτερη και για τον υπολογισμό του έχουμε:

- κατώτερο όριο της κλάσης που περιέχει το Q_1:

  $a_{i-1} = a_1 = 2$

- η αθροιστική συχνότητα της προηγούμενης κλάσης:

  $N_{i-1} = N_4 = 800$

  το εύρος της δεύτερης κλάσης:

  $c=2$

- η συχνότητα της κλάσης που περιέχει το Q_1 :

  $n_5 = 200$

- το πλήθος των παρατηρήσεων:

  n=1.200

Επομένως η τιμή του πρώτου τεταρτημορίου είναι:

$$Q_1 = a_1 + \left(\frac{n}{4} - N_1\right) \cdot \frac{c}{n_2} =$$

$$= 2 + (300 - 200) \cdot \frac{2}{200} =$$

$$= 2 + 100 \cdot \frac{1}{100} \Leftrightarrow$$

$$\Leftrightarrow Q_1 = 3$$

Όμοια η θέση του τρίτου τεταρτημορίου είναι:

$$\frac{3n}{4} = \frac{3 \cdot 1.200}{4} = \frac{3.600}{4} = 900$$

Δηλαδή η τιμή Q_3 θα πρέπει να είναι τέτοια ώστε 900 παρατηρήσεις να είναι μικρότερες και 300 μεγαλύτερες από αυτή. Η κλάση που περιέχει το τρίτο τεταρτημόριο είναι η πέμπτη και για τον υπολογισμό του έχουμε:

- κατώτερο όριο της κλάσης που περιέχει το Q_3:

$$\alpha_{i-1} = a_4 = 8$$

- η αθροιστική συχνότητα της προηγούμενης κλάσης:

$$N_{i-1} = N_4 = 800$$

- το εύρος της δεύτερης κλάσης:

  c=2

- η συχνότητα της κλάσης που περιέχει το Q1 :

$$n_5 = 200$$

- το πλήθος των παρατηρήσεων:

  n=1.200

Μπορούμε τώρα να υπολογίσουμε την τιμή του τρίτου τεταρτημορίου, που είναι:

$$Q_3 = a_4 + \left(\frac{3n}{4} - N_4\right) \cdot \frac{c}{n_5} =$$

$$= 8 + (900 - 800) \cdot \frac{2}{200} =$$

$$= 8 + 100 \cdot \frac{1}{100} \Leftrightarrow$$

$$\Leftrightarrow Q_3 = 9$$

Επομένως, το ενδοτεταρτημοριακό εύρος του δείγματος των τηλεφωνικών κλήσεων στο κέντρο εξυπηρέτησης πελατών, της εταιρείας κινητής τηλεφωνίας είναι ίσο με:

$$Q = Q_3 - Q_1 = 9 - 3 \Leftrightarrow$$

$$Q = 6 \text{ λεπτά της ώρας.}$$

.

2.4 Διακύμανση και Τυπική Απόκλιση

Ένας άλλος τρόπος για τον υπολογισμό του πόσο διασπαρμένες είναι οι τιμές μιας μεταβλητής X, γύρω από τη μέση τιμή τους, θα μπορούσε να προκύψει αν αφαιρεθεί η μέση τιμή x από κάθε παρατήρηση και υπολογιστεί η μέση τιμή του αθροίσματος των διαφορών αυτών. Όπως μπορεί όμως εύκολα να αποδειχθεί, η τιμή του στατιστικού αυτού, για οποιοδήποτε πλήθος n-παρατηρήσεων, θα ήταν πάντα ίση με το μηδέν, αφού:

$$\frac{(x_1 - \bar{x}) + (x_2 - \bar{x}) + \cdots + (x_n - \bar{x})}{n} = \frac{(x_1 + x_2 + \cdots + x_n) - n \cdot \bar{x}}{n} =$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{n \cdot \bar{x}}{n} = \bar{x} - \bar{x} = 0$$

Επειδή λοιπόν πάντα ο μέσος, του αθροίσματος των διαφορών των παρατηρήσεων με τη μέση τιμή του δείγματος, είναι ίσος με το 0, ως ένα μέτρο διασποράς παίρνουμε το μέσο όρο των τετραγώνων των αποκλίσεων των παρατηρήσεων x_i από τη μέση τιμή τους x.

Δηλαδή για τον υπολογισμό αυτού του μέτρου διασποράς, υψώνουμε τις διαφορές κάθε τιμής του δείγματος με τη μέση τιμή τους, στο τετράγωνο. Το πηλίκο του αθροίσματος των τετραγώνων των παραπάνω αποκλίσεων, με το πλήθος των παρατηρήσεων ελαττωμένο κατά ένα, είναι το μέτρο διασποράς που ονομάζεται διακύμανση.

Το μεγάλο μειονέκτημα της διακύμανσης, είναι το ότι εκφράζεται σε μονάδες οι οποίες είναι τα τετράγωνα των αρχικών μονάδων. Για παράδειγμα, αν το εισόδημα είναι σε ευρώ, η διακύμανση εκφράζεται με μονάδα το ευρώ στο τετράγωνο, ομοίως, αν η μεταβλητή εκφράζεται σε εκατοστά, η διακύμανση εκφράζεται σε εκατοστά στο τετράγωνο κ.λ.π.. Για να έχουμε όμως ένα δείκτη ο οποίος να μετράει τη διασπορά και να εκφράζεται στις ίδιες μονάδες που εκφράζεται η μεταβλητή μας, παίρνουμε την τετραγωνική ρίζα της διακύμανσης. Το μέτρο αυτό ονομάζεται τυπική απόκλιση και είναι το μέτρο διασποράς που χρησιμοποιούμε κυρίως στην πράξη.

Όσο μεγαλύτερη είναι η τυπική απόκλιση, τόσο μεγαλύτερη είναι η διασπορά των παρατηρήσεων από το μέσο αριθμητικό. Η διακύμανση ενός δείγματος, συμβολίζεται με $\mathbf{s^2}$ και η τυπική απόκλιση με **s**.

2.4.1 Υπολογισμός της διακύμανσης και της τυπικής απόκλισης σε μη ομαδοποιημένες παρατηρήσεις

Υποθέτουμε ότι έχουμε τις n σε πλήθος παρατηρήσεις x1,x2,...,xn, των οποίων η μέση τιμή είναι ίση με x. Η διακύμανση των παρατηρήσεων αυτών, δίνεται από τον τύπο:

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Στο σημείο αυτό θα πρέπει να αναφέρουμε ότι στη βιβλιογραφία συναντάται για τη διακύμανση και ο τύπος:

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

αλλά δεν προτιμάται, αφού δίνει μεγαλύτερο σφάλμα εκτίμησης της διακύμανσης του πληθυσμού.

Ο τύπος της διακύμανσης, μπορεί με κατάλληλες πράξεις να πάρει και την ισοδύναμη μορφή:

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

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \bar{x}^2\right] =$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2\right] =$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right] \Leftrightarrow$$

$$\boxed{s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{(\sum x_i)^2}{n}\right]}$$

Η τυπική απόκλιση, δίνεται από τον τύπο:

$$\boxed{s = \sqrt{s^2}}$$

Δηλαδή η τυπική απόκλιση είναι η τετραγωνική ρίζα της διακύμανσης.

Παράδειγμα 5

Στο παραδείγματος 1 της παραγράφου 2.1, είδαμε ότι τα δείγματα δυο μηχανών συσκευασίας ζυμαρικών, είχαν μέγεθος nA=nB=10 και την ίδια μέση τιμή, xA=xB=500 γραμμάρια. Για τα δεδομένα του παραδείγματος αυτού, θα εξετάσουμε τη διακύμανση και την τυπική απόκλιση των δυο δειγμάτων:

Για την μηχανή Α έχουμε διακύμανση

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_A)^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \bar{x}_A)^2 =$$

$$= \frac{(496-500)^2 + (497-500)^2 + (498-500)^2 + (499-500)^2 + (500-500)^2}{9}$$

$$+$$

$$+ \frac{(500-500)^2 + (501-500)^2 + (502-500)^2 + (503-500)^2 + (504-500)^2}{9}$$

$$=$$

$$= \frac{4^2 + 3^2 + 2^2 + 1^2 + 0^2 + 0^2 + 1^2 + 2^2 + 3^2 + 4^2}{9} =$$

$$= \frac{16+9+4+1+0+0+1+4+9+16}{9} = \frac{60}{9} \Leftrightarrow$$

$$\Leftrightarrow s^2 \cong 6{,}6667$$

και η τυπική απόκλιση για τη μηχανή Α είναι:

$$s_A = \sqrt{s^2} = \sqrt{6{,}6667} \cong 2{,}582$$

Δηλαδή η μηχανή Α έχει μέση τιμή $\bar{x}_A = 500$γραμμάρια, με τυπική απόκλιση sA=2,582 γραμμάρια.

Αντίστοιχα για τη μηχανή Β η διακύμανση είναι:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_B)^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \bar{x}_B)^2 =$$

$$= \frac{(492-500)^2 + (495-500)^2 + (497-500)^2 + (498-500)^2 + (500-500)^2}{9}$$

$$+$$

$$+ \frac{(500-500)^2 + (501-500)^2 + (503-500)^2 + (504-500)^2 + (510-500)^2}{9}$$

$$=$$

$$= \frac{8^2 + 5^2 + 3^2 + 2^2 + 0^2 + 0^2 + 1^2 + 3^2 + 4^2 + 10^2}{9} =$$

$$= \frac{64 + 25 + 9 + 4 + 0 + 0 + 1 + 9 + 16 + 100}{9} = \frac{228}{9} \Leftrightarrow$$

$$s^2 \cong 25,3333$$

και η τυπική απόκλιση για τη μηχανή Β είναι:

$$s_B = \sqrt{s^2} = \sqrt{25,3333} \cong 5,0332$$

Δηλαδή, ενώ και η μηχανή Β έχει μέση τιμή xB=500 γραμμάρια, έχει τυπική απόκλιση sB=5,0332 γραμμάρια.

Από την τυπική απόκλιση των δυο μηχανών, μπορούμε πλέον να απαντήσουμε στο αρχικό ερώτημα του παραδείγματος, ότι η καλύτερη επιλογή για την επιχείρηση παραγωγής των ζυμαρικών, είναι η μηχανή συσκευασίας Α, αφού αυτή έχει την μικρότερη τυπική απόκλιση στα πακέτα που συσκευάζει.

2.4.2 Υπολογισμός της διακύμανσης και της τυπικής απόκλισης σε ομαδοποιημένες παρατηρήσεις

Στην περίπτωση κατά την οποία οι παρατηρήσεις δίνονται ομαδοποιημένες σε πίνακα συχνοτήτων με k σε πλήθος τιμές της τυχαίας μεταβλητής, η διακύμανση υπολογίζεται με τον παρακάτω τύπο:

$$\boxed{s^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i (x_i - \bar{x})^2}$$

ενώ ο υπολογισμός της τυπικής απόκλισης δεν είναι διαφορετικός από αυτόν των μη ομαδοποιημένων παρατηρήσεων.

Ο τύπος της διακύμανσης, μπορεί με κατάλληλες πράξεις να πάρει και την ισοδύναμη μορφή:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i (x_i - \bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i (x_i^2 - 2x_i\bar{x} + \bar{x}^2) =$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{k} n_i x_i^2 - 2\bar{x}\sum_{i=1}^{k} n_i x_i + \sum_{i=1}^{k} n_i \bar{x}^2\right] =$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{k} n_i x_i^2 - 2n\bar{x}^2 + n\bar{x}^2\right] =$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{k} n_i x_i^2 - n\bar{x}^2\right] \Leftrightarrow$$

$$\boxed{s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} n_i x_i^2 - \frac{(\sum n_i x_i)^2}{n}\right]}$$

Παράδειγμα 6

Με χρήση των στοιχείων του παραδείγματος 4 της παραγράφου (2.4.2) θα υπολογίσουμε την μέση τιμή, τη διασπορά και την τυπική απόκλιση της χρονικής διάρκειας των τηλεφωνικών κλήσεων που δέχεται το τηλεφωνικό κέντρο της εταιρείας. Για το λόγο αυτό δημιουργούμε στον πίνακα των δεδομένων τις στήλες που χρειαζόμαστε για τον υπολογισμό των στατιστικών. Ακόμα σημειώνουμε ότι στα συνεχή δεδομένα, η τιμή του xi αντιστοιχεί στην κεντρική τιμή κάθε κλάσης, που υπολογίζεται ως το ημιάθροισμα των άκρων της κάθε κλάσης.

Πίνακας 4

Χρόνος διάρκειας κλήσεων	**xi**	**Αριθμός κλήσεων n_i**	**Αθροιστική Συχνότητα N_i**	**nixi**	**nixi2**
[0, 2)	1	100	100	100	100
[2, 4)	3	200	300	600	1.800
[4, 6)	5	250	550	1.250	6.250
[6, 8)	7	250	800	1.750	12.250
[8, 10)	9	200	1.000	1.800	16.200
[10, 12)	11	100	1.100	1.100	12.100
[12, 14)	13	60	1.160	780	10.140
[14, 16)	15	20	1.180	300	4.500
[16, 18)	17	10	1.190	170	2.890
[18, 20)	19	10	1.200	190	3.610
Σύνολο		**1.200**		**8.040**	**69.840**

Η μέση τιμή των τηλεφωνικών κλίσεων του τηλεφωνικού κέντρου είναι:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i \Leftrightarrow$$

$$\bar{x} = \frac{1}{1.200}\sum_{i=1}^{10} n_i x_i = \frac{8.040}{1.200} = 6{,}7$$

Δηλαδή η μέση διάρκεια κάθε κλήσης ήταν $\bar{x} = 6{,}7$ λεπτά.

Για τη διακύμανση των τηλεφωνικών κλήσεων του δείγματος έχουμε:

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} n_i x_i^2 - \frac{(\sum n_i x_i)^2}{n}\right] =$$

$$= \frac{1}{1.200-1}\left[\sum_{i=1}^{10} n_i x_i^2 - \frac{(\sum n_i x_i)^2}{1.200}\right] =$$

$$= \frac{1}{1.199}\left[69.840 - \frac{(8.040)^2}{1.200}\right] = \frac{1}{1.199}\left(69.840 - \frac{64.641.600}{1200}\right) =$$

$$\frac{69.840 - 53.868}{1.199} = 13{,}3211$$

Δηλαδή η διακύμανση των τηλεφωνικών κλήσεων του τηλεφωνικού κέντρου ήταν $s^2 = 13{,}3211$.

Τέλος η τυπική απόκλιση των τηλεφωνικών κλήσεων είναι:

$$s = \sqrt{s^2} = \sqrt{13{,}3211} \cong 3{,}65$$

Επομένως για τις τηλεφωνικές κλήσης που δέχεται το κέντρο εξυπηρέτησης πελατών της εταιρείας κινητής τηλεφωνίας, διαπιστώνουμε ότι έχουν μέση χρονική διάρκεια x=6,7 λεπτά και τυπική απόκλιση s=3,65 λεπτά.

2.4.3 Ιδιότητες της διακύμανσης και της τυπικής απόκλισης

Για τη διακύμανση και την τυπική απόκλιση των τιμών των δειγμάτων, ισχύουν οι ακόλουθες ιδιότητες:

1. Αν για τις τιμές μιας μεταβλητής, ενός δείγματος μεγέθους n ισχύει:

$$\boxed{x_1 = x_2 = \cdots = x_n}$$

τότε η διακύμανση, καθώς και η τυπική απόκλιση, είναι ίσες με το μηδέν. Δηλαδή η διακύμανση και η τυπική απόκλιση μιας σταθερής μεταβλητής, είναι πάντοτε ίση με το μηδέν.

2. Αν για όλες τις τιμές x_i και y_i δυο ίσου πλήθους τυχαίων μεταβλητών X και Y, ισχύει η σχέση:

$$y_i = c \cdot x_i$$

όπου c ένας σταθερός αριθμός, τότε η διακύμανση των τιμών της τυχαίας μεταβλητής Y ισούται με τη διακύμανση των τιμών της τυχαίας μεταβλητής X, πολλαπλασιαζόμενη με το τετράγωνο του αριθμού c, δηλαδή:

$$\boxed{s_Y^2 = c^2 \cdot s_X^2}$$

αφού ισχύει:

$$s_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \cdots + (y_n - \bar{y})^2}{n-1} =$$

$$= \frac{(c \cdot x_1 - c \cdot \bar{x})^2 + (c \cdot x_2 - c \cdot \bar{x})^2 + \cdots + (c \cdot x_n - c \cdot \bar{x})^2}{n-1} =$$

$$= \frac{c^2(x_1 - \bar{x})^2 + c^2(x_2 - \bar{x}) + \cdots + c^2(x_n - \bar{x})^2}{n-1} =$$

$$= c^2\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right] =$$

$$= c^2 \cdot s_X^2$$

Ακόμα, η τυπική απόκλιση των τιμών της τυχαίας μεταβλητής Y, είναι ίση με το γινόμενο της τυπικής απόκλισης της τυχαίας μεταβλητής X, με την απόλυτη τιμή του σταθερού αριθμού c:

$$\boxed{s_Y = |c| \cdot s_X}$$

αφού:

$$s_Y = \sqrt{s_Y^2} = \sqrt{c^2 \cdot s_X^2} = \sqrt{c^2} \cdot \sqrt{s_X^2} = |c| \cdot s_X$$

Παράδειγμα 7

Ένα εμπορικό κατάστημα ενδυμάτων, κατά την περίοδο των εκπτώσεων, προσφέρει όλα τα εμπορεύματά του με έκπτωση 30%. Η μέση τιμή πώλησης των εμπορευμάτων που έχει προς πώληση, χωρίς την έκπτωση, είναι x=100 ευρώ με τυπική απόκλιση sX=15 ευρώ. Ποια θα είναι η μέση τιμή και η τυπική απόκλιση των τιμών των εμπορευμάτων, μετά από την έκπτωση που προσφέρεται;

Για τον υπολογισμό της μέσης τιμής και της τυπικής απόκλισης των τιμών των εμπορευμάτων, μετά από την έκπτωση, ορίζουμε την τυχαία μεταβλητή Y οι τιμές της οποίας προκύπτουν από τις τιμές της τυχαίας μεταβλητής X, πολλαπλασιαζόμενες επί του 100%-30%=70%. Επομένως, για κάθε τιμή yi της τυχαίας μεταβλητής Y, ισχύει η σχέση:

$$y_i = 70\% \cdot x_i$$

Η μέση αξία των εμπορευμάτων μετά από την έκπτωση, θα είναι:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i =$$

$$= \frac{1}{n}\sum_{i=1}^{n} 70\% \cdot x_i = 70\% \cdot \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) =$$

$$= 70\% \cdot \bar{x} = 70\% \cdot 100 \Leftrightarrow$$

$$\bar{y} = 70$$

και η τυπική απόκλιση:

$$s_Y = 70\% \cdot s_X = 70\% \cdot 15 \Leftrightarrow$$

$$s_Y = 10{,}50$$

Δηλαδή η μέση τιμή πώλησης των εμπορευμάτων κατά τη διάρκεια των εκπτώσεων είναι 70 ευρώ, με τυπική απόκλιση 10,50 ευρώ.

3. Αν για όλες τις τιμές x_i και y_i δυο ίσου πλήθους τυχαίων μεταβλητών X και Y, ισχύει η σχέση:

$$y_i = x_i + a$$

όπου α σταθερός αριθμός, τότε η διακύμανση των τιμών της τυχαίας μεταβλητής Y είναι ίση με αυτή των τιμών της μεταβλητής X:

$$\boxed{s_Y^2 = s_X^2}$$

αφού:

$$s_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \cdots + (y_n - \bar{y})^2}{n-1} =$$

$$= \frac{[(x_1 + a) - (\bar{x} + a)]^2 + [(x_2 + a) - (\bar{x} + a)]^2 + \cdots + [(x_n + a) - (\bar{x} + a)]^2}{n-1} =$$

$$= \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1} = s_X^2$$

Όμοια και οι τυπικές αποκλίσεις των τιμών των δυο τυχαίων μεταβλητών X και Y, είναι και αυτές ίσες μεταξύ τους:

$$\boxed{s_Y = s_X}$$

4. Θεωρούμε ότι έχουμε k σε πλήθος ομάδες δεδομένων για την ίδια τυχαία μεταβλητή X, με πλήθος παρατηρήσεων n1, n2, ..., nk, με διακυμάνσεις s1, s2, ..., sk και μέσες τιμές x1, x2, ..., xk. Στην περίπτωση αυτή, η συνολική διακύμανση όλων των υποπληθυσμών, δίνεται από τη σχέση:

$$s^2 = \frac{1}{n}\sum_{i=1}^{k} n_i s_i^2 + \frac{1}{n}\sum_{i=1}^{n} n_i(\bar{x}_i - \bar{x})^2$$

όπου:

- Το συνολικό πλήθος όλων των δεδομένων είναι:

$$n = n_1 + n_2 + \cdots + n_k$$

- Η μέση τιμή όλων των δεδομένων είναι:

$$\bar{x} = \frac{n_1 \cdot \bar{x}_1 + n_2 \cdot \bar{x}_2 + \cdots + n_k \cdot \bar{x}_k}{n_1 + n_2 + \cdots + n_k}$$

Παράδειγμα 8

Να υπολογιστεί η μέση τιμή x και η διακύμανση s2, του συνολικού δείγματος που χωρίστηκε σε δυο υποπληθυσμούς για τους οποίους γνωρίζουμε ότι:

- $n_1 = 40, \quad \bar{x}_1 = 7, \quad s_1^2 = 1{,}3$
- $n_2 = 60, \quad \bar{x}_2 = 6{,}5 \quad s_2^2 = 1{,}5$

Η κοινή μέση τιμή των δυο δειγμάτων είναι:

$$\bar{x} = \frac{n_1 \cdot \bar{x}_1 + n_2 \cdot \bar{x}_2}{n_1 + n_2} =$$

$$= \frac{40 \cdot 7 + 60 \cdot 6{,}5}{40 + 60} = \frac{280 + 390}{100} = 6{,}7$$

Η κοινή διακύμανση των δυο πληθυσμών δίνεται από τη σχέση:

$$s^2 = \frac{1}{n}\sum_{i=1}^{2} n_i \cdot s_i^2 + \frac{1}{n}\sum_{i=1}^{2} n_i(\bar{x}_i - \bar{x})^2 =$$

$$= \frac{(40 \cdot 1{,}3 + 60 \cdot 1{,}5)}{40 + 60} + \frac{40 \cdot (7 - 6{,}7)^2 + 60 \cdot (6{,}5 - 6{,}7)^2}{40 + 60} =$$

$$= \frac{52 + 90}{100} + \frac{40 \cdot 0{,}09 + 60 \cdot 0{,}04}{100} =$$

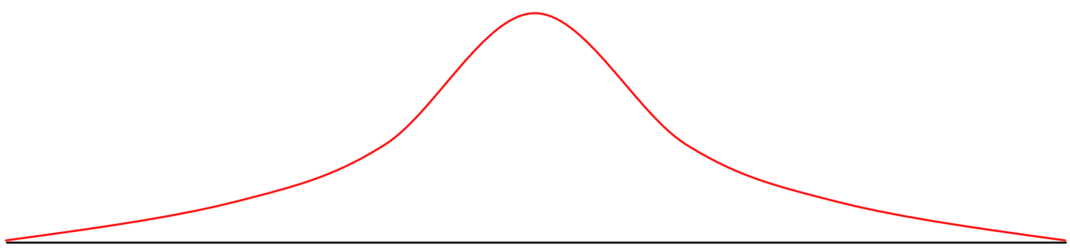
$$= 1{,}42 + \frac{3{,}6 + 2{,}4}{100} = 1{,}48$$

$$= 1{,}42 + \frac{3{,}6 + 2{,}4}{100} = 1{,}48$$

2.4.4 Τυπική απόκλιση στην κανονική κατανομή

Η τυπική απόκλιση είναι το σπουδαιότερο μέτρο (παράμετρος) διασποράς γιατί χρησιμοποιεί όλες τις τιμές της μεταβλητής και εκφράζεται με την ίδια μονάδα που εκφράζονται και οι τιμές. Η μέση τιμή $\bar{x}$ και η τυπική απόκλιση s χρησιμοποιούν όλες τις τιμές της τυχαίας μεταβλητής που υπάρχουν στο δείγμα. Η μέση τιμή δίνει το κέντρο της κατανομής και η τυπική απόκλιση δίνει ένα μέτρο της διασποράς της κατανομής.

Στην παράγραφο αυτή θα δούμε τη χρησιμότητα της μέσης τιμής και της τυπικής απόκλισης στην πιο γνωστή κατανομή συχνοτήτων, την κανονική κατανομή. Μια κατανομή συχνοτήτων λέμε ότι είναι κανονική όταν το διάγραμμα συχνοτήτων της έχει την κωδωνοειδή μορφή του ακόλουθου σχήματος:



Εικόνα 3: Καμπύλη κανονικής κατανομής

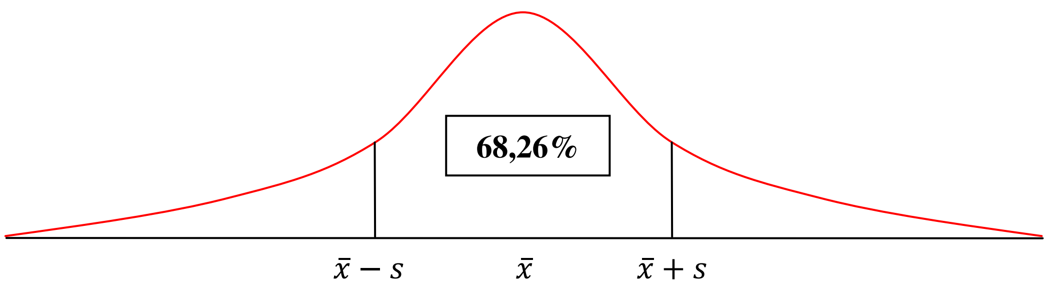
Όταν η κατανομή ενός δείγματος είναι κανονική με μέση τιμή $\bar{x}$ και τυπική απόκλιση s, τότε ισχύουν πάντοτε τα ακόλουθα:

- Το 68,26% των μεσαίων τιμών βρίσκεται μεταξύ του $\bar{x} - s$ και του $\bar{x} + s$
- Το 95,44% των μεσαίων τιμών βρίσκεται μεταξύ του $\bar{x} - 2 \cdot s$ και του $\bar{x} + 2 \cdot s$
- Το 99,7% των μεσαίων τιμών βρίσκεται μεταξύ του $\bar{x} - 3 \cdot s$ και του $\bar{x} + 3 \cdot s$
- Το 50% των μεσαίων τιμών βρίσκεται μεταξύ του $\bar{x} - \frac{2}{3} \cdot s$ και του $\bar{x} + \frac{2}{3} \cdot s$

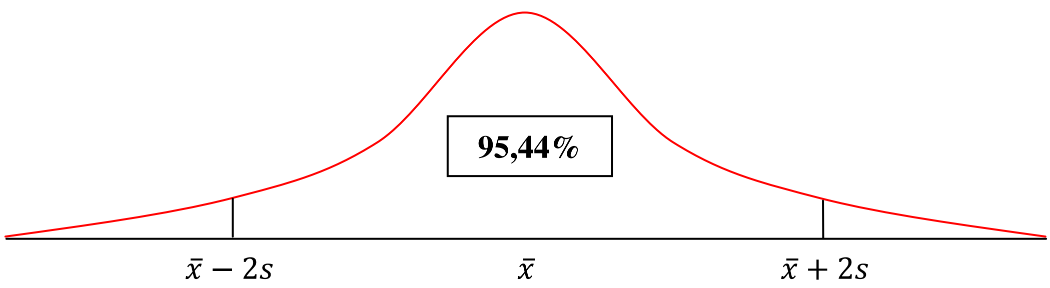
- Το εύρος του δείγματος ισούται περίπου με το εξαπλάσιο της τυπικής απόκλισης, δηλαδή:

$$R \cong 6 \cdot s$$

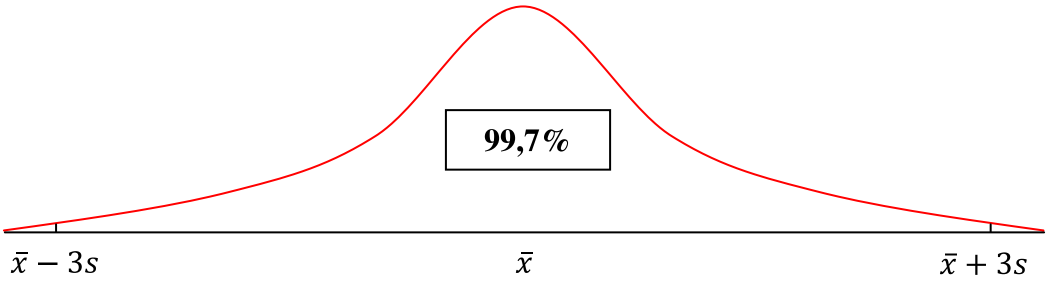
Τα συμπεράσματα αυτά αποδίδονται εποπτικά με τα επόμενα σχήματα:



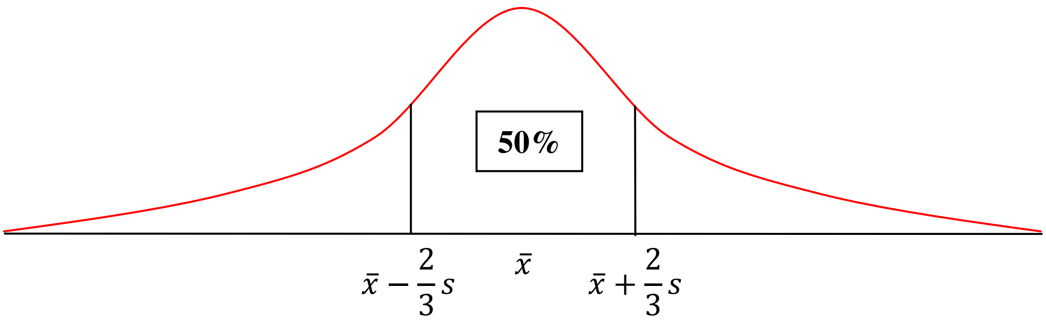
Εικόνα 4: Το διάστημα $[\bar{x} - s,\ \bar{x} + s]$ στην κανονική κατανομή



Εικόνα 5: Το διάστημα $[\bar{x} - 2s,\ \bar{x} + 2s]$ στην κανονική κατανομή



Εικόνα 6: Το διάστημα $[\bar{x} - 3s,\ \bar{x} + 3s]$ στην κανονική κατανομή



Εικόνα 7: Το διάστημα $\left[\bar{x} - \frac{2}{3}s,\ \bar{x} + \frac{2}{3}s\right]$ στην κανονική κατανομή

Παράδειγμα 9

Θεωρούμε όλους τους μηνιαίους μισθούς των 2.000 εργαζομένων σ' ένα εργοστάσιο και υπολογίζουμε ότι η μέση τιμή των μισθών είναι $\bar{x} = 1.015$ ευρώ, η τυπική απόκλιση $s = 174$ ευρώ και ότι η κατανομή των μισθών είναι κανονική.

Σύμφωνα με τα προηγούμενα, μπορούμε να αποφανθούμε με βεβαιότητα ότι:

i. Το 68,26% των εργαζομένων παίρνει μισθό 1.015 ± 174.

Δηλαδή μεταξύ 841 ευρώ και 1189 ευρώ αμείβονται οι 1.365 εργαζόμενοι του εργοστασίου.

ii. Το 95,44% των εργαζομένων παίρνει μισθό 1.015 ± 348.

Δηλαδή μεταξύ 667 ευρώ και 1363 ευρώ αμείβονται οι 1908 εργαζόμενοι του εργοστασίου.

iii. Το 99,7% των εργαζομένων παίρνει μισθό 1.015 ± 522.

Δηλαδή μεταξύ 493 ευρώ και 1537 ευρώ αμείβονται οι 1994 εργαζόμενοι του εργοστασίου.

iv. Το 50% των εργαζομένων παίρνει μισθό 1.015 ± 116.

Δηλαδή μεταξύ 899 ευρώ και 1131 ευρώ αμείβονται οι μισοί (1.000) εργαζόμενοι του εργοστασίου.

Στο σημείο αυτό θα πρέπει να τονιστεί ότι όλα τα παραπάνω ποσοστά ισχύουν όταν η κατανομή είναι κανονική. Όταν η κατανομή δεν είναι κανονική, τότε ο Chebyshev απέδειξε την ακόλουθη πρόταση, που είναι γνωστή και ως θεώρημα του Chebyshev:

«Για οποιοδήποτε σύνολο δεδομένων (πληθυσμού ή δείγματος) και για οποιαδήποτε σταθερά λ (με λ>1), ποσοστό τουλάχιστον $1 - \frac{1}{\lambda^2}$ των δεδομένων βρίσκεται εντός λ τυπικών αποκλίσεων εκατέρωθεν του μέσου όρου, δηλαδή μεταξύ $\bar{x} - \lambda s$ και $\bar{x} + \lambda s$».

Για παράδειγμα, αν $\lambda = 2$, έχουμε $1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} = 0{,}75$, δηλαδή ποσοστό 75% των τιμών της μεταβλητής, ανεξάρτητα από τη μορφή της κατανομής, βρίσκεται στο διάστημα $(\bar{x} - 2s, \bar{x} + 2s)$, ενώ για $\lambda = 3$, ποσοστό $1 - \frac{1}{3^2} = \frac{8}{9} = 0{,}89$, δηλαδή 89% των τιμών της μεταβλητής βρίσκεται στο διάστημα $(\bar{x} - 3s, \bar{x} + 3s)$.

Επομένως, σύμφωνα με το θεώρημα του Chebyshev, αν οι συσκευασίες της μηχανής Α του παραδείγματος 1 στην παράγραφο (2.1), έχουν μέσο βάρος $\bar{x} = 60$ και τυπική απόκλιση $s = 2{,}582$, τότε χωρίς καμία άλλη πληροφορία για την κατανομή του βάρους των συσκευασιών, μπορούμε να πούμε ότι το 75% των πακέτων των ζυμαρικών που προέρχονται από τη μηχανή αυτή, θα έχουν βάρος που θα περιέχεται στο διάστημα:

$$\begin{aligned}[(\bar{x} - 2s) - (\bar{x} + 2s)] &= [(500 - 2 \cdot 2{,}582), (500 + 2 \cdot 2{,}582)] \\ &= [(494{,}836), (505{,}164)]\end{aligned}$$

Δηλαδή το 75% των συσκευασιών της μηχανής Α έχουν βάρος από 494,8 μέχρι 505,2 γραμμάρια, ενώ το 89% των πακέτων έχουν βάρος που θα περιέχεται στο διάστημα:

$$[(\bar{x} - 3s), (\bar{x} + 3s)] = [(500 - 7{,}746), (500 + 7{,}746)] = [(492{,}254), (507{,}746)]$$

Δηλαδή το 89% των συσκευασιών, θα έχουν βάρος από 492,25 μέχρι 507,75 γραμμάρια.

Σημείωση:

Για λ=1 είναι $1 - \frac{1}{1^2} = 0$ δηλαδή το θεώρημα το Chebyshev δε μας δίνει πληροφορίες για το ποσοστό των τιμών της μεταβλητής που βρίσκεται στο διάστημα $(\bar{x} - S, \bar{x} + S)$.

2.5 Σύγκριση Μέτρων Διασποράς

Ανακεφαλαιώνοντας για τα μέτρα διασποράς και για να τα συγκρίνουμε μεταξύ τους, μπορούμε να εστιάσουμε στα ακόλουθα:

- Το εύρος είναι πολύ απλό στον υπολογισμό, χρησιμοποιείται στον έλεγχο ποιότητας και μπορεί να χρησιμοποιηθεί για την εκτίμηση της τυπικής απόκλισης, αλλά δεν θεωρείται αξιόπιστο μέτρο διασποράς, ακριβώς γιατί βασίζεται μόνο σε δύο ακραίες τιμές. Επίσης, δεν χρησιμοποιείται για περαιτέρω στατιστική ανάλυση.

- Το ενδοτεταρτημοριακό εύρος εκφράζει το εύρος του 50% των τιμών που βρίσκονται στο κέντρο της κατανομής με όρια το πρώτο και το τρίτο τεταρτημόριό της. Δε χρησιμοποιεί καθόλου τις πραγματικές τιμές της μεταβλητής αλλά τις θέσεις των τιμών αυτών. Είναι χρήσιμο και αντιπροσωπευτικό σε όλες τις περιπτώσεις.

- Η διακύμανση και η τυπική απόκλιση είναι τα σπουδαιότερα μέτρα διασποράς, γιατί για τον υπολογισμό τους λαμβάνονται υπόψη όλες οι παρατηρήσεις (τιμές) και έχουν μεγάλη εφαρμογή στη στατιστική συμπερασματολογία. Το μειονέκτημα της διακύμανσης είναι ότι δεν εκφράζεται στις ίδιες μονάδες με τις οποίες εκφράζονται οι τιμές της μεταβλητής, το οποίο όμως εξαλείφεται με την τυπική απόκλιση. Η τυπική απόκλιση δείχνει τη μεταβλητότητα των τιμών γύρω από τη μέση τιμή, λαμβάνοντας υπόψη όλες τις τιμές, και μπορεί να θεωρηθεί ότι αποτελεί ένα μέτρο τυπικής ή αντιπροσωπευτικής απόστασης. Μια μικρή τυπική απόκλιση δείχνει ότι οι τιμές είναι γενικά κοντά στη μέση τιμή, ενώ μια μεγάλη δείχνει ότι οι τιμές είναι απομακρυσμένες από τη μέση τιμή. Η τελευταία είναι το μέτρο που εφαρμόζεται συνήθως και είναι πολύ χρήσιμη στην Επαγωγική Στατιστική.

Αν θέλουμε να έχουμε μια σαφέστερη εικόνα για τα πλεονεκτήματα και τα μειονεκτήματα των κυριότερων μέτρων διασποράς, τότε μπορούμε να φτιάξουμε τον επόμενο συγκριτικό πίνακα:

Πίνακας 5: Σύγκριση μέτρων διασποράς

Πλεονεκτήματα	Μειονεκτήματα
Εύρος	
• **Είναι απλό στον υπολογισμό**	• Δεν θεωρείται αξιόπιστο μέτρο διασποράς, επειδή βασίζεται μόνο στις δύο ακραίες παρατηρήσεις
• **Χρησιμοποιείται αρκετά στον έλεγχο ποιότητας**	• Δε χρησιμοποιείται για περαιτέρω στατιστική ανάλυση
• **Μπορεί να χρησιμοποιηθεί για την εκτίμηση της τυπικής απόκλισης**	
• **Ενδοτεταρτημοριακό Εύρος**	
• **Δεν επηρεάζεται από τις ακραίες τιμές**	• Δεν λαμβάνει υπ' όψιν όλες τις τιμές του δείγματος παρά μόνο το 50% των κεντρικών τιμών
• **Υπολογίζεται όταν η κατανομή των δεδομένων είναι ανοικτή**	• Δε χρησιμοποιεί καθόλου τις πραγματικές τιμές της μεταβλητής αλλά τις θέσεις τους
• **Είναι χρήσιμο και αντιπροσωπευτικό σε όλες τις περιπτώσεις**	
• **Διακύμανση και τυπική απόκλιση**	
• **Λαμβάνονται υπόψη για τον υπολογισμό τους όλες οι παρατηρήσεις**	• Το κυριότερο μειονέκτημα στης διακύμανσης είναι ότι δεν εκφράζεται στις ίδιες μονάδες με τις τιμές
• **Έχουν μεγάλη εφαρμογή στη στατιστική συμπερασματολογία**	• (Το παραπάνω μειονέκτημα παύει να υπάρχει με τη χρησιμοποίηση της τυπικής απόκλισης)
• **Σε κανονικές κατανομές το 68,3%, το 95,4% και το 99,7% των παρατηρήσεων βρίσκονται στα διαστήματα και αντίστοιχα**	• Απαιτούνται περισσότερες αλγεβρικές πράξεις για τον υπολογισμό τους απ' ότι για τον υπολογισμό των άλλων μέτρων

Παρ' όλα αυτά, μια απλή περιγραφή των πλεονεκτημάτων, καθώς και των μειονεκτημάτων των μέτρων διασποράς, δεν είναι αρκετή για την επιλογή του κατάλληλου στατιστικού μέτρου. Υπάρχουν διάφοροι παράγοντες οι οποίοι επηρεάζουν τις τιμές ενός δείγματος με αποτέλεσμα διαφορετικά στατιστικά μεγέθη να θεωρούνται ως καταλληλότερα για κάθε μεταβλητή. Οι παράγοντες που επιδρούν στη μεταβλητότητα των τιμών και καθορίζουν την επιλογή του κατάλληλου μέτρου διασποράς είναι:

Οι ακραίες τιμές και έντονη ασυμμετρία.

Από τα μέτρα διασποράς που αναφέρθηκαν, το εύρος είναι αυτό που επηρεάζεται περισσότερο από την ύπαρξη ακραίων τιμών. Μια και μοναδική ακραία τιμή μπορεί να έχει δραματική επίδραση στο εύρος της μεταβλητής. Η τυπική απόκλιση και η διακύμανση, σε πολύ μικρότερο βαθμό, επηρεάζονται και αυτές από την ύπαρξη ακραίων τιμών στο δείγμα, αφού κάθε τιμή του συμμετέχει στον υπολογισμό τους με το τετράγωνο της απόκλισής της από τη μέση τιμή. Αντιθέτως το ενδοτεταρτημοριακό εύρος, επειδή αγνοεί το 25% των μικρότερων και το 25% των μεγαλύτερων τιμών των δεδομένων, κατά τον υπολογισμό του, δεν επηρεάζεται από την ύπαρξη ακραίων τιμών σε μια μεταβλητή. Γι' αυτό το λόγο, το ενδοτεταρτημοριακό εύρος θεωρείται το καλύτερο μέτρο διακύμανσης, όταν στο δείγμα υπάρχουν ακραίες τιμές ή μεγάλη ασυμμετρία στην κατανομή των δεδομένων.

Το μέγεθος του δείγματος.

Όταν το μέγεθος του δείγματος αυξάνεται, το εύρος έχει την τάση να αυξάνεται και αυτό, αφού κάθε καινούρια παρατήρηση μπορεί να πάρει τη θέση της μικρότερης ή της μεγαλύτερης τιμής των δεδομένων. Αυτή η σχέση ανάμεσα στη μεταβλητότητα και στο μέγεθος του δείγματος που παρουσιάζεται στο εύρος είναι λογικό να μην είναι επιθυμητή στην περιγραφή των δεδομένων. Τα υπόλοιπα μέτρα διασποράς είναι σχετικά ανεπηρέαστα από το μέγεθος του δείγματος με αποτέλεσμα να προσφέρουν καλύτερη προσέγγιση της μεταβλητότητας του.

Σταθερότητα της δειγματοληψίας.

Όταν σχηματίζονται διάφορες ομάδες δειγμάτων από τον ίδιο πληθυσμό, τότε αναμένεται η ύπαρξη μιας ομοιότητας μεταξύ αυτών των δειγμάτων. Οι μέσες τιμές των δειγμάτων αυτών πρέπει να είναι παρόμοιες, αφού όλα τα δείγματα αποτελούν μικρογραφίες του ίδιου πληθυσμού. Το ίδιο αναμένεται και για τις τιμές μεταβλητότητας των δειγμάτων. Όταν το μέτρο διασποράς που χρησιμοποιείται είναι η διακύμανση ή τυπική απόκλιση, τότε τα δείγματα έχουν παρόμοια μεταβλητότητα. Ακόμα, σταθερότητα κατά τη δειγματοληψία παρουσιάζει και το ενδοτεταρτημοριακό εύρος. Το εύρος, ωστόσο, είναι ένα μέτρο πολύ ασταθές από δείγμα σε δείγμα και δεν δίνει πάντοτε παρόμοιες τιμές, όταν οι ομάδες των δεδομένων προέρχονται από τον ίδιο πληθυσμό.

Κατανομή ανοικτών ορίων.

Όταν η μεγαλύτερη ή η μικρότερη τιμή των δεδομένων για μια τυχαία μεταβλητή δεν είναι σαφώς ορισμένες, τότε έχουμε μια κατανομή ανοικτών ορίων. Αυτό μπορεί να συμβεί, για παράδειγμα, όταν ένα άτομο αδυνατεί να ολοκληρώσει ένα έργο (στα χρονικά πλαίσια διεξαγωγής του πειράματος) και ο χρόνος που καταγράφεται από τον ερευνητή είναι ακαθόριστος. Όπως γίνεται εύκολα αντιληπτό, σε τέτοιες κατανομές δε μπορεί να υπολογιστεί ούτε η διακύμανση, ούτε η τυπική απόκλιση, αλλά ούτε και το εύρος. Το μοναδικό μέτρο διασποράς που μπορεί να υπολογιστεί στην περίπτωση αυτή, είναι το ενδοτεταρτημοριακό εύρος.

2.6 Συντελεστής Μεταβολής

Η τυπική απόκλιση, η οποία θεωρείται ως το κυρίως χρησιμοποιούμενο μέτρο για τη μέτρηση της διασποράς, εκφράζεται στις ίδιες μονάδες με τις οποίες εκφράζεται η τυχαία μεταβλητή και μας δίνει την απόλυτη διασπορά των τιμών της τυχαίας μεταβλητής από τη μέση τιμή.

Η χρησιμοποίηση όμως της τυπικής απόκλισης και των άλλων μέτρων διασποράς είναι σε αρκετές περιπτώσεις αδύνατη και σε άλλες περιπτώσεις περιορισμένη. Αυτό συμβαίνει όταν θέλουμε να συγκρίνουμε δύο δείγματα τα οποία εκφράζονται σε διαφορετικές μονάδες (μέτρα, κιλά, ευρώ, εκατοστά κ.λπ.), ή όταν οι μέσες τιμές δύο διαφορετικών τυχαίων μεταβλητών, έστω και αν εκφράζονται στις ίδιες μονάδες, διαφέρουν πάρα πολύ μεταξύ τους. Στις περιπτώσεις αυτές, τα μέτρα της απόλυτης διασποράς δεν μας εξυπηρετούν και χρησιμοποιούμε τη σχετική διασπορά. Το βασικό μέτρο της σχετικής διασποράς είναι ο συντελεστής μεταβλητότητας. Ο συντελεστής αυτός είναι ανεξάρτητος από τις μονάδες μέτρησης που χρησιμοποιούμε και, επομένως, επιτρέπει τη σύγκριση τόσο των ομοειδών όσο και των ετεροειδών δειγμάτων. Ο συντελεστής μεταβολής CV, δίνεται από το τύπο:

$$CV = \frac{s}{|\bar{x}|}$$

Δηλαδή, ο συντελεστής μεταβολής είναι το πηλίκο της τυπικής απόκλισης s

του δείγματος, προς το απόλυτο της μέσης τιμής αυτού και εκφράζει την τυπική απόκλιση ως ποσοστό επί τοις εκατό της μέσης τιμής $\bar{x}$ του δείγματος.

Ο συντελεστής μεταβολής είναι ένας καθαρός αριθμός, δηλαδή δεν εξαρτάται από τις μονάδες μέτρησης, συσχετίζει την τυπική απόκλιση με τη μέση τιμή των τιμών της μεταβλητής, εκφράζεται επί τοις εκατό και παριστάνει ένα μέτρο σχετικής διασποράς των τιμών (γι' αυτό ονομάζεται και σχετική τυπική απόκλιση) και όχι απόλυτης διασποράς, όπως έχουμε δει μέχρι τώρα. Εκφράζει δηλαδή τη μεταβλητότητα των δεδομένων απαλλαγμένη από την επίδραση της μέσης τιμής.

Συμπερασματικά μπορούμε να πούμε ότι όσο πιο μικρός είναι ο συντελεστής μεταβολής, τόσο μεγαλύτερη ομοιογένεια υπάρχει στις τιμές της μεταβλητής. Γενικά δεχόμαστε ότι ένα δείγμα τιμών μιας μεταβλητής θα είναι ομοιογενές, εάν ο συντελεστής μεταβολής δεν είναι μεγαλύτερος του 10%.

Θα πρέπει όμως να τονίσουμε ότι ο συντελεστής μεταβολής δεν ενδείκνυται να χρησιμοποιείται όταν η μέση τιμή είναι κοντά στο μηδέν.

Παράδειγμα 10

Για τις μηχανές Α και Β του παραδείγματος 1 στην παράγραφο (2.1), είδαμε ότι έχουν αντίστοιχα $\bar{x}_A = 500$, $s_A = 2{,}582$ και $\bar{x}_B = 500$, $s_B = 5{,}0332$. Ο συντελεστής μεταβολής για τις δυο μηχανές, είναι:

$$CV_A = \frac{s_A}{\bar{x}_A} = \frac{2{,}582}{500} \cong 0{,}0052 = 0{,}52\%$$

και

$$CV_B = \frac{s_B}{\bar{x}_B} = \frac{5{,}0332}{500} \cong 0{,}01 = 1\%$$

Όπως βλέπουμε από τους συντελεστές μεταβολής των δειγμάτων που πήραμε από τις δυο μηχανές, ενώ και τα δυο είναι ομοιογενή, αφού έχουν συντελεστή μεταβολής μικρότερο του 10%, το δείγμα από τη μηχανή Α παρουσιάζει μεγαλύτερη ομοιογένεια από αυτό της μηχανής Β.

Κεφάλαιο 3^{ο}: Στατιστική και επιχειρήσεις

3.1 Η σημασία της Στατιστικής στο χώρο των επιχειρήσεων

Όπως προκύπτει από τα προηγούμενα κεφάλαια, είναι πλέον προφανές, ότι, αν θέλουμε να έχουμε μία πληρέστερη και ασφαλέστερη εικόνα σχετικά με ένα πρόβλημα ή με ένα θέμα που εξετάζουμε, επιβάλλεται η χρησιμοποίηση των στατιστικών μέτρων θέσης μαζί με αυτά της διασποράς.

Για παράδειγμα, αν ένας οικονομικός ερευνητής μετά τη μελέτη των οικονομικών δεδομένων που έκανε για λογαριασμό ενός οργανισμού ή μιας επιχείρησης, διαπιστώσει την ύπαρξη υψηλού βαθμού διακύμανσης, αυτό θα πρέπει να τον προβληματίσει και να τον οδηγήσει σε περαιτέρω έρευνα και μελέτη των δεδομένων του. Τα μέτρα διασποράς, ως στατιστικά μέτρα που μας πληροφορούν πόσο διάσπαρτες (απλωμένες) γύρω από τα μέτρα θέσης είναι οι παρατηρήσεις μας, αποτελούν ένα επιπλέον εργαλείο στα χέρια των αναλυτών για να αντιληφθούν το βαθμό αντιπροσωπευτικότητας και αξιοπιστίας των μέτρων θέσης.

Στις επιχειρήσεις και στον οικονομικό χώρο γίνεται ευρύτατη χρησιμοποίηση των μέτρων διασποράς. Για παράδειγμα, ακούμε στην αγορά να γίνεται λόγος για διακύμανση των ποσοτήτων που πωλούνται από τις επιχειρήσεις. Τα τελευταία χρόνια πολύ συχνά ακούμε για το ύψος των διακυμάνσεων που παρουσιάζεται στις τιμές των μετοχών των εταιρειών που είναι στο Χρηματιστήριο και για τη διακύμανση στις συναλλαγματικές ισοτιμίες του Ευρώ σε σχέση με άλλα νομίσματα.

Στις επόμενες παραγράφους παραθέτουμε μερικές περιπτώσεις μελέτης (case study), ώστε να δούμε την εφαρμογή των όσων αναφέραμε στον τομέα των επιχειρηματικών δραστηριοτήτων.

3.2 Περίπτωση μελέτης αποτελέσματος μετά από αναδιοργάνωση επιχείρησης.

Στην παράγραφο αυτή παραθέτουμε ένα παράδειγμα εφαρμογής της στατιστικής στις επιχειρήσεις, για τον έλεγχο των αποτελεσμάτων που προέκυψαν από την αλλαγή (αναδιοργάνωση) στη λειτουργία της:

Παράδειγμα 11

Η Εταιρεία G and B Hellas – Neon έχει εγκαταστημένα 20 υποκαταστήματα (σημεία πώλησης των προϊόντων της) σε επιλεγμένες περιοχές, κατά το δυνατόν ισοδύναμες μεταξύ τους από πλευράς ευκαιριών και δυνατοτήτων ως προς το ετήσιο ύψος του κύκλου εργασιών τους (ετήσιου τζίρου).

Η εταιρεία προσέλαβε τη φετινή περίοδο ένα νέο γενικό διευθυντή, ο οποίος αναδιοργάνωσε την επιχείρηση και εφάρμοσε νέες επιστημονικές μεθόδους, στην οργάνωση και διοίκηση (management) καθώς και στον τομέα των πωλήσεων της εταιρείας. Μετά τη συμπλήρωση της νέας οικονομικής χρήσης ο γενικός διευθυντής αποφασίζει να μελετήσει τα έσοδα πωλήσεων των 20 υποκαταστημάτων. Έτσι θα αντλήσει συμπεράσματα, που θα τα συγκρίνει με αυτά της περσινής περιόδου, για να μπορέσει να εξάγει κάποια τελικά συμπεράσματα σχετικά με το τι προέκυψε μετά την αναδιοργάνωση της επιχείρησης.

Οι πωλήσεις (σε χιλιάδες ευρώ) των 20 υποκαταστημάτων της φετινής χρήσης ήταν:

105, 112, 115, 118, 123, 123, 124, 125, 127, 128,
132, 133, 134, 136, 138, 138, 142, 145, 149, 156

Ως προς την περσινή χρήση γνωρίζουμε ότι η μέση τιμή των πωλήσεων ήταν $\bar{x}_1 = 90.000$ ευρώ, και η διασπορά των πωλήσεων για τα 20 καταστήματα ήταν $s_1 = 10.000$ ευρώ. Ακόμη οι χαμηλότερες πωλήσεις υποκαταστήματος ήταν 70.000 ευρώ και οι υψηλότερες 140.000 ευρώ.

Ο γενικός διευθυντής συντάσσει αρχικά τον παρακάτω πίνακα και υπολογίζει τη νέα μέση τιμή και τη νέα τυπική απόκλιση:

Πίνακας 6: Πίνακας συχνοτήτων

Κλάσεις πωλήσεων σε χιλ. ευρώ	x_i	**Αριθμός υποκαταστημάτων** n_i	x^2	$n_i x_i$	$n_i x_i^2$
[100, 110)	105	1	11.025	105	11.025
[110, 120)	115	3	13.225	345	39.675
[120, 130)	125	6	15.625	750	93.750
[130, 140)	135	6	18.225	814	109.350
[140, 150)	145	3	21.025	435	63.075
[150, 160)	155	1	24.025	155	24.025
Σύνολο		**20**		**2.600**	**340.900**

Όπως προκύπτει από τα δεδομένα του έτους, η νέα μέση τιμή $\bar{x}_2$ των πωλήσεων της εταιρείας θα είναι:

$$\bar{x}_2 = \frac{1}{n}\sum_{i=1}^{20} n_i x_i = \frac{2.600}{20} = 130 \; \chi\iota\lambda. \varepsilon\upsilon\rho\omega$$

Επίσης η νέα τυπική απόκλιση s_2 είναι:

$$s_2^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} n_i x_i^2 - \frac{(\sum n_i x_i)^2}{n}\right] =$$

$$= \frac{1}{19}\left[340.900 - \frac{(2.600)^2}{20}\right] = \frac{340.900 - 338.000}{19} =$$

$$\cong 152{,}6316$$

και

$$s_2 = \sqrt{s_2^2} = \sqrt{152{,}6316} \cong 12{,}35$$

Δηλαδή η τυπική απόκλιση είναι $s_2 = 12$ χιλιάδες ευρώ.

Στο σημείο αυτό χρειάζεται να δείξουμε κάποια προσοχή. Δεν θα πρέπει να αφεθούμε στην πρώτη εικόνα των τυπικών αποκλίσεων και να πούμε ότι, επειδή η φετινή τυπική απόκλιση $s_2 = 12$ είναι μεγαλύτερη από την περσινή $s_1 = 10$, η διασπορά των εσόδων φέτος παρουσιάζει χειρότερη εικόνα. Για να είμαστε σίγουροι προχωράμε στον υπολογισμό και των αντίστοιχων CV_1 και CV_2 και έχουμε:

$$CV_1 = \frac{s_1}{\bar{x}_1} = \frac{10}{90} = 0{,}111 \text{ ή } CV_1 = 11{,}1\%$$

και

$$CV_2 = \frac{s_2}{\bar{x}_2} = \frac{12}{130} = 0{,}092 \text{ ή } CV_2 = 9{,}2\%$$

Όπως μπορούμε να δούμε όλες οι μεταβολές των στατιστικών μέτρων φαίνεται πως είναι ευνοϊκές τη νέα χρονιά. Συνοπτικά τα στατιστικά μεγέθη παρουσιάζονται στον ακόλουθο πίνακα:

Πίνακας 7: Τιμές των στατιστικών μέτρων των 2 διαχειριστικών χρήσεων

Στατιστικά μέτρα	**Προηγούμενη διαχειριστική χρήση (1)**	**Φετινή διαχειριστική χρήση (2)**	**Μεταβολή των χρήσεων 1 και 2**
I. Μέση τιμή $\bar{x}$	$\bar{x}_1 = 90$	$\bar{x}_2 = 130$	αύξηση
II. Διακύμανση s^2	$s_1^2 = 100$	$s_2^2 = 152$	αύξηση
III. Τυπική απόκλιση s	$s_1 = 10$	$s_2 = 12$	αύξηση
IV. Συντελεστής μεταβολής	$CV_1 = 11{,}1\%$	$CV_2 = 9{,}2\%$	μείωση

CV			
V. Εύρος μεταβολής Ε.Μ.	$E.M._1 = 70$	$E.M._2 = 51$	μείωση

Συμπέρασμα:

Από τη μέχρι τώρα μελέτη μπορούμε να πούμε ότι η επιχείρηση βελτίωσε την οικονομική της θέση, διότι:

1. έχουμε σημαντική αύξηση της μέσης τιμής των πωλήσεων από 90.000 ευρώ πέρυσι σε 130.000 ευρώ φέτος $(130 - 90 = 40)$, ποσοστό αύξησης περίπου 44,44%.
2. Η διασπορά των πωλήσεων φέτος παρουσιάζει καλύτερη εικόνα, δηλαδή έχουμε καλύτερη ομοιογένεια ως προς τα έσοδα, αφού ο φετινός συντελεστής μεταβολής $CV_2 = 9{,}2\%$ είναι μικρότερος από αυτόν της προηγούμενης χρήσης $CV_1 = 11{,}1\%$. Μάλιστα τη φετινή χρήση, οι πωλήσεις των υποκαταστημάτων είναι ομοιογενείς, σε αντίθεση με αυτές της προηγούμενης χρήσης που δεν παρουσίαζαν ομοιογένεια.

Ακόμη, έχουμε φέτος ένα εύρος μεταβολής ίσο με : 156 – 105 = 51 χιλιάδες, το οποίο είναι αρκετά μικρότερο από το περσινό, που ήταν 140 – 70 = 70 χιλιάδες. Φυσικά η μελέτη δεν σταματά εδώ, θα υπάρξει συνέχεια με την επεξεργασία και άλλων στοιχείων και με τη χρήση και άλλων στατιστικών μέτρων, όμως και τα μέχρι τώρα αποτελέσματα που βρέθηκαν είναι σημαντικά και ενδεικτικά του αποτελέσματος που είχε η αναδιοργάνωση στην λειτουργία της επιχείρησης.

3.3 Έλεγχος Ποιότητας

Σε όλες τις επιχειρήσεις, πραγματοποιείται μετασχηματισμός των πόρων που εισέρχονται σε αυτές, όπως για παράδειγμα οι πρώτες ύλες και το ανθρώπινο δυναμικό, σε στοιχεία τα οποία παράγονται από αυτές, όπως τα προϊόντα και οι υπηρεσίες. Τα προϊόντα και οι υπηρεσίες, παράγονται για την ικανοποίηση των ανθρώπινων αναγκών. Για το λόγο αυτό σχεδιάζονται λαμβάνοντας υπόψη τόσο τη φύση των αναγκών που προορίζονται να ικανοποιήσουν όσο και τις τεχνολογικές δυνατότητες που μπορούν να χρησιμοποιηθούν για την παραγωγή τους. Το αποτέλεσμα του σχεδιασμού αυτού θα πρέπει να ικανοποιεί τις γενικές καθώς και τις τεχνικές προδιαγραφές.

Με τις γενικές προδιαγραφές εξασφαλίζεται η λειτουργικότητα, η αξιοπιστία, το χαμηλό κόστος και διάφορα αισθητικά χαρακτηριστικά που θα πρέπει να χαρακτηρίζουν το προϊόν ή την υπηρεσία. Με άλλα λόγια, οι γενικές προδιαγραφές ορίζουν τα γενικά χαρακτηριστικά των προϊόντων καθώς και των υπηρεσιών. Με τις τεχνικές προδιαγραφές καθορίζονται τα λεπτομερή χαρακτηριστικά του προϊόντος όπως οι διαστάσεις των επιμέρους εξαρτημάτων, η σύσταση του υλικού τους κ.α. ή της υπηρεσίας, όπως η χρονική της διάρκεια, οι διάφορες παροχές κ.α.. Επίσης, οι τεχνικές προδιαγραφές καθορίζονται έτσι ώστε να εξασφαλίζουν τις γενικές προδιαγραφές του προϊόντος ή της υπηρεσίας με τον πιο οικονομικό τρόπο.

Για την παραγωγή των προϊόντων καθώς και την παροχή των υπηρεσιών, έτσι ώστε να ανταποκρίνονται στις προδιαγραφές που έχουν οριστεί, θα πρέπει σε κάθε φάση της παραγωγικής διαδικασίας να γίνεται προσπάθεια για σύγκλιση, αν όχι πλήρη ταύτιση, των χαρακτηριστικών των παραγόμενων προϊόντων και των προσφερόμενων υπηρεσιών με τις προδιαγραφές αυτές. Ο βαθμός σύγκλισης σε κάθε φάση προσδιορίζει την ποιότητα των προϊόντων και των υπηρεσιών. Με τον τρόπο αυτό μπορούμε να μιλάμε για ποιότητα πρώτων υλών, για ποιότητα στις φάσεις παραγωγής, για ποιότητα της υπηρεσίας σε σχέση με τις ανάγκες του καταναλωτή κ.τ.λ.

Η σύγχρονη τάση στη Διοίκηση Επιχειρήσεων, όπως αυτή εκφράζεται μέσω της φιλοσοφίας της Διοίκησης Ολικής Ποιότητας (Total Quality Management) επικεντρώνεται στη συνεχή βελτίωση:

1. των παραγόμενων προϊόντων ή υπηρεσιών και

2. της διαδικασίας παραγωγής τους

Αυτή η προσέγγιση στη διοίκηση απαιτεί τη γνώση των βασικών αρχών και τεχνικών της Στατιστικής. Στόχος είναι η κατανόηση, ο έλεγχος και η βελτίωση (με μείωση της υπάρχουσας μεταβλητότητας) των προϊόντων-υπηρεσιών και των διαδικασιών παραγωγής τους.

3.3.1 Μεταβλητότητα σε μια διαδικασία

Οποιαδήποτε επαναλαμβανόμενη διαδικασία υπόκειται σε μεταβολές. Πιο συγκεκριμένα, υπάρχουν μεταβολές:

- στα τεχνικά και άλλα μέσα που χρησιμοποιεί η διαδικασία
- στον τρόπο που ο χειριστής της επεμβαίνει σ' αυτήν,
- στο υλικό ή/και στην πληροφόρηση που υφίσταται επεξεργασία μέσω της συγκεκριμένης διαδικασίας.

Για παράδειγμα σε μια βιομηχανική διαδικασία υπάρχουν μεταβολές:

1. από πτώση στην απόδοση των μηχανών ή παρεμφερείς κατασκευαστικές αδυναμίες, όπως η διαφορετική απόδοση κάθε μηχανής
2. από ελάττωση ή αύξηση της προσοχής ή της πείρας των εργαζομένων
3. από αλλαγή των διαστάσεων των προϊόντων ή της χημικής σύστασης των υλικών, κ.α.
4. από συνδυασμούς των 1, 2, 3.

Με τις γενικές και τεχνικές προδιαγραφές καθορίζεται το εύρος των μεταβολών αυτών που είναι αποδεκτό σε κάθε σημείο της παραγωγικής διαδικασίας. Δηλαδή οι προδιαγραφές αυτές καθορίζουν τα όρια των ανοχών σε μια παραγωγική διαδικασία, τα οποία δεν μπορεί να είναι μικρότερα από τα φυσικά όρια των δυνατοτήτων της διαδικασίας, γιατί στην περίπτωση αυτή θα αυξηθεί η ποσότητα των ελαττωματικών προϊόντων που παράγονται.

Οι αιτίες της παρατηρούμενης μεταβλητότητας μπορούν να ομαδοποιηθούν στις ακόλουθες δύο κατηγορίες:

- **Κοινές αιτίες μεταβλητότητας**: Είναι συνήθως μικρές, τυχαίες μεταβολές, οι οποίες προκύπτουν από την ίδια τη λειτουργία του συγκεκριμένου συστήματος. Οι μικρές, τυχαίες αυτές αιτίες επιδρούν μόνιμα σε μια διαδικασία (ή σύστημα) προκαλώντας μία μεταβολή που συχνά ακολουθεί κάποια γνωστή στατιστική κατανομή (συνήθως την Κανονική). Εάν μόνο κοινές αιτίες προκαλούν τη μεταβλητότητα σε μια διαδικασία, τότε λέμε ότι η διαδικασία (ή το σύστημα) είναι σε κατάσταση στατιστικού ελέγχου. Για παράδειγμα, το βάρος της συσκευασίας ενός προϊόντος, ή ο χρόνος κατεργασίας μιας πρώτης ύλης, παρουσιάζουν μικρές αποκλίσεις σε κάθε μονάδα. Οι αποκλίσεις αυτές οφείλονται στις μικροδιαφορές που παρουσιάζουν οι συντελεστές της παραγωγικής διαδικασίας, όπως οι μηχανές συσκευασίας, οι εργαζόμενοι, οι πρώτες ύλες κ.τ.λ.

- **Ειδικές αιτίες μεταβλητότητας**: Είναι οι μεγάλες σχετικά μεταβολές που δεν οφείλονται στο συνήθη τρόπο λειτουργίας της διαδικασίας και που έχουν σημαντική επίδραση στην όλη διαδικασία (π.χ. παραγωγή ελαττωματικού προϊόντος). Σε περιπτώσεις εντοπισμού 'Ειδικών Αιτιών Μεταβλητότητας' λέμε ότι το σύστημα (ή διαδικασία) είναι εκτός ελέγχου και προσπαθούμε να εντοπίσουμε τη φύση των αιτίων αυτών και να επαναφέρουμε το σύστημα σε κατάσταση στατιστικού ελέγχου.

Έτσι, αν το μέγεθος που μας ενδιαφέρει (π.χ. το βάρος κάθε συσκευασίας ενός προϊόντος) κατά την παραγωγική διαδικασία, ακολουθεί κανονική κατανομή με μέση τιμή μ και τυπική απόκλιση σ μπορούν να παρουσιαστούν οι ακόλουθες δύο περιπτώσεις:

1. Η πρώτη είναι να επιδρούν μόνο κοινές αιτίες μεταβλητότητας οπότε τόσο η κατανομή όσο και η μέση τιμή με την τυπική απόκλιση παραμένουν σχεδόν αμετάβλητα με την πάροδο του χρόνου.
2. Αν όμως υπάρχει επίδραση ειδικών αιτιών μεταβλητότητας στην παραγωγική διαδικασία, τότε μπορεί να μεταβληθεί ή η μέση τιμή ή η τυπική απόκλισή της κατανομής ή και τα δύο μεγέθη μαζί. Είναι φανερό ότι στην περίπτωση αυτή η διαδικασία δεν θα παράγει προϊόντα με τα χαρακτηριστικά μ και σ που ορίζουν οι προδιαγραφές.

Από τα παραπάνω προκύπτει η ανάγκη για διαρκή έλεγχο της διαδικασίας έτσι ώστε να εντοπίζονται άμεσα και να διορθώνονται πιθανές παρεκκλίσεις. Ο καλύτερος τρόπος ελέγχου της διαδικασίας είναι ο τακτικός δειγματοληπτικός έλεγχος της παραγωγής. Με άλλα λόγια, για να εντοπιστούν οι πιθανές παρεκκλίσεις και να ελέγχεται η σταθερότητα της παραγωγικής διαδικασίας, λαμβάνονται τακτικά δείγματα και γίνονται συνεχείς στατιστικοί έλεγχοι ως προς τα διάφορα χαρακτηριστικά του πληθυσμού. Αυτό στην πράξη γίνεται με χρήση των διαγραμμάτων ελέγχου, που εξετάζονται παρακάτω.

3.3.2. Στατιστικά διαγράμματα (χάρτες) ελέγχου ποιότητας

Από τις προηγούμενες παραγράφους προκύπτει ότι οι γενικές και οι τεχνικές προδιαγραφές ενός προϊόντος ή μιας υπηρεσίας καθορίζουν την αποδεκτή κατανομή των μεγεθών που χαρακτηρίζουν το προϊόν ή την προσφερόμενη υπηρεσία. Για να γίνει επομένως ο έλεγχος της παραγωγικής διαδικασίας θα πρέπει να είναι γνωστό τόσο το είδος της κατανομής, όσο και οι τιμές των παραμέτρων της. Η γνώση αυτών των στατιστικών στοιχείων είναι απαραίτητη για το σχεδιασμό των κατάλληλων ελέγχων.

Στην πράξη όμως σχεδόν ποτέ δεν είναι γνωστό το είδος της κατανομής, που οι αρχικές φυσικές δυνατότητες της παραγωγικής διαδικασίας διαμορφώνουν, αλλά ούτε και οι παράμετροί τους. Για το λόγο αυτό υπάρχουν κάθε φορά δύο δυνατές επιλογές:

- Να αναζητηθεί η πραγματική κατανομή της παραγωγικής διαδικασίας, πράγμα όμως που είναι δύσκολο στην πράξη και έχει μεγάλο οικονομικό κόστος.
- Να χρησιμοποιηθούν διάφορες προσεγγίσεις που υποδεικνύει η Στατιστική.

Στον ποιοτικό έλεγχο ακολουθείται κατά κανόνα τη δεύτερη προσέγγιση.

3.3.3 Διάγραμμα ελέγχου μέσης τιμής

Ένας από τους σημαντικότερους ελέγχους που γίνονται κατά την παραγωγική διαδικασία αφορά τη μέση τιμή μ ενός χαρακτηριστικού του παραγόμενου προϊόντος ή της προσφερόμενης υπηρεσίας. Για παράδειγμα, ο μέσος χρόνος εξυπηρέτησης των πελατών σε μια τράπεζα ή το μέσο βάρος συσκευασίας ενός προϊόντος. Για την πραγματοποίηση του ελέγχου χρησιμοποιείται η παραδοχή ότι οι μέσες τιμές n σε πλήθος δειγμάτων ακολουθούν την κανονική κατανομή ανεξάρτητα από το είδος της κατανομής που ακολουθεί ο πληθυσμός από τον οποίο προέρχονται τα δείγματα. Επίσης, θεωρείται ότι η μέση τιμή της κατανομής των μέσων τιμών των δειγμάτων ισούται με τη μέση τιμή του πληθυσμού, δηλαδή

$$\mu_{\bar{x}} = \mu$$

και ότι η τυπική απόκλιση είναι ίση με

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

όπου:

- n το μέγεθος του δείγματος και
- σ η τυπική απόκλιση του πληθυσμού από τον οποίο προέρχονται τα δείγματα.

Εφόσον η δειγματική μέση τιμή $\bar{x}$ ακολουθεί κανονική κατανομή με μέση τιμή $\mu_{\bar{X}}$ και τυπική απόκλιση

$$\frac{\sigma}{\sqrt{n}}$$

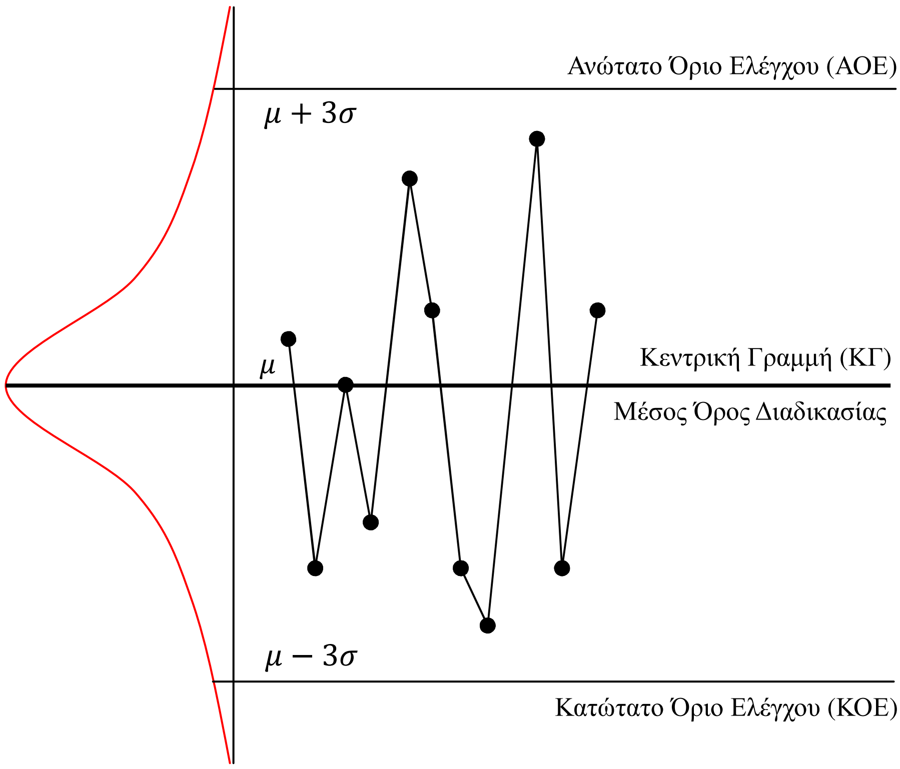
αν η παραγωγική διαδικασία παραμένει αμετάβλητη με την πάροδο του χρόνου, οι τιμές των μέσων τιμών των δειγμάτων πρέπει να βρίσκονται στην περιοχή:

$$\boldsymbol{\mu_{\bar{x}} - z_{1-\frac{1}{2}a} \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu_{\bar{x}} + z_{1-\frac{1}{2}a} \frac{\sigma}{\sqrt{n}}}$$

με πιθανότητα $1 - \alpha$. Αυτό σημαίνει ότι μόνο το $\alpha\%$ των μέσων τιμών μπορεί να βρίσκεται έξω από την παραπάνω περιοχή. Στον ποιοτικό έλεγχο έχει επικρατήσει να χρησιμοποιείτε ως $\alpha = 0{,}002$ ή $1 - \alpha = 0{,}998$, οπότε από τον πίνακα της κανονικής κατανομής για $\mu = 0$ και $\sigma = 1$ προκύπτει ότι

$$z_{1-\frac{\alpha}{2}} = 3$$

Με βάση τα παραπάνω μπορεί να σχεδιαστεί ένα διάγραμμα σαν αυτό που φαίνεται στο ακόλουθο σχήμα.

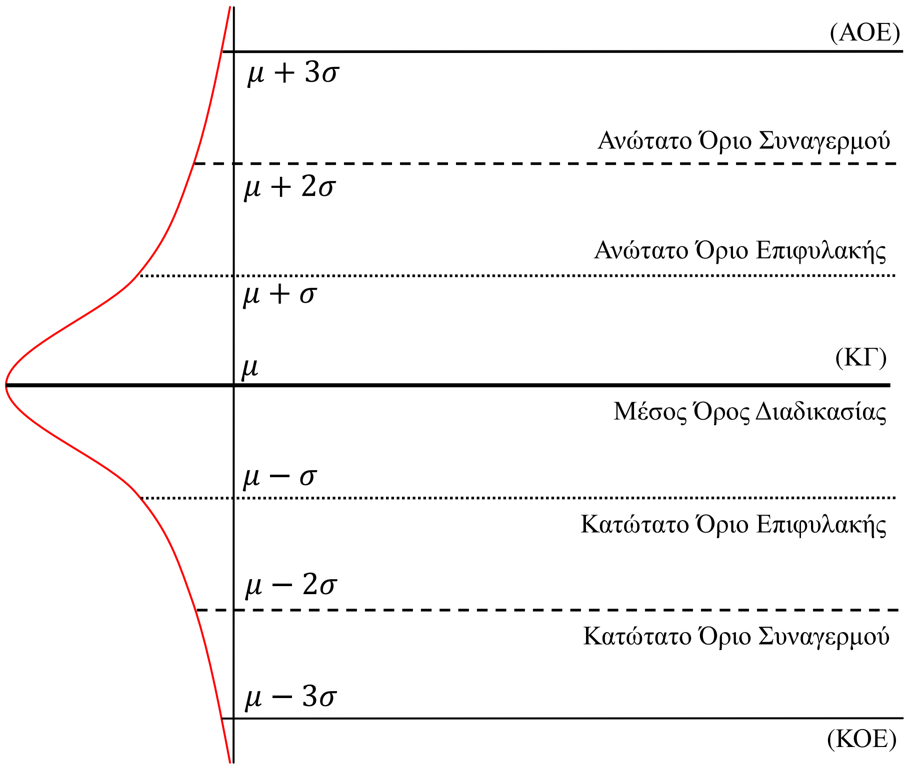


Εικόνα 8: Διάγραμμα ελέγχου μέσης τιμής

Το Ανώτατο όριο ελέγχου (ΑΟΕ) έχει τεθεί ίσο με $\mu + 3\sigma$, το Κατώτατο όριο ελέγχου (ΚΟΕ) είναι ίσο με $\mu - 3\sigma$ και ο Μέσος όρος της διαδικασίας ή η κεντρική γραμμή (ΚΓ) ίσος με μ. Η κεντρική γραμμή του διαγράμματος ελέγχου αντιστοιχεί στο μέσο όρο μ της κανονικής κατανομής και θεωρείται η ιδανική τιμή του χαρακτηριστικού μεγέθους που ελέγχεται κατά διάφορες χρονικές στιγμές.

Το Ανώτατο και το Κατώτατο Όριο Ελέγχου απέχουν 3σ από την ιδανική τιμή και ορίζουν το διάστημα των τιμών του μεγέθους που είναι αποδεκτές για την παραγωγική διαδικασία. Μια ή περισσότερες μετρήσεις εκτός των ορίων, είναι ενδείξεις ότι η διαδικασία είναι εκτός ελέγχου.

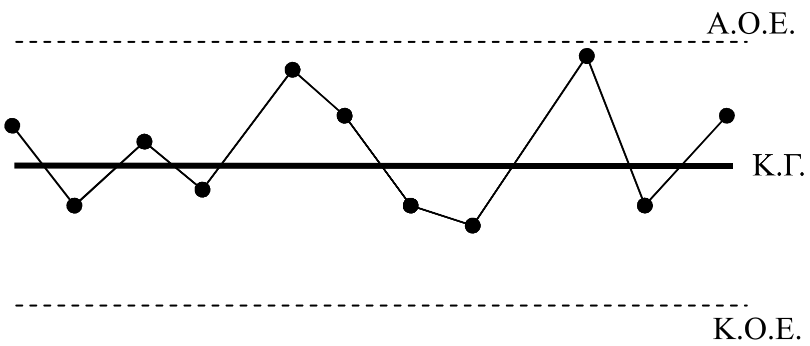
Όταν το κόστος από το να βρεθεί η διαδικασία εκτός ελέγχου είναι πολύ μεγάλο σε σχέση με το κόστος παρακολούθησης των μεταβολών της διαδικασίας, μπορούν να χρησιμοποιηθούν ακόμα τα «όρια επιφυλακής», όταν παρατηρούνται συστηματικά τιμές κοντά στα άκρα του διαστήματος $\mu \pm \sigma$ και τα «όρια συναγερμού» όταν αυτό συμβαίνει κοντά στα άκρα του διαστήματος $\mu \pm 2\sigma$. Τα όρια αυτά παρουσιάζονται στο σχήμα 9.



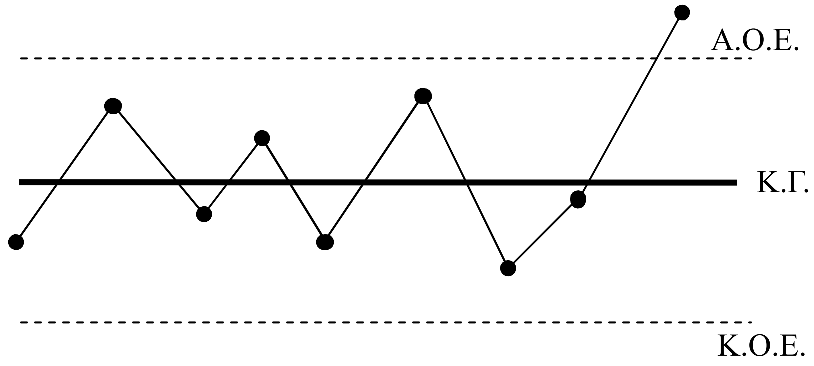
Εικόνα 9: Διάγραμμα ελέγχου ορίων επιφυλακής και συναγερμού

Μια τέτοια προσέγγιση επιτρέπει τη παρακολούθηση της παραγωγικής διαδικασίας καθώς και την έγκαιρη επέμβαση, όταν υπάρχει κίνδυνος να βρεθεί εκτός ελέγχου.

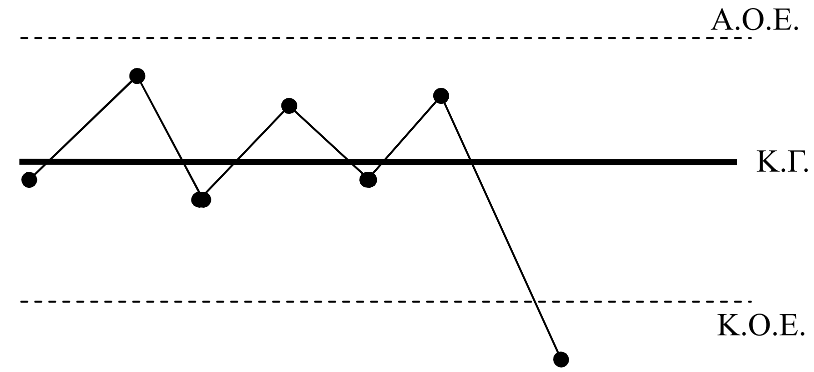
Στα επόμενα σχήματα εξετάζονται, μέσω συγκεκριμένων στατιστικών διαγραμμάτων ελέγχου, πιθανές καταστάσεις της παραγωγικής διαδικασίας όπου είναι σαφής η ανάγκη διερεύνησης αιτιών για επιφυλακή ή συναγερμό:



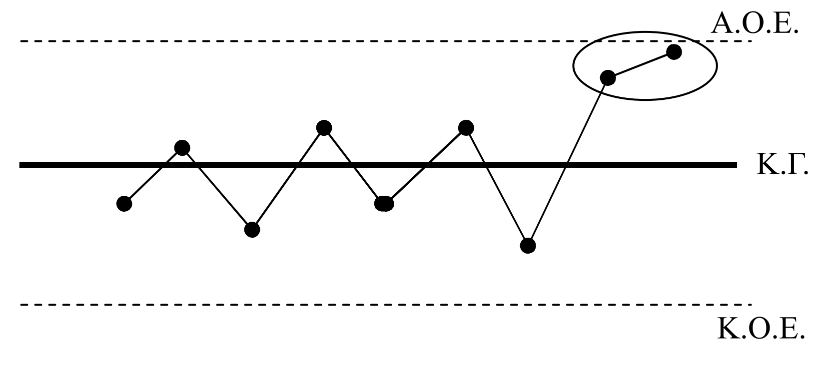
Εικόνα 10: Κανονική Παραγωγή



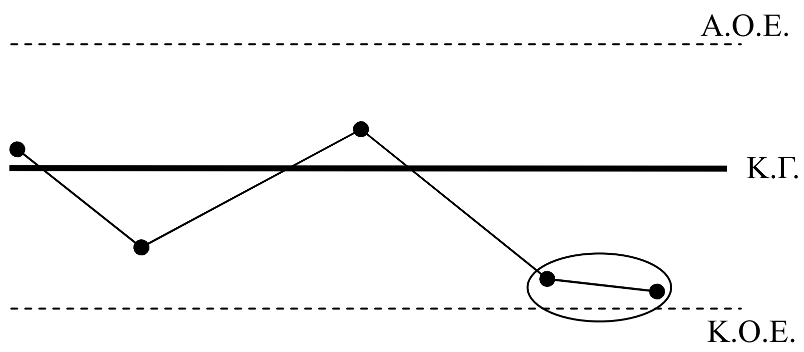
**Εικόνα 11: Ένα σημείο βρίσκεται πάνω από το ΑΟΕ.
Διερεύνηση αιτίας δυσλειτουργίας**



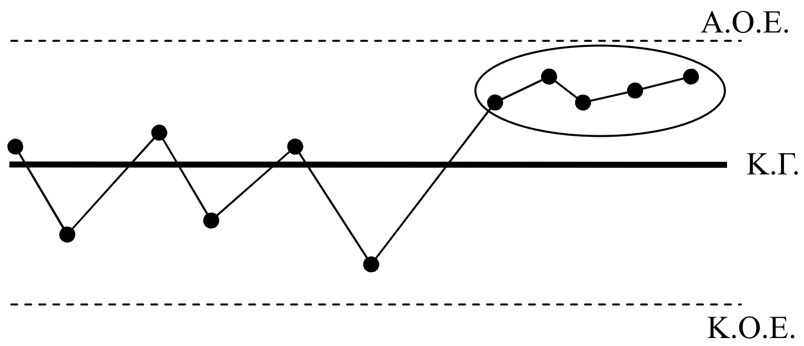
**Εικόνα 12: Ένα σημείο βρίσκεται κάτω από το ΚΟΕ.
Διερεύνηση αιτίας δυσλειτουργίας**



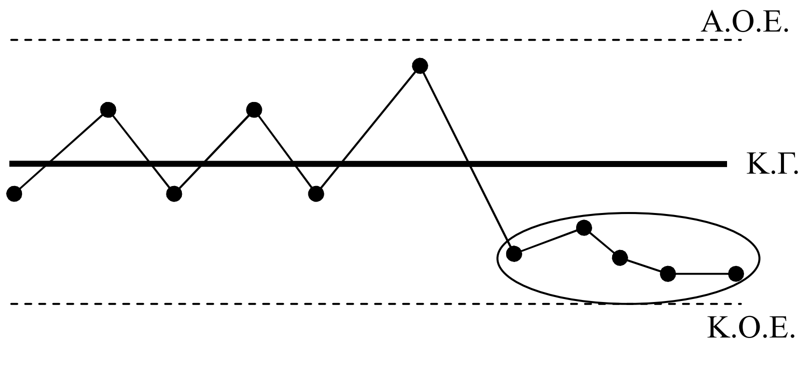
**Εικόνα 13: Δύο διαδοχικά σημεία βρίσκονται κοντά στο ΑΟΕ.
Διερεύνηση αιτίας δυσλειτουργίας**



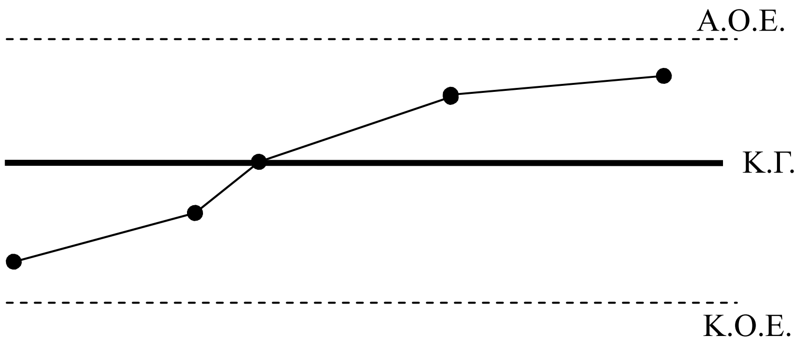
Εικόνα 14: Δύο διαδοχικά σημεία βρίσκονται κοντά στο Κ.Ο.Ε. Διερεύνηση της αιτίας δυσλειτουργίας



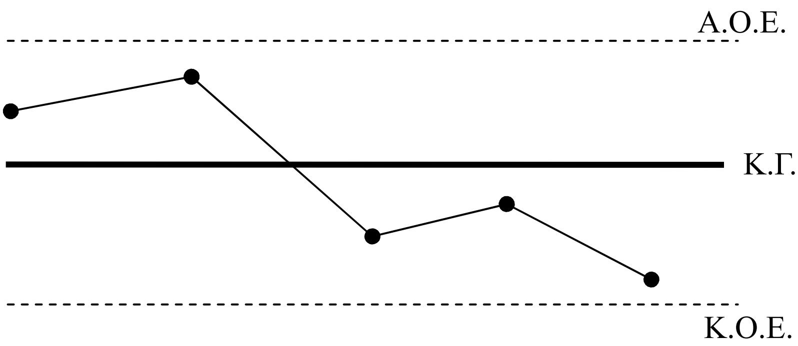
Εικόνα 15: Πέντε διαδοχικά σημεία βρίσκονται κοντά στο ΑΟΕ. Διερεύνηση αιτίας συνεχόμενης δυσλειτουργίας



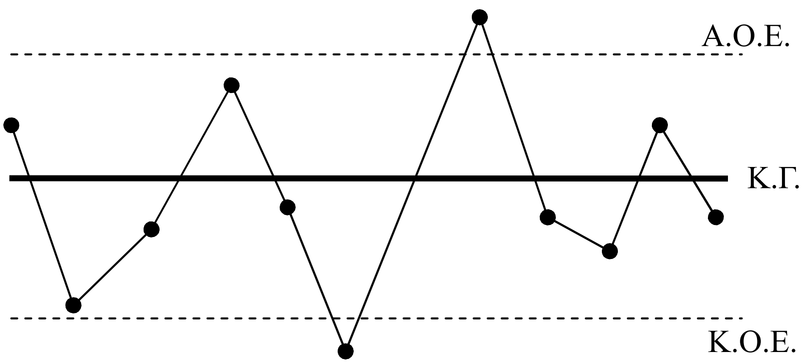
Εικόνα 16: Πέντε διαδοχικά σημεία βρίσκονται κοντά στο ΚΟΕ. Διερεύνηση αιτίας συνεχόμενης δυσλειτουργίας



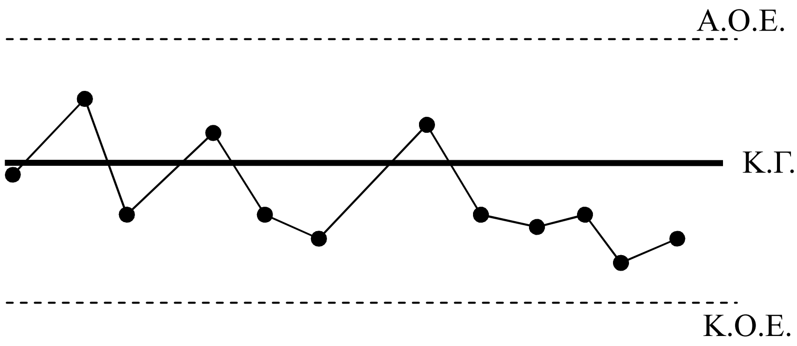
Εικόνα 17: Πέντε διαδοχικά σημεία παρουσιάζουν τάση προοδευτικής απόκλισης από ΚΓ προς ΑΟΕ



Εικόνα 18: Πέντε διαδοχικά σημεία παρουσιάζουν τάση προοδευτικής απόκλισης από ΚΓ προς ΚΟΕ



Εικόνα 19: Παραγωγή εκτός ελέγχου. Διερεύνηση αιτίας δυσλειτουργίας



Εικόνα 20: Συνεχής χειροτέρευση αποκλίσεων.
Διερεύνηση αιτίας δυσλειτουργίας

Στην πράξη, για να ελεγχθεί η κατάσταση της παραγωγικής διαδικασίας λαμβάνονται δείγματα σε διάφορες χρονικές στιγμές. Για να σχεδιαστεί το διάγραμμα ελέγχου των μέσων όρων λαμβάνονται οι μέσες τιμές $\bar{x}_i$ των δειγμάτων. Η κεντρική γραμμή, όταν υπάρχουν Ν πλήθους δείγματα, μεγέθους n_i, αντιστοιχεί στη μέση τιμή των μέσων τιμών των δειγμάτων και υπολογίζεται από τη σχέση:

$$\mu_{\bar{x}} = \frac{1}{N}\sum_{i=1}^{N} \bar{x}_i$$

Ακόμα γίνεται εκτίμηση της τυπικής απόκλισης σ του πληθυσμού. Αυτό γίνεται με την ακόλουθη διαδικασία:

Αρχικά υπολογίζεται η τυπική απόκλιση s_i για κάθε ένα από τα n δείγματα, από τη σχέση:

$$s_i^2 = \frac{1}{n_i - 1}\sum_{i=1}^{N} (x_i - \bar{x})^2$$

και στη συνέχεια υπολογίζεται η s^2, που είναι η εκτιμήτρια της σ^2 του πληθυσμού από τη σχέση:

$$s^2 = \frac{\sum_{i=1}^{N}(n_i - 1)s_i^2}{\sum_{i=1}^{N}(n_i - 1)}$$

Ακολουθεί η σχεδίαση του διαγράμματος σύμφωνα με την παραπάνω διαδικασία, και η τοποθέτηση των μέσων τιμών των δειγμάτων ανά χρονική στιγμή πάνω σε αυτό. Αν το γράφημα που δημιουργούν οι μέσες τιμές των διαδοχικών δειγμάτων είναι αυτή που φαίνεται στην εικόνα 8, τότε φαίνεται η παραγωγική διαδικασία να παραμένει στατιστικά αμετάβλητη. Αντίθετα, αν το γράφημα έχει κάποια από τις μορφές που φαίνονται στις εικόνες 11-20, συμπεραίνουμε ότι η διαδικασία βρίσκεται ή τείνει να βρεθεί εκτός ελέγχου και πρέπει να προσπαθήσουμε να άρουμε τις ειδικές αιτίες μεταβλητότητας που την οδήγησαν σ' αυτή την κατάσταση.

Όπως ειπώθηκε παραπάνω η τυπική απόκλιση του πληθυσμού τις περισσότερες φορές δεν είναι γνωστή και χρειάζεται να υπολογιστεί η εκτιμήτριά της s. Πολλές φορές όμως και ο υπολογισμός της εκτιμήτριας s είναι αρκετά δύσκολος σε μια διαδικασία ποιοτικού ελέγχου. Στην περίπτωση αυτή για τον υπολογισμό της τυπικής απόκλισης χρησιμοποιείται το εύρος R του δείγματος. Αυτό γίνεται με χρήση της σχέσης $A_2\bar{R} = 3\sigma_{\bar{x}}$, όπου A_2 είναι ένας εμπειρικός συντελεστής ο οποίος είναι ανεξάρτητος από την εκάστοτε παραγωγική διαδικασία και μεταβάλλεται μόνο από το μέγεθος του δείγματος. Τα όρια ελέγχου σ' αυτή την περίπτωση είναι:

$$AOE = \mu_{\bar{x}} + A_2\bar{R}$$

και

$$KOE = \mu_{\bar{x}} - A_2\bar{R}$$

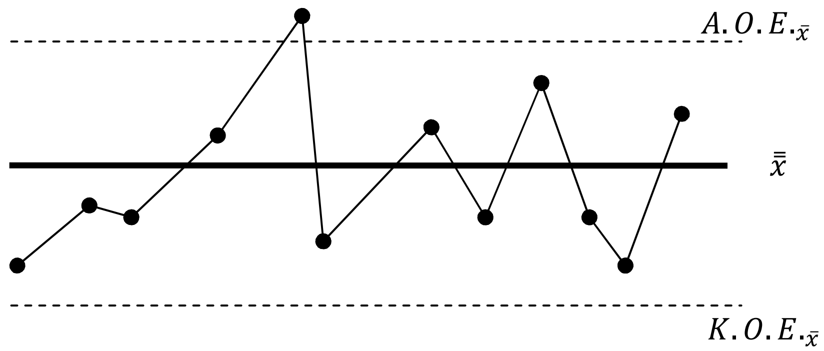
όπου:

- $\mu_{\bar{x}}$ είναι η μέση τιμή των μέσων τιμών των N δειγμάτων μεγέθους n,
- $\bar{R}$ η μέση τιμή των τιμών R του εύρους των N δειγμάτων και
- A_2 ο εμπειρικός συντελεστής του οποίου οι τιμές φαίνονται στον ακόλουθο πίνακα.

Πίνακας 8: Εμπειρικός συντελεστής A_2

Μέγεθος δείγματος	A_2
2	1.880
3	1.023
4	0.729
5	0.577
6	0.483
7	0.419
8	0.373
9	0.337
10	0.308
12	0.266
14	0.235
16	0.212
18	0.194
20	0.180
22	0.167
24	0.158

Για τον παραπάνω έλεγχο μπορεί να σχεδιαστεί προφανώς το διάγραμμα ελέγχου της μέσης τιμής.



Εικόνα 21: Στατιστικό διάγραμμα ελέγχου μέσου όρου δείγματος

Παράδειγμα 12

Ένα βιομηχανικό ψαλίδι κόβει σύρμα για τη δημιουργία βιδών μήκους 2,515 cm προοριζόμενες για κάποια ειδική χρήση. Έστω ότι λαμβάνουμε, με τυχαία δειγματοληψία 5 δείγματα από 4 βίδες κάθε φορά, όπως φαίνεται στις στήλες (1) και (2) του παρακάτω πίνακα.

Να υπολογιστούν τα ανώτατα όρια ελέγχου των μέσων όρων $\bar{x}$ των δειγμάτων και να σχεδιαστεί ο σχετικός στατιστικός χάρτης ελέγχου.

Απάντηση:

Οι τιμές που απαιτούνται για τον υπολογισμό των Ορίων Ελέγχου φαίνονται στις στήλες (3), (4) και (5) του παρακάτω πίνακα:

Πίνακας 9:Υπολογισμός ορίων ελέγχου

(1)	(2)				(3)	(4)	(5)
Αύξοντας αριθμός δείγματος	Ταξινόμηση των μετρήσεων στις 4 βίδες κάθε δείγματος				$\sum x$	$\bar{x}$	s^2
	1η βίδα	2η βίδα	3η βίδα	4η βίδα			
1	2,5180	2,5210	2,5240	2,5090	10,072	2,5180	0,0000315
2	2,5075	2,5135	2,5120	2,5150	10,048	2,5120	0,0000078
3	2,5045	2,5195	2,5075	2,5105	10,042	2,5105	0,0000315
4	2,5300	2,5270	2,5270	2,5240	10,108	2,5270	0,0000045
5	2,5060	2,5075	2,5015	2,5150	10,030	2,5075	0,0000236
Σύνολο					50,300	12,5750	0,0000990

Ο μέσος όρος του κάθε δείγματος συμβολίζεται με $\bar{x}_i$ για $i = 1, 2, 3, 4, 5$. Ο υπολογισμός για κάθε $\bar{x}_i$ φαίνεται στις στήλες (3) και (4). Αναλυτικά έχουμε:

$$\bar{x}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{10{,}072}{4} = 2{,}518$$

$$\bar{x}_2 = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{10{,}048}{4} = 2{,}512$$

$$\bar{x}_3 = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{10{,}042}{4} = 2{,}5105$$

$$\bar{x}_4 = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{10{,}108}{4} = 2{,}527$$

$$\bar{x}_5 = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{10{,}030}{4} = 2{,}5075$$

Η μέση τιμή των μέσων όρων των δειγμάτων, συμβολίζεται με $\bar{\bar{x}}$. Το άθροισμα των μέσων όρων των 5 δειγμάτων συμβολίζεται με ($\Sigma\bar{x}$) και το υπολογίζουμε από τη στήλη (4).

$$\bar{\bar{x}} = \frac{1}{N}\sum_{i=1}^{N}\bar{x} = \frac{12{,}575}{5} = 2{,}515$$

Η μέση τιμή $\bar{\bar{x}}$ μπορεί επίσης να υπολογιστεί και από το σύνολο όλων των μετρήσεων:

$$\bar{\bar{x}} = \frac{1}{n \cdot N}\sum_{i=1}^{n \cdot N} x_i = \frac{50{,}300}{20} = 2{,}515$$

Για τον υπολογισμό του Ανώτατου και του Κατώτατου ορίου ελέγχου πρέπει πρώτα να υπολογιστεί η διακύμανση s_i^2, όπου $i = 1, 2, 3, 4, 5$, για κάθε δείγμα.

Η τιμή του s_1^2 για το πρώτο δείγμα υπολογίζεται όπως φαίνεται στον ακόλουθο πίνακα:

Πίνακας 10:Υπολογισμός διακύμανσης 1ου δείγματος

$\boldsymbol{x}$	$\boldsymbol{x - \bar{x}}$ $\boldsymbol{\bar{x} = 2,518}$	$\boldsymbol{(x - \bar{x})^2}$
2,518	0	0
2,521	0,003	0,000009
2,524	0,006	0,000036
2,509	-0,009	0,000081
10,072		**0,000126**

και

$$s_1^2 = \frac{0{,}000126}{4} = 0{,}0000315$$

Η τιμή του s_1^2 για το δεύτερο δείγμα υπολογίζεται όπως φαίνεται στον ακόλουθο πίνακα:

Πίνακας 11:Υπολογισμός διακύμανσης 2ου δείγματος

x	$x-\bar{x}$ $\bar{x}=2,512$	$(x-\bar{x})^2$
2,5075	-0,0045	0,00002025
2,5135	0,0015	0,00000225
2,5120	0	0
2,515	0,003	0,000009
10,072		**0,0000315**

και

$$s_2^2 = \frac{0{,}0000315}{4} = 0{,}000007875$$

Η τιμή του s_3^2 για το τρίτο δείγμα υπολογίζεται όπως φαίνεται στον ακόλουθο πίνακα:

Πίνακας 12:Υπολογισμός διακύμανσης 3ου δείγματος

x	$x-\bar{x}$ $\bar{x}=2,5105$	$(x-\bar{x})^2$
2,5045	-0,006	0,000036
2,5195	0,009	0,000081
2,5075	-0,003	0,000009
2,5105	0	0
10,072		**0,000126**

και

$$s_3^2 = \frac{0{,}000126}{4} = 0{,}0000315$$

Η τιμή του s_4^2 για το τέταρτο δείγμα υπολογίζεται στον ακόλουθο πίνακα:

Πίνακας 13:Υπολογισμός διακύμανσης 4ου δείγματος

x	$x-\bar{x}$ $\bar{x}=2,527$	$(x-\bar{x})^2$
2,530	0,003	0,000009
2,527	0	0
2,527	0	0

2,524	-0,003	0,000009
10,072		**0,000018**

και

$$s_4^2 = \frac{0,000018}{4} = 0,0000045$$

Τέλος, η τιμή του s_5^2 για το πέμπτο δείγμα υπολογίζεται στον ακόλουθο πίνακα:

Πίνακας 14:Υπολογισμός διακύμανσης 5ου δείγματος

$\boldsymbol{x}$	$\boldsymbol{x-\bar{x}}$ $\boldsymbol{\bar{x}=2,5075}$	$\boldsymbol{(x-\bar{x})^2}$
2,506	-0,0015	0,00000225
2,5075	0	0
2,5015	-0,006	0,00003600
2,515	0,0075	0,00005625
10,072		**0,0000945**

και

$$s_5^2 = \frac{0,0000945}{4} = 0,000023625$$

Οπότε προκύπτει:

$$\sum_{i=1}^{5} s_i^2 =$$

$$= 0,0000315 + 0,000007875 + 0,0000315 + 0,0000045 + 0,000023625 =$$

$$= 0,000099$$

Ο μέσος όρος $\bar{s}^2$ (μέση διακύμανση) των πέντε δειγμάτων είναι:

$$\bar{s}^2 = \frac{0,000099}{5} = 0,0000198$$

και έτσι προκύπτει:

$$s = \sqrt{\bar{s}^2} = \sqrt{0{,}0000198}$$

Επομένως τα όρια ελέγχου είναι:

$$A.O.E._{\bar{x}} = \bar{\bar{x}} + 3\frac{s}{\sqrt{n-1}} = 2{,}515 + 3 \cdot \frac{\sqrt{0{,}0000198}}{\sqrt{4-1}} = 2{,}515 + 3 \cdot \sqrt{\frac{0{,}0000198}{3}} =$$

$$= 2{,}515 + 3 \cdot 0{,}00257 = 2{,}515 + 0{,}00771 = 2{,}52271$$

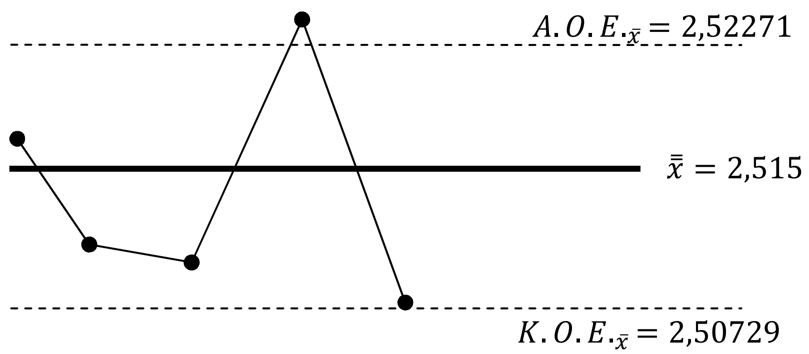
και

$$K.O.E._{\bar{x}} = \bar{\bar{x}} - 3\frac{s}{\sqrt{n-1}} = 2{,}515 - 3 \cdot \frac{\sqrt{0{,}0000198}}{\sqrt{4-1}} = 2{,}515 - 3 \cdot \sqrt{\frac{0{,}0000198}{3}} =$$

$$= 2{,}515 - 3 \cdot 0{,}00257 = 2{,}515 - 0{,}00771 = 2{,}50729$$

Δηλαδή $A.O.E._{\bar{x}} = 2{,}52271$ και $K.O.E._{\bar{x}} = 2{,}50729$

Ο χάρτης ελέγχου των $\bar{x}$ δίνεται από το ακόλουθο σχήμα:



Εικόνα 22: Στατιστικό διάγραμμα ελέγχου μέσου όρου δείγματος

Το στατιστικό διάγραμμα ελέγχου δείχνει ότι η διαδικασία είναι εκτός ελέγχου για το μέσο όρο του 4^{ου} δείγματος και πολύ κοντά στο ΚΟΕ για το μέσο όρο του 5^{ου} δείγματος.

Η «ειδική αιτία μεταβλητότητας» της διαδικασίας πρέπει να αναζητηθεί άμεσα και να γίνει η απαιτούμενη διόρθωση.

Παράδειγμα 13

Από μια διαδικασία που θεωρούμε ότι λειτουργούσε κανονικά μέχρι τώρα προέκυψε η παρακάτω κατανομή συχνότητας:

Πίνακας 15: Δεδομένα παραδείγματος

Μεταβλητή	5	4	3	2	1	0	-1	-2	-3	-4	-5
Συχνότητα	1	3	10	23	39	48	39	23	10	3	1

η οποία είναι κανονική με μέση τιμή 0 και τυπική απόκλιση 1,715.

Ζητείται να σχεδιαστεί ένα διάγραμμα ελέγχου μέσης τιμής του πληθυσμού χρησιμοποιώντας τη μέση τιμή δειγμάτων μεγέθους 5.

Για το διάγραμμα ελέγχου ισχύει:

- Κεντρική γραμμή:

$$\bar{\bar{x}} = \mu_{\bar{x}} = 0$$

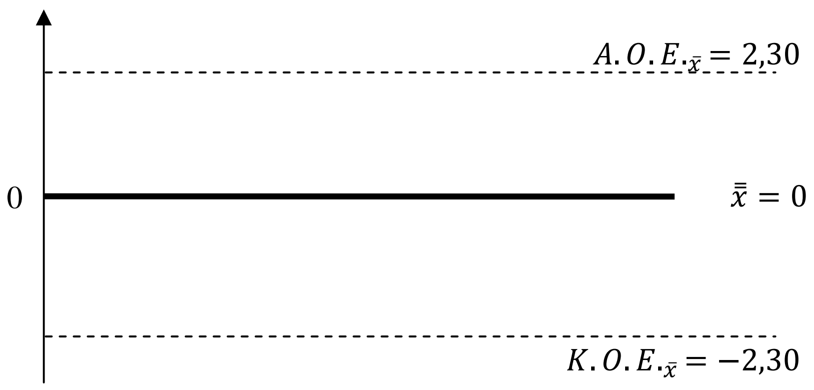
- ΑΟΕ:

$$\mu_{\bar{x}} + \frac{3\sigma}{\sqrt{n}} = 0 + \frac{3 \cdot 1{,}715}{\sqrt{5}} = 2{,}30$$

- ΚΟΕ:

$$\mu_{\bar{x}} - \frac{3\sigma}{\sqrt{n}} = 0 - \frac{3 \cdot 1{,}715}{\sqrt{5}} = -2{,}30$$

Το ζητούμενο διάγραμμα είναι το ακόλουθο:



Εικόνα 23: Διάγραμμα ελέγχου μέσης τιμής

3.3.4 Διάγραμμα ελέγχου εύρους

Όπως προαναφέρθηκε η διαδικασία μπορεί να μη μας δίνει τα χαρακτηριστικά που θέλουμε στα προϊόντα επειδή έχει μεταβληθεί η τυπική της απόκλιση. Αυτό μπορεί να γίνει ακόμη και αν η μέση τιμή δεν έχει μεταβληθεί. Συνεπώς, πρέπει να σχεδιάσουμε μια διαδικασία ελέγχου με την οποία να μπορούμε να ανιχνεύσουμε τις αδικαιολόγητες αποκλίσεις της διασποράς των τιμών, που δίνει η παραγωγική διαδικασία.

Όπως ξέρουμε, η πιο συνηθισμένη παράμετρος για τη μέτρηση της διασποράς είναι η τυπική απόκλιση. Για λόγους όμως που έχουν ήδη εξηγηθεί, καταφεύγουμε στο εύρος ως παράμετρο διασποράς. Οι υπολογισμοί γίνονται με χρήση των ακόλουθων σχέσεων:

$$D_4\bar{R} = \bar{R} + 3\sigma_R$$

και

$$D_3\bar{R} = \bar{R} - 3\sigma_R$$

Έτσι προκύπτει:

$$A.O.E._R = D_4\bar{R}$$

και

$$K.O.E._R = D_3\bar{R}$$

όπου:

- $\bar{R}$ είναι η μέση τιμή του εύρους των N δειγμάτων, δηλαδή

$$\bar{R} = \frac{1}{N}\sum_{i=1}^{N} R_i$$

- Οι συντελεστές D_3 και D_4 είναι ανάλογοι του A_2 και μεταβάλλονται με το μέγεθος του δείγματος. Οι τιμές των D_3 και D_4 δίνονται από πίνακες για διαφορετικά μεγέθη δειγμάτων και έτσι γίνονται άμεσα οι υπολογισμοί των ορίων.

Πίνακας 16: Συντελεστές για Υπολογισμό Ορίων Ελέγχου

Μέγεθος δείγματος	D_3	D_4
2	0	3.267
3	0	2.575
4	0	2.282
5	0	2.115
6	0	2.004
7	0.076	1.924
8	0.136	1.864
9	0.184	1.816
10	0.223	1.777
12	0.228	1.716
14	0.332	1.671
16	0.336	1.636
18	0.392	1.608
20	0.413	1.586
22	0.434	1.566
24	0.452	1.548

Παράδειγμα 14

Στον πίνακα 17 δίνονται τα εύρη 12 δειγμάτων μεγέθους 5 και ζητείται να υπολογίσουμε τα χαρακτηριστικά μεγέθη ενός διαγράμματος ελέγχου του εύρους.

Πίνακας 17: Δεδομένα

n	1	2	3	4	5	6	7	8	9	10	11	12
R	0,15	0,29	0,23	0,24	0,18	0,33	0,21	0,20	0,02	0,18	0,31	0,12

Αρχικά υπολογίζουμε τη μέση τιμή του εύρους των 12 δειγμάτων:

$$\bar{R} = \frac{1}{N}\sum_{i=1}^{N} R_i =$$

$$= \frac{0,15 + 0,29 + 0,23 + 0,24 + 0,18 + 0,33 + 0,21 + 0,20 + 0,02 + 0,18 + 0,31 + 0,12}{12}$$

$$=$$

$$= \frac{2,46}{12} = 0,205$$

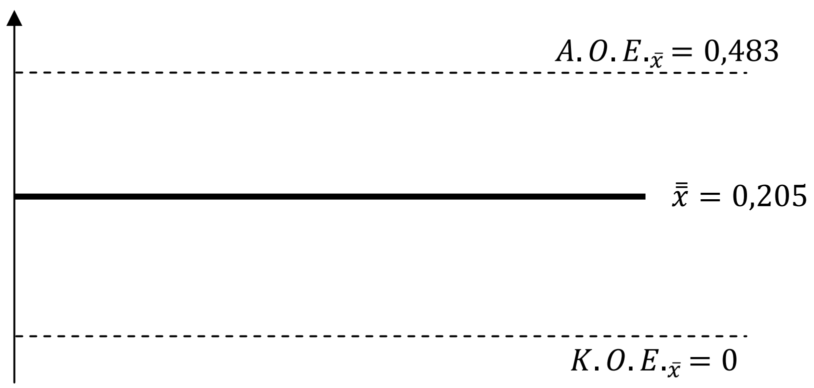
Η τιμή της $\bar{R}$ μας καθορίζει την κεντρική γραμμή του διαγράμματος. Για να βρούμε τα ΑΟΕ και ΚΟΕ που θα βάλουμε στο διάγραμμα, χρησιμοποιούμε τις τιμές από τον πίνακα 16. Έχουμε μέγεθος δείγματος $n = 5$, $D_3 = 0$ και $D_4 = 2{,}115$ οπότε παίρνουμε:

$$A.O.E._R = D_4\bar{R} = 2{,}115 \cdot 0{,}205 = 0{,}483$$

και

$$K.O.E._R = D_3\bar{R} = 0 \cdot 0{,}205 = 0$$

Τέλος το διάγραμμα ελέγχου παρουσιάζεται στην ακόλουθη εικόνα:



Εικόνα 24: Διάγραμμα ελέγχου εύρους

3.3.5 Διαγράμματα ελέγχου για ασυνεχείς μεταβλητές

Τα διαγράμματα ελέγχου για συνεχείς μεταβλητές, που χρησιμοποιούμε στον προληπτικό ποιοτικό έλεγχο, μπορούμε να τα χρησιμοποιήσουμε μόνο για εκείνα τα χαρακτηριστικά ενός προϊόντος που επιδέχονται μέτρηση. Συχνά όμως θέλουμε να ελέγξουμε χαρακτηριστικά κάποιου προϊόντος που η μέτρηση είναι αδύνατη ή δύσκολη. Σ' αυτές τις περιπτώσεις την ποιότητα του προϊόντος την καθορίζουμε με σύγκριση ως προς κάποιο πρότυπο. Το αποτέλεσμα της συγκρίσεως θα δίνουν το χαρακτηρισμό «καλό» ή «ελαττωματικό». Σ' αυτές τις περιπτώσεις χρησιμοποιούμε διαγράμματα ελέγχου για ασυνεχείς μεταβλητές.

Ο σκοπός του προληπτικού στατιστικού ποιοτικού ελέγχου είναι να προλάβουμε την παραγωγή μεγάλου αριθμού ελαττωματικών προϊόντων, δηλαδή την αύξηση της αναλογίας ελαττωματικών προϊόντων. Για να το κάνουμε αυτό ελέγχουμε την πιθανότητα να υπάρξει ένα ελαττωματικό είτε τον αριθμό των ελαττωματικών στο δείγμα.

3.3.6 Διάγραμμα ελέγχου ποσοστού ελαττωματικών ή διάγραμμα ελέγχου p.

Όταν η πιθανότητα να υπάρξει ένα ελαττωματικό είναι p ο αριθμός των ελαττωματικών σε ένα δείγμα n ακολουθεί τη διωνυμική κατανομή, η οποία έχει μέση τιμή

$$\mu_x = np$$

και μεταβλητότητα

$$\sigma_x^2 = np(1-p).$$

Εξάλλου ξέρουμε, ότι η εκτιμήτρια της p του πληθυσμού, που είναι η

$$\hat{p} = \frac{x}{n}$$

έχει μέση τιμή την

$$\mu_{\hat{p}} = p$$

και τυπική απόκλιση την

$$\sigma_{\hat{p}}^2 = \frac{p(1-p)}{n}$$

Εφόσον η κατανομή της x είναι διωνυμική και η κατανομή της $\hat{p}$ είναι επίσης διωνυμική αφού συνδέονται με τη σχέση

$$\hat{p} = \frac{x}{n}$$

Στην περίπτωση, όμως, που το δείγμα είναι μεγάλο, όταν δηλαδή $np \geq 10$ ή $n \geq 30$, τότε η κατανομή της $\hat{p}$ προσεγγίζει ικανοποιητικά την κανονική κατανομή.

Έτσι, αν εκλέξουμε το πλάτος των ορίων ελέγχου ίσο με έξι τυπικές αποκλίσεις θα έχουμε

$$AOE_p = \hat{p} + 3\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$KOE_p = max\begin{cases} \hat{p} - 3\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ 0 \end{cases}$$

Με τον τρόπο που σημειώνουμε το KOE_p, θέλουμε να εκφράσουμε ότι αυτό δε μπορεί να πάρει αρνητικές τιμές αφού είναι πιθανότητα και γι' αυτό ως κατώτερο όριο θέτουμε το μηδέν.

Στην πράξη, την τιμή της εκτιμήτριας $\hat{p}$ του p τη βρίσκουμε από k δείγματα μεγέθους n το καθένα. Έτσι, αν το δείγμα i περιέχει x_i ελαττωματικά για i=1,2,…k, την τιμή της $\hat{p}$ τη βρίσκουμε από τη σχέση

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{nk}$$

Παράδειγμα 15

Παίρνουμε 100 δείγματα με 25 μονάδες το καθένα από ένα ηλεκτρονικό στοιχείο. Στα δείγματα βρίσκουμε να περιέχονται τα εξής ελαττωματικά:

3, 2, 5, 1, 0, 2, 5, 6, 1, 9, 6, 2, 7, 9, 3, 2, 1, 9, 5, 3, 8, 9, 6, 5, 4.

Να σχεδιασθεί ένα διάγραμμα ελέγχου ποσοστού ελαττωματικών και να σημειωθούν σ' αυτό τα εξής αποτελέσματα δέκα επόμενων δειγμάτων:

5, 7, 4, 9, 11, 3, 7, 4, 6, 4.

Με χρήση των προηγούμενων τύπων βρίσκουμε ότι:

$$\hat{p} = \frac{113}{100 \cdot 25} = 0{,}0452$$

και

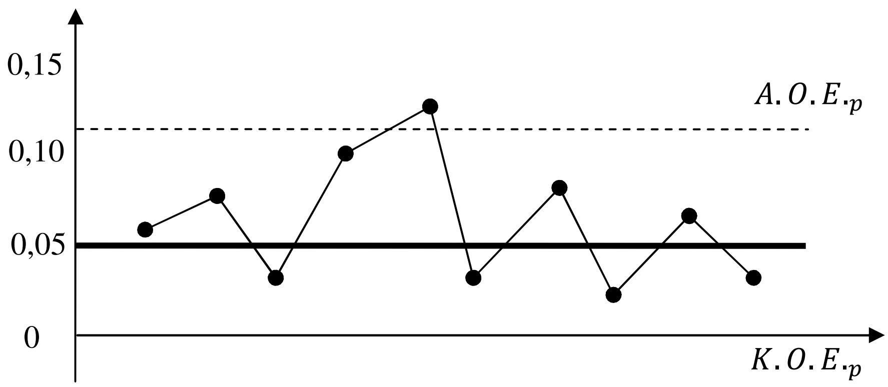
$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0{,}0452\ (1-0{,}0452)}{100}} = 0{,}0207$$

Οπότε,

$$AOE_p = 0{,}0452 + 3 \cdot 0{,}0207 = 0{,}1075$$

και

$$KOE_p = 0$$



Εικόνα 25: Διάγραμμα ελέγχου εύρους

Το σημείο που αντιστοιχεί στο δείγμα που έχει 11 ελαττωματικά, για το οποίο είναι $p = 0{,}11$ βρίσκεται πάνω από το AOE_p.

3.3.7 Διάγραμμα ελέγχου αριθμού ελαττωματικών (ή διάγραμμα np)

Συχνά χρησιμοποιούμε το διάγραμμα np, όπου n το μέγεθος του δείγματος και p η διωνυμική αναλογία, αντί του διαγράμματος p. Δηλαδή, ελέγχουμε τον αριθμό των ελαττωματικών σ' ένα δείγμα μεγέθους n και όχι την αναλογία των ελαττωματικών. Σ' αυτήν την περίπτωση, την κεντρική γραμμή τη θέτουμε στο σημείο

$$n\hat{p} = \frac{\sum_{i=1}^{k} x_i}{k}$$

όπου k ο αριθμός των δειγμάτων και x_i ο αριθμός των ελαττωματικών στο δείγμα i. Τα όρια ελέγχου είναι τα εξής:

$$AOE_{np} = n\hat{p} + 3\hat{\sigma}_{np} = n\hat{p} + 3\sqrt{n\hat{p}(1-\hat{p})}$$

$$KOE_{np} = n\hat{p} - 3\hat{\sigma}_{np} = n\hat{p} - 3\sqrt{n\hat{p}(1-\hat{p})}$$

όπου p είναι ο λόγος των ελαττωματικών που βρέθηκαν προς όσα εξετάσθηκαν.

3.3.8 Έλεγχος φυσικής δυνατότητας παραγωγικής διαδικασίας

Στις προηγούμενες παραγράφους είδαμε πως μπορούμε να ελέγχουμε κατά πόσο η παραγωγική διαδικασία παραμένει στατιστικά αμετάβλητη. Βέβαια, αυτό έχει πρακτικά ενδιαφέρον μόνο αν έχουμε καθορίσει τι θέλουμε και ακόμη έχουμε βεβαιωθεί ότι η παραγωγική διαδικασία μπορεί πράγματι να μας το δώσει. Συνεπώς, πριν από την ανάπτυξη οποιουδήποτε διαγράμματος ελέγχου πρέπει να αναγνωρίζουμε τις φυσικές δυνατότητες της παραγωγικής διαδικασίας και να διαπιστώνουμε ότι πράγματι αυτή μπορεί να μας δώσει τις προδιαγραφές που θέλουμε. Κάτι τέτοιο είναι χρήσιμο να το κάνουμε όταν αρχίζουμε μια ορισμένη παραγωγή για να καθορίσουμε του συγκεκριμένου συνδυασμού μηχανών – εργαλείων – υλικών και χειριστού. Αλλά όχι μόνο, αφού είναι σκόπιμο να το κάνουμε προτού παραλάβουμε νέο εξοπλισμό μιας παραγωγικής διαδικασίας. Και ακόμη, μπορούμε να το κάνουμε στους χώρους παραγωγής των προμηθευτών μας για να έχουμε προελέγξει ότι πρόκειται να παραλάβουμε κ.τ.λ.

Για να προσδιορίσουμε τις δυνατότητες της παραγωγικής διαδικασίας παράγουμε τουλάχιστον 30 διαδοχικές μονάδες προϊόντων. Με βάση τις τιμές αυτού του δείγματος, προσπαθούμε να διαπιστώσουμε το είδος της κατανομής που ακολουθεί ο πληθυσμός από τον οποίο προέρχεται το δείγμα. Αυτό το κάνουμε με τη γνωστή από τη Στατιστική διαδικασία του Ελέγχου Προσαρμογής. Υπενθυμίζουμε ότι έλεγχο προσαρμογής μιας κατανομής συχνότητας μπορούμε να κάνουμε και με γραφική μέθοδο. Είτε με τον ένα τρόπο είτε με τον άλλο θα βρούμε το είδος της κατανομής που ακολουθεί ο πληθυσμός και θα εκτιμήσουμε τις παραμέτρους του. Η φυσική δυνατότητα της παραγωγικής διαδικασίας ορίζεται από τα όρια $\mu \pm 3\sigma$. Βέβαια, τα μ και σ δεν τα ξέρουμε σχεδόν ποτέ γιατί αναφέρονται στον πληθυσμό. Γι’ αυτό αντί των μ και σ χρησιμοποιούμε τις εκτιμήτριές τους, που είναι η x και η σ. Η πρώτη είναι η μέση τιμή του δείγματος, ενώ η δεύτερη μπορεί να είναι είτε η τυπική απόκλιση του δείγματος, είτε η σ. Συνεπώς, τα όρια των φυσικών δυνατοτήτων της παραγωγικής διαδικασίας είναι τα

$$x \pm 3\hat{s}$$

ή

$$\bar{x} \pm 3\frac{\bar{R}}{\alpha_n}$$

Οι τιμές του α_n βρίσκονται στον ακόλουθο πίνακα:

Πίνακας 18: Συντελεστές για Υπολογισμό Ορίων Ελέγχου

Μέγεθος δείγματος	α_n	b_n
2	1,128	0,853
3	1,693	0,888
4	2,059	0,880
5	2,326	0,864
6	2,534	0,848
7	2,704	0,833
8	2,847	0,820
9	2,970	0,808
10	3,078	0,797
12	3,258	0,778
14	3,407	0,762
16	3,532	0,749
18	3,640	0,738
20	3,735	0,729
22	3,819	0,720
24	3,895	0,712

Πρέπει να σημειώσουμε ότι η εκτιμήτρια σ είναι των μεμονωμένων τιμών. Ενώ για τα όρια ελέγχου του διαγράμματος X χρησιμοποιούμε τη σx, που είναι η τυπική απόκλιση των μέσων τιμών των δειγμάτων. Η σ είναι μεγαλύτερη από τη σx, αφού συνδέεται με τη σχέση

$$\hat{\sigma} = \sigma_x\sqrt{n}$$

Όταν γνωρίζουμε τα όρια των φυσικών δυνατοτήτων της παραγωγικής διαδικασίας, μπορούμε να δούμε πως αυτά σχετίζονται με τις προδιαγραφές, δηλαδή με το 2T. Τρεις περιπτώσεις μπορούν να παρουσιαστούν:

1. 2T>6σ , οπότε η ανοχή βρίσκεται μέσα στα φυσικά όρια της διαδικασίας.
2. 2T=6σ, οπότε τα όρια των ανοχών συμπίπτουν με τα φυσικά όρια της διαδικασίας.
3. 2T<6σ, οπότε η ανοχή βρίσκεται έξω από τα φυσικά όρια της διαδικασίας και οπωσδήποτε θα παραχθούν ελαττωματικά.

Έτσι, γίνεται φανερό ότι φυσική δυνατότητα μιας παραγωγικής διαδικασίας λέμε την ελάχιστη ανοχή την οποία μπορεί να δώσει χωρίς να παράγει ελαττωματικά.

Παράδειγμα 16

Ένα εξάρτημα θέλουμε να έχει διάσταση ανάμεσα στα όριο 101,32±63 mm. Από ένα δείγμα μεγέθους 5 προέκυψε ότι η μέση τιμή τους είναι 101,25 mm και το εύρος ίσο με 0,55 mm. Να βρεθούν τα όρια των φυσικών δυνατοτήτων της διαδικασίας και να διαπιστωθεί κατά πόσο αυτή μπορεί να ανταποκριθεί στις απαιτήσεις των προδιαγραφών.

Τα φυσικά όρια της διαδικασίας τα υπολογίζουμε από τη σχέση

$$\bar{x} \pm 3\frac{\bar{R}}{\alpha_n}$$

Για n = 5 έχουμε ότι an=2,326

Συνεπώς,

$$\bar{x} \pm \frac{\bar{R}}{\alpha_n} = 101{,}25 \pm \frac{0{,}55}{2{,}326} = 101{,}25 \pm 0{,}70.$$

Έτσι, τα όρια των δυνατοτήτων της διαδικασίας είναι 101,95 και 100,55 mm. Επειδή τα όρια των ανοχών είναι 101,95 και 100,69 mm, η διαδικασία θα δημιουργεί διαρκώς ελαττωματικά. Ελαττωματικά θα είναι όλα όσα θα έχουν διάσταση που θα βρίσκεται ανάμεσα στις τιμές 100,55 και 100,69 mm. Κατά γνωστά από την κανονική κατανομή βρίσκουμε ότι το ποσοστό των παραγομένων προϊόντων ανάμεσα στα όρια αυτά θα είναι ίσο με 1,9%.

3.3.9 Χρήσιμες πρακτικές διαπιστώσεις

Στις προηγούμενες παραγράφους είδαμε τη στατιστική φύση των διαφόρων μεγεθών, που μας διαμορφώνουν οι παραγωγικές διαδικασίες, και τον τρόπο με τον οποίο μπορούμε να τα ελέγχουμε. Το γεγονός ότι μπορούμε τεχνικά να τα ελέγχουμε δε σημαίνει ότι θα το κάνουμε οπωσδήποτε. Αντίθετα, με μεγάλη προσοχή πρέπει να αποφασίσουμε τι θα ελέγξουμε και πως θα το ελέγξουμε. Για να αποφασίσουμε τι θα ελέγξουμε πρέπει να σκεφθούμε ότι κάθε προϊόν μπορεί να προσδιορίζεται με πάρα πολλά χαρακτηριστικά. Τα χαρακτηριστικά αυτά μπορεί να είναι φυσικά, χημικά, λειτουργικά κ.τ.λ. Παραδείγματος χάρη, ένα μηχανουργικό προϊόν χαρακτηρίζεται από τις 3 διαστάσεις των μερών του, από τη σύσταση του υλικού του, από την αντοχή του σε ορισμένου είδους καταπόνηση κ.τ.λ. Όλα αυτά τα χαρακτηριστικά βέβαια είναι μάλλον απίθανο να συμβάλουν εξίσου στο βαθμό που το προϊόν ικανοποιεί την ανάγκη την οποία προορίζεται να καλύψει. Συνεπώς, πρέπει να αναγνωρίσουμε εκείνα που είναι κρίσιμα στη λειτουργία του προϊόντος και αυτά να ελέγξουμε.

Για τον έλεγχο μπορούμε να χρησιμοποιήσουμε τα διαγράμματα X και R. Ο συνδυασμός τους διαμορφώνει μια πολύ ευαίσθητη διαδικασία ελέγχου, αφού τα διάγραμμα X μας ελέγχει το επίπεδο ποιότητας, ενώ το διάγραμμα R την ομοιομορφία της παραγωγής μας. Το διάγραμμα p και np ούτε είναι τόσο ευαίσθητα, ούτε μας λένε τι κάνει ένα προϊόν ελαττωματικό. Απλώς μας λένε ότι είναι ελαττωματικό. Πέρα όμως από τα διαγράμματα αυτά μπορούμε να σχεδιάσουμε κι άλλα, αφού πρέπει να έχει γίνει φανερό ότι τα διαγράμματα ελέγχου δεν είναι τίποτα άλλο από γραφικές παραστάσεις γνωστών στατιστικών αναλύσεων. Και συγκεκριμένα, δεν είναι τίποτα άλλο από γραφικές παραστάσεις των διαφόρων ελέγχων υποθέσεων. Τα όρια ελέγχου απλώς μας οριοθετούν την περιοχή αποδοχής της υποθέσεως που ελέγχουμε. Έτσι, αν για οποιοδήποτε λόγο θέλουμε να παρακολουθήσουμε κάποιο άλλο χαρακτηριστικό – πέρα από τη μέση τιμή, το εύρος, τη διωνυμική αναλογία ή τον αριθμό των ελαττωματικών σ’ ένα ορισμένο μέγεθος δείγματος- δεν έχουμε παρά να σχεδιάσουμε τον κατάλληλο έλεγχο υποθέσεων.

Τόσο για να σχεδιάσουμε, όσο και για να λειτουργήσουμε ένα από τα παραπάνω διαγράμματα ελέγχου χρειάζεται να χρησιμοποιήσουμε ένα ορισμένο μέγεθος δείγματος. Το μέγεθος του δείγματος κανονικά πρέπει να το προσδιορίσουμε έτσι ώστε να εξασφαλίζει γρήγορα τον εντοπισμό κάποιας αλλαγής στη διαδικασία. Αυτό θεωρητικά μπορούμε να το κάνουμε με τη βοήθεια της χαρακτηριστικής καμπύλης κάθε ελέγχου. Στην πράξη όμως, που εργαζόμαστε συνήθως με δύο διαγράμματα ελέγχου: ένα για να ελέγχουμε τη μέση τιμή και ένα τη μεταβλητότητα, είναι απίθανο ένα μέγεθος δείγματος να είναι το καλύτερο και για τα δύο διαγράμματα. Γι’ αυτό ακολουθούμε μια συμβιβαστική λύση: αυτή είναι να χρησιμοποιούμε μεγέθη δειγμάτων 4 ή 5 για τα διαγράμματα X και R και μεγέθη 60, 120 και 600 για το διάγραμμα p όταν το μέσο ποσοστό ελαττωματικών είναι 0,1 – 0,05 και 0,01 αντίστοιχα.

Η συχνότητα εξάλλου με την οποία παίρνουμε τα παραπάνω δείγματα πρέπει να είναι τέτοια που να περιλαμβάνουμε στους ελέγχους μας το 1 – 10% του συνόλου της παραγωγής, όταν χρησιμοποιούμε διαγράμματα X και R. Ενώ πρέπει να είναι τέτοια ώστε να περιλαμβάνουμε το 20 – 50% όταν χρησιμοποιούμε διάγραμμα p. Τα κατώτερα όρια των ποσοστών αυτών τα χρησιμοποιούμε όταν η διαδικασία είναι στατιστικά μάλλον σταθερή. Ενώ τα ανώτερα, όταν δεν είναι.

Τέλος, από την εικόνα των διαγραμμάτων συμπεραίνουμε ότι η διαδικασία βρίσκεται εκτός ελέγχου όταν παρατηρούμε τα παρακάτω:

Ένα ή περισσότερα σημεία πάνω από το ΑΟΕ ή κάτω από το ΚΟΕ.

Πολλά συνεχόμενα σημεία κοντά στο ένα όριο.

Σαφή τάση προς τα πάνω ή τα κάτω διαδοχικών σημείων

Γενικά, όταν τα σημεία στα διάφορα διαγράμματα, που εξετάσαμε βρίσκονται από τη μια μέσα στα όρια ελέγχου και από την άλλη η διαδοχή τους είναι τυχαία, δηλαδή δεν παρουσιάζουν μόνιμη τάση, συμπεραίνουμε ότι η διαδικασία είναι στατιστικά σταθερή, δηλαδή δε δημιουργεί αποκλίσεις εξαιτίας συστηματικών αιτίων.

Συμπεράσματα

Όπως γίνεται αντιληπτό από τα προηγούμενα κεφάλαια, η σύγχρονη επιστήμη της διοίκησης των επιχειρήσεων γίνεται σήμερα με επιστημονικές μεθόδους, ώστε να εξασφαλιστεί ο συντονισμός και ο έλεγχος της λειτουργίας των επιχειρήσεων για την επίτευξη του σκοπού τους, που είναι το κέρδος. Όπως διαπιστώνεται στην παρούσα διπλωματική εργασία, ο ρόλος της Στατιστικής είναι καθοριστικός σε πολλά στάδια της παραγωγικής διαδικασίας.

Κάποια από τα πλέον σημαντικά και ζωτικής σημασίας πεδία δραστηριότητας των σύγχρονων επιχειρήσεων είναι αυτά του ελέγχου ποιότητας των εμπορευμάτων, του τρόπου διοίκησης και λειτουργίας των παραγωγικών μονάδων, της έρευνας των αγορών, της διαφήμισης, καθώς και των διαφόρων τάσεων της αγοράς, ως προς την τιμή, τη διακίνηση και την προώθηση του προϊόντος. Σε κάθε περίπτωση οι στατιστικές έρευνες είναι αναγκαίες για να διαπιστώνουμε την ποιότητα και το πλήθος των ελαττωματικών παραγόμενων προϊόντων, το χρόνο και το κόστος παραγωγής με χρήση διαφορετικών διαδικασιών, μηχανών κ.α., τα αποτελέσματα μιας διαφήμισης, καθώς και ο έλεγχος ύπαρξης τρωτών σημείων στην όλη διακίνηση του προϊόντος, με αποτέλεσμα η κάθε επιχείρηση να παρεμβαίνει έγκαιρα, για να αποφεύγονται δυσάρεστες εξελίξεις.

Η συμβολή της Στατιστικής στη διαδικασία λήψης επιχειρηματικών αποφάσεων είναι ουσιώδης, ιδιαίτερα μάλιστα όταν αναφερόμαστε σε συλλογικό επίπεδο υψηλόβαθμων στελεχών και όχι τόσο στις ατομικές καθημερινές και συχνές αποφάσεις που λαμβάνονται από τους υπαλλήλους-στελέχη της επιχείρησης.

Επίσης η Στατιστική αναλύοντας σειρά δεδομένων προηγούμενων χρονικών περιόδων, της επιχείρησης και του κλάδου στον οποίο ανήκει, έχει τη δυνατότητα με ικανοποιητική προσέγγιση να προβλέψει και να πληροφορήσει τη διοίκηση της επιχείρησης για τις πωλήσεις του επόμενου έτους ή της επόμενης πενταετίας.

Τέλος, προαπαιτούμενο προσόν για την παροχή υπηρεσιών ή προϊόντων υψηλής ποιότητας από μέρους της επιχείρησης, ώστε να αναπτυχθεί, είναι να διακρίνεται και η ίδια για την ποιότητα της διαχείρισής της και την αποτελεσματική διοίκηση. Ένα προϊόν θεωρείται ποιοτικά καλό, όταν ανταποκρίνεται στις προδιαγραφές και στις επιθυμίες του καταναλωτή, τις οποίες οι επιστημονικά και ποιοτικά οργανωμένες επιχειρήσεις έχουν καταγράψει από προηγούμενες έρευνες αγορών. Για να επιτευχθούν τα παραπάνω, η επιχείρηση θα πρέπει να έχει μία καλή σχέση κόστους-απόδοσης. Αυτό σημαίνει ότι η επιχείρηση έχει μια διαχείριση ποιότητας, η οποία τη βοηθά να διακριθεί από τις υπόλοιπες του κλάδου της. Η επιτυχία των στόχων της (αποτελεσματικότητα) είναι συνάρτηση των ποιοτικών χαρακτηριστικών της.

Επομένως, μπορούμε συνοψίζοντας να πούμε ότι με τη βοήθεια των Στατιστικών μεθόδων που αναφέρθηκαν και των μέτρων διασποράς είναι εφικτός ο έλεγχος τόσο της ποιότητας των προϊόντων, όσο και των παρεχόμενων υπηρεσιών και της διαχείρισης των επιχειρήσεων, καθιστώντας τη Στατιστική ως ένα απαραίτητο εργαλείο στη σύγχρονη διοίκηση των επιχειρήσεων.

Ηλεκτρονικές πηγές

- http://users.auth.gr/~agpapana/StatLogistics
- http://digitalschool.minedu.gov.gr
- http://el.wikipedia.org
- www.stat-athens.aued.gr
- http://karagian.users.uth.gr/cscl/22-Katsanos-Avouris.pdf

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική

- Γναρδέλλης Χ. (2003). Εφαρμοσμένη Στατιστική, Εκδόσεις Παπαζήση
- Δαμιανού Χ. (1996). Στατιστικός Έλεγχος Ποιότητας και Αξιοπιστία, Εκδόσεις Συμμετρία.
- Δρόσος Γ., Καραπιστόλης Δ. (1994). Στατιστική των επιχειρήσεων, Εκδόσεις Έλλην.
- Καραγεώργιος Δ. Λ. (2001) Στατιστική Περιγραφική και Επαγωγική, Εκδόσεις Σαββάλας
- Καραπιστόλη Δ. Ν. (2005) Στατιστική Επιχειρήσεων, Εκδόσεις Ανικούλα
- Καρυώτης Β. (2001)Στατιστική επιχειρήσεων Α'τάξης-Τ.Ε.Ε. τομέας οικονομίας και διοίκησης Εκδόσεις Μέντορας
- Κιόχος Π. Α. (1993) Στατιστική, Εκδόσεις Interbooks
- Κονδύλης Ε.Κ. (1996). Στατιστικές Τεχνικές Διοίκησης Επιχειρήσεων, Εκδόσεις Interbooks
- Παλαμάς Β. (2009) Στατιστικές Τεχνικές και εφαρμογές στις Επιστήμες της Αγωγής, Εκδόσεις Πατάκη
- Ταγαράς Γ. Ν. (2001) Στατιστικός Έλεγχος Ποιότητας, Εκδόσεις Ζήτη
- Ψώνιου Δ. Π. (1986) Οργάνωση και Διοίκηση εργοστασίων. Δεύτερος Τόμος. Προγραμματισμός και έλεγχος παραγωγής, Εκδόσεις Ζήτη

Ξένη

- Douglas C. M. (2005). Introduction to Statistical Quality Control, Fifth Edition, John Wiley & Sons, Inc.

- Desrosières, Alain (2005): Η πολιτική των μεγάλων αριθμών. Ιστορία της στατιστικής αντίληψης. Επιστημονική επιμέλεια και μετάφραση: Ηλίας Αθανασιάδης. Αθήνα: Μεταίχμιο

- Hans-Joachim M. & Horst R. (1993). Statistical Methods for Quality Assurance, Chapman & Hall

- Practical Marketing Research, Pope. 1981 p.216

- Thomas P. R. (2000). Statistical Methods for Quality Improvementl, Second Edition,John Wiley & Sons, Inc.