

ΤΕΙ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ / ΜΕΣΟΛΟΓΓΙ

Πτυχιακή εργασία

ΣΤΑΤΙΣΤΙΚΑ ΜΕΤΡΑ ΚΑΙ ΟΙ ΕΦΑΡΜΟΓΕΣ ΤΟΥΣ
ΣΤΙΣ ΕΠΙΧΕΙΡΗΣΕΙΣ ΚΑΙ ΤΗΝ ΟΙΚΟΝΟΜΙΑ

Παυλίδου Βασιλική

Μεσολόγγι 2014

ΤΕΙ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ / ΜΕΣΟΛΟΓΓΙ

Πτυχιακή εργασία

ΣΤΑΤΙΣΤΙΚΑ ΜΕΤΡΑ ΚΑΙ ΟΙ ΕΦΑΡΜΟΓΕΣ ΤΟΥΣ
ΣΤΙΣ ΕΠΙΧΕΙΡΗΣΕΙΣ ΚΑΙ ΤΗΝ ΟΙΚΟΝΟΜΙΑ

Παυλίδου Βασιλική

Επιβλέπων καθηγητής
Αθανάσιος Μεγαρίτης

Μεσολόγγι 2014

Η έγκριση της πτυχιακής εργασίας από το Τμήμα Διοίκησης Επιχειρήσεων/Μεσολογγίου του ΤΕΙ Δυτικής Ελλάδας δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων του συγγραφέα εκ μέρους του Τμήματος.

ΠΕΡΙΛΗΨΗ

Βασικές έννοιες της Στατιστικής έχουν εισχωρήσει και ενσωματωθεί σε όλες τις επιστήμες (Ανθρωπιστικές, Κοινωνικές, Οικονομικές, Τεχνολογικές, Επιστήμες Υγείας, κ.ά.). Η ανάλυση σύνθετων φαινομένων απαιτεί από τους στατιστικούς τη χρησιμοποίηση των καταλληλότερων στατιστικών μεθόδων στις μεγάλες απαιτήσεις των παραπάνω επιστημονικών χώρων. Οι στατιστικές τεχνικές είναι απαραίτητες για την επιτυχή λειτουργία κάθε επιχείρησης γιατί παρέχουν στη Διοίκηση όλες τις απαραίτητες πληροφορίες που αφορούν την ίδια την επιχείρηση (εσωτερικό περιβάλλον) αλλά και τους ανταγωνιστές (εξωτερικό περιβάλλον). Η εργασία αναλύει όλα τα στατιστικά μέτρα που χρησιμοποιούμε, πιο συγκεκριμένα τα μέτρα θέσης ή κεντρικής τάσης, μέτρα διασποράς και τα μέτρα ασυμμετρίας και κύρτωσης. Εξηγούμε τη χρησιμότητα των στατιστικών μέτρων στις επιχειρήσεις. Τέλος, ενσωματώνουμε τη θεωρία στην πράξη με τη χρήση του στατιστικού προγράμματος SPSS και παρουσιάζουμε εφαρμογές οι οποίες μας βοηθούν να κατανοήσουμε τη χρησιμότητα της στατιστικής στις επιχειρήσεις.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΕΡΙΛΗΨΗ.....	iii
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ	iv
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ	vii
ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ	ix
ΑΠΟΔΟΣΗ ΟΡΩΝ	x
ΕΙΣΑΓΩΓΗ.....	xi
1 Στατιστική – Εννοιολογική Προσέγγιση.....	3
1.1 Τι είναι Στατιστική	3
1.2 Ιστορική Αναδρομή	4
1.3 Στατιστική Μονάδα - Στατιστικός Πληθυσμός.....	6
1.4 Άλλες Βασικές έννοιες Στατιστικής.....	7
1.5 Μέθοδοι Συλλογής Στοιχείων	12
1.6 Στατιστικοί πίνακες	16
Πίνακες Κατανομής Συχνοτήτων.....	17
Κατανομή Δεδομένων.....	20
1.7 Στατιστικά Διαγράμματα.....	21
1.7.1 Παραποιήσεις με γραφικές παραστάσεις	27
1.8 Αξιολόγηση Στατιστικών Μελετών	29
2 Αριθμητική Περιγραφή Δεδομένων.....	29
2.1 Μέτρα Θέσης ή Κεντρικής Τάσης.....	30
2.1.1 Αριθμητικός Μέσος.....	30
2.1.2 Παραλλαγές του Αριθμητικού Μέσου	32
2.1.3 Διάμεσος.....	32

2.1.4	Επικρατούσα τιμή.....	34
2.1.5	Γεωμετρικός Μέσος (GM)	34
2.1.6	Αρμονικός Μέσος (HM).....	35
2.1.7	Ποσοστιαία Σημεία (percentiles).....	36
2.2	Σύγκριση Μέτρων Θέσης.....	38
2.2.1	Πλεονεκτήματα και Μειονεκτήματα των μέτρων Κεντρικής Τάσεως.....	38
3	Μέτρα διασποράς.....	39
3.1	Μέτρα μεταβλητότητας ή διασποράς (Measures of variation or dispersion).....	39
3.1.1	Έκταση ή Εύρος (Range)	40
3.1.2	Ενδοτεταρτημοριακό Εύρος (Interquartile Range)	40
3.1.3	Μέση Απόλυτη Απόκλιση (Mean Absolute Deviation).....	40
3.1.4	Διακύμανση ή Διασπορά (Variance).....	41
3.1.5	Τυπική Απόκλιση (Standard Deviation).....	42
3.1.6	Βαθμοί Ελευθερίας (Degrees of Freedom).....	44
3.1.7	Χρήση της τυπικής απόκλισης	44
3.1.8	Εμπειρικός Κανόνας.....	45
3.1.9	Τυποποιημένες τιμές (Standardized Values).....	45
3.1.10	Μέτρα Σχετικής Μεταβλητότητας	47
3.1.11	Συντελεστής Μεταβλητότητας (Coefficient of Variation).....	47
3.1.12	Μέση Διαφορά του Gini (Gini's mean difference)	47
3.1.13	Καμπύλη Συγκέντρωσης ή καμπύλη Lorenz.....	48
3.2	Σύγκριση μέτρων διασποράς.....	50
3.3	Θηκόγραμμα.....	51
4	Μέτρα Ασυμμετρίας και Κύρτωσης	52
4.1	Συντελεστής λοξότητας.....	53

4.2	Συντελεστής κύρτωσης.....	54
5	Συντελεστή γραμμικής συσχέτισης.....	55
6	Στατιστική και Επιχειρήσεις	56
6.1	Αβεβαιότητα Οικονομικού Περιβάλλοντος	56
6.2	Έννοια της Χρεοκοπίας	59
6.3	Μεθοδολογίες πρόγνωσης της χρεοκοπίας	60
6.4	Τα Στοιχεία των Επιχειρήσεων.....	66
6.5	Εφαρμογές στατιστικών μέτρων στις επιχειρήσεις.....	74
	ΣΥΜΠΕΡΑΣΜΑΤΑ.....	99
	ΕΠΙΛΟΓΟΣ	103
	ΒΙΒΛΙΟΓΡΑΦΙΑ.....	104
	Πνευματικά δικαιώματα	106

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1: Ποσοστό στατιστικών ατυχημάτων 2011 - 2012.....	17
Πίνακας 2: Πίνακας συχνοτήτων αδερφών μίας τάξης.....	18
Πίνακας 3: Πίνακας συχνοτήτων και αθροιστικών συχνοτήτων αδερφών μιας τάξης.....	19
Πίνακας 4: Πίνακας συχνοτήτων παράδειγμα 1.....	32
Πίνακας 5: Chebyshev.....	44
Πίνακας 6: Παράδειγμα τυποποιημένων τιμών.....	46
Πίνακας 7: Πίνακας συχνοτήτων εισοδήματος.....	49
Πίνακας 8: Πίνακας πωλήσεων σε τεμάχια 2013.....	74
Πίνακας 9: Πίνακας αξίες πωλήσεων και επιστροφών.....	75
Πίνακας 10: Πίνακας στατιστικών στοιχείων εφαρμογής 1.....	75
Πίνακας 11: Σύγκριση στατιστικών μέτρων 2013-2014 εφαρμογής 1.....	78
Πίνακας 12: Πίνακας αξίες πωλήσεων - αξίες επιστροφών εφαρμογής 1.....	79
Πίνακας 13: Στατιστικά μέτρα εφαρμογής 1 με χρήση SPSS.....	81
Πίνακας 14: Πίνακας ενοικίων και ορόφων εφαρμογής 2.....	81
Πίνακας 15: Στατιστικά στοιχεία εφαρμογής 2.....	82
Πίνακας 16: Στατιστικά μέτρα εφαρμογής 2 με χρήση SPSS.....	83
Πίνακας 17: Πίνακας μισθών υπαλλήλων εφαρμογής 3.....	83
Πίνακας 18: Στατιστικά στοιχεία εφαρμογής 3.....	84
Πίνακας 19: Στατιστικά στοιχεία εφαρμογής 4.....	87
Πίνακας 20: Συσχέτιση (x) με (y).....	88
Πίνακας 21: Στατιστικά στοιχεία εφαρμογής 4.....	89

Πίνακας 22: Στατιστικά μέτρα (x) εφαρμογής 4.....	91
Πίνακας 23: Στατιστικά μέτρα (y) εφαρμογής 4.....	91
Πίνακας 24: Πίνακας κλάσεων εφαρμογής 5.....	95

ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ

Εικόνα 1: Κυκλικό διάγραμμα θανατηφόρων τροχαίων ατυχημάτων	23
Εικόνα 2: Χρονόγραμμα.....	24
Εικόνα 3: Διάγραμμα σημείων	25
Εικόνα 4: Διάγραμμα μίσχου φύλλου	26
Εικόνα 5: Γράφημα πωλήσεων.....	28
Εικόνα 6: Διαφορά διαγραμμάτων με τις ίδιες πωλήσεις.....	28
Εικόνα 7: Εικόνα 7: Καμπύλη Lorenz εισοδήματος	49
Εικόνα 8: Θηκόγραμμα	52
Εικόνα 9: Συντελεστής λοξότητας	53
Εικόνα 10: Συντελεστής κύρτωσης	54
Εικόνα 11: Θηκόγραμμα εφαρμογής 1.....	77
Εικόνα 12: Διάγραμμα συσχέτισης εφαρμογής 1.....	80
Εικόνα 13: Ραβδόγραμμα μισθοδοσίας εφαρμογής 3	85
Εικόνα 14: Κυκλικό διάγραμμα μισθοδοσίας εφαρμογής 3.....	85
Εικόνα 15: Στατιστικά μέτρα εφαρμογής 3 με χρήση SPSS.....	86
Εικόνα 16: Θηκόγραμμα (x) εφαρμογής 4.....	92
Εικόνα 17: Διάγραμμα μίσχου-φύλλου (x) εφαρμογής 4.....	92
Εικόνα 18: Θηκόγραμμα (y) εφαρμογής 4.....	93
Εικόνα 19: Διάγραμμα μίσχου φύλλου (y) εφαρμογής 4.....	93
Εικόνα 20: Ιστόγραμμα εφαρμογής 5.....	96

ΑΠΟΔΟΣΗ ΟΡΩΝ

Arithmetic mean	Αριθμητικός μέσος
Median	Διάμεσος
Mode	Επικρατούσα τιμή
a-trimmed mean	α-περικομμένος μέσος
Range	Εύρος
Mean absolute deviation	Μέση απόλυτη απόκλιση
Variance	Διασπορά
Standard deviation	Τυπική απόκλιση
Degrees of freedom	Βαθμοί ελευθερίας
Standardized values	Τυποποιημένες τιμές
Dot diagram	Διάγραμμα σημείων
Pie chart	Κυκλικό διάγραμμα
Slices	Κυκλικοί τομείς
Bar chart	Ραβδόγραμμα
Time series chart	Γράφημα χρονολογικής δειράς
Experimental design	Σχεδιασμός πειραμάτων
Descriptive statistics	Περιγραφική στατιστική

ΕΙΣΑΓΩΓΗ

Ο όρος «Στατιστική» χρησιμοποιείται, για να δηλώσει αριθμητικές πληροφορίες. Οι στατιστικές μελέτες παρουσιάζουν τα αποτελέσματα της συστηματικής συλλογής ενός συνόλου αριθμητικών πληροφοριών. Η λέξη «Στατιστική» μπορεί να προέρχεται από την αρχαία ελληνική λέξη στατίζω που σημαίνει (τοποθετώ, ταξινομώ, συμπεραίνω) ή από τη λατινική λέξη «status» που σημαίνει (πολιτεία, κράτος).

Η αρχαιότερη ίσως συλλογή στοιχείων θεωρείται η απογραφή πληθυσμού που έγινε το 2238 π.Χ. στην Κίνα από τον αυτοκράτορα Yao. Αργότερα στοιχειώδεις απογραφές φαίνεται να έχουν πραγματοποιηθεί από τους Αιγυπτίους και τους Πέρσες.

Στην αρχαιότητα, η συγκέντρωση στατιστικών στοιχείων είχε ως στόχο τον εντοπισμό των πολιτών που είχαν υποχρέωση: να υπηρετήσουν ως πολεμιστές ή να πληρώσουν φόρο.

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

Από το 16ο έως το 19ο αιώνα, η αλματώδης ανάπτυξη του εμπορίου ώθησε τις ηγεσίες των διαφόρων κρατών στη μελέτη οικονομικών δεδομένων, όπως την παραγωγικότητα των βιομηχανιών, το εξαγωγικό εμπόριο, κ.τ.λ. Ενώ αρχικά η Στατιστική ασχολείτο μόνο με την παράθεση τεραστίων αριθμητικών πινάκων, σήμερα μια στατιστική έρευνα χωρίζεται στα εξής στάδια:

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

Ο κλάδος της Στατιστικής που ασχολείται με τα δύο πρώτα στάδια λέγεται σχεδιασμός πειραμάτων (experimental design). Ο κλάδος που ασχολείται με το τρίτο στάδιο είναι η περιγραφική στατιστική (descriptive statistics).

Η επαγωγική στατιστική περιλαμβάνει τις μεθόδους με τις οποίες γίνεται η προσέγγιση των χαρακτηριστικών ενός μεγάλου συνόλου δεδομένων, από τη μελέτη των χαρακτηριστικών ενός μικρού υποσυνόλου των δεδομένων.

Σε κάθε χώρα έχουν δημιουργηθεί αυτοτελείς στατιστικοί οργανισμοί με σκοπό τον αποτελεσματικό συντονισμό όλων των στατιστικών εργασιών. Τέτοιος οργανισμός στην Ελλάδα είναι Ε.Σ.Υ.Ε. (Εθνική Στατιστική Υπηρεσία) (<http://www.statistics.gr/>). Οι στατιστικές που πραγματοποιεί η Ε.Σ.Υ.Ε. είναι μηνιαίες, τριμηνιαίες, ετήσιες, ανά 5ετία και ανά 10ετία, και καλύπτουν όλους σχεδόν τους τομείς δραστηριότητας. Πληθυσμιακά στοιχεία, στοιχεία απασχόλησης και ανεργίας, στοιχεία που αφορούν την υγεία, την κοινωνική ασφάλιση, την παιδεία, την παραγωγική διαδικασία, τις τιμές, το εθνικό εισόδημα κ.τ.λ. Βασικός χρήστης των στατιστικών και δεικτών που καταρτίζει η Ε.Σ.Υ.Ε. είναι το Κράτος, που χρησιμοποιεί τα στοιχεία αυτά για να χαράξει τις επιμέρους πολιτικές στους διάφορους τομείς. Η Ευρωπαϊκή Ένωση με τη βοήθεια του ευρωπαϊκού στατιστικού γραφείου EUROSTAT (<http://europa.eu.int/comm/eurostat/>) χρειάζεται τα επιμέρους στοιχεία των κρατών-μελών, για να συνθέσει τις ευρωπαϊκές στατιστικές.

Επίσης, υπάρχουν και διεθνείς οργανισμοί όπως η UNESCO με το στατιστικό της ινστιτούτο (<http://www.uis.unesco.org>) με αντικείμενο τη συλλογή, παρουσίαση και την επεξεργασία αριθμητικών πληροφοριών για τα επί μέρους κράτη και τις μεταξύ τους οικονομικές σχέσεις. Βασικές έννοιες της Στατιστικής έχουν εισχωρήσει και ενσωματωθεί σε όλες τις επιστήμες (Ανθρωπιστικές, Κοινωνικές, Οικονομικές, Τεχνολογικές, Επιστήμες Υγείας, κ.ά.).

Η ανάλυση σύνθετων φαινομένων απαιτεί από τους στατιστικούς τη χρησιμοποίηση των καταλληλότερων στατιστικών μεθόδων στις μεγάλες απαιτήσεις των παραπάνω στατιστικών χώρων. Μια σύντομη περιγραφή για το τι θα δούμε στην παρούσα εργασία:

Στο πρώτο κεφάλαιο αναφέρουμε τις βασικές έννοιες που έχουν δοθεί στη στατιστική μέχρι σήμερα, διάφορες μεθόδους με τις οποίες μπορούμε να συλλέξουμε στοιχεία και με ποιούς τρόπους μπορούμε να τα παρουσιάσουμε.

Στο δεύτερο κεφάλαιο αναλύουμε όλα τα μέτρα θέσης. Παρουσιάζουμε μια σύγκριση μεταξύ τους αναλύοντας τα πλεονεκτήματα-μειονεκτήματα τους.

Στο τρίτο κεφάλαιο θα αναλύσουμε αντίστοιχα τα μέτρα διασποράς και τα πλεονεκτήματα-μειονεκτήματα τους.

Το τέταρτο κεφάλαιο αναφέρεται στα μέτρα ασυμμετρίας και κύρτωσης.

Στο πέμπτο κεφάλαιο θα αναφερθούμε στο συντελεστή γραμμικής συσχέτισης και στη σημαντικότητα του σε μια επιχείρηση.

Στο τελευταίο κεφάλαιο που είναι και ο σκοπός της παρούσας εργασίας θα αναλύσουμε τη σημαντικότητα της στατιστικής και ειδικότερα των στατιστικών μέτρων στις επιχειρήσεις και την οικονομία εφαρμόζοντας παραδείγματα με όλα αυτά που είχαμε αναλύσει στα προηγούμενα κεφάλαια με τη χρήση του SPSS.

1 Στατιστική – Εννοιολογική Προσέγγιση

1.1 Τι είναι Στατιστική

Γενικά ο όρος Στατιστική φέρεται με διττή σημασία, αφενός υποδηλώνοντας μαθηματικές μεθόδους χειρισμού δεδομένων που λήφθηκαν με απαρίθμηση ή μέτρηση και αφετέρου αυτά τα ίδια τα δεδομένα που έχουν υποστεί αυτούς τους χειρισμούς.

Σύμφωνα με τον Α. Αλεξόπουλο εξετάζοντας τον ορισμό όπως αυτός καθορίστηκε στη δεκαετία του 1950 "Στατιστική είναι σύνολο μεθόδων που καθοδηγούν στη λήψη ορθών αποφάσεων σε περιπτώσεις αβεβαιότητας" τονίζει την εννοιολογική διάκριση του συνόλου των στοιχείων ενός φαινομένου και το σύνολο των μεθόδων που εξετάζουν αυτά προς τον κοινό σκοπό.

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

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

Ο όρος στατιστική είναι αρχαία ελληνική λέξη που ετυμολογείται από το αρχαίο ρήμα ίστημι και του εξ'αυτού παραγώγου ρήματος στατίζω που σημαίνει τοποθετώ, ταξινομώ, συμπεραίνω. Παράγωγο δε και ο στατήρας.

Τον όρο στατιστική αναφέρει ο Σωκράτης (Ξενοφώντας "Απομνημονεύματα") καθώς και ο Αριστοτέλης στο έργο του "Πολιτεία" απ' όπου και εισήλθε στη λατινική γλώσσα στη φράση *statisticum collegium* (διάλεξη για υποθέσεις της πολιτείας), από την οποία προήρθε με τη σειρά της η Ιταλική λέξη *statista*, που σημαίνει πολιτικός, και η Γερμανική λέξη *Statistik*, η οποία αρχικά αναφερόταν στην ανάλυση των δεδομένων για την πολιτεία. Τη σύγχρονη γενική έννοια της συλλογής και ταξινόμησης δεδομένων φέρεται να έλαβε στις αρχές του δεκάτου ένατου αιώνα.

1.2 Ιστορική Αναδρομή

Η Στατιστική ως έννοια εμφανίζεται από τους μυθικούς χρόνους από της πρώτης δημιουργίας οργανωμένων κοινωνιών. Μια πρώτη γραφή στατιστικής μορφής με αριθμητικά δεδομένα είναι ο νεών κατάλογος (κατάλογος των πλοίων) των Αχαιών στον Τρωικό πόλεμο από τον Όμηρο.

Από τον κατάλογο αυτό οι ιστορικοί απέσπασαν σημαντικές εκτιμήσεις της οικονομικής ευρωστίας και του πληθυσμού των πόλεων-κρατών που συμμετείχαν καθώς και σημαντικά στοιχεία για την τότε ναυπηγική, ναυτιλία και ναυτική τέχνη. Πρώτη ιστορική συλλογή καθαρά στατιστικών στοιχείων θεωρείται η απογραφή πληθυσμού από τον Αυτοκράτορα της Κίνας Γιάο (Υαο) το 2238 π.Χ. Στοιχειώδεις τέτοιες απογραφές είχαν πραγματοποιήσει και άλλοι αρχαίοι λαοί όπως οι αρχαίοι Αιγύπτιοι, οι Βαβυλώνιοι, οι Πέρσες οι αρχαίοι Έλληνες και οι Ρωμαίοι με χαρακτηριστικότερη την απογραφή του Οκταβιανού του Αυγούστου.

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

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

Η συστηματική όμως συλλογή δεδομένων για πληθυσμό και οικονομία (δημογραφική στατιστική) άρχισε στη διάρκεια της Αναγέννησης και ειδικότερα στη Βενετία και την Φλωρεντία όπου και γρήγορα επεκτάθηκε σ' όλα τα τότε Βασίλεια της Ευρώπης.

Περί το τέλος του 11ου αιώνα, επί εποχής Γουλιέλμου του Κατακτητή, πραγματοποιήθηκε μια σπουδαία στατιστική απογραφή που αφορούσε μονάδες παραγωγής της Αγγλίας, όπως μεταλλεία, ιχθυοτροφεία κ.λπ. Οι μεγάλοι όμως ρυθμοί θνησιμότητας που άρχισαν να παρατηρούνται λίγο αργότερα, από επιδημικές ασθένειες, πολέμους και λιμοκτονίες έδωσαν ιδιαίτερη ώθηση στη στατιστική έρευνα καταγράφοντας αιτίες και απώλειες.

Έτσι το 1348 ξεκίνησαν οι καταγραφές θανάτων από την πανώλη, την φοβερή ασθένεια που κράτησε τέσσερις αιώνες. Στις καταγραφές αυτές προστέθηκαν και θάνατοι από άλλες αιτίες. Το 1620 ο Άγγλος έμπορος Τζον Γκράουντ ξεκίνησε πρώτος τη δειγματοληπτική έρευνα σε οικογένειες του Λονδίνου όπου και διαπίστωσε ότι σε κάθε 88 άτομα υπήρχαν τρεις θάνατοι. Από το στοιχείο αυτό και χρησιμοποιώντας τους εν λόγω καταλόγους που έδιναν 13.200 θανάτους εκτιμήθηκε ότι ο πληθυσμός του Λονδίνου το 1620 αριθμούσε 387.000 κατοίκους.

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

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

1.3 Στατιστική Μονάδα - Στατιστικός Πληθυσμός

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

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

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

Ένα από τα βασικά στοιχεία που πρέπει να ορίζουμε για τον πληθυσμό είναι τα όριά του, δηλαδή ποιες είναι ακριβώς οι στατιστικές μονάδες του πληθυσμού που θα μελετηθούν.

Για παράδειγμα, αν θελήσει να μελετήσει ορισμένα χαρακτηριστικά των κατοίκων της Ραφήνας, δεν θα πρέπει να κάνει την έρευνα στο λιμάνι μια Παρασκευή απόγευμα ή Σάββατο πρωί, ή ένα Σαββατοκύριακο τους θερινούς μήνες. Δεν αρκεί, λοιπόν, μόνο ο προσδιορισμός των γεωγραφικών ορίων μιας πόλης.

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

1.4 Άλλες Βασικές έννοιες Στατιστικής

Πληθυσμός (population)

Είναι ένα σύνολο στοιχείων που μας ενδιαφέρει να μελετήσουμε. Χρησιμοποιείται και ο όρος ολότητα (universe).

Π.χ. Το σύνολο των ανθρώπων μιας περιοχής.

Διαδικασία (process)

Είναι ένα σύνολο περιορισμών που εμφανίζονται κατ' επανάληψη ώστε να μετατρέψουν πληροφορίες σε αποτελέσματα.

Παράδειγμα:

Η μελέτη φακέλων πελατών μιας ασφαλιστικής εταιρίας. Ο φάκελος παριστάνει αποτελέσματα πολλών διαδικασιών όπως η διαφημιστική διαδικασία της εταιρίας, η διαδικασία εξυπηρέτησης πελατών και η διοικητική διαδικασία. Κάθε διαδικασία περιλαμβάνει περιορισμούς που δρουν από κοινού για να δημιουργήσουν αποτελέσματα.

Πλαίσιο (frame)

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

Το πλαίσιο πρέπει να είναι όσο το δυνατό πιο αντιπροσωπευτικό του πληθυσμού που θέλουμε να μελετήσουμε.

Στο παράδειγμα με την ασφαλιστική εταιρία το πλαίσιο είναι το σύνολο των φακέλων των πελατών που ήταν διαθέσιμοι για μελέτη. Οι φάκελοι αυτοί προφανώς δεν περιλαμβάνουν νέες ασφάλειες που δεν έχουν ακόμα αρχειοθετηθεί ενώ ενδεχομένως περιλαμβάνουν ασφάλειες που έχουν σταματήσει αλλά οι φάκελοί τους δεν έχουν ακόμα ενημερωθεί.

Μεταβλητές

Τυχαία μεταβλητή ή απλά μεταβλητή ονομάζουμε οποιαδήποτε χαρακτηριστικό ως προς το οποίο εξετάζουμε ένα πληθυσμό και το οποίο παίρνει μια ή περισσότερες διαφορετικές τιμές. Για παράδειγμα το φύλο, η ηλικία, το ύψος, η τιμή του πετρελαίου σε διαφορετικές χρονικές στιγμές. Οι μεταβλητές συμβολίζονται με κεφαλαία γράμματα X, Ψ, Z και ο όρος παρατηρηθείσα τιμή ή παρατήρηση χρησιμοποιείται για την αριθμητική ή άλλη συμβολική της έκφραση.

Οι μεταβλητές διακρίνονται σε δύο μεγάλες κατηγορίες Ποιοτικές και Ποσοτικές. Ποιοτικές μεταβλητές είναι αυτές που αναφέρονται σε κάποιο ποιοτικό χαρακτηριστικό και οι τιμές τους δεν είναι αριθμητικές. Για παράδειγμα επίπεδο εκπαίδευσης, μητρική γλώσσα, βιοτικό επίπεδο, κ.τ.λ.

Επιπλέον οι ποιοτικές μεταβλητές χωρίζονται σε:

- Ονομαστικές μεταβλητές, οι οποίες επιδέχονται μόνο αυθαίρετη κατάταξη όπως π.χ. φύλο, θρησκεία, κ.τ.λ.
- Διατάξιμες μεταβλητές, οι οποίες επιδέχονται μέτρηση ανωτέρου επιπέδου που επιτρέπει την ιεράρχησή τους, όπως π.χ. χαρακτηρισμός πτυχίου (άριστα, λίαν καλώς, καλώς), σοβαρότητας μιας ασθένειας (ήπια, μέτρια, σοβαρή), της γνώμης για κάποιο μέτρο (διαφωνώ πλήρως, διαφωνώ σε κάποια σημεία, συμφωνώ, συμφωνώ πλήρως).

Ποσοτικές είναι οι μεταβλητές των οποίων οι τιμές είναι αριθμητικές και επιδέχονται μέτρηση. Για παράδειγμα το εισόδημα, το βάρος, το ύψος, ο αριθμός των παιδιών μιας οικογένειας.

Οι ποσοτικές μεταβλητές διακρίνονται σε δύο κατηγορίες, τις:

- Διακριτές, οι οποίες παίρνουν μόνο «μεμονωμένες» αριθμητικές τιμές όπως για παράδειγμα, το νούμερο των ανδρικών υποδημάτων, ο αριθμός των παιδιών μιας οικογένειας, ο αριθμός των ελαττωμάτων ενός προϊόντος.
- Συνεχείς οι οποίες μπορούν να πάρουν οποιαδήποτε τιμή μέσα από ένα συνεχές διάστημα, όπως για παράδειγμα το βάρος, το ύψος, η διάρκεια μιας τηλεφωνικής συνδιάλεξης.

Παρατήρηση: Ο διαχωρισμός μεταξύ διακριτών και συνεχών μεταβλητών είναι δύσκολος στην πράξη λόγω των περιορισμών που επιβάλλουν τα όργανα μέτρησης. Έτσι για π.χ. το βάρος ενός συνόλου ατόμων επειδή μετριέται με ακρίβεια γραμμαρίου μας δίνει διακριτές τιμές, όπως 63, 512 Kg 67, 383 Kg κ.τ.λ.

Οι ποσοτικές μεταβλητές διακρίνονται επιπλέον και σε δύο άλλες κατηγορίες, τις:

- Μεταβλητές διαστήματος, όπως π.χ. βαθμός πτυχίου (5, 10).
- Αναλογικές μεταβλητές, όπως για π.χ. χιλιομετρική απόσταση.

Εφαρμογή 1

Τα αποτελέσματα των εξετάσεων των φοιτητών του Μαθηματικού τμήματος στο μάθημα της Στατιστικής ήταν τα ακόλουθα: 2, 3, 3, 4, 4, 5, 7, 5, 5, 9. Να βρεθεί:

- i) Ποιος είναι ο πληθυσμός;
- ii) Ποια είναι τα άτομά;
- iii) Ποιες είναι οι παρατηρήσεις;
- iv) Ποια είναι η μεταβλητή και σε ποια κατηγορία ανήκει;
- v) Ποιες είναι οι τιμές των μεταβλητών;

ΛΥΣΗ

- i) Ο πληθυσμός είναι οι 10 φοιτητές του Μαθηματικού τμήματος.
- ii) Κάθε φοιτητής είναι ένα άτομο.
- iii) Οι παρατηρήσεις είναι: 2, 3, 3, 4, 4, 5, 7, 5, 5, 9.
- iv) Η μεταβλητή είναι «ο βαθμός στη Στατιστική» η οποία είναι ποσοτική, διακριτή και μεταβλητή διαστήματος.
- v) Οι τιμές της μεταβλητής είναι 2, 3, 4, 5, 7, 9.

Εφαρμογή 2

Εξετάζουμε τους κατοίκους μιας πόλης ως προς τα παρακάτω χαρακτηριστικά:

- i) Φύλο
- ii) ύψος
- iii) μορφωτικό επίπεδο
- iv) εισόδημα
- v) θρήσκευμα

Να χαρακτηρίσετε τις παραπάνω μεταβλητές.

ΛΥΣΗ

- i) Το «Φύλο» είναι ποιοτική ονομαστική μεταβλητή.
- ii) Το «ύψος» είναι ποσοτική συνεχής μεταβλητή διαστήματος.
- iii) Το «μορφωτικό επίπεδο» είναι ποιοτικά διατάξιμη μεταβλητή.
- iv) Το «εισόδημα» είναι ποσοτική συνεχής αναλογική μεταβλητή.
- v) Το «θρήσκευμα» είναι ποιοτική ονομαστική μεταβλητή.

Δείγμα (sample)

Είναι ένα υποσύνολο ενός πληθυσμού ή παρατηρηθέντων αποτελεσμάτων μιας διαδικασίας για μια χρονική περίοδο. Τα στοιχεία ενός δείγματος ονομάζονται δειγματικές ή δειγματοληπτικές μονάδες (sampling units). Δειγματοληπτικό σφάλμα λέγεται η διαφορά ανάμεσα στα αποτελέσματα μιας δειγματοληψίας και μιας απογραφής.

Παράμετρος (parameter)

Είναι μια αριθμητική ποσότητα που συνοψίζει κάποιο χαρακτηριστικό του πληθυσμού ή της ικανότητας μιας διαδικασίας.

Στατιστική συνάρτηση (statistic)

Είναι μια συνάρτηση των στοιχείων του δείγματος. Και στις δύο παραπάνω περιπτώσεις συνοψίζουμε τις πληροφορίες για ένα χαρακτηριστικό σε μια αριθμητική ποσότητα. Για

παράδειγμα ο μέσος του πληθυσμού είναι μια παράμετρος και ο δειγματικός μέσος είναι μια στατιστική συνάρτηση.

Στατιστική Συμπερασματολογία (Statistical Inference)

Είναι η διαδικασία χρησιμοποίησης πληροφοριών από το δείγμα με σκοπό την εξαγωγή συμπερασμάτων για τον πληθυσμό ή την ικανότητα μιας διαδικασίας.

Η εκτιμητική όπου χρησιμοποιούμε μια στατιστική συνάρτηση για να εκτιμήσουμε την τιμή μιας παραμέτρου του υπό μελέτη πληθυσμού.

Εμπιστοσύνη (confidence)

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

Παράδειγμα

Μια εταιρία που παράγει μύρα ενδιαφέρεται να διερευνήσει αν οι καταναλωτές προτιμούν μια καινούρια γεύση μύρας από αυτή που ήδη κυκλοφορεί στην αγορά. Προκειμένου να εξεταστεί αυτό δίνεται σε ένα τυχαίο δείγμα 125 καταναλωτών ένα δείγμα και από τις δύο γεύσεις. Οι καταναλωτές ερωτώνται ποια γεύση προτιμούν. 70 από τα 125 άτομα που ρωτήθηκαν δήλωσαν ότι προτιμούν τη νέα γεύση, ενώ 55 δήλωσαν ότι προτιμούν αυτή που ήδη κυκλοφορεί. Να καθορίσετε:

- A) τον πληθυσμό
- B) τη στατιστική μεταβλητή
- Γ) τις παραμέτρους που μας ενδιαφέρουν
- Δ) το δείγμα
- E) τη σχετική στατιστική συνάρτηση.

Λύση:

A) Ο πληθυσμός που μας ενδιαφέρει είναι οι προτιμήσεις σε γεύση των καταναλωτών της συγκεκριμένης περιοχής από την οποία επελέγη το δείγμα κατά τη συγκεκριμένη περίοδο που έγινε η μελέτη.

B) Η στατιστική μεταβλητή είναι η προτίμηση καθενός από τους καταναλωτές όπως αυτή εκφράστηκε στο δείγμα.

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

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

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

1.5 Μέθοδοι Συλλογής Στοιχείων

Οι διάφορες μέθοδοι συλλογής στατιστικών στοιχείων συνεχίζονται σε δύο μεγάλες ομάδες: τις απογραφές και τις δειγματοληπτικές έρευνες. Απογραφή ονομάζεται η διαδικασία με την οποία συλλέγονται οι παρατηρήσεις όλων των μονάδων ενός πληθυσμού σε μια συγκεκριμένη χρονική στιγμή. Μπορεί να έχουμε τις:

- Δημογραφικές απογραφές στις οποίες συλλέγονται στοιχεία σχετικά με το φύλο, την ηλικία, το επάγγελμα, κ.τ.λ.
- Οικονομικές απογραφές στις οποίες συγκεντρώνονται στοιχεία σχετικά με την οικονομική κατάσταση.
- Βιομηχανικές απογραφές, στις οποίες συλλέγουμε πληροφορίες σχετικά με την οικονομική δραστηριότητα των βιομηχανιών, το αριθμό των απασχολούμενων, το επίπεδο μηχανοργάνωσης, κ.τ.λ.

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

Στην Ελλάδα η Ε.Σ.Υ.Ε. πραγματοποιεί απογραφή του πληθυσμού κάθε 10 χρόνια (τελευταία τον Μάρτιο του 2001).

Τα μειονεκτήματα των απογραφών είναι:

- Το μεγάλο κόστος, καθώς χρειάζεται ειδική προεργασία και μεγάλο αριθμό απογραφέων.
- Τη μη ύπαρξη πολλών εξειδικευμένων ατόμων, που έχει ως συνέπεια την συγκέντρωση εσφαλμένων στοιχείων τα οποία μπορεί να δώσουν λανθασμένη εικόνα της διόρθωσης του πληθυσμού.
- Τη μη επίκαιρη έκδοση των αποτελεσμάτων, λόγω του μεγάλου όγκου των πληροφοριών.

Δειγματοληψία είναι απογραφή ορισμένων συγκεκριμένων χαρακτηριστικών ενός τμήματος του πληθυσμού. Το τμήμα του πληθυσμού που απογράφεται ονομάζεται δείγμα. Η δειγματοληψία πραγματοποιείται σε συγκεκριμένα στάδια ανάλογα με τη μέθοδο που επιλέγουμε.

Το περιοδικό Literary Digest χρησιμοποιώντας δείγμα 2.400.000 ατόμων πρόβλεψε νίκη του Landon με ποσοστό 57%. Αντίθετα, το δημοκοπικό γραφείο του G. Gallup χρησιμοποιώντας δείγμα 50.000 ατόμων πρόβλεψε το σωστό αποτέλεσμα που ήταν νίκη των Roosvelt με ποσοστό 62%. Οι αρχές και οι μέθοδοι επιλογής του δείγματος είναι αντικείμενο της Δειγματοληψίας.

Πιο αναλυτικά για τους τρόπους συλλογής στοιχείων μπορούμε να αναφέρουμε ότι ο βασικός στόχος είναι η συλλογή στοιχείων που χαρακτηρίζουν τον πληθυσμό όσο το δυνατό καλύτερα. Οι κύριες μέθοδοι συλλογής στοιχείων είναι οι εξής:

1. Επιλογή τυχαίου δείγματος (random sample)
2. Διεξαγωγή τυχαιοποιημένου πειράματος (randomized experiment)
3. Χρησιμοποίηση διαθέσιμων στοιχείων (convenience data)
4. Επιλογή λογικών υποομάδων (rational subgroups), δηλαδή σχεδιασμένα δείγματα που επιλέγονται στην πορεία του χρόνου.

Τυχαία Δείγματα

Η πιο αποτελεσματική μέθοδος επιλογής δείγματος για τον έλεγχο του δειγματοληπτικού λάθους. Η πιο απλή εξειδίκευση είναι η Απλή Τυχαία Δειγματοληψία όπου κάθε υποομάδα n στοιχείων του πληθυσμού έχει την ίδια πιθανότητα να επιλεγεί.

Τα πλεονεκτήματα της τυχαίας δειγματοληψίας σε σχέση με μη τυχαίες μεθόδους επιλογής είναι τα εξής:

1. Αποκλεισμός της μεροληψίας.
2. Καθορισμός εμπιστοσύνης.
3. Έλεγχος δειγματικού λάθους. Συνήθως με αντίστοιχο έλεγχο του μεγέθους του δείγματος.

Τυχαιοποιημένα πειράματα

Μελετάμε τις αιτίες μεταβλητότητας σχεδιάζοντας ένα πείραμα. Για παράδειγμα 50 σπόροι επιλέγονται τυχαία από ένα σύνολο 100 σπόρων και οι 25άδες τοποθετούνται σε δύο διαφορετικά μέρη ενός αγρού για να δούμε τη επίδραση στη σοδειά ενός λιπάσματος.

Υπάρχει τυχαιοποιημένη κατανομή των σπόρων, ωστόσο οι 100 σπόροι δεν είναι ο πληθυσμός μας. Αντίθετα ο πληθυσμός μας είναι η παραγωγή όλου του σταριού που θα

παίρναμε αν χρησιμοποιούσαμε τις συνθήκες μελέτης. Ο προσεκτικός σχεδιασμός των τυχαιοποιημένων πειραμάτων επιτρέπει να πάρουμε τη μέγιστη δυνατή πληροφορία για το φαινόμενο που μελετάμε με ελάχιστο κόστος.

Διαθέσιμα δεδομένα

Πολλές φορές δεν μπορούμε να χρησιμοποιήσουμε τυχαία δειγματοληψία αλλά στηριζόμαστε σε δεδομένα που απλώς είναι διαθέσιμα στον αναλυτή όπως συμβαίνει στην ιατρική ή στη γεωλογία. Τα μειονεκτήματά τους είναι τα εξής:

1. Δεν παρέχουν τις δυνατότητες που παρέχουν τα δεδομένα τα οποία προέρχονται από τυχαία δειγματοληψία.
2. Δεν μας δίνουν τη δυνατότητα να σχεδιάσουμε τη διαδικασία επιλογής ώστε να επωφεληθούμε από τα πλεονεκτήματα που προσφέρουν τα τυχαιοποιημένα πειράματα.

Λογικές Υποομάδες

Η μέθοδος αυτή χρησιμοποιείται όταν μας ενδιαφέρει να μελετήσουμε τη μεταβλητότητα που παρουσιάζεται στα αποτελέσματα της διαδικασίας με την πάροδο του χρόνου. Για το λόγο αυτό επιλέγονται μικρά δείγματα σε τακτά χρονικά διαστήματα με τα στοιχεία κάθε δείγματος να λαμβάνονται όσο το δυνατό κάτω από τις ίδιες συνθήκες και στον ίδιο περίπου χρόνο. Η μεταβλητότητα μεταξύ των δειγμάτων μας επιτρέπει να επισημάνουμε μεταβολές στη διαδικασία.

Είδη Αιτιών Διακύμανσης

A. Κοινές αιτίες (common causes): Είναι συνηθισμένοι παράγοντες που μεταβάλλονται φυσιολογικά με την πάροδο του χρόνου. Για παράδειγμα στο πρόβλημα χρόνου που απαιτείται για να πάει κάποιος στη δουλειά του τέτοιες είναι η διαδρομή που ακολουθεί, η ταχύτητα με την οποία οδηγεί, η ένταση κυκλοφορίας κτλ.

B. Ειδικές αιτίες (special causes): Είναι ασυνήθιστα περιστατικά. Στο πρόβλημα χρόνου μετάβασης στην εργασία τέτοια μπορεί να είναι ένα προβληματικό λάστιχο.

1.6 Στατιστικοί πίνακες

Μετά τη συλλογή των στατιστικών δεδομένων είναι αναγκαία η κατασκευή συνοπτικών πινάκων ή γραφικών παραστάσεων, ώστε να είναι εύκολη η κατανόησή τους και η εξαγωγή σωστών συμπερασμάτων. Η παρουσίαση των στατιστικών δεδομένων σε πίνακες γίνεται με την κατάλληλη τοποθέτηση των πληροφοριών σε γραμμές και στήλες, με τρόπο που να διευκολύνεται η σύγκριση των στοιχείων και η καλύτερη ενημέρωση του αναγνώστη σχετικά με τη δομή του πληθυσμού που ερευνάμε.

Οι πίνακες διακρίνονται στους:

- α) **Γενικούς πίνακες**, οι οποίοι περιέχουν όλες τις πληροφορίες που προκύπτουν από μία στατιστική έρευνα (συνήθως με αρκετά λεπτομερειακά στοιχεία) και αποτελούν πηγές στατιστικών πληροφοριών στη διάθεση των επιστημόνων-ερευνητών για παραπέρα ανάλυση και εξαγωγή συμπερασμάτων.
- β) **Ειδικούς πίνακες**, οι οποίοι είναι συνοπτικοί και σαφείς. Τα στοιχεία τους συνήθως έχουν ληφθεί από τους γενικούς πίνακες.

Κάθε πίνακας που έχει κατασκευαστεί σωστά πρέπει να περιέχει:

- α) τον τίτλο, που γράφεται στο επάνω μέρος του πίνακα και δηλώνει με σαφήνεια και συνοπτικά το περιεχόμενο του πίνακα,
- β) τις επικεφαλίδες των γραμμών και στηλών, που δείχνουν συνοπτικά τη φύση και τις μονάδες μέτρηστων δεδομένων,
- γ) το κύριο σώμα (κορμό), που περιέχει διαχωρισμένα μέσα στις γραμμές και στις στήλες τα στατιστικά δεδομένα,
- δ) την πηγή, που γράφεται στο κάτω μέρος του πίνακα και δείχνει την προέλευση των στατιστικών στοιχείων, έτσι ώστε ο αναγνώστης να ανατρέχει σ' αυτήν, όταν επιθυμεί, για επαλήθευση στοιχείων ή για λήψη περισσότερων πληροφοριών.

Ατυχήματα	2011	2012	Διαφορά	Ποσοστό
Θανατηφόρα	1.011	1.162	-151	-13,0%
Σοβαρά	1.397	1.454	-57	-3,9%
Ελαφρά	11.223	12.456	-1.233	-9,9%
Σύνολο	13.631	15.072	-1.441	-9,6%

Πίνακας 1: Ποσοστό στατιστικών ατυχημάτων 2011 - 2012

Πίνακες Κατανομής Συχνοτήτων

Ας υποθέσουμε ότι x_1, x_2, \dots, x_k είναι οι τιμές μιας μεταβλητής X , που αφορά τα στοιχεία ενός δείγματος μεγέθους $n, k \leq n$. Στην τιμή x_i αντιστοιχίζεται η συχνότητα v_i , δηλαδή ο φυσικός αριθμός που δείχνει πόσες φορές εμφανίζεται η τιμή x_i της εξεταζόμενης μεταβλητής X στο σύνολο των παρατηρήσεων. Για τις συχνότητες ισχύουν:

- $v_1 + v_2 + \dots + v_k = n$
- $0 \leq v_i \leq n, i=1,2,\dots, k$

Εάν διαιρέσουμε τη συχνότητα v_i με το μέγεθος n του δείγματος προκύπτει η σχετική συχνότητα f_i της τιμής x_i , δηλαδή:

$$f_i = \frac{v_i}{n}, i=1,2,\dots, k.$$

Για τις σχετικές συχνότητες ισχύουν:

- $0 \leq f_i \leq 1, i=1,2,\dots, k$

- $f_1 + f_2 + \dots + f_k = 1$

Συνήθως, τις σχετικές συχνότητες f_i τις εκφράζουμε επί τοις εκατό, οπότε συμβολίζονται με $f_i \%$.

Οι ποσότητες $x_i, v_i, f_i, i=1,2,\dots,k$ για ένα δείγμα συγκεντρώνονται σε ένα συνοπτικό πίνακα, που ονομάζεται πίνακας συχνοτήτων.

Αριθμός Αδερφών x_i	Συχνότητα v_i	Σχετική συχνότητα f_i
0	8	0,200
1	22	0,550
2	7	0,175
3	3	0,075
Σύνολο:	40	1

Πίνακας 2: Πίνακας συχνοτήτων αδερφών μίας τάξης

Στην περίπτωση των ποσοτικών μεταβλητών εκτός από τις συχνότητες v_i και f_i χρησιμοποιούνται συνήθως και οι λεγόμενες αθροιστικές συχνότητες N_i και οι αθροιστικές σχετικές συχνότητες F_i , οι οποίες εκφράζουν το πλήθος και το ποσοστό αντίστοιχα των παρατηρήσεων που είναι μικρότερες ή ίσες της τιμής x_i . Συχνά οι F_i πολλαπλασιάζονται επί 100 εκφραζόμενες έτσι επί τοις εκατό, δηλαδή $F_i \% = 100 F_i$.

x_i	v_i	f_i	$f_i \%$	N_i	F_i	$F_i \%$
0	8	0,200	20,0	8	0,200	20,0
1	22	0,550	55,0	30	0,750	75,0
2	7	0,175	17,5	37	0,925	92,5
3	3	0,075	7,5	40	1,000	100,0
Σύνολο	40	1	100			

Πίνακας 3: Πίνακας συχνοτήτων και αθροιστικών συχνοτήτων αδερφών μιας τάξης

Ομαδοποίηση

Στην περίπτωση διακριτής μεταβλητής, όταν το πλήθος των τιμών της είναι μεγάλο, αλλά πολύ περισσότερο σε συνεχή μεταβλητή, ταξινομούμε τα δεδομένα σε ίσα διαδοχικά διαστήματα της μορφής $[,)$ (κλάσεις ή τάξεις) και καταγράφουμε τις συχνότητες των παρατηρήσεων που ανήκουν σε κάθε κλάση. Για την i κλάση $[a, b)$ η κεντρική τιμή m_i είναι:

$$m_i = \frac{a + b}{2}.$$

Οι κλάσεις που επιλέγουμε πρέπει να είναι τέτοιες ώστε να δίνουν τη σωστή εικόνα της κατανομής των δεδομένων. Ο καθορισμός του αριθμού των κλάσεων είναι εν γένει αυθαίρετος, ωστόσο ως επί το πλείστον γίνεται χρήση του κανόνα του Sturges. Σύμφωνα με αυτό τον κανόνα $k = 1 + 3,322 \log_{10} N$ όπου k ο αριθμός των κλάσεων και N ο αριθμός των δεδομένων. Συνήθως χρησιμοποιούνται από 5 ως 20 κλάσεις. Όσο περισσότερα δεδομένα έχουμε τόσο περισσότερες κλάσεις πρέπει να χρησιμοποιούνται. Αυτό γιατί αν ο αριθμός των κλάσεων που θα χρησιμοποιηθούν είναι πολύ μικρός είναι ενδεχόμενο να αποκρύβουν σημαντικά χαρακτηριστικά των δεδομένων με την ομαδοποίησή τους. Απ' την άλλη πλευρά, αν ο αριθμός των κλάσεων είναι μεγάλος σε σχέση με τα δεδομένα θα έχουμε πολλές κλάσεις που θα είναι κενές ή με μικρό αριθμό παρατηρήσεων και η κατανομή που θα εμφανίζον θα δίνει μια όχι ικανοποιητική περιγραφή των δεδομένων.

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

Κατανομή Δεδομένων

Για την καλύτερη παρουσίαση των δεδομένων απαιτείται η σύνοψή τους. Ο πρώτος τρόπος είναι η παρουσίασή τους με κάποια γραφική μέθοδο. Ο τρόπος με τον οποίο ταξινομούνται τα δεδομένα ονομάζεται κατανομή (distribution) των δεδομένων αυτών. Η κατανομή αποκαλύπτει τρόπους με τους οποίους τα δεδομένα μεταβάλλονται.

Πολύγωνα Συχνότητας

Το πολύγωνο συχνότητας (frequency polygon) είναι ένας άλλος τρόπος παρουσίασης κατανομής συχνότητας. Κατασκευάζεται ως εξής: Σημειώνουμε τη συχνότητα κάθε κλάσης πάνω από το μέσο σημείο της κλάσης και στη συνέχεια ενώνουμε τα διαδοχικά σημεία που προκύπτουν με ευθύγραμμα τμήματα.

Το πολύγωνο 'κλείνει' με τη θεώρηση μιας πρόσθετης θεωρητικής κλάσης (με συχνότητα 0) σε κάθε ένα από τα άκρα του διαστήματος των τιμών της κατανομής και την προσθήκη, στη συνέχεια στην τεθλασμένη γραμμή δύο ευθυγράμμων τμημάτων που συνδέουν τα άκρα της τεθλασμένης με τα ενδιάμεσα σημεία των κλάσεων αυτών.

Τα πολύγωνα συχνότητας δίνουν μια γενική ιδέα για τη μορφή της κατανομής. Επίσης μπορούμε να κατασκευάσουμε πολύγωνο σχετικής συχνότητας. Τοποθετώντας ένα πολύγωνο σχετικής συχνότητας πάνω σε ένα άλλο μπορούμε να έχουμε μια οπτική σύγκριση δύο κατανομών.

Ακολούθως μπορούμε να κατασκευάσουμε το πολύγωνο αθροιστικής σχετικής συχνότητας (cumulative relative frequency polygon) το οποίο συνήθως ονομάζεται ακιδωτό (ogive). Η αθροιστική σχετική συχνότητα μιας κλάσης είναι το ποσοστό των μετρήσεων που είναι μικρότερες από το πάνω άκρο της κλάσης αυτής.

1.7 Στατιστικά Διαγράμματα

Οι στατιστικοί πίνακες, παρά την πληρότητα την οποία παρουσιάζουν και την ακρίβεια των πληροφοριών που περιέχουν, είναι σχεδόν πάντοτε χρήσιμοι οι πληροφορίες που περιέχουν να παρασταθούν με μορφή διαγραμμάτων ή γραφικών παραστάσεων.

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

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

Τέλος, πρέπει να αναφέρεται και η πηγή από την οποία πήραμε τα αριθμητικά δεδομένα των μεγεθών. Τα στατιστικά δεδομένα που συγκεντρώνουν σ' ένα πίνακα συχνοτήτων μπορούν να παρουσιαστούν με τη μορφή γραφικών παραστάσεων ή διαγραμμάτων.

Από τις γραφικές παραστάσεις ή τα διαγράμματα μπορούμε ν' αναλύσουμε συνοπτικά διάφορες χρήσιμες πληροφορίες. Με τα διαγράμματα διευκολύνεται η σύγκριση μεταξύ ομοειδών στοιχείων για τα ίδια ή διαφορετικά χαρακτηριστικά.

Τα κυριότερα είδη γραφικών παραστάσεων είναι:

- Ραβδόγραμμα
- Διάγραμμα συχνοτήτων
- Κυκλικό διάγραμμα
- Χρονόγραμμα
- Διαγράμματα Σημείων
- Ιστόγραμμα
- Διαγράμματα Μίσχου-Φύλλου

Ραβδόγραμμα

Το ραβδόγραμμα χρησιμοποιείται συνήθως όταν η μεταβλητή X είναι ποιοτική. Αποτελείται από ορθογώνιες στήλες που οι βάσεις τους έχουν κοινό μήκος που επιλέγεται αυθαίρετα είναι όμως τέτοιο ώστε να εξασφαλίζονται κενά μεταξύ δύο διαδοχικών ορθογωνίων. Οι στήλες υψώνονται πάνω σε ορθογώνιο ή κατακόρυφο άξονα. Το ύψος κάθε στήλης είναι ίσο με τη συχνότητα v_i ή τη σχετική συχνότητα f_i της αντίστοιχης τιμής της μεταβλητής.

Όταν θέλουμε να κάνουμε σύγκριση των αντίστοιχων τιμών της ίδιας μεταβλητής X για δύο διαφορετικά δείγματα, τότε κατασκευάζουμε δίπλα ορθογώνια για την ίδια τιμή της μεταβλητής X , ένα για το κάθε δείγμα.

Διάγραμμα συχνοτήτων

Στην περίπτωση ποσοτικών μεταβλητών αυτή του ραβδογράμματος χρησιμοποιείται το διάγραμμα συχνοτήτων και σχετικών συχνοτήτων. Αποτελείται από ευθύγραμμα τμήματα κάθετα στον άξονα της μεταβλητής X , ένα για κάθε τιμή x_i . Οι τιμές της μεταβλητής τοποθετούνται στο διάγραμμα σε αύξουσα σειρά. Το ύψος κάθε ευθύγραμμου τμήματος είναι ίσο με την αντίστοιχη συχνότητα v_i ή με την αντίστοιχη σχετική συχνότητα f_i .

Αν ενώσουμε τα σημεία (x_i, v_i) , $i=1,2,\dots,k$ με διαδοχικά ευθύγραμμα τμήματα δημιουργείται μια πολυγωνική γραμμή που ονομάζεται πολύγωνο συχνοτήτων. Ομοίως αν

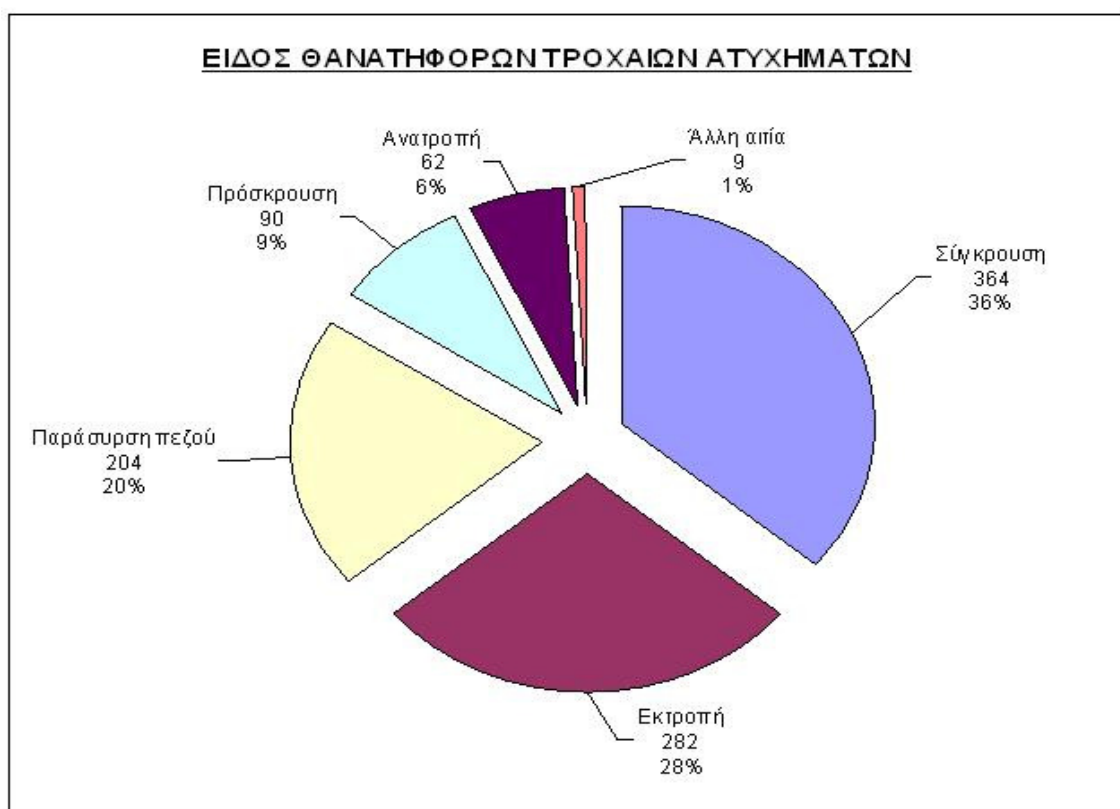
ενώσουμε τα σημεία (x_i, f_i) δημιουργείται το πολύγωνο σχετικών συχνοτήτων. Με τον ίδιο τρόπο κατασκευάζονται και τα διαγράμματα των αθροιστικών συχνοτήτων.

Κυκλικό διάγραμμα

Το κυκλικό διάγραμμα χρησιμοποιείται για την γραφική παράσταση ποιοτικών ή ποσοτικών μεταβλητών όταν οι τιμές της μεταβλητής X είναι σχετικά λίγες. Το κυκλικό διάγραμμα είναι ένας κυκλικός δίσκος χωρισμένος σε k κυκλικούς τομείς (όσες και οι τιμές της μεταβλητής X). Το εμβαδόν E_i κάθε κυκλικού τομέα είναι ανάλογο προς τις αντίστοιχες συχνότητες v_i ή τις σχετικές συχνότητες f_i .

Τα τόξα a_i που αντιστοιχούν σε κάθε κυκλικό τομέα δίνονται από τη

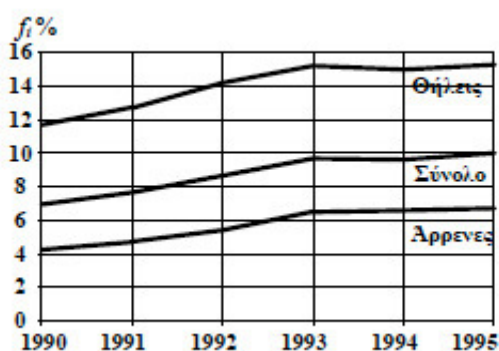
$$\text{σχέση: } \frac{a_i}{360^\circ} = \frac{v_i}{v} \Leftrightarrow a_i = \frac{v_i}{v} \cdot 360^\circ = f_i \cdot 360^\circ, \quad i=1,2,\dots,k.$$



Εικόνα 1: Κυκλικό διάγραμμα θανατηφόρων τροχαίων ατυχημάτων

Χρονόγραμμα

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



Εικόνα 2: Χρονόγραμμα

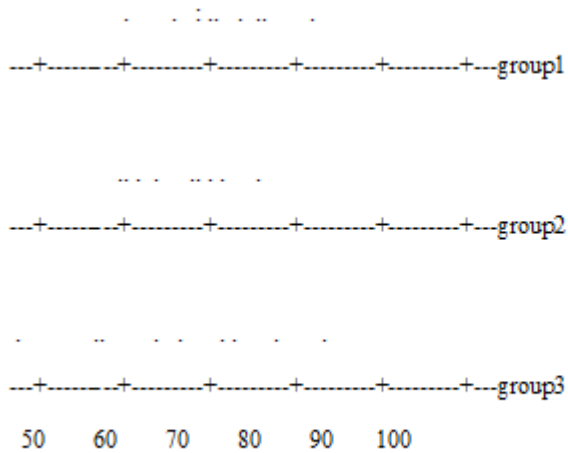
Διαγράμματα Σημείων

Όταν έχουμε ένα μικρό αριθμό παρατηρήσεων η κατανομή τους περιγράφεται εύκολα με ένα διάγραμμα σημείων (Dot Diagram). Όταν οι τιμές συμπίπτουν τοποθετούνται η μία πάνω στην άλλη. Για παράδειγμα έχουμε τους βαθμούς σε κάποιο μάθημα 66, 74, 78, 78, 80, 81, 85, 88, 89, 97. Για την κατασκευή του διαγράμματος σημείων χρειαζόμαστε ένα ευθύγραμμο τμήμα που να απεικονίζει τις τιμές από το 66 ως το 97 και σημειώνουμε τους βαθμούς πάνω σε αυτό χρησιμοποιώντας τελείες.

Έστω τώρα ότι έχουμε τρεις διαφορετικές ομάδες μαθητών όπου το μάθημα διδάσκεται με διαφορετική μέθοδο. Τα αποτελέσματα είναι τα εξής (η πρώτη ομάδα είναι η προηγούμενη):

Ομάδα 2: 65, 66, 68, 71, 77, 78, 80, 82, 88.

Ομάδα 3: 48, 61, 62, 71, 75, 82, 84, 91, 99.



Εικόνα 3: Διάγραμμα σημείων

Συγκρίνοντας τα διαγράμματα για τις τρεις ομάδες μαθητών μπορούμε να βγάλουμε τα εξής συμπεράσματα. Ο μέσος βαθμός των ομάδων 2 και 3 είναι περίπου ίδιος ενώ ο μέσος βαθμός της ομάδας 1 είναι υψηλότερος.

Επίσης η μεταβλητότητα των βαθμών της ομάδας 3 είναι η μεγαλύτερη και της ομάδας 2 η μικρότερη. Τέλος, αν τα δεδομένα αποτελούνται από πολλές τιμές τα διαγράμματα σημείων γίνονται πολύ πυκνά και όχι αποτελεσματικά. Σε αυτή την περίπτωση η καλύτερη λύση είναι να τοποθετήσουμε ομάδες με όμοιες τιμές δεδομένων. Ο κυριότερος τρόπος παρουσίασης αυτής της μορφής είναι το διάγραμμα μίσχου-φύλλου (stem and leaf plot).

Ιστόγραμμα

Η γραφική παράσταση ενός πίνακα συχνοτήτων με ομαδοποιημένα δεδομένα ονομάζεται ιστόγραμμα (histogram). Στον οριζόντιο άξονα ενός συστήματος ορθογωνίων αξόνων σημειώνουμε, με κατάλληλη κλίμακα, τα όρια των κλάσεων (θεωρούμε ως μονάδα μέτρησης το πλάτος κάθε κλάσης). Στη συνέχεια, κατασκευάζουμε διαδοχικά ορθογώνια (ιστούς), καθένα από τα οποία έχει βάση ίση με το πλάτος της κλάσης και ύψος ίσο με την αντίστοιχη συχνότητα (σχετική συχνότητα, αθροιστική συχνότητα, σχετική αθροιστική συχνότητα) της κλάσης αυτής.

Διαγράμματα Μίσχου-Φύλλου (Stem and Leaf Plots)

Το διάγραμμα μίσχου-φύλλου αποτελεί ένα χρήσιμο πρώτο βήμα στην κατασκευή μιας κατανομής συχνότητας και ενός ιστογράμματος. Έστω έχουμε τους βαθμούς 20 σπουδαστών σε ένα μάθημα στατιστικής και μας ενδιαφέρει να δούμε πως απέδωσαν 93 83 86 83 56 63 64 73 78 81 62 88 54 72 74 87 78 61 63 89. Το διάγραμμα μίσχου φύλλου κατασκευάζεται ως εξής: Χωρίζουμε κάθε αριθμό από τα δεδομένα σε δύο μέρη το μίσχο και το φύλλο. Ακολουθώς κάνουμε μια κατακόρυφη γραμμή και αριστερά της τοποθετούμε το μίσχο και δεξιά της το φύλλο οπότε έχουμε το παρακάτω αποτέλεσμα.

	M	Φ
2	5	46
7	6	12334
(5)	7	23488
8	8	1336789
1	9	3

Εικόνα 4: Διάγραμμα μίσχου φύλλου

Γενικά όλοι οι αριθμοί με την ίδια τιμή μίσχου τοποθετούνται στην ίδια γραμμή του διαγράμματος μίσχου-φύλλου με τιμές φύλλων τοποθετημένες στα δεξιά της κατακόρυφης γραμμής.

Η πρώτη στήλη στο παραπάνω διάγραμμα αναφέρεται στις αθροιστικές συχνότητες για κάθε μίσχο ξεκινώντας από τους ακραίους μίσχους με κατεύθυνση τον κεντρικό. Εξαιρέση αποτελεί ο κεντρικός μίσχος για τον οποίο δηλώνεται η απόλυτη συχνότητα σε παρένθεση. Ο κεντρικός μίσχος είναι αυτός που περιέχει τη διάμεσο του δείγματος.

1. Ένδειξη κατανομής: Οι βαθμοί κατανέμονται ομοιόμορφα μεταξύ των χαμηλών τιμών του 60 και των υψηλών τιμών του 80. Μόνο τρεις βαθμοί βρίσκονται έξω από την κλίμακα αυτή: ένας στα 90 και δύο στα 50.

2. Οργάνωση των δεδομένων: Δεν χάνονται λεπτομέρειες από τα δεδομένα. Οποιαδήποτε στατιστική συνάρτηση μπορεί να υπολογιστεί από ένα διάγραμμα μίσχου-φύλλου δοθέντος ότι όλα τα δεδομένα δίνονται στο διάγραμμα.

Για στοιχεία που εκφράζονται με μεγαλύτερους αριθμούς είναι απαραίτητο να διαιρεθούν οι αριθμοί αυτοί με διαφορετικούς τρόπους. Για παράδειγμα ο αριθμός 1268 μπορεί να έχει μίσχο το 12 και φύλλο το 68.

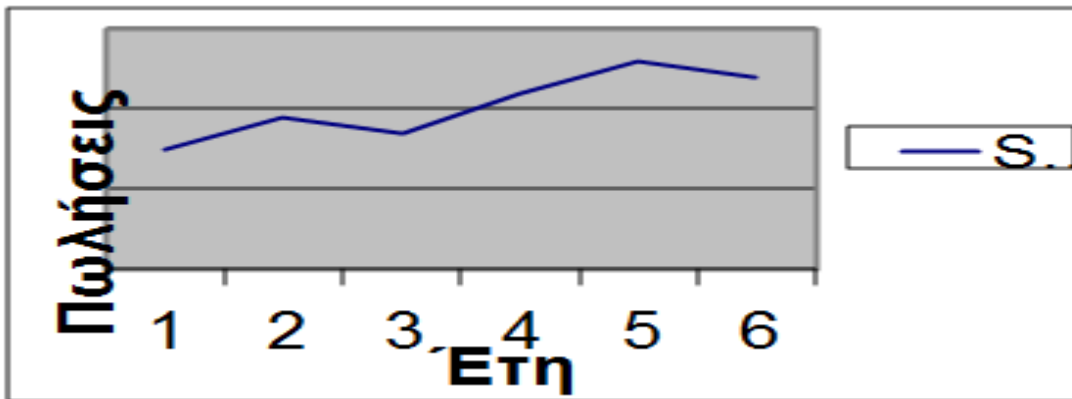
Η επιλογή γίνεται έτσι ώστε τα δεδομένα να παρουσιάζονται με τον καλύτερο δυνατό τρόπο. Τέλος πρέπει να αποφεύγουμε να έχουμε πολλά φύλλα σε σχέση με τον αριθμό των μίσχων. Αν στο παράδειγμα οι βαθμοί ήταν όλοι μεταξύ 70 και 90 η χρησιμοποίηση των ψηφίων των δεκάδων για τον καθορισμό των μίσχων θα μας έδινε μόνο δύο γραμμές. Σε αυτή την περίπτωση θα μπορούσαμε να χωρίσουμε τους βαθμούς με κοινό μίσχο σε δύο κατηγορίες: 'υψηλή' και 'χαμηλή'. Έτσι, οι μίσχοι για τους βαθμούς στη δεκάδα του 80 θα μπορούσαν να είναι 8^+ για βαθμούς από 85 ως 89 και 8^- για βαθμούς από 80 ως 84. Το ίδιο και για τη δεκάδα του 70 οπότε θα είχαμε τέσσερις γραμμές βαθμών αντί για δύο.

1.7.1 Παραποιήσεις με γραφικές παραστάσεις

Οι γραφικές παραστάσεις απλώς δημιουργούν μια οπτική εντύπωση, γεγονός που αφήνει περιθώρια για παραποιήσεις στην παρουσίαση αυτή. Ενδεικτικά αναφέρουμε:

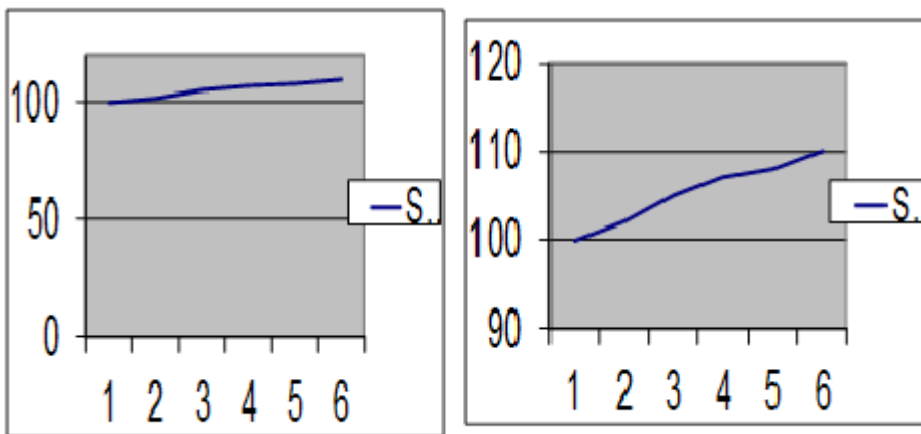
Παρουσίαση γραφημάτων χωρίς κλίμακα σε ένα από τους δύο άξονες.

Στο παρακάτω γράφημα των πωλήσεων (εικ. 5) μιας εταιρίας η αύξηση μπορεί να είναι 100% ή 1% για τα τελευταία 6 χρόνια ανάλογα με τη κλίμακα του κατακόρυφου άξονα. Τέτοια γραφήματα δεν πρέπει να λαμβάνονται υπόψη.



Εικόνα 5: Γράφημα πωλήσεων

Λανθασμένη εντύπωση μπορούμε να έχουμε όταν σε ένα γράφημα εμφανίζονται μόνο απόλυτες μεταβολές στις τιμές και όχι ποσοστιαίες. Η μεταβολή 3 ευρώ σε μια μετοχή 6 ευρώ είναι χειρότερη από τη μεταβολή σε μια μετοχή 20 ευρώ. Επίσης παραποίηση των δεδομένων επιτυγχάνεται με επιμήκυνση του κατακόρυφου ή του οριζόντιου άξονα. Δείτε τη διαφορά στα δύο παρακάτω διαγράμματα με τις ίδιες πωλήσεις.



Εικόνα 6: Διαφορά διαγραμμάτων με τις ίδιες πωλήσεις

1.8 Αξιολόγηση Στατιστικών Μελετών

Μια στατιστική μελέτη θεωρείται πλήρης όταν περιλαμβάνει ένα κομμάτι με όλες τις πιθανές πηγές λάθους. Τέτοιες είναι οι εξής:

- Αναντιστοιχία μεταξύ του πληθυσμού που μελετάται και του πληθυσμού στον οποίο θα εφαρμοσθούν οι όποιες αποφάσεις.
- Ακατάλληλη επεξεργασία των στοιχείων μιας διαδικασίας. Πρέπει πάντα να εξετάζουμε αν υπάρχουν συστηματικές αλλαγές στα δεδομένα με την πάροδο του χρόνου. Για παράδειγμα ο μέσος όρος αλλάζει αν υπάρχει μια τάση μείωσης ή αύξησης με την πάροδο του χρόνου.
- Αναντιστοιχία πλαισίου και πληθυσμού.

Συνοπτικά τα στάδια ενός στατιστικού προβλήματος είναι τα εξής:

1. Σαφής ορισμός του αντικειμένου του πειράματος και του πληθυσμού στον οποίο το πείραμα αναφέρεται.
2. Σχεδιασμός του πειράματος ή της δειγματοληπτικής διαδικασίας.
3. Συλλογή και ανάλυση δεδομένων.
4. Διαδικασία στατιστικής συμπερασματολογίας για τον υπό μελέτη πληθυσμό με βάση τις πληροφορίες του δείγματος.
5. Παροχή ενός μέτρου καταλληλότητας (αξιοπιστίας) της συμπερασματολογίας.

2 Αριθμητική Περιγραφή Δεδομένων

Οι γραφικές μέθοδοι αποτελούν μια χρήσιμη και γρήγορη παρουσίαση των δεδομένων ωστόσο δεν είναι εύκολο να χρησιμοποιηθούν για στατιστική συμπερασματολογία. Ως εκ τούτου για την αποφυγή αυτού του προβλήματος χρησιμοποιούμε αριθμητικά περιγραφικά μέτρα (numerical descriptive measures).

Εκείνο που μας ενδιαφέρει κυρίως είναι να χρησιμοποιήσουμε τα δειγματικά δεδομένα για να υπολογίσουμε ένα σύνολο τιμών στατιστικών συναρτήσεων που θα μας δώσουν μια καλή θεωρητική εικόνα της δειγματικής κατανομής σχετικής συχνότητας και τα οποία θα είναι

χρήσιμα για να βοηθηθούμε στη στατιστική συμπερασματολογία που αναφέρεται στην κατανομή της σχετικής συχνότητας του πληθυσμού.

2.1 Μέτρα Θέσης ή Κεντρικής Τάσης

Τα κυριότερα μέτρα κεντρικής τάσης είναι ο αριθμητικός μέσος (arithmetic mean) η διάμεσος (median) και η επικρατούσα τιμή (mode).

2.1.1 Αριθμητικός Μέσος

Αναφέρεται απλά ως μέσος. Έστω x_1, x_2, \dots, x_k οι τιμές της μεταβλητής X , που αφορά τα στοιχεία ενός δείγματος μεγέθους $v, k \leq v$. Ο αριθμητικός μέσος δίνεται από τον τύπο:

$$\bar{x} = \frac{v_1 \cdot x_1 + v_2 \cdot x_2 + \dots + v_k \cdot x_k}{v} = \frac{1}{v} \cdot \sum_{i=1}^k v_i \cdot x_i.$$

Στην περίπτωση που οι τιμές x_1, x_2, \dots, x_k της μεταβλητής X αφορούν τα στοιχεία ενός πληθυσμού με N στοιχεία, ο πληθυσμιακός μέσος δίνεται από τον τύπο:

$$\mu = \frac{v_1 \cdot x_1 + v_2 \cdot x_2 + \dots + v_k \cdot x_k}{N} = \frac{1}{N} \cdot \sum_{i=1}^k v_i \cdot x_i.$$

Η παραπάνω μορφή του μέσου ονομάζεται αστάθμητος μέσος αριθμητικός. Εκτός από αυτόν υπάρχει και ο σταθμικός μέσος αριθμητικός ο οποίος χρησιμοποιείται κυρίως στην κατασκευή αριθμοδεικτών. Σε κάθε παρατήρηση x_1, x_2, \dots, x_v αντιστοιχούμε ένα βάρος w_1, w_2, \dots, w_v και ο τύπος του είναι:

$$\bar{x}_w = \frac{w_1 x_1 + w_2 x_2 + \dots + w_k x_k}{w_1 + w_2 + \dots + w_k} = \frac{\sum_{i=1}^v w_i x_i}{\sum_{i=1}^v w_i}$$

Για παράδειγμα έστω οι παρατηρήσεις 5, 3 και 2 με αντίστοιχα βάρη 0,5 0,3 και 0,2 τότε

$$\bar{x}_w = \frac{5 \cdot 0,5 + 3 \cdot 0,3 + 2 \cdot 0,2}{0,5 + 0,3 + 0,2} = 3,8 \text{ ενώ ο αστάθμητος μέσος είναι } 3,333.$$

Τέλος αν έχουμε ομαδοποιημένα δεδομένα, δηλαδή σε κλάσεις τότε:

$$\bar{x} = \frac{v_1 m_1 + v_2 m_2 + \dots + v_k m_k}{v} = \frac{1}{v} \cdot \sum_{i=1}^k v_i \cdot m_i$$

όπου v_i η συχνότητα στην κλάση i ($i = 1, \dots, k$) και m_i το ενδιάμεσο σημείο κάθε κλάσης με την παραδοχή ότι αυτό προσεγγίζει ικανοποιητικά το μέσο των μετρήσεων που ανήκουν σε κάθε κλάση.

Παράδειγμα 1:

Έστω κάποιο πανεπιστήμιο ενδιαφέρεται να περιορίσει το κόστος των υπεραστικών τηλεφωνημάτων. Στην προσπάθεια αυτή το πανεπιστήμιο λαμβάνει ένα τυχαίο δείγμα 30 υπεραστικών τηλεφωνημάτων που έγιναν σε μια εβδομάδα. Οι χρόνοι διάρκειας των τηλεφωνημάτων είναι οι ακόλουθοι (σε λεπτά):

11,8 3,6 16,6 13,5 4,8 8,3 8,9 9,1 7,7 2,3 12,1 6,1 10,2 8,0 11,4 6,8 9,6 19,5 15,3 12,3 8,5
15,9 18,7 11,7 6,2 11,2 10,4 7,2 5,5 14,5

Κλάση	Όρια κλάσης	Συχνότητα v_i	Μέσο σημείο m_i	$v_i m_i$	$v_i m_i^2$
1	2-5	3	3,5	10,5	36,75
2	5-8	6	6,5	39	253,5
3	8-11	8	9,5	76	722
4	11-14	7	12,5	87,5	1093,75
5	14-17	4	15,5	62	961
6	17-20	2	18,5	37	684,5
Άθροισμα		$v = 30$		312	3751,5

Πίνακας 4: Πίνακας συχνοτήτων παράδειγμα 1

Ο μέσος χρόνος είναι:

$$\bar{x} = \frac{312}{30} = 10,4$$

2.1.2 Παραλλαγές του Αριθμητικού Μέσου

α-περικομμένος μέσος (a-trimmed mean) ορίζεται ο μέσος των παρατηρήσεων που απομένουν αφού παραλειφθεί το άνω 100α% και το κάτω 100α% των διατεταγμένων παρατηρήσεων.

Αν αντί να περικόψουμε το άνω και κάτω 25% των διατεταγμένων παρατηρήσεων αντικαταστήσουμε κάθε μία από τις παρατηρήσεις του άνω τετάρτου με το άνω τεταρτημόριο και κάθε μία από τις παρατηρήσεις του κάτω τετάρτου με το κάτω τεταρτημόριο, ο μέσος των παρατηρήσεων που προκύπτουν λέγεται Winsorised μέσος. Οι δύο αυτοί μέσοι χρησιμοποιούνται για να αποφύγουμε την επίδραση ακραίων τιμών.

2.1.3 Διάμεσος

Ορισμός: Η διάμεσος (M) ενός συνόλου μετρήσεων είναι η τιμή εκείνη με την ιδιότητα ότι το πολύ 50% των μετρήσεων είναι μικρότερες από την τιμή αυτή και το πολύ 50% των μετρήσεων είναι μεγαλύτερες από την τιμή αυτή.

Για να υπολογίσουμε τη διάμεσο ταξινομούμε τα δεδομένα κατά αύξουσα σειρά και στην περίπτωση που το πλήθος των δεδομένων είναι περιττό, η διάμεσος είναι η μεσαία παρατήρηση. Στην περίπτωση που το πλήθος των δεδομένων είναι άρτιο, η διάμεσος είναι το ημιάθροισμα των δύο μεσαίων παρατηρήσεων.

Όταν έχουμε ομαδοποιημένα δεδομένα, δεν είναι δυνατή η διάταξη των παρατηρήσεων σε αύξουσα σειρά. Στην περίπτωση ομαδοποιημένων δεδομένων σε v κλάσεις, ο υπολογισμός της διαμέσου γίνεται ως εξής:

1. Υπολογίζουμε τις αθροιστικές συχνότητες: N_1, N_2, \dots, N_v .
2. Εντοπίζουμε μεταξύ ποιων αθροιστικών συχνοτήτων N_{i-1} και N_i βρίσκεται ο αριθμός $v/2$. Εάν ο αριθμός $v/2$ συμπίπτει με μία από τις αθροιστικές συχνότητες, τότε ως διάμεσος ορίζεται το άνω άκρο της αντίστοιχης κλάσης.
3. Χρησιμοποιούμε τον τύπο:

$$M = L_{i-1} + \frac{\delta}{v_i} \left(\frac{v}{2} - N_{i-1} \right).$$

Όπου:

L_{i-1} : το κατώτερο όριο κλάσης i ,

v_i : η συχνότητα της κλάσης i ,

N_{i-1} : η αθροιστική συχνότητα της κλάσης $i-1$ η οποία προηγείται της κλάσης i ,

δ : το πλάτος της κλάσης i .

2.1.4 Επικρατούσα τιμή (M_0)

Είναι η τιμή εκείνη των δεδομένων που έχει τη μεγαλύτερη συχνότητα εμφάνισης. Στο παρακάτω σύνολο δεδομένων 1 3 4 3 4 5 4 2 η τιμή αυτή είναι το 4. Όταν υπάρχουν δύο τιμές με την ίδια συχνότητα εμφάνισης τότε λέμε ότι τα δεδομένα έχουν δύο επικρατούσες τιμές. Η κατανομή με μία επικρατούσα τιμή λέγεται μονοκόρυφη (unimodal) ενώ με δύο λέγεται δικόρυφη (bimodal). Η επικρατούσα τιμή χρησιμοποιείται κυρίως για ποιοτικά δεδομένα.

2.1.5 Γεωμετρικός Μέσος (GM)

Έστω x_1, x_2, \dots, x_k οι τιμές μιας μεταβλητής X , που αφορά τα στοιχεία ενός δείγματος μεγέθους $v, k \leq v$. Ο γεωμετρικός μέσος δίνεται από τον τύπο:

$$GM = \sqrt[v]{x_1^{v_1} \cdot x_2^{v_2} \cdot \dots \cdot x_k^{v_k}}.$$

Για τον υπολογισμό του γεωμετρικού μέσου χρησιμοποιούμε λογαρίθμους. Λογαριθμούμε και τα δύο μέλη της παραπάνω σχέσης και έχουμε:

$$\begin{aligned} \log GM &= \log \sqrt[v]{x_1^{v_1} \cdot x_2^{v_2} \cdot \dots \cdot x_k^{v_k}} = \log \left(x_1^{v_1} \cdot x_2^{v_2} \cdot \dots \cdot x_k^{v_k} \right)^{\frac{1}{v}} \\ &= \frac{1}{v} \cdot \log \left(x_1^{v_1} \cdot x_2^{v_2} \cdot \dots \cdot x_k^{v_k} \right) = \frac{1}{v} \cdot \sum_{i=1}^k \log \left(x_i^{v_i} \right) = \frac{1}{v} \cdot \sum_{i=1}^k v_i \log \left(x_i \right). \end{aligned}$$

Από τη σχέση αυτή προκύπτει ότι ο λογάριθμος του γεωμετρικού μέσου ισούτε με τον αριθμητικό μέσο των λογαρίθμων των τιμών της μεταβλητής X .

Έστω οι παρατηρήσεις 3 25 45 τότε $GM = \sqrt[3]{3 \cdot 25 \cdot 45} = 15$. Δεν μπορεί να οριστεί για αρνητικές ή μηδενικές τιμές ενώ πάντα είναι μικρότερος από τον αριθμητικό μέσο εκτός αν

οι παρατηρήσεις ταυτίζονται οπότε ταυτίζονται και οι δύο μέσοι. Ο γεωμετρικός μέσος είναι πιο σταθερός από τον αριθμητικό αφού επηρεάζεται λιγότερο από παρατηρήσεις με πολύ μεγάλες τιμές. Για παράδειγμα στο παρακάτω σύνολο δεδομένων 6 8 10 10 10 12 16 ο αριθμητικός μέσος είναι 10,286 και ο γεωμετρικός 9884. Αν ωστόσο αντί της τιμής 16 είχαμε την τιμή 48 τότε ο γεωμετρικός θα ήταν 11,564 και ο αριθμητικός 14,857.

Επίσης ο γεωμετρικός μέσος είναι κατάλληλο μέτρο για δεδομένα που αυξάνονται ή ελαττώνονται με γεωμετρική πρόοδο. Τέτοιο παράδειγμα είναι οι πληθυσμοί. Αν σε μια πόλη ο πληθυσμός το 1980 ήταν 270000 και το 1990 ήταν 510000 και θέλαμε να εκτιμήσουμε το μέγεθος του πληθυσμού το 1985 θα μπορούσαμε να ισχυρισθούμε ότι αυτό είναι σύμφωνα με τον αριθμητικό μέσο 390000.

Ωστόσο αυτό θα ήταν λογικό αν ο πληθυσμός αυξανόταν κατά σταθερό αριθμό κάθε χρόνο κάτι το οποίο δεν ισχύει. Η λύση στο πρόβλημα αυτό είναι η χρήση του γεωμετρικού μέσου ο οποίος λαμβάνει την τιμή 363730.

2.1.6 Αρμονικός Μέσος (HM)

Έστω x_1, x_2, \dots, x_k οι τιμές μιας μεταβλητής X , που αφορά τα στοιχεία ενός δείγματος μεγέθους $n, k \leq n$. Ο αρμονικός μέσος δίνεται από τον τύπο:

$$HM = \frac{n}{\frac{v_1}{x_1} + \frac{v_2}{x_2} + \dots + \frac{v_k}{x_k}} = \frac{n}{\sum_{i=1}^k v_i \frac{1}{x_i}}$$

Ο αρμονικός μέσος χρησιμοποιείται ως μέσος για ρυθμούς. Για παράδειγμα αν ένα αυτοκίνητο ταξιδεύει προς ένα μέρος με ταχύτητα 60km/h και επιστρέφει με ταχύτητα 40km/h τότε η μέση ταχύτητα δεν είναι 50km/h. Αυτή θα ήταν η μέση ταχύτητα αν το αυτοκίνητο είχε κινηθεί για μια ώρα με ταχύτητα 60km/h και στη συνέχεια για μια ώρα με ταχύτητα 40km/h. Για να καθορίσουμε τη μέση ταχύτητα όταν έχουν χρησιμοποιηθεί

διαφορετικές ταχύτητες πάνω στην ίδια απόσταση χρησιμοποιείται ο αρμονικός μέσος. Στο

παράδειγμά μας ο αρμονικός μέσος είναι $\frac{2}{\frac{1}{60} + \frac{1}{40}} = 48 \text{ km/h}$. Επίσης ισχύει ότι ο αρμονικός

μέσος είναι μικρότερος από τον γεωμετρικό ο οποίος είναι μικρότερος από τον αριθμητικό εκτός αν ταυτίζονται οι τιμές.

2.1.7 Ποσοστιαία Σημεία (percentiles)

Γενικά μπορούμε να προσδιορίσουμε ένα σύνολο $n-1$ τιμών οι οποίες διαιρούν τη συνολική συχνότητα σε n ίσα μέρη δηλαδή τα ποσοτικά σημεία (quantiles). Ορίζουμε ως k -ποσοστιαίο σημείο (k th quantile) ενός συνόλου τιμών την τιμή εκείνη που έχει την ιδιότητα ότι το πολύ $100k/n\%$ των τιμών είναι μικρότερες από την τιμή αυτή και το πολύ $100(n-k)/n\%$ των τιμών είναι μεγαλύτερες από την τιμή αυτή.

Τα ποσοστιαία σημεία που χωρίζουν ένα σύνολο διατεταγμένων μετρήσεων σε τέταρτα και δέκατα ονομάζονται αντίστοιχα τεταρτημόρια και δεκατημόρια. Έτσι έχουμε:

Πρώτο (κάτω) δεκατημόριο = 10ο εκατοστιαίο σημείο

Q_1 = πρώτο (κάτω) τεταρτημόριο = 25^ο εκατοστιαίο σημείο

Q_2 = δεύτερο (μεσαίο) τεταρτημόριο = διάμεσος (50ο εκατοστιαίο σημείο)

Q_3 = τρίτο (άνω) τεταρτημόριο = 75^ο ποσοστιαίο σημείο

Ένατο (άνω) δεκατημόριο = 90ο ποσοστιαίο σημείο

Γενικά τα ποσοστιαία σημεία βρίσκονται ως εξής: Έχοντας διατάξει τα δεδομένα (N το πλήθος) κατά αύξουσα σειρά το πρώτο δεκατημόριο είναι το $\frac{N+1}{10}$ σημείο, το κάτω

τεταρτημόριο είναι το $\frac{N+1}{4}$ σημείο, η διάμεσος είναι το $\frac{N+1}{2}$ σημείο, το άνω

τεταρτημόριο είναι το $\frac{3(N+1)}{4}$ σημείο και το άνω δεκατημόριο είναι το $\frac{9(N+1)}{10}$ σημείο.

Σε περίπτωση ομαδοποιημένων δεδομένων σε v κλάσεις, ο υπολογισμός του πρώτου τεταρτημορίου γίνεται ως εξής:

- 1) Υπολογίζουμε τις αθροιστικές συχνότητες: N_1, N_2, \dots, N_v .
- 2) Εντοπίζουμε μεταξύ ποιων αθροιστικών συχνοτήτων N_{i-1} και N_i βρίσκεται ο αριθμός $v/4$. Εάν ο αριθμός $v/4$ συμπίπτει με μία από τις αθροιστικές συχνότητες, τότε ως πρώτο τεταρτημόριο ορίζεται το άνω άκρο της αντίστοιχης κλάσης.
- 3) Χρησιμοποιούμε τον τύπο:

$$Q_1 = L_{i-1} + \frac{\delta}{v_i} \left(\frac{v}{4} - N_{i-1} \right).$$

Όπου:

L_{i-1} : το κατώτερο όριο κλάσης i ,

v_i : η συχνότητα της κλάσης i ,

N_{i-1} : η αθροιστική συχνότητα της κλάσης $i-1$ η οποία προηγείται της κλάσης i ,

δ : το πλάτος της κλάσης i .

Αντίστοιχα, στην περίπτωση ομαδοποιημένων δεδομένων σε v κλάσεις, ο υπολογισμός του τρίτου τεταρτημορίου γίνεται ως εξής:

- 1) Υπολογίζουμε τις αθροιστικές συχνότητες: N_1, N_2, \dots, N_v .
- 2) Εντοπίζουμε μεταξύ ποιων αθροιστικών συχνοτήτων N_{i-1} και N_i βρίσκεται ο αριθμός $3v/4$. Εάν ο αριθμός $3v/4$ συμπίπτει με μία από τις αθροιστικές συχνότητες, τότε ως τρίτο τεταρτημόριο ορίζεται το άνω άκρο της αντίστοιχης κλάσης.
- 3) Χρησιμοποιούμε τον τύπο:

$$Q_3 = L_{i-1} + \frac{\delta}{v_i} \left(\frac{3v}{4} - N_{i-1} \right).$$

Όπου:

L_{i-1} : το κατώτερο όριο κλάσης i ,

v_i : η συχνότητα της κλάσης i ,

N_{i-1} : η αθροιστική συχνότητα της κλάσης $i-1$ η οποία προηγείται της κλάσης i ,

δ : το πλάτος της κλάσης i .

2.2 Σύγκριση Μέτρων Θέσης

2.2.1 Πλεονεκτήματα και Μειονεκτήματα των μέτρων Κεντρικής Τάσεως

Διάμεσος

Πλεονεκτήματα

- Είναι πιο εύκολο να υπολογιστεί σε σχέση με τον μέσο όρο (εκτός αν χρησιμοποιηθεί ο μαθηματικός τύπος υπολογισμού της).
- Δεν επηρεάζεται από τις ακραίες τιμές, οπότε είναι καλύτερος δείκτης κεντρικής τάσεως όταν έχουμε ασύμμετρη κατανομή.
- Μπορεί να υπολογιστεί ακόμα και όταν δεν γνωρίζουμε τις ακραίες τιμές.

Μειονεκτήματα

- Δεν λαμβάνει υπόψη την ακριβή τιμή του κάθε στοιχείου.
- Δεν μπορεί να χρησιμοποιηθεί για να υπολογιστούν παράμετροι του πληθυσμού. Αν οι τιμές της κατανομής είναι λίγες, τότε η διάμεσος δεν μπορεί να τις

αντιπροσωπεύσει με ακρίβεια. Π.χ. Αν έχουμε τις τιμές 2,5,8,67 και 110 η διάμεσος είναι το 8.

Μέση τιμή

Πλεονεκτήματα

- Είναι εύκολος στον υπολογισμό του.
- Αντικατοπτρίζει πιο πιστά την κεντρική τιμή της κατανομής σε σχέση με τους άλλους δείκτες.
- Μπορεί να χρησιμοποιηθεί για τον υπολογισμό των παραμέτρων του πληθυσμού (παραμετρικά τεστ).

Μειονεκτήματα

- Είναι ευαίσθητη στις τιμές των δεδομένων της κατανομής. Π.χ. μια αλλαγή σε οποιαδήποτε τιμή προκαλεί διαφοροποίηση της μέσης τιμής.
- Επειδή υπολογίζεται αλγεβρικά, η τιμή της είναι πιθανόν να μην ανήκει στις τιμές της κατανομής.
- Είναι πολύ ευαίσθητη στις ακραίες τιμές.

3 Μέτρα διασποράς

3.1 Μέτρα μεταβλητότητας ή διασποράς (Measures of variation or dispersion)

Μεταβλητότητα είναι το 'άπλωμα' ή η διασπορά των τιμών σε ένα σύνολο δεδομένων. Η μεταβλητότητα μας δείχνει πόσο αντιπροσωπευτικό είναι το μέτρο θέσης σε σχέση με όλες τις μετρήσεις του συνόλου των δεδομένων. Δηλαδή πως 'απλώνονται' οι μετρήσεις γύρω από αυτό το μέτρο θέσης.

3.1.1 Έκταση ή Εύρος (Range)

Ορίζεται ως η διαφορά μεταξύ μεγαλύτερης και μικρότερης τιμής ενός συνόλου δεδομένων. Δηλαδή $R = X_{\max} - X_{\min}$. Το πλεονέκτημά της είναι η ευκολία στον υπολογισμό της ενώ το μειονέκτημα της είναι ότι εξαρτάται από δύο μόνο τιμές του συνόλου των παρατηρήσεων χωρίς να λαμβάνει υπόψη της τις υπόλοιπες. Για παράδειγμα στα παρακάτω σύνολα δεδομένων 1 2 3 7 12 και 1 1 1 12 12 η έκταση λαμβάνει την ίδια τιμή 12-1=11 ωστόσο το δεύτερο σύνολο δεδομένων εμφανίζει πολύ μεγαλύτερη μεταβλητότητα από το πρώτο.

3.1.2 Ενδοτεταρτημοριακό Εύρος (Interquartile Range)

Το ενδοτεταρτημοριακό εύρος (interquartile range) είναι η διαφορά του πρώτου από το τρίτο τεταρτημόριο:

$$H = Q_3 - Q_1$$

Στο μεταξύ τους διάστημα βρίσκεται το 50% των τιμών της κατανομής. Επομένως, όσο μικρότερο είναι αυτό το διάστημα, τόσο μεγαλύτερη θα είναι η συγκέντρωση των τιμών και άρα μικρότερη η διασπορά των τιμών της μεταβλητής. Το μισό του ενδοτεταρτημοριακού εύρους είναι γνωστό ως ημιενδοτεταρτημοριακό εύρος (semi-interquartile range) και συμβολίζεται με Q . Μετριέται με τις ίδιες μονάδες της μεταβλητής και δεν εξαρτάται από όλες τις τιμές αλλά μόνο από εκείνες που περιλαμβάνονται στον υπολογισμό του πρώτου και τρίτου τεταρτημορίου.

3.1.3 Μέση Απόλυτη Απόκλιση (Mean Absolute Deviation)

Το μέτρο αυτό θεωρεί τη μεταβλητότητα ως το βαθμό στον οποίο οι τιμές ενός συνόλου δεδομένων αποκλίνουν από το μέσο τους. Ως μέση απόλυτη απόκλιση ορίζεται ο μέσος των απολύτων τιμών των αποκλίσεων των παρατηρήσεων από το μέσο των παρατηρήσεων αυτών.

Έστω x_1, x_2, \dots, x_k οι τιμές μιας μεταβλητής X , που αφορά τα στοιχεία ενός δείγματος

μεγέθους $v, k \leq v$. Η μέση τιμή δίνεται από τον τύπο:

$$MAD = \frac{v_1 \cdot |x_1 - \bar{x}| + v_2 \cdot |x_2 - \bar{x}| + \dots + v_k \cdot |x_k - \bar{x}|}{v} = \frac{1}{v} \cdot \sum_{i=1}^k v_i \cdot |x_i - \bar{x}|.$$

Για παράδειγμα στο παρακάτω σύνολο δεδομένων 1 2 3 7 12 έχουμε

$$MAD = \frac{|1-5| + |2-5| + |3-5| + |7-5| + |12-5|}{5} = 3,6.$$

Μειονέκτημά της είναι ότι υπάρχει δυσκολία στην ανάπτυξη στατιστικών μεθόδων δοθέντος ότι είναι συνάρτηση απολύτων τιμών.

3.1.4 Διακύμανση ή Διασπορά (Variance)

Έστω x_1, x_2, \dots, x_k οι τιμές μιας μεταβλητής X , που αφορά τα στοιχεία ενός δείγματος μεγέθους $v, k \leq v$. Η διακύμανση δίνεται από τον τύπο:

$$s^2 = \frac{v_1 \cdot (x_1 - \bar{x})^2 + v_2 \cdot (x_2 - \bar{x})^2 + \dots + v_k \cdot (x_k - \bar{x})^2}{v-1} = \frac{1}{v-1} \sum_{i=1}^k v_i (x_i - \bar{x})^2.$$

Στην περίπτωση που οι τιμές της μεταβλητής X αφορούν τα στοιχεία ενός πληθυσμού με N στοιχεία, η πληθυσμιακή διακύμανση δίνεται από τον τύπο:

$$\sigma^2 = \frac{v_1 \cdot (x_1 - \mu)^2 + v_2 \cdot (x_2 - \mu)^2 + \dots + v_k \cdot (x_k - \mu)^2}{N} = \frac{1}{N} \sum_{i=1}^k v_i (x_i - \mu)^2.$$

Για παράδειγμα στον πληθυσμό με τις παρατηρήσεις:

11 12 13 17 22

με $\mu=15$ έχουμε

$$\sigma^2 = \frac{(11-15)^2 + (12-15)^2 + (13-15)^2 + (17-15)^2 + (22-15)^2}{5} = 16,4.$$

Ένας πιο εύχρηστος τύπος για τον υπολογισμό της δειγματικής διακύμανσης είναι ο εξής:

$$s^2 = \frac{1}{v-1} \left[\sum_{i=1}^k v_i x_i^2 - \frac{1}{v} \left(\sum_{i=1}^k v_i x_i \right)^2 \right].$$

Όταν έχουμε ομαδοποιημένα δεδομένα ο τύπος της δειγματικής διακύμανσης είναι

$$s^2 = \frac{1}{v-1} \left[\sum_{i=1}^k v_i m_i^2 - \frac{1}{v} \left(\sum_{i=1}^k v_i m_i \right)^2 \right].$$

Όπου v_i η συχνότητα στην κλάση i ($i=1,2,\dots,k$) και m_i το ενδιάμεσο σημείο κάθε κλάσης με την παραδοχή ότι αυτό προσεγγίζει ικανοποιητικά το μέσο των μετρήσεων που ανήκουν σε κάθε κλάση.

3.1.5 Τυπική Απόκλιση (Standard Deviation)

Η αριθμητική τιμή της διακύμανσης ενός πληθυσμού, ή ενός δείγματος, είναι δύσκολο να ερμηνευθεί δεδομένου ότι εκφράζεται σε τετράγωνα των μονάδων των παρατηρήσεων. Για να αποφευχθεί το πρόβλημα αυτό ορίστηκε ένα μέτρο μεταβλητότητας που να εκφράζεται στις ίδιες μονάδες όπως τα αρχικά δεδομένα.

Ορίζουμε ως τυπική απόκλιση την θετική τετραγωνική ρίζα της διακύμανσης. Δηλαδή $\sigma = +\sqrt{\sigma^2}$ και $s = +\sqrt{s^2}$. Προφανώς διακύμανση και τυπική απόκλιση δεν λαμβάνουν αρνητικές τιμές. Αυτό διότι η διακύμανση είναι ένας μέσος τετραγωνικών ποσοτήτων και η τυπική απόκλιση είναι η θετική τετραγωνική ρίζα της διακύμανσης.

Παράδειγμα:

Έστω ότι ο ετήσιος (ποσοστιαίος) ρυθμός απόδοσης για τα τελευταία δέκα χρόνια δύο αμοιβαίων κεφαλαίων δίνεται από τα στοιχεία που ακολουθούν:

A. 8,3 -6,2 20,9 -2,7 33,6 42,9 24,4 5,2 3,1 30,5

B. 12,1 -2,8 6,4 12,2 27,8 25,3 18,2 10,7 -1,3 11,4

Ποιο από τα δύο αμοιβαία κεφάλαια θεωρείτε ότι περικλείει το μεγαλύτερο βαθμό επενδυτικού κινδύνου;

Λύση:

Για το κεφάλαιο A έχουμε $\sum_{i=1}^{10} x_i = 160$ $\sum_{i=1}^{10} x_i^2 = 5083,06$

$$s_A^2 = \frac{1}{9} \left[5083,06 - \frac{160^2}{10} \right] = 280,34 \text{ και } s_A = 16,74$$

Για το κεφάλαιο B έχουμε $\sum_{i=1}^{10} x_i = 120$ $\sum_{i=1}^{10} x_i^2 = 2334,36$

$$s_B^2 = \frac{1}{9} \left[2334,36 - \frac{120^2}{10} \right] = 99,37 \text{ και } s_B = 9,97.$$

Παρατηρούμε ότι το αμοιβαίο κεφάλαιο A έχει μεγαλύτερο επίπεδο επικινδυνότητας όπως αυτό μετριέται με τη διακύμανση. Παρατηρούμε επίσης ότι το κεφάλαιο A έχει υψηλότερο μέσο ετήσιο ρυθμό απόδοσης από το κεφάλαιο B διότι $\bar{x}_A = \frac{160}{10} = 16$ και $\bar{x}_B = \frac{120}{10} = 12$. Το αποτέλεσμα συμβαδίζει με την αντίληψη ότι μια επένδυση με υψηλότερο κίνδυνο αποδίδει μεγαλύτερο ρυθμό απόδοσης.

3.1.6 Βαθμοί Ελευθερίας (Degrees of Freedom)

Η τιμή $n-1$ που απαιτείται για τον υπολογισμό της δειγματικής διακύμανσης αναφέρεται ως βαθμοί ελευθερίας. Η ονομασία οφείλεται στο ότι αν πρόκειται να διαλέξουμε n τιμές που θα πρέπει να έχουν ένα δεδομένο μέσο, ο αριθμός των τιμών που μπορούμε να διαλέξουμε ελεύθερα και αυθαίρετα είναι $n-1$.

Αν ο μέσος των n τιμών είναι καθορισμένος τότε και το άθροισμά τους είναι καθορισμένο οπότε η τελευταία τιμή θα προκύπτει ως η διαφορά του αθροίσματος των n τιμών μείον το άθροισμα των $n-1$ τιμών που έχουν επιλεγεί αυθαίρετα. Δοθέντος ότι η δειγματική διακύμανση υπολογίζεται μέσω των αποκλίσεων των n τιμών των δεδομένων από τον μέσο τους, λέμε ότι έχει $n-1$ βαθμούς ελευθερίας.

3.1.7 Χρήση της τυπικής απόκλισης

Θεώρημα Chebyshev:

Δοθέντος ενός συνόλου τιμών και ενός αριθμού $k \geq 1$, το ποσοστό των τιμών αυτών που περιλαμβάνεται σε απόσταση k τυπικών αποκλίσεων από τον αντίστοιχο μέσο τους είναι τουλάχιστον $1 - \frac{1}{k^2}$.

k	Διάστημα	Ποσοστό μετρήσεων στο διάστημα αυτό
1	$(\mu - \sigma, \mu + \sigma)$	$\geq 0\%$
2	$(\mu - 2\sigma, \mu + 2\sigma)$	$\geq 75\%$
2,5	$(\mu - 2,5\sigma, \mu + 2,5\sigma)$	$\geq 84\%$
3	$(\mu - 3\sigma, \mu + 3\sigma)$	$\geq 89\%$

Πίνακας 5: Chebyshev

Το θεώρημα αυτό μπορεί να εφαρμοσθεί για οποιοδήποτε σύνολο δεδομένων ανεξάρτητα από την κατανομή που τα δεδομένα αυτά έχουν. Γι' αυτό το λόγο έχουμε και συντηρητικά όρια. Το θεώρημα επίσης ισχύει αν αντικαταστήσουμε την μέση τιμή μ και την τυπική απόκλιση σ του πληθυσμού με την μέση τιμή \bar{x} και την τυπική απόκλιση s του δείγματος αντίστοιχα.

Παράδειγμα: Τα δεδομένα ενός δείγματος διάρκειας τηλεφωνημάτων έχουν μέσο $\bar{x} = 10,26$ και τυπική απόκλιση $s = 4,29$. Αν δεν είχαμε καμιά άλλη πληροφορία σε σχέση με την κατανομή της διάρκειας των τηλεφωνημάτων, από το θεώρημα του Chebyshev προκύπτει ότι τουλάχιστον 75% των τηλεφωνημάτων έχουν διάρκεια που βρίσκεται στο διάστημα $(\bar{x} - 2s, \bar{x} + 2s) = (1,68, 18,84)$.

Στην πραγματικότητα το ποσοστό αυτό είναι 96,7% κάτι που οφείλεται στη μορφή της κατανομής των δεδομένων. Όταν τα δεδομένα έχουν μια συμμετρική κατανομή με μια μόνο κορυφή, σχήμα το οποίο ταιριάζει στην κανονική κατανομή ισχύει ο εμπειρικός κανόνας.

3.1.8 Εμπειρικός Κανόνας

Για σύνολα δεδομένων με κατανομή συχνότητας που προσεγγίζει την κανονική κατανομή ή περιέχουν πολλά στοιχεία έχουμε ότι:

- Το 68% των παρατηρήσεων βρίσκονται στο διάστημα που τα άκρα του απέχουν μια τυπική απόκλιση από το μέσο.
- Το 95% των παρατηρήσεων βρίσκονται στο διάστημα που τα άκρα του απέχουν δύο τυπικές αποκλίσεις από το μέσο.
- Το 99% των παρατηρήσεων βρίσκονται στο διάστημα που τα άκρα του απέχουν τρεις τυπικές αποκλίσεις από τον μέσο.

Όταν τα δεδομένα ακολουθούν μια κατανομή που είναι έντονα μη συμμετρική ή έχει περισσότερες από μία κορυφές ο εμπειρικός κανόνας δεν θα πρέπει να εφαρμόζεται.

3.1.9 Τυποποιημένες τιμές (Standardized Values)

Οι τυποποιημένες τιμές είναι ένα άλλο μέτρο σχετικής θέσης και ονομάζονται αλλιώς Z-τιμές (Z-values). Ορίζεται ως τυποποιημένη τιμή ο αριθμός των τυπικών αποκλίσεων κατά τις οποίες μια παρατήρηση βρίσκεται πάνω ή κάτω από το μέσο.

Ο γενικός τρόπος υπολογισμού της Z-τιμής ενός συνόλου παρατηρήσεων είναι:

$$Z = \frac{X - \mu}{\sigma}$$

όπου, X η συγκεκριμένη παρατήρηση. Για το δείγμα η τυπική τιμή υπολογίζεται αντικαθιστώντας το μ με \bar{x} και το σ με s . Μπορούμε να παρατηρήσουμε ότι οι Z-τιμές είναι καθαροί αριθμοί (δεν εκφράζονται σε κάποια μονάδα μέτρησης) συνεπώς είναι ιδιαίτερα χρήσιμες όταν θέλουμε να συγκρίνουμε αποδόσεις που έχουν μετρηθεί σε διαφορετικές κλίμακες.

Παράδειγμα:

Έστω θέλουμε να κατατάξουμε 6 μαθητές σύμφωνα με τους βαθμούς τους οι οποίοι προέρχονται από 2 διαφορετικά σχολεία.

	X-τιμές	Κατάταξη σύμφωνα με X-τιμές	Z-τιμές	Κατάταξη σύμφωνα με Z-τιμές
$\mu_1 = 14$ $\sigma_1 = 2$	14	A	0	Γ
	13	B	-0,5	E
	12	Γ	-1	Z
$\mu_2 = 10$ $\sigma_2 = 4$	11	Δ	0,25	A
	10,5	E	0,125	B
	9	Z	-0,25	Δ

Πίνακας 6: Παράδειγμα τυποποιημένων τιμών

Από το παραπάνω πίνακα γίνεται κατανοητό ότι μια αρνητική Z-τιμή αποτελεί ένδειξη ότι μια παρατήρηση βρίσκεται κάτω από το μέσο ενώ μια θετική Z-τιμή αποτελεί ένδειξη ότι η αντίστοιχη παρατήρηση είναι μεγαλύτερη από το μέσο.

3.1.10 Μέτρα Σχετικής Μεταβλητότητας

Είναι γνωστό ότι μια απώλεια 5 ευρώ από μια επένδυση 100 ευρώ είναι σαφώς μεγαλύτερη από μια απώλεια 5 ευρώ μιας επένδυσης 200 ευρώ μόνο αν χρησιμοποιήσουμε ποσοστιαίες μεταβολές. Προβλήματα αυτής της μορφής οδηγούν στην ανάγκη ορισμού μέτρων που συνδυάζουν μέτρα θέσης με μέτρα απόκλισης.

3.1.11 Συντελεστής Μεταβλητότητας (Coefficient of Variation)

Ο συντελεστής μεταβλητότητας είναι ένα μέτρο σχετικής μεταβλητότητας και εκφράζεται συνήθως ως ποσοστό και όχι μέσω των μονάδων των δεδομένων στα οποία αναφέρεται. Ο συντελεστής μεταβλητότητας συμβολίζεται με CV και μετρά το 'άπλωμα' των δεδομένων σε σχέση με το μέσο και έχει τύπο $CV = \frac{s}{\bar{x}}$.

Μπορεί να χρησιμεύσει για σύγκριση συνόλων δεδομένων που έχουν μετρηθεί με τις ίδιες μονάδες αλλά διαφέρουν σε τέτοιο βαθμό ώστε μια άμεση σύγκριση των αντίστοιχων τυπικών αποκλίσεων να μην είναι χρήσιμη.

Για παράδειγμα αν θεωρήσουμε δύο σειρές A και B οι οποίες έχουν τυπικές αποκλίσεις 16 και 14 αντίστοιχα, θα επιλέγαμε τη σειρά B, αν ωστόσο γνωρίζαμε ότι είχαν αντίστοιχους μέσους 16 και 13 τότε οι συντελεστές μεταβλητότητας θα ήταν για την A 1,00 και για τη B σειρά 1,08 οπότε θα επιλέγαμε την A σειρά.

Ο συντελεστής μεταβλητότητας πολλαπλασιάζεται μερικές φορές με το 100 και δίνεται ως ποσοστό, πράγμα που σημαίνει ότι εκφράζει την τυπική απόκλιση σαν ένα ποσοστό του μέσου. Για παράδειγμα για τη σειρά A θα είχαμε την τιμή 100% που σημαίνει ότι η τυπική απόκλισή της είναι ίση με το 100% του μέσου της.

3.1.12 Μέση Διαφορά του Gini (Gini's mean difference)

Είναι ένα άλλο μέτρο σχετικής μεταβλητότητας. Έστω ότι έχουμε τις παρατηρήσεις x_1, x_2, \dots, x_n . Η μέση διαφορά του Gini δίνεται από τον τύπο:

$$g = \frac{1}{v(v-1)} \sum_{i=1}^v \sum_{j=1}^v |x_i - x_j|.$$

Ο δείκτης g εκφράζει τον μέσο των μεταξύ των μετρήσεων αποκλίσεων. Παράδειγμα: Το μηνιαίο εισόδημα 4 μελών μιας οικογένειας είναι 300 250 280 100. Η τιμή του g είναι

$$g = \frac{1}{4 \cdot 3} =$$

$$\left[|300 - 250| + |250 - 300| + |300 - 280| + |280 - 300| + |300 - 100| + |100 - 300| + |250 - 280| + |280 - 250| + |250 - 100| + |100 - 250| + |280 - 100| + |100 - 280| \right] = 83,3.$$

Επομένως ο μέσος όρος των απολύτων διαφορών του μισθού κάθε μέλους της οικογένειας αυτής από τους μισθούς των υπολοίπων μελών της οικογένειας είναι 83,3 χιλ.

Η διαφορά του δείκτη g από τα άλλα μέτρα σχετικής μεταβλητότητας είναι ότι ο g αναφέρεται σε μέση απόκλιση των μετρήσεων μεταξύ τους ενώ τα άλλα μέτρα κάνουν χρήση των αποκλίσεων των μετρήσεων από το μέσο τους. Ένας πιο απλοποιημένος τύπος είναι ο εξής:

$$g = \frac{2}{v(v-1)} \sum_{\substack{i,j=1 \\ i < j}}^n |x_i - x_j|.$$

3.1.13 Καμπύλη Συγκέντρωσης ή καμπύλη Lorenz

Κατά την εξέταση ορισμένων οικονομικών χαρακτηριστικών, όπως είναι το οικογενειακό εισόδημα, το μέγεθος των γεωργικών εκμεταλλεύσεων, το μέγεθος των βιομηχανικών επιχειρήσεων κλπ., πολλές φορές ιδιαίτερα ενδιαφέρει ο βαθμός συγκεντρώσεως της κατανομής τους.

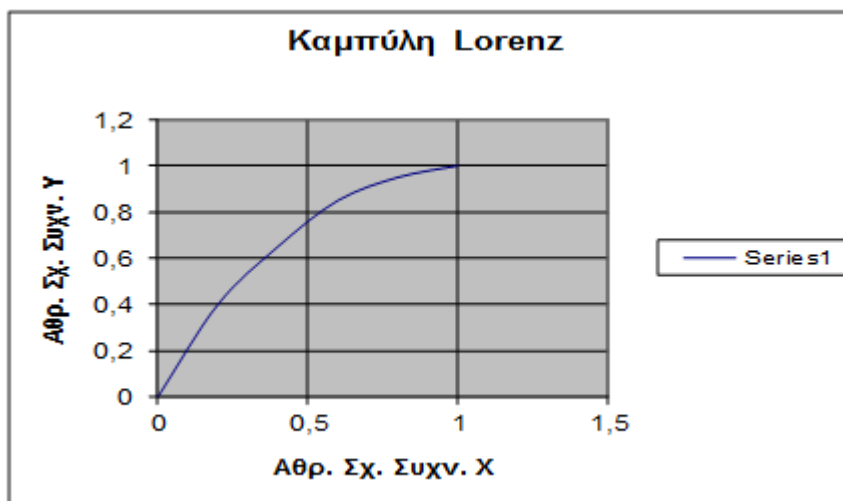
Η πιο απλή μέθοδος προσδιορισμού του βαθμού συγκεντρώσεως μιας κατανομής είναι η διαγραμματική και συνίσταται στην κατασκευή μιας καμπύλης που είναι γνωστή ως καμπύλη συγκεντρώσεως ή καμπύλη του Lorenz.

Παράδειγμα: Έστω 5 άτομα με τα εισοδήματά τους. Έχουμε:

Άτομα	Εισόδημα	X = αθροιστική σχετική συχνότητα στοιχείων	Y = αθροιστική σχετική συχνότητα εισοδήματος
A	800	0,2	0,4
B	500	0,4	0,65
Γ	400	0,6	0,85
Δ	200	0,8	0,95
E	100	1,0	1,00

Πίνακας 7: Πίνακας συχνοτήτων εισοδήματος

Και η καμπύλη Lorenz είναι η εξής:



Εικόνα 7: Εικόνα 7: Καμπύλη Lorenz εισοδήματος

3.2 Σύγκριση μέτρων διασποράς

Εύρος

Πλεονεκτήματα

- Είναι πολύ εύκολο στον υπολογισμό του.
- Περιλαμβάνει και τις ακραίες τιμές της κατανομής

Μειονεκτήματα

- Αλλοιώνεται από τις ακραίες τιμές, με αποτέλεσμα να μην παρουσιάζει σε πολλές περιπτώσεις, μια αντιπροσωπευτική εικόνα της διασποράς της κατανομής
- Δεν παρέχει καμία πληροφορία σχετικά με τη διασπορά των τιμών μεταξύ των άκρων της κατανομής. Π.χ., δεν μας «λέει» τίποτα για τη διασπορά των τιμών της κατανομής γύρω από τη μέση τιμή

Ενδοτεταρτημοριακό Εύρος

Πλεονεκτήματα

- Δεν επηρεάζεται από τις ακραίες τιμές
- Είναι σχετικά εύκολο στον υπολογισμό του
- Είναι αντιπροσωπευτικό των κεντρικών τιμών της κατανομής

Μειονεκτήματα

- Δεν λαμβάνει υπόψη τις ακραίες τιμές της Όπως και το εύρος,
- Δεν επιτρέπει την ακριβή ερμηνεία μιας συγκεκριμένης τιμής της κατανομής
- Δεν είναι ακριβές όταν τα δεδομένα είναι ομαδοποιημένα κατά μεγάλα διαστήματα τιμών
- Όπως και η διάμεσος δεν περιγράφει καμία από τις παραμέτρους του πληθυσμού που είναι βασικές για την επαγωγική στατιστική

Τυπική Απόκλιση

Πλεονεκτήματα

- Μπορεί να χρησιμοποιηθεί για τον υπολογισμό των παραμέτρων του πληθυσμού
- Λαμβάνει υπόψη όλες τις τιμές της κατανομής
- Είναι ο πιο ευαίσθητος από τους δείκτες διασποράς

Μειονεκτήματα

- Ο υπολογισμός της είναι σχετικά πιο περίπλοκος σε σχέση με τους υπόλοιπους δείκτες
- Είναι πολύ ευαίσθητη στις ακραίες τιμές της κατανομής.

3.3 Θηκόγραμμα

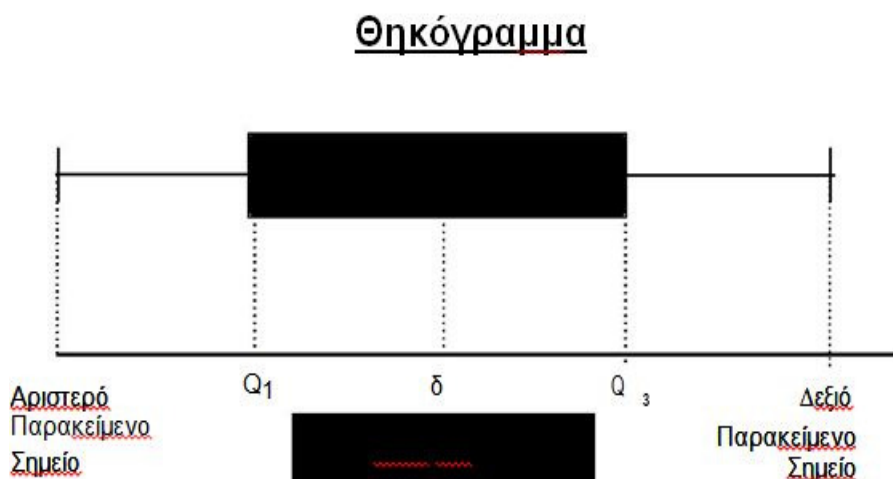
Το θηκόγραμμα είναι ένα γράφημα με το οποίο μπορούμε να παρουσιάσουμε τα κυριότερα χαρακτηριστικά της κατανομής του δείγματος. Μας βοηθάει να έχουμε μια εποπτική εικόνα της κατανομής της μεταβλητής και να εντοπίζουμε τυχόν παράτυπα σημεία.

Τρόπος Κατασκευής

- 1) Υπολογίζουμε: Διάμεσο, Q_1 , Q_3 και $Q_3 - Q_1$.
- 2) Κατασκευάζουμε ένα ορθογώνιο παραλληλόγραμμο με βάσεις τα Q_1 και Q_3 .
Χαράσσουμε στο παραλληλόγραμμο τη διάμεσο.
- 3) Υπολογίζουμε τις αποστάσεις: $Q_1 - 1,5(Q_3 - Q_1)$ και
 $Q_3 + 1,5(Q_3 - Q_1)$.
- 4) Ονομάζουμε αριστερό παρακείμενο σημείο (adjacent point) την παρατήρηση που είναι η αμέσως μεγαλύτερη της τιμής $Q_1 - 1,5(Q_3 - Q_1)$ και δεξιό παρακείμενο σημείο την παρατήρηση που είναι η αμέσως μικρότερη της τιμής $Q_3 + 1,5(Q_3 - Q_1)$.

Χαράσσουμε δύο ευθύγραμμα τμήματα από τα μέσα των βάσεων μέχρι τα παρακείμενα σημεία.

- 5) Οι τιμές που βρίσκονται έξω από τα όρια των παρακείμενων σημείων είναι πιθανά παράτυπα σημεία.



Εικόνα 8: Θηκόγραμμα

4 Μέτρα Ασυμμετρίας και Κύρτωσης

Ο τύπος της κατανομής συχνοτήτων, μπορεί να περιγραφεί ελέγχοντας κατά πόσο είναι συμμετρική ή ασύμμετρη και στην περίπτωση που δεν είναι συμμετρική ελέγχοντας το μεγαλύτερο μέρος της κατανομής εάν είναι προς τα δεξιά ή προς τα αριστερά.

Για τον παραπάνω έλεγχο χρησιμοποιούνται δύο μέτρα:

- Συντελεστής λοξότητας
- Συντελεστής κύρτωσης

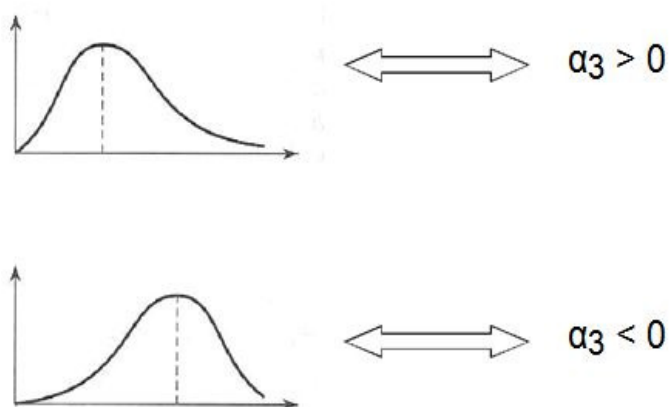
4.1 Συντελεστής λοξότητας

Έστω x_1, x_2, \dots, x_k οι τιμές μιας μεταβλητής X ,

που αφορά τα στοιχεία ενός δείγματος μεγέθους $n, k \leq n$. Ο συντελεστής λοξότητας ισούτε με:

$$a_3 = \frac{1}{n \cdot s^3} \sum_{i=1}^k v_i (x_i - \bar{x})^3.$$

Αν η κατανομή είναι συμμετρική τότε $a_3 = 0$. Συντελεστής λοξότητας θετικός (αρνητικός) σημαίνει ότι οι περισσότερες τιμές της μεταβλητής βρίσκονται δεξιά (αριστερά) της επικρατούσας τιμής.



Εικόνα 9: Συντελεστής λοξότητας

4.2 Συντελεστής κέρτωσης

Έστω x_1, x_2, \dots, x_k οι τιμές μιας μεταβλητής X , που αφορά τα στοιχεία ενός δείγματος μεγέθους $n, k \leq n$. Ο συντελεστής κέρτωσης ισούται με:

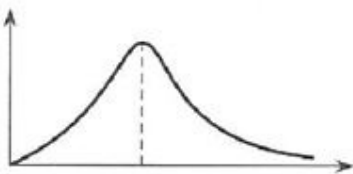
$$\alpha_4 = \frac{1}{n \cdot s^4} \sum_{i=1}^k v_i (x_i - \bar{x})^4.$$

Όταν ο συντελεστής κέρτωσης έχει τιμή μικρότερη του 3 η κατανομή λέγεται πλατύκυρτη ενώ όταν ο συντελεστής κέρτωσης έχει τιμή μεγαλύτερη του 3 η κατανομή λέγεται λεπτόκυρτη.



$$\alpha_4 < 3$$

πλατύκυρτη κατανομή



$$\alpha_4 > 3$$

λεπτόκυρτη κατανομή

Εικόνα 10: Συντελεστής κέρτωσης

5 Συντελεστή γραμμικής συσχέτισης

Ο συντελεστής γραμμικής συσχέτισης δύο μεταβλητών X και Y ορίζεται με βάση ένα δείγμα n ζευγών παρατηρήσεων (x_i, y_i) , $i=1,2,\dots,n$, συμβολίζεται με $r(X,Y)$ ή απλά με r και δίνεται από τον τύπο:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Ο συντελεστής συσχέτισης r μεταβάλλεται από -1 έως 1 δηλαδή, $-1 \leq r \leq 1$ και έχει τις εξής ιδιότητες:

Το πρόσημό του φανερώνει, αν η συσχέτιση είναι θετική ή αρνητική και είναι το ίδιο με του συντελεστή παλινδρόμησης $\hat{\beta}$. Οι δύο μεταβλητές X και Y θα λέμε ότι είναι θετικά συσχετισμένες, όταν η αύξηση των τιμών της μιας έχει ως συνέπεια και την αύξηση των τιμών της άλλης, ενώ θα λέμε ότι είναι αρνητικά συσχετισμένες, όταν η αύξηση των τιμών της μιας έχει συνέπεια τη μείωση των τιμών της άλλης. Το μέτρο του φανερώνει πόσο ισχυρά συσχετισμένες είναι οι μεταβλητές. Ισχυρή συσχέτιση εμφανίζουν οι μεταβλητές, όταν ο συντελεστής συσχέτισης πλησιάζει τις τιμές $+1$ ή -1 .

Αν $r = 1$, τότε έχουμε τέλεια θετική συσχέτιση. Όλες οι παρατηρήσεις βρίσκονται επάνω στην ευθεία παλινδρόμησης, η οποία έχει θετική κλίση ($\beta > 0$).

Αν $r = -1$, τότε έχουμε τέλεια αρνητική συσχέτιση. Όλες οι παρατηρήσεις βρίσκονται επάνω στην ευθεία παλινδρόμησης, η οποία έχει αρνητική κλίση ($\beta < 0$).

Αν $r = 0$, τότε οι μεταβλητές δεν εμφανίζουν γραμμική συσχέτιση και λέμε ότι οι μεταβλητές είναι γραμμικά ασυσχέτιστες.

Αποδεικνύεται ότι ο συντελεστής γραμμικής συσχέτισης r δίνεται ισοδύναμα και από τον παρακάτω τύπο, του οποίου διευκολύνει συχνά τους υπολογισμούς κυρίως στην περίπτωση που οι \bar{x}, \bar{y} δεν είναι ακέραιοι:

$$r = \frac{v \sum_{i=1}^v x_i y_i - \left(\sum_{i=1}^v x_i \right) \left(\sum_{i=1}^v y_i \right)}{\sqrt{v \sum_{i=1}^v x_i^2 - \left(\sum_{i=1}^v x_i \right)^2} \sqrt{v \sum_{i=1}^v y_i^2 - \left(\sum_{i=1}^v y_i \right)^2}}.$$

6 Στατιστική και Επιχειρήσεις

6.1 Αβεβαιότητα Οικονομικού Περιβάλλοντος

Η οικονομική δραστηριότητα συνίσταται στην αξιοποίηση των παραγωγικών συντελεστών, δηλαδή των φυσικών πόρων, της ανθρώπινης εργασίας και του διαθέσιμου κεφαλαίου, με στόχο την παραγωγή υλικών αγαθών και υπηρεσιών τα οποία καλύπτουν τις ανάγκες της κοινωνίας. Η αξιοποίηση αυτή, μάλιστα, πρέπει να είναι αποτελεσματική, με την έννοια ότι δεν επιτρέπεται να δαπανώνται πόροι πέραν των απολύτως αναγκαίων.

Στο καπιταλιστικό σύστημα, το οποίο κυριαρχεί σήμερα σε παγκόσμια βάση, η οικονομική δραστηριότητα έχει αναληφθεί κατά βάση από τις επιχειρήσεις, οι οποίες επιτρέπεται να δημιουργούνται χωρίς ουσιαστικούς περιορισμούς από οποιονδήποτε ενδιαφερόμενο. Στην ουσία, η αποτελεσματικότητα κάθε επιχείρησης αξιολογείται αυτομάτως από την Αγορά, με την έννοια ότι σε περιπτώσεις σπατάλης πόρων κατά την παραγωγική διαδικασία το αντίστοιχο κόστος είναι υψηλό και επομένως οι τιμές πώλησης των παραγόμενων προϊόντων διαμορφώνονται υψηλότερα σε σχέση με τους ανταγωνιστές. Έτσι, οι καταναλωτές τα αποφεύγουν (δεδομένου ότι έχουν φθηνότερες επιλογές) αναγκάζοντας τις

(αναποτελεσματικές) επιχειρήσεις να μειώσουν τα περιθώρια του κέρδους τους ή να πωλούν κάτω από το κόστος ή να σταματήσουν τη λειτουργία τους.

Στην πρώτη περίπτωση, η οποία είναι εφικτή μόνον όταν η σπατάλη πόρων είναι περιορισμένη, μειώνεται η αποδοτικότητα της επιχείρησης και επομένως αυτή δεν βρίσκει πρόθυμους χρηματοδότες να στηρίξουν τη λειτουργία της. Είναι προφανές ότι μια τέτοια επιχείρηση δεν μπορεί να σταθεί στην Αγορά για πολύ.

Στη δεύτερη περίπτωση, ο επιχειρηματίας καλείται να καλύπτει τη ζημία που προκαλεί η αρνητική διαφορά μεταξύ τιμής και κόστους. Αυτό δεν μπορεί, βέβαια, να συνεχίζεται για μεγάλη χρονική περίοδο, γι' αυτό η επιχείρηση σταματά (αναπόφευκτα) τη λειτουργία της, εκτός αν καταφέρει να αποκαταστήσει την αποτελεσματικότητά της.

Παρατηρούμε, δηλαδή, ότι υπάρχει ένας πολύ σοβαρός κίνδυνος για κάθε επιχείρηση, ο κίνδυνος μη αποτελεσματικής λειτουργίας, ο οποίος την οδηγεί στη χρεοκοπία.

Είναι δε τόσο πολλές οι παράμετροι της επιχειρηματικής αποτελεσματικότητας, ώστε ο πιο πάνω κίνδυνος να είναι εξαιρετικά μεγάλος και να μην εξαρτάται αποκλειστικά από την ικανότητα του επιχειρηματία και των στελεχών γενικότερα. Ενδεικτικά και μόνον, αναφερόμαστε σε μερικές από τις αιτίες που μπορούν να οδηγήσουν μια επιχείρηση σε αναποτελεσματική λειτουργία:

- Περιορισμένα κεφάλαια, τα οποία δεν επιτρέπουν την αγορά του κατάλληλου εξοπλισμού, την πρόσληψη ικανών στελεχών, την παροχή πιστώσεων στην πελατεία, την διείσδυση στην Αγορά μέσω ίδρυσης υποκαταστημάτων κλπ.
- Παρωχημένη τεχνολογία, η οποία έχει ως αποτέλεσμα αυξημένες αναλώσεις πρώτων υλών και εργασίας, δημιουργία προϊόντων κατώτερης ποιότητας, αδυναμία παραγωγής ειδών τα οποία έχουν συγκεκριμένες ιδιότητες κλπ.
- Έλλειψη τεχνογνωσίας (κύρια σε απομακρυσμένες ή υποανάπτυκτες περιοχές), που αναγκάζει την επιχείρηση να χρησιμοποιεί αναποτελεσματικές μεθόδους παραγωγής.

- Είσοδος πολλών νέων μονάδων στον κλάδο και επομένως διαμόρφωση συνθηκών σκληρού ανταγωνισμού ο οποίος περιορίζει τα περιθώρια κέρδους και επιδεινώνει την κατάσταση των λιγότερο ισχυρών ή οριακών οικονομικών μονάδων.

Πέρα, όμως, από τους κινδύνους που επηρεάζουν την αποτελεσματική λειτουργία μιας επιχείρησης και προέρχονται από τις επιχειρηματικές επιλογές ή τους περιορισμούς κάτω από τους οποίους αυτή λειτουργεί, υπάρχουν πρόσθετοι κίνδυνοι οι οποίοι πηγάζουν από το εξωτερικό (οικονομικό και κοινωνικό) περιβάλλον.

Ενδεικτικά και πάλι, αναφέρονται οι εξής κίνδυνοι αυτής της κατηγορίας :

- Ελλιπής υποδομή σε τηλεπικοινωνίες, οδικό δίκτυο, τράπεζες, δημόσιες υπηρεσίες κλπ (π.χ. επιχειρήσεις που ιδρύονται σε υποανάπτυκτες περιοχές).
- Σταδιακή μείωση της ζήτησης του προϊόντος, π.χ. λόγω αλλαγής προτιμήσεων ή υποκατάστασης, με αποτέλεσμα την μείωση του επιπέδου απασχόλησης της επιχείρησης και κατ'επέκταση την αύξηση της αδρανούς παραγωγικής δυναμικότητας της.
- Επιδείνωση των συνθηκών της εθνικής ή διεθνούς οικονομίας λόγω π.χ. ύφεσης. Σε μια τέτοια περίπτωση, η συνολική δαπάνη της κοινωνίας για τα προϊόντα του κλάδου περιορίζεται (εκτός από τα είδη πρώτης ανάγκης), με αποτέλεσμα την μείωση των εσόδων των αντίστοιχων επιχειρήσεων και την κατανομή του σταθερού κόστους λειτουργίας τους σε μικρότερο όγκο παραγωγής.
- Ασυνήθεις ή θεμελιακές μεταβολές, όπως π.χ. ισχυροποίηση νομίσματος (π.χ. είσοδος της Ελλάδας στην ΟΝΕ) ή έντονη εξασθένηση νομίσματος (π.χ. περιπτώσεις Αργεντινής, Βραζιλίας, Κορέας κλπ) ή διεύρυνση εθνικής οικονομίας (π.χ. ενοποίηση Γερμανιών) κλπ.
- Αλλαγή του πολιτικού κλίματος και πρόκληση αστάθειας, αβεβαιότητας κλπ.
- Εμπλοκή της χώρας σε εχθροπραξίες ή πολεμικές συγκρούσεις κλπ.
- Πρόκληση αναταραχής σε ευρύτερες γεωγραφικές περιοχές ή σε παγκόσμιο επίπεδο (π.χ. τρομοκρατικές ενέργειες 11/9/2001 στις ΗΠΑ).

Ο κίνδυνος που διατρέχουν οι οικονομικές μονάδες από τις συνθήκες και μεταβολές του εξωτερικού τους περιβάλλοντος, αναφέρονται συνοπτικά με τον όρο «συστηματικός κίνδυνος» ή «κίνδυνος αγοράς».

Είναι προφανές ότι, για να αντιμετωπίσει η επιχείρηση τα ενδεχόμενα που συνδέονται με το τεράστιο πλήθος εσωτερικών και εξωτερικών κινδύνων, είναι απαραίτητο να προγραμματίζει σωστά τη λειτουργία της και να αναπροσαρμόζει τη στρατηγική της κάθε φορά που ανατρέπονται τα δεδομένα στα οποία είχε βασίσει τον προηγούμενο σχεδιασμό της δράσης της.

6.2 Έννοια της Χρεοκοπίας

Από τους κινδύνους που διατρέχει η επιχείρηση, σοβαρότερος είναι ο κίνδυνος χρεοκοπίας που εκφράζει το ενδεχόμενο οριστικής λήξης της δραστηριότητάς της και (συνήθως) απώλειας των κεφαλαίων των μετόχων της, πολλές φορές δε και των δανειστών της.

Λόγω της μεγάλης σημασίας του, ο κίνδυνος χρεοκοπίας πρέπει να ανιχνεύεται και να μετριέται, ώστε να προστατεύεται (στο μέτρο του δυνατού) η βιωσιμότητα της επιχείρησης. Αυτός είναι ο λόγος που γίνεται εκτεταμένη αναφορά στη συνέχεια και δίνεται έμφαση στις μεθοδολογίες πρόγνωσης και μέτρησής του.

Η έννοια της χρεοκοπίας είναι πλατειά και για αυτό ο ορισμός της γίνεται δύσκολος αν όχι αδύνατος. Αρχικά, θα πρέπει να διακρίνουμε την ουσιαστική (οικονομική) από την τυπική (νομική) χρεοκοπία.

Η πρώτη περιλαμβάνει τις περιπτώσεις επιχειρήσεων που αντιμετωπίζουν αξεπέραστα προβλήματα ρευστότητας και επομένως οδηγούνται αναπότρεπτα σε παύση πληρωμών, στη συνέχεια δε σε διακοπή της λειτουργίας τους. Το πρόβλημα, στην περίπτωση αυτή, έγκειται στην αναγνώριση των συμπτωμάτων εκείνων που σηματοδοτούν το οριστικό αδιέξοδο

ρευστότητας και τον διαχωρισμό τους από τα συμπτώματα κακής μεν αλλά αντιστρέψιμης κατάστασης ελλιπούς ρευστότητας.

Για παράδειγμα, μια επιχείρηση που δεν μπορεί να ανταποκριθεί στις ληξιπρόθεσμες υποχρεώσεις της για μια περίοδο, είναι δυνατόν να τα καταφέρει μεσοπρόθεσμα, με (π.χ.) επαναρρύθμιση των χρεών της, αύξηση του μετοχικού κεφαλαίου της, πώληση και επαναμίσθωση παγίων κλπ. Στην περίπτωση της τυπικής χρεοκοπίας, η οριοθέτησή της είναι σε κάποιο βαθμό θέμα δικονομικό, αφού απαιτείται η συνδρομή των υπό του νόμου οριζόμενων προϋποθέσεων και η τήρηση της αντίστοιχης διαδικασίας για την τεκμηρίωση και δικαστική αποδοχή της.

Όμως, οι προϋποθέσεις που ορίζει ο νόμος δεν βασίζονται τόσο στη συλλογιστική ότι η κατάσταση της επιχείρησης είναι μη αντιστρέψιμη, όσο και τις γενικότερες διατάξεις περί άσκησης δικαιωμάτων αστικού χαρακτήρα.

Είναι προφανές ότι ο αναλυτής ενδιαφέρεται για την ουσιαστική χρεοκοπία και ιδιαίτερα για τον εντοπισμό των παραμέτρων που τη χαρακτηρίζουν. Έτσι, μέσω της παρακολούθησης των τιμών των παραμέτρων αυτών θα μπορεί να διαπιστώνει τα συμπτώματα χρεοκοπίας προ της διαμόρφωσης μη αντιστρέψιμης κατάστασης.

Με δεδομένο το πιο πάνω πλαίσιο, μια επιχείρηση βρίσκεται σε κατάσταση ουσιαστικής χρεοκοπίας, όταν δεν έχει τη δυνατότητα να ανταποκριθεί στις ληξιπρόθεσμες υποχρεώσεις της και, ακόμη, η ροή των κεφαλαίων της (περιλαμβανομένης της δυνατότητας δανεισμού ή έκδοσης νέων μετοχών) είναι τέτοια ώστε δεν επιτρέπει ουσιώδη βελτίωση στο μέλλον.

6.3 Μεθοδολογίες πρόγνωσης της χρεοκοπίας

Η πρόγνωση της χρεοκοπίας αποτελεί ένα προσφιλές θέμα για τους ακαδημαϊκούς αλλά και (κυρίως) για τον επαγγελματικό κόσμο. Πριν από αρκετά χρόνια, ο Beaver (1966), αναφέρθηκε αρκετά εκτεταμένα στο θέμα, αλλά ως πρώτη σημαντική συμβολή θα έπρεπε μάλλον να χαρακτηριστεί η δουλειά που παρουσίασε ο Altman λίγο αργότερα. Από τότε έχει

δημοσιευθεί μεγάλος αριθμός εργασιών, με αποτέλεσμα να εμπλουτισθεί σημαντικά η μεθοδολογία προσέγγισης του προβλήματος [Zmijewski (1984)].

Είναι αναγκαίο να τονισθεί, πάντως, ότι δεν έχει διαμορφωθεί κάποιο θεωρητικό πλαίσιο στο οποίο θα μπορούσαν να στηρίζονται οι επιλογές των ερευνητών κατά την πραγματοποίηση των σχετικών εμπειρικών ερευνών. Έτσι, οι εργασίες πουπραγματοποιούνται, αξιολογούνται από την πρακτική αποτελεσματικότητά τους [Jones (1987)].

Ένα κεντρικό πρόβλημα που αντιμετωπίζουν όλοι οι αναλυτές είναι ο ορισμός της χρεοκοπίας. Στη γενική της μορφή αυτή ορίζεται ως η αδυναμία της επιχείρησης να πραγματοποιήσει πληρωμές. Η αδυναμία αυτή οριοθετείται άλλοτε με νομικά κριτήρια και άλλοτε με οικονομικά κριτήρια (π.χ.de facto διακοπή πληρωμών).

Η πολυμορφία των ορισμών οδηγεί, όπως είναι φυσικό, σε υιοθέτηση δειγμάτων με αντίστοιχα διαφοροποιημένες ιδιότητες. Έτσι, οι σχετικές εμπειρικές έρευνες αποκτούν κάποια αυτοτέλεια, η οποία μάλιστα εντείνεται από τη μη χρησιμοποίηση των ίδιων ανεξάρτητων μεταβλητών [Beaver (1966), Altman (1968), Ohlson (1980)].

Συνήθως χρησιμοποιούνται δείκτες, οι οποίοι όμως είναι τόσο πολλοί σε αριθμό ώστε αρκετοί αναπτύσσουν μεθοδολογίες μείωσής τους σε ένα σχετικά μικρό αριθμό.

Στις λιγοστές εργασίες που αναζήτησαν άλλης μορφής μεταβλητές, περιλαμβάνονται εκείνες των Foster (1988), Altman (1982) & Rose, Andrews and Giroux (1982), οι οποίοι χρησιμοποίησαν μακροοικονομικά μεγέθη, καθώς και εκείνες των Altman, Haldeman and Narayanan (1977) και Diakoyiannis (1989), που αξιοποίησαν χρηματιστηριακά (κατά κύριο λόγο) στοιχεία.

Η παρατηρούμενη διασπορά τόσο στις μεθοδολογίες όσο και στις παραμέτρους, περιορίζει τη δυνατότητα αξιοποίησης του συσσωρευμένου ερευνητικού υλικού και εμποδίζει την κατάταξη των μεθοδολογιών πρόγνωσης της χρεοκοπίας, από άποψη αποτελεσματικότητας.

Για αυτό το λόγο δεν μπορεί να γίνει αξιολογική αναφορά σε αυτές παρά μόνο ομαδοποίησή τους με βάση τα γενικότερα χαρακτηριστικά τους. Στη βάση αυτή, διαμορφώνονται οι πιο κάτω γενικές κατηγορίες μεθοδολογιών:

- Μονομεταβλητές Αναλύσεις (Univariate Analysis)
- Πολυμεταβλητές Αναλύσεις
 - Ανάλυση Διαφοροποίησης (Discriminant Analysis)
 - Πιθανοτικά Υποδείγματα (Logit and Probit Analysis)
- Μη παραμετρικές μέθοδοι
 - Μέθοδος Recursive Partitioning
 - Σύγκριση των Μέσων Τιμών
 - Ανάλυση Επιβίωσης (Survival Analysis)

Μονομεταβλητή Ανάλυση

Χρησιμοποιήθηκε για πρώτη φορά από τον Beaver (1966), ο οποίος προσπάθησε να επιλέξει τον αποτελεσματικότερο από 30 χρηματοοικονομικούς δείκτες που περιλάμβανε στην εργασία του. Είναι σημαντικό να επισημανθεί ότι τελικά εντοπίσθηκε ο δείκτης ταμειακών ροών προς υποχρεώσεις (προβλεψιμότητα 87%), ακόμη δε ότι αποδείχθηκαν ιδιαίτερα αποτελεσματικοί και οι δείκτες αποδοτικότητας συνολικών κεφαλαίων και γενικής ρευστότητας.

Το πλεονέκτημα της εν λόγω μεθοδολογίας είναι η απλότητα και η ευκολία εφαρμογής της. Θεωρείται, πάντως ότι μπορεί να οδηγήσει σε αντιφατικά αποτελέσματα με τη χρήση διαφορετικών μεταβλητών, ενώ παραμένει και θεωρητικά “αδύναμη” αφού η πολυπλοκότητα μιας επιχείρησης δεν είναι δυνατόν να περιληφθεί έστω και προσεγγιστικά σε ένα μόνο δείκτη [Zavgren (1983)]. Για τους λόγους αυτούς η χρήση της έχει περιορισθεί σε μεμονωμένες εφαρμογές.

Πολυμεταβλητές Αναλύσεις

Η Ανάλυση Διαφοροποίησης (Discriminant Analysis) αποτελεί στατιστική τεχνική που επινοήθηκε το 1935 (Fisher) και αποτελεί την πρώτη πολυμεταβλητή τεχνική που χρησιμοποιήθηκε για πρόβλεψη της χρεοκοπίας [Altman (1986)]. Κατά τα τελευταία είκοσι χρόνια έχει εφαρμοστεί από μεγάλο αριθμό ερευνητών, όπως οι Altman, Halderman (1977), Dambolena & Khoury (1980).

Η Ανάλυση διαφοροποίησης διαχωρίζει δύο ή περισσότερα σύνολα με διαφορετικά χαρακτηριστικά (π.χ. προβληματικές και υγιείς επιχειρήσεις) και παράγει μια συνάρτηση διάκρισης (Discriminant Function), που αποτελεί γραμμικό συνδυασμό των χαρακτηριστικών αυτών (ανεξάρτητες μεταβλητές). Βάσει της συνάρτησης διάκρισης υπολογίζεται μια τιμή (Score), γνωστή ως z-score, για κάθε επιχείρηση.

Στη συνέχεια γίνεται διαχωρισμός των επιχειρήσεων (σε υγιείς και υπό χρεοκοπία) με βάση την τιμή αυτή και τις κριτικές τιμές της συνάρτησης διάκρισης (centroids). Οι τελευταίες αποτελούν τα σημεία πρόκρισης- απόρριψης (cut off points) που χρησιμοποιούνται από το υπόδειγμα για την κατάταξη των επιχειρήσεων.

Η αξιοπιστία πρόβλεψης του υποδείγματος που προκύπτει από την εφαρμογή αυτής της μεθόδου ελέγχεται με το “δείκτη σωστής ταξινόμησης”, ο οποίος προσδιορίζεται από το λόγο των σωστά ταξινομηθέντων επιχειρήσεων προς το σύνολο των επιχειρήσεων του δείγματος.

Η πιο πάνω μέθοδος έχει χρησιμοποιηθεί σε ευρεία κλίμακα στο παρελθόν με ικανοποιητικά αποτελέσματα, παρουσιάζει όμως στατιστικά προβλήματα, αφού δεν ικανοποιούνται συνήθως όλες οι υποθέσεις στις οποίες στηρίζεται και κυρίως η πολυμεταβλητή κανονικότητα (multivariate normality) και η ισότητα των μητρών συνδιακύμανσης των προς ταξινόμηση ομάδων δεδομένων.

Η συχνή παραβίαση των πιο πάνω υποθέσεων περιορίζει τη δυνατότητα αξιόπιστου υπολογισμού των πιθανοτήτων χρεοκοπίας μιας επιχείρησης, αλλά και τη δυνατότητα αξιολόγησης των μεταβλητών από πλευράς στατιστικής σημαντικότητας ακόμη και όταν

χρησιμοποιούνται ειδικοί στατιστικοί έλεγχοι, όπως το F- test σε συνδυασμό με το Wilks Lambda.

Ως αποτέλεσμα των πιο πάνω αδυναμιών, η Ανάλυση Διαφοροποίησης τείνει να χρησιμοποιείται σήμερα σε περιορισμένη κλίμακα, παρά το γεγονός ότι έχουν υποδειχθεί κάποιοι τρόποι (μερικής) αντιμετώπισης των στατιστικών προβλημάτων (λογαριθμική και τετραγωνική έκφραση των δεδομένων, απόλειψη ακραίων τιμών, εφαρμογή της quadratic discriminant analysis κλπ.).

Πολυμεταβλητές Αναλύσεις: Υποδείγματα Logit και Probit

Τα υποδείγματα αυτά είναι γνωστά και ως “υποδείγματα πιθανότητας υπό συνθήκη” ή “πολυμεταβλητές τεχνικές ποιοτικής επιλογής”. Συνήθως, η εξαρτημένη μεταβλητή είναι “ψευδομεταβλητή” (dummy variable), για να διευκολύνεται η χρήση ποιοτικών δεδομένων.

Πρόκειται για πιθανοτικά υποδείγματα τα οποία επιτρέπουν την κατάταξη επιχειρήσεων με βάση την πιθανότητα χρεοκοπίας τους, λαμβάνοντας υπ’ όψη τα οικονομικά τους δεδομένα. Και τα δύο υποδείγματα κάνουν χρήση της σωρευτικής συνάρτησης πιθανότητας και προσδιορίζονται με τη μεθοδολογία maximum likelihood. Η διαφορά των δύο μεθόδων έγκειται στο είδος της σωρευτικής καμπύλης πιθανότητας (Cumulative probability function), που στην περίπτωση της Logit είναι η Λογιστική Καμπύλη (Logistical Cumulative Function), ενώ στην περίπτωση της Probit, η κανονική καμπύλη (Normal Cumulative Function).

Οι καμπύλες των δυο υποδειγμάτων είναι σχεδόν ίδιες (διαφέρουν μόνο στις ακραίες περιοχές τους) και θεωρούνται ως κατάλληλες να περιγράψουν την πιθανότητα χρεοκοπίας λόγω του σχήματός τους (s-shape). Ειδικότερα, οι πολύ μικρές κλίσεις στην αρχή και στο τέλος της καμπύλης υποδηλώνουν ότι ακόμη και οι μεγάλες τιμές των μεταβλητών δεν αυξάνουν σημαντικά την πιθανότητα χρεοκοπίας, ενώ το αντίθετο συμβαίνει στην περίπτωση της περιοχής περί το μέσο της καμπύλης.

Τόσο στην περίπτωση της μεθοδολογίας Probit όσο και σε εκείνες της Logit χρησιμοποιούνται ως ανεξάρτητες μεταβλητές χρηματοοικονομικοί δείκτες για τις δυο κατηγορίες επιχειρήσεων (χρεοκοπημένες και υγιείς) και διαμορφώνεται μια συνάρτηση που περιλαμβάνει τις ίδιες ανεξάρτητες μεταβλητές, αλλά με συντελεστές που εκτιμώνται με δεδομένη διαδικασία. Στη συνέχεια γίνεται στάθμιση των συντελεστών αυτών με τρόπο που εξασφαλίζει τη μεγιστοποίηση της ενωμένης πιθανότητας (joint probability) χρεοκοπίας για τις αποτυχούσες επιχειρήσεις και μη χρεοκοπίας για τις υγιείς.

Η τελική συνάρτηση χρησιμοποιείται για τον προσδιορισμό μιας τιμής κατά επιχείρηση του δείγματος, η οποία συγκρίνεται με κάποιο σημείο αναφοράς, που αποτελεί το διαχωριστικό σημείο μεταξύ υγιών και χρεοκοπημένων επιχειρήσεων.

Οι πιο πάνω μέθοδοι (Logit and Probit) υπερέχουν θεωρητικά της Ανάλυσης Διαφοροποίησης και επιτρέπουν τη σχετικά εύκολη αξιολόγηση της αποτελεσματικότητάς τους σε κάθε εφαρμογή. Αυτός είναι, άλλωστε, ο λόγος της διαρκώς διευρυνόμενης χρήσης τους σε εμπειρικές έρευνες πρόγνωσης της χρεοκοπίας [Ohlson (1980)].

Θα πρέπει τέλος να σημειωθεί ότι, πέρα από τα αρχικά υποδείγματα Probit και Logit, έχουν αναπτυχθεί και κάποιες βελτιωμένες παραλλαγές τους, όπως η multi-logit Analysis.

Μη παραμετρικές μέθοδοι : Μέθοδος Recursive Partitioning

Πρόκειται για πρόσφατα ανεπτυγμένη, μη παραμετρική μέθοδο, που δεν στηρίζεται σε κάποιες υποθέσεις ως προς τις ιδιότητες των κατανομών των χρησιμοποιούμενων δεδομένων, γι αυτό και δεν πάσχει από θεωρητικά προβλήματα όπως οι άλλες πολυμεταβλητές μέθοδοι.

Η μέθοδος αυτή λειτουργεί με τη διαδικασία των βημάτων (stepwise procedure) και επιχειρεί τη διάκριση μεταξύ διαφορετικών ομάδων επιχειρήσεων μέσω της εισαγωγής μιας μόνο μεταβλητής σε κάθε βήμα της διαδικασίας, από την οποία προκύπτει η τελική μήτρα κατάταξης των ομάδων των επιχειρήσεων.

Παρά το πλεονέκτημα της έλλειψης προβλημάτων υποθέσεων, η Recursive Partitioning δεν δίνει πληροφορίες σε όρους πιθανότητας (για την χρεοκοπία μιας επιχείρησης), ούτε επιτρέπει την αξιολόγηση από πλευράς στατιστικής σημαντικότητας των μεταβλητών του μοντέλου.

Παρουσιάζει όμως μεγαλύτερη ακρίβεια στην πρόβλεψη της χρεοκοπίας και για το λόγο αυτό έχει προτιμηθεί από σημαντικό αριθμό μελετητών.

Μη παραμετρικές μέθοδοι: Σύγκριση των Μέσων Τιμών

Με βάση τη μέθοδο αυτή, υπολογίζονται οι μέσες τιμές των παραμέτρων που έχουν επιλεγεί για τις χρεοκοπημένες και μη χρεοκοπημένες επιχειρήσεις του δείγματος και εξετάζεται (π.χ. με τη μέθοδο analysis of variance) η σημαντικότητα των διαφορών τους. Αν οι παρατηρούμενες διαφορές είναι αξιόλογες και συστηματικές, επιτρέπουν κατατάξεις στη βάση των τιμών των χρησιμοποιούμενων μεταβλητών.

Μη παραμετρικές μέθοδοι: Ανάλυση Επιβίωσης

Πρόκειται για μεθοδολογία που αναπτύχθηκε στους τομείς της ιατρικής βιολογίας. Πρόσφατα άρχισε να χρησιμοποιείται σε εμπειρικές έρευνες σχετικές με την πρόγνωση χρεοκοπίας, όμως δεν υπάρχουν ακόμη ικανοποιητικές ενδείξεις για την αποτελεσματικότητά της.

6.4 Τα Στοιχεία των Επιχειρήσεων

Ως επιχείρηση εννοούμε την οργανωμένη δραστηριότητα η οποία έχει ως στόχο την παραγωγή αγαθών για την κάλυψη αναγκών. Στα πλαίσια αυτής της δραστηριότητάς της, η επιχείρηση προσπαθεί να εφαρμόσει τη θεμελιώδη οικονομική αρχή, δηλαδή να επιτύχει το μέγιστο δυνατό αποτέλεσμα με την ελάχιστη θυσία.

Τελικά, πάντως, η επιτυχία (ή αποτυχία) της προσπάθειας της κρίνεται από την Αγορά (όταν βέβαια λειτουργεί σε συνθήκες ελεύθερου ανταγωνισμού), μέσω του μηχανισμού των τιμών.

Όσο σημαντικότερο το αποτέλεσμα της δράσης της, τόσο υψηλότερες οι τιμές και τόσο μεγαλύτερο το άνοιγμα ανάμεσα στα έσοδα και το κόστος.

Σε μια ανταγωνιστική Αγορά οι τιμές των αγαθών διαμορφώνονται ανεξάρτητα από το κόστος παραγωγής των επιμέρους μονάδων.

Για παράδειγμα, έστω ότι υπάρχει επαρκής προσφορά του προϊόντος Π στην τιμή των 100 €, ακόμη δε ότι η επιχείρηση Ε παράγει το ίδιο προϊόν με κόστος 120 € .

Αν οι ποιοτικές προδιαγραφές του προϊόντος της Ε και οι όροι πώλησης του δεν διαφέρουν από εκείνα των ανταγωνιστών, τότε η επιχείρηση αυτή δεν μπορεί να πετύχει τιμή ανώτερη των 100€. Σε μια τέτοια περίπτωση, η Ε αναλώνει παραγωγικούς συντελεστές αξίας 120 €, ενώ δημιουργεί αγαθά αξίας 100€, λόγω αναποτελεσματικής οργάνωσης της δραστηριότητας της. Αφαιρεί, δηλαδή, από τι κοινωνικό σύνολο πόρους 20 € ανά παραγόμενη μονάδα.

Με τη ζημία αυτή επιβαρύνονται οι χρηματοδότες της επιχείρησης (μέτοχοι, δανειστές, προμηθευτές κλπ), στο βαθμό που δεν κατορθώνουν να εισπράξουν τα κεφάλαια που χορήγησαν (και τις αποδόσεις των κεφαλαίων αυτών). Οι χρηματοδότες, επομένως, για να περιορίσουν τον κίνδυνο απωλειών, έχουν ανάγκη να εκτιμήσουν τις προοπτικές της μονάδας την οποία εφοδιάζουν με κεφάλαια, πριν αποφασίσουν τη χρηματοδότησή της.

Βέβαια, η σύγχρονη επιχείρηση αποτελεί ένα πολυσύνθετο πλέγμα συμφερόντων, γι' αυτό υπάρχουν και άλλα μέρη τα οποία επιθυμούν να γνωρίζουν τη μελλοντική πορεία της, όπως π.χ. οι εργαζόμενοι σε αυτή (για να εκτιμούν το χρονικό ορίζοντα της απασχόλησης τους και το επίπεδο των αμοιβών τους) το management (για να προγραμματίσει τη δράση της μονάδας, ακόμη δε να εκτιμήσει τις επιπτώσεις στην αμοιβή και τη φήμη του), το Κράτος (να προστατέψει τους επενδυτές, πιστωτές κλπ και να εισπράξει φόρους) κ.ο.κ.

Η δυνατότητα, επομένως, πραγματοποίησης εκτιμήσεων σχετικά με τις προοπτικές της επιχείρησης, αποτελεί κρίσιμη παράμετρο για όλες τις πιο πάνω ομάδες ενδιαφερομένων.

Οι Οικονομικές Καταστάσεις της Επιχείρησης

Η ροή πόρων (κεφάλαια, εργασία, φυσικοί πόροι), η αξιοποίηση τους, η πώληση αγαθών, ο προσδιορισμός του επιχειρηματικού αποτελέσματος και η διανομή του (ή ο καταλογισμός του) καταγράφονται στα βιβλία της επιχείρησης και αποτελούν πολύτιμα στοιχεία για κάθε ενδιαφερόμενο. Ιδιαίτερη αξία αποκτούν οι ανακεφαλαιωτικές καταστάσεις όλων των οικονομικών δεδομένων και μεταβολών, διότι επιτρέπουν την ταχύτερη και ασφαλέστερη αξιολόγηση των επιχειρήσεων μέσω της τυποποίησης και ποσοτικοποίησης των οικονομικών τους στοιχείων. Οι γνωστότερες από τις καταστάσεις αυτές είναι οι εξής:

- Ισολογισμός
- Κατάσταση αποτελεσμάτων
- Κατάσταση διανομής κερδών

Εκτός, από τις πιο πάνω καταστάσεις υπάρχουν και άλλες, λιγότερο αναλυτικές, όπως οι τριμηνιαίες λογιστικές καταστάσεις (3μήνου, 6μήνου και 9μήνου), τα μηνιαία ισοζύγια κλπ. Τέλος, σημαντικές πληροφορίες περιέχουν οι εκθέσεις των ελεγκτών, οι ετήσιες εκθέσεις προς τους μετόχους, οι εταιρικές ανακοινώσεις, τα ενημερωτικά δελτία που καταρτίζονται έπ' ευκαιρία της αύξησης του μετοχικού κεφαλαίου κλπ.

Ο Ισολογισμός

Ο ισολογισμός μιας επιχείρησης συνοψίζει την οικονομική της κατάσταση σε μια δεδομένη χρονική στιγμή, γι' αυτό και η εικόνα που παρέχει είναι στατική. Ειδικότερα, παρουσιάζονται τα περιουσιακά στοιχεία της επιχείρησης ομαδοποιημένα (με κριτήριο το χρόνο παραμονής τους σε αυτή) σε πάγια, κυκλοφορούντα και διαθέσιμα και οι πηγές των κεφαλαίων της (ίδια κεφάλαια, μακροπρόθεσμες και βραχυπρόθεσμες υποχρεώσεις). Δηλαδή, ο ισολογισμός δείχνει ποια περιουσιακά στοιχεία έχουν αποκτηθεί και με τι

κεφάλαια έχει χρηματοδοτηθεί η απόκτησή τους. Όσο αφορά τις επιμέρους κατηγορίες μεγεθών, παρατηρούνται τα πιο κάτω:

Τα Πάγια περιλαμβάνουν μέσα που χρησιμοποιεί η επιχείρηση για την παραγωγή των προϊόντων και υπηρεσιών της. Η απόκτηση των μέσων αυτών αποτελεί “επένδυση”, η δε αξία τους επανεισρέει στην επιχείρηση κατά τη διάρκεια της ωφέλιμης ζωής τους.

Η απόκτηση πάγιων μεγαλύτερης δυναμικότητας από εκείνη που εξυπηρετεί τους μακροπρόθεσμους στόχους της μονάδας αποτελεί “υπερεπένδυση” και έχει ως αποτέλεσμα την σπατάλη κεφαλαίων και την άσκοπη διόγκωση των σταθερών δαπανών και των χρηματοοικονομικών εξόδων (ή την απώλεια χρηματοοικονομικών εσόδων).

Από την άλλη πλευρά, η ελλιπής δυναμικότητα δεν επιτρέπει την εκμετάλλευση των ευκαιριών της Αγοράς ή την αποτελεσματική χρησιμοποίηση των παραγωγικών πόρων.

Κρίσιμη παράμετρο κατά την εξέταση των πάγιων στοιχείων μιας επιχείρησης αποτελεί το τεχνολογικό τους επίπεδο. Όσο πιο σύγχρονη η τεχνολογία που ενσωματώνεται στα παραγωγικά μέσα της μονάδας, τόσο μεγαλύτερη η ευχέρεια μείωσης του κόστους και βελτίωσης της ποιότητας των προϊόντων της.

Τέλος, θα πρέπει να τονιστεί ότι υπάρχουν δραστηριότητες που χαρακτηρίζονται ως “εντάσεως κεφαλαίου”, δηλαδή προϋποθέτουν τη διάθεση μεγάλου μέρους των κεφαλαίων της επιχείρησης για απόκτηση πάγιων στοιχείων. Π.χ. αυτό συμβαίνει με τις τσιμεντοβιομηχανίες, τις ναυπηγικές επιχειρήσεις, τις αυτοκινητοβιομηχανίες κλπ, ενώ αντίστροφο ισχύει για τις επιχειρήσεις παροχής υπηρεσιών.

Το Κυκλοφορούν Ενεργητικό περιλαμβάνει τα στοιχεία εκείνα που παραμένουν για περιορισμένο χρονικό διάστημα στην επιχείρηση και στη συνέχεια ρευστοποιούνται.

Για παράδειγμα, τα παραγόμενα προϊόντα αποθεματοποιούνται μέχρι τη διοχέτευσή τους στην Αγορά, τα αποθέματα πρώτων και βοηθητικών υλών παραμένουν στις αποθήκες μέχρι να χρησιμοποιηθούν στην παραγωγική διαδικασία, οι απαιτήσεις ρευστοποιούνται σε ρυθμό αντίστοιχο των εισπράξεων από τους πελάτες και χρεώστες γενικότερα.

Ο προσδιορισμός του ύψους του κυκλοφορούντος ενεργητικού αποτελεί στην ουσία ένα πρόβλημα αριστοποίησης. Πιο συγκεκριμένα, οι μεγάλες ποσότητες αποθεμάτων έτοιμων προϊόντων και η γενναιόδωρη παροχή πιστώσεων στην πελατεία διευκολύνουν την επέκταση των πωλήσεων.

Επίσης, η διατήρηση υψηλών αποθεμάτων πρώτων και βοηθητικών υλών εξυπηρετεί τον άνετο ανεφοδιασμό της παραγωγής. Όμως, από την άλλη πλευρά, προκαλούν αύξηση των χρηματοοικονομικών δαπανών (λόγω της διόγκωσης των απασχολημένων κεφαλαίων) και των σταθερών εξόδων (κόστος αποθήκευσης, φύλαξης, διαχείρισης κλπ) και αυξάνουν τον κίνδυνο απωλειών από επισφάλειες (στην περίπτωση των απαιτήσεων) ή μείωσης της αξίας τους (στην περίπτωση των αποθεμάτων).

Τα Διαθέσιμα είναι περιουσιακά στοιχεία, με μορφή μετρητών ή τοποθετήσεων υψηλής ρευστότητας (π.χ. καταθέσεις όψεως, *repos*, καταθέσεις μιας ή λίγων ημερών κλπ) τα οποία χρησιμοποιούνται στα πλαίσια των τρεχουσών συναλλακτικών σχέσεων της επιχείρησης. Από τη φύση τους, τα διαθέσιμα μπορούν να αξιοποιηθούν μόνο σε εξαιρετικά βραχυπρόθεσμες τοποθετήσεις, με αποτέλεσμα να περιορίζεται αντίστοιχα η αποδοτικότητά τους. Έτσι, καταβάλλεται προσπάθεια ελαχιστοποίησης του μεγέθους τους μέσω προγραμματισμού των εισπράξεων και των πληρωμών (π.χ. κατάρτιση ταμειακών προγραμμάτων).

Τα Ίδια Κεφάλαια περιλαμβάνουν το μετοχικό κεφάλαιο και τα συσσωρευμένα κεφάλαια από παρακράτηση κερδών (= αποθεματικά κερδών) και από πώληση μετοχών σε τιμή ανώτερη της ονομαστικής τους αξίας (= αποθεματικό υπέρ το άρτιο), καθώς και την

υπεραξία των στοιχείων του ενεργητικού η οποία προκύπτει από αναπροσαρμογή της αξίας τους.

Τα ίδια κεφάλαια δεν είναι επιστρεπτά, ούτε δημιουργούν οποιεσδήποτε συμβατικές δεσμεύσεις σχετικά με την εξυπηρέτησή τους. Έτσι, συμβάλλουν στη βελτίωση της ρευστότητας της επιχείρησης και της επιτρέπουν να καλύπτει ενδεχόμενα αρνητικά ανοίγματα μεταξύ εσόδων και εξόδων, στα πλαίσια μιας ή περισσοτέρων χρήσεων.

Οι Υποχρεώσεις, διακρίνονται συνήθως σε δύο γενικές κατηγορίες:

- Τις Μακροπρόθεσμες
- Τις Βραχυπρόθεσμες

Οι Μακροπρόθεσμες Υποχρεώσεις περιλαμβάνουν κατά κύριο λόγο δάνεια τα οποία η επιχείρηση έχει συνάψει είτε με συγκεκριμένους δανειστές, είτε με πιστωτικά ιδρύματα, είτε με το ανώνυμο κοινό (= ομολογιακά δάνεια).

Η ευχέρεια επιστροφής των κεφαλαίων αυτής της κατηγορίας μετά από ένα μεγάλο χρονικό διάστημα (συντά δέκα ή περισσοτέρων ετών) επιτρέπει στην επιχείρηση να τα χρησιμοποιήσει για αγορά παγίων στοιχείων, γενικότερα δε επιδρά ευνοϊκά στη ρευστότητά της.

Ιδιαίτερη προσοχή απαιτείται όταν η επιχείρηση δανείζεται μακροπρόθεσμα με σταθερό επιτόκιο, διότι είναι πιθανό να ακολουθήσει πτώση των επιτοκίων προ της λήξεως του δανείου, με αποτέλεσμα την επιβάρυνσή της σε μικρό ή μεγάλο βαθμό, ανάλογα με την έκταση της μείωσης, τη διάρκεια της δανειακής σύμβασης και το ύψος του δανείου. Βέβαια, δεν αποκλείεται και η αντίστροφη περίπτωση.

Για να μην αφεθεί η κατάσταση στην τύχη, είναι αναγκαίο να αξιολογούνται οι παράμετροι του οικονομικού περιβάλλοντος που επηρεάζουν τα επιτόκια, για ένα βάθος χρόνου ανάλογο με τη διάρκεια της δανειακής σύμβασης.

Οι Βραχυπρόθεσμες Υποχρεώσεις περιλαμβάνουν τα τραπεζικά δάνεια μικρής διάρκειας και τις υποχρεώσεις προς τους προμηθευτές (από αγορές με πίστωση), το Δημόσιο (=πληρωτέοι φόροι), τα συνταξιοδοτικά ταμεία και ευρύτερα τους οργανισμούς κοινωνικής ασφάλιση (=κρατήσεις και εισφορές), τους εργαζόμενους (=δεδουλευμένες αποδοχές που δεν έχουν καταβληθεί ακόμη), τους μετόχους (=πληρωτέα μερίσματα) και γενικά κάθε πίστωση που παρέχεται στην επιχείρηση.

Οι δανειακές βραχυπρόθεσμες υποχρεώσεις μπορούν να δημιουργήσουν σοβαρά προβλήματα στην επιχείρηση όταν δεν υπάρχει δυνατότητα εξυπηρέτησής τους. Γι' αυτό θα πρέπει να χρησιμοποιούνται, κυρίως, για την κάλυψη προσωρινών ταμειακών ελλειμμάτων και όχι για επενδυτικούς σκοπούς.

Οι λοιπές βραχυπρόθεσμες υποχρεώσεις, δημιουργούν μικρότερους κινδύνους από άποψη ρευστότητας, θα πρέπει όμως να φροντίζει η επιχείρηση να διατηρεί επαρκή στοιχεία του κυκλοφορούντος και διαθέσιμου ενεργητικού ώστε να μπορεί να τις εξυπηρετεί με ευχέρεια. Θα πρέπει να τονισθεί ότι τα περιουσιακά στοιχεία και οι υποχρεώσεις είναι δυνατό να εμφανισθούν στον ισολογισμό με διάφορους βαθμούς ανάλυσης, που κυμαίνονται από την πολύ συνοπτική ως την ιδιαίτερα αναλυτική παρουσίαση.

Στις χώρες που υπάρχει και τηρείται κάποιο λογιστικό σχέδιο (όπως στην Ελλάδα τα τελευταία χρόνια), παρατηρείται μια σχετική ομοιομορφία στο βαθμό ανάλυσης και στην ονοματολογία των λογαριασμών, που περιορίζει σημαντικά την ανάγκη αποσαφήνισης και ερμηνείας των επιμέρους μεγεθών.

Οι Καταστάσεις Αποτελεσμάτων Χρήσης και Διανομής Κερδών

Η κατάσταση αποτελεσμάτων παρουσιάζει τα συνολικά έσοδα και έξοδα που πραγματοποίησε η επιχείρηση κατά τη διάρκεια της χρήσης. Επομένως, τα μεγέθη της δεν είναι στατικά ή συγκυριακά αλλά εκφράζουν, σε όρους αξίας, τη δραστηριότητά της σε ολόκληρη τη χρήση.

Όπως και στην περίπτωση του ισολογισμού, ο βαθμός ανάλυσης της πιο πάνω κατάστασης εξαρτάται από το βαθμό ανάλυσης των τηρούμενων λογαριασμών, τις επιλογές του management για την έκταση της ενημέρωσης που παρέχει και, από τους περιορισμούς της ισχύουσας νομοθεσίας.

Σε κάθε περίπτωση, πάντως, στην κατάσταση αποτελεσμάτων περιλαμβάνονται τα μεγέθη των εσόδων από πωλήσεις, του κόστους πωληθέντων, των λειτουργικών δαπανών, του φόρου εισοδήματος και των καθαρών κερδών. Δηλαδή, η κατάσταση αυτή περιέχει κρίσιμα στοιχεία, γι' αυτό βρίσκεται στο επίκεντρο του ενδιαφέροντος του αναλυτή, ο οποίος από τη μελέτη των μεγεθών της και το συσχετισμό τους με εκείνα του ισολογισμού, παράγει σημαντικές πληροφορίες σχετικά με τις δυνατότητες και αδυναμίες της επιχείρησης.

Η κατάσταση διανομής κερδών εμφανίζει τον τρόπο κατανομής του καθαρού επιχειρησιακού πλεονάσματος μεταξύ, κυρίως, της επιχείρησης και των μετόχων της. Τα μεγέθη της αποκαλύπτουν κατά βάση τη μερισματική πολιτική, αλλά και τον βαθμό αυτοχρηματοδότησης της επιχείρησης.

Έτσι, οδηγούν τον αναλυτή σε χρήσιμα συμπεράσματα, ιδιαίτερα όταν εξετάζεται η διαχρονική εξέλιξη τους καθώς και η σχέση τους με άλλα εσωτερικά μεγέθη (όπως τα έσοδα, τα απασχολημένα κεφαλαία, η χρηματιστηριακή αξία της επιχείρησης κλπ) ή με τα αντίστοιχα μεγέθη άλλων επιχειρήσεων.

6.5 Εφαρμογές στατιστικών μέτρων στις επιχειρήσεις

Σε αυτή την ενότητα θα συνοψίσουμε όλα οσα αναφέραμε για τα μέτρα θέσης και διασποράς σε εφαρμογές για να δείξουμε τη σημαντικότητα των μέτρων αυτών ως προς μια επιχείρηση.

ΕΦΑΡΜΟΓΗ 1

Έστω ότι η επιχείρηση Νικολάου ΑΕ λειτουργεί στο χώρο του εμπορίου πουλώντας ηλεκτρικές συσκευές. Ο νέος οικονομικός διευθυντής ζήτησε ένα στατιστικό έλεγχο για την ποσότητα συσκευών που πωλήθηκαν για το έτος 2013. Επίσης ζήτησε έναν συγκριτικό έλεγχο με το 2012 δεδομένου ότι $\bar{x} = 21$, $s = 18$ και $R=80$. Τέλος δεδομένου ότι η εταιρία είχε και κάποιες επιστροφές προϊόντων ζήτησε και ένα στατιστικό έλεγχο των τελευταίων 10 μηνών θέλοντας να εξακριβώσει αν οι επιστροφές των εμπορευμάτων γίνονται σε όρια επιτρεπτά. Οι πωλήσεις σε τεμάχια των 12 μηνών του έτους 2013 ήταν:

ΙΑΝΟΥΑΡΙΟΣ	20
ΦΕΦΡΟΥΑΡΙΟΣ	25
ΜΑΡΤΙΟΣ	27
ΑΠΡΙΛΙΟΣ	30
ΜΑΙΟΣ	56
ΙΟΥΝΙΟΣ	31
ΙΟΥΛΙΟΣ	48
ΑΥΓΟΥΣΤΟΣ	90
ΣΕΠΤΕΜΒΡΙΟΣ	14
ΟΚΤΩΜΒΡΙΟΣ	57
ΝΟΕΜΒΡΙΟΣ	52
ΔΕΚΕΜΒΡΙΟΣ	14

Πίνακας 8: Πίνακας πωλήσεων σε τεμάχια 2013

και ο πίνακας αξία πωλήσεων (σε χιλιάδες ευρώ) και αξία επιστροφών:

<u>ΑΞΙΑ</u> <u>ΠΩΛΗΣΕΩΝ(X)</u>	20	30	40	40	50	50	60	70	80	90
<u>ΑΞΙΑ</u> <u>ΕΠΙΣΤΡΩΦΩΝ(Y)</u>	1	3	3	4	5	6	6	8	9	10

Πίνακας 9: Πίνακας αξίες πωλήσεων και επιστροφών

Ο οικονομικός διευθυντής συντάσσει αρχικά τον παρακάτω πίνακα και υπολογίζει τη μέση τιμή και την τυπική απόκλιση:

i	x_i	v_i	x_i^2	$x_i v_i$	$v_i x_i^2$
1	14	2	196	28	392
2	20	1	400	20	400
3	25	1	625	25	625
4	27	1	729	27	729
5	30	1	900	30	900
6	31	1	961	31	961
7	48	1	2304	48	2304
8	52	1	2704	52	2704
9	56	1	3136	56	3136
10	57	1	3249	57	3249
11	90	1	8100	90	8100
ΣΥΝΟΛΟ		12		464	23500

Πίνακας 10: Πίνακας στατιστικών στοιχείων εφαρμογής 1

Η μέση τιμή θα είναι :

$$\bar{x} = \frac{1}{12} \sum_{i=1}^{11} v_i x_i = \frac{1}{12} 464 = 38.6$$

Βλέποντας τον πίνακα μπορούμε να παρατηρήσουμε ότι η επικρατούσα τιμή θα είναι:

$$M_0 = 14$$

Ταξινομώντας το δείγμα σε αύξουσα σειρά 14,14,20,25,27,30,31,48,52,56,57,90 η διάμεσος θα είναι:

$$M = \frac{30+31}{2} = 30,5$$

Το εύρος ισούται με τη διαφορά μεγαλύτερη παρατήρηση - μικρότερη παρατήρηση, δηλαδή :

$$R = x_{max} - x_{min} = 90 - 14 = 76$$

Το ενδοτεταρτομοριακό εύρος ισούται με :

$$H = Q_3 - Q_1$$

Άρα πρέπει πρώτα να βρούμε το πρώτο τεταρτημότιο και το τρίτο. Έχουμε λοιπόν:

$$\frac{1(12+1)}{4} = 3,25 \quad , \quad \frac{3(12+1)}{4} = 9,75.$$

Άρα,

$$Q_1 = 20 + 0,25 (25 - 20) = 21,25$$

και

$$Q_3 = 52 + 0,75 (56 - 52) = 55$$

Βλέπουμε λοιπόν ότι κατά 25% έχουν πωληθεί περίπου 22 προϊόντα ενώ το 50% είναι περίπου 32 προϊόντα. Τέλος υπάρχει ένα ποσοστό ακόμα της τάξης 25% που δείχνει ότι έχουν πωληθεί 55 προϊόντα. Έτσι έχουμε: $H = Q_3 - Q_1 = 55 - 21,25 = 33,75$

Υπολογίζουμε την τυπική απόκλιση. Για τον υπολογισμό της πρέπει πρώτα να υπολογίσουμε τη διασπορά.

Έτσι έχουμε:

$$s^2 = \frac{1}{12-1} \left[\sum_{i=1}^{11} v_i x_i^2 - \frac{1}{12} \left(\sum_{i=1}^{11} v_i x_i \right)^2 \right] = 505,3.$$

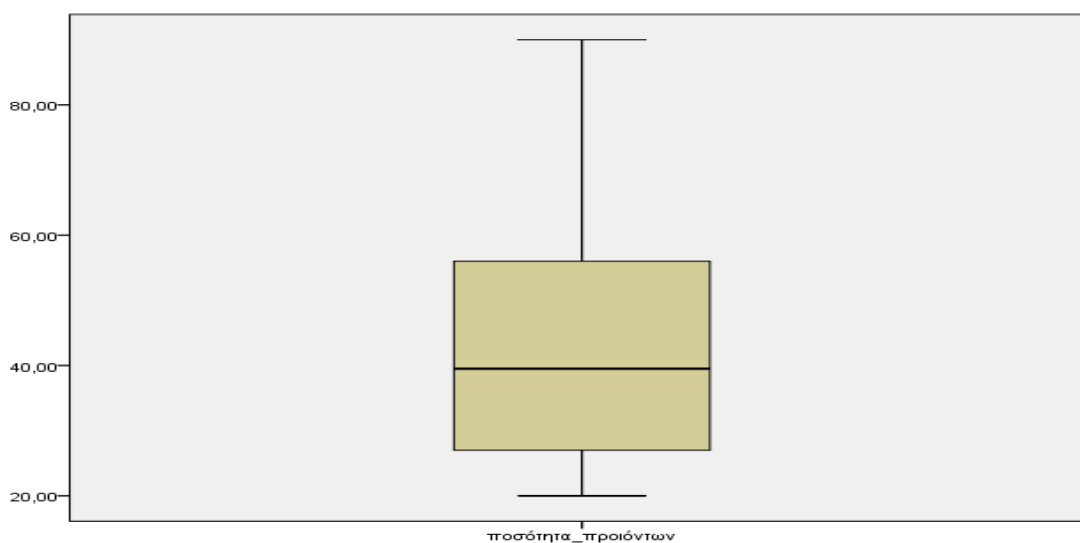
Άρα η διασπορά είναι:

$$s = \sqrt{s^2} = \sqrt{505,3} = 22,47$$

Συνεχίζοντας θα υπολογίσουμε το συντελεστή μεταβλητότητας και έχουμε:

$$CV = \frac{s}{\bar{x}} = \frac{22,47}{38,6} = 0,58 \text{ ή } 58,5 \%$$

Τέλος παρουσιάζουμε το θηκόγραμμα στο οποίο βλέπουμε ότι δεν υπάρχουν ακραίες τιμές.



Εικόνα 11: Θηκόγραμμα εφαρμογής 1

Συμπερασματικά έχοντας τη μέση τιμή του προηγούμενου έτους $\bar{x}=21$, την τυπική απόκλιση $s=18$ και το εύρος $R=80$ φτιάχνουμε τον παρακάτω πίνακα:

Στατιστικά Στοιχεία	Για το έτος 2012	Για το έτος 2013	Μεταβολή των χρήσεων 2012-2013
Μέση τιμή \bar{x}	21	38,6	αύξηση
Τυπική απόκλιση s	18	22,6	αύξηση
Συντελεστής μεταβληκότητας CV	0,85 ή 85%	0,585 ή 58,5%	μείωση
Ευρος μεταβολής R	80	76	μείωση

Πίνακας 11: Σύγκριση στατιστικών μέτρων 2013-2014 εφαρμογής 1

Μπορούμε να πούμε λοιπόν ότι η εταιρία από τη μέχρι τώρα μελέτη βελτίωσε την οικονομική της θέση, διότι

α) έχουμε σημαντική αύξηση της μέσης τιμής των πωλήσεων από 21 σε 38,6 φέτος ($38,6-21=17,6$), ποσοστό αύξησης περίπου 18%.

β) Η διασπορά των πωλήσεων φέτος παρουσιάζει καλύτερη εικόνα, δηλαδή έχουμε καλύτερη ομοιογένεια ως προς τις πωλήσεις, αφού ο φετινός $CV = 0,585 < CV = 0,85$

Ακόμη, έχουμε φέτος ένα εύρος μεταβολής ίσο με 76, το οποίο είναι αρκετά μικρότερο από το περσινό, που ήταν 80.

Τελειώνοντας για να δουμε ένα στατιστικό έλεγχο ανάμεσα στην αξία επιστροφών και στην αξία πωλήσεων υπολογίσουμε το συντελεστή γραμμικής συσχέτισης των μεταβλητών «αξία πωλήσεων» και «αξία επιστροφών» και συμπληρώνουμε τον πίνακα των δεδομένων.

<i>ΑΞΙΑ ΠΩΛΗΣΕΩΝ(X)</i>	<i>ΑΞΙΑ ΕΠΙΣΤΡΟΦΩΝ(Y)</i>	<i>xy</i>	<i>x²</i>	<i>y²</i>
20	1	20	400	1
30	3	90	900	9
40	3	120	1600	9
40	4	160	1600	16
50	5	250	2500	25
50	6	300	2500	36
60	6	360	360	36
70	8	560	490	64
80	9	720	6400	81
90	10	900	8100	100
ΣΥΝ. 530	ΣΥΝ. 55	ΣΥΝ.3480	ΣΥΝ.32500	377

Πίνακας 12: Πίνακας αξίες πωλήσεων - αξίες επιστροφών εφαρμογής 1

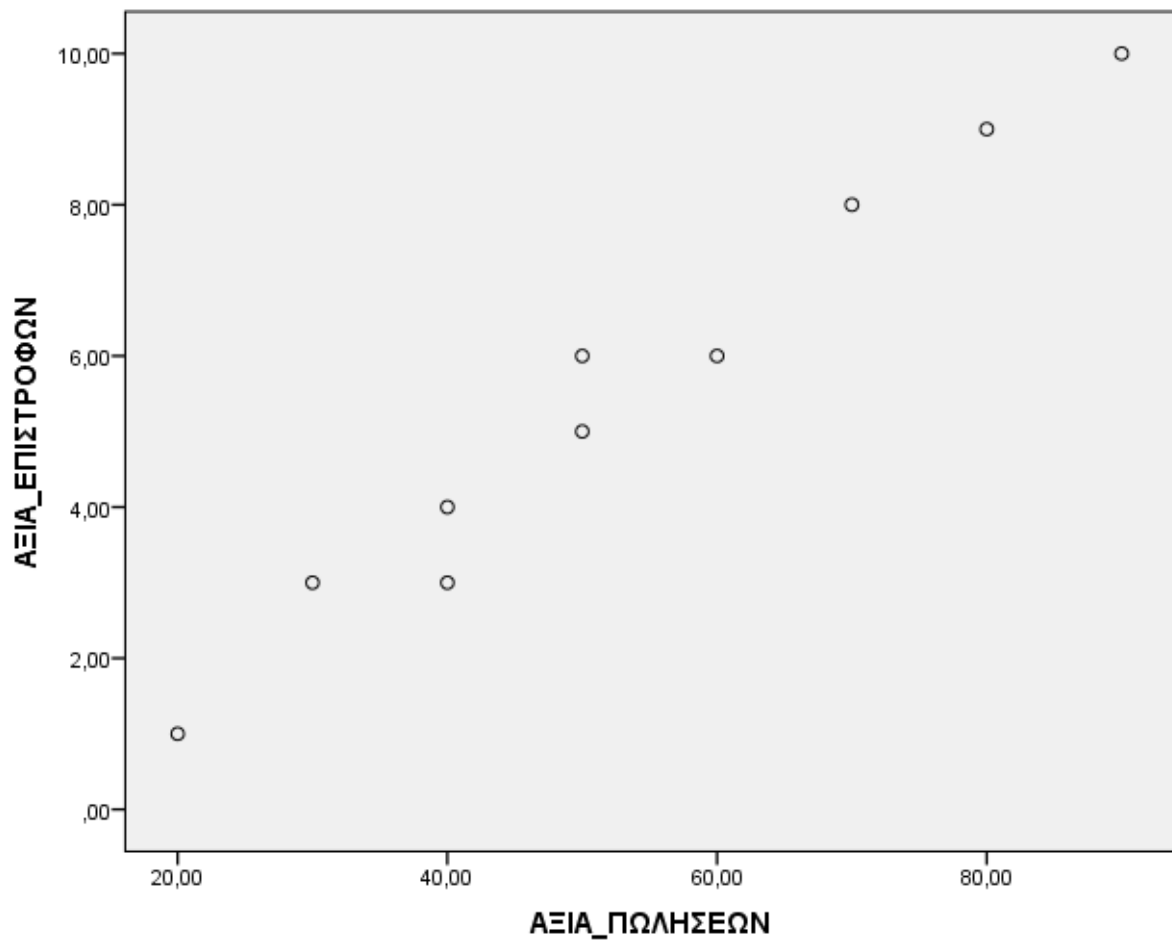
Αντικαθιστούμε στον τύπο υπολογισμού του συντελεστή συσχέτισης και έχουμε:

$$r = \frac{10 \sum_{i=1}^{10} x_i y_i - (\sum_{i=1}^{10} x_i)(\sum_{i=1}^{10} y_i)}{\sqrt{10 \sum_{i=1}^{10} x_i^2 - (\sum_{i=1}^{10} x_i)^2} \sqrt{10 \sum_{i=1}^{10} y_i^2 - (\sum_{i=1}^{10} y_i)^2}}$$

$$= \frac{10 \cdot 3480 - 530 \cdot 55}{\sqrt{10 \cdot 32500 - (530)^2} \sqrt{10 \cdot 377 - (55)^2}} =$$

$$= \frac{5650}{\sqrt{4410} \sqrt{745}} = \frac{5650}{210 \cdot 27,295} = \frac{5650}{5731,95} = 0,986.$$

Παρατηρούμε ότι υπάρχει ισχυρή συσχέτιση μεταξύ των μεταβλητών (X) και (Y).



Εικόνα 12: Διάγραμμα συσχέτισης εφαρμογής 1

Παρουσίαση στατιστικών μέτρων με χρήση SPSS:

Statistics		
VAR00001		
N	Valid	12
	Missing	0
Mean		38,6667
Median		30,5000
Mode		14,00
Std. Deviation		22,47962
Variance		505,333
Range		76,00
Percentiles	25	21,2500
	50	30,5000
	75	55,0000

Πίνακας 13: Στατιστικά μέτρα εφαρμογής 1 με χρήση SPSS

ΕΦΑΡΜΟΓΗ 2

Η εταιρία ARCADIA που δραστηριοποιείται στο χώρο της ναυτιλίας έχει την έδρα της στην Αθήνα σε ένα κτήριο 10 ορόφων. Για κάθε όροφο καταβάλλει ξεχωριστό ενοίκιο. Θέλουμε να βρούμε τα ακόλουθα μέτρα θέσης και διασποράς:

1. μέση τιμή
2. διάμεσο
3. εύρος
4. διακύμανση
5. τυπική απόκλιση

Παρακάτω έχουμε τον πίνακα ενοικίων (σε χιλιάδες ευρώ) που καταβάλλει η εταιρία για τους ορόφους .

ΕΝΟΙΚΙΑ(X)	1	2	3	4
ΟΡΟΦΟΙ(Y)	2	4	2	2

Πίνακας 14: Πίνακας ενοικίων και ορόφων εφαρμογής 2

Υπολογισμός μέσης τιμής \bar{x} :

Πρώτα φτιάχνουμε τον παρακάτω πίνακα:

x_i	v_i	$x_i v_i$
1	2	2
2	4	8
3	2	6
4	2	8
ΣΥΝΟΛΟ	10	24

Πίνακας 15: Στατιστικά στοιχεία εφαρμογής 2

Έπειτα εφαρμόζουμε τον τύπο της μέσης τιμης.

$$\bar{x} = \frac{1}{10} \sum_{i=1}^4 v_i x_i = \frac{24}{10} = 2,4.$$

Για να βρούμε τη διάμεσο κατατάσσουμε πρώτα το δείγμα σε αύξουσα σειρά:
1,1,2,2,2,2,3,3,4,4.

$$\text{Άρα } M = \frac{2+2}{2} = 2$$

Δηλαδή η επιχείρηση κατάβαλλει περίπου 2000 ευρώ για τους μισούς ορόφους.

Το εύρος μεταβολής είναι:

$$R = x_{\max} - x_{\min} = 4 - 1 = 3$$

Η διακύμανση είναι:

$$s^2 = \frac{1}{10-1} \sum_{i=1}^4 v_i (x_i - \bar{x})^2 = \frac{14}{9} = 1,56.$$

Η τυπική απόκλιση είναι:

$$s = \sqrt{s^2} = 1,07.$$

Παρουσίαση στατιστικών μέτρων σε χρήση SPSS:

Descriptives			Statistic	Std. Error
Mean			2,4000	,33993
95% Confidence Interval for Mean	Lower Bound		1,6310	
	Upper Bound		3,1690	
5% Trimmed Mean			2,3889	
Median			2,0000	
Variance			1,156	
Std. Deviation			1,07497	
Minimum			1,00	
Maximum			4,00	
Range			3,00	
Interquartile Range			1,50	
Skewness			,322	,687
Kurtosis			-,882	1,334

Πίνακας 16: Στατιστικά μέτρα εφαρμογής 2 με χρήση SPSS

ΕΦΑΡΜΟΓΗ 3

Η εταιρία ROW δραστηριοποιεί 20 υπαλλήλους σε διευθυντικές θέσεις.

Παρακάτω έχουμε τον πίνακα μισθοδοσίας (σε χιλιάδες ευρώ) των 20 υπαλλήλων.

ΜΙΣΘΟΣ	2	3	4	5
ΥΠΑΛΛΗΛΟΙ	5	5	8	2

Πίνακας 17: Πίνακας μισθών υπαλλήλων εφαρμογής 3

Θέλουμε να αναπαραστήσουμε γραφικά τις μισθοδοσίες με :

1. Ραβδόγραμμα
2. Κυκλικό διάγραμμα

Να βρεθεί επίσης η μέση τιμή, η τυπική απόκλιση, η επικρατούσα τιμή και ο συντελεστής μεταβλητότητας.

Φτιάχνουμε τον παρακάτω πίνακα:

x_i	v_i	$x_i v_i$
2	5	10
3	5	15
4	8	32
5	2	10
ΣΥΝΟΛΟ	20	67

Πίνακας 18: Στατιστικά στοιχεία εφαρμογής 3

Η μέση τιμή είναι:

$$\bar{x} = \frac{1}{20} \sum_{i=1}^4 v_i x_i = \frac{67}{20} = 3,35.$$

Άρα κατά μέσο όρο οι διευθυντές πληρώνονται 3500 ευρώ.

Η διακύμανση είναι:

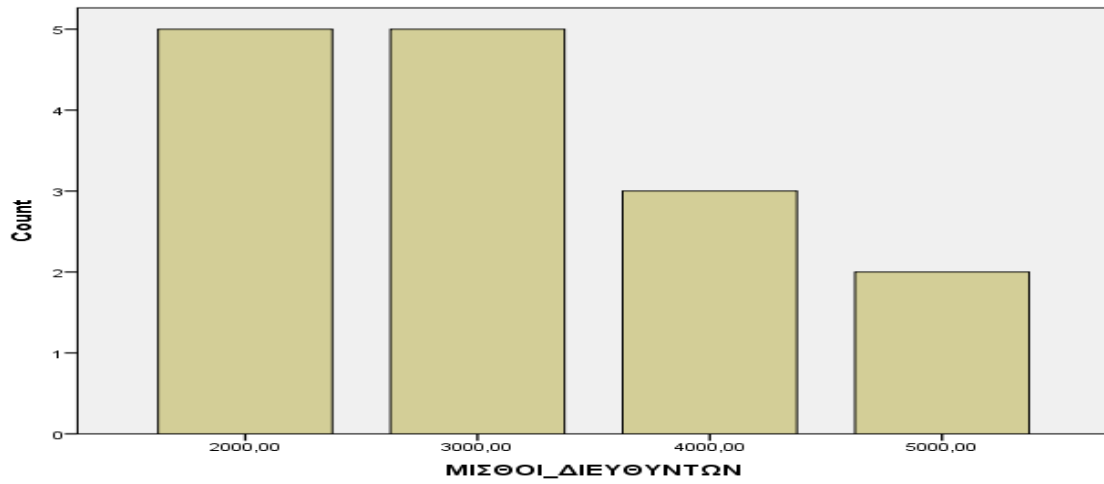
$$s^2 = \frac{1}{20-1} \sum_{i=1}^4 v_i (x_i - \bar{x})^2 = 0,97.$$

Η τυπική απόκλιση είναι:

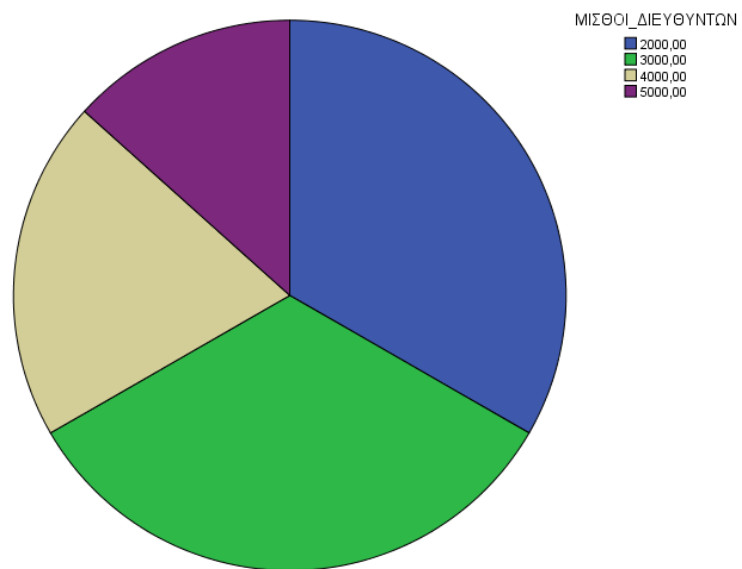
$$s = \sqrt{s^2} = 0,98.$$

Και τέλος, ο συντελεστής μεταβλητότητας:

$$CV = \frac{0,98}{3,35} = 0,29 \text{ ή } 29\%.$$



Εικόνα 13: Ραβδόγραμμα μισθοδοσίας εφαρμογής 3



Εικόνα 14: Κυκλικό διάγραμμα μισθοδοσίας εφαρμογής 3

Παρουσίαση στατιστικών μέτρων με χρήση SPSS:

N	Valid	20
	Missing	0
Mean		3,3500
Median		3,5000
Mode		4,00
Std. Deviation		,98809
Variance		,976
Range		3,00

Εικόνα 15: Στατιστικά μέτρα εφαρμογής 3 με χρήση SPSS

ΕΦΑΡΜΟΓΗ 4

Η Ανώνυμη Εταιρεία DIAPLUS διαθέτει δύο κλάδους εκμετάλλευσης, τον κλάδο (X) εμπορίας ηλεκτρικών συσκευών και τον κλάδο (Y) συντήρησης και επισκευής ηλεκτρικών συσκευών (service). Ο κύκλος εργασιών του κάθε κλάδου εκμετάλλευσης παρακολουθείται χωριστά. Στη διάρκεια του προηγούμενου οικονομικού έτους τα έσοδα του κλάδου εμπορίας ηλεκτρικών συσκευών (X) ήταν 50000 ευρώ του δε κλάδου συντήρησης – επισκευών (Y) ήταν 10000 ευρώ.

Ο γενικός διευθυντής της επιχείρησης θέλοντας να διερευνήσει το βαθμό αποτελεσματικότητας στη διοίκηση και οικονομική διαχείριση της εταιρείας ζήτησε, μεταξύ άλλων, από τους οικονομικούς αναλυτές της επιχείρησης να μελετήσουν τα αντίστοιχα δεδομένα πωλήσεων του προηγούμενου έτους των 12 πρώτων εταιρειών του κλάδου τους και να απαντήσουν, αν μπορούν, στα παρακάτω ερωτήματα:

1. Υπάρχει και σε ποιο βαθμό συσχέτιση μεταξύ των εσόδων από πωλήσεις ηλεκτρικών συσκευών (x) και των εσόδων της υπηρεσίας επισκευών – συντήρησης (y);
2. Να βρεθεί σ' αυτή την ομάδα των 12 επιχειρήσεων η τυπική απόκλιση εσόδων για την εκμετάλλευση X.
3. Να βρεθεί η τυπική απόκλιση εσόδων και για την εκμετάλλευση Y.
4. Να βρεθεί ο CV των εσόδων για την εκμετάλλευση X.

5. Να βρεθεί ο CV των εσόδων για την εκμετάλλευση Υ.
 6. Ποια από τις δύο εκμεταλλεύσεις παρουσιάζει μεγάλη διασπορά εσόδων;

Στήλες δεδομένων (σε χιλ. ευρώ)		Στήλες Υπολογισμών		
(1)	(2)	(3)	(4)	(5)
x	y	xy	x ²	y ²
35	7	245	1225	49
36	7	252	1296	49
38	8	304	1444	64
40	8	320	1600	64
40	8	320	1600	64
42	8,5	357	1764	72,25
44	8,5	374	1936	72,25
45	9	405	2025	81
48	10	480	2304	100
50	10	500	2500	100
55	10	550	3025	100
60	12	720	3600	144
Σύνολο 533	106	4827	24319	959,5

Πίνακας 19: Στατιστικά στοιχεία εφαρμογής 4

Είναι προφανές ότι το ερώτημα απαντάται, αν υπολογίσουμε το συντελεστή γραμμικής συσχέτισης r . Ακολουθώντας τα γνωστά βήματα υπολογισμού έχουμε:

$$\sum_{i=1}^{12} x_i = 533, \quad \sum_{i=1}^{12} y_i = 106, \quad \sum_{i=1}^{12} x_i y_i = 4827$$

$$\sum_{i=1}^{12} x_i^2 = 24319, \quad \sum_{i=1}^{12} y_i^2 = 959,5.$$

Υπολογισμός του r

$$r = \frac{12 \sum_{i=1}^{12} x_i y_i - \left(\sum_{i=1}^{12} x_i \right) \left(\sum_{i=1}^{12} y_i \right)}{\sqrt{12 \sum_{i=1}^{12} x_i^2 - \left(\sum_{i=1}^{12} x_i \right)^2} \sqrt{12 \sum_{i=1}^{12} y_i^2 - \left(\sum_{i=1}^{12} y_i \right)^2}}$$

ή

$$r = \frac{12 \cdot 4827 - 533 \cdot 106}{\sqrt{12 \cdot 24319 - 533^2} \cdot \sqrt{12 \cdot 959,5 - 106^2}}$$

ή

$$r = 0,97 \quad \text{ή} \quad r = 97\%$$

		X	Y
X	Pearson Correlation	1	,978**
	Sig. (2-tailed)		,000
	N	10	10
Y	Pearson Correlation	,978**	1
	Sig. (2-tailed)	,000	
	N	10	10

Πίνακας 20: Συσχέτιση (x) με (y)

Είναι λοιπόν φανερό ότι μεταξύ των δύο εκμεταλλεύσεων υπάρχει θετική συσχέτιση και μάλιστα εντονότατη, αφού το r είναι πολύ κοντά στη μονάδα.

Η τυπική απόκλιση των τιμών της X δίνεται από το γνωστό τύπο:

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

Σχηματίζουμε λοιπόν τον παρακάτω πίνακα κατά τα γνωστά, στη συνέχεια αντικαθιστούμε το άθροισμα των στοιχείων της στήλης (2) στον παραπάνω τύπο και βρίσκουμε:

(1)	(2)	(3)	(4)
$(x - \bar{x})$	$(x - \bar{x})^2$	$(y - \bar{y})$	$(y - \bar{y})^2$
35 - 44,41	88,54	7 - 8,8	3,24
36 - 44,41	70,72	7 - 8,8	3,24
38 - 44,41	41,09	8 - 8,8	0,64
40 - 44,41	19,45	8 - 8,8	0,64
40 - 44,41	19,45	8 - 8,8	0,64
42 - 44,41	5,81	8,5 - 8,8	0,09
44 - 44,41	0,17	8,5 - 8,8	0,09
45 - 44,41	0,35	9 - 8,8	0,04
48 - 44,41	12,88	10 - 8,8	1,44
50 - 44,41	31,25	10 - 8,8	1,44
55 - 44,41	112,15	10 - 8,8	1,44
60 - 44,41	243,05	12 - 8,8	10,24
	703,2		25,2

Πίνακας 21: Στατιστικά στοιχεία εφαρμογής 4

$$s_x^2 = \sqrt{\frac{703}{12}} = \sqrt{58.6}.$$

ή

$$s_x = 7,6.$$

Από τον ίδιο πίνακα βρίσκουμε το άθροισμα της στήλης (4) και το αντικαθιστούμε στον τύπο, οπότε έχουμε:

$$s_y = \sqrt{\frac{25,2}{12}} = \sqrt{2.1}.$$

ή

$$s_y = 1,45.$$

Παρατηρούμε ότι τα έσοδα της εκμετάλλευσης Y παρουσιάζουν πολύ μικρότερη διασπορά από αυτή που παρουσιάζουν τα έσοδα της εκμετάλλευσης X, αφού $s_x = 7,33 > s_y = 1,38$.

Με βάση το γνωστό τύπο $CV_x = \frac{s_x}{\bar{x}}$ θα έχουμε για την περίπτωση της X ότι:

$$CV_x = \frac{s_x}{\bar{x}} = \frac{7,33}{44,4} = 0,165.$$

ή

$$CV_x = 16,5\%.$$

Ομοίως και για την περίπτωση της Y, έχουμε:

$$CV_y = \frac{s_y}{\bar{y}} = \frac{1,38}{8,83} = 0,156.$$

ή

$$CV_y = 15,6\%$$

Παρατηρούμε ότι οι συντελεστές μεταβλητότητας CV_y και CV_x είναι πολύ κοντά ο ένας στον άλλο. Θα λέγαμε λοιπόν ότι η διασπορά των εσόδων της Y, είναι σχεδόν ίδια με αυτή της X. Προφανώς είναι λογικό και επόμενο να συμβαίνει αυτό, επειδή υπάρχει μια υψηλότερη συσχέτιση, όπως είδαμε, $r = 0,97$ με αποτέλεσμα οι τιμές της Y να επηρεάζονται εντονότατα από αυτές της X. Ακόμη παρατηρούμε ότι η διασπορά είναι μικρή και για τις δύο εκμεταλλεύσεις, αφού είναι μικρές οι τιμές του συντελεστή μεταβλητότητας (CV).

Παρουσίαση στατιστικών μέτρων με χρήση SPSS:

Statistics

X

N	Valid	12
	Missing	8
Mean		44,4167
Median		43,0000
Mode		40,00
Std. Deviation		7,65694
Variance		58,629
Sum		533,00

Πίνακας 22: Στατιστικά μέτρα (x) εφαρμογής 4

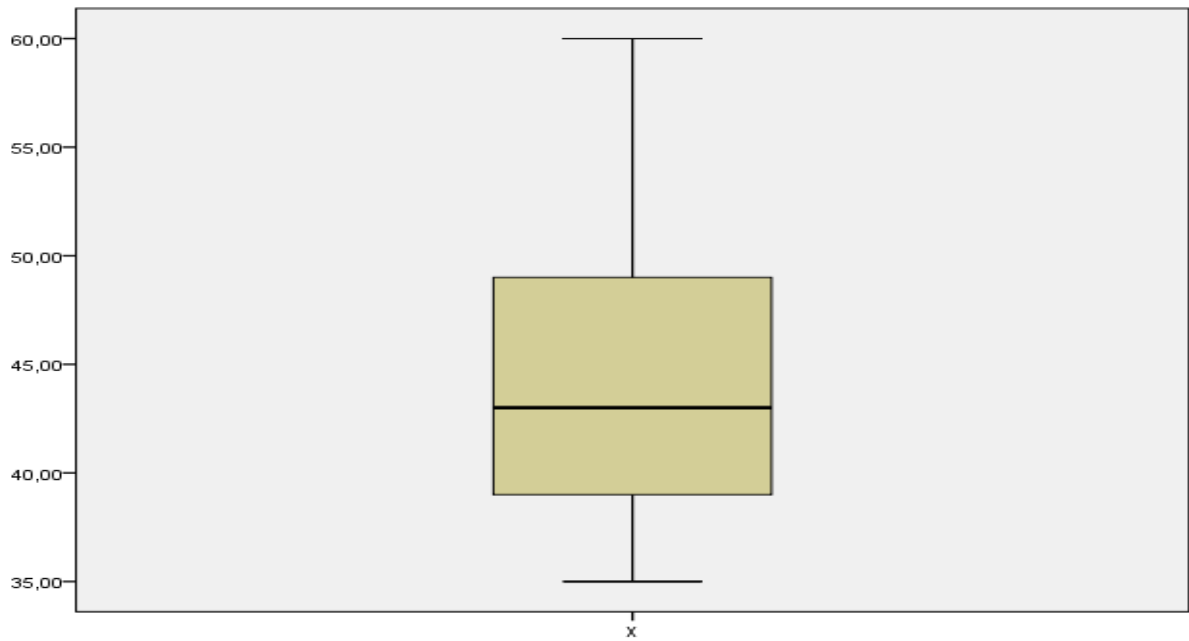
Statistics

y

N	Valid	12
	Missing	8
Mean		8,8333
Median		8,5000
Mode		8,00 ^a
Std. Deviation		1,45123
Variance		2,106
Sum		106,00

a. Multiple modes exist. The smallest value is shown

Πίνακας 23: Στατιστικά μέτρα (y) εφαρμογής 4



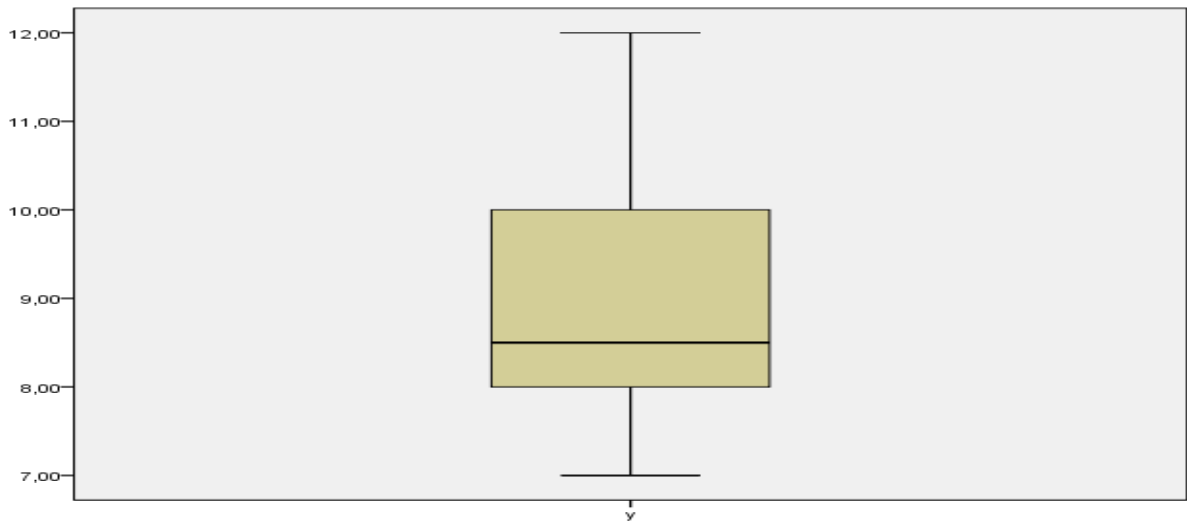
Εικόνα 16: Θηκόγραμμα (x) εφαρμογής 4

χ Stem-and-Leaf Plot

Frequency	Stem &	Leaf
,00	3 .	
3,00	3 .	568
4,00	4 .	0024
2,00	4 .	58
1,00	5 .	0
1,00	5 .	5
1,00	6 .	0

Stem width: 10,00
Each leaf: 1 case(s)

Εικόνα 17: Διάγραμμα μίσχου-φύλλου (x) εφαρμογής 4



Εικόνα 18: Θηκόγραμμα (y) εφαρμογής 4

y Stem-and-Leaf Plot

Frequency	Stem &	Leaf
2,00	7 .	00
5,00	8 .	00055
1,00	9 .	0
3,00	10 .	000
,00	11 .	
1,00	12 .	0

Stem width: 1,00
Each leaf: 1 case(s)

Εικόνα 19: Διάγραμμα μίσχου φύλλου (y) εφαρμογής 4

ΕΦΑΡΜΟΓΗ 5

Από τους ισολογισμούς 40 επιχειρήσεων πήραμε τα παρακάτω στοιχεία, που είναι οι οφειλές (υποχρεώσεις) αυτών των επιχειρήσεων στους προμηθευτές τους, σε χιλιάδες ευρώ:

2, 4, 3, 7, 18, 16, 13, 62, 66, 68, 68, 74, 76, 76, 77, 51, 52, 54, 54, 56, 58, 22, 26, 26, 28, 32, 32, 33, 34, 37, 38, 38, 42, 44, 46, 48, 48, 48, 48, 49.

Ζητούνται:

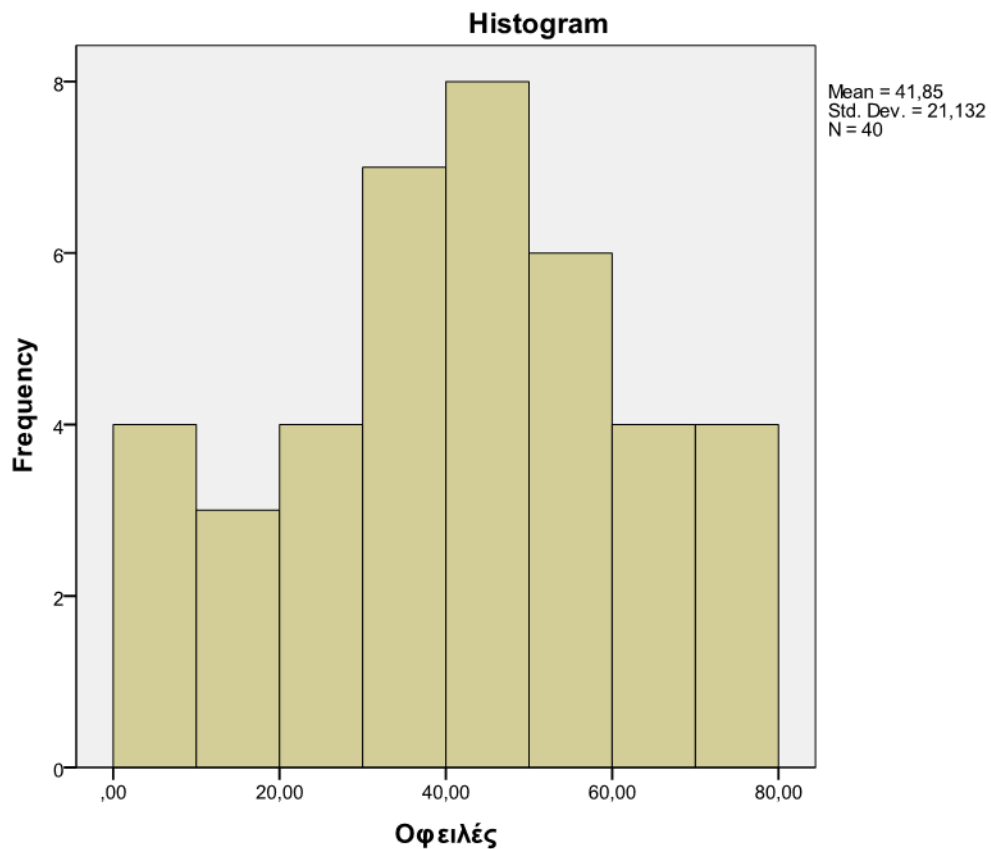
- A) Να βρεθεί η επικρατούσα τιμή.
- B) Να ομαδοποιηθούν τα παραπάνω δεδομένα σε κλάσεις ίσου πλάτους.
- Γ) Να υπολογισθεί η μέση τιμή.
- Δ) Να υπολογιστεί η διάμεσος.
- E) Να υπολογιστεί το ενδοτεταρτημοριακό εύρος.
- Στ) Να υπολογιστεί η διακύμανση.
- Z) Να υπολογιστεί η τυπική απόκλιση.
- H) Να βρεθεί το εύρος.
- Θ) Να υπολογιστεί ο συντελεστής μεταβλητότητας.

ΛΥΣΗ:

- A) Η επικρατούσα τιμή είναι 48, η οποία έχει εμφανιστεί περισσότερες φορές από κάθε άλλη τιμή στην ομάδα των στατιστικών στοιχείων που εξετάζουμε. Δηλαδή, 4 επιχειρήσεις οφείλουν από 48.000 ευρώ η κάθε μια.
- B) Ομαδοποιούμε τα δεδομένα σε οκτώ κλάσεις ίσου εύρους $c = 10$ και σχηματίζουμε τις στήλες:

(1) Κλάσεις οφειλών	(2) Αριθμός προμηθευτών v_i	(3) Κεντρικοί όροι κλάσεων x_i'	(4) $v_i x_i'$
[0-10)	4	5	20
[10-20)	3	15	45
[20-30)	4	25	100
[30-40)	7	35	245
[40-50)	8	45	360
[50-60)	6	55	330
[60-70)	4	65	260
[70-80)	4	75	300
Σύνολο	$v = 40$		$\sum_{i=1}^8 v_i x_i' = 1660$

Πίνακας 24: Πίνακας κλάσεων εφαρμογής 5



Εικόνα 20: Ιστόγραμμα εφαρμογής 5

Γ) Για τον υπολογισμό της μέσης τιμής κατασκευάζουμε τις στήλες (3) των κεντρικών όρων και (4) στον πίνακα 25, από τον οποίο βρίσκουμε $n = 40$ και $\sum_{i=1}^8 v_i x_i' = 1660$. Αντικαθιστώντας τα δύο αυτά αθροίσματα στον τύπο εύρεσης της μέσης τιμής έχουμε:

$$\bar{x} = \frac{\sum_{i=1}^8 v_i x_i'}{n} = \frac{1660}{40} = 41,5.$$

Παρατηρούμε ότι η μέση τιμή των οφειλών των 40 επιχειρήσεων στους προμηθευτές τους είναι 41.500 ευρώ.

Δ) Ο υπολογισμός της διαμέσου θα γίνει με τον εξής τύπο:

$$M = L_{i-1} + \frac{\delta}{v_i} \left(\frac{v}{2} - N_{i-1} \right) = 40 + \frac{10}{8} \left(\frac{40}{2} - 18 \right) = 42,5.$$

Το 50% των επιχειρήσεων οφείλει μέχρι και 42.500 ευρώ στους προμηθευτές.

Ε) Για να υπολογιστεί το ενδοτεταρτημοριακό εύρος θα πρέπει πρώτα να βρούμε το πρώτο τεταρτημόριο και το τρίτο τεταρτημόριο. Εργαζόμαστε ως εξής:

Ο υπολογισμός του πρώτου τεταρτημορίου θα γίνει με τον τύπο:

$$Q_1 = L_{i-1} + \frac{\delta}{v_i} \left(\frac{v}{4} - N_{i-1} \right) = 20 + \frac{10}{4} \left(\frac{40}{4} - 7 \right) = 27,5.$$

Αντίστοιχα ο υπολογισμός του τρίτου τεταρτημορίου θα γίνει με τον τύπο:

$$Q_3 = L_{i-1} + \frac{\delta}{v_i} \left(\frac{3v}{4} - N_{i-1} \right) = 50 + \frac{10}{6} \left(\frac{120}{4} - 26 \right) = 56,7.$$

Άρα το ενδοτεταρτημοριακό εύρος θα είναι:

$$H = Q_3 - Q_1 = 56,7 - 27,5 = 29,2.$$

Το 25% των επιχειρήσεων οφείλει μέχρι και 27.500 ευρώ ενώ υπάρχει και ένα 25% των επιχειρήσεων που οφείλει περισσότερο από 56.700 ευρώ.

Στ) Ο υπολογισμός της διακύμανσης θα γίνει από τον τύπο:

$$s^2 = \frac{1}{v-1} \left[\sum_{i=1}^8 v_i (x_i')^2 - \frac{1}{v} \left(\sum_{i=1}^8 v_i x_i' \right)^2 \right] = 457,89.$$

Θ) Η τυπική απόκλιση δίνεται από τον τύπο:

$$s = \sqrt{s^2} = 21,39.$$

H) Το εύρος δίνεται από το τύπο:

$$R = x_{\min} - x_{\max} = 49 - 2 = 47.$$

Z) Τέλος ο συντελεστής μεταβλητότητας θα υπολογιστεί με τον τύπο:

$$CV = \frac{s}{\bar{x}} = \frac{21,39}{41,5} = 0,51 \text{ ή } 51\%$$

ΣΥΜΠΕΡΑΣΜΑΤΑ

Μία απλή απαρίθμηση των εφαρμογών της Στατιστικής που είναι βασικά μια εφαρμοσμένη επιστήμη, αποδεικνύει ότι αυτή χρησιμοποιείται σε όλους σχεδόν τους τομείς της ανθρώπινης δραστηριότητας.

Μετά τη συλλογή δεδομένων θα πρέπει αυτά να παρουσιάζονται και να οπτικοποιούνται, αυτό επιτυγχάνεται με τη βοήθεια των Στατιστικών πινάκων.

Στατιστικοί πίνακες

- Γενικοί πίνακες
- Ειδικοί πίνακες

Μετά τη συγκέντρωση των στατιστικών στοιχείων που αναφέρονται σε μια μόνο ιδιότητα ενός πληθυσμού το πιο σημαντικό βήμα είναι η κατάλληλη κατάταξη και η συστηματική ομαδοποίηση των τιμών της μεταβλητής με αντικειμενικό σκοπό την παρουσίαση των στοιχείων με τέτοιο τρόπο ώστε να δίνει πληροφορίες για τη δομή του πληθυσμού που είναι προς μελέτη, και να διευκολύνεται το έργο της στατιστικής ανάλυσης και εξαγωγής χρήσιμων συμπερασμάτων.

Η ομαδοποίηση των τιμών της μεταβλητής χρησιμοποιούμε ειδικές κατατάξεις που τις ονομάζουμε κατανομές συχνοτήτων ή πίνακες συχνοτήτων.

Για την κατασκευή αυτών των πινάκων χρησιμοποιούμε τις:

- Συχνότητες
- Σχετικές συχνότητες
- Αθροιστικές συχνότητες

Σε περίπτωση μεγάλου αριθμού παρατηρήσεων αξιοποιούμε τη μέθοδο της ομαδοποίησης των παρατηρήσεων.

- Γραφικές παραστάσεις

Οι γραφικές παραστάσεις είναι το καλύτερο μέσο μιας στατιστικής παρουσίασης γιατί δίνουν στους αφηρημένους αριθμούς μια συγκεκριμένη μορφή που μας διευκολύνει να έχουμε μια άμεση αντίληψη της μορφής του φαινομένου που θέλουμε να μελετήσουμε. Υπάρχουν πολλές κατηγορίες διαγραμμάτων μεταξύ αυτών είναι:

- Ραβδόγραμμα
- Διάγραμμα συχνοτήτων
- Κυκλικό διάγραμμα
- Χρονοδιάγραμμα
- Ιστόγραμμα

Η χρήση πινάκων και διαγραμμάτων από τις επιχειρήσεις γίνεται και για εσωτερικούς, ερευνητικούς και επιστημονικούς λόγους που αφορούν την ίδια την επιχείρηση. Με τους πίνακες και τα διαγράμματα γίνεται πιο εύκολη και γρήγορη η εσωτερική ενημέρωση των στελεχών, των εργαζομένων της επιχείρησης και των πολύ στενών συνεργατών της, ιδιαίτερα στις τακτικές και έκτακτες ενημερωτικές συναντήσεις (meetings) που γίνονται.

Ειδικότερα για τα διαγράμματα, θα πρέπει να γνωρίζουμε ότι λόγω των τεράστιων δυνατοτήτων των ηλεκτρονικών υπολογιστών μπορούμε να κατασκευάζουμε εξαιρετικής ακρίβειας διαγράμματα με συνθέσεις χρωμάτων και πρωτοποριακά σχήματα που ελκύουν την προσοχή του κοινού.

Η γραφική απεικόνιση μιας ομαδοποιημένης κατανομής γίνεται με τη βοήθεια των ιστογραμμάτων συχνοτήτων ή σχετικών συχνοτήτων καθώς και με τα αντίστοιχα πολύγωνα. Οι πίνακες κατανομής συχνοτήτων δεν επιτρέπουν την εύκολη σύγκριση. Για το λόγο αυτό αντικαθίστανται από ορισμένους αντιπροσωπευτικούς αριθμούς που ονομάζονται στατιστικές παράμετροι και χαρακτηρίζουν τη θέση, τη διασπορά και τη μορφολογία του πληθυσμού που ερευνούμε και διευκολύνουν πάρα πολύ στις συγκρίσεις ομοειδών ομάδων.

- **Στατιστικές παράμετροι**
 - **Μέτρα θέσης**
 - αριθμητικός μέσος
 - διάμεσος
 - επικρατούσα τιμή
 - εκατοστημόρια
- **Μέτρα διασποράς**
 - εύρος
 - ενδοτεταρτημοριακό εύρος
 - διακύμανση
 - συντελεστής μεταβλητότητας

Τα μέτρα θέσης που μελετήσαμε, και ιδιαίτερα η μέση τιμή και η διάμεσος, είναι από τα πιο συχνά χρησιμοποιούμενα στατιστικά μέτρα και αποτελούν πολύτιμο πληροφοριακό υλικό για την άσκηση οποιασδήποτε επιχειρηματικής πολιτικής. Ακόμη και από την κρατική πλευρά υπάρχει μεγάλο ενδιαφέρον για αυτά, αφού η εφαρμογή τους βοηθάει πολύ στην άσκηση ορθής κοινωνικής, οικονομικής κ.τ.λ. πολιτικής. Παρόλο όμως που τα μέτρα θέσης είναι πολύ σημαντικά και απαραίτητα, δεν μας δίνουν πλήρη εικόνα του φαινομένου που εξετάζουμε. Εξίσου σημαντική και απαραίτητη είναι και η μελέτη για τα μέτρα διασποράς.

Είναι πλέον προφανές, μετά τη μελέτη που κάναμε πάνω στα μέτρα θέσης 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. Τελευταία 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.

ΕΠΙΛΟΓΟΣ

Οι οικονομικές μονάδες μοιάζουν με τους ζωντανούς οργανισμούς: Δημιουργούνται, αναπτύσσονται, ακμάζουν, παρακμάζουν και πεθαίνουν, μέσα σε ένα αβέβαιο περιβάλλον. Η νομοτελειακή πορεία τους, όμως, από τη γέννηση προς το θάνατο είναι δυνατό να διαφοροποιηθεί ποιοτικά, μέσω της παράτασης των φάσεων της ανάπτυξης και της ακμής.

Για να συμβεί αυτό θα πρέπει, στο μέτρο του δυνατού, να προγραμματίζεται η μελλοντική τους δράση, ώστε να εντοπίζονται έγκαιρα, να εκτιμώνται σωστά και να αντιμετωπίζονται αποτελεσματικά οι κίνδυνοι που τις απειλούν, παράλληλα δε να αξιοποιούνται οι ευκαιρίες που πηγάζουν από το ευρύτερο κοινωνικό και οικονομικό περιβάλλον. Είναι αναγκαία, επομένως, η διαρκής και συστηματική καταγραφή, παρατήρηση, ταξινόμηση, αξιολόγηση και αξιοποίηση των κάθε φορά διαθέσιμων δεδομένων, τα οποία εκφράζονται συνηθέστερα ως πιθανοτικές κατανομές και σπανιότερα ως μονοσήμαντα μεγέθη.

Η επιχείρηση πρέπει να χρησιμοποιεί τη Στατιστική για να μπορεί να προσεγγίσει τα μελλοντικά ποσοτικά μεγέθη της, που αντιστοιχούν σε εναλλακτικούς συνδυασμούς πιθανοτικών δεδομένων, με απλό και αποτελεσματικό τρόπο.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Γναρδέλης Χ., *Εφαρμοσμένη στατιστική*, Εκδόσεις Παπαζήση, 2003.
- [2] Γναρδέλης Χ., *Ανάλυση Δεδομένων με το PASW Statistics 17.0*, Εκδόσεις Παπαζήση, Αθήνα 2009.
- [3] Καραγεώργου Δ., Κόκλα Α., Παπακωσταντίνου Ε., *Στατιστική για τις επιχειρήσεις*, Οργανισμός έκδοσης διδακτικών βιβλίων, Αθήνα, 1999.
- [4] Κιόχος Π., Κιόχος Α., *Στατιστική για τις επιχειρήσεις και την οικονομία*, εκδ. Ελένη Κιόχου, 2010.
- [5] Μαμαλής Π., Καψή Θ., Τόλης Ε., Μιχαήλογλου Σ., Πρίντεζης Γ., *Στατιστική: Εφαρμογή στη διοίκηση – οικονομία – επιχειρήσεις, βασικές γνώσεις μαθηματικών –στατιστικής*, Ινστιτούτο Διαρκούς Εκπαίδευσης Ενηλίκων (ΙΔΕΚΕ).

ΠΗΓΕΣ ΑΠΟ ΤΟ ΔΙΑΔΙΚΤΥΟ

auth. Mamalis P., Kapsis Th., Tolis E., Mixailoglou S., Printezis G. *Βασικές γνώσεις μαθηματικών – Στατιστικής*.

Διαθέσιμο στο δικτυακό τόπο:

<http://repository.edulll.gr/edulll/retrieve/2866/878.pdf>

(τελευταία πρόσβαση στις 14/9/2014)

auth. Adamopoulos L. Damianou X., Sberkos A., *ΜΑΘΗΜΑΤΙΚΑ ΚΑΙ ΣΤΟΙΧΕΙΑ ΣΤΑΤΙΣΤΙΚΗΣ*

Διαθέσιμο στο δικτυακό τόπο:

<http://ebooks.edu.gr/modules/ebook/show.php/DSGL-C100/493/3201,13005/>.

(τελευταία πρόσβαση στις 14/9/2014)

auth. Karagiorgos D. Koklas A., *Στατιστική επιχειρήσεων*.

Διαθέσιμο στο δικτυακό τόπο:

<http://www.pi-schools.gr/lessons/tee/economic/>

(τελευταία πρόσβαση στις 14/9/2014)

auth. Rodos K. Papanis E., *Στατιστική έρευνα - μέθοδοι και εφαρμογές*

Διαθέσιμο στο δικτυακό τόπο:

http://users.sch.gr/epdiaman/images/stories/ergasies/biblia/statistics_with_calc_and_R_project.pdf.

(τελευταία πρόσβαση στις 14/9/2014)

Πνευματικά δικαιώματα

Copyright © ΤΕΙ Δυτικής Ελλάδας. Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Δηλώνω ρητά ότι, σύμφωνα με το άρθρο 8 του Ν. 1599/1988 και τα άρθρα 2,4,6 παρ. 3 του Ν. 1256/1982, η παρούσα εργασία αποτελεί αποκλειστικά προϊόν προσωπικής εργασίας και δεν προσβάλλει κάθε μορφής πνευματικά δικαιώματα τρίτων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον.

Παυλίδου Βασιλική, [2014]