ΦΛΩΡΟΥ ΓΙΑΝΝΟΥΛΑ
Καθηγήτρια ΔΙ.ΠΑ.Ε.

# *Στατιστικές Μέθοδοι Έρευνας*

Με εφαρμογή πιθανοτήτων στη λήψη αποφάσεων, θεωρία παιγνίων και αξιολόγηση επενδύσεων

ΚΑΛΛΙΠΟΣ
ανοικτές ακαδημαϊκές εκδόσεις

Εθνικό
Πρόγραμμα
Ανάπτυξης
2021-2025

# Στατιστικές Μέθοδοι Έρευνας
# Με εφαρμογή πιθανοτήτων στη λήψη αποφάσεων, θεωρία παιγνίων και αξιολόγηση επενδύσεων

***Συγγραφή***

Γιαννούλα Φλώρου

***Συντελεστές έκδοσης***

Γλωσσική Επιμέλεια: Μαρία Νιάρη

Γραφιστική Επιμέλεια: Ιωάννης Παλιόκας

ΚΑΛΛΙΠΟΣ
Εθνικό Μετσόβιο Πολυτεχνείο
Ηρώων Πολυτεχνείου 9, 15780 Ζωγράφου

www.kallipos.gr

ISBN: 978-618-228-000-3

**Βιβλιογραφική Αναφορά:** Φλώρου, Γ. (2023). *Στατιστικές Μέθοδοι Έρευνας* [Μεταπτυχιακό εγχειρίδιο]. Κάλλιπος, Ανοικτές Ακαδημαϊκές Εκδόσεις. http://dx.doi.org/10.57713/kallipos-230

*Στον σύζυγο μου Βύρωνα*

# Πίνακας Περιεχομένων


**Πίνακας συντομεύσεων-ακρωνυμίων** .......... **9**
**Πρόλογος** .......... **11**
**Εισαγωγή** .......... **13**
**Κεφάλαιο 1: Μεθοδολογία Επιστημονικής Έρευνας – Δεοντολογία** .......... **15**
1.1 Τι είναι επιστημονική έρευνα .......... 15
1.1.1 Τι δεν είναι επιστημονική έρευνα .......... 15
1.1.2 Μέθοδοι και μεθοδολογία .......... 15
1.2 Είδη ερευνών .......... 15
1.3 Τρόποι διεξαγωγής ερευνών .......... 16
1.3.1 Ερευνητική πορεία .......... 16
1.3.2 Βιβλιογραφική επισκόπηση .......... 17
1.4 Αρχές δεοντολογίας .......... 17
1.5 Εφαρμογές επιστημονικής έρευνας .......... 18
1.5.1 Επιλογή θέματος από φοιτητές .......... 18
1.5.2 Στατιστική έρευνα .......... 19
1.5.3 Παράδειγμα σχεδιασμού στατιστικής έρευνας .......... 19
1.5.4 Ανάλυση και αξιολόγηση των δεδομένων της στατιστικής έρευνας .......... 20
1.5.5 Συγγραφή εργασίας .......... 20
1.5.6 Προφορική παρουσίαση εργασίας .......... 21
**Βιβλιογραφία/Αναφορές** .......... **22**
**Κριτήρια αξιολόγησης** .......... **23**
**Κεφάλαιο 2: Τρόποι Συλλογής Δεδομένων – Δειγματοληψία** .......... **25**
2.1 Τρόποι συλλογής δεδομένων .......... 25
2.1.1 Ορισμοί .......... 25
2.1.2 Απογραφή, συνεχής καταγραφή, δειγματοληψία .......... 26
2.1.2.1 Απογραφή .......... 26
2.1.2.2 Δειγματοληψία .......... 27
2.1.2.3 Συνεχής καταγραφή .......... 27
2.2 Ερωτηματολόγιο, δομή, περιεχόμενο .......... 27
2.2.1 Μορφή ερωτηματολογίου .......... 27
2.2.2 Διατύπωση ερωτήσεων .......... 27
2.3 Σφάλμα επιλογής δείγματος .......... 28
2.4 Είδη δειγματοληψίας .......... 28
2.5 Παραδείγματα – Εφαρμογές .......... 29
2.5.1 Αξιολόγηση ερωτηματολογίου .......... 29

2.5.2 Δημιουργία ερωτηματολογίου ........ 29
2.5.3 Συλλογή στοιχείων ........ 29
**Βιβλιογραφία/Αναφορές ........ 30**
**Κριτήρια αξιολόγησης ........ 31**
**Κεφάλαιο 3: Περιγραφή Δεδομένων – Παρουσίαση με Παραμέτρους ........ 33**
3.1 Είδη μεταβλητών ........ 33
3.1.1 Κλίμακες μέτρησης ποσοτικών μεταβλητών ........ 34
3.1.2 Ποιοτικές μεταβλητές ........ 34
3.1.2.1 Μεταβλητές με πολλαπλές τιμές ........ 35
3.2 Περιγραφική παρουσίαση δεδομένων ........ 35
3.2.1 «Παράμετροι» ποσοτικών δεδομένων ........ 37
3.2.2 Παρουσίαση ποιοτικών δεδομένων ........ 37
3.2.3 Υπολογισμός παραμέτρων ποσοτικών μεταβλητών ........ 39
3.3 Κατανομές πιθανοτήτων ........ 41
3.3.1 Κανονική κατανομή ........ 42
3.3.1.1 Υπολογισμός πιθανοτήτων κανονικής κατανομής ........ 42
3.4 Εφαρμογές για κατανόηση ........ 44
3.4.1 Σχολιασμός αποτελεσμάτων ........ 44
3.4.2 Εφαρμογή σε κανονική κατανομή ........ 45
3.4.3 Σύγκριση τιμών ........ 46
3.5 Ασκήσεις για κατανόηση - εφαρμογή ........ 47
3.5.1 Άσκηση Α΄: Εισόδημα ........ 47
3.5.2 Άσκηση Β΄: Ημερήσια κατανάλωση ρεύματος ........ 48
3.5.3 Άσκηση Γ΄: Θηκόγραμμα και σχολιασμός ........ 49
3.5.4 Άσκηση Δ΄: Σχολιασμός αποτελεσμάτων ........ 50
3.5.5 Άσκηση Ε΄: Καθυστερήσεις πτήσεων ........ 50
3.5.6 Άσκηση ΣΤ΄: Κατανομή μισθωτών ........ 51
5.7 Άσκηση Ζ΄: Πριμ παραγωγικότητας εργαζομένων ........ 51
**Βιβλιογραφία/Αναφορές ........ 52**
**Κριτήρια αξιολόγησης ........ 53**
Παράρτημα: ........ 54
**Κεφάλαιο 4: Διαστήματα Εμπιστοσύνης – Έλεγχοι Υποθέσεων ........ 57**
4.1 Εκτίμηση (από το μερικό στο γενικό) ........ 57
4.2 Σημειακή εκτίμηση και εκτίμηση με διάστημα ........ 57
4.3 Στάθμη σημαντικότητας και κατανομή παραμέτρου ........ 57
4.3.1 Κατανομή (δειγματοληπτική) παραμέτρου ........ 58
4.4 Διάστημα εμπιστοσύνης για μέση τιμή, αναλογία ........ 58
4.4.1 Διάστημα εμπιστοσύνης για τη μέση τιμή πληθυσμού από δείγμα ........ 59
4.4.2 Διάστημα εμπιστοσύνης για την αναλογία πληθυσμού από δείγμα ........ 61

4.4.3 Διάστημα εμπιστοσύνης για τη διακύμανση πληθυσμού από δείγμα ........ 61
4.5 Έλεγχος υποθέσεων, μηδενική υπόθεση ........ 62
4.5.1 Πώς δημιουργείται και πώς ελέγχεται μία υπόθεση ........ 62
4.5.2 Τυχαίες και σημαντικές διαφορές ........ 62
4.5.3 Μηδενική υπόθεση H0 ........ 62
4.6 Σφάλματα στον έλεγχο υποθέσεων ........ 63
4.7 Υπολογισμός κριτηρίων για τη μηδενική υπόθεση ........ 63
4.7.1 Κριτήρια παραμετρικών ελέγχων ........ 64
4.7.1.1 Κριτήρια παραμετρικών ελέγχων ........ 64
4.7.1.2 Ανάλυση διακύμανσης ........ 65
4.7.2 Μη παραμετρικοί έλεγχοι ........ 65
4.8 Απόφαση για την αποδοχή της μηδενικής υπόθεσης ........ 66
4.9 Εφαρμογές για κατανόηση ........ 66
4.9.1 Έλεγχος επιβατών ........ 66
4.9.2 Έλεγχος οικονομικού ελεγκτή ........ 66
4.10 Ασκήσεις για κατανόηση - εφαρμογή ........ 67
4.10.1 Άσκηση Α΄: Αγορές με κατάτμηση διαγωνισμού ........ 67
4.10.2 Άσκηση Β΄: Αξία αποθεμάτων ........ 67
4.10.3 Άσκηση Γ΄: Πρόγραμμα αποζημιώσεων ........ 67
4.10.4 Άσκηση Δ΄: Μέσες τιμές δειγμάτων ........ 68
**Βιβλιογραφία/Αναφορές ........ 69**
**Κριτήρια αξιολόγησης ........ 70**
Παράρτημα ........ 71
**Κεφάλαιο 5: Συσχέτιση ........ 73**
5.1 Η έννοια της συσχέτισης ........ 73
5.1.1 Οπτικοποίηση συσχέτισης δύο ποιοτικών μεταβλητών ........ 73
5.1.2 Οπτικοποίηση γραμμικής συσχέτισης δύο ποσοτικών μεταβλητών ........ 74
5.2 Έλεγχος συσχέτισης ποιοτικών μεταβλητών ........ 76
5.2.1 Βήματα ελέγχου ανεξαρτησίας δύο ποιοτικών μεταβλητών ........ 76
5.2.2 Παράδειγμα ελέγχου ανεξαρτησίας δύο ποιοτικών μεταβλητών ........ 77
5.2.3 Μέτρα συνάφειας δύο ποιοτικών μεταβλητών ........ 78
5.3 Συσχέτιση ποσοτικών μεταβλητών ........ 79
5.3.1 Ερμηνεία των τιμών του ρ ........ 79
5.3.2 Περιορισμοί του συντελεστή γραμμικής συσχέτισης ........ 80
5.3.3 Παράδειγμα υπολογισμού συντελεστή γραμμικής συσχέτισης ........ 80
5.4 Έλεγχος ύπαρξης συσχέτισης στον πληθυσμό ........ 81
5.4.1 Παράδειγμα ελέγχου συντελεστή γραμμικής συσχέτισης ........ 81
5.5 Παρερμηνείες – Επηρεασμός συσχέτισης ........ 82
5.6 Συσχέτιση διατάξιμων μεταβλητών ........ 82

5.7 Εφαρμογές για κατανόηση ........ 83
5.7.1 Παράδειγμα ελέγχου ανεξαρτησίας ........ 83
5.8 Ασκήσεις για κατανόηση - εφαρμογή ........ 84
5.8.1 Έλεγχος ανεξαρτησίας ........ 84
5.8.2 Έλεγχος γραμμικής συσχέτισης ........ 84
5.8.3 Έλεγχος γραμμικής συσχέτισης πολλών διατάξιμων μεταβλητών ........ 86
**Βιβλιογραφία/Αναφορές ........ 88**
**Κριτήρια αξιολόγησης ........ 89**
**Κεφάλαιο 6: Παλινδρόμηση ........ 91**
6.1 Περιγραφή μοντέλου παλινδρόμησης ........ 91
6.1.1 Απλή παλινδρόμηση ........ 92
6.1.2 Πολλαπλή παλινδρόμηση ........ 93
6.1.3 Διαδικασία παλινδρόμησης ........ 94
6.2 Μετατροπή σε γραμμικό μοντέλο παλινδρόμησης ........ 94
6.3 Προϋποθέσεις – Έλεγχοι προϋποθέσεων ........ 95
6.3.1 Προϋποθέσεις δείγματος δεδομένων ........ 95
6.3.2 Προϋποθέσεις μεταβλητών μοντέλου ........ 95
6.3.3 Προϋποθέσεις καταλοίπων εκτίμησης ........ 96
6.3.4 Εξειδίκευση μοντέλου – Έλεγχος ........ 97
6.3.4.1 Έλεγχος πολυσυγγραμμικότητας μεταβλητών ........ 97
6.4 Υπολογισμός ευθείας ελαχίστων τετραγώνων ........ 98
6.4.1 Υπολογισμός παραμέτρων α και β στην απλή παλινδρόμηση ........ 98
6.4.1.1 Παράδειγμα υπολογισμού παραμέτρων απλής παλινδρόμησης ........ 98
6.4.2 Υπολογισμός παραμέτρων α και β1, β2,... βκ στην πολλαπλή παλινδρόμηση ........ 99
6.5 Αξιολόγηση της ευθείας παλινδρόμησης ........ 99
6.5.1 Παράδειγμα υπολογισμού $R^2$ στην απλή παλινδρόμηση ........ 100
6.6 Έλεγχος σημαντικότητας των συντελεστών ........ 101
6.7 Αξιοπιστία προβλέψεων διάστημα εμπιστοσύνης πρόβλεψης ........ 102
6.8 Εφαρμογές για κατανόηση ........ 102
6.8.1 Μοντέλο παλινδρόμησης για ζήτηση παγωτού ........ 102
6.8.2 Μοντέλο παλινδρόμησης για πωλήσεις αυτοκινήτων ........ 105
6.9 Ασκήσεις για κατανόηση - εφαρμογή ........ 113
6.9.1 Ερμηνεία αποτελεσμάτων παλινδρόμησης ........ 113
6.9.2 Ερμηνεία αποτελεσμάτων παλινδρόμησης ........ 114
**Βιβλιογραφία/Αναφορές ........ 115**
**Κριτήρια αξιολόγησης ........ 116**
**Κεφάλαιο 7: Πολυμεταβλητή Παραγοντική Ανάλυση ........ 117**
7.1 Εισαγωγή στην πολυμεταβλητή ανάλυση ........ 117
7.1.1 Αιτιότητα και συνάφεια ........ 117

7.1.2 Μείωση διαστάσεων ..... 118
7.1.3 Παραγοντική ανάλυση (Factor analysis) ..... 118
7.1.3.1 Χρησιμότητα ..... 119
7.1.3.1.1 Παράδειγμα ..... 119
7.2 Προϋποθέσεις παραγοντικής ανάλυσης ..... 120
7.3 Έλεγχοι καταλληλότητας δεδομένων ..... 120
7.4 Κριτήρια μέτρησης «αποστάσεων» ..... 120
7.5 Εύρεση κυρίων παραγόντων ..... 122
7.6 Επιλογή κυρίων παραγόντων ..... 122
7.7 Ερμηνεία παραγόντων ..... 123
7.7.1 Ονομασία παραγόντων ..... 123
7.8 Γραφική απεικόνιση παραγοντικών επιπέδων ..... 123
7.9 Εφαρμογές για κατανόηση ..... 123
7.9.1 Συνήθειες καταναλωτών ..... 123
7.9.2 Επισκέψεις σε λασπόλουτρα ..... 130
7.10 Ασκήσεις για κατανόηση - εφαρμογή ..... 134
**Βιβλιογραφία/Αναφορές ..... 136**
**Κριτήρια αξιολόγησης ..... 137**
**Κεφάλαιο 8: Λήψη Αποφάσεων Μελλοντικών Χρηματικών Ροών ..... 139**
8.1 Διαχρονική αξία χρήματος ..... 139
8.1.1 Ερωτήματα για μελλοντικά κεφάλαια ..... 140
8.1.2 Τόκος και επιτόκιο ..... 140
8.1.2.1 Απλός και σύνθετος τόκος ..... 141
8.1.2.2 Συχνότητα ανατοκισμού ..... 141
8.1.2.2.1 Πραγματικό επιτόκιο ..... 142
8.2 Επιτόκιο και μελλοντική αξία ..... 142
8.3 Προεξόφληση ..... 143
8.4 Γραμμάτια - συναλλαγματικές ..... 143
8.5 Παρούσα αξία πολλών μελλοντικών ροών ..... 144
8.6 Ράντες ..... 144
8.7 Δάνεια ..... 144
8.8 Εφαρμογές για κατανόηση ..... 145
8.8.1 Παράδειγμα υπολογισμού τελικού κεφαλαίου ..... 145
8.8.2 Παράδειγμα υπολογισμού τελικού κεφαλαίου ..... 145
8.8.3 Παράδειγμα πραγματικού επιτοκίου ..... 145
8.8.4 Παράδειγμα πραγματικού επιτοκίου ..... 145
8.8.5 Παράδειγμα υπολογισμού παρούσας αξίας ..... 146
8.8.6 Παράδειγμα αντικατάστασης γραμματίων ..... 146
8.9 Ασκήσεις για κατανόηση - εφαρμογή ..... 147

8.9.1 Αντικατάσταση γραμματίων ... 147
8.9.2 Προεξόφληση ... 147
**Βιβλιογραφία/Αναφορές ... 148**
**Κριτήρια αξιολόγησης ... 149**
**Κεφάλαιο 9: Αξιολόγηση Επενδύσεων ... 151**
9.1 Έννοια της επένδυσης ... 151
9.2 Χρηματοδότηση επενδύσεων ... 151
9.2.1 Πηγές χρηματοδότησης κεφαλαίου ... 151
9.2.2 Κόστος κεφαλαίου ... 151
9.3 Τεχνικές αξιολόγησης επενδύσεων ... 151
9.4 Καθαρή παρούσα αξία ... 152
9.5 Εσωτερικός βαθμός απόδοσης (IRR) ... 153
9.6 Περίοδος επανείσπραξης (PP) ... 153
9.7 Λογιστικός συντελεστής απόδοσης (ARR) ... 154
9.8 Δείκτης κερδοφορίας (PI) ... 154
9.9 Εφαρμογές για κατανόηση ... 154
9.9.1 Επιλογή με τη μέθοδο της Καθαρής Παρούσας Αξίας ... 155
9.9.2 Επιλογή με τη μέθοδο Εσωτερικού Ρυθμού Απόδοσης (IRR) ... 156
9.9.3 Επιλογή με τη μέθοδο της Περιόδου Επανείσπραξης ... 157
9.9.4 Επιλογή με τη μέθοδο του Λογιστικού Συντελεστή Απόδοσης (ARR) ... 157
9.9.5 Επιλογή με τη μέθοδο του Δείκτη Κερδοφορίας (PI) ... 158
9.10 Ασκήσεις για κατανόηση - εφαρμογή ... 159
9.10.1 Αντικατάσταση εξοπλισμού ... 159
9.10.2 Αξιολόγηση επενδυτικής πρότασης ... 160
**Βιβλιογραφία/Αναφορές ... 161**
**Κριτήρια αξιολόγησης ... 162**
**Κεφάλαιο 10: Θεωρία Παιγνίων ... 163**
10.1 Αποφάσεις σε συνθήκες αβεβαιότητας ... 163
10.1.1 Βήματα ανάλυσης αποφάσεων ... 163
10.1.2 Δραστηριότητες και εκβάσεις της «φύσης» ... 163
10.1.3 Συνθήκες αποφάσεων ... 164
10.2 Αποφάσεις σε συνθήκες κινδύνου ... 164
10.3 Αποφάσεις σε συνθήκες αβεβαιότητας ... 165
10.3.1 Παράδειγμα αποφάσεων σε συνθήκες αβεβαιότητας ... 165
10.3.2 Επιλογή με το κριτήριο maxmin σε συνθήκες αβεβαιότητας ... 166
10.3.3 Επιλογή με το κριτήριο maxmax σε συνθήκες αβεβαιότητας ... 166
10.3.4 Επιλογή με το κριτήριο minmax σε συνθήκες αβεβαιότητας ... 167
10.3.5 Παράδειγμα επιλογής απόφασης συμβιβασμού ή εκδίκασης ... 168
10.4 Δένδρο αποφάσεων σε συνθήκες κινδύνου ... 170

10.5 Αποφάσεις σε συνθήκες ανταγωνισμού ........ 171
10.5.1 Προϋποθέσεις λήψης απόφασης ........ 171
10.5.2 Παράδειγμα παιγνίου ........ 171
10.6 Κατηγορίες παιγνίων ........ 172
10.7 Προϋποθέσεις θεωρίας παιγνίων ........ 172
10.8 Έννοια της στρατηγικής ........ 173
10.8.1 Κυρίαρχες και υποδεέστερες στρατηγικές ........ 173
10.8.2 Παράδειγμα παιγνίου με υποδεέστερη στρατηγική ........ 173
10.8.3 Εντοπισμός υποδεέστερων στρατηγικών ........ 174
10.9 Αμιγής και μεικτή στρατηγική ........ 174
10.10 Επίλυση παιγνίων ........ 174
10.10.1 Εντοπισμός αμιγούς στρατηγικής (ισορροπίας) ........ 174
10.10.2 Εντοπισμός μεικτής στρατηγικής για πίνακα παιγνίου 2 × 2 ........ 175
10.11 Γραφική μέθοδος επίλυσης παιγνίων ........ 176
10.12 Εφαρμογές για κατανόηση ........ 177
10.12.1 Επιλογή συνεταίρου ........ 177
10.12.1.1 Παρουσίαση προβλήματος ........ 177
10.12.1.2 Λύση με αναμενόμενο όφελος ........ 177
10.12.1.3 Λύση με κριτήριο maxmin ........ 177
10.12.1.4 Λύση με κριτήριο maxmax ........ 178
10.12.1.5 Λύση με κριτήριο minmax ........ 178
10.12.2 Πρόσληψη βοηθών ........ 178
10.12.2.1 Λύση με κριτήριο αναμενόμενου οφέλους ........ 179
10.12.2.2 Λύση με κριτήριο maxmin ........ 179
10.12.2.3 Λύση με κριτήριο maxmax ........ 179
10.12.3 Συμβιβασμός σε αγωγή ........ 180
10.13 Ασκήσεις για κατανόηση - εφαρμογή ........ 182
10.13.1 Επιλογή νέας ιστοσελίδας ........ 182
10.13.2 Συμβιβασμός σε αγωγή ........ 182
10.13.3 Εντοπισμός ισορροπίας ........ 183
10.13.4 Εντοπισμός ισορροπίας ........ 183
10.13.5 Εντοπισμός μεικτής στρατηγικής για πίνακα παιγνίου 2 × 2 ........ 183
**Βιβλιογραφία/Αναφορές ........ 184**
**Κριτήρια αξιολόγησης ........ 185**

# Πίνακας συντομεύσεων-ακρωνυμίων

| | |
|---|---|
| ANOVA | Analysis of Variance |
| ARR | Accounting Rate of Return |
| EAR | Effective Annual Rate |
| IRR | Internal Rate of Return |
| LSD | Least Significant Difference |
| NPV | Net Present Value |
| PI | Profit Index |
| SPSS | Statistical Product and Service Solutions |
| STD | Standard |
| VIF | Variance Inflation Factor |
| ΔΙ.ΠΑ.Ε. | Διεθνές Πανεπιστήμιο της Ελλάδος |

# Πρόλογος

Όλες σχεδόν οι επιστημονικές έρευνες βασίζονται σε δεδομένα, δηλαδή σε πληροφορίες που έχουν στη διάθεσή τους οι ερευνητές για να μελετήσουν ένα φαινόμενο που τους ενδιαφέρει.

Τα δεδομένα ποικίλλουν και απαιτούν διαφορετική μέθοδο ανάλυσης ανάλογα με το είδος τους. Παρουσιάζονται στη συνέχεια οι πιο συχνά χρησιμοποιούμενες μέθοδοι και ο τρόπος συγγραφής μιας επιστημονικής εργασίας. Επίσης, παρουσιάζεται σύντομα η χρηματοοικονομική αξιολόγηση επενδύσεων, η θεωρία λήψης αποφάσεων με χρήση πιθανοτήτων και η θεωρία παιγνίων. Τα θέματα αυτά είναι χρήσιμα σε διάφορους επιστημονικούς κλάδους και θεωρήθηκε χρήσιμο να παρουσιαστούν σε μεταπτυχιακούς φοιτητές των κλάδων αυτών.

Το παρόν σύγγραμμα προέκυψε από την ανάγκη να μελετήσουν οι μεταπτυχιακοί φοιτητές οικονομικών και κοινωνικών κατευθύνσεων, βασικές κατευθύνσεις συγγραφής της μεταπτυχιακής τους εργασίας και τον τρόπο χρήσης των Πιθανοτήτων και της Στατιστικής στην ανάλυση των δεδομένων τους.

Πολλές ευχαριστίες αρμόζουν στην ομάδα Έργου Κάλλιππος+ – Ανοικτά Ακαδημαϊκά Ηλεκτρονικά Συγγράμματα για τη δυνατότητα συγγραφής του παρόντος και την ελεύθερη διάδοση και διακίνησή του. Το σύγγραμμα θα έχει πετύχει τον στόχο του αν παρακινήσει τον αναγνώστη να χρησιμοποιήσει τις προτεινόμενες μεθόδους σε πραγματικές εφαρμογές και να μελετήσει άλλα συγγράμματα που εμβαθύνουν στις αντίστοιχες μεθόδους.

# Εισαγωγή

Η Στατιστική είναι μια επιστήμη χρήσιμη σε όλες τις άλλες επιστήμες. Χρησιμοποιούνται μαθηματικά για την απόδειξη των θεωρημάτων της και την εφαρμογή των μεθόδων της. Αυτό προκαλεί κάποια αποστροφή σε όλους όσοι ασχολούνται με μη μαθηματικές επιστήμες και θεωρούν τη Στατιστική δύσκολη.

Στο παρόν σύγγραμμα γίνεται μια προσπάθεια να παρουσιαστούν οι κυριότερες μέθοδοι της Στατιστικής που εφαρμόζονται σε όλες σχεδόν τις άλλες επιστήμες. Παρουσιάζεται με απλό τρόπο η θεωρία, χωρίς μαθηματικές αποδείξεις, και ακολουθούν παραδείγματα εφαρμογής της. Δεν χρειάζεται προαπαιτούμε γνώση για τη μελέτη του βιβλίου, αλλά για την καλύτερη χρήση των μεθόδων καλό θα ήταν να έχει ο αναγνώστης ένα στατιστικό λογισμικό, ώστε να εφαρμόζει και μόνος του τις προτεινόμενες μεθόδους στα δικά του δεδομένα.

Τα δεδομένα είναι αυτά που έχουν όλη την απαραίτητη πληροφορία για ένα μελετώμενο φαινόμενο. Για να αποκαλυφθεί η πληροφορία αυτή, εφαρμόζεται η κατάλληλη στατιστική διαδικασία. Η μέθοδος πρέπει να ταιριάζει στην επεξεργασία των δεδομένων και όχι να προσαρμόζονται τα δεδομένα, ώστε να μπορεί να εφαρμοστεί μια συγκεκριμένη μέθοδος. Στο παρόν σύγγραμμα παρουσιάζονται οι στατιστικές διαδικασίες και μέθοδοι που αρμόζουν σε κάθε τύπο δεδομένων. Υπάρχουν πολλά χρήσιμα συγγράμματα Στατιστικής, τόσο σε απλή μορφή, όσο και με έμφαση σε κάποιες μεθόδους ή στη χρήση ενός στατιστικού λογισμικού, τα οποία μπορεί να μελετήσει ο αναγνώστης που επιθυμεί να εμβαθύνει σε κάποια μέθοδο.

Ακολουθούν στη συνέχεια κεφάλαια τόσο για τη χρήση στατιστικών μεθόδων, όσο και για την εφαρμογή των χρηματοοικονομικών στην αξιολόγηση επενδύσεων. Υπάρχει και ένα κεφάλαιο που αναφέρεται περιληπτικά στη θεωρία παιγνίων και στη χρήση της, κυρίως σε οικονομικά θέματα.

Στο Κεφάλαιο 1, αναλύεται η μεθοδολογία έρευνας και οι αρχές δεοντολογίας για τη συλλογή δεδομένων και πληροφοριών. Στο Κεφάλαιο 2, μελετώνται οι τρόποι συλλογής δεδομένων, δίνοντας έμφαση στις μεθόδους δειγματοληψίας και στη συλλογή απαντήσεων σε έρευνες με ερωτηματολόγιο. Στο Κεφάλαιο 3, παρουσιάζονται τα είδη των μεταβλητών και οι τρόποι περιγραφικής παρουσίασης των δεδομένων με πίνακες και διαγράμματα. Στο ίδιο κεφάλαιο αναλύονται οι στατιστικές κατανομές και η χρήση τους.

Στο Κεφάλαιο 4 εξετάζονται η εκτίμηση διαστήματος εμπιστοσύνης των άγνωστων παραμέτρων του πληθυσμού και ο έλεγχος υποθέσεων. Παρουσιάζονται παραμετρικοί και μη παραμετρικοί έλεγχοι υποθέσεων και οι συνθήκες για την εφαρμογή κάθε είδους ελέγχου. Σχολιάζεται το στατιστικό σφάλμα στην απόφαση του ελέγχου και εφαρμογές για την κατανόηση της διαδικασίας.

Στη συνέχεια, στα Κεφάλαια 5 και 6, αναλύονται η συσχέτιση μεταξύ δύο μεταβλητών και η παλινδρόμηση για δύο ή για περισσότερες μεταβλητές. Δίνεται περιγραφή των τρόπων αξιολόγησης της καταλληλότητας της παλινδρόμησης, καθώς και της εξαγωγής συμπερασμάτων από τα αποτελέσματά της.

Στην περίπτωση που τα δεδομένα περιγράφονται από πολλές ποιοτικές μεταβλητές για τις οποίες επιθυμούμε μία συνολική ανάλυση χωρίς εκ των προτέρων υποθέσεις και μοντέλα, ακολουθείται η μέθοδος της παραγοντικής ανάλυσης. Η μέθοδος αυτή περιγράφεται στο Κεφάλαιο 7 και ερμηνεύονται τα αποτελέσματά της σε πραγματικές απαντήσεις ερωτηματολογίων.

Στο τελευταίο μέρος του συγγράμματος, παρουσιάζονται τα Κεφάλαια 8, 9 και 10, για τη διαχρονική αξία του χρήματος, τους τρόπους αξιολόγησης επενδύσεων και τη θεωρία παιγνίων. Τα κεφάλαια αυτά δεν ανήκουν στη Στατιστική Επιστήμη, αλλά είναι πολύ χρήσιμα για την κατανόηση των οικονομικών δεδομένων που απαιτούνται για τη λήψη αποφάσεων.

Στο τέλος κάθε κεφαλαίου παρατίθενται εφαρμογές για την κατανόηση των θεμάτων που παρουσιάζονται, καθώς και ασκήσεις για εφαρμογή. Επίσης, υπάρχουν κριτήρια αξιολόγησης των γνώσεων μαζί με τις απαντήσεις τους.

Καταβλήθηκε προσπάθεια για τη σαφή και πλήρη παρουσίαση των θεμάτων, αλλά είναι αδύνατον να μην υπάρχουν λάθη ή παραλείψεις. Η συγγραφέας ζητά συγγνώμη για τις πιθανές αστοχίες, ελπίζοντας ότι το σύγγραμμα θα φανεί χρήσιμο στη μελέτη και κυρίως στην εφαρμογή των θεμάτων που αναλύει.

# Κεφάλαιο 1: Μεθοδολογία Επιστημονικής Έρευνας – Δεοντολογία

**Σύνοψη**
*Στο κεφάλαιο αυτό γίνεται μια εισαγωγή στα είδη των επιστημονικών ερευνών, στον τρόπο διεξαγωγής τους, στις αρχές δεοντολογίας, και παρουσιάζονται παραδείγματα για την καλύτερη κατανόησή τους.*

**Προαπαιτούμενη γνώση**
*Δεν απαιτείται.*

## 1.1 Τι είναι επιστημονική έρευνα

**Επιστήμη** είναι ένα σύνολο συστηματικών γνώσεων με στόχο να ερμηνεύει την ανθρώπινη συμπεριφορά και τα φυσικά και κοινωνικά φαινόμενα. Χρησιμοποιεί την κατασκευή μιας θεωρίας, ως άθροισμα εξηγητικών αρχών λογικά αλληλένδετων. Η θεωρία βασίζεται σε εμπειρικά δεδομένα, η συλλογή, η επεξεργασία και η ανάλυση τους περιλαμβάνονται στην **επιστημονική έρευνα** (Ρόντος & Παπάνης, 2006).

Η επιστημονική έρευνα είναι η μεθοδική αναζήτηση που διεξάγει κάποιος για να προσθέσει κάτι επιπλέον στις γνώσεις του και στις γνώσεις των άλλων, με την ανακάλυψη σημαντικών πραγμάτων ή απόψεων (Howard & Sharp, 1996). Η επιστημονική έρευνα όχι μόνο επαληθεύει, απορρίπτει, βελτιώνει ή τροποποιεί μία θεωρία, αλλά παράγει και νέες θεωρίες για την εξέλιξη των φυσικών και κοινωνικών φαινομένων, με σκοπό την πρόβλεψη προσεχών σταδίων αυτής της εξέλιξης (Καραγιώργος, 2002).

Τα χαρακτηριστικά της επιστημονικής έρευνας είναι η συστηματική συλλογή δεδομένων, η συστηματική ερμηνεία δεδομένων και η ύπαρξη σαφούς σκοπού για διαπίστωση πραγμάτων. Η έρευνα περιλαμβάνει εξήγηση της μεθόδου συλλογής δεδομένων, επιχειρηματολογία γιατί τα αποτελέσματα έχουν νόημα, εξήγηση περιορισμών και συγκεκριμένη προθεσμία ολοκλήρωσης (Saunderd et al., 2012).

### 1.1.1 Τι δεν είναι επιστημονική έρευνα

Κάποιες φορές χρησιμοποιείται καταχρηστικά η λέξη «έρευνα», αφήνοντας να εννοηθεί ότι πρόκειται για επιστημονική έρευνα, χωρίς να έχει τα χαρακτηριστικά της επιστημονικής έρευνας.

Δεν θεωρείται επιστημονική έρευνα η απλή συλλογή στοιχείων ή πληροφοριών χωρίς σαφή σκοπό, η αναδιοργάνωση και αναδιάταξη στοιχείων ή πληροφοριών χωρίς ερμηνεία. Εσωτερική δραστηριότητα με ελάχιστη ή καμία σχέση με την καθημερινότητα δεν ανήκει στην επιστημονική έρευνα. Η χρήση της λέξης «έρευνα» για την αύξηση του κύρους ενός προϊόντος ή μίας ιδέας, επίσης, δεν είναι επιστημονική έρευνα (Walliman, 2011).

### 1.1.2 Μέθοδοι και μεθοδολογία

Αρκετές φορές χρησιμοποιούνται για τον ίδιο λόγο οι λέξεις «μέθοδοι» και «μεθοδολογία». Η ακριβής σημασία των δύο αυτών λέξεων διαφέρει.

**Μέθοδοι** είναι οι τεχνικές και οι διαδικασίες που χρησιμοποιούνται για την απόκτηση και ανάλυση δεδομένων. **Μεθοδολογία** είναι η θεωρία του τρόπου πραγματοποίησης της έρευνας. Δηλαδή η περιγραφή και ανάλυση των μεθόδων που θα χρησιμοποιηθούν.

## 1.2 Είδη ερευνών

Η επιστημονική έρευνα διακρίνεται σε δύο μεγάλες κατηγορίες, στη *βασική έρευνα* και στην *εφαρμοσμένη έρευνα*. Στον Πίνακα 1.1 παρουσιάζονται οι κυριότερες διαφορές μεταξύ των δύο ειδών έρευνας.

Ανάλογα με το είδος των δεδομένων που συλλέγονται, διακρίνουμε την έρευνα σε ποσοτική (quantitative) και ποιοτική (qualitative). Η ποσοτική έρευνα βασίζεται στην καταμέτρηση εμφάνισης κάποιου φαινομένου. Από την άλλη, η ποιοτική έρευνα βασίζεται στην κατηγοριοποίηση και στην ερμηνεία των προσωπικών βιωμάτων και εμπειριών. Δίνει έμφαση στην εξήγηση και όχι στη συχνότητα εμφάνισης των φαινομένων.

**Πίνακας 1.1** *Χαρακτηριστικά βασικής και εφαρμοσμένης έρευνας.*

| **Βασική έρευνα** | **Εφαρμοσμένη έρευνα** |
|---|---|
| Διεύρυνση γνώσης | Κατανόηση συγκεκριμένων προβλημάτων |
| Διατύπωση γενικών αρχών που σχετίζονται με το αποτέλεσμα | Λύσεις σε πρόβλημα |
| Αξία ευρημάτων για την κοινωνία | Νέα γνώση για το πρόβλημα, πρακτική συνάφεια |
| Γίνεται από ερευνητές κυρίως σε πανεπιστήμια | Γίνεται από ανθρώπους σε διάφορες θέσεις |
| Επιλογή θέματος από ερευνητή | Στόχοι σε διαπραγμάτευση με τον δημιουργό της ιδέας |
| Ευέλικτα χρονικά περιθώρια | Αυστηρές προθεσμίες |

Δεν υπάρχει αντιπαράθεση μεταξύ ποσοτικής και ποιοτικής έρευνας. Στα κοινωνικά φαινόμενα τα ποσοτικά δεδομένα αποκαλύπτουν διαφαινόμενες τάσεις, ενώ τα ποιοτικά τις ερμηνεύουν, φωτίζοντας διαφορετικές πλευρές. Ο επιστήμονας ενσωματώνει στην ερμηνευτική διαδικασία και τις δύο πλευρές της έρευνας. Ανάλογα με τον τρόπο που γίνεται η έρευνα, φωτίζει ένα κομμάτι της πραγματικότητας, και έτσι πρέπει να βλέπουμε τα αποτελέσματά της σαν μέρος ενός άγνωστου συνόλου (Ρόντος & Παπάνης, 2006).

## 1.3 Τρόποι διεξαγωγής ερευνών

Για τη διεξαγωγή μιας επιστημονικής έρευνας έχει καθιερωθεί η **ερευνητική πορεία**, η οποία ξεκινάει με το θεωρητικό πλαίσιο (θεωρία) του θέματος, και ακολουθεί η διατύπωση των υποθέσεων. Οι υποθέσεις αφορούν στην πιθανή σχέση μεταξύ αντιλήψεων, βασίζονται στο θεωρητικό πλαίσιο και αποτελούν, στην ουσία, τους στόχους της έρευνας. Στη συνέχεια, επιλέγονται οι μεταβλητές που θα μελετηθούν για την επαλήθευση των υποθέσεων, καθώς και ο πληθυσμός που θα ερευνηθεί με τις επιλεγμένες μεταβλητές. Ακολουθούν η συλλογή δεδομένων των μεταβλητών, η ανάλυση των δεδομένων αυτών και η εξαγωγή συμπερασμάτων.

Για τη συλλογή των δεδομένων της επιστημονικής έρευνας υπάρχουν διάφορες τεχνικές. Οι συνηθέστερες από τις οποίες είναι

- η άμεση παρατήρηση,
- η συνέντευξη,
- η ανάλυση περιεχομένου,
- το πείραμα,
- η μελέτη περίπτωσης, και
- το ερωτηματολόγιο.

Οι τεχνικές αυτές μπορούν να χρησιμοποιηθούν τόσο στην ποιοτική όσο και στην ποσοτική έρευνα, αλλά κυρίως το ερωτηματολόγιο χρησιμοποιείται ως εργαλείο ποσοτικής έρευνας, ενώ οι άλλες τεχνικές χρησιμοποιούνται κυρίως στην ποιοτική έρευνα (Ρόντος & Παπάνης, 2006).

### 1.3.1 Ερευνητική πορεία

Τα στάδια της πορείας μιας έρευνας είναι:

- το θεωρητικό πλαίσιο, επισκόπηση της αντίστοιχης θεωρίας που θα βασισθεί η έρευνα,
- οι υποθέσεις που γίνονται από τον ερευνητή, αφορούν σε επαλήθευση της θεωρίας ή σε νέα θεωρία ή σε άλλες πιθανές σχέσεις,
- η επιλογή του πληθυσμού ή των αντικειμένων στα οποία θα γίνει η έρευνα,
- η συλλογή των δεδομένων,

- η ανάλυση των δεδομένων,
- η εξαγωγή των συμπερασμάτων.

Ανάλογα με το είδος των συμπερασμάτων που θα προκύψουν, πιθανώς να χρειαστούν νέες υποθέσεις και στη συνέχεια να επαναληφθεί η ερευνητική πορεία.

### 1.3.2 Βιβλιογραφική επισκόπηση

Για την παρουσίαση του θεωρητικού πλαισίου μιας έρευνας, παρουσιάζεται σύντομα η αντίστοιχη θεωρία. Για την κατανόηση των εννοιών που περιέχονται στην έρευνα θα πρέπει να αναφερθούν οι ορισμοί τους. Αναφέρονται, επίσης, προηγούμενες έρευνες που έχουν γίνει σχετικά με το μελετώμενο θέμα. Οι έρευνες αυτές παρατίθενται κριτικά, συντάσσονται τα πορίσματά τους και συγκρίνονται μεταξύ τους, παρουσιάζοντας τις ομοιότητες ή τις διαφορές τους. Σε κάθε αναφορά σε άλλες έρευνες γράφεται η πηγή τους σύντομα και αναλυτικά σε ειδικό κεφάλαιο, στο τέλος της εργασίας αναφέρονται όλες οι λεπτομέρειες των βιβλιογραφικών αναφορών.

Η βιβλιογραφική επισκόπηση έχει πολλά πλεονεκτήματα και για τον ερευνητή. Βοηθά στον προσδιορισμό και στη βελτίωση των ερωτημάτων και των στόχων της έρευνας, επισημαίνονται οι ερευνητικές δυνατότητες, λαμβάνονται ιδέες από προτάσεις για περαιτέρω έρευνα, αποφεύγονται επαναλήψεις εργασιών που έχουν γίνει, γίνονται γνωστές απόψεις και πτυχές της έρευνας. Η ανακάλυψη ερευνητικών προσεγγίσεων, στρατηγικών και κατάλληλων τεχνικών είναι αποτέλεσμα της βιβλιογραφικής επισκόπησης (Gall et al., 2006· Saunders et al., 2014).

Οι βιβλιογραφικές πηγές μπορεί να είναι πρωτογενείς, όπως άρθρα σε επιστημονικά περιοδικά, πρακτικά συνεδρίων, αναφορές, εκθέσεις, μπορεί να είναι δευτερογενείς, όπως εγχειρίδια, μονογραφίες ή ευρετήρια, μπορεί να είναι και τριτογενείς, όπως εγκυκλοπαίδειες, θεματικές βιβλιογραφίες ή οδηγοί στη χρήση ειδικευμένων βιβλιογραφιών.

Οι όροι αναζήτησης βιβλιογραφίας έχουν μεγάλη σημασία. Πρέπει να δίνεται ιδιαίτερη προσοχή στη σωστή ορθογραφία, στην κατάλληλη γλώσσα, στη σωστή ορολογία, στη χρήση τελεστών (AND, OR, *, ?) και σε εξειδικευμένες μηχανές αναζήτησης (π.χ. scholar.google.com, scopus).

## 1.4 Αρχές δεοντολογίας

Η επιστημονική έρευνα αποτελεί μέρος της ανθρώπινης ζωής και συμπεριφοράς. Όπως και σε κάθε πτυχή της κοινωνίας, επικράτησαν οι άγραφοι νόμοι που εκφράζουν την **ηθική** της κοινωνίας και κανόνες συμπεριφοράς που εκφράζουν το πώς πρέπει να γίνει κάτι και ονομάζονται **δεοντολογία**. Η ηθική και δεοντολογία της έρευνας επηρεάστηκαν από την ανθρώπινη συμπεριφορά των επιστημόνων αλλά και από το κοινωνικό πλαίσιο στο οποίο αυτοί δρουν και εργάζονται.

Επικράτησαν δύο αντικρουόμενες φιλοσοφικές θέσεις. Από τη μία πλευρά, η **δεοντολογική**, σύμφωνα με την οποία πρέπει να ακολουθούνται συγκεκριμένοι κανόνες που καθοδηγούν τη συμπεριφορά των ερευνητών. Δράση εκτός των κανόνων δεν δικαιολογείται.

Από την άλλη πλευρά, η **τελολογική**, σύμφωνα με την οποία η απόφαση για το αν μία συμπεριφορά είναι δικαιολογημένη καθορίζεται από τις συνέπειές της και όχι από προκαθορισμένους κανόνες. (Με απλά λόγια, η θέση αυτή εκφράζεται ως «ο σκοπός αγιάζει τα μέσα»).

Οι δεοντολογικές αρχές που γενικά εφαρμόζονται στη συλλογή δεδομένων είναι:

- η αποφυγή βλάβης, αμέλειας, κακών πρακτικών,
- το κοινό καλό,
- η ιδιωτικότητα όσων λαμβάνουν μέρος,
- η εθελοντική συμμετοχή και το δικαίωμα αποχώρησης,
- η ενήμερη συγκατάθεση των συμμετεχόντων,
- η διασφάλιση της εμπιστευτικότητας και της ανωνυμίας,
- η υπευθυνότητα στην ανάλυση των δεδομένων,

- η νομική συμμόρφωση στη διαχείριση δεδομένων,
- η διασφάλιση της ασφάλειας του ερευνητή.

Σε γενικές γραμμές οι ευρέως αποδεκτοί ηθικοί κανόνες επιστημονικής έρευνας μπορούν να εκφραστούν ως εξής για τον επιστήμονα, ο οποίος:

- είναι ελεύθερος και ανεξάρτητος από προκαταλήψεις ή συμφέροντα,
- δεν υπηρετεί καμία σκοπιμότητα, αλλά παρουσιάζει τα αντικειμενικά δεδομένα,
- ενθαρρύνει τις δημιουργικές δραστηριότητες στη διάθεση του κοινωνικού συνόλου,
- δεν προσπαθεί να προσηλυτίσει,
- δεν γίνεται οπαδός ή υπηρέτης κανενός ατόμου ή παράνομου θεσμού για αντικοινωνικούς σκοπούς,
- δεν είναι ρατσιστής, αλλά υπηρετεί το σύνολο της ανθρωπότητας,
- δεν χρησιμοποιεί βία ή απειλές, αλλά λύνει τις διαφορές με πολιτισμένο διάλογο,
- σέβεται την πολυφωνία και την ανεξαρτησία, δεν οικειοποιείται τα επιστημονικά ευρήματα των άλλων,
- είναι ανεκτικός, υποστηρίζει τον επιστημονικό ανταγωνισμό,
- παρουσιάζει τα ευρήματά του με συστηματικό τρόπο αποκαλύπτοντας τη μέθοδο,
- αναγνωρίζει τα σφάλματά του δημόσια,
- πιστεύει στη διαφάνεια και δηλώνει τις πηγές του,
- αποκτά την αναγνώριση και το κύρος με τη θετική συμβολή του στην πρόοδο,
- προσφέρει πρότυπο ή υπόδειγμα συμπεριφοράς στο κοινωνικό σύνολο,
- θέτει τον εαυτό του στην υπηρεσία όσων πάσχουν ή βρίσκονται σε μειονεκτική θέση.

## 1.5 Εφαρμογές επιστημονικής έρευνας

Οι επιστημονικές έρευνες που γίνονται περιλαμβάνουν συνήθως δύο μέρη: από τη διαδικασία του σχεδιασμού και τη διαδικασία της εκτέλεσης. Στη διαδικασία του σχεδιασμού προσδιορίζεται το πεδίο μελέτης, επιλέγεται το θέμα, η κατάλληλη προσέγγισή του, και διαμορφώνονται τα βήματα υλοποίησής του. Στη διαδικασία της εκτέλεσης συλλέγονται τα κατάλληλα δεδομένα, αναλύονται με τη βοήθεια των υπολογιστών και παρουσιάζονται τα ευρήματα της έρευνας.

### 1.5.1 Επιλογή θέματος από φοιτητές

Ειδικότερα, οι επιστημονικές έρευνες φοιτητών έχουν τα εξής χαρακτηριστικά σε ό,τι αφορά τους στόχους τους:

- επισκόπηση υφιστάμενης γνώσης,
- περιγραφή κατάστασης ή προβλήματος,
- δημιουργία καινοτομίας,
- ερμηνεία (αιτιατές σχέσεις, γενικεύσεις).

Επίσης, για τις έρευνες στο πλαίσιο ενός προπτυχιακού ή μεταπτυχιακού προγράμματος σπουδών, ισχύουν τα παρακάτω:

- Το θέμα έχει προκαθοριστεί.
- Πρέπει να ολοκληρωθεί σε προκαθορισμένη προθεσμία.
- Οι οικονομικοί πόροι είναι περιορισμένοι.
- Τα αποτελέσματα πρέπει να παρουσιάζονται με προκαθορισμένο τρόπο.

- Υποχρέωση αποδοχής συγκεκριμένου επιβλέποντος (Howard & Sharp, 1996).

Ευνοϊκοί παράγοντες για την επιτυχία μιας φοιτητικής επιστημονικής έρευνας είναι το εσωτερικό κίνητρο του φοιτητή, οι ατομικές ικανότητές του, η προηγούμενη συγγραφή άρθρων, η παρακολούθηση επιστημονικών συνεδρίων και η ποιότητα επίβλεψης. Αρνητικοί παράγοντες είναι ο κακός αρχικός σχεδιασμός, η υπερφόρτωση με υποχρεώσεις και η ανεπάρκεια επίβλεψης.

Σημαντικό ρόλο στη διενέργεια επιστημονικής έρευνας από φοιτητές κατέχει η επιλογή του θέματος και του επιβλέποντος. Η πιθανή πρόταση για το ερευνητικό θέμα μπορεί να επιλεγεί από άλλες ερευνητικές εργασίες ή επιστημονικά άρθρα, από βιβλία (μέσω της βιβλιογραφικής επισκόπησης), από επικοινωνία με ειδικούς, από συζήτηση με πιθανούς χρήστες, από συζήτηση με συναδέλφους ή από την επικαιρότητα των μέσων μαζικής ενημέρωσης. Η επιλογή του επιβλέποντος, συνήθως, βασίζεται στη γνώμη άλλων φοιτητών για τον χαρακτήρα του, στο αν είναι προσιτός, στο γνωστικό αντικείμενο που εξειδικεύεται, καθώς και στην ικανότητά του στην ερευνητική μεθοδολογία.

Φυσικά, ο κυρίαρχος λόγος στην επιτυχία είναι η ικανότητα και καταλληλότητα του ερευνητή για το συγκεκριμένο θέμα. Πριν από την επιλογή θα πρέπει ο ερευνητής να απαντήσει στις παρακάτω ερωτήσεις:

- Το θέμα με ενδιαφέρει πραγματικά;
- Έχω τις δεξιότητες για να εκπληρώσω τον στόχο;
- Η έρευνα μπορεί να ολοκληρωθεί εντός του διαθέσιμου χρόνου;
- Το θέμα θα εξακολουθεί να είναι επίκαιρο όταν ολοκληρωθεί;
- Θα μπορώ να έχω πρόσβαση στα δεδομένα που χρειάζονται;
- Το θέμα σχετίζεται άμεσα με τη θεωρία;
- Μπορώ να διατυπώσω με σαφήνεια τους στόχους;
- Θα φέρει νέες χρήσιμες γνώσεις;
- Σχετίζεται με τους στόχους που έχω θέσει για τη σταδιοδρομία μου;

### 1.5.2 Στατιστική έρευνα

Η έρευνα η οποία στηρίζεται σε αριθμητικά δεδομένα και διενεργείται συνήθως με ερωτηματολόγιο ονομάζεται στατιστική έρευνα ή εμπειρική έρευνα. Δεν αναφέρεται απλώς στην τυχαία επιλογή παρατηρήσεων, αλλά στη συστηματική μεθοδολογία για την παραγωγή αμερόληπτων και ακριβών στατιστικών δεδομένων και πληροφοριών. Η προσοχή δεν συγκεντρώνεται μόνο στην ανάλυση των δεδομένων, αλλά και στην ποιότητά τους, η οποία εξαρτάται από τους τρόπους συλλογής και επεξεργασίας τους (Ρόντος & Παπάνης, 2006).

Σε αυτή την περίπτωση θα πρέπει να δοθεί ιδιαίτερη προσοχή στο ερωτηματολόγιο της έρευνας. Οι ερωτήσεις του ερωτηματολογίου θα πρέπει να είναι σωστά ορισμένες και να έχει οριστεί το ενδιαφερόμενο κοινό στο οποίο θα απευθύνονται οι ερωτήσεις. Επίσης, η δειγματοληψία (επιλογή δείγματος στο οποίο θα δοθεί το ερωτηματολόγιο) θα πρέπει να πραγματοποιηθεί αποτελεσματικά και με όρους αντικειμενικούς/ αξιόπιστους. Στη συνέχεια, εφόσον συγκεντρωθούν τα απαντημένα ερωτηματολόγια, θα διεξαχθούν τα σχετικά στατιστικά αποτελέσματα και θα αναλυθούν τα δεδομένα με τη βοήθεια σχετικής στατιστικής ανάλυσης. Το πιο ουσιαστικό κομμάτι της εμπειρικής έρευνας είναι η παράθεση των συμπερασμάτων που προκύπτουν από τα στατιστικά αποτελέσματα.

### 1.5.3 Παράδειγμα σχεδιασμού στατιστικής έρευνας

Ως παράδειγμα σχεδιασμού μιας στατιστικής έρευνας ας θεωρηθεί μια στατιστική έρευνα για το θέμα «Ικανοποίηση πελατών από λογιστικές υπηρεσίες». Ο κύριος στόχος της έρευνας αυτής είναι η μελέτη της ικανοποίησης πελατών από λογιστικές υπηρεσίες. Υπάρχουν όμως διάφοροι επιμέρους στόχοι στους οποίους θα στραφεί η έρευνα όπως:

- ικανοποίηση ιδιωτών πελατών,
- ικανοποίηση πελατών επιχειρήσεων,

- ικανοποίηση από τακτικές λογιστικές εργασίες,
- ικανοποίηση από λογιστικές εργασίες που γίνονται μία φορά τον χρόνο,
- και διάφοροι άλλοι πιθανοί στόχοι για το ίδιο θέμα.

Ο προσδιορισμός του ερευνητέου πληθυσμού πιθανώς να αφορά τις επιχειρήσεις που έχουν δικό τους λογιστήριο ή τις επιχειρήσεις που χρησιμοποιούν εξωτερικό λογιστικό γραφείο.

Στη συνέχεια, αφού αποφασιστεί σε ποιον πληθυσμό θα γίνει η έρευνα, θα πρέπει να καθοριστούν το μέγεθος του δείγματος και η δειγματοληπτική μέθοδος για την επιλογή του. Περιγράφονται το είδος ερευνητικής μεθοδολογίας που θα ακολουθηθεί και ο σχεδιασμός των ερευνητικών εργαλείων - ερωτηματολογίων. Αυτό θα πρέπει να συνάδει με τους επιμέρους στόχους της έρευνας, αλλά και με τα χαρακτηριστικά του πληθυσμού στον οποίο θα απευθυνθεί.

Για τη σύνταξη των ερωτήσεων απαιτούνται συντομία, σαφήνεια, συνοχή, καλή δομή, καλή εμφάνιση, αντικειμενικότητα, πληρότητα, παράθεση όλων των δυνατών απαντήσεων, οδηγίες συμπλήρωσης και στοιχεία αυτού που διεξάγει την έρευνα. Μετά από τη σύνταξη του ερωτηματολογίου, αποφασίζεται ο τρόπος συλλογής των απαντήσεων, που μπορεί να είναι έντυπος ή προφορικός ή μέσω διαδικτύου. Απαντώνται τα ερωτήματα ποιος, πότε, πού και πώς. Γίνεται πιλοτική δοκιμή συλλογής απαντήσεων, για να εντοπιστούν πιθανά σφάλματα, διορθώνονται και συντάσσεται το χρονοδιάγραμμα της έρευνας, ενώ εκτιμάται και το πιθανό κόστος της.

### 1.5.4 Ανάλυση και αξιολόγηση των δεδομένων της στατιστικής έρευνας

Μετά από τη συλλογή των δεδομένων, αυτά καταγράφονται, ελέγχονται, γίνεται η ομαδοποίηση πληροφοριών και η ανάλυσή τους. Ερευνώνται πιθανές σχέσεις μεταξύ τους και ερμηνεύονται ως προς τους αρχικούς στόχους της έρευνας. Όλα τα παραπάνω καταγράφονται με λεπτομέρειες και παρουσιάζονται με πίνακες και διαγράμματα, ώστε να είναι πιο ευανάγνωστα και ελκυστικά. Παρουσιάζεται διεξοδικά η ανάλυση και ο τρόπος εξαγωγής των συμπερασμάτων. Τα συμπεράσματα γράφονται με απλά λόγια, ώστε να είναι εύκολη η ανάγνωσή τους και από μη ειδικούς.

### 1.5.5 Συγγραφή εργασίας

Για να είναι επιτυχής μια επιστημονική εργασία, θα πρέπει να προσφέρει στον αναγνώστη και στην επιστήμη κάτι νέο ή συμπληρωματικό, ή ενδιαφέρον σε σχέση με τη μέχρι τώρα βιβλιογραφία. Διότι δεν θα έχει νόημα απλώς να αναπαραχθεί η υπάρχουσα γνώση που είναι σχετική με το θέμα, χωρίς να προστεθεί κάτι νέο ή συμπληρωματικό από τον συγγραφέα. Κάτι τέτοιο σαφώς είναι δύσκολο και χρειάζεται κόπο και προσεκτική έρευνα για να επιτευχθεί.

Κύριος στόχος του συγγραφέα, ο οποίος είναι στην ουσία και το ζητούμενο μίας επιστημονικής έρευνας, είναι να εντοπίσει εκείνα τα σημεία στην υπάρχουσα βιβλιογραφία στα οποία θα μπορέσει να αναπτύξει μια διαφορετική επιχειρηματολογία. Σκοπός είναι να καταφέρει να αναλύσει τα πράγματα μέσα από μία διαφορετική οπτική γωνία και να διατυπώσει μία διαφορετική μεθοδολογία, ώστε τελικά να μπορέσει να διαφοροποιηθεί από τους υπόλοιπους συγγραφείς, προσθέτοντας ουσιαστικά το δικό του λιθαράκι στην έρευνα και στην ανάπτυξη της επιστήμης του αντικειμένου του.

Θα πρέπει πριν ξεκινήσει την εργασία να απαντηθούν τα εξής ερωτήματα:

- Τι ισχύει πάνω στο θέμα που πρόκειται να ερευνήσει και τι είναι γνωστό έως σήμερα από άλλες παραπλήσιες εργασίες/μελέτες;
- Ποια είναι τα κοινά τους σημεία, αλλά και πού εστιάζονται οι διαφορές αυτών των ερευνών/ μελετών;
- Η εργασία πρέπει να έχει σαφή σκοπό, να παρουσιάζεται στην αρχή και να φαίνεται στα επιμέρους τμήματά της. Γενικότητες πρέπει να αποφεύγονται, και κάθε άποψη που παρατίθεται πρέπει να τεκμηριώνεται είτε από στοιχεία της βιβλιογραφίας, είτε από στατιστικές έρευνες, είτε από τεκμηριωμένη προσωπική εμπειρία.

Σύντομα η οδηγία για τη συγγραφή θα ήταν «Παρουσιάστε τι πρόκειται να παρουσιάσετε, μετά παρουσιάστε το και, τέλος, πείτε τι έχετε παρουσιάσει».

Η εισαγωγή μίας επιστημονικής εργασίας περιέχει περιγραφή ερευνητικού αντικειμένου, αναφορά στο ερευνητικό πρόβλημα, εξήγηση των λόγων που οδήγησαν στην επιλογή του συγκεκριμένου θέματος, τον σκοπό και τη συνεισφορά της εργασίας, καθώς και τη δομή της με σύντομη περιγραφή.

Στο κύριο τμήμα της εργασίας αποτυπώνεται η μεθοδολογία έρευνας και ανάλυσης του θέματος και αναπτύσσονται όλες οι σημαντικές, κατά τον συγγραφέα, πλευρές του.

Στα συμπεράσματα περιλαμβάνεται μία σύνοψη των περιεχομένων της εργασίας, σχολιάζονται τα αποτελέσματα που προέκυψαν, ερμηνεύονται, συγκρίνονται με συμπεράσματα άλλων ανάλογων εργασιών. Παρουσιάζονται τρόποι ώστε τα συμπεράσματα να φανούν χρήσιμα, γίνονται προτάσεις χρήσης τους.

Χρήσιμο είναι στο τέλος της έρευνας να παρατίθενται οι περιορισμοί της και πιθανές αδυναμίες της ή άλλες δυσκολίες που προέκυψαν. Επίσης, δίνονται και κατευθύνσεις για ενδεχόμενη μελλοντική έρευνα σε σχέση με το μελετώμενο θέμα.

Ιδιαίτερη σημασία θα πρέπει να δοθεί και στη μορφοποίηση της εργασίας με σελιδαρίθμηση, ορθογραφικό έλεγχο, ενιαίο μέγεθος γραμματοσειράς σε όλο το κείμενο, λογικό μέγεθος φράσεων, χρήση συντομογραφιών μόνο εφόσον εξηγήθηκαν στην αρχή, παράθεση υποσημειώσεων, πινάκων, εικόνων με αρίθμηση.

Τυχόν παραρτήματα παρατίθενται στο τέλος και αφορούν υπάρχουσα γνώση, η οποία αναφέρεται στην εργασία, στοιχεία που χρησιμοποιήθηκαν ή λεπτομερή αποτελέσματα για περαιτέρω μελέτη κάποιου ενδιαφερόμενου.

Οι συχνότερες αστοχίες που συναντώνται σε μία εργασία και πρέπει να εξαλείφονται είναι οι αδυναμίες στη διαμόρφωση του κειμένου, τα γραμματικά, συντακτικά, ορθογραφικά σφάλματα, οι ελλείψεις στη δομή και στην πληρότητα περιεχομένου, τα λάθη σε παραπομπές και βιβλιογραφικές αναφορές, τα προβλήματα διαφοροποιήσεων στο ύφος και στιλ γραφής.

### 1.5.6 Προφορική παρουσίαση εργασίας

Μετά από τη συγγραφή μιας εργασίας, κάποιες φορές απαιτείται και η προφορική παρουσίασή της σε ακροατήριο ή σε επιτροπές κρίσης. Τα σημεία στα οποία πρέπει να δοθεί έμφαση είναι η προσέλκυση του ενδιαφέροντος του κοινού με τη σαφή αναφορά στους στόχους, η σύντομη επισκόπηση της εργασίας μέσα σε συγκεκριμένο χρόνο, ώστε να μην κουράζει, η χρήση οπτικών βοηθημάτων (διαφανειών) που βοηθούν στο να μην ξεχαστούν βασικά σημεία και ταυτόχρονα ενισχύουν την κατανόηση του ακροατηρίου.

Καλό είναι, επίσης, να γίνεται μια κάποια προετοιμασία από τον παρουσιαστή για τον χρόνο και τη γλώσσα παρουσίασης, δίνοντας προσοχή στην ένταση της φωνής, στον ρυθμό, στο στιλ και στο ύφος της παρουσίασης.

Τα σημεία που συνήθως αξιολογούνται σε μια προφορική παρουσίαση είναι κατά πόσο έχουν αφομοιωθεί οι γραπτοί ισχυρισμοί της εργασίας, κατά πόσο έχει αποκτήσει ο παρουσιαστής εξειδίκευση στο αντικείμενο, κατά πόσο είναι ειλικρινής και έχει αυτοπεποίθηση για την εργασία του, ποια είναι η δεξιοτεχνία του και πόσο καλά έχει προετοιμάσει την παρουσίαση και γνωρίζει τις απαντήσεις σε πιθανές ερωτήσεις που θα του υποβάλουν.

# Βιβλιογραφία/Αναφορές

Howard, K., & Sharp, J., (1996). *Η επιστημονική μελέτη*. Αθήνα: Gutenberg.

Καραγιώργος, Δ. (2002). *Μεθοδολογία Έρευνας στις Επιστήμες της Αγωγής*. Αθήνα: Σαββάλας.

Ρόντος, Κ., & Παπάνης, Ε. (2006). *Στατιστική έρευνα*. Αθήνα: Ι. Σιδέρης.

Saunders, M., Lewis, P., & Thornhill, A. (2014). *Μέθοδοι έρευνας στις επιχειρήσεις & την οικονομία*. Αθήνα: Δίσιγμα.

Walliman, N. (2011). *Research Methods the Basics*. Routledge Publications. https://doi.org/10.4324/9780203836071

# Κριτήρια αξιολόγησης

## Κριτήριο αξιολόγησης 1

**Ποια είναι η διαφορά της μεθοδολογίας από τις μεθόδους;**

**Απάντηση**
**Μέθοδοι** είναι οι τεχνικές και οι διαδικασίες που χρησιμοποιούνται για την απόκτηση και ανάλυση δεδομένων. **Μεθοδολογία** είναι η θεωρία του τρόπου πραγματοποίησης της έρευνας. Δηλαδή η περιγραφή και ανάλυση των μεθόδων που θα χρησιμοποιηθούν.

## Κριτήριο αξιολόγησης 2

**Ποια είναι τα χαρακτηριστικά της επιστημονικής έρευνας;**

**Απάντηση**
Τα χαρακτηριστικά της επιστημονικής έρευνας είναι η συστηματική συλλογή δεδομένων, η συστηματική ερμηνεία δεδομένων και η ύπαρξη σαφούς σκοπού για διαπίστωση πραγμάτων.

## Κριτήριο αξιολόγησης 3

**Ποιες είναι οι τεχνικές διεξαγωγής της επιστημονικής έρευνας;**

**Απάντηση**
Οι συνηθέστερες τεχνικές είναι: η άμεση παρατήρηση, η συνέντευξη, η ανάλυση περιεχομένου, το πείραμα, η μελέτη περίπτωσης και το ερωτηματολόγιο.

## Κριτήριο αξιολόγησης 4

**Ποιες είναι οι διαφορές μεταξύ ποιοτικής και ποσοτικής επιστημονικής έρευνας;**

**Απάντηση**
Η ποσοτική έρευνα βασίζεται στην καταμέτρηση εμφάνισης κάποιου φαινομένου. Η ποιοτική έρευνα βασίζεται στην κατηγοριοποίηση και στην ερμηνεία των προσωπικών βιωμάτων και εμπειριών. Δίνει έμφαση στην εξήγηση και όχι στη συχνότητα εμφάνισης των φαινομένων.

## Κριτήριο αξιολόγησης 5

**Ποια είναι τα στάδια μιας στατιστικής έρευνας;**

**Απάντηση**
Αρχικά ο σχεδιασμός - προετοιμασία της έρευνας, στη συνέχεια η διεξαγωγή της και κατόπιν η επεξεργασία των δεδομένων, η παρουσίαση και ανάλυση των αποτελεσμάτων και η έκθεση με τα τελικά συμπεράσματα.

## Κριτήριο αξιολόγησης 6

**Πώς θα περιγράφατε με μία φράση τη δομή μιας επιστημονικής εργασίας;**

**Απάντηση**
Παρουσιάστε τι πρόκειται να παρουσιάσετε, ύστερα παρουσιάστε το και, τέλος, πείτε τι έχετε παρουσιάσει.

# Κεφάλαιο 2: Τρόποι Συλλογής Δεδομένων – Δειγματοληψία

**Σύνοψη**
*Στο κεφάλαιο παρουσιάζονται οι διάφορες μέθοδοι συλλογής στατιστικών δεδομένων, οι τρόποι επιλογής δείγματος, και αναφέρονται παραδείγματα και εφαρμογές για την καλύτερη κατανόησή τους.*

**Προαπαιτούμενη γνώση**
*Δεν απαιτείται.*

## 2.1 Τρόποι συλλογής δεδομένων

Καθημερινά και σε πολλές περιστάσεις της ζωής μας, ακούμε τη λέξη «δεδομένα». «Προσωπικά δεδομένα», «δεδομένα κινητής συσκευής», «επιχειρηματικά δεδομένα», «ιατρικά δεδομένα» κ.λπ. Η καταγραφή παρατηρήσεων και δεδομένων ξεκίνησε πολύ χωρίς από τον άνθρωπο, όταν παρατηρούσε τις κινήσεις των άστρων, την παραγωγή της γης, την καταμέτρηση των εμπορευμάτων του. Τα δεδομένα πληθαίνουν και ακολουθούν την ανθρώπινη πορεία εξέλιξης. Στις μέρες μας, με τη μεγάλη ανάπτυξη όλων των επιστημών, έχουμε τεράστιο όγκο δεδομένων, τα λεγόμενα μεγάλα δεδομένα (Big Data) και την εμφάνιση νέας επιστήμης, αυτής της Ανάλυσης Μεγάλων Δεδομένων. Με το μεγάλο πλήθος δεδομένων αναπτύχθηκε ιδιαίτερα και ο κλάδος της Πληροφορικής, κυρίως των Βάσεων Δεδομένων.

Σε όλους τους τομείς της καθημερινής ζωής παράγονται δεδομένα τα οποία είναι χρήσιμα σε διάφορες επιστήμες. Δεδομένα καταγράφονται για την υγεία, την οικονομία, την καταναλωτική συμπεριφορά, την απήχηση ενός τραγουδιού ή μίας διαφήμισης, για την ικανοποίηση του πελάτη, τον τρόπο αναζήτησης στο διαδίκτυο, την εξιχνίαση εγκλημάτων, την ανακάλυψη του συγγραφέα ενός κειμένου, για την εύρεση χρονολογίας αρχαιολογικών ευρημάτων, για την πάταξη της παράνομης διακίνησης, και σε πολλές άλλες περιπτώσεις. Η επιστήμη της Στατιστικής είναι απαραίτητη για την εξαγωγή συμπερασμάτων από τα δεδομένα. Για τον λόγο αυτό, έχει εφαρμογές σε όλες τις επιστήμες και έχει ενταχθεί σε όλα τα προγράμματα σπουδών των πανεπιστημίων.

### 2.1.1 Ορισμοί

Η επιστήμη της Στατιστικής βασίζεται σε δεδομένα, τα οποία παρουσιάζει και αναλύει με στόχο την εξαγωγή χρήσιμων συμπερασμάτων. Τα δεδομένα είναι η βάση της, βρίσκονται σε πολλές διαφορετικές πηγές, μετριούνται με διαφορετικούς τρόπους, και μπορεί τα ίδια δεδομένα να έχουν διαφορετική ερμηνεία ανάλογα με τον χρόνο και τον τόπο που εμφανίζονται. Παρατίθενται στη συνέχεια κάποια παραδείγματα δεδομένων και κάποιοι χρήσιμοι ορισμοί σχετικοί με τα δεδομένα και τον τρόπο συλλογής τους.

Για την υγεία, καταγράφονται δεδομένα αριθμητικά, όπως η πίεση, το ποσοστό εμφάνισης ενός στοιχείου, αλλά και δεδομένα μη αριθμητικά, όπως ποιες αρρώστιες έχει περάσει κάποιος, το φύλο του, η γενετική του προδιάθεση κ.ά.

Για την οικονομία, κυρίως, τα δεδομένα είναι μέτρηση ποσότητας ή αξίας αλλά έχουν διαφορετική σημασία σε διαφορετικές χρονικές στιγμές, όπως λ.χ. η τιμή μίας μετοχής στο χρηματιστήριο, όταν την αγοράζει ή όταν την πουλάει κάποιος. Και στην οικονομία υπάρχουν μη ποσοτικά δεδομένα, όπως η ποιότητα ενός προϊόντος, η εμπιστοσύνη της αγοράς, η φήμη μιας επιχείρησης.

Για την καταναλωτική συμπεριφορά συνδυάζονται μη αριθμητικά δεδομένα (φύλο, οικογενειακή κατάσταση, τόπος διαμονής) με πλήθος, είδος και συχνότητα αγορών.

Για την απήχηση ενός τραγουδιού ή μίας διαφήμισης, για την ικανοποίηση του πελάτη, για τον τρόπο αναζήτησης στο διαδίκτυο, καταγράφονται ο χρόνος και η συχνότητα επανάληψης αλλά και οι εκφράσεις των συναισθημάτων του πελάτη ή καταναλωτή.

Για την εξιχνίαση εγκλημάτων, για την ανακάλυψη του συγγραφέα ενός κειμένου, για την εύρεση χρονολογίας αρχαιολογικών ευρημάτων, για την πάταξη της παράνομης διακίνησης, κυρίως αναλύονται και καταγράφονται λέξεις και εκφράσεις ή άλλες συνδεόμενες ενέργειες και συμπεριφορές.

Από τα παραπάνω παραδείγματα γίνεται κατανοητό ότι τα δεδομένα υπάρχουν παντού σε πολλές μορφές και αφορούν κάθε φαινόμενο της ανθρώπινης ζωής. Ονομάζουμε δεδομένα οποιαδήποτε πληροφορία έχουμε στη διάθεσή μας για κάποιο φαινόμενο που μελετάμε. Στατιστική έρευνα είναι οποιοσδήποτε τρόπος χρησιμοποιείται για να συλλέξουμε δεδομένα και να τα καταγράψουμε συστηματικά. Τα δεδομένα που συλλέγονται με στατιστική έρευνα ονομάζονται και παρατηρήσεις. Μεταβλητές ονομάζονται οι μελετώμενες ιδιότητες ενός φαινομένου. Τιμές μεταβλητής ονομάζονται οι αριθμοί ή οι εκφράσεις που αντιπροσωπεύουν τις διαφορετικές καταστάσεις μίας μεταβλητής.

Κάθε στατιστική έρευνα αφορά φαινόμενο για έναν πληθυσμό. Η στατιστική έρευνα για την πώληση ενός βιβλίου μελετάει τον αριθμό αντιτύπων που αγοράζει κάποιος, τον τρόπο που γνώρισε το βιβλίο, το φύλο του, τον σκοπό αγοράς του και αναφέρεται σε όλους τους αγοραστές του συγκεκριμένου βιβλίου, οι οποίοι αποτελούν τον πληθυσμό της έρευνας. Η στατιστική έρευνα για τα βιβλιοπωλεία μιας περιοχής, καταγράφει το είδος βιβλιοπωλείου, τον χρόνο ίδρυσής του, τον ιδιοκτήτη του, το πλήθος των πελατών του, το πλήθος ημερήσιων πωλήσεων, την αξία των πωλούμενων βιβλίων και άλλα. Πληθυσμός, στην περίπτωση αυτή, είναι όλα τα βιβλιοπωλεία της περιοχής.

Πληθυσμός ονομάζεται το σύνολο ατόμων ή αντικειμένων στα οποία αναφέρονται οι παρατηρήσεις μας. Τα στοιχεία του συνόλου αυτού ονομάζονται στατιστικές μονάδες ή άτομα ή αντικείμενα και συνήθως είναι πάρα πολλά. Η Στατιστική καταγράφει τις ιδιότητες του πληθυσμού και με την ανάλυσή τους, μετατρέπει τα δεδομένα σε χρήσιμα συμπεράσματα για τη λήψη αποφάσεων. Οι αποφάσεις πρέπει να ληφθούν στον κατάλληλο χρόνο. Αυτό συνεπάγεται ότι η καταγραφή δεδομένων και η ανάλυση πρέπει να γίνει γρήγορα για να είναι έγκαιρη. Για να είναι έγκυρη, θα πρέπει να ανταποκρίνεται στην πραγματικότητα, και αυτό συμβαίνει όταν μελετάται όσο το δυνατόν μεγαλύτερο μέρος του πληθυσμού. Οι δύο αυτές προϋποθέσεις, ταχύτητα και πληρότητα παλιότερα συγκρούονταν, αλλά με την ανάπτυξη της Πληροφορικής (επεξεργασία μεγάλου όγκου σε μικρό χρόνο) μπορούν να συμπορευτούν. Πάλι, όμως, η ανάλυση όλου του πληθυσμού απαιτεί μεγάλο υπολογιστικό χρόνο και μεγάλο κόστος, ενώ καθυστερεί την εξαγωγή των συμπερασμάτων. Για τους λόγους αυτούς, αντί για όλο τον πληθυσμό χρησιμοποιείται ένα μέρος του, κατάλληλα επιλεγμένο, που ονομάζεται δείγμα.

Δείγμα ονομάζεται ένα μέρος του μελετώμενου πληθυσμού. Χρησιμοποιείται το δείγμα αντί για τον πληθυσμό για τους παρακάτω λόγους:

- Απαιτεί λιγότερο χρόνο.
- Έχει μικρότερο κόστος.
- Εξάγονται συντομότερα τα αποτελέσματα.

## 2.1.2 Απογραφή, συνεχής καταγραφή, δειγματοληψία

Οι μέθοδοι συλλογής δεδομένων ποικίλλουν και κυρίως εξαρτώνται από τον χρόνο συλλογής και από το αν επιλέγεται όλος ο μελετώμενος πληθυσμός ή μόνο ένα δείγμα του. Τρεις είναι οι κύριες μέθοδοι συλλογής: απογραφή, δειγματοληψία, συνεχής καταγραφή.

### 2.1.2.1 Απογραφή

Σε μια δεδομένη χρονική στιγμή, μας ενδιαφέρει να μελετήσουμε και να συλλέξουμε δεδομένα από όλο τον πληθυσμό. Αυτό συμβαίνει σε συλλογή δεδομένων από το κράτος όπως η απογραφή πληθυσμού, αλλά και σε καταγραφή των δεδομένων από μία επιχείρηση, όπως η απογραφή εμπορευμάτων.

Τα πλεονεκτήματα της απογραφής είναι η ακρίβεια των στοιχείων και η πληρότητά τους, αφού συλλέγονται όλα. Επίσης, τα στοιχεία που συλλέγονται μπορούν να χρησιμοποιηθούν σε άλλες έρευνες.

Τα μειονεκτήματα της απογραφής είναι το μεγάλο κόστος συλλογής και η χρονοβόρα επεξεργασία όλων των στοιχείων, αφού απαιτούνται πολλά εξειδικευμένα άτομα. Επίσης, λόγω της χρονοβόρας επεξεργασία τους, κάποιες φορές τα αποτελέσματα δεν είναι επίκαιρα.

Η απογραφή δεν μπορεί να πραγματοποιηθεί στην περίπτωση που η καταγραφή των στατιστικών μεταβλητών απαιτεί την καταστροφή των μονάδων πληθυσμού, όπως συμβαίνει στη μελέτη διάρκειας ζωής λαμπτήρων ή στην αντοχή ζωνών ασφαλείας σε σύγκρουση αυτοκινήτων.

### 2.1.2.2 Δειγματοληψία

Επειδή πολλές φορές είναι δύσκολο, ακόμη και αδύνατον, να συλλέξουμε δεδομένα από όλο τον πληθυσμό που μας ενδιαφέρει σε μια χρονική στιγμή, περιοριζόμαστε στο να συλλέξουμε δεδομένα από ένα μικρό μέρος του πληθυσμού το οποίο ονομάζεται δείγμα.

Τα πλεονεκτήματα της δειγματοληψίας είναι η ευκολία, η ταχύτητα συλλογής των στοιχείων και το μικρό κόστος της διαδικασίας, αφού δεν συλλέγονται όλα τα δεδομένα αλλά μόνο ένα μέρος τους. Επίσης, τα στοιχεία που συλλέγονται δίνουν γρήγορα επίκαιρα αποτελέσματα.

Τα μειονεκτήματα της δειγματοληψίας είναι ότι δεν μπορούμε να έχουμε ακριβή και πλήρη εικόνα των δεδομένων του πληθυσμού, αφού πολλές φορές παρατηρούνται λάθη συλλογής αντιπροσωπευτικών στοιχείων. Για τον λόγο αυτό, έχουν προταθεί πολλές και διαφορετικές μέθοδοι επιλογής δείγματος, ανάλογα με τα χαρακτηριστικά του μελετώμενου πληθυσμού. Στον Πίνακα 2.1 αναφέρονται με συντομία οι κυριότερες μέθοδοι επιλογής δείγματος.

### 2.1.2.3 Συνεχής καταγραφή

Συλλέγουμε δεδομένα από όλο τον πληθυσμό, αλλά όλα τα δεδομένα δεν συλλέγονται την ίδια χρονική στιγμή. Κάθε φορά που συμβαίνει μια αλλαγή, καταγράφονται ο χρόνος και το είδος της αλλαγής (π.χ. δημοτολόγια, λογιστήρια).

Τα πλεονεκτήματα της συνεχούς καταγραφής είναι η ακρίβεια των στοιχείων και η πληρότητά τους, αφού συλλέγονται όλα τη χρονική στιγμή που συμβαίνουν. Έτσι, τα αποτελέσματα δίνουν επίκαιρες πληροφορίες για οποιαδήποτε χρονική στιγμή.

Τα μειονεκτήματα της συνεχούς καταγραφής είναι το μεγάλο κόστος συλλογής, η απαίτηση για καλή και συνεχή οργάνωση και η ανάγκη για πολλά εξειδικευμένα άτομα.

## 2.2 Ερωτηματολόγιο, δομή, περιεχόμενο

Για την οργανωμένη και εύκολη συλλογή στοιχείων χρησιμοποιείται το ερωτηματολόγιο, ιδιαίτερα σε έρευνες που αφορούν απόψεις ατόμων. Κάθε άτομο απαντά σε συγκεκριμένες ερωτήσεις της έρευνάς μας, μέσω ενός εντύπου ή ηλεκτρονικού ερωτηματολογίου, και καταγράφονται οι απαντήσεις του. Οι ερωτήσεις του ερωτηματολογίου πρέπει να συνάδουν με τον σκοπό και τον στόχο της έρευνας, χωρίς να είναι ασαφείς ή αδιάκριτες.

Αρχικά, προσδιορίζονται ο σκοπός και οι στόχοι της έρευνας, καθώς και ο πληθυσμός στον οποίο αυτή θα απευθυνθεί. Κατόπιν, συντάσσονται οι κατάλληλες ερωτήσεις, που θα είναι εύκολο να κατανοηθούν και να απαντηθούν από τον συγκεκριμένο πληθυσμό.

### 2.2.1 Μορφή ερωτηματολογίου

Το έντυπο ή ηλεκτρονικό ερωτηματολόγιο δεν είναι πυκνογραμμένο, ούτε πολυσέλιδο, πρέπει να είναι διατυπωμένο απλά και να προσελκύει το ενδιαφέρον του ερωτώμενου. Στην αρχή του υπάρχει σημείωση για τον σκοπό του ερωτηματολογίου, για τον τρόπο που θα χρησιμοποιηθούν οι απαντήσεις των ερωτηματολογίου και για τον εμπιστευτικό τους χαρακτήρα.

Στο πρώτο μέρος υπάρχουν κάποιες ερωτήσεις που αφορούν στοιχεία χαρακτηριστικά της ταυτότητας του ατόμου που απαντά (όχι το ονοματεπώνυμό του), αλλά ίσως το φύλο, η ηλικία, ο τόπος κατοικίας του και άλλα ανάλογα με τους στόχους της έρευνας. Στο δεύτερο μέρος αναφέρονται οι κύριες ερωτήσεις για την έρευνα, οι οποίες μπορούν να χωρίζονται σε ομάδες, ανάλογα με το περιεχόμενό τους. Στο τέλος του ερωτηματολογίου υπάρχει και σημείωση που ευχαριστεί των ερωτώμενο για τη βοήθειά του στη συμπλήρωση του ερωτηματολογίου και στη διεξαγωγή της έρευνας.

### 2.2.2 Διατύπωση ερωτήσεων

Το περιεχόμενο και η διατύπωση των ερωτήσεων θα πρέπει να μην θίγουν τον ερωτώμενο και να κεντρίζουν την προσοχή του, για να δώσει με ειλικρίνεια τις απαντήσεις του.

Οι ερωτήσεις του ερωτηματολογίου διατυπώνονται με απλότητα και σαφήνεια, χωρίς να θίγουν τα προσωπικά δεδομένα του ερωτώμενου. Στην περίπτωση που αναφέρονται και πιθανές απαντήσεις για επιλογή, αυτές δεν πρέπει να επηρεάζουν τον ερωτώμενο να επιλέξει συγκεκριμένη απάντηση. Ακόμη πρέπει να υπάρχουν όλες οι πιθανές δυνατές απαντήσεις, καθώς κι η δυνατότητα συμπλήρωσης κάποιας άλλης διαφορετικής απάντησης από τον ερωτώμενο, αν αυτός το επιθυμεί.

## 2.3 Σφάλμα επιλογής δείγματος

Στην περίπτωση που δεν γίνεται απογραφή ή συνεχής καταγραφή, επιλέγεται δείγμα. Ποιο είναι όμως το σωστό δείγμα και τι μέγεθος θα πρέπει αυτό να έχει; Η χρήση ικανοποιητικού δείγματος οδηγεί στην εξαγωγή σωστών συμπερασμάτων, αλλά το πολύ μεγάλο δείγμα έχει μεγάλο κόστος.

Η σωστή επιλογή μεγέθους δείγματος εξαρτάται από την ακρίβεια σφάλματος $\delta$ που γίνεται ανεκτό. Όσο μικρότερο το $\delta$, τόσο μεγαλύτερο το μέγεθος δείγματος που πρέπει να επιλεγεί.

Το δείγμα θα πρέπει να είναι αντιπροσωπευτικό του πληθυσμού. Ένα μεγάλο δείγμα που δεν αντιπροσωπεύει τον πληθυσμό, δίνει λανθασμένα συμπεράσματα παρά το μεγάλο μέγεθός του. Για παράδειγμα, αν στόχος μίας έρευνας είναι να καταγραφεί η ποδοσφαιρική ομάδα που υποστηρίζουν οι φίλαθλοι μιας πόλης και επιλεγεί δείγμα από έναν μεγάλο αγώνα μεταξύ δύο ομάδων της πόλης, αυτό δεν θα είναι αντιπροσωπευτικό. Κυρίως οι φίλαθλοι των δύο ομάδων που συμμετείχαν στον αγώνα θα συμπεριληφθούν στο δείγμα, και όχι οι φίλαθλοι των υπολοίπων ομάδων της πόλης. Στην περίπτωση αυτή, έχουμε σφάλμα εξαιτίας του τρόπου δειγματοληψίας και όχι εξαιτίας του μεγέθους του δείγματος.

## 2.4 Είδη δειγματοληψίας

Υπάρχουν διάφορες μέθοδοι συλλογής δείγματος, ώστε να έχουμε όσο γίνεται πληρέστερη και ακριβέστερη εικόνα της πραγματικότητας με τα λιγότερα λάθη (Φαρμάκης, 2015). Στον Πίνακα 2.1 παρουσιάζονται με συντομία οι κυριότερες μέθοδοι επιλογής δείγματος.

**Πίνακας 2.1** *Μέθοδοι επιλογής δείγματος.*

| **Μέθοδος δειγματοληψίας** | **Κύρια χαρακτηριστικά** |
|---|---|
| Απλή τυχαία | Κάθε στοιχείο του πληθυσμού έχει την ίδια πιθανότητα να επιλεγεί στο δείγμα. Η επιλογή γίνεται χρησιμοποιώντας τυχαίους αριθμούς. |
| Κατά στρώματα | Χρησιμοποιείται για ανομοιογενή πληθυσμό, ο οποίος είναι χωρισμένος σε διαφορετικά στρώματα. Επιλέγονται στοιχεία με απλή, τυχαία δειγματοληψία από κάθε στρώμα. |
| Κατά ομάδες | Χρησιμοποιείται για πληθυσμό, ο οποίος είναι χωρισμένος σε παρόμοιες ομάδες. Επιλέγεται ως δείγμα μία ή περισσότερες ομάδες. |
| Επιφανειακή | Χρησιμοποιείται για πληθυσμό, ο οποίος είναι χωρισμένος σε γεωγραφικές επιφάνειες. Επιλέγονται στοιχεία στο δείγμα από περιοχές ή από οικοδομικά τετράγωνα. |
| Συστηματική | Χρησιμοποιείται όταν ο πληθυσμός είναι αριθμημένος και ταξινομημένος. Επιλέγεται ένα στοιχείο στην τύχη και μετά τα υπόλοιπα στοιχεία του δείγματος με συστηματική σειρά. Π.χ. το 3ο, το 23ο, το 43ο κ.λπ. |
| Ποσοστιαία | Χρησιμοποιείται για πληθυσμό για τον οποίο διαθέτουμε λεπτομερή στοιχεία ως προς κάποια χαρακτηριστικά του. Το δείγμα επιλέγεται με αναλογία ποσοστών, ώστε να υπάρχουν σε αυτό όλα τα χαρακτηριστικά με την ίδια αναλογία που υπάρχουν και στον πληθυσμό. |
| Κατευθυνόμενη | Βασίζεται σε υποκειμενικά κριτήρια του ερευνητή, ο οποίος πρέπει να διαθέτει αντίστοιχη εμπειρία και καλή γνώση του πληθυσμού. |
| Ευκολίας | Δεν υπάρχει προφανής κανόνας οργάνωσης. Οι περιπτώσεις για το δείγμα επιλέγονται στην τύχη, είτε επειδή είναι εύκολα διαθέσιμες, είτε πιο βολικές. |

## 2.5 Παραδείγματα – Εφαρμογές

### 2.5.1 Αξιολόγηση ερωτηματολογίου

Σχολιάστε τη μορφή, το περιεχόμενο και τη διατύπωση των ερωτήσεων του ερωτηματολογίου της Εικόνας 2.1. Ποιος είναι ο στόχος της έρευνας; Τι βελτιώσεις θα προτείνατε;

Φύλο: Άνδρας ☐ Γυναίκα ☐
Ηλικία : …………..
Οικογενειακή Κατάσταση: Άγαμος ☐ Έγγαμος ☐ Χήρος ☐ Διαζευγμένος ☐
Αριθμός ατόμων οικογενείας : ……………..
Επίπεδο Μόρφωσης: Αγράμματος ☐ Α΄βάθμια ☐ Β΄βάθμια ☐ Γ΄βάθμια ☐
Ποια είναι η γνώμη σου για τη δημοτικοποίηση των αστικών συγκοινωνιών;
Πρέπει να γίνει ☐
Δεν πρέπει να γίνει☐
Δεν έχω γνώμη ☐

**Εικόνα 2.1** *Παράδειγμα ερωτηματολογίου.*

### 2.5.2 Δημιουργία ερωτηματολογίου

Προτείνετε το περιεχόμενο ενός ερωτηματολογίου για τις παρακάτω έρευνες:

- προτίμηση οδοντόκρεμας του πληθυσμού της Ελλάδας,
- λόγοι διακοπής ή μη διακοπής καπνίσματος,
- ικανοποίηση από τις σπουδές των φοιτητών.

### 2.5.3 Συλλογή στοιχείων

Εξηγήστε με ποια μέθοδο συλλέγονται τα παρακάτω στατιστικά δεδομένα:

- τα δεδομένα πληθυσμού των διαφόρων πόλεων της Ελλάδας,
- τα δεδομένα προτίμησης οδοντόκρεμας του πληθυσμού της Ελλάδας,
- τα δεδομένα για το πλήθος των καπνιστών της Ελλάδας,
- τα δεδομένα για την αύξηση του Εθνικού Εισοδήματος και της Συνολικής Κατανάλωσης,
- τα δεδομένα για το μέγεθος σεισμών της Ελλάδας,
- τα δεδομένα για τον χρόνο εξυπηρέτησης ενός πελάτη σε μια τράπεζα,
- τα δεδομένα για το πλήθος των ελαττωματικών προϊόντων σε μια βιομηχανία,
- τα δεδομένα για τη διάρκεια ζωής ηλεκτρικών λαμπτήρων,
- τα δεδομένα για την αντοχή των ζωνών ασφαλείας σε αυτοκινητιστικά ατυχήματα,
- τα δεδομένα για τους θανάτους σε μια πόλη της Ελλάδας,
- τα δεδομένα για την πρόβλεψη των εκλογικών αποτελεσμάτων.

# Βιβλιογραφία/Αναφορές

Φαρμάκης, Ν. (2015). Δειγματοληψία και εφαρμογές [Προπτυχιακό εγχειρίδιο]. Κάλλιπος, Ανοικτές Ακαδημαϊκές Εκδόσεις. https://hdl.handle.net/11419/4840

# Κριτήρια αξιολόγησης

## Κριτήριο αξιολόγησης 1

**Τι είναι πληθυσμός και τι δείγμα στη στατιστική έρευνα;**

**Απάντηση**
Πληθυσμός ονομάζεται το σύνολο ατόμων ή αντικειμένων στα οποία αναφέρονται οι παρατηρήσεις μας, ενώ δείγμα είναι ένα μέρος του πληθυσμού.

## Κριτήριο αξιολόγησης 2

**Πώς πρέπει να διατυπώνονται οι ερωτήσεις ενός ερωτηματολογίου;**

**Απάντηση**
Το περιεχόμενο και η διατύπωση των ερωτήσεων θα πρέπει να μην θίγουν τον ερωτώμενο και να κεντρίζουν την προσοχή του. Οι ερωτήσεις του ερωτηματολογίου διατυπώνονται με απλότητα και σαφήνεια χωρίς να θίγουν τα προσωπικά δεδομένα του ερωτώμενου.

## Κριτήριο αξιολόγησης 3

**Πότε χρησιμοποιείται η δειγματοληψία κατά ομάδες και πότε κατά στρώματα;**

**Απάντηση**
Η δειγματοληψία κατά ομάδες χρησιμοποιείται για πληθυσμό, ο οποίος είναι χωρισμένος σε παρόμοιες ομάδες. Η δειγματοληψία κατά στρώματα χρησιμοποιείται για πληθυσμό, ο οποίος είναι χωρισμένος σε ανομοιογενή στρώματα.

# Κεφάλαιο 3: Περιγραφή Δεδομένων – Παρουσίαση με Παραμέτρους

**Σύνοψη**
*Στο κεφάλαιο αυτό γίνεται μια εισαγωγή στους τρόπους περιγραφής των δεδομένων, στα είδη των στατιστικών μεταβλητών, στους τρόπους μέτρησης τους και στους τρόπους παρουσίασής τους με πίνακες, διαγράμματα, στατιστικές εκθέσεις και με τις παραμέτρους τους. Στόχος δεν είναι η αναλυτική παρουσίασή τους αλλά η επιλογή του κατάλληλου τρόπου παρουσίασης ανάλογα με το είδος των δεδομένων και τους σκοπούς της έρευνας. Αναφέρονται, επίσης, οι θεωρητικές κατανομές πιθανοτήτων και παρουσιάζονται παραδείγματα και εφαρμογές για την καλύτερη κατανόησή τους.*

**Προαπαιτούμενη γνώση**
*Δεν απαιτείται.*

## 3.1 Είδη μεταβλητών

Για τη στατιστική μελέτη ενός φαινομένου απαιτείται η ύπαρξη ενός συνόλου δεδομένων, το οποίο σχετίζεται με το υπό μελέτη φαινόμενο. Αυτό το σύνολο δεδομένων καταγράφει τις διάφορες ιδιότητες των αντικειμένων ή ατόμων του φαινομένου. Κάθε ιδιότητα αντιστοιχεί σε μία στατιστική μεταβλητή.

Όταν μελετώνται οι διατροφικές συνήθειες του πληθυσμού μιας πόλης, οι ιδιότητες που καταγράφονται για κάθε άτομο είναι η ηλικία, το φύλο, πόσα γεύματα καταναλώνει τη μέρα, πόσο συχνά καταναλώνει κρέας, πόσο συχνά λαχανικά, αν παίρνει φάρμακα, αν καπνίζει κ.ά. Δηλαδή, για τη μελέτη των διατροφικών συνηθειών συλλέγονται δεδομένα για τις στατιστικές μεταβλητές ηλικία, φύλο, πλήθος ημερήσιων γευμάτων, εβδομαδιαία κατανάλωση κρέατος, κατανάλωση λαχανικών, λήψη φαρμάκων, άτομο καπνιστής και διάφορες άλλες μεταβλητές που σχετίζονται με τις συνήθειες διατροφής.

Για τη μελέτη των επιχειρήσεων λιανικών πωλήσεων, οι μεταβλητές που θα καταγραφούν για κάθε επιχείρηση θα είναι η έδρα της επιχείρησης, τα έτη λειτουργίας της, το μέγεθός της, το πλήθος υπαλλήλων της, το αν έχει κάνει διαφήμιση, η διακόσμησή της κ.ά.

Οι στατιστικές μεταβλητές έχουν διαφορετικές τιμές σε κάθε άτομο ή αντικείμενο, και οι τιμές τους μπορεί να είναι αριθμοί (σε μονάδες μέτρησης), μπορεί να είναι λέξεις ή προτάσεις (όπως ΝΑΙ/ΟΧΙ, τόπος, γνώμη για κάτι). Κάθε στατιστική μεταβλητή μετριέται με τον ίδιο τρόπο για όλα τα άτομα ή αντικείμενα. Αν οι τιμές της αντιστοιχούν σε αριθμούς, ονομάζεται **ποσοτική** μεταβλητή. Αν οι τιμές της απεικονίζονται με λέξεις ή εκφράσεις, ονομάζεται **ποιοτική** μεταβλητή.

Οι μεταβλητές που αναφέρθηκαν προηγουμένως για τις συνήθειες διατροφής, ηλικία, πλήθος γευμάτων, εβδομαδιαία συχνότητα κατανάλωσης κρέατος είναι ποσοτικές, αφού στις τιμές τους αντιστοιχούν αριθμοί. Στις μεταβλητές φύλο, λήψη φαρμάκων, άτομο καπνιστής αντιστοιχούν λέξεις για τις τιμές τους. Αυτές οι μεταβλητές είναι ποιοτικές μεταβλητές. Στον Πίνακα 3.1 παρουσιάζονται οι κατηγορίες μεταβλητών με παραδείγματα.

**Πίνακας 3.1** *Κατηγορίες μεταβλητών.*

| **Μεταβλητές** | **Κατηγορίες μεταβλητών** | **Χαρακτηριστικά** | **Παράδειγμα** |
|---|---|---|---|
| ποσοτικές | συνεχείς | μπορούν να πάρουν και δεκαδικές τιμές | βάρος |
| | διακριτές | παίρνουν μόνο ακέραιες τιμές | πλήθος παιδιών |
| ποιοτικές | κατηγορικές | οι δυνατές τιμές τους δεν μπορούν να μπουν σε διάταξη | καταγωγή |
| | διατάξιμες | οι δυνατές τιμές τους μπορούν να μπουν σε διάταξη | προτίμηση |
| | Διχοτομικές ή και 0-1 | οι δυνατές τιμές τους είναι μόνο δύο, «Ναι ή Όχι», «Υπάρχει ή Δεν Υπάρχει» κ.ά. | φύλο |

### 3.1.1 Κλίμακες μέτρησης ποσοτικών μεταβλητών

Οι ποσοτικές μεταβλητές μετριούνται είτε με κάποια μονάδα μέτρησης, όπως έτη, μήνες, κιλά, εκατοστά κ.λπ., είτε με έναν ακέραιο αριθμό που αντιστοιχεί σε πλήθος. Στην πρώτη περίπτωση οι τιμές τους μπορεί να είναι και δεκαδικοί αριθμοί, ενώ στη δεύτερη είναι μόνο ακέραιοι αριθμοί.

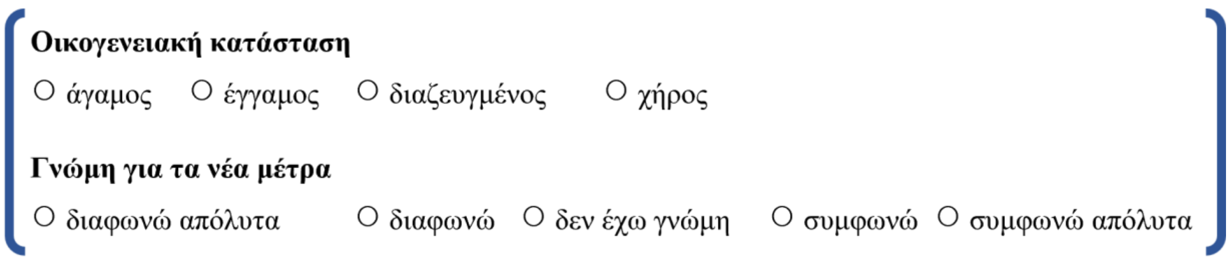
Ποσοτικές μεταβλητές με τιμές που μπορεί να είναι και δεκαδικοί αριθμοί, ονομάζονται **συνεχείς** μεταβλητές. Μπορούν να πάρουν οποιαδήποτε τιμή σε ένα διάστημα μέτρησής τους. Ανάλογα με τις ιδιότητες της μονάδας μέτρησής τους, διακρίνουμε τις εξής κλίμακες μέτρησης (Γναρδέλλης, 2019):

- Κλίμακα λόγου ή αναλογίας (ratio scale). Σε κλίμακα αναλογίας μετριέται ο χρόνος, το χρήμα, το βάρος το ύψος, το μήκος κ.ά. Οι τιμές έχει νόημα να αφαιρεθούν αλλά και να διαιρεθούν μεταξύ τους.
- Κλίμακα διαστήματος (interval scale). Στην κλίμακα αυτή μετριούνται βαθμολογία, μοριοδότηση, θερμοκρασία κ.ά. Οι τιμές έχει νόημα να αφαιρεθούν αλλά δεν έχει νόημα να διαιρεθούν μεταξύ τους.
- Κλίμακα ιεράρχησης (ordinal scale). Στην κλίμακα αυτή μετριούνται σειρά προτίμησης, κατάταξη διαγωνιζομένων, κλίμακα διαφωνίας ή συμφωνίας με κάποιο θέμα κ.ά. Οι τιμές δεν έχει νόημα να αφαιρεθούν αλλά ούτε και να διαιρεθούν μεταξύ τους.

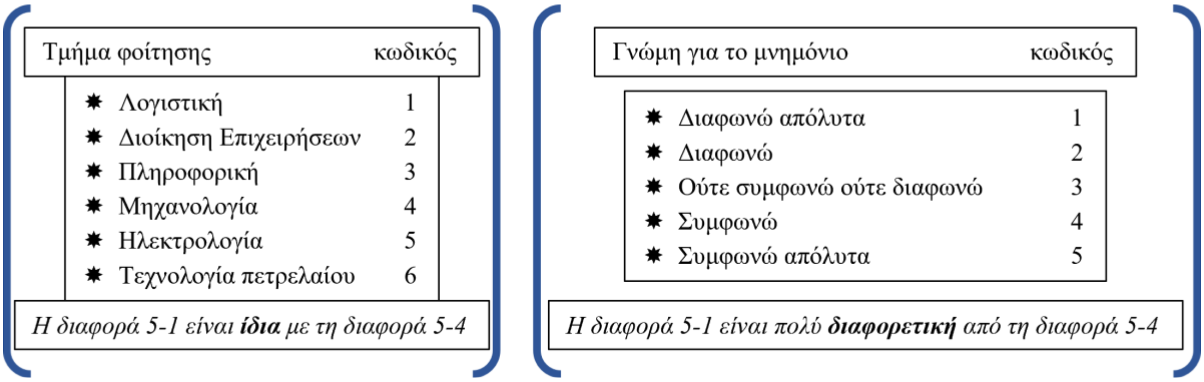
Ποσοτικές μεταβλητές με τιμές που είναι μόνο ακέραιοι αριθμοί, ονομάζονται **ασυνεχείς** μεταβλητές. Δεν μπορούν να πάρουν οποιαδήποτε τιμή σε ένα διάστημα μέτρησης τους. Οι μεταβλητές αυτές ονομάζονται και **διακριτές** μεταβλητές, επειδή οι τιμές τους είναι διακεκριμένες. Το πλήθος των παιδιών ενός ατόμου είναι διακριτή μεταβλητή. Το πόσες φορές την εβδομάδα ταξιδεύει, επίσης είναι διακριτή μεταβλητή. Το πλήθος των υπαλλήλων μιας επιχείρησης είναι διακριτή μεταβλητή.

### 3.1.2 Ποιοτικές μεταβλητές

Οι τιμές των ποιοτικών μεταβλητών είναι λέξεις ή εκφράσεις. Αν οι τιμές μιας ποιοτικής μεταβλητής μπορούν να καταταχθούν από τη χαμηλότερη προς την υψηλότερη, η ποιοτική μεταβλητή ονομάζεται **διατάξιμη** (ordinal). Αν οι τιμές της ποιοτικής μεταβλητής αντιστοιχούν σε κατηγορίες χωρίς κάποια διάταξη, ονομάζεται **κατηγορική** μεταβλητή (nominal). Παράδειγμα τιμών διατάξιμης ποιοτικής μεταβλητής είναι δεν συμφωνώ καθόλου, συμφωνώ λίγο, συμφωνώ πολύ. Παράδειγμα τιμών της κατηγορικής μεταβλητής, χώρα καταγωγής είναι Ελλάδα, Γερμανία, Γαλλία, Καναδάς, κ.λπ. Στην Εικόνα 3.1 υπάρχουν ερωτήσεις που αφορούν δύο ιδιότητες του ερωτώμενου ατόμου και είναι ποιοτικές μεταβλητές. Η οικογενειακή κατάσταση είναι μια κατηγορική μεταβλητή. Η γνώμη για τα νέα μέτρα είναι μια διατάξιμη ποιοτική μεταβλητή.

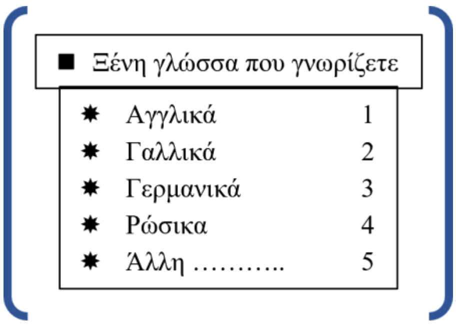



**Εικόνα 3.1** *Παράδειγμα δύο ερωτήσεων για μέτρηση ποιοτικών μεταβλητών.*




**Εικόνα 3.2** *Παράδειγμα κωδικοποίησης ποιοτικών μεταβλητών: κατηγορικής (αριστερά) και διάταξης (δεξιά).*

Επειδή είναι δύσκολος ο χειρισμός των λέξεων ή των εκφράσεων με ηλεκτρονικό υπολογιστή, συνήθως αντιστοιχεί ένας κωδικός αριθμός σε κάθε τιμή ποιοτικής μεταβλητής και ο κωδικός αυτός καταγράφεται στον υπολογιστή (Εικόνα 3.2). Δεν πρέπει όμως να συγχέονται οι κωδικοί αριθμοί που αντιστοιχούν στις τιμές με ποσοτικές τιμές και να θεωρούνται ποσοτικές οι ποιοτικές μεταβλητές. Υπάρχουν κάποιες περιπτώσεις διατάξιμων μεταβλητών που οι κωδικοί έχει νόημα να αφαιρεθούν μεταξύ τους και να υπολογιστεί η διαφορά τους που αντιστοιχεί σε διαφορά των ιδιοτήτων της ποιοτικής μεταβλητής (Εικόνα 3.2).




**Εικόνα 3.3** *Παράδειγμα για ποιοτική μεταβλητή με πολλαπλές τιμές.*

### 3.1.2.1 Μεταβλητές με πολλαπλές τιμές

Συνήθως μια ποιοτική μεταβλητή έχει μόνο μία τιμή για κάθε άτομο ή αντικείμενο. Υπάρχει όμως και η περίπτωση μία ποιοτική μεταβλητή να έχει πολλές τιμές για το ίδιο άτομο. Παράδειγμα, αν ερωτηθεί κάποιος ποια ξένη γλώσσα γνωρίζει, η απάντησή του μπορεί να είναι αγγλικά, γαλλικά και γερμανικά. Στην ερώτηση της Εικόνας 3.3 μπορούν να απαντηθούν από 0 έως και 5 επιλογές. Τέτοιες μεταβλητές ονομάζονται μεταβλητές με πολλαπλές απαντήσεις και χρειάζεται ιδιαίτερη προσοχή η επεξεργασία και παρουσίασή τους. Συνήθως, αναλύονται σε απλούστερες ποιοτικές μεταβλητές που δέχονται μία τιμή. Στο παράδειγμα της ξένης γλώσσας, θα μπορούσαν να καταγραφούν μεταβλητές, όπως γνωρίζετε αγγλικά, γνωρίζετε γαλλικά, γνωρίζετε γερμανικά, γνωρίζετε άλλη ξένη γλώσσα κ.λπ.

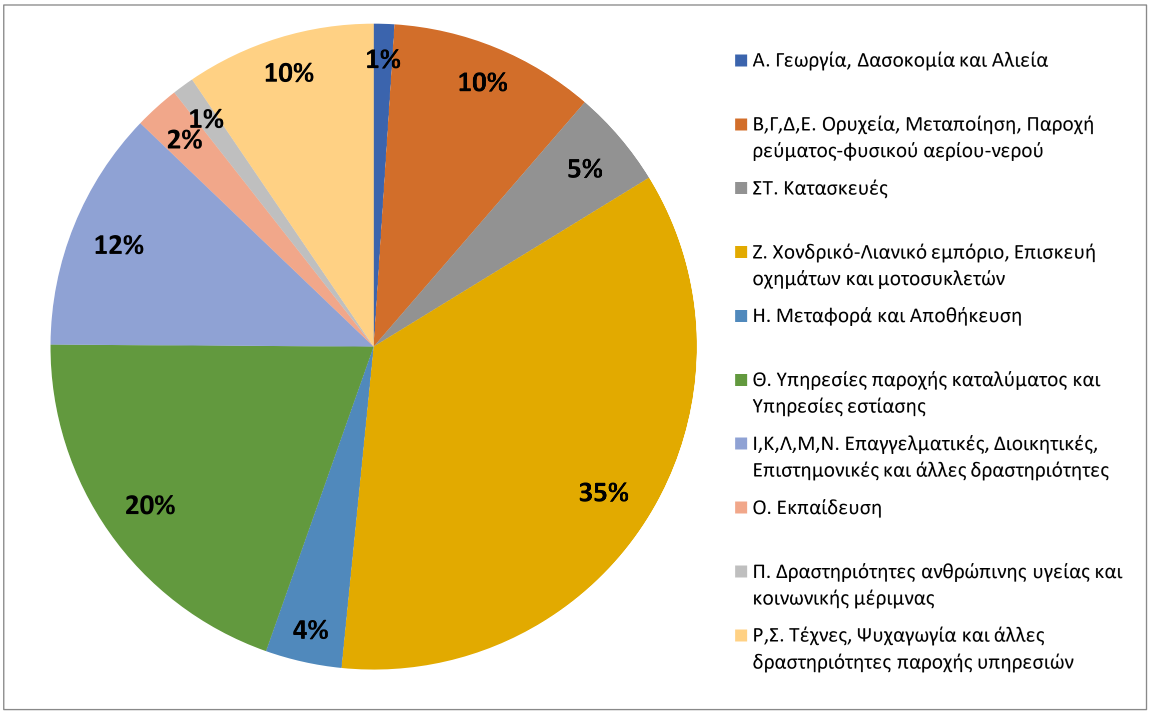
## 3.2 Περιγραφική παρουσίαση δεδομένων

Μετά από τη συλλογή και καταγραφή των δεδομένων, ακολουθεί η παρουσίασή τους, ώστε να φανούν τα κύρια χαρακτηριστικά τους. Η παρουσίαση των στατιστικών δεδομένων γίνεται με πίνακες, διαγράμματα ή με στατιστικές εκθέσεις που συνδυάζουν πίνακες και διαγράμματα με σχόλια για τα δεδομένα και συγκρίσεις (Κώστογλου & Αντωνίου, 2021). Ένα παράδειγμα παρουσίασης με πίνακα φαίνεται στον Πίνακα 3.2 και το αντίστοιχο διάγραμμα εμφανίζεται στην Εικόνα 3.4. Ένα παράδειγμα σύντομης στατιστικής έκθεσης από μια εφημερίδα απεικονίζεται στην Εικόνα 3.5.

**Πίνακας 3.2** *Παρουσίαση δεδομένων με πίνακα.*

| Κλάδος | Πλήθος επιχειρήσεων | Ποσοστό επιχειρήσεων |
|---|---|---|
| Α. Γεωργία, Δασοκομία και Αλιεία | 37 | 1,0% |
| Β,Γ,Δ,Ε. Ορυχεία, Μεταποίηση, Παροχή ρεύματος - φυσικού αερίου - νερού | 370 | 10,3% |
| ΣΤ. Κατασκευές | 176 | 4,9% |
| Ζ. Χονδρικό-Λιανικό εμπόριο, Επισκευή οχημάτων και μοτοσυκλετών | 1269 | 35,3% |
| Η. Μεταφορά και Αποθήκευση | 137 | 3,8% |
| Θ. Υπηρεσίες παροχής καταλύματος και Υπηρεσίες εστίασης | 707 | 19,7% |
| Ι,Κ,Λ,Μ,Ν. Επαγγελματικές, Διοικητικές, Επιστημονικές και άλλες δραστηριότητες | 433 | 12,1% |
| Ο. Εκπαίδευση | 81 | 2,3% |
| Π. Δραστηριότητες ανθρώπινης υγείας και κοινωνικής μέριμνας | 39 | 1,1% |
| Ρ,Σ. Τέχνες, Ψυχαγωγία και άλλες δραστηριότητες παροχής υπηρεσιών | 341 | 9,5% |

Ανάλογα με το είδος των μεταβλητών των δεδομένων έχουμε διαφορετικό είδος διαγράμματος. Τα πιο συνηθισμένα διαγράμματα είναι το ραβδόγραμμα και το κυκλικό διάγραμμα για ποιοτικές μεταβλητές, και το ιστόγραμμα για την παρουσίαση ποσοτικών μεταβλητών. Άλλα είδη διαγραμμάτων είναι το χρονοδιάγραμμα για την απεικόνιση μιας ποσοτικής μεταβλητής στη διάρκεια του χρόνου ή το χαρτόγραμμα για την απεικόνιση σε γεωγραφική επιφάνεια, το θηκόγραμμα και άλλα ειδικά διαγράμματα.

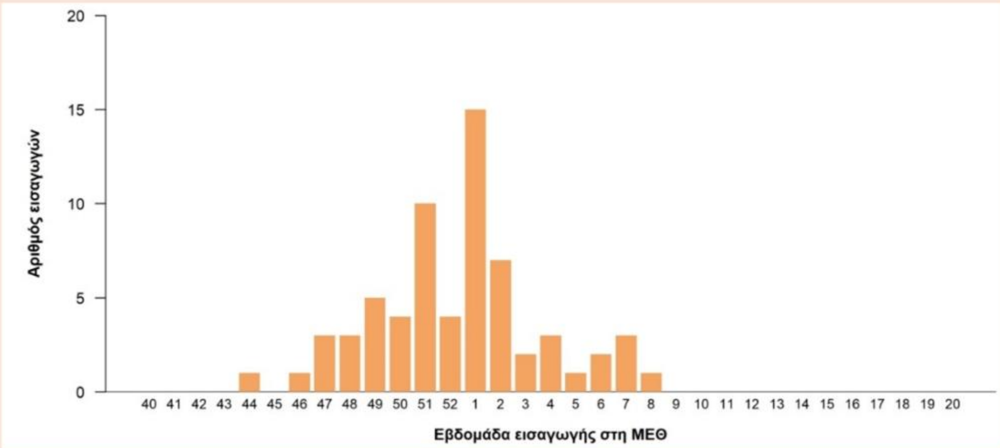



**Εικόνα 3.4** *Κυκλικό διάγραμμα για το ποσοστό επιχειρήσεων ανά κλάδο δραστηριότητας.*

**Επιτήρηση σοβαρών κρουσμάτων γρίπης**

Σε καθημερινή βάση πραγματοποιείται ενεργητική επιτήρηση όλων των κρουσμάτων με εργαστηριακά επιβεβαιωμένη λοίμωξη γρίπης που νοσηλεύονται σε ΜΕΘ από τη Διεύθυνση Επιδημιολογικής Επιτήρησης και Παρέμβασης για τα Λοιμώδη Νοσήματα.
Την εβδομάδα 11/2023 δεν καταγράφηκε νέο σοβαρό κρούσμα εργαστηριακά επιβεβαιωμένης γρίπης με νοσηλεία σε ΜΕΘ. Συνολικά, από την εβδομάδα 40/2022 έως και την εβδομάδα 11/2023 νοσηλεύτηκαν με γρίπη 65 άτομα σε ΜΕΘ. Στο Διάγραμμα 11 παρουσιάζεται ο εβδομαδιαίος αριθμός των νέων εισαγωγών ασθενών με γρίπη σε ΜΕΘ στο σύνολο της χώρας, από την εβδομάδα 40/2022 έως την εβδομάδα 11/2023.

**Διάγραμμα 11.** Εβδομαδιαίος αριθμός εισαγωγών σε Μονάδες Εντατικής Θεραπείας (ΜΕΘ) ασθενών με εργαστηριακά επιβεβαιωμένη γρίπη, σύνολο χώρας, εβδομάδα 40/2022 – εβδομάδα 11/2023 (n=65)




**Εικόνα 3.5** *Παράδειγμα στατιστικής έκθεσης (Πηγή: ΕΟΔΥ, 2023).*

## 3.2.1 «Παράμετροι» ποσοτικών δεδομένων

Οι στατιστικές παράμετροι είναι κάποιες τιμές που προκύπτουν από τα δεδομένα και επιτρέπουν να εκφραστεί ένα μεγάλο πλήθος δεδομένων με μία αντιπροσωπευτική τιμή, ώστε στη συνέχεια να είναι εύκολες οι συγκρίσεις, καθώς και η εξαγωγή συμπερασμάτων.

Υπάρχουν διάφορα είδη παραμέτρων ανάλογα με το τι αυτές αντιπροσωπεύουν.

- Παράμετροι θέσης (σε μια κλίμακα μετρήσεων) των ποσοτικών δεδομένων. Ονομάζονται και μέτρα θέσεως.
- Παράμετροι διασποράς για το εύρος των τιμών των ποσοτικών δεδομένων, οι οποίες ονομάζονται και μέτρα διασποράς.
- Παράμετροι που αντιπροσωπεύουν το σχήμα του διαγράμματος που περιγράφει όλα τα ποσοτικά δεδομένα και οι οποίες ονομάζονται μέτρα ασυμμετρίας και κύρτωσης.

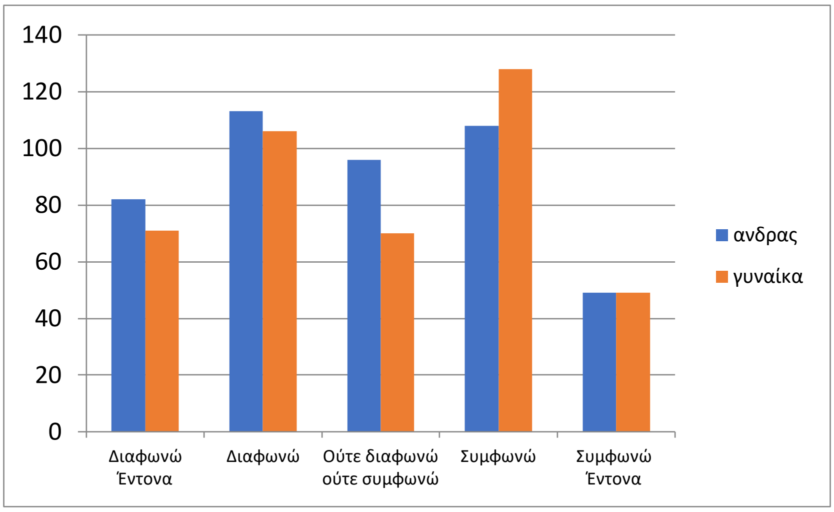
## 3.2.2 Παρουσίαση ποιοτικών δεδομένων

Ανάλογα με το είδος μεταβλητών, χρησιμοποιείται και ο αντίστοιχος τρόπος παρουσίασης. Στην περίπτωση που παρουσιάζονται ταυτόχρονα δύο ποιοτικές μεταβλητές, χρησιμοποιείται ο πίνακας διπλής εισόδου (Πίνακας 3.3) και το πολλαπλό ραβδόγραμμα (Εικόνα 3.6). Στην περίπτωση παρουσίασης δύο ποσοτικών μεταβλητών, κατάλληλο είναι το διάγραμμα διασποράς (Εικόνα 3.7). Το θηκόγραμμα χρησιμοποιείται για την παρουσίαση μιας ποσοτικής μεταβλητής σε σχέση με τις κατηγορίες μια ποιοτικής μεταβλητής (Εικόνα 3.8).

**Πίνακας 3.3** *Παρουσίαση γνώμης σε σχέση με το φύλο.*

| **Φύλο** | **Διαφωνώ έντονα** | **Διαφωνώ** | **Ούτε διαφωνώ Ούτε συμφωνώ** | **Συμφωνώ** | **Συμφωνώ έντονα** |
|---|---|---|---|---|---|
| άνδρας | 82 | 113 | 96 | 108 | 49 |
| | 18,3% | 25,2% | 21,4% | 24,1% | 10,9% |
| γυναίκα | 71 | 106 | 70 | 128 | 49 |
| | 16,7% | 25,0% | 16,5% | 30,2% | 11,6% |
| Σύνολο | 153 | 219 | 166 | 236 | 98 |
| | 17,5% | 25,1% | 19,0% | 27,1% | 11,2 |

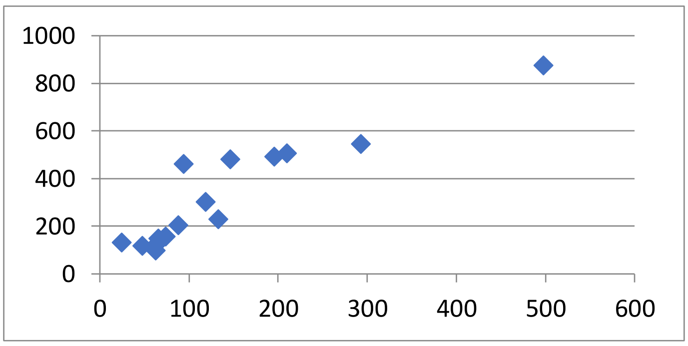
Στον Πίνακα 3.3 παρουσιάζονται ταυτόχρονα οι μεταβλητές «φύλο ατόμου» και η γνώμη του για το μελετώμενο θέμα. Στα κελιά του πίνακα εμφανίζονται τόσο η συχνότητα όσο και το ποσοστό κάθε γραμμής σε σχέση με το σύνολο της γραμμής.




**Εικόνα 3.6** *Ραβδόγραμμα γνώμης σε σχέση με το φύλο.*

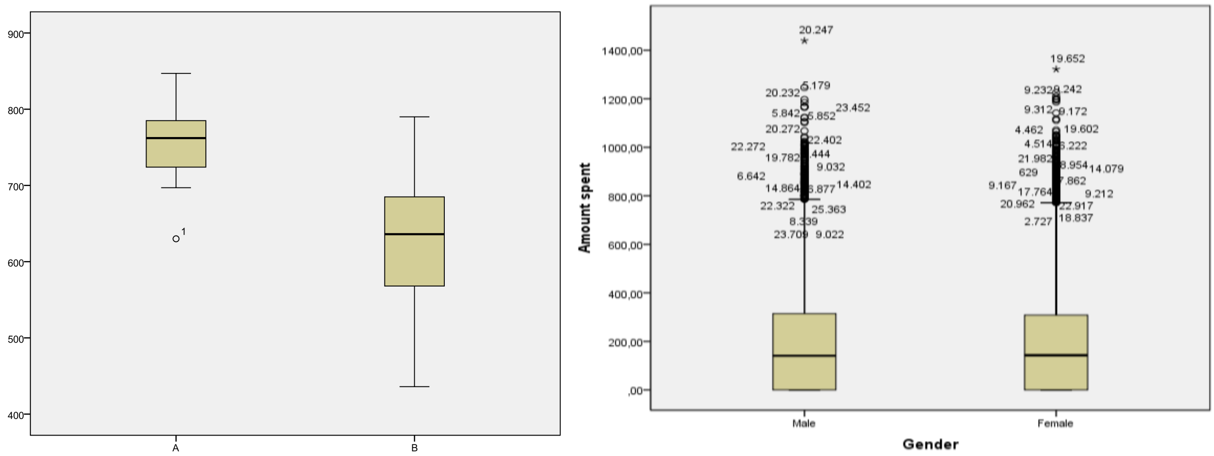
Για πολλές ποιοτικές μεταβλητές είναι δύσκολη η ταυτόχρονη παρουσίαση όλων μαζί των μεταβλητών με πίνακα ή με διάγραμμα. Όμως μπορούν να απεικονιστούν όλες οι μεταβλητές, χρησιμοποιώντας μεθόδους πολυμεταβλητής ανάλυσης δεδομένων, όπως:

- Παραγοντική ανάλυση, για την απεικόνιση των μεταβλητών στο επίπεδο των δύο κυριότερων παραγόντων.
- Ιεραρχική ταξινόμηση, για την εύρεση ομαδοποιήσεων μεταξύ των μεταβλητών και τη γραφική αναπαράστασή τους με δενδρόγραμμα.




**Εικόνα 3.7** *Διάγραμμα διασποράς δύο ποσοτικών μεταβλητών.*

Για την περιγραφική παρουσίαση μιας ποσοτικής και μιας ποιοτικής μεταβλητής μαζί, υπολογίζονται οι παράμετροι θέσεως ή διασποράς της ποσοτικής μεταβλητής, για κάθε κατηγορία της ποιοτικής μεταβλητής. Παρουσιάζεται το θηκόγραμμα της ποσοτικής μεταβλητής ξεχωριστά για κάθε κατηγορία της ποιοτικής μεταβλητής. Παράδειγμα θηκογράμματος για τα έξοδα σε σχέση με το φύλο ατόμου παρουσιάζεται στην Εικόνα 3.8 (δεξιά), όπου έγιναν υπολογισμοί ξεχωριστά για άνδρες και ξεχωριστά για γυναίκες και παρουσιάζονται τα μέτρα θέσεως, ώστε να γίνονται εύκολα οι συγκρίσεις μεταξύ ανδρών και γυναικών.




**Εικόνα 3.8** *Θηκόγραμμα για τα μέτρα θέσης των ομάδων Α και Β (αριστερά) και άλλο παράδειγμα (δεξιά).*

### 3.2.3 Υπολογισμός παραμέτρων ποσοτικών μεταβλητών

Στον Πίνακα 3.4 εμφανίζονται τα δεδομένα μοριοδότησης δύο ομάδων ατόμων που πήραν μέρος σε έναν διαγωνισμό. Στα ερωτήματα αν υπάρχει διαφορά στις δύο ομάδες ή ποια από τις δύο ομάδες είναι πιο ομογενής δεν υπάρχει άμεση απάντηση παρατηρώντας τα δεδομένα. Όταν υπολογιστούν τα μέτρα θέσεως και διασποράς, η απάντηση στα παραπάνω ερωτήματα γίνεται εύκολη. Τα κυριότερα μέτρα θέσης είναι ο μέσος όρος, η διάμεσος, τα τεταρτημόρια. Τα κυριότερα μέτρα διασποράς είναι η διακύμανση και η τυπική απόκλιση.

Ο **αριθμητικός μέσος** (μέσος όρος) υπολογίζεται αθροίζοντας όλα τα δεδομένα και διαιρώντας με το πλήθος τους. Η **διάμεσος** είναι η τιμή κάτω από την οποία βρίσκεται το 50% των δεδομένων και πάνω από την οποία βρίσκεται το υπόλοιπο 50%. Το **πρώτο τεταρτημόριο** αντιστοιχεί στην τιμή κάτω από την οποία βρίσκεται το 25% των δεδομένων και πάνω από την οποία βρίσκεται το υπόλοιπο 75%. Το **τρίτο τεταρτημόριο** αντιστοιχεί στην τιμή κάτω από την οποία βρίσκεται το 75% των δεδομένων και πάνω από την οποία βρίσκεται το υπόλοιπο 25%.

**Διακύμανση** είναι η τιμή που μετράει πόσο απέχουν τα δεδομένα από τον αριθμητικό μέσο. Υπολογίζεται αθροίζοντας τα τετράγωνα των διαφορών όλων των τιμών από τον αριθμητικό μέσο και διαιρώντας με το πλήθος τους. Επειδή για τη διακύμανση οι τιμές υψώνονται στο τετράγωνο, αυτή δεν μετριέται στις ίδιες μονάδες μέτρησης που μετριέται και η αντίστοιχη μεταβλητή και δεν μπορεί να δοθεί ερμηνεία στην τιμή της. Για να αποφευχθεί το πρόβλημα αυτό, υπολογίζεται η τυπική απόκλιση.

Η **τυπική απόκλιση** είναι η τιμή που υπολογίζεται από τη διακύμανση με την τετραγωνική ρίζα της διακύμανσης.

**Πίνακας 3.4** *Πίνακας με τη μοριοδότηση ατόμων δύο ομάδων.*

| **Α΄ ομάδα** | **Β΄ ομάδα** |
|---|---|
| 724 | 588 |
| 774 | 673 |
| 794 | 668 |
| 750 | 685 |
| 727 | 790 |
| 785 | 653 |
| 719 | 633 |
| 630 | 568 |
| 774 | 691 |
| 718 | 620 |
| 761 | 723 |
| 763 | 519 |
| 767 | 705 |
| 809 | 549 |
| 697 | 565 |
| 847 | 639 |
| 787 | 584 |
| 753 | 436 |

**Πίνακας 3.5** *Παράμετροι θέσης και διασποράς των δεδομένων του Πίνακα 3.4.*

| **Μέτρα θέσης και διασποράς** | **Ομάδα Α΄** | **Ομάδα Β΄** |
|---|---|---|
| Μέσος | 754,39 | 627,17 |
| Διάμεσος | 762 | 636 |
| Πρώτο τεταρτημόριο | 724 | 568 |
| Τρίτο τεταρτημόριο | 785 | 685 |
| Διακύμανση | 2.306,60 | 6.997,32 |
| Μέση απόκλιση τετραγώνου | 48,03 | 83,65 |

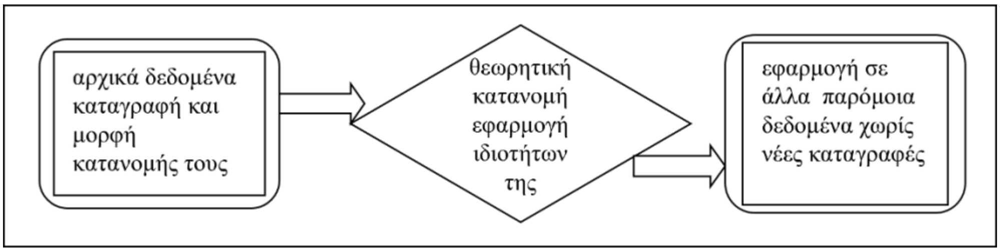
Στον Πίνακα 3.5 εμφανίζονται οι παράμετροι θέσης και διασποράς για τις δύο ομάδες του Πίνακα 3.4. Από τον Πίνακα 3.5 είναι ξεκάθαρο ότι η πρώτη ομάδα έχει μεγαλύτερο μέσο όρο, μεγαλύτερη διάμεσο και μικρότερη τυπική απόκλιση. Τα μόρια της πρώτης ομάδας είναι πιο συγκεντρωμένα γύρω από τον μέσο όρο 754,39 και επιπλέον οι μισοί της πρώτης ομάδας έχουν μόρια περισσότερα από 762. Στη δεύτερη ομάδα η ανομοιογένεια είναι μεγαλύτερη, αφού η τυπική απόκλιση είναι 83,65, η διάμεσος είναι μικρότερη από αυτήν της πρώτης ομάδας.

Στην πρώτη ομάδα το 75% των ατόμων έχει μόρια 724 και άνω, ενώ στη δεύτερη το 75% έχει λιγότερα από 685 μόρια. Στην Εικόνα 3.8 (αριστερά) παρουσιάζεται το θηκόγραμμα για τα μέτρα θέσης των ομάδων Α΄ και Β΄. Στο διάγραμμα του θηκογράμματος αυτού είναι εμφανής η διαφορά της ομάδας Α΄ σε σχέση με την ομάδα Β΄, η οποία έχει μεγαλύτερη ανομοιογένεια και οι τιμές της είναι πολύ χαμηλότερες από τις τιμές της Α΄ ομάδας. Στο ορθογώνιο που αντιστοιχεί στη Β΄ ομάδα, τόσο η κάτω πλευρά του (πρώτο τεταρτημόριο) όσο και η άνω πλευρά του (τρίτο τεταρτημόριο) βρίσκονται χαμηλότερα από την κάτω πλευρά (πρώτο τεταρτημόριο) του ορθογωνίου της Α΄ ομάδας.

## 3.3 Κατανομές πιθανοτήτων

Η απεικόνιση μιας ποσοτικής μεταβλητής με πίνακα ή ιστόγραμμα, απεικονίζει την **εμπειρική κατανομή** της. Παρατηρώντας ιστογράμματα από πολλά σύνολα διαφορετικών δεδομένων και διαφορετικών μεταβλητών, παρατηρήθηκε η επανάληψη κάποιων μορφών (σχήματος) ιστογράμματος. Αυτό έκανε τους επιστήμονες να μελετήσουν αναλυτικότερα τα συνηθέστερα σχήματα και να προσπαθήσουν να βρουν τις ιδιότητές τους. Έτσι, ορίστηκαν οι **θεωρητικές κατανομές** πιθανοτήτων. Γνωρίζοντας σε ποια θεωρητική κατανομή πλησιάζει μια μελετώμενη μεταβλητή πραγματικών δεδομένων είναι εύκολη η μελέτη της, ο υπολογισμός παραμέτρων της και η εύρεση πιθανοτήτων για τις διάφορες τιμές της, όπως θα παρουσιαστεί στη συνέχεια.

Οι θεωρητικές κατανομές αποτελούν τις γέφυρες, ώστε από τα εμπειρικά δεδομένα να γίνει η μετάβαση στη θεωρητική τους μελέτη και μετά να γίνει εφαρμογή σε άλλες παρόμοιες περιπτώσεις πραγματικών δεδομένων, όπως εμφανίζεται στο σχήμα της Εικόνας 3.9. Οι θεωρητικές κατανομές, υπολογίζονται με χρήση της θεωρίας, με κάποια συνάρτηση η οποία αντιστοιχεί σε γραφική παράσταση, και «προσομοιάζουν» ή «αντικαθιστούν» τις άγνωστες εμπειρικές κατανομές κάποιων μεταβλητών.




**Εικόνα 3.9** *Χρήση της θεωρητικής κατανομής για εφαρμογή σε πραγματικά δεδομένα.*

Οι θεωρητικές κατανομές ανάλογα με το είδος της μεταβλητής για την οποία χρησιμοποιούνται, διακρίνονται σε συνεχείς θεωρητικές κατανομές και διακριτές θεωρητικές κατανομές. Οι κυριότερες θεωρητικές κατανομές για συνεχείς μεταβλητές είναι η κανονική κατανομή, η κατανομή t student, η $X^2$ κατανομή, η F κατανομή, η εκθετική κατανομή, η ομοιόμορφη κατανομή και άλλες (Πίνακας 3.6). Για τη μέτρηση της συχνότητας εμφάνισης τιμών των ποιοτικών μεταβλητών, εφαρμόζεται πολλές φορές η διωνυμική κατανομή, η κατανομή Poisson και άλλες.

**Πίνακας 3.6** *Συνεχείς θεωρητικές κατανομές.*

| **Θεωρητική κατανομή συνεχής** | **Συμβολισμός** |
|---|---|
| Κανονική κατανομή | N(μ,σ) |
| t κατανομή | $t_n$ |
| $X^2$ κατανομή | $X^2_{n1,n2}$ |
| F κατανομή | $F_{n1,n2}$ |
| Εκθετική κατανομή | Exp |

Μια δύσκολη ερώτηση για τις θεωρητικές κατανομές και την προσαρμογή τους στα δεδομένα είναι «*Πώς καταλαβαίνουμε ποια κατανομή ακολουθούν τα δεδομένα;*» Η απάντηση εξαρτάται κυρίως από τους παράγοντες που εμφανίζονται στον Πίνακα 3.7. Για να βρεθεί ποια θεωρητική κατανομή ακολουθούν τα εμπειρικά δεδομένα, χρησιμοποιείται είτε αντίστοιχη θεωρία, είτε παλιότερες τιμές τους, αν υπάρχουν, είτε το σχήμα του ιστογράμματός τους, είτε υποτίθεται η θεωρητική κατανομή τους και ελέγχεται αν ισχύει η υπόθεση αυτή. Για την εφαρμογή μιας θεωρητικής κατανομής σε εμπειρικά δεδομένα, χρειάζεται να είναι γνωστές κάποιες παράμετροί της, δηλαδή κάποια βασικά στοιχεία που χαρακτηρίζουν τη θεωρητική κατανομή. Τα στοιχεία αυτά προκύπτουν από τα εμπειρικά δεδομένα. Γνωρίζοντας, λοιπόν, τις παραμέτρους αυτές και την αντίστοιχη θεωρητική κατανομή προκύπτει η «πλήρης γνώση» των δεδομένων.

Κάθε θεωρητική κατανομή, έχει «παραλλαγές» ανάλογα με τις τιμές κάποιων «παραμέτρων» της, δηλαδή τα βασικά στοιχεία της, που συνήθως αντιστοιχούν στη μέση τιμή ή και στην τυπική απόκλιση. Όταν οι παράμετροι αλλάζουν, τότε αλλάζουν και οι τιμές της θεωρητικής κατανομής, δηλαδή το «σχήμα» της ή οι πιθανότητες κάποιων τιμών της, αλλά παραμένουν στην ίδια «οικογένεια» της κατανομής. Γνωρίζοντας τις παραμέτρους της θεωρητικής κατανομής, υπολογίζεται μοναδικά η συγκεκριμένη κατανομή. Αν, για παράδειγμα, το ύψος από δένδρα λεύκες μπορεί περιγραφεί με τη θεωρητική κανονική κατανομή με μέση τιμή 5 μέτρα και τυπική απόκλιση 1,5 μέτρο, η κανονική κατανομή συμβολίζεται με N(5, 1,5). Το βάρος προβάτων μπορεί περιγραφεί με τη θεωρητική κανονική κατανομή με μέση τιμή 25 κιλά και τυπική απόκλιση 5 κιλά, η κανονική κατανομή συμβολίζεται με N(25, 5). Και οι δύο κατανομές ανήκουν στην οικογένεια της κανονικής κατανομής, αλλά με διαφορετικές παραμέτρους η καθεμία. Γνωρίζοντας τις παραμέτρους, υπάρχει η πλήρης γνώση για τη συγκεκριμένη κατανομή και αφού αυτή είναι «γνωστή», μπορούν να γίνουν οι υπολογισμοί βασιζόμενοι σε κάποιον τύπο ή σε κάποια ήδη υπολογισμένη τιμή και να εφαρμοστούν στα δεδομένα.

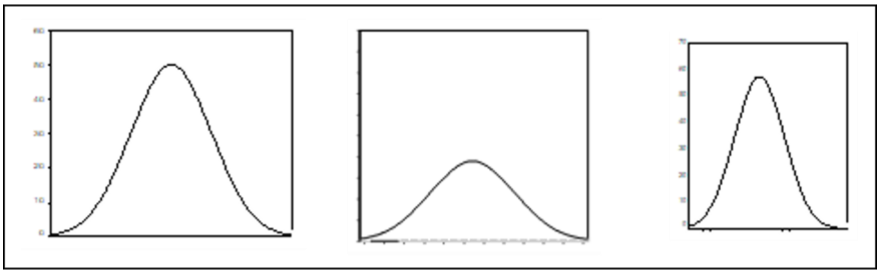
**Πίνακας 3.7** *Στοιχεία για την εύρεση της θεωρητικής κατανομής.*

| Α/Α | Παράγοντες δεομένων για την εύρεση της αντίστοιχης θεωρητικής κατανομής |
|---|---|
| 1 | Τι μετράνε (τι καταγράφουν) τα δεδομένα |
| 2 | Αντίστοιχη θεωρία |
| 3 | Σχήμα του ιστογράμματος, αν είναι σε ικανό πλήθος οι παρατηρήσεις |
| 4 | Υπόθεση που κάνουμε |
| 5 | Παλαιότερα δεδομένα για την ίδια μεταβλητή |

## 3.3.1 Κανονική κατανομή

Πολλές φορές, μελετώντας ποσοτικές μεταβλητές που αντιστοιχούν σε φυσικά μεγέθη, όπως ύψος ανθρώπων, βάρος ζώων, ύψος δένδρων, πληθυσμό μικροβίων κ.λπ., παρατηρήθηκε ότι το ιστόγραμμα τέτοιων μεταβλητών είναι συμμετρικό με κορυφή και σχήμα καμπάνας. **Κανονική κατανομή** ή φυσική κατανομή ή κατανομή Gauss ονομάζεται μία συνεχής συνάρτηση της οποίας η γραφική παράσταση είναι συμμετρική γύρω από ένα σημείο που αντιστοιχεί στη μέση τιμής της. Η μέση τιμή της ισούται με τη διάμεσό της και την επικρατούσα τιμή. Δεξιά και αριστερά από το κέντρο της, η γραφική της παράσταση επεκτείνεται το πολύ έως τρεις τυπικές αποκλίσεις και αλλάζει η κυρτότητά της σχηματίζοντας ένα σχήμα όπως το περίγραμμα μιας καμπάνας. Στο σχήμα της Εικόνας 3.10 παρουσιάζονται παραδείγματα κανονικής κατανομής.

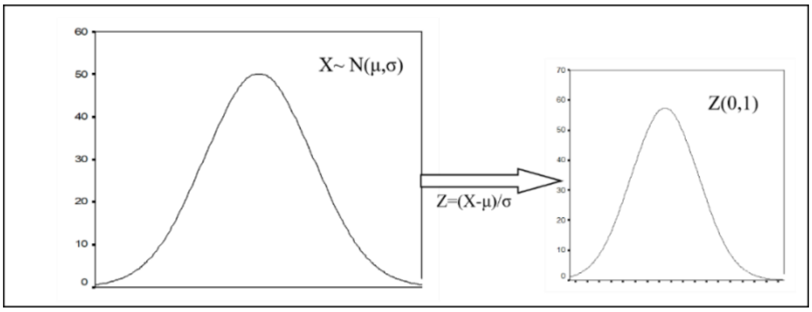
Η κανονική κατανομή δεν είναι μοναδική, αλλά αλλάζει ανάλογα με το κέντρο της που αντιστοιχεί στη μέση τιμή της μ και στο πλάτος της που αντιστοιχεί στην τυπική της απόκλιση σ. Συμβολίζεται με N(μ,σ), που αντιστοιχεί σε συνάρτηση κανονικής κατανομής με μέση τιμή μ και διακύμανση $\sigma^2$. Όταν η μέση τιμή είναι 0 και η διακύμανση 1, έχουμε την κανονική κατανομή N(0,1) η οποία ονομάζεται **τυπική κανονική** κατανομή και συμβολίζεται με Z. Ορίζεται για όλες τις τιμές από -3,9 έως και +3,9. Η τυπική κανονική κατανομή είναι πολύ χρήσιμη στον υπολογισμό πιθανοτήτων μιας οποιασδήποτε κανονικής κατανομής. Για τον λόγο αυτό μελετήθηκε διεξοδικά.



**Εικόνα 3.10** *Γραφικές παραστάσεις κανονικής κατανομής με διαφορετική τιμή διασποράς.*

### 3.3.1.1 Υπολογισμός πιθανοτήτων κανονικής κατανομής

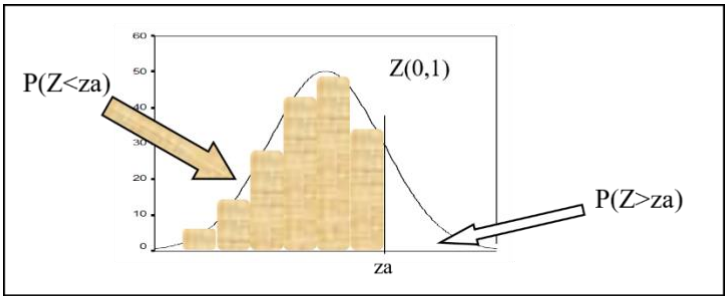
Η μετατροπή μίας μεταβλητής X που ακολουθεί την κανονική κατανομή με μέσο μ και τυπική απόκλιση σ, σε τυπική κανονική κατανομή Z(0,1), αποδεικνύεται ότι γίνεται με την πράξη αφαίρεσης της μέσης τιμής και κατόπιν διαίρεσης με το σ για όλες τις τιμές της X. Η νέα μεταβλητή που προκύπτει έχει μέση τιμή 0, τυπική απόκλιση 1 και ακολουθεί την τυπική κανονική κατανομή (Εικόνα 3.11).




**Εικόνα 3.11** *Μετατροπή κανονικής κατανομής σε τυπική κανονική κατανομή.*

Σε οποιαδήποτε κατανομή συνεχούς μεταβλητής, υπολογίζουμε την πιθανότητα η συνεχής μεταβλητή να πάρει τιμή μικρότερη ή ίση ενός δοσμένου αριθμού, με υπολογισμό του εμβαδού της καμπύλης της συνεχούς μεταβλητής, μεταξύ του οριζόντιου άξονα και του κατακόρυφου που αντιστοιχεί στον δοσμένο αριθμό. Το ίδιο ισχύει και για την τυπική κανονική κατανομή. Η πιθανότητα $P(Z<z_a)$ της συνεχούς μεταβλητής Z (η οποία ακολουθεί την τυπική κανονική κατανομή) υπολογίζεται από το εμβαδόν μεταξύ της καμπύλης της Z, του οριζόντιου άξονα και της κατακόρυφης ευθείας στο σημείο $z_a$ (Εικόνα 3.12). Έχουν βρεθεί τα εμβαδά όλων των τιμών από za = -3,9 έως za = +3,9, και υπάρχει πίνακας με τις πιθανότητες οι τιμές της να είναι μικρότερες από za δηλαδή P(Z<za). (Στο Παράρτημα του κεφαλαίου παρουσιάζεται ο πίνακας τυπικής κανονικής κατανομής).

Για τον υπολογισμό της πιθανότητας δηλαδή P(Z>za), δεν υπάρχει πίνακας με υπολογισμένες τιμές, αλλά ισχύει P(Z>za) = 1- P(Z<za), επειδή το συνολικό εμβαδόν μίας κατανομής είναι ίσο με 1. Ακόμη ισχύει P(zb<Z<za) = P(Z<za) - P(Z<zb). Δηλαδή αφαιρούνται τα αντίστοιχα εμβαδά (μικρότερο από zb και μικρότερο από za), ώστε να προκύψει το εμβαδόν ανάμεσα σε στις δύο τιμές za έως zb.




**Εικόνα 3.12** *Υπολογισμός πιθανότητας από το εμβαδόν σε τυπική κανονική κατανομή.*

Οι πιθανότητες για μια συνεχή μεταβλητή, η οποία ακολουθεί την κανονική κατανομή Ν(μ,σ), υπολογίζονται ως εξής με τη βοήθεια της μετατροπής μέσω της τυπικής κανονικής κατανομής:

P(X<κ) = P(Z<(κ-μ)/σ) = τιμή πιθανότητας από πίνακα τυπικής κανονικής κατανομής

P(X>κ) = 1-P(X<κ) = 1-P(Z< (κ-μ)/σ) = (προκύπτει αφαιρώντας από τη μονάδα την τιμή πιθανότητας από πίνακα τυπικής κανονικής κατανομής)

P(λ<X<κ) = P(X<κ)-P(X<λ) = P(Z< (κ-μ)/σ) - P(Z< (λ-μ)/σ) = (προκύπτει μετά από αφαίρεση μεταξύ τιμών πιθανότητας από πίνακα τυπικής κανονικής κατανομής).

Σημειώνουμε ότι για μία συνεχή μεταβλητή η πιθανότητα να πάρει ακριβώς μια συγκεκριμένη τιμή κ είναι μηδενική P(X=κ)= 0. Για τον λόγο αυτό, πάντα υπολογίζεται η πιθανότητα σε διάστημα τιμών γύρω από την τιμή κ, δηλαδή P(κ-0,01<X<κ+0,01).

Κάποιες φορές είναι ενδιαφέρον να βρεθεί η τιμή ω της συνεχούς μεταβλητής, πάνω από την οποία η πιθανότητα είναι 5% ή 10% κ.λπ. Στην περίπτωση αυτή η διαδικασία είναι ανάστροφη, ξεκινώντας από την εξίσωση P(X>ω) = 0,05 ⇔ P(X<ω) = 1-0,05 ⇔ P(X<ω) = 0,95 ⇔ P(Z<(ω-μ)/σ) = 0,95. Εντοπίζεται στον πίνακα τυπικής κανονικής κατανομής η τιμή $z_ω$, για την οποία η πιθανότητα αντιστοιχεί σε 0,95 και κατόπιν λύνεται η εξίσωση (ω-μ) /σ = $z_ω$ και προκύπτει η τιμή του ω.

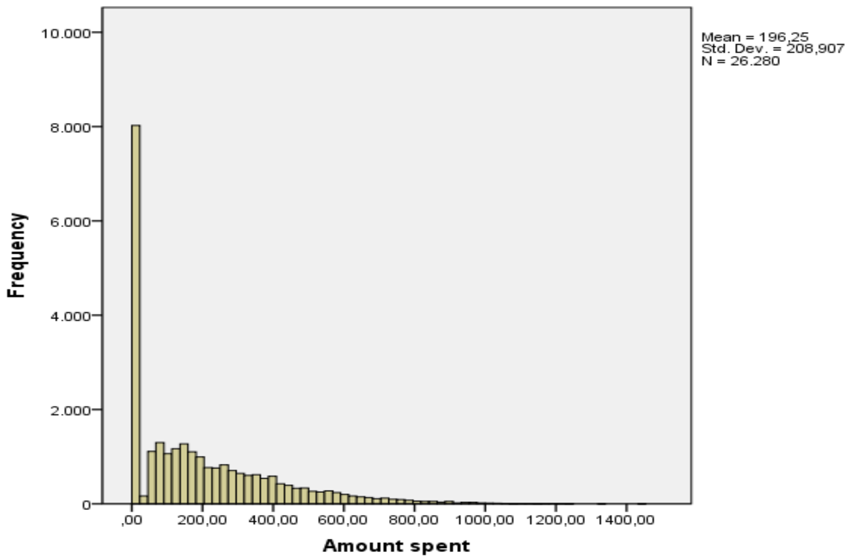
## 3.4 Εφαρμογές για κατανόηση

### 3.4.1 Σχολιασμός αποτελεσμάτων

Να σχολιάσετε τα αποτελέσματα, όπως προέκυψαν από το λογισμικό στατιστικής επεξεργασίας SPSS του Πίνακα 3.8 και της Εικόνας 3.13 για τις μεταβλητές «πλήθος αντικειμένων» και «ποσό που πληρώθηκε» για την αγορά τους.

**Πίνακας 3.8** *Παρουσίαση παραμέτρων για δύο ποσοτικές μεταβλητές.*

| | | Πλήθος αντικειμένων που αγοράστηκαν | Ποσό πληρωμής |
|---|---|---|---|
| N | Valid | 26.280 | 26.280 |
| | Missing | 0 | 0 |
| Mean | | 2,36 | 196,2524 |
| Median | | 2,00 | 141,7750 |
| Mode | | 0 | 0,00 |
| Minimum | | 0 | 0,00 |
| Maximum | | 13 | 1.439,37 |
| Percentiles | 25 | 0,00 | 0,0000 |
| | 50 | 2,00 | 141,7750 |
| | 75 | 4,00 | 311,3125 |




**Εικόνα 3.13** *Ιστόγραμμα.*

Απάντηση: Συνολικά εξετάστηκαν 26.280 αγορές (πλήθος δεδομένων). Η μέση τιμή του πλήθους αντικειμένων που αγοραζόταν σε κάθε αγορά υπολογίστηκε ως 2,36 αντικείμενα. Βέβαια, η τιμή αυτή δεν έχει φυσικό νόημα, αφού το πλήθος των αντικειμένων που αγοράζονται είναι ακέραιος αριθμός μεταξύ 0 και 13, όπως φαίνεται από τον Πίνακα 3.8 στις γραμμές minimum και maximum. Η τιμή 2,36 είναι υποθετική τιμή και αντιστοιχεί στο πλήθος αντικειμένων, αν υποτεθεί ότι σε κάθε αγορά υπήρχε το ίδιο ακριβώς πλήθος χωρίς να υπάρχουν διαφοροποιήσεις μεταξύ των αγορών.

Η διάμεσος του πλήθους αντικειμένων είναι 2 αντικείμενα, που σημαίνει ότι στις μισές αγορές αγοράστηκαν λιγότερα (ή ίσα) από 2 αντικείμενα και στις άλλες μισές περισσότερα (ή ίσα) από 2 αντικείμενα. Το πρώτο τεταρτημόριο είναι 0 αντικείμενα. Δηλαδή, στο 25% των αγορών δεν αγοράστηκε κανένα αντικείμενο. Το τρίτο τεταρτημόριο είναι 4 αντικείμενα. Δηλαδή, στο 25% των αγορών αγοράστηκαν περισσότερα από 4 αντικείμενα.

Η μέση τιμή του ποσού πληρωμής κάθε αγοράς υπολογίστηκε ως 196,3 μονάδες χρήματος. Η διάμεσος του ποσού πληρωμής είναι 141,8 μονάδες, που σημαίνει ότι στις μισές αγορές πληρώθηκαν λιγότερες από 141,8 μονάδες, στις άλλες μισές περισσότερες από 141,8 μονάδες. Το πρώτο τεταρτημόριο είναι 0 μονάδες. Δηλαδή, στο 25% των αγορών δεν πληρώθηκε ποσό. Το τρίτο τεταρτημόριο είναι 311,3 μονάδες. Δηλαδή, στο 25% των αγορών πληρώθηκαν περισσότερες από 311,3 μονάδες.

### 3.4.2 Εφαρμογή σε κανονική κατανομή

Καταγράφηκε το πλήθος των πελατών που εισέρχονται σε ένα πολυκατάστημα ανά ώρα, στις ώρες αιχμής και προέκυψαν οι παρακάτω τιμές.

203 185 223 206 199 214 231 160 178 187 170 210 220 207 180 213 219 194

Ο πίνακας συχνοτήτων για ομαδοποιημένα διαστήματα πλήθους πελατών παρουσιάζεται στον Πίνακα 3.9 και στο ιστόγραμμα στην Εικόνα 3.14. Στον Πίνακα 3.10 παρουσιάζονται τα περιγραφικά στατιστικά.

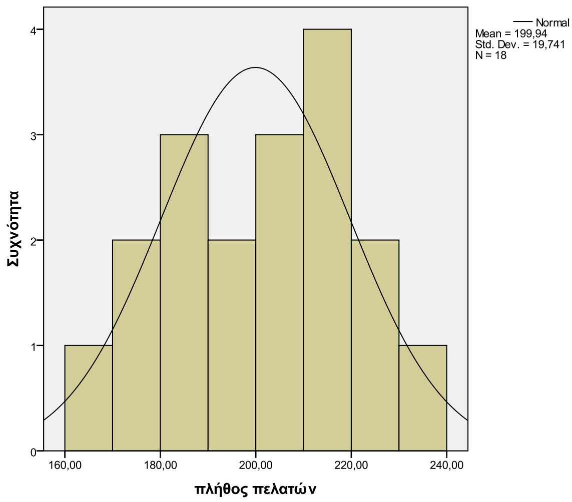
Ο υπεύθυνος του καταστήματος θεωρεί ότι το πλήθος πελατών στις ώρες αιχμής ακολουθεί την κανονική κατανομή με μέσο 200 πελάτες και τυπική απόκλιση 20 πελάτες. Να υπολογιστεί η πιθανότητα αύριο σε μια συγκεκριμένη ώρα αιχμής να εισέλθουν λιγότεροι από 230 πελάτες στο κατάστημα.

**Πίνακας 3.9** *Παρουσίαση δεδομένων με διαστήματα.*

| Πλήθος πελατών σε μια ώρα αιχμής | Συχνότητα | Ποσοστό |
|---|---|---|
| [160-180] | 3 | 16,7% |
| [180-200] | 5 | 27,8% |
| [200-220] | 6 | 38,9% |
| [220-240] | 4 | 16,7% |
| Σύνολο | 18 | |

**Πίνακας 3.10** *Παράμετροι δεδομένων εφαρμογής.*

| Παράμετρος | Τιμή |
|---|---|
| Μέση τιμή | 199,9 |
| Διάμεσος | 204,5 |
| Διακύμανση | 389,7 |
| Τυπική απόκλιση | 19,7 |




**Εικόνα 3.14** *Ιστόγραμμα με απεικόνιση της κανονικής κατανομής.*

<u>Απάντηση:</u> Αν υποτεθεί ότι η κατανομή της μεταβλητής X «πλήθος πελατών» πλησιάζει στην κανονική, τότε η ζητούμενη πιθανότητα θα υπολογιστεί μέσω του πίνακα της τυπικής κανονικής κατανομής ως εξής:

$P(X<230) = P((X-\mu)/\sigma < (230-200)/20) = P(Z<1{,}5) = 0{,}9332$, όπως εμφανίζεται στον πίνακα της τυπικής κανονικής κατανομής. Η πιθανότητα να έρθουν σε μία ώρα λιγότερο από 230 πελάτες την επόμενη μέρα είναι μεγάλη 93,32%. Η πιθανότητα να έρθουν περισσότεροι από 230 πελάτες θα είναι 1-0,9332 = 0,0668 = 6,68%.

Αν ο υπεύθυνος θέλει να δει ποιο πλήθος πελατών και άνω έχει πιθανότητα να εμφανιστεί μόνο 5%, θα πρέπει να υπολογίσει την τιμή ω, για την οποία $P(X>\omega) = 0{,}05$. Ισοδύναμα $P(X<\omega) = 1-0{,}05 \Leftrightarrow P(X<\omega)=0{,}95 \Leftrightarrow P(Z< (\omega-\mu)/\sigma) = 0{,}95$. Παρατηρώντας τον πίνακα της τυπικής κανονικής κατανομής, η πιθανότητα 0,95 αντιστοιχεί στην τιμή $Z = 1{,}64$ οπότε θα είναι $(\omega-\mu)/\sigma = 1{,}64 \Leftrightarrow \omega=\mu+1{,}64\sigma = 200+1{,}64\times20 = 232{,}8$. Υπάρχει πιθανότητα 5% οι πελάτες σε μια ώρα αιχμής να είναι περισσότεροι από 233.

### 3.4.3 Σύγκριση τιμών

Τα ημερήσια έσοδα περιπτέρων μιας πόλης ακολουθούν κανονική κατανομή με μέση τιμή 80 ευρώ και τυπική απόκλιση 20 ευρώ. Τα μηνιαία έσοδα από την πώληση μόνο προϊόντων καπνού ακολουθούν κανονική κατανομή με μέση τιμή 1.500 ευρώ και τυπική απόκλιση 200 ευρώ. Ένας ιδιοκτήτης περιπτέρου παρατήρησε ότι τον προηγούμενο μήνα είχε έσοδα από την πώληση προϊόντων καπνού 1.800 ευρώ. Οι ημερήσιες πωλήσεις του τον προηγούμενο μήνα ήταν κατά μέσο όρο 90 ευρώ. Ζητά να μάθει αν το περίπτερο του έχει δυνατότητα να αυξήσει και άλλο τις ημερήσιες πωλήσεις του ή αν είναι μεγαλύτερη η δυνατότητά του να αυξήσει τις μηνιαίες πωλήσεις του για τα προϊόντα καπνού, στη συγκεκριμένη πόλη.

<u>Απάντηση:</u> Για να δει ο ιδιοκτήτης αν υπάρχει δυνατότητα να αυξηθούν είτε οι ημερήσιες πωλήσεις του είτε οι μηνιαίες για τον καπνό, θα γίνει σύγκριση των πωλήσεων του περιπτέρου με τις αντίστοιχες πωλήσεις όλων των άλλων περιπτέρων της πόλης που είναι οι ανταγωνιστές του. Αν οι πωλήσεις του περιπτέρου είναι από τις καλύτερες πωλήσεις του κλάδου, δεν υπάρχει μεγάλη δυνατότητα αύξησης τους. Αν είναι στις χαμηλότερες ή στις μεσαίες του κλάδου, υπάρχει δυνατότητα να αυξηθούν στις υψηλότερες του κλάδου.

Οι ημερήσιες πωλήσεις μετριούνται διαφορετικά από ό,τι οι μηνιαίες πωλήσεις. Επειδή και οι δύο ακολουθούν κανονική κατανομή, θα μετατραπούν σε τυπική κανονική κατανομή (z-τιμές), ώστε να συγκριθούν.

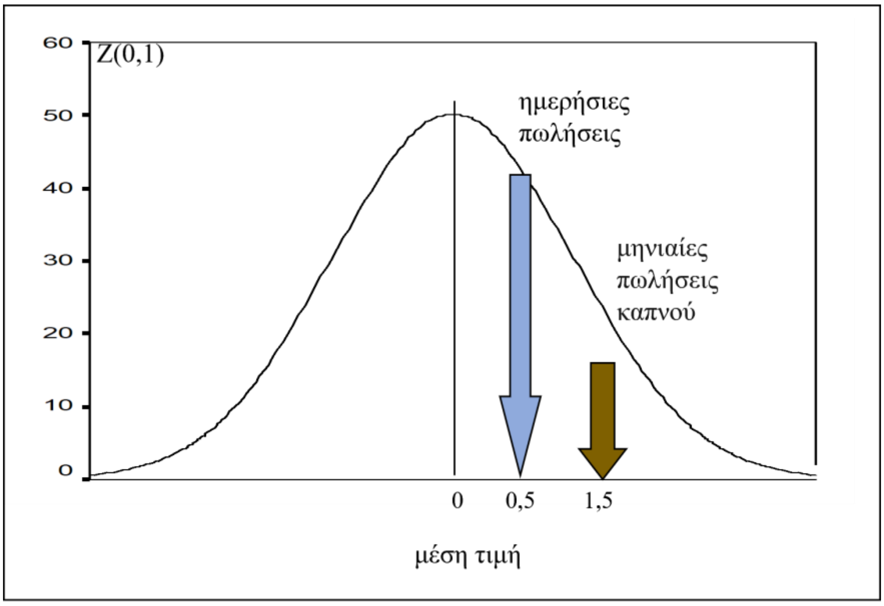
Η z-τιμή των ημερήσιων πωλήσεων του περιπτέρου είναι $(90-80)/20= 10/20= 0{,}5$.

Η z-τιμή των μηνιαίων πωλήσεων καπνού του περιπτέρου είναι $(1800-1500)/200= 300/200= 1{,}5$.

Στις ημερήσιες πωλήσεις του το περίπτερο βρίσκεται σε 0,5 τυπικές αποκλίσεις πάνω από τον μέσο όρο και μπορεί να μετακινηθεί έως και 3 αποκλίσεις πάνω από τον μέσο όρο, οπότε έχει πολλές δυνατότητες αύξησης των ημερήσιων πωλήσεών του.

Στις μηνιαίες πωλήσεις καπνού το περίπτερο βρίσκεται σε 1,5 τυπικές αποκλίσεις πάνω από τον μέσο όρο και μπορεί να μετακινηθεί έως και 3 αποκλίσεις πάνω από τον μέσο όρο, οπότε έχει πάλι δυνατότητες αύξησης των ημερήσιων πωλήσεών του, όχι τόσο πολλές όπως για τις ημερήσιες πωλήσεις, αφού ήδη είναι σε υψηλό βαθμό πωλήσεων καπνού σε σχέση με τα υπόλοιπα περίπτερα της πόλης του.

Στο σχήμα της τυπικής κανονικής κατανομής της Εικόνας 3.15 φαίνεται η θέση του περιπτέρου για τις ημερήσιες πωλήσεις και για τις μηνιαίες πωλήσεις καπνού.




**Εικόνα 3.15** *Υπολογισμός τιμών σε τυπική κανονική κατανομή.*

## 3.5 Ασκήσεις για κατανόηση - εφαρμογή

### 3.5.1 Άσκηση Α΄: Εισόδημα

Σας δίνονται οι παράμετροι θέσης και διασποράς για τη μεταβλητή «εισόδημα» σε τέσσερα δείγματα διαφορετικών ηλικιών, όπως προέκυψαν από το πρόγραμμα SPSS. Υπάρχει διαφορά στη θέση των δειγμάτων; Ποιο από τα δείγματα έχει μεγαλύτερη διασπορά; Ποιο μεγαλύτερη ομοιογένεια;

**Case Summaries**

**Εισόδημα (χιλιάδες)**

| **Ηλικία** | **N** | **Mean** | **Median** | **Std. Deviation** | **Minimum** | **Maximum** |
|---|---|---|---|---|---|---|
| Κάτω από 25 | 589 | 9,4630 | 9,7000 | 1,96623 | 4,20 | 12,60 |
| 25-49 | 1.130 | 17,6634 | 17,1000 | 3,58871 | 11,90 | 25,70 |
| 50-74 | 567 | 30,2125 | 29,8000 | 3,52136 | 23,60 | 37,90 |
| 75 και άνω | 824 | 60,9275 | 59,3500 | 16,16260 | 36,70 | 99,80 |
| Σύνολο | 3.110 | 29,8611 | 22,3000 | 21,57632 | 4,20 | 99,80 |

<u>Λύση</u>

Τα τέσσερα δείγματα έχουν διαφορετικό πλήθος και διαφορετικές μέσες τιμές. Μεγαλύτερη μέση τιμή ίση με 60,92 χιλιάδες έχει το δείγμα ηλικίας 75 και άνω. Μικρότερη μέση τιμή εισοδήματος 9,46 χιλιάδες έχει το δείγμα με τις ηλικίες κάτω των 25 ετών. Το δείγμα ηλικίας 25-49 ετών έχει μέση τιμή εισοδήματος 17,66 χιλιάδες και το δείγμα ηλικίας 50-74 ετών έχει μέση τιμή εισοδήματος 30,21 χιλιάδες. Το μέσο εισόδημα αυξάνεται καθώς αυξάνεται η ηλικία των ανθρώπων σε καθένα από τα 4 δείγματα. Παρατηρώντας τη στήλη με την τυπική απόκλιση, το τελευταίο δείγμα έχει μεγάλη τυπική απόκλιση ίση με 16,16 χιλιάδες και επομένως μεγάλη διασπορά τιμών γύρω από τη μέση τιμή. Τα υπόλοιπα τρία δείγματα έχουν περίπου ίδια τυπική απόκλιση 1,96 χιλιάδες, 3,58 χιλιάδες και 3,52 χιλιάδες αντίστοιχα. Για να συγκριθεί η ομοιογένεια των τεσσάρων δειγμάτων, επειδή αυτά έχουν διαφορετική μέση τιμή, θα υπολογιστεί ο συντελεστής μεταβλητότητας για το καθένα.

Ο συντελεστής μεταβλητότητας υπολογίζεται διαιρώντας την τυπική απόκλιση με τη μέση τιμή κάθε δείγματος. Για το πρώτο δείγμα είναι 1,96/9,46=0,21, για το δεύτερο 3,58/17,66=0,20, για το τρίτο δείγμα 3,52/30,21=0,12 και για το τέταρτο 16,16/60,93=0,26.

Όσο μικρότερος είναι ο συντελεστής μεταβλητότητας, τόσο μεγαλύτερη ομοιογένεια έχουν τα δεδομένα. Επομένως, μεγαλύτερη ομοιογένεια έχει το τρίτο δείγμα για το οποίο ο συντελεστής μεταβλητότητας είναι 0,12.

## 3.5.2 Άσκηση Β΄: Ημερήσια κατανάλωση ρεύματος

Σας δίνονται τα παρακάτω δεδομένα για την ημερήσια κατανάλωση ρεύματος 35 νοικοκυριών.

3 5 6 5 8 9 7 2 3 4 5 6 4 7 15 8 12 13 9 9 7 4 5 6 6 7 9 10 3 5 3 5 6 7 8

α) Υπολογίστε τον αριθμητικό μέσο και ερμηνεύστε τι σημαίνει η τιμή του.

β) Βρείτε τη διάμεσο και τα τεταρτημόρια και σχολιάστε την τιμή τους.

γ) Μπορείτε να προτείνετε μία θεωρητική κατανομή, η οποία ίσως προσαρμόζεται στα δεδομένα;

Λύση

α) Ο αριθμητικός μέσος υπολογίζεται με το άθροισμα όλων διά το πλήθος τους. Η μέση ημερήσια κατανάλωση είναι 231/35=6,6. Ερμηνεύεται ως η τιμή που θα είχαν όλα τα νοικοκυριά, αν όλα κατανάλωναν ακριβώς το ίδιο χωρίς διαφοροποιήσεις.

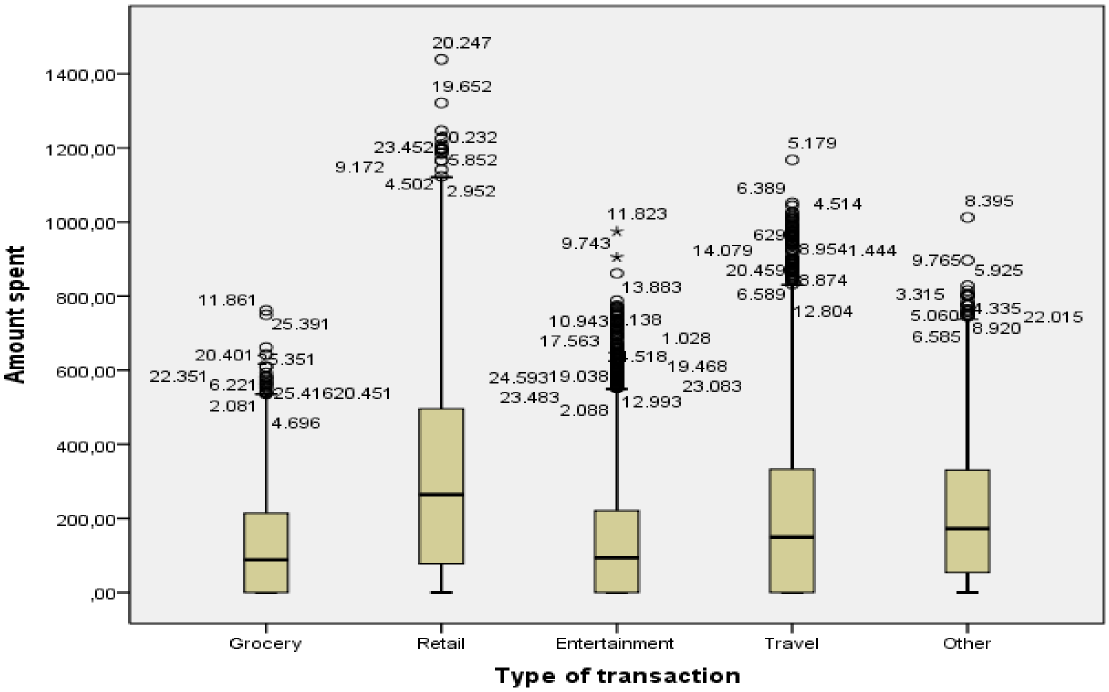
β) Η διάμεσος είναι η τιμή 6, κάτω από την οποία βρίσκονται οι καταναλώσεις των μισών νοικοκυριών και πάνω από την οποία βρίσκονται οι καταναλώσεις των άλλων μισών. Το πρώτο τεταρτημόριο είναι η τιμή 5, κάτω από την οποία βρίσκεται το 25% των νοικοκυριών με χαμηλή κατανάλωση. Το τρίτο τεταρτημόριο είναι η τιμή 8, πάνω από την οποία βρίσκεται το 25% των νοικοκυριών με υψηλή κατανάλωση. Οι «μεσαίες» τιμές ημερήσιας κατανάλωσης για τα νοικοκυριά που μελετήθηκαν είναι μεταξύ 6 και 8.

γ) Η διάμεσος πλησιάζει τον αριθμητικό μέσο, και ίσως μία κατανομή που ταιριάζει στα δεδομένα να είναι η κανονική κατανομή.

### 3.5.3 Άσκηση Γ΄: Θηκόγραμμα και σχολιασμός

Τα παρακάτω δεδομένα (πίνακας και διάγραμμα θηκόγραμμα) παρουσιάζουν το ποσό που πληρώθηκε μέσω πιστωτικής κάρτας για διάφορες κατηγορίες δαπανών. Τι συμπεράσματα μπορείτε να βγάλετε;

| Type of transaction | N | Mean | Median | Min | Maximum |
|---|---|---|---|---|---|
| Grocery (Τρόφιμα) | 5.256 | 129,2772 | 88,4150 | ,00 | 761,53 |
| Retail (Είδη σπιτιού) | 5.256 | 312,5793 | 264,3250 | ,00 | 1.439,37 |
| Entertainment (Διασκέδαση) | 5.256 | 135,4265 | 94,2100 | ,00 | 974,82 |
| Travel (Ταξίδια) | 5.256 | 199,1811 | 149,8850 | ,00 | 1.167,67 |
| Other (Λοιπά) | 5.256 | 204,7978 | 172,6150 | ,00 | 1.012,50 |
| Total (Σύνολο) | 26.280 | 196,2524 | 141,7750 | ,00 | 1.439,37 |




Λύση

Παρατηρώντας το θηκόγραμμα για τις διάφορες κατηγορίες δαπανών, διαπιστώνεται ότι τα είδη σπιτιού (δεύτερο είδος) έχουν τη μεγαλύτερη διάμεσο και τη μεγαλύτερη ανομοιογένεια σε σχέση με τις υπόλοιπες κατηγορίες δαπανών. Μικρότερη διασπορά έχουν τα είδη ψυχαγωγίας και τα είδη τροφίμων, τα οποία έχουν και τη μικρότερη μέση τιμή δαπανών. Οι κατηγορίες δαπάνης για ταξίδια ή για άλλα είδη έχουν παρόμοια διάμεσο και παρόμοια τυπική απόκλιση.

### 3.5.4 Άσκηση Δ΄: Σχολιασμός αποτελεσμάτων

Σχολιάστε τα παρακάτω αποτελέσματα, όπως προέκυψαν από το στατιστικό λογισμικό SPSS για τη μεταβλητή «ποσό που πληρώθηκε» από τους άνδρες και τις γυναίκες.

| Gender | | | | Statistic | Std. Error |
|---|---|---|---|---|---|
| Amount spent | Male | Mean | | 195,2478 | 1,79391 |
| | | 95% Confidence Interval for Mean | Lower Bound | 191,7315 | |
| | | | Upper Bound | 198,7642 | |
| | | 5% Trimmed Mean | | 175,2106 | |
| | | Median | | 140,8650 | |
| | | Variance | | 43.251,662 | |
| | | Std. Deviation | | 207,97034 | |
| | | Minimum | | ,00 | |
| | | Maximum | | 1.439,37 | |
| | | Range | | 1.439,37 | |
| | | Interquartile Range | | 314,38 | |

| Gender | | | | Statistic | Std. Error |
|---|---|---|---|---|---|
| | Female | Mean | | 197,3038 | 1,85225 |
| | | 95% Confidence Interval for Mean | Lower Bound | 193,6732 | |
| | | | Upper Bound | 200,9345 | |
| | | 5% Trimmed Mean | | 176,8384 | |
| | | Median | | 142,8800 | |
| | | Variance | | 44.051,703 | |
| | | Std. Deviation | | 209,88498 | |
| | | Minimum | | ,00 | |
| | | Maximum | | 1.321,55 | |
| | | Range | | 1.321,55 | |
| | | Interquartile Range | | 308,77 | |

Λύση

Η μέση τιμή του ποσού που πληρώθηκε για τους άνδρες είναι 195,24 με τυπική απόκλιση 207,97, ενώ για τις γυναίκες η μέση τιμή είναι 197,30 και η τυπική απόκλιση 209,88. Στα δύο φύλα, δεν παρατηρούνται σημαντικές διαφορές για το ποσό που πληρώθηκε, παρά μόνο μια μεγαλύτερη τιμή της διαμέσου (142,88) για τις γυναίκες από ό,τι για τους άνδρες (140,86).

### 3.5.5 Άσκηση Ε΄: Καθυστερήσεις πτήσεων

Διευθυντής αεροπορικής εταιρείας μελετά τις καθυστερήσεις των αναχωρήσεων. Από τα στατιστικά στοιχεία προκύπτει ότι ο μέσος χρόνος καθυστέρησης (μ) είναι 15 λεπτά και η τυπική απόκλιση (σ) είναι 6 λεπτά. Α. Να βρεθεί η πιθανότητα μια πτήση να καθυστερήσει

α) το πολύ μέχρι 10 λεπτά,

β) πάνω από 20 λεπτά,

γ) από 10 έως 20 λεπτά.

Β. Πάνω από πόσα λεπτά καθυστερεί το 10% των πιο αργοπορημένων πτήσεων;

Λύση

Δεν αναφέρεται το είδος της κατανομής για τη μεταβλητή Χ χρόνος καθυστέρησης αναχώρησης. Η μεταβλητή είναι συνεχής και για να βρεθούν οι ζητούμενες πιθανότητες υποτίθεται ότι ακολουθεί την κανονική κατανομή.

Α. Μετατρέπονται οι ζητούμενες πιθανότητες της μεταβλητής Χ σε πιθανότητες της τυπικής κανονικής κατανομής Ζ και υπολογίζονται με τον πίνακα για τις z τιμές:

α) $P(X<10) = P((X\text{-}\mu)/\sigma < (10\text{-}15)/6) = P(Z<-0{,}83) = 0{,}2033$

β) $P(X>20) = P((X\text{-}\mu)/\sigma > (20\text{-}15)/6) = P(Z>0{,}83) = 1\text{-} P(Z<0{,}83) = 1\text{-}0{,}7967 = 0{,}2033$

γ) $P(10<X<20) = P(X<20)\text{-}P(X<10) = 1\text{-}P(X>20) – P(X<10) = 1\text{-}0{,}2033\text{-}0{,}2033 = 0{,}5934$

Β. Στην περίπτωση αυτή είναι γνωστή η πιθανότητα 0,10 (10%) και αναζητείται η τιμή ω πάνω από την οποία η πιθανότητα θα είναι 0,10 και κάτω από την οποία θα είναι 1-0,10 = 0,90.

$$P(X<\omega) = 0{,}90 \Leftrightarrow P((X\text{-}\mu)/\sigma < (\omega\text{-}15)/6) = 0{,}90 \Leftrightarrow P(Z < (\omega\text{-}15)/6) = 0{,}90.$$

Από τον πίνακα τιμών της τυπικής κανονικής κατανομής προκύπτει $(\omega\text{-}15)/6 = 1{,}28 \Leftrightarrow \omega = 22{,}68$ λεπτά

## 3.5.6 Άσκηση ΣΤ΄: Κατανομή μισθωτών

Η κατανομή 2.000 μισθωτών ακολουθεί την κανονική κατανομή με μέσο μισθό μ = 1.000 ευρώ και τυπική απόκλιση σ = 160 ευρώ. Να βρεθεί ο αριθμός των μισθωτών που παίρνει μισθό: α) μικρότερο των 900 ευρώ, β) μεγαλύτερο των 1.300 ευρώ, γ) μεταξύ 1.000 και 1.200 ευρώ.

Λύση

α) $P(X<900) = P((X\text{-}\mu)/\sigma < (900\text{-}1.000)/160) = P(Z<-0{,}625) = 0{,}2643$

β) $P(X>1.300) = P((X\text{-}\mu)/\sigma > (1300\text{-}1.000)/160) = P(Z> 1{,}875) = 1\text{-} P(Z<1{,}875) = 1\text{-}0{,}9699 = 0{,}0301$

γ) $P(1.000<X<1.200) = P(X<1.200)\text{-}P(X<1.000) = P(Z<1{,}25) – P(X<0) = 0{,}8944 \text{-}0{,}50 = 0{,}3944$

## 5.7 Άσκηση Ζ΄: Πριμ παραγωγικότητας εργαζομένων

Διευθυντής μιας μονάδας παραγωγής σκέφτεται να δώσει πριμ παραγωγικότητας στους εργαζομένους αν παράγουν 5% και άνω της υψηλότερης απόδοσής τους σε σχέση με το παρελθόν. Τα στατιστικά στοιχεία του παρελθόντος μελετήθηκαν και έδειξαν ότι η μέση παραγωγή ήταν 4.000 μονάδες. Η κατανομή της πλησίαζε την κανονική με σ=600 μονάδες. Πόσες μονάδες πρέπει να παραγάγουν οι εργαζόμενοι για να πάρουν το πριμ;

Λύση

Στην περίπτωση αυτή είναι γνωστή η πιθανότητα 0,05 (5%) και αναζητείται η τιμή ω πάνω από την οποία η πιθανότητα θα είναι 0,05 και κάτω από την οποία θα είναι 1-0,05=0,95.

$P(X<\omega) = 0{,}95 \Leftrightarrow P((X\text{-}\mu)/\sigma < (\omega\text{-}4.000)/600) = 0{,}95 \Leftrightarrow P(Z < (\omega\text{-}4.000)/600) = 0{,}95$. Από τον πίνακα τιμών της τυπικής κανονικής κατανομής προκύπτει $(\omega\text{-}4.000)/600 = 1{,}645 \Leftrightarrow \omega = 4.987$ μονάδες

# Βιβλιογραφία/Αναφορές


Γναρδέλλης, Χ. (2019). *Εφαρμοσμένη Στατιστική*. Αθήνα: Παπαζήση.

Κώστογλου, Β., & Αντωνίου, Ε. ( 2021). *Πιθανότητες και Στατιστική*. Θεσσαλονίκη: Τζιόλα.

ΕΟΔΥ (2023). Εβδομαδιαία Έκθεση Επιδημιολογικής Επιτήρησης Αναπνευστικών Λοιμώξεων: Εβδομάδα 11/2023 (13 Μαρτίου 2023 – 19 Μαρτίου 2023), Εθνικός Οργανισμός Δημόσιας Υγείας, διαθέσιμο στο Διαδίκτυο: https://eody.gov.gr/wp-content/uploads/2023/03/anapneustikon-ion-report-2023-11.pdf

# Κριτήρια αξιολόγησης

## Κριτήριο αξιολόγησης 1

**Σε τι διαφέρουν οι ποιοτικές από τις ποσοτικές μεταβλητές;**

**Απάντηση**
Οι ποιοτικές μεταβλητές δεν μετριούνται με αριθμούς, ενώ οι ποσοτικές μετριούνται αριθμητικά με κάποια μονάδα μέτρησης.

## Κριτήριο αξιολόγησης 2

**Ποια ανάγκη καλύπτει ο υπολογισμός παραμέτρων για τις ποσοτικές μεταβλητές;**

**Απάντηση**
Οι παράμετροι αντιπροσωπεύουν το μεγάλο πλήθος δεδομένων και δίνουν την εικόνα της θέσης τους και της διασποράς τους με λίγους αριθμούς.

## Κριτήριο αξιολόγησης 3

**Ποια είναι τα χαρακτηριστικά της κανονικής κατανομής;**

**Απάντηση**
Η καμπύλη της είναι συμμετρική, έχει σχήμα καμπάνας, με κέντρο τη μέση τιμή της που είναι ίση και με τη διάμεσο. Το «άνοιγμα» δεξιά και αριστερά του κέντρου της είναι 3 φορές η τυπική της απόκλιση.

## Κριτήριο αξιολόγησης 4

**Οι μηνιαίες πωλήσεις γάλακτος σε 50 παντοπωλεία ακολουθούν κανονική κατανομή με μέση τιμή 1.200 λίτρα και τυπική απόκλιση 400 λίτρα. Να βρεθεί πόσα από τα παντοπωλεία αυτά πωλούν λιγότερα από 1.000 λίτρα γάλακτος τον μήνα;**

**Λύση**
Για να βρεθεί το πλήθος παντοπωλείων, θα υπολογιστεί πρώτα η πιθανότητα ένα παντοπωλείο να πωλεί λιγότερο από 1.000 λίτρα γάλα με χρήση της κανονικής κατανομής για τη μεταβλητή X «πωλούμενα λίτρα γάλακτος».

$$P(X<1.000) = P\left((X-\mu)/\sigma < (1.000-1.200)/400\right) = P(Z<-0{,}5) = 0{,}2776 = 27{,}76\%.$$

Η πιθανότητα ένα παντοπωλείο να πωλεί λιγότερο από 1.000 λίτρα είναι 27,76%. Στα 50 παντοπωλεία, ποσοστό 27,76% θα πωλεί λιγότερο από 1.000 λίτρα γάλα, δηλαδή $50\times0{,}2776 = 13{,}88$ περίπου 14 παντοπωλεία.

# Παράρτημα:

Πίνακας τιμών P(Z<z) της τυπικής κανονικής κατανομής για αρνητικές τιμές z

| **τιμές za** | **0,00** | **0,01** | **0,02** | **0,03** | **0,04** | **0,05** | **0,06** | **0,07** | **0,08** | **0,09** |
|---|---|---|---|---|---|---|---|---|---|---|
| -3,9 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 |
| -3,8 | 0,0001 | 0,0001 | 0,0001 | 0,0001 | 0,0001 | 0,0001 | 0,0001 | 0,0001 | 0,0001 | 0,0001 |
| -3,7 | 0,0001 | 0,0001 | 0,0001 | 0,0001 | 0,0001 | 0,0001 | 0,0001 | 0,0001 | 0,0001 | 0,0001 |
| -3,6 | 0,0002 | 0,0002 | 0,0001 | 0,0001 | 0,0001 | 0,0001 | 0,0001 | 0,0001 | 0,0001 | 0,0001 |
| -3,5 | 0,0002 | 0,0002 | 0,0002 | 0,0002 | 0,0002 | 0,0002 | 0,0002 | 0,0002 | 0,0002 | 0,0002 |
| -3,4 | 0,0003 | 0,0003 | 0,0003 | 0,0003 | 0,0003 | 0,0003 | 0,0003 | 0,0003 | 0,0003 | 0,0002 |
| -3,3 | 0,0005 | 0,0005 | 0,0005 | 0,0004 | 0,0004 | 0,0004 | 0,0004 | 0,0004 | 0,0004 | 0,0003 |
| -3,2 | 0,0007 | 0,0007 | 0,0006 | 0,0006 | 0,0006 | 0,0006 | 0,0006 | 0,0005 | 0,0005 | 0,0005 |
| -3,1 | 0,0010 | 0,0009 | 0,0009 | 0,0009 | 0,0008 | 0,0008 | 0,0008 | 0,0008 | 0,0007 | 0,0007 |
| -3 | 0,0013 | 0,0013 | 0,0013 | 0,0012 | 0,0012 | 0,0011 | 0,0011 | 0,0011 | 0,0010 | 0,0010 |
| -2,9 | 0,0019 | 0,0018 | 0,0018 | 0,0017 | 0,0016 | 0,0016 | 0,0015 | 0,0015 | 0,0014 | 0,0014 |
| -2,8 | 0,0026 | 0,0025 | 0,0024 | 0,0023 | 0,0023 | 0,0022 | 0,0021 | 0,0021 | 0,0020 | 0,0019 |
| -2,7 | 0,0035 | 0,0034 | 0,0033 | 0,0032 | 0,0031 | 0,0030 | 0,0029 | 0,0028 | 0,0027 | 0,0026 |
| -2,6 | 0,0047 | 0,0045 | 0,0044 | 0,0043 | 0,0041 | 0,0040 | 0,0039 | 0,0038 | 0,0037 | 0,0036 |
| -2,5 | 0,0062 | 0,0060 | 0,0059 | 0,0057 | 0,0055 | 0,0054 | 0,0052 | 0,0051 | 0,0049 | 0,0048 |
| -2,4 | 0,0082 | 0,0080 | 0,0078 | 0,0075 | 0,0073 | 0,0071 | 0,0069 | 0,0068 | 0,0066 | 0,0064 |
| -2,3 | 0,0107 | 0,0104 | 0,0102 | 0,0099 | 0,0096 | 0,0094 | 0,0091 | 0,0089 | 0,0087 | 0,0084 |
| -2,2 | 0,0139 | 0,0136 | 0,0132 | 0,0129 | 0,0125 | 0,0122 | 0,0119 | 0,0116 | 0,0113 | 0,0110 |
| -2,1 | 0,0179 | 0,0174 | 0,0170 | 0,0166 | 0,0162 | 0,0158 | 0,0154 | 0,0150 | 0,0146 | 0,0143 |
| -2 | 0,0228 | 0,0222 | 0,0217 | 0,0212 | 0,0207 | 0,0202 | 0,0197 | 0,0192 | 0,0188 | 0,0183 |
| -1,9 | 0,0287 | 0,0281 | 0,0274 | 0,0268 | 0,0262 | 0,0256 | 0,0250 | 0,0244 | 0,0239 | 0,0233 |
| -1,8 | 0,0359 | 0,0351 | 0,0344 | 0,0336 | 0,0329 | 0,0322 | 0,0314 | 0,0307 | 0,0301 | 0,0294 |
| -1,7 | 0,0446 | 0,0436 | 0,0427 | 0,0418 | 0,0409 | 0,0401 | 0,0392 | 0,0384 | 0,0375 | 0,0367 |
| -1,6 | 0,0548 | 0,0537 | 0,0526 | 0,0516 | 0,0505 | 0,0495 | 0,0485 | 0,0475 | 0,0465 | 0,0455 |
| -1,5 | 0,0668 | 0,0655 | 0,0643 | 0,0630 | 0,0618 | 0,0606 | 0,0594 | 0,0582 | 0,0571 | 0,0559 |
| -1,4 | 0,0808 | 0,0793 | 0,0778 | 0,0764 | 0,0749 | 0,0735 | 0,0721 | 0,0708 | 0,0694 | 0,0681 |
| -1,3 | 0,0968 | 0,0951 | 0,0934 | 0,0918 | 0,0901 | 0,0885 | 0,0869 | 0,0853 | 0,0838 | 0,0823 |
| -1,2 | 0,1151 | 0,1131 | 0,1112 | 0,1093 | 0,1075 | 0,1056 | 0,1038 | 0,1020 | 0,1003 | 0,0985 |
| -1,1 | 0,1357 | 0,1335 | 0,1314 | 0,1292 | 0,1271 | 0,1251 | 0,1230 | 0,1210 | 0,1190 | 0,1170 |
| -1 | 0,1587 | 0,1562 | 0,1539 | 0,1515 | 0,1492 | 0,1469 | 0,1446 | 0,1423 | 0,1401 | 0,1379 |
| -0,9 | 0,1841 | 0,1814 | 0,1788 | 0,1762 | 0,1736 | 0,1711 | 0,1685 | 0,1660 | 0,1635 | 0,1611 |
| -0,8 | 0,2119 | 0,2090 | 0,2061 | 0,2033 | 0,2005 | 0,1977 | 0,1949 | 0,1922 | 0,1894 | 0,1867 |
| -0,7 | 0,2420 | 0,2389 | 0,2358 | 0,2327 | 0,2296 | 0,2266 | 0,2236 | 0,2206 | 0,2177 | 0,2148 |
| -0,6 | 0,2743 | 0,2709 | 0,2676 | 0,2643 | 0,2611 | 0,2578 | 0,2546 | 0,2514 | 0,2483 | 0,2451 |
| -0,5 | 0,3085 | 0,3050 | 0,3015 | 0,2981 | 0,2946 | 0,2912 | 0,2877 | 0,2843 | 0,2810 | 0,2776 |
| -0,4 | 0,3446 | 0,3409 | 0,3372 | 0,3336 | 0,3300 | 0,3264 | 0,3228 | 0,3192 | 0,3156 | 0,3121 |
| -0,3 | 0,3821 | 0,3783 | 0,3745 | 0,3707 | 0,3669 | 0,3632 | 0,3594 | 0,3557 | 0,3520 | 0,3483 |
| -0,2 | 0,4207 | 0,4168 | 0,4129 | 0,4090 | 0,4052 | 0,4013 | 0,3974 | 0,3936 | 0,3897 | 0,3859 |
| -0,1 | 0,4602 | 0,4562 | 0,4522 | 0,4483 | 0,4443 | 0,4404 | 0,4364 | 0,4325 | 0,4286 | 0,4247 |
| 0 | 0,5000 | 0,4960 | 0,4920 | 0,4880 | 0,4840 | 0,4801 | 0,4761 | 0,4721 | 0,4681 | 0,4641 |

Πίνακας τιμών P(Z<z) της τυπικής κανονικής κατανομής για θετικές τιμές z.

| **τιμές za** | **0,00** | **0,01** | **0,02** | **0,03** | **0,04** | **0,05** | **0,06** | **0,07** | **0,08** | **0,09** |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0,5000 | 0,5040 | 0,5080 | 0,5120 | 0,5160 | 0,5199 | 0,5239 | 0,5279 | 0,5319 | 0,5359 |
| 0,1 | 0,5398 | 0,5438 | 0,5478 | 0,5517 | 0,5557 | 0,5596 | 0,5636 | 0,5675 | 0,5714 | 0,5753 |
| 0,2 | 0,5793 | 0,5832 | 0,5871 | 0,5910 | 0,5948 | 0,5987 | 0,6026 | 0,6064 | 0,6103 | 0,6141 |
| 0,3 | 0,6179 | 0,6217 | 0,6255 | 0,6293 | 0,6331 | 0,6368 | 0,6406 | 0,6443 | 0,6480 | 0,6517 |
| 0,4 | 0,6554 | 0,6591 | 0,6628 | 0,6664 | 0,6700 | 0,6736 | 0,6772 | 0,6808 | 0,6844 | 0,6879 |
| 0,5 | 0,6915 | 0,6950 | 0,6985 | 0,7019 | 0,7054 | 0,7088 | 0,7123 | 0,7157 | 0,7190 | 0,7224 |
| 0,6 | 0,7257 | 0,7291 | 0,7324 | 0,7357 | 0,7389 | 0,7422 | 0,7454 | 0,7486 | 0,7517 | 0,7549 |
| 0,7 | 0,7580 | 0,7611 | 0,7642 | 0,7673 | 0,7704 | 0,7734 | 0,7764 | 0,7794 | 0,7823 | 0,7852 |
| 0,8 | 0,7881 | 0,7910 | 0,7939 | 0,7967 | 0,7995 | 0,8023 | 0,8051 | 0,8078 | 0,8106 | 0,8133 |
| 0,9 | 0,8159 | 0,8186 | 0,8212 | 0,8238 | 0,8264 | 0,8289 | 0,8315 | 0,8340 | 0,8365 | 0,8389 |
| 1 | 0,8413 | 0,8438 | 0,8461 | 0,8485 | 0,8508 | 0,8531 | 0,8554 | 0,8577 | 0,8599 | 0,8621 |
| 1,1 | 0,8643 | 0,8665 | 0,8686 | 0,8708 | 0,8729 | 0,8749 | 0,8770 | 0,8790 | 0,8810 | 0,8830 |
| 1,2 | 0,8849 | 0,8869 | 0,8888 | 0,8907 | 0,8925 | 0,8944 | 0,8962 | 0,8980 | 0,8997 | 0,9015 |
| 1,3 | 0,9032 | 0,9049 | 0,9066 | 0,9082 | 0,9099 | 0,9115 | 0,9131 | 0,9147 | 0,9162 | 0,9177 |
| 1,4 | 0,9192 | 0,9207 | 0,9222 | 0,9236 | 0,9251 | 0,9265 | 0,9279 | 0,9292 | 0,9306 | 0,9319 |
| 1,5 | 0,9332 | 0,9345 | 0,9357 | 0,9370 | 0,9382 | 0,9394 | 0,9406 | 0,9418 | 0,9429 | 0,9441 |
| 1,6 | 0,9452 | 0,9463 | 0,9474 | 0,9484 | 0,9495 | 0,9505 | 0,9515 | 0,9525 | 0,9535 | 0,9545 |
| 1,7 | 0,9554 | 0,9564 | 0,9573 | 0,9582 | 0,9591 | 0,9599 | 0,9608 | 0,9616 | 0,9625 | 0,9633 |
| 1,8 | 0,9641 | 0,9649 | 0,9656 | 0,9664 | 0,9671 | 0,9678 | 0,9686 | 0,9693 | 0,9699 | 0,9706 |
| 1,9 | 0,9713 | 0,9719 | 0,9726 | 0,9732 | 0,9738 | 0,9744 | 0,9750 | 0,9756 | 0,9761 | 0,9767 |
| 2 | 0,9772 | 0,9778 | 0,9783 | 0,9788 | 0,9793 | 0,9798 | 0,9803 | 0,9808 | 0,9812 | 0,9817 |
| 2,1 | 0,9821 | 0,9826 | 0,9830 | 0,9834 | 0,9838 | 0,9842 | 0,9846 | 0,9850 | 0,9854 | 0,9857 |
| 2,2 | 0,9861 | 0,9864 | 0,9868 | 0,9871 | 0,9875 | 0,9878 | 0,9881 | 0,9884 | 0,9887 | 0,9890 |
| 2,3 | 0,9893 | 0,9896 | 0,9898 | 0,9901 | 0,9904 | 0,9906 | 0,9909 | 0,9911 | 0,9913 | 0,9916 |
| 2,4 | 0,9918 | 0,9920 | 0,9922 | 0,9925 | 0,9927 | 0,9929 | 0,9931 | 0,9932 | 0,9934 | 0,9936 |
| 2,5 | 0,9938 | 0,9940 | 0,9941 | 0,9943 | 0,9945 | 0,9946 | 0,9948 | 0,9949 | 0,9951 | 0,9952 |
| 2,6 | 0,9953 | 0,9955 | 0,9956 | 0,9957 | 0,9959 | 0,9960 | 0,9961 | 0,9962 | 0,9963 | 0,9964 |
| 2,7 | 0,9965 | 0,9966 | 0,9967 | 0,9968 | 0,9969 | 0,9970 | 0,9971 | 0,9972 | 0,9973 | 0,9974 |
| 2,8 | 0,9974 | 0,9975 | 0,9976 | 0,9977 | 0,9977 | 0,9978 | 0,9979 | 0,9979 | 0,9980 | 0,9981 |
| 2,9 | 0,9981 | 0,9982 | 0,9982 | 0,9983 | 0,9984 | 0,9984 | 0,9985 | 0,9985 | 0,9986 | 0,9986 |
| 3 | 0,9987 | 0,9987 | 0,9987 | 0,9988 | 0,9988 | 0,9989 | 0,9989 | 0,9989 | 0,9990 | 0,9990 |
| 3,1 | 0,9990 | 0,9991 | 0,9991 | 0,9991 | 0,9992 | 0,9992 | 0,9992 | 0,9992 | 0,9993 | 0,9993 |
| 3,2 | 0,9993 | 0,9993 | 0,9994 | 0,9994 | 0,9994 | 0,9994 | 0,9994 | 0,9995 | 0,9995 | 0,9995 |
| 3,3 | 0,9995 | 0,9995 | 0,9995 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9997 |
| 3,4 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9998 |
| 3,5 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 |
| 3,6 | 0,9998 | 0,9998 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 |
| 3,7 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 |
| 3,8 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 |
| 3,9 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 |

# Κεφάλαιο 4: Διαστήματα Εμπιστοσύνης – Έλεγχοι Υποθέσεων

**Σύνοψη**
*Στο κεφάλαιο αυτό γίνεται μια εισαγωγή στην εκτίμηση μιας άγνωστης παραμέτρου του πληθυσμού, μέσω του διαστήματος εμπιστοσύνης. Αναλύονται οι βασικές αρχές του ελέγχου υποθέσεων και παρουσιάζονται παραδείγματα και εφαρμογές για την καλύτερη κατανόησή τους. Στόχος είναι η εφαρμογή των ελέγχων σε πραγματικά δεδομένα και για τον λόγο αυτό, χρησιμοποιούνται τα αποτελέσματα από το λογισμικό SPSS.*

**Προαπαιτούμενη γνώση**
*Υπολογισμός παραμέτρων, κατανομές.*

## 4.1 Εκτίμηση (από το μερικό στο γενικό)

Από τους κύριους σκοπούς της επιστήμης της Στατιστικής, εκτός από την παρουσίαση των δεδομένων ενός δείγματος, είναι και η εκτίμηση άγνωστων παραμέτρων του μελετώμενου πληθυσμού. Γνωρίζοντας τις τιμές για τον μέσο όρο εσόδων ενός δείγματος επιχειρήσεων, να μπορεί να γίνει εκτίμηση για τον μέσο όρο εσόδων όλων των παρόμοιων επιχειρήσεων. Η γνώση, δηλαδή, μερικών στοιχείων μπορεί να οδηγήσει στην εκτίμηση για το σύνολο των στοιχείων. Από το μέρος να υπολογιστούν στοιχεία για το σύνολο. Η μέθοδος αυτή στα μαθηματικά ονομάζεται επαγωγή και για τον λόγο αυτό το αντίστοιχο τμήμα της Στατιστικής ονομάζεται επαγωγική στατιστική. Οι παράμετροι του πληθυσμού που κυρίως εκτιμώνται με την επαγωγική στατιστική είναι η μέση τιμή, η διακύμανση για ποσοτικές μεταβλητές και η αναλογία εμφάνισης για ποιοτικές μεταβλητές.

## 4.2 Σημειακή εκτίμηση και εκτίμηση με διάστημα

Ένας τρόπος εκτίμησης μίας άγνωστης παραμέτρου είναι η απευθείας εκτίμηση, δηλαδή ως εκτίμηση θεωρείται η τιμή του δείγματος, που ονομάζεται **σημειακή εκτίμηση**. Μοιάζει με την προσπάθεια για την επίτευξη ενός στόχου με μία μόνο βολή. Το μειονέκτημα του τρόπου αυτού είναι ο μεγάλος κίνδυνος αποτυχίας, ενώ το πλεονέκτημά του είναι η ταχύτητα εκτίμησης.

Άλλος τρόπος εκτίμησης μίας άγνωστης παραμέτρου είναι η πρόβλεψη για ένα διάστημα τιμών, μέσα στο οποίο θα βρίσκεται η άγνωστη παράμετρος, που ονομάζεται εκτίμηση με **διάστημα εμπιστοσύνης (confidence interval)**. Μοιάζει με την επίτευξη του στόχου με πολλές διαφορετικές βολές και όχι μόνο με μία. Το μειονέκτημα είναι οι πολλές προσπάθειες, ενώ το πλεονέκτημα είναι η σταθερότητα της εκτίμησης.

## 4.3 Στάθμη σημαντικότητας και κατανομή παραμέτρου

Για την εκτίμηση ενός διαστήματος εμπιστοσύνης, προσδιορίζεται ένα διάστημα τιμών μέσα στο οποίο *αναμένεται* να βρίσκεται η άγνωστη παράμετρος. Στη Στατιστική δεν υπάρχει βεβαιότητα ότι όντως η άγνωστη παράμετρος θα βρίσκεται στο διάστημα αυτό, εκτός εάν προσδιοριστεί διάστημα από το $-\infty$ έως το $+\infty$, αλλά κάτι τέτοιο δεν έχει νόημα εκτίμησης. Το διάστημα εμπιστοσύνης δεν είναι τόσο μεγάλο, αλλά ούτε και πολύ μικρό. Είναι τόσο, όσο η πιθανότητα να βρίσκεται μέσα σε αυτό η άγνωστη παράμετρος να είναι μεγάλη.

Συνήθως υπολογίζεται 95% διάστημα εμπιστοσύνης, το οποίο στο 95% των περιπτώσεων περιέχει την άγνωστη παράμετρο, έχοντας ένα περιθώριο 5% να μην είναι σωστό. Η πιθανότητα 95% μπορεί να διαφοροποιείται σε 90% ή 99% ή κάποια άλλη, ανάλογα με το είδος των δεδομένων. Η πιθανότητα αυτή γενικά ονομάζεται **στάθμη σημαντικότητας** ή στάθμη εμπιστοσύνης του διαστήματος και προσδιορίζει την ακρίβεια εκτίμησης. Συμβολίζεται με $(1\text{-}\alpha)\%$ και το αντίστοιχο ποσοστό $\alpha\%$ μετράει την πιθανότητα σφάλματος εκτίμησης. Συνηθέστερες τιμές του α είναι το 5% ή το 10% ή το 1%.

### 4.3.1 Κατανομή (δειγματοληπτική) παραμέτρου

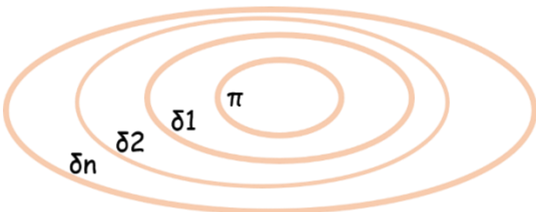
Βασικό βήμα για την εκτίμηση του διαστήματος εμπιστοσύνης είναι η γνώση της κατανομής της άγνωστης παραμέτρου, η οποία εξαρτάται από το είδος της παραμέτρου και από το μέγεθος δείγματος που χρησιμοποιείται για την εκτίμηση της.

Στη θεωρία των πιθανοτήτων, η εκτίμηση γίνεται με την εκτιμήτρια συνάρτηση. Η συνάρτηση αυτή υπολογίζεται από τις τιμές ενός δείγματος, και για να είναι σωστή θα πρέπει να έχει τις ιδιότητες της αμεροληψίας, της αποτελεσματικότητας της συνέπειας και της επάρκειας. Αμεροληψία της εκτιμήτριας συνάρτησης σημαίνει ότι η αναμενόμενη τιμή της θα είναι ίση με την τιμή της άγνωστης παραμέτρου. Αποτελεσματικότητα υπάρχει σε μία αμερόληπτη εκτιμήτρια, όταν αυτή έχει τη μικρότερη δυνατή διασπορά. Συνεπής λέγεται μια εκτιμήτρια συνάρτηση όταν, όσο μεγαλώνει το μέγεθος δείγματος, τόσο περισσότερο η τιμή της πλησιάζει στην άγνωστη παράμετρο. Επαρκής λέγεται μια εκτιμήτρια η οποία περιέχει όλες τις πληροφορίες για μία άγνωστη παράμετρο, χωρίς να υπάρχει άλλη εκτιμήτρια που να δίνει πρόσθετες πληροφορίες (Χάλκος, 2020).

Ανάλογα με το είδος της άγνωστης παραμέτρου, χρησιμοποιείται διαφορετική εκτιμήτρια συνάρτηση. Η συνάρτηση αυτή παίρνει τιμές από το δείγμα και θα πρέπει να χρησιμοποιηθούν πολλά δείγματα για να εκτιμηθεί. Μελετήθηκαν θεωρητικά οι εκτιμήτριες συναρτήσεις για τη μέση τιμή, για τη διακύμανση και για την αναλογία, και προέκυψαν οι κατανομές των εκτιμητριών τους. Οι κατανομές αυτές ονομάζονται **δειγματοληπτικές κατανομές**. Γνωρίζοντας τη θεωρητική δειγματοληπτική κατανομή, μπορεί να βρεθεί το διάστημα εμπιστοσύνης της αντίστοιχης παραμέτρου μόνο από ένα δείγμα, όπως θα δούμε στη συνέχεια.

## 4.4 Διάστημα εμπιστοσύνης για μέση τιμή, αναλογία

Η προσπάθεια εκτίμησης μίας άγνωστης παραμέτρου π ενός πληθυσμού μοιάζει σαν την προσπάθεια επίτευξης ενός στόχου σκοποβολής (Εικόνα 4.1), όπου στο κέντρο βρίσκεται το π και με τις εκτιμήσεις από διάφορα δείγματα δ1, δ2, …, δn άλλη φορά η βολή είναι κοντά και άλλη μακριά από τον στόχο. Δυστυχώς, στη Στατιστική (όπως και στην πραγματικότητα), το κέντρο του στόχου δεν φαίνεται. Υποτίθεται ότι βρίσκεται εκεί και αφού γίνει η εκτίμηση από κάποιο δείγμα, ελέγχεται η υπόθεση για την τοποθεσία του κέντρου του στόχου.

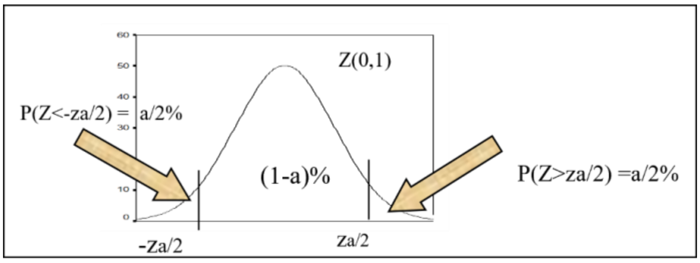



**Εικόνα 4.1** *Προσπάθεια εκτίμησης της άγνωστης παραμέτρου π από διαφορετικά δείγματα.*

Για τον προσδιορισμό του διαστήματος εμπιστοσύνης της άγνωστης παραμέτρου με στάθμη σημαντικότητας (1-α)%, χρησιμοποιείται η συμμετρική ιδιότητα της τυπικής κανονικής κατανομής:

$$P(-z_{a/2}<Z<z_{a/2}) = (1\text{-}a)\%$$

Δηλαδή ότι στο διάστημα ($-z_{a/2}$, $+z_{a/2}$) περιέχεται το (1-α)% των τιμών της και έξω από το διάστημα αυτό υπάρχει το α% των τιμών, κατανεμημένο εξίσου στο δεξί και αριστερό άκρο της καμπύλης της (Εικόνα 4.2).




**Εικόνα 4.2** *Εμβαδόν τυπικής κανονικής κατανομής με πιθανότητα (1-α)%.*

Γενικότερα για την τυχαία μεταβλητή Z που ακολουθεί τυπική κανονική κατανομή ισχύει το εξής διάστημα εμπιστοσύνης με στάθμη σημαντικότητας (1-α)% : $[-z_{a/2}, +z_{a/2}]$. Μία οποιαδήποτε άλλη μεταβλητή X που ακολουθεί κανονική κατανομή με μέσο E(X) και διακύμανση Var(X), μετατρέπεται σε τυπική κανονική κατανομή με τον μετασχηματισμό X-E(X)/√Var(X). Ισχύει:

$$P(-z_{a/2} < X-E(X)/\sqrt{Var(X)}. < z_{a/2}) = (1-a)\% \Leftrightarrow$$

$$P(E(X)-z_{a/2}\sqrt{Var(X)} < X < E(X)+z_{a/2}\sqrt{Var(X)}) = (1-a)\%$$

Οπότε, το διάστημα εμπιστοσύνης (1-α)% για τη μεταβλητή X θα είναι:

$$[E(X)-z_{a/2}\sqrt{Var(X)} , E(X)+z_{a/2}\sqrt{Var(X)}]$$

ή πιο απλά:

[μέση τιμή X-$z_{a/2}$ επί τυπική απόκλιση X, μέση τιμή X +$z_{a/2}$ επί τυπική απόκλιση X]

**Πίνακας 4.1** *Τιμές τυπικής κανονικής κατανομής ανάλογα με τη στάθμη σημαντικότητας.*

| Στάθμη σημαντικότητας | Σφάλμα α% | α/2 % | Τιμή $z_{a/2}$ |
|---|---|---|---|
| 95% | 5% | 2,5% | 1,96 |
| 90% | 10% | 5,0% | 1,65 |
| 99% | 1% | 0,5% | 2,58 |

Γνωρίζοντας τη μέση τιμή και την τυπική απόκλιση της X και τις τιμές $z_{a/2}$ είναι εύκολο να υπολογιστεί το αντίστοιχο διάστημα εμπιστοσύνης με την παραπάνω σχέση. Οι τιμές $z_{a/2}$ βρίσκονται από τον πίνακα πιθανοτήτων της τυπικής κανονικής κατανομής και οι συνηθέστερες εμφανίζονται στον Πίνακα 4.1.

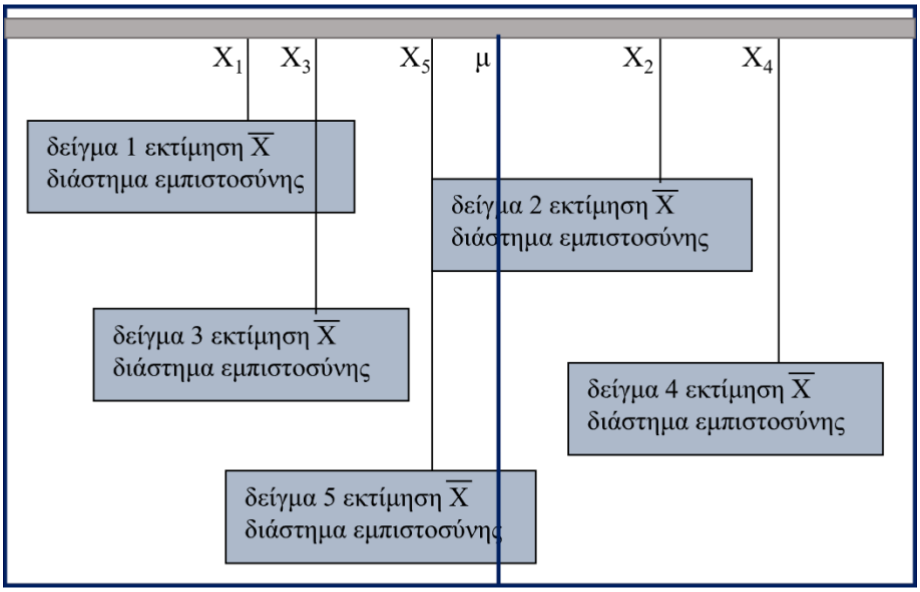
## 4.4.1 Διάστημα εμπιστοσύνης για τη μέση τιμή πληθυσμού από δείγμα

Αποδεικνύεται ότι δειγματοληπτική κατανομή της μέσης τιμής $\bar{X}$ δείγματος από έναν πληθυσμό (με πλήθος N), είναι μια κανονική κατανομή με μέση τιμή μ (άγνωστη) και διακύμανση $\sigma^2(N-n)/n(N-1)$, όπου $\sigma^2$ είναι η διακύμανση του πληθυσμού και n το μέγεθος του δείγματος που χρησιμοποιήθηκε. Στην περίπτωση που το N είναι πολύ μεγάλο σε σχέση με το μέγεθος του δείγματος, η διακύμανση της δειγματοληπτικής κατανομής είναι απλά $\sigma^2/n$. Η απόδειξη των παραπάνω, όταν το μέγεθος δείγματος είναι μεγάλο, έγινε με το λεγόμενο Κεντρικό Οριακό Θεώρημα.

Το διάστημα εμπιστοσύνης για την εκτίμηση της μέσης τιμής μ του πληθυσμού θα είναι:

$[\bar{X}-z_{a/2} \frac{\sigma}{\sqrt{n}}, \bar{X}+z_{a/2} \frac{\sigma}{\sqrt{n}}]$ ή για πεπερασμένο N : $[\bar{X}-z_{a/2} \frac{\sigma}{\sqrt{n}} \frac{N-n}{N-1}, \bar{X}+z_{a/2} \frac{\sigma}{\sqrt{n}} \frac{N-n}{N-1}]$

Η χρήση του παραπάνω τύπου δίνει διάστημα εμπιστοσύνης για την εκτίμηση μίας τιμής, αλλά όχι τη βεβαιότητα (100%) ότι το διάστημα αυτό περιέχει την αντίστοιχη τιμή. Στην Εικόνα 4.3 απεικονίζονται 5 διαφορετικά διαστήματα εμπιστοσύνης, που υπολογίστηκαν με 5 διαφορετικά δείγματα, για την πραγματική μέση τιμή μ, με την ίδια στάθμη σημαντικότητας.




**Εικόνα 4.3** *Διαστήματα εμπιστοσύνης (1-α)% για το μ από 5 διαφορετικά δείγματα.*

Έστω ότι το πραγματικό μ είναι γνωστό, στην εικόνα φαίνεται πόσο κοντά στο μ είναι ο δειγματικός μέσος κάθε δείγματος. Τα διαστήματα εμπιστοσύνης που υπολογίζονται από τα δείγματα 2 και 3 περιέχουν τα πραγματικό μ, ενώ αυτά που υπολογίζονται από τα δείγματα 1, 4 και 5 δεν περιέχουν το πραγματικό μ. Όσο κοντύτερα στην άγνωστη πραγματική τιμή είναι η εκτιμημένη τιμή του δείγματος, τόσο πιθανότερο είναι το διάστημα εμπιστοσύνης να περιέχει την άγνωστη πραγματική τιμή.

Στη συνήθη περίπτωση που η διακύμανση $\sigma^2$ του πληθυσμού είναι άγνωστη αλλά το μέγεθος δείγματος είναι μεγάλο, δηλαδή n>30 παρατηρήσεις, αποδεικνύεται πάλι με το Κεντρικό Οριακό Θεώρημα ότι η δειγματοληπτική κατανομή είναι πάλι η κανονική κατανομή με μέσο μ και διακύμανση $s^2/n$ όπου $s^2$ είναι η διακύμανση του δείγματος.

Το διάστημα εμπιστοσύνης για την εκτίμηση της μέσης τιμής μ του πληθυσμού θα είναι:

$$[\bar{X}\text{-}z_{a/2}\,\frac{s}{\sqrt{n}},\ \bar{X}\text{+}z_{a/2}\,\frac{s}{\sqrt{n}}]$$

Στην περίπτωση που η διακύμανση $\sigma^2$ του πληθυσμού είναι άγνωστη αλλά το μέγεθος δείγματος είναι μικρό, δηλαδή n<30 παρατηρήσεις, τότε η δειγματοληπτική κατανομή δεν είναι κανονική κατανομή, αλλά είναι η κατανομή t (student) με n-1 βαθμούς ελευθερίας.

Στην περίπτωση αυτή, το διάστημα εμπιστοσύνης για την εκτίμηση της μέσης τιμής μ του πληθυσμού θα είναι:

$$[\bar{X}\text{-}t_{a/2,n-1}\,\frac{s}{\sqrt{n}},\ \bar{X}\text{+}t_{a/2,n-1}\frac{s}{\sqrt{n}}]$$

Οι τιμές $t_{a/2,n-1}$ εντοπίζονται στον πίνακα της κατανομής t για το αντίστοιχο α και για βαθμούς ελευθερίας n-1. Για n >30 οι τιμές της κατανομής t, συμπίπτουν με τις τιμές της τυπικής κανονικής κατανομής.

Στους τύπους των διαστημάτων εμπιστοσύνης για τη μέση τιμή, το μέγεθος δείγματος n είναι στον παρονομαστή, που σημαίνει ότι όσο μεγαλώνει το μέγεθος δείγματος τόσο μικραίνει το εύρος του διαστήματος εμπιστοσύνης και τόσο ακριβέστερη γίνεται η εκτίμηση.

Για να προσδιοριστεί το μέγεθος δείγματος ώστε η ακρίβεια του διαστήματος εμπιστοσύνης να είναι E (Χάλκος, 2020), δηλαδή |X-μ|=E, χρησιμοποιείται η σχέση:

$$|X-\mu|/\sigma/\sqrt{n} = z_{a/2} \Leftrightarrow E = z_{a/2}\, \sigma /\sqrt{n} \Leftrightarrow n = ( z_{a/2}\, \sigma /E)^2$$

## 4.4.2 Διάστημα εμπιστοσύνης για την αναλογία πληθυσμού από δείγμα

Αποδεικνύεται ότι δειγματοληπτική κατανομή της αναλογίας p δείγματος από έναν πληθυσμό είναι μία κανονική κατανομή με μέση τιμή την άγνωστη αναλογία π του πληθυσμού και διακύμανση π(1-π)/n, όπου n το μέγεθος του δείγματος που χρησιμοποιήθηκε. Επειδή το π είναι άγνωστο, στην περίπτωση που ισχύει np>5 και n(1-p)>5 η διακύμανση μιας δειγματοληπτικής κατανομής υπολογίζεται με τη σχέση p(1-p)/n.

Το διάστημα εμπιστοσύνης για την εκτίμησης της αναλογίας π του πληθυσμού θα είναι:

$$[p\text{-}z_{a/2} \sqrt{(p(1-p))/n},\ \ p+ z_{a/2} \sqrt{(p(1-p))/n}]$$

ή για πεπερασμένο πλήθος N του πληθυσμού:

$$[p\text{-}z_{a/2} \sqrt{(p(1-p))/n}\sqrt{\frac{N-n}{N-1}},\ \ p+ z_{a/2} \sqrt{(p(1-p))/n}\sqrt{\frac{N-n}{N-1}}]$$

Στον τύπο του διαστήματος εμπιστοσύνης για την αναλογία, το μέγεθος δείγματος n είναι στον παρονομαστή, που σημαίνει ότι όσο μεγαλώνει το μέγεθος δείγματος τόσο μικραίνει το εύρος του διαστήματος εμπιστοσύνης και τόσο ακριβέστερη γίνεται η εκτίμηση.

Για να προσδιοριστεί το μέγεθος δείγματος ώστε η ακρίβεια του διαστήματος εμπιστοσύνης να είναι E (Χάλκος, 2020), δηλαδή |p-π|=E, χρησιμοποιείται η σχέση:

$$|\ p-\pi|/\sqrt{(p(1-p))/n} = z_{a/2} \ \Leftrightarrow E = z_{a/2} \sqrt{(p(1-p))/n} \Leftrightarrow\ n = p(1-p)\ ( z_{a/2} /E)^2$$

Επειδή p(1-p) < = ¼ το μέγεθος δείγματος για ακρίβεια E θα είνα¼ = 1/4 ( $z_{a/2}$ /E)$^2$.

Στην περίπτωση που αντί για την αναλογία ενδιαφέρει να βρεθεί το πλήθος φορών που εμφανίζεται μία τιμή, το διάστημα εμπιστοσύνης πολλαπλασιάζεται με το μέγεθος N του πληθυσμού, όταν αυτό είναι γνωστό. Στην περίπτωση αυτή, το διάστημα εμπιστοσύνης για το πλήθος εμφανίσεων θα είναι:

$$N[p\text{-}z_{a/2} \sqrt{(p(1-p))/n},\ \ p+ z_{a/2} \sqrt{(p(1-p))/n}]$$

## 4.4.3 Διάστημα εμπιστοσύνης για τη διακύμανση πληθυσμού από δείγμα

Αποδεικνύεται ότι δειγματοληπτική κατανομή της τυχαίας μεταβλητής $s^2(n-1)/\sigma^2$ όπου $\sigma^2$ είναι η διακύμανση του πληθυσμού και n το μέγεθος του δείγματος που χρησιμοποιήθηκε, ακολουθεί την κατανομή $X^2$ με n-1 βαθμούς ελευθερίας. Η κατανομή $X^2$ κατανομή δεν είναι συμμετρική. Υπάρχει πίνακας τιμών για πιθανότητες α% ανάλογα με τους βαθμούς ελευθερίας της n, έτσι ώστε $P(X<X^2_{n,a/2})$ = α/2% και $P(X<X^2_{n,1-a/2})$ = (1-α/2)%. Στην περίπτωση αυτή, το διάστημα εμπιστοσύνης για την εκτίμηση της διακύμανσης $\sigma^2$ του πληθυσμού θα υπολογιστεί από τη σχέση (Χάλκος, 2020):

$$P( \frac{(n-1)s^2}{X^2_{n,1-a/2}} < \sigma^2 < \frac{(n-1)s^2}{X^2_{n,a/2}}) = (1-\alpha)\%$$

# 4.5 Έλεγχος υποθέσεων, μηδενική υπόθεση

## 4.5.1 Πώς δημιουργείται και πώς ελέγχεται μία υπόθεση

Ως παράδειγμα για το πώς διατυπώνεται και πώς ελέγχεται μία υπόθεση στη Στατιστική αναφέρεται η «διαρροή» νερού στους υδραυλικούς σωλήνες ενός σπιτιού. Έστω ότι αυτοί δεν είναι ορατοί, καθώς βρίσκονται κάτω από τα πλακάκια του πατώματος. Γίνεται μία υπόθεση ότι η διαρροή προέρχεται από τις υδραυλικές εγκαταστάσεις πόσιμου νερού. Κλείνεται ο γενικός διακόπτης, ώστε να φανεί αν σταματάει η διαρροή. Αν ναι, αποφασίζεται ότι βρέθηκε το πρόβλημα της διαρροής. Αν όχι, γίνεται νέα υπόθεση ότι η διαρροή προέρχεται από τις υδραυλικές εγκαταστάσεις θέρμανσης. Κλείνεται ο γενικός διακόπτης ώστε να φανεί αν σταματάει η διαρροή. Επαναλαμβάνεται η παρατήρηση για τον τερματισμό ή όχι της διαρροής, γίνεται προσπάθεια με άλλη εγκατάσταση έως ότου αναγνωριστεί το προβληματικό σύστημα υδραυλικών εγκαταστάσεων.

Μην έχοντας πλήρη στοιχεία για τον πληθυσμό που μελετάται, υποτίθεται ότι οι παράμετροι του πληθυσμού έχουν κάποιες τιμές και με βάση την υπόθεση μετριέται πόσο «πιθανό» είναι να βρεθεί η συγκεκριμένη τιμή ενός δείγματος που ελήφθη. Το δείγμα δίνει πάντα ακριβείς υπολογισμούς, αλλά η υπόθεση για τον πληθυσμό γίνεται «στα τυφλά» και επιβεβαιώνεται (αν ισχύει) με τις τιμές του δείγματος.

Αν η πιθανότητα να βρεθεί η συγκεκριμένη τιμή του δείγματος είναι μεγάλη, μπορεί να εξαχθεί το συμπέρασμα ότι η τιμή του πληθυσμού που υποτέθηκε είναι σωστή. Αν όχι, θα πρέπει να ξανατεθεί νέα υπόθεση για την παράμετρο του πληθυσμού.

## 4.5.2 Τυχαίες και σημαντικές διαφορές

Ας υποθέσουμε ότι χωρίζεται ένας πληθυσμός σε δύο ομάδες-δείγματα με τυχαίο τρόπο. Υπολογίζεται η μέση τιμή για το πρώτο δείγμα και η μέση τιμή για το δεύτερο δείγμα. Αναμένεται οι δύο αυτές μέσες τιμές να είναι πολύ κοντά η μία στην άλλη, αλλά να είναι ακριβώς ίσες είναι πάρα πολύ σπάνιο. Τέτοιες μικρές διαφορές που οφείλονται στην τύχη ονομάζονται «τυχαίες» διαφορές. Αν, όμως, τα δύο δείγματα χωριστούν όχι με τυχαίο τρόπο, π.χ. με βάση κάποιο άλλο χαρακτηριστικό τους, οι μέσες τιμές των δύο δειγμάτων μπορεί να διαφέρουν πολύ μεταξύ τους, οπότε υπάρχουν «σημαντικές» διαφορές.

Το ερώτημα που τίθεται είναι, πότε μία διαφορά μεταξύ δύο δειγμάτων θεωρείται τυχαία και πότε θεωρείται σημαντική; Ποτέ δεν υπάρχει βεβαιότητα απάντησης στην παραπάνω ερώτηση. Η απάντηση στο ερώτημα αυτό θα δοθεί χρησιμοποιώντας πιθανότητα. Όταν η θεωρητική πιθανότητα να συμβεί μία διαφορά είναι μικρότερη από το 5%, θεωρείται ότι η διαφορά συμβαίνει με τυχαίο τρόπο, είναι δηλαδή απίθανο να οφείλεται σε σημαντική διαφορά. Παρ' όλα αυτά, όμως, υπάρχει η μικρή πιθανότητα 5% η διαφορά να είναι σημαντική στην πραγματικότητα. Όταν η θεωρητική πιθανότητα να συμβεί μία διαφορά είναι μεγαλύτερη από το 5%, θεωρείται ότι η διαφορά δεν συμβαίνει με τυχαίο τρόπο, είναι δηλαδή σημαντική διαφορά.

## 4.5.3 Μηδενική υπόθεση H0

Η αρχική υπόθεση που γίνεται σε έναν έλεγχο υποθέσεων ονομάζεται **μηδενική υπόθεση** και συμβολίζεται με H0. Η αντίθετη της υπόθεσης αυτής (όχι H0) ονομάζεται **εναλλακτική υπόθεση** και συμβολίζεται με H1. Η μηδενική υπόθεση H0 συνήθως είναι μία συγκεκριμένη τιμή της άγνωστης παραμέτρου του πληθυσμού, εκφράζεται με μία ισότητα και όχι με ένα διάστημα τιμών. Η εναλλακτική υπόθεση H1 συνήθως εκφράζεται με διάφορο ($\neq$) ή με μία ανισότητα ($<$ ή $>$) ή απλά «Δεν ισχύει η H0».

Αφού διατυπωθούν η μηδενική και η εναλλακτική υπόθεση, θεωρείται προσωρινά ότι ισχύει η μηδενική υπόθεση. Στη συνέχεια, επιλέγεται το κατάλληλο κριτήριο για τον έλεγχό της, και με βάση το κριτήριο λαμβάνεται η κατάλληλη απόφαση.

Η κατάλληλη απόφαση είναι είτε η απόρριψη της H0 είτε ή μη απόρριψή της. Η σωστή έκφραση για την απόφαση είναι: «Δεν μπορούμε να απορρίψουμε την H0» σε περίπτωση μη απόρριψης. Για την περίπτωση της απόρριψής της, η έκφραση της απόφασης είναι: «Η H0 απορρίπτεται».

## 4.6 Σφάλματα στον έλεγχο υποθέσεων

Στον έλεγχο υποθέσεων στη Στατιστική, γίνεται υπόθεση (συμβολίζεται με H0) που αναφέρεται στην εκτίμηση για την τιμή μιας παραμέτρου του πληθυσμού. Η απόφαση αναφέρεται στο αν θα γίνει δεκτή ή όχι η υπόθεση. Τα σφάλματα που μπορεί να γίνουν είναι δύο ειδών:

- Η υπόθεση H0 να είναι σωστή και η απόφαση να είναι λανθασμένη. (Λάθος επειδή απορρίπτεται μία σωστή υπόθεση.)
- Η υπόθεση H0 να είναι λανθασμένη και η απόφαση να είναι λανθασμένη. (Λάθος επειδή γίνεται δεκτή μία λανθασμένη υπόθεση)

Το πρώτο είδος σφάλματος ονομάζεται σφάλμα τύπου α και το δεύτερο είδος σφάλμα τύπου β. Το σφάλμα **τύπου α** χρησιμοποιείται σχεδόν πάντα, είναι εύκολο να υπολογιστεί και αντιστοιχεί στην **πιθανότητα να μην γίνει δεκτή μία σωστή υπόθεση**. Ερμηνεύεται ως σφάλμα α% και η πιθανότητα 1-α ονομάζεται στάθμη εμπιστοσύνης ή επίπεδο εμπιστοσύνης του ελέγχου.

Το επίπεδο εμπιστοσύνης (confidence level) είναι το ποσοστό των φορών που μία διαδικασία εκτίμησης είναι σωστή. Το επίπεδο σημαντικότητας (significance level) μετρά πόσο συχνά το αποτέλεσμα θα είναι εσφαλμένο. Πολύ συχνά το ανεκτό σφάλμα α είναι 5%, οπότε η συνηθέστερη στάθμη (επίπεδο) εμπιστοσύνης θα είναι 1-0,05=95%. Κάποιες φορές το ανεκτό σφάλμα α=10%, οπότε η στάθμη (επίπεδο) εμπιστοσύνης θα είναι 1-0,10=90%. Άλλες φορές η ανοχή σφάλματος είναι α=1%, οπότε η στάθμη (επίπεδο) εμπιστοσύνης θα είναι 1-0,01=99%.

## 4.7 Υπολογισμός κριτηρίων για τη μηδενική υπόθεση

Στον Πίνακα 4.2 παρουσιάζονται τα βήματα της διαδικασίας ελέγχου υποθέσεων για οποιονδήποτε στατιστικό έλεγχο υποθέσεων, είτε αυτός αφορά μία παράμετρο πληθυσμού είτε κάποια άλλη υπόθεση, όπως ισότητα μέσων τιμών ή αναλογιών δύο πληθυσμών, ανεξαρτησία δύο μεταβλητών, είδος κατανομής μίας μεταβλητής κ.ά. Τα βήματα 1, 2, 4 και 5 είναι κοινά για όλα τα είδη ελέγχων. Αυτό που διαφοροποιείται ανάλογα με το είδος της υπόθεσης H0 είναι το βήμα 3, δηλαδή η επιλογή του κατάλληλου κριτηρίου.

**Πίνακας 4.2** *Βήματα ελέγχου υποθέσεων.*

| Βήμα | Διαδικασία |
|---|---|
| 1 | Ορίζεται η H0 |
| 2 | Ορίζεται η H1 |
| 3 | Επιλέγεται το κατάλληλο κριτήριο ανάλογα με το είδος ελέγχου. |
| 4 | Υπολογίζεται η περιοχή απόρριψης ανάλογα με το ανεκτό σφάλμα α% ή το p value, sig. level από το στατιστικό πρόγραμμα. |
| 5 | Λαμβάνεται η απόφαση για την απόρριψη ή μη απόρριψη της H0. |

Στο βήμα 3 του Πίνακα 4.2, η επιλογή του κριτηρίου εξαρτάται από το είδος της υπόθεσης H0, από το μέγεθος του δείγματος ή των δειγμάτων που χρησιμοποιούνται, από την υποτιθέμενη κατανομή του πληθυσμού, από το είδος της μεταβλητής για την οποία γίνεται η υπόθεση H0 και από το είδος του ελέγχου που εφαρμόζεται. Το κατάλληλο κριτήριο εφαρμόζεται για τις εκτιμήσεις που προκύπτουν από ένα δείγμα μεγέθους n ή από δύο ή περισσότερα δείγματα όταν πρόκειται για σύγκριση μέσων τιμών πληθυσμών ή ομάδων.

Τα βήματα του Πίνακα 4.2 εφαρμόζονται τόσο σε παραμετρικούς όσο και σε μη παραμετρικούς ελέγχους.

### 4.7.1 Κριτήρια παραμετρικών ελέγχων

Τους ελέγχους υποθέσεων για την τιμή κάποιας παραμέτρου πληθυσμού, τους ονομάζουμε παραμετρικούς ελέγχους. Η υπόθεση Η0 αφορά μία παράμετρο του πληθυσμού, συνήθως τη μέση τιμή. Υπολογίζεται η κατανομή που πρέπει να ακολουθείται αν η Η0 ισχύει, και συγκρίνεται η τιμή του κριτηρίου που προκύπτει από κάποιο δείγμα με την αντίστοιχη δειγματοληπτική κατανομή που ακολουθεί το κριτήριο με βάση τη θεωρία και την ισχύ της Η0. Αν η πιθανότητα (υπολογίζεται από την αντίστοιχη κατανομή) της εμφάνισης τιμής που προέκυψε από το κριτήριο είναι μεγάλη, δεν μπορεί να απορριφθεί η Η0. Αν η πιθανότητα εμφάνισης της τιμής του κριτηρίου είναι μικρή (μικρότερη από το σφάλμα τύπου α), η υπόθεση Η0 μπορεί να απορριφθεί.

#### 4.7.1.1 Κριτήρια παραμετρικών ελέγχων

Στον Πίνακα 4.3 περιγράφονται τα κριτήρια για περιπτώσεις ελέγχου υποθέσεων παραμέτρων πληθυσμού, καθώς και η κατανομή που ακολουθεί κάθε κριτήριο.

**Πίνακας 4.3** *Κριτήρια παραμετρικών ελέγχων υποθέσεων.*

| **Είδος ελέγχου - υπόθεσης** | **Μέγεθος δείγματος** | **Κριτήριο και κατανομή** |
|---|---|---|
| μέση τιμή μ ενός πληθυσμού<br>Η0 $\mu=\mu_0$<br>Η1 $\mu \neq \mu_0$ | n >30 | $\dfrac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}} \sim N(0,1)$ |
| μέση τιμή μ ενός πληθυσμού<br>Η0 $\mu=\mu_0$<br>Η1 $\mu \neq \mu_0$ | n <30 | $\dfrac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}} \sim t_{\alpha/2,n-1}$ |
| Αναλογία π ενός πληθυσμού<br>Η0 $\pi = p_o$<br>Η1 $\pi \neq p_o$ | n >30 | $\dfrac{\tilde{p}-p_0}{\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}} \sim N(0,1)$ |
| Σύγκριση δύο μέσων τιμών<br>Η0 $\mu_1=\mu_2$<br>Η1 $\mu_1 \neq \mu_2$ | Μεγάλα δείγματα | $\dfrac{(\overline{X}_1-\overline{X}_2)}{s} \sim N(0,1)$ $\quad s=\sqrt{\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}}$ |
| Σύγκριση δύο μέσων τιμών<br>Η0 $\mu_1=\mu_2$<br>Η1 $\mu_1 \neq \mu_2$ | Μικρά δείγματα | $\dfrac{(\overline{X}_1-\overline{X}_2)}{s_t} \sim t_{\alpha/2,n1+n2-2}$ $\quad s_t = s\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}$ $\quad s=\sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$ |
| Σύγκριση δύο αναλογιών<br>Η0 $p_1=p_2$<br>Η1 $p_1 \neq p_2$ | Μεγάλα δείγματα | $\dfrac{(\hat{p}_1-\hat{p}_2)}{s_p} \sim N(0,1)$ $\quad s_p=\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}$ $\quad \hat{p}=\dfrac{n_1p_1+n_2p_2}{n_1+n_2}$ |
| Σύγκριση πολλών k μέσων τιμών (Ανάλυση Διακύμανσης)<br>Η0 $\mu_1=\mu_2=\mu_{3=..}=\mu_k$<br>Η1 όχι Η0 | k δείγματα ( ή ομάδες), ανεξάρτητα μεταξύ τους από κανονικές κατανομές, που έχουν ίδια διακύμανση | $\dfrac{s_b^2}{s_e^2} \sim F_{k-1,n-k}$ $\quad$ $S_b^2$ Διακύμανση μεταξύ των μέσων τιμών $\quad$ $S_e^2$ Διακύμανση υπολοίπων, άθροισμα μέσα σε κάθε δείγμα (ή ομάδα) |

### 4.7.1.2 Ανάλυση διακύμανσης

Ειδικότερα στην περίπτωση της σύγκρισης πολλών μέσων τιμών, η μέθοδος ονομάζεται **ανάλυση διακύμανσης** (ANOVA) και προϋπόθεσή της είναι οι ομάδες-δείγματα να προέρχονται από κανονικούς πληθυσμούς με ίδιες διακυμάνσεις. Στην περίπτωση αυτή, αν ισχύει η H0, συγκρίνεται η διακύμανση μεταξύ των δειγμάτων με τη διακύμανση των υπολοίπων που προκύπτουν μέσα από κάθε δείγμα. Ο λόγος αυτών των διακυμάνσεων πρέπει να ακολουθεί την F κατανομή.

Προσοχή πρέπει να δοθεί στο ότι η μέθοδος ονομάζεται ανάλυση διακύμανσης λόγω χρήσης κριτηρίου για τις διακυμάνσεις, αλλά ο έλεγχος υπόθεσης γίνεται για την ισότητα των μέσων τιμών. Στην περίπτωση που απορριφθεί η H0, σημαίνει ότι μία τουλάχιστον μέση τιμή διαφέρει από τις υπόλοιπες.

Για να βρεθεί ποια ή ποιες μέσες τιμές διαφέρουν, γίνεται έλεγχος ανά δύο μέσων τιμών των δειγμάτων-ομάδων. Ο έλεγχος αυτός δεν γίνεται με το t κριτήριο για δύο μέσες τιμές. Για μεγαλύτερη αξιοπιστία, γίνεται με διάφορα κριτήρια πολλαπλών συγκρίσεων, όπως Fisher (LSD), Tukey, Scheffe, Duncan και άλλα. Από τον έλεγχο πολλαπλών συγκρίσεων, προκύπτουν ομάδες-δείγματα με παρόμοιες μέσες τιμές.

## 4.7.2 Μη παραμετρικοί έλεγχοι

Τους ελέγχους που αφορούν το είδος κατανομής που ακολουθεί κάποια μεταβλητή ή τους ελέγχους για την ανεξαρτησία δύο μεταβλητών (Κεφάλαιο 5) ή τους ελέγχους για την ομοιογένεια πληθυσμών και άλλους, που δεν υποθέτουν τιμή παραμέτρου, τους ονομάζουμε μη παραμετρικούς ελέγχους. Οι μη παραμετρικοί έλεγχοι είναι πολυποίκιλοι και επίσης εφαρμόζονται όταν δεν ισχύουν κάποιες προϋποθέσεις ή όταν τα δείγματα είναι μικρά.

Κάποιοι μη παραμετρικοί έλεγχοι χρησιμοποιούνται σε δεδομένα που μπορούν να διαταχθούν από το μικρότερο στο μεγαλύτερο. Υπάρχουν μη παραμετρικοί έλεγχοι που χρησιμοποιούνται για δύο δείγματα ανεξάρτητα ή για δύο δείγματα εξαρτημένα ή για περισσότερα δείγματα ανεξάρτητα ή εξαρτημένα.

Στον Πίνακα 4.4 εμφανίζονται οι πιο συχνά χρησιμοποιούμενοι μη παραμετρικοί έλεγχοι.

**Πίνακας 4.4** *Κριτική τιμή και απόφαση απόρριψης H0.*

| **Πλήθος και είδος δειγμάτων** | **Μη παραμετρικός έλεγχος** | **H0** | **Διαδικασία** |
|---|---|---|---|
| Δύο ανεξάρτητα δείγματα | Wilcoxon ή Mann-Whitney | Ίδια θέση δεδομένων | Σύγκριση «τάξεων» δεδομένων |
| Δύο εξαρτημένα δείγματα | Wilcoxon | Ίδια θέση δεδομένων | Σύγκριση πρόσημων διαφορών |
| Περισσότερα από δύο ανεξάρτητα δείγματα | Kruskal Wallis | Ίδια κατανομή δεδομένων | Σύγκριση «τάξεων» δεδομένων |
| Περισσότερα από δύο εξαρτημένα δείγματα | Friedman | Ίδια κατανομή δεδομένων | Σύγκριση «τάξεων» διαφορών |
| Ένα δείγμα | Run test | Τυχαία κατανεμημένα δεδομένα | Σύγκριση εμφανίσεων για τιμές μικρότερες ή μεγαλύτερες από τη διάμεσο ή άλλη τιμή |
| Ένα δείγμα με δίτιμη μεταβλητή | Binomial | Η αναλογία στον πληθυσμό $=p_0$ | Αναλογία εμφανίσεων της μίας τιμής |

Όταν η εξαρτημένη μεταβλητή ελέγχου δεν είναι ποσοτική αλλά είναι διατάξιμη (διαβαθμιζόμενη κλίμακα τιμών), δεν έχει νόημα η μέση τιμή αλλά οι θέσεις των δεδομένων. Στην περίπτωση αυτή ελέγχεται η υπόθεση H0 για το αν οι θέσεις των δεδομένων είναι ίδιες.

Όταν οι κατανομές των δειγμάτων-ομάδων δεν είναι κανονικές με ίσες διακυμάνσεις, δεν μπορεί να εφαρμοστεί ο έλεγχος υποθέσεων ισότητας μέσων τιμών. Στην περίπτωση αυτή τα δεδομένα θεωρούνται ως διατάξιμα δεδομένα και έτσι ελέγχεται η μηδενική υπόθεση H0 για το εάν προέρχονται από την ίδια κατανομή ή όχι.

**Πίνακας 4.5** *Κριτική τιμή και απόφαση απόρριψης H0.*

| **Στάθμη σημαντικότητας** | **Ανεκτό σφάλμα α%** | **Sig level ή p-value (Πιθανότητα λάθους για απόρριψη σωστής H0)** | **Απόφαση για H0** |
|---|---|---|---|
| (1-α)% | α% | Sig <α | απόρριψη |
| (1-α)% | α% | Sig >α | Όχι απόρριψη |

## 4.8 Απόφαση για την αποδοχή της μηδενικής υπόθεσης

Με τη μεγάλη ανάπτυξη των υπολογιστών, αντί για το σφάλμα τύπου α, υπολογίζεται αυτόματα από τα στατιστικά προγράμματα η κριτική τιμή, δηλαδή η πιθανότητα να απορριφθεί η υπόθεση αν και είναι σωστή. Η πιθανότητα αυτή ονομάζεται είτε **Sig. level** (significant level) είτε **p value** (probability value).

Αν η κριτική τιμή αυτή είναι μικρότερη από το σφάλμα τύπου α, σημαίνει ότι μπορεί να απορριφθεί η υπόθεση H0. Αν η τιμή αυτή είναι μεγαλύτερη από το σφάλμα τύπου α, τότε δεν μπορεί να απορριφθεί η υπόθεση H0 (Πίνακας 4.5).

## 4.9 Εφαρμογές για κατανόηση

### 4.9.1 Έλεγχος επιβατών

Ένας οργανισμός αστικών συγκοινωνιών κατήργησε τα μηχανήματα έκδοσης εισιτηρίων μέσα στα οχήματα, και οι επιβάτες πρέπει να έχουν το εισιτήριό τους πριν ανέβουν στο όχημα. Σε μία διαδρομή ελέγχθηκαν 300 επιβάτες και βρέθηκαν 63 χωρίς εισιτήριο. Να εκτιμήσετε ένα 95% διάστημα εμπιστοσύνης για την αναλογία των επιβατών που επιβιβάζονται χωρίς εισιτήριο. Αν υποθέσουμε ότι το εισιτήριο στοιχίζει 1,20 ευρώ και κάθε μήνα μετακινούνται 10.000 επιβάτες, να εκτιμήσετε διάστημα εμπιστοσύνης 95% για τη ζημιά του οργανισμού.

<u>Λύση</u>

Η εκτίμηση της αναλογίας από το δείγμα είναι 63/300=0,21. Εφαρμόζοντας τον τύπο, το διάστημα εμπιστοσύνης για την άγνωστη αναλογία του πληθυσμού όλων των επιβατών, υπολογίζεται:

$$\left[\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}\right] =$$
$$\left[0,21 \pm 1,96\sqrt{(0,21(1-0,21)/300}\right] =$$
$$\left[0,21 \pm 0.0472\right] = \left[0.1628, 0.2572\right]$$

Δηλαδή θα είναι από 16,28% έως 25,72%. Συνολικά, οι επιβάτες που μετακινούνται τον μήνα χωρίς εισιτήριο θα είναι 10.000*[0,1628,0,2572] = [1.628,2.572]. Το διάστημα εμπιστοσύνης για τη ζημιά του οργανισμού θα είναι 1,20* [1.628,2.572] =[1.953,6 έως 3.086,4] ευρώ

### 4.9.2 Έλεγχος οικονομικού ελεγκτή

Ένας ελεγκτής πραγματοποίησε έλεγχο για να διαπιστώσει αν καταχωρίζονται σωστά τα παραστατικά. Ελέγχθηκαν 96 παραστατικά από τα 866 και βρέθηκαν διαφορές με μέσο όρο 6,46 ευρώ και τυπική απόκλιση 17,58 ευρώ. Να βρεθεί το 95% διάστημα εμπιστοσύνης για το σύνολο των διαφορών στις καταχωρίσεις.

Λύση

Το 95% διάστημα εμπιστοσύνης για το σύνολο των διαφορών στις καταχωρίσεις είναι:

$$N\left[\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}\right] = 866\left[6.46 \pm 1.984 \frac{17.58}{\sqrt{96}} \sqrt{\frac{866-96}{866-1}}\right] =$$

$$\left[5.594.36 \pm 2.909.76\right] = \left[2.684,6 \, έως \, 8.504,12\right]$$

# 4.10 Ασκήσεις για κατανόηση - εφαρμογή

## 4.10.1 Άσκηση Α΄: Αγορές με κατάτμηση διαγωνισμού

Οι ελεγκτές μίας υπηρεσίας διαπίστωσαν ότι 1.550 συναλλαγές αφορούσαν αγορές με κατάτμηση διαγωνισμού. Επέλεξαν τυχαίο δείγμα 200 τέτοιων συναλλαγών και υπολόγισαν την αξία τους. Στο δείγμα η μέση τιμή ήταν 320,5 ευρώ και η τυπική απόκλιση 55,5 ευρώ. Να εκτιμήσετε διάστημα εμπιστοσύνης 95% για τη συνολική αξία των αγορών με κατάτμηση διαγωνισμού.

Λύση

Το 95% διάστημα εμπιστοσύνης για τη συνολική αξία αγορών με κατάτμηση διαγωνισμού είναι:

$$N\left[\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}\right] = 1.550\left[320,5 \pm 1.96 \frac{55,5}{\sqrt{200}} \sqrt{\frac{1.550-200}{1.550-1}}\right] =$$

$$\left[496.775 \pm 11.131.94\right] = \left[485.643,06 \text{ έως } 507.906,94\right]$$

## 4.10.2 Άσκηση Β΄: Αξία αποθεμάτων

Όταν μία τράπεζα απαίτησε από μια εταιρεία ανταλλακτικών αυτοκινήτων την πληρωμή δανείου της, η τελευταία κήρυξε πτώχευση. Ο εκκαθαριστής βρήκε 83.544 ανταλλακτικά και σε έλεγχο που έκανε σε τυχαίο δείγμα 400 ανταλλακτικών βρήκε μέση αξία 209,7 ευρώ με τυπική απόκλιση 67,4 ευρώ. Να εκτιμήσετε διάστημα εμπιστοσύνης 95% για τη συνολική αξία των αποθεμάτων ανταλλακτικών.

Λύση

Το 95% διάστημα εμπιστοσύνης για τη συνολική αξία αποθεμάτων ανταλλακτικών είναι:

$$N\left[\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}\right] = 83.544\left[209,7 \pm 1.96 \frac{67,4}{\sqrt{400}} \sqrt{\frac{83.544-400}{83.544-1}}\right] =$$

$$\left[17.519.176,8 \pm 550.505,5\right] = \left[16.968.671,3 \, έως \, 18.069.682,3\right]$$

## 4.10.3 Άσκηση Γ΄: Πρόγραμμα αποζημιώσεων

Οικονομικός έλεγχος ενός προγράμματος αποζημιώσεων έγινε σε 110.113 εγκεκριμένες αιτήσεις. Επελέγη τυχαίο δείγμα 539 αιτήσεων και βρέθηκαν 29 που δεν έπρεπε να είχαν εγκριθεί. Να εκτιμήσετε διάστημα εμπιστοσύνης 99% για το σύνολο των αποζημιώσεων που δόθηκαν ενώ δεν θα έπρεπε να είχαν εγκριθεί.

Λύση

Το 95% διάστημα εμπιστοσύνης για τη συνολική αξία αποζημιώσεων χωρίς έγκριση είναι:

$$N\left[p \pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}\right] = 110.113\left[\frac{29}{539} \pm 2,57\sqrt{\frac{0,054(1-0,054)}{539}}\sqrt{\frac{110.113-539}{110.113-1}}\right] =$$

$$[59.461,02 \pm 2.752,825] = [53.708,195 \text{ έως } 62.213,85]$$

### 4.10.4 Άσκηση Δ΄: Μέσες τιμές δειγμάτων

Δύο δείγματα προέρχονται από πληθυσμούς που ακολουθούν κανονική κατανομή, ελέγχεται η υπόθεση ότι οι μέσες τιμές των αντίστοιχων πληθυσμών δεν διαφέρουν. Δικαιολογήστε την απάντησή σας με στάθμη σημαντικότητας 95%. Μπορείτε να χρησιμοποιήσετε και τον παρακάτω πίνακα του SPSS για να δικαιολογήσετε την απάντησή σας.

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | 95% Confidence Interval of the Difference | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | Lower | Upper |
| Βασικός μισθός | Equal variances assumed | 21,484 | ,000 | -7,856 | 109 | ,000 | -15,180.079 | 1,932.252 | -19,009.740 | -11,350.419 |
| | Equal variances not assumed | | | -13,563 | 91,664 | ,000 | -15,180.079 | 1,119.186 | -17,402.989 | -12,957.170 |

Λύση

Αρχικά, ελέγχεται η ισότητα των διακυμάνσεων με το Levene's Test και προκύπτει sig level=0,000, οπότε απορρίπτεται η υπόθεση της ισότητας διακυμάνσεων. Κατόπιν, ελέγχεται η ισότητα των μέσων τιμών των δύο δειγμάτων χρησιμοποιώντας την τελευταία γραμμή του παραπάνω πίνακα (Equal variances not assumed). Η τιμή του κριτηρίου είναι -13,563 και προκύπτει sig level=0,000, οπότε απορρίπτεται η υπόθεση της ισότητας των μέσων τιμών των δύο δειγμάτων. Το διάστημα εμπιστοσύνης της διαφοράς μεταξύ μέσης τιμής πρώτου και δευτέρου δείγματος είναι [-17.402,98 έως -12.957,17], που σημαίνει ότι η μέση τιμή του πρώτου δείγματος είναι μικρότερη κατά 12,957,17 έως και 17.402,98 από τη μέση τιμή του δευτέρου δείγματος.

# Βιβλιογραφία/Αναφορές

Χάλκος, Γ. (2020). *Στατιστική Θεωρία και Πράξη*. Αθήνα: Δίσιγμα.

Tintle, N. L., Chance, B. L., Cobb, G. W., Cobb, R., Allan, J., Roy, S., Swanson, T. M., & VanderStoep J. L. (2021). *Εισαγωγή στις Στατιστικές Έρευνες*. Αθήνα: Gutenberg.

# Κριτήρια αξιολόγησης

## Κριτήριο αξιολόγησης 1

**Ποιος τύπος χρησιμοποιείται για τον προσδιορισμό του διαστήματος εμπιστοσύνης της μέσης τιμής πληθυσμού με στάθμη σημαντικότητας (1-α)%;**

**Απάντηση**
Για τον προσδιορισμό του διαστήματος εμπιστοσύνης της άγνωστης παραμέτρου μ της μέσης τιμής με στάθμη σημαντικότητας (1-α)%, χρησιμοποιείται ένα δείγμα μεγέθους n (>30), στο οποίο υπολογίζεται η μέση τιμή του και κατόπιν εφαρμόζεται ο τύπος:

$$[\bar{X} \text{ - za/2} \frac{\sigma}{\sqrt{n}}, \bar{X} \text{ + za/2} \frac{\sigma}{\sqrt{n}}]$$

Αν το σ (τυπική απόκλιση πληθυσμού) είναι άγνωστο, αντικαθίσταται με την τυπική απόκλιση s του δείγματος.

## Κριτήριο αξιολόγησης 2

**Ποιος τύπος χρησιμοποιείται για τον προσδιορισμό του διαστήματος εμπιστοσύνης της αναλογίας πληθυσμού με στάθμη σημαντικότητας (1-α)%;**

**Απάντηση**
Για τον προσδιορισμό του διαστήματος εμπιστοσύνης της άγνωστης αναλογίας π με στάθμη σημαντικότητας (1-α)%, χρησιμοποιείται ένα δείγμα μεγέθους n (>30), στο οποίο υπολογίζεται η αναλογία του και κατόπιν εφαρμόζεται ο τύπος.

$$[\text{p-z}_{a/2} \sqrt{(p(1-p))/n}, \quad \text{p+ z}_{a/2} \sqrt{(p(1-p))/n}]$$

# Παράρτημα

Πίνακας τιμών κατανομής t

| **Β.ε.** | **α=10%** | **α=5%** | **α=2,5%** | **α=1,0%** | **α=0,5%** |
|---|---|---|---|---|---|
| 1 | 3,078 | 6,314 | 12,706 | 31,821 | 63,657 |
| 2 | 1,886 | 2,920 | 4,303 | 6,965 | 9,925 |
| 3 | 1,638 | 2,353 | 3,182 | 4,541 | 5,841 |
| 4 | 1,533 | 2,132 | 2,776 | 3,747 | 4,604 |
| 5 | 1,476 | 2,015 | 2,571 | 3,365 | 4,032 |
| 6 | 1,440 | 1,943 | 2,447 | 3,143 | 3,707 |
| 7 | 1,415 | 1,895 | 2,365 | 2,998 | 3,499 |
| 8 | 1,397 | 1,860 | 2,306 | 2,896 | 3,355 |
| 9 | 1,383 | 1,833 | 2,262 | 2,821 | 3,250 |
| 10 | 1,372 | 1,812 | 2,228 | 2,764 | 3,169 |
| 11 | 1,363 | 1,796 | 2,201 | 2,718 | 3,106 |
| 12 | 1,356 | 1,782 | 2,179 | 2,681 | 3,055 |
| 13 | 1,350 | 1,771 | 2,160 | 2,650 | 3,012 |
| 14 | 1,345 | 1,761 | 2,145 | 2,624 | 2,977 |
| 15 | 1,341 | 1,753 | 2,131 | 2,602 | 2,947 |
| 16 | 1,337 | 1,746 | 2,120 | 2,583 | 2,921 |
| 17 | 1,333 | 1,740 | 2,110 | 2,567 | 2,898 |
| 18 | 1,330 | 1,734 | 2,101 | 2,552 | 2,878 |
| 19 | 1,328 | 1,729 | 2,093 | 2,539 | 2,861 |
| 20 | 1,325 | 1,725 | 2,086 | 2,528 | 2,845 |
| 21 | 1,323 | 1,721 | 2,080 | 2,518 | 2,831 |
| 22 | 1,321 | 1,717 | 2,074 | 2,508 | 2,819 |
| 23 | 1,319 | 1,714 | 2,069 | 2,500 | 2,807 |
| 24 | 1,318 | 1,711 | 2,064 | 2,492 | 2,797 |
| 25 | 1,316 | 1,708 | 2,060 | 2,485 | 2,787 |
| 26 | 1,315 | 1,706 | 2,056 | 2,479 | 2,779 |
| 27 | 1,314 | 1,703 | 2,052 | 2,473 | 2,771 |
| 28 | 1,313 | 1,701 | 2,048 | 2,467 | 2,763 |
| 29 | 1,311 | 1,699 | 2,045 | 2,462 | 2,756 |
| >=30 | 1,282 | 1,645 | 1,960 | 2,326 | 2,576 |
| za | 1,282 | 1,645 | 1,960 | 2,326 | 2,576 |

# Κεφάλαιο 5: Συσχέτιση

**Σύνοψη**
*Στο κεφάλαιο αυτό παρουσιάζονται τρόποι ελέγχου ύπαρξης συσχέτισης μεταξύ των στατιστικών μεταβλητών.*

**Προαπαιτούμενη γνώση**
*Έλεγχος υποθέσεων (Κεφάλαιο 4).*

## 5.1 Η έννοια της συσχέτισης

Ο άνθρωπος από τα πρώτα χρόνια της επιστημονικής ιστορίας του, προσπάθησε να εξηγήσει τα φαινόμενα που παρατηρούσε και να βρει πιθανές εξαρτήσεις τους από άλλα φαινόμενα. Άρχισε να μελετά αλλαγές σε κάποια φαινόμενα και τη συχνότητα των αλλαγών αυτών σε σχέση με μεταβολές σε άλλα φαινόμενα. Χαρακτηριστική περίπτωση αποτελεί ο σεληνιασμός, για τον οποίο οι άνθρωποι πίστευαν ότι η σελήνη ήταν αιτία για κάποιες νευρολογικές ασθένειες και ότι επηρέαζε τη ζωή τους.

Διάφορες στατιστικές μεταβλητές καταγράφονται παράλληλα, σε μια προσπάθεια να διερευνηθεί αν οι μεταβολές σε μία μεταβλητή, συνοδεύονται ή προκαλούνται από μεταβολές σε κάποια άλλη μεταβλητή. Έτσι, μελετήθηκε η επίδραση των καιρικών συνθηκών σε γεωργικές καλλιέργειες, η απόδοση σε σχέση με τα κίνητρα, η επίδραση φαρμάκων σε συμπτώματα ασθενειών κ.ά. Στατιστικές μεταβλητές, ποιοτικές ή ποσοτικές, που παρατηρήθηκε ότι μεταβάλλονται παράλληλα, ονομάστηκαν συσχετισμένες μεταβλητές και η μεταξύ τους σχέση ονομάστηκε **συσχέτιση**. Η συσχέτιση μεταξύ ποσοτικών μεταβλητών μπορεί να μετρηθεί αν μετρηθεί η ένταση της συν-μεταβολής τους, ενώ η συσχέτιση μεταξύ ποιοτικών μεταβλητών δεν μπορεί να μετρηθεί, παρά μόνο να διαπιστωθεί αν υπάρχει ή όχι.

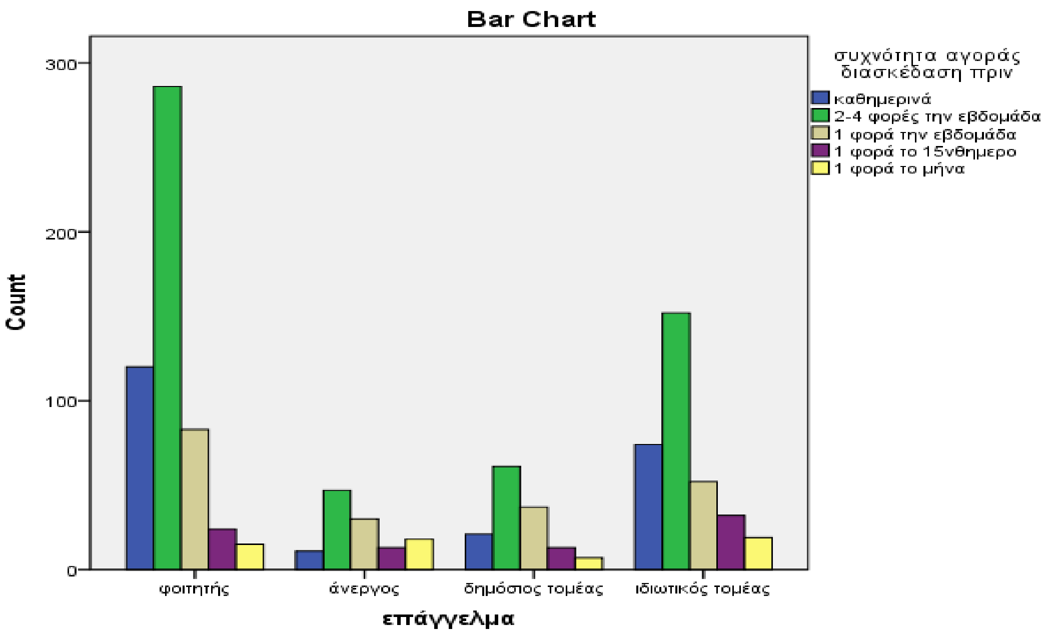
### 5.1.1 Οπτικοποίηση συσχέτισης δύο ποιοτικών μεταβλητών

Η ταυτόχρονη παρουσίαση δύο ποιοτικών μεταβλητών γίνεται με πίνακα διπλής εισόδου και πολλαπλό ραβδόγραμμα. Ο πίνακας διπλής εισόδου ή πίνακας συμπτώσεων, ή πίνακας συνάφειας, είναι ένας πίνακας με γραμμές που αντιστοιχούν στις κατηγορίες για την πρώτη μεταβλητή και με στήλες που αντιστοιχούν στις κατηγορίες της δεύτερης μεταβλητής. Στη διασταύρωση γραμμής και στήλης υπάρχει το πλήθος (συχνότητα) των παρατηρήσεων με αντίστοιχη κατηγορία για την πρώτη και δεύτερη μεταβλητή.

**Πίνακας 5.1** *Συχνότητα διασκέδασης και επάγγελμα.*

| | **Καθημερινά** | **2-4 /Εβδομάδα** | **1 /Εβδομάδα** | **1/ 15μερο** | **1/Μήνα** |
|---|---|---|---|---|---|
| Φοιτητής | 120 | 286 | 83 | 24 | 15 |
| Άνεργος | 12 | 47 | 30 | 13 | 18 |
| Δημ. Τομέας | 21 | 61 | 37 | 13 | 7 |
| Ιδιωτ. Τομέας | 74 | 152 | 52 | 32 | 19 |

Στον Πίνακα 5.1 υπάρχουν οι συχνότητες απαντήσεων της ιδιότητας ατόμου και της συχνότητας διασκέδασης. Τα ίδια δεδομένα απεικονίζονται στην Εικόνα 5.1. Παρατηρώντας τον πίνακα και το διάγραμμα, μπορούν να εξαχθούν κάποια συμπεράσματα για τη σχέση συχνότητας διασκέδασης και ιδιότητας ατόμου, τα οποία όμως είναι υποκειμενικά αν δεν βασίζονται σε κάποιους κανόνες διαπίστωσης της συσχέτισης. Ο αντικειμενικός έλεγχος ύπαρξης συσχέτισης γίνεται με τον **έλεγχο ανεξαρτησίας $X^2$**.




**Εικόνα 5.1** *Ραβδόγραμμα για συχνότητα διασκέδασης.*

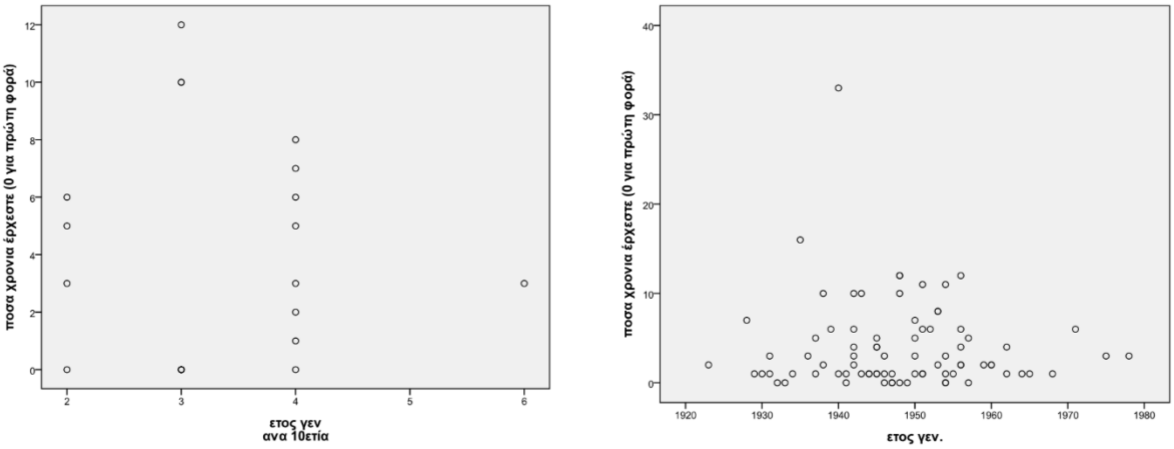
### 5.1.2 Οπτικοποίηση γραμμικής συσχέτισης δύο ποσοτικών μεταβλητών

Έχοντας καταγράψει μετρήσεις για δύο ποσοτικές μεταβλητές, μπορούμε να τοποθετήσουμε τις μετρήσεις αυτές σε δύο άξονες (οριζόντιος και κατακόρυφος) και να αντιστοιχίσουμε κάθε μέτρηση με μία τελεία στο επίπεδο των δύο αξόνων. Αυτό είναι το διάγραμμα διασποράς που περιγράφεται στο Κεφάλαιο 3. Στον Πίνακα 5.2 υπάρχουν οι μετρήσεις για τις μεταβλητές X (ηλικία) και Y (πλήθος επισκέψεων σε λασπόλουτρα), οι οποίες απεικονίζονται στο αριστερό διάγραμμα της Εικόνας 5.2.

**Πίνακας 5.2** *Καταγραφή μετρήσεων για δύο ποσοτικές μεταβλητές.*

| Ηλικία (σε δεκαετία) | Συχνότητα επίσκεψης στα λασπόλουτρα |
|---|---|
| 4 | 2 |
| 4 | 0 |
| 3 | 0 |
| 4 | 1 |
| 4 | 6 |
| 2 | 0 |
| 2 | 5 |
| 3 | 0 |
| 4 | 7 |
| 4 | 8 |
| 3 | 0 |
| 3 | 12 |
| 3 | 10 |
| 2 | 3 |
| 3 | 0 |

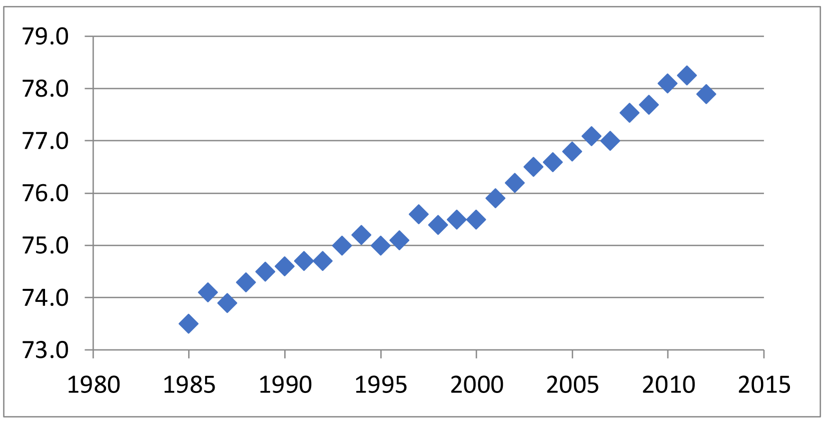
| | |
|---|---|
| 4 | 3 |
| 4 | 5 |
| 3 | 10 |
| 2 | 6 |
| 3 | 0 |
| 6 | 3 |



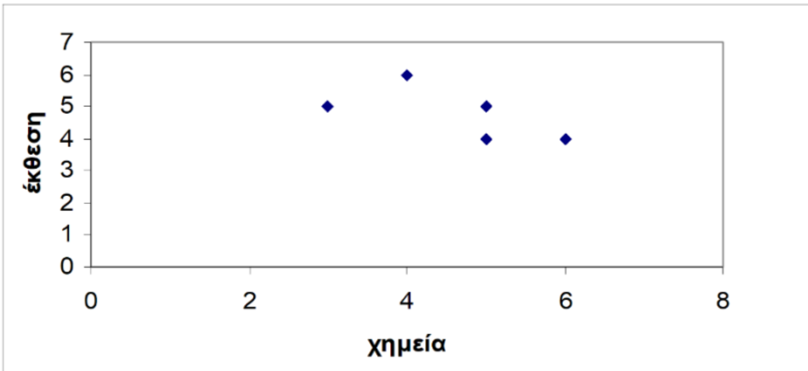

**Εικόνα 5.2** *Διαγράμματα διασποράς (αριστερά: με λίγες παρατηρήσεις, δεξιά: με περισσότερες παρατηρήσεις).*

Παρατηρώντας το αριστερό διάγραμμα διασποράς (Εικόνα 5.2), με τις 21 τελείες-σημεία, που αντιστοιχούν στις καταγραφές του Πίνακα 5.2, εμφανίζονται κάποια κατακόρυφα σχήματα στις οριζόντιες τιμές 2 και 4. Στο δεξί διάγραμμα της Εικόνας 5.2 απεικονίζονται τα έτη γέννησης και το πλήθος επισκέψεων στα λασπόλουτρα για 83 καταγραφές ατόμων. Στο σχήμα αυτό δεν μπορεί να παρατηρηθεί κάποια ιδιαίτερη «τάση» που σχηματίζεται από τις τελείες-σημεία. Στις περιπτώσεις αυτές, το συμπέρασμα στο οποίο καταλήγουμε είναι ότι δεν υπάρχει σχέση εξάρτησης μεταξύ ηλικίας και πλήθους επισκέψεων.

Στην Εικόνα 5.3 παρουσιάζονται καταγραφές δύο άλλων μεταβλητών, του έτους (οριζόντιος άξονας) και του προσδόκιμου χρόνου ζωής σε έτη (κατακόρυφος άξονας). Παρατηρείται οι τελείες-σημεία να έχουν μια ανοδική τάση (όσο μεγαλώνει η χρονολογία, τόσο αυξάνεται το προσδόκιμο ζωής). Στην περίπτωση αυτή, είναι πολύ πιθανό να υπάρχει μια σχέση «γραμμής» μεταξύ του χρόνου και του προσδόκιμου ζωής.



**Εικόνα 5.3** *Διάγραμμα για το προσδόκιμο ζωής ανά έτος.*




**Εικόνα 5.4** *Σημεία μαθητών με βαθμούς στην έκθεση και στη χημεία.*

Στην Εικόνα 5.4 παρουσιάζονται καταγραφές δύο άλλων μεταβλητών, του βαθμού στη χημεία (οριζόντιος άξονας) και του βαθμού στην έκθεση (κατακόρυφος άξονας) για κάποιους μαθητές. Παρατηρείται ότι οι τελείες-σημεία έχουν μια καθοδική τάση (όσο μεγαλώνει ο βαθμός της χημείας, τόσο μειώνεται ο βαθμός της έκθεσης). Στην περίπτωση αυτή, φαίνεται ότι υπάρχει μια σχέση φθίνουσας «γραμμής».

## 5.2 Έλεγχος συσχέτισης ποιοτικών μεταβλητών

Από τον πίνακα συμπτώσεων μπορεί να μελετηθεί η κατανομή της μίας ποιοτικής μεταβλητής σε σχέση με μία άλλη μεταβλητή, εξάγοντας υποκειμενικά συμπεράσματα. Για αντικειμενικά συμπεράσματα χρησιμοποιείται ο **έλεγχος ανεξαρτησίας $X^2$**.

Ο έλεγχος αυτός υποθέτει αρχικά ότι οι δύο μεταβλητές δεν έχουν κάποια σχέση. Αν ισχύει η υπόθεση αυτή, στη διασταύρωση γραμμής και στήλης, θα αντιστοιχεί μια θεωρητική συχνότητα που προκύπτει με βάση το πλήθος κατηγορίας γραμμής και κατηγορίας στήλης σε σχέση με το συνολικό πλήθος δεδομένων. Δηλαδή, θεωρείται τυχαία η κατανομή σε κατηγορία γραμμής και κατηγορία στήλης. Εξαρτάται μόνο από το συνολικό πλήθος εμφανίσεων κάθε κατηγορίας στα δεδομένα. Μια κατηγορία γραμμής που εμφανίστηκε σε μεγάλο πλήθος δεδομένων είναι φυσικό να εμφανίστηκε και σε αντίστοιχο μεγάλο πλήθος κατηγορίας στήλης, αν η κατανομή είναι τυχαία. Για να είναι αξιόπιστος ο έλεγχος, θα πρέπει το πλήθος δεδομένων να είναι αρκετά μεγάλο, ώστε τουλάχιστον 80% των θεωρητικών συχνοτήτων να είναι μεγαλύτερες από τον αριθμό 5. Στις περιπτώσεις που αυτό δεν ισχύει, πιθανώς να μπορούν να συνενωθούν δύο κατηγορίες γραμμών ή στηλών και κατόπιν να γίνει ο έλεγχος. Συγκρίνεται στη συνέχεια, η διαφορά των θεωρητικών από τις πραγματικές συχνότητες όλων των διασταυρώσεων (κελιών) του πίνακα, ώστε να ελεγχθεί (να γίνει δεκτή ή να απορριφθεί) η υπόθεση της ανεξαρτησίας των δύο μεταβλητών.

Αν η μία από τις δύο μεταβλητές είναι ποσοτική και η άλλη ποιοτική, μπορεί να γίνει έλεγχος συσχέτισης αν η ποσοτική μεταβλητή μετατραπεί σε ποιοτική με τον χωρισμό της σε μικρό πλήθος κλάσεων.

### 5.2.1 Βήματα ελέγχου ανεξαρτησίας δύο ποιοτικών μεταβλητών

Αναλυτικά τα βήματα ελέγχου της υπόθεσης ανεξαρτησίας δύο ποιοτικών μεταβλητών είναι:

Διατύπωση υποθέσεων

H0: Οι δύο μεταβλητές είναι ανεξάρτητες. Δεν έχουν σχέση μεταξύ τους.

H1: Οι δύο μεταβλητές δεν είναι ανεξάρτητες. Έχουν σχέση μεταξύ τους.

Υπολογισμός θεωρητικών συχνοτήτων και κριτηρίου ελέγχου της υπόθεσης H0

Υποτίθεται ότι ισχύει η H0 και υπολογίζονται οι θεωρητικές συχνότητες, με βάση το γεγονός ότι οι δύο μεταβλητές είναι ανεξάρτητες και επομένως τυχαία κατανεμημένες οι παρατηρήσεις στις διασταυρώσεις γραμμών και στηλών. Σε κάθε γραμμή και στήλη υπολογίζεται η θεωρητική συχνότητα με βάση το πλήθος γραμμής και στήλης σε σχέση με το συνολικό πλήθος παρατηρήσεων.

Ο τύπος για τη θεωρητική συχνότητα της i γραμμής και j στήλης είναι θij=Γινόμενο άθροισμα γραμμής x άθροισμα στήλης και διαίρεση με το συνολικό πλήθος.

Αφού υπολογιστούν οι θεωρητικές συχνότητες θij για κάθε διασταύρωση γραμμής και στήλης, συγκρίνονται με τις αντίστοιχες πραγματικές (παρατηρούμενες) συχνότητες νij. Η σύγκριση δεν γίνεται μόνο ως προς το μέγεθος της διαφοράς, αλλά ως προς το σχετικό μέγεθος εμφάνισης κάθε διασταύρωσης. Μια διαφορά 3 μονάδων σε πλήθος 10 τιμών δεν είναι ίδια με μια διαφορά 3 μονάδων σε πλήθος 75 τιμών. Η δεύτερη θεωρείται μικρότερη σχετική διαφορά (3/75 = 4%) από την πρώτη (3/10 = 30%). Προκειμένου στη σύγκριση να μην υπάρχει επηρεασμός ούτε από το πρόσημο της διαφοράς ούτε και από το πλήθος παρατηρήσεων κάθε κατηγορίας, υπολογίζεται η διαφορά στο τετράγωνο και διαιρείται με το πλήθος παρατηρήσεων. Το κριτήριο για τον υπολογισμό των διαφορών είναι:

$$X^2= \sum(\nu ij-\theta ij)^2 / \theta ij$$

Το κριτήριο αυτό συμβολίζεται $X^2$ επειδή αποδεικνύεται ότι ακολουθεί τη $X^2$ κατανομή με βαθμούς ελευθερίας (γινόμενο πλήθους γραμμών-1) × (πλήθος στηλών -1) (Χάλκος, 2020, σελ. 463)

Απόφαση, εξαγωγή συμπεράσματος για την ισχύ της H0

Η τιμή του $X^2$ που προκύπτει από τα δεδομένα συγκρίνεται με την αντίστοιχη τιμή της $X^2$ κατανομής (συνήθως με πιθανότητα σφάλματος a = 5%), ώστε να γίνει δεκτή ή όχι η υπόθεση H0. Το άθροισμα $X^2$ όλων των διαφορών είναι αμελητέο ή σημαντικό, ανάλογα με την τιμή που προκύπτει από πίνακα της κατανομής $X^2$, για a% περιθώριο σφάλματος. Αν η τιμή είναι μικρότερη από την τιμή της $X^2$ κατανομής, δίνεται δεκτή η H0. Αν η τιμή είναι μεγαλύτερη από την τιμή της $X^2$ κατανομής, μπορεί να απορριφθεί η H0, αφού δεν οδηγεί σε αποδεκτές τιμές $X^2$ κατανομής.

Το στατιστικό πρόγραμμα SPSS, κάνει μόνο του τον έλεγχο για τη σύγκριση με την τιμή της $X^2$ κατανομής και δίνει **την πιθανότητα να κάνουμε λάθος απορρίπτοντας την H0** γνωστή στον έλεγχο υποθέσεων ως sig-level ή p-value.

Στην πράξη, η απόφαση βασίζεται στην τιμή για το sig-level.

Αν το sig-level <0,05 **απορρίπτεται η H0**, κάνοντας ένα λάθος το πολύ 5%.

Αν το sig-level >0,05 **δεν μπορεί να απορριφθεί η H0**, κάνοντας ένα λάθος το πολύ 5%.

## 5.2.2 Παράδειγμα ελέγχου ανεξαρτησίας δύο ποιοτικών μεταβλητών

Ο Πίνακας διπλής εισόδου 5.1 που απεικονίζει το πλήθος ατόμων ανά επάγγελμα και συχνότητα διασκέδασης, θα χρησιμοποιηθεί για τον έλεγχο ανεξαρτησίας της συχνότητας διασκέδασης σε σχέση με το επάγγελμα. Η υπόθεση H0 είναι ότι επάγγελμα και συχνότητα διασκέδασης είναι ανεξάρτητα. Η H1 είναι ότι σχετίζονται μεταξύ τους. Στη συνέχεια υπολογίζονται τα αθροίσματα στηλών και γραμμών και οι θεωρητικές συχνότητες μέσα σε παρένθεση, οι οποίες παρουσιάζονται στον Πίνακα 5.3.

**Πίνακας 5.3** *Συχνότητα διασκέδασης και επάγγελμα.*

| | **Καθημερινά** | **2-4 ανά εβδομάδα** | **1 ανά εβδομάδα** | **1 ανά 15ήμερο** | **1 ανά μήνα** | **Σύνολο** |
|---|---|---|---|---|---|---|
| Φοιτητής | 120 (107,4) | 286 (258,3) | 83 (95,6) | 24 (38,8) | 15 (27,9) | 528 |
| Άνεργος | 12 (24,4) | 47 (58,7) | 30 (21,7) | 13 (8,8) | 18 (6,3) | 120 |
| Δημ. Τομέας | 21 (28,3) | 61 (68,0) | 37 (25,2) | 13 (10,2) | 7 (7,3) | 139 |
| Ιδιωτ. Τομέας | 74 (66,9) | 152 (161,0) | 52 (59,6) | 32 (24,2) | 19 (17,4) | 329 |
| Σύνολο | 227 | 546 | 202 | 82 | 59 | 1.116 |

Οι θεωρητικές συχνότητες υπολογίστηκαν ως γινόμενο συνόλου γραμμής επί σύνολο στήλης, διά το σύνολο 1.116 ατόμων. Στη διασταύρωση πρώτης γραμμής και πρώτης στήλης υπολογίστηκε η θεωρητική συχνότητα με τις πράξεις $528 \times 227 / 1116 = 107{,}4$. Παρόμοιοι υπολογισμοί έγιναν και για τις υπόλοιπες θεωρητικές συχνότητες που εμφανίζονται στον Πίνακα 5.3.

Στη συνέχεια υπολογίστηκε το άθροισμα όλων των διαφορών θεωρητικών από τις πραγματικές συχνότητες του Πίνακα 5.3, με τον τύπο:

$$X^2 = \sum(\nu ij-\theta ij)^2 / \theta ij = (120-107{,}4)^2/107{,}4+(286-258{,}3)^2/258{,}3+\ldots+$$

$$+\ldots+ (32-24{,}2)^2/24{,}2+(19-17{,}4)^2/17{,}4= 66{,}74$$

Η τιμή 66,74 δεν μπορεί να θεωρηθεί ούτε μικρή ούτε μεγάλη, αφού εξαρτάται από το πλήθος και το μέγεθος των διαφορών που αθροίστηκαν για τον υπολογισμό της. Η σύγκριση της τιμής 66,74 γίνεται με την τιμή του πίνακα κατανομής $X^2$ με στάθμη σημαντικότητας a% (συνήθως 5%) και βαθμούς ελευθερίας γινόμενο (πλήθους γραμμών -1) επί (πλήθους στηλών -1), δηλαδή $(4-1) \times (5-1)=12$.

Από τον στατιστικό πίνακα της $X^2$ κατανομής για 12 βαθμούς ελευθερίας και 5% επίπεδο σφάλματος προκύπτει η ανεκτή τιμή έως 21,026, η οποία είναι μικρότερη από την τιμή 66,74 που προέκυψε από τα δεδομένα. Επομένως, η υπόθεση $H0$ δεν μπορεί να γίνει δεκτή και συμπεραίνουμε ότι υπάρχει κάποια σχέση μεταξύ επαγγέλματος και συχνότητας διασκέδασης.

Τα αντίστοιχα αποτελέσματα με το στατιστικό πρόγραμμα SPSS εμφανίζονται στον Πίνακα 5.4. Στον πίνακα με πιθανότητα σφάλματος 5% εμφανίζεται το sig-level = $0{,}000001 < 0{,}05$. Το συμπέρασμα από την τιμή αυτή είναι ότι απορρίπτεται η υπόθεση της ανεξαρτησίας των δύο μεταβλητών. Επομένως, δεν μπορεί να γίνει δεκτός ο ισχυρισμός ότι επάγγελμα και συχνότητα διασκέδασης είναι ανεξάρτητα. Άρα, σχετίζονται μεταξύ τους.

**Πίνακας 5.4** *Αποτελέσματα ελέγχου ανεξαρτησίας στο SPSS.*

| Chi-Square Tests | Value | df | Asymp. Sig. (2-sided) |
|---|---|---|---|
| Pearson Chi-Square | 66,743 | 12 | 0,000 |
| Likelihood Ratio | 61,870 | 12 | 0,000 |
| Linear-by-Linear Association | 8,050 | 1 | 0,005 |
| N of Valid Cases | 1.116 | | |

### 5.2.3 Μέτρα συνάφειας δύο ποιοτικών μεταβλητών

Δυστυχώς, ο έλεγχος ανεξαρτησίας δεν δίνει περισσότερες πληροφορίες για τη σχέση των δύο μεταβλητών. Διερεύνηση της σχέσης μπορεί να γίνει μελετώντας περισσότερο τα ποσοστά της μίας μεταβλητής σε σχέση με την άλλη. Στο παράδειγμα του Πίνακα 5.3 μπορεί να γίνει διερεύνηση κάθε κατηγορίας επαγγέλματος σε σχέση με τη συχνότητα διασκέδασης.

Έχουν προταθεί διάφορα **μέτρα συνάφειας** για τον προσδιορισμό του βαθμού συνάφειας μεταξύ ποιοτικών μεταβλητών (Κώστογλου & Αντωνίου, 2021, σελ. 335). Τα μέτρα αυτά διαφοροποιούνται ανάλογα με το αν οι μεταβλητές είναι διατάξιμες ή όχι, και ανάλογα με το πλήθος των κατηγοριών τους.

Για τη μέτρηση της συνάφειας δύο δίτιμων, μη διατάξιμων ποιοτικών μεταβλητών κυριότερα μέτρα συνάφειας είναι το Φ και το Q του Jule. Υπολογίζονται από τις συχνότητες του $2\times2$ πίνακα συμπτώσεων των δύο μεταβλητών. Οι τιμές τους κυμαίνονται από -1 έως 1. Το -1 αντιστοιχεί σε αρνητική συνάφεια (αντίστροφη σχέση), το 1 σε θετική συνάφεια (ανάλογη σχέση), ενώ η τιμή κοντά στο 0 δηλώνει απουσία συνάφειας. (Κώστογλου & Αντωνίου, 2021, σελ. 336).

Η μέτρηση της συνάφειας για πίνακες συμπτώσεων κ×λ (όπου κ >2 ή λ >2), μη διατάξιμων μεταβλητών, γίνεται με το μέτρο Cramer ή τον συντελεστή συνάφειας Pearson ή το Φ, χρησιμοποιώντας την τιμή του $X^2$ που προκύπτει από τον έλεγχο ανεξαρτησίας και το πλήθος των κατηγοριών του πίνακα συμπτώσεων (Κώστογλου & Αντωνίου, 2021, σελ. 337).

## 5.3 Συσχέτιση ποσοτικών μεταβλητών

Στην περίπτωση ποσοτικών μεταβλητών, επειδή πρόκειται για αριθμητικές τιμές, είναι εύκολο να κάνουμε υπολογισμούς και να μετρήσουμε το πόσο επηρεάζεται η μία από την άλλη. Η μέτρηση της συσχέτισης δύο ποσοτικών μεταβλητών γίνεται κυρίως με τον υπολογισμό του γραμμικού συντελεστή συσχέτισης. Ο συντελεστής αυτός ονομάζεται και συντελεστής Pearson, συμβολίζεται με r ή ρ και υπολογίζεται με τον τύπο:

$$\rho = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sqrt{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}\sqrt{n\sum_{i=1}^{n} \mathrm{Y}_i^2 - \left(\sum_{i=1}^{n} \mathrm{Y}_i\right)^2}}$$

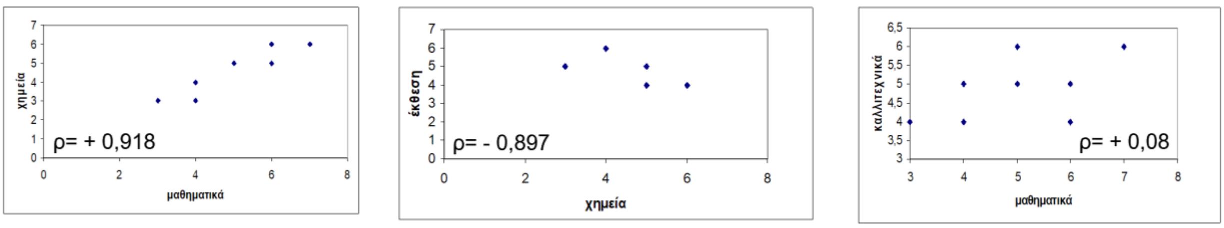
Με Xi, Yi, συμβολίζονται οι αντίστοιχες τιμές των δύο ποσοτικών μεταβλητών Χ και Υ. Το n είναι ίσο με το πλήθος των παρατηρήσεων. Οι τιμές του ρ βρίσκονται στο διάστημα από -1 έως +1. Το πρόσημό του δηλώνει το είδος της συσχέτισης, ενώ η τιμή του την ένταση της γραμμικής συσχέτισης των μεταβλητών Χ και Υ.

**Πίνακας 5.5** *Ένταση συσχέτιση ανάλογα με την τιμή του ρ.*

| Τιμές του ρ | Συμπέρασμα |
|---|---|
| ±(0 έως 0,29) | απουσία γραμμικής συσχέτισης |
| ±(0,30 έως 0,49) | χαμηλή γραμμική συσχέτιση |
| ±(0,50 έως 0,69) | μέτρια γραμμική συσχέτιση |
| ±(0,70 έως 0,79) | έντονη γραμμική συσχέτιση |
| ±(0,80 έως 1) | πολύ έντονη γραμμική συσχέτιση |

### 5.3.1 Ερμηνεία των τιμών του ρ

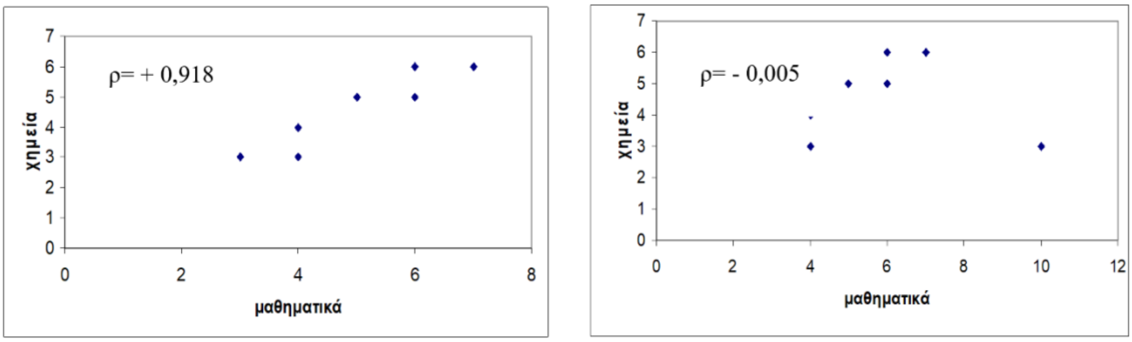
Όταν η τιμή του συντελεστή γραμμικής συσχέτισης είναι κοντά στο μηδέν, αυτό δηλώνει την απουσία γραμμικής συσχέτισης. Θετικές τιμές του συντελεστή γραμμικής συσχέτισης εκφράζουν θετική κατεύθυνση στη συσχέτιση, δηλαδή όταν η μία μεταβλητή αυξάνεται και η άλλη μεταβλητή αυξάνεται επίσης. Αρνητικές τιμές του συντελεστή γραμμικής συσχέτισης εκφράζουν αρνητική κατεύθυνση στη συσχέτιση· δηλαδή όταν η μία μεταβλητή αυξάνεται η άλλη μεταβλητή μειώνεται. Στον Πίνακα 5.5 παρουσιάζονται τα συμπεράσματα ανάλογα με την τιμή του ρ. Στην εικόνα εμφανίζονται κάποια διαγράμματα διασποράς δύο μεταβλητών και η αντίστοιχη τιμή του συντελεστή γραμμικής συσχέτισης.




**Εικόνα 5.5** *.Παραδείγματα σχέσεων δύο μεταβλητών και συντελεστές γραμμικής συσχέτισης.*

## 5.3.2 Περιορισμοί του συντελεστή γραμμικής συσχέτισης

Μία ακραία τιμή είναι πιθανό να επηρεάσει την τιμή του συντελεστή γραμμικής συσχέτισης και να οδηγήσει σε λανθασμένα συμπεράσματα. Στην Εικόνα 5.6 παρατηρούμε ότι η κάτω δεξιά ακραία τιμή, επηρεάζει πολύ τον συντελεστή γραμμικής συσχέτισης.




**Εικόνα 5.6** *Επίδραση ακραίας τιμής στον συντελεστή γραμμικής συσχέτισης (αριστερά: χωρίς την ακραία τιμή, δεξιά: με την ακραία τιμή).*

Ο συντελεστής γραμμικής συσχέτισης μετράει μόνο σχέσεις δύο μεταβλητών, οι οποίες στο διάγραμμα διασποράς εκφράζονται με ευθεία γραμμή. Σχέσεις οι οποίες στο διάγραμμα διασποράς πιθανώς να εκφράζονται με καμπύλη ή άλλο σχήμα έχουν συντελεστή γραμμικής συσχέτισης σχεδόν 0. Αυτό όμως δεν σημαίνει ότι οι δύο μεταβλητές δεν έχουν κάποια συσχέτιση. Σημαίνει μόνο ότι δεν έχουν γραμμική συσχέτιση.

## 5.3.3 Παράδειγμα υπολογισμού συντελεστή γραμμικής συσχέτισης

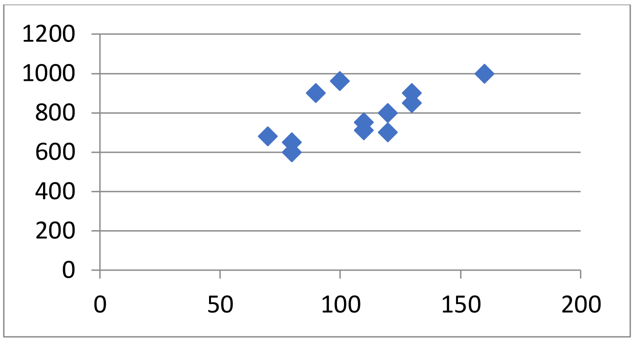
Στον Πίνακα 5.6 καταγράφεται ένα παράδειγμα μηνιαίων δαπανών μιας επιχείρησης και αντίστοιχων εισπράξεων. Πιθανολογείται ότι υπάρχει γραμμική σχέση μεταξύ των δαπανών και των εισπράξεων και παρουσιάζονται στο διάγραμμα της Εικόνας 5.7 τα σημεία που αντιστοιχούν στις τιμές του Πίνακα 5.6.

**Πίνακας 5.6** *Παράδειγμα δεδομένων για δαπάνες και αντίστοιχες εισπράξεις.*

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Δαπάνες Χ** | 80 | 120 | 70 | 110 | 130 | 100 | 90 | 160 | 120 | 80 | 110 | 130 |
| **Εισπράξεις Υ** | 600 | 700 | 680 | 710 | 900 | 960 | 900 | 1.000 | 800 | 650 | 750 | 850 |

Υπολογίζονται τα αθροίσματα ΣXi = 1.300, ΣYi = 9.500, ΣXi2 = 148.200, ΣYi2 = 7.705.600, ΣXiYi = 1.052.700 και κατόπιν, κάνοντας αντικατάσταση στον τύπο υπολογισμού του ρ για πλήθος n = 12, υπολογίζεται η τιμή:

$$\rho = \frac{12\times 1.300\times 9.500}{\sqrt{12\times 148.200-(1300)^2}\,\sqrt{12\times 77.056.000-9.500^2}} = 0,64$$




**Εικόνα 5.7** *Παράδειγμα δαπανών (X) και εισπράξεων (Y).*

## 5.4 Έλεγχος ύπαρξης συσχέτισης στον πληθυσμό

Έχοντας ένα πλήθος n σημείων δεδομένων για δύο ποσοτικές μεταβλητές X και Y, υπολογίζεται ο δειγματικός συντελεστής συσχέτισης r (συντελεστής συσχέτισης στις τιμές του δείγματος) και διαπιστώνεται η ύπαρξη ή όχι της μεταξύ τους σχέσης. Συνήθως, όμως, ο στόχος είναι η διαπίστωση ή όχι συσχέτισης για το σύνολο του πληθυσμού και όχι μόνο για το δείγμα.

Ο έλεγχος ύπαρξης συσχέτισης στον πληθυσμό γίνεται με την αρχική υπόθεση ότι ο συντελεστής συσχέτισης ρ μεταξύ των δύο μεταβλητών στον πληθυσμό είναι μηδενικός. Υποτίθεται δηλαδή η απουσία συσχέτισης.

$$H0: \ \rho = 0$$

$$H1: \ \rho \neq 0$$

Με βάση την υπόθεση αυτή υπολογίζεται το κατάλληλο κριτήριο, και ανάλογα με την τιμή του κριτηρίου λαμβάνεται η απόφαση αν ισχύει ή όχι η αρχική υπόθεση H0. Υποθέτοντας ότι οι δύο μελετώμενες μεταβλητές είναι κανονικά κατανεμημένες, η κατανομή του δειγματικού συντελεστή συσχέτισης r είναι η t-κατανομή με βαθμούς ελευθερίας n-2 και το κριτήριο για τον έλεγχο της H0 είναι:

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

το οποίο συγκρίνεται με την κατανομή t και το αντίστοιχο επίπεδο εμπιστοσύνης α% (Γναρδέλλης, 2019).

### 5.4.1 Παράδειγμα ελέγχου συντελεστή γραμμικής συσχέτισης

Στο παράδειγμα των δεδομένων του Πίνακα 5.6 μηνιαίων δαπανών μιας επιχείρησης και αντίστοιχων εισπράξεων, υπολογίσθηκε ο συντελεστής συσχέτισης r = 0,64.

Για να ελεγχθεί η ύπαρξη γραμμικής σχέση ρ μεταξύ των δαπανών και των εισπράξεων στην επιχείρηση γενικότερα, γίνονται οι εξής υποθέσεις:

$$H0: \ \rho = 0$$

$$H1: \ \rho \neq 0$$

και υπολογίζεται το ακόλουθο κριτήριο:

$$t = r\sqrt{\frac{n-2}{1-r^2}} = 0{,}64\sqrt{\frac{12-2}{1-0{,}41}} = 2{,}6333$$

Στον πίνακα της t κατανομής για βαθμούς ελευθερίας 10 και στάθμη σημαντικότητας 5% εντοπίζεται η τιμή ελέγχου 2,228. Η τιμή του κριτηρίου 2,63 είναι μεγαλύτερη από την τιμή του πίνακα t-κατανομής, οπότε η H0 μπορεί να απορριφθεί. Ως συμπέρασμα προκύπτει ότι υπάρχει συσχέτιση στον πληθυσμό μεταξύ δαπανών και εισπράξεων.

## 5.5 Παρερμηνείες – Επηρεασμός συσχέτισης

Η συσχέτιση μετράει την ύπαρξη σχέσης μεταξύ των δύο μεταβλητών, αλλά στην περίπτωση ύπαρξης συσχέτισης δεν μπορεί να προσδιοριστεί η μία μεταβλητή ως αιτία και η άλλη ως αποτέλεσμα. Κάποιες φορές, το αποτέλεσμα της συσχέτισης εμφανίζεται εξαιτίας της ύπαρξης μιας τρίτης μεταβλητής.

Μελετώντας τη σχέση μεταξύ κατανάλωσης ειδών ένδυσης και φύλου βρίσκουμε υψηλή συσχέτιση με το ποσοστό γυναικών να καταναλώνουν περισσότερο από ό,τι οι άνδρες. Αυτό πιθανώς να οφείλεται και στη μεταβλητή «εισόδημα» που δεν συμπεριλάβαμε στην ανάλυση, αλλά σχετίζεται με την κατανάλωση και το φύλο.

Άλλο παράδειγμα μπορεί να είναι η συσχέτιση μεταξύ συχνότητας εξόδου για διασκέδαση και κατηγορίας επαγγέλματος. Η ύπαρξη συσχέτισης μπορεί να οδηγήσει στο λανθασμένο συμπέρασμα ότι επιλέγοντας κάποια κατηγορία επαγγέλματος θα διασκεδάζει συχνότερα κάποιος. Η συσχέτιση πιθανώς να οφείλεται στον χαρακτήρα του ατόμου ή στις κλίσεις του, που προκαλούν τόσο την επιλογή επαγγέλματος όσο και τη συχνότητα εξόδου διασκέδασης.

Κάποιες άλλες φορές, μπορεί να υπάρχει συσχέτιση μεταξύ δύο μεταβλητών αλλά πιθανώς μια τρίτη μεταβλητή επηρεάζει και τις δύο μεταβλητές και εξουδετερώνει την ύπαρξη εμφάνισης της συσχέτισής τους. Παράδειγμα σε ένα μεγάλο δείγμα ατόμων, δεν εμφανίζεται συσχέτιση μεταξύ εισοδήματος και επιπέδου εκπαίδευσης. Χωρίζοντας το δείγμα αυτό σε δύο μικρότερα δείγματα, ένα με μικρές ηλικίες και ένα με μεγάλες ηλικίες, εμφανίζεται συσχέτιση μεταξύ εισοδήματος και επιπέδου εκπαίδευσης. Αυτό συμβαίνει γιατί η μεταβλητή ηλικία δρα αρνητικά με το επίπεδο εκπαίδευσης στο συνολικό δείγμα (μεγαλύτεροι σε ηλικία έχουν χαμηλότερο επίπεδο εκπαίδευσης), ενώ δρα θετικά με το εισόδημα (μεγαλύτεροι σε ηλικία έχουν μεγαλύτερο εισόδημα). Αν δεν ληφθεί, όμως, υπόψη η ηλικία, στο συνολικό δείγμα δεν εμφανίζεται συσχέτιση μεταξύ εισοδήματος και εκπαίδευσης, αφού σε αυτό περιλαμβάνονται τόσο νέοι με χαμηλό εισόδημα και υψηλό επίπεδο εκπαίδευσης, όσο και μεγαλύτεροι με υψηλό εισόδημα και χαμηλότερο επίπεδο εκπαίδευσης.

Επίδραση, επίσης, στον υπολογισμό της συσχέτισης έχει το μέγεθος του δείγματος, το οποίο θα πρέπει να είναι κατάλληλο για να μετρηθεί σωστά η συσχέτιση αλλά και αντιπροσωπευτικό, ώστε να περιλαμβάνει ικανό αριθμό από κάθε κατηγορία ή διάστημα τιμών των μεταβλητών.

Στην περίπτωση ποσοτικών ή διατάξιμων μεταβλητών που χωρίζονται σε κατηγορίες και κατόπιν ελέγχεται η συσχέτισή τους με μία άλλη μεταβλητή, επίδραση στον υπολογισμό της τιμής του $X^2$ έχει και το πλήθος κατηγοριών που επιλέγονται, καθώς και το εύρος των διαστημάτων κάθε κατηγορίας. Θα πρέπει να ελέγχεται η ύπαρξη συνάφειας και με διαφορετική κατηγοριοποίηση, για να επιβεβαιώνονται τα αποτελέσματα.

## 5.6 Συσχέτιση διατάξιμων μεταβλητών

Η συσχέτιση ποσοτικών μεταβλητών ελέγχεται με τον συντελεστή συσχέτισης ρ, που ονομάζεται συντελεστής Pearson. Αρκετές φορές, όμως, οι παρατηρούμενες μεταβλητές δεν είναι ποσοτικές, αλλά μεταβλητές που εκφράζουν βαθμό συμφωνίας με μια άποψη (καθόλου, λίγο, μέτρια, πολύ, πάρα πολύ) ή συχνότητα κάποιας ενέργειας (ποτέ, λίγο, μέτρια, συχνά, πολύ συχνά). Στις περιπτώσεις αυτές, οι τιμές των μεταβλητών κωδικοποιούνται με τους αριθμούς 1, 2, 3, 4, 5, αλλά οι αριθμοί αυτοί εκφράζουν διάταξη και όχι μέτρηση.

Μελετώντας τέτοιου είδους μεταβλητές, δεν έχει νόημα ο μέσος όρος των τιμών, αλλά οι θέσεις (τάξεις) των τιμών και η σχετική τους διάταξη. Οι μεταβλητές αυτές ονομάζονται διατάξιμες και δεν είναι σωστό να αντιμετωπίζονται ως ποσοτικές μεταβλητές.

Για τον υπολογισμό της συσχέτισης μεταξύ ανεξάρτητων διατάξιμων μεταβλητών, προτάθηκε ο συντελεστής Spearman, ο οποίος υπολογίζει τη συσχέτιση όχι των τιμών, αλλά των τάξεων των δεδομένων.

Παίρνει τιμές στο διάστημα -1 έως +1. Αρνητικές τιμές εκφράζουν αρνητική συσχέτιση, ενώ θετικές τιμές εκφράζουν θετική συσχέτιση. Τιμές κοντά στο 0 εκφράζουν απουσία συσχέτισης μεταξύ των μεταβλητών.

Επίσης, για τον υπολογισμό της συσχέτισης μεταξύ εξαρτημένων διατάξιμων μεταβλητών (ζευγαρωτές μεταβλητές), προτάθηκε ο συντελεστής Kendal, ο οποίος υπολογίζει τη συσχέτιση όχι των τιμών, αλλά των διαφορών τάξεων των δεδομένων. Παίρνει τιμές στο διάστημα -1 έως +1. Αρνητικές τιμές εκφράζουν αρνητική συσχέτιση, ενώ θετικές τιμές εκφράζουν θετική συσχέτιση. Τιμές κοντά στο 0 εκφράζουν απουσία συσχέτισης μεταξύ των μεταβλητών.

## 5.7 Εφαρμογές για κατανόηση

### 5.7.1 Παράδειγμα ελέγχου ανεξαρτησίας

Για τα δεδομένα του Πίνακα 5.7, να εξεταστεί αν υπάρχει σχέση μεταξύ του εισοδήματος των ατόμων και της γνώμης τους. Το εισόδημα είναι ποσοτική μεταβλητή αλλά με τον χωρισμό του σε 5 κλάσεις τιμών του μετατράπηκε σε ποιοτική μεταβλητή. Οι κλάσεις εισοδήματος αντιστοιχούν στις στήλες του Πίνακα 5.7.

**Πίνακας 5.7** *Παράδειγμα δεδομένων για εισόδημα και γνώμη πριν από την κρίση.*

| | | **Εισόδημα πριν από την κρίση** | | | | | |
|---|---|---|---|---|---|---|---|
| | | **<300** | **300-600** | **600-900** | **900-1200** | **>1200** | **Σύνολο** |
| Γνώμη πριν από την κρίση | πολύ καλή | 14 | 40 | 21 | 15 | 42 | 132 |
| | καλή | 98 | 172 | 96 | 68 | 47 | 481 |
| | μέτρια | 98 | 152 | 77 | 50 | 17 | 394 |
| | δύσκολη | 38 | 32 | 10 | 11 | 3 | 94 |
| | πολύ δύσκολη | 9 | 3 | 3 | 0 | 2 | 17 |
| σύνολο | | 257 | 399 | 207 | 144 | 111 | 1.118 |

**Πίνακας 5.8** *Αποτελέσματα ελέγχου $X^2$ για εισόδημα και γνώμη πριν από την κρίση.*

| **Chi-Square Tests** | **Value** | **df** | **Asymp. Sig. (2-sided)** |
|---|---|---|---|
| Pearson Chi-Square | 121,641a | 16 | ,000 |
| Likelihood Ratio | 104,237 | 16 | ,000 |
| Linear-by-Linear Association | 61,593 | 1 | ,000 |
| N of Valid Cases | 1.118 | | |

a. 4 cells (16,0%) have expected count less than 5. The minimum expected count is 1,69.

Ο έλεγχος ανεξαρτησίας γίνεται με τον έλεγχο $X^2$.

Η αρχική υπόθεση H0 είναι ότι οι μεταβλητές «γνώμη» και «κατηγορία εισοδήματος» είναι ανεξάρτητες. Η εναλλακτική υπόθεση H1 είναι ότι οι μεταβλητές «γνώμη» και «κατηγορία εισοδήματος» συσχετίζονται. Το κριτήριο για τον έλεγχο αυτό, είναι η τιμή Pearson Chi-Square = 121,641 που εμφανίζεται στον Πίνακα 5.8 αποτελεσμάτων του ελέγχου ανεξαρτησίας που έγινε με το SPSS.

Επειδή το sig = 0,000001 < 0,05, η υπόθεση H0 (ανεξαρτησία) μπορεί να απορριφθεί σε επίπεδο εμπιστοσύνης 95% ή με σφάλμα τύπου α = 5%. Οπότε, συμπεραίνεται ότι οι δύο μεταβλητές «γνώμη» και «κατηγορία εισοδήματος» συσχετίζονται, δηλαδή αλληλοεπηρεάζονται.

## 5.8 Ασκήσεις για κατανόηση - εφαρμογή

### 5.8.1 Έλεγχος ανεξαρτησίας

Στον παρακάτω πίνακα παρουσιάζονται τα αποτελέσματα ελέγχου υποθέσεων μέσω του SPSS.

α) Ποιος έλεγχος εφαρμόστηκε; Σε τι είδους μεταβλητές; Ποια αρχική υπόθεση ελέγχθηκε;

β) Τι συμπέρασμα προκύπτει από τον παρακάτω πίνακα σχετικά με την αρχική υπόθεση;

| **Chi-Square Tests** | **Value** | **df** | **Asymp. Sig. (2-sided)** |
|---|---|---|---|
| Pearson Chi-Square | 76,85 | 12 | 0,108 |
| Likelihood Ratio | 67,70 | 12 | 0,103 |
| Linear-by-Linear Association | 9,10 | 1 | 0,106 |
| N of Valid Cases | 754 | | |

Λύση

α) Ο έλεγχος που εφαρμόστηκε είναι ο έλεγχος ανεξαρτησίας $X^2$ (Chi-Square Tests**).** Ο έλεγχος αυτός εφαρμόζεται σε ποιοτικές μεταβλητές με περισσότερες από 2 κατηγορίες τιμών, για να ελεγχθεί αν είναι ή όχι ανεξάρτητες. Οι υποθέσεις είναι:

H0: υπάρχει ανεξαρτησία

H1: δεν ισχύει η H0

β) Η τιμή του κριτηρίου $X^2$ είναι 76,85 με πιθανότητα sig = 0,108, όπως φαίνεται στην πρώτη γραμμή του παραπάνω πίνακα. Η πιθανότητα sig = 0,108 εκφράζει την πιθανότητα λάθους, αν η H0 απορριφθεί. Σε επίπεδο σημαντικότητας 5%, δεν μπορεί να απορριφθεί η H0, αφού τότε η πιθανότητα λάθους θα είναι 10,8% μεγαλύτερη από το 5%. Επομένως, θα πρέπει να γίνει δεκτό το ότι υπάρχει ανεξαρτησία μεταξύ των δύο ποιοτικών μεταβλητών.

### 5.8.2 Έλεγχος γραμμικής συσχέτισης

Οι αγορές τροφίμων X σε ένα εστιατόριο σχετίζονται με τις εισπράξεις Y όπως φαίνεται στον πίνακα.

| **Αγορές X (σε ευρώ)** | 180 | 220 | 170 | 210 | 240 | 270 | 250 | 260 |
|---|---|---|---|---|---|---|---|---|
| **Εισπράξεις Y (σε ευρώ)** | 600 | 700 | 680 | 710 | 900 | 980 | 930 | 1.000 |

α) Προκειμένου να ερευνηθεί αν υπάρχει γραμμική συσχέτιση μεταξύ των αγορών και των εισπράξεων, να υπολογιστεί ο συντελεστής γραμμικής συσχέτισης στο δείγμα.

β) Να εξεταστεί για τον πληθυσμό η υπόθεση H0: δεν υπάρχει σχέση μεταξύ αγορών και εισπράξεων, με α = 5%.

Λύση

α) Ο υπολογισμός του συντελεστή συσχέτισης στο δείγμα των 8 παρατηρήσεων γίνεται με τον τύπο:

$$\rho = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sqrt{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}\sqrt{n\sum_{i=1}^{n} Y_i^2 - \left(\sum_{i=1}^{n} Y_i\right)^2}} = \frac{8\times 1.499.800 - 1.800\times 6.500}{\sqrt{8\times 414.400 - 1.800^2}\sqrt{8\times 5.451.800 - 6.500^2}} = 0,9315$$

β) Για να ελεγχθεί η ύπαρξη γραμμικής σχέση ρ μεταξύ των αγορών και των εισπράξεων στο εστιατόριο γενικότερα (πληθυσμός), γράφονται οι υποθέσεις ως εξής:

$$H0: \ \rho = 0$$

$$H1: \ \rho \neq 0$$

και υπολογίζεται το κριτήριο ελέγχου:

$$t = r\sqrt{\frac{n-2}{1-r^2}} = 0{,}9315\sqrt{\frac{8-2}{1-0{,}868}} = 6{,}276$$

Στον πίνακα της t κατανομής για βαθμούς ελευθερίας 6 (8-2) και στάθμη σημαντικότητας 5%, εντοπίζεται η τιμή ελέγχου 2,447. Η τιμή του κριτηρίου 6,276 είναι μεγαλύτερη από την τιμή ελέγχου του πίνακα t-κατανομής, οπότε η H0 μπορεί να απορριφθεί. Ως συμπέρασμα προκύπτει ότι υπάρχει συσχέτιση στον πληθυσμό μεταξύ αγορών και εισπράξεων.

## 5.8.3 Έλεγχος γραμμικής συσχέτισης πολλών διατάξιμων μεταβλητών

Σε ένα δείγμα απαντήσεων ερωτηματολογίου για τις συνήθειες διατροφής, μελετήθηκαν πολλές μεταβλητές σε μια προσπάθεια να διαπιστωθεί ποιες σχετίζονται μεταξύ τους. Οι μεταβλητές παίρνουν διατεταγμένες τιμές στην κλίμακα 1 έως 5. Υπολογίστηκαν οι συντελεστές συσχέτισης Spearman και Kendal. Τα αποτελέσματα του SPSS εμφανίζονται στην επόμενη εικόνα. Σε κάθε κελί του πίνακα αποτελεσμάτων (διασταύρωση γραμμής και στήλης), εμφανίζεται η τιμή του συντελεστή συσχέτισης, το sig. level για τον έλεγχο ανεξαρτησίας (στον πληθυσμό) με α = 1% και το πλήθος των δεδομένων που χρησιμοποιήθηκαν. Στην περίπτωση που η αντίστοιχη υπόθεση «H0: Οι μεταβλητές είναι ανεξάρτητες ή ο μεταξύ τους συντελεστής συσχέτισης=0» δεν μπορεί να απορριφθεί, υπάρχουν δύο ** πάνω από την τιμή του συντελεστή συσχέτισης. Ποια είναι τα συμπεράσματα σας;

| Spearman | | Προσέχω να είμαι υγιής | Γυμνάζομαι | Καπνίζω | Παραλείπω το πρωινό γεύμα | Παραλείπω το βραδινό γεύμα | Λαμβάνω βιταμίνες | Λαμβάνω συμπληρώματα διατροφής |
|---|---|---|---|---|---|---|---|---|
| Προσέχω να είμαι υγιής | Correlation Coefficient | 1,000 | ,458** | -,145 | ,065 | ,214 | ,248 | ,205 |
| | Sig. (2-tailed) | . | ,004 | ,391 | ,701 | ,204 | ,140 | ,224 |
| | N | 37 | 37 | 37 | 37 | 37 | 37 | 37 |
| Γυμνάζομαι | Correlation Coefficient | ,458** | 1,000 | ,109 | ,040 | ,289 | ,138 | ,180 |
| | Sig. (2-tailed) | ,004 | . | ,520 | ,813 | ,083 | ,415 | ,287 |
| | N | 37 | 37 | 37 | 37 | 37 | 37 | 37 |
| Καπνίζω | Correlation Coefficient | -,145 | ,109 | 1,000 | -,152 | ,107 | ,294 | ,165 |
| | Sig. (2-tailed) | ,391 | ,520 | . | ,369 | ,529 | ,077 | ,328 |
| | N | 37 | 37 | 37 | 37 | 37 | 37 | 37 |
| Παραλείπω το πρωινό γεύμα | Correlation Coefficient | ,065 | ,040 | -,152 | 1,000 | ,235 | ,215 | ,161 |
| | Sig. (2-tailed) | ,701 | ,813 | ,369 | . | ,161 | ,201 | ,341 |
| | N | 37 | 37 | 37 | 37 | 37 | 37 | 37 |
| Παραλείπω το βραδινό γεύμα | Correlation Coefficient | ,214 | ,289 | ,107 | ,235 | 1,000 | ,278 | ,257 |
| | Sig. (2-tailed) | ,204 | ,083 | ,529 | ,161 | . | ,095 | ,124 |
| | N | 37 | 37 | 37 | 37 | 37 | 37 | 37 |
| Λαμβάνω βιταμίνες | Correlation Coefficient | ,248 | ,138 | ,294 | ,215 | ,278 | 1,000 | ,446** |
| | Sig. (2-tailed) | ,140 | ,415 | ,077 | ,201 | ,095 | . | ,006 |
| | N | 37 | 37 | 37 | 37 | 37 | 37 | 37 |
| Λαμβάνω συμπληρώματα διατροφής | Correlation Coefficient | ,205 | ,180 | ,165 | ,161 | ,257 | ,446** | 1,000 |
| | Sig. (2-tailed) | ,224 | ,287 | ,328 | ,341 | ,124 | ,006 | . |
| | N | 37 | 37 | 37 | 37 | 37 | 37 | 37 |

| Kendall's tau_b | | Προσέχω να είμαι υγιής | Γυμνάζο-μαι | Καπνίζω | Παραλείπω το πρωινό γεύμα | Παραλείπω το βραδινό γεύμα | Λαμβάνω βιταμίνες | Λαμβάνω συμπληρώματα διατροφής |
|---|---|---|---|---|---|---|---|---|
| Προσέχω να είμαι υγιής | Correlation Coefficient | 1,000 | ,390** | -,133 | ,058 | ,179 | ,211 | ,187 |
| | Sig. (2-tailed) | . | ,006 | ,375 | ,685 | ,203 | ,146 | ,215 |
| | N | 37 | 37 | 37 | 37 | 37 | 37 | 37 |
| Γυμνάζομαι | Correlation Coefficient | ,390** | 1,000 | ,099 | ,034 | ,245 | ,122 | ,160 |
| | Sig. (2-tailed) | ,006 | . | ,507 | ,812 | ,080 | ,399 | ,285 |
| | N | 37 | 37 | 37 | 37 | 37 | 37 | 37 |
| Καπνίζω | Correlation Coefficient | -,133 | ,099 | 1,000 | -,137 | ,096 | ,270 | ,158 |
| | Sig. (2-tailed) | ,375 | ,507 | . | ,363 | ,521 | ,078 | ,321 |
| | N | 37 | 37 | 37 | 37 | 37 | 37 | 37 |
| Παραλείπω το πρωινό γεύμα | Correlation Coefficient | ,058 | ,034 | -,137 | 1,000 | ,206 | ,187 | ,145 |
| | Sig. (2-tailed) | ,685 | ,812 | ,363 | . | ,145 | ,198 | ,336 |
| | N | 37 | 37 | 37 | 37 | 37 | 37 | 37 |
| Παραλείπω το βραδινό γεύμα | Correlation Coefficient | ,179 | ,245 | ,096 | ,206 | 1,000 | ,240 | ,229 |
| | Sig. (2-tailed) | ,203 | ,080 | ,521 | ,145 | . | ,095 | ,126 |
| | N | 37 | 37 | 37 | 37 | 37 | 37 | 37 |
| Λαμβάνω βιταμίνες | Correlation Coefficient | ,211 | ,122 | ,270 | ,187 | ,240 | 1,000 | ,413** |
| | Sig. (2-tailed) | ,146 | ,399 | ,078 | ,198 | ,095 | . | ,007 |
| | N | 37 | 37 | 37 | 37 | 37 | 37 | 37 |
| Λαμβάνω συμπληρώματα διατροφής | Correlation Coefficient | ,187 | ,160 | ,158 | ,145 | ,229 | ,413** | 1,000 |
| | Sig. (2-tailed) | ,215 | ,285 | ,321 | ,336 | ,126 | ,007 | . |
| | N | 37 | 37 | 37 | 37 | 37 | 37 | 37 |

<u>Λύση</u>
Σε ό,τι αφορά τις τιμές Sig στους πίνακες αποτελεσμάτων, σχεδόν όλες ξεπερνούν το 0,01 και το 0,05. Στις περιπτώσεις αυτές απορρίπτεται η υπόθεση H0 ανεξαρτησίας των αντίστοιχων μεταβλητών.

Συμπεραίνεται ότι όλες οι μεταβλητές σχετίζονται μεταξύ τους ανά δύο θετικά, εκτός από τα ζεύγη μεταβλητών «Καπνίζω» - «Προσέχω να είμαι υγιής» και «Καπνίζω» - «Παραλείπω το πρωινό γεύμα», που σχετίζονται αρνητικά μεταξύ τους.

Για τα ζεύγη μεταβλητών «Προσέχω να είμαι υγιής» - «Γυμνάζομαι» και «Λαμβάνω βιταμίνες» - «Λαμβάνω συμπληρώματα διατροφής» ο συντελεστής συσχέτισής τους μπορεί να θεωρηθεί 0, και επομένως πρόκειται για ζεύγη με μεταβλητές ανεξάρτητες μεταξύ τους.

# Βιβλιογραφία/Αναφορές


Κώστογλου, Β., & Αντωνίου, Ε. ( 2021). Θεσσαλονίκη: Τζιόλα.

Γναρδέλλης, Χ., (2019). *Εφαρμοσμένη Στατιστική*. Αθήνα: Παπαζήση.

Χάλκος, Γ. Ε., (2020). *Στατιστική: Θεωρία και Πράξη*. Αθήνα: Δίσιγμα.

# Κριτήρια αξιολόγησης

## Κριτήριο αξιολόγησης 1

**Με ποιον τρόπο ελέγχεται η συσχέτιση ποιοτικών μεταβλητών και με ποιον τρόπο η συσχέτιση ποσοτικών μεταβλητών;**

**Απάντηση**
Για τον έλεγχο συσχέτισης μεταξύ δύο ποιοτικών μεταβλητών, εφαρμόζεται ο έλεγχος $X^2$ ανεξαρτησίας, με τον οποίο διαπιστώνεται αν ισχύει ή όχι η αρχική υπόθεση H0 ότι οι μεταβλητές είναι ανεξάρτητες. Για τον έλεγχο συσχέτισης μεταξύ δύο ποσοτικών μεταβλητών, υπολογίζεται ο συντελεστής συσχέτισής τους και ερμηνεύεται η τιμή του, ανάλογα αν είναι κοντά στο -1, στο 0 ή στο +1.

## Κριτήριο αξιολόγησης 2

**Με ποιον τρόπο ελέγχεται η συσχέτιση μεταξύ μίας ποιοτικής μεταβλητής και μίας ποσοτικής μεταβλητής;**

**Απάντηση**
Δεν υπάρχει συγκεκριμένος τρόπος ελέγχου της συσχέτισης μεταξύ μίας ποιοτικής μεταβλητής και μίας ποσοτικής μεταβλητής. Συνήθως η ποσοτική μετασχηματίζεται σε ποιοτική με τον χωρισμό της σε κλάσεις και κατόπιν γίνεται έλεγχος ανεξαρτησίας $X^2$.

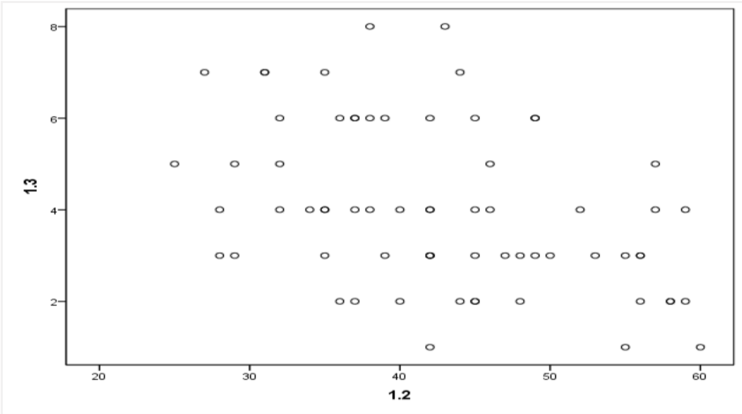
## Κριτήριο αξιολόγησης 3

**Αν ο συντελεστής γραμμικής συσχέτισης δύο μεταβλητών είναι -0,76, τι συμπεραίνετε για τη σχέση των δύο μεταβλητών;**

**Απάντηση**
Είναι έντονα αρνητικά συσχετισμένες. Όταν η μία μεταβλητή αυξάνεται, η άλλη μεταβλητή μειώνεται.

## Κριτήριο αξιολόγησης 4

**Ποια περιμένετε να είναι η τιμή του συντελεστή γραμμικής συσχέτισης στο παρακάτω διάγραμμα διασποράς;**




**Απάντηση**

Η τιμή του συντελεστή γραμμικής συσχέτισης αναμένεται κοντά στο μηδέν, αφού από το διάγραμμα δεν διακρίνεται κάποια γραμμική σχέση μεταξύ των δύο μεταβλητών.

# Κεφάλαιο 6: Παλινδρόμηση

**Σύνοψη**
*Στο κεφάλαιο αυτό αναλύεται η προσαρμογή πραγματικών δεδομένων σε υποδείγματα παλινδρόμησης και παρουσιάζονται παραδείγματα για την εφαρμογή πραγματικών δεδομένων σε θεωρητικά υποδείγματα και την αξιολόγηση της προσαρμογής τους σε αυτά.*

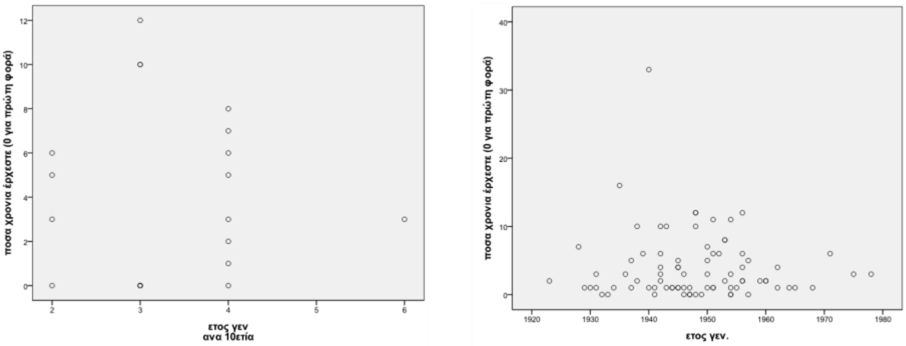
**Προαπαιτούμενη γνώση**
*Περιγραφική παρουσίαση, έλεγχος υποθέσεων.*

## 6.1 Περιγραφή μοντέλου παλινδρόμησης

Ο άνθρωπος από τα πρώτα χρόνια της επιστημονικής ιστορίας του προσπάθησε να εξηγήσει τα φαινόμενα που παρατηρούσε και να βρει τις αιτίες τους, ώστε αργότερα να μπορεί να κάνει προβλέψεις για την εξέλιξή τους. Άρχισε να καταγράφει με μετρήσεις και προσπαθούσε να εντοπίζει σχέσεις μεταξύ των μετρήσεων αυτών. Ο εντοπισμός των πιθανών σχέσεων μπορεί να γίνει είτε παρατηρώντας προσεκτικά τα δεδομένα είτε απεικονίζοντάς τα με ένα διάγραμμα, είτε βρίσκοντας μία μαθηματική σχέση που συνδέει τις τιμές τους.

**Πίνακας 6.1** *Καταγραφή μετρήσεων για δύο ποσοτικές μεταβλητές.*

| Ηλικία (σε δεκαετία) | Συχνότητα επίσκεψης στα λασπόλουτρα |
|---|---|
| 4 | 2 |
| 4 | 0 |
| 3 | 0 |
| 4 | 1 |
| 4 | 6 |
| 2 | 0 |
| 2 | 5 |
| 3 | 0 |
| 4 | 7 |
| 4 | 8 |
| 3 | 0 |
| 3 | 12 |
| 3 | 10 |
| 2 | 3 |
| 3 | 0 |
| 4 | 3 |
| 4 | 5 |
| 3 | 10 |
| 2 | 6 |
| 3 | 0 |
| 6 | 3 |

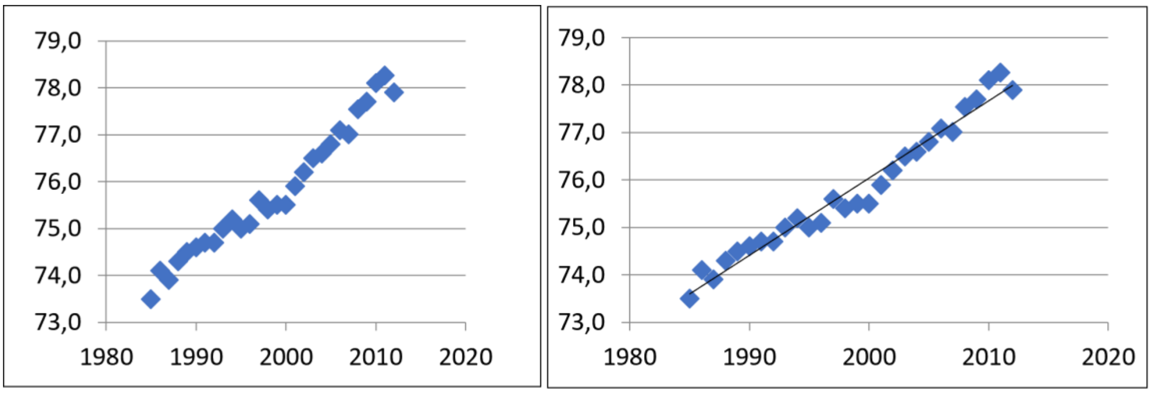


**Εικόνα 6.1** *Δύο διαγράμματα διασποράς (αριστερά: με λίγες τιμές, δεξιά: με περισσότερες τιμές).*

### 6.1.1 Απλή παλινδρόμηση

Έχοντας καταγράψει μετρήσεις για δύο ποσοτικές μεταβλητές, μπορούμε να τοποθετήσουμε τις μετρήσεις αυτές σε δύο άξονες (οριζόντιο και κατακόρυφο) και να αντιστοιχίσουμε κάθε μέτρηση με μία τελεία στο επίπεδο των δύο αξόνων. Αυτό είναι το διάγραμμα διασποράς που περιγράφηκε στο Κεφάλαιο 3. Στον Πίνακα 6.1 υπάρχουν οι μετρήσεις για τις μεταβλητές Χ (ηλικία) και Υ (πλήθος επισκέψεων σε λασπόλουτρα), οι οποίες απεικονίζονται στο αριστερό μέρος της Εικόνας 6.1.

Παρατηρώντας το αριστερό διάγραμμα διασποράς με τις 21 τελείες-σημεία, που αντιστοιχούν στις καταγραφές του Πίνακα 6,1, εμφανίζονται κάποια κατακόρυφα σχήματα στις οριζόντιες τιμές 2 και 4. Στο δεξί διάγραμμα του σχήματος της Εικόνας 6.1 απεικονίζονται τα έτη γέννησης και το πλήθος επισκέψεων στα λασπόλουτρα για 83 καταγραφές ατόμων. Και στο σχήμα αυτό δεν μπορούμε να παρατηρήσουμε κάποια ιδιαίτερη «τάση» που σχηματίζεται από τις τελείες-σημεία. Στις περιπτώσεις αυτές, το συμπέρασμα στο οποίο καταλήγουμε είναι ότι δεν υπάρχει σχέση εξάρτησης μεταξύ ηλικίας και πλήθους επισκέψεων.




**Εικόνα 6.2** *Απεικόνιση του προσδόκιμου ζωής σε σχέση με το έτος (αριστερά: καταγραφές έτους - προσδόκιμου ζωής, δεξιά: το ίδιο γράφημα αλλά και με γραμμικό μοντέλο ως ευθεία γραμμή).*

Στην Εικόνα 6.2 παρουσιάζονται καταγραφές δύο άλλων μεταβλητών, του έτους (οριζόντιος άξονας) και του προσδόκιμου χρόνου ζωής σε έτη (κατακόρυφος άξονας). Παρατηρούμε ότι οι τελείες-σημεία έχουν μια ανοδική τάση (όσο μεγαλώνει η χρονολογία, τόσο αυξάνεται το προσδόκιμο ζωής). Στην περίπτωση αυτή, μπορούμε να ισχυριστούμε ότι υπάρχει μια σχέση «γραμμής» μεταξύ του χρόνου και του προσδόκιμου ζωής, η οποία εμφανίζεται στο δεξί σχήμα της Εικόνας 6.2 και θα μπορούσε να περιγραφεί με την εξίσωση ευθείας ενός επιπέδου που είναι:

$$Y = \alpha+\beta X$$

όπου Χ αντιστοιχεί στο έτος και Υ στο προσδόκιμο ζωής. Επειδή όπως η ευθεία πλησιάζει στα σημεία χωρίς να είναι όλα πάνω σε αυτή, αλλά κάποια να είναι λίγο πιο ψηλά ή πιο χαμηλά από την ευθεία, σωστότερα θα ήταν να γράψουμε τη σχέση μεταξύ Χ και Υ με την εξής εξίσωση:

$$Yi = α+βXi+ei$$

Η εξίσωση αυτή ονομάζεται γραμμικό μοντέλο, που ερμηνεύει την τιμή Y του προσδόκιμου ζωής, όταν γνωρίζουμε την τιμή χρονολογίας $X_i$ αλλά πάντα υπάρχει και ένα σφάλμα το οποίο συμβολίζουμε με $e_i$. Δυστυχώς, δεν υπάρχει τρόπος να μηδενίσουμε το σφάλμα αυτό, αφού στην ανθρώπινη ζωή τίποτα δεν μπορεί να προβλεφθεί με απόλυτη ακρίβεια.

Οι παράμετροι α και β του γραμμικού μοντέλου είναι σταθεροί αριθμοί που αντιστοιχούν στην τομή της ευθείας γραμμής με τον κατακόρυφο άξονα και στην κλίση της ευθείας γραμμής αντίστοιχα.

## 6.1.2 Πολλαπλή παλινδρόμηση

Το μοντέλο παλινδρόμησης μπορεί να γενικευτεί ανάλογα, στην περίπτωση που τα δεδομένα περιγράφονται από περισσότερες μεταβλητές και στόχος είναι να προσδιοριστεί μία από τις μεταβλητές αυτές σε σχέση με τις υπόλοιπες. Η μορφή του μοντέλου στην περίπτωση αυτή γίνεται

$$Yi = α+β1X1i+β2X2i+β3X3i+…+βκXκi +ei.$$

Οι μεταβλητές X1,X2,…Xκ ονομάζονται ανεξάρτητες μεταβλητές και η μεταβλητή Y εξαρτημένη, αφού σχετίζεται με τις ανεξάρτητες X1,X2,…Xκ. Για όλες τις μεταβλητές αυτές υπάρχουν n σε πλήθος μετρήσεις, οι οποίες όμως δεν είναι εύκολο να παρασταθούν γραφικά, όπως στην απλή παλινδρόμηση, γιατί θα χρειάζονταν κ+1 άξονες. Δεν υπάρχει η δυνατότητα να «εμφανιστούν» οι τελείες που αντιστοιχούν στα δεδομένα και να φανεί η μεταξύ τους γραμμική ή άλλου είδους σχέση.

Στις περιπτώσεις αυτές, για να υπάρχει αντικειμενική εκτίμηση της ύπαρξης γραμμικής σχέσης υπολογίζεται ο **συντελεστής γραμμικής συσχέτισης** μεταξύ της μεταβλητής Y και καθεμίας από τις Xκ μεταβλητές. Αν η τιμή του συντελεστή γραμμικής συσχέτισης διαφέρει σημαντικά από το μηδέν, τότε υπάρχει γραμμική σχέση μεταξύ των δύο μεταβλητών και έχει ερμηνεία το γραμμικό μοντέλο παλινδρόμησης. (Γναρδέλης, 2019, σελ. 503).

Στην περίπτωση πολλών μεταβλητών, ελέγχουμε τον συντελεστή γραμμικής συσχέτισης τους ανά δύο, ή υπολογίζουμε και τον **συντελεστή μερικής συσχέτισης,** προκειμένου να μετρηθεί η συσχέτιση μεταξύ εξαρτημένης και ανεξάρτητης μεταβλητής, αφαιρώντας πιθανές επιδράσεις από άλλες ανεξάρτητες μεταβλητές, που προκύπτουν από τις συγχυτικές λεγόμενες μεταβλητές (Γναρδέλης, 2019, σελ. 572). Στην εκτίμηση της παλινδρόμησης καλό είναι να σχετίζεται γραμμικά η εξαρτημένη μεταβλητή με τις ανεξάρτητες μεταβλητές. Αν δύο ή περισσότερες από τις ανεξάρτητες μεταβλητές είναι γραμμικά συσχετισμένες μεταξύ τους, προκαλείται το πρόβλημα της **πολυσυγγραμμικότητας** στο μοντέλο της παλινδρόμησης. Στην πράξη, εκτιμάται το μοντέλο πολλαπλής παλινδρόμησης με τις πολλές μεταβλητές και κατόπιν αξιολογείται η καταλληλότητά του, όπως θα παρουσιαστεί σε επόμενες παραγράφους.

**Πίνακας 6.2** *Βήματα παλινδρόμησης.*

| | **Βήμα** | **Διαδικασία** |
|---|---|---|
| 1 | Προετοιμασία για δημιουργία μοντέλου | Προϋποθέσεις των μεταβλητών, του δείγματος και της μορφής του υποδείγματος (μοντέλου). |
| 2 | Αξιολόγηση συνολική του μοντέλου μετά από την εκτίμηση | Εγκυρότητα του μοντέλου.<br>Η υπόθεση για την παλινδρόμηση συνολικά είναι<br>H0 : β1 = β2 = ... = βk = 0 με<br>H1 : κάποιο βi ≠ 0 .<br>Το μοντέλο της παλινδρόμησης είναι έγκυρο μόνο όταν η H0 απορριφθεί. |
| 3 | Αξιολόγηση ξεχωριστά για κάθε εκτιμημένο συντελεστή | Ερμηνεία των συντελεστών του μοντέλου. Η υπόθεση για κάθε συντελεστή ξεχωριστά είναι H0 : β j = 0 H1 : β j ≠ 0 για j = 1,..., k όταν απορρίπτεται η H0, για κάποιο j = 1,..., k , τότε η ανεξάρτητη μεταβλητή X j έχει σημαντική συμβολή στην ερμηνεία της Y |

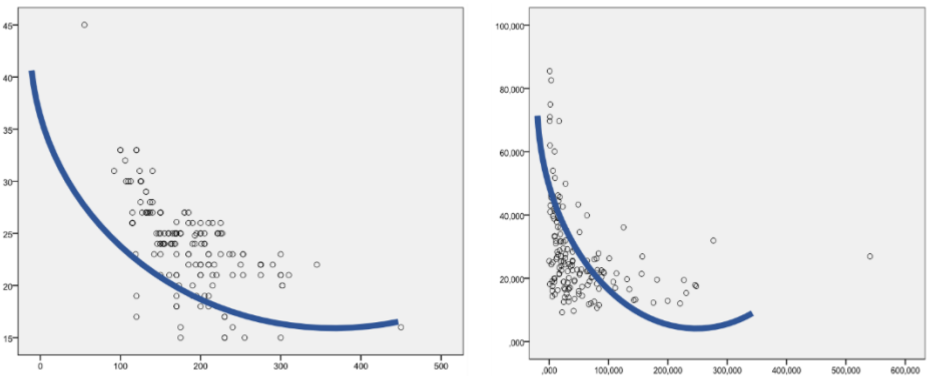
| | | |
|---|---|---|
| 4 | Εύρεση ποσοστού ερμηνείας δεδομένων από την παλινδρόμηση | Υπολογισμός συντελεστή προσδιορισμού |
| 5 | Εκ των υστέρων έλεγχος των προϋποθέσεων | Έλεγχοι κατανομής υπολοίπων, ετεροσκεδαστικότητας, πολυσυγγραμμικότητας αυτοσυσχέτισης κ.ά. |
| 6 | Δημιουργία νέου μοντέλου αν χρειαστεί | Αλλαγή - τροποποίηση μεταβλητών, εισαγωγή νέων μεταβλητών |

### 6.1.3 Διαδικασία παλινδρόμησης

Τα βήματα της διαδικασίας της παλινδρόμησης στη γενική της μορφή παρουσιάζονται σύντομα στον Πίνακα 6.2. Σε κάθε βήμα αντιστοιχεί μία διαδικασία ελέγχου ή υπολογισμού.

## 6.2 Μετατροπή σε γραμμικό μοντέλο παλινδρόμησης

Μερικές φορές υπάρχει σχέση μεταξύ δύο μεταβλητών X και Y, αλλά η σχέση αυτή δεν είναι γραμμική. Στο διάγραμμα διασποράς της Εικόνας 6.3, τα σημεία που αντιστοιχούν στα δεδομένα σχηματίζουν μία καμπύλη και όχι μία τάση ευθείας γραμμής. Η παλινδρόμηση εκτιμά μόνο γραμμικές σχέσεις μεταξύ των μεταβλητών. Για να μπορεί να γίνει εκτίμηση και άλλου είδους σχέσεων με παλινδρόμηση, θα πρέπει η μη γραμμική σχέση να μετατραπεί σε κάποια γραμμική σχέση, η οποία θα εκτιμηθεί με παλινδρόμηση. Σε παρόμοιες περιπτώσεις, κάνοντας μετασχηματισμό στις μεταβλητές, είναι δυνατόν να δημιουργηθεί γραμμικό μοντέλο μεταξύ των μετασχηματισμένων μεταβλητών, οι παράμετροι του οποίου εκτιμώνται με παλινδρόμηση. Στον Πίνακα 6.3 υπάρχουν παραδείγματα είδους μετασχηματισμών μη γραμμικών μοντέλων σε γραμμικά μοντέλα.



**Εικόνα 6.3** *Διαγράμματα διασποράς σε σχήμα καμπύλης.*

**Πίνακας 6.3** *Είδη μετασχηματισμών για μετατροπή σε γραμμικό μοντέλο.*

| **Είδος σχέσης εξαρτημένης μεταβλητής** | **Μετατροπή σε γραμμικό μοντέλο** |
|---|---|
| $Y=\alpha+\beta X^2$ καμπύλη παραβολής κατακόρυφης | $Y=\alpha+\beta Z$ όπου $Z=X^2$ |
| $X=\alpha+\beta Y^2$ καμπύλη παραβολής οριζόντιας | $Y=\gamma+\delta Z$ όπου $Z=\sqrt{X}$ |
| $Y=\alpha X^{\beta}$ | $Y'=\gamma+\beta Z$ όπου $Z=\log X$ κι $Y'=\log Y$ |
| $Y=e^{\alpha+\beta X}$ | $Y'=\alpha+\beta X$ όπου $Y'=\log Y$ |
| $Y=\alpha+\beta 1X+\beta 2X^2$ πολυώνυμο 2ου βαθμού | $Y=\alpha+\beta 1X+\beta 2Z$ όπου $Z=X^2$ |
| $Y=\alpha X1^{\beta 1}\ X2^{\beta 1}$ | $Y'=\gamma+\beta 1Z1+\beta 2Z2$ όπου $Y'=\log y$, $Z1=\log X1$ κι $Z2=\log X2$ |

Ιδιαίτερα στην πολλαπλή παλινδρόμηση, υπάρχει πιθανότητα πολλών μη γραμμικών σχέσεων τόσο για κάθε ανεξάρτητη μεταβλητή, όσο και για συνδυασμούς των ανεξάρτητων μεταβλητών. Μία νέα μετασχηματισμένη μεταβλητή θα μπορούσε να είναι το γινόμενο δύο ανεξάρτητων μεταβλητών ή το γινόμενο μίας ανεξάρτητης μεταβλητής με το τετράγωνο μιας άλλης ανεξάρτητης μεταβλητής.

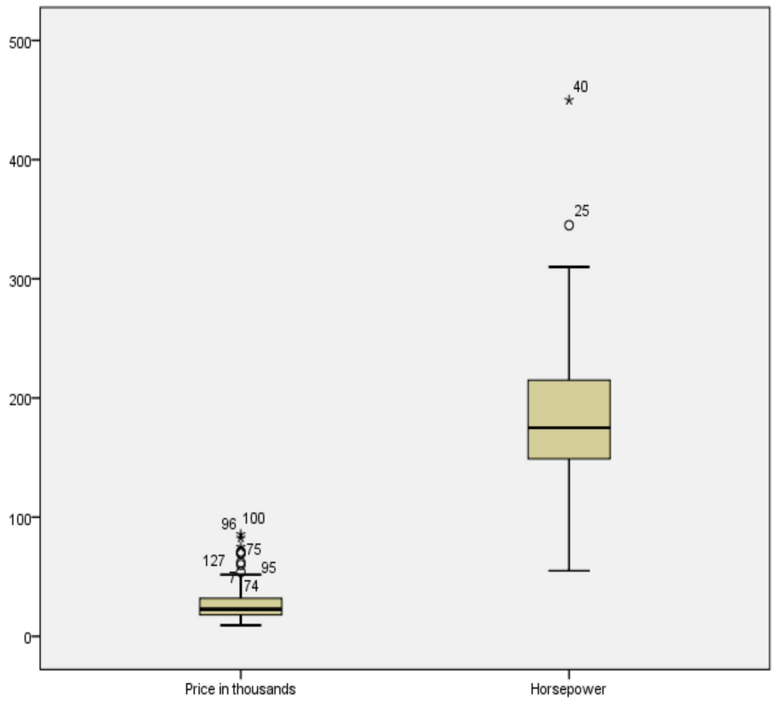
## 6.3 Προϋποθέσεις – Έλεγχοι προϋποθέσεων

Η χρησιμότητα της παλινδρόμησης είναι μεγάλη στην εκτίμηση και γραμμικών και μη γραμμικών μοντέλων. Για να είναι, όμως, αξιόπιστα τα αποτελέσματά της, θα πρέπει να ισχύουν κάποιες προϋποθέσεις στα δεδομένα. Οι προϋποθέσεις αυτές αφορούν το μέγεθος του δείγματος, τις μεταβλητές που χρησιμοποιούνται και τις πιθανές μεταξύ τους σχέσεις, τα κατάλοιπα (υπόλοιπα) της εκτίμησης και τις πιθανές τους σχέσεις με τις μεταβλητές, και την εξειδίκευση του μοντέλου, δηλαδή τον τρόπο με τον οποίο γράφεται το αρχικό μοντέλο που θα εκτιμηθεί. Στη συνέχεια παρουσιάζονται αναλυτικότερα οι προϋποθέσεις αυτές.

### 6.3.1 Προϋποθέσεις δείγματος δεδομένων

Το μέγεθος του δείγματος πρέπει να είναι ικανοποιητικό, π.χ. το πλήθος των παρατηρήσεων να είναι τουλάχιστον πενταπλάσιο από το πλήθος των ανεξάρτητων μεταβλητών.

Πρέπει να δίνεται ιδιαίτερη προσοχή στις εξαιρετικές τιμές (outliers). Αυτές είναι τιμές ιδιαίτερα μεγάλες ή μικρές σε σχέση με τις υπόλοιπες. Κάποιες φορές είναι χρήσιμες (σε μερικές περιπτώσεις είναι οι μόνες τιμές που μας ενδιαφέρουν), και επομένως πρέπει να συμπεριληφθούν στην ανάλυση. Σε άλλες περιπτώσεις, πρόκειται για λανθασμένες τιμές ή μη κανονικές, και πρέπει να αφαιρεθούν από την ανάλυση. Εξαιρετικές και ασυνήθιστες τιμές εντοπίζονται με τη χρήση των μέτρων θέσεων κάθε μεταβλητής και την εμφάνιση του θηκογράμματος (όπως στην Εικόνα 6.4).




**Εικόνα 6.4** *Θηκόγραμμα με εμφάνιση ακραίων τιμών.*

### 6.3.2 Προϋποθέσεις μεταβλητών μοντέλου

Πρέπει να διερευνώνται όλες οι εμπλεκόμενες ανεξάρτητες μεταβλητές. Πρέπει να υπάρχει μια σχέση αιτίας – αποτελέσματος (cause – effect), κατά την οποία οι ανεξάρτητες μεταβλητές ερμηνεύουν την εξαρτημένη. Πρέπει να υπάρχει μια γραμμική σχέση μεταξύ της εξαρτημένης και των ανεξάρτητων μεταβλητών. Αν η σχέση είναι μη γραμμική, τότε πρέπει είτε κάποιες μεταβλητές να μετασχηματιστούν (π.χ. με την τετραγωνική ρίζα ή τον λογάριθμο) είτε να προστεθεί κάποια επιπλέον μεταβλητή.

Όλες οι μεταβλητές (εξαρτημένες ή/και ανεξάρτητες) πρέπει να ανήκουν τουλάχιστον στη διαστηματική (interval) κλίμακα. Όταν η εξαρτημένη μεταβλητή είναι δίτιμη (binary), τότε εφαρμόζεται μια ειδική περίπτωση παλινδρόμησης που λέγεται λογιστική (logistic). Όταν οι ανεξάρτητες μεταβλητές είναι κατηγορικές (categorical), τότε μπορούν να μετατραπούν σε εικονικές δίτιμες (dummy) μεταβλητές.

## 6.3.3 Προϋποθέσεις καταλοίπων εκτίμησης

Για τον έλεγχο τόσο των προϋποθέσεων όσο και της αξιοπιστίας του υποδείγματος, πρέπει να υπάρχει δείγμα παρατηρήσεων και να γίνει η εκτίμηση της παλινδρόμησης. Κατόπιν ελέγχονται οι προϋποθέσεις για τα σφάλματα, τις μεταβλητές καθώς και η αξιοπιστία εκτίμησης του μοντέλου.

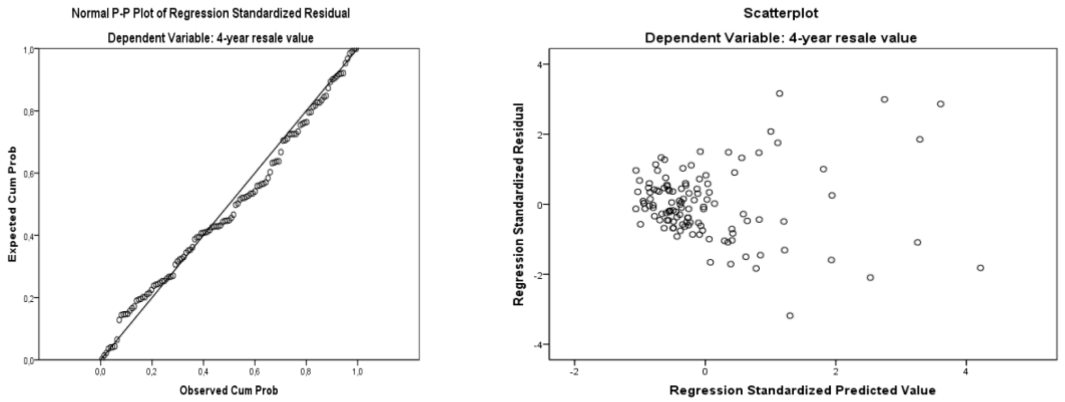
Τα υπόλοιπα ή κατάλοιπα (residuals) θεωρούνται τυχαίες μεταβλητές που πρέπει να έχουν τα παρακάτω χαρακτηριστικά:

Είναι ανεξάρτητα.

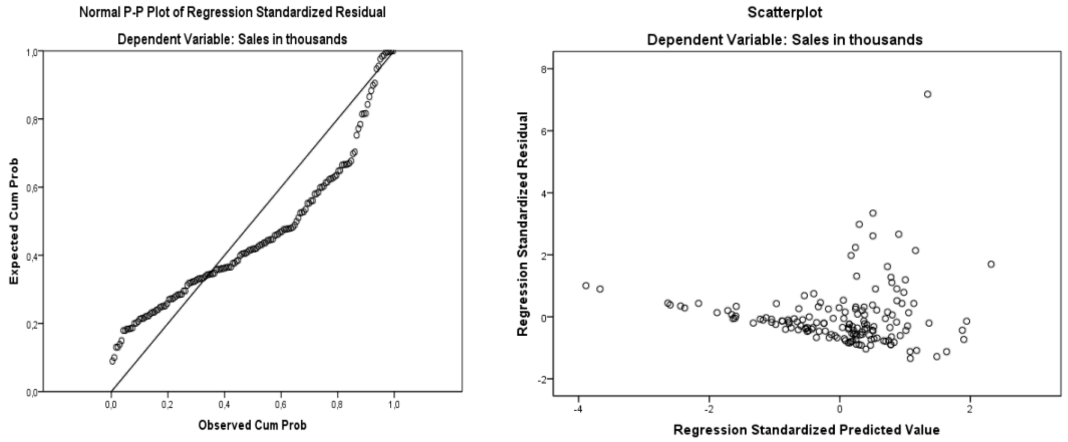
Προέρχονται από την κανονική κατανομή με μέσο όρο 0.

Έχουν την ίδια διακύμανση (ομοσκεδαστικότητα).

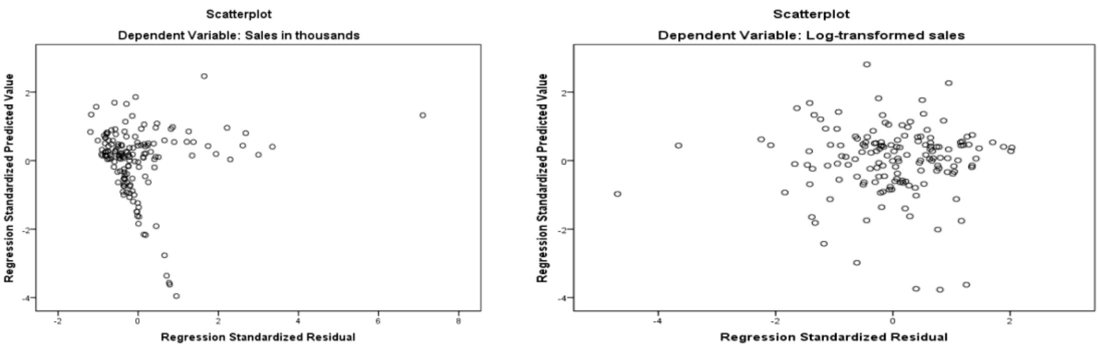
Ο έλεγχος των καταλοίπων γίνεται μετά από την εκτίμηση της παλινδρόμησης και την παρουσίαση του ιστογράμματος των τυποποιημένων υπολοίπων (ή διαγράμματος ZRESID-ZPRED, κανονικοποιημένα κατάλοιπα ως προς κανονικοποιημένες εκτιμώμενες τιμές της εξαρτημένης μεταβλητής Y), ώστε να ελεγχθεί η κανονική κατανομή τους. Επίσης, με το διάγραμμα σε αντίστροφη σειρά (ZPRED-ZRESID) ελέγχεται η σταθερότητα της διακύμανσης ή πιθανή ύπαρξη ετεροσκεδαστικότητας των σφαλμάτων ή η μη γραμμικότητα, αν σχηματίζεται σχήμα κάποιας καμπύλης (Εικόνες 6.5α, 6.5β και 6.6). Πιθανή αυτοσυσχέτιση των υπολοίπων (κυρίως για χρονικές σειρές) ελέγχεται με τη στατιστική Durbin-Watson.




**Εικόνα 6.5α** *Διαγράμματα υπολοίπων παλινδρόμησης για τη μεταβλητή «τιμή μεταπώλησης» ως προς κανονική κατανομή (αριστερά) και ως προς κανονικοποιημένες εκτιμώμενες τιμές (δεξιά).*




**Εικόνα 6.5β** *Διαγράμματα υπολοίπων παλινδρόμησης για τη μεταβλητή «πωλήσεις» ως προς κανονική κατανομή (αριστερά) και ως προς κανονικοποιημένες εκτιμώμενες τιμές (δεξιά).*




**Εικόνα 6.6** *Αριστερά ύπαρξη ετεροσκεδαστικότητας, η οποία εξαλείφεται στο δεξί διάγραμμα με λογαρίθμηση της Υ.*

## 6.3.4 Εξειδίκευση μοντέλου – Έλεγχος

Πρέπει να υπάρχει μια γραμμική σχέση μεταξύ της εξαρτημένης και των ανεξάρτητων μεταβλητών. Αν η σχέση είναι μη γραμμική, τότε είτε πρέπει κάποιες μεταβλητές να μετασχηματιστούν (π.χ. με την τετραγωνική ρίζα ή τον λογάριθμο) είτε να προστεθεί κάποια επιπλέον μεταβλητή (π.χ. τετραγωνίζοντας μία από τις υπάρχουσες).

Δεν πρέπει να υπάρχει πολυσυγγραμμικότητα (multicollinearity), δηλαδή υψηλού βαθμού συσχέτιση μεταξύ των ανεξάρτητων μεταβλητών. Οι γραμμικές σχέσεις μεταξύ των μεταβλητών εντοπίζονται υπολογίζοντας τους συντελεστές γραμμικής συσχέτισης ανά δύο των μεταβλητών. Άλλος τρόπος είναι η εμφάνιση του διαγράμματος διασποράς ανά δύο των μεταβλητών.

Η σχέση της εξαρτημένης και των ανεξάρτητων μεταβλητών πρέπει να είναι προσθετική. Η προϋπόθεση αυτή δεν ισχύει όταν υπάρχει αλληλεπίδραση μεταβλητών (interaction) στο μοντέλο, που είναι ίση με το γινόμενο δύο ή περισσότερων άλλων ανεξάρτητων μεταβλητών. Τότε, δημιουργείται μια νέα μεταβλητή γινόμενο και εισάγεται στο μοντέλο προσθετικά.

Με τα διαγράμματα των κανονικοποιημένων υπολοίπων (ZRESID) και των κανονικοποιημένων προβλεπόμενων τιμών (ZPRED), ελέγχεται η καταλληλότητα των μεταβλητών που υπάρχουν στο μοντέλο ή αν λείπει κάποια μεταβλητή από το μοντέλο. Στην περίπτωση κατά την οποία στο διάγραμμα (ZRESID-ZPRED) όλα τα σημεία εμφανίζονται σε οριζόντιο ορθογώνιο, χωρίς να εμφανίζεται κάποιο σχήμα (τρίγωνο ή καμπύλη), η εξειδίκευση του υποδείγματος είναι καλή. Αν εμφανίζεται κάποια μορφή σχήματος, θα πρέπει να ελεγχθούν και να προσαρμοστούν ανάλογα οι μεταβλητές του μοντέλου (σχήμα της Εικόνας 6.6).

### 6.3.4.1 Έλεγχος πολυσυγγραμμικότητας μεταβλητών

Στην περίπτωση που εισάγονται πολλές ανεξάρτητες μεταβλητές στο μοντέλο, οι οποίες σχετίζονται γραμμικά μεταξύ τους (έχουν υψηλό συντελεστή γραμμικής συσχέτισης), έχουμε το πρόβλημα της πολυσυγγραμικότητας. Το πρόβλημα αυτό δεν μας δίνει καλές εκτιμήσεις των συντελεστών των μεταβλητών, και επομένως όχι καλό μοντέλο εκτίμησης της εξαρτημένης μεταβλητής.

Ο έλεγχος γραμμικών συσχετίσεων μεταξύ των ανεξάρτητων μεταβλητών γίνεται με τους αντίστοιχους συντελεστές γραμμικής συσχέτισης ή με τον ειδικό έλεγχο VIF (Variance Inflation Factor). Ο έλεγχος αυτός γίνεται από τα στατιστικά προγράμματα εκτίμησης της παλινδρόμησης και παρουσιάζονται οι τιμές του για κάθε μεταβλητή. Μεγάλη τιμή VIF για μία μεταβλητή δηλώνει ότι αυτή έχει ισχυρή γραμμική συσχέτιση με άλλες μεταβλητές και δεν χρειάζεται στο μοντέλο (ανεκτές τιμές VIF είναι οι τιμές <10).

Το αντίστροφο της τιμής (1/VIF) ονομάζεται tolerance, και αν η τιμή του είναι μικρότερη του 0,1 υπάρχει πρόβλημα συσχέτισης με άλλες μεταβλητές.

## 6.4 Υπολογισμός ευθείας ελαχίστων τετραγώνων

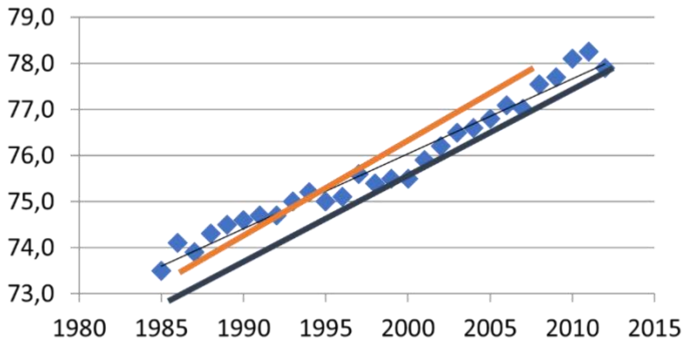
### 6.4.1 Υπολογισμός παραμέτρων α και β στην απλή παλινδρόμηση

Η τιμή του β (κλίση της ευθείας τάσης) μπορεί να προσδιοριστεί αν γνωρίζουμε δύο σημεία της ευθείας. Αν η ευθεία έχει τάση ανόδου, όπως στο σχήμα της Εικόνας 6.2, η τιμή του β είναι θετική, ενώ, αν έχει τάση καθόδου, η τιμή του β είναι αρνητική. Μηδενική τιμή του β σημαίνει ότι η ευθεία είναι οριζόντια και δεν υπάρχει σχέση μεταξύ των δύο μεταβλητών (όπως στο σχήμα της Εικόνας 6.1).

Τίθεται τώρα το ερώτημα, ποια δύο σημεία της ευθείας θα χρησιμοποιηθούν για τον υπολογισμό του β; Τα δύο ακραία σημεία; Δύο ενδιάμεσα σημεία; Δύο τυχαία σημεία; Υπολογίζοντας το β κάθε φορά με δύο διαφορετικά σημεία, προκύπτει διαφορετική τιμή και διαφορετική ευθεία όπως παρουσιάζεται στο σχήμα της Εικόνας 6.7, με σκούρο μπλε χρώμα ή με πορτοκαλί χρώμα. Το βασικό ερώτημα που πρέπει να απαντηθεί είναι: Ποια από όλες τις δυνατές ευθείες (μαύρη, μπλε, πορτοκαλί) είναι σωστότερη, για να ερμηνεύει την ανοδική τάση των δεδομένων μας; Η απάντηση αυτή δόθηκε με τη βελτιστοποίηση του αθροίσματος τετραγώνων διαφορών μεταξύ πραγματικής τιμής Yi και εκτιμημένης από την ευθεία τιμής Yi (Κώστογλου & Αντωνίου, 2021, σελ. 382), ώστε να βρεθεί η βέλτιστη ευθεία, που είναι κοντά σε όσο το δυνατόν περισσότερα σημεία-δεδομένα. Η ευθεία αυτή ονομάζεται ευθεία ελαχίστων τετραγώνων, επειδή για να βρεθεί ελαχιστοποιείται το άθροισμα τετραγώνων των κάθετων αποστάσεων των σημείων από την ευθεία. Οι τιμές των παραμέτρων α και β υπολογίζονται από τη συνθήκη μηδενισμού των πρώτων μερικών παραγώγων, με τη χρήση των εξής εξισώσεων:

$$\beta = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2} \qquad a = \overline{Y} - \beta\overline{X} \tag{6.1}$$

Ο υπολογισμός των παραμέτρων α και β, με τη χρήση των τύπων (Εικόνα 6.7) και όλων των n σε πλήθος σημείων (Xi, Yi) που έχουμε στη διάθεσή μας, δίνει μία μόνο ευθεία που προσαρμόζεται όσο το δυνατόν καλύτερα στα δεδομένα μας. Η ευθεία αυτή ονομάζεται ευθεία παλινδρόμησης. Το πρόσημο της παραμέτρου β δίνει πληροφορίες για τη φορά της ευθείας (ανοδική ή καθοδική). Ο αριθμός δίνει πληροφορίες για την κατεύθυνση της ευθείας (κλίση οριζόντια, όταν είναι κοντά στο 0, ή κατακόρυφη, όσο απομακρύνεται από το 0). Επίσης, η τιμή του β εκφράζει τη μεταβολή της μεταβλητής Y, όταν η μεταβλητή X μεταβληθεί κατά 1 μονάδα. Η τιμή του α εκφράζει την τιμή της μεταβλητής Y, όταν η μεταβλητή X=0.



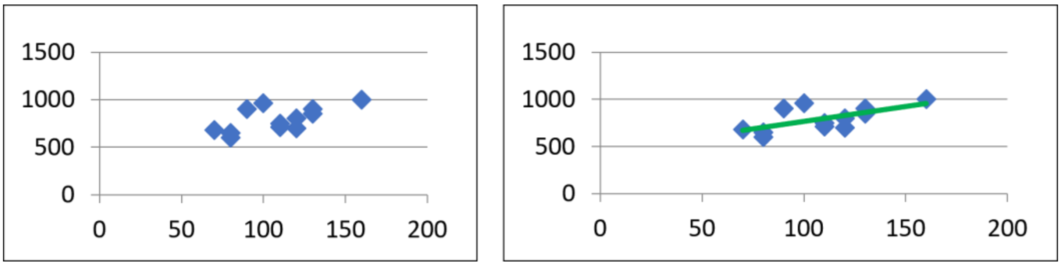

**Εικόνα 6.7** *Προσαρμογή ευθειών στα δεδομένα και εξισώσεις εύρεσης συντελεστών.*

#### 6.4.1.1 Παράδειγμα υπολογισμού παραμέτρων απλής παλινδρόμησης

Στον Πίνακα 6.4 καταγράφεται ένα παράδειγμα μηνιαίων δαπανών μιας επιχείρησης και αντίστοιχων εισπράξεων. Πιθανολογείται ότι υπάρχει γραμμική σχέση μεταξύ των δαπανών και των εισπράξεων και παρουσιάζονται στο Διάγραμμα 6.8 (αριστερό μέρος) τα σημεία. Υπολογίζονται τα αθροίσματα ΣXi = 1.300, ΣYi = 9.500, ΣXi2 = 148.200, ΣXiYi = 1.052.700, και κατόπιν, κάνοντας αντικατάσταση στους τύπους για πλήθος n = 12, υπολογίζονται οι παράμετροι που είναι α = 446 και β = 3,19. Η εξίσωση της ευθείας ελαχίστων τετραγώνων για το παράδειγμα αυτό είναι η Y = 446 +3,19 X και παρουσιάζεται με πράσινο χρώμα στο δεξί μέρος του σχήματος της Εικόνας 6.8. Παρατηρούνται αρκετά σημεία πολύ κοντά στην ευθεία αλλά και κάποια σημεία που απέχουν από αυτήν. Γενικά, η τάση είναι ανοδική, όπως προκύπτει από τη θετική τιμή του β = 3,19. Για κάθε αύξηση των δαπανών κατά 1 μονάδα, προκύπτει αύξηση των εισπράξεων κατά 3,19 μονάδες.

**Πίνακας 6.4** *Παράδειγμα δεδομένων για δαπάνες και αντίστοιχες εισπράξεις.*

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Δαπάνες X** | 80 | 120 | 70 | 110 | 130 | 100 | 90 | 160 | 120 | 80 | 110 | 130 |
| **Εισπράξεις Y** | 600 | 700 | 680 | 710 | 900 | 960 | 900 | 1.000 | 800 | 650 | 750 | 850 |




**Εικόνα 6.8** *Αριστερά: παράδειγμα δαπανών (X) και εισπράξεων (Y), δεξιά: το ίδιο διάγραμμα με αποτύπωση της ευθείας ελαχίστων τετραγώνων.*

### 6.4.2 Υπολογισμός παραμέτρων α και β1, β2,... βκ στην πολλαπλή παλινδρόμηση

Κατά τρόπο ανάλογο με τον υπολογισμό των παραμέτρων της απλής παλινδρόμησης υπολογίζονται και οι παράμετροι της πολλαπλής παλινδρόμησης, αλλά με τη χρήση υπολογιστή, μιας που οι πράξεις είναι πιο περίπλοκες.

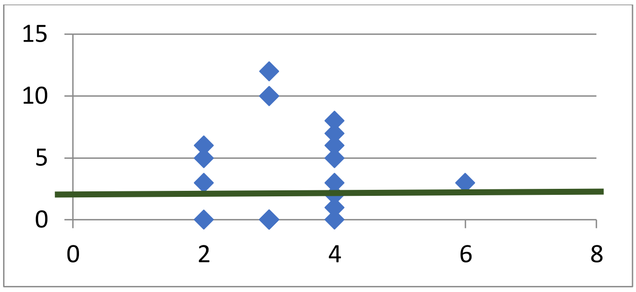
Η τιμή της καθεμίας από τις παραμέτρους β1, β2,... βκ εκφράζει τη σχέση που συνδέει την εξαρτημένη μεταβλητή Y με την αντίστοιχη ανεξάρτητη μεταβλητή X1 ή X2 ή Xκ. Αν η τιμή του αντίστοιχου β είναι θετική, η σχέση Y και X είναι ανάλογη, ενώ, αν η τιμή του β είναι αρνητική, η σχέση είναι αντιστρόφως ανάλογη. Μηδενική τιμή του β σημαίνει ότι δεν υπάρχει σχέση μεταξύ των δύο μεταβλητών.

Η ακριβής ερμηνεία καθενός βj είναι ότι, αν όλες οι υπόλοιπες ανεξάρτητες μεταβλητές δεν μεταβληθούν και μόνο η τιμή της μεταβλητής Xj μεταβληθεί κατά μία μονάδα, τότε η μεταβολή της εξαρτημένης μεταβλητής Y θα είναι ίση με βj μονάδες. Αυτός είναι και ο λόγος που οι παράμετροι β ονομάζονται και «καθαροί συντελεστές παλινδρόμησης» (Berenson et al., 2018, σελ. 592).

## 6.5 Αξιολόγηση της ευθείας παλινδρόμησης

Έχοντας, λοιπόν, ένα πλήθος n σημείων δεδομένων για δύο ποσοτικές μεταβλητές X και Y, πάντα μπορεί με χρήση των τύπων να προσδιοριστούν οι τιμές των παραμέτρων α και β και η καλύτερη δυνατή ευθεία παλινδρόμησης Y = α+βX.

Δυστυχώς, όμως, η καλύτερη δυνατή ευθεία που προσδιορίζεται δεν είναι και η ευθεία που ανταποκρίνεται στα δεδομένα, αν αυτά δεν έχουν «τάση» ευθείας γραμμής. Για τα δεδομένα του παραδείγματος της Εικόνας 6.9, η καλύτερη δυνατή ευθεία, παρουσιάζεται στο σχήμα της Εικόνας 6.9, είναι σχεδόν οριζόντια. Πολλά σημεία δεδομένων δεν είναι κοντά στην ευθεία αυτή. Χρειάζεται, λοιπόν, κάποιο «μέτρο» να αξιολογεί πόσο καλή είναι η ευθεία. Το μέτρο αυτό υπολογίζεται, αφού υπολογιστούν οι παράμετροι α και β της ευθείας, και ονομάζεται ο **συντελεστής προσδιορισμού**. Συμβολίζεται με $R^2$.



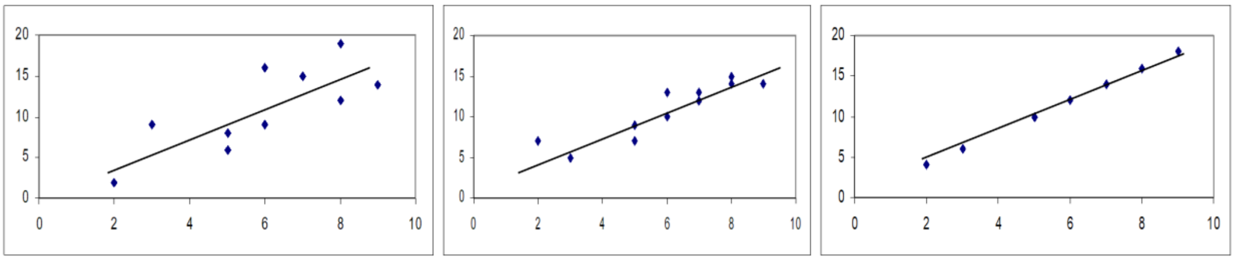

**Εικόνα 6.9** *Παράδειγμα με ευθεία παλινδρόμησης που δεν πλησιάζει τα δεδομένα.*

Ο συντελεστής προσδιορισμού $R^2$ υπολογίζεται ως το πηλίκο του αθροίσματος των τετραγώνων των διαφορών των εκτιμημένων από την ευθεία τιμών Y από τον μέσο όρο τους προς το άθροισμα των τετραγώνων των διαφορών των τιμών Y από τον μέσο όρο τους.

$$R^2 = \frac{\sum_1^n (\hat{y} - \bar{y})^2}{\sum_1^n (y - \bar{y})^2} \tag{6.2}$$

Είναι πάντα θετικός αριθμός μεταξύ 0 και 1 και εκφράζει σε ποσοστό το πλήθος των πραγματικών δεδομένων που ερμηνεύονται από την ευθεία παλινδρόμησης.

Στο σχήμα της Εικόνας 6.7 εμφανίζονται τρία διαφορετικά σύνολα δεδομένων που περιγράφονται από δύο μεταβλητές, από τα οποία προκύπτει η ίδια ευθεία παλινδρόμησης. Ο συντελεστής προσδιορισμού της όμως, διαφέρει για κάθε σύνολο και είναι μεγαλύτερος στην περίπτωση που περισσότερα σημεία-δεδομένα πλησιάζουν την ευθεία. Η τιμή του είναι $R^2 = 0{,}73$ για το αριστερό σύνολο του σχήματος της Εικόνας 6.10, $R^2 = 0{,}95$ για το μεσαίο σύνολο και $R^2 = 1$ για το σύνολο στα δεξιά, όπου όλα τα σημεία είναι πάνω στην ευθεία παλινδρόμησης.




**Εικόνα 6.10** *Τρεις ευθείες παλινδρόμησης με διαφορετικούς συντελεστές προσδιορισμού.*

Ο συντελεστής προσδιορισμού υπολογίζεται με τον ίδιο τύπο και στην πολλαπλή παλινδρόμηση, και είναι το κύριο μέτρο για την αξιολόγηση της παλινδρόμησης. Λόγω του γεγονότος ότι όσο περισσότερες ανεξάρτητες μεταβλητές υπάρχουν στο μοντέλο τόσο αυξάνεται ο συντελεστής προσδιορισμού, για να μετρηθεί η ερμηνευτικότητα του μοντέλου ανεξάρτητα από το πλήθος των μεταβλητών και σε σχέση με το μέγεθος του δείγματος, υπολογίζεται και ο **διορθωμένος συντελεστής προσδιορισμού** (Ζαφειρόπουλος, 2022, σελ. 50).

### 6.5.1 Παράδειγμα υπολογισμού $R^2$ στην απλή παλινδρόμηση

Στο παράδειγμα μηνιαίων δαπανών (Πίνακας 6.4) υπολογίζουμε τις εκτιμημένες τιμών των Yi με την εξίσωση Yi=446+3,19Xi και κατόπιν τη διαφορά των τιμών αυτών από τη μέση τιμή των Y που είναι 9.500/12 = 791,67. Κατόπιν, ο συντελεστής προσδιορισμού της παλινδρόμησης υπολογίζεται με τον τύπο:

$$R^2 = \frac{\sum_1^n (\hat{y} - \bar{y})^2}{\sum_1^n (y - \bar{y})^2} = \frac{74.964{,}02}{184.766{,}7} = 0{,}405 \tag{6.3}$$

Η τιμή 0,405 σχολιάζεται ως εξής: το 40,5% των δεδομένων ερμηνεύεται από την ευθεία παλινδρόμησης. Το υπόλοιπο σχεδόν 60% εξαρτάται από άλλους παράγοντες και όχι από τη σχέση Y = α+βX.

## 6.6 Έλεγχος σημαντικότητας των συντελεστών

Η εκτίμηση των παραμέτρων α, β του μοντέλου της παλινδρόμησης, προκύπτει από ένα σύνολο n παρατηρήσεων. Αν αλλάξει το σύνολο των παρατηρήσεων, θα προκύψει μία άλλη εκτίμηση, διαφορετική από την προηγούμενη. Για κάθε εκτίμηση του μοντέλου, δεν υπάρχει βεβαιότητα για τις τιμές των παραμέτρων. Για τον λόγο αυτό, αναφέρεται ως «εκτίμηση» και όχι ως υπολογισμός των παραμέτρων του μοντέλου.

Η μέθοδος ελαχίστων τετραγώνων που χρησιμοποιείται, δίνει εκτιμήσεις «αμερόληπτες», οι οποίες δηλαδή πλησιάζουν κατά μέσο όρο τις πραγματικές και άγνωστες τιμές της εξαρτημένης μεταβλητής, με ένα τυπικό σφάλμα εκτίμησης. Το **τυπικό** αυτό **σφάλμα εκτίμησης**, εκφράζεται σε μονάδες μέτρησης της εξαρτημένης μεταβλητής και υπολογίζεται όπως ανάλογα με την τυπική απόκλιση των δεδομένων από τη μέση τιμή τους. Υπολογίζεται ως το άθροισμα τετραγώνων των διαφορών πραγματικών τιμών της εξαρτημένης μεταβλητής από τις εκτιμώμενες τιμές της με το μοντέλο παλινδρόμησης, διαιρεμένο με το πλήθος (μειωμένο κατά 2) των τιμών του δείγματος που χρησιμοποιήθηκε για την εκτίμηση (Berenson et al., 2018, σελ. 550). Με τον ίδιο τύπο υπολογίζεται και η τυπική απόκλιση των σφαλμάτων της παλινδρόμησης.

$$s_{XY} = \frac{\sum_1^n (Y_i - \widehat{Y_i})^2}{n-2} \qquad (6.4)$$

Με τη θεώρηση αυτή, τα α και β δεν θεωρούνται μόνο σταθεροί αριθμοί, αλλά καθεμία από τις παραμέτρους α, β του μοντέλου θεωρείται ως τυχαία στατιστική μεταβλητή. Η εκτίμηση κάθε τέτοιας μεταβλητής, έχει ως μέση τιμή την άγνωστη πραγματική τιμή της παραμέτρου αλλά και τυπική απόκλιση από την πραγματική τιμή. Η τυπική απόκλιση της εκτίμησης κάθε παραμέτρου βj αποδεικνύεται ότι είναι ίση με το πηλίκο του τυπικού σφάλματος εκτίμησης διά του αθροίσματος τετραγώνων διαφορών όλων των n τιμών της αντίστοιχης ανεξάρτητης μεταβλητής Xj από τη μέση τιμή τους. Υπολογίζεται με τον τύπο:

$$s_{\beta j} = \frac{s_{XY}}{\sqrt{\sum_1^n \left(X_{jk} - \bar{X}_j\right)^2}} \qquad (6.5)$$

Τα στατιστικά προγράμματα εκτίμησης παλινδρόμησης μαζί με τις εκτιμήσεις των παραμέτρων, εκτιμούν και τις τυπικές τους αποκλίσεις. Για κάθε τυχαία μεταβλητή $\beta_j$ μπορεί να γίνει έλεγχος υποθέσεων αν η μέση τιμή της είναι 0 με το t-test.

Οι υποθέσεις για την πραγματική άγνωστη παράμετρο βj είναι

Αρχική: H0: βj = 0 και

εναλλακτική: H1 βj ≠ 0.

Η H0 γίνεται δεκτή όταν η τιμή του t στατιστικού είναι μεταξύ των τιμών το πίνακα t κατανομής. Διαφορετικά απορρίπτεται. Τα στατιστικά προγράμματα κατά την εκτίμηση των παραμέτρων βj, υπολογίζουν τόσο την τιμή του, όσο και την τυπική απόκλιση, καθώς και την τιμή p-value (ή sig level), που εκφράζει την πιθανότητα λάθους, αν απορριφθεί η H0. Αν η πιθανότητα αυτή είναι μικρότερη από 0,05, απορρίπτεται η H0 με σφάλμα 5%.

Αν με τον έλεγχο t-test, δεν απορριφθεί η H0 θα προκύψει ότι η πραγματική άγνωστη τιμή $\beta_j$ μπορεί να θεωρηθεί ίση με 0. Δηλαδή στην εξίσωση του μοντέλου παλινδρόμησης η αντίστοιχη μεταβλητή $X_j$ πολλαπλασιάζεται με μηδενικό συντελεστή. Αυτό είναι σαν να μην υπάρχει στην εξίσωση του μοντέλου η μεταβλητή αυτή. Η συνεισφορά της αντίστοιχης ανεξάρτητης μεταβλητής $X_j$ στην ερμηνεία της εξαρτημένης μεταβλητής Y μπορεί να θεωρηθεί μηδενική, οπότε η μεταβλητή αυτή αφαιρείται από το μοντέλο παλινδρόμησης.

## 6.7 Αξιοπιστία προβλέψεων διάστημα εμπιστοσύνης πρόβλεψης

Η εκτιμημένη ευθεία παλινδρόμησης χρησιμοποιείται για να προβλεφθούν οι τιμές της εξαρτημένης μεταβλητής Υ όταν γνωρίζουμε τις τιμές των ανεξάρτητων μεταβλητών Χ. Για να είναι αξιόπιστες οι προβλέψεις, οι τιμές των ανεξάρτητων μεταβλητών πρέπει να είναι μέσα στα όρια των τιμών του δείγματος που χρησιμοποιήθηκαν για την εκτίμηση του μοντέλου. Για την καλύτερη εκτίμηση των προβλέψεων, αντί για τον υπολογισμό μόνο μίας τιμής για την ανεξάρτητη μεταβλητή, υπολογίζεται διάστημα εμπιστοσύνης για την πρόβλεψη, χρησιμοποιώντας τον γενικό τύπο για τα διαστήματα εμπιστοσύνης α%:

(σημειακή εκτίμηση +/- τιμή πίνακα t × τυπικό σφάλμα εκτίμησης) (6.5)

Το τυπικό σφάλμα εκτίμησης εξαρτάται από τις τιμές των ανεξάρτητων μεταβλητών και από το πλήθος των δεδομένων του δείγματος εκτίμησης. Είναι μικρότερο κοντά στη μέση τιμή τους και μεγαλύτερο μακριά από αυτή (Γναρδέλλης, 2019, σελ. 532). Έτσι, οι προβλέψεις είναι ακριβέστερες για τιμές των ανεξάρτητων μεταβλητών κοντά στη μέση τιμή τους. Επίσης, γίνονται ακριβέστερες, όταν οι εκτιμήσεις γίνουν με μεγάλο πλήθος παρατηρήσεων.

## 6.8 Εφαρμογές για κατανόηση

**Πίνακας 6.5** *Παράδειγμα δεδομένων για ζήτηση παγωτού.*

| Μήνας | Ζήτηση παγωτού Y | Τιμή X1 | Τιμή X2 (υποκατάστατου) | Θερμοκρασία X3 |
|---|---|---|---|---|
| M1 | 50 | 10 | 8 | 34 |
| M2 | 53 | 12 | 9 | 35 |
| M3 | 45 | 11 | 10 | 29 |
| M4 | 40 | 11 | 7 | 28 |
| M5 | 48 | 12 | 8 | 32 |
| M6 | 51 | 13 | 9 | 33 |
| M7 | 55 | 11 | 7 | 35 |
| M8 | 42 | 14 | 11 | 30 |
| M9 | 38 | 13 | 10 | 25 |
| M10 | 60 | 10 | 12 | 35 |
| M11 | 59 | 10 | 11 | 31 |

### 6.8.1 Μοντέλο παλινδρόμησης για ζήτηση παγωτού

Στον Πίνακα 6.5 παρουσιάζονται δεδομένα 11 μηνών για τη ζήτηση παγωτού (Y) σε σχέση με την τιμή του (X1), την τιμή ενός υποκατάστατου αγαθού (X2) και τη θερμοκρασία περιβάλλοντος (X3). Υποτίθεται ότι η ζήτηση ακολουθεί το γραμμικό μοντέλο:

$$Y = \alpha+\beta 1X1+\beta 2X2+\beta 3X3 \qquad (6.6)$$

οι συντελεστές του οποίου θα εκτιμηθούν με τη βοήθεια της παλινδρόμησης. Το πρόσημο των συντελεστών β1, β2, β3 δίνει πληροφορίες για την κατεύθυνση της σχέσης των μεταβλητών Y και X1, X2, X3 αντίστοιχα. Η εκτιμημένη τιμή των β1, β2, β3 δίνει πληροφορίες για τη μεταβολή του Y, όταν το X1 ή X2 ή X3 αντίστοιχα μεταβληθεί κατά μία μονάδα.

Με την εκτίμηση του γραμμικού μοντέλου θα απαντηθούν και οι ερωτήσεις:

- Η ζήτηση παγωτού πόσο επηρεάζεται από την τιμή του;

- Η ζήτηση παγωτού επηρεάζεται από τη θερμοκρασία;
- Η ζήτηση παγωτού επηρεάζεται από την τιμή του υποκατάστατου;
- Υπάρχει σταθερός αριθμός διάφορος από το μηδέν στο υποτιθέμενο μοντέλο ζήτησης παγωτού;

**Πίνακας 6.5** *Παράδειγμα δεδομένων για ζήτηση παγωτού.*

| **Μήνας** | **Ζήτηση παγωτού Y** | **Τιμή X1** | **Τιμή X2 (υποκατάστατου)** | **Θερμοκρασία X3** |
|---|---|---|---|---|
| M1 | 50 | 10 | 8 | 34 |
| M2 | 53 | 12 | 9 | 35 |
| M3 | 45 | 11 | 10 | 29 |
| M4 | 40 | 11 | 7 | 28 |
| M5 | 48 | 12 | 8 | 32 |
| M6 | 51 | 13 | 9 | 33 |
| M7 | 55 | 11 | 7 | 35 |
| M8 | 42 | 14 | 11 | 30 |
| M9 | 38 | 13 | 10 | 25 |
| M10 | 60 | 10 | 12 | 35 |
| M11 | 59 | 10 | 11 | 31 |

**Πίνακας 6.6** *Στατιστικά αποτελέσματα παλινδρόμησης για τα δεδομένα του Πίνακα 6.5.*

| **Regression Statistics** | |
|---|---|
| Multiple R | 0,927734 |
| R Square | 0,86069 |
| Adjusted R Square | 0,800986 |
| Standard Error | 3,295292 |
| Observations | 11 |

**Πίνακας 6.7** *Αποτελέσματα ανάλυσης διακύμανσης παλινδρόμησης για τα δεδομένα του Πίνακα 6.5.*

| **ANOVA** | **df** | **SS** | **MS** | **F** | **Significance F** |
|---|---|---|---|---|---|
| Regression | 3 | 469,6237 | 156,5412 | 14,41587 | 0,002218 |
| Residual | 7 | 76,01267 | 10,85895 | | |
| Total | 10 | 545,6364 | | | |

**Πίνακας 6.8** *Εκτίμηση συντελεστών παλινδρόμησης για τα δεδομένα του Πίνακα 6.5.*

| | Coefficients | Standard Error | t Stat | P-value |
|---|---|---|---|---|
| Intercept | 6,491425 | 17,8546 | 0,363571 | 0,72691 |
| X Variable 1 | -1,86462 | 0,816655 | -2,28324 | 0,056366 |
| X Variable 2 | 1,456883 | 0,624169 | 2,334118 | 0,052296 |
| X Variable 3 | 1,607491 | 0,340097 | 4,72656 | 0,002141 |

Τα αποτελέσματα εκτίμησης του μοντέλου με το excel εμφανίζονται στους Πίνακες 6.6, 6.7 και 6.8. Η τιμή του συντελεστή προσδιορισμού $R^2$ είναι 0,86 (Πίνακας 6.6), που σημαίνει ότι η παλινδρόμηση ερμηνεύει το 86% των δεδομένων.

Στη δεύτερη στήλη του Πίνακα 6.8 εμφανίζονται οι εκτιμήσεις των συντελεστών του μοντέλου και στην τελευταία στήλη η τιμή p-value που εκφράζει την πιθανότητα λάθους, αν απορριφτεί η υπόθεση H0, με την οποία υποτίθεται ότι ο αντίστοιχος συντελεστής βi = 0.

Παρατηρώντας τη στήλη με τις p-value τιμές, μία μεγάλη τιμή p = 0,72691 αντιστοιχεί στην υπόθεση H0 για τον σταθερό όρο του μοντέλου. Αυτό σημαίνει ότι δεν μπορεί να απορριφθεί η υπόθεση H0: α = 0, οπότε στο μοντέλο δεν χρειάζεται να υπάρχει σταθερός όρος και πρέπει να εκτιμηθεί από την αρχή νέο μοντέλο χωρίς σταθερό όρο.

Τα αποτελέσματα της νέας αυτής εκτίμησης φαίνονται στους Πίνακες 6.9, 6.10, 6.11. Η νέα τιμή του συντελεστή προσδιορισμού $R^2$ είναι 0,99 (Πίνακας 6.9), που σημαίνει ότι η παλινδρόμηση ερμηνεύει το 99% των δεδομένων και είναι πολύ καλή.

Στη δεύτερη στήλη του Πίνακα 6.11 εμφανίζονται οι εκτιμήσεις των συντελεστών του μοντέλου και στην τελευταία στήλη οι τιμές p-value. Στη στήλη αυτή, κάθε p-value που εκφράζει την πιθανότητα λάθους αν απορριφθεί η υπόθεση (H0: αντίστοιχος συντελεστής β=0), είναι πολύ μικρή, μικρότερη από το 0,05. Αυτό οδηγεί στην απόρριψη των υποθέσεων β1 = 0, β2 = 0 και β3 = 0 με σφάλμα 5%. Επομένως, όλες οι μεταβλητές χρειάζονται στο μοντέλο, αφού οι συντελεστές τους δεν μπορούν να θεωρηθούν μηδέν (δεν είναι ασήμαντοι).

Η εκτίμηση του συντελεστή β1 για την τιμή του παγωτού είναι -1,647 και αυτό σημαίνει ότι για καθεμία μονάδα μεταβολής της τιμής, η ζητούμενη ποσότητα μεταβάλλεται προς την αντίστροφη κατεύθυνση κατά 1,647 μονάδες.

Η εκτίμηση του συντελεστή β2 για την τιμή του υποκατάστατου αγαθού είναι 1,539 και αυτό σημαίνει ότι για καθεμία μονάδα μεταβολής, η ζητούμενη ποσότητα μεταβάλλεται προς την ίδια κατεύθυνση κατά 1,539 μονάδες.

Η εκτίμηση του συντελεστή β3 για τη θερμοκρασία είναι 1,708, και αυτό σημαίνει ότι για καθεμία μονάδα μεταβολής της θερμοκρασίας, η ζητούμενη ποσότητα μεταβάλλεται προς την ίδια κατεύθυνση κατά 1,708 μονάδες.

**Πίνακας 6.9** *Στατιστικά αποτελέσματα παλινδρόμησης χωρίς σταθερό όρο για τα δεδομένα του Πίνακα 6.5.*

| Regression Statistics | |
|---|---|
| Multiple R | 0,998573 |
| R Square | 0,997148 |
| Adjusted R Square | 0,871435 |
| Standard Error | 3,111432 |
| Observations | 11 |

**Πίνακας 6.10** *Αποτελέσματα ανάλυσης διακύμανσης παλινδρόμησης χωρίς σταθερό όρο για τα δεδομένα του Πίνακα 6.5.*

| ANOVA | df | SS | MS | F | Significance F |
|---|---|---|---|---|---|
| Regression | 3 | 27075,55 | 9025,184 | 932,2568 | 0,00000000181 |
| Residual | 8 | 77,44805 | 9,681006 | | |
| Total | 11 | 27153 | | | |

**Πίνακας 6.11** *Εκτίμηση συντελεστών παλινδρόμησης χωρίς σταθερό όρο για τα δεδομένα του Πίνακα 6.5.*

| | Coefficients | Standard Error | t Stat | P-value |
|---|---|---|---|---|
| X Variable 1 | -1,64647 | 0,523082 | -3,14764 | 0,013647 |
| X Variable 2 | 1,539298 | 0,549103 | 2,803294 | 0,02308 |
| X Variable 3 | 1,708567 | 0,184971 | 9,236921 | 0,0000153 |

### 6.8.2 Μοντέλο παλινδρόμησης για πωλήσεις αυτοκινήτων

Εφαρμόζεται βήμα βήμα η μέθοδος της παλινδρόμησης σε ένα υπάρχον αρχείο δεδομένων (car sales.sav) με το SPSS.

Οι μεταβλητές του αρχείου και τα περιγραφικά τους δεδομένα παρουσιάζονται στον Πίνακα 6.12.

**Πίνακας 6.12** *Περιγραφή των μεταβλητών του αρχείου car sales.sav.*

| | Sales in thousands | 4-year resale value | Vehicle type | Price in thousands | Engine size | Horsepower | Wheelbase |
|---|---|---|---|---|---|---|---|
| N | 157 | 121 | 157 | 155 | 156 | 156 | 156 |
| Mean | 52,99808 | 18,07298 | 0,26 | 27,39075 | 3,061 | 185,95 | 107,487 |
| Median | 29,45000 | 14,18000 | 0,00 | 22,79900 | 3,000 | 177,50 | 107,000 |
| Std. Deviation | 68,029422 | 11,453384 | 0,441 | 14,351653 | 1,0447 | 56,700 | 7,6413 |
| Minimum | 0,110 | 5,160 | Automobile | 9,235 | 1,0 | 55 | 92,6 |
| Maximum | 540,561 | 67,550 | Truck | 85,500 | 8,0 | 450 | 138,7 |

Πιστεύεται ότι υπάρχει σχέση αλληλεπίδρασης για τις ποσοτικές μεταβλητές, “Sales in thousands”, “4-year resale value”, “Price in thousands”, “Horsepower”, “Fuel efficiency”, για τις οποίες υπολογίζεται ο συντελεστής συσχέτισης τους ανά δύο.

Τα αποτελέσματα των συσχετίσεων εμφανίζονται στον Πίνακα 6.13. Οι σημαντικές και πολύ σημαντικές συσχετίσεις δηλώνονται με ένα ή δύο αστεράκια στην τιμή του συντελεστή συσχέτισης. Αν η τιμή sig >0,05 σε κάποιο κελί του Πίνακα 6.13, ο συντελεστής συσχέτισης μεταξύ των αντίστοιχων μεταβλητών μπορεί να θεωρηθεί 0. Αυτό σημαίνει ότι δεν υπάρχει γραμμική σχέση μεταξύ τους.

Με βάση τους μη μηδενικούς συντελεστές συσχέτισης του Πίνακα 6.13, προκύπτει ότι η εξαρτημένη μεταβλητή Y = “4-year resale value”, σχετίζεται με τις ανεξάρτητες μεταβλητές X1 = “Price in thousands”, X2 = “Horsepower” και X3 = “Fuel efficiency”. Το μοντέλο για την Y μεταβλητή μπορεί να εκτιμηθεί από τις τρεις μεταβλητές X1, X2 και X3.

**Πίνακας 6.13** *Συντελεστές συσχέτισης μεταξύ των μεταβλητών του αρχείου car sales.sav.*

| | | Sales in thousands | 4-year resale value | Price in thousands | Horsepower | Fuel efficiency |
|---|---|---|---|---|---|---|
| **Sales in thousands** | **Pearson Correlation** | 1 | -,279** | -,305** | -,198* | -0,017 |
| | **Sig. (2-tailed)** | | 0,002 | 0,000 | 0,013 | 0,837 |
| | **N** | 157 | 121 | 155 | 156 | 154 |
| **4-year resale value** | **Pearson Correlation** | -,279** | 1 | ,954** | ,769** | -,401** |
| | **Sig. (2-tailed)** | 0,002 | | 0,000 | 0,000 | 0,000 |
| | **N** | 121 | 121 | 119 | 120 | 119 |
| **Price in thousands** | **Pearson Correlation** | -,305** | ,954** | 1 | ,840** | -,492** |
| | **Sig. (2-tailed)** | 0,000 | 0,000 | | 0,000 | 0,000 |
| | **N** | 155 | 119 | 155 | 155 | 153 |
| **Horsepower** | **Pearson Correlation** | -,198* | ,769** | ,840** | 1 | -,611** |
| | **Sig. (2-tailed)** | 0,013 | 0,000 | 0,000 | | 0,000 |
| | **N** | 156 | 120 | 155 | 156 | 154 |
| **Fuel efficiency** | **Pearson Correlation** | -0,017 | -,401** | -,492** | -,611** | 1 |
| | **Sig. (2-tailed)** | 0,837 | 0,000 | 0,000 | 0,000 | |
| | **N** | 154 | 119 | 153 | 154 | 154 |
| | | | | | | |

Το μοντέλο προς εκτίμηση υποτίθεται ότι είναι $Y = βo+β1X1+β2X2+β3X3+u$, όπου u τα υπόλοιπα-σφάλματα. Επιλέγονται στο SPSS διαδοχικά οι εντολές από τα αντίστοιχα μενού επιλογών, Analyze, Regression, Linear και στη θέση Dependent και Independents οι αντίστοιχες μεταβλητές.

Προέκυψε το output αποτελεσμάτων των Πινάκων 6.14, 6.15, 6.16.

**Πίνακας 6.14** *Αποτελέσματα συντελεστή προσδιορισμού παλινδρόμησης του αρχείου car sales.sav.*

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | ,958[a] | ,918 | ,916 | 3,352784 |

a. Predictors: (Constant), Fuel efficiency, Price in thousands, Horsepower

**Πίνακας 6.15** *Αποτελέσματα ανάλυσης διακύμανσης παλινδρόμησης του αρχείου car sales.sav.*

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 14.424,169 | 3 | 4.808,056 | 427,719 | ,000[a] |
| | Residual | 1.281,492 | 114 | 11,241 | | |
| | Total | 15.705,661 | 117 | | | |

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 14.424,169 | 3 | 4.808,056 | 427,719 | ,000[a] |
| | Residual | 1.281,492 | 114 | 11,241 | | |
| | Total | 15.705,661 | 117 | | | |

a. Predictors: (Constant), Fuel efficiency, Price in thousands, Horsepower

b. Dependent Variable: 4-year resale value

**Πίνακας 6.16** *Αποτελέσματα συντελεστών παλινδρόμησης του αρχείου car sales.sav.*

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | -3,275 | 3,076 | | -1,065 | ,289 |
| | Price in thousands | ,887 | ,042 | 1,087 | 20,973 | ,000 |
| | Horsepower | -,026 | ,011 | -,130 | -2,300 | ,023 |
| | Fuel efficiency | ,118 | ,088 | ,045 | 1,343 | ,182 |

a. Dependent Variable: 4-year resale value

Η τιμή του $R^2$ = 0,918 είναι πολύ καλή και δείχνει ότι το μοντέλο που εκτιμήθηκε ερμηνεύει το 91,8% των παρατηρήσεων. Στον Πίνακα 6.15 ANOVA είναι sig = 0.00. Αυτό σημαίνει ότι συνολικά δεν μπορεί να απορριφθεί το μοντέλο παλινδρόμησης, αφού απορρίπτεται η υπόθεση:

$$H0 : \beta 1 = \beta 2 = ... = \beta k = 0$$

Από τον Πίνακα 6.16 των αποτελεσμάτων (coefficients), το εκτιμημένο μοντέλο θα είναι:

$$Y = -3,275 +0,887X1 -0,026X2 +0,118X3$$

Επειδή στη γραμμή του σταθερού όρου το sig = 0,289 η υπόθεση H0: β0 = 0 δεν μπορεί να απορριφθεί και επομένως ο σταθερός όρος είναι 0 και δεν χρειάζεται στο μοντέλο. Επίσης, στη γραμμή του συντελεστή της X3, το sig = 0,182 η υπόθεση H0: β3 = 0 δεν μπορεί να απορριφθεί, και επομένως ο β3 = 0 και δεν χρειάζεται στο μοντέλο η μεταβλητή X3.

Εκτιμάται ξανά το μοντέλο παλινδρόμησης, αυτή τη φορά χωρίς τη μεταβλητή X3 και χωρίς σταθερό όρο. Στο πλήκτρο options του μενού παλινδρόμησης του SPSS, απενεργοποιούμε την επιλογή “Include constant in equation”, ώστε να μην έχουμε εκτίμηση σταθερού όρου που υπάρχει σαν προεπιλογή στο SPSS.

**Πίνακας 6.17** *Αποτελέσματα συντελεστή προσδιορισμού παλινδρόμησης του αρχείου car sales.sav.*

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | ,988[a] | ,976 | ,976 | 3,344761 |

a. Predictors: Horsepower, Price in thousands

**Πίνακας 6.18** *Αποτελέσματα συντελεστών παλινδρόμησης του αρχείου car sales.sav.*

| Model | | Unstandardized Coefficients B | Unstandardized Coefficients Std. Error | Standardized Coefficients Beta | t | Sig. |
|---|---|---|---|---|---|---|
| 1 | Price in thousands | ,884 | ,038 | 1,224 | 23,342 | ,000 |
| | Horsepower | -,028 | ,006 | -,248 | -4,725 | ,000 |

a. Dependent Variable: 4-year resale value

b. Linear Regression through the Origin

Τα αποτελέσματα της νέας παλινδρόμησης (χωρίς τη μεταβλητή X3 και χωρίς σταθερό όρο) εμφανίζονται στους Πίνακες 6.17 και 6.18. Το νέο $R^2 = 0,976$ είναι πολύ καλό και τα sig των αντίστοιχων δύο μεταβλητών μάς οδηγούν στην απόρριψη των υποθέσεων β1 = 0 και β2 = 0. Το νέο εκτιμημένο μοντέλο είναι:

$$Y = 0,884X1 -0,028X2$$

Στη συνέχεια, θα ελεγχθεί αν ισχύουν οι προϋποθέσεις εφαρμογής της παλινδρόμησης. Στο μενού παλινδρόμησης του SPSS ξαναγίνεται η εκτίμηση, ώστε να υπολογιστούν και τα νέα στοιχεία για τον έλεγχο των προϋποθέσεων. Αυτή τη φορά το πλήκτρο Statistics στο παράθυρο linear regression επιλέγεται Estimates, Model fit, Collinearity diagnostics και Casewise diagnostics/Outliers outside 2 standard deviations. Στην επιλογή plots, επιλέγεται ZPRED, ZRESID για X και Y. Επίσης, επιλέγονται οι επιλογές histogram, produce all partial plots. Τα αποτελέσματα της παλινδρόμησης με τους διαγνωστικούς ελέγχους εμφανίζονται στους Πίνακες 6.19, 6.20 και 6.21.

**Πίνακας 6.19** *Αποτελέσματα ελέγχων νέας παλινδρόμησης του αρχείου car sales.sav.*

| Model | Dimension | Eigenvalue | Condition Index | Variance Proportions: Price in thousands | Variance Proportions: Horsepower |
|---|---|---|---|---|---|
| 1 | 1 | 1,962 | 1,000 | ,02 | ,02 |
| | 2 | ,038 | 7,184 | ,98 | ,98 |

a. Dependent Variable: 4-year resale value

b. Linear Regression through the Origin

**Πίνακας 6.20** *Αποτελέσματα ελέγχων νέας παλινδρόμησης του αρχείου car sales.sav.*

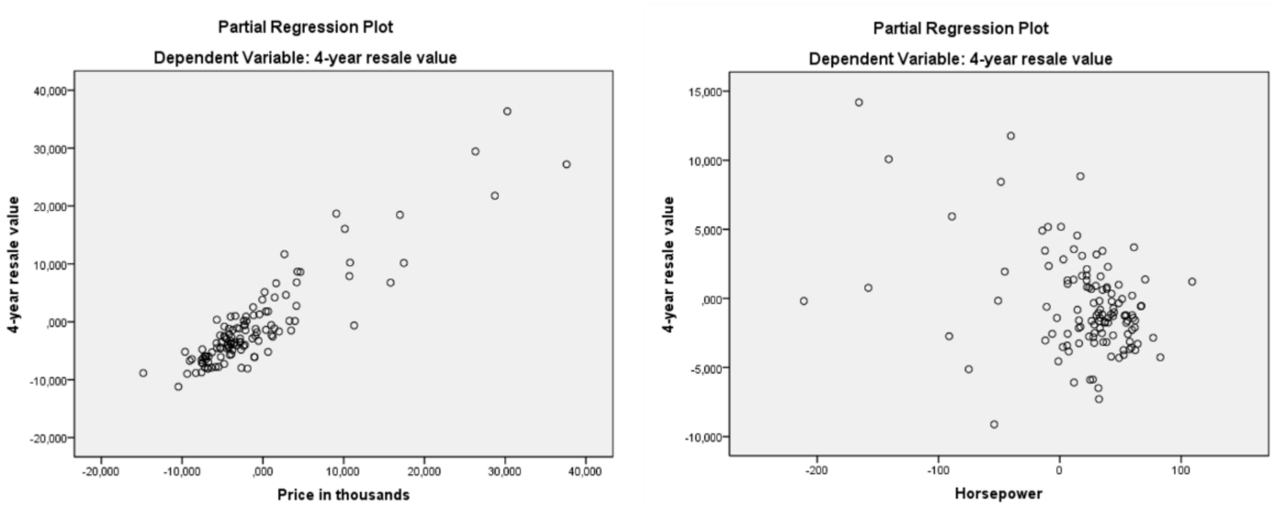
| Model | | Unstandardized Coefficients B | Unstandardized Coefficients Std. Error | Standardized Coefficients Beta | t | Sig. | Collinearity Statistics Tolerance | Collinearity Statistics VIF |
|---|---|---|---|---|---|---|---|---|
| 1 | Price in thousands | ,884 | ,038 | 1,224 | 23,342 | ,000 | ,075 | 13,406 |
| | Horsepower | -,028 | ,006 | -,248 | -4,725 | ,000 | ,075 | 13,406 |

a. Dependent Variable: 4-year resale value

b. Linear Regression through the Origin

Ο έλεγχος (Πίνακας 6.20) γίνεται για τον αποκλεισμό της συγγραμμικότητας μεταξύ των ανεξάρτητων μεταβλητών. Από τις τιμές VIF συμπεραίνουμε ότι υπάρχει πρόβλημα συγγραμμικότητας μεταξύ των δύο μεταβλητών, κάτι που ήταν εμφανές και από τον υπολογισμό του συντελεστή συσχέτισης των δύο μεταβλητών που εμφανίζεται στον Πίνακα 6.11.

Στο αριστερό διάγραμμα διασποράς της Εικόνας 6.11 παρατηρούμε μια γραμμική σχέση με αύξουσα τάση μεταξύ της εξαρτημένης μεταβλητής «τιμή μεταπώλησης σε 4 έτη» και της ανεξάρτητης μεταβλητής «τιμή». Στο δεξί διάγραμμα διασποράς της Εικόνας 6.11 παρατηρούμε μία (όχι άμεσα εμφανή) γραμμική σχέση με φθίνουσα τάση μεταξύ της εξαρτημένης μεταβλητής «τιμή μεταπώλησης σε 4 έτη» και της ανεξάρτητης μεταβλητής «ιπποδύναμη». Για τους λόγους αυτούς, στην εκτίμηση του μοντέλου ο συντελεστής της πρώτης μεταβλητής είναι θετικός και της δεύτερης αρνητικός.




**Εικόνα 6.11** *Διαγράμματα διασποράς για τις δύο μεταβλητές και την εξαρτημένη μεταβλητή της νέας παλινδρόμησης του αρχείου car sales.sav.*

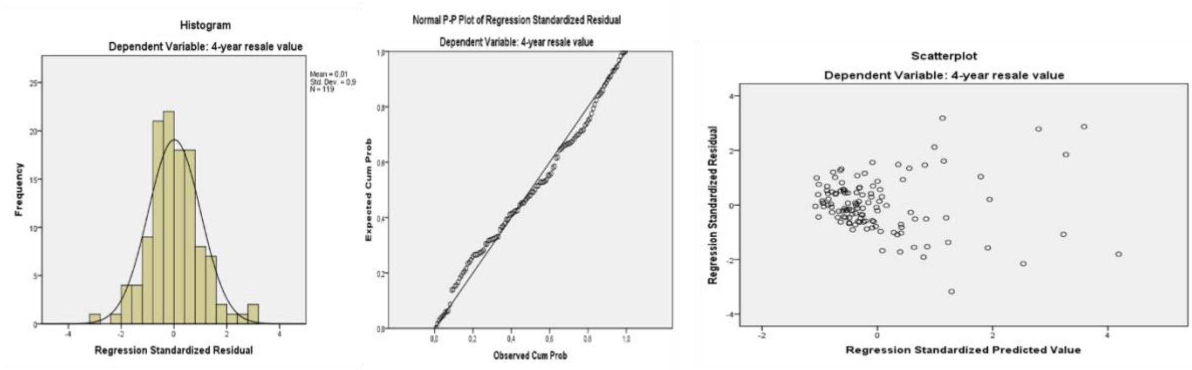
Στο ιστόγραμμα υπολοίπων της Εικόνας 6.12, διακρίνουμε την κανονική κατανομή των υπολοίπων, όπως και στο διάγραμμα p-plot. Στο διάγραμμα διασποράς των τυποποιημένων υπολοίπων (δεξί διάγραμμα της Εικόνας 6.12) σε σχέση με τις εκτιμημένες τιμές της εξαρτημένης μεταβλητής δεν διακρίνεται κάποιο σχήμα, αλλά διάσπαρτα τυχαία τα σημεία, με εξαίρεση ίσως στο δεξί πάνω και κάτω μέρος του διαγράμματος όπου εντοπίζονται κάποιες μικρές ομαδοποιήσεις τιμών. Αυτό μας οδηγεί στο συμπέρασμα ότι τα υπόλοιπα έχουν την ίδια διασπορά (ισχύει η ομοσκεδαστικότητα).

Στη συνέχεια, για το ίδιο αρχείο δεδομένων γίνεται προσπάθεια για την εκτίμηση της μεταβλητής “sales in thousands” ως προς τις ανεξάρτητες μεταβλητές “Vehicle type”, “Price in thousands”, “Engine size”, “Horsepower”, “Wheelbase”, “Width”, “Length”, “Curb weight”, “Fuel capacity”, “Fuel efficiency”, χρησιμοποιώντας το γραμμικό μοντέλο. Η τιμή του συντελεστή προσδιορισμού $R^2$ αυτή τη φορά είναι μόνο 0,339 και το μοντέλο δεν είναι ικανοποιητικό στην εκτίμηση.

Μετασχηματίζεται η εξαρτημένη μεταβλητή “sales in thousands”, δημιουργώντας μια νέα μεταβλητή όπου σε κάθε τιμή αντιστοιχεί ο λογάριθμός της και εκτιμάται εκ νέου το γραμμικό μοντέλο. Η τιμή του συντελεστή προσδιορισμού $R^2$ αυτή τη φορά είναι 0,486 και οι εκτιμήσεις των συντελεστών των ανεξάρτητων μεταβλητών εμφανίζονται στον Πίνακα 6.21. Επειδή στις μεταβλητές “Horsepower”, “Width”, “Length”, “Curb weight”, “Fuel capacity”, το sig είναι μεγαλύτερο από 0,05, γίνεται δεκτή η υπόθεση (H0: ο αντίστοιχος συντελεστής =0) και επομένως η αντίστοιχη μεταβλητή δεν χρειάζεται στο μοντέλο και μπορεί να αφαιρεθεί από αυτό.

Στις μεταβλητές “Engine size” και “Wheelbase” το sig είναι μεγαλύτερο από 0,05 αλλά μικρότερο από 0,10. Επειδή η σημαντικότητα του συντελεστή απορρίπτεται με σφάλμα 5% αλλά γίνεται δεκτή με σφάλμα 10%, αυτές οι μεταβλητές προσωρινά παραμένουν στο μοντέλο, ώστε η σημαντικότητά τους να εξεταστεί και στο νέο μοντέλο.

Εκτιμάται εκ νέου η μεταβλητή “λογάριθμος Sales in thousands” με τις σημαντικές μεταβλητές και τα αποτελέσματα εμφανίζονται στις Εικόνες 6.21 και 6.22. Στη νέα εκτίμηση του μοντέλου (Εικόνα 6.21), όλες οι μεταβλητές είναι σημαντικές, αφού για όλες το sig <0,05. Επιπλέον από τις τιμές VIF δεν φαίνεται να υπάρχει πρόβλημα πολυσυγγραμμικότητας για καμία μεταβλητή.




**Εικόνα 6.12** *Αποτελέσματα υπολοίπων νέας παλινδρόμησης του αρχείου car sales.sav.*

**Πίνακας 6.21** *Αποτελέσματα συντελεστών νέας παλινδρόμησης με λογάριθμο της εξαρτημένης μεταβλητής του αρχείου car sales.sav.*

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | -3,017 | 2,741 | | -1,101 | ,273 |
| | Vehicle type | ,883 | ,331 | ,293 | 2,670 | ,008 |
| | Price in thousands | -,046 | ,013 | -,502 | -3,596 | ,000 |
| | Engine size | ,356 | ,190 | ,281 | 1,871 | ,063 |
| | Horsepower | -,002 | ,004 | -,092 | -,509 | ,611 |
| | Wheelbase | ,042 | ,023 | ,241 | 1,785 | ,076 |
| | Width | -,028 | ,042 | -,073 | -,676 | ,500 |
| | Length | ,015 | ,014 | ,148 | 1,032 | ,304 |
| | Curb weight | ,156 | ,350 | ,075 | ,447 | ,655 |
| | Fuel capacity | -,057 | ,047 | -,167 | -1,203 | ,231 |
| | Fuel efficiency | ,081 | ,040 | ,262 | 2,023 | ,045 |

**Πίνακας 6.22** *Αποτελέσματα συντελεστή προσδιορισμού εκ νέου παλινδρόμησης με λογάριθμο της εξαρτημένης μεταβλητής του αρχείου car sales.sav.*

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | ,689[a] | ,475 | ,457 | ,97999 |

a. Predictors: (Constant), Vehicle type, Price in thousands, Wheelbase, Engine size, Fuel efficiency

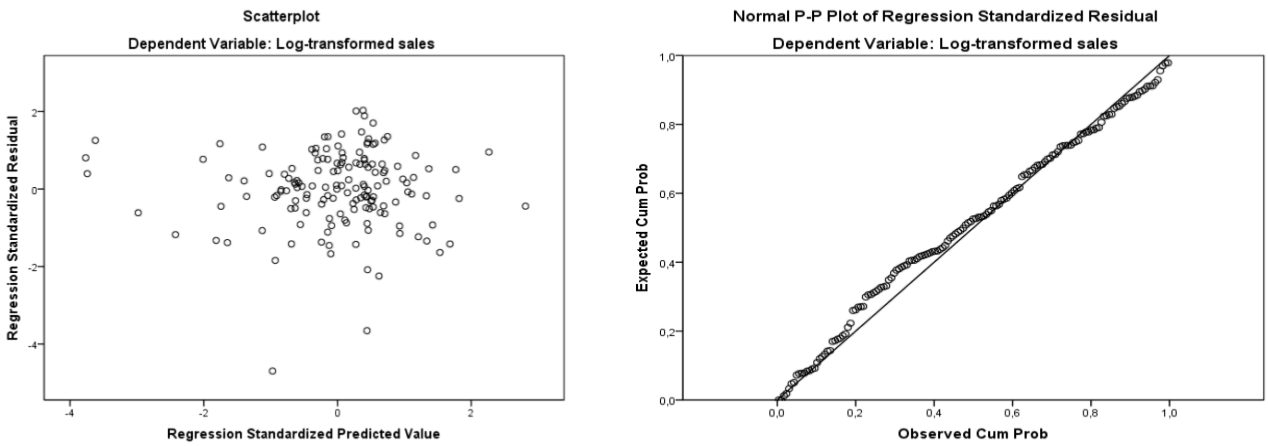
b. Dependent Variable: Log-transformed sales

**Πίνακας 6.23** *Αποτελέσματα σημαντικών συντελεστών εκ νέου παλινδρόμησης με λογάριθμο της εξαρτημένης μεταβλητής του αρχείου car sales.sav.*

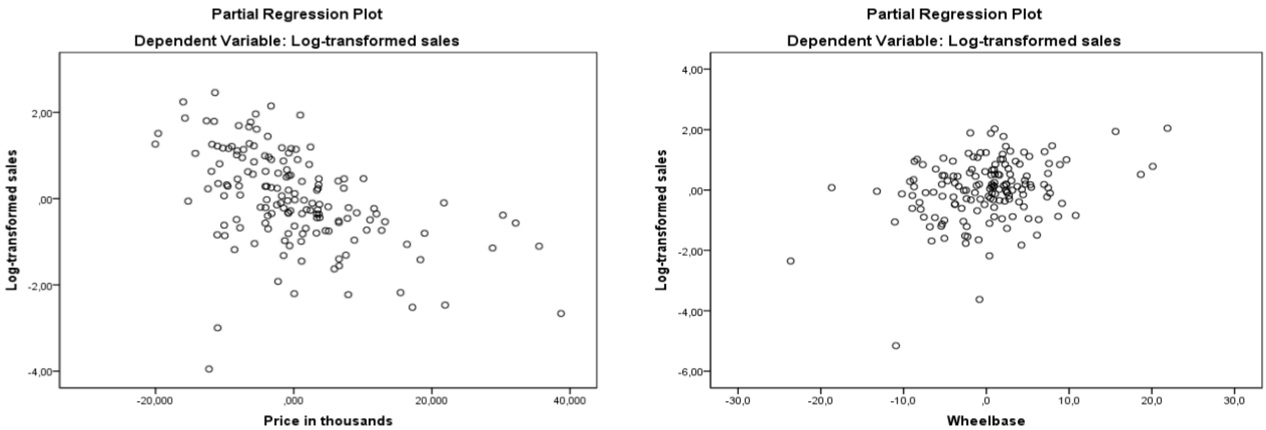
| Model | | Unstandardized Coefficients B | Unstandardized Coefficients Std. Error | Standardized Coefficients Beta | t | Sig. | Collinearity Statistics Tolerance | Collinearity Statistics VIF |
|---|---|---|---|---|---|---|---|---|
| 1 | (Constant) | -4,088 | 1,841 | | -2,220 | ,028 | | |
| | Price in thousands | -,053 | ,008 | -,579 | -6,741 | ,000 | ,485 | 2,060 |
| | Wheelbase | ,051 | ,013 | ,294 | 3,978 | ,000 | ,654 | 1,529 |
| | Fuel efficiency | ,094 | ,036 | ,304 | 2,645 | ,009 | ,271 | 3,686 |
| | Engine size | ,310 | ,132 | ,246 | 2,349 | ,020 | ,327 | 3,060 |
| | Vehicle type | ,732 | ,251 | ,243 | 2,913 | ,004 | ,514 | 1,944 |

a. Dependent Variable: Log-transformed sales

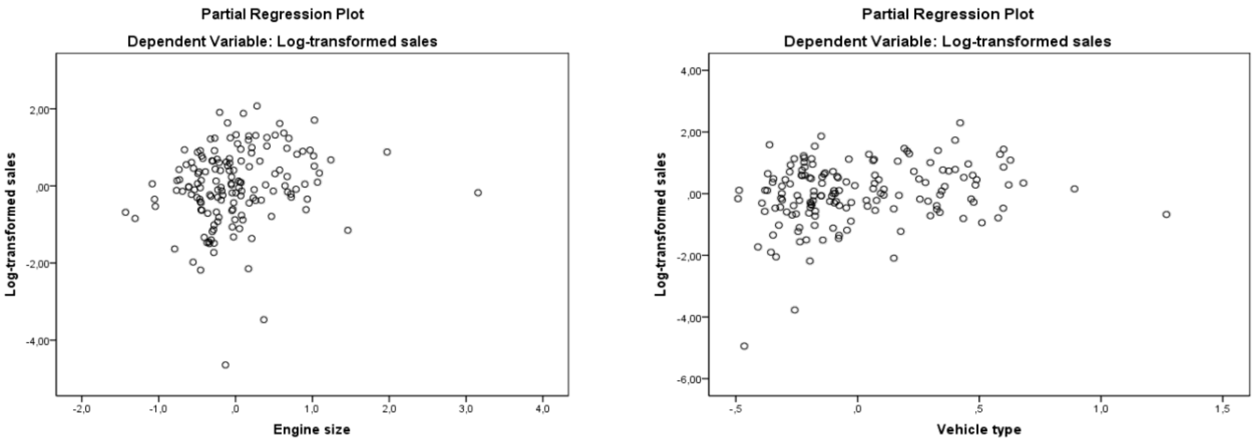
Τα διαγράμματα διασποράς υπολοίπων και ανεξάρτητων μεταβλητών για τη νέα εκτίμηση του μοντέλου εμφανίζονται στις Εικόνες 6.13, 6.14 και 6.15. Παρατηρείται σχεδόν κανονική κατανομή των υπολοίπων (Εικόνα 6.13). Επίσης, στα διαγράμματα διασποράς των σημείων της εξαρτημένης μεταβλητής με καθεμία από τις ανεξάρτητες μεταβλητές (Εικόνες 6.14 και 6.15) παρατηρείται η θετική ή αρνητική τάση των σημείων, σε συμφωνία και με το πρόσημο της εκτίμησης των αντίστοιχων συντελεστών του εκτιμημένου μοντέλου.




**Εικόνα 6.13** *Διαγράμματα υπολοίπων της εκ νέου παλινδρόμησης με λογάριθμο της εξαρτημένης μεταβλητής του αρχείου car sales.sav.*




**Εικόνα 6.14** *Διαγράμματα υπολοίπων της εκ νέου παλινδρόμησης με λογάριθμο της εξαρτημένης μεταβλητής του αρχείου car sales.sav, ως προς την πρώτη και δεύτερη ανεξάρτητη μεταβλητή.*




**Εικόνα 6.15** *Διαγράμματα υπολοίπων της εκ νέου παλινδρόμησης με λογάριθμο της εξαρτημένης μεταβλητής του αρχείου car sales.sav, ως προς την τρίτη και τέταρτη ανεξάρτητη μεταβλητή.*

## 6.9 Ασκήσεις για κατανόηση - εφαρμογή

### 6.9.1 Ερμηνεία αποτελεσμάτων παλινδρόμησης

Εφαρμόζοντας τη μέθοδο της παλινδρόμησης με το πρόγραμμα SPSS, προέκυψε το παρακάτω output.

**Coefficients[a]**

| Model | Unstandardized Coefficients B | Unstandardized Coefficients Std. Error | Standardized Coefficients Beta | t | Sig. |
|---|---|---|---|---|---|
| (Constant) | 23.844,424 | 3.712,367 | | 6,423 | ,000 |
| antikeimeniki axia oikopedou | 3,160 | ,090 | ,746 | 35,211 | ,000 |
| antikeimeniki axia ktismatos | ,170 | ,049 | ,074 | 3,477 | ,001 |

a. Dependent Variable: timi polisis

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | ,745[a] | ,555 | ,554 | 37.373,966 |

α) Ποιες είναι οι εκτιμήσεις των συντελεστών του μοντέλου παλινδρόμησης; Γράψτε το μοντέλο που προκύπτει.

β) Μπορούμε να θεωρήσουμε ότι κάποιος συντελεστής είναι μηδέν;

γ) Πόσο αξιόπιστο είναι το μοντέλο της παλινδρόμησης; Τι θα προτείνατε να γίνει ώστε να βελτιωθεί η εκτίμηση του μοντέλου της παλινδρόμησης;

Λύση

α) Οι εκτιμήσεις των συντελεστών του μοντέλου παλινδρόμησης εμφανίζονται στη στήλη B του πίνακα Coefficients και είναι 23.844,424 για τον σταθερό όρο, 3.160 για την πρώτη μεταβλητή και 0,170 για τη δεύτερη ανεξάρτητη μεταβλητή. Το εκτιμημένο μοντέλο είναι:

τιμή πώλησης = 23.844,424 + 3.160 × Αντικειμενική αξία οικοπέδου + 0,170 × Αντικειμενική αξία κτιρίου

β) Για να ελεγχθεί αν κάποιος συντελεστής μπορεί να θεωρηθεί μηδέν, γίνεται έλεγχος με το κριτήριο t-test. Στην τελευταία στήλη του πίνακα Coefficients εμφανίζεται η πιθανότητα λάθους, αν απορριφθεί η μηδενική υπόθεση H0 ότι ο αντίστοιχος συντελεστής είναι 0. Επειδή όλες οι πιθανότητες είναι μικρότερες από 0,05, η κάθε υπόθεση H0 απορρίπτεται, και επομένως κάθε συντελεστής είναι σημαντικός στο μοντέλο και δεν μπορεί να θεωρηθεί 0.

γ) Η αξιοπιστία του μοντέλου μετριέται με τον συντελεστή προσδιορισμού $R^2$ (R square). Η τιμή του είναι 0,555. Αυτό αντιστοιχεί σε ποσοστό 55,5% των δεδομένων που ερμηνεύεται από το μοντέλο παλινδρόμησης. Για να βελτιωθεί το μοντέλο, θα πρέπει να εισαχθούν σε αυτό και άλλες ανεξάρτητες μεταβλητές, ώστε να μεγαλώσει η τιμή του $R^2$.

## 6.9.2 Ερμηνεία αποτελεσμάτων παλινδρόμησης

Εφαρμόζοντας τη μέθοδο της παλινδρόμησης με το πρόγραμμα SPSS προέκυψε το παρακάτω output. Είναι καλή η εκτίμηση; Ποιες είναι οι εκτιμήσεις των συντελεστών; Μπορούμε να θεωρήσουμε ότι κάποιος συντελεστής είναι μηδέν; Γράψτε την εξίσωση του μοντέλου παλινδρόμησης.

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | ,376[a] | ,141 | ,138 | 15,12892 |

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 69,921 | 2,004 | | 34,890 | ,000 |
| | μηνιαίο εισόδημα | -,206 | ,477 | -,015 | -,433 | ,665 |
| | εκπαίδευση | -1,335 | ,348 | -,127 | -3,839 | ,000 |
| | ηλικία | -3,958 | ,377 | -,353 | -10,509 | ,000 |

a. Dependent Variable: Μόρια βαθμολόγησης

Λύση

Οι εκτιμήσεις των συντελεστών του μοντέλου παλινδρόμησης εμφανίζονται στη στήλη B του πίνακα Coefficients και είναι 69,921 για το σταθερό όρο, -0,206 για την πρώτη μεταβλητή, -1,335 για τη δεύτερη ανεξάρτητη μεταβλητή και -3,958 για την τρίτη ανεξάρτητη μεταβλητή. Το εκτιμημένο μοντέλο είναι:

$$\text{Μόρια βαθμολόγησης} = 69{,}921 - 0{,}206 \times \text{μηνιαίο εισόδημα} - 1{,}335 \times \text{εκπαίδευση} - 3{,}958 \times \text{ηλικία}$$

Για να ελεγχθεί αν κάποιος συντελεστής μπορεί να θεωρηθεί μηδέν, γίνεται έλεγχος με το κριτήριο t-test. Στην τελευταία στήλη του πίνακα Coefficients εμφανίζεται η πιθανότητα λάθους, αν απορριφθεί η μηδενική υπόθεση H0 ότι ο αντίστοιχος συντελεστής είναι 0. Όλες οι πιθανότητες είναι μικρότερες από 0,05, εκτός από την τιμή για το «μηνιαίο εισόδημα» που είναι 0,665. Η υπόθεση H0 για τη μεταβλητή «μηνιαίο εισόδημα» δεν μπορεί να απορριφθεί, και επομένως προκύπτει ως συμπέρασμα ότι ο συντελεστής της είναι μηδέν και δεν είναι σημαντικός στο μοντέλο. Δηλαδή η μεταβλητή μηνιαίο εισόδημα δεν εξηγεί τη μεταβλητή «Μόρια βαθμολόγησης» και δεν χρειάζεται στο μοντέλο. Οι υπόλοιπες μεταβλητές είναι χρήσιμες στο μοντέλο.

Η αξιοπιστία του μοντέλου μετριέται με τον συντελεστή προσδιορισμού $R^2$ (R square). Η τιμή του είναι 0,141. Αυτό αντιστοιχεί σε μικρό ποσοστό 14,1% των δεδομένων που ερμηνεύεται από το μοντέλο παλινδρόμησης. Για να βελτιωθεί το μοντέλο, θα πρέπει να εισαχθούν σε αυτό και άλλες ανεξάρτητες μεταβλητές, ώστε να μεγαλώσει η τιμή του $R^2$.

# Βιβλιογραφία/Αναφορές

Κώστογλου, Β., & Αντωνίου, Ε. (2021). *Πιθανότητες και Στατιστική*. Θεσσαλονίκη: Τζιόλα.

Berenson, M. L, Levine, D. M., Szabat, K. A. ( 2018). *Βασικές Αρχές Στατιστικής*. Κύπρος: Broken Hill.

Ζαφειρόπουλος, Κ., (2022). *Εφαρμογές Ανάλυσης Παλινδρόμησης*. Θεσσαλονίκη: Τζιόλα.

Γναρδέλλης, Χ., (2019). *Εφαρμοσμένη Στατιστική*. Αθήνα: Παπαζήση.

# Κριτήρια αξιολόγησης

## Κριτήριο αξιολόγησης 1

**Σε τι βοηθά η τιμή του $R^2$ σε μία πολλαπλή παλινδρόμηση;**

**Απάντηση**
Η τιμή του $R^2$ αξιολογεί πόσο καλή είναι η παλινδρόμηση. Εκφράζει το ποσοστό των δεδομένων που ερμηνεύεται από το μοντέλο της παλινδρόμησης.

## Κριτήριο αξιολόγησης 2

**Πώς θα μετασχηματιστεί το μοντέλο $Y = αX+βX^2+γ$, ώστε να εκτιμηθούν οι συντελεστές α και β;**

**Απάντηση**
Το μοντέλο αυτό δεν είναι γραμμικό, αφού η τιμή της X είναι υψωμένη στο τετράγωνο. Για να εκτιμηθεί με την παλινδρόμηση, πρέπει να υπάρχουν μόνο γραμμικές σχέσεις μεταξύ των μεταβλητών. Θα δημιουργηθεί μια νέα μεταβλητή $Z = X^2$ και στη συνέχεια θα εκτιμηθεί το γραμμικό μοντέλο παλινδρόμησης $Y = αX+βZ+γ$.

Οι τιμές των συντελεστών α, β, γ που θα εκτιμηθούν από το νέο μοντέλο θα προσαρμόζονται και στο αρχικό μοντέλο.

## Κριτήριο αξιολόγησης 3

**Πότε υπάρχει το πρόβλημα πολυσυγγραμμικότητας σε μια παλινδρόμηση;**

**Απάντηση**
Το πρόβλημα πολυσυγγραμμικότητας εμφανίζεται όταν υπάρχει γραμμική συσχέτιση μεταξύ των ανεξάρτητων μεταβλητών σε ένα μοντέλο παλινδρόμησης.

# Κεφάλαιο 7: Πολυμεταβλητή Παραγοντική Ανάλυση

**Σύνοψη**
*Στο κεφάλαιο αυτό γίνεται μια εισαγωγή στην πολυμεταβλητή ανάλυση στατιστικών μεταβλητών, κυρίως με τη μεθοδολογία της παραγοντικής ανάλυσης. Παρουσιάζονται παραδείγματα πραγματικών δεδομένων για την καλύτερη ερμηνεία τους και εύρεσης των μεταξύ τους αλληλοεξαρτήσεων.*

**Προαπαιτούμενη γνώση**
*Δεν απαιτείται.*

## 7.1 Εισαγωγή στην πολυμεταβλητή ανάλυση

Η **πολυμεταβλητή ανάλυση** είναι ένα σύνολο μεθόδων με στόχο την ταυτόχρονη ανάλυση και επεξεργασία πολλών μεταβλητών μαζί, ώστε να αναδειχθούν οι κυρίαρχες τάσεις. Στόχος είναι η ανάδειξη των κύριων τάσεων των μεταβλητών, με την ταυτόχρονη διατήρηση της πληροφορίας και τη μείωση των διαστάσεων του συνόλου των δεδομένων.

Εφαρμόζεται στην ερμηνεία κάποιας ή κάποιων μεταβλητών, χρησιμοποιώντας μεγάλο αριθμό άλλων επεξηγηματικών μεταβλητών. Η πολλαπλή παλινδρόμηση είναι μια περίπτωση πολυμεταβλητής ανάλυσης, αφού γίνεται προσπάθεια μία μεταβλητή (εξαρτημένη) να ερμηνευτεί από άλλες (ανεξάρτητες). Ο ερευνητής, όμως, ορίζει την εξαρτημένη μεταβλητή. Γενικότερα, στην πολυμεταβλητή ανάλυση ερευνώνται και αποτυπώνονται οι κυρίαρχες τάσεις που προκύπτουν από τα δεδομένα, χωρίς κάποια εκ των προτέρων υπόθεση από τον ερευνητή. Βρίσκει, κυρίως, εφαρμογή σε περιπτώσεις όπως η κατανόηση χαρακτηριστικών ή τάσεων ενός φαινομένου με έρευνα με τη χρήση ερωτηματολογίου με πολλές ερωτήσεις.

### 7.1.1 Αιτιότητα και συνάφεια

Στο σημείο αυτό κρίνεται σκόπιμο να γίνει η διάκριση των εννοιών «αιτιότητα» και «συνάφεια». **Συνάφεια** μεταξύ δύο μεταβλητών δηλώνεται με την ύπαρξη μιας σχέσης μεταξύ τους. Δηλαδή η μία προκαλεί την άλλη και αντίστροφα. Όταν αλλάζει η τιμή της μίας μεταβλητής, αλλάζει και η τιμή της άλλης και αντίστροφα. Για παράδειγμα, σε μελέτη σε σχολεία πόλεων βρέθηκε εντονότερο το φαινόμενο του εκφοβισμού από ό,τι σε σχολεία επαρχίας. Υπάρχει συνάφεια μεταξύ του είδους σχολείου και της συχνότητας εκφοβισμού.

Η ύπαρξη συνάφειας όμως δεν σημαίνει και αιτιώδη σχέση (Agresti, 2021). Δηλαδή, στο παράδειγμα, δεν σημαίνει ότι σε κάθε σχολείο πόλης θα υπάρχει εκφοβισμός μεταξύ των μαθητών. Άλλο παράδειγμα συνάφειας που μπορεί να αναφερθεί είναι ότι η προσεκτική μελέτη ενός φοιτητή προκαλεί επιτυχία στις εξετάσεις και η επιτυχία στις εξετάσεις εξαρτάται από την προσεκτική του μελέτη. Στην περίπτωση αυτή η σχέση είναι αιτιώδης, από την προσεκτική μελέτη προς την επιτυχία στις εξετάσεις. Αν κάποιος φοβάται να κυκλοφορήσει τη νύχτα, μένει στο σπίτι τις βραδινές ώρες. Είναι μία αιτιώδης σχέση, αλλά δεν ισχύει το αντίστροφο, κάποιος που μένει στο σπίτι τις βραδινές ώρες δεν είναι επειδή φοβάται να κυκλοφορήσει τη νύχτα. Για να θεωρηθεί μία σχέση μεταξύ μεταβλητών αιτιώδης, θα πρέπει να υπάρχει συνάφεια μεταξύ των μεταβλητών, να υπάρχει χρονική αλληλουχία στην παρατήρησή τους, και να μην υπάρχουν άλλες εναλλακτικές εξηγήσεις για τη σχέση τους.

Χρειάζεται ιδιαίτερη προσοχή στις ερμηνείες των σχέσεων συνάφειας μεταξύ δύο μεταβλητών. Γενικά, δεν πρέπει να συμπεραίνεται αιτιώδης σχέση χωρίς ανάλυση, γιατί αυτό οδηγεί πολλές φορές σε παρερμηνείες. Παρατηρώντας το βάρος φοιτητών και την επίδοσή τους σε κάποιο μάθημα, μπορεί να υπάρχει μεγάλη συσχέτιση. Αυτό όμως δεν πρέπει να μας οδηγήσει στο συμπέρασμα ότι αν ένας φοιτητής βάλει κιλά θα έχει καλύτερη επίδοση! Η σχέση μεταξύ βάρους και επίδοσης ίσως να προέκυψε από κάποια άλλη μεταβλητή, όπως η πολύωρη μελέτη, η οποία προκάλεσε αύξηση βάρους και ταυτόχρονα καλύτερη επίδοση.

Στις περιπτώσεις αυτές απαιτείται συνολική ανάλυση όλων των μεταβλητών του φαινομένου και όχι μόνο ανάλυση ανά δύο μεταβλητών. Με την παραγοντική ανάλυση πραγματοποιείται η εύρεση όλων των συσχετιζόμενων μεταβλητών, χωρίς όμως να διευκρινίζονται σχέσεις αιτιώδεις ή μη αιτιώδεις.

### 7.1.2 Μείωση διαστάσεων

Έχοντας για παράδειγμα, 30 μεταβλητές για ένα φαινόμενο που μελετάται και προσπαθώντας να υπολογίζουμε τη μεταξύ τους σχέση, θα πρέπει να υπολογιστούν και να ερμηνευτούν 435 συντελεστές συσχέτισης ανά δύο των μεταβλητών. Για την πρώτη από τις 30 μεταβλητές υπολογίζεται συσχέτιση με τις υπόλοιπες 29 μεταβλητές. Για την επόμενη μεταβλητή με τις υπόλοιπες 28, για τη μεθεπόμενη με τις υπόλοιπες 27 και, τέλος, για την προτελευταία υπολογίζεται η συσχέτισή της με την τελευταία μεταβλητή. Δηλαδή 30.29.28.27.26….3.2 = 435 συσχετίσεις.

Όλες αυτές οι ανά δύο συσχετίσεις δεν είναι δύσκολο να υπολογιστούν με έναν ηλεκτρονικό υπολογιστή. Είναι όμως περίπλοκο να βρεθούν οι κυρίαρχες και οι σχέσεις μεταξύ περισσότερων από δύο μεταβλητών, μελετώντας μόνο τις συσχετίσεις ανά δύο μεταβλητών. Ο ερευνητής προσπαθεί να μειώσει τον αριθμό των μεταβλητών, μετά από τη συλλογή των δεδομένων, ομαδοποιώντας όσες συμπεριφέρονται παρόμοια. Πάντα όμως με τη φροντίδα να αναδειχθούν οι κύριες τάσεις των δεδομένων, δηλαδή να μειωθούν οι διαστάσεις του φαινομένου, χωρίς να χαθεί η πληροφορία του.

Μία από τις μεθόδους μείωσης των διαστάσεων είναι η παραγοντική ανάλυση που θα παρουσιαστεί στη συνέχεια.

### 7.1.3 Παραγοντική ανάλυση (Factor analysis)

Η **παραγοντική ανάλυση** είναι μια πολυμεταβλητή τεχνική, η οποία προσπαθεί να ερμηνεύσει συσχετίσεις μεταξύ παρατηρούμενων μεταβλητών, υποθέτοντας ότι αυτές οφείλονται στην ύπαρξη κάποιων κοινών παραγόντων, οι οποίοι δεν φαίνονται άμεσα ούτε μπορούν να μετρηθούν άμεσα. Ως αποτέλεσμα η εφαρμογή της μεθόδου πετυχαίνει τη μείωση των αρχικών μεταβλητών και την αντικατάστασή τους με νέες που ονομάζονται **παράγοντες**. Οι παράγοντες αναδεικνύονται μέσα από τα δεδομένα, χωρίς αρχικές υποθέσεις, γίνονται «μετρήσιμοι» και είναι αυτοί που περιγράφουν πλήρως το μελετώμενο φαινόμενο.

Η παραγοντική ανάλυση, εξαιτίας τις ικανότητάς της να εντοπίζει τους «αφανείς» παράγοντες, βρήκε ευρεία εφαρμογή κυρίως στις κοινωνικές επιστήμες, στην οικονομία, στις έρευνες αγοράς. Χρησιμοποιείται για μεγάλη ποικιλία σκοπών (Agresti, 2021), όπως:

- αποκάλυψη μοτίβων σχέσεων μεταξύ των μεταβλητών,
- αποκάλυψη συστάδων μεταβλητών, που καθεμία περιέχει αλληλοσχετιζόμενες μεταβλητές, οπότε αρκεί η μία για να περιγράψει τη συστάδα,
- μείωση μεγάλου αριθμού μεταβλητών σε μικρότερο αριθμό μη συσχετιζόμενων μεταβλητών που ονομάζονται παράγοντες (factors).

Στις κοινωνικές επιστήμες δεν μπορούν να μετρηθούν άμεσα κάποιες μεταβλητές που ενδιαφέρουν την έρευνα, είτε γιατί δεν είναι παρατηρήσιμες είτε γιατί είναι πολυδιάστατες. Οι μεταβλητές αυτές ονομάζονται λανθάνουσες μεταβλητές (latent variables). Παραδείγματα λανθανουσών μεταβλητών είναι η ευφυΐα ενός ατόμου, η κοινωνική του κατάσταση, οι απόψεις του για ένα συγκεκριμένο θέμα (Bartholomew et al., 2011).

Για την εύρεση των κύριων παραγόντων αναζητούνται γραμμικοί συνδυασμοί των μεταβλητών, έτσι ώστε το αποτέλεσμα του γραμμικού συνδυασμού να αποτελεί τον παράγοντα με τη μέγιστη δυνατή διασπορά. Στη συνέχεια, αναζητείται νέος γραμμικός συνδυασμός ο οποίος να είναι ασυσχέτιστος με τον προηγούμενο και να ερμηνεύει την επόμενη μέγιστη δυνατή διασπορά των δεδομένων.

Η διαδικασία επαναλαμβάνεται έως ότου ερμηνευτεί όλη η διασπορά των δεδομένων μέσω των ασυσχέτιστων παραγόντων, οι οποίοι σε πλήθος μπορεί να είναι όσες οι αρχικές μεταβλητές. Οι πρώτοι παράγοντες συνήθως ερμηνεύουν μεγάλο ποσοστό της διασποράς και χρησιμοποιούνται μόνο αυτοί αντί όλων των μεταβλητών. Έτσι, επιτυγχάνεται η μείωση των διαστάσεων και η περιγραφή των δεδομένων με τους κύριους παράγοντες, που είναι πολύ λιγότεροι σε πλήθος σε σχέση με τις αρχικές μεταβλητές. Ένα ακόμη πλεονέκτημα των κυρίων παραγόντων είναι ότι είναι ποσοτικές μεταβλητές που μπορούν να χρησιμοποιηθούν και σε άλλες στατιστικές μεθόδους για την ανάλυση των δεδομένων, όπως π.χ. σε παλινδρόμηση ή σε ιεραρχική ταξινόμηση.

### 7.1.3.1 Χρησιμότητα

Χρησιμοποιείται η παραγοντική ανάλυση όταν πρόκειται να αναλυθούν πολλές ποιοτικές μεταβλητές, έχοντας διαθέσιμο μεγάλο πλήθος δεδομένων.

Με την παραγοντική ανάλυση επιδιώκεται μία συνολική ανάλυση των μεταβλητών και όχι απλώς μία ανάλυση κάθε μεταβλητής ξεχωριστά ή ανά ζεύγη. Στόχος είναι η αναζήτηση των κυρίαρχων τάσεων των δεδομένων. Κυρίως εφαρμόζεται σε έρευνες με ερωτηματολόγιο, όπου οι απαντήσεις εκφράζονται με την πενταβάθμια κλίμακα Likert. Θεωρούνται ποιοτικές μεταβλητές. Διακρίνουμε τη διερευνητική παραγοντική ανάλυση και την επιβεβαιωτική παραγοντική ανάλυση. Η πρώτη εφαρμόζεται όταν αναζητούνται οι κύριοι παράγοντες ενός φαινομένου, ενώ η δεύτερη όταν στόχος είναι η επιβεβαίωση για τις υποθέσεις των κυρίων παραγόντων.

#### 7.1.3.1.1 Παράδειγμα

Στην πρώτη στήλη του Πίνακα 7.1 παρουσιάζονται οι ερωτήσεις ενός ερωτηματολογίου που αφορά τις διατροφικές συνήθεις ατόμων. Οι απαντήσεις στις ερωτήσεις αυτές δίνονται στην κλίμακα Likert 1 έως 5. Στον παρακάτω Πίνακα 7.1 παρουσιάζονται τα ποσοστά των απαντήσεων σε κάθε κατηγορία ερώτησης.

**Πίνακας 7.1** *Παράδειγμα ερωτηματολογίου με τα ποσοστά απαντήσεων ανά κατηγορία.*

| **Ερώτηση** | **Καθόλου** | **Κάποιες φορές** | **Μέτρια** | **Συχνά** | **Πολύ συχνά** |
|---|---|---|---|---|---|
| Κωδικός | 1 | 2 | 3 | 4 | 5 |
| 1. Προσέχω να είμαι υγιής. | 2,7 | 10,8 | 35,1 | 37,8 | 13,5 |
| 2. Γυμνάζομαι | 24,3 | 35,1 | 29,7 | 8,1 | 2,7 |
| 3. Καπνίζω | 83,8 | 5,4 | 8,1 | 2,7 | ,0 |
| 4. Καταναλώνω την απαραίτητη ποσότητα νερού καθημερινά | 10,8 | 2,7 | 32,4 | 29,7 | 24,3 |
| 5. Εγώ αγοράζω τα τρόφιμά μου | 5,4 | 16,2 | 18,9 | 27,0 | 32,4 |
| 6. Παραλείπω το πρωινό γεύμα | 51,4 | 18,9 | 10,8 | 2,7 | 16,2 |
| 7. Παραλείπω το μεσημεριανό γεύμα | 48,6 | 35,1 | 13,5 | 2,7 | ,0 |
| 8. Παραλείπω το βραδινό γεύμα | 32,4 | 35,1 | 10,8 | 16,2 | 5,4 |
| 9. Λαμβάνω φάρμακα | 54,1 | 21,6 | 2,7 | 5,4 | 16,2 |
| 10. Λαμβάνω βιταμίνες | 56,8 | 24,3 | 2,7 | 8,1 | 8,1 |
| 11. Λαμβάνω συμπληρώματα διατροφής | 81,1 | 2,7 | 13,5 | ,0 | 2,7 |
| 12. Φροντίζω να καταναλώνω τρόφιμα πλούσια σε φυτικές ίνες | 8,1 | 10,8 | 32,4 | 24,3 | 24,3 |
| 13. Φροντίζω να καταναλώνω τρόφιμα με χαμηλά λιπαρά | 13,5 | 32,4 | 27,0 | 21,6 | 5,4 |
| 14. Καταναλώνω τρόφιμα εμπλουτισμένα (ω3, ω6, προβιοτικά, φυτοστερόλες) | 10,8 | 18,9 | 27,0 | 32,4 | 10,8 |
| 15. Καταναλώνω τρόφιμα χωρίς γλουτένη ή χωρίς λακτόζη | 43,2 | 29,7 | 13,5 | 10,8 | 2,7 |
| 16. Καταναλώνω εκχυλίσματα βοτάνων | 40,5 | 16,2 | 27,0 | 10,8 | 5,4 |
| 17. Καταναλώνω δυναμωτικά ποτά | 83,8 | 13,5 | 2,7 | ,0 | 0 |

Μετά από τη συλλογή των απαντήσεων, μπορεί να γίνει η παρουσίαση μίας μίας των ερωτήσεων, καθώς και η παρουσίασή τους ανά δύο. Κυρίως ενδιαφέρει η συνολική ανάλυση των δεδομένων, ώστε να προκύψουν οι ομάδες ατόμων που απαντούν με παρόμοιο τρόπο στις ερωτήσεις, καθώς και οι μεταβλητές που χαρακτηρίζουν τις ομάδες αυτές. Η μέθοδος για τη συνολική ανάλυση των μεταβλητών είναι η παραγοντική ανάλυση. Δεν απαιτεί να ορίσουμε κάποια μεταβλητή ως εξαρτημένη από άλλες μεταβλητές, αλλά γίνεται συνολική ανάλυση όλων των μεταβλητών. Έχει στόχο να εξηγήσει τις συσχετίσεις που υπάρχουν στο σύνολο των μεταβλητών. Αρκετές φορές οι συσχετίσεις αυτές δεν εξηγούν αιτιώδη σχέση μεταξύ μεταβλητών, αλλά οφείλονται στο γεγονός ότι οι συσχετισμένες μεταβλητές έχουν κοινή εξάρτηση από κάποια ή κάποιες άλλες μεταβλητές (Bartholomew et al., 2011, σελ. 258).

Οι μεταβλητές, όμως, που επηρεάζουν τις υπόλοιπες δεν είναι γνωστές ή δεν μπορούν να καταγραφούν. Για τον λόγο αυτό υπολογίζονται οι «παράγοντες» που αντιστοιχούν στις άγνωστες μεταβλητές. Οι παράγοντες εντοπίζονται με τη σειρά σπουδαιότητάς τους (πρώτος, δεύτερος) και είναι ανεξάρτητοι μεταξύ τους. Για κάθε παράγοντα υπολογίζονται οι συσχετίσεις του με τις αρχικές κ.λπ. μεταβλητές. Οι μεταβλητές με μεγάλες συσχετίσεις «ερμηνεύουν» τον συγκεκριμένο παράγοντα.

## 7.2 Προϋποθέσεις παραγοντικής ανάλυσης

Δεν υπάρχουν εκ των προτέρων προϋποθέσεις για να εφαρμοστεί η παραγοντική ανάλυση. Αρκεί να υπάρχει ένα σύνολο δεδομένων με καταγεγραμμένες τις τιμές των εξεταζόμενων μεταβλητών. Το καταλληλότερο μέγεθος δείγματος για μια αξιόπιστη παραγοντική ανάλυση προτείνεται να είναι περίπου 10-20 παρατηρήσεις ανά μεταβλητή. Δηλαδή για 15 μεταβλητές να υπάρχουν τουλάχιστον 150 παρατηρήσεις.

## 7.3 Έλεγχοι καταλληλότητας δεδομένων

Για την εύρεση των κυρίων παραγόντων αναζητείται γραμμικός συνδυασμός συσχετισμένων μεταβλητών και κατόπιν αναζητείται ο επόμενος κύριος παράγοντας ασυσχέτιστος με τους προηγούμενους παράγοντες. Αν όλες οι αρχικές μεταβλητές είναι ασυσχέτιστες, δεν υπάρχει λόγος να προχωρήσει η διαδικασία. Ο έλεγχος για υπόθεση καταλληλότητας των δεδομένων για την παραγοντική ανάλυση, δηλαδή ο έλεγχος των ασυσχέτιστων αρχικών μεταβλητών, γίνεται με το κριτήριο Bartlett's sphericity test.

Η αρχική υπόθεση H0 είναι ότι όλες οι μεταβλητές είναι μεταξύ τους ασυσχέτιστες και η εναλλακτική H1 είναι ότι δεν ισχύει η H0. Αν το αποτέλεσμα (significant level) του ελέγχου αυτού είναι μικρότερο του 0,05, η υπόθεση H0 απορρίπτεται και μπορεί να πραγματοποιηθεί η παραγοντική ανάλυση.

Ένας έλεγχος για την καταλληλότητα των δεδομένων είναι ο δείκτης KMO (Kaiser-Meyer-Olkin), που εκτιμά το μέγεθος της ομοιογένειας των μεταβλητών (Πετρίδης, 2015). Ο δείκτης αυτός παίρνει τιμές από 0 έως 1. Ο δείκτης συγκρίνει το μέγεθος των συντελεστών συσχέτισης σχετικά με τους μερικούς συντελεστές συσχέτισης. Όσο πιο κοντά στη μονάδα είναι η τιμή του, τόσο καταλληλότερα είναι τα δεδομένα για παραγοντική ανάλυση. Καλές τιμές του δείκτη θεωρούνται οι τιμές που είναι μεγαλύτερες από το 0,8 (Καρλής, 2005).

## 7.4 Κριτήρια μέτρησης «αποστάσεων»

Οι μεταβλητές που κυρίως χρησιμοποιούνται στην παραγοντική ανάλυση είναι κατηγορικές ή διατάξιμες μεταβλητές, χωρισμένες σε κλάσεις. Για τις μεταβλητές αυτές δημιουργούνται πίνακες διπλής εισόδου (πίνακες συμπτώσεων) με το πλήθος (συχνότητα) των δεδομένων που ανήκουν ταυτόχρονα και στις δύο κλάσεις της μιας και της άλλης μεταβλητής. Βάζοντας στη σειρά (δίπλα δίπλα και έναν κάτω από τον άλλο) πολλούς πίνακες συμπτώσεων, σχηματίζουμε έναν πίνακα συμπτώσεων πολλών μεταβλητών, όπως αυτόν στον Πίνακα 7.2. Στον πίνακα αυτό παρουσιάζονται ταυτόχρονα οι απαντήσεις 37 ατόμων για 3 ερωτήσεις, καθώς και το πλήθος ατόμων που απάντησε σε δύο κατηγορίες ανά δύο των ερωτήσεων. Ο συνολικός πίνακας είναι συμμετρικός και στους διαγώνιους υποπίνακες τα στοιχεία εκτός των διαγωνίων είναι μηδενικά, επειδή πρόκειται για κατηγορίες της ίδιας ερώτησης και κάθε άτομο απάντησε μόνο μία κατηγορία κάθε ερώτησης.

Κατηγορίες μεταβλητών με «παρόμοια κατανομή» γραμμών πλησιάζουν μεταξύ τους, δηλαδή το ίδιο πλήθος ατόμων απάντησε παρόμοια στις υπόλοιπες κατηγορίες όλων των άλλων μεταβλητών. Για να βρεθούν οι κατηγορίες μεταβλητών που είναι κοντά, και επομένως σχετίζονται, υπολογίζεται η «απόσταση» ή η «ομοιότητα» ανά δύο όλων των κατηγοριών.

**Πίνακας 7.2** *Πίνακας συμπτώσεων για 3 μεταβλητές ταυτόχρονα.*

| | | **Προσέχω να είμαι υγιής** | | | | | **Καταναλώνω τρόφιμα πλούσια σε φυτικές ίνες** | | | | | **Καταναλώνω τρόφιμα με χαμηλά λιπαρά** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **καθόλου** | **κάποιες φορές** | **μέτρια** | **συχνά** | **πολύ συχνά** | **καθόλου** | **κάποιες φορές** | **μέτρια** | **συχνά** | **πολύ συχνά** | **καθόλου** | **κάποιες φορές** | **μέτρια** | **συχνά** | **πολύ συχνά** |
| **Προσέχω να είμαι υγιής** | **καθόλου** | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| | **κάποιες φορές** | 0 | 4 | 0 | 0 | 0 | 1 | 1 | 2 | 0 | 0 | 1 | 3 | 0 | 0 | 0 |
| | **μέτρια** | 0 | 0 | 13 | 0 | 0 | 0 | 2 | 9 | 1 | 1 | 1 | 6 | 5 | 1 | 0 |
| | **συχνά** | 0 | 0 | 0 | 14 | 0 | 1 | 0 | 1 | 6 | 6 | 1 | 3 | 3 | 7 | 0 |
| | **πολύ συχνά** | 0 | 0 | 0 | 0 | 5 | 1 | 0 | 0 | 2 | 2 | 1 | 0 | 2 | 0 | 2 |
| **Καταναλώνω τρόφιμα πλούσια σε φυτικές ίνες** | **καθόλου** | 0 | 1 | 0 | 1 | 1 | 3 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 0 |
| | **κάποιες φορές** | 1 | 1 | 2 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 1 | 3 | 0 | 0 | 0 |
| | **μέτρια** | 0 | 2 | 9 | 1 | 0 | 0 | 0 | 12 | 0 | 0 | 1 | 7 | 4 | 0 | 0 |
| | **συχνά** | 0 | 0 | 1 | 6 | 2 | 0 | 0 | 0 | 9 | 0 | 0 | 2 | 6 | 1 | 0 |
| | **πολύ συχνά** | 0 | 0 | 1 | 6 | 2 | 0 | 0 | 0 | 0 | 9 | 1 | 0 | 0 | 6 | 2 |
| **Καταναλώνω τρόφιμα με χαμηλά λιπαρά** | **καθόλου** | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 0 | 1 | 5 | 0 | 0 | 0 | 0 |
| | **κάποιες φορές** | 0 | 3 | 6 | 3 | 0 | 0 | 3 | 7 | 2 | 0 | 0 | 12 | 0 | 0 | 0 |
| | **μέτρια** | 0 | 0 | 5 | 3 | 2 | 0 | 0 | 4 | 6 | 0 | 0 | 0 | 10 | 0 | 0 |
| | **συχνά** | 0 | 0 | 1 | 7 | 0 | 1 | 0 | 0 | 1 | 6 | 0 | 0 | 0 | 8 | 0 |
| | **πολύ συχνά** | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 2 |

Ένας πίνακας συμπτώσεων πολλών μεταβλητών είναι γνωστός και ως πίνακας Burt. Οι γραμμές του και οι στήλες του είναι ίσες σε πλήθος και ίσες με το άθροισμα των κατηγοριών όλων των μελετώμενων μεταβλητών. Ο πίνακας είναι συμμετρικός, και στην κύρια διαγώνιό του βρίσκονται τετράγωνοι διαγώνιοι υποπίνακες, που παριστάνουν την κατανομή των κατηγοριών μιας μεταβλητής στις παρατηρήσεις μας. Οι υπόλοιποι υποπίνακες είναι πίνακες συμπτώσεων ανά δύο όλων των μεταβλητών. Ο πίνακας Burt, χρησιμοποιείται στην παραγοντική ανάλυση.

Αν διαιρεθεί κάθε γραμμή ενός πίνακα συμπτώσεων με το άθροισμα όλων των στοιχείων της, προκύπτει ένας νέος πίνακας που παριστάνει τα ποσοστά των παρατηρήσεων της κατηγορίας που αντιστοιχεί στη γραμμή αυτή, τα οποία κατανέμονται στις κατηγορίες της άλλης μεταβλητής (στήλες του πίνακα). Παρόμοια, διαιρώντας κάθε στήλη του πίνακα με το άθροισμα των γραμμών της, προκύπτει ένας νέος πίνακας που παριστάνει την κατανομή των παρατηρήσεων μιας κατηγορίας (που αντιστοιχεί στη στήλη) στις κατηγορίες της άλλης μεταβλητής (γραμμές του πίνακα). Με τον τρόπο αυτό, ξεκινώντας από έναν πίνακα συμπτώσεων παράγονται 2 άλλοι πίνακες, χρήσιμοι για την εξαγωγή συμπερασμάτων μεταξύ κατηγοριών των 2 μεταβλητών.

Ο πίνακας Burt είναι πίνακας με στοιχεία αριθμούς (συχνότητες) των ποιοτικών μεταβλητών και μπορεί να χρησιμοποιηθεί σε μέτρηση καταλλήλων αποστάσεων μεταξύ κατηγοριών μεταβλητών. Από τον υποπίνακα για δύο μεταβλητές v1,v2 του πίνακα Burt, βρίσκεται η μεταξύ τους «απόσταση» με χρήση κάποιας κατάλληλης μετρικής αποστάσεων. Για την εύρεση σχέσης μεταξύ ενός ζεύγους (v1,v2) και μιας τρίτης μεταβλητής v3 χρησιμοποιείται ο υποπίνακας v3x(v1,v2). Οι διαγώνιοι υποπίνακες που αφορούν μόνο μία μεταβλητή, δεν χρησιμοποιούνται. Επίσης, με τον πίνακα Burt είναι δυνατόν να μελετηθούν κατηγορίες ποιοτικών μεταβλητών σαν ξεχωριστές μεταβλητές, οπότε προκύπτουν περισσότερο ερμηνεύσιμα συμπεράσματα αφού διακρίνονται και οι σχέσεις μεταξύ κατηγοριών.

## 7.5 Εύρεση κυρίων παραγόντων

Η ολική διασπορά των δεδομένων (άθροισμα από όλες τις διασπορές των μεταβλητών) διαχωρίζεται, βρίσκοντας έναν παράγοντα (δηλαδή γραμμικό συνδυασμό μεταβλητών) που έχει τη μέγιστη δυνατή διασπορά. Κατόπιν, βρίσκουμε άλλον παράγονται (νέο γραμμικό συνδυασμό μεταβλητών) με μέγιστη δυνατή διασπορά, αλλά ασυσχέτιστο με τον προηγούμενο παράγοντα, και συνεχίζουμε με στόχο να εντοπίσουμε όλους τους «αφανείς» παράγοντες.

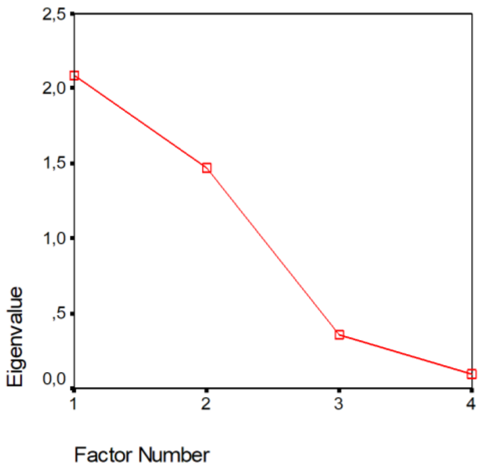
Οι παράγοντες ερμηνεύονται – περιγράφονται χρησιμοποιώντας γραμμικό συνδυασμό των αρχικών μεταβλητών. Η συσχέτιση μίας αρχικής μεταβλητής με έναν παράγοντα ονομάζεται βαρύτητα στον παράγοντα (factor loading). Το άθροισμα των τετραγώνων των φορτίων μίας μεταβλητής ονομάζεται communality της μεταβλητής και αντιπροσωπεύει την αναλογία της μεταβλητότητάς της που εξηγείται από τους παράγοντες (Agresti, 2021).

Για κάθε παράγοντα το άθροισμα των communality όλων των μεταβλητών είναι η διακύμανση (ή το ποσοστό της) που εξηγείται από τον συγκεκριμένο παράγοντα (Bartholomew et al., 2011, σελ. 268).

## 7.6 Επιλογή κυρίων παραγόντων

Σημαντικό μέρος στην ερμηνεία των αποτελεσμάτων της παραγοντικής ανάλυσης είναι η επιλογή του πλήθους των παραγόντων που θα ερμηνευτούν. Επιλέγεται μικρό πλήθος, αλλά με όσο γίνεται μεγαλύτερο ποσοστό ερμηνευμένης διασποράς των δεδομένων. Συνήθως, οι κύριες συνιστώσες που προκύπτουν είναι 2 έως 6 και ερμηνεύουν μεγάλο μέρος της διασποράς των δεδομένων (περίπου 75%).

Η επιλογή τους γίνεται από το διάγραμμα scree plot, το οποίο εμφανίζει το ποσοστό διασποράς σε σχέση με το πλήθος παραγόντων. Το σημείο που παρατηρείται απότομη πτώση (μεταβολή) στο διάγραμμα αντιστοιχεί στο πλήθος των επιλεγμένων παραγόντων. Ένα παράδειγμα scree plot εμφανίζεται στην Εικόνα 7.1. Στην εικόνα αυτή, πτώση του ποσοστού διασποράς συμβαίνει μεταξύ 2 και 3 παραγόντων, οπότε επιλέγονται δύο παράγοντες για ερμηνεία.




**Εικόνα 7.1** *Διάγραμμα scree plot για την επιλογή πλήθους παραγόντων.*

## 7.7 Ερμηνεία παραγόντων

Οι παράγοντες ερμηνεύονται ως σύνθεση των μεταβλητών που έχουν υψηλό φορτίο σε αυτούς. Δηλαδή των μεταβλητών που σχετίζονται υψηλά με κάθε παράγοντα.

### 7.7.1 Ονομασία παραγόντων

Ανάλογα με τις μεταβλητές που έχουν τη μεγαλύτερη βαρύτητα σε κάποιον παράγοντα, ονομάζεται ο παράγοντας με κάποιο γενικό όνομα που περιλαμβάνει καλύτερα τις μεταβλητές με τη μεγάλη βαρύτητα. Για παράδειγμα ένας παράγοντας που αποτελείται από βαθμολογία μαθητών στα καλλιτεχνικά, στην έκθεση και στη γλώσσα, θα μπορούσε να ονομαστεί παράγοντας έκφρασης.

Δεν είναι πάντα εύκολο να βρεθεί ονομασία για κάθε παράγοντα. Η διαδικασία αυτή απαιτεί πολύ καλή γνώση του αντικειμένου της έρευνας και των πιθανών «αφανών» μεταβλητών-παραγόντων. Είναι ευκολότερη στην επιβεβαιωτική παραγοντική ανάλυση, όπου υποτίθενται εκ των προτέρων οι παράγοντες και η παραγοντική ανάλυση επιβεβαιώνει τις υποθέσεις. Στη διερευνητική παραγοντική ανάλυση, πολλές φορές απλώς περιγράφονται οι μεταβλητές με μεγάλη βαρύτητα στον αντίστοιχο παράγοντα.

## 7.8 Γραφική απεικόνιση παραγοντικών επιπέδων

Έστω ότι οι μεταβλητές των δεδομένων είναι k σε πλήθος και οι παράγοντες που θα προκύψουν έστω m σε πλήθος, όπου το m<k. Στόχος είναι οι k μεταβλητές να αντικατασταθούν από τους το m παράγοντες. Οι k τιμές (συντεταγμένες) κάθε αντικειμένου για κάθε μεταβλητή, μετατρέπονται σε m ποσοτικές τιμές για τους παράγοντες. Στη συνέχεια, οι νέες ποσοτικές μεταβλητές (παράγοντες) μπορούν να χρησιμοποιηθούν αντί των αρχικών μεταβλητών και να γίνει μια νέα ανάλυση με μεθόδους κατάλληλες για ποσοτικές μεταβλητές.

Μπορούν, επίσης, να απεικονιστούν τα αντικείμενα σε ένα σύστημα συντεταγμένων των παραγόντων και να διακριθούν ομάδες αντικειμένων με παρόμοιες συντεταγμένες. Συνήθως, παρουσιάζονται τα αντικείμενα στο σύστημα των δύο ή τριών πρώτων παραγόντων. Ομάδες αντικειμένων που γειτνιάζουν στον χώρο ή στο επίπεδο των πρώτων παραγοντικών αξόνων αφορούν αντικείμενα με παρόμοια χαρακτηριστικά.

Προκειμένου να απεικονιστούν και τα χαρακτηριστικά τους, απεικονίζονται στο ίδιο επίπεδο και οι αρχικές μεταβλητές των δεδομένων. Μεταβλητές που γειτνιάζουν με κάποια ομάδα ατόμων αποτελούν χαρακτηριστικό της ομάδας αυτής.

## 7.9 Εφαρμογές για κατανόηση

### 7.9.1 Συνήθειες καταναλωτών

Σε μία έρευνα για τις συνήθειες καταναλωτών, πριν και μετά από την κρίση, ρωτήθηκε ένα δείγμα 660 καταναλωτών και συλλέχθηκαν απαντήσεις στις παρακάτω ερωτήσεις:

- Ηλικία (σε έτη)
- Φύλο (άνδρας - γυναίκα)
- Τόπος κατοικίας (μεγάλη πόλη - άλλες περιοχές)
- Επάγγελμα (φοιτητής - άνεργος - δημόσιος τομέας - ιδιωτικός τομέας)
- Κατηγορία μηνιαίου εισοδήματος (πριν από το 2010, μετά από το 2010)
- 1: <300 2: 300-600 3: 600-900 4: 900-1.200 5: >1.200

Υπήρχαν συνολικά 26 ερωτήσεις με 5 πολλαπλές απαντήσεις, κλίμακας 1 έως 5, στη μορφή:

1-πάρα πολύ, 2-πολύ, 3-μέτρια, 4-λίγο, 5-καθόλου, για ερωτήσεις γνώμης ή σκέψης.

Για ερωτήσεις που αφορούσαν το πόσο συχνά κάνουν κάτι, οι απαντήσεις ήταν κλίμακας 1 έως 5, στη μορφή: 1: πολύ συχνά, 2: 1φορά/εβδομάδα, 3: 1φορά/15μερο, 4: 1φορά/μήνα, 5: σπάνια.

Για ερωτήσεις που αφορούσαν τη δαπάνη αγοράς κάποιων προϊόντων, οι απαντήσεις ήταν κλίμακας 1 έως 5, στη μορφή: 1: <150  2: 150-300  3: 300-450  4: 450-600  5: >600 (δαπάνη).

Όλες οι ερωτήσεις πενταβάθμιας κλίμακας, μαζί με τον μέσο όρο απαντήσεων και την τυπική απόκλιση, εμφανίζονται στον Πίνακα 7.3.

**Πίνακας 7.3** *Παρουσίαση μεταβλητών και μέτρα θέσεως-διασποράς.*

| | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| εισόδημα πριν | 660 | 1 | 5 | 2,47 | 1,230 |
| εισόδημα μετά | 660 | 1 | 5 | 2,27 | 1,204 |
| γνώμη πριν | 660 | 1 | 5 | 2,46 | ,853 |
| γνώμη μετά | 660 | 1 | 5 | 2,96 | 1,010 |
| συχνότητα αγοράς ένδυσης πριν | 660 | 1 | 5 | 3,90 | 1,123 |
| σκέψη αγοράς ένδυσης πριν | 660 | 1 | 5 | 3,23 | 1,107 |
| δαπάνη αγοράς ένδυσης πριν | 660 | 1 | 5 | 1,58 | ,880 |
| συχνότητα αγοράς ένδυσης μετά | 660 | 1 | 5 | 4,22 | 1,034 |
| σκέψη αγοράς ένδυσης μετά | 660 | 1 | 5 | 2,43 | 1,110 |
| δαπάνη αγοράς ένδυσης μετά | 660 | 1 | 5 | 1,33 | ,693 |
| συχνότητα αγοράς καλλυντικών πριν | 660 | 1 | 5 | 3,97 | 1,136 |
| σκέψη αγοράς καλλυντικών πριν | 660 | 1 | 5 | 3,26 | 1,184 |
| δαπάνη αγοράς καλλυντικών πριν | 660 | 1 | 5 | 1,29 | ,715 |
| συχνότητα αγοράς καλλυντικών μετά | 659 | 1 | 5 | 4,32 | 1,008 |
| σκέψη αγοράς καλλυντικών μετά | 659 | 1 | 5 | 2,57 | 1,231 |
| δαπάνη αγοράς καλλυντικών μετά | 660 | 1 | 5 | 1,19 | ,645 |
| συχνότητα αγοράς τεχνολογίας πριν | 660 | 1 | 5 | 4,49 | ,919 |
| σκέψη αγοράς τεχνολογίας πριν | 660 | 1 | 5 | 2,65 | 1,195 |
| δαπάνη αγοράς τεχνολογίας πριν | 660 | 1 | 5 | 1,56 | ,939 |
| συχνότητα αγοράς τεχνολογίας μετά | 660 | 1 | 5 | 4,74 | ,673 |
| σκέψη αγοράς τεχνολογίας μετά | 660 | 1 | 5 | 2,08 | 1,195 |
| δαπάνη αγοράς τεχνολογίας μετά | 660 | 1 | 5 | 1,24 | ,590 |
| συχνότητα αγοράς διασκέδασης πριν | 658 | 1 | 5 | 2,34 | 1,057 |
| σκέψη αγοράς διασκέδασης πριν | 659 | 1 | 5 | 3,68 | 1,101 |
| δαπάνη αγοράς διασκέδασης πριν | 659 | 1 | 5 | 2,07 | 1,102 |
| συχνότητα αγοράς διασκέδασης μετά | 659 | 1 | 5 | 2,90 | 1,145 |
| σκέψη αγοράς διασκέδασης μετά | 658 | 1 | 5 | 2,73 | 1,213 |
| δαπάνη αγοράς διασκέδασης μετά | 659 | 1 | 5 | 1,59 | ,830 |
| Valid N (listwise) | 655 | | | | |

Στον πίνακα των απαντήσεων εφαρμόστηκε η παραγοντική ανάλυση με το SPSS, και τα εξαγόμενα του προγράμματος εμφανίζονται στους Πίνακες 7.4, 7.5, 7.6 στις Εικόνες 7.2, 7.3, 7.4, 7.5 και 7.6.

**Πίνακας 7.4** *Παρουσίαση αποτελεσμάτων KMO.*

| | |
|---|---|
| **Kaiser-Meyer-Olkin Measure of Sampling Adequacy.** | ,795 |
| **Bartlett's Test of Sphericity Approx. Chi-Square** | 10.340,798 |
| **df** | 378 |
| **Sig.** | ,000 |

Ο έλεγχος με το κριτήριο Bartlett’s sphericity test (Πίνακας 7.4), για την υπόθεση H0 ότι όλες οι μεταβλητές είναι μεταξύ τους ασυσχέτιστες, έδωσε sig <0,05 οπότε η H0 απορρίπτεται. Επομένως, τα δεδομένα είναι κατάλληλα για παραγοντική ανάλυση. Από το διάγραμμα Scree plot (Εικόνα 7.2), η απότομη μεταβολή στην ερμηνευόμενη διασπορά συμβαίνει μετά από τον δεύτερο και μετά από τον τέταρτο παράγοντα. Επιλέγεται η ερμηνεία για τους 4 κύριους παράγοντες που ερμηνεύουν το 52,5% της συνολικής μεταβλητότητας των δεδομένων, όπως φαίνεται στην προτελευταία στήλη (% of variance) του Πίνακα 7.5

**Πίνακας 7.5** *Πίνακας με ποσοστό ερμηνευόμενης διασποράς ανά παράγοντα.*

| | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | |
|---|---|---|---|---|---|---|
| **Component** | **Total** | **% of Variance** | **Cumulative %** | **Total** | **% of Variance** | **Cumulative %** |
| 1 | 7,680 | 27,430 | 27,430 | 7,680 | 27,430 | 27,430 |
| 2 | 2,991 | 10,683 | 38,114 | 2,991 | 10,683 | 38,114 |
| 3 | 2,277 | 8,130 | 46,244 | 2,277 | 8,130 | 46,244 |
| 4 | 1,766 | 6,306 | 52,550 | 1,766 | 6,306 | 52,550 |
| 5 | 1,466 | 5,235 | 57,786 | 1,466 | 5,235 | 57,786 |
| 6 | 1,390 | 4,964 | 62,750 | 1,390 | 4,964 | 62,750 |
| 7 | 1,173 | 4,190 | 66,940 | 1,173 | 4,190 | 66,940 |
| 8 | ,997 | 3,561 | 70,501 | | | |
| 9 | ,922 | 3,294 | 73,795 | | | |
| 10 | ,882 | 3,149 | 76,944 | | | |
| 11 | ,736 | 2,627 | 79,572 | | | |
| 12 | ,696 | 2,485 | 82,057 | | | |
| 13 | ,662 | 2,366 | 84,422 | | | |
| 14 | ,572 | 2,043 | 86,465 | | | |
| 15 | ,504 | 1,800 | 88,265 | | | |
| 16 | ,450 | 1,605 | 89,870 | | | |
| 17 | ,372 | 1,327 | 91,197 | | | |
| 18 | ,361 | 1,291 | 92,488 | | | |
| 19 | ,328 | 1,171 | 93,660 | | | |
| 20 | ,289 | 1,033 | 94,693 | | | |
| 21 | ,229 | ,817 | 95,510 | | | |
| 22 | ,227 | ,810 | 96,320 | | | |
| 23 | ,206 | ,735 | 97,055 | | | |
| 24 | ,203 | ,725 | 97,780 | | | |
| 25 | ,189 | ,677 | 98,457 | | | |
| 26 | ,182 | ,649 | 99,106 | | | |
| 27 | ,164 | ,587 | 99,692 | | | |
| 28 | ,086 | ,308 | 100,000 | | | |

| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | |
|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 7,680 | 27,430 | 27,430 | 7,680 | 27,430 | 27,430 |
| 2 | 2,991 | 10,683 | 38,114 | 2,991 | 10,683 | 38,114 |
| 3 | 2,277 | 8,130 | 46,244 | 2,277 | 8,130 | 46,244 |
| 4 | 1,766 | 6,306 | 52,550 | 1,766 | 6,306 | 52,550 |
| 5 | 1,466 | 5,235 | 57,786 | 1,466 | 5,235 | 57,786 |
| 6 | 1,390 | 4,964 | 62,750 | 1,390 | 4,964 | 62,750 |
| 7 | 1,173 | 4,190 | 66,940 | 1,173 | 4,190 | 66,940 |
| 8 | ,997 | 3,561 | 70,501 | | | |
| 9 | ,922 | 3,294 | 73,795 | | | |
| 10 | ,882 | 3,149 | 76,944 | | | |
| 11 | ,736 | 2,627 | 79,572 | | | |
| 12 | ,696 | 2,485 | 82,057 | | | |
| 13 | ,662 | 2,366 | 84,422 | | | |
| 14 | ,572 | 2,043 | 86,465 | | | |
| 15 | ,504 | 1,800 | 88,265 | | | |
| 16 | ,450 | 1,605 | 89,870 | | | |
| 17 | ,372 | 1,327 | 91,197 | | | |
| 18 | ,361 | 1,291 | 92,488 | | | |
| 19 | ,328 | 1,171 | 93,660 | | | |
| 20 | ,289 | 1,033 | 94,693 | | | |
| 21 | ,229 | ,817 | 95,510 | | | |
| 22 | ,227 | ,810 | 96,320 | | | |
| 23 | ,206 | ,735 | 97,055 | | | |
| 24 | ,203 | ,725 | 97,780 | | | |
| 25 | ,189 | ,677 | 98,457 | | | |
| 26 | ,182 | ,649 | 99,106 | | | |
| 27 | ,164 | ,587 | 99,692 | | | |
| 28 | ,086 | ,308 | 100,000 | | | |

Extraction Method: Principal Component Analysis.

Οι συσχετίσεις των μεταβλητών στους τέσσερις κύριους παράγοντες εμφανίζονται στον Πίνακα 7.6. Παρατηρώντας τις συσχετίσεις κάθε παράγοντα (στήλης), δίνεται η εξής ερμηνεία των παραγόντων:

1ος παράγοντας σχετίζεται θετικά με τις μεταβλητές

- Δαπάνη αγοράς ένδυσης πριν και μετά
- Δαπάνη διασκέδασης μετά
- Σκέψη αγοράς ένδυσης μετά
- Σκέψη αγοράς καλλυντικών πριν και μετά

2ος παράγοντας σχετίζεται θετικά με τις μεταβλητές

- Εισόδημα πριν
- Δαπάνη αγοράς καλλυντικών πριν και μετά
- Συχνότητα διασκέδασης πριν και μετά

- Αρνητικά με τη Σκέψη διασκέδασης μετά

3ος παράγοντας σχετίζεται θετικά με τις μεταβλητές

- Εισόδημα πριν και μετά
- Συχνότητα αγοράς ένδυσης πριν και μετά
- Σκέψη αγοράς καλλυντικών μετά

4ος παράγοντας σχετίζεται θετικά με τις μεταβλητές

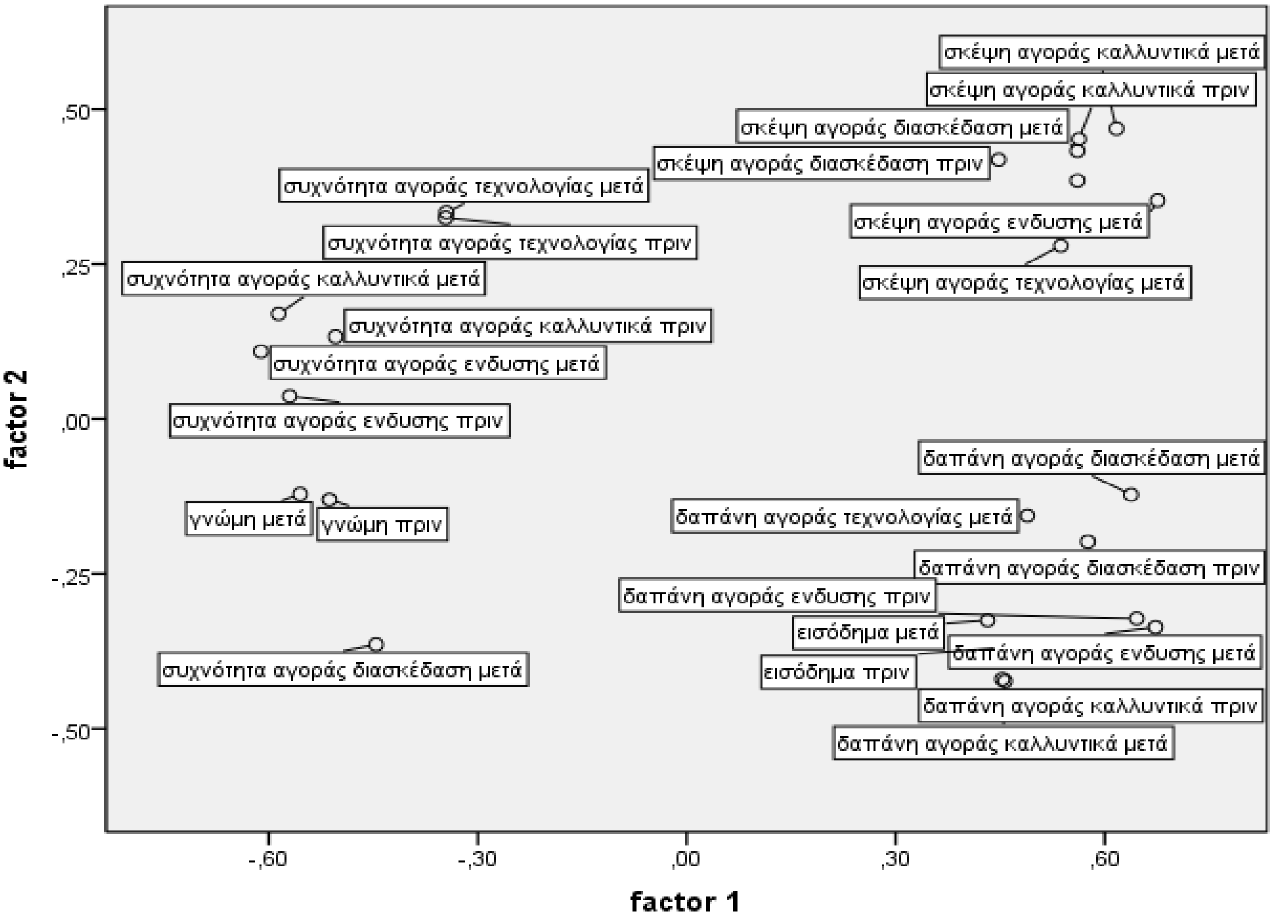
- Συχνότητα αγοράς τεχνολογίας μετά
- Αρνητικά με τη Σκέψη αγοράς τεχνολογίας μετά

**Πίνακας 7.6** *Πίνακας με συσχετίσεις μεταβλητών ανά παράγοντα.*

| | **Component** | | | |
|---|---|---|---|---|
| | **1** | **2** | **3** | **4** |
| εισόδημα πριν | ,463 | ,406 | ,531 | ,196 |
| εισόδημα μετά | ,428 | ,363 | ,518 | ,128 |
| γνώμη πριν | -,535 | ,217 | -,033 | ,180 |
| γνώμη μετά | -,511 | ,165 | ,055 | ,333 |
| συχνότητα αγοράς ένδυσης πριν | -,530 | ,058 | ,496 | -,280 |
| σκέψη αγοράς ένδυσης πριν | ,592 | -,344 | ,192 | ,238 |
| δαπάνη αγοράς ένδυσης πριν | ,633 | ,393 | ,125 | ,218 |
| συχνότητα αγοράς ένδυσης μετά | -,527 | -,060 | ,609 | -,017 |
| σκέψη αγοράς ένδυσης μετά | ,616 | -,297 | -,016 | -,172 |
| δαπάνη αγοράς ένδυσης μετά | ,613 | ,421 | ,000 | ,033 |
| συχνότητα αγοράς καλλυντικών πριν | -,493 | -,040 | ,517 | -,361 |
| σκέψη αγοράς καλλυντικών πριν | ,609 | -,369 | ,181 | ,116 |
| δαπάνη αγοράς καλλυντικών πριν | ,482 | ,467 | -,218 | ,135 |
| συχνότητα αγοράς καλλυντικών μετά | -,560 | -,127 | ,561 | -,079 |
| σκέψη αγοράς καλλυντικών μετά | ,610 | -,376 | ,088 | -,242 |
| δαπάνη αγοράς καλλυντικών μετά | ,408 | ,456 | -,181 | -,150 |
| συχνότητα αγοράς τεχνολογίας πριν | -,388 | -,219 | ,131 | ,391 |
| σκέψη αγοράς τεχνολογίας πριν | ,528 | -,311 | ,294 | -,233 |
| δαπάνη αγοράς τεχνολογίας πριν | ,461 | ,350 | ,156 | -,138 |
| συχνότητα αγοράς τεχνολογίας μετά | -,356 | -,285 | ,138 | ,499 |
| σκέψη αγοράς τεχνολογίας μετά | ,536 | -,292 | ,069 | -,481 |
| δαπάνη αγοράς τεχνολογίας μετά | ,498 | ,281 | ,075 | -,233 |
| συχνότητα αγοράς διασκέδασης πριν | -,391 | ,440 | ,005 | -,359 |
| σκέψη αγοράς διασκέδασης πριν | ,513 | -,389 | ,232 | ,341 |
| δαπάνη αγοράς διασκέδασης πριν | ,584 | ,271 | ,276 | ,286 |
| συχνότητα αγοράς διασκέδασης μετά | -,438 | ,450 | ,181 | -,094 |
| σκέψη αγοράς διασκέδασης μετά | ,545 | -,407 | ,033 | -,080 |
| δαπάνη αγοράς διασκέδασης μετά | ,650 | ,236 | ,169 | ,048 |

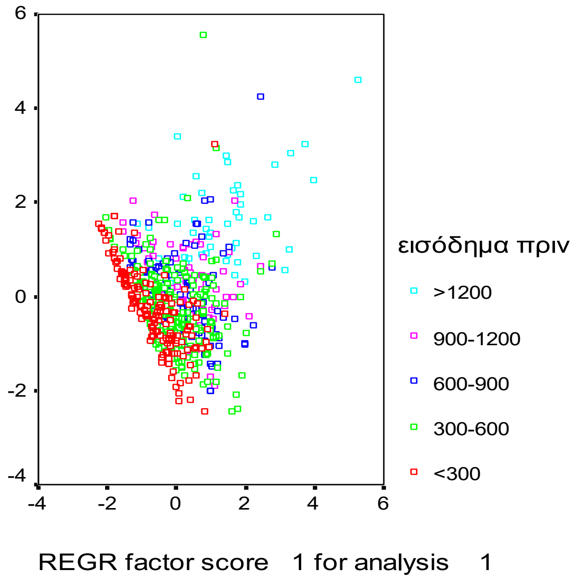
Ονομάζεται ο κάθε παράγοντας με γενικό όνομα που περιλαμβάνει όσο γίνεται καλύτερα τις μεταβλητές με τη μεγάλη βαρύτητα. Έτσι, προτείνεται η παρακάτω ονομασία για τους παράγοντες:

- 1ος παράγοντας – Μεγάλες δαπάνες πριν και μετά από την κρίση, χωρίς πολλή σκέψη αγοράς.
- 2ος παράγοντας – Μεγάλες δαπάνες πριν και μετά από την κρίση χωρίς αλλαγή συχνότητας διασκέδασης αλλά και χωρίς σκέψη διασκέδασης μετά.
- 3ος παράγοντας – Μέτρια προς υψηλά εισοδήματα με περισσότερη σκέψη αγοράς καλλυντικών μετά από την κρίση.
- 4ος παράγοντας – Μέτρια προς χαμηλά εισοδήματα με περισσότερη σκέψη αγοράς τεχνολογίας μετά από την κρίση.

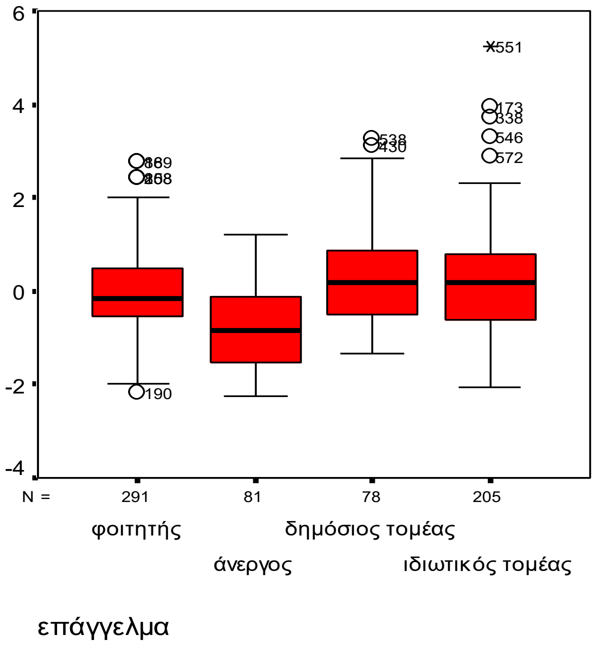



**Εικόνα 7.3** *Εμφάνιση των μεταβλητών στον πρώτο και δεύτερο παράγοντα.*

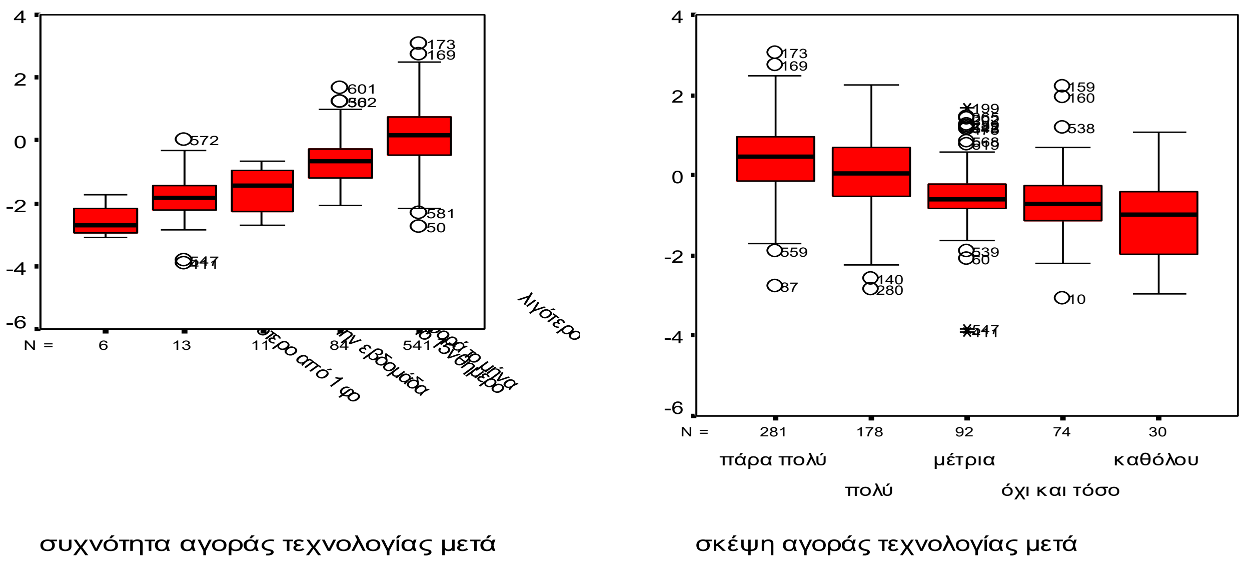
Στην Εικόνα 7.3, που αναπαριστά τις συντεταγμένες των μεταβλητών στο επίπεδο του πρώτου και δεύτερου παράγοντα, διακρίνονται ποιες μεταβλητές πλησιάζουν μεταξύ τους, δηλαδή τα άτομα απάντησαν παρόμοια σε αυτές, και ποιες διαχωρίζονται μεταξύ τους. Στο ίδιο επίπεδο των αξόνων πρώτου και δεύτερου, μπορούν να αναπαρασταθούν και οι κλάσεις του εισοδήματος με διαφορετικό χρώμα, ώστε να φανεί η σχέση εισοδήματος και ομάδων μεταβλητών που προέκυψαν από την παραγοντική ανάλυση. Αυτό φαίνεται στο επάνω αριστερό σχήμα της Εικόνας 7.4. Παρατηρείται ότι όσο μεγαλύτερο είναι το εισόδημα τόσο δεξιά κι επάνω βρίσκονται τα αντίστοιχα σημεία. Δηλαδή, έχουν θετικές και μεγάλες τιμές τόσο για τον πρώτο παράγοντα όσο και για τον δεύτερο παράγοντα.




**Εικόνα 7.4** *Εμφάνιση εισοδήματος στον πρώτο και δεύτερο παράγοντα.*




**Εικόνα 7.5** *Θηκογράμματα πρώτου παράγοντα ως προς το επάγγελμα.*




**Εικόνα 7.6** *Θηκογράμματα τέταρτου παράγοντα ως συχνότητα αγορών (αριστερά) και σκέψης αγορών τεχνολογίας (δεξιά).*

Άλλος τρόπος αναπαράστασης είναι η δημιουργία θηκογράμματος για το *επάγγελμα* και τις συντεταγμένες του στον πρώτο παραγοντικό άξονα (Εικόνα 7.4), όπου παρατηρούνται αρνητικές τιμές για τους ανέργους και φοιτητές. Το θηκόγραμμα για τις μεταβλητές «συχνότητα αγοράς τεχνολογίας μετά από την κρίση» και «σκέψη για την αγορά τεχνολογίας μετά από την κρίση», εμφανίζει ξεκάθαρα τις διαφοροποιήσεις στον 4ο παραγοντικό άξονα (σχήμα αριστερά και δεξιά στην Εικόνα 7.5 αντίστοιχα).

### 7.9.2 Επισκέψεις σε λασπόλουτρα

Σε μία έρευνα για τους λόγους επισκέψεων σε λασπόλουτρα, ρωτήθηκε ένα δείγμα 83 επισκεπτών και συλλέχθηκαν απαντήσεις στις παρακάτω ερωτήσεις:

- Φύλο (άνδρας - γυναίκα)
- Ηλικία (σε δεκαετία με τιμές 1,2,3,4,5,6)
- Πλήθος επισκέψεων (3 κατηγορίες)
- Χρόνος παραμονής σε λασπόλουτρο (λίγο ή πολύ)
- Συχνότητα ανά ημέρα (μία ή δύο)
- Διενέργεια ασκήσεων (3 κατηγορίες)
- Χρόνος ξεκούρασης (3 κατηγορίες)
- Μέρες διακοπής (0,1,2,3,…)
- Εξέταση από γιατρό (ναι ή όχι)
- Ωφέλεια (3 κατηγορίες)
- Ευχάριστο περιβάλλον (3 κατηγορίες)
- Ευχάριστη διαμονή (3 κατηγορίες)
- Σύσταση σε γνωστό (ναι ή όχι)

**Πίνακας 7.7** *Παρουσίαση αποτελεσμάτων KMO.*

| | |
|---|---|
| **Kaiser-Meyer-Olkin Measure of Sampling Adequacy.** | ,395 |
| **Bartlett's Test of Sphericity Approx. Chi-Square** | 92,844 |
| **df** | 78 |
| **Sig.** | ,120 |

**Πίνακας 7.8** *Πίνακας με ποσοστό ερμηνευόμενης διασποράς ανά παράγοντα.*

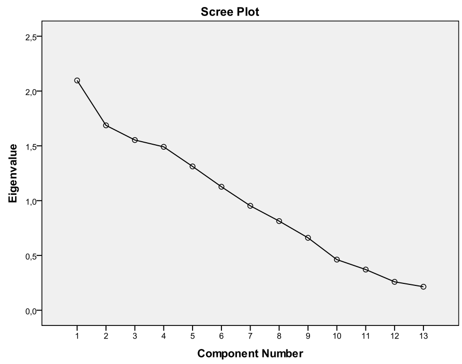
| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | |
|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 2,096 | 16,127 | 16,127 | 2,096 | 16,127 | 16,127 |
| 2 | 1,687 | 12,975 | 29,102 | 1,687 | 12,975 | 29,102 |
| 3 | 1,553 | 11,946 | 41,048 | 1,553 | 11,946 | 41,048 |
| 4 | 1,491 | 11,468 | 52,516 | 1,491 | 11,468 | 52,516 |
| 5 | 1,312 | 10,093 | 62,608 | 1,312 | 10,093 | 62,608 |
| 6 | 1,127 | 8,667 | 71,275 | 1,127 | 8,667 | 71,275 |
| 7 | ,953 | 7,329 | 78,605 | | | |
| 8 | ,813 | 6,257 | 84,861 | | | |
| 9 | ,661 | 5,083 | 89,945 | | | |
| 10 | ,462 | 3,555 | 93,500 | | | |
| 11 | ,371 | 2,856 | 96,356 | | | |
| 12 | ,259 | 1,994 | 98,350 | | | |
| 13 | ,215 | 1,650 | 100,000 | | | |

Extraction Method: Principal Component Analysis.

Υπήρχαν συνολικά 11 ερωτήσεις, με απαντήσεις που αντιστοιχούν σε ποιοτικές μεταβλητές, αλλά με διαφορετική κλίμακα μέτρησης, άλλες ερωτήσεις με δύο πιθανές απαντήσεις, κάποιες με 3 πιθανές απαντήσεις, και 2 ερωτήσεις με ακέραιες τιμές που αντιστοιχούν σε διακριτές ποσοτικές μεταβλητές.

Πραγματοποιήθηκε παραγοντική ανάλυση στις απαντήσεις αυτών των ερωτήσεων και τα αποτελέσματα εμφανίζονται στους Πίνακες 7.7, 7.8 και στην Εικόνα 7.6.

Σε ό,τι αφορά τον έλεγχο με το κριτήριο Bartlett's sphericity test, για την υπόθεση H0 ότι όλες οι μεταβλητές είναι μεταξύ τους ασυσχέτιστες (Πίνακας 7.7), η πιθανότητα λάθους, αν απορρίψουμε την H0 είναι sig=0,12 >0,05, οπότε η H0 δεν απορρίπτεται. Τα δεδομένα δεν είναι κατάλληλα για παραγοντική ανάλυση, αφού ήδη οι μεταβλητές είναι ασυσχέτιστες. Από το διάγραμμα Scree plot (Εικόνα 7.6) δεν παρατηρείται απότομη μεταβολή στην ερμηνευόμενη διασπορά. Παρ' όλα αυτά, το SPSS δίνει τα ποσοστά συσχετίσεων των μεταβλητών με τους 6 κύριους παράγοντες που ερμηνεύουν το 71,3% της συνολικής μεταβλητότητας των δεδομένων, όπως φαίνεται στην προτελευταία στήλη (% of variance) του Πίνακα 7.8. Τα αποτελέσματα, όμως, δεν είναι αξιόπιστα λόγω μη καταλληλότητας των δεδομένων.




**Εικόνα 7.6** *Scree plot παραγοντικής ανάλυσης για επισκέψεις σε λασπόλουτρα.*

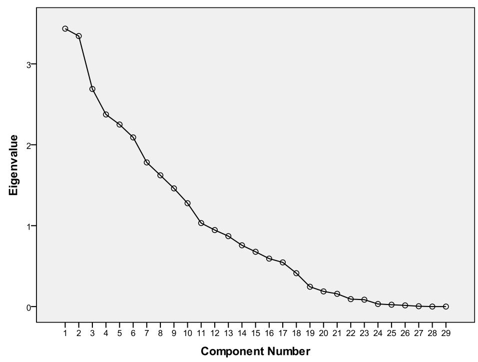
Σε παρόμοιες περιπτώσεις όπου οι μεταβλητές δεν έχουν το ίδιο πλήθος κλάσεων (δυνατότητες απαντήσεων διαφορετικές σε πλήθος), μία λύση θα ήταν κάθε μεταβλητή να αναλυθεί σε τόσες νέες μεταβλητές όσες και οι κλάσεις της. Κάθε νέα μεταβλητή θα παίρνει μόνο δύο τιμές, την τιμή 1 αν απαντήθηκε η αντίστοιχη κατηγορία, και την τιμή 0 διαφορετικά.

Στη συνέχεια, πραγματοποιείται παραγοντική ανάλυση στις νέες αυτές μεταβλητές όπου όλες έχουν ίσο αριθμό δύο κλάσεων. Τα αποτελέσματα της εμφανίζονται στον Πίνακα 7.9, 7.10 και στην Εικόνα 7.7.

**Πίνακας 7.9** *Παράγοντες της παραγοντικής ανάλυσης νέων μεταβλητών 0-1 για επισκέψεις σε λασπόλουτρα.*

| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | |
|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 3,435 | 11,845 | 11,845 | 3,435 | 11,845 | 11,845 |
| 2 | 3,344 | 11,531 | 23,376 | 3,344 | 11,531 | 23,376 |
| 3 | 2,689 | 9,274 | 32,650 | 2,689 | 9,274 | 32,650 |
| 4 | 2,374 | 8,187 | 40,837 | 2,374 | 8,187 | 40,837 |
| 5 | 2,250 | 7,759 | 48,596 | 2,250 | 7,759 | 48,596 |
| 6 | 2,091 | 7,210 | 55,806 | 2,091 | 7,210 | 55,806 |
| 7 | 1,782 | 6,143 | 61,949 | 1,782 | 6,143 | 61,949 |
| 8 | 1,622 | 5,595 | 67,544 | 1,622 | 5,595 | 67,544 |
| 9 | 1,461 | 5,037 | 72,581 | 1,461 | 5,037 | 72,581 |
| 10 | 1,278 | 4,405 | 76,987 | 1,278 | 4,405 | 76,987 |
| 11 | 1,032 | 3,559 | 80,545 | 1,032 | 3,559 | 80,545 |
| 12 | ,946 | 3,261 | 83,807 | | | |
| 13 | ,870 | 3,002 | 86,808 | | | |
| 14 | ,757 | 2,611 | 89,419 | | | |
| 15 | ,677 | 2,335 | 91,755 | | | |
| 16 | ,593 | 2,044 | 93,798 | | | |
| 17 | ,546 | 1,882 | 95,681 | | | |
| 18 | ,412 | 1,422 | 97,102 | | | |
| 19 | ,245 | ,844 | 97,947 | | | |
| 20 | ,187 | ,646 | 98,593 | | | |
| 21 | ,158 | ,544 | 99,137 | | | |
| 22 | ,092 | ,317 | 99,454 | | | |
| 23 | ,085 | ,293 | 99,747 | | | |
| 24 | ,032 | ,110 | 99,857 | | | |
| 25 | ,023 | ,079 | 99,936 | | | |
| 26 | ,015 | ,050 | 99,986 | | | |
| 27 | ,004 | ,014 | 100,000 | | | |
| 28 | 1,868E-16 | 6,443E-16 | 100,000 | | | |
| 29 | -6,198E-16 | -2,137E-15 | 100,000 | | | |

Extraction Method: Principal Component Analysis.




**Εικόνα 7.8** *Scree plot παραγοντικής ανάλυσης για επισκέψεις σε λασπόλουτρα (μεταβλητές με τιμές 0-1).*

**Πίνακας 7.10** *Συσχετίσεις με τους κύριους παράγοντες της παραγοντικής ανάλυσης των νέων μεταβλητών 0-1 για επισκέψεις σε λασπόλουτρα.*

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| άνδρας | -,274 | -,539 | ,382 | -,242 | ,289 | -,105 |
| γυναίκα | ,274 | ,539 | -,382 | ,242 | -,289 | ,105 |
| έτος2 | ,043 | -,111 | ,024 | ,396 | ,415 | ,169 |
| έτος3 | -,253 | -,334 | ,354 | -,467 | -,023 | ,093 |
| έτος4 | ,165 | ,402 | -,267 | ,041 | -,203 | -,134 |
| πλήθος1 | ,123 | ,046 | -,047 | ,581 | -,189 | -,074 |
| πλήθος2 | -,491 | ,136 | ,233 | -,172 | -,022 | -,409 |
| πλήθος3 | ,408 | -,173 | -,202 | -,258 | ,162 | ,470 |
| παραμονή2 | ,103 | ,044 | ,575 | ,590 | -,101 | -,122 |
| συχνότητα1 | -,359 | -,293 | -,236 | ,473 | ,288 | ,371 |
| συχνότητα2 | ,551 | ,329 | ,323 | -,430 | -,156 | -,141 |
| κίνηση1 | ,278 | -,278 | ,273 | ,185 | -,572 | ,307 |
| κίνηση2 | -,263 | ,404 | ,044 | ,012 | ,581 | ,014 |
| κίνηση3 | ,150 | -,208 | -,348 | -,188 | ,241 | -,100 |
| ξεκούραση1 | -,177 | -,059 | ,644 | -,098 | -,077 | ,281 |
| ξεκούραση2 | ,545 | ,041 | -,324 | -,234 | -,042 | -,023 |
| ξεκούραση3 | -,123 | ,023 | -,371 | ,369 | ,271 | -,027 |
| γιατρός1 | -,205 | -,346 | -,161 | ,049 | -,318 | ,603 |
| γιατρός2 | ,357 | ,268 | ,241 | ,084 | ,458 | -,370 |
| ωφέλεια1 | -,234 | ,226 | -,027 | ,201 | ,318 | ,114 |
| ωφέλεια2 | ,086 | ,180 | ,366 | -,123 | ,279 | ,365 |
| ωφέλεια3 | ,676 | -,103 | ,077 | ,017 | -,162 | -,157 |
| περιβάλλον2 | -,041 | ,714 | ,278 | -,139 | ,107 | ,414 |
| περιβάλλον3 | ,622 | -,522 | -,022 | ,107 | ,277 | -,142 |
| διαμονή2 | ,203 | ,572 | ,101 | -,039 | ,103 | ,392 |
| διαμονή3 | ,493 | -,470 | ,064 | ,045 | ,353 | -,139 |
| σύσταση1 | ,597 | ,221 | ,165 | -,006 | ,359 | ,340 |
| 1 μέρα διακοπή | ,278 | -,523 | ,125 | ,077 | ,017 | ,299 |
| παραμονή1 | -,138 | -,060 | -,631 | -,561 | ,127 | ,150 |

Στη νέα αυτή παραγοντική ανάλυση ερμηνεύονται οι 6 πρώτοι παραγοντικοί άξονες. Τα αποτελέσματά τους εμφανίζονται στον Πίνακα 7.10 και αυτή τη φορά είναι ευκολότερη η ερμηνεία τους. Ο πρώτος παράγοντας σχετίζεται με όσους είναι ευχαριστημένοι από την ωφέλεια και το περιβάλλον των λασπόλουτρων και θα σύστηναν τη λασποθεραπεία σε άλλους. Ο δεύτερος άξονας σχετίζεται με γυναίκες της 4ης δεκαετίας ηλικίας, που κάνουν μέτριες ασκήσεις κίνησης, δηλώνουν μέτρια ικανοποιημένες από το περιβάλλον και τη διαμονή τους και δεν διέκοψαν λόγω αδιαθεσίας. Ο τρίτος άξονας σχετίζεται με όσους παραμένουν πάνω από 30 λεπτά στο λουτρό και κατόπιν ξεκουράζονται 1-2 ώρες. Ο τέταρτος σχετίζεται με όσους κάνουν λίγες λασποθεραπείες και παραμένουν λίγο χρόνο στο λουτρό, κυρίως είναι σε μικρή ηλικία. Ο πέμπτος σχετίζεται με όσους τους εξέτασε γιατρός, κάνουν μέτριες κινήσεις και είναι σε μικρή ηλικία. Ο έκτος άξονας σχετίζεται με όσους κάνουν πολλές λασποθεραπείες, αλλά μία κάθε μέρα και τους έχει εξετάσει γιατρός.

## 7.10 Ασκήσεις για κατανόηση - εφαρμογή

Σε μία έρευνα για τους αποφοίτους και της δημοσιεύσεις μελών ΔΕΠ κάποιων ΑΕΙ, εφαρμόστηκε η παραγοντική ανάλυση και προέκυψαν τα αποτελέσματα που εμφανίζονται στους παρακάτω πίνακες.

Πόσοι παραγοντικοί άξονες προέκυψαν; Τι ποσοστό διασποράς ερμηνεύει ο καθένας; Ποιες μεταβλητές σχετίζονται με κάθε παραγοντικό άξονα; Πώς θα μπορούσε να ονομαστεί ο καθένας;

**Total Variance Explained**

| | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | |
|---|---|---|---|---|---|---|
| Component | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 2,840 | 35,504 | 35,504 | 2,840 | 35,504 | 35,504 |
| 2 | 2,007 | 25,090 | 60,594 | 2,007 | 25,090 | 60,594 |
| 3 | 1,476 | 18,454 | 79,048 | 1,476 | 18,454 | 79,048 |
| 4 | 1,041 | 13,008 | 92,056 | 1,041 | 13,008 | 92,056 |
| 5 | ,590 | 7,381 | 99,437 | | | |
| 6 | ,026 | ,322 | 99,759 | | | |
| 7 | ,015 | ,191 | 99,950 | | | |
| 8 | ,004 | ,050 | 100,000 | | | |

Extraction Method: Principal Component Analysis.

**Component Matrix[a]**

| | Component | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| ΔΗΜΟΣΙΕΥΣΕΙΣ/ ΛΕΙΤΟΥΡΓΙΚΕΣ_ΔΑΠΑΝΕΣ | ,744 | -,463 | ,245 | -,233 |
| ΔΗΜΟΣΙΕΥΣΕΙΣ/ ΕΠ | ,589 | -,475 | ,089 | ,642 |
| ΔΗΜΟΣΙΕΥΣΕΙΣ /ΣΥΜΒΑΣΗ | ,688 | ,071 | -,714 | -,034 |
| ΔΗΜΟΣΙΕΥΣΕΙΣ /ΦΟΙΤΗΤΕΣ | ,692 | -,622 | ,191 | -,150 |
| ΑΠΟΦΟΙΤΟΙ/ ΛΕΙΤΟΥΡΓΙΚΕΣ_ΔΑΠΑΝΕΣ | ,466 | ,608 | ,394 | -,312 |
| ΑΠΟΦΟΙΤΟΙ/ ΕΠ | ,320 | ,632 | ,288 | ,642 |
| ΑΠΟΦΟΙΤΟΙ/ ΣΥΜΒΑΣΗ | ,593 | ,443 | -,663 | -,042 |
| ΑΠΟΦΟΙΤΟΙ/ ΦΟΙΤΗΤΕΣ | ,563 | ,458 | ,430 | -,199 |

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

<u>Λύση</u>

Προέκυψαν 8 παραγοντικοί άξονες, όσες και οι μεταβλητές που χρησιμοποιήθηκαν. Όμως για τους πρώτους 4 παραγοντικούς άξονες υπολογίστηκαν οι συσχετίσεις με τις αντίστοιχες μεταβλητές. Το ποσοστό διασποράς που ερμηνεύει ο πρώτος παράγοντας είναι 35,5%, ο δεύτερος παράγοντας ερμηνεύει 25%, ο τρίτος

παράγοντας 18,5% και ο τέταρτος 13%. Συνολικά, οι 4 παράγοντες ερμηνεύουν 92% της διασποράς των δεδομένων.

Στον δεύτερο από τους παραπάνω πίνακες φαίνονται οι συσχετίσεις των μεταβλητών με τους παράγοντες. Ο πρώτος παράγοντας ερμηνεύει τις μεταβλητές «Δημοσιεύσεις» σε σχέση με τις «Λειτουργικές δαπάνες», «Δημοσιεύσεις» σε σχέση με τις «Συμβάσεις προσωπικού» και «Δημοσιεύσεις» σε σχέση με τους «Φοιτητές».

Ο δεύτερος παράγοντας ερμηνεύει θετικά τις μεταβλητές «Απόφοιτοι» σε σχέση με τις «Λειτουργικές δαπάνες» και «Απόφοιτοι» σε σχέση με το «Εκπαιδευτικό Προσωπικό» και αρνητικά τη μεταβλητή «Δημοσιεύσεις» σε σχέση με τους «Φοιτητές».

Ο τρίτος παράγοντας ερμηνεύει αρνητικά τις μεταβλητές «Δημοσιεύσεις» σε σχέση με το «Εκπαιδευτικό Προσωπικό» και «Απόφοιτοι» σε σχέση με το «Εκπαιδευτικό Προσωπικό».

Ο τέταρτος παράγοντας ερμηνεύει τις μεταβλητές «Δημοσιεύσεις» σε σχέση με τις «Συμβάσεις προσωπικού» και «Απόφοιτοι» σε σχέση σε σχέση με τις «Συμβάσεις προσωπικού».

Ο πρώτος παραγοντικός άξονας θα μπορούσε να ονομαστεί «Επιστημονικές δημοσιεύσεις του ΑΕΙ». Ο δεύτερος παραγοντικός άξονας θα μπορούσε να ονομαστεί «Απόφοιτοι του ΑΕΙ». Ο τρίτος παραγοντικός άξονας θα μπορούσε να ονομαστεί «Συμβασιούχοι του ΑΕΙ». Ο τέταρτος παραγοντικός άξονας θα μπορούσε να ονομαστεί «Εκπαιδευτικό προσωπικό του ΑΕΙ».

# Βιβλιογραφία/Αναφορές


Agresti, A. (2021). *Στατιστικές μέθοδοι για κοινωνικές επιστήμες*. Θεσσαλονίκη: Τζιόλα.

Καρλής, Δ. (2005). *Πολυμεταβλητή Στατιστική Ανάλυση*. Αθήνα: Σταμούλη.

Πετρίδης, Δ. (2015). *ΑΝΑΛΥΣΗ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ - ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ* [Κεφάλαιο]. Στο Πετρίδης, Δ. 2015. *Ανάλυση πολυμεταβλητών τεχνικών* [Προπτυχιακό εγχειρίδιο]. Κάλλιπος, Ανοικτές Ακαδημαϊκές Εκδόσεις. https://hdl.handle.net/11419/2129

Bartholomew, D., Steele, F., Moustaki, I., & Galbraith, J. (2011). *Ανάλυση Πολυμεταβλητών Τεχνικών στις Κοινωνικές Επιστήμες*. Αθήνα: Κλειδάριθμος.

# Κριτήρια αξιολόγησης

## Κριτήριο αξιολόγησης 1

**Τι είναι η παραγοντική ανάλυση;**

**Απάντηση**
Η **παραγοντική ανάλυση** είναι μια πολυμεταβλητή τεχνική, η οποία προσπαθεί να ερμηνεύσει συσχετίσεις μεταξύ παρατηρούμενων μεταβλητών, υποθέτοντας ότι αυτές οφείλονται στην ύπαρξη κάποιων κοινών παραγόντων, οι οποίοι δεν φαίνονται άμεσα, ούτε μπορούν να μετρηθούν άμεσα.

## Κριτήριο αξιολόγησης 2

**Ποιος είναι ο στόχος της παραγοντικής ανάλυσης;**

**Απάντηση**
Στόχος της είναι η ανάδειξη των κύριων παραγόντων των μεταβλητών, με την ταυτόχρονη διατήρηση της πληροφορίας και τη μείωση των διαστάσεων του συνόλου των δεδομένων.

## Κριτήριο αξιολόγησης 3

**Πώς γίνεται ο έλεγχος για υπόθεση καταλληλότητας των δεδομένων για την παραγοντική ανάλυση;**

**Απάντηση**
Ο έλεγχος για υπόθεση καταλληλότητας των δεδομένων για την παραγοντική ανάλυση, δηλαδή ο έλεγχος των ασυσχέτιστων αρχικών μεταβλητών, γίνεται με το κριτήριο Bartlett's sphericity test. Επίσης, ένας έλεγχος για την καταλληλότητα των δεδομένων είναι ο δείκτης KMO (Kaiser-Meyer-Olkin) που εκτιμά το μέγεθος της ομοιογένειας των μεταβλητών.

## Κριτήριο αξιολόγησης 4

**Πώς επιλέγεται το πλήθος των παραγόντων που θα ερμηνευτεί στην παραγοντική ανάλυση;**

**Απάντηση**
Η επιλογή τους γίνεται από το διάγραμμα scree plot, το οποίο εμφανίζει το ποσοστό διασποράς σε σχέση με το πλήθος παραγόντων. Το σημείο στο οποίο παρατηρείται απότομη πτώση (μεταβολή) στο διάγραμμα αντιστοιχεί στο πλήθος των επιλεγμένων παραγόντων.

## Κριτήριο αξιολόγησης 5

**Πώς επιλέγονται οι μεταβλητές που σχετίζονται με κάθε παράγοντα;**

**Απάντηση**
Οι μεταβλητές που σχετίζονται υψηλά με κάθε παράγοντα είναι αυτές που έχουν υψηλό φορτίο σε αυτόν.

# Κεφάλαιο 8: Λήψη Αποφάσεων Μελλοντικών Χρηματικών Ροών

**Σύνοψη**

*Στο κεφάλαιο αυτό περιγράφεται η διαχρονική αξία του χρήματος, ώστε να μπορούν οι αναγνώστες να υπολογίζουν και να συγκρίνουν μελλοντικές χρηματικές ροές που χρησιμοποιούνται σε δάνεια και ράντες. Παρουσιάζονται με απλό τρόπο παραδείγματα με έμφαση στην εφαρμογή υπολογισμού παρούσας αξίας.*

**Προαπαιτούμενη γνώση**

*Δεν απαιτείται.*

## 8.1 Διαχρονική αξία χρήματος

**Χρήμα** είναι το μέσο ανταλλαγής και ταυτόχρονα το μέτρο αξίας αγαθών και υπηρεσιών. Η ιστορία του είναι μεγάλη. Οι μορφές του πολλές κατά τη διάρκεια των χρόνων, έως ότου φτάσει στη σημερινή του μορφή (Εικόνα 8.1) με την εκτύπωση τραπεζογραμματίων και με τη νέα του άυλη μορφή.

***Εικόνα 8.1*** *Νομίσματα και τραπεζογραμμάτια.*

«Η πρώτη μορφή χρήματος που χρησιμοποιήθηκε στην οικονομική ιστορία, γύρω στο 9.000 π.Χ. με 6.000 π.Χ., ήταν τα βοοειδή και το καλαμπόκι! Στη συνέχεια, ως χρήμα χρησιμοποιήθηκαν κάποια σπάνια κοχύλια, τα δέρματα ορισμένων ζώων, το κεχριμπάρι, τα φτερά, οι χάντρες, τα αυγά, τα όστρακα, το τσάι, οι σιδερένιες ράβδοι, ο καπνός και το αλάτι. Με την πάροδο του χρόνου, οι οικονομίες γίνονταν πιο πολύπλοκες και δημιουργήθηκε η ανάγκη για έναν πιο σύγχρονο τρόπο συναλλαγών. Έτσι, δημιουργήθηκαν τα νομίσματα. Το πρώτο επίσημο νόμισμα (η λέξη προέρχεται από τον νόμο) στην ιστορία εμφανίστηκε στην αρχαία Ελλάδα, στη Λυδία της Μικράς Ασίας, κατά το τέλος του 7ου με αρχές του 6ου αιώνα προ Χριστού. Ήταν κατασκευασμένο από ένα κράμα χρυσού και αργύρου, γνωστό ως ήλεκτρον, που βρισκόταν σε αφθονία στον ποταμό Πακτωλό. Το πρώτο επίσημο χαρτονόμισμα, νόμισμα δηλαδή από κομμάτι χαρτιού, το οποίο έγινε αποδεκτό ως μέσο πληρωμής, εμφανίστηκε στην Κίνα τον 11ο αιώνα μετά Χριστόν. Ο Φοίνικας ήταν το πρώτο νόμισμα της νεότερης Ελλάδας, το οποίο εισήχθη το 1828 από τον Ιωάννη Καποδίστρια. Πέντε χρόνια αργότερα, ο βασιλιάς Όθων, ως λάτρης της αρχαίας Ελλάδος, αντικατέστησε τον Φοίνικα με τη Δραχμή, η οποία παρέμεινε το επίσημο νόμισμα της χώρας μας μέχρι και το 2002, όταν η Ελλάδα υιοθέτησε ως νόμισμα το Ευρώ. Σήμερα ζούμε σε μια νέα εποχή, την ψηφιακή (digital). Ψηφιακό χρήμα είναι οποιοδήποτε μέσο πληρωμής υπάρχει σε ηλεκτρονική μορφή. Το πρώτο είδος ψηφιακού χρήματος ήταν οι πιστωτικές κάρτες (credit cards). Τα τελευταία χρόνια έχουν γνωρίσει μεγάλη άνθηση τα κρυπτονομίσματα (cryptocurrencies), ένα νέο είδος ψηφιακού νομίσματος, τα οποία έχουν ήδη ξεπεράσει τα 10.000 με πιο γνωστά το Bitcoin και το Ethereum» (https://www.gfli.gr/programma-axia/to-xrima-kai-i-istoria-tou/).

«Στις αρχές του 20ου αιώνα όλες σχεδόν οι χώρες υιοθέτησαν αυτό το σύστημα, όπου για τα πιστοποιητικά που εξέδιδαν υπήρχε προκαθορισμένη ποσότητα χρυσού προς εξαργύρωση. Μετά από τον Β΄

Παγκόσμιο Πόλεμο, με τη διάσκεψη του Bretton Woods, οι περισσότερες χώρες υιοθέτησαν τα χρήματα Fiat, των οποίων η τιμή είχε καθοριστεί σύμφωνα με το δολάριο ΗΠΑ. Το αμερικανικό δολάριο με τη σειρά του καθορίστηκε σε σχέση με τον χρυσό. Το 1971, η κυβέρνηση των ΗΠΑ έπαυσε τη μετατρεψιμότητα του δολαρίου ΗΠΑ σε χρυσού. Μετά από αυτό, πολλές χώρες ακολούθησαν το παράδειγμα των ΗΠΑ και η πλειονότητα των χρημάτων παγκοσμίως σταμάτησε να υποστηρίζεται από αποθέματα χρυσού» (http://www.businesslife.gr/istoria-tou-chrimatos/).

Σε ό,τι αφορά την εξέλιξη του χρήματος: «Στην αρχή, το χρήμα ήταν κατά βάση εμπορευματικό, δηλαδή ένα αντικείμενο κατασκευασμένο από κάποιο υλικό που είχε αγοραία αξία, όπως ένα χρυσό κέρμα. Αργότερα, το χρήμα ήταν αντιπροσωπευτικό, αποτελούνταν δηλαδή από τραπεζογραμμάτια, τα οποία μπορούσαν να ανταλλάσσονται έναντι συγκεκριμένης ποσότητας χρυσού ή αργύρου. Οι σύγχρονες οικονομίες στηρίζονται στο υποχρεωτικό χρήμα. Πρόκειται για το χρήμα το οποίο ορίζεται ως νόμιμο και εκδίδεται από μια κεντρική τράπεζα αλλά, σε αντίθεση με το αντιπροσωπευτικό χρήμα, δεν μπορεί να μετατραπεί σε καθορισμένη ποσότητα χρυσού. Σήμερα, το νόμισμα μπορεί επίσης να υπάρχει και σε άλλη μορφή εκτός από τη φυσική. Το χρήμα μπορεί να υπάρχει σε τραπεζικό λογαριασμό ως ηλεκτρονική καταχώριση ή σε λογαριασμό ταμιευτηρίου ως κατάθεση. Το ψηφιακό ή ηλεκτρονικό χρήμα αποτελεί νομισματική αξία, η οποία είναι αποθηκευμένη, για παράδειγμα, σε μια προπληρωμένη κάρτα ή σε ένα έξυπνο κινητό τηλέφωνο. Οι άμεσες χρεώσεις, οι πληρωμές μέσω διαδικτύου και οι μεταφορές κεφαλαίων μέσω καρτών αποτελούν και αυτές μορφές πληρωμής στις οποίες δεν χρησιμοποιούνται μετρητά. Έχουν μάλιστα κάνει την εμφάνισή τους νέα αποκεντρωμένα ψηφιακά νομίσματα ή συστήματα εικονικών πληρωμών, όπως το Bitcoin, τα οποία λειτουργούν χωρίς να ελέγχονται από μια κεντρική αρχή, για παράδειγμα από μια κεντρική τράπεζα. Νομικά, ωστόσο, δεν θεωρούνται χρήμα» (https://www.ecb.europa.eu/ecb/educational/explainers/tell-me-more/html/what_is_money.el.html).

**Κεφάλαιο** είναι το χρήμα που έχει παραγωγική αξία. Είναι η βάση δημιουργίας επιχείρησης, αφού για να λειτουργήσει χρειάζεται κάποια αρχικά χρήματα τα οποία θα μετατραπούν (ή θα χρησιμοποιηθούν) για την παραγωγή αγαθών και υπηρεσιών. Αναρωτιέται κάποιος αν υπάρχει χρήμα χωρίς παραγωγική αξία, αφού πάντα αυτό μπορεί να χρησιμοποιηθεί για ανταλλαγή με αγαθά ή υπηρεσίες. Η απάντηση στο ερώτημα αυτό είναι απλή. Όταν το χρήμα ανταλλάσσεται έχει παραγωγική αξία. Η κατάθεσή του σε τράπεζα ή σε κάποια επένδυση επίσης έχει παραγωγική αξία, αφού με τον τρόπο αυτό ανταλλάσσεται. Όταν το χρήμα «κρύβεται» (σε μυστική κρύπτη) κυρίως από τον φόβο του ρίσκου, τότε δεν έχει παραγωγική αξία, ίσως θα μπορεί να έχει στο μέλλον παραγωγική αξία όταν βγει από την «κρυψώνα» του και χρησιμοποιηθεί. Στη συνέχεια θα γίνεται αναφορά μόνο σε κεφάλαιο, δηλαδή σε χρήμα που «κυκλοφορεί» στην οικονομική ζωή.

### 8.1.1 Ερωτήματα για μελλοντικά κεφάλαια

Για τα χρηματικά κεφάλαια, θα γίνει προσπάθεια να απαντηθούν τα παρακάτω ερωτήματα:

1. Ποια η μελλοντική αξία κεφαλαίων που τοκίζονται ή δανείζονται;
2. Ποιο το ποσό αποζημίωσης με καθυστέρηση 4 ετών;
3. Ποια είναι η αξία σήμερα μελλοντικών περιοδικών καταθέσεων;
4. Με ποιον τρόπο συμφέρει να αποσβεστεί δάνειο;
5. Ποια η σημερινή αξία γραμματίων ή συναλλαγματικών;
6. Τι είναι η εγγραφή (δημόσια) για αγορά ομολόγων;
7. Μεταξύ τριών δυνατοτήτων επένδυσης ποια να επιλεγεί;
8. Συμφέρει να επενδυθούν διαθέσιμα χρήματα ή να κατατεθούν σε προθεσμιακή κατάθεση;

### 8.1.2 Τόκος και επιτόκιο

Το κεφάλαιο έχει παραγωγική αξία και παράγει αγαθά ή υπηρεσίες. Το ποσό που παράγει ένα κεφάλαιο σε ορισμένο χρόνο ή διαφορετικά η αμοιβή για τη χρήση του ονομάζεται **τόκος**. Προέκυψε από τη λέξη τίκτω (γεννώ, παράγω). Ο τόκος είναι ανάλογος του κεφαλαίου και ανάλογος του χρόνου που αυτό χρησιμοποιείται.

Ως μονάδα για τη μέτρηση του τόκου χρησιμοποιείται το **επιτόκιο**. Είναι ο τόκος που παράγει μία μονάδα κεφαλαίου στην περίοδο μίας μονάδας χρόνου. Εκφράζεται σε ποσοστό ανά μονάδα χρόνου, ώστε να είναι εύκολη η σύγκριση μεταξύ διαφορετικών επιτοκίων. Όταν έχουμε ετήσιο επιτόκιο 2%, αυτό σημαίνει ότι η χρήση 100 μονάδων κεφαλαίου για ένα έτος, θα παράγει τόκο 2 μονάδων κεφαλαίου. Αν το επιτόκιο εκφράζεται ως 2% τον μήνα, αυτό σημαίνει ότι η χρήση 100 μονάδων κεφαλαίου για έναν μήνα, θα παράγει τόκο 2 μονάδων κεφαλαίου.

#### 8.1.2.1 Απλός και σύνθετος τόκος

Στις βραχυχρόνιες (μικρότερες από ένα έτος) οικονομικές πράξεις, υπολογίζεται ο **απλός τόκος**, πολλαπλασιάζοντας κεφάλαιο επί επιτόκιο επί χρόνο. Ο τόκος συμβολίζεται συνήθως με I και ο τύπος του είναι:

$$I = K * i * n \qquad (8.1)$$

όπου i συμβολίζεται το επιτόκιο στη μονάδα του χρόνου, και με n συμβολίζονται οι μονάδες του χρόνου για τον οποίο χρησιμοποιήθηκε το κεφάλαιο K.

Στις υπόλοιπες οικονομικές πράξεις, το κεφάλαιο μεταβάλλεται στη διάρκεια του χρόνου ενσωματώνοντας τον τόκο της προηγούμενης χρονικής περιόδου. Στην περίπτωση αυτή, και ο τόκος της προηγούμενης χρονικής περιόδου γίνεται κεφάλαιο και παράγει νέο τόκο με τη σειρά του. Ο συνολικός τόκος που προκύπτει ονομάζεται σύνθετος τόκος και όλη η διαδικασία ανατοκισμός. Στον ανατοκισμό δεν υπολογίζουμε τον τόκο αλλά το συνολικά τελικό κεφάλαιο που προκύπτει και το οποίο συμβολίζεται με $K_n$, όταν το αρχικό κεφάλαιο είναι $K_0$ και ο τύπος του είναι:

$$K_n = K_0(1 + i)^n \qquad (8.2)$$

Επειδή ο χρόνος n είναι στον εκθέτη στον τύπο (8.2), με τον ανατοκισμό το τελικό κεφάλαιο μεγαλώνει πολύ γρήγορα στη διάρκεια του χρόνου. Ο συντελεστής $(1+i)^n$ ονομάζεται συντελεστής ανατοκισμού. Για ευκολία στις πράξεις είναι ήδη υπολογισμένος σε πίνακα ανάλογα με τις τιμές i, n (αρχείο excel συντελεστής ανατοκισμού_PINAKAS_YPOLOGISMOU.xls) για εύκολο υπολογισμό του. Διαφορετικά γίνεται η πράξη ύψωση σε δύναμη για να υπολογιστεί.

Παραδείγματα υπολογισμού τελικού κεφαλαίου με ανατοκισμό υπάρχουν στην ενότητα 8.8 προς το τέλος του κεφαλαίου.

#### 8.1.2.2 Συχνότητα ανατοκισμού

Στον σχηματισμό του τελικού κεφαλαίου επιδρά το αρχικό κεφάλαιο, το επιτόκιο, η χρονική περίοδος αλλά και το πόσο συχνά ενσωματώνεται ο τόκος κάθε περιόδου στο κεφάλαιο και γίνεται εκ νέου ανατοκισμός. Συνήθως, η συχνότητα ανατοκισμού στις τραπεζικές καταθέσεις είναι το εξάμηνο. Στον υπολογισμό των χρεών πιστωτικών καρτών, ο ανατοκισμός γίνεται κάθε μήνα.

Προκειμένου να φανεί η επίδραση της χρονικής συχνότητας του ανατοκισμού στο τελικό κεφάλαιο, υπολογίζεται τελικό κεφάλαιο για αρχικό κεφάλαιο 3.000 ευρώ που ανατοκίζεται με επιτόκιο 2% κάθε εξάμηνο για 4 έτη. Επειδή η συχνότητα ανατοκισμού είναι το εξάμηνο, ο χρόνος θα υπολογιστεί σε εξάμηνα και τα 4 έτη αντιστοιχούν σε 8 εξάμηνα. Σύμφωνα με τον τύπο (8.2) υπολογίζεται το τελικό κεφάλαιο:

$$K_n = K_0(1 + i)^n = 3.000(1 + 0{,}02)^8 = 3.000 * 1{,}1717 = 3.515{,}1 \text{ ευρώ}$$

Αν το ίδιο κεφάλαιο ανατοκιστεί με ετήσιο ανατοκισμό και επιτόκιο 4% κάθε έτος, για 4 έτη θα προκύψει τελικό κεφάλαιο που υπολογίζεται με τον τύπο (8.2):

$$K_n = K_0(1 + i)^n = 3.000(1 + 0{,}04)^4 = 3.000 * 1{,}1699 = 3.509{,}7 \text{ ευρώ}$$

Παρατηρείται ότι το ίδιο αρχικό κεφάλαιο (3.000 ευρώ) για τον ίδιο χρόνο (4 έτη) με ανάλογο επιτόκιο (2% το εξάμηνο και 4% το έτος) δίνει διαφορετικό τελικό κεφάλαιο. Η διαφορά οφείλεται στη συχνότητα ανατοκισμού, δηλαδή σε ποια χρονική περίοδο ο τόκος ενσωματώνεται στο κεφάλαιο και παράγει με τη σειρά του νέο τόκο. Όσο συντομότερη είναι η συχνότητα ανατοκισμού, τόσο πιο μεγάλο γίνεται το τελικό κεφάλαιο.

#### 8.1.2.2.1 Πραγματικό επιτόκιο

Για να μετρηθεί και η επίδραση της συχνότητας ανατοκισμού, όταν το επιτόκιο εκφράζεται σε διαφορετικές χρονικές περιόδους, δεν μετατρέπεται σε ανάλογο επιτόκιο αλλά σε πραγματικό επιτόκιο περιόδου. Αν το ετήσιο <u>ονομαστικό</u> επιτόκιο είναι 8% και ο ανατοκισμός γίνεται κάθε εξάμηνο (αντιστοιχεί ανάλογο εξαμηνιαίο επιτόκιο 8/2= 4%). **Πραγματικό ετήσιο επιτόκιο** EIR ονομάζεται το ετήσιο επιτόκιο που θα δώσει ίδιο τελικό κεφάλαιο σε ετήσιο ανατοκισμό με το τελικό κεφάλαιο εξαμηνιαίου ανατοκισμού. Υπολογίζεται από τη σχέση:

$$Ko\ (1+EIR) = Ko(1+0{,}04)2 \Leftrightarrow (1+EIR) = (1+0{,}04)2 \Leftrightarrow$$

$$EIR = (1+0{,}04)2\text{-}1 = 1{,}0816\text{-}1 = 0{,}0816 = 8{,}16\%$$

Επομένως το πραγματικό ετήσιο επιτόκιο για εξαμηνιαίο επιτόκιο 4% είναι 8,16% και όχι 8% που είναι το ανάλογο ονομαστικό επιτόκιο.

Παραδείγματα υπολογισμού πραγματικού επιτοκίου υπάρχουν στην ενότητα 8.8 προς το τέλος του κεφαλαίου.

## 8.2 Επιτόκιο και μελλοντική αξία

Στην οικονομική ζωή, όταν χρησιμοποιείται κεφάλαιο, πάντα παράγει τόκο για τον χρόνο χρήσης του. Είτε ο τόκος αυτός προκύπτει από την κατάθεσή του στην τράπεζα, είτε προκύπτει ως αμοιβή για τον δανεισμό του, είτε ως απόδοση του για κάποια επένδυση.

Αν θεωρηθεί ως i% ή r% το επιτόκιο απόδοσης τόκου για κεφάλαιο (με αρχικό ποσό $K_0$), καθώς περνάει ο χρόνος, το νέο κεφάλαιο θα αλλάζει και θα αυξάνεται με τον αντίστοιχο τόκο του. Διαχρονικά θα προκύπτει όλο και μεγαλύτερη αξία κεφαλαίου στη διάρκεια του χρόνου, εξ ου και ρήση: ο χρόνος είναι χρήμα.

Στην περίπτωση κατά την οποία πρέπει να συγκριθούν ή να υπολογιστούν αξίες κεφαλαίων σε διαφορετικές χρονικές στιγμές, θα ισχύει η εξής βασική αρχή: Σε κάθε χρονική στιγμή, οι παρούσες αξίες (της συγκεκριμένης χρονικής στιγμής) είναι συγκρίσιμες μεταξύ τους, ενώ οι μελλοντικές αξίες (σε άλλες χρονικές στιγμές) δεν είναι άμεσα συγκρίσιμες. Για να γίνει σύγκριση κεφαλαίων διαφορετικών χρονικών στιγμών, χρησιμοποιείται η εξίσωση της οικονομικής ισοδυναμίας μεταξύ δύο χρηματικών ποσών Kn (σε n χρονικές περιόδους από το παρόν) και K0 (αξία στο παρόν).

Η εξίσωση της οικονομικής ισοδυναμίας γενικά περιγράφεται από τη σχέση:

**Αξία παρόντος κεφαλαίου = Αξία μελλοντικού κεφαλαίου χωρίς τον ενδιάμεσο τόκο** (για το χρονικό διάστημα μεταξύ παρόντος και μέλλοντος)

και ειδικότερα αν το επιτόκιο είναι i% και n το χρονικό διάστημα από το παρόν έως το μελλοντικό κεφάλαιο, θα ισχύει σύμφωνα με τον τύπο (8.2)

$$K_n = K_0(1+i)^n$$

οπότε η αξία παρόντος κεφαλαίου $K_0 = K_n\ 1/(1+i)^n$ ή αν συμβολίσουμε με $U^n = 1/(1+i)^n$ έχουμε:

$$K_0 = K_n\ U^n \qquad (8.3)$$

Η αξία του παρόντος κεφαλαίου K0 ονομάζεται παρούσα αξία. Ο συντελεστής $U^n$ ονομάζεται συντελεστής προεξόφλησης. Για ευκολία στις πράξεις είναι ήδη υπολογισμένος σε πίνακα ανάλογα με τις τιμές i, n (παράρτημα με αρχείο excel). Μπορεί να εντοπιστεί εκεί ή διαφορετικά να γίνει η πράξη ύψωση σε δύναμη και διαίρεση, για να υπολογιστεί. Παραδείγματα υπολογισμού παρούσας αξίας υπάρχουν στην ενότητα 8.8 προς το τέλος του κεφαλαίου.

Η εξίσωση (8.3) είναι γνωστή ως τύπος υπολογισμού της παρούσας αξίας μελλοντικού κεφαλαίου.

## 8.3 Προεξόφληση

Κάποιες φορές χρειάζεται να υπολογιστεί η αξία στο παρόν ενός κεφαλαίου που θα πληρωθεί ή θα εισπραχθεί στο μέλλον. Π.χ., όταν πρέπει να εξοφληθεί ένα δάνειο νωρίτερα ή όταν πρόκειται να πληρωθεί νωρίτερα κάποια μελλοντική υπόσχεση πληρωμής ή όταν χρειαστεί αποτίμηση στο παρόν μελλοντικών εσόδων ή εξόδων.

Η διαδικασία αυτή ονομάζεται **προεξόφληση**. Υπολογίζεται δηλαδή η παρούσα αξία ενός μελλοντικού κεφαλαίου με τη σχέση (8.3).

## 8.4 Γραμμάτια - συναλλαγματικές

Στις εμπορικές συναλλαγές κάποιες φορές αγοράζονται προϊόντα ή υπηρεσίες και ο αγοραστής, μην έχοντας αρκετό κεφάλαιο ώστε να τα πληρώσει, υπόσχεται να εξοφλήσει την αξία τους στο μέλλον. Ο πωλητής (πιστωτής) για να εξασφαλίσει την απαίτησή του, υποχρεώνει τον αγοραστή (χρεώστη) να υπογράψει ειδικό έγγραφο, με το οποίο μπορεί στο μέλλον να εισπράξει το οφειλόμενο χρηματικό ποσό. Η υπόσχεση του αυτή γίνεται με την υπογραφή κατάλληλου νομικού εγγράφου.

Με τον Νόμο 5325 του 1932 έχουν καθιερωθεί δύο τύποι τέτοιων εγγράφων, τα γραμμάτια εις διαταγή και οι συναλλαγματικές. Το γραμμάτιο εις διαταγή συντάσσεται από τον οφειλέτη, ο οποίος υπόσχεται να πληρώσει συγκεκριμένο χρηματικό ποσό, σε ορισμένο τόπο και χρόνο στον πιστωτή. Συναλλαγματική είναι το έγγραφο που υπογράφει ο πιστωτής (εκδότης) και με το οποίο δίνει εντολή στον οφειλέτη (αποδέκτη) να πληρώσει στον κομιστή του εγγράφου (ή για λογαριασμό του), σε ορισμένο χρόνο και τόπο, το ποσό που είναι γραμμένο στη συναλλαγματική. Θεωρούνται πιστωτικοί τίτλοι και νομικά είναι όμοιοι. Είναι αυστηρά τυπικά έγγραφα. Στο γραμμάτιο ο οφειλέτης υπόσχεται στον πιστωτή ότι θα εξοφλήσει το αναγραφόμενο ποσό, ενώ με τη συναλλαγματική ο πιστωτής διατάσσει τον οφειλέτη να πληρώσει το αναγραφόμενο ποσό. Η διαφορά τους είναι τυπική και στις εμπορικές συναλλαγές αναφέρονται συνήθως με κοινό όνομα «εμπορικά γραμμάτια» ή απλώς «γραμμάτια».

Κάποιες φορές, έμποροι που έχουν στην κατοχή τους γραμμάτια ή συναλλαγματικές χρειάζονται χρήματα για να πληρώσουν άλλες οφειλές τους. Θα πρέπει να «προεξοφλήσουν» τα γραμμάτια και να πάρουν ποσό μικρότερο από την ονομαστική τους αξία. Αυτό είναι δίκαιο, εφόσον το ποσό που θα πάρουν, δυνητικά μπορούν να το καταθέσουν σε τράπεζα και να εισπράξουν τόκο. Πόσο μικρότερο ποσό εισπράττουν; Εξαρτάται από τον χρόνο που απομένει έως ότου «λήξουν», δηλαδή πόσος χρόνος μένει μέχρι να έρθει η ημερομηνία που αναγράφουν ότι πρέπει να πληρωθούν. Όσο μεγαλύτερος είναι ο χρόνος, τόσο μικραίνει το ποσό που θα πάρουν στην προεξόφληση, αφού αφαιρείται ο τόκος του αντίστοιχου χρονικού διαστήματος.

Ο επιχειρηματίας που κατέχει μία ή περισσότερες συναλλαγματικές (ή γραμμάτια) έχει τις εξής επιλογές: Να την κρατήσει μέχρι την ημερομηνία λήξης και να εισπράξει το αναγραφόμενο ποσό από τον οφειλέτη. Να τη μεταβιβάσει σε κάποιον άλλο οπισθογραφώντας την. Να αναθέσει σε μια τράπεζα την είσπραξη, πληρώνοντας και κάποια προμήθεια. Να την προεξοφλήσει (ρευστοποιήσει) στην τράπεζά του, πληρώνοντας διάφορα έξοδα και τους τόκους από την ημέρα προεξόφλησης έως την ημέρα λήξης. Μια συναλλαγματική (ή γραμμάτιο) έχει δύο αξίες, την **ονομαστική αξία** (αυτή που αναγράφεται) και την **παρούσα αξία** (αυτή με την οποία γίνεται η προεξόφληση πριν από τη λήξη της). Στην ημερομηνία λήξης, ονομαστική και παρούσα αξία είναι ίσες, ενώ πριν από την ημερομηνία λήξης η παρούσα αξία είναι πάντα μικρότερη από την ονομαστική αξία. Η παρούσα αξία υπολογίζεται με τη σχέση (8.3), όπου i είναι το προεξοφλητικό επιτόκιο (όπως έχει συμφωνηθεί συνήθως από την τράπεζα) και n είναι ο χρόνος από την ημερομηνία προεξόφλησης έως την ημερομηνία λήξης που αναγράφεται στο νομικό έντυπο.

## 8.5 Παρούσα αξία πολλών μελλοντικών ροών

Μπορεί να γίνει υπολογισμός πολλών διαφορετικών κεφαλαίων σε διαφορετικές χρονικές στιγμές, υπολογίζοντας για το καθένα την παρούσα αξία του και κατόπιν αθροίζοντας όλες τις παρούσες αξίες τους (Φλώρου, 2015).

Συμβαίνει μερικές φορές ο οφειλέτης που οφείλει πολλά διαφορετικά γραμμάτια να ζητήσει από τον δανειστή την αντικατάσταση δύο ή περισσότερων γραμματίων με ένα ενιαίο νέο γραμμάτιο. Στην περίπτωση αυτή, για να μην χάσουν ούτε ο δανειστής αλλά ούτε και ο οφειλέτης, για να υπάρχει δίκαιη αντικατάσταση, το νέο ενιαίο κοινό γραμμάτιο μπορεί να έχει διαφορετική ημερομηνία πληρωμής, αλλά θα πρέπει να είναι **οικονομικώς ισοδύναμο** με τα γραμμάτια που αντικαθιστά. Η εύρεση του οικονομικώς ισοδυνάμου γίνεται με χρήση της αρχής της οικονομικής ισοδυναμίας. Η αρχή αυτή διατυπώνεται ως εξής: «Το άθροισμα των παρουσών αξιών των αντικαθιστάμενων γραμματίων ισούται με την παρούσα αξία του ενιαίου γραμματίου σε ορισμένη χρονική στιγμή και με το ίδιο επιτόκιο». Εκφράζεται ως εξίσωση με τη σχέση:

Παρούσα αξία ενιαίου νέου γραμματίου = άθροισμα παρουσών αξιών αντικαθιστάμενων γραμματίων

Στην παραπάνω σχέση, εποχή ισχύος της ισοδυναμίας (ισότητας) είναι η χρονική στιγμή που υπολογίζουμε τις αξίες αντικατάστασης. Η περίοδος μέτρησης του επιτοκίου (συχνότητα ανατοκισμού) θα πρέπει να συμπίπτει με τις περιόδους μέτρησης του χρόνου από την εποχή της ισοδυναμίας έως την ημερομηνία ή τις ημερομηνίες λήξης του ή των γραμματίων που αντικαθίστανται.

## 8.6 Ράντες

Κάποιες φορές ένα κεφάλαιο σχηματίζεται όχι άμεσα, αλλά καταθέτοντας ή αποταμιεύοντας χρηματικά ποσά σε τακτά χρονικά διαστήματα. Εφαρμογές τέτοιων περιοδικών κεφαλαίων είναι ο σχηματισμός κεφαλαίου με ισόποσες καταθέσεις, η εξόφληση χρέους με δόσεις, οι μηνιαίες κρατήσεις μισθωτών. Στην περίπτωση αυτή, μας ενδιαφέρει η συνολική αξία όλου του πλήθους χρηματικών ποσών τα οποία αντιστοιχούν σε διαφορετικές χρονικές περιόδους. Η σειρά κεφαλαίων τα οποία κατατίθενται ή καταβάλλονται σε διαφορετικά χρονικά διαστήματα ονομάζεται **ράντα**. Όρος ή δόση (Rent) είναι το ποσό που καταβάλλεται σε κάθε χρονικό διάστημα, συμβολίζεται με R και συνήθως είναι το ίδιο αριθμητικά. Ο χρόνος μεταξύ δύο διαδοχικών δόσεων ονομάζεται περίοδος της ράντας, και το πλήθος των δόσεων συμβολίζεται με n. Λήξη της περιόδου είναι η χρονική στιγμή που κατατίθεται ή πληρώνεται η δόση.

Η ράντα δεν αντιμετωπίζεται σαν διαφορετικά κεφάλαια, αλλά όλες οι δόσεις μαζί έχουν μία συνολική αξία. Στην αρχή της ονομάζεται αρχική αξία, στο τέλος της ονομάζεται τελική αξία. Η αρχική και τελική αξία υπολογίζονται με βάση τη δόση R, το πλήθος n των δόσεων και το επιτόκιο i της ράντας. Είναι αριθμητικά διαφορετικές μεταξύ τους, αλλά οι δύο αξίες είναι οικονομικά ισοδύναμες, αφού ισχύει για αυτές η αρχή της οικονομικής ισοδυναμίας. Φυσικά, με βάση την αρχή της οικονομικής ισοδυναμίας, μπορεί να υπολογιστεί η αξία της ράντας και σε οποιαδήποτε άλλη χρονική στιγμή (Φλώρου, 2015).

## 8.7 Δάνεια

Δάνειο είναι κεφάλαιο που παραχωρείται με ορισμένους όρους, οι οποίοι λέγονται όροι δανεισμού. Δανειστής είναι όποιος παραχωρεί το κεφάλαιο, ενώ δανειζόμενος (οφειλέτης) όποιος παίρνει το κεφάλαιο. Οι όροι δανεισμού αναφέρονται στον τόπο, στον χρόνο επιστροφής του δανείου και στον τόκο που πρέπει να καταβάλλει ο δανειζόμενος στον δανειστή. Υπάρχουν πολλές κατηγορίες δανείων. Διακρίνονται σε σχέση με τη διάρκειά τους σε βραχυπρόθεσμα και μακροπρόθεσμα δάνεια, σε σχέση με το πλήθος των δανειστών σε ενιαία και ομολογιακά δάνεια. Ανάλογα με τον τρόπο εξόφλησής τους διακρίνονται σε εξοφλητέα εφάπαξ και σε εξοφλητέα τοκοχρεωλυτικά δάνεια.

Το ποσό που καταβάλλεται από τον οφειλέτη για την εξόφληση του δανείου διαχωρίζεται ανάλογα με τον σκοπό εξόφλησης σε ποσό που καταβάλλεται για το χρέος και λέγεται χρεολύσιο και στο ποσό που καταβάλλεται για την εξόφληση του τόκου. Το άθροισμα των δύο ποσών ονομάζεται **τοκοχρεολύσιο**. Ο τρόπος υπολογισμού του τόκου και χρεολυσίου ονομάζεται **σύστημα απόσβεσης** δανείου (Φλώρου, 2015).

## 8.8 Εφαρμογές για κατανόηση

### 8.8.1 Παράδειγμα υπολογισμού τελικού κεφαλαίου

Πόσο θα γίνει ένα κεφάλαιο 10.000 ευρώ, το οποίο ανατοκίζεται για 5 έτη με ετήσιο επιτόκιο 3%;

Λύση

Έστω Κο = 10.000 το αρχικό κεφάλαιο, i = 3% = 0,03 ετήσιο επιτόκιο, n = 5 έτη. Ο τύπος που μας δίνει το τελικό κεφάλαιο είναι:

$$K_n = K_o\,(1+i)^n = 10.000*\,(1+0,03)^5=$$

από τον πίνακα συντελεστή ανατοκισμού (αρχείο excel) για i = 3% και n = 5 εντοπίζεται συντελεστής ανατοκισμού = 1,15927. Έτσι, $K_n$ = 10.000*1,15927 = 11.592,7 ευρώ θα είναι το τελικό κεφάλαιο.

### 8.8.2 Παράδειγμα υπολογισμού τελικού κεφαλαίου

Πόσο θα γίνει στο τέλος αρχικό κεφάλαιο 10.000 το οποίο ανατοκίζεται για 5 έτη με εξαμηνιαίο επιτόκιο 3%;

Λύση

Έστω Κο = 10.000 το αρχικό κεφάλαιο, i = 3% = 0.03 επιτόκιο εξαμήνου και ο ανατοκισμός γίνεται κάθε εξάμηνο. Στην περίπτωση αυτή και ο χρόνος θα πρέπει να μετρηθεί σε εξάμηνα. Ισχύει n = 5 έτη = 10 εξάμηνα.

Ο τύπος που δίνει το τελικό κεφάλαιο είναι:

$K_n = K_o\,(1+i)^n = 10.000*\,(1+0,03)^{10}$ = από τον πίνακα συντελεστή ανατοκισμού (αρχείο excel) για i = 3% και n = 10 εντοπίζεται συντελεστής ανατοκισμού = 1,3439167.

Έτσι $K_n$ = 10.000*1,343916 = 13.439,16 ευρώ θα είναι το τελικό κεφάλαιο. Δηλαδή κατά πολύ μεγαλύτερο τελικό κεφάλαιο από ό,τι στο παράδειγμα 8.1, επειδή ο ανατοκισμός γίνεται σε μικρότερη χρονική περίοδο (εξάμηνο) και το αντίστοιχο επιτόκιο είναι μεγαλύτερο ανά εξάμηνο 3% (ανάλογο ετήσιο 6%).

### 8.8.3 Παράδειγμα πραγματικού επιτοκίου

Αν το ετήσιο ονομαστικό επιτόκιο είναι 8% και ο ανατοκισμός γίνεται κάθε τρίμηνο (τριμηνιαίο επιτόκιο 8/4 = 2%), ποιο είναι το πραγματικό ετήσιο επιτόκιο EIR;

Λύση

Θα πρέπει να προκύπτει το ίδιο τελικό κεφάλαιο για αξία 1 ευρώ, τόσο με το πραγματικό επιτόκιο όσο και με το τριμηνιαίο επιτόκιο. Δηλαδή πρέπει να ισχύει:

$$(1+EIR) = (1+0,02)^4 \Leftrightarrow EIR= (1+0,002)^4\text{-}1= 0,0824 = 8,24\%$$

Επομένως, το πραγματικό ετήσιο επιτόκιο EIR θα είναι 8,24%.

### 8.8.4 Παράδειγμα πραγματικού επιτοκίου

Αν το ετήσιο ονομαστικό επιτόκιο είναι 8% και ο ανατοκισμός γίνεται κάθε μήνα (μηνιαίο επιτόκιο 8/12 = 0,67%), ποιο είναι το πραγματικό ετήσιο επιτόκιο EIR;

Λύση

Θα πρέπει να προκύπτει το ίδιο τελικό κεφάλαιο για αξία 1 ευρώ, τόσο με το πραγματικό επιτόκιο όσο και με το μηνιαίο επιτόκιο. Δηλαδή πρέπει να ισχύει:

$$(1+EIR) = (1+0{,}0067)12 \Leftrightarrow EIR= (1+0{,}0067)12-1 = 0{,}0834 = 8{,}34\%$$

Επομένως, το πραγματικό ετήσιο επιτόκιο EIR θα είναι 8,34%.

## 8.8.5 Παράδειγμα υπολογισμού παρούσας αξίας

Πόσο κεφάλαιο πρέπει να κατατεθεί σήμερα (παρούσα αξία) το οποίο ανατοκίζεται για 5 έτη με ετήσιο επιτόκιο 3% και γίνεται 15.000 ευρώ;

Λύση

Kn = 15.000 το τελικό κεφάλαιο

i = 3% = 0,03

n = 5 έτη

$K_0$ = το άγνωστο αρχικό κεφάλαιο

Η παρούσα αξία είναι το τελικό κεφάλαιο διά του συντελεστή ανατοκισμού ή διαφορετικά το αρχικό επί τον συντελεστή προεξόφλησης. Επομένως, θα είναι:

$$K_0 = K_n/(1+i)^n = 15.000/ (1+0{,}03)^5 = 15.000*U^5$$

(εντοπίζουμε στον πίνακα προεξόφλησης την τιμή για n = 5 και i = 3%), οπότε προκύπτει $K_0$ = 15.000*0,8626 = 12.939 ευρώ.

## 8.8.6 Παράδειγμα αντικατάστασης γραμματίων

Γραμμάτια 1.000 ευρώ που λήγει σε 2 έτη και 2.000 ευρώ που λήγει σε 5 έτη αντικαθίστανται με ενιαίο γραμμάτιο το οποίο λήγει σε 3 έτη με ετήσιο επιτόκιο 5%. Πόσο είναι το ενιαίο γραμμάτιο;

Λύση

K1 = 1.000

K2 = 2.000

K3 = άγνωστο

n1 = 2 έτη, n2 = 5 έτη, n3 = 3 έτη

1ος τρόπος: Χρησιμοποιώντας το «σήμερα» ως ημέρα υπολογισμού για την οικονομική ισοδυναμία θα έχουμε:

1. Εξίσωση οικονομική ισοδυναμίας $K_nU^3= K_1U^2 + K_2U^5$

Με αντικατάσταση:

$$K_nU^3= K_1U^2 + K_2U^5 \Leftrightarrow$$

$$K_3 = (K_1U^2 + K_2U^5)/U^3 = (1.000*0{,}9070+2.000*0{,}7835)/0{,}8638 = 2.864{,}08$$

2ος τρόπος: Χρησιμοποιώντας την ημέρα λήξης (το οποίο λήγει σε 3 έτη) του νέου γραμματίου, ως ημέρα υπολογισμού για την οικονομική ισοδυναμία, για το πρώτο γραμμάτιο (το οποίο λήγει σε 2 έτη) θα έχει περάσει ήδη ένας χρόνος καθυστέρησης από τη λήξη του και για το δεύτερο (το οποίο λήγει σε 5 έτη) θα χρειάζονται δύο χρόνια ακόμη για να λήξει:

2. Εξίσωση οικονομική ισοδυναμίας $K_n= K_1*(1+i)^{3-2}+ K_2U^{5-3}$

Με αντικατάσταση:

$$K_3 = K_1*(1+i)^1+ K_2U^2 \Leftrightarrow$$

$$K_3 = 1.000*(1+i)^1+2.000*U^2 \Leftrightarrow$$

$$K_3 = 1.000*1,05+2.000*0,90703 = 2.864,06$$

# 8.9 Ασκήσεις για κατανόηση - εφαρμογή

## 8.9.1 Αντικατάσταση γραμματίων

Γραμμάτιο 2.000 ευρώ που λήγει σε 3 έτη αντικαθίσταται με γραμμάτιο το οποίο λήγει σε 2 έτη με ετήσιο επιτόκιο 5%. Ποια θα είναι η ονομαστική αξία του νέου γραμματίου;

Λύση

Επειδή το νέο γραμμάτιο θα πληρωθεί σε 2 έτη αντί για 3 έτη, θα έχει μικρότερο ποσό. Το ποσό $K_2$ του νέου γραμματίου θα αντιστοιχεί στην προεξόφληση του παλαιού $K_2$ κατά ένα έτος με 5%.

Η ονομαστική του αξία θα υπολογιστεί με τον τύπο $K_2= K_1U^1 \Leftrightarrow K_2=2000/(1+0,05) \Leftrightarrow K_2=1904,76$ ευρώ

## 8.9.2 Προεξόφληση

Συναλλαγματική αξίας 10.000 ευρώ η οποία λήγει σε 2 έτη. Το επιτόκιο σήμερα είναι 4% και είναι γνωστό ότι σε ένα έτος το επιτόκιο θα γίνει 4,5%. Πόσα χρήματα θα εισπραχθούν αν προεξοφληθεί η συναλλαγματική σήμερα και πόσα χρήματα θα εισπραχθούν αν προεξοφληθεί σε ένα έτος;

Λύση

α) Αν προεξοφληθεί η συναλλαγματική σήμερα, το ονομαστικό ποσό της K1, θα μειωθεί με τον τόκο που αντιστοιχεί στο 4% για 2 έτη. Έτσι θα εισπραχθεί ποσό:

$K_2= K_1U^2 \Leftrightarrow K_2=10.000/(1+0.04)^2 \Leftrightarrow K_2=9.245,56$ ευρώ

β) Αν προεξοφληθεί η συναλλαγματική σε έναν χρόνο, το ονομαστικό ποσό της K1 θα μειωθεί με τον τόκο που αντιστοιχεί στο 4,5% για 1 έτος. Έτσι σε έναν χρόνο θα εισπραχθεί ποσό:

$K_3= K_1U^1 \Leftrightarrow K_2=10.000/(1+0.045) \Leftrightarrow K_2=9.569,38$ ευρώ

# Βιβλιογραφία/Αναφορές

Φλώρου, Γ. (2015). *Ποσοτικές μέθοδοι στα χρηματοοικονομικά* [Προπτυχιακό εγχειρίδιο]. Κάλλιπος, Ανοικτές Ακαδημαϊκές Εκδόσεις. https://hdl.handle.net/11419/5066

Φλώρου, Γ. (2015). *Ποσοτικές μέθοδοι στα χρηματοοικονομικά* [Προπτυχιακό εγχειρίδιο]. Κάλλιπος, Ανοικτές Ακαδημαϊκές Εκδόσεις. συντελεστής ανατοκισμού_PINAKAS_YPOLOGISMOU.xls (live.com)

Ιστορία χρήματος. Ανακτήθηκε 30 Μαρτίου, 2022, https://www.gfli.gr/programma-axia/to-xrima-kai-i-istoria-tou/

Ιστορία του χρήματος. Ανακτήθηκε 30 Μαρτίου, 2022, http://www.businesslife.gr/istoria-tou-chrimatos/

Η εξέλιξη της φύσης του χρήματος. Ανακτήθηκε 30 Μαρτίου, 2022, https://www.ecb.europa.eu/ecb/educational/explainers/tell-me-more/html/what_is_money.el.html

# Κριτήρια αξιολόγησης

## Κριτήριο αξιολόγησης 1

**Τι είναι το επιτόκιο;**

**Απάντηση**
Είναι ο τόκος που παράγει μια μονάδα κεφαλαίου στην περίοδο μίας μονάδας χρόνου. Εκφράζεται σε ποσοστό ανά μονάδα χρόνου, ώστε να είναι εύκολη η σύγκριση μεταξύ διαφορετικών επιτοκίων.

## Κριτήριο αξιολόγησης 2

**Πώς υπολογίζεται το συνολικό τελικό κεφάλαιο στον ανατοκισμό;**

**Απάντηση**
Στον ανατοκισμό το συνολικό τελικό κεφάλαιο που προκύπτει και το οποίο συμβολίζεται με $K_n$ όταν το αρχικό κεφάλαιο είναι $K_0$ υπολογίζεται με τον τύπο:

$$K_{\mathrm{n}} = K_0(1 + \mathrm{i})^{\mathrm{n}}$$

## Κριτήριο αξιολόγησης 3

**Τι είναι η προεξόφληση;**

**Απάντηση**
Ο υπολογισμός της αξίας στο παρόν ενός κεφαλαίου που θα πληρωθεί ή θα εισπραχθεί στο μέλλον.

## Κριτήριο αξιολόγησης 4

**Τι είναι η ράντα;**

**Απάντηση**
Η σειρά κεφαλαίων τα οποία κατατίθενται ή καταβάλλονται σε διαφορετικά χρονικά διαστήματα ονομάζεται ράντα.

# Κεφάλαιο 9: Αξιολόγηση Επενδύσεων

**Σύνοψη**
*Στο κεφάλαιο αυτό περιγράφονται οι τρόποι αξιολόγησης επενδύσεων και εύρεσης της καλύτερης επένδυσης μεταξύ εναλλακτικών επιλογών. Παρουσιάζονται με απλό τρόπο παραδείγματα, με έμφαση στην εφαρμογή τους στα χρηματοοικονομικά.*

**Προαπαιτούμενη γνώση**
*Κεφάλαιο 8: Λήψη αποφάσεων μελλοντικών χρηματικών ροών, διαχρονική αξία χρήματος.*

## 9.1 Έννοια της επένδυσης

**Επένδυση** είναι η πράξη απόκτησης διαρκούς αγαθού με σκοπό τη χρησιμοποίησή του, κατά τη διάρκεια ζωής του, προς παραγωγή νέων αγαθών ή υπηρεσιών. Είναι η βάση δημιουργίας επιχείρησης. Δεσμεύει κεφάλαια για πολλά χρόνια με την ελπίδα μελλοντικών ωφελειών και συνεπάγεται κινδύνους.

Η χρηματοοικονομική αξιολόγηση των επενδύσεων αναπόφευκτα βασίζεται σε πάρα πολλές οικονομικές, εμπορικές και παραγωγικές παραδοχές, ενώ τα συμπεράσματά της πρέπει με τη σειρά τους να δικαιολογούν με χρηματοοικονομικούς όρους τη σκοπιμότητα της επένδυσης.

## 9.2 Χρηματοδότηση επενδύσεων

### 9.2.1 Πηγές χρηματοδότησης κεφαλαίου

Για κάθε επένδυση συνήθως χρειάζεται κεφάλαιο, κυρίως σε χρήμα ή κάποιες φορές σε ακίνητο ή ανθρώπινο δυναμικό. Οι πηγές για την εξεύρεση κεφαλαίου μπορεί να είναι ίδια κεφάλαια, δανειακά κεφάλαια από ιδιώτες ή τραπεζικός δανεισμός.

Στην περίπτωση επιχειρήσεων μπορεί ακόμη να βρεθούν κεφάλαια από:

- έκδοση ομολογιών,
- προνομιούχες μετοχές,
- ίδια κεφάλαια,
- παρακρατηθέντα κέρδη,
- έκδοση νέων μετοχών.

### 9.2.2 Κόστος κεφαλαίου

Ανάλογα με την πηγή χρηματοδότησης για το κεφάλαιο που χρησιμοποιείται σε μια επένδυση, θα πρέπει να δοθεί κάποια αμοιβή για τη χρήση του. Η αμοιβή αυτή ονομάζεται κόστος κεφαλαίου.

Το κόστος κεφαλαίου είναι η υψηλότερη απόδοση την οποία μπορεί να επιτύχει κάποιος επενδυτής εάν δεν επενδύσει τα χρήματά του στη συγκεκριμένη επένδυση αλλά σε κάποια άλλη. Ή εναλλακτικά είναι το κόστος ευκαιρίας των κεφαλαίων που έχουν όλοι οι επενδυτές της εταιρείας (μέτοχοι και δανειστές). Αν η μόνη εναλλακτική είναι η κατάθεση στην τράπεζα, το κόστος κεφαλαίου είναι το επιτόκιο καταθέσεων (μηδενικό ρίσκο). Όσο μεγαλώνει το ρίσκο της επένδυσης, τόσο μεγαλώνει το κόστος κεφαλαίου.

## 9.3 Τεχνικές αξιολόγησης επενδύσεων

Προκειμένου να γίνει επιλογή μιας επένδυσης, θα πρέπει να προηγηθεί η αξιολόγησή της, δηλαδή η αποτίμηση των μελλοντικών ωφελειών της. Αν υπάρχουν πολλές δυνατές επενδύσεις, θα πρέπει να βρεθεί η βέλτιστη. Για την αξιολόγηση επενδύσεων υπάρχουν δύο βασικές διαδικασίες. Καταρχάς γίνεται ο εντοπισμός όλων των εσόδων (εισροών) και εξόδων (εκροών) που σχετίζονται με τη σχεδιαζόμενη επένδυση (cash flow analysis). **Ταμειακές ροές** ονομάζονται τα έσοδα που αποφέρει η επένδυση, αλλά και τα έξοδα που αυτή απαιτεί. Η διαφορά έσοδα – έξοδα κάθε έτους είναι η καθαρή ταμειακή ροή του αντίστοιχου έτους. Στη συνέχεια ακολουθεί η χρήση μεθόδων και κριτηρίων με βάση τα οποία οι παραπάνω εισροές και εκροές να μπορούν να αξιολογούνται (capital budgeting decision methods).

Η πρώτη διαδικασία, του εντοπισμού των αναμενόμενων εσόδων και εξόδων της επένδυσης, είναι ασφαλώς και η περισσότερο δύσκολη, αυτή που εμπεριέχει τη μεγαλύτερη αβεβαιότητα για τα συμπεράσματα της αξιολόγησης. Στη φάση αυτή καταρτίζονται όλες οι παραδοχές για την επένδυση, πράγμα ιδιαίτερα δύσκολο και με μεγάλη αβεβαιότητα. Στη φάση αυτή εμπλέκονται άτομα διάφορων ετερόκλητων ειδικοτήτων, προκειμένου να αποδοθούν με τον μεγαλύτερο δυνατό πραγματισμό οι πιθανές υποθέσεις της εξεταζόμενης επένδυσης.

Η δεύτερη διαδικασία έχει έναν μεθοδολογικό/αναλυτικό χαρακτήρα, που σκοπό έχει την επεξεργασία των δεδομένων και των παραδοχών της πρώτης φάσης, ώστε η λήψη απόφασης να βοηθείται, βασιζόμενη, ανάμεσα στα άλλα, σε αντικειμενικούς και κατανοητούς τρόπους μέτρησης.

Η βασική παραδοχή στη διαδικασία υπολογισμού των κριτηρίων και των δεικτών αυτών είναι η εξέλιξη του κόστους κεφαλαίου σε όρους του παρόντος μέσα στον χρόνο. Κατά τα άλλα, ο υπολογισμός των δεικτών είναι μια απλή κατά βάση υπόθεση, ενώ τα συμπεράσματα στα οποία αυτοί οδηγούν είναι τελικά τόσο αξιόπιστα όσο ακριβείς και οι υποθέσεις που καταστρώθηκαν στην πρώτη φάση της ανάλυσης (eclass.teiion.gr/modules/course_metadata/info.php?course=DE-DE163).

Οι κυριότερες τεχνικές αξιολόγησης επενδύσεων (με καθαρά χρηματοοικονομικά κριτήρια) είναι:

- καθαρή παρούσα αξία (NPV),
- εσωτερικός βαθμός απόδοσης (IRR),
- περίοδος επανείσπραξης (PP),
- λογιστικός συντελεστής απόδοσης (ARR),
- δείκτης κερδοφορίας (PI).

Οι τεχνικές αυτές θα παρουσιαστούν στη συνέχεια και θα γίνει η εφαρμογή τους μέσω ενός παραδείγματος.

## 9.4 Καθαρή παρούσα αξία

Υπολογίζεται η αξία όλων των μελλοντικών ταμειακών εκροών και εισροών στον χρόνο μηδέν (πριν ξεκινήσει η επένδυση). Για όλες τις μελλοντικές εισροές και εκροές υπολογίζεται η αξία τους στον χρόνο μηδέν, ώστε να μπορεί να γίνει σύγκριση και με άλλες πιθανές επενδύσεις. Η αξία στον χρόνο μηδέν θα ονομάζεται παρούσα αξία στη συνέχεια. Η παρούσα αξία κάθε μελλοντικής εισροής έχει θετικό πρόσημο (έσοδα), ενώ η παρούσα αξία κάθε μελλοντικής εκροής έχει αρνητικό πρόσημο (έξοδα). Το άθροισμα, για μία χρονική περίοδο, όλων των παρουσών αξιών εισροών και εκροών ονομάζεται **Καθαρή Παρούσα Αξία (ΚΠΑ) ή Net Present Value (NPV)** της επένδυσης. Μετράει το πλεόνασμα ή την έλλειψη ταμειακών ροών, σε όρους παρούσας αξίας, σε σχέση με το κόστος κεφαλαίων (cost of funds) που χρησιμοποιήθηκαν για μια επένδυση. Η Καθαρή Παρούσα Αξία (ΚΠΑ) είναι ένα χρήσιμο εργαλείο που χρησιμοποιείται στην οικονομική επιστήμη (economics), στα χρηματοοικονομικά (finance) και στη λογιστική (accounting), για να καθοριστεί αν μια επένδυση ή ένα έργο, κρίνεται συμφέρουσα για να χρηματοδοτηθεί ή όχι. Η παρούσα αξία των αναμενόμενων ταμειακών ροών υπολογίζεται με την προεξόφλησή τους, χρησιμοποιώντας το κατάλληλο προεξοφλητικό επιτόκιο.

Κριτήριο για την αξιολόγηση των εναλλακτικών επενδύσεων αποτελεί ο υπολογισμός της καθαρής παρούσας αξίας καθεμιάς με τη σχέση:

**Καθαρή Παρούσα Αξία = Παρούσα αξία όλων των καθαρών ταμειακών ροών – Κόστος επένδυσης**

Μια επένδυση με θετική καθαρή παρούσα αξία είναι κερδοφόρα, ενώ μια επένδυση με αρνητική καθαρή παρούσα αξία αποφέρει ζημιά και δεν πρέπει να γίνει, αφού θα εισπραχθούν λιγότερα από τα χρήματα που θα επενδυθούν. Καλύτερη από πολλές εναλλακτικές επενδύσεις είναι αυτή με τη μεγαλύτερη θετική Καθαρή Παρούσα Αξία.

Τα πλεονεκτήματα της μεθόδου της καθαρής παρούσας αξίας είναι ότι είναι δυνατόν να συγκρίνονται διαφορετικές επενδύσεις βασιζόμενοι σε ποσοτικά στοιχεία. Επίσης, η μέθοδος μπορεί να χρησιμοποιηθεί και για περισσότερες επενδύσεις που αλληλοσυμπληρώνονται.

Μειονεκτήματα της μεθόδου είναι ότι δεν λαμβάνεται υπόψη η χρονική διάρκεια κάθε επένδυσης. Υπολογίζεται η παρούσα αξία όλων των μελλοντικών εισροών στον χρόνο 0 χωρίς να γίνεται σύγκριση αν στη μία επένδυση χρειάζονται 10 έτη για να εισπραχθούν εισροές, ενώ στην άλλη μπορεί να ληφθούν οικονομικά ισοδύναμες εισροές σε 3 έτη. Η μέθοδος Καθαρής Παρούσας Αξίας δεν διαφοροποιεί τις επενδύσεις ως προς τις απαιτήσεις τους σε αρχικό κεφάλαιο, αφού δεν συγκρίνει το ύψος αρχικού κεφαλαίου μεταξύ επενδύσεων. Μια επένδυση που απαιτεί 20.000 ευρώ αρχικό κεφάλαιο ίσως να έχει μεγαλύτερη Καθαρή Παρούσα Αξία από μία άλλη που απαιτεί 1.000 ευρώ αρχικό κεφάλαιο. Η δεύτερη όμως επένδυση ίσως έχει μικρότερο ρίσκο και μικρότερες ανάγκες δανεισμού.

Άλλα μειονεκτήματα είναι ο τρόπος εκτίμησης των ταμειακών ροών και η δυσκολία να εκτιμηθούν με ακρίβεια, καθώς εξαρτώνται από πάρα πολλές μεταβλητές και από τις υποκείμενες εκτιμήσεις των αναλυτών. Επίσης, το προεξοφλητικό επιτόκιο (discount rate) που θα πρέπει να χρησιμοποιηθεί δεν είναι πάντοτε σαφές και συνήθως θεωρείται σταθερό κατά τη διάρκεια ζωής της επένδυσης. Αυτό όμως δεν ισχύει για μακροχρόνιες και υψηλού ρίσκου επενδύσεις, καθώς στη διάρκεια ζωής τους αλλάζει το κόστος κεφαλαίου.

## 9.5 Εσωτερικός βαθμός απόδοσης (IRR)

Μια άλλη μέθοδος αξιολόγησης επενδύσεων είναι η μέθοδος του εσωτερικού βαθμού απόδοσης της επένδυσης. Διεθνώς ονομάζεται **Internal Rate of Return** και συμβολίζεται με τα αρχικά **IRR**. Με τη μέθοδο αυτή δεν χρησιμοποιείται επιτόκιο υπολογισμού της αξίας της επένδυσης, αλλά υπολογίζεται το επιτόκιο r, το οποίο ονομάζεται **εσωτερικός βαθμός απόδοσης** της επένδυσης, χρησιμοποιώντας την αρχική αξία της επένδυσης και την τελική της αξία. Το επιτόκιο που αντιστοιχεί σε μηδενική παρούσα αξία είναι ο εσωτερικός ρυθμός απόδοσης. Αν ο εσωτερικός βαθμός απόδοσης της επένδυσης είναι μεγαλύτερος από το ισχύον κόστος κεφαλαίου (επιτόκιο), τότε αξιολογείται η επένδυση ως κερδοφόρα. Διαφορετικά, η απόδοσή της είναι αρνητική. Καλύτερη από πολλές εναλλακτικές επενδύσεις είναι αυτή με τον μεγαλύτερο εσωτερικό βαθμό απόδοσης (Φλώρου, 2015).

«Το Internal Rate of Return (IRR) είναι το προεξοφλητικό επιτόκιο που μηδενίζει την καθαρή παρούσα αξία μιας επένδυσης. Επομένως, όσο μεγαλύτερη είναι η καθαρή παρούσα αξία της επένδυσης, τόσο μεγαλύτερο θα είναι και το IRR. Το IRR χρησιμοποιείται για να συγκρίνει επενδύσεις και διαφέρει από την καθαρή παρούσα αξία, διότι μετράει και τον χρόνο των ταμειακών ροών, όχι μόνο τον συνολικό όγκο. Για παράδειγμα, δύο επενδύσεις με καθαρή παρούσα αξία 1.000.000€ μπορούν να έχουν διαφορετικά IRR, και η επένδυση με το μεγαλύτερο IRR επιστρέφει χρήματα γρηγορότερα από αυτή με το χαμηλότερο IRR» (IRR ependysopedia, 2015).

«Το πλεονέκτημα της χρήσης της μεθόδου του εσωτερικού βαθμού απόδοσης είναι ότι δεν εξαρτάται από κάποιο εξωτερικό επιτόκιο. Βασικό μειονέκτημα της μεθόδου είναι ότι δεν διαφοροποιείται μεταξύ επενδύσεων διαφορετικής κλίμακας. Μια επένδυση μπορεί να περιλαμβάνει χρηματικές ροές σε χιλιάδες ευρώ, ενώ μία άλλη σε δεκάδες ευρώ, αλλά και οι δύο να έχουν τον ίδιο εσωτερικό βαθμό απόδοσης» (Bradley, 2014).

Στον υπολογισμό του εσωτερικού βαθμού απόδοσης, δεν χρησιμοποιείται το κόστος κεφαλαίου, αλλά αναζητείται το επιτόκιο εκείνο (με δοκιμές) με το οποίο θα προκύψει μηδενική συνολική παρούσα αξία επένδυσης. Αυτό έχει κάποιο υπολογιστικό κόστος, αλλά ταυτόχρονα η μέθοδος είναι ανεξάρτητη από το πρόσκαιρο κόστος κεφαλαίου.

## 9.6 Περίοδος επανείσπραξης (PP)

Υπολογίζεται η παρούσα αξία όλων των ταμειακών εισροών μιας επένδυσης στην αρχή της. Κατόπιν, εντοπίζεται η χρονική στιγμή κατά την οποία εισπράττονται οι αρχικές εκροές της. Δηλαδή σε ποια χρονική στιγμή οι ταμειακές εισροές θα είναι ίσες με την αρχική εκροή της επένδυσης. Μετά από τη χρονική αυτή στιγμή αρχίζει να έχει κέρδος η επένδυση.

Με τη μέθοδο αυτή γίνεται προσπάθεια να απαλειφθεί το μειονέκτημα της μεθόδου της Καθαρής Παρούσας Αξίας και να ληφθεί υπόψη ο χρόνος που απαιτεί μια επένδυση για να δώσει κέρδος. Αξιολογείται η ταχύτητα επιστροφής του αρχικού ποσού της επένδυσης και δίνεται βαρύτητα στη δέσμευση χρημάτων για μικρό χρονικό διάστημα. Καλύτερη μεταξύ διαφορετικών επενδύσεων θεωρείται αυτή που έχει τον μικρότερο χρόνο επανείσπραξης.

Μειονέκτημα της μεθόδου της περιόδου επανείσπραξης είναι ότι δεν αξιολογούνται καθόλου εισροές ή εκροές που συμβαίνουν μετά από την περίοδο επανείσπραξης. Δηλαδή μία επένδυση με χρόνο επανείσπραξης δύο έτη ίσως να αξιολογηθεί πιο συμφέρουσα από μία άλλη με χρόνο επανείσπραξης τρία έτη, χωρίς να ληφθεί υπόψη ότι η δεύτερη επένδυση στα τέσσερα έτη μπορεί να φέρει συνολικά τις διπλάσιες ταμειακές εισροές από ό,τι η πρώτη επένδυση.

## 9.7 Λογιστικός συντελεστής απόδοσης (ARR)

Με τη μέθοδο αυτή εκτιμάται η συνολική απόδοση μιας επένδυσης, λαμβάνοντας υπόψη τις ετήσιες ταμειακές ροές και τον χρόνο ζωής της επένδυσης. Χρησιμοποιείται η λογιστική αποτίμηση και δεν λαμβάνεται υπόψη η μετατροπή σε παρούσα αξία των ταμειακών εισροών.

Η **μέση καθαρή εισροή** μιας επένδυσης είναι το άθροισμα εισροών-εκροών διά τον χρόνο ζωής της επένδυσης. Αντιστοιχεί στον ετήσιο κατά μέσο όρο κέρδους της επένδυσης.

Η υπολειμματική αξία είναι η αξία της επένδυσης στη λήξη του χρόνου ζωής της, όταν μπορεί να μεταβιβαστεί και να εισπραχθεί κάποιο ποσό. Παράδειγμα, στην αγορά ενός αυτοκινήτου, το αρχικό του κόστος μπορεί να είναι 20.000 ευρώ, να χρησιμοποιηθεί για 8 χρόνια και κατόπιν να πωληθεί για 2.000 ευρώ. Τα 2.000 ευρώ αντιστοιχούν στην υπολειμματική του αξία μετά από τα 8 έτη που θεωρείται ο χρόνος ζωής του. Πολλές επενδύσεις έχουν μηδενική υπολειμματική αξία όταν μετά από τον χρόνο ζωής τους δεν μπορούν να πωληθούν ή να χρησιμοποιηθούν με άλλον τρόπο.

Το **μέσο κόστος** μιας επένδυσης είναι το άθροισμα αρχικής αξίας (αρνητικό πρόσημο) και υπολειμματικής αξίας της (θετικό πρόσημο) διά τον χρόνο ζωής της επένδυσης. Αντιστοιχεί στο ετήσιο κατά μέσο όρο κόστος της επένδυσης.

Ο **λογιστικός συντελεστής απόδοσης** είναι το κλάσμα της μέσης καθαρής εισροής διά το μέσο κόστος της επένδυσης. Συμβολίζεται με τα αρχικά ARR, δεν έχει μονάδες μέτρησης και υπολογίζεται με μία διαίρεση.

Πλεονέκτημα της μεθόδου του λογιστικού συντελεστή απόδοσης είναι ότι εστιάζει στη συνολική απόδοση της επένδυσης. Μειονέκτημα είναι ότι δεν λαμβάνει υπόψη το κόστος κεφαλαίου, αφού δεν χρησιμοποιεί καθόλου τη διαχρονική αξία του χρήματος με το ανάλογο επιτόκιο. Επίσης, μεταξύ επενδύσεων με ίδιο λογιστικό συντελεστή απόδοσης δεν γίνεται διάκριση, αν η μία απαιτεί πολύ χρόνο για την πραγματοποίησή της ενώ η άλλη γίνεται γρηγορότερα.

## 9.8 Δείκτης κερδοφορίας (PI)

Ο δείκτης κερδοφορίας συμβολίζεται τα αρχικά PI. Είναι το κλάσμα της παρούσας αξίας των εισροών διά της παρούσας αξίας των εκροών της επένδυσης. Μια επένδυση είναι κερδοφόρα όταν PI>1. Όποια επένδυση έχει μεγαλύτερο δείκτη κερδοφορίας είναι καλύτερη.

## 9.9 Εφαρμογές για κατανόηση

Έστω ότι αναζητείται η επιλογή μεταξύ δύο επενδύσεων Κ και Λ. Οι καθαρές ταμειακές ροές των επενδύσεων αυτών για τα επόμενα έτη, παρουσιάζονται στον Πίνακα 9.1.

**Πίνακας 9.1** *Παράδειγμα καθαρών ταμειακών ροών δύο επενδύσεων.*

| Χρόνος | Κ επένδυση | Λ επένδυση |
|---|---|---|
| 0 αρχή | -20.000 (απαιτούμενο αρχικό κεφάλαιο) | -24.000 (απαιτούμενο αρχικό κεφάλαιο) |
| 1 | 10.000 | 3.000 |
| 2 | 7.000 | 4.000 |
| 3 | 6.000 | 9.000 |
| 4 | 3.000 | 12.000 |
| 5 | 0 | 2.000 |
| 6 | 0 | 0 |

Υποθέτοντας ότι οι μελλοντικές ταμειακές ροές των δύο επενδύσεων Κ και Λ έχουν εκτιμηθεί σωστά, τα πιθανά ερωτήματα για την επιλογή της καλύτερης επένδυσης μεταξύ των δύο είναι:

- Ποια θέλει λιγότερα χρήματα;
- Ποια έχει περισσότερα έσοδα;
- Ποια αποφέρει κέρδη πιο γρήγορα;
- Ποια είναι πιο «σίγουρη»;
- Ποια είναι πιο «απαραίτητη»;

Προφανώς, η απάντηση σε καθένα από τα ερωτήματα αυτά δεν δίνει πάντα την ίδια επένδυση ως καλύτερη. Λιγότερα χρήματα αρχικά χρειάζεται η Κ, αλλά τα περισσότερα έσοδα τα έχει η Λ. Η απάντηση στο ποια είναι πιο απαραίτητη ή πιο σίγουρη έχει να κάνει με ποιοτικά κριτήρια και την κρίση του επενδυτή. Στη συνέχεια θα επιλεγεί η βέλτιστη επένδυση, επί τη βάσει μόνο χρηματοοικονομικών κριτηρίων και με τις μεθόδους που αναπτύχθηκαν στις ενότητες 9.4, 9.5, 9.6, 9.7 και 9.8.

### 9.9.1 Επιλογή με τη μέθοδο της Καθαρής Παρούσας Αξίας

Για την εφαρμογή της μεθόδου της καθαρής παρούσας αξίας θα πρέπει να είναι γνωστό το κόστος κεφαλαίου. Ας υποτεθεί ότι το επιτόκιο είναι 10%. Μετατρέπονται όλες οι ταμειακές ροές σε ισοδύναμες τιμές της αρχής του χρόνου, σύμφωνα με τον τύπο $K_0 = K_n/(1+i)^n$

Επειδή υποτέθηκε επιτόκιο 10%, είναι 1+i = 1+0,10 = 1,10. Διαιρείται κάθε ταμειακή ροή των επενδύσεων Κ και Λ, με $(1,10)^1$ για το πρώτο έτος, με $(1,10)^2$ για το δεύτερο έτος, $(1,10)^3$ για το τρίτο έτος, κ.λπ. και ο Πίνακας 9.1 γίνεται Πίνακας 9.2 με τις παρούσες αξίες όλων των ροών στον χρόνο μηδέν.

**Πίνακας 9.2** *Παράδειγμα με τις παρούσες αξίες των καθαρών ταμειακών ροών Πίνακα 9.1.*

| **Χρόνος** | **Κ επένδυση** | | **Λ επένδυση** | |
|---|---|---|---|---|
| 0 αρχή | -20.000 | -20.000 | -24.000 | -24.000 |
| 1 | $10.000 /(1+i)^1$ | =9.091 | $3.000 /(1+i)^1$ | =2.727 |
| 2 | $7.000 /(1+i)^2$ | =5.785 | $4.000 /(1+i)^2$ | =6.612 |
| 3 | $6.000 /(1+i)^3$ | =4.508 | $9.000 /(1+i)^3$ | =6.762 |
| 4 | $3.000 /(1+i)^4$ | =2.049 | $12.000 /(1+i)^4$ | =8.196 |
| 5 | 0 | 0 | $2.000 /(1+i)^5$ | =1.242 |
| 6 | 0 | 0 | 0 | 0 |
| Συνολική Καθαρή Παρούσα Αξία | | -20.000 +9.091+5.785+ 4.508+2.049= 1.433 | | -24.000 +2.727+6.612+6.762+8.196+.1242= 1.539 |

Μεγαλύτερη Καθαρή Παρούσα Αξία έχει η επένδυση Λ (1.539) από την επένδυση Κ (1.433). Χωρίς να ληφθεί υπόψη ότι η Λ απαιτεί περισσότερο αρχικό κεφάλαιο από την Κ και έχει διάρκεια περισσότερα χρόνια (πέντε), μόνο με κριτήριο τη μεγαλύτερη Καθαρή Παρούσα Αξία, θα πρέπει να επιλεγεί η Λ επένδυση.

## 9.9.2 Επιλογή με τη μέθοδο Εσωτερικού Ρυθμού Απόδοσης (IRR)

Στη μέθοδο αυτή δεν χρησιμοποιείται το κόστος κεφαλαίου (10%), επειδή ίσως δεν είναι γνωστό. Αναζητείται το επιτόκιο εκείνο (με δοκιμές), ώστε να μηδενιστεί η συνολική παρούσα αξία κάθε επένδυσης ξεχωριστά. Αυτό το επιτόκιο ονομάζεται εσωτερικός ρυθμός απόδοσης.

Στο παράδειγμα, δοκιμάζοντας διαιρέσεις με $(1+i)^n$ για διάφορες τιμές του επιτοκίου, όπως έγινε με τη μέθοδο της παρούσας αξίας, υπολογίζεται η συνολική παρούσα αξία κάθε επένδυσης. Για την Κ επένδυση το επιτόκιο που δίνει σχεδόν μηδενική συνολική παρούσα αξία (για την ακρίβεια 1) είναι το 13,95% και για τη Λ το επιτόκιο 12,35% δίνει σχεδόν μηδενική συνολική παρούσα αξία (για την ακρίβεια 3). Ο Πίνακας 9.1 γίνεται Πίνακας 9.3 με τις παρούσες αξίες όλων των ροών στον χρόνο μηδέν για τους αντίστοιχους εσωτερικούς βαθμούς απόδοσης των επενδύσεων Κ και Λ.

**Πίνακας 9.3** *Παράδειγμα με τις παρούσες αξίες των καθαρών ταμειακών ροών Πίνακα 9.1.*

| **Χρόνος** | **Κ επένδυση** | **i=13,95%** | **Λ επένδυση** | **i=12,35%** |
|---|---|---|---|---|
| 0 αρχή | -20.000 | -20.000 | -24.000 | -24.000 |
| 1 | $10.000/(1+i)^1$ | 8.776 | $3.000/(1+i)^1$ | 2.670 |
| 2 | $7.000/(1+i)^2$ | 5.391 | $4.000/(1+i)^2$ | 6.338 |
| 3 | $6.000/(1+i)^3$ | 4.055 | $9.000/(1+i)^3$ | 6.346 |
| 4 | $3.000/(1+i)^4$ | 1.779 | $12.000/(1+i)^4$ | 7.532 |
| 5 | 0 | 0 | $2.000/(1+i)^5$ | 1.117 |
| 6 | 0 | 0 | 0 | 0 |
| Συνολική Καθαρή Παρούσα Αξία | | -20.000 +8.776+5.391+ 4.055+1.779=1 | | -24.000 +2.670+6.338+6.346+7.532+1.117=3 |

Μεγαλύτερο εσωτερικό βαθμό απόδοσης έχει η επένδυση Κ (13,95%) και θα πρέπει να επιλεγεί ως η βέλτιστη επένδυση.

### 9.9.3 Επιλογή με τη μέθοδο της Περιόδου Επανείσπραξης

Για την εφαρμογή της μεθόδου της περιόδου επανείσπραξης θα πρέπει να είναι γνωστό το κόστος κεφαλαίου. Ας υποτεθεί ότι το επιτόκιο είναι 10%. Μετατρέπονται όλες οι ταμειακές ροές σε ισοδύναμες τιμές της αρχής του χρόνου, σύμφωνα με τον τύπο $K_0=K_n/(1+i)^n$

Επειδή υποτέθηκε επιτόκιο 10%, είναι 1+i=1+0,10=1,10. Διαιρείται κάθε ταμειακή ροή των επενδύσεων Κ και Λ, με $(1,10)^1$ για το πρώτο έτος, με $(1,10)^2$ για το δεύτερο έτος, $(1,10)^3$ για το τρίτο έτος, κ.λπ. και ο Πίνακας 9.1 γίνεται πίνακας 9.4 με τις παρούσες αξίες όλων των ροών στον χρόνο μηδέν.

**Πίνακας 9.4** *Παράδειγμα με τις παρούσες αξίες των καθαρών ταμειακών ροών Πίνακα 9.1 και περιόδου επανείσπραξης.*

| **Χρόνος** | **Κ επένδυση** | | **Λ επένδυση** | |
|---|---|---|---|---|
| 0 αρχή | -20.000 | -20.000 | -24.000 | -24.000 |
| 1 | $10.000/(1+i)^1$ | =9.091 | $3.000/(1+i)^1$ | =2.727 |
| 2 | $7.000/(1+i)^2$ | =5.785 | $4.000/(1+i)^2$ | =6.612 |
| 3 | $6.000/(1+i)^3$ | =4.508 | $9.000/(1+i)^3$ | =6.762 |
| 4 | $3.000/(1+i)^4$ | =2.049 | $12.000/(1+i)^4$ | =8.196 |
| 5 | 0 | 0 | $2.000/(1+i)^5$ | =1.242 |
| 6 | 0 | 0 | 0 | 0 |
| Έτη επανείσ-πραξης | | 3 έτη και κάτι… | | 4 έτη |

Αθροίζονται οι καθαρές ταμειακές ροές κάθε έτους με τα προηγούμενα, ώστε να βρεθεί πότε (σε ποιον χρόνο) θα υπάρχει επανείσπραξη του αρχικού κεφαλαίου ή θετικό άθροισμα καθαρών παρουσών αξιών για πρώτη φορά. Για την επένδυση Κ η περίοδος επανείσπραξης είναι λίγο περισσότερο από 3 έτη, ενώ για την επένδυση Λ είναι 4 έτη. Με το κριτήριο της μικρότερης περιόδου επανείσπραξης, θα πρέπει να επιλεγεί η Κ επένδυση, με την οποία ανακτάται το αρχικό κεφάλαιο της επένδυσης γρηγορότερα.

### 9.9.4 Επιλογή με τη μέθοδο του Λογιστικού Συντελεστή Απόδοσης (ARR)

Για την εφαρμογή της μεθόδου του Λογιστικού Συντελεστή Απόδοσης (ARR) δεν μετατρέπονται όλες οι ταμειακές ροές σε ισοδύναμες τιμές της αρχής του χρόνου, αφού δεν λαμβάνεται υπόψη η διαχρονική αξία του χρήματος. Υπολογίζεται η μέση ετήσια καθαρή εισροή και το μέσο ετήσιο κόστος και συγκρίνονται διαιρώντας τη μέση εισροή με το μέσο κόστος.

Η μέση ετήσια καθαρή εισροή είναι το άθροισμα εισροών-εκροών διά τον χρόνο ζωής της επένδυσης. Το μέσο κόστος είναι το άθροισμα αρχικής και υπολειμματικής αξίας διά τον χρόνο ζωής επένδυσης. Στον Πίνακα 9.5 εμφανίζεται η μέση ετήσια καθαρή εισροή και το μέσο ετήσιο κόστος για τις δύο επενδύσεις του Πίνακα 9.1.

**Πίνακας 9.5** *Παράδειγμα με μέση καθαρή εισροή και μέσο κόστος για τα δεδομένα του Πίνακα 9.1.*

| **Χρόνος** | **Κ επένδυση** | **Λ επένδυση** |
|---|---|---|
| 0 αρχή | -20.000 | -24.000 |
| 1 | $10.000/(1+i)^1$ | $3.000/(1+i)^1$ |
| 2 | $7.000/(1+i)^2$ | $4.000/(1+i)^2$ |
| 3 | $6.000/(1+i)^3$ | $9.000/(1+i)^3$ |
| 4 | $3.000/(1+i)^4$ | $12.000/(1+i)^4$ |
| 5 | 0 | $2.000/(1+i)^5$ |
| 6 | 0 | 0 |
| Μέση ετήσια καθαρή εισροή | (26.000-20.000)/4=1.500 | (34.000-24.000)/5=2.000 |
| Μέσο κόστος | (20.000-0)/4=5.000 | (24.000-0)/5=4.800 |
| ARR | 1.500/5.000=0,333 | 2.000/4.800=0,417 |

Μεγαλύτερο Λογιστικό Συντελεστή Απόδοσης έχει η επένδυση Λ (41,7%) από την επένδυση Κ (33,3%). Εστιάζοντας στη συνολική λογιστική απόδοση, θα πρέπει να επιλεγεί η Λ επένδυση.

## 9.9.5 Επιλογή με τη μέθοδο του Δείκτη Κερδοφορίας (PI)

Για την εφαρμογή της μεθόδου του Δείκτης κερδοφορίας (PI) θα πρέπει να είναι γνωστό το κόστος κεφαλαίου. Ας υποτεθεί ότι το επιτόκιο είναι 10%. Μετατρέπονται όλες οι ταμειακές ροές σε ισοδύναμες τιμές της αρχής του χρόνου, σύμφωνα με τον τύπο $K_0 = K_n/(1+i)^n$

Επειδή υποτέθηκε επιτόκιο 10%, είναι 1+i = 1+0,10 = 1,10. Διαιρείται κάθε ταμειακή ροή των επενδύσεων Κ και Λ, με $(1,10)^1$ για το πρώτο έτος, με $(1,10)^2$ για το δεύτερο έτος, $(1,10)^3$ για το τρίτο έτος, κ.λπ. και ο Πίνακας 9.1 γίνεται Πίνακας 9.6 με τις παρούσες αξίες όλων των ροών στον χρόνο μηδέν.

**Πίνακας 9.6** *Παράδειγμα με τις παρούσες αξίες των καθαρών ταμειακών ροών Πίνακα 9.1.*

| **Χρόνος** | **Κ επένδυση** | | **Λ επένδυση** | |
|---|---|---|---|---|
| 0 αρχή | -20.000 | -20.000 | -24.000 | -24.000 |
| 1 | $10.000/(1+i)^1$ | =9.091 | $3.000/(1+i)^1$ | =2.727 |
| 2 | $7.000/(1+i)^2$ | =5.785 | $4.000/(1+i)^2$ | =6.612 |
| 3 | $6.000/(1+i)^3$ | =4.508 | $9.000/(1+i)^3$ | =6.762 |
| 4 | $3.000/(1+i)^4$ | =2.049 | $12.000/(1+i)^4$ | =8.196 |
| 5 | 0 | 0 | $2.000/(1+i)^5$ | =1.242 |
| 6 | 0 | 0 | 0 | 0 |
| Δείκτης κερδοφορίας | | (9.091+5.785+4.508+2.049)/20.000=1,072 | | (2.727+6.612+6.762+8.196+1.242)/24.000=1,064 |

Δείκτης Κερδοφορίας (PI) είναι το πηλίκο της παρούσας αξίας των εισροών διά της παρούσας αξίας των εκροών της επένδυσης. Μεγαλύτερο δείκτη κερδοφορίας έχει η επένδυση Κ (1,072) από την επένδυση Λ (1,064). Σύμφωνα με τον δείκτη κερδοφορίας και οι δύο επενδύσεις είναι κερδοφόρες, αφού οι τιμές του δείκτη κερδοφορίας είναι μεγαλύτερες από 1. Καλύτερο δείκτη κερδοφορίας έχει η Κ επένδυση.

## 9.10 Ασκήσεις για κατανόηση - εφαρμογή

### 9.10.1 Αντικατάσταση εξοπλισμού

Μια επιχείρηση μεταφορών εξετάζει το ενδεχόμενο να επισκευάσει ένα παλιό φορτηγό της ή να αγοράσει καινούργιο. Το κόστος επισκευής είναι 20.000 ευρώ σήμερα και αναμένεται σε 4 έτη να χρειαστεί επισκευή μηχανής, που θα κοστίσει 8.000 ευρώ. Υπολογίζεται ότι με τις επισκευές το φορτηγό θα μπορεί να χρησιμοποιηθεί για τα επόμενα 8 έτη και στα 8 έτη θα έχει υπολειμματική αξία 6.000 ευρώ. Αν δεν επισκευαστεί και πωληθεί σήμερα, η αξία του είναι 7.000 ευρώ. Τα ετήσια έξοδα χρήσης του φορτηγού είναι 30.000, ενώ τα ετήσια έσοδα θα είναι 40.000 ευρώ. Η αγορά νέου φορτηγού θα κοστίσει 36.000 ευρώ, η διάρκεια ζωής του εκτιμάται σε 8 έτη και θα χρειαστεί κάποια συντήρηση ύψους 3.000 στα 4 έτη. Στο τέλος της οκταετίας η υπολειμματική του αξία θα είναι 6.000. Τα ετήσια έξοδα χρήσης του νέου φορτηγού είναι 21.000 ενώ τα ετήσια έσοδα θα είναι 40.000 ευρώ.

Αν η επιχείρηση απαιτεί απόδοση 12% για τις επενδύσεις της, θα πρέπει να προχωρήσει στην αντικατάσταση του φορτηγού της ή να επισκευάσει το παλιό;

Υπόδειξη. Συμπληρώνεται καταρχήν ο πίνακας ταμειακών ροών της επένδυσης για τα 8 έτη και γίνεται σύγκριση μεταξύ των δύο επιλογών επισκευή του υπάρχοντος και αγορά νέου.

**Πίνακας 9.7** *Δεδομένα Άσκησης 9.10.1.*

| **Χρόνος** | **Επισκευή παλαιού φορτηγού** | **Αγορά νέου φορτηγού** |
|---|---|---|
| 0 αρχή | -20.000 | 7.000-36.000 |
| 1 | +40.000-30.000 | +40.000-21.000 |
| 2 | +40.000-30.000 | +40.000-21.000 |
| 3 | +40.000-30.000 | +40.000-21.000 |
| 4 | +40.000-30.000-8.000 | +40.000-21.000 – 3.000 |
| 5 | +40.000-30.000 | +40.000-21.000 |
| 6 | +40.000-30.000 | +40.000-21.000 |
| 7 | +40.000-30.000 | +40.000-21.000 |
| 8 | +40.000-30.000 +6.000 | +40.000-21.000 + 6.000 |

Λύση

Ο πίνακας ταμειακών ροών της επένδυσης για τα 8 έτη και η καθαρή παρούσα αξία κάθε ποσού, εμφανίζεται στον Πίνακα 9.8

**Πίνακας 9.8** *Λύση Άσκησης 9.10.1.*

| **Χρόνος** | **Επισκευή παλαιού φορτηγού** | **Καθαρή παρούσα αξία παλαιού** | **Αγορά νέου φορτηγού** | **Καθαρή παρούσα αξία νέου** |
|---|---|---|---|---|
| 0 αρχή | -20.000 | -20.000 | 7.000-36.000 | -29.000 |
| 1 | +40.000-30.000 | 8.929 | +40.000-21.000 | 16.964 |
| 2 | +40.000-30.000 | 7.972 | +40.000-21.000 | 15.147 |
| 3 | +40.000-30.000 | 7.118 | +40.000-21.000 | 13.524 |
| 4 | +40.000-30.000-8.000 | 1.271 | +40.000-21.000 – 3.000 | 10.168 |
| 5 | +40.000-30.000 | 5.674 | +40.000-21.000 | 10.781 |
| 6 | +40.000-30.000 | 5.066 | +40.000-21.000 | 9.626 |
| 7 | +40.000-30.000 | 4.523 | +40.000-21.000 | 8.595 |
| 8 | +40.000-30.000 +6.000 | 6.462 | +40.000-21.000 + 6.000 | 10.097 |
| | Συνολική αξία | 27.015,55 | | 65.901,9 |

Συγκρίνοντας τη συνολική καθαρή παρούσα αξία μεταξύ των δύο επιλογών, επισκευή του υπάρχοντος 27.015,55 ευρώ και αγορά νέου 65.901,9 ευρώ, προκύπτει ως καλύτερη επιλογή η αγορά νέου φορτηγού, που δίνει τη μεγαλύτερη καθαρή παρούσα αξία.

### 9.10.2 Αξιολόγηση επενδυτικής πρότασης

Μια επιχείρηση αξιολογεί την παρακάτω επενδυτική πρόταση. Το κόστος αγοράς μηχανημάτων είναι 60.000 ευρώ σήμερα και απαιτείται αρχικό κεφάλαιο 100.000 ευρώ, το οποίο αποδεσμεύεται σε 5 έτη. Για τη συντήρηση των μηχανημάτων πληρώνουμε εφάπαξ 5.000 ευρώ συνδρομή για τα 5 πρώτα έτη λειτουργίας τους. Η υπολειμματική αξία των μηχανημάτων είναι 10.000 ευρώ στο τέλος της πενταετίας. Τα ετήσια έξοδα λειτουργίας θα είναι 35.000 ευρώ, οι άλλες δαπάνες θα είναι 125.000 ευρώ, ενώ τα ετήσια έσοδα θα είναι 200.000 ευρώ.

Αν η επιχείρηση απαιτεί απόδοση 9% για τις επενδύσεις της, θα πρέπει να προχωρήσει στην προτεινόμενη επένδυση;

Υπόδειξη. Συμπληρώνεται καταρχήν ο πίνακας ταμειακών ροών της επένδυσης για τα 5 έτη και γίνεται μετατροπή σε καθαρές παρούσες αξίες, ώστε να υπολογιστεί η συνολική Καθαρή Παρούσα Αξία.

Λύση

**Πίνακας 9.9** *Λύση Άσκησης 9.10.2.*

| **Χρόνος** | | | **Ταμειακές ροές επένδυσης** | **Καθαρή παρούσα αξία** |
|---|---|---|---|---|
| 0 αρχή | 100.000+60.000 | | -160.000 | -160.000 |
| 1 | 125.000+35.000 | 200.000 | 40.000 | 36.697 |
| 2 | 125.000+35.000 | 200.000 | 40.000 | 33.667 |
| 3 | 125.000+35.000 | 200.000 | 40.000 | 30.887 |
| 4 | 125.000+35.000 | 200.000 | 40.000 | 28.337 |
| 5 | 125.000+35.000+5.000 | 200.000+ | 145.000 | 94.240 |
| | | | Συνολική αξία | 63.829 |

Στον Πίνακα 9.9 εμφανίζεται η συνολική καθαρή παρούσα αξία 63.829 ευρώ. Άρα συμφέρει η επένδυση.

# Βιβλιογραφία/Αναφορές


Bradley, T. (2014). *Μαθηματικά για την Οικονομία και τη Διοίκηση*. Αθήνα: Κριτική.

Σακκάς, Ν. (2012). *Αξιολόγηση επενδύσεων Βασική θεωρία*, https://eclass.hmu.gr/modules/document/index.php?...

Φλώρου, Γ. (2015). *Ποσοτικές μέθοδοι στα χρηματοοικονομικά* [Προπτυχιακό εγχειρίδιο]. Κάλλιπος, Ανοικτές Ακαδημαϊκές Εκδόσεις. https://hdl.handle.net/11419/5066

Φώτης Π. (2015). *Χρηματοοικονομική Ανάλυση Επενδύσεων*. Αθήνα: Προπομπός.

IRR ependysopedia (2015). Ανακτήθηκε 30 Ιουνίου, 2015, από http://www.ependysopedia.gr/internal-rate-of return-irr

https://eclass.teiath.gr/modules/document/index.php?course=DE138&openDir=/53859d5675hC

https://eclass.hmu.gr/modules/document/index.php?...

# Κριτήρια αξιολόγησης

## Κριτήριο αξιολόγησης 1

**Τι είναι επένδυση;**

**Απάντηση**
**Επένδυση** είναι η πράξη απόκτησης διαρκούς αγαθού με σκοπό τη χρησιμοποίησή του κατά τη διάρκεια ζωής του προς παραγωγή νέων αγαθών ή υπηρεσιών.

## Κριτήριο αξιολόγησης 2

**Ποιες είναι οι κυριότερες τεχνικές αξιολόγησης επενδύσεων;**

**Απάντηση**
Οι κυριότερες τεχνικές αξιολόγησης επενδύσεων (με καθαρά χρηματοοικονομικά κριτήρια) είναι Καθαρή Παρούσα Αξία (NPV), Εσωτερικός Βαθμός Απόδοσης (IRR), Περίοδος Επανείσπραξης (PP), Λογιστικός Συντελεστής Απόδοσης (ARR), Δείκτης Κερδοφορίας (PI).

## Κριτήριο αξιολόγησης 3

**Πώς συγκρίνονται οι επενδύσεις με την τεχνική της παρούσας αξίας;**

**Απάντηση**
Για κάθε επένδυση και για όλες τις μελλοντικές εισροές και εκροές της υπολογίζεται η αξία τους στον χρόνο μηδέν. Η σύγκριση γίνεται για τις παρούσες αξίες όλων των επενδύσεων.

## Κριτήριο αξιολόγησης 4

**Πώς βρίσκεται ο εσωτερικός βαθμός απόδοσης μιας επένδυσης;**

**Απάντηση**
Αναζητείται το επιτόκιο εκείνο (με δοκιμές), ώστε να μηδενιστεί η συνολική παρούσα αξία της επένδυσης.

# Κεφάλαιο 10: Θεωρία Παιγνίων

**Σύνοψη**
*Στο κεφάλαιο αυτό γίνεται εισαγωγή στη θεωρία παιγνίων και στη χρήση της σε συνθήκες κινδύνου ή ανταγωνισμού. Παρουσιάζονται με απλό τρόπο παραδείγματα από την επιχειρηματική ζωή με έμφαση στον ορισμό και στην επίλυση παιγνίων.*

**Προαπαιτούμενη γνώση**
*Δεν απαιτείται.*

## 10.1 Αποφάσεις σε συνθήκες αβεβαιότητας

**Απόφαση** χρειάζεται να ληφθεί σε διάφορες καταστάσεις που απαιτούν κάποια ενέργεια και υπάρχουν εναλλακτικές επιλογές, όπως ποια τράπεζα να επιλεγεί για δανεισμό, με ποιον τρόπο να επενδυθούν κεφάλαια, σε ποια τοποθεσία να γίνει η έδρα μια επιχείρησης, με ποια εταιρεία ταχυμεταφορών θα συνεργαστεί κάποιος και άλλα. Η επιλογή της σωστής απόφασης είναι δύσκολη σε προβλήματα στα οποία υπεισέρχεται η αβεβαιότητα, και θα πρέπει να γίνει μια ενέργεια η οποία θα είναι επικερδής. Στις περιπτώσεις αυτές, έχει αναπτυχθεί η θεωρία πιθανοτήτων, αλλά δεν μπορεί να εφαρμοστεί μόνο αυτή. Η αβεβαιότητα για το αποτέλεσμα μιας ενέργειας δεν είναι μόνο αντικειμενική αλλά είναι και υποκειμενική, αφού εξαρτάται από το ποια επιλογή θα γίνει και σε ποιον χρόνο.

Η ανάλυση όλων των δυνατών αποφάσεων, ώστε να επιλεγεί η σωστότερη κάθε φορά, είναι μια διαδικασία την οποία θα περιγράψουμε στη συνέχεια.

### 10.1.1 Βήματα ανάλυσης αποφάσεων

Για κάθε απόφαση θα χρειαστούν μελέτη της υπάρχουσας κατάστασης, πρόβλεψη των αποτελεσμάτων, εκτίμηση της πιθανότητας κάθε αποτελέσματος και εντοπισμός της βέλτιστης επιλογής. Τα βήματα της ανάλυσης αποφάσεων είναι:

1. Απαρίθμηση εναλλακτικών/επιτρεπτών δραστηριοτήτων. (Προϋπόθεση για τη σωστή απόφαση είναι να έχουν καταγραφεί όλες οι εναλλακτικές.)
2. Απαρίθμηση εναλλακτικών εκβάσεων για τα αβέβαια γεγονότα που επηρεάζουν την απόφαση.
3. Καταγραφή πίνακα αποτελεσμάτων ή διάγραμμα αποφάσεων – δένδρο αποφάσεων.
4. Εκτίμηση πόσο πιθανή είναι κάθε εναλλακτική έκβαση.
5. Σύγκριση ελκυστικότητας για τις εναλλακτικές δραστηριότητες.

### 10.1.2 Δραστηριότητες και εκβάσεις της «φύσης»

Οι εναλλακτικές δραστηριότητες που είναι διαθέσιμες εξαρτούν το αποτέλεσμά τους από τις εκβάσεις της «φύσης», δηλαδή από τυχαίους παράγοντες που δεν είναι γνωστό αν τελικά συμβούν ή όχι.

Οι δραστηριότητες-αποφάσεις συμβολίζονται με Α1, Α2, ... Αν και οι φυσικές εκβάσεις / τυχαία φαινόμενα με Φ1, Φ2, … Φκ. Για κάθε απόφαση Αi και κάθε έκβαση Φj αντιστοιχεί ένα μετρήσιμο αποτέλεσμα. Αν το αποτέλεσμα δεν είναι μετρήσιμο, με κάποιον τρόπο (αντιστοίχιση κέρδους ή ζημιάς ή χρόνου κ.λπ.), γίνεται μετρήσιμο, ώστε να μπορεί να συγκριθεί με τα υπόλοιπα.

Όλα τα αποτελέσματα καταγράφονται σε πίνακα εκβάσεων, όπως φαίνεται στον Πίνακα 10.1. Στον Πίνακα 10.1 παρουσιάζονται τα κέρδη για καθεμία από τις ν σε πλήθος Αποφάσεις, Α1, Α2, … Αν, ανάλογα με τις κ εκβάσεις της «φύσης» Φ1, Φ2, … Φκ. Στη διασταύρωση της Α1 απόφασης και της Φ1 έκβασης, η τιμή 100 αντιστοιχεί στο όφελος, αν ληφθεί η Α1 απόφαση και συμβεί η Φ1 έκβαση. Αν ληφθεί η Α1 απόφαση και συμβεί η Φ2 έκβαση, το όφελος θα είναι 150, ενώ αν συμβεί η Φκ έκβαση το όφελος θα είναι 50 για την απόφαση Α1. Στην περίπτωση της απόφασης Αν και της Φκ έκβασης θα υπάρχει ζημιά 200, η οποία αντιστοιχεί σε τιμή με αρνητικό πρόσημο στον Πίνακα 10.1.

**Πίνακας 10.1** *Παράδειγμα διαμόρφωσης πίνακα αποφάσεων και αποτελεσμάτων ανάλογα με τις εκβάσεις της φύσης.*

| **Αποφάσεις Αi και φυσικές εκβάσεις Φj** | **Φ1** | **Φ2** | **…** | **Φκ** |
|---|---|---|---|---|
| Α1 | 100 | 150 | | 30 |
| Α2 | 20 | | | |
| … | | | | |
| Αν | 30 | | | -200 |

### 10.1.3 Συνθήκες αποφάσεων

Η γνώση που υπάρχει για τις εκβάσεις της φύσης καθορίζει και τις συνθήκες στις οποίες λαμβάνεται η βέλτιστη απόφαση. Όταν είναι γνωστό ποια από τις εκβάσεις θα συμβεί (μία αντίστοιχη στήλη στον Πίνακα 10.1), η απόφαση λαμβάνεται σε **συνθήκες βεβαιότητας**. Στην περίπτωση αυτή, επιλέγεται η απόφαση με το μεγαλύτερο όφελος για την αντίστοιχη στήλη της βέβαιης έκβασης. Όταν δεν είναι γνωστό ποια από τις εκβάσεις θα συμβεί (κάποια αντίστοιχη στήλη στον Πίνακα 10.1) αλλά η πιθανότητα της κάθε έκβασης έχει εκτιμηθεί, η απόφαση λαμβάνεται σε **συνθήκες κινδύνου**. Όταν δεν είναι γνωστό ποια από τις εκβάσεις θα συμβεί (κάποια αντίστοιχη στήλη στον Πίνακα 10.1), ούτε είναι γνωστή η πιθανότητα κάθε έκβασης, η απόφαση λαμβάνεται σε **συνθήκες αβεβαιότητας**. Επειδή η φύση είναι απρόβλεπτη, είναι πολύ σπάνιες οι συνθήκες βεβαιότητας. Οι συνθήκες κινδύνου εκτιμώνται αν υπάρχει αντίστοιχη εκτίμηση της πιθανότητας κάθε έκβασης. Στις συνθήκες αβεβαιότητας, που είναι και η πιο συνηθισμένη περίπτωση, η απόφαση λαμβάνεται ανάλογα με την εμπειρία ή κάνοντας υποθέσεις για τις εκβάσεις που πρόκειται να συμβούν.

## 10.2 Αποφάσεις σε συνθήκες κινδύνου

Στον Πίνακα 10.2 εμφανίζεται ένα παράδειγμα όπου δεν είναι γνωστό ποια από τις εκβάσεις θα συμβεί (κάποια στήλη στον Πίνακα 10.2), αλλά η πιθανότητα της κάθε έκβασης έχει εκτιμηθεί και εμφανίζεται στη δεύτερη γραμμή του Πίνακα 10.2. Οι εκτιμήσεις των πιθανοτήτων των εκβάσεων μπορούν να πραγματοποιηθούν με χρήση παλαιότερων δεδομένων ή αντίστοιχων γνωστών καταστάσεων και γίνονται με τη χρήση της Στατιστικής Επιστήμης. Στην περίπτωση αυτή ορίζεται ότι η απόφαση λαμβάνεται σε **συνθήκες κινδύνου**.

Για κάθε απόφαση Α1, Α2,…Αν, υπολογίζεται το **αναμενόμενο όφελος** της απόφασης, πολλαπλασιάζοντας τα οφέλη της αντίστοιχης γραμμής με τις πιθανότητες κάθε έκβασης, όπως αυτές εκτιμήθηκαν για τις εκβάσεις Φ1, Φ2,…Φκ και προσθέτοντας το άθροισμα των γινομένων. Το αναμενόμενο όφελος κάθε απόφασης αντιστοιχεί στη μέση τιμή του οφέλους της αντίστοιχης απόφασης, όταν αυτή επαναλαμβάνεται και εμφανίζονται διαφορετικές εκβάσεις σε κάθε επανάληψη. Τα αναμενόμενα οφέλη όλων των αποφάσεων συγκρίνονται μεταξύ τους και επιλέγεται ως βέλτιστη η απόφαση με το μεγαλύτερο αναμενόμενο όφελος.

Για παράδειγμα, το αναμενόμενο όφελος της Α1 θα είναι το άθροισμα γινομένων της γραμμής που αντιστοιχεί στην Α1 με τις αντίστοιχες πιθανότητες που εμφανίζονται στη δεύτερη γραμμή (Πίνακας 10.2). Δηλαδή το αναμενόμενο όφελος της Α1, υπολογίζεται ως εξής: 100*0,10+150*0,05+…+30*0,20. Με τον ίδιο τρόπο υπολογίζονται τα αναμενόμενα οφέλη όλων των αποφάσεων.

**Πίνακας 10.2** *Παράδειγμα πίνακα αποφάσεων και αποτελεσμάτων σε συνθήκες κινδύνου.*

| Αποφάσεις Ai και φυσικές εκβάσεις Φj | Φ1 | Φ2 | … | Φκ |
|---|---|---|---|---|
| Πιθανότητες εκβάσεων | 10% | 5% | | 20% |
| A1 | 100 | 150 | | 30 |
| A2 | 20 | | | |
| … | | | | |
| Aν | 30 | | | -200 |
| | | | | |

Στη διαδικασία αποφάσεων σε συνθήκες κινδύνου, επιλέγεται η απόφαση που κατά μέσο όρο έχει το μεγαλύτερο αναμενόμενο όφελος, άσχετα με το ποια θα είναι η έκβαση Φ1, Φ2,...Φκ. Είναι πιθανό να επιλεγεί μία απόφαση η οποία για κάποια συγκεκριμένη έκβαση Φj να έχει μικρότερο όφελος από μια άλλη απόφαση. Επειδή, όμως, όλες οι εκβάσεις είναι πιθανές και δεν είναι γνωστό αν θα συμβεί η έκβαση Φj, επιλέγεται η απόφαση εκείνη που θα έχει μεγαλύτερο όφελος, άσχετα με το ποια από τις εκβάσεις θα συμβεί.

## 10.3 Αποφάσεις σε συνθήκες αβεβαιότητας

Για την αξιολόγηση αποφάσεων σε συνθήκες αβεβαιότητας, όπου δεν είναι γνωστές ούτε μπορούν να εκτιμηθούν οι πιθανότητες εμφάνισης των εκβάσεων χρησιμοποιούνται διάφορα κριτήρια σύγκρισης. Η επιλογή για το ποιο από τα κριτήρια αυτά θα εφαρμοστεί επηρεάζεται από υποκειμενικούς παράγοντες, όπως η εμπειρία και ο χαρακτήρας όποιου αποφασίζει. Όπως θα φανεί και στα διάφορα παραδείγματα, κάθε διαφορετικό κριτήριο δεν οδηγεί πάντα στην ίδια επιλογή βέλτιστης απόφασης για το ίδιο πρόβλημα. Τα κυριότερα κριτήρια που θα παρουσιαστούν στη συνέχεια είναι:

1 maxmin

2 maxmax

3 minmax

Η ονομασία των κριτηρίων προκύπτει από τον συνδυασμό των max για την επιλογή του μεγίστου και min για την επιλογή του ελαχίστου. Η σειρά καταγραφής τους είναι πρώτα δεξιά η επιλογή για τη γραμμή και κατόπιν αριστερά η επιλογή μεταξύ των αποτελεσμάτων που προέκυψαν από όλες τις γραμμές. Δηλαδή:

1 maxmin (Επιλέγεται το μεγαλύτερο από τα χειρότερα αποτελέσματα). Καταγράφεται αρχικά το χειρότερο αποτέλεσμα κάθε απόφασης και στη συνέχεια επιλέγεται το μεγαλύτερο από αυτά.

2 maxmax (Επιλέγεται το μεγαλύτερο από τα καλύτερα αποτελέσματα). Καταγράφεται το καλύτερο αποτέλεσμα κάθε απόφασης και στη συνέχεια επιλέγεται το μεγαλύτερο από αυτά.

3 minmax (Επιλέγεται το μικρότερο από τις μεγαλύτερες απώλειες ευκαιρίας). Καταγράφεται νέος πίνακας με το κόστος χαμένης ευκαιρίας για κάθε απόφαση. Στη συνέχεια καταγράφεται η μεγαλύτερη απώλεια ευκαιρίας για κάθε απόφαση και επιλέγεται η απόφαση με τη μικρότερη από τις μεγαλύτερες απώλειες ευκαιρίας.

Η βασική παραδοχή στη διαδικασία επιλογής κριτηρίου είναι η δημιουργία πίνακα αποφάσεων (όπως ο Πίνακας 10.1), με την καταγραφή των διαφορετικών επιλογών αποφάσεων, την καταγραφή όλων των πιθανών εκβάσεων, καθώς και την εκτίμηση του οφέλους που προκύπτει από κάθε απόφαση για κάθε διαφορετική έκβαση.

### 10.3.1 Παράδειγμα αποφάσεων σε συνθήκες αβεβαιότητας

Ένα παράδειγμα σε συνθήκες αβεβαιότητας περιγράφεται στον Πίνακα 10.3. Πρόκειται να επιλεγεί χώρος δημιουργίας ενός γραφείου ταχυμεταφορών και προτείνονται 3 διαφορετικές επιλογές (αποφάσεις). **Α** η δημιουργία του στο κέντρο της πόλης, **Β** η δημιουργία του στα περίχωρα της πόλης και **Γ** η δημιουργία του στην επαρχία. Η ζήτηση σε ταχυμεταφορές ίσως είναι χαμηλή, μέτρια ή υψηλή στο μέλλον. Πρόκειται για τρεις διαφορετικές εκβάσεις $Φ_Α$, $Φ_Β$, $Φ_Γ$, για τις οποίες δεν έχουμε στατιστικά στοιχεία και δεν μπορούμε να εκτιμήσουμε τις πιθανότητές τους να συμβούν. Το γραφείο, ανάλογα με τη μελλοντική ζήτηση και την περιοχή εγκατάστασης, εκτιμά τα κέρδη του, τα οποία παρουσιάζονται στα αντίστοιχα κελιά του Πίνακα 10.3. Θετικές τιμές αντιστοιχούν σε κέρδη, ενώ αρνητικές τιμές σε ζημιές.

Ο επιχειρηματίας είναι άτομο αισιόδοξο, αλλά ο σύμβουλος του είναι πιο σκεφτικός και διστάζει να προτείνει τόπο επιλογής που δεν θα φέρει μεγάλα κέρδη.

**Πίνακας 10.3** *Παράδειγμα πίνακα αποφάσεων σε συνθήκες αβεβαιότητας.*

| **Κέρδη για χώρο επιλογής γραφείου και μελλοντική ζήτηση** | **$Φ_Α$ χαμηλή** | **$Φ_Β$ μέτρια** | **$Φ_Γ$ υψηλή** |
|---|---|---|---|
| Α κέντρο της πόλης | -100 | 250 | 300 |
| Β περίχωρα της πόλης | -200 | 350 | 400 |
| Γ επαρχία | -250 | 300 | 500 |

Στη συνέχεια θα παρουσιαστεί η βέλτιστη επιλογή για τον τόπο γραφείου, ανάλογα με τα κριτήρια maxmin, maxmax, minmax.

## 10.3.2 Επιλογή με το κριτήριο maxmin σε συνθήκες αβεβαιότητας

Ο σύμβουλος του επιχειρηματία, καταγράφει καταρχάς το ελάχιστο δυνατό αποτέλεσμα για κάθε απόφαση Α, Β, Γ όπως απεικονίζεται στον Πίνακα 10.4. Θέλοντας να περιορίσει το ελάχιστο αποτέλεσμα, επιλέγει την απόφαση η οποία αντιστοιχεί στο μέγιστο από τα ελάχιστα. Πρόκειται για την απόφαση Α, κέντρο της πόλης. Προτείνει αυτή την επιλογή ως τόπο εγκατάστασης, ώστε ο επιχειρηματίας να μην έχει αποτέλεσμα μικρότερο από το -100. Να περιορίσει δηλαδή πιθανή ζημία του στη χαμηλή ζήτηση στο -100 ή να έχει κέρδος 250 ή 300 ανάλογα με το αν η ζήτηση γίνει μέτρια ή υψηλή.

**Πίνακας 10.4** *Παράδειγμα πίνακα αποφάσεων σε συνθήκες αβεβαιότητας με το κριτήριο maxmin.*

| **Κέρδη για χώρο επιλογής γραφείου και μελλοντική ζήτηση** | **$Φ_Α$ χαμηλή** | **$Φ_Β$ μέτρια** | **$Φ_Γ$ υψηλή** | **Ελάχιστο γραμμής** | **Μέγιστο από τα ελάχιστα γραμμών maxmin** |
|---|---|---|---|---|---|
| Α κέντρο της πόλης | -100 | 250 | 300 | -100 | -100 |
| Β περίχωρα της πόλης | -200 | 350 | 400 | -200 | |
| Γ επαρχία | 250 | -300 | 500 | -300 | |

Το κριτήριο αυτό ονομάζεται και κριτήριο απαισιοδοξίας, λόγω του γεγονότος ότι δεν χρησιμοποιεί την ελπίδα για το μέγιστο αποτέλεσμα, αλλά θεωρεί ότι θα συμβεί η χειρότερη έκβαση για κάθε απόφαση και προσπαθεί να περιορίσει τα αποτελέσματα της έκβασης αυτής.

## 10.3.3 Επιλογή με το κριτήριο maxmax σε συνθήκες αβεβαιότητας

Ο επιχειρηματίας, επειδή είναι αισιόδοξος και παρατηρεί ότι στον Πίνακα 10.5 υπάρχει και η δυνατότητα μεγαλύτερων κερδών αν δεν επιλέξει την απόφαση Α κέντρο πόλης, καταγράφει καταρχάς το μέγιστο δυνατό αποτέλεσμα για κάθε απόφαση Α, Β, Γ (Πίνακας 10.5). Θέλοντας να αυξήσει το μέγιστο πιθανό αποτέλεσμα, επιλέγει την απόφαση η οποία αντιστοιχεί στο μέγιστο από τα μέγιστα. Πρόκειται για την απόφαση Γ, επαρχία. Προτείνει αυτό ως τόπο εγκατάστασης, ώστε ο επιχειρηματίας να έχει το μέγιστο δυνατό αποτέλεσμα, προσδοκώντας ότι η ζήτηση θα είναι υψηλή στο μέλλον. Με την απόφαση όμως αυτή, στην περίπτωση χαμηλής ζήτησης το κέρδος του περιορίζεται στο 250, ενώ στην περίπτωση μέτριας ζήτησης μπορεί και να έχει ζημία -300.

**Πίνακας 10.5** *Παράδειγμα πίνακα αποφάσεων σε συνθήκες αβεβαιότητας με το κριτήριο maxmax.*

| **Κέρδη για χώρο επιλογής γραφείου και μελλοντική ζήτηση** | **$Φ_Α$ χαμηλή** | **$Φ_Β$ μέτρια** | **$Φ_Γ$ υψηλή** | **Μέγιστο γραμμής** | **Μέγιστο από τα μέγιστα γραμμών maxmax** |
|---|---|---|---|---|---|
| Α κέντρο της πόλης | -100 | 250 | 300 | 300 | |
| Β περίχωρα της πόλης | -200 | 350 | 400 | 400 | |
| Γ επαρχία | 250 | -300 | 500 | 500 | 500 |

Το κριτήριο αυτό ονομάζεται και κριτήριο αισιοδοξίας, λόγω του γεγονότος ότι χρησιμοποιεί την ελπίδα ότι θα συμβεί η καλύτερη έκβαση για κάθε απόφαση και προσπαθεί να μεγιστοποιήσει τα αποτελέσματα της έκβασης αυτής.

### 10.3.4 Επιλογή με το κριτήριο minmax σε συνθήκες αβεβαιότητας

Ο σύμβουλος, στην προσπάθειά του να πείσει τον επιχειρηματία να μην είναι τόσο ριψοκίνδυνος, προτείνει να υπολογιστεί και το κόστος χαμένης ευκαιρίας για κάθε απόφαση. Το κόστος χαμένης ευκαιρίας αντιστοιχεί στη διαφορά του αποτελέσματος κάθε απόφασης από το μέγιστο δυνατό αποτέλεσμα κάθε έκβασης. Δηλαδή σε κάθε στήλη-έκβασης του Πίνακα 10.6 υπολογίζεται μία νέα στήλη έκβασης, όπου αναγράφεται η διαφορά του μεγίστου της στήλης με το αντίστοιχο κελί κάθε γραμμής. Πρόκειται για υποθετικό κόστος χαμένης ευκαιρίας, το οποίο «χάνει» η επιχείρηση επειδή δεν γνωρίζει ποια έκβαση θα γίνει και επιλέγει απόφαση που δεν δίνει το καλύτερο αποτέλεσμα για τη συγκεκριμένη έκβαση.

Δηλαδ,ή αν επιλεγεί η απόφαση Α κέντρο πόλης και η ζήτηση είναι χαμηλή, η επιχείρηση χάνει 100 ενώ θα μπορούσε να κερδίσει 250, αν γνώριζε τη χαμηλή ζήτηση και επέλεγε την απόφαση Γ επαρχία. Η διαφορά μεταξύ όσων θα μπορούσε να κερδίσει και όσων κερδίζει με την απόφασή της ονομάζεται κόστος χαμένης ευκαιρίας.

Αν επιλεγεί η απόφαση Β περίχωρα πόλης και η ζήτηση είναι χαμηλή, η επιχείρηση χάνει 200, ενώ θα μπορούσε να κερδίσει 250 αν επέλεγε την απόφαση Γ επαρχία. Αν επιλεγεί η απόφαση Γ επαρχία και η ζήτηση είναι χαμηλή, το κόστος χαμένης ευκαιρίας είναι 0, αφού η επιχείρηση ήδη έκανε την καλύτερη επιλογή για την έκβαση χαμηλή ζήτηση.

Αν επιλεγεί η απόφαση Α κέντρο πόλης και η ζήτηση είναι μέτρια, η επιχείρηση κερδίζει 250 ενώ θα μπορούσε να κερδίσει 350 αν γνώριζε τη μέτρια ζήτηση και επέλεγε την απόφαση Β περίχωρα πόλης. Η διαφορά μεταξύ όσων θα μπορούσε να κερδίσει (350) και όσων κερδίζει (250) με την απόφασή της ονομάζεται κόστος χαμένης ευκαιρίας.

Αν επιλεγεί η απόφαση Β περίχωρα πόλης και η ζήτηση είναι μέτρια, η επιχείρηση κερδίζει 350 και το κόστος χαμένης ευκαιρίας είναι 0, αφού η επιχείρηση ήδη έκανε την καλύτερη επιλογή για την έκβαση μέτρια ζήτηση. Αν επιλεγεί η απόφαση Γ επαρχία και η ζήτηση είναι μέτρια, το κόστος χαμένης ευκαιρίας είναι 650, αφού η επιχείρηση θα μπορούσε να κερδίσει 350 και με την απόφασή της Γ χάνει 300.

Αν επιλεγεί η απόφαση Α κέντρο πόλης και η ζήτηση είναι υψηλή, η επιχείρηση κερδίζει 300 ενώ θα μπορούσε να κερδίσει 500, αν γνώριζε την υψηλή ζήτηση και επέλεγε την απόφαση Γ επαρχία. Η διαφορά 200 όσων θα μπορούσε να κερδίσει (500) και όσων κερδίζει (300) με την απόφασή της ονομάζεται κόστος χαμένης ευκαιρίας.

Αν επιλεγεί η απόφαση Β περίχωρα πόλης και η ζήτηση είναι υψηλή, η επιχείρηση κερδίζει 400, ενώ θα μπορούσε να κερδίσει 500 αν επέλεγε την απόφαση Γ επαρχία. Αν επιλεγεί η απόφαση Γ επαρχία και η ζήτηση είναι υψηλή, το κόστος χαμένης ευκαιρίας είναι 0, αφού η επιχείρηση ήδη έκανε την καλύτερη επιλογή για την έκβαση υψηλή ζήτηση.

Μετά από τη δημιουργία του πίνακα με τα κόστη χαμένης ευκαιρίας, επιλέγεται η απόφαση που δίνει το ελάχιστο δυνατό κόστος χαμένης ευκαιρίας όποια έκβαση και να συμβεί. Για να βρεθεί η απόφαση αυτή, καταγράφεται το μέγιστο κόστος χαμένης ευκαιρίας για κάθε γραμμή απόφασης και στη συνέχεια επιλέγεται το ελάχιστο από τα μέγιστα.

Στον Πίνακα 10.6 το ελάχιστο κόστος χαμένης ευκαιρίας από τα μέγιστα είναι 450 και συμπίπτει για τις δύο αποφάσεις Α κέντρο πόλης και Β περίχωρα. Οπότε προτείνεται μία από τις δύο αυτές αποφάσεις, αφού η απόφαση Γ επαρχία είναι πιθανό να δώσει μεγαλύτερο (650) κόστος χαμένης ευκαιρίας.

**Πίνακας 10.6** *Παράδειγμα πίνακα αποφάσεων σε συνθήκες αβεβαιότητας με το κριτήριο minmax.*

| **Κέρδη για χώρο επιλογής γραφείου και μελλοντική ζήτηση** | **$Φ_Α$ χαμηλή** | | **$Φ_Β$ μέτρια** | | **$Φ_Γ$ υψηλή** | | **Μέγιστο γραμμής χαμένης ευκαιρίας** | **Ελάχιστο από τα μέγιστα γραμμών minmax** |
|---|---|---|---|---|---|---|---|---|
| | | Χαμένη ευκαιρία | | Χαμένη ευκαιρία | | Χαμένη ευκαιρία | | |
| Α κέντρο της πόλης | -100 | 250-(-100) = 350 | 250 | 350-(-100) = 450 | 300 | 500-300 = 200 | 450 | 450 |
| Β περίχωρα της πόλης | -200 | 250-(-200) = 450 | 350 | 350-350 = 0 | 400 | 500-400 = 100 | 450 | 450 |
| Γ επαρχία | 250 | 250-250 = 0 | -300 | 350-(-300) = 650 | 500 | 500-500 = 0 | 650 | |

Το κριτήριο αυτό ονομάζεται και κριτήριο υπερηφάνειας, βασίζεται στον περιορισμό της δυνατότητας απώλειας ευκαιρίας-κέρδους. Χρησιμοποιείται, επίσης, για να προσδιοριστεί η αξία (αμοιβή) της πληροφόρησης για τις πιθανότητες των εκβάσεων που θα συμβούν στο μέλλον, ώστε να λάβει η επιχείρηση την κατάλληλη απόφαση.

Αν οι πιθανότητες των εκβάσεων μπορούν να εκτιμηθούν μέσω κάποιας μελέτης, υπολογίζεται το αναμενόμενο κόστος χαμένης ευκαιρίας για κάθε απόφαση, πολλαπλασιάζοντας το κόστος χαμένης ευκαιρίας με την πιθανότητα κάθε έκβασης και προσθέτοντας όλα τα γινόμενα. Το ελάχιστο αναμενόμενο κόστος χαμένης ευκαιρίας αντιστοιχεί στην αξία της άριστης πληροφόρησης, στο μέγιστο ποσό που μπορεί να διατεθεί για την πληροφόρηση των πιθανοτήτων των εκβάσεων.

### 10.3.5 Παράδειγμα επιλογής απόφασης συμβιβασμού ή εκδίκασης

Ένας δικηγόρος αντιμετωπίζει μία διαμάχη με έναν διάδικο και μπορεί να προτείνει είτε συμβιβασμό είτε εκδίκαση της διαμάχης. Ο συμβιβασμός ή θα γίνει δεκτός από τον διάδικο ή όχι. Αν δεν γίνει δεκτός, θα υπάρξει δίκη με αποτέλεσμα είτε θετικό είτε αρνητικό για τον δικηγόρο. Οι δυνατές εκβάσεις είναι:

Φ1 Συμβιβασμός. Στην περίπτωση αποδοχής συμβιβασμού κερδίζει 300, αν τον προτείνει ο δικηγόρος, ή 200, αν τον επιβάλλει ο δικαστής.

Φ2 Εκδίκαση με θετική έκβαση. Στην περίπτωση που ο δικηγόρος προτείνει συμβιβασμό αλλά γίνει εκδίκαση (υποχρεωτική) με θετική έκβαση, κερδίζει 150. Στην περίπτωση που ο δικηγόρος προτείνει εκδίκαση και η έκβασή της είναι θετική, κερδίζει 400.

Φ3 Εκδίκαση με αρνητική έκβαση. Στην περίπτωση που ο δικηγόρος προτείνει συμβιβασμό αλλά γίνει εκδίκαση (υποχρεωτική) με αρνητική έκβαση, χάνει 50. Στην περίπτωση που ο δικηγόρος προτείνει εκδίκαση και η έκβασή της είναι αρνητική, χάνει 50.

Κατ’ αρχάς, θα πρέπει να καταγραφεί ο πίνακας αποφάσεων, ώστε στη συνέχεια να χρησιμοποιηθεί το κατάλληλο κριτήριο για την επιλογή της βέλτιστης απόφασης. Για την καταγραφή του πίνακα αποφάσεων (Πίνακας 10.7) εφαρμόζονται τα δεδομένα που δίνονται παραπάνω για το κέρδος του δικηγόρου.

Το κριτήριο που θα χρησιμοποιηθεί θα εξαρτηθεί από τα παρακάτω ερωτήματα.

α) Ποια απόφαση θα ληφθεί αν ο δικηγόρος γνωρίζει ότι ο δικαστής πάντα δέχεται πρόταση συμβιβασμού και πάντα κάνει συμβιβασμό;

β) Ποια απόφαση θα ληφθεί όταν η πιθανότητα αποδοχής συμβιβασμού είναι 75% και η πιθανότητα θετικής έκβασης όταν γίνει δίκη είναι 30%;

γ) Ποια απόφαση θα πάρει ο δικηγόρος όταν δεν γνωρίζει τίποτα από τα παραπάνω (α, β);

**Πίνακας 10.7** *Πίνακας αποφάσεων παραδείγματος 3.5.*

| **Κέρδη για δικηγόρο** | **$Φ_Α$ συμβιβασμός δεκτός** | **$Φ_Β$ εκδίκαση με θετική έκβαση** | **$Φ_Γ$ εκδίκαση με αρνητική έκβαση** |
|---|---|---|---|
| Α προτείνει συμβιβασμό | 300 | 150 | -50 |
| Β προτείνει εκδίκαση | 200 | 400 | -50 |

Στο α) ερώτημα (συνθήκες βεβαιότητας) υπάρχει η βεβαιότητα ότι η έκβαση θα είναι συμβιβασμός, επομένως θα χρησιμοποιηθεί μόνο η στήλη $Φ_Α$ του Πίνακα 10.7. Οι άλλες δύο εκβάσεις είναι σαν να μην υπάρχουν, αφού ποτέ δεν γίνεται εκδίκαση. Γνωρίζοντας ο δικηγόρος με σιγουριά ότι θα γίνει συμβιβασμός, θα επιλέξει την απόφαση που θα του δώσει το μεγαλύτερο κέρδος, δηλαδή την απόφαση να προτείνει συμβιβασμό και να κερδίσει 300.

**Πίνακας 10.8** *Πίνακας αποφάσεων παραδείγματος 3.5 γνωρίζοντας τις πιθανότητες των εκβάσεων.*

| **Κέρδη για δικηγόρο** | **$Φ_Α$ συμβιβασμός δεκτός** | **$Φ_Β$ εκδίκαση με θετική έκβαση** | **$Φ_Γ$ εκδίκαση με αρνητική έκβαση** | **Αναμενόμενο όφελος γραμμής** | **Μέγιστο από τα αναμενόμενα οφέλη** |
|---|---|---|---|---|---|
| πιθανότητες | 75% | 30% *25%= 7,5% | 70%*25%= 17,5% | | |
| Α προτείνει συμβιβασμό | 300 | 150 | -50 | 300*0,75+150*0,075 -50*0,175 = 227,5 | 227,5 |
| Β προτείνει εκδίκαση | 200 | 400 | -50 | 200*0,75+400*0,075 -50*0,175 = 171,25 | |

Στο β) ερώτημα, γνωρίζοντας τις πιθανότητες των εκβάσεων (συνθήκες κινδύνου), καταγράφονται οι πιθανότητες αυτές στη δεύτερη γραμμή του Πίνακα 10.8 και υπολογίζεται το αναμενόμενο όφελος κάθε απόφασης. Προσοχή χρειάζεται στον υπολογισμό των πιθανοτήτων. Η έκβαση ΦΑ έχει πιθανότητα 75%, οπότε οι άλλες δύο εκβάσεις μαζί (εκδίκαση) θα έχουν πιθανότητα 25% (100%-75%). Επειδή η πιθανότητα θετικής έκβασης αν γίνει δίκη είναι 30%, η πιθανότητα της έκβασης ΦΒ θα είναι 30% του 25% = 7,5%. Αντίστοιχα η πιθανότητα της έκβασης ΦΓ θα είναι το υπόλοιπο (100%-30%) 70% του 25% = 17,5%. Γενικά, οι πιθανότητες όλων των πιθανών εκβάσεων θα πρέπει να αθροίζουν στο 100%. Το αναμενόμενο όφελος για την πρόταση συμβιβασμού υπολογίζεται σε 227,5, ενώ για την πρόταση εκδίκασης σε 171,25. Επειδή το πρώτο είναι μεγαλύτερο, ο δικηγόρος θα επιλέξει την απόφαση να προτείνει συμβιβασμό.

Στο γ) ερώτημα (συνθήκες αβεβαιότητας), θα χρησιμοποιηθούν τα κριτήρια maxmin, maxmax και minmax. Ο Πίνακας 10.9 απεικονίζει την επιλογή με το κριτήριο maxmin, από όπου προκύπτουν και οι δύο επιλογές με το ίδιο μέγιστο ελαχίστων -50. Όποια απόφαση και να επιλέξει ο δικηγόρος, το μικρότερο δυνατό αποτέλεσμά του θα περιοριστεί στο -50.

**Πίνακας 10.9** *Πίνακας αποφάσεων παραδείγματος 3.5 κριτήριο maxmin.*

| **Κέρδη για δικηγόρο** | **$Φ_Α$ συμβιβασμός δεκτός** | **$Φ_Β$ εκδίκαση με θετική έκβαση** | **$Φ_Γ$ εκδίκαση με αρνητική έκβαση** | **ελάχιστο γραμμής** | **Μέγιστο από τα ελάχιστα maxmin** |
|---|---|---|---|---|---|
| Α προτείνει συμβιβασμό | 300 | 150 | -50 | -50 | -50 |
| Β προτείνει εκδίκαση | 200 | 400 | -50 | -50 | -50 |

Ο Πίνακας 10.10 απεικονίζει την επιλογή με το κριτήριο maxmax (αισιοδοξίας), από όπου προκύπτει ως βέλτιστη η απόφαση της εκδίκασης με μέγιστο αποτέλεσμα 400.

**Πίνακας 10.10** *Πίνακας αποφάσεων παραδείγματος 3.5 κριτήριο maxmax.*

| **Κέρδη για δικηγόρο** | **$Φ_A$ συμβιβασμός δεκτός** | **$Φ_B$ εκδίκαση με θετική έκβαση** | **$Φ_Γ$ εκδίκαση με αρνητική έκβαση** | **μέγιστο γραμμής** | **Μέγιστο από τα μέγιστα maxmax** |
|---|---|---|---|---|---|
| Α προτείνει συμβιβασμό | 300 | 150 | -50 | 300 | |
| Β προτείνει εκδίκαση | 200 | 400 | -50 | 400 | 400 |

Ο Πίνακας 10.11 απεικονίζει την επιλογή με το κριτήριο minmax (κόστος χαμένης ευκαιρίας), από όπου προκύπτει ως βέλτιστη η απόφαση της εκδίκασης με μικρότερο κόστος χαμένης ευκαιρίας ίσο με 100.

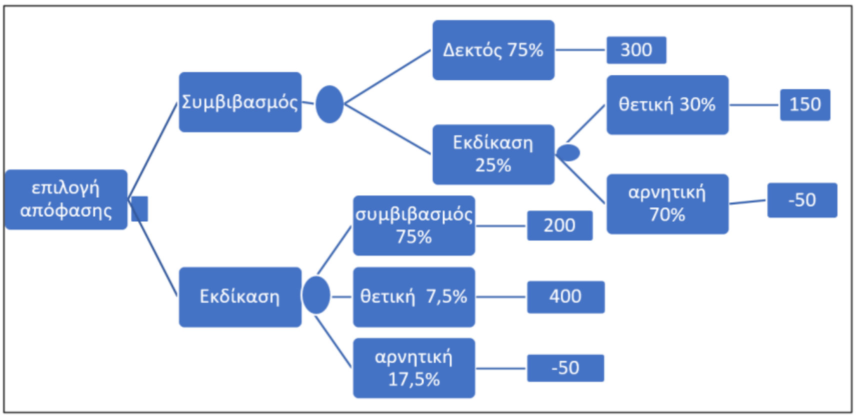
**Πίνακας 10.11** *Πίνακας αποφάσεων παραδείγματος 3.5 κριτήριο minmax.*

| **Κέρδη για δικηγόρο** | **$Φ_A$ συμβιβασμός δεκτός** | **$Φ_B$ εκδίκαση με θετική έκβαση** | **$Φ_Γ$ εκδίκαση με αρνητική έκβαση** | **μέγιστο γραμμής (κόστος χαμένης ευκαιρίας)** | **ελάχιστο από τα μέγιστα minmax** |
|---|---|---|---|---|---|
| Α προτείνει συμβιβασμό | 300-300=0 | 400-150=250 | -50-(-50)=0 | 250 | |
| Β προτείνει εκδίκαση | 300-200=100 | 400-400=0 | -50-(-50)=0 | 100 | 100 |

## 10.4 Δένδρο αποφάσεων σε συνθήκες κινδύνου

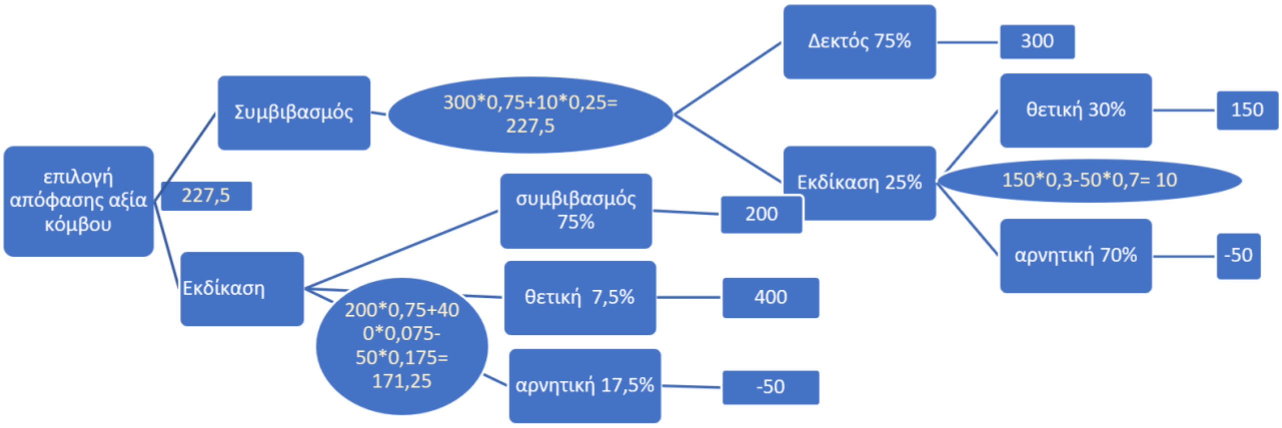
Ένας άλλος τρόπος παρουσίασης των επιλογών αποφάσεων και των εκβάσεων σε συνθήκες κινδύνου είναι η διαγραμματική παρουσίαση με κόμβους (σύμβολο ▫) που αντιστοιχούν σε σημεία αποφάσεων και κλάδους που αντιστοιχούν σε εναλλακτικές εκβάσεις. Τα σημεία που έχουμε τυχαία γεγονότα τα συμβολίζουμε με (◦), και έχουν τόσους κλάδους όσα και τα τυχαία ενδεχόμενα. Σε κάθε κλάδο έκβασης αναγράφεται η πιθανότητα της αντίστοιχης έκβασης. Σε κάθε κόμβο αποφάσεων υπολογίζεται το αναμενόμενο όφελος κάθε απόφασης.

Το δένδρο αποφάσεων του παραδείγματος 3.5 παρουσιάζεται στο Διάγραμμα 3.1 και ο υπολογισμός της αξίας των κόμβων (αναμενόμενο όφελος) παρουσιάζεται στο Διάγραμμα 3.2.




**Εικόνα 10.1** *Απεικόνιση αποφάσεων με δένδρο.*

Σε κάθε κόμβο πιθανών εκβάσεων υπολογίζεται το αναμενόμενο αποτέλεσμα, πολλαπλασιάζοντας την πιθανότητα κάθε έκβασης με το αντίστοιχο αποτέλεσμα. Σε κάθε κόμβο απόφασης επιλέγεται η απόφαση με το μεγαλύτερο αποτέλεσμα και η αξία της αντιστοιχεί στον κόμβο απόφασης. Πλεονέκτημα του δένδρου αποφάσεων είναι η εύκολη απεικόνισή του, η κατανόηση των σημείων αποφάσεων και η γρήγορη προσθήκη νέων κόμβων αποφάσεων ή εκβάσεων, αν χρειαστεί.




**Εικόνα 10.2** *Απεικόνιση αξίας κόμβων στο δένδρο αποφάσεων.*

## 10.5 Αποφάσεις σε συνθήκες ανταγωνισμού

Τα προβλήματα στα οποία υπεισέρχεται η αβεβαιότητα και θα πρέπει να ληφθεί μια απόφαση (να γίνει μια ενέργεια) σε σχέση με μία απόφαση που θα λάβει ο ανταγωνιστής (μπορεί να είναι ένας ή πολλοί), αναφέρονται με τον όρο **αποφάσεις σε συνθήκες ανταγωνισμού**. Το όφελος αυτού που λαμβάνει την απόφαση είναι σε σύγκρουση με το όφελος του ανταγωνιστή ή των ανταγωνιστών. Στις περιπτώσεις αυτές εφαρμόζεται η **θεωρία παιγνίων**. Αναπτύχθηκε σχετικά πρόσφατα τη δεκαετία 1950-60. Εφαρμόζεται στην περίπτωση λήψης μιας απόφασης, λαμβάνοντας υπόψη πιθανές αποφάσεις άλλων (κυρίως ανταγωνιστών). Δεν απαιτούνται ιδιαίτερες μαθηματικές γνώσεις για την εφαρμογή της.

**Παίγνιο** είναι μια κατάσταση με «αντίπαλο», στην οποία θα πρέπει να ληφθεί μια απόφαση με σκοπό τη βελτιστοποίηση της «ωφέλειας» του ενός **παίκτη**. Απεικονίζεται με έναν πίνακα πληρωμών για τον ένα παίκτη σε σχέση με τον ανταγωνιστή. Στα κελιά του πίνακα απεικονίζεται το κέρδος του ενός παίκτη, το οποίο είναι ταυτόχρονα ζημιά για τον ανταγωνιστή του. Οι επιλογές του παίκτη απεικονίζονται στις γραμμές του πίνακα και οι επιλογές του ανταγωνιστή του στις στήλες του πίνακα.

### 10.5.1 Προϋποθέσεις λήψης απόφασης

Στη θεωρία παιγνίων θεωρείται ότι οι αποφάσεις λαμβάνονται με βάση τη λογική. Δηλαδή μία απόφαση δεν λαμβάνεται στην τύχη. Για κάθε δυνατή απόφαση είναι γνωστές από πριν οι πιθανές συνέπειές της. Επιδίωξη της απόφασης είναι η βελτιστοποίηση του αποτελέσματος και, τέλος, η αντικειμενικότητα, αφού σε παρόμοια περίπτωση λαμβάνεται ακριβώς η ίδια απόφαση.

### 10.5.2 Παράδειγμα παιγνίου

Έστω σαν παράδειγμα παιγνίου δύο υδραυλικοί Α και Β στην ίδια περιοχή που προσπαθούν να κερδίσουν μερίδιο από τους κατοίκους της περιοχής. Ο υδραυλικός Α εφαρμόζει 3 επιλογές για την προσέλκυση πελατών. Επιλογή Α1 «συνεργασία με ηλεκτρολόγο», επιλογή Α2 «διαφημιστικά φυλλάδια», επιλογή Α3 «έκπτωση σε συχνούς πελάτες». Ο Β εφαρμόζει 3 επιλογές για να εμποδίσει τον Α στην προσέλκυση πελατών. Επιλογή Β1: κάνει διαφήμιση με μπρελόκ, επιλογή Β2: κάνει έκπτωση, και επιλογή Β3: παρέχει δωρεάν συντήρηση στις υδραυλικές εγκαταστάσεις.

**Πίνακας 10.12** *Παράδειγμα παιγνίου για δύο υδραυλικούς Α και Β.*

| **Κέρδη για Α, ανάλογα με τις στρατηγικές του Α και Β** | **Β1 διαφήμιση με μπρελόκ** | **Β2 έκπτωση** | **Β3 δωρεάν συντήρηση** |
|---|---|---|---|
| Α1 συνεργασία με ηλεκτρολόγο | 3 | 5 | 4 |
| Α2 διαφημιστικά φυλλάδια | -2 | 0 | 2 |
| Α3 έκπτωση σε συχνούς πελάτες | 2 | 4 | -3 |

Τα κέρδη του Α (σε ποσοστό επιπλέον πελατών) για κάθε συνδυασμό επιλογών παρουσιάζονται στον Πίνακα 10.12. Σημειώνεται ότι στα κελιά του πίνακα όλες οι θετικές τιμές μετρούν αύξηση ποσοστού πελατών για τον Α και οι αρνητικές τιμές μείωση του ποσοστού πελατών του Α.

Ας υποτεθεί ότι ο Α ξεκινά την έκπτωση σε συχνούς πελάτες (απόφαση Α3). Ο Β, αν κάνει έκπτωση (απόφαση Β2), θα επιτρέψει στον Α να κερδίσει 4, ενώ, αν κάνει διαφήμιση με μπρελόκ (απόφαση Β1), ο Α θα κερδίσει 2. Ο Β επιλέγει να κάνει δωρεάν συντήρηση (απόφαση Β3), ώστε τα κέρδη του Α να περιοριστούν σε -3. Στην περίπτωση της απόφασης Β3 του Β, ο Α αλλάζει τη στρατηγική του και επιλέγει τη συνεργασία με ηλεκτρολόγο (απόφαση Α1), ώστε να κερδίσει 4, που είναι και το μέγιστο της αντίστοιχης στήλης Β3.

Μετά από την απόφαση Α1 του Α, ο Β αν κάνει έκπτωση (Β2) θα επιτρέψει στον Α να κερδίσει 5, ενώ αν συνεχίσει τη δωρεάν συντήρηση (Β3) ο Α θα κερδίσει 4. Ο Β επιλέγει να κάνει διαφήμιση με μπρελόκ (απόφαση Β1), ώστε να περιορίσει τα κέρδη του Α σε 3. Στην περίπτωση της Β1 επιλογής από τον Β, ο Α δεν αλλάζει τη στρατηγική του και συνεχίζει να επιλέγει την Α1, ώστε να κερδίσει 3 που είναι και το μέγιστο της αντίστοιχης στήλης Β1. Ο Β δεν μπορεί να αλλάξει στρατηγική για να περιορίσει τον Α, αφού οι άλλες δύο στρατηγικές του Β2 και Β3 επιτρέπουν στον Α μεγαλύτερα κέρδη, 5 και 4 αντίστοιχα.

Το παίγνιο του παραδείγματος μετά από κάποιες επαναλήψεις καταλήγει σε **ισορροπία**, με τον Α να εφαρμόζει τη στρατηγική Α1 και τον Β τη Β1. Ο Α είναι κερδισμένος κατά 3 ποσοστιαίες μονάδες από τις αποφάσεις αυτές.

## 10.6 Κατηγορίες παιγνίων

Υπάρχουν πολλές κατηγορίες παιγνίων:

Ταυτόχρονα ή διαδοχικά παίγνια.

Παίγνια σύγκρουσης (κερδίζω - χάνεις) ή απόκτησης πόντων.

Επαναλαμβανόμενα ή μη επαναλαμβανόμενα παίγνια.

Παίγνια με πλήρη πληροφόρηση ή με ελλιπή πληροφόρηση.

Στη συνέχεια θα παρουσιαστούν μόνο τα παίγνια με δύο παίκτες (Α και Β) και μηδενικού αθροίσματος. Στην περίπτωση αυτή, κατά τη διάρκεια του παιγνίου οι παίκτες γνωρίζουν τόσο τις δικές τους στρατηγικές, όσο και τις στρατηγικές του αντιπάλου τους. Επίσης, κάθε παίκτης γνωρίζει ότι ο αντίπαλος γνωρίζει τα στοιχεία για τις πληρωμές του πίνακα του παιγνίου. Επιπλέον, κάθε παίκτης γνωρίζει ότι ο αντίπαλός του ξέρει ότι αυτός γνωρίζει τις πληρωμές του πίνακα κ.ο.κ. (Common Knowledge – κοινή γνώση). Ένα παράδειγμα τέτοιου παιγνίου περιγράφεται στην υποενότητα 10.8.2.

## 10.7 Προϋποθέσεις θεωρίας παιγνίων

Ο αντικειμενικός σκοπός του παίκτη Α είναι η μεγιστοποίηση του κέρδους του, ενώ του Β η ελαχιστοποίηση της ζημιάς του. Με άλλα λόγια, κάθε παίκτης προσπαθεί να ελαχιστοποιήσει το χειρότερο που μπορεί να πάθει! Αυτό για τον παίκτη Α ερμηνεύεται ως εντοπισμός του μεγαλύτερου από τα ελάχιστα (από κάθε γραμμή στρατηγικής), δηλαδή maxmin, ενώ για τον παίκτη Β ερμηνεύεται ως επιλογή του μικρότερου από τα μέγιστα (κάθε στήλη στρατηγικής), δηλαδή minmax.

Ο Α θα ακολουθεί τη λεγόμενη στρατηγική maxmin, δηλαδή επιλέγει από κάθε στρατηγική τη μικρότερη τιμή και κατόπιν επιλέγει τη μέγιστη τιμή από αυτές τις ελάχιστες τιμές. Αντίστοιχα, ο Β ακολουθεί τη στρατηγική minmax, δηλαδή επιλέγει το ελάχιστο των μεγίστων που προκύπτουν από κάθε στρατηγική.

Εάν ισχύει η σχέση minmax = maxmin = V, τότε το V αντιπροσωπεύει την τιμή του παιγνίου και το στοιχείο του πίνακα το οποίο είναι ίσο με το V ονομάζεται σημείο ισορροπίας ή σαγματικό σημείο (saddle point).

## 10.8 Έννοια της στρατηγικής

Οι διάφορες τεχνικές με τις οποίες αξιοποιούνται οι γνώσεις για το πρόβλημα, οι ικανότητες και το περιβάλλον, ώστε να επιτευχθεί το καλύτερο δυνατό αποτέλεσμα με βάση τις πληροφορίες που διατίθενται για το **τι μπορεί** να κάνει ο άλλος και όχι για το τι **πρόκειται** να κάνει.

### 10.8.1 Κυρίαρχες και υποδεέστερες στρατηγικές

Σε ένα παίγνιο πιθανόν να υπάρχουν στρατηγικές τις οποίες ο παίκτης δεν πρόκειται να επιλέξει ποτέ, γιατί το αποτέλεσμά τους είναι πάντα μικρότερο από κάποια άλλη στρατηγική. Αυτές ονομάζονται υποδεέστερες, ενώ όσες έχουν πάντα καλύτερο αποτέλεσμα ονομάζονται κυρίαρχες. Για διευκόλυνση απαλείφονται οι υποδεέστερες στρατηγικές και απλοποιείται ο πίνακας του παιγνίου.

### 10.8.2 Παράδειγμα παιγνίου με υποδεέστερη στρατηγική

Έστω σαν παράδειγμα παιγνίου δύο ξενοδοχεία Α και Β στην ίδια περιοχή που προσπαθούν να κερδίσουν μερίδιο από τους επισκέπτες της περιοχής. Δεν υπάρχει κοντά τους τρίτο ξενοδοχείο, οπότε οι επισκέπτες που πηγαίνουν στο Α ξενοδοχείο δεν πηγαίνουν στο Β, και όσοι επισκέπτες δεν πηγαίνουν στο Α ξενοδοχείο πάνε στο Β ξενοδοχείο. Οι διευθυντές των δύο ξενοδοχείων δρουν ανταγωνιστικά μεταξύ τους, με στόχο την προσέλκυση των επισκεπτών της περιοχής.

Ο παίκτης Α (ως προς τον οποίο αναζητείται η βέλτιστη λύση) έχει στη διάθεσή του 4 στρατηγικές. Την Α1 «διαφήμιση σε ταξιδιωτικά γραφεία», την Α2 «έκπτωση διανυκτέρευσης», την Α3 «προσφορά σε συχνούς πελάτες» και την Α4 «κλήρωση αναμνηστικών δώρων στους πελάτες».

Ο ανταγωνιστής του Β έχει στη διάθεσή του 3 στρατηγικές, προκειμένου να περιορίσει τον Α να κερδίσει. Τη Β1 «διαφήμιση με αφίσες στην περιοχή», τη Β2 «έκπτωση διανυκτέρευσης», και τη Β3 «δωρεάν περιήγηση στην περιοχή».

**Πίνακας 10.13** *Παράδειγμα παιγνίου για δύο ανταγωνιστικά ξενοδοχεία.*

| **Κέρδη για Α, ανάλογα με τις στρατηγικές του Α και Β** | **Β1 διαφήμιση με αφίσες** | **Β2 έκπτωση διανυκτέρευσης** | **Β3 δωρεάν περιήγηση** |
|---|---|---|---|
| Α1 διαφήμιση σε ταξιδιωτικά γραφεία | 3 | 2 | 4 |
| Α2 έκπτωση διανυκτέρευσης | 5 | 1 | 0 |
| Α3 προσφορά σε συχνούς πελάτες | 3 | 0 | 5 |
| Α4 κλήρωση αναμνηστικών | 4 | 5 | 5 |

Τα κέρδη του Α (σε δεκάδες επισκεπτών) για κάθε συνδυασμό στρατηγικών παρουσιάζονται στον Πίνακα 10.13. Σημειώνεται ότι στα κελιά του πίνακα όλες οι τιμές μετρούν κέρδη μεριδίου επισκεπτών για τον Α διευθυντή.

Ας υποθέσουμε ότι ο διευθυντής Α ξεκινά τη διαφήμιση στα ταξιδιωτικά γραφεία. Ο Β, αν κάνει διαφήμιση με αφίσες, θα επιτρέψει στον Α να κερδίσει 3 δεκάδες, ενώ αν κάνει δωρεάν περιήγηση ο Α θα κερδίσει 4 δεκάδες με τη διαφήμισή του σε ταξιδιωτικά γραφεία. Ο Β επιλέγει να κάνει έκπτωση διανυκτέρευσης (Β2), ώστε να περιορίσει τους πελάτες του Α σε 2 δεκάδες. Στην περίπτωση της απόφασης Β2, όπου ο Β κάνει έκπτωση, ο Α αλλάζει τη στρατηγική του και επιλέγει Α4 να κληρώσει αναμνηστικά, ώστε να κερδίσει 5 δεκάδες πελατών.

Ως αντίδραση ο Β αλλάζει στρατηγική και επιλέγει τη Β1 (διαφήμιση με αφίσες), ώστε να περιορίσει τον Α σε 4 δεκάδες αντί για 5. Στη συνέχεια, ο Α επίσης αλλάζει στρατηγική και επιλέγει Α2 (έκπτωση διανυκτέρευσης), ώστε να κερδίσει 5 δεκάδες με τη στρατηγική του Β1 (διαφήμιση). Είναι η σειρά του Β να αλλάξει τώρα στρατηγική επιλέγοντας Β3 (δωρεάν περιήγηση) και περιορίζοντας τον Α στο 0. Η αντίδρασή του Α είναι να επιλέξει την Α3 ή Α4 στρατηγική, αφού και με τις δύο κερδίζει 5 δεκάδες. Έστω ότι επιλέγει την Α3. Τότε, ο Β αλλάζει και αυτός στρατηγική και επιλέγει τη Β2. Στη συνέχεια, ο Α επιλέγει την Α4 και ο Β τη Β1 στρατηγική. Η διαδικασία αυτή αλλαγής στρατηγικών επαναλαμβάνεται. Στις επαναλήψεις αυτές παρατηρούμε ότι ο Α δεν έχει κέρδος να επιλέξει την Α1 στρατηγική και δεν την επιλέγει στη συνέχεια. Η στρατηγική Α1 είναι υποδεέστερη και μπορεί να διαγραφεί από τον πίνακα του παιγνίου.

### 10.8.3 Εντοπισμός υποδεέστερων στρατηγικών

Μία στρατηγική του Α (γραμμή του πίνακα) είναι υποδεέστερη μίας άλλης, όταν όλα τα στοιχεία της γραμμής της πρώτης στρατηγικής είναι μικρότερα από τα στοιχεία της γραμμής της δεύτερης στρατηγικής. Μία στρατηγική του Β (στήλη του πίνακα) είναι υποδεέστερη μίας άλλης, όταν όλα τα στοιχεία της στήλης της πρώτης στρατηγικής είναι μεγαλύτερα από τα στοιχεία της στήλης της δεύτερης στρατηγικής. Οι υποδεέστερες στρατηγικές μπορούν να διαγραφούν από τον πίνακα του παιγνίου, αφού ποτέ δεν συμφέρει η επιλογή τους στον παίκτη Α ή στον παίκτη Β. Στον Πίνακα 10.12 του παραδείγματος, η στρατηγική Β2 είναι υποδεέστερη της Β1, αφού όλα τα στοιχεία της Β2 είναι μεγαλύτερα από τα στοιχεία της Β1. Η Β1 ονομάζεται και κυρίαρχη έναντι της Β2. Επίσης, η στρατηγική Α2 είναι υποδεέστερη της Α1, αφού όλα τα στοιχεία της Α2 είναι μικρότερα από τα στοιχεία της Α1. Επίσης, η στρατηγική Α3 είναι υποδεέστερη της Α1, αφού όλα τα στοιχεία της Α3 είναι μικρότερα από τα στοιχεία της Α1. Μετά από τη διαγραφή των στρατηγικών Β2, Α2, Α3, προκύπτει και η στρατηγική Β3 υποδεέστερη της Β1, αφού το στοιχείο της είναι μεγαλύτερο. Καταλήγουμε, επομένως, σε μία μόνο στρατηγική Α1 για τον Α και μία στρατηγική Β1 για τον Β, που οδηγούν το παίγνιο σε ισορροπία.

## 10.9 Αμιγής και μεικτή στρατηγική

Όταν κάθε παίκτης μπορεί να εντοπίσει μία στρατηγική για να ισορροπήσει, τότε η στρατηγική του αυτή ονομάζεται **αμιγής** στρατηγική. Είναι δυνατόν σε ένα παίγνιο να υπάρχουν περισσότερα του ενός «σαγματικά» σημεία (σημεία με την ίδια φυσικά τιμή) και, επομένως, περισσότερες από μία αμιγείς στρατηγικές. Στην υποενότητα 10.5.2 παρουσιάστηκε ένα παράδειγμα παιγνίου με ισορροπία, όπου η αμιγής στρατηγική του Α υδραυλικού ήταν η Α1 και του Β η Β1.

Υπάρχουν, ωστόσο, και παίγνια δύο παικτών μηδενικού αθροίσματος χωρίς σημείο ισορροπίας, όπως αυτό που παρουσιάστηκε στην υποενότητα 10.8.2. Σε αυτά εφαρμόζεται η **μεικτή** στρατηγική, ώστε σε πολλές επαναλήψεις του παιγνίου να επιτευχθεί το μέγιστο αναμενόμενο όφελος. Όταν δεν υπάρχουν αμιγείς στρατηγικές, δηλαδή δεν υπάρχει ισορροπία, κάθε παίκτης εναλλάσσει τυχαία τις στρατηγικές του, ώστε ο αντίπαλος να μην ξέρει τις κινήσεις του. Στην περίπτωση αυτή, υπολογίζεται η πιθανότητα κάθε παίκτης να εφαρμόζει μία στρατηγική, θεωρώντας ότι η αναμενόμενη τιμή παιγνίου είναι ίδια άσχετα με το τι θα κάνει ο άλλος παίκτης.

## 10.10 Επίλυση παιγνίων

Επίλυση παιγνίων ονομάζεται η διαδικασία εύρεσης των στρατηγικών που οδηγούν στη βέλτιστη λύση για τον παίκτη Α. Η επίλυση μπορεί να γίνει με την εύρεση αμιγούς στρατηγικής, αν υπάρχει ή με την εύρεση των μεικτών στρατηγικών. Για παίγνια με πίνακα 2×2 η διαδικασία επίλυσης απαιτεί τη λύση ενός συστήματος εξισώσεων. Αν οι διαστάσεις του πίνακα (πλήθος γραμμών ή στηλών) είναι μεγαλύτερες από 2 μετά από την αφαίρεση υποδεέστερων στρατηγικών, απαιτείται η επίλυση με τη γραφική μέθοδο.

### 10.10.1 Εντοπισμός αμιγούς στρατηγικής (ισορροπίας)

Η διερεύνηση της ισορροπίας ενός παιγνίου μπορεί να γίνει με τον εντοπισμό και τη διαγραφή όλων των υποδεέστερων στρατηγικών. Μπορεί όμως να γίνει και ευκολότερα, αν εντοπίσουμε το μέγιστο από τα ελάχιστα κάθε γραμμής (maxmin) και το ελάχιστο από τα μέγιστα κάθε στήλης (minmax). Αν οι δύο αυτοί αριθμοί συμπίπτουν, το παίγνιο έχει ισορροπία που αντιστοιχεί στη γραμμή και στήλη που εντοπίστηκαν οι 2 ίσοι αριθμοί.

Στον Πίνακα 10.14 έχουμε τα στοιχεία για το παράδειγμα της υποενότητας 10.5.2. Στην τελευταία στήλη εντοπίσαμε το ελάχιστο κάθε γραμμής, ενώ στην τελευταία γραμμή καταγράψαμε το μέγιστο κάθε στήλης.

**Πίνακας 10.14** *Παράδειγμα εντοπισμού ισορροπίας παιγνίου για δύο υδραυλικούς Α και Β.*

| **Κέρδη για Α, ανάλογα με τις στρατηγικές του Α και Β** | **Β1 διαφήμιση με μπρελόκ** | **Β2 έκπτωση** | **Β3 δωρεάν συντήρηση** | **min** |
|---|---|---|---|---|
| Α1 συνεργασία με ηλεκτρολόγο | 3 | 5 | 4 | 3 |
| Α2 διαφημιστικά φυλλάδια | -2 | 0 | 2 | -2 |
| Α3 έκπτωση σε συχνούς πελάτες | 2 | 4 | -3 | -3 |
| Max | 3 | 5 | 4 | |

Το maxmin (μέγιστο των ελαχίστων) της τελευταίας στήλης είναι ο αριθμός 3 (Α1 στρατηγικής) και το minmax (ελάχιστο των μεγίστων) της τελευταίας γραμμής είναι ο αριθμός 3 (Β1 στρατηγική). Επειδή maxmin=minmax=3, συμπεραίνουμε ότι το παίγνιο θα έχει ισορροπία με τη στρατηγική Α1 για τον Α υδραυλικό και Β1 για τον Β υδραυλικό και το κέρδος του Α θα είναι 3, όπως και αναλυτικά περιγράφηκε στην υποενότητα 10.5.2.

## 10.10.2 Εντοπισμός μεικτής στρατηγικής για πίνακα παιγνίου 2 × 2

Έστω σαν παράδειγμα παιγνίου δύο παντοπωλεία Α και Β στην ίδια περιοχή που προσπαθούν να κερδίσουν μερίδιο πελατών. Οι ιδιοκτήτες δρουν ανταγωνιστικά μεταξύ τους με στόχο την προσέλκυση πελατών, εφαρμόζοντας και οι δύο στρατηγική διαφήμισης και στρατηγική έκπτωσης. Ο πίνακας κερδών του παντοπωλείου Α παρουσιάζεται στον Πίνακα 10.15.

**Πίνακας 10.15** *Παράδειγμα για παντοπωλεία Α και Β.*

| **Κέρδη για Α, ανάλογα με τις στρατηγικές του Α και Β** | **Β1 διαφήμιση** | **Β2 έκπτωση** |
|---|---|---|
| Α1 διαφήμιση | 2 | 4 |
| Α2 έκπτωση | 5 | 3 |

Δεν υπάρχουν υποδεέστερες στρατηγικές και δεν υπάρχει και ισορροπία στο παίγνιο. Έτσι, οι δύο ιδιοκτήτες εφαρμόζουν μεικτή στρατηγική, εναλλάσσοντας τις στρατηγικές τους. Ας υποθέσουμε ότι ο Α εφαρμόζει σε ποσοστό χ% τη διαφήμιση και στο υπόλοιπο (1-χ)% την έκπτωση. Ας υποθέσουμε ότι ο Β εφαρμόζει σε ποσοστό ψ% τη διαφήμιση και στο υπόλοιπο (1-ψ)% την έκπτωση. Προφανώς τα χ και ψ δεν είναι ίδια και δεν γνωρίζει ο ένας ποια στρατηγική θα κάνει ο άλλος κάθε φορά.

Στόχος του Α είναι να βρει τέτοιο ποσοστό χ%, ώστε, όποια στρατηγική και να εφαρμόσει ο Β, ο Α να έχει το ίδιο αναμενόμενο κέρδος. Αν ο Β ακολουθήσει τη Β1 στρατηγική, το αναμενόμενο κέρδος του Α θα είναι 2χ+5(1-χ), το οποίο υπολογίζεται πολλαπλασιάζοντας τα στοιχεία της Β1 στήλης με τα ποσοστά χ% και (1-χ)% των επιλογών του Α για την Α1 και Α2 στρατηγική του. Αν ο Β ακολουθήσει τη Β2 στρατηγική, το αναμενόμενο κέρδος του Α θα είναι 4χ+3(1-χ), το οποίο υπολογίζεται πολλαπλασιάζοντας τα στοιχεία της Β2 στήλης με τα ποσοστά χ% και (1-χ)% των επιλογών του Α για την Α1 και Α2 στρατηγική του. Επειδή ο στόχος του Α είναι να έχει το ίδιο αναμενόμενο κέρδος είτε ο Β ακολουθήσει τη Β1 είτε τη Β2, πρέπει να ισχύει η εξίσωση:

$$2\chi+5(1-\chi) = 4\chi+3(1-\chi)$$

Η εξίσωση αυτή είναι πρώτου βαθμού, λύνεται αν βγάλουμε τις παρενθέσεις, χωρίσουμε γνωστούς από αγνώστους και διαιρέσουμε με τον συντελεστή του αγνώστου. Προκύπτει:

$$2\chi+5(1-\chi) = 4\chi+3(1-\chi) \Leftrightarrow 2\chi+5-5\chi = 4\chi+3-3\chi \Leftrightarrow 2\chi-5\chi-4\chi+3\chi = 3-5 \Leftrightarrow -4\chi = -2 \Leftrightarrow \chi = 1/2 = 0{,}5 = 50\%$$

Η λύση είναι χ = 50%, που σημαίνει ότι ο Α κατά 50% επιλέγει την Α1 στρατηγική και στο υπόλοιπο 50% την Α2 στρατηγική.

Με παρόμοιο τρόπο υπολογίζεται και το ψ% της στρατηγικής Β1 που επιλέγεται από τον Β. Στόχος του Β είναι να βρει το κατάλληλο ψ%, ώστε, όποια στρατηγική και να εφαρμόσει ο Α, να τον περιορίζει στο ίδιο αναμενόμενο κέρδος. Αν ο Α ακολουθήσει την Α1 στρατηγική, το αναμενόμενο κέρδος του Β θα είναι 2ψ+4(1-ψ), το οποίο υπολογίζεται πολλαπλασιάζοντας τα στοιχεία της Α1 γραμμής με τα ποσοστά ψ% και (1-ψ)% των επιλογών του Β για τη Β1 και Β2 στρατηγική του. Αν ο Α ακολουθήσει την Α2 στρατηγική, το αναμενόμενο κέρδος του Β θα είναι 5ψ+3(1-ψ), το οποίο υπολογίζεται πολλαπλασιάζοντας τα στοιχεία της Α2 γραμμής με τα ποσοστά ψ% και (1-ψ)% των επιλογών του Β για τη Β1 και Β2 στρατηγική του. Επειδή ο στόχος του Β είναι το ίδιο αναμενόμενο κέρδος είτε ο Α ακολουθήσει την Α1 είτε την Α2, πρέπει να ισχύει η εξίσωση:

$$2ψ+4(1-ψ) = 5ψ+3(1-ψ)$$

Η εξίσωση αυτή είναι πρώτου βαθμού, λύνεται αν βγάλουμε τις παρενθέσεις, χωρίσουμε γνωστούς από αγνώστους και διαιρέσουμε με τον συντελεστή του αγνώστου. Προκύπτει:

$$2ψ+4(1-ψ) = 5ψ+3(1-ψ) \Leftrightarrow 2ψ+4-4ψ=5ψ+3-3ψ \Leftrightarrow 2ψ-4ψ-5ψ+3ψ=3-4 \Leftrightarrow -4ψ=-1 \Leftrightarrow χ=1/4=0,25 = 25\%$$

Η λύση είναι ψ = 25%, που σημαίνει ότι ο Β κατά 25% επιλέγει τη Β1 στρατηγική και στο υπόλοιπο 75% τη Β2 στρατηγική.

**Τιμή V του παιγνίου** ονομάζεται το αναμενόμενο κέρδος του Α από τη μεικτή στρατηγική και υπολογίζεται αν αντικατασταθεί η τιμή του χ στη σχέση 2χ+5(1-χ) ή στην ισοδύναμη σχέση 4χ+3(1-χ), από όπου προκύπτει:

$$V = 2χ+5(1-χ) = 2*0,5+5(1-0,5) = 1+2,5 = 3,5 \text{ και } V = 4χ+3(1-χ) = 4*0,5+3(1-0,5) = 2+1,5 = 3,5.$$

Η ίδια τιμή **V** προκύπτει και αν αντικατασταθεί η τιμή ψ στη σχέση 2ψ+4(1-ψ) ή στην ισοδύναμη σχέση 5ψ+3(1-ψ), από όπου προκύπτει:

$$V = 2ψ+4(1-ψ) = 2*0,25+4(1-0,25) = 0,5+3 = 3,5 \text{ και } V = 5ψ+3(1-ψ) = 5*0,25+3(1-0,25) = 1,25+2,25 = 3,5.$$

## 10.11 Γραφική μέθοδος επίλυσης παιγνίων

Στην περίπτωση που δεν υπάρχει ισορροπία σε ένα παίγνιο με διαστάσεις nx2 (2 στρατηγικές για τον Β και περισσότερες από 2 στρατηγικές για τον Α), χρησιμοποιείται η γραφική μέθοδος, ώστε να μετατραπεί το παίγνιο σε 2×2 και κατόπιν να βρεθεί η μεικτή στρατηγική. Σχεδιάζονται γραφικά οι στρατηγικές του Α σε σχέση με τις 2 στρατηγικές του Β και επιλέγεται το χαμηλότερο σημείο του υψηλότερου τεθλασμένου ευθύγραμμου τμήματος minmax. Αυτό γίνεται γιατί ο Α προσπαθεί για το μέγιστο κέρδος και ο Β προσπαθεί να περιορίσει τον Α στο χαμηλότερο σημείο των μέγιστων κερδών του.

Στην περίπτωση που δεν υπάρχει ισορροπία σε ένα παίγνιο με διαστάσεις 2xn (2 στρατηγικές για τον Α και περισσότερες από 2 στρατηγικές για τον Β, χρησιμοποιείται η γραφική μέθοδος, ώστε να μετατραπεί το παίγνιο σε 2×2 και κατόπιν να βρεθεί η μεικτή στρατηγική. Σχεδιάζονται γραφικά οι στρατηγικές του Β σε σχέση με τις 2 στρατηγικές του Α και επιλέγεται το υψηλότερο σημείο του χαμηλότερου τεθλασμένου ευθύγραμμου τμήματος maxmin. Αυτό γίνεται γιατί ο Β προσπαθεί για το ελάχιστο κέρδος του Α και ο Α προσπαθεί για το μέγιστο σημείο των κερδών του.

Το σημείο που επιλέγεται αντιστοιχεί στο σημείο τομής δύο στρατηγικών του Α ή του Β, ανάλογα με το είδος του πίνακα του παιγνίου. Επομένως, στη συνέχεια, οι υπόλοιπες στρατηγικές διαγράφονται και προκύπτει πίνακας 2 στρατηγικών για τον Α και 2 στρατηγικών για τον Β. Εφαρμόζεται τότε ο υπολογισμός της μεικτής στρατηγικής, όπως περιγράφηκε στην υποενότητα 10.10.2.

## 10.12 Εφαρμογές για κατανόηση

### 10.12.1 Επιλογή συνεταίρου

#### 10.12.1.1 Παρουσίαση προβλήματος

Ο παίκτης Α σκοπεύει να συνεργαστεί με έναν από τρεις υποψήφιους συνεταίρους. Η μελλοντική συνεργασία τους μπορεί να έχει πολύ καλή κατάληξη, μέτρια κατάληξη ή άσχημη κατάληξη, με πιθανότητα 20%, 50% και 30% αντίστοιχα. Το όφελος από τη συνεργασία με τον καθένα διαφέρει ανάλογα με τις ικανότητες του καθενός και ανάλογα με το μέλλον της συνεργασίας, όπως φαίνεται στον Πίνακα 10.16. Ποιον συνεταίρο θα επιλέξει; Αν οι δοσμένες πιθανότητες ήταν άγνωστες, ποιον θα επέλεγε;

**Πίνακας 10.16** *Πίνακας αποφάσεων παραδείγματος 10.12.1.*

| Συνεταίρος | Καλή | Μέτρια | Άσχημη |
|---|---|---|---|
| Σ1 | 3.000 | 1.500 | -4.000 |
| Σ2 | 10.000 | 8.000 | -6.000 |
| Σ3 | 6.000 | 4.000 | -1.500 |

#### 10.12.1.2 Λύση με αναμενόμενο όφελος

Στον Πίνακα 10.17 παρουσιάζονται οι υπολογισμοί για το αναμενόμενο όφελος. Καλύτερη επιλογή προκύπτει να είναι η επιλογή του συνεταίρου Σ2.

**Πίνακας 10.17** *Πίνακας αποφάσεων παραδείγματος 10.12.1 με αναμενόμενο όφελος.*

| Συνεταίρος | Καλή | Μέτρια | Άσχημη | Αναμενόμενο όφελος | Βέλτιστο |
|---|---|---|---|---|---|
| **Πιθανότητες** | **20%** | **50%** | **30%** | | |
| Σ1 | 3.000 | 1.500 | -4.000 | =3.000*0,20+1.500*0,50-4.000*0,30 =150 | |
| Σ2 | 10.000 | 8.000 | -6.000 | =10.000*0,20+8.000*0,50-6.000*0,30 =4.200 | 4.200 |
| Σ3 | 6.000 | 4.000 | -1.500 | =6.000*0,20+4.000*0,50-1.500*0,30 =2.750 | |

#### 10.12.1.3 Λύση με κριτήριο maxmin

Στον Πίνακα 10.18 παρουσιάζονται οι υπολογισμοί για το κριτήριο maxmin. Καλύτερη επιλογή προκύπτει να είναι η επιλογή του συνεταίρου Σ3.

**Πίνακας 10.18** *Πίνακας αποφάσεων παραδείγματος 10.12.1 με κριτήριο maxmin.*

| Συνεταίρος | Καλή | Μέτρια | Άσχημη | Min | Βέλτιστο maxmin |
|---|---|---|---|---|---|
| Σ1 | 3.000 | 1.500 | -4.000 | -4.000 | |
| Σ2 | 10.000 | 8.000 | -6.000 | -6.000 | |
| Σ3 | 6.000 | 4.000 | -1.500 | -1.500 | -1.500 |

α

### 10.12.1.4 Λύση με κριτήριο maxmax

Στον Πίνακα 10.19 παρουσιάζονται οι υπολογισμοί για το κριτήριο maxmax. Καλύτερη επιλογή προκύπτει να είναι η επιλογή του συνεταίρου Σ2.

**Πίνακας 10.19** *Πίνακας αποφάσεων παραδείγματος 10.12.1 με κριτήριο maxmax.*

| Συνεταίρος | Καλή | Μέτρια | Άσχημη | Max | Βέλτιστο maxmax |
|---|---|---|---|---|---|
| Σ1 | 3.000 | 1.500 | -4.000 | 3.000 | |
| Σ2 | 10.000 | 8.000 | -6.000 | 10.000 | 10.000 |
| Σ3 | 6.000 | 4.000 | -1.500 | 6.000 | |

### 10.12.1.5 Λύση με κριτήριο minmax

Στον Πίνακα 10.20 παρουσιάζονται οι υπολογισμοί για το κριτήριο minmax. Υπολογίζεται για κάθε στήλη το κόστος χαμένης ευκαιρίας και κατόπιν το μέγιστο κόστος χαμένης ευκαιρίας για κάθε γραμμή-απόφαση. Καλύτερη επιλογή προκύπτει να είναι η επιλογή του συνεταίρου Σ3, στον οποίο αντιστοιχεί το μικρότερο από τα μέγιστα κόστη χαμένης ευκαιρίας.

**Πίνακας 10.20** *Πίνακας αποφάσεων παραδείγματος 10.12.1 με κριτήριο minmax.*

| Συνεταίρος | Καλή (Κόστος χαμένης ευκαιρίας) | | Μέτρια (Κόστος χαμένης ευκαιρίας) | | Άσχημη (Κόστος χαμένης ευκαιρίας) | | Max | Βέλτιστο minmax |
|---|---|---|---|---|---|---|---|---|
| | όφελος | αφαίρεση από το 10.000 | όφελος | αφαίρεση από το 8.000 | όφελος | αφαίρεση από το -1.500 | Κόστος χαμένης ευκαιρίας | |
| Σ1 | 3.000 | 7.000 | 1.500 | 6.500 | -4.000 | 2.500 | 7.000 | |
| Σ2 | 10.000 | 0 | 8.000 | 0 | -6.000 | 4.500 | 4.500 | |
| Σ3 | 6.000 | 4.000 | 4.000 | 4.000 | -1.500 | 0 | 4.000 | 4.000 |

## 10.12.2 Πρόσληψη βοηθών

Κατά την περίοδο των φορολογικών δηλώσεων η δουλειά ενός λογιστικού γραφείου αυξάνεται και ο υπεύθυνος μπορεί να προσλάβει και βοηθούς. Το ημερήσιο πλήθος των επιπλέον πελατών που έρχονται την περίοδο αυτή είναι 6 ή 10 ή 14 ή 18 πελάτες με πιθανότητες 30%, 40%, 20%, 10% αντίστοιχα. Το ημερομίσθιο ενός βοηθού είναι 25 ευρώ και τα επιπλέον έσοδα από την εργασία του είναι 30 ευρώ ανά πελάτη και μπορεί να εξυπηρετήσει μέχρι 6 πελάτες την ημέρα. Ο πελάτης που δεν μπορεί να εξυπηρετηθεί και φύγει κοστίζει 20 ευρώ (δυσφήμιση – δεν ξανάρχεται). Πόσοι βοηθοί θα επιλεγούν (ένας, δύο ή τρεις); Ποιο το αναμενόμενο κέρδος; Αν οι δοσμένες πιθανότητες ήταν άγνωστες, πόσοι βοηθοί θα επιλεγούν;

Στον Πίνακα 10.21 παρουσιάζονται οι υπολογισμοί για τα οφέλη που προκύπτουν ανάλογα με το πιθανό πλήθος πελατών που θα προσέλθουν και το πλήθος βοηθών. Για παράδειγμα, αν έρθουν 14 πελάτες και έχουμε 2 βοηθούς, θα εξυπηρετηθούν οι 12 (6 πελάτες ανά βοηθό), αφήνοντας κέρδος 30*12, θα υπάρχει ζημία 2*25 από τους πελάτες που θα φύγουν χωρίς να εξυπηρετηθούν και τα έξοδα για την αμοιβή των δύο βοηθών θα είναι 2*20, οπότε το συνολικό όφελος θα είναι 30*12-2*25-2*20= 270. Παρόμοιοι υπολογισμοί γίνονται και στα υπόλοιπα κελιά.

**Πίνακας 10.21** *Πίνακας οφέλους αποφάσεων παραδείγματος 10.12.2.*

| Πλήθος βοηθών | 6 πελάτες | 10 πελάτες | 14 πελάτες | 18 πελάτες |
|---|---|---|---|---|
| 1 | 30*6-25= 155 | 30*6-25-4*20= 75 | 30*6-25-8*20= -5 | 30*6-25-12*20= -85 |
| 2 | 30*6-2*25= 130 | 30*10-2*25= 250 | 30*12-2*25-2*20= 270 | 30*12-2*25-6*20= 190 |
| 3 | 30*6-3*25= 105 | 30*10-3*25= 225 | 30*14-3*25= 345 | 30*18-3*25= 465 |

### 10.12.2.1 Λύση με κριτήριο αναμενόμενου οφέλους

Στον Πίνακα 10.22 παρουσιάζονται οι υπολογισμοί για το κριτήριο αναμενόμενου οφέλους. Καλύτερη επιλογή προκύπτει να είναι η πρόσληψη 3 βοηθών, με αναμενόμενο όφελος 237.

**Πίνακας 10.22** *Πίνακας αποφάσεων παραδείγματος 10.12.2 με αναμενόμενο όφελος.*

| Πλήθος βοηθών | 6 πελάτες | 10 πελάτες | 14 πελάτες | 18 πελάτες | Αναμενόμενο όφελος |
|---|---|---|---|---|---|
| 1 | 155 | 75 | -5 | -85 | =155*0,3+75*0,4-5*0,2-85*0,1= 67 |
| 2 | 130 | 250 | 270 | 190 | =130*0,3+250*0,4+270*0,2+190*0,1= 212 |
| 3 | 105 | 225 | 345 | 465 | =105*0,3+225*0,4+345*0,2+465*0,1= 237 |

### 10.12.2.2 Λύση με κριτήριο maxmin

Στον Πίνακα 10.23 παρουσιάζονται οι υπολογισμοί για το κριτήριο maxmin. Καλύτερη επιλογή προκύπτει να είναι η πρόσληψη 2 βοηθών, όπου ο όφελος θα είναι σίγουρα άνω του 130.

**Πίνακας 10.23** *Πίνακας αποφάσεων παραδείγματος 10.12.2 με maxmin.*

| Πλήθος βοηθών | 6 πελάτες | 10 πελάτες | 14 πελάτες | 18 πελάτες | min | maxmin |
|---|---|---|---|---|---|---|
| 1 | 155 | 75 | -5 | -85 | -85 | |
| 2 | 130 | 250 | 270 | 190 | 130 | 130 |
| 3 | 105 | 225 | 345 | 465 | 105 | |

### 10.12.2.3 Λύση με κριτήριο maxmax

Στον Πίνακα 10.24 παρουσιάζονται οι υπολογισμοί για το κριτήριο maxmax. Καλύτερη επιλογή προκύπτει να είναι η πρόσληψη 3 βοηθών, με την ελπίδα να επιτευχθεί το μεγαλύτερο δυνατό όφελος 465.

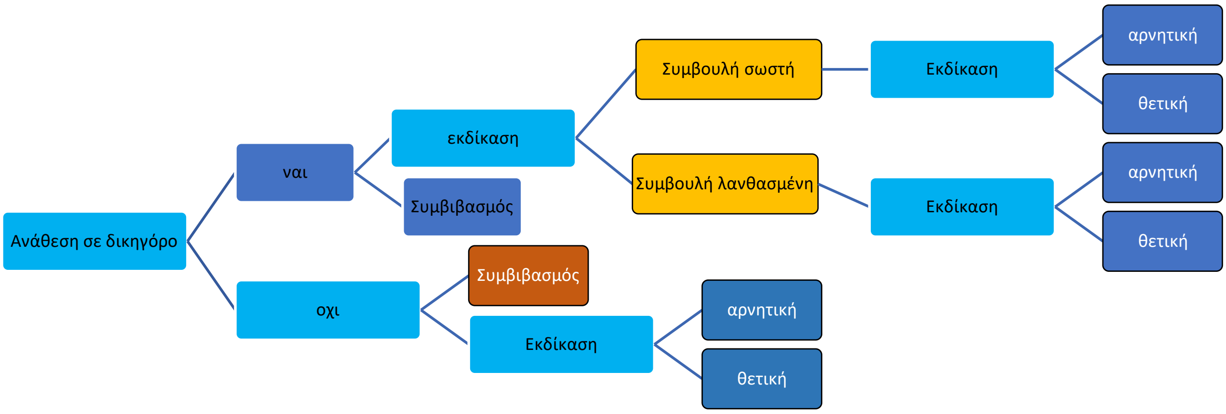
**Πίνακας 10.24** *Πίνακας αποφάσεων παραδείγματος 10.12.2 με maxmax.*

| Πλήθος βοηθών | 6 πελάτες | 10 πελάτες | 14 πελάτες | 18 πελάτες | max | maxmax |
|---|---|---|---|---|---|---|
| 1 | 155 | 75 | -5 | -85 | 155 | |
| 2 | 130 | 250 | 270 | 190 | 270 | |
| 3 | 105 | 225 | 345 | 465 | 465 | 465 |

### 10.12.3 Συμβιβασμός σε αγωγή

Ένας δικηγόρος αντιμετωπίζει μία αγωγή που έγινε σε πελάτη για παράβαση. Οι επιλογές είναι: Ο πελάτης να συμβιβαστεί πληρώνοντας 5.000 ευρώ ή να γίνει δίκη με πιθανότητα 50% να τη χάσει και να πληρώσει 15.000 ευρώ ή να την κερδίσει και να μην πληρώσει τίποτα. Εναλλακτικά, μπορεί να αναθέσει στον δικηγόρο την εξέταση της παράβασης, και αφού εκείνος του δώσει τη συμβουλή του, να αποφασίσει αν θα συμβιβαστεί ή όχι. Η αμοιβή του δικηγόρου θα είναι 2.000 ευρώ και οι συμβουλές-εκτιμήσεις του δικηγόρου σε παρόμοιες περιπτώσεις που προέβλεπαν τη θετική έκβαση της δίκης ήταν σωστές κατά 80%, ενώ αυτές που προέβλεπαν την αρνητική έκβαση ήταν σωστές στο 90% των περιπτώσεων. Ο δικηγόρος στο 10% παρόμοιων περιπτώσεων προτείνει συμβιβασμό και επιστρέφει την αμοιβή του. Αν η συμβουλή του βγει τελικά λανθασμένη, επιστρέφει τη μισή αμοιβή του. Η εκδίκαση έχει δύο πιθανές εκβάσεις, να κερδηθεί (θετική) ή να μην κερδηθεί (αρνητική). Αν η εκδίκαση κερδηθεί, ο αντίδικος πληρώνει τα έξοδα του δικηγόρου. Σχεδιάστε το δένδρο αποφάσεων. Υπολογίστε τις πιθανότητες για κάθε κλάδο. Ποιο το αναμενόμενο ποσό που θα πληρώσετε σε κάθε περίπτωση;

Στην Εικόνα 10.3 παρουσιάζονται οι αποφάσεις και οι εκβάσεις με το δένδρο αποφάσεων. Οι αρχικές επιλογές είναι δύο, να ανατεθεί η υπόθεση σε δικηγόρο ή όχι.




**Εικόνα 10.3** *Απεικόνιση εφαρμογής 10.12.3 με δένδρο αποφάσεων.*

Στην Εικόνα 10.4 παρουσιάζονται τα οφέλη για κάθε περίπτωση αποφάσεων και εκβάσεων πάνω στο δένδρο αποφάσεων. Για την επιλογή της βέλτιστης απόφασης, από δεξιά προς αριστερά, υπολογίζεται η αξία κάθε κόμβου.

Θα πρέπει να υπολογιστούν και οι πιθανότητες για τους κόμβους που δίνει συμβουλή ο δικηγόρος και για την έκβαση της δίκης με βάση τη συμβουλή του. Πιθανότητες για συμβουλή για θετική έκβαση με βάση τα δεδομένα στο 80% είναι σωστή και στο 20% λανθασμένη, δηλαδή παρά τη συμβουλή ότι θα είναι θετική η έκβαση, στην πραγματικότητα είναι αρνητική. Ισχύουν οι εξής δεσμευμένες πιθανότητες:

P(θετική |συμβουλή θετική) = 0,80  P(αρνητική |συμβουλή θετική) = 0,20

Αντίστοιχα, όταν ο δικηγόρος συμβουλεύει για αρνητική έκβαση, στο 90% επαληθεύεται και στο 10% διαψεύδεται. Οι δεσμευμένες πιθανότητες για συμβουλή αρνητική έκβαση είναι:

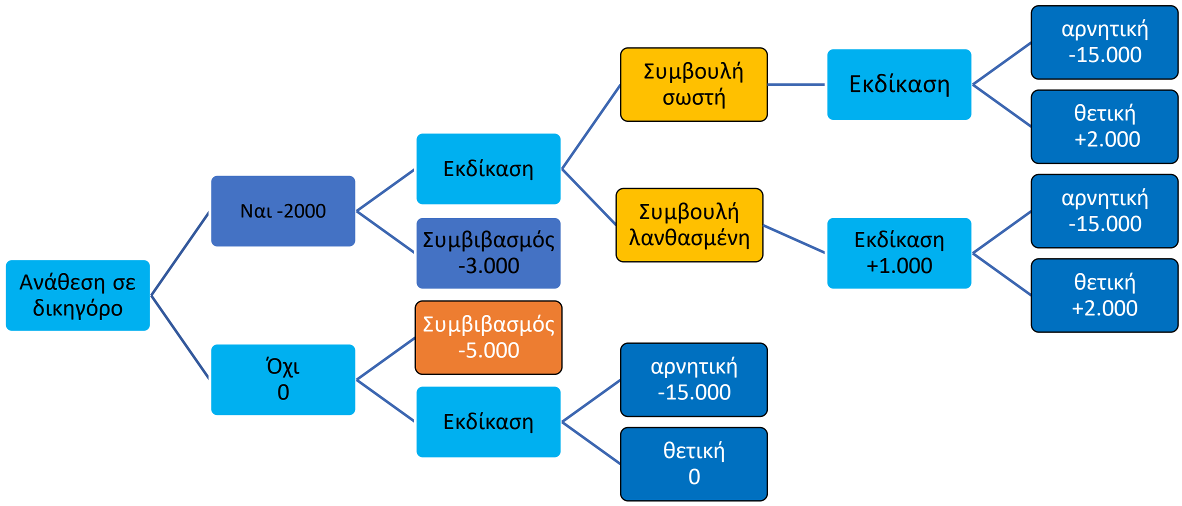
P(αρνητική |συμβουλή αρνητική) = 0,90  P(θετική |συμβουλή αρνητική) = 0,10

Από την αρχή δεν είναι γνωστή η έκβαση της εκδίκασης και δεν μπορούμε να ξέρουμε αν η συμβουλή του δικηγόρου είναι σωστή ή λανθασμένη. Η εκδίκαση οδηγεί σε θετική εκδίκαση με πιθανότητα 50% και σε αρνητική με πιθανότητα επίσης 50%. Με χρήση του θεωρήματος Bayes υπολογίζεται η συνολική πιθανότητα της σωστής συμβουλής του δικηγόρου και η πιθανότητα της λανθασμένης συμβουλής (Εικόνα 10.5) ως εξής:

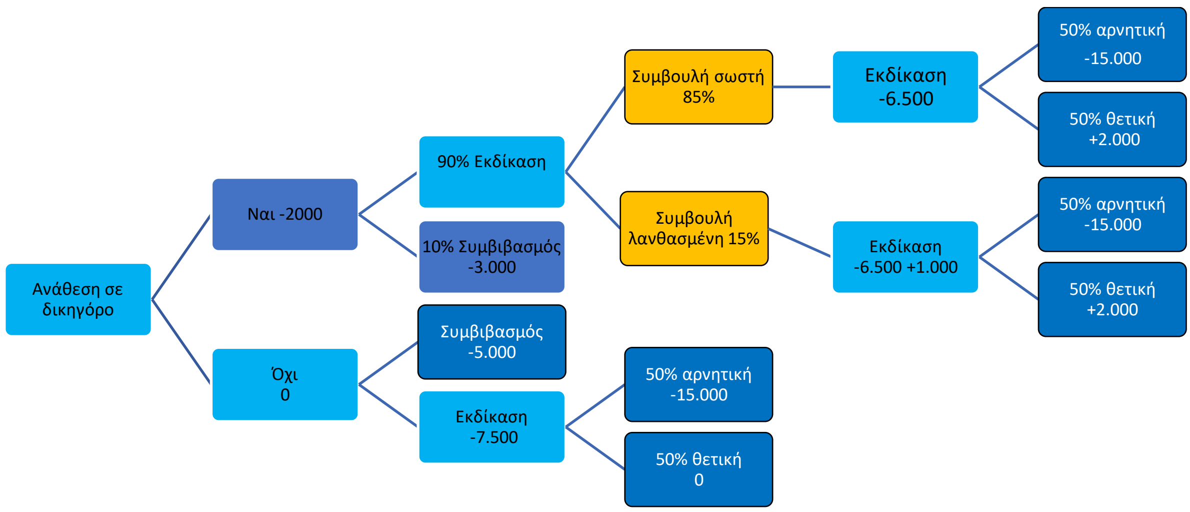
P(Συμβουλή σωστή) = P(θετική |συμβουλή θετική)*P(θετική) + P(αρνητική |συμβουλή αρνητική) *P(αρνητική) = 0,8*0,50+0,9*0,50 = 0,85

P(Συμβουλή λανθασμένη) = P(αρνητική | συμβουλή θετική)*P(αρνητική) + P(θετική | συμβουλή αρνητική)*P(θετική) = 0,20*0,50+0,1*0,50 = 0,15.

Αν ο δικηγόρος προτείνει συμβιβασμό, δεν παίρνει αμοιβή, οπότε θέτουμε το κόστος του συμβιβασμού στην περίπτωση αυτή να είναι -5.000+2.000 = -3.000, ώστε να συνεχίσουμε να έχουμε στον κόμβο απόφασης «ανάθεση σε δικηγόρο» την αμοιβή του 2.000, για τις πράξεις στους υπόλοιπους κόμβους.



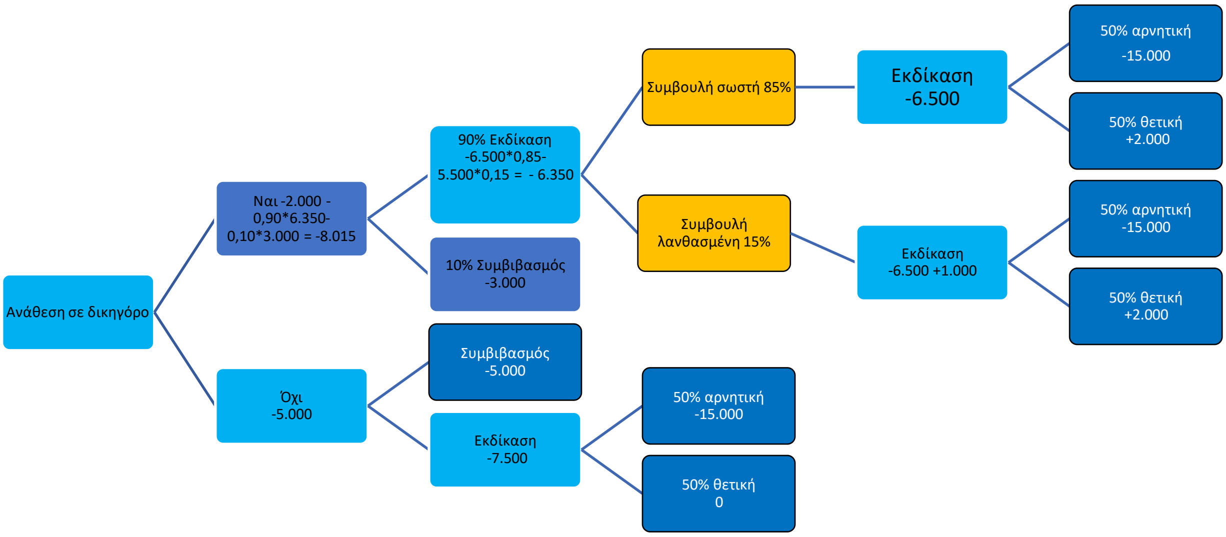

**Εικόνα 10.4** *Απεικόνιση εφαρμογής 12.3 με δένδρο αποφάσεων και όφελος επιλογών.*




**Εικόνα 10.5** *Απεικόνιση εφαρμογής 12.3 με δένδρο αποφάσεων και πιθανότητες εκβάσεων.*

Στην Εικόνα 10.5 προστίθενται στους αντίστοιχους κόμβους οι πιθανότητες των εκβάσεων και τα χρήματα που θα πληρώσει ο αντίδικος σε περίπτωση αρνητικής έκβασης αν έχουμε δικηγόρο, καθώς και τα χρήματα που θα επιστρέψει ο δικηγόρος για λανθασμένη συμβουλή του.

Στην Εικόνα 10.6, έχει υπολογιστεί και η αξία κάθε κόμβου. Στις εναλλακτικές εκβάσεις, η αξία υπολογίζεται με το άθροισμα γινομένων των πιθανοτήτων, ενώ στους κόμβους απόφασης, επιλέγεται η απόφαση με το μεγαλύτερο όφελος. Όπως μπορούμε να δούμε στην Εικόνα 10.6, η αξία του κόμβου «ανάθεση σε δικηγόρο» είναι -8.350 και η αξία του κόμβου «μη ανάθεση» είναι -5.000 (απευθείας συμβιβασμός). Βέλτιστη λύση, επομένως, είναι ο απευθείας συμβιβασμός και όχι η ανάθεση σε δικηγόρο.




**Εικόνα 10.6** *Απεικόνιση εφαρμογής 12.3 με αξία κόμβων.*

## 10.13 Ασκήσεις για κατανόηση - εφαρμογή

### 10.13.1 Επιλογή νέας ιστοσελίδας

Μια επιχείρηση σκοπεύει να αυξήσει την πελατεία της κάνοντας μια νέα ιστοσελίδα η οποία θα κοστίσει 700 ευρώ. Αυτή τη στιγμή η μελλοντική κατανομή της πελατείας προβλέπεται 100 άτομα σε πολύ καλές οικονομικές συνθήκες, 80 άτομα σε μέτριες οικονομικές συνθήκες και 60 άτομα σε άσχημες οικονομικές συνθήκες. Οι πιθανότητες για τις οικονομικές συνθήκες είναι 10%, 60% και 30% αντίστοιχα. Το όφελος από κάθε πελάτη είναι 50 ευρώ. Με τη νέα ιστοσελίδα το πλήθος των πελατών θα αυξηθεί κατά 20% για κάθε οικονομική συνθήκη. Τι θα επιλεγεί (υπάρχουσα κατάσταση ή νέα ιστοσελίδα); Αν οι δοσμένες πιθανότητες ήταν άγνωστες, τι θα επιλεγόταν;

**Πίνακας 10.25**. *Πίνακας αποφάσεων παραδείγματος 13.1.*

| **Ιστοσελίδα** | **Καλές** | **Μέτριες** | **Άσχημες** |
|---|---|---|---|
| πιθανότητες | 10% | 60% | 30% |
| ΟΧΙ | 5.000 | 4.000 | 3.000 |
| ΝΑΙ | 5.000*1,20-700= 5.300 | 4.000*1,20-700= 4.100 | 3.000*1,20-700= 2.900 |

### 10.13.2 Συμβιβασμός σε αγωγή

Ένας δικηγόρος αντιμετωπίζει μία αγωγή που έγινε σε πελάτη για παράβαση. Ο πελάτης μπορεί να συμβιβαστεί πληρώνοντας 5.000 ευρώ ή να δεχθεί να γίνει δίκη, με πιθανότητα 60% να τη χάσει και να πληρώσει 10.000 ευρώ ή να την κερδίσει και να μην πληρώσει τίποτα. Εναλλακτικά, μπορεί να αναθέσει στον δικηγόρο την εξέταση της παράβασης, και αφού εκείνος του δώσει τη συμβουλή του, να αποφασίσει αν θα συμβιβαστεί ή όχι. Η αμοιβή του δικηγόρου θα είναι 1.000 ευρώ και οι συμβουλές-εκτιμήσεις του δικηγόρου σε παρόμοιες περιπτώσεις που προέβλεπαν τη θετική έκβαση της δίκης ήταν σωστές κατά 80%, ενώ αυτές που προέβλεπαν την αρνητική έκβαση ήταν σωστές στο 90% των περιπτώσεων. Ο δικηγόρος στο 10% παρόμοιων περιπτώσεων προτείνει συμβιβασμό και επιστρέφει την αμοιβή του. Αν η συμβουλή του βγει τελικά λανθασμένη, επιστρέφει τη μισή αμοιβή του. Η εκδίκαση έχει δύο πιθανές εκβάσεις, να κερδηθεί (θετική) ή να μην κερδηθεί (αρνητική). Αν η εκδίκαση κερδηθεί, ο αντίδικος πληρώνει τα έξοδα του δικηγόρου. Να σχεδιαστεί το δένδρο αποφάσεων. Να υπολογιστούν οι πιθανότητες για κάθε κλάδο. Ποιο το αναμενόμενο ποσό που θα πληρωθεί σε κάθε περίπτωση;

### 10.13.3 Εντοπισμός ισορροπίας

Στον Πίνακα 10.26 απεικονίζονται τα στοιχεία ενός παιγνίου. Να εντοπιστεί αν υπάρχει ισορροπία και αν υπάρχουν υποδεέστερες στρατηγικές. Ποια είναι η λύση του παιγνίου;

**Πίνακας 10.26** *Παράδειγμα παιγνίου.*

| **Κέρδη για Α, ανάλογα με τις στρατηγικές του Α και Β** | **Β1 διαφήμιση με μπρελόκ** | **Β2 έκπτωση** | **Β3 δωρεάν συντήρηση** | **min** |
|---|---|---|---|---|
| Α1 συνεργασία με ηλεκτρολόγο | 3 | 2 | 4 | |
| Α2 διαφημιστικά φυλλάδια | 5 | 1 | 0 | |
| Α3 έκπτωση σε συχνούς πελάτες | 3 | 0 | 5 | |
| Max | | | | |

**Πίνακας 10.27** *Παράδειγμα εντοπισμού ισορροπίας παιγνίου.*

| **Κέρδη για Α** | **Β1** | **Β2** | **Β3** | **min** |
|---|---|---|---|---|
| Α1 | 2 | 4 | 2 | |
| Α2 | 5 | 3 | 7 | |
| Α3 | 1 | 2 | 3 | |
| Max | | | | |

### 10.13.4 Εντοπισμός ισορροπίας

Στον Πίνακα 10.27 απεικονίζονται τα στοιχεία ενός παιγνίου. Να εντοπιστεί αν υπάρχει ισορροπία και αν υπάρχουν υποδεέστερες στρατηγικές. Ποια είναι η λύση του παιγνίου;

### 10.13.5 Εντοπισμός μεικτής στρατηγικής για πίνακα παιγνίου 2 × 2

Έστω σαν παράδειγμα παιγνίου δύο επιχειρήσεις Α και Β που παράγουν το ίδιο προϊόν και προσπαθούν να κερδίσουν μερίδιο πελατών. Οι ιδιοκτήτες δρουν ανταγωνιστικά μεταξύ τους με στόχο την προσέλκυση πελατών, εφαρμόζοντας και οι δύο στρατηγική υψηλής τιμής και χαμηλής τιμής. Ο πίνακας κερδών του Α παρουσιάζεται στον Πίνακα 10.28.

**Πίνακας 10.28** *Επιχειρήσεις Α και Β.*

| **Κέρδη για Α, ανάλογα με τις στρατηγικές του Α και Β** | **Β1 υψηλή τιμή** | **Β2 χαμηλή τιμή** |
|---|---|---|
| Α1 υψηλή τιμή | 2.000 | -400 |
| Α2 χαμηλή τιμή | 2.500 | 1.300 |

# Βιβλιογραφία/Αναφορές

Δινοπούλου, Β., & Χιωτίδης, Γ. (2007). *Εισαγωγή στην επιχειρησιακή Έρευνα Γρ. Προγραμματισμός και θεωρία Αποφάσεων*. Αθήνα: Γκιούρδας.

Μάγειρος, Ε. (2012). *Παίγνια και Αποφάσεις Μια εισαγωγική προσέγγιση*. Αθήνα: Κριτική.

# Κριτήρια αξιολόγησης

## Κριτήριο αξιολόγησης 1

**Πότε μία απόφαση λαμβάνεται σε συνθήκες κινδύνου;**

**Απάντηση**
Όταν δεν είναι γνωστό ποια από τις πιθανές εκβάσεις θα συμβεί αλλά η πιθανότητα της κάθε έκβασης έχει εκτιμηθεί, τότε η απόφαση λαμβάνεται σε **συνθήκες κινδύνου.**

## Κριτήριο αξιολόγησης 2

**Πότε μία απόφαση λαμβάνεται σε συνθήκες αβεβαιότητας;**

**Απάντηση**
Όταν δεν είναι γνωστό ποια από τις εκβάσεις θα συμβεί, ούτε είναι γνωστή η πιθανότητα κάθε έκβασης, η απόφαση λαμβάνεται σε **συνθήκες αβεβαιότητας**.

## Κριτήριο αξιολόγησης 3

**Σε τι διαφέρει το κριτήριο maxmin από το κριτήριο maxmax;**

**Απάντηση**
Το κριτήριο maxmin επιλέγει το μέγιστο από τα χειρότερα αποτελέσματα, ενώ το κριτήριο maxmax επιλέγει το μέγιστο από τα μέγιστα αποτελέσματα. Το δεύτερο ονομάζεται και κριτήριο της αισιοδοξίας.

## Κριτήριο αξιολόγησης 4

**Πώς υπολογίζεται το αναμενόμενο αποτέλεσμα κάθε έκβασης ενός δένδρου αποφάσεων;**

**Απάντηση**
Σε κάθε κόμβο πιθανών εκβάσεων ενός δένδρου αποφάσεων υπολογίζεται το αναμενόμενο αποτέλεσμα, πολλαπλασιάζοντας την πιθανότητα κάθε έκβασης με το αντίστοιχο αποτέλεσμα της έκβασης και προσθέτοντας όλα τα επιμέρους γινόμενα.

## Κριτήριο αξιολόγησης 5

**Σε ένα παίγνιο ποια στρατηγική ονομάζεται υποδεέστερη;**

**Απάντηση**
Η στρατηγική την οποία ένας παίκτης δεν πρόκειται να επιλέξει ποτέ, γιατί το αποτέλεσμά της είναι πάντα μικρότερο από κάποια άλλη στρατηγική, ονομάζεται υποδεέστερη.

Στο παρόν σύγγραμμα γίνεται προσπάθεια να παρουσιαστούν οι κυριότερες μέθοδοι της Στατιστικής, οι οποίες εφαρμόζονται σε όλες σχεδόν τις άλλες επιστήμες. Παρουσιάζεται με απλό τρόπο η θεωρία, καθώς και παραδείγματα εφαρμογής της. Ειδικότερα, στο πρώτο μέρος παρουσιάζονται η μεθοδολογία και η δεοντολογία έρευνας. Στη συνέχεια, περιγράφονται οι τρόποι συλλογής δεδομένων μέσω ερωτηματολογίων και η περιγραφική παρουσίαση των δεδομένων, η εύρεση διαστημάτων εμπιστοσύνης για τις παραμέτρους τους και οι έλεγχοι υποθέσεων για τις πραγματικές τιμές των παραμέτρων. Επίσης, δίνονται κατευθύνσεις για την εξαγωγή συμπερασμάτων για ποσοτικά ή ποιοτικά δεδομένα. Η έμφαση δίνεται στους τρόπους στατιστικής ανάλυσης ανάλογα με το είδος δεδομένων, μέσω του στατιστικού λογισμικού SPSS, και στην ερμηνεία των αποτελεσμάτων του λογισμικού για την εξαγωγή αξιόπιστων συμπερασμάτων. Αναλύονται ιδιαίτερα οι τρόποι μελέτης ποσοτικών μεταβλητών, με τον υπολογισμό συσχετίσεων και την εκτίμηση μοντέλων παλινδρόμησης. Στην περίπτωση ύπαρξης μεγάλου πλήθους δεδομένων που περιγράφονται από πολλές ποιοτικές μεταβλητές, παρουσιάζεται η εφαρμογή της μεθόδου της παραγοντικής ανάλυσης μέσω εφαρμογών σε πραγματικά δεδομένα, και ακολούθως σχολιάζονται τα αποτελέσματα της μεθόδου και προτείνονται τρόποι ερμηνείας των παραγοντικών αξόνων. Στο δεύτερο μέρος παρουσιάζονται οι έννοιες της διαχρονικής αξίας χρήματος, της προεξόφλησης και του ανατοκισμού ραντών και δανείων. Επιπλέον, αναφέρονται οι τρόποι αξιολόγησης επενδύσεων, ερμηνεύεται η χρήση πιθανοτήτων στη διαδικασίας λήψη αποφάσεων και περιγράφονται η θεωρία παιγνίων, η έννοια των στρατηγικών και τρόποι επίλυσης απλών παιγνίων. Καταβλήθηκε προσπάθεια ώστε το παρόν σύγγραμμα να είναι απλό και κατανοητό, χωρίς να απαιτεί ιδιαίτερες μαθηματικές γνώσεις, προκειμένου να βοηθήσει τους αναγνώστες να κατανοήσουν την εφαρμογή των στατιστικών μεθόδων σε διάφορες επιστημονικές περιοχές.


| Το παρόν σύγγραμμα δημιουργήθηκε στο πλαίσιο του Έργου ΚΑΛΛΙΠΟΣ+ | |
|---|---|
| **Χρηματοδότης** | Υπουργείο Παιδείας και Θρησκευμάτων, Προγράμματα ΠΔΕ, ΕΠΑ 2020-2025 |
| **Φορέας υλοποίησης** | ΕΛΚΕ ΕΜΠ |
| **Φορέας λειτουργίας** | ΣΕΑΒ/Παράρτημα ΕΜΠ/Μονάδα Εκδόσεων |
| **Διάρκεια 2ης Φάσης** | 2020-2023 |
| **Σκοπός** | Η δημιουργία ακαδημαϊκών ψηφιακών συγγραμμάτων ανοικτής πρόσβασης (περισσότερων από 700)<br>• Προπτυχιακών και μεταπτυχιακών εγχειριδίων<br>• Μονογραφιών<br>• Μεταφράσεων ανοικτών textbooks<br>• Βιβλιογραφικών Οδηγών |
| **Επιστημονικά Υπεύθυνος** | Νικόλαος Μήτρου, Καθηγητής ΣΗΜΜΥ ΕΜΠ |
| **ISBN:** 978-618-228-000-3 | **DOI:** http://dx.doi.org/10.57713/kallipos-230 |

Το παρόν σύγγραμμα χρηματοδοτήθηκε από το Πρόγραμμα Δημοσίων Επενδύσεων του Υπουργείου Παιδείας