
ΠΕΡΙΕΧΟΜΕΝΑ

Πρόλογος	13
Κεφάλαιο 1. Εισαγωγή	17
1.1. Επιστημονική μέθοδος και στατιστική	18
1.2. Πληθυσμός και δείγμα	24
1.2.1. Επιλογή δειγμάτων	25
Ένθετο 1.1. Η ετυμολογία της λέξης «στατιστική»	29
1.3. Παράμετρος και στατιστικό στοιχείο	32
Ένθετο 1.2. Πόσο μεγάλο πρέπει να είναι το δείγμα;	34
Περίληψη	35
Ασκήσεις	35
Κεφάλαιο 2. Οργάνωση και παρουσίαση δεδομένων	39
2.1. Πίνακες συχνοτήτων	39
2.2. Μεταβλητή, τιμή, μέτρηση	41
2.2.1. Είδη μεταβλητών	42
2.2.2. Επίπεδα μέτρησης	43
2.3. Τα τρία βήματα για την κατασκευή του πίνακα συχνοτήτων	45
2.3.1. Ομαδοποιημένοι πίνακες συχνοτήτων	49
2.4. Απόλυτη και σχετική συχνότητα	52
2.5. Αθροιστικές συχνότητες	53
2.6. Γραφικές παραστάσεις	54
2.6.1. Ποιοτικές μεταβλητές	55
2.6.2. Ποσοτικές μεταβλητές	58
Ένθετο 2.1. Θεία τύχη(!)	62
2.7. Παραδείγματα «κακών» διαγραμμάτων	64
2.7.1. «Παραφορτωμένα» διαγράμματα	65
2.7.2. «Παραμορφωμένα» διαγράμματα	66

2.8. Μορφολογία κατανομών	70
Περίληψη	74
Ασκήσεις	75
Κεφάλαιο 3. Δείκτες κεντρικής τάσης	77
3.1. Επικρατούσα ή δεσπόζουσα τιμή	77
3.2. Διάμεσος	78
3.3. Αριθμητικός μέσος όρος	79
3.4. Δείκτες κεντρικής τάσης σε ομαδοποιημένη κατανομή	83
3.5. Ποιος δείκτης κεντρικής τάσης είναι καταλληλότερος;	86
Ένθετο 3.1. Το παράδοξο του Simpson	90
Περίληψη	91
Ασκήσεις	92
Κεφάλαιο 4. Δείκτες διασποράς	95
4.1. Εύρος	96
4.2. Ενδοτεταρτημοριακό εύρος	97
4.2.1. Το box-plot	98
4.3. Διακύμανση και τυπική απόκλιση	104
4.3.1. Συντελεστής μεταβλητότητας	112
Ένθετο 4.1. «Δεδομένα! Δεδομένα! Δεδομένα!	
Δεν μπορώ να κάνω τούβλα χωρίς πηλό»	114
Περίληψη	114
Ασκήσεις	115
Κεφάλαιο 5. Μετασχηματισμοί μεταβλητών και z-τιμές	119
5.1. Γραμμικοί μετασχηματισμοί μεταβλητών	119
5.2. Η z-τιμή – Η τυπική απόκλιση ως σπάντα μονάδα μέτρησης	124
Ένθετο 5.1. Στατιστική, ανέκδοτα και ψεύδη	131
Περίληψη	132
Ασκήσεις	132
Κεφάλαιο 6. Μοντέλα κατανομών – Η κανονική κατανομή	135
6.1. Διακριτά μοντέλα	136
6.2. Η καμπύλη πυκνότητας	139
6.3. Η κανονική κατανομή	142
6.3.1. Ιδιότητες της $N(\mu, \sigma^2)$	145
6.3.2. Υπολογισμός εμβαδού/ποσοστού κάτω από την κανονική καμπύλη $N(\mu, \sigma^2)$	149

6.3.3. Ακολουθούν τα δεδομένα μου την κανονική κατανομή;	157
6.4. Η σπουδαιότητα της κανονικής κατανομής – Το Κεντρικό Οριακό Θεώρημα	159
6.5. Κατανομές που απορρέουν από την κανονική	163
Ένθετο 6.1. Οι άνθρωποι πίσω από τα μοντέλα κατανομών	166
Περίληψη	169
Ασκήσεις	169
Κεφάλαιο 7. Σχέσεις μεταξύ δύο μεταβλητών	171
7.1. Συσχέτιση δύο ποσοτικών μεταβλητών	172
7.1.1. Συντελεστής συσχέτισης	178
7.1.2. Ιδιότητες και ερμηνεία συντελεστή συσχέτισης	188
7.1.3. Ευθεία παλινδρόμησης και πρόβλεψη	190
7.1.4. Ακραία σημεία και σημεία επιρροής	194
Ένθετο 7.1. <i>Sir Francis Galton, ο πατέρας της συσχέτισης</i>	200
7.2. Συσχέτιση δύο ποιοτικών μεταβλητών	201
7.2.1. Συσχέτιση μεταβλητών με τακτική κλίμακα μέτρησης	202
7.2.2. Συσχέτιση μεταβλητών με ονομαστική κλίμακα μέτρησης	206
Περίληψη	215
Ασκήσεις	216
Βιβλιογραφία	219
Παραρτήματα	221
Παράρτημα Α: Υπολογιστικοί τύποι	222
Παράρτημα Β: Πίνακας τυπικής κανονικής κατανομής	224
Παράρτημα Γ: Στατιστική στο Διαδίκτυο	226
Ευρετήριο όρων και θεμάτων	227

ΠΡΟΛΟΓΟΣ

Η *Εισαγωγή στη Στατιστική* αποτελεί πρόκληση τόσο για τον διδάσκοντα όσο και για τους φοιτητές. Η πρόκληση είναι ακόμα μεγαλύτερη στις σχολές θεωρητικών επιστημών, όπου ο διδάσκων καλείται να διασκεδάσει: α) τη φοβία των φοιτητών για τα μαθηματικά και β) τη διάχυτη εντύπωση ότι η στατιστική είναι ένα ανιαρό, στρυφνό ή ακόμα και άχρηστο μάθημα, που απαιτεί στέρεες μαθηματικές γνώσεις.

Η εντύπωση αυτή οφείλεται εν μέρει και στο γεγονός ότι τα πρώτα διδακτικά εγχειρίδια της στατιστικής (δεκαετίες '50 και '60) ήταν γραμμένα σε αυστηρό, απρόσιτο και ξερό ύφος, με περισσότερη έμφαση στις αποδείξεις και στις στείρες απομνημονεύσεις πολύπλοκων υπολογιστικών τύπων και με λιγότερη (ή καθόλου) έμφαση στην κατανόηση της έννοιας. Τίποτα δεν είναι πιο άχαρο και απωθητικό στα μαθηματικά (ακόμα και για τους ίδιους τους μαθηματικούς) από τους υπολογισμούς για χάρη των υπολογισμών. Το αντιπαιδαγωγικό βιβλίο, σε συνδυασμό με τη φορμαλιστική διδασκαλία και την τυπολαγνεία, αποτελούσε τραυματική εμπειρία για τους περισσότερους φοιτητές, τους αποξένωνε από το μάθημα και αναπόφευκτα τους προκαλούσε τη διά βίου απέχθεια της στατιστικής. Οι φοιτητές δεν κατανοούσαν πώς, πότε και γιατί εφαρμόζονται οι παράξενοι τύποι με τα μυστηριώδη σύμβολα, και πώς χρησιμοποιούνται οι μέθοδοί της για την επίλυση πραγματικών προβλημάτων στο δικό τους επιστημονικό πεδίο.

Η αποτυχία της επίτευξης του στόχου του μαθήματος έχει αλυσιδωτές συνέπειες: οι φοιτητές ξεχνούν σε σύντομο χρονικό διάστημα όσα απομνημονεύουν μηχανικά για να περάσουν τις εξετάσεις, δεν μπορούν να εφαρμόσουν στοιχειώδεις στατιστικές αναλύσεις στις πτυχιακές τους εργασίες, αδυνατούν να κατανοήσουν ή να αξιολογήσουν με κριτική ματιά τα αποτελέσματα δημοσιευμένων πειραματικών ερευνών. Η αρνητική εμπειρία τους αποθαρρύνει να ακολουθήσουν μεταπτυχιακές σπουδές, αφού γνωρίζουν ότι αργά ή γρήγορα θα συναντήσουν μπροστά τους τη στατιστική.

Τη δεκαετία του '70 και κατά τη διάρκεια των τελευταίων 20 ετών τα εκσυγχρονιστικά ρεύματα στη διδακτική της στατιστικής πρότειναν αλλαγή τόσο του περιεχομένου όσο και του τρόπου διδασκαλίας: πολύ λιγότερη έμφαση στη μαθηματική διάσταση της στατιστικής και στους «απωθητικούς» υπολογισμούς και περισσότερη στις έννοιες-κλειδιά που αναπτύσσουν στατιστική σκέψη. Η προσέγγιση αυτή, που συνεχίζεται και βελτιώνεται διαρκώς, έχει αποφέρει ενθαρρυντικά αποτελέσματα. Οι φοιτητές των θεωρητικών σχολών προσεγγίζουν το μάθημα με θετικότερα συναισθήματα, συμμετέχουν και προσέχουν περισσότερο την ώρα της διδασκαλίας. Τα τελευταία χρόνια η χρήση εξειδικευμένων στατιστικών προγραμμάτων στον Η/Υ, που κρύβει όμως κάποιους κινδύνους αν δεν γίνει σωστά, απαλλάσσει από το «βάρος» των υπολογισμών φοιτητές και διδάσκοντες, καθιστώντας εφικτή την ανάλυση μεγάλου όγκου δεδομένων, και επιτρέπει την εστίασή τους στην ουσία της στατιστικής μεθόδου κάνοντας τις παρουσιάσεις πιο ενδιαφέρουσες και τις συζητήσεις πιο γόνιμες.

Η συνεχής βελτίωση της διδακτικής μεθοδολογίας με σκοπό την καλύτερη εκμάθηση των μαθηματικών και της στατιστικής, σε όλα τα επίπεδα της εκπαίδευσης, είναι επιβεβλημένη και αποτελεί αντικείμενο πολλών ερευνητικών μελετών. Εκπαιδευτικοί και γνωστικοί ψυχολόγοι ερευνούν, με διαφορετικές μεθόδους ο καθένας, πώς συντελείται και βελτιώνεται η μάθηση. Οι πρώτοι ερευνούν το θέμα κυρίως από την πρακτική και οι δεύτεροι κυρίως από τη θεωρητική σκοπιά. Οι Lovett και Greenhouse (2000) συνδύασαν τα συγκλίνοντα αποτελέσματα αυτών των δύο προσεγγίσεων – οι οποίες δυστυχώς, όπως αναφέρουν, δεν επικοινωνούν συχνά μεταξύ τους – και τα εφάρμοσαν στη διδασκαλία της στατιστικής προσφέροντας πρακτικές οδηγίες σε όσους διδάσκουν το αντικείμενο αυτό. Συνιστούν δυναμικές διδακτικές τεχνικές, στις οποίες οι φοιτητές δεν είναι απλοί θεατές, αλλά συμμετέχουν στη συλλογή και ανάλυση δεδομένων *μαθαίνοντας ενεργά* (active learning), συνεργαζόμενοι σε μικρές ομάδες εργασίας (collaborative learning). Η χρήση των νέων τεχνολογιών μπορεί να βοηθήσει, με ένα προσομοιωμένο περιβάλλον διερεύνησης και ανακάλυψης, στην κατανόηση δύσκολων εννοιών.

Το παρόν βιβλίο φιλοδοξεί να βοηθήσει τους φοιτητές να ξεπεράσουν τη φοβία τους και να τους αποδείξει ότι η στατιστική είναι μια δυναμική, ενδιαφέρουσα και συνεχώς αναπτυσσόμενη επιστήμη, ένα πολύτιμο εργαλείο στα χέρια κάθε επιστήμονα. Σκοπός του εγχειριδίου αυτού είναι να ανατρέψει την αρνητική εικόνα που πολλοί φοιτητές (και όχι μόνο) έχουν για τη στατιστική, ότι είναι ένα στρυφνό και αφηρημένο αντικείμενο, κατανοητό μόνο σε μαθηματικά μυαλά. Το σύγγραμμα αυτό δεν απευθύνεται

σε φοιτητές θετικών επιστημών που έχουν γερό μαθηματικό υπόβαθρο και απαιτήσεις για αυστηρότερη θεμελίωση των εννοιών και για αποδείξεις των διατυπωμένων θεωρημάτων.

Στο παρόν βιβλίο οι περισσότερες στατιστικές έννοιες δεν παρουσιάζονται αξιωματικά, αλλά ως προβλήματα που επιζητούν λύση. Έχει γίνει προσπάθεια να χρησιμοποιηθούν ενδιαφέροντα παραδείγματα με πραγματικά, ως επί το πλείστον, δεδομένα από επίκαιρα θέματα στον χώρο της εκπαίδευσης και της ψυχολογίας, τα οποία ελπίζουμε να κεντρίσουν το ενδιαφέρον και να ερεθίσουν την περιέργεια των φοιτητών. Παραδείγματα και ασκήσεις λαμβάνονται από την τρέχουσα επικαιρότητα, προκειμένου να καταδειχθεί η *πρακτική* χρησιμότητα της στατιστικής, αλλά και να στιγματίζεται η κακή χρήση της που, δυστυχώς, γίνεται συχνά. Χρησιμοποιούνται επίσης παραδείγματα με «φτιαχτά» δεδομένα για την αποσαφήνιση ορισμένων εννοιών. Οι τύποι δίνονται τις περισσότερες φορές και με λεκτικό και με συμβολικό τρόπο, προκειμένου οι φοιτητές να μάθουν να τους αποκωδικοποιούν, αλλά και για να εξοικειωθούν με τα σύμβολα. Οι υπολογιστικοί τύποι εξοικονομούν μεν χρόνο, αλλά δεν φανερώνουν τον τρόπο της γέννησης της ιδέας ή του ορισμού της έννοιας και, επειδή μπερδεύουν τους περισσότερους αναγνώστες, δεν παρουσιάζονται στα κεφάλαια, αλλά εντάσσονται στο παράρτημα Α, ακριβώς για να μην τους αποπροσανατολίσουν. Οι κυριότεροι στατιστικοί όροι δίνονται και στα αγγλικά, ώστε η χρήση των διάφορων στατιστικών πακέτων στον Η/Υ καθώς και η μελέτη αγγλικών συγγραμμάτων να καταστούν ευκολότερες. Σε κάθε κεφάλαιο υπάρχουν αυτόνομα σχετικά ένθετα, με ενδιαφέρουσες κυρίως ιστορικές πληροφορίες, τα οποία σκιαγραφούν το γενικότερο πλαίσιο στο οποίο οι διάφορες στατιστικές έννοιες αναπτύχθηκαν.

Ελπίζουμε ότι ο φοιτητής με «μη μαθηματικό» μυαλό θα βρει το βιβλίο ελκυστικό, ακόμα και αν συναντήσει κάποια δυσκολία στην κατανόηση ορισμένων από τις νέες έννοιες. Ο αναγνώστης φοιτητής καλείται να μάθει την τέχνη της στατιστικής σκέψης και της αναζήτησης αντιμετώπιζοντας το παρόν σύγγραμμα με κριτική διάθεση. Αν το επιτύχει, θα νιώσει τη χαρά που κρύβεται πίσω από κάθε ανακάλυψη, μικρή ή μεγάλη.

Τέλος, θα ήθελα να ευχαριστήσω τους καθηγητές του Πανεπιστημίου Κρήτης κυρίους Γεώργιο Κρασανάκη και Παναγιώτη Γεωργούση για τις εύστοχες παρατηρήσεις τους. Τυχόν σφάλματα ή παραλείψεις αυτού του τόμου βαρύνουν αποκλειστικά τον συγγραφέα.