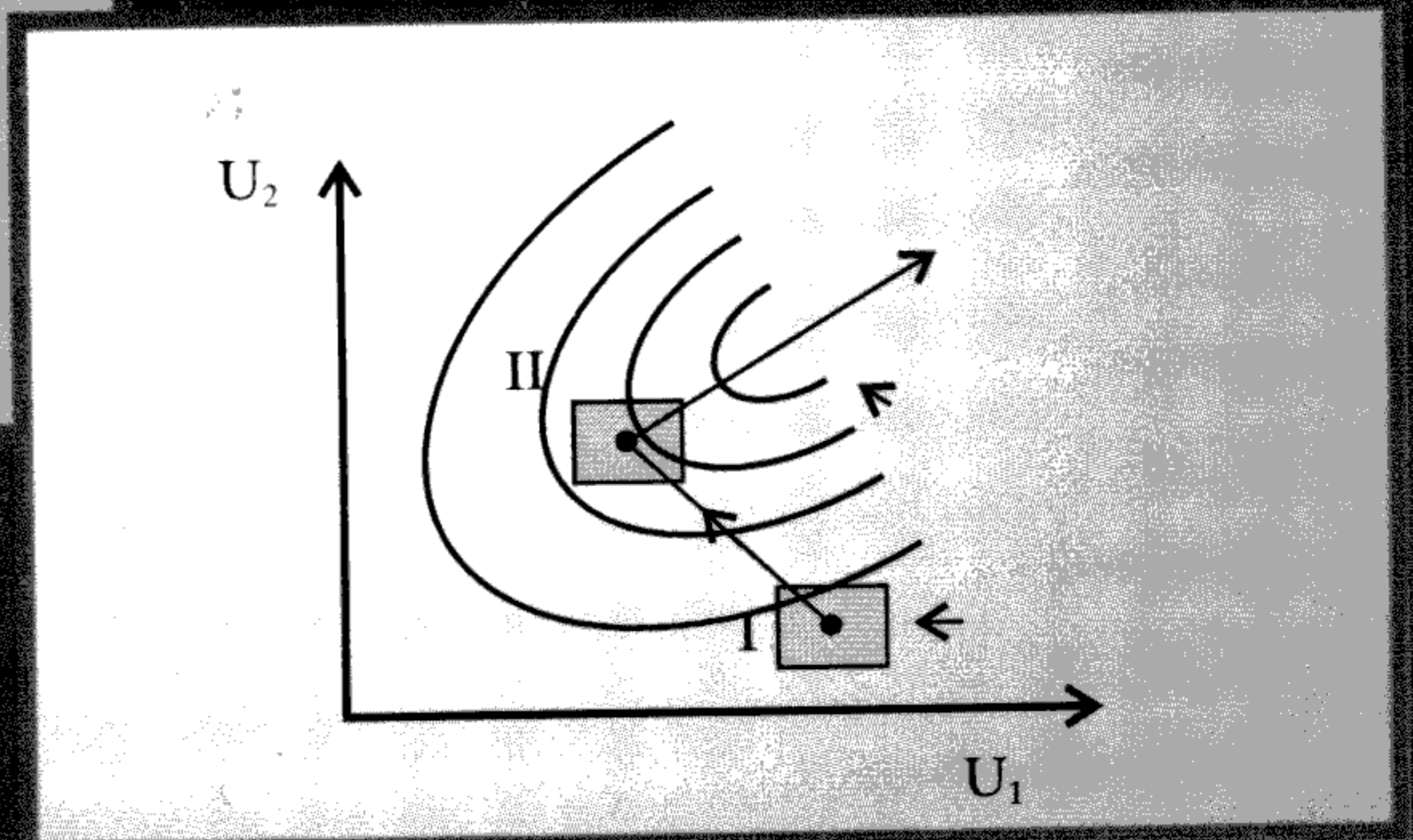
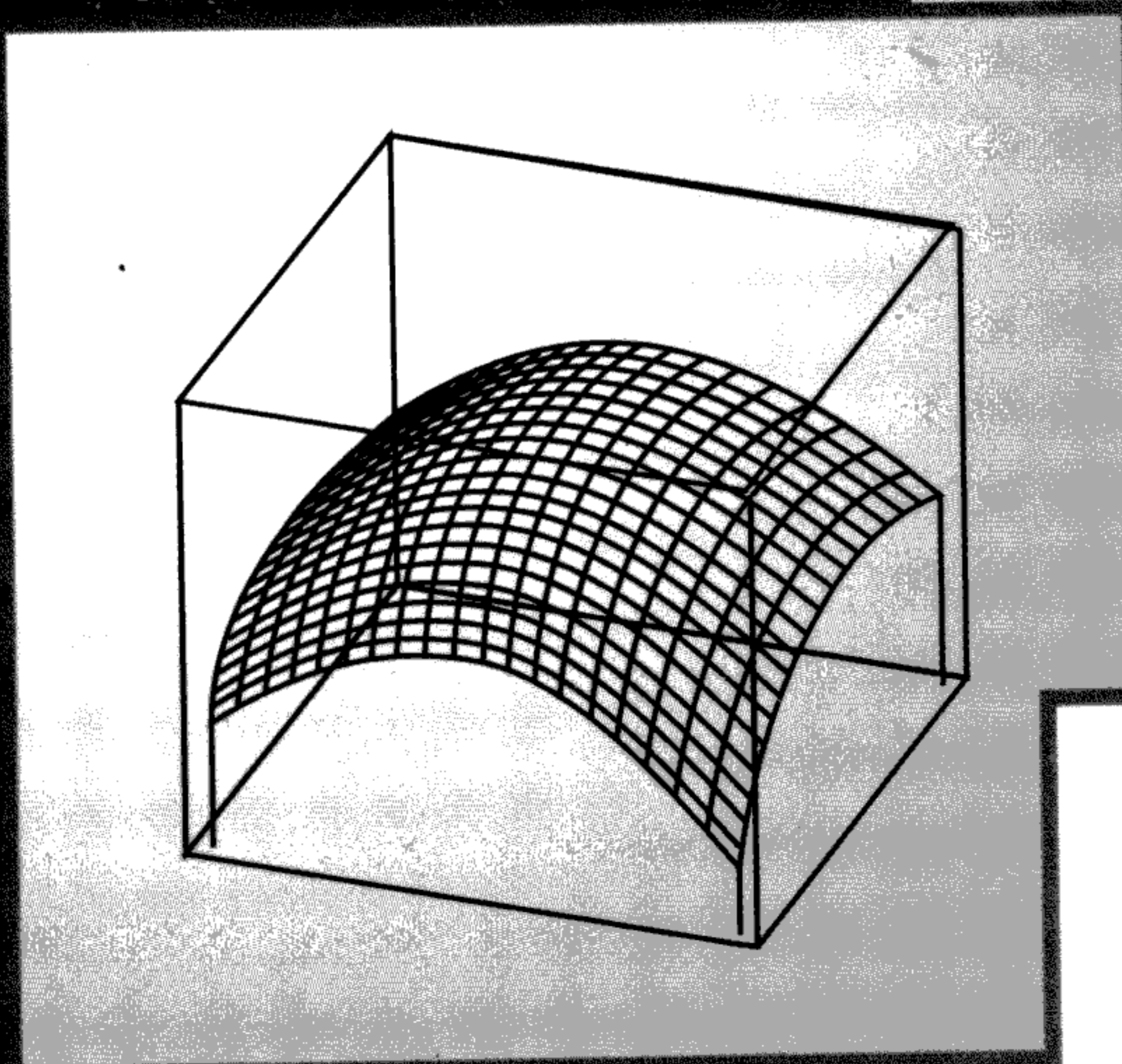
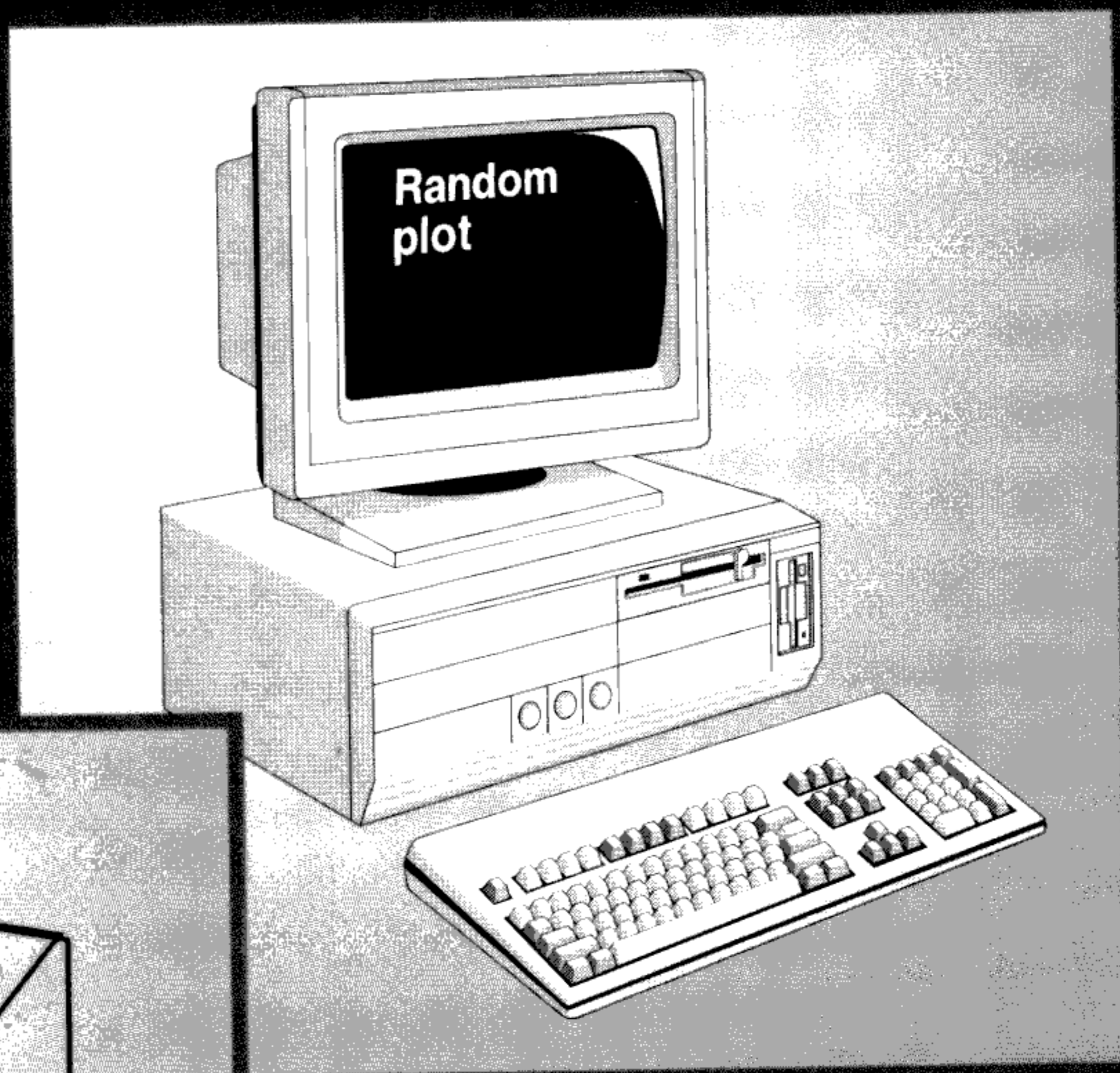


ΧΡΗΣΤΟΣ ΠΑΡ. ΚΙΤΣΟΣ
Στατιστικός ΜΑ, Ph.D.

ΥΠΟΛΟΓΙΣΤΙΚΗ ΣΤΑΤΙΣΤΙΚΗ



ΑΘΗΝΑ

Τ.Ε.Ι. ΑΘΗΝΑΣ
ΒΙΒΛΙΟΘΗΚΗ

Χ. Π. Κίτσος, Ph.D.,
Επίκουρος Καθηγητής
Οικονομικό Πανεπιστήμιο Αθηνών

Τ.Ε.Ι. ΑΘΗΝΑΣ
ΒΙΒΛΙΟΘΗΚΗ
Αρ. εισ. 63014

519 502 85
KIT

Έργο: Ενίσχυση και εμπλουτισμός
των Ακαδημαϊκών Βιβλιοθηκών
MIS 9 755

Η συλλογή της Βιβλ. αυτής συγχρηματοδοτείται απ. την Εργασία 41. Το 75% υποποσστό 75% και 25% από εθνικούς πόρους.

ΥΠΟΛΟΓΙΣΤΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Αθήνα

Κάθε γνήσιο αντίτυπο φέρει την υπογραφή του συγγραφέα.



ΕΚΔΟΣΕΙΣ ΝΕΩΝ ΤΕΧΝΟΛΟΓΙΩΝ

Στουρνάρη 49Α, 106 82, Αθήνα

Τηλ.: 38 45 594, Fax: 38 08 009

ISBN: 960-7235-72-X

Απαγορεύεται η με οποιονδήποτε τρόπο ανατύπωση, καταχώριση σε σύστημα αποθήκευσης και επανάκτησης, ή μετάδοση σε κάθε μορφή και μέσο (ηλεκτρονικό, μηχανικό, φωτοαντιγραφικό κ.λπ.) του συνόλου ή μέρους του βιβλίου αυτού χωρίς έγγραφη άδεια του συγγραφέα.

*Τι μας έμεινε Νικήτα
γύρνα πίσω σου και κοίτα
χιλιάδες χρόνια πάνω στον τροχό
ποιός θυμάται πες μου το φτωχό;*

N. Γκάτσος: Μελαγχολικό Εμβατήριο

ΠΕΡΙΕΧΟΜΕΝΑ

Περιεχόμενα	5
Πρόλογος	7
Εισαγωγή	9
1ο Κεφάλαιο: ΓΕΝΙΚΗ ΘΕΩΡΗΣΗ	11
2ο Κεφάλαιο: ΑΡΧΕΙΑ ΔΕΔΟΜΕΝΩΝ	53
3ο Κεφάλαιο: ΤΥΧΑΙΟΙ ΑΡΙΘΜΟΙ	91
4ο Κεφάλαιο: STATGRAPHICS	129
5ο Κεφάλαιο: ΘΕΩΡΙΑ ΣΦΑΛΜΑΤΩΝ	171
6ο Κεφάλαιο: ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΣΤΑΤΙΣΤΙΚΗ	219
7ο Κεφάλαιο: ΥΠΟΛΟΓΙΣΤΙΚΑ ΘΕΜΑΤΑ ΓΡΑΜΜΙΚΩΝ ΜΟΝΤΕΛΩΝ	269
8ο Κεφάλαιο: ΕΥΡΩΣΤΗ ΑΝΑΛΥΣΗ	309
9ο Κεφάλαιο: ΕΙΔΙΚΕΣ ΜΟΡΦΕΣ ΠΟΛΥΩΝΥΜΙΚΗΣ ΠΡΟΣΕΓΓΙΣΗΣ ΔΕΔΟΜΕΝΩΝ	321
10ο Κεφάλαιο: ΕΙΣΑΓΩΓΗ ΣΤΗ ΜΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ	339
Παράρτημα I: ΠΡΟΓΡΑΜΜΑΤΑ PASCAL	357
Παράρτημα II: ΜΕΓΙΣΤΑ-ΕΛΑΧΙΣΤΑ	369
Επίλογος	387
Βιβλιογραφία	389
Αναλυτικά Περιεχόμενα	395

ΠΡΟΛΟΓΟΣ

*Στην ερημιά πού χα βρεθεί
με τό 'να χέρι στο σπαθί
και τ' άλλο στο βαγγέλιο.*

N. Γκάτσος: Ένας ευαίσθητος ληστής

Ηταν το 1976 στο Πανεπιστήμιο του New Brunswick (UNB) στον Καναδά, όταν, ως πτυχιούχος Μαθηματικός, πρωτάκουσα τον όρο Υπολογιστική Στατιστική (Computational Statistics).

Αργότερα, το 1986, άρχισα εκ νέου μια ερευνητική προσπάθεια, στο Glasgow University του Ηνωμένου Βασιλείου, που, πιστεύω, συνεχίζεται σε παρεμφερή θέματα: στην ερημιά της έρευνας με τό 'να χέρι στη Στατιστική και τ' άλλο στον Υπολογιστή.

Δέκα χρόνια αργότερα (και είκοσι από την αρχική προσπάθεια) μου δίνεται η ευκαιρία να παρουσιάσω αυτό το βιβλίο. Είναι προϊόν συγκομιδής διαφόρων θεμάτων που για κύριο άξονα ανάπτυξης της ύλης, ελήφθη το περιεχόμενο των συνεδρίων COMPSTAT (COMPUTATIONAL STATISTICS), που άρχισαν το 1974 και πραγματοποιούνται έκτοτε κάθε δύο χρόνια. Η Στατιστική συνδέεται με διάφορες επιστήμες και η σύνδεσή της με την Αριθμητική Ανάλυση είναι η βάση της Υπολογιστικής Στατιστικής.

Στην τελική εμφάνιση βοήθησαν πολλοί φοιτητές του τμήματος Στατιστικής του Οικονομικού Πανεπιστημίου Αθηνών, που καθοδηγήθησαν στην επίλυση των παραδειγμάτων. Αισθάνομαι την υπο-

χρέωση να ευχαριστήσω ιδιαίτερος τους ήδη πτυχιούχους Στατιστικούς, κ.κ. Φ. Σταυρόπουλο, Ν. Παξιμαδάκη, Μ. Περράκη και τον φοιτητή κ. Γ. Αδραχτά. Την χημικό Λέκτορα του Πανεπιστημίου κα Β. Νινή θερμά ευχαριστώ που διήλθε τα παραδείγματα Χημείας, καθώς και τον πληροφορικό κ. Χρήστο Κοΐλια, καθηγητή ΤΕΙ, για τις χρήσιμες συζητήσεις σε θέματα προγραμματισμού.

Ευελπιστώ ότι όλη αυτή η προσπάθεια θα βρει ανταπόκριση από εκείνους που θέλουν να ασχοληθούν με εφαρμοσμένα θέματα Στατιστικής.

Χ. Π. Κ.

ΕΙΣΑΓΩΓΗ

*Με τα λόγια παιδιά τίποτε δε γίνεται.
Τί να κάνουμε; Μας χρειάζεται πομονή.*

Ψυχάρης: Το ταξίδι μου

Ο Στατιστικός με δίκαιη θλίψη βλέπει ότι πλείστοι όσοι οικειοποιούνται τα πεδία έρευνας του και τους χώρους εργασίας του. Οι πάντες γεγόνασιν Στατιστικοί πλην των αληθών Στατιστικών, για να παραφράσουμε μια έκφραση της Βίβλου. Ολοι οι χρήστες των διάφορων στατιστικών πακέτων αισθάνονται, πιστεύω, την ανάγκη του κατάλληλου υπόβαθρου, που δεν μπορεί πάντα να καλύψει η εντολή HELP. Ο Στατιστικός όμως δεν πρέπει να γίνει ένας απλός χρήστης. Πρέπει να γνωρίζει αφενός την ανάπτυξη της μεθοδολογίας, που οδηγεί σε θαυμαστό χρόνο στις μέρες μας, στα αποτελέσματα της στατιστικής που θέλει να υπολογίσει και αφετέρου πρέπει να αντιλαμβάνεται την ευαισθησία που εγκλείεται στο αριθμητικό αποτέλεσμα.

Στη δεκαετία του '70 αναπτύχθησαν διάφοροι αλγόριθμοι για την επίτευξη λύσης των διαφόρων προβλημάτων της Στατιστικής. Κυρίως το ενδιαφέρον εστιάστηκε στα θέματα εκείνα που ελύοντο με υπολογιστικό κόπο, π.χ. όλα εκείνα στα οποία η εύρεση αντιστρόφου πίνακα είναι απαραίτητη. Αναγκαστικά λοιπόν δανείσθησαν αλγόριθμους από την Αριθμητική Ανάλυση, π.χ. LU-ανάλυση. Ο συνδυασμός αυτός, οικειποίηση μεθόδων, αλγορίθμων της Αριθμητικής Ανάλυσης από την Στατιστική και ανάπτυξη άλλων αλγορίθμων απέβη επωφελής, αφού στο μεταξύ ανεπτύσσοντο και οι Η/Υ. Έτσι άρχισε μια σειρά συνεδρίων με

το όνομα COMSTAT (COMPUtational STATistics), όπου κύριος στόχος είναι η σύνδεση της Στατιστικής με τους όμορους κλάδους της επιστήμης των Η/Υ. Ο σπόρος που έρριξε ο Fisher στην Αγγλία συνδυαζόταν με εκείνον του Zuse στην Γερμανία. Και αν οι πειραματικοί σχεδιασμοί, από τα πρώτα θέματα της Υπολογιστικής Στατιστικής, βρήκαν γόνιμο έδαφος ανάπτυξης στο Ηνωμένο Βασίλειο, στην Ιαπωνία, ήταν όπου βρήκαν πλήρη εφαρμογή στον ποιοτικό έλεγχο και στην βιομηχανική παραγωγή γενικώτερα.

Το βιβλίο αυτό προσπαθεί να καλύψει πολλά από τα θέματα της Υπολογιστικής Στατιστικής σε εισαγωγικό επίπεδο. Παραθέτει πλήθος παραδειγμάτων στην αντίστοιχη θεωρία και επιδιώκεται να είναι πλήρως λυμένα με κάποιο στατιστικό πακέτο (Minitab, Statgraphics ή σπανιότερα SPSS). Ορισμένα θέματα, π.χ. εύρωστη ανάλυση, παρουσιάζουν ιδιομορφία στο λογισμικό, ενώ τα μη γραμμικά μοντέλα έχουν ισχυρότατο μαθηματικό υπόβαθρο. Επί πλέον και οι δύο κλάδοι αποτελούν ανεξάρτητους τομείς έρευνας όπως, π.χ. τα γραμμικά μοντέλα. Όμως στο παρόν αναφέρεται το τμήμα εκείνο που είναι πλησιέστερα στην υπολογιστική ανάπτυξη, στην Αρχιμήδεια διάσταση της έρευνας, παρά στη θεωρητική κάλυψη, στην Ευκλείδεια διάσταση της έρευνας.

Πιστεύεται ότι ο αναγνώστης θα αντιληφθεί τη χρησιμότητα του παρόντος, όταν θελήσει να επιλύσει προβλήματα στην πράξη. Γιατί άλλωστε η ανάγκη της επίλυσης τέτοιων πρακτικών προβλημάτων από τη βιομηχανία κυρίως, έδωσε ώθηση στον κλάδο αυτό της εφαρμοσμένης Στατιστικής.

Με τα λόγια λοιπόν τίποτε δεν γίνεται. Χρειάζεται υπομονή στην εφαρμογή και επιμονή να επιτευχθεί το τελικό αποτέλεσμα σε μια εφαρμογή. Και η Υπολογιστική Στατιστική θα σταθεί αρωγός!

ΑΝΑΛΥΤΙΚΑ ΠΕΡΙΕΧΟΜΕΝΑ

Περιεχόμενα.....	5
Πρόλογος	7
Εισαγωγή	9
1ο Κεφάλαιο: ΓΕΝΙΚΗ ΘΕΩΡΗΣΗ	11
1.1. Εισαγωγή.....	11
1.2. Ιστορική Αναδρομή	13
1.3. Γλώσσες Προγραμματισμού	25
1.4. Συζήτηση	27
1.5. Στοιχειώδη επί των λειτουργικών συστημάτων	34
1.6. Πακέτα.....	37
1.7. Στατιστικά πακέτα	38
1.8. Σχόλια	43
1ο - ΑΣΚΗΣΕΙΣ	44
1ο - ΠΑΡΑΡΤΗΜΑ: εντολές MINITAB	44
2ο Κεφάλαιο: ΑΡΧΕΙΑ ΔΕΔΟΜΕΝΩΝ.....	53
2.1. Εισαγωγή.....	53
2.2. Γενικά περί αρχείων.....	54
2.3. Γραμμογράφηση και οργάνωση	58
2.4. Επεξεργασία αρχείων	61
2.5. Μέγεθος αρχείου δεδομένων	63
2.6. Δημιουργία αρχείων στο Minitab	65
2.7. Προγράμματα στο Minitab	76
2.8. Ανάγνωση αρχείου στο SPSS	82
2ο - ΑΣΚΗΣΕΙΣ	86

2ο - ΠΑΡΑΡΤΗΜΑ	87
I) Διαχείριση αρχείων στο DOS.....	87
II) SPSS for Windows: ελάχιστα βασικά	88
3ο Κεφάλαιο: ΤΥΧΑΙΟΙ ΑΡΙΘΜΟΙ.....	91
3.1. Εισαγωγή.....	91
3.2. Βασικές έννοιες.....	92
3.3. Ειδικές περιπτώσεις.....	95
3.3.1. Εκθετική κατανομή.....	95
3.3.2. Κανονική κατανομή.....	95
3.3.3. Διδιάστατη κανονική.....	96
3.3.4. Άλλες μεθοδολογίες για την κανονική.....	98
3.4. Δημιουργία τυχαίων αριθμών στην πράξη.....	103
3.5. Χρήση των τυχαίων αριθμών.....	113
3.6. Βηματική προσομοίωση (Bootstrap).....	117
3.7. Αριθμοί Fibonacci.....	119
3.8. Εφαρμογές αριθμών Fibonacci.....	122
3.9. Γεννήτορες τυχαίων αριθμών.....	124
3ο - ΑΣΚΗΣΕΙΣ.....	125
4ο Κεφάλαιο: STATGRAPHICS	129
4.1. Εισαγωγή.....	129
4.2. Γενική περιγραφή.....	130
4.3. Οδηγίες για την εισαγωγή δεδομένων.....	149
4.4. Ειδικά θέματα.....	151
4.5. Περιληπτική σύγκριση Minitab και Statgraphics.....	155
4ο - ΑΣΚΗΣΕΙΣ.....	167
5ο Κεφάλαιο: ΘΕΩΡΙΑ ΣΦΑΛΜΑΤΩΝ	171
5.1. Εισαγωγή.....	171
5.2. Τα σύμβολα σ και Ω	172
5.3. Διάδοση σφαλμάτων.....	174
5.4. Οικογένεια μετασχηματιστών.....	177
5.5. Συντελεστής ασταθείας.....	178
5.6. Χρήση συντελεστών ασταθείας-πολυσυγγραμμικότητα.....	181
5.7. Ετερο(ομο)σκεδαστικότητα.....	185

5.8. Συντελεστές ασταθείας για συναρτήσεις	187
5.9. Διαταραχή σε Ax και β στο $Ax=\beta$	189
5ο - ΑΣΚΗΣΕΙΣ	212
5ο - ΠΑΡΑΡΤΗΜΑ	214
6ο Κεφάλαιο: ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΣΤΑΤΙΣΤΙΚΗ	219
6.1. Εισαγωγή	219
6.2. Βασικές Έννοιες	222
6.3. Διάστημα εμπιστοσύνης για το $\hat{\chi}_0$	233
6.4. Μέθοδος βαθέος βήματος	236
6.5. Κανονική παρουσίαση δευτεροβάθμιου μοντέλου	248
6.6. Χρήση Πολλαπλασιαστών Lagrange	261
6ο - ΑΣΚΗΣΕΙΣ	264
6ο - ΠΑΡΑΡΤΗΜΑ	266
7ο Κεφάλαιο: ΥΠΟΛΟΓΙΣΤΙΚΑ ΘΕΜΑΤΑ ΓΡΑΜΜΙΚΩΝ	
 ΜΟΝΤΕΛΩΝ	269
7.1. Εισαγωγή	269
7.2. Τα βασικά για το γενικό γραμμικό μοντέλο	270
7.3. Υπολογιστικοί μέθοδοι επιλογής του 'βέλτιστου' γενικού γραμμικού μοντέλου	275
7.4. Υπολογιστικά κριτήρια επιλογής του 'βέλτιστου' γραμμικού μοντέλου	279
7.5. Υποστήριξη με Minitab και SPSS for Windows	282
7.5.1. Επίλυση του παραδείγματος με το Minitab	283
7.5.2. Επίλυση του παραδείγματος με το SPSS for Windows	286
7.6. Τριγωνομετρική προσαρμογή	295
7.7. Γενικευμένα ελάχιστα τετράγωνα (γετ)	300
7.8. Αυτοσυσχέτιση	301
7.9. Ο έλεγχος Durbin-Watson	302
7ο - ΑΣΚΗΣΕΙΣ	304
7ο - ΠΑΡΑΡΤΗΜΑ: Μετασχηματισμός <i>Householder</i>	306
8ο Κεφάλαιο: ΕΥΡΩΣΤΗ ΑΝΑΛΥΣΗ	309
8.1. Εισαγωγή	309
8.2. Εύρωστες παρατηρήσεις	310
8.3. Εύρωστοι εκτιμητές	312

8.4. Άλλες προσεγγίσεις	315
8.5. Λογισμικό	317
8.6. Ασθενείς υποθέσεις για τα δεδομένα.....	318
8ο - ΑΣΚΗΣΕΙΣ	320

9ο Κεφάλαιο: ΕΙΔΙΚΕΣ ΜΟΡΦΕΣ ΠΟΛΥΩΝΥΜΙΚΗΣ ΠΡΟΣΕΓΓΙΣΗΣ ΔΕΔΟΜΕΝΩΝ..... 321

9.1. Εισαγωγή.....	321
9.2. Ορισμός της συνάρτησης Spline	322
9.3. Splines στη Στατιστική	324
9.4. Άλλες εφαρμογές Splines.....	326
9.5. Προσέγγιση ποσοστιαίων σημείων.....	328
9.6. Βέλτιστος σχεδιασμός και ορθογώνια πολυώνυμα	329
9.7. Συζήτηση	331
9ο - ΑΣΚΗΣΕΙΣ	336
9ο - ΠΑΡΑΡΤΗΜΑ: Ορισμοί	337

10ο Κεφάλαιο: ΕΙΣΑΓΩΓΗ ΣΤΗ ΜΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ 339

10.1. Εισαγωγή.....	339
10.2. Βελτιστοποίηση (ξανά).....	340
10.3. Στατιστική αντιμετώπιση του προβλήματος	346
10.4. Γραμμική προσέγγιση.....	349
10.5. Συζήτηση	351
10ο ΑΣΚΗΣΕΙΣ	352

Παράρτημα I: ΠΡΟΓΡΑΜΜΑΤΑ PASCAL 357

Παράρτημα II: ΜΕΓΙΣΤΑ-ΕΛΑΧΙΣΤΑ 369

Επίλογος..... 387

Βιβλιογραφία

Αναλυτικά περιεχόμενα

395

Το βιβλίο αυτό είναι το τέταρτο της σειράς Εφαρμοσμένης Στατιστικής, που έχει γράψει ο Επίκ. Καθηγητής Χ. Κίτσος.

Περιέχει θέματα Υπολογιστικής Στατιστικής με πλήθος παραδειγμάτων και προγραμμάτων σε Minitab, SPSS, Statgraphics και Pascal.

Κατεβλήθη ιδιαίτερη προσπάθεια τα παραδείγματα και οι εφαρμογές να είναι άμεσα κατανοητά και πλήρως λυμένα ώστε ο φιλομαθής αναγνώστης να τα χρησιμοποιεί στο PC του.

Τα θέματα Υπολογιστικής Στατιστικής ενδιαφέρουν τους:

- Στατιστικούς
- Γεωπόνους, Χημικούς, Βιολόγους
- Γιατρούς
- Μαθηματικούς



ΕΚΔΟΣΕΙΣ
ΝΕΩΝ

ΤΕΧΝΟΛΟΓΙΩΝ

ISBN: 960-7235-72-x