


Τσάντας Νίκος
Μωυσιάδης Χρόνης
Μπαγιάτης Ντίνος
Χατζηπαντελής Θόδωρος



ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΜΕ ΤΗ ΒΟΗΘΕΙΑ ΣΤΑΤΙΣΤΙΚΩΝ ΠΑΚΕΤΩΝ

S P S S

Excel

S-Plus

Πρόλογος

Η ανάγκη των ερευνητών οποιοδήποτε επιστημονικού χώρου να επιβεβαιώσουν πειραματικά τις θεωρίες που διατυπώνονται ή που θεωρούνται ότι ισχύουν ισχυροποιώντας με τον τρόπο αυτό τα συμπεράσματά τους, οδήγησε στην ανάπτυξη της αποκαλούμενης *ανάλυσης δεδομένων*. Με τον όρο αυτό εννοούνται, όχι μόνον οι τεχνικές και οι μέθοδοι επεξεργασίας πληροφοριών που προέκυψαν από πραγματική ή εικονική εκτέλεση κατάλληλων πειραμάτων ή παρακολούθησης φαινομένων, αλλά και η θεσμοθέτηση κοινά παραδεκτών μονάδων μέτρησης και μεθόδων συλλογής ή διαχείρισης αυτών των πληροφοριών. Εξάλλου η μεγάλη ανάπτυξη της τεχνολογίας των υπολογιστικών μηχανών καθώς και των δυνατοτήτων μηχανοργάνωσης και αποθήκευσης δεδομένων, τροφοδότησε ακόμη περισσότερο την ανάγκη αξιολόγησης και χρήσης των υπαρχόντων πληροφοριών για την καλύτερη κατανόηση της μεταβολής των φαινομένων, την πρόβλεψη μελλοντικών καταστάσεων και τη διερεύνηση δυνατοτήτων παρέμβασης. Πολύ δε περισσότερο που η πληροφορία, ως κατεχόμενο προϊόν και μέσο άσκησης πολιτικής, υποκατέστησε σε μεγάλο βαθμό τα υλικά προϊόντα, οδηγώντας σε μια νέα μορφή συγκέντρωσης δύναμης.

Στις μέρες μας, οι μέθοδοι της *περιγραφικής στατιστικής* αποτελούν το επιστημονικό εργαλείο για τη συγκέντρωση, ταξινόμηση και παρουσίαση των πρωτογενών δεδομένων σε εύληπτη μορφή, ενώ, οι τεχνικές της *στατιστικής συμπερασματολογίας*, καθιστούν δυνατή την προσέγγιση χαρακτηριστικών του συνόλου των δεδομένων από τη μελέτη ενός (μικρού) υποσυνόλου τους.

Ως συνέπεια της ευρείας χρήσης των παραπάνω μεθόδων σε οποιαδήποτε περίπτωση, παρατηρείται, δυστυχώς, ολοένα και πιο συχνά το φαινόμενο της λαθε-

μένης χρήσης και της εξαγωγής επισφαλών συμπερασμάτων. Το γεγονός αυτό, οφείλεται σε μεγάλο βαθμό στην ευκολία με την οποία οι διάφοροι ερευνητές μπορούν σήμερα να διεκπεραιώσουν μια στατιστική ανάλυση. Δεν έχουν παρά να προμηθευτούν ένα στατιστικό πακέτο από τα δεκάδες που κυκλοφορούν και να «τρέξουν» το κατάλληλο υποπρόγραμμα. Το υποπρόγραμμα αυτό, λειτουργώντας ως «συνταγή», θα δώσει τα αναμενόμενα συμπεράσματα.

Οι συγγραφείς αυτού του βιβλίου, εργαζόμενοι σε διαφορετικούς χώρους ο καθένας, με πολύχρονη πείρα από στατιστικές αναλύσεις που έχουμε διεξαγάγει σε ποικίλα προβλήματα, πιστεύουμε ότι η Στατιστική δεν είναι ένα σύνολο συνταγών για την ανάλυση των καταγραμμένων δεδομένων. Δεν αρκεί να έχουμε δεδομένα και κάποιο λογισμικό για να κάνουμε την πρόποσα επεξεργασία. Αποτελεί μεγάλη πλάνη η άποψη ότι Στατιστική δεν είναι παρά χειρισμός προγραμμάτων, πινακοποίηση αποτελεσμάτων και παρουσίαση δεδομένων με όμορφα γραφήματα. Αντίθετα πιστεύουμε ότι η Στατιστική, απαιτεί πρώτα απ' όλα την καλή γνώση του παρατηρούμενου φαινομένου, την οργάνωση της παρατήρησης και τη σωστή καταγραφή των δεδομένων. Απαιτεί επιπλέον, θεωρητική γνώση ώστε να επιλεγεί η πλέον κατάλληλη μέθοδος για τη στατιστική ανάλυση που επιδιώκουμε. Απαιτεί τέλος, το σχολαστικό έλεγχο της ικανοποίησης των απαραίτητων προϋποθέσεων για την εφαρμογή των διάφορων στατιστικών τεχνικών και μεθόδων. Οι βασικές γνώσεις Στατιστικής που διδάσκονται στους μαθητές της πρωτοβάθμιας και δευτεροβάθμιας εκπαίδευσης έχουν ως στόχο την ανάπτυξη της δυνατότητας ανάγνωσης στατιστικών αποτελεσμάτων κι όχι την ανάλυση δεδομένων. Ανάλογα ισχύουν και για πολλά πανεπιστημιακά Τμήματα στα οποία η Στατιστική χρησιμοποιείται μόνο ως εργαλείο.

Στο βιβλίο αυτό περιγράφουμε μεθόδους επεξεργασίας και ελέγχου υποθέσεων για διάφορα σύνολα δεδομένων, καθώς και τους τρόπους οργάνωσης της παρατήρησης και της καταγραφής που απαιτεί η κάθε μέθοδος. Οι μέθοδοι αυτές εφαρμόζονται σε δεδομένα, είτε πραγματικά, είτε φανταστικά, είτε γνωστά από τη διεθνή βιβλιογραφία, με τη βοήθεια τριών αντιπροσωπευτικών προγραμμάτων, που το καθένα τους έχει σχεδιαστεί για την αντιμετώπιση διαφορετικού επιπέδου προβλημάτων. Τα προγράμματα αυτά είναι το EXCEL, το S.P.S.S. και το S-Plus.

Το **EXCEL** αποτελεί εξελιγμένη μορφή προγράμματος «λογιστικού φύλλου» που στοχεύει στην καταγραφή δεδομένων στη μορφή πίνακα διπλής εισόδου σύμφωνα με το σχήμα:

Γραμμή-Υποκείμενο(περίπτωση), Στήλη-Μεταβλητή

Με τη χρήση μακροεντολών και ενσωματωμένων προγραμμάτων δίνει την δυνα-

τότητα υπολογισμού στατιστικών μέτρων και σχεδιασμού γραφικών παραστάσεων, μετατροπών και αριθμητικών υπολογισμών.

Το **S.P.S.S.** είναι ένα από τα πιο εξελιγμένα υπολογιστικά προγράμματα Στατιστικής που προσπαθεί να καλύψει το σύνολο των «γνωστότερων» στατιστικών τεχνικών. Ιδιαίτερα στις εκδόσεις του σε γραφικό περιβάλλον, δίνει τη δυνατότητα γραφικής επεξεργασίας και αναδραστικής λειτουργίας με πολύ μεγάλη επιτυχία.

Το **S-Plus** είναι ένα υπολογιστικό πρόγραμμα Στατιστικής με μεγάλη δυνατότητα προγραμματισμού διαδικασιών. Με τον τρόπο αυτό, προσφέρει στον έμπειρο χρήστη υψηλούς βαθμούς ελευθερίας στον χειρισμό συνόλων δεδομένων.

Στο βιβλίο, γίνεται προσπάθεια να περιγραφεί η κάθε στατιστική διαδικασία και η χρήση των παραπάνω στατιστικών προγραμμάτων σε αυτή. Οι στατιστικές διαδικασίες παρουσιάζονται όχι με τις τεχνικές τους λεπτομέρειες αλλά με τη λειτουργία και τα όρια τους. Για να επιλέξει κάποιος την κατάλληλη διαδικασία πρέπει να κινηθεί επιλέγοντας εκείνη που προσαρμόζεται στα δεδομένα του και όχι να προσαρμόσει τα δεδομένα του σε κάποια γνωστή διαδικασία. Έμφαση δίνεται επίσης στην καταγραφή των κριτηρίων για την επιλογή των διαφόρων στατιστικών τεχνικών, καθώς επίσης και των ορίων εφαρμογής της.

Θεσσαλονίκη, 1999

Νίκος Τσάντας
Χρόνης Μωυσιάδης
Ντίνος Μπαγιάτης
Θόδωρος Χατζηπαντελής

Περιεχόμενα

Πρόλογος

Κεφάλαιο 1. Εισαγωγή στη Στατιστική Ανάλυση

1.1 Στατιστική	
Η πορεία της επιστήμης	1-3
Λειτουργίες της Στατιστικής	1-11
Σκοπός του βιβλίου	1-12
1.2 Επιστημονική Έρευνα και Στατιστική	
Πληθυσμός	1-16
Δείγμα	1-18
“Μετρώντας...”	1-24
Garbage In, Garbage Out ...	1-32
Πειραματικός Σχεδιασμός	1-41
Στατιστικές Προσεγγίσεις στην Ανάλυση Δεδομένων	1-45
1.3 Στατιστικές Τεχνικές	
Ανάλυση μιας μεταβλητής	1-50
Σύγκριση δύο δειγμάτων μιας (της ίδιας) μεταβλητής	1-53
Σύγκριση περισσότερων δειγμάτων μιας (της ίδιας) μεταβλητής	1-56
Σχέση – Σύγκριση δύο μεταβλητών του ίδιου δείγματος	1-62

Κεφάλαιο 2. Ανάλυση μιας μεταβλητής

2.1 Περιγραφική Στατιστική

2.1.1 S.P.S.S

Ποιοτικά χαρακτηριστικά

Η διαδικασία <i>Frequencies</i>	2-4
Πίνακες Συχνοτήτων, Ραβδογράμματα, Κυκλικά Διαγράμματα.	

Ποσοτικά χαρακτηριστικά

Η διαδικασία <i>Descriptives</i>	2-7
Συνήθη Στατιστικά Μέτρα.	
Η διαδικασία <i>Frequencies</i>	2-10
Στατιστικά Μέτρα, Ιστογράμματα, Πίνακες Συχνοτήτων (<i>Recode, Aggregate</i>), Πολύγωνο Συχνοτήτων.	
Η διαδικασία <i>Explore</i>	2-15
Στατιστικά μέτρα, Μ-εκτιμητές, Παράτυπα και Ακραία Σημεία, Θηκογράμματα, Φυλλογραφήματα, Ιστογράμματα, Πιθανοθεωρητικά Γραφήματα, Έλεγχος Κανονικότητας.	

2.1.2 Excel

Ποιοτικά χαρακτηριστικά

Η συνάρτηση <i>Frequency</i>	2-23
Πίνακες Συχνοτήτων, Ραβδογράμματα, Κυκλικά Διαγράμματα	
Συγκεντρωτικοί Πίνακες	2-25
Πίνακες Συχνοτήτων, Ραβδογράμματα, Κυκλικά Διαγράμματα, Pareto Διαγράμματα	

Ποσοτικά χαρακτηριστικά

Το εργαλείο <i>Descriptive Statistics</i>	2-29
Συνήθη Στατιστικά Μέτρα.	
Ενσωματωμένες Στατιστικές Συναρτήσεις	2-31
Συνήθη Στατιστικά Μέτρα (<i>average, median, mode, stdev, var, kurt, skew, min, max, sum, count, large, small, percentile, quartile</i>).	
Το εργαλείο <i>Histogram</i>	2-33
Πίνακες Συχνοτήτων, Ιστογράμματα, Πολύγωνο Αθροιστικών Συχνοτήτων.	
Κατασκευή Πολυγώνου Συχνοτήτων	2-36
Κατασκευή Θηκογράμματος	2-38

2.1.3 S-Plus

Προτεινόμενα Γραφήματα	2-41
Παρουσίαση των Δεδομένων	2-43
Ιστόγραμμα - Πολύγωνο Συχνοτήτων	2-48
Εμπειρική Συνάρτηση Κατανομής	2-52
Φυλλογράφημα - Θηκόγραμμα	2-53

Ραβδόγραμμα – Κυκλικό διάγραμμα	2-55
2.2 Στατιστική Συμπερασματολογία	
2.2.1 S.P.S.S	
Έλεγχος τυχαιότητας του δείγματος Η διαδικασία <i>Runs</i> στα <i>Nonparametric Tests</i> .	2-60
Έλεγχος καλής προσαρμογής (χ^2 test) Η διαδικασία <i>Chi-Square</i> στα <i>Nonparametric Tests</i> .	2-63
Έλεγχος προσαρμογής με το κριτήριο των K-S Η διαδικασία <i>1-Sample K-S</i> στα <i>Nonparametric Tests</i> .	2-69
Διωνυμικός έλεγχος (Binomial test) Η διαδικασία <i>Binomial</i> στα <i>Nonparametric Tests</i> .	2-74
Πιθανοθεωρητικά γραφήματα (P-P, Q-Q) Η επιλογή <i>P-P</i> και <i>Q-Q</i> του καταλόγου <i>Graphs</i> .	2-77
Έλεγχος υποθέσεων για τη μέση τιμή (t test) Η διαδικασία <i>One-Sample T Test</i> .	2-84
Έλεγχος υποθέσεων για το ποσοστό (Binomial test) Η διαδικασία <i>Binomial</i> στα <i>Nonparametric Tests</i> .	2-86
Έλεγχος υποθέσεων για τη διάμεσο (Sign test)	2-90
Έλεγχος αυτοσυσχέτισης Η επιλογή <i>Time Series Autocorrelations</i> του καταλόγου <i>Graphs</i> .	2-92
2.2.2 Excel	
Παραγωγή συνόλων τυχαίων αριθμών με συγκεκριμένη κατανομή Το εργαλείο <i>Random Number Generation</i> .	2-93
Πιθανοθεωρητικά γραφήματα (Normal Q-Q)	2-96
Έλεγχος υποθέσεων για τη μέση τιμή (t test)	2-99
Έλεγχος υποθέσεων για τη μέση τιμή ενός πληθυσμού όταν η διακύμανση είναι γνωστή	2-105
Έλεγχος υποθέσεων για το ποσοστό (Binomial test)	2-109
2.2.3 S-Plus	
Έλεγχος τυχαιότητας του δείγματος Μέση Τετραγωνική Διαφορά, Κριτήριο των Ροών, Κριτήριο Διαφοράς Προσήμου.	2-115
Έλεγχος καλής προσαρμογής (Goodness Of Fit tests) Κριτήριο χ^2 Προσαρμογής για ένα δείγμα, Κριτήριο Kolmogorov-Smirnov για ένα δείγμα, Γραφικοί Έλεγχοι Προσαρμογής σε Γνωστές Κατανομές	2-120
Έλεγχος αναλογιών	2-131
Έλεγχοι Υποθέσεων για ένα Δείγμα Έλεγχος Υποθέσεων για τη Μέση Τιμή του Πληθυσμού, Έλεγχος για τη Διάμεσο του Πληθυσμού, Έλεγχος για τη Διασπορά του Πληθυσμού	2-134

Κεφάλαιο 3. Ανάλυση δύο μεταβλητών

3.1 Περιγραφική Στατιστική

3.1.1 S.P.S.S

Ποιοτικά χαρακτηριστικά

Η διαδικασία <i>Crosstabs</i> Πίνακες Συνάφειας, Βαθμός και Φύση της Σχέσης Μεταξύ δύο Μεταβλητών.	3-4
---	-----

Ποσοτικά χαρακτηριστικά

Η διαδικασία <i>Case Summaries</i>	3-25
------------------------------------	------

Η διαδικασία <i>Means</i>	3-29
---------------------------	------

Γραφικές παραστάσεις	3-32
----------------------	------

Ραβδογράμματα, Γραμμογραφήματα, Εμβοδογραφήματα, Κυκλικά Διαγράμματα και Διαγράμματα Σφαλμάτων για τις Τιμές Διαφόρων Στατιστικών Μέτρων.

Η διαδικασία <i>Explore</i>	3-33
-----------------------------	------

Στατιστικά μέτρα, Μ-εκτιμητές, Παράτυπα και Ακραία Σημεία, Θηκογράμματα, Φυλλογραφήματα, Ιστογράμματα, Πιθανοθεωρητικά Γραφήματα, Έλεγχος Κανονικότητας και Ισότητας των Διακυμάνσεων.

Διαγράμματα Διασποράς	3-45
-----------------------	------

Απλά, Επικαλυπτόμενα, Πολλαπλά και Τρισδιάστατα Διαγράμματα Διασποράς. Επεξεργασία και Τροποποίηση ενός Διαγράμματος Διασποράς.

Η διαδικασία <i>Correlate</i>	3-60
-------------------------------	------

Συντελεστές Συσχέτισης του Pearson και του Spearman, Πίνακες Συσχέτισης, Μερικός Συντελεστής Συσχέτισης, Αντιστοιχία Διαγραμμάτων Διασποράς και Συντελεστών Συσχέτισης.

3.1.2 Excel

Ποιοτικά χαρακτηριστικά

Συγκεντρωτικοί Πίνακες	3-70
------------------------	------

Ποσοτικά χαρακτηριστικά

Συγκεντρωτικοί Πίνακες	3-78
------------------------	------

Διαγράμματα Διασποράς	3-82
-----------------------	------

Κατασκευή Διαγραμμάτων Διασποράς. Προσθήκη Γραμμών Τάσης. Ο Συντελεστής Συσχέτισης του Pearson. Το εργαλείο *Correlation*. Συνδιακύμανση. Το εργαλείο *Covariance*.

3.1.3 S-Plus

Πίνακες Συνάφειας	3-89
-------------------	------

Διαγράμματα Διασποράς	3-101
-----------------------	-------

Διαγράμματα Χρονοσειρών	3-104
-------------------------	-------

Γραφήματα Trellis	3-106
-------------------	-------

3.2 Στατιστική Συμπερασματολογία (σύγκριση δύο δειγμάτων)

3.2.1 S.P.S.S

Ανεξάρτητα δείγματα

Έλεγχος υποθέσεων για τις μέσες τιμές δύο πληθυσμών	3-112
H διαδικασία <i>Independent-Sample T-test</i> του καταλόγου <i>Compare Means</i> . H διαδικασία <i>2 Independent Samples</i> στα <i>Nonparametric Tests</i>	

Εξαρτημένα δείγματα

Έλεγχος υποθέσεων για τις μέσες τιμές δύο πληθυσμών	3-120
H διαδικασία <i>Paired-Samples T-test</i> του καταλόγου <i>Compare Means</i> . H διαδικασία <i>2 Related Samples</i> στα <i>Nonparametric Tests</i> .	

Έλεγχος αναλογιών	3-126
Έλεγχος υποθέσεων για τις αναλογίες δύο ανεξάρτητων πληθυσμών. Έλεγχος υποθέσεων για τις αναλογίες δύο εξαρτημένων πληθυσμών.	

Έλεγχος διαμέσων	3-129
Median test.	

Έλεγχος διακυμάνσεων	3-131
----------------------	-------

3.2.2 Excel

Ανεξάρτητα δείγματα

Έλεγχος υποθέσεων για τις μέσες τιμές δύο πληθυσμών	3-137
Τα εργαλεία <i>F-Test</i> και <i>t-Test</i> . Οι συναρτήσεις <i>FTEST</i> και <i>TTEST</i> .	

Εξαρτημένα δείγματα

Έλεγχος υποθέσεων για τις μέσες τιμές δύο πληθυσμών	3-147
Το εργαλείο <i>t-Test</i> .	

Έλεγχος αναλογιών	3-149
-------------------	-------

Έλεγχος διακυμάνσεων	3-153
----------------------	-------

3.2.3 S-Plus

Έλεγχος μέσων τιμών δύο δειγμάτων από κανονικούς πληθυσμούς	3-156
---	-------

Έλεγχος διασπορών δύο δειγμάτων από κανονικούς πληθυσμούς	3-158
---	-------

Μικρά δείγματα	3-161
----------------	-------

Έλεγχοι καλής προσαρμογής δύο δειγμάτων	3-168
---	-------

Έλεγχος αναλογιών	3-169
-------------------	-------

3.3 Στατιστική Συμπερασματολογία (σχέση μεταξύ δύο μεταβλητών)

3.3.1 S.P.S.S

Ποιοτικά χαρακτηριστικά

Η διαδικασία <i>Crosstabs</i>	3-173
χ^2 test ανεξαρτησίας, Βαθμός και Φύση της Σχέσης Μεταξύ δύο Μεταβλητών.	
Ποσοτικά χαρακτηριστικά	
Η διαδικασία <i>Correlate</i>	3-186
Συντελεστές Συσχέτισης των Pearson, Kendall και Spearman, Πίνακες Συσχέτισης, Μερικός Συντελεστής Συσχέτισης.	
Η διαδικασία <i>Regression</i>	3-195
Απλή Γραμμική Παλινδρόμηση, Έλεγχος του Μοντέλου, Ανάλυση Καταλοίπων.	
3.3.2 S-Plus	
Έλεγχοι Ανεξαρτησίας σε Πίνακες Συνάφειας	3-221
Απλή Γραμμική Παλινδρόμηση	3-225
Λείανση (Smoothing)	3-236

Παράρτημα Α. Η γλώσσα S

A.1 Το στατιστικό πακέτο S-Plus	A-3
A.2 Εκκίνηση	A-3
A.3 Ονομασία αντικειμένων στην S	A-4
A.4 ΣύNTAXη	A-4
A.5 Διανύσματα και πίνακες	A-6
A.6 Λίστες	A-9
A.7 Παράγοντες	A-10
A.8 Πλαίσια δεδομένων	A-11
A.9 Εκφράσεις και τελεστές	A-13
A.10 Ο κανόνας ανακύκλωσης	A-15
A.11 Συναρτήσεις	A-16
A.12 Συστηματικές ακολουθίες αριθμών	A-18
A.13 Τυχαίοι αριθμοί	A-20
A.14 Ανάγνωση δεδομένων	
Από αρχεία ASCII	A-22
Η συνάρτηση <code>read.table</code>	A-24
Η συνάρτηση <code>scan</code>	A-24

Ανάγνωση αρχείων διαφορετικών τύπων	A-26
A.15 Διάταξη	A-29
A.16 Έξοδος δεδομένων – Εκτύπωση	A-31
A.17 Πράξεις με διανύσματα και πίνακες	
Οι συναρτήσεις <code>crossprod</code> και <code>outer</code>	A-34
Οι συναρτήσεις <code>apply</code> και <code>sweep</code>	A-35
Συναρτήσεις που εφαρμόζονται σε πίνακες	A-36
A.18 Τροποποίηση του S–περιβάλλοντος	A-39
A.19 Εργαλεία του S-Plus	
Διαχειριστής αντικειμένων (Object Manager)	A-41
Άνοιγμα – Κλείσιμο Συσκευών για γραφήματα	A-43
Εκτέλεση εντολών DOS ή Windows	A-45
Προηγηθείσες Εντολές	A-46
A.20 Γραφήματα στο S-Plus	A-47

Παράρτημα B. S.P.S.S.

B.1 Προεπισκόπηση	B-3
B.2 Τα αρχεία δεδομένων στο SPSS	
Ανάγνωση SPSS αρχείων δεδομένων	B-6
Ανάγνωση ASCII αρχείων δεδομένων Συστηματική Μορφή. Ελεύθερη Μορφή.	B-7
Ανάγνωση διαφόρων formats αρχείων δεδομένων	B-10
Αποθήκευση αρχείων δεδομένων	B-10
B.3 Εισαγωγή δεδομένων	B-11
Εισαγωγή νέων περιπτώσεων ή/και μεταβλητών	B-13
Διαγραφή περιπτώσεων ή/και μεταβλητών	B-14
Ορισμός μεταβλητών	B-14
Αναζήτηση (μεταβλητών, περιπτώσεων, τιμών)	B-17
B.4 Μετασχηματισμοί μεταβλητών	
Δημιουργία νέων μεταβλητών Η εντολή COMPUTE.	B-19
(Επανα)κωδικοποίηση τιμών Η εντολή RECODE.	B-24

B.5 Αναμόρφωση των δεδομένων	
Ταξινόμηση των τιμών μιας μεταβλητής	B-27
Αναστροφή δεδομένων	B-28
Ranking	B-29
Αντιστοίχιση βαρών	B-31
B.6 Επιλογή περιπτώσεων	B-31
B.7 Το αρχείο αποτελεσμάτων του SPSS	B-33
Outline Pane	B-34
Display Pane	B-35
Pivot Tables	B-36
Γραφικές Παραστάσεις	B-41
Εκτύπωση Αποτελεσμάτων	B-43
Αποθήκευση Αποτελεσμάτων	B-43
Έξοδος Αποτελεσμάτων	B-44
B.8 Ανακατασκευή αρχείων στο SPSS	
Κατακόρυφη συνένωση αρχείων	B-46
Οριζόντια (παράλληλη) συνένωση αρχείων	B-47
Διάσπαση αρχείου	B-48
Συνάθροιση δεδομένων	B-49
B.9 Το αρχείο εντολών του SPSS	B-51
Εκτέλεση του περιεχομένου ενός αρχείου εντολών	B-54
B.10 Οι γραφικές παραστάσεις του SPSS	B-54
Δημιουργία γραφικών παραστάσεων	B-55
Ραβδογράμματα, Γραμμογραφήματα, Εμβοδογραφήματα, Κυκλικά διαγράμματα, Θηκογράμματα, Διαγράμματα σφαλμάτων, Διαγράμματα διασποράς, Ιστογράμματα, Πιθανοθεωρητικά γραφήματα.	
Επεξεργασία-τροποποίηση γραφικών παραστάσεων	B-72
Αποθήκευση του γραφήματος	B-73
B.11 Αναζητώντας βοήθεια	B-73

Παράρτημα Γ. Excel

Γ.1 Εισαγωγή	Γ-3
Γ.2 Προεπισκόπηση	Γ-3
Καταχώριση Δεδομένων, Η έννοια της Περιοχής, Εισαγωγή Συναρτήσεων σ' ένα Φύλο Εργασίας, Τα Εργαλεία Analysis ToolPak.	

Γ.3 Οι στατιστικές συναρτήσεις του Excel	
Γενικές Συναρτήσεις COUNT, COUNTA, COUNTIF, MIN, MINA, MAX, MAXA, LARGE, SMALL, SUM, SUMIF, SUMXMY2, SUMX2MY2, SUMX2PY2, SUMSQ, SUMPRODUCT, INDEX.	Γ-9
Στατιστικές Συναρτήσεις AVERAGE, AVERAGEA, TRIMMEAN, CONFIDENCE, MEDIAN, MODE, GEOMEAN, HARMEAN, AVEDEV, DEVSQ, STDEV, STDEVA, VAR, VARA, STDEVP, STDEVPA, VARP, VARPA, KURT, SKEW, PERCENTILE, QUARTILE, FREQUENCY, STARDIZE, RANK, PERCENTRANK, COVAR, CORREL, PEARSON, LINEST, RSQ, SLOPE, INTERCEPT, STEYX, TREND, FORECAST, LOGEST, GROWTH.	Γ-20
Συναρτήσεις Κατανομών BINOMDIST, CRITBINOM, HYPGEOMDIST, NEGBINOMDIST, POISSON, BETADIST, BETAINV, CHIDIST, CHIINV, EXPONDIS, FDIST, FINV, GAMMADIST, GAMMAINV, GAMMALN, NORMDIST, NORMINV, NORMSDIST, NORMSINV, LOGNORMDIST, LOGINV, TDIST, TINV, WEIBULL, FISHER, FISHERINV, COMBIN, PERMUT.	Γ-55
Συναρτήσεις Στατιστικών Ελέγχων (tests) CHITEST, FTEST, TTEST, ZTEST.	Γ-74
Γ.4 Δημιουργώντας γραφήματα	Γ-77
Ο οδηγός γραφημάτων του Excel.	
Γ.5 Συγκεντρωτικοί πίνακες (Pivot Tables)	Γ-79
Ο οδηγός συγκεντρωτικού πίνακα του Excel.	

1.1 Στατιστική

Η Στατιστική στις μέρες μας, την εποχή της πληροφοριακής έκρηξης δηλαδή, ορίζεται ως η *“επιστήμη που ασχολείται με τη συγκέντρωση, παρουσίαση, αξιολόγηση και την εν συνεχεία επεξεργασία (εξαγωγή συμπερασμάτων) της πληροφορίας”*. Η κοινωνία ολοένα και επιτακτικότερα απαιτεί πληροφορίες πλήρεις, ποιοτικά ελεγμένες, απαλλαγμένες από “θορύβους” που σκόπιμα ή λόγω λάθους περιέχονται σ’ αυτές. Ιδιαίτερη ευαισθησία αναπτύσσεται πια για θέματα που αφορούν ευρύτερα κοινωνικά φαινόμενα, όπως για παράδειγμα, οι μετρήσεις ακροαματικότητας των ΜΜΕ, των εκπαιδευτικών παραμέτρων, των οικονομικών μεγεθών, της γνώμης για κόμματα ή/και πρόσωπα, κλπ.

Η ιδιαιτερότητα της Στατιστικής ως επιστήμης είναι ότι, ενώ στο επίπεδο της θεωρίας είναι μαθηματικά, στο επίπεδο της εφαρμογής χρησιμοποιεί το πλαίσιο όλων σχεδόν των άλλων γνωστικών περιοχών. Οι κοινωνικο-οικονομικοί δείκτες, η εξέλιξη των φυσικών φαινομένων, οι επιδημιολογικές μελέτες, οι κλινικές δοκιμές φαρμάκων, ο ποιοτικός έλεγχος των προϊόντων, η εκτίμηση της στάσης των πολιτών πάνω σ’ ένα θέμα, η αξιόπιστη ‘καθαρή’ εικόνα μιας δορυφορικής φωτογραφίας, η αναζήτηση διαδικασιών αξιολόγησης οικολογικών συστημάτων, είναι μόνον μερικές από τις περιπτώσεις που η Στατιστική συνεργάστηκε επιτυχώς με άλλες επιστήμες (Πληροφορική, Οικονομία, Βιολογία, Ιατρική, Χημεία, Ψυχολογία, κλπ). Ακόμη και στη Νομική αναγνωρίζεται ότι η Στατιστική αποτελεί “μαχητό τεκμήριο απόδειξης” (*prima facie*), ως αποδεικτική δηλαδή διαδικασία που για να την αντικρούσεις πρέπει να χρησιμοποιήσεις αντίστοιχα επιχειρήματα και αντίστοιχες διαδικασίες (Ανώτατο Δικαστήριο των ΗΠΑ 1964).

Η πορεία της επιστήμης

Η ρίζα της λέξης *“Στατιστική”* αναζητείται στη λατινική λέξη *“STATUS”* (καθεστώς, κοινωνία, κράτος) και παραπέμπει στην καταγραφή και αρχειοθέτηση ποσοτικών μεγεθών. Υπό αυτή την έννοια, η “ηλικία” της λέξης είναι προγενέστερη. Βαβυλώνιοι, Αιγύπτιοι και Κινέζοι είχαν προχω-

ρήσει (περί το 3000 π.Χ.) σε συλλογή στοιχείων για φορολογικούς, στρατιωτικούς (άντρες σε στρατεύσιμη ηλικία), αλλά και άλλους σκοπούς (π.χ. στην Αίγυπτο για την κατασκευή των πυραμίδων και τη διανομή της γης μετά τις ετήσιες πλημμύρες του Νείλου). Το τέταρτο βιβλίο της Παλαιάς Διαθήκης, ξεκινά με την εντολή του Θεού προς τον Μωυσή να προχωρήσει σε απογραφή όλων των Ισραηλινών που είχαν τη δυνατότητα να πολεμήσουν :

«Και ελάλησε Κύριος προς Μωυσήν εν τη ερήμω τη Σινά, εν τη σπηνή του μαρτυρίου, εν μια του μηνός του δευτέρου, έτους δευτέρου εξελθόντων αυτών εκ γης Αιγύπτου λέγων ²Λάβετε αρχήν πάσης συναγωγής Ισραήλ κατά συγγενείας, κατ' οίκους πατριών αυτών, κατ' αριθμόν εξ' ονόματος αυτών, κατά κεφαλήν αυτών. ³Πάς άρσρην από εικοσαετούς και επάνω, πως ο επιπορευόμενος εν δυνάμει Ισραήλ, επισκέψασθε αυτούς συν δυνάμει αυτών, συ και Λαρών επισκέψασθε αυτούς.»

Αλλά και αργότερα, περί το 1500 π.Χ., ο βασιλιάς Δαυίδ προέβη σε απογραφή :

«Και προσέθετο αρχή κυρίου επικαήναι εν Ισραήλ, και επέσεισεν τον Δαυίδ εν αυτοίς λέγων Βάδιζε αριθμησον τον Ισραήλ και τον Ιουδα ²Και είπεν ο βασιλεύς προς Ιωάβ άρχοντα της ισχύος του μετ' αυτού Δίελθε δη πάσας φυλάς Ισραήλ από Δαν και έως Βηρσαβες και επισκεψαι τον λαόν, και γνώσομαι τον αριθμόν του λαού ... ⁹Και έδωκεν Ιωάβ τον αριθμόν της επισκέψεως του λαού προς τον βασιλέα, και εγένετο Ισραήλ οκτακόσιοι χιλιάδες ανδρών δυνάμειως σπιωμένων ρομφαίαν και ανήρ Ιουδα πεντακόσιοι χιλιάδες ανδρών μαχητών ¹⁰Και επάταξεν καρδιά Δαυίδ αυτόν μετά το αριθμήσαι τον λαόν, και είπεν Δαυίδ προς κύριον Ήμαρτον σφόδρα ό έποίησα νυν, κύριε, παραβίβασον δη την ανομίαν του δούλου σου, όπι εμαράνθη σφόδρα.»

(δέκατο βιβλίο της Παλαιάς Διαθήκης). Κατά την παράδοση, η απογραφή προκάλεσε την οργή του Θεού, με αποτέλεσμα, μετά την ολοκλήρωσή της, να αποδκατιστεί από πανούκλα ο πληθυσμός.

Πηγές που να αναφέρονται στην καταγραφή ή την αξιολόγηση μεγεθών στην αρχαία Ελλάδα δεν έχουν ανακαλυφθεί. Στην Πολιτεία όμως του Αριστοτέλη, περιέχεται η συγκριτική περιγραφή 158 πολιτειών (ακριβώς το θέμα το οποίο, στα μέσα του 17ου αιώνα, Άγγλοι συγγραφείς, χαρακτήρισαν με τον όρο “Στατιστική”). Αντίθετα, πολλές φαίνεται να είναι οι αναφορές στην έννοια της πιθανότητας (εικός). Ο Πλάτωνας εκφράζεται αρνητικά γράφοντας στο Φαίδωνα :

«...Εγώ δε τοις δια των εκότων τας αποδείξεις ποιούμενοις λόγοις ξύνκειδα ούσιν αλαζόσιν, και, αν τις αυτούς μη φυλάττηται, ευ μάλα εξαπατώσι, και εν γεωμετρία και εν τοις άλλοις άπκοσι.»

Από την άλλη μεριά, στο αριστοτελικό φιλοσοφικό σύστημα συναντάμε μια θετικότερη στάση :

«...Ο γαρ ως επί το πολύ ίσασιν ούτω γινόμενον ή μη γινόμενον ή
ον ή μή ον, τούτ' εστίν εκός, οίον το μισείν τους φθονούντας ή το
φιλείν τους ερωμένους.»

(*Αναλυτικά Πρότερα*). Κατά πολλούς, στον Αριστοτέλη οφείλεται ο εμπειρικός ορισμός της πιθανότητας (σχετική συχνότητα). Στο έργο του *Φυσικά*, ο Αριστοτέλης προσπαθεί να τεκμηριώσει τη δυνατότητα εξαγωγής συμπερασμάτων από την παρατήρηση της εξέλιξης των φαινομένων. Κατά τον Αριστοτέλη, τα φαινόμενα διακρίνονται σε “γεγονότα” των οποίων η πραγματοποίηση καθορίζεται από την τύχη (δεν μπορεί δηλαδή να προβλεφτούν) και σε “γεγονότα” των οποίων η πραγματοποίηση καθορίζεται από “φυσικούς νόμους”. Μερικά από αυτά γίνονται πάντα με τον ίδιο τρόπο, ενώ για άλλα, η πραγματοποίησή τους μπορεί να προβλεφτεί με “επαναληπτική συχνότητα” (αντιστοιχώντας δηλαδή, όπως θα λέγαμε σήμερα, μία πιθανότητα σε καθένα εξ' αυτών).

Η Ρωμαϊκή αυτοκρατορία διατηρούσε ένα οργανωμένο σύστημα καταγραφής των κρατικών αγαθών (που ανήκαν στον αυτοκράτορα). Οι Ρωμαίοι, ξεκινώντας από την εποχή του δού βασιλιά της Ρώμης Servius Tullius (578-534 π.Χ.), πραγματοποιούσαν τακτικές απογραφές κάθε πέντε χρόνια. Μάλιστα, η λέξη “census” (απογραφή) προέρχεται από τη λατινική λέξη “censere” (προς φορολόγηση). Τη διαδικασία αυτή, επέκτεινε αργότερα σ' ολόκληρη τη Ρωμαϊκή αυτοκρατορία, ο Καίσαρας Αύγουστος (63 π.Χ. – 14 μ.Χ.). Γνωστή είναι σε όλους μας η γενική απογραφή πληθυσμού την εποχή της γέννησης του Ιησού Χριστού :

«¹Εγένετο δε εν ταις ημέραις εκείναις εξήλθε δόγμα παρά Καίσαρος Αυγούστου απογράφεσθαι πάσαν την οικουμένην ²Αυτή η απογραφή πρώτη εγένετο ηγεμονεύοντος της Συρίας Κυρηνίου ³Και επισρεύοντα απογράφεσθαι, έκαστος εις την ιδίαν πόλιν. Ανέβη δε και Ιωσήφ από της Γαλιλαίας εκ πόλεως Ναζαρέτ εις την Ιουδαίαν εις πόλιν Δαβίδ, ήτις καλείται Βηθλεέμ, δια το είναι εξ' οίκου και πατριάς Δαβίδ. ⁴απογράφεσθαι συν Μαριάμ τη μεμνηστευμένη αυτώ γυναίκι, ούση εγκύω.»

(*Κατά Λουκάν Ευαγγέλιο*).

Ο φεουδαρχισμός κατά τη διάρκεια του Μεσαίωνα, κατέστησε πρακτικά αδύνατη τη διενέργεια απογραφών, αν και υπήρξαν κάποιες απόπειρες (σαν αυτή του Καρλομάγνου για την εκκλησιαστική περιουσία το 762).

Τα Χριστούγεννα του 1085, ο Γουλιέλμος ο Κατακτητής (1027-1087) διέταξε την πραγματοποίηση γενικής απογραφής στην Αγγλία. Τα αποτελέ-

Εικόνα 1.1 *Domesday Book* (1086): στατιστική απογραφή όλων των αγροτικών μονάδων (αγροτεμάχια, καλλιέργειες, μεταλλεία, ιχθυοτροφεία, κλπ) της Αγγλίας ύστερα από διαταγή του βασιλέα Γουλιέλμου του Κατακτητή (1027-1087).



σματά της, τα οποία περιέχονται στο γνωστό ως *Domesday Book* (*Domesday* είναι παράφραση της λέξης *Doomsday* – ημέρα της κρίσης), χρησιμοποιήθηκαν σαν βάση για τη φορολογία μέχρι το 1522, οπότε ένα αντίστοιχο βιβλίο δημιουργήθηκε.

Γνωστό είναι επίσης ότι, ο Ισπανός αντιβασιλέας Don Pedro de la Fasca είχε οργανώσει απογραφή των ιθαγενών στο Περού το 1548. Αλλά και οι Ίνκας, πριν από τον ερχομό των Ισπανών είχαν αναπτύξει ένα σύστημα καταγραφής στατιστικών στοιχείων, το *quipus*. Το σύστημα αυτό, χρησιμοποιούσε χρωματιστές κλωστές και κόμπους για να καταγράψει τις γεννήσεις και τους θανάτους.

Στις αρχές του 16ου αιώνα αναφορές με τον γενικό τίτλο *Bills of Mortality* άρχισαν να εμφανίζονται στο Λονδίνο. Σ' αυτές καταγραφόταν αρχικά οι θάνατοι που προερχόταν από την πανούκλα. Αργότερα, οι αναφορές αυτές επεκτάθηκαν ώστε να συμπεριλάβουν τις γεννήσεις και τους πάσης φύσεως θανάτους. Στα 1662, ένας ευφυής Εγγλέζος καταστηματάρχης, ο John Graunt (1620-1674), δημοσίευσε μια, εξαιρετικά πρωτότυπη για την εποχή, εργασία με τίτλο *Natural and Political Observations upon the Bills of Mortality*. Πρόκειται για μια μελέτη επί των εβδομαδιαίων περιστατικών

(γεννήσεις/θάνατοι) 122 ενοριών και ενός νοσοκομείου του Λονδίνου. Η εργασία του, που ξέφευγε από την απλή καταγραφή περιστατικών και προχωρούσε στην ανάλυσή τους, θεωρείται η απαρχή της δημογραφικής επιστήμης. Ο Graunt κατάφερε να αποδείξει ότι τεράστιοι όγκοι δεδομένων, με όχι ιδιαίτερα απλή μορφή, μπορούσαν να συμπτυχθούν σε λίγους μόνο πίνακες. Το γεγονός αυτό άλλαξε έκτοτε δραματικά, τον τρόπο με τον οποίο εκρατούντο τα στοιχεία στις διάφορες αγγλικές δημόσιες υπηρεσίες.

Αλλά, η μεγάλη συνεισφορά στην επιστήμη της εργασίας του Graunt φάνηκε στα 1693, όταν ο γνωστός αστρονόμος Edmond Halley (1656-1742), βασιζόμενος σ' αυτή, δημιούργησε τους πίνακες επιβίωσης (life tables). Με τους πίνακες αυτούς, ο Halley υπέδειξε την 'τύχη' της επιβίωσης των αν-

Εικόνα 1.2
'Bills of Mortality' αναφορές του 1841.

TO THE WORSHIPFUL
WILLIAM TURNER, ESQ., MAYOR,
The Aldermen, Municipal Councillors,
AND THE BORN OF THE
WORTHY INHABITANTS of the TOWN of NORTHAMPTON,
This Yearly Bill of Mortality
IS PRESENTED BY THEIR MOST OBEYANT HUMBLE SERVANT,
JOHN WEIGHT.

The Bill of Mortality
Within the Parish of All Saints, from the 1st of December, 1840, to the 31st December, 1841.

DECEASED, &c. IN THE PARISH OF ALL SAINTS.											
April	27	Deaths	2	Deaths	2	Sex	4	Wholesome	4		
May	30	Deaths	2	Deaths	2	Infant mortality	7	Deaths	7		
June	30	Deaths	2	Deaths	2	Deaths	4	Deaths	4	Went to the	
July	31	Deaths	2	Deaths	2	Deaths	11	Deaths	11	Dead	4
Aug.	31	Deaths	2	Deaths	2	Deaths	11	Deaths	11		
Sept.	30	Deaths	2	Deaths	2	Deaths	11	Deaths	11		
Oct.	31	Deaths	2	Deaths	2	Deaths	11	Deaths	11		
Nov.	30	Deaths	2	Deaths	2	Deaths	11	Deaths	11		
Dec.	31	Deaths	2	Deaths	2	Deaths	11	Deaths	11		
Total for Year	306	Deaths	20	Deaths	20	Deaths	67	Deaths	67	Deaths	24
Total for Year	306	Deaths	20	Deaths	20	Deaths	67	Deaths	67	Deaths	24

WHICH HAVE BEEN

CHURCHES.											
		Male			Female			Total			
		Male	Female	Total	Male	Female	Total	Male	Female	Total	
All Saints		10	10	20	10	10	20	20	20	40	
St. Andrew		10	10	20	10	10	20	20	20	40	
St. George		10	10	20	10	10	20	20	20	40	
St. Peter		10	10	20	10	10	20	20	20	40	
St. John's Hospital		10	10	20	10	10	20	20	20	40	
Independent Chapel in King Street (All Saints)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. Peter's Church (Formerly Church of St. Andrew)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. George's Church (Formerly Church of St. Peter)		10	10	20	10	10	20	20	20	40	
Independent Meeting in St. John's Church (Formerly Church of St. George)		10	10								

θρώπων κάτω από διαφορετικές συνθήκες. (Η τεχνική αυτή χρησιμοποιείται μέχρι και στις μέρες μας από τους δημογράφους και τις ασφαλιστικές εταιρείες).

Το τέλος του 17ου αιώνα συνεπώς, σηματοδοτεί την απαρχή μιας νέας εποχής για τη Στατιστική και επαναπροσδιορίζει το αντικείμενο της ενασχόλησής της από την μέχρι τότε απλή καταγραφή στοιχείων. Είναι η εποχή κατά την οποία η Στατιστική ταυτίζεται με την αποκαλούμενη "**Πολιτική Αριθμητική**" (*η τέχνη της εξαγωγής συμπερασμάτων από σχήματα για θέματα που ενδιαφέρουν το κράτος*). Για τα χρόνια αυτά, αξίζει να μνημονεύσει κάποιος και τη συνεισφορά των William Petty (1623-1687) και John Arbuthnot (1667-1735). Ο William Petty υπήρξε ένας εκ των θεμελιωτών της Πολιτικής Αριθμητικής και των ιδρυτών της Βασιλικής Ακαδημίας. Σε μια σειρά εργασιών του με το γενικό τίτλο *Essays in Political Arithmetic*, δημοσιευμένες στην περίοδο 1683-1690, χρησιμοποίησε την τεχνική του Graunt για να παρουσιάσει μια συγκριτική μελέτη της εξέλιξης του πληθυσμού, του εμπορίου και της βιομηχανίας σε Λονδίνο, Ρώμη και Παρίσι. Ο John Arbuthnot, γνωστός μαθηματικός, γιατρός και είρων της εποχής του, υποστήριξε την ανάγκη επιβολής των κανόνων της Πολιτικής Αριθμητικής γύρω στα 1710 γράφοντας χαρακτηριστικά :

"...all the visible works of God Almighty are made in number, weight, and measure; therefore to consider them we ought to understand arithmetic, geometry, and statics."

Η λέξη "**Στατιστική**" αποδίδεται στο γερμανό πολυμαθή Gottfried Achenwall (1719-1772). Φυσικά προήλθε από τη λέξη status και τη γερμανική Πολιτική Αριθμητική. Ο πρώτος Άγγλος συγγραφέας που υιοθετεί τον όρο είναι ο John Sinclair (1754-1835), ο οποίος σε μια σειρά τευχιδίων δημοσιευμένα μεταξύ των ετών 1791 και 1799, παρουσιάζει μια εκτενή επισκόπηση της Σκωτίας βασιζόμενος σε στοιχεία των κοινοτικών συμβουλίων. Ο Sinclair παρατηρεί στον πρόλογό του, ότι ενώ η *στατιστική* στη Γερμανία έχει να κάνει με την πολιτική δύναμη του κράτους, εκείνος θεώρησε πιο ενδιαφέρον στο να καταγράψει :

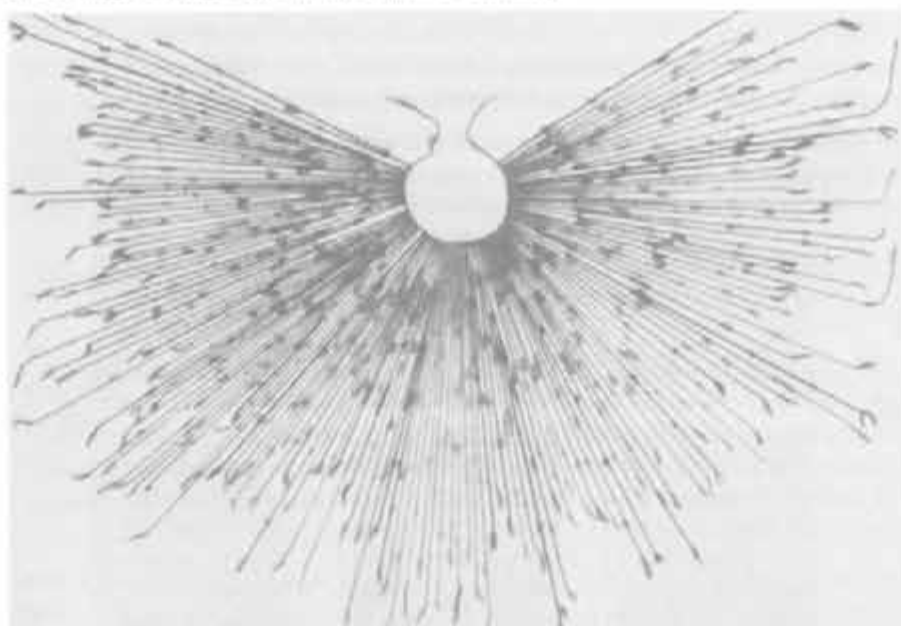
"...the quantum of happiness enjoyed by the inhabitants and the means of its improvement; yet as I thought a new word (statistics) might attract more public attention, I resolved to adopt it."

Στην εγκυκλοπαίδεια *Britannica* η πρώτη αναφορά στο λήμμα «στατιστική» εμφανίζεται το 1797. Η πρώτη απογραφή πληθυσμού στην Αγγλία πραγματοποιήθηκε το 1801 και από τότε, επαναλαμβάνεται συστηματικά

κάθε δέκα χρόνια. Αλλά και στο *Σύνταγμα* των Ηνωμένων Πολιτειών της Αμερικής (1787) γίνεται αναφορά στην ανάγκη ύπαρξης απογραφικών διαδικασιών :

“...The actual Enumeration shall be made within three years after the first meeting of the Congress of the United States, and within every subsequent term of ten years, in such manner as they shall by law direct.”

(1ο άρθρο, 2η παράγραφος). Η πρώτη απογραφή εδώ, χρονολογείται από το 1790 (επαναλαμβανόμενη επίσης ανά δεκαετία).



Εικόνα 1.3 *Κιούρι* : το σύστημα με το οποίο οι *Ινκας* κατέγραφαν γεννήσεις και θανάτους.

Ξεκινά πια η εποχή (τέλη του 18ου αιώνα), που για τη στατιστική συμβαίνει αυτό που περιγράφεται ως ταυτόχρονη οριζόντια και κάθετη ανάπτυξη. Οριζόντια υπό την έννοια της εξάπλωσης των μεθόδων σε άλλους επιστημονικούς κλάδους, από την αστρονομία και τη γεωδαισία μέχρι τη ψυχολογία, τη βιολογία και τις κοινωνικές επιστήμες. Κάθετη υπό την έννοια ότι η αναγνώριση του ρόλου της πιθανότητας άνοιξε το δρόμο στα πιθανοθεωρητικά μοντέλα μέτρησης και οδήγησε τελικά στο ξεκίνημα της στατιστικής συμπερασματολογίας.

Η ανάπτυξη μεθόδων που επιτρέπουν την εξαγωγή συμπερασμάτων από τα δεδομένα ξεκινά με τις εργασίες των Francis Galton (1822-1911) και

Karl Pearson (1857-1936) στα τέλη του 19ου αιώνα. Η πρώτη αυτή περίοδος χαρακτηρίζεται ακριβώς από αυτή την προαναφερόμενη αλλαγή ως προς τη στατιστική και την αναγνώριση της σημασίας της για τις άλλες επιστήμες. Επιπρόσθετα μεγάλες πρόοδοι συντελούνται στην επιστήμη. Μεταξύ των εφευρεθέντων εργαλείων σημειώνουμε αυτά της τυπικής απόκλισης, του συντελεστή συσχέτισης και του χ^2 test.

Η δεύτερη περίοδος ανάπτυξης της στατιστικής θεωρίας ξεκινά περί το 1915 και συνίσταται κυρίως στη δουλειά του R.A. Fisher (1890-1962) και των διαδόχων του. Στη συμβολή τους περιλαμβάνονται οι τεχνικές για μικρά δείγματα, η ανακάλυψη των πιο πολλών δειγματικών κατανομών, η διατύπωση των αρχών στον έλεγχο υποθέσεων, η ανάλυση διασποράς καθώς επίσης και η εισαγωγή κριτηρίων για την επιλογή μεταξύ των διαφόρων εκτιμητών μιας εκ των παραμέτρων του πληθυσμού.



Η δημοσίευση γύρω στα 1928 μιας σειράς κοινών εργασιών από τους Jerzy Neyman και Egon Pearson οριοθετεί την αρχή της τρίτης περιόδου. Στις εργασίες αυτές εισάγονται σημαντικές στατιστικές έννοιες όπως, το σφάλμα τύπου II, η ισχύς του test και τα διαστήματα εμπιστοσύνης. Ακριβώς αυτή την περίοδο, οι βιομηχανίες ξεκινούν μια γενικευμένη εφαρμογή των στατιστικών τεχνικών, και κυρίως αυτών που έχουν να κάνουν με τον ποιοτικό έλεγχο. Μεγάλο ενδιαφέρον αναπτύσσεται για τις στατιστικές έρευνες σε συνδυασμό με τη θεωρία και τις εφαρμογές της δειγματοληψίας.

Η τέταρτη περίοδος ξεκινά το 1933 με την εργασία του Abraham Wald για τη θεωρία αποφάσεων. Σ' αυτή, η Στατιστική θεωρείται σαν ένα

παιγνίδι με τη φύση. Πρόκειται για μια πολύ γενική θεωρία, η οποία, αν και οδήγησε σε δύσκολα μαθηματικά, υιοθετήθηκε από ένα μεγάλο αριθμό σύγχρονων στατιστικών.

Η εφεύρεση των ηλεκτρονικών υπολογιστών στα 1940, μαζί με την απίστευτη αύξηση της υπολογιστικής δυνατότητας που σημειώθηκε χάρη σ' αυτούς, είχε προφανή αντίκτυπο και στη στατιστική: οι υπολογιστές κατέστησαν δυνατή την εφαρμογή τεχνικών που θα ήταν αδιανόητες χωρίς την ύπαρξή τους, και μάλιστα πάνω σε τεραστίων όγκων δεδομένων. Ο όγκος των υπολογισμών που απαιτείται από τις τεχνικές αυτές, σε συνδυασμό με το μέγεθος των δεδομένων που έπρεπε να εφαρμοστούν, τις καθιστούσε πρακτικά άχρηστες πριν την εμφάνιση των υπολογιστών.

Η στατιστική θεωρία και εφαρμογή συνεχίζει να είναι σήμερα μια εξαιρετικά σημαντική περιοχή έρευνας. Στο κάτω-κάτω καμία σημαντική ανακάλυψη της στατιστικής επιστήμης δεν έγινε νωρίτερα από το τέλος του 19ου αιώνα.

Λειτουργίες της Στατιστικής

Στις μέρες μας, η ανάγκη της καταγραφής οδήγησε στη δημιουργία οργανωμένων κρατικών στατιστικών υπηρεσιών, υπεύθυνων για τις μετρήσεις του κράτους, καθώς επίσης και αρχείων για την αποθήκευση όλων αυτών των στοιχείων, οργανωμένων σε μορφή συγκρίσιμη (γεωγραφική και θεματική). Παράλληλα, λόγω της χρήσης των επιστημών σαν εργαλείων ανάλυσης και αισθητοποίησης του περιβάλλοντος κόσμου, προέκυψε η ανάγκη θεσμοθέτησης κοινά παραδεκτών μεθόδων συλλογής πληροφοριών και μονάδων μέτρησης, καθώς επίσης και η ανάγκη παραγωγής επιστημονικών τεχνικών εξαγωγής συμπερασμάτων από τις καταγραφόμενες μετρήσεις (γνωστές ως **δεδομένα**¹). Από την άλλη μεριά, η ανάπτυξη της τεχνολογίας των υπολογιστικών μηχανών έφερε μια έκρηξη στη συγκέντρωση πληροφοριακού υλικού και τροφοδότησε την ανάγκη αξιολόγησης και χρήσης του. Η πληροφορία, ως κατεχόμενο προϊόν και μέσο άσκησης πολιτικής, υποκατέστησε στις μέρες μας σε μεγάλο βαθμό τα υλικά προϊόντα, οδηγώντας σε μια νέα μορφή συγκέντρωσης δύναμης.

Οι μέθοδοι της **περιγραφικής στατιστικής** αποτελούν το επιστημονικό εργαλείο για τη συγκέντρωση, ταξινόμηση και παρουσίαση των πρωτογε-

¹ Δεδομένα : συγκεκριμένες πληροφορίες (στοιχεία, γεγονότα) οι οποίες μπορούν να χρησιμοποιηθούν για περαιτέρω σκέψεις και αποφάσεις.

νών δεδομένων σε εύληπτη μορφή, ενώ, τα εργαλεία της **στατιστικής συμπερασματολογίας**, καθιστούν δυνατή τη προσέγγιση χαρακτηριστικών του συνόλου των δεδομένων από τη μελέτη ενός (μικρού) υποσυνόλου τους (επαγωγική επιστήμη).

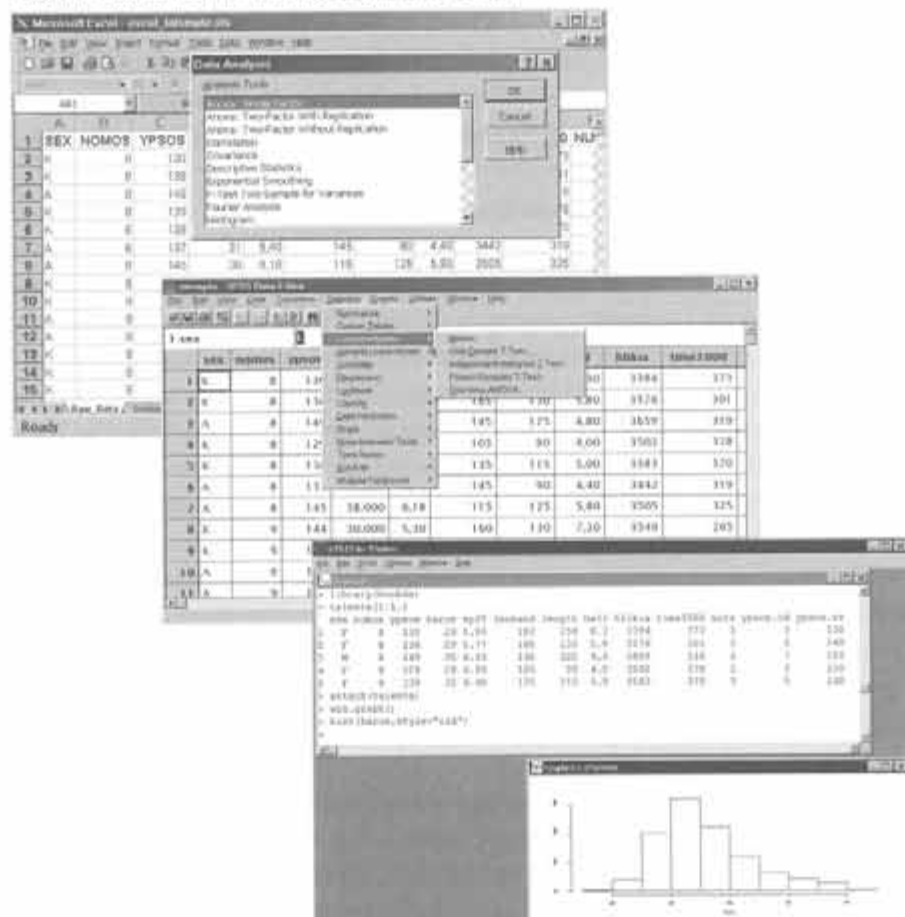
Δυστυχώς, ολοένα και πιο συχνά παρατηρείται το φαινόμενο της λαθεμένης χρήσης των στατιστικών μεθόδων στην ανάλυση και την παρουσίαση στοιχείων και αποτελεσμάτων. Το γεγονός αυτό, οφείλεται σε μεγάλο βαθμό στη σύγχυση ερευνητών ότι, αφού χρησιμοποιούν δεδομένα και χειρίζονται κάποιο από τα πολλά στατιστικά λογισμικά, κατέχουν και τη Στατιστική. Η Στατιστική όμως, δεν είναι ένα σύνολο συνταγών για την ανάλυση των καταγραμμένων δεδομένων. Δεν αρκεί να έχουμε δεδομένα και κάποιο λογισμικό για να κάνουμε την πρέπουσα επεξεργασία. Αποτελεί μεγάλη πλάνη η άποψη ότι Στατιστική δεν είναι παρά χειρισμός προγραμμάτων, πινακοποίηση αποτελεσμάτων και παρουσίαση δεδομένων με όμορφα γραφήματα.

Η Στατιστική, απαιτεί πρώτα απ' όλα την καλή γνώση του παρατηρούμενου φαινομένου, την οργάνωση της παρατήρησης και τη σωστή καταγραφή των δεδομένων. Απαιτεί επιπλέον, θεωρητική γνώση ώστε να επιλεγεί η πλέον κατάλληλη μέθοδος για τη στατιστική ανάλυση που επιδιώκουμε. Απαιτεί τέλος, το σχολαστικό έλεγχο της ικανοποίησης των απαραίτητων προϋποθέσεων για την εφαρμογή των διάφορων στατιστικών τεχνικών και μεθόδων. Αν και η Στατιστική εισάγεται στο αναλυτικό πρόγραμμα εκπαίδευσης της χώρας μας στην 3η Δημοτικού, οι γνώσεις στατιστικής που διδάσκονται (ακόμη και στο πανεπιστήμιο) αποδεικνύονται ελλιπείς. Βασικός λόγος είναι, ότι οι γνώσεις αυτές έχουν ως κύριο στόχο την ανάγνωση στατιστικών αποτελεσμάτων κι όχι την ανάλυση δεδομένων σε ερευνητικό επίπεδο.

Σκοπός του βιβλίου

Στο βιβλίο αυτό περιγράφουμε μεθόδους επεξεργασίας και ελέγχου υποθέσεων για διάφορα σύνολα δεδομένων, καθώς και τους τρόπους οργάνωσης της παρατήρησης και της καταγραφής που απαιτεί η κάθε μέθοδος. Οι μέθοδοι αυτές εφαρμόζονται σε δεδομένα, είτε πραγματικά, είτε φανταστικά, είτε γνωστά από τη διεθνή βιβλιογραφία, με τη βοήθεια τριών αντιπροσωπευτικών προγραμμάτων, που το καθένα τους έχει σχεδιαστεί για την αντιμετώπιση διαφορετικού επιπέδου προβλημάτων. Τα προγράμματα αυτά

είναι το EXCEL, το S.P.S.S. και το S-Plus.



Εικόνα 1.4 Οι αρχικές οθόνες των λογισμικών Excel, S.P.S.S. και S-Plus.

Το EXCEL αποτελεί εξελιγμένη μορφή προγράμματος «λογιστικού φύλλου» που στοχεύει στην καταγραφή δεδομένων στη μορφή πίνακα διπλής εισόδου σύμφωνα με το σχήμα:

Γραμμή-Υποκείμενο(περίπτωση), Στήλη-Μεταβλητή

Με τη χρήση μακροεντολών και ενσωματωμένων προγραμμάτων δίνει την δυνατότητα υπολογισμού στατιστικών μέτρων και σχεδιασμού γραφικών παραστάσεων, μετατροπών και αριθμητικών υπολογισμών.

Το S.P.S.S. είναι ένα από τα πιο εξελιγμένα υπολογιστικά προγράμματα Στατιστικής που προσπαθεί να καλύψει το σύνολο των «γνωστότερων» στατιστικών τεχνικών. Ιδιαίτερα στις εκδόσεις του σε γραφικό περιβάλλον,

δίνει τη δυνατότητα γραφικής επεξεργασίας και αναδραστικής λειτουργίας με πολύ μεγάλη επιτυχία.

Το **S-Plus** είναι ένα υπολογιστικό πρόγραμμα Στατιστικής με μεγάλη δυνατότητα προγραμματισμού διαδικασιών. Με τον τρόπο αυτό, προσφέρει στον έμπειρο χρήστη υψηλούς βαθμούς ελευθερίας στον χειρισμό συνόλων δεδομένων.

Στο βιβλίο, γίνεται προσπάθεια να περιγραφεί η κάθε στατιστική διαδικασία και η χρήση των παραπάνω στατιστικών προγραμμάτων σε αυτή. Οι στατιστικές διαδικασίες παρουσιάζονται όχι με τις τεχνικές τους λεπτομέρειες αλλά με τη λειτουργία και τα όρια τους. Για να επιλέξει κάποιος την κατάλληλη διαδικασία πρέπει να κινηθεί επιλέγοντας εκείνη που προσαρμόζεται στα δεδομένα του και όχι να προσαρμόσει τα δεδομένα του σε κάποια γνωστή διαδικασία. Σ' αυτό το εισαγωγικό κεφάλαιο, προσπαθούμε να σχολιάσουμε και να περιγράψουμε κριτήρια για την επιλογή των κατάλληλων στατιστικών τεχνικών. Παράλληλα, είναι απαραίτητο να γνωρίζει ο χρήστης τα όρια και την ερμηνεία των αποτελεσμάτων κάθε στατιστικής τεχνικής. Κανένα υπολογιστικό πρόγραμμα δεν μπορεί να δώσει απαντήσεις για ερωτήματα για τα οποία δεν είναι σχεδιασμένο.

1.2 Επιστημονική Έρευνα και Στατιστική

Το γεγονός ότι η διάδοση των επιστημονικών ερευνών αποτέλεσε έναν από τους βασικότερους λόγους εξάπλωσης της Στατιστικής, έχει ήδη επισημανθεί. Στις μέρες μας, είναι γενικά παραδεκτό ότι **όλες οι ερευνητικές προσπάθειες οι οποίες παράγουν δεδομένα επιδέχονται στατιστική επεξεργασία**. Η επεξεργασία αυτή διακρίνεται σε δύο (κύριες) φάσεις.

- ☞ *Σκοπός της πρώτης φάσης*, κι αρκετές φορές δυστυχώς μοναδικής, είναι η *άθροιση και σύνοψη των δεδομένων*. Αποτελεί το σκοπό της **Περιγραφικής Στατιστικής** και γίνεται με τη χρήση πινάκων (συχνοτήτων, διπλής εισόδου), γραφημάτων (ραβδογράμματα, θηκογράμματα, διαγράμματα διασποράς, κλπ), και στατιστικών αριθμών (μέτρα κεντρικής τάσης και μεταβλητότητας).
- ☞ *Στη δεύτερη φάση*, με τη βοήθεια της **Στατιστικής Συμπερασματολογίας**, γνωστής και ως **Επαγωγική ή Μαθηματική Στατιστική**, γίνεται προσπάθεια να *εξαχθούν από τα δεδομένα νόμοι, κανόνες και συμπεράσματα των οποίων η ισχύς να ξεπερνά το επίπεδο των παρατη-*

ρήσεων που έχουμε. Το βασικό εργαλείο εδώ είναι το μαθηματικό μοντέλο, η όλη δε διαδικασία βασίζεται στη **Θεωρία Πιθανοτήτων**.

Τα μαθηματικά μοντέλα αποτελούν σήμερα την πιο διαδεδομένη μέθοδο για τη μελέτη ενός φυσικού ή κοινωνικού φαινομένου (συστήματος). Το μοντέλο δεν είναι παρά μια προσομοίωση της πραγματικότητας, όπου οι συσχετισμοί μεταξύ των πραγματικών στοιχείων έχουν αντικατασταθεί από ανάλογους συσχετισμούς ανάμεσα σε μαθηματικές οντότητες. Συχνά όμως η πολυπλοκότητα των μελετούμενων συστημάτων μας αναγκάζει να προβούμε σε απλοποιήσεις και παραδοχές. Η βασική μας απαίτηση είναι το μαθηματικό και πραγματικό σύστημα να είναι «ισόμορφα» σε όλα τα σημαντικά, για το πρόβλημα που μελετάμε, θέματα. Η επιτυχία στην κατασκευή ενός μαθηματικού μοντέλου, έγκειται στην κατάλληλη χρήση της απλοποίησης και της παραδοχής, στο σωστό χρόνο και στο σωστό στάδιο, έτσι ώστε να περιέχει όσο το δυνατόν περισσότερη από την «πραγματικότητα» και, κατά συνέπεια, να δημιουργεί ένα εξυπηρετικό μαθηματικό υποκατάστατο του συστήματος, στο οποίο να μπορούμε να πειραματιστούμε.

Η κατασκευή ενός μοντέλου στηρίζεται αρχικά στην παρατήρηση, την εμπειρία και τη διαίσθηση που οδηγούν στη διατύπωση των θεωριών τις οποίες αυτό (το μοντέλο) παριστά και περιγράφει. Μετά τον έλεγχο τους, ακολουθεί η αναπροσαρμογή τους, η ανατροφοδότησή τους με καινούρια στοιχεία και η επαναδιατύπωσή τους, σε μια διαρκή κίνηση που οδηγεί συνήθως σε στοχαστικά μοντέλα ερμηνείας και λιγότερο σε μοντέλα ντετερμινιστικού χαρακτήρα. Εξάλλου, είναι παραδεκτό πια, ότι ο χρόνος και η αλλαγή αποτελούν κύρια χαρακτηριστικά όλων των συστημάτων, κι ότι επιπλέον, στα περισσότερα από αυτά ενυπάρχει τυχαιότητα.

Συνήθη αναφερόμενα παραδείγματα στοχαστικών φαινομένων, είναι το ριζίμο ενός ζαριού, τα διάφορα παιγνίδια τύχης, το φύλο ενός νεογέννητου παιδιού, η καθημερινή θερμοκρασία, η σοδειά ενός χωραφιού, ο χρόνος ζωής ενός λαμπτήρα, το μήκος των φύλλων ενός καλλωπιστικού θάμνου, κλπ. Με το μοντέλο που κατασκευάζουμε για το καθένα από αυτά, διατυπώνουμε θεωρίες που μπορεί να αφορούν την πιθανότητα εμφάνισης ενός γεγονότος, την εξάρτηση μεταξύ διαφόρων από τους εμπλεκόμενους παράγοντες, την τάση τους, σχέσεις ομοιότητας ή διαφοράς, κλπ.

Κάθε φαινόμενο εξελίσσεται (συμβαίνει) όχι αφηρημένα, αλλά σε

υποκείμενα, αυτόνομες δηλαδή μονάδες παρατήρησης, το σύνολο των οποίων ορίζει τον **πληθυσμό αναφοράς**. Το φαινόμενο αναλύεται σε επιμέρους **μετρήσιμα** χαρακτηριστικά, **τις μεταβλητές**, στις οποίες αντιστοιχούμε **τιμές**. Η αντιστοίχιση αυτή, ονομάζεται **μέτρηση**, και γίνεται με την χρήση εργαλείων γενικού χαρακτήρα, είτε «**αντικειμενικών**» (επιστημονικά όργανα) είτε «**αποκειμενικών**» (ερωτηματολόγια). Σημειώστε επίσης ότι, οι επιμέρους ιδιότητες των μεταβλητών προσδιορίζονται από την κλίμακα μέτρησής της.

Όμως, στην κατασκευή ενός μοντέλου, η μελέτη που θα κάνουμε είναι σίγουρο ότι είναι μερική. Για παράδειγμα, δεν είναι δυνατό να μελετήσουμε όλους τους ανθρώπους για να κατασκευάσουμε ένα μοντέλο –ελέγχουμε μια θεωρία– για το «ύψος» τους ή στιδήποτε άλλο. Η **ΣΤΑΤΙΣΤΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ**, έρχεται ακριβώς για να επιβεβαιώσει ή να διαψεύσει τη θεωρία (υπόθεση) την οποία διατυπώσαμε με το μοντέλο, στη βάση των όσων υποδεικνύει η τεκμηριωμένη εμπειρία από το υποσύνολο του πληθυσμού που μελετήσαμε (**δείγμα**).

Συνοψίζοντας, η συνηθισμένη προσέγγιση σ' ένα πρόβλημα που απαιτεί στατιστική ανάλυση είναι στην κατεύθυνση της οργάνωσης της παρατήρησης, της καταγραφής των δεδομένων (αποτελεσμάτων) και της διατύπωσης της υπόθεσης (θεωρίας). Συστατικά αυτής της προσέγγισης είναι ο **πληθυσμός**, το **υποκείμενο**, το **δείγμα**, η **μεταβλητή**, η **μέτρηση**, τα **όργανα μέτρησης** και το **μοντέλο**.

Πληθυσμός

Με τον όρο **πληθυσμός** εννοούμε ένα σύνολο ατόμων ή αντικειμένων ή οποιονδήποτε άλλων οντοτήτων, για τα οποία ενδιαφερόμαστε να βγάλουμε συμπεράσματα σε σχέση με κάποιες ιδιότητες που αφορούν τα στοιχεία του. Ο πληθυσμός πρέπει να είναι καλά ορισμένος, να περιγράφεται δηλαδή με κάποιες ιδιότητες των στοιχείων που τον αποτελούν, σε τρόπο ώστε να μπορεί οποιοσδήποτε να αποφανθεί μονοσήμαντα αν κάποιο στοιχείο είναι μέλος του (: ικανοποιεί τις ιδιότητες) ή όχι (: δεν ικανοποιεί τις ιδιότητες). Έτσι για παράδειγμα, το σύνολο όλων των Ελλήνων, οι μαθητές της έκτης τάξης, τα νοικοκυριά της Θεσσαλονίκης, τα αγόρια ηλικίας 12-16 ετών, το σύνολο των ηλεκτρονικών μηνυμάτων που φθάνουν στον κεντρικό υπολογιστή του ΑΠΘ, κλπ είναι πληθυσμοί.

3.1 Περιγραφική Στατιστική

Η ύπαρξη δύο ή περισσότερων μεταβλητών σε μια έρευνα, οδηγεί εύλογα στην αναζήτηση της (πιθανής) μεταξύ τους σχέσης. Αρχικά σ' αυτή την παράγραφο, πρόκειται να αναλυθούν οι τεχνικές οι οποίες χρησιμοποιούνται για την παρουσίαση της συμπεριφοράς των τιμών μιας μεταβλητής σε σχέση με τις τιμές κάποιας άλλης. Στη συνέχεια θα απαντηθούν τα ερωτήματα που αφορούν τη μεταξύ τους σχέση: μήπως κάποιες συγκεκριμένες τιμές της μιας μεταβλητής εμφανίζονται πιο συχνά όταν εμφανίζονται συγκεκριμένες τιμές της άλλης;

Η επιλογή της στατιστικής τεχνικής για την παραπάνω ανάλυση εξαρτάται αποκλειστικά από τη διάκριση των μεταβλητών σε ποιοτικές ή ποσοτικές. Υπάρχει σαφής διαφοροποίηση των εργαλείων που είναι διαθέσιμα στην κάθε περίπτωση.

Θα πρέπει να αναφερθεί ίσως εδώ, ότι οι μεταβλητές αυτές, που πρέπει να είναι ή και οι δύο ποιοτικές ή και οι δύο ποσοτικές, εμφανίζονται στα δεδομένα με την εξής δομή:

A/A Παρατήρησης	X μεταβλητή	Y μεταβλητή
1	x_1	y_1
2	x_2	y_2
3	x_3	y_3
\vdots	\vdots	\vdots
n	x_n	y_n

Στον πίνακα αυτό, X και Y είναι τα ονόματα των δύο μεταβλητών ενώ x_i , y_i οι τιμές τους για την i -παρατήρηση. Όταν οι μεταβλητές είναι ποσοτικές, τα x_i , y_i ορίζουν ποσότητες (π.χ. ύψος και βάρος). Αν όμως οι μεταβλητές X και Y είναι ποιοτικές, x_i , y_i είναι τα ονόματα των κατηγοριών στις οποίες αντιστοιχούν (π.χ. φύλο και είδος τηλεοπτικής εκπομπής). Σημειώστε τέλος ότι, προκειμένου να μελετήσουμε τη σχέση μεταξύ δύο μεταβλητών X και Y, θα πρέπει οπωσδήποτε το ζευγάρι των τιμών (x_i , y_i) να προέρχεται από το ίδιο αντικείμενο του δείγματος.

3.1.1 S.P.S.S.

Το SPSS ενσωματώνει τόσο γραφικές όσο και υπολογιστικές διαδικασίες για την αναζήτηση της σχέσης μεταξύ δύο μεταβλητών. Πριν εφαρμόσουμε κάποια από αυτές θα πρέπει να λάβουμε υπόψη τον χαρακτήρα των μεταβλητών, τη διάκρισή τους δηλαδή σε ποιοτικές και ποσοτικές.

Ποιοτικά χαρακτηριστικά

Οι μέθοδοι ταυτόχρονης παρουσίασης δύο τουλάχιστον ποιοτικών χαρακτηριστικών (μεταβλητών) περιορίζονται στους πίνακες συνάφειας και τις γραφικές παραστάσεις. Με τη διαδικασία 'Crosstabs' του S.P.S.S. μπορούμε να πετύχουμε όχι μόνο την άμεση κατασκευή τους, αλλά επιπλέον να προχωρήσουμε και στην αναζήτηση της έντασης και φύσης της (πιθανής) σχέσης τους.

⇒ Από τη βασική ράβδο προτιμήσεων του λογισμικού επιλέγοντας
 Statistics
 Summarize
 Crosstabs...

εμφανίζεται το πλαίσιο διαλόγου (οθόνη) της εικόνας 3.1.

Εικόνα 3.1

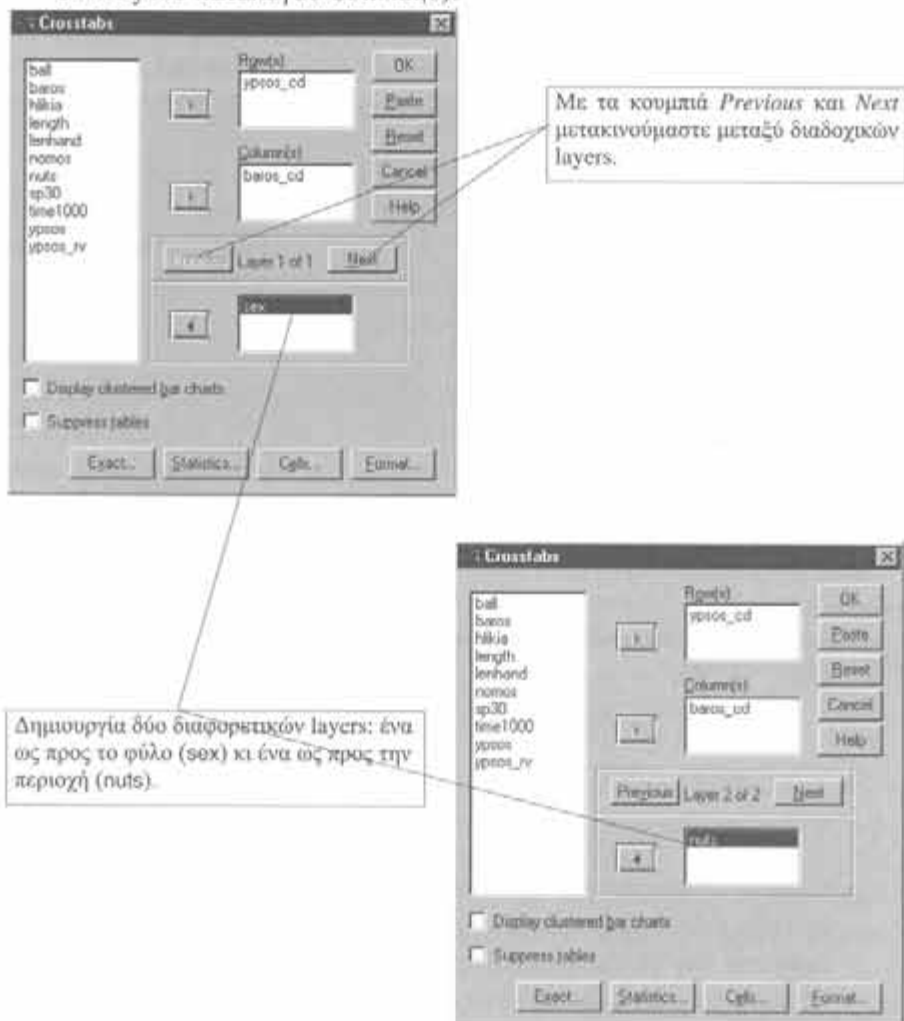
Το πλαίσιο διαλόγου στο SPSS για την κατασκευή ενός πίνακα συνάφειας.



⇒ Διαλέγουμε την (ποιοτική) μεταβλητή, τις κατηγορίες της οποίας θέλουμε να έχουμε στις γραμμές του πίνακα συνάφειας και τη μετακινούμε στο παράθυρο *Row(s)*. Διαλέγουμε κάποια άλλη (ποιοτική) μεταβλητή, τις κατηγορίες της οποίας θέλουμε να έχουμε στις στήλες του πίνακα συνάφειας και τη μετακινούμε στο παράθυρο *Column(s)*.

Σημειώστε ότι η διαδικασία μπορεί να χρησιμοποιηθεί και στην περίπτωση των ποσοτικών μεταβλητών των οποίων οι τιμές προηγουμένως έχουν κωδικοποιηθεί και αντιστοιχηθεί σε διαστήματα τιμών (βλέπε και παρατήρηση σελίδων 1-28, 1-29).

Φυσικά, μπορούμε να μετακινήσουμε περισσότερες από μια μεταβλητές τόσο στον κατάλογο *Row(s)* όσο και στον κατάλογο *Column(s)*. Το S.P.S.S. θα κατασκευάσει από ένα διαφορετικό πίνακα συνάφειας για κάθε δυνατό συνδυασμό των μεταβλητών του καταλόγου *Row(s)* με εκείνες του καταλόγου *Column(s)*.



Εικόνα 3.2 Πλαίσιο διαλόγου (οθόνη) για τη δημιουργία δύο layers.

⇒ Το S.P.S.S. παρέχει, με τη βοήθεια της επιλογής 'Layers', τη δυνατότητα να ορίσουμε μία ή περισσότερες ποιοτικές μεταβλητές σαν μεταβλητές ελέγχου των ζητούμενων πινάκων συνάφειας: αρκεί να τη μετακινήσουμε στον κατάλογο *Layers 1*. Το λογισμικό, για κάθε κατηγορία της πρώτης εκ των μεταβλητών αυτών θα κατασκευάσει όλους τους ζητούμενους στην προηγούμενη επιλογή πίνακες συνάφειας. Στη συνέχεια θα κάνει το ίδιο και για κάθε κατηγορία της δεύτερης ποιοτικής μεταβλητής που μετακινήσαμε στον κατάλογο *Layers 1*, κλπ.

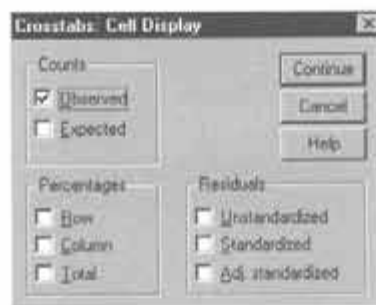
Προαιρετικά, μπορούμε να δημιουργήσουμε περισσότερους από έναν καταλόγους *Layers*. Τότε το S.P.S.S., θα κατασκευάσει όλους τους ζητούμενους πίνακες συνάφειας για κάθε δυνατό συνδυασμό κατηγοριών των μεταβλητών του καταλόγου *Layers 1*, με εκείνους των μεταβλητών του καταλόγου *Layers 2*.

Στην εικόνα 3.2 φαίνεται ο τρόπος δημιουργίας δύο διαφορετικών *Layers*. Θα δημιουργηθούν 6 (δύο από τις κατηγορίες του φύλου επί τρεις από τις κατηγορίες της περιοχής) πίνακες συνάφειας οι οποίοι θα έχουν για γραμμές τις κατηγορίες (διαστήματα) του ύψους και για στήλες τις κατηγορίες (διαστήματα) του βάρους.

⇒ Επιλέγουμε τις πληροφορίες που θέλουμε να εμφανίζονται σε κάθε κελί του πίνακα συνάφειας (εικόνα 3.3). Εξ ορισμού, το λογισμικό θα εμφανίσει μόνον τις παρατηρούμενες συχνότητες. Με την επιλογή 'Cells...' μπορούμε να εμπλουτίσουμε τις εμφανιζόμενες πληροφορίες με τις αναμενόμενες συχνότητες, τα ποσοστά των παρατηρήσεων στα κελιά της κάθε γραμμής ή στήλης, κλπ.

Εικόνα 3.3

Οι πληροφορίες που μπορούμε να έχουμε διαθέσιμες σ' έναν πίνακα συνάφειας.



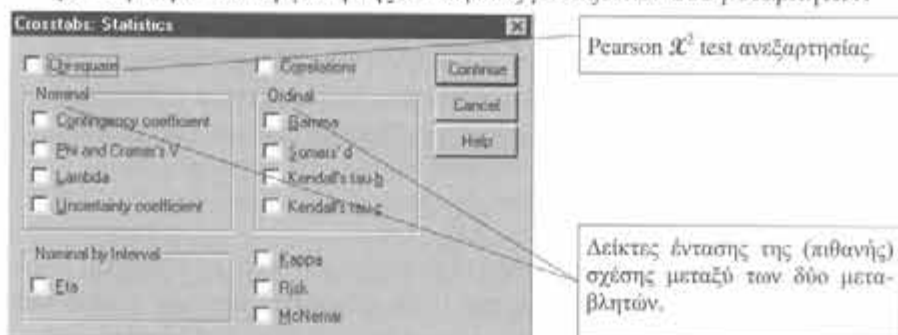
- **Counts:** Παρατηρούμενες συχνότητες (Observed) είναι οι μετρήσεις σε κάθε κελί, ενώ αναμενόμενες (Expected) είναι ο αριθμός των περιπτώσεων που θα ήταν σε κάθε κελί αν οι μεταβλητές που ορίζουν τις γραμμές και στήλες του πίνακα είναι στατιστικά ανεξάρτητες.
- **Percentages:** Τα ποσοστά των γραμμών/στηλών αθροίζουν στο 100%

κατά μήκος της κάθε γραμμής/στήλης του πίνακα συνάφειας, ενώ τα συνολικά ποσοστά αθροίζουν στο 100% μέσα σ' όλα τα κελιά του πίνακα.

- **Residuals:** Τα υπόλοιπα δεν είναι παρά οι διαφορές μεταξύ παρατηρούμενων και αναμενόμενων τιμών του κάθε κελιού.

Μπορούμε επιπλέον, ενεργοποιώντας την επιλογή 'Suppress tables', να ακυρώσουμε την εμφάνιση του πίνακα συνάφειας.

⇒ Επιλέγουμε τα στατιστικά μέτρα (Statistics...) που θέλουμε να εμφανιστούν μαζί με τον πίνακα συνάφειας. Μπορούμε όχι μόνο να πραγματοποιήσουμε έλεγχο ανεξαρτησίας (βλέπε §3.3), αλλά και να αναζητήσουμε το βαθμό και τη φύση της συνάφειας μεταξύ των δύο μεταβλητών.



Εικόνα 3.4 Τα στατιστικά μέτρα της διαδικασίας Crosstabs.

- **Chi-square:** Αφορά την πραγματοποίηση ελέγχου ανεξαρτησίας των δύο μεταβλητών. Περισσότερες πληροφορίες δίνονται στην §3.3.
- **Correlations:** Αναφέρεται στον υπολογισμό του συντελεστή συσχέτισης για την καταγραφή της φύσης και του βαθμού συνάφειας δύο μεταβλητών (περισσότερες πληροφορίες δίνονται στη σελίδα 3-59):
 - ✓ Ο συντελεστής συσχέτισης του Pearson μετρά την ένταση της γραμμικής σχέσης δύο ποσοτικών μεταβλητών ορίζοντας παράλληλα και τη φύση της.
 - ✓ Ο συντελεστής συσχέτισης του Spearman υπολογίζεται σαν ένα μέτρο της έντασης της σχέσης μεταξύ δύο μεταβλητών στην περίπτωση που και οι δύο είναι διάταξης (ordinal).

Το S.P.S.S. μαζί με τον υπολογισμό του συντελεστή συσχέτισης πραγματοποιεί και στατιστικό έλεγχο για το αν αυτός διαφέρει σημαντικά από το μηδέν (παράδειγμα 3.43).

- **Nominal:** Στην περίπτωση που και οι δύο ποιοτικές μεταβλητές είναι κατηγορίας (nominal) δεν μπορούμε βέβαια να μιλάμε για τη φύση

της σχέσης μεταξύ των δύο μεταβλητών, μόνο η ένταση έχει νόημα. Στην περίπτωση αυτή, υπάρχουν μέτρα που βασίζονται στο χ^2 στατιστικό (συντελεστής συνάφειας, phi, V του Cramer), αλλά και μέτρα αναλογικής μείωσης του σφάλματος πρόβλεψης (Lambda, συντελεστής αβεβαιότητας). Η πρώτη κατηγορία των μέτρων αυτών δύσκολα ερμηνεύεται, η πρακτική τους σημασία περιορίζεται στη σύγκριση του βαθμού συνάφειας μεταξύ διαφορετικών πινάκων (για παράδειγμα όταν έχουμε χρησιμοποιήσει στην στατιστική ανάλυση και μεταβλητές ελέγχου). Αντίθετα η σημασία των μέτρων της δεύτερης κατηγορίας είναι ξεκάθαρη: ψάχνουν να βρουν το βαθμό στον οποίο η γνώση των τιμών της μιας μεταβλητής βελτιώνει τις προβλέψεις μας για την άλλη. Για την εφαρμογή τους είναι απαραίτητο να θεωρήσουμε τη μια μεταβλητή ως ανεξάρτητη και την άλλη ως εξαρτημένη, χωρίς βέβαια ο χαρακτηρισμός αυτός να ορίζει σχέση αιτίας και αποτελέσματος.

- ✓ Ο συντελεστής συνάφειας C του Pearson ορίζεται από τη σχέση

$$C = \sqrt{\frac{X^2}{X^2 + n}}$$

και παίρνει τιμές στο διάστημα 0 (ανεξάρτητες μεταβλητές) μέχρι κάποια τιμή μικρότερη του 1 που εξαρτάται από τον αριθμό των γραμμών και στηλών του πίνακα (π.χ. σε έναν 4x4 πίνακα η μεγαλύτερη τιμή είναι 0.87).

- ✓ Ο συντελεστής phi (ϕ) ορίζεται μόνο σε 2x2 πίνακες συνάφειας από τη σχέση

$$\phi = \sqrt{X^2 / n}$$

και παίρνει τιμές στο διάστημα 0 (ανεξάρτητες μεταβλητές) μέχρι 1 (απόλυτη συνάφεια).

- ✓ Ο συντελεστής V του Cramer είναι μια επέκταση του συντελεστή phi σε πίνακες συνάφειας μεγαλύτερης διάστασης. Ορίζεται από τη σχέση

$$V = \sqrt{\frac{X^2}{n * \min\{r-1, c-1\}}}$$

και παίρνει τιμές στο διάστημα 0 (ανεξάρτητες μεταβλητές) μέχρι 1 (απόλυτη συνάφεια).

- ✓ Οι Lambda (λ) συντελεστές στηρίζονται στην άποψη ότι αν ξέρουμε τις τιμές (κατηγορίες) για τη μια μεταβλητή (ανεξάρτη-

τη) η καλύτερη πολιτική για να προβλέψουμε τις τιμές της άλλης (εξαρτημένη), είναι να επιλέξουμε εκείνη με τη μεγαλύτερη συχνότητα. Με τον τρόπο αυτό θα ελαχιστοποιήσουμε τον αριθμό των λαθεμένων προβλέψεων. Ο συντελεστής παίρνει τη μεγαλύτερη τιμή 1 όταν υπάρχει πλήρης σχέση μεταξύ των δύο μεταβλητών (οπότε αν γνωρίζουμε την τιμή της ανεξάρτητης μπορούμε να προβλέψουμε επακριβώς την τιμή της εξαρτημένης). Η μικρότερη τιμή του συντελεστή είναι το 0. Τότε οι συχνότητες των κατηγοριών της εξαρτημένης είναι ίδιες για κάθε τιμή της ανεξάρτητης και συνεπώς οι δύο μεταβλητές είναι ανεξάρτητες.

- ✓ Οι συντελεστές αβεβαιότητας έχουν την ίδια λογική με τους Lambda συντελεστές αλλά είναι γενικά πιο αξιόπιστοι γιατί χρησιμοποιούν για τον υπολογισμό τους όλες τις τιμές κι όχι μόνο τις επικρατούσες (που χρησιμοποιούν οι Lambda).
- **Ordinal:** Οι συντελεστές Lambda και αβεβαιότητας μπορούν να χρησιμοποιηθούν σαν μέτρα του βαθμού συνάφειας δύο μεταβλητών ανεξάρτητα από τη κλίμακα μέτρησής τους. Στην περίπτωση όμως μεταβλητών διάταξης, η υπάρχουσα διάταξη δεν επηρεάζει σε κανένα σημείο τους υπολογισμούς για την εύρεσή τους. Εναλλακτικά όμως στην περίπτωση αυτή (των μεταβλητών διάταξης δηλαδή), μπορούμε να χρησιμοποιήσουμε μέτρα τα οποία θα προσδιορίσουν όχι μόνο την ένταση, αλλά και τη φύση της συνάφειας (θετική ή αρνητική). Τέτοια μέτρα είναι ο συντελεστής gamma (γ), οι συντελεστές Tau-b (T_b) και Tau-c (T_c) του Kendall, καθώς επίσης και ο συντελεστής d του Somers.

Για τον υπολογισμό τους θα πρέπει να θεμελιώσουμε τις έννοιες της *συμφωνίας* (concordant), *ασυμφωνίας* (discordant) και *ισοπαλίας* (tied). Ένα ζευγάρι περιπτώσεων λέμε ότι είναι σε συμφωνία όταν οι τιμές και των δύο μεταβλητών της μιας περίπτωσης είναι μεγαλύτερες από τις αντίστοιχες τιμές των μεταβλητών της άλλης περίπτωσης. Αντίθετα, ένα ζευγάρι περιπτώσεων λέμε ότι είναι σε ασυμφωνία όταν η τιμή για τη μια μεταβλητή της μιας περίπτωσης είναι μεγαλύτερη από την αντίστοιχη τιμή της μεταβλητής της άλλης περίπτωσης, ενώ για τις τιμές της άλλης μεταβλητής ισχύει το αντίθετο. Τέλος, ισοπαλία έχουμε όταν οι δύο περιπτώσεις έχουν ίδιες τιμές για τουλάχιστον μία των μεταβλητών.

Γενικά, υπάρχουν πέντε δυνατά αποτελέσματα όταν συγκρίνουμε δύο

περιπτώσεις: είναι δυνατόν να έχουμε συμφωνία, ασυμφωνία, ισοπαλία στην πρώτη μεταβλητή, ισοπαλία στη δεύτερη μεταβλητή, ή τέλος ισοπαλία και στις δύο μεταβλητές.

Η ιδέα όλων των παραπάνω μέτρων κρύβεται πίσω από τη σύγκριση του πλήθους P των ζευγών που είναι σε συμφωνία με το πλήθος Q αυτών που είναι σε ασυμφωνία. Αν το P είναι πολύ μεγαλύτερο του Q τότε υπάρχει σχέση μεταξύ των δύο μεταβλητών και μάλιστα όταν αυξάνεται η μια αυξάνεται και η άλλη (θετική συσχέτιση). Όμοια όταν το Q είναι πολύ μεγαλύτερο του P έχουμε αρνητικά συσχετισμένες μεταβλητές. Αν όμως $P \approx Q$ τότε οι μεταβλητές είναι μάλλον ασυσχέτιστες.

Επειδή θέλουμε το μέτρο που θα κατασκευάσουμε να παίρνει τιμές στο διάστημα από -1 (τέλεια αρνητική συσχέτιση) μέχρι 1 (τέλεια θετική συσχέτιση) η διαφορά $P-Q$ θα πρέπει να τυποποιηθεί. Ο τρόπος με τον οποίο επιχειρείται κάτι τέτοιο είναι και η διαφορά των προαναφερθέντων μέτρων συνάφειας. Ο συντελεστής Tau-b είναι κατάλληλος για συμμετρικούς πίνακες, ενώ ο Tau-c για μη συμμετρικούς. Και στους δύο λαμβάνονται υπόψη οι ισοπαλίες, ενώ στον συντελεστή gamma όχι. Τέλος με τον συντελεστή d του Somers μπορούμε να κάνουμε υπολογισμούς θεωρώντας τη μια μεταβλητή εξαρτημένη και την άλλη ανεξάρτητη (αποτελεί μια επέκταση του συντελεστή gamma ώστε να λαμβάνονται υπόψη οι ισοπαλίες).

- **Nominal by Interval:** Στην περίπτωση που η μια μεταβλητή είναι κατηγορίας (nominal) και η άλλη διαστημική (interval) θα πρέπει να χρησιμοποιήσουμε τον συντελεστή eta (η). Πρόκειται για ένα δείκτη της ασυμφωνίας των μέσων τιμών της εξαρτημένης διαστημικής μεταβλητής μέσα στις στάθμες της ανεξάρτητης κατηγορίας. Παίρνει τιμές στο διάστημα από 0 (πλήρης συμφωνία) μέχρι 1 (πλήρης ασυμφωνία). Η ποσότητα η^2 αναφέρεται συχνά σαν *λόγος συσχέτισης* (correlation ratio) και εκφράζει το ποσοστό της διασποράς της εξαρτημένης μεταβλητής που εξηγείται από την ανεξάρτητη.
- **Kappa:** Εδώ έχουμε τις ίδιες κατηγορίες στις γραμμές και στις στήλες του πίνακα συνάφειας. Αν ζητήσουμε από δύο ανθρώπους να αποτιμήσουν το ίδιο αντικείμενο (σε μια μικρή κλίμακα διακριτών τιμών) εύλογα στη συνέχεια θα αναρωτηθούμε πόσο συχνά συμφωνούν. Ο συντελεστής kappa (κ) του Cohen μετρά το βαθμό συμφωνίας των δύο κριτών λαμβάνοντας υπόψη τον παράγοντα της τύχης. Παίρνει τιμές

στο διάστημα από -1 (πλήρης ασυμφωνία) μέχρι 1 (πλήρης ασυμφωνία) με την τιμή 0 να υποδεικνύει ότι η όποια ομοιότητα υπάρχει δεν διαφέρει από αυτή που περιμένουμε από την τύχη.

- **Risk:** Πρόκειται για ένα μέτρο της έντασης της σχέσης μεταξύ της παρουσίας κάποιου παράγοντα και της εμφάνισης ενός περιστατικού. Τιμή 1 του δείκτη υποδεικνύει ανυπαρξία της υποτιθέμενης σχέσης.
- **McNemar:** Το κριτήριο του McNemar χρησιμοποιείται όταν θέλουμε να ελέγξουμε τη μεταβολή στη συμπεριφορά των ατόμων ενός πληθυσμού πριν και μετά τη μεσολάβηση κάποιου γεγονότος. Πρόκειται για έναν πολύ γνωστό μη παραμετρικό έλεγχο που αναλύεται περισσότερο στην §3.2.1

Σημειώστε τέλος ότι, το λογισμικό μαζί με τον υπολογισμό του όποιου μέτρου συνάφειας ζητηθεί πραγματώνει και στατιστικό έλεγχο για το αν αυτό διαφέρει σημαντικά από το μηδέν (παράδειγμα 3.41).

⇒ Καθορίζουμε τον τρόπο εμφάνισης (*Format...*) του πίνακα συνάφειας. Ο πίνακας μπορεί να ταξινομηθεί (εικόνα 3.5) ως προς τη σειρά εμφάνισης (κατά αύξουσα ή φθίνουσα σειρά) των διαφορετικών κατηγοριών της μεταβλητής που ορίζει τις γραμμές.

Εικόνα 3.5

Καθορισμός του τρόπου εμφάνισης των κατηγοριών στις γραμμές του πίνακα συνάφειας.



⇒ Αποφασίζουμε αν επιθυμούμε να έχουμε και γραφική αναπαράσταση των αποτελεσμάτων (*Display clustered bar charts*). Εξ ορισμού θα κατασκευαστούν ομαδοποιημένα (clustered) ραβδογράμματα. Μπορούμε όμως να δημιουργήσουμε αποτελεσματικότερα γραφήματα ενεργοποιώντας την αντίστοιχη επιλογή του λογισμικού (εικόνα B.71).

Παράδειγμα 3.1. Θα χρησιμοποιήσουμε τα δεδομένα από το αρχείο *talssmple*. Συγκεκριμένα, το ενδιαφέρον μας αφορά την αναζήτηση σχέσης μεταξύ των μεταβλητών *ypsos* και *baros* τις οποίες κωδικοποιήσαμε στις μεταβλητές *ypsos_nh* και *baros_nh* αντίστοιχα (εικόνα 3.6).

Επιλέγοντας διαδοχικά

```
Statistics
  Summarize
    Crosstabs...
```

και δίνοντας στη συνέχεια