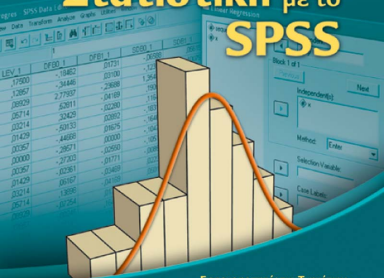


Δρ. Βασίλης Δαφέρμος
Τμήμα Πολιτικής Επιστήμης Πανεπιστημίου Κρήτης

Κοινωνική Στατιστική με το SPSS



Για τους φοιτητές των Τμημάτων :

- Πολιτικής Επιστήμης • Κοινωνιολογίας • Ψυχολογίας
- Παιδαγωγικών • Οικονομίας • Ιατρικής • Νοσηλευτικής • Βιολογίας
- Εφαρμοσμένων Μαθηματικών • Χημείας • Φαρμακευτικής • Φυσικής Αγωγής
- Και για τους νέους ερευνητές των Κοινωνικών Επιστημών



Το βιβλίο συνοδεύεται από CD

στο οποίο περιέχονται όλα τα αρχεία δεδομένων (Data Sets)



ΕΚΔΟΣΕΙΣ
ΖΗΤΗ

Κάθε γνήσιο αντίτυπο φέρει την υπογραφή του συγγραφέα

ISBN 960-431-960-4

© Copyright: Δαφέρμος Βασίλης, Εκδόσεις Ζήτη, Μάιος 2005, Θεσσαλονίκη

SPSS

SPSS Inc., <http://www.spss.com/>

Το παρόν έργο πνευματικής ιδιοκτησίας προστατεύεται κατά τις διατάξεις του Ελληνικού νόμου (Ν.2121/1993 όπως έχει τροποποιηθεί και ισχύει σήμερα) και τις διεθνείς συμβάσεις περί πνευματικής ιδιοκτησίας. Απαγορεύεται απολύτως η άνευ γραπτής άδειας του εκδότη και συγγραφέα κατά οποιοδήποτε τρόπο ή μέσο αντιγραφή, φωτοανατύπωση και εν γένει αναπαραγωγή, εκμίσθωση ή δανεισμός, μετάφραση, διασκευή, αναμετάδοση στο κοινό σε οποιαδήποτε μορφή (ηλεκτρονική, μηχανική ή άλλη) και η εν γένει εκμετάλλευση του συνόλου ή μέρους του έργου.



**Φωτοστοιχειοθεσία
Εκτύπωση**

Π. ΖΗΤΗ & ΣΙΑ ΟΕ

18ο χλμ Θεσ/νίκης-Περαίας

Τ.Θ. 4171 • Περαία Θεσσαλονίκης • Τ.Κ. 570 19

Τηλ.: 23920 72.222 (5 γραμ.) - Fax: 23920 72.229

e-mail: info@ziti.gr

Βιβλιοπωλείο

ΕΚΔΟΣΕΙΣ ΖΗΤΗ

Αρμενοπούλου 27 • 546 35 Θεσσαλονίκη

Τηλ. 2310 203.720, Fax 2310 211.305

e-mail: sales@ziti.gr

www.ziti.gr

Στα παιδιά μου

Πάρι,

Ευγενία,

Γιάννη,

Και στον ανιψιό μου

Νικόλα.

Ευχαριστίες

Ευχαριστώ θερμά τη Διευθύντρια της Βιβλιοθήκης του Μαθηματικού Τμήματος του Πανεπιστημίου Ιωαννίνων, κ. Μαρία Κοντογιάννη, για την καθοριστική, άμεση και πάντοτε πρόθυμη βοήθειά της στην εξεύρεση της ελληνικής και ξένης βιβλιογραφίας. Η κ. Κοντογιάννη, βέβαια δεν είναι μια συνηθισμένη περίπτωση βιβλιοθηκονόμου. Έχει άποψη αναφορικά με την οργάνωση των Μαθηματικών Βιβλιοθηκών, αναμφισβήτητες ικανότητες, σπάνιο ήθος και αξιοπρέπεια. Την ευχαριστώ βαθιά, και της εύχομαι υγεία και δύναμη στο δύσκολο έργο της εκεί στα Γιάννενα.

Ευχαριστώ επίσης, την υποψήφια διδάκτορα κ. Βικτώρια Βιβιλάκη για την παραχώρηση ερευνητικών δεδομένων για τις ανάγκες του 18^ο κεφαλαίου αυτού του βιβλίου.

Τέλος, θα ήθελα να ευχαριστήσω τον εκδοτικό Οίκο Ζήτη, για την άποψη εμφάνιση αυτής της έκδοσης.

Αναφορικά με το SPSS θα πρέπει να πω τα εξής:

- Το Πανεπιστήμιό μας, το Πανεπιστήμιό μας Κρήτης, είναι νόμιμος χρήστης του Στατιστικού Προγράμματος SPSS, με ελεύθερο αριθμό αντιγράφων (Free Copies). Αυτό σημαίνει ότι τα Εργαστήρια, οι Κλινικές, τα Τμήματα και οι Τομείς, οι προπτυχιακοί και οι μεταπτυχιακοί Φοιτητές, οι Διδάσκοντες, οι Ερευνητές και γενικά σε κάθε μέλος του Πανεπιστημίου μας, μπορεί να χρησιμοποιεί το SPSS, νόμιμα, για ερευνητικούς και διδακτικούς σκοπούς.
- Θα ήθελα να ευχαριστήσω θερμά όλους εκείνους που συνέβαλλαν στην απόκτηση αυτού του ιδιαίτερα χρήσιμου Προγράμματος, αλλά και εκείνους που φρόντισαν και φροντίζουν για την ετήσια συντήρηση και ανανέωσή του. Στην πρώτη κατηγορία είναι τα εξής μέλη του Π.Κ.: Η κ. Σταυρούλα Τσινόρεμα, Αν. Καθηγήτρια Φιλοσοφίας, ο κ. Τάσσος Ξεπαπαδέας, Καθηγητής του Οικονομικού Τμήματος, ο κ. Περικλής Δράκος, μέλος ΕΤΕΠ του Οικονομικού Τμήματος, και ο υπογράφων. Στη δεύτερη κατηγορία είναι ο κ. Κώστας Λάβδας, Αν. Καθηγητής του Τμήματος Πολιτικής Επιστήμης, Αντιπρύτανης του Π.Κ. και οι συνάδελφοί του στο σημερινό Πρυτανικό σχήμα, κ.κ. Ιωάννης Παλλήκαρης, Καθηγητής Ιατρικής, Πρύτανης του Π.Κ., Μαρουδιώ Κεντούρη, Καθηγήτρια Βιολογίας, Αντιπρύτανης, και Ιωάννα Γιατρομανωλάκη, Αν. Καθηγήτρια Φιλολογίας, Αντιπρύτανης.
- Οι πολυάριθμοι χρήστες του SPSS, στο Πανεπιστήμιό μας, οφείλουν ιδιαίτερες ευχαριστίες τον Αντιπρύτανη κ. Λάβδα: Υπέβαλλε, ως Κοσμήτορας τότε της Σχολής Κοινωνικών Επιστημών την πρόταση για ανάληψη της ετήσιας συνδρομής του SPSS από το Πανεπιστήμιο, φρόντισε να συμπεριληφθεί στο προεκλογικό Πρόγραμμα του παρόντος Πρυτανικού Σχήματος ρητή δέσμευση για τη μόνιμη οικονομική υποστήριξη του SPSS, παρεμβαίνει σε κάθε ζήτημα αναφορικά με το SPSS, και μεριμνά σε σταθερή βάση, για την έγκαιρη παραλαβή και διανομή της εκάστοτε νέας έκδοσης του SPSS, στα μέλη του Πανεπιστημίου μας.



Πρόλογος

Το βιβλίο αυτό είναι καρπός συγγραφικής δουλειάς δύο χρόνων, και φιλοδοξεί να καλύψει ένα κενό στην υπάρχουσα ελληνική βιβλιογραφία. Δεν ήταν στις προθέσεις μας να γράψουμε ένα ακόμη βιβλίο θεωρίας, στο οποίο θα βρίθουν οι φόρμουλες, οι μαθηματικοί τύποι, οι εκτενείς μαθηματικές θεμελιώσεις, ο αυστηρός συμβολισμός, οι ορισμοί, οι με χαρτί και μολύβι ασκήσεις, και οι ενίοτε, άγονες θεωρητικές συζητήσεις. Διότι ακριβώς, δεν είναι αυτά που χρειάζεται ένας Κοινωνικός επιστήμονας σήμερα, και οπωσδήποτε δεν είναι αυτά που θα βοηθήσουν ένα ερευνητή στο χώρο των Κοινωνικών Επιστημών να εργασθεί πρακτικά και αποτελεσματικά.

Αντίθετα, αυτό που έχουν ανάγκη οι Κοινωνικοί Επιστήμονες σήμερα, είναι να τους δοθούν συγκεκριμένα εργαλεία, μέθοδοι και τεχνικές που θα βοηθήσουν αποφασιστικά τη **Στατιστική Ανάλυση** των ερευνητικών τους δεδομένων. Ακόμη είναι ανάγκη να προσδιορισθούν στο θεωρητικό και στο τεχνικό επίπεδο, οι στρατηγικές εκείνες οι οποίες θα πρέπει να αναπτυχθούν για να εξυπηρετήσουν το σκοπό και τους στόχους, που στην αρχή της θέτει, η έρευνα. Με άλλα λόγια, είναι ορατό ζητούμενο σήμερα, να αναπτυχθούν με ακρίβεια τα κατάλληλα **στατιστικά- μαθηματικά μοντέλα**, τα οποία θα βοηθήσουν την αναζήτηση σχέσεων ανάμεσα σε μεταβλητές και παράγοντες που εμπλέκονται και περιγράφουν, τη συγκεκριμένη κοινωνική έρευνα.

Έτσι, θελήσαμε να χαράξουμε ένα άλλο, πιο πρακτικό, πιο βατό, και οπωσδήποτε πιο χρήσιμο για τους κοινωνικούς επιστήμονες, δρόμο. Ο δρόμος αυτός επιβάλλει να κρατήσουμε τη θεωρία ως αναγκαίο προαπαιτούμενο, γιατί χωρίς θεωρία δεν υπάρχει επιστήμη, αλλά ταυτόχρονα αυτή τη θεωρία να την εντάξουμε μέσα στη στατιστική διαδικασία. Αυτό σημαίνει ότι κάθε φορά που αποφασίζουμε να χρησιμοποιήσουμε μια στατιστική διαδικασία, μια συγκεκριμένη τεχνική, ή ένα μαθηματικό-στατιστικό μοντέλο, θα πρέπει να αναφερθούμε άμεσα στις θεωρητικές του παραδοχές, να πούμε δηλ. πότε αυτό το μοντέλο εφαρμόζεται, και κάτω από ποιες συγκεκριμένες συνθήκες. Παραπέρα, αυτό σημαίνει, ότι θα πρέπει να περιγραφούν με ακρίβεια, οι εναλλακτικές λύσεις

που πιθανά θα πρέπει να μετέλθουμε, στην περίπτωση που οι αρχικές μας παραδοχές, οι παραδοχές που διατυπώθηκαν για το στατιστικό μοντέλο που αρχικά κατασκευάσαμε, δεν ικανοποιούνται.

Από την άλλη, τίποτε δεν είναι δυνατόν να γίνει, με χαρτί και μολύβι. Δεν μπορούμε να εφαρμόσουμε τεχνικές, στατιστικές διαδικασίες ή τέλος δεν μπορούμε σχεδιάσουμε και να δοκιμάσουμε ακόμα και το πιο απλό στατιστικό μοντέλο, με τα χέρια μας. Με αλγεβρικό δηλ. τρόπο. Αν θέλουμε να είμαστε αποτελεσματικοί, γρήγοροι και ακριβείς, θα πρέπει να εξοικειωθούμε από νωρίς στη χρήση, ενός τουλάχιστον στατιστικού προγράμματος για Κοινωνικές Επιστήμες. Ένα τέτοιο στατιστικό πρόγραμμα, με ιδιαίτερα φιλικό για το χρήστη περιβάλλον, εύκολο στη χρήση και στην εκμάθησή του, ευέλικτο, και με μεγάλες υπολογιστικές δυνατότητες, φαίνεται να είναι το SPSS (Statistical Package for Social Sciences). Ακόμη θα πρέπει να αναφέρουμε ότι το SPSS συγκριτικά με άλλα στατιστικά προγράμματα, διαθέτει την πιο πλατιά βιβλιογραφία, αναφορικά με τις Κοινωνικές Επιστήμες. Το Πανεπιστήμιό μας, το Πανεπιστήμιο Κρήτης, είναι νόμιμος χρήστης αυτού του προγράμματος και παρέχει τη δυνατότητα στα Εργαστήρια, στις Κλινικές, στους Τομείς, στους Φοιτητές, στους Διδάσκοντες και γενικά σε κάθε μέλος του, να το χρησιμοποιεί για ερευνητικούς και διδακτικούς σκοπούς.

Το βασικό μέλημα αυτού του βιβλίου, είναι η αναζήτηση τρόπων και μέσων τα οποία θα υπηρετήσουν, με τον καλύτερο δυνατό τρόπο, τη Στατιστική Ανάλυση των στοιχείων που προέρχονται από τις Κοινωνικές Επιστήμες. Η Στατιστική Ανάλυση, έτσι όπως εκτελείται στο πλαίσιο αυτού του βιβλίου, περιλαμβάνει τα παρακάτω βήματα:

- Καθορισμός του προβλήματος.
- Οικοδόμηση βάσης δεδομένων στο SPSS.
- Κωδικοποίηση και άμεση εισαγωγή δεδομένων στο SPSS.
- Επιστράτευση της κατάλληλης στατιστικής διαδικασίας.
- Παραδοχές γι' αυτή τη στατιστική διαδικασία.
- Έλεγχος των παραδοχών.
- Εναλλακτικές λύσεις σε περίπτωση μη ικανοποίησης των παραδοχών.
- Εκτέλεση της παραπάνω διαδικασίας.
- Διατύπωση μηδενικής και εναλλακτικής υπόθεσης.
- Στατιστικό Συμπέρασμα.
- Τελικό Συμπέρασμα για τη συγκεκριμένη Κοινωνική Επιστήμη από την οποία προέρχονται τα δεδομένα.

Τελικό συμπέρασμα από όλα τα παραπάνω είναι ότι η Στατιστική μας ενδίδει στο βαθμό που μπορεί να προσφέρει βοήθεια στην ανάλυση των κοινωνικών δεδομένων, στον εντοπισμό και την ανάδειξη σχέσεων ανάμεσα σε παράγοντες και μεταβλητές της κοινωνικής έρευνας, και επομένως, μας ενδιαφέρει κυρίως από την άποψη του περιεκτικού, αποτελεσματικού, κατανοητού και αξιόπιστου *εργαλείου ανάλυσης*.

Η γλώσσα που χρησιμοποιούμε στο βιβλίο αυτό, όπως εύκολα μπορεί να διαπιστωθεί από τον αναγνώστη, είναι μια γλώσσα απλή, *μια γλώσσα επικοινωνίας* με τους Κοινωνικούς Επιστήμονες. Δεν απευθυνόμαστε μόνο σε φοιτητές, αλλά και σε διδάσκοντες. Ακόμη πιστεύουμε ότι έχουμε να προσφέρουμε σημαντικές σκέψεις στους νέους ερευνητές. Τέλος, πουθενά δεν έχουμε την αίσθηση, ότι θυσιάσαμε την επιστημονική ακρίβεια χάριν της απλότητας. Καταβάλλαμε κάθε προσπάθεια ώστε το βιβλίο να κινηθεί μέσα σε διεθνείς προδιαγραφές.

Κάποτε ίσως φανεί υπερβολική η χρήση της ξένης ορολογίας, μέσα στο κείμενο. Ωστόσο, αυτό ήταν αναπόφευκτο. Γιατί το SPSS είναι από μόνο του μια γλώσσα. Μια γλώσσα δομημένη, σαφέστατα επιστημονική, ακριβής, με μεγάλες υπολογιστικές δυνατότητες, μια γλώσσα που θα πρέπει κάποια στιγμή, να ‘μιλήσουμε’. Αλλά και για όσους προτιμούν με πάθος τη χρήση ελληνικής και μόνον ορολογίας, έχει ληφθεί πρόνοια: Δεν υπάρχει κανένας ξενόγλωσσος όρος, που να μην έχει μεταφραστεί στα ελληνικά, όταν εισάγεται στο κείμενο για πρώτη φορά.

Τέλος, θα πρέπει να τονίσουμε ότι το βιβλίο αυτό έχει γραφτεί κατά τέτοιο τρόπο, ώστε να μπορεί να παρακολουθήσει την εξέλιξη του SPSS. Προς το παρόν αναφέρεται στην έκδοση SPSS 12.0. Ωστόσο, στην αναθεωρημένη του έκδοση, όποτε αυτή κριθεί αναγκαίο να υπάρξει, οπωσδήποτε θα συμπεριληφθούν σημαντικά στοιχεία από νεότερες εκδόσεις του SPSS. Ας μην ανησυχούν όμως οι αναγνώστες του βιβλίου και οι χρήστες του SPSS. Η πολύχρονη εμπειρία μας στη χρήση αυτού του προγράμματος έδειξε με σαφή τρόπο, ότι η εκάστοτε νέα έκδοση του SPSS, δεν αποτελεί άρνηση της προηγούμενης, ούτε και επανάσταση. Αντίθετα, αποτελεί μια βελτίωση αυτής, όχι πάντοτε σημαντική.



Περιεχόμενα

Κεφάλαιο 1: Βασικές Στατιστικές Έννοιες

1.1 Η έννοια της μεταβλητής	21
1.2 Είδη μεταβλητών	22
1.3 Τι εννοούμε με τον όρο παρατηρησιακή μονάδα	26
1.4 Πληθυσμός και Δείγμα	27
1.5 Κλίμακες μέτρησης	29
1.6 Η κατανόηση των κλιμάκων μέτρησης μέσα από την κωδικοποίηση	36
1.7 Αξιοπιστία (Reliability) και Εγκυρότητα (Validity)	42

Κεφάλαιο 2: Βασικές Έννοιες και Διαδικασίες του SPSS

2.1 Ξεκίνημα	45
2.2 Τι είναι και πως δημιουργούμε μια βάση δεδομένων στο SPSS	47
2.3 Επιλογή μεταβλητών και cases	63
2.4 Διαγραφή ή Παρεμβολή μεταβλητών και cases	64
2.5 Πού βρίσκεται κάποια τιμή που είναι λάθος ή μας ενδιαφέρει	65
2.6 Επανακωδικοποίηση των δεδομένων μας (η διαδικασία Recode)	65
2.7 Μετασχηματισμοί δεδομένων (η διαδικασία Compute)	71
2.8 Η ταξινόμηση των δεδομένων μας (Sort Cases)	76
2.9 Επιλέγοντας περιπτώσεις (Select cases)	77
2.10 Μετακινώντας μεταβλητές (Moving Variables)	79
2.11 Καταμετρώντας τιμές από σύνολα μεταβλητών (η διαδικασία Count)	81
2.12 Ο έλεγχος τυχαιότητας του δείγματός μας	83
2.13 Πως λαμβάνουμε ένα μικρότερο σύνολο δεδομένων από ένα μεγαλύτερο, ή ισοδύναμα, πως τεμαχίζουμε, με τυχαίο τρόπο, μια βάση δεδομένων του SPSS. Η διαδικασία Random Sample	89

Κεφάλαιο 3: Περιγραφική Στατιστική

3.1	Δείκτες κεντρικής τάσης (measures of central tendency)	93
3.2	Τα ζητούμενα ‘χαρίσματα’ από τους δείκτες κεντρικής τάσης	99
3.3	Η αλγεβρική σχέση των δεικτών κεντρικής τάσης	100
3.4	Δείκτες διασποράς (measures of variation), ή μεταβλητότητας (variability), ή διασκόρπισης (dispersion)	101
3.5	Συζήτηση	107
3.6	Μέτρα ασυμμετρίας	108
3.7	Δείκτες ομοιογένειας	115
3.8	Η προϊδέαση της κανονικής κατανομής	116
3.9	Παράδειγμα – Πρόβλημα 1	119
3.10	Ομαδοποιημένα δεδομένα	140
3.11	Ασκήσεις Περιγραφικής Στατιστικής που επιλύονται αλγεβρικά, δηλ. χωρίς τη βοήθεια υπολογιστή	149

Κεφάλαιο 4: Στοιχεία από τη Θεωρία Πιθανοτήτων

4.1	Εισαγωγή	153
4.2	Η έννοια του πειράματος τύχης	153
4.3	Η έννοια του δειγματοχώρου ή δειγματικού χώρου	154
4.4	Η έννοια του ενδεχομένου	154
4.5	Πράξεις με ενδεχόμενα	154
4.6	Ασυμβίβαστα ενδεχόμενα	156
4.7	Κλασικός ορισμός πιθανότητας	156
4.8	Κανόνες λογισμού πιθανοτήτων	157
4.9	Ασκήσεις	157
4.10	Ανεξάρτητα ενδεχόμενα	166
4.11	Ασκήσεις	167
4.12	Δεσμευμένη πιθανότητα	170
4.13	Ασκήσεις	170

Κεφάλαιο 5: Η Κανονική Κατανομή (Normal Distribution)

5.1	Εισαγωγικές έννοιες	175
5.2	Ορισμός της κανονικής κατανομής	180

5.3	Η αξία και η χρησιμότητα της κανονικής κατανομής – Σύντομη ιστορική αναδρομή	181
5.4	Οι ιδιότητες της κανονικής κατανομής	182
5.5	Η μετατροπή των τιμών της κανονικής κατανομής σε z-τιμές	183
5.6	Πώς το SPSS δημιουργεί z-τιμές	184
5.7	Ιδιότητες της τυπικής ή τυποποιημένης κανονικής κατανομής (Standardized Normal Distribution)	187
5.8	Η συνάρτηση αθροιστικής κατανομής της τυπικής κανονικής κατανομής	187
5.9	Σύγκριση τιμών που ανήκουν σε διαφορετικές κανονικές κατανομές	190
5.10	Δειγματοληπτικές κατανομές. Η κατανομή του μέσου όρου	190
5.11	Το Θεώρημα του Κεντρικού Ορίου	195
5.12	ΓΕΝΙΚΕΣ ΑΣΚΗΣΕΙΣ	198

Κεφάλαιο 6: Έλεγχοι Υποθέσεων και Διαστήματα Εμπιστοσύνης

6.1	Εισαγωγικά. Η μηδενική και η εναλλακτική υπόθεση	203
6.2	Παραδείγματα μηδενικών και εναλλακτικών υποθέσεων	204
6.3	Μονόπλευρος και αμφίπλευρος έλεγχος υποθέσεων	206
6.4	Σφάλματα στους ελέγχους υποθέσεων	208
6.5	Η διαφοροποιητική δύναμη ενός στατιστικού κριτηρίου (Power)	209
6.6	Το περιεχόμενο των όρων ‘στατιστικώς σημαντική διαφορά’, ‘στατιστικώς ασήμαντη διαφορά’ και ‘σπουδαιότητα’	210
6.7	Η έννοια του διαστήματος εμπιστοσύνης	213
6.8	Η κατασκευή διαστήματος εμπιστοσύνης για τον άγνωστο μέσο όρο μ ενός πληθυσμού με τη βοήθεια ενός τυχαίου δείγματος	215

Κεφάλαιο 7: Έλεγχοι Κανονικότητας

7.1	Εισαγωγικά	221
7.2	Έλεγχος κανονικότητας με τη βοήθεια στατιστικών κριτηρίων	222
7.3	Έλεγχος κανονικότητας με τη βοήθεια γραφικών αναπαραστάσεων	230
7.3.1	Έλεγχος κανονικότητας με τη βοήθεια του ονομαζόμενου Normal Q-Q Plot	230

7.3.2	Με τη βοήθεια του ονομαζόμενου Detrended Normal Q-Q Plot	231
7.3.3	Έλεγχος κανονικότητας με τη βοήθεια του ονομαζόμενου Box Plot	233
7.3.4	Ο έλεγχος κανονικότητας με τη βοήθεια του ονομαζόμενου Histogram	236
7.3.5	Ο έλεγχος κανονικότητας με τη βοήθεια του ονομαζόμενου Stem and Leaf	237
7.4	Ο έλεγχος σημαντικότητας με τη βοήθεια περιγραφικών στατιστικών δεικτών	238
7.5	Τελικές Επισημάνσεις	239

Κεφάλαιο 8: Γραφικές Αναπαραστάσεις

8.1	Εισαγωγικά	243
8.2	Απλά γραφήματα σκέδασης (Simple Scatter Plots)	244
8.3	Τα γραφήματα γραμμής (Line Charts)	255
8.4	Τα ραβδογράμματα σφάλματος (Error Bar Charts)	259
8.5	Τα διαγράμματα PARETO	263

Κεφάλαιο 9: Απλή Τυχαία Δειγματοληψία

9.1	Εισαγωγικά	267
9.2	Βασικές έννοιες και ορισμοί	268
9.3	Το μέγεθος του δείγματος	269
9.4	Απλή τυχαία δειγματοληψία για την εκτίμηση της μέσης τιμής μ ενός πεπερασμένου ή άπειρου κανονικού πληθυσμού, με γνωστή διασπορά σ^2	272
9.5	Απλή τυχαία δειγματοληψία για την εκτίμηση ποσοστού	277
9.6	Οι ‘αρετές’ ενός ερωτηματολογίου έρευνας: περιεχόμενο, δομή, εμφάνιση	283

Κεφάλαιο 10: Η Δοκιμασία χ^2

10.1	Εισαγωγικά	289
10.2	Η δοκιμασία χ^2 ως τεστ ομοιογένειας	290
10.2.1	Παράδειγμα – Πρόβλημα 1	293
10.2.2	Η υπολογιστική λύση του παραδείγματος 1	297

10.3	Η δοκιμασία χ^2 ως τεστ ανεξαρτησίας	309
10.3.1	Παράδειγμα – Πρόβλημα 2	310
10.3.2	Λύση του προβλήματος 2 με το SPSS	310
10.4	Πότε το χ^2 , ως τεστ ανεξαρτησίας ή ως τεστ ομοιογένειας, δεν εφαρμόζεται;	314
10.5	Η δοκιμασία χ^2 ως τεστ καλής προσαρμογής	314
10.5.1	Παράδειγμα – Πρόβλημα 3	316
10.5.2	Παράδειγμα – Πρόβλημα 4	320
10.5.3	Παράδειγμα – Πρόβλημα 5	328
10.5.4	Παράδειγμα – Πρόβλημα 6	331
10.5.5	Παράδειγμα – Πρόβλημα 7	334
10.5.6	Παράδειγμα – Πρόβλημα 8	335
10.6	Πλεονεκτήματα του κριτηρίου των $K-S$ έναντι του μη παραμετρί- κού χ^2	340
10.7	Τι είναι τα Exact tests, πότε τα εφαρμόζουμε, και πόσο φερέγγυα είναι, αναφορικά με το χ^2	340
10.7.1	Παράδειγμα – Πρόβλημα 9	340
10.8	Πώς αντιμετωπίζεται η παραβίαση της παραδοχής του 20%, με τη μέθοδο της ενοποίησης των κατηγοριών των κατηγορικών μεταβλητών. Η παρέμβαση της εντολής Recode	349
10.8.1	Παράδειγμα – Πρόβλημα 10	349
10.9	Τελικά τι κάνουμε; Σκέψεις και σύνοψη στρατηγικών και εναλλα- κτικών λύσεων	356

Κεφάλαιο 11: Το Στατιστικό Κριτήριο T -Test

11.1	Γενικές παραδοχές για όλες τις μορφές T -Test	359
11.2	T -Test για δύο ανεξάρτητα δείγματα (Two independent samples T -Test)	360
11.2.1	Παράδειγμα – Πρόβλημα 1	364
11.3	T -test για ένα δείγμα (one sample T -test)	373
11.3.1	Παράδειγμα – Πρόβλημα 2	373

Κεφάλαιο 12: Γραμμική Διμεταβλητή Συσχέτιση (Linear Bivariate Correlation)

12.1	Το νόημα και το περιεχόμενο της συσχέτισης	379
------	--------------------------------------------------	-----

12.2	Οι παραδοχές για την Διμεταβλητή Ανάλυση Συσχέτισης	381
12.3	Στοιχεία για το συντελεστή συσχέτισης του Pearson	384
12.3.1	Παράδειγμα – Πρόβλημα 1	385
12.4	Ο έλεγχος της κανονικότητας	389
12.5	Ο έλεγχος της Γραμμικότητας (Linearity) και της Ομοσκεδαστικότητας (Homoscedasticity)	390
12.6	Ο συντελεστής ρ του Spearman και οι εναλλακτικές λύσεις στην περίπτωση που η κατά Pearson Ανάλυση Συσχέτισης είναι αδιέξοδη	392
12.6.1	Παράδειγμα – Πρόβλημα 2	393

Κεφάλαιο 13: Μερική Συσχέτιση (Partial Correlation)

13.1	Σκοπός και περιεχόμενο της ανάλυσης μερικής συσχέτισης	403
13.2	Η τάξη της ανάλυσης μερικής συσχέτισης (order of partial correlation)	404
13.3	Οι παραδοχές	405
13.3.1	Παράδειγμα – Πρόβλημα 1	405
13.4	Τρέχοντας ταυτόχρονα Zero-order correlation analysis και First- order partial correlation analysis, με μεταβλητή ελέγχου τη μετα- βλητή EXPER	406
13.5	Τρέχοντας ταυτόχρονα Zero-order correlation analysis και First- order partial correlation analysis, με μεταβλητή ελέγχου τη μετα- βλητή AGE	410
13.6	Τρέχοντας ταυτόχρονα Zero-order correlation analysis και Second- order partial correlation analysis, με δύο μεταβλητές ελέγχου, EXPER και AGE	412

Κεφάλαιο 14: Απλή Γραμμική Παλινδρόμηση (Simple Linear Regression)

14.1	Εισαγωγικά	415
14.2	Τι είναι η ευθεία παλινδρόμησης (regression line)	416
14.3	Πώς κατασκευάζεται η ευθεία παλινδρόμησης	417
14.4	Η γραφική αναπαράσταση της ευθείας παλινδρόμησης	421
14.5	Οι παραδοχές για την Απλή Παλινδρομική Ανάλυση	423

14.6	Το νόημα και η σπουδαιότητα των υπολοίπων ή καταλοίπων (residuals)	427
14.6.1	Παράδειγμα – Πρόβλημα 1	430
14.7	Τρέχοντας απλή παλινδρομική ανάλυση	432
14.8	Τα αποτελέσματα της παλινδρομικής ανάλυσης	437
14.9	Η γραφική αναπαράσταση της εξίσωσης παλινδρόμησης	442
14.10	Ο έλεγχος της παραδοχής της ανεξαρτησίας	446
14.11	Ο έλεγχος της παραδοχής της κανονικότητας	448
14.12	Ο έλεγχος της παραδοχής της γραμμικότητας	451
14.13	Ο έλεγχος της παραδοχής της ισότητας των διασπορών	454
14.14	Η φροντίδα των ακραίων ή επιδραστικών τιμών (Outliers or Influential Points)	454

Κεφάλαιο 15: Πολλαπλή Γραμμική Παλινδρόμηση (Multiple Linear Regression)

15.1	Νόημα και σκοπός της πολλαπλής παλινδρόμησης	459
15.2	Οι παραδοχές της πολλαπλής παλινδρομικής ανάλυσης	460
15.2.1	Παράδειγμα – Πρόβλημα 1	463
15.3	Η εκτέλεση της ανάλυσης παλινδρόμησης	465
15.4	Τα αποτελέσματα της ανάλυσης παλινδρόμησης	472
15.4.1	Η παλινδρομική εξίσωση	480
15.4.2	Η ερμηνεία της παλινδρομικής εξίσωσης	482
15.4.3	Η ‘αξία’ των μεταβλητών	483
15.5	Ο έλεγχος των παραδοχών	485
15.5.1	Ο έλεγχος της παραδοχής της ανεξαρτησίας	485
15.5.2	Ο έλεγχος της παραδοχής της κανονικότητας	489
15.5.3	Ο έλεγχος της παραδοχής της γραμμικότητας	494
15.5.4	Ο έλεγχος της παραδοχής της ισότητας των διασπορών	503
15.5.5	Ο έλεγχος της πολυσυγγραμμικότητας	504
15.5.6	Η ανίχνευση ακραίων παρατηρήσεων (outliers) και παρατηρήσεων επίδρασης (influential points)	508
15.5.7	Ειδικά σημαντικά γραφήματα γύρω από τις ακραίες τιμές και τις παρατηρήσεις επίδρασης	525
15.5.8	Τα διαγράμματα μερικών υπολοίπων (Partial Residual Plots)	532
15.6	Οι παραβιάσεις των παραδοχών, οι ‘θεραπείες’ και οι εναλλακτικές λύσεις	534
15.6.1	Η παραβίαση της κανονικότητας	535

15.6.2	Η παραβίαση της γραμμικότητας και της ισότητας των διασπορών	536
15.6.3	Η αντιμετώπιση του προβλήματος της πολυσυγγραμμικότητας	542
15.7	Η επιλογή της μεθόδου στο πλαίσιο της πολλαπλής γραμμικής παλινδρόμησης	544
15.7.1	Πόσες μεταβλητές να συμπεριλάβουμε στο μοντέλο μας;	547
15.8	Η επικύρωση του παλινδρομικού μας μοντέλου (model validation)	550
15.9	Με ποια σειρά εισάγουμε τις μεταβλητές σε ένα παλινδρομικό μοντέλο;	561
15.10	Ο Πίνακας ANOVA και η ερμηνεία του	562
15.11	Η ερμηνεία λογαριθμικής παλινδρομικής εξίσωσης	564
15.12	Πώς εισάγουμε κατηγορικές μεταβλητές μέσα σε ένα γραμμικό παλινδρομικό μοντέλο	565
15.13	Συζήτηση – Γενικά συμπεράσματα	568

Κεφάλαιο 16: Απλή Ανάλυση Διασποράς με Ένα Παράγοντα (One Way Analysis of Variance)

16.1	Νόημα και σκοπός της ανάλυσης διασποράς	577
16.2	Οι παραδοχές της απλής ανάλυσης διασποράς	580
16.2.1	Παράδειγμα – Πρόβλημα 1	581
16.3	Τα αποτελέσματα της ανάλυσης	585
16.4	Οι γραφικές αναπαραστάσεις	588
16.5	Ο έλεγχος των παραδοχών	591
16.5.1	Ο έλεγχος της κανονικότητας	591
16.5.2	Ο έλεγχος της ισότητας των διασπορών	593
16.6	Παρατηρήσεις πάνω στις 3 σπουδαιότερες μεθόδους πολλαπλών συγκρίσεων (Tukey, Bonferroni και Scheffe)	593

Κεφάλαιο 17: Μη Παραμετρικά Στατιστικά Κριτήρια (Nonparametric Statistics)

17.1	Γενικά	597
17.2	Το μη παραμετρικό κριτήριο των Mann - Witney και τα συναφή, για δύο ανεξάρτητα δείγματα, κριτήρια Kolmogorov - Smirnov Z, Moses extreme reactions, και Wald - Wolfowitz runs	598

17.2.1	Παράδειγμα – Πρόβλημα 1	599
17.2.2	Δημιουργία τάξης (rank) για τις αρχικές μας τιμές	601
17.2.3	Η εκτέλεση του μη παραμετρικού κριτηρίου Mann - Whitney	604
17.2.4	Τα αποτελέσματα	606
17.3	Τα συναφή με το Mann - Whitney τεστ	609
17.4	Το μη παραμετρικό κριτήριο των Kruskal - Wallis	611
17.4.1	Παράδειγμα – Πρόβλημα 2	612
17.4.2	Τρέχοντας Kruskal - Wallis ANOVA	613
17.4.3	Τα αποτελέσματα της Kruskal - Wallis ANOVA	614
17.5	Τα συναφή με το Kruskal - Wallis στατιστικά κριτήρια	616
17.6	Το διωνυμικό κριτήριο (Binomial Test)	618
17.6.1	Παράδειγμα – Πρόβλημα 3	619

Κεφάλαιο 18: Kaplan-Meier Ανάλυση Επιβίωσης (Kaplan-Meier Survival Analysis)

18.1	Εισαγωγικά	625
18.2	Τι είναι οι censored cases και πως διαφοροποιούν τα Kaplan-Meier μοντέλα, έναντι των παραδοσιακών στατιστικών μοντέλων	626
18.3	Η συνάρτηση επιβίωσης (survival function)	627
18.4	Πότε χρησιμοποιούμε μοντέλα Kaplan - Meier - Ερευνητικές Υποθέσεις	627
18.5	Η δομή ενός μοντέλου Kaplan - Meier	628
18.6	Οι παραδοχές για το Kaplan - Meier στατιστικό μοντέλο	631
18.7	Στατιστικά τεστ σύγκρισης καμπυλών επιβίωσης, στο πλαίσιο των Kaplan - Meier στατιστικών μοντέλων	632
18.8	Ποιο νόημα έχουν η μέση τιμή, η διάμεσος, και το τυπικό σφάλμα στα Kaplan - Meier μοντέλα	634
18.8.1	Παράδειγμα – Πρόβλημα 1	635
18.8.2	Παράδειγμα – Πρόβλημα 2	645
18.8.3	Η εκτέλεση του παραπάνω Kaplan - Meier μοντέλου	646
18.8.4	Τα αποτελέσματα	647
18.8.5	Παράδειγμα – Πρόβλημα 3	650
18.8.6	Η εκτέλεση του παραπάνω Kaplan - Meier μοντέλου	651
18.8.7	Τα αποτελέσματα	652
18.8.8	Παράδειγμα – Πρόβλημα 4	658
18.8.9	Η εκτέλεση του παραπάνω Kaplan - Meier μοντέλου	659
18.8.10	Τα αποτελέσματα	661

ΒΙΒΛΙΟΓΡΑΦΙΑ

ΕΛΛΗΝΙΚΗ ΒΙΒΛΙΟΓΡΑΦΙΑ669

ΞΕΝΟΓΛΩΣΣΗ ΒΙΒΛΙΟΓΡΑΦΙΑ670

ΠΑΡΑΡΤΗΜΑ

ΣΤΑΤΙΣΤΙΚΟΙ ΠΙΝΑΚΕΣ675

Ενρετήριο Όρων683



Βασικές Στατιστικές Έννοιες

1.1 Η έννοια της μεταβλητής

Με τον όρο «μεταβλητή» ονομάζουμε κάθε τι το οποίο μεταβάλλεται, ποικίλλει ή παραλλάσσει. Και αυτό το «κάθε τι» μπορεί να είναι ένα χαρακτηριστικό, μια ιδιότητα, μια ικανότητα, ή τέλος ένας παράγοντας που μας ενδιαφέρει στο πλαίσιο της κοινωνικής έρευνας που διεξάγουμε. Στις Κοινωνικές Επιστήμες υπάρχει ολόένα και μεγαλύτερη ανάγκη να μετρήσουμε. Αλλά η μέτρηση χωρίς τη χρήση μεταβλητών, πουθενά, δηλ. σε κανένα ερευνητικό πλαίσιο, δεν είναι δυνατή. Θέλουμε για παράδειγμα, να μετρήσουμε το δείκτη ευφυΐας κάποιων ατόμων, να παρακολουθήσουμε την εξέλιξη ασθενών, να καταγράψουμε τις στάσεις και γενικά την πολιτική συμπεριφορά κάποιων κοινωνικών ομάδων, να αξιολογήσουμε τις ικανότητες κάποιων μαθητευόμενων ατόμων, να μετρήσουμε το χρόνο αντίδρασης σε κάποιο ερέθισμα, να καταγράψουμε τη θέληση του εκλογικού σώματος κτλ. Και όλα αυτά, φυσικά επιβάλλουν να εισαχθούμε σε μια διαδικασία μέτρησης με τη χρήση μεταβλητών.

Πρακτικά, όταν κάνουμε λόγο για μια μεταβλητή, ξεχωρίζουμε σε αυτήν δύο πράγματα: το όνομα και την τιμή της. Για παράδειγμα όταν δηλώνουμε: $V1=3$ είναι σαφές ότι πρόκειται για μια μεταβλητή της οποίας το όνομα είναι $V1$ και η τιμή που τώρα της αποδίδεται, είναι 3.

Οι μεταβλητές έχουν, επομένως, ονόματα και λαμβάνουν τιμές, που όπως θα δούμε παρακάτω, προέρχονται από μια συγκεκριμένη κλίμακα μέτρησης. Ωστόσο, τα ονόματα των μεταβλητών δεν μπορεί να είναι οποιαδήποτε. Υπάρχουν περιορισμοί και μάλιστα στο πλαίσιο του SPSS (που ιδιαίτερα μας ενδιαφέρει), είναι αυστηροί (βλ. επόμενο κεφάλαιο). Από τώρα όμως είναι ανάγκη να επισημάνουμε ότι το όνομα μιας μεταβλητής είναι μια αδιαίρετη, ενιαία και συμπαγής οντότητα αλφαβητικών και αριθμητικών χαρακτήρων, η οποία δεν

διακόπτεται από κενά διαστήματα. Στην μαθηματική- υπολογιστική γλώσσα αυτή η οντότητα ονομάζεται ορμανθός, δηλ. αλληλουχία αλφαριθμητικών χαρακτήρων. Για παράδειγμα, μια τέτοια οντότητα αποτελεί το όνομα CAT_REG1. Τη συνέχεια που υπαινίσσεται ο ορμανθός, την εξασφαλίζει στην περίπτωση μας, η κάτω μπάρα, η οποία «συνδέει» κατά κάποιο τρόπο ή για να το πούμε καλύτερα, ανάγει τα δύο μέρη CAT και REG1, σε μια ενιαία ονοματολογική οντότητα που είναι αποδεκτή σαν όνομα μεταβλητής στο υπολογιστικό πλαίσιο του SPSS.

Τι τις χρειαζόμαστε όμως τις μεταβλητές; Και τι ακριβώς μας ενδιαφέρει από αυτές; Μήπως ο τρόπος ή ο ρυθμός με τον οποίο κινούνται ή αλλάζουν τιμές; Ο τρόπος που επηρεάζονται ή αλληλεπιδρούν; Ή μήπως ο βαθμός, η κατεύθυνση συσχέτισής τους ή η ένταση αλληλεξάρτησής τους; Ή τέλος η συνεισφορά τους σε κάποιο στατιστικό μοντέλο για τον επιμερισμό της συνολικής πληροφορίας;

Η απάντηση δεν μπορεί να είναι μονολεκτική. Εξαρτάται από το είδος, την ποιότητα και το βάθος της στατιστικής ανάλυσης που επιχειρούμε. Όσο καλύτερα γνωρίζουμε ένα -ένα από τα παραπάνω, όσο βαθύτερα καταφέρουμε να υπεισέλθουμε στην υφή και στις σχέσεις των εμπλεκόμενων στην έρευνά μας μεταβλητών, τόσο πιο πλούσια θα είναι και η πληροφορία στο επίπεδο της Κοινωνικής Επιστήμης που καλλιεργούμε.

1.2 Είδη μεταβλητών

Ανάλογα με το πώς μεταβάλλεται μια μεταβλητή διακρίνεται και το είδος στο οποίο ανήκει. Έτσι, αν μεταβάλλεται από την άποψη της ποιότητας τη διακρίνουμε σε **ποιοτική** (qualitative), ενώ αν μεταβάλλεται από την άποψη της ποσότητας τη διακρίνουμε σε **ποσοτική** (quantitative).

▮ Στις **ποιοτικές μεταβλητές** κατατάσσονται το φύλο, (άνδρας ή γυναίκα), η θρησκεία (Χριστιανός, Μουσουλμάνος, κτλ.), ο τόπος διαμονής (αστικός, ημιαστικός, αγροτικός), η οικογενειακή κατάσταση (έγγαμος, άγαμος, διαζευγμένος, χήρος), το μορφωτικό επίπεδο (απόφοιτος Δημοτικού, Γυμνασίου, Λυκείου, Πανεπιστημίου κτλ.), η πολιτική τοποθέτηση (δεξιά, κέντρο, αριστερά) καθώς και πλήθος άλλων μεταβλητών οι οποίες συγκροτούν κατηγορίες **κατά το δυνατόν** ευδιάκριτες και **οπωσδήποτε** αμοιβαία απο-

κλειόμενες, αν η μεταβλητή διαθέτει δύο επίπεδα (διχοτομική). Για παράδειγμα, ένας μαθητής ανάλογα με την απόδοσή του *ίσως* γίνει κατορθωτό να ενταχθεί σε μια από τις παρακάτω κατηγορίες (κακός, μέτριος, καλός, άριστος), οι οποίες ούτε σαφείς, κατά την άποψή μας είναι, ούτε βέβαια αμοιβαία αποκλειόμενες αφού είναι πάνω από δύο. Στη διαδρομή του χρόνου ίσως ο «κακός», δεν είναι πια και τόσο «κακός», ή από δοκιμασία σε δοκιμασία ή από μάθημα σε μάθημα, τίποτε δεν αποκλείεται, δηλ. τίποτε δεν είναι δεδομένο ή σαφές αναφορικά με την επίδοσή του, και επομένως η ένταξη ενός τέτοιου ατόμου σε μια κατηγορία, ούτε σαφής, ούτε ακριβής, ούτε δίκαιη, πιθανώς να μην είναι. Σε κάθε περίπτωση κατηγορικής μεταβλητής εξετάζεται ποιος εντάσσει ποιόν, και πού, δηλ. σε ποια κατηγορία, τον εντάσσει. Τελικά, αυτό που απαιτεί η επιστημονική έρευνα είναι η ακρίβεια, αναφορικά με την ένταξη ενός ερευνητικού υποκειμένου, η οποία όταν απουσιάζει εύκολα η έρευνα εκφυλίζεται σε δημοσιογραφία αμφίβολης επιστημονικής αξίας. Να σημειώσουμε τέλος, ότι οι ποιοτικές μεταβλητές λίγη πληροφορία γενικώς «κουβαλάνε» ή είναι σε θέση να μας προσφέρουν, και ακόμη, σε μικρό βαθμό είναι στατιστικώς επεξεργάσιμες, όπως θα δείξουμε στα επόμενα.

▣▣▣▣ ▣ Στις **ποσοτικές μεταβλητές** κατατάσσονται η ηλικία, το βάρος, το ύψος, το εισόδημα, ο βαθμός ευφυΐας κτλ. Στην περίπτωση των ποσοτικών μεταβλητών αυτό που αναζητούμε επειγόντως είναι **η μονάδα** και μάλιστα **η κατάλληλη** μονάδα μέτρησης. Για παράδειγμα, ένας φοιτητής πήρε 80 μονάδες στα Μαθηματικά και 65 μονάδες στη Βοτανική. Διόλου απίθανο ο συγκεκριμένος φοιτητής να είναι καλύτερος στη Βοτανική. Εξαρτάται από το είδος των χρησιμοποιούμενων μονάδων. Και για να είμαστε πιο ακριβείς το ερώτημα πού είναι καλύτερος ο εν λόγω φοιτητής, θα το απαντήσουμε σωστά μόνο αν ανάγουμε τις μονάδες των Μαθηματικών και της Βοτανικής σε ένα ενιαίο (και άρα αντικειμενικό ως ένα βαθμό) σύστημα μέτρησης. Οι ποσοτικές μεταβλητές διακρίνονται παραπέρα σε **συνεχείς (continuous) και σε διακριτές (discrete)**.

▣▣▣▣ ▣ Μια **συνεχής ποσοτική μεταβλητή** συνήθως κινείται ανάμεσα σε μια ελάχιστη και σε μια μέγιστη τιμή από το σύνολο των πραγματικών αριθμών (real numbers). Αυτό που τη χαρακτηρίζει «συνεχή» είναι η ικανότητά της να μπορεί να λάβει κάθε τιμή ανάμεσα στην ελάχιστη και στη μέγιστη τιμή της. Για παράδειγμα, το βάρος ενός πειραματόζωου είναι δυνατόν να είναι 2 κιλά, 45 γραμμάρια και 9 χιλιοστά του γραμμαρίου και κτλ.... Άλλο παράδειγμα, η ηλικία ενός μικροβίου στον αέρα μπορεί να είναι 25 δευτερόλεπτα,

0,3 εκατοστά του δευτερολέπτου κτλ.. Με άλλα λόγια, μπορεί να γίνει λόγος για ατελείωτα δεκαδικά ψηφία και επομένως υπάρχει «συνέχεια» στις τιμές μιας μεταβλητής. Τελειώνοντας με αυτές τις μεταβλητές θα πρέπει να επισημάνουμε κάτι το οποίο, λίγο-πολύ είναι φανερό, από τα προηγούμενα. Δεν υπάρχει μέγιστη ακρίβεια στις μετρήσεις με συνεχείς ποσοτικές μεταβλητές.

- ▣ Αντίθετα με τις συνεχείς, οι **διακριτές ποσοτικές μεταβλητές** δεν «κτενίζουν» όλες τις δυνατές τιμές ανάμεσα σε ένα ελάχιστο και σε μέγιστο, αλλά λαμβάνουν **ορισμένες, διακριτές ή απαριθμητές** τιμές, οι οποίες προκύπτουν από τις μετρήσεις και είναι πάντα ακέραιοι αριθμοί. Για παράδειγμα, μια οικογένεια δεν μπορεί να έχει 2,5 παιδιά. Ωστόσο, ο μέσος όρος για την ελληνική οικογένεια ίσως είναι 1,7 παιδιά!! Στην περίπτωση των διακριτών ποσοτικών μεταβλητών είναι σαφής η ύπαρξη ακρίβειας, σε αντίθεση με της συνεχείς ποσοτικές μεταβλητές. Ένα άλλο παράδειγμα διακριτής μεταβλητής, θα μπορούσε να είναι η μεταβλητή που εκφράζει τον αριθμό των ηλεκτρονίων τα οποία «χωράνε» οι ηλεκτρονικές στοιβάδες στη δομή ενός ατόμου. Από τη Φυσική είναι γνωστό ότι μια συγκεκριμένη ηλεκτρονική στοιβάδα δεν μπορεί να φιλοξενήσει οποιονδήποτε αριθμό ηλεκτρονίων, αλλά έναν αριθμό ο οποίος εξαρτάται από την τάξη της.

Ανάλογα με τη θέση της σε ένα στατιστικό μοντέλο μια μεταβλητή είναι δυνατόν να διακριθεί σε ανεξάρτητη ή εξαρτημένη.

- ▣ Θα χαρακτηρίζουμε μια μεταβλητή ως **ανεξάρτητη** (independent), όταν αυτή έχει εισαχθεί από τον ερευνητή για να εκτιμηθεί η επίδρασή της πάνω σε μια άλλη και σε άλλες μεταβλητές οι οποίες χαρακτηρίζονται ως εξαρτημένες. Με άλλα λόγια, η ανεξάρτητη μεταβλητή αποτελεί την αιτία και η αυτόνομη δράση της επηρεάζει άλλη ή άλλες μεταβλητές. Σε ορισμένα στατιστικά μοντέλα την ανεξάρτητη μεταβλητή τη συναντάμε και με το όνομα **predictor variable**, για το λόγο ότι προϋδεάζει ή προβλέπει την τιμή της εξαρτημένης μεταβλητής.
- ▣ Ανάλογα, θα χαρακτηρίζουμε μια μεταβλητή ως εξαρτημένη (dependent), όταν η τιμή της επηρεάζεται ή προσδιορίζεται από τις μεταβολές που συμβαίνουν στην ανεξάρτητη ή στις ανεξάρτητες μεταβλητές. Με άλλα λόγια, η εξαρτημένη μεταβλητή εκφράζει το αποτέλεσμα της δράσης άλλης ή άλλων μεταβλητών οι οποίες την επηρεάζουν. Σε ορισμένα στατιστικά μοντέλα η εξαρτημένη μεταβλητή απαντάται και με τον όρο response variable (βλ. στατιστική διαδικασία Categorical Regression).

▮▮▮▮ *Παράδειγμα:* Ας υποθέσουμε ότι ένας ερευνητής θέλει να διερευνήσει τη σχέση ανάμεσα στην ικανότητα ενός πωλητή να πραγματοποιεί πωλήσεις (από τη μια), και στα χρόνια υπηρεσίας του, στην μόρφωση, στην ηλικία και στην πείρα του (από την άλλη). Θα πρέπει να οικοδομήσει ένα στατιστικό μοντέλο με μία εξαρτημένη και 3 ανεξάρτητες μεταβλητές. Σε αυτό το μοντέλο η εξαρτημένη μεταβλητή θα εκφράζει το βαθμό ικανότητας του πωλητή, ενώ οι ανεξάρτητες μεταβλητές θα αφορούν την ηλικία, τη μόρφωση και την πείρα του. Οι τελευταίες, δηλ. οι ανεξάρτητες μεταβλητές, θα πρέπει να είναι σε θέση να προβλέψουν την τιμή της εξαρτημένης μεταβλητής, όταν είναι γνωστή η εξίσωση παλινδρόμησης που προσδιορίζεται από τα δεδομένα της έρευνας.

▮▮▮▮ Όταν διεξάγουμε μια έρευνα, εκτός από τις ανεξάρτητες και τις εξαρτημένες μεταβλητές, εκτός δηλ. από εκείνες που επηρεάζουν και εκείνες που επηρεάζονται, και που κατά κάποιο τρόπο είναι ‘ορατές’, υπάρχουν και άλλες ‘αόρατες’ μεταβλητές, που υπεισέρχονται στην έρευνα, την επηρεάζουν προς τη μια ή την άλλη κατεύθυνση, αλλά πιθανά δεν είναι συχνά εύκολο να εντοπισθούν, να απομονωθούν ή να ελεγχθούν. Αυτές οι μεταβλητές ονομάζονται στην ελληνική βιβλιογραφία *παρασιτικές ή αστάθμητες μεταβλητές*, ενώ στη διεθνή βιβλιογραφία αποδίδονται με τον όρο *confounding variables*. Είναι όλοι εκείνοι οι παράγοντες που στο πλαίσιο μιας έρευνας έχουν ‘υπόγεια’, γενικά μη ελέγξιμη, αλλά κάποτε όμως καταλυτική δράση. Ας δούμε ένα παράδειγμα. Υποθέτουμε ότι θέλουμε να μετρήσουμε την απόδοση μιας συγκεκριμένης ομάδας φοιτητών σε κάποιο μάθημα. Αν η εξέταση έγινε το πρωί ή αργά το απόγευμα, αυτό είναι ένας παράγων ο οποίος προφανώς επηρέασε την απόδοση των φοιτητών. Αν η αίθουσα ήταν παγωμένη, δηλ. χωρίς θέρμανση ή τα θέματα δόθηκαν με αρκετή καθυστέρηση, ή οι φοιτητές ήταν κουρασμένοι από τις εξετάσεις προηγούμενων μαθημάτων, αν αντίθετα όλες οι προηγούμενες συνθήκες ήταν καλές έως ιδανικές, προφανώς όλοι αυτοί είναι λίγο ως πολύ αστάθμητοι παράγοντες οι οποίοι όμως προφανώς επηρέασαν την απόδοση των φοιτητών.

Όταν διεξάγεται μια Κοινωνική Έρευνα, συχνά δεν είμαστε βέβαιοι για τις τιμές που μπορεί να πάρει μια συγκεκριμένη μεταβλητή. Αυτό εκ πρώτης όψεως φαίνεται παράδοξο, αλλά γίνεται σαφές από το εξής παράδειγμα: Ας υποθέσουμε ότι είναι στόχος μας να προσδιορίσουμε των αριθμό των εισακτέων στο Πανεπιστήμιο Κρήτης κάθε χρόνο. Θεωρητικά αυτός ο αριθμός, μπορεί να είναι από μηδέν μέχρι άπειρο. Αν ορίσουμε λοιπόν μια μεταβλητή με το όνομα

EIS, αυτή είναι φυσικά μια ποσοτική διακριτή μεταβλητή, τις τιμές της οποίας όμως δεν μπορούμε επακριβώς να προσδιορίσουμε από την αρχή, από ΠΡΙΝ ! Και λέμε από ΠΡΙΝ, διότι κάθε χρόνο, δεν είναι γνωστή η βούληση του Υπουργείου Παιδείας ή καλύτερα η πολιτική βούληση της συντεταγμένης Πολιτείας, δεν είναι γνωστές οι ανάγκες της αγοράς εργασίας ή οι δυνατότητες των διάφορων Τμημάτων του Πανεπιστημίου Κρήτης να εκπαιδεύσουν ένα μεγαλύτερο ή μικρότερο αριθμό φοιτητών, και επομένως, μόνο εκτιμήσεις πιθανότητας θα μπορούσαμε να κάνουμε για τις τιμές που θα μπορούσε να λάβει η παραπάνω μεταβλητή. Έτσι,

- ▮ Θα λέμε **τυχαία μεταβλητή** (random variable), τη μεταβλητή της οποίας οι τιμές δεν μπορούν να προσδιορισθούν με ακρίβεια, αλλά μέσω μιας διαδικασίας στην οποία σε κάθε τιμή της μεταβλητής αντιστοιχεί μια τιμή πιθανότητας.
- ▮ Ανάλογα, θα λέμε **μη τυχαία¹ μεταβλητή** (fixed variable), τη μεταβλητή της οποίας οι τιμές μπορούν με ακρίβεια να προσδιοριστούν πριν από τη μέτρησή τους. Για παράδειγμα, η ταχύτητα ενός αυτοκινήτου το οποίο κινείται με σταθερή² επιτάχυνση είναι γνωστή ανά πάσα στιγμή από γνωστούς μαθηματικούς τύπους. Επομένως, δεν χρειάζεται καν να τη μετρήσουμε σε κανένα σημείο της διαδρομής του αυτοκινήτου ή σε κάποιες καθορισμένες χρονικές στιγμές, οπότε η αντίστοιχη μεταβλητή μπορεί να λάβει (τελείως) προκαθορισμένες τιμές.

1.3 Τι εννοούμε με τον όρο παρατηρησιακή μονάδα

Αυτό που παρατηρούμε, ως ερευνητές, μέσα σε μια έρευνα, είναι δυνατόν να είναι ένας άνθρωπος, ένα μικρόβιο, ένα πειραματόζωο, ένα πολιτικό υποκείμενο, ένας ασθενής, ένα μαθητευόμενο άτομο, ένα φυτό, ένας αθλητής ή ένα βρέφος. Όλα αυτά ονομάζονται **παρατηρησιακές μονάδες**, διότι οι επιδόσεις τους (σκόρ), ως προς κάποιο χαρακτηριστικό ή ιδιότητά τους, είναι αυτό που ενδιαφέρει την έρευνά μας.

-
- 1 Στη διεθνή βιβλιογραφία, σπάνια συναντά κανείς την μη τυχαία μεταβλητή με τον όρο non random variable, ενώ συχνότατα ή κατά κόρον τη συναντά με τον όρο fixed variable.
 - 2 Εννοούμε σταθερή διανυσματικά δηλ. κατά μέτρο, διεύθυνση και φορά.

1.4 Πληθυσμός και Δείγμα

Οι οπαδοί μιας ποδοσφαιρικής ομάδας, οι πάσχοντες από μεσογειακή αναιμία, οι πιστοί μιας ενός θρησκευτικού δόγματος, οι καπνιστές, το γυναικείο φύλο, το ανδρικό φύλο, οι μαθητές μιας συγκεκριμένης εκπαιδευτικής βαθμίδας, οι έλληνες αρσιβαρίστες, οι οπαδοί μιας συγκεκριμένης πολιτικής παράταξης, οι δίδυμοι σε όλο τον κόσμο, τα βρέφη μέχρι 6 μηνών, κ.τ.λ. προφανώς συγκροτούν συγκεκριμένες πληθυσμιακές ομάδες, δηλ. **πληθυσμούς** λιγότερο ή περισσότερο προσδιορισίμους σε ό,τι αφορά το μέγεθός τους, δηλ. τον ακριβή αριθμό των μελών τους.

Ωστόσο, αυστηρά μιλώντας, θα λέγαμε ότι δεν είναι δυνατόν να υπάρξει ορισμός για τον πληθυσμό, ούτε για το δείγμα, αν προηγουμένως δεν ξεκαθαρίσουμε πιο πράγμα μας ενδιαφέρει να μελετήσουμε. Αυτό είναι το πρώτο μας μέλημα, σε μια έρευνα. Δηλ. να καθορίσουμε **τι θέλουμε να μελετήσουμε**. Για παράδειγμα, αν θέλουμε να μελετήσουμε την συμπεριφορά του εκλογικού σώματος στη Β' περιφέρεια Αθηνών, τότε όλοι οι πολίτες αυτής της περιφέρειας άνω των 18 ετών, συγκροτούν ένα πληθυσμό. Αλλά όλους αυτούς τους πολίτες είναι αδύνατον να τους παρατηρήσουμε, να τους ρωτήσουμε και τελικά να καταγράψουμε τις πολιτικές τους στάσεις, προθέσεις και συμπεριφορές. Γι' αυτό είναι ανάγκη να πάρουμε από αυτούς ένα δείγμα. Ένα δείγμα όμως που να τους αντιπροσωπεύει και το οποίο θα έχει ληφθεί, οπωσδήποτε με τυχαίο τρόπο. Αυτό το δείγμα, προφανώς είναι δυνατόν να το παρατηρήσουμε και να το εξετάσουμε διεξοδικά και εξονυχιστικά, διότι κάτι τέτοιο μας επιτρέπει το σχετικά μικρό του μέγεθος.

Αντικειμενικός σκοπός μας, είναι να προβούμε σε συμπεράσματα για την **ευρύτερη** ομάδα των ατόμων της έρευνάς μας (πληθυσμός), την οποία έτσι κι αλλιώς δεν μπορούμε να παρατηρήσουμε και να ελέγξουμε, μελετώντας, παρατηρώντας και ελέγχοντας, τη **στενότερη** ομάδα των ατόμων που ενδιαφέρει την έρευνά μας (δείγμα). Με απλά λόγια, όλα όσα θα παρατηρήσουμε, θα καταγράψουμε και θα μετρήσουμε στο δείγμα, θα θέλαμε να αποδοθούν ως χαρακτηριστικά και ιδιότητες του υπό μελέτη πληθυσμού, ο οποίος στην περίπτωση μας είναι όλοι οι ψηφοφόροι της Β' Αθηνών. Τελικά, τα πορίσματα της έρευνας, οι ισχυρισμοί και οι αιτιάσεις που θα διατυπωθούν μέσα από αυτήν, ενώ θα έχουν εξαχθεί από τη μελέτη του δείγματος, θα αφορούν τον πληθυσμό! Όμως, πως είναι δυνατόν να πετύχουμε κάτι τέτοιο; Να μιλήσουμε δηλ. στο όνομα ενός πληθυσμού, ο οποίος σε ορισμένες περιπτώσεις μπορεί να είναι υπερβολι-

κά ευρύς ή άπειρος³, όπως για παράδειγμα οι πολίτες της Ευρωπαϊκής Ένωσης, βασιζόμενοι στις παρατηρήσεις ενός, ενίοτε πολύ μικρού, δείγματος; Και με την ευκαιρία, **πόσο μεγάλο** θα πρέπει να είναι αυτό το δείγμα και πόσο είναι το σφάλμα της μέτρησής μας με αυτό το συγκεκριμένο μέγεθος δείγματος;

Δεν υπάρχει κανένας λόγος ανησυχίας. Η επιστήμη της Στατιστικής έχει την απάντηση σε κάθε περίπτωση, αρκεί να ακολουθήσουμε αυστηρά τους κανόνες και τη δεοντολογία της. Οι στατιστικομαθηματικές τεχνικές, που θα αναπτύξουμε λεπτομερώς στα επόμενα, αναλαμβάνουν **το 'πέρασμα' από το δείγμα στον πληθυσμό** και προσδιορίζουν με ακρίβεια το πιθανό σφάλμα των μετρήσεών μας. Ωστόσο είναι ανάγκη να επισημάνουμε τα εξής:

- ▣ Το δείγμα θα πρέπει να είναι τυχαίο, πράγμα που σημαίνει ότι κάθε στοιχείο (παρατηρησιακή μονάδα) του πληθυσμού, θα πρέπει να έχει ίσες δυνατότητες (πιθανότητα) να συμπεριληφθεί στο δείγμα. Με άλλα λόγια, η επιλογή ενός στοιχείου του πληθυσμού, είναι ανεξάρτητη από την επιλογή ενός άλλου στοιχείου του πληθυσμού.
- ▣ Το δείγμα θα πρέπει να είναι αντιπροσωπευτικό, που σημαίνει ότι το δείγμα θα πρέπει να έχει τα χαρακτηριστικά και τις ιδιότητες του πληθυσμού από τον οποίο προέρχεται. Κι αυτό εξασφαλίζεται στην περίπτωση που έχει επιλεγεί στη βάση των επιστημονικών μεθόδων δειγματοληψίας.
- ▣ Ο πληθυσμός θα πρέπει να καθορίζεται στην αρχή της κάθε έρευνας με τη μεγαλύτερη δυνατή ακρίβεια. Μόνο τότε, είναι δυνατή η επιλογή αντιπροσωπευτικού δείγματος.

Κάθε συγκεκριμένος τρόπος λήψης ενός δείγματος συνιστά και μια τεχνική δειγματοληψίας, δηλ. μια επιστημονική μέθοδο συγκρότησης του δείγματός μας από ένα καθορισμένο πληθυσμό. Στις Κοινωνικές Επιστήμες συνήθως χρησιμοποιούνται πέντε⁴ τεχνικές δειγματοληψίας ίσων, όπως λέγεται, πιθανοτήτων:

- α) η απλή τυχαία δειγματοληψία,
- β) η συστηματική δειγματοληψία,
- γ) η δειγματοληψία κατά στρώματα,
- δ) η δειγματοληψία κατά ομάδες, και τέλος,

3 Εννοούμε εδώ ότι ο πληθυσμός είναι τόσο μεγάλος και τόσο δύσκολα καθορίσιμος, ώστε να χαρακτηρίζεται, χωρίς μεγάλη προσέγγιση, άπειρος.

4 Στο πλαίσιο του παρόντος εγχειριδίου αναπτύσσουμε, για ευνόητους λόγους, μόνο μία τεχνική δειγματοληψίας: την απλή τυχαία δειγματοληψία (βλ. Κεφ. 9).

ε) η δειγματοληψία κατά στάδια⁵.

Για το μέγεθος του δείγματος, δεν μπορούμε αυθαίρετα να αποφασίσουμε. Αυτό εξαρτάται από το μέγεθος του πληθυσμού και τη συγκεκριμένη τεχνική δειγματοληψίας που θα εφαρμόσουμε. Σε κάθε περίπτωση, υπάρχουν μαθηματικές συναρτήσεις, οι οποίες είναι σε θέση να μας προσδιορίσουν **το επαρκές** μέγεθος του δείγματος που απαιτεί η έρευνά μας, συνυπολογίζοντας και το αντίστοιχο σφάλμα δειγματοληψίας.

Κάθε στατιστικός δείκτης ή μέγεθος, που αφορά το δείγμα, ονομάζεται **στατιστική** (statistics), ενώ κάθε δείκτης που αφορά τον πληθυσμό, ονομάζεται **παράμετρος** (parameter).

1.5 Κλίμακες μέτρησης

Οι Stevens (1951), Torgenson (1958), κατατάσσουν σε τέσσερις κατηγορίες, τις κλίμακες μέτρησης που χρησιμοποιούνται στις ανθρωπολογικές μελέτες και γενικά στις Κοινωνικές Επιστήμες, για να μετρήσουν διάφορες μεταβλητές:

- ▣▣▣▣ Ονομαστικές ή Κατηγορικές Κλίμακες (nominal scales)
- ▣▣▣▣ Τακτικές κλίμακες (ordinal scales)
- ▣▣▣▣ Αριθμητικές ή ισοδιαστημικές κλίμακες (interval scales)
- ▣▣▣▣ Αναλογικές κλίμακες (ratio scales)

Θα λέμε **ονομαστικές τις κλίμακες μέτρησης** στις οποίες η κατάταξη των υποκειμένων γίνεται σε καλά προσδιορισμένες, σαφώς διακρίσιμες μεταξύ τους, ισοδύναμες, και οπωσδήποτε αμοιβαία αποκλειόμενες, στην περίπτωση διχοτομικής κλίμακας μέτρησης, κατηγορίες. Με άλλα λόγια, ένα υποκείμενο, ή ορθότερα μια παρατηρησιακή μονάδα, δεν είναι δυνατό να ανήκει σε περισσότερες από μία κατηγορίες, αλλά μόνο σε μία, η οποία διακρίνεται από τις υπόλοιπες με σαφή τρόπο (ως προς ένα χαρακτηριστικό ή μια ιδιότητα). Παραδείγματα κατηγορικών κλιμάκων ή καλύτερα μεταβλητών που παίρνουν τιμές από κατηγορικές κλίμακες είναι:

5 Για μια εισαγωγή στο πνεύμα και τις τεχνικές δειγματοληψίας, παραπέμπουμε τον αναγνώστη σε δύο ξένες και σε δύο ελληνικές, πολύ καλές κατά την άποψή μας, αναφορές: Williams (1978), Judd, Smith, and Kidder (1991), και Ξεκαλάκη (1995), Χαρίσης, Κιόχος, (1997).

- ▣ το φύλο (1 = άνδρας, 2 = γυναίκα),
- ▣ ο τόπος διαμονής (1 = αστικό, 2 = ημιαστικό, 3 = αγροτικό),
- ▣ το θρήσκευμα (1 = Χριστιανός, 2 = Βουδιστής, 3 = Μουσουλμάνος, 4 = Ινδουϊστής, κτλ.),
- ▣ η πολιτική τοποθέτηση (1 = δεξιός, 2 = κεντρώος, 3 = αριστερός κτλ.).

Διευκρινήσεις

- i. Ισχυριστήκαμε παραπάνω ότι οι κατηγορίες ‘είναι αμοιβαία αποκλειόμενες’, στην περίπτωση της κατηγορικής μεταβλητής με δύο κατηγορίες, και αυτό φαίνεται από το εξής παράδειγμα:
Ας θεωρήσουμε τη μεταβλητή *pregnant* με τιμές 1 = NAI, και 2 = OXI. Τότε κάθε γυναίκα της έρευνάς μας θα ενταχθεί υποχρεωτικά σε **μία** από τις δύο κατηγορίες που περιλαμβάνει η μεταβλητή, δηλ. αν εγκυμονεί θα ενταχθεί στην πρώτη κατηγορία, στην κατηγορία του NAI, ενώ αν δεν εγκυμονεί, θα ενταχθεί στη δεύτερη κατηγορία, στην κατηγορία του OXI, και φυσικά καμιά άλλη επιλογή ή κατηγορία⁶ δεν υπάρχει, οπότε εδώ έχουμε να κάνουμε με μια **διχοτομική κατηγορική** μεταβλητή.
- ii. Ο αναγνώστης θα πρέπει να έχει υπόψη του, ότι στις παραπάνω αναφερθείσες κατηγορικές μεταβλητές, όπως για παράδειγμα, το θρήσκευμα, οι αριθμοί 1, για την περίπτωση ‘Χριστιανός’, 2 για την περίπτωση ‘Βουδιστής’, 3 για την περίπτωση ‘Μουσουλμάνος’ κτλ., τέθηκαν από εμάς αυθαίρετα και **δεν** αποτελούν αριθμούς δηλ. τιμές, αλλά **απλά σύμβολα**. Με την ευκαιρία να επισημάνουμε ότι στο πλαίσιο του SPSS και γενικά στο πλαίσιο οποιασδήποτε στατιστικής ανάλυσης, ‘βολεύει’ να θέσουμε αριθμούς ως κατηγορίες μιας κατηγορικής μεταβλητής, και μάλιστα αυτές οι τιμές **δεν** είναι επιστημονικά ανακριβείς. Ο Howell (1997), αλλά και οι Dancey, & Reidy (1999), υποστηρίζουν ότι, μια τέτοια ‘κλιμάκωση’, δηλ. μια τέτοια μετατροπή κατηγοριών σε αριθμούς, δίνει τη δυνατότητα στον Κοινωνικό επιστήμονα να διεξάγει οποιεσδήποτε αναλύσεις επιθυμεί. Οι ίδιοι όμως υποστηρίζουν ότι, «ωστόσο, τέτοιες αναλύσεις μπορεί και να μην είναι και τόσο σπουδαίες...» και δίνουν διάφορα παραδείγματα για να στηρίξουν τον τελευταίο αυτό ισχυρισμό για τις «όχι και τόσο σπουδαίες αναλύσεις» που θα μπορούσαν να παραχθούν.

6 Προφανώς, όχι μόνο οι επιστήμονες της Στατιστικής, αλλά και κάθε λογικός άνθρωπος, είναι λίγο δύσκολο να φανταστεί μια γυναίκα λίγο –λίγο έγκυο ...

- iii. Αντίθετα με τους παραπάνω συγγραφείς, εμείς θεωρούμε ότι σε κάθε περίπτωση θα πρέπει να εκχωρούνται αριθμοί- σύμβολα σε κατηγορίες, όχι μόνο γιατί δεν έχουμε τίποτε να χάσουμε ή γιατί δεν είναι αντιεπιστημονικό, αλλά αντίθετα, σε συγκεκριμένες περιπτώσεις προχωρημένης κατηγορικής ανάλυσης, όπως Homogeneity analysis, Correspondence analysis, Hierarchical cluster analysis, Non linear canonical correlation analysis, Multidimensional Scaling, Categorical regression analysis, τουλάχιστον στο πλαίσιο του SPSS, μια τέτοια πολιτική, είναι επιβεβλημένη. Σε όλες αυτές τις περιπτώσεις αναλύσεων, όχι μόνο πρέπει να εκχωρούμε αριθμητικές τιμές στις κατηγορίες μιας ονομαστικής μεταβλητής, αλλά θα πρέπει να έχουμε υπόψη μας ότι η ‘κλιμάκωση’ των όποιων κατηγοριών μιας ονομαστικής μεταβλητής, δεν μπορεί να ξεκινά από το μηδέν, αλλά υποχρεωτικά από το ένα. Με απλούστερα λόγια, αν έχουμε την κατηγορική μεταβλητή *φύλο*, θα πρέπει να συμβολίσουμε με 1 τη μια κατηγορία, για παράδειγμα το άρρεν φύλο και 2 το θήλυ, ή αντίστροφα. Δεν επιτρέπεται δηλ. σε αυτές τις περιπτώσεις, να συμβολίσουμε με 0 το άρρεν φύλο και 1 το θήλυ ή αντίστροφα. Ή τέλος, αν έχουμε τη μεταβλητή *θρήσκευμα* δεν επιτρέπεται να συμβολίσουμε με μηδέν την κατηγορία ‘Χριστιανός’, 1 την κατηγορία ‘Βουδιστής’, 3 την κατηγορία ‘Μουσουλμάνος’, κτλ.

Θα λέμε **τακτικές κλίμακες μέτρησης** εκείνες στις οποίες η ένταξη των υποκειμένων γίνεται σε κατηγορίες σαφείς, ισοδύναμες, αλλά και διατεταγμένες μεταξύ τους. Με άλλα λόγια, οι τακτικές κλίμακες έχουν όλα τα χαρακτηριστικά των κατηγορικών κλιμάκων μέτρησης συν το στοιχείο της διάταξης. Για παράδειγμα, ας πάρουμε την περίπτωση μιας 5μελούς οικογένειας, που τα μέλη της διατάσσονται σύμφωνα με το ύψος τους, και ας υποθέσουμε πως έχουμε την παρακάτω σειρά, από το ψηλότερο μέχρι το χαμηλότερο μέλος:

- 1 = Πατέρας
- 2 = Μητέρα
- 3 = πρώτο παιδί
- 4 = δεύτερο παιδί
- 5 = τρίτο παιδί

Προφανώς εδώ οι κατηγορίες είναι διατεταγμένες, αλλά **αυτό που δεν υπάρχει** είναι το στοιχείο της ίσης απόστασης, ανάμεσα σε δύο διαδοχικές κατηγορίες. Διότι κανείς δεν μας είπε, ότι όση διαφορά ηλικίας έχει ο Πατέρας

από τη Μητέρα, έχει και το δεύτερο από το τρίτο παιδί, έχει και το πρώτο από το δεύτερο παιδί, και γενικά έχουν οι διάδοχες κατηγορίες μεταξύ τους. Υπάρχει επομένως, μια κλιμάκωση των κατηγοριών ως προς ένα ιδιαίτερο, συγκεκριμένο χαρακτηριστικό, το ύψος, αλλά οι κατηγορίες δεν ισαπέχουν, ως προς το ιδιαίτερο αυτό χαρακτηριστικό. Ένα άλλο παράδειγμα τακτικής κλίμακας, θα μπορούσε να είναι μια ομάδα αθλητών στην άρση βαρών, ας πούμε στην κατηγορία των 60 κιλών. Οι αθλητές, ανάλογα με τα κιλά που κατάφεραν να σηκώσουν κατετάγησαν 1^{ος}, 2^{ος}, 3^{ος}, 4^{ος}, κτλ., αλλά η διαφορά του πρώτου από το δεύτερο, του δευτέρου από τον τρίτο, του 7 από τον 8^ο κ.ο.κ., δεν είναι ίδια. Οι διαδοχικές βαθμίδες της κλίμακας εδώ, *δεν* ισαπέχουν. Τέλος, ένα τρίτο παράδειγμα τακτικής κλίμακας, θα μπορούσε να είναι η σημερινή διάταξη των κομμάτων στο Ελληνικό Κοινοβούλιο, ανάλογα με τη δύναμη εδρών που αυτά διαθέτουν. Όπως όλοι ξέρουμε, η αριθμητική υπεροχή ανάμεσα στο πρώτο και στο δεύτερο κόμμα, δεν είναι του ίδιου μεγέθους με αυτήν, του δευτέρου και του τρίτου κόμματος.

Διευκρινήσεις

Όλα τα παραπάνω παραδείγματα τακτικών κλιμάκων που προαναφέραμε, έχουν ένα κοινό χαρακτηριστικό. Οι πράξεις ανάμεσα στις βαθμίδες της κλίμακας δεν έχουν νόημα. Για παράδειγμα, ο μέσος όρος του πρώτου, του δευτέρου και του νι-οστού αθλητή, δεν σημαίνει κάτι. Με άλλα λόγια, δεν είναι ερμηνεύσιμος.

Ας λάβουμε όμως το εξής παράδειγμα: Σε ένα ερωτηματολόγιο πολιτικής έρευνας υπάρχει η ερώτηση: Θεωρείτε επιτυχημένη την Ελληνική Προεδρία στην Ευρωπαϊκή Ένωση ; Και με δυνατές απαντήσεις από τους ερωτώμενους :

1 = Καθόλου, 2 = λίγο, 3 = αρκετά, 4 = πολύ, 5 = πάρα πολύ.

Προφανώς η ερώτηση εντάσσεται στις τακτικές κλίμακας μέτρησης, αλλά επιπλέον ο μέσος όρος έχει νόημα, δηλ. είναι ερμηνεύσιμος. Π.χ. ένας μέσος όρος 3,9 σημαίνει ότι το σώμα των ερωτώμενων θεωρεί ότι η Ελληνική Προεδρία ήταν αρκετά ως πολύ, επιτυχημένη.

Στις τακτικές κλίμακες μέτρησης εντάσσονται και οι λεγόμενες κλίμακες Likert των οποίων οι απαντήσεις διαβαθμίζονται από το απόλυτα αρνητικό μέχρι το απόλυτα θετικό ανάλογα με το θέμα που εξετάζεται:

1 = διαφωνώ απόλυτα, 2 = διαφωνώ, 3 = αβέβαιος, 4 = συμφωνώ,
και 5 = συμφωνώ απόλυτα.

Και εδώ ο μέσος όρος είναι σαφές ότι κάτι σημαίνει, είναι ερμηνεύσιμος και επίσης είναι σαφές και εδώ, όπως και σε όλες τις περιπτώσεις τακτικών κλιμάκων, ότι οι κατηγορίες των δεδομένων κλιμακώνονται ανάλογα με την ‘ποσότητα’ του ιδιαίτερου χαρακτηριστικού που κατέχουν.

Θα λέμε *αριθμητικές ή (ισο)διαστημικές* τις κλίμακες μέτρησης στις οποίες τα υποκείμενα εντάσσονται σε σαφώς καθορισμένες, αμοιβαία αποκλειόμενες, διατεταγμένες κατηγορίες, και οι οποίες όμως έχουν και το εξής, επιπλέον χαρακτηριστικό: χρησιμοποιούν σταθερή μονάδα μέτρησης. Ο χρόνος (με μονάδες μέτρησης το λεπτό, το δευτερόλεπτο, την ώρα κτλ., η απόσταση, η ηλικία, η θερμοκρασία, σε κλίμακα Celsius ή Fahrenheit κτλ., είναι κλασικά παραδείγματα διαστημικών κλιμάκων. Ωστόσο, αν κάνουμε λόγο για μηδέν βαθμούς Κελσίου (0°C), τότε προφανώς αυτό το μηδέν σε καμιά περίπτωση δεν σημαίνει την απουσία του μετρούμενου χαρακτηριστικού, που είναι η θερμοκρασία, διότι όλοι ξέρουμε ότι στους 0°C το νερό γίνεται πάγος, επομένως έχει συγκεκριμένη τιμή το μετρούμενο χαρακτηριστικό, άρα δεν απουσιάζει. Οι μεταβλητές που μετριοούνται με μια διαστημική κλίμακα μέτρησης, έχουν όλες τις ιδιότητες των μεταβλητών που μετριοούνται με τακτικές κλίμακες, με μια πρόσθετη ιδιότητα: η διαφορά μεταξύ δύο τυχαίων διαβαθμίσεων μέσα στην κλίμακα αυτή αντιπροσωπεύει και ίσες διαφορές στο μετρούμενο χαρακτηριστικό. Για παράδειγμα, ας υποθέσουμε ότι κάνουμε έρευνα στο Δημοτικό Σχολείο και σε κάποιο τεστ αναγνωστικής ικανότητας ο μαθητής Α πήρε βαθμό 5, ο μαθητής Β πήρε βαθμό 6, ο μαθητής Γ πήρε βαθμό 8 και ο μαθητής Δ πήρε βαθμό 9. Προφανώς η βαθμολογική διαφορά του μαθητή Β από το μαθητή Α είναι η ίδια με τη διαφορά ανάμεσα στο μαθητή Δ και στο μαθητή Γ: μία μονάδα. Σε ό,τι αφορά ωστόσο την αναγνωστική ικανότητα, μπορούμε να πούμε ότι όσο καλύτερος είναι ο Β από τον Α, τόσο καλύτερος είναι και ο Δ έναντι του Γ.

Διευκρινήσεις

Στο πλαίσιο της διαστημικής κλίμακας μπορούμε να κάνουμε λόγο για διαφορές, αλλά δεν μπορούμε να κάνουμε λόγο για διπλάσιο⁷, τριπλάσιο, και γενι-

7 Στο πλαίσιο της αριθμητικής ή ισοδιαστημικής κλίμακας, είναι σαφές ότι σε καμιά περίπτωση δεν ισχύει το γνωστό «Μια σαραντάρα ίσον δύο εικοσάρες!».

κά πολλαπλάσιο, διότι η κλίμακά μας προφανώς χρειάζεται ένα γνήσιο σημείο αφετηρίας, το απόλυτο μηδέν, το οποίο όπως θα δούμε αμέσως παρακάτω, δεν έχει. Για παράδειγμα, στο τεστ αναγνωστικής ικανότητας που προαναφέραμε είναι δυνατόν να γίνει λόγος για διπλάσια βαθμολογική διαφορά ανάμεσα σε δύο ζεύγη μαθητών, αλλά δεν μπορεί να γίνει λόγος για διπλάσια επίδοση ενός μαθητή έναντι ενός άλλου. Ωστόσο, βαθμός μηδέν στο τεστ, φυσικά και δεν σημαίνει πλήρη απώλεια της αναγνωστικής ικανότητας ενός μαθητή. Η αναγνωστική ικανότητα, όσο κι αν είναι μειωμένη, δεν μπορεί να καταστεί ανύπαρκτη. Επομένως, η κλίμακά μας, στερείται απολύτου μηδενός, που σημαίνει ότι το μετρούμενο χαρακτηριστικό, η αναγνωστική ικανότητα δεν είναι δυνατόν, να μην υπάρχει παντελώς. Αυτό έχει σαν συνέπεια την αδυναμία αυτών των κλιμάκων να εξυπηρετήσουν συγκρίσεις τιμών.

Οι διαστημικές κλίμακες ομαδοποιούν τα δεδομένα μας, όπως οι ονομαστικές και οι τακτικές κλίμακες.

Οι διαστημικές κλίμακες ταξινομούν τα δεδομένα μας, όπως και οι τακτικές κλίμακες.

Διαθέτουν σταθερή μονάδα μέτρησης και επομένως είναι σε θέση να προσδιορίσουν αποστάσεις ανάμεσα στα δεδομένα μας, δυνατότητα που δεν έχουν ούτε οι τακτικές, ούτε οι ονομαστικές κλίμακες μέτρησης.

Στο πλαίσιο μιας διαστημικής κλίμακας η εκτέλεση πράξεων, όπως είναι η πρόσθεση και η διαίρεση, αποκτούν ουσιαστικό νόημα. Να τονίσουμε όμως και πάλι ότι, αυτό που μπορεί να προστεθεί και να διαιρεθεί σε μια διαστημική κλίμακα είναι τα διαστήματα δηλ. διαφορές τιμών ή ποσοτήτων και όχι αυτές καθαυτές οι ποσότητες. Οι λόγοι που συμβαίνει τούτο είναι δύο:

- α) Ενώ υπάρχει σημείο μηδέν στη διαστημική κλίμακα, τούτο ορίζεται εντελώς αυθαίρετα (ο μαθητής που δεν έλυσε κανένα πρόβλημα, δεν σημαίνει ότι έχει μηδέν αριθμητική ικανότητα)
- β) η μονάδα μέτρησης σε μια διαστημική κλίμακα, ενώ προσδιορίζεται και είναι σταθερή, είναι κι εκείνη, (όπως και το σημείο μηδέν) αυθαίρετη.

Θα λέμε **αναλογικές κλίμακες μέτρησης**, εκείνες που διατηρούν όλα τα χαρακτηριστικά των διαστημικών κλιμάκων, και επιπλέον διαθέτουν πραγματικό σημείο αναφοράς το οποίο αντιστοιχεί στο απόλυτο μηδέν. Το σημείο αυτό είναι **ένα γνήσιο σημείο**, είναι ένα εναρκτήριο σημείο, με την έννοια ότι το χαρακτηριστικό ή η ιδιότητα που μετράει η κλίμακα στο ση-

μείο αυτό, δεν υπάρχει. Κλασικά παραδείγματα αναλογικής κλίμακας είναι η ταχύτητα, η απόλυτη θερμοκρασία (βαθμοί Kelvin), η πίεση του αίματος, το βάρος, η επιτάχυνση, η μάζα κτλ.

Διευκρινήσεις

Οι αναλογικές κλίμακες μέτρησης είναι οι τελειότερες κλίμακες μέτρησης, αλλά στις Κοινωνικές Επιστήμες λίγο ή σπάνια, χρησιμοποιούνται.

Οι ιδιότητες των αναλογικών κλιμάκων είναι πέντε:

- α) Οι κατηγορίες στις οποίες εντάσσουν τα υποκείμενα της έρευνας είναι καθορισμένες και σαφώς διακρίσιμες μεταξύ τους, πράγμα που σημαίνει ότι ένα και το αυτό υποκείμενο, μπορεί να ανήκει σε μία και μόνο κατηγορία
- β) Οι κατηγορίες των δεδομένων είναι διατεταγμένες
- γ) Οι κατηγορίες των δεδομένων κλιμακώνονται ανάλογα με την 'ποσότητα' του ιδιαίτερου χαρακτηριστικού που κατέχουν
- δ) Ίσες διαφορές στις αριθμητικές τιμές της κλίμακας, αντανakλούν επίσης ίσες διαφορές στο χαρακτηριστικό ή στην ιδιότητα που μετράει η κλίμακα
- ε) Το σημείο μηδέν αντικατοπτρίζει την απουσία της ιδιότητας ή του χαρακτηριστικού που μετράει η κλίμακα.

Τόσο η διαστημική όσο και η αναλογική κλίμακα, κάνουν χρήση του σημείου μηδέν. Η διαφορά τους όμως έγκειται στο γεγονός ότι, ενώ η διαστημική κλίμακα χρησιμοποιεί αυθαίρετα ένα σημείο μηδέν, το οποίο δεν σημαίνει απουσία του μετρούμενου χαρακτηριστικού, αντίθετα η αναλογική κλίμακα χρησιμοποιεί ένα γνήσιο, απόλυτο μηδέν το οποίο είναι στην πραγματικότητα και το εναρκτήριο σημείο μέτρησης γι' αυτή την κλίμακα και ασφαλώς αντανakλά την παντελή απουσία του μετρούμενου χαρακτηριστικού.

Η αναλογική κλίμακα λαμβάνει το όνομά της από το γεγονός ότι επιτρέπει τη χρήση μαθηματικών αναλογιών ανάμεσα στις βαθμίδες της.

Στο πλαίσιο της αναλογικής κλίμακας είναι δυνατόν να γίνει λόγος για διπλάσιο, τριπλάσιο, τετραπλάσιο κτλ., επειδή ακριβώς υπάρχει σε αυτήν απόλυτο μηδέν, δηλ. έχουμε πραγματική αφετηρία μέτρησης. Με άλλα λόγια, στη γνησιότητα του μηδενός βασίζεται η χρήση των μαθηματικών αναλογιών ανάμεσα στις βαθμίδες της, πράγμα που κάνει τελειότερη όλων την αναλογική κλίμακα.

1.6 Η κατανόηση των κλιμάκων μέτρησης μέσα από την κωδικοποίηση

Σε αυτή την παράγραφο παραθέτουμε αυτούσιες ερωτήσεις από διάφορες πραγματικές έρευνες, τις οποίες θα κωδικοποιήσουμε δίνοντας ονόματα και τιμές σε αντίστοιχες μεταβλητές. Η χρήση των μεταβλητών και ο καθορισμός των τιμών που αυτές θα μπορούν να λαμβάνουν, θα βοηθήσει στη βαθύτερη κατανόηση των κλιμάκων μέτρησης.

Ερώτηση 1. Που κατατάσσετε πολιτικά τον εαυτό σας;

- ☐ Δεξιά
- ☐ Κεντροδεξιά
- ☐ Κέντρο
- ☐ Κεντροαριστερά
- ☐ Αριστερά

Για να κωδικοποιήσουμε αυτή την ερώτηση ορίζουμε μια μεταβλητή με το όνομα R1 (για παράδειγμα) και με δυνατές τιμές 1 = Δεξιά, 2 = Κεντροδεξιά, 3 = Κέντρο, 4 = Κεντροαριστερά, 5 = Αριστερά. Πρόκειται για μια **καθαρά κατηγορική μεταβλητή**, όπου το 1, το 2, το 3 κτλ. είναι απλά σύμβολα και όχι αριθμοί, τους οποίους εισάγουμε στον υπολογιστή, αντί των λέξεων Δεξιά, Κεντροδεξιά κτλ., για δική μας και μόνο διευκόλυνση. Στις πέντε κατηγορίες πολιτικής τοποθέτησης, δεν υπάρχει το στοιχείο της διάταξης, φαινομενικά μόνο αυτές είναι διατεταγμένες, καθώς ο Δεξιός πολίτης ούτε μεγαλύτερος, ούτε μικρότερος είναι από τον Κεντροδεξιό πολίτη, ούτε ο Κεντρώος είναι μεγαλύτερος ή μικρότερος του Κεντροδεξιού ή του Κεντροαριστερού πολίτη κτλ.

Ερώτηση 2. Ποια είναι η καταγωγή της συζύγου σας;

- ☐ Ελληνική
- ☐ Αλβανική
- ☐ Ανατολική Ευρώπη
- ☐ Δυτική Ευρώπη
- ☐ Άλλη χώρα (προσδιορίστε)

Όμοια σκεπτόμενοι, θα μπορούσαμε να ορίσουμε μια μεταβλητή με το όνομα R2 και με πέντε δυνατές τιμές: 1 = Ελληνική, 2 = Αλβανική, 3 = Ανατολική Ευρώπη, 4 = Δυτική Ευρώπη, και 5 = άλλη χώρα (δηλ. μια χώρα εκτός από τις παραπάνω αναφερόμενες). Η μεταβλητή R2 είναι προφανώς μια **κατηγορική μεταβλητή** και ισχύουν όλα όσα είπαμε στην περίπτωση της προηγούμενης μεταβλητής R1.

Ερώτηση 3. Ποια είναι η ηλικία σας;

- ☐ Κάτω των 18
- ☐ Από 18 μέχρι 45
- ☐ Από 46 μέχρι 67
- ☐ Πάνω από 67

Εδώ πρόκειται για μια **τακτική μεταβλητή**, καθώς οι κατηγορίες είναι σαφώς προσδιορισμένες και διακρίσιμες μεταξύ τους, αλλά επιπλέον έχουν το στοιχείο της διάταξης. Διότι τα υποκείμενα της έρευνας που εντάσσονται στην πρώτη κατηγορία (κάτω των 18 ετών), σαφώς έχουν μικρότερη ηλικία, από εκείνα εντάσσονται στη δεύτερη κατηγορία (από 18 μέχρι 45) κτλ., οπότε μπορούμε να ορίσουμε μια μεταβλητή με το όνομα R3, με δυνατές τιμές: 1 = κάτω των 18, 2 = από 18 μέχρι 45, 3 = από 45 μέχρις 67 και 4 = από 67 και άνω.

Ερώτηση 4. Ποια είναι η ηλικία σας; (σε έτη)

.....

Προφανώς δεν είναι δυνατόν η έρευνα να ρωτά ένα ανύπαρκτο άτομο, οπότε για ηλικία μηδέν δεν μπορεί να γίνει λόγος. Επομένως, η αναλογική κλίμακα μέτρησης αποκλείεται εμμέσως πλην σαφώς, οπότε στην περίπτωση αυτή μπορούμε να ορίσουμε μια **αριθμητικού τύπου μεταβλητή** με το όνομα R4, η οποία θα παίρνει τιμές = αριθμούς από μια ηλικία και πάνω (αυτή που ακριβώς ενδιαφέρει τη συγκεκριμένη έρευνα). Θα πρέπει, ωστόσο να πληροφορήσουμε τον αναγνώστη, ότι όπως θα δούμε και παρακάτω, το στατιστικό πρόγραμμα SPSS δεν κάνει διάκριση ανάμεσα σε διαστημικές και σε αναλογικές κλίμακες μέτρησης, σε αριθμητικές ή αναλογικές μεταβλητές. Βάζει τόσο τις αριθμητικές, όσο και τις αναλογικές μεταβλητές σε ένα τσουβάλι, τις θεωρεί δηλ. το ίδιο πράγμα από πλευράς τύπου, και τους προσάπτει το χαρακτηρισμό **numeric variables**.

Ερώτηση 5. Πόσο ευχαριστημένος είστε από την εξωτερική πολιτική της κυβέρνησης;

- ☐ Καθόλου
- ☐ Λίγο
- ☐ Αρκετά
- ☐ Πολύ
- ☐ Πάρα πολύ

Προφανώς στην ερώτηση αυτή οι κατηγορίες είναι σαφείς, και αμοιβαία αποκλειόμενες, και επίσης έχουν το στοιχείο της διάταξης. Επομένως είναι δυνατόν να ορίσουμε μια μεταβλητή, για παράδειγμα με το όνομα R5 η οποία είναι σαφώς *τύπου ordinal* και οι δυνατές της τιμές είναι 1 = καθόλου, 2 = λίγο, 3 = αρκετά, 4 = πολύ, 5 = πάρα πολύ. Τέλος, να επισημάνουμε ότι, εάν στη συνέχεια θελήσουμε να κάνουμε πράξεις με αυτή τη μεταβλητή, πράξεις σαν κι αυτές που περιγράψαμε στην παράγραφο 1.5 ή να υπολογίσουμε το μέσο όρο ή και ακόμη να υπολογίσουμε διάφορους δείκτες περιγραφικής στατιστικής, που θα εξετάσουμε στα παρακάτω, δεν θα έχουμε κανένα πρόβλημα ερμηνείας αυτών των δεικτών.

Ερώτηση 6. Πόσα θερμοκήπια έχετε;

- 0. κανένα
- 1. ένα
- 2. δύο
- 3. τρία
- 4. τέσσερα
- 5. πέντε
- 6. έξι

Γι αυτήν την ερώτηση ορίζουμε μια μεταβλητή με το όνομα R6, με δυνατές τιμές τις 0, 2, 3, 4, 5 και 6. Η μεταβλητή R6 έχει όλα τα χαρακτηριστικά για να μας ξεγελάσει, να πιστέψουμε ότι πρόκειται για κατηγορική ή για τακτική μεταβλητή ή ακόμη περισσότερο για αριθμητική μεταβλητή. Ωστόσο η R6 ου-

σιαστικά είναι μια **αναλογική μεταβλητή**, διότι το απόλυτο και γνήσιο μηδέν (που είναι και η ειδοποιός της διαφορά με την αριθμητική μεταβλητή), είναι και υπαρκτό και σαφές. Πράγματι το μηδέν (0), ως τιμή στη μεταβλητή R6 σημαίνει **κανένα** θερμοκήπιο, και δηλώνει με σαφήνεια την απουσία του μετρούμενου χαρακτηριστικού (θερμοκήπια). Δηλ. ο απαντών με αυτό τον τρόπο (κανένα θερμοκήπιο), είναι ένας ακτήμονας σε ό,τι αφορά την ιδιοκτησία θερμοκηπίου.

Ερώτηση 7. Κατά τη γνώμη σας, όταν υπάρχει πολιτική κρίση από τι, **κυρίως**, δημιουργείται;

- ☐ Από εμάς τους ίδιους
- ☐ Από τα οργανωμένα συμφέροντα
- ☐ Από την έλλειψη κρατικής πολιτικής
- ☐ Από την Ιστορία
- ☐ Άλλο (συμπληρώστε)

Κατ' αρχήν θα πρέπει να διευκρινίσουμε ότι ο ερωτώμενος, καλείται να επιλέξει από μηδέν, μέχρι πέντε λόγους για τους οποίους προκαλείται πολιτική κρίση. Αλλά ενώ η διατύπωση της ερώτησης δεν παρουσιάζει κανένα (εμφανές) πρόβλημα, κατά την άποψη μας, θα ήταν ορθότερο ο απαντών να καλείται να απαντήσει με ένα ΝΑΙ ή ένα ΟΧΙ σε κάθε ένα από τους παραπάνω λόγους/κατηγορίες. Έτσι θα είχαμε αντίστοιχα τις εξής **5 διχοτομικές κατηγορικές μεταβλητές**:

R7_1, R7_2, R7_3, R7_4 και R7_5 ,

όλες με δύο τιμές 1 = ΝΑΙ και 2 = ΟΧΙ.

Να σημειώσουμε επίσης ότι οι ερωτήσεις σαν την ερώτηση 7, καλούνται ερωτήσεις πολλαπλής απάντησης δηλ. οι απαντήσεις μπορεί να είναι δίδυμες, τρίδυμες κτλ. ή να είναι μιας απάντησης ή τέλος να απουσιάζουν παντελώς και το SPSS διαθέτει ειδική διαδικασία για την επεξεργασία τους η οποία ονομάζεται MULTI RESPONSE ANALYSIS. (Βλ. SPSS Base 11.5 User's Guide, 2002 chapter 35 pp. 447-456).

Ερώτηση 8. Πόσο έντονα είναι τα προβλήματα πολιτικής κρίσης, που έχουν εμφανιστεί στον τόπο σας τα τελευταία δέκα χρόνια;

- ☐ Πολύ έντονα
- ☐ Μέτρια
- ☐ Δεν υπάρχουν σημαντικά πολιτικά προβλήματα

Αρχικά να επισημάνουμε ότι λέξη ‘έντονα’ από μόνη της συχνά είναι ικανή να προϊδεάσει, τουλάχιστον τακτική κλίμακα. Επίσης να σημειώσουμε ότι εμείς δεν θα θέταμε τις τρεις κατηγορίες της ερώτησης 8 με αυτήν τη σειρά. Θα προτιμούσαμε την εξής σειρά:

1 = δεν υπάρχουν σημαντικά πολιτικά προβλήματα

2 = υπάρχουν μέτρια πολιτικά προβλήματα

3 = υπάρχουν έντονα πολιτικά προβλήματα

Έτσι ο μέσος όρος, θα ήταν ευκολότερα ερμηνεύσιμος. Υψηλός μέσος όρος στην περίπτωση της διάταξης και της κωδικοποίησης που προτείνουμε, π.χ. 2,7 ασφαλώς θα υπαινίσσονταν έντονα πολιτικά προβλήματα, ενώ χαμηλός τοιούτος θα έδειχνε ανυπαρξία έντονων και σημαντικών πολιτικών προβλημάτων.

Έτσι λοιπόν, είμαστε σε θέση, να ορίσουμε μια μεταβλητή R8, σαφέστατα τακτικού τύπου και με δυνατές τιμές 1,2 και 3.

Στο σημείο αυτό θα δείξουμε ότι ένα και το αυτό ζήτημα, είναι δυνατόν να διατρέξει τις γνωστές κλίμακες μέτρησης, και να συμβάλλει στο ξεκαθάρισμα τους, αρκεί κάθε φορά που διατυπώνεται ως ερώτημα να έχει την κατάλληλη δομή. Ας πάρουμε για παράδειγμα το κάπνισμα. Οι επόμενες τρεις ερωτήσεις αντιστοιχούν, όπως είναι εύκολο να διαπιστώσουμε, σε τρεις διαφορετικές κλίμακες μέτρησης και βέβαια αντανακλούν τρεις διαφορετικούς τύπους μεταβλητών.

Ερώτηση 9. Καπνίζετε;

- ☐ ΝΑΙ
- ☐ ΟΧΙ

Για αυτή την ερώτηση θα μπορούσαμε να ορίσουμε μια κατηγορική, διχοτομική μεταβλητή με το όνομα R9, με δύο δυνατές τιμές 1=ΝΑΙ και 2=ΟΧΙ, όπου τα 1 και 2 αντανakλούν κατηγορίες, είναι απλά σύμβολα, όπως έχουμε και παραπάνω επισημάνει, αλλά που εισάγονται με ευκολία και οικονομία στον Η/Υ.

Ερώτηση 10. Πόσο καπνίζετε;

- ☐ Καθόλου
- ☐ Λίγο
- ☐ Μέτρια
- ☐ Πολύ
- ☐ Πάρα Πολύ

Κωδικοποιούμε αυτή την ερώτηση ορίζοντας μια τακτική μεταβλητή, για παράδειγμα με το όνομα R10, και με 5 δυνατές τιμές:

- 1 = καθόλου
- 2 = λίγο
- 3 = μέτρια
- 4 = πολύ
- 5 = πάρα πολύ

Ερώτηση 11. Πόσα τσιγάρα καπνίζετε την ημέρα;

Η ερώτηση μπορεί να κωδικοποιηθεί με τον ορισμό μιας αριθμητικής μεταβλητής, για παράδειγμα με το όνομα R11, και προφανώς μπορεί να λαμβάνει τιμές θεωρητικά από μηδέν μέχρι άπειρο. Είναι κι εδώ σαφές ότι έχουμε να κάνουμε με μια αριθμητική ή πιο σωστά με μια αναλογική κλίμακα μέτρησης, καθώς κάλλιστα μπορεί να λάβει τιμή μηδέν, πράγμα το οποίο είναι δυνατόν να ερμηνευτεί ως απουσία του μετρούμενου χαρακτηριστικού (κάπνισμα).

1.7 Αξιοπιστία (Reliability) και Εγκυρότητα (Validity)

Πριν αναφερθούμε σε αυτές τις δύο έννοιες, (αξιοπιστία και εγκυρότητα) είναι ανάγκη να ξεκαθαρίσουμε τι εννοούμε με τον όρο μέτρηση και τις συναφείς με αυτόν έννοιες και διαδικασίες.

Θα λέμε **μέτρηση**, κάθε εμπειρική διαδικασία η οποία καταλήγει στην αντιστοίχιση υποκειμένων με αριθμούς, ή γενικότερα με σύμβολα, σύμφωνα με ορισμένους κανόνες. Φυσικά, η αντιστοίχιση αυτή, θα πρέπει να είναι όσο το δυνατόν σαφής και ακριβής. Αυτό όμως, πως μπορούμε να το εξασφαλίσουμε; Προφανώς, θα πρέπει να αναζητήσουμε το σωστό εργαλείο μέτρησης, δηλ. το εργαλείο μέτρησης που διακρίνεται από τα παρακάτω χαρακτηριστικά:

- ▀ Ικανότητα να διακρίνει διαφορές στην τάξη μεγέθους η οποία είναι επιθυμητή από την έρευνά μας. Για παράδειγμα, αν θέλουμε να μετρήσουμε κλάσματα δευτερολέπτου προφανώς δεν είναι κατάλληλο ένα κοινό ρολόϊ.
- ▀ Αντικειμενικότητα, δηλ. όποιος κι αν το χειριστεί να βγάλει την ίδια τιμή, δηλ. το ίδιο αποτέλεσμα, αν η εργασία γίνεται κάτω από τις ίδιες συνθήκες. Για παράδειγμα, δεν είναι δυνατόν να μετρήσουμε την επίδοση των φοιτητών στη Στατιστική με θέματα διαφορετικής δυσκολίας στις διάφορες εξεταστικές αίθουσες.
- ▀ Πρόληψη της στρέβλωσης/ παραμόρφωσης των τιμών. Από το όνομα ενός ανθρώπου και προ παντός από τα 2-3 τελευταία γράμματα δεν είναι δυνατόν πάντοτε να προβλέψουμε το φύλο του.
- ▀ Και τέλος να είναι απλό στη χρήση του, ώστε να αποφεύγονται τα λάθη χειρισμού.

Τις παραπάνω όμως αρετές του οργάνου μέτρησης, όσο κι αν τις επιδιώξουμε είναι αδύνατο να τις πετύχουμε όλες. Στον ένα ή στον άλλο βαθμό, τη μία ή την άλλη αρετή θα πετύχουμε, σε βάρος πιθανά των υπολοίπων. Έτσι είναι αναπόφευκτα τα σφάλματα. Σε κανένα επιστημονικό κλάδο έρευνας, δεν είναι δυνατόν να ελεγχθούν πλήρως οι συνθήκες κάτω από τις οποίες γίνεται μια μέτρηση. Τι κάνουμε επομένως; Ποιος θα πρέπει να είναι ο βασικός μας στόχος όταν εργαζόμαστε στο πλαίσιο μιας έρευνας; Μα προφανώς, η όσο το δυνατόν μείωση των σφαλμάτων.

Γενικά μιλώντας, θα διακρίνουμε δύο είδη σφαλμάτων: το συστηματικό και το τυχαίο.

▮ **Συστηματικό** είναι το σφάλμα το οποίο κάνει τις ατομικές βαθμολογίες να μεταβάλλονται με σταθερό και ομοιόμορφο τρόπο από τη μια στην άλλη (επόμενη) μέτρηση. Έτσι από μέτρηση σε μέτρηση συστηματικά μπορεί να πηγαίνουμε σε υψηλότερες ή σε χαμηλότερες επιδόσεις. Ανάλογα. Διότι, αν έχουμε για παράδειγμα, να εξετάσουμε ως προς την αναγνωστική ικανότητα κάποια άτομα, είναι σαφές ότι μπορεί, από εξέταση σε εξέταση να επιφέρει η εξάσκηση μια σταθερή και προοδευτικά αυξανόμενη βελτίωση στην επίδοσή τους. Αντίθετα, ίσως να έχουμε από εξέταση σε εξέταση πλήξη ή κόπωση, παράγοντες που μπορεί να προκαλέσουν μια αύξηση των σφαλμάτων του εξεταζόμενου και συνακόλουθα μια σταθερά μειούμενη επίδοση.

▮ **Τυχαίο** είναι το σφάλμα το οποίο αποδίδεται σε αστάθμητους και γενικά απρόβλεπτους παράγοντες, ξένους προς τη μελέτη μας. Δεν είναι γνωστή από πριν η ένταση και η έκταση αυτών των παραγόντων. Άλλοτε λαμβάνουν υψηλές και άλλοτε χαμηλές τιμές, ανάλογα με την περίπτωση και τις ιδιαίτερες συνθήκες. Για παράδειγμα, η μεταβαλλόμενη προσωπική διάθεση του εξεταζόμενου, ο αυξομειούμενος περιβαλλοντικός θόρυβος, ο φωτισμός του χώρου, η θερμοκρασία του περιβάλλοντος κτλ., είναι περιστασιακοί τυχαίοι παράγοντες που επιδρούν στις μετρήσεις, προκαλώντας αυτά που αποκαλέσαμε τυχαία σφάλματα.

Συχνότατα στις έρευνες κάνουμε την παραδοχή ότι τα σφάλματά μας δεν είναι συστηματικά και σε ό,τι αφορά τα τυχαία σφάλματα προσπαθούμε κατά το δυνατόν να τα ελέγξουμε.

Με βάση τα παραπάνω και μιλώντας (πάλι) γενικά θα μπορούσαμε να ορίσουμε τις έννοιες της αξιοπιστίας και της εγκυρότητας, πότε δηλ. οι μετρήσεις μας είναι αξιόπιστες⁸ και έγκυρες.

⁸ Ο Μακράκης (1997), κάνει μια αξιοπρόσεκτη (θεωρητική) ανάλυση της έννοιας της αξιοπιστίας με τη χρήση μεταβλητών που ορθά αποκαλεί **προσδιοριστικές**, οι οποίες αποτελούν μέτρα προσδιορισμού για μια έννοια, προκύπτουν από μια σειρά παρατηρήσεων και όσο μεγαλύτερος είναι ο αριθμός τους τόσο περισσότερο αξιόπιστη είναι μια μέτρηση (όπ. παρ. σελ. 39). Η ανάλυση καταλήγει στον προσδιορισμό του συντελεστή αξιοπιστίας Cronbach α, οποίος ελέγχει την αξιοπιστία των συνθετικών μεταβλητών που



Θα λέμε ότι μια μέτρηση *είναι αξιόπιστη*, αν φανερώνει την παραγωγή των ίδιων αποτελεσμάτων, μέσα από διαδικασίες επαναλαμβανόμενων μετρήσεων, με τη χρήση ίδιων οργάνων, με τα ίδια πειραματικά υποκείμενα, και κάτω από τις ίδιες συνθήκες.

Θα λέμε ότι μια μέτρηση *είναι έγκυρη*, αν μετράει αυτό που θέλει να μετρήσει. Οι Σίμος και Κομίλη (2003), ορθά αναφέρουν ότι η έννοια της εγκυρότητας δύσκολα μπορεί να ορισθεί με μία⁹ μόνο προσέγγιση. Ο Cronbach (1971, pp. 443-507), διακρίνει τρία είδη εγκυρότητας: την εγκυρότητα περιεχομένου, την εγκυρότητα εννοιολογικής κατασκευής, και τη σχετιζόμενη με κριτήριο εγκυρότητα.

δημιουργούνται από περισσότερες από δύο προσδιοριστικές μεταβλητές και η τιμή του κυμαίνεται από μηδέν μέχρι 1. Πολύ ικανοποιητικός συντελεστής αξιοπιστίας θεωρείται ένας συντελεστής της τάξης του 80%.

Οι Σίμος και Κομίλη (2003), επίσης κάνουν μια αξιοπρόσεκτη (θεωρητική, τεχνική και μαθηματική) ανάλυση της έννοιας της αξιοπιστίας, καταδεικνύοντας με σαφή τρόπο τις διάφορες εκδοχές ή εκφάνσεις της, αλλά και τις χρησιμοποιούμενες τεχνικές στο μαθηματικό επίπεδο που οδηγούν στον υπολογισμό των διάφορων εκδοχών αξιοπιστίας. Στα στενά πλαίσια αυτού του βιβλίου, δεν είναι δυνατή η ανάπτυξη αυτών των τεχνικών. Δυστυχώς δεν είναι δυνατή και η σύνθεση των απόψεων των τελευταίων ερευνητών με εκείνες του Μακράκη (οπ. Παρ), γεγονός που θα είχε ζωηρό ενδιαφέρον για τους νέους ερευνητές. Γι' αυτό θα περιοριστούμε απλά στο γενικό μαθηματικό ορισμό της αξιοπιστίας, γύρω από τον οποίο περιστρέφεται όλη η υπονοούμενη ανάλυση:

$$\text{Αξιοπιστία} = \frac{\text{διακύμανση των πραγματικών τιμών της εξαρτημένης μεταβλητής}}{\text{διακύμανση των παρατηρούμενων τιμών της εξαρτημένης μεταβλητής}}$$

Κλείνουμε την παραπομπή μας με την πληροφορία ότι το πρόγραμμα SPSS διαθέτει μια διαδικασία που φέρει το όνομα RELIABILITY ANALYSIS, με τη βοήθεια της οποίας είναι δυνατός ο προσδιορισμός του συντελεστή αξιοπιστίας Cronbach α και των συναφών τεχνικών υπολογισμού της αξιοπιστίας.

9 Οι παραπάνω ερευνητές τεκμηριώνουν την άποψή τους αυτή στις σελίδες 74-75 του βιβλίου τους (Σίμος, Κομίλη, 2003).