

- 1.1 Γιατί χρησιμοποιούμε τη στατιστική;
- 1.2 Τι είναι η στατιστική;
- 1.3 Περισσότερα για την επαγωγική στατιστική
- 1.4 Τρεις τύποι δεδομένων
- 1.5 Επίπεδα μέτρησης
- 1.6 Τύποι μεταβλητών
- 1.7 Πώς θα χρησιμοποιήσετε αυτό το βιβλίο

Περίληψη / Σημαντικοί όροι / Ερωτήσεις επανάληψης

Πρόλογος

Η στατιστική ασχολείται με τη μεταβλητότητα. Ο καθένας είναι διαφορετικός απ' οποιονδήποτε άλλον (και ελπίζουμε περήφανος γι' αυτό). Το σήμερα διαφέρει τόσο από το χθες όσο και από το αύριο. Σε ένα πείραμα που σχεδιάζεται έτσι ώστε να ανιχνεύει αν η ψυχοθεραπεία ενισχύει την αυτοεκτίμηση, οι βαθμοί αυτοεκτίμησης θα διαφέρουν μεταξύ των αντικειμένων του πειράματος, ανεξάρτητα από το αν η ψυχοθεραπεία βελτιώνει την αυτοεκτίμηση.

Ξεκινώντας από το Κεφάλαιο 2, η περιγραφική στατιστική θα παράσχει εργαλεία, όπως πίνακες, διαγράμματα και μέσους όρους, τα οποία σας βοηθούν να περιγράψετε και να οργανώσετε την αναπόφευκτη μεταβλητότητα μεταξύ των παρατηρήσεων. Για παράδειγμα, οι βαθμοί αυτοεκτίμησης (σε μια κλίμακα από 0 ως 50) για μια ομάδα φοιτητών θα μπορούσαν να σχηματίζουν μια κωδωνοειδή καμπύλη με μέσο βαθμό 32 και εύρος βαθμών από 18 ως 49.

Ξεκινώντας από το Κεφάλαιο 8, η επαγωγική στατιστική θα παράσχει σημαντικές έννοιες οι οποίες, μέσα από την απαραίτητη προσαρμογή για τις διεισδυτικές επιδράσεις της μεταβλητότητας, θα σας επιτρέπουν να γενικεύετε τα περιορισμένα σύνολα παρατηρήσεων. Για παράδειγμα, η επαγωγική στατιστική μπορεί να μας βοηθήσει να αποφασίσουμε αν –μετά από μια προσαρμογή που επιβάλλεται για τη μεταβλητότητα ιστορικού (ή την πιθανότητα)– μια παρατηρηθείσα βελτίωση στους βαθμούς αυτοεκτίμησης θα μπορούσε να αποδοθεί στην ψυχοθεραπεία και όχι στην πιθανότητα.

Το Κεφάλαιο 1 παρέχει μια επισκόπηση τόσο της περιγραφικής όσο και της επαγωγικής στατιστικής και παρουσιάζει επίσης αρκετούς όρους –μερικούς από τη στατιστική και μερικούς από τα μαθηματικά αλλά και τις ερευνητικές μεθόδους– τους οποίους ενδεχομένως να γνωρίζετε ήδη. Αυτοί οι όροι θα δώσουν απαντήσεις σε αρκετές σημαντικές απορίες και θα σας συνδράμουν στην κατανόηση καθώς θα προχωράτε στο παρόν σύγγραμμα.

1.1 Γιατί χρησιμοποιούμε τη στατιστική;

Το γεγονός ότι παρακολουθείτε το μάθημα της στατιστικής πιθανώς οφείλεται στο ότι είναι υποχρεωτικό και μάλλον δεν νιώθετε ιδιαίτερα ενθουσιασμένοι γι' αυτό. Θα προσπαθήσουμε να εξερευνήσουμε μερικούς από τους λόγους για τους οποίους θα σας ωφελήσει να μελετήσετε τη στατιστική. Για παράδειγμα, σε μια ημερήσια εφημερίδα ανακαλύψαμε πρόσφατα τα εξής θέματα:

- Τα ετήσια κέρδη των αποφοίτων πανεπιστημίου υπερβαίνουν, κατά μέσο όρο, εκείνα των αποφοίτων λυκείου κατά \$20.000.

- Βάσει της υπάρχουσας έρευνας, δεν υφίστανται αποδείξεις για την ύπαρξη σχέσης μεταξύ του μεγέθους μιας οικογένειας και της βαθμολογίας ανηλίκων σε μια δοκιμασία ψυχολογικής προσαρμογής.
- Οι καπνιστές υποφέρουν σημαντικά περισσότερο από αναπνευστικά προβλήματα απ' όσα άτομα δεν καπνίζουν.

Όταν θα έχετε μάθει λίγη στατιστική, δεν θα αναρωτιέστε για τις φράσεις που εμφανίζονται με πλάγια γράμματα, αλλά ούτε και θα διστάζετε να θέσετε κάποιες διευκρινίσεις ρωτώντας «ποιος μέσος όρος δείχνει υψηλότερα ετήσια κέρδη;» ή «τι συνιστά έλλειψη αποδείξεων για μια σχέση;» ή «πόσο περισσότερο είναι τα σημαντικά περισσότερα αναπνευστικά προβλήματα;».

Μια γνώση της στατιστικής είναι απολύτως αναγκαία για την κατανόηση εκθέσεων ερευνητικών εργασιών ή reports στο πεδίο του ενδιαφέροντός σας. Στατιστικές αναφορές αναφέρονται σε αρκετές ενότητες στα τμήματα αποτελεσμάτων των περισσότερων εκθέσεων ή/και ερευνών. Αυτές οι αναφορές, οι οποίες συχνά εκφράζονται εν συντομία, παρέχουν στατιστική υποστήριξη για τα συμπεράσματα των ερευνητών:

- Άτομα που ασκούνται καθημερινά παίρνουν μεγαλύτερους βαθμούς σε δοκιμασίες αυτοεκτίμησης από άτομα που δεν ασκούνται [$p < 0,05$].
- Οι υπερβολικά αγχωμένοι φοιτητές θεωρούνται λιγότερο ελκυστικοί σε σχέση με τους πιο ήρεμους φοιτητές [$t(48) = 3,21, p < 0,01, d = 0,42$].
- Η στάση απέναντι στο σεξ εκτός γάμου εξαρτάται από την κοινωνικοοικονομική κατάσταση [$\chi^2(4, n = 185) = 11,49, p < 0,05, \varphi^2_c = 0,03$].

Αφού μάθετε κάποια πράγματα για τη στατιστική, θα είστε σε θέση να αποκρυπτογραφείτε το νόημα αυτών των συμβόλων και τελικά να κατανοείτε μελετώντας αυτές τις αναφορές.

Κάποια στιγμή στο μέλλον –πιθανώς πιο σύντομα απ' όσο πιστεύετε– ίσως χρειαστεί να σχεδιάσετε μια στατιστική ανάλυση για ένα δικό σας ερευνητικό έργο. Αφού γνωρίσετε βασικές έννοιες της στατιστικής, θα μπορείτε να σχεδιάσετε τη στατιστική ανάλυση για έργα μεσαίου επιπέδου στα οποία περιλαμβάνονται άμεσα ερευνητικά ερωτήματα. Αν το έργο σας απαιτεί πιο προχωρημένη στατιστική ανάλυση, θα ξέρετε αρκετά ώστε να συμβουλευτείτε κάποιον που γνωρίζει περισσότερα πράγματα για τη στατιστική. Και όταν αρχίσετε να καταλαβαίνετε τις βασικές στατιστικές έννοιες, θα ανακαλύψετε ότι, με κάποια καθοδήγηση, οι δικές σας προσπάθειες συχνά θα σας δίνουν τη δυνατότητα να χρησιμοποιείτε και να ερμηνεύετε την πιο προχωρημένη στατιστική ανάλυση που απαιτείται για την έρευνά σας.

1.2 Τι είναι η στατιστική;

Είναι δύσκολο να φανταστούμε, ακόμα και στο πλαίσιο μιας δοκιμασίας, έναν κόσμο χωρίς μεταβλητότητα – όπου, για παράδειγμα, όλοι έχουν τα ίδια φυσικά χαρακτηριστικά, ευφυΐα, αντιλήψεις κ.λπ. Γνωρίζοντας ότι κάποιος έχει ύψος 1,80, δείκτη νοημοσύνης (IQ) 125 και διάκειται θετικά απέναντι στη θανατική ποινή, θα μπορούσαμε να συμπεράνουμε αμέσως ότι όλοι οι άλλοι θα έχουν επίσης αυτά τα χαρακτηριστικά. Ένας τέτοιος κόσμος που δεν θα έθετε τη σκέψη σε λειτουργία δεν θα είχε ανάγκη καμίας σύστασης πέραν του γεγονότος ότι δεν θα υπήρχε λόγος ύπαρξης της στατιστικής (και ταυτόχρονα πολλοί από εμάς θα αναζητούσαν εργασία).

Περιγραφική στατιστική

Η στατιστική υπάρχει εξαιτίας της επικράτησης της μεταβλητότητας στην πραγματικότητα. Στην απλούστερη μορφή της, την αποκαλούμενη **περιγραφική στατιστική**, η στατιστική μάς παρέχει εργαλεία –πίνακες, διαγράμματα, μέσους όρους, διαστήματα, συσχετίσεις– για την οργάνωση και τη σύνοψη της αναπόφευκτης μεταβλητότητας σε συλλογές πραγματικών παρατηρήσεων ή αποτελεσμάτων. Μεταξύ των παραδειγμάτων έχουμε:

1. Μια λίστα σε μορφή πίνακα του συνολικού αριθμού των ερωτικών σχέσεων κατά τη διάρκεια των σπουδών, ταξινομημένων κατά φθίνουσα σειρά, όπως αυτές δηλώθηκαν ανώνυμα από κάθε φοιτητή που παρακολούθει το μάθημα της στατιστικής.

2. Ένα διάγραμμα που δείχνει την ετήσια μεταβολή της θερμοκρασίας παγκοσμίως τα τελευταία 30 χρόνια.
3. Μια έκθεση που περιγράφει τη μέση διαφορά στον μέσο όρο μαθημάτων (GPA) μεταξύ των φοιτητών που καταναλώνουν τακτικά αλκοόλ, καθώς όμως και των υπολοίπων.

Επαγωγική στατιστική

Η στατιστική παρέχει επίσης εργαλεία –διάφορες δοκιμασίες, ελέγχους και εκτιμήσεις– για τη γενίκευση πέρα από συλλογές πραγματικών παρατηρήσεων. Αυτό το πιο προχωρημένο πεδίο ονομάζεται **επαγωγική στατιστική**. Τα εργαλεία της επαγωγικής στατιστικής μάς επιτρέπουν να χρησιμοποιούμε μια σχετικά μικρή συλλογή πραγματικών παρατηρήσεων προκειμένου να αξιολογήσουμε, για παράδειγμα:

1. Τον ισχυρισμό ενός δημοσκόπου ότι η πλειοψηφία των αμερικανών ψηφοφόρων είναι υπέρ πιο αυστηρών νόμων για τον έλεγχο της οπλοκατοχής.
2. Την υπόθεση ενός ερευνητή ότι, κατά μέσο όρο, όσοι διαλογίζονται έχουν λιγότερους πονοκεφάλους από τους υπόλοιπους.
3. Έναν ισχυρισμό για τη σχέση ανάμεσα στην ικανοποίηση από την εργασία και τη συνολική ευτυχία.

Σ' αυτό το βιβλίο θα συναντήσετε τα πιο απαραίτητα εργαλεία της περιγραφικής στατιστικής (Μέρος 1), ξεκινώντας από το Κεφάλαιο 2, και εκείνα της επαγωγικής στατιστικής (Μέρος 2), ξεκινώντας από το Κεφάλαιο 8.

Έλεγχος προόδου *1.1 Δείξτε αν καθεμία από τις παρακάτω προτάσεις εμπίπτει στο αντικείμενο της περιγραφικής στατιστικής (επειδή περιγράφει σύνολα πραγματικών παρατηρήσεων) ή της επαγωγικής στατιστικής (επειδή μπορεί να γενικευτεί πέρα από σύνολα πραγματικών παρατηρήσεων).

(α) Οι φοιτητές στην τάξη στατιστικής μου είναι, κατά μέσο όρο, 23 ετών.

(β) Ο παγκόσμιος πληθυσμός υπερβαίνει τα 7 δισεκατομμύρια (δηλαδή 7.000.000.000 ή 1 εκατομμύριο επί 7.000).

(γ) Τέσσερα ή οκτώ χρόνια είναι η πιο συχνή διάρκεια της θητείας των αμερικανών προέδρων.

(δ) Το 64% όλων των φοιτητών είναι θετικοί απέναντι στους νόμους υπέρ του δικαιώματος έκτρωσης.

Απαντήσεις στη σελίδα 529.

1.3 Περισσότερα για την επαγωγική στατιστική

Πληθυσμοί και δείγματα

Η επαγωγική στατιστική αφορά τη γενίκευση πέρα από σύνολα πραγματικών παρατηρήσεων, δηλαδή τη γενίκευση από ένα δείγμα σε έναν πληθυσμό. Στη στατιστική, ο **πληθυσμός** είναι οποιαδήποτε ολοκληρωμένη «συλλογή» παρατηρήσεων ή πιθανών παρατηρήσεων, ενώ το **δείγμα** είναι οποιαδήποτε μικρότερη «συλλογή» πραγματικών παρατηρήσεων που εξάγονται από έναν πληθυσμό και αποτελεί υποσύνολό του. Στην καθημερινότητα, οι πληθυσμοί πολλές φορές θεωρούνται συλλογές πραγματικών αντικειμένων (π.χ., αριθμός ατόμων, φώκιες, αυτοκίνητα), ενώ, στη στατιστική, οι πληθυσμοί θεωρούνται, με πιο αφηρημένο τρόπο, ιδιότητες ή μετρήσεις (π.χ., η εθνική καταγωγή ανθρώπων, η διάρκεια ζωής της φώκιας, η κατανάλωση καυσίμων των αυτοκινήτων).

Ανάλογα με την οπτική σας, ένα δεδομένο σύνολο παρατηρήσεων μπορεί να είναι πληθυσμός ή δείγμα. Για παράδειγμα, τα βάρη που αναφέρουν 53 άρρενες φοιτητές στατιστικής στον Πίνακα 1.1 μπορούν να θεωρηθούν πληθυσμός, επειδή ανησυχείτε για την υπέρβαση του ορίου του φορτίου που μπορεί να μεταφέρει ένα σκάφος (το οποίο έχει μισθωθεί από τους 53 φοιτητές για τον εορτασμό της περαίωσης των μαθημάτων τους), ή δείγμα ενός πληθυσμού, επειδή θέλετε να γενικεύσετε τα βάρη όλων των αρρένων φοιτητών στατιστικής ή όλων των αρρένων φοιτητών.

Κανονικά, οι πληθυσμοί είναι αρκετά μεγάλοι σε μέγεθος και υπάρχουν μόνο ως εν δυνάμει παρατηρήσεις (π.χ., τα εν δυνάμει αποτελέσματα όλων των φοιτητών στις ΗΠΑ σε μια εξέταση που υπολογίζει το άγχος). Από την άλλη πλευρά, τα δείγματα είναι σχετικά μικρά και υπάρχουν ως πραγματικές παρατηρήσεις (τα πραγματικά αποτελέσματα 100 φοιτητών στην εξέταση άγχους). Όταν χρησιμοποιείται ένα δείγμα (100 πραγματικά αποτε-

Πληθυσμός

Οποιαδήποτε ολοκληρωμένη «συλλογή» παρατηρήσεων ή πιθανών παρατηρήσεων.

Δείγμα

Οποιαδήποτε μικρότερη «συλλογή» πραγματικών παρατηρήσεων από έναν πληθυσμό.

Πίνακας 1.1
ΠΟΣΟΤΙΚΑ ΔΕΔΟΜΕΝΑ: ΒΑΡΗ (ΣΕ ΛΙΒΡΕΣ) ΑΡΡΕΝΩΝ ΦΟΙΤΗΤΩΝ ΣΤΑΤΙΣΤΙΚΗΣ

160	168	133	170	150	165	158	165
193	169	245	160	152	190	179	157
226	160	170	180	150	156	190	156
157	163	152	158	225	135	165	135
180	172	160	170	145	185	152	
205	151	220	166	152	159	156	
165	157	190	206	172	175	154	

λέσματα) για να υπάρξει μια γενίκευση σε έναν πληθυσμό (εκατομμύρια εν δυνάμει αποτελεσμάτων), είναι σημαντικό το δείγμα να αντιπροσωπεύει τον πληθυσμό, αλλιώς οποιαδήποτε γενίκευση μπορεί να είναι λανθασμένη.¹ Αν και η πρόσβαση σ' αυτά τα στοιχεία είναι εύκολη, τα αποτελέσματα της εξέτασης άγχους για τους 100 φοιτητές σε μαθήματα στατιστικής στη σχολή σας πιθανώς να μην είναι αντιπροσωπευτικά των αποτελεσμάτων για όλους τους φοιτητές. Αυτοί οι 100 φοιτητές στατιστικής μάλιστα θα μπορούσαν να δίνουν υψηλότερα ή χαμηλότερα αποτελέσματα από τον πληθυσμό-στόχο της μελέτης για διάφορους λόγους, όπως, για παράδειγμα, για το γεγονός ότι οι 100 φοιτητές είναι πιθανό να σπουδάζουν ψυχολογία και να παρακολουθούν ένα υποχρεωτικό μάθημα στατιστικής στη συγκεκριμένη σχολή.

Τυχαία δειγματοληψία (έρευνα)

Όποτε είναι δυνατόν, τα δείγματα θα πρέπει να επιλέγονται τυχαία από έναν πληθυσμό έτσι ώστε να αυξάνεται η πιθανότητα ότι το δείγμα θα αντιπροσωπεύει επακριβώς τον πληθυσμό. Η **τυχαία δειγματοληψία** είναι μια διαδικασία που σχεδιάζεται έτσι ώστε να διασφαλίζεται πως κάθε εν δυνάμει παρατήρηση στον

Τυχαία δειγματοληψία

Μια διαδικασία που σχεδιάζεται έτσι ώστε να διασφαλίζεται πως κάθε εν δυνάμει παρατήρηση στον πληθυσμό έχει ίση πιθανότητα να επιλεγεί σε μια έρευνα.

πληθυσμό έχει ίση πιθανότητα να επιλεγεί σε μια έρευνα. Κλασικά παραδείγματα τυχαίων δειγμάτων είναι μια κλήρωση λαχείου όπου κάθε αριθμός από το 1 ως το 99 στον πληθυσμό έχει ίση πιθανότητα να επιλεγεί ως ένας από τους πέντε νικητήριους αριθμούς, ή μια έρευνα γνώμης στην οποία κάθε αριθμός τηλεφώνου έχει ίση πιθανότητα να επιλεγεί ως αποτέλεσμα μιας σειράς τυχαίων επιλογών, ξεκινώντας με έναν τριψήφιο αριθμό περιοχής και τελειώνοντας με έναν συγκεκριμένο επταψήφιο αριθμό τηλεφώνου.

Η τυχαία δειγματοληψία μπορεί να είναι ιδιαίτερα δύσκολη όταν ένας πληθυσμός στερείται δομής (π.χ., όλα τα άτομα που υποβάλλονται τη δεδομένη στιγμή σε ψυχοθεραπεία) ή συγκεκριμένων ορίων (π.χ., όλοι οι εθελοντές που θα μπορούσαν θεωρητικά να συμμετάσχουν σε ένα πείραμα). Σ' αυτήν την περίπτωση, ένα τυχαίο δείγμα αποτελεί ιδανική περίπτωση για την οποία μπορεί να υπάρχει μόνο κατά προσέγγιση κρίση – πάντα με μια προσπάθεια να αφαιρεθούν προφανείς μεροληψίες που θα μπορούσαν να οδηγήσουν σε εσφαλμένη αντιπροσώπηση του πληθυσμού από το συγκεκριμένο δείγμα. Για παράδειγμα, η έλλειψη πόρων για να υπάρξει τυχαία δειγματοληψία του ζητούμενου πληθυσμού όλων των αμερικανών φοιτητών θα μπορούσε να οδηγήσει στη λήψη αποτελεσμάτων με τυχαία επιλογή των 100 φοιτητών, όχι μόνο από μαθήματα στατιστικής στη σχολή σας, αλλά και από άλλες σχολές, χρησιμοποιώντας πιθανώς μερικές από τις πιο προχωρημένες τεχνικές που περιγράφονται στο Κεφάλαιο 8. Στον βαθμό που το δείγμα σας προσεγγίζει μόνο ένα πραγματικά τυχαίο δείγμα, οποιαδήποτε γενίκευση προκύπτει θα πρέπει να αξιολογείται αναλόγως. Για παράδειγμα, αν οι 100 φοιτητές επιλέχθηκαν τυχαία μόνο από αρκετές δημόσιες ανώτατες σχολές στη Βόρεια Καλιφόρνια, αυτό το γεγονός θα πρέπει να επισημανθεί και οποιαδήποτε γενίκευση σε όλους τους φοιτητές των ΗΠΑ θα ήταν πρόσκαιρη αλλά και ανοιχτή στην κριτική.

Τυχαία ανάθεση (εκχώρηση) (πειράματα)

Η εκτίμηση του μέσου βαθμού άγχους για όλους τους φοιτητές πιθανώς δεν θα είχε μεγάλο ενδιαφέρον. Αντίθετα, ίσως είχε περισσότερο ενδιαφέρον να διαπιστώσουμε αν εκπαιδευτικά προγράμματα τεχνικών χαλάρωσης

1. Εισάγοντας μεροληψία. (Σ.τ.Ε.)

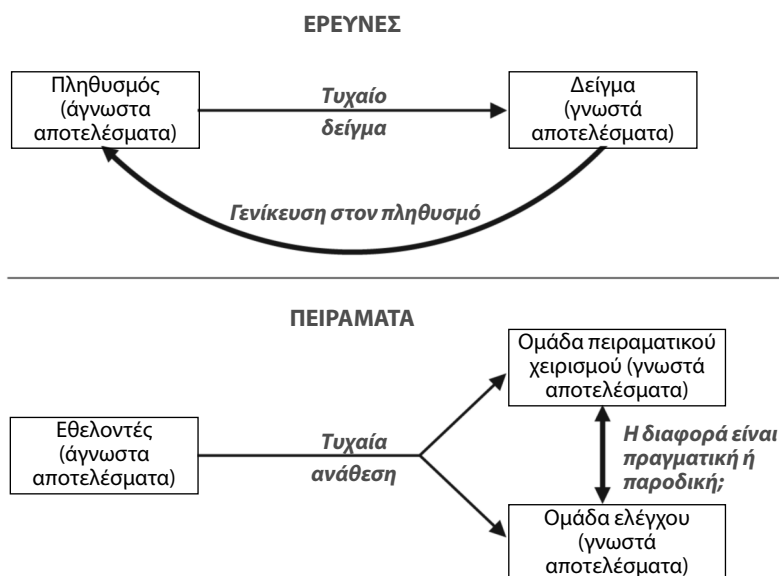
προκαλούν, κατά μέσο όρο, μείωση του άγχους σε δύο ομάδες κατά τ' άλλα παρόμοιων φοιτητών. Ακόμα κι αν τα εκπαιδευτικά προγράμματα τεχνικών χαλάρωσης δεν έχουν επίδραση στο άγχος, θα περιμέναμε ότι τα μέσα αποτελέσματα για τις δύο ομάδες θα ήταν διαφορετικά εξαιτίας της αναπόφευκτης μεταβλητότητας μεταξύ των ομάδων. Το ερώτημα διαμορφώνεται ως εξής: Πώς θα πρέπει να ερμηνεύσουμε την προφανή διαφορά μεταξύ της ομάδας πειραματικού χειρισμού και της ομάδας ελέγχου; Όταν λαμβάνεται υπόψη η μεταβλητότητα, η διαφορά θα πρέπει να θεωρείται πραγματική (και να αποδίδεται στα εκπαιδευτικά προγράμματα τεχνικών χαλάρωσης) ή παροδική (και να αποδίδεται απλώς στη μεταβλητότητα ή στην τύχη);

Οι φοιτητές στο πείραμα χαλάρωσης μάλλον δεν είναι ένα τυχαίο δείγμα από κάποιον άθικτο πληθυσμό ενδιαφέροντος, αλλά, αντίθετα, ένα *πρακτικό δείγμα* που αποτελείται από εθελοντές από μια περιορισμένη δεξαμενή φοιτητών που παρακολουθούν ένα υποχρεωτικό μάθημα. Αναλόγως, η εστίασή μας μεταπηδά από την τυχαία δειγματοληψία στην τυχαία ανάθεση (εκχώρηση) εθελοντών σε δύο ομάδες. Η **τυχαία ανάθεση (εκχώρηση)** συνεπάγεται ότι *κάθε άτομο έχει ίση πιθανότητα να τοποθετηθεί σε οποιαδήποτε ομάδα σε ένα πείραμα*. Χρησιμοποιώντας διαδικασίες που περιγράφονται στο Κεφάλαιο 8, η τυχαία ανάθεση (εκχώρηση) θα πρέπει να εφαρμόζεται όποτε αυτό είναι εφικτό. Επειδή η τύχη υπαγορεύει ποιος θα είναι μέλος σε κάθε ομάδα, η τυχαία ανάθεση (εκχώρηση) όχι απλώς ελαχιστοποιεί τη μεροληψία που μπορεί να ευνοήσει τη μία ομάδα, αλλά επίσης αποτελεί μια βάση για την αξιολόγηση του ρόλου της μεταβλητότητας σε οποιοδήποτε παρατηρούμενο αποτέλεσμα. Η τυχαία ανάθεση (εκχώρηση) μας επιτρέπει να αξιολογούμε κάθε εύρημα, όπως την πραγματική μέση διαφορά μεταξύ δύο ομάδων, να προσδιορίζουμε αν αυτή η διαφορά είναι μεγαλύτερη από την αναμενόμενη μόνο λόγω τύχης, όταν λαμβάνεται υπόψη η μεταβλητότητα. Με άλλα λόγια, μας επιτρέπει να γενικεύουμε πέρα από απλές εμφανίσεις και να διαπιστώνουμε αν η μέση διαφορά αξίζει περαιτέρω προσοχής επειδή *πιθανώς είναι πραγματική* ή αν πρέπει να αγνοηθεί επειδή *μπορεί να αποδοθεί στη μεταβλητότητα ή στην τύχη*.

Τυχαία ανάθεση (εκχώρηση)
Μια διαδικασία που σχεδιάζεται έτσι ώστε να διασφαλίζει ότι κάθε άτομο έχει ίση πιθανότητα να τοποθετηθεί σε οποιαδήποτε ομάδα σε ένα πείραμα.

Επισκόπηση: Έρευνες και πειράματα

Το Σχήμα 1.1 συγκρίνει έρευνες και πειράματα. Με βάση τυχαία δείγματα από πληθυσμούς, οι έρευνες επιτρέπουν τις γενικεύσεις από δείγματα σε πληθυσμούς. Με βάση την τυχαία ανάθεση εθελοντών σε ομάδες, τα πειράματα επιτρέπουν τη λήψη αποφάσεων σχετικά με το αν οι διαφορές μεταξύ ομάδων είναι πραγματικές ή πρόσκαιρες.



ΣΧΗΜΑ 1.1

Επισκόπηση: Έρευνες και πειράματα.

Έλεγχος προόδου *1.2 Με τι συσχετίζονται κυρίως οι παρακάτω όροι, με μια έρευνα (Ε) ή με ένα πείραμα (Π);

(α) τυχαία ανάθεση (εκχώρηση)

(β) αντιπροσωπευτικός

(γ) γενίκευση στον πληθυσμό

(δ) ομάδα ελέγχου

Απαντήσεις στη σελίδα 529.

(ε) πραγματική διαφορά

(στ) τυχαία επιλογή

(ζ) πρακτικό δείγμα

(η) εθελοντές

Απαντήσεις στη σελίδα 529.

1.4 Τρεις τύποι δεδομένων

Δεδομένα

Μια συλλογή πραγματικών παρατηρήσεων ή αποτελεσμάτων σε μια έρευνα ή ένα πείραμα.

Ποιοτικά δεδομένα

Ένα σύνολο παρατηρήσεων όπου οποιαδήποτε μεμονωμένη παρατήρηση είναι μια λέξη, ένα γράμμα ή ένας αριθμητικός κωδικός που αναπαριστά μια κλάση ή κατηγορία.

Διατεταγμένα δεδομένα

Ένα σύνολο παρατηρήσεων όπου οποιαδήποτε μεμονωμένη παρατήρηση είναι ένας αριθμός που δείχνει τη σχετική θέση σε ένα σύνολο.

Ποσοτικά δεδομένα

Ένα σύνολο παρατηρήσεων όπου οποιαδήποτε μεμονωμένη παρατήρηση είναι ένας αριθμός που αναπαριστά μια ποσότητα ή μια μέτρηση.

Οποιαδήποτε στατιστική ανάλυση πραγματοποιείται με τη χρήση **δεδομένων**, δηλαδή με μια συλλογή πραγματικών παρατηρήσεων ή αποτελεσμάτων σε μια έρευνα ή ένα πείραμα.

Η ακριβής μορφή μιας στατιστικής ανάλυσης εξαρτάται συχνά από το αν τα δεδομένα είναι ποιοτικά, διατεταγμένα ή ποσοτικά.

Γενικά, τα **ποιοτικά δεδομένα** αποτελούνται από εκφράσεις (Ναι ή Όχι), γράμματα (N ή O) ή αριθμητικούς κωδικούς (0 ή 1) που αναπαριστούν μια κλάση ή κατηγορία. Τα **διατεταγμένα δεδομένα** αποτελούνται από αριθμούς (1η, 2η, ..., 40ή θέση) που αντιπροσωπεύουν τη σχετική κατάταξη μέσα σε μια ομάδα. Τα **ποσοτικά δεδομένα** αποτελούνται από αριθμούς (βάρη 238, 170, ..., 185 λιβρών) που αναπαριστούν μια ποσότητα ή μια μέτρηση. Για να διαπιστώσετε τον τύπο των δεδομένων, εστιάζετε σε μία μεμονωμένη παρατήρηση που ανήκει σε οποιαδήποτε συλλογή παρατηρήσεων. Για παράδειγμα, τα βάρη που αναφέρονται από 53 άρρενες φοιτητές στον Πίνακα 1.1 είναι ποσοτικά δεδομένα, επειδή οποιαδήποτε μεμονωμένη παρατήρηση, όπως το 160 λίβρες, αναπαριστά ένα βάρος. Αν τα βάρη στον Πίνακα 1.1 είχαν αντικατασταθεί από βαθμούς, ξεκινώντας από τον βαθμό 1 για το μικρότερο βάρος των 133 λιβρών έως τον βαθμό 53 για το μεγαλύτερο βάρος των 245 λιβρών, αυτοί οι αριθμοί θα ήταν διατεταγμένα δεδομένα, επειδή οποιαδήποτε μεμονωμένη παρατήρηση αντιπροσωπεύει όχι μια ποσότητα αλλά μόνο τη σχετική θέση μέσα στην ομάδα των 53 φοιτητών. Τέλος, οι απαντήσεις N (για το ναι) και O (για το όχι) των φοιτητών στον Πίνακα 1.2 είναι ποιοτικά δεδομένα, επειδή οποιαδήποτε μεμονωμένη παρατήρηση είναι ένα γράμμα που αναπαριστά μια κατηγορία απαντήσεων.

Έλεγχος προόδου *1.3 Απαντήστε αν ο καθένας από τους παρακάτω όρους είναι *ποιοτικός* (επειδή είναι μια λέξη, ένα γράμμα ή ένας αριθμητικός κωδικός που αναπαριστά μια κλάση ή κατηγορία), *διατεταγμένος* (επειδή είναι ένας αριθμός που δείχνει τη σχετική θέση) ή *ποσοτικός* (επειδή είναι ένας αριθμός που αναπαριστά μια ποσότητα ή μια μέτρηση).

(α) εθνικότητα

(β) ηλικία

(γ) μέγεθος οικογένειας

(δ) ακαδημαϊκός τίτλος σπουδών

(ε) σεξουαλική προτίμηση

(στ) δείκτης νοημοσύνης (IQ)

(ζ) καθαρή αξία (σε δολάρια)

(η) τρίτη θέση σε αγώνα

(θ) φύλο

(ι) θερμοκρασία

Απαντήσεις στη σελίδα 529.

Πίνακας 1.2
ΠΟΙΟΤΙΚΑ ΔΕΔΟΜΕΝΑ: «ΔΙΑΤΗΡΕΙΤΕ ΠΡΟΦΙΛ ΣΤΟ FACEBOOK;»
ΑΠΑΝΤΗΣΕΙΣ ΝΑΙ (Ν) Ή ΟΧΙ (Ο) ΦΟΙΤΗΤΩΝ ΣΤΑΤΙΣΤΙΚΗΣ

N	N	N	O	O	N	N	N
N	N	N	O	O	N	N	N
O	N	O	N	N	N	N	N
N	N	O	N	O	N	O	N
N	O	N	O	O	N	N	N
N	N	O	N	N	N	N	N
O	O	O	O	N	O	O	N
N	N	N	N	N	O	N	O
N	N	N	N	O	O	N	N
O	N	O	O	N	N	N	N
	N	N	O				

1.5 Επίπεδα μέτρησης

Όπως μάθατε στο δημοτικό σχολείο, η ισότητα ότι $2 + 2 = 4$ θεωρείται μία από τις σταθερές της ζωής, όπως είναι οι φόροι και ο θάνατος. Δεν επιδέχονται όμως όλοι οι αριθμοί την ίδια ερμηνεία. Για παράδειγμα, δεν θα είχε νόημα να βρούμε το άθροισμα δύο αριθμών μητρώου κοινωνικής ασφάλισης ή να ισχυριστούμε ότι, όταν θεωρούνται δείκτες ακαδημαϊκής επιτυχίας, δύο ξεχωριστοί γενικοί βαθμοί 5 ισούνται με έναν γενικό βαθμό 10. Για να αποσαφηνίσουμε καλύτερα τις διαφορές μεταξύ των τριών τύπων δεδομένων, θα περιγράψουμε την έννοια του επιπέδου μέτρησης. Προβάλλοντας πίσω από όλα τα δεδομένα, το **επίπεδο μέτρησης** καθορίζει τον βαθμό (έκταση) στον οποίο ένας αριθμός (ή λέξη ή γράμμα) πράγματι αναπαριστά κάποια ιδιότητα και, επομένως, επηρεάζει την καταλληλότητα διάφορων αριθμητικών πράξεων και στατιστικών διαδικασιών.

Για δικούς μας σκοπούς, υπάρχουν τρία επίπεδα μέτρησης –ονομαστικό, διατάξιμο και διαστήματος/λόγου–, και αυτά τα επίπεδα αντιστοιχούν στα ποιοτικά, στα διατεταγμένα και στα ποσοτικά δεδομένα. Οι ιδιότητες αυτών των επιπέδων –και η χρησιμότητα των σχετικών αριθμών τους– διαφέρουν από το ονομαστικό, το απλούστερο επίπεδο με μία μόνο ιδιότητα, στο διαστήματος/λόγου, το πιο περίπλοκο επίπεδο με τέσσερις ιδιότητες. Όλο και πιο περίπλοκα επίπεδα περιέχουν όλες τις ιδιότητες των απλούστερων επιπέδων και, επιπλέον, μία ή δύο νέες ιδιότητες.

Πιο περίπλοκα επίπεδα μέτρησης συσχετίζονται με αριθμούς οι οποίοι, επειδή αντιπροσωπεύουν καλύτερα τις ιδιότητες, επιτρέπουν μεγαλύτερη ποικιλία αριθμητικών πράξεων και στατιστικών διαδικασιών.

Ποιοτικά δεδομένα και ονομαστική μέτρηση

Αν οι άνθρωποι κατατάσσονται σε άντρες ή γυναίκες (ή, κωδικοποιημένα, ως 1 ή 2), τα δεδομένα θεωρούνται ποιοτικά και η μέτρηση είναι ονομαστική. Η **μοναδική ιδιότητα της ονομαστικής μέτρησης είναι η ταξινόμηση** – δηλαδή η κατάταξη των παρατηρήσεων σε διαφορετικές κλάσεις ή κατηγορίες. Οι λέξεις, τα γράμματα ή οι αριθμητικοί κωδικοί εκφράζουν μόνο διαφορές σε είδος και όχι σε ποσότητα. Παραδείγματα ονομαστικής μέτρησης είναι η ταξινόμηση των διαταραχών διάθεσης, όπως μανιακές, διπολικές ή καταθλιπτικές, των σεξουαλικών προτιμήσεων ως ετεροφυλόφιλων, ομοφυλόφιλων, αμφιφυλόφιλων ή ασεξουαλικών και των απόψεων απέναντι σε αυστηρότερους ελέγχους της ρύπανσης ως υπέρ, κατά ή αναποφάσιτων.

Ένα διακριτό γνώρισμα της ονομαστικής μέτρησης είναι η θεμελιώδης αναπαράσταση οποιουδήποτε χαρακτηριστικού. Για παράδειγμα, ένας φοιτητής είναι άντρας ή γυναίκα. Ακόμα και μετά την εισαγωγή αυθαίρετων αριθμητικών κωδικών, όπως 1 για τους άντρες και 2 για τις γυναίκες, δεν θα ήταν σωστό να ισχυριστούμε ότι,

Επίπεδο μέτρησης

Καθορίζει τον βαθμό (έκταση) στον οποίο ένας αριθμός (ή λέξη ή γράμμα) πράγματι αναπαριστά κάποια ιδιότητα και, επομένως, επηρεάζει την καταλληλότητα διάφορων αριθμητικών πράξεων και στατιστικών διαδικασιών.

Ονομαστική μέτρηση

Λέξεις, γράμματα ή αριθμητικοί κωδικοί ποιοτικών δεδομένων που εκφράζουν διαφορές στο είδος με βάση την ταξινόμηση.

επειδή οι γυναίκες είναι το 2 και οι άντρες το 1, τότε οι γυναίκες έχουν διπλάσιο φύλο από τους άντρες. Ομοίως, ο υπολογισμός ενός μέσου όρου με τέτοιους αριθμούς δεν θα είχε κανένα νόημα. Εξαιτίας αυτών των περιορισμών, λίγες μόνο ενότητες αυτού του βιβλίου και το Κεφάλαιο 19 αφιερώνονται αποκλειστικά σε μια ανάλυση των ποιοτικών δεδομένων με ονομαστική μέτρηση.

Διατάξιμη μέτρηση

Σχετική θέση διατεταγμένων δεδομένων που εκφράζουν διαφορές στον βαθμό με βάση την τάξη.

Διατεταγμένα δεδομένα και διατάξιμη μέτρηση

Όταν οποιοσδήποτε ξεχωριστός αριθμός υποδεικνύει μόνο *σχετική θέση*, όπως η πρώτη, η δεύτερη ή η δέκατη θέση σε μια ιπποδρομία ή σε μια τάξη τελειοφοίτων, τα δεδομένα θεωρούνται *διατεταγμένα* και το επίπεδο μέτρησης *διατάξιμο*. Η *διακριτή ιδιότητα της διατάξιμης μέτρησης είναι η τάξη ή σειρά*. Συγκριτικά, μια πρώτη θέση σημαίνει τον πιο γρήγορο τερματισμό σε μια ιπποδρομία ή τον υψηλότερο γενικό βαθμό μεταξύ των τελειοφοίτων. Αν και η πρώτη θέση σε μια ιπποδρομία δείχνει έναν *ταχύτερο* τερματισμό από τη δεύτερη θέση, δεν γνωρίζουμε το *πόσο* ταχύτερη είναι αυτή.

Επειδή η διατάξιμη μέτρηση δεν καταφέρνει να αποτυπώσει την πραγματική απόσταση μεταξύ γειτονικών βαθμών, δεν προσφέρονται οι απλές αριθμητικές πράξεις με τους βαθμούς. Για παράδειγμα, δεν είναι σωστό να συμπεράνουμε ότι ο αριθμητικός μέσος των βαθμών 1 και 3 ισούται με τον βαθμό 2, επειδή αυτό προϋποθέτει ότι η πραγματική απόσταση μεταξύ των βαθμών 1 και 2 ισούται με την απόσταση μεταξύ των βαθμών 2 και 3. Αντίθετα, αυτές οι αποστάσεις μπορεί να είναι πολύ διαφορετικές. Για παράδειγμα, ο βαθμός 2 μπορεί να συνδέεται είτε με τον βαθμό 1 είτε με τον βαθμό 3. Λίγες μόνο ενότητες αυτού του βιβλίου και το Κεφάλαιο 20 ασχολούνται αποκλειστικά με μια ανάλυση διατεταγμένων δεδομένων με διατάξιμη μέτρηση.²

Ποσοτικά δεδομένα και μέτρηση διαστήματος/λόγου

Μέτρηση διαστήματος/λόγου

Οι ποσότητες ή οι μετρήσεις ποσοτικών δεδομένων αναπαριστούν διαφορές στον βαθμό με βάση τα ίσα διαστήματα και ένα πραγματικό μηδέν.

Συχνά σε προϊόντα οικείων «συσκευών» μέτρησης, όπως χάρακες, ρολόγια ή μέτρα, οι *χαρακτηριστικές ιδιότητες της μέτρησης διαστήματος/λόγου είναι ίσα διαστήματα και ένα πραγματικό μηδέν*.³ Όταν ζυγίζετε σε μια οικιακή ζυγαριά, πραγματοποιείτε ουσιαστικά μια μέτρηση διαστήματος/λόγου. Η φράση *ίσα διαστήματα* σημαίνει ότι, όταν σηκώνετε ένα βάρος 5 κιλών εφόσον είστε πάνω στη ζυγαριά, το βάρος σας θα δείχνει 5 κιλά επιπλέον. Τα ίσα διαστήματα σημαίνουν ότι η διαφορά μεταξύ 60 και 65 κιλών αναπαριστά μια

ποσότητα βάρους ίση με τη διαφορά μεταξύ των 60 και των 65 και αρμόζει για την περιγραφή του βάρους ενός ατόμου ως μια συγκεκριμένη ποσότητα μεγαλύτερη από το βάρος κάποιου άλλου.

Το *πραγματικό μηδέν* πιστοποιεί το γεγονός ότι η ζυγαριά αναγράφει 0 όταν δεν χρησιμοποιείται, ότι δηλαδή το βάρος είναι εντελώς απόν. Επειδή η ζυγαριά έχει ένα πραγματικό μηδέν, οι αριθμητικές ενδείξεις αναπαριστούν τη *συνολική ποσότητα* του βάρους ενός ατόμου και είναι κατάλληλη για την περιγραφή του βάρους ενός ατόμου ως συγκεκριμένο ποσοστό του βάρους ενός άλλου. Μπορούμε να πούμε ότι το βάρος ενός ατόμου 70 κιλών είναι διπλάσιο από το βάρος ενός ατόμου που ζυγίζει 35 κιλά.

Όταν δεν υπάρχει «πραγματικό μηδέν», οι αριθμοί –περίπου όπως οι φανερός κορυφές των παγόβουνων– δεν καταφέρνουν να αναπαραστήσουν τη συνολική υπό μέτρηση ποσότητα. Για παράδειγμα, μια ένδειξη 0 στην κλίμακα θερμοκρασίας με βαθμούς Fahrenheit δεν αποτυπώνει την απόλυτη απουσία θερμότητας, δηλαδή την απουσία οποιασδήποτε μοριακής κίνησης. Για την ακρίβεια, σ' αυτήν την κλίμακα το πραγματικό μηδέν ισούται με $-459,4^{\circ}\text{F}$. Δεν θα ήταν, επομένως, ορθό να ισχυριστούμε ότι οι 80°F σημαίνουν δύο φορές περισσότερη ζέση από τους 40°F . Ένας ορθός ισχυρισμός θα μπορούσε να γίνει αν προσθέταμε το $459,4^{\circ}\text{F}$ σ' αυτούς τους αριθμούς: το 80° γίνεται $539,4^{\circ}$ και το 40° γίνεται $499,4^{\circ}$. Είναι σαφές ότι οι $539,4^{\circ}\text{F}$ σημαίνουν δύο φορές περισσότερη ζέση από τους $499,4^{\circ}\text{F}$.

2. Με αυστηρά τεχνικούς όρους, η διατάξιμη μέτρηση μπορεί επίσης να συσχετιστεί με ποιοτικά δεδομένα των οποίων οι κλάσεις είναι ταξινομημένες. Παραδείγματα *ταξινομημένων* ποιοτικών δεδομένων είναι η ταξινόμηση εξειδικευμένων εργαζομένων, όπως αρχιμαστόρων, εμπειρών τεχνιτών ή μαθητευόμενων, κοινωνικοοικονομικής κατάστασης, όπως χαμηλή, μεσαία ή υψηλή, και βαθμών που παίρνετε στα μαθήματά σας. Αξίζει να διακρίνουμε τα ποιοτικά δεδομένα με ονομαστική και διατάξιμη μέτρηση επειδή, όπως περιγράφεται στα Κεφάλαια 3 και 4, υπάρχουν κάποιες επιπλέον στατιστικές διαδικασίες για τα ταξινομημένα ποιοτικά δεδομένα.

3. Η ύπαρξη του μηδενός ως υφισταμένου. (Σ.τ.Ε.)

Η μέτρηση διαστήματος/λόγου εμφανίζεται σε συμπεριφορικές και κοινωνικές επιστήμες, όπως, για παράδειγμα, ο ρυθμός που πατάνε τον μοχλό τα ποντίκια σε κλουβιά Skinner, τα λεπτά του φιλικού προς τα όνειρα ύπνου του σταδίου REM σε ένα πείραμα στέρησης ύπνου και ο συνολικός αριθμός των οπτικών επαφών κατά τη διάρκεια καβγάδων ερωτικών συντρόφων. Χάρη στη σημαντική ποσότητα πληροφοριών που παράγεται από κάθε παρατήρηση, η μέτρηση διαστήματος/λόγου επιτρέπει την εκτέλεση ουσιαστικών αριθμητικών πράξεων, όπως είναι ο υπολογισμός του αριθμητικού μέσου, αλλά και των πολλών στατιστικών διαδικασιών για τα ποσοτικά δεδομένα που περιγράφονται σ' αυτό το βιβλίο.

Μέτρηση μη φυσικών χαρακτηριστικών

Όταν οι αριθμοί αναπαριστούν μη φυσικά χαρακτηριστικά, όπως είναι η πνευματική ικανότητα, η ψυχοπαθητική τάση ή η συναισθηματική ωριμότητα, η επίτευξη μέτρησης διαστήματος/λόγου πολλές φορές αμφισβητείται. Για παράδειγμα, δεν υπάρχει εξωτερικό πρότυπο (όπως το βάρος των 5 κιλών) που μπορεί να δείξει ότι η πρόσθεση μιας σταθερής ποσότητας πνευματικής ικανότητας παράγει πάντα ίση αύξηση στα τεστ IQ (ίσα διαστήματα). Δεν υπάρχει επίσης κανένα όργανο (όπως η ζυγαριά όταν δεν είναι κανείς πάνω της) που καταγράφει αποτέλεσμα IQ 0 όταν η πνευματική ικανότητα είναι εντελώς απύουσα (πραγματικό μηδέν).

Όταν δεν υπάρχουν ίσα διαστήματα, δεν θα ήταν ορθό να ισχυριστούμε ότι η διαφορά μεταξύ αποτελεσμάτων σε τεστ IQ 120 και 130 αναπαριστά ίση ποσότητα πνευματικής ικανότητας με τη διαφορά μεταξύ αποτελεσμάτων 130 και 140. Ομοίως, όταν δεν υπάρχει πραγματικό μηδέν, δεν θα ήταν ορθό να ισχυριστούμε ότι ένα αποτέλεσμα 140 σε τεστ IQ συνεπάγεται διπλάσια πνευματική ικανότητα από ένα αποτέλεσμα 70.

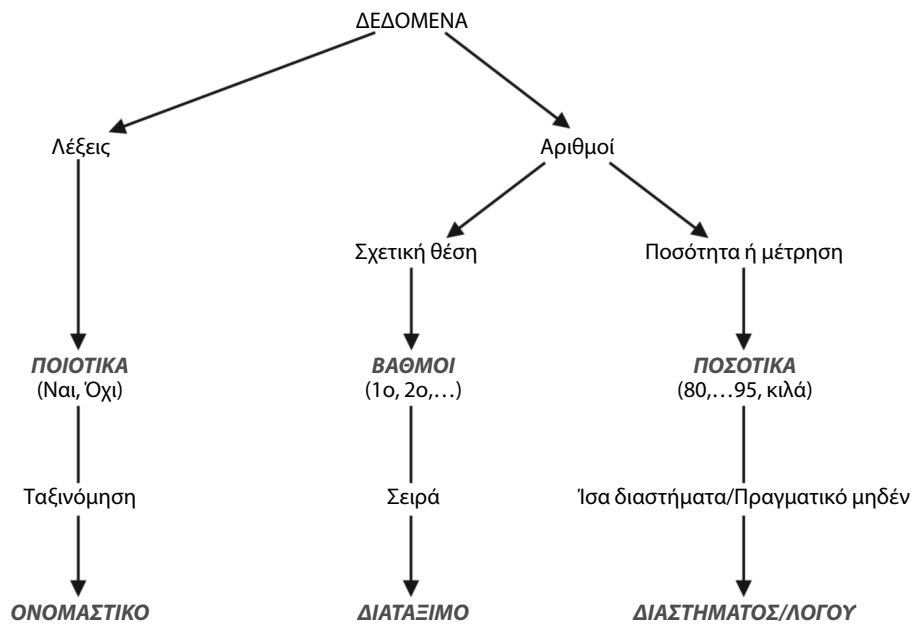
Υπάρχουν επίσης κι άλλες ερμηνείες. Μια πιθανότητα είναι να θεωρούμε ότι τα αποτελέσματα σε τεστ IQ επιτυγχάνουν μόνο διατάξιμη μέτρηση –δηλαδή, για παράδειγμα, το αποτέλεσμα 140 αντιπροσωπεύει μεγαλύτερη πνευματική ικανότητα από το 130–, χωρίς όμως να προσδιορίζεται το πραγματικό μέγεθος αυτής της διαφοράς. Αυτή η αυστηρή ερμηνεία θα περιόριζε σημαντικά τον αριθμό των στατιστικών διαδικασιών που θα μπορούσαν να εφαρμοστούν σε συμπεριφορικά και κοινωνικά δεδομένα. Μια χαλαρή (και πολύ πιο συνηθισμένη) ερμηνεία, την οποία υιοθετούμε σ' αυτό το βιβλίο, θεωρεί ότι, ακόμα κι αν στερούνται πραγματικού μηδενός, τα αποτελέσματα σε τεστ IQ παρέχουν ένα ακατέργαστο μέτρο αντιστοίχισης διαφορών στην πνευματική ικανότητα (ίσα διαστήματα). Ως εκ τούτου, η διαφορά μεταξύ αποτελεσμάτων 120 και 130 σε τεστ IQ αναπαριστά μια σχεδόν *παρόμοια* ποσότητα πνευματικής ικανότητας όση είναι η διαφορά μεταξύ των αποτελεσμάτων 130 και 140.

Στον βαθμό που τα αριθμητικά μέτρα μη φυσικών χαρακτηριστικών προσεγγίζουν τη μέτρηση διαστήματος, αντιμετωπίζονται στατιστικά με τον ίδιο τρόπο με τα αριθμητικά μέτρα φυσικών χαρακτηριστικών. Με άλλα λόγια, αυτά τα μέτρα υποστηρίζουν τις αριθμητικές πράξεις και τα στατιστικά εργαλεία που αρμόζουν στα ποσοτικά δεδομένα.

Σ' αυτό το σημείο μπορεί να ευχόσασταν να ήταν δυνατή η προσθήκη σε ένα άτομο 10 βαθμών πνευματικής ικανότητας (ή ψυχοπαθητικής τάσης ή συναισθηματικής ωριμότητας) ως ένα πρώτο βήμα προς μια κλίμακα δείκτη νοημοσύνης με ίσα διαστήματα και πραγματικό μηδέν. Η έλλειψη όμως αυτής της εναλλακτικής λύσης σας αναγκάζει να εκπαιδευτείτε κατάλληλα ώστε να εξετάζετε τους αριθμούς ως προϊόντα μέτρησης και να χρησιμοποιείτε τους αριθμητικούς ισχυρισμούς σας αναλόγως – ιδιαίτερα όταν τα αριθμητικά δεδομένα φαίνεται ότι απλώς καταφέρνουν να προσεγγίσουν τη μέτρηση διαστήματος.

Επισκόπηση: Τύποι δεδομένων και επίπεδα μέτρησης

Ενώ διαβάζετε αυτήν την παράγραφο, παρατηρήστε το Σχήμα 1.2. Δεδομένου κάποιου συνόλου παρατηρήσεων, αποφασίστε αν οποιαδήποτε μεμονωμένη παρατήρηση μπορεί να θεωρηθεί λέξη ή αριθμός. Αν είναι λέξη (ή γράμμα ή αριθμητικός κωδικός), τα δεδομένα είναι ποιοτικά και το επίπεδο μέτρησης ονομαστικό. Οι αριθμητικές πράξεις δεν έχουν νόημα και οι στατιστικές διαδικασίες είναι περιορισμένες. Από την άλλη πλευρά, αν η παρατήρηση είναι ένας αριθμός, τα δεδομένα είναι διατεταγμένα ή ποσοτικά, ανάλογα με το αν οι αριθμοί αναπαριστούν μόνο σχετική θέση ή μια ποσότητα/μέτρηση. Αν τα δεδομένα είναι διατεταγμένα, το επίπεδο μέτρησης είναι διατάξιμο και, όπως συμβαίνει επίσης με τα ποιοτικά δεδομένα, οι αριθμητικές πράξεις και οι στατιστικές διαδικασίες είναι περιορισμένες. Αν τα δεδομένα είναι ποσοτικά, το επίπεδο μέτρησης είναι διαστήματος/λόγου –ή κατά προσέγγιση διαστήματος όταν οι αριθμοί αναπαριστούν μη φυσικά χαρακτηριστικά– και υπάρχει διαθέσιμη μια ολόκληρη σειρά από αριθμητικές πράξεις και στατιστικές διαδικασίες.



ΣΧΗΜΑ 1.2

Επισκόπηση: Τύποι δεδομένων και επίπεδα μέτρησης.

Έλεγχος προόδου *1.4 Ποιο είναι το επίπεδο μέτρησης –ονομαστικό, διατάξιμο ή διαστήματος/λόγου– που επιτυγχάνεται από τα παρακάτω σύνολα παρατηρήσεων ή δεδομένων; Όπου αρμόζει, δηλώστε ότι η μέτρηση είναι μόνο κατά προσέγγιση διαστήματος.

ΣΗΜΕΙΩΣΗ: Θα πρέπει να ορίζετε το υψηλότερο επιτρεπόμενο επίπεδο μέτρησης σε ένα συγκεκριμένο σύνολο παρατηρήσεων. Για παράδειγμα, μια λίστα με ετήσια εισοδήματα θα πρέπει να οριστεί ως διαστήματος/λόγου επειδή μια διαφορά 1.000€ σηματοδοτεί πάντα την ίδια ποσότητα εισοδήματος (ίσα διαστήματα) και επειδή η ποσότητα 0€ σημαίνει παντελή απουσία εισοδήματος. Θα ήταν λάθος να περιγράψουμε το ετήσιο εισόδημα ως διατεταγμένο δεδομένο, ακόμα κι αν διαφορετικά εισοδήματα μπορούν πάντα να κατατάσσονται ως περισσότερα ή λιγότερα (σειρά), ή ως ονομαστικό δεδομένο, ακόμα κι αν διαφορετικά εισοδήματα μπορούν πάντοτε να αντιπροσωπεύουν διαφορετικές κοινωνικές τάξεις (ταξινόμηση).

- (α) ύψος
- (β) σχέση με τη θρησκεία
- (γ) αποτέλεσμα εξέτασης ψυχοπαθητικής τάσης
- (δ) έτη εκπαίδευσης
- (ε) στρατιωτικός βαθμός
- (στ) επαγγελματικός στόχος
- (ζ) μέσος όρος βαθμολογίας
- (η) οικογενειακή κατάσταση

Απαντήσεις στη σελίδα 529.

Μεταβλητή

Ένα χαρακτηριστικό ή μια ιδιότητα που μπορεί να πάρει διαφορετικές τιμές.

1.6 Τύποι μεταβλητών

Γενικός ορισμός

Ακόμα μια χρήσιμη διάκριση βασίζεται σε διαφορετικούς τύπους μεταβλητών. **Μεταβλητή** είναι ένα χαρακτηριστικό ή μια ιδιότητα που μπορεί να πάρει διαφορετικές τιμές. Αναλό-

γως, οι τιμές των βαρών στον Πίνακα 1.1 μπορούν να περιγραφούν όχι μόνο ως ποσοτικά δεδομένα αλλά επίσης ως παρατηρήσεις για μια ποσοτική μεταβλητή, επειδή τα διαφορετικά βάρη παίρνουν διαφορετικές αριθμητικές τιμές. Με το ίδιο σκεπτικό, οι απαντήσεις στον Πίνακα 1.2 μπορούν να περιγραφούν ως παρατηρήσεις για μια ποιοτική μεταβλητή, επειδή οι απαντήσεις στην ερώτηση σχετικά με το προφίλ στο Facebook παίρνουν διαφο-

ρετικές τιμές, Ναι ή Όχι. Υπό αυτήν την έννοια, οποιαδήποτε *μεμονωμένη παρατήρηση* στον Πίνακα 1.1 ή στον Πίνακα 1.2 μπορεί να περιγραφεί ως μια **σταθερά**, επειδή *παίρνει μόνο μία τιμή*.

Σταθερά

Ένα χαρακτηριστικό ή μια ιδιότητα που μπορεί να πάρει μόνο μία τιμή.

Διακριτές και συνεχείς μεταβλητές

Οι ποσοτικές μεταβλητές μπορούν να διακριθούν περαιτέρω ανάλογα με το αν είναι διακριτές ή συνεχείς. Μια **διακριτή μεταβλητή** αποτελείται από μεμονωμένους αριθμούς που χωρίζονται με κενά. Παραδείγματα αποτελούν οι περισσότερες αριθμήσιμες μεταβλητές, όπως ο αριθμός των παιδιών μιας οικογένειας (1, 2, 3 κ.λπ., αλλά ποτέ το 1½, ακόμα κι αν μπορεί να σκέφτεστε κάπως έτσι για έναν αδελφό ή αδελφή), ο αριθμός των ξένων χωρών που έχετε επισκεφτεί και το τρέχον μέγεθος του πληθυσμού των ΗΠΑ. Μια **συνεχής μεταβλητή** αποτελείται από αριθμούς των οποίων οι τιμές, τουλάχιστον θεωρητικά, δεν έχουν περιορισμούς. Παραδείγματα είναι διάφορες ποσότητες, όπως τα βάρη αρρένων φοιτητών στατιστικής, διάρκειες, όπως οι χρόνοι αντίδρασης παιδιών δημοτικών σχολείων σε έναν συναγερμό για πυρκαγιά, και οι βαθμολογίες τυποποιημένων εξετάσεων, όπως η εξέταση σχολικών δεξιοτήτων SAT.

Διακριτή μεταβλητή

Μια μεταβλητή που αποτελείται από μεμονωμένους αριθμούς που χωρίζονται με κενά.

Συνεχής μεταβλητή

Μια μεταβλητή που αποτελείται από αριθμούς των οποίων οι τιμές, τουλάχιστον θεωρητικά, δεν έχουν περιορισμούς.

Προσεγγιστικοί αριθμοί

Στη θεωρία, οι τιμές για συνεχείς μεταβλητές μπορούν να συνεχίζονται επ' αόριστον. Το βάρος κάποιου, σε κιλά, μπορεί να είναι 79,02352 και αυτός ο αριθμός να συνεχίζεται στο άπειρο! Πρακτικά ζητήματα επιβάλλουν τη στρογγυλοποίηση των τιμών των συνεχών μεταβλητών. Όποτε οι τιμές στρογγυλοποιούνται, όπως συμβαίνει πάντα με πραγματικές τιμές για συνεχείς μεταβλητές, οι αριθμοί που προκύπτουν είναι **προσεγγιστικοί** και ποτέ δεν είναι ακριβείς. Για παράδειγμα, τα βάρη των αρρένων φοιτητών στατιστικής στον Πίνακα 1.1 (σε λίβρες) είναι προσεγγιστικοί, καθώς έχουν στρογγυλοποιηθεί στον πλησιέστερο ακέραιο αριθμό. Ένας φοιτητής του οποίου το βάρος εμφανίζεται ως 150 λίβρες μπορεί στην πραγματικότητα να ζυγίζει μεταξύ 149,5 και 150,5 λίβρες. Τυπικά, οποιαδήποτε τιμή για μια συνεχή μεταβλητή, όπως οι 150 λίβρες, πρέπει να προσδιορίζεται με ένα διάστημα τιμών από 149,5 ως 150,5 αντί με μία μόνο τιμή. Όπως θα δούμε, αυτή η ιδιότητα των συνεχών μεταβλητών έχει αρκετές επιπτώσεις, όπως την επιλογή διαγραμμάτων στο Κεφάλαιο 2 και τους τύπους ουσιαστικών ερωτήσεων για κανονικές κατανομές στο Κεφάλαιο 5.

Προσεγγιστικοί αριθμοί

Αριθμοί που στρογγυλοποιούνται, όπως συμβαίνει πάντοτε με τιμές για συνεχείς μεταβλητές.

Εξαιτίας των διαδικασιών στρογγυλοποίησης, μεταξύ των τιμών συνεχών μεταβλητών εμφανίζονται κενά. Για παράδειγμα, επειδή τα βάρη στρογγυλοποιούνται στον εγγύτερο ακέραιο αριθμό, κανένας άντρας φοιτητής στατιστικής στον Πίνακα 1.1 δεν εμφανίζεται με βάρος μεταξύ 150 και 151 λιβρών. Αυτά τα κενά είναι πιο προφανή από ό,τι πραγματικά και τίθενται υπεράνω σε μια συνεχή μεταβλητή λόγω της ανάγκης μας να χειριζόμαστε πεπερασμένους (και, επομένως, προσεγγιστικούς) αριθμούς.

Έλεγχος προόδου *1.5 Οι παρακάτω ποσοτικές παρατηρήσεις είναι διακριτές ή συνεχείς;

- (α) άμμος για γάτες
 - (β) χρόνος μαγειρέματος ζυμαρικών
 - (γ) παραβιάσεις αναστολής από καταδικασμένους εγκληματίες
 - (δ) IQ
 - (ε) ηλικία
 - (στ) πληθυσμός της πόλης σας
 - (ζ) ταχύτητα αεροσκάφους
- Απαντήσεις στη σελίδα 529.

Ανεξάρτητες και εξαρτημένες μεταβλητές

Αντίθετα από τις απλές έρευνες που παρήγαγαν τα δεδομένα των Πινάκων 1.1 και 1.2, οι περισσότερες μελέτες εγείρουν ερωτήματα σχετικά με την παρουσία ή απουσία μιας σχέσης μεταξύ δύο (ή περισσότερων) μεταβλητών. Για παράδειγμα, ένας ψυχολόγος, στο πλαίσιο μιας έρευνας, ερευνά εάν τα ζευγάρια που υποβάλλονται σε

Πείραμα

Μια μελέτη στην οποία ο ερευνητής αποφασίζει ποιος θα υποβάλλεται σε ειδικό πειραματικό χειρισμό.

ειδική εκπαίδευση στην «ενεργητική ακρόαση» παρουσιάζουν λιγότερα προβλήματα επικοινωνίας από τα ζευγάρια που δεν έχουν τέτοια εκπαίδευση. Για να μελετήσει αυτό το ζήτημα, ο ψυχολόγος μπορεί να εκθέσει ζευγάρια σε δύο διαφορετικές συνθήκες, εντάσσοντάς τους τυχαία είτε σε μια ομάδα πειραματικού χειρισμού που υποβάλλεται σε ειδική εκπαίδευση στην ενεργητική ακρόαση είτε σε μια ομάδα ελέγχου που δεν λαμβάνει ειδική εκπαίδευση. Τέτοιες μελέτες αναφέρονται ως πειράματα. Ένα **πείραμα** είναι μια μελέτη στην οποία ο ερευνητής αποφασίζει ποιος θα υποβάλλεται σε ειδικό πειραματικό χειρισμό. Όταν σχεδιάζονται καλά, τα πειράματα δίνουν τα πιο ενημερωτικά και σαφή συμπεράσματα για σχέσεις αιτίου-αιτιατού.

Ανεξάρτητη μεταβλητή

Το αντικείμενο που χειρίζεται ένας ερευνητής σε ένα πείραμα.

Ανεξάρτητη μεταβλητή

Επειδή η εκπαίδευση θεωρείται ότι επηρεάζει την επικοινωνία, είναι μια ανεξάρτητη μεταβλητή. Σε ένα πείραμα, μια **ανεξάρτητη μεταβλητή** είναι το αντικείμενο που χειρίζεται ένας ερευνητής.

Η αμερόληπτη δημιουργία διακριτών ομάδων, οι οποίες διαφέρουν μόνο ως προς την ανεξάρτητη μεταβλητή, έχει μια εξαιρετικά επιθυμητή επίπτωση. Αφού συλλεχθούν τα δεδομένα, οποιαδήποτε διαφορά μεταξύ των ομάδων (η οποία επιβιώνει από μια στατιστική ανάλυση, όπως περιγράφεται στο Μέρος 2 του βιβλίου) μπορεί να ερμηνευτεί ότι προκαλείται από την ανεξάρτητη μεταβλητή, ή, αλλιώς, ότι η ανεξάρτητη μεταβλητή είναι το αίτιο της διαφοράς.

Αν, για παράδειγμα, υπάρχει διαφορά υπέρ της ομάδας ενεργητικής ακρόασης, ο ψυχολόγος μπορεί να συμπεράνει ότι η εκπαίδευση στην ενεργητική ακρόαση προκαλεί λιγότερα προβλήματα επικοινωνίας μεταξύ των ζευγαριών. Αφού παρατηρήσει αυτήν τη σχέση, ο ψυχολόγος μπορεί να αναμένει ότι, αν τα νέα ζευγάρια εκπαιδευτούν στην ενεργητική ακρόαση, θα υπάρχουν λιγότερα προβλήματα στη μεταξύ τους επικοινωνία.

Εξαρτημένη μεταβλητή

Προκειμένου να δοκιμάσει αν η εκπαίδευση επηρεάζει την επικοινωνία, ο ψυχολόγος μετρά τον αριθμό των προστριβών που υπάρχουν σε κάθε ζευγάρι, όπως αποκαλύπτονται από ανάρμοστες απαντήσεις, επιθετικά σχόλια,

Εξαρτημένη μεταβλητή

Μια μεταβλητή που θεωρείται ότι επηρεάζεται από την ανεξάρτητη μεταβλητή.

προφορικές διακοπές κ.λπ., κατά τη συζήτηση θεμάτων που θα μπορούσαν να προκαλέσουν προστριβή, όπως αν είναι αποδεκτό να έχουν στενή σχέση με ένα τρίτο πρόσωπο. Όταν μια μεταβλητή θεωρείται ότι επηρεάζεται από την ανεξάρτητη μεταβλητή, ονομάζεται **εξαρτημένη μεταβλητή**. Σε ένα πείραμα, η εξαρτημένη μεταβλητή μετριέται ή καταγράφεται από τον ερευνητή.

Αντίθετα από την ανεξάρτητη μεταβλητή, η εξαρτημένη μεταβλητή δεν είναι αντικείμενο χειρισμού του ερευνητή, αλλά αναπαριστά ένα αποτέλεσμα: τα δεδομένα που παράγονται από το πείραμα. Συνεπώς, οι τιμές που εμφανίζονται για την εξαρτημένη μεταβλητή δεν μπορούν να προσδιοριστούν εκ των προτέρων. Αν και ο ψυχολόγος εικάζει ότι τα ζευγάρια με ειδική εκπαίδευση θα έχουν λιγότερες προστριβές, πρέπει να δει και στην πράξη το πόσες, ακριβώς, προστριβές θα παρατηρηθούν σε κάθε ζευγάρι.

Ανεξάρτητη ή εξαρτημένη μεταβλητή;

Αρκεί λίγη μόνο εξάσκηση για να μπορείτε να αναγνωρίσετε αυτούς τους τύπους μεταβλητών. Σε ένα πείραμα, ποιο είναι το αντικείμενο που χειρίζεται ο ερευνητής στην αρχή της μελέτης και, επομένως, μπορεί αυτό να θεωρηθεί ως η ανεξάρτητη μεταβλητή; Τι μετριέται ή καταγράφεται από τον ερευνητή όταν ολοκληρωθεί η μελέτη και, επομένως, μπορεί να θεωρηθεί ως εξαρτημένη μεταβλητή; Αφού αναγνωριστούν αυτές οι δύο μεταβλητές, μπορούν να χρησιμοποιηθούν για την περιγραφή του προβλήματος που θέτει η μελέτη, δηλαδή το εξής ερώτημα: Η ανεξάρτητη μεταβλητή προκαλεί αλλαγή στην εξαρτημένη μεταβλητή;⁴

4. Στο συγκεκριμένο παράδειγμα, σημειώστε ότι η ανεξάρτητη μεταβλητή (είδος εκπαίδευσης) είναι ποιοτική με ονομαστική μέτρηση, ενώ η εξαρτημένη μεταβλητή (πλήθος προβλημάτων επικοινωνίας) είναι ποσοτική. Στον βαθμό που το πλήθος των προβλημάτων επικοινωνίας χρησιμοποιείται για να προσδιοριστεί η ποιότητα της σχέσης ζευγαριών, το επίπεδο μέτρησης είναι κατά προσέγγιση διαστήματος.