

1

Η αναζήτηση της αιτιότητας

Πώς ξέρουμε αυτά που ξέρουμε; Ή τουλάχιστον, γιατί σκεφτόμαστε ό,τι σκεφτόμαστε; Η σύγχρονη απάντηση είναι τα αποδεικτικά στοιχεία.. Για να πείσουμε τους άλλους –και για να πείσουμε τον εαυτό μας– θα πρέπει να παρέχουμε πληροφορίες που μπορούν να επαληθευτούν. Κάτι που γνωρίζουμε διαισθητικά ή κάτι που απλώς «ξέρουμε» μπορεί να είναι σημαντικό, αλλά δεν αποτελεί αποδεικτικό στοιχείο, όπως αυτά που καθοδηγούν τη σύγχρονη επιστημονική έρευνα.

Ποια είναι η βάση των αποδείξεών μας; Σε ορισμένες περιπτώσεις, μπορούμε να δούμε την αιτία και το αποτέλεσμα. Βλέπουμε ένα αναμμένο κερί να αναποδογυρίζει και να ανάβει φωτιά. Τώρα ξέρουμε τι προκάλεσε τη φωτιά. Αυτή είναι μια πολύ καλή γνώση. Μερικές φορές στις πολιτικές επιστήμες και στην πολιτική εντοπίζουμε μια αλυσίδα αιτιότητας με παρόμοιο τρόπο. Ωστόσο, αυτή η διαδικασία μπορεί να



γίνει περίπλοκη. Γιατί ορισμένες οικονομίες μένουν στάσιμες ενώ άλλες ευδοκιμούν; Ποιες είναι οι οικονομικές και κοινωνικές επιπτώσεις του διεθνούς εμπορίου; Γιατί ο Donald Trump κέρδισε τις προεδρικές εκλογές το 2016; Γιατί έχει μειωθεί η εγκληματικότητα στις Ηνωμένες Πολιτείες; Για τέτοιου είδους ερωτήσεις, δεν μπορούμε να λαμβάνουμε υπόψη μόνο ένα αναμμένο κερί: υπάρχουν και οι κεραυνοί, τα ελαττωματικά καλώδια, οι εμπρηστές, και ποιος ξέρει τι άλλο. Σαφώς, σε αυτές τις σύνθετες περιπτώσεις είναι πολύ πιο δύσκολο να εντοπιστούν η αιτία και το αποτέλεσμα.

Όταν είναι δυνατή η άμεση παρατήρηση της αιτίας και του αποτελέσματος, στρεφόμαστε φυσικά στα δεδομένα. Και τα δεδομένα υπόσχονται πολλά. Κάποιο κτίριο κα-

Συσχέτιση \neq Αιτιότητα

ΣΧΗΜΑ 1.1: Κανόνας #1

ταρρέει κατά τη διάρκεια ενός σεισμού. Τι οδήγησε το κτίριο -και όχι άλλα στην ίδια πόλη- στην κατάρρευση; Ευθύνονται τα οικοδομικά υλικά; Το ύψος του; Ο σχεδιασμός του; Η ηλικία του; Η εσφαλμένη τοποθεσία; Παρότι ενδέχεται να μην είναι δυνατό να δούμε άμεσα την αιτία, μπορούμε να συλλέξουμε πληροφορίες για τα κτίρια που κατέρρευσαν και για εκείνα που έμειναν όρθια. Αν τα παλαιότερα κτίρια είναι πιο πιθανό να καταρρεύσουν, θα μπορούσαμε εύλογα να υποψιαστούμε ότι η ηλικία του κτιρίου έπαιξε ρόλο στην κατάρρευσή του. Αν τα κτίρια που κατασκευάζονταν χωρίς ενίσχυση από χάλυβα είναι πιο πιθανό να καταρρεύσουν ανεξάρτητα από την ηλικία τους, θα μπορούσαμε εύλογα να υποψιαστούμε ότι η κατασκευή του κτιρίου σχετίζεται με την κατάρρευσή του.

Κι όμως, δεν πρέπει να νιώθουμε απόλυτη σιγουριά. Ακόμα κι αν τα παλιά κτίρια είναι πιο πιθανό να καταρρεύσουν, δεν γνωρίζουμε με βεβαιότητα ότι η ηλικία του κτιρίου είναι η κύρια εξήγηση για την κατάρρευση. Θα μπορούσε να ευθύνεται το γεγονός ότι τα περισσότερα κτίρια σχεδιάζονταν με έναν συγκεκριμένο τρόπο μια ορισμένη χρονική περίοδο. Θα μπορούσε να ευθύνεται το γεγονός ότι υπάρχουν περισσότερα παλιά κτίρια σε μια γειτονιά όπου η σεισμική δραστηριότητα είναι πιο έντονη. Ή η κατάρρευση πολλών κτιρίων που έτυχε να είναι παλιά θα μπορούσε να αποτελεί μια τεράστια σύμπτωση. Με άλλα λόγια, η συσχέτιση δεν είναι το ίδιο με την αιτιότητα. Επισημαίνουμε αυτό το γεγονός με μεγάλα μπλε γράμματα στο Σχήμα 1.1 επειδή είναι ένα θεμελιώδες σημείο εκκίνησης σε κάθε σοβαρή ανάλυση δεδομένων.

Η οικονομετρία που μαθαίνουμε σε αυτό το βιβλίο θα μας βοηθήσει να εντοπίσουμε τις αιτίες και να διατυπώσουμε επιχειρήματα σχετικά με τις πραγματικές αιτίες ενός γεγονότος. Αν η συσχέτιση δεν είναι αιτιότητα, τι σημαίνει αιτιότητα; Θα χρειαστεί ολόκληρο το βιβλίο για να δοθεί μια πλήρης απάντηση, αλλά εδώ δίνουμε μια σύντομη εκδοχή: *αν μπορούμε να βρούμε εξωγενείς μεταβολές, τότε η συσχέτιση είναι πιθανώς αιτιότητα*. Επομένως, θα πρέπει να καταλάβουμε τι σημαίνει εξωγενής μεταβολή και πώς να διακρίνουμε την τυχαιότητα από την αιτιότητα όσο καλύτερα μπορούμε.

Σε αυτό το κεφάλαιο, εισάγουμε τρεις έννοιες που βρίσκονται στον πυρήνα του βιβλίου. Η Ενότητα 1.1 εξηγεί το βασικό υπόδειγμα που χρησιμοποιούμε σε όλο το σύγγραμμα. Η Ενότητα 1.2 παρουσιάζει δύο μεγάλες προκλήσεις που μπορεί να δυσχεράνουν τη χρήση δεδομένων για την κατανόηση του κόσμου. Δεν είναι τα μαθηματικά. (Αλήθεια!) Η πρώτη πρόκληση είναι η τυχαιότητα: μερικές φορές η στοχαστικότητα της δειγματοληψίας μας ωθεί να παρατηρούμε σχέσεις που δεν είναι πραγματικές. Άλλες φορές η τυχαία πιθανότητα μας κάνει να αγνοούμε σχέσεις που είναι πραγματικές. Η δεύτερη πρόκληση είναι η ενδογένεια, ένα φαινόμενο που μπορεί να μας κάνει να πιστεύουμε λανθασμένα ότι μια μεταβλητή προκαλεί κάποιο αποτέλεσμα ενώ δεν το κάνει. Στην Ενότητα 1.3 παρουσιάζονται τα τυχαίοποιημένα πειράματα ως ο ιδανικός τρό-

πος για να ξεπεραστεί η ενδογένεια. Συνήθως, αυτά τα πειράματα δεν είναι δυνατό να διεξαχθούν, αλλά ακόμη και όταν υπάρχει αυτή η δυνατότητα, τα πράγματα μπορεί να πάνε στραβά. Ως εκ τούτου, το υπόλοιπο βιβλίο αφορά την ανάπτυξη μιας εργαλειοθήκης που μας βοηθά να ανταποκριθούμε (ή να προσεγγίσουμε) στο εξιδανικευμένο πρότυπο των τυχαιοποιημένων πειραμάτων.

1.1 Το βασικό υπόδειγμα

Όταν μιλάμε για αιτία και αποτέλεσμα, θα αναφερόμαστε στο αποτέλεσμα ενδιαφέροντος ως την **εξαρτημένη μεταβλητή**. Θα αναφερόμαστε σε μια πιθανή αιτία ως την **ανεξάρτητη μεταβλητή**. Η εξαρτημένη μεταβλητή, που συνήθως συμβολίζεται ως Y , ονομάζεται έτσι επειδή η τιμή της εξαρτάται από την ανεξάρτητη μεταβλητή. Η ανεξάρτητη μεταβλητή, που συνήθως συμβολίζεται με X , ονομάζεται έτσι γιατί κάνει ό,τι θέλει. Είναι δυνατικά η αιτία κάποιας μεταβολής στην εξαρτημένη μεταβλητή.

Στον πυρήνα τους, οι κοινωνικές επιστημονικές θεωρίες υποστηρίζουν ότι μια μεταβολή σε κάτι (την ανεξάρτητη μεταβλητή) οδηγεί σε μεταβολή σε κάτι άλλο (την εξαρτημένη μεταβλητή). Θα εκφράζουμε με πιο τυπικό τρόπο αυτή τη σχέση στη συνέχεια. Τώρα, ας ξεκινήσουμε με ένα παράδειγμα. Ας υποθέσουμε ότι μας ενδιαφέρει η μεγάλη αύξηση της παχυσαρκίας στις ΗΠΑ και θέλουμε να αναλύσουμε την επίδραση της διατροφής στην υγεία. Μπορεί να αναρωτηθούμε, για παράδειγμα, αν τα ντόνατ προκαλούν προβλήματα υγείας. Στο υπόδειγμά μας η κατανάλωση ντόνατ (Donuts, μεταβλητή X , η ανεξάρτητη μεταβλητή μας) προκαλεί κάποια μεταβολή στο βάρος (Weight, μεταβλητή Y , η εξαρτημένη μεταβλητή μας). Αν καταφέρουμε να βρούμε δεδομένα για την ποσότητα ντόνατ που κατανάλωσαν οι άνθρωποι και το βάρος τους, μπορεί να βρεθούμε στα πρόθυρα μιας επιστημονικής ανακάλυψης.

Ας φανταστούμε μια μικρή μεσοδυτική πόλη των ΗΠΑ και ας κάνουμε μια μικρή έρευνα. Το Σχήμα 1.2 απεικονίζει την κατανάλωση ντόνατ και το βάρος 13 ατόμων από μια τυχαία επιλεγμένη πόλη: το Σπρίνγκφιλντ των ΗΠΑ. Τα ακατέργαστα δεδομένα μας εμφανίζονται στον Πίνακα 1.1. Κάθε άτομο έχει μια γραμμή στον πίνακα. Ο Homer είναι η παρατήρηση 1. Εφόσον έφαγε 14 ντόνατ την εβδομάδα, Donuts₁ = 14. Θα αναφερόμαστε συχνά στο X_i ή το Y_i , που είναι οι τιμές των X και Y για το άτομο i στο σύνολο δεδομένων. Το βάρος του έβδομου ατόμου στο σύνολο δεδομένων, του Smithers, είναι 160 λίβρες,¹ που σημαίνει Weight₇ = 160, Κ.Ο.Κ.

Το Σχήμα 1.2 είναι ένα **διάγραμμα διασποράς** δεδομένων, με κάθε παρατήρηση να βρίσκεται στις συντεταγμένες

► εξαρτημένη μεταβλητή

Το αποτέλεσμα ενδιαφέροντος, που συνήθως συμβολίζεται ως Y .

► ανεξάρτητη μεταβλητή

Μια μεταβλητή που πιθανώς επηρεάζει την τιμή της εξαρτημένης μεταβλητής.

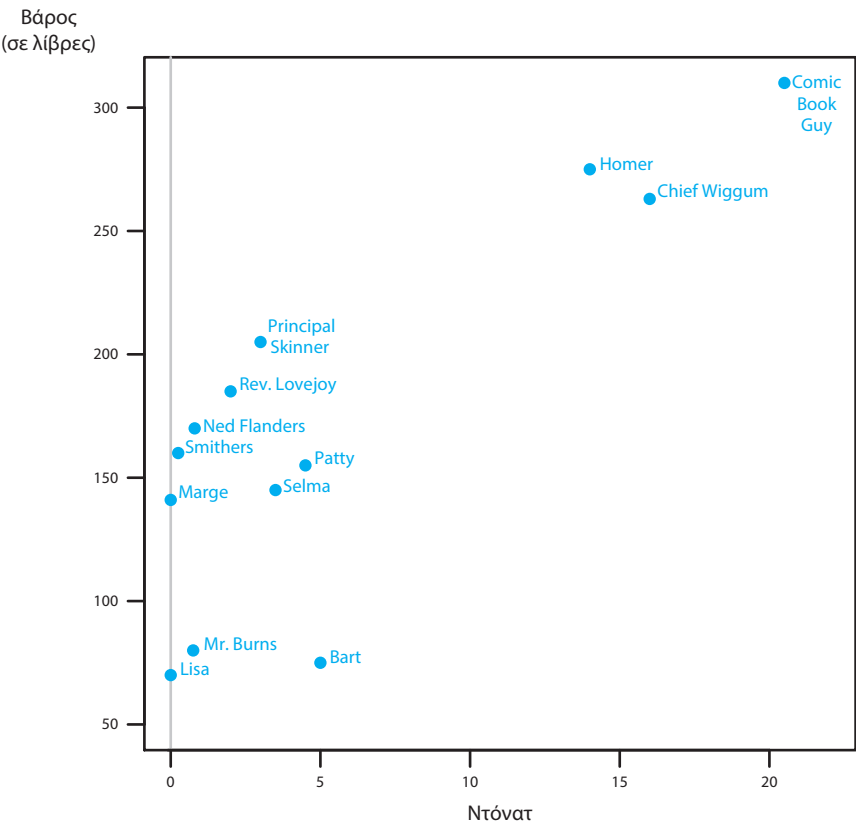
► διάγραμμα διασποράς

Ένα διάγραμμα δεδομένων στο οποίο κάθε παρατήρηση βρίσκεται στις συντεταγμένες που ορίζονται από τις ανεξάρτητες και τις εξαρτημένες μεταβλητές

1. Η 1 λίβρα είναι περίπου ίση με 0.454 κιλά. (ΣτΕ)

ΔΕΙΓΜΑ ΠΡΙΝ ΤΙΣ ΔΙΟΡΘΩΣΕΙΣ

46 / ΕΦΑΡΜΟΣΜΕΝΗ ΟΙΚΟΝΟΜΕΤΡΙΑ



ΣΧΗΜΑ 1.2: Βάρος και ντόνατ στο Σπρίνγκφιλντ

ΠΙΝΑΚΑΣ 2.2: Κατανάλωση ντόνατς και βάρος

Αριθμός παρατήρησης	Όνομα	Ντόνατ ανά εβδομάδα	Βάρος (λίβρες)
1	Homer	14	275
2	Marge	0	141
3	Lisa	0	70
4	Bart	5	75
5	Comic Book Guy	20	310
6	Mr. Burns	0.75	80
7	Smithers	0.25	160
8	Chief Wiggum	16	263
9	Principal Skinner	3	205
10	Rev. Lovejoy	2	185
11	Ned Flanders	0.8	170
12	Patty	5	155
13	Selma	4	145

που ορίζονται από τις ανεξάρτητες και τις εξαρτημένες μεταβλητές. Η τιμή των ντόνατ ανά εβδομάδα βρίσκεται στον άξονα X και το βάρος στον άξονα Y . Κοιτάζοντας αυτό το διάγραμμα, αντιλαμβανόμαστε ότι υπάρχει μια θετική σχέση μεταξύ των ντόνατ και του βάρους, επειδή όσο περισσότερα ντόνατ καταναλώνονται από κάποιο άτομο τόσο υψηλότερο τείνει να είναι το βάρος του.

Χρησιμοποιούμε μια απλή εξίσωση για να χαρακτηρίσουμε τη σχέση μεταξύ των δύο μεταβλητών:

$$Weight_i = \beta_0 + \beta_1 Donuts_i + \varepsilon_i \quad (1.1)$$

- Η εξαρτημένη μεταβλητή, $Weight_i$, είναι το βάρος του ατόμου i .
- Η ανεξάρτητη μεταβλητή, $Donuts_i$, είναι ο αριθμός των ντόνατ που καταναλώνει αυτό το άτομο σε μία εβδομάδα.
- Το β_1 είναι ο **συντελεστής κλίσης** στα ντόνατ, που δείχνει πόσο περισσότερο² ζυγίζει ένα άτομο για κάθε ντόνατ που τρώει.
- Το β_0 είναι ο **σταθερός όρος** ή **σημείο τομής**, που δείχνει το αναμενόμενο βάρος των ατόμων που τρώνε μηδέν ντόνατ.
- ε_i είναι ο **όρος σφάλματος** που καταγράφει οτιδήποτε άλλο επηρεάζει το βάρος.

► **συντελεστής κλίσης** Ο συντελεστής μιας ανεξάρτητης μεταβλητής. Δείχνει πόσο αυξάνεται η εξαρτημένη μεταβλητή όταν η ανεξάρτητη μεταβλητή αυξάνεται κατά μία μονάδα

► **σταθερός όρος** Η παράμετρος β_0 σε ένα υπόδειγμα παλινδρόμησης. Είναι το σημείο στο οποίο μια γραμμή παλινδρόμησης τέμνει τον άξονα Y . Επίσης αναφέρεται ως σημείο τομής.

► **όρος σφάλματος** Ο όρος που σχετίζεται με μη μετρήσιμους παράγοντες σε ένα υπόδειγμα παλινδρόμησης, τυπικά συμβολίζεται με ε .

Αυτή η εξίσωση θα μας βοηθήσει να εκτιμήσουμε τις δύο παραμέτρους που είναι απαραίτητες για να χαρακτηριστεί μια γραμμή. Θυμάστε τη σχέση $Y = mX + b$ από το γυμνάσιο; Αυτή είναι η εξίσωση για μια γραμμή όπου Y είναι η τιμή της γραμμής στον κατακόρυφο άξονα, X είναι η τιμή στον οριζόντιο άξονα, m είναι η κλίση και b είναι ο σταθερός όρος ή η τιμή του Y όταν το X είναι μηδέν. Η Εξίσωση 1.1 είναι ουσιαστικά η ίδια, μόνο που αναφερόμαστε στον όρο « b » ως β_0 και ονομάζουμε τον όρο « m » β_1 .

Το Σχήμα 1.3 παρουσιάζει ένα παράδειγμα μιας πιθανής γραμμής από αυτό το υπόδειγμα για τα δεδομένα μας στο Σπρίνγκφιλντ. Το σημείο τομής (β_0) είναι η τιμή του βάρους όταν η κατανάλωση ντόνατ είναι μηδέν ($X = 0$). Η κλίση (β_1) είναι η ποσότητα στην οποία αυξάνεται το βάρος για κάθε ντόνατ που καταναλώνεται. Σε αυτή την περίπτωση, ο σταθερός όρος είναι περίπου 123, το οποίο σημαίνει ότι το αναμενόμενο βάρος για όσους δεν τρώνε κανένα ντόνατ είναι περίπου 123 λίβρες. Η κλίση είναι περίπου 9.1, το οποίο σημαίνει ότι για κάθε ντόνατ που καταναλώνεται την εβδομάδα, το βάρος αυξάνεται κατά περίπου 9.1 λίβρες.

2. Ή λιγότερο – αν είμαστε αισιόδοξοι!