

Περιγραφή δεδομένων: Γραφική

- 1.1 Λήψη αποφάσεων σε αβέβαιο περιβάλλον
Τυχαία και συστηματική δειγματοληψία
Δειγματοληπτικά και μη δειγματοληπτικά σφάλματα
- 1.2 Ταξινόμηση μεταβλητών
Κατηγορικές και αριθμητικές μεταβλητές
Κλίμακες μέτρησης
- 1.3 Γραφήματα περιγραφής κατηγορικών μεταβλητών
Πίνακες και διαγράμματα
Πίνακες διασταύρωσης
Διαγράμματα πίτας
Διαγράμματα Pareto
- 1.4 Γραφήματα περιγραφής δεδομένων χρονοσειράς
- 1.5 Γραφήματα περιγραφής αριθμητικών μεταβλητών
Κατανομές συχνότητας
Ιστογράμματα και γραφήματα αθροιστικών συχνοτήτων
Σχήμα κατανομής
Διαγράμματα μίσχου-φύλλου
Διαγράμματα διασποράς
- 1.6 Σφάλματα παρουσίασης δεδομένων
Παραπλανητικά ιστογράμματα
Παραπλανητικά διαγράμματα χρονοσειράς

Εισαγωγή

Ποιες είναι οι προβλεπόμενες πωλήσεις ενός νέου προϊόντος; Η τιμή των μετοχών της Google θα συνεχίσει την ανοδική πορεία της; Ποια ομάδα θα κερδίσει το επόμενο UEFA Champions League; Πόσο ικανοποιημένοι είστε με την τελευταία αγορά σας στα Starbucks, alibaba.com ή IKEA; Εάν εργαζόσασταν στο Εθνικό Συμβούλιο Διατροφής της χώρας σας, πώς θα διαπιστώνατε αν ακολουθούνται οι κατευθύνσεις του Συμβουλίου για την κατανάλωση φρούτων, λαχανικών, πρόχειρου φαγητού και αναψυκτικών; Τα άτομα που ασκούνται ακολουθούν πιο υγιεινή διατροφή από όσους δεν ασκούνται; Ποιοι παράγοντες (ενδεχομένως το διαθέσιμο εισόδημα ή επιχορηγήσεις) είναι σημαντικοί για την πρόβλεψη της συνολικής κατανάλωσης διαρκών αγαθών; Ποια επίδραση θα έχει μια αύξηση των επιτοκίων κατά 2% σε επενδύσεις σε κατοικίες; Η πιστοληπτική ικανότητα, το τρέχον υπόλοιπο ή τα ανεξόφλητα ποσά συνεισφέρουν σε μια αύξηση στο ποσοστό των

εκπρόθεσμων λογαριασμών μιας εταιρείας παροχής ενυπόθηκων δανείων; Οι απαντήσεις σε τέτοιες ερωτήσεις προκύπτουν από την κατανόηση της στατιστικής, των διακυμάνσεων στην αγορά, των καταναλωτικών προτιμήσεων, των τάσεων κ.ο.κ.

Η στατιστική χρησιμοποιείται για την πρόβλεψη των πωλήσεων ενός νέου προϊόντος, του κόστους παραγωγής, των επιπέδων ικανοποίησης πελατών, του καιρού, των εκλογικών αποτελεσμάτων, του αριθμού εγγραφών σε ένα πανεπιστήμιο, του μέσου όρου της βαθμολογίας των σπουδαστών, των επιτοκίων, των συναλλαγματικών ισοτιμιών και πολλών άλλων μεταβλητών που επηρεάζουν την καθημερινότητά μας. Υπάρχει η ανάγκη να αφομοιώνουμε και να ερμηνεύουμε σημαντικές ποσότητες δεδομένων. Οι κυβερνήσεις, οι επιχειρήσεις και οι επιστημονικοί ερευνητές δαπανούν δισεκατομμύρια για τη συλλογή δεδομένων. Αλλά αφού συλλεχθούν τα δεδομένα, τι κάνουμε με αυτά; Πώς επηρεάζουν τα δεδομένα τη λήψη αποφάσεων;

Κατά τη μελέτη της *στατιστικής* μαθαίνουμε να χρησιμοποιούμε πολλά εργαλεία που μας βοηθούν να επεξεργαστούμε, να συνοψίσουμε, να αναλύσουμε και να ερμηνεύσουμε δεδομένα προκειμένου να λάβουμε καλύτερες αποφάσεις σε ένα αβέβαιο περιβάλλον. Ουσιαστικά, η κατανόηση της στατιστικής θα μας βοηθήσει να αντιληφθούμε καλύτερα όλα αυτά τα δεδομένα.

Σε αυτό το κεφάλαιο παρουσιάζουμε πίνακες και γραφήματα τα οποία υποστηρίζουν την καλύτερη κατανόηση των δεδομένων και παρέχουν μια οπτική βάση για την καλύτερη λήψη αποφάσεων. Οι εκθέσεις εμπλουτίζονται με την ενσωμάτωση κατάλληλων πινάκων και γραφημάτων, όπως κατανομές συχνότητας, ραβδογράμματα, διαγράμματα πίτας, διαγράμματα Pareto, διαγράμματα γραμμής, ιστογράμματα, διαγράμματα μίσχου-φύλλου ή γραφήματα αθροιστικών συχνοτήτων. Η απεικόνιση των δεδομένων είναι σημαντική και πρέπει πάντοτε να θέτουμε τις εξής ερωτήσεις: Τι δείχνει το γράφημα για τα δεδομένα; Τι μπορούμε να δούμε;

1.1 Λήψη αποφάσεων σε αβέβαιο περιβάλλον

Οι αποφάσεις λαμβάνονται συχνά βάσει περιορισμένων πληροφοριών. Οι λογιστές μπορεί να χρειαστεί να επιλέξουν ένα μέρος των εγγραφών για έλεγχο. Οι χρηματοοικονομικοί επενδυτές πρέπει να κατανοούν τις διακυμάνσεις των αγορών και να επιλέγουν μεταξύ διαφορετικών επενδύσεων χαρτοφυλακίου. Οι διευθυντές μπορεί να χρησιμοποιούν έρευνες για να διαπιστώνουν αν οι πελάτες είναι ικανοποιημένοι με τα προϊόντα ή τις υπηρεσίες της εταιρείας τους. Ένα στέλεχος του μάρκετινγκ ενδεχομένως να χρειάζεται πληροφορίες που αφορούν τις προτιμήσεις των καταναλωτών, τις αγοραστικές τους συνήθειες ή τα δημογραφικά στοιχεία των διαδικτυακών αγοραστών. Ένας επενδυτής δεν γνωρίζει με βεβαιότητα αν οι χρηματοοικονομικές αγορές θα είναι ανοδικές, σταθερές ή καθοδικές. Ωστόσο, πρέπει να αποφασίσει πώς θα κατανείμει ένα χαρτοφυλάκιο μεταξύ μετοχών, ομολόγων και τίτλων της χρηματαγοράς τη στιγμή που οι μελλοντικές κινήσεις των αγορών παραμένουν άγνωστες.

Σε όλες αυτές τις περιπτώσεις, πρέπει να ορίζουμε προσεκτικά το πρόβλημα, να προσδιορίζουμε τα δεδομένα που χρειαζόμαστε, να τα συλλέγουμε και να χρησιμοποιούμε τη στατιστική για να συνοψίσουμε τα δεδομένα, να εξαγάγουμε συμπεράσματα και να λάβουμε αποφάσεις βάσει των κτηθέντων δεδομένων. Η στατιστική σκέψη είναι απαραίτητη από τον αρχικό ορισμό του προβλήματος έως την τελική απόφαση, η οποία μπορεί να συμβάλει στη μείωση του κόστους, στην αύξηση του κέρδους, στη βελτίωση των διαδικασιών και την αυξημένη ικανοποίηση των πελατών.

Τυχαία και συστηματική δειγματοληψία

Πριν από την εισαγωγή ενός νέου προϊόντος στην αγορά, ο παραγωγός του θα ήθελε να έχει κάποια εκτίμηση για το πιθανό επίπεδο ζήτησης και για αυτόν τον λόγο μπορεί να διεξάγει μια σχετική έρευνα αγοράς. Ο παραγωγός ενδιαφέρεται στην πραγματικότητα για όλους τους υποψήφιους αγοραστές (ο πληθυσμός). Ωστόσο, οι πληθυσμοί πολλές φορές είναι τόσο μεγάλοι, ώστε η ανάλυσή τους είναι δύσκολη και η συλλογή ολοκληρωμένων πληροφοριών για έναν πληθυσμό θα μπορούσε να είναι αδύνατη ή απαγορευτικά ακριβή. Ακόμα και σε περιπτώσεις στις οποίες φαίνεται να υπάρχουν επαρκείς πόροι, οι χρονικοί περιορισμοί καθιστούν απαραίτητη την εξέταση ενός υποσυνόλου (το δείγμα).

Πληθυσμός και Δείγμα

Πληθυσμός (population) είναι το πλήρες σύνολο όλων των στοιχείων που ενδιαφέρουν έναν ερευνητή. Το μέγεθος του πληθυσμού, N , μπορεί να είναι πολύ μεγάλο ή ακόμα και άπειρο. **Δείγμα** (sample) είναι ένα παρατηρούμενο υποσύνολο (ή τμήμα) ενός πληθυσμού με το μέγεθος του δείγματος να συμβολίζεται ως n .

Παραδείγματα πληθυσμών είναι τα εξής:

- Όλοι οι υποψήφιοι αγοραστές ενός νέου προϊόντος
- Όλες οι μετοχές που τελούν υπό διαπραγμάτευση στο Χρηματιστήριο του Λονδίνου (LSE)
- Όλοι οι εγγεγραμμένοι ψηφοφόροι ενός συγκεκριμένου δήμου ή περιφέρειας
- Όλοι οι εισπρακτέοι λογαριασμοί για μια επιχείρηση

Ο τελικός στόχος είναι να διατυπώσουμε προτάσεις βάσει των δεδομένων του δείγματος που έχουν κάποια εφαρμογή στον πληθυσμό γενικά. Χρειαζόμαστε επομένως ένα αντιπροσωπευτικό δείγμα του πληθυσμού. Πώς μπορούμε να το πετύχουμε; Μια σημαντική αρχή που πρέπει να ακολουθούμε κατά τη διαδικασία επιλογής του δείγματος είναι η τυχαιότητα.

Τυχαία Δειγματοληψία

Η **απλή τυχαία δειγματοληψία** (simple random sampling) είναι μια μέθοδος που χρησιμοποιείται για την επιλογή ενός δείγματος n αντικειμένων από έναν πληθυσμό, έτσι ώστε κάθε μέλος του πληθυσμού να επιλέγεται αυστηρώς τυχαία, η επιλογή ενός μέλους να μην επηρεάζει την επιλογή οποιουδήποτε άλλου μέλους, κάθε μέλος του πληθυσμού να είναι εξίσου πιθανό να επιλεγεί και κάθε πιθανό δείγμα δεδομένου μεγέθους, n , να έχει την ίδια πιθανότητα επιλογής. Αυτή η μέθοδος είναι τόσο κοινή ώστε το επίθετο *απλή* παραλείπεται γενικά και το τελικό δείγμα αναφέρεται ως **τυχαίο δείγμα** (random sample).

Ακόμα μια μέθοδος δειγματοληψίας είναι η συστηματική δειγματοληψία (η στρωματοποιημένη δειγματοληψία και η δειγματοληψία κατά ομάδες περιγράφονται στο Κεφάλαιο 17).

Συστηματική Δειγματοληψία

Έστω ότι τα στοιχεία του πληθυσμού διατάσσονται με κάποιον τρόπο που δεν συνδέεται με το θέμα του ενδιαφέροντος. Η **συστηματική δειγματοληψία** (systematic sampling) περιλαμβάνει την επιλογή κάθε j -οστού στοιχείου του πληθυσμού, όπου j είναι ο λόγος του μεγέθους του πληθυσμού, N , προς το επιθυμητό μέγεθος του δείγματος, n , δηλαδή $j = N/n$. Επιλέγετε τυχαία έναν αριθμό από το 1 έως το j για να πάρετε το πρώτο στοιχείο που θα συμπεριληφθεί στο συστηματικό δείγμα σας.

Έστω ότι το ζητούμενο μέγεθος δείγματος είναι 100 και ότι ο πληθυσμός αποτελείται από 5.000 ονόματα σε αλφαβητική σειρά. Τότε $j = 50$. Επιλέγετε τυχαία έναν αριθμό από το 1 έως το 50. Εάν ο αριθμός σας είναι το 20, επιλέγετε αυτόν και κάθε διαδοχικό 50ό αριθμό, σχηματίζοντας το συστηματικό δείγμα με τα στοιχεία με αριθμό 20, 70, 120, 170 κ.ο.κ., μέχρι να επιλεγθούν τα 100 ζητούμενα στοιχεία. Το συστηματικό δείγμα αναλύεται με τον ίδιο τρόπο όπως ένα απλό τυχαίο δείγμα βάσει του ότι, σε σχέση με το θέμα ενδιαφέροντος, ο κατάλογος του πληθυσμού είναι ήδη σε τυχαία σειρά. Ο κίνδυνος είναι ότι θα μπορούσε να υπάρχει κάποια ανεπαίσθητη, απαρτήρητη σύνδεση μεταξύ της διάταξης του πληθυσμού και του υπό μελέτη θέματος. Εάν ίσχυε αυτό, θα υπήρχε μεροληψία εφόσον εφαρμοζόταν συστηματική δειγματοληψία. Τα συστηματικά δείγματα παρέχουν μια καλή αντιπροσώπευση του πληθυσμού αν δεν υπάρχει οποιαδήποτε κυκλική διακύμανση στον πληθυσμό.

Δειγματοληπτικά και μη δειγματοληπτικά σφάλματα

Έστω ότι θέλουμε να μάθουμε τη μέση ηλικία των εγγεγραμμένων ψηφοφόρων στις Ηνωμένες Πολιτείες Αμερικής. Προφανώς το μέγεθος του πληθυσμού είναι τόσο μεγάλο που μπορούμε να πάρουμε μόνο ένα τυχαίο δείγμα, ίσως 500 εγγεγραμμένους ψηφοφόρους, και να υπολογίσουμε τη μέση ηλικία τους. Επειδή αυτός ο μέσος όρος βασίζεται στα δεδομένα ενός δείγματος, ονομάζεται *στατιστική*. Εάν ήμασταν σε θέση να υπολογίσουμε τη μέση ηλικία ολόκληρου του πληθυσμού, τότε ο μέσος όρος που θα παίρναμε θα λεγόταν *παράμετρος*.

Παράμετρος και στατιστική

Παράμετρος (parameter) είναι ένα αριθμητικό μέτρο που περιγράφει ένα συγκεκριμένο χαρακτηριστικό ενός πληθυσμού. **Στατιστική** (statistic) είναι ένα αριθμητικό μέτρο που περιγράφει ένα συγκεκριμένο χαρακτηριστικό ενός δείγματος.

Σε όλο το βιβλίο μελετούμε τρόπους λήψης αποφάσεων για μια παράμετρο του πληθυσμού, με βάση μια στατιστική του δείγματος. Οφείλουμε να αναγνωρίσουμε ότι θα υπάρχει πάντα ένα στοιχείο αβεβαιότητας, καθώς δεν γνωρίζουμε την ακριβή τιμή της παραμέτρου. Με άλλα λόγια, όταν παίρνουμε ένα δείγμα από έναν πληθυσμό, δεν μπορούμε να γνωρίζουμε *επακριβώς* την τιμή οποιασδήποτε παραμέτρου του. Μια πηγή σφάλματος, το **δειγματοληπτικό σφάλμα** (sampling error), προκύπτει από το γεγονός ότι υπάρχουν διαθέσιμες πληροφορίες για ένα μόνο υποσύνολο όλων των μελών του πληθυσμού. Στα Κεφάλαια 6, 7 και 8 αναπτύσσουμε τη στατιστική θεωρία που μας επιτρέπει να χαρακτηρίζουμε τη φύση του δειγματοληπτικού σφάλματος και να διατυπώνουμε συγκεκριμένες προτάσεις για τις παραμέτρους του πληθυσμού.

Σε πρακτικές αναλύσεις υπάρχει η πιθανότητα σφάλματος που δεν συνδέεται με τον τύπο της μεθόδου δειγματοληψίας που εφαρμόζεται. Πράγματι, τέτοια σφάλματα θα μπορούσαν εξίσου πιθανώς να προκύψουν ακόμα κι αν κάναμε μια πλήρη απογραφή του πληθυσμού. Αυτά αναφέρονται ως **μη δειγματοληπτικά σφάλματα** (nonsampling error). Παραδείγματα μη δειγματοληπτικών σφαλμάτων είναι τα εξής:

1. *Ο πληθυσμός από τον οποίο δημιουργείται το δείγμα δεν είναι ο σχετικός.* Μια περίφημη τέτοια περίπτωση σημειώθηκε το 1936, όταν το περιοδικό *Literary Digest* προέβλεψε με βεβαιότητα ότι ο Άλφρεντ Λάντον θα κέρδιζε στις προεδρικές εκλογές τον Φράνκλιν Ρούσβελτ. Ωστόσο, ο Ρούσβελτ κέρδισε πολύ άνετα. Αυτή η εσφαλμένη πρόβλεψη ήταν το αποτέλεσμα σχηματισμού του δείγματος του περιοδικού από τηλεφωνικούς και άλλους καταλόγους, όπως καταλόγους συνδρομητών περιοδικών και ιδιοκτητών αυτοκινήτων. Σε αυτές τις πηγές υπήρχε σοβαρή υποεκπροσώπηση των φτωχών στρωμάτων του πληθυσμού, τα οποία ήταν κυρίως ψηφοφόροι των Δημοκρατικών. Για να εξαχθεί ένα επαγωγικό συμπέρασμα σχετικά με έναν πληθυσμό (εν προκειμένω το εκλογικό σώμα στις ΗΠΑ) είναι σημαντικό το δείγμα να λαμβάνεται από το σύνολο του πληθυσμού και όχι από κάποια υποομάδα του, όσο βολικό κι αν είναι αυτό.
2. *Τα υποκείμενα των ερευνών μπορεί να δώσουν ανακριβείς ή ανειλικρινείς απαντήσεις.* Αυτό είναι κάτι που θα μπορούσε να συμβεί επειδή οι ερωτήσεις διατυπώνονται με τρόπο δυσνόητο ή έτσι ώστε να δίνεται η εντύπωση ότι μια συγκεκριμένη απάντηση είναι πιο ευπρόσδεκτη ή επιθυμητή. Επίσης, πολλές ερωτήσεις που θα μπορούσαν να τεθούν είναι τόσο ευαίσθητες ώστε θα ήταν αφελές εκ μέρους του ερευνητή να θεωρήσει ότι θα λάβει ειλικρινείς απαντήσεις. Για παράδειγμα, έστω ότι ένας διευθυντής εργοστασίου θέλει να αξιολογήσει τις ετήσιες απώλειες που έχει η εταιρεία εξαιτίας κλοπών από το προσωπικό. Θεωρητικά, θα μπορούσε να επιλεχθεί ένα τυχαίο δείγμα εργαζομένων και στα μέλη του δείγματος να τεθεί το ερώτημα, «Τι έχετε κλέψει από το εργοστάσιο τους τελευταίους 12 μήνες;». Προφανώς αυτή η ερώτηση δεν είναι το πιο αξιόπιστο μέσο για τη λήψη των ζητούμενων πληροφοριών!
3. *Ενδεχομένως να μην υπάρχει καμία απάντηση στις ερωτήσεις της έρευνας.* Τα υποκείμενα της έρευνας μπορεί να μην απαντήσουν καθόλου ή να μην απαντήσουν σε ορισμένες ερωτήσεις. Εάν ο βαθμός της αποχής από απαντήσεις είναι μεγάλος, θα μπορούσε να προκαλέσει πρόσθετα δειγματοληπτικά και μη δειγματοληπτικά σφάλματα. Το δειγματοληπτικό σφάλμα προκύπτει επειδή το τελικό μέγεθος του δείγματος θα είναι μικρότερο από το επιθυμητό. Το μη δειγματοληπτικό σφάλμα συμβαίνει πιθανώς επειδή ο πληθυσμός από τον οποίο εξάγεται το δείγμα δεν είναι ο πληθυσμός ενδιαφέροντος. Τα ληφθέντα αποτελέσματα μπορούν να θεωρηθούν ως ένα τυχαίο δείγμα από τον πληθυσμό αυτών που είναι πρόθυμοι να απαντήσουν. Αυτά τα άτομα θα μπο-

ρούσαν να παρουσιάζουν σοβαρές διαφορές από τον ευρύτερο πληθυσμό και, σε αυτή την περίπτωση, οι τελικές εκτιμήσεις θα είναι μεροληπτικές.

Δεν υπάρχει κάποια γενική μέθοδος για την αναγνώριση και ανάλυση μη δειγματοληπτικών σφαλμάτων. Αλλά τα μη δειγματοληπτικά σφάλματα θα μπορούσαν να είναι σημαντικά. Ο ερευνητής οφείλει να προσδιορίζει τον σχετικό πληθυσμό, να σχεδιάζει το ερωτηματολόγιο και να χειρίζεται την αποχή από απαντήσεις προκειμένου να ελαχιστοποιεί τη βαρύτητα μη δειγματοληπτικών σφαλμάτων. Στο υπόλοιπο βιβλίο, θεωρούμε ότι ο ερευνητής φροντίζει αυτά τα θέματα και η ανάλυσή μας επικεντρώνεται στην αντιμετώπιση των δειγματοληπτικών σφαλμάτων.

Η στατιστική σκέψη ξεκινά από τον ορισμό του προβλήματος: (1) Ποιες πληροφορίες απαιτούνται; (2) Ποιος είναι ο σχετικός πληθυσμός; (3) Πώς πρέπει να επιλεγθούν τα μέλη του δείγματος; (4) Πώς πρέπει να ληφθούν οι πληροφορίες από τα μέλη του δείγματος; Στη συνέχεια, θα πρέπει να γνωρίζουμε πώς να χρησιμοποιήσουμε τις πληροφορίες από το δείγμα για να λάβουμε αποφάσεις για τον πληθυσμό που μας ενδιαφέρει. Τέλος, θα χρειαστεί να γνωρίζουμε ποια συμπεράσματα μπορούμε να εξαγάγουμε για τον πληθυσμό.

Αφού αναγνωριστεί και οριστεί ένα πρόβλημα, ξεκινά η συλλογή δεδομένων που παράγονται με διάφορες μεθόδους σύμφωνα με ένα σχέδιο και ακολούθως αυτά τα δεδομένα αναλύονται χρησιμοποιώντας μία ή περισσότερες στατιστικές μεθόδους. Από αυτή την ανάλυση λαμβάνονται πληροφορίες, οι οποίες με τη σειρά τους μετατρέπονται σε γνώση βάσει εμπειρίας, θεωρίας, βιβλιογραφίας και άλλων στατιστικών μεθόδων. Τόσο περιγραφικά όσο και επαγωγικά στατιστικά στοιχεία χρησιμοποιούνται για τη μετατροπή των δεδομένων σε γνώση που οδηγεί στη λήψη καλύτερων αποφάσεων και, προκειμένου να γίνει αυτό, εφαρμόζεται στην πράξη η περιγραφική στατιστική και η επαγωγική στατιστική.

Περιγραφική και επαγωγική στατιστική

Η **περιγραφική στατιστική** (descriptive statistics) επικεντρώνεται σε γραφικές και αριθμητικές μεθόδους που χρησιμοποιούνται για τη σύνοψη και την επεξεργασία δεδομένων. Η **επαγωγική στατιστική** (inferential statistics) εστιάζεται στη χρήση των δεδομένων για την πραγματοποίηση προβλέψεων και εκτιμήσεων που συμβάλλουν στη λήψη καλύτερων αποφάσεων.

1.2 Ταξινόμηση μεταβλητών

Οι μεταβλητές είναι συγκεκριμένα χαρακτηριστικά (όπως ηλικία ή βάρος) ενός ατόμου ή ενός αντικειμένου. Οι μεταβλητές μπορούν να ταξινομηθούν με πολλούς τρόπους. Μια μέθοδος ταξινόμησης αναφέρεται στον τύπο και τον όγκο των πληροφοριών που περιέχουν τα δεδομένα. Τα δεδομένα είναι κατηγορικά ή αριθμητικά. Μια άλλη μέθοδος, η οποία προτάθηκε το 1946 από τον Αμερικανό ψυχολόγο Stanley Smith Stevens, είναι η ταξινόμηση των δεδομένων κατά επίπεδα μέτρησης, δίνοντας ποιοτικές ή ποσοτικές μεταβλητές. Η ορθή ταξινόμηση των δεδομένων είναι ένα καίριο πρώτο βήμα για την επιλογή των ορθών στατιστικών μεθόδων που απαιτούνται για την ανάλυση και ερμηνεία δεδομένων.

Κατηγορικές και αριθμητικές μεταβλητές

Οι **κατηγορικές μεταβλητές** (categorical variable) παράγουν απαντήσεις που ανήκουν σε ομάδες ή κατηγορίες. Για παράδειγμα, οι απαντήσεις σε ερωτήσεις ναι/όχι είναι κατηγορικές. Οι ερωτήσεις «Έχετε πτυχίο στη διοίκηση επιχειρήσεων;» και «Έχετε αυτοκίνητο;» περιορίζονται σε απαντήσεις ναι ή όχι. Μια ασφαλιστική εταιρεία μπορεί να ταξινομή τις εσφαλμένες απαιτήσεις ανάλογα με τον τύπο του σφάλματος, όπως διαδικαστικά και διαγνωστικά λάθη, λάθη στα στοιχεία των ασθενών και συμβατικά λάθη. Άλλα παραδείγματα κατηγορικών μεταβλητών συμπεριλαμβάνουν ερωτήσεις για το φύλο ή την οικογενειακή κατάσταση. Οι κατηγορικές μεταβλητές περιλαμβάνουν ενίοτε ένα εύρος επιλογών, όπως «διαφωνώ κατηγορηματικά» έως «συμφωνώ κατηγορηματικά». Για παράδειγμα, αναλογιστείτε ένα έντυπο αξιολόγησης διδακτικού προσωπικού όπου οι φοιτητές καλούνται να απαντήσουν σε προτάσεις όπως: Ο διδάσκων σε αυτό το μάθημα ήταν αποτελεσματικός (1: διαφωνώ κατηγορηματικά, 2: διαφωνώ ελαφρώς, 3: ούτε συμφωνώ ούτε διαφωνώ, 4: συμφωνώ ελαφρώς, 5: συμφωνώ κατηγορηματικά).