

Κατανομές συχνοτήτων

Ο στόχος οποιασδήποτε συλλογής δεδομένων είναι κατανόηση του προβλήματος το οποίου ζητάμε την απάντηση και η εξαγωγή συμπερασμάτων τα οποία στη συνέχεια θα μπορέσουν με τον έναν ή τον άλλο τρόπο να τροφοδοτήσουν την πράξη. Πριν όμως το κάνουμε αυτό, σχεδόν πάντα, ξεκινάμε από μία απόπειρα κατανόησης των δεδομένων που μαζεύτηκαν. Δεν μπορεί κανείς να βγάλει εύκολα συμπεράσματα απλά κοιτώντας έναν πίνακα με μερικές δεκάδες στήλες και μερικές εκατοντάδες ή χιλιάδες γραμμές.

Μία από τις πιο απλές και χρήσιμες μεθόδους περιγραφής των δεδομένων που συλλέχθηκαν για ένα σύνολο μεταβλητών που μας ενδιαφέρει στηρίζεται στην έννοια της συχνότητας. Με απλά λόγια, η συχνότητα μίας τιμής μίας μεταβλητής που έχει μετρηθεί σε ένα δείγμα μας μετράει πόσες φορές εμφανίζεται η συγκεκριμένη μεταβλητή σε σχέση τις υπόλοιπες.

Ξεκινάμε από την περίπτωση των διακριτών μεταβλητών. Έστω ότι έχουμε τη μεταβλητή X , το πλήθος των μαθητών ανά τάξη, και ένα δείγμα από 20 νηπιαγωγεία της χώρας με τιμές x_1, \dots, x_{20} :

20	23	20	24	20
18	17	25	25	19
20	17	17	16	17
15	20	19	23	25

Βλέπουμε ότι οι τιμές κειμένονται από 15 μέχρι 25. Οπότε μπορούμε εύκολα να φτιάξουμε έναν πίνακα συχνοτήτων

x	f
15	1
16	1
17	5
18	1
19	2
20	5
23	2
24	1
25	3

Ως άσκηση, επιβεβαιώστε ότι $\sum X = \sum xf$.

Χρησιμοποιώντας τις συχνότητες εμφάνισης των τιμών μιας μεταβλητής μπορούμε να ορίσουμε τις αναλογίες. Η αναλογία q ορίζεται ως

$$q = \frac{f}{N},$$

όπου N είναι το μέγεθος του δείγματος. Επειδή η αναλογία περιγράφει τη συχνότητα μίας τιμής σε σχέση με μέγεθος του δείγματος, συχνά αποκαλείται και σχετική συχνότητα. Τέλος, μία αναλογία q μπορεί πολύ εύκολα να μετατραπεί σε ποσοστό p πολλαπλασιάζοντας το αποτέλεσμα με το 100

$$p = 100 \times q = 100 \times \frac{f}{N} \%$$

Στην προηγούμενη περίπτωση, το πλήθος των τιμών που τελικά πήρε η μεταβλητή X ήταν σχετικά μικρό (συνολικά 9 διαφορετικές τιμές). Τι γίνεται όμως αν το πλήθος των πιθανών τιμών είναι πολύ μεγάλο; Για παράδειγμα, έστω ότι ένα δείγμα 30 μαθητών δημιούργησε μία ζωγραφιά με τελείες. Έστω Y , το πλήθος των στιγμάτων που ζωγράφισε κάθε παιδί και οι τιμές που παρατηρήσαμε:

24	37	41	28	35
22	31	56	42	33
58	45	64	83	38
94	49	53	70	84
66	72	90	47	59
71	88	63	75	95

Για να δημιουργήσουμε έναν πίνακα συχνοτήτων ο οποίος να είναι χρήσιμος, μπορούμε να χωρίσουμε τις τιμές σε κατηγορίες. Για παράδειγμα, ένας πιθανός διαχωρισμός θα ήταν 21-40, 41-60, 61-80, 81-100, δημιουργώντας τον πίνακα συχνοτήτων

y	f
21-40	7
41-60	8
61-80	7
81-100	8