

Μεταβλητές

Ονομάζεται **μεταβλητή** οποιοδήποτε χαρακτηριστικό που παίρνει περισσότερες από μία διαφορετικές τιμές. Το εισόδημα, η ηλικία, το ύψος και το βάρος ατόμων είναι παραδείγματα μεταβλητών. Άλλα παραδείγματα είναι η τιμή μιας μετοχής, του πετρελαίου, του δολαρίου σε διαφορετικές χρονικές στιγμές.

Οι μεταβλητές διακρίνονται σε δύο κατηγορίες:

Ποιοτικές μεταβλητές είναι αυτές που αναφέρονται σε κάποιο ποιοτικό χαρακτηριστικό, όπως το φύλο, το επίπεδο μόρφωσης, η περιοχή καταγωγής κ.λ.π. Για παράδειγμα από το παραπάνω ερωτηματολόγιο οι μεταβλητές 1, 3-11, 13-14, 16-23, 26-27 είναι ποιοτικές.

Ποσοτικές είναι οι μεταβλητές των οποίων οι τιμές έχουν αριθμητικές ιδιότητες και εκφράζονται με μία μονάδα μέτρησης. Παραδείγματα ποσοτικών μεταβλητών είναι το εισόδημα, το βάρος, το ύψος, ο αριθμός των παιδιών μιας οικογένειας, το πλήθος των ελαττωμάτων σε ορισμένο προϊόν. Για παράδειγμα από το παραπάνω ερωτηματολόγιο οι μεταβλητές 12, 24 και 25 είναι ποσοτικές. Η μεταβλητή 2 αποτελεί ένα πεδίο ημερομηνίας από το οποίο θα προκύψει μία ποσοτική μεταβλητή που θα περιγράφει την ηλικία του ερωτώμενου με τη βοήθεια μίας αφαίρεσης της ημερομηνίας γέννησης από την ημερομηνία διεξαγωγής της μελέτης.

Οι ποσοτικές μεταβλητές ειδικότερα διακρίνονται σε δύο κατηγορίες:

Συνεχείς είναι οι μεταβλητές οι οποίες μπορούν να πάρουν οποιαδήποτε τιμή μέσα σε ένα συνεχές διάστημα ενώ **ασυνεχείς** ή **διακριτές** είναι οι μεταβλητές οι οποίες μπορούν να πάρουν μόνον διακεκριμένες τιμές. Παραδείγματα συνεχών μεταβλητών είναι το βάρος, το ύψος, η πίεση, η αντοχή ενώ παραδείγματα διακριτών μεταβλητών είναι ο αριθμός των παιδιών μιας οικογένειας, οι ζητούμενες μονάδες προϊόντος, ο αριθμός των ελαττωμάτων ενός προϊόντος.

Στην πράξη, οι πραγματικές μετρήσεις όλων των μεταβλητών είναι ασυνεχείς λόγω της περιορισμένης ακρίβειας των οργάνων μέτρησης. Έτσι π.χ. ενώ η μεταβλητή «βάρος» είναι θεωρητικά συνεχής στην πράξη και επειδή συνήθως μετριέται με ακρίβεια γραμμαρίου είναι διακριτή. Εξάλλου, ενώ θεωρητικά η μεταβλητή «εισόδημα» είναι διακριτή αφού ορίζεται για λεπτά μόνον, μπορεί να θεωρηθεί ως συνεχής αφού το πλήθος των δυνατών τιμών της είναι εξαιρετικά μεγάλο. Έτσι, για λόγους πρακτικούς, θεωρούνται ως συνεχείς οι μεταβλητές οι οποίες μπορούν να πάρουν πολλές τιμές και ως διακριτές αυτές που μπορούν να πάρουν λίγες, σχετικά, τιμές. Έτσι η μεταβλητή 25 του ερωτηματολογίου (μηνιαίο εισόδημα) καθώς και η μεταβλητή της ηλικίας που θα προκύψει από τη μεταβλητή 2 θεωρούνται συνεχείς, ενώ οι μεταβλητές 12 (μήνες ανεργίας τον τελευταίο χρόνο) και 24 (πλήθος ατόμων που κατοικούν μόνιμα στο σπίτι) θεωρούνται διακριτές αφού μπορούν να πάρουν πεπερασμένο πλήθος τιμών.

Ανάλογα με την κλίμακα μέτρησης η οποία επιλέγεται στη συλλογή δεδομένων από ορισμένη μεταβλητή προκύπτουν διαφορετικά δεδομένα. Έτσι π.χ. όταν συλλέγουμε δεδομένα για την ποσοτική μεταβλητή «μηνιαίο εισόδημα» μπορεί να επιλέξουμε να προσδιορίσουμε το ακριβές δραχμικό εισόδημα ή απλώς να το κατατάξουμε με μία από τις ακόλουθες κατηγορίες: «χαμηλό», «μέτριο», «υψηλό», «πολύ υψηλό» με σαφή προσδιορισμό της κάθε κατηγορίας. Αυτό συνέβη για τη μεταβλητή 15 του ερωτηματολογίου (διάστημα ανεργίας) που ενώ θα μπορούσε να είναι ποσοτική, εμφανίζεται μετά την ομαδοποίηση των τιμών της ως μία ποιοτική μεταβλητή με

δυνατότητα τριών μόνο διαφορετικών απαντήσεων. Είναι προφανές ότι η στατιστική ανάλυση θα είναι διαφορετική στα δύο είδη δεδομένων.

Γενικά, ανάλογα με την κλίμακα μέτρησης τα δεδομένα διακρίνονται σε **κατηγορικά** δεδομένα όταν αυτά είναι απλώς επίπεδα ή κατηγορίες και σε **μετρήσεις**.

Τα κατηγορικά δεδομένα διακρίνονται ακόμη σε:

Ονομαστικά, όταν οι παρατηρήσεις είναι κατηγορίες που η σειρά τους δεν έχει καμιά σημασία. Τα ονομαστικά δεδομένα προκύπτουν από την παρατήρηση ποιοτικών μεταβλητών. Παραδείγματα ονομαστικών δεδομένων είναι όσα προκύπτουν από την παρατήρηση της φυλής (λευκή, μαύρη, ασιατική, μικτή κτλ.), του χρώματος των μαλλιών (ξανθό, καστανό, κόκκινο, μαύρο, λευκό κτλ.), του μέσου μεταφοράς από και προς την εργασία (Ι.Χ., τραίνο, λεωφορείο, πόδια κτλ.). Στα ονομαστικά δεδομένα υπολογίζονται ποσοστά και η περαιτέρω ανάλυση βασίζεται στα ποσοστά αυτά. Από το ερωτηματολόγιο ονομαστικά δεδομένα είναι αυτά των μεταβλητών 1, 4-11, 13-14, 16-23.

Διατακτικά είναι εκείνα τα κατηγορικά δεδομένα στα οποία η διάταξη έχει σημασία. Παραδείγματα διατακτικών δεδομένων είναι οι παρατηρήσεις της σοβαρότητας μιας ασθένειας (καμία, ήπια, μέτρια, σοβαρή κτλ.) της γνώμης ως προς κάποιο μέτρο (διαφωνώ πλήρως, διαφωνώ σε κάποια σημεία, συμφωνώ, συμφωνώ πλήρως) της συχνότητας συμμετοχής σε ορισμένη δραστηριότητα (ποτέ, σπάνια, κάπου-κάπου, συχνά, πολύ συχνά). Από το ερωτηματολόγιο διατακτικά δεδομένα είναι αυτά των μεταβλητών 3, 15, 26 και 27.

Τα διατακτικά δεδομένα συχνά προκύπτουν όταν κατηγοριοποιούνται μεταβλητές, οι οποίες είναι εγγενώς συνεχείς όπως η ηλικία, η εκπαίδευση, το επίπεδο του εισοδήματος. Θεωρούνται ποσοτικά δεδομένα, αφού οι τιμές τους είναι συγκρίσιμες, μπορούμε δηλαδή, να προσδιορίσουμε ποια ανήκει σε ανώτερο μέγεθος ορισμένου χαρακτηριστικού, παρόλο που οι απόλυτες διαφορές ανάμεσα στις κατηγορίες δεν έχουν νόημα ή δεν είναι γνωστές. Έτσι π.χ. ενώ αντιλαμβανόμαστε ότι το «ανώτερο» επίπεδο εισοδήματος είναι μεγαλύτερο από το «μεσαίο» δεν μπορούμε να συγκρίνουμε τις διαφορές του «μεσαίου» από το «κατώτερο» και από το «ανώτερο».

Τα δεδομένα τα οποία είναι **μετρήσεις** προκύπτουν από την παρατήρηση κάποιας ποσοτικής μεταβλητής και επομένως είναι αριθμοί. Ανάλογα με το αν η μεταβλητή είναι συνεχής ή διακριτή τα δεδομένα διακρίνονται αντίστοιχα σε συνεχή και διακριτά.

Παραδείγματα διακριτών δεδομένων είναι ο αριθμός των οδικών ατυχημάτων στη διάρκεια ενός 24ώρου, ο αριθμός των παιδιών ανά οικογένεια, ο αριθμός των φοιτητών που μπαίνουν μετά τον διδάσκοντα στο μάθημα. Γενικά τα διακριτά δεδομένα προκύπτουν από απαρίθμηση.

Παραδείγματα συνεχών δεδομένων είναι ο χρόνος που χρειάζεται ένα ταξί για να κάνει ορισμένη διαδρομή, το επίπεδο χοληστερίνης, η ηλικία. Γενικά τα συνεχή δεδομένα προκύπτουν από μια διαδικασία μέτρησης η οποία μπορεί να είναι σχετικά απλή, όπως η μέτρηση της ηλικίας ή να απαιτεί ειδικές γνώσεις και εργαλεία όπως η μέτρηση της χοληστερίνης.

Η **κωδικοποίηση** των απαντήσεων των ερωτηματολογίων έχει σκοπό να μας διευκολύνει στη διαλογή των στοιχείων που περιέχονται σε αυτά, καθώς και στην κατάταξη των αποτελεσμάτων σε μορφή πινάκων. Για τις ποιοτικές μεταβλητές πρέπει στο ερωτηματολόγιο να αναγράφονται όλες οι πιθανές απαντήσεις. Σε κάθε μια από αυτές αντιστοιχούμε έναν αριθμό ώστε να είναι εύκολη η μεταφορά του κάθε

ερωτηματολογίου σε μια σειρά της συγκεντρωτικής κατάστασης. Αν π.χ. θέλουμε να κωδικοποιήσουμε τη μεταβλητή επίπεδο μόρφωσης, τότε μια κωδικοποίηση αυτής της μεταβλητής θα μπορούσε να είναι όπως αυτή του ερωτηματολογίου:

Επίπεδο Εκπαίδευσης	Κωδικός
Δεν πήγα σχολείο	1
Μερικές τάξεις Δημοτικού	2
Απόφοιτος Δημοτικού	3
Απόφοιτος Γυμνασίου	4
Απόφοιτος Λυκείου	5
Απόφοιτος ΙΕΚ	6
Απόφοιτος ΤΕΙ	7
Απόφοιτος ΑΕΙ	8
Κάτοχος Μεταπτυχιακού Διπλώματος	9
Κάτοχος Διδακτορικού Διπλώματος	10

Αν μία ερώτηση αναφέρεται σε ποσοτικό χαρακτηριστικό, δηλαδή έχουμε ποσοτική μεταβλητή, τότε μπορούμε να τη χωρίσουμε σε κλάσεις. Ο χωρισμός αυτός μπορεί να γίνει με διάφορους τρόπους:

- Οι n κλάσεις έχουν το ίδιο εύρος.
- Οι n κλάσεις έχουν τον ίδιο αριθμό αντικειμένων.
- Η κάθε μία από τις n κλάσεις περιλαμβάνει ένα συγκεκριμένο ποσοστό αντικειμένων.
- Η κάθε μία από τις n κλάσεις έχει όρια που εμείς ορίζουμε.

Έστω ότι θέλουμε να κωδικοποιήσουμε την ποσοτική μεταβλητή "ηλικία" και ο χωρισμός σε κλάσεις είναι: μέχρι 17 ετών, από 18 έως 24, από 25 έως 30, από 31 έως 40, από 40 έως 60 και μεγαλύτερος των 60 χρόνων. Στην περίπτωση αυτή σε κάθε κλάση αντιστοιχούμε και έναν αριθμό 1,2,3,4,5,6.

Ηλικία	Κωδικός
Μέχρι 17	1
Από 18 έως 24	2
Από 25 έως 30	3
Από 31 έως 40	4
Από 41 έως 60	5
Μεγαλύτερος από 60	6

Αφού αρχικά προδιαγράψουμε την κωδικοποίηση της κάθε μεταβλητής μεταφέρουμε τις απαντήσεις των ερωτηματολογίων σε καταστάσεις. Σε αυτές τις καταστάσεις κάθε ερωτηματολόγιο αντιστοιχεί και μεταφέρεται σε μία σειρά χωρισμένη σε στήλες. Η κάθε στήλη αντιστοιχεί σε μία μεταβλητή. Στην κατάσταση όπου θα μεταφέρουμε τις απαντήσεις θα πρέπει να υπάρχουν τόσες γραμμές όσα είναι και τα ερωτηματολόγια και τόσες στήλες όσες είναι και οι ερωτήσεις και μια στήλη στην αρχή όπου γράφουμε τον αριθμό του ερωτηματολογίου, για να μπορούμε να κάνουμε την αντιπαραβολή.

Συνήθως στην κωδικοποίηση των ποιοτικών μεταβλητών χρησιμοποιούμε τον αριθμό

ένα και όχι τον αριθμό μηδέν ως κωδικό της πρώτης κατηγορίας. Υπάρχει όμως μία ειδική κατηγορία ποιοτικών μεταβλητών τις οποίες ονομάζουμε **δίτιμες**. Οι μεταβλητές αυτές προκύπτουν όταν η ερώτηση αναφέρεται στην διερεύνηση της ύπαρξης ή όχι μίας ιδιότητας και η απάντηση στην ερώτηση μπορεί να γίνει είτε με ένα «Ναι» είτε με ένα «Όχι». Για παράδειγμα η ερώτηση 21 του ερωτηματολογίου (Έχετε Ελληνική υπηκοότητα;). Στις περιπτώσεις αυτές είναι προτιμότερο η κωδικοποίηση να γίνει αντιστοιχίζοντας την απάντηση «Ναι» στον αριθμό 1 και την απάντηση «Όχι» στον αριθμό 0.

Τέλος, πολλές φορές είναι δυνατό ο ερωτώμενος να απαντά σε δύο ή και περισσότερες κλάσεις μιας ερώτησης, όπως για παράδειγμα στην ερώτηση 22 του ερωτηματολογίου (Αν δεν έχετε ελληνική υπηκοότητα έχετε: Άδεια παραμονής, Πράσινη Κάρτα), όπου κάποιος μπορεί να έχει και τα δύο, ένα από τα δύο ή κανένα από αυτά.

Η κωδικοποίηση είχε γίνει:

1: Άδεια παραμονής

2: Πράσινη Κάρτα

Αν κρατήσουμε μια στήλη θα βρεθούμε σε αδυναμία, όταν κάποιος έχει και πράσινη κάρτα και άδεια παραμονής. Το πρόβλημα αυτό λύνεται, αν αντιστοιχίσουμε μια στήλη για κάθε μία από τις άδειες και βάλουμε τον αριθμό 1 αν την έχει και τον αριθμό 0 αν δεν την έχει. Μετατρέπουμε, δηλαδή, την κάθε δυνατότητα απάντησης σε μία δίτιμη λογική ψευδομεταβλητή (NAI-OXI ή 1-0).

Για παράδειγμα, έστω ότι το ερωτηματολόγιο 1 προέρχεται από άτομο που δήλωσε ότι έχει μόνο άδεια παραμονής, το 2 από άτομο που έχει μόνο πράσινη κάρτα και το 3 από άτομο που έχει και τα δύο. Η κατάσταση αυτών θα είναι:

α/α	Άδεια παραμονής	Πράσινη Κάρτα
1	1	0
2	0	1
3	1	1