

Μάθημα 3

§ 2.2 Μεταβλητές

Στα παραδείγματα που έχουμε δει μέχρι τώρα, οι πληροφορίες σχετικά με το ύψος, την ηλικία, τις εκπαιδευτικές συνήθειες κ.λπ. των ερωτώμενων αναφέρονται σε ορισμένα χαρακτηριστικά ή ιδιότητες που παρουσιάζουν τα άτομα του υπό εξέταση πληθυσμού τα οποία στη Στατιστική καλούνται πληθυσμιακά χαρακτηριστικά (population characteristics). Προφανώς, χαρακτηριστικά και ιδιότητες παρατηρούνται όχι μόνο στους ανθρώπινους αλλά και στους μη-ανθρώπινους πληθυσμούς. Π.χ. η θερμοκρασία, το ύψος των βροχοπτώσεων μιας περιοχής ή το μέγεθος και το χρώμα ορισμένου κατηγορίας φυτών, αποτελούν παραδείγματα χαρακτηριστικών και ιδιοτήτων μη ανθρώπινων πληθυσμών.

Ας δούμε ένα παράδειγμα. Στην έρευνα με σκοπό την μελέτη της εκπαιδευτικής συνήθειας του Ελληνικού πληθυσμού ποιάστηκαν ερωτηματολόγια προς συμπλήρωση. Επιλέχουμε τυχαία ένα από τα ερωτηματολόγια που έχουμε συλλέξει. Από τα στοιχεία που αναγράφονται στο συμπληρωμένο ερωτηματολόγιο αντιλαμβανόμαστε ότι πρόκειται για άνδρα, 27 ετών, 72 κιλών, άγαμο και εκπαιδευτή. Η πληροφορία αυτή μας λέει ότι η ερευνώμενη στατιστική μονάδα είναι το άτομο, τα χαρακτηριστικά που συλλέξαμε και θέλουμε

στη συνέχεια να μετρήσουμε είναι κατά σειρά, το -15-
φύλο, η ηλικία, το βάρος, η οικογενειακή κατάσταση
και η μαθησιακή συνήθεια.

Οι αριθμικές περιγραφές και μετρήσεις των χαρακτη-
ριστικών και ιδιοτήτων των στατιστικών μονάδων
καλούνται τιμές των δεδομένων (data values).

Προφανώς οι τιμές των εξεταζόμενων χαρακτηριστι-
κών διαφοροποιούνται, μεταβαλλόμενες μεταξύ των ατό-
μων ή γενικά μεταξύ των στατιστικών μονάδων. Κάθε πη-
θυσιακό χαρακτηριστικό καλείται μεταβλητή (variable).

Οι μετρήσεις μιας μεταβλητής ενός συνόλου στατιστι-
κών μονάδων είναι πραγματικοί αριθμοί (π.χ. βάρος, 70 κιλά)
ή συμβολικές εκφράσεις (π.χ. φύλο: άνδρας) και καλού-
νται τιμές της μεταβλητής (variable values). Οι μετα-
βλητές συμβολίζονται με κεφαλαίους λατινικούς χαρα-
κτήρες.

Για παράδειγμα:

X = ηλικία Y = εισόδημα Z = φύλο W = επάγγελμα

Η πρώτη, δεύτερη, κ.ο.κ. ερευνόμενη στατιστική μονάδα συμβο-
λίζεται με i . Ο αύξων αυτός αριθμός κινείται μεταξύ
του 1 και του N όταν αναφερόμαστε σε στοιχεία πη-
θυσικού μετέθους $N, (i=1, 2, \dots, N)$ και μεταξύ του 1
και του n όταν πρόκειται για δεδομένα ενός δεί-
γματος μετέθους $n, (i=1, 2, \dots, n)$. Η τιμή της

μεταβλητής X για το i -οστό ερευνώμενο άτομο ή γενικά την i -στη στατιστική μονάδα, θα συμβολίζεται με x_i . Για παράδειγμα, οι τιμές του βάρους (X) και του φύλου (Y) πέντε ατόμων ($i=1,2,3,4,5$) είναι:

άτομα	βάρος (μίλα)	φύλο
i	x_i	y_i
1	$x_1 = 72$	$y_1 = \text{άνδρας}$
2	$x_2 = 78$	$y_2 = \text{άνδρας}$
3	$x_3 = 65$	$y_3 = \text{γυναίκα}$
4	$x_4 = 70$	$y_4 = \text{άνδρας}$
5	$x_5 = 64$	$y_5 = \text{γυναίκα}$

§2.3 Τύποι Μεταβλητών

Η φύση των μεταβλητών καθορίζει το είδος των δεδομένων. Οι μεταβλητές διακρίνονται σε ποσοτικές (numeric, quantitative variables) και ποιοτικές (qualitative variables) ανάλογα με το αν οι τιμές τους εκφράζονται αριθμητικά ή ονομαστικά.

Οι ποσοτικές μεταβλητές διακρίνονται σε συνεχείς και σε διακριτές ανάλογα με το εάν οι τιμές είναι συνεχείς ή διακριτές αντίστοιχα.

Οι ποιοτικές μεταβλητές διακρίνονται σε κατηγορημένες (categorical) και σε διατάξιμες (ordinal).

Για παράδειγμα, οι τιμές της μεταβλητής "φύλο" κατατάσσονται με βάση τον πληθυσμό ή του δείγματος

σε δύο κατηγορίες ή υπο-οράδες, άνδρες και γυναίκες. -17-

Όπως το "επίπεδο εκπαίδευσης" κατατάσσει τους ερευνώ-
μένους σε διακριτές κατηγορίες π.χ. αυτούς που έχουν
τελειώσει πρωτοβάθμια, δευτεροβάθμια και τριτοβάθμια
εκπαίδευση. Όπως επιπλέον μας δίνει τη δυνατότητα
να διατάξουμε τις κατηγορίες αυτές σε μία σειρά από
το χαμηλότερο προς το ανώτερο εκπαιδευτικό επίπεδο
ή αντίστροφα. Οι ποιοτικές μεταβλητές που δίνουν
τη δυνατότητα στον ερευνητή να διατάξει και να
διαβαθρίσει τις κατηγορίες που προκύπτουν από
τις τιμές τους καλούνται διατάξιμες ή ταξιτικές. Οι
υπόλοιπες οι οποίες από τη φύση τους δεν παρέχουν
αυτή τη δυνατότητα διάταξης αλλά, με βάση τα
ποιοτικά χαρακτηριστικά που εκφράζουν οι τιμές
τους, επιτρέπουν απλά και μόνο τη διάκριση ορισμέ-
νων κατηγοριών ή υπο-οράδων καλούνται κατηγορητικές.
Παραδείγματα κατηγορητικών μεταβλητών είναι το
φύλο, η οικογενειακή κατάσταση, το θρήσκευμα
κλπ., και διατάξιμων το επίπεδο εκπαίδευσης,
ο βαθμός προτίμησης του καταναλωτή για κάποιο
προϊόν, το επίπεδο υγείας κλπ.

Οι μεταβλητές που λαμβάνουν την ίδια τιμή για
όλες τις στατιστικές μονάδες καλούνται σταθερές
μεταβλητές (constant variables).

Οι τιμές μιας ονομαστικής μεταβλητής μπορούν να κωδικοποιηθούν και να αντικατασταθούν από αριθμικές τιμές καθοριζόμενες από τον ερευνητή. Έτσι προκύπτει μία "νέα" αριθμητική διακριτή μεταβλητή η οποία καλείται ψευδομεταβλητή (dummy variable). Π.χ. έστω ότι έχουμε συλλέξει στοιχεία για το φύλο (μεταβλητή X) πέντε ατόμων. Θέτουμε "0" για τους άνδρες και "1" για τις γυναίκες λαμβάνουμε την μεταβλητή Y που εμφανίζεται παρακάτω.

i	x_i	y_i
1	γυναίκα	1
2	άνδρας	0
3	γυναίκα	1
4	άνδρας	0
5	άνδρας	0

Άσκηση Καθορίστε το είδος των δεδομένων-μεταβλητών

- επίπεδο προτίμησης κάποιων αναφυτιών
(ποιοτική-διάταξη)
- αποτελέσματα πανελληνίων εξετάσεων
(ποσοτική-συνεχής)
- επίδοση αθλητή δρόμου ταχύτητας
(ποσοτική-συνεχής)
- θερμοκρασίες Αστεροσκοπίου Αθηνών
(ποσοτική-διακριτή)
- βάρος ταχυδρομικού δέματος
(ποσοτική-συνεχής)
- τιμή βενζίνης ανά λίτρο (ποσοτική-συνεχής)