

Βασικές έννοιες

Σε ένα ερωτηματολόγιο έχουμε ένα σύνολο ερωτήσεων. Μπορούμε να πούμε ότι σε κάθε ερώτηση αντιστοιχεί μία μεταβλητή. Αν θεωρήσουμε μια ερώτηση, τα άτομα δίνουν κάποιες απαντήσεις στην ερώτηση αυτή. Οι απαντήσεις αυτές είναι οι τιμές της αντίστοιχης μεταβλητής. Με τον τρόπο αυτό τιμή είναι το 35 για τη μεταβλητή «ηλικία», αλλά τιμή είναι και το κορίτσι για τη μεταβλητή «φύλο».

Μπορούμε να θεωρήσουμε τα εξής είδη μεταβλητών:

1. Κατηγορικές μεταβλητές.

Μια μεταβλητή είναι κατηγορική, όταν η μεταβλητή αυτή διαχωρίζει τα άτομα σε κατηγορίες. Για παράδειγμα, η μεταβλητή «φύλο», με δύο δυνατές τιμές, αγόρι και κορίτσι, διαχωρίζει τα άτομα σε αγόρια και κορίτσια. Η μεταβλητή «πτυχίο στο οποίο αποβλέπεις», με τρεις δυνατές τιμές, βασικό πτυχίο, μεταπτυχιακό, διδακτορικό, διαχωρίζει τα άτομα σε αυτούς που αποβλέπουν στο βασικό πτυχίο, σε αυτούς που αποβλέπουν στο μεταπτυχιακό και σε αυτούς που αποβλέπουν στο διδακτορικό.

2. Ποσοτικές μεταβλητές.

Μια μεταβλητή είναι ποσοτική όταν η μεταβλητή αυτή αναφέρεται σε μετρήσεις ποσοτήτων. Για παράδειγμα, η μεταβλητή «βάρος» ή η μεταβλητή «ύψος», αλλά επίσης η μεταβλητή «ηλικία».

3. Μεταβλητές που αντιστοιχούν σε κλίμακες αξιολόγησης ή αλλιώς διπολικές κλίμακες ή κλίμακες Likert.

Οι μεταβλητές αυτές αντιστοιχούν σε ερωτήσεις στις οποίες τα άτομα καλούνται να δηλώσουν το βαθμό αποδοχής ή απόρριψης για μια σει-

ρά απόψεις, φράσεις, θέματα, πρόσωπα κ.λπ. στη βάση μιας αριθμητικής κλίμακας, η οποία μπορεί να είναι από 1 έως 5 ή από 1 έως 7 ή από 1 έως 10. Για παράδειγμα, η μεταβλητή «μ' αρέσει η στατιστική», στην οποία το άτομο καλείται να τοποθετηθεί σε μια κλίμακα από το 1 μέχρι το 7, όπου το 1 αντιστοιχεί στο «διαφωνώ έντονα», το 7 στο «συμφωνώ έντονα», το 4 στο «ούτε συμφωνώ ούτε διαφωνώ» και οι ενδιάμεσες τιμές στις ενδιάμεσες απόψεις.

Αφού συγκεντρώσουμε τα στοιχεία για το σύνολο των μεταβλητών και για το σύνολο των ατόμων, εισάγουμε τα στοιχεία αυτά στον ηλεκτρονικό υπολογιστή και δημιουργούμε τον πίνακα των δεδομένων. Στον πίνακα αυτό εφαρμόζουμε τη στατιστική επεξεργασία. Ανάλογα με το είδος των μεθόδων και των τεχνικών που χρησιμοποιούμε για την επεξεργασία των δεδομένων, μπορούμε να διακρίνουμε τη στατιστική σε τρία μέρη:

1. Την **περιγραφική στατιστική**, η οποία περιλαμβάνει απλές τεχνικές για την περιγραφή των δεδομένων.
2. Την **επαγωγική στατιστική**, με την οποία επιχειρούμε να γενικεύσουμε στον πληθυσμό αποτελέσματα που παίρνουμε από ένα δείγμα.
3. Την **πολυδιάστατη στατιστική**, με την οποία επιχειρούμε να διερευνήσουμε τα δεδομένα και να βρούμε τους βασικούς παράγοντες ή τις κύριες ομάδες ατόμων στη βάση ενός συνόλου μεταβλητών.

Η ομάδα από την οποία έχουν ληφθεί τα δεδομένα λέγεται δείγμα. Το ευρύτερο σύνολο ομοειδών περιπτώσεων από το οποίο λαμβάνεται το δείγμα λέγεται πληθυσμός. Για παράδειγμα, οι 500 μαθητές μιας έρευνας είναι το δείγμα. Το σύνολο των μαθητών της χώρας είναι ο αντίστοιχος πληθυσμός. Ο αριθμός των ατόμων του δείγματος ονομάζεται μέγεθος του δείγματος.

Περιγραφική στατιστική

Η επεξεργασία των δεδομένων ξεκινά με τις τεχνικές της περιγραφικής στατιστικής. Με την περιγραφική στατιστική μπορούμε να εξετάσουμε τις μεταβλητές ανά μία ή ανά δύο. Η εξέταση των μεταβλητών ανά μία περιλαμβάνει:

1. Την **κατανομή συχνότητων μιας μεταβλητής**, η οποία δείχνει τη συχνότητα εμφάνισης των διάφορων τιμών της μεταβλητής.

2. Τους **δείκτες κεντρικής τάσης μιας μεταβλητής**, με τους οποίους επιχειρούμε να προσδιορίσουμε την κεντρική τιμή της μεταβλητής.
3. Τους **δείκτες διασποράς μιας μεταβλητής**, με τους οποίους επιχειρούμε να προσδιορίσουμε ένα μέτρο τού πόσο μεταβάλλονται από την κεντρική τιμή οι τιμές της μεταβλητής.

Η εξέταση των μεταβλητών ανά δύο περιλαμβάνει:

1. Τους **πίνακες διασταύρωσης**, με τους οποίους επιχειρούμε να μελετήσουμε τη συμμεταβολή δύο κατηγορικών μεταβλητών.
2. Τη **γραμμική συσχέτιση**, με την οποία επιχειρούμε να μελετήσουμε τη συμμεταβολή δύο ποσοτικών μεταβλητών ή δύο μεταβλητών που αντιστοιχούν σε κλίμακες αξιολόγησης.
3. Τη **γραμμική παλινδρόμηση**, με την οποία επιχειρούμε να διατυπώσουμε με μαθηματικούς όρους τη σχέση δύο ποσοτικών μεταβλητών.

Τέλος, θα αναφερθούμε στην **κανονική κατανομή**, κατανομή η οποία έχει σημαντικό ρόλο ιδιαίτερα στην επαγωγική στατιστική.



ΚΑΤΑΝΟΜΕΣ ΣΥΧΝΟΤΗΤΩΝ

Η κατανομή συχνοτήτων μιας μεταβλητής δείχνει τη συχνότητα εμφάνισης των τιμών της μεταβλητής. Τις κατανομές συχνοτήτων τις χρησιμοποιούμε για να περιγράψουμε τις κατηγορικές μεταβλητές, τις ποσοτικές μεταβλητές και τις μεταβλητές που αντιστοιχούν στις κλίμακες αξιολόγησης.

Κατανομές κατηγορικών μεταβλητών

Οι κατανομές συχνοτήτων των κατηγορικών μεταβλητών κατασκευάζονται εύκολα. Για κάθε τιμή της κατηγορικής μεταβλητής υπολογίζουμε τη συχνότητά της στο σύνολο του δείγματος και στη συνέχεια υπολογίζουμε το ποσοστό της τιμής ως προς το σύνολο του δείγματος. Μια κατανομή συχνοτήτων μπορούμε να την απεικονίσουμε με ένα διάγραμμα το οποίο ονομάζεται ραβδόγραμμα. Το ραβδόγραμμα περιλαμβάνει για κάθε τιμή της μεταβλητής μια στήλη. Τα μήκη των στηλών του ραβδογράμματος είναι ανάλογα με τις συχνότητες ή τα ποσοστά των διάφορων τιμών της μεταβλητής.

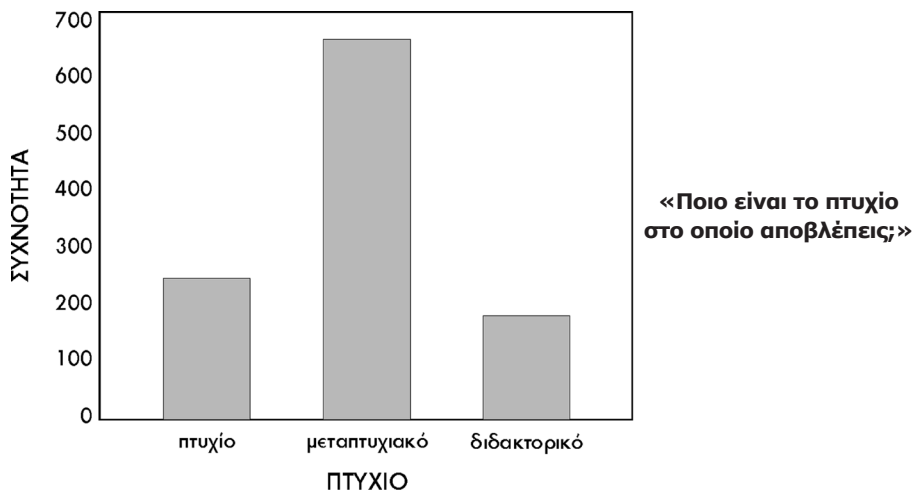
Για παράδειγμα, για τη μεταβλητή «ποιο είναι το πτυχίο στο οποίο αποβλέπεις» έχουμε 241 άτομα ή 22.5% του δείγματος που επιλέγουν το βασικό πτυχίο, 650 άτομα ή 60.7% του δείγματος που επιλέγουν το μεταπτυχιακό και 179 άτομα ή 16.7% του δείγματος που επιλέγουν το διδακτορικό.

Πίνακας 1 (κατανομή συχνοτήτων)

«Ποιο είναι το πτυχίο στο οποίο αποβλέπεις;»	ΣΥΧΝΟΤΗΤΑ	ΠΟΣΟΣΤΟ
Πτυχίο	241	22.5
Μεταπτυχιακό	650	60.7
Διδακτορικό	179	16.7
<i>ΣΥΝΟΛΟ</i>	<i>1070</i>	<i>100.0</i>

Στο ραβδόγραμμα που ακολουθεί μπορούμε να δούμε ότι η μεγαλύτερη στήλη είναι αυτή που αντιστοιχεί στο μεταπτυχιακό και ακολουθούν το βασικό πτυχίο και το διδακτορικό.

Διάγραμμα 1 (ραβδόγραμμα του πίνακα 1)



Κατανομές ποσοτικών μεταβλητών

Όταν έχουμε ποσοτικές μεταβλητές κάνουμε ό,τι κάνουμε και στην περίπτωση των κατηγορικών μεταβλητών. Για κάθε μία τιμή, δηλαδή, υπολογίζουμε τη συχνότητα, το ποσοστό και τέλος κατασκευάζουμε την αντίστοιχη γραφική παράσταση, η οποία στην περίπτωση αυτή ονομάζεται ιστόγραμμα.

Επειδή στις ποσοτικές μεταβλητές η ποικιλία των τιμών είναι μεγάλη, συνήθως, αντί για τις τιμές χρησιμοποιούμε διαστήματα τιμών. Για παράδειγμα, για τη μεταβλητή «ηλικία» μπορούμε να δημιουργήσουμε τα διαστήματα τιμών 18-24, 25-34, 35-44, 45-54, 55-64 και πάνω από 65. Έτσι, για παράδειγμα, το διάστημα 18-24 περιλαμβάνει τα άτομα ηλικίας από 18 ως 24 κ.λπ.

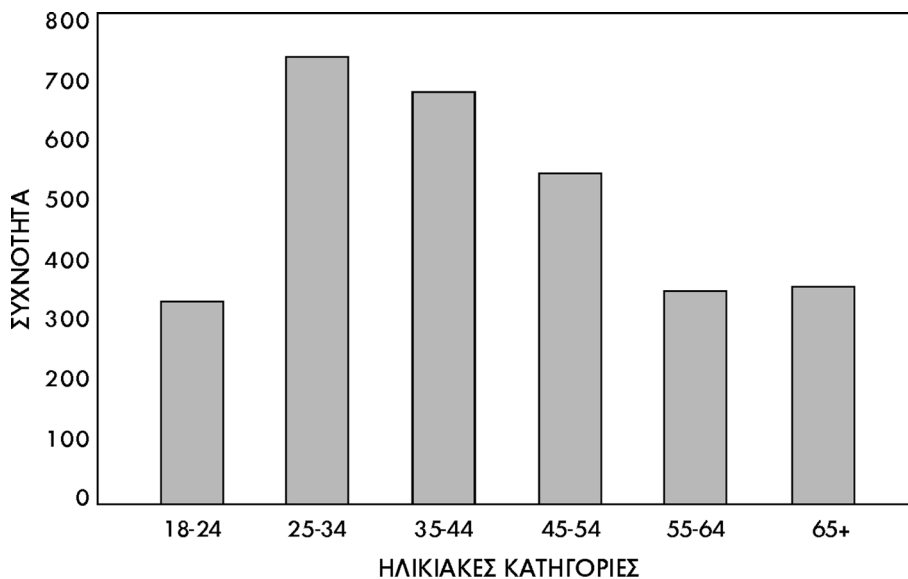
Με τον τρόπο αυτό υπολογίζουμε τη συχνότητα για κάθε διάστημα, το ποσοστό για κάθε διάστημα και κατασκευάζουμε το ιστόγραμμα με βάση τα διαστήματα.

Στον επόμενο πίνακα μπορούμε να δούμε ότι το μεγαλύτερο ποσοστό του δείγματος το έχει η ηλικιακή κατηγορία από 25 έως 34 (24.47%), ακολουθεί η ηλικιακή κατηγορία από 35 έως 44 (22.7%) κ.λπ.

Πίνακας 2 (κατανομή συχνοτήτων)

Ηλικιακές κατηγορίες	ΣΥΧΝΟΤΗΤΑ	ΠΟΣΟΣΤΟ
18-24	336	11.20
25-34	734	24.47
35-44	681	22.70
45-54	540	18.00
55-64	351	11.70
65+	358	11.93
ΣΥΝΟΛΟ	3000	100.0

Διάγραμμα 2 (ιστόγραμμα του πίνακα 2)



Κατανομές για μεταβλητές που αντιστοιχούν σε κλίμακες αξιολόγησης

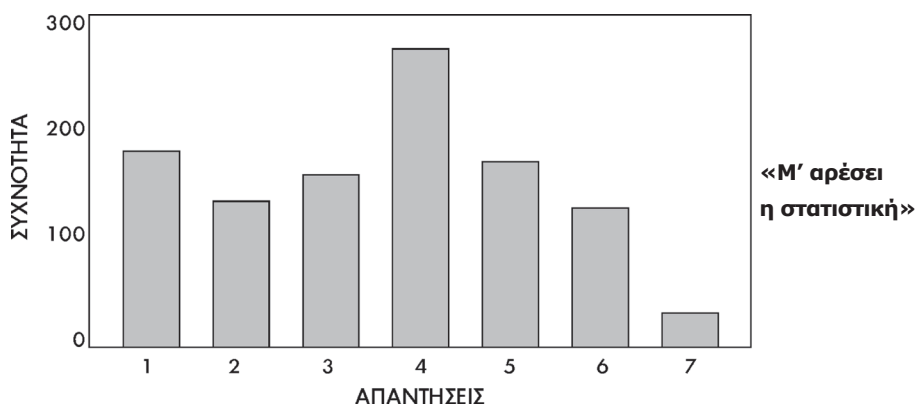
Στην περίπτωση των μεταβλητών που αντιστοιχούν σε κλίμακες αξιολόγησης κάνουμε ό,τι κάνουμε και στην περίπτωση των κατηγορικών μεταβλητών. Για κάθε τιμή, δηλαδή, υπολογίζουμε τη συχνότητα, το ποσοστό και τέλος κατασκευάζουμε την αντίστοιχη γραφική παράσταση. Στον πίνακα που ακολουθεί μπορούμε να δούμε ότι για τη μεταβλητή «μ' αρέσει η στατιστική» την τιμή 1 (διαφωνώ έντονα) την επιλέγει το 17.2% του δείγματος, την τιμή 7 (συμφωνώ έντονα) την επιλέγει το 3.1% του δείγματος, την τιμή 4 (ούτε συμφωνώ ούτε διαφωνώ) την επιλέγει το 25.2% του δείγματος κ.λπ.

Πίνακας 3 (κατανομή συχνοτήτων)

«Μ' αρέσει η στατιστική»	ΣΥΧΝΟΤΗΤΑ	ΠΟΣΟΣΤΟ
1	184	17.2
2	132	12.3
3	155	14.5
4	270	25.2
5	169	15.8
6	128	12.0
7	33	3.1
ΣΥΝΟΛΟ	1071	100.0

Στο ραβδόγραμμα που ακολουθεί παρατηρούμε ότι τα περισσότερα άτομα τα συγκεντρώνει η τιμή 4 (ούτε συμφωνώ ούτε διαφωνώ), αλλά συγκριτικά οι τιμές 1 έως 3 (διαφωνία με την πρόταση) συγκεντρώνουν περισσότερα άτομα από τις τιμές 5 έως 7 (συμφωνία με την πρόταση).

Διάγραμμα 3 (ραβδόγραμμα του πίνακα 3)



ΣΥΝΟΨΗ

Κατανομές συχνοτήτων

Η κατανομή συχνοτήτων μιας μεταβλητής δείχνει τη συχνότητα εμφάνισης των τιμών της μεταβλητής. Τις κατανομές συχνοτήτων τις χρησιμοποιούμε για να περιγράψουμε τις κατηγορικές μεταβλητές, τις ποσοτικές μεταβλητές και τις μεταβλητές που αντιστοιχούν στις κλίμακες αξιολόγησης. Σε όλες τις περιπτώσεις, για κάθε τιμή της μεταβλητής υπολογίζουμε τη συχνότητα, το ποσοστό και τέλος κατασκευάζουμε την αντίστοιχη γραφική απεικόνιση, το ραβδόγραμμα ή το ιστόγραμμα ανάλογα με το είδος της μεταβλητής. Σε ένα ραβδόγραμμα ή σε ένα ιστόγραμμα έχουμε με μια ματιά την εικόνα της μεταβλητής.



Εκτός από τις κατανομές συχνότητας και τη γραφική τους απεικόνιση, είναι δυνατό να περιγράψουμε μια μεταβλητή με κάποιους αριθμητικούς δείκτες. Οι πλέον συνηθισμένοι δείκτες είναι οι δείκτες κεντρικής τάσης και οι δείκτες διασποράς. Με τους δείκτες κεντρικής τάσης επιχειρούμε να προσδιορίσουμε το κέντρο της κατανομής της μεταβλητής, δηλαδή το σημείο στο οποίο τείνουν να συγκεντρωθούν οι τιμές της μεταβλητής, όταν δημιουργήσουμε την κατανομή της μεταβλητής. Με τους δείκτες διασποράς επιχειρούμε να προσδιορίσουμε ένα μέτρο τού πόσο οι τιμές της μεταβλητής μεταβάλλονται γύρω από το κέντρο της κατανομής της μεταβλητής.

Οι δείκτες κεντρικής τάσης είναι η δεσπόζουσα τιμή, η διάμεσος και η μέση τιμή. Η μέση τιμή και η διάμεσος χρησιμοποιούνται για ποσοτικές μεταβλητές και για μεταβλητές που αντιστοιχούν σε κλίμακες αξιολόγησης. Η δεσπόζουσα τιμή χρησιμοποιείται κυρίως για κατηγορικές μεταβλητές και μεταβλητές που αντιστοιχούν σε κλίμακες αξιολόγησης, αλλά μπορεί να υπολογιστεί και στην περίπτωση ποσοτικών μεταβλητών.

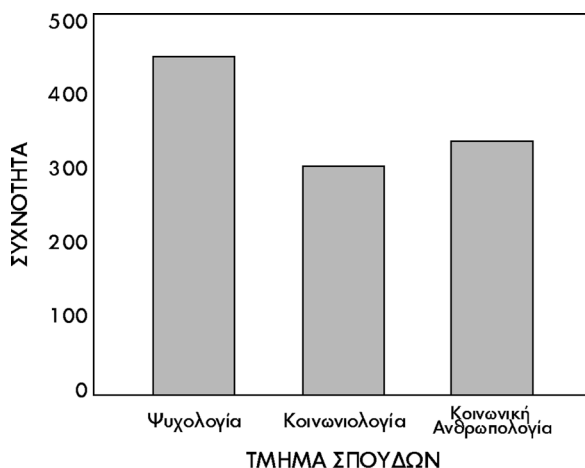
Δεσπόζουσα τιμή

Ο πρώτος και απλούστερος δείκτης κεντρικής τάσης με τον οποίο θα ασχοληθούμε είναι η δεσπόζουσα τιμή. Δεσπόζουσα τιμή μιας μεταβλητής είναι η τιμή της μεταβλητής με τη μεγαλύτερη συχνότητα. Μπορούμε να συμβολίσουμε τη δεσπόζουσα τιμή με δ .

Προϋπόθεση για να βρούμε τη δεσπόζουσα τιμή είναι να υπολογίσουμε πρώτα τη συχνότητα κάθε τιμής. Γι' αυτό μπορούμε να κατασκευάσουμε πρώτα την κατανομή της μεταβλητής και στη συνέχεια να προσδιορίσουμε την τιμή με τη μεγαλύτερη συχνότητα.

Στο επόμενο παράδειγμα, όπου παρουσιάζουμε το ραβδόγραμμα της μεταβλητής «τμήμα σπουδών» με τρεις δυνατές τιμές, Ψυχολογία, Κοινωνιολογία, Κοινωνική Ανθρωπολογία, η δεσπόζουσα τιμή είναι η Ψυχολογία, γιατί έχει τη μεγαλύτερη συχνότητα.

Διάγραμμα 4 (ραβδόγραμμα)



Διάμεσος

Η διάμεσος είναι μία τιμή, είτε από το σύνολο των τιμών της μεταβλητής είτε εκτός του συνόλου των τιμών της μεταβλητής, η οποία χωρίζει το σύνολο των τιμών της μεταβλητής σε δύο ίσα μέρη: Οι μισές τιμές είναι μεγαλύτερες της διαμέσου και οι άλλες μισές είναι μικρότερες της διαμέσου. Όταν ο αριθμός των τιμών είναι περιττός αριθμός, η διάμεσος είναι μία από τις τιμές της μεταβλητής. Όταν ο αριθμός των τιμών είναι άρτιος αριθμός, η διάμεσος είναι μία τιμή εκτός του συνόλου των τιμών της μεταβλητής. Μπορούμε να συμβολίσουμε τη διάμεσο με M .

Παράδειγμα: Α) Περίπτωση περιττού αριθμού τιμών.

Έχουμε τις τιμές: 45 64 67 72 82 82 89

Η τιμή η οποία χωρίζει το σύνολο των τιμών σε δύο ίσα μέρη είναι η τιμή 72. Άρα η διάμεσος είναι $M = 72$.

Παράδειγμα: Β) Περίπτωση άρτιου αριθμού τιμών.

Έχουμε τις τιμές: 46 48 58 72 74 84 91 116

Η τιμή η οποία χωρίζει το σύνολο των τιμών σε δύο ίσα μέρη βρίσκεται εκτός του συνόλου των τιμών αυτών, μεταξύ των τιμών 72 και 74. Στην περίπτωση αυτή ορίζουμε ως διάμεσο το $\frac{1}{2}$ του αθροίσματος των τιμών 72 και 74. Άρα η διάμεσος είναι $M = \frac{1}{2} (72+74) = 73$.