

ΣΥΝΟΨΙΣΗ ΚΑΙ ΠΑΡΟΥΣΙΑΣΗ ΔΕΔΟΜΕΝΩΝ

Μολονότι κατά τη διεξαγωγή της επιστημονικής έρευνας μελετούμε σχεδόν πάντοτε μόνο δείγματα των πληθυσμών που μας ενδιαφέρουν, τα αριθμητικά δεδομένα που συγκεντρώνονται είναι συνήθως πολυάριθμα και δεν είναι δυνατό ούτε να ερμηνευτούν ούτε να παρουσιαστούν, προτού συνοψιστούν με τον κατάλληλο τρόπο. Για το σκοπό αυτό χρησιμοποιούνται συνηθέστερα οι κατανομές συχνότητας, τα διαγράμματα και τα μέτρα κεντρικής τάσεως και διασποράς.

Τόσο οι κατανομές συχνότητας και τα διαγράμματα, όσο και τα μέτρα κεντρικής τάσεως και διασποράς, «περιγράφουν» τα αποτελέσματα της έρευνας. Γι' αυτό, η συνόψιση και παρουσίαση των δεδομένων λέγεται και **Περιγραφική Στατιστική**, σε αντιδιαστολή με την **Επαγωγική Στατιστική**, η οποία βγάζει συμπεράσματα από το δείγμα για τον πληθυσμό και η οποία αποτελεί το κύριο μέρος της Στατιστικής.

2.1. Κατανομές συχνότητας

Σ' ένα δείγμα τιμών μιας μεταβλητής η κάθε τιμή είναι δυνατό να απαντάται περισσότερες από μία φορές. Ο αριθμός των επαναλήψεων της κάθε τιμής ονομάζεται **απόλυτη συχνότητα** της τιμής. Αν καταχωρήσουμε σε μια στήλη τις τιμές της μεταβλητής και σε μια δεύτερη στήλη τη συχνότητα της κάθε τιμής, σχηματίζουμε έναν πίνακα **κατανομής συχνότητας**. Οι λεπτομέρειες της κατασκευής του πίνακα διαφέρουν ανάλογα με τον αριθμό των τιμών που παίρνει η μεταβλητή.

2.1.1. Αριθμός τιμών μεταβλητής μικρός

Όταν ο αριθμός των τιμών που παίρνει η μεταβλητή είναι περιορισμένος, όπως συχνά συμβαίνει στην περίπτωση των ασυνεχών μεταβλητών, για να κατασκευάσουμε έναν πίνακα κατανομής συχνότητας, γράφουμε σε μια στήλη όλες τις τιμές της μεταβλητής και απέναντι από κάθε τιμή γράφουμε τη συχνότητά της.

Ο πίνακας 2.1 δίνει την κατανομή συχνότητας των βαθμών που πήραν

ΠΙΝΑΚΑΣ 2.1

Κατανομή των βαθμών 140 φοιτητών στο μάθημα της Στατιστικής

Βαθμός	Αριθμός φοιτητών	Σχετική συχνότητα	Σχετική συχνότητα %
3	8	0,057	5,7
4	15	0,107	10,7
5	8	0,057	5,7
6	44	0,314	31,4
7	32	0,229	22,9
8	25	0,179	17,9
9	8	0,057	5,7
Σύνολο	140	1,000	100,0

στο μάθημα της Στατιστικής 140 φοιτητές μιας τάξεως. Η πρώτη στήλη δίνει τις τιμές της μεταβλητής «βαθμός στο μάθημα της Στατιστικής», ενώ η δεύτερη τη συχνότητα της κάθε τιμής της μεταβλητής. Θα μπορούσαμε να περιλάβουμε στην πρώτη στήλη και τους βαθμούς 1, 2 και 10 και να γράψουμε μηδέν στις αντίστοιχες θέσεις της δεύτερης στήλης, αλλά αυτό δεν είναι απαραίτητο.

Για να κατασκευάσουμε τον πίνακα 2.1, όταν μας δοθούν οι 140 βαθμοί, ρίχνουμε μια ματιά στα δεδομένα και βρίσκουμε τον κατώτερο και ανώτερο βαθμό, 3 και 9, αντίστοιχα, στο παράδειγμά μας. Γράφουμε τους αριθμούς 3 έως 9 στην πρώτη στήλη του πίνακα. Για να βρούμε πόσοι φοιτητές πήραν τον κάθε βαθμό υπάρχουν διάφοροι τρόποι: Ένας είναι να διαβάσουμε έναν έναν τους 140 βαθμούς, να σημειώσουμε για τον καθένα μια κάθετη γραμμή απέναντι από τον αντίστοιχο βαθμό της πρώτης στήλης και μετά να μετρήσουμε πόσες γραμμές αντιστοιχούν σε κάθε βαθμό. Ένας άλλος τρόπος είναι να γράψουμε καθένα από τους 140 βαθμούς σε μία κάρτα, να ταξινομήσουμε τις κάρτες σε ομάδες, κάθε μια από τις οποίες θα περιλαμβάνει κάρτες με τον ίδιο βαθμό, και κατόπιν να μετρήσουμε πόσες κάρτες περιλαμβάνει η κάθε ομάδα. Ένας τρίτος τρόπος είναι να κάνουμε την ταξινόμηση των βαθμών με τη βοήθεια ενός ηλεκτρονικού υπολογιστή.

Διαιρώντας κάθε αριθμό της δεύτερης στήλης του πίνακα 2.1 με το άθροισμα όλων των αριθμών της στήλης, το 140, σχηματίζουμε την τρίτη στήλη του πίνακα, η οποία μας δίνει τη σχετική συχνότητα του κάθε βαθμού. Το άθροισμα των σχετικών συχνοτήτων ισούται με τη μονάδα. Πολλαπλασιάζοντας τις σχετικές αυτές συχνότητες με το 100, παίρνουμε την τέταρτη στήλη του πίνακα 2.1, η οποία δίνει τις σχετικές συχνότητες επί

τους εκατό. Το άθροισμα αυτών των συχνοτήτων ισούται με 100.

Ο πίνακας 2.1 δεν είναι μόνο πολύ πιο συνοπτικός από έναν πίνακα π.χ. με 14 γραμμές και 10 στήλες, που θα κατασκευάζαμε αν δίναμε τους 140 βαθμούς των φοιτητών χωρίς καμιά ταξινόμηση, αλλά παρουσιάζει και πολλά άλλα πλεονεκτήματα. Έχοντας τον πίνακα αυτόν, μπορούμε να βρούμε με μεγάλη ευκολία ποιος είναι ο μικρότερος και ο μεγαλύτερος βαθμός, ποιος βαθμός παρουσιάζει τη μεγαλύτερη συχνότητα, τι ποσοστό των φοιτητών πήρε βαθμό μικρότερο από 5 ή μεγαλύτερο από 7 κ.ο.κ.

2.1.2. Αριθμός τιμών μεταβλητής μεγάλος

Όταν ο αριθμός των τιμών που παίρνει η μεταβλητή είναι μεγάλος, ο πίνακας κατανομής συχνότητας αποκτά μεγάλο μήκος, που είναι ανεπιθύμητο. Αυτό μπορεί να συμβεί και με ασυνεχείς μεταβλητές, αλλά αποτελεί τον κανόνα για τις συνεχείς μεταβλητές. Για να περιορίσουμε το μήκος του πίνακα στις περιπτώσεις αυτές, διαιρούμε το εύρος των τιμών της μεταβλητής σε διαστήματα, που ονομάζονται κλάσεις, και μετρούμε κατόπιν τον αριθμό των μελών του δείγματος που εμπίπτουν σε κάθε κλάση, βρίσκουμε δηλαδή τη συχνότητα της κάθε κλάσεως, με έναν από τους τρόπους που περιγράψαμε στην προηγούμενη υποδιαίρεση.

Ο πίνακας 2.2 δίνει την κατανομή συχνότητας των αποδόσεων 168 μεμονωμένων φυτών μηδικής. Στην πρώτη στήλη του πίνακα τα διαστήματα

ΠΙΝΑΚΑΣ 2.2

Κατανομή της αποδόσεως (kg χλωρού βάρους) 168 μεμονωμένων φυτών μηδικής

Κλάση	Μέση τιμή	Συχνότητα	Σχετ. συχν. %
0,6 - 1,0	0,8	5	2,98
1,1 - 1,5	1,3	8	4,76
1,6 - 2,0	1,8	20	11,90
2,1 - 2,5	2,3	20	11,90
2,6 - 3,0	2,8	27	16,07
3,1 - 3,5	3,3	30	17,86
3,6 - 4,0	3,8	20	11,90
4,1 - 4,5	4,3	13	7,74
4,6 - 5,0	4,8	11	6,55
5,1 - 5,5	5,3	7	4,17
5,6 - 6,0	5,8	7	4,17
Σύνολο		168	100,00

(0,6-1,0), (1,1-1,5) κ.λ.π. αποτελούν τις κλάσεις. Η κατώτερη τιμή του κάθε διαστήματος αποτελεί το **κατώτερο όριο** της κλάσεως, ενώ η ανώτερη τιμή το **ανώτερο όριο**. Το **εύρος των κλάσεων** είναι η διαφορά μεταξύ δύο διαδοχικών κατώτερων ή ανώτερων ορίων. Στο παράδειγμά μας είναι 0,5. Ας σημειωθεί ότι το εύρος μιας κλάσεως δεν μπορεί πάντοτε να βρεθεί με την αφαίρεση του κατώτερου από το ανώτερο όριό της. Αυτό ισχύει και για το παράδειγμά μας. Η δεύτερη στήλη του πίνακα 2.2 δίνει τις μέσες τιμές των κλάσεων. Η **μέση τιμή** της κλάσεως είναι ο μέσος όρος των ορίων της. Η διαφορά μεταξύ δύο διαδοχικών μέσων τιμών ισούται με το εύρος των κλάσεων. Όλες οι τιμές που εμπίπτουν σε μια κλάση θεωρείται ότι είναι ίσες με τη μέση τιμή της κλάσεως. Η τρίτη στήλη του πίνακα δίνει την απόλυτη συχνότητα κάθε κλάσεως, ενώ η τέταρτη τη σχετική συχνότητα επί τοις εκατό.

Για να διαιρέσουμε το εύρος των τιμών της μεταβλητής σε κλάσεις, βρίσκουμε πρώτα τη μικρότερη και τη μεγαλύτερη τιμή μέσα στα δεδομένα μας. Στα δεδομένα που χρησιμοποιήθηκαν για την κατασκευή του πίνακα 2.2 η μικρότερη τιμή ήταν 0,8 και η μεγαλύτερη 5,7. Συνήθως χρησιμοποιούμε για κατώτερο όριο της πρώτης κλάσεως έναν αριθμό λίγο μικρότερο από τη μικρότερη τιμή των δεδομένων μας, και για ανώτερο όριο της τελευταίας κλάσεως έναν αριθμό λίγο μεγαλύτερο από τη μεγαλύτερη τιμή τους. Για το παράδειγμά μας οι αριθμοί 0,6 και 6,0 είναι κατάλληλοι για τα δύο παραπάνω όρια. Το διάστημα μεταξύ των δύο αυτών ορίων το διαιρούμε σε ίσα τμήματα, που αποτελούν τις κλάσεις. Στο παράδειγμά μας το διάστημα 0,6-6,0 διαιρέθηκε σε 11 ίσα τμήματα, τις κλάσεις, κάθε μια από τις οποίες έχει εύρος 0,5.

Ο αριθμός των κλάσεων κυμαίνεται συνήθως μεταξύ 10 και 20, και εξαρτάται από διάφορους παράγοντες όπως είναι ο συνολικός αριθμός των δεδομένων, το εύρος των τιμών τους, ο βαθμός στον οποίο θέλουμε να συνοψίσουμε τα δεδομένα μας, η ακρίβεια των υπολογισμών που θέλουμε να κάνουμε με βάση τον πίνακα. Οι δύο τελευταίοι παράγοντες αντιστρατεύονται ο ένας τον άλλο: όσο μεγαλύτερος είναι ο αριθμός των κλάσεων, τόσο ακριβέστεροι είναι οι υπολογισμοί μας, αλλά είναι μικρότερος ο βαθμός συνοψίσεως των δεδομένων.

Επειδή όλες οι τιμές που εμπίπτουν σε μια κλάση θεωρούνται ίσες με τη μέση τιμή, πρέπει να καθορίζουμε τα όρια των κλάσεων κατά τέτοιο τρόπο, ώστε, αν είναι δυνατόν, οι περισσότερες τιμές που εμπίπτουν σε κάθε κλάση να είναι κοντά στη μέση τιμή και οπωσδήποτε να μη συμπίπτουν με τα όρια της κλάσεως: διαφορετικά, θα έχουμε μεγάλο σφάλμα κάνοντας υπολογισμούς με βάση τον πίνακα.