

Μάθημα 4

ΚΕΦΑΛΑΙΟ 3 ΑΡΙΘΜΗΤΙΚΕΣ ΚΑΙ ΓΡΑΦΙΚΕΣ ΤΕΧΝΙΚΕΣ
ΠΕΡΙΓΡΑΦΗΣ ΟΝΟΜΑΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

§ 3.1 Ταξινόηση και Καταρέτρωση Ονομαστικών Δεδομένων

Τα ονομαστικά δεδομένα (διατάξιμα ή κατηγορικά) μπορούν να απαριθμηθούν δεν είναι όμως μετρήσιμα. Η μόνη δυνατότητα που μας δίνεται είναι να καταετρήσουμε τις συχνοτικές τιμές των μεταβλητών για να βρούμε το πλήθος των στατιστικών μονάδων που ερπύπτουν σε κάθε μία από τις διαφορετικές κατηγορίες που προσδιορίζονται από την φύση της κάθε εξεταζόμενης μεταβλητής.

Τα αποτελέσματα της απαρίθμησης παρουσιάζονται συνοπτικά σε στατιστικούς πίνακες (statistical tables) που περιγράφουν πως κατανέμονται οι παρατηρήσεις σε σχέση με το ποιοτικό χαρακτηριστικό που μελετάμε.

Συνήθως το περιεχόμενο των πινάκων παρουσιάζεται διαγραμματικά με ειδικές γραφικές παραστάσεις (graphs) που παρέχουν μία άρεση και εύχλητη οπτική εικόνα της δομής του υπό μελέτη πληθυσμού ή δείγματος.

Ας δούμε ένα παράδειγμα. Έστω ότι δίνονται οι απαντήσεις 25 ανδρών σχετικά με την οικογενειακή τους κατάσταση όπου Α: Άγαθοι, Ε: Έγγαμοι, Χ: Χήροι και Δ: Διαζευγμένοι.

Ε Ε Ε Α Χ Δ Ε Ε Χ Α Ε Α Χ Ε Ε Δ Χ Α Ε Α Ε
Α Ε Α Α

Μετράμε το πλήθος των Α (άδαρων), των Ε (εσδάρων), των Χ (χήρων) και των Δ (διαζευγμένων). Προκύπτει ο ακόλουθος πίνακας:

		πλήθος f_i
Άδαροι	Α Α Α Α Α Α Α	$f_1 = 8$
Εσδαροι	Ε Ε Ε Ε Ε Ε Ε Ε Ε Ε	$f_2 = 11$
Χήροι	Χ Χ Χ Χ	$f_3 = 4$
Διαζευγμένοι	Δ Δ	$f_4 = 2$
Συνολικός αριθμός Παρατηρήσεων		$n = 25$

Στα 25 άτομα που εξετάσαμε βρέθηκαν 8 άδαροι, 11 έσδαροι, 4 χήροι και 2 διαζευγμένοι. Αυτά τα αποτελέσματα αυτά μας δίνουν την εμπειρική ή πειραματική κατανομή (empirical distribution) των στοιχείων του δείγματος ανάλογα με την οικογενειακή τους κατάσταση.

Τα στοιχεία της τελευταίας στήλης δηλώνουν πόσο συχνά ερπύπτουν οι τιμές της μεταβλητής σε κάθε μία από τις 4 οικογενειακές κατηγορίες. Οι αριθμητικές τιμές που προέκυψαν από τη διαδικασία καταμέτρησης καλούνται συχνότητες (frequencies). Επειδή πρόκειται για πλήθος στατιστικών μονάδων καλούνται απόλυτες συχνότητες (absolute frequencies). Αυτές τις συχνότητες τις συμβολίζουμε με f_i , $i=1, \dots, k$ όπου με i

συμβολίζουμε την πρώτη, δεύτερη κ.ο.κ. κατηγορία της ποιοτικής μεταβλητής. Στο παράδειγμά μας είναι $i=1, 2, 3, 4$.

Εναλλακτικά, θα μπορούσαμε να πούμε ότι στους 25 άνδρες του δείγματος αναλογούν 8 άγαμοι. Δηλαδή η αναλογία των άγαμων στο δείγμα είναι $\frac{8}{25} = 0.32$ δηλαδή οι άγαμοι αντιπροσωπεύουν το 32% του συνόλου των παρατηρούμενων στατιστικών μονάδων. Η αναλογία των έγγαμων είναι $\frac{11}{25}$ δηλαδή 44% των χήρων $\frac{4}{25}$ δηλαδή 16% και των διαβευχθέντων $\frac{2}{25}$ ή ποσοστό 8%. Αυτοί οι αριθμοί εκφράζουν τη σχετική (αναλογική ή ποσοστιαία) κατανομή των στοιχείων του υπό μελέτη δείγματος και για το λόγο αυτό καλούνται σχετικές συχνότητες (relative frequencies).

Οι σχετικές συχνότητες μπορούν να εκφραστούν είτε ανά μονάδα π.χ. ανά άτομο οπότε έχουν τη μορφή της αναλογίας δηλαδή $\frac{f_i}{n}$ είτε ως ποσοστό επί 100 ατόμων δηλαδή $\frac{f_i}{n} \cdot 100$. Αν η ανάλυση αφορά επεξεργασία στοιχείων πληθυσμού οι σχετικές συχνότητες θα είναι $\frac{f_i}{N}$ και $\frac{f_i}{N} \cdot 100$ αντίστοιχα. Συνήθως τις σχετικές συχνότητες τις συμβολίζουμε με $f_i(\%)$.

Στο παράδειγμά μας, το άθροισμα των αποχρώσεων, των αναλογιών και των ποσοστιαίων συχνοτήτων είναι

αντίστοιχα: $8 + 11 + 4 + 2 = 25$

$$0.32 + 0.44 + 0.16 + 0.08 = 1$$

$$32\% + 44\% + 16\% + 8\% = 100\%$$

Το άθροισμα των απόλυτων συχνοτήτων είναι ίσο με τον συνολικό αριθμό παρατηρήσεων που επεξεργαστήσατε:

$$\sum_{i=1}^K f_i = n \quad \text{ή} \quad \sum_{i=1}^K f_i = N \quad \text{ανάλογα αν η ανάλυση}$$

βασίζεται σε στοιχεία δείγματος μεγέθους n ή σε στοιχεία πληθυσμού μεγέθους N .

Το άθροισμα των σχετικών συχνοτήτων είναι πάντα ίσο με 1 ή με 100 ανάλογα με το αν οι σχετικές συχνότητες εκφράζονται ανά μονάδα ή ως ποσοστό επί τοις εκατό. Δηλαδή

$$\sum_{i=1}^K \frac{f_i}{n} = \frac{1}{n} \sum_{i=1}^K f_i = \frac{1}{n} n = 1 \quad \text{ή} \quad \sum_{i=1}^K \frac{f_i}{n} \cdot 100 =$$

$$= 100 \sum_{i=1}^K \frac{f_i}{n} = 100 \quad \text{για τα στοιχεία δείγματος ενώ για}$$

τα στοιχεία ενός πληθυσμού

$$\sum_{i=1}^K \frac{f_i}{N} = \frac{1}{N} \sum_{i=1}^K f_i = \frac{1}{N} N = 1 \quad \text{ή} \quad \sum_{i=1}^K \frac{f_i}{N} \cdot 100 = 100 \sum_{i=1}^K \frac{f_i}{N} = 100.$$

ΠΙΝΑΚΑΣ 3.1

Ομογενειακή κατάσταση	Απόλυτες συχνότητες	Αναλογίες	Ποσοστά
	f_i	$\frac{f_i}{n}$	$\frac{f_i}{n} \cdot 100$
Άγαροι	8	0.32	32
Έγγατοι	11	0.44	44
Χήροι	4	0.16	16
Διαζευγμένοι	2	0.08	8
Σύνολο	25	1	100

Ο Πίνακας 3.1 παρουσιάζει την κατανομή κατά οικογενειακή κατάσταση τυχαίου δείγματος $n=25$ ανδρών

Οι σχετικές συχνότητες χρησιμοποιούν για συζητήσεις μεταξύ δειγμάτων ή πληθυσμών και για την καλύτερη κατανόηση της σύνθεσης και της δομής ενός δείγματος ή ενός πληθυσμού. Οι απόλυτες συχνότητες είναι απαραίτητες για να έχουμε μία σαφή εικόνα των πραγματικών μεγεθών μιας κατάστασης ή ενός φαινομένου που μελετάμε. Για παράδειγμα, για τον προϋπολογισμό των δαπανών για το σχεδιασμό των υπηρεσιών που θα απαιτηθούν για την επέμβαση σε ορισμένες περιοχές του δικτύου κεντρικού ρεύματος δεν αρκεί μόνο η διαθεσιμότητα του ποσοστού των νοικοκυριών που στερούνται τέτοιων παροχών αλλά και η γνώση του **απολύτου** αριθμού τους.