

# Η θεωρία της στατιστικής σε ερωτήσεις-απαντήσεις

## Μέρος 1<sup>ον</sup> (έως ομαδοποίηση δεδομένων)

1) Πώς ορίζεται η Στατιστική επιστήμη;

**Στατιστική** είναι ένα σύνολο αρχών και μεθοδολογιών για:

- το σχεδιασμό της διαδικασίας συλλογής δεδομένων
- τη συνοπτική και αποτελεσματική παρουσίασή τους
- την ανάλυση και εξαγωγή αντίστοιχων συμπερασμάτων.

2) Ποιοι είναι οι κλάδοι της Στατιστικής επιστήμης; Με τι ασχολείται ο κάθε κλάδος;

- **σχεδιασμός πειραμάτων**
- **περιγραφική στατιστική**
- **επαγωγική στατιστική ή στατιστική συμπερασματολογία**

Ο κλάδος της Στατιστικής που ασχολείται με το σχεδιασμό της διαδικασίας συλλογής δεδομένων λέγεται **σχεδιασμός πειραμάτων**. Με τη συνοπτική και αποτελεσματική παρουσίασή τους ασχολείται η **περιγραφική στατιστική** που αποτελεί και το αντικείμενο μελέτης μας. Τέλος, με την ανάλυση και εξαγωγή συμπερασμάτων ασχολείται η **επαγωγική στατιστική**

3) Τι ονομάζουμε πληθυσμό;

Είναι ένα σύνολο του οποίου θέλουμε να εξετάσουμε τα στοιχεία του ως προς ένα ή περισσότερα χαρακτηριστικά τους.

Τα στοιχεία του πληθυσμού συχνά αναφέρονται και ως **μονάδες ή άτομα** του πληθυσμού.

4) Τι ονομάζουμε μεταβλητές; Τι είναι οι τιμές της μεταβλητής;

Τα χαρακτηριστικά ως προς τα οποία εξετάζουμε έναν πληθυσμό λέγονται **μεταβλητές** και τις συμβολίζουμε συνήθως με τα κεφαλαία γράμματα  $X, \Psi, Z, \dots$

Οι δυνατές τιμές (**διαφορετικές μεταξύ τους**) που μπορεί να πάρει μια μεταβλητή λέγονται **τιμές της μεταβλητής** και συμβολίζονται ως  $x_1, x_2, \dots, x_k$ ,  $k \leq n$  όπου  $n$  το πλήθος των στοιχείων του συνόλου.

# Η θεωρία της στατιστικής σε ερωτήσεις-απαντήσεις

## Μέρος 1<sup>ον</sup> (έως ομαδοποίηση δεδομένων)

---

5) Τι λέμε στατιστικά δεδομένα ή παρατηρήσεις ;

Από τη μελέτη των ατόμων του πληθυσμού ως προς ένα χαρακτηριστικό τους προκύπτει μια σειρά από δεδομένα, που λέγονται στατιστικά δεδομένα ή παρατηρήσεις και συμβολίζονται  $t_1, t_2, \dots, t_n$ .

6) Σε τι διακρίνονται οι μεταβλητές

Τις μεταβλητές τις διακρίνουμε:

- Σε **ποιοτικές** ή **κατηγορικές** μεταβλητές, των οποίων οι τιμές τους δεν είναι αριθμοί.
- Σε **ποσοτικές** μεταβλητές, των οποίων οι τιμές είναι αριθμοί

7) Σε τι διακρίνονται οι ποσοτικές μεταβλητές ;

- Σε **διακριτές** μεταβλητές, που παίρνουν μόνο «μεμονωμένες» τιμές.
- Σε **συνεχείς** μεταβλητές, που μπορούν να πάρουν οποιαδήποτε τιμή ενός διαστήματος πραγματικών αριθμών  $(\alpha, \beta)$ .

8) Τι ονομάζουμε απογραφή ;

Ένας τρόπος για να πάρουμε τις απαραίτητες πληροφορίες που χρειαζόμαστε για κάποιο πληθυσμό είναι να εξετάσουμε όλα τα άτομα του πληθυσμού ως προς το χαρακτηριστικό που μας ενδιαφέρει. Η μέθοδος αυτή ονομάζεται **απογραφή**.

9) Τι ονομάζουμε δείγμα ; Τι ονομάζουμε δημοσκόπηση ;

Σε πολλές περιπτώσεις η απογραφή όλων των μονάδων του πληθυσμού είναι

# Η θεωρία της στατιστικής σε ερωτήσεις-απαντήσεις

## Μέρος 1<sup>ον</sup> (έως ομαδοποίηση δεδομένων)

---

δύσκολη, ή ασύμφορη, ή ακόμα και αδύνατη. Μαζεύονται, τότε, πληροφορίες από κάποια μικρή ομάδα, υποσύνολο του πληθυσμού, το οποίο καλείται **δείγμα**. Γίνονται παρατηρήσεις στο δείγμα αυτό και μετά γενικεύονται τα συμπεράσματα για ολόκληρο τον πληθυσμό. Τα συμπεράσματα αυτά ισχύουν με ικανοποιητική ακρίβεια για όλο τον πληθυσμό, αν το δείγμα είναι, όπως λέμε, **αντιπροσωπευτικό** του πληθυσμού. Όταν το δείγμα λαμβάνεται από ανθρώπινο πληθυσμό, τότε η στατιστική μελέτη και έρευνα λέγεται δημοσκόπηση

10) Πότε λέμε ότι ένα δείγμα είναι αντιπροσωπευτικό;

Ένα δείγμα είναι αντιπροσωπευτικό ενός πληθυσμού, εάν έχει επιλεγεί κατά τέτοιο τρόπο, ώστε κάθε μονάδα του πληθυσμού να έχει την ίδια δυνατότητα να επιλεγεί.

11) Τι είναι η Δειγματοληψία;

Οι αρχές και οι μέθοδοι για τη συλλογή και ανάλυση δεδομένων από πεπερασμένους πληθυσμούς είναι το αντικείμενο της **Δειγματοληψίας**, που αποτελεί τη βάση της Στατιστικής.

12) Τι είναι οι στατιστικοί πίνακες Σε τι διακρίνονται; Τι πρέπει να περιέχει ώστε να είναι πλήρης ένας στατιστικός πίνακας;

Μετά τη συλλογή των στατιστικών δεδομένων είναι αναγκαία η κατασκευή συνοπτικών **πινάκων** ή **γραφικών παραστάσεων**, ώστε να είναι εύκολη η κατανόησή τους και η εξαγωγή σωστών συμπερασμάτων

Οι πίνακες διακρίνονται στους:

- **γενικούς πίνακες**, οι οποίοι περιέχουν όλες τις πληροφορίες που προκύπτουν από μία στατιστική έρευνα.
- **ειδικούς πίνακες**, οι οποίοι περιέχουν συνοπτικά μερικά στοιχεία της έρευνας που έχουν

## Η θεωρία της στατιστικής σε ερωτήσεις-απαντήσεις Μέρος 1<sup>ον</sup> (έως ομαδοποίηση δεδομένων)

ληφθεί από τους γενικούς πίνακες.

Κάθε πίνακας περιέχει:

- α) τον **τίτλο**, που περιγράφει το περιεχόμενο του πίνακα,
- β) τις **επικεφαλίδες** των στηλών, που δείχνουν τη φύση και τις μονάδες μέτρησης των δεδομένων,
- γ) το **κύριο σώμα**, που περιέχει τα στατιστικά δεδομένα,
- δ) την **πηγή**, που γράφεται στο κάτω μέρος του πίνακα και δείχνει την προέλευση των στατιστικών στοιχείων.

13) Τι ονομάζουμε απόλυτη συχνότητα

Ας υποθέσουμε ότι  $x_1, x_2, \dots, x_k$  είναι οι τιμές μιας μεταβλητής  $X$ , που αφορά τα άτομα ενός δείγματος μεγέθους  $n$ ,  $k \leq n$ . Στην τιμή  $x_i$  αντιστοιχίζεται η (απόλυτη) συχνότητα  $v_i$   $i=1, 2, \dots, k$ , δηλαδή, ο φυσικός αριθμός που δείχνει πόσες φορές εμφανίζεται η τιμή  $x_i$  της εξεταζόμενης μεταβλητής  $X$  στο σύνολο των παρατηρήσεων.

Είναι φανερό ότι το άθροισμα όλων των συχνοτήτων είναι ίσο με το μέγεθος  $n$  του δείγματος, δηλαδή:

$$v_1 + v_2 + \dots + v_k = n \quad \text{ή} \quad \sum_{i=1}^k v_i = n$$

14) Τι ονομάζουμε σχετική συχνότητα  $f_i$  ;  
Τι ονομάζουμε σχετική συχνότητα επί τοις εκατό  $f_i$  % ;

Αν διαιρέσουμε τη συχνότητα  $v_i$  με το μέγεθος  $n$  του δείγματος, προκύπτει η **σχετική συχνότητα**  $f_i$  της τιμής  $x_i$ , δηλαδή

$$f_i = \frac{v_i}{n}, \quad i=1, 2, \dots, k.$$

Όταν τις σχετικές συχνότητες  $f_i$  τις εκφράζουμε επί τοις εκατό, τότε συμβολίζονται με  $f_i$ %, δηλαδή  $f_i\% = 100f_i$ .

# Η θεωρία της στατιστικής σε ερωτήσεις-απαντήσεις

## Μέρος 1<sup>ον</sup> (έως ομαδοποίηση δεδομένων)

15) Να αποδείξετε ότι για την σχετική συχνότητα ισχύουν :

(i)  $0 \leq f_i \leq 1$  για  $i = 1, 2, \dots, \kappa$ .

(ii)  $f_1 + f_2 + \dots + f_\kappa = 1$

ι) αφού  $0 \leq v_i \leq v$  είναι  $0 \leq \frac{v_i}{v} \leq 1$

ή  $0 \leq f_i \leq 1$

ιι) Είναι :

$$f_1 + f_2 + \dots + f_\kappa = \frac{v_1}{v} + \frac{v_2}{v} + \dots + \frac{v_\kappa}{v} =$$

$$\frac{v_1 + v_2 + \dots + v_\kappa}{v} = \frac{v}{v} = 1$$

16) Τι λέμε πίνακα κατανομής συχνοτήτων ;

Οι ποσότητες  $x_i, v_i, f_i$  για ένα δείγμα συγκεντρώνονται σε ένα συνοπτικό πίνακα, που ονομάζεται **πίνακας κατανομής συχνοτήτων** ή απλά **πίνακας συχνοτήτων**

17) Τι είναι η κατανομή συχνοτήτων ή η κατανομή σχετικών συχνοτήτων

Για μια μεταβλητή, το σύνολο των ζευγών  $(x_i, v_i)$  λέμε ότι αποτελεί την **κατανομή συχνοτήτων** και το σύνολο των ζευγών  $(x_i, f_i)$ , ή των ζευγών  $(x_i, f_i \%)$ , την **κατανομή των σχετικών συχνοτήτων**.

18) Τι είναι η αθροιστική συχνότητα, η σχετική αθροιστική συχνότητα, σε ποιο είδος μεταβλητής χρησιμοποιούνται ;

Οι **αθροιστικές συχνότητες**  $N_i$  εκφράζουν το πλήθος των παρατηρήσεων που είναι μικρότερες ή ίσες της τιμής  $x_i, i = 1, 2, \dots, \kappa$ .

Οι **αθροιστικές σχετικές συχνότητες**  $F_i$ , εκφράζουν το ποσοστό των παρατηρήσεων που είναι μικρότερες ή ίσες της τιμής  $x_i, i = 1, 2, \dots, \kappa$ .

Συχνά οι  $F_i$  πολλαπλασιάζονται επί 100 εκφραζόμενες έτσι επί τοις εκατό, δηλαδή

$$F_i \% = 100F_i.$$

Χρησιμοποιούνται για **ποσοτικές**

## Η θεωρία της στατιστικής σε ερωτήσεις-απαντήσεις Μέρος 1<sup>ον</sup> (έως ομαδοποίηση δεδομένων)

---

μεταβλητές

Ισχύει:  $N_i = v_1 + v_2 + \dots + v_i \quad i=1,2,\dots,\kappa$

Και  $F_i = f_1 + f_2 + \dots + f_i$ , για  $i=1,2,\dots,\kappa$  μόνο όταν οι τιμές  $x_1, x_2, \dots, x_\kappa$  μιας ποσοτικής μεταβλητής είναι σε αύξουσα σειρά. Προφανώς ισχύει :

$v_1 = N_1, v_2 = N_2 - N_1, \dots, v_\kappa = N_\kappa - N_{\kappa-1}, N_\kappa = v$

Και

$f_1 = F_1, f_2 = F_2 - F_1, \dots, f_\kappa = F_\kappa - F_{\kappa-1}, F_\kappa = 1$

19) Τι είναι το ραβδόγραμμα συχνοτήτων ;

Το **ραβδόγραμμα συχνοτήτων** χρησιμοποιείται για τη γραφική παράσταση των τιμών μιας **ποιοτικής μεταβλητής**. Το ραβδόγραμμα αποτελείται από ορθογώνιες στήλες που οι βάσεις τους βρίσκονται πάνω στον οριζόντιο ή τον κατακόρυφο άξονα και το μήκος των βάσεων ή των ορθογωνίων στηλών είναι αυθαίρετο. Σε κάθε τιμή της μεταβλητής  $X$  αντιστοιχεί μια ορθογώνια στήλη της οποίας το ύψος είναι ίσο με την αντίστοιχη συχνότητα. Έτσι έχουμε το **ραβδόγραμμα συχνοτήτων**. Αντίστοιχα ορίζεται και το **ραβδόγραμμα σχετικών συχνοτήτων**.

20) Τι είναι το διάγραμμα συχνοτήτων;

Στην περίπτωση που έχουμε μια **ποσοτική μεταβλητή** αντί του ραβδογράμματος χρησιμοποιείται το **διάγραμμα συχνοτήτων**. Αυτό μοιάζει με το ραβδόγραμμα με μόνη διαφορά ότι αντί να χρησιμοποιούμε συμπαγή ορθογώνια υψώνουμε σε κάθε  $x_i$  μία κάθετη γραμμή με μήκος ίσο προς την αντίστοιχη συχνότητα. Αν αντί των συχνοτήτων  $v_i$  στον κάθετο άξονα να βάλουμε τις σχετικές συχνότητες  $f_i$ , έχουμε το **διάγραμμα σχετικών συχνοτήτων**.

## Η θεωρία της στατιστικής σε ερωτήσεις-απαντήσεις Μέρος 1<sup>ον</sup> (έως ομαδοποίηση δεδομένων)

---

21) Τι είναι το διάγραμμα αθροιστικών συχνοτήτων  $N_i$  και τι το διάγραμμα σχετικών αθροιστικών συχνοτήτων  $F_i$ ;

Τοποθετούμε τα  $x_i$  με σειρά μεγέθους στον οριζόντιο άξονα  $X$  και υψώνουμε σε κάθε  $x_i$  μία κάθετη γραμμή με μήκος ίσο προς την αντίστοιχη αθροιστική συχνότητα  $N_i$ . Αντίστοιχα υψώνουμε σε κάθε  $x_i$  μία κάθετη γραμμή με μήκος ίσο προς την αντίστοιχη σχετική αθροιστική συχνότητα  $F_i$ .

22) Τι είναι το πολύγωνο συχνοτήτων;

Ενώνοντας τα σημεία  $(x_i, v_i)$  ή  $(x_i, f_i)$  έχουμε το λεγόμενο **πολύγωνο συχνοτήτων** ή **πολύγωνο σχετικών συχνοτήτων**, αντίστοιχα, που μας δίνουν μια γενική ιδέα για τη μεταβολή της συχνότητας ή της σχετικής συχνότητας όσο μεγαλώνει η τιμή της μεταβλητής που εξετάζουμε.

23) Τι είναι το κυκλικό διάγραμμα;

Το **κυκλικό διάγραμμα** χρησιμοποιείται για τη γραφική παράσταση τόσο των ποσοτικών όσο και των ποσοτικών δεδομένων συνήθως όταν οι τιμές είναι λίγες.

Το κυκλικό διάγραμμα είναι ένας κυκλικός δίσκος χωρισμένος σε κυκλικούς τομείς, τα τόξα των οποίων είναι ανάλογα προς τις αντίστοιχες συχνότητες  $v_i$  ή τις σχετικές συχνότητες  $f_i$  των τιμών  $x_i$  της μεταβλητής.

24) Με ποιο τύπο υπολογίζω το μέτρο της επίκεντρης γωνίας ή του αντίστοιχου τόξου στο κυκλικό διάγραμμα;

Αν συμβολίσουμε με  $\alpha_i$  το αντίστοιχο τόξο ενός κυκλικού τμήματος στο κυκλικό διάγραμμα συχνοτήτων, τότε

$$\alpha_i = v_i \frac{360^\circ}{v} = 360^\circ f_i \quad \text{για } i=1,2,\dots,k.$$