

ΣΗΜΕΙΩΣΕΙΣ ΣΤΑΤΙΣΤΙΚΗΣ

1 Ορισμοί

Στατιστική είναι το σύνολο αρχών και μεθοδολογίας για:

1. Σχεδιασμός της διαδικασίας συλλογής δεδομένων
2. Συνοπτική και αποτελεσματική παρουσίαση των δεδομένων
3. Ανάλυση των δεδομένων και εξαγωγή συμπερασμάτων

Χρησιμότητα της Στατιστικής

1. Κατανόηση του παρελθόντων (παρελθόντων δεδομένων)
2. Πρόβλεψη του μέλλοντος

Πληθυσμός ονομάζεται το σύνολο των οντοτήτων του οποίου τα στοιχεία εξετάζουμε ως προς ένα ή περισσότερα χαρακτηριστικά

Μονάδες ή στοιχεία του πληθυσμού ονομάζονται τα στοιχεία του πληθυσμού

Μεταβλητή ονομάζεται το χαρακτηριστικό ως προς το οποίο εξετάζουμε τις μονάδες ή άτομα ενός πληθυσμού

Στατιστικά δεδομένα ή παρατηρήσεις ονομάζονται τα δεδομένα (αποτελέσματα) από την παρατήρηση του χαρακτηριστικού (μεταβλητής) σε κάθε μονάδα ή άτομο του πληθυσμού. Συνήθως συμβολίζονται ως t_1, t_2, t_3, \dots

Οι μεταβλητές χωρίζονται σε δύο μεγάλες κατηγορίες, **ποιοτικές** ή **ποσοτικές**. **Ποιοτική** ή **κατηγορική** (από το **κατηγορία**) είναι η μεταβλητή που

οι τιμές της δεν είναι αριθμοί, ή εάν είναι, δεν μπορούν να εννοηθούν μέτρα θέσης (πχ μέση τιμή). **Ποσοτική** είναι η μεταβλητή που οι τιμές της είναι αριθμοί, και εννοούνται μέτρα θέσης (πχ μέση τιμή)

Οι ποσοτικές μεταβλητές διακρίνονται σε δυο κατηγορίες, τις **διακριτές** και τις **συνεχείς**. **Συνεχείς** μεταβλητές είναι αυτές που μπορούν να πάρουν οποιαδήποτε τιμή ενός διαστήματος πραγματικών αριθμών. **Διακριτές** μεταβλητές είναι αυτές οι οποίες "παίρνουν" μεμονωμένες τιμές (δηλαδή το σύνολο των υπαρκτών διαφορετικών τιμών είναι μικρό, πχ μικρότερο από 10-20 διαφορετικές τιμές)

Δείγμα ονομάζεται ένα υποσύνολο του πληθυσμού το οποίο εξετάζεται ως προς το χαρακτηριστικό που μας ενδιαφέρει.

Αντιπροσωπευτικό δείγμα ονομάζεται το δείγμα ενός πληθυσμού, εάν έχει επιλεγεί με τέτοιο τρόπο ώστε κάθε μονάδα του πληθυσμού να έχει την ίδια πιθανότητα να επιλεγεί.

Δειγματοληψία ονομάζεται η διαδικασία επιλογής ενός δείγματος.

Μέγεθος δείγματος ονομάζεται το πλήθος των παρατηρήσεων του δείγματος. Είναι πάντα θετικός αριθμός και συμβολίζεται με το γράμμα n

Στατιστικοί Πίνακες Για να

είναι σωστά κατασκευασμένος ένας στατιστικός πίνακας πρέπει να περιέχει:

1. **Τίτλος.** Γράφεται στο πάνω μέρος του πίνακα και δηλώνει με σαφήνεια και συνοπτικά τι περιέχει ο πίνακας
2. **Επικεφαλίδα.** Γράφεται σε κάθε γραμμή ή στήλη και δείχνει συνοπτικά την φύση και την μονάδα μέτρησης των δεδομένων
3. **Κύριο σώμα.** Περιέχει, διαχωρισμένα μέσα σε γραμμές και στήλες τα στατιστικά δεδομένα
4. **Πηγή.** Γράφεται στο κάτω μέρος του πίνακα και δείχνει την προέλευση των στατιστικών στοιχείων (πχ για επαλήθευση ή περαιτέρω έρευνα).

Ο φυσικός αριθμός που δείχνει πόσες φορές εμφανίζεται η τιμή $x_i, i = 1, 2, \dots, \kappa$ της εξεταζόμενης μεταβλητής, λέγεται **συχνότητα** και συμβολίζεται $\nu_i, i = 1, 2, \dots, \kappa$

Σχετική συχνότητα f_i της τιμής x_i ορίζεται ως $f_i = \frac{\nu_i}{\nu}, i = 1, 2, \dots, \kappa$

Πίνακας Κατανομής Συχνοτήτων ονομάζεται ο συνοπτικός πίνακας που έχει τα στοιχεία όπως οι διαφορετικές περιπτώσεις της μεταβλητής x_i καθώς και τις στήλες ν_i και f_i .

Οι ομάδες στις οποίες ταξινομούνται οι τιμές ονομάζονται **κλάσεις** και είναι της μορφής $[\alpha, \beta)$. Τα άκρα της κλάσης, δηλαδή τα α και β ονομάζονται **όρια των κλάσεων**.

Το κέντρο κάθε κλάσης λέγεται **κεντρική τιμή** x_i . Το κέντρο της κλάσης

$[\alpha, \beta)$ δίδεται από τον τύπο $x_i = \frac{\alpha + \beta}{2}$

Πλάτος c μιας κλάσης ονομάζεται η διαφορά του κατώτερου από το ανώτερο όριο της κλάσης. Δηλαδή για μια κλάση $[\alpha, \beta)$ αυτό είναι $c = \beta - \alpha$

Ευρος R του δείγματος ονομάζεται η διαφορά της μικρότερης παρατήρησης από την μεγαλύτερη παρατήρηση του συνολικού δείγματος. Δηλαδή $R = t_{max} - t_{min}$ (ισχύει μόνο για μεταβλητές ποσοτικές)

Διάμεσος δ του δείγματος ονομάζεται η μεσαία παρατήρηση εφόσον οι παρατηρήσεις έχουν διαταχθεί σε αύξουσα ή φθίνουσα σειρά, και είναι το πλήθος τους περιττός αριθμός. Εάν το πλήθος είναι άρτιος αριθμός, τότε διάμεσος είναι ο μέσος όρος των δυο μεσαίων παρατηρήσεων εφόσον έχουν διαταχθεί οι παρατηρήσεις. (ισχύει μόνο για μεταβλητές ποσοτικές)

Επικρατούσα τιμή είναι η τιμή της μεταβλητής που εμφανίζεται περισσότερες φορές από κάθε άλλη τιμή. Μπορούν να υπάρχουν περισσότερες από μια επικρατούσες τιμές, αλλά πρέπει να είναι σίγουρα λιγότερες από το δείγμα.

ΔΙΑΓΡΑΜΜΑΤΑ

Ραβδόγραμμα είναι ένα διάγραμμα όπου χρησιμοποιούνται συνήθως για ποιοτικές μεταβλητές, με στήλες που βρίσκονται σε οριζόντιο (κάθετο ραβδόγραμμα) ή κάθετο άξονα (οριζόντιο ραβδόγραμμα). Μπορεί το ραβδόγραμμα να είναι ραβδόγραμμα συχνοτήτων, σε αυτή την περίπτωση, το μέγεθος (μήκος ή ύψος) της κάθε

ράβδου είναι η αντίστοιχη συχνότητα. Η' μπορεί να είναι ραβδόγραμμα σχετικών συχνοτήτων (ή σχετικών συχνοτήτων επί τοις εκατό), οπότε σε αυτή την περίπτωση το μέγεθος (μήκος ή ύψος) είναι αντίστοιχα η σχετική συχνότητα (ή σχετική συχνότητα επί τοις εκατό).

Το κυκλικό διάγραμμα χρησιμοποιείται για την γραφική παράσταση ποιοτικών και ποσοτικών δεδομένων. Χρησιμοποιείται όταν οι διαφορετικές τιμές είναι σχετικά λίγες. Στο κυκλικό διάγραμμα επειδή είναι δύσκολο να αντιληφθούμε πολλές φορές ποιος τομέας είναι μεγαλύτερος, δίδονται και άλλες πληροφορίες για κάθε τομέα, είτε η συχνότητα, είτε η σχετική συχνότητα είτε η σχετική συχνότητα επί τοις εκατό. Για τον υπολογισμό του κάθε τομέα, βρίσκουμε πόσες μοίρες πρέπει να είναι στον κύκλο, όπου το σύνολο είναι 360^0 . Κάθε τομέας a_i έχει μοίρες $a_i = f_i \cdot 360^0$

2 Μεθοδολογία

Πως κατασκευάζουμε ένα πίνακα συχνοτήτων, οι διάφορες στήλες, και πως υπολογίζονται

1. **Μεταβλητή X** Βρίσκουμε ποιά είναι η μεταβλητή X
2. **Δυνατές τιμές της μεταβλητής X** Εντοπίζουμε τις διαφορετικές τιμές που παίρνει η μεταβλητή X . Για κάθε διαφορετική λοιπόν τιμή της X κάνουμε απο μια γραμμή στον πίνακα συχνοτήτων. Γράφουμε τις τιμές στην στήλη X_i κατα αύξουσα σειρά (εαν είναι ποσοτική μεταβλητή) ή με όποια σειρά θέλουμε (εαν είναι ποιοτική μεταβλητή)
3. **Συχνότητα ν_i** Σε κάθε γραμμή, μετράμε πόσες φορές εμφανίζεται η αντίστοιχη μεταβλητή, και αυτό το γράφουμε στην ίδια γραμμή στην στήλη ν_i . Το άθροισμα όλων των συχνοτήτων στο τέλος ισούτε με τον πληθυσμό.
4. **Αθροιστική Συχνότητα N_i** Στην στήλη αυτή έχουμε το άθροισμα των συχνοτήτων ν_i μέχρι την συγκεκριμένη γραμμή. Ενας άλλος τρόπος υπολογισμού είναι ότι για κάθε i , $N_i = N_{i-1} + \nu_i$. Με απλά λόγια δηλαδή, ισουται με το επάνω N_i συν το ν_i στην ίδια γραμμή.
5. **Σχετική συχνότητα f_i** Για τον υπολογισμό της σχετικής συχνότητας χρειαζόμαστε το αντίστοιχο ν_i και τον πληθυσμό ν . Ο υπολογισμός είναι $f_i = \frac{\nu_i}{\nu}$. Η σχέση αυτή χρησιμοποιείτε επίσης οποτεδήποτε γνωρίζουμε δυο απο τα τρία στοιχεία, και θέλουμε το τρίτο.
6. **Αθροιστική Σχετική συχνότητα F_i** Λειτουργεί με την ίδια λογική με την αθροιστική συχνότητα, αλλά χρησιμοποιώντας την στήλη f_i και όχι την ν_i .
7. **Σχετική συχνότητα επι τοις εκατό $f_i\%$**

Πως βρίσκουμε την **διάμεσο** (ΜΟΝΟ ΓΙΑ ΠΟΣΟΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ)

1. Εαν έχουμε παρατηρήσεις
 - (a) Βάζουμε στην σειρά τις παρατηρήσεις, απο την μικρότερη τιμή προς την μεγαλύτερη
 - (b) Μετράμε τον πληθυσμό των παρατηρήσεων, έστω ότι είναι ν
 - (c) Εαν ο πληθυσμός είναι άρτιος αριθμός (ζυγός), τότε παίρνουμε την $\frac{\nu}{2}$ παρατήρηση, δηλαδή την $t_{\frac{\nu}{2}}$ και την επόμενη της, δηλαδή την $t_{\frac{\nu}{2}+1}$ και τις προσθέτουμε και το αποτέλεσμα το διαιρούμε δια δύο. Αυτό είναι η διάμεσος δ . Δηλαδή $\delta = \frac{t_{\frac{\nu}{2}} + t_{\frac{\nu}{2}+1}}{2}$
 - (d) Εαν όμως ο πληθυσμός είναι περιττός αριθμός (μονός), τότε παίρνουμε ακριβώς την παρατήρηση που είναι στην μέση, και αυτή είναι η ακέραια τιμή του $\frac{\nu}{2}$ συν 1. Δηλαδή $\delta = t_{[\frac{\nu}{2}]+1}$

2. Εάν έχουμε πίνακα συχνοτήτων, η λογική είναι ανάλογη. Υποθέτουμε ότι είναι σε αύξουσα διάταξη ως προς την τιμή της μεταβλητής x_i . Βοηθάει επίσης να έχουμε την στήλη της αθροιστικής συχνότητας N_i .

- Βρίσκουμε τον πληθυσμό. Εάν ο πληθυσμός είναι άρτιος αριθμός (ζυγός) τότε:

(a) Βρίσκουμε την μεσαία τιμή, δηλαδή το $\frac{\nu}{2}$

(b) Από την στήλη των αθροιστικών συχνοτήτων, στην οποία κάθε αριθμός δείχνει ποιά είναι η τελευταία παρατήρηση κατά αύξοντα αριθμό, βρίσκουμε την γραμμή στην οποία πρέπει να ανήκει το $\frac{\nu}{2}$, δηλαδή να είναι ίσο ή μικρότερο από το N_i της γραμμής αλλά μεγαλύτερο από αυτό της προηγούμενης. Την γραμμή στην οποία το βρήκαμε αυτό, παίρνουμε την τιμή x_i και αυτό είναι η μια παρατήρηση. Το ίδιο κάνουμε και για την αμέσως επόμενη παρατήρηση, βρίσκουμε και το δικό της x_i που μπορεί να είναι στην ίδια γραμμή, ή στην επόμενη. Από τις δυο αυτές παρατηρήσεις παίρνουμε τον μέσο όρο, αθροίζοντας τις και διαιρώντας με 2. Αυτό είναι η διάμεσος.

- Εάν ο πληθυσμός είναι περιττός αριθμός (μονός), τότε:

(a) Βρίσκουμε ποιά είναι η μεσαία παρατήρηση, υπολογίζοντας $\lfloor \frac{\nu}{2} \rfloor + 1$

(b) Εκτελούμε την ίδια διαδικασία, όπως πιο πάνω, για να βρούμε σε ποιά γραμμή αντιστοιχεί η συγκεκριμένη παρατήρηση μέσω της στήλης N_i , και όταν βρούμε ποιά γραμμή είναι, παίρνουμε την τιμή της x_i . Αυτή είναι η διάμεσος

3 Προβλήματα

1. Ρωτήσαμε 10 μέλη ενός γυμναστηρίου πόσες ώρες γυμνάζονται καθημερινά, και οι απαντήσεις που λάβαμε ήταν οι εξής: 2,1,3,3,2,1,3,4,5,1

- (a) Ποιές είναι οι παρατηρήσεις;
(b) Ποιά είναι η μεταβλητή που εξετάζουμε;
(c) Να κατασκευαστεί ο πίνακας συχνοτήτων με όλες τις στήλες

2. Να συμπληρώσετε τον παρακάτω πίνακα

x_i	ν_i	f_i	$f_i\%$	N_i	F_i	$F_i\%$
-1	3					
0	8					
2	5					
4	4					
ΣΥΝΟΛΟ						

3. Να συμπληρώσετε τον παρακάτω πίνακα

x_i	ν_i	f_i	$f_i\%$	N_i	F_i	$F_i\%$
3					0,2	
7	15	0.3	30			
9				45		
10						
ΣΥΝΟΛΟ						

4. Εξετάστηκαν 30 μαθητές ως προς την ομάδα αίματος και είχαμε τα παρακάτω αποτελέσματα: 0,0,AB,A,B,A,B,0,A,0,AB,0,B,0,A,B,A,0,A,B,A,0,AB,0,A,0,B,0,A,0

- (a) Να κατασκευαστεί ο πίνακας συχνοτήτων καθώς και οι στήλες με τις σχετικές συχνότητες
(b) Να κατασκευαστεί ραβδόγραμμα συχνοτήτων
(c) Να κατασκευαστεί κυκλικό διάγραμμα και να βρεθεί πόσες μοίρες αντιστοιχούν σε κάθε τομέα

5. Δίνεται ο παρακάτω πίνακας

x_i	ν_i	$f_i\%$	$F_i\%$	N_i
15	32	40		
17	22	27,5		
18	16	20		
20	10	12,5		
ΣΥΝΟΛΟ				

- (a) Να συμπληρωθεί ο πίνακας
- (b) Να κατασκευαστεί το διάγραμμα συχνοτήτων και το αντίστοιχο πολύγωνο
- (c) Να κατασκευάσετε το πολύγωνο των αθροιστικών σχετικών συχνοτήτων επι τοις %.
- (d) Ποιό ποσοστό των παρατηρήσεων είναι το πολύ 17;

6. Δίδεται ο παρακάτω πίνακας

x_i	ν_i	f_i	$f_i\%$
x_1			10
x_2			
x_3	8	0,2	
x_4		0,35	
ΣΥΝΟΛΟ		1	100

7. Σε ένα κυκλικό διάγραμμα παριστάνεται ο αριθμός των θηραμάτων που πέτυχε ένας κυνηγός στην διάρκεια της κυνηγετικής περιόδου.

- Το 20% των θηραμάτων ήταν τρυγόνια
- 75 θηράματα ήταν τσίχλες
- Η γωνία του κυκλικού τομέα που αντιστοιχεί στις πάπιες, είναι ίση με αυτή που αντιστοιχεί στα τρυγόνια
- Το εμβαδόν του κυκλικού τομέα που αντιστοιχεί στους λαγούς είναι τετραπλάσιο από αυτό που αντιστοιχεί στα αγριογούρουνα
- Η γωνία που αντιστοιχεί στις τσίχλες είναι ίση με 180°

- (a) Να βρείτε πόσα θηράματα πέτυχε ο κυνηγός
- (b) Να βρείτε πόσα τρυγόνια και πόσες πάπιες πέτυχε ο κυνηγός
- (c) Να κατασκευάσετε το κυκλικό διάγραμμα σχετικών συχνοτήτων επι τοις εκατό

8. Σε ένα κυκλικό διάγραμμα συχνοτήτων παρουσιάζονται οι βαθμοί που είχαν οι συμμετέχοντες (4,5,6,7)

- Η γωνία του κυκλικού τομέα που αντιστοιχεί στον βαθμό 4, ισούται με 36^0
- Το $\frac{1}{4}$ των υποψήφιων έγραψε 5
- 80 υποψήφιοι είχαν βαθμό 6
- Οι υποψήφιοι που έγραψαν 7 είναι τετραπλάσιοι από αυτούς που έγραψαν 8

(a) Να κατασκευάσετε πίνακα συχνοτήτων με τους βαθμούς που έγραψαν

(b) Εάν για να πετύχεις στις εξετάσεις πρέπει να είσαι στο άνω 25%, ποιος είναι ο ελάχιστος βαθμός που πρέπει να γράψεις;