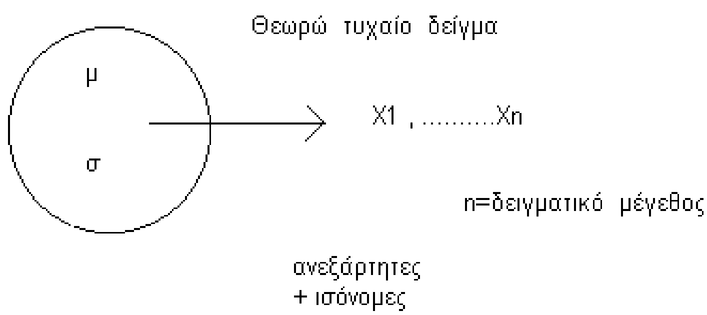
LECTURE 17:

Περιγραφική Στατιστική

Ο πληθυσμός :



Με βάση το δείγμα , μπορώ να «εκτιμήσω» τις άγνωστες ποσότητες του πληθυσμού ;

Θα ορίσουμε διάφορες <u>στατιστικές συναρτήσεις</u> (συναρτήσεις του δείγματος) , οι οποίες εκτιμούν τις αντίστοιχες πληθυσμιακές .

(1) <u>Επικρατούσα τιμή (mode)</u>

Είναι εκείνη η τιμή , η οποία παρατηρείται πιο συχνά σε ένα σύνολο αριθμών .

(2) <u>Δειγματική Διάμεσος</u>

Είναι η τιμή του δείγματος που βρίσκεται στην μέση , όταν αυτό είναι διατεταγμένο .

(*) Αν έχουμε περιττό αριθμό παρατηρήσεων , η δειγματική διάμεσος βρίσκεται στην μέση του διατεταγμένου δείγματος .

(**) Αν έχουμε άρτιο αριθμό παρατηρήσεων , η δειγματική διάμεσος είναι ο μέσος όρος των παρατηρήσεων που βρίσκονται στην μέση του διατεταγμένου δείγματος .

π.χ. (α) -3,-2,-1,1,3 n = 5 διάμεσος = -1

(β) -3,-2,-1,2,3,4 n = 6 διάμεσος= $\frac{-1+2}{2} = \frac{1}{2}$

(3) Μέσος όρος

Ο μέσος όρος των παρατηρήσεων (κέντρο βαρύτητας των παρατηρήσεων)

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Είναι γεγονός ότι ο μέσος όρος επηρεάζεται από μεγάλες τιμές ενός συνόλου δεδομένων .

π.χ. Οι παρακάτω αποτελούν τιμές σπιτιών :

67,000	105,000	148,000	5,260,000
91,000	116,000	167,000	
95,000	122,000	189,000	

διάμεσος = $\frac{116,000 + 122,000}{2}$ = 119,000

$\overline{x}$ = 635.000

(4) Δειγματικά ποσοστημόρια

Είναι εκείνα τα σημεία του δείγματος , τα οποία χωρίζουν την συλλογή δεδομένων στα 100 .

25^o	ποσοστημόριο	λέγεται	1^o τέταρτο(Q1)
50^o	»	»	διάμεσος
75^o	»	»	3^o τέταρτο(Q3)

<u>Πώς υπολογίζουμε ποσοστημόρια :</u>

Διατάσσουμε την συλλογή δεδομένων κατά αύξουσα σειρά .

Η θέση του ποσοστημορίου είναι το σημείο I , όπου $I = \frac{P}{100} \cdot n$

p =ποσοστημόριο n = δειγματικό μέγεθος

(α) Αν το i είναι ακέραιος , τότε το ποσοστημόριο βρίσκεται στον μέσο όρο των αριθμών που είναι στην i και στην i+1 θέση.

(β) Αν το i δεν είναι ακέραιος , τότε το ποσοστημόριο βρίσκεται στην θέση ακέραιο μέρος του (i+1) .

π.χ. 30,32,33,33,34,35,42,50

n = 8

- 30^o ποσοστημόριο : p = 30

  το σημείο $i = \frac{30}{100} \cdot 8 = 2,4 \rightarrow 2,4+1=3,4$

  στην θεση 3 : 33.

- 75^{o} ποσοστημόριο : $p = 75$

  το σημείο $i = \frac{75}{100} \cdot 8 = 6 \rightarrow$ το μέσο όρο των αριθμών

  στις θέσεις 6,7 : $\frac{35+42}{2} = 38{,}5$

Μέτρα Διασποράς

(1) Δειγματικό Εύρος

$R = \max X_i - \min X_i$ (το εύρος των παρατηρήσεων)

(2) Δειγματική Διακύμανση (μετρά την πληθυσμιακή διασπορά)

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})$$

$x_1 - \bar{x}, \ldots\ldots\ldots\ldots x_n - \bar{x} \longrightarrow$ αποκλίσεις απο $\bar{\chi}$

(n - 1) βαθμοί ελευθερίας

$$\sum_{i=1}^{n} (\chi_i - \bar{\chi}) = \sum_{i=1} \chi_i - \sum_{i=1} \bar{\chi} = \sum_{i=1} \chi_i - \bar{\chi} \cdot n = 0$$

(3) Δειγματική Τυπική Απόκλιση

$$S = \sqrt{S^2}$$

Εμπειρικός Κανόνας

ΓΙΑ ΣΥΜΜΕΤΡΙΚΕΣ ΣΥΛΛΟΓΕΣ ΔΕΔΟΜΕΝΩΝ

$(X - S, \quad X+S)$ 68% των δεδομένων

$(X - 2S, X+2S)$ 95% »

$(X-3S, X+3S)$ 99,7% »

π.χ. για κάποια συμμετρική συλλογή δεδομένων

$$\chi = 16{,}32 \qquad S = 4{,}7$$

$(\chi-2S, X+2S) = (6.92, 25.77)$

↓

→ 95% των δεδομένων