Τμήμα Επιστήμης Φυσικής Αγωγής & Αθλητισμού

Δημοκρίτειο Πανεπιστήμιο Θράκης

Στατιστική
στη Φυσική Αγωγή

(N161)

4η Διάλεξη

Κοκκότης Χρήστος, PhD

Στόχοι Διάλεξης

Κατανόηση βασικών εννοιών:

- Εκτίμηση παραμέτρων

Για την περιγραφή μιας μεταβλητής, που μετριέται σε έναν πληθυσμό ή σε ένα δείγμα, απαιτούνται τρία στοιχεία της:

- η κεντρική τιμή τάσης της, όπως π.χ. ο αριθμητικός μέσος
- η τιμή της διασποράς της , όπως π.χ. η τυπική απόκλιση
- το είδος της κατανομής της, όπως π.χ. η κανονική κατανομή.

Σε μια έρευνα από τα στοιχεία του δείγματος

στατιστικά:

- μέσος όρος δείγματος $\bar{X}$
- τυπική απόκλιση δείγματος s

προσπαθούμε να βγάλουμε συμπεράσματα (να προσδιορίσουμε) για τα στοιχεία του πληθυσμού

παράμετροι:

- μέσος όρος πληθυσμού μ
- τυπική απόκλιση πληθυσμού σ

Ερώτημα:

Μπορούμε να προσδιορίσουμε (**με ακρίβεια** !)

τον αριθμητικό μέσο ενός πληθυσμού (μ)

από τον αντίστοιχο αριθμητικό μέσο του δείγματος ($\bar{X}$)

ΟΧΙ!!!

Διαφορετικά δείγματα

οδηγούν

σε **διαφορετικές εκτιμήσεις**

των παραμέτρων του πληθυσμού

Υπάρχει δηλαδή πάντα **μια απόκλιση, ένα σφάλμα.**

Ερώτημα:

Πόσο είναι το σφάλμα δειγματοληψίας;

Δηλαδή, πόσο κοντά βρίσκεται ο μέσος όρος του δείγματος ($\bar{\mathbf{X}}$) στο μέσο όρο του πληθυσμού (μ);

$$\bar{\mathbf{X}} - \boldsymbol{\mu} = ;$$

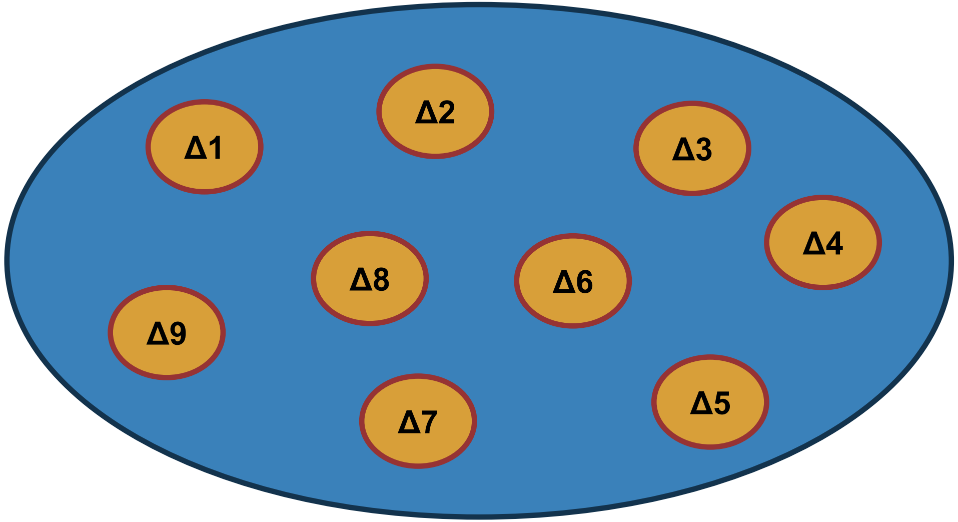
Από έναν πληθυσμό πάρουμε **πολλά δείγματα ίδιου μεγέθους N** και υπολογίσουμε ένα «στατιστικό»,

π.χ. τον **αριθμητικό μέσο του κάθε δείγματος**, που καλείται δειγματικός μέσος, προκύπτει

η κατανομή των στατιστικών των δειγμάτων,

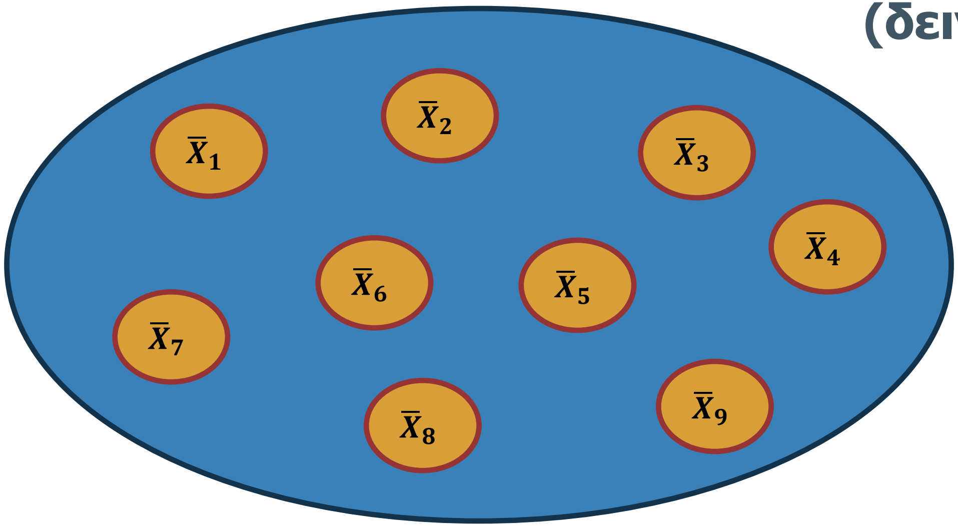
δηλαδή η **κατανομή των μέσων όρων των δειγμάτων**, η οποία καλείται δειγματική κατανομή .

Πληθυσμός



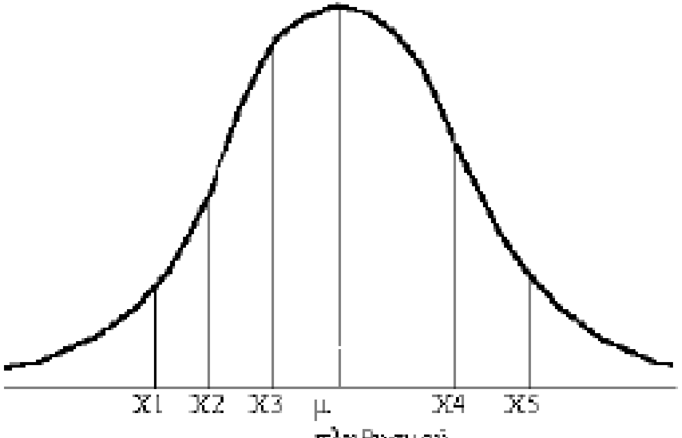
Πολλά δείγματα ίδιου μεγέθους Ν

Πληθυσμός



αριθμητικός μέσος του κάθε δείγματος **(δειγματικός μέσος)**

κατανομή των μέσων όρων των δειγμάτων **(δειγματική κατανομή)**



Παράδειγμα:

Έστω ότι ένας πληθυσμός αποτελείται από τους αριθμούς 2, 3 και 4.

α/α	X_i
1	2
2	3
3	4
N=3	$\sum_{i=1}^{3} X_i = 9$

Αριθμητικός μέσος πληθυσμού: $\mu = \dfrac{\sum_{i=1}^{3} X_i}{N} = \dfrac{9}{3} = 3$

Διακύμανση πληθυσμού: $\sigma^2 = \dfrac{\sum_{i=1}^{3} (X_i - \mu)^2}{N}$

Παράδειγμα:

Έστω ότι ένας πληθυσμός αποτελείται από τους αριθμούς 2, 3 και 4.

α/α	X_i	X_i - μ	$(X_i - \mu)^2$
1	2	-1	1
2	3	0	0
3	4	+1	1
N=3	$\sum_{i=1}^{3} X_i = 9$		$\sum_{i=1}^{3} (X_i - \mu)^2 = 2$

Αριθμητικός μέσος πληθυσμού: $\mu = \dfrac{\sum_{i=1}^{3} X_i}{N} = \dfrac{9}{3} = 3$

Διακύμανση πληθυσμού: $\sigma^2 = \dfrac{\sum_{i=1}^{3} (X_i - \mu)^2}{N} = \dfrac{2}{3} = 0.66$

Παράδειγμα:

Έστω ότι ένας πληθυσμός αποτελείται από τους αριθμούς 2, 3 και 4.

α/α	X_i	X_i - μ	$(X_i - \mu)^2$
1	2	-1	1
2	3	0	0
3	4	+1	1
N=3	$\sum_{i=1}^{3} X_i = 9$		$\sum_{i=1}^{3} (X_i - \mu)^2 = 2$

Τυπική απόκλιση πληθυσμού: $\sigma = \sqrt{\frac{\sum_{i=1}^{3}(X_i - \mu)^2}{N}} = \sqrt{\frac{2}{3}} = 0.817$

Δείγμα	$\overline{X}$	s^2	s
(2,2)	2.0	0.0	0.00
(2,3)	2.5	0.5	0.71
(2,4)	3.0	2.0	1.41
(3,4)	3.5	0.5	0.71
(3,2)	2.5	0.5	0.71
(3 3)	3.0	0.0	0.00
(4,2)	3.0	2.0	1.41
(4,3)	3.5	0.5	0.71
(4,4)	4.0	0.0	0.00
	$\sum_{i=1}^{9} \overline{X}_i = 27$	$\sum_{i=1}^{9} s^2 = 6.0$	$\sum_{i=1}^{9} s = 5.66$

Μέση τιμή των δειγματικών μέσων: $\overline{\overline{X}} = \frac{\sum_{i=1}^{9} \overline{X}_i}{N} = \frac{27}{9} = 3$

Μέση τιμή των διακυμάνσεων των δειγμάτων:

$$\overline{s^2} = \frac{\sum_{i=1}^{9} s^2}{N} = \frac{6}{9} = 0.66$$

Μέση τιμή των τυπικών αποκλίσεων των δειγμάτων:

$$\bar{s} = \frac{\sum_{i=1}^{9} s}{N} = \frac{5.66}{9} = 0.628$$

Αριθμητικός μέσος πληθυσμού:

$$\mu = \frac{\sum_{i=1}^{3} X_i}{N} = \frac{9}{3} = 3$$

Μέση τιμή των δειγματικών μέσων:

$$\bar{X} = \frac{\sum_{i=1}^{9} \bar{X}_i}{N} = \frac{27}{9} = 3$$

ο δειγματικός μέσος είναι ένας **αμερόληπτος εκτιμητής**

του μέσου όρου του πληθυσμού

ένας συγκεκριμένος δειγματικός μέσος

ΔΕΝ συμπίπτει υποχρεωτικά

με τον αριθμητικό μέσο του πληθυσμού

Όμως

αν επαναλάβουμε τη δειγματοληψία πολλές φορές

(με τις ίδιες συνθήκες)

η αναμενόμενη τιμή **του μέσου όρου των δειγματικών μέσων**

θα συμπίπτει

με τον αριθμητικό μέσο του πληθυσμού

Διακύμανση πληθυσμού:

$$\sigma^2 = \frac{\sum_{i=1}^{3}(X_i - \mu)^2}{N} = \frac{2}{3} = 0.66$$

Μέση τιμή των διακυμάνσεων των δειγμάτων:

$$\overline{s^2} = \frac{\sum_{i=1}^{9} s^2}{N} = \frac{6}{9} = 0.66$$

η δειγματική διακύμανση είναι ένας **αμερόληπτος εκτιμητής**
της διακύμανσης του πληθυσμού

αν επαναλάβουμε τη δειγματοληψία πολλές φορές
(με τις ίδιες συνθήκες)
η αναμενόμενη τιμή **του μέσου όρου των διακυμάνσεων των δειγμάτων**
θα συμπίπτει
με τη **διακύμανση του πληθυσμού**

Τυπική απόκλιση πληθυσμού:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{3}(X_i - \mu)^2}{N}} = \sqrt{\frac{2}{3}} = 0.817$$

Μέση τιμή των τυπικών αποκλίσεων των δειγμάτων:

$$\bar{s} = \frac{\sum_{i=1}^{9} s}{N} = \frac{5.66}{9} = 0.628$$

Η δειγματική τυπική απόκλιση είναι ένας **μερόληπτος εκτιμητής** της τυπικής απόκλισης του πληθυσμού

Δειγματική κατανομή του δειγματικού μέσου

Σ΄ ότι αφορά

τη **δυνατότητα εκτίμησης του αριθμητικού μέσου ενός πληθυσμού**

από τον αριθμητικό μέσο ενός δείγματος,

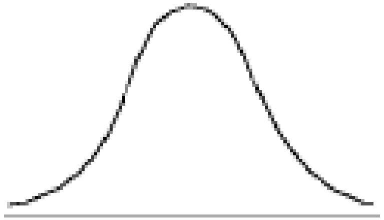
που προέρχεται από τον συγκεκριμένο πληθυσμό,

υπάρχει στην θεωρία της Στατιστικής ένα πολύ σημαντικό θεώρημα,

το «**Κεντρικό Οριακό Θεώρημα**».

Κεντρικό Οριακό Θεώρημα

Αν ένας πληθυσμός ακολουθεί την κανονική κατανομή,

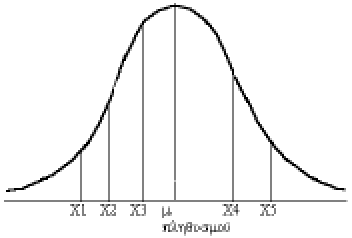


τότε

οι **δειγματικοί αριθμητικοί μέσοι τυχαίων δειγμάτων, ίσου μεγέθους, ακολουθούν και αυτοί την κανονική κατανομή**

γύρω από τον αριθμητικό μέσο του πληθυσμού,

ανεξάρτητα από το μέγεθος των δειγμάτων.



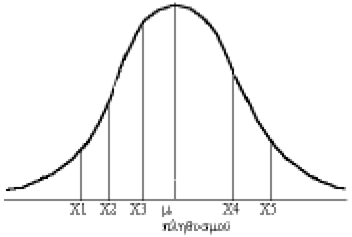
Κεντρικό Οριακό Θεώρημα

Αν ένας πληθυσμός ΔΕΝ ακολουθεί την κανονική κατανομή,

αλλά το μέγεθος των δειγμάτων **είναι μεγάλο** (N>30),

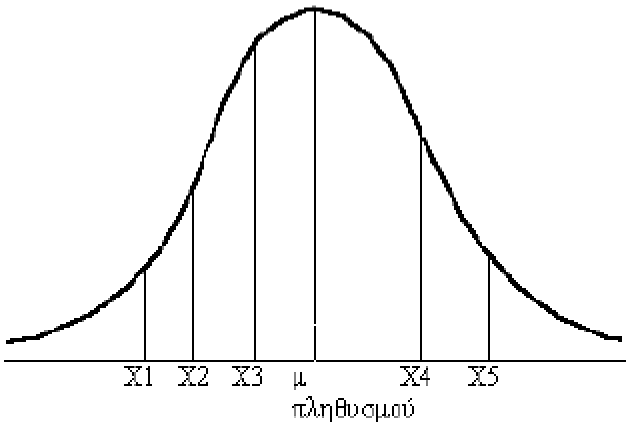
τότε

οι **δειγματικοί αριθμητικοί μέσοι τυχαίων δειγμάτων, ίσου μεγέθους, ακολουθούν και αυτοί την κανονική κατανομή** γύρω από τον αριθμητικό μέσο του πληθυσμού.



Κεντρικό Οριακό Θεώρημα

Ο αριθμητικός μέσος των δειγματικών μέσων, δηλαδή ο αριθμητικός μέσος των μέσων όρων των επιμέρους δειγμάτων, θα είναι ίσος με τον αριθμητικό μέσο του πληθυσμού.



Σύμφωνα με το Κ.Ο.Θ.

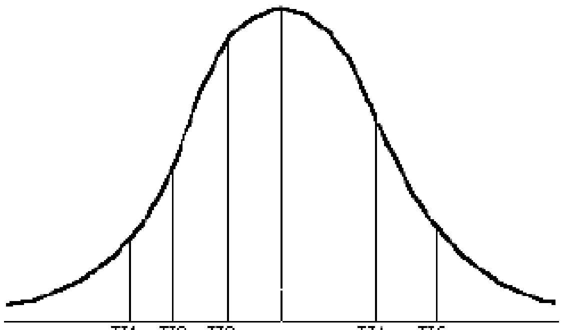
αν επιλέξουμε ένα δείγμα (μεγέθους Ν) από κάποιον πληθυσμό,

ο αριθμητικός μέσος του δείγματος (δειγματικός μέσος)

θα είναι

μια τιμή από τη δειγματική (θεωρητική) κατανομή

(δηλ. την κατανομή των μέσων όρων των τυχαία επιλεγμένων δειγμάτων)



είναι απίθανο

η (συγκεκριμένη) μέση τιμή του δείγματος να συμπίπτει με τη μέση τιμή του πληθυσμού

Πάντα θα υπάρχει κάποια "απόκλιση" (σφάλμα δειγματοληψίας)

Σφάλμα δειγματοληψίας = Τυπικό σφάλμα του μέσου όρου=

“πόσο απέχει ο μέσος όρος του δείγματος από τον πραγματικό μέσο όρο του πληθυσμού”

$$\bar{X} - \mu = ;$$

εκφράζεται από

την τυπική απόκλιση της κατανομής των δειγματικών μέσων

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}$$

$\sigma_{\bar{X}}$ = τυπική απόκλιση δείγματος

σ = τυπική απόκλιση πληθυσμού

N= μέγεθος δείγματος

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}}$$

- αν η διασπορά των τιμών του πληθυσμού είναι μεγάλη,

αν δηλαδή είναι μεγάλη η τυπική απόκλιση σ του πληθυσμού,

Τότε αναμένονται **μεγάλες αποκλίσεις** των δειγματικών μέσων από την μέση τιμή του πληθυσμού.

- όσο πιο μεγάλο είναι το μέγεθος του δείγματος (N),

τόσο πιο μικρή θα είναι και η διασπορά των δειγματικών μέσων, δηλαδή τόσο πιο κοντά στην μέση τιμή του πληθυσμού θα βρίσκονται οι δειγματικοί αριθμητικοί μέσοι.

Ερωτήσεις & Συζήτηση