



Δειγματοληπτική κατανομή¹
&
Διάστημα εμπιστοσύνης

Κωνσταντίνος Μπουγιούκας, MSc, PhD
mpougioukas@auth.gr

2025-2026



Στόχοι του σημερινού μαθήματος

- Ορισμός πληθυσμού, δείγματος
- Σφάλμα εκτίμησης
- Δειγματοληπτική Κατανομή μέσης τιμής
- Διάστημα Εμπιστοσύνης (ΔΕ) μέσης τιμής



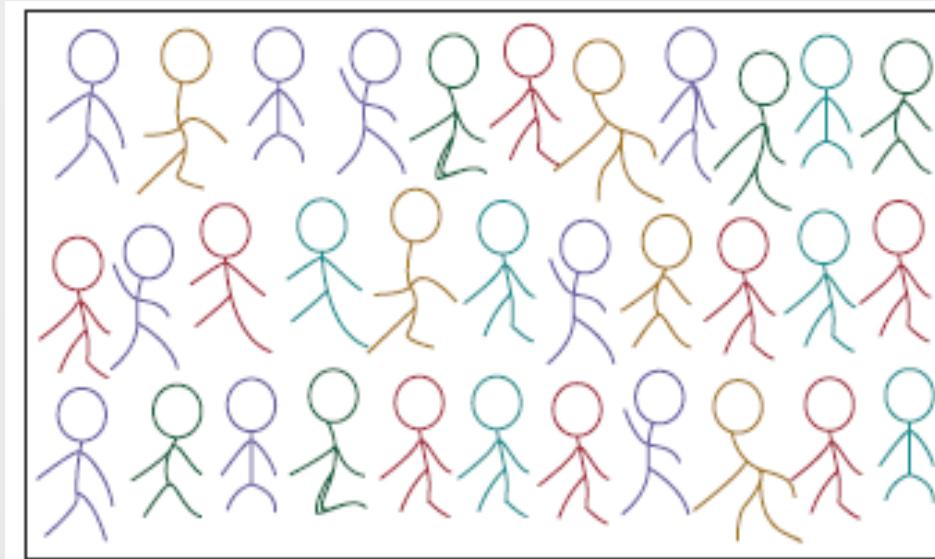


Πληθυσμός ορίζεται ως το πλήρες σύνολο των στατιστικών μονάδων (όχι απαραίτητα ανθρώπων) που μοιράζονται συγκεκριμένα χαρακτηριστικά που μας ενδιαφέρουν να μελετήσουμε.

- Όλοι οι ασθενείς με σακχαρώδη διαβήτη
- Όλα τα άτομα που έχουν διαγνωστεί με κατάθλιψη
- Όλες γυναίκες μέσης ηλικίας

Αλλά και:

- όλα τα σχολεία μιας περιφέρειας
- όλα τα νοσοκομεία σε μια περιοχή

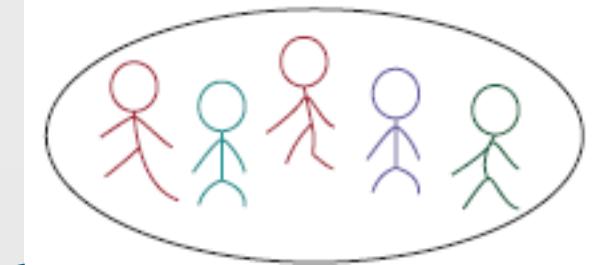


Πληθυσμός

**Μέθοδος
δειγματοληψίας**

(π.χ. απλή τυχαία
δειγματοληψία)

Αντιπροσωπευτικό ?



Δείγμα

Το Δείγμα αποτελεί ένα υποσύνολο του πληθυσμού που μελετάμε.

Το πλήθος των στατιστικών μονάδων από τα οποία αποτελείται το δείγμα ονομάζεται μέγεθος του δείγματος και συμβολίζεται συνήθως ως n .

Στατιστικό πλαίσιο για την εκτίμηση των παραμέτρων του πληθυσμού



Σφάλμα της εκτίμησης (estimation error)

Είναι η διαφορά μεταξύ της σημειακής εκτίμησης και της αντίστοιχης παραμέτρου του πληθυσμού (σφάλμα εκτίμησης = $\hat{\theta} - \theta$).

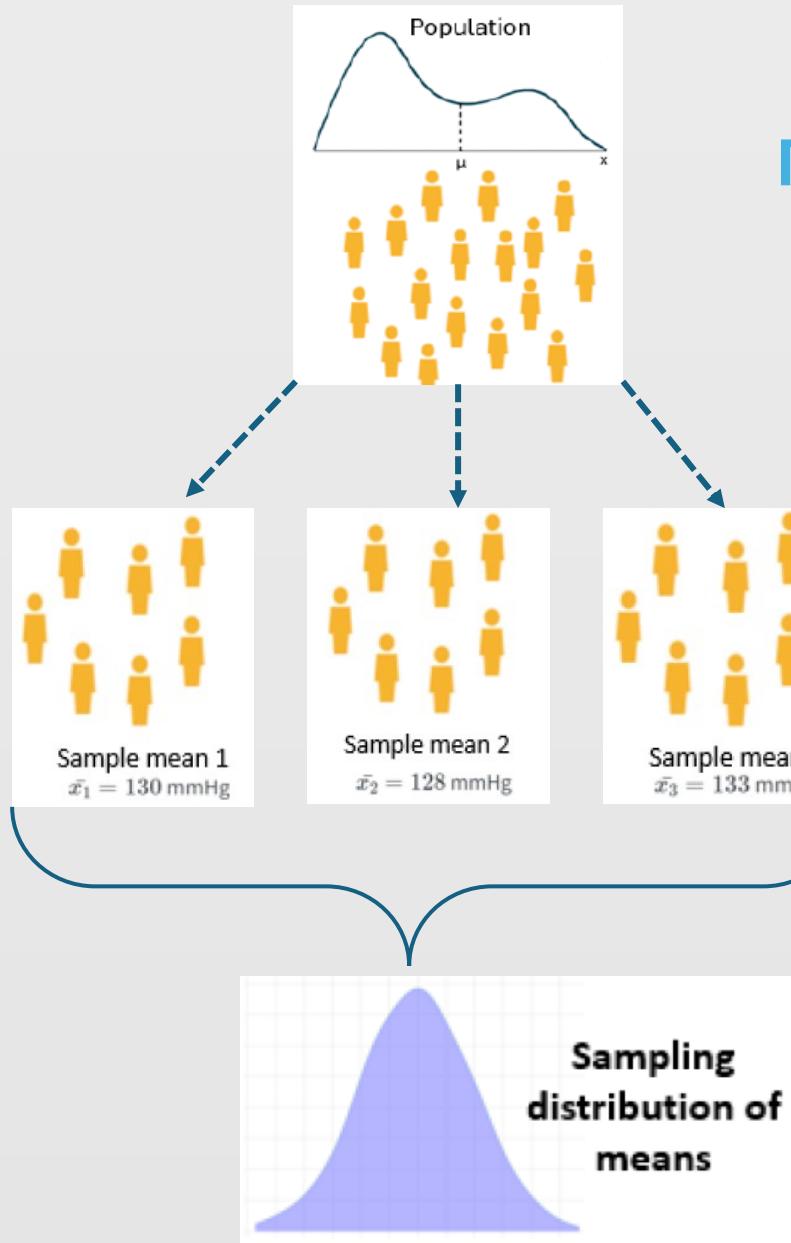
Παράδειγμα:

Αν η πραγματική μέση τιμή IQ στον πληθυσμό είναι $\mu = 100$ (δηλ. η παράμετρος θ που μας ενδιαφέρει) και η μέση τιμή IQ του δείγματος είναι $\bar{x} = 103$ (δηλ. η σημειακή εκτίμηση $\hat{\theta}$), τότε:

$$\text{Σφάλμα εκτίμησης} = \bar{x} - \mu = 103 - 100 = +3$$

Τύποι σφάλματος:

- **Σφάλμα δειγματοληψίας (sampling error):** είναι η διαφορά μεταξύ της εκτίμησης που προκύπτει από το δείγμα και της αντίστοιχης πραγματικής παραμέτρου του πληθυσμού, η οποία οφείλεται στο ότι χρησιμοποιούμε ένα υποσύνολο του πληθυσμού και όχι ολόκληρο τον πληθυσμό.
- **Μη-δειγματοληπτικό σφάλμα (non-sampling error):** Το σφάλμα που οφείλεται **σε πηγές που δεν ανήκουν στο σφάλμα δειγματοληψίας** (π.χ. σφάλμα επιλογής-selection bias, σφάλμα ανταπόκρισης-non-response bias κ.ά.).



Πληθυσμός

Μέση τιμή
για κάθε
δυνατό¹
δείγμα
μεγέθους n

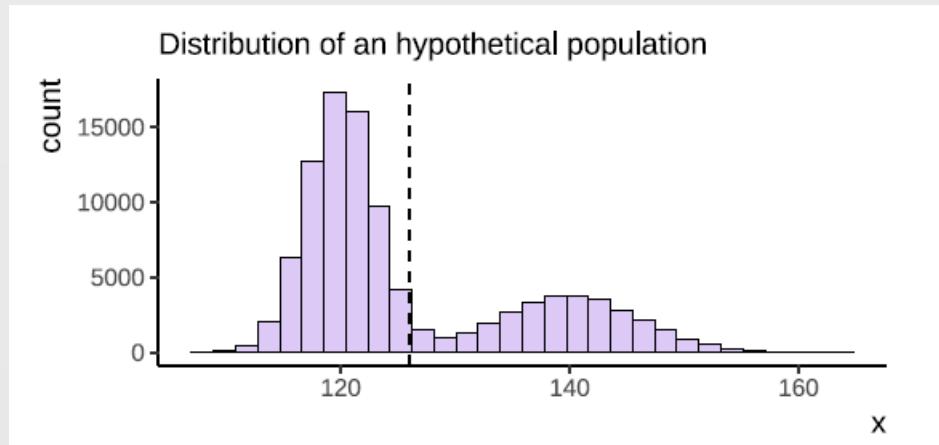
Δειγματοληπτική
Κατανομή μέσης τιμής

Τυπικό σφάλμα μέσης
τιμής (standard error of
mean, SEM)

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Είναι η τυπική απόκλιση της
δειγματοληπτικής κατανομής
της μέσης τιμής.

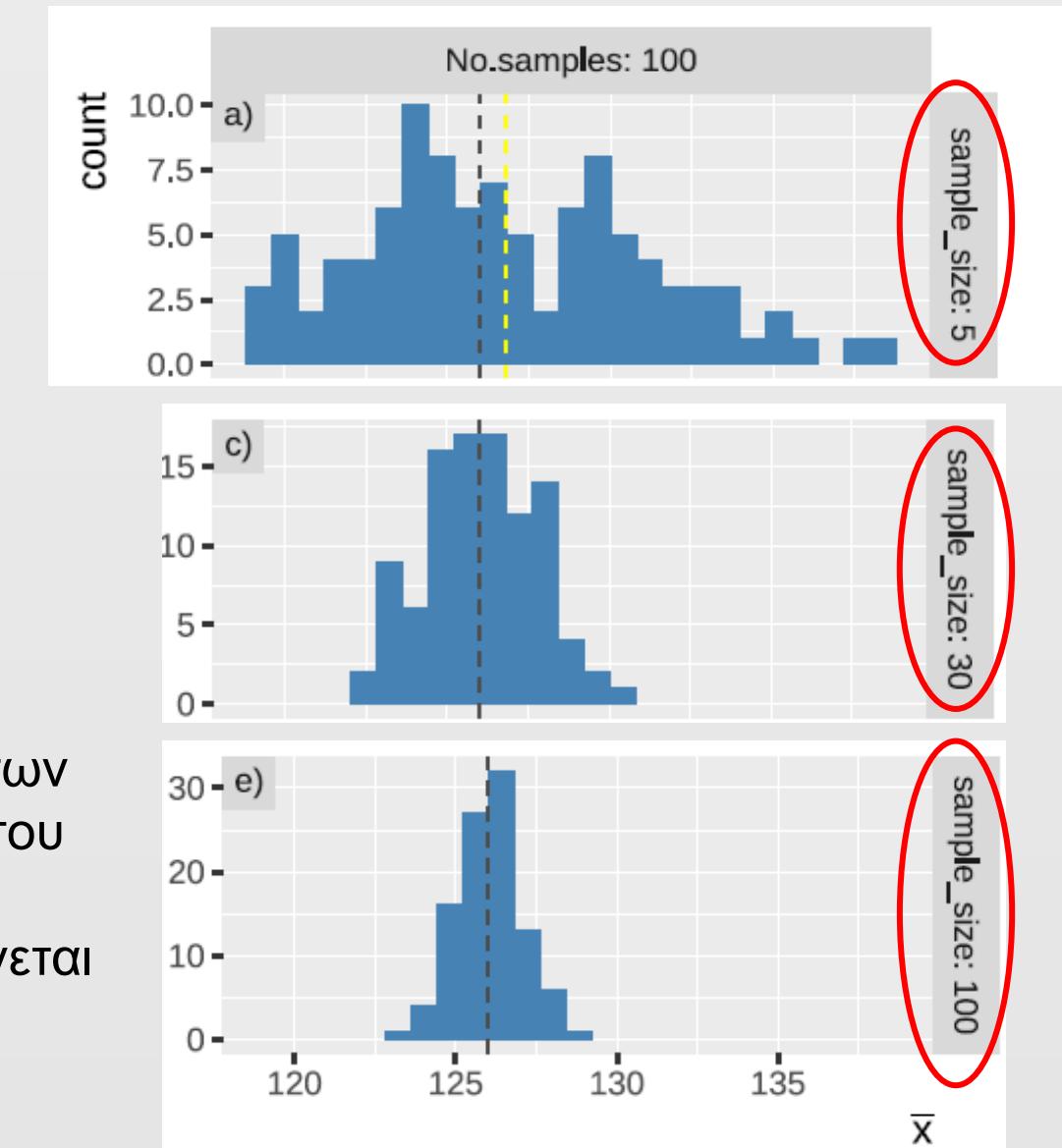
Πληθυσμός π.χ. με Συστολική Αρτηριακή Πίεση (ΣΑΠ), όπου $\mu=130$ και $\sigma=10$.



Ιδιότητες της Δειγματοληπτικής Κατανομής μέσης τιμής

- Καθώς αυξάνεται το μέγεθος δείγματος, η **μέση τιμή** των δειγματικών μέσων συγκλίνει με την πραγματική τιμή του πληθυσμού
- Καθώς αυξάνεται το μέγεθος δείγματος, το **SEM** μειώνεται
- Καθώς αυξάνεται το μέγεθος δείγματος, η κατανομή πλησιάζει την **κανονική** κατανομή

Δειγματοληπτική Κατανομή μέσης τιμής





Κεντρικό Οριακό Θεώρημα για την μέση τιμή

Η δειγματοληπτική κατανομή της μέσης τιμής όλων των δυνατών δειγμάτων του πληθυσμού είναι κατά προσέγγιση **κανονική** ανεξαρτήτως της κατανομής του πληθυσμού, εφόσον το μέγεθος n των δειγμάτων είναι επαρκώς μεγάλο.

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

(τυπική απόκλιση
δειγματοληπτικής κατανομής)

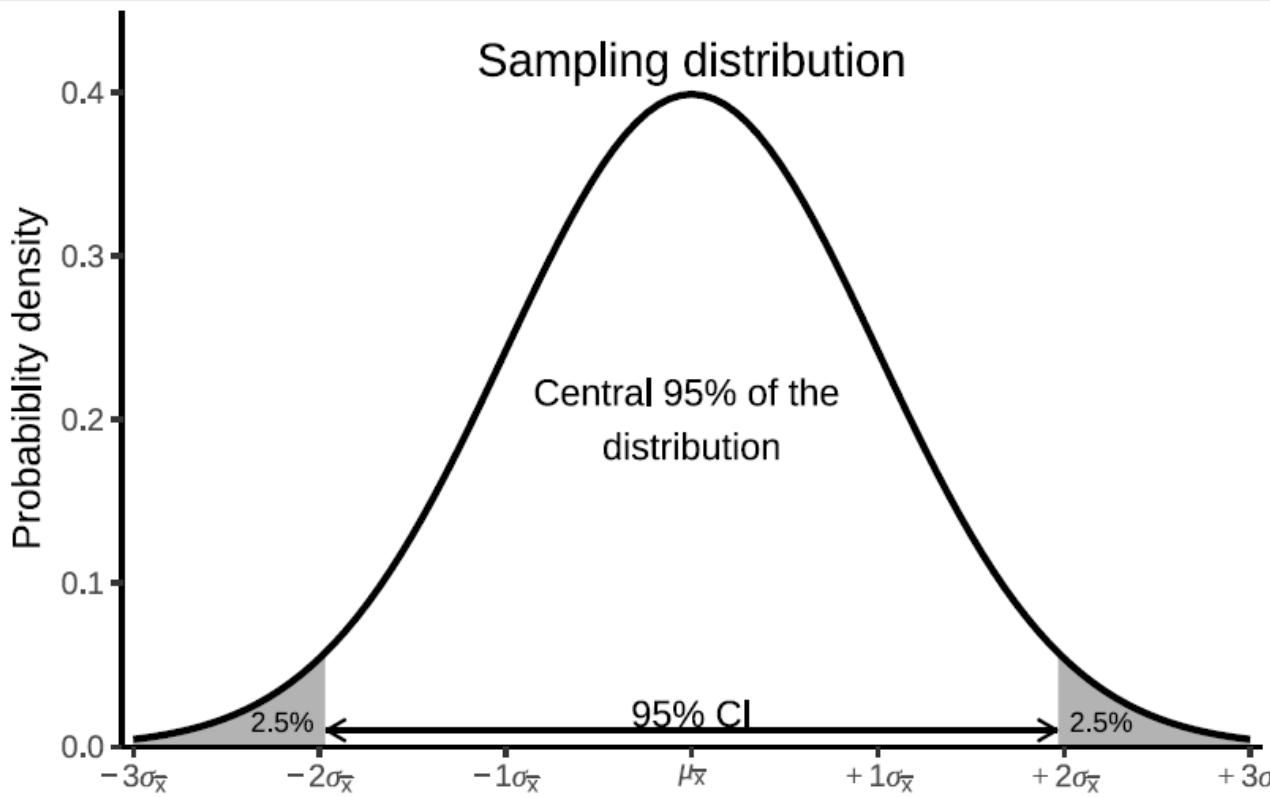
Πως θα κατασκευάσω ένα 95% διάστημα εμπιστοσύνης (ΔΕ) για την μέση τιμή;

1. Το ΔΕ θα είναι γύρω από την **εκτίμηση** για την μέση τιμή, που είναι η καλύτερη εκτίμηση για την μέση τιμή του πληθυσμού: $\mu_{\bar{x}} = \mu$
2. Το **τυπικό σφάλμα (SEM)** θα αποτελεί μέτρο της μεταβλητότητας των τιμών γύρω από την εκτίμηση:
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

(τυπική απόκλιση
δειγματοληπτικής κατανομής)
3. Χρήση του κανόνα 68-95-99.7 της κανονικής κατανομής

95% Διάστημα Εμπιστοσύνης (95% ΔΕ) μέσης τιμής

Αγγλική ορολογία: 95% Confidence Interval (95% CI)



Εκτίμηση μέσης τιμής $\pm 2 \cdot$ τυπική απόκλιση δειγματολ. κατανομής

$$\mu_{\bar{x}}$$

1.96
(πιο ακριβές)

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$95\% CI : \mu_{\bar{x}} \pm 1.96 \sigma_{\bar{x}} = \mu_{\bar{x}} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$



95% Διάστημα Εμπιστοσύνης (95% ΔΕ) μέσης τιμής

Αγγλική ορολογία: 95% Confidence Interval (95% CI)

$$95\% CI : \mu_{\bar{X}} \pm 1.96 \sigma_{\bar{X}} = \mu_{\bar{X}} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

Στην πράξη όμως έχουμε ένα μόνο δείγμα!

Άρα:

$$\begin{array}{ccc} \mu_{\bar{x}} & \xrightarrow{\text{-----}} & \bar{x} \\ \sigma & \xrightarrow{\text{-----}} & s \end{array} \Rightarrow$$

$$95\% CI : \bar{x} \pm 1.96 \frac{s}{\sqrt{n}}$$

Κάτω και άνω όριο του ΔΕ



95% Διαστήματος Εμπιστοσύνης μέσης τιμής: υπολογιστικό παράδειγμα

Η μέση συστολική αρτηριακή πίεση (ΣΑΠ) ενός τυχαίου δείγματος 100 ατόμων βρέθηκε ίση με 125 mmHg με τυπική απόκλιση 12 mmHg. Το 95% διάστημα εμπιστοσύνης (ΔΕ) είναι:

$n = 100$

Δειγματική μέση τιμή, $\bar{x} = 125 \text{ mmHg}$

Δειγματική τυπική απόκλιση, $s = 12 \text{ mmHg}$

Τύπος του 95% ΔΕ μέσης τιμής

$$\bar{x} \pm 1.96 \cdot \frac{s}{\sqrt{n}}$$

Κάτω όριο
του 95% ΔΕ:

$$125 - 1.96 \cdot \frac{12}{\sqrt{100}} = 125 - 1.96 \cdot \frac{12}{10} = 125 - 2.352 = 122.6 \text{ mmHg}$$

Άνω όριο του
95% ΔΕ:

$$125 + 1.96 \cdot \frac{12}{\sqrt{100}} = 125 + 1.96 \cdot \frac{12}{10} = 125 + 2.352 = 127.4 \text{ mmHg}$$

Ερμηνεία του 95% διαστήματος εμπιστοσύνης μέσης τιμής

- Η πραγματική μέση τιμή του πληθυσμού (μ) θεωρείται μια σταθερή, συγκεκριμένη, αλλά άγνωστη τιμή.

Αν ένας μεγάλος αριθμός τυχαίων δειγμάτων λαμβάνονταν **επανειλημμένα** από τον πληθυσμό και κατασκευαζόταν ένα 95% διάστημα εμπιστοσύνης για κάθε δείγμα, περίπου το 95% αυτών των διαστημάτων θα περιείχε τη σταθερή, πραγματική μέση τιμή του πληθυσμού.

95% confidence intervals

