

Διαστήματα Εμπιστοσύνης

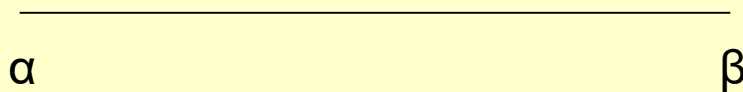
- α) για τη μέση τιμή (ποσοτικά)
- β) για ένα ποσοστό (ποιοτικά)

Τι είναι «διάστημα»

- Διάστημα στα μαθηματικά, είναι ένα σύνολο αριθμών, που βρίσκονται ανάμεσα σε ΔΥΟ αριθμούς, το α και το β ,

π.χ.

$[3, 8]$ ή $[12, 25]$ ή $(7, 9)$ κ.λ.π.



Διάστημα ΔΕΝ είναι η Απόσταση!!!

Διάστημα (πως γράφεται)

- Συνήθως σημειώνοντας τα άκρα, δηλαδή το κάτω και το πάνω όριο π.χ. (11 , 18)

δεν μας απασχολεί στη Στατιστική να διαφοροποιήσουμε Κλειστά ή Ανοικτά Διαστήματα

- Αν το διάστημα είναι συμμετρικό ως προς μια κεντρική τιμή, το γράφουμε π.χ. 5 ± 2 που είναι το διάστημα (3, 7)

αυτό σημαίνει ότι το κέντρο του διαστήματος είναι το 5

Πληθυσμός - Δείγμα

- Θυμίζουμε ότι για να μελετήσουμε έναν πληθυσμό, παίρνουμε ένα δείγμα n ατόμων
- Υπολογίζουμε τη μέση τιμή του δείγματος, και λέμε ότι αυτή εκτιμάει τη μέση τιμή του πληθυσμού που δεν την γνωρίζουμε

Τι είναι η «εμπιστοσύνη»

- Εκτιμώντας την μέση τιμή (μ) του πληθυσμού, παίρνουμε ένα δείγμα n μετρήσεων $x_1, x_2, \dots, x_n$ και υπολογίζουμε τη μέση τιμή του δείγματος $\bar{X}$.
- Όμως ΠΟΣΟ ΣΩΣΤΗ και ΠΟΣΟ ΑΚΡΙΒΗΣ είναι η εκτίμηση που κάνουμε για το μ , με τον τρόπο αυτό?
- Δεν γνωρίζουμε ούτε το ένα, ούτε το άλλο!!

Τι είναι η «εμπιστοσύνη»

- Πόσο σίγουροι είμαστε πως το $\bar{X}$ έχει εκτιμήσει σωστά το μ ? Ή αλλιώς, πόσο κοντά είναι το $\bar{X}$ στο μ ?
- ΔΕΝ ΜΠΟΡΟΥΜΕ να είμαστε «σίγουροι» 100% αν δεν μετρήσουμε ΟΛΟ τον πληθυσμό.
- Μπορούμε όμως να υπολογίσουμε το μέγιστο δυνατό λάθος (σφάλμα) που κάνουμε σε μια τέτοια εκτίμηση

Παράδειγμα

- Μετρώντας το βάρος 400 ανδρών μιας περιοχής, βρήκαμε μέση τιμή 75 κιλά.
- Αν αυτοί οι 400 είναι ένα δείγμα από τους άνδρες που κατοικούν στην περιοχή, τότε
- Πόσο σίγουροι είμαστε ότι αν μετράγαμε το βάρος ΟΛΩΝ των ανδρών, θα βγάzaμε ακριβώς 75 κιλά;
- **Φυσικά δεν μπορούμε να είμαστε ΚΑΘΟΛΟΥ σίγουροι για κάτι τέτοιο !!!**

Σημειακή εκτίμηση – Εκτίμηση Διαστήματος

Το να πούμε ότι προβλέπω (εκτιμώ) ότι ο μέσος όρος του πληθυσμού είναι 75 κιλά είναι μια σημειακή εκτίμηση

Αν δημιουργήσουμε ένα διάστημα (με κάποιον τύπο που θα δούμε), π.χ. από 72 κιλά μέχρι 78 κιλά δηλ. (72, 78)

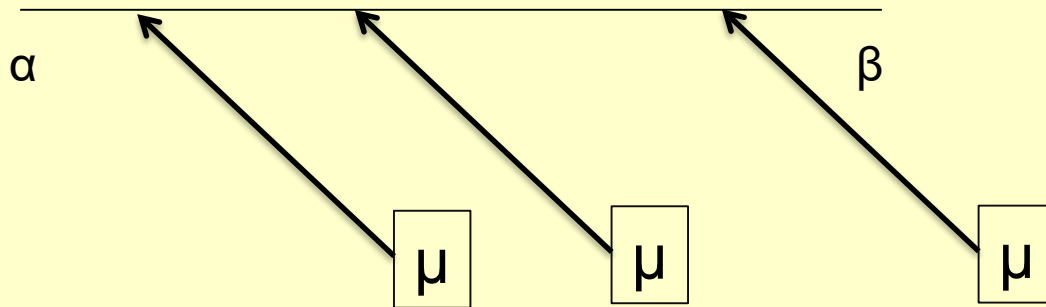
και πούμε ότι η μέση τιμή του πληθυσμού θα είναι εκεί ανάμεσα, τότε έχουμε εκτίμηση διαστήματος

Εκτίμηση με Διάστημα Εμπιστοσύνης

- Αν δημιουργήσουμε ένα τέτοιο διάστημα για την εκτίμηση μπορούμε να το ονομάζουμε Διάστημα Εμπιστοσύνης για τη μέση τιμή
- Θα λέμε τότε πως «έχουμε μεγάλη εμπιστοσύνη, πως το μέσο βάρος των ανδρών της περιοχής θα είναι κάπου ανάμεσα στα 72 και στα 78 κιλά»
- Μεγάλη $\rightarrow$ συνήθως 95% στην Στατιστική

Τι είναι λοιπόν το 95% Διάστημα εμπιστοσύνης για τη μέση τιμή;

- Είναι ένα διάστημα (α, β) όπου έχουμε 95% σιγουριά (εμπιστοσύνη), ότι η ΠΡΑΓΜΑΤΙΚΗ μέση τιμή του πληθυσμού βρίσκεται ανάμεσα στο α και στο β .

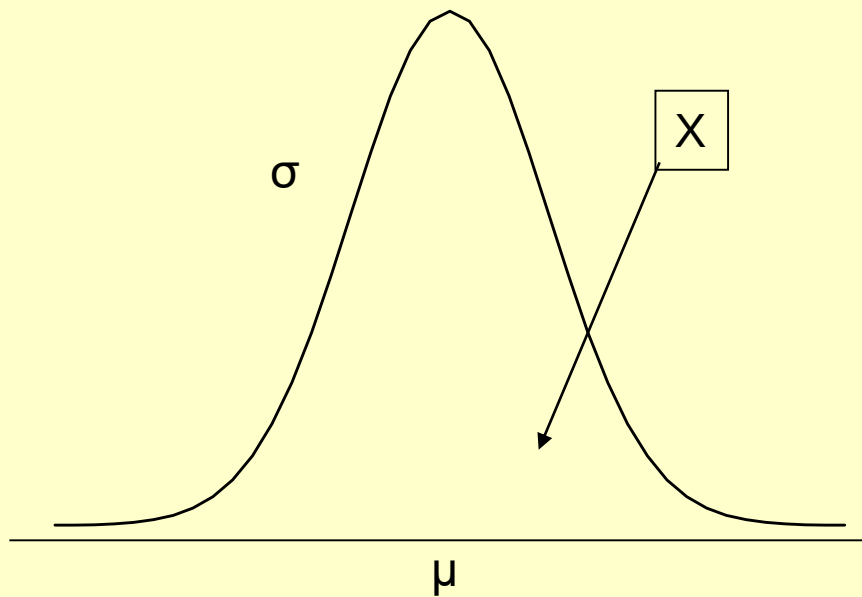


Πολλαπλά δείγματα – Κατανομή του μέσου

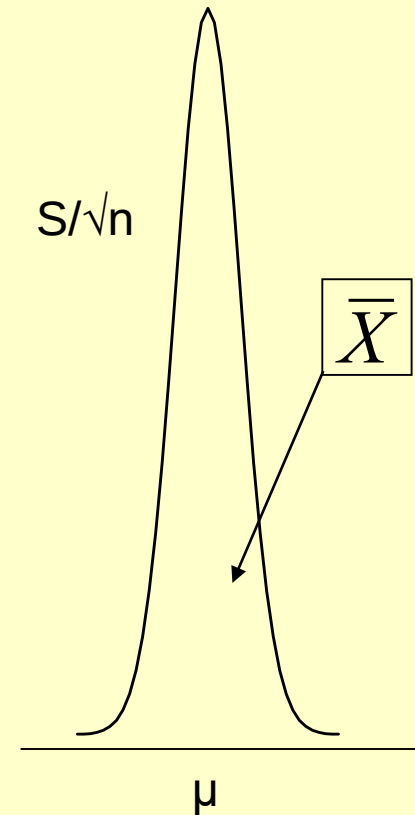
- Αν κάποιος άλλος πήρε πάλι δείγμα 400 ανδρών και βρήκε μέση τιμή 74 κιλά
- Ή κάποιος άλλος πήρε κάποιους άλλους 400 και βρήκε μέση τιμή 76,5 κιλά
-
- Πως μπορούμε να ξέρουμε ποιος έχει δίκιο ή ποιος έχει πέσει πιο κοντά στην πραγματική μέση τιμή του πληθυσμού;

Αν τελικά πάρουμε δεκάδες δείγματα και βρίσκουμε κάθε φορά τη μέση τιμή δημιουργείται ένα ιστόγραμμα και μια κατανομή για τους μέσους όρους

Θα παρατηρήσουμε ότι θα έχουμε μια συμμετρική κατανομή – κανονική!



Κατανομή του χαρακτηριστικού X



Το τυπικό σφάλμα της μέσης τιμής είναι η τυπική απόκλιση αυτής της κατανομής, που είναι ίση με $S/\sqrt{n}$

Κατανομή της μέσης τιμής του X , για συγκεκριμένο μέγεθος δείγματος n .

Τυπικό Σφάλμα της μέσης τιμής – standard error

$$SE(\bar{X}) = \frac{S}{\sqrt{n}}$$


Συμβολισμός για το τυπικό σφάλμα

Με λόγια (αν δεν σας αρέσουν τα ...μαθηματικά)
Το τυπικό σφάλμα είναι η τυπική απόκλιση αφού
διαιρεθεί με την τετραγωνική ρίζα του n

95% Διάστημα Εμπιστοσύνης για τη μέση τιμή (μ)
--- προσεγγιστικό, για μεγάλα δείγματα ---

$$\bar{X} \pm 2 \cdot SE(\bar{X})$$

$$(\bar{X} - 2 \cdot SE(\bar{X}), \bar{X} + 2 \cdot SE(\bar{X}))$$

Με λόγια (αν δεν σας αρέσουν τα ...μαθηματικά)

Στη μέση τιμή προσθέτουμε

2 φορές το τυπικό σφάλμα και βγάζουμε το πάνω άκρο,

Στη μέση τιμή αφαιρούμε

2 φορές το τυπικό σφάλμα και βρίσκουμε το κάτω άκρο

Ερμηνεία (τι σημαίνει 95% ΔΕ)

Έχουμε 95% Εμπιστοσύνη, ότι η πραγματική μέση τιμή (μ) του πληθυσμού βρίσκεται ανάμεσα σε αυτά τα όρια

Αν δούμε σε κάποια έρευνα, ότι π.χ. το 95% ΔΕ για τη μέση τιμή του βάρους ήταν (72,78) τότε είμαστε 95% σίγουροι, ότι η πραγματική μέση τιμή του βάρους του πληθυσμού που μελετήσαμε βρίσκεται ανάμεσα σε 72 και 78 κιλά.

Παράδειγμα - Άσκηση

- Αν η μέση τιμή ενός δείγματος με 60 μετρήσεις, βρέθηκε 7,2 και η τυπική απόκλιση βρέθηκε 1,6, τότε

Η ρίζα του 60 είναι 7,75 (παρανομαστής)

Και το τυπικό σφάλμα της μέσης τιμής είναι η τυπική απόκλιση δια τη ρίζα του 60

- Δηλαδή είναι $SE = 1,6/7,72$
- Άρα $SE=0,206$

....αν σας αρέσουν οι μαθηματικοί
τύποι.....

$$SE(\bar{X}) = \frac{1,6}{\sqrt{60}} = \frac{1,6}{7,75} = 0,206$$

Οι 0,206 μονάδες σφάλματος πολλαπλασιάζονται
με το 2 για να προστεθούν και να αφαιρεθούν από
τη μέση τιμή.....

Οπότε το 95% διάστημα εμπιστοσύνης για τη μέση τιμή γίνεται

$$7,2 \pm 2 \cdot 0,206 = 7,2 \pm 0,412$$

δηλ. έχουμε 95% εμπιστοσύνη ότι η πραγματική μέση τιμή του πληθυσμού που μελετήσαμε είναι μεταξύ

6,788 και 7,612

Από τι εξαρτάται το μέγεθος του Δ.Ε. για τη μέση τιμή?

1. Από το μέγεθος δείγματος n . Όσο πιο μεγάλο είναι το n , τόσο πιο μικρό είναι το Δ.Ε., άρα και η εκτίμηση της μέσης τιμής είναι καλύτερη
2. Από την τυπική απόκλιση S . Όσο πιο μικρή είναι η τυπική απόκλιση, τόσο πιο μικρό είναι το Δ.Ε., άρα και η εκτίμηση της μέσης τιμής είναι καλύτερη

S ή SE ?

- Η τυπική απόκλιση (S) αναφέρεται στο χαρακτηριστικό που μετρήσαμε (στο δείγμα ή σε ολόκληρο τον πληθυσμό)
- Το τυπικό σφάλμα (SE) αναφέρεται στη μέση τιμή του χαρακτηριστικού. Μας δείχνει πόσο κοντά στην μέση τιμή του πληθυσμού (μ), μπορεί να βρίσκεται η εκτίμησή μας. με άλλα λόγια μας δείχνει πόσο καλή εκτίμηση μπορούμε να έχουμε κάνει.
- Σε πολλά επιστημονικά άρθρα παρουσιάζεται η μέση τιμή και η τυπική απόκλιση, ενώ σε άλλα παρουσιάζεται η μέση τιμή με το τυπικό της σφάλμα.

Όταν έχουμε ποιοτικά χαρακτηριστικά;

Εδώ δεν υπάρχουν μέσες τιμές ή τυπικές αποκλίσεις!

Εκτιμάμε μόνο **ΠΟΣΟΣΤΑ** στον πληθυσμό

Π.χ. αν έχουμε μια ερώτηση με τρεις
δυνατές απαντήσεις A, B ή Γ

Εκτιμάμε ποσοστό για το A, ποσοστό για το
B και ποσοστό για το Γ

Διάστημα Εμπιστοσύνης για ένα ποσοστό

- Ένα ποσοστό σε ένα πληθυσμό, εκτιμάται από το δείγμα με τη «σχετική συχνότητα», που συνήθως συμβολίζουμε με f ή με π

$$\pi = \frac{n_i}{n}$$

Δεν πολλαπλασιάζω με 100%
Το αφήνω στην κλίμακα 0-1

- Πόσο καλή είναι η εκτίμηση ενός ποσοστού, εξαρτάται από το τυπικό σφάλμα του ποσοστού

$$SE(\pi) = \sqrt{\frac{\pi \cdot (1 - \pi)}{n}}$$

Συμβολισμός για το τυπικό σφάλμα

Διάστημα Εμπιστοσύνης για μια πιθανότητα ή ένα ποσοστό (π)

--- προσεγγιστικό, για μεγάλα δείγματα ---

$$\pi \pm 2 \cdot SE(\pi)$$

$$(\pi - 2 \cdot SE(\pi), \pi + 2 \cdot SE(\pi))$$

ΕΡΜΗΝΕΙΑ:

Έχουμε 95% Εμπιστοσύνη, ότι το πραγματικό ποσοστό του πληθυσμού βρίσκεται ανάμεσα σε αυτά τα όρια

Παράδειγμα 2

Σε κάποια δημοσκόπηση πριν τις εκλογές ερωτήθηκαν 585 ψηφοφόροι και δήλωσαν ότι ψηφίζουν το κόμμα Α οι 213.

Εχουμε

$$\pi = \frac{213}{585} = 0.364$$

Το τυπικό σφάλμα είναι

$$SE(\pi) = 0,0199$$

Οπότε τα όρια εμπιστοσύνης για το ποσοστό που θα πάρει το κόμμα Α στις εκλογές είναι

$$0,364 \pm 2 \cdot 0,0199 = 0,364 \pm 0,0398$$

Δηλ. 0,324 μέχρι 0,404 ή

32,4% μέχρι **40,4%**

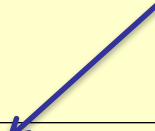
Από τι εξαρτάται το μέγεθος του Δ.Ε. για ένα ποσοστό?

Από το μέγεθος δείγματος n . Όσο πιο μεγάλο είναι το n , τόσο πιο μικρό είναι το Δ.Ε., άρα και η εκτίμηση του ποσοστού είναι καλύτερη

Συνήθως για να έχουμε μικρό Δ.Ε. ($\pm 2\%$ π.χ.) πρέπει να πάρουμε πολύ μεγάλο δείγμα, πάνω από 1000 ή 2000 άτομα)

Παράδειγμα (διερεύνηση)

Αν είχαμε αυξήσει το μέγεθος του δείγματος
(δηλαδή το συνολικό αριθμό των ψηφοφόρων)



			95% Δ.Ε.	
Κόμμα Α	Ψηφοφόροι	π	από	μέχρι
213	585	36,4%	32,4%	40,4%
745	2047	36,4%	34,3%	38,5%
1713	4708	36,4%	35,0%	37,8%
3426	9416	36,4%	35,4%	37,4%
13704	37664	36,4%	35,9%	36,9%

Μια δημοσκόπηση...

Σε 1094 άτομα, στην ερώτηση «Πόσο ικανοποιημένοι είστε με την υγειονομική διαχείριση της πανδημίας από την κυβέρνηση», διαβάσαμε τον πίνακα

Λίγο/Καθόλου	62,6%
Πολύ/Αρκετά	37,4%

Πως θα δημιουργήσω ΔΕ;

Βρίσκω το SE των δύο ποσοστών με τον τύπο

$$SE = 0,0147 \quad \text{---} \quad SE = 1,47\%$$

Οπότε τα 95% Δ.Ε. για τα ποσοστά γίνονται

	από	μέχρι
Λίγο/Καθόλου	59,7%	65,5%
Πολύ/Αρκετά	34,4%	40,3%

Άσκηση



Αν γνωρίζουμε ότι ρωτήθηκαν 1120 άτομα, να βρείτε το 95% Δ.Ε. για όσους δήλωσαν ότι αξιολογούν την εξέλιξη του προγράμματος εμβολιασμών ΣΙΓΟΥΡΑ/ΜΑΛΛΟΝ ΘΕΤΙΚΑ (δηλ του 58,3%)