



Στατιστική στη Φυσική Αγωγή

(N161)

5^η Διάλεξη

Κοκκότης Χρήστος, PhD

Στόχοι Διάλεξης

Κατανόηση βασικών εννοιών:

- Εκτίμηση παραμέτρων και διαστημάτων
- t-κατανομή

Εφαρμογές Κεντρικού Οριακού Θεωρήματος (Κ.Ο.Θ.)

Σύμφωνα με το «Κεντρικό Οριακό Θεώρημα»:

όταν ένας πληθυσμός ακολουθεί την κανονική κατανομή,
οι δειγματικοί μέσοι όροι ακολουθούν επίσης την κανονική κατανομή.

Κάθε κανονική κατανομή μπορεί να μετασχηματιστεί
σε τυπική κανονική κατανομή

$$z = \frac{X - \mu}{\sigma}$$

Εφαρμογές Κεντρικού Οριακού Θεωρήματος (Κ.Ο.Θ.)

Η δειγματική κατανομή, που είναι μια κανονική κατανομή, μπορεί να μετασχηματιστεί σε **z – τυπική κατανομή**, μετασχηματίζοντας τις τιμές X σε z τιμές.

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \rightarrow z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{N}}}$$

Όπου:

\bar{x} = η μέση τιμή του δείγματος

μ = η μέση τιμή του πληθυσμού

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$ = η τυπική απόκλιση του δειγματικού μέσου (τυπικό σφάλμα)

σ = η τυπική απόκλιση του πληθυσμού

N = το μέγεθος του δείγματος

Εφαρμογές Κεντρικού Οριακού Θεωρήματος (Κ.Ο.Θ.)

Σύμφωνα λοιπόν

- ▶ με το «κεντρικό οριακό θεώρημα» και
- ▶ τη δυνατότητα μετασχηματισμού της δειγματικής κατανομής σε z – τυπική κατανομή,
μπορεί να υπολογιστεί
- ▶ πόση είναι η πιθανότητα η μέση τιμή ενός δείγματος να βρίσκεται μεταξύ συγκεκριμένων ορίων, βάσει της μέσης τιμής του πληθυσμού και της τυπικής απόκλισης του πληθυσμού.

Εφαρμογές Κεντρικού Οριακού Θεωρήματος (Κ.Ο.Θ.)

Παράδειγμα:

Ένας προπονητής γνωρίζει ότι ο χρόνος αντίδρασης ολόκληρου του πληθυσμού

- ▶ ακολουθεί την κανονική κατανομή
- ▶ η μέση τιμή ολόκληρου του πληθυσμού είναι $\mu = 100$ msec και
- ▶ η τυπική απόκλιση του πληθυσμού είναι $\sigma = 20$ msec

Ο συγκεκριμένος προπονητής προπονεί μια ομάδα η οποία αποτελεί ένα δείγμα 16 ατόμων ($N = 16$).

Τι πιθανότητα υπάρχει η μέση τιμή του χρόνου αντίδρασης του συγκεκριμένου δείγματος \bar{X} να είναι μεταξύ 90 και 110 msec ($90 < \bar{X} < 110$);

Εφαρμογές Κεντρικού Οριακού Θεωρήματος (Κ.Ο.Θ.)

Εφόσον ο πληθυσμός ακολουθεί την **κανονική κατανομή**, τότε και η δειγματική κατανομή θα ακολουθεί επίσης την **κανονική κατανομή**.

Συνεπώς, η δειγματική κατανομή μπορεί να μετατραπεί σε z – τυπική κατανομή:

$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{N}}}$$

Εφαρμογές Κεντρικού Οριακού Θεωρήματος (Κ.Ο.Θ.)

- ▶ $\mu = 100$ msec \bar{X} μεταξύ 90 και 110 msec
- ▶ $\sigma = 20$ msec
- ▶ $N = 16$ άτομα

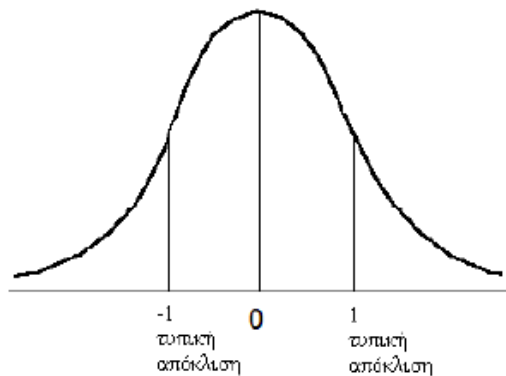
$$\bar{X} = 90: \quad z_{90} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{N}}} = \frac{90 - 100}{\frac{20}{\sqrt{16}}} = \frac{-10}{5} = -2$$

$$\bar{X} = 110: \quad z_{110} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{N}}} = \frac{110 - 100}{\frac{20}{\sqrt{16}}} = \frac{10}{5} = 2$$

Διαστήματα εμπιστοσύνης

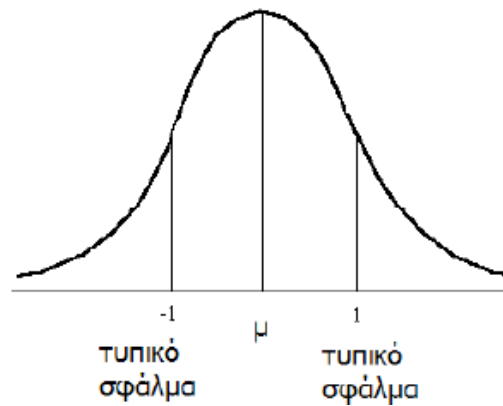
Στη z – τυπική κατανομή

- ο μέσος όρος είναι $\mu=0$
- η τυπική απόκλιση είναι $\sigma=1$
- οι τιμές στον άξονα των x (z – τιμές) εκφράζονται σε μονάδες τυπικής απόκλισης (σ)



Στη δειγματική κατανομή η τυπική απόκλιση εκφράζεται από το τυπικό σφάλμα

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



Διαστήματα εμπιστοσύνης

Συνεπώς, αν γνωρίζουμε την τυπική απόκλιση της δειγματικής κατανομής,

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

μπορούμε να προβλέψουμε «πόσο μακριά από τη μέση τιμή του πληθυσμού βρίσκεται η μέση τιμή του δείγματος»

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

Διαστήματα εμπιστοσύνης

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \rightarrow \bar{X} - \mu = z \cdot \sigma_{\bar{X}} \rightarrow \bar{X} - \mu = z \cdot \frac{\sigma}{\sqrt{N}}$$

Άρα, η μέση τιμή του δείγματος απέχει από τη μέση τιμή του πληθυσμού $\pm z \cdot \frac{\sigma}{\sqrt{N}}$ με πιθανότητα α %.

Διαστήματα εμπιστοσύνης

Παράδειγμα:

Ένας προπονητής καταγράφει τις βολές ενός ρίπτη ακοντίου όλη τη διάρκεια του έτους κάτω από συνθήκες αγώνων και γνωρίζει ότι το σύνολο των επιδόσεων του αθλητή του, δηλαδή ολόκληρος ο πληθυσμός των βολών του, ακολουθούν την **κανονική κατανομή**, με μέση τιμή $\mu = 50 \text{ m}$ και τυπική απόκλιση $\sigma = 6 \text{ m}$.

Ο προπονητής ενδιαφέρεται να μάθει **μεταξύ ποιών τιμών (ορίων) θα βρίσκεται η μέση τιμή των βολών του ρίπτη, με πιθανότητα 95% σε έναν αγώνα όπου ο αθλητής του θα εκτελέσει 6 βολές ($N = 6$)**.

Διαστήματα εμπιστοσύνης

Έχουμε

- ▶ πληθυσμός βολών που ακολουθεί την κανονική κατανομή
- ▶ $\mu = 50$ m
- ▶ $\sigma = 6$ m
- ▶ πιθανότητα 95%

Ψάχνουμε

- ▶ πληροφορίες για το δείγμα ($N = 6$ βολές)

Διαστήματα εμπιστοσύνης

Εφόσον ο πληθυσμός των βολών ακολουθεί την κανονική κατανομή,

Τότε

σύμφωνα με το κεντρικό οριακό θεώρημα και η δειγματική κατανομή θα ακολουθεί επίσης την κανονική κατανομή.

Συνεπώς

η δειγματική κατανομή μπορεί να μετατραπεί σε z – τυπική κατανομή.

Διαστήματα εμπιστοσύνης

Η τυπική απόκλιση του δειγματικού μέσου $\sigma_{\bar{x}}$

- ▶ με τυπική απόκλιση του πληθυσμού $\sigma = 6$ m και
- ▶ μέγεθος δείγματος $N = 6$ βολές, θα είναι

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}} = \frac{6}{\sqrt{6}} = \frac{6}{2,45} = 2,45$$

Διαστήματα εμπιστοσύνης

Άρα, η μέση τιμή των επιδόσεων του αθλητή σε έναν αγώνα όπου θα εκτελέσει 6 βολές (μέγεθος δείγματος $N=6$), υπάρχει πιθανότητα 95 % να κυμαίνεται μεταξύ

$$\mu \pm z \cdot \sigma_{\bar{x}}$$

Εφόσον

- η μέση τιμή του πληθυσμού είναι $\mu=50$ m
- η z – τιμή που αντιστοιχεί στην πιθανότητα 95% είναι $z=1.96$ και
- η τυπική απόκλιση του δειγματικού μέσου είναι $\sigma_{\bar{x}}=2.45$

$$\mu \pm z \cdot \sigma_{\bar{x}} = 50 \pm 1.96 \cdot 2.45 = 50 \pm 4.802$$

Διαστήματα εμπιστοσύνης

Άρα η μέση τιμή του δείγματος των βολών σε έναν αγώνα όπου εκτελούνται $N=6$ βολές,

- ▶ υπάρχει πιθανότητα 95% να κυμαίνεται από 45.198 m έως 54.802 m:

$$50 - 4.802 < \bar{X} < 50 + 4.802 \rightarrow 45.198 < \bar{X} < 54.802$$

t-κατανομή

Αν δεν είναι γνωστή η τυπική απόκλιση του πληθυσμού (σ),
Τότε θα πρέπει να χρησιμοποιηθεί ένας εκτιμητής της.

η τυπική απόκλιση του δείγματος $s = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1}}$

είναι ένας μεροληπτικός εκτιμητής της τυπικής απόκλισης
του πληθυσμού (σ)

t-κατανομή

- ▶ Όταν το μέγεθος του δείγματος είναι μεγάλο ($N > 30$).

η μεταβλητή

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{N}}} = \frac{\bar{X} - \mu}{s_{\bar{x}}}$$

ακολουθεί την κανονική κατανομή
γιατί οι διαφορές που προκαλούνται στον
τύπο

$$s = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1}}$$

από την αντικατάσταση του N με $N - 1$,
δεν δημιουργούν σοβαρά προβλήματα

- ▶ Όταν το μέγεθος του δείγματος είναι μικρό ($N < 30$).

η μεταβλητή

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{N}}} = \frac{\bar{X} - \mu}{s_{\bar{x}}}$$

ΔΕΝ ακολουθεί την κανονική κατανομή

t-κατανομή

Όταν το μέγεθος του δείγματος
είναι μικρό ($N < 30$)

αν

- ▶ δεν είναι γνωστή η τυπική απόκλιση του πληθυσμού (σ) και
- ▶ είναι γνωστή η τυπική απόκλιση του δείγματος (s)

για τον καθορισμό των διαστημάτων εμπιστοσύνης **δεν χρησιμοποιείται η z – τυπική κατανομή**, όπως στην περίπτωση όπου

- ▶ ο πληθυσμός ακολουθεί την κανονική κατανομή ή
- ▶ όπου ο πληθυσμός δεν ακολουθεί την κανονική κατανομή αλλά είναι μεγάλο το μέγεθος του δείγματος,

αλλά

- ▶ χρησιμοποιείται μια άλλη μορφή κατανομής, η **t-κατανομή**.

t-κατανομή

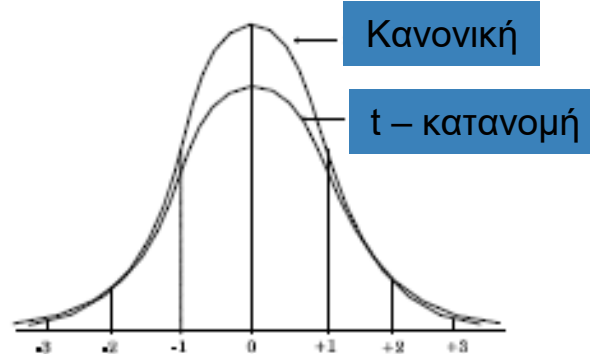
- ▶ Σ' αυτές τις περιπτώσεις αντί για την z - τιμή χρησιμοποιείται η t - τιμή:

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{N}}} = \frac{\bar{X} - \mu}{s_{\bar{x}}}$$

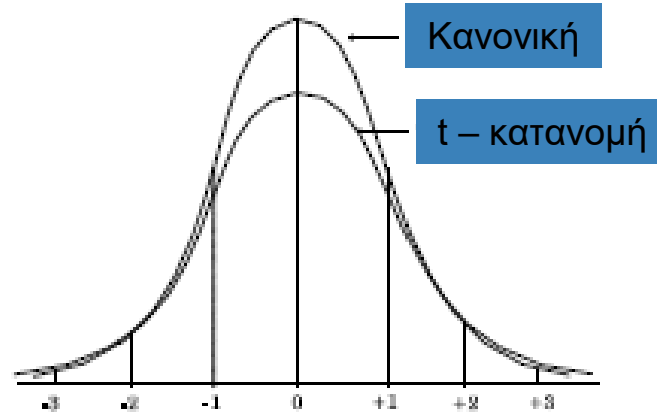
t-κατανομή

Η t – κατανομή έχει πολλές ομοιότητες με τη z – τυπική κατανομή.

- ▶ Η t – κατανομή εξαρτάται από το μέγεθος του δείγματος (N) και επομένως πρόκειται για μια για οικογένεια κατανομών.
- ▶ Όλα τα σημεία της t – κατανομής βρίσκονται πιο κάτω από τα σημεία της z – τυπικής κατανομής.
- ▶ Όταν όμως το μέγεθος του δείγματος είναι μεγαλύτερο από 30 ($N > 30$), οι δύο κατανομές αρχίζουν να ταυτίζονται.
- ▶ Για μικρά δείγματα ωστόσο, η διασπορά που παρατηρείται στην t – κατανομή είναι μεγαλύτερη από αυτή της z – τυπικής κατανομής.



t-κατανομή



Στην t – κατανομή, οι αποστάσεις των τιμών μιας μεταβλητής από το μέσο όρο, μετριοούνται σε μονάδες τυπικής απόκλισης $s_{\bar{x}} = \frac{s}{\sqrt{N}}$

και

εκφράζονται με t – τιμές, οι οποίες βρίσκονται από τον πίνακα της t – κατανομής για **N-1 βαθμούς ελευθερίας**, όπου N = το μέγεθος του δείγματος.

t-κατανομή

Πίνακας t – κατανομής

<i>df</i> <i>p</i>	.60	.70	.80	.90	.95	.975	.99	.995
1	.325	.727	1.367	3.078	6.314	12.706	31.821	63.657
2	.289	.617	1.061	1.886	2.920	4.303	6.965	9.925
3	.277	.584	.978	1.638	2.353	3.182	4.541	5.841
4	.271	.569	.941	1.533	2.132	2.776	3.747	4.604
5	.267	.559	.920	1.476	2.015	2.571	3.365	4.032
6	.265	.553	.906	1.440	1.943	2.447	3.143	3.707
7	.263	.549	.896	1.415	1.895	2.365	2.998	3.499
8	.262	.546	.889	1.397	1.860	2.306	2.896	3.355
9	.261	.543	.883	1.383	1.833	2.262	2.821	3.250
10	.260	.542	.879	1.372	1.812	2.228	2.764	3.169
11	.260	.540	.876	1.363	1.796	2.201	2.718	3.106
12	.259	.539	.873	1.356	1.782	2.179	2.681	3.055
13	.259	.538	.870	1.350	1.771	2.160	2.650	3.012
14	.258	.537	.868	1.345	1.761	2.145	2.624	2.977
15	.258	.536	.866	1.341	1.753	2.131	2.602	2.947
16	.258	.535	.865	1.337	1.746	2.120	2.583	2.921
17	.257	.534	.863	1.333	1.740	2.110	2.567	2.898
18	.257	.534	.862	1.330	1.734	2.101	2.552	2.878
19	.257	.533	.861	1.328	1.729	2.093	2.539	2.861
20	.257	.533	.860	1.325	1.725	2.086	2.528	2.845

Διάστημα εμπιστοσύνης του μέσου όρου ενός πληθυσμού

Αν

- ▶ ο πληθυσμός ακολουθεί κανονική κατανομή και
- ▶ είναι γνωστή η τυπική απόκλιση του πληθυσμού (σ),

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} \Rightarrow z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{N}}}$$

Διάστημα εμπιστοσύνης του μέσου όρου ενός πληθυσμού

Αν

- ▶ δεν είναι γνωστή η τυπική απόκλιση του πληθυσμού (σ),
- ▶ και το δείγμα είναι μεγάλο ($N > 30$)

$$z = \frac{\bar{X} - \mu}{s_{\bar{x}}} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{N}}}$$

Διάστημα εμπιστοσύνης του μέσου όρου ενός πληθυσμού

Αν

- ▶ δεν είναι γνωστή η τυπική απόκλιση του πληθυσμού (σ),
- ▶ και το δείγμα είναι μικρό ($N < 30$).

$$t = \frac{\bar{X} - \mu}{s_{\bar{x}}} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{N}}}$$

Διάστημα εμπιστοσύνης του μέσου όρου ενός πληθυσμού

Αν

- ▶ ο πληθυσμός ακολουθεί κανονική κατανομή και
- ▶ είναι γνωστή η τυπική απόκλιση του πληθυσμού (σ),

Θέλουμε να εκτιμήσουμε μεταξύ ποιών τιμών θα βρίσκεται ο αριθμητικός μέσος του πληθυσμού, βάσει του αριθμητικού μέσου του δείγματος

- ▶ πληθυσμός ακολουθεί κανονική κατανομή

ΣΥΜΦΩΝΑ με το Κ.Ο.Θ

οι δειγματικοί αριθμητικοί μέσοι τυχαίων δειγμάτων, ακολουθούν και αυτοί την κανονική κατανομή γύρω από τον αριθμητικό μέσο του πληθυσμού, ανεξάρτητα από το μέγεθος των δειγμάτων.

Διάστημα εμπιστοσύνης του μέσου όρου ενός πληθυσμού

Η δειγματική κατανομή, που είναι μια κανονική κατανομή, μπορεί να μετασχηματιστεί σε **z – τυπική κατανομή**, μετασχηματίζοντας τις τιμές X σε z τιμές.

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \rightarrow z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{N}}}$$

Όπου:

\bar{x} = η μέση τιμή του δείγματος

μ = η μέση τιμή του πληθυσμού

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$ = η τυπική απόκλιση του δειγματικού μέσου (τυπικό σφάλμα)

σ = η τυπική απόκλιση του πληθυσμού

N = το μέγεθος του δείγματος

Διάστημα εμπιστοσύνης του μέσου όρου ενός πληθυσμού

Παράδειγμα:

Ένας προπονητής καταγράφει τις επιδόσεις 5 αθλητών του ($N=5$) σε ένα τεστ ευστοχίας. Οι επιδόσεις των αθλητών του είναι: 3, 5, 5, 4 και 3.

Αν ο πληθυσμός από τον οποίο προέρχεται το συγκεκριμένο δείγμα

- ▶ ακολουθεί την κανονική κατανομή και
- ▶ η τυπική απόκλιση του πληθυσμού είναι $\sigma = 1.5$,

ο συγκεκριμένος προπονητής θέλει να προσδιοριστεί το διάστημα των τιμών μεταξύ των οποίων θα βρίσκεται ο μέσος όρος του πληθυσμού, με πιθανότητα 95%.

Θέλει δηλαδή να προσδιορίσει το διάστημα εμπιστοσύνης του μέσου όρου του πληθυσμού (μ) με πιθανότητα επιτυχούς προσδιορισμού 95%.

Διάστημα εμπιστοσύνης του μέσου όρου ενός πληθυσμού

Η μέση τιμή των επιδόσεων του συγκεκριμένου δείγματος των πέντε αθλητών είναι:

$$\bar{X} = \frac{3+5+5+4+3}{5} = \frac{20}{5} = 4$$

Εφόσον

- ✓ ο πληθυσμός από τον οποίο προέρχεται το δείγμα ακολουθεί την κανονική κατανομή και
- ✓ γνωρίζουμε την τυπική απόκλιση του πληθυσμού (σ)

τότε

χρησιμοποιούμε τις τυπικές τιμές της δειγματικής κατανομής.

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} \Rightarrow z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{N}}}$$

Διάστημα εμπιστοσύνης του μέσου όρου ενός πληθυσμού

«το διάστημα των τιμών
μεταξύ των οποίων θα βρίσκεται ο μέσος όρος του
πληθυσμού,
με πιθανότητα 95%»



Πίνακας z – τυπικής κατανομής

Διάστημα εμπιστοσύνης του μέσου όρου ενός πληθυσμού

Συνεπώς το διάστημα εμπιστοσύνης του μέσου όρου του πληθυσμού, με πιθανότητα 95%, θα βρίσκεται μεταξύ των τιμών

$$\bar{X} - z \cdot \frac{\sigma}{\sqrt{N}} < \mu < \bar{X} + z \cdot \frac{\sigma}{\sqrt{N}} \Rightarrow$$

$$\bar{X} = 4$$

$$z = 1.96$$

$$\sigma = 1.5$$

$$N = 5$$

$$4 - 1.96 \cdot \frac{1.5}{\sqrt{5}} < \mu < 4 + 1.96 \cdot \frac{1.5}{\sqrt{5}} \Rightarrow$$

$$2.685 < \mu < 5.315$$

Διάστημα εμπιστοσύνης του μέσου όρου ενός πληθυσμού

Αν

- ▶ δεν είναι γνωστή η τυπική απόκλιση του πληθυσμού (σ),
- ▶ και το δείγμα είναι **μεγάλο** ($N > 30$)

Θέλουμε να εκτιμήσουμε μεταξύ ποιών τιμών θα βρίσκεται ο αριθμητικός μέσος του πληθυσμού, βάσει του αριθμητικού μέσου του δείγματος

- ▶ Δεν γνωρίζουμε την τυπική απόκλιση του πληθυσμού
- ▶ ο πληθυσμός μπορεί να μην ακολουθεί την κανονική κατανομή
- ▶ το μέγεθος του δείγματος είναι μεγάλο ($N > 30$)

ΣΥΜΦΩΝΑ με το Κ.Ο.Θ

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

Διάστημα εμπιστοσύνης του μέσου όρου ενός πληθυσμού

Παράδειγμα:

Μια εταιρεία αθλητικών παπουτσιών καταγράφει το χρόνο ζωής ενός δείγματος $N = 120$ αθλητικών παπουτσιών, τα οποία έχουν χρησιμοποιηθεί κατά μέσο όρο $\bar{X} = 335$ ώρες

Η τυπική απόκλιση του δείγματος είναι $s = 70$ ώρες

Η συγκεκριμένη εταιρεία θέλει να προσδιορίσει μεταξύ ποιών τιμών θα βρίσκεται ο μέσος όρος του πληθυσμού ($\mu = ?$), με πιθανότητα 95%;

Διάστημα εμπιστοσύνης του μέσου όρου ενός πληθυσμού

Εφόσον το δείγμα είναι μεγάλο ($N=120 > 30$) για το καθορισμό του διαστήματος εμπιστοσύνης του μέσου όρου του πληθυσμού χρησιμοποιούμε την z – τυπική κατανομή:

$$z = \frac{\bar{X} - \mu}{s_{\bar{x}}} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{N}}}$$

$$\bar{X} - z \cdot \frac{s}{\sqrt{N}} < \mu < \bar{X} + z \cdot \frac{s}{\sqrt{N}}$$

Διάστημα εμπιστοσύνης του μέσου όρου ενός πληθυσμού

Συνεπώς το διάστημα εμπιστοσύνης του μέσου όρου του πληθυσμού, με πιθανότητα 95%, θα βρίσκεται μεταξύ των τιμών

$$\begin{aligned} \bar{X} &= 335 & \bar{X} - z \cdot \frac{s}{\sqrt{N}} < \mu < \bar{X} + z \cdot \frac{s}{\sqrt{N}} &\Rightarrow \\ z &= 1.96 & 335 - 1.96 \cdot \frac{70}{\sqrt{120}} < \mu < 335 + 1.96 \cdot \frac{70}{\sqrt{120}} &\Rightarrow \\ s &= 70 & & \\ N &= 120 & 335 - 1.96 \cdot 6.39 < \mu < 335 + 1.96 \cdot 6.39 &\Rightarrow \\ & & 335 - 12.52 < \mu < 335 + 12.52 &\Rightarrow \\ & & 322.48 < \mu < 347.52 & \end{aligned}$$

Διάστημα εμπιστοσύνης του μέσου όρου ενός πληθυσμού

Αν

- ▶ δεν είναι γνωστή η τυπική απόκλιση του πληθυσμού (σ),
- ▶ και το δείγμα είναι μικρό ($N < 30$)

Εφόσον το δείγμα είναι μικρό ($N < 30$), αντί για τις z – τιμές (z – τυπική κατανομή) χρησιμοποιούνται οι t – τιμές (t – κατανομή).

Διάστημα εμπιστοσύνης του μέσου όρου ενός πληθυσμού

Παράδειγμα:

Μια εταιρεία αθλητικών παπουτσιών καταγράφει το χρόνο ζωής ενός δείγματος $N = 16$ αθλητικών παπουτσιών, τα οποία έχουν χρησιμοποιηθεί κατά μέσο όρο $\bar{X} = 320$ ώρες.

Η τυπική απόκλιση του δείγματος είναι $s=120$ ώρες.

Η συγκεκριμένη εταιρεία θέλει να προσδιοριστεί μεταξύ ποιών τιμών θα βρίσκεται ο μέσος όρος του πληθυσμού ($\mu=;$), με πιθανότητα 95%;

Διάστημα εμπιστοσύνης του μέσου όρου ενός πληθυσμού

Εφόσον το δείγμα είναι μικρό ($N=16 < 30$) για το καθορισμό του διαστήματος εμπιστοσύνης του μέσου όρου του πληθυσμού δεν χρησιμοποιούμε την z - τυπική κατανομή, αλλά την t - κατανομή:

$$\bar{X} - t \cdot \frac{s}{\sqrt{N}} < \mu < \bar{X} + t \cdot \frac{s}{\sqrt{N}}$$

Για τον υπολογισμό των διαστημάτων εμπιστοσύνης θα πρέπει να καθοριστεί η t - τιμή που αντιστοιχεί σε $N - 1$ βαθμούς ελευθερίας και σε επίπεδο σημαντικότητας που αντιστοιχεί στην πιθανότητα των 95%, το οποίο ορίζεται ως $\alpha = 0.05$.

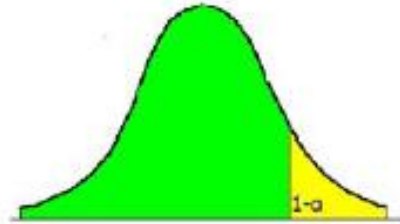
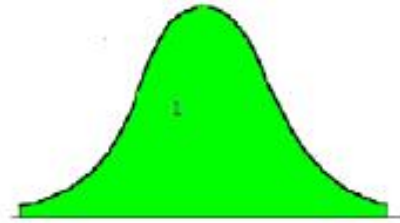
Διάστημα εμπιστοσύνης του μέσου όρου ενός πληθυσμού

Το συνολικό εμβαδόν κάτω από την καμπύλη της t - κατανομής ισούται με 1.

Κατά συνέπεια για να εντοπιστεί η t - τιμή που αντιστοιχεί στο επίπεδο σημαντικότητας $\alpha = 0.05$ θα πρέπει να δημιουργηθεί η διαφορά:

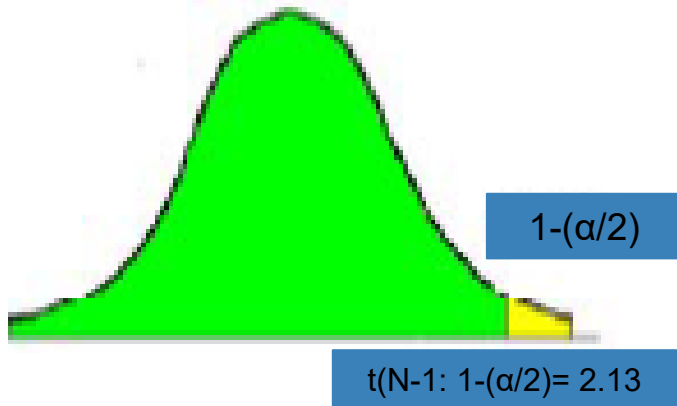
$$1 - \alpha = 1 - 0.05.$$

Επειδή όμως μας ενδιαφέρουν οι τιμές οι οποίες βρίσκονται και δεξιά και αριστερά από τον μέσο όρο, θα πρέπει να υπολογιστεί η διαφορά: $1 - \alpha/2 = 1 - 0.05/2 = 1 - 0.025 = 0.975$.



Διάστημα εμπιστοσύνης του μέσου όρου ενός πληθυσμού

- ▶ $1 - (\alpha/2) = 1 - (0.05/2) = 1 - 0.025 = 0.975$
- ▶ $N - 1 = 16 - 1 = 15$



<i>df</i> <i>p</i>	.60	.70	.80	.90	.95	.975	.99	.995
1	.325	.727	1.367	3.078	6.314	12.706	31.821	63.657
2	.289	.617	1.061	1.886	2.920	4.303	6.965	9.925
3	.277	.584	.978	1.638	2.353	3.182	4.541	5.841
4	.271	.569	.941	1.533	2.132	2.776	3.747	4.604
5	.267	.559	.920	1.476	2.015	2.571	3.365	4.032
6	.265	.553	.906	1.440	1.943	2.447	3.143	3.707
7	.263	.549	.896	1.415	1.895	2.365	2.998	3.499
8	.262	.546	.889	1.397	1.860	2.306	2.896	3.355
9	.261	.543	.883	1.383	1.833	2.262	2.821	3.250
10	.260	.542	.879	1.372	1.812	2.228	2.764	3.169
11	.260	.540	.876	1.363	1.796	2.201	2.718	3.106
12	.259	.539	.873	1.356	1.782	2.179	2.681	3.055
13	.259	.538	.870	1.350	1.771	2.160	2.650	3.012
14	.258	.537	.868	1.345	1.761	2.145	2.624	2.977
15	.258	.536	.866	1.341	1.753	2.131	2.602	2.947
16	.258	.535	.865	1.337	1.746	2.120	2.583	2.921
17	.257	.534	.863	1.333	1.740	2.110	2.567	2.898
18	.257	.534	.862	1.330	1.734	2.101	2.552	2.878
19	.257	.533	.861	1.328	1.729	2.093	2.539	2.861
20	.257	.533	.860	1.325	1.725	2.086	2.528	2.845

Διάστημα εμπιστοσύνης του μέσου όρου ενός πληθυσμού

Συνεπώς, το διάστημα εμπιστοσύνης στο οποίο θα βρίσκεται ο μέσος όρος του πληθυσμού με πιθανότητα 95%, θα είναι

$$\begin{aligned}
 \bar{X} &= 320 \text{ ώρες} & \bar{X} - t \cdot \frac{s}{\sqrt{N}} < \mu < \bar{X} + t \cdot \frac{s}{\sqrt{N}} &\Rightarrow \\
 t(N-1; 1 - (\alpha/2)) &= 2.13. & & \\
 N &= 16 & 320 - 2.13 \cdot \frac{120}{\sqrt{16}} < \mu < 320 + 2.13 \cdot \frac{120}{\sqrt{16}} &\Rightarrow \\
 s &= 120 \text{ ώρες} & 320 - 2.13 \cdot 30 < \mu < 320 + 2.13 \cdot 30 &\Rightarrow \\
 & & 320 - 63.9 < \mu < 320 + 63.9 &\Rightarrow \\
 & & 256.1 < \mu < 383.9 &
 \end{aligned}$$

Ερωτήσεις & Συζήτηση

