Θεώρημα του Chebyshev
Νόμος των Μεγάλων Αριθμών
Κεντρικό Οριακό Θεώρημα
Διαστήματα Εμπιστοσύνης

Καθηγητής Χρήστος Γ. Μασούρος

Θεώρημα του **Chebyshev**

Διαισθητικά αντιλαμβανόμαστε ότι όταν η τυπική απόκλιση είναι μικρή, οι μετρήσεις τείνουν να συσσωρευθούν πέριξ της μέσης τιμής και όταν η τυπική απόκλιση είναι μεγάλη τείνουν να διασπαρούν. Το ερώτημα το οποίο γεννιέται είναι το εξής:

> *Τι κλάσμα των μετρήσεων ευρίσκεται εντός δοθέντος διαστήματος με κέντρο την μέση τιμή;*

ή ακόμη:

> *Ποιο διάστημα πρέπει να λάβουμε εκατέρωθεν της μέσης τιμής ώστε τούτο να περικλείει π.χ. τα ¾ των μετρήσεων;*

Θεώρημα του **Chebyshev**

Διατυπώνουμε το φημισμένο θεώρημα του Chebyshev:

Αν δίνεται η κατανομή πιθανότητας της τυχαίας μεταβλητής X με μέση τιμή **μ** και τυπική απόκλιση **σ**, τότε η πιθανότητα να παρατηρήσουμε μια τιμή της X να διαφέρει από τη μέση τιμή **μ** κατά **k** ή περισσότερες τυπικές αποκλίσεις, **δεν** μπορεί να είναι μεγαλύτερη από $\mathbf{1/k^2}$

$$\mathbf{P[|X-\mu| \geq k\sigma] \leq 1/k^2}$$

Θεώρημα του **Chebyshev**

Εναλλακτική μορφή του θεωρήματος του Chebyshev:

Αν δίνεται η κατανομή πιθανότητας της τυχαίας μεταβλητής X με μέση τιμή **μ** και τυπική απόκλιση **σ**, τότε η πιθανότητα να παρατηρήσουμε μια τιμή της X εντός του διαστήματος το οποίο ορίζεται εάν εκατέρωθεν της μέσης τιμής **μ** λάβουμε **k** τυπικές αποκλίσεις είναι τουλάχιστον **$1 - (1/k^2)$**, δηλαδή:

$$P[|X-\mu| < k\sigma] \geq 1 - (1/k^2)$$

Για παράδειγμα, εάν λάβουμε διάστημα **2** τυπικών αποκλίσεων εκατέρωθεν της μέσης τιμής **μ**, τότε τουλάχιστον τα **¾** των μετρήσεων περικλείονται σε αυτό το διάστημα.

Παραδείγματα στο Θεώρημα του **Chebyshev**

1. Εάν οι μετρήσεις είναι:

-8, -1, -1, 0, 0, 0, 0, 1, 1, 8

τότε η μέση τιμή είναι **0** και τυπική απόκλιση περίπου **3,6**. Σε απόσταση 2 τυπικών αποκλίσεων, ήτοι μεταξύ **-7,2** και **7,2** υπάρχουν 8 μετρήσεις, ή το 80% των μετρήσεων. Οι 10 μετρήσεις βρίσκονται σε απόσταση 3 αποκλίσεων από τον μέσο όρο.

2. Στην περίπτωση της κανονικής κατανομής, το θεώρημα του Chebyshev δίνει «ασθενή πληροφορία», διότι αντί των πιθανοτήτων 0,95 και 0,997 που γνωρίζουμε ότι ισχύουν, μας οδηγεί στις πιθανότητες τουλάχιστον

$1-(1/2^2)=0,75$ και $1-(1/2^3)=0,899$ αντίστοιχα

Παράδειγμα δειγματοληψίας χωρίς επανάθεση

Έστω ότι ένας **πληθυσμός** συνίσταται από 4 φοιτητές, οι οποίοι έχουν τα παρακάτω χρηματικά ποσά:

40€, 80€, 100€, 160€

Τότε η μέση τιμή του πληθυσμού είναι

$$\mu = \frac{40+80+100+160}{4} = 95$$

και τυπική απόκλιση σε ολόκληρο τον πληθυσμό είναι

$$\sigma^2 = \frac{1}{4}\left[(40-95)^2 + (80-95)^2 + (100-95)^2 + (160-95)^2\right] = 1875$$

$$\sigma = \sqrt{1875} \approx 43{,}3$$

όπου αμφότερα τα μ και σ εκφράζονται σε ευρώ

Παράδειγμα δειγματοληψίας χωρίς επανάθεση

Στη συνέχεια λαμβάνουμε από τον **πληθυσμό** των 4ων φοιτητών, όλα τα δυνατά **δείγματα** μεγέθους 2:

		Χρηματικό ποσό του δεύτερου φοιτητή που επιλέγεται			
		40	80	100	160
Χρηματικό ποσό του πρώτου φοιτητή που επιλέγεται	40	-	(40, 80)	(40, 100)	(40, 160)
	80	(80 ,40)	-	(80 ,100)	(100, 160)
	100	(100, 40)	(100, 80)	-	(100, 160)
	160	(160, 40)	(160, 80)	(160, 100)	-

Η διαγώνιος του πίνακα είναι κενή, διότι η λήψη γίνεται χωρίς επανάθεση.

Παράδειγμα δειγματοληψίας χωρίς επανάθεση

Για κάθε ένα από τα προηγούμενα δείγματα υπάρχει ένας δειγματικός μέσος. Στον πίνακα που ακολουθεί υπολογίζονται όλοι αυτοί οι δειγματικοί μέσοι

Τιμές των Δειγματικών μέσων $\bar{X}$			
-	60	70	100
60	-	90	120
70	90	-	130
100	120	130	-

Έχουμε τώρα ένα νέο πληθυσμό, τον πληθυσμό των δειγματικών μέσων.

Παράδειγμα δειγματοληψίας χωρίς επανάθεση

Η μέση τιμή του νέου αυτού πληθυσμού είναι $\mu_{\overline{X}} = 95$

και η διακύμανση $\sigma^2_{\overline{X}} = 625$

Παρατηρούμε ότι οι μέσοι δειγματικοί έχουν δύο αξιοσημείωτες ιδιότητες:

1. Ο μέσος τους ισούται με τον μέσο του αρχικού πληθυσμού
2. Η διακύμανση είναι μικρότερη (το ένα τρίτο) της διακύμανσης του αρχικού πληθυσμού. Δηλαδή οι δειγματικοί μέσοι είναι συγκεντρωμένοι πέριξ του μέσου του πληθυσμού περισσότερο παρά οι ατομικές μετρήσεις.

Θεώρημα 1

Έστω πληθυσμός μεγέθους **N** με μέση τιμή **μ** και διακύμανση **σ^2**. Από τον πληθυσμό αυτό λαμβάνουμε δείγματα μεγέθους **n** **χωρίς επανάθεση**. Τότε:

- η μέση τιμή του πληθυσμού των δειγματικών μέσων ισούται με την μέση τιμή του αρχικού πληθυσμού, δηλαδή

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

- η διακύμανση του πληθυσμού των δειγματικών μέσων είναι

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$$

Θεώρημα 2

Έστω πληθυσμός μεγέθους **N** με μέση τιμή **μ** και διακύμανση **σ^2**. Από τον πληθυσμό αυτό λαμβάνουμε δείγματα μεγέθους **n** **με επανάθεση**. Τότε:

- η μέση τιμή του πληθυσμού των **δειγματικών μέσων** ισούται με την μέση τιμή του αρχικού πληθυσμού, δηλαδή

$$\mu_{\overline{X}} = \mu$$

- η διακύμανση του πληθυσμού των **δειγματικών μέσων** είναι

$$\sigma^2_{\overline{x}} = \frac{\sigma^2}{n}$$

Νόμος των Μεγάλων Αριθμών

Από όσα προηγήθηκαν έχουμε δει ότι όταν αυξάνεται το μέγεθος του δείγματος, ελαττώνεται η τυπική απόκλιση της κατανομής συχνοτήτων των δειγματικών μέσων. Έτσι δημιουργείται το ερώτημα:

Αν κάνουμε το μέγεθος του δείγματος αρκετά μεγάλο, μπορούμε στην πράξη να είμαστε βέβαιοι ότι ο δειγματικός μέσος θα βρίσκεται πλησίον του μέσου του πληθυσμού;

Για δείγματα που λαμβάνονται χωρίς επανάθεση η απάντηση είναι καταφατική και δίνεται από το Θεώρημα 1, από το οποίο προκύπτει ότι όταν n=N τότε $\sigma_{\bar{x}}^2 = 0$

Για δείγματα που λαμβάνονται με επανάθεση η απάντηση είναι επίσης καταφατική και η απάντηση δίνεται από το Θεώρημα που ακολουθεί και φέρει το όνομα ***Νόμος των Μεγάλων Αριθμών***.

Νόμος των Μεγάλων Αριθμών

Θεώρημα (***Νόμος των Μεγάλων Αριθμών)***. **Αν ένας πληθυσμός έχει μέσο μ και διακύμανση σ^2 και αν $\bar{x}$ είναι ο μέσος ενός τυχαίου δείγματος μεγέθους n, που λαμβάνεται με επανάθεση από τον πληθυσμό αυτό, τότε για τυχόν $\delta>0$, είναι**

$$\lim_{n\to\infty} P\left(\mu-\delta\le\bar{x}\le\mu+\delta\right)=1$$

Δηλαδή, αν λάβουμε ένα δείγμα με μέγεθος n αρκετά μεγάλο, τότε μπορούμε να είμαστε όσον επιθυμούμε βέβαιοι ότι ο δειγματικός μέσος θα προσεγγίσει τον αληθή μέσο όσο θέλουμε.

Το πρακτικό συμπέρασμα αυτού του Θεωρήματος είναι ότι τα μεγάλα δείγματα τείνουν να έχουν μέσους οι οποίοι προσεγγίζουν τον μέσο του πληθυσμού.

Εφαρμογές του Νόμου των Μεγάλων Αριθμών

Αν ορίσουμε ως δίκαιο παιχνίδι, ένα παιχνίδι στο οποίο το μέσο κέρδος είναι μ=0, τότε ο Νόμος των Μεγάλων Αριθμών λέει ότι αν ένα τέτοιο παιχνίδι παιχθεί για αρκετό χρονικό διάστημα τα αποτελέσματα θα δώσουν ισοπαλία. Π.χ. στο παιχνίδι κεφάλι-γράμματα το κέρδος είναι +1 ή -1 μονάδες. Αν οι δύο πλευρές του νομίσματος είναι ισοπίθανες (αμερόληπτο νόμισμα) το θεώρημα λέει ότι αν το παιχνίδι αυτό παιχθεί επί μεγάλο διάστημα, το **μέσο** κέρδος (ή ζημία) είναι σχεδόν βέβαιο ότι θα είναι πολύ μικρό.

Στην καθημερινότητα όμως περισσότερο ενδιαφέρον έχουν τα συνολικά αθροίσματα παρά οι μέσοι όροι. Αποδεικνύεται ότι αν παίξουμε ένα δίκαιο παιχνίδι επί αρκετό διάστημα θα έχουμε μεγάλη πιθανότητα να κερδίσουμε ή να χάσουμε πολλά. Επίσης αν παίξουμε επί μακρόν ένα παιχνίδι το οποίο είναι μεροληπτικό εις βάρος μας (π.χ. ρουλέτα) είναι εξόχως πιθανόν η ζημιά μας να είναι μεγάλη.

Κεντρικό Οριακό Θεώρημα (ΚΟΘ)

Αν οι τυχαίες μεταβλητές $X_1, X_2, \ldots, X_n$ είναι ανεξάρτητες και ακολουθούν την ίδια κατανομή (όποια και αν είναι αυτή), με μέση τιμή $E(X_i) = \mu$ και διακύμανση $Var(X_i) = \sigma^2$, τότε ο δειγματικός μέσος

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

ακολουθεί για μεγάλο n ($n \to \infty$) την κανονική κατανομή

$$N\left(\mu, \frac{\sigma^2}{n}\right)$$

Κεντρικό Οριακό Θεώρημα (ΚΟΘ)

Πολλές φορές το ΚΟΘ διατυπώνεται και ως εξής:

Αν $X_1, X_2, \ldots, X_n$ ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με μέση τιμή $E(X_i)=\mu$ και διακύμανση $Var(X_i)=\sigma^2$, τότε η συνάρτηση κατανομής της τυχαίας μεταβλητής

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n \cdot \mu}{\sigma\sqrt{n}}$$

συγκλίνει, για $n \to \infty$ στην συνάρτηση κατανομής της τυποποιημένης κανονικής κατανομής **N(0,1).**

Κεντρικό Οριακό Θεώρημα (ΚΟΘ)

Το ΚΟΘ είναι από τα πιο σπουδαία Θεωρήματα της Στατιστικής και αναδεικνύει έναν από τους κύριους λόγους για τους οποίους η κανονική κατανομή είναι η πιο σπουδαία από όλες τις κατανομές.

Να τονισθεί ότι αν ο αρχικός πληθυσμός ακολουθεί κανονική κατανομή, τότε η κατανομή του $\bar{x}$ είναι ακριβώς κανονική, ανεξάρτητα από το μέγεθος n του δείγματος. Αν ο αρχικός πληθυσμός δεν ακολουθεί κανονική κατανομή, τότε η κατανομή του $\bar{x}$ είναι θα είναι κατά προσέγγιση κανονική.

Στην πράξη εφαρμόζουμε το ΚΟΘ οποτεδήποτε $n \geq 30$

Παράδειγμα

Έστω ότι ο μέσος όρος των μηνιαίων εξόδων διατροφής των φοιτητών της Φοιτητικής Εστίας είναι μ=120€ με τυπική απόκλιση σ=12€. Έστω τυχαίο δείγμα 100 φοιτητών της Εστίας. Ποια η πιθανότητα ώστε ο μέσος όρος του δείγματος να είναι μεταξύ 119€ και 121€;

Λύση: Επειδή το μέγεθος του δείγματος είναι n=100(>30) έχει εφαρμογή το ΚΟΘ, συνεπώς το $\bar{x}$ ακολουθεί κατά προσέγγιση κανονική κατανομή με

μέση τιμή $\boldsymbol{\mu_{\bar{x}} = \mu = 120}$ και τυπική απόκλιση $\boldsymbol{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{100}} = 1,2}$

Ζητάμε την πιθανότητα:

$$P(119 \le \bar{x} \le 121) = P\left(\frac{119-120}{1,2} \le \frac{\bar{x}-120}{1,2} \le \frac{121-120}{1,2}\right) =$$

$$= P(-0,83 \le \bar{z} \le 0,83) = 2 \cdot P(0 \le \bar{z} \le 0,83) = 2 \cdot F(0,83) =$$

$$= 2 \cdot 0,2967 = 0,5934$$

Ο στόχος των περισσότερων στατιστικών αναλύσεων είναι η έγκυρη γενίκευση των συμπερασμάτων τους σε πληθυσμούς, στηριζόμενες σε δείγματα προερχόμενα από τους πληθυσμούς αυτούς.

Το Κεντρικό Οριακό Θεώρημα μας λέει ότι:

- αν θα μπορούσαμε να πάρουμε όλα τα δυνατά τυχαία δείγματα μεγέθους ***n*** και υπολογίζαμε τους δειγματικούς μέσους όρους, τότε αυτοί έχουν μέσο όρο την άγνωστη σε εμάς μέση τιμή ***μ*** του πληθυσμού και
- οι δειγματικοί μέσοι όροι κατανέμονται συμμετρικά εκατέρωθέν της μέσης τιμής ***μ*** του πληθυσμού σε σχήμα καμπάνας (από την κανονική κατανομή). Έχουν τυπική απόκλιση $\sigma_{\bar{x}}$ ίση με την τυπική απόκλιση του πληθυσμού **σ** διαιρεμένη με την τετραγωνική ρίζα του **n**

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

Έστω ένας πληθυσμός από τον οποίο έχουμε **ένα μόνο δείγμα** μεγέθους ***n***.

Υπολογίζουμε
- τον μέσο όρο **του δείγματος** $\bar{\mathbf{x}}$
- την τυπική απόκλιση **του δείγματος** $\mathbf{s}$

Η μέση τιμή $\boldsymbol{\mu_{\bar{x}}}$ των δειγματικών μέσων είναι όση και η μέση τιμή $\boldsymbol{\mu}$ του πληθυσμού (δηλαδή της Χ). Έτσι η εκτίμηση της $\boldsymbol{\mu_{\bar{x}}}$ είναι $\bar{\mathbf{x}}$

Η τυπική απόκλιση $\boldsymbol{\sigma_{\bar{x}}}$ των δειγματικών μέσων και η την τυπική απόκλιση $\boldsymbol{\sigma}$ του πληθυσμού σχετίζονται ως εξής: $\boldsymbol{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$

Προσεγγίζουμε την τυπική απόκλιση $\boldsymbol{\sigma}$ του πληθυσμού με την εκτίμηση της $\mathbf{s}$

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

Έτσι η εκτίμηση της $\boldsymbol{\sigma_{\bar{x}}}$ είναι:

$$\boldsymbol{s_{\bar{x}} = \frac{s}{\sqrt{n}}}$$

Η ποσότητα αυτή ονομάζεται *Τυπικό Σφάλμα του Μέσου Όρου* ή απλά *Τυπικό Σφάλμα*

Διαστήματα Εμπιστοσύνης

Σύμφωνα με το ΚΟΘ ο δειγματικός μέσος ακολουθεί **για μεγάλο *n*** ($n \to \infty$) την κανονική κατανομή.

Η κανονική κατανομή έχει την ιδιότητα ότι αν αφαιρέσουμε δύο τυπικές αποκλίσεις από την μέση τιμή και προσθέσουμε δύο τυπικές αποκλίσεις στην μέση τιμή, τότε δημιουργείται ένα διάστημα μέσα στο οποίο περιλαμβάνεται περίπου το 95% των τιμών της τυχαίας μεταβλητής Χ που ακολουθεί την κανονική κατανομή. Έτσι σε μια ακτίνα 2 τυπικών αποκλίσεων πέριξ της μέσης τιμής του δειγματικό μέσου βρίσκεται περίπου το 95% όλων των δειγματικών μέσων (για την ακρίβεια το 95,4%).

Αντίστοιχα σε μια ακτίνα 3 τυπικών αποκλίσεων πέριξ της μέσης τιμής του δειγματικό μέσου βρίσκεται περίπου το 99% όλων των δειγματικών μέσων, ενώ σε ακτίνα 1 τυπικής απόκλισης περίπου το 68% όλων των δειγματικών μέσων

Διαστήματα Εμπιστοσύνης

Η μέση τιμή $\boldsymbol{\mu_{\bar{x}}}$ των δειγματικών μέσων $\boldsymbol{\bar{x}}$ είναι όση και η μέση τιμή $\boldsymbol{\mu}$ του πληθυσμού (δηλαδή της Χ).

Η τυπική απόκλιση $\boldsymbol{\sigma_{\bar{x}}}$ των δειγματικών μέσων $\boldsymbol{\bar{x}}$ είναι $\boldsymbol{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$ που την προσεγγίζουμε με $\boldsymbol{s_{\bar{x}} = \frac{s}{\sqrt{n}}}$

Έτσι το διάστημα

$$(\mu_{\bar{x}} - 2\, s_{\bar{x}},\ \mu_{\bar{x}} + 2\, s_{\bar{x}}) = (\mu - 2\frac{s}{\sqrt{n}}, \mu + 2\frac{s}{\sqrt{n}})$$

περιλαμβάνει περίπου το 95% των δειγματικών μέσων όρων, δηλαδή αν πάρουμε οποιοδήποτε μέσο όρο, αυτός θα βρίσκεται με πιθανότητα σχεδόν 95% στο παραπάνω διάστημα.

Επειδή όμως δεν ξέρουμε την μέση τιμή $\boldsymbol{\mu}$ του πληθυσμού και προσπαθούμε να την εκτιμήσουμε με την βοήθεια του δείγματος που έχουμε λάβει, σκεφτόμαστε ως εξής:

Διαστήματα Εμπιστοσύνης

Ο μέσος $\bar{x}$ του δείγματός μας απέχει το πολύ $2\frac{s}{\sqrt{n}}$ από την μέση τιμή μ.

Άρα η μέση τιμή απέχει το πολύ $2\frac{s}{\sqrt{n}}$ από τον μέσο $\bar{x}$ του δείγματός. Αν λοιπόν προσθέσουμε και αφαιρέσουμε στον μέσο $\bar{x}$ του δείγματός μας την ποσότητα $2\frac{s}{\sqrt{n}}$, δημιουργείται το διάστημα

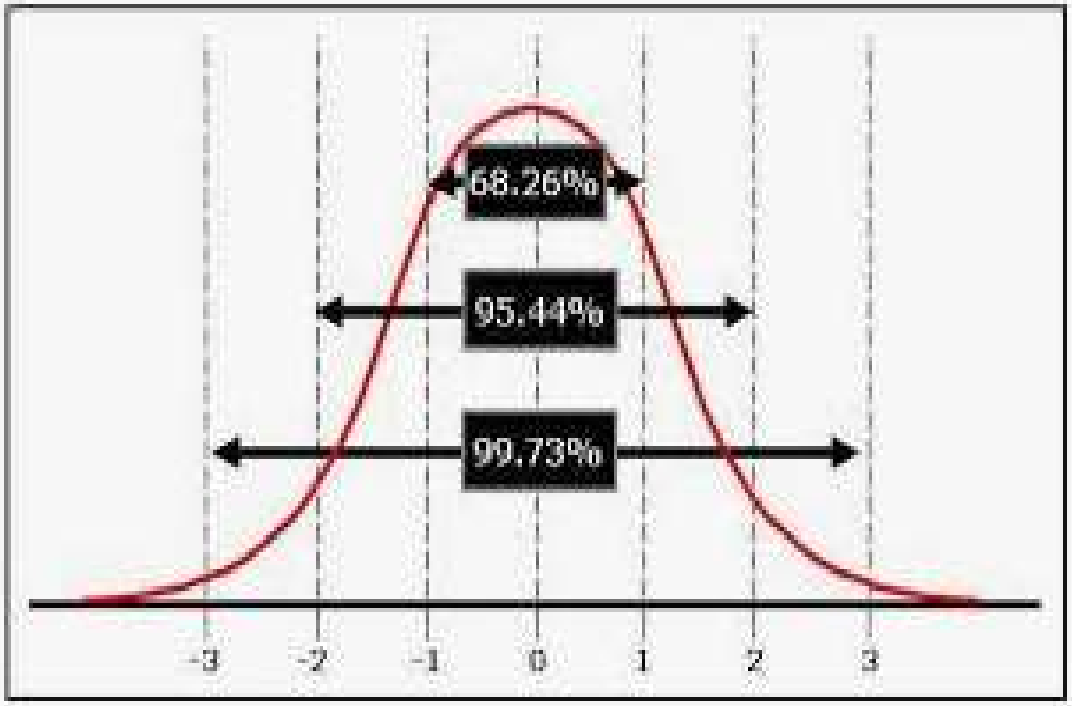
$$\left(\bar{x} - 2\frac{s}{\sqrt{n}},\ \ \bar{x} + 2\frac{s}{\sqrt{n}}\right)$$

Στο διάστημα αυτό θα έχουμε συμπεριλάβει με πιθανότητα 95% (περίπου) τη μέση τιμή του πληθυσμού. Όμοια στο διάστημα

$$\left(\bar{x} - 3\frac{s}{\sqrt{n}},\ \ \bar{x} + 3\frac{s}{\sqrt{n}}\right)$$

θα έχουμε συμπεριλάβει με πιθανότητα 99% (περίπου) τη μέση τιμή του πληθυσμού

Ακρίβεια στους υπολογισμούς



Σύμφωνα με τους πίνακες της κανονικής κατανομής στο διάστημα

(μ-2σ, μ+2σ)

βρίσκεται το 95,4% της κατανομής και όχι το 95%

ενώ

το 95% της κατανομής βρίσκεται στο διάστημα

(μ-1,96σ, μ+1,96σ)

Να προσδιορισθεί η πιθανότητα

$$P\left(\mu-2\frac{\sigma}{\sqrt{n}}\le\bar{x}\le\mu+2\frac{\sigma}{\sqrt{n}}\right)$$

$$P\left(\mu-2\frac{\sigma}{\sqrt{n}}\le\bar{x}\le\mu+2\frac{\sigma}{\sqrt{n}}\right)=$$

$$=P\left(\frac{\left(\mu-2\frac{\sigma}{\sqrt{n}}\right)-\mu}{\frac{\sigma}{\sqrt{n}}}\le\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\le\frac{\left(\mu+2\frac{\sigma}{\sqrt{n}}\right)-\mu}{\frac{\sigma}{\sqrt{n}}}\right)=$$

$$=P\left(-2\le\bar{z}\le2\right)=P\left(\bar{z}\le2\right)-\left[1-P\left(\bar{z}\le2\right)\right]=2\cdot F\left(2\right)-1\ \square\ 0{,}9544$$

Αν ο δειγματικός μέσος $\bar{x}$ δείγματος μεγέθους ***n*** έχει μέση τιμή ***μ*** και διακύμανση $\boldsymbol{\sigma^2}$ να υπολογισθεί η τιμή του ***z*** έτσι ώστε

$$P\left(\mu - z\frac{\sigma}{\sqrt{n}} \le \bar{x} \le \mu + z\frac{\sigma}{\sqrt{n}}\right) = 0,95$$

$$P\left(\mu - z\frac{\sigma}{\sqrt{n}} \le \bar{x} \le \mu + z\frac{\sigma}{\sqrt{n}}\right) = 0,95 \Leftrightarrow$$

$$\Leftrightarrow P\left(\frac{\left(\mu - z\frac{\sigma}{\sqrt{n}}\right) - \mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{\left(\mu + z\frac{\sigma}{\sqrt{n}}\right) - \mu}{\frac{\sigma}{\sqrt{n}}}\right) = 0,95 \Leftrightarrow$$

$$\Leftrightarrow P\left(-z \le \bar{Z} \le z\right) = 0,95 \Leftrightarrow P\left(\bar{Z} \le z\right) - \left[1 - P\left(\bar{Z} \le z\right)\right] = 0,95 \Leftrightarrow$$

$$\Leftrightarrow 2 \cdot F\left(z\right) - 1 = 0,95 \Leftrightarrow F\left(z\right) = \frac{1,95}{2} \Leftrightarrow F\left(z\right) = 0,9750 \Leftrightarrow z = 1,96$$

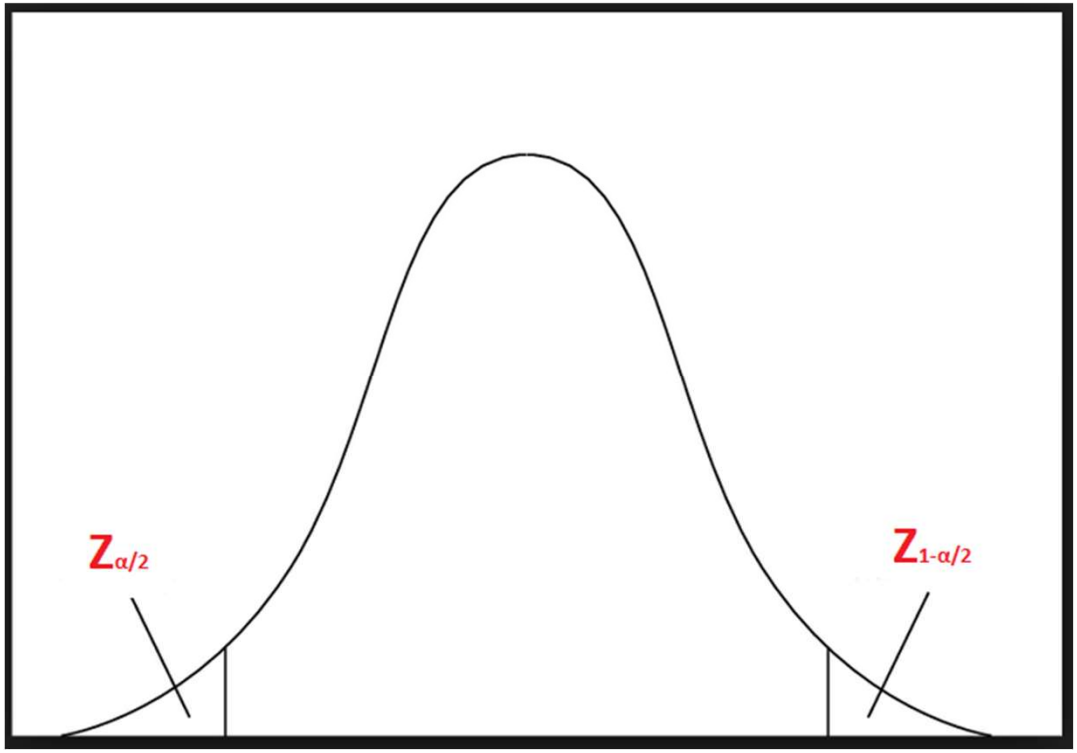
Διάστημα εμπιστοσύνης του μέσου με πιθανότητα 1-α

Η παραπάνω ανάλυση οδηγεί σε ένα γενικό κανόνα για την εκτίμηση του διαστήματος εμπιστοσύνης του μέσου με πιθανότητα 1-α. Με **1-α** συμβολίζουμε το **επίπεδο εμπιστοσύνης** (level of confidence) και με **α** το **περιθώριο** (την πιθανότητα) **σφάλματος**.

Για παράδειγμα αν επίπεδο εμπιστοσύνης είναι 95% τότε 1-α=95%, και επομένως το περιθώριο σφάλματος είναι α=5%. Όμοια αν το επίπεδο εμπιστοσύνης είναι 99% τότε το σφάλμα είναι α=1%

Διάστημα εμπιστοσύνης του μέσου με πιθανότητα 1-α

Επειδή το **α** μοιράζεται στα δύο άκρα της κατανομής δειγματοληψίας, δηλαδή **α/2** στο αριστερό και **α/2** στο δεξιό άκρο της κατανομής, οι αντίστοιχες τιμές της τυποποιημένης κανονικής μεταβλητής **Z** συμβολίζονται με $Z_{\alpha/2}$ και $Z_{1-\alpha/2}$ αντίστοιχα.



Εφαρμογή

Ας υποθέσουμε ότι η διασπορά σ^2 ενός πληθυσμού είναι γνωστή και ότι θέλουμε να κατασκευάσουμε ένα 95% διάστημα εμπιστοσύνης για τη μέση τιμή μ. Δηλαδή αναζητούμε ένα διάστημα πραγματικών αριθμών το οποίο να περιέχει το μ με πιθανότητα 0,95. Λαμβάνουμε από τον πληθυσμό αυτό ένα δείγμα $n>30$. Γνωρίζουμε από το ΚΟΘ ότι ο μέσος $\bar{X}$ του δείγματος ακολουθεί κανονική κατανομή με μέση τιμή μ και διασπορά $\frac{\sigma^2}{n}$. Επομένως

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Εφαρμογή

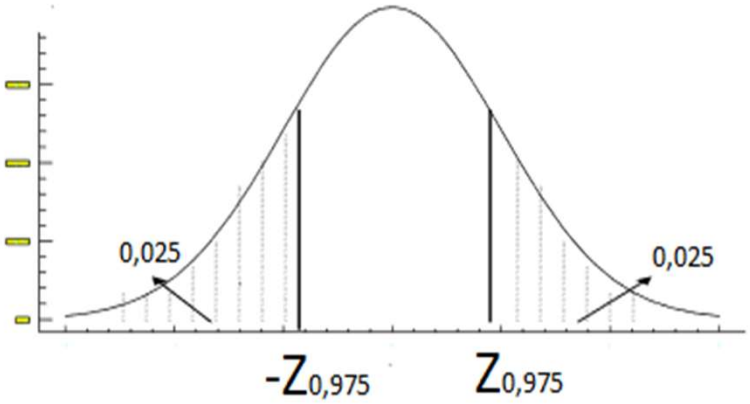
Συνεπώς αναζητούμε δύο σημεία Z_L και Z_U από τους πίνακες της τυποποιημένης κανονικής κατανομής τέτοια ώστε,

$$P\left(Z_L \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq Z_U\right) = 0{,}95$$

Είναι 1-α = 0,95 άρα α = 0,5. Επομένως

$Z_L = Z_{0,05/2} = Z_{0,025} = -1{,}96$ και

$Z_U = Z_{1-0,05/2} = Z_{0,975} = +1{,}96$



$Z_L = Z_{0,025} = -Z_{0,975}$ και $Z_U = Z_{0,975}$

Εφαρμογή

Επομένως

$$P\left(-Z_{0,975} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq Z_{0,975}\right) = 0,95$$

Λύνοντας τη σχέση αυτή ως προς $\boldsymbol{\mu}$ παίρνουμε:

$$P\left(\bar{X} - Z_{0,975}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{0,975}\frac{\sigma}{\sqrt{n}}\right) = 0,95$$

Κατά συνέπεια το 95% διάστημα εμπιστοσύνης για τη μέση τιμή $\boldsymbol{\mu}$ είναι το:

$$\left(\bar{X} - Z_{0,975}\frac{\sigma}{\sqrt{n}},\ \ \bar{X} + Z_{0,975}\frac{\sigma}{\sqrt{n}}\right) = \left(\bar{X} - 1,96\frac{\sigma}{\sqrt{n}}, \bar{X} + 1,96\frac{\sigma}{\sqrt{n}}\right)$$

Παράδειγμα

Μια ναυτιλιακή εταιρεία έχει 310 υπαλλήλους. Η εταιρεία δεν αποκαλύπτει τον πίνακα των αποδοχών του προσωπικού της. Συντάκτης οικονομικού περιοδικού που διεξάγει έρευνα σχετικά με τον μέσο μισθό των υπαλλήλων των ναυτιλιακή εταιρειών έχει την πληροφορία ότι οι μισθοί των υπαλλήλων της συγκεκριμένης εταιρείας έχουν τυπική απόκλιση **σ=14,837**. Έτσι η μόνη λύση που έχει ο συντάκτης προκειμένου να πραγματοποιήσει τη έρευνά του, είναι να επιλέξει ένα τυχαίο δείγμα εργαζομένων και να τους ρωτήσει για το επίπεδο των ετησίων αποδοχών τους. Προκειμένου να είναι σε θέση να εφαρμόσει το ΚΟΘ επιλέγει δείγμα **40** υπαλλήλων. Αν ο μέσος του δείγματος είναι **29500€** ποιο είναι **95%** διάστημα εμπιστοσύνης για τη μέση τιμή των μισθών της εταιρείας;

Παράδειγμα

Λύση: Η πραγματική μέση τιμή ***μ*** των μισθών των υπαλλήλων της εταιρείας αναμένεται με πιθανότητα 95% να βρίσκεται στο διάστημα

$$\left(\bar{X} - 1{,}96\frac{\sigma}{\sqrt{n}}, \qquad \bar{X} + 1{,}96\frac{\sigma}{\sqrt{n}}\right)$$

ή $$\left(29.500 - 1{,}96\frac{14{,}837}{\sqrt{40}}, \quad 29.500 + 1{,}96\frac{14{,}837}{\sqrt{40}}\right)$$

ή $$(24.902€, \quad 34.098€)$$

Ο **ΤΥΠΟΣ** για το Διάστημα εμπιστοσύνης σε μεγάλα δείγματα (n>30)

Για οποιοδήποτε **επίπεδο εμπιστοσύνης 1-α**, ο τύπος διαμορφώνεται ως εξής:

$$\left(\overline{x} - z_{1-\frac{a}{2}} \cdot \frac{s}{\sqrt{n}}, \quad \overline{x} + z_{1-\frac{a}{2}} \cdot \frac{s}{\sqrt{n}} \right)$$

Παράδειγμα

Ποιο είναι το διάστημα εμπιστοσύνης 98% ;

Λύση: Αφού 1-α = 0,98 τότε α = 0,02.

Άρα α/2 = 0,01 και 1 - α/2 = 0,99

Ψάχνουμε στον πίνακα της τυπικής κανονικής κατανομής το $z_{0,99}$.

Αυτό βρίσκεται στη διασταύρωση της γραμμής 2,3 με τη στήλη 0,03.

Άρα $z_{0,99} = 2,33$. Οπότε το διάστημα εμπιστοσύνης είναι:

$$\left(\bar{x} - 2,33 \frac{s}{\sqrt{n}}, \bar{x} + 2,33 \frac{s}{\sqrt{n}} \right)$$

Εκτίμηση Διαστήματος Εμπιστοσύνης Ποσοστού

Ο λόγος του πλήθους των στοιχείων ενός πληθυσμού που πληρούν μια συνθήκη προς το σύνολο των στοιχείων του πληθυσμού λέγεται αναλογία της p. Στην περίπτωση αυτή έχουμε μια δίτιμη τυχαία μεταβλητή που ακολουθεί την κατανομή Bernoulli. «Επιτυχία» υπάρχει όταν πληρούται η συνθήκη και «αποτυχία» όταν δεν πλητούται η συνθήκη. Γνωρίζουμε ότι η μέση τιμή της κατανομής Bernoulli είναι p και η τυπική απόκλιση $\sqrt{p(1-p)}$.

Εκτίμηση Διαστήματος Εμπιστοσύνης Ποσοστού

Έτσι για οποιοδήποτε **επίπεδο εμπιστοσύνης 1-α**, ο τύπος του διαστήματος εμπιστοσύνης διαμορφώνεται ως εξής:

$$\left(p - z_{1-\frac{a}{2}} \cdot \sqrt{\frac{p(1-p)}{n}}, \quad p + z_{1-\frac{a}{2}} \cdot \sqrt{\frac{p(1-p)}{n}} \right)$$

Παράδειγμα

Σε ένα ερωτηματολόγιο για την διερεύνηση της κοινής γνώμης, από τα 1500 άτομα που ρωτήθηκαν τα 1050 ήταν υπέρ του υποψηφίου Α. Να βρεθεί το διάστημα εμπιστοσύνης του πληθυσμού που προτιμούν τον υποψήφιο Α, με πιθανότητα 98%.

Λύση: Αφού 1-α = 0,98 τότε α = 0,02. Άρα α/2 = 0,01, 1-α/2 = 0,99

Και από τον πίνακα της τυπικής κανονικής κατανομής το $z_{0,99} = 2,33$.

Ακόμη p=1050/1500=0,7. Οπότε το διάστημα εμπιστοσύνης είναι:

$$\left(0,7-2,33\cdot\sqrt{\frac{0,7(1-0,7)}{1500}},\quad 0,7+2,33\cdot\sqrt{\frac{0,7(1-0,7)}{1500}}\right)=$$

$$=(0,6725,\quad 0,7275)$$

Διάστημα εμπιστοσύνης σε μικρά δείγματα

Όταν το μέγεθος του δείγματος που χρησιμοποιούμε είναι μικρό (πρακτικά n≤30) τότε αντί της κανονικής χρησιμοποιούμε την κατανομή Student. Το διάστημα εμπιστοσύνης σε επίπεδο 1-α για την μέση τιμή είναι:

$$\left(\overline{x} - t_{n-1;\frac{a}{2}} \cdot \frac{s}{\sqrt{n}}, \quad \overline{x} + t_{n-1;\frac{a}{2}} \cdot \frac{s}{\sqrt{n}} \right)$$

Όπου n-1 είναι οι βαθμοί ελευθερίας όταν έχουμε δείγμα μεγέθους n και α/2 είναι το ποσοστό των τιμών της δεξιάς ουράς της t με n-1 βαθμούς ελευθερίας

Παράδειγμα

Κατά τη διάρκεια μιας ημερήσιας εκδρομής μαθητών Γυμνασίου ο κάθε μαθητής έχει μαζί του ορισμένα χρήματα. Ας υποθέσουμε ότι από τη μελέτη ενός τυχαίου δείγματος 20 μαθητών προέκυψε ότι ο μέσος όρος των χρημάτων που έχουν μαζί τους είναι 7€ και η τυπική απόκλιση 3€. Υπολογίστε ένα διάστημα εμπιστοσύνης 98% για τη μέση τιμή των χρημάτων που έχουν μαζί τους οι μαθητές.

Λύση: Υποθέτουμε ότι η κατανομή των χρημάτων που φέρουν μαζί τους οι μαθητές ακολουθεί την κανονική κατανομή. Αφού $1-\alpha = 0,98$ τότε $\alpha = 0,02$. Άρα $\alpha/2 = 0,01$. Επειδή το δείγμα είναι μικρό χρησιμοποιούμε τον τύπο με την κατανομή Student για $n-1=20-1=19$ βαθμούς ελευθερίας. Οπότε το διάστημα εμπιστοσύνης είναι:

Παράδειγμα

$$\left(\bar{x} - t_{19;0,01} \cdot \frac{s}{\sqrt{n}},\quad \bar{x} + t_{19;0,01} \cdot \frac{s}{\sqrt{n}}\right) =$$

$$= \left(7 - 2{,}539 \cdot \frac{3}{\sqrt{20}},\quad 7 + 2{,}539 \cdot \frac{3}{\sqrt{20}}\right) =$$

$$= (5{,}27,\quad 8{,}7)$$

Δηλαδή, με βεβαιότητα 98% η μέση τιμή των χρημάτων που έχουν μαζί τους οι μαθητές κυμαίνεται από 5,27€ έως 8,7€.

Προσδιορισμός μεγέθους δείγματος

Ζητούμε να προσδιορίσουμε το μέγεθος του δείγματος *n* που απαιτείται, για να επιτύχουμε την επιθυμητή ακρίβεια στο διάστημα εμπιστοσύνης του μέσου.

Έχουμε $Z_{\alpha/2} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ απ' όπου $\bar{X} - \mu = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Το $e = \bar{X} - \mu$ ονομάζεται **δειγματοληπτικό σφάλμα** (sampling error) και σε κάθε έρευνα ορίζουμε την μέγιστη τιμή που επιθυμούμε να έχει σε συγκεκριμένο επίπεδο εμπιστοσύνης 1-α.

Λύνοντας την σχέση $e = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ ως προς *n* έχουμε: $n = \frac{Z_{a/2}^2 \cdot \sigma^2}{e^2}$

Το επίμαχο σημείο στην πρακτική εφαρμογή της προηγούμενης σχέσης είναι η εκτίμηση της σ. Αυτό συνήθως γίνεται διεξάγοντας πρώτα μιας μικρής κλίμακας έρευνα, που ονομάζεται *πιλοτική έρευνα*, με σκοπό να εκτιμήσουμε την άγνωστη σ από την τυπική απόκλιση *s* του δείγματος.