ΑΡΙΣΤΟΤΕΛΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΟΝΙΚΗΣ

ΑΝΟΙΚΤΑ
ΑΚΑΔΗΜΑΪΚΑ
ΜΑΘΗΜΑΤΑ

Μαθηματικά Και Στατιστική Στη Βιολογία

Ενότητα 5 : Εκτιμήσεις

Ι. Αντωνίου, Χ. Μπράτσας
Τμήμα Μαθηματικών

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Εκτιμήσεις

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Περιεχόμενα Ενότητας (1 από 2)

Περιεχόμενα Ενότητας (2 από 2)

12. Διάστημα Εμπιστοσύνης για Διασπορά: Κανονική Κατανομή
13. X^2 Κατανομή
14. Διάστημα Εμπιστοσύνης X^2
15. Διάστημα Εμπιστοσύνης για την Αναλόγια p
16. Εκτίμηση διαστήματος εμπιστοσύνης μέσης τιμής
17. Διάστημα Εμπιστοσύνης για την Διαφορά των Μέσων Τιμών
18. Υπολογισμός του Τετραγώνου Διαφοράς
19. Διάστημα Εμπιστοσύνης για τον Λόγο Διασπορών
20. F Κατανομή
21. Διάστημα Εμπιστοσύνης Λόγου Διασπορών
22. Διάστημα Εμπιστοσύνης για τη Διαφορά Αναλογιών

Σκοποί Ενότητας

- Στην Ενότητα 5 παρουσιάζεται μια Εισαγωγή στην εκτίμηση των Παραμέτρων του Πληθυσμού από το Δείγμα.

Εκτιμήσεις των Παραμέτρων του Πληθυσμού από το Δείγμα

- **ΕΚΤΙΜΗΣΕΙΣ των Παραμέτρων του Πληθυσμού από το Δείγμα**

  = ο Δειγματοχώρος

  θ η παράμετρος του Πληθυσμού με τιμές στον Παραμετρικό Χώρο Θ

  $\tilde{\theta}$ **η αντίστοιχη παράμετρος του Δείγματος**

  $\hat{\theta}$ **η Εκτίμηση της παραμέτρου θ του Πληθυσμού από το Δείγμα.**

  Η Εκτίμηση της Παραμέτρου θ από το Δείγμα μπορεί να είναι Μια τιμή $\hat{\theta}$ η οποία πρέπει να είναι «κοντά» στην θ.

- **Εκτίμηση Σημείων = Point Estimation** ή

  Ένα διάστημα $\hat{\mathcal{D}}$ στο οποίο ανήκει η θ “με μεγάλη πιθανότητα”

- **Εκτίμηση Διαστημάτων = Interval Estimation**
- **Εκτιμήτρια (Estimator)**: Αλγόριθμος Υπολογισμού Εκτιμήσεων

  Κάθε Εκτιμήτρια είναι ένας Κανόνας Λήψης Αποφάσεων, μια Στρατηγική Παιγνίου.

Εκτίμηση Σημείων

- **Εκτίμηση Σημείων = Point Estimation**
- **Σημειακή Εκτιμήτρια (Συνάρτηση) = Point Estimator**
- : $\longrightarrow$: $(\xi_1, \xi_2,\ldots, \xi_N) \longmapsto$ $(\xi_1, \xi_2,\ldots, \xi_N)= \hat{\theta}$

  η Εκτίμηση Estimation της Παραμέτρου θ από το δείγμα μέσω της Εκτιμήτριας (Συνάρτησης)

  η Εκτιμήτρια είναι δειγματοσυνάρτηση - Τυχαία Μεταβλητή

  από τον Δειγματικό Χώρο στον Παραμετρικό Χώρο Θ
- **Παράδειγμα**

  Η Μέση τιμή του δείγματος

$$\widetilde{m}\,(\xi_1, \xi_2,\ldots, \xi_N) = \frac{\xi_1+\cdots+\xi_N}{N} = \widehat{m}$$

Είναι εκτίμηση της μέσης Τιμής m της Κατανομής του Πληθυσμού

$\widetilde{m}\,(\xi_1, \xi_2,\ldots, \xi_N)$ Εκτιμήτρια της μέσης Τιμής m της Κατανομής του Πληθυσμού.

Κριτήρια Αξιολόγησης- Επιλογής Σημειακών Εκτιμητριών (1 από 3)

- **Κριτήρια Αξιολόγησης- Επιλογής Σημειακών Εκτιμητριών**
- **Μεροληψία Ελάχιστη = Minimal Bias**
- **Μεροληψία : Bias( ) = E[(X_1,…, X_N)] − θ**
- **Παράδειγμα**

1. Η Τυπική Απόκλιση δείγματος δεν είναι Αμερόληπτη Εκτιμήτρια της Τυπικής Απόκλισης του Πληθυσμού

$$\tilde{\sigma} = \sqrt{\tilde{\sigma}_2} = \sqrt{\frac{(\xi_1-\tilde{m})^2+\cdots+(\xi_N-\tilde{m})^2}{N}}$$

2. Η διόρθωση Bessel 1830: (N – 1) αντί N, δίδει την Αμερόληπτη Εκτιμήτρια της Τυπικής Απόκλισης του Πληθυσμού

$$s = \sqrt{\frac{(\xi_1-\tilde{m})^2+\cdots+(\xi_N-\tilde{m})^2}{N-1}}$$

$\tilde{m}$ = η Μεση Τιμη του Δειγματος

Απόδειξη Άσκηση

Κριτήρια Αξιολόγησης- Επιλογής Σημειακών Εκτιμητριών (2 από 3)

- **Επάρκεια = Sufficiency, Fisher**

  Επιπλέον πληροφορία από την εκτίμηση (ξ_1, ξ_2,…, ξ_N)= δεν βελτιώνει την εκτίμηση της θ.

- Αποτελεσματικότης = Efficiency ⟺ Διασπορά Ελάχιστη = Minimal Dispersion

  E[((X_1,…, X_N) – θ)^2] minimal

  Ποια η Ελάχιστη τιμή της Διασποράς;

- **Θεώρημα**

  Ανισότητα Cramer – Rao

  Η διασπορά του Αμερόληπτης Εκτιμήτριας (X_1,…, X_N) είναι:

$$\mathrm{E}[(\partial(\mathrm{X}_1,\ldots,\mathrm{X}_\nu)-\theta)^2] \geq \frac{1}{J_{FISHER}}$$

η Πληροφορία Fisher

$$J_{FISHER} = J_{FISHER}(\theta) = \int dx \rho(\mathrm{x};\theta)(\frac{\partial}{\partial\theta}\ln\rho(\mathrm{x};\theta))^2$$

- Η Αμερόληπτη Εκτιμήτρια ϑ(X_1,…, X_N) είναι Αποτελεσματική ⟺

$$\mathrm{E}[(\partial(\mathrm{X}_1,\ldots,\mathrm{X}_\mathrm{M})-\theta)^2] = \frac{1}{J_{FISHER}}$$

Κριτήρια Αξιολόγησης- Επιλογής Σημειακών Εκτιμητριών (3 από 3)

- **Ασυμπτωτική Αποτελεσματικότητα = Efficiency ⟺ Διασπορά Ελάχιστη = Minimal Dispersion**

Η Αμερόληπτη Εκτιμήτριας (X1,…, XN) είναι Ασυμπτωτικά Αποτελεσματική

$$\Leftrightarrow \lim_{M \to +\infty} E(\partial(\mathrm{X}_1, \ldots, X_{\mathrm{N}}) - \theta)^2] = 0$$

- **Συνέπεια = Σύγκλιση = Consistency = Convergence**

$$\lim_{M \to +\infty} \partial(\mathrm{X}_1, \ldots, X_{\mathrm{N}}) = \theta$$

- **Ασυμπτωτικά Κανονική Εκτιμήτρια της θ**

⟺ Η Ασυμπτωτική Κατανομή της √ ($_{N}$ – θ) είναι η Τυποποιημένη κανονική N(0,σ^2(θ)).

Μέθοδοι Κατασκευής Σημειακών Εκτιμητριών

- **Μέθοδος Ροπών Pearson**

Οι ροπές του δείγματος ως εκτιμήτριες:

$$\widetilde{m}_r(\xi_1, \xi_2, \ldots, \xi_N) = \frac{(\xi_1)^r + \cdots + (\xi_N)^r}{N} = \widehat{m}_r$$

1. Η επίλυση των εξισώσεων είναι σχετικά εύκολη.

$$\widetilde{m}_1 = g_1(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n)$$

$$\widetilde{m}_2 = g_1(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n)$$

$$\vdots$$

$$\widetilde{m}_n = g_1(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n)$$

2. Οι εκτιμήτριες $\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_n$ ως συναρτήσεις των $\widetilde{\mu}_1, \widetilde{\mu}_2, \ldots, \widetilde{\mu}_n$ είναι **συνεπείς.**
Κάθε εκτιμήτρια $\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_n$ εκφράζεται ως συνεχής συνάρτηση.

3. Ο Fisher έδειξε ότι οι εκτιμήτριες $\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_n$ δεν είναι **αποτελεσματικές (efficient)**.

Μέθοδος Ελαχίστων Τετραγώνων

- **Μέθοδος Ελαχίστων τετραγώνων Gauss 1809**

Η Εκτίμηση $\hat{\theta}$ ης παραμέτρου θ του Πληθυσμού από το Δείγμα Καθιστά Ελάχιστο το άθροισμα των τετραγωνικών σφαλμάτων

$(\psi_1 - g(\xi_1,\theta))^2 + \ldots + (\psi_N - g(\xi_N,\theta))^2$

Της Προσέγγισης της Υ από την Χ

Εκτίμηση Διαστημάτων

- **Εκτίμηση Διαστημάτων = Interval Estimation**

  Διάστημα Εμπιστοσύνης βαθμού c για την Εκτίμηση της παραμέτρου θ από το Δείγμα καλείται κάθε διάστημα

  (A,B): $P[A < \theta < B] = c$

  Τα άκρα A, B του Διαστήματος είναι Μεταβλητές που εξαρτώνται από το Δείγμα

  $A = A(\xi_1, \xi_2, \ldots, \xi_N)$
  $B = B(\xi_1, \xi_2, \ldots, \xi_N)$
  Ο βαθμός Εμπιστοσύνης c του διαστήματος (A,B) (Confidence Level) πρέπει να είναι κοντά στο 1 για να είναι αξιόπιστη η Εκτίμηση.

- Neyman J. 1937, Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability, Philosophical Transactions of the Royal Society of London A236, 333–380.

Interval Estimation Methods

Interval Estimation Methods		
confidence intervals	Frequentistic method Cannot readily deal with prior information	Statistical methods
credible intervals	Bayesian method Can readily deal with prior information	
Prediction intervals	Regression Analysis	
Likelihood intervals		
fuzzy logic	application-specific	Non-statistical method

Διάστημα Εμπιστοσύνης (1 από 10)

- **Δείγμα από ένα Πληθυσμό**
- **Διάστημα Εμπιστοσύνης για την Εκτίμηση**
  Μέσης Τιμής μ
  Διασποράς σ^2
  Αναλογίας p μελών Πληθυσμού που έχουν επιθυμητή Ιδιότητα
- **Δείγμα από 2 Πληθυσμούς**
- **Διάστημα Εμπιστοσύνης για την Εκτίμηση**
- **Διαφοράς των Μέσων Τιμών $\mu_X - \mu_Y$**
- Λόγου $\frac{\sigma_X^2}{\sigma_Y^2}$ των διασπορών $\sigma_X^2,\ \sigma_Y^2$
- **Διαφοράς $p_X - p_Y$ των Αναλογιών p_X , p_Y των μελών που έχουν επιθυμητές Ιδιότητες**

Διάστημα Εμπιστοσύνης (2 από 10)

- **Τα προβλήματα ανάγονται στα ποσοστημόρια των Κατανομών**
  Τυποποιημένη Κανονικη (μ=0, σ=1)
  X^2 (chi square)
  Student – t
  F (Fisher)

Διάστημα Εμπιστοσύνης (3 από 10)

- **Διάστημα Εμπιστοσύνης για την Εκτίμηση της Μεσης Τιμής**
- **Δείγμα από Πληθυσμό με Κανονική Κατανομή**
- **άγνωστη Μέση Τιμή μ και γνώστη διασπορά σ^2**

$$\rho(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

E[X]=μ
v[X]=σ^2

Διάστημα Εμπιστοσύνης (4 από 10)

- **Λήμμα 1**

  Έστω $\xi_1, \ldots, \xi_N$ τυχαίο δείγμα από Πληθυσμό με κανονική κατανομή με μέσο μ και διασπορά σ^2

  Η δειγματική Μέση Τιμή

$$\widetilde{m} = \frac{\xi_1 + \cdots + \xi_N}{N}$$

  ακολουθεί την κανονική κατανομή με Μέση Τιμή μ και διασπορά $\dfrac{\sigma^2}{N}$

Διάστημα Εμπιστοσύνης (5 από 10)

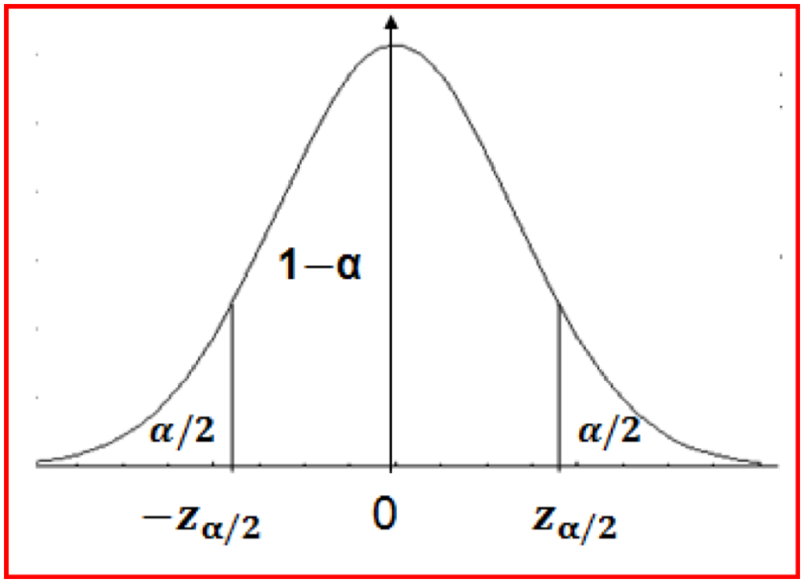
- **Λήμμα 2**

  Η Μεταβλητή $z = \dfrac{x-\mu}{\sigma\sqrt{\mathrm{N}}}$ ακολουθεί την τυποποιημένη κανονική κατανομή

$$\rho(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \qquad x \in \mathbb{R}$$

E[X]=0
v[X]=1

Διάστημα Εμπιστοσύνης (6 από 10)



- Λήμμα 3

P[− /2 < z < /2] = 1− α

/ το άνω ποσοστημόριο της τυποποιημένης κανονικής Κατανομής

Διάστημα Εμπιστοσύνης (7 από 10)

- **Από την σχέση των ποσοστημορίων προκύπτει το Διάστημα Εμπιστοσύνης**

$$P\left[-z_{\alpha/2} < \frac{\tilde{m}-\mu}{\sigma\sqrt{N}} < z_{\alpha/2}\right]=1-a$$

- **Θεώρημα**

$$P\left[\tilde{m} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}} < \mu < \tilde{m} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}\right] = 1-\alpha$$

όπου

$$A = \tilde{m} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}}$$

$$B = \tilde{m} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}$$

οι εκτιμήσεις των ακρών του διαστήματος από το Δείγμα.

Διάστημα Εμπιστοσύνης (8 από 10)

- Το Εύρος του διαστήματος Εμπιστοσύνης:

  $$|B - A| = 2z_{a/2}\frac{\sigma}{\sqrt{N}}$$

  Είναι εκτίμηση της ακριβείας εκτίμησης

  $|\tilde{m} - \mu| \leq z_{\alpha/2}\frac{\sigma}{\sqrt{N}}$ με εμπιστοσύνη c = (1− α)%.

Διάστημα Εμπιστοσύνης (9 από 10)

Παραδειγμα

c=0.95	c=0.99
$\alpha = 1 - c = 0.05$	$\alpha = 1 - c = 0.01$
$z_{\alpha/2} = z_{0.025} = 1.96$	$z_{\alpha/2} = z_{0.005} = 2.58$
$A = \tilde{m} - 1,96 \frac{\sigma}{\sqrt{N}}$	$A = \tilde{m} - 2.58 \frac{\sigma}{\sqrt{N}}$
$B = \tilde{m} + 1,96 \frac{\sigma}{\sqrt{N}}$	$B = \tilde{m} + 2.58 \frac{\sigma}{\sqrt{N}}$
Ευρος $\|A-B\| = 3,92 \frac{\sigma}{\sqrt{N}}$	Ευρος $\|A-B\| = 5,16 \frac{\sigma}{\sqrt{N}}$

Αύξηση Εμπιστοσύνης από 95% σε 99% **(κατά 4.4%)**
Αυξάνει το Διάστημα Εμπιστοσύνης από 3,92 σε 5,16 **(κατά 31.63%)**

Διάστημα Εμπιστοσύνης (10 από 10)

- Πόσες μετρήσεις Κ πέραν (από τις Ν) πρέπει να πάρω για να έχω Διάστημα Εμπιστοσύνης του αυτού Εύρους $3.92\frac{\sigma}{\sqrt{N}}$ με βαθμό Εμπιστοσύνης 99%;

Αρκεί:

$$5.16\frac{\sigma}{\sqrt{N+K}} = 3.92\frac{\sigma}{\sqrt{N}}$$

$$\Leftrightarrow \frac{5.16}{3.92} = \frac{\sqrt{N+K}}{\sqrt{N}}$$

$$\Leftrightarrow (\frac{5.16}{3.92})^2 = \frac{N+K}{N}$$

$$\Leftrightarrow K = N[(\frac{5.16}{3.92})^2 - 1]$$

Παράδειγμα

- Δηλαδή $K = N[(\frac{5.16}{3.92})^2 - 1]\% \cong 0.73$ = 73% αύξηση των μετρήσεων. Για αύξηση Εμπιστοσύνης κατά 4.4% στο αυτό Διάστημα Εμπιστοσύνης.

Άγνωστη Μέση Τιμή μ και άγνωστη διασπορά σ^2

Δείγμα “Μεγάλο” N≥30

- **Θεώρημα**

$$P[\, \tilde{m} - z_{\alpha/2}\frac{s}{\sqrt{N}} < \mu < \tilde{m} + z_{\alpha/2}\frac{s}{\sqrt{N}} \,] \approx 1-\alpha$$

$s = \sqrt{\frac{(\xi_1-\tilde{m})^2+\cdots+(\xi_N-\tilde{m})^2}{N-1}}$ η Αμερόληπτη Τυπική Απόκλιση του δείγματος για μεγάλα δείγματα: σ≈s

Κεντρικό Οριακό Θεώρημα

- **Κεντρικό Οριακό Θεώρημα**

  Η κατανομή της Μέσης Τιμής $\frac{x_1 + x_2,....,x_N}{N}$ ανεξάρτητων μεταβλητών X_1 ,X_2 ,…,X_N προσεγγίζει την Κανονική, όσο αυξάνει ο αριθμός Ν των Μεταβλητών.

Κατανομή Student (1 από 2)

- **Άγνωστη Μέση Τιμή μ και άγνωστη διασπορά σ2**

Λήμμα

Έστω $\xi_1, \ldots, \xi_N$ τυχαίο δείγμα από Πληθυσμό με κανονική κατανομή με μέσο μ και διασπορά σ^2

Η μεταβλητή $T = \frac{\widetilde{m}-\mu}{s\sqrt{N}}$ ακολουθεί την κατανομή Student – t βαθμού ν= N − 1

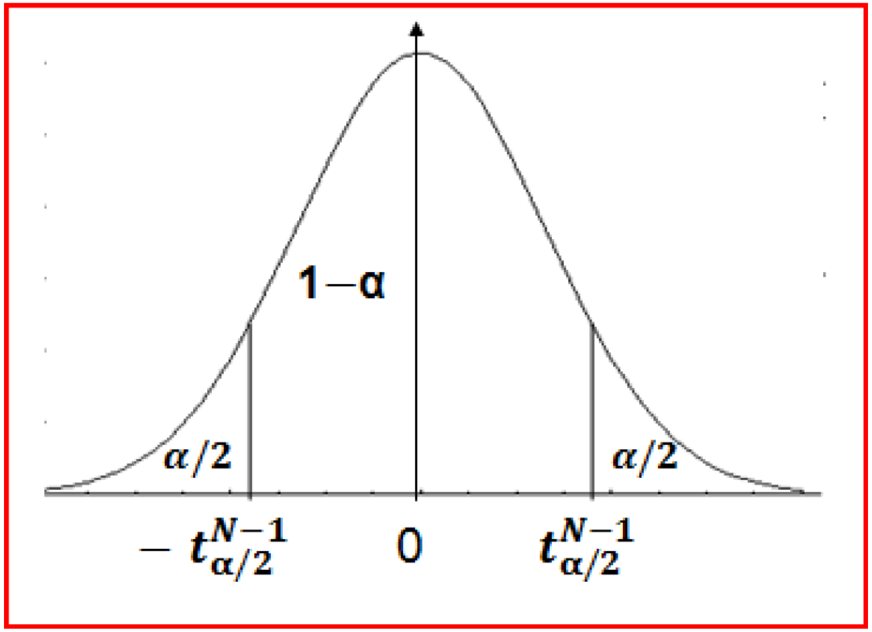
$$s = \sqrt{\frac{(\xi_1-\widetilde{m})^2+\cdots+(\xi_N-\widetilde{m})^2}{N-1}}$$ **η Αμερόληπτη Τυπική Απόκλιση του δείγματος**

$$\rho_\nu(t) = \frac{1}{\sqrt{\nu\pi}}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, t \in \mathbb{R}$$

Γ(z) η συνάρτηση Γ

Η κατανομή $\rho_\nu(t)$ δεν εξαρτάται από τις παραμέτρους μ και σ

Κατανομή Student
(2 από 2)



- **Λήμμα 2**

P[$- t_{\alpha/2}^{N-1}$ < t < $t_{\alpha/2}^{N-1}$] = 1− α

t= η τιμη της μεταβλητης T = $\frac{m-\mu}{s\sqrt{N}}$

$t_{\alpha/2}^{N-1}$ το ανω ποσοστιαιο σημειο

Κατανομή Student: Διάστημα Εμπιστοσύνης

- **Από την σχέση των ποσοστημορίων προκύπτει το Διάστημα Εμπιστοσύνης**

$$\mathrm{P}[- t_{\alpha/2}^{N-1} < \frac{\widetilde{m}-\mu}{s\sqrt{N}} < t_{\alpha/2}^{N-1}] = 1-\alpha$$

- **Θεώρημα**

$$\mathrm{P}\left[\widetilde{m} - t_{\alpha/2}^{N-1}\frac{s}{\sqrt{N}} < \mu < \widetilde{m} + t_{\alpha/2}^{N-1}\frac{s}{\sqrt{N}}\right] = 1-\alpha$$

Διάστημα Εμπιστοσύνης για Διασπορά: Κανονική Κατανομή

- **Διάστημα Εμπιστοσύνης για την Διασπορά Πληθυσμού με Κανονική Κατανομή**
  Λημμα 1
  Έστω $\xi_1, \ldots, \xi_N$ τυχαίο δείγμα από Πληθυσμό με κανονική κατανομή με μέσο μ και διασπορά σ^2

  Η μεταβλητή $X^2 = \frac{(N-1)s^2}{\sigma^2}$ ακολουθεί την κατανομή Χι – τετράγωνο βαθμού $\nu = N - 1$

  $$s^2 = \frac{(\xi_1 - \tilde{m})^2 + \cdots + (\xi_N - \tilde{m})^2}{N-1}$$

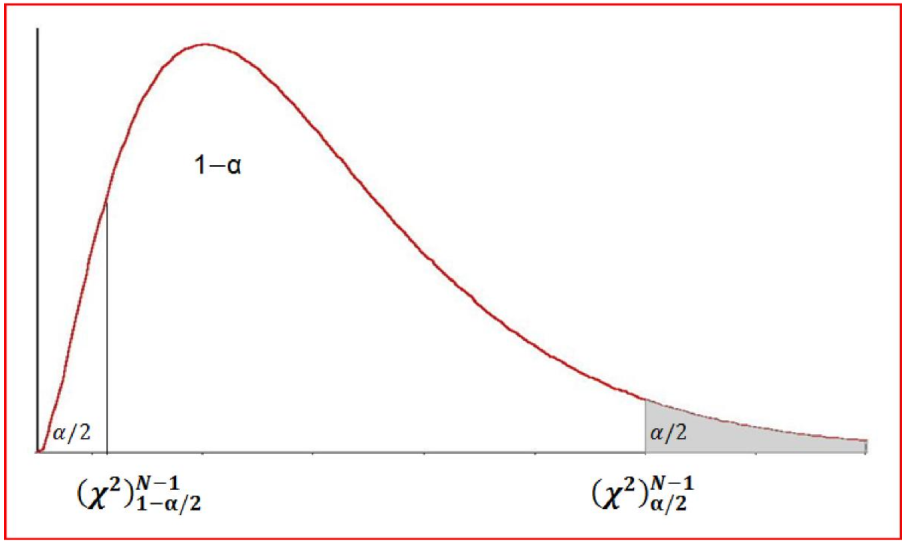
  η αμερόληπτη διασπορά του Δείγματος

  $$\rho_\nu(x) = \left(\frac{1}{2}\right)^{\frac{\nu}{2}} \frac{1}{\Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} e^{-\frac{\nu}{2}}$$

  Γ(z) η συνάρτηση Γ
  Η κατανομή $\rho_\nu(x)$ δεν εξαρτάται από τις παραμέτρους μ και σ

X^2 Κατανομή

- **Λήμμα 2**



x^2 = η τιμη της μεταβλητης $X^2 = \frac{(N-1)s^2}{\sigma^2}$

$\chi^2 \, {}^{N-1}_{\alpha/2}$ **το ανω ποσοστημοριο**

Διάστημα Εμπιστοσύνης Χ^2

Από την σχέση των ποσοστιαίων σημείων προκύπτει το Διάστημα Εμπιστοσύνης

$$P \left((\chi^2)_{1-\alpha/2}^{N-1} < \frac{(N-1)s^2}{\sigma^2} < (\chi^2)_{\alpha/2}^{N-1} \right) = 1 - a$$

Θεωρημα

$$P \left(\frac{(N-1)s^2}{\chi^2{}_{1-\alpha/2}^{N-1}} < \sigma^2 < \frac{(N-1)s^2}{\chi^2{}_{\alpha/2}^{N-1}} \right) = 1-\alpha$$

$$P \left(\sqrt{\frac{(N-1)s^2}{\chi^2{}_{1-\alpha/2}^{N-1}}} < \sigma < \sqrt{\frac{(N-1)s^2}{\chi^2{}_{\alpha/2}^{N-1}}} \right) = 1-\alpha$$

Διάστημα Εμπιστοσύνης για την Αναλόγια p

- **Διάστημα Εμπιστοσύνης για την Αναλόγια p μελών Πληθυσμού που έχουν επιθυμητή Ιδιότητα Q**

Δείγμα από Πληθυσμό μεγάλο (N>30)

Ληmμα

1) η δειγματική αναλογία $\tilde{p} = \frac{N_Q}{N}$ είναι Αμερόληπτη Εκτιμήτρια $E[\tilde{p}]=p$

με διασπορά $\sigma^2 = \frac{p(1-p)}{N}$

2) για μεγάλο N η Μεταβλητή

$$Z = \frac{x-\mu}{\sigma\sqrt{N}} = \frac{\tilde{p}-p}{\sqrt{\frac{p(1-p)}{N}}}$$

είναι Τυποποιημένη Κανονική

Εκτίμηση διαστήματος εμπιστοσύνης μέσης τιμής

- **Κεντρικό Οριακό Θεώρημα**

Η εκτίμηση ανάγεται στην Εκτίμηση Διαστημάτων Εμπιστοσύνης για την Εκτίμηση της Μέσης Τιμής.

$$\mathrm{P}\left[\, \boldsymbol{m} - \boldsymbol{z_{a/2}} \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{N}}} < \mathrm{p} < \boldsymbol{m} + \boldsymbol{z_{a/2}} \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{N}}} \,\right] = 1 - \alpha$$

$\mu = \mathrm{p}$

$$\boldsymbol{m} = \boldsymbol{p}$$

$$\frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{N}}} = \sqrt{\frac{\boldsymbol{p}(\mathbf{1}-\boldsymbol{p})}{\boldsymbol{N}}}$$

Διάστημα Εμπιστοσύνης για την Διαφορά των Μέσων Τιμών (1 από 8)

- **Διάστημα Εμπιστοσύνης για την Διαφορά των Μέσων Τιμών 2 Ανεξάρτητων Πληθυσμών με Κανονικές Κατανομές**

Λαμβάνονται 2 ανεξάρτητα Δείγματα
$(X_1, \ldots, X_N) = (\xi_1, \ldots, \xi_N)$, $(Y_1, \ldots, Y_M) = (\eta_1, \ldots, \eta_M)$
Από 2 διαφορετικές Κανονικές κατανομές (μ_X, σ_X), (μ_Y, σ_Y)
$\tilde{m}_X$, $\tilde{m}_Y$ οι δειγματικοί Μεσοί Όροι.

- **Λήμμα**

Η διαφορά $\tilde{m} = \tilde{m}_X - \tilde{m}_Y$ ακολουθεί Κανονική Κατανομή
με παραμέτρους $\mu = \mu_X - \mu_Y$, $\sigma = \sqrt{\frac{\sigma_X^2}{N} + \frac{\sigma_Y^2}{M}}$
Η εκτίμηση ανάγεται στην Εκτίμηση Διαστημάτων Εμπιστοσύνης για την Εκτίμηση της Μέσης Τιμής.

Διάστημα Εμπιστοσύνης για την Διαφορά των Μέσων Τιμών (2 από 8)

- **Για γνώστες διασπορές** σ_X^2, σ_Y^2

$$P[\ \tilde{m} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}} < \mu < \tilde{m} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}\ ] = 1-\alpha$$

$$\mu = \mu_X - \mu_Y$$

$$\frac{\sigma}{\sqrt{N}} = \sqrt{\frac{\sigma_X^2}{N} + \frac{\sigma_Y^2}{M}}$$

$$P\left[\tilde{m}_X - \tilde{m}_Y - z_{a/2}\sqrt{\frac{\sigma_X^2}{N} + \frac{\sigma_Y^2}{M}} < \mu_X - \mu_Y < \tilde{m}_X - \tilde{m}_Y + z_{a/2}\sqrt{\frac{\sigma_X^2}{N} + \frac{\sigma_Y^2}{M}}\right] = 1-\alpha$$

Διάστημα Εμπιστοσύνης για την Διαφορά των Μέσων Τιμών (3 από 8)

- **Για άγνωστες διασπορές $\sigma_X^2,\ \sigma_Y^2$, μεγάλα δείγματα N,M≥ 30,**

$$P\left[\tilde{m} - z_{\alpha/2}\frac{s}{\sqrt{N}} < \mu < \tilde{m} + z_{\alpha/2}\frac{s}{\sqrt{N}}\right] \approx 1 - a$$

$$P\left[\tilde{m}_X - \tilde{m}_Y - z_{a/2}\sqrt{\frac{s_X^2}{N} + \frac{s_Y^2}{M}} < \mu_X - \mu_Y < \tilde{m}_X - \tilde{m}_Y + z_{a/2}\sqrt{\frac{s_X^2}{N} + \frac{s_Y^2}{M}}\right] = 1 - \alpha$$

Διάστημα Εμπιστοσύνης για την Διαφορά των Μέσων Τιμών (4 από 8)

- **Για άγνωστες διασπορές** $\sigma_X^2 = \sigma_Y^2$

$$\mathrm{P}\left[\tilde{m} - t_{\alpha/2}^{\nu}\frac{s}{\sqrt{N}} < \mu < \tilde{m} + t_{\alpha/2}^{\nu}\frac{s}{\sqrt{N}}\right] = 1 - \alpha$$

$$\tilde{m} = \tilde{m}_X - \tilde{m}_Y$$

$$\mu = \mu_X - \mu_Y$$

$$\frac{s}{\sqrt{N}} = \sqrt{\left(\frac{1}{N} + \frac{1}{M}\right)\frac{(N-1)s_X^2 + (M-1)s_Y^2}{N+M-2}}$$

$$\nu = N + M - 2$$

Διάστημα Εμπιστοσύνης για την Διαφορά των Μέσων Τιμών (5 από 8)

- **Για άγνωστες διασπορές** $\sigma_X^2 \neq \sigma_Y^2$ **και N=M**

$$P\left[\tilde{m} - t_{\alpha/2}^{\nu}\sqrt{\frac{s_X^2}{\nu}+\frac{s_Y^2}{\nu}} < \mu < \tilde{m} + t_{\alpha/2}^{\nu}\sqrt{\frac{s_X^2}{\nu}+\frac{s_Y^2}{\nu}}\right] = 1-\alpha$$

$$\tilde{m} = \tilde{m}_X - \tilde{m}_Y$$

$$\mu=\mu_X - \mu_Y$$

$$\frac{s}{\sqrt{N}} = \sqrt{\frac{s_X^2}{\nu}+\frac{s_Y^2}{\nu}}$$

$$\nu=2(N-1)$$

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Διάστημα Εμπιστοσύνης για την Διαφορά των Μέσων Τιμών (6 από 8)

- **Για άγνωστες διασπορές $\sigma_X^2 \neq \sigma_Y^2$ και N≠M**

$$P\left[\tilde{m} - t_{\alpha/2}^{\nu}\frac{s}{\sqrt{N}} < \mu < \tilde{m} + t_{\alpha/2}^{\nu}\frac{s}{\sqrt{N}}\right] = 1 - a$$

$$\tilde{m} = \tilde{m}_X - \tilde{m}_Y$$

$$\mu = \mu_X - \mu_Y$$

$$\frac{s}{\sqrt{N}} = \sqrt{\frac{s_X^2}{\nu} + \frac{s_Y^2}{\nu}}$$

$$\nu = \frac{\left(\frac{s_X^2}{N} + \frac{s_Y^2}{M}\right)^2}{\frac{\frac{s_X^2}{N}}{N-1} + \frac{\frac{s_Y^2}{M}}{M-1}}$$

Διάστημα Εμπιστοσύνης για την Διαφορά των Μέσων Τιμών (7 από 8)

- **Διάστημα Εμπιστοσύνης για την Διαφορά των Μέσων Τιμών 2 Αλληλοεξαρτώμενων Πληθυσμών με Κανονικές Κατανομές**

  Λαμβάνονται 2 αλληλοεξαρτώμενα Δείγματα

  $(X_1, \ldots, X_N)$, $(Y_1, \ldots, Y_M)$

  Από 2 διαφορετικές Κανονικές κατανομές (μ_X, σ_X) , (μ_Y, σ_Y), $\tilde{m}_X$, $\tilde{m}_Y$ οι δειγματικοί Μεσοί Όροι

  Λήμμα

  Οι διάφορες $Z_1 = X_1 - Y_1$, $Z_2 = X_2 - Y_2$, … $Z_N = X_N - Y_N$ είναι ανεξάρτητες.

Διάστημα Εμπιστοσύνης για την Διαφορά των Μέσων Τιμών (8 από 8)

Η εκτίμηση ανάγεται στην Εκτίμηση Διαστημάτων Εμπιστοσύνης για την Εκτίμηση της Μέσης Τιμής Δείγματος από κανονική Κατανομή άγνωστης Μέσης Τιμής μ και άγνωστης διασποράς $σ^2$

$$\mathrm{P}\left[\tilde{m} - t_{\alpha/2}^{N-1}\frac{s}{\sqrt{N}} < \mu < \tilde{m} + t_{\alpha/2}^{N-1}\frac{s}{\sqrt{N}}\right] = 1 - a$$

ΟΠΟΥ

$$\tilde{m} = \frac{\zeta_1 + \cdots + \zeta_N}{N}$$

$\zeta_1, \ldots, \zeta_N$ δειγμα μετρησης των μεταβλητων $Z_1 = X_1 - Y_1$, $Z_2 = X_2 - Y_2$, …, $Z_N = X_N - Y_N$

$$s = \sqrt{\frac{(\zeta_1 - \tilde{m})^2 + \cdots + (\zeta_N - \tilde{m})^2}{N-1}} = \sqrt{\frac{(\zeta_1)^2 + \cdots + (\zeta_N)^2 - N\tilde{m}^2}{N-1}}$$ η Αμεροληπτη Τυπικη Αποκλιση

Υπολογισμός του Τετραγώνου Διαφοράς

- **Λήμμα**

$$(\zeta_1 - \widetilde{m})^2 + \cdots + (\zeta_N - \widetilde{m})^2 = (\zeta_1)^2 + \cdots + (\zeta_N)^2 - N\widetilde{m}^2$$

Απόδειξη

$$(\zeta_1 - \widetilde{m})^2 + \cdots + (\zeta_N - \widetilde{m})^2 = (\zeta_1)^2 + \cdots + (\zeta_N)^2 + \mathrm{N}(\widetilde{m})^2 - 2(\zeta_1 + \cdots + \zeta_N)\widetilde{m}$$

$$= (\zeta_1)^2 + \cdots + (\zeta_N)^2 + \mathrm{N}(\widetilde{m})^2 - 2(N\widetilde{m})\widetilde{m} =$$

$$= (\zeta_1)^2 + \cdots + (\zeta_N)^2 - N\widetilde{m}^2$$

Διάστημα Εμπιστοσύνης για τον Λόγο Διασπορών

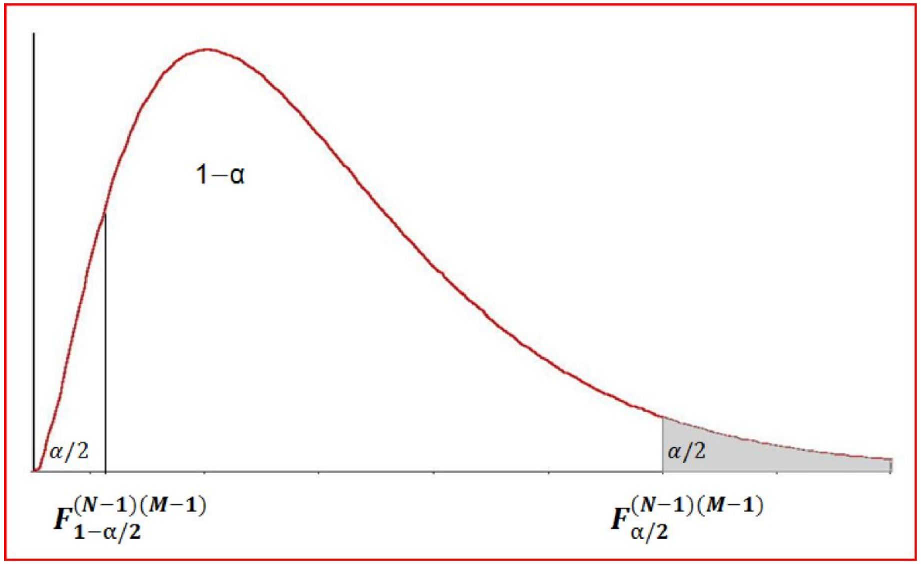
• **Διάστημα Εμπιστοσύνης για τον Λόγο $\frac{\sigma_X^2}{\sigma_Y^2}$ των Διασπορών 2 Ανεξάρτητων Πληθυσμών με Κανονικές Κατανομές $\sigma_X^2, \ \sigma_Y^2$**

Λήμμα
Η Μεταβλητή $\frac{s_X^2}{s_Y^2}\frac{\sigma_Y^2}{\sigma_X^2}$ των Διασπορών $\sigma_X^2, \ \sigma_Y^2$
2 Ανεξάρτητων Πληθυσμών με Κανονικές Κατανομές
Ακολουθεί Κατανομή F (Fisher) με βαθμούς $\nu_1 = N - 1$, $\nu_2 = N - 1$

$$\rho_{\nu_1\nu_2}(\mathrm{x}) = \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)+\Gamma\left(\frac{\nu_2}{2}\right)} \frac{x^{\frac{\nu_1-2}{2}}}{\left(1+\frac{\nu_1}{\nu_2}x\right)^{\frac{\nu_1+\nu_2}{2}}}, x \geq 0$$

Γ(z) η συναρτηση Γ
Η κατανομη δεν εξαρταται από τις παραμετρους μ_X , μ_Y και σ_X , σ_X

F Κατανομή



- **Λήμμα 2**

$$P\left(F_{1-\alpha/2}^{(N-1)(M-1)} < \mathbf{x} < F_{\alpha/2}^{(N-1)(M-1)} \right) = 1-\alpha$$

x^2 = η τιμη της μεταβλητης $\frac{s_X^2}{s_Y^2}\frac{\sigma_Y^2}{\sigma_X^2}$

$F_{\alpha/2}^{(N-1)(M-1)}$ **το ανω ποσοστημοριο**

Διάστημα Εμπιστοσύνης Λόγου Διασπορών

- **Από την σχέση των ποσοστιαίων σημείων προκύπτει το Διάστημα Εμπιστοσύνης**

$$P\left[F_{1-\alpha/2}^{(N-1)(\mathrm{M}-1)} < \frac{s_X^2}{s_Y^2}\frac{\sigma_Y^2}{\sigma_X^2} < F_{\alpha/2}^{(N-1)(\mathrm{M}-1)}\right] = 1-\alpha$$

- **Θεώρημα**

$$P\left[\frac{s_{\mathrm{Y}}^2}{s_{\mathrm{X}}^2}F_{1-\alpha/2}^{(N-1)(\mathrm{M}-1)} < \frac{\sigma_Y^2}{\sigma_X^2} < \frac{s_{\mathrm{Y}}^2}{s_{\mathrm{X}}^2}F_{\alpha/2}^{(N-1)(\mathrm{M}-1)}\right] = 1-a$$

Διάστημα Εμπιστοσύνης για τη Διαφορά Αναλογιών

Διάστημα Εμπιστοσύνης για τη Διαφορά $p_X - p_Y$ των Αναλογιών p_X, p_Y των μελών
2 Ανεξάρτητων Πληθυσμών που έχουν επιθυμητές Ιδιότητες.
Δείγματα από Πληθυσμό μεγάλο (N, M >30)

Κεντρικό Οριακό Θεώρημα
Η εκτίμηση ανάγεται στην Εκτίμηση Διαστημάτων Εμπιστοσύνης
για την Εκτίμηση της Μέσης Τιμής

$$P\left[\tilde{p}_X - \tilde{p}_Y - z_{\alpha/2}\frac{\sigma}{\sqrt{N}} < p_X - p_Y < \tilde{p}_X - \tilde{p}_Y + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}\right] = 1-\alpha$$

$$\frac{\sigma}{\sqrt{N}} = \sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{N} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{M}}$$

Σημείωμα Αδειοδότησης

Τέλος ενότητας

Επεξεργασία: Βασιλική Αλμπανίδου

Θεσσαλονίκη, Χειμερινό Εξάμηνο 2014-2015

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.