ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΘΕΩΡΙΑ ΑΠΟΦΑΣΕΩΝ

ΔΙΑΛΕΞΗ: ΔΕΙΓΜΑΤΙΚΕΣ ΚΑΤΑΝΟΜΕΣ - ΕΚΤΙΜΗΤΙΚΗ

Διδάσκουσα: Ε. Γάκη, Επίκ. Καθηγήτρια

Περιεχόμενα Μαθήματος

- Δειγματικές Κατανομές
- Σημειακές Εκτιμήσεις
- Μέθοδοι Εκτίμησης
- Διαστήματα Εμπιστοσύνης
- Παραδείγματα

Δειγματική Κατανομή

Δειγματική Κατανομή?????????

Γιατί χρειάζεται η δειγματοληψία?

Αντικειμενικός στόχος της δειγματοληψίας είναι η επιλογή αντιπροσωπευτικών δειγμάτων από ένα πληθυσμό και η εκτίμηση των τιμών μιας συγκεκριμένης παραμέτρου του πληθυσμού

Πως προκύπτει νοητικά η «δειγματική κατανομή?»

Υποθέτουμε ότι

1. μπορούμε να επιλέξουμε όλα τα δυνατά δείγματα από των πληθυσμό
2. υπολογίζουμε για καθένα από αυτά τη στατιστική συνάρτηση που χρειαζόμαστε
3. δημιουργούμε την κατανομή πιθανότητας των αποτελεσμάτων

Έτσι προκύπτει η **δειγματική κατανομή** (*sampling distribution*) της μιας συγκεκριμένης στατιστικής συνάρτησης.

Κάθε κατανομή πιθανότητας περιγράφετε ικανοποιητικά από
τον **μέσο** της &
την **τυπική** της **απόκλιση**

Στην περίπτωση της δειγματικής κατανομής μιας στατιστικής συνάρτησης
- ο μέσος αναφέρεται **ως μέσος της δειγματικής κατανομής**
- η τυπική απόκλιση ως **τυπικό σφάλμα** της στατιστικής συνάρτησης.

Πληθυσμός ή Δείγμα

X = τιμή του πληθυσμού που προκύπτει κατά την επιλογή του στοιχείου.

Η X είναι τυχαία μεταβλητή και η κατανομή πιθανότητας της ονομάζεται **κατανομή του πληθυσμού**.

Έστω ότι παίρνουμε από τον πληθυσμό με τρόπο τυχαίο n στοιχεία και ορίζουμε τις ακόλουθες μεταβλητές:

X = Η τιμή του πληθυσμού που προκύπτει από την i λήψη στοιχείου, i = 1, 2, … ,n

Αν η δειγματοληψία γίνεται από πεπερασμένο πληθυσμό με επανάθεση ή, ισοδύναμα, από άπειρο πληθυσμό τότε οι τυχαίες μεταβλητές $X_1, X_2, \ldots, X_n$ είναι ανεξάρτητες και έχουν κατανομή πιθανότητας την κατανομή του πληθυσμού, ονομάζεται δε **τυχαίο δείγμα** (*random sample*).
Επομένως, μπορούμε γενικά να πούμε ότι αν X είναι μία τυχαία μεταβλητή με ορισμένη κατανομή πιθανότητας και $X_1, X_2, \ldots, X_n$ ανεξάρτητες τυχαίες μεταβλητές κάθε μία από τις οποίες έχει την ίδια κατανομή πιθανότητας με την X τότε $X_1, X_2, \ldots, X_n$ ονομάζονται **τυχαίο δείγμα μεγέθους n από την μεταβλητή X.**

Μία τιμή του τυχαίου δείγματος συμβολίζεται με $x_1, x_2, \ldots, x_n$ και ονομάζεται **δειγματικό σημείο** (*sample point*). Πολλές φορές το δειγματικό σημείο ονομάζεται και **υλοποίηση** (*realization*) του τυχαίου δείγματος .

Σημειακές Εκτιμήσεις

Σημειακές Εκτιμήσεις

Οι περισσότεροι στατιστικοί έλεγχοι που κάνουμε αφορούν μια παράμετρο του πληθυσμού

Έστω θ η υπό εξέταση παράμετρος του πληθυσμού.

Από το δείγμα υπολογίζουμε μια εκτίμηση $\hat{\theta}$ της παραμέτρου θ του πληθυσμού και αποφασίζουμε με κάποια κριτήρια για την ορθότητα της υπόθεσης θ=θο Εξετάζοντας την απόσταση μεταξύ $\hat{\theta}$ και θο.

Η παραπάνω διαδικασία εξαρτάται από τη μέθοδο που χρησιμοποιούμε για την εκτίμηση του θ.

Το πιο γνωστό σας παράδειγμα είναι η τιμή του μέσου όρου του πληθυσμού μ και η εκτίμηση της τιμής $\bar{X}$ από το δείγμα.
Στην συνέχεια ο έλεγχος γίνεται με την βοήθεια της τυποποιημένης κανονικής κατανομής

$$z = \frac{(\bar{X} - \mu)}{\sigma}$$

Έννοιες

Εκτίμηση: εννοούμε την τιμή που παίρνουμε για έναν παράγοντα που υπολογίστηκε από ένα δείγμα πληθυσμού.

π.χ Η τιμή $\bar{x}=20$ που υπολογίστηκε από ένα δείγμα πληθυσμού, αποτελεί μια εκτίμηση του μέσου όρου του πληθυσμού

ΠΙΟ ΣΥΓΚΕΚΡΙΜΕΝΑ αποτελεί μια **σημειακή εκτίμηση** αφού μας δίνει μια μόνο τιμή και όχι ένα εύρος τιμών, όπως κάνουν άλλες μέθοδοι εκτίμησης

Εκτιμήτρια ή εκτιμητής είναι η τυχαία μεταβλητή οι τιμές της οποίας καθορίζονται όταν διαθέτουμε τα δεδομένα ενός δείγματος. Οι διάφορες τιμές που μπορεί να πάρει η εκτιμήτρια ονομάζονται **εκτιμήσεις**

Η εκτιμήτρια ως τυχαία μεταβλητή συμβολίζεται με κεφαλαίο χαρακτήρα, ενώ η εκτίμηση με τον αντίστοιχο μικρό χαρακτήρα.

$$\bar{X} = \bar{x}$$

Διαστήματα Εμπιστοσύνης

Διαστήματα Εμπιστοσύνης

Σε αντίθεση με την σημειακή εκτιμήτρια θ της παραμέτρου θ που μας δίνει μια μοναδική τιμή ως εκτίμηση, το διάστημα εμπιστοσύνης μας δίνει ένα εύρος τιμών (θ1,θ2) μέσα στο οποίο βρίσκεται η παράμετρος.

Εξαιτίας των δειγματοληπτικών σφαλμάτων, η πιθανότητα να έχει η θ την πραγματική τιμή της θ στον πληθυσμό είναι ελάχιστη αν όχι μηδέν.

Το Διάστημα Εμπιστοσύνης καθορίζει ένα εύρος τιμών μέσα στο οποίο βρίσκεται η θ και αποφεύγεται η σχεδόν σίγουρα εσφαλμένη σημειακή εκτίμηση. Λόγω των σφαλμάτων δεν μπορούμε με βεβαιότητα να υποστηρίξουμε ότι η θ βρίσκεται μέσα στο εύρος τιμών.

Για αυτό το λόγο το ΔΕ εκφράζεται με τη μορφή πιθανότητας.

Συγκεκριμένα λέμε ότι το Δ.Ε. (θ1,θ2) είναι ένα 95% Δ.Ε. για την παράμετρο θ, όταν η Πιθανότητα το διάστημα αυτό να καλύπτει την αληθινή τιμή της θ είναι 95%.

Οι συνηθισμένοι συντελεστές εμπιστοσύνης είναι 95% και 99% με επιλογή επιπέδων Σημαντικότητας 5% και 1% αντίστοιχα.

Διαστήματα εμπιστοσύνης για τους μέσους κανονικών πληθυσμών στην περίπτωση που η διακύμανση του πληθυσμού είναι γνωστή

Θέλουμε να εκτιμήσουμε το μέσο μ ενός κανονικού πληθυσμού.

Η διακύμανση του υπό μελέτη πληθυσμού είναι <u>γνωστή</u>.

Ο δειγματικός μέσος ακολουθεί την κανονική κατανομή με μέσο αριθμητικό μ και η διακύμανση $\sigma_{\bar{X}}^2 = \sigma^2 / n$

$$\frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \sim N(0,1)$$

Άρα

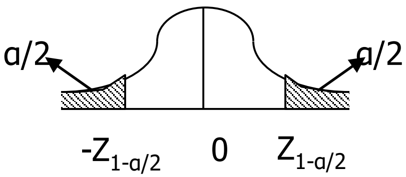
Έστω $Z_{1-\alpha/2}$ η τιμή του Z για την οποία η πιθανότητα να υπερβληθεί είναι α/2 και επομένως λόγω συμμετρίας $-Z_{1-\alpha/2}$ είναι η τιμή του Z για την οποία η πιθανότητα να βρεθεί τιμή ίση ή μικρότερη είναι α/2.

Διαστήματα εμπιστοσύνης για τους μέσους κανονικών πληθυσμών στην περίπτωση που η διακύμανση του πληθυσμού είναι γνωστή

Άρα η πιθανότητα να βρεθεί τιμή του Z που να περιλαμβάνεται μεταξύ $-Z_{1-\alpha/2}$ και $Z_{1-\alpha/2}$ είναι 1-α και γράφεται: $P\left(-Z_{1-\alpha/2} \le Z \le Z_{1-\alpha/2}\right) = 1-\alpha$

Επειδή $Z = \dfrac{\bar{X}-\mu}{\sigma_{\bar{X}}} \sim N(0,1)$ $\quad P\left(-Z_{1-\alpha/2} \le \dfrac{\bar{X}-\mu}{\sigma_{\bar{X}}} \le Z_{1-\alpha/2}\right) = 1-\alpha$

Επομένως, αν προκαθορίσουμε την πιθανότητα 1-α μπορούμε να βρούμε τις τιμές $-Z_{1-\alpha/2}$ και $Z_{1-\alpha/2}$ από τους πίνακες της **τυποποιημένης κανονικής κατανομής**.



$$P\left(\bar{X} - Z_{1-\alpha/2}\sigma_{\bar{X}} \le \mu \le \bar{X} + Z_{1-\alpha/2}\sigma_{\bar{X}}\right) = 1-\alpha$$

Επομένως τα άκρα του 100(1-α)% διαστήματος εμπιστοσύνης για τον μέσο μ ενός κανονικού πληθυσμού με γνωστή διακύμανση σ^2 είναι

$$\boxed{\left(\bar{X} - Z_{1-\frac{\alpha}{2}}\sigma_{\bar{X}},\ \bar{X} + Z_{1-\frac{\alpha}{2}}\sigma_{\bar{X}}\right)}$$

$$\sigma_{\bar{X}} = \sigma/\sqrt{n}$$

Παράδειγμα 1

Ένα μηχάνημα κατασκευάζει μπαταρίες για ειδικές χρήσεις των οποίων ο χρόνος ζωής ακολουθεί, σύμφωνα με τις τεχνικές προδιαγραφές του μηχανήματος, κανονική κατανομή με τυπική απόκλιση 3 ώρες. Επειδή στις χρήσεις για τις οποίες προορίζονται οι μπαταρίες ο χρόνος ζωής τους είναι ιδιαίτερα κρίσιμος, η λειτουργία του μηχανήματος παρακολουθείται συνεχώς με δειγματοληπτικό έλεγχο του χρόνου ζωής των παραγόμενων μπαταριών. Από τυχαίο δείγμα 9 μπαταριών διαπιστώνεται ότι οι χρόνοι ζωής τους (σε ώρες) είναι:

16	18	22	20	21	20	16	22	25

Με τη βοήθεια της πληροφορίας αυτής να εκτιμηθεί ένα 95% διάστημα εμπιστοσύνης για το μέσο χρόνο ζωής του συνόλου των παραγόμενων μπαταριών.

$$\overline{X} \pm Z_{1-a/2} \frac{\sigma}{\sqrt{n}}$$

Τυπολόγιο Σελ. 21!!!

Παράδειγμα 1

Για να σχηματίσουμε το 95% διάστημα εμπιστοσύνης για το μέσο χρόνο ζωής των μπαταριών θα εφαρμόσουμε τον τύπο Error! Reference source not found.**. Εύκολα βρίσκουμε ότι η τιμή του δειγματικού μέσου είναι:**

$$\bar{x} = \frac{\sum_{i=1}^{9} x_i}{n} = 20 \text{ ώρες.}$$

Το τυπικό σφάλμα εκτίμησης του μέσου είναι $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = 1$.

Στην προκειμένη περίπτωση $1-\alpha = 0,95 \Rightarrow \alpha = 0,05$ και κατά συνέπεια από τον Πίνακα 4 της τυποποιημένης κανονικής κατανομής βρίσκουμε ότι: $Z_{1-\alpha/2} = Z_{0,975} = 1,96$.

Παράδειγμα 1

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817

Παράδειγμα 1

Επομένως τα άκρα του 95% διαστήματος εμπιστοσύνης για το μέσο χρόνο ζωής των μπαταριών είναι, με βάση τον τύπο, 18,04 ώρες και 21,96 ώρες. Με απλά λόγια αυτό σημαίνει ότι με πιθανότητα 0,95 ο μέσος χρόνος ζωής του συνόλου των παραγομένων μπαταριών θα είναι μεταξύ 18,04 και 21,96 ωρών.

Διαστήματα εμπιστοσύνης για τους μέσους κανονικών πληθυσμών στην περίπτωση που η διακύμανση του πληθυσμού είναι άγνωστη

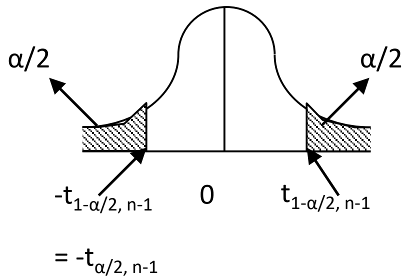
Το μέτρο που χρησιμοποιούμε για την εκτίμηση του σ^2 είναι η δειγματική διακύμανση η οποία ορίζεται ως

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

Στην περίπτωση αυτή όμως, η τυχαία μεταβλητή $\frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{\bar{X}-\mu}{S_{\bar{X}}}$ ακολουθεί την **κατανομή t του Student** με v = n -1 βαθμούς ελευθερίας

Επομένως, αν προκαθορίσουμε την πιθανότητα 1-α μπορούμε να βρούμε από τον πίνακα t τις τιμές $-t_{1-\alpha/2,\, n-1}$ και $t_{1-\alpha/2,\, n-1}$ έτσι ώστε,

$$P\left(t_{1-\alpha/2,\, n-1} \le \frac{\bar{X}-\mu}{S_{\bar{X}}} \le t_{1-\alpha/2,\, n-1} \right) = 1-\alpha$$



Λύνοντας τη σχέση αυτή ως προς μ θα έχουμε,

$$P\left(\bar{X} - t_{1-\alpha/2,\, n-1} S_{\bar{X}} \le \mu \le \bar{X} + t_{1-\alpha/2,\, n-1} S_{\bar{X}} \right) = 1-\alpha$$

Επομένως τα άκρα του 100(1-α)% διαστήματος εμπιστοσύνης για τον μέσο μ ενός κανονικού πληθυσμού με άγνωστη διακύμανση σ^2 είναι

$$\boxed{\left(\bar{X} - t_{1-\frac{\alpha}{2},\, n-1} S_{\bar{X}},\ \bar{X} - t_{1-\frac{\alpha}{2},\, n-1} S_{\bar{X}} \right)}$$

$$S_{\bar{X}} = S/\sqrt{n}$$

Παράδειγμα 2

Ένα μηχάνημα κατασκευάζει μπαταρίες για ειδικές χρήσεις των οποίων ο χρόνος ζωής ακολουθεί, σύμφωνα με τις τεχνικές προδιαγραφές του μηχανήματος, κανονική κατανομή . Επειδή στις χρήσεις για τις οποίες προορίζονται οι μπαταρίες ο χρόνος ζωής τους είναι μία κρίσιμη παράμετρος η λειτουργία του μηχανήματος παρακολουθείται συνεχώς με δειγματοληπτικό έλεγχο του χρόνου ζωής των παραγόμενων μπαταριών.

Από τυχαίο δείγμα 9 μπαταριών διαπιστώνεται ότι οι χρόνοι ζωής τους (*σε ώρες*) είναι:

16	18	22	20	21	20	16	22	25

Με τη βοήθεια της πληροφορίας αυτής να εκτιμηθεί ένα 95% διάστημα εμπιστοσύνης για το μέσο χρόνο ζωής του συνόλου των παραγόμενων μπαταριών.

Παράδειγμα 2

Χρησιμοποιούμε τα δεδομένα της άσκησης 1 μόνο που τώρα δεν δίνονται πληροφορίες για την πληθυσμιακή διακύμανση.

Ζητείται να εκτιμηθεί εκ νέου ένα 95% διάστημα εμπιστοσύνης για το μέσο χρόνο ζωής του συνόλου των παραγόμενων μπαταριών.

Τυπολόγιο Σελ. 21!!!

$$\boxed{\begin{array}{l} \overline{X} \pm t_{\nu,1-\alpha/2} \dfrac{s}{\sqrt{n}} \\ \nu = n-1 \end{array}}$$

Παράδειγμα 2

Για να προσδιορίσουμε τα άκρα ενός 95% διαστήματος εμπιστοσύνης για το μέσο χρόνο ζωής των μπαταριών θα εφαρμόσουμε τον τύπο.

Εύκολα βρίσκουμε ότι η τιμή του δειγματικού μέσου είναι:

$$\overline{x} = \frac{\sum_{i=1}^{9} x_i}{n} = 20 \text{ ώρες.}$$

Δεδομένου ότι η πληθυσμιακή διακύμανση είναι άγνωστη θα πρέπει να εκτιμηθεί από το δείγμα με τη χρήση του τύπου.

$$S^2 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n-1}$$

Παράδειγμα 2

Εύκολα προκύπτει ότι η δειγματική τυπική απόκλιση είναι $s = 2{,}958$ ώρες και κατά συνέπεια το τυπικό σφάλμα εκτίμησης του μέσου είναι $s_{\bar{X}} = 0{,}986$.

Στην προκειμένη περίπτωση $1-\alpha = 0{,}95 \Rightarrow \alpha = 0{,}05$ και $\nu = n-1 = 9-1 = 8$. Κατά συνέπεια από τον Πίνακα 5 των ποσοστιαίων σημείων της κατανομής t βρίσκουμε ότι: $t_{n-1,1-\alpha/2} = t_{8,0,975} = 2{,}306$. Επομένως τα άκρα του 95% διαστήματος εμπιστοσύνης για το μέσο χρόνο ζωής του συνόλου των παραγόμενων μπαταριών είναι, 17,726 ώρες και 22,274 ώρες.

	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	0.975	0.990
1	0.158	0.325	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821
2	0.142	0.289	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965
3	0.137	0.277	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541
4	0.134	0.271	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747
5	0.132	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365
6	0.131	0.265	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143
7	0.130	0.263	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998
8	0.130	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	$\boxed{2.306}$	2.896
9	0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821

Διαστήματα εμπιστοσύνης για τη διαφορά των μέσων κανονικών πληθυσμών στην περίπτωση που οι διακυμάνσεις των πληθυσμών είναι γνωστές

Έστω δύο ανεξάρτητοι κανονικοί πληθυσμοί $N(\mu_X, \sigma_X^2)$ και $N(\mu_Y, \sigma_Y^2)$ έστω ότι θέλουμε να συγκρίνουμε τους μέσους τους

Η τυχαία μεταβλητή $\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, \sigma^2_{\bar{X}-\bar{Y}}\right)$

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sigma_{\bar{X}-\bar{Y}}} \sim N(0.1)$$

όπου $\sigma_{\bar{x}-\bar{y}} = \sqrt{\sigma^2_{\bar{x}} + \sigma^2_{\bar{y}}} = \sqrt{\sigma^2_x / n + \sigma^2_Y / m}$

Επομένως τα άκρα του 100(1-a)% διαστήματος εμπιστοσύνης για τον μέσο μ_X-μ_Y δύο κανονικών πληθυσμών με γνωστές διακυμάνσεις είναι

$$\boxed{\left(\left(\bar{X} - \bar{Y}\right) - Z_{1-a/2}\, \sigma_{\bar{x}-\bar{y}}, \left(\bar{X} - \bar{Y}\right) + Z_{1-a/2}\, \sigma_{\bar{x}-\bar{y}}\right)}$$

$\sigma_{\bar{x}-\bar{y}} = \sqrt{\sigma^2_{\bar{x}} + \sigma^2_{\bar{y}}} = \sqrt{\sigma^2_x / n + \sigma^2_Y / m}$

Παράδειγμα 3

Ο υπεύθυνος εκπαίδευσης ενός ναυπηγείου θέλει να συγκρίνει την αποτελεσματικότητα δύο εναλλακτικών μεθόδων που χρησιμοποιούνται για την εκπαίδευση των ηλεκτροσυγκολητών της εταιρείας. Για το λόγο αυτό χωρίζει τυχαία 26 υποψηφίους για εκπαίδευση ηλεκτροσυγκολητές σε 2 ισοπληθείς ομάδες Α και Β για να εκπαιδευτούν με τις μεθόδους 1 και 2 αντίστοιχα. Κατά τη διάρκεια της εκπαίδευσης κάποιοι εκπαιδευόμενοι αποχώρησαν για λόγους ανεξάρτητους με το πρόγραμμα εκπαίδευσης και τελικά το πρόγραμμα ολοκλήρωσαν 12 άτομα της ομάδας Α και 10 άτομα της ομάδας Β. Μετά την ολοκλήρωση του προγράμματος οι εκπαιδευθέντες υποβλήθηκαν σε μια κοινή για όλους πρακτική άσκηση προκειμένου να αξιολογηθούν οι γνώσεις και οι ικανότητες που απέκτησαν. Η βαθμολογία της απόδοσης τους στην πρακτική άσκηση δίνεται στον παρακάτω Πίνακα:

ΒΑΘΜΟΛΟΓΙΑ ΕΚΠΑΙΔΕΥΘΕΝΤΩΝ

ΟΜΑΔΑ Α (xi) 70 93 82 90 77 86 79 84 98 73 81 85

ΟΜΑΔΑ Β (yi) 89 78 94 83 88 80 91 92 87 97

Με βάση τα στοιχεία αυτά να κατασκευασθεί ένα 90% διάστημα εμπιστοσύνης για τη διαφορά της μέσης βαθμολογίας.

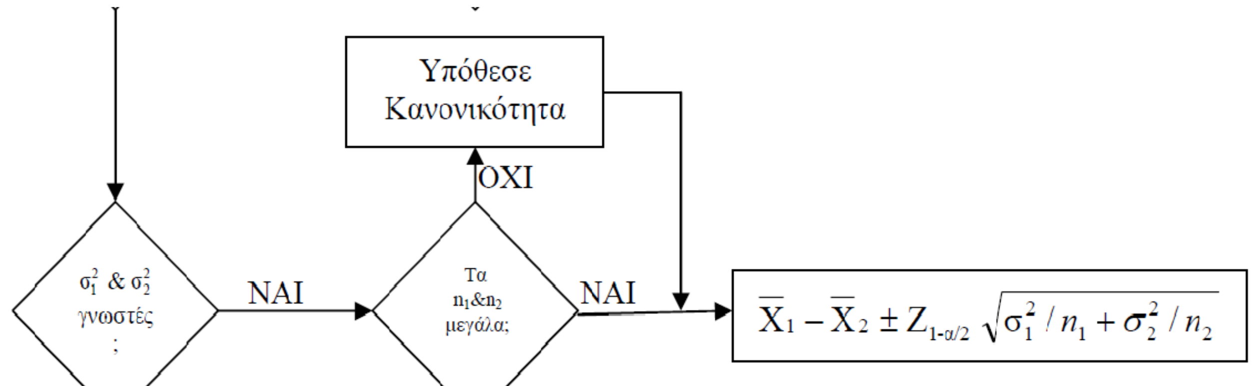
Δίνεται ότι η βαθμολογία ακολουθεί την κανονική κατανομή. Επίσης από προηγούμενη εμπειρία είναι γνωστό ότι η τυπική απόκλιση της βαθμολογίας για όσους εκπαιδεύονται με την πρώτη μέθοδο είναι $\sigma_X = 8$, ενώ η τυπική απόκλιση της βαθμολογίας για όσους εκπαιδεύονται με τη δεύτερη μέθοδο είναι $\sigma_Y = 6$.

Παράδειγμα 3

α. **Δίνεται ότι η βαθμολογία ακολουθεί την κανονική κατανομή. Επίσης από προηγούμενη εμπειρία είναι γνωστό ότι η τυπική απόκλιση της βαθμολογίας για όσους εκπαιδεύονται με την πρώτη μέθοδο είναι** $\sigma_X = 8$**, ενώ η τυπική απόκλιση της βαθμολογίας για όσους εκπαιδεύονται με τη δεύτερη μέθοδο είναι** $\sigma_Y = 6$**.**

α. Θέλουμε να κατασκευάσουμε ένα 90% διάστημα εμπιστοσύνης για τη διαφορά των μέσων των δύο πληθυσμών οι διακυμάνσεις των οποίων είναι γνωστές. Επίσης, από το πρόβλημα προκύπτει ότι τα δύο δείγματα είναι ανεξάρτητα. Επομένως θα εφαρμόσουμε τον τύπο.

Τυπολόγιο Σελ. 22!



Παράδειγμα 3

Εύκολα βρίσκουμε ότι οι τιμές των δειγματικών μέσων είναι, $\overline{x} = 83{,}17$ και $\overline{y} = 87{,}90$. Το τυπικό σφάλμα της εκτίμησης της διαφοράς των μέσων είναι

$$\sigma_{\overline{X}-\overline{Y}} = \sqrt{\sigma_X^2 / n + \sigma_Y^2 / m} = \sqrt{8^2 / 12 + 6^2 / 10} = 2{,}99\,.$$

Στην προκειμένη περίπτωση $1-\alpha = 0{,}90 \Rightarrow \alpha = 0{,}10$ και κατά συνέπεια από τον Πίνακα 4 της τυποποιημένης κανονικής κατανομής βρίσκουμε ότι: $Z_{1-\alpha/2} = Z_{0,95} = 1{,}64$.

Άρα, με αντικατάσταση στον τύπο προκύπτει ότι τα άκρα του 90% διαστήματος εμπιστοσύνης για τη διαφορά της μέσης βαθμολογίας είναι –9,63 και 0,17.

Διαστήματα εμπιστοσύνης για τη διαφορά των μέσων κανονικών πληθυσμών στην περίπτωση που οι διακυμάνσεις των πληθυσμών είναι άγνωστες

Έστω δύο ανεξάρτητοι κανονικοί πληθυσμοί $N(\mu_X, \sigma_X^2)$ και $N(\mu_Y, \sigma_Y^2)$ έστω ότι θέλουμε να συγκρίνουμε τους μέσους τους

Η τυχαία μεταβλητή $\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, \sigma^2_{\bar{X}-\bar{Y}}\right)$

Αν οι διακυμάνσεις είναι άγνωστες αλλά ίσες

$$T = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{S_X^2/n + S_Y^2/m}} \sim t_{n+m-2}$$

$$S_p^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{m}(Y_i - \bar{Y})^2}{n+m-2}$$

Επομένως τα άκρα του 100(1-α)% διαστήματος εμπιστοσύνης για τον μέσο μ_X-μ_Y δύο κανονικών πληθυσμών με άγνωστές ίσες διακυμάνσεις είναι

$$\left(\left(\bar{X} - \bar{Y}\right) - t_{1-a/2,\, n+m-2} S_{\bar{x}-\bar{y}}, \left(\bar{X} - \bar{Y}\right) + t_{1-a/2,\, n+m-2} S_{\bar{x}-\bar{y}}\right)$$

$$S_{\bar{x}-\bar{y}} = S_p\sqrt{1/n + 1/m}$$

Παράδειγμα 4

Ο υπεύθυνος εκπαίδευσης ενός ναυπηγείου θέλει να συγκρίνει την αποτελεσματικότητα δύο εναλλακτικών μεθόδων που χρησιμοποιούνται για την εκπαίδευση των ηλεκτροσυγκολητών της εταιρείας. Για το λόγο αυτό χωρίζει τυχαία 26 υποψηφίους για εκπαίδευση ηλεκτροσυγκολητές σε 2 ισοπληθείς ομάδες Α και Β για να εκπαιδευτούν με τις μεθόδους 1 και 2 αντίστοιχα. Κατά τη διάρκεια της εκπαίδευσης κάποιοι εκπαιδευόμενοι αποχώρησαν για λόγους ανεξάρτητους με το πρόγραμμα εκπαίδευσης και τελικά το πρόγραμμα ολοκλήρωσαν 12 άτομα της ομάδας Α και 10 άτομα της ομάδας Β. Μετά την ολοκλήρωση του προγράμματος οι εκπαιδευθέντες υποβλήθηκαν σε μια κοινή για όλους πρακτική άσκηση προκειμένου να αξιολογηθούν οι γνώσεις και οι ικανότητες που απέκτησαν. Η βαθμολογία της απόδοσης τους στην πρακτική άσκηση δίνεται στον παρακάτω Πίνακα:

ΒΑΘΜΟΛΟΓΙΑ ΕΚΠΑΙΔΕΥΘΕΝΤΩΝ

ΟΜΑΔΑ Α (x_i) 70 93 82 90 77 86 79 84 98 73 81 85

ΟΜΑΔΑ Β (y_i) 89 78 94 83 88 80 91 92 87 97

Με βάση τα στοιχεία αυτά να κατασκευασθεί ένα 90% διάστημα εμπιστοσύνης για τη διαφορά της μέσης βαθμολογίας.

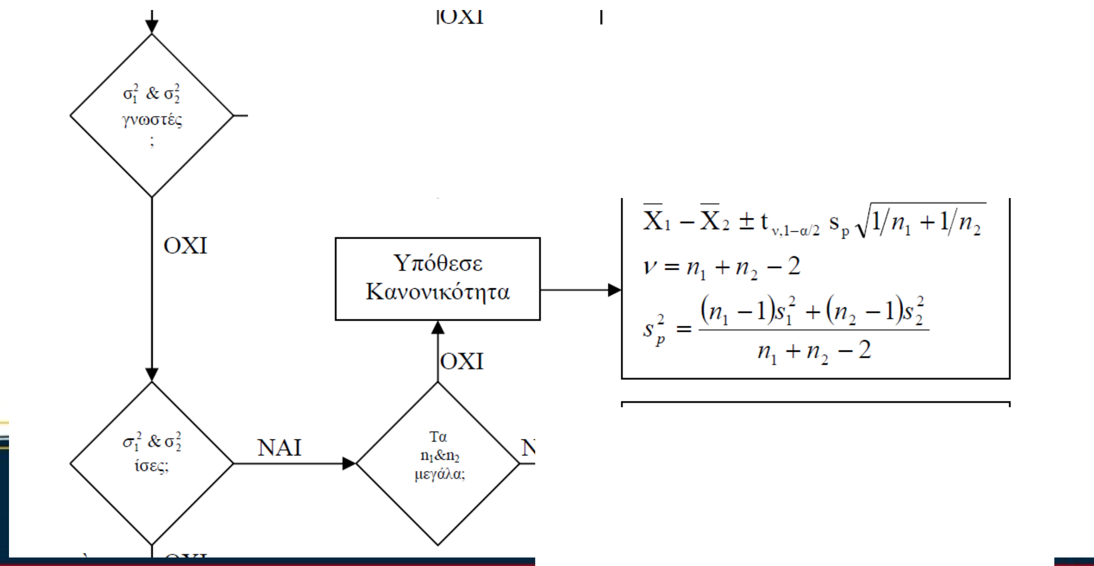
Δίνεται ότι η βαθμολογία ακολουθεί την κανονική κατανομή και οι διακυμάνσεις των δύο πληθυσμών μπορούν να υποτεθούν ίσες.

Παράδειγμα 4

β. **Δίνεται ότι η βαθμολογία ακολουθεί την κανονική κατανομή και υποθέτουμε ότι οι διακυμάνσεις των δύο πληθυσμών είναι ίσες.**

β. Θέλουμε να κατασκευάσουμε ένα 90% διάστημα εμπιστοσύνης για τη διαφορά των μέσων των δύο πληθυσμών για τις άγνωστες διακυμάνσεις των οποίων έχουμε υποθέσει ότι είναι ίσες. Επίσης, από το πρόβλημα προκύπτει ότι τα δύο δείγματα είναι ανεξάρτητα. Επομένως θα εφαρμόσουμε τον τύπο.

Τυπολόγιο Σελ. 22!!!



Παράδειγμα 4

Έχουμε ήδη υπολογίσει τις τιμές των δειγματικών μέσων ότι είναι: $\overline{x}=83,17$ και $\overline{y}=87,90$. Επίσης, υπολογίζονται οι δειγματικές διακυμάνσεις οπότε προκύπτει ότι οι δειγματικές τυπικές αποκλίσεις είναι $s_X=8,05$ και $s_Y=6,08$.

Αντικαθιστώντας στον τύπο υπολογίζουμε τη συνδυασμένη διακύμανση και κατά συνέπεια βρίσκουμε ότι η συνδυασμένη τυπική απόκλιση είναι, $s_p=7,23$.

Άρα, το τυπικό σφάλμα της εκτίμησης της διαφοράς των μέσων με βάση τον τύπο είναι,

$$s_{\overline{X}-\overline{Y}}=7,23\sqrt{1/12+1/10}=3,096.$$

Στην προκειμένη περίπτωση $1-\alpha=0,90\Rightarrow\alpha=0,10$ οι δε βαθμοί ελευθερίας δίνονται από τη σχέση $\nu=n+m-2=12+10-2=20$. Κατά συνέπεια από τον Πίνακα 5 των ποσοστιαίων σημείων της κατανομής t βρίσκουμε ότι: $t_{n+m-2,1-\alpha/2}=t_{20,0,95}=1,725$.

Άρα, προκύπτει ότι τα άκρα του 90% διαστήματος εμπιστοσύνης για τη διαφορά της μέσης βαθμολογίας είναι –10,07 και 0,61.

Διαστήματα εμπιστοσύνης για τη διαφορά των μέσων κανονικών πληθυσμών στην περίπτωση που οι διακυμάνσεις των πληθυσμών είναι άγνωστες

Έστω δύο ανεξάρτητοι κανονικοί πληθυσμοί $N(\mu_X, \sigma_X^2)$ και $N(\mu_Y, \sigma_Y^2)$ έστω ότι θέλουμε να συγκρίνουμε τους μέσους τους

Η τυχαία μεταβλητή $\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, \sigma^2_{\bar{X}-\bar{Y}}\right)$

Αν οι διακυμάνσεις είναι άγνωστες και άνισες

$$T = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{S_X^2 / n + S_Y^2 / m}} \sim t_{n+m-2}$$

Επομένως τα άκρα του 100(1-α)% διαστήματος εμπιστοσύνης για τον μέσο μ_X-μ_Y δύο κανονικών πληθυσμών με άγνωστές άνισες διακυμάνσεις είναι

$$\boxed{\left(\left(\bar{X} - \bar{Y}\right) - t_{1-a/2,\ n+m-2} S_{\bar{x}-\bar{y}}, \left(\bar{X} - \bar{Y}\right) + t_{1-a/2,\ n+m-2} S_{\bar{x}-\bar{y}}\right)}$$

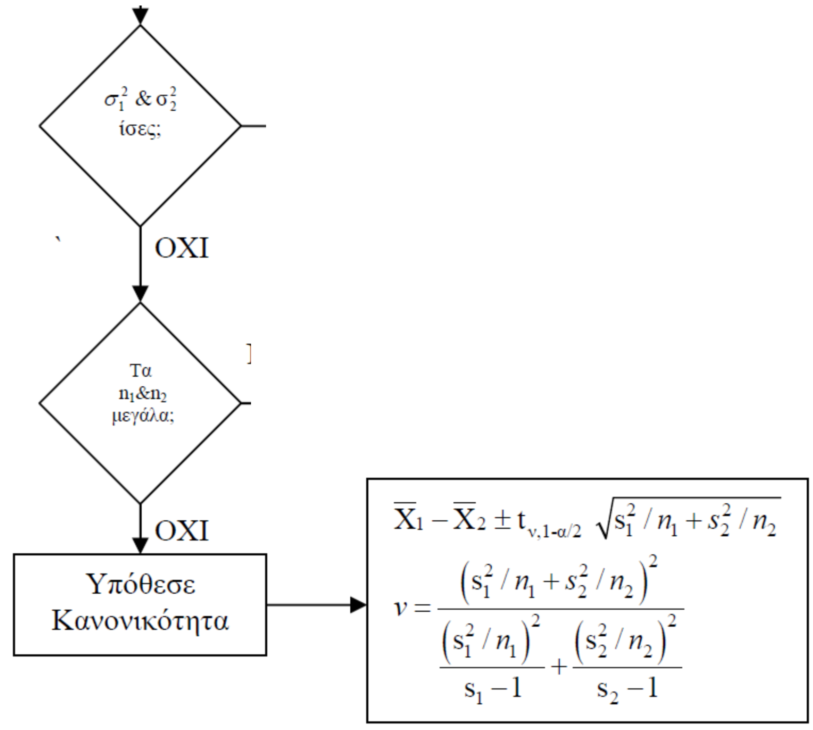
$$S_{\bar{x}-\bar{y}} = \sqrt{S_X^2 / n + S_Y^2 / m}$$

Παράδειγμα 5

γ. **Δίνεται ότι η βαθμολογία ακολουθεί την κανονική κατανομή και υποθέτουμε ότι οι διακυμάνσεις των δύο πληθυσμών είναι άνισες.**

γ. Θέλουμε να κατασκευάσουμε ένα 90% διάστημα εμπιστοσύνης για τη διαφορά των μέσων των μέσων των δύο πληθυσμών για τις διακυμάνσεις των οποίων έχουμε υποθέσει ότι είναι άνισες.

Τυπολόγιο Σελ. 22!!!



Παράδειγμα 5

Επίσης, από το πρόβλημα προκύπτει ότι τα δύο δείγματα είναι ανεξάρτητα. Έχουμε ήδη υπολογίσει ότι: $\overline{x} = 83,17$, $\overline{y} = 87,90$ και $s_X = 8,05$, $s_Y = 6,08$

Επομένως το τυπικό σφάλμα της εκτίμησης της διαφοράς των μέσων είναι, με βάση τον τύπο

$$S_{\overline{X}-\overline{Y}} = \sqrt{(8,05)^2/12 + (6,08)^2/10} = 3,016 .$$

Από τον τύπο προκύπτει ότι οι βαθμοί ελευθερίας στην περίπτωση αυτή είναι

$$\nu \cong \frac{\left(\frac{8,05^2}{12} + \frac{6,08^2}{10}\right)^2}{\frac{\left(8,05^2/12\right)^2}{12-1} + \frac{\left(6,08^2/10\right)^2}{10-1}} \cong 16 .$$

Επιπλέον, στην προκειμένη περίπτωση $1-\alpha = 0,90 \Rightarrow \alpha = 0,10$ οι δε βαθμοί ελευθερίας είναι 16. Κατά συνέπεια από τον Πίνακα 5 των ποσοστιαίων σημείων της κατανομής t βρίσκουμε ότι: $t_{\nu,1-\alpha/2} = t_{17,0,95} = 1,74$. Άρα, με αντικατάσταση προκύπτει ότι τα άκρα του 90% διαστήματος εμπιστοσύνης για τη διαφορά της μέσης περιεκτικότητας είναι -9,98 και 0,52.

Διαστήματα εμπιστοσύνης – Δείγματα εξαρτημένα (Ζεύγη)

Οι παρατηρήσεις ενός δείγματος δεν είναι ανεξάρτητες/

Η εξάρτηση εμφανίζεται:

- Είτε διότι οι παρατηρήσεις έχουν επιλεγεί κατά ζεύγη με βάση κάποιο χαρακτηριστικό
- Είτε διότι έχουμε επαναλαμβανόμενες παρατηρήσεις πάνω στα ίδια άτομα ή στοιχεία

$$\bar{D} = \frac{\sum_{i=1}^{n} D_i}{n}$$

$$S_D = \sqrt{\frac{\sum_{i=1}^{n} (D_i - \bar{D})^2}{n-1}}.$$

$$\bar{D} \pm t_{n-1,1-\alpha/2} S_{\bar{D}}$$

όπου

$$S_{\bar{D}} = \frac{S_D}{\sqrt{n}}$$

$$S_{\bar{x}-\bar{y}} = \sqrt{S_X^2 / n + S_Y^2 / m}$$

Παράδειγμα 6

Για να μετρηθεί η αποτελεσματικότητα μιας μεθόδου αδυνατίσματος επελέγη ένα τυχαίο δείγμα πέντε ατόμων το βάρος των οποίων καταγράφτηκε με τα εξής αποτελέσματα.

Α.Χ.	**Κ.Μ.**	**Π.Π.**	**Β.Α.**	**Κ.Ζ.**
84	98	78	92	85

Τέσσερις βδομάδες μετά από την έναρξη της δίαιτας παρατηρήθηκε ότι το βάρος τους ελαττώθηκε με εξαίρεση του Κ.Μ. ο οποίος ζύγιζε περισσότερο. Συγκεκριμένα, τα βάρη μετά τη δίαιτα διαμορφώθηκαν ως εξής:

Κ.Μ.	**Κ.Ζ.**	**Β.Α.**	**Π.Π**	**Α.Χ.**
99	82	90	75	80

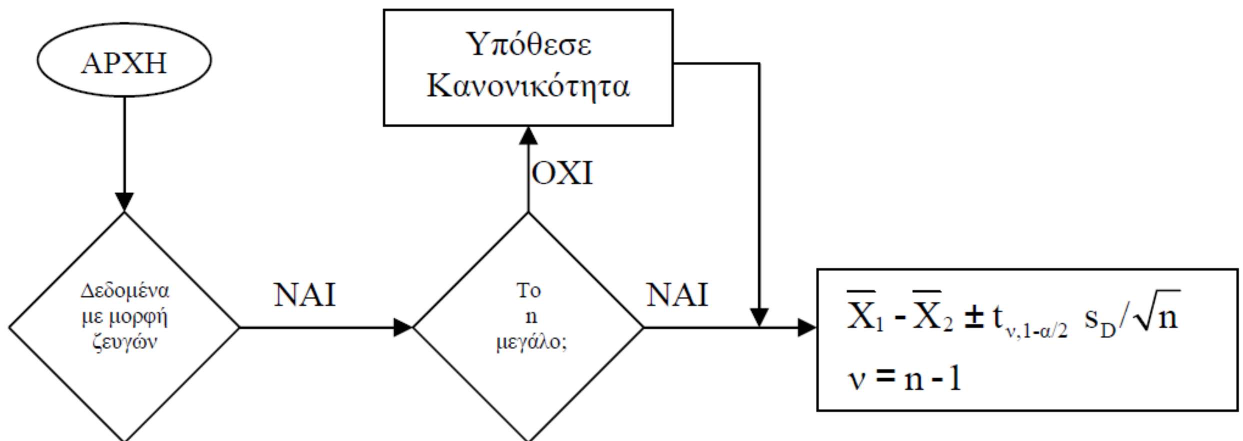
Να κατασκευαστεί ένα 95% διάστημα εμπιστοσύνης για τη διαφορά του μέσου βάρους πριν και μετά τη δίαιτα.

Παράδειγμα 6

Θέλουμε να κατασκευάσουμε ένα 95% διάστημα εμπιστοσύνης για τη διαφορά των μέσων δύο πληθυσμών στηριζόμενοι στις πληροφορίες δύο εξαρτημένων δειγμάτων.

Τα δείγματα είναι εξαρτημένα αφού αναφέρονται στα ίδια άτομα, δηλαδή τα στοιχεία μας αποτελούν ζεύγη.

Τυπολόγιο Σελ. 22!!!



Παράδειγμα 6

Για το λόγο αυτό το πρώτο βήμα είναι η αντιστοίχηση των τιμών.

	Α.Χ.	**Κ.Μ.**	**Π.Π.**	**Β.Α.**	**Κ.Ζ.**
ΠΡΙΝ	84	98	78	92	85
ΜΕΤΑ	80	99	75	90	82
D_i	4	-1	3	2	3

Με βάση τον τύπο προκύπτει ότι:

$$\bar{D} = 2,2$$.

Αντικατάσταση στον τύπο (8.45) δίνει:

$$S_D = 1,924$$.

Το τυπικό σφάλμα εκτίμησης της διαφοράς των μέσων είναι,

$$S_{\bar{D}} = 0,860$$.

Τέλος από τους πίνακες της κατανομής t βρίσκουμε ότι $t_{4,0,975} = 2,776$.
Επομένως με αντικατάσταση στον τύπο προκύπτει ότι τα άκρα ενός 95% διαστήματος εμπιστοσύνης για τη διαφορά των μέσων είναι –0,189 και 4,589.