Εφαρμοσμένη Στατιστική

Δημήτριος Μπάγκαβος

Τμήμα Μαθηματικών και Εφαρμοσμένων Μαθηματικών
Πανεπιστήμιο Κρήτης

14 Μαρτίου 2018

Διαστήματα Εμπιστοσύνης.

- Έχουμε δει εκτενώς μέχρι τώρα τρόπους εκτίμησης παραμέτρων διαφόρων κατανομών.
- Π.χ. για να εκτιμήσουμε την μέση τιμή ενός κανονικού πληθυσμού (δηλ. δείγματος $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$) χρησιμοποιούμε τον δειγματικό μέσο $\bar{X}$.
- Εκτίμηση μέσω μιας συνάρτησης (όπως ο $\bar{X}$) καλείται σημειακή εκτίμηση διότι το αποτέλεσμα της εκτίμησης είναι ένα σημείο.
- Προκειμένου να έχει νόημα η εκτίμηση, το αποτέλεσμα της θεωρητικά πρέπει να είναι «κοντά» στην παράμετρο που εκτιμούμε με «μεγάλη» πιθανότητα.
  - Όπως είπαμε και στον έλεγχο υποθέσεων, αν πάρουμε αρκετά δείγματα τότε τα αντίστοιχα X που θα υπολογίζουμε θα παίρνουν τιμές «κοντά» και «γύρω» από το μ με «μεγάλη» πιθανότητα.
- Θα ήταν χρήσιμο να έχουμε και μια διαδικασία που μας δίνει κάποια ιδέα για την ακρίβεια ή το σφάλμα της εκτίμησης που κάνουμε κατά την εκτίμηση παραμέτρων.
- Θα ήταν συνεπώς προτιμότερο αν μπορούσαμε να πούμε ότι, βάσει του συγκεκριμένου τυχαίου δείγματος, η υπό εκτίμηση παράμετρος βρίσκεται με κάποια «πιθανότητα» μεταξύ δύο τιμών.

Διαστήματα Εμπιστοσύνης.

Ορισμός: Διάστημα Εμπιστοσύνης

Έστω τυχαίο δείγμα $X_1, \ldots, X_n$ από κατανομή $f(x; \theta)$, $g(\theta)$ μία παραμετρική συνάρτηση που θέλουμε να εκτιμήσουμε και έστω δυο στατιστικές συναρτήσεις $L(\mathbf{X}) = L(X_1, \ldots, X_n), U(\mathbf{X}) = U(X_1, \ldots, X_n)$. Το τυχαίο διάστημα $[L(\mathbf{X}), U(\mathbf{X})]$ καλείται διάστημα εμπιστοσύνης για την $g(\theta)$ σε επίπεδο σημαντικότητας $1 - \alpha$ αν ισχύει ότι

$$P\left(g(\theta) \in [L(\mathbf{X}), U(\mathbf{X})]\right) = P\left(L(\mathbf{X}) \leq g(\theta) \leq U(\mathbf{X})\right) \geq 1 - \alpha.$$

Όταν η πιο πάνω σχέση ισχύει ως ισότητα τότε το $1 - \alpha$ καλείται συντελεστής εμπιστοσύνης.

- Επομένως, αν δεδομένου ενός τ.δ. $X_1, \ldots, X_n$, βρούμε στατιστικές συναρτήσεις $L(\mathbf{X})$ και $U(\mathbf{X})$ όπως παραπάνω, τότε μπορούμε να πούμε ότι η παραμετρική συνάρτηση την οποία επιθυμούμε να εκτιμήσουμε βρίσκεται μέσα στο διάστημα $[L(\mathbf{X}), U(\mathbf{X})]$ με πιθανότητα (τουλάχιστον) $1 - \alpha$.
- Αν δηλαδή παίρναμε πάρα πολλά δείγματα και υπολογίζαμε κάθε φορά το $[L(\mathbf{X}), U(\mathbf{X})]$ τότε θεωρητικά το $g(\theta)$ θα βρισκόταν μέσα σε τουλάχιστον $100(1 - \alpha)\%$ των διαστημάτων αυτών.

(α) Εκτίμηση του μ όταν $X \sim N(\mu, \sigma^2)$, σ^2 γνωστό.

- Δεδομένου ενός τ.δ. ανεξάρτητων, ισόνομων παρατηρήσεων $X_1, \ldots, X_n$, από κατανομή $N(\mu, \sigma^2)$ όπου σ^2 γνωστό,
- Ζητάμε να βρούμε ένα διάστημα μέσα στο οποίο βρίσκεται το μ σε τουλάχιστον $100(1-\alpha)\%$ των φορών.
- Επειδή ο δειγματικός μέσος $\bar{X}$ είναι μία αμερόληπτη εκτιμήτρια του μ, ψάχνουμε ένα διάστημα της μορφής $[\bar{X}-d, \bar{X}+d]$.
- Σύμφωνα με τον ορισμό θέλουμε να ισχύει

$$P\left(\mu \in [\bar{X}-d, \bar{X}+d]\right) = P\left(\bar{X}-d \leq \mu \leq \bar{X}+d\right) \geq 1-\alpha.$$

- Ξέρουμε ότι $\bar{X} \sim N(\mu, n^{-1}\sigma^2)$ (γιατί;)
- Ισοδύναμα

$$Z = \frac{\bar{X}-\mu}{\sqrt{n^{-1}\sigma^2}} \sim N(0,1). \text{ (γιατί;)}$$

- Επομένως το d πρέπει να είναι τέτοιο ώστε

$$P\left(\bar{X}-d \leq \mu \leq \bar{X}+d\right) = 1-\alpha \Leftrightarrow P\left(-d \leq \bar{X}-\mu \leq d\right) = 1-\alpha.$$

(α) Εκτίμηση του μ όταν $X \sim N(\mu, \sigma^2)$, σ^2 γνωστό.

- Οπότε

$$
\begin{aligned}
P\left(-\frac{d\sqrt{n}}{\sigma} \le \frac{\bar{X}-\mu}{\sqrt{n^{-1}\sigma^2}} \le \frac{d\sqrt{n}}{\sigma}\right) &= P\left(-\frac{d\sqrt{n}}{\sigma} \le Z \le \frac{d\sqrt{n}}{\sigma}\right) \\
= \Phi\left(\frac{d\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{d\sqrt{n}}{\sigma}\right) &= \Phi\left(\frac{d\sqrt{n}}{\sigma}\right) - 1 + \Phi\left(\frac{d\sqrt{n}}{\sigma}\right) \\
&= 1-\alpha \Leftrightarrow \Phi\left(\frac{d\sqrt{n}}{\sigma}\right) = 1-\frac{\alpha}{2}.
\end{aligned}
$$

- Έστω Φ^{-1} η αντίστροφη της Φ (η Φ είναι γνήσια αύξουσα και άρα είναι 1-1 και άρα αντιστρέφεται). Οπότε

$$
\frac{d}{\sqrt{n^{-1}\sigma^2}} = \Phi^{-1}\left(1-\frac{\alpha}{2}\right) \Leftrightarrow d = \sqrt{n^{-1}\sigma^2}\Phi^{-1}\left(1-\frac{\alpha}{2}\right).
$$

- Οπότε ένα δ.ε. για την παράμετρο μ με συντελεστή $1-\alpha$ είναι

$$
\left[\bar{X} - \sqrt{n^{-1}\sigma^2}\Phi^{-1}\left(1-\frac{\alpha}{2}\right), \bar{X} + \sqrt{n^{-1}\sigma^2}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right]. \qquad (1)
$$

(α) Εκτίμηση του μ όταν $X \sim N(\mu, \sigma^2)$, σ^2 γνωστό.

- Αξίζει να παρατηρήσουμε ότι το παραπάνω δ.ε. εξακολουθεί να ισχύει και στην περίπτωση που το δείγμα προέρχεται από οποιονδήποτε πληθυσμό (όχι απαραίτητα κανονικό), υπό την προϋπόθεση ότι το n είναι σχετικά μεγάλο (γιατί;).

Παράδειγμα 1

Έχουμε ένα δείγμα βαρών φοιτητών με τιμές: 73, 81, 84, 77, 71, 75, 71, 76, 63, 69, 85, 77, 71, 81, 71, 76, 79, 68, 72, 71 το οποίο γνωρίζουμε ότι προέρχεται από κανονική κατανομή με $\sigma = 5$. Θέλουμε ένα δ.ε. για τη μέση τιμή του πληθυσμού με συντελεστή εμπιστοσύνης 95% ($\alpha = 5\%$).

Λύση: Έχουμε $n = 20, \bar{X} = 74.55$. Με αντικατάσταση στην (1), το ζητούμενο δ.ε. είναι:

$$\left[74.55 - \sqrt{20^{-1}5^2}\Phi^{-1}(0.975), 74.55 + \sqrt{20^{-1}5^2}\Phi^{-1}(0.975)\right].$$

Από πίνακες της κανονικής ξέρουμε ότι $\Phi^{-1}(0.975) = 1.96 (= z_{1-\alpha/2})$ (ποια στατιστική συνάρτηση είναι; - μπορείτε να την υπολογίσετε στην R;) οπότε το δ.ε. είναι $[72.36, 76.74]$.

(α) Εκτίμηση του μ όταν $X \sim N(\mu, \sigma^2)$, σ^2 γνωστό.

- Μπορούμε λοιπόν να πούμε ότι το μέσο βάρος των φοιτητών βρίσκεται μεταξύ του 72.36 και 76.74 με συντελεστή εμπιστοσύνης 95%.

Διατύπωση

- Σε ότι ακολουθεί χρησιμοποιούμε εκφράσεις της μορφής: «το μ βρίσκεται μεταξύ του 72.36 και 76.74 με συντελεστή εμπιστοσύνης 95%».
- Αυτό υποδηλώνει ότι αν παίρναμε ένα μεγάλο πλήθος από δείγματα, και για το καθένα κατασκευάζαμε ένα δ.ε. για το μ, τότε θα αναμέναμε ότι το 95% των δ.ε. θα συμπεριλάμβανε το μ.

- Παρατηρούμε ότι η γνώση του σ είναι προαπαιτούμενη διότι η τιμή της είναι αναγκαία για τον υπολογισμό των άκρων του διαστήματος.
- Επίσης παρατηρούμε ότι όσο το δείγμα μεγαλώνει, τόσο το εύρος του διαστήματος μικραίνει (στενεύει), δηλαδή έχουμε καλύτερη εκτίμηση του μ. (γιατί;)

(α) Εκτίμηση του μ όταν $X \sim N(\mu, \sigma^2)$, σ^2 γνωστό.

- Αντίστοιχα, αν αυξήσουμε το συντελεστή εμπιστοσύνης (θέλουμε π.χ. να έχουμε ασφαλέστερη πρόβλεψη) τότε το εύρος του δ.ε. αυξάνεται.
- Αναφερόμενοι στο Παράδειγμα 1 με την εκτίμηση του μέσου βάρους, αν πάρουμε ως $1-\alpha = 99\%$ τότε το δ.ε. για το μέσο βάρος θα είναι:

$$\left[74.55 - \sqrt{\frac{25}{20}} z_{1-\frac{0.01}{2}}, 74.55 + \sqrt{\frac{25}{20}} z_{1-\frac{0.01}{2}}\right] \overset{z_{1-\frac{0.01}{2}}=2.58}{=} [71.66, 77.34]$$

όπου ισχύει ότι $-z_{\frac{0.01}{2}} = z_{1-\frac{0.01}{2}}$ λόγω συμμετρικής κατανομής.

- Το νέο δ.ε. είναι ευρύτερο από το $[72.36, 76.74]$ που είχαμε βρει για δ.ε. $1-\alpha = 95\%$.
- Αυτό συμβαίνει διότι με το ίδιο δείγμα θέλουμε να έχουμε ένα ασφαλέστερο άνω και κάτω όριο για το μ.
- Για να ελαττωθεί λοιπόν η πιθανότητα το μ να μην βρίσκεται εντός των ορίων του δ.ε., αυτό που γίνεται είναι ότι αυξάνεται το εύρος του δ.ε.

(β) Εκτίμηση του σ^2 όταν $X \sim N(\mu, \sigma^2)$, μ γνωστό.

- Δεδομένου ενός τ.δ. ανεξάρτητων, ισόνομων παρατηρήσεων $X_1, \ldots, X_n$, από κατανομή $N(\mu, \sigma^2)$ όπου μ γνωστό θέλουμε ένα δ.ε. για το σ^2 με επίπεδο εμπιστοσύνης $1-\alpha$.
- Είδαμε στα προηγούμενα ότι η δειγματική διασπορά

$$T = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

είναι ένας αμερόληπτες εκτιμητής της σ^2 και μάλιστα είναι ε.μ.π.

- Παρατηρούμε ότι

$$\frac{nT}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n.$$

- Το σημαντικό εδώ είναι ότι η χ^2_n κατανομή δεν εξαρτάται από το σ^2.
- Οπότε για να υπολογίσουμε το δ.ε. πρέπει να υπολογίσουμε τις σταθερές c_1 και c_2 ώστε

$$P\left(c_1 \leq \frac{nT}{\sigma^2} \leq c_2\right) = 1 - \alpha \qquad (2)$$

(β) Εκτίμηση του σ^2 όταν $X \sim N(\mu, \sigma^2)$, μ γνωστό.

- Για να ισχύει η (2), Θέλουμε τα c_1, c_2 να ικανοποιούν τις

$$P\left(\frac{nT}{\sigma^2} > c_2\right) = \frac{\alpha}{2}, \; P\left(\frac{nT}{\sigma^2} < c_1\right) = \frac{\alpha}{2}$$

- οπότε

$$c_2 = \chi^2_{n,1-\frac{\alpha}{2}}, c_1 = \chi^2_{n,\frac{\alpha}{2}}$$

  όπου γενικά με $\chi^2_{n,\alpha}$ συμβολίζουμε το α ποσοστιαίο σημείο της χ^2 κατανομής.
- Αντικαθιστώντας στην (2) και λύνοντας ως προς σ^2 παίρνουμε

$$P\left(\frac{nT}{\chi^2_{n,1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{nT}{\chi^2_{n,\frac{\alpha}{2}}}\right) = 1 - \alpha.$$

- Οπότε στην πράξη, σαν ένα $100(1-\alpha)\%$ δ/μα εμπιστοσύνης του σ^2 όταν το μ είναι γνωστό, χρησιμοποιούμε το

$$\left[\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi^2_{n,1-\frac{\alpha}{2}}}, \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi^2_{n,\frac{\alpha}{2}}}\right]. \tag{3}$$

(β) Εκτίμηση του σ^2 όταν $X \sim N(\mu, \sigma^2)$, μ γνωστό.

Συνέχεια παραδείγματος 1.

Να βρεθεί ένα διάστημα εμπιστοσύνης συντελεστού $1 - \alpha = 95\%$ για τη διασπορά σ^2 του βάρους των φοιτητών στο παράδειγμα 1 δεδομένου ότι το μέσο βάρος είναι $\mu = 75$.

Λύση: Καταρχήν

$$\sum_{i=1}^{n} (X_i - \mu)^2 = (73 - 75)^2 + \cdots + (71 - 75)^2 = 601.$$

Επίσης, $\chi^2_{20,0.025} = 9.5, \chi^2_{20,0.975} = 34.1$, οπότε αντικαθιστώντας στην (3)

$$\left[\frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\chi^2_{n,1-\frac{\alpha}{2}}}, \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\chi^2_{n,\frac{\alpha}{2}}} \right] = \left[\frac{601}{34.17}, \frac{601}{9.59} \right] = [17.59, 62.67].$$

Συμπερασματικά, η τυπική απόκλιση του βάρους των φοιτητών θα είναι μεταξύ του 4.19 και του 7.91 με συντελεστή εμπιστοσύνης 95%.

(γ) Εκτίμηση του σ^2 όταν $X \sim N(\mu, \sigma^2)$, μ άγνωστο.

- Έστω τ.δ. ανεξάρτητων, ισόνομων παρατηρήσεων $X_1, \ldots, X_n$, από κατανομή $N(\mu, \sigma^2)$ όπου μ άγνωστο.
- Για να βρούμε ένα διάστημα στο οποίο βρίσκεται το σ^2, αντικαθιστούμε στην (3) όπου στη θέση του μ χρησιμοποιούμε το δειγματικό μέσο.
- Δουλεύοντας ακριβώς με τον ίδιο τρόπο με τον οποίο πήραμε την (3) καταλήγουμε στο το δ/μα

$$\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right] \tag{4}$$

  που περιμένουμε να περιλαμβάνει την υποεκτίμηση παράμετρο σε ποσοστό $100(1-\alpha)\%$
- Π.χ. επαναλαμβάνοντας την εκτίμηση του παραδείγματος 1 με άγνωστη μέση τιμή το εκτιμώμενο δ/μα είναι

$$[18.17, 70.00]$$ επιβεβαιώστε τις πράξεις!!

(δ) Εκτίμηση του μ όταν $X \sim N(\mu, \sigma^2)$, σ^2 άγνωστο.

- Δεδομένου ενός τ.δ. ανεξάρτητων, ισόνομων παρατηρήσεων $X_1, \ldots, X_n$, από κατανομή $N(\mu, \sigma^2)$ όπου σ^2 άγνωστό θέλουμε ένα δ.ε. για το μ με επίπεδο εμπιστοσύνης $1 - \alpha$.
- Όπως και πριν, ένα δ/μα για το μ μπορεί να προκύψει από την εκτιμήτρια
$$\frac{\bar{X} - \mu}{\sqrt{n^{-1}\sigma^2}}$$
όπου όμως το σ^2 είναι άγνωστο.
- Μπορούμε αντικαταστήσουμε τη σ^2 με την εκτιμήτρια μέγιστης πιθανόφάνειας (τη δειγματική διασπορά S^2).
- Η διαφορά είναι τώρα ότι ο παρανομαστής της εκτιμήτριας συνάρτησης είναι τ.μ. και έτσι
$$\frac{\bar{X} - \mu}{\sqrt{n^{-1}S^2}} \sim t_{n-1}.$$
- Υπολογίζοντας τις σταθερές c_1, c_2 ώστε να ισχύει
$$P\left(c_1 \leq \frac{\bar{X} - \mu}{\sqrt{n^{-1}S^2}} \leq c_2\right) = 1 - \alpha$$

(δ) Εκτίμηση του μ όταν $X \sim N(\mu, \sigma^2)$, σ^2 άγνωστο.

- παίρνουμε

$$P\left(\frac{\bar{X}-\mu}{\sqrt{n^{-1}S^2}} > c_2\right) = 1-\frac{\alpha}{2} \Rightarrow c_2 = t_{n-1,1-\frac{\alpha}{2}} = -t_{n-1,a/2},$$

$$P\left(\frac{\bar{X}-\mu}{\sqrt{n^{-1}S^2}} < c_1\right) = \frac{\alpha}{2} \Rightarrow c_1 = t_{n-1,\alpha/2}.$$

- Έτσι,

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

- οπότε, λύνοντας ως προς μ,

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

- Τελικά, ένα δ.ε. για τη μέση τιμή κανονικού πληθυσμού με άγνωστη διακύμανση είναι το

$$\left[\bar{X} - \frac{S}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}}, \bar{X} + \frac{S}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}}\right]. \tag{5}$$

(δ) Εκτίμηση του μ όταν $X \sim N(\mu, \sigma^2)$, σ^2 άγνωστο.

- Για μεγάλο n, $t_{n,\alpha} \approx z_\alpha$. Σε αυτή την περίπτωση η (5) γίνεται

$$\left[\bar{X} - n^{-1/2} S z_{1-\frac{\alpha}{2}}, \bar{X} + n^{-1/2} S z_{1-\frac{\alpha}{2}}\right].$$

- Το παραπάνω δ/μα ισχύει για οποιοδήποτε πληθυσμό για μεγάλο n.

Παράδειγμα 2.

Να βρεθεί ένα διάστημα εμπιστοσύνης συντελεστού $1 - \alpha = 95\%$ για τη μέση τιμή των δεδομένων του παραδείγματος 1 όταν όμως τώρα η διακύμανση θεωρείται άγνωστη.

Λύση: Υπολογίζοντας τη δειγματική διακύμανση

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{1}{19}((73-74.55)^2 + \cdots + (71-74.55)^2) = 31.42,$$

υπολογίζοντας $t_{19,0.025} = 2.093$ και αντικαθιστώντας στην (5),

$$\left[74.55 - 20^{-\frac{1}{2}} 31.42 \cdot 2.093, 74.55 + 20^{-\frac{1}{2}} 31.42 \cdot 2.093\right] = [71.92, 77.17].$$

Άρα το μέσο βάρος είναι μεταξύ 71.92 και 77.17 κιλών με σ. ε. 95%.

Άσκηση 1.

Θέλοντας να εκτιμήσουμε τη μέση τιμή μ του λίτρου της βενζίνης στα πρατήρια των Αθηνών επισκεφτήκαμε τυχαία $n = 10$ βενζινάδικα από όπου καταγράψαμε τις τιμές: 280.3, 282.8, 278.5, 283.1, 290.0, 284.9, 284.4, 279.8, 291.1, 286.7. α) Να δώσετε ένα δ.ε. συντελεστού 95% για τη μέση τιμή μ του λίτρου της βενζίνης στα πρατήρια του λεκανοπεδίου. Ποίο θα ήταν το αντίστοιχο δ.ε. αν ήταν γνωστό ότι $\sigma = 4$; β) Να δώσετε δ.ε. συντελεστού 95% για τη διασπορά και την τυπική απόκλιση της τιμής στο λεκανοπέδιο. Ποίο θα ήταν το αντίστοιχο δ.ε. αν ήταν γνωστό ότι $\mu = 280$; (Υποθέστε ότι οι τιμές της βενζίνης στα διάφορα πρατήρια ακολουθούν κανονική κατανομή).

Λύση: Από τα δεδομένα βρίσκουμε αμέσως $\bar{X} = 284.16$ και $S^2 = 17.58$. Έχουμε μικρό δείγμα και άγνωστη διακύμανση οπότε χρησιμοποιούμε τη σχέση (5). Αντικαθιστώντας, παίρνουμε,

$$\left[284.16 - \sqrt{\frac{17.58}{10}}t_{n-1,1-\frac{\alpha}{2}}, 284.16 + \sqrt{\frac{17.58}{10}}t_{n-1,1-\frac{\alpha}{2}}\right] = [281.16, 287.16]$$

γιατί έχουμε ότι $\alpha = 5\%$ οπότε $t_{n-1,1-\frac{\alpha}{2}} = t_{9,0.975} = 2.262$

Άσκηση 1: συνέχεια.

- Στην περίπτωση που είναι γνωστό ότι $\sigma = 4$, τότε η κατάλληλη σχέση είναι η (1) οπότε έχουμε

$$\left[\bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \bar{X} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right] \overset{z_{1-\frac{\alpha}{2}}=1.96,\ \alpha=5\%}{=} \left[284.16 - \frac{4 \cdot 1.96}{\sqrt{10}}, 284.16 + \frac{4 \cdot 1.96}{\sqrt{10}}\right] = [281.68, 286.64]$$

- Ένα δ.ε. για το σ^2 όταν μ άγνωστο θα προκύψει από την (4). Αντικαθιστώντας παίρνουμε

$$\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right] = \left[\frac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right] = [8.32, 58.609].$$

- Για το σ ένα δ/μα θα προκύψει από: $\sqrt{8.32} \leq \sigma \leq \sqrt{58.609}$
- Για μ γνωστό, αντικαθιστώντας στην (3),

$$\left[\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\chi^2_{n,1-\frac{\alpha}{2}}}, \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\chi^2_{n,\frac{\alpha}{2}}}\right] = [4.02^2, 10.101^2].$$

Άσκηση 2.

Έστω ότι επιθυμούμε να εκτιμήσουμε το μέσο χρόνο που κάνει ένα τρένο του Μετρό για να μεταβεί από το σταθμό Α στο σταθμό Β. Χρονομετρώντας τη διαδρομή αυτή 10 φορές σημειώνουμε τους χρόνους (σε δευτερόλεπτα) 357, 337, 351, 357, 350, 352, 360, 353, 377, 372. α) Να δοθεί ένα δ.ε. συντελεστού 99% για το μέσο χρόνο μετάβασης. β) Να δοθεί ένα δ.ε. συντελεστού 99% για την τυπική απόκλιση του χρόνου μετάβασης. (Υποθέστε ότι οι χρόνοι είναι κανονικοί).

Λύση: Από τα δεδομένα βρίσκουμε αμέσως $\bar{X} = 356.6$ και $S^2 = 128.711$. Έχουμε μικρό δείγμα και άγνωστη διακύμανση οπότε χρησιμοποιούμε τη σχέση (5) με $\alpha = 1\%$. Αντικαθιστώντας,

$$\left[356.6 - \sqrt{\frac{128.711}{10}}\, t_{9,0.995}, 356.6 + \sqrt{\frac{128.711}{10}}\, t_{9,0.995}\right] = [344.9, 368.2]$$

Ένα 99%, δ.ε. για το σ^2 όταν μ άγνωστο προκύπτει από την (4)

$$\left[\frac{(n-1)S^2}{\chi^2_{9,0.995}}, \frac{(n-1)S^2}{\chi^2_{9,0.005}}\right] = [49.10, 669.59].$$

Άσκηση 3.

Έστω ότι θέλουμε να εκτιμήσουμε το μέσο χρόνο ζωής των αρρένων κατοίκων μιας συγκεκριμένης περιοχής. Για το σκοπό αυτό ελήφθη τ.δ. μεγέθους $n = 200$ ανδρών. Βρέθηκε ότι,

$$\sum_{i=1}^{200} X_i = 13959.6, \ \sum_{i=1}^{200} X_i^2 = 977265$$

Υποθέτοντας ότι οι χρόνοι ζωής είναι κανονικοί, α) Να βρείτε ένα δ.ε. συντελεστού 95% για το μέσο χρόνο ζωής και β) Να βρείτε ένα δ.ε. συντελεστού 95% για την τυπική απόκλιση του χρόνου ζωής των ανδρών της περιοχής.

Λύση: Από τα δεδομένα βρίσκουμε αμέσως $\bar{X} = 69.79$.
Για τη διακύμανση χρησιμοποιούμε

$$S^2 = (n-1)^{-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) = 14.63.$$

Έχουμε άγνωστη διακύμανση οπότε χρησιμοποιούμε τη σχέση (5) με $\alpha = 5\%$.

Άσκηση 3: συνέχεια.

- Αντικαθιστώντας,

$$\left[69.79 - \sqrt{\frac{14.63}{200}} t_{199,0.975}, 69.79 + \sqrt{\frac{14.63}{200}} t_{199,0.975}\right]$$

- Το μέγεθος του δείγματος είναι μεγάλο οπότε $t_{199,0.9755} \approx z_{0.975} = 1.96$ οπότε τελικά το ζητούμενο δ/μα βγαίνει $[69.27, 70.33]$.
- Ένα δ.ε. για το σ^2 όταν μ άγνωστο προκύπτει από την (4)

$$\left[\frac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right]$$

- Το μέγεθος του δείγματος είναι μεγάλο οπότε $\chi^2_{n,\alpha} \approx n + \sqrt{2n} z_\alpha$.
- Επίσης, $z_{1-\alpha} = -z_\alpha$,
- οπότε τελικά το ζητούμενο δ/μα βγαίνει $[12.22, 18.20]$.
- Αντίστοιχα ένα δ/μα για το σ είναι $[\sqrt{12.22}, \sqrt{18.20}] = [3.49, 4.26]$.

Άσκηση 4.

Έστω ότι θέλουμε να κατασκευάσουμε δ.ε. συντελεστού $1-\alpha$ για το μέσο μ κανονικής κατανομής. α) Να βρείτε το ελάχιστο μέγεθος του δείγματος n που πρέπει να πάρουμε ώστε το δ.ε. να έχει πλάτος το πολύ c (το σ είναι γνωστό). Να γίνει εφαρμογή για $1-\alpha = 99\%, \sigma = 1, c = 0.1$. β) Να βρείτε το ελάχιστο n ώστε το δ.ε. να έχει πλάτος το πολύ σ. $(1-\alpha = 99\%)$. γ) Αν στο (α) το σ είναι άγνωστο, εκτιμήστε το μέγεθος n του δείγματος που πρέπει να πάρουμε χρησιμοποιώντας ένα αρχικό βοηθητικό δείγμα μεγέθους n_1 (υποθ. ότι n: μεγάλο)

Λύση: Για το πρώτο ερώτημα, από την (1) έχουμε

$$\left[\bar{X} - \sqrt{n^{-1}\sigma^2} z_{1-\frac{\alpha}{2}}, \bar{X} + \sqrt{n^{-1}\sigma^2} z_{1-\frac{\alpha}{2}}\right]$$

Ψάχνουμε το n ώστε το πλάτος του δ/τος να είναι το πολύ c άρα

$$2\sqrt{n^{-1}\sigma^2} z_{1-\frac{\alpha}{2}} \leq c \Leftrightarrow n \geq 4\sigma^2 c^{-2} z^2_{1-\frac{\alpha}{2}}.$$

Οπότε για $1-\alpha = 99\%, \sigma = 1, c = 0.1$ παίρνουμε $n \geq 2662.56 = 2663$.

Άσκηση 4: συνέχεια.

- Για το δεύτερο ερώτημα, για $c = \sigma$, από το πρώτο ερώτημα προκύπτει

$$n \geq 4\sigma^2\sigma^{-2}z^2_{1-\frac{\alpha}{2}} = 4z^2_{1-\frac{\alpha}{2}} = 4 \cdot 2.58^2 \approx 27.$$

- Για το τρίτο ερώτημα, στην ουσία θέλουμε να εκτιμήσουμε την

$$n_0 \geq 4\sigma^2 c^{-2} z^2_{1-\frac{\alpha}{2}}$$

  η οποία μπορεί να θεωρηθεί ως παραμετρική συνάρτηση του άγνωστου σ^2.

- Χρησιμοποιώντας τη δειγματική διασπορά s^2 σαν εκτιμητή της σ^2, ο εκτιμητής της ζητούμενης συνάρτησης είναι

$$\hat{n}_0 = 4s^2 c^{-2} z^2_{1-\frac{\alpha}{2}}.$$

- Ο εκτιμητής s^2 προκύπτει από ένα βοηθητικό δείγμα μεγέθους $n_1 < n$

Άσκηση 5.

Μία εταιρία συσκευασίας ενός προϊόντος (π.χ. ζάχαρης) επιθυμεί να εκτιμήσει το μέσο βάρος της συσκευασίας ενός ορισμένου τύπου (π.χ. συσκευασία που αναγράφει ότι περιέχει 100γρ) η οποία εξέρχεται από την παραγωγική διαδικασία. Για να μπορέσει η εταιρία αξιόπιστα να κρίνει αν η παραγωγή γίνεται ορθά, επιθυμεί να εκτιμήσει το μέσο βάρος έχοντας ακρίβεια δέκατου του γραμμαρίου με συντελεστή εμπιστοσύνης 95%. Λαμβάνοντας ένα αρχικό (βοηθητικό) δείγμα μεγέθους $n_1 = 100$ (και βρίσκοντας $S_1 = 2.12$) να βρεθεί μία σημειακή εκτίμηση και ένα δ.ε. 90% για το τελικό μέγεθος n του δείγματος που πρέπει να ληφθεί (υποθ. ότι τα βάρη κατανέμονται κανονικά).

Λύση: Από την προηγούμενη άσκηση, ξέρουμε ότι πρέπει να πάρουμε μέγεθος δείγματος

$$n = 4s^2c^{-2}z^2_{1-\alpha/2}$$

με $c = 0.1, \alpha = 0.05$. Ως εκτίμηση της διακύμανσης χρησιμοποιούμε την $S_1 = 2.12$. Οπότε

$$\hat{n} = 42.12^2 0.1^{-2} 1.96^2 = 6906.$$

Άσκηση 5: συνέχεια.

- Εφόσον έχουμε ήδη 100 παρατηρήσεις από το πρώτο δείγμα τελικά θέλουμε 6806 παρατηρήσεις.
- Επειδή $P(L \leq \sigma^2 \leq U) = 1 - \alpha$ όπου

$$\left[\frac{(n_1-1)S^2}{\chi^2_{n_1-1,1-\frac{\alpha}{2}}}, \frac{(n_1-1)S^2}{\chi^2_{n_1-1,\frac{\alpha}{2}}}\right] \equiv [L, U]$$

- για το δ.ε. του $\hat{n}$, βασιζόμενοι στο δ.ε. του σ^2, παίρνουμε

$$P\left(4Lc^{-2}z^2_{1-\frac{\alpha}{2}} \leq 4\sigma^2 c^{-2} z^2_{1-\frac{\alpha}{2}} \leq 4Uc^{-2}z^2_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

- Συνεπώς το ζητούμενο δ/μα είναι

$$[4Lc^{-2}z^2_{1-\frac{\alpha}{2}}, 4Uc^{-2}z^2_{1-\frac{\alpha}{2}}] =$$

$$[4\frac{(n_1-1)S^2}{\chi^2_{n_1-1,1-\frac{\alpha}{2}}}c^{-2}z^2_{1-\frac{\alpha}{2}}, 4\frac{(n_1-1)S^2}{\chi^2_{n_1-1,\frac{\alpha}{2}}}c^{-2}z^2_{1-\frac{\alpha}{2}}] = [5498, 8773]$$

  αφού αντικαταστήσαμε $\alpha = 0.1$. Άρα το τελικό μέγεθος του δείγματος δεν μπορεί να είναι μικρότερο από 5498 η μεγαλύτερο του 8773 με συντελεστή εμπιστοσύνης 90%.

Άσκηση 6.

Έστω $X_1, \ldots, X_n$ τ.δ. από μία άγνωστη κατανομή με μέσο μ και διασπορά σ^2. Δεδομένου ότι το μέγεθος n του δείγματος είναι αρκετά μεγάλο, να κατασκευάσετε προσεγγιστικό δ.ε. για το μ συντελεστού $1 - \alpha$.

Λύση: Παρατηρούμε ότι εδώ το τ.δ. δεν είναι απαραίτητα κανονικό και για αυτό δεν μπορούμε να χρησιμοποιήσουμε απευθείας το γνωστό δ.ε. για το μέσο κανονικής κατανομής όταν το σ^2 είναι γνωστό. Μπορούμε όμως μέσω κανονικής προσέγγισης να φτάσουμε σε παρόμοιο δ.ε. Από την (1) έχουμε ότι ένα προσεγγιστικό δ.ε. συντ. $1 - \alpha$ για το μ όταν το σ^2 είναι γνωστό είναι το

$$\left[\bar{X} - \sqrt{n^{-1}\sigma^2} z_{1-\frac{\alpha}{2}}, \bar{X} + \sqrt{n^{-1}\sigma^2} z_{1-\frac{\alpha}{2}}\right].$$

Τέλος, στην περίπτωση που το σ^2 είναι άγνωστο, υποθέτοντας ότι $\sigma^2 \approx s^2$ (η s^2 είναι συνεπής εκτιμήτρια του σ^2 και άρα $s^2 \to \sigma^2$ για $n \to +\infty$) προκύπτει το (προσεγγιστικό) δ.ε. για το μ συντελεστού $1 - \alpha$,

$$\left[\bar{X} - \sqrt{n^{-1}s^2} z_{1-\frac{\alpha}{2}}, \bar{X} + \sqrt{n^{-1}s^2} z_{1-\frac{\alpha}{2}}\right].$$

(ζ) Δ.Ε. για τη διαφορά των μέσων δύο ανεξάρτητων κανονικών πληθυσμών - γνωστές διακυμάνσεις.

- Έχουμε ανεξάρτητα δείγματα από δύο κανονικούς πληθυσμούς $X_1, \ldots, X_{n_1} \sim N(\mu_1, \sigma_1^2)$ και $Y_1, \ldots, Y_{n_2} \sim N(\mu_2, \sigma_2^2)$, με γνωστές διασπορές. Θέλουμε ένα δ.ε. για τη διαφορά $\mu_1 - \mu_2$.
- Η συνάρτηση για τον έλεγχο είναι η $T = \bar{X} - \bar{Y}$. Εφόσον σ_1^2, σ_2^2 γνωστά έχουμε

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$
$$\Rightarrow \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \equiv z \sim N(0,1). \quad (6)$$

- Όπως και στις προηγούμενες περιπτώσεις,

$$P\left(z_{\alpha/2} \leq \left\{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)\right\}\left\{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right\}^{-1} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

- Γιατί και πάλι, η διαφορά $\bar{X} - \bar{Y}$ είναι γραμμικός συνδυασμός ανεξάρτητων κανονικών τ.μ.

(ζ) Δ.Ε. για τη διαφορά των μέσων δύο ανεξάρτητων κανονικών πληθυσμών - γνωστές διακυμάνσεις.

- Λύνοντας ως προς $\mu_1 - \mu_2$, το ζητούμενο $(1-\alpha)100\%$ δ.ε. είναι

$$(\bar{X} - \bar{Y}) - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} z_{1-\frac{\alpha}{2}} \leq \mu_1 - \mu_2 \leq (\bar{X} - \bar{Y}) + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} z_{1-\frac{\alpha}{2}}. \quad (7)$$

- Εφαρμογή: Σε περίπτωση ίσων και γνωστών διακυμάνσεων, από την (6) με $\sigma_1 = \sigma_2 = \sigma$ παίρνουμε

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1).$$

- Οπότε ένα $(1-\alpha)100\%$ δ.ε. είναι

$$\left((\bar{X} - \bar{Y}) - \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} z_{1-\frac{\alpha}{2}}, (\bar{X} - \bar{Y}) + \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} z_{1-\frac{\alpha}{2}} \right). \quad (8)$$

(η) Δ.Ε. για τη διαφορά των μέσων δύο ανεξάρτητων κανονικών πληθυσμών - άγνωστες, ίσες διακυμάνσεις.

- Για να κατασκευάσουμε δ.ε. για άγνωστες αλλά ίσες διακυμάνσεις, σκεφτόμαστε ως εξής:
- Ξεκινάμε θεωρώντας το σ γνωστό. Πάλι από την (6) με $\sigma_1 = \sigma_2 = \sigma$ παίρνουμε

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1). \tag{9}$$

- Επίσης αν S_1^2 και S_2^2 δειγματικές διασπορές από τα 2 δείγματα τότε, όπως είδαμε στην (4),

$$\frac{(n_1 - 1)S_1^2}{\sigma^2} \sim \chi^2_{n_1-1}, \frac{(n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2_{n_2-1}.$$

- Οπότε, παίρνοντας υπ΄ όψιν ότι το άθροισμα ανεξάρτητων τ.μ. που ακολουθούν χ^2 κατανομή είναι και αυτό τ.μ. με χ^2 κατανομή, έχουμε

$$\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}. \tag{10}$$

(η) Δ.Ε. για τη διαφορά των μέσων δύο ανεξάρτητων κανονικών πληθυσμών - άγνωστες, ίσες διακυμάνσεις.

- Διαιρώντας την (9) με την (10), τελικά θα έχουμε

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\left(\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{\sigma^2(n_1+n_2-2)}}\right)^{-1} \sim t_{n_1+n_2-2}.$$

- Συνεπώς,

$$P\left(t_{n_1+n_2-2,\frac{\alpha}{2}} \leq \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{(n_1+n_2-2)}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \leq t_{n_1+n_2-2,1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

- Λύνοντας ως προς $\mu_1-\mu_2$, το ζητούμενο δ.ε. όταν $\sigma_1=\sigma_2$ είναι

$$\Bigg[\bar{X}-\bar{Y}-\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{(n_1+n_2-2)}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\,t_{n_1+n_2-2,1-\frac{\alpha}{2}},$$

$$\bar{X}-\bar{Y}+\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{(n_1+n_2-2)}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\,t_{n_1+n_2-2,1-\frac{\alpha}{2}}\Bigg] \qquad (11)$$

Άσκηση 7.

Έστω μ_1 και μ_2 οι μέσοι χρόνοι εξυπηρέτησης των πελατών από δύο ταμίες μιας τράπεζας. Αν 234, 99, 234, 174, 188, 107, 173, 172 και 105, 194, 77, 33, 159, 150, 167, 127, 169, 166 είναι δειγματοληπτικά κάποιοι χρόνοι (σε sec) εξυπηρέτησης των δύο αυτών υπαλλήλων αντίστοιχα, να βρείτε δ.ε. συντελεστού 95% για τη διαφορά $\mu_1 - \mu_2$ υποθέτοντας ότι οι χρόνοι εξυπηρέτησης είναι κανονικοί με $\sigma_1 = \sigma_2 = 40$ sec. Με βάση το συγκεκριμένο δ.ε. μπορούμε να πούμε ότι οι δύο υπάλληλοι έχουν διαφορετική απόδοση;

Λύση: Το δ.ε. στην περίπτωση που οι διακυμάνσεις είναι γνωστές και ισες είναι εφαρμογή της (8) οπότε έχουμε να υπολογίσουμε τις

$$L = (\bar{X} - \bar{Y}) - \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}z_{1-\frac{\alpha}{2}}, R = (\bar{X} - \bar{Y}) + \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}z_{1-\frac{\alpha}{2}}$$

όπου μετά από πράξεις $\bar{X} = 172.62$, $\bar{Y} = 134.7$, $z_{1-\alpha/2} = 1.96$ οπότε αν αντικαταστήσουμε παίρνουμε $[0.73, 75.11]$

Το συμπέρασμα είναι ότι εφόσον $\mu_1 - \mu_2 > 0.73$ με συντελεστή 95% μπορούμε να πούμε ότι οι υπάλληλοι έχουν διαφορετική απόδοση και μάλιστα ο πρώτος φαίνεται να είναι πιο αποδοτικός από τον δεύτερο.

Άσκηση 8.

Έστω μ_1 η μέση τιμή πώλησης ενός προϊόντος σε μία περιοχή A και μ_2 η μέση τιμή πώλησης του ίδιου προϊόντος σε μία περιοχή B. Η μέση τιμή και η διασπορά ενός τ.δ. 10 τιμών πώλησης από την περιοχή A βρέθηκε 100.9 και 8.76667 αντίστοιχα. Επίσης, η μέση τιμή και η διασπορά ενός τ.δ. 20 τιμών πώλησης από την περιοχή B βρέθηκε 104.45 και 12.9974 αντίστοιχα. Αν υποθέσουμε ότι οι τιμές κατανέμονται κανονικά και με ίση (αλλά άγνωστη) διασπορά και στις δύο περιοχές, να βρείτε δ.ε. συντελεστού 95% για τη διαφορά $\mu_1 - \mu_2$. Μπορούμε με βάση το δ.ε. να πούμε ότι η μέση τιμή πώλησης στην περιοχή A είναι διαφορετική από την αντίστοιχη στην περιοχή B;

Λύση: Η άσκηση είναι απευθείας αντικατάσταση στην (11) όπου $t_{n_1+n_2-2,0.975} = 2.048$ μετά από τις πράξεις παίρνουμε

$$-6.25 \leq \mu_1 - \mu_2 \leq -0.845.$$

Μπορούμε τελικά να πούμε ότι η μέση τιμή πώλησης στην περιοχή A είναι διαφορετική και μάλιστα χαμηλότερη από τη μέση τιμή πώλησης στην περιοχή B με συντελεστή εμπιστοσύνης τουλάχιστον 95%.

(θ) Διάστημα εμπιστοσύνης για το λόγο των διασπορών δύο ανεξάρτητων κανονικών πληθυσμών.

- Με δεδομένα ανεξάρτητα δείγματα από δύο κανονικούς πληθυσμούς $X_1, \ldots, X_{n_1} \sim N(\mu_1, \sigma_1^2)$ και $Y_1, \ldots, Y_{n_2} \sim N(\mu_2, \sigma_2^2)$,
- Θέλουμε ένα δ.ε. συντελεστή $1 - \alpha$ για το πηλίκο σ_2^2/σ_1^2.
- Διαστήματα αυτής της μορφής χρησιμοποιούνται συνήθως για τη σύγκριση των δύο διασπορών.
- Όταν οι μέσες τιμές μ_1, μ_2 είναι γνωστές, ξέρουμε ότι

$$\frac{nT_k}{\sigma_k^2} = \sum_{i=1}^{n} \left(\frac{R_i - \mu_k}{\sigma_k} \right)^2 \sim \chi_{n_k}^2, k = 1, 2, R_1 = X_i, R_2 = Y_i$$

  και ότι T_1, T_2 είναι μεταξύ τους ανεξάρτητες αφού προέρχονται από διαφορετικά και ανεξάρτητα δείγματα, όπου στο κάθε δείγμα οι παρατηρήσεις είναι ισόνομες.
- Παρατηρούμε ότι

$$\frac{\frac{nT_1}{\sigma_1^2}}{\frac{nT_2}{\sigma_2^2}} = \frac{\sigma_2^2 T_1}{\sigma_1^2 T_2} \sim \frac{\chi_{n_1}^2 n_1^{-1}}{\chi_{n_2}^2 n_2^{-1}} \equiv F_{n_1, n_2}$$

(ϑ) Διάστημα εμπιστοσύνης για το λόγο των διασπορών δύο ανεξάρτητων κανονικών πληθυσμών.

- Όπως και πριν,

$$P\left(F_{n_1,n_2,\frac{\alpha}{2}} \leq \sigma_2^2 T_1 \left(\sigma_1^2 T_2\right)^{-1} \leq F_{n_1,n_2,1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

- Οπότε λύνοντας ως προς σ_2^2/σ_1^2 ένα δ.ε. συντελεστού $1-\alpha$ για το εν λόγω πηλίκο με μ_1, μ_2 γνωστά είναι το

$$\left[\frac{n_1\sum_{i=1}^{n_2}(Y_i-\mu_2)^2}{n_2\sum_{i=1}^{n_1}(X_i-\mu_1)^2}F_{n_1,n_2,\frac{\alpha}{2}}, \frac{n_1\sum_{i=1}^{n_2}(Y_i-\mu_2)^2}{n_2\sum_{i=1}^{n_1}(X_i-\mu_1)^2}F_{n_1,n_2,1-\frac{\alpha}{2}}\right].$$

- Για μ_1, μ_2 άγνωστα, χρησιμοποιώντας τον δειγματικό μέσο έναντι του πραγματικού στον τύπο των T_1, T_2 παίρνουμε ότι ένα δ.ε. $(1-\alpha)100\%$ για έλεγχο ισότητας διακυμάνσεων είναι το

$$\left[\frac{(n_1-1)\sum_{i=1}^{n_2}(Y_i-\bar{Y})^2}{(n_2-1)\sum_{i=1}^{n_1}(X_i-\bar{X})^2}F_{n_1-1,n_2-1,\frac{\alpha}{2}}, \frac{(n_1-1)\sum_{i=1}^{n_2}(Y_i-\bar{Y})^2}{(n_2-1)\sum_{i=1}^{n_1}(X_i-\bar{X})^2}F_{n_1-1,n_2-1,1-\frac{\alpha}{2}}\right] \quad (12)$$

- και χρησιμοποιείται για έλεγχο της υπόθεσης στα πιο πάνω τεστ.

Άσκηση 9.

Στην Άσκηση 8 που αφορούσε τις τιμές πώλησης ενός προϊόντος σε δύο περιοχές A και B υποθέσαμε ότι οι διασπορές των τιμών στις περιοχές αυτές είναι ίσες. Βρείτε ένα δ.ε. συντελεστού 95% για το πηλίκο σ_2^2/σ_1^2. Μπορούμε να πούμε με συντελεστή εμπιστοσύνης 95% ότι οι διασπορές αυτές είναι άνισες;

Λύση: Έχουμε άγνωστη μέση τιμή άρα χρης. την σχέση (12) για την οποία έχουμε

$$\frac{(n_1-1)\sum_{i=1}^{n_2}(Y_i-\bar{Y})^2}{(n_2-1)\sum_{i=1}^{n_1}(X_i-\bar{X})^2}F_{n_1-1,n_2-1,\alpha/2} = \frac{S_2^2}{S_1^2}F_{n_1-1,n_2-1,\alpha/2},$$

$$\frac{(n_1-1)\sum_{i=1}^{n_2}(Y_i-\bar{Y})^2}{(n_2-1)\sum_{i=1}^{n_1}(X_i-\bar{X})^2}F_{n_1-1,n_2-1,1-\alpha/2} = \frac{S_2^2}{S_1^2}F_{n_1-1,n_2-1,1-\alpha/2}$$

όπου

$$F_{n_1-1,n_2-1,1-\alpha/2} = F_{9,19,0.975} = 2.88, F_{9,19,0.025} = 0.271.$$

Με αντικατάσταση, το δ/μα θα γίνει $[0.402, 4.27]$.
Το διάστημα αυτό περιέχει το 1 και άρα δεν μπορούμε να αποκλείσουμε ότι $\sigma_2^2/\sigma_1^2 = 1$ με συντελεστή εμπιστοσύνης 95%.