

Διαστήματα Εμπιστοσύνης (*Confidence Intervals*)

Διαστήματα Εμπιστοσύνης (*Confidence Intervals*)

- **Εκτίμηση παραμέτρου:** «Στόχος είναι η κατασκευή ενός μέτρου αβεβαιότητας που να σχετίζεται με την εκτίμηση της τιμής μιας παραμέτρου ενός πληθυσμού»
- Μια σημειακή εκτίμηση συνήθως δεν είναι πολύ ακριβής
- Μια **εκτίμηση διαστήματος** εμπεριέχει μεγαλύτερη πληροφορία παρέχοντας παράλληλα ένα **μέτρο** (ή έναν **βαθμό**) **εμπιστοσύνης** της εκτίμησης.
- Το διάστημα αυτό καλείται **διάστημα εμπιστοσύνης**

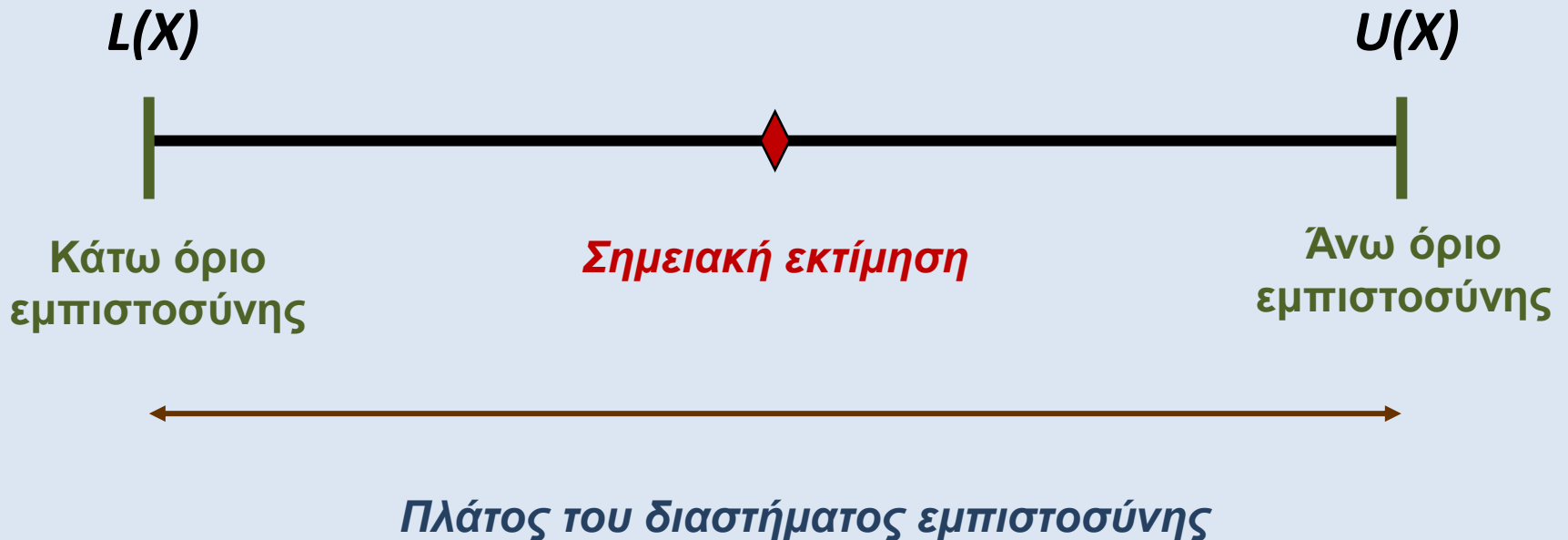
Ορισμός

- Έστω $\mathbf{X}=\{X_1, X_2, \dots, X_n\}$ ένα τυχαίο δείγμα πληθυσμού με κατανομή με σ.π.π. $f(x|\theta)$, όπου θ μια **άγνωστη παράμετρο** με τιμές σε έναν παραμετρικό χώρο Θ , $\theta \in \Theta$.

- Το τυχαίο διάστημα $[L, U]$ με $L \leq U$ και $L(\mathbf{X})$, $U(\mathbf{X})$ δύο **στατιστικές συναρτήσεις** με τιμές στον χώρο Θ , ονομάζεται **διάστημα εμπιστοσύνης (δ.ε.)** για την παράμετρο θ με **συντελεστή εμπιστοσύνης $1-\alpha$** ($0 < \alpha < 1$) αν ισχύει ότι:

$$P\left(L(\mathbf{X}) \leq \theta \leq U(\mathbf{X})\right) = 1 - \alpha$$

Διάστημα εμπιστοσύνης



$1-\alpha$: μέτρο αβεβαιότητας του διαστήματος,
συντελεστής εμπιστοσύνης της εκτίμησης

α : εκφράζει το **σφάλμα** της εκτίμησης

Μεθοδολογία κατασκευής διαστημάτων εμπιστοσύνης

1. Ορίζουμε μία **εκτιμήτρια συνάρτηση** $T(X)$, συνήθως μία εκτιμήτρια της παραμέτρου θ , με κατανομή που εξαρτάται από την θ .
2. Κατασκευάζουμε μία **συνάρτηση** $Y=g(T, \theta)$ η κατανομή της οποίας είναι **γνωστή** και δεν εξαρτάται από το θ ή άλλες παραμέτρους.

3. Υπολογίζουμε **δύο σταθερές** c_1, c_2 : $P(c_1 \leq Y \leq c_2) = 1 - \alpha$

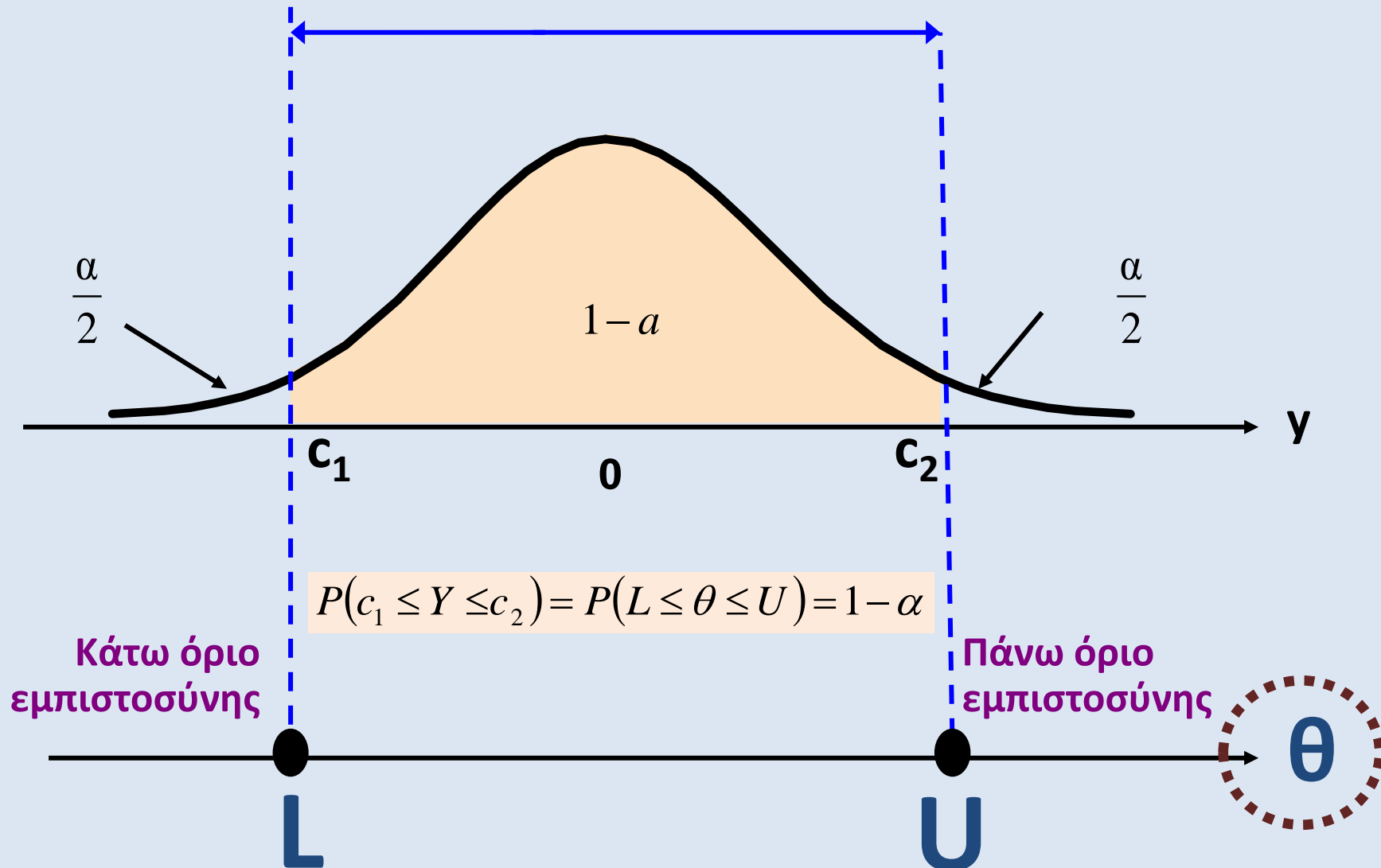
Υπόθεση $P(Y < c_1) = P(Y > c_2) = \alpha/2$, δηλ. οι ουρές της κατανομής στις δύο σταθερές έχουν ίσο μέτρο πιθανότητας ή ίσα εμβαδά, $\alpha/2$.

4. Λύνουμε τη σχέση $c_1 < Y=g(T, \theta) < c_2$ ως προς θ και έτσι προκύπτει

$$L(X) < \theta < U(X)$$

5. Το $[L, U]$ είναι το **διάστημα εμπιστοσύνης** της παραμέτρου θ με **συντελεστή εμπιστοσύνης** $1-\alpha$.

Γενική μεθοδολογία κατασκευής δ.ε.



Περιπτώσεις διαστήματος εμπιστοσύνης (δ.ε.)

(Α)	Μέσο με γνωστή διασπορά του πληθυσμού, σ (ή ποσοστό)	
(Β)	Μέσο με άγνωστη διασπορά του πληθυσμού, σ (ή ποσοστό)	
(Γ)	Διασπορά με γνωστό μέσο του πληθυσμού	
(Δ)	Διασπορά με άγνωστο μέσο του πληθυσμού	
(Ε1)	Διαφορά μέσων με γνωστές διασπορές	
(Ε2)	Διαφορά μέσων με άγνωστές διασπορές	
(ΣΤ1)	Λόγος διασπορών με άγνωστα μέσα	
(ΣΤ2)	Λόγος διασπορών με γνωστά μέσα	

(Α) δ.ε. για το μέσο μ όταν η διασπορά του πληθυσμού σ^2 , είναι γνωστή

- Έστω τυχαίο δείγμα $X=\{X_1, X_2, \dots, X_n\}$ όπου υποθέτουμε ότι προέρχεται από την κανονική κατανομή $N(\mu, \sigma_0^2)$, με **άγνωστο μέσο μ** αλλά **γνωστή διασπορά $\sigma^2=\sigma_0^2$** .
- Για την **εύρεση του δ.ε.** για το μ ακολουθούμε τα βήματα της γενικής μεθοδολογίας

(A) δ.ε. **για το μέσο μ** όταν η **διασπορά σ^2** είναι γνωστή

1. Χρησιμοποιούμε ως **εκτιμήτρια συνάρτηση $T(X)$** τον δειγματικό μέσο:

$$T(X) = \bar{X} = \frac{1}{n} \sum_i X_i$$

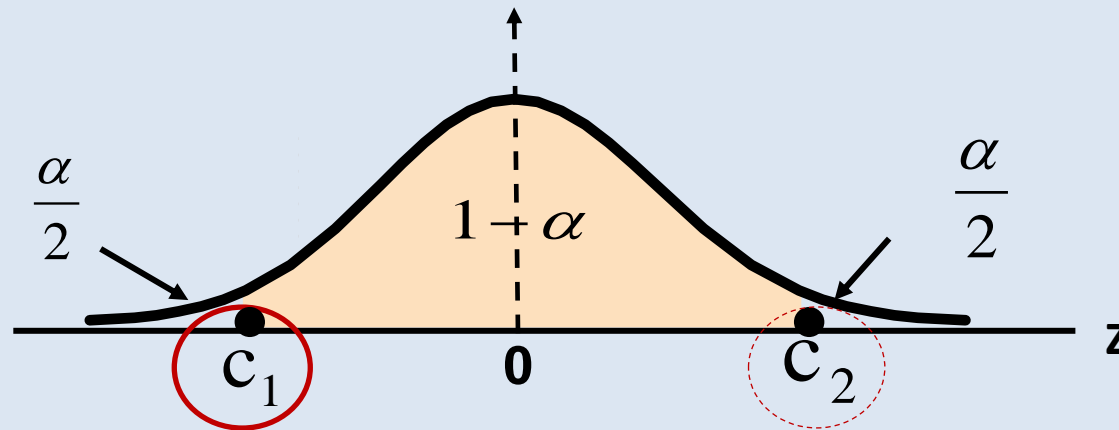
2. Εύρεση **στατιστικής συνάρτησης (Y)** για την οποία γνωρίζουμε την κατανομή της:

$$Y = \frac{\bar{X} - \mu}{\sigma_0 / \sqrt{n}} \sim N(0,1)$$

3. Εύρεση των **σταθερών c_1 και c_2** έτσι ώστε:

$$P(c_1 \leq Y \leq c_2) = 1 - \alpha$$

$$Y \sim N(0,1)$$



$$P(Y > c_2) = \frac{\alpha}{2} \Rightarrow c_2 = z_{\alpha/2}$$

$$P(Y < c_1) = \frac{\alpha}{2} \Rightarrow P(Y > c_1) = 1 - \frac{\alpha}{2} \Rightarrow c_1 = z_{1-\alpha/2} = -z_{\alpha/2}$$

**λόγω
συμμετρίας**

4. Αντικαθιστώντας τις τιμές των c_1 και c_2 και λύνοντας ως προς τη **παράμετρο μ** έχουμε:

$$P\left(c_1 \leq \frac{\bar{X} - \mu}{\sigma_0 / \sqrt{n}} \leq c_2\right) = 1 - a \Rightarrow$$

$$P\left(-z_{a/2} \leq \frac{\bar{X} - \mu}{\sigma_0 / \sqrt{n}} \leq z_{a/2}\right) = 1 - a \Rightarrow$$

$$P\left(\bar{X} - \frac{\sigma_0}{\sqrt{n}} z_{a/2} \leq \mu \leq \bar{X} + \frac{\sigma_0}{\sqrt{n}} z_{a/2}\right) = 1 - a$$

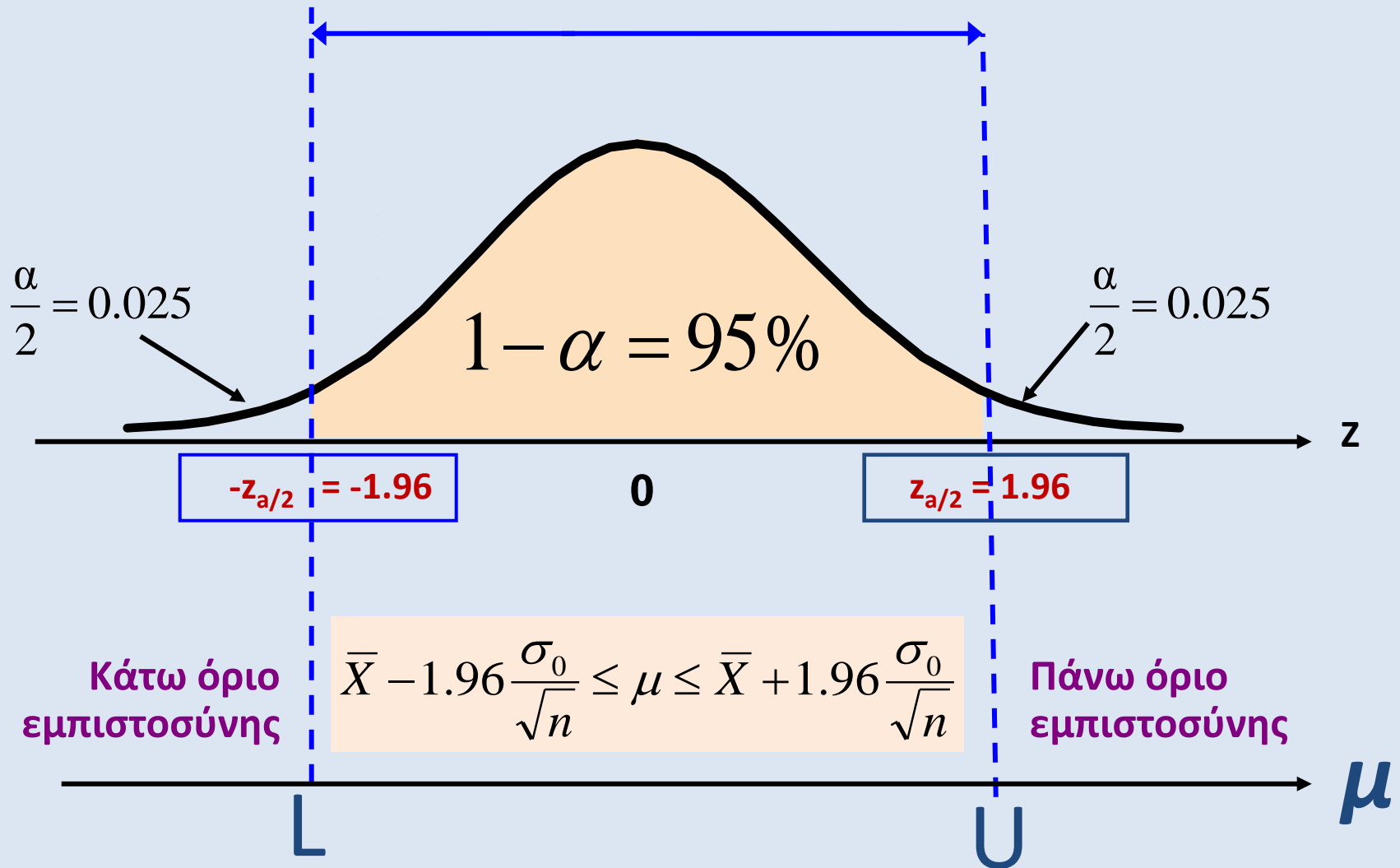
Διάστημα εμπιστοσύνης για το **μέσο μ** όταν το **σ** είναι γνωστό (**σ_0**), με **συντελεστή εμπιστοσύνης $1-\alpha$**

$$\bar{X} - \frac{\sigma_0}{\sqrt{n}} z_{\alpha/2} \leq \mu \leq \bar{X} + \frac{\sigma_0}{\sqrt{n}} z_{\alpha/2}$$

- **Παρατήρηση:** το **πλάτος** του διαστήματος εμπιστοσύνης γίνεται **ελάχιστο** στις **περιπτώσεις** όπου:
 - το **n** (μέγεθος δείγματος) **μεγαλώνει**
 - η (γνωστή) διακύμανση **σ_0** **μικραίνει**
 - το **$z_{\alpha/2}$** **μικραίνει**, δηλ. το **σφάλμα α** **μεγαλώνει**

π.χ. **95%** διάστημα εμπιστοσύνης:

$$1 - \alpha = 0.95 \quad \alpha = 0.05 \quad \alpha / 2 = 0.025$$



(B) δ.ε. για το μέσο μ όταν η διασπορά σ^2 είναι άγνωστη

- Έστω τυχαίο δείγμα $X=\{X_1, X_2, \dots, X_n\}$ όπου υποθέτουμε ότι προέρχεται από την κανονική κατανομή $N(\mu, \sigma^2)$, με **άγνωστο μέσο μ** και **άγνωστη διασπορά σ^2** .
- Για την εύρεση **δ.ε. για το μ** ακολουθούμε τα βήματα της γενικής μεθοδολογίας:

(B) δ.ε. **για το μέσο μ** όταν η **διασπορά σ^2** είναι άγνωστή

1. Χρησιμοποιούμε τον **δειγματικό μέσο ως εκτιμήτρια** συνάρτηση $T(X)$:

$$T(X) = \bar{X} = \frac{1}{n} \sum_i X_i$$

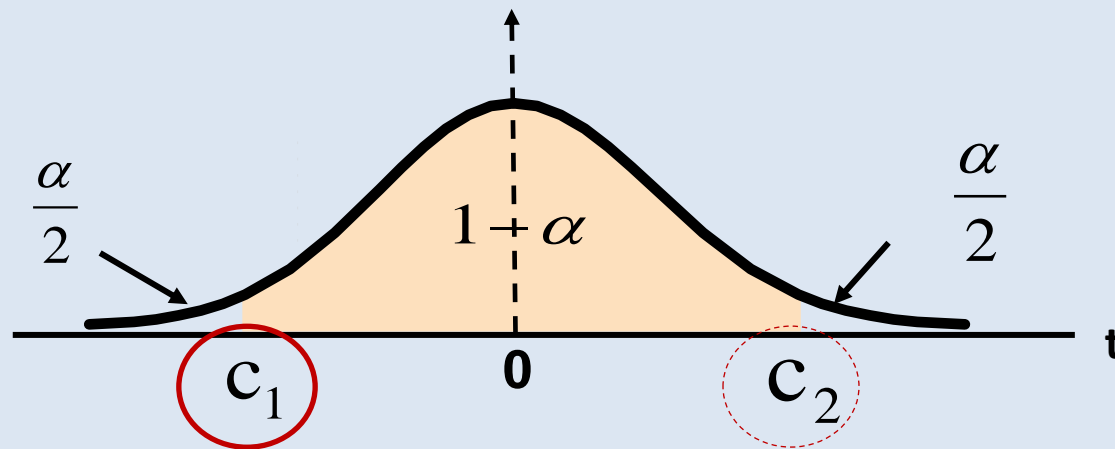
2. Ορίζουμε ως **στατιστική συνάρτηση (Y)** με γνωστή κατανομή:

$$Y = \frac{\bar{X} - \mu}{S / \sqrt{n}} \sim t_{n-1}$$

όπου S η **δειγματική διασπορά**

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

$$Y \sim t_{n-1}$$



3. Εύρεση των **σταθερών c_1 και c_2** έτσι ώστε:

$$P(c_1 \leq Y \leq c_2) = 1 - \alpha$$

$$P(Y > c_2) = \frac{\alpha}{2} \Rightarrow c_2 = t_{n-1}\left(\frac{\alpha}{2}\right)$$

**λόγω
συμμετρίας**

$$P(Y < c_1) = \frac{\alpha}{2} \Rightarrow P(Y > c_1) = 1 - \frac{\alpha}{2} \Rightarrow c_1 = t_{n-1}\left(1 - \frac{\alpha}{2}\right) = -t_{n-1}\left(\frac{\alpha}{2}\right)$$

4. Έτσι το **δ.ε.** κατασκευάζεται από την ανίσωση λύνοντας ως προς τη **παράμετρο μ** :

$$P\left(c_1 \leq Y = \frac{\bar{X} - \mu}{S / \sqrt{n}} \leq c_2\right) = 1 - a \Rightarrow$$

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

$$P\left(\bar{X} - \frac{S}{\sqrt{n}} t_{n-1}(a/2) \leq \mu \leq \bar{X} + \frac{S}{\sqrt{n}} t_{n-1}(a/2)\right) = 1 - a$$

Διάστημα εμπιστοσύνης για το μέσο μ με συντελεστή εμπιστοσύνης $1-\alpha$

- Με **γνωστή** διασπορά $\sigma=\sigma_0$

$$\bar{X} - \frac{\sigma_0}{\sqrt{n}} z_{\alpha/2} \leq \mu \leq \bar{X} + \frac{\sigma_0}{\sqrt{n}} z_{\alpha/2}$$

- Με **άγνωστη** διασπορά

$$\bar{X} - \frac{S}{\sqrt{n}} t_{n-1} \left(\frac{\alpha}{2} \right) \leq \mu \leq \bar{X} + \frac{S}{\sqrt{n}} t_{n-1} \left(\frac{\alpha}{2} \right)$$

(Γ) δ.ε. για τη διασπορά σ^2 όταν ο μέσος μ είναι γνωστός

- Έστω τυχαίο δείγμα $X=\{X_1, X_2, \dots, X_n\}$ όπου υποθέτουμε ότι προέρχεται από την κανονική κατανομή $N(\mu, \sigma^2)$, με γνωστό μέσο μ και άγνωστη διασπορά σ^2 .
- Για την εύρεση δ.ε. για το σ^2 ακολουθούμε τα βήματα της γενικής μεθοδολογίας:

(Γ) δ.ε. για τη διασπορά σ^2 όταν ο μέσος μ είναι γνωστός

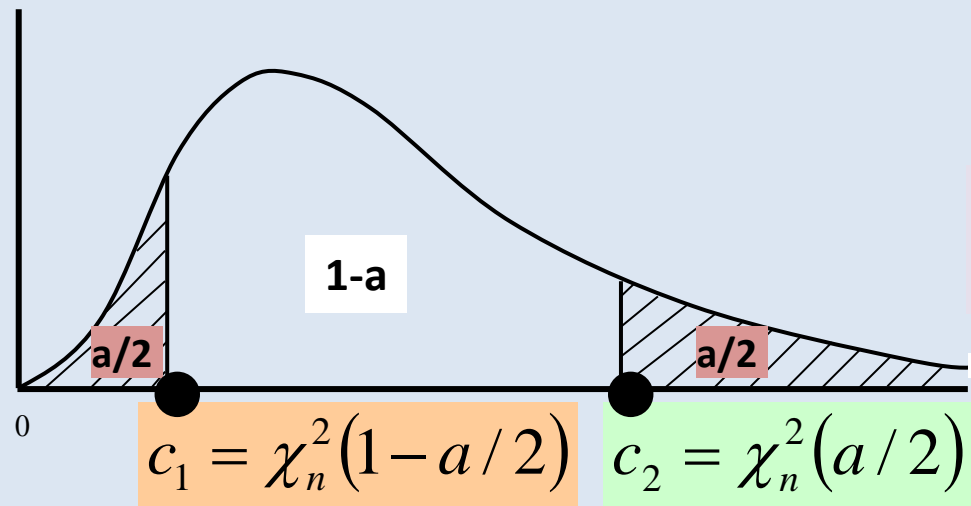
1. Χρησιμοποιούμε την **δειγματική διακύμανση με μ γνωστό** ως εκτιμήτρια συνάρτηση

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

2. Ορίζουμε ως **στατιστική συνάρτηση (Y)** με γνωστή κατανομή:

$$Y = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi_n^2$$

$$Y \sim \chi_n^2$$



όχι συμμετρική
κατανομή

3. Εύρεση των **σταθερών c_1 και c_2** έτσι ώστε:


$$P(c_1 \leq Y \leq c_2) = 1 - a$$

$$P(Y > c_2) = a/2 \Rightarrow c_2 = \chi_n^2(a/2)$$

$$P(Y < c_1) = a/2 \Rightarrow P(Y > c_1) = 1 - a/2 \Rightarrow c_1 = \chi_n^2(1 - a/2)$$

4. Το δ.ε. κατασκευάζεται από την ανίσωση λύνοντας ως προς την **παράμετρο σ^2** :

$$c_1 < Y < c_2 \Rightarrow \chi_n^2 \left(1 - \frac{a}{2} \right) \leq \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \leq \chi_n^2 \left(\frac{a}{2} \right) \Rightarrow$$


$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_n^2 \left(\frac{a}{2} \right)} \leq \sigma^2 \leq \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_n^2 \left(1 - \frac{a}{2} \right)}$$

L(X) **U(X)**

δ.ε. για την διασπορά σ^2 με **γνωστό μέσο μ**

(Δ) δ.ε. για τη διασπορά σ^2 όταν ο μέσος μ είναι άγνωστος

- Έστω τυχαίο δείγμα $X=\{X_1, X_2, \dots, X_n\}$ όπου υποθέτουμε ότι προέρχεται από την κανονική κατανομή $N(\mu, \sigma^2)$, με **άγνωστό μέσο μ** και **άγνωστη διασπορά σ^2** .
- Για την εύρεση δ.ε. για το σ^2 ακολουθούμε τα βήματα της γενικής μεθοδολογίας.

(Δ) δ.ε. για τη διασπορά σ^2 όταν ο μέσος μ είναι άγνωστος

1. Χρησιμοποιούμε την **δειγματική διακύμανση** ως εκτιμήτρια συνάρτηση:

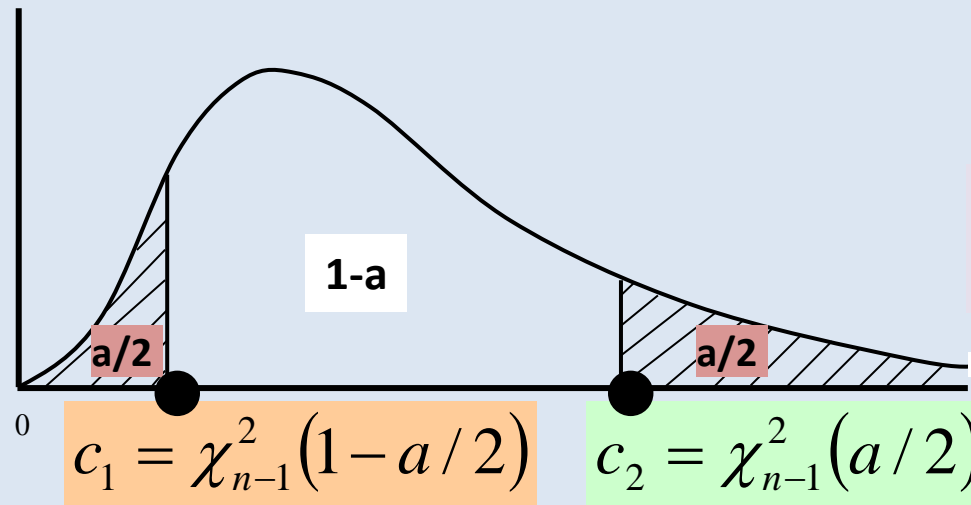
$$T(X) = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

2. Ορίζουμε ως **στατιστική συνάρτηση (Y)** με γνωστή κατανομή:

$$Y = \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim X_{n-1}^2$$

$$Y \sim \chi_{n-1}^2$$

όχι συμμετρική
κατανομή



3. Εύρεση των **σταθερών c_1 και c_2** έτσι ώστε:

$$P(c_1 \leq Y \leq c_2) = 1 - a$$

$$P(Y > c_2) = a/2 \Rightarrow c_2 = \chi_{n-1}^2(a/2)$$

$$P(Y < c_1) = a/2 \Rightarrow P(Y > c_1) = 1 - a/2 \Rightarrow c_1 = \chi_{n-1}^2(1 - a/2)$$

$$Y = \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2}$$

4. Το δ.ε. κατασκευάζεται από την ανίσωση:

$$c_1 < Y < c_2 \Rightarrow \chi_{n-1}^2(1-a/2) \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi_{n-1}^2(a/2) \Rightarrow$$



$$\frac{(n-1)S^2}{\chi_{n-1}^2(a/2)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{n-1}^2(1-a/2)}$$

L(X)

U(X)

(ή)

$$Y = \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2}$$

$$c_1 < Y < c_2 \Rightarrow \chi_{n-1}^2(1-a/2) \leq \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \leq \chi_{n-1}^2(a/2) \Rightarrow$$



$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi_{n-1}^2(a/2)} \leq \sigma^2 \leq \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi_{n-1}^2(1-a/2)}$$

L(X)

U(X)

δ.ε. για την διασπορά σ^2

- Με **γνωστό** μέσο, μ

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_n^2\left(\frac{a}{2}\right)} \leq \sigma^2 \leq \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_n^2\left(1 - \frac{a}{2}\right)}$$

- Με **άγνωστο** μέσο, μ

$$\frac{(n-1)S^2}{\chi_{n-1}^2(a/2)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{n-1}^2(1-a/2)}$$

ή

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi_{n-1}^2(a/2)} \leq \sigma^2 \leq \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi_{n-1}^2(1-a/2)}$$