ΑΝΑΛΥΣΗ ΧΡΟΝΟΣΕΙΡΩΝ ΚΑΙ ΔΕΔΟΜΕΝΩΝ ΠΕΡΙΒΑΛΛΟΝΤΙΚΩΝ ΚΙΝΔΥΝΩΝ

Καθ. Θεόδωρος Καρακασίδης
Δρ Αθανάσιος Φράγκου

Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών
«Βιώσιμη Διαχείριση Περιβαλλοντικών Αλλαγών και Κυκλική Οικονομία»

Στατιστική Συμπερασματολογία

- Συμπερασματολογία: Εξαγωγή συμπερασμάτων (αποφάσεις ή προβλέψεις για το μέλλον)

  - Στατιστική Συμπερασματολογία: Εκτίμηση παραμέτρων και έλεγχος τιμών τους με έλεγχο υποθέσεων.

- Πρακτικά στη Στατιστική Συμπερασματολογία: εξάγουμε το συμπέρασμα με μια μικρή αβεβαιότητα (διότι από ένα δείγμα εξάγουμε συμπέρασμα για ολόκληρο πληθυσμό) και μετράμε την ορθότητά του.

Εκτιμητική

- Εκτιμητική: Εκτίμηση παραμέτρων πληθυσμού

πχ Εκτίμηση του μέσου ύψους των Ελλήνων (θεωρούμε ένα δείγμα από όλο τον πληθυσμό).

- Παράμετρος: Ύψος
- Πληθυσμός: Έλληνες

- Τρόποι Εκτίμησης:

-Σημειακή Εκτίμηση: Δείγμα από πληθυσμό με μια άγνωστη παράμετρο

-Εκτίμηση σε διάστημα: Διάστημα εμπιστοσύνης στο οποίο ανήκει η πληθυσμιακή παράμετρος με κάποια προκαθορισμένη πιθανότητα.

Σημειακή Εκτίμηση

- Σημειακή Εκτίμηση: Εκτίμηση της μέσης τιμής μ ενός πληθυσμού

- Εκτιμήτρια $\overline{X}$, εκτίμηση της μέσης τιμής $\bar{x}$

- Αξιολόγηση εκτιμητών

- *Αμεροληψία* -> Η μέση τιμή της εκτιμήτριας ισούται με την παράμετρο

- *Αποτελεσματικότητα* -> Η Αποτελεσματική εκτιμήτρια έχει την ελάχιστη διακύμανση

- *Επάρκεια* -> Χρήση όλης της πληροφορίας που περιέχεται στο δείγμα για την παράμετρο

- *Συνέπεια* -> Όσο αυξάνει το μέγεθος του δείγματος, τόσο μεγαλώνει η πιθανότητα η εκτίμηση να είναι κοντά στην πραγματική τιμή της παραμέτρου

Εκτίμηση σε διάστημα
Διαστήματα εμπιστοσύνης

Διάστημα εμπιστοσύνης μιας παραμέτρου της κατανομής ενός πληθυσμού ονομάζουμε ένα διάστημα με κέντρο ένα σημειακό εκτιμητή της παραμέτρου αυτής που περιέχει με δεδομένη πιθανότητα την παράμετρο.

Παράδειγμα: Θεωρούμε δείγμα 120 αγοριών, ηλικίας 15 ετών. Εκτιμούμε ένα διάστημα με κέντρο την μέση τιμή του δείγματος στο οποίο βρίσκεται η μέση τιμή του ύψους όλων των αγοριών ηλικίας 15 ετών της χώρας.

Εκτίμηση σε διάστημα
Διαστήματα εμπιστοσύνης

Εύρεση διαστήματος μέσα στο οποίο θα περιέχεται η πραγματική αλλά άγνωστη τιμή της παραμέτρου (άγνωστη παράμετρος θ) γίνεται με τη βοήθεια στατιστικού που περιέχει την εκτιμώμενη παράμετρο και μια στοχαστικής πρόταση για το στατιστικό.

Παράδειγμα: Για την εξαγωγή συμπερασμάτων και τον έλεγχο υποθέσεων για τη μέση τιμή ενός πληθυσμού από ένα δείγμα μεγέθους n γίνεται με τη βοήθεια του στατιστικού:

$$Z = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

Όπου μ μέση τιμή του πληθυσμού, $\bar{X}$ η δειγματική μέση τιμή, *s* τυπική απόκλιση για το δεδομένο δείγμα που ακολουθεί την τυποποιημένη κανονική κατανομή.

Εκτίμηση σε διάστημα
Διαστήματα εμπιστοσύνης

Η στοχαστική πρόταση για το στατιστικό είναι:

$$P\left(|Z| \leq z_{a/2}\right) = 1 - a$$

Δηλαδή η μέση τιμή μ του πληθυσμού βρίσκεται με πιθανότητα 1-α στο διάστημα εμπιστοσύνης

$$100(1 - a)\%$$

1-α συντελεστής εμπιστοσύνης

α στάθμη σημαντικότητας

Διαστήματα εμπιστοσύνης για τη μέση τιμή του πληθυσμού
1. Γνωστή διασπορά του πληθυσμού
Μικρά και μεγάλα δείγματα

Γνωστή διασπορά: Για κάθε μ, η τυχαία μεταβλητή

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

ακολουθεί την τυπική κανονική κατανομή N(0,1).

Ένα 100(1-α)% διάστημα εμπιστοσύνης για τη μέση τιμή μ του πληθυσμού με σ^2 γνωστό:

$$\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{a/2}, \qquad \bar{X} + \frac{\sigma}{\sqrt{n}} z_{a/2}\right)$$

Δηλαδή η μέση τιμή μ του πληθυσμού βρίσκεται μέσα σε αυτό το διάστημα

Διαστήματα εμπιστοσύνης για τη μέση τιμή του πληθυσμού 2. Άγνωστη διασπορά του πληθυσμού Μεγάλα δείγματα

Άγνωστη διασπορά: Για κάθε μ, η τυχαία μεταβλητή

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

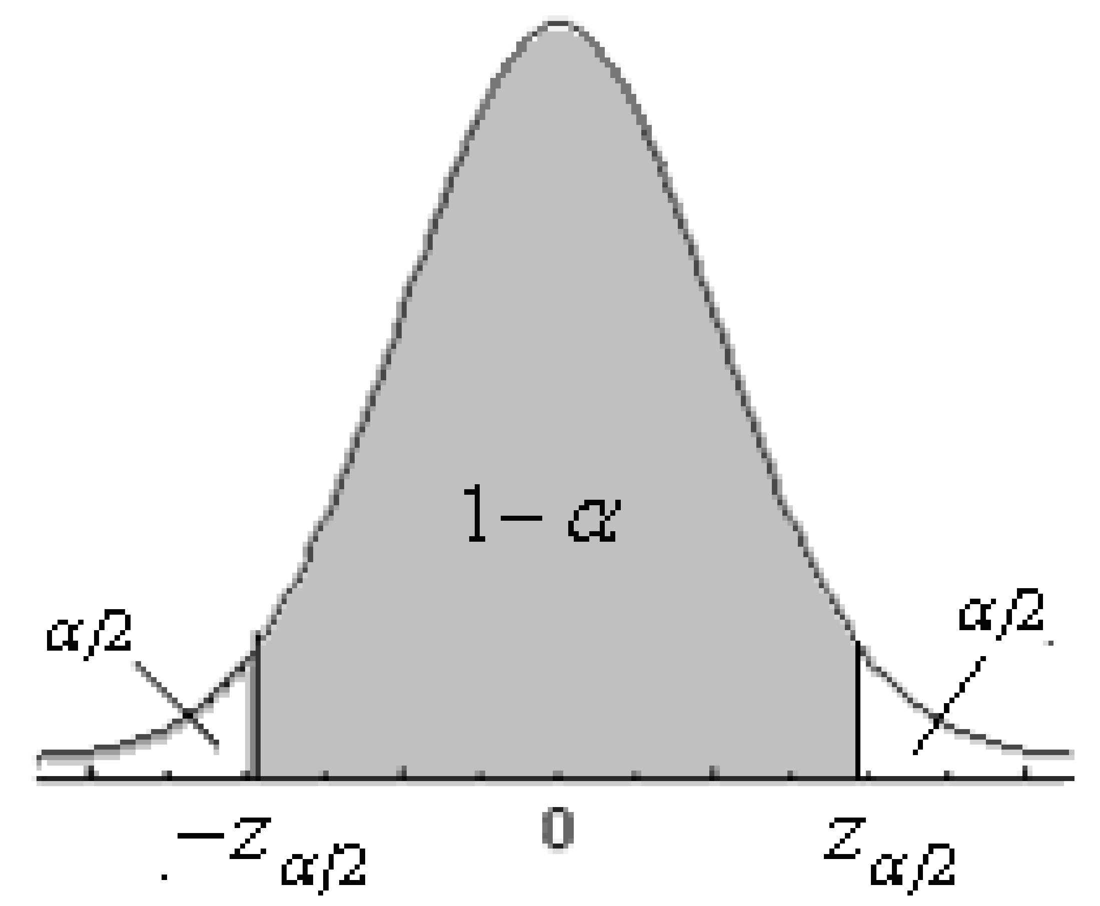
ακολουθεί την τυπική κανονική κατανομή N(0,1).

Ένα 100(1-α)% διάστημα εμπιστοσύνης για τη μέση τιμή μ του πληθυσμού με σ^2 άγνωστο:

$$\left(\bar{X} - \frac{s}{\sqrt{n}} z_{a/2}, \qquad \bar{X} + \frac{s}{\sqrt{n}} z_{a/2}\right)$$

Δηλαδή η μέση τιμή μ του πληθυσμού βρίσκεται μέσα σε αυτό το διάστημα . Η διαφορά με την προηγούμενη περίπτωση είναι ότι στο διάστημα αντικαθιστούμε την σ δειγματική με το s λόγω μεγάλου δείγματος)

Διαστήματα εμπιστοσύνης για τη μέση τιμή του πληθυσμού (κανονική κατανομή περίπτωση 1 και 2)



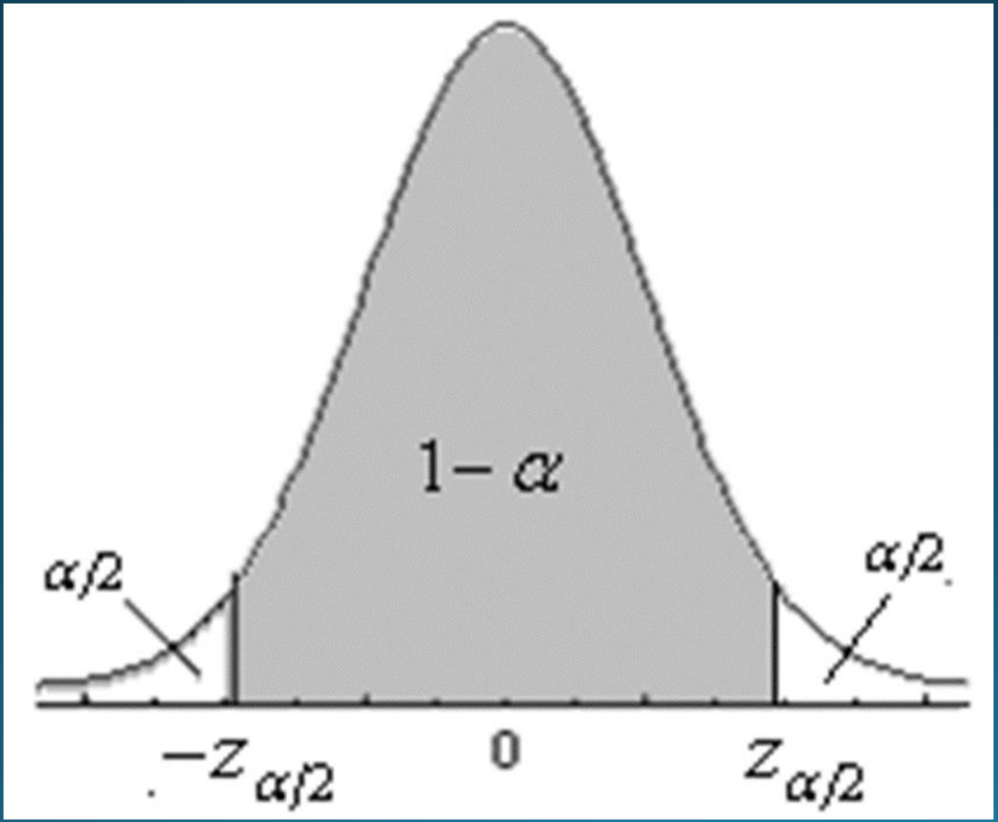
Περιοχή στην οποία βρίσκεται ένα 100(1-α)% διάστημα εμπιστοσύνης για τη μέση τιμή μ του πληθυσμού (κανονική κατανομή).

Διαστήματα εμπιστοσύνης (υπολογισμός ορίου)

Ποσοστιαία σημεία: Σημεία υπολογισμού των ορίων του διαστήματος

Άνω α-σημείο h: $P(X > h) = a$.

Στην τυπική κανονική κατανομή συμβολίζεται z_a (Εύρεση από πίνακες) ή $z_{a/2}$ λόγω συμμετρίας της κανονικής κατανομής.



Διαστήματα εμπιστοσύνης για τη μέση τιμή του πληθυσμού
3. Άγνωστη διασπορά του πληθυσμού
Μικρά δείγματα

Άγνωστή διασπορά: Για κάθε μ, η τυχαία μεταβλητή

$$t = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

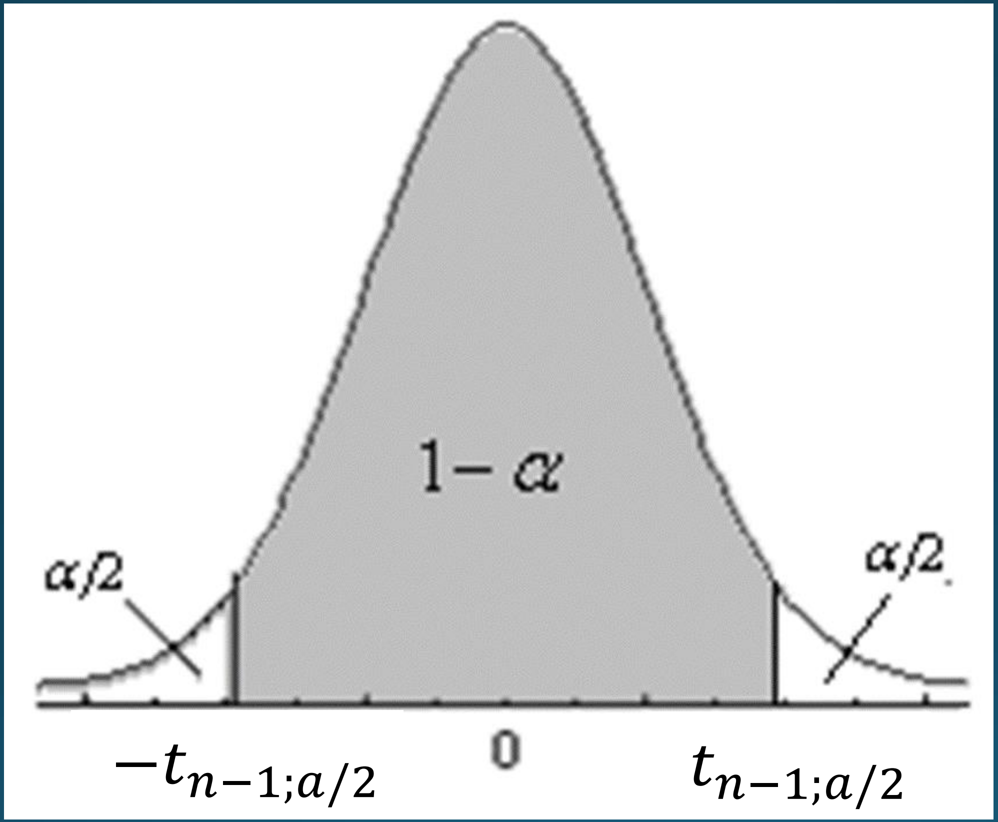
ακολουθεί την t- κανονική κατανομή (Κατανομή Student) με n-1 βαθμούς ελευθερίας.

Ένα 100(1-α)% διάστημα εμπιστοσύνης για τη μέση τιμή μ του πληθυσμού με σ^2 άγνωστο:

$$\left(\bar{X} - \frac{s}{\sqrt{n}} t_{n-1;\, a/2}, \qquad \bar{X} + \frac{s}{\sqrt{n}} t_{n-1;\, a/2}\right)$$

Δηλαδή η μέση τιμή μ του πληθυσμού βρίσκεται μέσα σε αυτό το διάστημα .

Διαστήματα εμπιστοσύνης για τη μέση τιμή



Το $t_{n-1;a/2}$ είναι η τιμή που στα δεξιά της αφήνει εμβαδόν α/2 με n-1 βαθμούς ελευθερίας

Άσκηση

Σε ένα μετεωρολογικό εργαστήριο με όργανα που μετρούν την ηλιακή ακτινοβολία στην ατμόσφαιρα, ο ερευνητής θέλει να υπολογίσει την απόδοση των οργάνων σε μέσους χρόνους μεταξύ αδυναμιών εκτέλεσης της μέτρησης (αποτυχία μέτρησης). Για να υπολογίσει αυτήν την τιμή το εργαστήριο κατέγραψε το χρόνο αποτυχιών για ένα τυχαίο δείγμα 45 αποτυχιών των οργάνων, όπου υπολογίστηκαν τα ακόλουθα δειγματικά στατιστικά:

$$\bar{x} = 1762 \text{ δευτερόλεπτα}, s = 215 \text{ δευτερόλεπτα}$$

Α. Εκτιμήστε την πραγματική μέση τιμή του χρόνου μεταξύ αποτυχιών με ένα 90% διάστημα εμπιστοσύνης.

Β. Εάν το όργανο λειτουργεί σωστά, η πραγματική μέση τιμή των αποτυχιών μέτρησης ξεπερνά τα 1700 δευτερόλεπτα. Βάσει του διαστήματος εμπιστοσύνης που εκτιμήσατε, τι μπορείτε να συμπεράνετε για το όργανο;

Γ (εξάσκηση). Εάν τα παραπάνω δειγματικά στατιστικά, συγκεντρώθηκαν από 25 αποτυχίες του οργάνου απαντήστε στα Α και Β.

Άσκηση - Λύση

- Δεν γνωρίζουμε τη διασπορά όλου του πληθυσμού (έχουμε μόνο δειγματικά στατιστικά).
- Το δείγμα μας είναι άνω των 30 (n=45).
- Αναζητείται ένα 90% διάστημα εμπιστοσύνης.

Άρα:

- Βρίσκουμε το a που είναι 1-0.90=0.10 και συνεπώς το a/2 είναι 0.10/2=0.05

Έτσι η πραγματική μέση τιμή θα ανήκει το διάστημα:

$$\left(\bar{X}-\frac{s}{\sqrt{n}}z_{a/2}, \qquad \bar{X}+\frac{s}{\sqrt{n}}z_{a/2}\right)=\left(\bar{X}-\frac{215}{\sqrt{45}}z_{0.05}, \qquad \bar{X}+\frac{215}{\sqrt{45}}z_{0.05}\right)$$

Άσκηση - Λύση

- Αντικαθιστούμε και το δειγματικό μέσο:

$$\left(1762 - \frac{215}{\sqrt{45}} z_{0.05}, \qquad 1762 + \frac{215}{\sqrt{45}} z_{0.05}\right)$$

- Μένει να βρούμε το z (πίνακας Ια: Διαιρούμε το δ.ε. δια 2 δηλαδή 0.90/2=0.45 και αντιστοιχίζουμε τη συντεταγμένη x,y για το z, δηλαδή z=1.65)

- Το $z_{0.05}$ είναι η τιμή z που αντιστοιχεί στο κομμάτι της δεξιάς ουράς της κανονικής κατανομής, εμβαδού ίσου με 0.05 και βρίσκεται ίση με 1.65.

- Αντικαθιστώντας έχουμε τελικά το 90% δ.ε. (1709 , 1814) στο οποίο βρίσκεται η πραγματική μέση τιμή του χρόνου αποτυχιών μέτρησης.

Άσκηση- Λύση β ερώτημα

Γνωστό ότι το όργανο λειτουργεί σωστά, εάν η πραγματική μέση τιμή του χρόνου των αποτυχιών μέτρησης ξεπερνά τις 1700 δευτερόλεπτα.

Επειδή η μέση τιμή μ βάσει του διαστήματος εμπιστοσύνης ξεπερνά τα 1700 δευτερόλεπτα, μπορούμε με βεβαιότητα 90% να θεωρήσουμε ότι το όργανο λειτουργεί σωστά.