



z – or t – distribution

z – ή t – κατανομή

Ζιντζαράς Ηλίας, M.Sc., Ph.D.

*Καθηγητής Βιομαθηματικών-Βιομετρίας
Εργαστήριο Βιομαθηματικών
Τμήμα Ιατρικής
Πανεπιστήμιο Θεσσαλίας*

*Institute for Clinical Research and Health Policy Studies
Tufts University School of Medicine
Boston, MA, USA*

*Θεόδωρος Μπρότσης, MSc, PhD
Εντεταλμένος Διδάσκων
(<http://biomath.med.uth.gr>)
Πανεπιστήμιο Θεσσαλίας
Email: tmprotsis@uth.gr*



z – ή t – σχέσεις

$$\bar{x} \pm z \frac{s}{\sqrt{n}}$$

$$\bar{x} \pm t \frac{s}{\sqrt{n}}$$



Παραδείγματα

Παράδειγμα 1

- Επιλέχθηκαν τυχαία 10 άτομα ηλικίας 65+ (δείγμα) από εκατομμύρια πιθανόν άτομα ηλικίας 65+ (πληθυσμός)
- Βρείτε τη μέση τιμή της πίεσης
- Αναφέρετε το αποτέλεσμα ως «Η μέση τιμή της πίεσης είναι ...»

Παράδειγμα 2

- Επιλέχθηκαν τυχαία 5 μαθητές της 3^{ης} λυκείου
- Βρείτε τη μέση τιμή του ύψους
- Αναφέρετε το αποτέλεσμα «Η μέση τιμή είναι του ύψους είναι ...»

Αυτά τα παραδείγματα φαίνονται να έχουν ένα πρόβλημα



Όρια δεδομένων στην έρευνα ή ανάλυση

- Όταν γίνεται ποσοτική έρευνα ή ανάλυση, συνήθως ενδιαφερόμαστε για ένα πληθυσμό
 - Ο επιπολασμός της κατάθλιψης μεταξύ ατόμων ηλικίας 65+
 - Το μέσο ύψος των πρωτοετών φοιτητών
- Ωστόσο λόγω μη διαθέσιμου χρόνου και υψηλού κόστους, πάντα κάνουμε χρήση δείγματος αντιπροσωπευτικού του πληθυσμού
- Το δείγμα αποτελεί όμως μία εκτίμηση ή προσέγγιση ενός μεγαλύτερου πληθυσμού από το οποίο επιλέχτηκε



Όρια δεδομένων στην έρευνα ή ανάλυση

- Όσο μικραίνει ο αριθμός του δείγματος n , τόσο πιο αβέβαιοι είμαστε ότι αντιπροσωπεύει τον ευρύτερο πληθυσμό. Ο κίνδυνος σφάλματος μεγαλώνει
- Επιπρόσθετα, συνήθως δεν γνωρίζουμε τίποτα για τον πληθυσμό· τη μέση τιμή, τη διακύμανση, ή την τυπική απόκλιση

Επομένως, χρησιμοποιούμε δείγματα από ένα πληθυσμό για τον οποίο έχουμε λίγες ή καθόλου πληροφορίες



Μέγεθος δείγματος

- Επομένως καταλαβαίνουμε, πως όσο πιο μεγάλο είναι ένα δείγμα, τόσο πιο βέβαιοι (confident) είμαστε ότι αντιπροσωπεύει τον ευρύτερο πληθυσμό
- Σ' ένα μεγάλο δείγμα, είναι πιθανό να συλλάβουμε την φυσική διακύμανση και ποικιλία των δεδομένων του ευρύτερου πληθυσμού
- Υπάρχει και ένα σημείο όπου αυξάνοντας περισσότερο το δείγμα δεν προσφέρει επιπλέον στατιστικά οφέλη

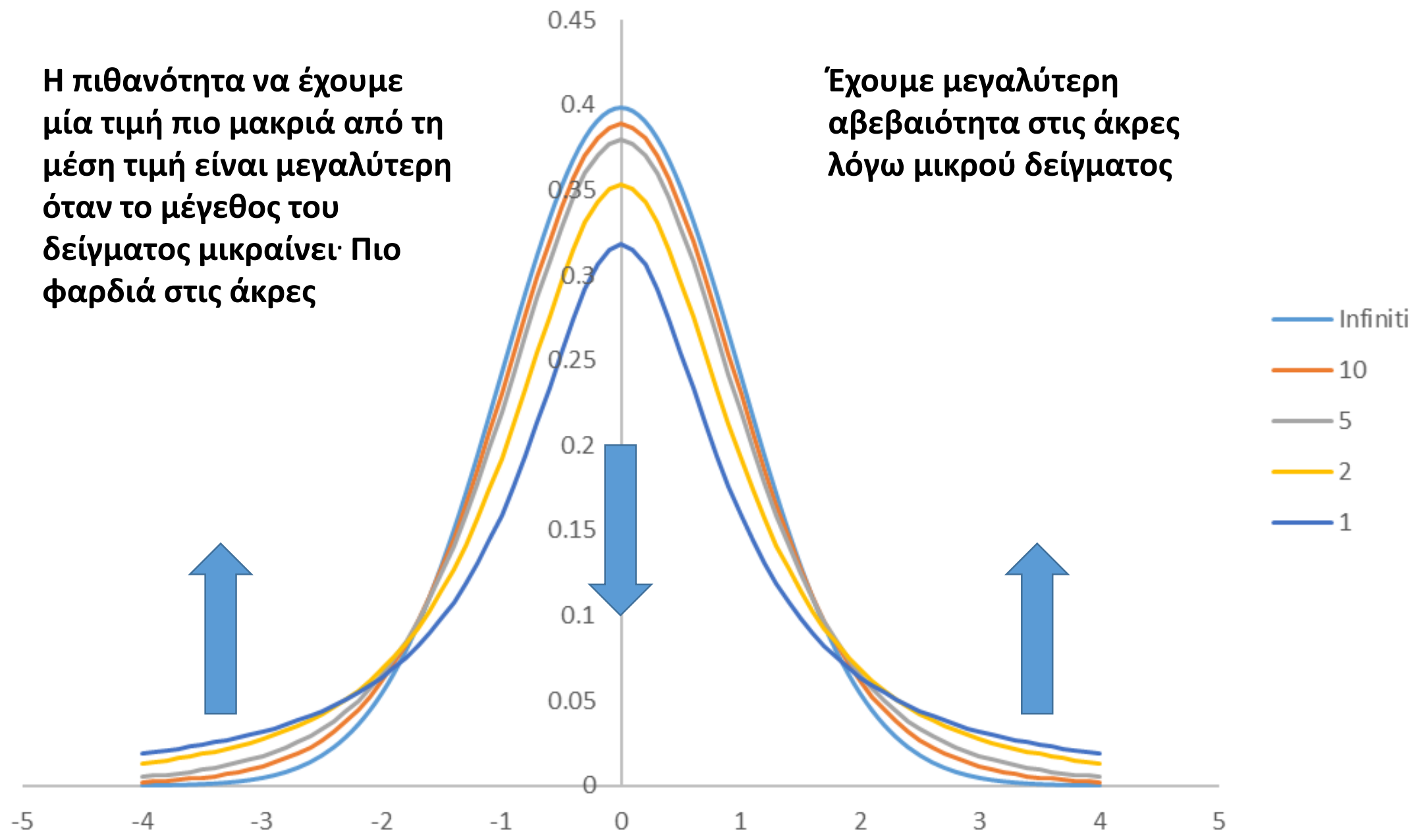


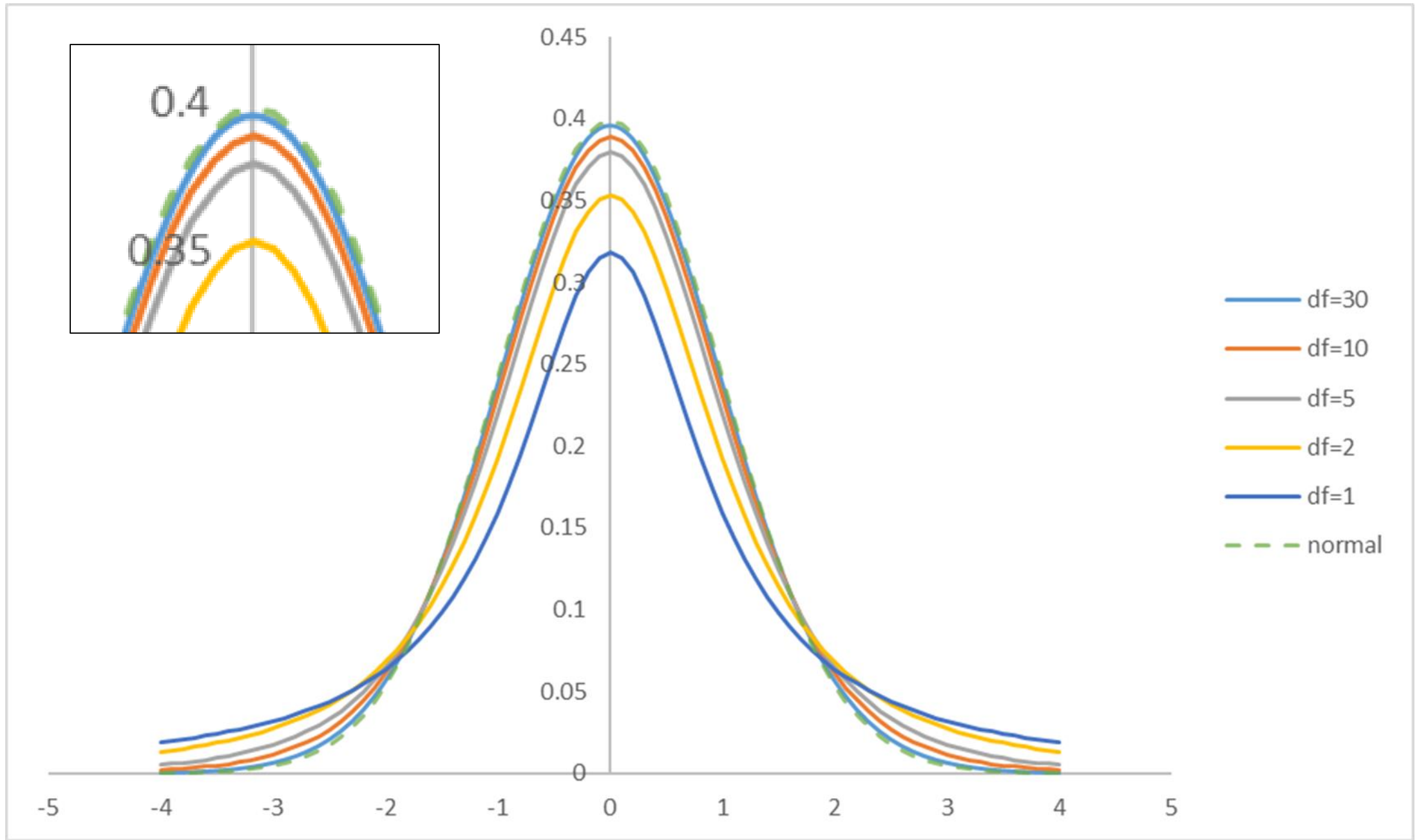
t – Κατανομή(ες)

- Όταν το μέγεθος του δείγματος είναι $n \leq 30$ ή όταν δεν γνωρίζουμε την διακύμανση/τυπική απόκλιση του πληθυσμού τότε χρησιμοποιούμε την t – κατανομή έναντι της z – κατανομής (τυπική κανονική κατανομή)
- Τι μας επιτρέπει αυτό;
 - Η t – κατανομή μας επιτρέπει να κάνουμε χρήση μικρών δειγμάτων· $n \leq 30$
 - Κάνοντας όμως αυτό θυσιάζουμε κάποια βεβαιότητα στους υπολογισμούς μας·
Περιθώριο σφάλματος
 - Λαμβάνει υπόψη το μέγεθος του δείγματος χρησιμοποιώντας $n - 1$ βαθμούς ελευθερίας· η t – κατανομή είναι διαφορετική για κάθε μέγεθος δείγματος
 - Η καμπύλη καμπάνας «τραβιέται» στη μέση προς το κάτω και «ψηλώνει» στις άκρες· όσο πιο μικρό το δείγμα τόσο πιο πολύ «τραβιέται» προς τα κάτω στη μέση και «ψηλώνει» στις άκρες
 - Όσο το $n > 30$ και ειδικά όσο το $n \geq 100$, η t – κατανομή και η z – κατανομή γίνονται όλο και πιο δυσδιάκριτες

Η πιθανότητα να έχουμε μία τιμή πιο μακριά από τη μέση τιμή είναι μεγαλύτερη όταν το μέγεθος του δείγματος μικραίνει. Πιο φαρδιά στις άκρες

Έχουμε μεγαλύτερη αβεβαιότητα στις άκρες λόγω μικρού δείγματος





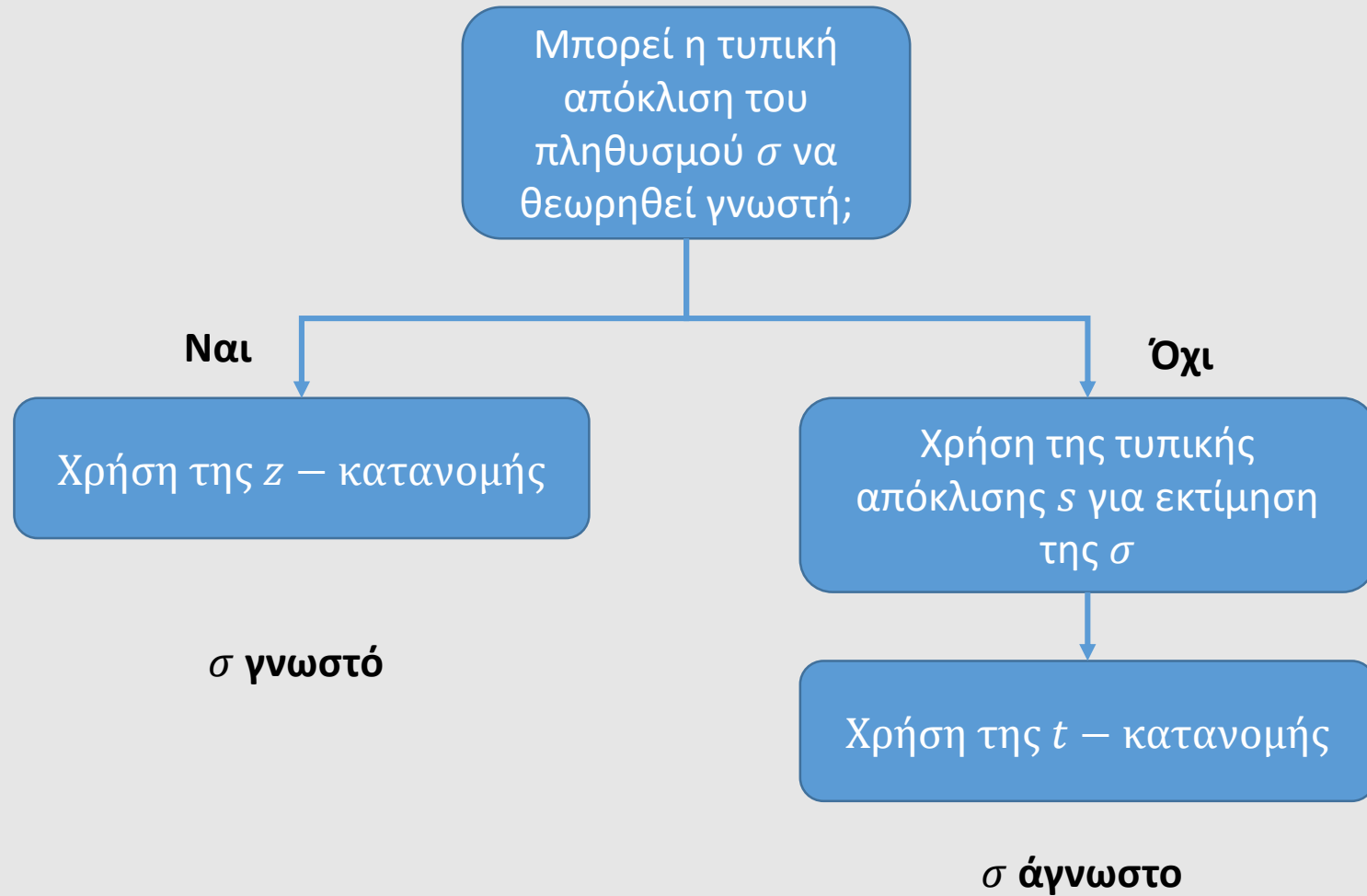


Βαθμοί ελευθερίας

- Οι βαθμοί ελευθερίας δεν είναι τίποτα άλλο από μία προσαρμογή στο μέγεθος του δείγματος ($n - 1$) ή σε άλλες περιπτώσεις ($n - 2$) ή περισσότερο
- Συνδέεται με την ιδέα ότι εκτιμούμε κάτι για τον ευρύτερο πληθυσμό· τις περισσότερες φορές τη διακύμανση/τυπική απόκλιση του πληθυσμού
- Μας δίνει ελαφρώς ένα μεγαλύτερο περιθώριο σφάλματος στην εκτίμηση
- Αποτελεί μία στατιστική προσαρμογή



Δέντρο απόφασης





Δέντρο απόφασης

