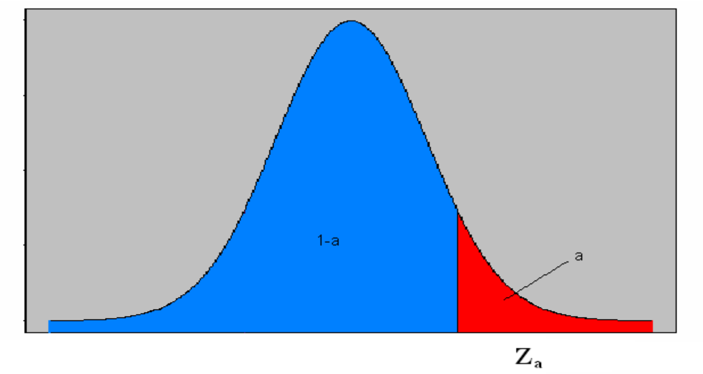


Βιοστατιστική

Έλεγχοι Υποθέσεων & Διαστήματα Εμπιστοσύνης

Αριστοτέλης Παπαγεωργίου, Τμ. ΜΒΓ ΔΠΘ, apapage@mbg.duth.gr

Πηγές υλικού

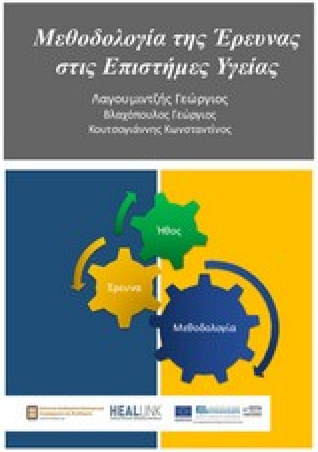
- Διαφάνειες και ασκήσεις του Dr Theophanis Tsandilas (National Institute for Research in Digital Science and Technology, INRIA, Γαλλία)
- Διαφάνειες και ασκήσεις του καθηγητή Σπύρου Γαλατσίδα (Τμ. Δασολογίας & Διαχείρισης Περιβάλλοντος & Φυσικών Πόρων, ΔΠΘ)

R tutorials

- There is a large collection of online tutorials for learning R. The following list may not be representative. It will be updated during the course:
  - An Introduction to R
  - Using R for Introductory Statistics
  - R Tutorial
  - Learn R Programming

Εισαγωγή στον προγραμματισμό και στη στατιστική ανάλυση με R

Author(s) :*Ntzoufras, Ioannis; Karlis, Dimitrios*
Type : Undergraduate textbook



Μεθοδολογία της έρευνας στις επιστήμες υγείας

Author(s) :*Lagoumintzis, Georgios; Vlachopoulos, Georgios; Koutsogiannis, Konstantinos*
Type : Undergraduate textbook

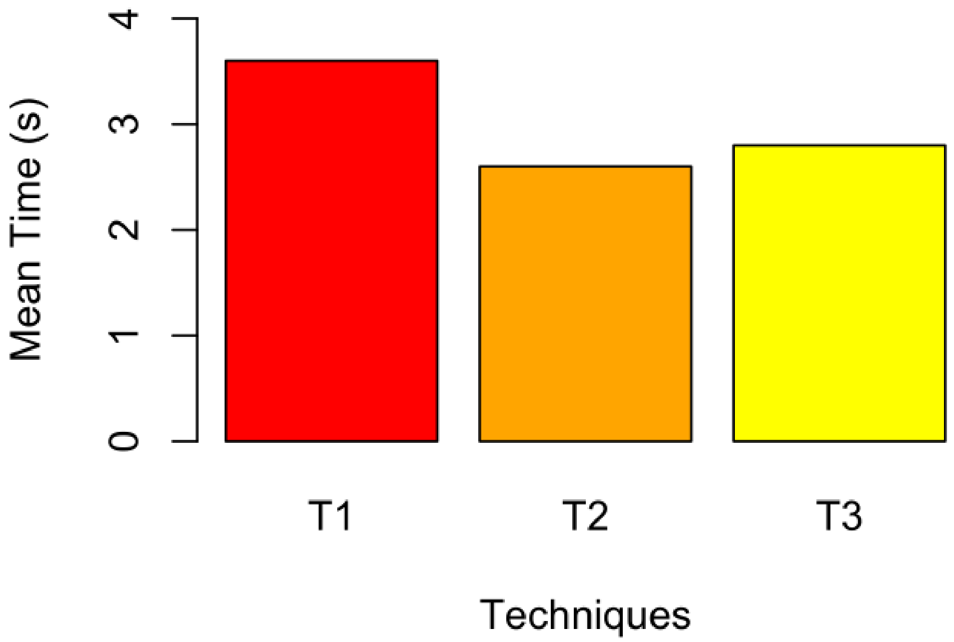
- https://repository.kallipos.gr/

Στατιστικό συμπέρασμα

- Η διαδικασία της εξαγωγής των παραμέτρων μιας υποκείμενης κατανομής πιθανοτήτων από ένα δείγμα
- Τέσσερις κατηγορίες:
  - Σημειακές εκτιμήσεις
  - Εκτιμήσεις διαστήματος
  - Έλεγχος υποθέσεων
  - Πρόβλεψη

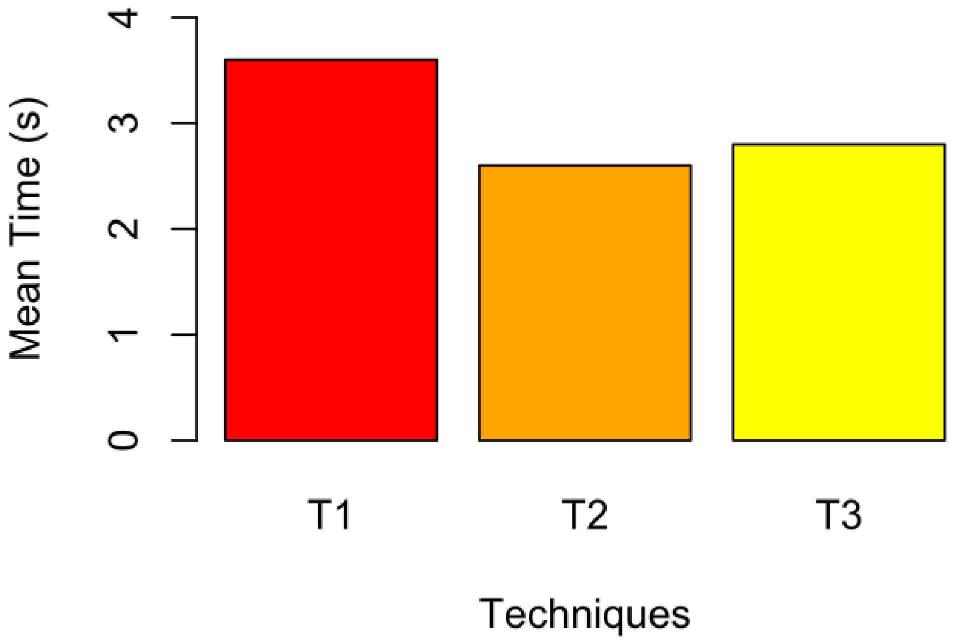
Σημειακές εκτιμήσεις

Τι μας δείχνει αυτό το γράφημα;

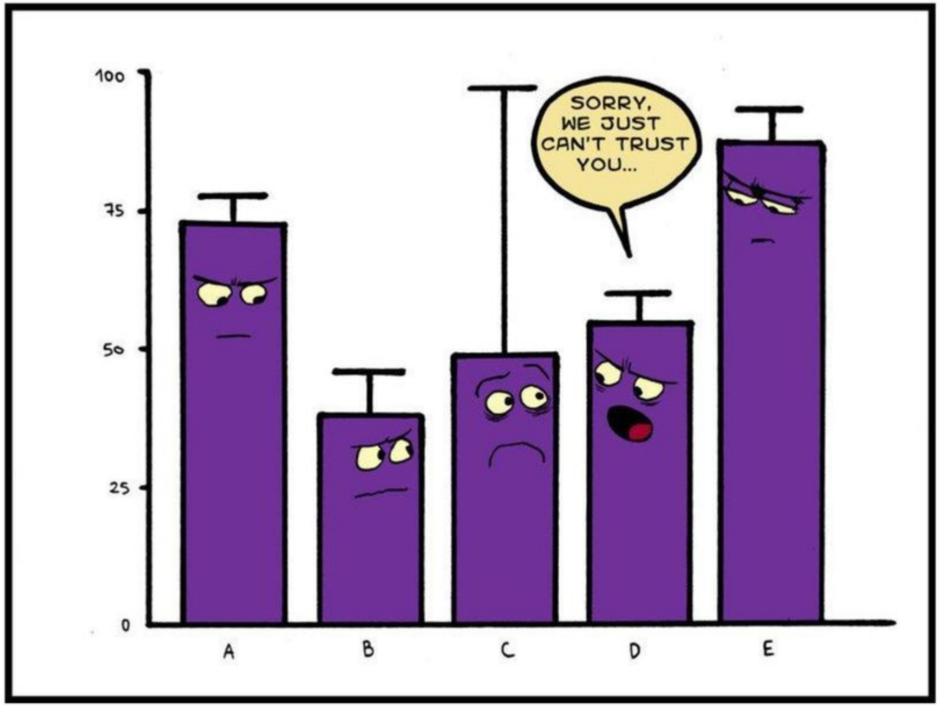


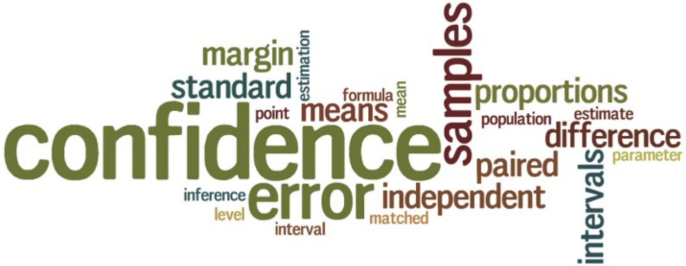
Σημειακές εκτιμήσεις

- Μια σημειακή εκτίμηση είναι η “καλύτερη εικασία” που κάνουμε σχετικά με την παράμετρο του πληθυσμού
- Τα περιγραφικά στατιστικά (μέσος, διάμεσος) είναι σημειακές εκτιμήσεις
- Δεν μας παρέχουν καμία πληροφορία για την αβεβαιότητα ή την ποιότητα της εκτίμησης



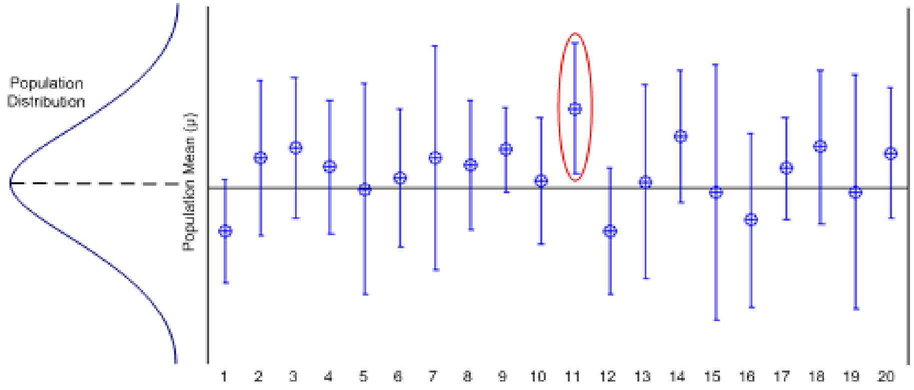
Εκτίμηση διαστήματος





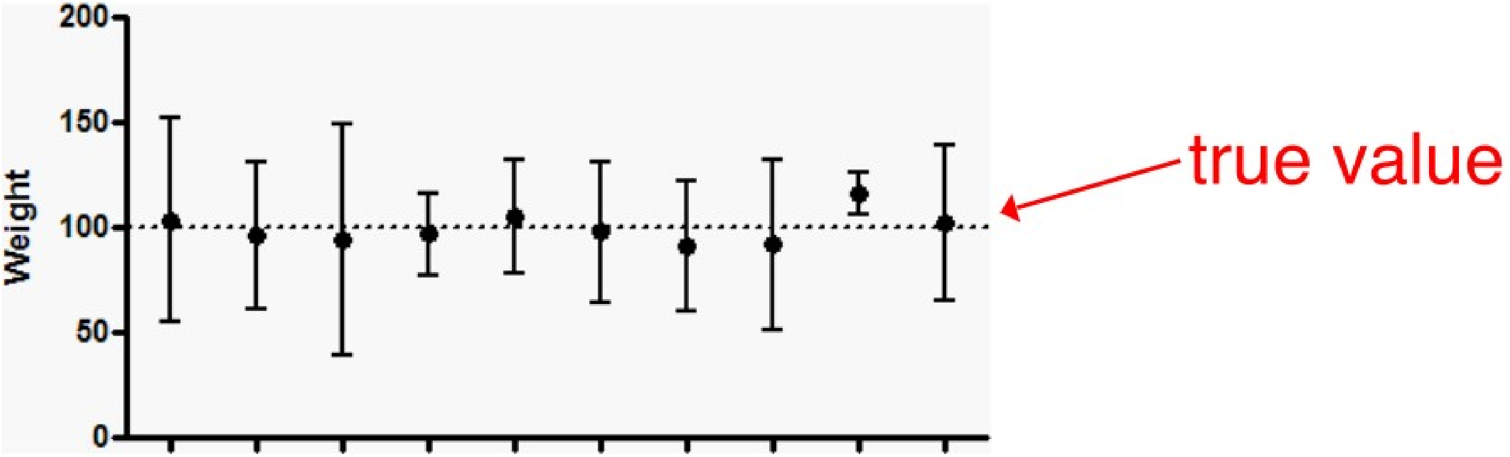
Εκτίμηση διαστήματος

- Μια εκτίμηση διαστήματος δεν παρέχει μία συγκεκριμένη τιμή, αλλά ένα εύρος τιμών που μπορεί εύλογα να πάρει μια παράμετρος
  - Η πιο κοινή μέθοδος είναι το **διάστημα εμπιστοσύνης** (confidence interval – CI)



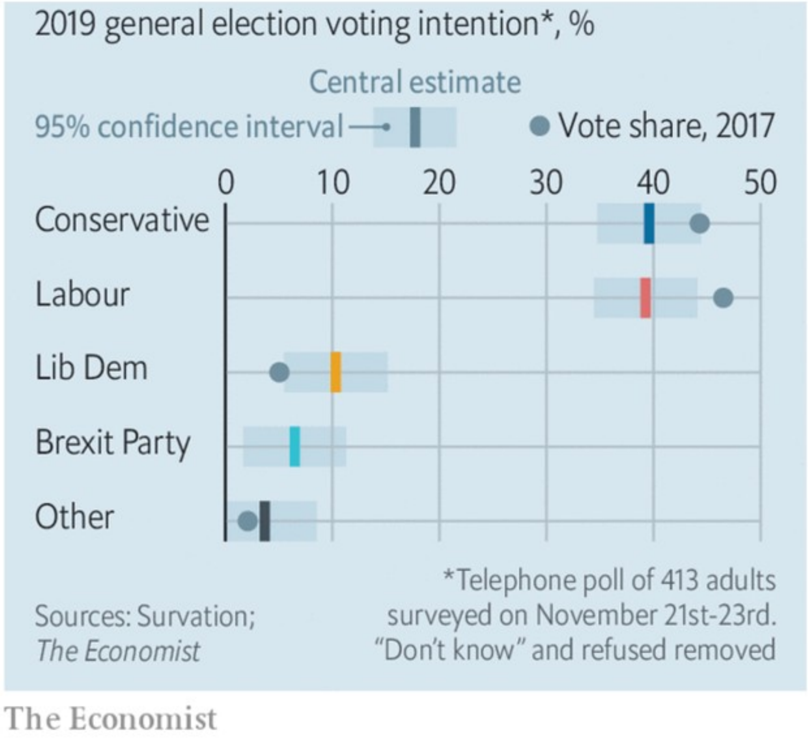
Διάστημα εμπιστοσύνης (CI)

- Προσδιορίζει ένα εύρος τιμών, που αναμένεται να περιέχει την πραγματική τιμή της παραμέτρου του πληθυσμού (ή και όχι)
- Συνδέεται με ένα **επίπεδο εμπιστοσύνης**, συνήθως ένα ποσοστό (π.χ. 95% CI ή 99% CI)



Διάστημα εμπιστοσύνης (CI)

What do 95% confidence intervals represent here?



A. 95% των ψήφων των Conservatives είναι ανάμεσα στο 35% και το 45%

B. Είμαστε βέβαιοι κατά 95% ότι η πραγματική τιμή του εκλογικού σώματος για τους Conservatives θα είναι μέσα στο διάστημα εμπιστοσύνης

Ποιο από τα δύο είναι αληθές;

Επίπεδο εμπιστοσύνης

- Ένα διάστημα εμπιστοσύνης (CI) 100% θα περιλαμβάνει όλο το εύρος των πιθανών τιμών
- Ένα CI 0% είναι ένα σημείο
- Το CI = 95% είναι η πιο συνηθισμένη επιλογή (παραδοσιακά…)

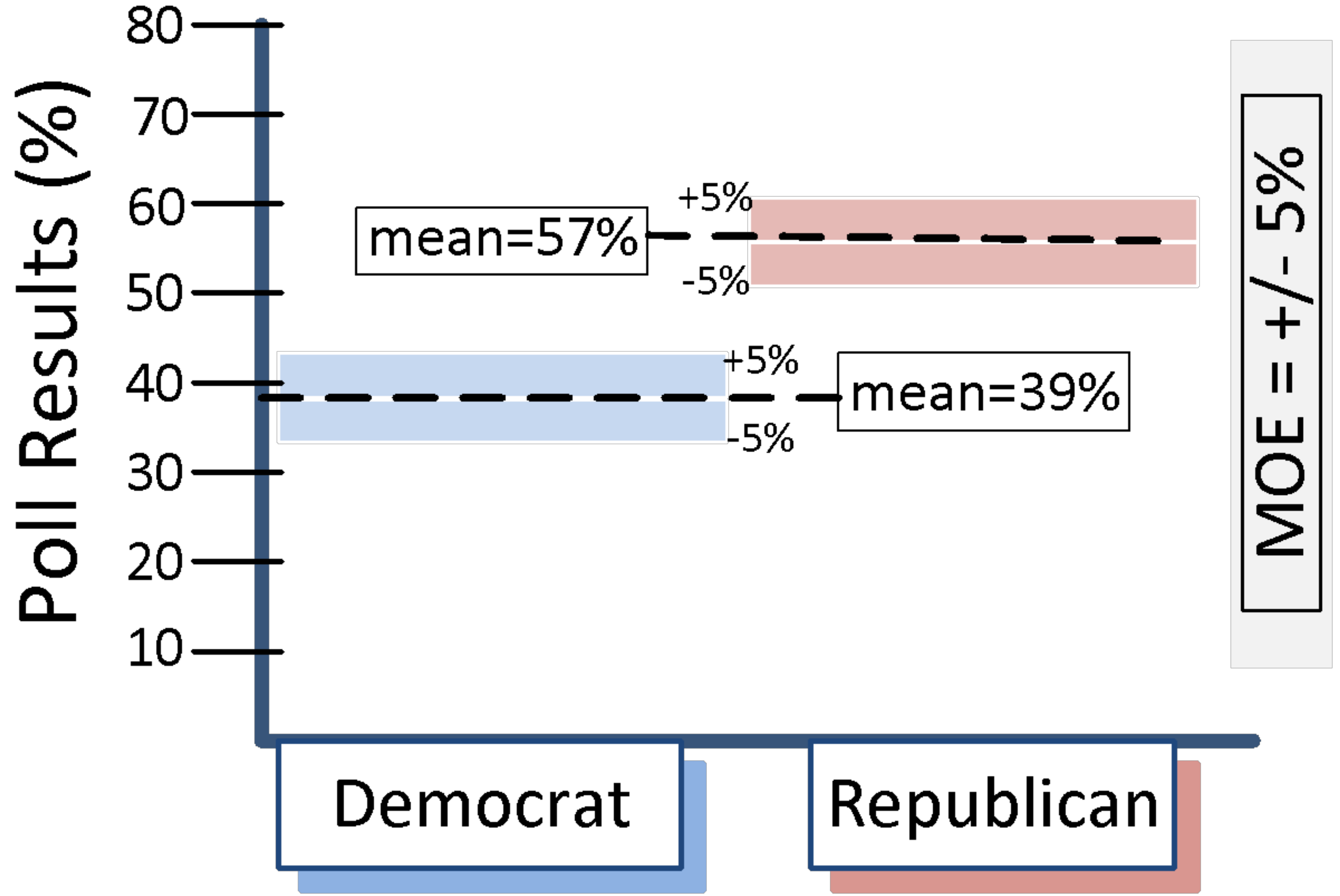
- Το επίπεδο εμπιστοσύνης ενός διαστήματος εμπιστοσύνης (CI) το συμβολίζουμε με C
  - π.χ. $C = 95\%$

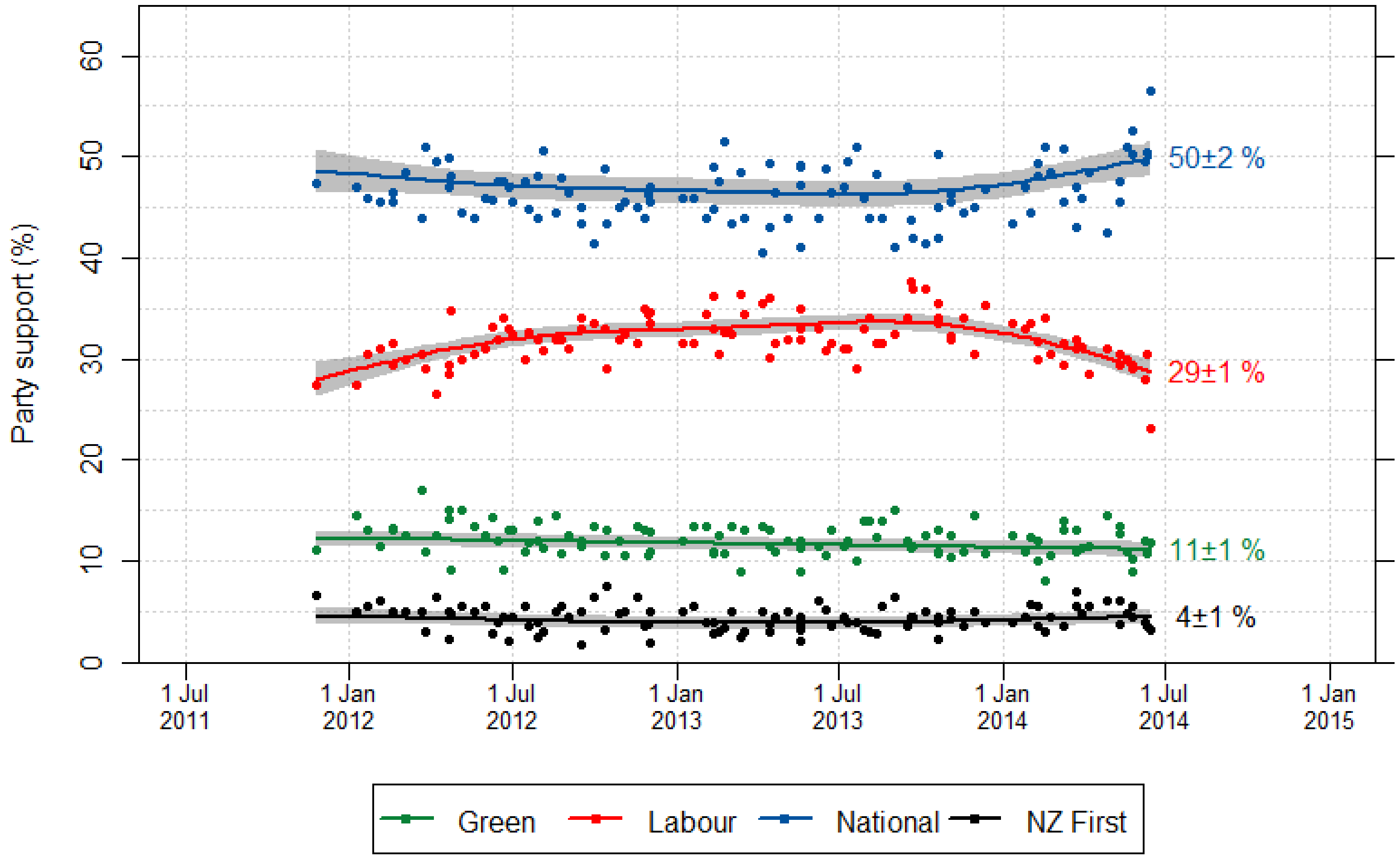
Επίπεδο alpha

- Για το επίπεδο εμπιστοσύνης *C* ενός διαστήματος εμπιστοσύνης ισχύει
  - $C = 100\,(1 - \alpha)$
- Το ***α*** (ή **alpha**) αντιπροσωπεύει την πιθανότητα ότι το διάστημα εμπιστοσύνης θα αποτύχει
  - Δηλαδή ότι το διάστημα εμπιστοσύνης σε *α* ποσοστό θα αποτύχει να συμπεριλάβει την πραγματική τιμή του πληθυσμού
- Για $C = 95\%$, τότε $\alpha = 0.05$

Δομή διαστήματος εμπιστοσύνης

- Ένα διάστημα εμπιστοσύνης ορίζεται από δύο όρια, το **άνω** και το **κάτω** όριο
  - Μπορεί να είναι συμμετρικό, όταν η σημειακή εκτίμηση βρίσκεται στο κέντρο του CI
  - Είναι ασύμμετρο όταν η σημειακή εκτίμηση δεν βρίσκεται στο κέντρο του CI
- Τα συμμετρικά CI μπορούν να περιγραφούν από τη σημειακή εκτίμηση και ± το μισό του εύρους του, π.χ. 165 ± 6 cm (το διάστημα εμπιστοσύνης για το ύψος μιας ομάδας ανθρώπων)
  - Το μισό εύρος του διαστήματος εμπιστοσύνης είναι γνωστό και ως **περιθώριο σφάλματος** (Margin of Error – MOE)


Poll Results (%)
80
70
60
50
40
30
20
10
mean=57%
+5%
-5%
+5%
-5%
mean=39%
MOE = +/- 5%
Democrat
Republican



Πώς γίνεται να έχουμε διαφορετικά MOE στο ίδιο δείγμα;

Τι καθορίζει το εύρος ενός CI

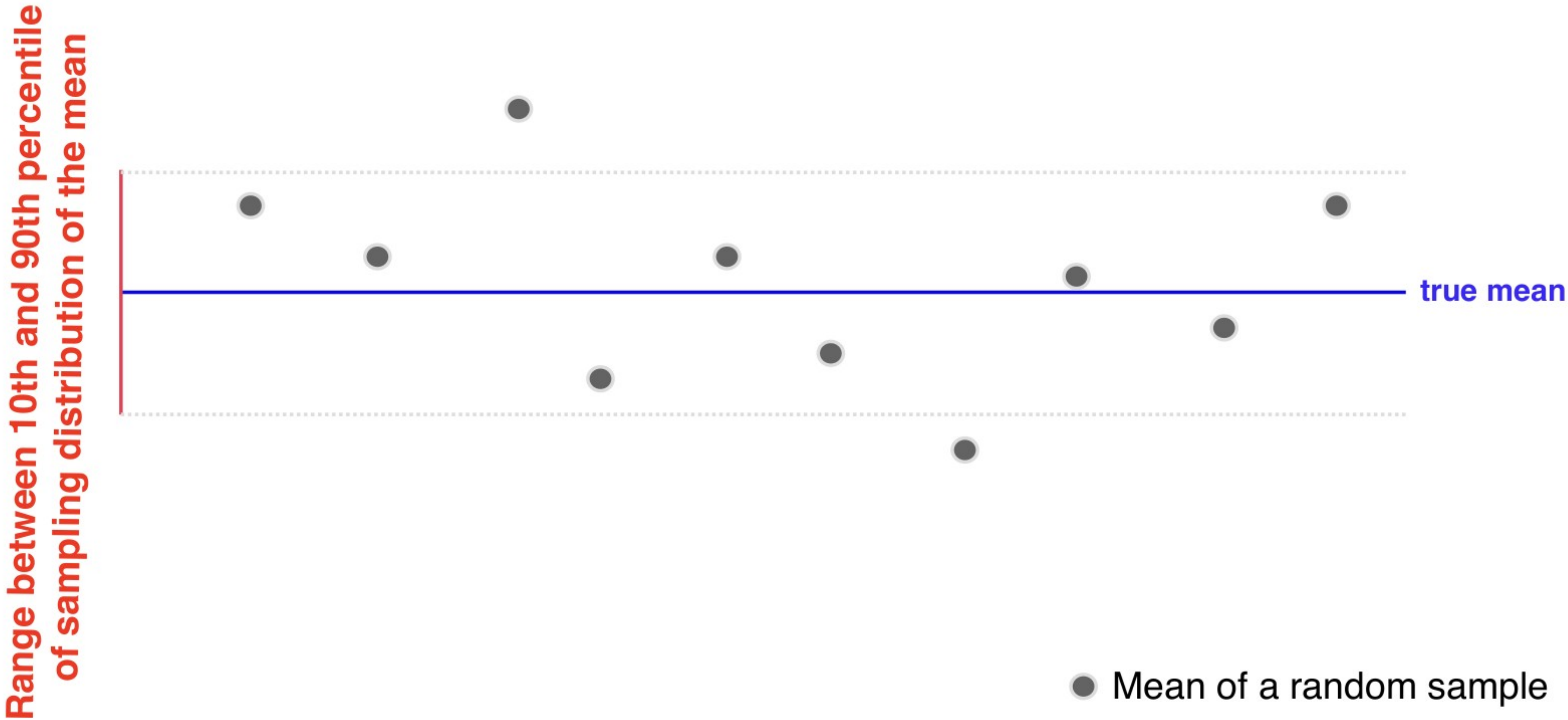
- Το εύρος ενός διαστήματος εμπιστοσύνης εξαρτάται από το **επίπεδο εμπιστοσύνης** που εμείς ορίζουμε
  - Ένα CI 99% είναι ευρύτερο από ένα 95%
- Επίσης εξαρτάται από το **μέγεθος του δείγματος**
  - Τα μικρά δείγματα συνήθως δίνουν ευρύτερο CI
- Η **διακύμανση** του πληθυσμού για μια μεταβλητή
  - Όσο πιο μεγάλη η διακύμανση, τόσο πιο ευρύ το CI
- **Σχεδιασμός** της έρευνας
  - Οτιδήποτε μειώνει την αντιπροσωπευτικότητα του δείγματος απέναντι στον πληθυσμό, αυξάνει το εύρος του CI

Κατανομές δειγματοληψίας και CI

- Για να προσδιορίσουμε το εύρος ενός διαστήματος εμπιστοσύνης (*C*%), πρέπει πρώτα να εξετάσουμε την **κατανομή** της **δειγματοληψίας** του στατιστικού
  - Η κατανομή της δειγματοληψίας μας δίνει την κατανομή των πιθανοτήτων όλων των τιμών που μπορεί να έχει ένα στατιστικό
  - Σκοπός μας είναι να ορίσουμε το *C*% αυτών των πιθανών τιμών

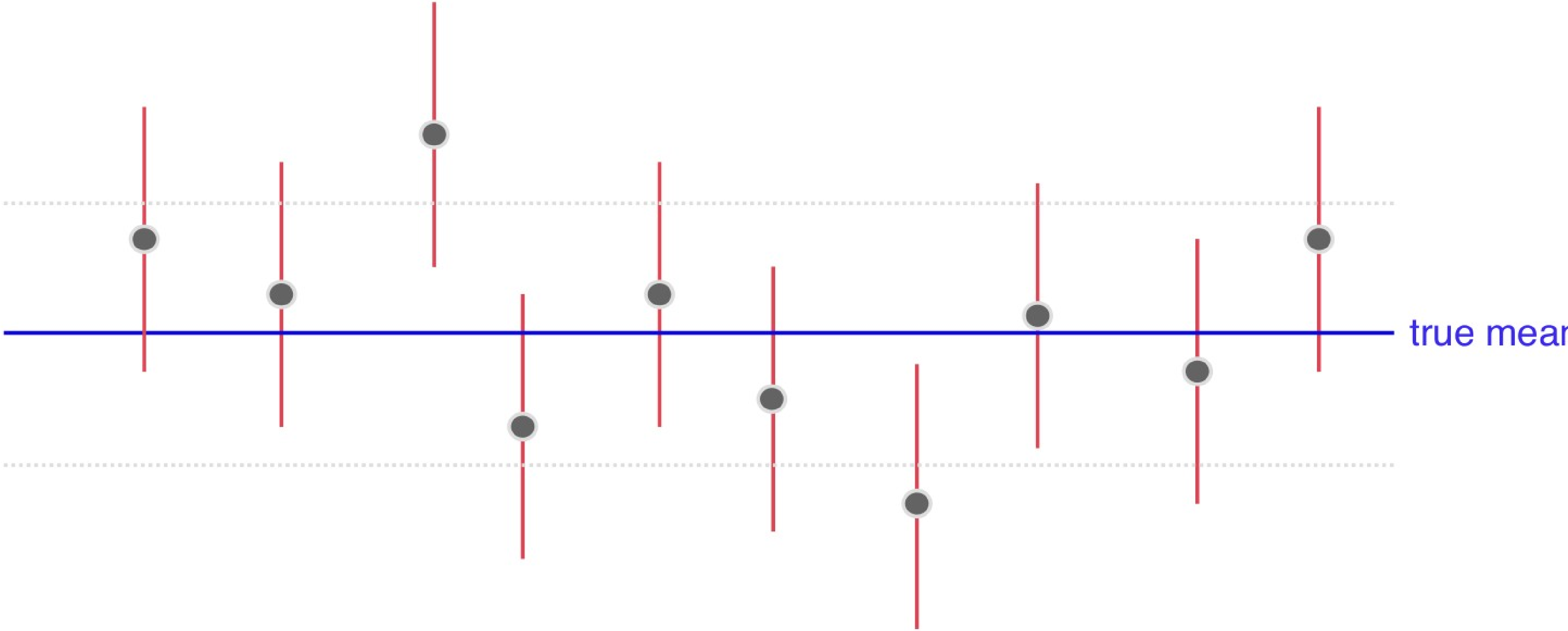
Παράδειγμα

- Για μία κατανομή δειγματοληψίας ενός πληθυσμού, παίρνουμε πολλά δείγματα μεγέθους *n* = 20
  - Εξετάζουμε το εύρος τιμών ανάμεσα στο 10% και το 90& και υπολογίζουμε τον μέσο για κάθε δείγμα
- Ερώτηση 1
  - Πόσες τιμές της κατανομής περιέχονται στο εύρος αυτό;
- Ερώτηση 2
  - Κατά μέσο όρο, πόσοι δειγματικοί μέσοι πέφτουν μέσα στο εύρος αυτό;


Range between 10th and 90th percentile of sampling distribution of the mean
true mean
Mean of a random sample

Παράδειγμα

- Ερώτηση 3
  - Κατά μέσο όρο, πόσα διαστήματα εμπιστοσύνης 80% περιέχουν τον πραγματικό μέσο του πληθυσμού;


true mear

Τυπικό σφάλμα (SE) ενός στατιστικού

- Υπάρχει γνωστή σχέση ανάμεσα στην τυπική απόκλιση (SD) μιας κατανομής δειγματοληψίας και του διαστήματος εμπιστοσύνης (CI)
- **Τυπικό σφάλμα** (Standard Error of Mean – SEM) ενός στατιστικού (π.χ. μέσου)
  - Η τυπική απόκλιση μιας κατανομής δειγματοληψίας

Υπολογισμός του SEM

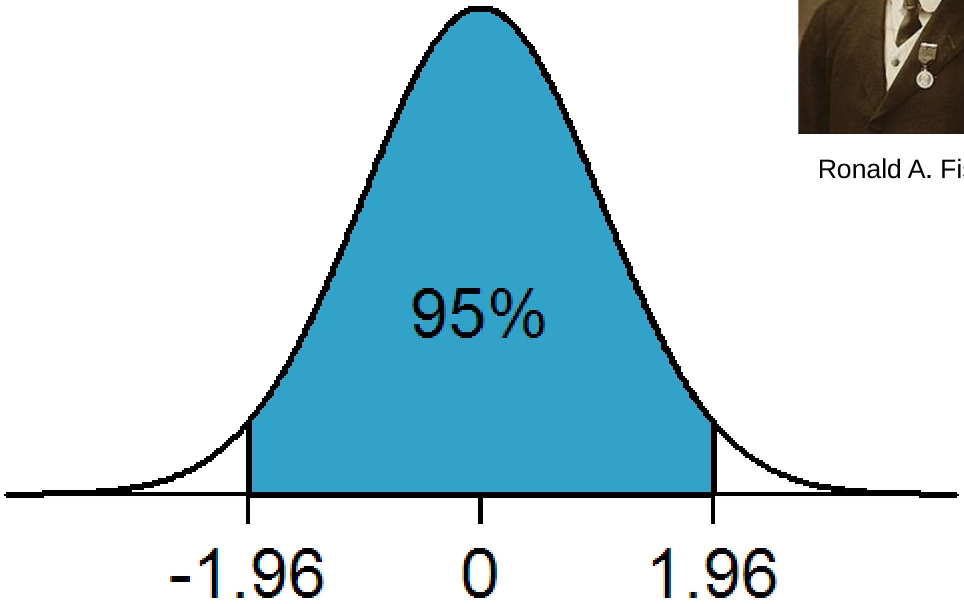
- Το τυπικό σφάλμα του μέσου (SEM) που συμβολίζεται με $\sigma_{\hat{\mu}}$ υπολογίζεται από την τυπική απόκλιση του πληθυσμού και από το μέγεθος του δείγματος

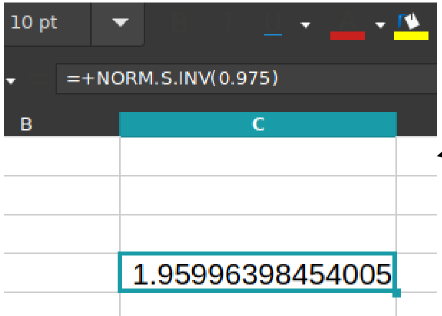
$$\sigma_{\hat{\mu}} = \frac{\sigma}{\sqrt{n}}$$

Κατανομή δειγματοληψίας

Ronald A. Fisher

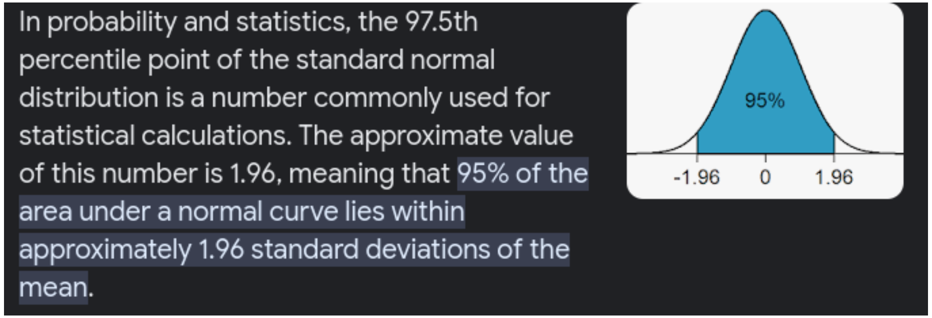
- Μια κατανομή δειγματοληψίας του μέσου είναι συνήθως κανονική
- **Πώς μπορούμε να εξάγουμε το διάστημα εμπιστοσύνης για τον μέσο αν γνωρίζουμε το τυπικό σφάλμα;**



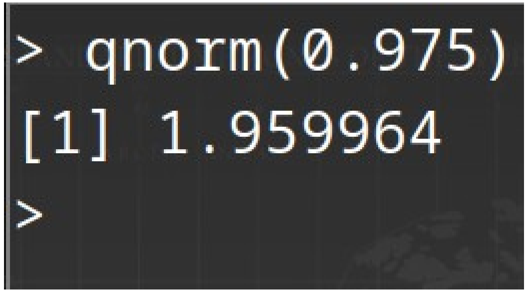


Ο φίλος μας το libreoffice calc (και το excel)

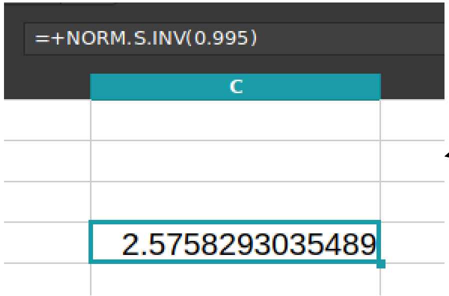
Ο φίλος μας το google



Πώς μπορούμε να βρίσκουμε το σημείο πάνω στον άξονα των x που διαχωρίζει το κεντρικό 95% των τιμών από τα δύο άκρα;



Ο φίλος μας το R

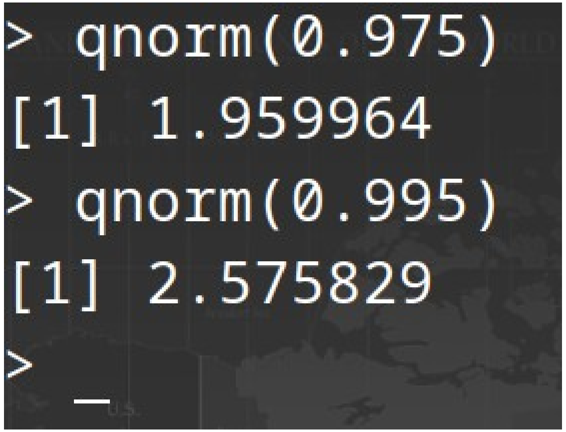


Ο φίλος μας το libreoffice calc (και το excel)

Ο φίλος μας το google

Δηλαδή μπορούμε να κάνουμε το ίδιο και για το 99%;

Ναι, τα εξωτερικά άκρα θα περιέχουν το (1-(0.99/2)) προς κάθε κατεύθυνση, άρα **0.995** και **0.005**



Ο φίλος μας το R

Υπολογισμός συμμετρικού διαστήματος εμπιστοσύνης (CI)

Συνήθως είναι **1,96**

$$CI = \hat{\mu} \pm \phi_{a/2} \times \sigma_{\hat{\mu}}$$

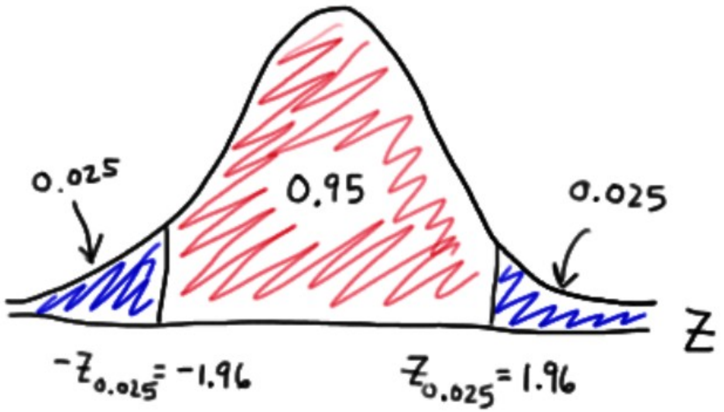
$\phi_{a/2}$ → **το όριο της καμπύλης του ποσοστού *α*/2 x 100 για συμμετρική κατανομή** (π.χ. για *α*=0.05 είναι 1,96)

Υπολογισμός συμμετρικού διαστήματος εμπιστοσύνης (CI)

Συνήθως είναι **1,96**

$$CI = \hat{\mu} \pm z_{a/2} \times \sigma_{\hat{\mu}}$$

a/2 x 100 percentile of the standard normal distribution



standard normal (z) distribution

Σε μια κανονική κατανομή **z**, ο μέσος είναι 0 και η τυπική απόκλιση είναι 1

Για 95% CI παίρνουμε το διάστημα ανάμεσα στα όρια του 2,5% και από τις δύο πλευρές της καμπύλης, δηλαδή +/- 1,96

Άσκηση

- Ένας ερευνητής παίρνει ένα τυχαίο δείγμα 30 ανθρώπων, τους κρατά ξύπνιους για 24 ώρες και μετά τους υποβάλλει σε τεστ IQ. Ο ερευνητής βρίσκει ένα μέσο σκορ IQ ίσο με M = 94,6.
- Γνωρίζει από τη βιβλιογραφία ότι η τυπική απόκλιση του IQ στον ανθρώπινο πληθυσμό είναι *SD* = 15. Υπολογίστε το 95% CI για τον μέσο

Άσκηση

- Ένας ερευνητής παίρνει ένα τυχαίο δείγμα 30 ανθρώπων, τους κρατά ξύπνιους για 24 ώρες και μετά τους υποβάλλει σε τεστ IQ. Ο ερευνητής βρίσκει ένα μέσο σκορ IQ ίσο με M = 94,6.
- Γνωρίζει από τη βιβλιογραφία ότι η τυπική απόκλιση του IQ στον ανθρώπινο πληθυσμό είναι *SD* = 15. Υπολογίστε το 95% CI για τον μέσο

$$CI = \hat{\mu} \pm z_{a/2} \times \sigma_{\hat{\mu}} \implies CI = 94.6 \pm 5.4$$

$$94.6 \qquad \pm 1.96 \qquad \frac{15}{\sqrt{30}} = 2.74$$

Έχουμε πρόβλημα...

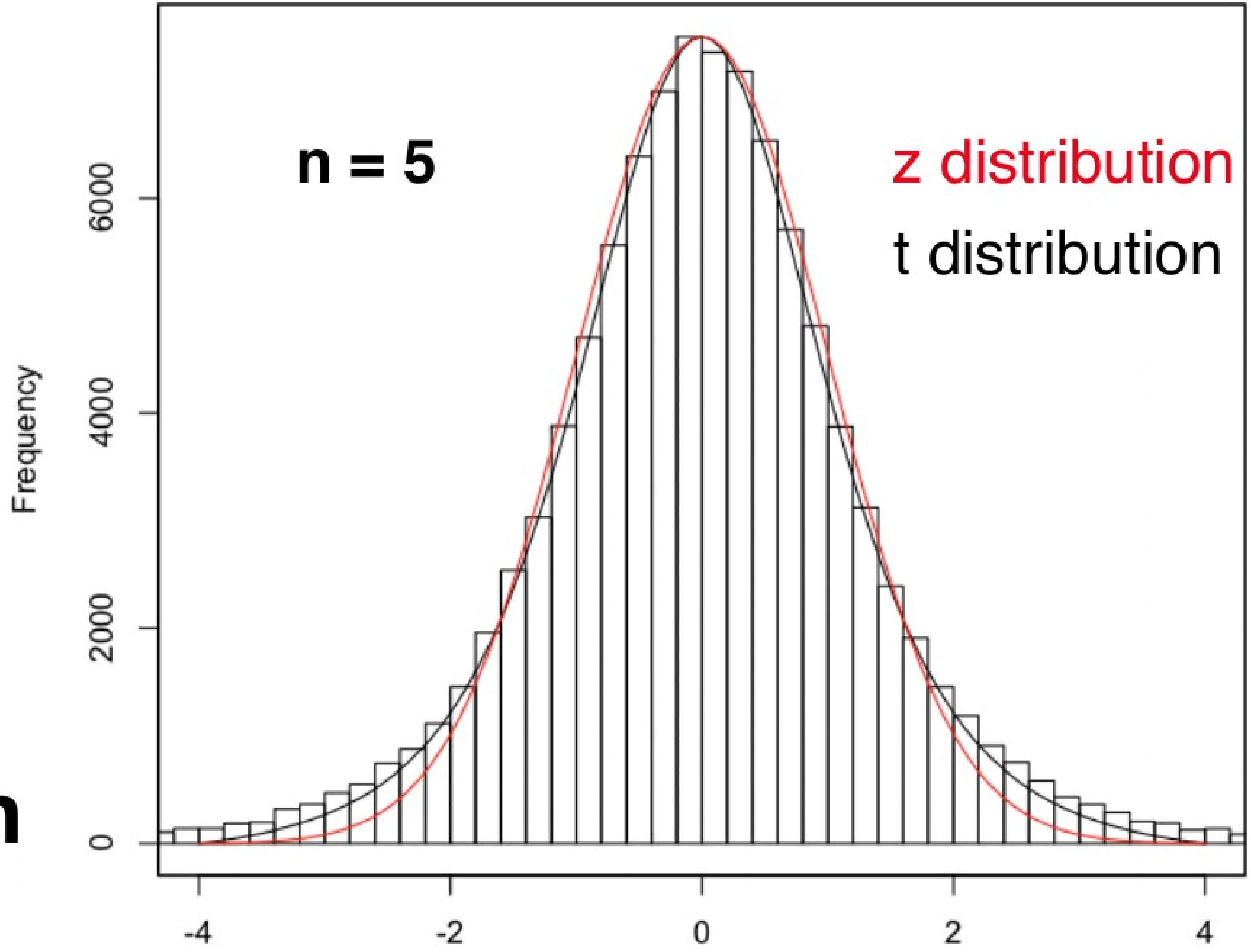
- Συνήθως δεν γνωρίζουμε την τυπική απόκλιση του πληθυσμού
- Αναγκαστικά χρησιμοποιούμε την τυπική απόκλιση του δείγματος για τον υπολογισμό του SEM

$$\hat{\sigma}_{\hat{\mu}} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{\sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{\mu})^2}{n-1}}}{\sqrt{n}}$$

unbiased estimate
of population stand. deviation

Κατανομή t (Student)

- Όταν χρησιμοποιούμε, αντί για την τυπική απόκλιση του πληθυσμού, αυτή του δείγματος, τότε ο **μέσος όρος** των δειγμάτων από τον πληθυσμό αυτόν δεν ακολουθεί την κανονική κατανομή ***z***, αλλά την κατανομή ***t*** με $v=n-1$ βαθμούς ελευθερίας
  - Δημοσιεύτηκε από τον William Gosset (1908) με το ψευδώνυμο “Student”

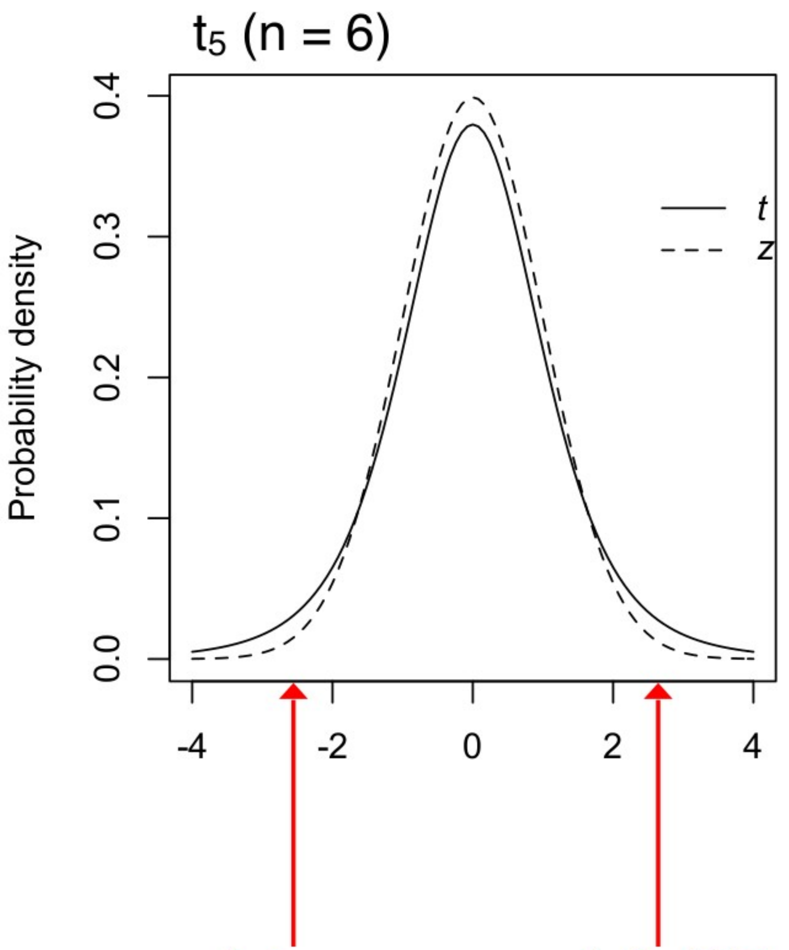


It is a **t distribution**

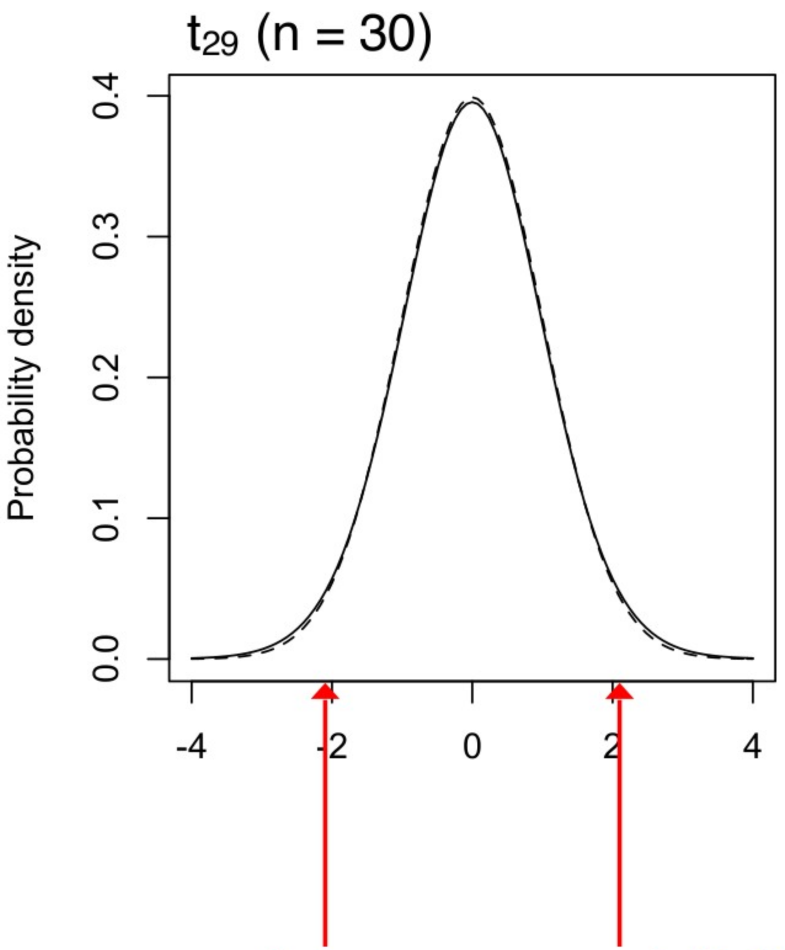
Διαστήματα εμπιστοσύνης για την t κατανομή

- Το σχήμα της καμπύλης της κατανομής εξαρτάται από τους βαθμούς ελευθερίας της **t** κατανομής: ***ν* = *n* – 1**
- Έτσι, τα όρια της καμπύλης για *α*=0,05 εξαρτώνται και αυτά από τους βαθμούς ελευθερίας (και δεν είναι πλέον +/- 1,96)
- Το διάστημα εμπιστοσύνης για την **t** κατανομή υπολογίζεται με τον ίδιο τρόπο, όπως και για την **z** κατανομή, αλλά χρησιμοποιείται η τυπική απόκλιση του δείγματος για τον υπολογισμό του SEM και τα όρια είναι διαφορετικά

$$\hat{\mu} \pm t_{\nu,\alpha/2} \times \hat{\sigma}_{\hat{\mu}}$$



$$\pm t_{5,.025} = \pm 2.57$$



$$\pm t_{29,.025} = \pm 2.05$$

Άσκηση

- Λαμβάνουμε το σκορ IQ 10 υγειών ανθρώπων: 96 112 103 89 105 112 96 118 102 107
  - Να βρεθεί το διάστημα εμπιστοσύνης του μέσου 95% CI, υποθέτοντας ότι ο πληθυσμός ακολουθεί την κανονική κατανομή

Degrees of Freedom	90th Percentile (a = .10)	95th Percentile (a = .05)	97.5th Percentile (a = .025)	98th Percentile (a = .02)	99th Percentile (a = .01)
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.333	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.989
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845

Τα παλιά τα χρόνια…

Άσκηση

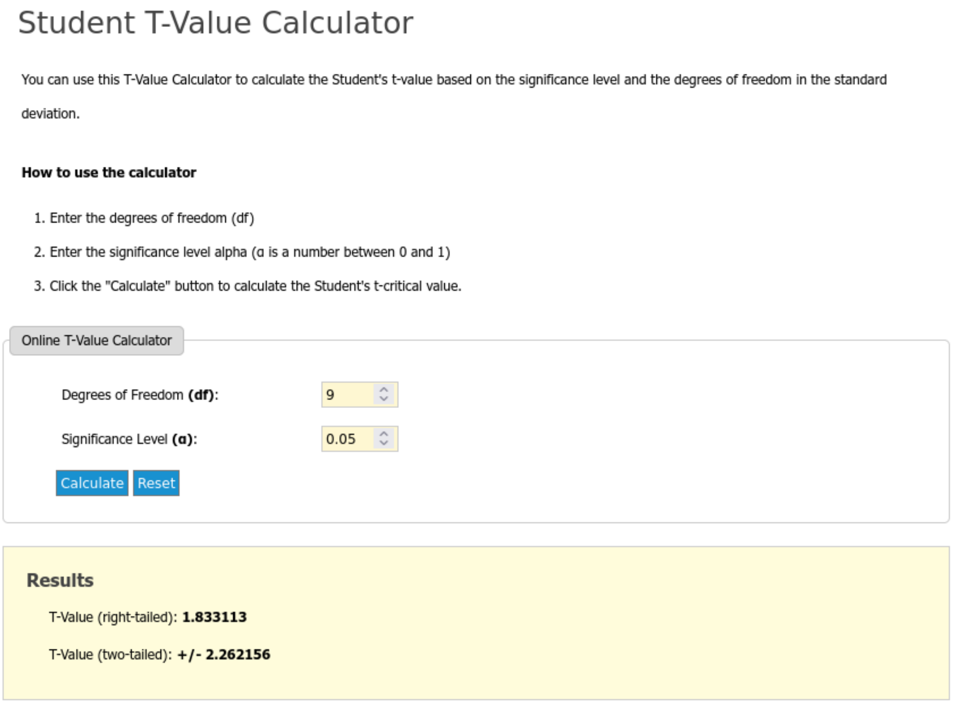
- Λαμβάνουμε το σκορ IQ 10 υγειών ανθρώπων: 96 112 103 89 105 112 96 118 102 107
  - Να βρεθεί το διάστημα εμπιστοσύνης του μέσου 95% CI, υποθέτοντας ότι ο πληθυσμός ακολουθεί την κανονική κατανομή

Ο φίλος μας το google

The sample size is n=10, the degrees of freedom (df) = n-1 = 9. The t value for 95% confidence with df = 9 is t = 2.262.

Άσκηση

- Λαμβάνουμε το σκορ IQ 10 υγειών ανθρώπων: 96 112 103 89 105 112 96 118 102 107
  - Να βρεθεί το διάστημα εμπιστοσύνης του μέσου 95% CI, υποθέτοντας ότι ο πληθυσμός ακολουθεί την κανονική κατανομή



https://goodcalculators.com/student-t-value-calculator/

Άσκηση

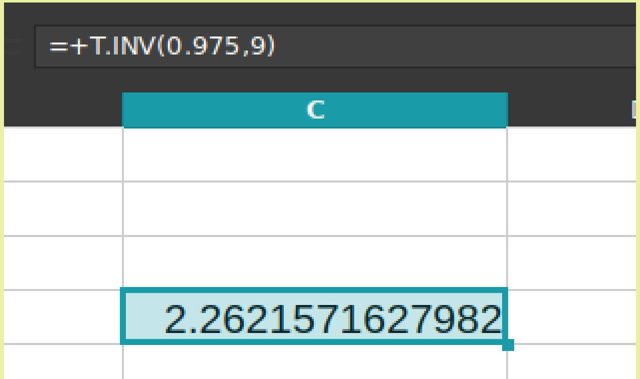
- Λαμβάνουμε το σκορ IQ 10 υγειών ανθρώπων: 96 112 103 89 105 112 96 118 102 107
  - Να βρεθεί το διάστημα εμπιστοσύνης του μέσου 95% CI, υποθέτοντας ότι ο πληθυσμός ακολουθεί την κανονική κατανομή

```
> qt(0.025, 9)
[1] -2.262157
> qt(0.975, 9)
[1] 2.262157
> _
```

Ο φίλος μας το R

Άσκηση

- Λαμβάνουμε το σκορ IQ 10 υγειών ανθρώπων: 96 112 103 89 105 112 96 118 102 107
  - Να βρεθεί το διάστημα εμπιστοσύνης του μέσου 95% CI, υποθέτοντας ότι ο πληθυσμός ακολουθεί την κανονική κατανομή



Ο φίλος μας το libreoffice calc (και το excel)

Άσκηση

- Λαμβάνουμε το σκορ IQ **10** υγιών ανθρώπων: 96 112 103 89 105 112 96 118 102 107
  - Να βρεθεί το διάστημα εμπιστοσύνης του μέσου 95% CI, υποθέτοντας ότι ο πληθυσμός ακολουθεί την κανονική κατανομή
- Υπολογίζουμε τον μέσο και τυπική απόκλιση στο δείγμα

$$\hat{\mu} = 104.0 \quad \hat{\sigma} = 8.8$$

- Υπολογίζουμε το SEM

$$\hat{\sigma}_{\hat{\mu}} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{8.8}{\sqrt{10}} = 2.77$$

Άσκηση

- Αφού γνωρίζουμε την τιμή t για τα όρια του *α*=0,05%, (~2,262), μπορούμε να βρούμε το περιθώριο του σφάλματος (ΜΟΕ) για 95% διάστημα εμπιστοσύνης, που είναι

$$MOE = t_{9,.975} \times \hat{\sigma}_{\hat{\mu}} \simeq 2.262 \times 2.77 = 6.27$$

- Έτσι, το διάστημα εμπιστοσύνης του μέσου του πειράματος είναι:

104 +/- 6,27

Πότε χρησιμοποιούμε z και πότε t test;

- Όταν **γνωρίζουμε** τη διακύμανση του πληθυσμού, χρησιμοποιούμε το z-test γιατί οι μέσοι των δειγμάτων μας ακολουθούν την κανονική κατανομή
  - Και το όριο για 0,025 και 0,975 είναι -1,96 και +1,96 αντίστοιχα
- Όταν **δεν γνωρίζουμε** τη διακύμανση του πληθυσμού, αξιοποιούμε αυτήν του δείγματος και χρησιμοποιούμε το t-test γιατί οι μέσοι των δειγμάτων μας ακολουθούν την κατανομή Student
  - Και το όριο για 0,025 και 0,975 αλλάζει και εξαρτάται και από τους βαθμούς ελευθερίας (ν=n-1)
- **Όταν το δείγμα είναι μεγαλύτερο από 30, τότε ακόμα και αν δεν γνωρίζουμε τη διακύμανση του πληθυσμού χρησιμοποιούμε το z-test**

Έλεγχος σημαντικότητας

- Ένα συμπέρασμα με τη χρήση διαστήματος εμπιστοσύνης (CI) λέγεται **έλεγχος σημαντικότητας** (significance testing)
- Ένας έλεγχος σημαντικότητας αφορά μια **υπόθεση** που κάνουμε για την τιμή μιας παραμέτρου ενός πληθυσμού

Στατιστικό συμπέρασμα

- Η διαδικασία της εξαγωγής των παραμέτρων μιας υποκείμενης κατανομής πιθανοτήτων από ένα δείγμα
- Τέσσερις κατηγορίες:
  - Σημειακές εκτιμήσεις
  - Εκτιμήσεις διαστήματος
  - **Έλεγχος υποθέσεων**
  - Πρόβλεψη

Έλεγχος υποθέσεων

Μια υπόθεση

- Μια υπόθεση αφορά την τιμή που λαμβάνει η παράμετρος ενός πληθυσμού
  - Η διάμεσος του ύψους των γυναικών στην Ελλάδα είναι 1,66
  - Ο μέσος του ύψους των ανδρών στην Ελλάδα είναι μικρότερος από αυτόν των ανδρών στην Κροατία
  - Το ύψος των γυναικών στην Ελλάδα ακολουθεί την κανονική κατανομή
  - Το φάρμακο Χ βελτιώνει την κλινική εικόνα των ασθενών μετά από χρήση δύο εβδομάδων

Έλεγχος υποθέσεων

- Η διαδικασία προσδιορισμού αν μια δεδομένη υπόθεση ισχύει ή όχι, με τη χρήση της στατιστικής
  - Βρίσκει εφαρμογή σε στοχαστικά προβλήματα απόφασης μεταξύ δύο εναλλακτικών υποθέσεων
- Η μία υπόθεση συμβολίζεται με H_0 και ονομάζεται **μηδενική υπόθεση** (null hypothesis)
  - Η άλλη με H_1 και ονομάζεται **εναλλακτική υπόθεση** (alternative hypothesis)

Διατύπωση μηδενικής υπόθεσης

- Η μηδενική υπόθεση πρέπει να είναι διατυπωμένη με σαφήνεια

Διατύπωση μηδενικής υπόθεσης

- Η μηδενική υπόθεση είναι συνήθως η πιο “βαρετή”
- $\mathbf{H_0}$: *το βάρος του πτηνού δεν σχετίζεται με τη μορφή της μεταλλικής βέργας*

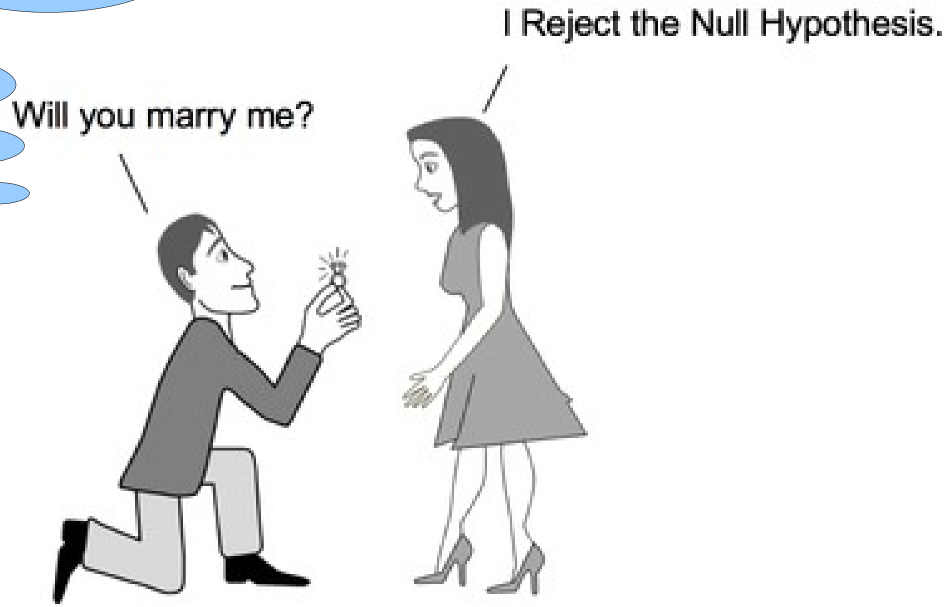
Παραδείγματα μηδενικών υποθέσεων

- Το μέσο ύψος των ανδρών είναι ίσο με το μέσο ύψος στις γυναίκες
- Ο μέσος χρόνος επιλογής με το ποντίκι είναι εξίσου γρήγορος με τον μέσο όρο χρόνου επιλογής με trackpad
- Όλες οι ποικιλίες καλαμποκιού μεγαλώνουν με την ίδια ταχύτητα στο ίδιο χωράφι
- Ένα φάρμακο Χ δεν έχει καμία επίδραση στην κλινική εικόνα των ασθενών

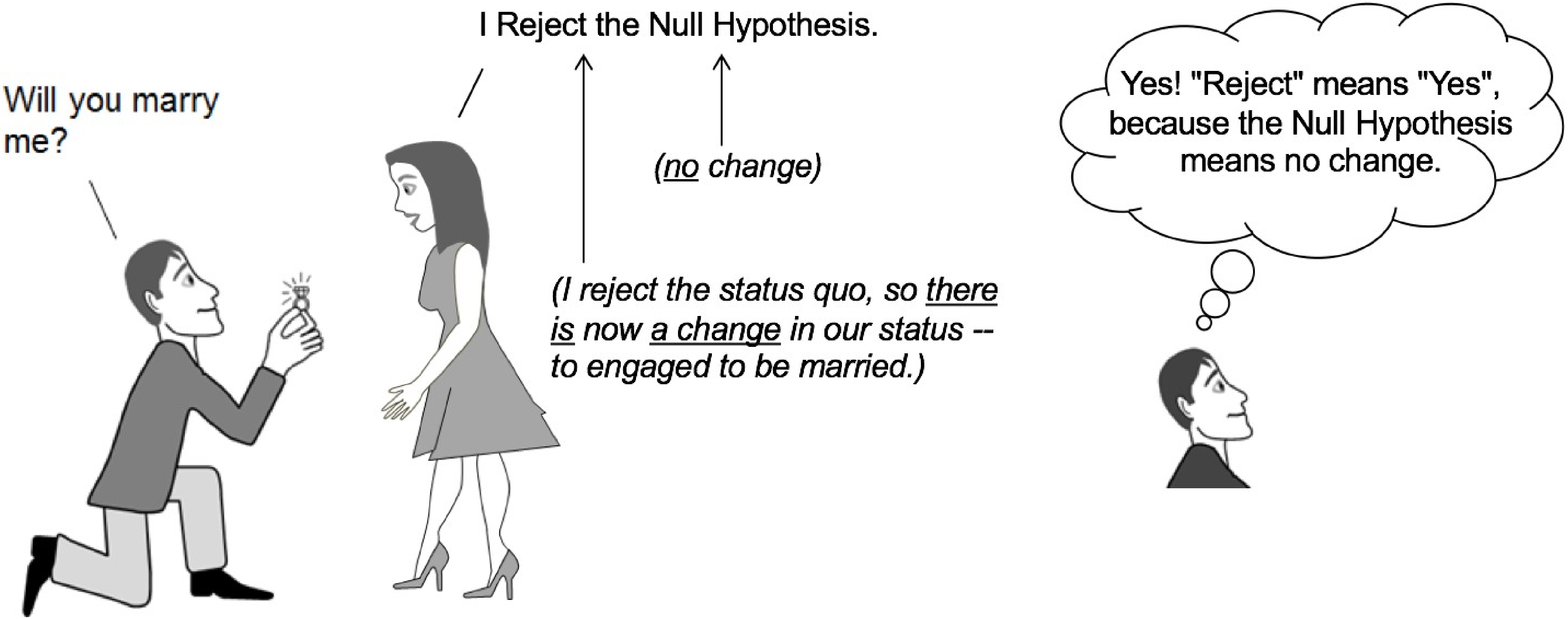
Μηδενική υπόθεση (H_0)

- Θέτουμε ως μηδενική υπόθεση (H_0) αυτήν που αμφισβητείται
  - Εξετάζουμε αν ένα τυχαίο δείγμα του πληθυσμού μπορεί να στηρίξει την απόρριψή της, έναντι της εναλλακτικής (H_1)
- Παραδείγματα:
  - H_0: $\mu=1000$ & H_1: $\mu\neq1000$
  - H_0: $\mu>200$ & H_1: $\mu\leq200$

Τελικά δέχεται ή όχι;



From the book *Statistics from A to Z – Confusing Concepts Clarified.*


Will you marry me?
I Reject the Null Hypothesis.
(no change)
(I reject the status quo, so there is now a change in our status -- to engaged to be married.)
Yes! "Reject" means "Yes", because the Null Hypothesis means no change.

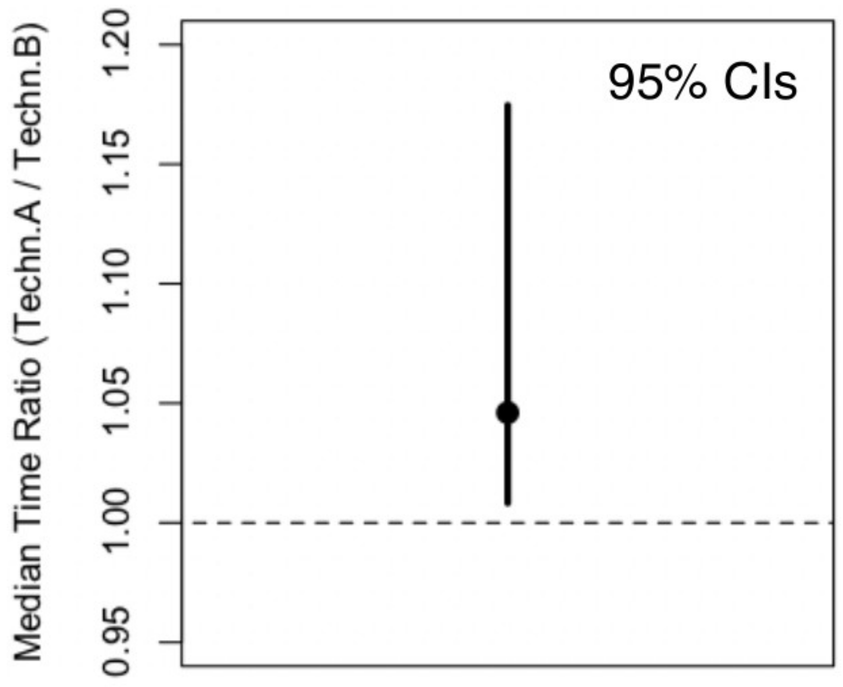
Μηδενική υπόθεση (H_0)

- Η H_0 απορρίπτεται ή δεν απορρίπτεται με βάση το τι παρατηρούμε στο τυχαίο δείγμα και υποθέτοντας ότι η H_0 είναι αληθής
  - Αν αυτό που παρατηρείται στο δείγμα έχει **πολύ μικρή πιθανότητα** να συμβεί, τότε απορρίπτουμε την H_0
  - Αν αυτό που παρατηρείται στο δείγμα **δεν είναι ακραίο-σπάνιο**, τότε δεν έχουμε δίνει αρκετές ενδείξεις για την απόρριψη της H_0
- Προσοχή: Δεν δεχόμαστε την H_0, απλώς “αποτυγχάνουμε να την απορρίψουμε”
  - Το δείγμα μας δεν μας το επιτρέπει
  - Αντίθετα, όταν απορρίπτουμε την H_0 δεχόμαστε την H_1

Πότε απορρίπτουμε τη μηδενική υπόθεση;

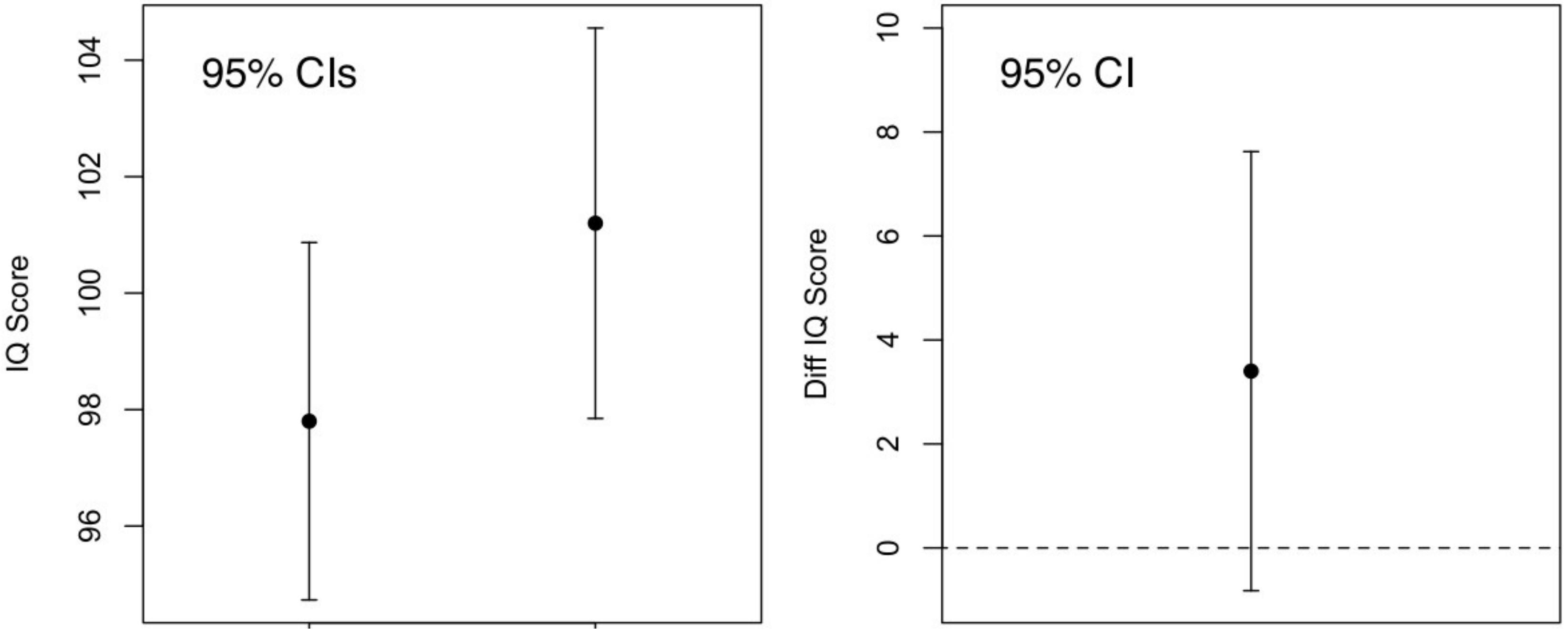
- Εάν το διάστημα εμπιστοσύνης *C*% αποκλείει την υποθετική τιμή της παραμέτρου του πληθυσμού, τότε η υπόθεση απορρίπτεται
  - για *C*% επίπεδο εμπιστοσύνης, π.χ. 95%

Παράδειγμα



- Η H_0 είναι ότι η παράμετρος που μελετούμε σε έναν πληθυσμό είναι ίση με 1
- Όμως στο πείραμα που κάναμε, βλέπουμε ότι το CI για 95% δεν περιλαμβάνει το 1
- Η H_0 απορρίπτεται

Παράδειγμα



Εδώ, η H_0 είναι ότι δύο ομάδες έχουν το ίδιο IQ, άρα η διαφορά τους θα είναι 0. Η H_0 δεν μπορεί να απορριφθεί στατιστικά. Αυτό δεν σημαίνει ότι δεν υπάρχουν διαφορές ανάμεσα στις δύο ομάδες, αυτές όμως δεν είναι **στατιστικά σημαντικές**

Ερωτήματα

- Τι εννοούμε όταν λέμε **παρατήρηση στο δείγμα**;
  - Πώς εκφράζεται;
  - Μπορεί να μετρηθεί - ποσοτικοποιηθεί;
- Πώς κρίνουμε ότι μια παρατήρηση στο δείγμα είναι ή όχι **ακραία**;
  - Με ποιον κανόνα θεωρείται η παρατήρηση στο δείγμα ακραία;

Σφάλμα

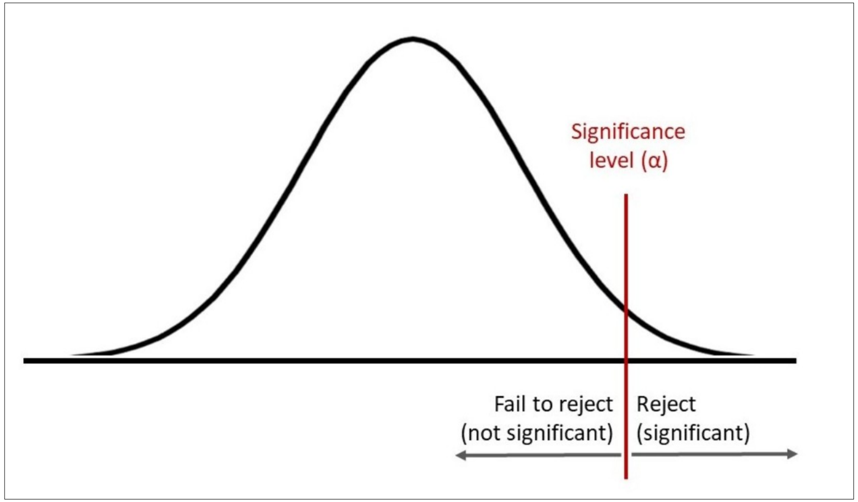
- Στον στατιστικό έλεγχο υποθέσεων υπάρχει η δυνατότητα σφάλματος
  - Τύπου I, όπου κάνω απόρριψη μιας σωστής μηδενικής υπόθεσης
  - Τύπου II, όπου κάνω αποδοχή μιας λανθασμένης μηδενικής υπόθεσης

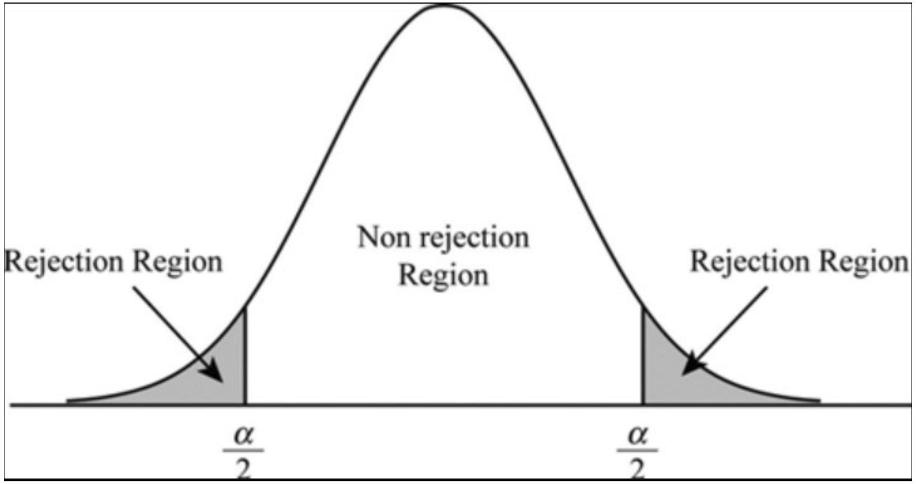
	Null Hypothesis is TRUE	Null Hypothesis is FALSE
Reject null hypothesis	Type I Error (False positive)	Correct Outcome! (True positive)
Fail to reject null hypothesis	Correct Outcome! (True negative)	Type II Error (False negative)

Πιθανότητες σφαλμάτων

- Σφάλμα τύπου Ι έχουμε όταν απορρίπτουμε μια ορθή H_0, με **πιθανότητα *α***
  - Η πιθανότητα ενός σφάλματος τύπου Ι (*α*) ονομάζεται **επίπεδο σημαντικότητας** ενός ελέγχου υποθέσεων
  - Μπορούμε να το ορίσουμε εμείς ανάλογα με την εμπειρία και τα δεδομένα μας, π.χ. σε τι ποσοστό μπορούμε να δεχτούμε ένα σφάλμα τύπου Ι για μια συγκεκριμένη υπόθεση;
- Σφάλμα τύπου ΙΙ έχουμε όταν δεχόμαστε μια εσφαλμένη H_0, με **πιθανότητα *β***
  - Η πιθανότητα να απορρίψουμε ορθά μια εσφαλμένη H_0 είναι $1-\beta$ και συμβολίζει τη **στατιστική ισχύ** ενός ελέγχου υποθέσεων
    - Η ισχύς ενός ελέγχου είναι η πιθανότητα ότι μια λανθασμένη μηδενική υπόθεση θα εντοπιστεί από τον έλεγχο

Επίπεδο σημαντικότητας α



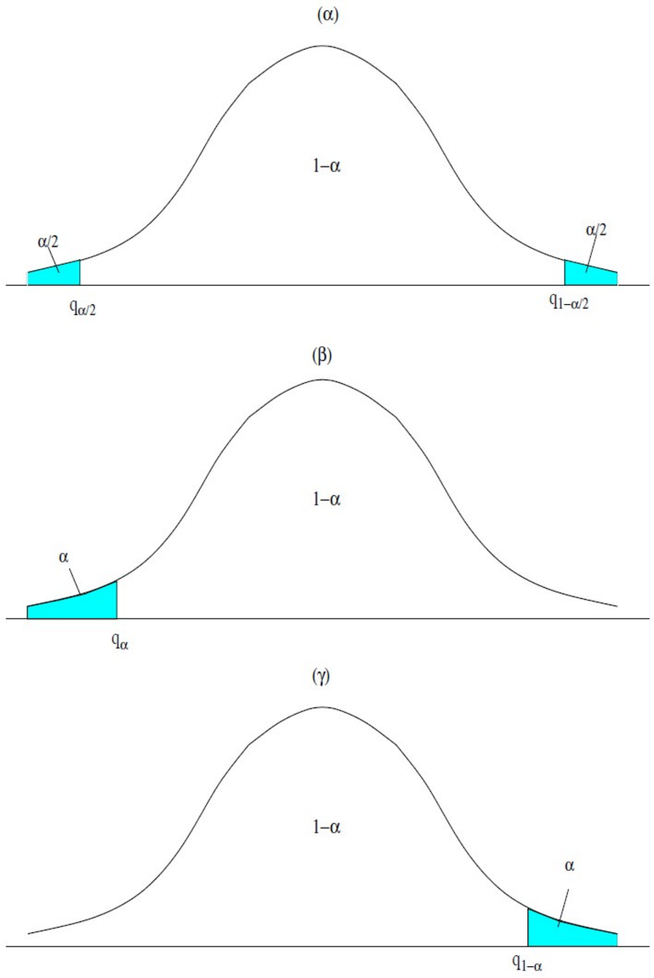


Έλεγχος υποθέσεων

Δίπλευρος

Μονόπλευρος προς τα αριστερά

Μονόπλευρος προς τα δεξιά



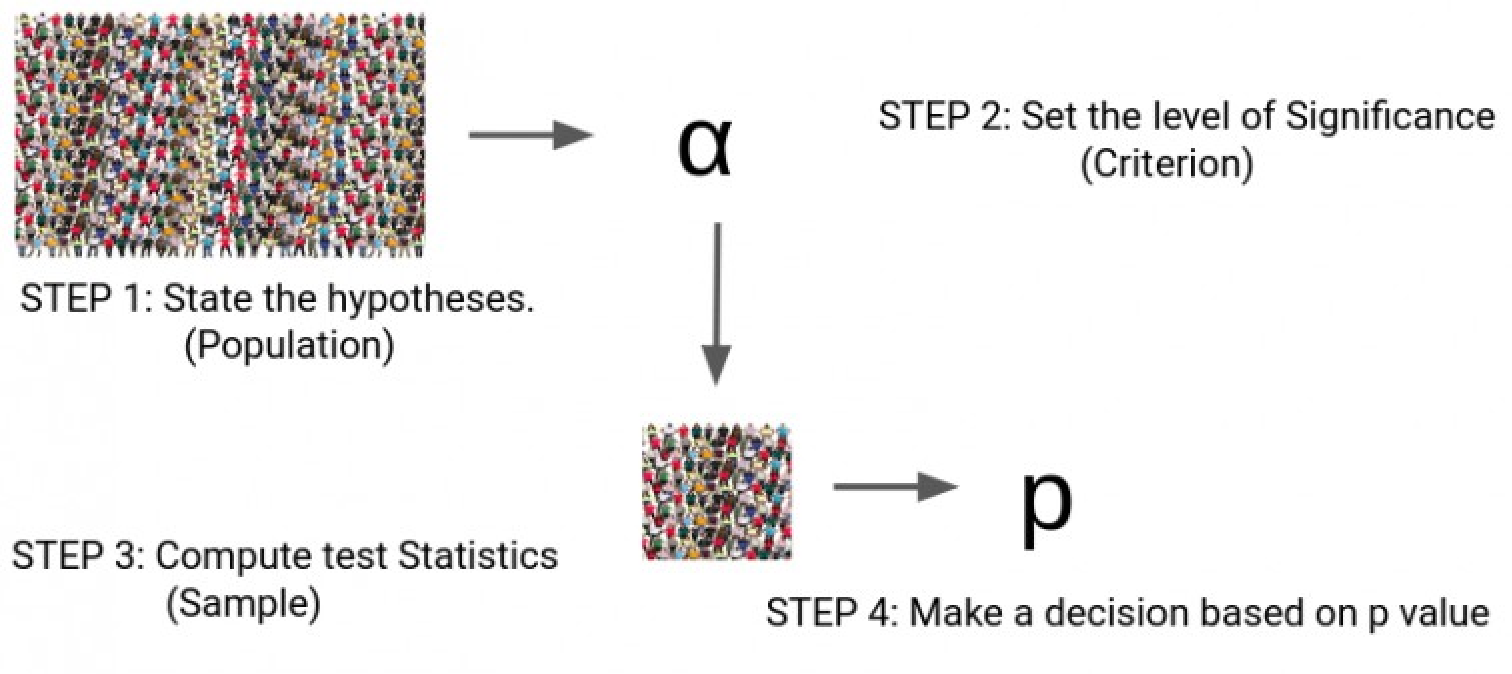
Γιατί 95%;

- Συνήθως χρησιμοποιείται ως όριο για την απόρριψη μιας μηδενικής υπόθεσης (επίπεδο σημαντικότητας)
- Δεν είναι μαγικός αριθμός και δεν υπάρχει λόγος να μην χρησιμοποιείται διαφορετικό επίπεδο (π.χ. 92% ή 97%)
  - Αλλά η χρήση του αντανακλά μια μακρά παράδοση στην επιστήμη
- Υπάρχουν και άλλα επίπεδα σημαντικότητας στη βιολογία
  - 95% *
  - 99% **
  - 99,9% ***

Έλεγχος σημαντικότητας

- Ο έλεγχος της σημαντικότητας μιας μηδενικής υπόθεσης H0 γίνεται μέσω διαστημάτων εμπιστοσύνης (CI)
- Ο έλεγχος σημαντικότητας είναι πολύ κοινό εργαλείο έρευνας, αλλά υπάρχουν περιορισμοί
  - Συχνά χρησιμοποιούνται υπερβολικά ή κακώς
- Το αποτέλεσμα ενός ελέγχου σημαντικότητας είναι μια τιμή πιθανότητας, που λέγεται **p value**

Μεθοδολογία ελέγχου υποθέσεων

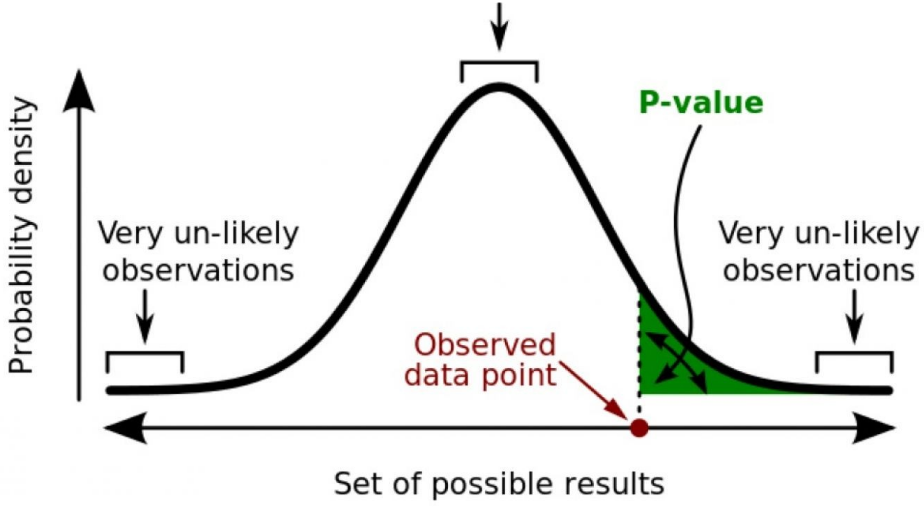


Και τι είναι τελικά μια **p-value**;

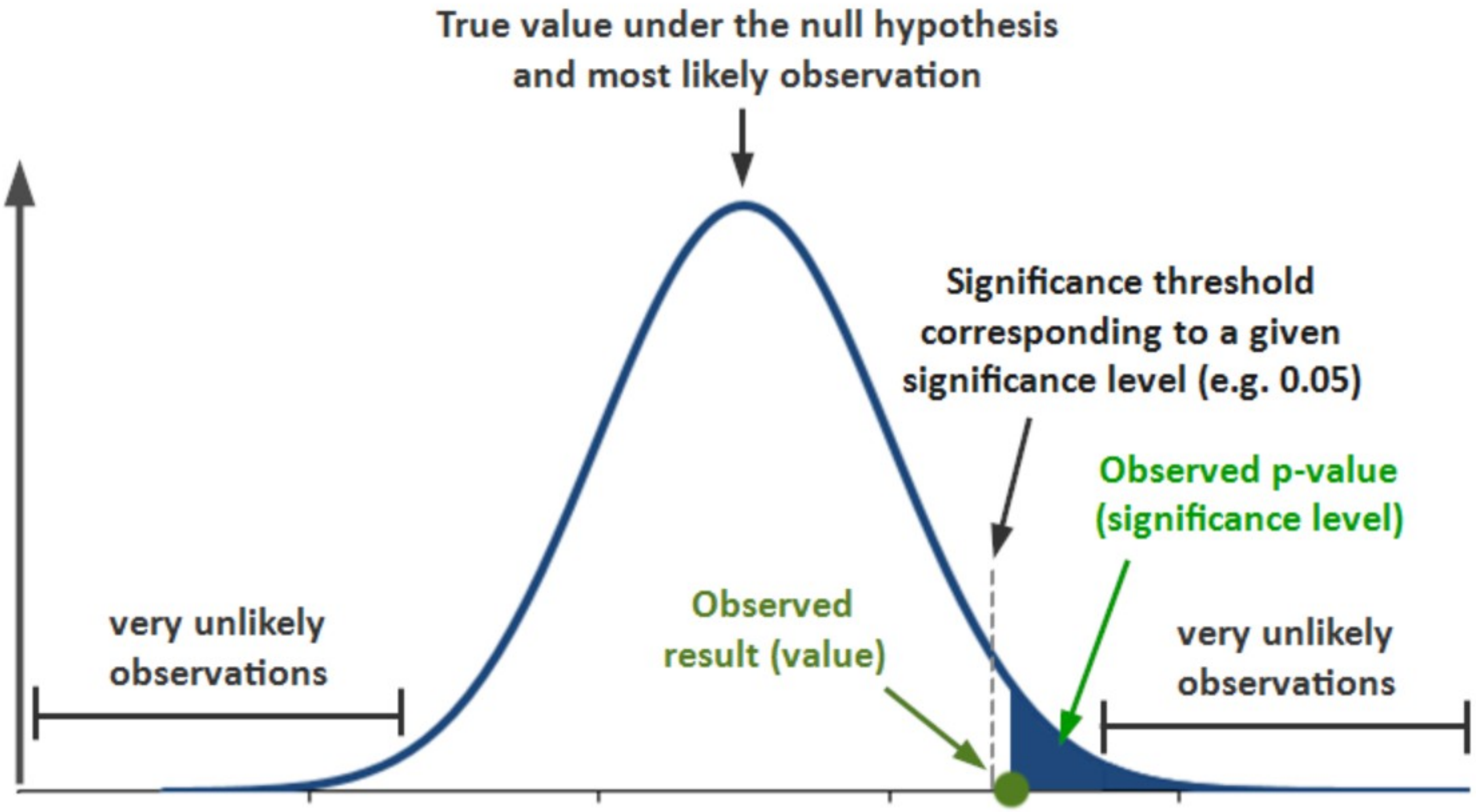
- Είναι η πιθανότητα να προκύψει μια εξίσου ή πλέον απομακρυσμένη τιμή στο δείγμα σε σχέση με μια συγκεκριμένη παρατήρηση (στατιστική τιμή), όταν η μηδενική υπόθεση H_0 είναι ορθή
- *Θέλουμε κάτι πιο απλό!*

Τι είναι το p-value;

- Ένα συγκεκριμένο στατιστικό τοποθετείται σε μια κατανομή πιθανοτήτων
  - Ισχύει η H_0
- Η πιθανότητα για οποιοδήποτε άλλο στατιστικό άλλου δείγματος στην ίδια κατανομή να λάβει τιμή εξίσου ή πιο απομακρυσμένη από τον μέσο όρο (πιο σπάνια δηλαδή) από την αρχική τιμή, είναι το **p-value**
- Είναι η πιθανότητα, η τιμή του στατιστικού που πήραμε να είναι τυχαία

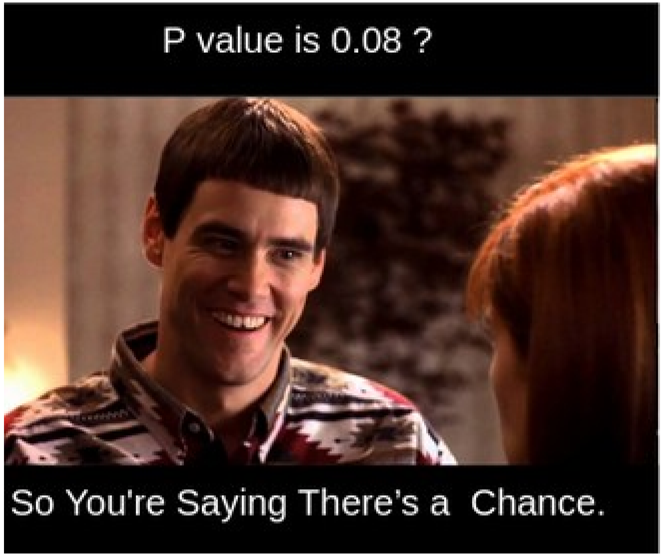


p-value και α



p-value και *α*

- Όταν η p-value είναι μικρότερη από το επίπεδο σημαντικότητας *α*, τότε η μηδενική υπόθεση H_0 απορρίπτεται
- Το *α* είναι η μέγιστη p-value για την οποία απορρίπτεται η H_0
  - Είναι η μέγιστη πιθανότητα να κάνουμε σφάλμα τύπου I



Ένα παράδειγμα θα μας πείσει...

- Θέλουμε να δούμε αν ο μισθός των ανδρών και των γυναικών στην Ελλάδα είναι ίδιος
    - H_0: Η μέση διαφορά μισθού ανδρών και γυναικών είναι 0
    - H_1: Η μέση διαφορά μισθού ανδρών και γυναικών είναι πάνω από το 0

Παράδειγμα

- Παίρνουμε ένα δείγμα 15 ανδρών και ένα 15 γυναικών παρόμοιας ηλικίας σε μια πόλη
  - Η μέση διαφορά μισθού ανδρών και γυναικών είναι 150 ευρώ
  - Από τον έλεγχο σημαντικότητας προκύπτει ότι η **p-value** είναι ***p=0.006***

Τι σημαίνει αυτή η p-value;

Παράδειγμα

- Μια p-value ίση με ***p=0.006***, σημαίνει ότι:
  - αν ισχύει η υπόθεση H_0 και δεν υπάρχει μισθολογική διαφορά ανάμεσα σε άνδρες και γυναίκες, τότε
  - σε ένα νέο δείγμα 15 μισθολογικών διαφορών ανδρών και γυναικών από τον ίδιο πληθυσμό, **η πιθανότητα ο μέσος να είναι ίσος ή μεγαλύτερος από 150 ευρώ είναι 0.006**

Τώρα μάλιστα!

Παράδειγμα

- Μια p-value ίση με ***p=0.006***, σημαίνει ότι:
  - η πιθανότητα σε έναν πληθυσμό που ισχύει η υπόθεση H_0 και δεν υπάρχει μισθολογική διαφορά ανάμεσα σε άνδρες και γυναίκες, να προκύψει δείγμα 15 μισθολογικών διαφορών ανδρών και γυναικών από τον ίδιο πληθυσμό με μέσο Μ=150 ευρώ, **είναι 0.006**

Παράδειγμα

- Είναι αυτή η διαφορά 150 ευρώ στον μισθό ανδρών και γυναικών τυχαία για ένα δείγμα 15 ατόμων;
  - **Όχι**, γιατί η p-value είναι ίση με ***p=0.006*** και είναι μικρότερη του ορίου σημαντικότητας ***α=0.05*** για επίπεδο σημαντικότητας 95%
  - Άρα απορρίπτουμε την H_0

Θα έχουμε την ευκαιρία να δούμε πολλά υπολογιστικά παραδείγματα με p-value στα επόμενα μαθήματα...

https://www.youtube.com/watch?v=ukcFrzt6cHk

Thank you