



ΔΗΜΟΚΡΙΤΕΙΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΘΡΑΚΗΣ

DEMOCRITUS  
UNIVERSITY  
OF THRACE

ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ &  
ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗΣ  
ΠΜΣ ΣΤΗ ΛΟΓΙΣΤΙΚΗ & ΕΛΕΓΚΤΙΚΗ

# Στατιστικές Μέθοδοι Έρευνας

## Έλεγχος A/B



Καθηγητής Παύλος Δελιάς

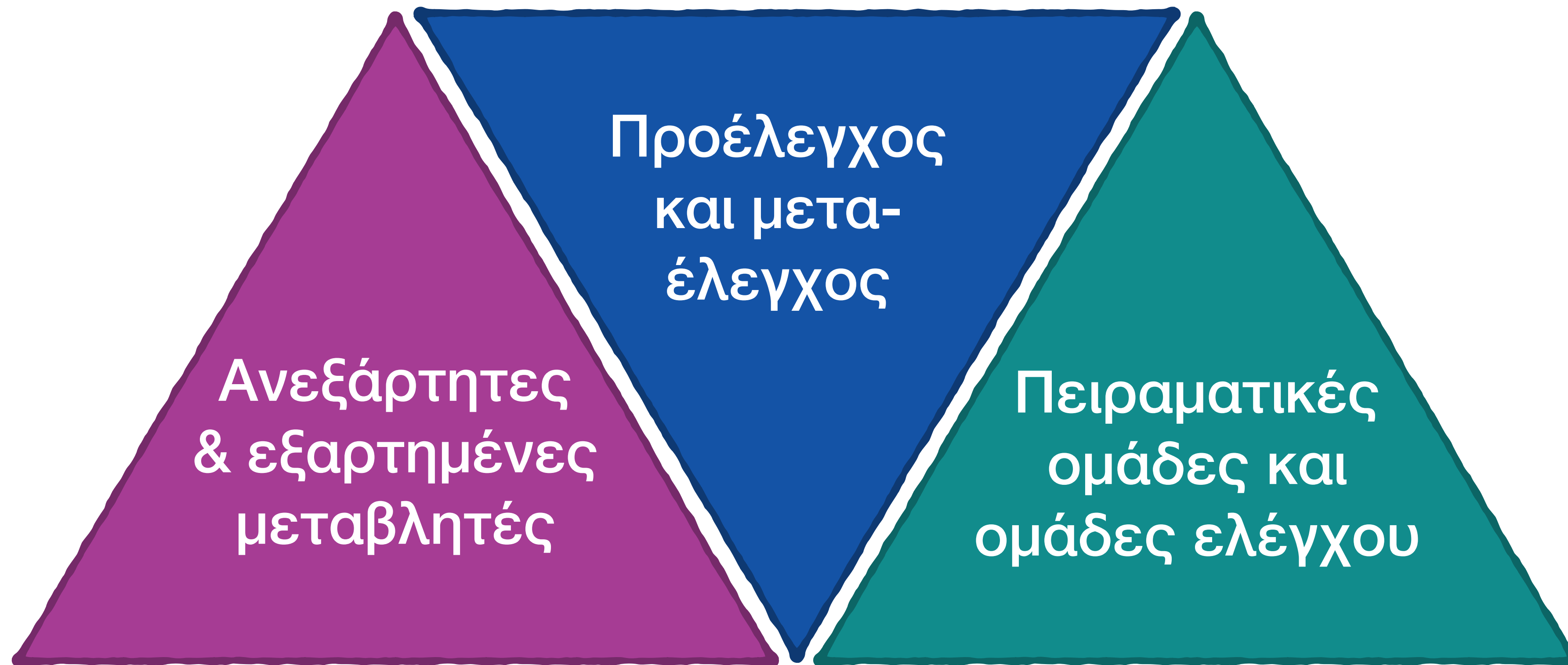


[pdelias@af.duth.gr](mailto:pdelias@af.duth.gr)



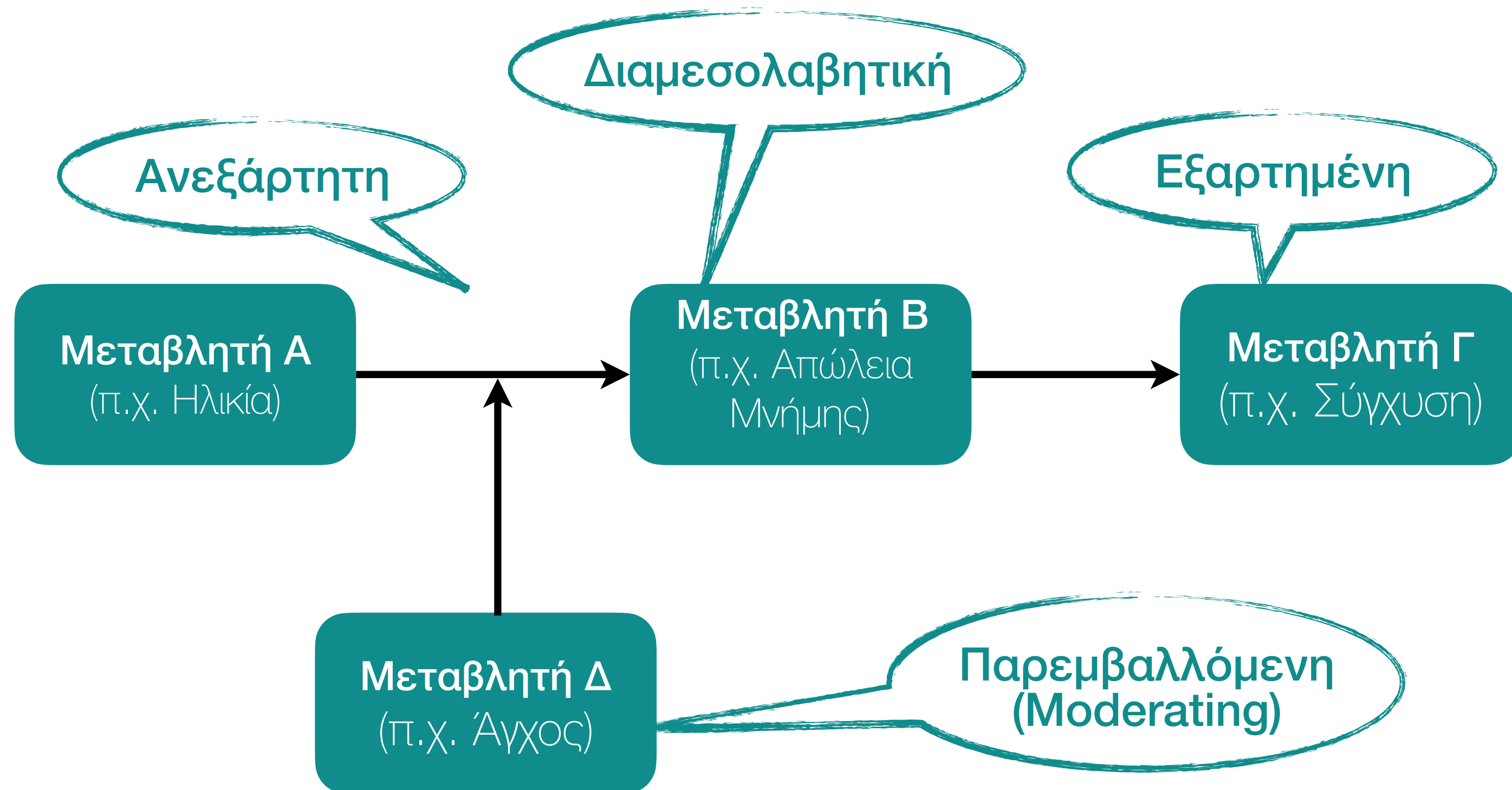
@PavlosDelias

# Κλασικό Πείραμα - Βασικά Συστατικά



# Είδη Μεταβλητών

## Νομολογικά Δίκτυα



# Έλεγχος εξωγενών μεταβλητών

- Γενικότεροι παράγοντες στο περιβάλλον που να επιδρούν στις αλλαγές στα επίπεδα της εξαρτημένης μεταβλητής
- Μπορούν να ελεγχτούν μέσω:
  - Εξάλειψης
  - Τυχαιοποίησης

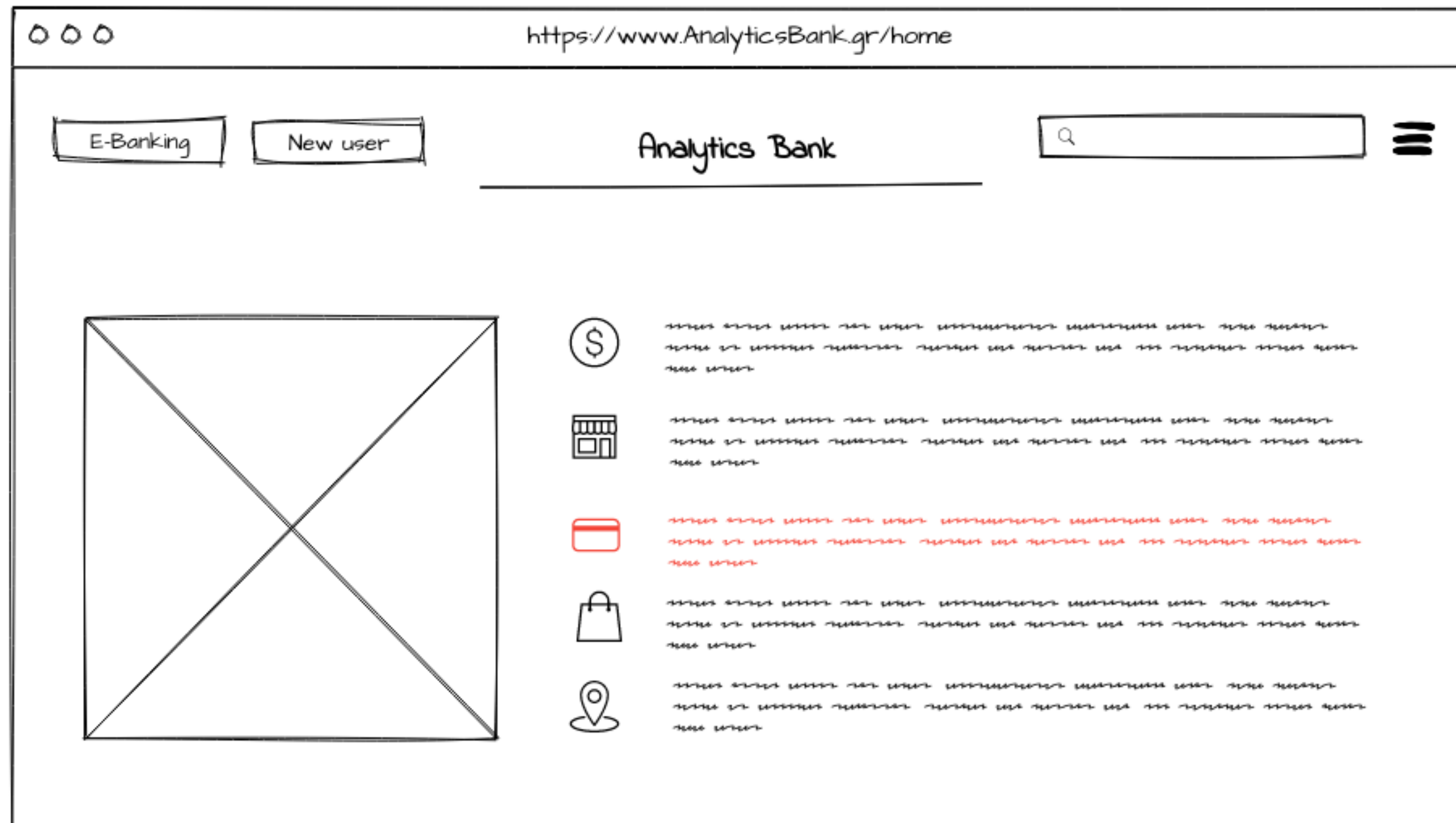
# Προ-έλεγχος & Μέτα-έλεγχος

- **Προέλεγχος** – Η μέτρηση μιας εξαρτημένης μεταβλητής προτού εκτεθούν τα υποκείμενα της έρευνας σε ένα ερέθισμα που αντιπροσωπεύει μια ανεξάρτητη μεταβλητή
- **Μετα-έλεγχος** – Η μέτρηση εκ νέου μιας εξαρτημένης μεταβλητής αφότου τα υποκείμενα της έρευνας εκτεθούν σε κάποιο ερέθισμα που αντιπροσωπεύει μια ανεξάρτητη μεταβλητή

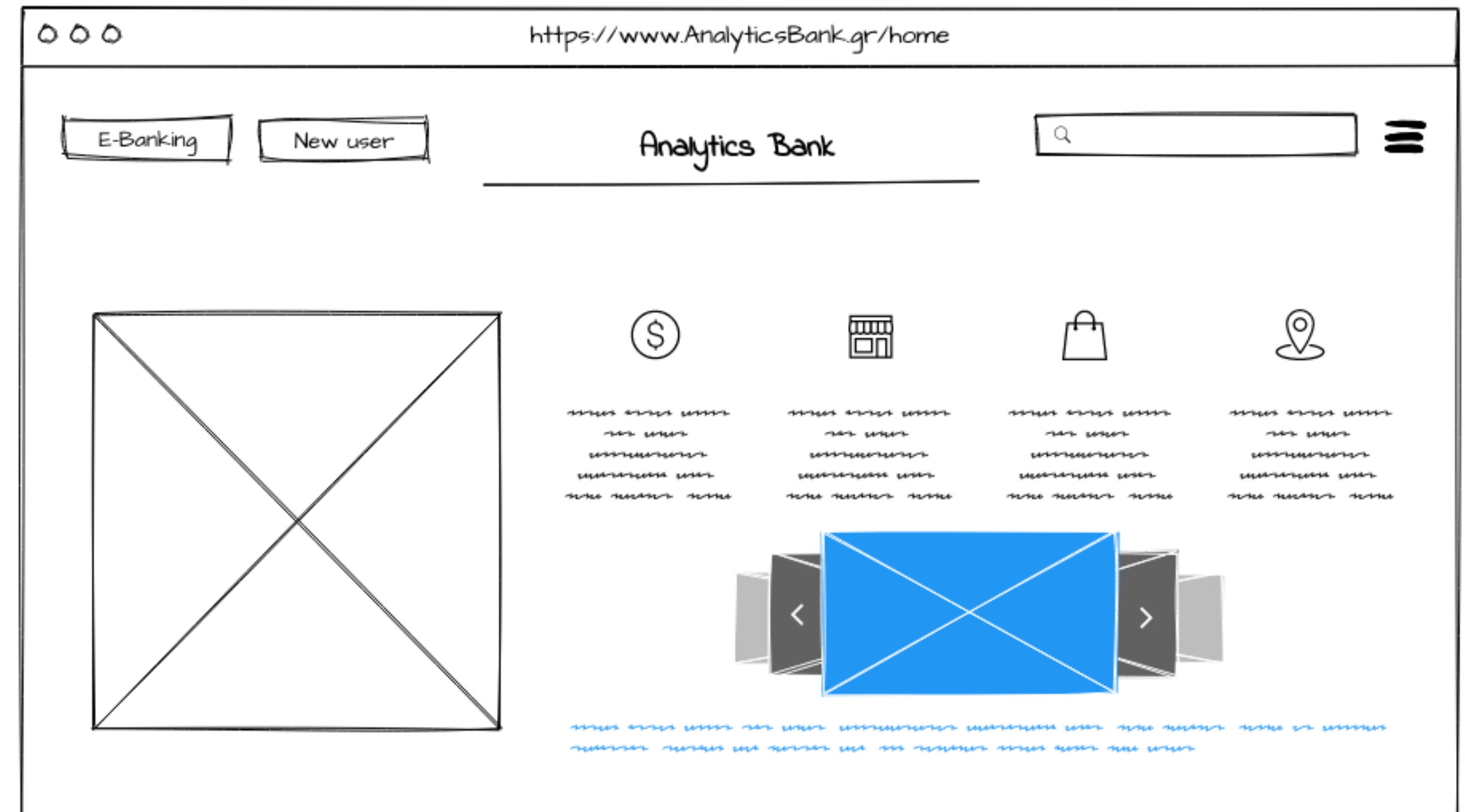
# Πειραματική Ομάδα & Ομάδα Ελέγχου

- **Πειραματική** ομάδα – Μια ομάδα υποκειμένων στα οποία ασκείται ένα πειραματικό ερέθισμα
- Ομάδα **ελέγχου** – μια ομάδα υποκειμένων έρευνας, τα οποία δεν υπόκεινται σε κανένα πειραματικό ερέθισμα και τα οποία έχουν, κατά τα άλλα, κοινά γνωρίσματα με την πειραματική ομάδα

# Ένα Επιχειρηματικό Παράδειγμα



Πριν



Μετά



# Που χρησιμεύει ο Έλεγχος A/B;

- Εκτίμηση επιτυχίας / απόδοσης μιας αλλαγής
  - Κατάλληλη για δοκιμή νέων επιμέρους στοιχείων
  - Δεν είναι επιθυμητή η άμεση αντικατάσταση του παλιού προϊόντος
- ...και που όχι**
- Δοκιμές για εντελώς νέα προϊόντα
  - Δοκιμές για μακρόχρονες αλλαγές
  - Όταν ψάχνουμε την γενικά βέλτιστη λύση / Πληρότητα χαρακτηριστικών



# Ερώτηση Εκμάθησης

## Που μπορούμε να εφαρμόσουμε τον Έλεγχο A/B και που όχι;

- A. Ένα κατάστημα ηλεκτρονικού εμπορίου θέλει να διαπιστώσει αν ο κατάλογος με τα προϊόντα που προσφέρει είναι πλήρης είναι αν πρέπει να εμπλουτιστεί και με νέα είδη.
- B. Μία εταιρεία προσφέρει δωρεάν ένα application για φορητές συσκευές. Θέλει να εξετάσει τη δυνατότητα να προσφέρει έναν λογαριασμό premium (επί πληρωμή) για χρήστες οι οποίοι θέλουν αυξημένη λειτουργικότητα και ειδικές επιπλέον παροχές.
- C. Μία εταιρεία προβολής ταινιών μέσω internet παρέχει συστάσεις για ταινίες στους συνδρομητές της. Θέλει να ελέγξει αν ο νέος αλγόριθμος συστάσεων που έχουν αναπτύξει ο οποίος προτείνει με διαφορετική λογική ταινίες είναι πιο αποτελεσματικός από τον υφιστάμενο αλγόριθμο.
- D. Μία εταιρεία θέλει να διαπιστώσει αν αξίζει να αλλάξει πάροχο για τις ιστοσελίδες που διαθέτει, κάτι που αναμένεται να επηρεάσει την ταχύτητα με την οποία εμφανίζονται οι ιστοσελίδες της στους χρήστες.

# Μέτρο επιτυχίας

## Το χωνί της πορείας των επισκεπτών



# Μέτρα Επιτυχίας

## Κατηγορίες

### Πλήθη και Αθροίσματα

Πόσα cookies επισκέφτηκαν την ιστοσελίδα;

### Λόγοι - Αναλογίες

### Μέτρα θέσης

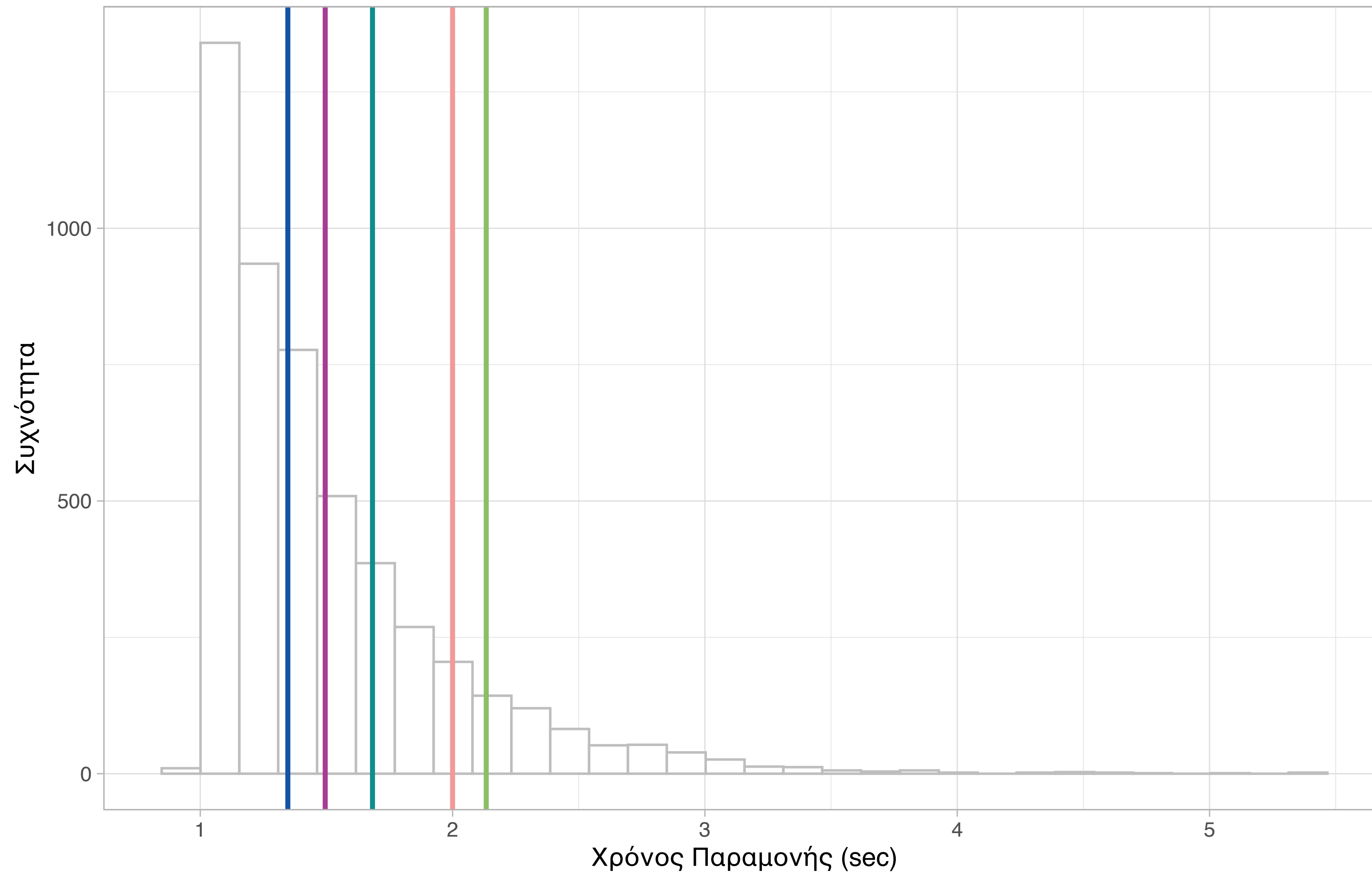
Μέση τιμή, διάμεσος, εκατοστημόρια, κτλ.

### Πιθανότητες και ρυθμοί

Πιθανότητα κλικ



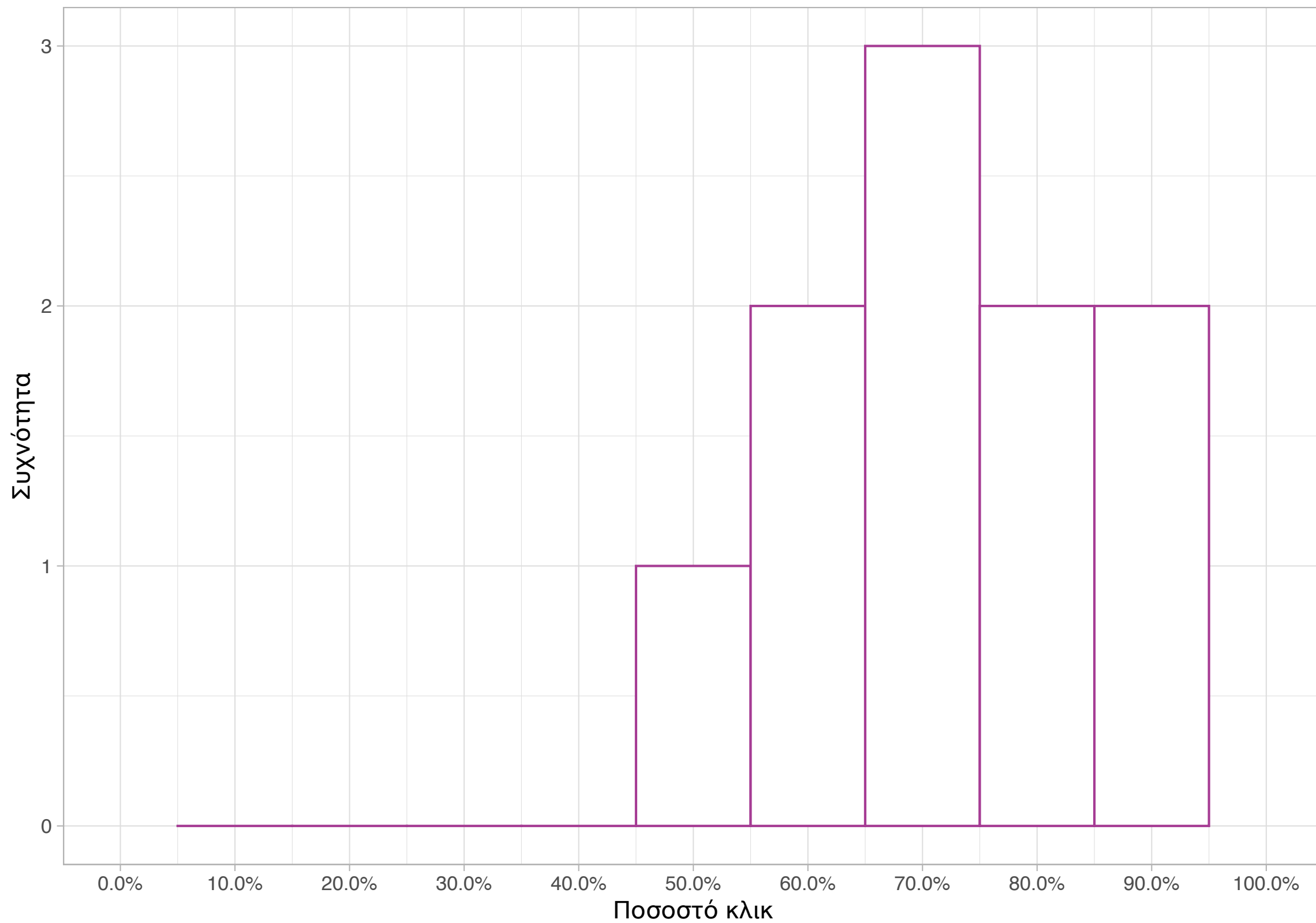
# Αξιολόγηση αποτελεσμάτων πειράματος με διάφορα μέτρα θέσης



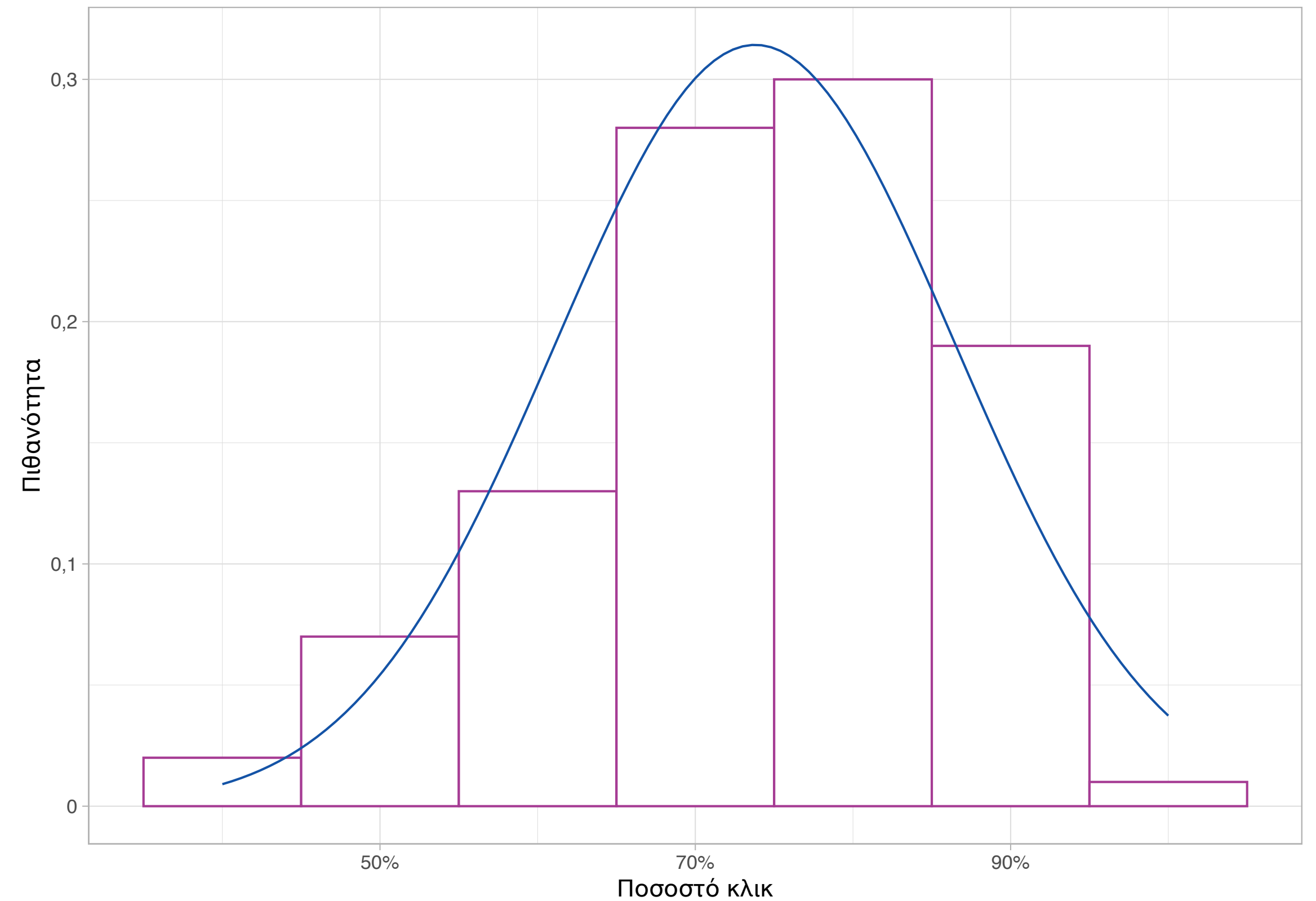
Μέτρα Μέση τιμή Διάμεσος 75ο εκατοστημόριο 90ο εκατοστημόριο τουλάχιστον 2 sec

# Μέτρηση της πιθανότητας επίσκεψης

## Διωνυμική κατανομή



10 πειράματα



100 πειράματα

# Διωνυμική κατανομή

## Βασικές παραδοχές

Ακριβώς δύο εναλλακτικές,  
αμοιβαία αποκλεισμένες



Οι περιπτώσεις είναι  
ανεξάρτητες

Ίδια πιθανότητα επιτυχίας για  
όλες τις περιπτώσεις

# Ερωτήσεις Εκμάθησης

**Που μπορούμε να χρησιμοποιήσουμε την διωνυμική κατανομή και που όχι;**

- A. Επιλέγουμε 20 χαρτιά από μία ανακατεμένη τράπουλα. Μεταβλητή αποτελέσματος: Κόκκινο χαρτί ή Μαύρο χαρτί.
- B. Παρακολουθούμε τα κλικ σε μία σελίδα αποτελεσμάτων μιας μηχανής αναζήτησης. Μεταβλητή αποτελέσματος: Κλικ ή χωρίς κλικ.
- C. Παρακολουθούμε την επιτυχία των φοιτητών σε ένα μάθημα. Μεταβλητή αποτελέσματος: Πέρασε ο φοιτητής το μάθημα ή όχι.
- D. Αγορά προϊόντων στο διάστημα μιας εβδομάδας. Μεταβλητή αποτελέσματος: Αγοράστηκε το προϊόν ή όχι.



# Διάστημα Εμπιστοσύνης για το μέτρο επιτυχίας

Σε ένα πείραμα για μία αλλαγή σε μία ιστοσελίδα, μαζεύουμε δεδομένα από 2000 επισκέπτες και βρίσκουμε πως 300 από αυτούς προχώρησαν στο πάτημα ενός συνδέσμου που μας ενδιαφέρει. Θέλουμε να υπολογίσουμε το διάστημα εμπιστοσύνης για την πιθανότητα πατήματος σε επίπεδο 99%

Μέση τιμή πιθανότητας

$$\hat{p} = \frac{X}{N}$$

Εύρος διαστήματος

$$m = z \times \sigma$$

Τυπική απόκλιση δείγματος

$$\sigma = \sqrt{\frac{\hat{p}(1 - \hat{p})}{N}}$$

Διάστημα εμπιστοσύνης

$$\hat{p} \pm m$$

# Έλεγχος Υποθέσεων στον Έλεγχο AB

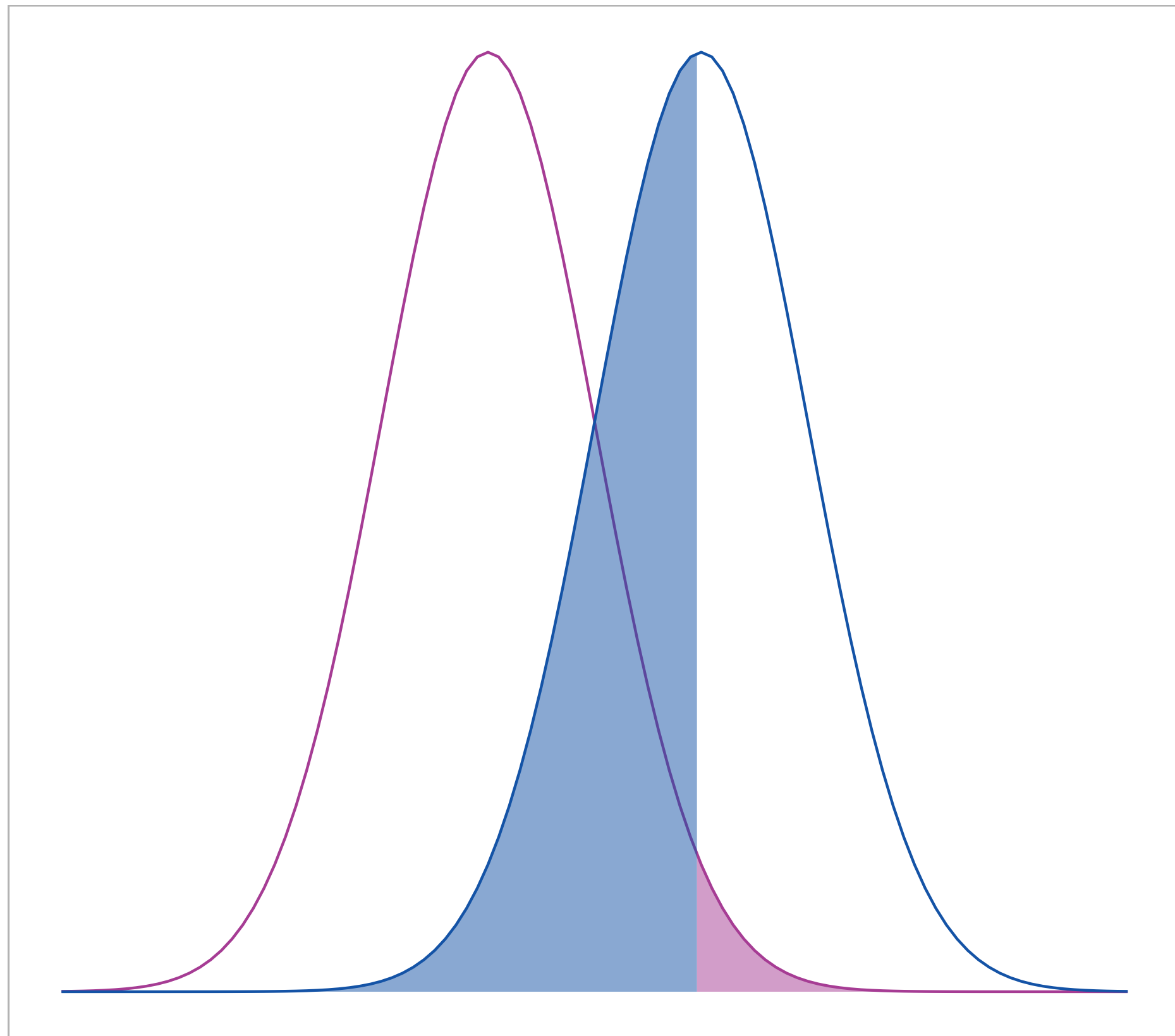
- **Μηδενική υπόθεση:** Τίποτα σημαντικό δεν άλλαξε με την αλλαγή που επιχειρήσαμε ( $H_0 : \delta = 0$ )
- **Εναλλακτική υπόθεση :** Οι πιθανότητες στις δύο ομάδες είναι διαφορετικές (αμφίπλευρος έλεγχος) ( $H_0 : \delta \neq 0$ )

$$\hat{P}_{ομαδ} = \frac{X_{ελ} + X_{\pi}}{N_{ελ} + N_{\pi}} \quad \hat{\sigma}_{ομαδ} = \sqrt{\hat{P}_{ομαδ} \times (1 - \hat{P}_{ομαδ}) \times \left(\frac{1}{N_{ελ}} + \frac{1}{N_{\pi}}\right)}$$

**Απορρίπτουμε την μηδενική υπόθεση:** Αν  $\hat{\delta} > z \times \hat{\sigma}_{ομαδ}$  ή αν  $\hat{\delta} < -z \times \hat{\sigma}_{ομαδ}$

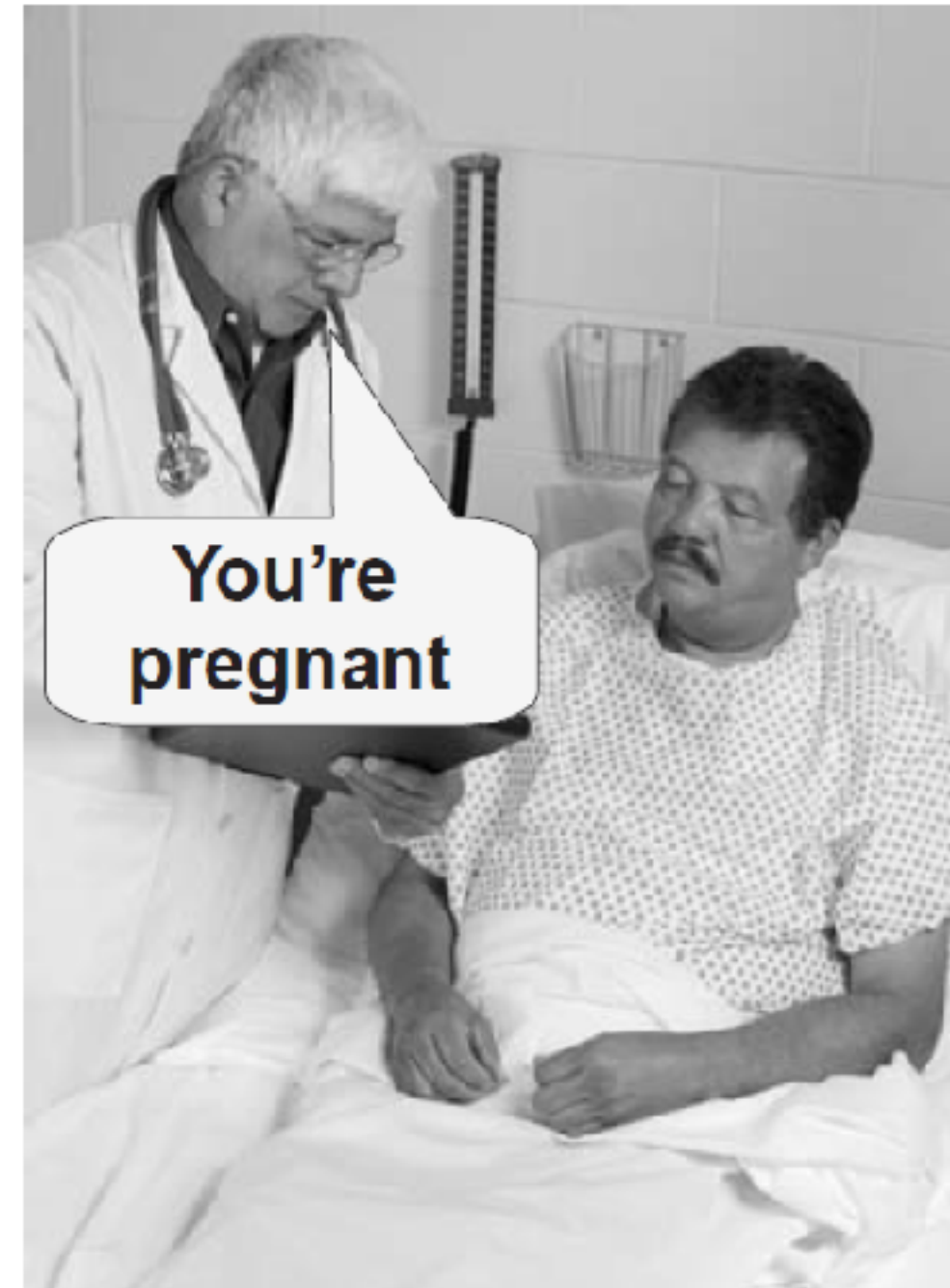
# Ισχύς ελέγχου του πειράματος

## Σφάλματα τύπου I - II



■ σφάλμα τύπου II ( $\beta$ )  
■ σφάλμα τύπου I ( $\alpha$ )

**Type I error**  
(false positive)



**Type II error**  
(false negative)



# Εκτιμήσεις και σφάλματα για τον έλεγχο υποθέσεων

		ΠΡΑΓΜΑΤΙΚΟΤΗΤΑ	
		Αληθής μηδενική υπόθεση (δεν αξίζει να προχωρήσουμε στην αλλαγή)	Ψευδής μηδενική υπόθεση (αξίζει να προχωρήσουμε στην αλλαγή)
<b>ΕΚΤΙΜΗΣΕΙΣ</b>	Δεχόμαστε την μηδενική υπόθεση (υποστηρίζουμε πως δεν αξίζει να προχωρήσουμε στην αλλαγή)	<b>1-α επίπεδο εμπιστοσύνης</b> (1-α)% φορές που δεν πρέπει να προχωρήσουμε στην αλλαγή υποστηρίζουμε πως δεν πρέπει να προχωρήσουμε	<b>β - Σφάλμα τύπου II</b> β% φορές που πρέπει να προχωρήσουμε στην αλλαγή υποστηρίζουμε πως δεν πρέπει να προχωρήσουμε
	Απορρίπτουμε την μηδενική υπόθεση (υποστηρίζουμε πως αξίζει να προχωρήσουμε στην αλλαγή)	<b>α - Σφάλμα τύπου I</b> α% φορές που δεν πρέπει να προχωρήσουμε στην αλλαγή υποστηρίζουμε πως πρέπει να προχωρήσουμε	<b>1-β ευαισθησία</b> (1-β)% φορές που πρέπει να προχωρήσουμε στην αλλαγή υποστηρίζουμε πως πρέπει να προχωρήσουμε

# Υπολογισμός μεγέθους δείγματος

Ίσοι υπό-πληθυσμοί

$$N = \frac{2\hat{p}(1 - \hat{p})(z_{1-\beta} - z_{\frac{\alpha}{2}})^2}{(\hat{p}_\pi - \hat{p}_{\varepsilon\lambda})^2}$$

- <http://powerandsamplesize.com>
- <https://www.evanmiller.org/ab-testing/sample-size.html>

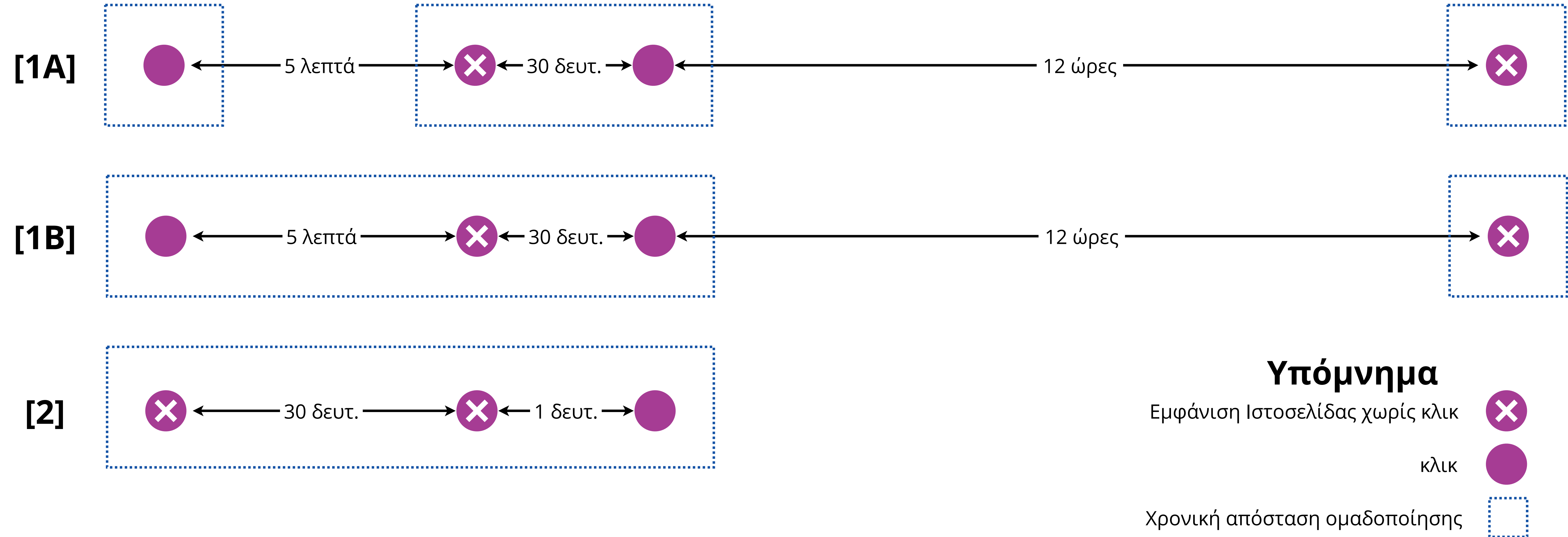
# Ερωτήσεις εκμάθησης

## Μέγεθος δείγματος

$$N = \frac{2\hat{p}(1 - \hat{p})(z_{1-\beta} - z_{\frac{\alpha}{2}})^2}{(\hat{p}_{\pi} - \hat{p}_{\varepsilon\lambda})^2}$$

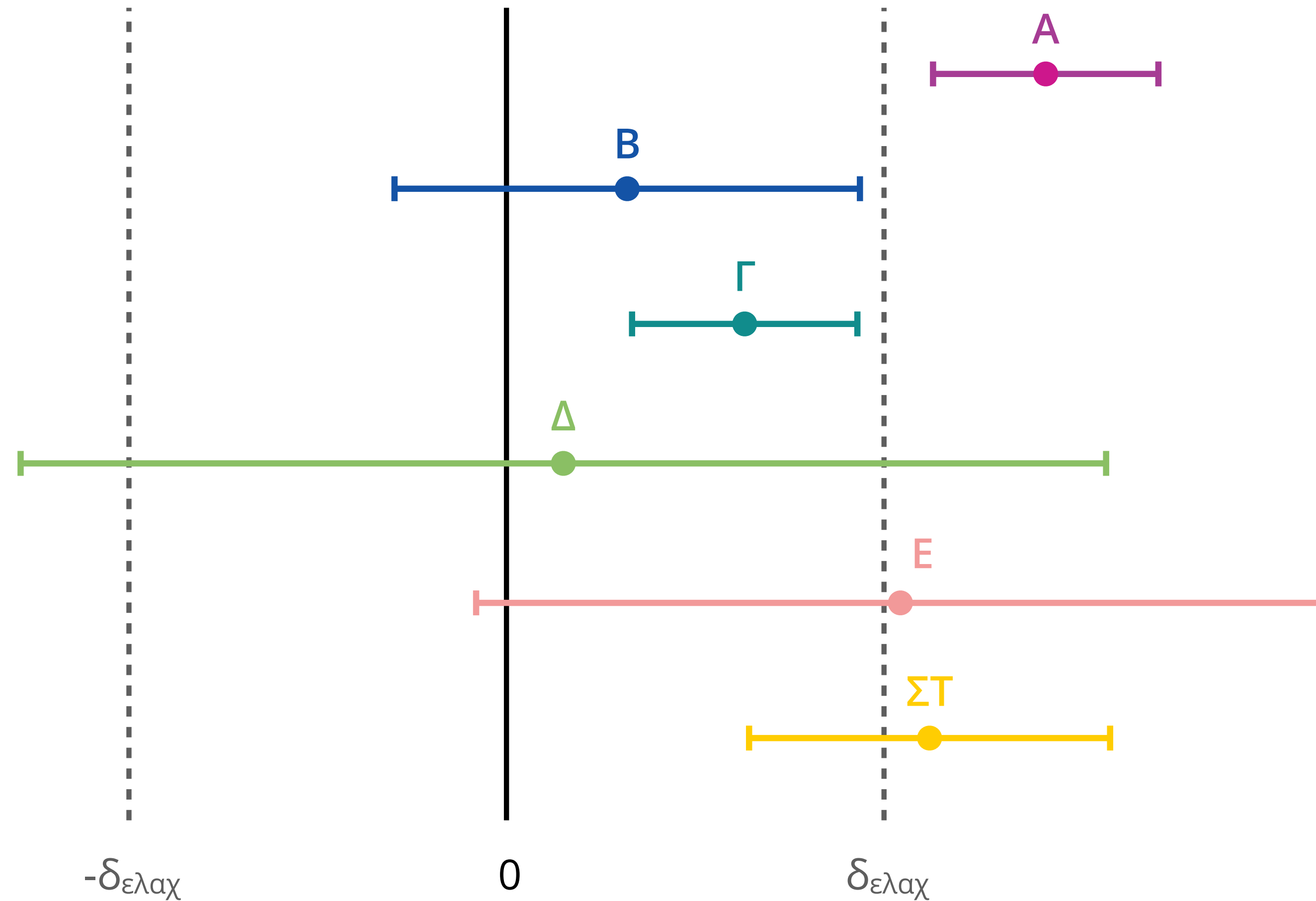
- Αν η αρχική πιθανότητα κλικ (πριν την αλλαγή), είναι μεγαλύτερη, θα χρειαζόμαστε περισσότερους ή λιγότερους επισκέπτες για να έχουμε το ίδιο επίπεδο ευαισθησίας;
- Αν θέλουμε η αλλαγή που σχεδιάζουμε να έχει μία διαφορά τουλάχιστον 5% για να έχουμε επιχειρηματική ουσία, θέλουμε μεγαλύτερο ή μικρότερο δείγμα (όλες οι άλλες παράμετροι παραμένουν ίδιες);
- Αν θέλουμε να μειώσουμε το επίπεδο σημαντικότητας στο 1% (δηλ. επίπεδο εμπιστοσύνης 99%), κρατώντας όλες τις άλλες παραμέτρους ίδιες, θα χρειαζόμαστε μεγαλύτερο ή μικρότερο δείγμα;

# Λειτουργικοποίηση Μέτρου Επιτυχίας





# Η επιχειρηματική απόφαση με βάση το πείραμα



# Παράγοντες που επιδρούν

## Επιδράσεις ωρίμανσης

Οι άνθρωποι μαθαίνουν όσο περνάει ο καιρός ή μπορεί να επέλθει κόπωση με την πάροδο του χρόνου

## Διαδικασίες μέτρησης

Η ίδια η δοκιμασία προ-έλεγχου μπορεί να έκανε τα υποκείμενα περισσότερο ευαίσθητα

## Εργαλεία μέτρησης

Αλλαγές στα εργαλεία μέτρησης, όπως για παράδειγμα στους παρατηρητές, ή στα εργαλεία καταγραφής επιδόσεων

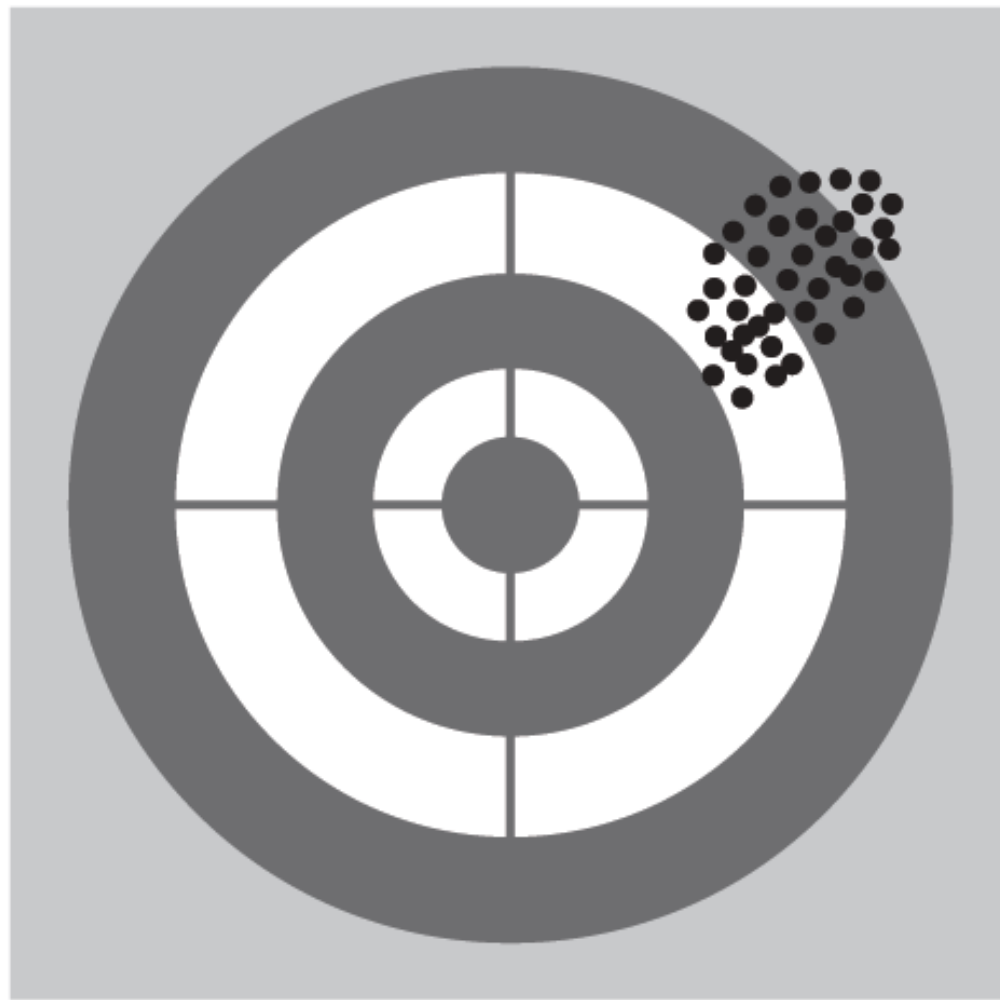
## Πειραματική θνησιμότητα

Η απώλεια των υποκειμένων μιας ομάδας σε σχέση με την άλλη, για παράδειγμα λόγω απουσίας, ασθένειας, ή εγκατάλειψης

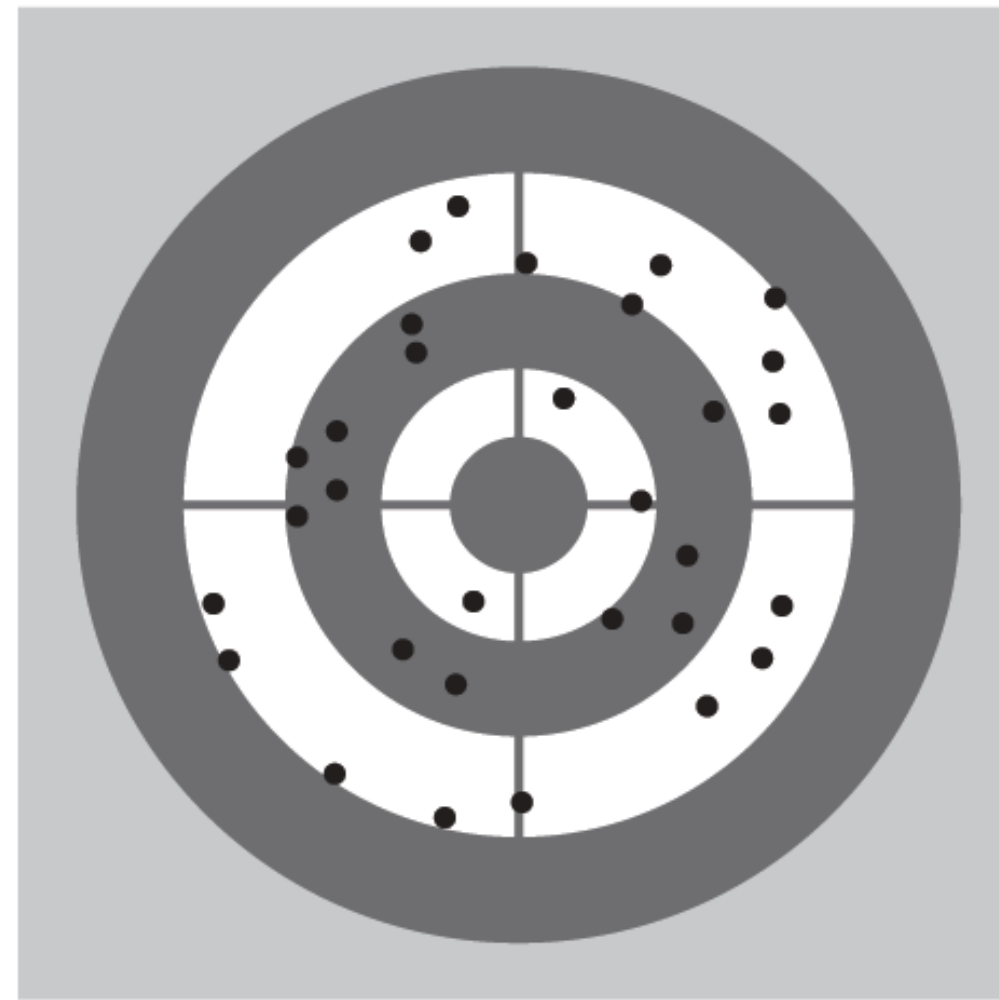
## Μεταβλητές περιβάλλοντος

Επηρεάζουν τα αποτελέσματα, ειδικά αν έχει μεσολαβήσει μεγάλο χρονικό διάστημα ανάμεσα στον προ-έλεγχο και τον μετα-έλεγχο

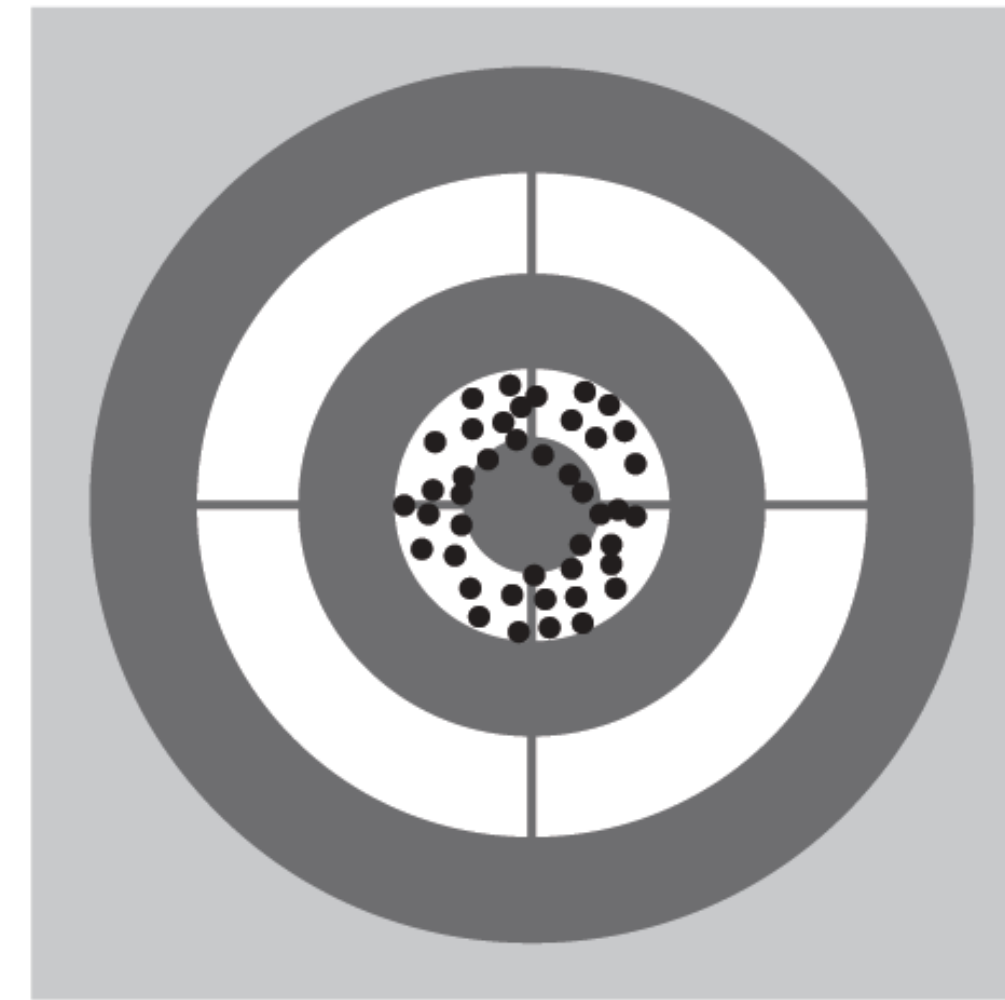
# Εγκυρότητα και Αξιοπιστία



Αξιόπιστο, αλλά όχι έγκυρο

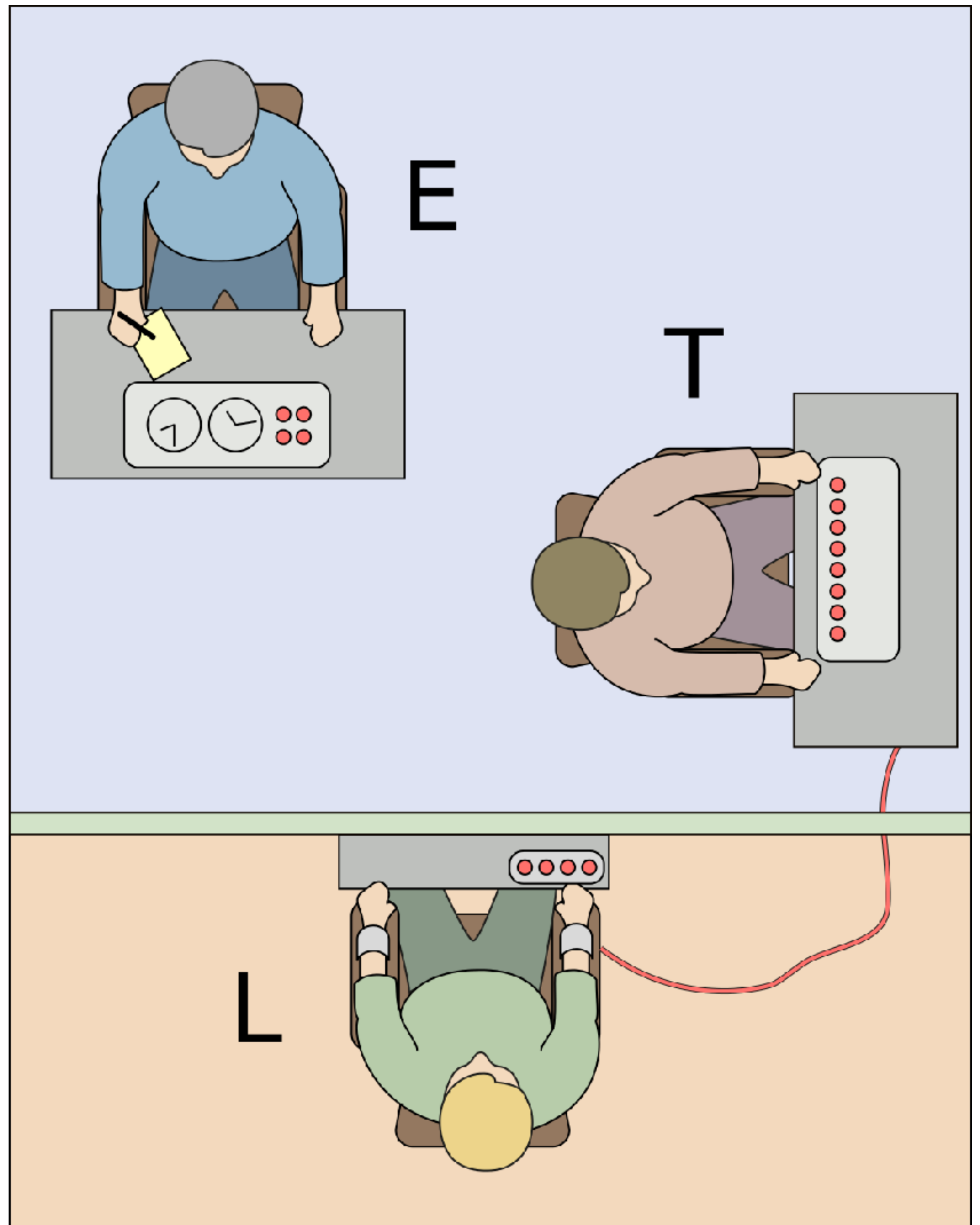


Έγκυρο, αλλά όχι αξιόπιστο



Έγκυρο και αξιόπιστο

# Το πείραμα του Milgram



Picture By Fred the Oyster, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=35182994>