



Στατιστική Ι

Άγγελος Μάρκος, Καθηγητής Ανάλυσης Δεδομένων στις Κοινωνικές Επιστήμες
Τμήμα Ψυχολογίας, Δημοκρίτειο Πανεπιστήμιο Θράκης

Πληροφορίες για το μάθημα

Διαλέξεις: Παρασκευή 9:00 – 12:00

Εργαστήριο; Παρουσίες;

Πώς θα εξεταστώ:

- Δοκιμασίες μέσω eclass (3 μονάδες)
- Τελική γραπτή εξέταση (7 μονάδες)

Από που να διαβάσω:

- Διαφάνειες του μαθήματος (pdf)
- Ηλεκτρονικό Εγχειρίδιο



<http://www.amarkos.gr/statbook/>

Πληροφορίες για το μάθημα

Ποιο σύγγραμμα να πάρω:

Εισαγωγή στις Στατιστικές Έρευνες
(Tintle, Chance, Cobb και συν.)

Ώρες συνεργασίας: Πέμπτη 16:30 – 18:00, Παρασκευή 16:30 – 18:00

email: amarkos@eled.duth.gr

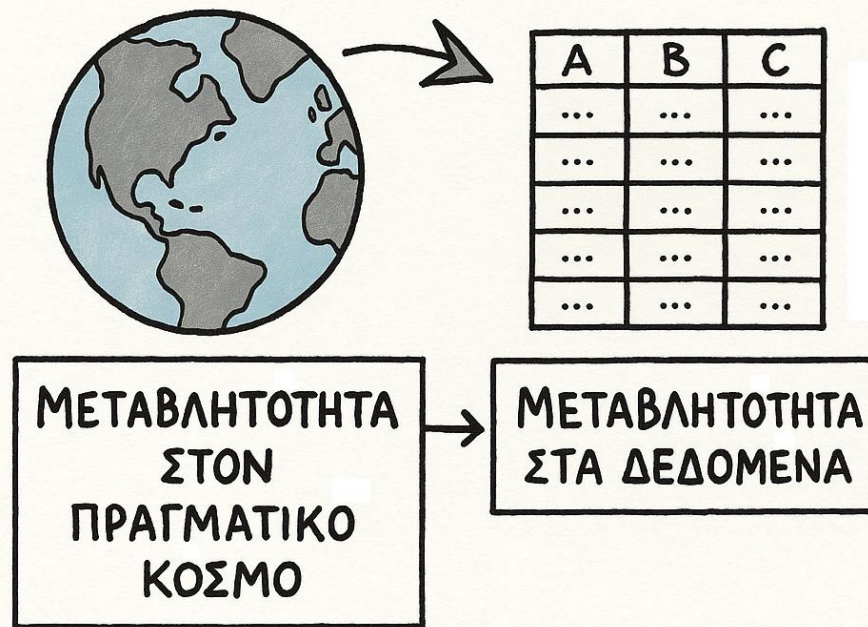


Τι είναι η Στατιστική;

Η Στατιστική είναι η επιστήμη που μελετά τη μεταβλητότητα

Περιλαμβάνει ένα σύνολο εργαλείων και εννοιών που έχουν εξελιχθεί διαχρονικά για να μας βοηθήσουν να κατανοήσουμε τη μεταβλητότητα

Χωρίς μεταβλητότητα γύρω μας, η Στατιστική θα ήταν περιττή



Τι είναι η Στατιστική;

Η Στατιστική είναι η επιστήμη που
μελετά τη μεταβλητότητα

Περιλαμβάνει ένα σύνολο εργαλείων και
εννοιών που έχουν εξελιχθεί διαχρονικά για
να μας βοηθήσουν να κατανοήσουμε τη
μεταβλητότητα

*Χωρίς μεταβλητότητα γύρω μας, η Στατιστική
θα ήταν περιττή*

Παράδειγμα:

Άτομα που λαμβάνουν μια θεραπεία
βελτιώνονται, ενώ άλλα όχι. Κάποιοι
που δεν τη λαμβάνουν αναρρώνουν
ούτως ή άλλως.

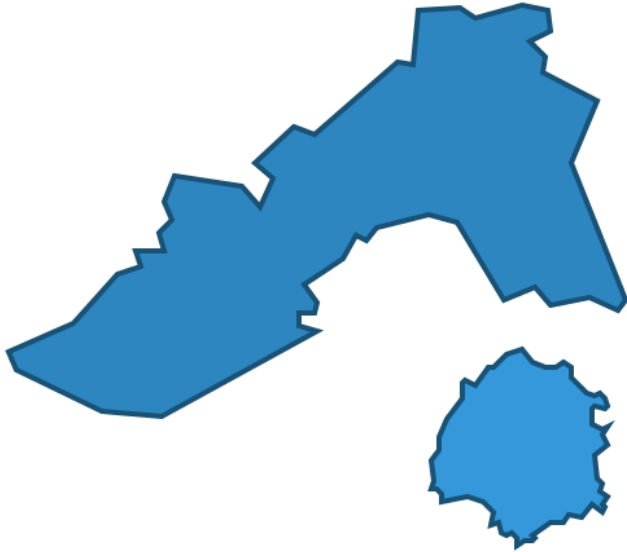
**Η Στατιστική μας βοηθά να
διακρίνουμε αν η θεραπεία
θεραπεύει πραγματικά!**

Το Στατιστικό Μοντέλο

Η βασική ιδέα που ενσωματώνει όλες τις έννοιες της στατιστικής

- 1 Κατανόηση:** Μας βοηθά να κατανοήσουμε μοτίβα στα δεδομένα και τη Διαδικασία Παραγωγής των Δεδομένων (ΔΠΔ)
- 2 Πρόβλεψη:** Μας βοηθά να προβλέψουμε το μέλλον (π.χ. το Netflix προτείνει ταινίες που ενδέχεται να μας αρέσουν)
- 3 Βελτιστοποίηση:** Βοηθά στη βελτιστοποίηση πολύπλοκων συστημάτων (π.χ. μείωση χρόνου αναμονής σε νοσοκομεία)

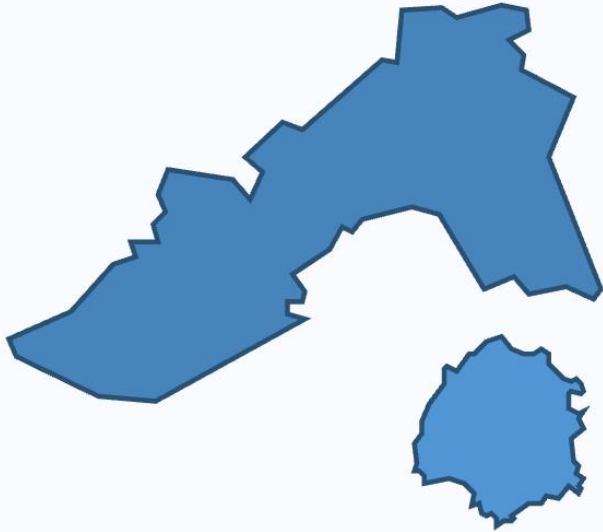
Ένα Απλό Μοντέλο



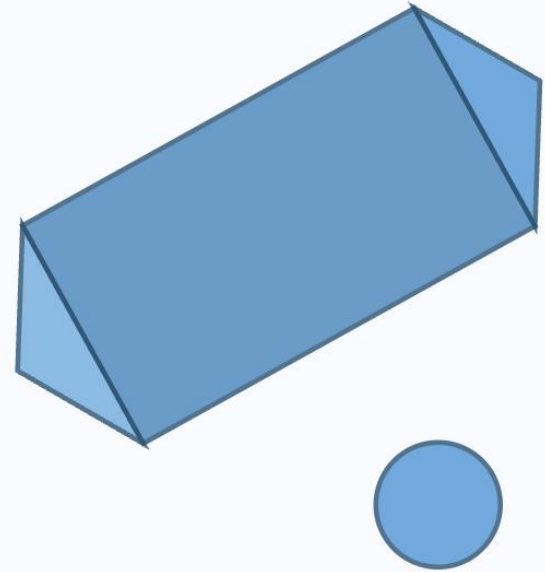
- Ας υποθέσουμε ότι θέλετε να εκτιμήσετε το εμβαδόν σε τετραγωνικά χιλιόμετρα του Νομού Καβάλας.

Ένα Απλό Μοντέλο

Αρχικός Χάρτης



Γεωμετρική Προσέγγιση



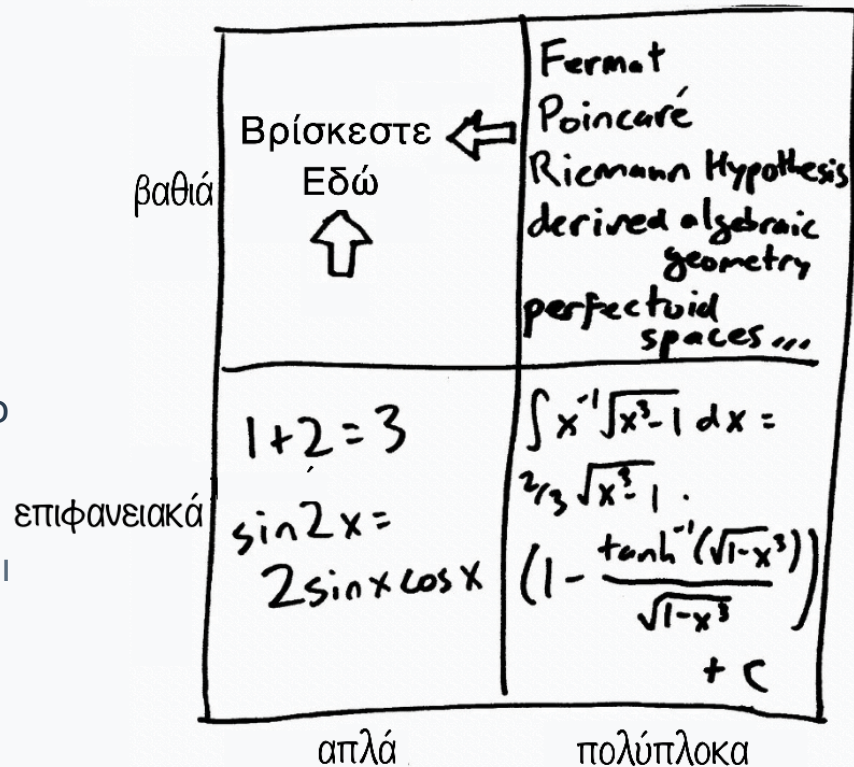
Απλά και σε Βάθος

Δύο Διαστάσεις (Ellenberg, 2016):

- Από τα **απλά** στα **πολύπλοκα**
- Από τα **επιφανειακά** στα **βαθιά**

Η στατιστική που θα μελετήσουμε εμπίπτει στο τεταρτημόριο: **απλά και σε βάθος**

Οι ιδέες είναι αρκετά απλές, αλλά είναι ισχυρές και μπορούν να αλλάξουν ριζικά τον τρόπο που σκέφτεστε!



Ο Συλλογισμός Πιο Σημαντικός από τους Υπολογισμούς

Τα μαθηματικά πίσω από τη βασική στατιστική είναι **απλά**:

Πρόσθεση, αφαίρεση, πολλαπλασιασμός, διαίρεση που επαναλαμβάνονται

Οι υπολογισμοί είναι τετριμμένοι, αλλά χρονοβόροι. Γι' αυτό χρησιμοποιούμε υπολογιστές!

Η στατιστική δεν έχει να κάνει με τον υπολογισμό, αλλά με τον συλλογισμό

Το δύσκολο είναι να καταλάβετε πώς όλα συνδέονται μεταξύ τους

Η κατανόηση σε αυτό το μάθημα

1. Η Κατανόηση Καλλιεργείται με Εξάσκηση

Όπως το παίξιμο πιάνου, η κατανόηση βελτιώνεται μέσω της εξάσκησης στην κριτική σκέψη

2. Η Σύγχυση Είναι Μέρος της Μαθησιακής Διαδικασίας

Η σύγχυση δεν σημαίνει έλλειψη ικανοτήτων - είναι φυσιολογική στην αρχή

3. Η Κατανόηση Απαιτεί Χρόνο

Δεν είναι στιγμιαία αποκάλυψη - αναπτύσσεται σταδιακά με υπομονή

Θα μάθουμε στατιστική με την R

Γιατί με την R;

Δύο Βασικοί Λόγοι:



Είναι δωρεάν

Ανοικτού κώδικα και προσβάσιμη σε όλους



Είναι γλώσσα προγραμματισμού

Βοηθά στην κατανόηση της στατιστικής

Bonus: Νέα δεξιότητα για το CV σας!

Πλεονεκτήματα:

- Πολλαπλές αναπαραστάσεις: λέξεις, διαγράμματα, κώδικας
- Δημιουργία συνδέσεων μεταξύ αναπαραστάσεων → Εμβάθυνση της κατανόησης

Δεν είναι απαραίτητη η εγκατάσταση στον υπολογιστή σας!

Μαθαίνοντας R

Παραδείγματα Κώδικα:

```
print("Hello world!")
```

```
sum(1, 5, 10)
```

```
my_number <- 20
```

Η R είναι απαιτητική - προσοχή στα κεφαλαία/πεζά!

Κουλτούρα Προγραμματισμού:

Δοκιμή και Σφάλμα

- Πειραματιστείτε
- Μάθετε από τα λάθη
- Εξερευνήστε λύσεις

Η απογοήτευση είναι μέρος της διαδικασίας!

Στόχοι του Μαθήματος

1. Επάρκεια στην Ανάλυση Δεδομένων με την R

Απόκτηση εμπειρίας στην ανάλυση δεδομένων

2. Κατανόηση Στατιστικών Εννοιών

Στέρεα βάση για ερμηνεία και πλαισίωση αναλύσεων

3. Εφόδια για Συνεχή Μάθηση

Βάση για περαιτέρω γνώση πέρα από το μάθημα

Οποιοσδήποτε μπορεί να μάθει στατιστική

Όταν κάτι φαίνεται δύσκολο, αυτό σημαίνει ότι καλείστε να
αντιμετωπίσετε απαιτητικές έννοιες

Όχι ότι δεν είστε ικανοί/ικανές να τις κατακτήσετε!

Αναγνώριση Μοτίβων σε Αριθμούς

Χωρίς μεταβλητότητα:

2, 2, 2, 2, 2, 2, 2, 2, 2

Απλή περιγραφή: «Εννέα φορές ο αριθμός 2»

Με μεταβλητότητα:

2, 1, 3, 3, 2, 3, 1, 2, 1

Χρειάζονται στατιστικές μέθοδοι!

Δύο Βασικές Μέθοδοι:

- Ταξινόμηση - `sort()`
- Πίνακες Συχνοτήτων - `table()`

Παράδειγμα:

Ταξινομημένα: 1, 1, 1, 2, 2, 2, 3, 3, 3

Σε πίνακα συχνοτήτων:

1 2 3

3 3 3

Από τους Αριθμούς στα Δεδομένα

Δεδομένα (data): Αριθμοί ή κατηγορίες που αντικατοπτρίζουν πτυχές του πραγματικού κόσμου

Δύο Θεμελιώδεις Διαδικασίες:

- **Μέτρηση:** Αναπαράσταση χαρακτηριστικών με αριθμούς ή κατηγορίες
- **Δειγματοληψία:** Επιλογή αντικειμένων για μελέτη

Παράδειγμα: Ύψος Ανθρώπων

Μετρούμε το ύψος σε εκατοστά για να μετατρέψουμε τη μεταβλητότητα σε αριθμούς που μπορούμε να αναλύσουμε

Οργάνωση Δεδομένων

Πλαίσιο Δεδομένων (Data Frame)

Πίνακας με γραμμές και στήλες

- **Γραμμές:** Παρατηρήσεις (άτομα, αντικείμενα)
- **Στήλες:** Μεταβλητές (χαρακτηριστικά)

Κάθε κελί περιέχει την τιμή μιας παρατήρησης σε μια μεταβλητή

	Method	Age	PreScore	PostScore
1	Traditional	20	68	72
2	Traditional	19	75	78
3	Interactive	21	66	82
4	Interactive	18	71	89
5	Traditional	20	73	74
6	Interactive	19	69	85

Χρήσιμες Εντολές R:

- `head()` - Πρώτες γραμμές
- `tail()` - Τελευταίες γραμμές
- `str()` - Δομή δεδομένων
- `table()` - Πίνακας συχνοτήτων
- `arrange()` - Ταξινόμηση

Σύμβολο \$: Πρόσβαση σε μεταβλητή

`TeachingMethods$Age`

Είδη Μεταβλητών

Ποσοτικές

Αναπαριστούν ποσότητες

- Ύψος (εκατοστά)
- Αριθμός μελών οικογένειας
- Βαθμολογία (0-100)

Οι διαφορές έχουν νόημα

Ποιοτικές / Κατηγορικές

Αναπαριστούν κατηγορίες

- Φύλο (1=άνδρας, 2=γυναίκα)
- Μέθοδος διδασκαλίας
- Χρώμα ματιών

Οι αριθμοί είναι αυθαίρετοι

Αντικείμενα τύπου Factor στην R

Η R χρειάζεται βοήθεια να διακρίνει ποιοτικές από ποσοτικές μεταβλητές

Δήλωση ποιοτικής μεταβλητής:

```
factor(Fingers$Gender)
```

```
as.numeric() # Μετατροπή σε αριθμητική
```

Προσοχή: Να μην αθροίζετε κατηγορικές μεταβλητές! (Δεν έχει νόημα)

Σφάλμα Μέτρησης

Οι μετρήσεις συνήθως εμπεριέχουν σφάλμα

Αμερόληπτο Σφάλμα

Τυχαίες αποκλίσεις προς τα πάνω ή προς τα κάτω

Μεροληπτικό Σφάλμα

Συστηματικές αποκλίσεις προς μία κατεύθυνση

Παράδειγμα:

Δύο άτομα μετρούν έναν αντίχειρα:

- Άτομο 1: 56mm
- Άτομο 2: 57mm

Το πραγματικό μήκος είναι κάπου ανάμεσα

Όσο πιο δύσκολη η μέτρηση, τόσο μεγαλύτερο το σφάλμα

Δειγματοληψία από Πληθυσμό

Δύο Κρίσιμες Αποφάσεις:

- **Δείγμα:** Ποιες παρατηρήσεις θα συμπεριληφθούν;
- **Δειγματοληψία:** Πώς θα επιλεγούν από τον πληθυσμό;

Τυχαία Δειγματοληψία (Random)

Κάθε μέλος του πληθυσμού έχει την ίδια πιθανότητα επιλογής

Ανεξαρτησία (Independence)

Η επιλογή μιας παρατήρησης δεν επηρεάζει την επιλογή άλλης

Μεταβλητότητα Δειγματοληψίας

Σφάλμα Δειγματοληψίας (Sampling Error):

Κάθε δείγμα διαφέρει από τον πληθυσμό και από άλλα δείγματα

- Κανένα δείγμα δεν αντιπροσωπεύει τέλεια τον πληθυσμό
- Διαφορετικά δείγματα → Διαφορετικά αποτελέσματα
- Μπορεί να είναι αμερόληπτο ή μεροληπτικό

Σημαντικό: Τα τυχαία δείγματα είναι αμερόληπτα - δεν είναι συστηματικά λανθασμένα προς μία κατεύθυνση

Στην πράξη: Μη τυχαία δειγματοληψία, στρωματοποιημένη, συστηματική και δειγματοληψία κατά συστάδες

Βασικά Σημεία

Μεταβλητότητα $\rightarrow$ Δεδομένα $\rightarrow$ Ανάλυση

Μέτρηση + Δειγματοληψία = Δεδομένα

**Κατανοώντας τα δεδομένα, κατανοούμε καλύτερα
τον κόσμο!**