

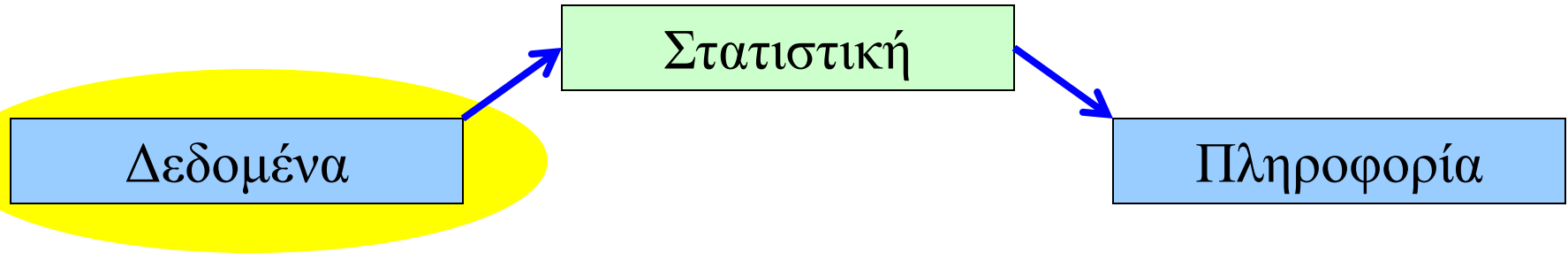
# ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΣΤΑΤΙΣΤΙΚΗ ΤΩΝ ΕΠΙΧΕΙΡΗΣΕΩΝ

## Κεφάλαιο 5 Συλλογή Δεδομένων & Δειγματοληψία

*Επιμέλεια παρουσιάσεων: Δρ. Αλέκα Καλαπόδη*

# Ας θυμηθούμε ...

Η Στατιστική είναι ένα εργαλείο για τη μετατροπή *δεδομένων* σε *πληροφορία*:



Αλλά από πού προέρχονται τα *δεδομένα*; Πώς έγινε η συλλογή τους; Πώς διασφαλίζουμε ότι είναι ακριβή; Είναι τα δεδομένα αξιόπιστα; Είναι αντιπροσωπευτικά του πληθυσμού από τον οποίο έχουν εξαχθεί; Το κεφάλαιο αυτό εξετάζει μερικά από αυτά τα ζητήματα.

# Μέθοδοι Συλλογής Δεδομένων ...

Υπάρχουν πολλές μέθοδοι που χρησιμοποιούνται για τη συλλογή ή την απόκτηση δεδομένων για στατιστική ανάλυση.

Τρεις από τις πιο δημοφιλείς μεθόδους είναι οι ακόλουθες:

- Άμεση παρατήρηση (π.χ. αριθμός πελατών που μπαίνουν σε μια τράπεζα ανά ώρα)
- Πειράματα (π.χ. νέοι τρόποι παραγωγής αγαθών για να ελαχιστοποιηθεί το κόστος)
- Δημοσκοπήσεις

# Δημοσκοπήσεις ...

---

Μια *δημοσκόπηση* ζητά πληροφορίες από τον κόσμο.  
Π.χ. γκάλοπ, προεκλογικές δημοσκοπήσεις, έρευνες αγοράς.

Το *Ποσοστό Απαντήσεων* (δηλ. το ποσοστό αυτών που απάντησαν από το σύνολο αυτών που ερωτήθηκαν) είναι μια κομβική παράμετρος της έρευνας.

Οι δημοσκοπήσεις μπορούν να διεξαχθούν με πολλούς τρόπους, όπως:

- Προσωπική συνέντευξη
- Τηλεφωνική συνέντευξη
- Ερωτηματολόγιο

# Σχεδίαση Ερωτηματολογίου ...

Τα τελευταία χρόνια, έχει αναπτυχθεί αρκετά η επιστήμη σχεδίασης ερωτηματολογίων. Βασικές αρχές σχεδίασης:

1. Το ερωτηματολόγιο να είναι όσο το δυνατόν πιο σύντομο.
2. Οι ερωτήσεις να είναι σύντομες, απλές και σαφείς.
3. Να αρχίζει με δημογραφικές ερωτήσεις ώστε να βοηθάει τους ερωτώμενους να ξεκινήσουν άνετα τις απαντήσεις.
4. Να χρησιμοποιούνται διχοτομικές (ναι|όχι) και πολλαπλής επιλογής ερωτήσεις.
5. Να χρησιμοποιούνται με προσοχή ερωτήσεις ανοιχτού τύπου.
6. Να αποφεύγονται ερωτήσεις που καθοδηγούν σε συγκεκριμένες απαντήσεις.
7. Να δοκιμάζετε το ερωτηματολόγιο σε μια μικρή ομάδα ανθρώπων.
8. Να έχετε σκεφτεί τον τρόπο επεξεργασίας των δεδομένων που θα συλλέξετε, όταν ετοιμάζετε το ερωτηματολόγιο.

# Δειγματοληψία ...

Η επαγωγική στατιστική μας επιτρέπει να εξάγουμε συμπεράσματα για έναν πληθυσμό εξετάζοντας ένα δείγμα του.

Δειγματοληψία (δηλ. επιλογή ενός υποσυνόλου του πληθυσμού) γίνεται συχνά για λόγους **κόστους** (είναι οικονομικότερο να εξετάσεις 1.000 τηλεθεατές από ότι 100 εκατομμύρια τηλεθεατές) και **πρακτικότητας** (είναι αδύνατο να γίνει δοκιμή πρόσκρουσης σε κάθε παραγόμενο αυτοκίνητο).

Σε κάθε περίπτωση, ο *πληθυσμός του δείγματος* και ο *πληθυσμός-στόχος* θα πρέπει να **είναι όμοιοι**.

# Σχέδια Δειγματοληψίας ...

---

Ένα *σχέδιο δειγματοληψίας* είναι απλά μια μέθοδος επιλογής δείγματος από έναν πληθυσμό.

Θα επικεντρωθούμε στις ακόλουθες τρεις μεθόδους:

- Απλή τυχαία δειγματοληψία
- Κατηγοριοποιημένη τυχαία δειγματοληψία
- Ομαδοποιημένη δειγματοληψία

# Απλή Τυχαία Δειγματοληψία ...

Ένα *απλό τυχαίο δείγμα* είναι ένα δείγμα που επιλέγεται έτσι ώστε κάθε δυνατό δείγμα ίδιου μεγέθους να έχει την ίδια πιθανότητα επιλογής.

Η εξαγωγή τριών ονομάτων από ένα καπέλο που περιέχει όλα τα ονόματα των φοιτητών του τμήματος είναι ένα παράδειγμα απλού τυχαίου δείγματος: κάθε ομάδα τριών ονομάτων έχει ίση πιθανότητα επιλογής με οποιαδήποτε άλλη ομάδα τριών ονομάτων.



# Απλή Τυχαία Δειγματοληψία ...

Παράδειγμα 5.1: Ένας ελεγκτής της εφορίας πρέπει να επιλέξει από 1000 δηλώσεις ένα δείγμα 40 δηλώσεων για έλεγχο ...

Random Number Generation

Input

Number of variables: 1

Number of random numbers: 50

Random seed: |

Distribution: Uniform

Between 1 and 1000

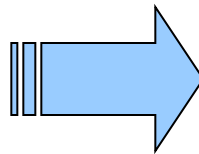
Output Options

Output range: \$B\$1

New worksheet ply:

New workbook

Help Cancel OK



	A	B	C
1		<b>Random #:</b>	<b>Rounded Up:</b>
2		800.791	801
3		655.516	656
4		305.514	306
5		675.303	676
6		107.647	108
7		517.070	518
8		800.857	801
9		602.863	603
10		370.575	371
11		257.404	258
12		374.813	375
13		825.761	826
14		173.532	174
15		298.502	299

Μπορούν να χρησιμοποιηθούν επιπλέον # εάν παραχθούν ίδιοι τυχαίοι αριθμοί

# Κατηγοριοποιημένη Τυχαία Δειγματοληψία ...

Ένα *κατηγοριοποιημένο τυχαίο δείγμα* λαμβάνεται διαχωρίζοντας τον πληθυσμό σε ξένα μεταξύ τους σύνολα, ή αλλιώς στρώματα, και επιλέγοντας στη συνέχεια απλά τυχαία δείγματα από κάθε στρώμα.

Στρώμα 1 : Φύλο

Άνδρες

Γυναίκες

Στρώμα 2 : Ηλικία

< 20

20-30

31-40

41-50

51-60

> 60

Στρώμα 3 : Επάγγελμα

ελεύθ. επαγγελματίας

υπάλληλος

εργάτης

άλλο

Μπορούμε να εξαγάγουμε συμπεράσματα για το συνολικό πληθυσμό, χρησιμοποιώντας στοιχεία **μέσα από ένα στρώμα** ή κάνοντας συγκρίσεις **μεταξύ των στρωμάτων**

# Κατηγοριοποιημένη Τυχαία Δειγματοληψία ...

Μετά την κατηγοριοποίηση του πληθυσμού, μπορούμε να χρησιμοποιήσουμε *απλή τυχαία δειγματοληψία* για να δημιουργήσουμε το πλήρες δείγμα:

Κατηγορία Εισοδήματος	Ποσοστό Πληθυσμού	Μέγεθος Δείγματος	
		n=400	n=1000
under \$25,000	25%	100	250
\$25,000 - \$39,999	40%	160	400
\$40,000 - \$60,000	30%	120	300
over \$60,000	5%	20	50

Εάν οι πόροι μας επαρκούν για δείγμα μεγέθους 400 ατόμων, θα επιλέξουμε 100 από την κατηγορία χαμηλού εισοδήματος ...

... εάν έχουμε δείγμα 1000 ατόμων, θα επιλέξουμε 50 από την κατηγορία υψηλού εισοδήματος.

# Ομαδοποιημένη Δειγματοληψία ...

Ένα *ομαδοποιημένο δείγμα* είναι ένα απλό τυχαίο δείγμα ομάδων ή συστοιχιών (έναντι ενός απλού τυχαίου δείγματος μεμονωμένων αντικειμένων).

Η μέθοδος αυτή χρησιμοποιείται όταν είναι δύσκολο ή ασύμφορο να σχηματιστεί ο πλήρης κατάλογος των μελών του πληθυσμού ή όταν τα μέλη του πληθυσμού είναι διασκορπισμένα σε μεγάλη γεωγραφική έκταση.

# Μέγεθος Δείγματος ...

---

Θα περιγράψουμε στη συνέχεια αριθμητικές τεχνικές για τον καθορισμό του μεγέθους των δειγμάτων, ωστόσο αρκεί να πούμε ότι όσο μεγαλύτερο είναι το δείγμα τόσο πιο αξιόπιστα περιμένουμε να είναι τα αποτελέσματα.

# Σφάλματα Δειγματοληψίας και Δεδομένων ...

Δύο κύριοι τύποι σφαλμάτων εμφανίζονται όταν ένα δείγμα παρατηρήσεων εξάγεται από έναν πληθυσμό:

*σφάλμα δειγματοληψίας και σφάλμα δεδομένων.*

Το *σφάλμα δειγματοληψίας* αναφέρεται σε διαφορές μεταξύ του δείγματος και του πληθυσμού οι οποίες οφείλονται στις τυχαίες παρατηρήσεις που έτυχε να εξαχθούν από το δείγμα.

Αυξάνοντας το μέγεθος του δείγματος, μειώνεται αυτό το σφάλμα.

# Σφάλματα Δεδομένων ...

---

Τα *σφάλματα δεδομένων* είναι πιο σοβαρά και οφείλονται σε λάθη που έγιναν στην καταγραφή των δεδομένων ή σε παρατηρήσεις που λανθασμένα επιλέχθηκαν.

Τρεις τύποι σφαλμάτων δεδομένων:

Σφάλματα καταγραφής

Σφάλματα μη απάντησης

Σφάλματα επιλογής

Σημείωση: αυξάνοντας το μέγεθος του δείγματος **δεν** μειώνονται τα σφάλματα δεδομένων.

# Σφάλματα καταγραφής ...

... προκύπτουν από την καταγραφή λάθος απαντήσεων, που οφείλονται σε:

- λανθασμένες μετρήσεις εξαιτίας ελαττωματικού εξοπλισμού,
- λάθη στη διαδικασία μεταγραφής δεδομένων από πρωτογενείς πηγές,
- ανακριβή καταγραφή δεδομένων οφειλόμενη σε παρερμηνεία των όρων
- ανακριβείς απαντήσεις σε ερωτήσεις σχετικές με ευαίσθητα θέματα.



# Σφάλματα μη απάντησης ...

... αναφέρονται σε σφάλματα που δημιουργούνται όταν κάποια μέλη του δείγματος δεν δίνουν απαντήσεις, δηλ. οι παρατηρήσεις του δείγματος που έχουμε μπορεί να μην είναι αντιπροσωπευτικές για το σύνολο του πληθυσμού.

Όπως είπαμε και νωρίτερα, το *Ποσοστό Απαντήσεων* (δηλ. το ποσοστό αυτών που απάντησαν από το σύνολο αυτών που ερωτήθηκαν) είναι μια κομβική παράμετρος της έρευνας και βοηθάει στην κατανόηση όσον αφορά την εγκυρότητα της έρευνας και των πηγών των σφαλμάτων μη απάντησης.

# Σφάλματα επιλογής ...

... εμφανίζονται όταν το σχέδιο δειγματοληψίας είναι τέτοιο ώστε κάποια μέλη του πληθυσμού να είναι αδύνατον να συμπεριληφθούν στο δείγμα.