



Εισαγωγή στην Στατιστική και στην Ανάλυση Δεδομένων

Κωνσταντίνος Μπουγιούκας, MSc, PhD
mpougioukas@auth.gr

2024-2025



Βοηθητικό υλικό του μαθήματος (στα αγγλικά): <https://statsocial.netlify.app/>

Basic Statistics for the Social Sciences

1st Edition

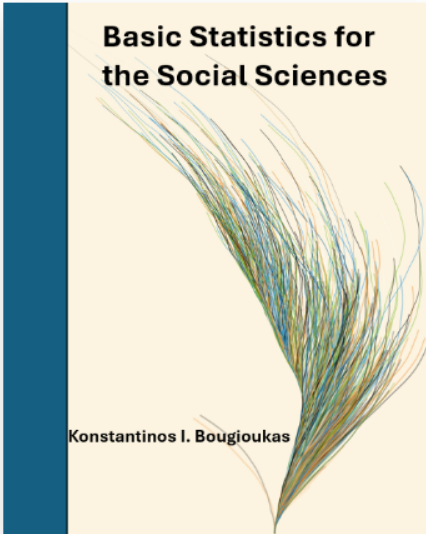
AUTHOR
Konstantinos I. Bougioukas, PhD 

PUBLISHED
October 2, 2024

Welcome!

This open course introduces students to key concepts and statistical methods used in the quantitative social and behavioral sciences to describe and test hypotheses about the social world and human behavior. Students will learn to:

- Describe and summarize data distributions.
- Formulate and test various types of research hypotheses.
- Analyze associations between factors, characteristics, or events.



On this page

Welcome!

Basic Statistics for the Social Sciences

Q

Welcome!

- 1 Introduction
- 2 Summary
- References



Στόχοι του σημερινού μαθήματος

- ☑ Η κατανόηση της **σημασίας** της ανάλυσης δεδομένων στον χώρο της ψυχολογίας και των κοινωνικών επιστημών
- ☑ Ορισμός της Στατιστικής Επιστήμης και διάκριση ανάμεσα στους κλάδους της κλασικής στατιστικής (**περιγραφική** Vs **επαγωγική** στατιστικής)
- ☑ Η ταξινόμηση των **πηγών** από τις οποίες προέρχονται τα δεδομένα
- ☑ Η κατανόηση της έννοιας της **μεταβλητής** και η αναγνώριση του τύπου των μεταβλητών





Γιατί να μάθω βασική Στατιστική;

Η εξοικείωση με τις βασικές αρχές της Στατιστικής, της ανάλυσης δεδομένων και της μεθοδολογίας έρευνας για μια/έναν ψυχολόγο κρίνεται σήμερα αναγκαία για τρεις βασικούς λόγους:

- για να κατανοεί βασικούς **στατιστικούς-κοινωνικούς δείκτες** (π.χ. στατιστικά στοιχεία από ΕΛΣΤΑΤ, EUROSTAT, WHO), να αξιολογεί και να διασταυρώνει στατιστικές πληροφορίες
- για να είναι σε θέση να **μελετάει κριτικά** και να **αξιολογεί** μια έρευνα και τα ευρήματά της
- για να είναι σε θέση να πραγματοποιήσει κάποια έρευνα και να **συνεργαστεί** αποδοτικότερα με στατιστικούς επιστήμονες





Ορίζοντας τη Στατιστική Επιστήμη

Η στατιστική είναι μια **εφαρμοσμένη** μαθηματική επιστήμη που ασχολείται με τη συλλογή, ανάλυση και ερμηνεία αριθμητικών **δεδομένων**.

(Corxton και Cowden)

Συνεπώς θα ασχοληθούμε με την **Ποσοτική Έρευνα**.



Περιγραφική Στατιστική (descriptive statistics)

Ασχολείται με τη συλλογή, την περιγραφή και την οργάνωση των δεδομένων. Σκοπός της είναι να παρουσιάσει και να συνοψίσει τα δεδομένα με τρόπο που να είναι εύκολα κατανοητός, χωρίς να γίνεται προσπάθεια για εξαγωγή γενικότερων συμπερασμάτων.

- Μέτρα κεντρικής θέσης
- Μέτρα διασποράς
- Γραφήματα, πίνακες

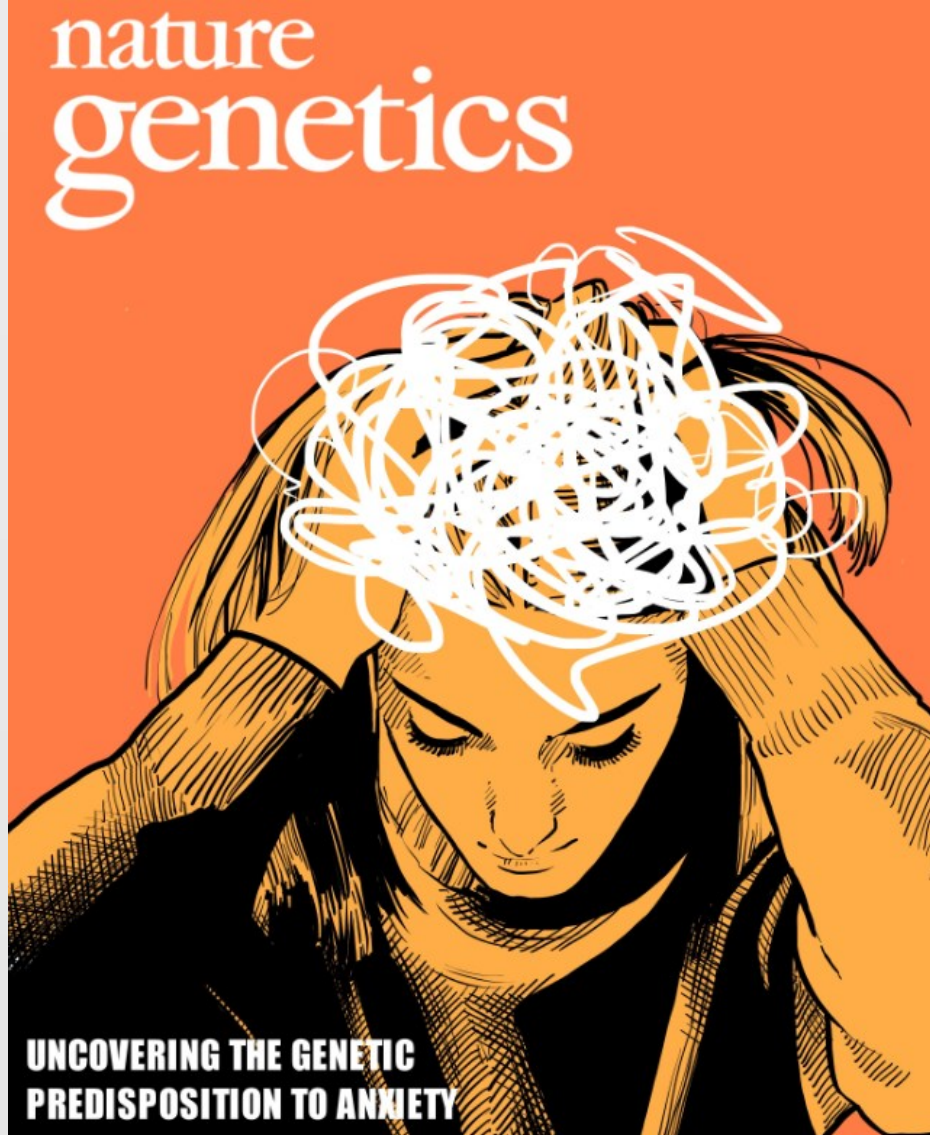
Επαγωγική Στατιστική (inferential statistics)

Μας επιτρέπει να κάνουμε εκτιμήσεις και να επάγουμε συμπεράσματα από το **δείγμα** στον **πληθυσμό**.

- Εκτίμηση παραμέτρων και Διαστήματα Εμπιστοσύνης
- Έλεγχος υποθέσεων
- P-values



Κλασική Στατιστική



Η εποχή των δεδομένων

- Ερευνητική ομάδα από το Πανεπιστήμιο του Yale διερεύνησε το γενετικό προφίλ περισσότερων από **1 εκ. ανθρώπων**.
- Βρέθηκαν περισσότερα από **100 γονίδια** να συσχετίζονται με το **άγχος**.

<https://doi.org/10.1038/s41588-024-01908-2>



Από το ELIZA (1964-7) στο ChatGPT (2022)

Welcome to

EEEEEE	LL	IIII	ZZZZZZ	AAAAA
EE	LL	II	ZZ	AA AA
EEEEEE	LL	II	ZZZ	AAAAAAA
EE	LL	II	ZZ	AA AA
EEEEEE	LLLLLL	IIII	ZZZZZZ	AA AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU: Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU: They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU: Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU: He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU: It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:



ChatGPT

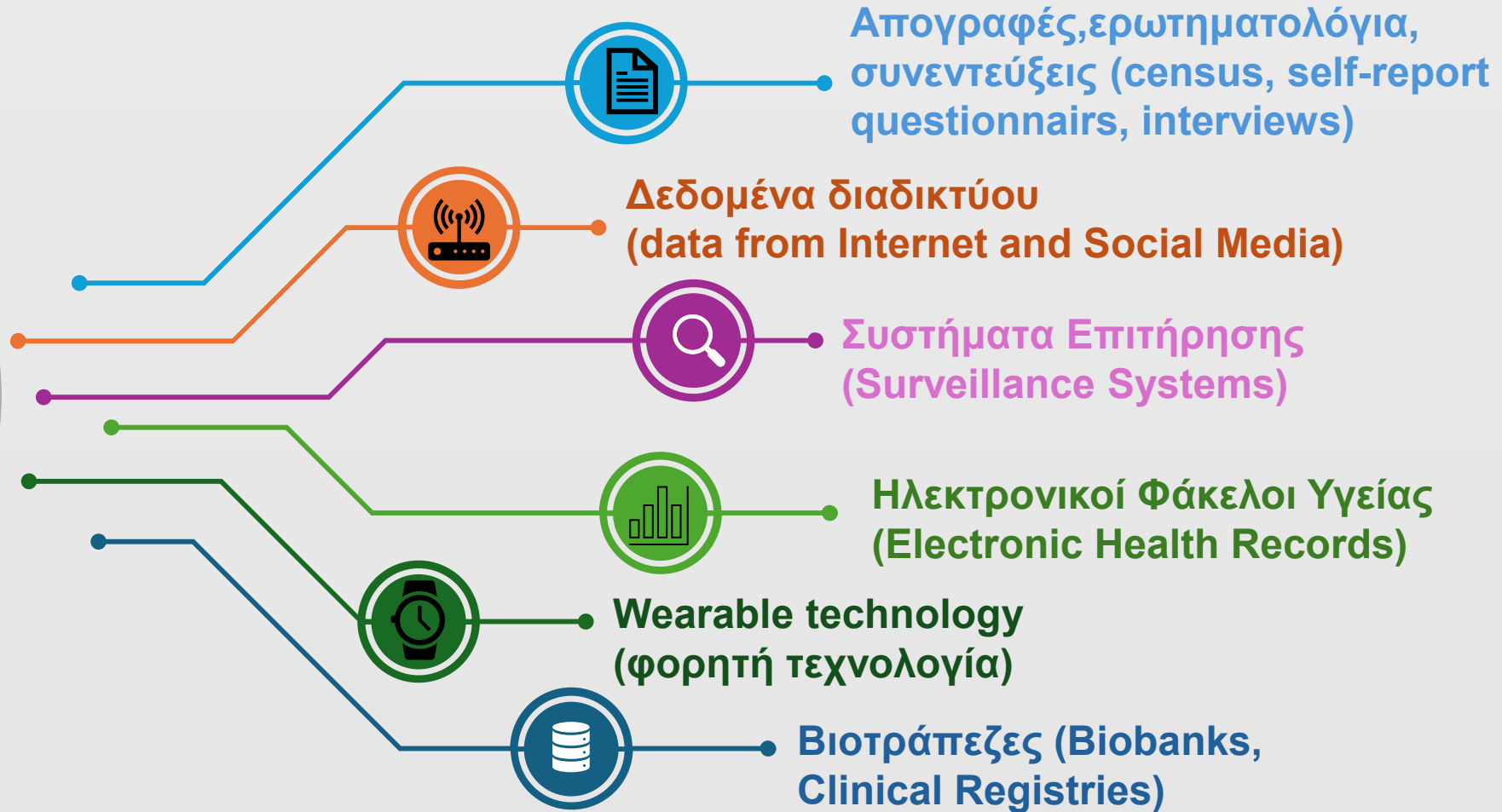
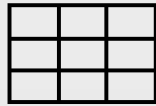
Το πιο διάσημο σενάριο,
“DOCTOR”, προσομοίωσε
έναν ψυχοθεραπευτή της
σχολής Rogerian.



Πηγές δεδομένων

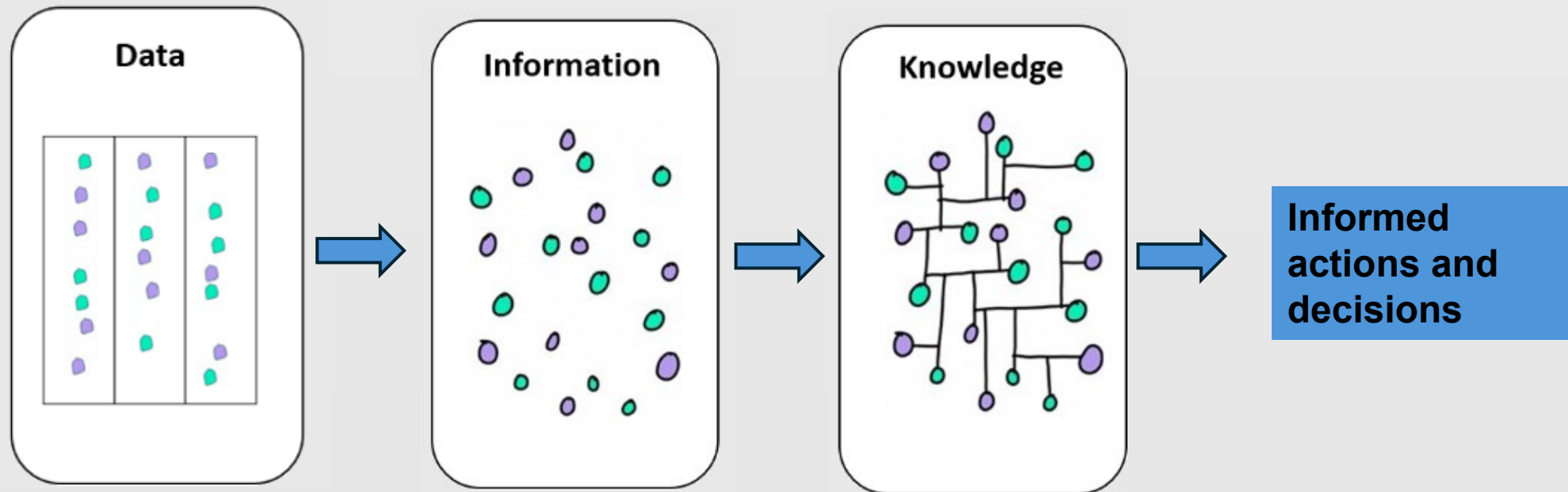
Δεδομένα (Data)

- Δομημένα (Structured)
- Ημι-δομημένα (semi-structured)
- Αδόμητα (Unstructured)





Τα κοινωνικά, συμπεριφορικά και βιοϊατρικά δεδομένα (data) μπορούν να μετατραπούν σε πληροφορία (information). Αυτή η πληροφορία μπορεί να γίνει **γνώση** (knowledge) όταν οι επιστήμονες και τα ενδιαφερόμενα μέρη την κατανοούν, επιτρέποντάς τους να λαμβάνουν ενημερωμένες **αποφάσεις, δράσεις** (actions) και πολιτικές για την σωματική και ψυχική υγεία του πληθυσμού.





Κοινωνική Στατιστική

- Είναι ο τομέας της στατιστικής που εφαρμόζεται στις κοινωνικές επιστήμες για τη μελέτη κοινωνικών φαινομένων και τάσεων.
- Χρησιμοποιεί στατιστικές μεθόδους συλλογής δεδομένων, όπως οι **απογραφές**, οι **κοινωνικές έρευνες**, και τα **διοικητικά αρχεία**, και συχνά χρησιμοποιείται από κυβερνητικούς οργανισμούς, ιδρύματα και ερευνητές για να αναλύσει δεδομένα που σχετίζονται με την κοινωνική ζωή, την ανθρώπινη συμπεριφορά και τη λειτουργία των κοινωνικών θεσμών.
- Ο σκοπός της είναι να παρέχει αντικειμενικά και **ποσοτικά στοιχεία** που βοηθούν στην κατανόηση και την ερμηνεία **κοινωνικών ζητημάτων** με στόχο την διαμόρφωση κατάλληλων **κοινωνικών πολιτικών**.



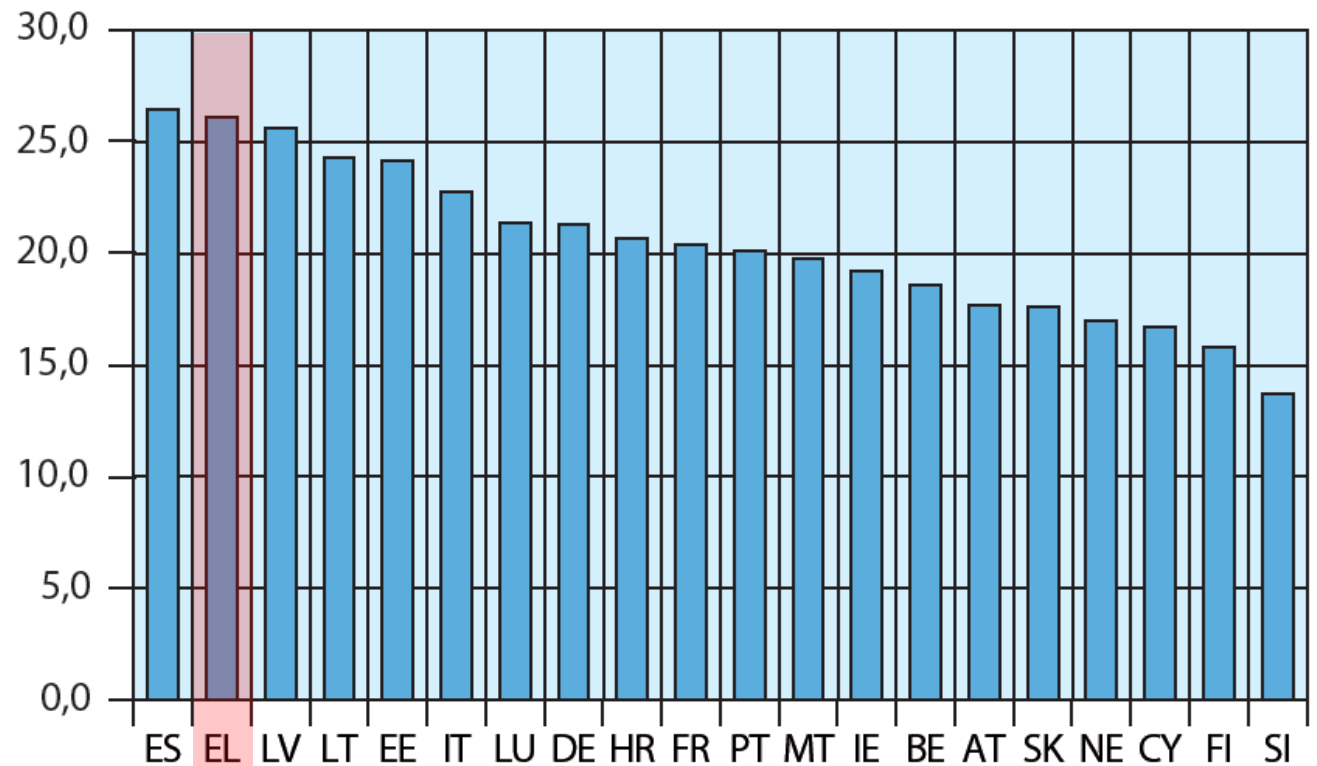
ΕΛΛΗΝΙΚΗ ΣΤΑΤΙΣΤΙΚΗ ΑΡΧΗ ΕΛΛΑΣ ΜΕ ΑΡΙΘΜΟΥΣ

Ιούλιος - Σεπτέμβριος 2024



<https://www.statistics.gr/greece-in-figures> (σελ.131)

Χώρες Ευρωζώνης: Πληθυσμός σε κίνδυνο φτώχειας ή κοινωνικό αποκλεισμό (%), 2023



Το **26.1%** του ελληνικού πληθυσμού βρίσκεται σε κίνδυνο φτώχειας ή κοινωνικού αποκλεισμού.



Global and regional estimates
of violence against women:
**prevalence and health effects of
intimate partner violence and
non-partner sexual violence**



Έκθεση Παγκόσμιου
Οργανισμού Υγείας (ΠΟΥ) για
την βία που υφίστανται οι
γυναίκες (2018):

Παγκοσμίως, το **30% (95% CI= 27.8% to 32.2%)** των γυναικών
έχουν υποστεί βία από κάποιον
σύντροφό τους.

The global prevalence of physical and/or sexual intimate partner violence among all ever-partnered women was 30.0% (95% confidence interval [CI] = 27.8% to 32.2%). The prevalence was highest in the WHO African, Eastern Mediterranean and South-East Asia Regions, where approximately 37% of ever-partnered women reported having experienced physical and/or sexual intimate partner violence at some point in their lives (see Table 2). Respondents in the Region of the Americas reported the next highest prevalence, with approximately 30% of women reporting lifetime exposure. Prevalence was lower in the high-income region (23%) and in the European and the Western Pacific Regions, where 25% of ever-partnered women reported lifetime intimate partner violence experience (see Figure 2).⁴

The global lifetime prevalence of intimate partner violence among ever-partnered women is 30.0% (95% CI = 27.8% to 32.2%).

(Published in May 2021)

SOCIAL DETERMINANTS OF HEALTH

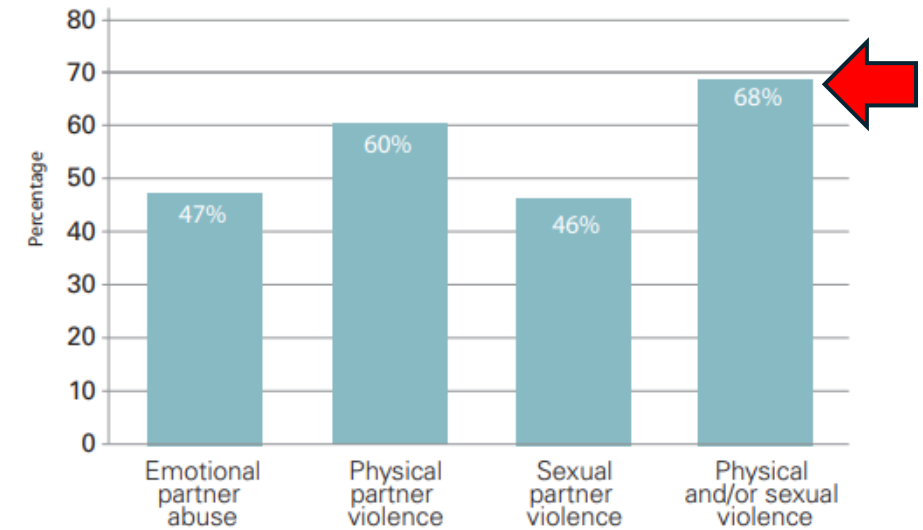
COUNTRY CASE

Measuring and responding to violence against women in Kiribati

Action on gender inequality as a social
determinant of health

<https://iris.who.int/handle/10665/207758> (σελ.1-2)

Figure 1. Percentage of women aged 15–49, who have ever been in a relationship, reporting emotional, physical and sexual partner violence ($n = 1527$)



Source: Kiribati Family Health and Support Study 2010

Το **68%** των γυναικών ηλικίας 15–49 ετών, που είχαν βρεθεί σε σχέση, είχαν βιώσει κάποια μορφή βίας (συναισθηματική, σωματική ή/και σεξουαλική) από κάποιον σύντροφό τους.



Θέτοντας ερευνητικά ερωτήματα



<https://doi.org/10.1002/14651858.CD013480.pub2>

Απαντώντας ερευνητικά ερωτήματα

Psychological and educational interventions for preventing falls in older people living in the community

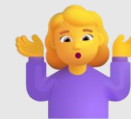
✉ Amy Drahota, Julie E Udell, Heather Mackenzie, Mark T Pugh Authors' declarations of interest

Version published: 03 October 2024 Version history

<https://doi.org/10.1002/14651858.CD013480.pub2>



Rate ratio 0.65, 95% CI 0.43 to 0.99; 1 study, 430 participants; moderate-certainty evidence



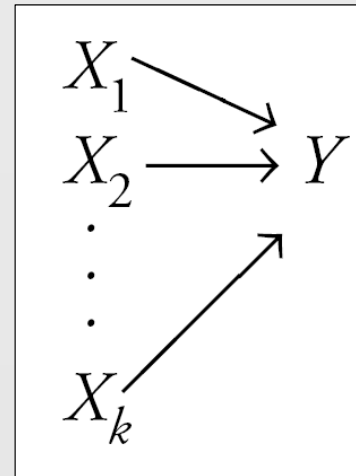
Τα στοιχεία δείχνουν ότι μια συνδυασμένη ψυχολογική και εκπαιδευτική παρέμβαση πιθανότατα μειώνει το ποσοστό πτώσεων.



Μεταβλητές

Μια μεταβλητή είναι μια ποσότητα ή ένα χαρακτηριστικό που μπορεί ελεύθερα να ποικίλλει ή να λάβει διαφορετικές τιμές.

- **Ανεξάρτητη μεταβλητή (X)** είναι η μεταβλητή που αλλάζει ή ελέγχεται σε ένα επιστημονικό πείραμα, προκειμένου να μελετηθεί η επίδρασή της σε μια άλλη μεταβλητή, που χαρακτηρίζεται ως εξαρτημένη μεταβλητή του πειράματος (έκβαση Y).
- **Εξαρτημένη μεταβλητή (Y)** είναι η μεταβλητή που μετριέται και αξιολογείται σε ένα επιστημονικό πείραμα και επηρεάζεται από τις ανεξάρτητες μεταβλητές (X) που μελετώνται (δηλαδή είναι η έκβαση του πειράματος).



Ανεξάρτητες μεταβλητές X
(independent variable)



Εξαρτημένη μεταβλητή Y
(dependent variable)



Psychological and educational interventions for preventing falls in older people living in the community

✉ Amy Drahota, Julie E Udell, Heather Mackenzie, Mark T Pugh Authors' declarations of interest

Version published: 03 October 2024 Version history

<https://doi.org/10.1002/14651858.CD013480.pub2>

Ανεξάρτητη μεταβλητή: το είδος της παρέμβασης (ψυχολογική, εκπαιδευτική, συνδυασμός) που λαμβάνεται από τους ηλικιωμένους

Εξαρτημένη μεταβλητή: η μέτρηση του ποσοστού των πτώσεων

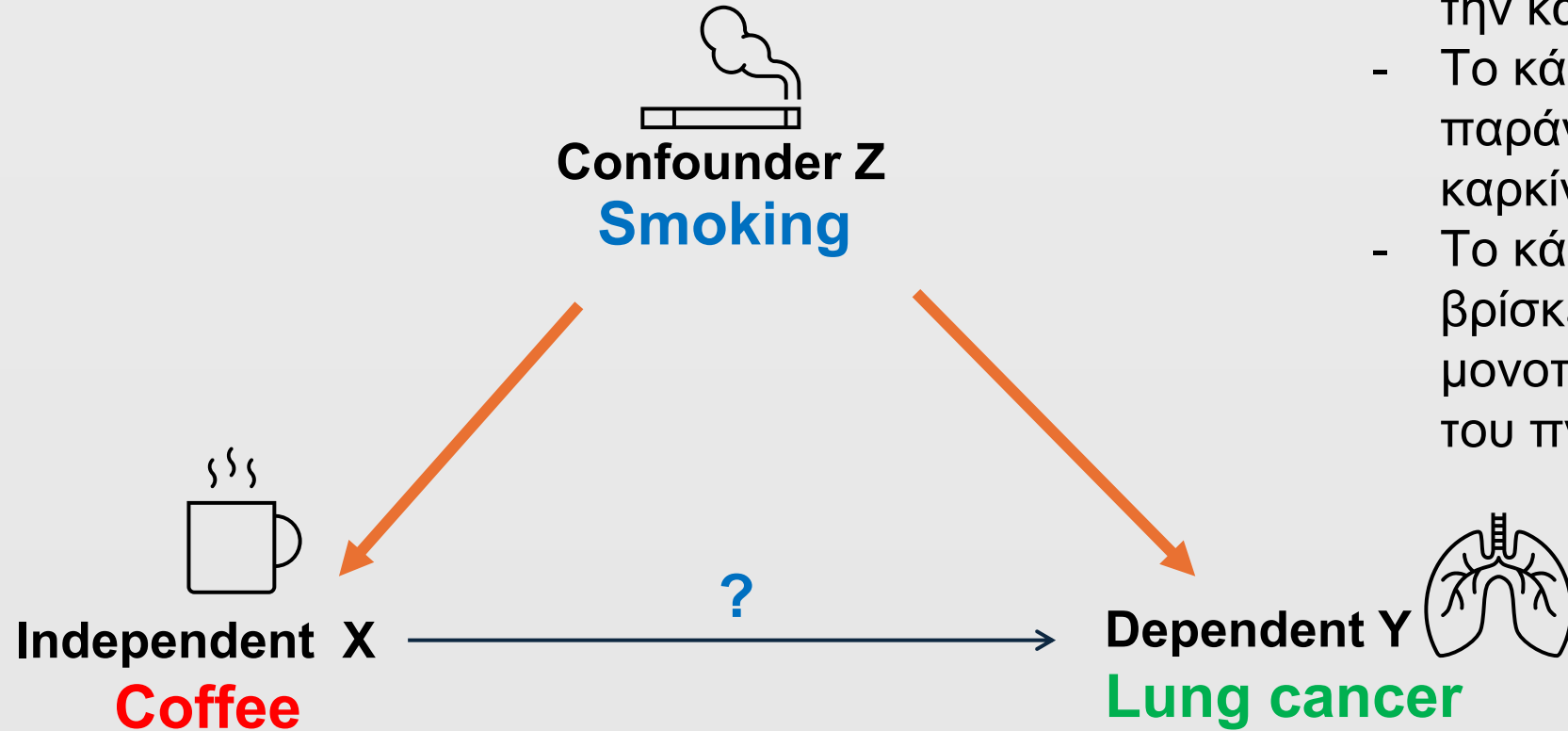
- Παρέμβαση (ψυχολογική κτλ.)
- Φύλο
- Προβλήματα αστάθειας



**Πτώσεις
ηλικιωμένων**



Παράδειγμα συγχυτικής μεταβλητής



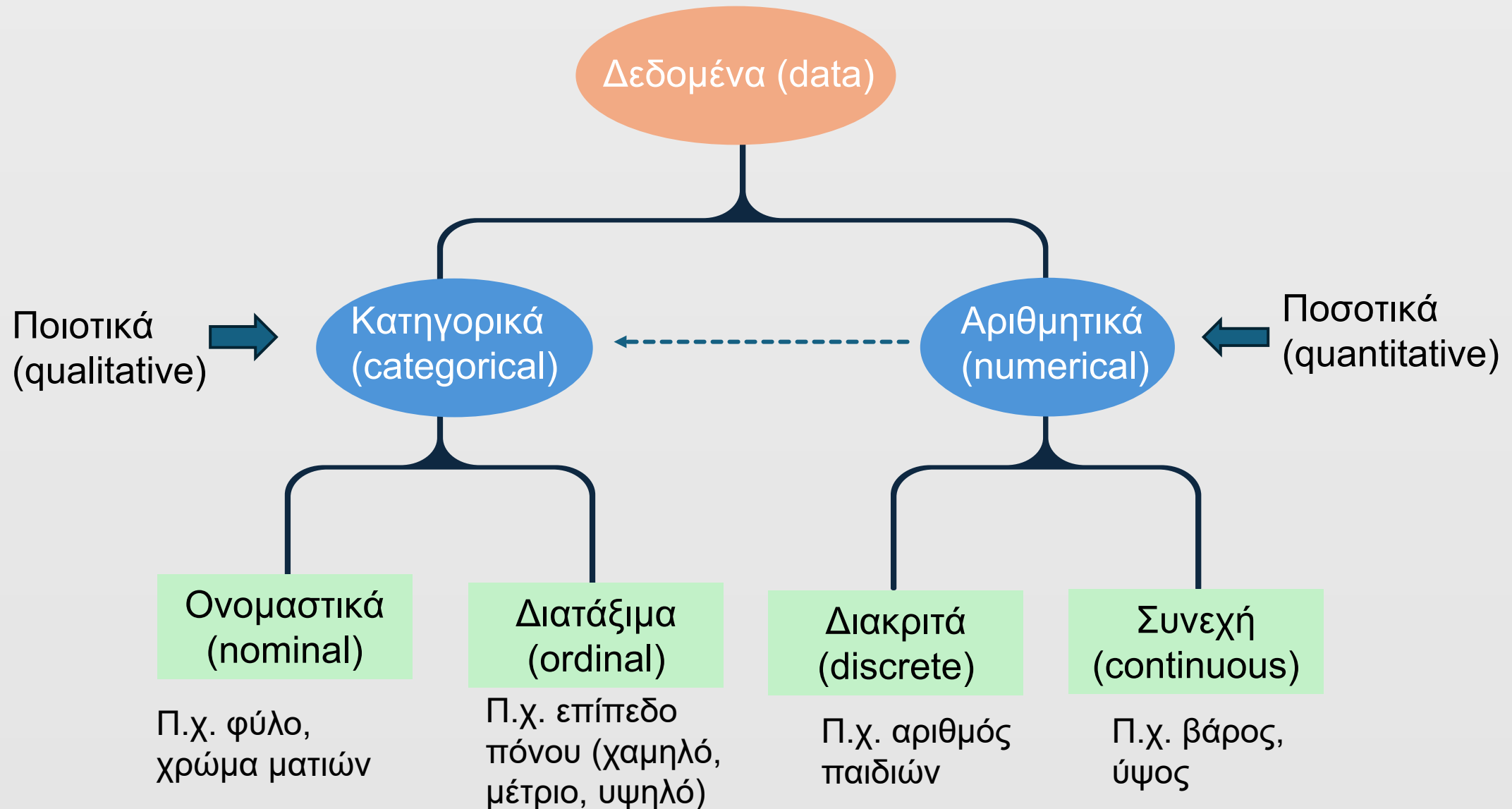
Τρίτος παράγοντας

- Το κάπνισμα συνδέεται με την κατανάλωση καφεΐνης
- Το κάπνισμα είναι σημαντικός παράγοντας κινδύνου για καρκίνο του πνεύμονα
- Το κάπνισμα δεν πρέπει να βρίσκεται στο αιτιολογικό μονοπάτι καφεΐνης –καρκίνου του πνεύμονα

Συγχυτική (confounding) είναι η κατάσταση όπου η σχέση μεταξύ της ανεξάρτητης μεταβλητής-εξαρτημένης μεταβλητής οφείλεται ολικώς ή μερικώς στην έκθεση σε έναν άλλο τρίτο παράγοντα (μεταβλητή), που ονομάζεται συγχυτικός παράγοντας.



Τύποι δεδομένων





Τα δεδομένα μας

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	sex	age	time_spent	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	
2	2	18	9	3	3	0	2	0	2	3	1	1	0	
3	2	19	3	1	2	1	2	1	1	0	2	2	2	
4	1	18	3	2	2	0	1	2	2	1	1	1	1	
5	1	17	12	2	2	0	1	1	2	1	1	1	1	
6	1	17	3	2	2	1	2	2	2	1	1	1	1	
7	1	18	3	2	2	1	3	1	2	1	1	1	1	
8	2	18	7	1	1	1	1	1	2	2	1	1	1	
9	1	17	3	2	1	1	2	1	2	2	1	1	1	
10	2	18	2	2	1	1	1	1	2	3	1	1	1	
11	2	14	3	3	2	1	2	0	3	3	1	1	1	
12	2	17	3	3	3	1	3	1	3	2	0	0	0	
13	2	20	1	3	2	1	2	1	2	2	1	1	1	
14	1	17	3	2	1	1	1	2	2	1	1	2	1	
15	2	18	7	1	1	2	1	1	1	1	1	1	1	
16	2	17	3	1	0	2	1	2	0	2	2	2	2	
17	2	17	3	1	1	3	1	1	2	0	1	3	3	
18	2	18	9	2	1	1	1	1	2	3	1	1	1	
19	2	18	13	0	1	2	1	2	1	0	2	3	3	
20	2	22	1	1	1	2	1	1	1	1	1	1	1	
21	2	14	1	2	2	1	2	2	2	1	1	1	1	

Παρατήρηση
(observation)

Μεταβλητή
(variable)



Όσο υψηλότερο score τόσο μεγαλύτερη αυτοεκτίμηση.

		STRONGLY AGREE	AGREE	DISAGREE	STRONGLY DISAGREE
1	I feel that I'm a person of worth, at least on an equal plane with others.	SA 3	A 2	D 1	SD 0
2	I feel that I have a number of good qualities.	SA	A	D	SD
3	All in all, I am inclined to feel that I am a failure.**	SA 0	A 1	D 2	SD 3
4	I am able to do things as well as most other people.	SA	A	D	SD
5	I feel I do not have much to be proud of.**	SA	A	D	SD

Codes

- Strongly Disagree=0
- Disagree=1,
- Agree=2
- Strongly Agree=3

Positively worded Items 1,2,4,6,7

Negatively worded Items 3,5,8,9,10
(codes should be reversed)