

Συγχαρητήρια

ΚΑΛΩΣΗΡΘΑΤΕ

Ιατρική Στατιστική

Η Στατιστική όπως εφαρμόζεται για να απαντά στα προβλήματα της Ιατρικής πράξης

Ιατρική Στατιστική

Σκοπός

- Η παροχή βασικών γνώσεων και εισαγωγική κατανόηση της Στατιστικής όπως εφαρμόζεται στην Ιατρική και στις Βιοϊατρικές επιστήμες.
- Περισσότερο από το 90% των ερευνητικών εργασιών που δημοσιεύονται σε έγκυρα βιοϊατρικά περιοδικά επικαλούνται βιοστατιστικές έννοιες ή εφαρμόζουν βιοστατιστικές μεθόδους.
- Ο στατιστικός έλεγχος αποτελεί βασικό κριτήριο στη διαδικασία κρίσης όλων αυτών των επιστημονικών εργασιών, στις οποίες βασίζεται το σύνολο των γνώσεων της Ιατρικής και των άλλων Επιστημών Υγείας.

Στόχοι

- Η Ιατρική Στατιστική αποτελεί κλάδο της εφαρμοσμένης Βιοστατιστικής.
- Το συγκεκριμένο μάθημα επιδιώκει οι φοιτητές να εξοικειωθούν με τις βασικές στατιστικές έννοιες και αρχές παρουσίασης και ανάλυσης ιατρικών δεδομένων.

« ΙΑΤΡΙΚΗ ΣΤΑΤΙΣΤΙΚΗ »

2024 - 2025

- **Διδάσκοντες:**

- Γ. ΤΟΥΛΟΥΜΗ, Καθηγήτρια - Συντονίστρια Μαθήματος
- Β. ΣΥΨΑ, Καθηγήτρια - Συντονίστρια Μαθήματος
- Ε. ΣΑΜΟΛΗ, Καθηγήτρια
- Ν. ΠΑΝΤΑΖΗΣ, Επίκουρος Καθηγητής.
- Φ. ΟΡΦΑΝΟΣ, Ε.ΔΙ.Π.

- **Ασκήσεις**

- Α. ΑΝΑΛΥΤΗΣ, Ε.Τ.Ε.Π.
- Ν. ΔΕΣΥΠΡΗΣ, Ε.Τ.Ε.Π.
- Μ. ΚΑΣΔΑΓΛΗ, μετα-διδακτορική ερευνήτρια
- Β. ΜΠΑΡΑΛΟΥ, υποψήφια διδακτόρισα
- Α. ΣΤΑΜΟΥΛΟΠΟΥΛΟΣ, υποψήφιος διδάκτορας

Οργάνωση του μαθήματος

- **Μαθήματα Δευτέρα-Τετάρτη 10-12 σε ΔΥΟ ομάδες φοιτητών**
 - Επίθετο από Α-Κ αμφιθέατρο Φυσιολογίας
 - Επίθετο από Λ-Ω αμφιθέατρο ΝΑΒΟ
- **Ασκήσεις σε Η/Υ στην αίθουσα της Ιατρικής Σχολής πάνω από το Αναγνωστήριο (11πμ-1μμ)**
- **Σύνολο: 5 ασκήσεις ανά φοιτητή**
- **Ημέρες: Τρίτη, Πέμπτη, Παρασκευή**

1^Η ΑΣΚΗΣΗ – ΕΙΣΑΓΩΓΗ ΣΤΟ SPSS ΑΙΘΟΥΣΑ ΠΟΛΥΜΕΣΩΝ – ΔΙΑ ΖΩΣΗΣ

ΟΜΑΔΑ	ΗΜΕΡΑ	ΟΜΑΔΑ
A	Πέμπτη 10/10/2024	1-50
B	Τρίτη 15/10/2024	51-100
Γ	Πέμπτη 17/10/2024	101-150
Δ	Τρίτη 22/10/2024	151-200
E (extra)	Παρασκευή 6/12/2024	201+

Βιβλιογραφία

- Petrie A & Sabin C. (2008) Ιατρική Στατιστική με μια Ματιά. (Μετάφραση: Τζώνου Α). Παρισιάνος, Αθήνα.
- B. Rosner. (2023). Θεμελιώδεις Αρχές Βιοστατιστικής. (Επ. Επ.: Ντζούφρας Ι). Προπομπός Ι.Κ.Ε.
- Τριχόπουλος Δ, Τζώνου Α, Κατσουγιάννη Κ. (2000). Βιοστατιστική. Παριζιάνος, Αθήνα.
- Medical Statistics at a glance 2nd edition, Blackwell, Oxford.
- Altman GD. (1991) Practical Statistics for Medical Research. Chapman & Hall, London.
- Armitage P, Berry G. (1994) Statistical Methods in Medical Research. Blackwell Science Ltd, Oxford.
- Pagano M & Gauvreau K. Principles of Biostatistics. Wandsworth. (μετ.: Δαφνή Ο)

E-class

- Πληροφορίες για το μάθημα: οδηγός σπουδών
- Όλες οι ανακοινώσεις του μαθήματος
- Μετά τις παραδόσεις αναρτώνται οι διαφάνειες
- Εκφωνήσεις ακήσεων
- Σημειώσεις

Τεκμηριωμένη Ιατρική (Evidence based)

- Το μεγαλύτερο μέρος των προληπτικών, διαγνωστικών και θεραπευτικών αποφάσεων σήμερα στηρίζεται στα «τεκμήρια», δηλαδή τα αποτελέσματα των ερευνών που δημοσιεύονται και αξιολογούνται ως σημαντικά.
- Ανάγκη για εργασίες υψηλής ποιότητας σε περιοδικά που θα διαβαστούν ευρέως.
- Διακύβευμα την υγεία και την ποιότητα ζωής των ανθρώπων.

Αναζήτηση επιστημονικής ιατρικής Βιβλιογραφίας

- <http://www.ncbi.nlm.nih.gov/pubmed>
- Αναζήτηση με λέξεις κλειδιά
- Διαθέσιμες πάντοτε οι περιλήψεις/ πολλές φορές ολόκληρη η εργασία

ORIGINAL ARTICLE

10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer

F.C. Hamdy, J.L. Donovan, J.A. Lane, M. Mason, C. Metcalfe, P. Holding, M. Davis, T.J. Peters, E.L. Turner, R.M. Martin, J. Oxley, M. Robinson, J. Staffurth, E. Walsh, P. Bollina, J. Catto, A. Doble, A. Doherty, D. Gillatt, R. Kockelbergh, H. Kynaston, A. Paul, P. Powell, S. Prescott, D.J. Rosario, E. Rowe, and D.E. Neal, for the ProtecT Study Group*

2016; 375:1415-1424

CONCLUSIONS

At a median of 10 years, prostate-cancer–specific mortality was low irrespective of the treatment assigned, with no significant difference among treatments. Surgery and radiotherapy were associated with lower incidences of disease progression and metastases than was active monitoring. (Funded by the National Institute for Health Research; Current Controlled Trials number, ISRCTN20141297; ClinicalTrials.gov number, NCT02044172.)

N ENGL J MED NEJM.ORG

The New England Journal of Medicine

Downloaded from nejm.org on September 25, 2016. For personal use only. No other uses without permission.

Copyright © 2016 Massachusetts Medical Society. All rights reserved.

STATISTICAL ANALYSIS

A prespecified statistical analysis plan was developed before the data for the primary analysis were accessed¹⁶ (see the Supplementary Appendix). The primary outcome of prostate-cancer mortality (the rate of death due to prostate cancer or its treatment) was compared among the three assigned treatment groups on an intention-to-treat basis with the use of Cox proportional-hazards

Prostate-cancer–specific mortality is reported with 95% confidence intervals for each treatment group, and pairwise significance tests were planned if a test of the null hypothesis of no difference in 10-year disease-specific risk of death across all three groups yielded a P value of less than 0.05.

[CANCER RESEARCH 63, 6556–6562, October 1, 2003]

Lung Cancer Mortality in Relation to Age, Duration of Smoking, and Daily Cigarette Consumption: Results from Cancer Prevention Study II¹

W. Dana Flanders,² Cathy A. Lally, Bao-Ping Zhu,³ S. Jane Henley, and Michael J. Thun

Department of Epidemiology and Surveillance, American Cancer Society, Atlanta, Georgia 30329 [W. D. F., C. A. L., S. J. H., M. J. T.]; Office of Epidemiology, Missouri Department of Health and Senior Services, Jefferson City, Missouri 65102 [B-P. Z.]; and Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia 30322 [W. D. F., C. A. L.]

Fig. 1. Logarithm of lung cancer mortality among currently smoking men by logarithm of duration of smoking, by age, CPS II 1982-1988.

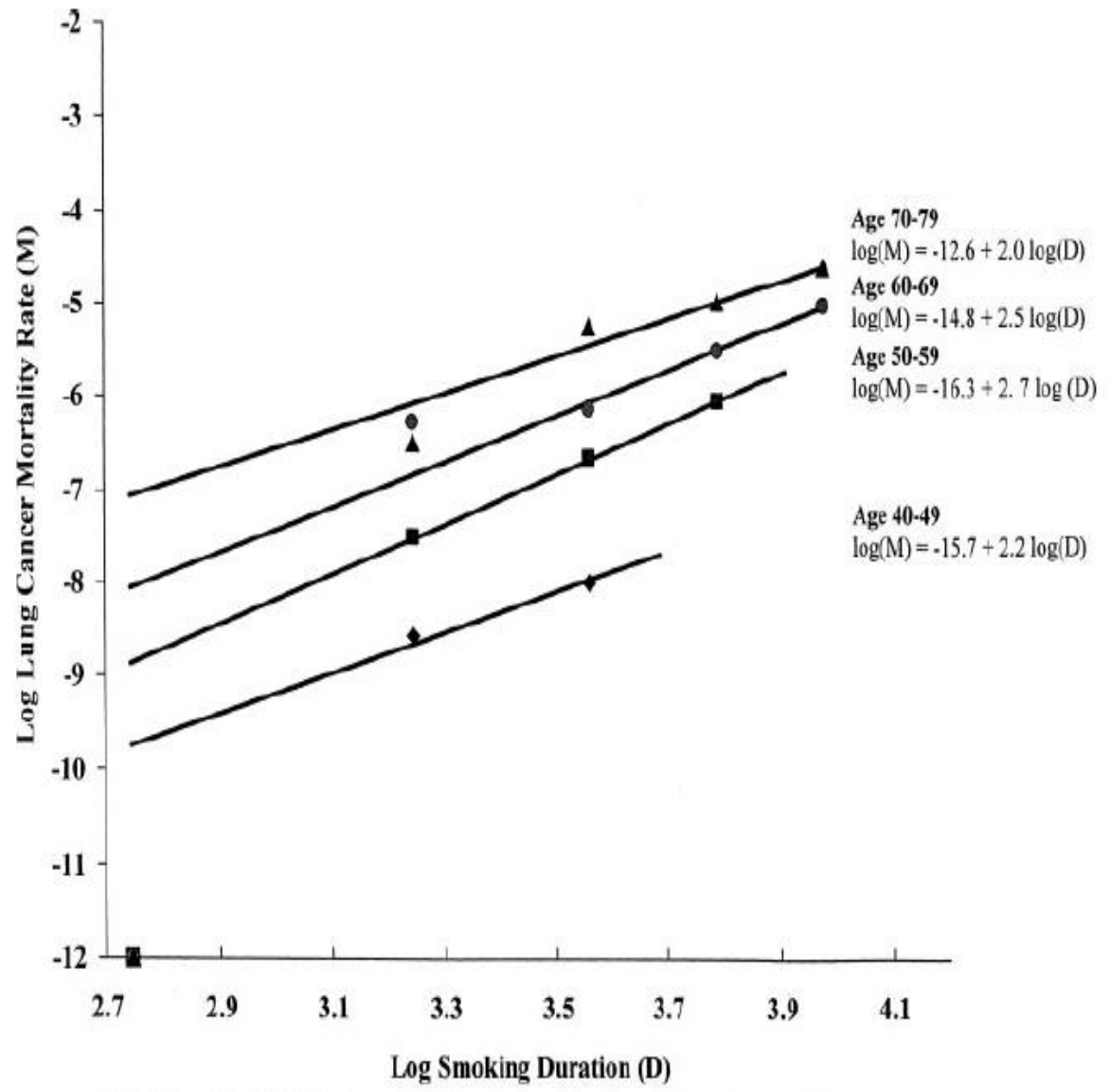
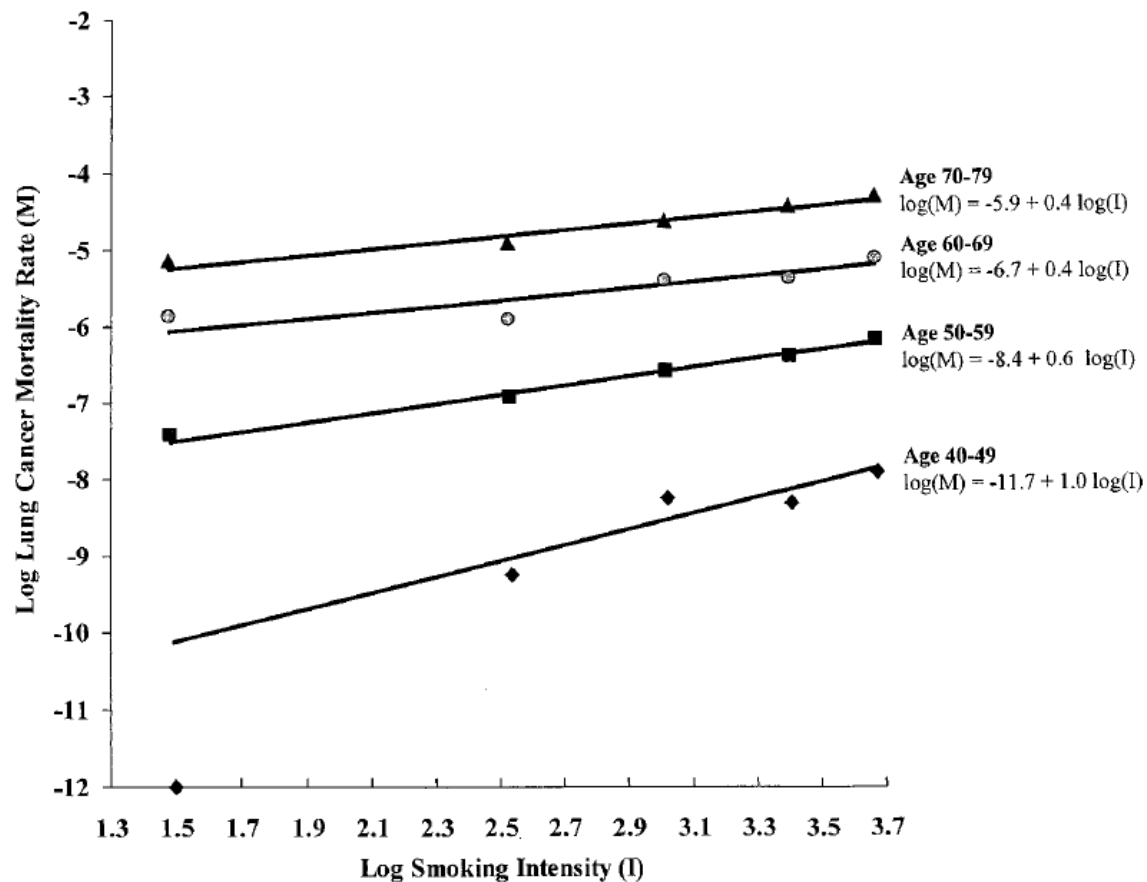


Fig. 2. Logarithm of lung cancer mortality among currently smoking men by logarithm of intensity of smoking, by age, CPS II 1982-1988.



Note: Categories with 0 deaths are plotted as though the log (mortality rate) were -12.

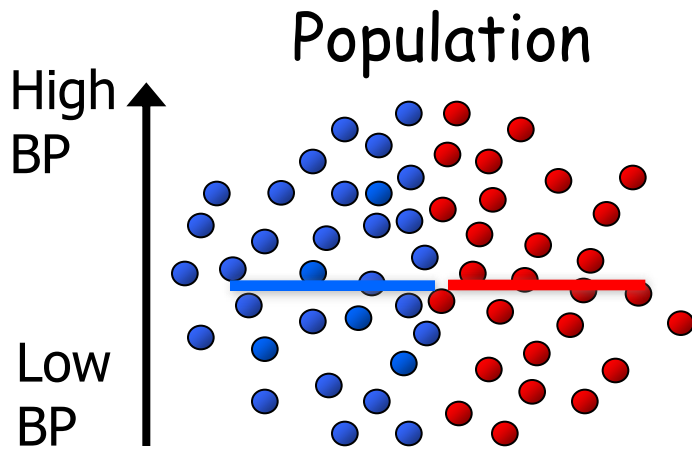
- Our study confirms that years of cigarette smoking is far more important than the number of cigarettes smoked per day in predicting lung cancer risk in United States men, regardless of age, and provides new evidence that a qualitatively similar pattern holds for women.

Γιατί υπάρχει η Στατιστική;

- διαφορετικότητα
ποικιλομορφία
μεταβλητότητα
- αβεβαιότητα



Population vs. sample



Random sample



Calculation

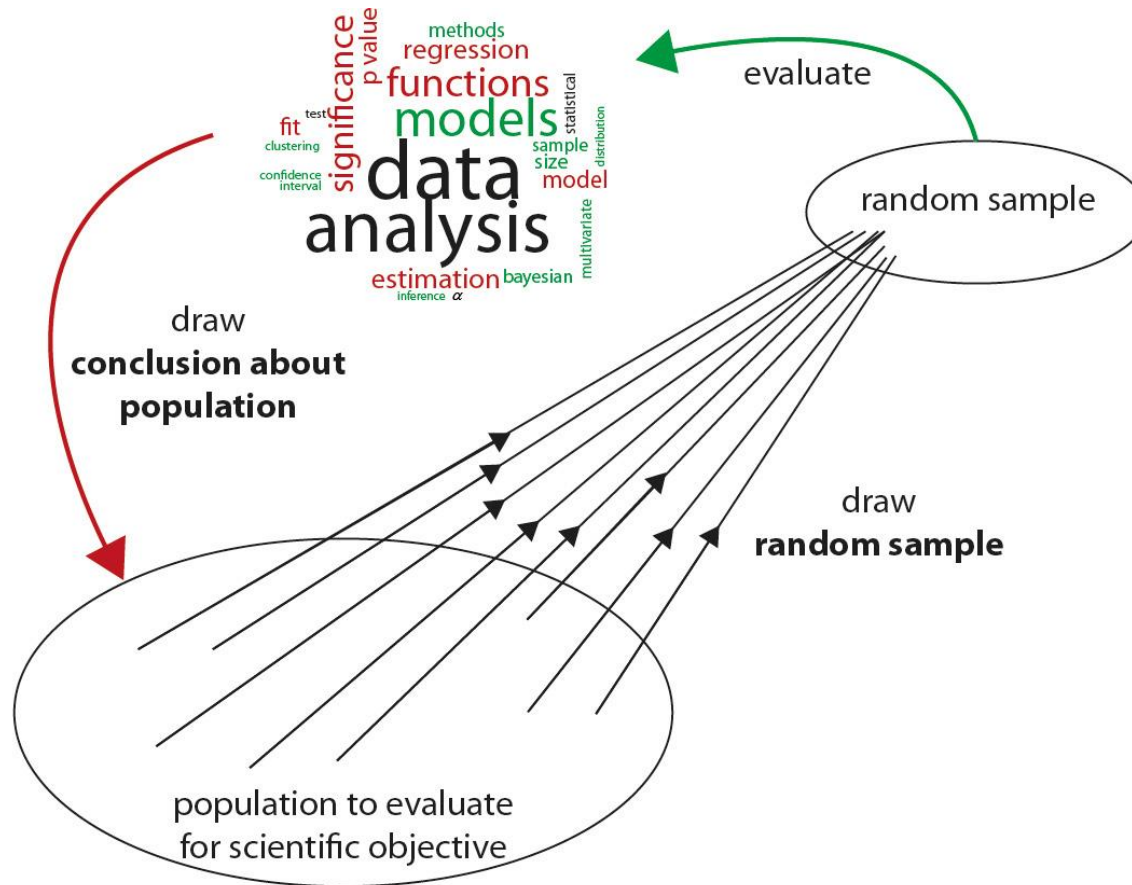
Population difference
between groups in
mean blood pressure

Inference

Sample difference
between groups in
average blood
pressure



Η Επαγωγική φύση της Στατιστικής

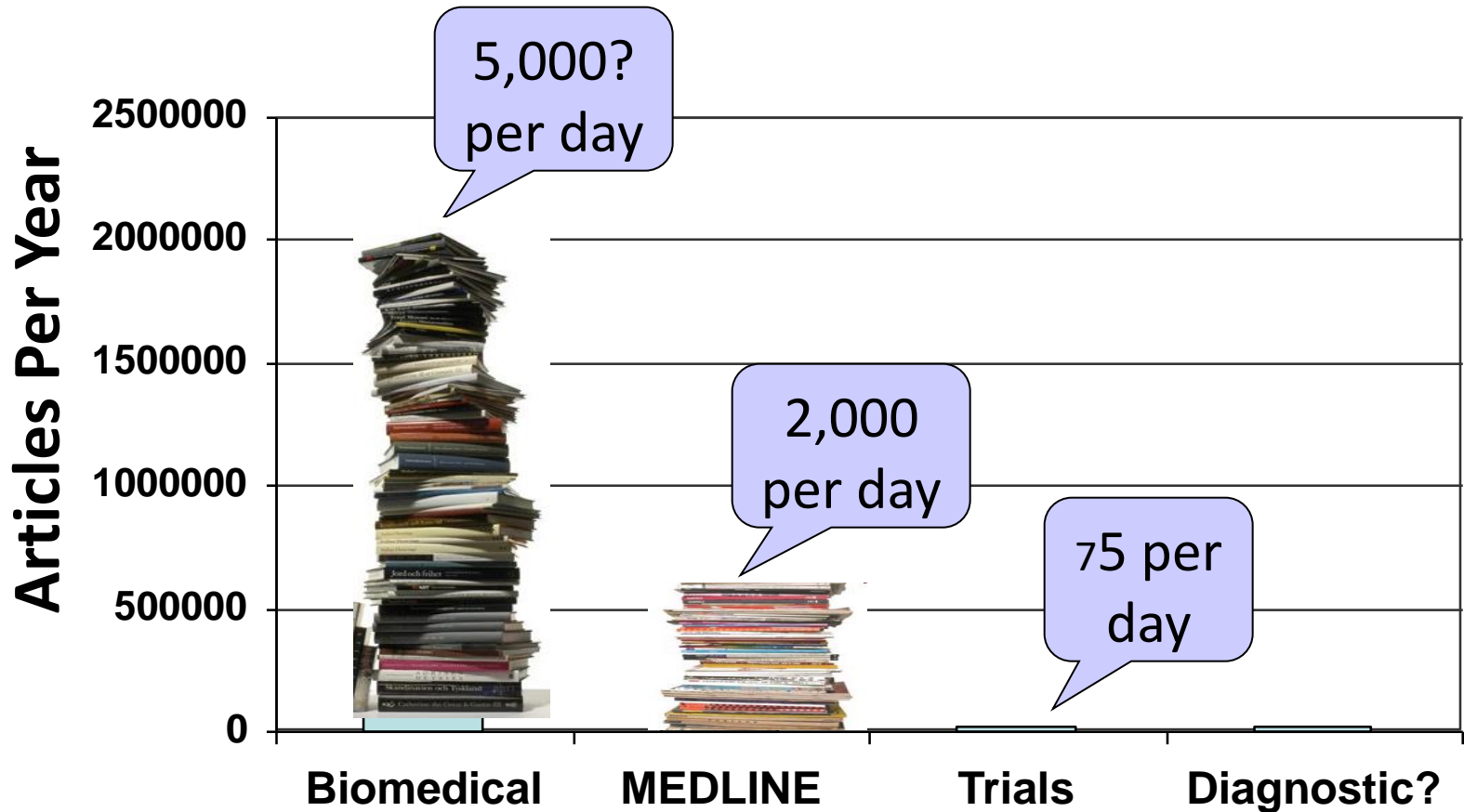


The scandal of poor medical research

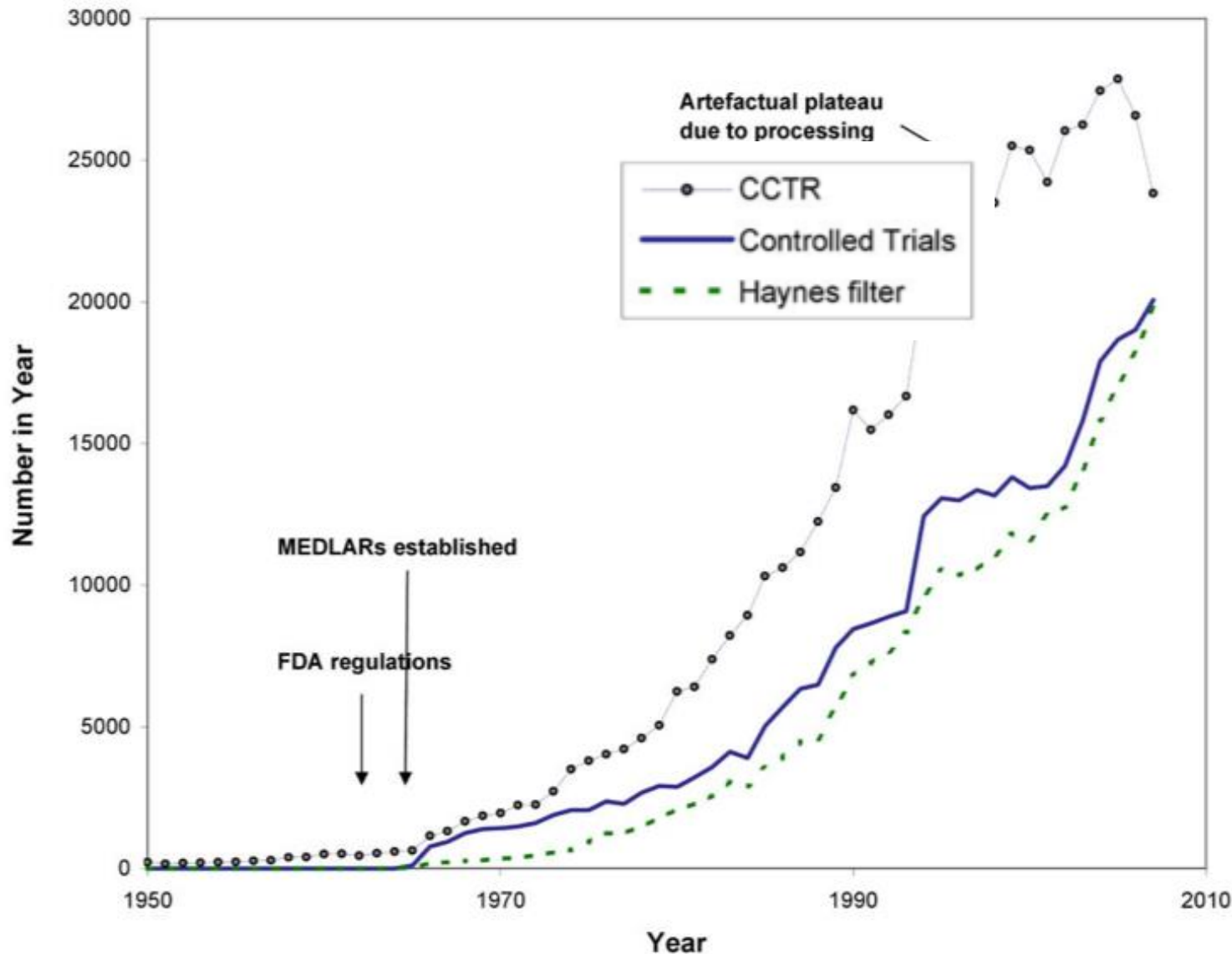
"When I tell friends outside medicine that many papers published in medical journals are misleading because of methodological weaknesses they are rightly shocked....Why are errors so common? Put simply, much poor research arises because researchers feel compelled for career reasons to carry out research that they are ill equipped to perform, and nobody stops them."

DG Altman, BMJ 1994;308:283

why do we need to use evidence efficiently?



Clinical evidence increasing so rapidly we need better skills to keep up-to-date more efficiently than previous generations of clinicians



using evidence more critically

validity: most articles should be ignored

EBM Journal Process

- 140+ journals scanned
 - 60,000 articles
- Is it **valid**? (<5%)
 - Intervention: RCT
 - Prognosis: inception cohort
 - Etc
- Is it **relevant**?
 - 6-12 GPs & specialists asked: Relevant? Newsworthy?
- < 0.5% selected

Number Needed to Read
to find 1 valid is 20+



Number Needed to Read
to find 1 valid & relevant is 200+

Avoidable waste in the production and reporting of research evidence

Iain Chalmers, Paul Glasziou

Lancet 2009; 374: 86-89 Without accessible and usable reports, research cannot research involving patients have been powerful

bmj.com

The scandal of poor epidemiological research

Erik von Elm and Matthias Egger

BMJ 2004;329;868-869
doi:10.1136/bmj.329.7471.868



Research: increasing value, reducing waste 2

Increasing value and reducing waste in research design, conduct, and analysis

John P A Ioannidis, Sander Greenland, Mark A Hlatky, Muin J Khoury, Malcolm R Macleod, David Moher, Kenneth F Schulz, Robert Tibshirani

Lancet 2014; 383: 166-75 Correctable weaknesses in the design, conduct, and analysis of biomedical and public health research studies can produce misleading results and waste valuable resources. Small effects can be difficult to distinguish from bias

Currently, many published research findings are false or exaggerated, and an estimated 85% of research resources are wasted.

Ioannidis, PLoS Medicine, 2014:e1001747

ΙΑΤΡΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Περιγραφική Στατιστική



Ιατρικές έρευνες

ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ



ΣΧΕΔΙΑΣΜΟΣ



ΕΚΤΕΛΕΣΗ (ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ)



ΔΙΟΡΘΩΣΗ ΣΤΟΙΧΕΙΩΝ



ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ



ΕΡΜΗΝΕΙΑ ΕΥΡΗΜΑΤΩΝ



ΠΑΡΟΥΣΙΑΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ



ΔΗΜΟΣΙΕΥΣΗ



Συλλογή δεδομένων **από δείγμα**
πληθυσμού – Αρχική υπόθεση

Κατασκευή **βάσης δεδομένων**

Περιγραφική στατιστική

Συνοπτική, αλλά εμπειριστατωμένη
παρουσίαση των δεδομένων

Συνοπτική παρουσίαση με απλές
περιγραφικές μεθόδους

Εφαρμογή κατάλληλου στατιστικού
κριτηρίου

Στατιστική
Συμπερασματολογία

Συναγωγή
συμπερασμάτων που
βασίζονται
στα ευρήματα
μελέτης

Αξιολόγηση κατάλληλου
στατιστικού κριτηρίου

Εξειδικευμένες **στατιστικές αναλύσεις**

Συμπεράσματα για τον **πληθυσμό**

Περιεχόμενα Περιγραφικής Στατιστικής

- ❑ Διάκριση μεταβλητών
- ❑ Παρουσίαση ποιοτικών μεταβλητών
- ❑ Παρουσίαση ποσοτικών μεταβλητών
- ❑ Κατανομές συχνοτήτων
- ❑ Αντιπροσωπευτικές τιμές θέσης και διασποράς
- ❑ Φυσιολογικές τιμές

Table 1 – Demographic characteristics of the study participants (n = 4393).

Variable	N (% [weighted])
Gender (males/females) [n (%)]	1888 (48.5)/2505 (51.5)
BMI [kg/m ² , mean (SD)]	28.2 (5.8)
Age [years, mean (SD)]	49.3 (18.6)
Age distribution (yrs)	
18–29	410 (17.7)
30–39	583 (18.3)
40–49	785 (17.7)
50–59	870 (15.6)
60–69	848 (12.7)
70+	897 (17.9)
Family income (€/month)	
<900	1831 (40.1)
900–1.700	1199 (28.6)
>1.700	459 (12.1)
Unknown	904 (19.2)
Educational level	
Primary	1637 (29.0)
Secondary	1856 (45.7)
Tertiary	829 (23.9)
Unknown	71 (1.5)
Employment status	
Employed	1573 (39.1)
Unemployed	602 (15.3)
Pensioner/Household chores	1903 (35.3)
Other	254 (9.1)
Unknown	61 (1.2)
Urbanization degree	
Urban (>10,000 inhabitants)	2218 (64.1)
Semi-rural (2000–9999 inhabitants)	815 (16.1)
Rural (<2000 inhabitants)	1360 (19.8)
Nationality	
Greek/Western Europe	4165 (93)
Africa	10 (0.4)
Eastern Europe/former Soviet Union	203 (6.1)
Asia	11 (0.4)
America/Canada	4 (0.1)

BMI: Body mass index

Σε μία έρευνα για κάθε ασθενή συλλέγουμε δεδομένα για :

1. Δείκτη υγείας
2. Θεραπευτική αγωγή/ παρέμβαση / έκθεση
3. Προγνωστικοί και άλλοι παράγοντες (ιατρικό ιστορικό κα)

Μεγέθη
Ηλικία, βάρος

Χαρακτηριστικά
Φύλο, επάγγελμα

Έννοιες
Άγχος, μόρφωση

Υπάρχει ανάγκη ποσοτικής αποτίμησης



ΜΕΤΑΒΛΗΤΕΣ

ΔΥΟ βασικά κριτήρια για...

- Συλλογή δεδομένων,
- Γραφική παράσταση,
- Περιγραφή
- Στατιστική ανάλυση

είναι

- 1. Είδος δεδομένων**
- 2. Κατανομή**

Μεταβλητή

**Αριθμητική
(ποσοτική)**

**Περιγραφική
(ποιοτική)**

Συνεχής

Μπορεί να λάβει οποιοσδήποτε τιμές ή εύρος τιμών
πχ Βάρος (kg)

Διακριτή

Ακέραιες τιμές, κλάσματος ή κτηθως αριθμ, μετρήσεων
πχ. Αριθμός καταγμάτων ή θανάτων

Διατάξιμη

Κατηγορίες ποιοτικά και ποσοτικά
επιλεκτικές και ιεραρχημένες
Πχ Βαρύτητα ασθένειας (ήπια/μεσαία/σοβαρή), APGAR score

Κατηγορική

Ποιοτικά επιλεκτικές και ποσοτικά
κατηγορίες, χωρίς ταξινόμηση
Πχ. Φύλο (άντρας/γυναίκα);
Θάνατος (ναι/οχι)



	id	age	race	smoke	bwt	htm	wkg	var
1	4	28	3	1	709	1.59	58.96	
2	10	29	1	0	1021	1.50	63.49	
3	11	34	2	1	1135	1.62	68.03	
4	13	25	3	0	1330	1.63	50.79	
5	15	25	3	0	1474	1.64	49.89	
6	16	27	3	0	1588	1.65	72.56	
7	17	23	3	0	1588	1.66	62.59	
8	18	24	2	0	1701	1.55	57.00	
9	19	24	3	0	1729	1.52	40.00	
10	20	21	1	1	1790	1.68	54.42	
11	22	32	1	1	1818	1.65	67.12	
12	23	19	1	1	1885	1.55	54.42	
13	24	25	3	0	1893	1.63	52.15	
14	25	16	3	0	1899	1.60	50.79	
15	26	25	1	1	1928	1.63	52.15	
16	27	20	1	1	1928	1.60	49.89	
17	28	21	2	0	1928	1.65	56.69	
18	29	24	1	1	1936	1.63	47.62	
19	30	21	3	0	1970	1.70	70.00	
20	31	20	3	0	2055	1.88	90.00	
21	32	25	3	0	2055	1.70	58.96	
22	22	19	1	1	2092	1.60	72.56	

Ποιοτικές μεταβλητές

Κάθε παρατήρηση-άτομο συμπεριλαμβάνεται σε μία και μόνο μία από τις υπάρχουσες κατηγορίες.

Στην πιο απλή περίπτωση μόνο 2 κατηγορίες:

πχ ασθενής, υγιής ή επιτυχία, αποτυχία (πχ. Έμφραγμα μυοκαρδίου: επιτυχία = επιβίωση για 1 χρόνο ή Λέμφωμα Hodgkins: επιτυχία = εξαφάνιση όγκων σε 1 μήνα)

Περιοχή κατοικίας (για μια έρευνα που έγινε σε κατοίκους 4 Δήμων):
Δήμος Αθηναίων, Δήμος Ζωγράφου, Δήμος Καισαριανής, Δήμος Καλλιθέας

Υγιή, πάσχοντα από καρκίνο του πνεύμονα, έμφραγμα ;;;;

Ποιοτικές μεταβλητές (2)

Παραδείγματα διατάξιμων μεταβλητών:

Ταξινόμηση της βαρύτητας μιας νόσου σε ένα δείγμα ασθενών:
ελαφρά, μέτρια, βαριά, πολύ βαριά.

Ανάστημα: μικρό, κανονικό, μεγάλο.

Ποιοτικές μεταβλητές (3)

Παρενέργειες ως αιματολογική τοξικότητα

Βαθμός	#Λευκών		Αιμοπετάλια
0 (απουσία)	≥ 4.000	και	≥ 90.000
1 ήπιες	3.000-4.000	ή	70.000-90.0000
2 μέτριες	2.000-3.000	ή	50.000-70.000
3 σοβαρές	1.000-2.000	ή	30.000-50.000
4 απειλητικές	< 1.000	ή	< 30.000

Περιγραφή ποιοτικών και διατάξιμων μεταβλητών

Κατανομή συχνοτήτων : αριθμός παρατηρήσεων ανά κατηγορία (και ποσοστό)

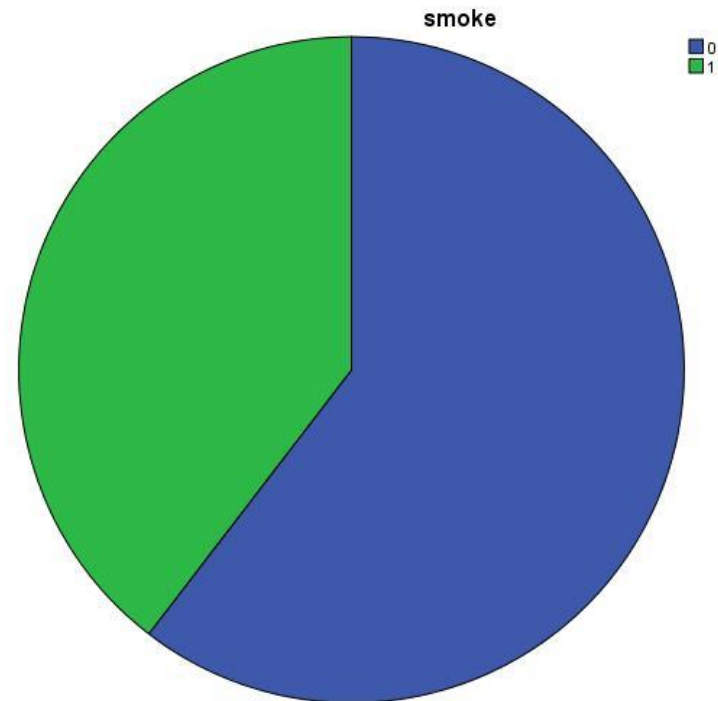
		Σύνολο
Ολοκληρωτική θεραπεία		57 (21%)
Μερική θεραπεία		110 (40%)
Εξέλιξη της νόσου		32 (12%)
Απουσία αλλαγής		74 (27%)
Σύνολο		273 (100%)

↓
Σημαντικό το μέγεθος δείγματος

Γραφικές Παραστάσεις

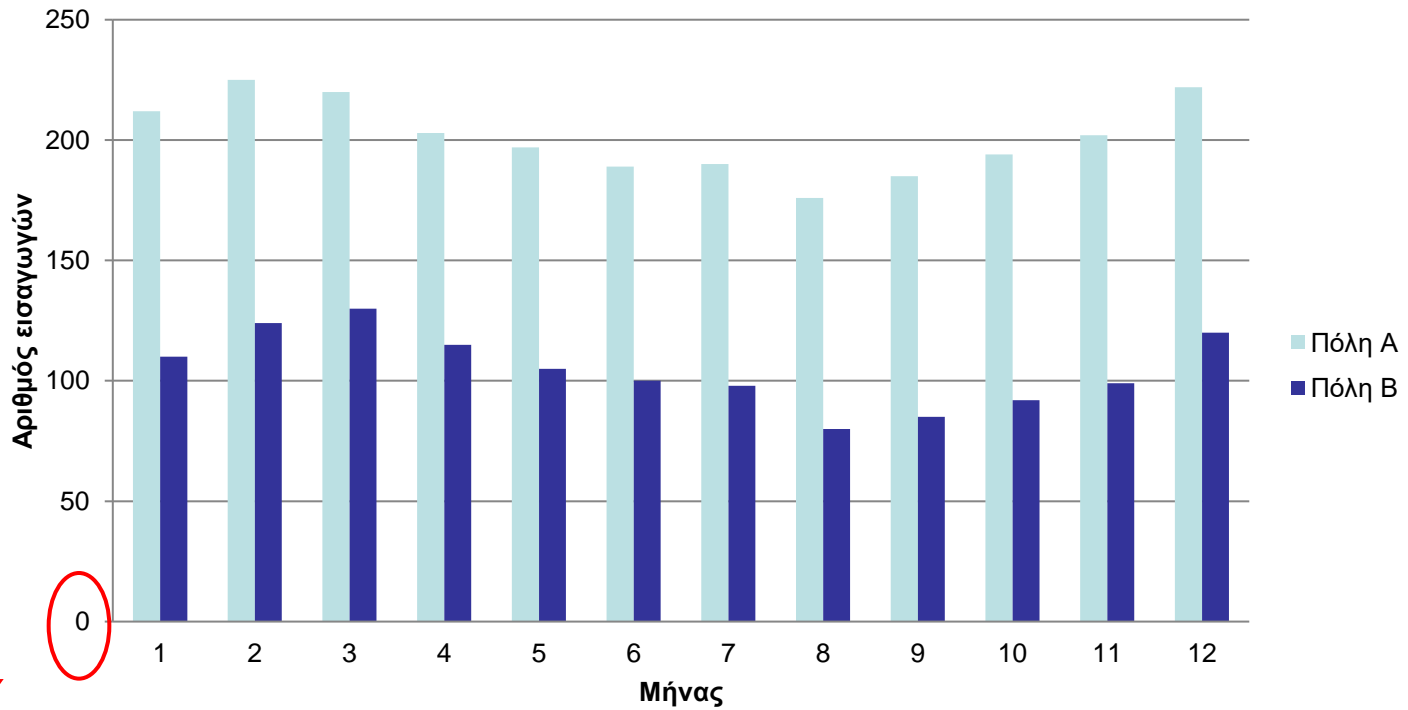
Επιτρέπουν την καλύτερη 1) κατανόηση
2) παρουσίαση

smoke					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	113	60.4	60.4	60.4
	1	74	39.6	39.6	100.0
	Total	187	100.0	100.0	



Ραβδόγραμμα – Διάγραμμα στηλών (barchart)

Αριθμός εισαγωγών με καρδιολογικό πρόβλημα στα νοσοκομεία 2 πόλεων



Ποσοτικές μεταβλητές

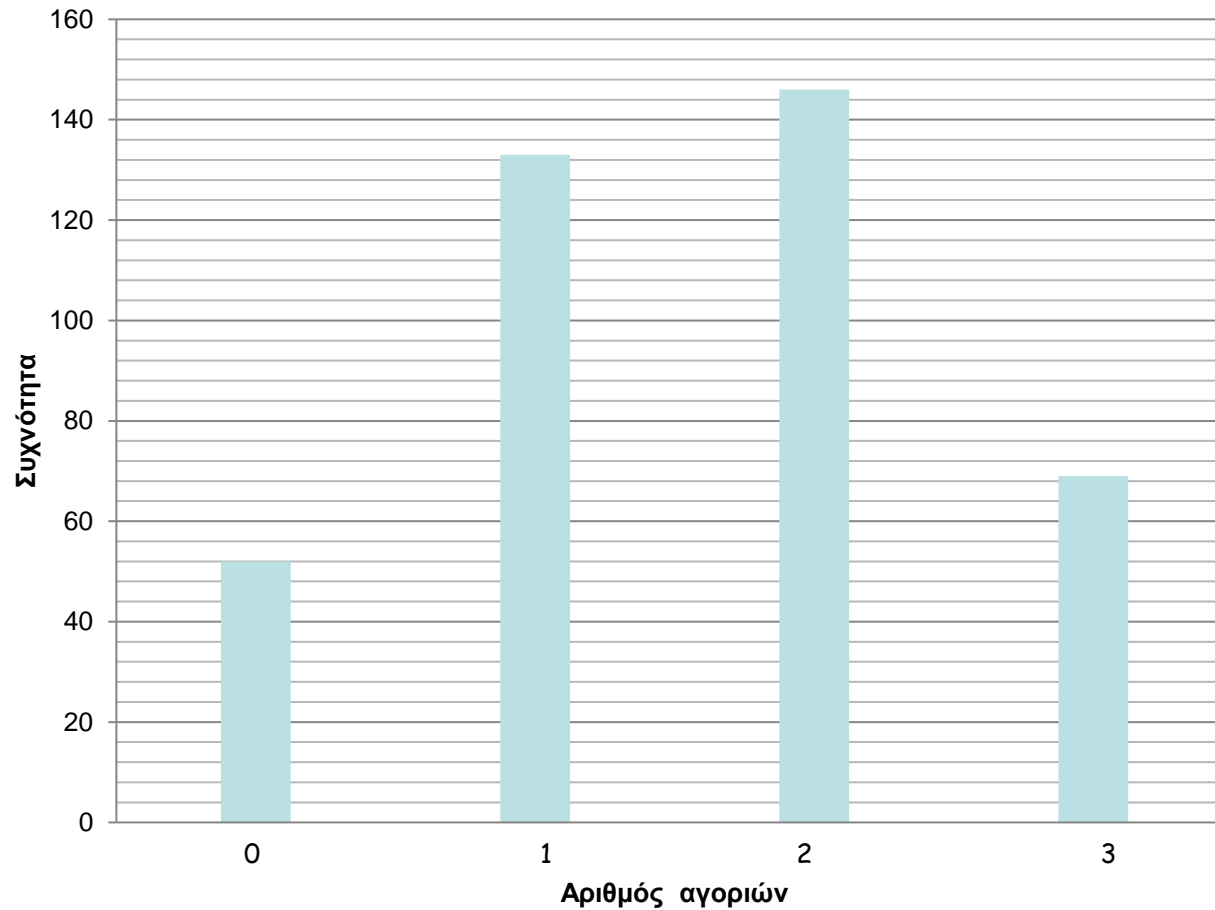
Κατανομή συχνοτήτων (frequency distribution).

Με την κατανομή συχνοτήτων καταγράφεται για κάθε τιμή ή εύρος τιμών του ποσοτικού μεγέθους ο αντίστοιχος αριθμός παρατηρήσεων. Αρχική διερεύνηση ολόκληρου του δείγματος (εντόπιση- έλεγχος ακραίων τιμών)

Κατανομή συχνοτήτων 400 οικογενειών με 3 παιδιά ανάλογα με τον αριθμό των αγοριών

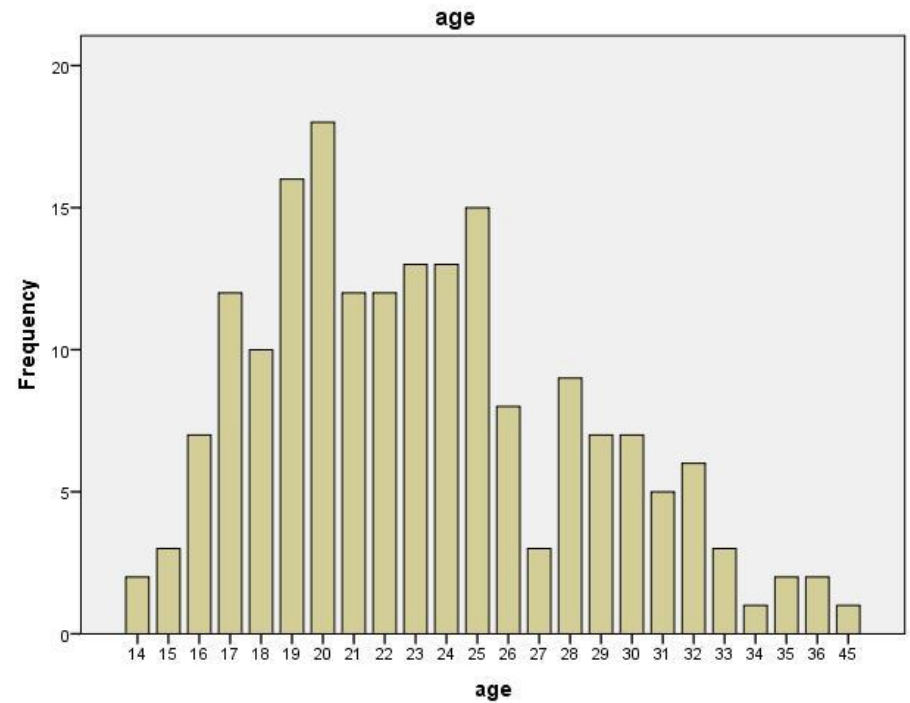
Αγόρια	Συχνότητα	Σχετική Συχνότητα %
0	52	13
1	133	33,25
2	146	36,5
3	69	17,25

Κατανομή συχνοτήτων 400 οικογενειών με 3 παιδιά ανάλογα με τον αριθμό των αγοριών



age


	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 14	2	1.1	1.1	1.1
15	3	1.6	1.6	2.7
16	7	3.7	3.7	6.4
17	12	6.4	6.4	12.8
18	10	5.3	5.3	18.2
19	16	8.6	8.6	26.7
20	18	9.6	9.6	36.4
21	12	6.4	6.4	42.8
22	12	6.4	6.4	49.2
23	13	7.0	7.0	56.1
24	13	7.0	7.0	63.1
25	15	8.0	8.0	71.1
26	8	4.3	4.3	75.4
27	3	1.6	1.6	77.0
28	9	4.8	4.8	81.8
29	7	3.7	3.7	85.6
30	7	3.7	3.7	89.3
31	5	2.7	2.7	92.0
32	6	3.2	3.2	95.2
33	3	1.6	1.6	96.8
34	1	.5	.5	97.3
35	2	1.1	1.1	98.4
36	2	1.1	1.1	99.5
45	1	.5	.5	100.0
Total	187	100.0	100.0	



Ομαδοποίηση

Κατανομή συχνοτήτων 111 δειγμάτων ούρων ανάλογα με τον αριθμό πυοσφαιρίων κατά οπτικό πεδίο

Αριθμός Πυοσφαιρίων	Αριθμός δειγμάτων
0 - 9	25
10 - 19	40
20 - 29	18
30 - 39	14
40 - 49	6
50 +	8

- ✓ Αριθμός ομάδων  6-20
- ✓ Κεντρική τιμή κάθε ομάδος $(\max + \min) / 2$
- ✓ Ανάγκη διατήρησης σταθερού εύρους όλων των ομάδων

Ομαδοποίηση τιμών (2)

Υπολογισμός κεντρικής τιμής

-Φύση μεγέθους (συνεχές ή ασυνεχές)

Βάρος στο διάστημα [30-35) kg

$$(30+35)/2 = 32,5$$

Αριθμός μαθητών ανά τάξη, για την κατηγορία 20-24

$$(20+24)/2 = 22$$

Ηλικία στο διάστημα 10-14 ποια είναι η κεντρική τιμή;

- $(10+14)/2=12$ αν ασυνεχές
- $(10+15)/2=12,5$ αν συνεχές

Εύρος ομάδων

Ηλικία σε μήνες	Αριθμός θανάτων
0	3031
1	518
2	514
3	491
4	366
5	308
6-8	583
9-11	330
Σύνολο	6141

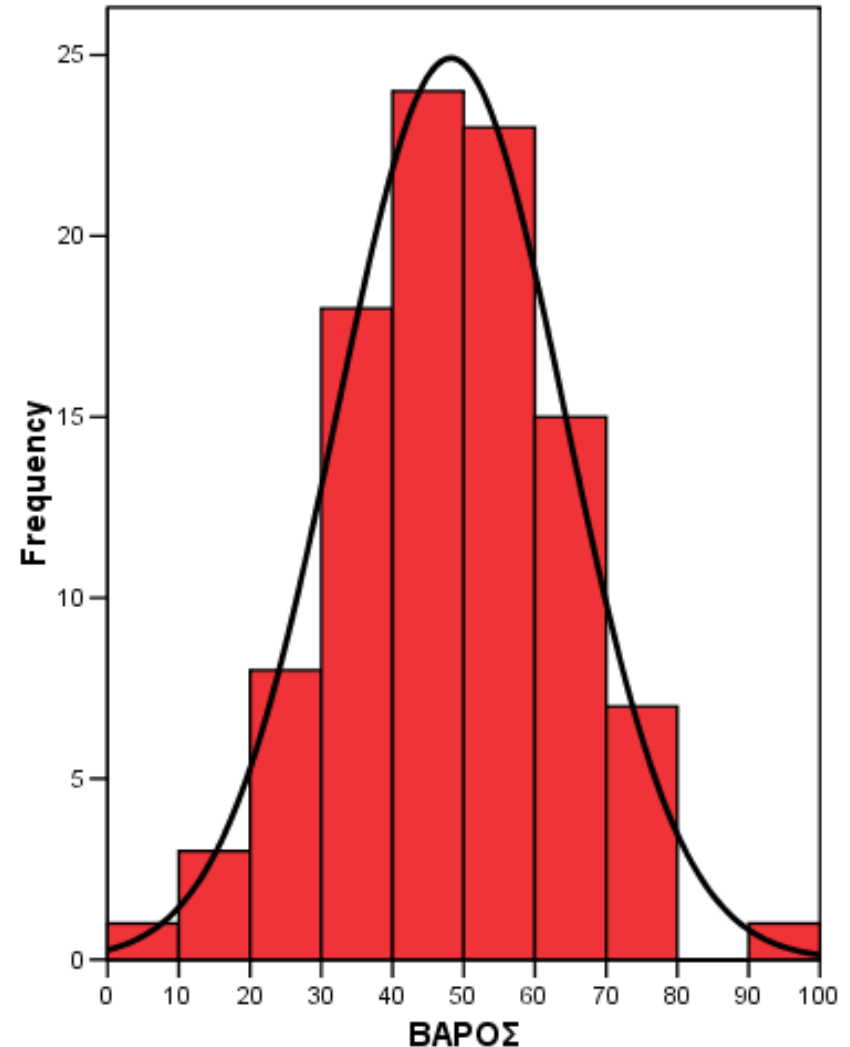
Ιστογράμματα

$$\text{Εμβαδό} = \text{Βάση} * \text{Ύψος}$$

$$\text{Ύψος} = \text{Εμβαδό} / \text{Βάση} = \text{Συχνότητα} / \text{Μετρούμενο μέγεθος}$$

Συχνότητες

Διάστημα τιμών	Τιμή	Απόλυτη συχνότητα	Σχετική συχνότητα	Απόλυτη αθροιστική συχνότητα	Σχετική αθροιστική συχνότητα
0.00- 9.99	5	1	0.01	1	0.01
10.00-19.99	15	3	0.03	4	0.04
20.00-29.99	25	8	0.08	12	0.12
30.00-39.99	35	18	0.18	30	0.30
40.00-49.99	45	24	0.24	54	0.54
50.00-59.99	55	22	0.22	76	0.76
60.00-69.99	65	15	0.15	91	0.91
70.00-79.99	75	8	0.08	99	0.99
80.00-89.99	85	0	0.00	99	0.99
90.00-99.99	95	1	0.01	100	

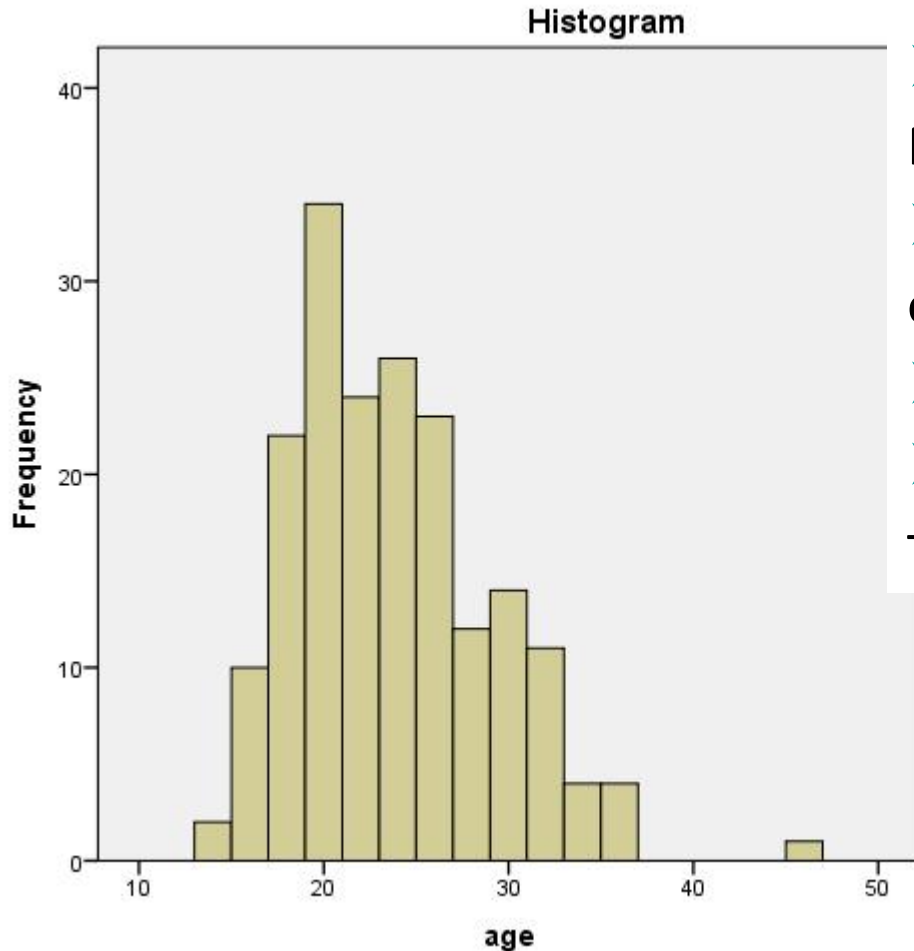


○ Ξεκινάει από το 0



○ Εύρος παραλληλογράμμου

Ιστόγραμμα



- Πόσο συμμετρικά είναι τα δεδομένα; Πόσο διεσπαρμένα είναι τα δεδομένα;
- Υπάρχουν διαστήματα με υψηλή συγκέντρωση δεδομένων;
- Υπάρχουν κενά στα δεδομένα;
- Υπάρχουν παρατηρήσεις μακριά από τις υπόλοιπες (ακραίες - outliers);

Ιστόγραμμα κλινικής δοκιμής για πρόληψη βρεφικής υπο-ασβεστιαμίας

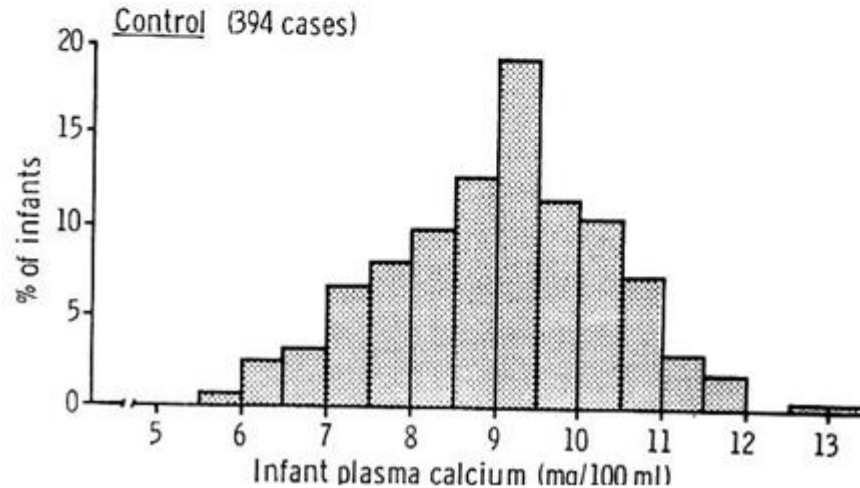
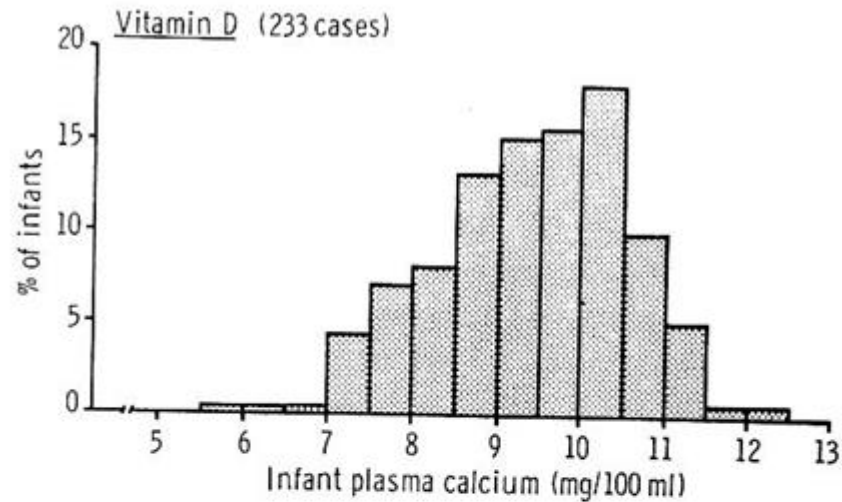
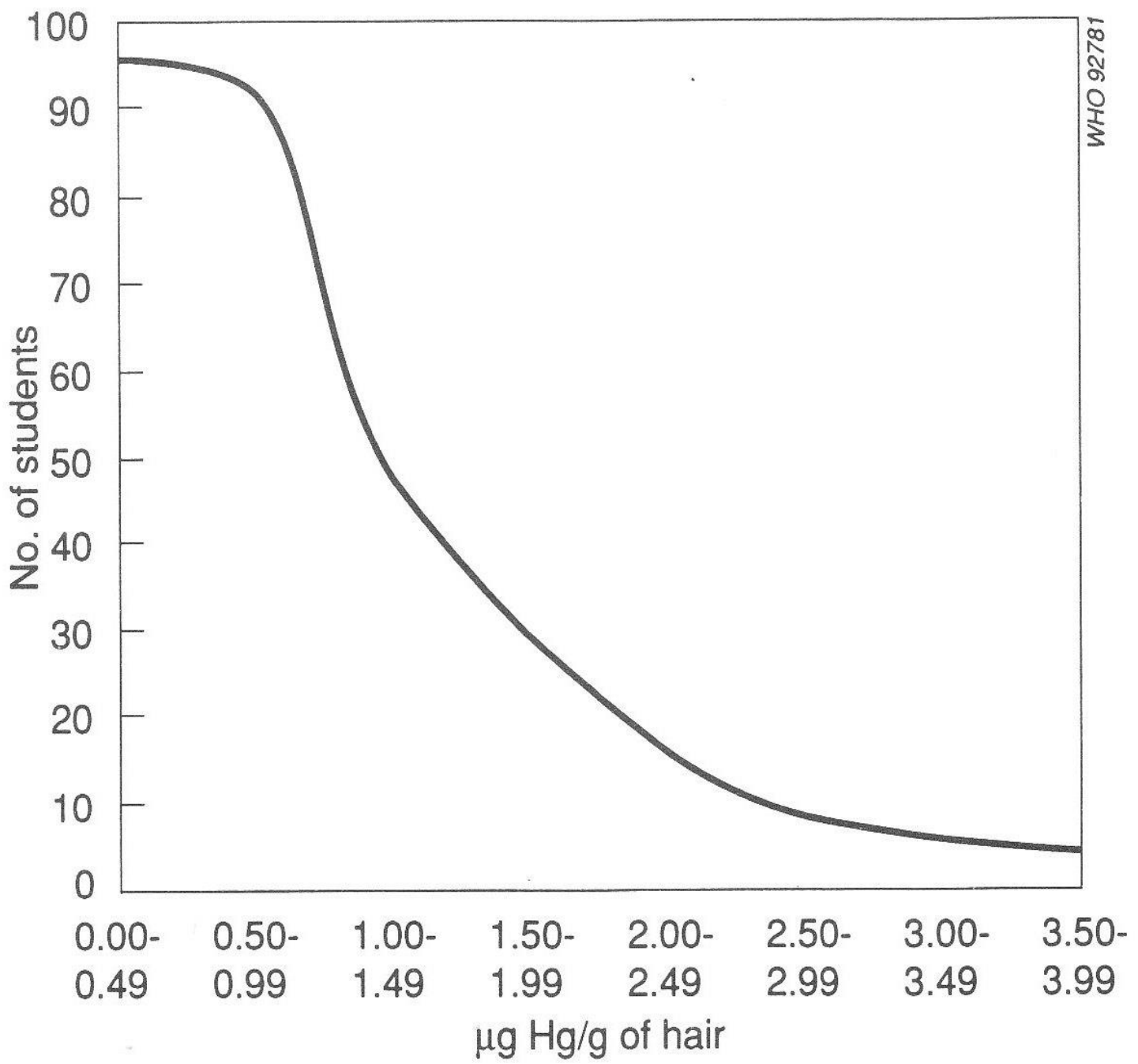


Table 4.1. Distribution of mercury concentrations in hair of 300 high school students

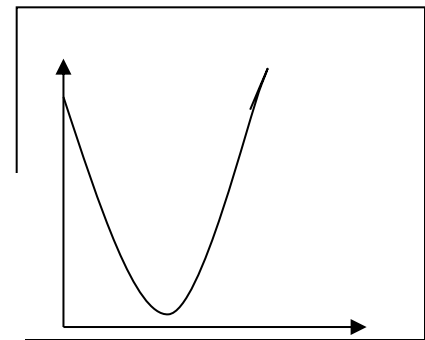
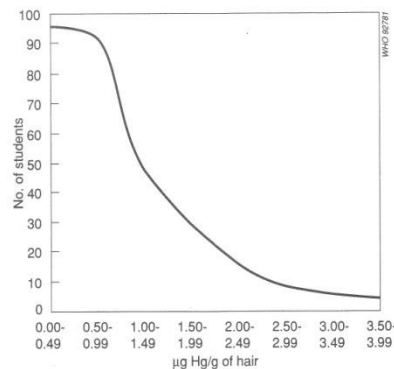
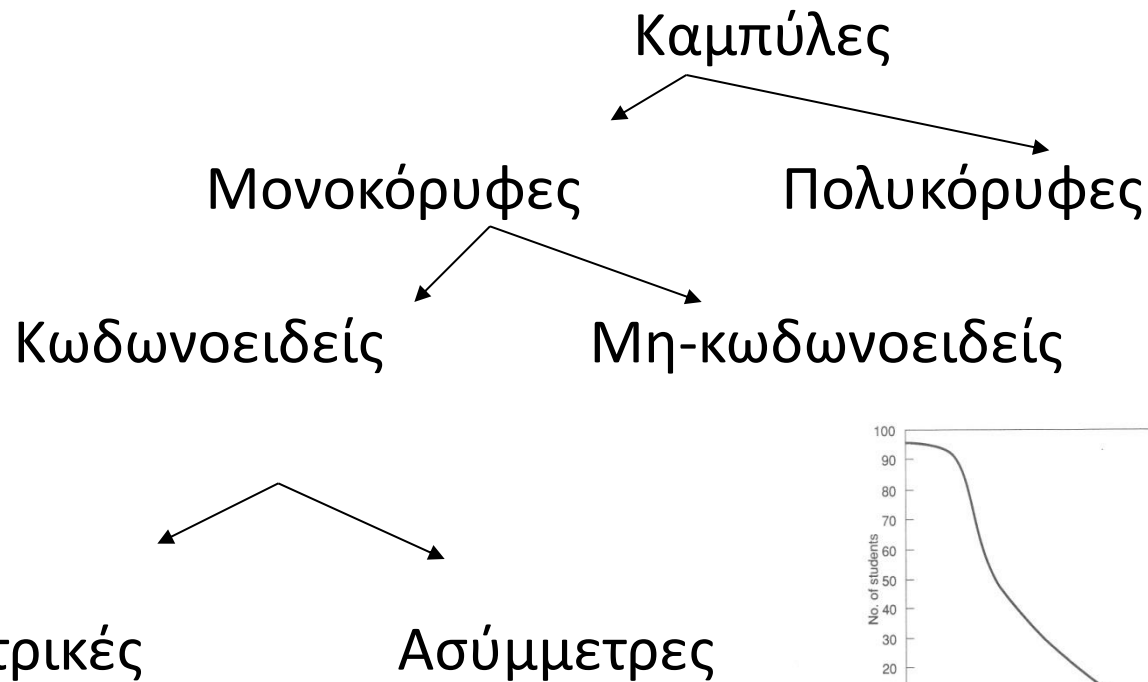
Mercury concentration ($\mu\text{g/g}$)	No. of children
0–0.49	95
0.5–0.99	91
1.0–1.49	47
1.5–1.99	30
2.0–2.49	16
2.5–2.99	8
3.0–3.49	9
3.5–3.99	4

Adapted from Kjellström et al., 1982.



WHO 92781

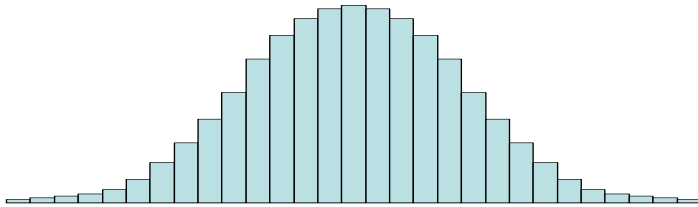
Κατανομή Ποσοτικών Δεδομένων



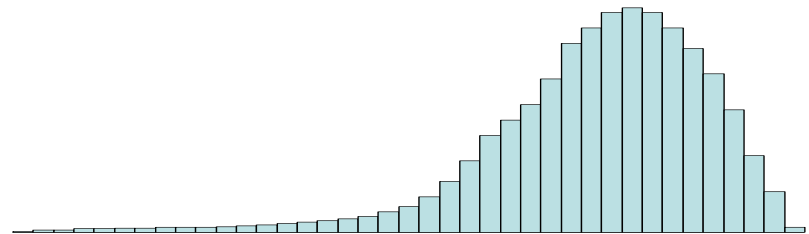
(Κατά προσέγγιση συμμετρία -> σχετική με μέγεθος δείγματος)

Κατανομή Ποσοτικών Δεδομένων

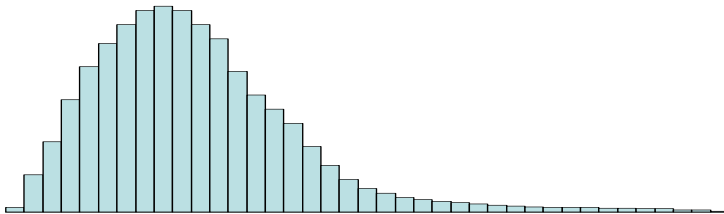
Συμμετρική, σχήμα καμπάνας
Κανονική (Gaussian) κατανομή



Αρνητικά λοξή
Λιγότερο συνηθισμένη



Θετικά λοξή
Σε εργαστηριακά δεδομένα
πχ. CD4 counts in HIV+



Ομοιόμορφη κατανομή
Ίδια πιθανότητα οποιασδήποτε τιμής
μέσα στο εύρος



Παράδειγμα κανονικής κατανομής

