

Βιοστατιστική

Περιγραφική Στατιστική (με πίνακες και γραφικές μεθόδους)

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

Φίλιππος Ορφανός, PhD
Εργαστήριο Υγιεινής, Επιδημιολογίας
και Ιατρικής Στατιστικής, Ιατρική Σχολή Αθηνών
phorfanos@med.uoa.gr

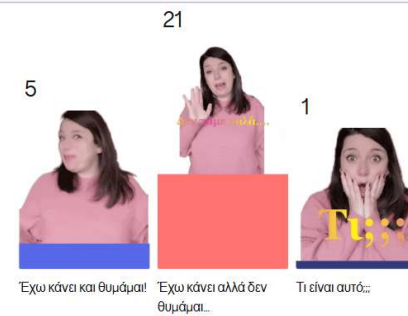
Τετάρτη 01/10/2025

Menti.com

Τι έρχεται πρώτο στο μυαλό σας στο άκουσμα της λέξης "Στατιστική";

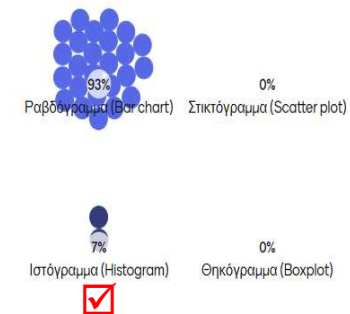
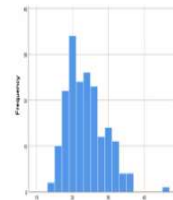
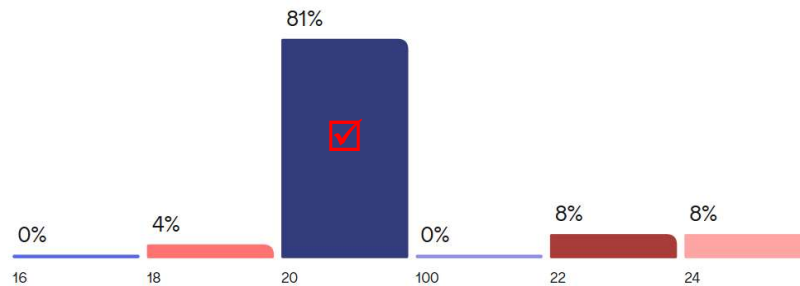


Ποια η εμπειρία σας με το μάθημα Στατιστικής;



Πως ονομάζεται αυτό το γράφημα;

Ποιος είναι ο μέσος όρος των αριθμών : 16, 18, 20, 22, 24 ;



Τι είναι Στατιστική;

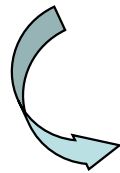
“Statistics is the art of learning from data. It is concerned with the **collection of data**, their subsequent description, and **their analysis**, which often leads to the **drawing of conclusions**”

Sheldon M. Ross, 2005

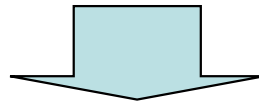
Στατιστική και Μαθηματικά

- Αρνητική προκατάληψη με το άκουσμα και μόνο της λέξης «ΣΤΑΤΙΣΤΙΚΗ»

Γιατί;



Είναι Μαθηματικά



- η αντίληψη που επικρατεί ότι η Στατιστική είναι κλάδος των Μαθηματικών **δεν είναι απολύτως σωστή**

Στατιστική και Μαθηματικά



Στατιστική είναι η μελέτη της **ΑΒΕΒΑΙΟΤΗΤΑΣ**.

- Τα Μαθηματικά ακολουθούν μια σταθερή δομή (ορισμός, θεώρημα, απόδειξη)
- Με τη Στατιστική ορίζουμε πράγματα με τη διαίσθηση και με παραδείγματα (χρήση δεδομένων)

Στατιστική και Μαθηματικά

Πολλές στατιστικές μέθοδοι συμπεριφέρονται άσχημα όταν τα δεδομένα περιέχουν «ακραίες τιμές» (outliers)

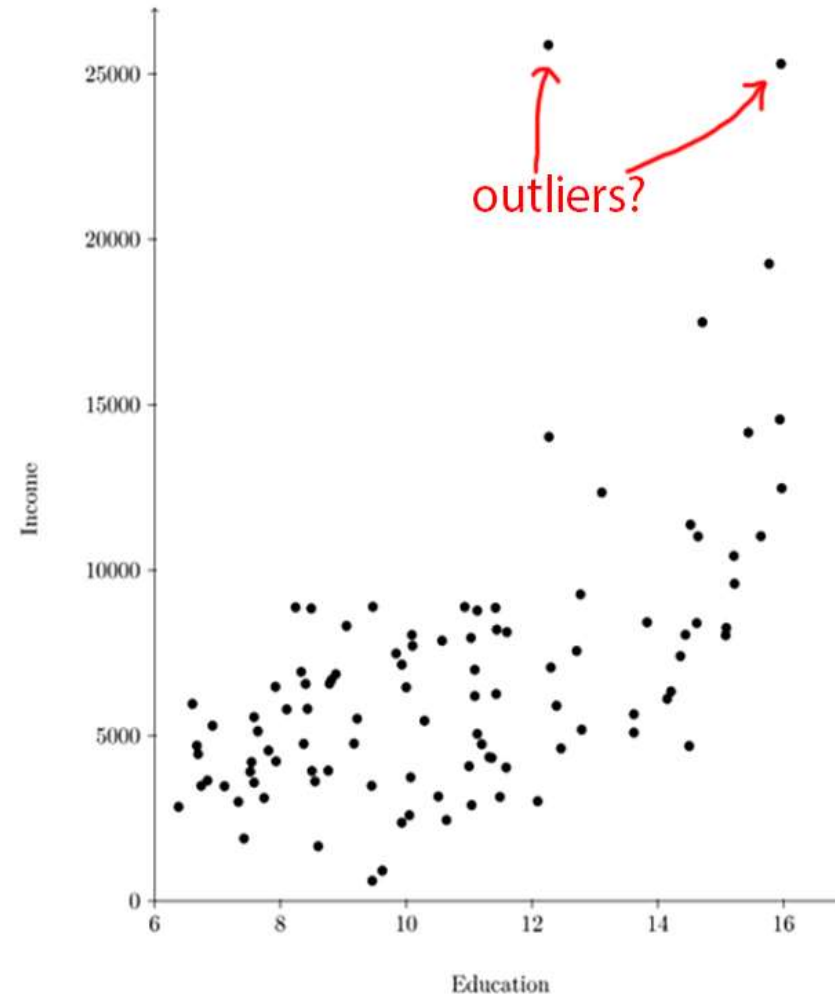
Κοινή πρακτική:

Ο εντοπισμός και αφαίρεσή τους.

Αλλά τι συνιστά μια ακραία τιμή;

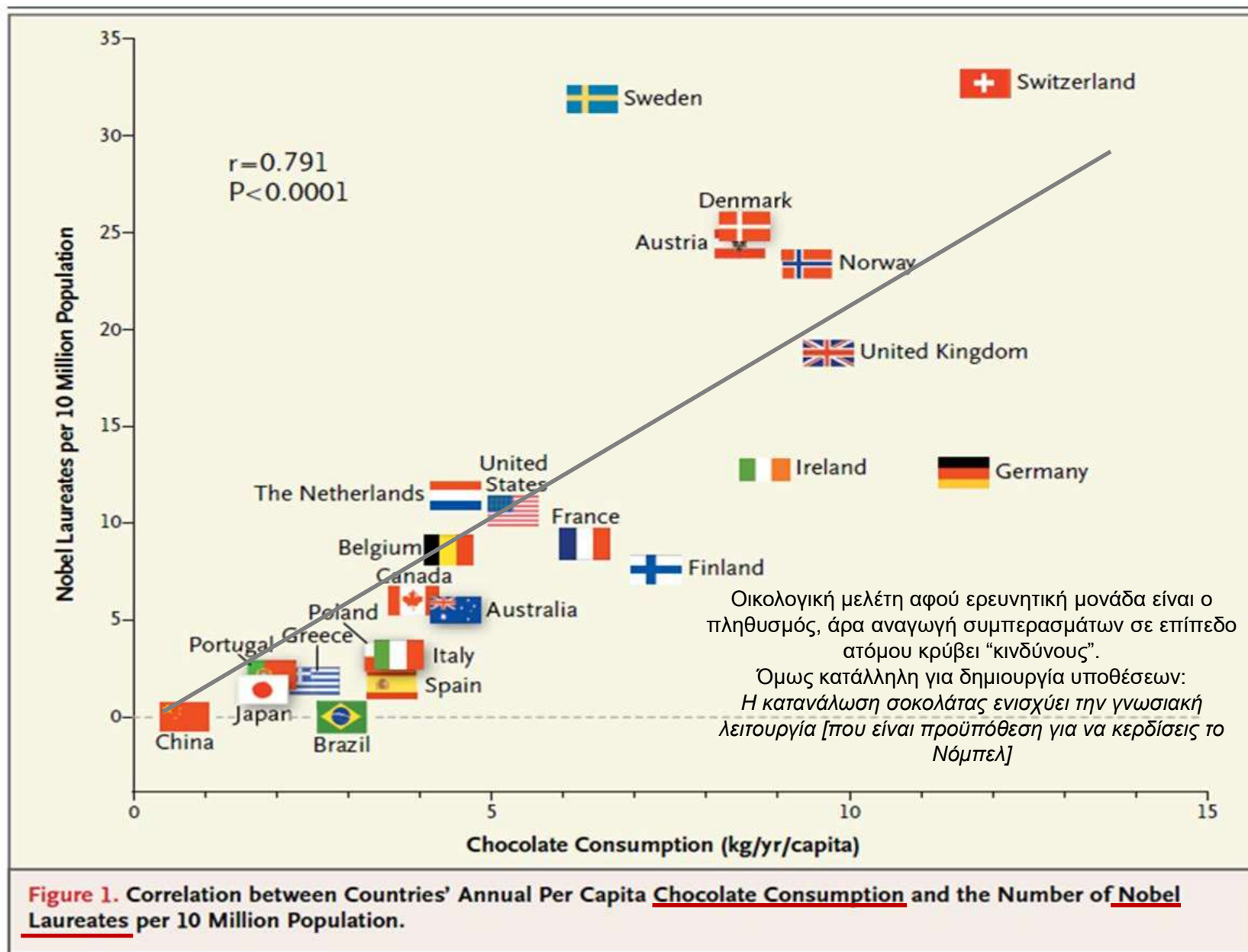
Εξαρτάται από διάφορα κριτήρια:

- πόσα σημεία δεδομένων έχετε,
- πόσο μακριά είναι από τα υπόλοιπα σημεία και
- τι είδους μοντέλο χρησιμοποιείτε.



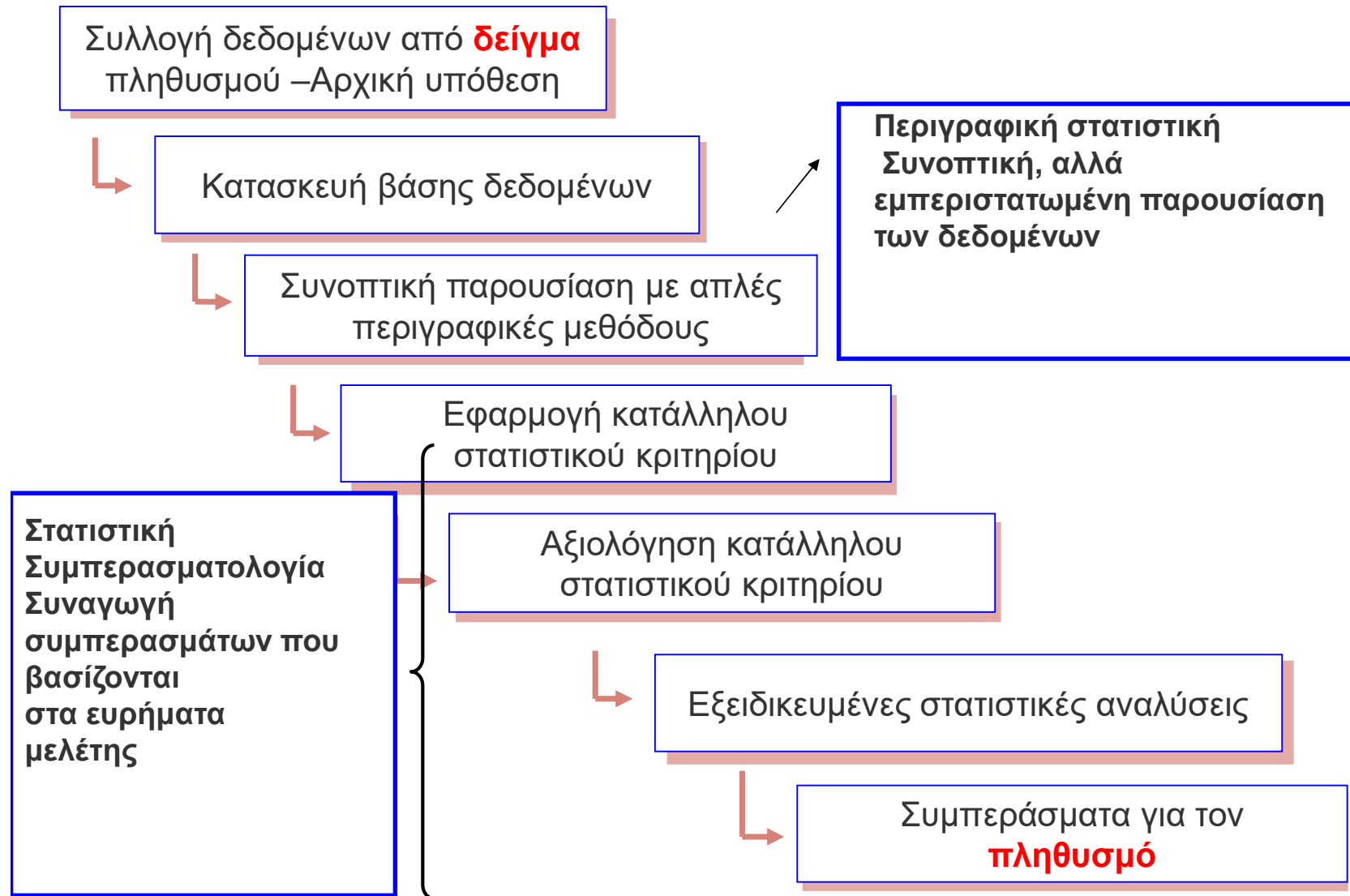
Δομή μαθήματος / Διδάσκοντες

α/α	Ημ/νία	Διάλεξη	Διδάσκων
1ο	1/10/2025	Εισαγωγικά: δομή του μαθήματος, Περιγραφική Στατιστική (1)	Ορφανός
2ο	8/10/2025	Περιγραφική Στατιστική (2): Μέτρα θέσης & διασποράς	Κωστάκη
3ο	15/10/2025	Ανάλυση ποιοτικών δεδομένων: χ^2 για ανάλυση πινάκων, αναλογίες, έλεγχος υποθέσεων για αναλογίες	Ρούσσος
4ο	22/10/2025	Ανάλυση ποσοτικών δεδομένων (1): t-test για ανεξάρτητα δείγματα	Ρούσσος
5ο	29/10/2025	Ανάλυση ποσοτικών δεδομένων (2): t-test για δείγματα με αντιστοιχία κατά ζεύγη	Κωστάκη
6ο	5/11/2025	Χρήση στατιστικών πακέτων και εισαγωγή στο SPSS. Επαναληπτικό I	Κωστάκη
7ο	12/11/2025	Ανάλυση ποσοτικών και ποιοτικών δεδομένων. Επαναληπτικό II	Ορφανός
8ο	19/11/2025	Εξάρτηση ποσοτικών χαρακτηριστικών: η έννοια της εξάρτησης δύο ποσοτικών μεταβλητών	Ρούσσος
9ο	26/11/2025	Πολλαπλή γραμμική εξάρτηση - Ερμηνεία στατιστικών ευρημάτων	Ρούσσος
10ο	3/12/2025	Απλή και πολλαπλή λογαριθμική εξάρτηση - Ερμηνεία στατιστικών ευρημάτων	Ορφανός
11ο	10/12/2025	Γενική επανάληψη / απορίες	Ορφανός
12ο	17/12/2025	<i>Παρουσίαση ερευνητικών πρωτοκόλλων I</i>	
13ο	7/1/2026	<i>Παρουσίαση ερευνητικών πρωτοκόλλων II</i>	



Περιγραφική Στατιστική

Ιατρική Έρευνα



Περιεχόμενα Περιγραφικής Στατιστικής

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

Συλλογή δεδομένων

Για να μάθεις κάτι, πρέπει πρώτα **να συλλέξεις δεδομένα**.

Για παράδειγμα, το πρώτο βήμα για να μάθεις για:

- τη σημερινή κατάσταση της οικονομίας
- το ποσοστό των πολιτών που ψηφίζουν υπέρ μιας συγκεκριμένης πρότασης
- την κατανάλωση βενζίνης σε λίτρα/100 χλμ του αυτοκινήτου
- την αποτελεσματικότητα ενός νέου εμβολίου

είναι η συλλογή δεδομένων.

Συλλογή δεδομένων

- **Ερευνητικές μονάδες:** «Οντότητες» στις οποίες αναφέρεται η πληροφορία που συλλέγουμε (π.χ. άτομα, νοικοκυριά, χώρες κτλ)
- **Μεταβλητή:** χαρακτηριστικό το οποίο συλλέγουμε για κάθε ερευνητική μονάδα και παίρνει περισσότερες από μία διαφορετικές τιμές ή χαρακτηρίζει διαφορετικά άτομα

Μεταβλητές

```
graph TD; A[Μεταβλητές] --> B[Ποιοτικές<br/>(π.χ. Φύλο,<br/>Απουσία/Παρουσία<br/>παθολογικού<br/>χαρακτηριστικού,<br/>ναυτία/μη κατά την<br/>εγκυμοσύνη κλπ)]; A --> C[Ποσοτικές<br/>συνεχείς ή διακριτές<br/>(Ηλικία, ύψος,<br/>επίπεδα σακχάρου, αριθμός<br/>καταγμάτων κλπ)]; B --> D[Διατάξιμες: θέση σε μια<br/>λίστα (Ordinal. Οι<br/>κατηγορίες ιεραρχούνται,<br/>π.χ. κατηγορίες ΔΜΣ)]; B --> E[Ονομαστικές: αριθμοί που<br/>προσδιορίζουν κατηγορίες ή<br/>τύπους πραγμάτων (αλλά δεν έχει<br/>νόημα η διάταξή τους. Nominal)]; C --> F[Εξακρίβωση μεταβλητών υπό μελέτη:<br/>Αν αναγνωρίσετε τι είδους μεταβλητές<br/>έχετε, τότε μπορείτε να τις περιγράψετε.];
```

Ποιοτικές
(π.χ. Φύλο,
Απουσία/Παρουσία
παθολογικού
χαρακτηριστικού,
ναυτία/μη κατά την
εγκυμοσύνη κλπ)

Διατάξιμες: θέση σε μια
λίστα (Ordinal. Οι
κατηγορίες ιεραρχούνται,
π.χ. κατηγορίες ΔΜΣ)

Ονομαστικές: αριθμοί που
προσδιορίζουν κατηγορίες ή
τύπους πραγμάτων (αλλά δεν έχει
νόημα η διάταξή τους. Nominal)

Ποσοτικές
συνεχείς ή διακριτές
(Ηλικία, ύψος,
επίπεδα σακχάρου, αριθμός
καταγμάτων κλπ)

Εξακρίβωση μεταβλητών υπό μελέτη:
Αν αναγνωρίσετε τι είδους μεταβλητές
έχετε, τότε μπορείτε να τις περιγράψετε.

Ποιοτικές Μεταβλητές (διατάξιμες)

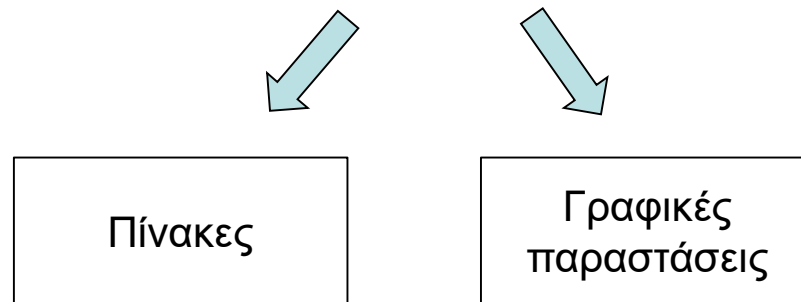
Κατηγορίες ατόμων	Τιμή Δείκτη Μάζας Σώματος	Κίνδυνος νοσηρότητας
Λιποβαρείς	<18,50	Χαμηλός
Φυσιολογικού βάρους	18,50 – 24,99	Συνήθης
Υπέρβαροι	≥ 25,00	
Προπαχυσαρκία	25,00 – 29,99	Αυξημένος
Παχύσαρκοι	30,00 +	Σοβαρά αυξημένος

Παρενέργειες ως αιματολογική τοξικότητα

Βαθμός	#Λευκών		Αιμοπετάλια
0 (απουσία)	≥4.000	και	≥90.000
1 ήπιες	3.000-4.000	ή	70.000-90.000
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

Βασικές μέθοδοι σύνοψης δεδομένων

1. Χρήση αριθμητικών ποσοτήτων (στατιστικές συναρτήσεις, π.χ. μέσος όρος)
2. Κατανομή (τρόπος με τον οποίο συνοψίζονται τα δεδομένα) – distribution



Levels and correlates of alpha-fetoprotein in normal pregnancies among Caucasian and Chinese women

Pagona Lagiou^{a,c}, Evangelia Samoli^a, Areti Lagiou^{a,b}, Mats Lambe^e,
Dimitrios Trichopoulos^{c,e}, Hans-Olov Adami^{c,e} and Chung-Cheng Hsieh^{c,e,d}

Table 1 Distribution of 212 Caucasian and 196 Chinese pregnant women by maternal and newborn characteristics^a

	Boston ^b	Shanghai ^b
Age (years)	31.0 (3.2)	25.2 (3.5)
Parity		
1	123 (58.0)	193 (98.5)
2	89 (42.0)	3 (1.5)
Duration of gestation (weeks)	40.0 (1.2)	40.0 (1.1)
Gender of offspring		
Male	112 (52.8)	122 (62.2)
Female	100 (47.2)	74 (37.8)
Nausea		
Yes (with/without vomiting)	166 (78.3)	119 (60.7)
No	46 (21.7)	77 (39.3)
Gestational week at 1st measurement	16.7 (1.1)	17.0 (1.7)
Gestational week at 2nd measurement	27.1 (1.8)	26.8 (1.3)
BMI before pregnancy (kg/m ²)	22.1 (3.1)	19.8 (2.0)

BMI, body mass index.

^aOnly pregnancies of healthy offsprings with duration of 37–42 (inclusive) weeks included.

^bIn parentheses: for quantitative variables, standard deviations; for qualitative variables, percentages.

lbw3- EX1.sav [DataSet1] - IBM SPSS Statistics Data Editor

	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	33	16	1	0	2082	1.68	72.56	

Γραφήματα για ποιοτικά δεδομένα

Κυκλικό διάγραμμα ή πίττα (pie chart)

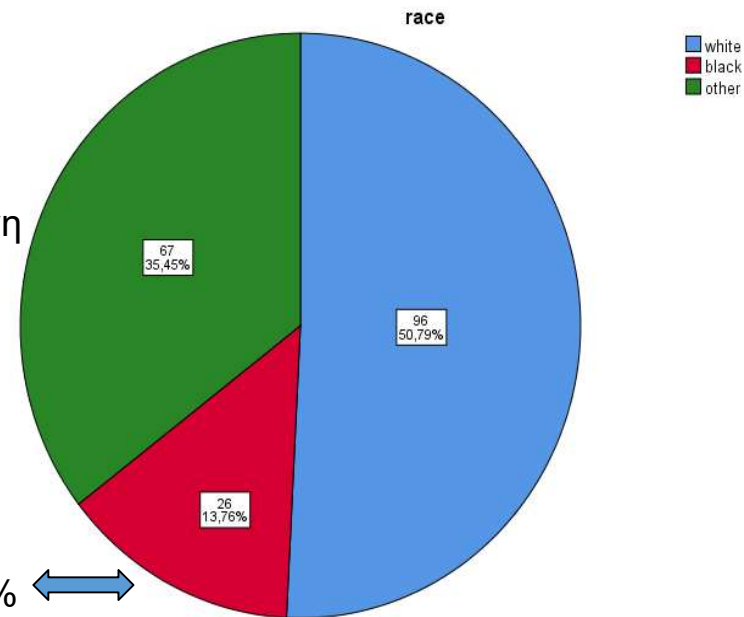
		race			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	white	96	50,8	50,8	50,8
	black	26	13,8	13,8	64,6
	other	67	35,4	35,4	100,0
Total		189	100,0	100,0	

Το εμβαδόν κάθε φέτας είναι ανάλογο του ποσοστού που αντιστοιχεί στη συγκεκριμένη κατηγορία.

Π.χ. κύκλος = 360° :

Κόκκινη φέτα (black): $26/189=0,138 = 13,8\%$ $\longleftrightarrow$

$0,138 \cdot 360^\circ = 50^\circ$.

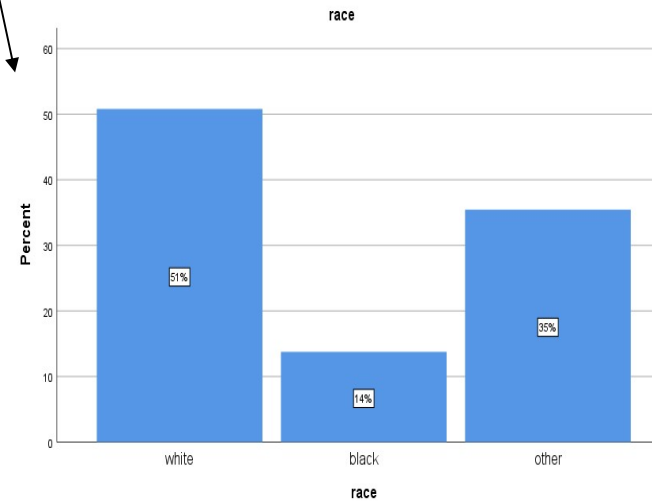
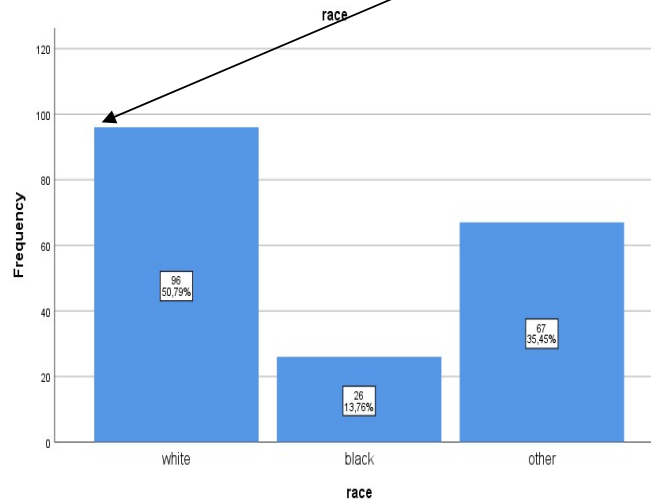


Γραφήματα για ποιοτικά δεδομένα

Ραβδόγραμμα (bar chart)

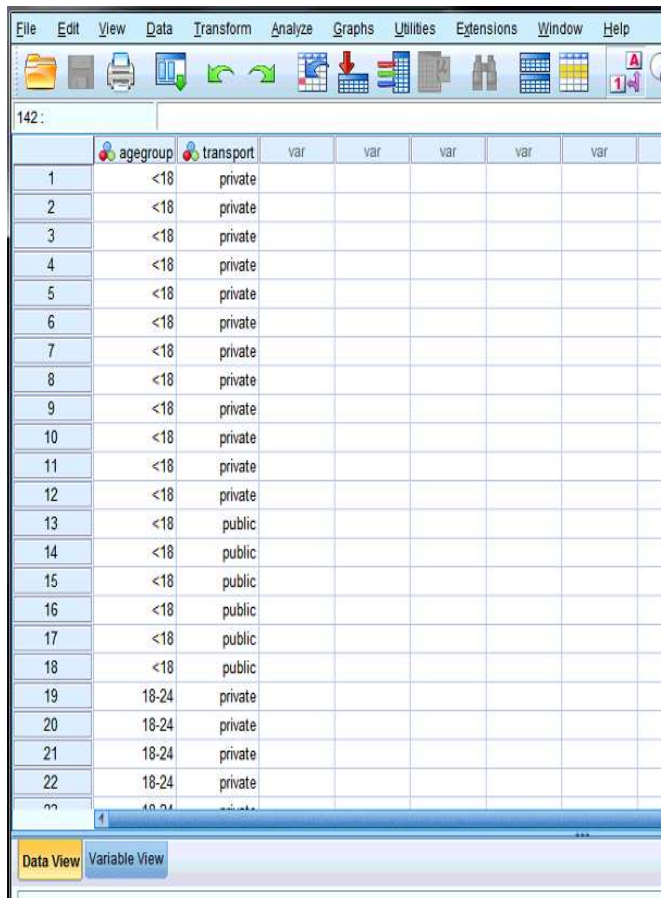
		race			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	white	96	50,8	50,8	50,8
	black	26	13,8	13,8	64,6
	other	67	35,4	35,4	100,0
	Total	189	100,0	100,0	

Προσοχή στον κατακόρυφο άξονα



Γραφήματα για ποιοτικά δεδομένα

Για σύγκριση υποομάδων (2 μεταβλητές)



	agegroup	transport	var	var	var	var	var
1	<18	private					
2	<18	private					
3	<18	private					
4	<18	private					
5	<18	private					
6	<18	private					
7	<18	private					
8	<18	private					
9	<18	private					
10	<18	private					
11	<18	private					
12	<18	private					
13	<18	public					
14	<18	public					
15	<18	public					
16	<18	public					
17	<18	public					
18	<18	public					
19	18-24	private					
20	18-24	private					
21	18-24	private					
22	18-24	private					

Έστω δείγμα 149 υπαλλήλων γραφείου με πληροφορίες για

- Χ1: τον τρόπο που πηγαίνουν στη δουλειά (ιδιωτικό ή δημόσιο μέσο) και
- Χ2: την ηλικιακή ομάδα που ανήκουν

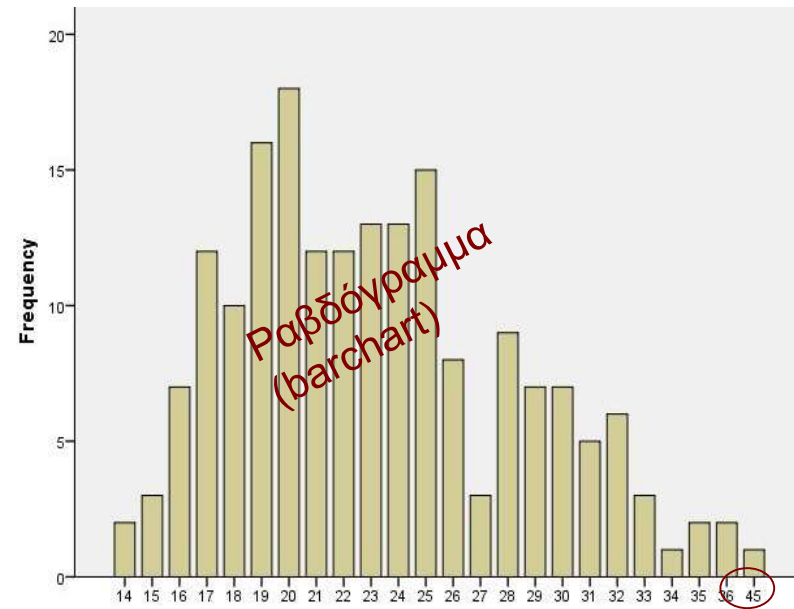
Table

Count

		transport		Total
		private	public	
agegroup	<18	12	6	18
	18-24	15	10	25
	25-39	25	18	43
	40-54	13	32	45
	55+	6	12	18
Total		71	78	149

Γραφήματα για ποσοτικά δεδομένα

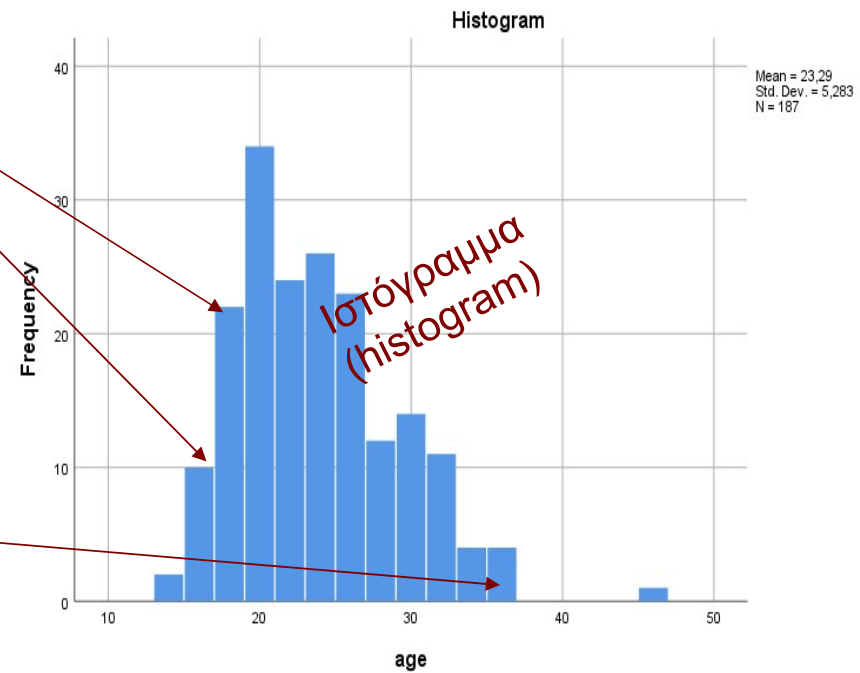
age					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	14	2	1.1	1.1	1.1
	15	3	1.6	1.6	2.7
	16	7	3.7	3.7	6.4
	17	12	6.4	6.4	12.8
	18	10	5.3	5.3	18.2
	19	16	8.6	8.6	26.7
	20	18	9.6	9.6	36.4
	21	12	6.4	6.4	42.8
	22	12	6.4	6.4	49.2
	23	13	7.0	7.0	56.1
	24	13	7.0	7.0	63.1
	25	15	8.0	8.0	71.1
	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	



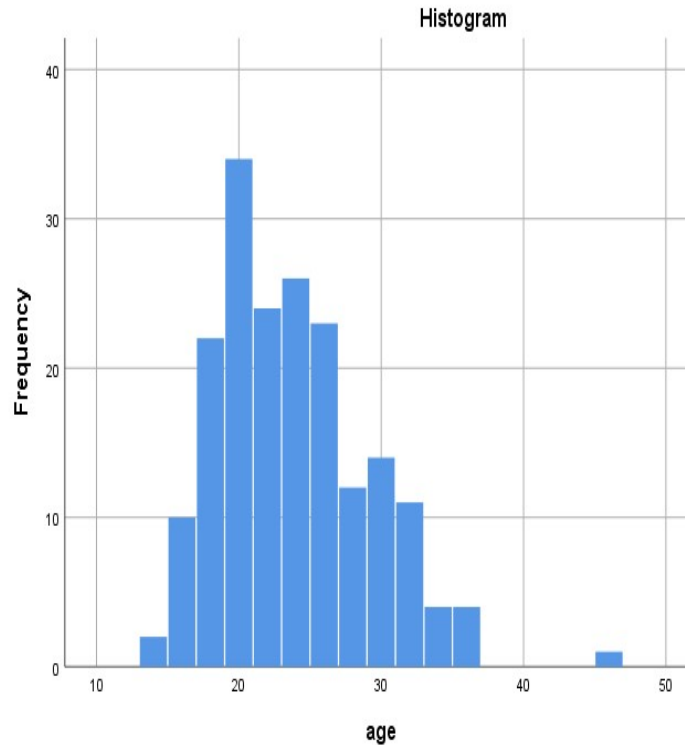
age
Το λαμβάνει σαν κατηγορία και το κολλά δίπλα στην προηγούμενη

Γραφήματα για ποσοτικά δεδομένα

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	



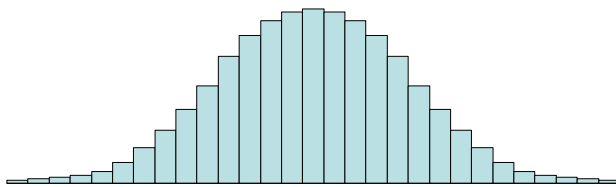
Ιστόγραμμα



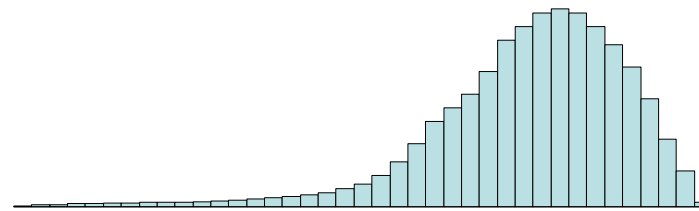
- ☐ Πόσο συμμετρικά είναι τα δεδομένα;
Πόσο διεσπαρμένα είναι τα δεδομένα;
- ☐ Υπάρχουν διαστήματα με υψηλή συγκέντρωση δεδομένων;
- ☐ Υπάρχουν κενά στα δεδομένα;
- ☐ Υπάρχουν παρατηρήσεις μακριά από τις υπόλοιπες (ακραίες - outliers);

Κατανομή Ποσοτικών Δεδομένων

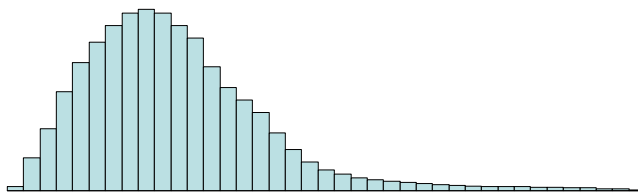
Συμμετρική, σχήμα καμπάνας
Κανονική (Gaussian) κατανομή



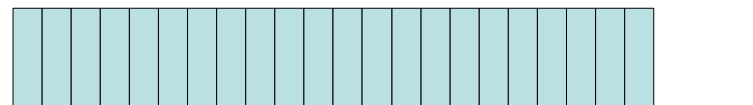
Αρνητικά λοξή
Λιγότερο συνηθισμένη



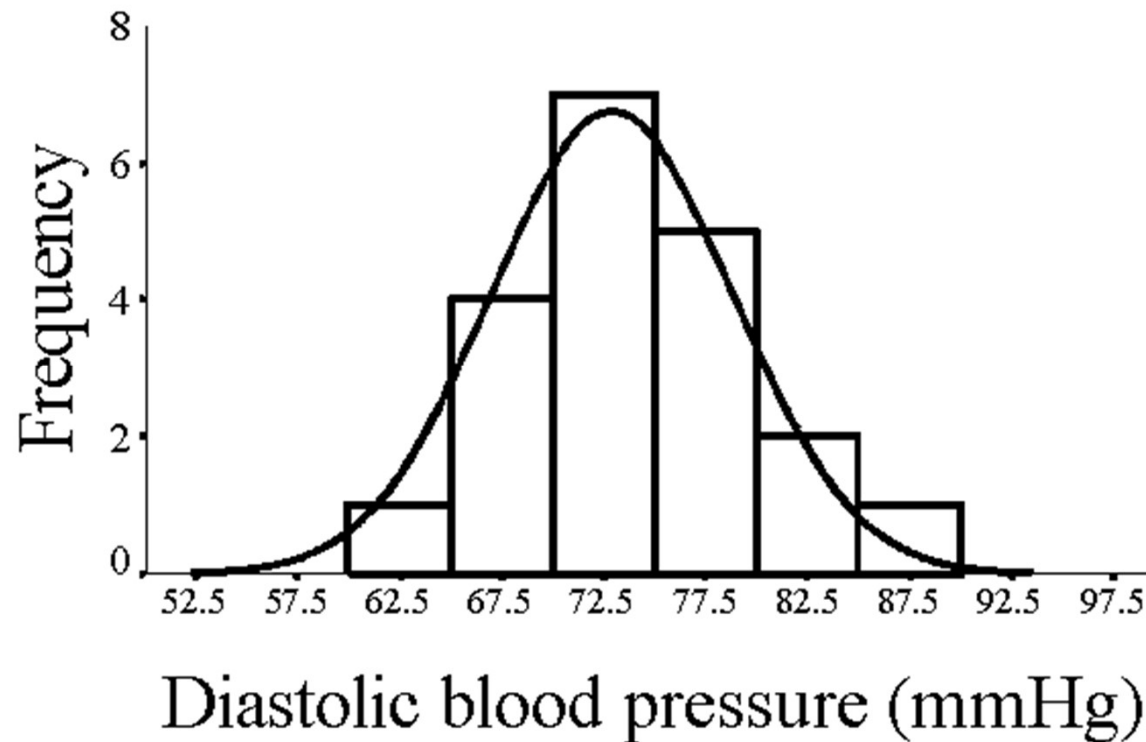
Θετικά λοξή
Σε εργαστηριακά δεδομένα
πχ. CD4 counts in HIV+



Ομοιόμορφη κατανομή
Ίδια πιθανότητα οποιασδήποτε τιμής
μέσα στο εύρος



Παράδειγμα κανονικής κατανομής



Πως συνεχίζουμε;

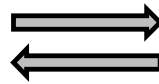
Ακολουθεί μία διάλεξη στην Περιγραφική Στατιστική

Και οι επόμενες διαλέξεις καλύπτουν την ενότητα της Στατιστικής Συμπερασματολογίας

Παράδειγμα



Καπνιστές και
μη-καπνιστές



Καταναλωτές και
μη-καταναλωτές καφέ

Σχετίζεται το κάπνισμα με την κατανάλωση καφέ;

Η κατάλληλη δοκιμασία είναι η χ^2

Παράδειγμα

Σύγκριση δύο περιοχών με χαμηλά και υψηλά, αντίστοιχα, επίπεδα μόλυβδου στο περιβάλλον με σκοπό τη διερεύνηση της πιθανής διαφοράς στο ανάστημα των παιδιών.

Διαφέρει το ανάστημα;

Αν δεν ήξερα ΚΑΘΟΛΟΥ στατιστική, τι θα έκανα?

Σαν πρώτο βήμα, θα υπολόγιζα το μέσο ύψος στις δύο περιοχές

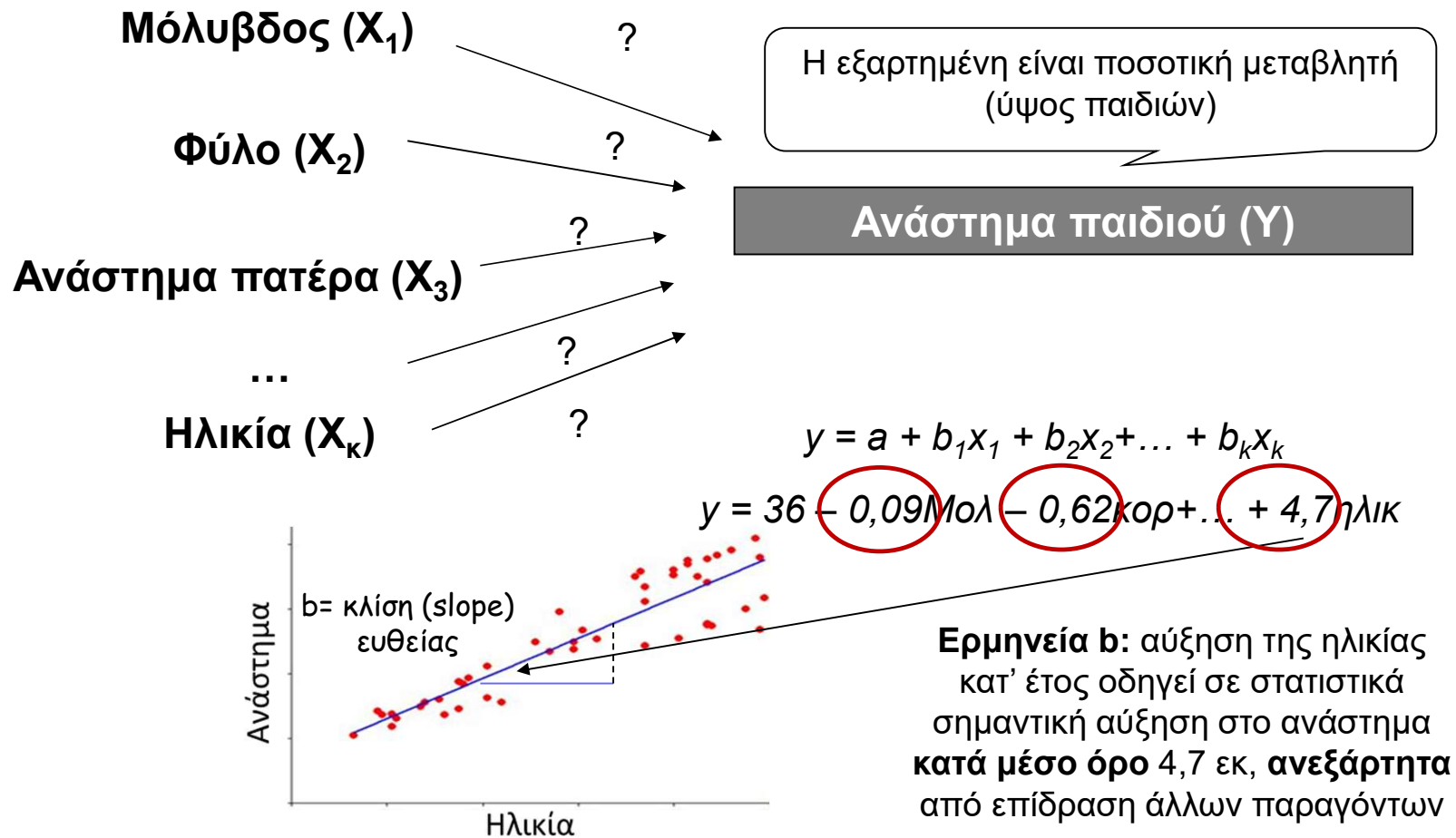
$$\bar{X}_1 = 123,7 \text{ cm}$$

$$\bar{X}_2 = 120,6 \text{ cm}$$

Είναι αυτή η διαφορά πραγματική; Οφείλεται στην τύχη?

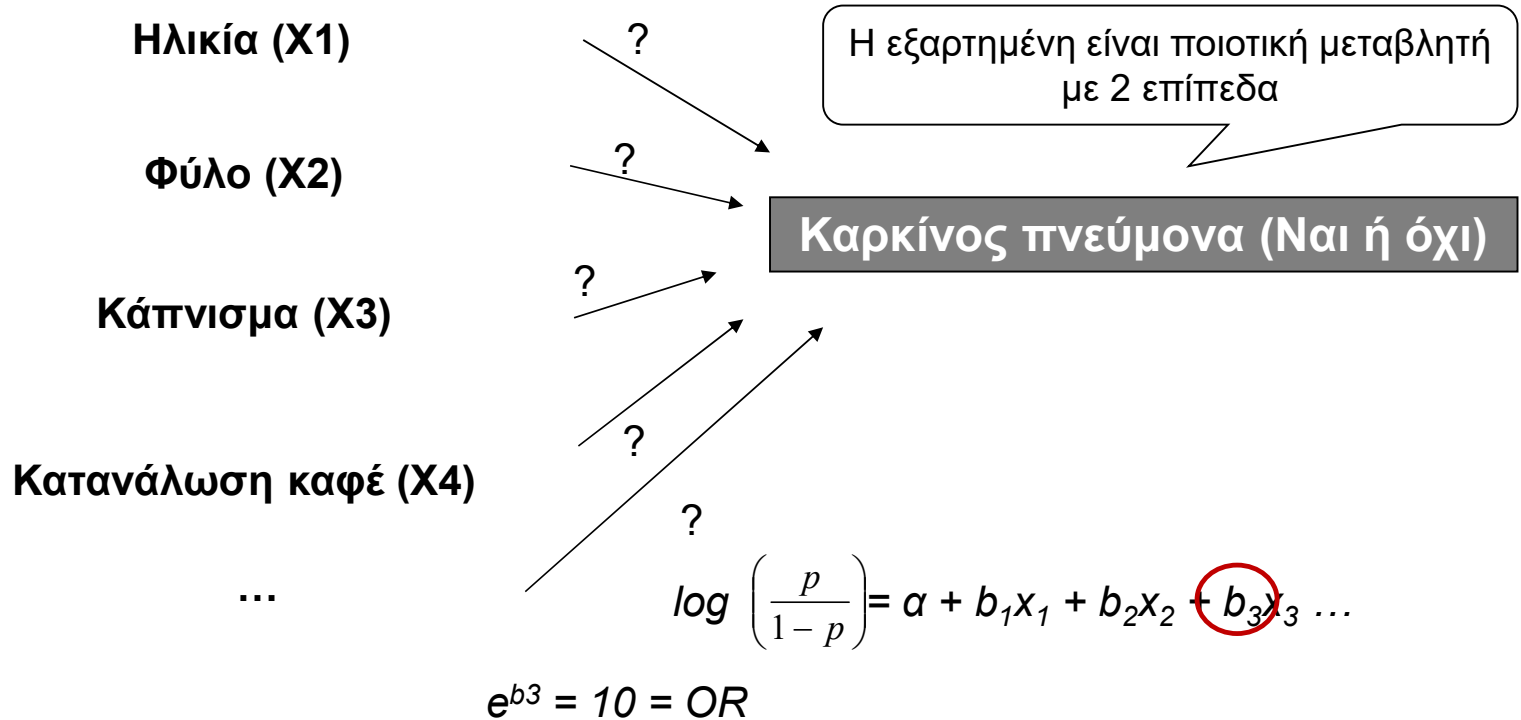
Η κατάλληλη δοκιμασία είναι το **t-test**

Από τι εξαρτάται το ανάστημα του παιδιού



Η κατάλληλη δοκιμασία είναι η **γραμμική εξάρτηση**

Από τι εξαρτάται η πιθανότητα εμφάνισης καρκίνου του πνεύμονα



Ερμηνεία OR: οι καπνιστές έχουν π.χ. 10 φορές > κίνδυνο εμφάνισης...

$b < 0$

→

$OR < 1$

Προστατευτικός παράγοντας

$b > 0$

→

$OR > 1$

Επιβαρυντικός παράγοντας

Η κατάλληλη δοκιμασία είναι η **λογαριθμική εξάρτηση**

Καλώς ήρθατε!

Καλή αρχή

Και υπομονή!

**Manuscript
submitted**



**Manuscript
accepted**

