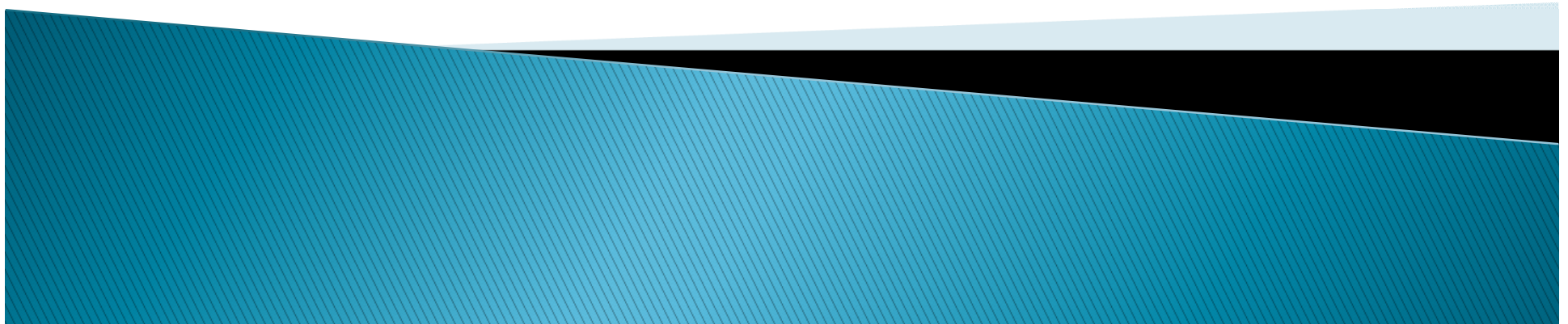


Βασικές αρχές στατιστικής

Νίκος Πανταζής
Βιοστατιστικός, PhD
ΕΔΙΠ Ιατρικής Σχολής ΕΚΠΑ
Εργαστήριο Υγιεινής, Επιδημιολογίας & Ιατρικής Στατιστικής



Εισαγωγή

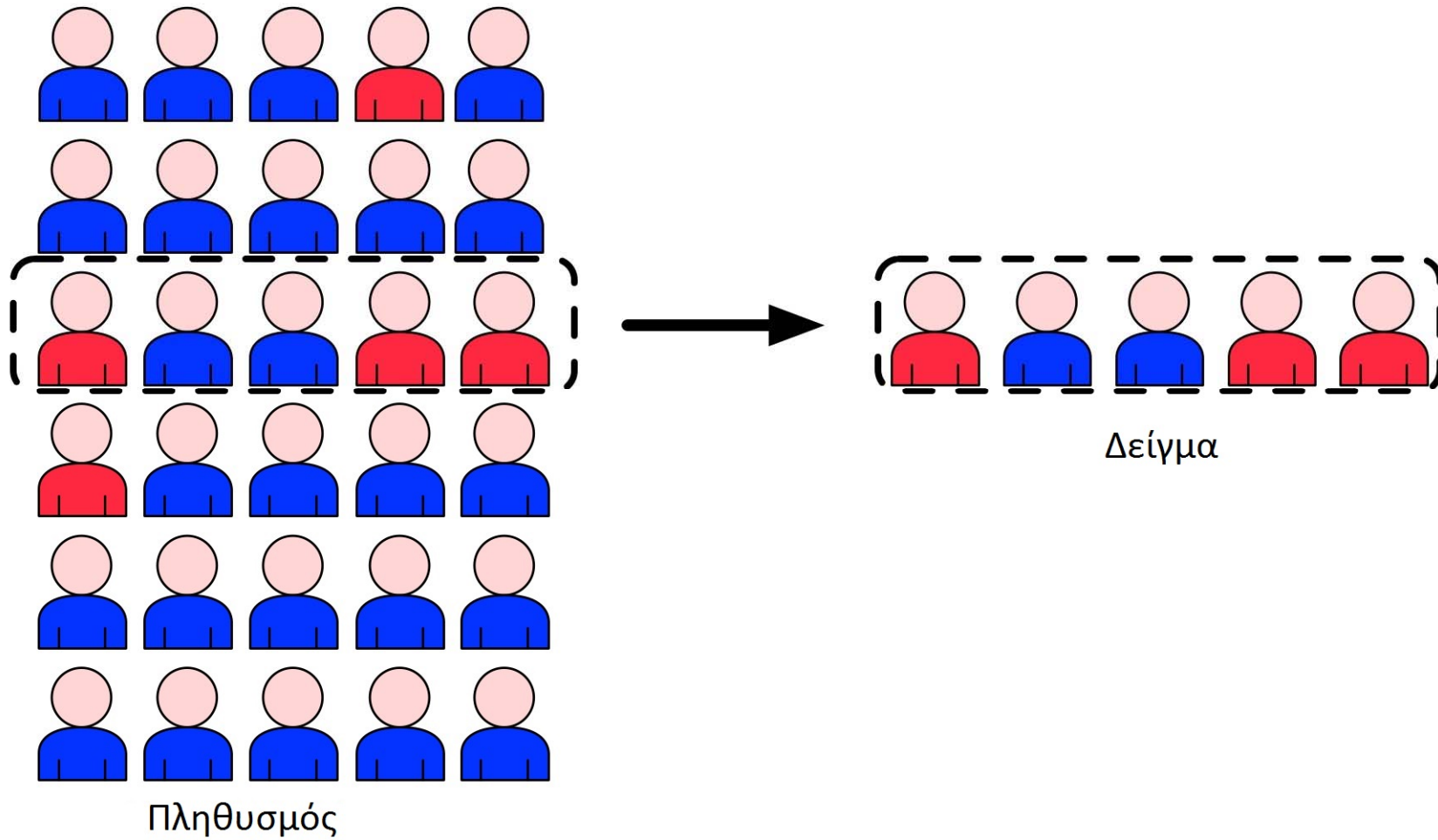
- ▶ Η έννοια της στατιστικής για τον απλό άνθρωπο:
 - Πολιτικές δημοσκοπήσεις (π.χ. το Χ κόμμα προηγείται κατά γ ποσοστιαίες μονάδες από το Ζ κόμμα)
 - Κοινωνικό-οικονομικές έρευνες [π.χ. Οι Έλληνες εργάζονται 5,5 ώρες περισσότερο κάθε εβδομάδα σε σχέση με τους Ευρωπαίους (κατά μέσο όρο)]
 - Έρευνες υγείας (π.χ. οι γυναίκες που καταναλώνουν κρασί διατρέχουν 44% χαμηλότερο κίνδυνο από καρδιαγγειακά νοσήματα σε σχέση με γυναίκες που δεν καταναλώνουν καθόλου κρασί)
- 

Έννοιες, ορισμός και κύριοι κλάδοι

- ▶ Στατιστικές έννοιες στα προηγούμενα παραδείγματα:
 - Περιγραφή-Σύνοψη αποτελεσμάτων από ένα δείγμα του πληθυσμού
 - Εκτίμηση παραμέτρων-συσχετίσεων και συμπεράσματα για τον πληθυσμό
- ▶ Επιστήμη που επιχειρεί να εξαγάγει έγκυρη γνώση χρησιμοποιώντας εμπειρικά δεδομένα
- ▶ Αντικείμενα έρευνας: συλλογή, επεξεργασία, παρουσίαση, ανάλυση, ερμηνεία δεδομένων με στόχο την εξαγωγή συμπερασμάτων ή/και λήψη αποφάσεων
- ▶ Κύριοι κλάδοι στατιστικής:
 - Περιγραφική στατιστική
 - Εκτιμητική/Συμπερασματολογία

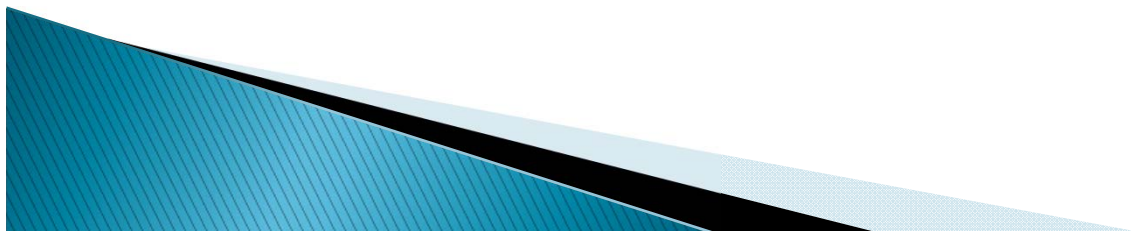


Πληθυσμός-Δείγμα



Πληθυσμός-Δείγμα

- ▶ Θεωρητικός πληθυσμός (theoretical/target population): αυτοί στους οποίους θέλουμε να γενικεύσουμε τα συμπεράσματά μας
- ▶ Ο πληθυσμός της μελέτης (study population): αυτοί στους οποίους έχουμε πρόσβαση
- ▶ Δειγματοληπτικό πλαίσιο (sampling frame): ο τρόπος με τον οποίο αποκτούμε πρόσβαση
- ▶ Δείγμα (sample): αυτοί που συμπεριλαμβάνονται στη μελέτη



Στάδια μιας τυπικής μελέτης

- ▶ Ορισμός ερευνητικού ερωτήματος
- ▶ Καθορισμός πληθυσμού
- ▶ Δειγματοληψία
- ▶ Περιγραφή/σύνοψη χαρακτηριστικών στο δείγμα
- ▶ Περαιτέρω στατιστική ανάλυση
- ▶ Εξαγωγή συμπερασμάτων για τον πληθυσμό
- ▶ Παρουσίαση-Σχολιασμός αποτελεσμάτων και συμπερασμάτων



ΣΤΑΤΙΣΤΙΚΗ

```
graph TD; A[ΣΤΑΤΙΣΤΙΚΗ] --> B[Περιγραφική στατιστική: Συνοπτική παρουσίαση δεδομένων]; A --> C[Στατιστική συμπερασματολογία: Συναγωγή συμπερασμάτων για τη σχέση π.χ. μίας μεταβλητής με μία άλλη];
```

Περιγραφική στατιστική:
Συνοπτική παρουσίαση
δεδομένων

π.χ.

- Μέσα επίπεδα CD4 σε ένα πληθυσμό ενηλίκων HIV+
- Ποσοστό θεραπευομένων HIV+ με μη ανιχνεύσιμο φορτίο

Στατιστική συμπερασματολογία:
Συναγωγή συμπερασμάτων για τη
σχέση π.χ. μίας μεταβλητής με
μία άλλη

π.χ.

- Διαφέρει η ανταπόκριση στη θεραπεία ανάλογα με τον υπότυπο του HIV;
- Μειώθηκαν τα ποσοστά καθυστερημένης διάγνωσης μέσα στην τελευταία 10-ετία;

Περιγραφική στατιστική

Μεταβλητές

```
graph TD; A[Μεταβλητές] --> B[Ποιοτικές<br/>(π.χ. Φύλο, Ομάδα<br/>κινδύνου κλπ)]; A --> C[Ποσοτικές<br/>(Ηλικία, BMI,<br/>CD4 count, Viral load<br/>κλπ)];
```

Ποιοτικές
(π.χ. Φύλο, Ομάδα
κινδύνου κλπ)

Ποσοτικές
(Ηλικία, BMI,
CD4 count, Viral load
κλπ)

Ποσοτικές μεταβλητές (μέτρα θέσης/διασποράς)

	Mean (SD)
Age	43.4 (11.4)
CD4 count (cells/microL)	639.5 (337.1)
	Median (IQR)
Age	42.4 (34.8, 50.3)
CD4 count (cells/microL)	604 (418, 822)



Different Means
Same Standard Deviation



Same Mean
Different Standard Deviations



Different Means
Different Standard Deviations

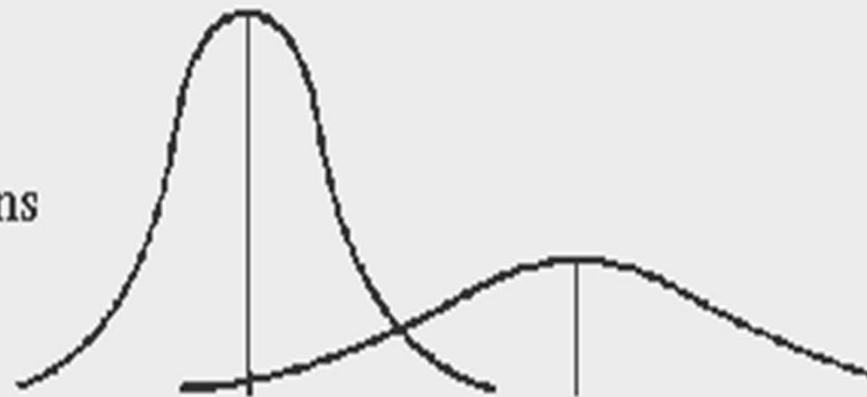
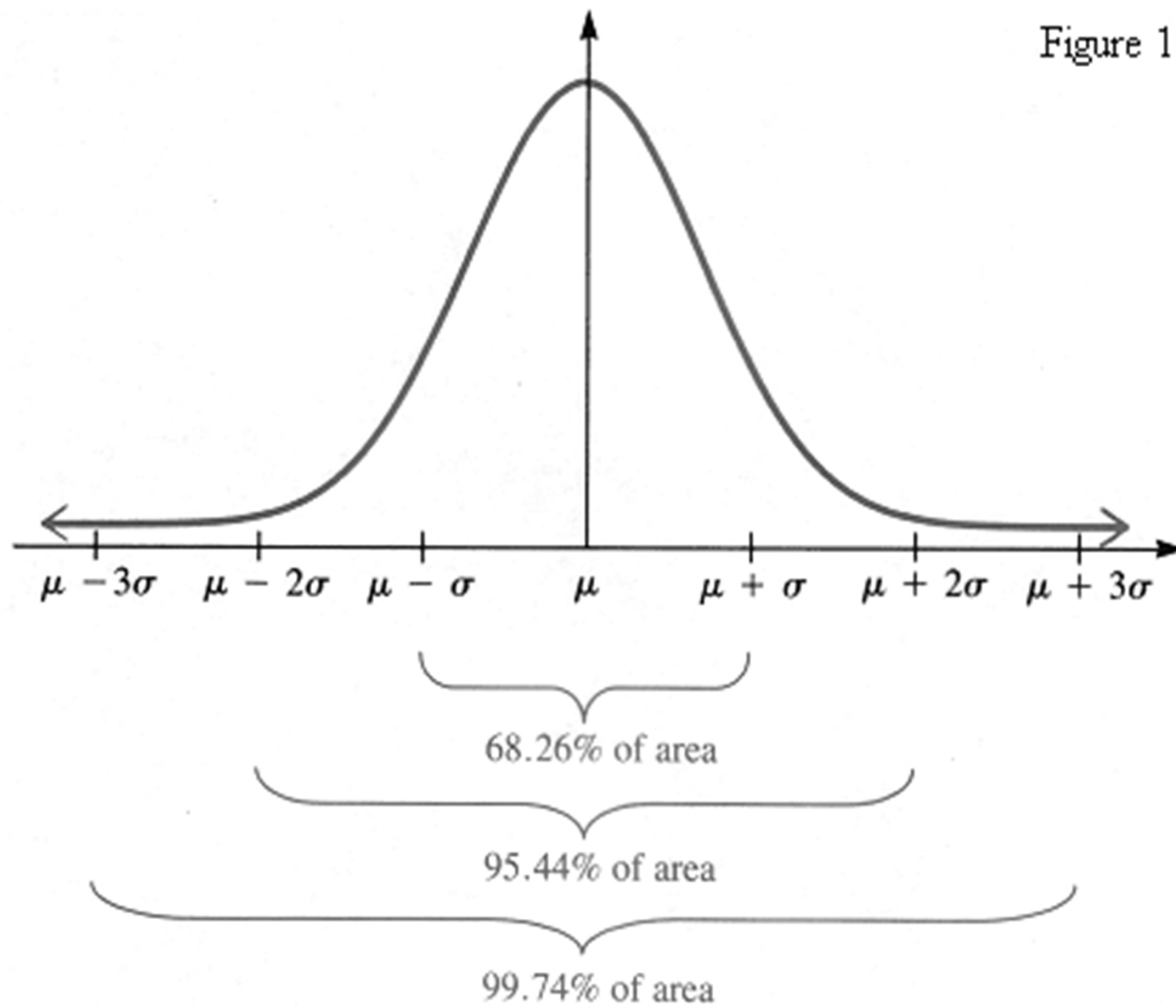
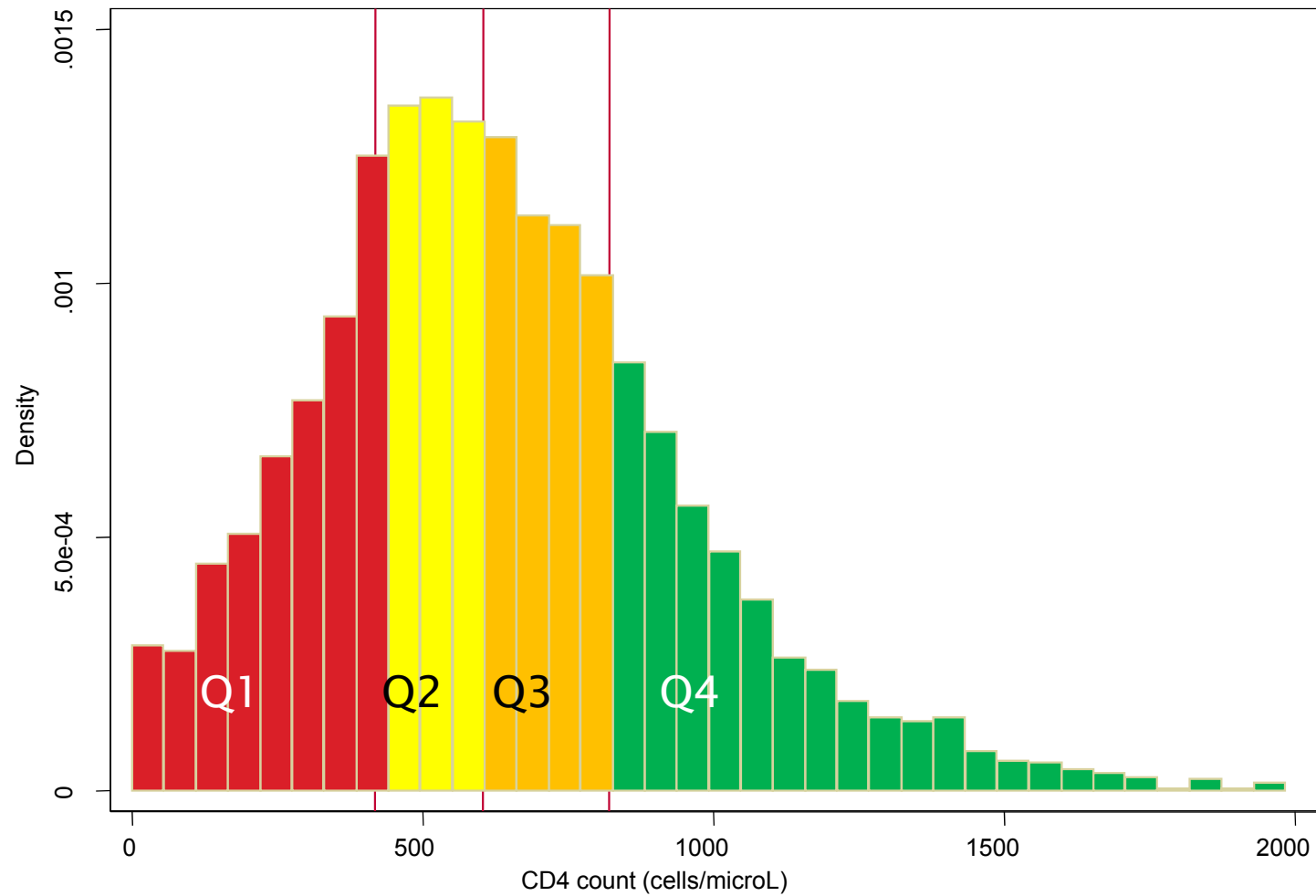


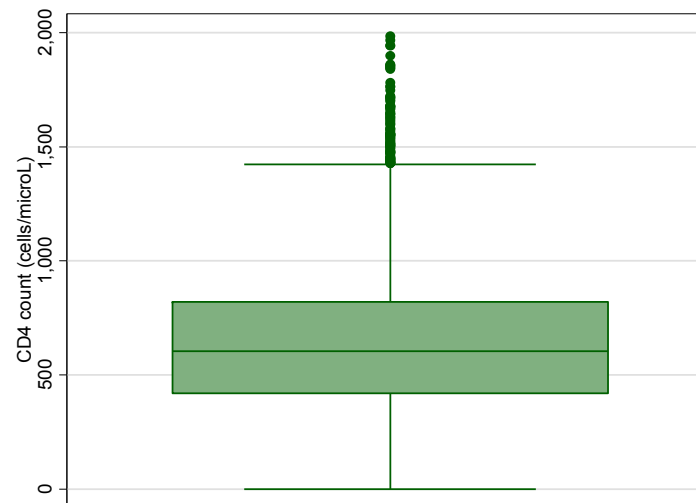
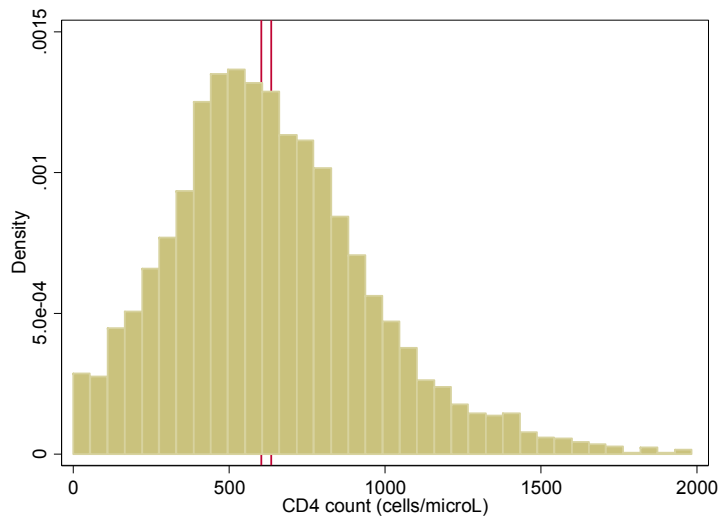
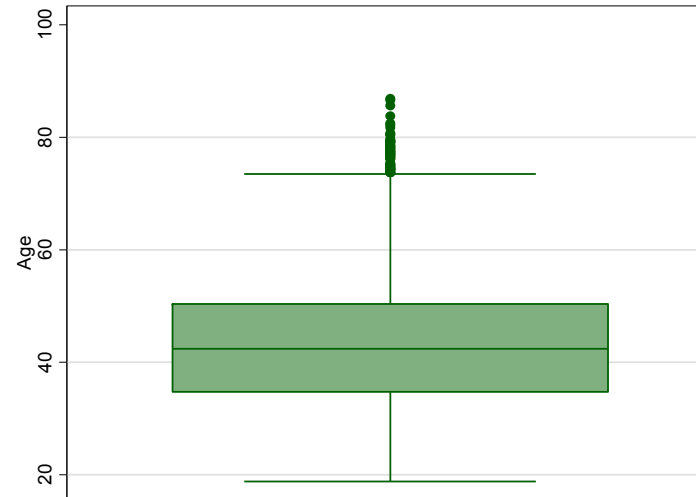
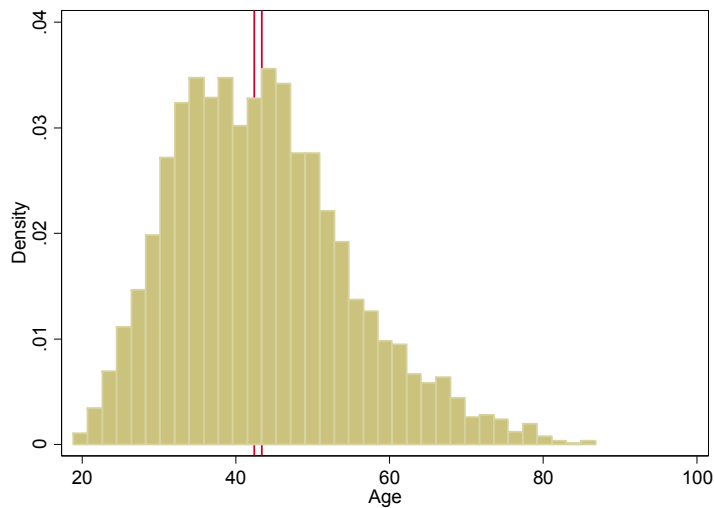
Figure 1



Τεταρτημόρια

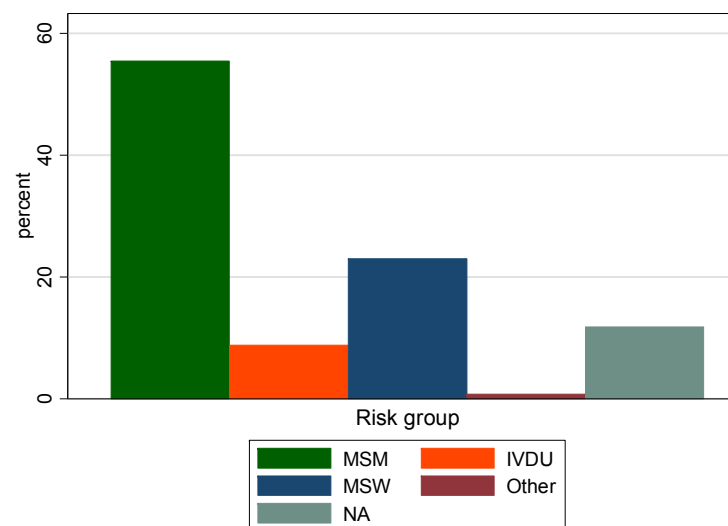
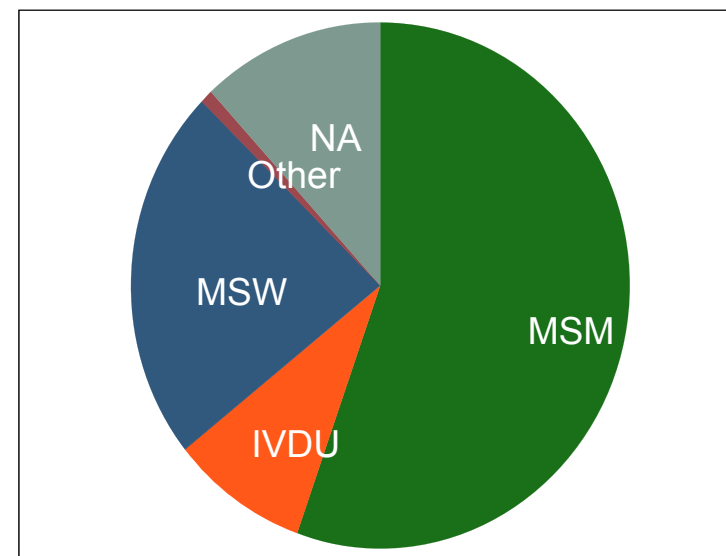


Ποσοτικές μεταβλητές (γραφήματα)

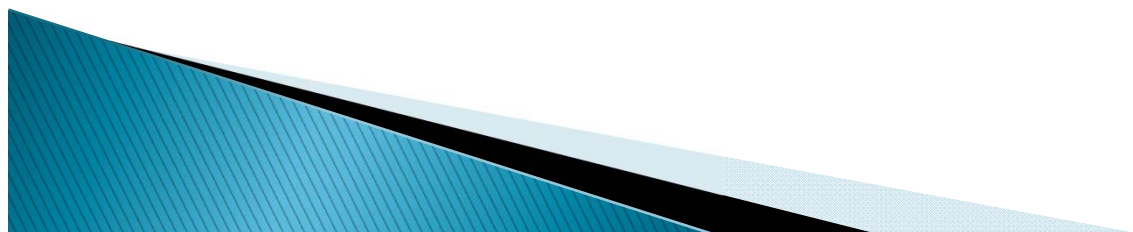
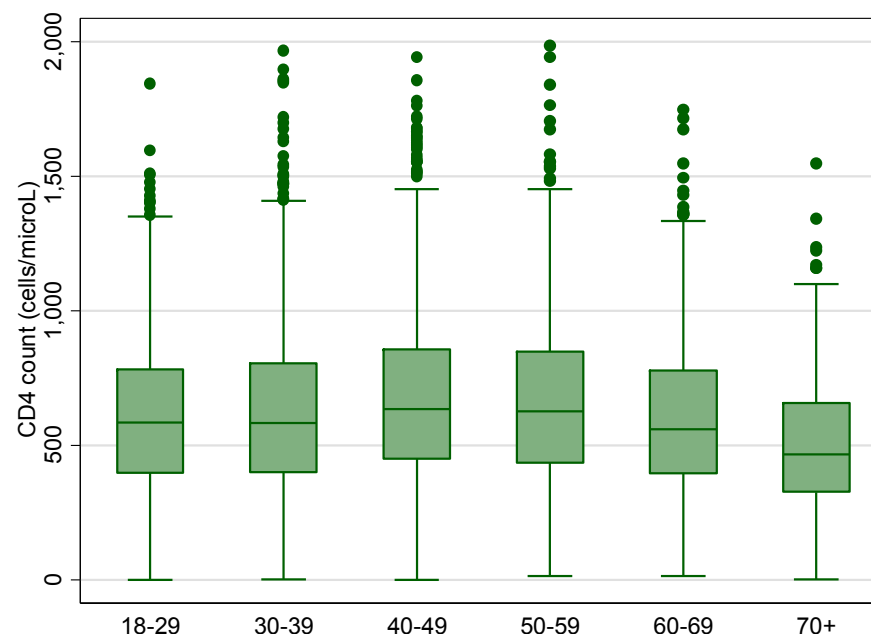
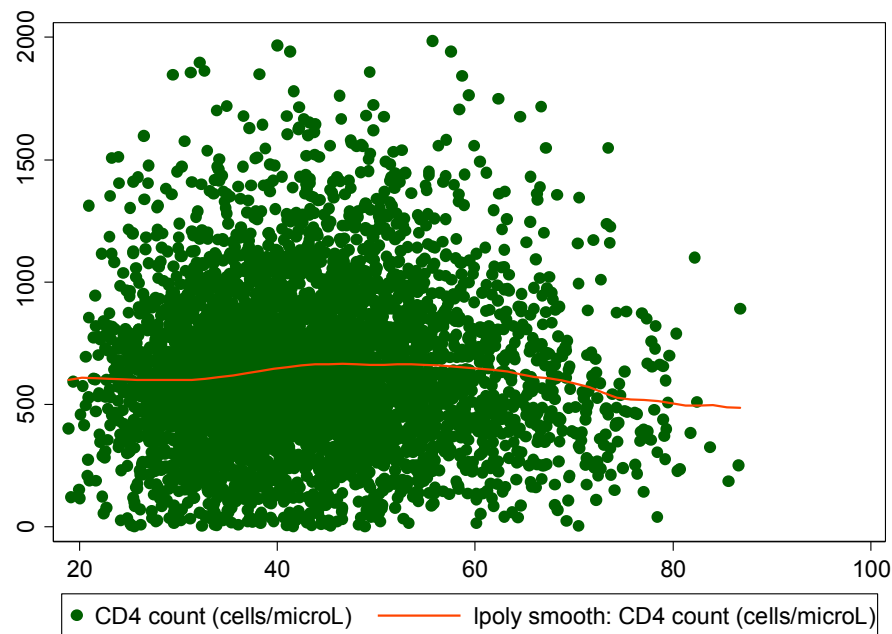


Ποιοτικές μεταβλητές

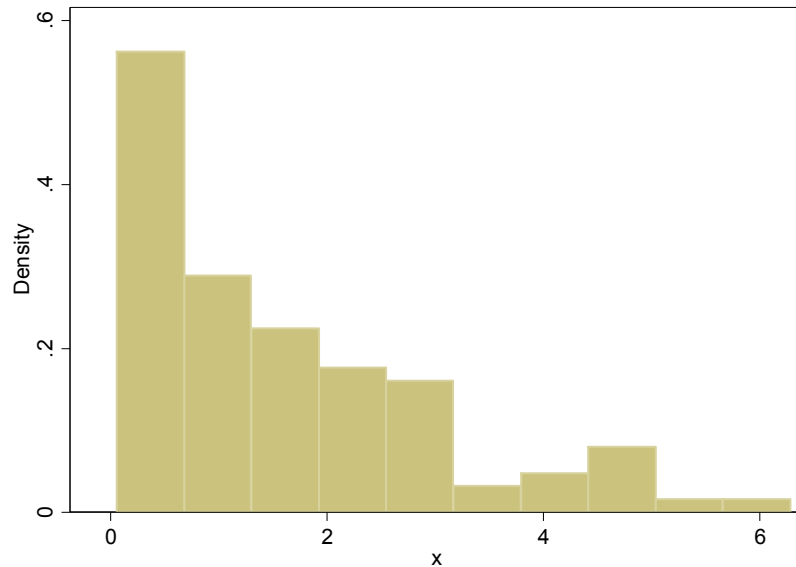
	N (%)
Risk group	
MSM	2722 (55.4)
IVDU	435 (8.9)
MSW	1130 (23.0)
Other	40 (0.8)
NA	583 (11.9)
Total	4910 (100.0)



Σχέσεις μεταξύ μεταβλητών



Ερώτηση 1



- ▶ Κάποια ποσοτική μεταβλητή έχει στο δείγμα μας την κατανομή που φαίνεται αριστερά.
- ▶ Τι θα χρησιμοποιούσατε για να την περιγράψετε;
 - A) Mean (SD)
 - B) Median (IQR)
 - Γ) N(%)



Ερώτηση 2

- ▶ Η κατανομή των CD4 σε ένα δείγμα έχει median (IQR) 604 (418, 822) cells/ μ L
- ▶ Αν σας έλεγαν ότι 27% των ατόμων του δείγματος είχαν $CD4 < 350$ cells/ μ L
 - Α) Θα σας φαινόταν λογικό
 - Β) Όχι, θα περιμένατε το ποσοστό αυτών με $CD4 < 350$ cells/ μ L να είναι μικρότερο από 27%
 - Γ) Όχι, θα περιμένατε το ποσοστό αυτών με $CD4 < 350$ cells/ μ L να είναι μεγαλύτερο από 27%

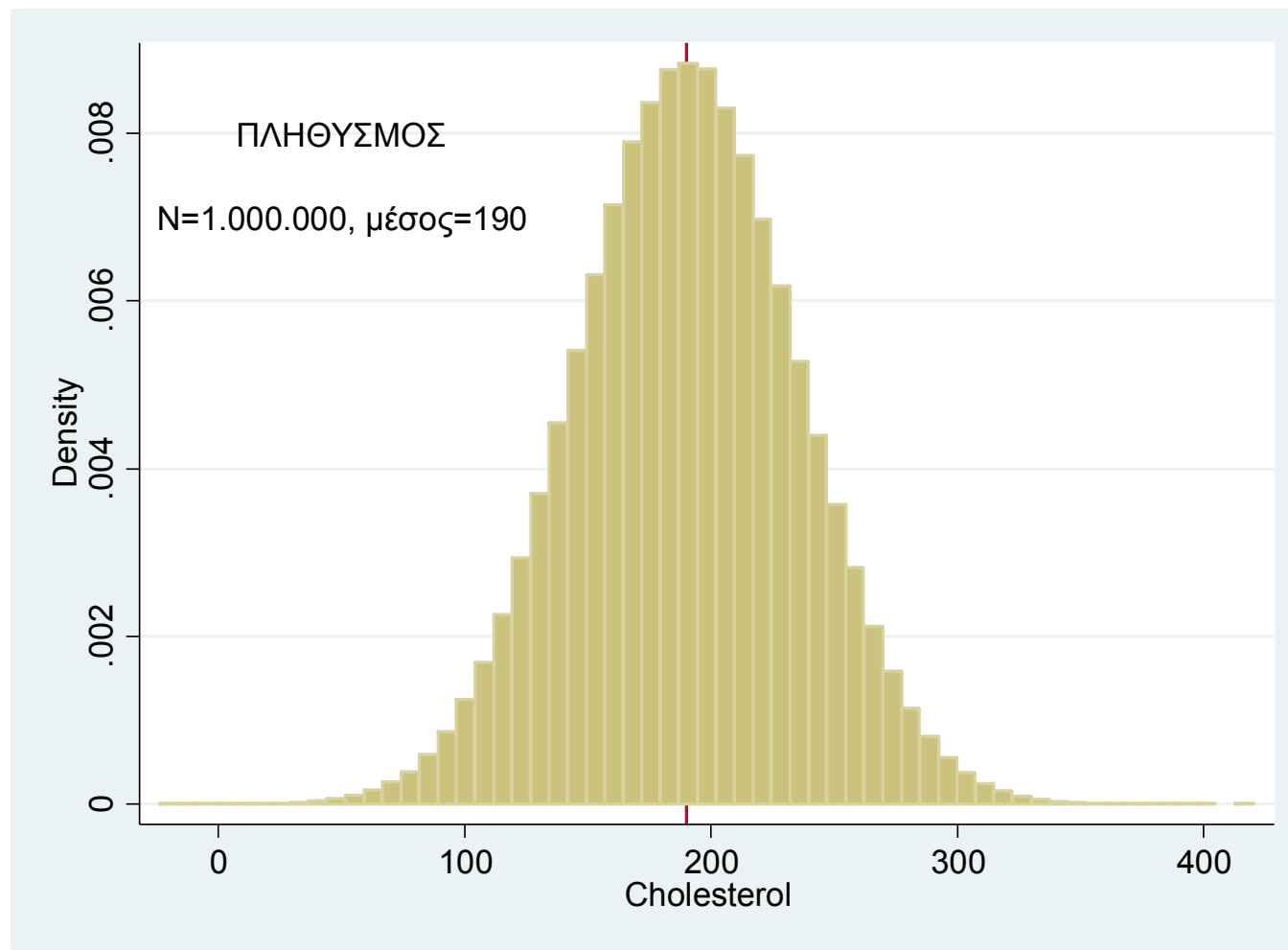


Εκτιμητική: από το δείγμα στον πληθυσμό

- ▶ Τα μέτρα θέσης (και διασποράς) ή τα ποσοστά μιας κατηγορικής μεταβλητής μας δίνουν πληροφορίες για το δείγμα
- ▶ Ωστόσο:
 - Η μέση τιμή μιας ποσοτικής μεταβλητής στο δείγμα είναι (υπό προϋποθέσεις) μια «καλή» εκτίμηση της μέσης τιμής του πληθυσμού
 - Τα ποσοστά μιας κατηγορικής στο δείγμα είναι (υπό προϋποθέσεις) «καλές» εκτιμήσεις των αντίστοιχων στον πληθυσμό
- ▶ Οι εκτιμήσεις έχουν πάντα αβεβαιότητα και αυτή πρέπει να δηλώνεται – 95% Δ.Ε.

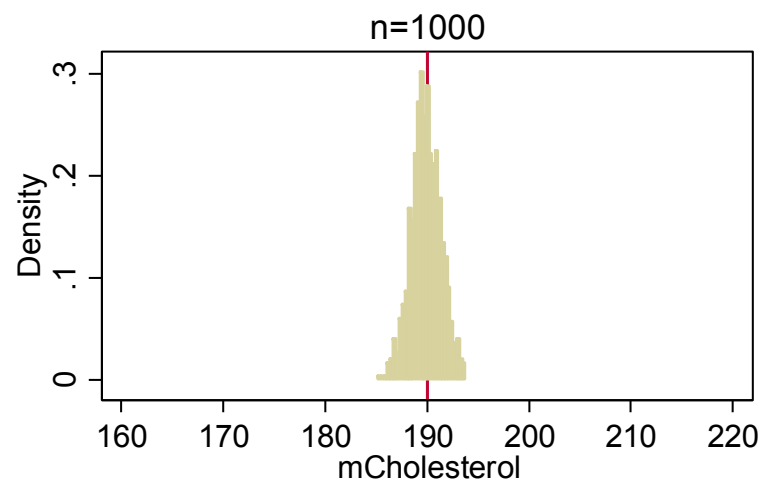
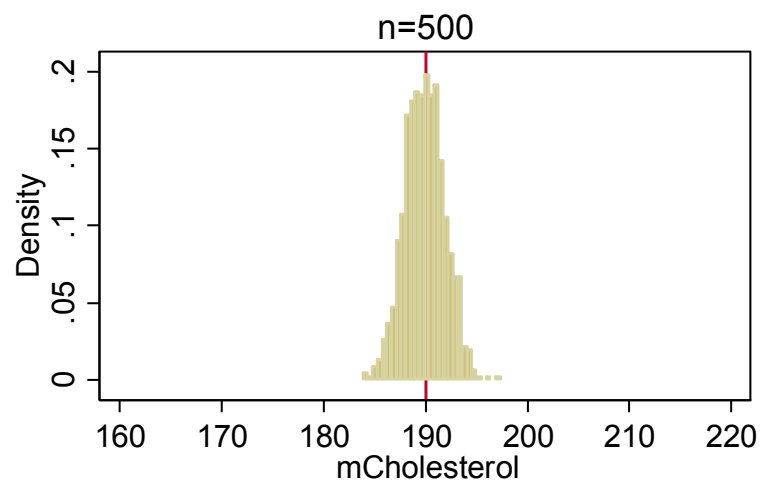
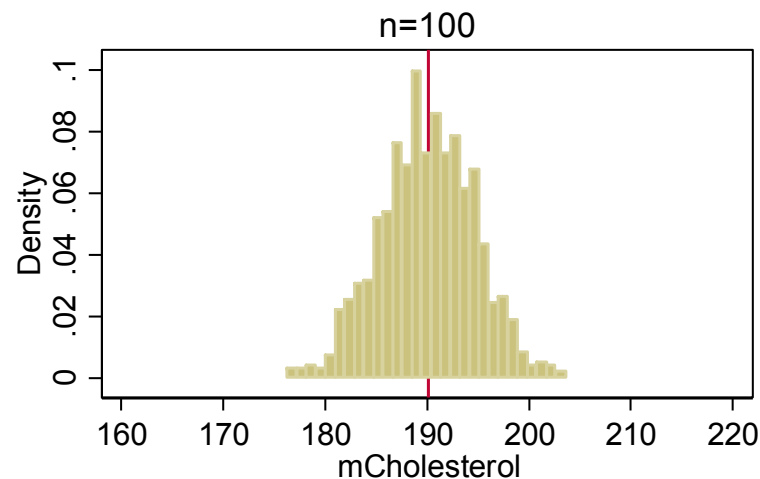
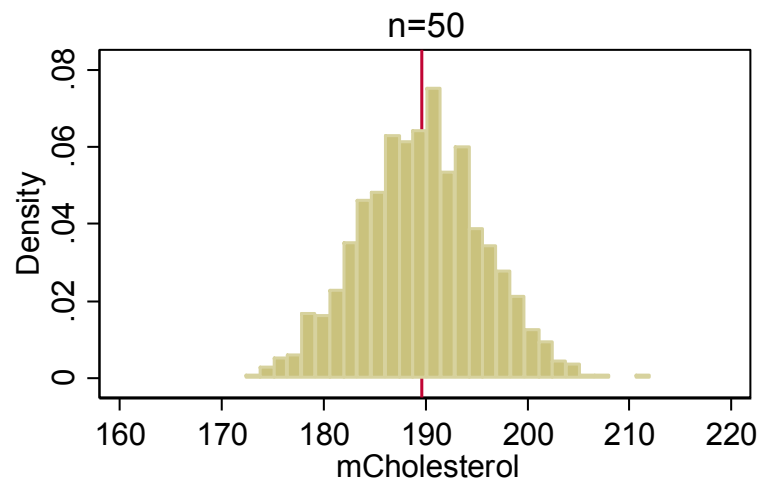


Παράδειγμα – Μέση τιμή



Κατανομή μέσης τιμής

Μέση τιμή 1000 δειγμάτων



Παραδείγματα για διαστήματα εμπιστοσύνης

n	mean	SE(mean)	95% CI
50	180.1	6.5	166.8–193.3
100	195.5	4.4	186.7–204.4
500	190.9	2.0	186.9–194.9
1000	188.8	1.4	186.0–191.6

Standard Error of the mean = $SD/\sqrt{n}$

95% C.I.: $\text{mean} \pm 1.96 * SE$



Ερώτηση 3

- ▶ Έχουμε δύο εντελώς τυχαία δείγματα A και B από έναν πληθυσμό, με μέγεθος δείγματος $N\alpha=100$ και $N\beta=400$, αντίστοιχα. Εκτιμούμε το μέσο κάποιας ποσοτικής μεταβλητής από αυτά τα δείγματα.
- ▶ Περιμένετε το πλάτος του 95% Διαστήματος Εμπιστοσύνης για το μέσο να είναι:
 - Α) Περίπου 4 φορές μικρότερο στο δείγμα B
 - Β) Περίπου 2 φορές μικρότερο στο δείγμα B
 - Γ) Περίπου το ίδιο και στα δύο δείγματα
 - Δ) Περίπου 2 φορές μεγαλύτερο στο δείγμα B
 - Ε) Περίπου 4 φορές μεγαλύτερο στο δείγμα B



Στατιστική συμπερασματολογία

- ▶ Έλεγχος υποθέσεων
 - π.χ. διαφέρουν τα CD4 κατά τη διάγνωση μεταξύ ανδρών και γυναικών;
- ▶ Μηδενικές υποθέσεις (H_0)
 - Η μέση τιμή των CD4 στον πληθυσμό των νεοδιαγνωσθέντων είναι ίδια για άνδρες και γυναίκες
- ▶ Γενικός τρόπος ελέγχου (κλασική στατιστική)
 - Έστω ότι ισχύει η H_0 (ίδια CD4 σε άνδρες γυναίκες). Ποια η πιθανότητα να δω, εκ τύχης, διαφορές σαν κι αυτές που έχω στο δείγμα μου (ή μεγαλύτερες): p-value
 - Αν $p\text{-value} < 0.05 \Rightarrow$ απορρίπτω H_0 και αποδέχομαι την εναλλακτική (H_a) δηλαδή ότι τα CD4 ανδρών διαφέρουν από αυτά των γυναικών στον πληθυσμό των νεοδιαγνωσθέντων



Σφάλματα τύπου I και II

- ▶ Σφάλμα τύπου I (α): η απόρριψη της μηδενικής υπόθεσης ενώ αυτή ισχύει
 - Δέχομαι ότι η θεραπεία A υπερτερεί της B ενώ οι θεραπείες είναι ισοδύναμες
- ▶ Σφάλμα τύπου II (β): η αποδοχή της μηδενικής υπόθεσης ενώ αυτή δεν ισχύει
 - Δέχομαι ότι οι θεραπείες A και B δε διαφέρουν ενώ η A υπερτερεί της B (ή το αντίστροφο)
- ▶ Ισχύς ($\text{Power} = 1 - \beta$): Η πιθανότητα να απορρίψω τη μηδενική υπόθεση όταν αυτή δεν ισχύει
 - Μέτρο της ικανότητας μιας μελέτης να ανιχνεύσει μια πραγματική διαφορά/σχέση (εντονότερη από κάποιο ελάχιστο «κλινικά σημαντικό» επίπεδο)



Σχετικά με τα p-values

- ▶ Τα p-values δεν δίνουν κάποια πληροφορία σχετικά με την εναλλακτική υπόθεση (π.χ. ύπαρξη πραγματικής διαφοράς/σχέσης)
- ▶ Τα p-values δείχνουν τη συμβατότητα των δεδομένων με τη μηδενική υπόθεση
- ▶ Στατιστική σημαντικότητα $\neq$ κλινική σημαντικότητα
- ▶ Με μεγάλα μεγέθη δείγματος μπορούμε να απορρίψουμε οποιαδήποτε υπόθεση
- ▶ Το $p=0.049$ δε διαφέρει ιδιαίτερα από το $p=0.051$
- ▶ Συνιστάται πάντα: συνυπολογισμός διαστημάτων εμπιστοσύνης, σχεδιασμού μελέτης, προηγούμενης γνώσης, βιολογικών μηχανισμών, αποτελεσμάτων άλλων σχετικών μελετών ...



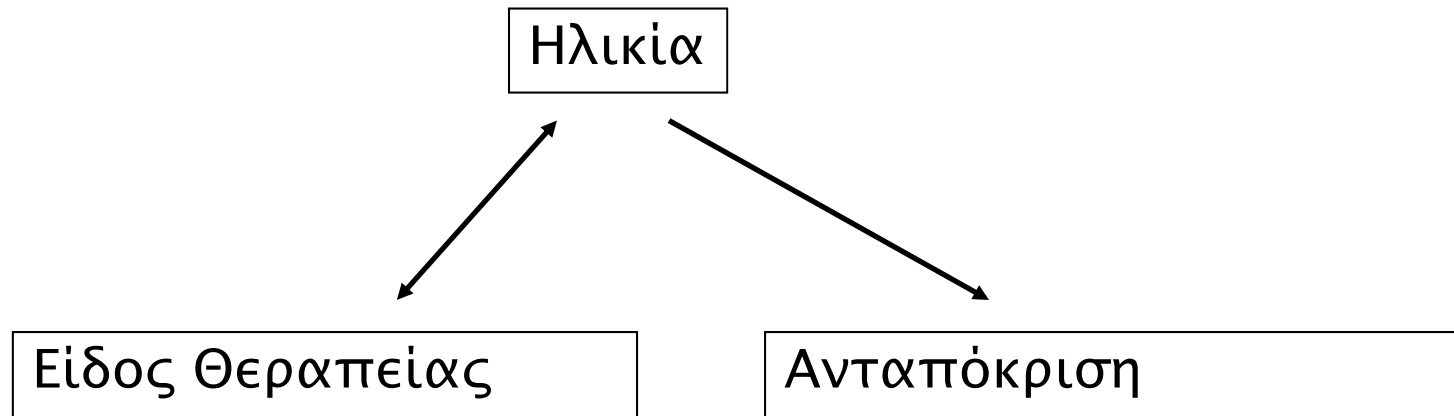
Τυπικά παραδείγματα «συσχετίσεων» προς διερεύνηση

- ▶ Διαφέρουν τα επίπεδα μιας ποσοτικής μεταβλητής μεταξύ δύο ή περισσότερων ομάδων;
 - (*CD4 ~ φύλο*)
 - ▶ Διαφέρει η κατανομή μιας ποιοτικής μεταβλητής σε δύο ή περισσότερες ομάδες;
 - (*HIV subtype ~ Ομάδα κινδύνου*)
 - ▶ Η αύξηση μιας ποσοτικής μεταβλητής συνδέεται με αύξηση (ή μείωση μιας άλλης);
 - (*επίπεδα Χοληστερόλης ~ έτη λήψης θεραπείας*)
 - ▶ Διαφέρει η πιθανότητα παρουσίας κάποιου χαρακτηριστικού μεταξύ ομάδων ή ανάλογα με τα επίπεδα μιας ποσοτικής μεταβλητής;
 - (*Καθυστερημένη διάγνωση ~ Τρόπος μεταδοσης, ηλικία, φύλο, έτος διάγνωσης*)
- 

Συνήθεις τρόποι ανάλυσης

«Συσχέτιση»	Μέθοδοι ανάλυσης
Ποσοτική ~ Κατηγορική (2 επίπεδα)	t-test (<i>Mann-Whitney U-test</i>)
Ποσοτική ~ Κατηγορική (n επίπεδα)	One-way ANOVA (<i>Kruskal-Wallis test</i>)
Ποσοτική ~ Ποσοτική	Pearson (r) / <i>Spearman (ρ)</i> correlation
Ποσοτική ~ Ποσοτικές και κατηγορικές μεταβλητές	Linear regression (median regression)
Ποσοτική ~ Ποσοτικές και κατηγορικές μεταβλητές (επαναλαμβανόμενες μετρήσεις)	Mixed (random effects) linear models
Κατηγορική ~ Κατηγορική	Chi-square test (Fisher's exact)
Δίτιμη κατηγορική ~ Ποσοτικές και κατηγορικές μεταβλητές	Logistic regression (conditional logistic regression σε matched-case control μελέτες)
Χρόνος έως εμφάνιση event ~ Κατηγορική	Log-rank test, Kaplan-Meier
Χρόνος έως εμφάνιση event ~ Ποσοτικές και κατηγορικές μεταβλητές	Cox models

Συγχυτικοί παράγοντες



Στην πιο πάνω περίπτωση η ηλικία μπορεί να δρα ως συγχυτικός παράγοντας στη σχέση Θεραπείας-ανταπόκρισης αν σχετίζεται με το είδος της θεραπείας και ταυτόχρονα είναι ανεξάρτητος προγνωστικός παράγοντας για ανταπόκριση



Παραδείγματα κατεύθυνσης συγχυτικού σφάλματος

- Σύγκριση θεραπειών A και B (έκθεση)
- Έκβαση: πιθανότητα 5-ετούς επιβίωσης
- Συγχυτικός παράγοντας: ηλικία (χαμηλότερες πιθανότητες επιβίωσης σε πιο ηλικιωμένους & η θεραπεία A έτεινε να δίνεται σε πιο ηλικιωμένους)

Πραγματική σχέση:
A αποτελεσματικότερη της B

Θεραπεία A Μέση ηλικία 60 έτη	Θεραπεία B Μέση ηλικία 45 έτη
↓	↓
Επιβίωση 70%	Επιβίωση 70%

Συμπέρασμα:
A ισοδύναμη με B
(ψευδώς αρνητικό εύρημα)

Πραγματική σχέση:
A ισοδύναμη της B

Θεραπεία A Μέση ηλικία 60 έτη	Θεραπεία B Μέση ηλικία 45 έτη
↓	↓
Επιβίωση 50%	Επιβίωση 70%

Συμπέρασμα :
B αποτελεσματικότερη της A
(ψευδώς θετικό εύρημα)

Πραγματική σχέση:
B αποτελεσματικότερη της A

Θεραπεία A Μέση ηλικία 60 έτη	Θεραπεία B Μέση ηλικία 45 έτη
↓	↓
Επιβίωση 30%	Επιβίωση 80%

Συμπέρασμα :
B αποτελεσματικότερη της A
(ορθή κατεύθυνση αλλά
υπερεκτίμηση της υπεροχής της B)

Ερώτηση 4

- ▶ Ενδιαφερόμαστε να διαπιστώσουμε εάν η ηλικία των νεοδιαγνωσθέντων HIV+ ήταν διαφορετική για το διάστημα 2005-09 σε σχέση με το διάστημα 2010-15 και διαθέτουμε αντίστοιχα τυχαία δείγματα ικανοποιητικού μεγέθους
- ▶ Τι έλεγχο/μοντέλο/μέθοδο θα χρησιμοποιούσαμε για να απαντήσουμε στο πιο πάνω ερώτημα;
 - A) Chi-square test
 - B) Pearson correlation coefficient
 - Γ) t-test (δύο ανεξάρτητων δειγμάτων)
 - Δ) Linear regression



Ερώτηση 5

- ▶ Ενδιαφερόμαστε να διαπιστώσουμε εάν η ηλικία των νεοδιαγνωσθέντων HIV+ ήταν διαφορετική για το διάστημα 2005-09 σε σχέση με το διάστημα 2010-15, έχοντας σταθμίσει για πιθανές αλλαγές στην κατανομή της ομάδας μετάδοσης στα δύο διαστήματα
- ▶ Τι έλεγχο/μοντέλο/μέθοδο θα χρησιμοποιούσαμε για να απαντήσουμε στο πιο πάνω ερώτημα;
 - A) Chi-square test
 - B) Pearson correlation coefficient
 - Γ) t-test (δύο ανεξάρτητων δειγμάτων)
 - Δ) Linear regression



Student's t-test

VOLUME VI

MARCH, 1908

No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

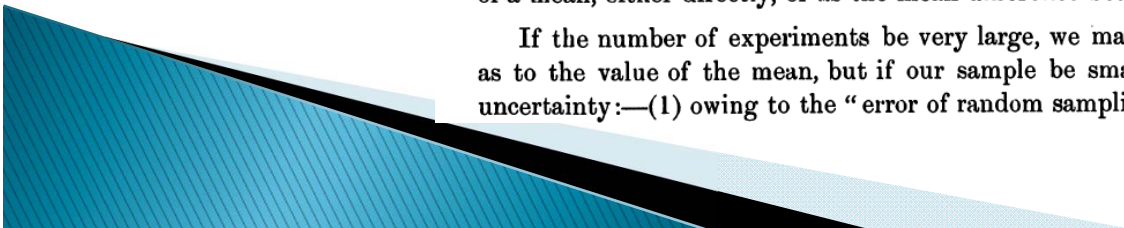
By STUDENT.

Introduction.

ANY experiment may be regarded as forming an individual of a "population" of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong. In a great number of cases the question finally turns on the value of a mean, either directly, or as the mean difference between the two quantities.

If the number of experiments be very large, we may have precise information as to the value of the mean, but if our sample be small, we have two sources of uncertainty:—(1) owing to the "error of random sampling" the mean of our series



t-test (ηλικία ~ φύλο)

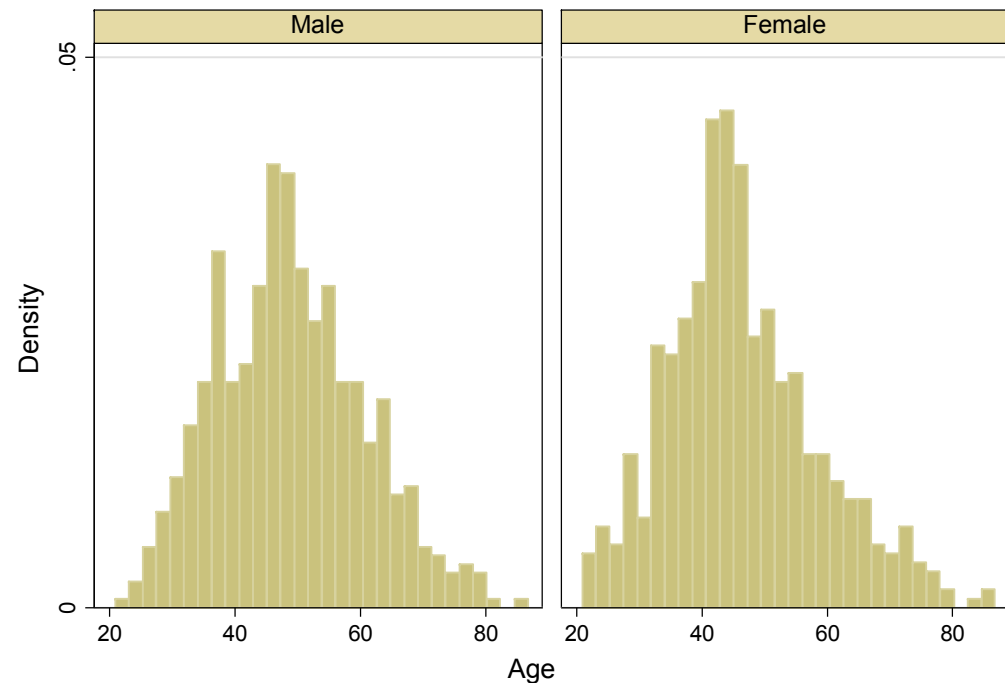
Μηδενική υπόθεση

$$H_0: \mu_1 = \mu_2$$

Εναλλακτική υπόθεση

$$H_1: \mu_1 \neq \mu_2$$

$$t = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{SE_1^2 + SE_2^2}}$$



Graphs by Sex

t-test (2 ανεξάρτητων δειγμάτων)

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
Male	576	48.86173	.4875729	11.70175	47.90409	49.81937
Female	554	46.48874	.5116005	12.04165	45.48382	47.49365
combined	1130	47.69833	.3547043	11.92355	47.00238	48.39429
diff		2.372993	.7063325		.9871198	3.758866

diff = mean(Male) - mean(Female) t = 3.3596
Ho: diff = 0 degrees of freedom = 1128

Ha: diff < 0
Pr(T < t) = 0.9996

Ha: diff != 0
Pr(|T| > |t|) = 0.0008

Ha: diff > 0
Pr(T > t) = 0.0004

Η ηλικία διέφερε στατιστικά σημαντικά ($p < 0.001$) μεταξύ ανδρών και γυναικών. Οι άνδρες είχαν κατά μέσο όρο 2.37 (95% Δ.Ε. 0.99 - 3.76) έτη μεγαλύτερη ηλικία.

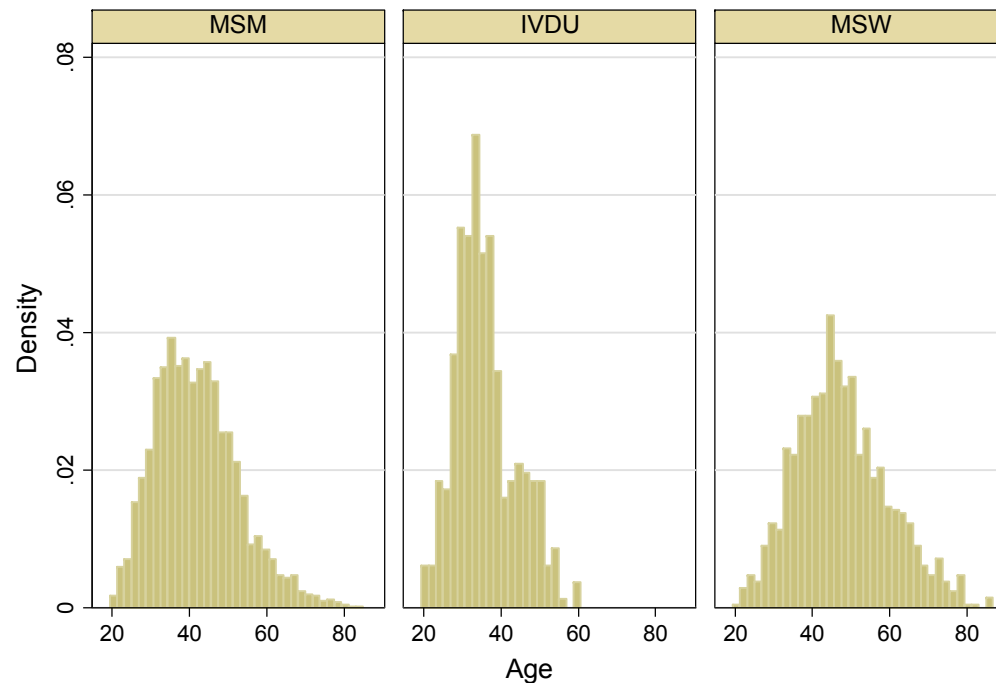
One-way ANOVA (ηλικία ~ομάδα μετάδοσης)

Μηδενική υπόθεση

$$H_0: \mu_1 = \mu_2 = \mu_3$$

Εναλλακτική υπόθεση

$$H_1: \mu_1 \neq \mu_2 \text{ ή } \mu_1 \neq \mu_3 \text{ ή } \mu_2 \neq \mu_3$$



Graphs by Risk group

One-way ANOVA

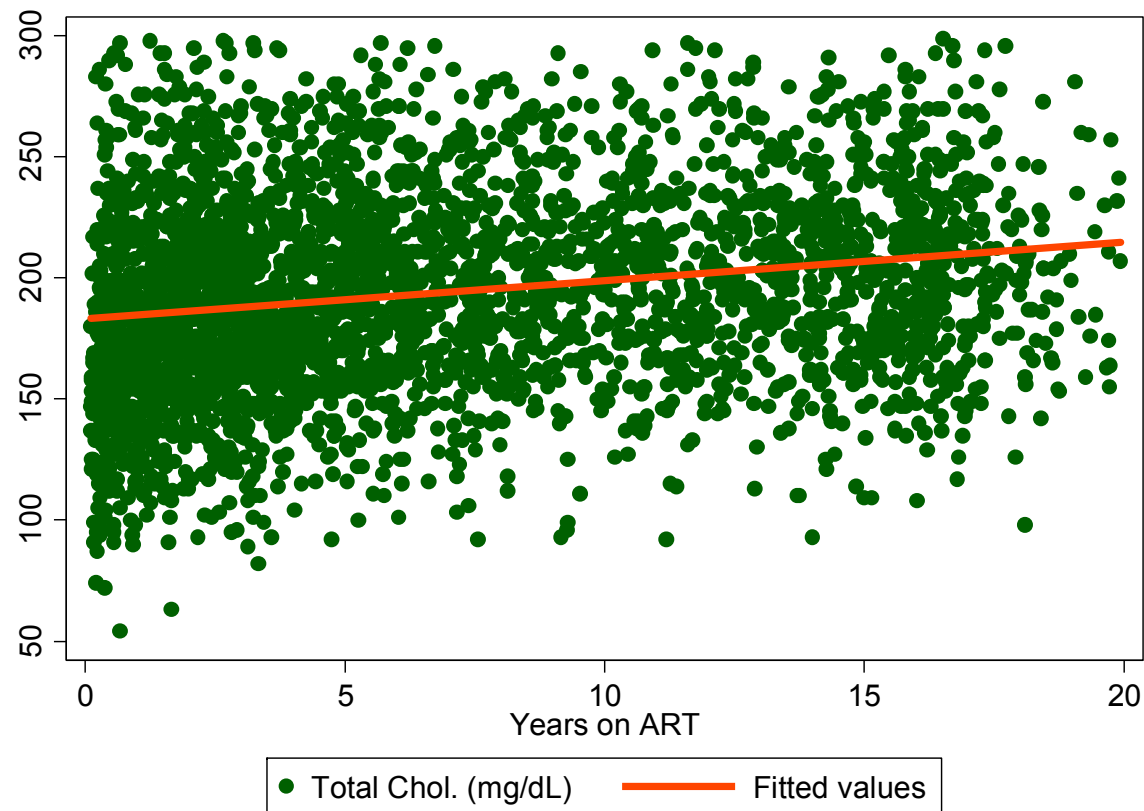
Risk group	Summary of Age		
	Mean	Std. Dev.	Freq.
MSM	41.955144	10.67977	2722
IVDU	35.905434	7.7130766	435
MSW	47.698332	11.923553	1130
Total	42.855116	11.290933	4287

Source	Analysis of Variance				
	SS	df	MS	F	Prob > F
Between groups	49720.4693	2	24860.2347	214.43	0.0000
Within groups	496680.959	4284	115.938599		
Total	546401.428	4286	127.485168		

Bartlett's test for equal variances: $\chi^2(2) = 103.7463$ Prob> $\chi^2 = 0.000$

Η ηλικία διέφερε στατιστικά σημαντικά ($p < 0.001$) μεταξύ των τριών ομάδων μετάδοσης.

Συσχέτιση (correlation) και εξάρτηση/παλινδρόμηση (regression)



$$r = \frac{\sum_i X_i Y_i - \frac{\sum_i X_i \sum_i Y_i}{n}}{\sqrt{(\sum_i X_i^2 - \frac{(\sum_i X_i)^2}{n})(\sum_i Y_i^2 - \frac{(\sum_i Y_i)^2}{n})}}$$

Pearson's $r=0.208$,
 $p<0.001$

$$Y=b_0+b_1*X+\epsilon$$

$$\text{Χοληστερόλη}=b_0 + b_1*\text{Έτη ART} + \epsilon$$

Source	SS	df	MS
Model	243813.922	1	243813.922
Residual	5413197.17	3332	1624.60899
Total	5657011.09	3333	1697.27305

Number of obs = 3334
 F(1, 3332) = 150.08
 Prob > F = 0.0000
 R-squared = 0.0431
 Adj R-squared = 0.0428
 Root MSE = 40.306

chol_2013	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
yrs_on_ART_2013	1.587603	.1295946	12.25	0.000	1.333509	1.841696
_cons	183.1414	1.142038	160.36	0.000	180.9022	185.3805

Βρέθηκε ασθενής θετική συσχέτιση μεταξύ επιπέδων ολικής χοληστερόλης και χρόνου λήψης ART (Pearson $r=0.208$; $p<0.001$).

Ανάλυση παλινδρόμησης έδειξε ότι η εκτιμώμενη (95% Δ.Ε.) αύξηση της ολικής χοληστερόλης ήταν 1.59 (1.33–1.84) mg/dL για κάθε έτος αύξησης στο χρόνο λήψης ART

Chi-square (AIDS και ομάδα μετάδοσης)

Risk group	AIDS (CDC C)		Total
	No	Yes	
MSM	1,322	258	1,580
	1,302.1	277.9	1,580.0
	83.67	16.33	100.00
IVDU	25	8	33
	27.2	5.8	33.0
	75.76	24.24	100.00
MSW	626	155	781
	643.7	137.3	781.0
	80.15	19.85	100.00
Total	1,973	421	2,394
	1,973.0	421.0	2,394.0
	82.41	17.59	100.00

Pearson chi2(2) = 5.4843 Pr = 0.064
Fisher's exact = 0.059

*Βρέθηκε ενδεικτική ($p=0.059$)
διαφορά στα ποσοστά HIV+
ατόμων με κλινικό AIDS
μεταξύ των τριών ομάδων
μετάδοσης.*



Logistic regression (AIDS ~ Ομάδα μετάδοσης)

- $\log[\text{Pr}/(1 - \text{Pr})] = b_0 + b_1 * X_1 + \dots$
- $\log[\text{Odds of AIDS}] = b_0 + b_1 * \text{IVDU} + b_2 * \text{MSW}$
- $\exp(b_1) = \text{Odds Ratio of AIDS (IVDU vs. MSM)}$
- $\exp(b_2) = \text{Odds Ratio of AIDS (MSW vs. MSM)}$

Covariate	Odds Ratio	Std. Err.	P> z	95% Conf. Interval
Risk group				
MSM*	1			
IVDU	1.640	0.675	0.230	(0.731 to 3.676)
MSW	1.269	0.143	0.035	(1.017 to 1.582)

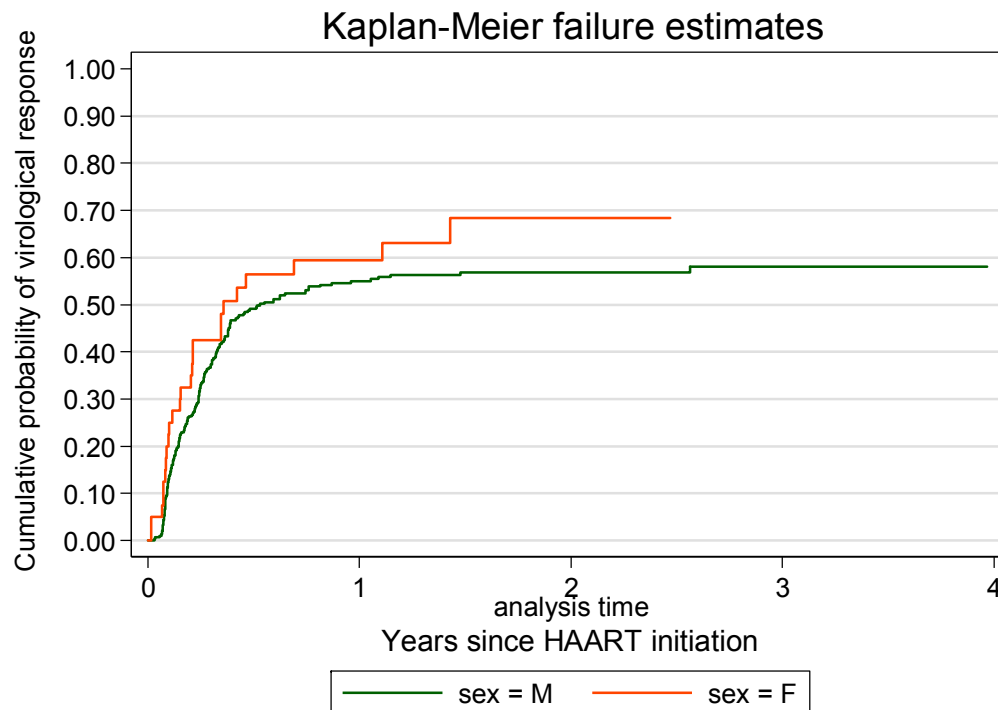
* Baseline category

Η εκτιμώμενη πιθανότητα ανάπτυξης AIDS ήταν αυξημένη κατά 27% (95% Δ.Ε. 2%-58%) σε MSW σε σχέση με MSM ($p=0.035$)



Kaplan-Meier curves / Log-rank

Ανάλυση χρόνου μέχρι ιολογική ανταπόκριση (διαφορές ανά φύλο)



Log-rank test for equality of survivor functions

sex	Events observed	Events expected
M	168	173.08
F	25	19.92
Total	193	193.00

chi2(1) = 1.45
Pr>chi2 = 0.2281

Cox model (μονο/πολυ-παράγοντικό)

Covariate	Haz. Ratio	Std. Err.	P> z	95% Conf. Interval

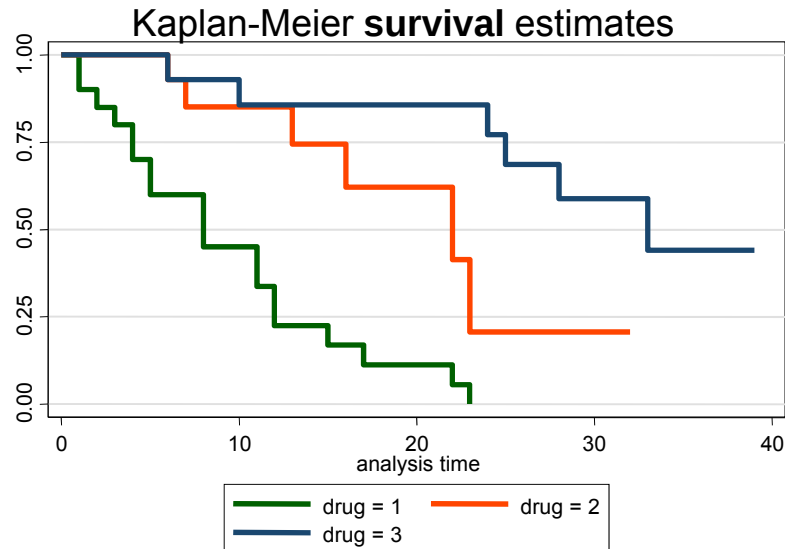
Sex				
M*	1			
F	1.293	0.277	0.231	(0.849 to 1.969)

Covariate	Haz. Ratio	Std. Err.	P> z	95% Conf. Interval

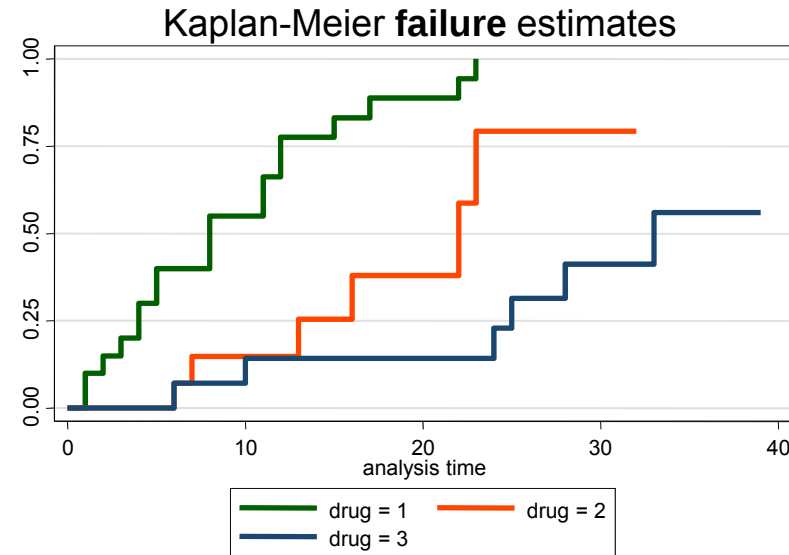
Sex				
M*	1			
F	1.565	0.470	0.135	(0.870 to 2.818)
Risk grp				
Homosexual	1.034	0.240	0.886	(0.656 to 1.630)
Heterosexual*	1			
Drug user	0.109	0.111	0.030	(0.015 to 0.806)
Haemophilia	1.634	0.508	0.114	(0.888 to 3.007)
Other	0.372	0.286	0.199	(0.083 to 1.680)
NAIVE prior to HAART				
Yes/No	2.991	0.505	<0.001	(2.149 to 4.163)
Age at HAART				
per 10 yrs	1.156	0.081	0.038	(1.008 to 1.327)
Baseline CD4				
per 100 cells/ μ L	1.085	0.039	0.023	(1.011 to 1.164)
Baseline VL				
per 1 log10	0.679	0.077	0.001	(0.544 to 0.848)

Ερώτηση 6

Γράφημα 1



Γράφημα 2



- Α) Τα γραφήματα 1 και 2 φαίνονται να είναι συμβατά μεταξύ τους και συμβατά με αποτελέσματα Cox μοντέλου που εκτιμά $HR(2 \text{ vs. } 1) = 0.23$ και $HR(3 \text{ vs. } 1) = 0.07$
- Β) Τα γραφήματα 1 και 2 φαίνονται να είναι συμβατά μεταξύ τους και συμβατά με αποτελέσματα Cox μοντέλου που εκτιμά $HR(2 \text{ vs. } 1) = 1.4$ και $HR(3 \text{ vs. } 1) = 2.3$
- Γ) Τα γραφήματα 1 και 2 δεν φαίνονται να είναι συμβατά μεταξύ τους