

Στοιχεία στατιστικής

Δημ. Θ. Καραμήτσος

Κατηγορίες στατιστικής

- Περιγραφική
- Συγκριτική

Περιγραφική

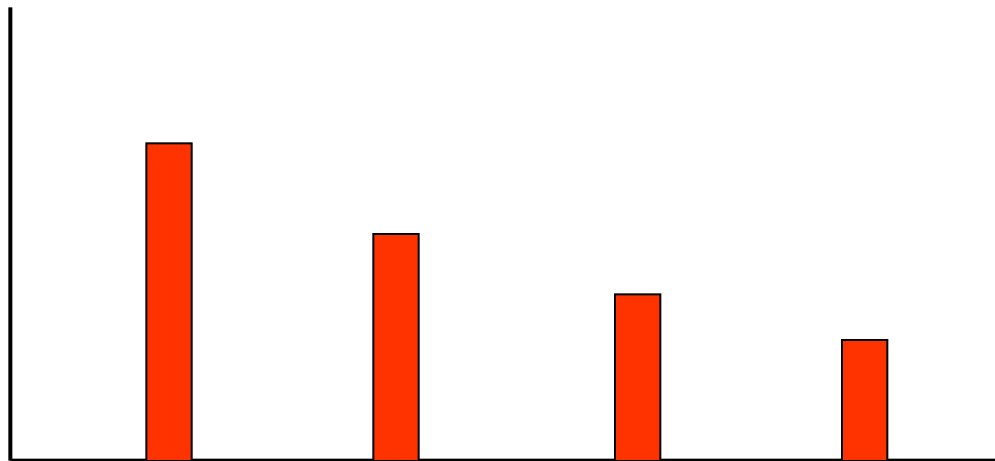
Απεικονίσεις

- Ιστογράμματα
- Καμπύλες
- Φέτες

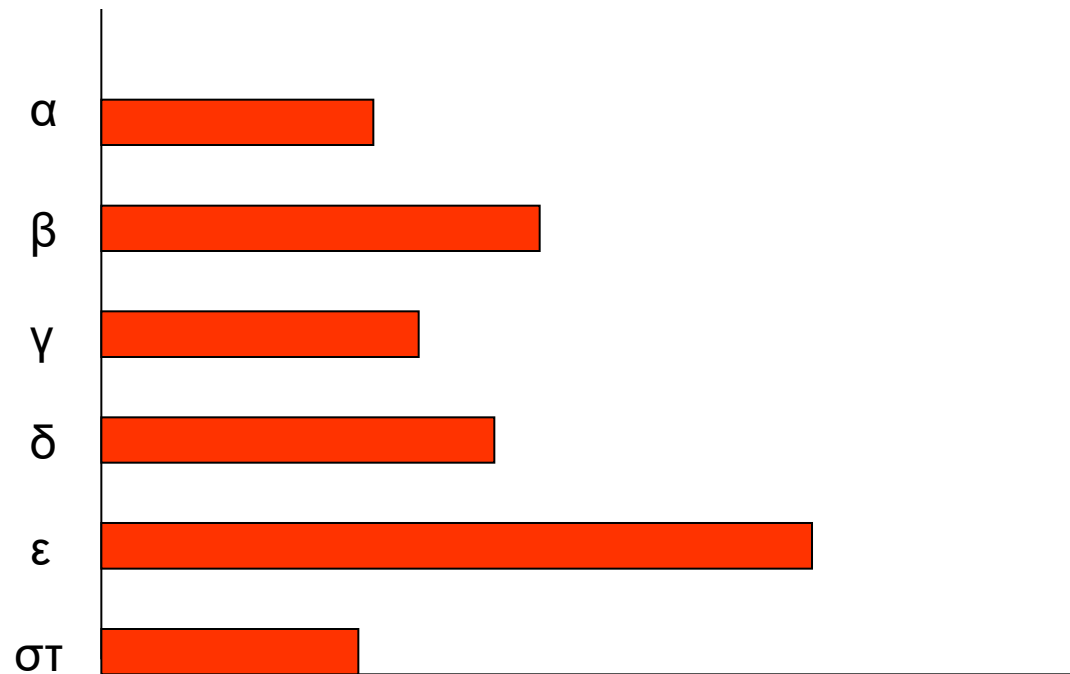
Αναγραφές

- Μέσες τιμές
- Διάμεσες τιμές
- Σταθερή απόκλιση
- Σταθερό σφάλμα

Ιστογράμματα



Ιστογράμματα



Πλεονεκτήματα και Μειονεκτήματα της Θεραπείας με Αντλία Ινσουλίνης*

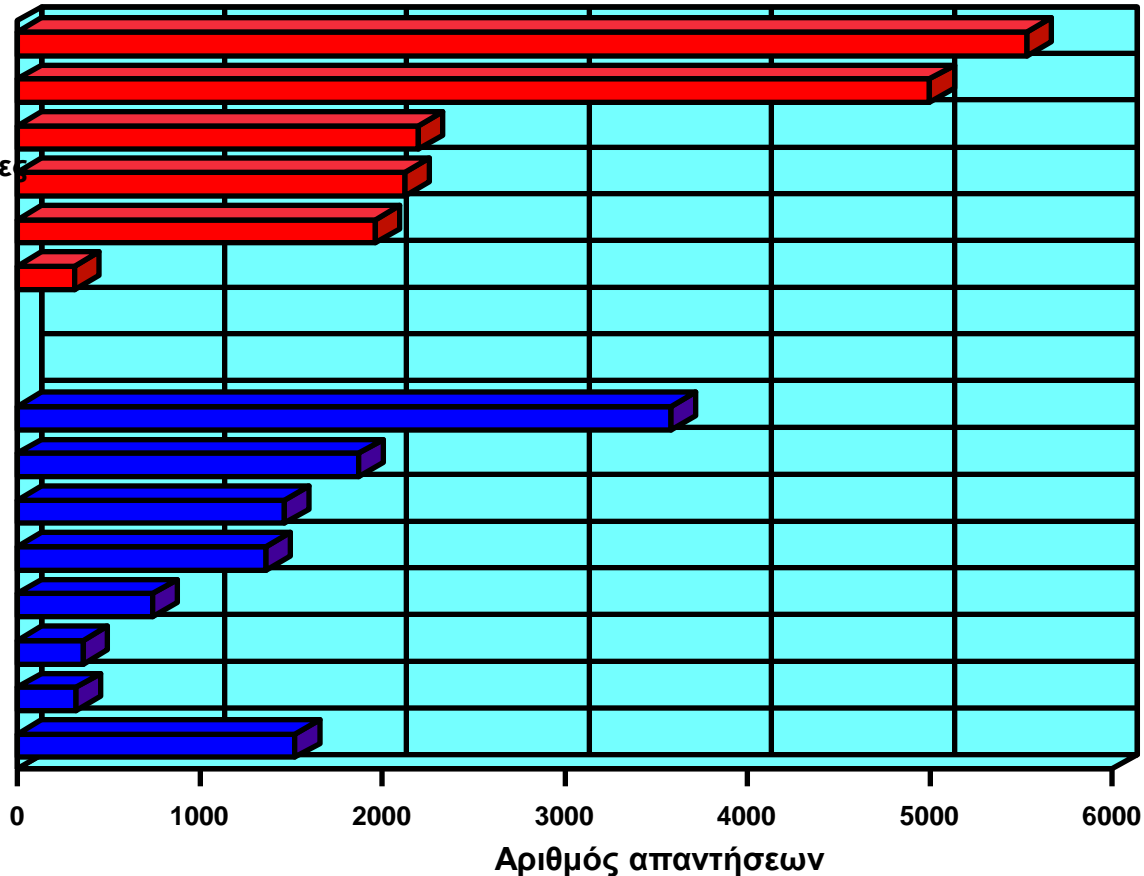
6.890 απαντήσεις, το πολύ 3 υπέρ και 3 κατά ανά ασθενή.

Πλεονεκτήματα

- Καλύτερος έλεγχος μεταβολισμού
- Περισσότερη ευελιξία
- Καλύτερη ποιότητα ζωής
- Δεν χρειάζονται πλέον ενέσεις/σύριγγες
- Προαιρετική παροχή ινσουλίνης
- Άλλα

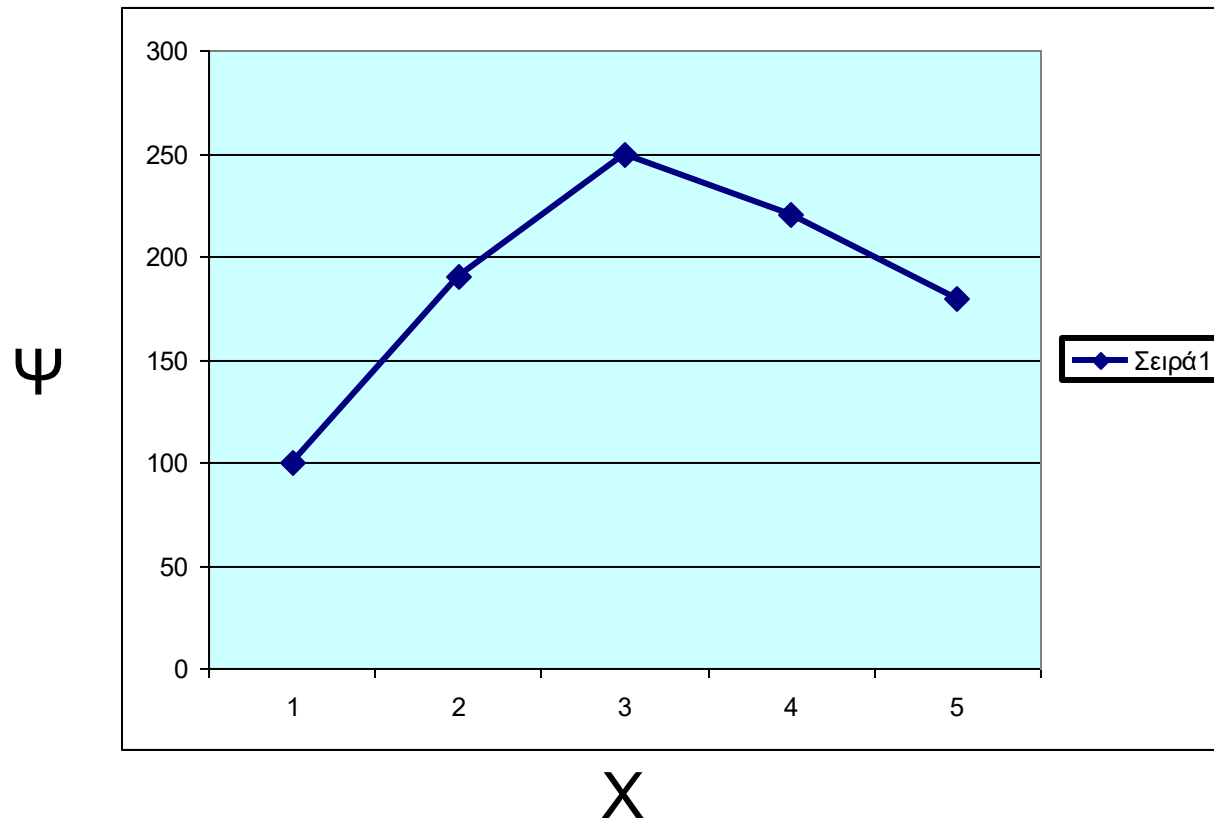
Μειονεκτήματα

- Η αντλία ως ξένο σώμα
- Προβλήματα με εξαρτήματα /θεραπεία
- Ενοχλήσεις (σπορ, προσωπικές στιγμές.)
- Η αντλία απαιτεί περισσότερη εκπαίδευση
- Υψηλό κόστος
- Τακτικότεροι έλεγχοι σακχάρου αίματος
- Μεγαλύτερος κίνδυνος κετοξέωσης
- άλλα

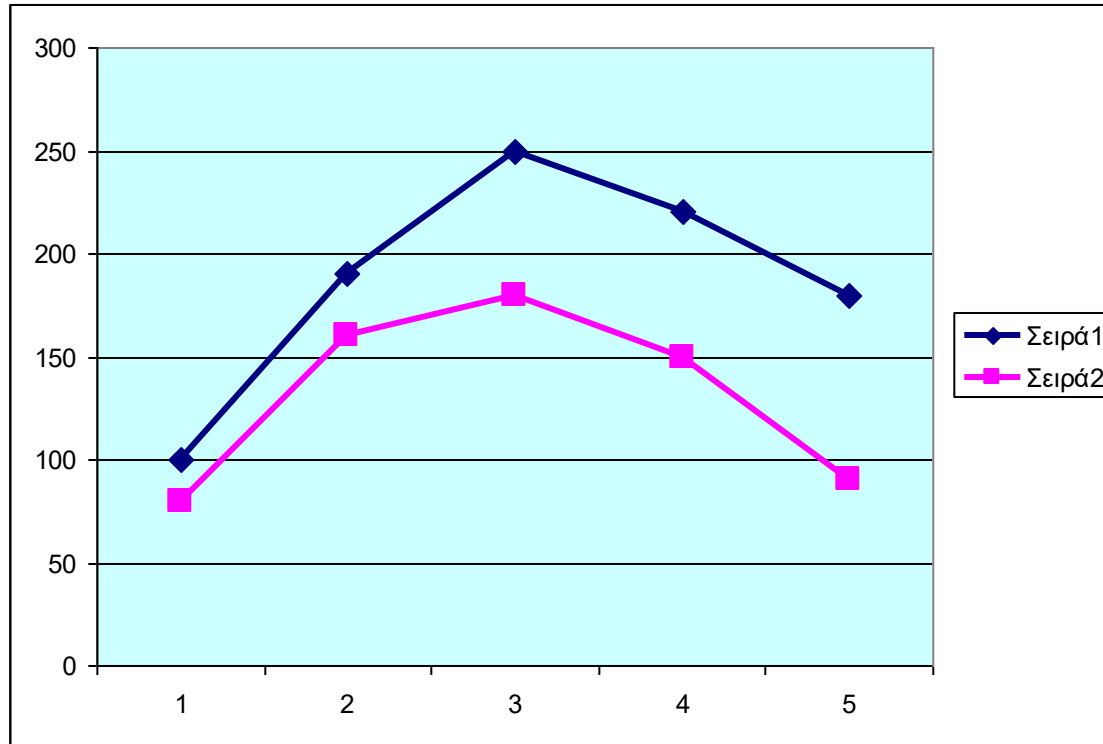


* Αποτελέσματα διεθνούς έρευνας με ασθενείς που χρησιμοποιούσαν αντλία ινσουλίνης 1998; EASD.

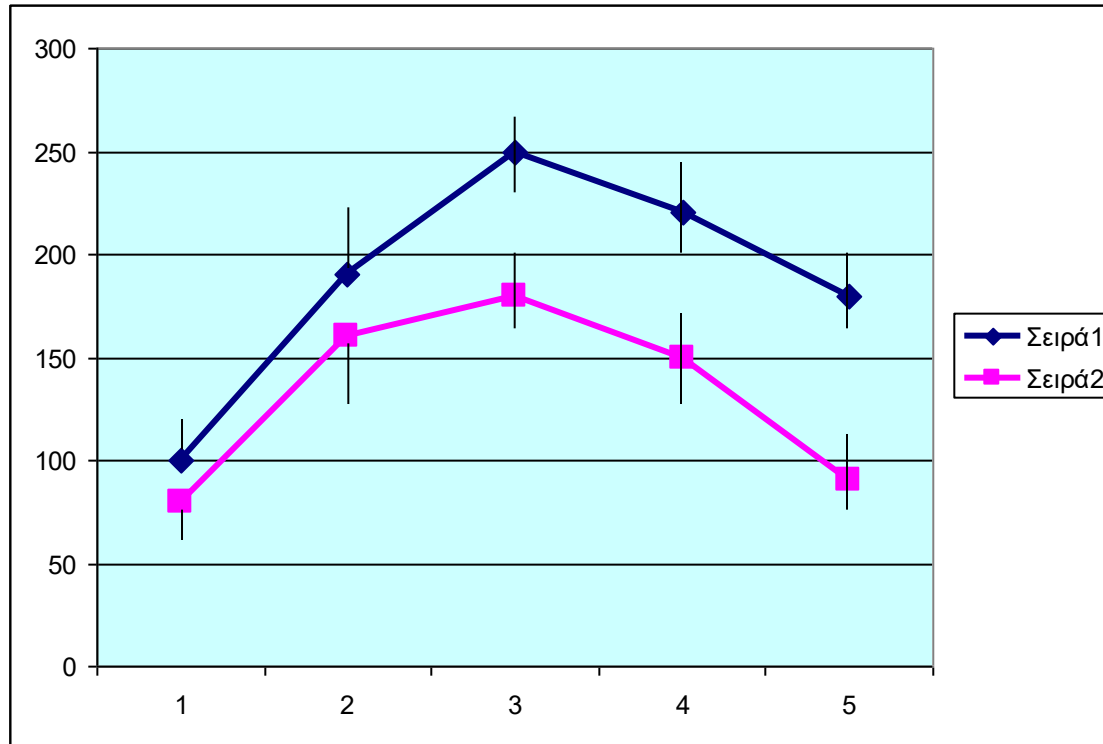
Καμπύλη



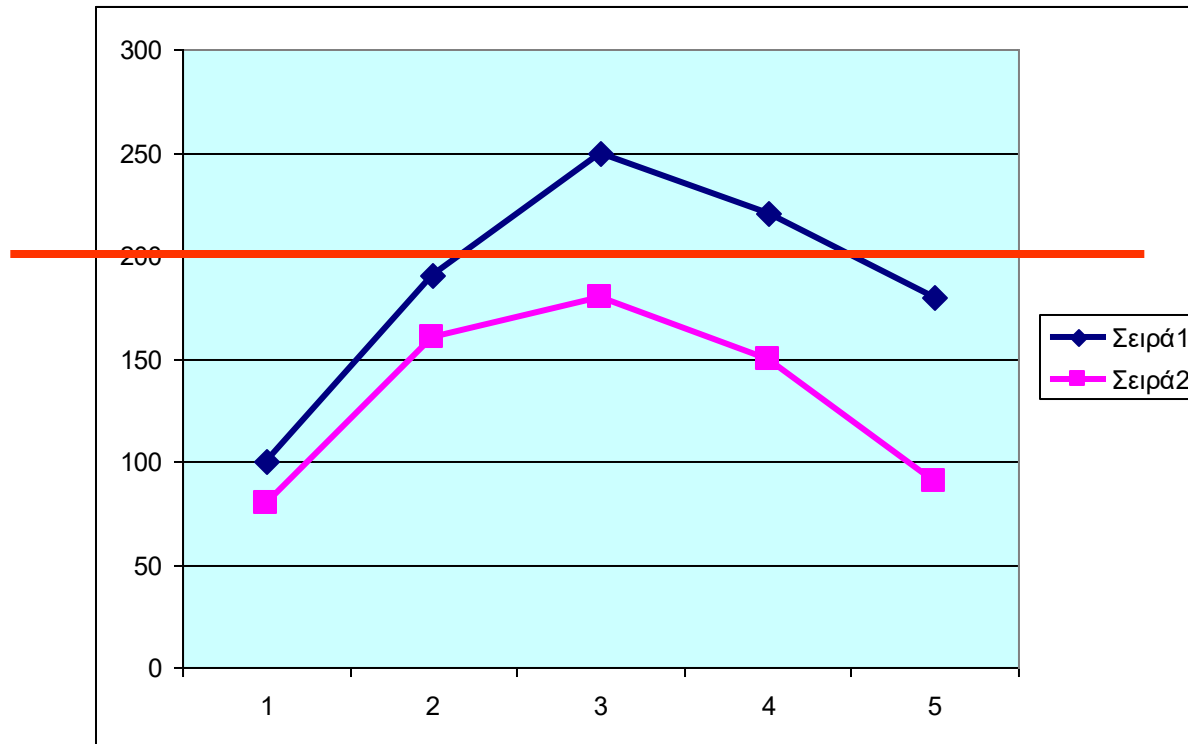
Καμπύλες περιγραφικά



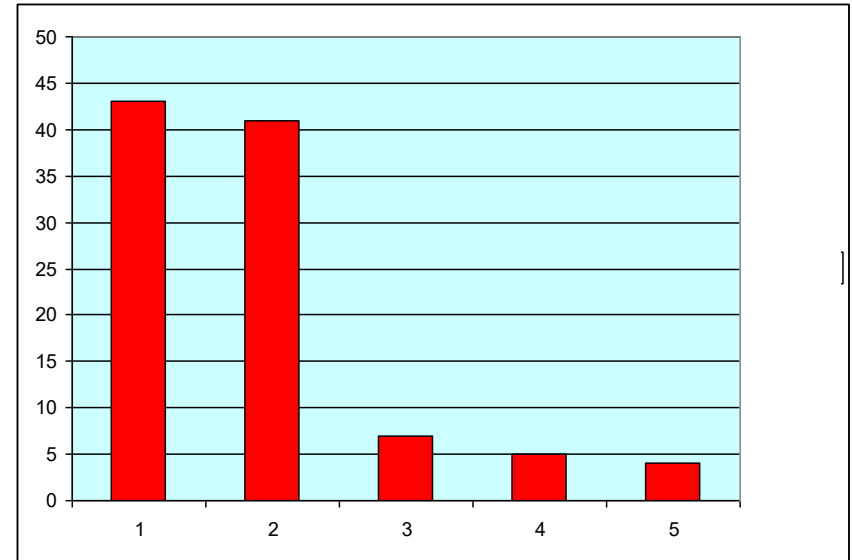
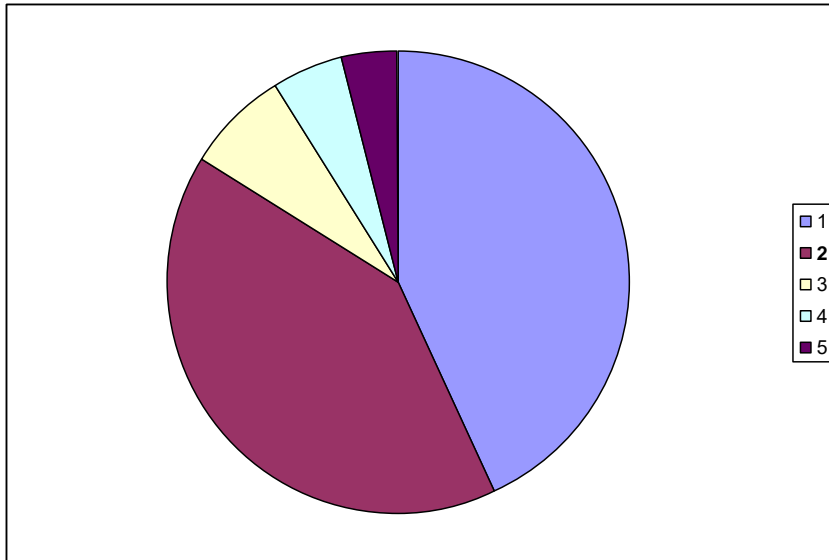
Καμπύλες περιγραφής και σύγκρισης



Καμπύλες περιγραφικά με ανώτατο όριο

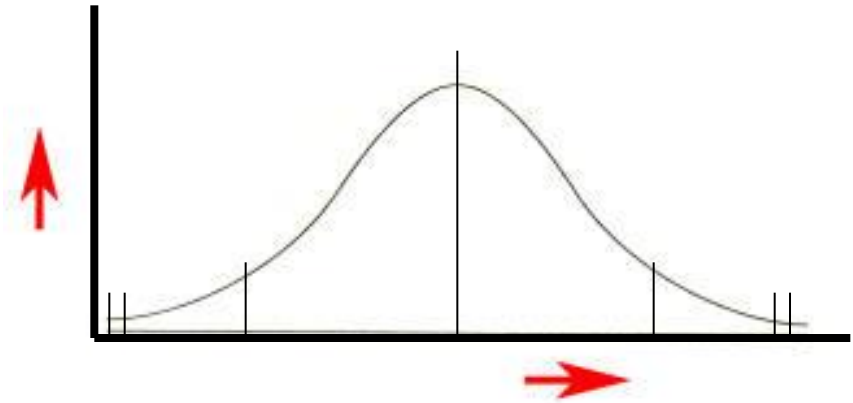


Κατανομές % Ποιο προτιμάτε;



Κατανομή συχνοτήτων

- Ομαλή (καμπύλη Gause)
- Αριθμητικός μέσος (mean)
- Διάμεση τιμή ίδια (median)
- Σταθερή απόκλιση (SD)



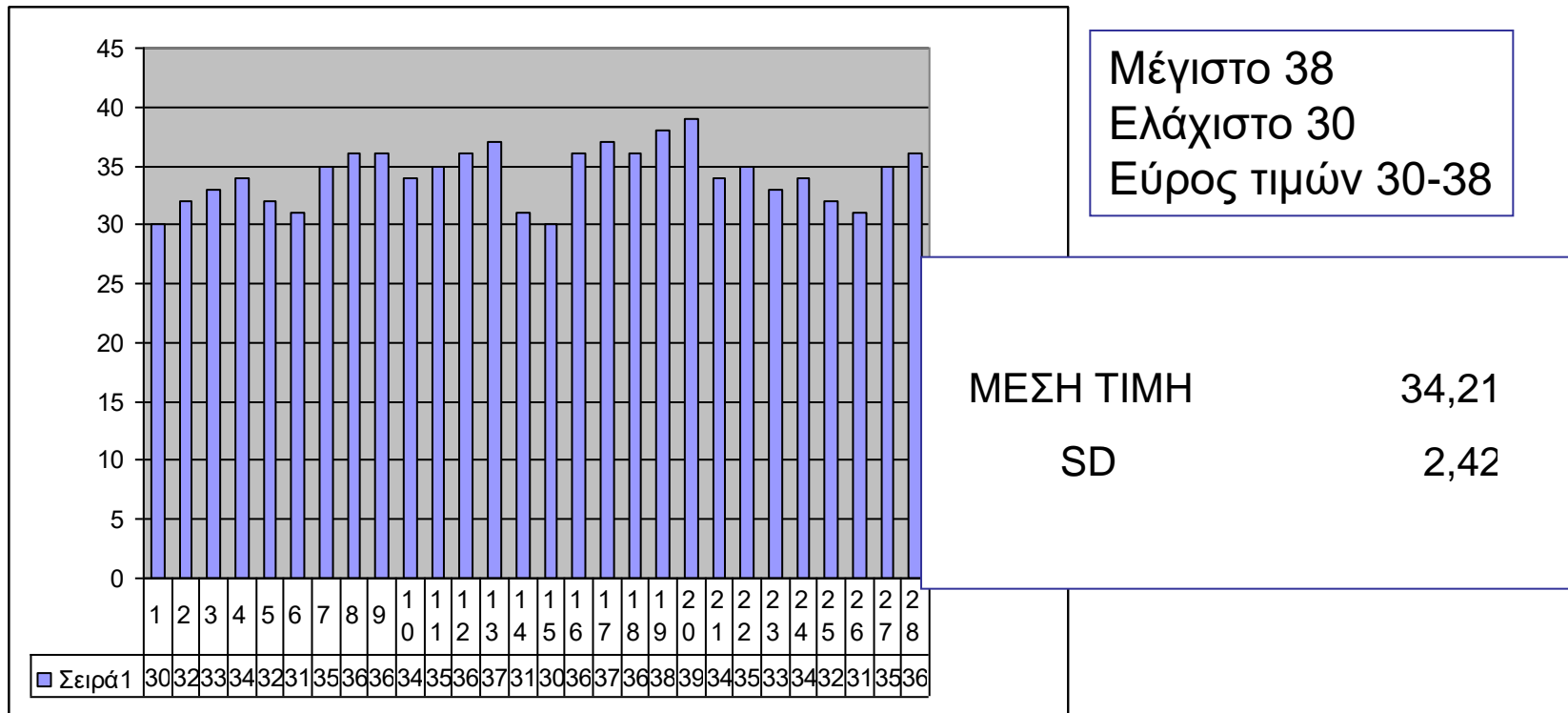
Κωδωνοειδής καμπύλη

Π.χ. $52 \pm 16 = 1SD$ = περιλαμβάνει το 67% των τιμών

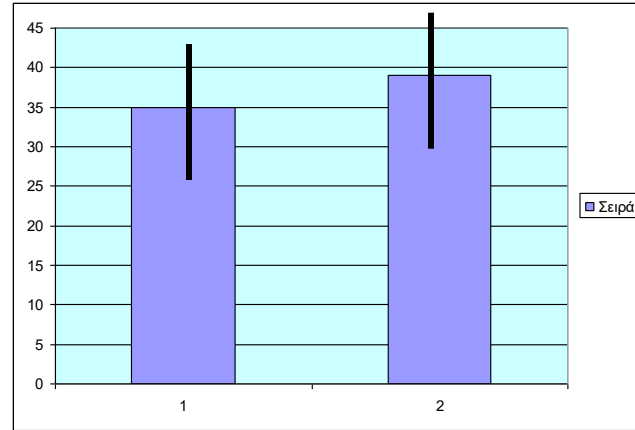
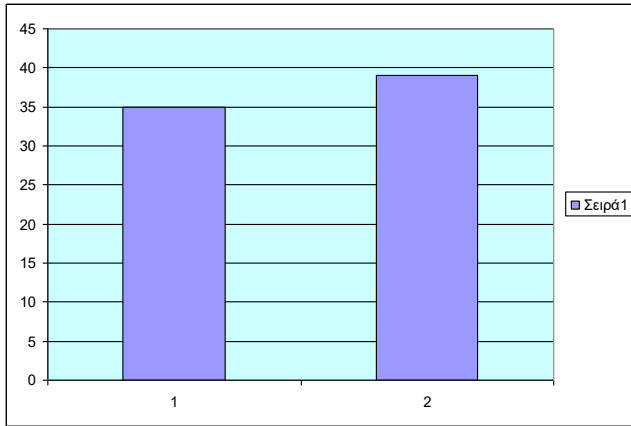
$2SD$ = περιλαμβάνει το 95% των τιμών

$3SD$ = περιλαμβάνει το 99% των τιμών

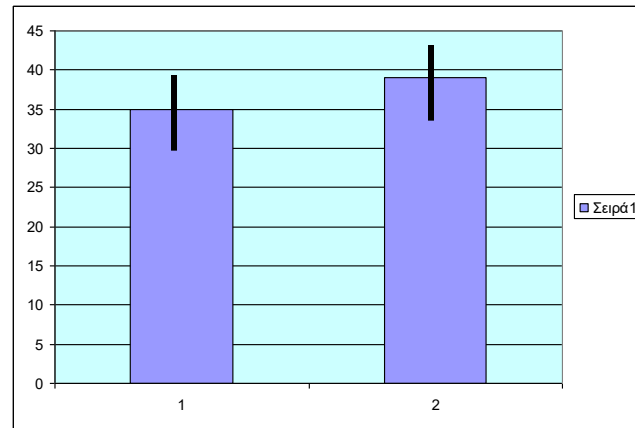
Μέση τιμή +/- SD



Σύγκριση δύο ομάδων τιμών



Στήλες με SD



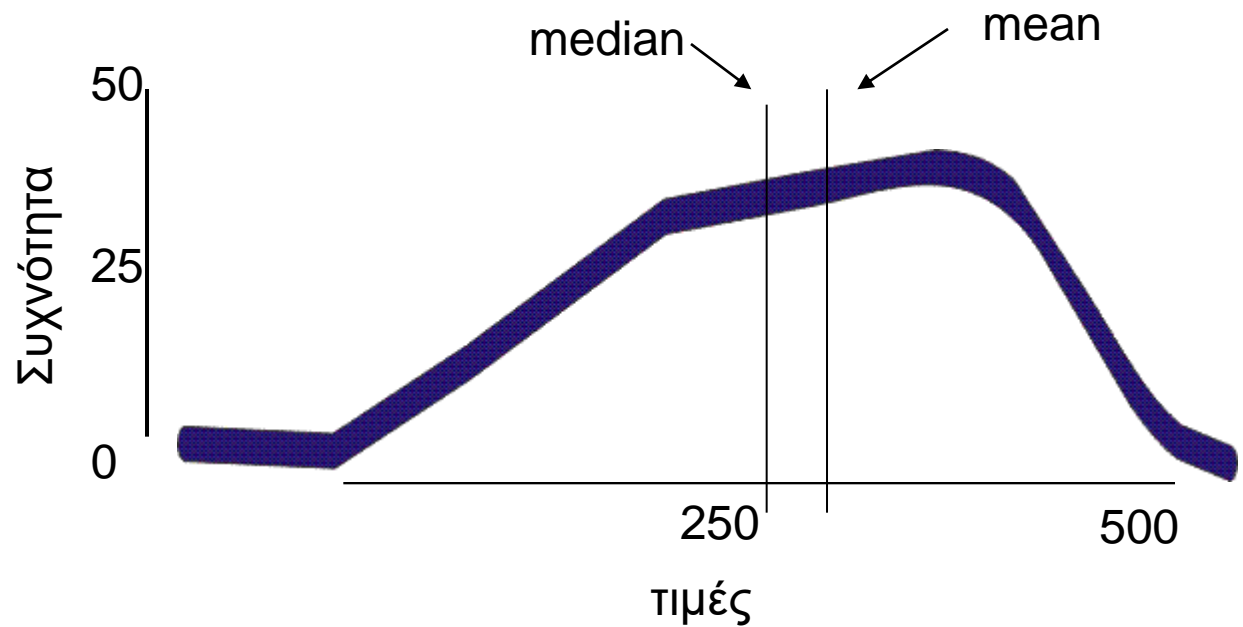
Στήλες με SE

Μη κανονική κατανομή

Αριθμητικός μέσος (**mean**)

Διάμεση τιμή (**median**)

Σταθερή απόκλιση (**SD**) **δεν εκφράζει ακρίβεια**
όπως με κανονική κατανομή



Σταθερό σφάλμα = Standar Error (SE)

$$SE = \frac{SD}{\sqrt{n}}$$

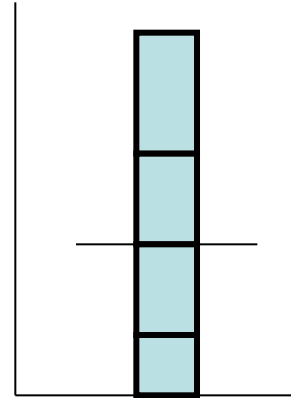
Αδρός στατιστικός έλεγχος

- Διαφορά μέσων τιμών ($\bar{X}_1 - \bar{X}_2$)
 - Άθροιση των δύο SE
- Αν η διαφορά των μέσων τιμών είναι μεγαλύτερη από τα δύο SE τότε είναι σημαντική

$$(\bar{X}_1 - \bar{X}_2) > SE_1 + SE_2$$

Κατανομή σε αναλογίες

- Πλεονέκτημα ότι δείχνει την κατανομή άσχετα με το σχήμα της κατανομής συχνοτήτων
- Δεν βασίζεται στο κωδωνοειδές σχήμα της ομαλής κατανομής



Παλινδρόμηση από τη μέση τιμή

- Είναι η διακύμανση τιμών μετρήσεων η οποία οφείλεται σε μικρά σφάλματα μέτρησης
- Προσοχή στον χαρακτηρισμό «διαφοράς»
γιατί μπορεί οι μετρήσεις να θεωρηθεί ότι διαφέρουν ενώ είναι στο όριο του λάθους μέτρησης

Coefficient variation

(Συντελεστής μεταβλητότητας)

$$CV = \frac{SD}{n} \times 100$$

Π.χ. μετρούμε τη χοληστερίνη 20 φορές στο ίδιο δείγμα αίματος

$$SD = +/-6 \quad N=20$$

$$CV = 6/20 \times 100 = 30 \%$$

$$CV = 0,6/20 \times 100 = 3 \%$$

Μεταβλητότητα μετρήσεων

- Μέθοδος μέτρησης ουρίας
 - Intra assay 1,2%
 - Between assay 1,5%

- Μέθοδος μέτρησης ινσουλίνης
 - Intra assay 3,2%
 - Between assay 4,5%

Επιδημιολογικές έννοιες

Επίπτωση=Incidence

Αριθμός νέων περιπτώσεων ανά έτος/100.000 άτομα.

Π.χ. 10 νέες περιπτώσεις
ΣΔ τύπου 1
ετησίως/100.000 πληθυσμού

Επιπολασμός=Prevalence

Εκατοστιαία αναλογία (%) αρρώστων από ένα νόσημα στον γενικό πληθυσμό.

Π.χ. 0,22 % ΣΔ τύπου 1 στον πληθυσμό σε δεδομένη στιγμή

Επιδημιολογικές έννοιες

Θνησιμότητα

Οι θάνατοι που συνέβησαν -ως ποσοστό επί συνολικού ζώντος πληθυσμού- από όλα τα αίτια θανάτου.

Π.χ. **Θνησιμότητα** των ατόμων που είναι >70 ετών.

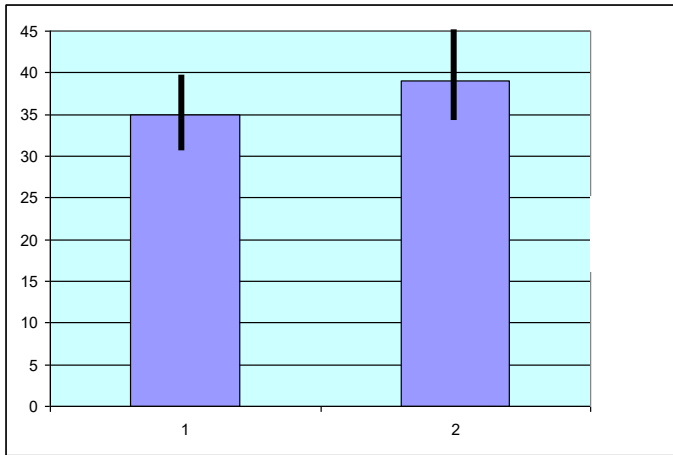
Θνητότητα

Οι θάνατοι που συνέβησαν –ως ποσοστό επί συνολικού ζώντος πληθυσμού- από μία νόσο ή ιατρική παρέμβαση ή επιπλοκή ή παρενέργεια.

Π.χ. Νοσοκομειακή **Θνητότητα** των ατόμων που νοσηλεύθηκαν για έμφραγμα μυοκαρδίου.

Συγκριτική στατιστική

Διαστήματα εμπιστοσύνης (Confidence intervals)



Αν το διάστημα εμπιστοσύνης είναι ευρύ, πολλά πιθανά αποτελέσματα πρέπει να εξεταστούν

Μερικές φορές το διάστημα εμπιστοσύνης περιλαμβάνει μηδενικές διαφορές π.χ. μεταξύ δύο θεραπειών

Ευρύ διάστημα εμπιστοσύνης δείχνει αβεβαιότητα ως προς την αληθινή τιμή ενός αποτελέσματος

Μηδενική υπόθεση και στατιστική διαφορά

Σύμφωνα με τη μηδενική υπόθεση σε δύο ομάδες παρατηρήσεων δεν υπάρχουν διαφορές

Αν υπάρχουν σημαντικές διαφορές δεν ισχύει η μηδενική υπόθεση

$P > 0,05 = 5\%$ πιθανότητες να μην είναι σωστή η μηδενική υπόθεση

$P < 0,01 = 1\%$ >> >> >>

$P < 0,001 = 1/‰$ >> >> >>

Αξιώματα

- Το μέγεθος της τιμής του p δεν σημαίνει διαφορετική σημασία στο αποτέλεσμα
- Διαφορές που δεν είναι στατιστικά σημαντικές δεν είναι αναγκαστικά ασήμαντες
- Τα αποτελέσματα μπορεί να είναι στατιστικά σημαντικά αλλά κλινικά ασήμαντα
 - Π.χ. 2% έναντι 5% αποτυχία θεραπευτική όταν χρησιμοποιούνται δύο φάρμακα έναντι ενός.
Δηλαδή αξίζει τον κόπο να υποβάλλονται όλοι σε διπλή θεραπευτική αγωγή για τόσο μικρή διαφορά (cost and effect)
- Με τη στατιστική εμφανίζονται και μεγάλα ψεύδη

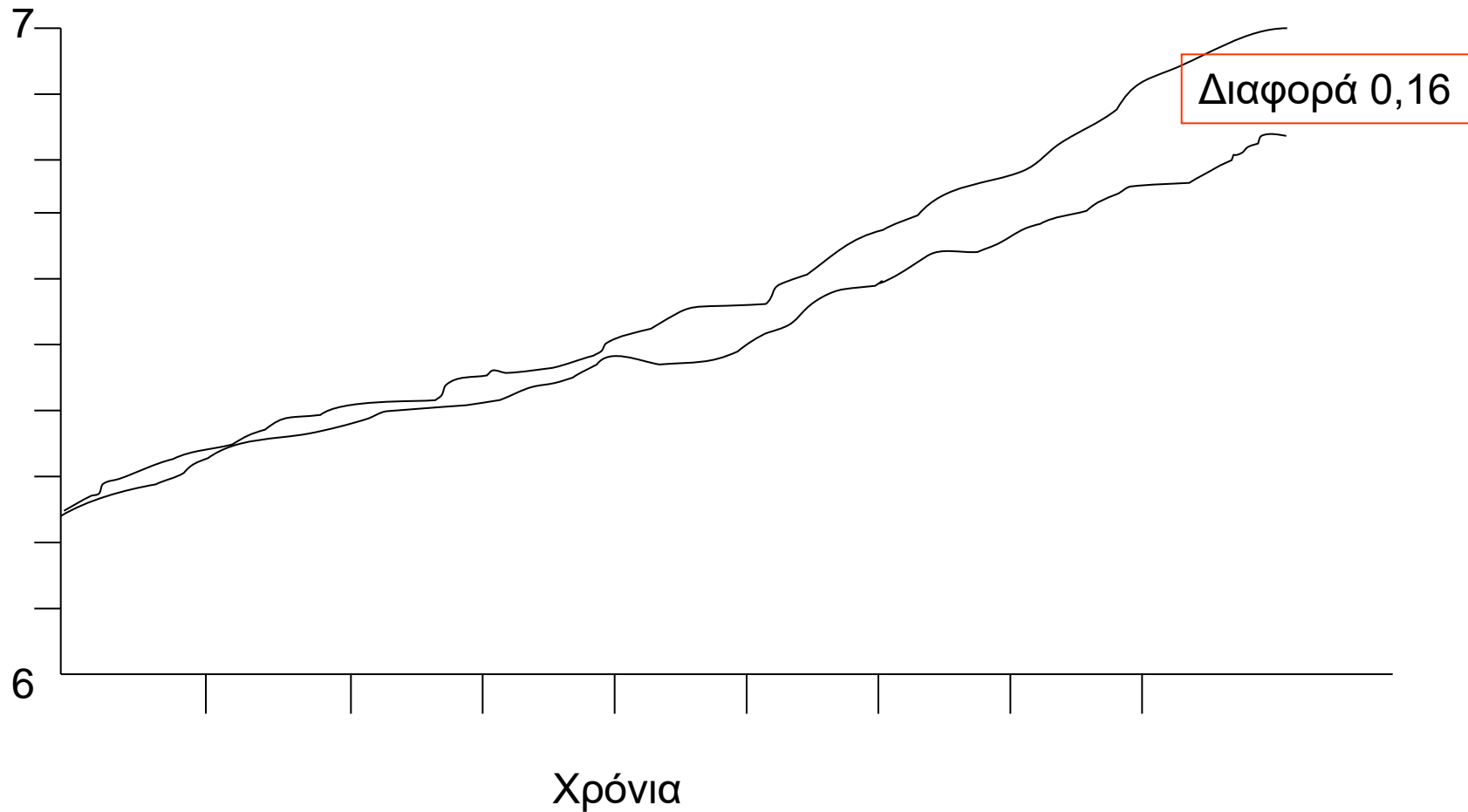
Αξιώματα

- Η στατιστική δύναμη είναι μεγαλύτερη όταν ο αριθμός των παρατηρήσεων είναι μεγάλος
- Η στατιστική δύναμη είναι μεγαλύτερη όταν οι συγκρινόμενες διαφορές είναι μεγάλες

Στατιστικά σφάλματα στη διερεύνηση δραστηκότητας θεραπειας

- Απορρίπτουμε τη μηδενική υπόθεση ενώ είναι αληθής (τύπου 1 σφάλμα)
(δηλαδή βρίσκουμε διαφορές ενώ δεν υπάρχουν)
- Συμφωνούμε με τη μηδενική υπόθεση (τύπου 2 σφάλμα)
(δηλαδή δεν βρίσκουμε στατιστικές διαφορές ενώ υπάρχουν διαφορές)

Παράδειγμα ψεύδους ή ψευδούς εντύπωσης



Σχετικός κίνδυνος

relative risk, risk ratio, odds ratio

	Νόσος	Έλλειψη νόσου	
Παράγων (+)	A	B	A + B
Παράγων (-)	C	D	C + D

Σχετικός κίνδυνος = $\frac{\text{Ποσοστό νόσου σε άτομα με παράγοντα (+)}}{\text{Ποσοστό νόσου σε άτομα με παράγοντα (-)}}$

$$\mathbf{RR} = \frac{A / (A+B)}{C / (C+D)}$$

Σχετικός κίνδυνος

relative risk, risk ratio, odds ratio

A= Υπέρταση(+) και ΑΕΕ (+) =120

C= Χωρίς υπέρταση(-) και ΑΕΕ (-) = 5

B= Υπέρταση (+) και ΑΕΕ (-) = 55

D= Όχι ΑΕΕ (-) χωρίς υπέρταση(-) =400

$$\text{RR} = \frac{A / (A+B)}{C / (C+D)} = \frac{120 / (120+55)}{5 / (5+400)} = 5,56$$

$$\text{Odds ratio} = \frac{A/B}{C/D} = \frac{A \times D}{B \times C} =$$

Η στατιστική και οι βιοχημικές ή αιματολογικές φυσιολογικές τιμές

- Κριτική στην τακτική στατιστικού ορισμού φυσιολογικών τιμών
- Για τον ορισμό φυσιολογικής τιμής πρέπει να υπάρχει βιολογική σημασία

Είδη δεδομένων και στατιστικά τεστ



Δεδομένα Κατηγοριών
(Categorical data)

Συνεχείς μεταβλητές
(Continuous data)

Δεδομένα Κατηγοριών (Categorical data)

- Ονομαστικά δεδομένα (Nominal)
 - Φύλο,
 - ομάδες αίματος,
 - Ναι-όχι,
 - επιτυχία -αποτυχία
- Τακτικά δεδομένα (Ordinal)
 - Κοινωνικές τάξεις,
 - στρατιωτικοί βαθμοί,
 - Ιατρική κατάσταση
 - (σταθερή, ασταθής, βαριά, κριτική)

Συνεχείς μεταβλητές (Continuous data)

- Θερμοκρασία
- Βιολογικά χαρακτηριστικά
 - ύψος,
 - Βάρος
 - Περίμετρος μέσης
- Αιματολογικές τιμές
- Βιοχημικές τιμές

Τεστ σε δεδομένα κατηγοριών

χ^2 ΤΕΣΤ

και

Fisher's exact test

Συγκρίνουν αναλογίες σε ομάδες παρατήρησης

Εφαρμόζεται σε **μεγάλους** αριθμούς

Εφαρμόζεται σε **μικρούς** αριθμούς

Σύγκριση αναλογιών % = χ^2 test

Εμβολιάστηκε ένας πληθυσμός και μετά από 5 χρόνια καταμετρήθηκαν οι νοσήσεις. Ήταν αποτελεσματικό το εμβόλιο;

	Νοσήσαντες	Μη νοσήσαντες	Σύνολα
Εμβολιασθέντες	5	95	100
Μη εμβολιασθέντες	10	85	95
Σύνολα	15	180	195

Υπολογισμός χ^2 test

	Νοσήσαντες	Μη Νοσήσαντες	Σύνολα
Εμβολιο (-)	a	c	e
Εμβολιο (+)	b	d	f
Σύνολα	g	h	K

$$\chi^2 = \frac{(b \times c - a \times d)^2 \times K}{e \times f \times g \times h}$$

Υπολογισμός χ^2 test

	Νοσήσαντες	Μη Νοσήσαντες	Σύνολα
Εμβολιο (-)	a=10	c=117	e=127
Εμβολιο (+)	b=3	d=144	f=147
Σύνολα	g=13	h= 261	K=274

$$\chi^2 = \frac{(b \times c - a \times d)^2 \times K}{e \times f \times g \times h} = 3,9$$

Για $\chi^2 = 3,9$ $p < 0,05$

$P < 0,05 = \chi^2 3,84-6,64$

$P < 0,01 = \chi^2 6,64-10,83$

$P < 0,001 = \chi^2 > 10,83$

ΣΤΑΤΙΣΤΙΚΕΣ ΣΥΝΕΧΩΝ ΜΕΤΑΒΛΗΤΩΝ

Ομαλής κατανομής
Παραμετρικές μέθοδοι



-Student's t – test
paired

-Student's t – test
unpaired

t

u

Μη ομαλής κατανομής
Μη παραμετρικές μέθοδοι



-Wilcoxon test
paired

-Mann-Whitney
unpaired

Συγκρίνονται οι μέσες τιμές από δύο σειρές (ομάδες, γκρουπ) τιμών (μετρήσεων)

Στο αποτέλεσμα της σύγκρισης παίζουν ρόλο
-το μέγεθος των δειγμάτων
-η διαφορά των μέσων τιμών
-η μεταβλητότητα των τιμών κάθε ομάδας

Υπολογισμός στατιστικής σημαντικότητας

- Για το t test → **t**
- Για μη παραμετρικές → **u ή z**

Σε πίνακες

το **t** ή ανάλογα το **u** αντιστοιχούν σε **p**

Προηγουμένως λαμβάνονται υπόψη βαθμοί ελευθερίας

Student's t test

Sample Variance $s^2 = \frac{\sum x^2}{n} - \bar{X}^2$

Η Sample Variance είναι το τετράγωνο της σταθερής απόκλισης s

Συνολική εκτίμηση Variance $\hat{\sigma} = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$

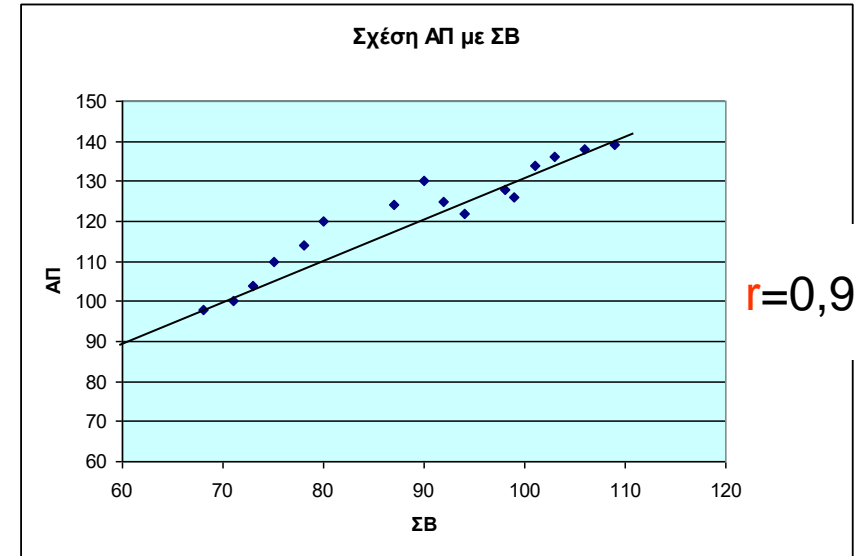
Best estimate of the standard error for difference of the means of two samples of this size

$$\hat{\sigma}_\omega = \hat{\sigma} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Student's t $t = \frac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}_\omega}$

Συσχετίσεις

- Η συσχέτιση δείχνει τη δύναμη μιας γραμμικής σχέσης ή σύνδεσης μεταξύ μεταβλητών
 - Π.χ. έχει σχέση η ΑΠ με το ΣΒ
 - Π.χ αυξάνει ο κίνδυνος για ηπατίτιδα με την αύξηση της συχνότητας των μεταγγίσεων



Εξαρτάται από:

- Αριθμό παρατηρήσεων
- Μεταβλητότητα δεδομένων

Συντελεστής συσχέτισης = Coefficient correlation = r

-Κυμαίνεται από -1 έως +1

-Όσο πλησιέστερος προς το 1 τόσο καλύτερη η συσχέτιση

-Αρνητικό ή θετικό πρόσημο

-Τελικά βγαίνει και συντελεστής p

Τεστ συσχετίσεων

Regression analysis

Παραμετρικές



test Pearson

Μη παραμετρικές



test Spearman

Ορολογίες για Regression analysis

- logistic regression
- Log linear models
- Stepwise multiple regression
- Discriminate function analysis

Τεστ συσχετίσεων

Regression analysis

$$y = a + bx$$

a το θεωρητικό σημείο της τομής (intercept) στον άξονα y όταν το X είναι μηδέν

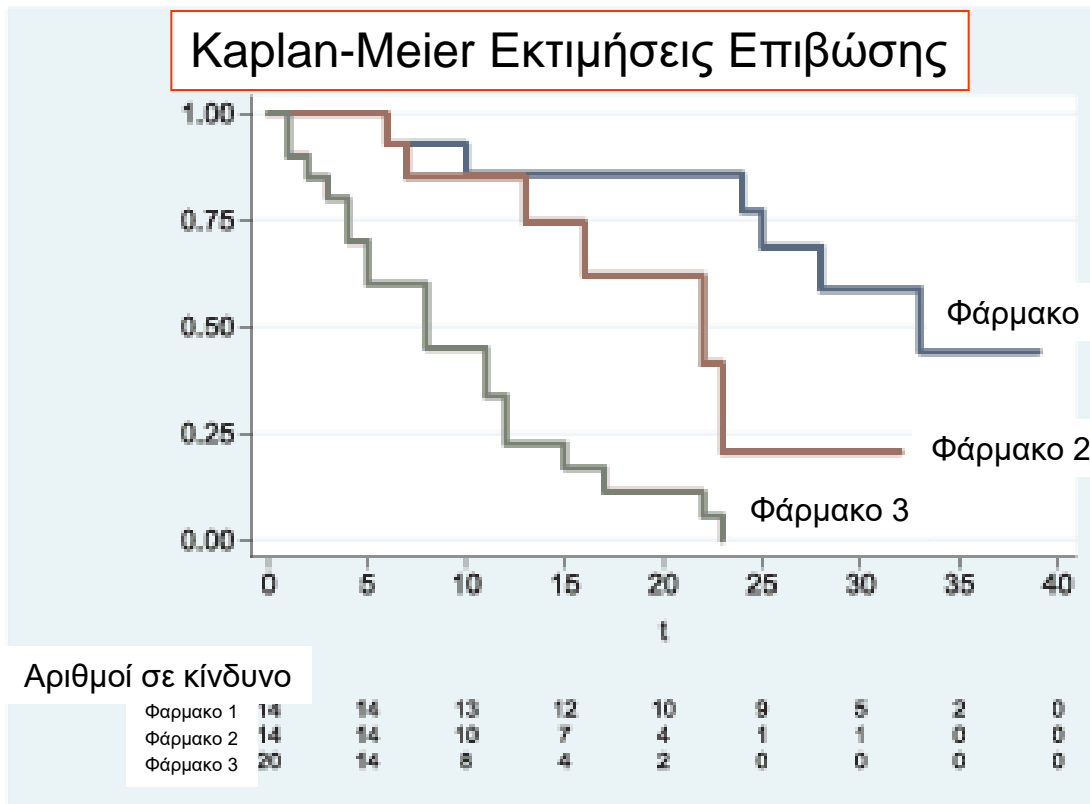
b το ποσό της μεταβολής της y μεταβλητής για κάθε μονάδα μεταβολής της X

r^2 είναι ο συντελεστής συσχέτισης υψωμένος στο τετράγωνο

Π.χ. ένα $r = 0,69$ γίνεται $r^2 = 0,48$

πράγμα που δείχνει ότι 48% της μεταβλητότητας (variance) μιας μεταβλητής οφείλεται στην άλλη μεταβλητή. Το υπόλοιπο της μεταβλητότητας οφείλεται σε άλλους παράγοντες.

Kaplan-Meier Survival estimates



Τέλος μαθήματος

Ευχαριστώ για την προσοχή σας