

Ανάλυση ποσοτικών και ποιοτικών δεδομένων

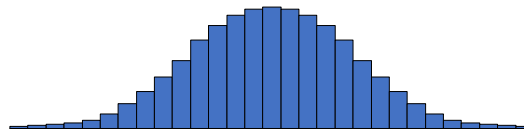
Επαναληπτικές ασκήσεις II

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

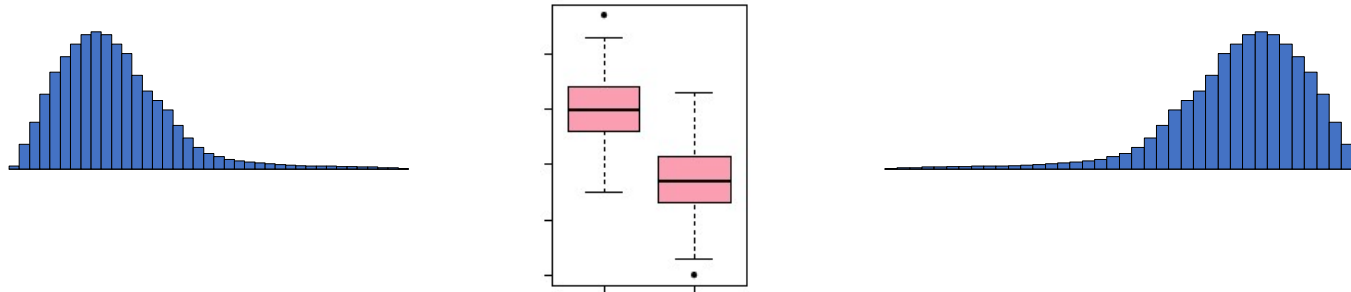
12/11/2025, 15:30–17:00

Κατανομές και μέτρα θέσης και διασποράς

Αν συμμετρικές κατανομές (π.χ. κανονική κατανομή) τότε καταλληλότερα μέτρα είναι η **μέση τιμή** και **σταθερή απόκλιση** και χρησιμοποιείται κυρίως το **ιστόγραμμα**

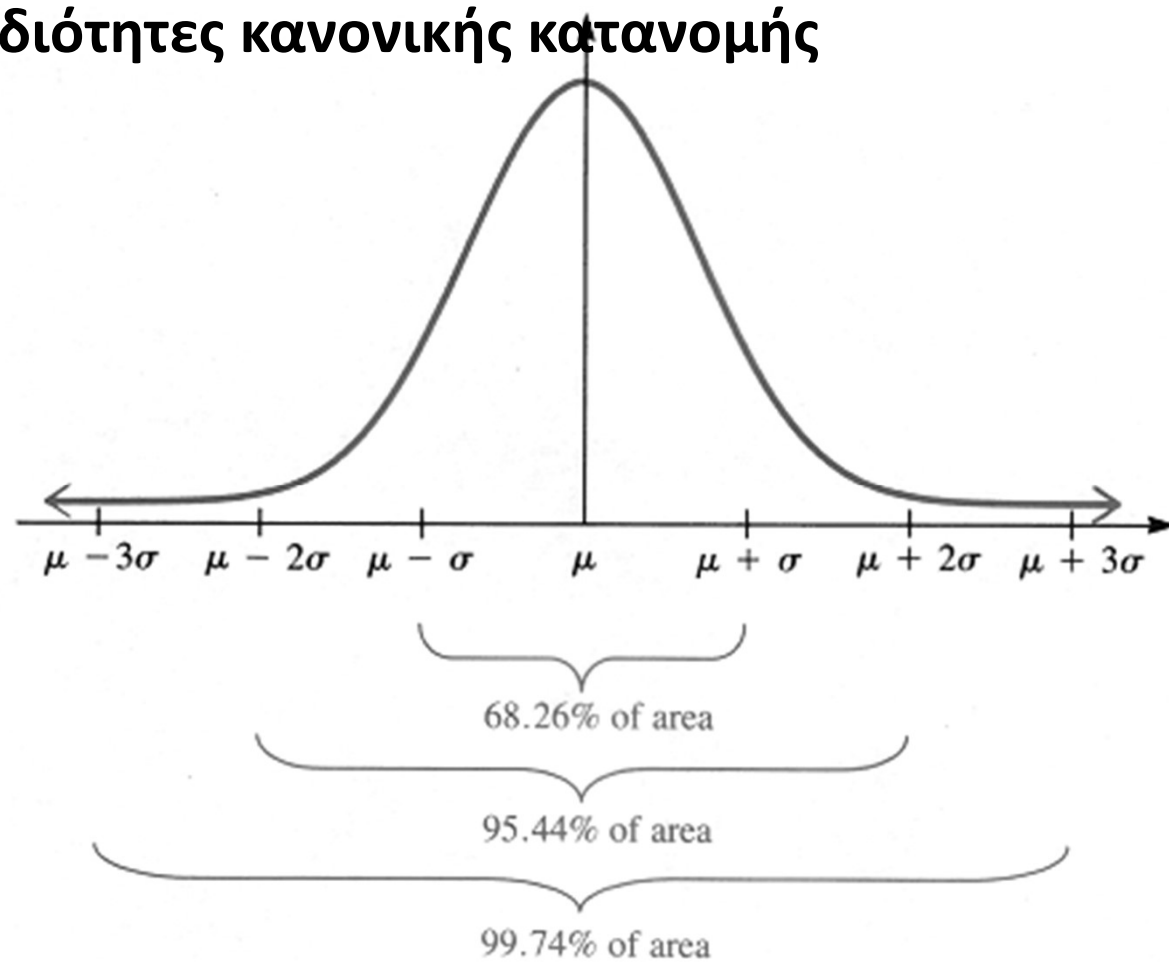


Αν ασύμμετρες κατανομές (π.χ. λοξή θετικά ή αρνητικά) τότε καταλληλότερα μέτρα είναι η **διάμεσος** (που δεν επηρεάζεται από ακραίες τιμές) και **τα εκατοστημόρια (25ο και 75ο)** και χρησιμοποιείται κυρίως το **θηκόγραμμα**



Αν ποιοτικές μεταβλητές τότε χρησιμοποιούνται **ραβδογράμματα** ή κυκλικά διαγράμματα (**πίττες**)

Ιδιότητες κανονικής κατανομής



ΠΙΝΑΚΑΣ ΠΙ

Κανονική κατανομή. Το εκατοστιαίο ποσοστό των παρατηρήσεων των οποίων οι τιμές περιλαμβάνονται μεταξύ των τιμών $\bar{X} + Zs$ και $\bar{X} - Zs$ και το εκατοστιαίο ποσοστό των παρατηρήσεων των οποίων οι τιμές βρίσκονται εκτός των παραπάνω τιμών (δηλαδή είναι μεγαλύτερες της $\bar{X} + Zs$ ή μικρότερες της $\bar{X} - Zs$). Με $\bar{X}$ παριστάνεται η μέση τιμή, με s η σταθερή απόκλιση και με Z το κάθε φορά υποπολλαπλάσιο ή πολλαπλάσιο της σταθερής απόκλισης (standard normal deviate).

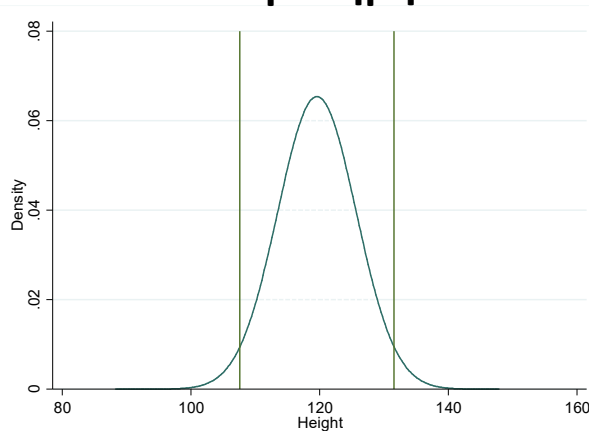
Z	Μεταξύ των $\bar{x} + Zs$ και $\bar{x} - Zs$	Εκτός των $\bar{x} + Zs$ και $\bar{x} - Zs$
0,00	0,00	100,00
0,10	7,97	92,03
0,20	15,85	84,15
0,30	23,58	76,42
0,40	31,08	68,92
0,50	38,29	61,71
0,60	45,15	54,85
0,70	51,61	48,39
0,80	57,63	42,37
0,90	63,19	36,81
1,00	68,27	31,73
1,10	72,87	27,13
1,20	76,99	23,01
1,30	80,64	19,36
1,40	83,85	16,15
1,50	86,64	13,36
1,60	89,04	10,96
1,645	90,00	10,00
1,70	91,09	8,91
1,80	92,81	7,19
1,90	94,26	5,74
1,960	95,00	5,00
2,00	95,44	4,56
2,10	96,43	3,57
2,20	97,22	2,78
2,30	97,85	2,15
2,40	98,36	1,64
2,50	98,76	1,24
2,576	99,00	1,00
2,60	99,07	0,93
2,70	99,31	0,69
2,80	99,49	0,51
2,90	99,63	0,37
3,00	99,73	0,27
3,10	99,81	0,19
3,20	99,86	0,14
3,30	99,90	0,10
3,40	99,93	0,07
3,50	99,95	0,05
3,891	99,99	0,01

95% Διάστημα εμπιστοσύνης και φυσιολογικές τιμές

Συχνά γίνεται σύγχυση μεταξύ των φυσιολογικών τιμών και του διαστήματος εμπιστοσύνης της μέσης τιμής

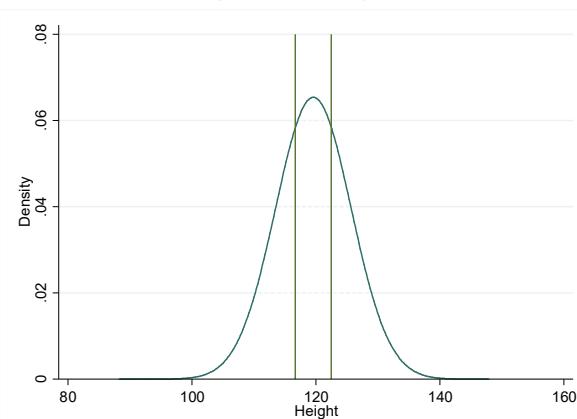
$$\bar{X} \pm 1.96 * SD$$

Εκφράζει το εύρος που βρίσκεται το 95% των παρατηρήσεων



$$\bar{X} \pm 1.96 * SE$$

Εκφράζει το εύρος που βρίσκεται με 95% βεβαιότητα η μέση τιμή του Πληθυσμού



Παράδειγμα: Τιμές επιθηλιακής πρόσφυσης

Σε ένα τυχαίο δείγμα 1098 ατόμων ηλικίας 20-39 ετών από έναν πληθυσμό, η μέση τιμή της απώλειας επιθηλιακής πρόσφυσης είναι 0.82 mm και η σταθερή απόκλιση 0.23 mm.

(α) Πόσα περίπου άτομα θα έχουν τιμές απώλειας επιθηλιακής πρόσφυσης μεταξύ 1.05 και 1.28 mm;

(β) Πόσα άτομα θα έχουν τιμές μεταξύ 0.59 και 1.28 mm;

Παράδειγμα: Τιμές επιθηλιακής πρόσφυσης

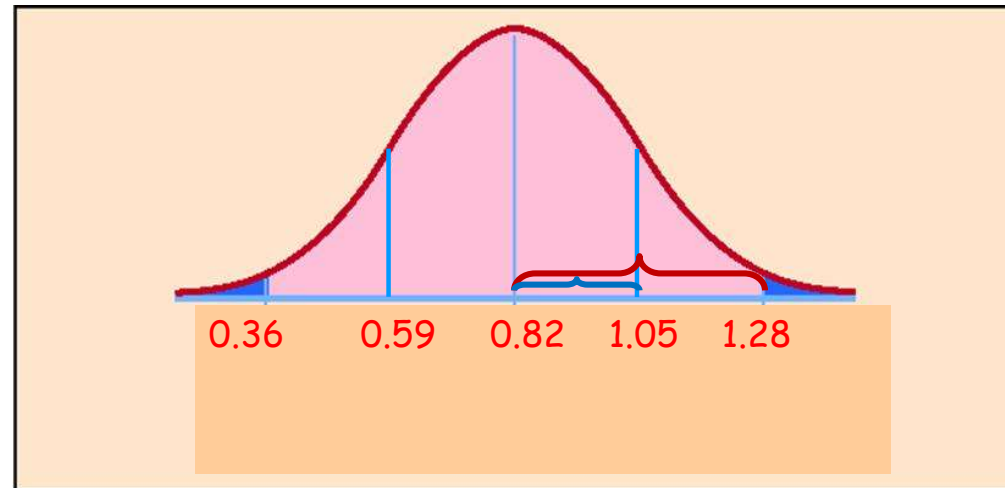
- N=1098 άτομα
- Επιθηλιακή πρόσφυση
 - Μέση τιμή ($\bar{X}$) 0.82 mm
 - Σταθερή απόκλιση (SD) 0.23 mm

Πως μεταφράζεται η σταθερή απόκλιση;

π.χ. Περίπου το 95% των ατόμων έχουν επιθηλιακή πρόσφυση στο εύρος $\bar{X} \pm 2*SD$ (δηλαδή από 0.36 – 1.28 mm)

(υποθέτοντας ότι το μέγεθος ακολουθεί την κανονική κατανομή)

Πόσα άτομα έχουν τιμές απώλειας επιθηλ. πρόσφ. 1.05 έως 1.28



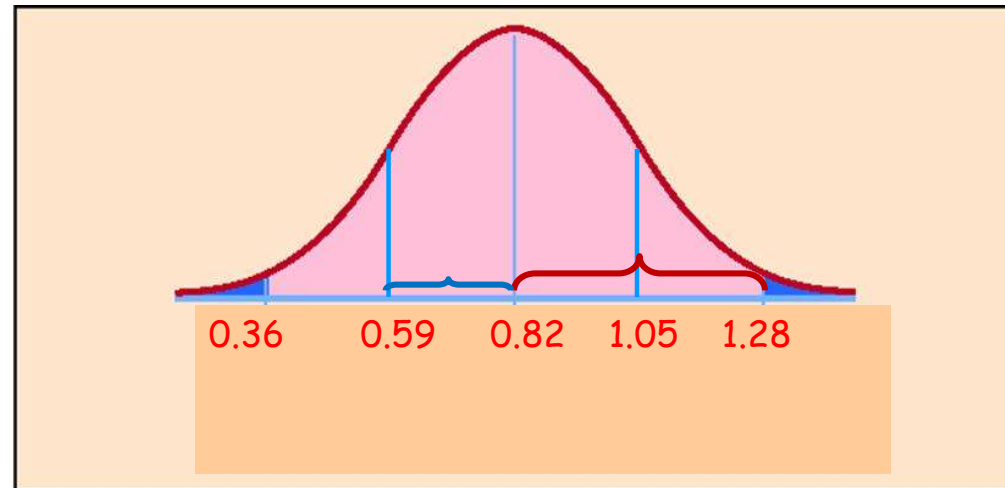
$[\bar{X} + s, \bar{X} + 2s]$

- α) $0.36 - 1.28 \rightarrow$ **95%** των παρατηρήσεων
 $0.82 - 1.28 \rightarrow \frac{95}{2} \% =$ **47.5%** των παρατηρήσεων
 $0.59 - 1.05 \rightarrow$ **68%** των παρατηρήσεων
 $0.82 - 1.05 \rightarrow \frac{68}{2} \% =$ **34%** παρατηρήσεων

Τελικά από 1.05 έως 1.28: **47.5%-34%** = 13.5 % των παρατ.

$$n_1 = 13.5 \% * 1098 \cong 148 \text{ άτομα}$$

Πόσα άτομα έχουν τιμές απώλειας επιθηλ. πρόσφ. 0.59 έως 1.28



$[\bar{X} - s, \bar{X} + 2s]$

- α) $0.36 - 1.28 \rightarrow$ **95%** των παρατηρήσεων
 $0.82 - 1.28 \rightarrow \frac{95}{2} \% =$ **47.5%** των παρατηρήσεων
 $0.59 - 1.05 \rightarrow$ **68%** των παρατηρήσεων
 $0.59 - 0.82 \rightarrow \frac{68}{2} \% =$ **34%** παρατηρήσεων

Τελικά από 0.59 έως 1.28: **34% + 47.5% = 81.5 %** των παρατ.

$$n1 = 81.5 \% * 1098 \cong 895 \text{ άτομα}$$

Παράδειγμα: σχέση στοματικής υγείας & ΔΜΣ

Έστω 2 ομάδες: ομάδα A (n_1 άτομα) και ομάδα B (n_2 άτομα)

Δίνονται οι τιμές του Δείκτη μάζας σώματος (ΔΜΣ) σε κάθε ομάδα:

Ομάδα A: 225 άτομα με πλήρη νωδότητα

Ομάδα B: 110 άτομα με 11 έως 32 δόντια

Έστω μ_1 η μέση τιμή του ΔΜΣ στην ομάδα A και μ_2 στην B

Θέλουμε να ελέγξουμε αν οι 2 μέσες τιμές είναι ίσες ή διαφέρουν

$H_0: \mu_1 = \mu_2$ (Μηδενική υπόθεση)

$H_a: \mu_1 \neq \mu_2$ (Εναλλακτική υπόθεση)

bmi_edent.sav [DataSet1] - IBM SPSS Statistics Data Editor								
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help								
	id	edentulism	bmi	var	var	var	var	var
1	1	νωδότητα	23,41					
2	2	νωδότητα	20,31					
3	3	νωδότητα	18,14					
4	4	νωδότητα	15,06					
5	5	νωδότητα	19,25					
6	6	νωδότητα	21,45					
7	7	νωδότητα	18,46					
8	8	νωδότητα	16,14					
9	9	νωδότητα	20,50					
10	10	νωδότητα	20,39					
11	11	νωδότητα	21,48					
12	12	νωδότητα	17,44					
13	13	νωδότητα	19,01					
14	14	νωδότητα	18,19					
15	15	νωδότητα	18,15					
16	16	νωδότητα	23,40					
17	17	νωδότητα	15,45					
18	18	νωδότητα	19,35					
19	19	νωδότητα	15,12					
20	20	νωδότητα	22,84					
21	21	νωδότητα	18,77					
22	22	νωδότητα	18,34					
Data View Variable View								

- Σαν πρώτο βήμα, θα υπολόγιζα το ΔΜΣ στις δύο ομάδες

$$\bar{X}_1 = \frac{23.41 + .31 + 18.14 + 15.06 + \dots}{225} = 19.6 \text{ kg/m}^2$$

$$\bar{X}_2 = \frac{26.05 + 2 .05 + 2 .59 + 32.98 \dots}{110} = 21.8 \text{ kg/m}^2$$

- Είναι αυτή η διαφορά πραγματική ή είναι αποτέλεσμα τυχαίας διακύμανσης?
 → **t-test:** θα υπολογίσω την πιθανότητα να βρω μία τέτοιας τάξης διαφορά **στην τύχη** δηλαδή όταν ο ΔΜΣ στις δύο ομάδες δεν διαφέρει
- αν η πιθανότητα να βρω μία τέτοιας τάξης διαφορά όταν οι ομάδες είναι παρόμοιες είναι πολύ μικρή (<0.05) τότε η διαφορά είναι στατιστικά σημαντική

Στατιστική αξιολόγηση

- ο Ελέγχουμε την τιμή t στον πίνακα με τις οριακές τιμές (συνήθως με την τιμή που αντιστοιχεί στο **5%** επίπεδο σημαντικότητας) σε n_1+n_2-2 (degrees of freedom – df ή BE) όπου n_1 και n_2 ο αριθμός ατόμων στις 2 ομάδες

- ο **Αν $t \geq$ οριακή τιμή $\rightarrow$ απορρίπτω H_0 και συμπεραίνω ότι η σχέση είναι στατιστικά σημαντική**

Θυμηθείτε:

Κάθε t αντιστοιχεί σε μια πιθανότητα η διαφορά να οφείλεται σε τυχαίους λόγους,
 $p\text{-value} \leq 0.05$, άρα απορρίπτουμε τη H_0

- ο **Αν $t <$ οριακή τιμή $\rightarrow$ η σχέση δεν είναι στατ. σημαντική**

Θυμηθείτε:

Κάθε t αντιστοιχεί σε μια πιθανότητα η διαφορά να οφείλεται σε τυχαίους λόγους,
 $p\text{-value} > 0.05$, άρα **δεν** απορρίπτουμε τη H_0

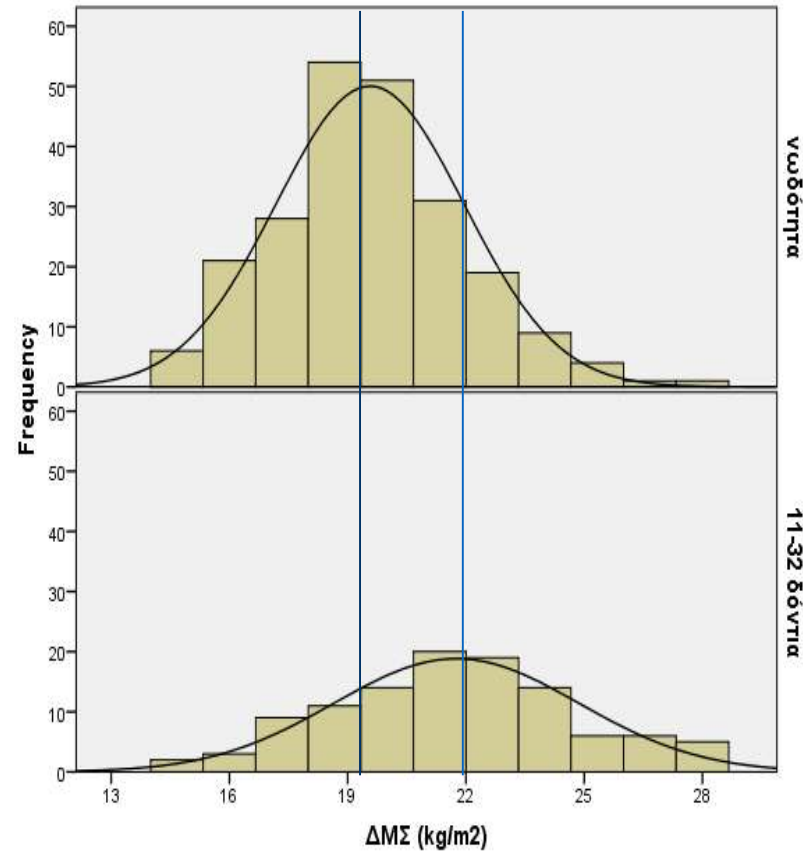
Παράδειγμα: σχέση στοματικής υγείας & ΔΜΣ

Ομάδα Α: $n_1=225$ άτομα με
νωδότητα

$\bar{x}_1=19.6$, $SD_1=2.4$

Ομάδα Β: $n_2=110$ άτομα
με 11 έως 32 δόντια

$\bar{x}_2=21.8$, $SD_2=3.1$



$$SE(\bar{X}_1 - \bar{X}_2) = S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad \text{όπου} \quad S = \sqrt{\left[\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)} \right]}$$

$$SE(\bar{X}_1 - \bar{X}_2) = 0.336$$

$$t = \frac{|\bar{X}_1 - \bar{X}_2|}{SE(\bar{X}_1 - \bar{X}_2)} = \frac{|19.6 - 21.8|}{0.336} = \frac{2.2}{0.336} = 6.56$$

και για BE = 225+110-2 = 333 ανατρέχουμε στους πίνακες

Αξιολόγηση στους πίνακες του t –test

$$t=6.56 > 1.97 \rightarrow$$

η διαφορά είναι στατιστικά σημαντική στο 5%

Συμπέρασμα: Ο ΔΜΣ διαφέρει σε βαθμό στατιστικά σημαντικό στις δύο ομάδες.

Ερμηνεία: Τα άτομα με 11-32 δόντια έχουν κατά μέσο όρο υψηλότερο ΔΜΣ από τα άτομα με πλήρη νωδότητα.

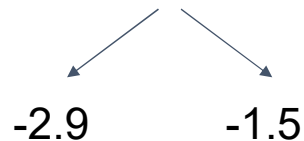
Degrees of freedom	Significance level		
	10%	5%	1%
1	6.314	12.706	63.657
2	2.920	4.303	9.925
3	2.353	3.182	5.841
4	2.132	2.776	4.604
5	2.015	2.571	4.032
6	1.943	2.447	3.707
7	1.894	2.365	3.499
8	1.860	2.306	3.355
9	1.833	2.262	3.250
10	1.812	2.228	3.169
.....			
60	1.671	2.000	2.660
70	1.667	1.994	2.648
80	1.664	1.990	2.639
90	1.662	1.987	2.632
100	1.660	1.984	2.626
120	1.658	1.980	2.617
150	1.655	1.976	2.609
200	1.653	1.972	2.601
300	1.650	1.968	2.592
400	1.649	1.966	2.588
500	1.648	1.965	2.586
600	1.647	1.964	2.584
∞	1.645	1.960	2.576

95% Διάστημα αξιοπιστίας της διαφοράς των δύο μέσων τιμών

Μέση Διαφορά στον ΔΜΣ: $\bar{X}_1 - \bar{X}_2 = 19.6 - 21.8 = -2.2$

$$(\bar{X}_1 - \bar{X}_2) \pm t_{n_1+n_2-2, 0.05} * SE_d$$

$$= -2.2 \pm 1.97 * 0.336 =$$


$$\begin{matrix} \swarrow & \searrow \\ -2.9 & -1.5 \end{matrix}$$

Ερμηνεία: Τα άτομα με πλήρη νωδότητα έχουν χαμηλότερο ΔΜΣ κατά μέσο όρο κατά -2.2 kg/m^2 από τα άτομα με 11-32 δόντια, με 95 % όρια αξιοπιστίας της διαφοράς $(-2.9, -1.5)$

Προσοχή: Δεν περιλαμβάνει το 0 (συμβαδίζει με τον έλεγχο υπόθεσης)

Παράδειγμα: Takemura et al. *Relation between Breastfeeding and the Prevalence of Asthma*. The Tokorozawa Childhood Asthma and Pollinosis Study. Am. Journal of Epidemiology, 2001.

TABLE 2. Comparison between subjects with asthma ($n = 2,315$) and controls ($n = 21,513$), according to selected risk factors, Tokorozawa Childhood Asthma and Pollinosis Study, Japan, February 2–17, 1998

Variables	Asthma (mean (SD*))		Controls (mean (SD))		p value
Age (years)	10.71 (2.45)		10.77 (2.53)		0.30†
	No.	%	No.	%	
Gender					
Male	1,429	61.7	10,676	49.6	<0.01‡
Female	886	38.3	10,837	50.4	
Parental smoking					
(+)	1,254	54.2	11,997	55.8	0.14‡
(–)	1,061	45.8	9,516	44.2	
Parental history of asthma					
(+)	522	22.6	1,898	8.8	<0.01‡
(–)	1,793	77.4	19,615	91.2	
Feeding pattern					
Breastfeeding only	992	42.9	8,620	40.1	<0.01‡
Mixed	966	41.7	9,134	42.4	
Artificial feeding	357	15.4	3,759	17.5	

* SD, standard deviation.

† p value for t test.

‡ p value for chi-squared test.

Έχει σχέση το άσθμα με φύλο του παιδιού?			
	Άσθμα		
Φύλο	Ναι	Όχι	Σύνολο
Αγόρι	1429 E=1176	10676 E=10929	12105
Κορίτσι	886 E=1139	10837 E=10584	11723
Σύνολο	2315	21513	23828

$$1429/2315 = 61.7\% \quad 10676/21513 = 49.6\% \quad 12105/23828 = 50.8\%$$

$$\chi^2 = \frac{(1429-1176)^2}{1176} + \frac{(10676-10929)^2}{10929} + \frac{(886-1139)^2}{1139} + \frac{(10837-10584)^2}{10584}$$

$$\chi^2 = 54.4 + 5.9 + 56.2 + 6.1 = 122.6$$

ΒΕ	10%	5%	1%	0,1%
1	2.71	3.84	6.64	10.83

Ωστόσο η δοκιμασία χ^2 δεν
δίνει το μέγεθος της σχέσης...

$$P < 0.001$$

Σχετικός λόγος (μέτρο σχετικού κινδύνου)

- Πολλές φορές στην ιατρική μελετάται η σχέση ανάμεσα σε ένα παράγοντα (έκθεση) και ένα αποτέλεσμα (π.χ. νόσημα). Η σχέση αυτή εκφράζεται με τους δείκτες σχετικού κινδύνου.
- Ένας τέτοιος δείκτης είναι ο Σχετικός Λόγος (**Odds Ratio, OR**) ο οποίος αποτελεί ένα μέτρο σχετικού κινδύνου.

Τρόπος υπολογισμού του σχετικού λόγου σε 2x2 πίνακες

	Νόσημα		
Παράγοντας	Ναι	Όχι	Σύνολο
Ναι	a	b	a+b
Όχι	c	d	c+d
Σύνολο	a+c	b+d	n

Σχετικός λόγος (odds ratio): πηλίκο διαγωνίων γινομένων:
 $OR = (a \cdot d) / (b \cdot c)$

Μας λέει πόσες φορές μεγαλύτερη «πιθανότητα» έχουν να νοσήσουν οι εκτεθειμένοι σε σχέση με τους μη εκτεθειμένους.

Παράδειγμα

	Καρκίνος ωοθηκών		
Αριθμός παιδιών	Ναι	Όχι	Σύνολο
0	48	30	78
1+	152	150	302
Σύνολο	200	180	380

$$\text{Σχετικός Λόγος (OR)} = (48 \times 150) / (30 \times 152) = 1.58$$

Συγκρίνει εκείνες που δεν έχουν παιδιά με εκείνες που έχουν ως προς την πιθανότητα να έχουν καρκίνο ωοθηκών

Γυναίκες που δεν είχαν παιδιά είχαν κατά 58% αυξημένο κίνδυνο καρκίνου των ωοθηκών σε σύγκριση με εκείνες που είχαν παιδιά

Παράδειγμα: Takemura et al. *Relation between Breastfeeding and the Prevalence of Asthma*. The Tokorozawa Childhood Asthma and Pollinosis Study. Am. Journal of Epidemiology, 2001.

TABLE 2. Comparison between subjects with asthma ($n = 2,315$) and controls ($n = 21,513$), according to selected risk factors, Tokorozawa Childhood Asthma and Pollinosis Study, Japan, February 2–17, 1998

Variables	Asthma (mean (SD*))		Controls (mean (SD))		p value
Age (years)	10.71 (2.45)		10.77 (2.53)		0.30†
	No.	%	No.	%	
Gender					
Male	1,429	61.7	10,676	49.6	<0.01‡
Female	886	38.3	10,837	50.4	
Parental smoking					
(+)	1,254	54.2	11,997	55.8	0.14‡
(–)	1,061	45.8	9,516	44.2	
Parental history of asthma					
(+)	522	22.6	1,898	8.8	<0.01‡
(–)	1,793	77.4	19,615	91.2	
Feeding pattern					
Breastfeeding only	992	42.9	8,620	40.1	<0.01‡
Mixed	966	41.7	9,134	42.4	
Artificial feeding	357	15.4	3,759	17.5	

* SD, standard deviation.

† p value for t test.

‡ p value for chi-squared test.

OR για άσθμα των αγοριών συγκριτικά με τα κορίτσια;

$$OR = \frac{1429 \times 10837}{886 \times 10676} = 1.64$$

Τα αγόρια είχαν κατά 64% αυξημένο κίνδυνο (ή 1,64 φορές μεγαλύτερο κίνδυνο) για άσθμα σε σύγκριση με τα κορίτσια

OR για άσθμα παιδιών με ασθματικούς γονείς συγκριτικά με παιδιά των οποίων οι γονείς δεν είχαν άσθμα;

$$OR = \frac{522 \times 19615}{1793 \times 1898} = 3.01$$

Τα παιδιά με ασθματικούς γονείς είχαν κατά 3 φορές μεγαλύτερο κίνδυνο για άσθμα σε σύγκριση με τα παιδιά που δεν είχαν γονείς με ιστορικό άσθματος