Ανάλυση ποσοτικών δεδομένων (III)

Ε. Σαμόλη
Καθηγήτρια Επιδημιολογίας και Ιατρικής Στατιστικής
Εργ. Υγιεινής, Επιδημιολογίας και Ιατρικής Στατιστικής
Ιατρική Σχολή, ΕΚΠΑ
esamoli@med.uoa.gr

Βάνα Σύψα
Καθηγήτρια Επιδημιολογίας και Ιατρικής Στατιστικής
Εργ. Υγιεινής, Επιδημιολογίας και Ιατρικής Στατιστικής
Ιατρική Σχολή, ΕΚΠΑ
vsipsa@med.uoa.gr

Πιθανά ερωτήματα προς διερεύνηση

1. Ποιο είναι το ύψος των αγοριών ηλικίας 6 ετών στην Ελλάδα;
2. Διαφέρει το ύψος των αγοριών σήμερα σε σχέση με το αντίστοιχο ύψος το 1981;
3. Διαφέρει το ύψος των αγοριών στην Ελλάδα σε σχέση με το ύψος των κοριτσιών;
4. **Μεταβάλλεται σημαντικά το ύψος των αγοριών από την ηλικία των 15 στην ηλικία των 17 ετών;**

Ερώτημα

- Μεταβάλλεται σημαντικά το ύψος των αγοριών από την ηλικία των 15 στην ηλικία των 17 ετών;
- Για να απαντήσω σε αυτό το ερώτημα, επιλέγω ένα δείγμα αγοριών ηλικίας 15 ετών και μετρώ το ανάστημά τους. Δύο χρόνια μετά, μετρώ ξανά το ανάστημά τους
- Με ενδιαφέρει να συγκρίνω τις μέσες τιμές στις 2 αυτές χρονικές στιγμές

- **Η διαφορά σε σχέση με τα προηγούμενα παραδείγματα:**
    - Οι μετρήσεις γίνονται στα ίδια άτομα (στην ίδια ερευνητική μονάδα) → **αντιστοιχία κατά ζεύγη**
    - Πλεονέκτημα: ένα σημαντικό μέρος της βιολογικής μεταβλητότητας που υπάρχει μεταξύ των ερευνητικών μονάδων εξαφανίζεται. Αυτό έχει σαν αποτέλεσμα πιο ακριβείς συγκρίσεις

t-test κατά ζεύγη (paired t-test)

Εφαρμόζεται όταν έχω ένα ποσοτικό μέγεθος όπου οι παρατηρήσεις συγκρινόμενες ομάδων εμφανίζουν ατομική αντιστοιχία (δεν είναι ανεξάρτητες)

1. Συνήθως **ίδια ερευνητική μονάδα**, π.χ.:
    - Μέτρηση της συστολικής αρτηριακής πίεσης στα ίδια άτομα, πριν και μετά από χρήση αντισυλληπτικών.
    - Σύγκριση αποτελεσματικότητας 2 φαρμάκων στους ίδιους ασθενείς (χορήγηση φαρμάκων σε 2 περιόδους).
2. Ή, οι μετρήσεις δε γίνονται στα ίδια άτομα, αλλά σε **εξομοιωμένα (matched) άτομα**, π.χ. σε δίδυμα αδέρφια, ή σε άτομα του ίδιου φύλου, ηλικίας, βάρους, κτλ.

- Τότε οι συγκρίσεις πρέπει να γίνονται **κατά ζεύγη**.
- **Προϋπόθεση**: Η κατανομή της διαφοράς να είναι κανονική.

Μορφή δεδομένων

Ύψος πριν	Ύψος μετά	Διαφορά d_i (μετά-πριν)
[redacted]		
174	180	6
[redacted]		
180	185	5
[redacted]		
173	176	3

Μέση διαφορά $\boldsymbol{\bar{d}}= \frac{\sum d_i}{n}=4.7$

Αν το ύψος δεν μεταβάλλεται από τα 15 στα 17 έτη ηλικίας, πώς θα αναμέναμε αυτή τη μέση διαφορά;

t-test κατά ζεύγη

Επειδή έχουμε παρατηρήσεις με ατομική αντιστοιχία, μας ενδιαφέρουν οι **διαφορές μεταξύ των μετρήσεων**. Συμβολίζουμε με $\boldsymbol{\mu_\delta}$ τη μέση διαφορά

H_0: $\mu_\delta = 0$

H_1: $\mu_\delta \neq 0$

- Τα **δείγματα** θα έχουν το ίδιο μέγεθος **n**

1. Υπολογίζουμε από το δείγμα μας: t=$|\bar{d}|/SE_d$

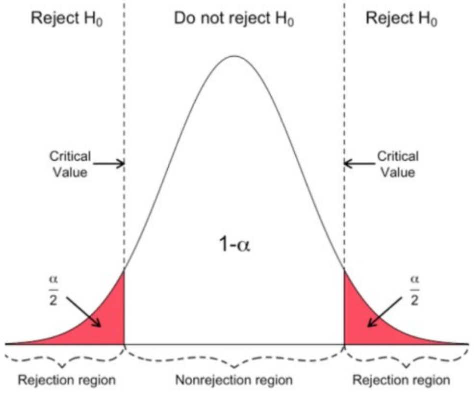
Υπολογισμός $\bar{d} = \frac{\sum d_i}{n}$ (ο μέσος όρος των διαφορών των ατόμων i=1,..,n)

Υπολογισμός $SE_d = \frac{S_d}{\sqrt{n}}$ όπου $S_d = \sqrt{\frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{n-1}}$

2. Κάτω από τη μηδενική υπόθεση: $|\bar{d}|/SE_d$ ακολουθεί την **t κατανομή με n-1 βαθμούς ελευθερίας** → Βρίσκουμε από πίνακες την τιμή $t_{n-1,\ 0.05}$

3. **Αν $|\bar{d}|/SE_d$ < t** $_{n\text{-}1,\ 0.05}$ → η διαφορά <u>δεν είναι</u> στατιστικά σημαντική (δεν απορρίπτω τη μηδενική υπόθεση)

   Αν $|\bar{d}|/SE_d$ > t $_{n\text{-}1,\ 0.05}$ → διαφορά <u>είναι</u> στατιστικά σημαντική (απορρίπτω τη μηδενική υπόθεση)



Παράδειγμα: Μεταβολή αναστήματος

- Σε υποθετικά δεδομένα, μετρήθηκε το ανάστημα σε δείγμα 500 αγοριών σε 2 χρονικές στιγμές: σε ηλικία 15 ετών και σε ηλικία 17 ετών

		Mean (SD) height (cm)	
	n	Ηλικία 15 ετών	Ηλικία 17 ετών
Αγόρια	500	172.4 (7.1)	176.6 (7.2)

- Μεταβλήθηκε σημαντικά το ύψος τους;

- Ανάστημα: ποσοτική μεταβλητή
- Μετρήθηκε το ανάστημα στα ίδια άτομα
  → κατά ζεύγη αντιστοιχία

t-test κατά ζεύγη

Παράδειγμα - συνέχεια

- t=$|\bar{d}|/SE_d$= … = 4.2/0.44 = 9.5
- $t_{n\text{-}1,\,0.05} = t_{499,\,0.05}$ =1.97

- Άρα t= 9.5 > 1.97

- Επομένως, η διαφορά είναι στατιστικά σημαντική στο 5% επίπεδο σημαντικότητας
- Είναι στατιστικά σημαντική και στο 1% (9.5>2.586)

→Το ανάστημα των αγοριών μεταβάλλεται σημαντικά από τα 15 στα 17 έτη

- Από στατιστικό πρόγραμμα: p<0.001

Η πιθανότητα η παρατηρούμενη διαφορά να έχει προκύψει κατά τύχη (ενώ δηλαδή ισχύει η μηδενική υπόθεση ότι δεν υπάρχει διαφορά στο ύψος) **είναι πάρα πολύ μικρή**

→απίθανο να έχει συμβεί τυχαία

→στατιστικά σημαντική διαφορά

	Significance level		
Degrees of freedom	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

Παράδειγμα: Είναι το φάρμακο Α αποτελεσματικό στη μείωση της συστολικής πίεσης;

Άτομα	Συστολική Πίεση πριν θεραπεία	Συστολική Πίεση μετά θεραπεία
1	160	164
2	170	134
3	170	150
4	152	140
5	142	121
6	144	132
7	136	141
8	147	153
9	148	132
10	126	99
11	147	126
12	139	133

Παρατηρηθείσα διαφορά d_i (πριν-μετά)
-4
36
20
12
21
12
-5
-6
16
27
21
6

$$\overline{d} = \frac{\sum_i d_i}{n} = \frac{156}{12} = 13mmHg$$

$$SD_d = \sqrt{\frac{\sum_i d_i^2 - \frac{(\sum_i d_i)^2}{n}}{n-1}} = \sqrt{\frac{3964 - \frac{(156)^2}{12}}{12-1}} = 13.27mmHg$$

$$SE = \frac{SD_d}{\sqrt{n}} = \frac{13.27}{\sqrt{12}} = 3.83mmHg$$

$$t = \frac{|\overline{d}|}{SE_d} = \frac{13}{3.83} = 3.39$$

Παράδειγμα: Είναι το φάρμακο Α αποτελεσματικό στη μείωση της συστολικής πίεσης;

- t= 3.39 > 2.201 (n=12-1=11)
- Στατιστικά σημαντική διαφορά στο επίπεδο 5% (και στο 1%)
- Άρα το φάρμακο Α μειώνει τα επίπεδα συστολικής πίεσης στους ασθενείς (κατά μέσο όρο κατά 13 mmHg)
- 95% ΔΕ της διαφοράς πριν-μετά:

$$\bar{d} \pm \boxed{t_{n\text{-}1,\,0.05} * SE_{\bar{d}}} =$$

13 ± 2,20 * 3.83 = (4.6, 21.6) mmHg

	Significance level		
Degrees of freedom	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
11	1.796	2.201	3.106

Ερμηνεία: Η χορήγηση του φαρμάκου μειώνει την συστολική πίεση κατά μέσο όρο 13 mmHg

Είματσε 95% σίγουροι ότι πραγματική μέση μείωση βρίσκεται μεταξύ 4,6 και 21,6 mmHg

Η τιμή 0 δεν περιλαμβάνεται στο 95% ΔΕ → σε συμφωνία με το αποτέλεσμα paired t-test

Παράδειγμα: Είναι αποτελεσματικό ένα υπνωτικό χάπι Α στην αύξηση των ωρών ύπνου;

- Κλινική δοκιμή για τη σύγκριση της αποτελεσματικότητας ενός υπνωτικού χαπιού σε σχέση με placebo
- Καταγράφηκαν οι συνεχόμενες ώρες ύπνου σε 10 άτομα με προβλήματα ύπνου ένα βράδυ που έλαβαν υπνωτικό και ένα βράδυ που έλαβαν placebo

Results of a placebo-controlled clinical trial to test the effectiveness of a sleeping drug.

	Hours of sleep		
Patient	Drug	Placebo	Difference
1	6.1	5.2	0.9
2	6.0	7.9	−1.9
3	8.2	3.9	4.3
4	7.6	4.7	2.9
5	6.5	5.3	1.2
6	5.4	7.4	−2.0
7	6.9	4.2	2.7
8	6.7	6.1	0.6
9	7.4	3.8	3.6
10	5.8	7.3	−1.5

- Μέση διάρκεια ύπνου με το Placebo: 5.58 h
- Μέση διάρκεια ύπνου με το υπνωτικό: 6.66 h
- Μέσο όρος επιπλέον ωρών ύπνου: 1.08 h
- $SE_d = SD/\sqrt{n} = 2.31/\sqrt{10} = 0.73$ h

- Είναι αυτή η διαφορά σημαντική;

Kirkwood & Sterne

Παράδειγμα: Είναι αποτελεσματικό ένα υπνωτικό χάπι Α στην αύξηση των ωρών ύπνου;

- 95% ΔΕ της διαφοράς στις ώρες ύπνου: $\boxed{\bar{d} \pm t_{n\text{-}1,\,0.05} * SE_{\bar{d}}}$

  $1.08 \pm 2.26 * 0.73 \rightarrow (-0.57, 2.73)$ h

- Τι συμπεραίνετε;

- Αντίστοιχα με εφαρμογή paired t-test:

  $t = |\bar{d}| / SE_d = 1.08/0.73 = 1.48$

- $t_{n\text{-}1,\,0.05} = t_{9,\,0.05} = 2.26$

- Άρα t= 1.48 <2.26

- Επομένως, η διαφορά δεν είναι στατιστικά σημαντική στο 5% επίπεδο σημαντικότητας
- Το υπνωτικό χάπι δεν φαίνεται να αυξάνει τις ώρες ύπνου
- Από στατιστικό πρόγραμμα: p=0.17

	Significance level		
Degrees of freedom	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
11	1.796	2.201	3.106

Αν αντί για t-test κατά ζεύγη χρησιμοποιούσαμε απλό t-test θα ήταν λάθος;

- Σε περιπτώσεις παρατηρήσεων κατά ζεύγη ενδείκνυται το t-test κατά ζεύγη γιατί η δοκιμασία αυτή είναι πιο **ισχυρή** από το απλό t-test.
  - **ισχύς → η πιθανότητα να ανιχνευθεί η αναμενόμενη διαφορά όταν υπάρχει**

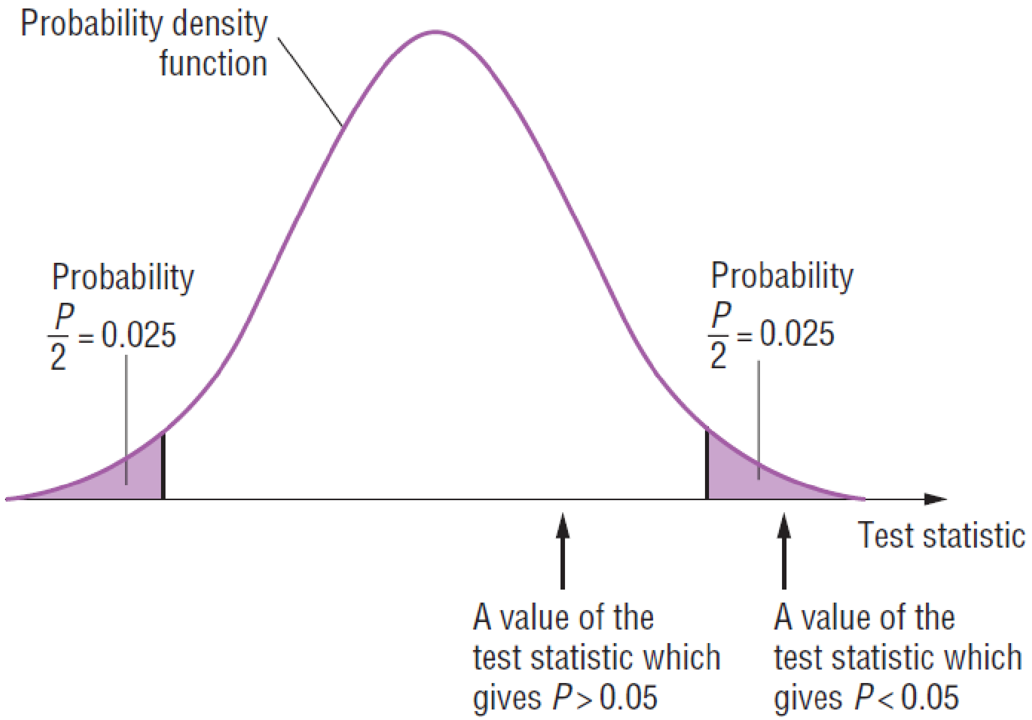
    ισχύς = 1-σφάλμα τύπου ΙΙ

- Δηλαδή, **τεκμηριώνει με μικρότερο αριθμό παρατηρήσεων την ενδεχόμενη στατιστική σημαντικότητα μιας πραγματικής διαφοράς.**

- Δηλαδή, αν εφαρμόσουμε απλό t-test
  - αν το αποτέλεσμα είναι στατιστικά σημαντικό τότε ερμηνεύουμε κανονικά το εύρημά μας
  - αν δεν προκύψει σημαντική διαφορά, αυτό μπορεί να οφείλεται στο ότι η δοκιμασία που επιλέχθηκε δεν έχει επαρκή ισχύ

Συνοπτικά τα 3 είδη t-test που συζητήσαμε

One sample t-test	$\frac{\lvert \bar{x} - \mu_0 \rvert}{\mathrm{SE}}$
t-test για ανεξάρτητα δείγματα	$\frac{\lvert \bar{X}_1 - \bar{X}_2 \rvert}{SE_{\bar{X}_1 - \bar{X}_2}}$
t-test για κατά ζεύγη παρατηρήσεις	$\frac{\lvert \bar{d} \rvert}{\mathrm{SE_d}}$



Συγκρίνοντας μέσες τιμές με τη βοήθεια των 95% ΔΕ

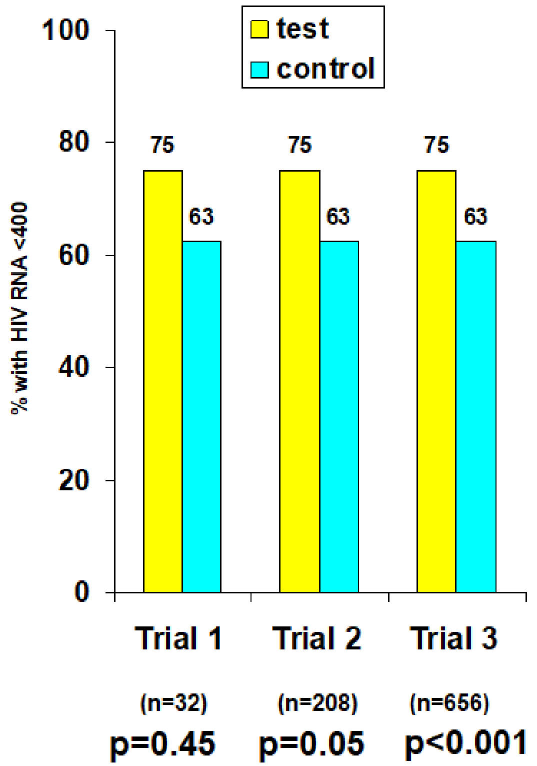
Σύγκριση μέσης τιμής με μία δοσμένη μ_0	$\bar{x} \pm 1.96 * SE$	• Αν το 95% ΔΕ δεν περιλαμβάνει τη δοσμένη τιμή μ_0, απορρίπτουμε τη μηδενική υπόθεση στο επίπεδο σημαντικότητας 5% (έχουμε βρει διαφορά) • Διαφορετικά, δεν απορρίπτουμε τη μηδενική υπόθεση If the C.I. includes μ0, then we fail to reject the null hypothesis at that level.
Σύγκριση 2 μέσων τιμών από ανεξάρτητα δείγμα	$(\bar{x}_1 - \bar{x}_2) \pm 1.96 * SE_{(\bar{x}1 - \bar{x}2)}$	• Αν το 95% ΔΕ δεν περιλαμβάνει το 0, απορρίπτουμε τη μηδενική υπόθεση στο επίπεδο σημαντικότητας 5%
Σύγκριση 2 μέσων τιμών από παρατηρήσεις κατά ζεύγη	$\bar{d} \pm 1.96 * SE_{\bar{d}}$	

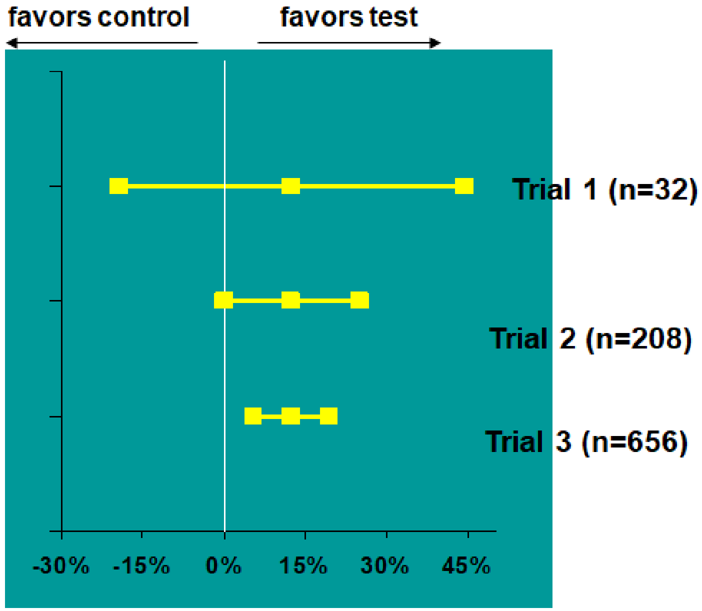
Παρατηρήσεις στον έλεγχο υποθέσεων και στο t-test

Παρατηρήσεις στον έλεγχο υποθέσεων

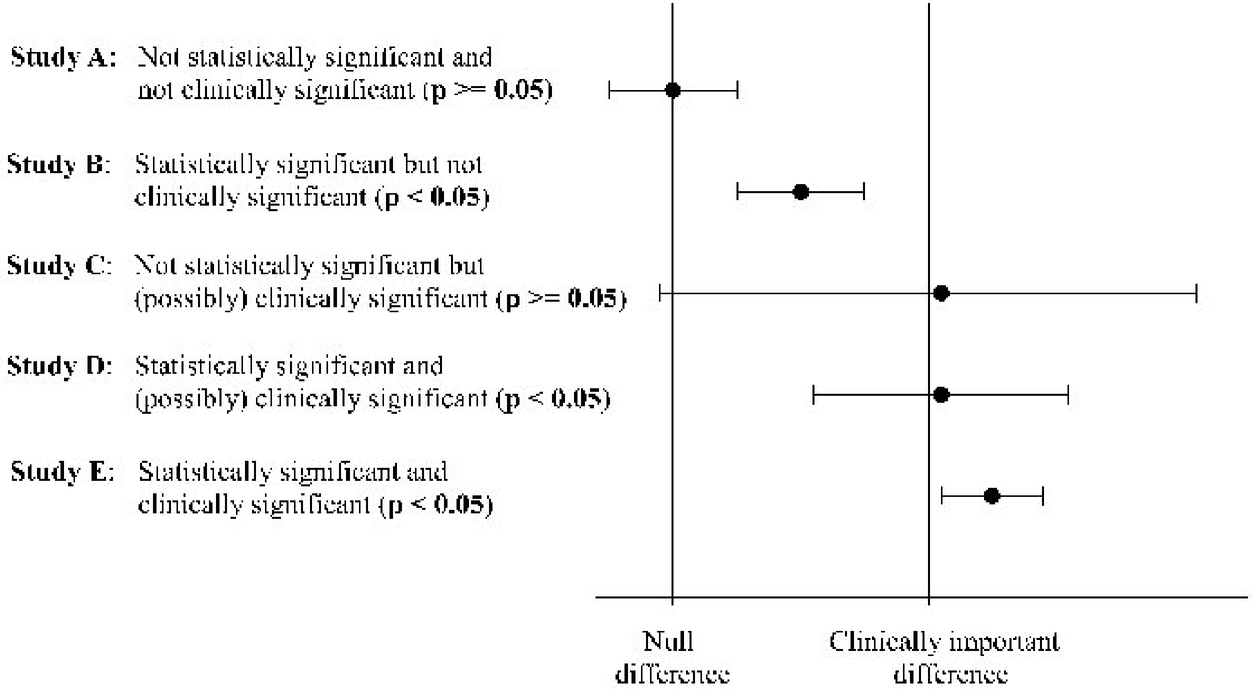
1. **Ένα στατιστικά σημαντικό εύρημα δεν είναι απαραίτητα και κλινικά σημαντικό**
   - Η p-value εξαρτάται από το μέγεθος του δείγματος της μελέτης μας
     - Το τυπικό σφάλμα μειώνεται όσο αυξάνεται το δείγμα (άρα η συνάρτηση t μεγαλώνει)
     - Επίσης, θυμηθείτε πώς μεταβάλλεται η κρίσιμη τιμή ανάλογα με τους βαθμούς ελευθερίας (όσο πιο μεγάλο δείγμα, τόσο μικρότερη)

- Αν το δείγμα της μελέτης είναι πολύ μεγάλο, ακόμα και πολύ μικρές διαφορές μπορεί να είναι στατιστικά σημαντικές
- Μία πολύ μικρή διαφορά σε μία μεγάλη μελέτη μπορεί να έχει το ίδιο p-value με μία μεγάλη διαφορά σε μία μικρή μελέτη





95% CI for difference in response rates



Παρατηρήσεις στον έλεγχο υποθέσεων

2. **Αν το αποτέλεσμα <u>δεν είναι στατιστικά σημαντικό</u> δεν σημαίνει απαραίτητα ότι δεν υπάρχει διαφορά**
    - Ισχύς της μελέτης & σφάλμα τύπου ΙΙ

3. **Το συμπέρασμα από την κατασκευή του 95% Διαστήματος Εμπιστοσύνης <u>συμπίπτει</u> με αυτό του αμφίπλευρου ελέγχου υπόθεσης**

Αν θέλουμε να συγκρίνουμε τις μέσες τιμές σε περισσότερες από 2 ομάδες;

- Το t-test για ανεξάρτητα δείγματα επιτρέπει τη σύγκριση **δύο** μέσων τιμών

- Αν υπάρχουν >2 ομάδες προς σύγκριση, εφαρμόζεται άλλη μεθοδολογία (**one-way analysis of variance**). Η μηδενική και εναλλακτική υπόθεση για π.χ. 3 ομάδες:

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

  H_1: τουλάχιστον μία μέση τιμή διαφέρει από τις υπόλοιπες

- Για δύο ομάδες: αντίστοιχο με t-test για ανεξάρτητα δείγματα

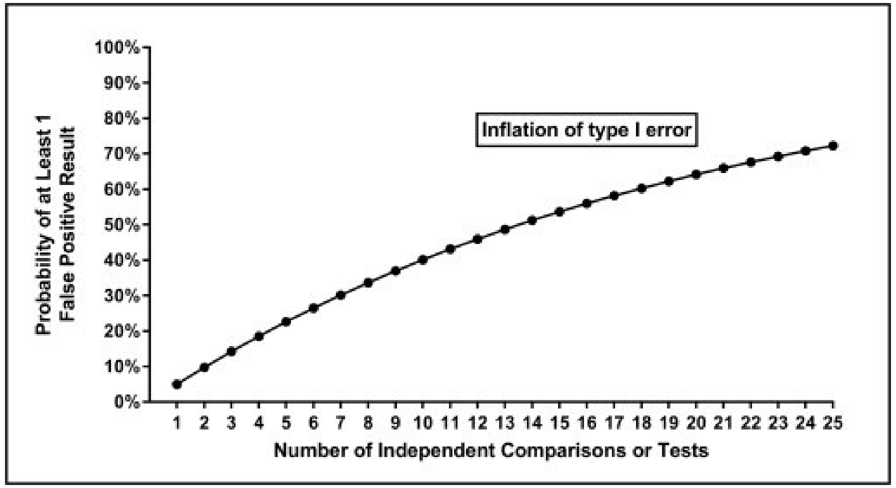
- <u>Δεν περιλαμβάνεται στην ύλη του μαθήματος</u>

Πολλαπλοί έλεγχοι

- Ας υποθέσουμε ότι αξιολογούμε τα επίπεδα πίεσης σε ασθενείς με υπέρταση μετά τη λήψη ενός από 4 αξιολογούμενα φάρμακα A, B, C, D
- Αν θέλω να συγκρίνω όλους τους δυνατούς συνδυασμούς:

A vs. B	A vs. C
A vs. D	B vs. C
B vs. D	C vs. D

- Εφαρμόζοντας π.χ. t-test για τη σύγκριση κάθε συνδυασμού με επίπεδο σημαντικότητας α=5%
  → έως 5% πιθανότητα να απορρίψω τη μηδενική υπόθεση όταν στην πραγματικότητα δεν υπάρχει διαφορά (**σφάλμα τύπου Ι**)
  Αν πραγματοποιήσουμε περισσότερα από ένα τεστ, η πιθανότητα να καταλήξω ότι υπάρχει διαφορά ενώ δεν υπάρχει θα είναι >5% (→ αυξάνεται το σφάλμα τύπου Ι)



- Στην πράξη αποφεύγουμε τις πολλαπλές συγκρίσεις. Υπάρχουν μέθοδοι που διορθώνουν τις συνέπειες αυτές των πολλαπλών συγκρίσεων π.χ. Bonferroni:

  significance level=5%/Number of comparisons

Π.χ. για 10 πολλαπλές συγκρίσεις με α=5%:

$1 - (1 - \alpha)^{10} =$

$1 - (1 - 0.05)^{10} = 0.40$

Υπολογισμός μεγέθους δείγματος

- Πόσα άτομα χρειάζομαι στη μελέτη μου για να μπορέσω να απαντήσω το ερώτημα που έχω θέσει;

- Γιατί πρέπει να εκτιμήσουμε το απαιτούμενο μέγεθος δείγματος πριν την έναρξη της μελέτης;
    - Μικρό δείγμα → Χαμηλή ισχύς της μελέτης
    - Μεγάλο δείγμα → Κόστος, χρόνος, ηθικά ζητήματα

- Ο σκοπός είναι να εκτιμηθεί ο «κατάλληλος» αριθμός ασθενών που απαιτούνται

Από τι επηρεάζεται το μέγεθος του δείγματος;

- Την απαιτούμενη ισχύ της μελέτης (τουλάχιστον 80%)

  → άρα σφάλμα τύπου ΙΙ ≤20%

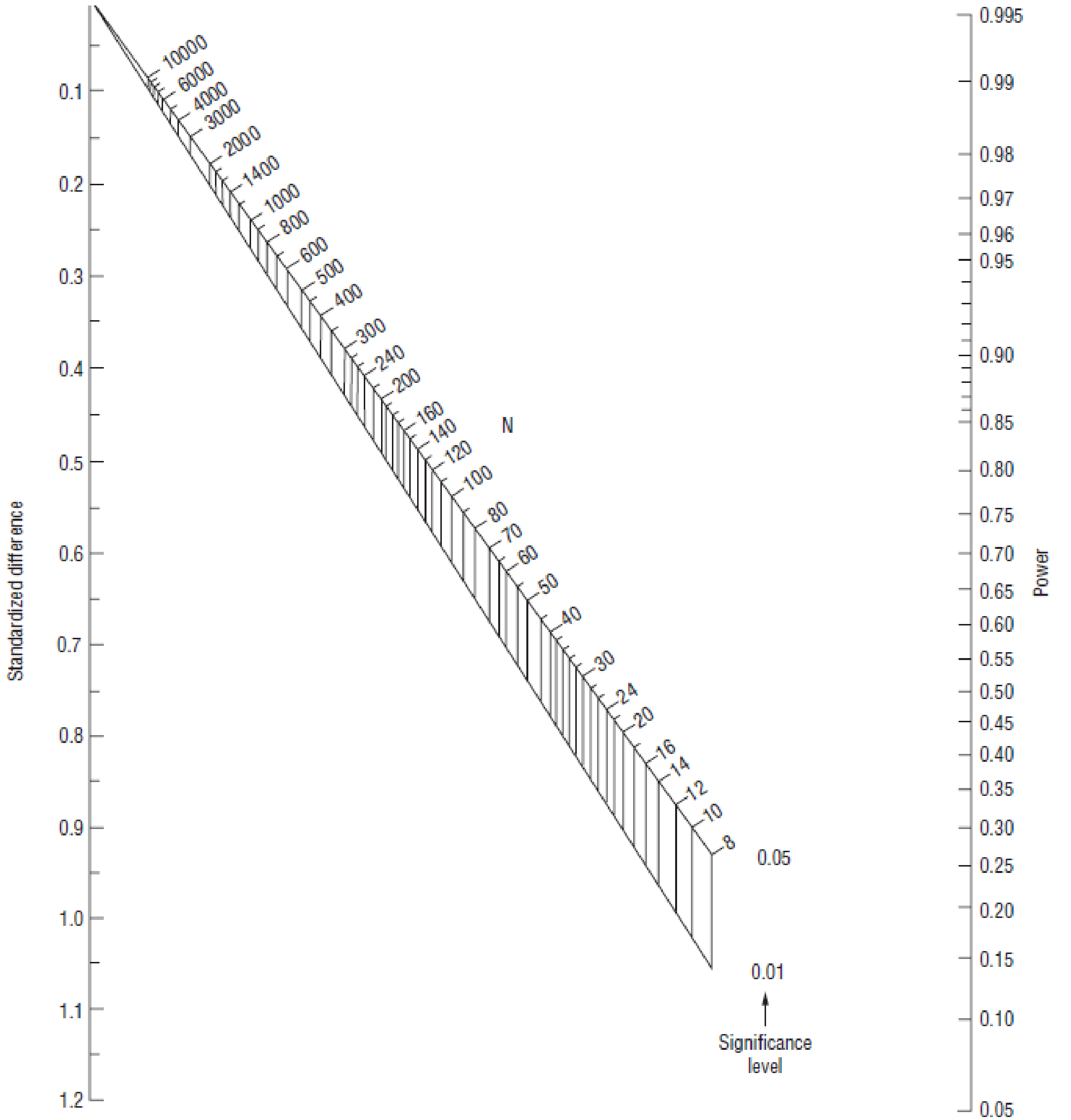
- Το επίπεδο σημαντικότητας που θα χρησιμοποιήσουμε στον έλεγχο υποθέσεων (5%)

- Τη μεταβλητότητα του μετρούμενου μεγέθους (πχ. τη σταθερή απόκλιση αν μας ενδιαφέρει ένα ποσοτικό χαρακτηριστικό)

- Το μέγεθος της αναμενόμενης διαφοράς ή της ελάχιστης διαφοράς που θεωρούμε ότι έχει κλινική σημασία (και άρα θα θέλαμε να είμαστε σε θέση να την εντοπίσουμε αν υπάρχει)

Πώς υπολογίζεται το απαιτούμενο μέγεθος δείγματος:

- Τύποι
- Ειδικοί πίνακες
- Νομογράμματα
- Στατιστικά πακέτα/προγράμματα



Extracted from: Altman, D.G. (1982) How large a sample? In: Gore, S.M. and Altman, D.G., editors. *Statistics in Practice*. London: BMA. With permission from Blackwell Publishing Ltd.

Παράδειγμα (αποτελεσματικότητα ακυκλοβίρης στην ερπητική λοίμωξη στόματος σε παιδιά)

Objective – to examine the effectiveness of aciclovir suspension (15 mg/kg) for treating 1- to 7-year-old children with herpetic gingivostomatitis lasting less than 72 hours.

Design – randomized double-blind placebo-controlled trial with 'treatment' administered five times a day for 7 days.
Main outcome measure for the determination of sample size – duration of oral lesions.

Sample size question – how many children are required in order to have a 90% power of detecting a 2.5 day difference in mean duration of oral lesions between the two groups at the 5% level of significance? The authors assume that the standard deviation of duration of oral lesions is approximately 5 days.

- Στο νομόγραμμα χρειάζομαι την τυποποιημένη διαφορά (standardised difference): διαφορά/SD

- Στο παράδειγμά μας: 2.5/5 = 0.50

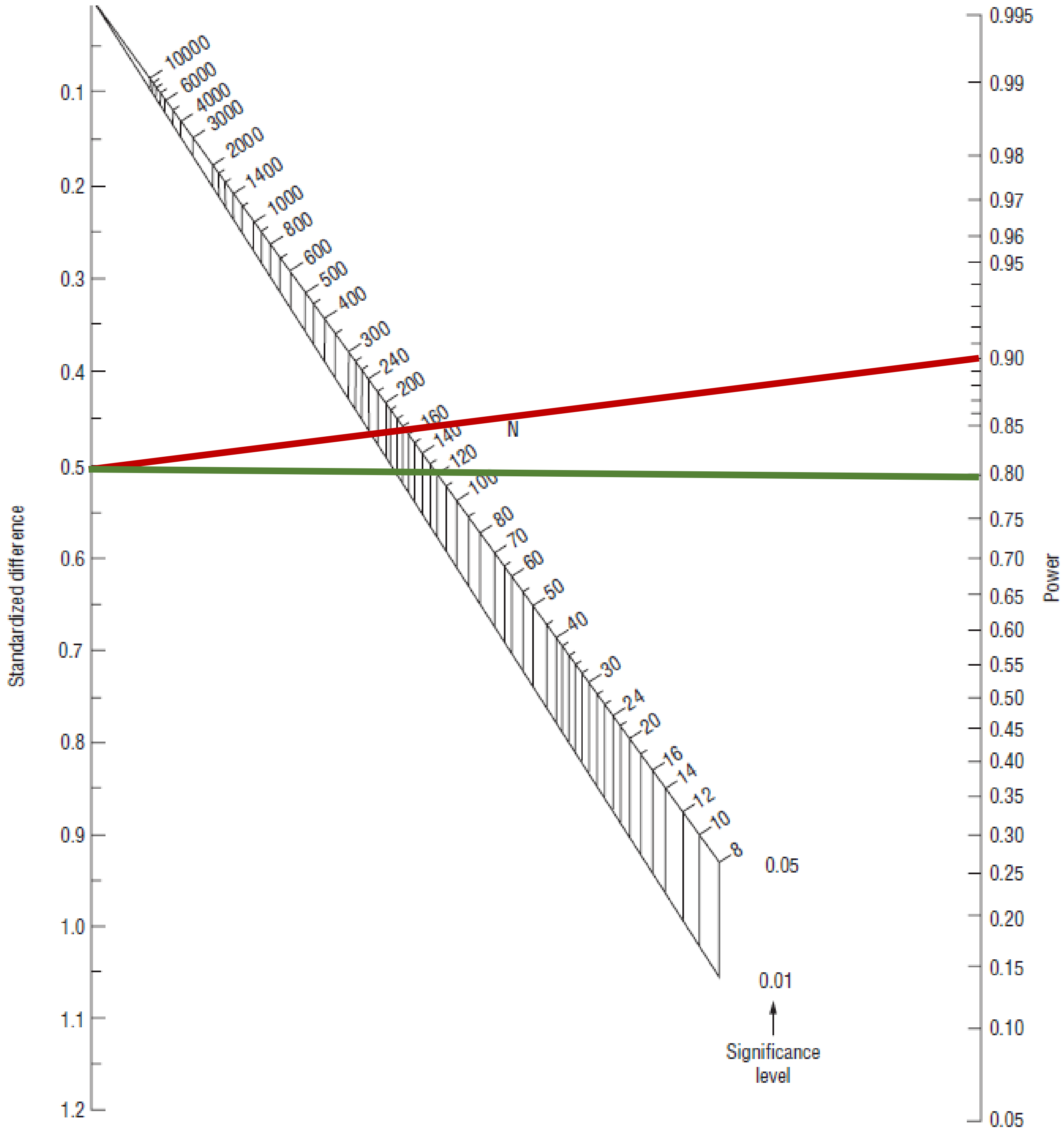
Από Petrie & Sabin

Standardized difference: 0.50
Επιθυμητή ισχύς: 90%
Επίπεδο σημαντικότητας: 5%

Περίπου 160 παιδιά απαιτούνται (80+80)

Αν αρκεστώ σε μικρότερη ισχύ; (80%)

- ~120 παιδιά (60+60)



Extracted from: Altman, D.G. (1982) How large a sample? In: Gore, S.M. and Altman, D.G., editors. *Statistics in Practice*. London: BMA. With permission from Blackwell Publishing Ltd.