

Απλή και πολλαπλή λογαριθμική εξάρτηση

Φίλιππος Ορφανός, PhD, Μέλος ΕΔΙΠ

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

03/12/2025, 17:05–18:30

Είδη δεδομένων

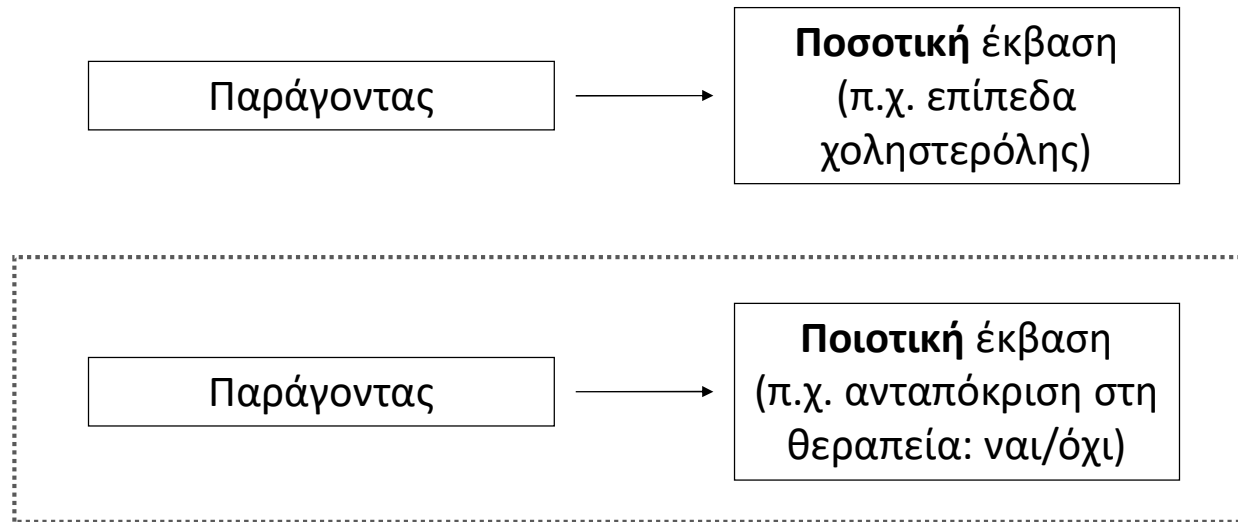
Σε μία επιδημιολογική έρευνα, καταγράφονται τα παρακάτω δεδομένα για κάθε άτομο:

- Έκθεση στον υπό μελέτη παράγοντα (π.χ. θεραπεία, κάπνισμα κλπ.)
- Άλλοι παράγοντες (ηλικία, φύλο κλπ.)
- Έκβαση (π.χ. ανταπόκριση στη θεραπεία, εμφάνιση νόσου κλπ.)

Είδη δεδομένων

- Ποσοτικές μεταβλητές (quantitative)
 - Π.χ. ηλικία, δείκτης μάζας σώματος, επίπεδα χοληστερόλης, διαστολική/συστολική πίεση κλπ.
- Ποιοτικές μεταβλητές (qualitative ή categorical)
 - Διχοτομικές: φύλο, επιβίωση
 - Μεταβλητές με περισσότερες από δύο κατηγορίες: ομάδα αίματος, φυλή
 - Διατάξιμες (ordinal): επίπεδο εκπαίδευσης, διαβάθμιση σοβαρότητας μίας ασθένειας

Πιθανά ερωτήματα σε μία επιδημιολογική έρευνα

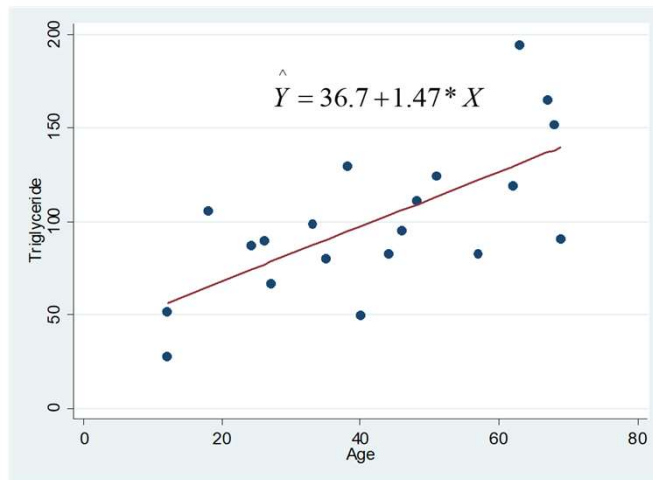


Στατιστικές δοκιμασίες για τη διερεύνηση σχέσης μεταξύ μίας
έκβασης και ενός ή περισσότερων παραγόντων

Παράγοντας	Έκβαση	
	Ποσοτική	Ποιοτική (με δύο επίπεδα)
Ποιοτική (με δύο επίπεδα)	t-test	χ^2
Ποσοτική	Συντελεστής συσχέτισης με αξιολόγηση Απλή γραμμική εξάρτηση	t-test
Πολλοί παραγόντες	Πολλαπλή γραμμική εξάρτηση	;

Διερεύνηση εξάρτησης ενός **ποσοτικού** μεγέθους από έναν άλλο παράγοντα (π.χ. εξάρτηση επιπέδων τριγλυκεριδίων από την ηλικία)

Απλή γραμμική εξάρτηση



- Η χρήση αυτής της μεθόδου επιτρέπει:
 - Να αξιολογήσουμε αν η ηλικία **επηρεάζει** τα επίπεδα τριγλυκεριδίων
 - Να κάνουμε **προβλέψεις** για τα αναμενόμενα επίπεδα τριγλυκεριδίων με βάση την ηλικία του ατόμου
- Προϋπόθεση:
 - Η εξαρτημένη μεταβλητή Y να έχει κανονική κατανομή
 - Η ευθεία γραμμή να είναι ικανοποιητική προσέγγιση της σχέσης Y και X

Διερεύνηση εξάρτησης ενός **ποιοτικού** χαρακτηριστικού από έναν άλλο παράγοντα
[π.χ. εξάρτηση παρουσίας/απουσίας στεφανιαίας νόσου από την ηλικία]

Αύξηση της ηλικίας σχετίζεται με αύξηση ή μείωση της
πιθανότητας να έχει το άτομο τη νόσο ($Y=0$ ή $Y=1$);

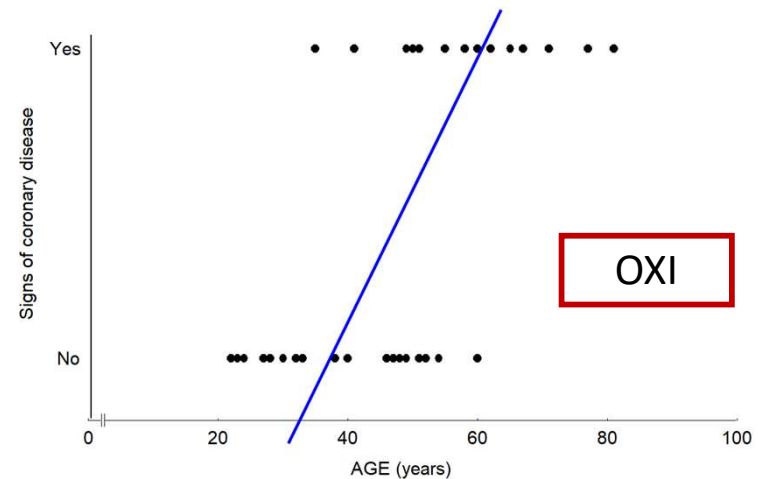
Age	CHD	Age	CHD	Age	CHD
22	0	40	0	54	0
23	0	41	1	55	1
24	0	46	0	58	1
27	0	47	0	60	1
28	0	48	0	60	0
30	0	49	1	62	1
30	0	49	0	65	1
32	0	50	1	67	1
33	0	51	0	71	1
35	1	51	1	77	1
38	0	52	0	81	1

Μέση ηλικία σε υγιείς: 38.6 έτη

Μέση ηλικία σε ασθενείς: 58.7 έτη

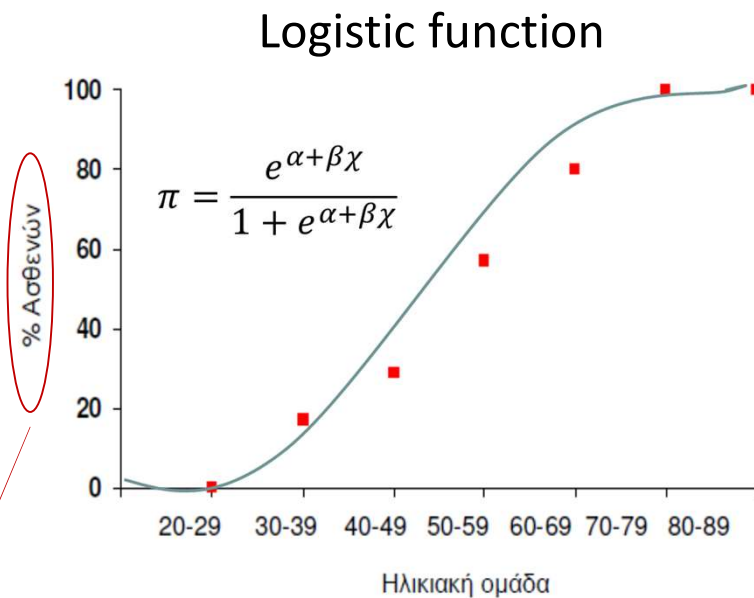
t-test: $p < 0.001$

Απλή γραμμική εξάρτηση;



Διερεύνηση εξάρτησης ενός **ποιοτικού** χαρακτηριστικού από έναν άλλο παράγοντα
[π.χ. εξάρτηση παρουσίας/απουσίας στεφανιαίας νόσου από την ηλικία]

Age group	# in group	Diseased	
		#	%
20 - 29	5	0	0
30 - 39	6	1	17
40 - 49	7	2	29
50 - 59	7	4	57
60 - 69	5	4	80
70 - 79	2	2	100
80 - 89	1	1	100



Πιθανότητα παρουσίας της νόσου = $\Pr(Y=1) = \pi$

Απλή λογαριθμική εξάρτηση

Πώς μπορώ να μετασχηματίσω τον τύπο σε μία πιο «οικεία» μορφή όπως αυτή της γραμμικής εξάρτησης ($Y = a + \beta X$);

$$\pi = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}$$

$$\Rightarrow \frac{\pi}{1 - \pi} = \frac{\frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}}{1 - \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}}$$

$$\Rightarrow \frac{\pi}{1 - \pi} = e^{\alpha + \beta x}$$

$$\Rightarrow \ln\left(\frac{\pi}{1 - \pi}\right) = \alpha + \beta x$$

Odds (λόγος συμπληρωματικών πιθανοτήτων)

$$odds = \frac{\text{πιθανότητα να συμβεί}}{\text{πιθανότητα να μη συμβεί}} = \frac{\pi}{1 - \pi}$$

Δεν έχει νόημα η ερμηνεία του β αλλά η ερμηνεία το e^b (το υψώνω στην εκθετική):

Τι είναι το e^b ;;;

Είναι το Odds Ratio (σχετικός λόγος) και ερμηνεύεται όπως ήδη γνωρίζουμε

Ερώτημα

- Υπάρχει σχέση μεταξύ μίας μεταβλητής (π.χ. θεραπεία A ή θεραπεία B) με μία ποιοτική μεταβλητή με δύο κατηγορίες (π.χ. ανταπόκριση στη θεραπεία ή όχι);



- Αν ναι, η σχέση αυτή είναι **ανεξάρτητη** από την ηλικία, το στάδιο νόσου ή άλλα χαρακτηριστικά;
→ Δηλαδή εξακολουθεί να υπάρχει ακόμα και αν λάβω υπόψη αυτά τα χαρακτηριστικά των ασθενών;

Παράδειγμα

Treatment	Response		Total
	Yes	No	
A	180 (60.0%)	120 (40.0%)	300 (100.0%)
B	130 (43.3%)	170 (56.7%)	300 (100.0%)

- Μπορώ να εφαρμόσω **δοκιμασία χ^2** σύγκρισης αναλογιών **χ^2 -test: $p < 0.001$**

- Μπορώ επίσης να υπολογίσω **το σχετικό λόγο:**

Εκφράζει πόσες φορές τα άτομα που λαμβάνουν τη θεραπεία A έχουν μεγαλύτερη (ή μικρότερη) πιθανότητα ανταπόκρισης σε σχέση με τα άτομα που λαμβάνουν τη θεραπεία B.

Odds ratio (σχετικός λόγος)

Treatment	Response	
	Yes	No
A	α	β
B	γ	δ

$$OR = \frac{\alpha\delta}{\beta\gamma}$$

- $OR \approx 1$: Τα άτομα που παίρνουν τη θεραπεία A έχουν παρόμοια πιθανότητα ανταπόκρισης σε σχέση με τα άτομα που παίρνουν τη θεραπεία B
- $OR > 1$: Τα άτομα που παίρνουν τη θεραπεία A έχουν μεγαλύτερη πιθανότητα ανταπόκρισης σε σχέση με τα άτομα που παίρνουν τη θεραπεία B
- $OR < 1$: Τα άτομα που παίρνουν τη θεραπεία A έχουν μικρότερη πιθανότητα ανταπόκρισης σε σχέση με τα άτομα που παίρνουν τη θεραπεία B

Παράδειγμα (συνέχεια)

Treatment	Response		Total
	Yes	No	
A	180 (60.0%)	120 (40.0%)	300 (100.0%)
B	130 (43.3%)	170 (56.7%)	300 (100.0%)

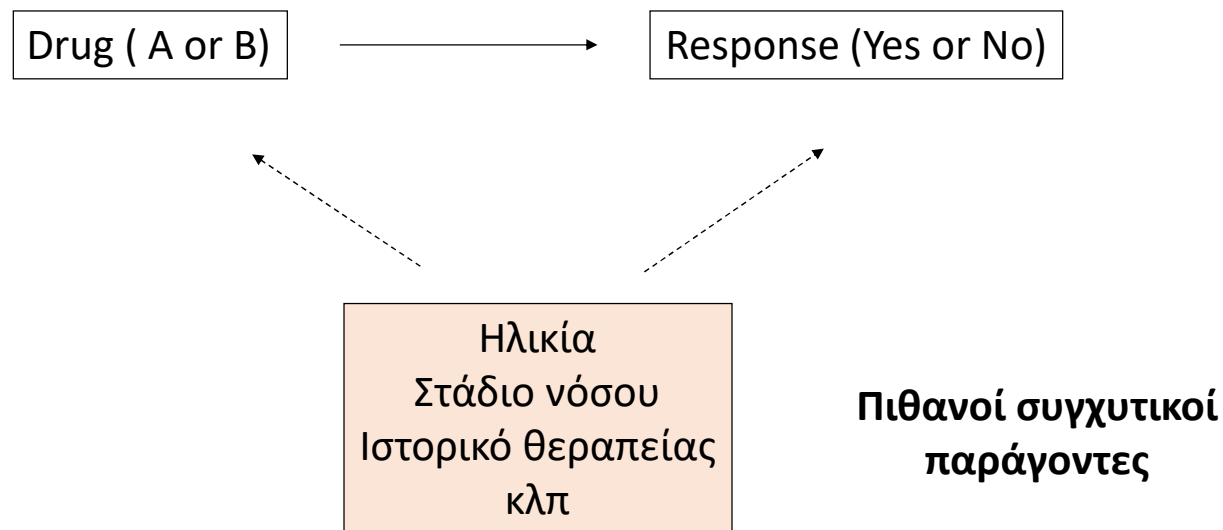
$$OR = \frac{180 \times 170}{130 \times 120} = 1.96$$

Ερμηνεία: Τα άτομα που λαμβάνουν τη **θεραπεία A** έχουν **σχεδόν διπλάσια πιθανότητα (ή 96% μεγαλύτερη πιθανότητα)** ανταπόκρισης στη θεραπεία σε **σχέση με** τα άτομα που λαμβάνουν τη θεραπεία B

**Μήπως το εύρημα οφείλεται σε
συγκυτικούς παράγοντες;**

Πιθανοί συγχυτικοί παράγοντες;

- Το ότι υπάρχει στατιστικά σημαντική σχέση μεταξύ δύο μεταβλητών δε σημαίνει απαραίτητα ότι υπάρχει μία αιτιολογική σχέση μεταξύ τους.
- Μπορεί π.χ. να υπάρχει μία τρίτη (ή 4^η, 5^η ...) μεταβλητή που επηρεάζει το αποτέλεσμα.



Έλεγχος συγχυτικών παραγόντων

- Στην απλή περίπτωση που υπάρχει ένας πιθανός συγχυτικός παράγοντας (π.χ. κάπνισμα), η επίδραση του μπορεί να εξουδετερωθεί με **διαστρωμάτωση (π.χ. η σχέση επανελέγχεται χωριστά στους καπνιστές και στους μη-καπνιστές)**
- **Όταν υπάρχουν πολλοί πιθανοί συγχυτικοί παράγοντες;**

Μοντέλα πολλαπλής εξάρτησης

- Τα μοντέλα πολλαπλής εξάρτησης επιτρέπουν:
 - να διερευνηθεί η επίδραση πολλών μεταβλητών ταυτόχρονα
 - να ελεγχθούν οι πιθανοί συγχυτικοί παράγοντες
- **Πολλαπλή λογαριθμική εξάρτηση**
(Multiple logistic regression model):
 - Χρησιμοποιείται όταν η **έκβαση** είναι δίτιμη **ποιοτική** μεταβλητή (π.χ. ναι/όχι)
Οι παράγοντες των οποίων την επίδραση μελετάμε μπορούν να είναι **ποιοτικές και ποσοτικές** μεταβλητές

Πολλαπλή λογαριθμική εξάρτηση

- **Γραμμική** εξάρτηση (όταν η έκβαση **ποσοτικό** χαρακτηριστικό)

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

- **Λογαριθμική** εξάρτηση (όταν η έκβαση **ποιοτικό** χαρακτηριστικό με 2 επίπεδα)

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

όπου p : η αναλογία των ατόμων με το υπό μελέτη συμβάν (π.χ. % ασθενών που ανταποκρίθηκαν)

Πολλαπλή λογαριθμική εξάρτηση: Ερμηνεία συντελεστών μερικής εξάρτησης β

- Odds ratio (OR)

$$OR = e^{\beta}$$

Συγκρίνει την πιθανότητα εμφάνισης του υπό μελέτη συμβάντος

- μεταξύ δύο ομάδων (όταν έχω ποιοτική ανεξάρτητη)

ή

- όταν η μεταβλητή X αυξηθεί κατά μία μονάδα (όταν έχω ποσοτική ανεξάρτητη)
και οι υπόλοιπες μεταβλητές είναι ίδιες

OR $\begin{cases} \nearrow > 1 & \text{αύξηση της } X_i \text{ συνεπάγεται αύξηση της πιθανότητας} \\ \rightarrow \cong 1 & \text{μεταβολή της } X_i \text{ δεν επιφέρει μεταβολή της πιθανότητας} \\ \searrow < 1 & \text{αύξηση της } X_i \text{ συνεπάγεται μείωση της πιθανότητας} \end{cases}$

Παράδειγμα: Γέννηση μωρού χαμηλού βάρους & βάρος, καπνιστικές συνήθειες μητέρας Απλή λογαριθμική εξάρτηση (2 χωριστά μοντέλα)

Low Birth Weight (1/0)	b	OR (ή exp(b))	P-value	95% Confidence Interval	
Βάρος μητέρας ανά 1 κγκ αύξηση)	-0.014	0.98	0.023	0.97	0.99

$$OR = e^{-0.014} = 0.98$$

Για κάθε 1 Kg αύξηση βάρους της μητέρας, η πιθανότητα γέννησης μωρού χαμηλού βάρους μειώνεται κατά $(1 - 0.98) = 0.02 = 2\%$

Low Birth Weight (1/0)	b	OR (ή exp(b))	P-value	95% Confidence Interval	
Κάπνισμα Ναι/Όχι	0.704	2.02	0.028	1.08	3.78

$$OR = e^{0.704} = 2.02$$

Οι καπνίστριες μητέρες έχουν σχεδόν διπλάσια πιθανότητα γέννησης μωρού χαμηλού βάρους

OR για την κατηγορία που δηλώνει η X (π.χ. ΝΑΙ) σε σχέση με την κατηγορία αναφοράς (π.χ. ΌΧΙ)

Εάν **δεν περιλαμβάνεται το 1** εντός των ορίων του **95% CI** του Odds Ratio, τότε το OR διαφέρει **στατιστικά σημαντικά** από το 1, επομένως η έκθεση συσχετίζεται στατιστικά σημαντικά με την έκβαση

Παράδειγμα: Γέννηση μωρού χαμηλού βάρους & βάρος, καπνιστικές συνήθειες μητέρας

Πολλαπλή λογαριθμική εξάρτηση (ένα μοντέλο με 2 ανεξάρτητες)

Συγκρίνει 2 μητέρες
με ίδιες καπνιστικές
συνήθειες

Low Birth Weight (1/0)	b	OR (ή exp(b))	P-value	95% Confidence Interval	
Βάρος μητέρας ανά 1 κγ αύξηση)	-0.013	0.98	0.029	0.97	0.99
Κάπνισμα Ναι/Όχι	0.677	1.97	0.037	1.04	3.72

Συγκρίνει 2 μητέρες
με ίδιο βάρος

$$OR = e^{-0.013} = 0.98$$

Για κάθε 1 Kg αύξηση βάρους της μητέρας, η πιθανότητα γέννησης μωρού χαμηλού βάρους μειώνεται κατά $(1 - 0.98) = 0.02 = 2\%$
ανεξάρτητα από την επίδραση του καπνίσματος

$$OR = e^{0.677} = 1.97$$

Οι καπνίστριες μητέρες έχουν σχεδόν διπλάσια πιθανότητα (ή 97%) γέννησης μωρού χαμηλού βάρους
ανεξάρτητα από την επίδραση του βάρους

Σύνοψη (I)

- ✓ Η **λογαριθμική εξάρτηση** μας επιτρέπει να διερευνήσουμε πώς διάφοροι παράγοντες (ανεξάρτητες μεταβλητές) επηρεάζουν την πιθανότητα κάποιου γεγονότος.
- ✓ Η εξαρτημένη μεταβλητή που εκφράζει την παρουσία ή απουσία του γεγονότος είναι **δίτιμη**.
- ✓ Οι ανεξάρτητες μεταβλητές μπορεί να είναι είτε **ποσοτικές** είτε **ποιοτικές** με δύο ή περισσότερες κατηγορίες.
- ✓ Στα μοντέλα λογαριθμικής εξάρτησης (όπως και σε αυτά της γραμμικής εξάρτησης) μπορούμε να έχουμε **μία ανεξάρτητη μεταβλητή** (μονοπαραγοντικά μοντέλα) ή **περισσότερες** (πολυπαραγοντικά μοντέλα).

Σύνοψη (II)

- ✓ Οι συντελεστές των ανεξάρτητων μεταβλητών (β), μετά από **αντιλογαρίθμηση (e^β)** εκφράζουν **Odds Ratios** είτε ως προς την κατηγορία αναφοράς (κατηγορικές μεταβλητές) είτε για μία μονάδα αύξησης (ποσοτικές μεταβλητές).
- ✓ Σε πολυπαραγοντικά μοντέλα εκτιμάμε επίδραση παραγόντων **ανεξάρτητα από τις υπόλοιπες μεταβλητές** που υπάρχουν στο συγκεκριμένο μοντέλο.
- ✓ Πιθανές **συγχυτικές** επιδράσεις διορθώνονται με την είσοδο των αντίστοιχων παραγόντων στο μοντέλο.