

**Παράδειγμα: Τα δεδομένα χρήσης
αντισυληπτικών (Little, 1998, Rodriguez, 2000)**

Age	Education	Desires more Children?	<u>Contraceptive use</u>		Total
			No	Yes	
<25	Lower	Yes	53	6	59
<25		No	10	4	14
<25	Higher	Yes	212	52	264
<25		No	50	10	60
25-29	Lower	Yes	60	14	74
25-29		No	19	10	29
25-29	Higher	Yes	155	54	209
25-29		No	65	27	92
30-39	Lower	Yes	112	33	145
30-39		No	77	80	157
30-39	Higher	Yes	118	46	164
30-39		No	68	78	146
40-49	Lower	Yes	35	6	41
40-49		No	46	48	94
40-49	Higher	Yes	8	8	16
40-49		No	12	31	43
Total			1100	507	1607

Πιθανά ερωτήματα

- Σχέση επιθυμίας για περισσότερα παιδιά με χρήση αντισυλληπτικών
- Σχέση ηλικίας με χρήση αντισυλληπτικών
- Σχέση επιθυμίας για περισσότερα παιδιά με χρήση αντισυλληπτικών, διορθωμένη για ηλικία

Πίνακας συνάφειας

Desires more children?	Contrac. use	
	Yes	No
Yes	219	753
No	288	347

- Αξιολόγηση σχέσης χρήσης αντισυλληπτικών με επιθυμία για περισσότερα παιδιά: μέσω χ^2

$$\chi^2=92.04$$

Βαθμοί ελευθερίας: $(K-1)(L-1) = (2-1)*(2-1)=1$

Αξιολόγηση στους πίνακες της χ^2 κατανομής:

BE	10%	5%	1%	1‰
1	2,71	3,84	6,64	10,83

$$92.04 > 10.83$$

Στατιστικά σημαντικό στο **1‰**
($p < 1‰$)

Πως θα ερμηνεύσω καλύτερα τη σχέση που υπάρχει;

Desires more children?	Contrac. use	
	Yes	No
Yes	219	753
No	288	347

$$OR = \frac{219 * 347}{288 * 753} = 0.35$$

Οι γυναίκες που επιθυμούν περισσότερα παιδιά έχουν 65% μικρότερη πιθανότητα ($1-0.35=0.65$) να χρησιμοποιούν αντισυλληπτικά από τις γυναίκες που δεν επιθυμούν παιδιά

Ανάλυση με λογαριθμιστική εξάρτηση

- π : πιθανότητα να χρησιμοποιεί μία γυναίκα αντισυλληπτικά
→ $\pi/(1-\pi)$: odds
- X : επιθυμεί (ναι/όχι) περισσότερα παιδιά
 - Θέλω να συγκρίνω αυτές που επιθυμούν σε σχέση με αυτές που δεν επιθυμούν, άρα:
 - $X=1$ → επιθυμούν παιδιά
 - $X=0$ → δεν επιθυμούν παιδιά (κατηγορία αναφοράς)

- Μοντέλο: $\ln\left(\frac{\pi}{1-\pi}\right) = \alpha + bX$ και εκτιμάται από

τα δεδομένα:

$$\ln\left(\frac{\pi}{1-\pi}\right) = -0.186 - 1.049X$$

Σύγκριση γυναικών που θέλουν παιδιά ($X=1$) σε σχέση με αυτούς που δεν θέλουν ($X=0$)

Ο συντελεστής $b = 1.049$

$$e^{-1.049} = 0.35$$

Όμοια με πριν

Πώς προκύπτει αυτό το OR από τη λογαριθμική εξάρτηση;

$$\left. \begin{aligned} \ln\left(\frac{\pi_{yes}}{1-\pi_{yes}}\right) &= a + \beta * 1 = a + \beta \\ \ln\left(\frac{\pi_{no}}{1-\pi_{no}}\right) &= a + \beta * 0 = a \end{aligned} \right\} \Rightarrow \ln\left(\frac{\pi_{yes}}{1-\pi_{yes}}\right) - \ln\left(\frac{\pi_{no}}{1-\pi_{no}}\right) = a + \beta - a$$

$$\Rightarrow \ln\left(\frac{\pi_{yes}}{1-\pi_{yes}}\right) - \ln\left(\frac{\pi_{no}}{1-\pi_{no}}\right) = \beta$$

$$\Rightarrow \ln\left(\frac{\frac{\pi_{yes}}{1-\pi_{yes}}}{\frac{\pi_{no}}{1-\pi_{no}}}\right) = \beta$$

Odds να χρησιμοποιεί
αντισυλληπτικά γυναίκα
που επιθυμεί παιδιά

Odds να χρησιμοποιεί
αντισυλληπτικά γυναίκα
που δεν επιθυμεί παιδιά

$$\Rightarrow \frac{\frac{\pi_{yes}}{1-\pi_{yes}}}{\frac{\pi_{no}}{1-\pi_{no}}} = \text{Odds Ratio (OR)} = e^{-1.049} = 0.35$$

Αξιολόγηση συντελεστών

- $\frac{\beta}{SE_{\beta}}$: Αξιολογείται στους πίνακες της κανονικής κατανομής ή t κατανομής στους **άπειρους** Β.Ε.
- 95% όρια αξιοπιστίας του OR:

$$e^{\beta \pm 1.96 SE_{\beta}}$$

Αξιολόγηση συντελεστών

- Από το μοντέλο δίνεται $SE_b=0.111$, άρα:

$$\frac{b}{SE_b} = \frac{-1.049}{0.111} = 89.78$$

t κατανομή στους άπειρους Β.Ε.

BE	10%	5%	1%	1‰
∞	1,65	1,96	2,58	3,29

89,78 > 3,29
p < 1‰

Επομένως, το b διαφέρει στατιστικά σημαντικά από το 0

(ή αντίστοιχα το odds ratio e^b διαφέρει από το 1)

→ στατιστικά σημαντική σχέση χρήσης αντισυλληπτικών και επιθυμίας παιδιών

95% OA

95%CI:

$$e^{\beta \pm 1.96 SE_{\beta}}$$

$$\exp(-1.049 \pm 1.96 * 0.111)$$

$$\exp(-1.267)$$

0.282

$$\exp(-0.831)$$

0.435

Επομένως OR (95% CI): 0.35 (0.28, 0.44)
(δεν περιλαμβάνεται η μονάδα)

Η επίδραση της ηλικίας

Η επίδραση του παράγοντα ηλικία μπορεί να διαπιστωθεί με παρόμοιο τρόπο θεωρώντας τον παρακάτω 2x4 πίνακα

Age	Contraceptive use		Total
	Yes	No	
<25	72	325	397
25-29	105	299	404
30-39	237	375	375
40-49	93	101	101
Total	507	1100	1607

Age	Contraceptive use	
	Yes	No
25-29	105	299
<25	72	325

$$OR_{25-29/<25} = \frac{105 * 325}{72 * 299} = 1.585$$

Age	Contraceptive use	
	Yes	No
30-39	237	375
<25	72	325

$$OR_{30-39/<25} = \frac{237 * 325}{72 * 375} = 2.853$$

Age	Contraceptive use	
	Yes	No
40-49	93	101
<25	72	325

$$OR_{40-49/<25} = \frac{93 * 325}{72 * 101} = 4.156$$

Η επίδραση της ηλικίας (με λογαριθμιστική εξάρτηση)

- Εφόσον η ηλικία είναι ένας παράγοντας με τέσσερα επίπεδα, για να εισαχθεί στο μοντέλο απαιτείται η δημιουργία ψευδομεταβλητών. Έτσι, θα δημιουργήσουμε 3 ψευδομεταβλητές X_1 , X_2 και X_3 όπως φαίνεται παρακάτω (εδώ η ομάδα των γυναικών ηλικίας <25 ετών χρησιμοποιείται ως κατηγορία αναφοράς).

Dummy variable	Age factor			
	<25	25-29	30-39	40-49
X_1	0	1	0	0
X_2	0	0	1	0
X_3	0	0	0	1

Το μοντέλο είναι της μορφής:

$$\log\left(\frac{\pi}{1-\pi}\right) = a + b_1 X_1 + b_2 X_2 + b_3 X_3$$

Αποτελέσματα και ερμηνεία

$$\ln\left(\frac{\pi}{1-\pi}\right) = -1.507 + 0.461X_1 + 1.048X_2 + 1.425X_3$$

Ερμηνεία:

$$\exp(0.461)=1.585$$

Γυναίκες ηλικίας 25-29 ετών είναι 1.6 φορές πιο πιθανό να χρησιμοποιούν αντισυλληπτικά σε σχέση με γυναίκες <25 ετών (ή 58.5% πιο πιθανό)

$$\exp(1.425)=4.156$$

Γυναίκες ηλικίας 40-49 ετών είναι 4.2 φορές πιο πιθανό να χρησιμοποιούν αντισυλληπτικά σε σχέση με γυναίκες <25 ετών.

Συνδυάζοντας 2x2 πίνακες συνάφειας: σχέση μεταξύ επιθυμίας παιδιών και χρήσης αντισύλληψης ανάλογα με την ηλικιακή ομάδα

-> age = <25			
Contraceptive use (Yes/No)	Desires more children?		Total
	No	Yes	
No	60	265	325
Yes	14	58	72
Total	74	323	397

-> age = 25-29			
Contraceptive use (Yes/No)	Desires more children?		Total
	No	Yes	
No	84	215	299
Yes	37	68	105
Total	121	283	404

-> age = 30-39			
Contraceptive use (Yes/No)	Desires more children?		Total
	No	Yes	
No	145	230	375
Yes	158	79	237
Total	303	309	612

-> age = 40-49			
Contraceptive use (Yes/No)	Desires more children?		Total
	No	Yes	
No	58	43	101
Yes	79	14	93
Total	137	57	194

OR=0.94 →

Στις νέες γυναίκες, αυτές που επιθυμούν παιδιά έχουν παρόμοια πιθανότητα χρήσης αντισυλληπτικών με αυτές που δεν επιθυμούν

OR=0.72

OR=0.32

OR=0.24

Πώς εξετάζω τη σχέση 2 μεταβλητών λαμβάνοντας υπόψη ένα πιθανό συγχυτικό παράγοντα;

- Στρωματοποιημένη ανάλυση (Stratified analysis): Εξετάζω τη σχέση αντισυλληπτικών-επιθυμίας παιδιών με διάστρωση κατά ηλικιακή ομάδα
 - Ένα odds ratio για τη σχέση αντισυλληπτικών-επιθυμίας παιδιών για κάθε ηλικιακή ομάδα
 - Σύνθεση των αποτελεσμάτων για εκτίμηση ενός «συνολικού» odds ratio το οποίο λαμβάνει υπόψη (διορθώνει) την ηλικία → Mantel-Haenszel:

$$OR_{M-H} = \frac{\sum a_i d_i / N_i}{\sum b_i c_i / N_i} = 0.43$$

Αξιολόγηση H_0 : $OR_{MH} = 1$ μέσω του τεστ M-H.

Ανάλυση μέσω λογαριθμιστικής παλινδρόμησης

Ανάλυση μέσω λογαριθμιστικής παλινδρόμησης: για να ελέγξω τη σχέση αντισυλληπτικών-επιθυμίας παιδιών διορθώνοντας για την ηλικία, αρκεί να προσθέσω στο μοντέλο και την ηλικία:

$$\ln\left(\frac{\pi}{1-\pi}\right) = -870 + 0.368AGE1 + 0.808AGE2 + 1.023AGE3 - 0.824CHILDREN$$

→ Διορθωμένος OR για τη σχέση αντισυλληπτικών-επιθυμίας παιδιών:
 $\exp(-0.824) = 0.44$

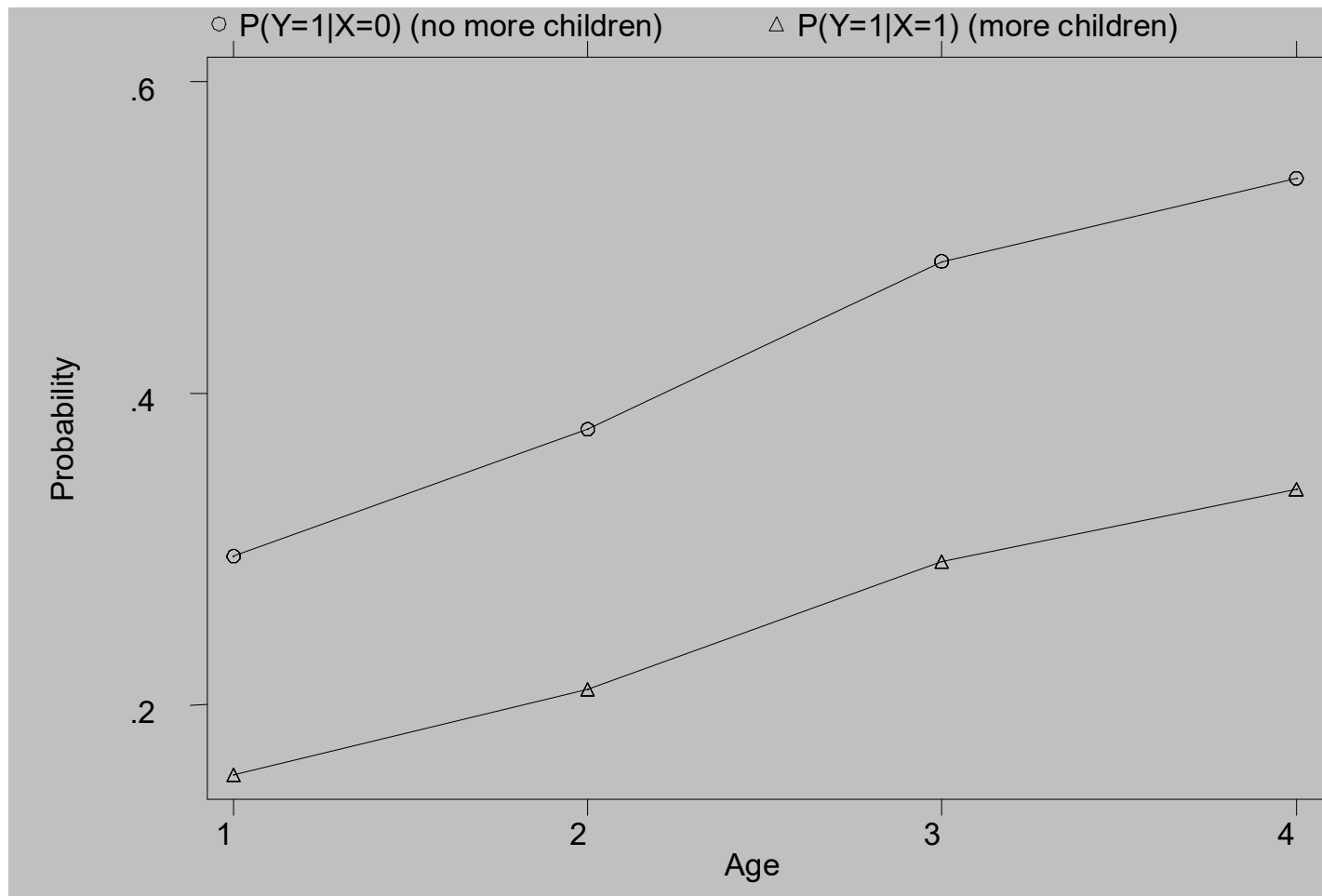
→ όπως και μέσω στρωματοποιημένης ανάλυσης κατά M-H.

Άρα: Αδρός (unadjusted) **OR = 0.35**

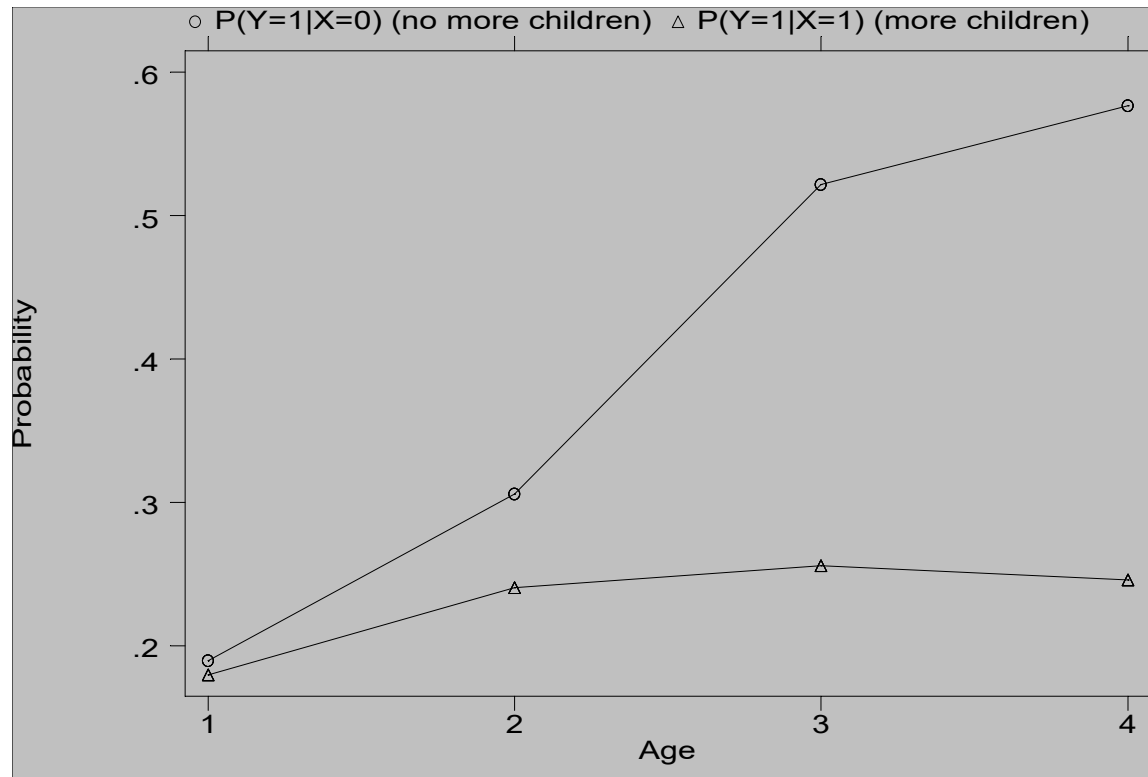
Διορθωμένος ως προς ηλικία (age-adjusted) **OR = 0.44**

$P < 0.001$

Γραφική αναπαράσταση του μοντέλου με διόρθωση



Γραφική αναπαράσταση του μοντέλου με αλληλεπίδραση



Αυτό σημαίνει ότι ενώ η πιθανότητα χρήσης αντισυλληπτικών αυξάνει με την ηλικία, η αύξηση αυτή είναι εντονότερη μεταξύ των γυναικών που δεν επιθυμούν περισσότερα παιδιά