

ΙΑΤΡΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

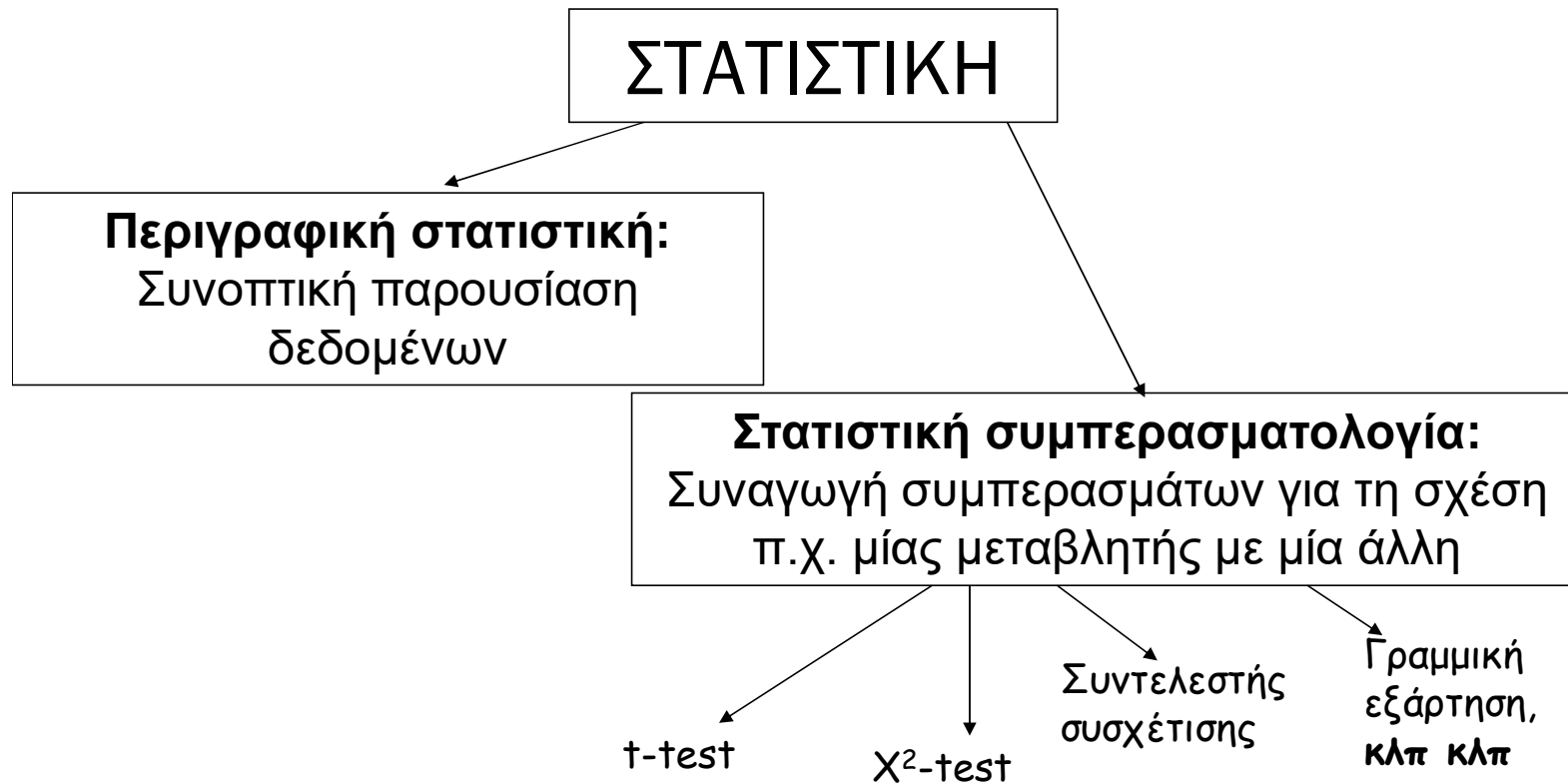
Επαναληπτικό III

Πολλαπλή γραμμική και λογαριθμιστική εξάρτηση

Φίλιππος Ορφανός, PhD

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

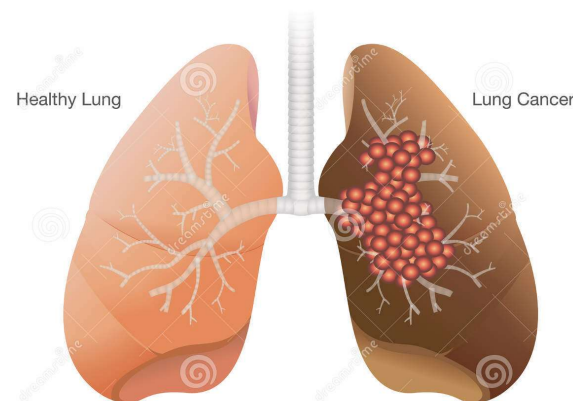
11/12/2024, 17:05–18:30



Στατιστικές δοκιμασίες σε επίπεδο σύγκρισης δύο ποιοτικών μεταβλητών



Καταναλωτές και
μη-καταναλωτές καφέ



Καρκίνος του πνεύμονα
(ναι ή όχι)

Σχετίζεται η κατανάλωση καφέ με τον καρκίνο του πνεύμονα;

Η κατάλληλη δοκιμασία είναι η χ^2

Στατιστικές δοκιμασίες σε επίπεδο σύγκρισης ποσοτικής με ποιοτική μεταβλητή

Σύγκριση δύο περιοχών με χαμηλά και υψηλά, αντίστοιχα, επίπεδα μόλυβδου στο περιβάλλον με σκοπό τη διερεύνηση της πιθανής διαφοράς στο ανάστημα των παιδιών.

Διαφέρει το ανάστημα;

Σαν πρώτο βήμα, θα υπολόγιζα το μέσο ύψος στις δύο περιοχές

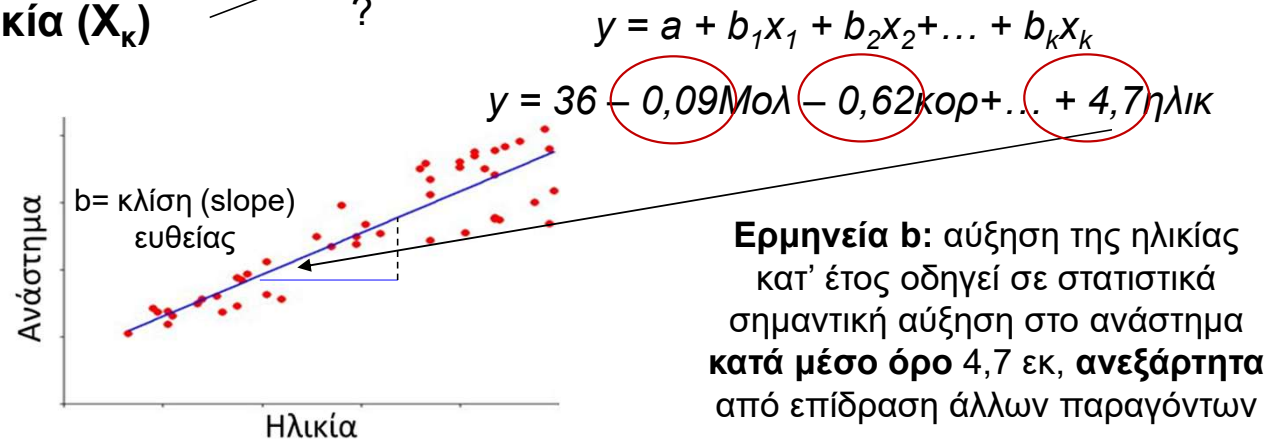
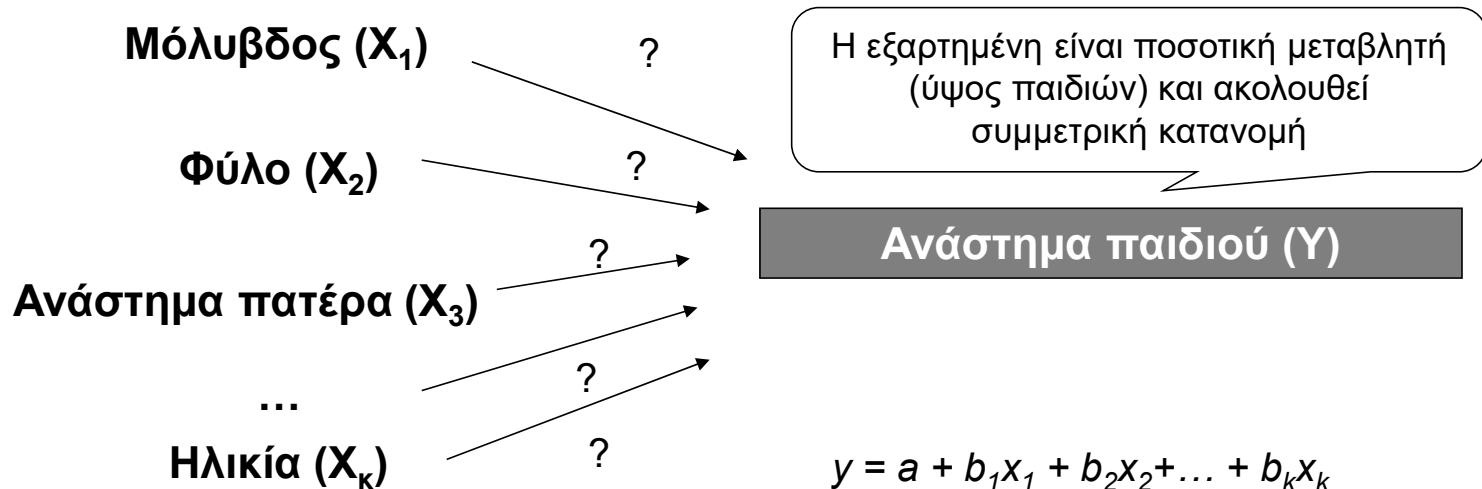
$$\bar{X}_1 = 123,7 \text{ cm}$$

$$\bar{X}_2 = 120,6 \text{ cm}$$

Είναι αυτή η διαφορά πραγματική;
ή οφείλεται στην τύχη;

Η κατάλληλη δοκιμασία είναι η **t (t-test)**

Από τι εξαρτάται το ανάστημα του παιδιού



Η κατάλληλη δοκιμασία είναι η **γραμμική εξάρτηση**

Παράδειγμα γραμμικής εξάρτησης

- Σε μελέτη για τη διερεύνηση της επίδρασης του μολύβδου στην σωματομετρική ανάπτυξη των παιδιών, μελετήθηκαν παιδιά σχολικής ηλικίας από τρεις περιοχές:
 - Λαύριο, Ελευσίνα και Λουτράκι
- Το συνολικό δείγμα αποτελείται από 522 παιδιά, 274 αγόρια και 248 κορίτσια ηλικίας 6-9 χρονών. Μέρος των δεδομένων παρουσιάζεται στον πίνακα που ακολουθεί

Πηγή: Kafourou et al, 1997.

(α) Ερμηνεύστε τους συντελεστές εξάρτησης.

(β) Επηρεάζει η έκθεση σε μόλυβδο το ανάστημα των παιδιών;

Κωδικός	Πόλη	Ηλικία (έτη)	Ανάστημα πατέρα (cm)	Μόλυβδος (μg/mL)	Ανάστημα παιδιού (cm)
353	2	8	172	23.42	116
419	2	.	165	51.17	107
19	1	8	152	.	114
26	1	7	177	5.94	122
506	2	7	155	20.21	119
683	3	8	170	4.16	117
612	3	7	164	9.78	112
97	1	8	164	.	121
504	2	7	172	17.29	113
469	2	9	170	26.98	124
498	2	7	160	13.24	110
565	2	8	168	22.94	123
140	1	8	162	2.86	115
374	2	6	155	26.59	112
673	3	6	172	5.69	119
644	3	8	167	11.87	123
507	2	8	177	10.19	125
711	3	7	165	4.15	124

Όπου υπάρχει . υποδεικνύει ελλείπουσα τιμή. Για την πόλη: 1 σημαίνει Λουτράκι, 2 Λαύριο και 3 Ελευσίνα.

Τελικό πολυπαραγοντικό μοντέλο για ύψος παιδιών

<u>Παράγοντες</u>	<u>Συντ. Εξ. (b)</u>	<u>SE(b)</u>	<u>t-test</u>	<u>P</u>
<i>Μόλυβδος</i> (μg/dl)	-0,086	0,037	-2,34	0,020
<i>Φύλο</i>				
Αγόρι		Κατηγορία αναφοράς		
Κορίτσι	-0,616	0,426	-1,45	0,149
<i>Ανάστημα πατέρα</i> (cm)	0,233	0,032	7,27	<0,001
<i>Επάγγελμα πατέρα</i>				
Ανειδίκευτος		Κατηγορία αναφοράς		
Ειδικευμένος	2,038	0,487	4,18	<0,001
<i>Ηλικία παιδιού</i> (έτη)	4,654	0,275	16,93	<0,001
<i>Αιμοσφαιρίνη</i> (g/dl)	0,736	0,235	3,14	0,002
<i>Πόλη διαμονής</i>				
Λουτράκι		Κατηγορία αναφοράς		
Λαύριο	1,179	0,719	1,64	0,102
Ελευσίνα	-0,214	0,612	-0,35	0,726
<i>Σταθερά</i>	36,280	6,607	5,49	<0,001
$R^2 = 0,45$				

$$\begin{aligned}
 E[\text{Υψος}] = & \\
 = & b_0 + \\
 & + b_1 * \text{Μόλυβδος} + \\
 & + b_2 * \text{Κορίτσι (1/0)} + \\
 & + b_3 * \text{Ανάστημα} + \\
 & + b_4 * \text{Ειδικευμ. (1/0)} + \\
 & + b_5 * \text{Ηλικία} + \\
 & + b_6 * \text{Αιμοσφαιρίνη} \\
 & + b_7 * \text{Λαύριο (1/0)} + \\
 & + b_8 * \text{Ελευσίνα (1/0)}
 \end{aligned}$$

Παράγοντες	Συντ. Εξ. (b)	SE(b)	t-test	P
<i>Μόλυβδος</i> (μg/dl)	-0,086	0,037	-2,34	0,020
Φύλο				
Αγόρι	Κατηγορία αναφοράς			
Κορίτσι	-0,616	0,426	-1,45	0,149
Ανάστημα πατέρα				
(cm)	0,233	0,032	7,27	<0,001
Επάγγελμα πατέρα				
Ανειδίκευτος	Κατηγορία αναφοράς			
Ειδικευμένος	2,038	0,487	4,18	<0,001
Ηλικία παιδιού				
(έτη)	4,654	0,275	16,93	<0,001
Αιμοσφαιρίνη				
(g/dl)	0,736	0,235	3,14	0,002
Πόλη διαμονής				
Λουτράκι	Κατηγορία αναφοράς			
Λαύριο	1,179	0,719	1,64	0,102
Ελευσίνα	-0,214	0,612	-0,35	0,726
Σταθερά	36,280	6,607	5,49	<0,001
R ² = 0,45				

Ποσοτικές μεταβλητές: Μόλυβδος

Για **1** µg/dl αύξηση στα επίπεδα **μολύβδου** παρατηρείται στατιστικά σημαντική μέση μείωση του αναστήματος του παιδιού **κατά 0,086 cm**

ανεξάρτητα από το φύλο, την ηλικία, την αιμοσφαιρίνη, τον τόπο διαμονής του παιδιού, καθώς και το ανάστημα και το επάγγελμα του πατέρα

ή αντίστοιχα

Για **10** µg/dl αύξηση στα επίπεδα **μολύβδου** παρατηρείται στατιστικά σημαντική μέση μείωση του αναστήματος του παιδιού **κατά 0,86 cm**

ανεξάρτητα από το φύλο, την ηλικία, την αιμοσφαιρίνη, τον τόπο διαμονής του παιδιού, καθώς και το ανάστημα και το επάγγελμα του πατέρα.

Παράγοντες	Συντ. Εξ. (b)	SE(b)	t-test	P
<i>Μόλυβδος</i> (μg/dl)	-0,086	0,037	-2,34	0,020
<i>Φύλο</i>				
Αγόρι				
Κορίτσι	-0,616	0,426	-1,45	0,149
<i>Ανάστημα πατέρα</i> (cm)	0,233	0,032	7,27	<0,001
<i>Επάγγελμα πατέρα</i>				
Ανειδίκευτος				
Ειδικευμένος	2,038	0,487	4,18	<0,001
<i>Ηλικία παιδιού</i> (έτη)	4,654	0,275	16,93	<0,001
<i>Αιμοσφαιρίνη</i> (g/dl)	0,736	0,235	3,14	0,002
<i>Πόλη διαμονής</i>				
Λουτράκι				
Λαύριο	1,179	0,719	1,64	0,102
Ελευσίνα	-0,214	0,612	-0,35	0,726
<i>Σταθερά</i>	36,280	6,607	5,49	<0,001
$R^2 = 0,45$				

Ποσοτικές μεταβλητές: Ανάστημα πατέρα

Για **1 cm** αύξηση στο **ανάστημα του πατέρα** παρατηρείται *στατιστικά σημαντική μέση* αύξηση του αναστήματος του παιδιού κατά **0,233 cm** ανεξάρτητα από τα επίπεδα μόλυβδου, το φύλο, την ηλικία, την αιμοσφαιρίνη και τον τόπο διαμονής του παιδιού, καθώς και το επάγγελμα του πατέρα.

Παράγοντες	Συντ. Εξ. (b)	SE(b)	t-test	P
<i>Μόλυβδος</i> (μg/dl)	-0,086	0,037	-2,34	0,020
<i>Φύλο</i>				
Αγόρι				
Κορίτσι	-0,616	0,426	-1,45	0,149
<i>Ανάστημα πατέρα</i> (cm)	0,233	0,032	7,27	<0,001
<i>Επάγγελμα πατέρα</i>				
Ανειδίκευτος				
Ειδικευμένος	2,038	0,487	4,18	<0,001
<i>Ηλικία παιδιού</i> (έτη)	4,654	0,275	16,93	<0,001
<i>Αιμοσφαιρίνη</i> (g/dl)	0,736	0,235	3,14	0,002
<i>Πόλη διαμονής</i>				
Λουτράκι				
Λαύριο	1,179	0,719	1,64	0,102
Ελευσίνα	-0,214	0,612	-0,35	0,726
<i>Σταθερά</i>	36,280	6,607	5,49	<0,001
$R^2 = 0,45$				

Ποσοτικές μεταβλητές: Ηλικία παιδιού

Για κάθε έτος αύξησης της **ηλικίας** του παιδιού παρατηρείται *στατιστικά σημαντική μέση* αύξηση του αναστήματος του παιδιού κατά **4,654** cm ανεξάρτητα από τα επίπεδα μόλυβδου, το φύλο, την αιμοσφαιρίνη και τον τόπο διαμονής του παιδιού, καθώς και το ανάστημα και το επάγγελμα του πατέρα.

<u>Παράγοντες</u>	<u>Συντ. Εξ. (b)</u>	<u>SE(b)</u>	<u>t-test</u>	<u>P</u>
<i>Μόλυβδος</i> (μg/dl)	-0,086	0,037	-2,34	0,020
Φύλο				
Αγόρι		Κατηγορία αναφοράς		
Κορίτσι	-0,616	0,426	-1,45	0,149
<i>Ανάστημα πατέρα</i> (cm)	0,233	0,032	7,27	<0,001
<i>Επάγγελμα πατέρα</i>				
Ανειδίκευτος		Κατηγορία αναφοράς		
Ειδικευμένος	2,038	0,487	4,18	<0,001
<i>Ηλικία παιδιού</i> (έτη)	4,654	0,275	16,93	<0,001
<i>Αιμοσφαιρίνη</i> (g/dl)	0,736	0,235	3,14	0,002
<i>Πόλη διαμονής</i>				
Λουτράκι		Κατηγορία αναφοράς		
Λαύριο	1,179	0,719	1,64	0,102
Ελευσίνα	-0,214	0,612	-0,35	0,726
<i>Σταθερά</i>	36,280	6,607	5,49	<0,001
R ² = 0,45]				

Ποιοτικές μεταβλητές με 2 κατηγορίες: Φύλο

- Το ανάστημα των κοριτσιών είναι χαμηλότερο από αυτό των αγοριών κατά μέσο όρο **κατά 0,616 cm** ανεξάρτητα από *τα επίπεδα μόλυβδου, την αιμοσφαιρίνη, την ηλικία, τον τόπο διαμονής του παιδιού, καθώς και το επάγγελμα του πατέρα*
- Η διαφορά μεταξύ των δύο φύλων **δεν** είναι στατιστικά σημαντική ($p=0,149$).

Παράγοντες	Συντ. Εξ. (b)	SE(b)	t-test	P
<i>Μόλυβδος</i> (μg/dl)	-0,086	0,037	-2,34	0,020
<i>Φύλο</i>				
Αγόρι				
Κορίτσι	-0,616	0,426	-1,45	0,149
<i>Ανάστημα πατέρα</i> (cm)	0,233	0,032	7,27	<0,001
<i>Επάγγελμα πατέρα</i>				
Ανειδίκευτος				
Ειδικευμένος	2,038	0,487	4,18	<0,001
<i>Ηλικία παιδιού</i> (έτη)	4,654	0,275	16,93	<0,001
<i>Αιμοσφαιρίνη</i> (g/dl)	0,736	0,235	3,14	0,002
<i>Πόλη διαμονής</i>				
Λουτράκι				
Λαύριο	1,179	0,719	1,64	0,102
Ελευσίνα	-0,214	0,612	-0,35	0,726
<i>Σταθερά</i>	36,280	6,607	5,49	<0,001
$R^2 = 0,45$				

Ποιοτικές μεταβλητές με >2 κατηγορίες: Πόλη διαμονής

- **Λαύριο με Λουτράκι:**
 - Τα παιδιά του **Λαυρίου** είναι υψηλότερα από τα παιδιά του **Λουτρακίου** κατά μέσο όρο κατά 1,179 cm, μετά τον έλεγχο για διαφορές ανάλογα με το φύλο, την ηλικία, τα επίπεδα μολύβδου και αιμοσφαιρίνης καθώς και το ανάστημα και το επάγγελμα του πατέρα
 - Η διαφορά **δεν** είναι στατιστικά σημαντική
- **Ελευσίνα με Λουτράκι**
 - Η μέση διαφορά στο ανάστημα των παιδιών του Λουτρακίου και των παιδιών της Ελευσίνας είναι -0,214 cm (με τα παιδιά της Ελευσίνας να είναι χαμηλότερα), αφού ελήφθησαν υπόψη οι διαφορές που οφείλονται στο φύλο, στην ηλικία, στα επίπεδα μολύβδου και αιμοσφαιρίνης καθώς και το ανάστημα και το επάγγελμα του πατέρα
 - Η διαφορά **δεν** είναι στατιστικά σημαντική

Παράγοντες	Συντ. Εξ. (b)	SE(b)	t-test	P
<i>Μόλυβδος</i> (μg/dl)	-0,086	0,037	-2,34	0,020
<i>Φύλο</i>				
Αγόρι				
Κορίτσι	-0,616	0,426	-1,45	0,149
<i>Ανάστημα πατέρα</i> (cm)	0,233	0,032	7,27	<0,001
<i>Επάγγελμα πατέρα</i>				
Ανειδίκευτος				
Ειδικευμένος	2,038	0,487	4,18	<0,001
<i>Ηλικία παιδιού</i> (έτη)	4,654	0,275	16,93	<0,001
<i>Αιμοσφαιρίνη</i> (g/dl)	0,736	0,235	3,14	0,002
<i>Πόλη διαμονής</i>				
Λουτράκι				
Λαύριο	1,179	0,719	1,64	0,102
Ελευσίνα	-0,214	0,612	-0,35	0,726
<i>Σταθερά</i>	36,280	6,607	5,49	<0,001
$R^2 = 0,45$				

Αν δεν σας δινόταν η τιμή του ρ , πως θα αξιολογούσατε αν οι μεταβλητές είναι στατιστικά σημαντικές;

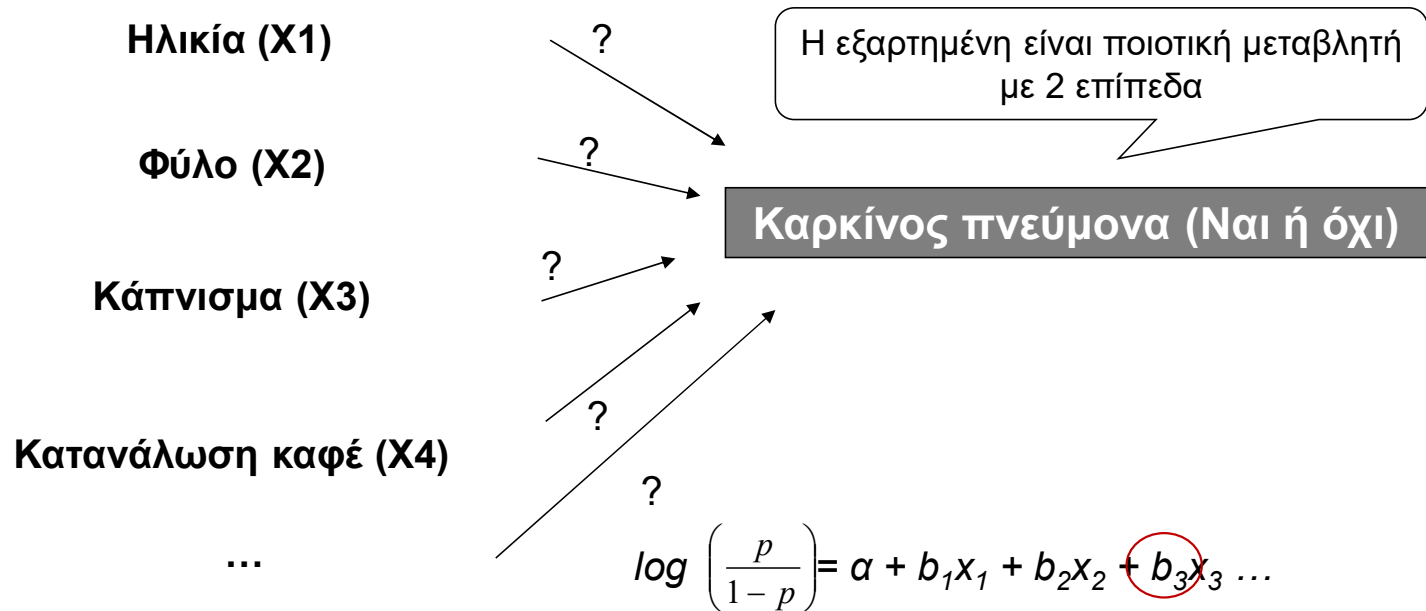
- Υπολογισμός 95%CI του b :

$$b \pm 1.96 * SE(b)$$

- Αν δεν περιλαμβάνεται το 0 εντός των ορίων, τότε ο συντελεστής b διαφέρει στατιστικά σημαντικά από το 0 \rightarrow επομένως η αντίστοιχη μεταβλητή συσχετίζεται στατιστικά σημαντικά με την εξαρτημένη (ανάστημα παιδιού)

<u>Παράγοντες</u>	<u>Συντ. Εξ. (b)</u>	<u>SE(b)</u>	<u>95% Δ.Α. του b</u>
<i>Μόλυβδος</i> (μg/dl)	-0,086	0,037	-0,086 ± 1,96 *0,037 [-0,159 έως -0,013]
<i>Φύλο</i>			
Αγόρι		Κατηγορία αναφοράς	
Κορίτσι	-0,616	0,426	[-1,451 έως 0,219]
<i>Ανάστημα πατέρα</i> (cm)	0,233	0,032	[0,170 έως 0,296]
<i>Επάγγελμα πατέρα</i>			
Ανειδίκευτος		Κατηγορία αναφοράς	
Ειδικευμένος	2,038	0,487	[1,083 έως 2,993]
<i>Ηλικία παιδιού</i> (έτη)	4,654	0,275	[4,115 έως 5,193]
<i>Αιμοσφαιρίνη</i> (g/dl)	0,736	0,235	[0,275 έως 1,197]
<i>Πόλη διαμονής</i>			
Λουτράκι		Κατηγορία αναφοράς	
Λαύριο	1,179	0,719	[-0,230 έως 2,589]
Ελευσίνα	-0,214	0,612	[-1,414 έως 0,986]
<i>Σταθερά</i>	36,280	6,607	
$R^2 = 0,45$			

Από τι εξαρτάται η πιθανότητα εμφάνισης καρκίνου του πνεύμονα



$$e^{b_3} = 10 = OR$$

Ερμηνεία OR: οι καπνιστές έχουν π.χ. 10 φορές > κίνδυνο εμφάνισης...

$$b < 0$$

→

OR < 1 Προστατευτικός παράγοντας

$$b > 0$$

→

OR > 1 Επιβαρυντικός παράγοντας

Η κατάλληλη δοκιμασία είναι η **λογαριθμιστική εξάρτηση**

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
	No.	%	No.	%	
Age (years)	10.71 (2.45)		10.77 (2.53)		0.30†
Gender					<0.01‡
Male	1,429	61.7	10,676	49.6	
Female	886	38.3	10,837	50.4	
Parental smoking					0.14‡
(+)	1,254	54.2	11,997	55.8	
(-)	1,061	45.8	9,516	44.2	
Parental history of asthma					<0.01‡
(+)	522	22.6	1,898	8.8	
(-)	1,793	77.4	19,615	91.2	
Feeding pattern					<0.01‡
Breastfeeding only	992	42.9	8,620	40.1	
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;

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

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

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

OR να μην παρουσιάσουν άσθμα παιδιών με ασθματικούς γονείς συγκριτικά με παιδιά των οποίων οι γονείς δεν είχαν άσθμα;

$$OR = \frac{1898 \cdot 1793}{522 \cdot 19616} = 0.33$$

Παράδειγμα λογαριθμιστικής εξάρτησης

Το Ελληνικό Ίδρυμα Γαστρεντερολογίας και Διατροφής διεξήγαγε έρευνα για την εκτίμηση της συχνότητας και τη διερεύνηση προγνωστικών παραγόντων της παχυσαρκίας σε μαθητές Δημοτικών και Γυμνασίων. Στη μελέτη συμμετείχαν συνολικά 1818 μαθητές ηλικίας 6 έως 16 ετών. Για τη διερεύνηση προγνωστικών παραγόντων παχυσαρκίας εφαρμόστηκε λογαριθμιστική εξάρτηση, τα αποτελέσματα της οποίας παρουσιάζονται στον Πίνακα που ακολουθεί:

Ερμηνεύστε τους Σχετικούς Λόγους.

Μοντέλα για δίτιμες μεταβλητές

- Όταν η εξαρτημένη είναι **συνεχής** μεταβλητή, μας ενδιαφέρει να εκτιμήσουμε πως μεταβάλλονται τα επίπεδα της κατά μέσο όρο αν αυξηθεί η ανεξάρτητη μεταβλητή κατά μία μονάδα. π.χ.
 - αύξηση της διάρκειας βάρδισης κατά 1 ώρα την εβδομάδα συντελεί σε μείωση του βάρους κατά 300γρ. κατά μέσο όρο
- Όταν η εξαρτημένη είναι **δίτιμη** μεταβλητή (π.χ. απουσία ή παρουσία νόσου), μας ενδιαφέρει να εκτιμήσουμε αν η αύξηση της ανεξάρτητης μεταβλητής κατά μία μονάδα (π.χ. αριθμός τσιγάρων) σχετίζεται με τη μέση αύξηση (ή μείωση) της πιθανότητας να έχει ένα άτομο τη νόσο

Ερμηνεύστε τους σχετικούς λόγους (odds ratios)

<u>Παράγοντας</u>	<u>OR</u>	<u>95% CI</u>	<u>P</u>
<i>Ηλικία μαθητή</i>			
Παιδική ηλικία	Κατηγορία αναφοράς		
Πρώτη εφηβεία	0,86	0,67-1,11	0,244
Εφηβεία	0,35	0,23-0,52	<0,001
<i>Φύλο μαθητή</i>			
Αγόρι	Κατηγορία αναφοράς		
Κορίτσι	0,73	0,58-0,92	0,007
<i><u>ΔΜΣ γονέων</u></i>			
Δύο γονείς φυσιολογικού βάρους	Κατηγορία αναφοράς		
1 γονέας υπέρβαρος /παχύσαρκος	1,92	1,39-2,64	<0,001
Δύο γονείς υπέρβαροι /παχύσαρκοι	4,29	3,00-6,13	<0,001
<i>Ηλικία μητέρας (έτη)</i>	0,96	0,94-0,99	0,002

Έφηβοι vs παιδική ηλικία

- Οι έφηβοι έχουν **0,35 φορές** (μικρότερο) κίνδυνο για την εμφάνιση παχυσαρκίας σε σχέση με τα **παιδιά**, **ανεξάρτητα από** *το φύλο τους, το ΔΜΣ των γονιών και την ηλικία της μητέρας*, με τη διαφορά αυτή να είναι **στατιστικά σημαντική**.
- Ισοδύναμα ο κίνδυνος των εφήβων για παχυσαρκία είναι **το 0.35 του αντίστοιχου κινδύνου** στα παιδιά **ανεξάρτητα από...**
- Ισοδύναμα, οι έφηβοι, σε σχέση με τα παιδιά, έχουν μειωμένο **κατά $(1-0,35=)$ 65% κίνδυνο** να εμφανίσουν παχυσαρκία **ανεξάρτητα από...**