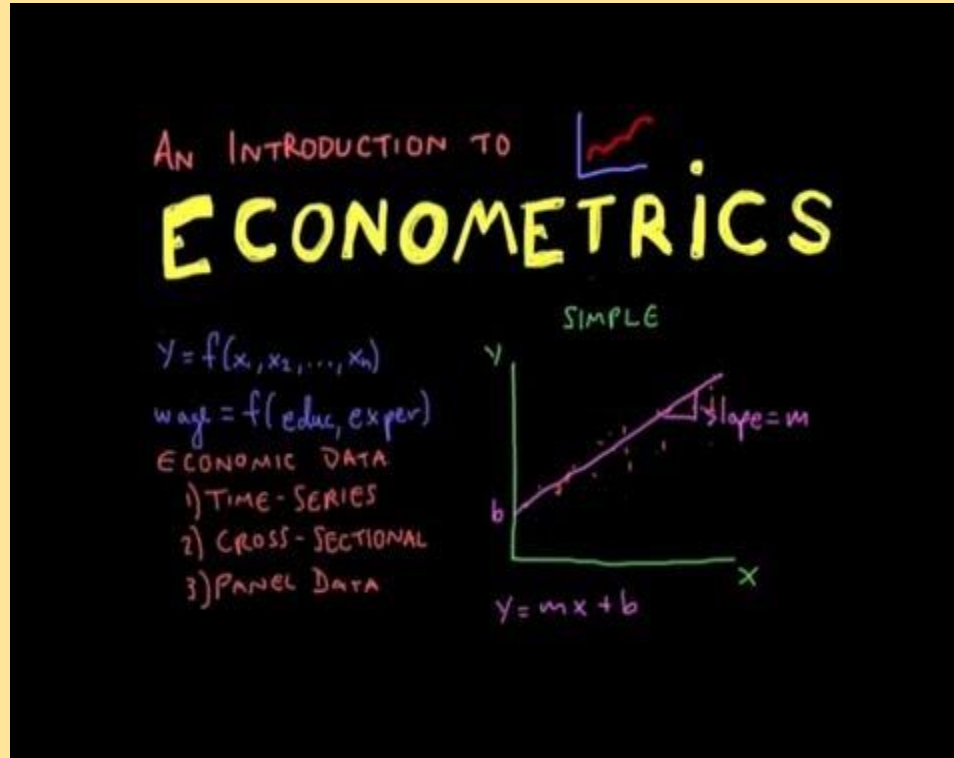


Οικονομικό Πανεπιστήμιο Αθηνών

Μάθημα: "Εφαρμογές Στατιστικών Μεθόδων σε Επιχειρησιακά Προβλήματα"

Συνάντηση: 10η



Οικονομικό Πανεπιστήμιο Αθηνών

Λεμοντζόγλου

Μάθημα: "Εφαρμογές Στατιστικών Μεθόδων σε Επιχειρησιακά Προβλήματα"

Συνάντηση: 10η

Δρ. Τρύφωνας

Προβλέψεις

πρόσβ. ηλεκτρ. αστικ. πληθ.

1	7	15
2	18	22
3	64	59
4	92	78
5	16	49
6	44	41
7	36	43
8	23	23
9	66	43
10	5	18

Έστω ένα μικρό δείγμα 10 χωρών ($n=10$) με παρατηρήσεις για την **πρόσβαση του πληθυσμού στον ηλεκτρισμό (%)** και τον **αστικό πληθυσμό (%)**. Αναμένω μια θετική σχέση ανάμεσα στις τιμές των δύο μεταβλητών. **Μπορώ να προβλέψω το επίπεδο της πρόσβασης στον ηλεκτρισμό με δεδομένο το επίπεδο του αστικού πληθυσμού (X_0) ;;;**

πρόσβ. ηλεκτρ. αστικ. πληθ.

1	7	15
2	18	22
3	64	59
4	92	78
5	16	49
6	44	41
7	36	43
8	23	23
9	66	43
10	5	18

Χρησιμοποίησε το περιβάλλον της εφαρμογής →

<https://www.statskingdom.com/linear-regression-calculator.html>

**Κάνε εισαγωγή των τιμών του πίνακα στις στήλες Ψ και Χ της εφαρμογής.
Κάνε εισαγωγή των σεναρίων σου εδώ:**

X values for prediction: (You may leave empty)

Έστω Χο: 95, 25, 85, 8

Simple Linear regression

Multiple Linear regression

Logistic regression

Multinomial logistic regression

X	Y
15	7
22	18
59	64
78	92
49	16
41	44
43	36
23	23
43	66
18	5

Εισάγεις τις τιμές του
X₀ (σενάρια)

X values for prediction: (You may leave empty)

95 25 85 8

Regression line equation

$$\hat{Y} = -11.9508 + 1.2545X$$

Μια αύξηση του αστικού πληθυσμού κατά 1% θα αυξήσει την πρόσβαση στον ηλεκτρισμό κατά περίπου 1,25%

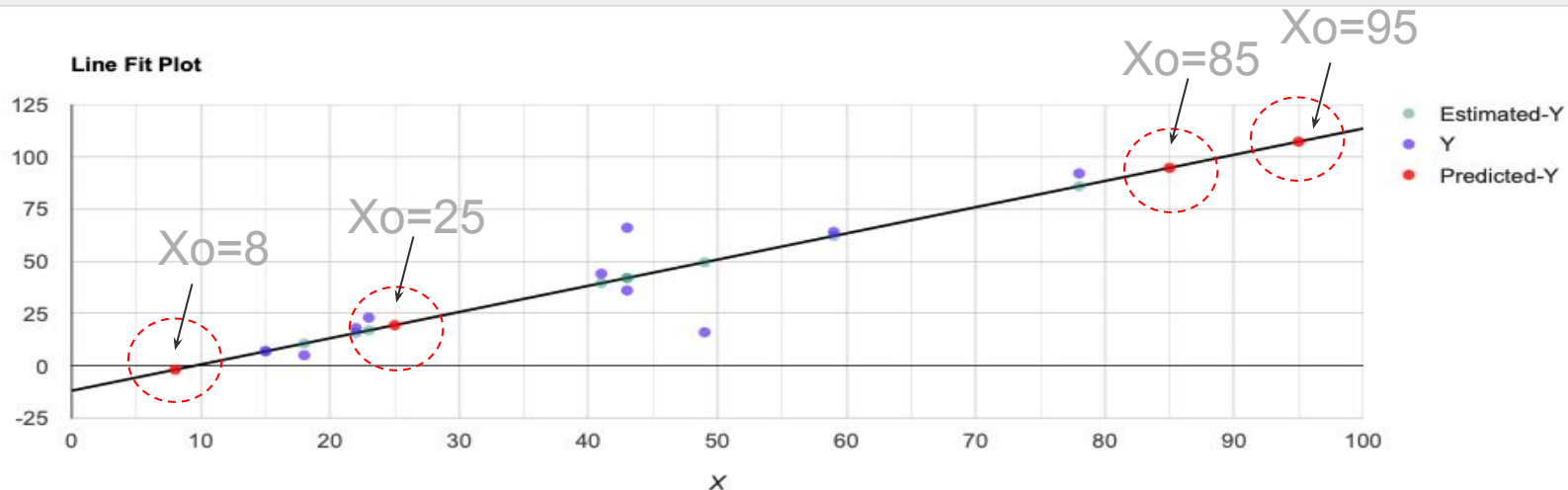
Reporting linear regression in APA style

X predicted Y, $R^2 = .75$, $F(1,8) = 24.34$, $p = .001$.

$\beta = 1.25$, $p = .001$, $\alpha = -11.95$, $p = .311$.

Η εφαρμογή δίνει σε κόκκινες κουκίδες τις προβλέψεις για κάθε τιμή του X_0 που εισάγεις

Line Fit Plot



Για κάθε τιμή του X_0 που εισάγεις η εφαρμογή σου παραθέτει και το διάστημα εμπιστοσύνης της πρόβλεψης !!!

τιμές του X_0

προβλέψεις

διάστημα

εμπιστοσύνης

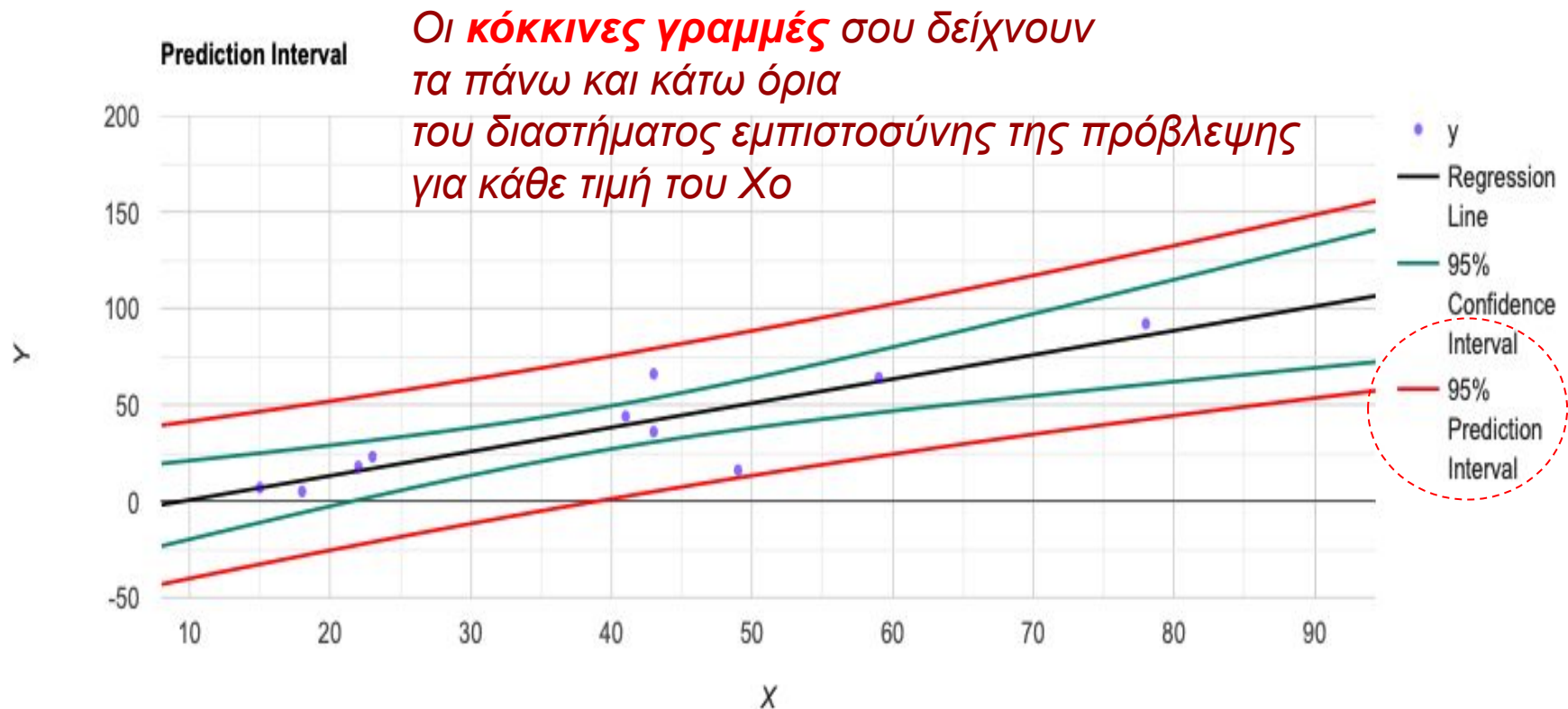
πρόβλεψης

τυπικό σφάλμα

πρόβλεψης

Prediction

X	Y prediction	Confidence Interval	S.E Confidence Int	Prediction Interval	S.E Prediction Int
95	107.2264	[72.6052, 141.8475]	15.0135	[57.8028, 156.65]	21.4326
25	19.4116	[5.5281, 33.2951]	6.0206	[-18.4939, 57.317]	16.4377
85	94.6814	[65.5495, 123.8133]	12.6331	[48.935, 140.4279]	19.838
8	-1.9149	[-23.2902, 19.4605]	9.2694	[-43.1578, 39.328]	17.885



Πως υπολογίζω ένα διάστημα εμπιστοσύνης πρόβλεψης με το χέρι ;;;

τιμή της t από
τους πίνακες

τιμή της t από
τους πίνακες

$$P(\widehat{\Psi}_0 - t_{\alpha/2, n-k-1} * s.e(\widehat{\Psi}_0) < \Psi_0 < \widehat{\Psi}_0 + t_{\alpha/2, n-k-1} * s.e(\widehat{\Psi}_0)) = 95\%$$

πρόβλεψη

τυπικό σφάλμα
πρόβλεψης

πρόβλεψη

τυπικό σφάλμα
πρόβλεψης

$$s.e(\widehat{\Psi}_0) = s * \sqrt{1 + \frac{1}{n} + \frac{(X_o - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

$$s = \sqrt{\frac{\sum_{i=1}^n (\Psi_i - \widehat{\Psi}_i)^2}{n-k-1}}$$

τυπικό σφάλμα
παλινδρόμησης

Λογαριθμικό Μοντέλο (log)

	year	cons	wag
1	2026	148	69
2	2025	145	66
3	2024	143	63
4	2023	141	60
5	2022	138	56
6	2021	127	53
7	2020	121	49
8	2019	129	52
9	2018	126	50
10	2017	123	48
11	2016	120	48
12	2015	119	47

Παίρνουμε τιμές για την ελληνική οικονομία από τη βάση δεδομένων της ευρωπαϊκής επιτροπής (**n=12**).

Όπου

“**cons**”: τελικές δαπάνες νοικοκυριών (σε δισ. ευρώ)

“**wag**”: ακαθάριστο εθνικό εισόδημα (σε δισ. ευρώ)

Τρέξτε το μοντέλο στην εφαρμογή !!!

Regression line equation

$$\hat{Y} = 56.9275 + 1.3568X$$

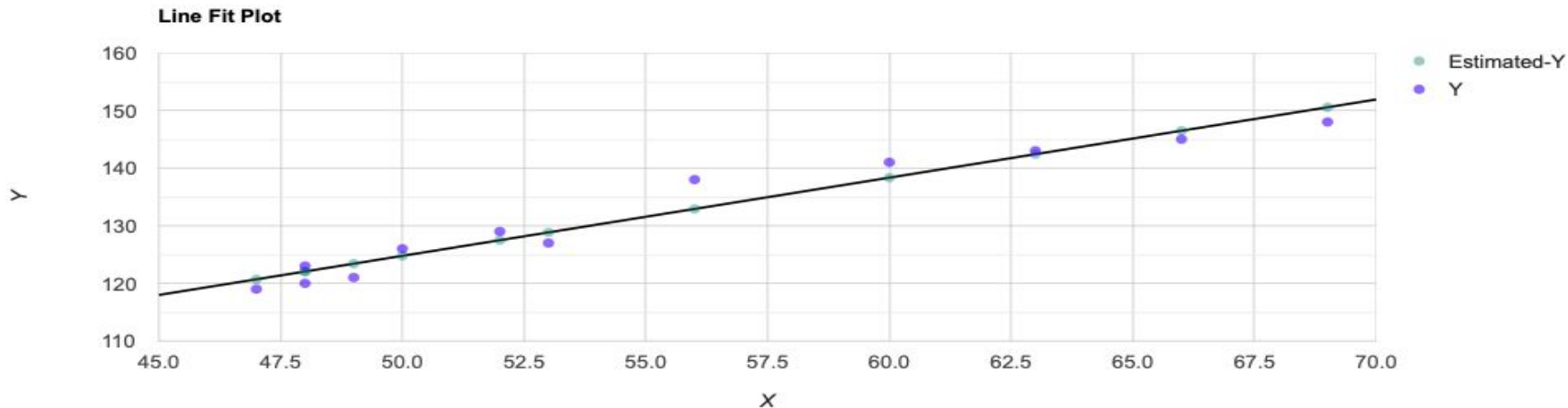
Reporting linear regression in APA style

X predicted Y , $R^2 = .95$, $F(1,10) = 187.72$, $p < .001$.

$\beta = 1.36$, $p < .001$, $\alpha = 56.93$, $p < .001$.

Μια αύξηση του ακαθάριστου εθνικού εισοδήματος κατά μια μονάδα (δηλ. κατά 1 δισ. ευρώ) θα αυξήσει τις τελικές δαπάνες των νοικοκυριών κατά περίπου 1,35 μονάδες (δηλ. κατά περίπου 1,35 δισ. ευρώ)

Line Fit Plot



	year	cons	wag		year	log_cons	log_wag
1	2026	148	69		2026	2,170261715	1,836956737
2	2025	145	66		2025	2,161368002	1,820201459
3	2024	143	63		2024	2,155336037	1,800717078
4	2023	141	60		2023	2,149219113	1,776701184
5	2022	138	56		2022	2,139879086	1,748962861
6	2021	127	53		2021	2,103803721	1,726727209
7	2020	121	49		2020	2,08278537	1,693726949
8	2019	129	52		2019	2,11058971	1,711807229
9	2018	126	50		2018	2,100370545	1,698100546
10	2017	123	48		2017	2,089905111	1,683947131
11	2016	120	48		2016	2,079181246	1,683947131
12	2015	119	47		2015	2,075546961	1,673020907

Αν εισάγεις στο μοντέλο τις λογαριθμικές τιμές τότε θα αλλάξει η ερμηνεία των συντελεστών της παλινδρόμησης !!! Δοκίμασε να τρέξει το μοντέλο με τις λογαριθμικές τιμές !!!

Simple Linear regression

Multiple Linear regression

Logistic regression

Multinomial logistic regression

log_X

1.836956737
1.820201459
1.800717078
1.776701184
1.748962861
1.726727209
1.693726949
1.711807229
1.698100546
1.683947131
1.683947131
1.673020907

log_Y

2.170261715
2.161368002
2.155336037
2.149219113
2.139879086
2.103803721
2.08278537
2.11058971
2.100370545
2.089905111
2.079181246
2.075546961

Χρήση λογαριθμικών τιμών

Regression line equation

$$\hat{Y} = 1.0945 + 0.589X$$

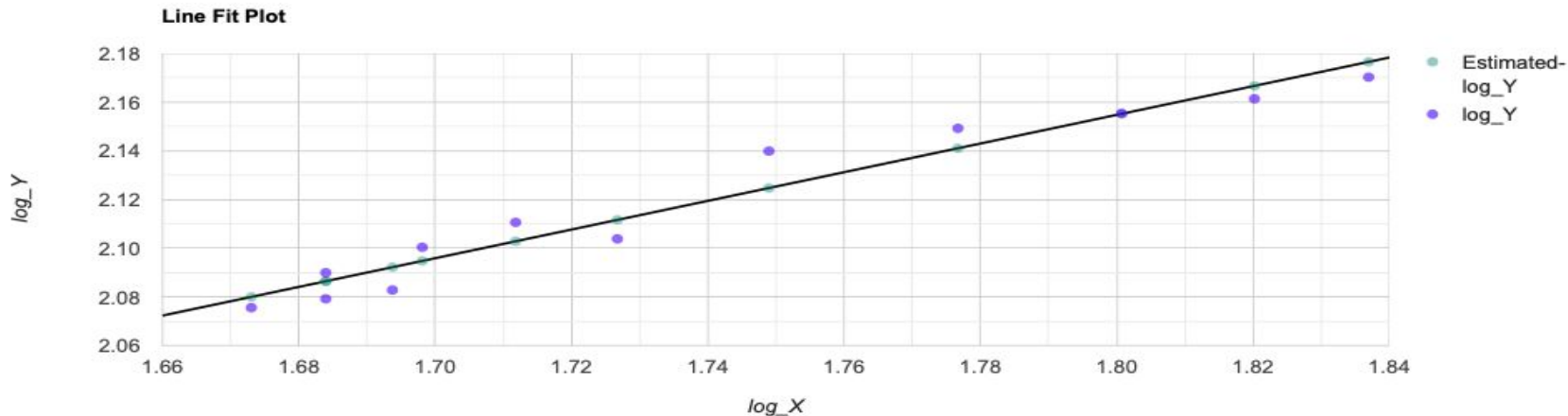
Reporting linear regression in APA style

$\log_X$ predicted $\log_Y$, $R^2 = .95$, $F(1,10) = 183.81$, $p < .001$.

$\beta = .59$, $p < .001$, $\alpha = 1.09$, $p < .001$.

Ερμηνεία του συντελεστή στο λογαριθμικό μοντέλο: **Μια αύξηση του εθνικού εισοδήματος κατά 1% θα αυξήσει την κατανάλωση των νοικοκυριών κατά περίπου 0,58% !!!**

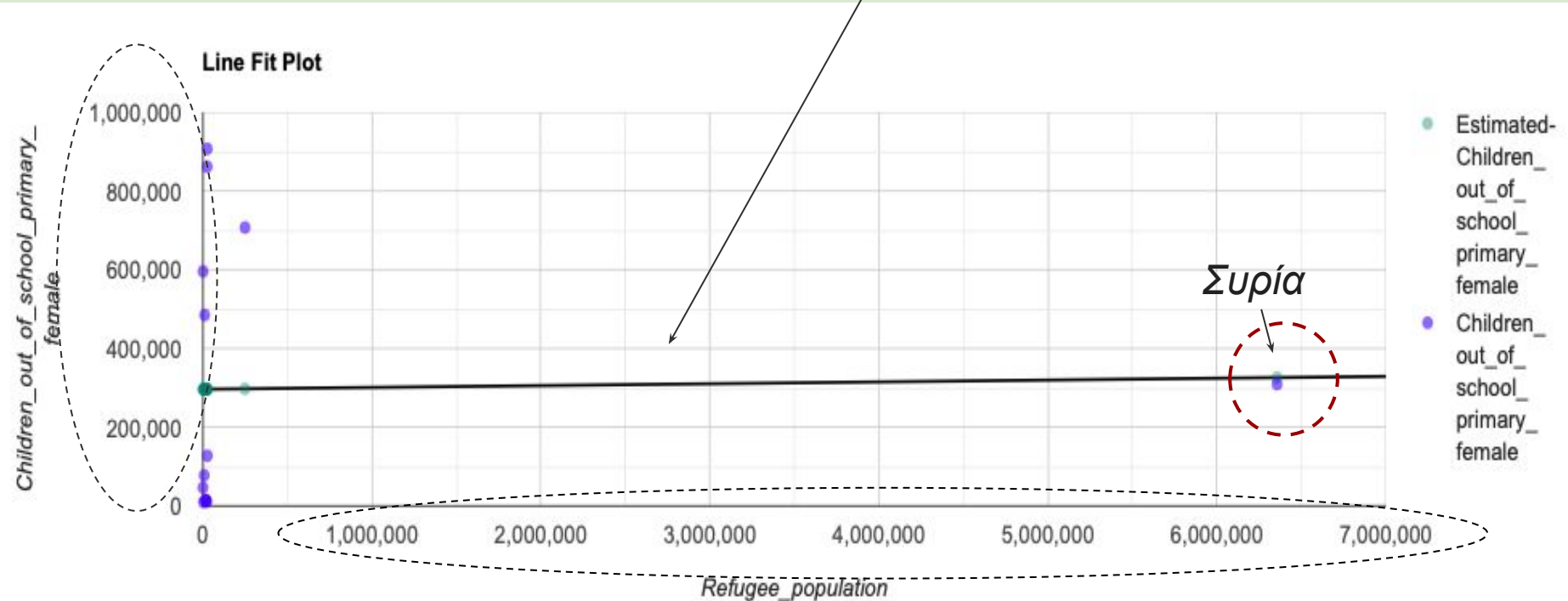
Line Fit Plot



Country Name		Children out of school, primary, female	Refugee population by country or territory of origin	log_y	log_x
Albania	2023	10813	21919	4,033946203	4,340820737
Bosnia and Herz.	2023	12522	19318	4,097673699	4,285962162
Belarus	2023	13423	15071	4,12784959	4,17814207
Cote d'Ivoire	2023	127461	25994	5,105377321	4,414873115
Cuba	2023	15377	18560	4,186871614	4,268577972
India	2023	907275	25319	5,957738944	4,403446549
Lebanon	2023	77820	7392	4,891091226	3,868761958
Mali	2023	707138	249046	5,849504176	5,396279571
Niger	2023	861789	24794	5,935400946	4,394346597
Philippines	2023	595798	645	5,775099041	2,809559715
Paraguay	2023	46641	155	4,668767853	2,190331698
Syrian Arab R.	2023	309430	6355788	5,490562417	6,803169403
Chad	2023	485022	10196	5,685761438	4,008429827
Tunisia	2023	10001	2800	4,000043427	3,447158031

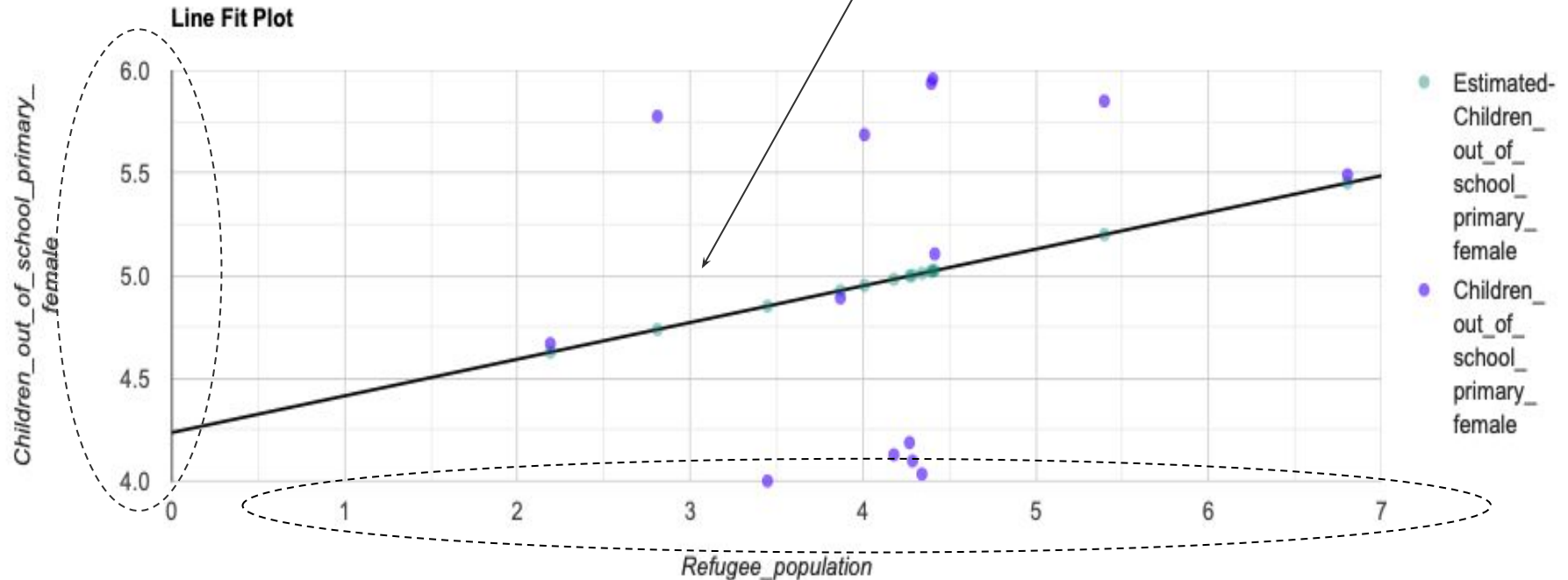
$$\hat{Y} = 296346.7272 + 0.004671X$$

Μοντέλο **ΧΩΡΙΣ** χρήση λογαρίθμων !!!



$$\hat{Y} = 4.2359 + 0.1786X$$

Μοντέλο **ME** χρήση λογαρίθμων !!!



Country Name		Urban population	Rail lines (total route-km)
United Arab Emirates	2020	8183416	261
Austria	2020	5238479	4966
Azerbaijan	2020	5692217	2239
Belarus	2020	7455467	5454
Canada	2020	31016918	48309
Switzerland	2020	6384901	3672
Chile	2020	16993267	2396
China	2020	866810508	106235
Cote d'Ivoire	2020	14951022	639
Cameroon	2020	15086797	884
Congo, Dem. Rep.	2020	43807915	3641
Czechia	2020	7922941	9377
Germany	2020	64410589	33399
Spain	2020	38276186	15848
Finland	2020	4728699	5918
France	2020	54739999	27872
Gabon	2020	2092422	648
United Kingdom	2020	56283168	16351
Georgia	2020	2213266	1411
Greece	2020	8528388	2345

Regression line equation

$$\hat{Y} = 7182.0615 + 0.0001176X$$

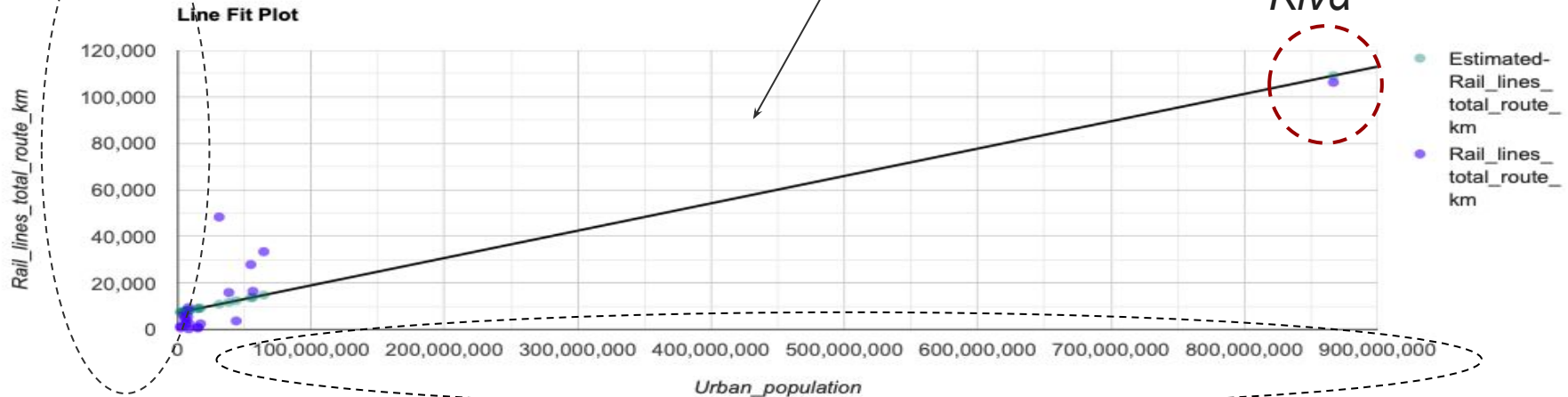
Μοντέλο **ΧΩΡΙΣ** χρήση λογαρίθμων !!!

Reporting linear regression in APA style

Urban_population predicted Rail_lines_total_route_km, $R^2 = .79$, $F(1,18) = 68.67$, $p < .001$.

$\beta = .00012$, $p < .001$, $\alpha = 7182.06$, $p = .019$.

Line Fit Plot



Country Name		Urban population	Rail lines (total route-km)		log_y	log_x
United Arab Emirates	2020	8183416	261		6,912934629	2,416640507
Austria	2020	5238479	4966		6,719205207	3,696006715
Azerbaijan	2020	5692217	2239		6,755281448	3,350054094
Belarus	2020	7455467	5454		6,872474852	3,736715134
Canada	2020	31016918	48309		7,491598642	4,684028048
Switzerland	2020	6384901	3672		6,805154168	3,564902673
Chile	2020	16993267	2396		7,230276881	3,379486814
China	2020	866810508	106235		8,937924167	5,026267622
Cote d'Ivoire	2020	14951022	639		7,174670881	2,805500858
Cameroon	2020	15086797	884		7,178597047	2,946452265
Congo, Dem. Rep.	2020	43807915	3641		7,641552584	3,561220679
Czechia	2020	7922941	9377		6,898886422	3,972063916
Germany	2020	64410589	33399		7,808957271	4,523733464
Spain	2020	38276186	15848		7,582928656	4,199974463
Finland	2020	4728699	5918		6,67474167	3,772174961
France	2020	54739999	27872		7,738304785	4,445168133
Gabon	2020	2092422	648		6,320649278	2,811575006
United Kingdom	2020	56283168	16351		7,750378535	4,213544319
Georgia	2020	2213266	1411		6,345033612	3,149527014
Greece	2020	8528388	2345		6,93086695	3,370142847

Regression line equation

$$\hat{Y} = 4.8558 + 0.6337X$$

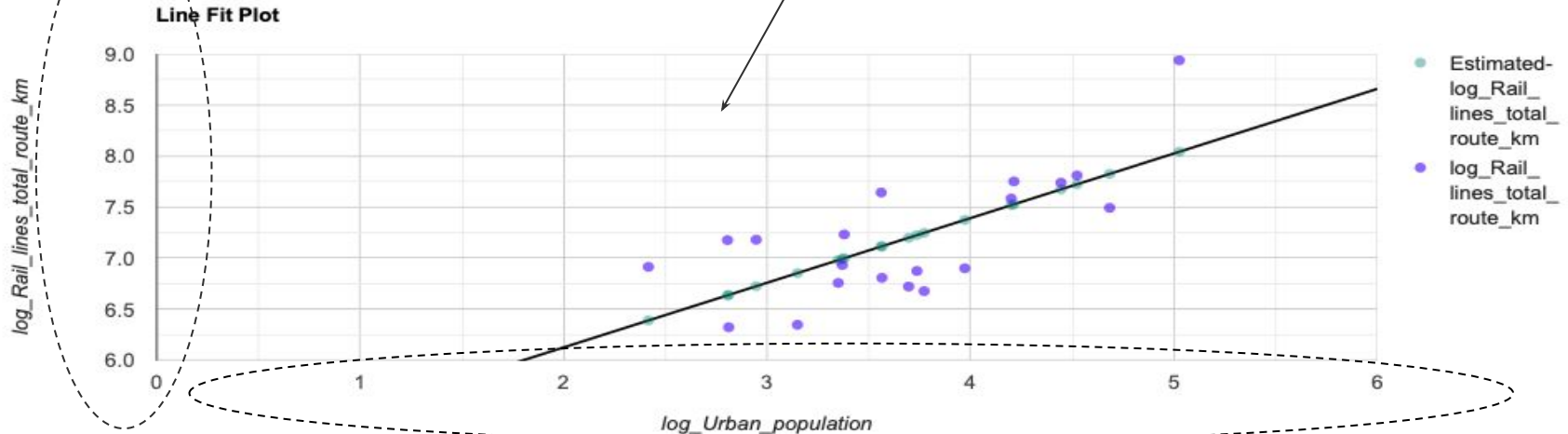
Μοντέλο **ME** χρήση λογαρίθμων !!!

Reporting linear regression in APA style

log_Urban_population predicted log_Rail_lines_total_route_km, $R^2 = .51$, $F(1,18) = 18.63$, $p < .001$.

$\beta = .63$, $p < .001$, $\alpha = 4.86$, $p < .001$.

Line Fit Plot



Ο ρόλος των ψευδομεταβλητών (dummies)

Υπάρχουν μισθολογικές διαφορές
στη βάση του φύλου ; ; ;

ψευδομεταβλητή
φύλου

Δίνουμε την
τιμή 1 κάθε
φορά που
συναντάμε
γυναίκα. Σε
κάθε άλλη
περίπτωση
δίνουμε την
τιμή 0.

i	Άτομα	Μηνιαίος Μισθός	Έτη Σπουδών	Βαθμός Πτυχίου	Φύλο
1	Maria L.	800	7	7	1
2	Makis E.	580	4	6	0
3	Giorgos M.	1.500	12	9,5	0
4	Marios K.	430	3	6,5	0
5	Ioannis P.	760	3	7,2	0
6	Georgia P.	580	4	6	1
7	Ilias K.	2.300	9	8,9	0
8	Maria O.	780	6	5,8	1
9	Giota L.	650	4	6,2	1
10	Nancy T.	950	8	8,2	1

**Υπάρχουν μισθολογικές διαφορές
στη βάση του φύλου ; ; ;**

$$\widehat{\text{μισθός}}_i = \hat{\alpha} + \hat{\beta}1 * \text{βαθμός}_i + \hat{\beta}2 * \text{έτη}_i + \hat{\beta}3 * \text{φύλο}_i$$

εξαρτημένη
μεταβλητή

ανεξάρτητη
μεταβλητή

ανεξάρτητη
μεταβλητή

**ψευδομεταβλητή
φύλου**

Ψευδομεταβλητή φύλου: παίρνει την τιμή 1 αν συναντάμε γυναίκα. Σε κάθε άλλη περίπτωση λαμβάνει την τιμή 0.

**Υπάρχουν μισθολογικές διαφορές
στη βάση του φύλου ; ; ;**

$$\widehat{\text{μισθός}}_i = \hat{\alpha} + \hat{\beta}1 * \text{βαθμός}_i + \hat{\beta}2 * \text{έτη}_i + \hat{\beta}3 * \text{φύλο}_i$$

εξαρτημένη
μεταβλητή

ανεξάρτητη
μεταβλητή

ανεξάρτητη
μεταβλητή

**ψευδομεταβλητή
φύλου**

Αν το $\hat{\beta}3$ είναι μη στατιστικά σημαντικό

τότε δεν υπάρχει καμία έμφυλη μισθολογική διάκριση ...

**Υπάρχουν μισθολογικές διαφορές
στη βάση του φύλου ; ; ;**

$$\widehat{\text{μισθός}}_i = \hat{\alpha} + \hat{\beta}1 * \text{βαθμός}_i + \hat{\beta}2 * \text{έτη}_i + \hat{\beta}3 * \text{φύλο}_i$$

εξαρτημένη
μεταβλητή

ανεξάρτητη
μεταβλητή

ανεξάρτητη
μεταβλητή

**ψευδομεταβλητή
φύλου**

Αν το $\hat{\beta}3$ είναι **στατιστικά σημαντικό** και έχει **αρνητικό πρόσημο**

τότε υπάρχει έμφυλη μισθολογική διάκριση (εις βάρος των γυναικών) ...

Country Name	Indicator Name	2010	2023	nordic	western	balkan_east	non_european
Albania	GDP growth (annual %)	4	4	0	0	1	0
Australia	GDP growth (annual %)	2	3	0	0	0	1
Austria	GDP growth (annual %)	2	-1	0	1	0	0
Bolivia	GDP growth (annual %)	4	3	0	0	0	1
Brazil	GDP growth (annual %)	8	3	0	0	0	1
Chile	GDP growth (annual %)	6	0	0	0	0	1
China	GDP growth (annual %)	11	5	0	0	0	1
Cyprus	GDP growth (annual %)	2	3	0	0	1	0
Czechia	GDP growth (annual %)	3	-0	0	0	1	0
Germany	GDP growth (annual %)	4	-0	0	1	0	0
Denmark	GDP growth (annual %)	2	2	1	0	0	0
Spain	GDP growth (annual %)	0	3	0	0	0	0
France	GDP growth (annual %)	2	1	0	1	0	0
Greece	GDP growth (annual %)	-6	2	0	0	1	0
Ireland	GDP growth (annual %)	2	-6	0	1	0	0
North Macedonia	GDP growth (annual %)	3	2	0	0	1	0
Montenegro	GDP growth (annual %)	3	6	0	0	1	0
Netherlands	GDP growth (annual %)	1	0	0	1	0	0
Norway	GDP growth (annual %)	1	0	1	0	0	0
Philippines	GDP growth (annual %)	7	6	0	0	0	1
Portugal	GDP growth (annual %)	2	3	0	0	0	0
Romania	GDP growth (annual %)	-4	2	0	0	1	0
Serbia	GDP growth (annual %)	2	4	0	0	1	0
Sweden	GDP growth (annual %)	6	-0	1	0	0	0

Το 2010 στα Βαλκάνια έχουμε
Ο ρόλος της γεωγραφίας !!!
ρυθμούς ανάπτυξης **μειωμένους**
κατά 2,87%

	Σκανδιναβία	δυτική ευρώπη	βαλκάνια και ανατολική ευρώπη	εκτός ευρώπη
ρυθμοί ανάπτυξης (2010)	+0,2381	-0,74	-2,87 *	+4,72 ***
ρυθμοί ανάπτυξης (2023)	-1,38	-3,88 ***	+1,5	+1,24

Το 2023 στις χώρες της δυτικής ευρώπης έχουμε ρυθμούς ανάπτυξης **μειωμένους** κατά 3,88%

Το 2023 στις χώρες εκτός ευρώπης έχουμε ρυθμούς ανάπτυξης **αυξημένους** κατά 4,72%

Η επίδραση μιας νομοθετικής διάταξης (βλ. πολιτικές) !!!

Εξαρτημένη μεταβλητή: *γεννήσεις παιδιών εκτός γάμου (%)*

Χρονική ψευδομεταβλητή: δίνουμε *την τιμή 0* πριν την αναθεώρηση του οικογενειακού δικαίου το 1982. Για κάθε έτος μετά το 1981 δίνουμε *την τιμή 1*.

Δες πίσω τις τιμές ...

	1960	1,2	0	1982	1,5	1	2004	5,1	1	
	1961	1,2	0	1983	1,6	1	2005	5,1	1	
	1962	1,2	0	1984	1,7	1	2006	5,3	1	
	1963	1,2	0	1985	1,8	1	2007	5,8	1	
	1964	1,1	0	1986	1,8	1	2008	5,9	1	
	1965	1,1	0	1987	2,1	1	2009	6,6	1	
	1966	1	0	1988	2,1	1	2010	7,3	1	
	1967	1	0	1989	2,1	1	2011	7,4	1	
	1968	1,1	0	1990	2,2	1	2012	7,6	1	
	1969	1,1	0	1991	2,4	1	2013	7	1	
	1970	1,1	0	1992	2,6	1	2014	8,2	1	
	1971	1,2	0	1993	2,8	1	2015	8,8	1	
	1972	1,2	0	1994	2,9	1	2016	9,4	1	
	1973	1,3	0	1995	3	1	2017	10,3	1	
	1974	1,2	0	1996	3,3	1	2018	11,1	1	
	1975	1,3	0	1997	3,5	1	2019	12,4	1	
	1976	1,3	0	1998	3,8	1	2020	13,8	1	
	1977	1,3	0	1999	3,9	1				
	1978	1,4	0	2000	4	1				
	1979	1,4	0	2001	4,3	1				
	1980	1,5	0	2002	4,4	1				
	1981	1,6	0	2003	4,8	1				

Η επίδραση μιας νομοθετικής διάταξης (βλ. πολιτικές) !!!

Regression line equation

$$\hat{Y} = 1.2273 + 3.8932X$$

Reporting linear regression in APA style

1982_dummy predicted BoM, $R^2 = .35$, $F(1,59) = 32.16$, $p < .001$.

$\beta = 3.89$, $p < .001$, $\alpha = 1.23$, $p = .029$.

Παρατηρούμε πως η **χρονική ψευδομεταβλητή** που εκφράζει την πιθανή **επίδραση της μεταρρύθμισης του οικογενειακού δικαίου το έτος 1982** είναι **στατιστικά σημαντική** και έχει **θετικό πρόσημο**.

Μετά το 1981 παρατηρούμε **αυξημένες γεννήσεις εκτός γάμου κατά περίπου 3,89%**