



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών

Μεθοδολογία των Επιστημών του Ανθρώπου: Στατιστική

Ενότητα 2: Επαγωγική Στατιστική

Βασίλης Γιαλαμάς
Σχολή Επιστημών της Αγωγής
Τμήμα Εκπαίδευσης και Αγωγής στην Προσχολική
Ηλικία

Περιεχόμενα ενότητας

Παρουσιάζονται οι βασικές έννοιες του ελέγχου υποθέσεων που αφορούν στη σύγκριση μέσων τιμών και τη συνάφεια μεταξύ δυο μεταβλητών. Γίνεται επίσης εισαγωγή στην εκτιμητική της μέσης τιμής με τη βοήθεια των διαστημάτων εμπιστοσύνης.



Πού χρησιμοποιείται ο συντελεστής συσχέτισης (r) pearson (1 από 3)

- **Εγκυρότητα**

Η συσχέτιση ανάμεσα σε δυο διαφορετικές κλίμακες στάσεων απέναντι στα Μαθηματικά πρέπει να είναι θετική και ισχυρή αφού και οι δύο κλίμακες αναφέρονται στην ίδια έννοια.

- **Αξιοπιστία**

Μια από τις απαραίτητες μορφές της αξιοπιστίας ενός εργαλείου μέτρησης αναφέρεται στην Προ-τεστ μετά-τεστ αξιοπιστία.

Η αξιοπιστία αυτή ελέγχεται, στην ίδια ομάδα περιπτώσεων, με τον υπολογισμό του συντελεστή συσχέτισης των δύο μετρήσεων με την ίδια κλίμακα (εργαλείο) σε δύο διαφορετικές χρονικές στιγμές (π.χ. με διαφορά 6 μηνών). Η τιμή του συντελεστή οφείλει να είναι θετική και ισχυρή.



Πού χρησιμοποιείται ο συντελεστής συσχέτισης (r) pearson (2 από 3)

- **Επαλήθευση της θεωρίας**

Οι ερευνητικές υποθέσεις που αναφέρονται στη σχέση δυο ποσοτικών μεταβλητών εξετάζονται με τη βοήθεια του στατιστικού ελέγχου υποθέσεων για τον συντελεστή συσχέτισης ($H_0: \rho = 0, H_A: \rho \neq 0$), όπως είδαμε στα παραδείγματα των διαφανειών 16 και 17.

Για παράδειγμα η συσχέτιση μεταξύ «Άγχους» και «Γονικού Ελέγχου Συμπεριφοράς» επαληθεύεται στους εφήβους αφού βρέθηκε στατιστικά σημαντική ($r = -0,4$, $p < 0.001$, $N = 240$). Όσο υψηλότερο το επίπεδο του ελέγχου από τον γονέα τόσο χαμηλότερο το επίπεδο άγχους του εφήβου.



Πού χρησιμοποιείται ο συντελεστής συσχέτισης (r) pearson (3 από 3)

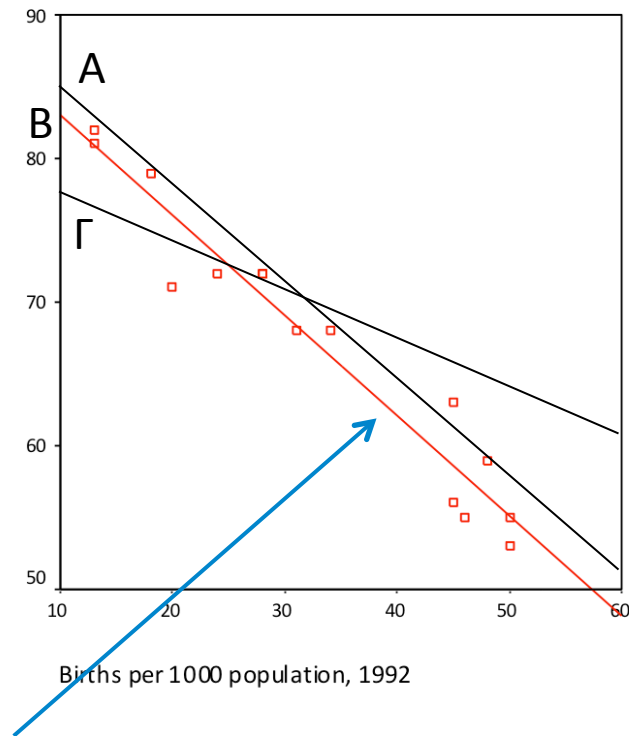
- **Πρόβλεψη**

Όπως θα δούμε παρακάτω μπορούμε να προβλέψουμε την τιμή μιας μεταβλητής με την τιμή μιας άλλης όταν διαπιστωθεί γραμμική και στατιστικά σημαντική συσχέτιση μεταξύ δύο ποσοτικών μεταβλητών. πχ. Μπορεί να προβλέψουμε την επίδοση ενός μαθητή στην πρώτη δημοτικού από τον βαθμό αναγνωστικής ετοιμότητας που παρουσίασε (μέτρηση με κατάλληλη κλίμακας) στο νηπιαγωγείο.

Αυτό γίνεται με τη βοήθεια της εξίσωσης παλινδρόμησης που μπορεί να υπολογιστεί από τα δεδομένα ενός δείγματος μαθητών για τους οποίους υπολογίστηκε η αναγνωστική ετοιμότητα στο νηπιαγωγείο και η επίδοσή τους στο τέλος της πρώτης δημοτικού.



Τρεις πιθανές γραμμές



Η γραμμή B προσαρμόζεται καλύτερα στα δεδομένα.

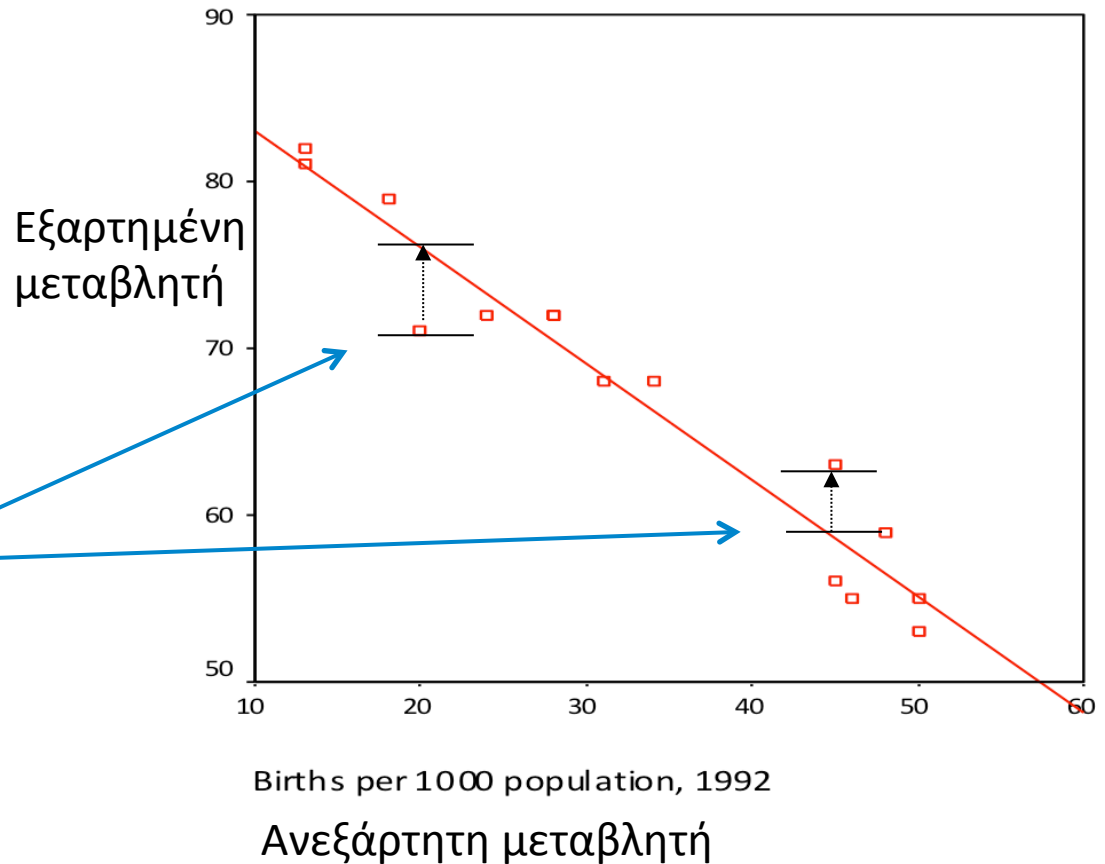


Γραμμή παλινδρόμησης με τη μέθοδο των ελαχίστων τετραγώνων

Η γραμμή παλινδρόμησης είναι αυτή για την οποία το άθροισμα των τετραγώνων των αποκλίσεων ($\sum d^2$) γίνεται ελάχιστο.

«Απόκλιση σημείου»

d = Απόσταση μεταξύ σημείου και γραμμής παλινδρόμησης



Η εξίσωση της (ευθείας) γραμμής παλινδρόμησης (1 από 2)

$$Y = a + bX$$

Εξαρτημένη μεταβλητή σταθερά κλίση Ανεξάρτητη μεταβλητή

Προσδόκιμο επιβίωσης = 90 – (0.70 x αριθμός γεννήσεων)

Η κλίση είναι -0.70 Όταν αριθμός γεννήσεων αυξάνει κατά ένα παιδί (στους 1000 κατοίκους) τότε το προσδόκιμο επιβίωσης ελαττώνεται κατά 0.7 έτη.

Μπορούμε να προβλέψουμε για μια χώρα που δεν ανήκει στο δείγμα που χρησιμοποιήθηκε για τον υπολογισμό του a και b το προσδόκιμο επιβίωσης των γυναικών αν γνωρίζουμε τον αριθμό των γεννήσεων.

Αν ο ο αριθ. γεννήσεων είναι 10 τότε το προσδόκιμο επιβίωσης προβλέπεται $90 - 0,7 \times 10 = 90 - 7 = 83$ χρόνια.



Η εξίσωση της (ευθείας) γραμμής παλινδρόμησης (2 από 2)

Η συντελεστής της εξίσωσης παλινδρόμησης που προκύπτουν με τη μέθοδο των ελαχίστων τετραγώνων υπολογίζονται με του παρακάτω τύπους:

$$b = \sum_{i=1}^N \frac{(x_i - \bar{x})(y_i - \bar{y})}{(N-1)s_x^2} \quad a = \bar{y} - b\bar{x}$$

Ο υπολογισμός του b είναι ευκολότερος με τον ισοδύναμο τύπο.

$$b = \frac{(\sum_{i=1}^N x_i y_i) - N \cdot \bar{x} \cdot \bar{y}}{(N-1)s_x^2}$$



Υπολογισμός συντελεστών γραμμής παλινδρόμησης: «Προσδόκιμο επιβίωσης Γυναικών» και «Αριθμός γεννήσεων»

$$\bar{X} = \frac{493}{15} = 32,87 \quad \bar{Y} = \frac{1006}{15} = 67,07$$

$$S_x^2 = \frac{\sum_{i=1}^N X_i^2 - \left(\sum_{i=1}^N X_i\right)^2 / N}{N - 1} =$$

$$\frac{18793 - 493^2 / 15}{14} = 184,98$$

$$b = \frac{\left(\sum_{i=1}^N x_i y_i\right) - N \cdot \bar{x} \cdot \bar{y}}{(N - 1) s_x^2} =$$

$$= \frac{31258 - 15 \cdot 67,07 \cdot 32,87}{14 \cdot 184,98} = -0,697$$

$$a = \bar{y} - b\bar{x} =$$

$$= 67,07 - (-0,697) \cdot 32,87$$

$$= 89,932$$

Y	X	Y ²	X ²	XY
68	31	4624	961	2108
53	50	2809	2500	2650
79	18	6241	324	1422
72	28	5184	784	2016
82	13	6724	169	1066
68	34	4624	1156	2312
63	45	3969	2025	2835
81	13	6561	169	1053
72	24	5184	576	1728
55	46	3025	2116	2530
55	50	3025	2500	2750
71	20	5041	400	1420
72	28	5184	784	2016
56	45	3136	2025	2520
59	48	3481	2304	2832
1006	493	68812	18793	31258



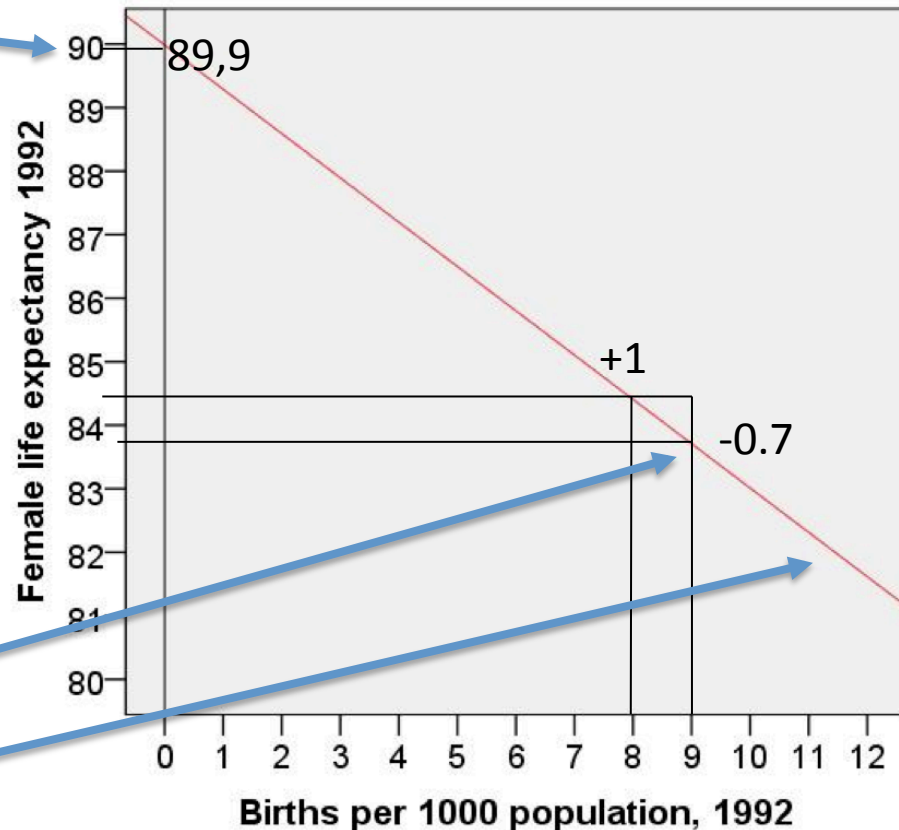
Γραφική απεικόνιση των συντελεστών της γραμμής παλινδρόμησης

Σταθερά

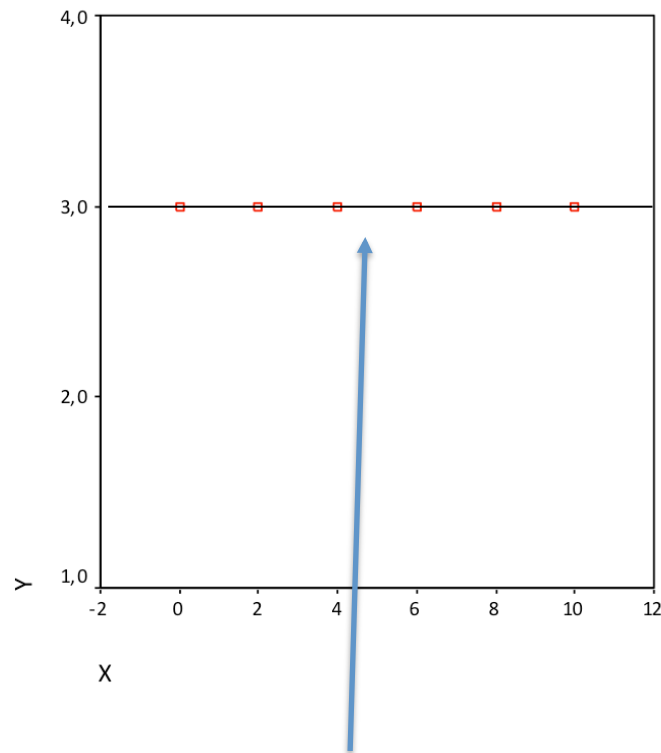
Η τιμή του προσδόκιμου επιβίωσης όταν ο αριθμός των γεννήσεων είναι 0 δηλ. το σημείο στο οποίο η γραμμή παλινδρόμησης τέμνει τον άξονα (y) του προσδόκιμου επιβίωσης.

Καθώς ο αριθμός γεννήσεων αυξάνει κατά 1, η επιβίωση ελαττώνεται κατά 0.7 έτη.

κλίση



Μη γραμμική σχέση (κλίση $b=0$)



Αλλαγές της x δεν έχουν επίδραση στη μεταβλητή y



Προβλήματα ερμηνείας του συντελεστή συσχέτισης (1 από 5)

Το συμπέρασμα ότι υπάρχει μια σχέση αίτιου - αποτελέσματος ανάμεσα στις δυο μεταβλητές πρέπει να εξάγεται μετά από τεκμηρίωση.

π.χ. η σχέση «αριθμού γεννήσεων» και «ποσοστού αστικοποίησης» μιας χώρας είναι πιθανό να οφείλεται στην ισχυρή συσχέτιση και των δυο με το επίπεδο ανάπτυξης της χώρας ή/και το επίπεδο του εθνικού συστήματος υγείας.

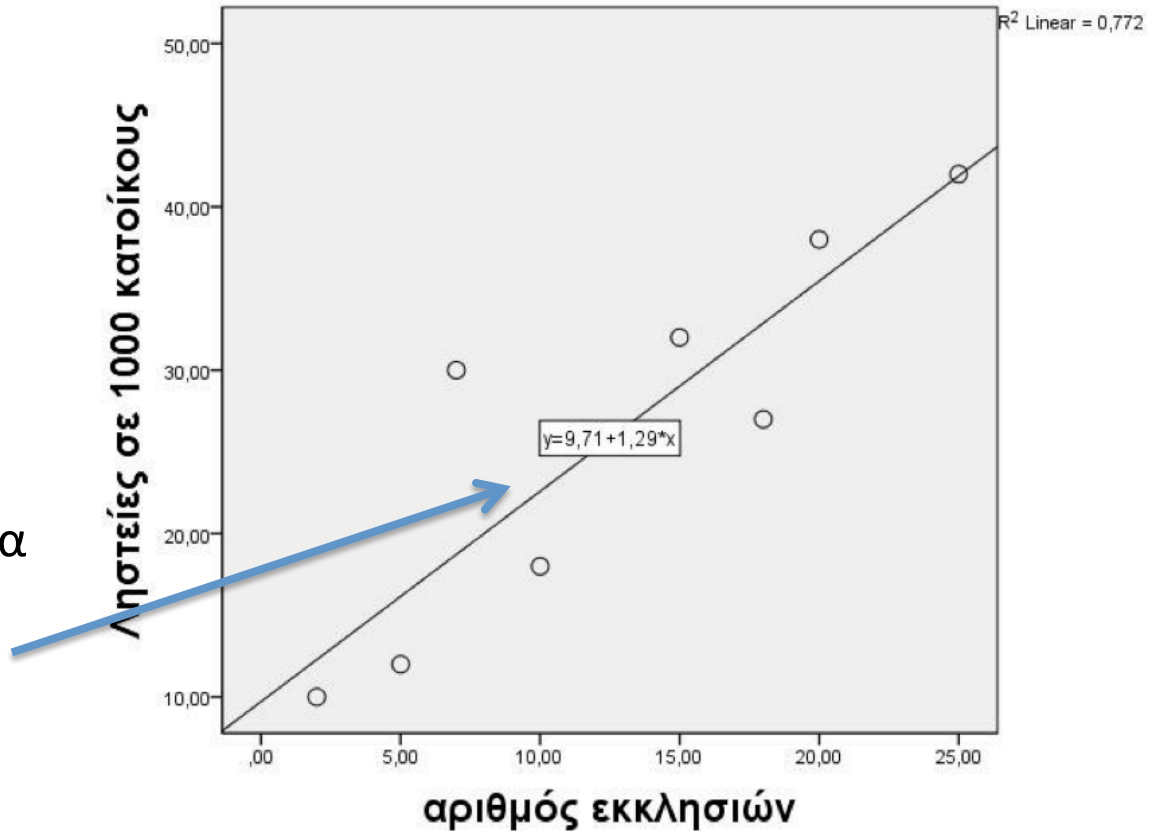


Προβλήματα ερμηνείας του συντελεστή συσχέτισης (2 από 5)

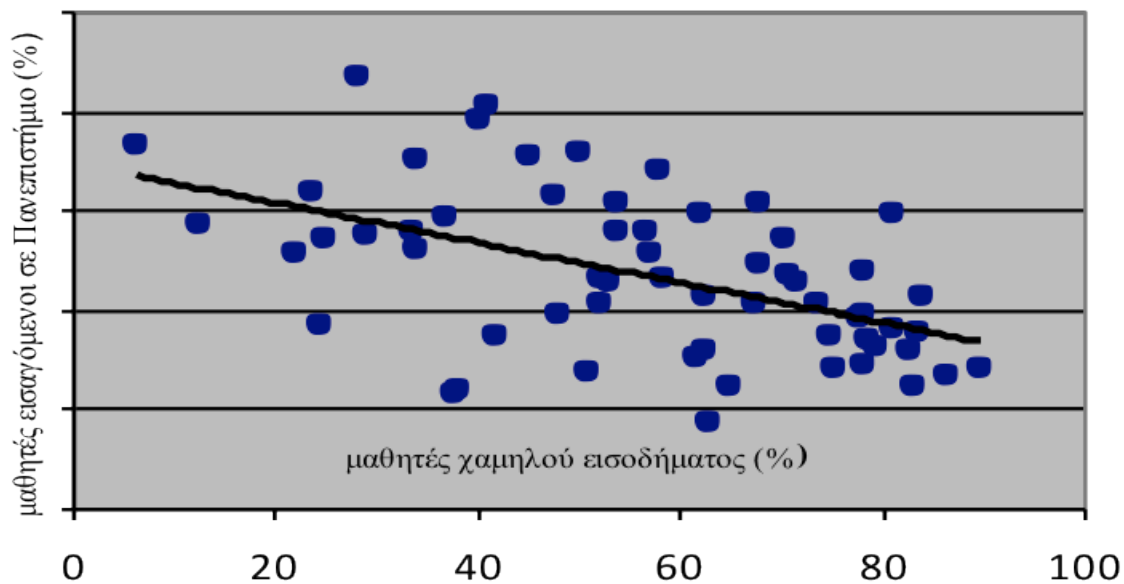
Στο διάγραμμα απεικονίζονται πόλεις.

Μπορεί να υποστηριχτεί ότι υφίσταται αιτιατή συσχέτιση θρησκευτικότητας και εγκληματικότητας μιας πόλης;

Η συσχέτισή τους είναι μια **απλή συμμεταβολή** λόγω της ισχυρής τους θετικής συσχέτισης με ένα τρίτο παράγοντα το **μέγεθος της πόλης**.



Προβλήματα ερμηνείας του συντελεστή συσχέτισης (3 από 5)

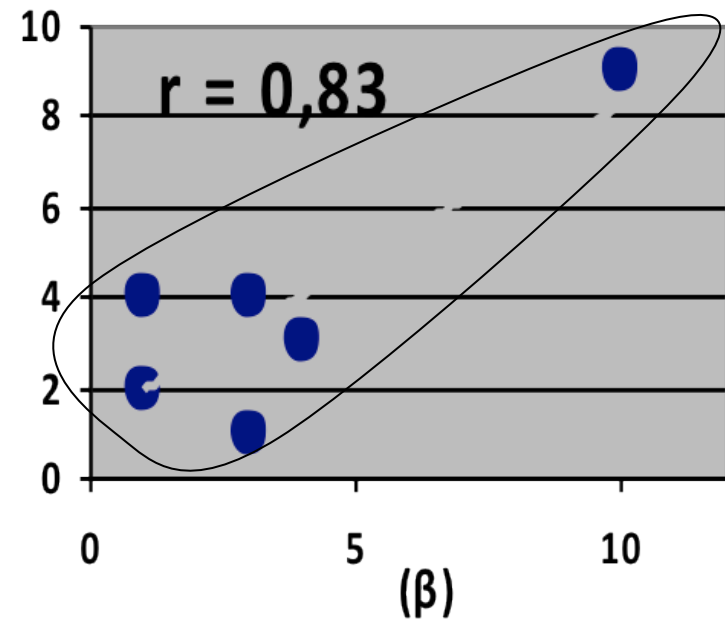
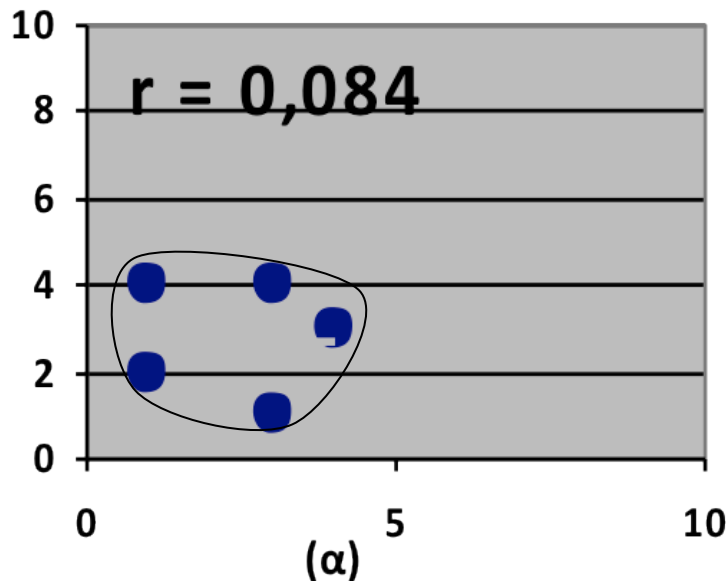


Η τιμή του συντελεστή συσχέτισης μπορεί να αλλοιώνεται από το εύρος των τιμών.



Προβλήματα ερμηνείας του συντελεστή συσχέτισης (4 από 5)

Η ύπαρξη ακραίων ατόμων-σημείων μπορεί να επιδράσει έντονα στη τιμή της συσχέτισης.

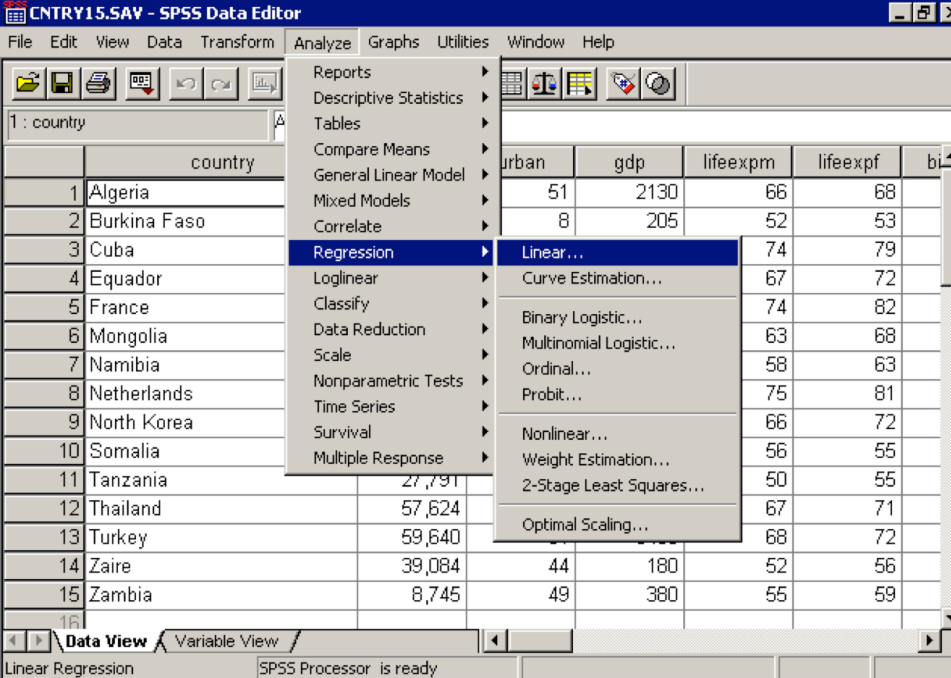


Προβλήματα ερμηνείας του συντελεστή συσχέτισης (5 από 5)

Η συσχέτιση **δεν** πρέπει να ερμηνεύεται ως **αναλογία**. Χρησιμοποιείται η τιμή r^2 (συντελεστής προσδιορισμού) που εκφράζει το ποσοστό της μεταβλητότητας της μιας μεταβλητής που ερμηνεύεται από τη σχέση της με την άλλη μεταβλητή.



Υπολογισμός της γραμμής παλινδρόμησης (1 από 2)

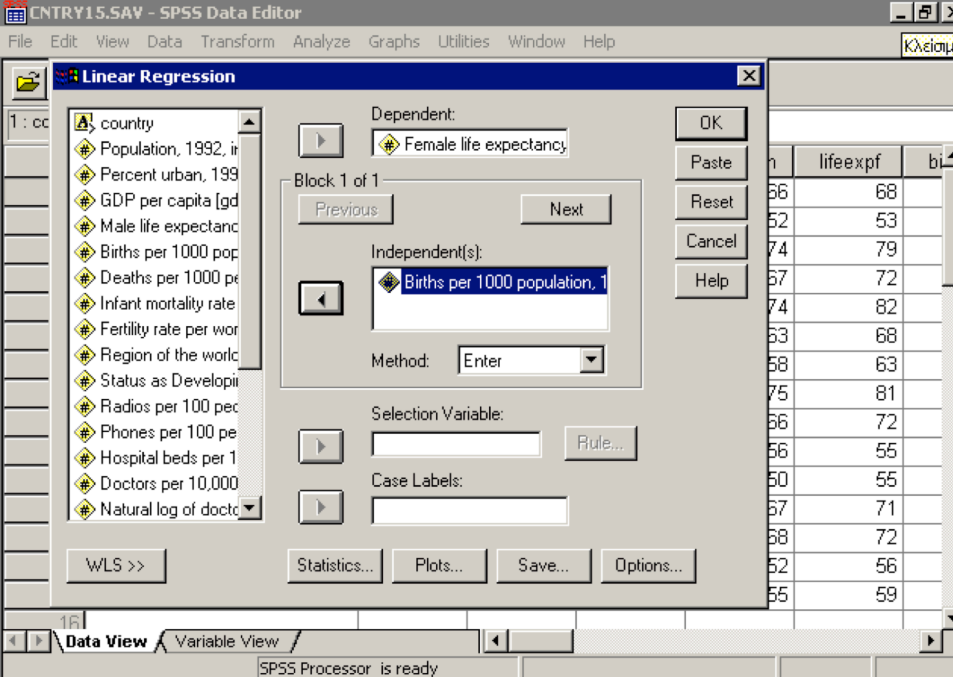


The screenshot shows the SPSS Data Editor window for a file named 'COUNTRY15.SAV'. The 'Analyze' menu is open, and the 'Regression' option is selected. The 'Linear...' option is highlighted. The data table below shows the following variables and values:

country	urban	gdp	lifeexpm	lifeexpf	bi
1 Algeria	51	2130	66	68	
2 Burkina Faso	8	205	52	53	
3 Cuba			74	79	
4 Equador			67	72	
5 France			74	82	
6 Mongolia			63	68	
7 Namibia			58	63	
8 Netherlands			75	81	
9 North Korea			66	72	
10 Somalia			56	55	
11 Tanzania	27,791		50	55	
12 Thailand	57,624		67	71	
13 Turkey	59,640		68	72	
14 Zaire	39,084	44	180	52	56
15 Zambia	8,745	49	380	55	59



Υπολογισμός της γραμμής παλινδρόμησης (2 από 2)



The screenshot shows the SPSS Data Editor window with the 'Linear Regression' dialog box open. The dependent variable is 'Female life expectancy' and the independent variable is 'Births per 1000 population, 1'. The method is set to 'Enter'. The dialog box includes buttons for 'OK', 'Paste', 'Reset', 'Cancel', 'Help', 'WLS >>', 'Statistics...', 'Plots...', 'Save...', and 'Options...'. The background shows a data table with columns 'lifeexpf' and 'births'.

lifeexpf	births
66	68
52	53
74	79
67	72
74	82
63	68
58	63
75	81
66	72
56	55
50	55
67	71
68	72
52	56
55	59



Γραμμική παλινδρόμηση του προσδόκιμου επιβίωσης γυναικών και του αριθμού γεννήσεων

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	89,985	1,765		50,995	,000
	Births per 1000 population, 1992	-,697	,050	-,968	-13,988	,000

a. Dependent Variable: Female life expectancy 1992

Κλίση

Σταθερά

Προσδόκιμο επιβίωσης γυναικών = $89.99 - (0.697 \times \text{αρ. γεννήσεων})$

Ασκήσεις επανάληψης: Στάσεις μαθητών απέναντι στις φυσικές επιστήμες (1 από 5)

1. Ποιος από τους παρακάτω ελέγχους είναι καταλληλότερος για την διερεύνηση της σχέσης μεταξύ φύλου και κλιμάκων σχετικών με τις στάσεις των μαθητών απέναντι στις φυσικές επιστήμες; (Οι τιμές κάθε κλίμακας στάσεων κυμαίνονται στο διάστημα 12-60. Υψηλές τιμές δηλώνουν θετική στάση)
 1. Έλεγχος ανεξαρτησίας με Χι-τετράγωνο κατανομή
 2. Έλεγχος καλής προσαρμογής με Χι-τετράγωνο κατανομή
 3. Έλεγχος t ανεξάρτητων δειγμάτων
 4. Έλεγχος t σχετιζόμενων δειγμάτων
 5. Με τον έλεγχο του συντελεστή pearson
 6. Έλεγχος Z για τη μέση τιμή ενός πληθυσμού



Ασκήσεις επανάληψης: Στάσεις μαθητών απέναντι στις φυσικές επιστήμες (2 από 5)

2. Με τη βοήθεια των πινάκων της επόμενης διαφάνειας και των στοιχείων σχετικά με τις κλίμακες που δίνονται στην προηγούμενη άσκηση, να γίνει έλεγχος t ανεξαρτήτων δειγμάτων και επιλεχθεί η κατάλληλη απάντηση ($\alpha=0,05$)
1. Δεν υπάρχει σχέση φύλου και άγχους στην εκμάθηση των Φυσικών Επιστημών.
 2. Τα αγόρια εκδηλώνουν σημαντικά υψηλότερο επίπεδο άγχους από τα κορίτσια και δεν ισχύει η προϋπόθεση των ίσων διακυμάνσεων στους πληθυσμούς των δύο φύλων
 3. Τα αγόρια εκδηλώνουν σημαντικά χαμηλότερο επίπεδο άγχους από τα κορίτσια και δεν ισχύει η προϋπόθεση των ίσων διακυμάνσεων στους πληθυσμούς των δύο φύλων



Ασκήσεις επανάληψης: Στάσεις μαθητών απέναντι στις φυσικές επιστήμες (3 από 5)

4. Τα αγόρια εκδηλώνουν σημαντικά χαμηλότερο επίπεδο άγχους από τα κορίτσια και ισχύει η προϋπόθεση των ίσων διακυμάνσεων στους πληθυσμούς των δύο φύλων
5. Τα αγόρια εκδηλώνουν σημαντικά υψηλότερο επίπεδο άγχους από τα κορίτσια και ισχύει η προϋπόθεση των ίσων διακυμάνσεων στους πληθυσμούς των δύο φύλων



Ασκήσεις επανάληψης: Στάσεις μαθητών απέναντι στις φυσικές επιστήμες (4 από 5)

Group Statistics

	sex ΦΥΛΟ	N	Mean	Std. Deviation	Std. Error Mean
ΑΓΧΟΣ ΣΤΙΣ ΦΥΣΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	ΑΓΟΡΙ	183	39,0899	7,83000	,57881
	ΚΟΡΙΤΣΙ	177	35,3510	8,68709	,65296
ΧΡΗΣΙΜΟΤΗΤΑ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ	ΑΓΟΡΙ	183	42,8557	7,68591	,56816
	ΚΟΡΙΤΣΙ	177	42,5708	8,48325	,63764



Ασκήσεις επανάληψης: Στάσεις μαθητών απέναντι στις φυσικές επιστήμες (5 από 5)

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2- tailed)
ΑΓΧΟΣ ΣΤΙΣ ΦΥΣΙΚΕΣ ΕΠΙΣΤΗΜΕΣ	Equal variances assumed	,956	,329	4,292	358	,000
	Equal variances not assumed			4,285	351,429	,000
ΧΡΗΣΙΜΟΤΗΤΑ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ	Equal variances assumed	,351	,554	,334	358	,738
	Equal variances not assumed			,334	351,902	,739



Τέλος Ενότητας

Επαγωγική Στατιστική

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αθηνών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Σημειώματα

Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση 1.0.



Σημείωμα Αναφοράς

Copyright Εθνικών και Καποδιστριακών Πανεπιστημίων Αθηνών 2015, Βασίλης Γιαλαμάς 2015. Βασίλης Γιαλαμάς. «Μεθοδολογία των Επιστημών του Ανθρώπου: Στατιστική. Επαγωγική Στατιστική». Έκδοση: 1.0. Αθήνα 2015. Διαθέσιμο από τη δικτυακή διεύθυνση: <http://opencourses.uoa.gr/courses/ECD102/>.



Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



[1] <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Ως **Μη Εμπορική** ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο

Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.



Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.



Σημείωμα Χρήσης Έργων Τρίτων (1/4)

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Εικόνες/Σχήματα/Διαγράμματα/Φωτογραφίες

Εικόνα 1, Σελίδα 6: Εικόνα γραφικής παράστασης τριών πιθανών γραμμών /
Copyrighted

Εικόνα 2, Σελίδα 7: Εικόνα γραφικής παράστασης ανεξάρτητης προς εξαρτημένη
μεταβλητή / Copyrighted

Εικόνα 3, Σελίδα 11: Εικόνα γραφικής απεικόνισης των συντελεστών της γραμμής
παλινδρόμησης / Copyrighted

Εικόνα 4, Σελίδα 12: Εικόνα γραφικής παράστασης μη γραμμικής σχέσης /
Copyrighted

Εικόνα 5, Σελίδα 14: Εικόνα γραφικής παράστασης αριθμού εκκλησιών προς
ληστείες / Copyrighted



Σημείωμα Χρήσης Έργων Τρίτων (2/4)

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Εικόνες/Σχήματα/Διαγράμματα/Φωτογραφίες

Διάγραμμα 1, Σελίδα 15: Διάγραμμα μαθητών χαμηλού εισοδήματος που εισάγονται στο Πανεπιστήμιο / Copyright Σ. Πατάκης ΑΕΕΔΕ (Εκδόσεις Πατάκη) και Β. Γιαλαμάς, 2004 / Πηγή: «Στατιστικές Τεχνικές και Εφαρμογές στις Επιστήμες της Αγωγής» Β. Γιαλαμάς, Εκδόσεις Πατάκη

Διάγραμμα 2, Σελίδα 16: Διάγραμμα με $r = 0,084$ / Copyright Σ. Πατάκης ΑΕΕΔΕ (Εκδόσεις Πατάκη) και Β. Γιαλαμάς, 2004 / Πηγή: «Στατιστικές Τεχνικές και Εφαρμογές στις Επιστήμες της Αγωγής» Β. Γιαλαμάς, Εκδόσεις Πατάκη

Διάγραμμα 3, Σελίδα 16: Διάγραμμα με $r=0,83$ / Copyright Σ. Πατάκης ΑΕΕΔΕ (Εκδόσεις Πατάκη) και Β. Γιαλαμάς, 2004 / Πηγή: «Στατιστικές Τεχνικές και Εφαρμογές στις Επιστήμες της Αγωγής» Β. Γιαλαμάς, Εκδόσεις Πατάκη



Σημείωμα Χρήσης Έργων Τρίτων (3/4)

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Εικόνες/Σχήματα/Διαγράμματα/Φωτογραφίες

Εικόνα 6, Σελίδα 18: Εικόνα οθόνης ηλεκτρονικού υπολογιστή που γίνεται χρήση του SPSS / Copyrighted

Εικόνα 7, Σελίδα 19: Εικόνα οθόνης ηλεκτρονικού υπολογιστή που γίνεται χρήση του SPSS / Copyrighted



Σημείωμα Χρήσης Έργων Τρίτων (4/4)

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Πίνακες

Πίνακας 1, Σελίδα 10: Πίνακας δεδομένων / Copyrighted

Πίνακας 2, Σελίδα 20: Πίνακας δεδομένων για προσδόκιμο επιβίωσης γυναικών / Copyrighted

Πίνακας 3, Σελίδα 24: Πίνακας δεδομένων για τη στάση των μαθητών απέναντι στις φυσικές επιστήμες / Copyrighted

Πίνακας 4, Σελίδα 25: Πίνακας δεδομένων για τη στάση των μαθητών απέναντι στις φυσικές επιστήμες / Copyrighted

