

ΑΝΑΛΥΣΗ ΧΡΟΝΟΣΕΙΡΩΝ ΚΑΙ ΔΕΔΟΜΕΝΩΝ ΠΕΡΙΒΑΛΛΟΝΤΙΚΩΝ ΚΙΝΔΥΝΩΝ

Καθ. Θεόδωρος Καρακασίδης
Δρ Αθανάσιος Φράγκου

Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών
«Βιώσιμη Διαχείριση Περιβαλλοντικών Αλλαγών και
Κυκλική Οικονομία»

Γραμμική Παλινδρόμηση

Εάν δύο μεταβλητές X , Y συσχετίζονται γραμμικά μεταξύ τους - \rightarrow όλες οι τιμές της μίας Y (εξαρτημένη) μπορούν να προβλεφθούν εάν είναι γνωστές οι τιμές της άλλης X (ανεξάρτητη)

Ευθεία παλινδρόμησης $Y = a + bX \rightarrow$ Ο απλούστερος τύπος μοντέλου που συνδέει τις X και Y

a σταθερά

b η κλίση της ευθείας

Ένα μοντέλο που δίνει τη δυνατότητα στην τιμή Y να μη βρίσκεται ακριβώς στην ευθεία $Y = a + bX$ είναι το $Y = a + bX + e$ όπου e το **σφάλμα** και ισούται με τη **διαφορά** της παρατηρημένης τιμής Y από τη θεωρητική τιμή $a + bX$

Γραμμική Παλινδρόμιση

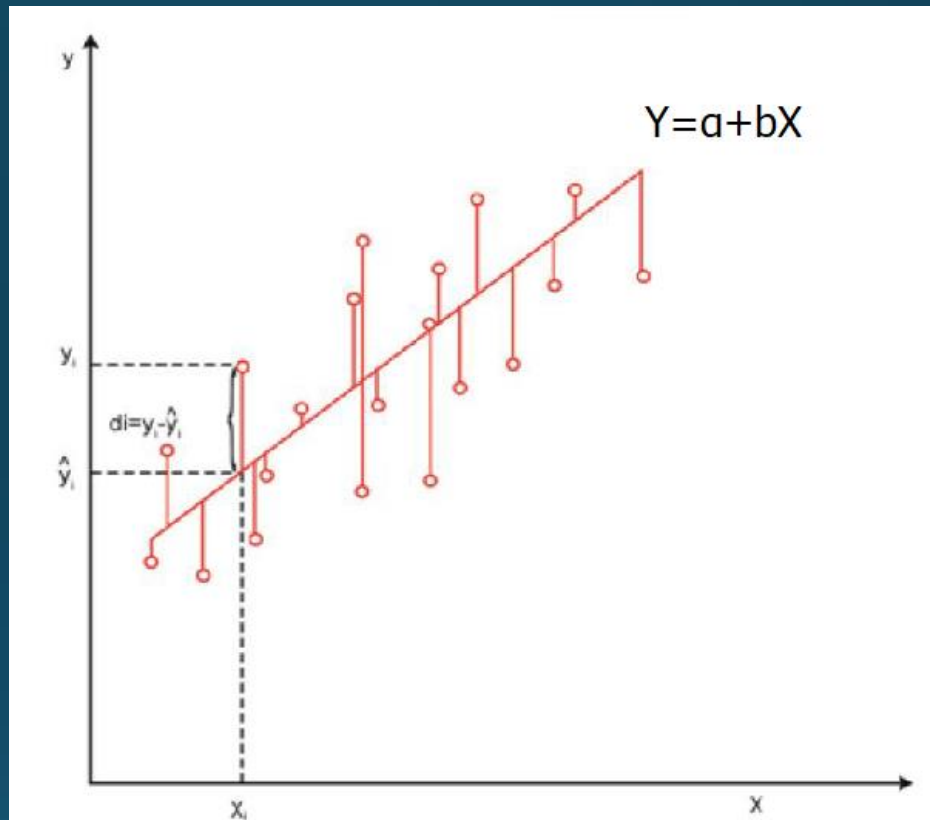
Ποια είναι όμως η ευθεία που προσαρμόζεται καλύτερα στα δεδομένα;

Η μέθοδος των ελαχίστων τετραγώνων αποκαλύπτει ότι η ευθεία που προσαρμόζεται καλύτερα είναι αυτή που ελαχιστοποιεί το άθροισμα των τετραγώνων των σφαλμάτων

$$\sum_{i=1}^n e_i = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Γραμμική Παλινδρόμηση

Η ευθεία
παλινδρόμησης



Οι συντελεστές
δίνονται από τις
σχέσεις

$$b = \frac{N(\sum_{i=1}^N x_i y_i) - (\sum_{i=1}^N x_i)(\sum_{i=1}^N y_i)}{N \sum_{i=1}^N x_i^2 - (\sum_{i=1}^N x_i)^2}$$

$$a = \frac{(\sum_{i=1}^N x_i^2)(\sum_{i=1}^N y_i) - (\sum_{i=1}^N x_i)(\sum_{i=1}^N x_i y_i)}{N \sum_{i=1}^N x_i^2 - (\sum_{i=1}^N x_i)^2}$$

Γραμμική Παλινδρόμηση

Το κατά πόσο **καλά** η ευθεία προσαρμόζεται στις μετρήσεις, μας το δίνει ο **συντελεστής προσδιορισμού** $R^2 = 1 - \frac{SSE}{SST}$

Με SST την ολική διακύμανση των Y που είναι

$$SST = \sum_{i=1}^n (Y_i - \bar{Y})^2 = SSE + SSR$$

όπου

$$SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2, SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Παράδειγμα 5.11: Σε ένα εργαστηριακό πείραμα φυσικής σχετικά με την ομαλά επιταχυνόμενη κίνηση σώματος καταγράφηκαν οι ακόλουθες μετρήσεις ταχύτητας ως συνάρτηση του χρόνου (Πίνακας 5.28). Να γίνει προσαρμογή καμπύλης πρώτου βαθμού.

a/a	$t(s)$	$v(cm/s)$
1	0.1	4.5
2	0.2	9.75
3	0.3	15.25
4	0.4	20.5
5	0.5	25.75
6	0.6	30.75
7	0.7	35.5
8	0.8	40.75
9	0.9	45.75
10	1.0	50.25
11	1.1	54.75
12	1.2	60.25
13	1.3	64.25
14	1.4	68.75
15	1.5	72.75
16	1.6	77.05

Γραμμική Παλινδρόμηση (Excel)

Καταχωρούμε τα δεδομένα

A	B	C
A/A	ΧΡΟΝΟΣ	ΤΑΧΥΤΗΤΑ
1	0,1	4,5
2	0,2	9,75
3	0,3	15,25
4	0,4	20,5
5	0,5	25,75
6	0,6	30,75
7	0,7	35,5
8	0,8	40,75
9	0,9	45,75
10	1	50,25
11	1,1	54,75
12	1,2	60,25
13	1,3	64,25
14	1,4	68,75
15	1,5	72,75
16	1,6	77,05

Γραμμική Παλινδρόμηση (Excel)

Κάνουμε τους υπολογισμούς

Μέση τιμή X

μ.τ X
0,85

Διασπορά X

διασπορά X
0,22666667

Μέση τιμή Y

μ.τ.Y
42,284375

Διασπορά Y

διασπορά Y
534,64724

Γραμμική Παλινδρόμηση (Excel)

Τα απαραίτητα αθροίσματα
και γινόμενα

χίγι	Σχίγι	16*μ.τ Χ επί μ.τ.Υ	1/(n-1)
0,45	740,08	575,0675	0,06666667

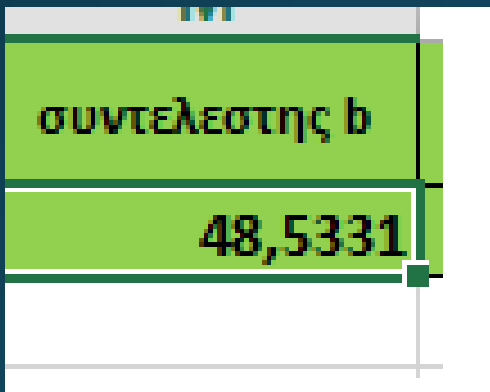
Διασπορά XY που προκύπτει
από τα υπολογισμένα
αθροίσματα και γινόμενα

7	διασπορά XY
	11,000833

Γραμμική Παλινδρόμηση (Excel)

Υπολογισμός των δύο συντελεστών

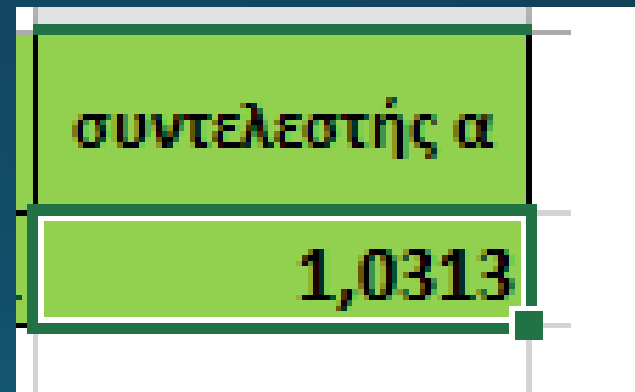
Συντελεστής b



A screenshot of an Excel spreadsheet showing the calculation of coefficient b. The cell contains the text "συντελεστής b" in the top row and the numerical value "48,5331" in the bottom row. The cell is highlighted with a green border.

συντελεστής b
48,5331

Συντελεστής α

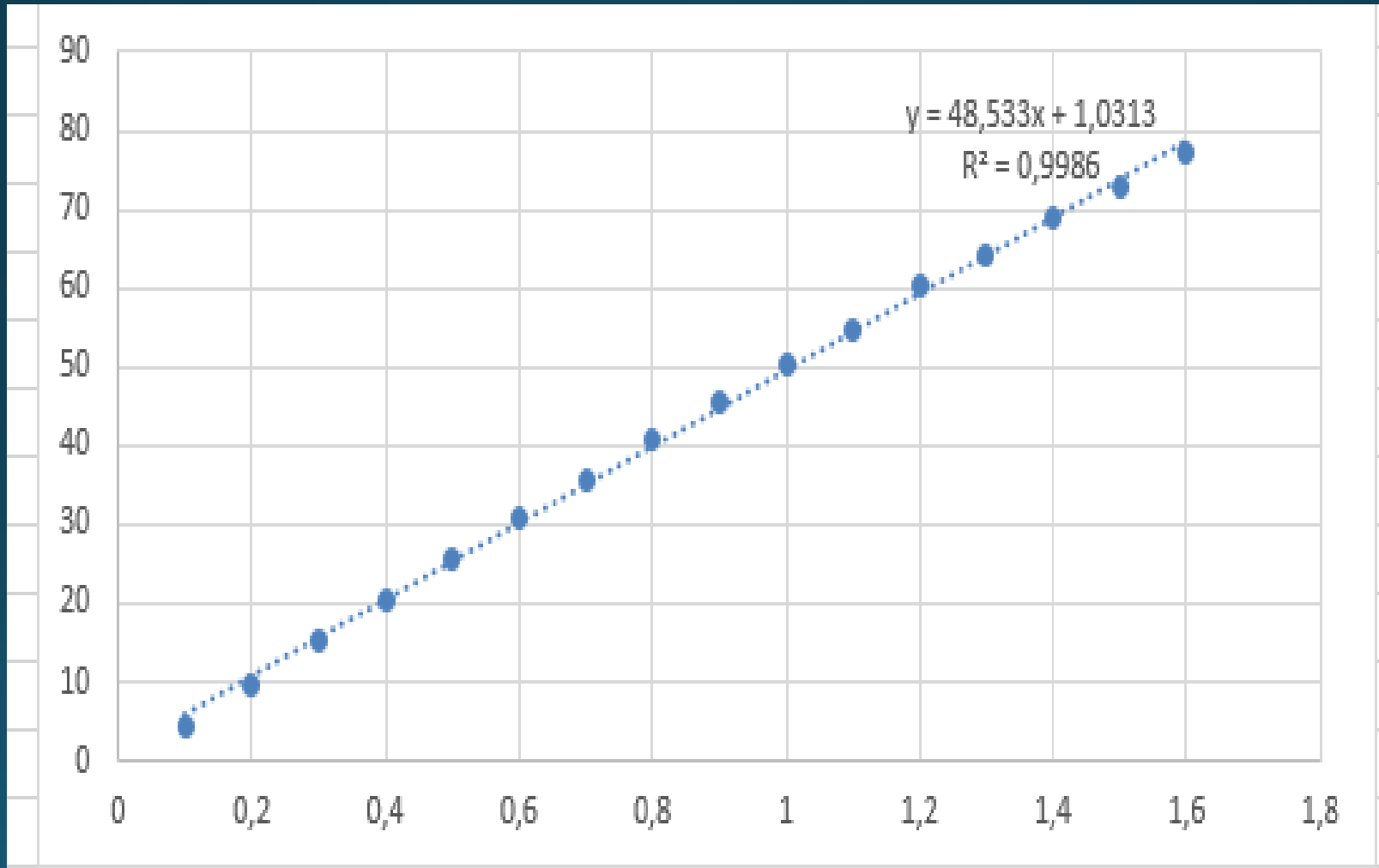


A screenshot of an Excel spreadsheet showing the calculation of coefficient a. The cell contains the text "συντελεστής α" in the top row and the numerical value "1,0313" in the bottom row. The cell is highlighted with a green border.

συντελεστής α
1,0313

Γραμμική Παλινδρόμηση (Excel)

Υπολογισμός της ευθείας παλινδρόμησης

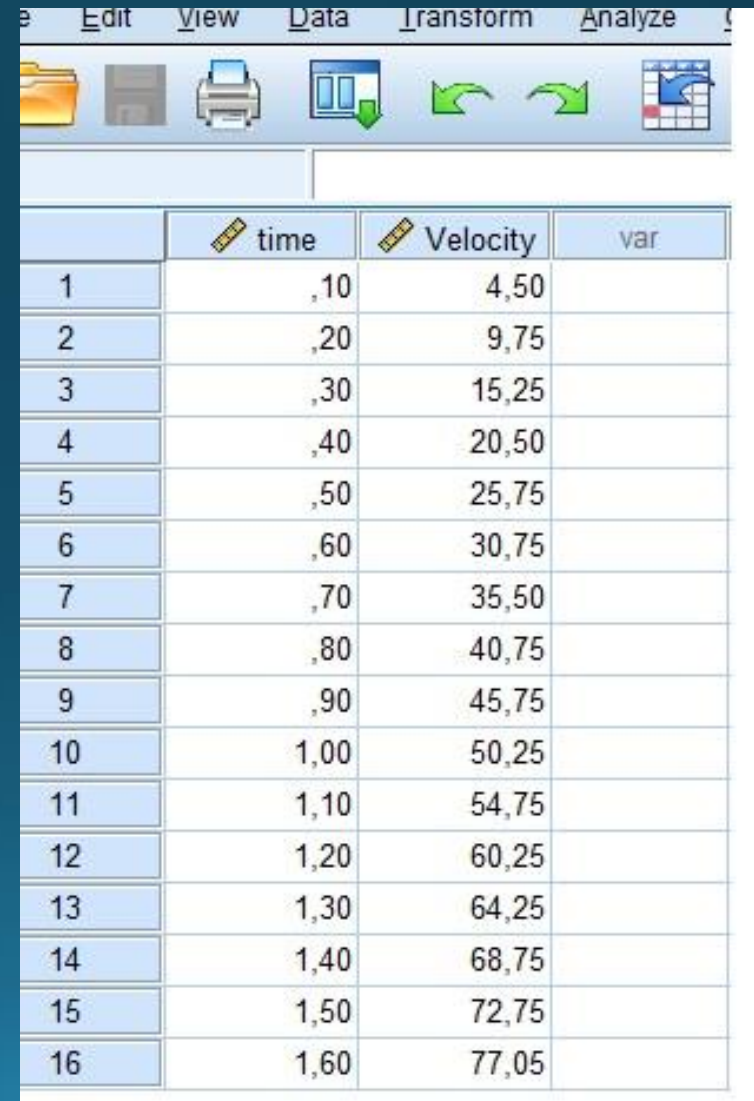


Γραμμική Παλινδρόμηση (SPSS)

Στα δεδομένα θα προσαρμοστεί ευθεία $Y = a + bX + e$ με

$Y =$ Ταχύτητα (Εξαρτημένη μεταβλητή)

$X =$ Χρόνος (Ανεξάρτητη μεταβλητή)



The screenshot shows the SPSS software interface with a data table. The table has three columns: 'time', 'Velocity', and 'var'. The 'time' column contains values from 0,10 to 1,60 in increments of 0,10. The 'Velocity' column contains values from 4,50 to 77,05 in increments of 5,25. The 'var' column is empty. The table is displayed in a grid format with a light blue header and alternating row colors.

	time	Velocity	var
1	,10	4,50	
2	,20	9,75	
3	,30	15,25	
4	,40	20,50	
5	,50	25,75	
6	,60	30,75	
7	,70	35,50	
8	,80	40,75	
9	,90	45,75	
10	1,00	50,25	
11	1,10	54,75	
12	1,20	60,25	
13	1,30	64,25	
14	1,40	68,75	
15	1,50	72,75	
16	1,60	77,05	

Γραμμική Παλινδρόμηση (SPSS)

Από την επιλογή Analyze

The screenshot shows the SPSS interface with the 'Analyze' menu open. A red arrow points from the text 'Από την επιλογή Analyze' to the 'Analyze' menu. Another red arrow points from the 'Regression' option in the 'Analyze' menu to the 'Linear...' option in the sub-menu. The data view shows two columns: 'time' and 'Velocity'.

	time	Velocity
3	,30	15,25
4	,40	20,50
5	,50	25,75
6	,60	30,75
7	,70	35,50
8	,80	40,75
9	,90	45,75
10	1,00	50,25
11	1,10	54,75
12	1,20	60,25
13	1,30	64,25

The 'Analyze' menu options are:

- Reports
- Descriptive Statistics
- Bayesian Statistics
- Tables
- Compare Means
- General Linear Model
- Generalized Linear Models
- Mixed Models
- Correlate
- Regression**
- Loglinear
- Neural Networks
- Classify
- Dimension Reduction
- Scale
- Nonparametric Tests
- Forecasting
- Survival
- Multiple Response
- Missing Value Analysis...

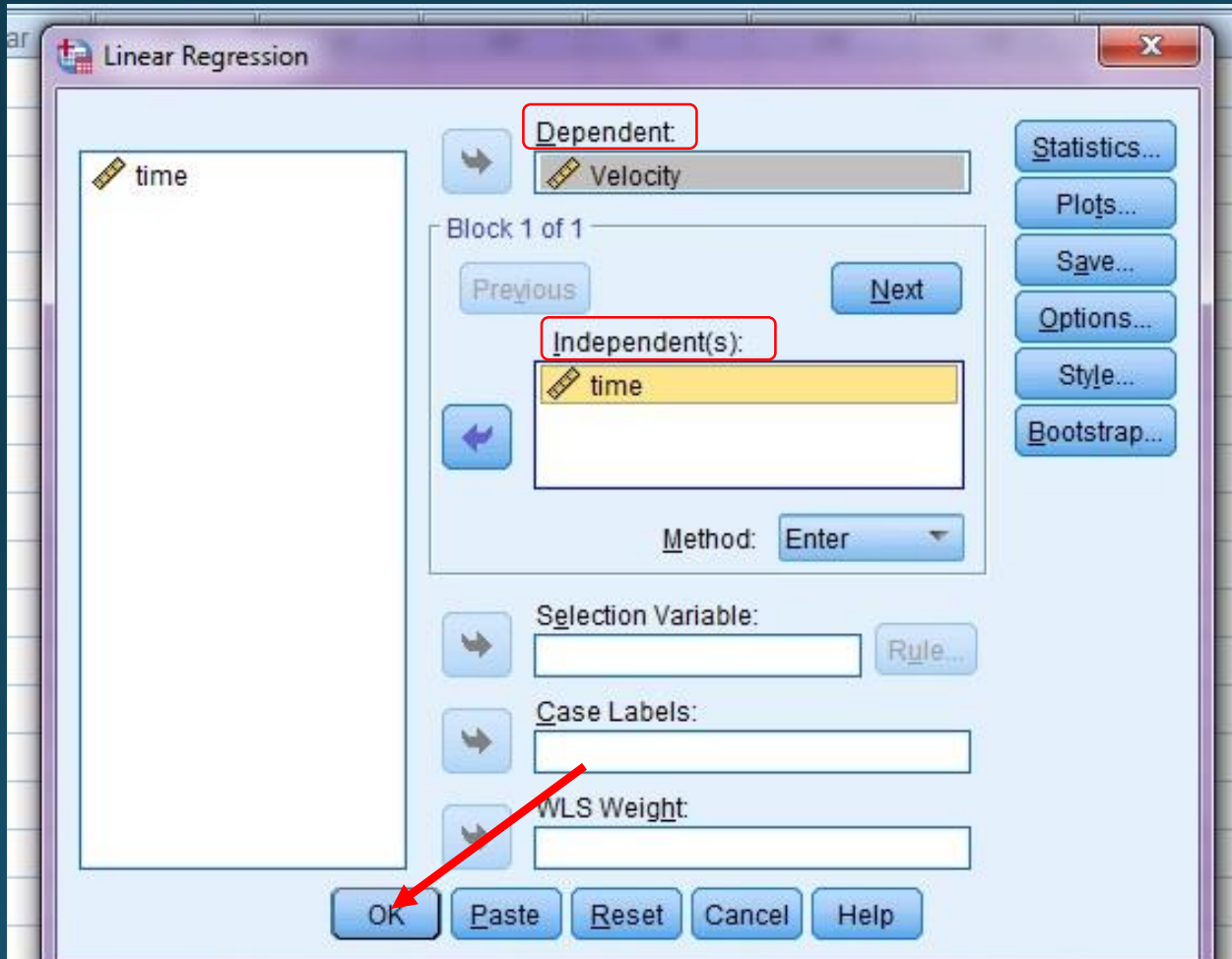
The 'Regression' sub-menu options are:

- Automatic Linear Modeling...
- Linear...**
- Curve Estimation...
- Partial Least Squares...
- Binary Logistic...
- Multinomial Logistic...
- Ordinal...
- Probit...
- Nonlinear...
- Weight Estimation...

Γραμμική Παλινδρόμηση (SPSS)

Επιλογή Y = Ταχύτητα (Εξαρτημένη μεταβλητή)

X= Χρόνος (Ανεξάρτητη μεταβλητή)



Γραμμική Παλινδρόμηση (SPSS)


Ενημερωτικός πίνακας μεταβλητών

Model	Variables Entered	Variables Removed	Method
1	time ^b	.	Enter

a. Dependent Variable: Velocity
b. All requested variables entered.

Γραμμική Παλινδρόμηση (SPSS)

Εύρεση συντελεστή προσδιορισμού R^2



Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,999 ^a	,999	,999	,89212

a. Predictors: (Constant), time

Υψηλός συντελεστής προσδιορισμού R^2 -> πολύ καλή προσαρμογή της ευθείας παλινδρόμησης που σημαίνει ότι ο χρόνος (ανεξ) ερμηνεύει το 99.9% της ταχύτητας (εξαρτ).

Γραμμική Παλινδρόμηση (SPSS)

Οι συντελεστές της ευθείας παλινδρόμησης

Model		Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.
		B	Std. Error			
1	(Constant)	1,031	,468		2,204	,045
	time	48,533	,484	,999	100,312	,000

a. Dependent Variable: Velocity

Η ευθεία παλινδρόμησης είναι η $Y = 1.031 + 48.533X$

Πολλαπλή Γραμμική Παλινδρόμηση

Συχνά μια εξαρτημένη μεταβλητή δεν εξαρτάται μόνο από μια ανεξάρτητη αλλά από πολλές, πχ το ύψος βροχής ενός τόπου μπορεί να εξαρτάται όχι μόνο από τη θερμοκρασία του αέρα αλλά και από την πίεση.

Πρόβλημα: Εύρεση μιας γραμμικής συνάρτησης που να συνδέει την εξαρτημένη μεταβλητή με τις ανεξάρτητες.

Πολλαπλή γραμμική παλινδρόμηση -> συνδέει την εξαρτημένη μεταβλητή Y με τις ανεξάρτητες X_1, X_2, \dots, X_n

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p$$

Πολλαπλή Γραμμική Παλινδρόμηση

Ενδιαφερόμαστε για το «καλύτερο» επίπεδο παλινδρόμησης λόγω πολλών μεταβλητών.

Στόχος : Η μείωση της ολικής μεταβλητότητας που δεν έχει ερμηνευτεί.

Συντελεστής προσδιορισμού $R^2 = 1 - \frac{SSE}{SST}$

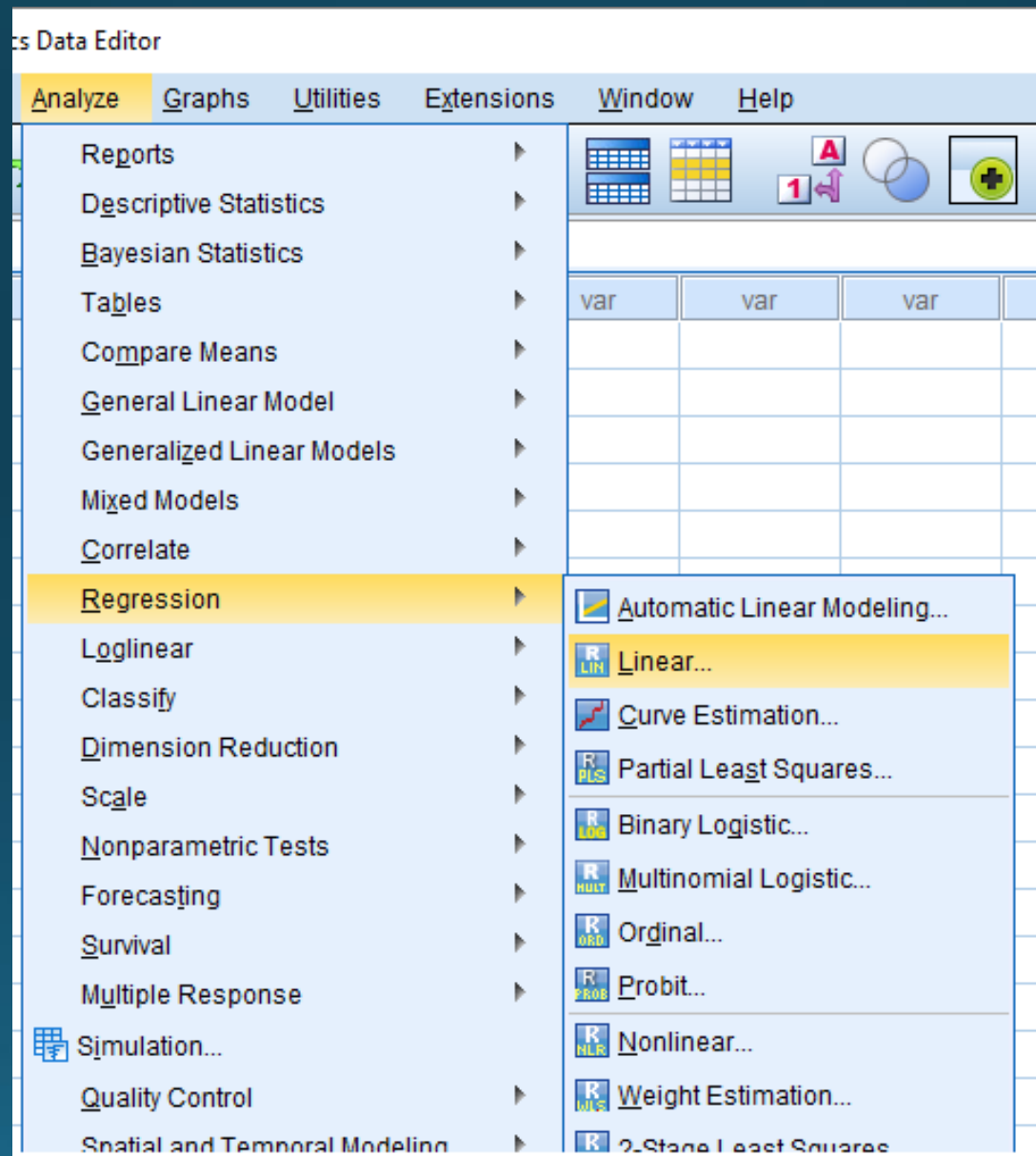
- Εάν το $R^2 = 1$, τότε οι εκτιμημένες τιμές της Y ταυτίζονται με τις παρατηρημένες των X_i και έχουμε τέλεια προσαρμογή.
- Εάν το $R^2 = 0$, τότε οι εκτιμημένες τιμές της Y ισούνται με τη μέση τιμή της Y και η Y δεν επηρεάζεται από τις X_i .

Πολλαπλή Γραμμική Παλινδρόμηση

Να βρεθεί το γραμμικό μοντέλο παλινδρόμησης: Ένας χημικός ενδιαφέρεται να προσδιορίσει το χάσιμο βάρους Y μιας χημικής ένωσης, σαν συνάρτηση του χρόνου που η ουσία έμεινε εκτεθειμένη στον αέρα και στην υγρασία του περιβάλλοντος με τα εξής αποτελέσματα

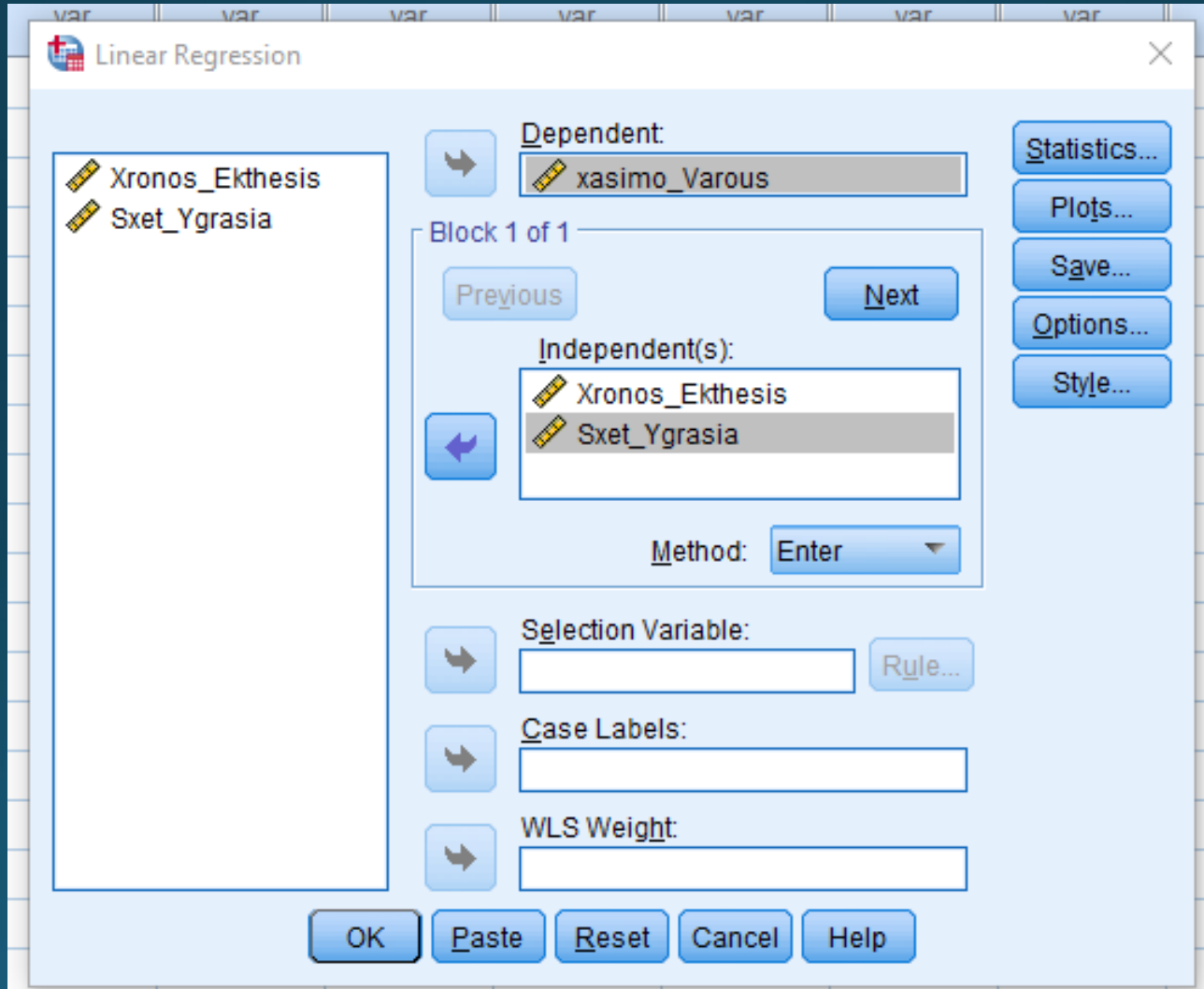
Χάσιμο Βάρους Y	Χρόνος έκθεσης X_1	Σχετική υγρασία X_2
4,3	4	0,2
5,5	5	0,2
6,8	6	0,2
8	7	0,2
4	4	0,3
5,2	5	0,3
6,6	6	0,3
7,5	7	0,3
2	4	0,4
4	5	0,4
5,7	6	0,4
6,5	7	0,4

Πολλαπλή Γραμμική Παλινδρόμηση (SPSS)



Πολλαπλή Γραμμική Παλινδρόμηση (SPSS)

Επιλογή ανεξάρτητης και εξαρτημένων μεταβλητών



Πολλαπλή Γραμμική Παλινδρόμηση (SPSS)

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Sxet_Ygrasia, Xronos_Ekthesis ^b	.	Enter

a. Dependent Variable: xasimo_Varous

b. All requested variables entered.

Ενημερωτικός πίνακας
μεταβλητών

Πολλαπλή Γραμμική Παλινδρόμηση (SPSS)

Εύρεση συντελεστή προσδιορισμού R^2

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,979 ^a	,959	,949	,38658

a. Predictors: (Constant), Sxet_Ygrasia, Xronos_Ekthesis
b. Dependent Variable: xasimo_Varous

- Το $R^2 = 0.959$, τότε οι εκτιμημένες τιμές της Y σχεδόν ταυτίζονται με τις παρατηρημένες των X_i και έχουμε εξαιρετική προσαρμογή και οι ανεξάρτητες μεταβλητές σχετική υγρασία και χρόνος έκθεσης ερμηνεύουν σε πολύ καλό βαθμό το χάσιμο βάρους της ουσίας.

Πολλαπλή Γραμμική Παλινδρόμηση (SPSS)

Οι συντελεστές της ευθείας παλινδρόμησης

Model		Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.
		B	Std. Error			
1	(Constant)	,667	,694		,960	,362
	Xronos_Ekthesis	1,317	,100	,895	13,191	,000
	Sxet_Ygrasia	-8,000	1,367	-,397	-5,853	,000

a. Dependent Variable: xasimo_Varous

Η εξίσωση παλινδρόμησης είναι η $Y = -0,667 + 1,317X_1 - 8X_2$

Άρα εάν θέλουμε να **προβλέψουμε** το χάσιμο βάρους μιας ουσίας μετά από 6,5 λεπτά ώρες έκθεσης στον αέρα και σε σχετική υγρασία 0.35 έχουμε:

$$Y = 0,63 + 1,32 * 6,5 - 8 * 0,35 = 6,41$$

Γραμμική Παλινδρόμηση (SPSS) Άσκηση

Για τη μείωση των αερίων που δημιουργούν το φαινόμενο του θερμοκηπίου σε ένα ερευνητικό εργαστήριο μελετήθηκε η ποσότητα βελτιωτικής ουσίας στα καύσιμα σε σχέση με τις μειώσεις των ρυπογόνων ουσιών και βρέθηκε ότι είναι γραμμική. Να προσαρμοστεί ένα κατάλληλο γραμμικό μοντέλο που συνδέει τη μείωση των ρυπογόνων ουσιών (εξαρτημένη Y) με την ποσότητα βελτιωτικής ουσίας (X ανεξάρτητη) στα καύσιμα.

Στα δεδομένα θα προσαρμοστεί ευθεία $Y = a + bX + e$ με

Y = μείωση ρυπογόνων ουσιών

X = ποσότητα βελτιωτικής ουσίας