ΑΡΙΣΤΟΤΕΛΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΟΝΙΚΗΣ

ΓΕΩΠΟΝΙΚΗ ΣΧΟΛΗ
Α.Π.Θ.

Απλή Ευθύγραμμη Συμμεταβολή

Επιστημονική Επιμέλεια

Δρ. Γεώργιος Μενεξές

Τομέας Φυτών Μεγάλης Καλλιέργειας και Οικολογίας,
Εργαστήριο Γεωργίας

Εισαγωγή

Στην ενότητα αυτή θα ασχοληθούμε με προβλήματα στα οποία εμπλέκονται δύο τυχαίες μεταβλητές, όπως για παράδειγμα ύψος και βάρος φοιτητών, αμοιβές εργαζομένων και έτη προϋπηρεσίας, οικογενειακό εισόδημα και αντίστοιχες δαπάνες. Η μεταβλητή της οποίας οι τιμές καθορίζονται από τον ερευνητή ονομάζεται **ανεξάρτητη** ή **ερμηνευτική** ενώ η άλλη **εξαρτημένη**.

Ανάλυση Παλινδρόμησης και Συσχέτιση

Η **Ανάλυση Παλινδρόμησης** έχει ως στόχο την επιλογή κατάλληλου μοντέλου το οποίο θα χρησιμοποιηθεί για την εκτίμηση των τιμών της εξαρτημένης μεταβλητής για δεδομένες τιμές της ανεξάρτητης.

Η **Συσχέτιση** έχει ως στόχο τη μέτρηση του βαθμού εξάρτησης, δηλαδή της φοράς και της έντασης της συμμεταβολής που υπάρχει μεταξύ των τυχαίων μεταβλητών Χ και Υ.

Απλή Γραμμική Παλινδρόμηση

Απλή Γραμμική Παλινδρόμηση

Ας υποθέσουμε ότι θέλουμε να μελετήσουμε τη σχέση που υπάρχει ανάμεσα σε δύο μεταβλητές X και Y. Ας υποθέσουμε επίσης ότι η συναρτησιακή σχέση ανάμεσα στις δύο μεταβλητές είναι γραμμική. Στην περίπτωση αυτή η σχέση μεταξύ των δύο μεταβλητών εκφράζεται από το μοντέλο της απλής γραμμικής παλινδρόμησης:

$$Y_i = \mathrm{a} + \mathrm{b}\ X_i + \varepsilon_i, \quad i = 1, \ldots, n$$

Όπου

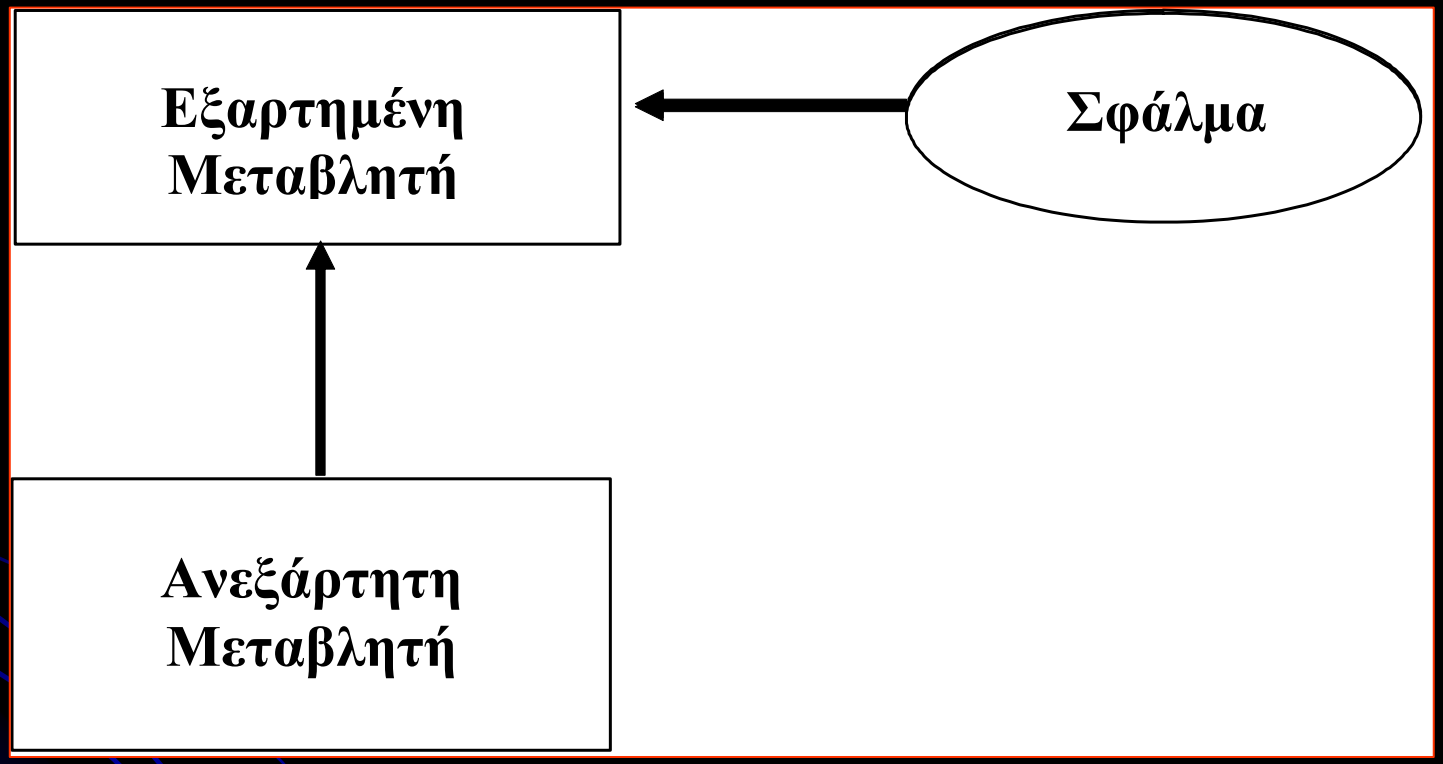
X_i : η i παρατήρηση της ανεξάρτητης μεταβλητής

Y_i : η i παρατήρηση της εξαρτημένης μεταβλητής

a , b : οι συντελεστές της ευθείας παλινδρόμησης

ε_i : το σφάλμα της εξίσωσης

Το Υπόδειγμα Διαγραμματικά



Οι Συντελεστές Παλινδρόμησης

Οι συντελεστές a και b υπολογίζονται από τους παρακάτω τύπους, στους οποίους καταλήγουμε με τη μέθοδο των ελαχίστων τετραγώνων.

$$b = \frac{\sum\left[(Y_i - \bar{Y})(X_i - \bar{X})\right]}{\sum\left[(X_i - \bar{X})^2\right]} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$a = \frac{\sum Y_i}{n} - b\,\frac{\sum X_i}{n} = \bar{Y} - b\,\bar{X}$$

ένας εναλλακτικός τρόπος υπολογισμού του b είναι

$$b = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}$$

Η Ευθεία Παλινδρόμησης

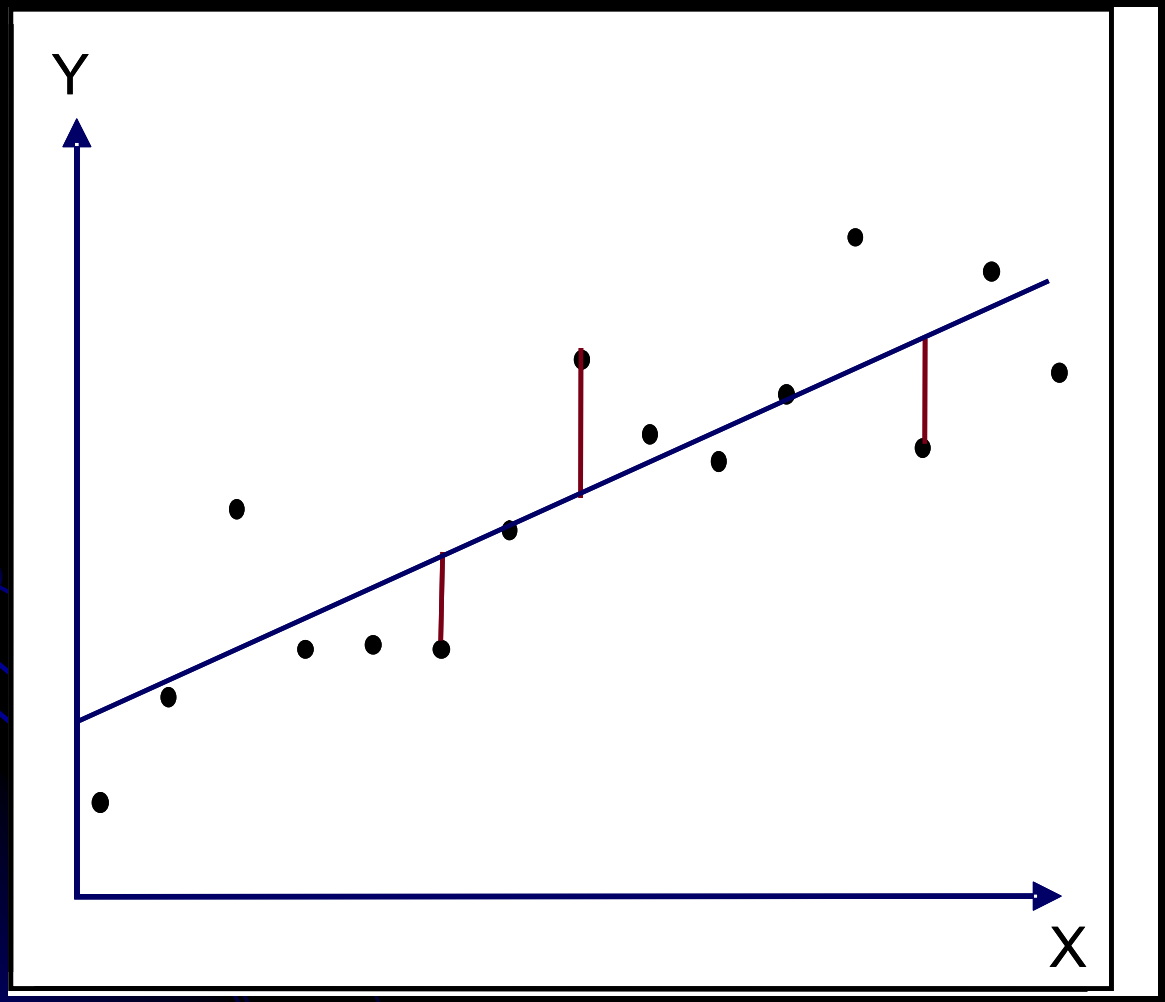
Έχοντας υπολογίσει τους συντελεστές της μπορούμε τώρα να ορίσουμε την **Ευθεία Παλινδρόμησης**

$$\hat{Y} = \mathrm{a} + \mathrm{b}\,X$$

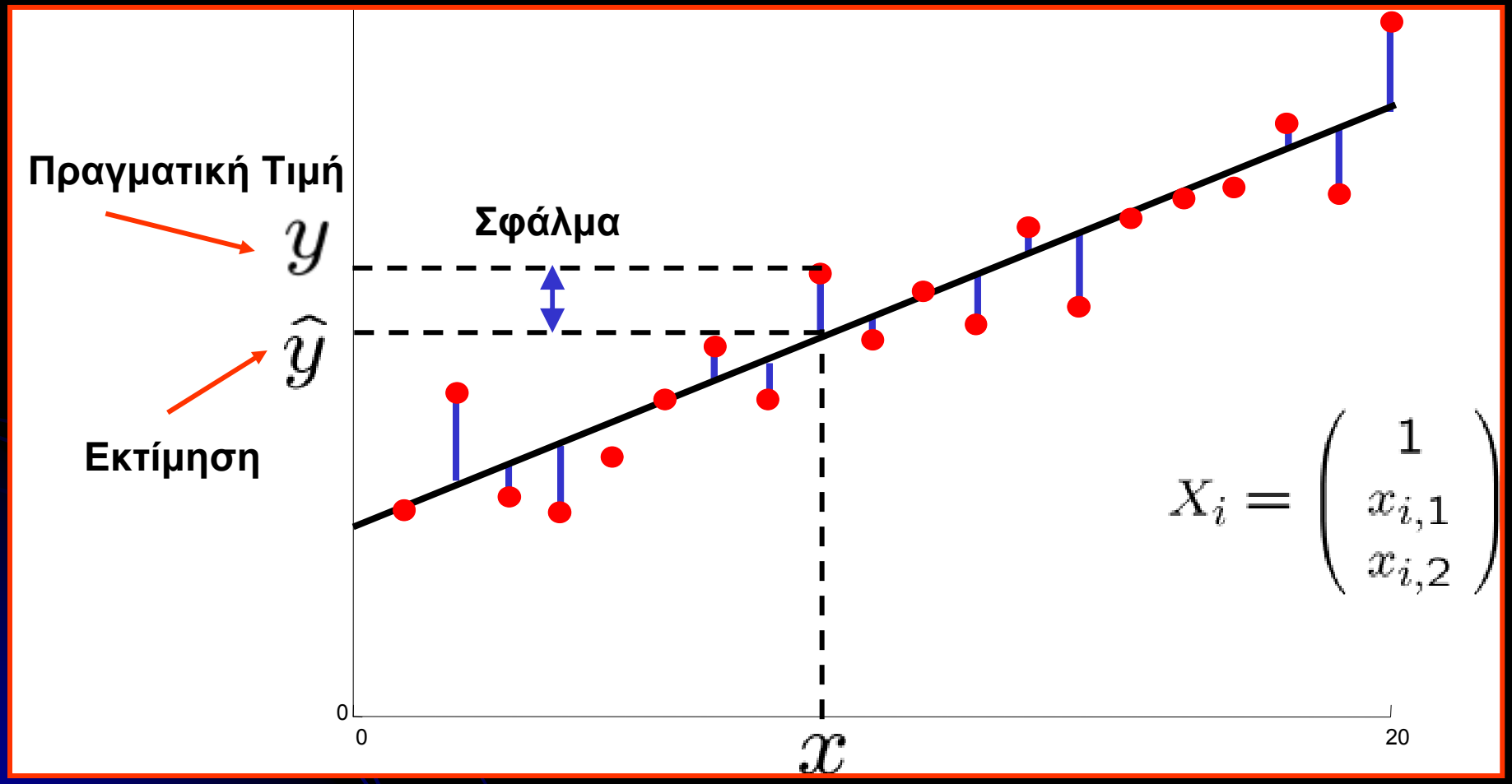
Ο συντελεστής a εκφράζει την τιμή της εξαρτημένης μεταβλητής Y όταν η ανεξάρτητη μεταβλητή X παίρνει την τιμή μηδέν.

Ο συντελεστής b εκφράζει τη μεταβολή της εξαρτημένης μεταβλητής Y όταν η ανεξάρτητη μεταβλητή X αυξηθεί κατά μία μονάδα.

Η Ευθεία Ελαχίστων Τετραγώνων (1)



Η Ευθεία Ελαχίστων Τετραγώνων (2)



Ο Συντελεστής Γραμμικής Συσχέτισης (Pearson)

Συντελεστής Συσχέτισης

Ο Συντελεστής Συσχέτισης εκφράζει τη φορά και την ένταση της συμμεταβολής των δύο μεταβλητών και μπορεί να χρησιμοποιηθεί ως μέτρο ισχυρότητας της γραμμικής εξάρτησής τους. Ορίζεται από τον τύπο:

$$r = \frac{\sum\left[\left(Y_i - \bar{Y}\right)\left(X_i - \bar{X}\right)\right]}{\sqrt{\sum\left(Y_i - \bar{Y}\right)^2 \sum\left(X_i - \bar{X}\right)^2}} = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sqrt{\left[\sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}\right]\left[\sum Y_i^2 - \frac{\left(\sum Y_i\right)^2}{n}\right]}}$$

και παίρνει τιμές στο διάστημα [-1, +1].

Ερμηνεία του Συντελεστή Γραμμικής Συσχέτισης

Ειδικότερα, όταν

- r = 1 τότε υπάρχει τέλεια θετική συσχέτιση μεταξύ των δύο μεταβλητών
- r = -1 τότε υπάρχει τέλεια αρνητική συσχέτιση μεταξύ των δύο μεταβλητών
- r = 0 τότε οι δύο μεταβλητές είναι γραμμικά ασυσχέτιστες

Θα πρέπει να τονιστεί ότι ο συντελεστής συσχέτισης είναι μέτρο γραμμικής εξάρτησης μόνο και όχι μέτρο εξάρτησης οποιασδήποτε μορφής. Αν δηλαδή $r = 0$ αυτό δεν συνεπάγεται ότι τα X και Y είναι ανεξάρτητα. Είναι μεν γραμμικά ασυσχέτιστα αλλά μπορεί να έχουν άλλη μορφή εξάρτησης.

Στατιστική Σημαντικότητα του Συντελεστή Συσχέτισης

H_0: $r=0$ (στον πληθυσμό)

H_1: $r \neq 0$

Υπολογίζουμε το στατιστικό:

$$|t| = \frac{\left|r\sqrt{n-2}\right|}{\sqrt{1-r^2}}$$

Το συγκρίνουμε με την κρίσιμη τιμή της *t*-Κατανομής για (*n*-2) β.ε. σε επίπεδο σημαντικότητας *α*/2.

Ο Συντελεστής Προσδιορισμού

Συντελεστής Προσδιορισμού

Ο Συντελεστής Προσδιορισμού χρησιμοποιείται ως κριτήριο καλής προσαρμογής των δεδομένων στο γραμμικό μοντέλο και εκφράζει το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής *Y* που ερμηνεύεται από την ανεξάρτητη μεταβλητή *X*.

Ορίζεται από τον τύπο

$$R^2 = \frac{\left(\sum\left[\left(Y_i - \bar{Y}\right)\left(X_i - \bar{X}\right)\right]\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2 \sum\left(X_i - \bar{X}\right)^2} = \frac{\left(\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}\right)^2}{\left[\sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}\right]\left[\sum Y_i^2 - \frac{\left(\sum Y_i\right)^2}{n}\right]}$$

και παίρνει τιμές μεταξύ *0* και *1* $(0 \le R^2 \le 1)$,

Ερμηνεία του Συντελεστή Προσδιορισμού

Κατά συνέπεια, όσο μεγαλύτερη είναι η τιμή του R^2, τόσο καλύτερη είναι η προσαρμογή των δεδομένων στο γραμμικό μοντέλο ή αλλιώς τόσο καλύτερα το γραμμικό μοντέλο εκφράζει τα δεδομένα,

Σημειώνεται ότι $r = (\text{προσημο } a_1)\left(+\sqrt{R^2}\right)$

Θα πρέπει να τονιστεί ότι χρειάζεται προσοχή στη χρησιμοποίηση του r και του R^2 . Το R^2 μπορεί να χρησιμοποιηθεί σαν μέτρο ισχυρότητας της γραμμικής σχέσης ανεξάρτητα από το αν το X παίρνει καθορισμένες τιμές ή αν είναι τυχαία μεταβλητή, Αντίθετα το r μπορεί να χρησιμοποιηθεί μόνο αν το Y και το X είναι τυχαίες μεταβλητές,

Πίνακας Ανάλυσης Παραλλακτικότητας-Διασποράς (ANOVA) μετά την Παλινδρόμηση

Πηγές Διακύμανσης	Αθροίσματα Τετραγώνων	Βαθμοί Ελευθερίας	Μέσα Αθροίσματα Τετραγώνων	F
Παλιδρόμησης	$\sum_{t=1}^{n}(\hat{Y}_t-\bar{Y})^2$	1	$\frac{\sum_{t=1}^{n}(\hat{Y}_t-\bar{Y})^2}{1}$	
Καταλοίπων	$\sum_{t=1}^{n}(Y_t-\hat{Y}_t)^2$	n - 2	$\frac{\sum_{t=1}^{n}(Y_t-\hat{Y}_t)^2}{n-2}$	$F=\frac{\frac{\sum_{t=1}^{n}(\hat{Y}_t-\bar{Y})^2}{1}}{\frac{\sum_{t=1}^{n}(Y_t-\hat{Y}_t)^2}{n-2}}$
Σύνολο	$\sum_{t=1}^{n}(Y_t-\bar{Y})^2$	n - 1		

Εκτιμώμενες Τιμές της Y

Σύγκριση με κρίσιμη τιμή της F(1, n-2)

Από τον Πίνακα ANOVA

- Στην ουσία ελέγχουμε την μηδενική υπόθεση:

H_0: b=0 (στον πληθυσμό)
H_1: b≠0

- Η τετραγωνική ρίζα του Μέσου Αθροίσματος Τετραγώνων που αντιστοιχεί στα Σφάλματα εκφράζει το Τυπικό Σφάλμα της εκτίμησης (S.E. of Estimate). Είναι ίση με την τυπική απόκλιση των σφαλμάτων. Εκφράζει την ακρίβεια της πρόβλεψης σε σύγκριση με το μέσο όρο της εξαρτημένης μεταβλητής.

Παράδειγμα

Παράδειγμα

Ένα Ιατρικό Κέντρο δοκιμάζει μια νέα δίαιτα αδυνατίσματος, Ο Πίνακας 4,1 περιέχει, για 5 τυχαία επιλεγμένα άτομα (Α – Ε) από αυτά που υποβλήθηκαν στη συγκεκριμένη δίαιτα, τον αριθμό των εβδομάδων (X) που την ακολούθησαν και το βάρος (Y) σε κιλά που έχασαν

Άτομα	**Α**	**Β**	**Γ**	**Δ**	**Ε**
X	2	3	5	6	8
Y	1	2	5	6	7

Με βάση τα παραπάνω στοιχεία και αν επιπλέον γνωρίζουμε ότι οι μεταβλητές X και Y συνδέονται με γραμμική σχέση:

Ερωτήματα

α. Να εκτιμηθεί η ευθεία παλινδρόμησης της Y πάνω στη X και να ερμηνευτεί ο συντελεστής της ανεξάρτητης μεταβλητής

β. Να εκτιμηθεί ο συντελεστής συσχέτισης και να ερμηνευτεί,

γ. Να εκτιμηθεί ο συντελεστής προσδιορισμού και να ερμηνευτεί

δ. Να εκτιμηθεί το βάρος που θα χάσει ένα άτομο αν ακολουθήσει τη δίαιτα για 4 εβδομάδες.

Λύση

Λύση

Με βάση τα δεδομένα της άσκησης δημιουργείται ο παρακάτω πίνακας

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
X_i	Y_i	X_i^2	Y_i^2	$(X_i-\bar{X})$	$(Y_i-\bar{Y})$	$(X_i-\bar{X})(Y_i-\bar{Y})$	$(X_i-\bar{X})^2$	$(Y_i-\bar{Y})^2$
2	1	4	1	-2,8	-3,2	8,96	7,84	10,24
3	2	9	4	-1,8	-2,2	3,96	3,24	4,84
5	5	25	25	0,2	0,8	0,16	0,04	0,64
6	6	36	36	1,2	1,8	2,16	1,44	3,24
8	7	64	49	3,2	2,8	8,96	10,24	7,84
24	**21**	**138**	**115**	**0**	**0**	**24,2**	**22,8**	**26,8**

Αθροίσματα

Υπολογισμοί

Βοηθητικά Στοιχεία

$$\sum X_i = 24,\ \sum Y_i = 21,\ \overline{X} = \frac{\sum X_i}{n} = \frac{24}{5} = 4,8,\ \overline{Y} = \frac{\sum Y_i}{n} = \frac{21}{5} = 4,2$$

$$\sum (X_i - \overline{X})^2 = 22,8,\ \sum (Y_i - \overline{Y})^2 = 26,8,\ \sum (X_i - \overline{X})(Y_i - \overline{Y}) = 24,2$$

α. Ευθεία Παλινδρόμησης

$$\mathbf{b} = \frac{\sum\left[(Y_i - \overline{Y})(X_i - \overline{X})\right]}{\sum\left[(X_i - \overline{X})^2\right]} = 1,06$$

Εναλλακτικά

$$\mathrm{b} = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}} = 1,06$$

$$\mathrm{a} = \overline{Y} - \mathrm{b}\,\overline{X} = -0,89$$

Ευθεία Παλινδρόμησης

Άρα η ευθεία παλινδρόμησης είναι:

$$\hat{Y} = -0,89 + 1,06X$$

Η τιμή του b δηλώνει ότι αν η διάρκεια της δίαιτας ενός ατόμου αυξηθεί κατά μία βδομάδα, η απώλεια βάρους του θα αυξηθεί κατά 1,06 κιλά

Βοηθητικοί Υπολογισμοί

Y	$Y-\bar{Y}$	$(Y-\bar{Y})^2$	$\hat{Y}$	$\hat{Y}-\bar{Y}$	$(\hat{Y}-\bar{Y})^2$
1	-3.20	10.24	1.23	-2.97	8.83
2	-2.20	4.84	2.29	-1.91	3.65
5	0.80	0.64	4.41	0.21	0.05
6	1.80	3.24	5.47	1.27	1.62
7	2.80	7.84	7.60	3.40	11.54
Σύνολο		**26.80**			**25.69**
Άθροισμα Τετραγώνων για τα Σφάλματα	26,80-25,69=1,11		Κρίσιμη Τιμή της $F(1, 3)$ σε α=0,05=10,128		
$\bar{Y}=4,2$					

Πίνακας Ανάλυσης Παραλλακτικότητας-Διασποράς μετά την Παλινδρόμηση

ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ

	β.ε.	*SS*	*MS*	*F*	*p*
Παλινδρόμηση	1	25.686	25.686	69.170	0.004
Σφάλματα	3	1.114	0.371		
Σύνολο	4	26.800			

H_0: b=0 (στον πληθυσμό)

H_1: b≠0

Το υπόδειγμα της Γραμμικής Παλινδρόμησης είναι στατιστικά σημαντικό σε επίπεδο σημαντικότητας *α*=0,05

Συντελεστής Συσχέτισης

β. Συντελεστής Συσχέτισης

$$r = \frac{\sum\left[\left(Y_i - \bar{Y}\right)\left(X_i - \bar{X}\right)\right]}{\sqrt{\sum\left(Y_i - \bar{Y}\right)^2 \sum\left(X_i - \bar{X}\right)^2}} = 0,98$$

Εναλλακτικά

$$r = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sqrt{\left[\sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}\right]\left[\sum Y_i^2 - \frac{\left(\sum Y_i\right)^2}{n}\right]}} = 0,98$$

Η τιμή του r δηλώνει ότι υπάρχει έντονη θετική συσχέτιση μεταξύ του χρόνου δίαιτας ενός ατόμου και της απώλειας βάρους του.

Συντελεστής Προσδιορισμού

γ. Συντελεστής Προσδιορισμού

$$R^2 = \frac{\left(\sum\left[\left(Y_i - \bar{Y}\right)\left(X_i - \bar{X}\right)\right]\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2 \sum\left(X_i - \bar{X}\right)^2} = 0,96$$

Εναλλακτικά

$$R^2 = \frac{\left(\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}\right)^2}{\left[\sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}\right]\left[\sum Y_i^2 - \frac{\left(\sum Y_i\right)^2}{n}\right]} = 0,96$$

Η τιμή του R^2 δείχνει ότι 96% της μεταβλητότητας της απώλειας βάρους ερμηνεύεται από τη χρονική διάρκεια της δίαιτας.

Εκτίμηση

δ. Εκτίμηση της τιμής του *Y* για *X*=4

Η ευθεία παλινδρόμησης είναι:

$$\hat{Y} = -0,89 + 1,06 * X$$

Για $X = 4$ έχουμε $\hat{Y} = -0,89 + 1,06 * 4 \Rightarrow \hat{Y} = 3,35$

Άρα η απώλεια βάρους μετά από 4 εβδομάδες δίαιτας θα είναι 3,35 κιλά.

Αποτελέσματα με το SPSS (1)

Descriptive Statistics

	Mean	Std. Deviation	N
Απώλεια Βάρους (κιλά)	4.2	2.6	5
Αρ. Εβδομάδων	4.8	2.4	5

Correlations

		Απώλεια Βάρους (κιλά)	Αρ. Εβδομάδων
Pearson Correlation	Απώλεια Βάρους (κιλά)	1.000	.979
	Αρ. Εβδομάδων	.979	1.000
Sig. (1-tailed)	Απώλεια Βάρους (κιλά)	.	.002
	Αρ. Εβδομάδων	.002	.
N	Απώλεια Βάρους (κιλά)	5	5
	Αρ. Εβδομάδων	5	5

Αποτελέσματα με το SPSS (2)

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.979[a]	.958	.945	.61

a. Predictors: (Constant), Αρ. Εβδομάδων

b. Dependent Variable: Απώλεια Βάρους (κιλά)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	25.686	1	25.686	69.170	.004[a]
	Residual	1.114	3	.371		
	Total	26.800	4			

a. Predictors: (Constant), Αρ. Εβδομάδων

b. Dependent Variable: Απώλεια Βάρους (κιλά)

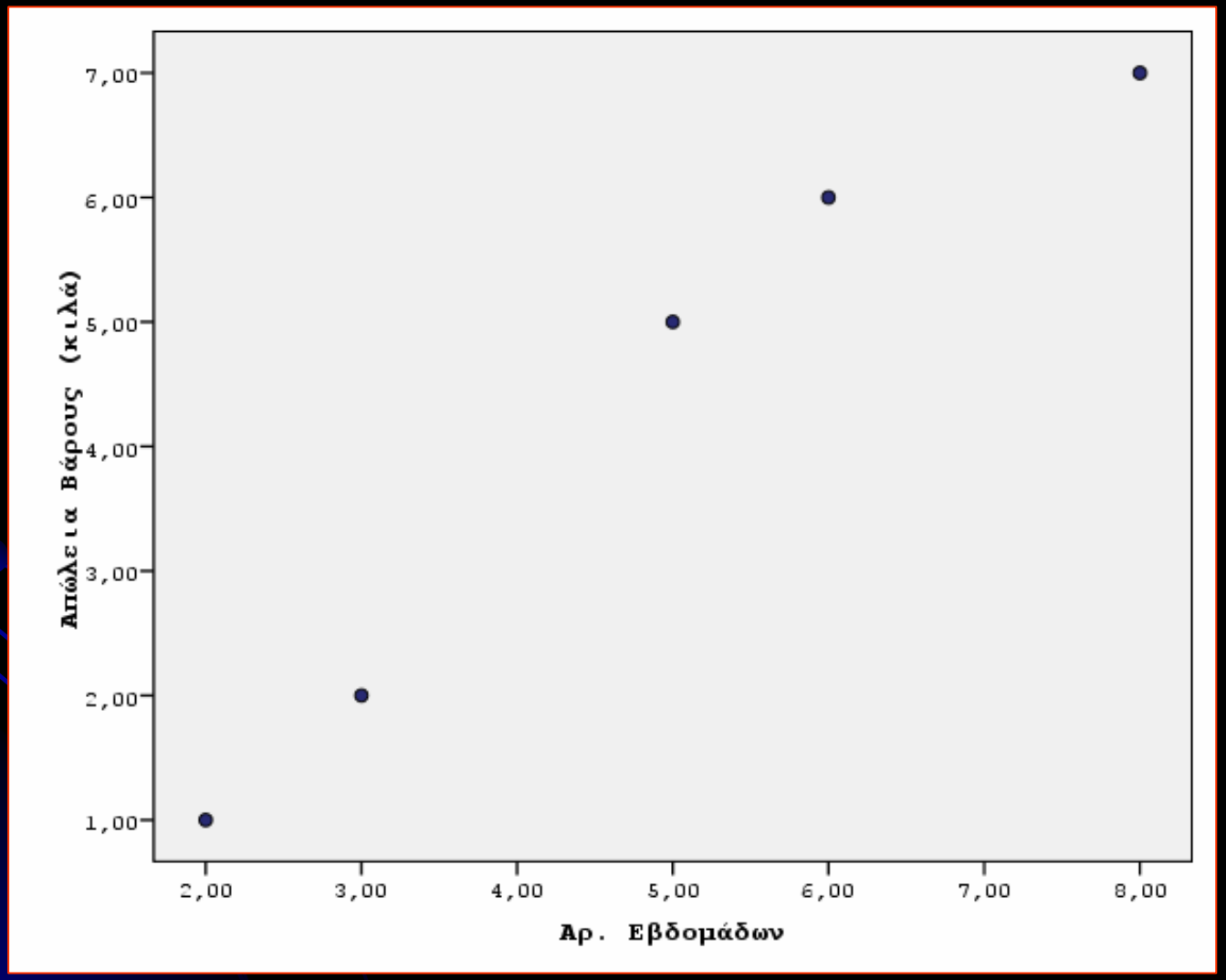
Αποτελέσματα με το SPSS (3)

Coefficients[a]

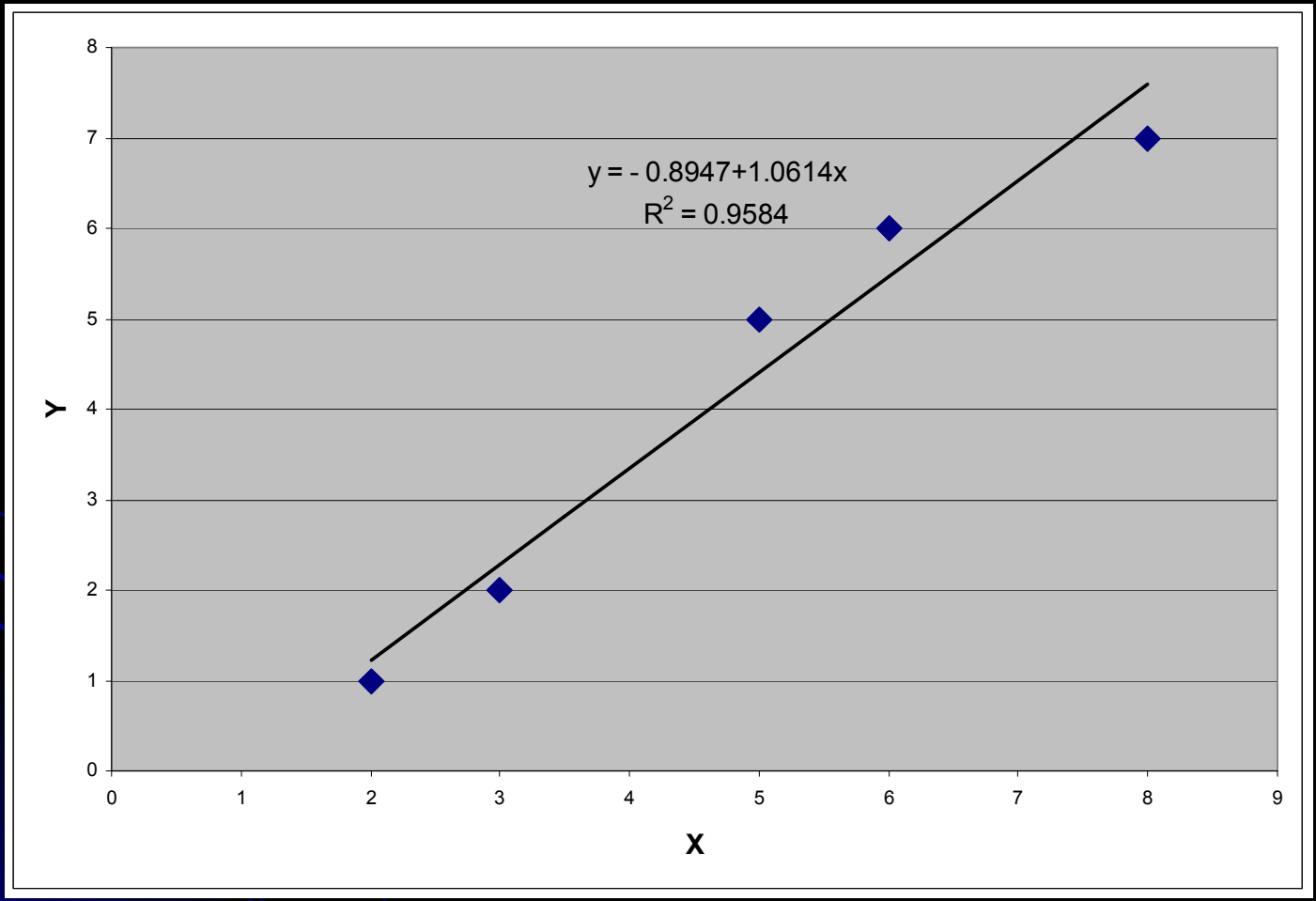
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	5% Confidence Interval for	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-.895	.670		-1.335	.274	-3.028	1.239
	Αρ. Εβδομάδα	1.061	.128	.979	8.317	.004	.655	1.468

a. Dependent Variable: Απώλεια Βάρους (κιλά)

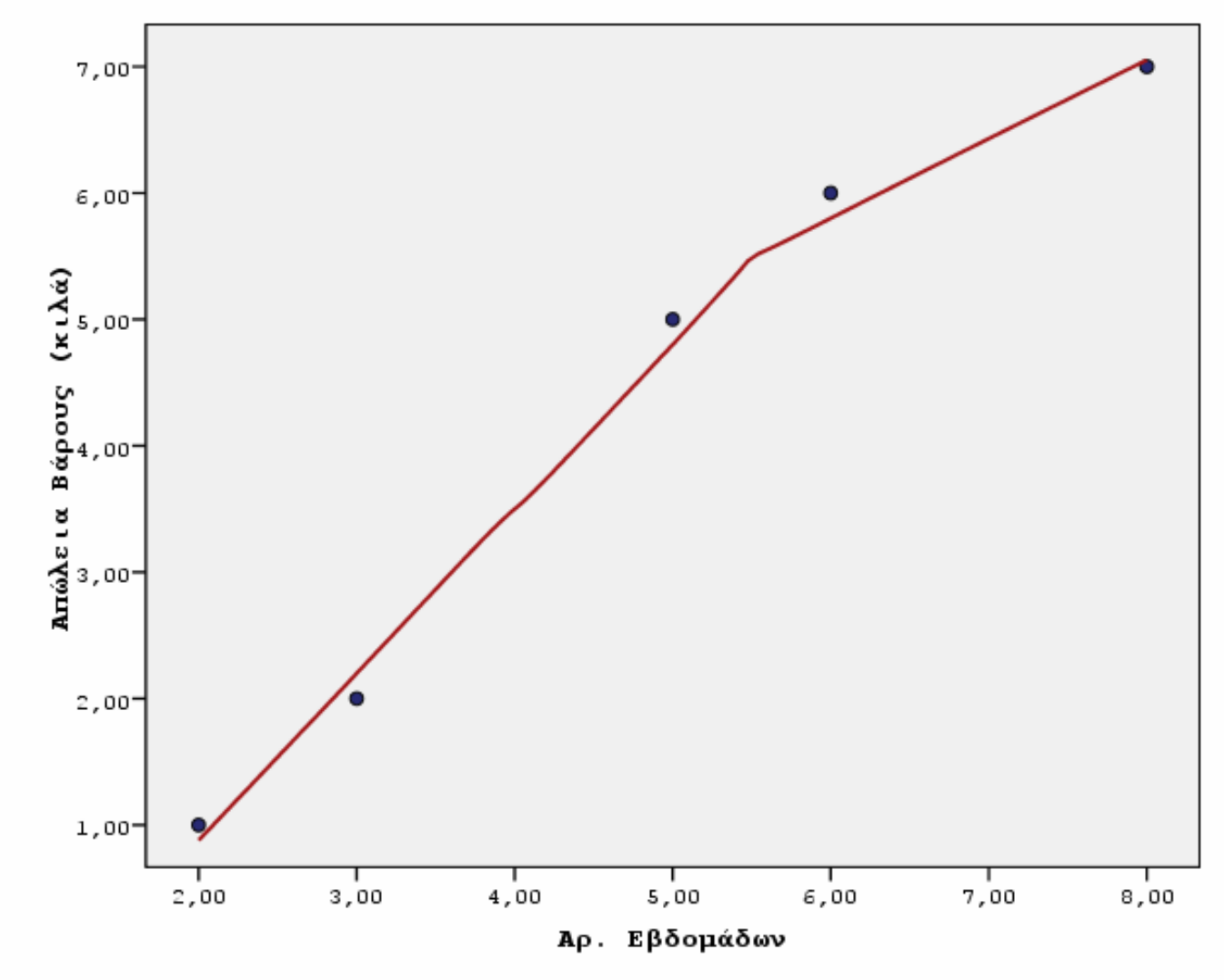
Διάγραμμα Διασποράς (*Scatter Plot*)



Ευθεία Ελαχίστων Τετραγώνων (*Least Squares Regression Line*)



Βέλτιστη Καμπύλη (μέθοδος *Loess*)



Γραμμική Παλινδρόμηση και Συσχέτιση (Ι)

Για το μοντέλο $\hat{y} = \hat{\alpha} + \hat{\beta}x$ ισχύουν:

$$\hat{\beta} = \frac{s_{xy}}{s_x^2}, \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

$$s^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{n-1}{n-2}\left(s_y^2 - \frac{s_{xy}^2}{s_x^2}\right) = \frac{n-1}{n-2}\left(s_y^2 - \hat{\beta}^2 s_x^2\right)$$

$$s_\alpha^2 = s^2\left(\frac{1}{n} + \frac{\bar{x}^2}{(n-1)s_x^2}\right), \quad s_\beta^2 = \frac{s^2}{(n-1)s_x^2}$$

100(1–a)% δ.ε. για το α: $(\hat{\alpha} \pm s_\alpha \cdot t_{n-2;\, a/2})$

100(1–a)% δ.ε. για το β: $(\hat{\beta} \pm s_\beta \cdot t_{n-2;\, a/2})$

100(1–a)% δ.ε. για το E(Y): $\left(\hat{y} \pm s \cdot t_{n-2;\, a/2}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{(n-1)s_x^2}}\right)$

100(1–a)% δ.ε. για τη διαφορά $\beta_1 - \beta_2$ $\left(\hat{\beta}_1 - \hat{\beta}_2 \pm s_0 t_{n_1+n_2-4;\, a/2}\sqrt{\frac{n_1-1}{s_{x_1}^2} + \frac{n_2-1}{s_{x_2}^2}}\right)$

όπου $s_0^2 = \dfrac{(n_1-2)\, s_{x_1}^2 + (n_2-2) s_{x_2}^2}{n_1+n_2-4}$

Γραμμική Παλινδρόμηση και Συσχέτιση (ΙΙ)

Για το γενικό γραμμικό μοντέλο $\underset{\sim}{Y} = X \cdot \underset{\sim}{\beta} + e$ ισχύουν:

$$\hat{\underset{\sim}{\beta}} = (X'X)^{-1} X' \underset{\sim}{Y} \quad \text{με} \quad s^2 = \frac{\underset{\sim}{Y}'\underset{\sim}{Y} - \hat{\underset{\sim}{\beta}}'XX}{n-(\varkappa+1)}$$

Συντελεστής συσχέτισης: $\varrho = \dfrac{\mathrm{cov}(x,y)}{\sqrt{\mathrm{var}x \cdot \mathrm{var}y}}$, $\quad r = \dfrac{s_{xy}}{s_x s_y}$

$$\left.\begin{array}{l} H_0: \varrho = 0 \\ H_1: \varrho \neq 0 \end{array}\right\} \Rightarrow R = \left\{ |t| = \frac{|r\sqrt{n-2}|}{\sqrt{1-r^2}} > t_{n-2;\, a/2} \right\}$$

Συντελεστής μερικής συσχέτισης: $r_{y1,2} = \dfrac{r_{y1} - r_{y2}\, r_{12}}{\sqrt{(1-r_{y2})^2 (1-r_{12}^2)}}$

$$r_{y2,1} = \frac{r_{y2} - r_{y1}\, r_{12}}{\sqrt{(1-r_{y1})^2 (1-r_{12})^2}}$$

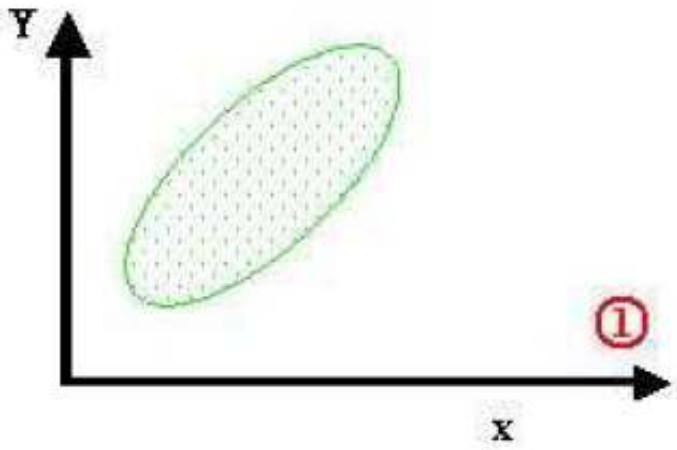
Πίνακας Ανάλυσης Διακύμανσης μετά την Παλινδρόμηση

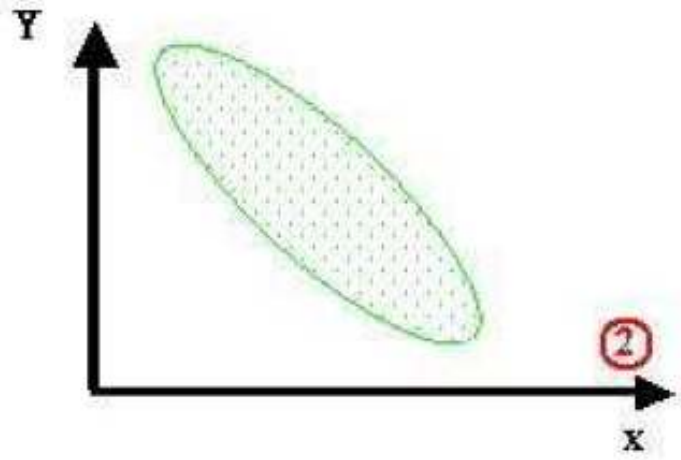
Πηγές Διακύμανσης	Αθροίσματα Τετραγώνων	Βαθμοί Ελευθερίας	Μέσα Αθροίσματα Τετραγώνων	F
Παλιδρόμησης	$\sum_{t=1}^{n}(\hat{Y}_t-\bar{Y})^2$	1	$\frac{\sum_{t=1}^{n}(\hat{Y}_t-\bar{Y})^2}{1}$	
Καταλοίπων	$\sum_{t=1}^{n}(Y_t-\hat{Y}_t)^2$	n - 2	$\frac{\sum_{t=1}^{n}(Y_t-\hat{Y}_t)^2}{n-2}$	$F=\frac{\frac{\sum_{t=1}^{n}(\hat{Y}_t-\bar{Y})^2}{1}}{\frac{\sum_{t=1}^{n}(Y_t-\hat{Y}_t)^2}{n-2}}$
Σύνολο	$\sum_{t=1}^{n}(Y_t-\bar{Y})^2$	n - 1		

Παραδείγματα (1)

ΔΙΑΓΡΑΜΜΑΤΑ ΔΙΑΣΠΟΡΑΣ

Η γραφική παράσταση των δειγματικών σημείων στο σύστημα καρτεσιανών συντεταγμένων ονομάζεται διάγραμμα διασποράς (scatter diagram) και δίνει χρήσιμες πληροφορίες για την εξειδίκευση του μοντέλου





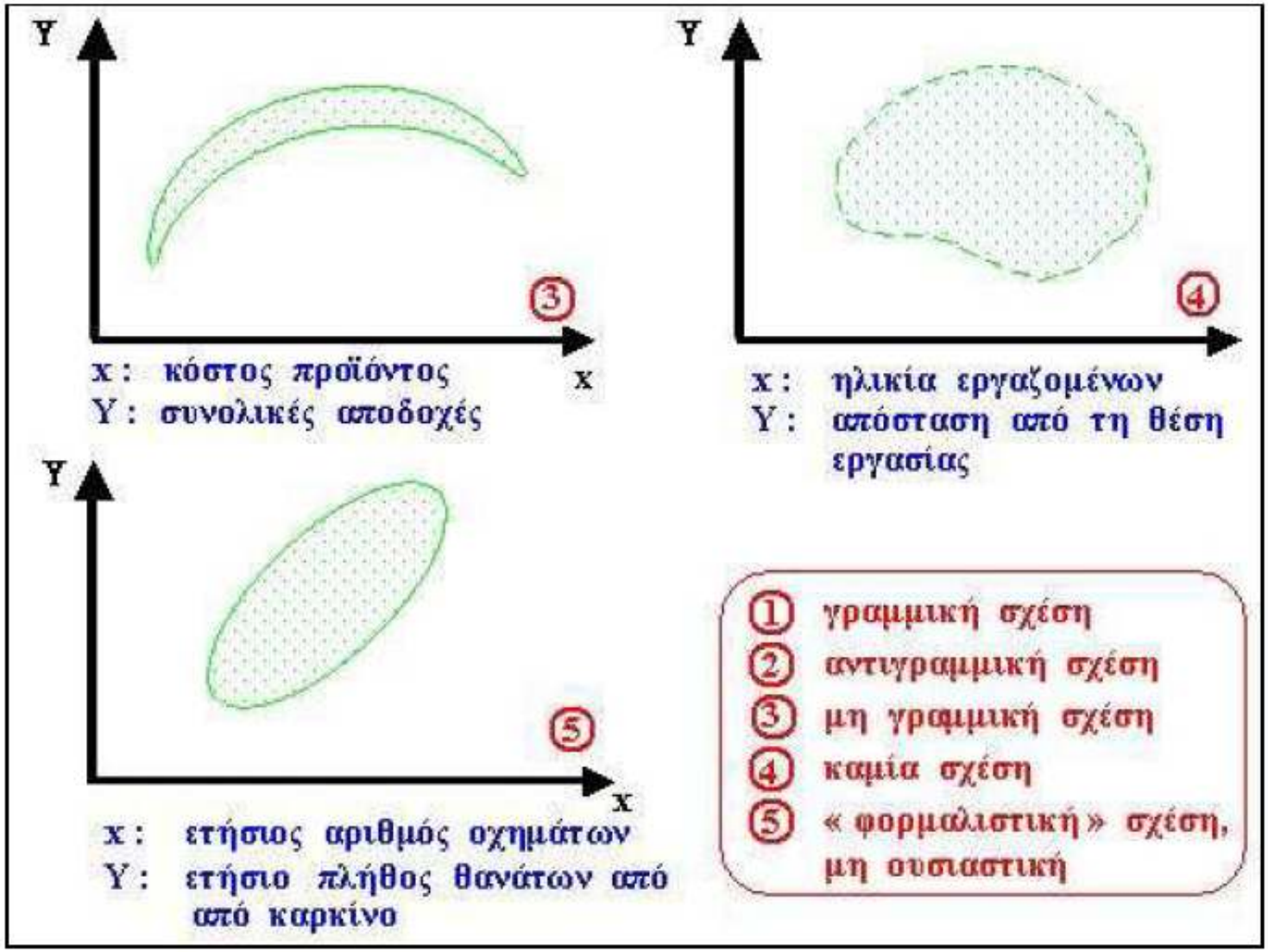
x : ετήσιο εισόδημα οικογένειας

Y : ετήσια έξοδα διατροφής

x : καταπόνηση υλικού

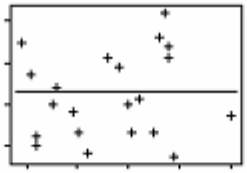
Y : ωφέλιμη διάρκεια ζωής

Παραδείγματα (2)

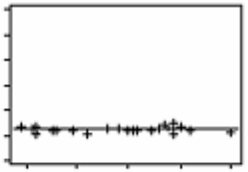


Παραδείγματα 3

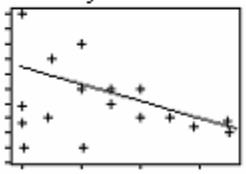
(i) No association strong relationship R^2 zero, misleading



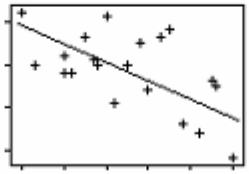
(ii) No relationship at all R^2 usefully zero



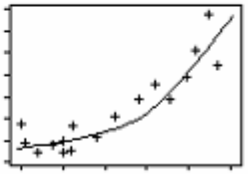
(iii) The same as (ii) on a different scale



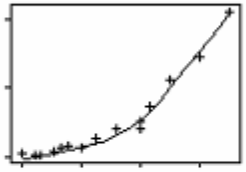
(iv) High y values only when x is small, R^2=24%



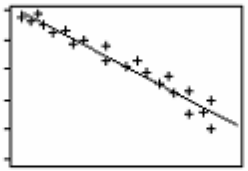
(v) Moderate linear association, R^2=40%



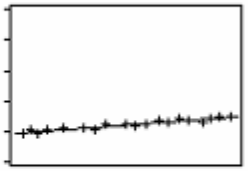
(vi) Moderate association, R^2=50% but misleading



(vii) Strong curvilinear relation, R^2=87% misleading

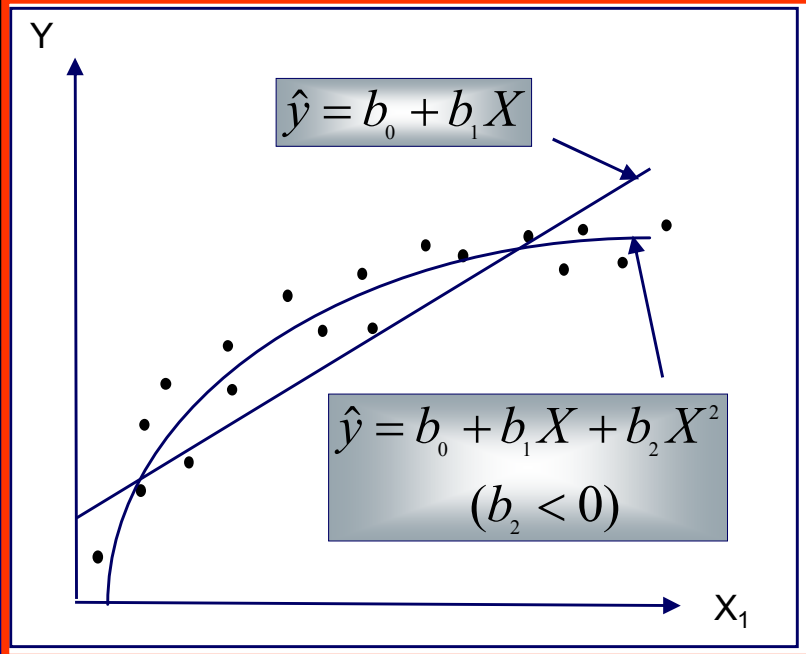


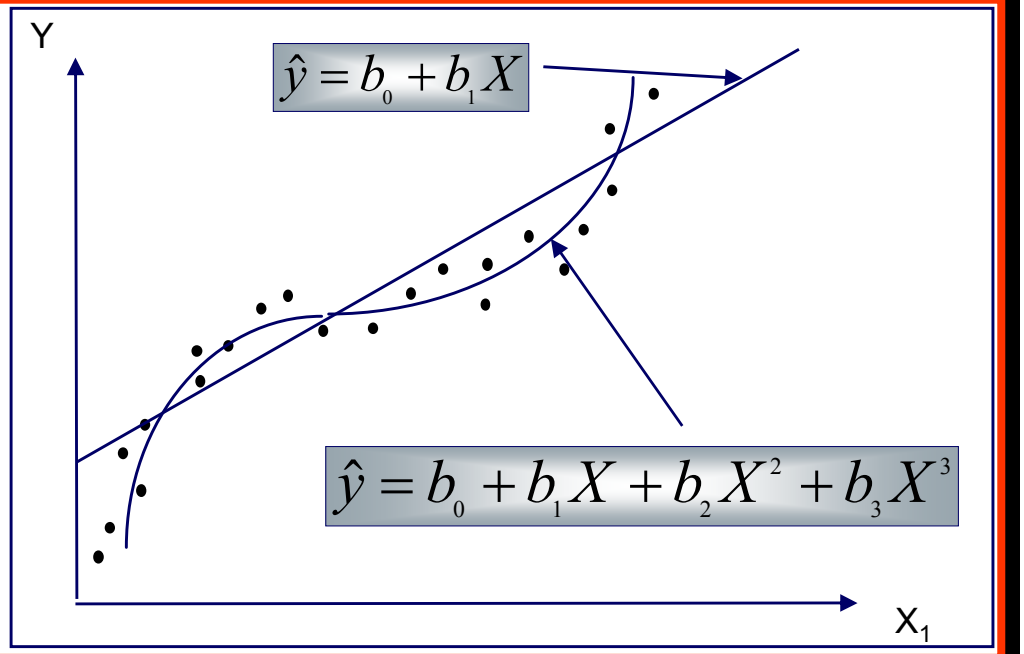
(viii) A strong relation and association R^2=99%



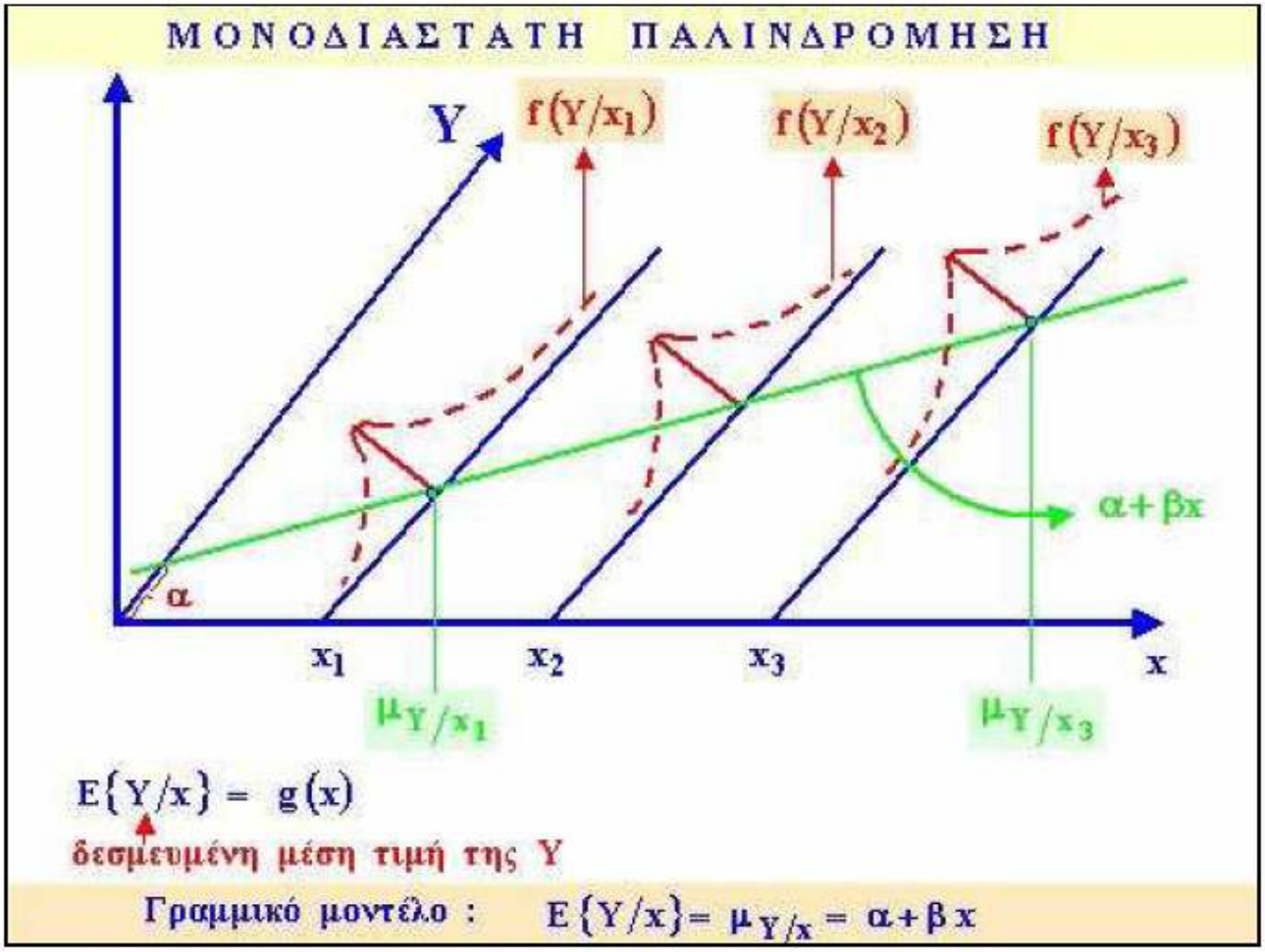
(ix) Similar to (viii) but on a compressed scale and sloping up R^2=99%

Γραμμική και Μη Γραμμική Παλινδρόμηση

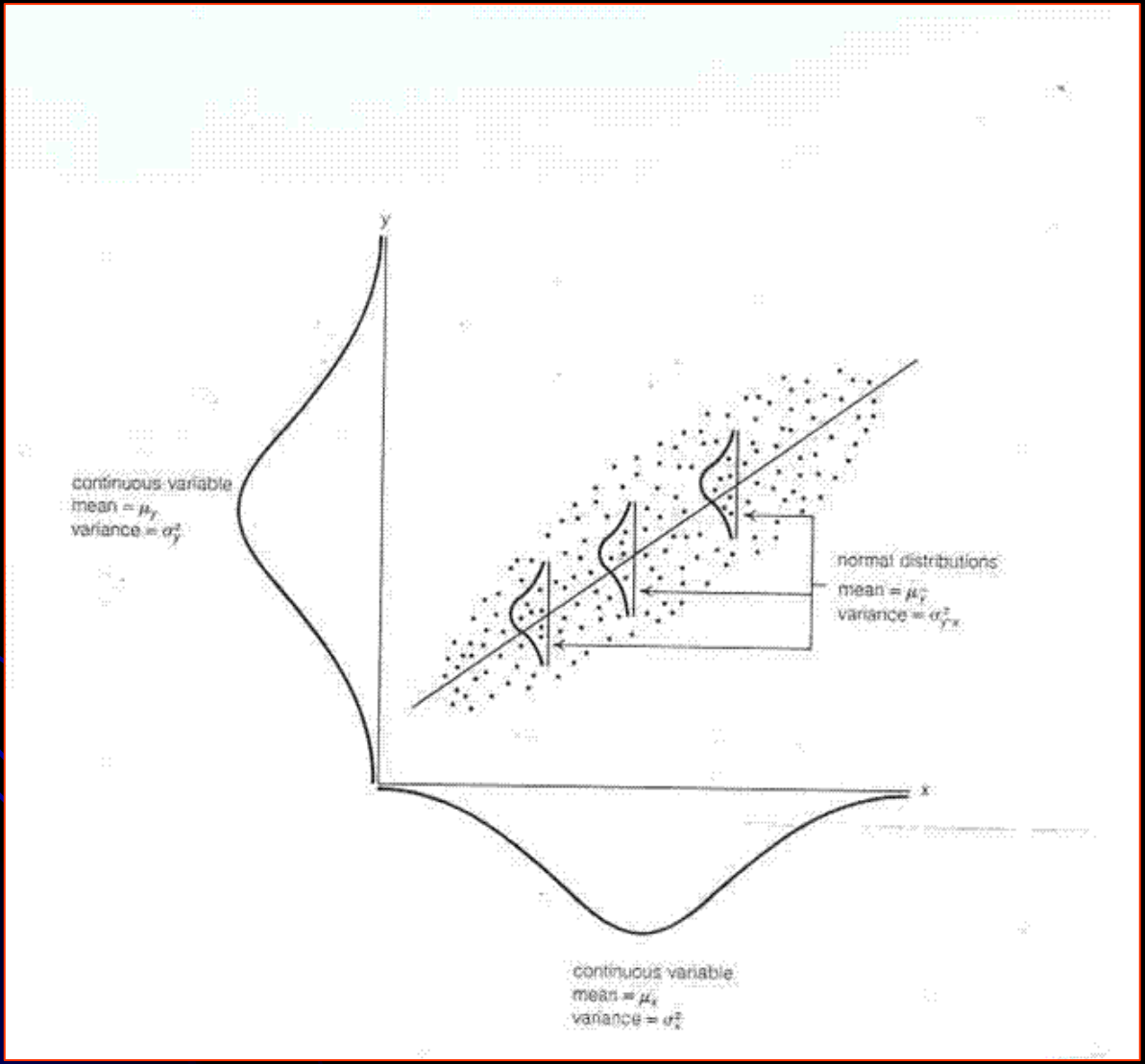




Βασική Υπόθεση (Ι)

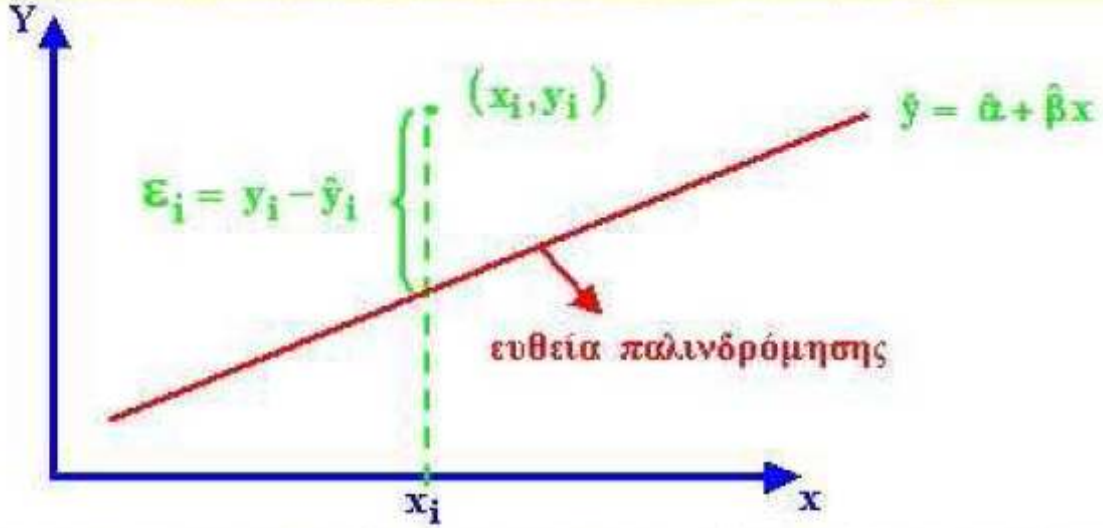


Βασική Υπόθεση (ΙΙ)



Μηχανισμός

ΕΚΤΙΜΗΣΗ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ α και β



Η εκτίμηση των α και β γίνεται με το σκεπτικό της ελαχιστοποίησης των τετραγώνων των αποκλίσεων $\varepsilon_i = y_i - \hat{y}_i$ των παρατηρήσεων από την ευθεία

(Μέθοδος ελαχίστων τετραγώνων)

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \min$$

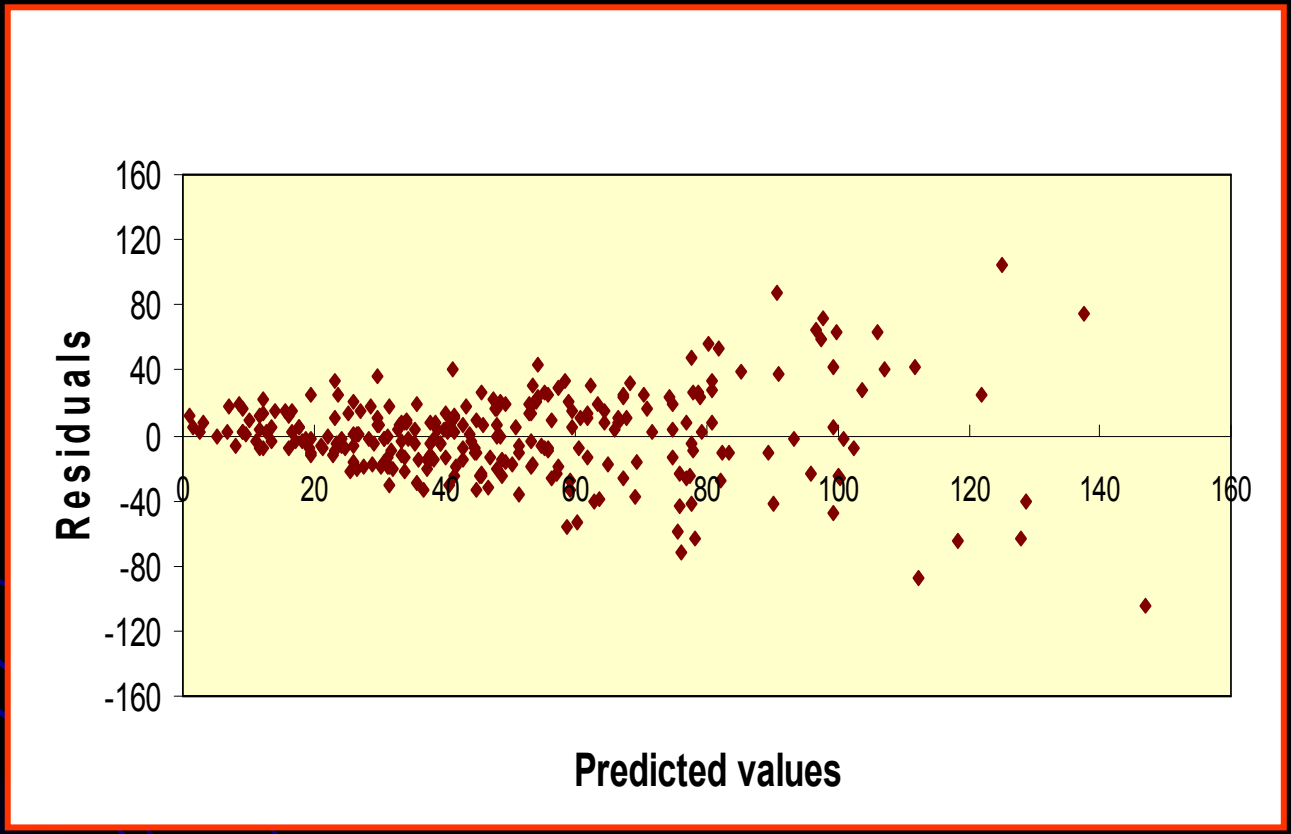
παρατήρηση

σημείο στην ευθεία παλινδρόμησης

Βασικές Προϋποθέσεις

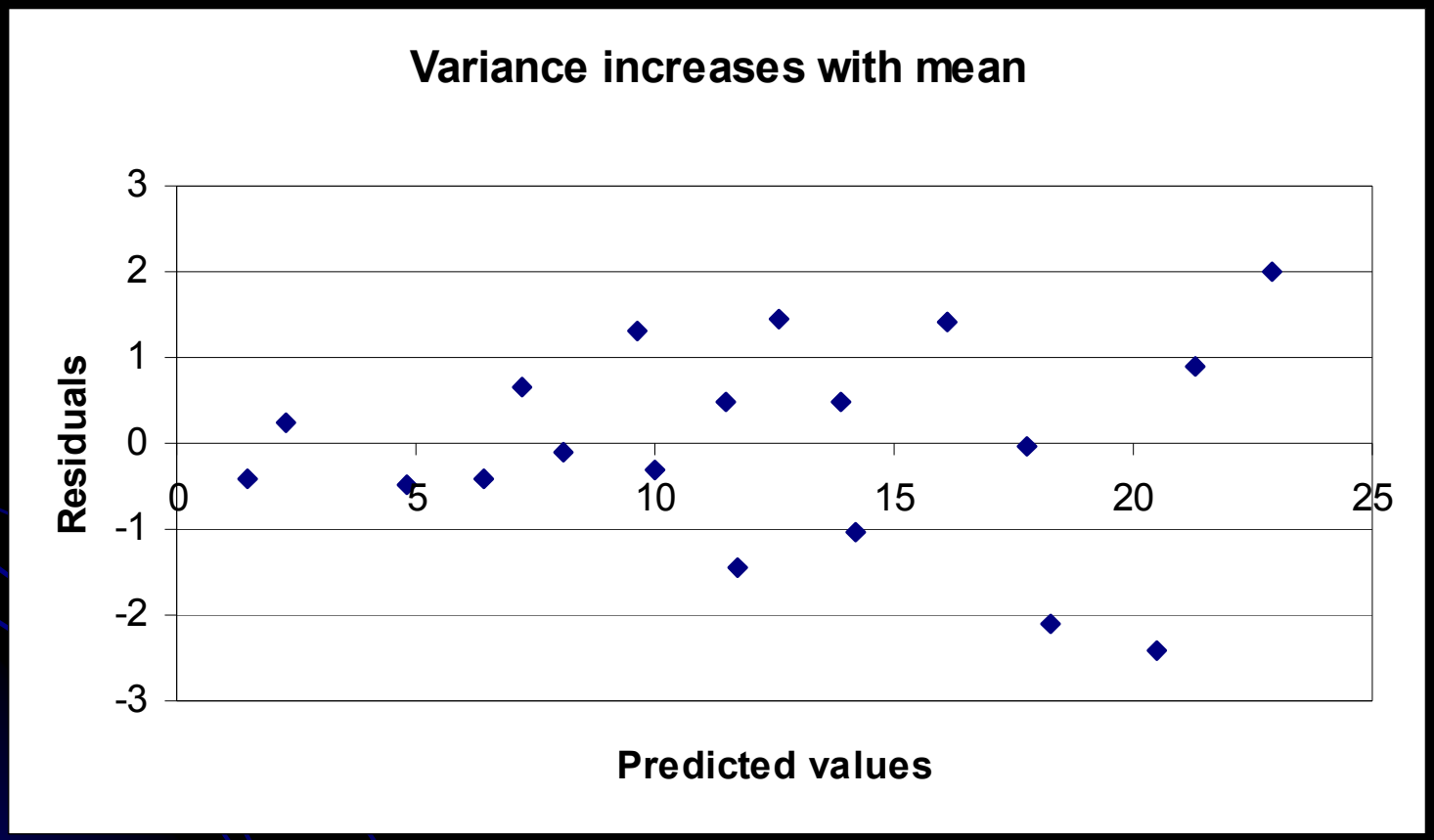
- Τα σφάλματα ακολουθούν Κανονική Κατανομή
- Ισχύει η Ομοσκεδαστικότητα των Σφαλμάτων
- Δεν υπάρχουν παράτυπα σημεία (outliers) δηλ. τιμές με z-score σε απόλυτη τιμή >3.
- Δεν υπάρχουν σημεία μόχλευσης.

Αποκλίσεις από την Ομοσκεδαστικότητα (1)

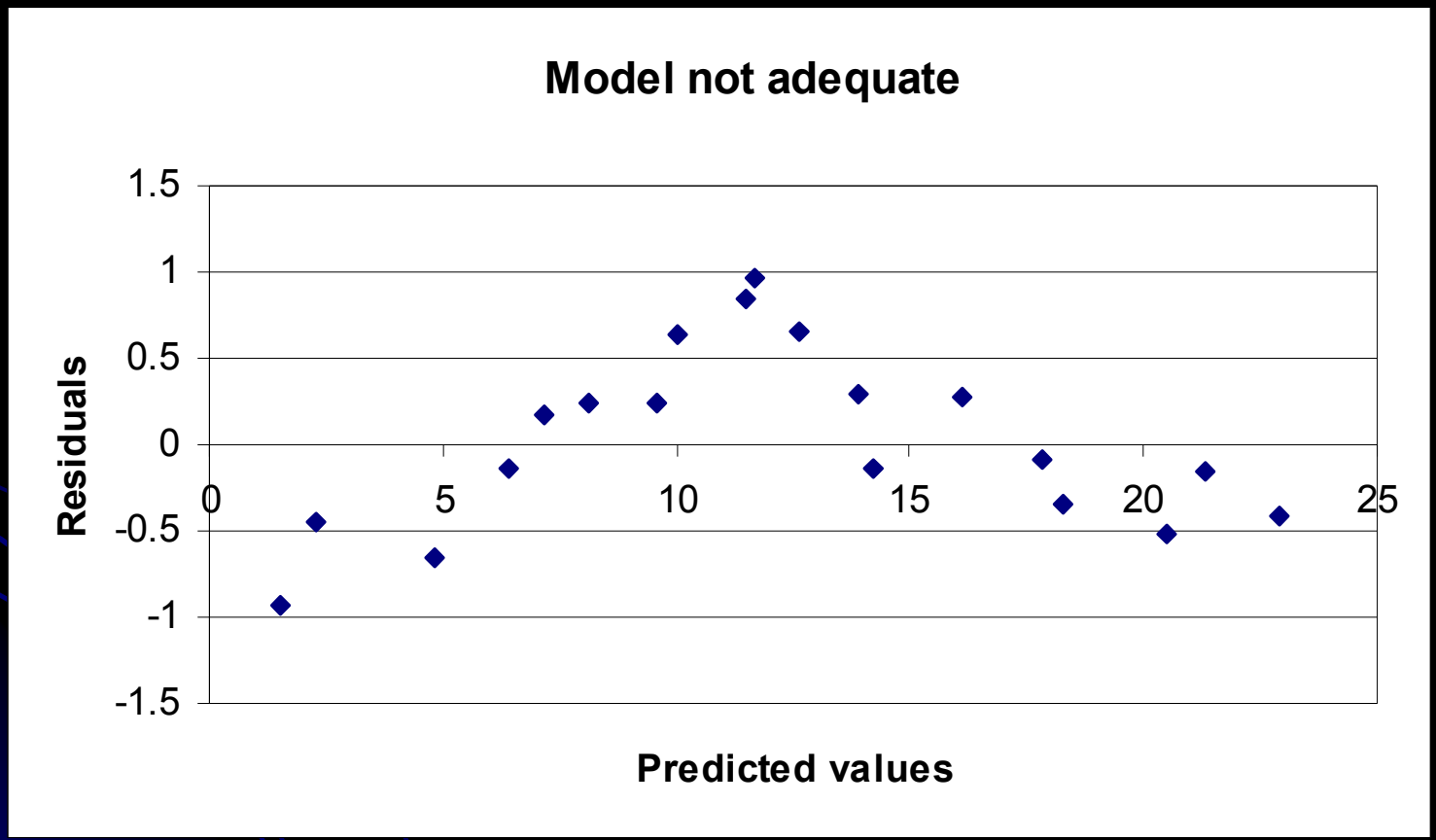


Για μεγάλες τιμές της εξαρτημένης μεταβλητής μεγαλώνει και η διασπορά των σφαλμάτων

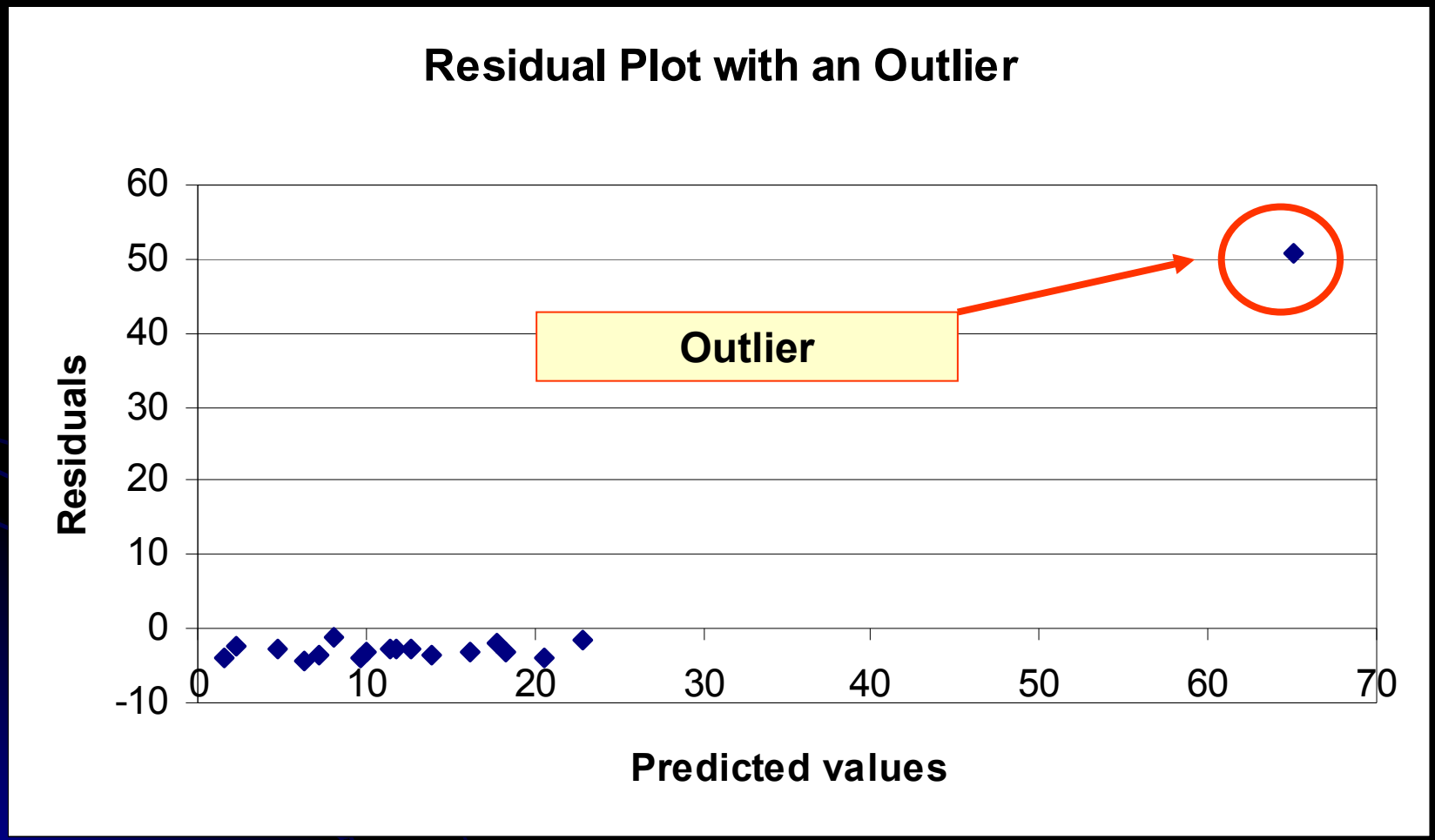
Αποκλίσεις από την Ομοσκεδαστικότητα (2)



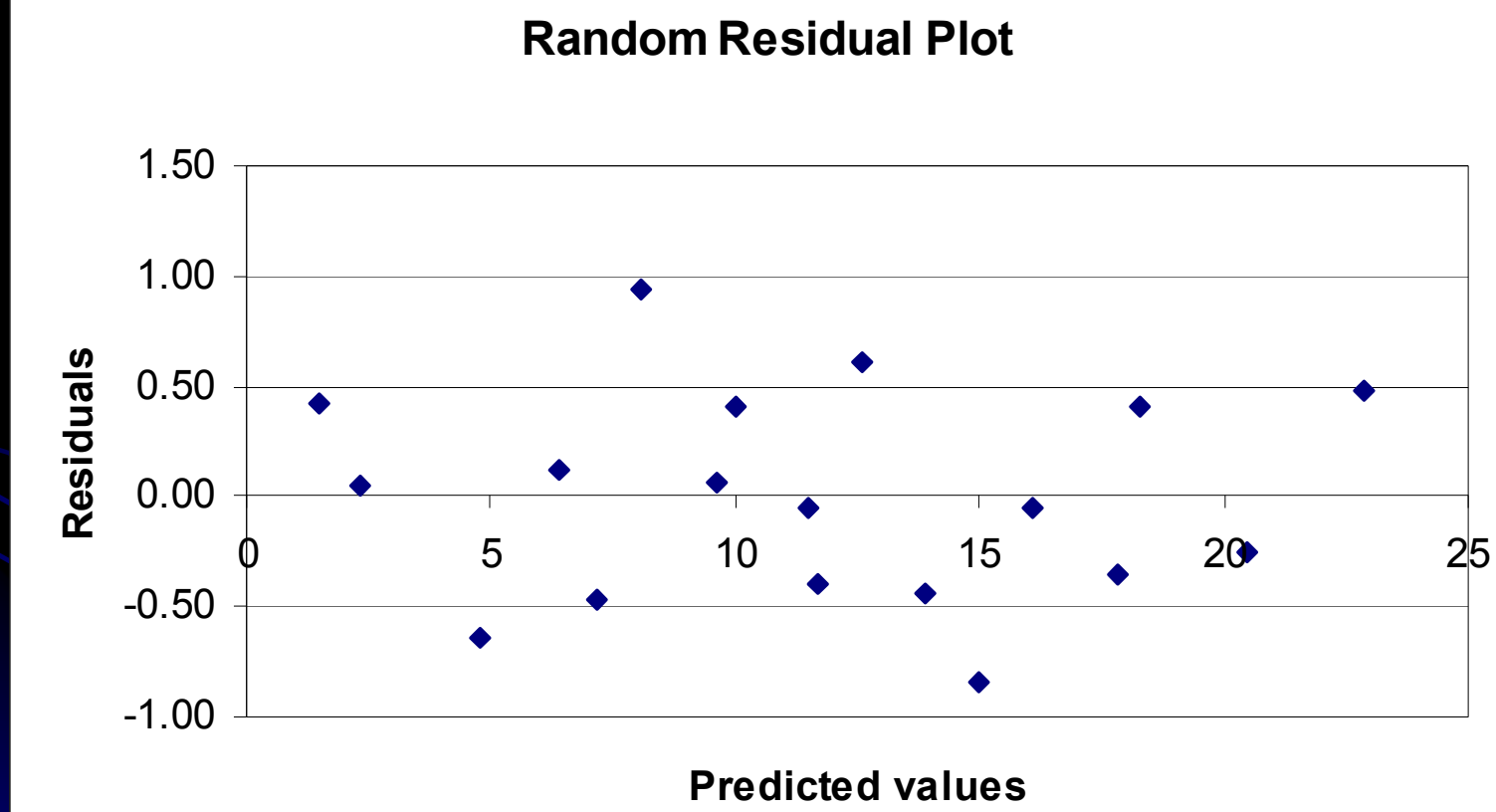
Αποκλίσεις από την Ομοσκεδαστικότητα (3)



Αποκλίσεις από την Ομοσκεδαστικότητα (4)



Ομοσκεδαστικότητα των Σφαλμάτων



Σύνδεση της Απλής Ευθύγραμμης Συμμεταβολής με την Ανάλυση Παραλλακτικότητας (1)

Στον πίνακα δίνονται οι μέσοι όροι ενός βιολογικού χαρακτηριστικού για 6 επεμβάσεις (περιεκτικότητα % του εδάφους σε θρεπτικά συστατικά) από ένα RCBD

Report

y

Treatmt	Mean	Std. Deviation	N
0%	4.75	1.500	4
5%	5.75	.957	4
10%	5.75	.957	4
15%	6.70	.258	4
20%	7.00	.860	4
25%	6.88	.750	4
Total	6.14	1.174	24

Σύνδεση της Απλής Ευθύγραμμης Συμμεταβολής με την Ανάλυση Παραλλακτικότητας (2)

Ο έλεγχος του Tukey (HSD test) έδειξε ότι οι 6 μέσοι όροι δεν διαφέρουν στατιστικά σημαντικά σε *α*=0,05

y

Tukey HSD[a,b]

Treatmt	N	Subset 1
0%	4	4.75
5%	4	5.75
10%	4	5.75
15%	4	6.70
25%	4	6.88
20%	4	7.00
Sig.		.061

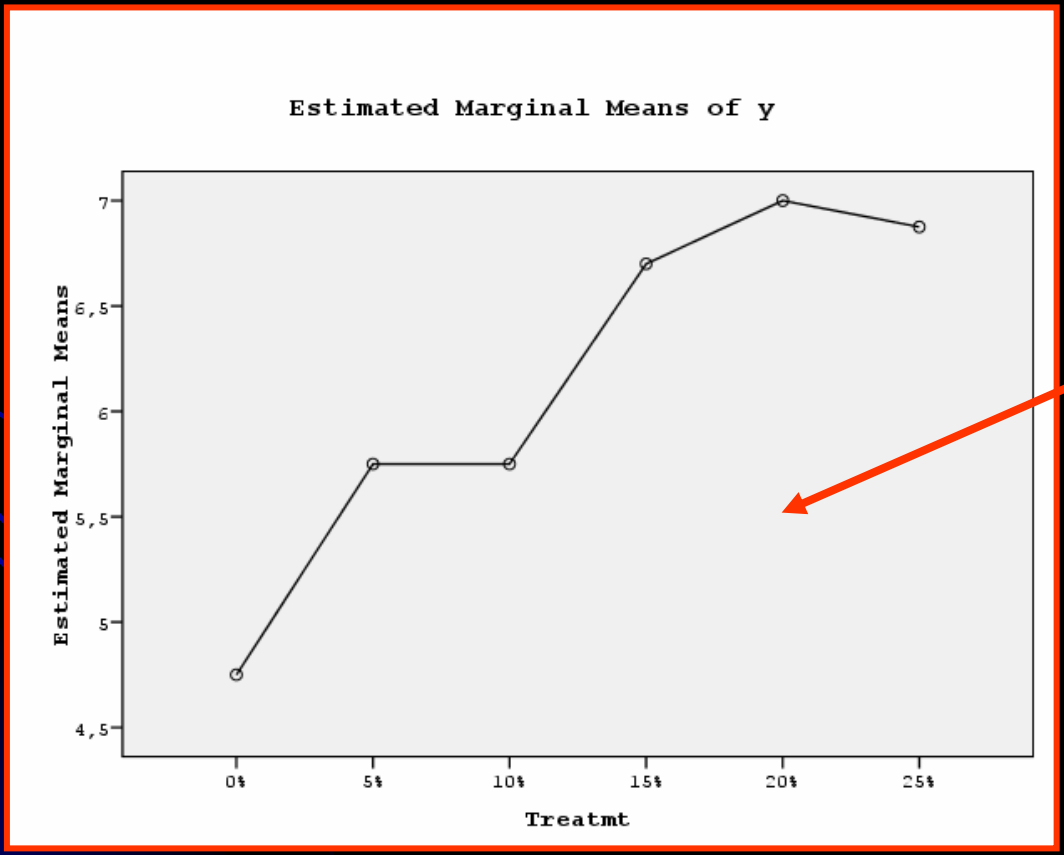
Means for groups in homogeneous subsets are displayed.
Based on Type III Sum of Squares
The error term is Mean Square(Error) = 1.026.

a. Uses Harmonic Mean Sample Size = 4.000.

b. Alpha = .05.

Σύνδεση της Απλής Ευθύγραμμης Συμμεταβολής με την Ανάλυση Παραλλακτικότητας (3)

Τίθεται το ερώτημα, εάν υπάρχει στατιστικά σημαντική τάση-συμμεταβολή (θετική, ανοδική) των μέσων όρων του βιολογικού χαρακτηριστικού και της περιεκτικότητας του εδάφους σε θρεπτικά συστατικά



Διάγραμμα Μέσων Όρων

Σύνδεση της Απλής Ευθύγραμμης Συμμεταβολής με την Ανάλυση Παραλλακτικότητας (4)

Η εφαρμογή της μεθόδου της Απλής Ευθύγραμμης Συμμεταβολής στα 6 ζεύγη σημείων έδωσε τα παρακάτω αποτελέσματα:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.937[a]	.877	.847	.34316

a. Predictors: (Constant), Treat

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.042	.248		20.300	.000
	Treat	8.771	1.641	.937	5.346	.006

a. Dependent Variable: Y

Σύνδεση της Απλής Ευθύγραμμης Συμμεταβολής με την Ανάλυση Παραλλακτικότητας (5)

- Συμπέρασμα:

  Ενώ δεν υπάρχουν στατιστικά σημαντικές διαφορές μεταξύ των μέσων όρων των 6 επεμβάσεων ωστόσο υπάρχει στατιστικά σημαντική συμμεταβολή (θετική, ανοδική τάση) (R=0,937, p=0,006) μεταξύ των 6 επιπέδων περιεκτικότητας του εδάφους σε θρεπτικά συστατικά και των αντίστοιχων μέσων όρων του βιολογικού χαρακτηριστικού

Βιβλιογραφία

- Ελληνικό Ανοικτό Πανεπιστήμιο (2003). *Σημειώσεις Στατιστικής*.
- Φασούλας, Α. Κ. (ανατ. 2008). *Στοιχεία Πειραματικής Στατιστικής*. Θεσσαλονίκη: Άγις-Σάββας Δ. Γαρταγάνης.
- Steel, R. & Torrie, J. (1986). *Principles and Procedures of Statistics: A Biometrical Approach*. Singapore: McGraw-Hill Book Company.
- Gomez, K. & Gomez, A. (1984). *Statistical Procedures for Agricultural Research*. Singapore: John Willey & Sons, Inc.
- Kuehl, R. (2000). *Designs of Experiments: Statistical Principles of Research Design and Analysis*. Pacific Grove: Duxbury Thomson Learning.
- Jacoby, W. (2000). Loess: a nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies*, 19, 577-613.
- Κολυβά, Φ. και Μπόρα-Σέντα, Ε. (1995). *Στατιστική: Θεωρία-Εφαρμογές*. Θεσσαλονίκη: Εκδόσεις ΖΗΤΗ.

Σας ευχαριστώ για την προσοχή σας
Viola adorata