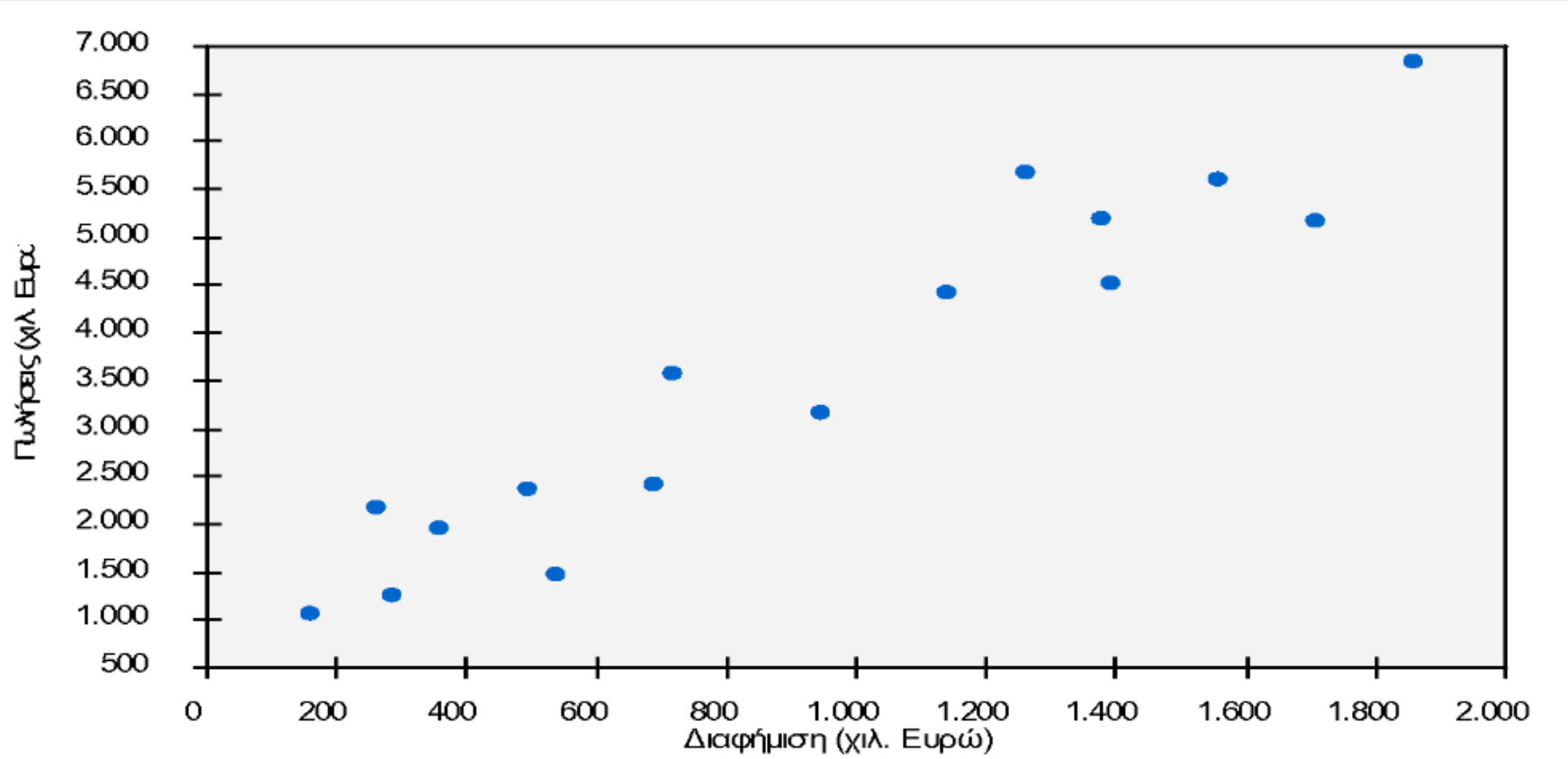
Απλή Παλινδρόμηση και Συσχέτιση

Πωλήσεις, Δαπάνες Διαφήμισης και Αριθμός Πωλητών

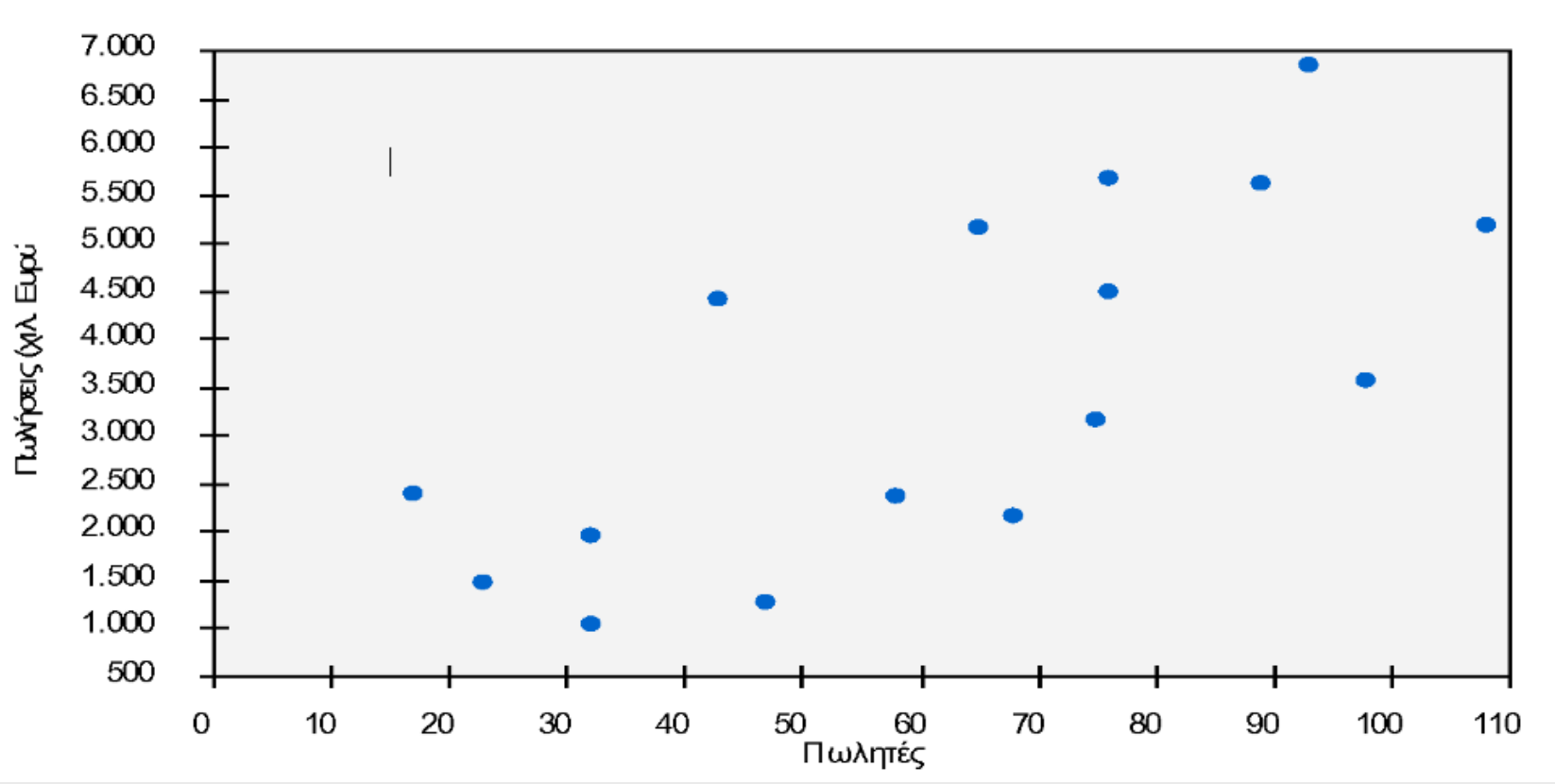
Έτος	Πωλήσεις (χιλ €)	Διαφήμιση (χιλ €)	Πωλητές (Άτομα)
1981	1050	162	32
1982	1260	285	47
1983	1470	540	23
1984	2160	261	68
1985	1950	360	32
1986	2400	690	17
1987	2370	495	58
1988	3150	948	75

Έτος	Πωλήσεις (χιλ €)	Διαφήμιση (χιλ €)	Πωλητές (Άτομα)
1989	3570	720	98
1990	4410	1140	43
1991	4500	1395	76
1992	5610	1560	89
1993	5190	1380	108
1994	5670	1260	76
1995	5160	1710	65
1996	6840	1860	93

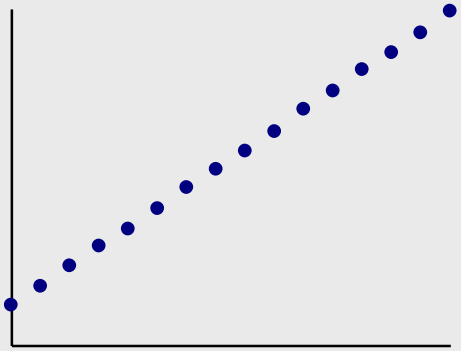
Διάγραμμα Διασποράς Μεταξύ Πωλήσεων και Διαφήμισης



Διάγραμμα Διασποράς Μεταξύ Πωλήσεων και Αριθμού Πωλητών

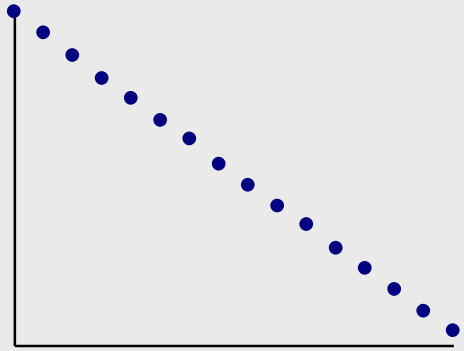


Διαγράμματα Διασποράς



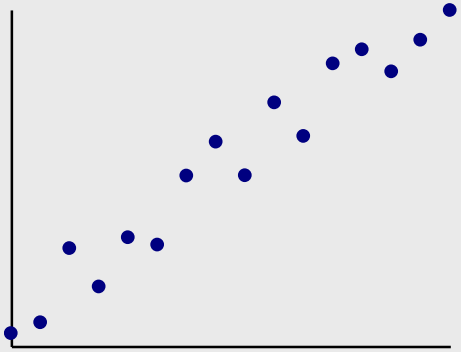
Τέλεια θετική συσχέτιση

(r = +1,0)



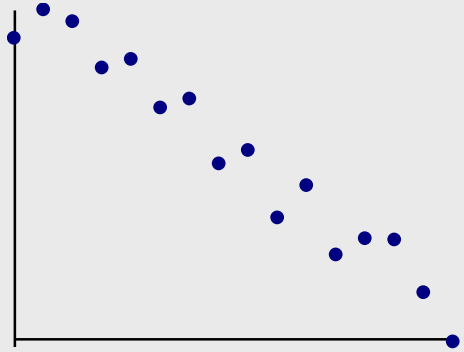
Τέλεια αρνητική συσχέτιση

(r = -1,0)



Έντονη θετική συσχέτιση

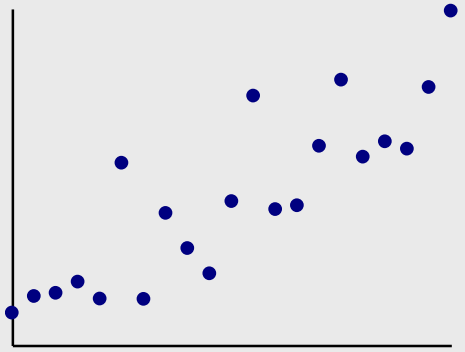
(r = +0,9)



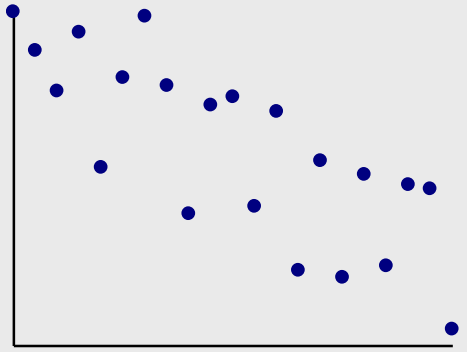
Έντονη Αρνητική Συσχέτιση

(r= - 0,9)

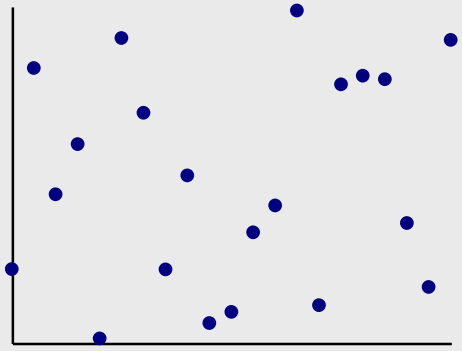
Διαγράμματα Διασποράς

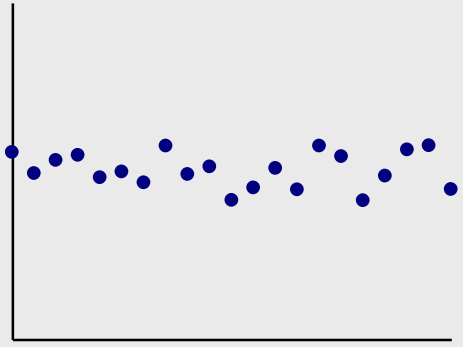


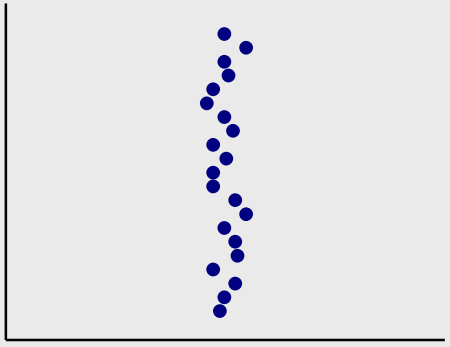
Ασθενής θετική συσχέτιση (r = +0,7)



Ασθενής αρνητική συσχέτιση (r= - 0,7)







Μηδενική συσχέτιση (r= 0)

Συμμεταβολή

Την ύπαρξη γραμμικής σχέσης μεταξύ δύο μεταβλητών, την κατεύθυνση συμμεταβολής των τιμών τους καθώς και την από κοινού διασπορά των τιμών των δύο μεταβλητών από τις αντίστοιχες μέσες τιμές τους δίνεται από ένα πολύ σημαντικό στατιστικό μέτρο που καλείται **συνδιακύμανση** (**Covariance**)

$$\sigma_{XY} = \operatorname{cov}(X, Y) = \mathrm{E}((\mathrm{X} - \mu_X)(\mathrm{Y} - \mu_Y))$$

$$s_{xy} = \frac{\sum_{i=1}^{n} (x_i - \overline{X})(y_i - \overline{Y})}{n-1}$$

Συσχέτιση

Την κατεύθυνση της συμμεταβολής των τιμών δύο μεταβλητών και τον βαθμό της γραμμικής σχέσης των μεταβλητών, δίνει ο ***συντελεστής συσχέτισης*** (***correlation coefficient***) που χρησιμοποιείται κυρίως αντί της συνδιακύμανσης

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n} (x_i - \overline{X})(y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} [(x_i - \overline{X})^2][(y_i - \overline{Y})^2]}}$$

Έλεγχος Στατιστικής σημαντικότητας του r

Λόγω το ότι η τιμή του **r** βασίζεται σε δείγμα παρατηρήσεων, υπόκειται στα σφάλματα της δειγματοληψίας. Το **r** αποτελεί μία εκτίμηση του άγνωστου συντελεστή του πληθυσμού **ρ**

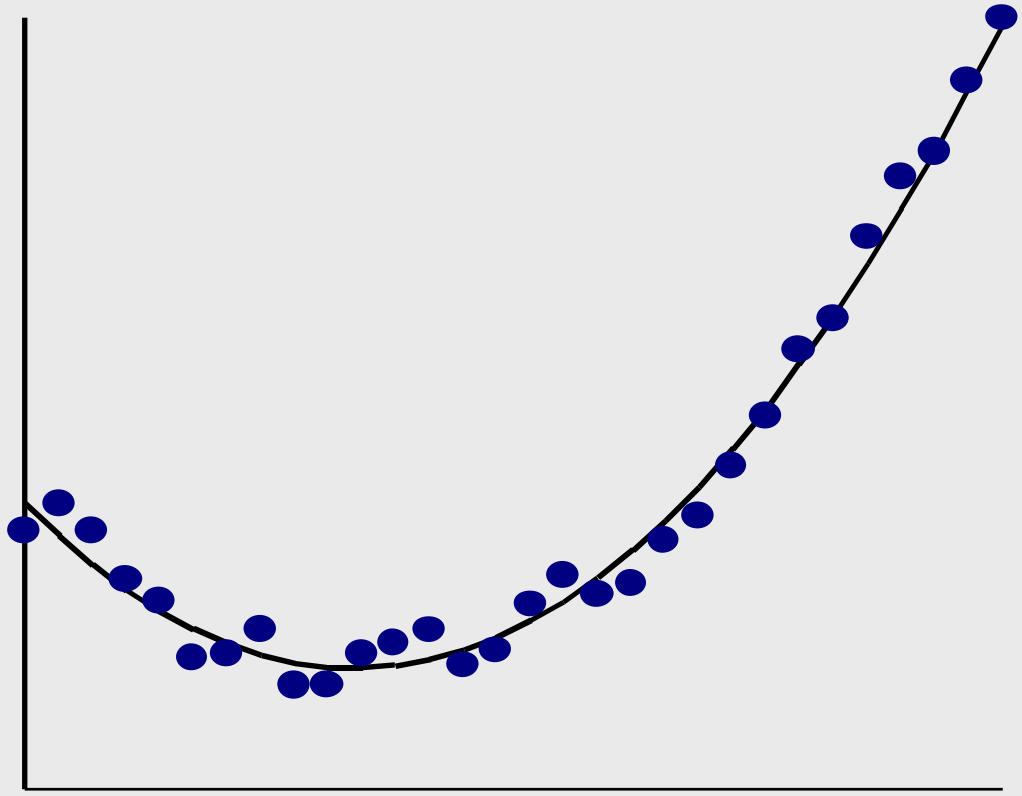
H_0: ρ=0 H_1: ρ≠0

Ο Έλεγχος γίνεται με την γνωστή κατανομή t

$$t_{n-2} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

Εάν η τιμή $|t_{n-2}|$ > της κριτικής τιμής $|t_{n-2,\alpha/2}|$ η H_0 απορρίπτεται και αντίστροφα

Διάγραμμα Διασποράς Καμπυλόγραμμης Σχέσης (Συντελεστής Γραμμικής Συσχέτισης = 0,75)



Το υπόδειγμα της Απλής Γραμμικής Παλινδρόμησης

The Simple Linear Regression Model

Εισαγωγή

Οικονομική θεωρία

Διατύπωση της αιτιώδους σχέσης μεταξύ των οικονομικών μεταβλητών

Οικονομικά μαθηματικά

Καθορισμός της μορφής της συνάρτησης που εκφράζει την συγκεκριμένη σχέση

Ανάλυση παλινδρόμησης

(α) Καθορισμός συγκεκριμένων τιμών για τις παραμέτρους των συναρτήσεων με βάση τα στοιχεία ενός δείγματος (*εκτιμητική*)

(β) Διατύπωση συμπερασμάτων για τις πραγματικές τιμές των παραμέτρων (*στατιστική επαγωγή*)

Απλή παλινδρόμηση
(simple regression)

Διερεύνηση της σχέσης δύο μεταβλητών

Πολλαπλή παλινδρόμηση
(multiple regression)

Διερεύνηση της σχέσης μεταξύ περισσότερων μεταβλητών

Σχέσεις μεταβλητών

Συναρτησιακές ή μαθηματικές
(functional relations)

Σε κάθε τιμή της X αντιστοιχεί μια τιμή της Y

Στοχαστικές ή στατιστικές
(Stochastic or statistical relations)

Σε κάθε τιμή της X αντιστοιχεί μια κατανομή τιμών της Y

*Η ανάλυση παλινδρόμησης ασχολείται με **στοχαστικές σχέσεις** και έχει σαν σκοπό να διερευνήσει την **συναρτησιακή σχέση** όχι πλέον ανάμεσα στην X και την Y αλλά ανάμεσα στην X και τον **μέσο** της κατανομής του Y για **δεδομένο** X , ανάμεσα δηλαδή στο X και στον **δεσμευμένο μέσο** της κατανομής του Y η οποία διατυπώνεται μαθηματικά σαν*

$$E(Y|X_i) = f(X_i)$$

Οταν η σχέση μεταξύ των μεταβλητών είναι ***γραμμική*** *στις* ***παραμέτρους*** *η ανάλυση παλινδρόμησης ονομάζεται* ***γραμμική*** *(linear),*
π.χ. $E(Y|X_i) = a + bX_i$ *ή* $E(Y|X_i) = a + bX_i^2$

Στην αντίθετη περίπτωση ονομάζεται ***μη γραμμική*** *(non-linear),*
π.χ. $E(Y|X_i) = a + X_i^b$

Υπόθεση

Στον πληθυσμό που μας ενδιαφέρει υπάρχει μια γραμμική σχέση ανάμεσα στις τιμές του X και στους δεσμευμένους μέσους του Y. Δηλαδή

$$E(Y|X_i) = a + bX_i$$

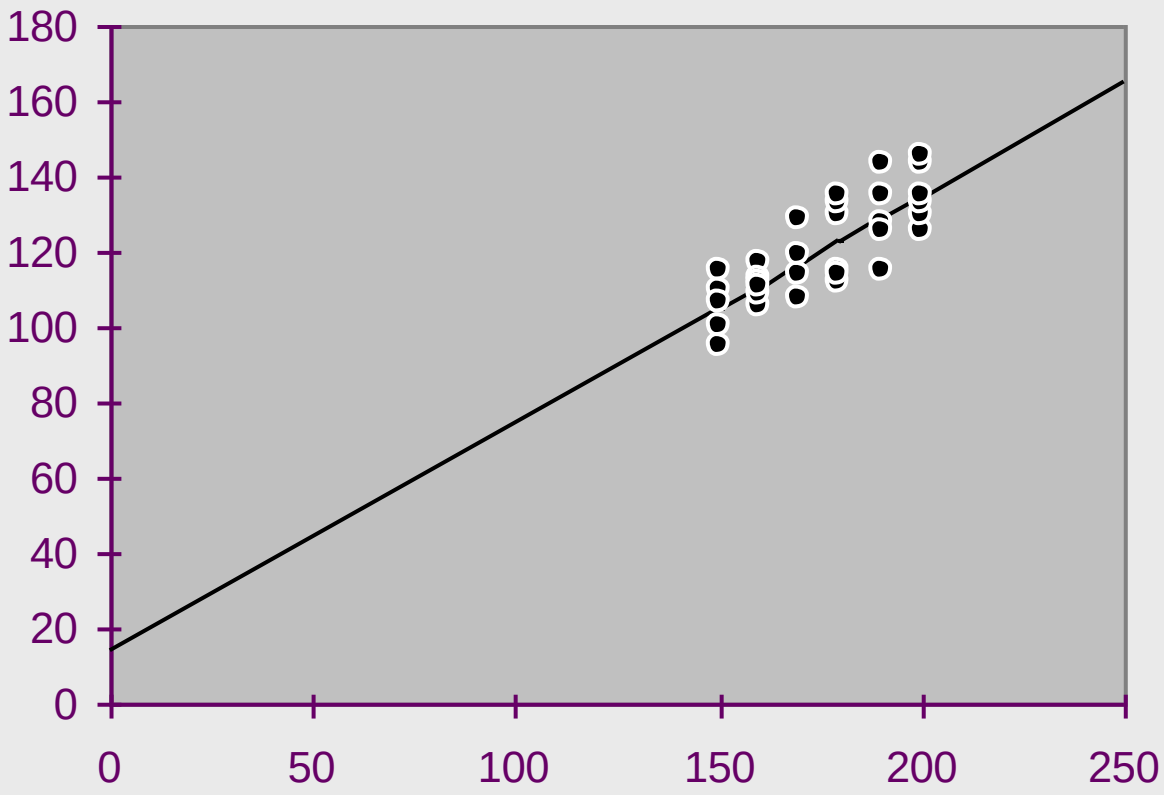
Απλή Γραμμική Παλινδρόμηση στον Πληθυσμό και στο Δείγμα

Παράδειγμα: Πληθυσμός όπου η σχέση ανάμεσα στο X και τους δεσμευμένους μέσους του Y είναι γραμμική.

Μηνιαίο Εισόδημα

Μηνιαία κατανάλωση

150	**160**	**170**	**180**	**190**	**200**
115	105	107	112	115	125
100	117	114	115	127	129
95	113	119	114	135	133
109	108	128	129	143	143
106	112		133	125	145
	111		135		135
105	**111**	**117**	**123**	**129**	**135**


180
160
140
120
100
80
60
40
20
0
0
50
100
150
200
250

?

Αν η σχέση ανάμεσα στο X και στους δεσμευμένους μέσους του Y είναι:

$$E(Y|X_i) = a + bX_i$$

ποια είναι η σχέση ανάμεσα στο Y και στο X ;

Ορίζουμε u_i την διαφορά $Y_i - E(Y|X_i)$

$$Y_i = E(Y|X_i) + u_i = a + bX_i + u_i$$

Y = *Συστηματικός όρος* + *Τυχαίος όρος*

Το u_i είναι *τυχαία μεταβλητή* και ονομάζεται
όρος σφάλματος (error term)
ή *στοχαστικός (stochastic) όρος*
ή *διαταρακτικός (disturbance) όρος*

Από τα προηγούμενα προκύπτει ότι

$$E(Y_i|X_i) = E(a + bX_i + u_i|X_i) = a + bX_i + E(u_i|X_i) = E(Y|X_i) + E(u_i|X_i)$$

και αφού $E(Y_i|X_i) = E(Y|X_i) \quad \Rightarrow \quad E(u_i|X_i) = 0$

Η υπόθεση δηλαδή ότι η γραμμή παλινδρόμησης περνάει από όλους τους **δεσμευμένους μέσους του** Y συνεπάγεται ότι όλοι οι ***δεσμευμένοι μέσοι* του** u **είναι ίσοι με το μηδέν**.

Το u, αντιπροσωπεύει όλες τις μεταβλητές που επηρεάζουν το Y και δεν λήφθηκαν υπόψη στο υπόδειγμα.

Γιατί δεν λήφθηκαν υπόψη ?

- Δεν υπάρχουν διαθέσιμα στατιστικά στοιχεία
- Ο ρόλος τους δεν είναι και τόσο σημαντικός και σ' ένα βαθμό μπορεί να θεωρηθεί τυχαίος
- Η συμμετοχή τους στο υπόδειγμα οδηγεί σε πολύπλοκη μορφή συνάρτησης
- Η προσέγγιση της πραγματικής μεταβλητής είναι δύσκολη ή αδύνατη
- Η μορφή της συνάρτησης δεν είναι η σωστή
- κ.λ.π.

Ο πληθυσμός όμως δεν είναι συνήθως γνωστός !

Στην αντίθετη περίπτωση η Στατιστική θα είχε ελάχιστο ενδιαφέρον

Οι μόνες πληροφορίες που έχει ο ερευνητής είναι αυτές που προέρχονται από το *δείγμα*

Παράδειγμα:

Y_i	X_i
100	150
112	160
114	170
129	180
115	190
145	200

Σκοπός:

Η καλύτερη δυνατή *προσέγγιση* της ευθείας του πληθυσμού

$\hat{Y}_i = \hat{a} + \hat{b}X_i$ ⇨ Η εξίσωση της ευθείας που προκύπτει από τα στοιχεία του δείγματος

$\hat{Y}_i$ Η εκτιμήτρια *(estimator)* του $E(Y|X_i)$ ή θεωρητική τιμή του Y

$\hat{a}$ Η εκτιμήτρια του a

$\hat{b}$ Η εκτιμήτρια του b

Εκτιμήτρια - Εκτίμηση:
estimator - estimate

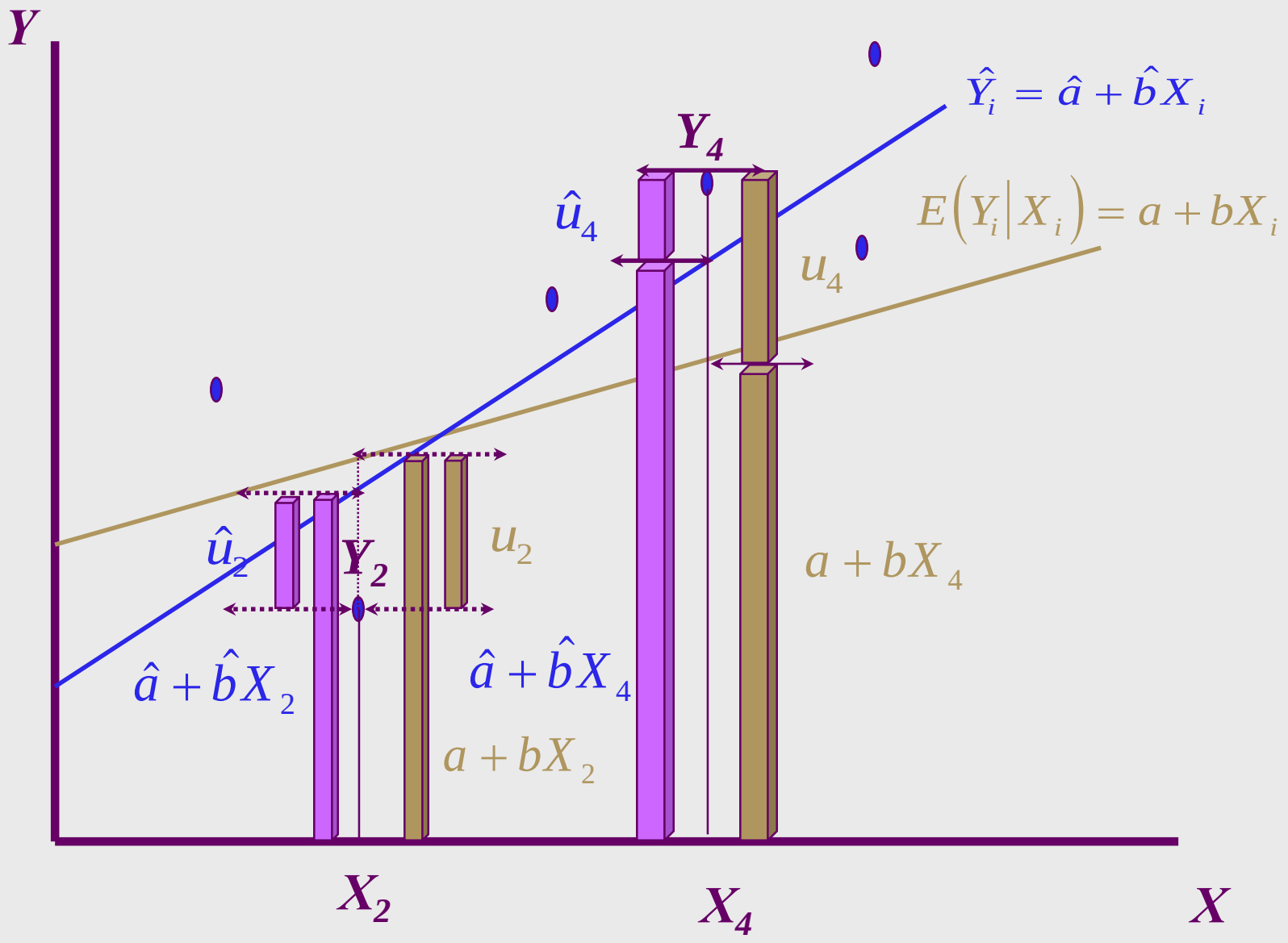
Ο όρος ***εκτιμήτρια*** αναφέρεται σε ένα ***μαθηματικό τύπο*** που δίνει τον τρόπο εκτίμησης μιας παραμέτρου του πληθυσμού με βάση τα στοιχεία του δείγματος.

Η ***εκτίμηση*** αντιπροσωπεύει μια ***συγκεκριμένη*** τιμή που προκύπτει από την εφαρμογή αυτού του τύπου.

$Y_i - \hat{Y}_i = \hat{u}_i$ Η εκτιμήτρια του u
ή κατάλοιπο *(residual)*

έτσι

$$Y_i = a + bX_i + u_i = \hat{a} + \hat{b}X_i + \hat{u}_i$$


Y
$\hat{Y}_i = \hat{a} + \hat{b}X_i$
$E(Y_i|X_i) = a + bX_i$
Y_4
$\hat{u}_4$
u_4
$\hat{u}_2$
Y_2
u_2
$a + bX_4$
$\hat{a} + \hat{b}X_2$
$\hat{a} + \hat{b}X_4$
$a + bX_2$
X_2
X_4
X

Η μέθοδος των Ελαχίστων Τετραγώνων

Η μέθοδος των ***Ελαχίστων Τετραγώνων*** *(Least Squares)* είναι μια από τις μεθόδους που χρησιμοποιούνται για την εκτίμηση της γραμμής παλινδρόμησης.

Είναι η μέθοδος που χρησιμοποιείται περισσότερο επειδή (α) έχει σημαντικές στατιστικές ιδιότητες

(β) Είναι εύκολη στην εφαρμογή της.

Κριτήριο:

Επιλογή των $\hat{a}$ και $\hat{b}$ που ελαχιστοποιούν τα τετράγωνα των αποκλίσεων της ευθείας παλινδρόμησης από τις πραγματικές τιμές

$$\sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{a} - \hat{b} X_i \right)^2 = f\left(\hat{a}, \hat{b}\right)$$

Ελαχιστοποίηση $\sum_{i=1}^{n} \hat{u}_i^2$

$$\frac{\partial \sum_{i=1}^{n} \hat{u}_i^2}{\partial \hat{a}} = 0 \qquad \frac{\partial \sum_{i=1}^{n} \hat{u}_i^2}{\partial \hat{b}} = 0$$

$$\frac{\partial \sum_{i=1}^{n} \hat{u}_i^2}{\partial \hat{a}} = 2 \sum_{i=1}^{n} \left(Y_i - \hat{a} - \hat{b}X_i\right)(-1) = 0$$

$$\frac{\partial \sum_{i=1}^{n} \hat{u}_i^2}{\partial \hat{b}} = 2 \sum_{i=1}^{n} \left(Y_i - \hat{a} - \hat{b}X_i\right)(-X_i) = 0$$

$$\sum_{i=1}^{n} Y_i = n\hat{a} + \hat{b}\sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} X_i Y_i = \hat{a}\sum_{i=1}^{n} X_i + \hat{b}\sum_{i=1}^{n} X_i^2$$

Σύστημα Κανονικών Εξισώσεων

(Normal equations)

Επίλυση κανονικών εξισώσεων

$$\hat{a} = \frac{\begin{vmatrix} \sum_{i=1}^{n} Y_i & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i Y_i & \sum_{i=1}^{n} X_i^2 \end{vmatrix}}{\begin{vmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{vmatrix}} \Rightarrow \hat{a} = \frac{\sum_{i=1}^{n} X_i^2 \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} X_i Y_i}{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}$$

$$\hat{b} = \frac{\begin{vmatrix} n & \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i Y_i \end{vmatrix}}{\begin{vmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{vmatrix}} \Rightarrow \hat{b} = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}$$

$$\hat{a} = \bar{Y} - \hat{b}\bar{X}$$

$$\hat{b} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

$$b = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}$$

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}$$

$$\hat{b} = \frac{\sum_{i=1}^{n} y_i X_i}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}$$

Παράδειγμα:

	X_i	Y_i	$X_i Y_i$	X_i^2	$\hat{Y}_i$	$\hat{u}_i$
	150	100	15000	22500	101,4	-1,4
	160	112	17920	25600	108,5	3,5
	170	114	19380	28900	115,6	-1,6
	180	129	23220	32400	122,8	6,3
	190	115	21850	36100	129,8	-14,8
	200	145	290000	40000	137,0	8,0
Αθροίσματα	**1050**	**715**	**126370**	**185500**	**715**	**0,0**
Α. Μέσοι	175	119				

$$b = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2} = \frac{126370 - 6 \cdot 175 \cdot 119}{185500 - 6 \cdot 175^2} = 0.7114$$

$$\hat{a} = 119 - 0.7114 \cdot 175 = -5.333$$

Μια αύξηση του εισοδήματος κατά μια νομισματική μονάδα επιφέρει μια αύξηση στην κατανάλωση κατά 0.7114 νομισματικές μονάδες. Η Οριακή Ροπή προς Κατανάλωση δηλαδή είναι 0.7114.

Οι υποθέσεις του Κλασσικού Γραμμικού Υποδείγματος Παλινδρόμησης

Το υπόδειγμα είναι *γραμμικό στις παραμέτρους*

$$Y_i = a + bX_i + u_i$$

Οι τιμές της ανεξάρτητης μεταβλητής X παραμένουν *σταθερές* σε επαναλαμβανόμενα δείγματα. Έτσι

$$E\left(Y_i \middle| X_i\right) = a + bX_i \quad \Rightarrow \quad E\left(u_i \middle| X_i\right) = 0$$

Η διακύμανση του όρου σφάλματος για κάθε X_i δεν εξαρτάται από το Y_i και είναι σταθερή, δηλαδή

$$\mathrm{var}\left(u_i \middle| X_i\right) = E\left[u_i - E\left(u_i\right)\middle| X_i\right]^2 = E\left(u_i^2 \middle| X_i\right) = \sigma^2$$

Η ιδιότητα αυτή ονομάζεται ***ομοσκεδαστικότητα*** *(homoscedasticity)*.

Στην αντίθετη περίπτωση έχουμε ***ετεροσκεδαστικότητα*** *(heteroscedasticity)* και

$$\mathrm{var}\left(u_i \middle| X_i\right) = \sigma_i^2$$

Μεταξύ των διαφόρων τιμών του όρου σφάλματος δεν υπάρχει συσχέτιση ή όπως αλλιώς λέγεται δεν υπάρχει ***αυτοσυσχέτιση*** *(autocorrelation)*.

$$\operatorname{cov}\left(u_i u_j \middle| X_i, X_j\right) = E\left[u_i - E\left(u_i\right)\middle| X_i\right]\left[u_j - E\left(u_j\right)\middle| X_j\right] =$$
$$= E\left[\left(u_i \middle| X_i\right)\left(u_j \middle| X_j\right)\right] = 0 \qquad \forall i \neq j$$

Δεν υπάρχει συσχέτιση μεταξύ ***u*** και ***X***. Δηλαδή

$$\operatorname{cov}\left(u_i X_i\right) = E\left[\left(u_i - E\left(u_i\right)\middle| X_i\right)\left(X_i - E\left(X_i\right)\middle| X_i\right)\right] =$$

$$E\left[u_i\left(X_i - E\left(X_i\right)\right)\middle| X_i\right] = E\left(u_i X_i\right) - E\left(u_i\right)E\left(X_i\right) = E\left(u_i X_i\right) = 0$$

Ιδιότητες των εκτιμητριών $\hat{a}$ και $\hat{b}$

Οι εκτιμήτριες $\hat{a}$ και $\hat{b}$ είναι *τυχαίες μεταβλητές* και τα χαρακτηριστικά της κατανομής τους καθώς και οι ιδιότητές τους είναι σημαντικές πληροφορίες για την εξαγωγή συμπερασμάτων για τις αντίστοιχες τιμές των παραμέτρων του πληθυσμού

Ο ***μέσος*** *(mean)* του $\hat{b}$ $\quad E\left(\hat{b}\right) = b$

Ο ***μέσος*** του $\hat{a}$ $\quad E(\hat{a}) = a$

Η ***διακύμανση*** *(variance)* του $\hat{b}$

$$\mathrm{var}\left(\hat{b}\right) = \sigma_{\hat{b}}^2 = E\left(\hat{b} - E\left(\hat{b}\right)\right)^2 = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2} \qquad \sigma^2 = E\left(u_i^2\right)$$

Η ***τυπική απόκλιση*** *(standard deviation)* του $\hat{b}$

$$\sigma_{\hat{b}} = \frac{\sigma}{\sqrt{\sum_{i=1}^{n} x_i^2}}$$

Ο υπολογισμός τόσο της διακύμανσης όσο και της τυπικής απόκλισης βασίζεται στην διακύμανση του όρου σφάλματος στον ***πληθυσμό*** (***σ²***).
Συνήθως η πληροφορία αυτή δεν είναι γνωστή γιαυτό καταφεύγουμε στην εκτίμησή της από τις πληροφορίες του **δείγματος**

τυπικό σφάλμα της εξίσωσης
(standard error of the regression)

$$\hat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-2}$$

$$\hat{\sigma} = S = \sqrt{\frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-2}}$$

Εκτιμήτριες της *διακύμανσης* και του *τυπικού σφάλματος* του $\hat{b}$

$$S_{\hat{b}}^2 = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n} x_i^2} \qquad S_{\hat{b}} = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n} x_i^2}}$$

Η ***διακύμανση*** *(variance)* του $\hat{a}$ $$\sigma_{\hat{a}}^2 = \left(\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n} x_i^2} \right) \sigma^2$$

Η ***τυπική απόκλιση*** του $\hat{a}$ *(standard deviation)* $$\sigma_{\hat{a}} = \sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n} x_i^2}}\,\sigma$$

Εκτιμήτριες της *διακύμανσης* και του *τυπικού σφάλματος* του $\hat{a}$

$$S_{\hat{a}}^2 = \left(\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n} x_i^2} \right) \hat{\sigma}^2 \qquad S_{\hat{a}} = \hat{\sigma} \sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n} x_i^2}}$$

Η συνδιακύμανση του $\hat{a}$ *και* $\hat{b}$

$$\operatorname{cov}\left(\hat{a},\hat{b}\right) = E\left(\left(\hat{a} - E(\hat{a})\right)\left(\hat{b} - E\left(\hat{b}\right)\right)\right) = -\overline{X}\,\frac{\sigma^2}{\sum x_i^2}$$

ή $$\operatorname{cov}\left(\hat{a},\hat{b}\right) = -\overline{X}\,\frac{\hat{\sigma}^2}{\sum x_i^2}$$

Παράδειγμα:

X_i	Y_i	$\hat{Y}_i$	$\hat{u}_i^2$	x_i^2
150	100	101.38	1.907	625
160	112	108.50	12.283	225
170	114	115.61	2.590	25
180	129	122.72	39.391	25
190	115	129.84	220.170	225
200	145	136.95	64.764	625
			341.105	1750

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-2} = \frac{341.105}{4} = 85.276 \quad \hat{\sigma} = 9.234$$

$$\operatorname{cov}(\hat{a},\hat{b}) = -175\frac{85.276}{1750} = -8.527$$

$$S_{\hat{b}}^2 = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n} x_i^2} = \frac{85.276}{1750} = 0.0487 \quad S_{\hat{b}} = 0.2207$$

$$S_{\hat{a}}^2 = \left(\frac{1}{6} + \frac{175^2}{1750}\right)85.276 = 1506.55 \quad S_{\hat{a}} = 38.81$$

Το $\hat{b}$ που προκύπτει από την εφαρμογή της μεθόδου των ελαχίστων τετραγώνων είναι ***αμερόληπτη εκτιμήτρια*** *(unbiased estimator)* του b αφού $E\left(\hat{b}\right) = b$

Το $\hat{a}$ που προκύπτει από την εφαρμογή της μεθόδου των ελαχίστων τετραγώνων είναι ***αμερόληπτη εκτιμήτρια*** *(unbiased estimator)* του a αφού $E(\hat{a}) = a$

Αποδεικνύεται ότι τόσο η διακύμανση του $\hat{a}$ όσο και του $\hat{b}$ είναι οι ***μικρότερες*** μεταξύ όλων των αμερόληπτων εκτιμητριών του a και b.
Τα $\hat{a}$ και $\hat{b}$ είναι δηλαδή οι ***πιο αποτελεσματικές*** εκτιμήτριες γι' αυτό και ονομάζονται ***άριστες εκτιμήτριες*** *(best estimators)*.

Άριστες Γραμμικές Αμερόληπτες Εκτιμήσεις
Best Linear Unbiased Estimators - BLUE