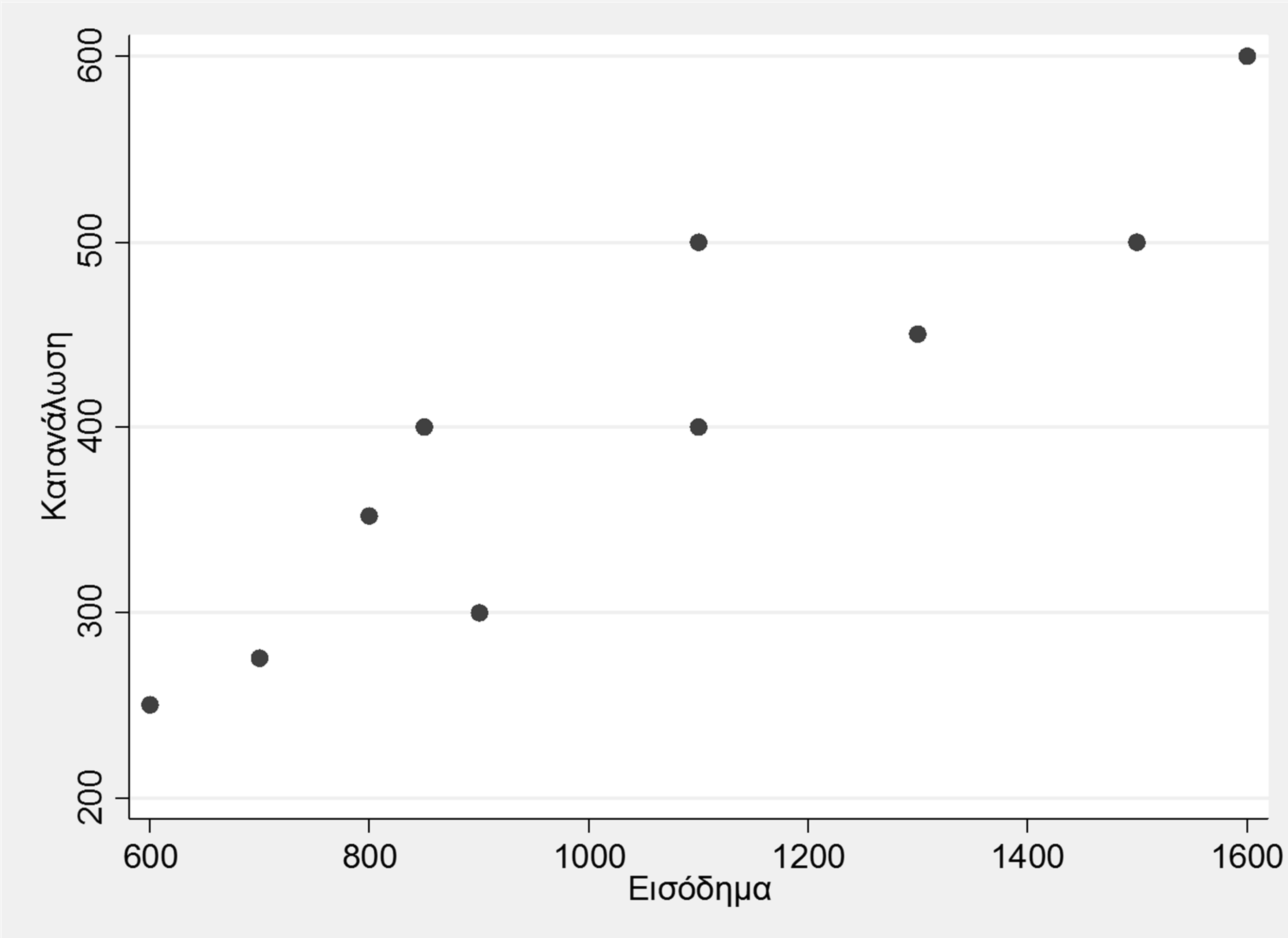
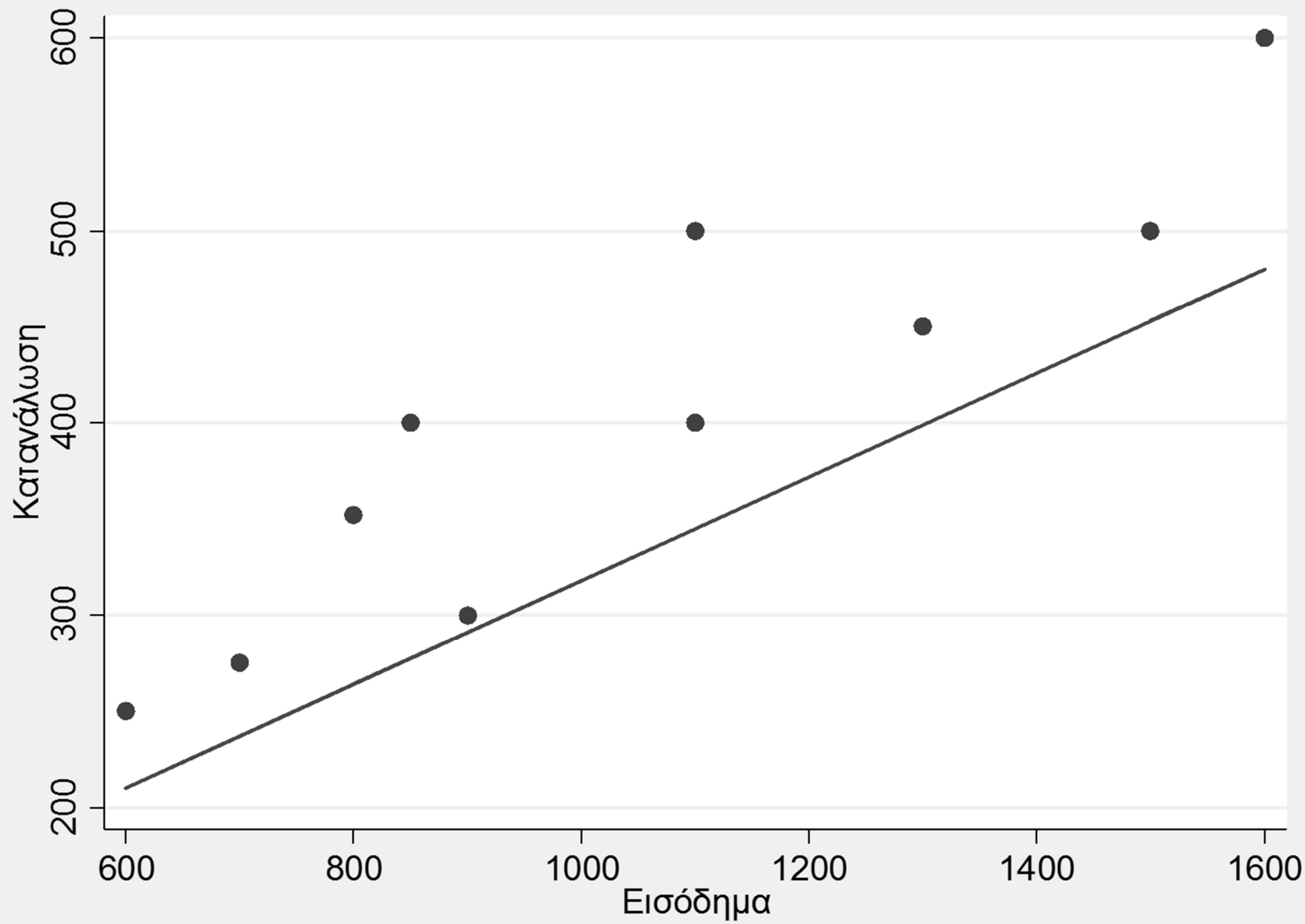
Το Κλασσικό Γραμμικό Υπόδειγμα Απλής Παλινδρόμησης

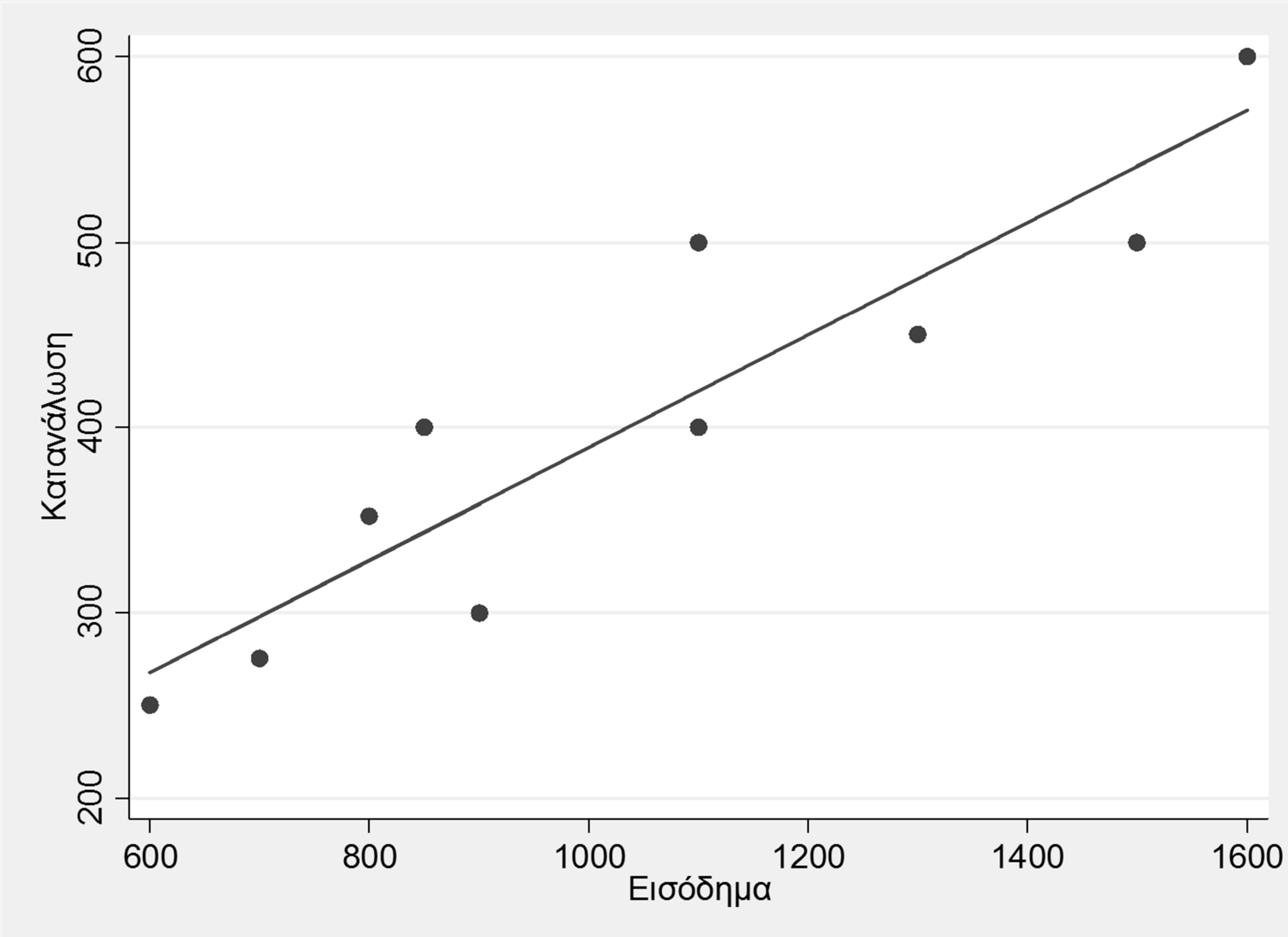
Εισόδημα	Κατανάλωση
1500	500
1600	600
1300	450
1100	400
600	250
700	275
900	300
800	352
850	400
1100	500

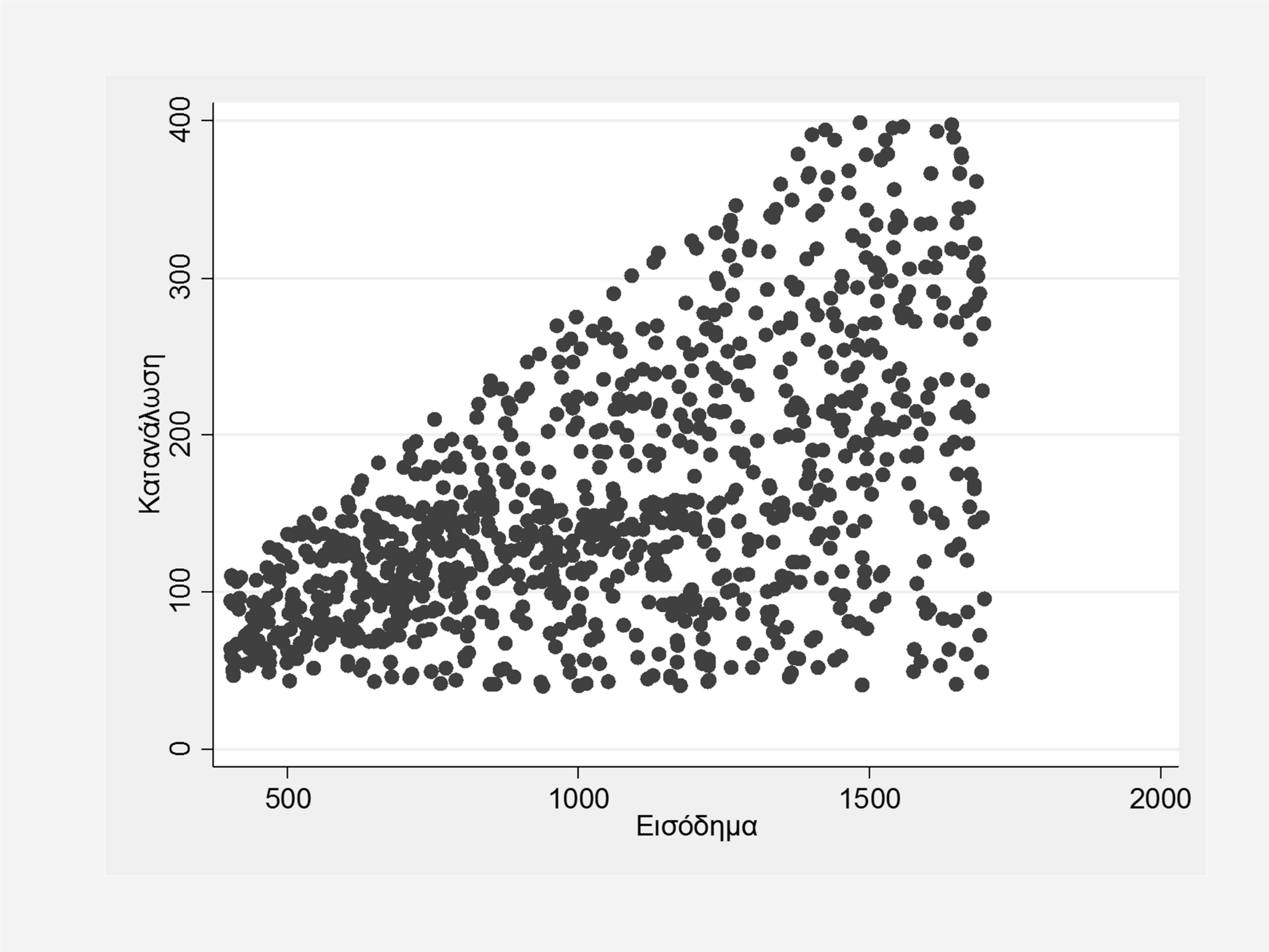


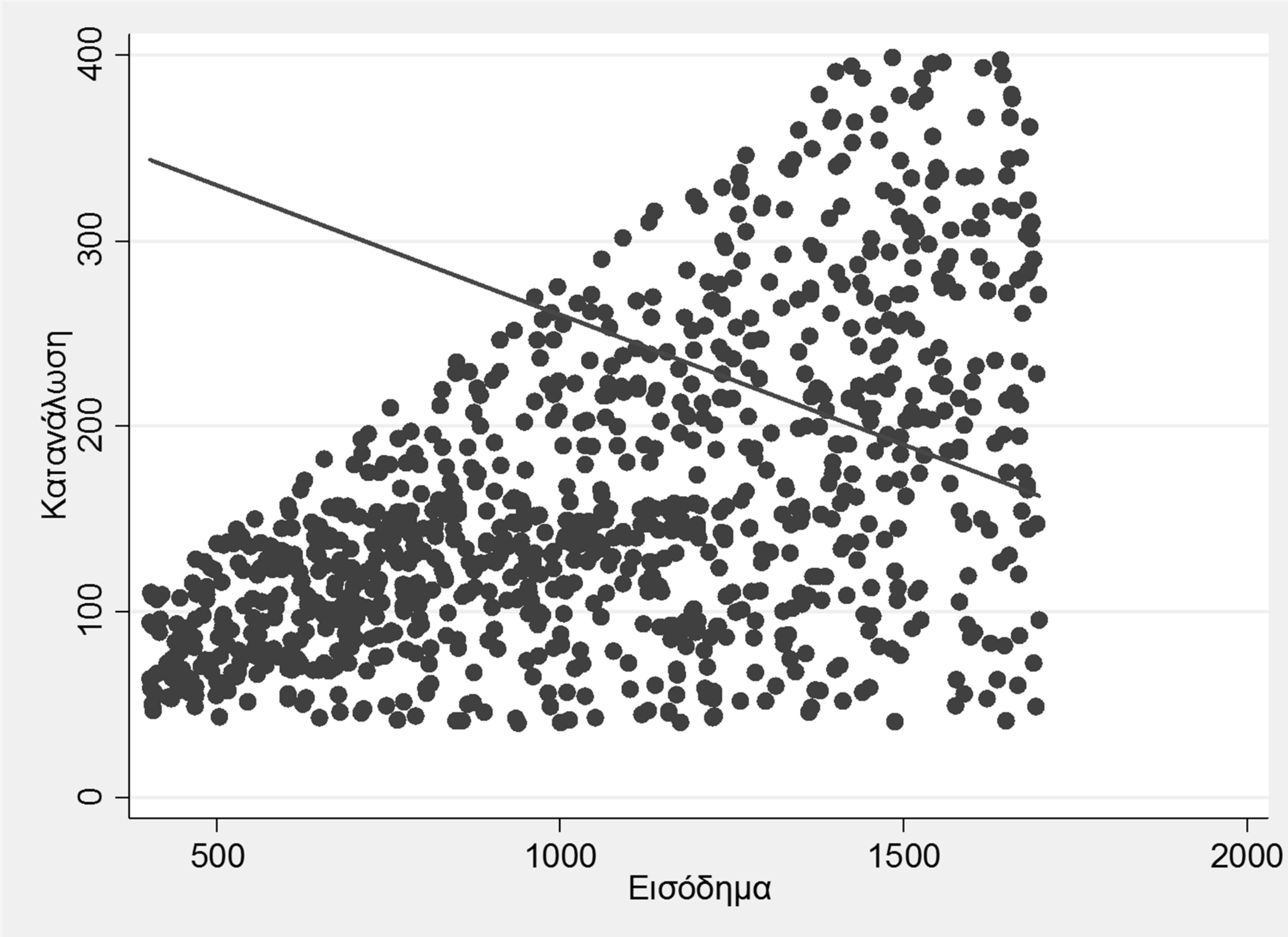
Κατανάλωση
600
500
400
300
200
600
800
1000
1200
1400
1600
Εισόδημα



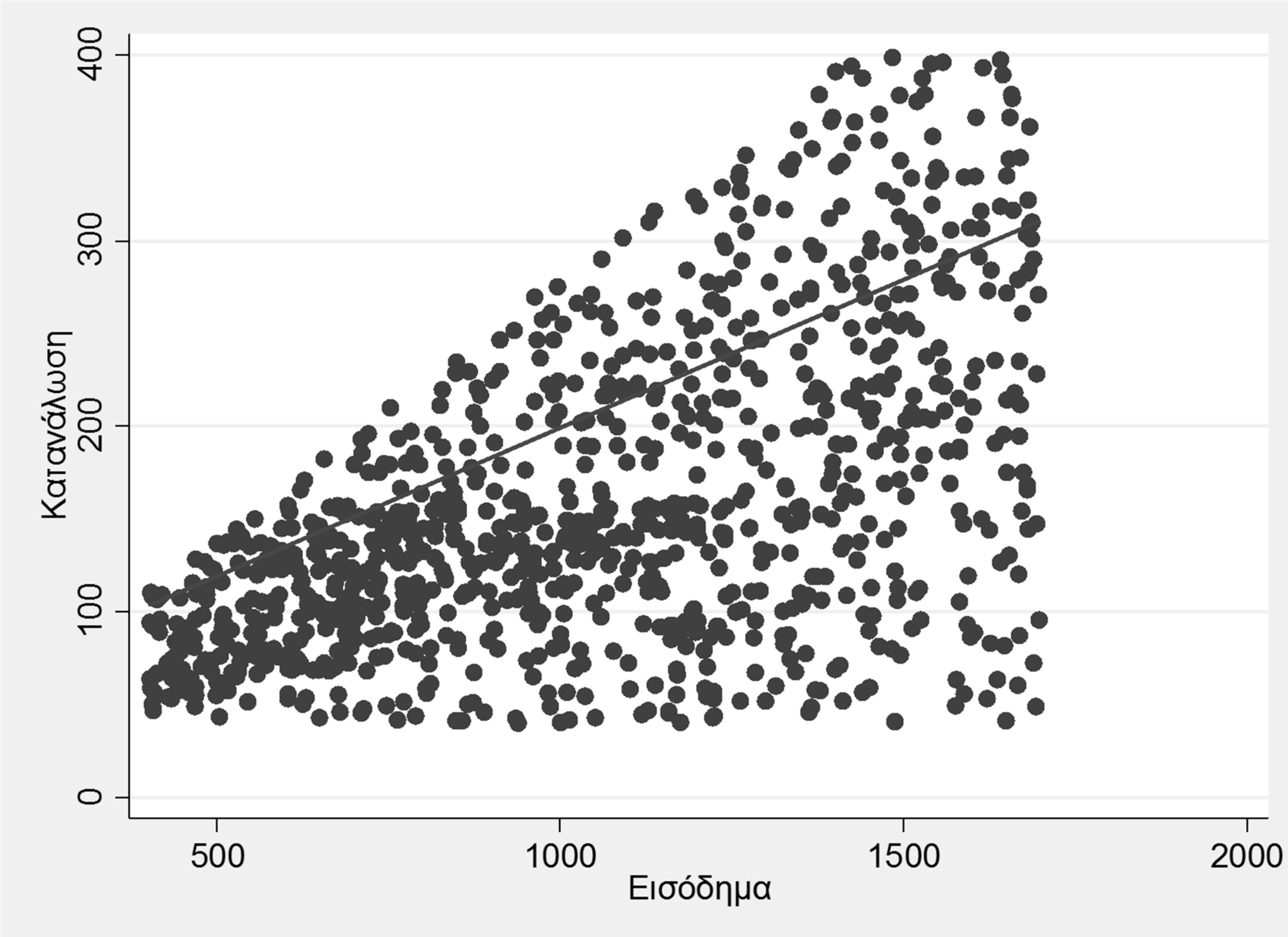
600
500
400
300
200
Κατανάλωση
600
800
1000
1200
1400
1600
Εισόδημα


Κατανάλωση
200
300
400
500
600
600
800
1000
1200
1400
1600
Εισόδημα

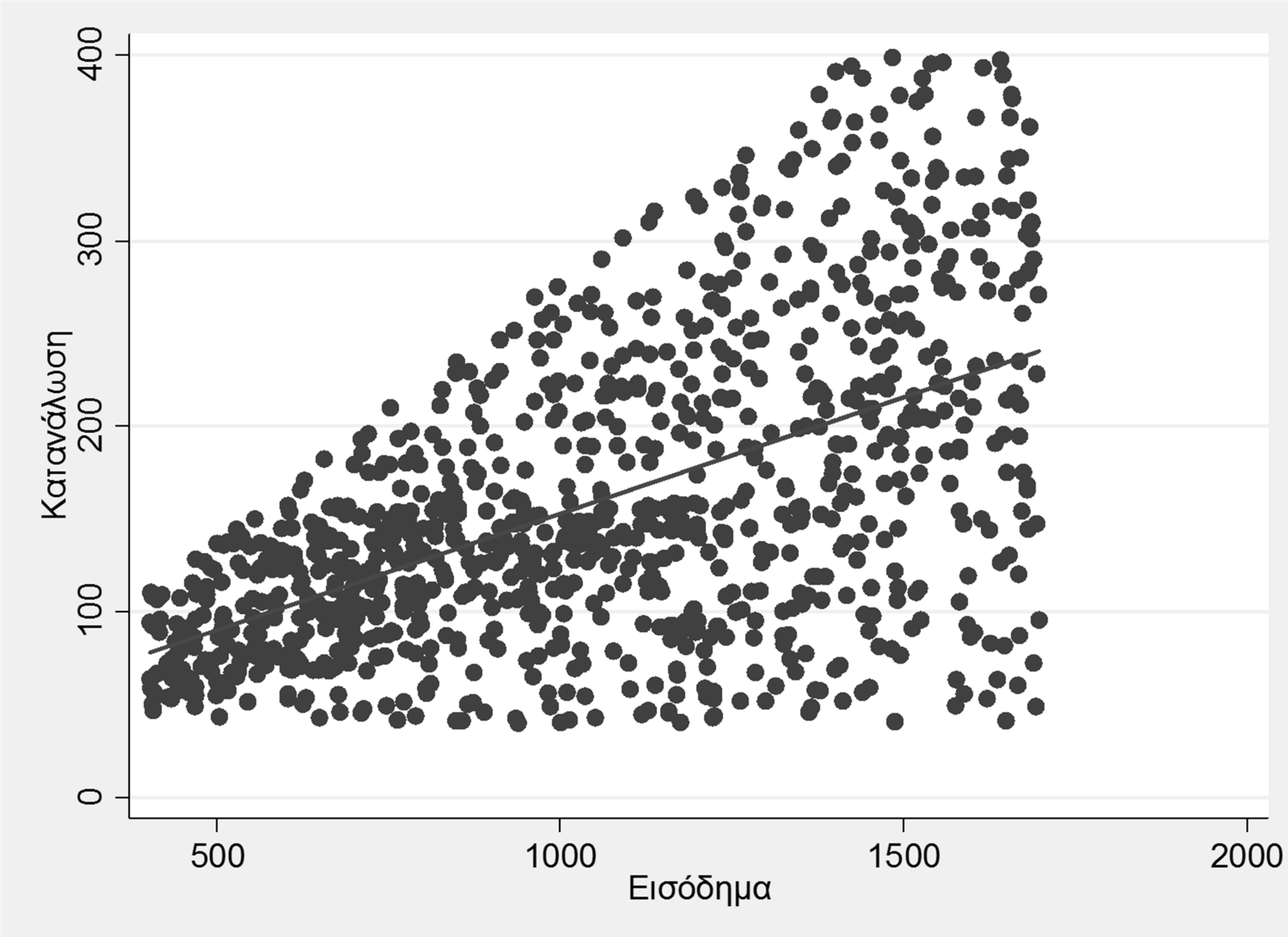

Κατανάλωση
0
100
200
300
400
500
1000
1500
2000
Εισόδημα



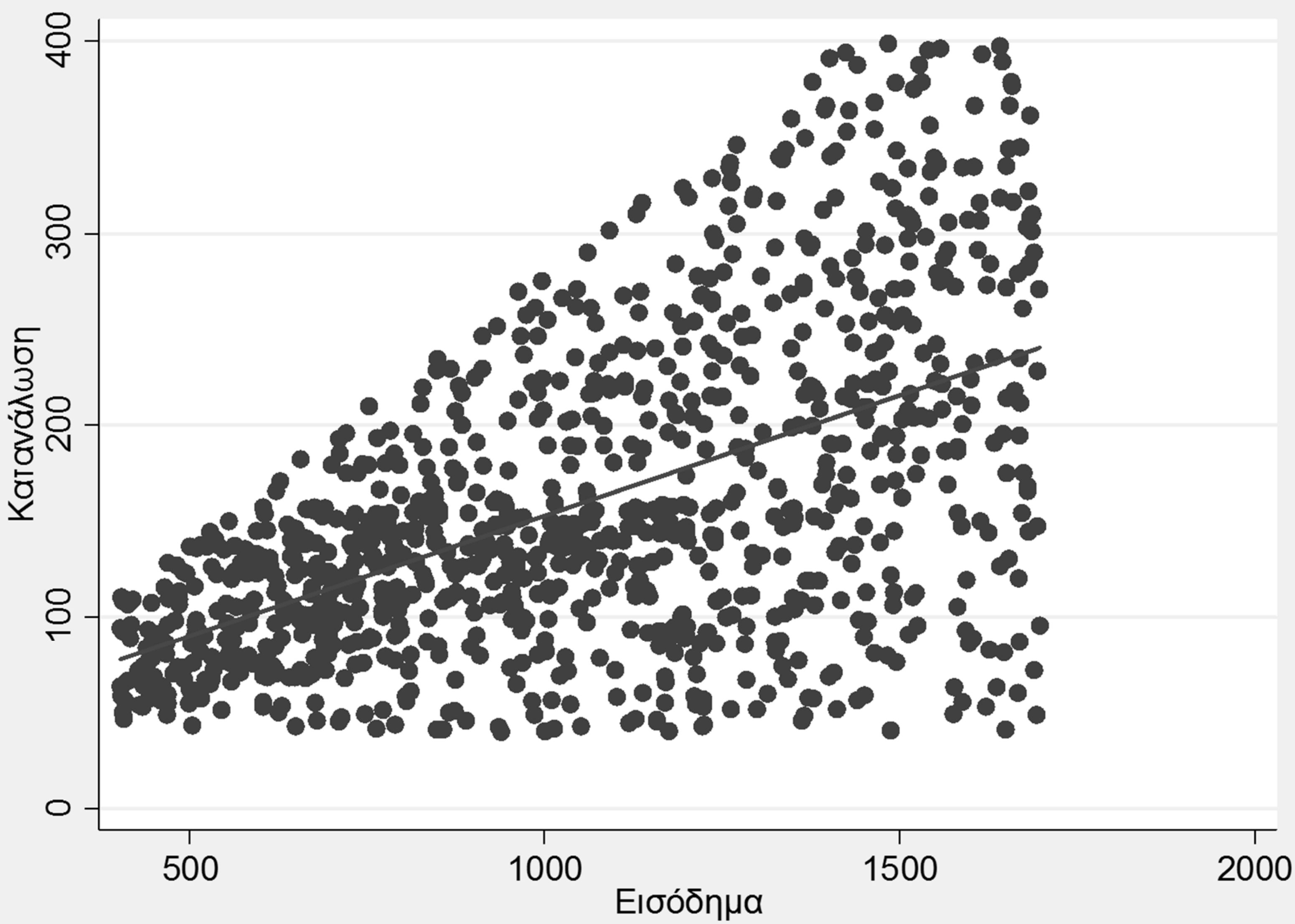
Κατανάλωση
0
100
200
300
400
500
1000
1500
2000
Εισόδημα



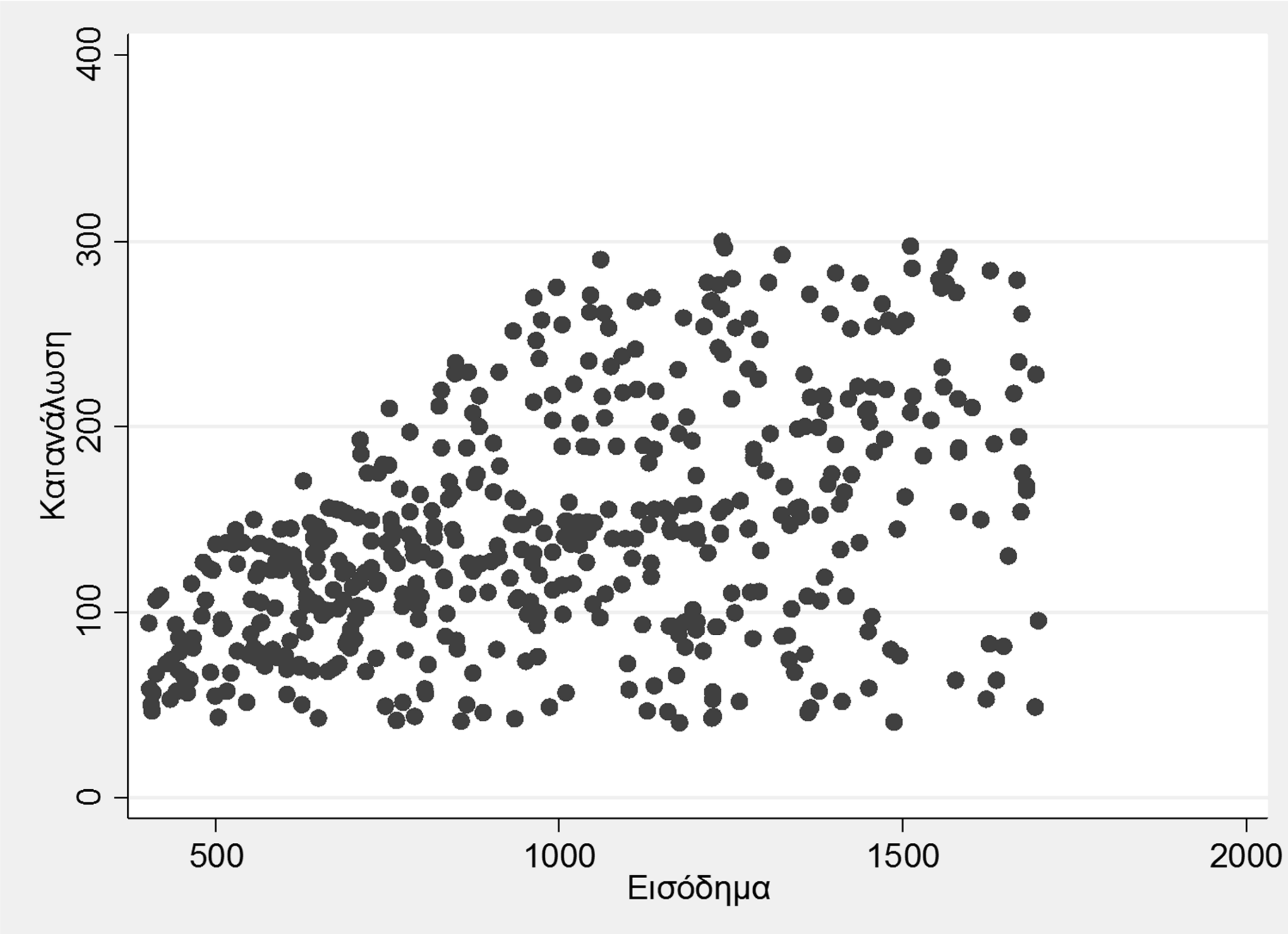
400
300
200
100
0
Κατανάλωση
500
1000
1500
2000
Εισόδημα


Κατανάλωση
0
100
200
300
400
500
1000
1500
2000
Εισόδημα

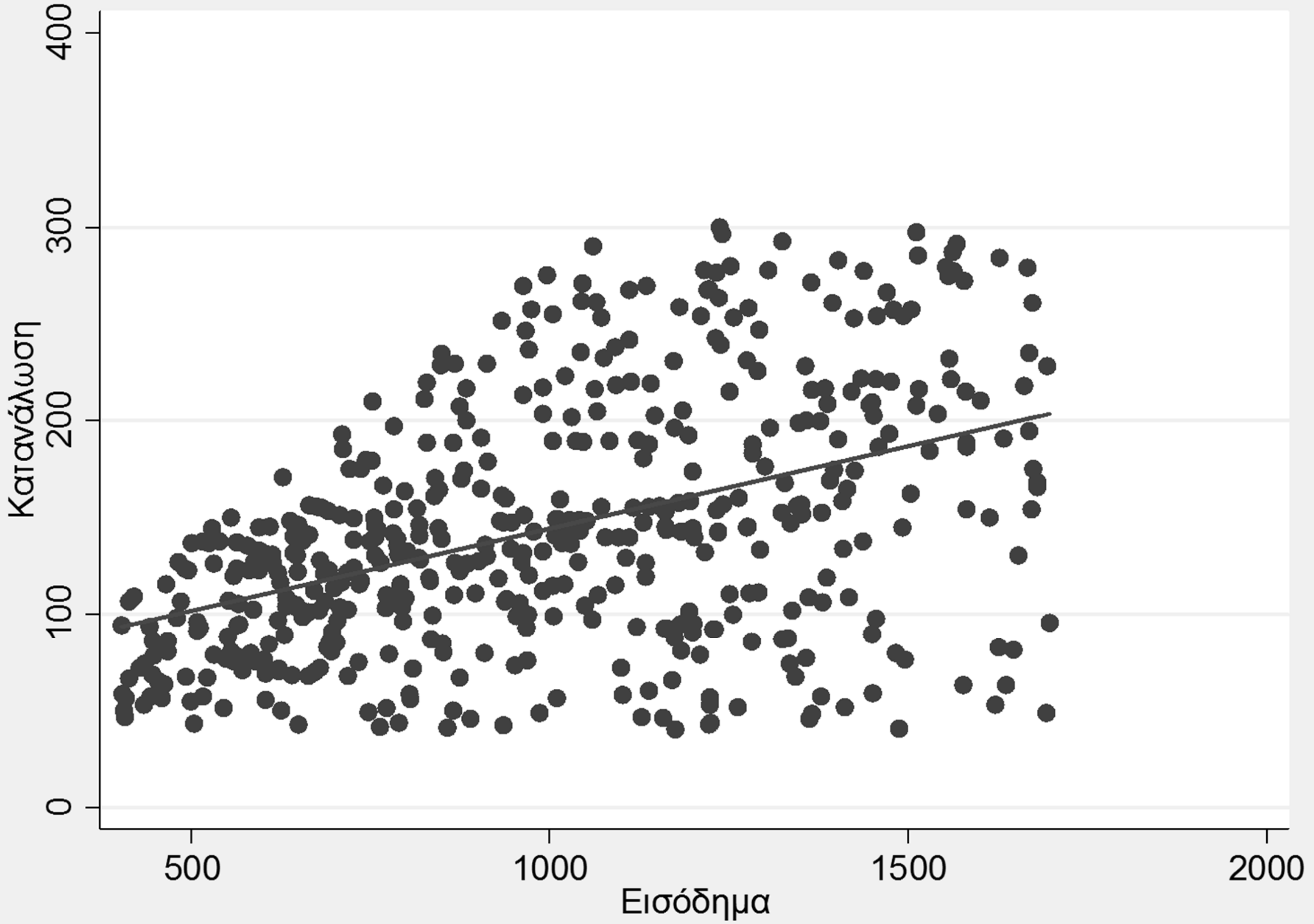
Πληθυσμός



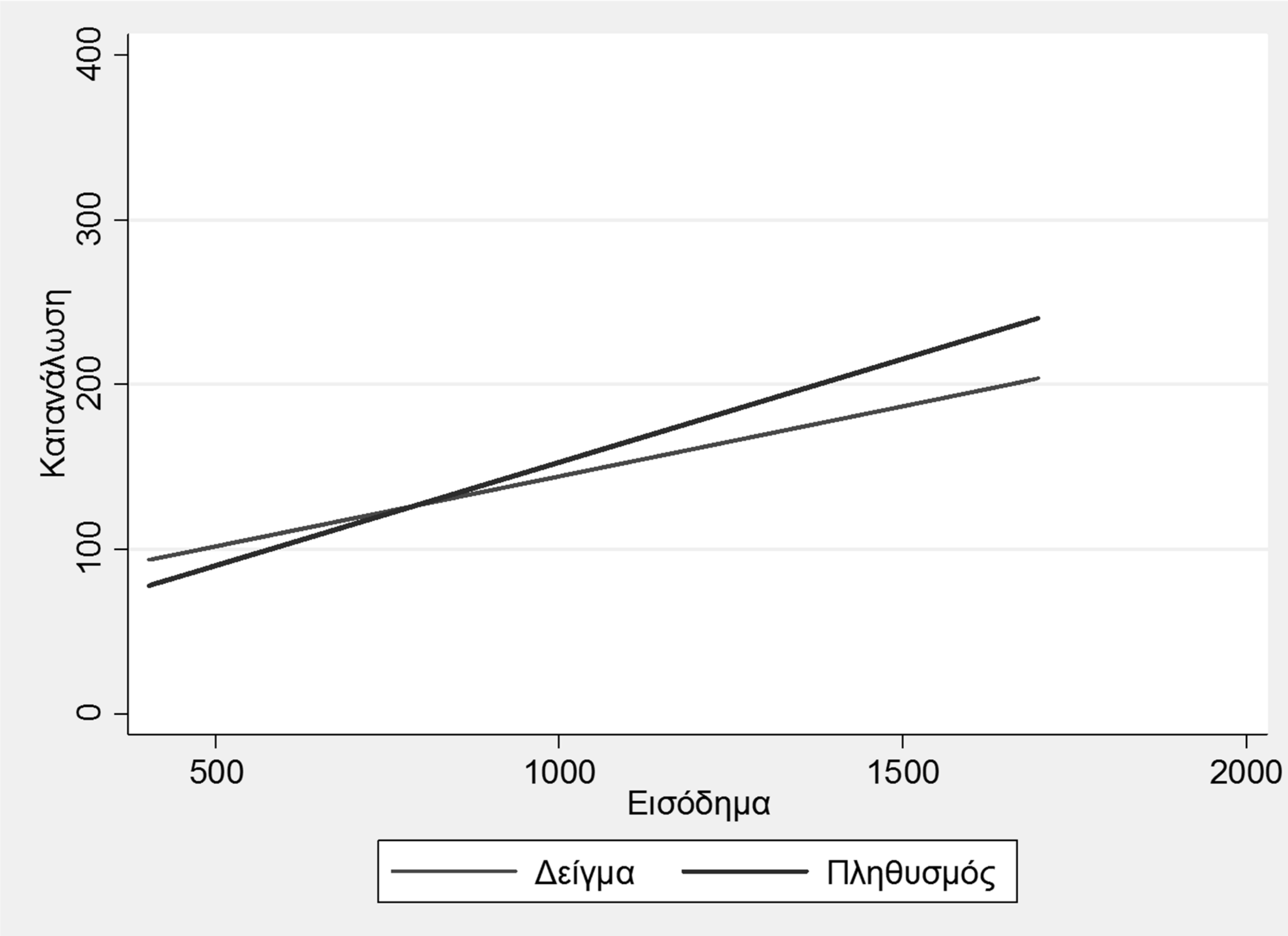
Δείγμα



Δείγμα



Δείγμα



400
300
200
100
0
Κατανάλωση
500
1000
1500
2000
Εισόδημα
Δείγμα
Πληθυσμός

Ο ρόλος της Οικονομετρίας

Οικονομική Θεωρία

Διατύπωση της αιτιώδους σχέσης μεταξύ των οικονομικών μεταβλητών

Παράδειγμα: $Q_D^A = f\left(\overset{-}{P^A}, \overset{+-}{I}, \overset{+-}{P^B}\right)$

Μια θεωρία δεν μπορεί όμως να σταθεί χωρίς εμπειρικό έλεγχο.

Οι οικονομικές θεωρίες ή προτάσεις πρέπει να ελέγχονται και να αξιολογούνται με βάση τα πραγματικά δεδομένα της οικονομικής ζωής.

Οικονομετρία

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

<u>Παράδειγμα:</u>

$$Q_D^A = \beta_0 + \beta_1 I + \beta_2 I^2 + \beta_3 P^A + \beta_4 P^B$$
$$(24) \ (2.4) \ (0.12) \ (-1.6) \ \ (0.8)$$

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

<u>Παράδειγμα:</u> $H_0 : \ \beta_4 = 0$

(γ) Προβλέψεις των μελλοντικών τιμών των οικονομικών μεταβλητών

Διαδικασία της οικονομετρικής ανάλυσης

Α) Εξειδίκευση του υποδείγματος

Καθορισμός των μεταβλητών που θα περιληφθούν, διαχωρισμός σε ενδογενείς/εξωγενείς, μαθηματική διατύπωση.

Η οικονομική θεωρία μπορεί να υποδείξει ποιες μεταβλητές είναι σημαντικές ή ίσως σχετικές αλλά δεν καθορίζει την μαθηματική μορφή που συνδέει τις μεταβλητές.

Συνήθως η επιλογή της μαθηματικής μορφής της συναρτησιακής σχέσεως είναι συνδυασμός των πληροφοριών από την οικονομική θεωρία και τα πραγματικά δεδομένα.

Β) Εκτίμηση του υποδείγματος

Εφαρμογή των κατάλληλων οικονομετρικών μεθόδων για την εκτίμηση των παραμέτρων του υποδείγματος.

Διαδικασία της οικονομετρικής ανάλυσης

Οι στατιστικές παρατηρήσεις που μπορούν να χρησιμοποιηθούν για την εκτίμηση είναι:

Χρονολογικές σειρές (time series data): Διαχρονικές παρατηρήσεις για μια σειρά ετών, μηνών κτλ για μια οικονομική μονάδα.

Διαστρωματικά στοιχεία (cross-sectional data): Παρατηρήσεις για έναν αριθμό οικονομικών μονάδων για μια χρονική στιγμή.

Δυναμικά διαστρωματικά στοιχεία (panel data): Διαχρονικές παρατηρήσεις για μια σειρά οικονομικών μονάδων.

Γ) Έλεγχος του υποδείγματος

Αξιολόγηση και έλεγχος των αποτελεσμάτων της εκτιμήσεως. Επιβεβαιώνεται ή αμφισβητείται η οικονομική θεωρία;

Ανάλυση Παλινδρόμησης

Στατιστική μέθοδος που προσπαθεί να ερμηνεύσει και να ποσοτικοποιήσει τις μεταβολές μιας μεταβλητής (***εξαρτημένη***) σε σχέση με τις μεταβολές άλλων μεταβλητών (***ανεξάρτητες***).

Ο ερευνητής με βάση την θεωρία και την εμπειρία επιλέγει:

- Την εξαρτημένη μεταβλητή
- Τις ανεξάρτητες μεταβλητές
- Την μορφή της συνάρτησης

Η **Ανάλυση Παλινδρόμησης** υπολογίζει (εκτιμά) τα ***β***

<u>Παράδειγμα:</u>

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2^2 + X_3^{\beta_3} + \beta_4 \log \frac{X_3}{X_5}$$

Παλινδρόμηση

Απλή

Μια ανεξάρτητη μεταβλητή

Πολλαπλή

Δύο ή περισσότερες ανεξάρτητες μεταβλητές

Γραμμική

Μη Γραμμική

στις παραμέτρους

<u>Παραδείγματα:</u>

Απλή Γραμμική $Y = \beta_0 + \beta_1 X$ $\quad Y = \beta_0 + \beta_1 X^2$

Απλή μη Γραμμική $Y = \beta_0 + X^{\beta_1}$

Πολλαπλή Γραμμική $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$

$$Y = \beta_0 + \beta_1 \ln X_1 + \beta_2 X_2 + \beta_3 \left(X_3 / X_2 \right)$$

Πολλαπλή Γραμμική $Y = \beta_0 X_1^{\beta_1} X_2^{\beta_2} X_3^{\beta_3}$

Προσδιοριστικές και Στοχαστικές Σχέσεις

Προσδιοριστικές

Σε κάθε τιμή της *X* αντιστοιχεί μια τιμή της *Y* π.χ. Y=α+βX

C=α+βY, σε κάθε επίπεδο διαθέσιμου εισοδήματος Y αντιστοιχεί μόνο ένα επίπεδο καταναλώσεως. Στην πραγματικότητα υπάρχουν αποκλίσεις.

Στοχαστικές ή στατιστικές

Σε κάθε τιμή της X αντιστοιχεί μια ***κατανομή*** τιμών της Y
π.χ. Y=α+βX+u

Η ανάλυση **παλινδρόμησης** ασχολείται με ***στοχαστικές σχέσεις***
Διερευνά την συναρτησιακή σχέση ανάμεσα στην X και τον μέσο της κατανομής του Y για δεδομένο X ***(δεσμευμένος μέσος)***

$$E(Y \mid X_i) = g(X_i)$$

Βασική υπόθεση του υποδείγματος της απλής γραμμικής παλινδρόμησης

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

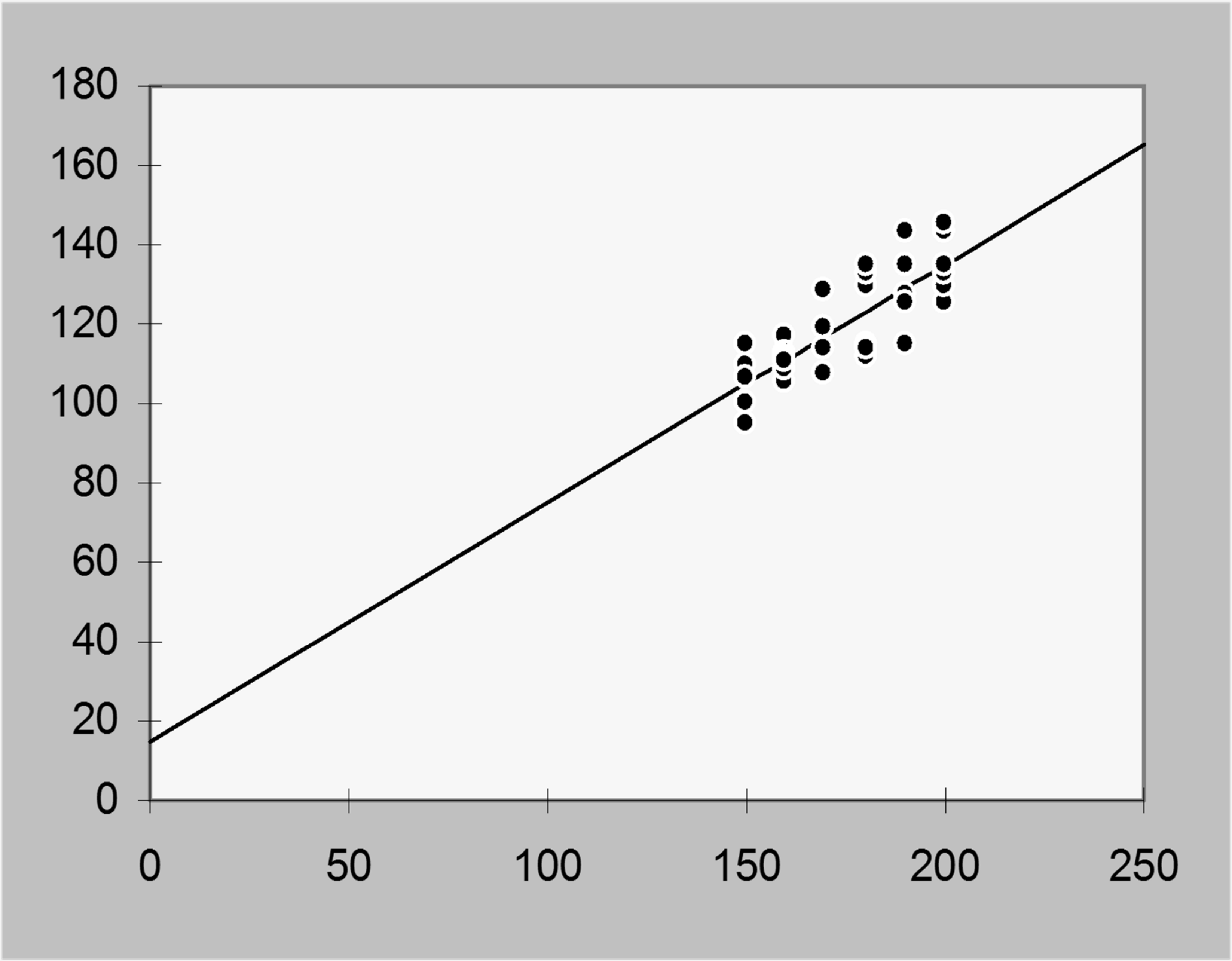
$$E(Y|X_i) = \beta_0 + \beta_1 X_i$$

Παράδειγμα

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

Η έννοια του τυχαίου σφάλματος

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

$$E(Y|X_i) = \beta_0 + \beta_1 X_i$$

ποιά είναι η σχέση ανάμεσα στο Y και στο X ?

Ορίζουμε $u_i = Y_i - E(Y|X_i)$

$$Y_i = E(Y|X_i) + u_i = \beta_0 + \beta_1 X_i + u_i$$

Y = ***Συστηματικός όρος*** + ***Τυχαίος όρος***

Το u_i είναι τυχαία μεταβλητή και ονομάζεται ***όρος σφάλματος*** ή ***στοχαστικός όρος*** ή ***διαταρακτικός όρος*** (διαταράσσει την προσδιοριστική σχέση ανάμεσα στις Y και X)

Η υπόθεση 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.

$$E(Y_i|X_i) = E(\beta_0 + \beta_1 X_i + u_i|X_i) = \beta_0 + \beta_1 X_i + E(u_i|X_i)$$

$$= E(Y_i|X_i) + E(u_i|X_i) \quad \Rightarrow \quad E(u_i|X_i) = 0$$

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
- Δεν είναι γνωστές ή δεν μπορούν να μετρηθούν στατιστικά π.χ. ψυχολογικοί παράγοντες
- Είναι αστάθμητοι ή απρόβλεπτοι παράγοντες που χαρακτηρίζουν την ανθρώπινη συμπεριφορά

Πληθυσμός και Δείγμα

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

Παράδειγμα:

Y_i	X_i
100	150
112	150
114	170
129	170
115	200
145	200

Σκοπός:

Η καλύτερη δυνατή ***προσέγγιση*** της ευθείας του πληθυσμού

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ ➡ Η εξίσωση της ευθείας που προκύπτει από τα στοιχεία του ***δείγματος***

$\hat{Y}_i$ Ο εκτιμητής του $E(Y|X_i)$ ή θεωρητική τιμή του $\boldsymbol{Y}$

$\hat{\beta}_0$ Ο εκτιμητής του β_0

$\hat{\beta}_1$ Ο εκτιμητής του β_1

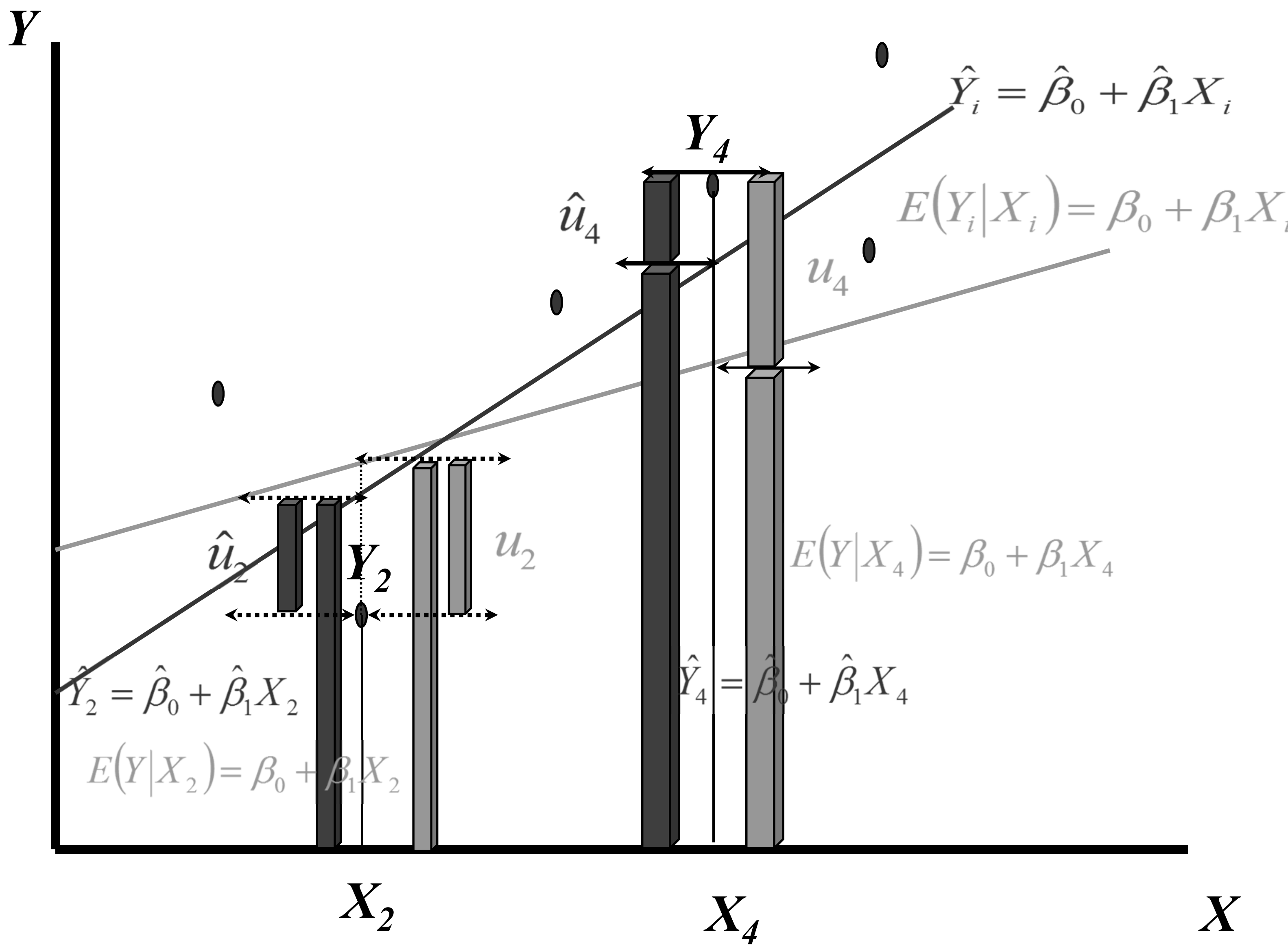
Εκτιμητής και Εκτίμηση

Εκτιμητής: μαθηματικός τύπος που δίνει τον τρόπο εκτίμησης μιας παραμέτρου του πληθυσμού με βάση τα στοιχεία του δείγματος.

Εκτίμηση: Μια συγκεκριμένη τιμή που προκύπτει από την εφαρμογή του τύπου.

$Y_i - \hat{Y}_i = \hat{u}_i$ Ο εκτιμητής του u ή ***κατάλοιπο***

έτσι $Y_i = \beta_0 + \beta_1 X_i + u_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i$



Y
$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$
Y_4
$\hat{u}_4$
$E(Y_i|X_i) = \beta_0 + \beta_1 X_i$
u_4
$\hat{u}_2$
Y_2
u_2
$E(Y|X_4) = \beta_0 + \beta_1 X_4$
$\hat{Y}_2 = \hat{\beta}_0 + \hat{\beta}_1 X_2$
$\hat{Y}_4 = \hat{\beta}_0 + \hat{\beta}_1 X_4$
$E(Y|X_2) = \beta_0 + \beta_1 X_2$
X_2
X_4
X

Η μέθοδος των Ελαχίστων Τετραγώνων

Μια από τις μεθόδους που χρησιμοποιούνται για την εκτίμηση της γραμμής παλινδρόμησης.

Είναι η μέθοδος που χρησιμοποιείται περισσότερο επειδή
(α) οι εκτιμητές έχουν πολλές επιθυμητές ιδιοτήτες
(β) Είναι εύκολη στην εφαρμογή της.

Ο αριθμός των εκτιμητών είναι ουσιαστικά άπειρος δηλαδή μπορούμε να κατασκευάσουμε και άπειρες γραμμές παλινδρομήσεως. Χρειαζόμαστε ένα κριτήριο.

Κριτήριο:
Επιλογή των $\hat{\beta}_0$ και $\hat{\beta}_1$ που ελαχιστοποιούν τα τετράγωνα των αποκλίσεων της ευθείας παλινδρόμησης από τις πραγματικές τιμές

$$\sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i \right)^2 = f\left(\hat{\beta}_0, \hat{\beta}_1 \right)$$

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

$$\frac{\partial \sum_{i=1}^{n} \hat{u}_i^2}{\partial \hat{\beta}_0} = 0 \qquad \frac{\partial \sum_{i=1}^{n} \hat{u}_i^2}{\partial \hat{\beta}_1} = 0$$

$$\frac{\partial \sum_{i=1}^{n} \hat{u}_i^2}{\partial \hat{\beta}_0} = 2\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)(-1) = 0$$

$$\frac{\partial \sum_{i=1}^{n} \hat{u}_i^2}{\partial \hat{\beta}_1} = 2\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)(-X_i) = 0$$

$$\left.\begin{aligned} \sum_{i=1}^{n} Y_i &= n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i Y_i &= \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 \end{aligned}\right\}$$

Σύστημα Κανονικών Εξισώσεων

Επίλυση κανονικών εξισώσεων

$$\hat{\beta}_1 = \frac{\begin{vmatrix} n & \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i Y_i \end{vmatrix}}{\begin{vmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{vmatrix}} \Rightarrow \hat{\beta}_1 = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}$$

$$\hat{\beta}_0 = \frac{\begin{vmatrix} \sum_{i=1}^{n} Y_i & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i Y_i & \sum_{i=1}^{n} X_i^2 \end{vmatrix}}{\begin{vmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{vmatrix}} \Rightarrow \hat{\beta}_0 = \frac{\sum_{i=1}^{n} X_i^2 \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} X_i Y_i}{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}$$

Διαιρούμε την πρώτη κανονική εξίσωση με n και έχουμε:

$$\frac{\sum_{i=1}^{n} Y_i}{n} = \frac{n\hat{\beta}_0}{n} + \frac{\hat{\beta}_1 \sum_{i=1}^{n} X_i}{n}$$

$$\overline{Y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{X} \;\Rightarrow\; \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2} = \frac{\sum_{i=1}^{n} X_i Y_i - n\frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n^2}}{\sum_{i=1}^{n} X_i^2 - n\frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n^2}} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}$$

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

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

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

Υπολογίζουμε δηλαδή το $\hat{\beta}_1$ ως συνάρτηση των αποκλίσεων των τιμών των Y και X από τους μέσους τους.

Είναι επίσης φανερό γιατί θέλουμε οι τιμές της X να μην είναι όλες ίδιες μεταξύ τους, γιατί αλλιώς $\Sigma x_i^2=0$.

Παράδειγμα:

	X_i	Y_i	X_iY_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**				

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_iY_i - n\overline{X}\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2} = \frac{126370 - 6\cdot 175\cdot 119}{185500 - 6\cdot 175^2} = 0,7114$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X} = 119 - 0,7114\cdot 175 = -5,333$$

$$\hat{Y}_i = -5,333 + 0,7114X_i \qquad Y_i = -5,333 + 0,7114X_i + \hat{u}_i$$

Μια αύξηση του εισοδήματος κατά μια νομισματική μονάδα επιφέρει μια αύξηση στην κατανάλωση κατά 0,7114 νομισματικές μονάδες.

Οριακή Ροπή προς Κατανάλωση = 0,7114.

Άσκηση:

X_i	Y_i
5	12
2	10
3	8
4	10

Άσκηση:

	X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$\hat{Y}_i$	$\hat{u}_i$
	5	12	1.5	2	2.25	3	11.2	0.8
	2	10	-1.5	0	2.25	0	8.8	1.2
	3	8	-0.5	-2	0.25	1	9.6	-1.6
	4	10	0.5	0	0.25	0	10.4	-0.4
Αθροίσματα					**5**	**4**		
Α. Μέσοι	**3.5**	**10**						

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{4}{5} = 0,8$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 10 - 0,8 \cdot 3,5 = 7,2$$

$$\hat{Y}_i = 7,2 + 0,8 X_i$$

$$Y_i = 7,2 + 0,8 X_i + \hat{u}_i$$

Οι υποθέσεις του Κλασσικού Γραμμικού Υποδείγματος Παλινδρόμησης

Για την εκτίμηση χρησιμοποιούμε τα ζεύγη παρατηρήσεων (Y, X) αλλά εφόσον δεν έχουμε παρατηρήσεις για το u, πρέπει να κάνουμε υποθέσεις για την συμπεριφορά του.

Απαιτούμενες υποθέσεις προκειμένου η μέθοδος των ελαχίστων τετραγώνων να αποτελεί την ιδανική μέθοδο

Το υπόδειγμα είναι γραμμικό στις παραμέτρους και έχει όρο σφάλματος που λειτουργεί προσθετικά

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Η μεταβλητή u_i είναι τυχαία μεταβλητή που παίρνει θετικές και αρνητικές τιμές αλλά κατά μέσο όρο η τιμή της είναι μηδέν

$$E\left(u_i\right) = 0$$

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

$$\operatorname{var}(u_i \mid X_i) = E\left[u_i - E(u_i) \mid X_i\right]^2 = E\left[u_i^2 \mid X_i\right] = \sigma^2$$

Η ιδιότητα αυτή ονομάζεται ***ομοσκεδαστικότητα.*** Στην αντίθετη περίπτωση έχουμε ***ετεροσκεδαστικότητα*** και

$$\operatorname{var}(u_i \mid X_i) = \sigma_i^2$$

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

$$\operatorname{cov}\left(u_i u_j \mid X_i, X_j\right) = E\left[u_i - E(u_i) \mid X_i\right]\left[u_j - E(u_j) \mid X_j\right] =$$

$$= E\left[(u_i \mid X_i)(u_j \mid X_j)\right] = 0 \qquad \forall i \neq j$$

Οι τιμές της ανεξάρτητης μεταβλητής X παραμένουν σταθερές σε επαναλαμβανόμενα δείγματα.
Η X δεν είναι στοχαστική.

Δεν υπάρχει συσχέτιση μεταξύ u και X. Δηλαδή

$$\operatorname{cov}\left(u_i, X_i\right)=0$$

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

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

Οι τιμές της X πρέπει να είναι διαφορετικές μεταξύ τους.

Ιδιότητες των εκτιμητών $\hat{\beta}_0$ και $\hat{\beta}_1$

Οι εκτιμητές $\hat{\beta}_0$ και $\hat{\beta}_1$ 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

Αποδεικνύεται ότι οι εκτιμητές που προκύπτουν από την μέθοδο των Ελαχίστων Τετραγώνων είναι Άριστοι Γραμμικοί Αμερόληπτοι Εκτιμητές:

Α) Είναι γραμμικές συναρτήσεις των παρατηρήσεων της εξαρτημένης μεταβλητής Y

Β) Είναι αμερόληπτοι, δηλαδή:

Ο **μέσος** του $\hat{\beta}_0$ $E\left(\hat{\beta}_0\right) = \beta_0$

Ο **μέσος** του $\hat{\beta}_1$ $E\left(\hat{\beta}_1\right) = \beta_1$

Γ) Μεταξύ όλων των αμερόληπτων εκτιμητών έχουν την μικρότερη διακύμανση που δίνεται από τις σχέσεις:

Η **διακύμανση** του $\hat{\beta}_0$: $\mathrm{V}\left(\hat{\beta}_0\right)=\sigma^2_{\hat{\beta}_0} = E\left(\hat{\beta}_0 - E\left(\hat{\beta}_0\right)\right)^2 = \dfrac{\sigma^2\sum_{i=1}^{n} X_i^2}{n\sum_{i=1}^{n} x_i^2}$

Η **διακύμανση** του $\hat{\beta}_1$: $\mathrm{V}\left(\hat{\beta}_1\right)=\sigma^2_{\hat{\beta}_1} = E\left(\hat{\beta}_1 - E\left(\hat{\beta}_1\right)\right)^2 = \dfrac{\sigma^2}{\sum_{i=1}^{n} x_i^2}$

Η **συνδιακύμανση** των $\hat{\beta}_0, \hat{\beta}_1$:

$$\mathrm{Cov}\left(\hat{\beta}_0, \hat{\beta}_1\right) = -\frac{\sigma^2 \bar{X}}{\sum_{i=1}^{n} x_i^2} = -\mathrm{V}\left(\hat{\beta}_1\right)\bar{X} = -\mathrm{V}\left(\hat{\beta}_0\right)\frac{n\bar{X}}{\sum_{i=1}^{n} X_i^2}$$

Όπου $\sigma^2 = E\left(u_i^2\right)$ αναφέρεται στον άγνωστο **πληθυσμό**

Για να εκτιμήσουμε τις διακυμάνσεις και την συνδιακύμανση των συντελεστών $\hat{\beta}_0, \hat{\beta}_1$ θα πρέπει να έχουμε μια εκτίμηση για την σ^2.

Στο **δείγμα** $\hat{\sigma}^2 = S^2 = \dfrac{\sum_{i=1}^{n} \hat{u}_i^2}{n-2}$ είναι αμερόληπτος εκτιμητής της σ^2.

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

Τυπικό Σφάλμα της εξίσωσης ή
Τυπικό σφάλμα της εκτιμήσεως της Y

Το $\hat{\sigma}^2$ μπορεί να υπολογιστεί και με τον παρακάτω τύπο:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} y_i^2 - \hat{\beta}_1 \sum_{i=1}^{n} x_i y_i}{n-2}$$

Παράδειγμα:

	X_i	Y_i	$X_i^{\,2}$	$\hat{Y}_i$	$\hat{u}_i^2$	x_i^2
	150	100	22500	101,38	1,907	625
	160	112	25600	108,50	12,283	225
	170	114	28900	115,61	2,590	25
	180	129	32400	122,72	39,391	25
	190	115	36100	129,84	220,169	225
	200	145	40000	136,95	64,764	625
Α. μέσος	**175**	**119,16**				
Άθροισμα			**185500**		**341,104**	**1750**

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\hat{u}_i^2}{n-2} = \frac{341,104}{4} = 85,276 \qquad \hat{\sigma} = 9,2345$$

$$\hat{\mathrm{cov}}\left(\hat{\beta}_0, \hat{\beta}_1\right) = -175\frac{85,276}{1750} = -8,528$$

$$S_{\hat{\beta}_0}^2 = \frac{\hat{\sigma}^2 \sum_{i=1}^{n} X^2}{n\sum_{i=1}^{n} x_i^2} = \left(\frac{185500}{6 \cdot 1750}\right) 85,276 = 1506,543 \qquad S_{\hat{\beta}_0} = 38,814$$

$$S_{\hat{\beta}_1}^2 = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n} x_i^2} = \frac{85,276}{1750} = 0,0487 \qquad S_{\hat{\beta}_1} = 0,2207$$

Ιδιότητες της ευθείας παλινδρόμησης

Η ευθεία παλινδρόμησης περνάει από το σημείο $\overline{Y}, \overline{X}$
αφού $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$

$$\overline{\hat{Y}} = \overline{Y} \qquad \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i = \left(\overline{Y} - \hat{\beta}_1 \overline{X}\right) + \hat{\beta}_1 X_i = \overline{Y} + \hat{\beta}_1 \left(X_i - \overline{X}\right)$$

$$\sum_{i=1}^{n} \hat{Y}_i = n\overline{Y} + \hat{\beta}_1 \sum_{i=1}^{n} \left(X_i - \overline{X}\right) = n\overline{Y} \quad \frac{\sum_{i=1}^{n} \hat{Y}_i}{n} = \overline{Y} \Rightarrow \quad \overline{\hat{Y}} = \overline{Y}$$

$$\sum_{i=1}^{n} \hat{u}_i = 0 \qquad \sum_{i=1}^{n} \hat{u}_i = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = \sum_{i=1}^{n} Y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} X_i = 0$$

Αυτό όμως προϋποθέτει την ύπαρξη του $\hat{\beta}_0$.
Αν υποχρεώσουμε την ευθεία παλινδρόμησης να περάσει από την ***αρχή των αξόνων*** τότε η ιδιότητα αυτή ***δεν ισχύει***

$$\hat{y}_i = \hat{\beta}_1 x_i$$

$$\left.\begin{aligned} Y_i &= \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i \\ \overline{Y} &= \hat{\beta}_0 + \hat{\beta}_1 \overline{X} \end{aligned}\right\} \Rightarrow Y_i - \overline{Y} = \hat{\beta}_1\left(X_i - \overline{X}\right) + \hat{u}_i$$

$$y_i = \hat{\beta}_1 x_i + \hat{u}_i \Rightarrow \hat{y}_i = \hat{\beta}_1 x_i$$

$$\sum_{i=1}^{n} \hat{y}_i \hat{u}_i = 0$$

$$\sum_{i=1}^{n} \hat{y}_i \hat{u}_i = \sum_{i=1}^{n} \left(\hat{\beta}_1 x_i\right) \hat{u}_i = \hat{\beta}_1 \sum_{i=1}^{n} x_i \hat{u}_i = \hat{\beta}_1 \sum_{i=1}^{n} x_i \left(y_i - \hat{\beta}_1 x_i\right) =$$

$$= \hat{\beta}_1 \sum_{i=1}^{n} x_i y_i - \hat{\beta}_1^2 \sum_{i=1}^{n} x_i^2 = \hat{\beta}_1 \left(\hat{\beta}_1 \sum_{i=1}^{n} x_i^2\right) - \hat{\beta}_1^2 \sum_{i=1}^{n} x_i^2 = 0$$

$$\sum_{i=1}^{n} \hat{u}_i X_i = 0$$

$$\sum_{i=1}^{n} \hat{u}_i X_i = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) X_i = 0$$

Συντελεστής προσδιορισμού

Μέρος της μεταβλητότητας που παρατηρείται στις τιμές της Y οφείλεται στις μεταβολές της X και μέρος στις επιδράσεις των τυχαίων παραγόντων (u).

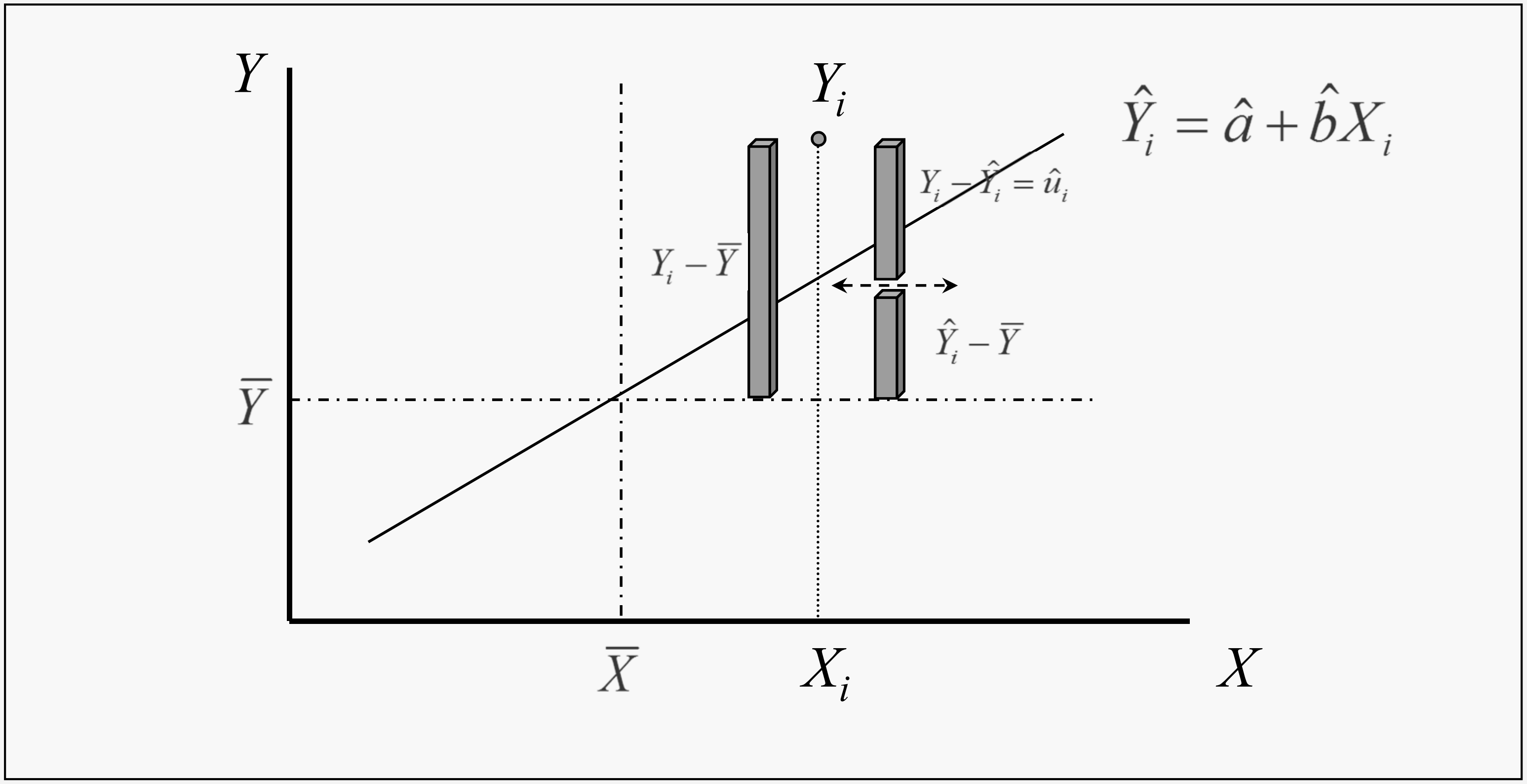
Πόση μεταβλητότητα της Y εξηγείται από την X και πόση από τυχαίους παράγοντες;

Μέτρο του βαθμού προσαρμογής της ευθείας παλινδρόμησης στις παρατηρήσεις του δείγματος.

Συμβολίζεται με R^2.

Ο αριθμητικός μέσος μιας μεταβλητής $(\bar{Y})$ είναι η καλύτερη πρόβλεψη όταν η μόνη διαθέσιμη πληροφορία είναι οι τιμές της ίδιας της μεταβλητής.

Η X μπορεί να θεωρηθεί ότι ερμηνεύει την Y στον βαθμό που συμβάλλει στην πρόβλεψή της πέρα από τον μέσο $\bar{Y}$



$$\sum \left(Y_i - \overline{Y}\right)^2$$

Συνολικό Άθροισμα Τετραγώνων *(Total Sum of Squares)* - ***TSS***. Αντιπροσωπεύει την συνολική μεταβλητότητα της ***Y***.

$$\sum\left(\hat{Y}_i - \overline{Y}\right)^2$$

Ερμηνευόμενο Άθροισμα Τετραγώνων
(Regression Sum of Squares) - ***RSS*** .
Αντιπροσωπεύει την μεταβλητότητα της Y που ερμηνεύεται από την ευθεία παλινδρόμησης.

$$\sum\left(Y_i - \hat{Y}_i\right)^2 = \sum \hat{u}_i^2$$

Άθροισμα Τετραγώνων των Σφαλμάτων
(Error Sum of Squares) - ***ESS*** .
Αντιπροσωπεύει την μεταβλητότητα της Y που δεν ερμηνεύεται από την ευθεία παλινδρόμησης.

Αποδεικνύεται ότι

$$TSS=RSS+ESS$$

Απόδειξη:

$$\sum\left(Y_i-\overline{Y}\right)^2=\sum\left(\hat{Y}_i-\overline{Y}+\hat{u}_i\right)^2$$

$$=\sum\left(\hat{Y}_i-\overline{Y}\right)^2+\sum\hat{u}_i^{\,2}+2\sum\hat{u}_i\left(\hat{Y}_i-\overline{Y}\right)$$

$$=\sum\left(\hat{Y}_i-\overline{Y}\right)^2+\sum\hat{u}_i^{\,2}+2\sum\hat{u}_i\hat{Y}_i-2\overline{Y}\sum\hat{u}_i$$

Αλλά: $\sum\hat{u}_i=0$ και $\sum\hat{u}_i\hat{Y}_i=\sum\hat{u}_i\left(\hat{\beta}_0+\hat{\beta}_1X_i\right)=$

$$=\hat{\beta}_0\sum\hat{u}_i+\hat{\beta}_1\sum\hat{u}_iX_i=0$$

Άρα: $\sum\left(Y_i-\overline{Y}\right)^2=\sum\left(\hat{Y}_i-\overline{Y}\right)^2+\sum\hat{u}_i^{\,2}$

$$\Rightarrow TSS=RSS+ESS$$

Συντελεστής Προσδιορισμού

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}$$

$$R^2 = \frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2} = 1 - \frac{\sum\left(Y_i - \hat{Y}_i\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2}$$

$$0 \le R^2 \le 1$$

$$R^2 = \frac{\sum_{i=1}^{n} \hat{y}_i^2}{\sum_{i=1}^{n} y_i^2} = \frac{\sum_{i=1}^{n}\left(\hat{\beta}_1 x_i\right)^2}{\sum_{i=1}^{n} y_i^2} = \hat{\beta}_1^{\,2} \frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} y_i^2} = \hat{\beta}_1^{\,2} \frac{\sum_{i=1}^{n} x_i^2 \Big/ (n-1)}{\sum_{i=1}^{n} y_i^2 \Big/ (n-1)} = \hat{\beta}_1^{\,2} \frac{S_X^2}{S_Y^2}$$

Παράδειγμα:

	X_i	Y_i	$\hat{Y}_i$	$\left(Y_i - \hat{Y}_i\right)^2$	$\left(Y_i - \overline{Y}\right)^2$	$\left(X_i - \overline{X}\right)^2$
	150	100	101,38	1,907	367,361	625
	160	112	108,50	12,283	51,361	225
	170	114	115,61	2,590	26,694	25
	180	129	122,72	39,391	96,694	25
	190	115	129,84	220,170	17,361	225
	200	145	136,95	64,764	667,361	625
Α. μέσος	**175**	**119,6**				
Άθροισμα				**341,104**	**1226,8**	**1750**

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2} = 1 - \frac{341,104}{1226,8} = 0,722$$

$$R^2 = \hat{\beta}_1^{\,2}\,\frac{S_X^2}{S_Y^2} = 0,711^2\,\frac{1750/5}{1226,8\,/\,5} = 0,722$$

Συντελεστής συσχέτισης

Αντιπροσωπεύει τον βαθμό ***γραμμικής συσχέτισης*** ανάμεσα σε δύο μεταβλητές χωρίς να ενδιαφέρεται για την αιτιώδη σχέση που μπορεί να υπάρχει μεταξύ τους.

Συμβολίζεται με το ρ και συνδέεται άμεσα με τον συντελεστή προσδιορισμού αφού $\rho = \sqrt{R^2}$

$$\rho = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2}} = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sqrt{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}\sqrt{n\sum_{i=1}^{n} Y_i^2 - \left(\sum_{i=1}^{n} Y_i\right)^2}}$$

$$\rho = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)}\sqrt{\operatorname{var}(Y)}}$$

Ιδιότητες

1. $-1 \leq \rho \leq 1$

2. $\rho_{X,Y} = \rho_{Y,X}$

3. Είναι ένα μέτρο ***γραμμικής συσχέτισης***. Έτσι όταν $\rho=0$ δεν συνεπάγεται αναγκαστικά ότι οι δύο μεταβλητές είναι *ανεξάρτητες*.

Παράδειγμα:

	X_i	Y_i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
	150	100	-25	-19.17	479,167	625	367,361
	160	112	-15	-7.17	107,500	225	51,361
	170	114	-5	-5.17	25,833	25	26,694
	180	129	5	9.83	49,167	25	96,694
	190	115	15	-4.17	-62,500	225	17,361
	200	145	25	25.83	645,833	625	667,361
Άθροισμα					**1245**	**1750**	**1226,833**

$$\rho = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2}} = \frac{1245}{\sqrt{1750 \cdot 1226,833}} = 0,8496$$

Έλεγχος Υποθέσεων

Έλεγχος υπόθεσης

Με βάση τις εκτιμήσεις που προκύπτουν από τα στοιχεία ενός ***δείγματος*** ελέγχουμε διάφορες ***υποθέσεις*** που αναφέρονται στις πραγματικές τιμές του ***πληθυσμού***.

Παράδειγμα : Αν $\hat{\beta}_1 = 2,3$

μπορεί να ισχύει στον πληθυσμό $\beta_1 = 1$

Αν η απόσταση μεταξύ της εκτίμησης και της υποθετικής τιμής είναι μικρή τότε δεν απορρίπτουμε την υπόθεση. Αν η απόσταση είναι μεγάλη την απορρίπτουμε.

- ***Διατύπωση της υπόθεσης***
- Πώς μετράμε την απόσταση ?
  Υπολογισμός της κατάλληλης στατιστικής
- ***Κριτήρια απόρριψης***

Η μηδενική και η εναλλακτική υπόθεση

Η ***μηδενική υπόθεση*** (H_0) αναφέρεται σε μια τιμή ή σε ένα διάστημα τιμών ενός συντελεστή που συνήθως δεν είναι οι αναμενόμενες.

Η ***εναλλακτική υπόθεση*** (H_1) αναφέρεται στην τιμή ή στο διάστημα τιμών του συντελεστή που ισχύουν αν δεν γίνει αποδεκτή η μηδενική υπόθεση.

Παράδειγμα:

$$H_0 : \beta_1 = 0 \qquad\qquad H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 < 0 \qquad\qquad H_1 : \beta_1 \neq 0$$

Υπολογισμός της κατάλληλης στατιστικής

Ως στατιστική ελέγχου χρησιμοποιείται η στατιστική t

$$\frac{\hat{\beta}_i - \beta_i}{S_{\hat{\beta}_i}}$$

Όσο μεγαλύτερο είναι το μέγεθος αυτό τόσο μικρότερη είναι η ***πιθανότητα*** να διαπράξουμε σφάλμα τύπου I. Δηλαδή, η πιθανότητα να κάνουμε λάθος αν απορρίψουμε την H_0 γίνεται πολύ μικρή.

Για τον καθορισμό της πιθανότητας πρέπει να γνωρίζουμε την κατανομή του $\hat{\beta}_i$

Υπόθεση: $u_i \sim N\left(0, \sigma^2\right)$

Κάθε ***γραμμικός μετασχηματισμός*** μιας τυχαίας μεταβλητής που ακολουθεί την κανονική κατανομή είναι επίσης τυχαία μεταβλητή με κανονική κατανομή

- Το Y ακολουθεί την ***κανονική κατανομή*** αφού είναι γραμμικός μετασχηματισμός του u
- Τα $\hat{\beta}_0$ και $\hat{\beta}_1$ είναι ***τυχαίες μεταβλητές*** που ακολουθούν την ***κανονική κατανομή*** αφού είναι γραμμικοί μετασχηματισμοί του Y

Αφού $\hat{\beta}_i \sim N\left(\beta_i, \sigma^2_{\hat{\beta}_1}\right)$

$\sigma^2_{\hat{\beta}_1}$ άγνωστο

Στη θέση του χρησιμοποιείται $S^2_{\hat{\beta}_1}$

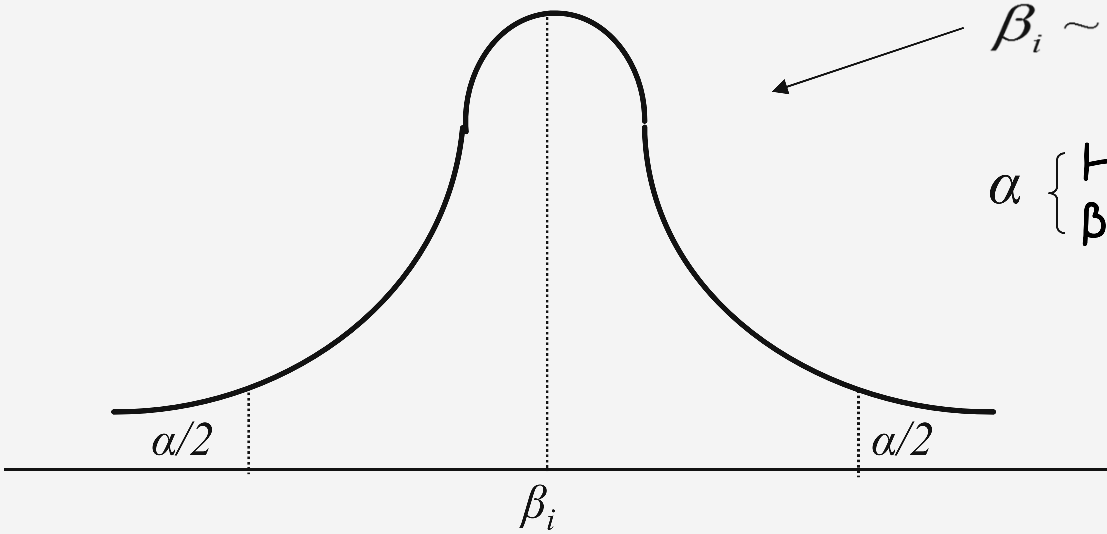
$$\boxed{\frac{\hat{\beta}_i - \beta_i}{S_{\hat{\beta}_i}} \sim t_{n-2}}$$

Κριτήρια αποδοχής και απόρριψης

Περιοχή αποδοχής : Το απαραίτητο εύρος τιμών της στατιστικής για να γίνει αποδεκτή η H_0

Περιοχή απόρριψης : Το απαραίτητο εύρος τιμών της στατιστικής για να απορριφθεί η H_0

$$\hat{\beta}_i \sim N\left(\beta_i, \sigma^2_{\hat{\beta}_1}\right)$$



α $\begin{cases} \text{Η πιθανότητα ένα } \hat{\beta}_i \text{ να} \\ \text{βρεθεί εκτός ορίου} \end{cases}$

Η πιθανότητα να απορρίψουμε μια σωστή υπόθεση

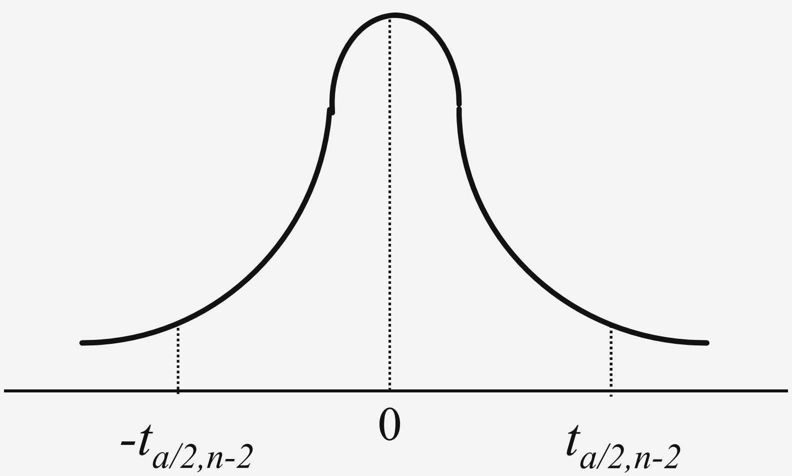
Διαδικασία ελέγχου

Διατυπώνεται η H_o και η H_1

Υπολογίζεται $t = \dfrac{\hat{\beta}_i - \beta_i}{S_{\hat{\beta}_i}}$ [Η απόσταση της εκτίμησης από την υποθετική τιμή]

Αποφασίζεται το επίπεδο σημαντικότητας α

Καθορίζεται το $t_{a/2,n-2}$ από τον σχετικό πίνακα



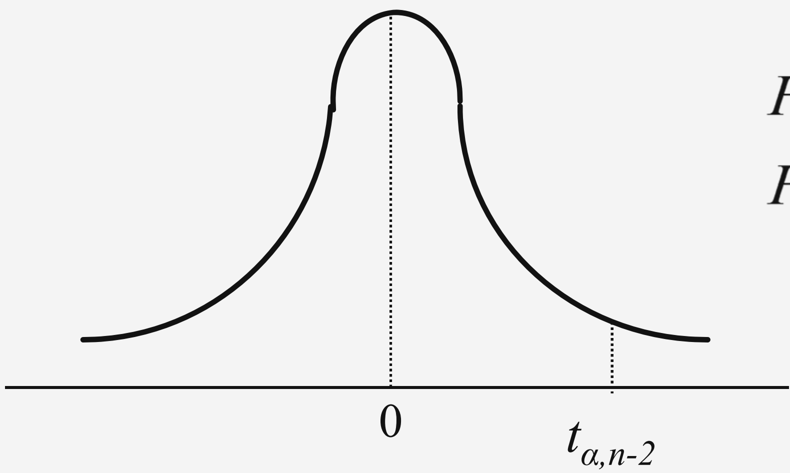
Αν $|t| > t_{a/2,n-2}$

Η H_0 απορρίπτεται

Αν $|t| \leq t_{a/2,n-2}$

Η H_0 δεν απορρίπτεται

Μονόπλευρος έλεγχος



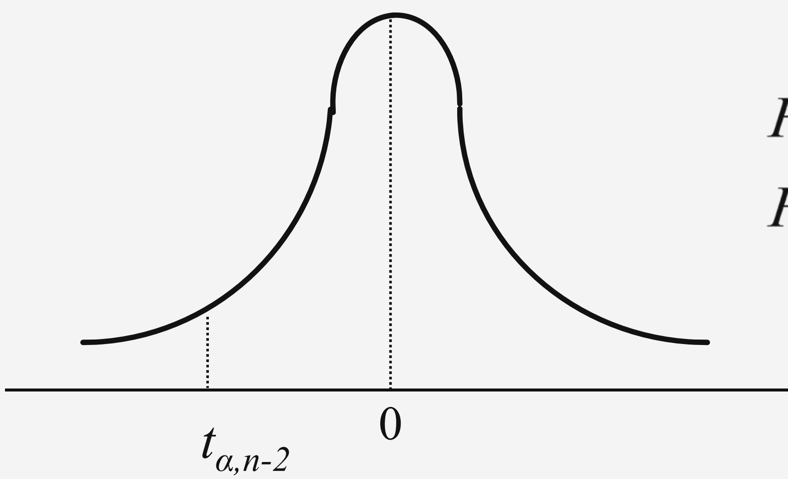
$$H_0 : \beta_i = 0$$
$$H_1 : \beta_i > 0$$

Αν $t > t_{a,n-2}$

Η H_0 απορρίπτεται

Αν $t \le t_{a,n-2}$

Η H_0 δεν απορρίπτεται



$$H_0 : \beta_i = 0$$
$$H_1 : \beta_i < 0$$

Αν $t < -t_{a,n-2}$

Η H_0 απορρίπτεται

Αν $t > -t_{a,n-2}$

Η H_0 δεν απορρίπτεται

$$u \sim N\left(0, \sigma^2\right)$$

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

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2_{\hat{\beta}_0}\right)$$

$$S^2_{\hat{\beta}_0} = \frac{\sum_{i=1}^{n} X^2}{n \sum_{i=1}^{n} x_i^2} \hat{\sigma}^2$$

$$\hat{\beta}_1 \sim N\left(\beta_1, \sigma^2_{\hat{\beta}_1}\right)$$

$$S^2_{\hat{\beta}_1} = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n} x_i^2}$$

Στατιστικός έλεγχος του β_0

$$t_{n-2} = \frac{\hat{\beta}_0 - \beta_0}{S_{\hat{\beta}_0}} = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\sigma}^2 \left(\sum_{i=1}^{n} X^2 \Big/ n\sum_{i=1}^{n} x_i^2 \right)}}$$

Παράδειγμα:

$$\hat{Y}_i = -5,333 + 0,7114 X_i$$
$$(38,81) \quad (0,221)$$

$$H_0 : \beta_0 = 0$$
$$H_1 : \beta_0 \neq 0$$

$(\alpha=0,05)$

$$t = \frac{\hat{\beta}_0 - \beta_0}{S_{\hat{\beta}_0}} = \frac{-5,33 - 0}{38,81} = -0,1374$$

Από τους πίνακες $t_{0,025/4} = 2,776$

Η υπόθεση H_0 **δεν** απορρίπτεται

Στατιστικός έλεγχος του β_1

$$t_{n-2}=\frac{\hat{\beta}_1-\beta_1}{S_{\hat{\beta}_1}}=\frac{\hat{\beta}_1-\beta_1}{\sqrt{\hat{\sigma}^2\Big/\sum_{i=1}^{n}x_i^2}}$$

Παράδειγμα:

$$\hat{Y}_i=-5,333+0,7114X_i$$
$$(38,81)\quad(0,221)$$

$$H_0:\beta_1=0$$
$$H_1:\beta_1\neq 0$$

$$t=\frac{\hat{b}-b}{S_{\hat{b}}}=\frac{0,7114-0}{0,2207}=3,222$$

Από τους πίνακες $t_{0,025/4}=2,776$

Η υπόθεση H_0 απορρίπτεται

Διαστήματα εμπιστοσύνης για το β_0 και β_1

$$P\left(-t_{\alpha/2,n-2} \le t \le t_{\alpha/2,n-2}\right) = 1-\alpha$$

$$P\left(\hat{\beta}_i - S_{\hat{\beta}_i} \cdot t_{\alpha/2,n-2} \le \beta_i \le \hat{\beta}_i + S_{\hat{\beta}_i} \cdot t_{\alpha/2,n-2}\right) = 1-\alpha$$

$$\hat{\beta}_i \pm S_{\hat{\beta}_i} \cdot t_{\alpha/2,n-2}$$

Παράδειγμα:

$$\hat{Y}_i = -5,333 + 0,7114 X_i$$

$$(38,81) \quad (0,221)$$

$$\hat{\beta}_1 - S_{\hat{\beta}_1} \cdot t_{\alpha/2,n-2} \le \beta_1 \le \hat{\beta}_1 + S_{\hat{\beta}_1} \cdot t_{\alpha/2,n-2}$$

$$0,7114 - 0,2207 \cdot 2,776 \le \beta_1 \le 0,7114 + 0,2207 \cdot 2,776$$

$$0,0987 \le \beta_1 \le 1,324$$

$$\hat{\beta}_0 - S_{\hat{\beta}_0} \cdot t_{\alpha/2,n-2} \le \beta_0 \le \hat{\beta}_0 + S_{\hat{\beta}_0} \cdot t_{\alpha/2,n-2}$$

$$-5,33 - 38,81 \cdot 2,776 \le \beta_0 \le -5,33 + 38,81 \cdot 2,776$$

$$-113,07 \le \beta_0 \le 102,41$$

Έλεγχος κανονικότητας διαταρακτικού όρου

Η βασική υπόθεση για τον έλεγχο υποθέσεων είναι ότι ο διαταρακτικός όρος κατανέμεται κανονικά.

Ένας έλεγχος της υπόθεσης αυτής γίνεται με τον έλεγχο *Jarque-Bera* με βάση την στατιστική:

$$JB = n\left(\frac{S^2}{6} + \frac{(k-3)^2}{24}\right)$$

Όπου S είναι η ασσυμετρία και k η κύρτωση της κατανομής των καταλοίπων, όπου:

$$S = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^3}{n \cdot s^3}, \quad k = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^4}{n \cdot s^4} \text{ και } s = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n}}$$

$JB \sim \chi^2_{2,\alpha}$

π.χ. για S=0 και k=1,7314

$\chi^2_{2,(0,05)} = 5{,}99 > JB = 0{,}402$ άρα δεν απορρίπτεται η μηδέν υπόθεση ότι τα κατάλοιπα ακολουθούν την κανονική κατανομή.

Άσκηση

Έστω η ζήτηση ενός αγαθού:

Υ = ποσότητα σε κιλά	100	82	86	90	60	70
Χ = τιμή κιλού €	13	13,5	13,2	13	14	13,7

Α) Εάν η συνάρτηση ζήτησης είναι γραμμικής μορφής $Y=\beta_0+\beta_1 X+u$ να υπολογιστούν τα β_0 και β_1.

$$\hat{\beta}_1 = \frac{\sum xy}{\sum x^2} = \frac{-28}{0.82} = -34.146$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} = 81.33 - (-34.146) \cdot 13.4 = 538.89$$

	Y	X	y	x	xy	x^2
	100	13	18.67	-0.40	-7.47	0.16
	82	13.5	0.67	0.10	0.07	0.01
	86	13.2	4.67	-0.20	-0.93	0.04
	90	13	8.67	-0.40	-3.47	0.16
	60	14	-21.33	0.60	-12.80	0.36
	70	13.7	-11.33	0.30	-3.40	0.09
Α. Μέσοι	**81.33**	**13.40**				
Αθροίσματα					**-28.00**	**0.82**

Β) Να υπολογιστεί η ελαστικότητα ζήτησης στο σημείο $\left(\bar{X},\bar{Y}\right)$.

Η παλινδρόμηση στο δείγμα είναι: $\hat{Y} = 538.94 - 34.146X$

$$\varepsilon_{YX} = \frac{dY}{dX}\frac{\bar{X}}{\bar{Y}} = \hat{\beta}_1 \frac{\bar{X}}{\bar{Y}} = -34.146\frac{13.4}{81.33} = -5.63$$

Γ) Σχολιάστε τα αποτελέσματα από οικονομική άποψη

Όταν μας ζητάνε να σχολιάσουμε τα αποτελέσματα από οικονομική άποψη αναφερόμαστε στα πρόσημα των εκτιμήσεων και το μέγεθος τους.

Τα πρόσημα είναι αυτά που αναμένονται από την θεωρία. Θετικό πρόσημα για τον σταθερό όρο και αρνητική κλίση της ευθείας.

Δηλαδή, για τιμή μηδενική ($X=0$) η ζητούμενη ποσότητα είναι θετική (και πολύ μεγαλύτερη από το μέσο όρο 538.94>>81.33). Επίσης η αρνητική κλίση της ευθείας ($\hat{\beta}_1$) δείχνει ότι αύξηση της τιμής οδηγεί σε μείωση της ζητούμενης ποσότητας π.χ. αύξηση της τιμής κατά 1 € μειώνει την ζητούμενη ποσότητα κατά 34.146 κιλά.

Επίσης η ζήτηση του αγαθού είναι ελαστική αφού η ελαστικότητα είναι μεγαλύτερη από την μονάδα.

Δ) Υπολογίστε τον συντελεστή προσδιορισμού, τα τυπικά σφάλματα των συντελεστών και κάντε ελέγχους σημαντικότητας των συντελεστών.

$$R^2 = \frac{\sum_{i=1}^{n} \hat{y}_i^2}{\sum_{i=1}^{n} y_i^2} = \frac{956.09}{1029.33} = 0.9288$$

	Y	X	x^2	y^2	X^2	$\hat{Y}$	$\hat{y}^2$	$\hat{u}^2$
	100	13	0.16	348.44	169	95.04	187.93	24.58
	82	13.5	0.01	0.44	182.25	77.97	11.32	16.25
	86	13.2	0.04	21.78	174.24	88.21	47.33	4.90
	90	13	0.16	75.11	169	95.04	187.93	25.42
	60	14	0.36	455.11	196	60.90	417.68	0.80
	70	13.7	0.09	128.44	187.69	71.14	103.91	1.30
Α. Μέσοι	**81.33**	**13.40**						
Αθροίσματα			**0.82**	**1029.33**	**1078.18**		**956.09**	**73.25**

Επίσης η ζήτηση του αγαθού είναι ελαστική αφού η ελαστικότητα είναι μεγαλύτερη από την μονάδα.

Δ) Υπολογίστε τον συντελεστή προσδιορισμού, τα τυπικά σφάλματα των συντελεστών και κάντε ελέγχους σημαντικότητας των συντελεστών.

$$S_{\hat{\beta}_0} = \sqrt{\frac{\sum_{i=1}^{n} X^2}{n\sum_{i=1}^{n} x_i^2}\hat{\sigma}^2} = \sqrt{\frac{1078.18}{6 \cdot 0.82} \cdot 18.31} = \sqrt{4012.49} = 63.34$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-2} = \frac{73.25}{4} = 18.31$$

$$S_{\hat{\beta}_1} = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n} x_i^2}} = \sqrt{\frac{18.31}{0.82}} = \sqrt{22.33} = 4.72$$

	Y	X	x^2	y^2	X^2	$\hat{Y}$	$\hat{y}^2$	$\hat{u}^2$
	100	13	0.16	348.44	169	95.04	187.93	24.58
	82	13.5	0.01	0.44	182.25	77.97	11.32	16.25
	86	13.2	0.04	21.78	174.24	88.21	47.33	4.90
	90	13	0.16	75.11	169	95.04	187.93	25.42
	60	14	0.36	455.11	196	60.90	417.68	0.80
	70	13.7	0.09	128.44	187.69	71.14	103.91	1.30
Α. Μέσοι	**81.33**	**13.40**						
Αθροίσματα			**0.82**	**1029.33**	**1078.18**		**956.09**	**73.25**

Επίσης η ζήτηση του αγαθού είναι ελαστική αφού η ελαστικότητα είναι μεγαλύτερη από την μονάδα.

Δ) Υπολογίστε τον συντελεστή προσδιορισμού, τα τυπικά σφάλματα των συντελεστών και κάντε ελέγχους σημαντικότητας των συντελεστών.

Η παλινδρόμηση στο δείγμα μπορεί να γραφεί: $\hat{Y} = 538.94 - 34.146X$

$$\begin{array}{cc} (63.34) & (4.72) \end{array}$$

$$H_0 : \beta_0 = 0 \qquad\qquad H_0 : \beta_1 = 0$$

$$H_1 : \beta_0 \neq 0 \qquad\qquad H_1 : \beta_1 \neq 0$$

$$|t| = \left| \frac{\hat{\beta}_0 - \beta_0}{s_{\hat{\beta}_0}} \right| = \left| \frac{538.88 - 0}{63.33} \right| = 8.5 > t_{4,(0.05)} = 2.132$$

$$|t| = \left| \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}} \right| = \left| \frac{-34.146 - 0}{4.72} \right| = 7.23 > t_{4,(0.05)} = 2.132$$

Ε) Να κατασκευάσετε ένα 95% διάστημα εμπιστοσύνης για τον συντελεστή β_1.

$$\hat{\beta}_1 - S_{\hat{\beta}_1} \cdot t_{\alpha/2,n-2} \le \beta_1 \le \hat{\beta}_1 + S_{\hat{\beta}_1} \cdot t_{\alpha/2,n-2} \qquad t_{0.025,4} = 2.776$$

$$-34,146 - 4.72 \cdot 2.776 \le \beta_1 \le -34,146 + 4.72 \cdot 2.776$$

$$-47.248 \le \beta_1 \le -21.043$$

ΣΤ) Ελέγξτε αν ο διαταρακτικός όρος ακολουθεί την κανονική κατανομή.

$$S = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^3}{n \cdot s^3}, \quad k = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^4}{n \cdot s^4} \text{ και } s = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n}}$$

$$JB = n\left(\frac{S^2}{6} + \frac{(k-3)^2}{24}\right)$$

ΣΤ) Ελέγξτε αν ο διαταρακτικός όρος ακολουθεί την κανονική κατανομή.

$$S=\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^3}{n\cdot s^3}=\frac{0.11}{6\cdot 0.37^3}=0.36$$

$$s=\sqrt{\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{n}}=\sqrt{\frac{0.82}{6}}=0.37$$

$$k=\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^4}{n\cdot s^4}=\frac{0.19}{6\cdot 0.37^4}=1.69$$

$$JB=n\left(\frac{S^2}{6}+\frac{(k-3)^2}{24}\right)=6\left(\frac{0.36^2}{6}+\frac{(1.69-3)^2}{24}\right)=6(0.0216+0.109)=0.7836$$

$\chi^2_{2,\,(0{,}05)}$=5.99 > JB=0.7836 άρα δεν απορρίπτεται η μηδενική υπόθεση ότι τα κατάλοιπα ακολουθούν την κανονική κατανομή.

	Y	X	x^2	x^3	x^4
	100	13	0.16	-0.06	0.03
	82	13.5	0.01	0.00	0.00
	86	13.2	0.04	-0.01	0.00
	90	13	0.16	-0.06	0.03
	60	14	0.36	0.22	0.13
	70	13.7	0.09	0.03	0.01
Α. Μέσοι	**81.33**	**13.40**			
Αθροίσματα			**0.82**	**0.11**	**0.19**