ΟΙΚΟΝΟΜΕΤΡΙΑ

Θεωρία: 01

Βιολέττα Δάλλα

Τμήμα Οικονομικών Επιστημών

Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

Εισαγωγή

- Οικονομετρία (Econometrics) είναι ο τομέας των Οικονομικών Επιστήμων που αναλύει οικονομικά φαινόμενα με τη χρήση στατιστικών ή άλλων ποσοτικών μεθόδων.
- Οι βασικοί σκοποί της Οικονομετρίας είναι
    - Να ποσοτικοποιηθούν και να ελεγχθούν αιτιακές σχέσεις μεταξύ οικονομικών μεταβλητών, π.χ. αν το εισόδημα μεταβληθεί πόσο μεταβάλλεται η κατανάλωση, αν η ανεργία μεταβληθεί πόσο μεταβάλλεται το ΑΕΠ, αν το εισόδημα εξαρτάται από το φύλο, κ.λπ.
    - Να γίνουν προβλέψεις οικονομικών μεταβλητών, π.χ. για δεδομένες δαπάνες πόσο είναι το εισόδημα, ποιο θα είναι το ΑΕΠ το επόμενο τρίμηνο, ποια θα είναι η ανεργία τα επόμενα δύο έτη, κ.λπ.
- Το βασικότερο εργαλείο στην Οικονομετρία είναι η Ανάλυση Παλινδρόμησης (Regression Analysis).

- Ανάλογα με την πηγή, τα στατιστικά στοιχεία ή δεδομένα χωρίζονται σε
  - Πειραματικά δεδομένα (experimental data)
  - Παρατηρήσιμα δεδομένα (observational data)
- Ανάλογα με τη μεταβλητή, οι βασικές κατηγορίες δεδομένων είναι
  - Διαστρωματικά δεδομένα (cross-sectional data), π.χ. το ετήσιο εισόδημα 100 ατόμων για το 2020, το τριμηνιαίο ΑΕΠ 30 χωρών για το 2021Q1, η ημερήσια απόδοση 20 μετοχών για την 01/03/2021.
  - Δεδομένα χρονολογικών σειρών (time series data), π.χ. το ετήσιο εισόδημα ενός ατόμου για την περίοδο 2000 – 2020, το τριμηνιαίο ΑΕΠ μίας χώρας για την περίοδο 2010Q1 – 2021Q1, η ημερήσια απόδοση μίας μετοχής για την περίοδο 04/01/2021 – 01/03/2021.
  - Μεικτά δεδομένα (panel data), π.χ. το ετήσιο εισόδημα 100 ατόμων για την περίοδο 2000 – 2020, το τριμηνιαίο ΑΕΠ 30 χωρών για την περίοδο 2010Q1 – 2021Q1, η ημερήσια απόδοση 20 μετοχών για την περίοδο 04/01/2021 – 01/03/2021.

Ανάλυση παλινδρόμησης

Υπόδειγμα πολλαπλής γραμμικής παλινδρόμησης (multiple linear regression model)

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_K X_K + u$$

όπου

- Y είναι η εξαρτημένη μεταβλητή (dependent variable).
- $X_1, ..., X_K$ είναι οι ανεξάρτητες ή ερμηνευτικές μεταβλητές (independent variables or regressors). K είναι ο αριθμός των ερμηνευτικών μεταβλητών.
- $\beta_0, \beta_1, ..., \beta_K$ είναι οι συντελεστές παλινδρόμησης (regression coefficients).
- β_0 είναι ο σταθερός όρος (constant term or intercept) και $\beta_1, ..., \beta_K$ είναι οι συντελεστές κλίσης (slope coefficients).
- $\beta_0 + \beta_1 X_1 + ... + \beta_K X_K$ είναι η γραμμή ή εξίσωση παλινδρόμησης (regression line or equation).
- u είναι το σφάλμα ή ο διαταρακτικός όρος (error or disturbance term).

Για $K = 1$: απλό γραμμικό υπόδειγμα παλινδρόμησης (simple linear regression model).

- Π.χ. Y = εισόδημα, X_1 = εκπαίδευση, X_2 = προϋπηρεσία, X_3 = φύλο, X_4 = ηλικία, X_5 = τόπος διαμονής

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + u$$

Π.χ. Y = επιτόκιο, X_1 = πληθωρισμός, X_2 = παραγωγικό κενό

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

Π.χ. Y = ζήτηση, X_1 = τιμή, X_2 = τιμή ανταγωνισμού, X_3 = εισόδημα, X_4 = διαφημιστική δαπάνη

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u$$

Π.χ. Y = οριακό κόστος, X_1 = ποσότητα, $X_2 = X_1^2$ = τετραγωνική ποσότητα

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

Π.χ. Y = κατανάλωση, X_1 = εισόδημα, X_2 = κατανάλωση προηγούμενης περιόδου

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

- Δεδομένα δείγματος μεγέθους T

  $Y_1, Y_2, ..., Y_T$ για την Y,

  $X_{11}, X_{21}, ..., X_{T1}$ για την X_1,

  $\vdots$

  $X_{1K}, X_{2K}, ..., X_{TK}$ για την X_K

- Υπόδειγμα παλινδρόμησης με δεδομένα

$$Y_t = \beta_0 + \beta_1 X_{t1} + ... + \beta_K X_{tK} + u_t, \quad t = 1, 2, ..., T$$

Συμβολισμός με πίνακες

$$Y = X\beta + u$$

όπου

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}, \quad X = \begin{pmatrix} 1 & X_{11} & \dots & X_{1K} \\ 1 & X_{21} & \dots & X_{2K} \\ \vdots & \vdots & & \vdots \\ 1 & X_{T1} & \dots & X_{TK} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix}$$

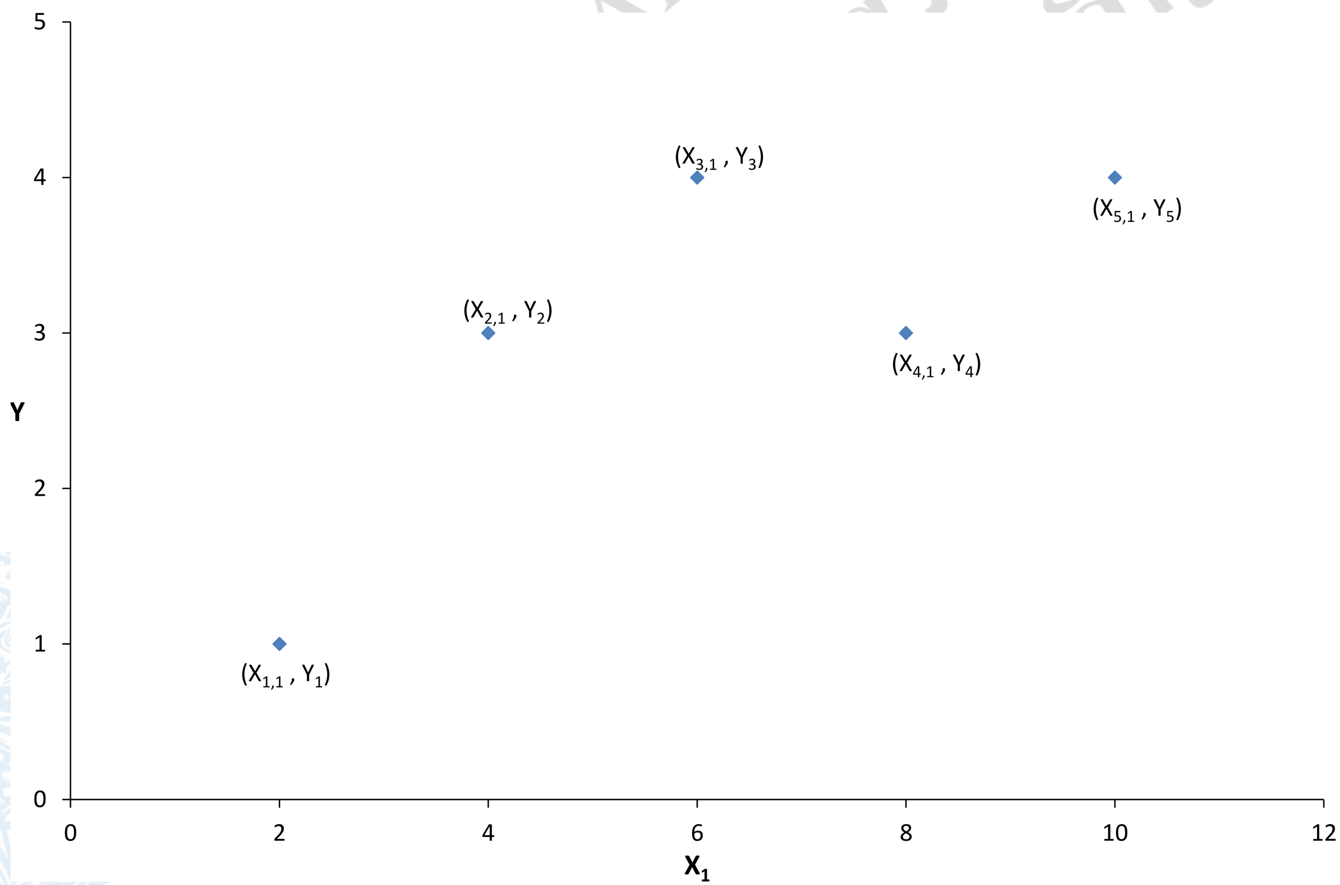
Π.χ. Απλή γραμμική παλινδρόμηση $Y_t = \beta_0 + \beta_1 X_{t1} + u_t$, $t = 1, ..., 5$

Δεδομένα

	A	B	C
1	t	X_1	Y
2	1	2	1
3	2	4	3
4	3	6	4
5	4	8	3
6	5	10	4

Δεδομένα σε πίνακες

$$Y = \begin{pmatrix} 1 \\ 3 \\ 4 \\ 3 \\ 4 \end{pmatrix} \quad X = \begin{pmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 6 \\ 1 & 8 \\ 1 & 10 \end{pmatrix}$$



Βασικές υποθέσεις με μη στοχαστικές ερμηνευτικές μεταβλητές

A.1 $Y = X\beta + u$ είναι το σωστό υπόδειγμα και $E(u) = 0$ ή

$Y_t = \beta_0 + \beta_1 X_{t1} + ... + \beta_K X_{tK} + u_t$, $t = 1, ..., T$, είναι το σωστό υπόδειγμα και $E(u_t) = 0$, $t = 1, ..., T$.

A.2 X είναι πλήρους βαθμού και $T > K + 1$ ή

δεν υπάρχουν ακριβείς γραμμικές σχέσεις μεταξύ των $X_1, ..., X_K$ και $T > K + 1$.

A.3 X είναι μη στοχαστικός ή

$X_1, ..., X_K$ είναι μη στοχαστικές.

A.4 Ισχύει ότι $V(u) = \sigma^2 I$ ή

ισχύει ότι $V(u_t) = \sigma^2$, $t = 1, ..., T$ και $Cov(u_t, u_s) = 0$, $t, s = 1, ..., T$, $t \neq s$.

A.5 u ακολουθεί την κανονική κατανομή ή

u_t, $t = 1, ..., T$ ακολουθούν την κανονική κατανομή.

- Ερμηνεία συντελεστών παλινδρόμησης

Ισχύει ότι

$$
\begin{aligned}
E(Y) &\overset{\mathbf{A.1}}{=} E(\beta_0 + \beta_1 X_1 + ... + \beta_K X_K + u) \overset{\mathbf{A.3}}{=} \beta_0 + \beta_1 X_1 + ... + \beta_K X_K + E(u) \\
&\overset{\mathbf{A.1}}{=} \beta_0 + \beta_1 X_1 + ... + \beta_K X_K
\end{aligned}
$$

Άρα

$$E(Y) = \beta_0 \ \text{ αν } \ X_1 = ... = X_K = 0$$

και για $j = 1, ..., K$

$$\frac{\partial E(Y)}{\partial X_j} = \beta_j \text{ σταθ.} \Rightarrow \frac{\Delta E(Y)}{\Delta X_j} = \beta_j \Rightarrow \Delta E(Y) = \beta_j \ \text{ αν } \ \Delta X_j = 1$$

β_0 : Αν οι $X_1, ..., X_K$ είναι μηδενικές, η (μέση) Y είναι β_0 μονάδες μέτρησης.

β_j, $j = 1, ..., K$: Αν η X_j αυξηθεί κατά 1 μονάδα μέτρησης, ενώ οι υπόλοιπες $X_1, ..., X_{j-1}, X_{j+1}, ..., X_K$ παραμείνουν σταθερές, τότε η (μέση) Y αυξάνεται κατά β_j μονάδες μέτρησης.

Μέθοδος ελαχίστων τετραγώνων (OLS)

Η μέθοδος ελαχίστων τετραγώνων OLS (ordinary least squares) ελαχιστοποιεί το άθροισμα των τετραγώνων

$$S(\widetilde{\beta}) = \sum_{t=1}^{T} \left(Y_t - \widetilde{\beta}_0 - \widetilde{\beta}_1 X_{t1} - ... - \widetilde{\beta}_K X_{tK} \right)^2 = \left(Y - X\widetilde{\beta} \right)' \left(Y - X\widetilde{\beta} \right)$$

ως προς όλες τις πιθανές τιμές $\widetilde{\beta}$.

Η συνθήκη $1^{ου}$ βαθμού δίνει τις κανονικές εξισώσεις (normal equations)

$$\frac{dS(\widehat{\beta})}{d\widetilde{\beta}} = 0 \Leftrightarrow X'X\widehat{\beta} = X'Y$$

Αν η υπόθεση A.2 ισχύει, τότε ο πίνακας $X'X$ είναι αντιστρέψιμος και άρα

$$\widehat{\beta} = \left(X'X \right)^{-1} X'Y$$

Η συνθήκη $2^{ου}$ βαθμού ισχύει

$$\frac{d^2 S(\widehat{\beta})}{d\widetilde{\beta}' d\widetilde{\beta}} = 2X'X > 0$$

- Εκτιμητής OLS του β

$$\widehat{\beta} = \left(X'X\right)^{-1} X'Y$$

- Υπολογισμένες τιμές (fitted values)

$$\widehat{Y} = X\widehat{\beta} \quad \text{ή} \quad \widehat{Y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 X_{t1} + \ldots + \widehat{\beta}_K X_{tK},\ t = 1, 2, \ldots, T$$

- Κατάλοιπα (residuals)

$$\widehat{u} = Y - \widehat{Y} \quad \text{ή} \quad \widehat{u}_t = Y_t - \widehat{Y}_t,\ t = 1, 2, \ldots, T$$

Ισχύει ότι

$$\overline{\widehat{u}} = 0 \quad \text{και} \quad X'\widehat{u} = 0$$

- Εκτιμητής OLS του σ^2

$$s^2 = \frac{1}{T - K - 1} \widehat{u}'\widehat{u} = \frac{1}{T - K - 1} \sum_{t=1}^{T} \widehat{u}_t^2$$

- Εκτιμητής OLS του πίνακα διακυμάνσεων-συνδιακυμάνσεων του $\widehat{\beta}$

$$\widehat{V}\left(\widehat{\beta}\right) = s^2 \left(X'X\right)^{-1}$$

- Τυπικό σφάλμα (standard error) του $\widehat{\beta}_j$

$$s_{\widehat{\beta}_j} = \sqrt{\widehat{V}\left(\widehat{\beta}_j\right)} = \sqrt{\left[\widehat{V}\left(\widehat{\beta}\right)\right]_{j+1,j+1}}$$

όπου $\left[\widehat{V}\left(\widehat{\beta}\right)\right]_{j+1,j+1}$ είναι το $j+1$ διαγώνιο στοιχείο του $\widehat{V}\left(\widehat{\beta}\right)$.

- Γενικά, η διακύμανση του $\widehat{\beta}_j$ μειώνεται όταν
  - Το μέγεθος του δείγματος T αυξάνεται.
  - Η διακύμανση της ερμηνευτικής μεταβλητής X_j αυξάνεται.
  - Η διακύμανση του σφάλματος u μειώνεται.
  - Η συσχέτιση μεταξύ της ερμηνευτικής μεταβλητής X_j και των υπολοίπων ερμηνευτικών μεταβλητών μειώνεται.

- Συνολικό άθροισμα τετραγώνων SST (total sum of squares)

$$SST = \sum_{t=1}^{T}(Y_t - \overline{Y})^2$$

- Άθροισμα τετραγώνων παλινδρόμησης SSR (explained sum of squares)

$$SSR = \sum_{t=1}^{T} (\widehat{Y}_t - \overline{Y})^2$$

- Άθροισμα τετραγώνων καταλοίπων SSE (residual sum of squares)

$$SSE = \sum_{t=1}^{T} \widehat{u}_t^2$$

- Ισχύει ότι

$$SST = SSR + SSE$$

- Συντελεστής προσδιορισμού (coefficient of determination)

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

  - Είναι ένα μέτρο προσαρμογής του υποδείγματος στα δεδομένα.
  - Ισχύει ότι $0 \leq R^2 \leq 1$.
  - Γενικά, ο R^2 δεν μειώνεται όταν το μέγεθος του δείγματος T ή ο αριθμός των

ερμηνευτικών μεταβλητών K αυξάνεται. Ο R^2 δεν είναι κατάλληλο μέτρο σύγκρισης υποδειγμάτων για την εξαρτημένη μεταβλητή που έχουν διαφορετικό αριθμό ερμηνευτικών μεταβλητών.

- Ερμηνεία συντελεστή προσδιορισμού
  R^2 : Το $(100 \cdot R^2)\,\%$ της συνολικής μεταβλητότητας της Y ερμηνεύεται από τη γραμμική επίδραση των $X_1, ..., X_K$.

- Διορθωμένος συντελεστής προσδιορισμού (adjusted coefficient of determination)

$$\overline{R}^2 = 1 - \frac{SSE/(T-K-1)}{SST/(T-1)}$$

  - Ισχύει ότι $\overline{R}^2 \leq 1$ και ο $\overline{R}^2$ μπορεί να πάρει αρνητικές τιμές.
  - Ισχύει ότι $\overline{R}^2 < R^2$ και ισχύει ότι $\overline{R}^2 = R^2$ μόνο όταν $R^2 = 1$ ή όταν $T \to \infty$.
  - Ο $\overline{R}^2$ λαμβάνει υπόψη το μέγεθος του δείγματος T και τον αριθμό των ερμηνευτικών μεταβλητών K. Ο $\overline{R}^2$ είναι κατάλληλο μέτρο σύγκρισης υποδειγμάτων για την εξαρτημένη μεταβλητή που έχουν διαφορετικό αριθμό ερμηνευτικών μεταβλητών.

Στατιστικές ιδιότητες των OLS εκτιμητών $\widehat{\beta}, s^2$ και $\widehat{V}(\widehat{\beta})$ (ΣΙπα)

1. Αν οι Α.1-Α.3 ισχύουν, ο $\widehat{\beta}$ είναι αμερόληπτος και συνεπής εκτιμητής του β.

2. Αν οι Α.1-Α.4 ισχύουν, ο πίνακας διακυμάνσεων-συνδιακυμάνσεων του $\widehat{\beta}$ είναι ο $V(\widehat{\beta}) = \sigma^2(X'X)^{-1}$.

3. Θεώρημα Gauss-Markov: Αν οι Α.1-Α.4 ισχύουν, ο $\widehat{\beta}$ είναι άριστος γραμμικός, αμερόληπτος και συνεπής εκτιμητής του β.

4. Αν οι Α.1-Α.4 ισχύουν, ο s^2 είναι αμερόληπτος και συνεπής εκτιμητής του σ^2.

5. Αν οι Α.1-Α.4 ισχύουν, ο $\widehat{V}(\widehat{\beta})$ είναι αμερόληπτος και συνεπής εκτιμητής του $V(\widehat{\beta})$.

6. Αν οι Α.1-Α.5 ισχύουν, $\widehat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$, $(T-K-1)\frac{s^2}{\sigma^2} \sim \chi^2_{T-K-1}$ και οι $\widehat{\beta}, s^2$ είναι ανεξάρτητοι.

7. Αν οι Α.1-Α.5 ισχύουν, ο $\widehat{\beta}$ είναι άριστος, αμερόληπτος και συνεπής εκτιμητής του β.

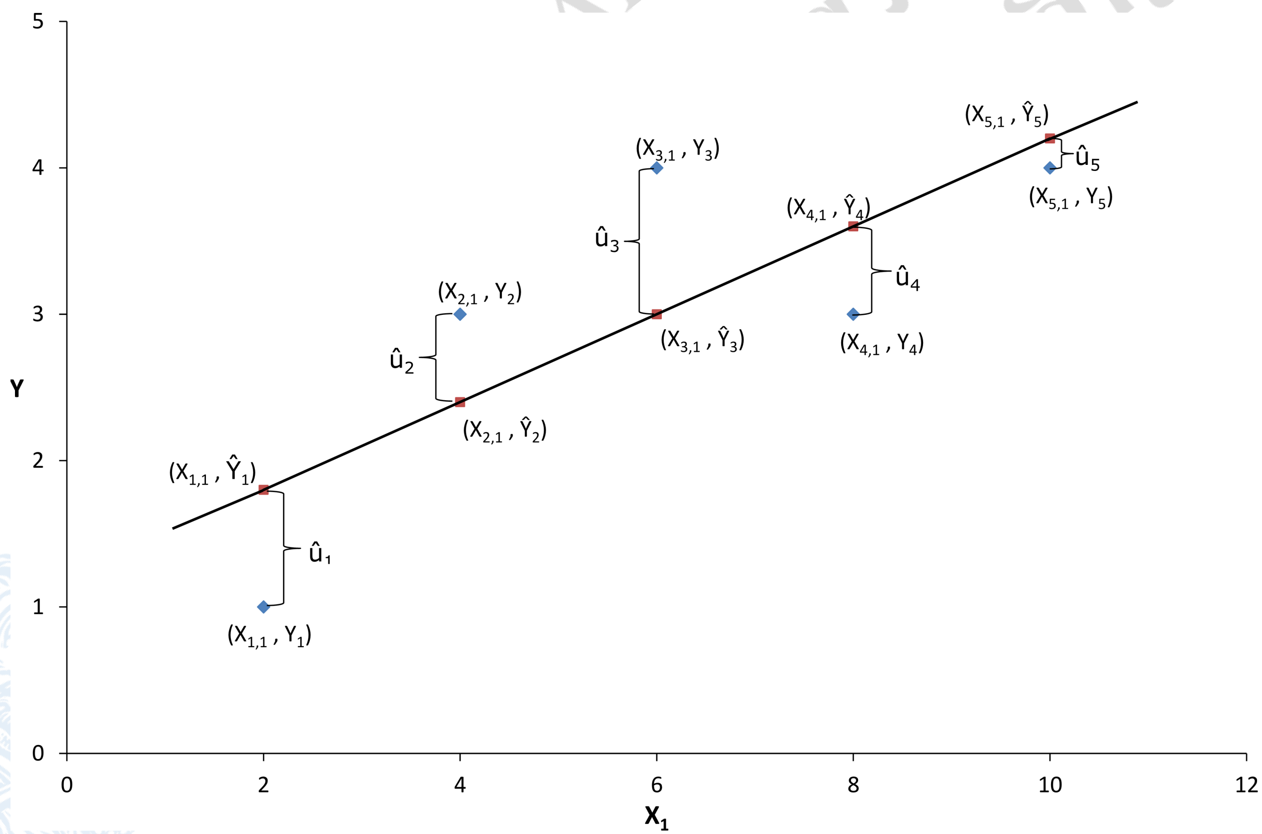
Π.χ. Απλή γραμμική παλινδρόμηση $Y_t = \beta_0 + \beta_1 X_{t1} + u_t$, $t = 1, \ldots, 5$

Δεδομένα

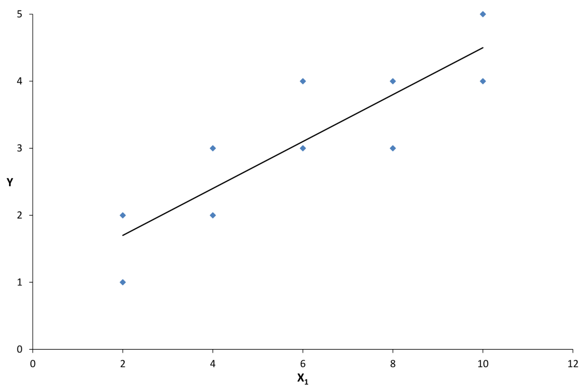
	A	B	C
1	t	X_1	Y
2	1	2	1
3	2	4	3
4	3	6	4
5	4	8	3
6	5	10	4

Δεδομένα σε πίνακες

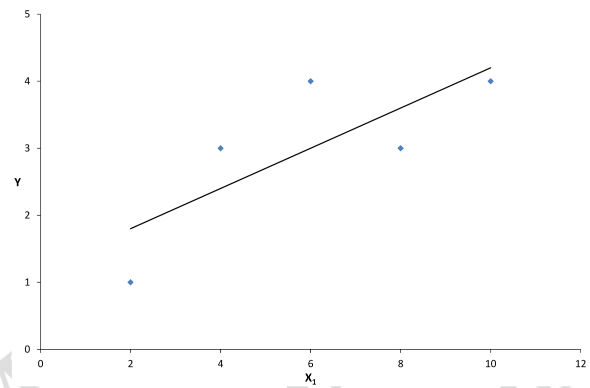
$$Y = \begin{pmatrix} 1 \\ 3 \\ 4 \\ 3 \\ 4 \end{pmatrix} \quad X = \begin{pmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 6 \\ 1 & 8 \\ 1 & 10 \end{pmatrix}$$



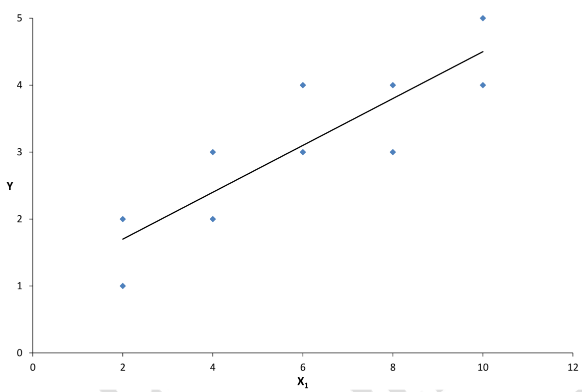
Μεγαλύτερο T



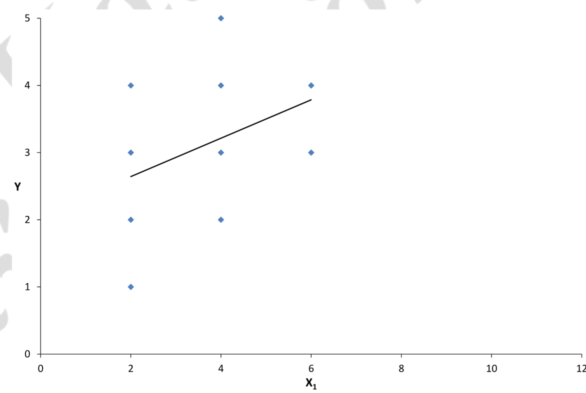
Μικρότερο T



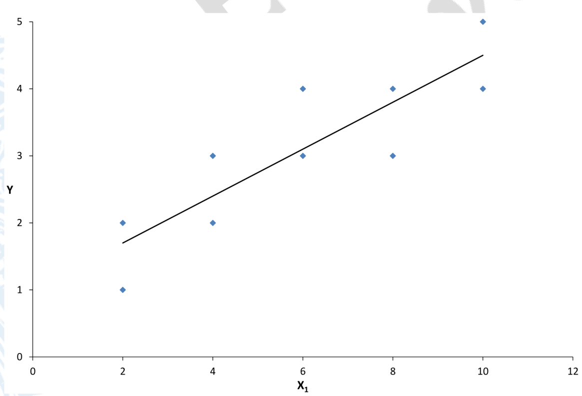
Μεγαλύτερη $V(X_1)$



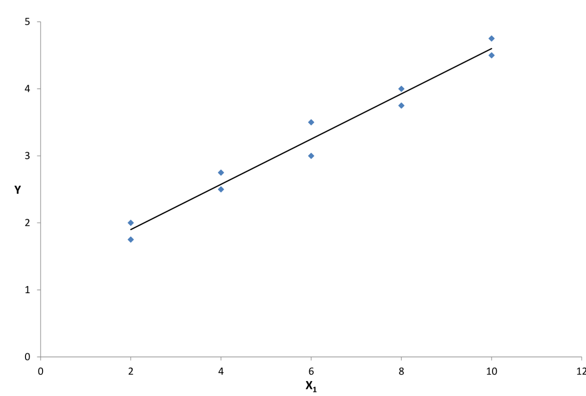
Μικρότερη $V(X_1)$



Μεγαλύτερη $V(u)$



Μικρότερη $V(u)$



Παράρτημα

Έστω διάνυσμα τυχαίων μεταβλητών $z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$, π.χ. $u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix}$, $\widehat{\beta} = \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \vdots \\ \widehat{\beta}_K \end{pmatrix}$.

Η μέση τιμή του z είναι

$$E(z) = \begin{pmatrix} E(z_1) \\ E(z_2) \\ \vdots \\ E(z_n) \end{pmatrix}$$

Ο πίνακας διακυμάνσεων-συνδιακυμάνσεων του z είναι

$$V(z) = \begin{pmatrix} V(z_1) & Cov(z_1, z_2) & \cdots & \cdots & Cov(z_1, z_n) \\ Cov(z_1, z_2) & V(z_2) & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & V(z_{n-1}) & Cov(z_{n-1}, z_n) \\ Cov(z_1, z_n) & \cdots & \cdots & Cov(z_{n-1}, z_n) & V(z_n) \end{pmatrix}$$