ΟΙΚΟΝΟΜΕΤΡΙΑ

Ενότητα 2: Απλό γραμμικό υπόδειγμα και η μέθοδος ελαχίστων τετραγώνων

Ιωάννης Βενέτης
Αναπληρωτής Καθηγητής

Τμήμα Οικονομικών Επιστημών
Σχολή Οργάνωσης και Διοίκησης Επιχειρήσεων
Πανεπιστήμιο Πατρών

Σκοποί Ενότητας

⇒ «Αιτιότητα» και συσχέτιση στα πλαίσια της εισαγωγής στην Οικονομετρία

⇒ Παρουσίαση του βασικού οικονομετρικού υποδείγματος $y_i = \alpha + \beta x_i + u_i$ και των «κλασσικών» υποθέσεων/παραδοχών πριν την εκτίμηση των παραμέτρων α, β, σ^2 και ειδικά της παραμέτρου (οικονομικού) ενδιαφέροντος β

⇒ Εισαγωγή και παρουσίαση της μεθόδου εκτίμησης ελαχίστων τετραγώνων (εκτίμησης των παραμέτρων του απλού γραμμικού υποδείγματος)

⇒ Εισαγωγή σε μέτρα προσαρμογής του υποδείγματος και ειδικά στον συντελεστή προσδιορισμού R^2 ως μέτρο επεξήγησης της μεταβλητότητας της εξαρτημένης μεταβλητής y_i

Περιεχόμενα ενότητας

2.1 Αιτιότητα και πλασματικές συσχετίσεις

- Ένα χαρακτηριστικό λάθος στο οποίο μπορεί να υποπέσουμε στην εμπειρική οικονομετρική πρακτική, είναι η αναγνώριση σχέσεων αιτιότητας μεταξύ μεταβλητών (κυρίως χρονοσειρών) που εμφανίζουν γραμμική συσχέτιση.

- Η ύπαρξη συσχέτισης μεταξύ δύο χρονοσειρών (ή και μεταξύ δύο μεταβλητών διαστρωματικών δεδομένων) μπορεί να θεωρηθεί είτε ως επιβεβαίωση της θεωρίας η οποία αναπτύσσει και εξηγεί την πιθανή σχέση αιτίας - αιτιατού δύο μεταβλητών είτε, αντίστροφα και σύμφωνα με προϋποθέσεις, μπορεί να δώσει το έναυσμα για την εξέλιξη μιας θεωρίας.

- Σε καμμία όμως περίπτωση δεν δύναται η ύπαρξη και μόνο γραμμικής συσχέτισης να αποτελέσει θεωρία ή αλλιώς επιβεβαίωση σχέσεων αιτίας-αιτιατού, αφού συχνά η εμφανιζόμενη δειγματική γραμμική συσχέτιση είναι «πλασματική».

2.1 Αιτιότητα και πλασματικές συσχετίσεις

- Τα παραδείγματα πλασματικών συσχετίσεων είναι άφθονα ειδικότερα αν οι υπο-εξέταση μεταβλητές εμφανίζουν **«τάσεις»** (δεδομένα χρονοσειρών) και γενικότερα αν οι υπο-εξέταση μεταβλητές εμπίπτουν στην κατηγορία των **μη στάσιμων χρονοσειρών** (θέμα στο οποίο θα επανέλθουμε).

Χαρακτηριστικά αναφέρουμε

(α) την περίπτωση χρονοσειρών που υπόκεινται σε «στοχαστικές τάσεις» και όπου δύο ή περισσότερες ανεξάρτητες μεταξύ τους μεταβλητές εμφανίζουν (γραμμικούς) διαχρονικούς δεσμούς λόγω της ύπαρξης «στοχαστικών τάσεων»

(β) την περίπτωση χρονοσειρών που εμφανίζουν άλλου είδους τάση και πιο συγκεκριμένα προσδιοριστικές τάσεις (ο χρόνος εισέρχεται ως ερμηνευτική μεταβλητή) (γ) την περίπτωση χρονοσειρών που εμφανίζουν πλασματική συσχέτιση λόγω παράλειψης μίας κοινής «βοηθητικής» μεταβλητής που επιδρά στις εξεταζόμενες μεταβλητές

(δ) την περίπτωση μεταβλητών διαστρωματικών δεδομένων όπου υπάρχει αμφίδρομη σχέση μεταξύ των μεταβλητών η οποία δυσκολεύει την ακριβή μέτρηση (ταυτοποίηση) της σχέσης των μεταβλητών ενδιαφέροντος.

- Στις περιπτώσεις αυτές, η εύρεση μίας «βοηθητικής» μεταβλητής που συσχετίζεται με μία από τις μεταβλητές ενδιαφέροντος και η υιοθέτηση της πληροφόρησης που μεταφέρει στο υπόδειγμα καθίσταται ζωτικής σημασίας

2.2 Το απλό διμεταβλητό υπόδειγμα παλινδρόμησης και οι «κλασσικές υποθέσεις»

Το απλούστερο οικονομετρικό υπόδειγμα, το απλό γραμμικό υπόδειγμα ή υπόδειγμα παλινδρόμησης, λαμβάνει τη μορφή

$$Y_i = \alpha + \beta X_i + u_i$$

όπου u_i είναι ο διαταρακτικός όρος.
Ένα πρώτο (και βασικό) μέτρο περίληψης ή περιγραφής μιας κατανομής είναι ο υπό συνθήκη ή δεσμευμένος μέσος[1] $E(Y|\mathbb{X})$ τον οποίο θεωρούμε συνάρτηση τουλάχιστον των τυχαίων μεταβλητών $X_1, \ldots, X_n$, δηλαδή το σύνολο πληροφόρησης $\mathbb{X}$ δίνεται τουλάχιστον ως

$$\mathbb{X} = \{X_1, \ldots, X_n\}$$

[1] Αποτελεί μέτρο της κεντρικής ροπής της δεσμευμένης κατανομής.

2.2 Το απλό διμεταβλητό υπόδειγμα παλινδρόμησης και οι «κλασσικές υποθέσεις»

Μάλιστα - απλοποιώντας σημαντικά την ανάλυσή μας - υποθέτουμε ότι η υπό συνθήκη αναμενόμενη τιμή $E(Y_i|\mathbb{X})$ είναι γραμμική συνάρτηση της ερμηνευτικής μεταβλητής X_i με τις παραμέτρους του σταθερού όρου α και της κλίσης β να είναι σταθερές για κάθε i. Δηλαδή, υποθέτουμε ότι

$$E\left(Y_i\,|\mathbb{X}\right) = \alpha + \beta X_i \tag{1}$$

με

$$\alpha, \beta \text{ αμετάβλητα } \forall i \tag{2}$$

Σύμφωνα με την υπόθεση (1), ο υπό συνθήκη μέσος $E(Y_i|\mathbb{X})$ περιγράφεται ακριβώς από μία συνάρτηση ευθείας. Το **γραμμικό σφάλμα παλινδρόμησης** (linear regression error) ή **διαταρακτικός όρος** ορίζεται ως η διαφορά της Y_i από τον υπό συνθήκη μέσο της, οπότε

2.2 Το απλό διμεταβλητό υπόδειγμα παλινδρόμησης και οι «κλασσικές υποθέσεις»

$$Y_i - E(Y_i|\mathbb{X}) = u_i \tag{3}$$

και με συνδυασμό των (1), (2) και (3) έχουμε

$$\begin{aligned} Y_i &= E(Y_i|\mathbb{X}) + u_i \\ &= \alpha + \beta X_i + u_i \end{aligned} \tag{4}$$

Παρατηρήστε ότι το **γραμμικό σφάλμα παλινδρόμησης** - εξ'ορισμού - θα ικανοποιεί την παρακάτω συνθήκη **ισχυρής εξωγένειας** της ερμηνευτικής μεταβλητής

$$E(u_i|\mathbb{X}) = 0 \tag{5}$$

2.2 Το απλό διμεταβλητό υπόδειγμα παλινδρόμησης και οι «κλασσικές υποθέσεις»

αφού

$$\begin{aligned} E(u_i|\mathbb{X}) &= E([Y_i - E(Y_i|\mathbb{X})]|\mathbb{X}) \\ &= E(Y_i|\mathbb{X}) - E(E(Y_i|\mathbb{X})|\mathbb{X}) \\ &= E(Y_i|\mathbb{X}) - E(Y_i|\mathbb{X}) \\ &= 0 \end{aligned}$$

κάτι που συνεπάγεται και την ισότητα

$$E(u_i) = 0 \tag{6}$$

αφού από τον **νόμο των επαναλαμβανόμενων προσδοκιών**[2] **(ν.ε.π)**

$$E(u_i) = E(E(u_i|\mathbb{X})) = E(0) = 0$$

[2] Law of iterated expectations. Δείτε το παράρτημα στατιστικής για κατανόηση του βασικού αυτού νόμου.

Άρα, συνοψίζοντας, στο απλούστερο δυνατό (γραμμικό) διμεταβλητό υπόδειγμα

$$Y_i = \alpha + \beta X_i + u_i$$

εμπεριέχονται οι παρακάτω υποθέσεις οι οποίες στο εξής θα ονομάζονται **«κλασσικές υποθέσεις»:**

Περίπτωση 1.

Η ερμηνευτική μεταβλητή του υποδείγματος, X_i, είναι σταθερή σε επαναλαμβανόμενα δείγματα

Υπόθεση 1. Ο μη δεσμευμένος μέσος της εξαρτημένης μεταβλητής $E(Y_i)$ είναι γραμμική συνάρτηση της ερμηνευτικής μεταβλητής

2.2 Το απλό διμεταβλητό υπόδειγμα παλινδρόμησης και οι «κλασσικές υποθέσεις»

Υπόθεση 2. Οι παράμετροι α,β (συντελεστές υποδείγματος) είναι σταθερές δηλαδή αμετάβλητες ή αλλιώς δεν εξαρτώνται από τον υποδείκτη i ή t

Υπόθεση 3. $E(u_i) = 0$, μηδενική αναμενόμενη τιμή του διαταρακτικού όρου ή του σφάλματος παλινδρόμησης

Υπόθεση 4. $Var(u_i) = E(u_i^2) = \sigma^2$, $\forall i$, ομοσκεδαστικότητα

Υπόθεση 5. $Cov(u_i, u_j) = 0$, $\forall i \neq j$, απουσία συσχέτισης (ή αυτοσυσχέτισης όταν έχουμε δεδομένα χρονοσειρών) του διαταρακτικού όρου

Υπόθεση 6. από κοινού κανονικότητα των διαταρακτικών όρων

$$u_i \sim N(E(u_i), Var(u_i))$$

άρα

$$u_i \sim N(0, \sigma^2)$$

2.2 Το απλό διμεταβλητό υπόδειγμα παλινδρόμησης και οι «κλασσικές υποθέσεις»

Περίπτωση 2.

Η ερμηνευτική μεταβλητή του υποδείγματος, X_i, είναι στοχαστική (δηλαδή είναι τυχαία μεταβλητή)

Υπόθεση 1. Ο δεσμευμένος μέσος της εξαρτημένης μεταβλητής $E(Y_i|\mathbb{X})$ είναι γραμμική συνάρτηση της ερμηνευτικής μεταβλητής

Υπόθεση 2. Οι παράμετροι α, β (συντελεστές υποδείγματος) είναι σταθερές δηλαδή αμετάβλητες ή αλλιώς δεν εξαρτώνται από τον υποδείκτη i ή t

Υπόθεση 3. $E(u_i|\mathbb{X}) = 0$, μηδενική δεσμευμένη αναμενόμενη τιμή του διαταρακτικού όρου ή του σφάλματος παλινδρόμησης ως προς την X_i, $\forall i$.

Υπόθεση 4. $Var(u_i|\mathbb{X}) = E(u_i^2|\mathbb{X}) = \sigma^2$, $\forall i$, δεσμευμένη ομοσκεδαστικότητα ως προς την X_i, $\forall i$

2.2 Το απλό διμεταβλητό υπόδειγμα παλινδρόμησης και οι «κλασσικές υποθέσεις»

Υπόθεση 5. $Cov(u_i, u_j|\mathbb{X}) = 0$, $\forall i \neq j$, απουσία δεσμευμένης συσχέτισης (ή δεσμευμένης αυτοσυσχέτισης όταν έχουμε δεδομένα χρονοσειρών) του διαταρακτικού όρου ως προς την X_i, $\forall i$

Υπόθεση 6. από κοινού δεσμευμένη κανονικότητα των διαταρακτικών όρων

$$u_i|\mathbb{X} \sim N(E(u_i), Var(u_i))$$

άρα

$$u_i|\mathbb{X} \sim N(0, \sigma^2)$$

2.2 Το απλό διμεταβλητό υπόδειγμα παλινδρόμησης και οι «κλασσικές υποθέσεις»

Οι τελευταίες τρεις υποθέσεις (υποθέσεις 4, 5 και 6) στην περίπτωση 1 συνοψίζονται στο συμβολισμό

$$u_i \sim N.i.d(0, \sigma^2)$$

δηλαδή οι διαταρακτικοί όροι κατανέμονται ως κανονικές και ανεξάρτητες τυχαίες μεταβλητές (normally and independently distributed) με μέσο μηδέν και διακύμανση σ^2.

2.3 Μέθοδος εκτίμησης ελαχίστων τετραγώνων (ΕΤ)

- Έστω το απλό γραμμικό υπόδειγμα,
$$Y_i = \alpha + \beta X_i + u_i, \ \ i = 1, \ldots, n \tag{7}$$
όπου Y_i η εξαρτημένη μεταβλητή και X_i η ερμηνευτική μεταβλητή.

- Αρχικός σκοπός μας είναι να εκτιμήσουμε τις παραμέτρους α, β αφού σε αυτές συνοψίζονται οι επιδράσεις της ερμηνευτικής μεταβλητής X_i επί της Y_i. Οι τιμές των παραμέτρων α, β είναι άγνωστες.

- Με βάση λοιπόν ένα περιορισμένο δείγμα τιμών που έχουμε στη διάθεσή μας για τις μεταβλητές Y_i, X_i θα προσπαθήσουμε να εκτιμήσουμε τις άγνωστες παραμέτρους.

- Οι εκτιμητές των α, β συμβολίζονται με $\hat{\alpha}$, $\hat{\beta}$ και αποτελούν μαθηματικούς τύπους (συναρτήσεις) που βασίζονται στα δεδομένα του δείγματος.

2.3 Μέθοδος εκτίμησης ελαχίστων τετραγώνων (ΕΤ)

- Θεωρητικά υπάρχει ένας τεράστιος αριθμός δειγμάτων που θα μπορούσαμε να λάβουμε υπόψη. Επιλέγοντας ένα νέο δείγμα (ίδιου μεγέθους) θα άλλαζε και η τιμή του εκτιμητή. Ουσιαστικά λοιπόν, αντιμετωπίζουμε τους εκτιμητές ως τυχαίες μεταβλητές και τις κατανομές στις οποίες υπόκεινται τις ονομάζουμε κατανομές δειγματοληψίας.
- Σε εισαγωγικό επίπεδο υπάρχουν σοβαροί στατιστικοί λόγοι να ξεκινήσουμε την «περιπέτεια» στην εκτίμηση υιοθετώντας την **μέθοδο εκτίμησης Ελαχίστων Τετραγώνων (ΕΤ)**. Η μέθοδος θα δώσει εκτιμητές των παραμέτρων α, β που θα συμβολίζουμε με $\hat{\alpha}_{ET}$, $\hat{\beta}_{ET}$.
- Η διαδικασία εκτίμησης ονομάζεται «ελάχιστα τετράγωνα» αφού βασίζεται στην ελαχιστοποίηση του αθροίσματος των τετραγώνων των **καταλοίπων** (residuals) $\hat{u}_i$,

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

2.3 Μέθοδος εκτίμησης ελαχίστων τετραγώνων (ΕΤ)

Σημείωση 1: Αν γνωρίζουμε δύο εκτιμητές $\hat{\alpha}$, $\hat{\beta}$ τότε οι όροι

$$\hat{u}_i = Y_i - \hat{\alpha} - \hat{\beta} X_i \ , \ \forall i$$

μπορούν να υπολογιστούν από τις ισότητες

$$\begin{aligned} \hat{u}_1 &= Y_1 - \hat{\alpha} - \hat{\beta} X_1 \\ \hat{u}_2 &= Y_2 - \hat{\alpha} - \hat{\beta} X_2 \\ &\vdots \\ \hat{u}_n &= Y_n - \hat{\alpha} - \hat{\beta} X_n \end{aligned}$$

και ονομάζονται **κατάλοιπα** (όχι διαταρακτικοί όροι). Στην ουσία, τα κατάλοιπα $\hat{u}_i$, $i = 1, \ldots, n$ αποτελούν εκτιμήσεις των διαταρακτικών όρων.

2.3 Μέθοδος εκτίμησης ελαχίστων τετραγώνων (ΕΤ)

Σημείωση 2: οι εκτιμητές ΕΤ λύνουν το πρόβλημα ελαχιστοποίησης

$$\min_{\hat{\alpha},\hat{\beta}} S\left(\hat{\alpha},\hat{\beta}\right) \text{ όπου } S\left(\hat{\alpha},\hat{\beta}\right) = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\alpha} - \hat{\beta} X_i\right)^2$$

Η λύση των συνθηκών πρώτης τάξης δίνει

$$\hat{\alpha}_{ET} = \bar{Y} - \hat{\beta}\bar{X}$$

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

Ο υποδείκτης ΕΤ δηλώνει ότι ο εκτιμητής εξάγεται μέσω της μεθόδου των ελαχίστων τετραγώνων.

2.4 Συντελεστής προσδιορισμού R^2

- Παρατηρήστε ότι μετά την εκτίμηση το υπόδειγμα (7) γράφεται ως
$$Y_i = \hat{\alpha}_{ET} + \hat{\beta}_{ET} X_i + \hat{u}_i, \ \ i = 1, \ldots, n$$
και μπορούμε να γράψουμε ταυτοτικά ότι
$$Y_i = \hat{Y}_i + \hat{u}_i \tag{8}$$

- όπου $\hat{Y}_i = \hat{\alpha}_{ET} + \hat{\beta}_{ET} X_i$ συμβολίζει τις **προσαρμοσμένες ή εκτιμημένες τιμές** της Y_i (fitted values).

- Η σχέση (8) διαχωρίζει την εξαρτημένη μεταβλητή σε δύο συστατικά, το **προσαρμοσμένο ή ερμηνευμένο** μέρος $\hat{Y}_i$ και το **ανερμήνευτο μέρος** ή κατάλοιπα $\hat{u}_i$ (fitted values)

2.4 Συντελεστής προσδιορισμού R^2

- Αφαιρώντας $\bar{Y}$ και από τα δύο σκέλη της (8), υψώνοντας και τα δύο σκέλη στο τετράγωνο και προβαίνοντας στην απαραίτητη άλγεβρα καταλήγουμε στην ισότητα,

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2 + \sum_{i=1}^{n}\hat{u}_i^2$$

- η οποία συχνά γράφεται και ως

$$TSS = ESS + RSS$$

- δηλαδή το συνολικό άθροισμα τετραγώνων της εξαρτημένης μεταβλητής (total sum of squares) είναι ίσο με το άθροισμα των προσαρμοσμένων «(επεξηγημένων» ή ερμηνευμένων τετραγώνων) (explained sum of squares, TSS) και το άθροισμα των τετραγώνων των καταλοίπων (residual sum of squares, RSS).

Διαιρώντας και τα δύο σκέλη με TSS έχουμε

$$1 = \frac{ESS}{TSS} + \frac{RSS}{TSS} \Rightarrow \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

Ο λόγος $\frac{ESS}{TSS}$ συμβολίζεται με R^2 και ονομάζεται **συντελεστής προσδιορισμού**. Εκπροσωπεί το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής που «εξηγείται» από την παλινδρόμηση και ισχύει ότι $0 < R^2 < 1$. Ο **συντελεστής προσδιορισμού** είναι ένα μέτρο προσαρμογής του υποδείγματος στα δεδομένα.

Παράδειγμα. Έστω ότι με βάση ένα δείγμα και συγκεκριμένο υπόδειγμα υπολογίσαμε ότι $R^2 = 0.846$. Άρα το 84.6% της μεταβλητότητας της εξαρτημένης μεταβλητής εξηγείται από την παλινδρόμηση.

2.4 Συντελεστής προσδιορισμού R^2

Παρατήρηση 1

Καθώς το R^2 απομακρύνεται από το 0 προς την τιμή 1 θεωρούμε ότι η προσαρμογή είναι ολοένα και καλύτερη. Ωστόσο, δεν πρέπει να είμαστε εξαιρετικά αυστηροί στην κρίση μας ειδικότερα όταν αντιμετωπίζουμε μικρο-οικονομικές μεταβλητές, για παράδειγμα ωρομίσθια και έτη εκπαίδευσης ή ωρομίσθια και ηλικία. Η ετερογένεια των μικροοικονομικών μονάδων είναι τέτοια που καθιστά εξαιρετικά δύσκολο για μία και μόνο ανεξάρτητη μεταβλητή να ερμηνεύει «μεγάλο» μέρος της μεταβλητότητας της εξαρτημένης μεταβλητής.

Παρατήρηση 2

Επιπλέον, πρέπει να είμαστε ιδιαίτερα προσεκτικοί με την ερμηνεία του συντελεστή προσδιορισμού R^2 όταν έχουμε στη διάθεσή μας δεδομένα χρονοσειρών. Αποδεικνύεται ότι στο διμεταβλητό υπόδειγμα $R^2 = r^2$, δηλαδή ο συντελεστής προσδιορισμού R^2 είναι ίσος με το τετράγωνο της δειγματικής συσχέτισης εξαρτημένης και ανεξάρτητης μεταβλητής r^2. Συνεπώς, σε περιπτώσεις πλασματικής συσχέτισης ο συντελεστής R^2 θα εμφανίζεται εξαιρετικά υψηλός οδηγώντας σε εσφαλμένα συμπεράσματα.

Παρατήρηση 3
Επίσης, πρέπει να έχουμε υπόψιν ότι όταν η εξίσωση που εκτιμήθηκε με τη μέθοδο των ελαχίστων τετραγώνων δεν περιλαμβάνει σταθερό όρο, τότε το R^2 μπορεί να λάβει αρνητικές τιμές και δεν χρησιμοποιείται ως μέτρο προσαρμογής της εκτιμημένης γραμμής στα δεδομένα.

Παρατήρηση 4
Τέλος, ο συντελεστής προσδιορισμού R^2 μπορεί να χρησιμοποιηθεί για σύγκριση της προσαρμοστικότητας διαφορετικών υποδειγμάτων στα δεδομένα ή αλλιώς να συγκριθεί πόσο καλά εξηγούν διαφορετικά υποδείγματα τη μεταβλητότητα της **ίδιας** εξαρτημένης μεταβλητής. Για παράδειγμα, **δεν συγκρίνουμε** το R^2 δύο υποδειγμάτων με εξαρτημένη μεταβλητή την Y_i και $ln(Y_i)$ αντίστοιχα.

2.5 Ασκήσεις

Χρησιμοποιώντας βασική άλγεβρα και μερικές «έξυπνες» αντικαταστάσεις αθροισμάτων όπως

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \Leftrightarrow n\bar{X} = \sum_{i=1}^{n} X_i$$

μπορούμε να δείξουμε ότι
(α) όταν το γραμμικό υπόδειγμα περιλαμβάνει σταθερό όρο, τότε η μέθοδος ΕΤ δίνει κατάλοιπα με μηδενικό συνολικό άθροισμα άρα και με μηδενικό αριθμητικό μέσο ή αλλιώς

$$\sum_{i=1}^{n} \hat{u}_i = 0 \text{ άρα και } \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i = 0$$

(β) η ερμηνευτική μεταβλητή X_i είναι «ορθογώνια» με τα κατάλοιπα δηλαδή $\sum\limits_{i=1}^{n} X_i \hat{u}_i = 0$ ενώ ισχύει το ίδιο και για τις αποκλίσεις της ερμηνευτικής μεταβλητής από τον δειγματικό της μέσο, δηλαδή

$$\sum_{i=1}^{n} \left(X_i - \bar{X} \right) \hat{u}_i = \sum_{i=1}^{n} x_i \hat{u}_i = 0$$

Απάντηση
(α) Από την συνθήκη πρώτης τάξης (δηλαδή από την κανονική εξίσωση) του σταθερού όρου, έχουμε ότι

$$\sum_{i=1}^{n}\left(Y_i - \hat{\alpha}_{ET} - \hat{\beta}_{ET} X_i\right) = 0 \Leftrightarrow \sum_{i=1}^{n} \hat{u}_i = 0 \Leftrightarrow \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i = 0$$

Απάντηση

(β) Από την συνθήκη πρώτης τάξης για το συντελεστή κλίσης β έχουμε

$$\sum_{i=1}^{n}\left(Y_i-\hat{\alpha}_{ET}-\hat{\beta}_{ET}X_i\right)(X_i)=0 \Leftrightarrow \sum_{i=1}^{n}X_i\hat{u}_i=0$$

και

$$\begin{aligned}
\sum_{i=1}^{n}X_i\hat{u}_i &= 0 \Leftrightarrow \\
\sum_{i=1}^{n}\left(X_i-\bar{X}+\bar{X}\right)\hat{u}_i &= 0 \Leftrightarrow \\
\sum_{i=1}^{n}x_i\hat{u}_i+\bar{X}\sum_{i=1}^{n}\hat{u}_i &= 0 \Leftrightarrow \\
\sum_{i=1}^{n}x_i\hat{u}_i &= 0
\end{aligned}$$

αφού

$$\bar{X}\sum_{i=1}^{n}\hat{u}_i=\bar{X}\times 0=0$$

Σας δίνεται ένα δείγμα $n = 6$ παρατηρήσεων για τις μεταβλητές Y_i και X_i

Y_i	X_i
4	4
1.876	2
6.846	5
1.041	2
6.515	6
2.715	1

Χρησιμοποιήστε το πρόγραμμα Excel ή Gretl και
(α) Σχεδιάστε το διάγραμμα διασποράς των Y_i (κάθετος άξονας) και X_i (οριζόντιος άξονας)
(β) Υπολογίστε τους εκτιμητές ελαχίστων τετραγώνων $\hat{\alpha}_{ET}$ και $\hat{\beta}_{ET}$ ενός απλού διμεταβλητού υποδείγματος $Y_i = \alpha + \beta X_i + u_i$
(γ) Υπολογίστε και σχεδιάστε τα κατάλοιπα $\hat{u}_i$, $i = 1, \ldots, 6$.

Επιβεβαιώστε υπολογιστικά ότι $\sum_{i=1}^{n} \hat{u}_i = 0$

2.5 Ασκήσεις: Άσκηση 2

(δ) Επιβεβαιώστε υπολογιστικά ότι $\sum_{i=1}^{n} X_i \hat{u}_i = 0$

(ε) Υπολογίστε τις **προσαρμοσμένες ή προβλεπόμενες τιμές**

$$\hat{Y}_i = \hat{\alpha}_{ET} + \hat{\beta}_{ET} X_i$$

και σχεδιάστε σε ένα διάγραμμα μαζί τις τιμές των Y_i και $\hat{Y}_i$

(στ) Εισάγετε στο διάγραμμα διασποράς των Y_i (κάθετος άξονας) και X_i (οριζόντιος άξονας) τις προσαρμοσμένες τιμές $\hat{Y}_i$ (επίσης οριζόντιος άξονας). Θα εμφανιστούν ως μία ευθεία γραμμή

(ζ) Υπολογίστε και σχολιάστε το συντελεστή προσδιορισμού R^2

Τέλος ενότητας

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο **«Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Πατρών»** έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Σημειώματα

Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση **1.0**.

Σημείωμα Αναφοράς

Σημείωμα Αδειοδότησης

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.

Σημείωμα Χρήσης Έργων Τρίτων

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

- Ιωάννης Α. Βενέτης (2013). **Εισαγωγή στην Οικονομετρία**, GOTSIS Εκδόσεις, Πάτρα, ISBN 978-960-9427-25-8