Εισαγωγή στην οικονομετρία (1404)

Διάλεξη 11

26-27/10/2022

Άγγελος Αλεξόπουλος
Οικονομικό Πανεπιστήμιο Αθηνών -
Τμήμα Οικονομικής Επιστήμης
angelos@aueb.gr

Περίγραμμα Διάλεξης

- ✓Ιδιότητες εκτιμητών
  - ✓Γραμμικότητα
  - ✓Αμεροληψία
  - ✓Αποτελεσματικότητα
- ✓Θεώρημα Gauss-Markov
- ✓Εκτίμηση της διακύμανσης του διαταρακτικού όρου σ^2
- ✓Συντελεστής προσδιορισμού R^2
- ✓Άσκηση

Επανάληψη βασικών εννοιών

Απλό γραμμικό υπόδειγμα: $Y_i = \beta_0 + \beta_1 X_i + u_i, i = 1, \dots, n$

Εκτιμημένο γραμμικό υπόδειγμα: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \; \forall i$

Εκτιμητές ΕΤ: $\hat{\beta}_0 = \overline{\mathrm{Y}} - \hat{\beta}_1 \overline{\mathrm{X}}$ και $\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$ ή $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}$

Κατάλοιπα: $\hat{u}_i = Y_i - \hat{Y}_i$

Ιδιότητες: $\sum_{i=1}^{n} \hat{u}_i = 0, \sum_{i=1}^{n} X_i \hat{u}_i = 0, \sum_{i=1}^{n} \hat{Y}_i \hat{u}_i = 0$

Ισχύουν οι κλασικές υποθέσεις

Σημείωση: X_i προκαθορισμένες

Κλασικές υποθέσεις

Αν X_i είναι σταθερή σε επαναλαμβανόμενα δείγματα

1. Γραμμικότητα ως προς τις παραμέτρους
2. $E(u_i) = 0\ or\ E(u_i|x_i) = 0\ \forall i$
3. $var(u_i) = E(u_i^2) = \sigma^2$ or $var(u_i|x_i) = E(u_i^2|x_i) = \sigma^2, \forall i$
4. $cov(u_i, u_j) = 0, \forall i \neq j$
5. $u_i \sim N(0, \sigma^2), \forall i$

Στατιστικές ιδιότητες του εκτιμητή ΕΤ

✓ $\hat{\beta}_0$ και $\hat{\beta}_1$ τυχαίες μεταβλητές

- οι τιμές τους αλλάζουν από δείγμα σε δείγμα καθώς οι τιμές της εξαρτημένης μεταβλητής είναι τυχαίες
- κατανομή δειγματοληψίας, απαραίτητη για τη διεξαγωγή στατιστικών υποθέσεων για τους συντελεστές του υποδείγματος

✓ διαθέτουν στατιστικές ιδιότητες που τους καθιστούν ελκυστικούς έναντι άλλων εκτιμητών

- Γραμμικότητα
- Αμεροληψία
- Αποτελεσματικότητα

Γραμμικότητα (Linearity)

✓Ο εκτιμητής LS $\hat{\beta}_1$ αποτελεί γραμμική συνάρτηση των τιμών της εξαρτημένης μεταβλητής του υποδείγματος Y_i για όλα τα δείγματα

Απόδειξη: Γράψτε τον εκτιμητή $\hat{\beta}_1$ σε αποκλίσεις από τους μέσους ως εξής:

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

καθώς $\sum_{i=1}^{n}(X_i-\bar{X}) = 0$,

$$= \sum_{i=1}^{n}\left[\frac{(X_i-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right]Y_i = \sum_{i=1}^{n} c_i Y_i, \qquad (1)$$

όπου $c_i = \frac{(X_i-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$ θεωρούνται ως σταθερές τιμές. Η τελευταία σχέση (1) δείχνει ότι $\hat{\beta}_1$ αποτελεί γραμμικό συνδυασμό των παρατηρήσεων της εξαρτημένης μεταβλητής Y_i.

Αμεροληψία (unbiasedness)

✓Η μέση τιμή της κατανομής δειγματοληψίας του εκτιμητή $\hat{\beta}_1$ ισούται με την αληθινή (θεωρητική) τιμή αυτού στον πληθυσμό β_1, δηλαδή ισχύει:

$$E\left(\hat{\beta}_1\right) = \beta_1$$

Απόδειξη:

Πρώτα, γράψτε τον εκτιμητή $\hat{\beta}_1$ ως εξής:

$\hat{\beta}_1 = \sum_{i=1}^{n} c_i Y_i$ (βλέπε σχέση (1), γραμμικότητα εκτιμητή)

$= \sum_{i=1}^{n} c_i(\beta_0 + \beta_1 X_i + u_i)$, αντικαθιστώντας $Y_i = \beta_0 + \beta_1 X_i + u_i$,

$= \beta_0 \sum_{i=1}^{n} c_i + \beta_1 \sum_{i=1}^{n} c_i X_i + \sum_{i=1}^{n} c_i u_i$

Αμεροληψία (unbiasedness)

$$\hat{\beta}_1 = \beta_0 \sum_{i=1}^{n} c_i + \beta_1 \sum_{i=1}^{n} c_i X_i + \sum_{i=1}^{n} c_i u_i$$
$$= \beta_1 + \sum_{i=1}^{n} c_i u_i, \qquad (2)$$

καθώς ισχύουν τα ακόλουθα αποτελέσματα:

$$\sum_{i=1}^{n} c_i = \sum_{i=1}^{n} \left[\frac{(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right] = \frac{1}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \sum_{i=1}^{n}(X_i - \bar{X}) = 0$$

και

$$\sum_{i=1}^{n} c_i (X_i - \bar{X}) = \sum_{i=1}^{n} c_i X_i - \bar{X} \sum_{i=1}^{n} c_i = \sum_{i=1}^{n} c_i X_i \text{ καθώς } \sum_{i=1}^{n} c_i = 0$$

$$\sum_{i=1}^{n} c_i (X_i - \bar{X}) = \sum_{i=1}^{n} \left[\frac{(X_i - \bar{X})(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right] = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = 1 \qquad \text{επομένως}$$

$$\sum_{i=1}^{n} c_i X_i = 1$$

Αμεροληψία (unbiasedness)

$$\hat{\beta}_1 = \beta_1 + \sum_{i=1}^{n} c_i u_i \qquad (2)$$

Παίρνοντας την αναμενόμενη τιμή της σχέσης (2) έχουμε την ακόλουθη σχέση:

$$E(\hat{\beta}_1) = \beta_1 + \sum_{i=1}^{n} c_i E(u_i) = \beta_1, \qquad \text{καθώς } E(u_i) = 0$$

που αποδεικνύει ότι ο εκτιμητής ΕΤ $\hat{\beta}_1$ είναι ένας αμερόληπτος εκτιμητής του συντελεστή κλίσης β_1

- $E(\hat{\beta}_0) = \beta_0$ ο εκτιμητής ΕΤ $\hat{\beta}_0$ είναι ένας αμερόληπτος εκτιμητής του σταθερού όρου β_0

Αποτελεσματικότητα (efficiency)

Ένας εκτιμητής λέγεται αποτελεσματικός, αν αυτός έχει μικρότερη διακύμανση από κάποιον άλλον

$$var(\hat{\beta}_1) < var(\hat{\beta}_1^*)$$

όπου $\hat{\beta}_1$ ο εκτιμητής ΕΤ και $\hat{\beta}_1^*$ κάποιος άλλος εκτιμητής ($\hat{\beta}_1$ και $\hat{\beta}_1^*$ αμερόληπτοι εκτιμητές)

- Ισχύει και για τον εκτιμητή του σταθερού όρου

$$var(\hat{\beta}_0) < var(\hat{\beta}_0^*)$$

Διακύμανση του $\hat{\beta}_1$

Απόδειξη:

Για να προχωρήσουμε στην ιδιότητα της αποτελεσματικότητας πρέπει να βρούμε την διακύμανση του εκτιμητή ΕΤ, $var(\hat{\beta}_1)$ και $var(\hat{\beta}_0)$

- Από την σχέση (1) έχουμε

$$\hat{\beta}_1 = \sum_{i=1}^{n} c_i Y_i$$

$$var(\hat{\beta}_1) = \sum_{i=1}^{n} c_i^2 var(Y_i) \qquad Y_i \text{ είναι μη συσχετιζόμενες μεταβλητές}$$

$$= \sum_{i=1}^{n} c_i^2 \sigma^2 \qquad \text{καθώς } var(Y_i) = \sigma^2, \forall i$$

$$= \sigma^2 \sum_{i=1}^{n} c_i^2 = \sigma^2 \sum_{i=1}^{n} \left[\frac{(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \qquad (3)$$

Διακύμανση του $\hat{\beta}_0$

✓Η διακύμανση του εκτιμητή ΕΤ του σταθερού όρου δίνεται ως εξής:

$$var(\hat{\beta}_0) = \sigma^2 \frac{\sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n} (X_i - \bar{X})^2}$$

Αποτελεσματικότητα (efficiency)

Απόδειξη:

Τώρα, υποθέστε τώρα ότι υπάρχει κάποιος άλλος γραμμικός και αμερόληπτος εκτιμητής του β_1, που συμβολίζεται ως $\hat{\beta}_1^*$ και ορίζεται ως

$$\hat{\beta}_1^* = \sum_{i=1}^n d_i Y_i, \qquad \text{όπου } d_i = \alpha_i + c_i$$

Λόγω αμεροληψίας του εκτιμητή $\hat{\beta}_1^*$ θα πρέπει να έχουμε:

$$E\left(\hat{\beta}_1^*\right) = E\left[\sum_{i=1}^n (\alpha_i + c_i)(\beta_0 + \beta_1 X_i + u_i)\right]$$

$$= \sum_{i=1}^n (\alpha_i + c_i)\beta_0 + \sum_{i=1}^n (\alpha_i + c_i)\beta_1 X_i + \sum_{i=1}^n (\alpha_i + c_i) E(u_i)$$

$$= \beta_0 \sum_{i=1}^n \alpha_i + \beta_1 \sum_{i=1}^n \alpha_i X_i + \beta_1$$

καθώς $\sum_{i=1}^n c_i = 0, \sum_{i=1}^n c_i X_i = 1$ και $E(u_i) = 0$.

Αποτελεσματικότητα (efficiency)

$$E(\hat{\beta}_1^*) = \beta_0 \sum_{i=1}^{n} \alpha_i + \beta_1 \sum_{i=1}^{n} \alpha_i X_i + \beta_1$$

για να είναι αμερόληπτός ο εκτιμητής $\hat{\beta}_1^*$, θα πρέπει να ισχύουν οι ακόλουθες δύο συνθήκες:

$$\sum_{i=1}^{n} \alpha_i = 0 \text{ και } \sum_{i=1}^{n} \alpha_i X_i = 0 \qquad (4)$$

καθώς μόνο τότε θα ισχύει $E(\hat{\beta}_1^*) = \beta_1$.

Αποτελεσματικότητα (efficiency)

Με βάση τις συνθήκες της σχέσης (4) η διακύμανση του εκτιμητή $\hat{\beta}_1^*$ βρίσκεται ως εξής:

Πρώτα, παρατηρήστε ότι ο εκτιμητής $\hat{\beta}_1^*$ γράφεται:

$$\begin{aligned}\hat{\beta}_1^* &= \sum_{i=1}^{n} d_i Y_i = \sum_{i=1}^{n}(\alpha_i + c_i)(\beta_0 + \beta_1 X_i + u_i) \\ &= \sum_{i=1}^{n}(\alpha_i + c_i)\beta_0 + \sum_{i=1}^{n}(\alpha_i + c_i)\beta_1 X_i + \sum_{i=1}^{n}(\alpha_i + c_i)(u_i) \\ &= \beta_1 + \sum_{i=1}^{n}(\alpha_i + c_i)u_i\end{aligned}$$

όπου $\sum_{i=1}^{n} c_i = 0, \sum_{i=1}^{n} c_i X_i = 1, \sum_{i=1}^{n} \alpha_i = 0$ και $\sum_{i=1}^{n} \alpha_i X_i = 0$

Αποτελεσματικότητα (efficiency)

$$\hat{\beta}_1^* = \beta_1 + \sum_{i=1}^{n}(\alpha_i + c_i)u_i$$

Παίρνοντας τη διακύμανση της τελευταίας σχέσης έχουμε:

$$\begin{aligned} var(\hat{\beta}_1^*) &= var[\beta_1 + \textstyle\sum_{i=1}^{n}(\alpha_i + c_i)u_i] = \sum_{i=1}^{n}(\alpha_i + c_i)^2 var(u_i) \\ &= \sigma^2 \textstyle\sum_{i=1}^{n}(\alpha_i + c_i)^2 = \sigma^2 \sum_{i=1}^{n}(c_i^2 + \alpha_i^2 + 2\alpha_i c_i) \\ &= \sigma^2 \textstyle\sum_{i=1}^{n} c_i^2 + \sigma^2 \sum_{i=1}^{n} \alpha_i^2 + 2\sigma^2 \sum_{i=1}^{n} \alpha_i c_i = var(\hat{\beta}_1) + \sigma^2 \sum_{i=1}^{n} \alpha_i^2 \geq var(\hat{\beta}_1) \end{aligned}$$

καθώς $\sum_{i=1}^{n} \alpha_i c_i = 0$ και $\sum_{i=1}^{n} \alpha_i^2 \geq 0$

Η τελευταία σχέση δείχνει ότι, αν $\alpha_i \neq 0$, τότε

$$var(\hat{\beta}_1^*) > var(\hat{\beta}_1)$$

Αποτελεσματικότητα (efficiency)

Σημείωση:

$\sum_{i=1}^{n} \alpha_i c_i = 0$

$c_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$

$$\sum_{i=1}^{n} \alpha_i c_i = \frac{\sum_{i=1}^{n} \alpha_i (X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n} \alpha_i X_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} - \frac{\bar{X}\sum_{i=1}^{n} \alpha_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = 0$$

Λόγω των συνθηκών της σχέσης (4)

Θεώρημα Gauss-Markov

✓Εφόσον ισχύουν οι υποθέσεις του απλού γραμμικού υποδείγματος της παλινδρόμησης, οι εκτιμητές ΕΤ $\hat{\beta}_0$ και $\hat{\beta}_1$ αποτελούν τους **καλύτερους γραμμικούς αμερόληπτους εκτιμητές** και αναφέρονται ως **BLUE** (best linear unbiased estimator) δηλαδή:

i. Είναι γραμμικές συναρτήσεις των παρατηρήσεων της εξαρτημένης μεταβλητής Y_i

ii. Είναι αμερόληπτοι εκτιμητές, $E(\hat{\beta}_0) = \beta_0$ και $E(\hat{\beta}_1) = \beta_1$

iii. Είναι αποτελεσματικοί εκτιμητές, δηλαδή μεταξύ όλων των γραμμικών αμερόληπτων εκτιμητών έχουν την μικρότερη διακύμανση

Η κατανομή του εκτιμητή $\hat{\beta}_1$

Όταν $u_i \sim N(0, \sigma^2)\ \forall i,$ η κατανομή του $\hat{\beta}_1$ είναι κανονική και δίνεται ως εξής

$$\hat{\beta}_1 \sim N\left[E(\hat{\beta}_1), var(\hat{\beta}_1)\right] \sim N\left[\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right] \qquad (5)$$

Η κατανομή του εκτιμητή $\hat{\beta}_1$

Απόδειξη: Προκύπτει κατευθείαν από τη γραμμική σχέση του ΕΤ εκτιμητή $\hat{\beta}_1$ ως προς τις παρατηρήσεις της εξαρτημένης μεταβλητής ή του διαταρακτικού όρου

$$\hat{\beta}_1 = \beta_1 + \sum_{i=1}^{n} c_i u_i \quad (\text{ή } \hat{\beta}_1 = \sum_{i=1}^{n} c_i Y_i)$$

Όταν $u_i \sim N(0, \sigma^2)$ (οπότε $Y_i \sim N(E(Y_i) = \beta_0 + \beta_1 X_i, var(Y_i) = \sigma^2)$) ισχύει το αποτέλεσμα της σχέσης (5)

- ο εκτιμητής $\hat{\beta}_1$ αποτελεί γραμμικό συνδυασμό κανονικά κατανεμημένων μεταβλητών και έτσι, και αυτός (που αποτελεί τυχαία μεταβλητή) κατανέμεται κανονικά

Η κατανομή του εκτιμητή $\hat{\beta}_0$

Όταν $u_i \sim N(0, \sigma^2)\ \forall i$, η κατανομή του $\hat{\beta}_0$ είναι κανονική και δίδεται ως εξής

$$\hat{\beta}_0 \sim N\left[E(\hat{\beta}_0), var(\hat{\beta}_0)\right] \sim N\left[\beta_0, \sigma^2 \frac{\sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n} (X_i - \bar{X})^2}\right]$$

«Ακρίβεια» εκτιμητή ΕΤ

✓Οι εκτιμητές ΕΤ αποτελούν συνάρτηση των δεδομένων του δείγματος & επομένως είναι πιθανόν να διαφέρουν μεταξύ διαφορετικών δειγμάτων

✓ Τυπικό σφάλμα (standard error, se): Μέτρο «αξιοπιστίας» (reliability) ή «ακρίβειας» (precision) ενός εκτιμητή

$$se\left(\hat{\beta}_1\right) = \frac{\sigma}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}} \text{ και } se\left(\hat{\beta}_0\right) = \sigma\sqrt{\frac{\sum_{i=1}^{n} X_i^2}{n\sum_{i=1}^{n}(X_i - \bar{X})^2}}$$

Εκτίμηση της διακύμανσης του διαταρακτικού όρου σ^2

Από την κατανομή του $\hat{\beta}_1 \sim N[E(\hat{\beta}_1), var(\hat{\beta}_1)] \sim N\left[\beta_1, \frac{\sigma^2}{\sum_{i=}^{n}(X_i-\bar{X})^2}\right]$ έχουμε ότι για να υπολογιστεί η διακύμανση του εκτιμητή πρέπει να εκτιμηθεί σ^2(διακύμανση διαταρακτικού όρου)

Ένας αμερόληπτος εκτιμητής αυτής είναι ο

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-k} \qquad (6)$$

- $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$
- n είναι ο αριθμός των παρατηρήσεων του δείγματος και
- k είναι ο αριθμός των συντελεστών του υποδείγματος που εκτιμώνται (δηλ. η σταθερά και ο συντελεστής κλίσης, δηλ. $k = 2$)
- $n - k$ **βαθμοί ελευθερίας** που διαθέτουμε στην εκτίμηση της σ^2
- $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ τυπικό σφάλμα της παλινδρόμησης

Συντελεστής προσδιορισμού R^2

✓ Συντελεστής προσδιορισμού η συντελεστής προσαρμογής του υποδείγματος R^2

- Μέτρο για να κρίνουμε την ικανότητα του γραμμικού υποδείγματος να προβλέπει τις μεταβολές της εξαρτημένης μεταβλητής για οποιαδήποτε τιμή των ανεξάρτητων μεταβλητών
- Ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής που εξηγείται από την παλινδρόμηση
- Αποτελεί μέτρο προσαρμογής του υποδείγματος στα δεδομένα

Συντελεστής προσδιορισμού R^2

Απόδειξη:

$$\hat{u}_i = Y_i - \hat{Y}_i \Longrightarrow Y_i = \hat{Y}_i + \hat{u}_i$$

Αφαιρέστε και από τα δύο μέλη της σχέσης το δειγματικό μέσο της μεταβλητής Y_i, $\overline{\mathrm{Y}}$ και υψώνουμε στο τετράγωνο

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

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

Παίρνουμε το άθροισμα

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

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

Καθώς από τις ιδιότητες καταλοίπων έχουμε $\sum_{i=1}^{n}\left(\hat{Y}_i - \overline{\mathrm{Y}}\right)\hat{u}_i = 0$

Συντελεστής προσδιορισμού R^2

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

$\sum_{i=1}^{n}(Y_i - \overline{Y})^2$: TSS (total sum of squares) συνολικό άθροισμα των τετραγώνων των αποκλίσεων του Y_i από το δειγματικό μέσο $\overline{Y}$

$\sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2$: ESS (explained sum of squares)το ερμηνευμένο άθροισμα των τετραγώνων των αποκλίσεων $Y_i - \overline{Y}$ από το υπόδειγμα με βάση τις μεταβολές της ανεξάρτητης μεταβλητής X_i

$\sum_{i=1}^{n}\hat{u}^2$: RSS (residual sum of squares) το ανερμήνευτο άθροισμα των τετραγώνων των αποκλίσεων $Y_i - \overline{Y}$ που οφείλεται στην ύπαρξη του διαταρακτικού όρου

Συντελεστής προσδιορισμού R^2

$$TSS = ESS + RSS$$

Διαιρώντας τα δύο μέλη της σχέσης με το TSS έχουμε

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

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n} \hat{u}^2}{\sum_{i=1}^{n}(Y_i - \overline{\mathrm{Y}})^2}$$

$$0 < R^2 < 1$$

$$R^2 = r_{\hat{Y}Y}^2$$

όπου $r_{\hat{Y}Y}$ ο συντελεστής γραμμικής συσχέτισης ανάμεσα στις παρατηρούμενες τιμές της εξαρτημένης μεταβλητής Y_i και τις προβλέψεις της $\hat{Y}_i$

Άσκηση

Σας δίνεται ένα δείγμα 13 παρατηρήσεων για τις μεταβλητές Y_i (μέσο ωρομίσθιο) X_i (έτη εκπαίδευσης). Η γραμμική σχέση που τις συνδέει με βάση την οικονομική θεωρία είναι η εξής $Y_i = \beta_0 + \beta_1 X_i + u_i$

a. Υπολογίστε την εκτίμηση της διακύμανση του διαταρακτικού όρου

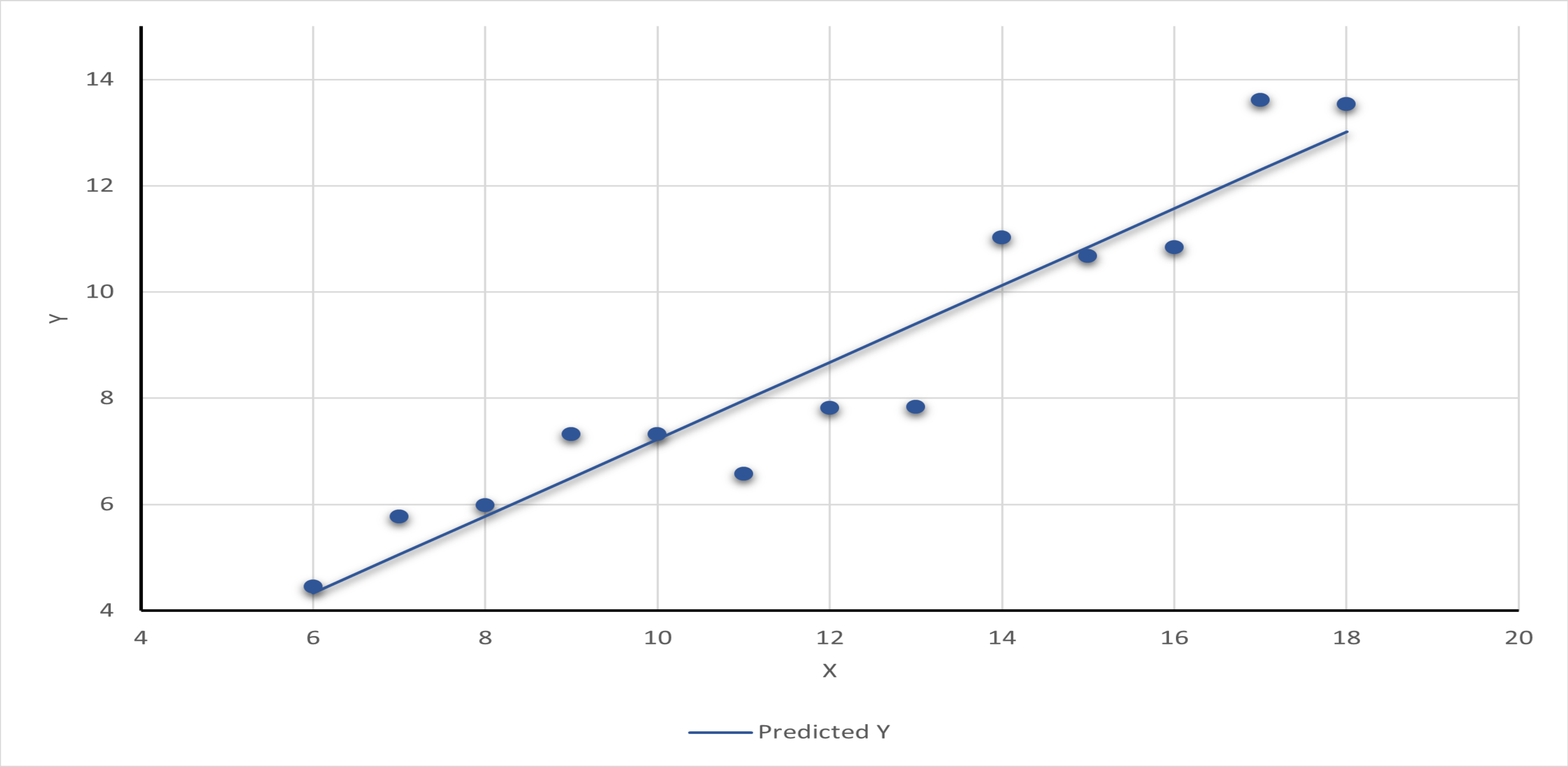
b. Υπολογίστε τη διακύμανση των εκτιμητών ΕΤ $\hat{\beta}_0$ και $\hat{\beta}_1$

c. Υπολογίστε το συντελεστή προσδιορισμού

n	Y_i	X_i
1	4.4567	6
2	**5.77**	7
3	5.9787	8
4	7.3317	9
5	7.3182	10
6	6.5844	11
7	7.8182	12
8	7.8351	13
9	11.0223	14
10	10.6738	15
11	10.8361	16
12	13.615	17
13	13.531	18

Άσκηση

$$\hat{Y}_i = -0.01445 + 0.724097X_i$$



$$\hat{Y}_i = -0.01445 + 0.724097X_i$$

Άσκηση

αν $y_i = Y_i - \bar{Y}$, $x_i = X_i - \bar{X}$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-k} = \frac{9.69}{13-2} = 0.8813$$

$$var(\hat{\beta}_0) = \sigma^2 \frac{\sum_{i=1}^{n} X_i^2}{n\sum_{i=1}^{n}(X_i-\bar{X})^2} = 0.8813\frac{2054}{13*182} = 0.7649$$

$$var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=}^{n}(X_i-\bar{X})^2} = \frac{0.8813}{182} = 0.004842$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \hat{u}_i^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2} = 1 - \frac{9.69}{105.1183} = 0.9078$$

n	Y_i	X_i	y_i	x_i	x_i^2	y_i^2	X_i^2	$\hat{Y}_i$	$\hat{u}_i$	$\hat{u}_i^2$
1	4.4567	6	-4.2180	-6	36	17.7916	36	4.3301	0.1266	0.0160
2	5.77	7	-2.9047	-5	25	8.4373	49	5.0542	0.7158	0.5123
3	5.9787	8	-2.6960	-4	16	7.2685	64	5.7783	0.2004	0.0402
4	7.3317	9	-1.3430	-3	9	1.8037	81	6.5024	0.8293	0.6877
5	7.3182	10	-1.3565	-2	4	1.8401	100	7.2265	0.0917	0.0084
6	6.5844	11	-2.0903	-1	1	4.3694	121	7.9506	-1.3662	1.8665
7	7.8182	12	-0.8565	0	0	0.7336	144	8.6747	-0.8565	0.7336
8	7.8351	13	-0.8396	1	1	0.7049	169	9.3988	-1.5637	2.4452
9	11.0223	14	2.3476	2	4	5.5112	196	10.123	0.8994	0.8089
10	10.6738	15	1.9991	3	9	3.9964	225	10.847	-0.1732	0.0300
11	10.8361	16	2.1614	4	16	4.6716	256	11.571	-0.7350	0.5402
12	13.615	17	4.9403	5	25	24.4065	289	12.295	1.3198	1.7419
13	13.531	18	4.8563	6	36	23.5836	324	13.019	0.5117	0.2618
sum	112.7712	156	0.00	0.00	182.00	105.1183	2054.00	112.77	0.00	9.69

Περίληψη

- ✓Ιδιότητες εκτιμητών
  - ✓Γραμμικότητα
  - ✓Αμεροληψία
  - ✓Αποτελεσματικότητα

- ✓Θεώρημα Gauss-Markov

- ✓Εκτίμηση της διακύμανσης του διαταρακτικού όρου σ^2

- ✓Συντελεστής προσδιορισμού R^2

Βιβλιογραφία

- ✓ Stock & Watson, κεφ. 4 και Παραρτήματα, 5.5 και Παράρτημα 5.2
- ✓ Gujarati & Porter, κεφ. 3
- ✓ Τζαβαλής, κεφ. 2 & 5

Ευχαριστώ!

Ερωτήσεις?