Εισαγωγή στην οικονομετρία (1404)

Διάλεξη 13

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

Περίγραμμα Διάλεξης

- Πρόβλεψη
  - Σφάλμα πρόβλεψης
  - Διάστημα εμπιστοσύνης για την πρόβλεψη
- Άσκηση

Σκοπός Οικονομετρίας

- **Εκτίμηση οικονομικών σχέσεων (ποσοτικοποίηση)**

- Έλεγχος υποθέσεων (οικονομικών σχέσεων) – Στατιστική Επαγωγή

- Πρόβλεψη οικονομικών μεταβλητών

Εκτίμηση οικονομικών σχέσεων

Κλασικές υποθέσεις

Αν X_i είναι σταθερή σε επαναλαμβανόμενα δείγματα (προκαθορισμένη - fixed)

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

Εκτίμηση οικονομικών σχέσεων

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

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

Εκτίμηση οικονομικών σχέσεων

✓Θεώρημα Gauss-Markov: οι εκτιμητές ΕΤ $\hat{\beta}_0$ και $\hat{\beta}_1$ αποτελούν τους **καλύτερους γραμμικούς αμερόληπτους εκτιμητές** και αναφέρονται ως **BLUE** (best linear unbiased estimator)

✓ Όταν $u_i \sim N(0,\sigma^2)$, $\forall i$ και X_i προκαθορισμένες τότε

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

και

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

✓Αμερόληπτος εκτιμητής της διακύμανση διαταρακτικού όρου $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-k}$

✓Συντελεστής προσδιορισμού (coefficient of determination)

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

$$0 < R^2 < 1$$

Σκοπός Οικονομετρίας

- Εκτίμηση οικονομικών σχέσεων (ποσοτικοποίηση)
- **Έλεγχος υποθέσεων (οικονομικών σχέσεων) – Στατιστική Επαγωγή**
- Πρόβλεψη οικονομικών μεταβλητών

Στατιστική επαγωγή

✓ Διάστημα εμπιστοσύνης:

- η διακύμανση του διαταρακτικού όρου σ^2 είναι γνωστή

$$\Pr\left(\hat{\beta}_1 - z_c \cdot se\left(\hat{\beta}_1\right) \leq \beta_1 \leq \hat{\beta}_1 + z_c \cdot se\left(\hat{\beta}_1\right)\right) = 1 - \alpha$$

όπου z_c (ή $z_{a/2}$) και $-z_c$ (ή $-z_{a/2}$) οι κριτικές τιμές της τυποποιημένης κανονικής κατανομής

- η διακύμανση του διαταρακτικού όρου σ^2 είναι άγνωστη

$$\Pr\left(\hat{\beta}_1 - t_c \cdot se\left(\hat{\beta}_1\right) \leq \beta_1 \leq \hat{\beta}_1 + t_c \cdot se\left(\hat{\beta}_1\right)\right) = 1 - \alpha$$

όπου $-t_c$ (ή $-t_{a/2,n-2}$) και t_c (ή $t_{a/2,n-2}$) οι κριτικές τιμές της t-student με $n-2$ βαθμούς ελευθερίας

Στατιστική επαγωγή

✓Έλεγχοι στατιστικών υποθέσεων:

- $H_0: \beta_1 = \beta_1^0$ vs. $H_1: \beta_1 \neq \beta_1^0$ ή $H_1: \beta_1 > \beta_1^0$ ή $H_1: \beta_1 < \beta_1^0$
  - Έλεγχος σημαντικότητας, $\beta_1 = 0$
  - Ή οποιαδήποτε άλλη τιμή

$$z = \frac{\hat{\beta}_1 - \beta_1^0}{se(\hat{\beta}_1)} \sim N(0,1) \text{ ή } t = \frac{\hat{\beta}_1 - \beta_1^0}{se(\hat{\beta}_1)} \sim t_{(n-2)}$$

- z (μεγάλα δείγματα & γνωστή σ^2) και t (μικρά δείγματα & άγνωστη σ^2)

Σκοπός Οικονομετρίας

- Εκτίμηση οικονομικών σχέσεων (ποσοτικοποίηση)
- Έλεγχος υποθέσεων (οικονομικών σχέσεων)
- **Πρόβλεψη οικονομικών μεταβλητών**

Πρόβλεψη με το απλό γραμμικό υπόδειγμα

- **Πρόβλεψη**: εκτίμηση της εξαρτημένης μεταβλητής ενός γραμμικού υποδείγματος που αντιστοιχεί σε μια τιμή της ανεξάρτητης μεταβλητής και προκύπτει με βάση τις εκτιμήσεις ΕΤ των συντελεστών του υποδείγματος για κάποιο δείγμα παρατηρήσεων
  - Έχουν το ελάχιστο δυνατό σφάλμα πρόβλεψης καθώς στηρίζονται στους εκτιμητές ΕΤ που είναι αμερόληπτοι & αποτελεσματικοί εκτιμητές των συντελεστών του υποδείγματος

- Διάστημα εμπιστοσύνης για την πρόβλεψη ή **διάστημα πρόβλεψης** (prediction interval)

Πρόβλεψη με το απλό γραμμικό υπόδειγμα

- Έστω το απλό γραμμικό υπόδειγμα

$$Y_i = \beta_0 + \beta_1 X_i + u_i, i = 1, \ldots, n$$

- Αν υποθέσουμε ότι X_0 είναι μια τιμή της ανεξάρτητης μεταβλητής του γραμμικού υποδείγματος για την οποία θέλουμε να βρούμε την πρόβλεψη της εξαρτημένης μεταβλητής Y_i, Y_0
  - Πρόβλεψή στο σημείο 0 της ανεξάρτητης μεταβλητής X_i
- Η X_0 μπορεί να πάρει κάποια τιμή από το δείγμα των παρατηρήσεων και τότε αναφέρεται ως <u>πρόβλεψη εντός δείγματος</u> (in-sample) ή

κάποια άλλη τιμή και αναφέρεται ως <u>πρόβλεψη εκτός δείγματος</u> (out-of-sample)

Πρόβλεψη με το απλό γραμμικό υπόδειγμα

- Με βάση το απλό γραμμικό υπόδειγμα η αληθινή τιμή της μεταβλητής Y_i στο σημείο 0 θα δίνεται ως εξής

$$Y_0 = \beta_0 + \beta_1 X_0 + u_0$$

ενώ η πρόβλεψη του υποδείγματος για αυτή στον πληθυσμό δίνεται ως η αναμενόμενη τιμής στο σημείο 0

$$E(Y_0) = \beta_0 + \beta_1 X_0$$

καθώς $E(u_0) = 0$

- Μια <u>αμερόληπτη εκτίμηση της πρόβλεψής</u> αυτής δίνεται αν αντικαταστήσουμε στους συντελεστές του υποδείγματος τις εκτιμήσεις ΕΤ με βάση κάποιο υπόδειγμα

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_0$$

η σχέση αυτή αναφέρεται ως **εκτιμητής της πρόβλεψης** καθώς βασίζεται στους εκτιμητές ΕΤ $\hat{\beta}_0$ και $\hat{\beta}_1$ και η $\hat{Y}_0$ ως **πρόβλεψη**

Σφάλμα πρόβλεψης

- **Σφάλμα πρόβλεψης**: η διαφορά ανάμεσα στην πρόβλεψη $\hat{Y}_0$ και την αληθινή τιμή της Y_0
    - Είναι ανάλογο των καταλοίπων του εκτιμημένου υποδείγματος

$$\hat{Y}_0 - Y_0 = (\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)X_0 - u_0$$

- Παίρνοντας την αναμενόμενη τιμή του σφάλματος πρόβλεψης έχουμε

$$E(\hat{Y}_0 - Y_0) = E(\hat{\beta}_0 - \beta_0) + X_0 E(\hat{\beta}_1 - \beta_1) - E(u_0) = 0$$

καθώς $E(\hat{\beta}_0 - \beta_0) = 0, E(\hat{\beta}_1 - \beta_1) = 0$ και $E(u_0) = 0$

- Η πρόβλεψη $\hat{Y}_0$ αποτελεί αμερόληπτη εκτίμηση της πρόβλεψης του υποδείγματος στον πληθυσμό, $E(Y_0) = \beta_0 + \beta_1 X_0$

Διακύμανση σφάλματος πρόβλεψης

- Για να έχουμε την καλύτερη δυνατή πρόβλεψη θα πρέπει να ελαχιστοποιήσουμε το σφάλμα της πρόβλεψης

- Η πιθανότητα να λάβουμε μια πρόβλεψη με μικρό σφάλμα εξαρτάται από τη διακύμανση του του σφάλματος πρόβλεψης

$$
\begin{aligned}
var\left(\hat{Y}_0 - Y_0\right) &= var\left[(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)X_0 - u_0\right] \\
&= var(\hat{\beta}_0 - \beta_0) + X_0^2 var(\hat{\beta}_1 - \beta_1) + var(u_0) + 2X_0 cov(\hat{\beta}_0 - \beta_0; \hat{\beta}_1 - \beta_1) \\
&= var(\hat{\beta}_0) + X_0^2 var(\hat{\beta}_1) + \sigma^2 + 2X_0 cov(\hat{\beta}_0, \hat{\beta}_1)
\end{aligned}
$$

Διακύμανση σφάλματος πρόβλεψης

(συνέχεια)

$$var(\hat{Y}_0 - Y_0) = var(\hat{\beta}_0) + X_0^2 var(\hat{\beta}_1) + \sigma^2 + 2X_0 cov(\hat{\beta}_0, \hat{\beta}_1)$$

όπου $var(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n}(X_i - \bar{X})^2}$, $var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$ και $cov(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\sigma^2 \bar{X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$

Άρα

$$var(\hat{Y}_0 - Y_0) = \frac{\sigma^2 \sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n}(X_i - \bar{X})^2} + \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} X_0^2 + \sigma^2 - 2\frac{\sigma^2 \bar{X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2} X_0$$

Διακύμανση σφάλματος πρόβλεψης

(συνέχεια) $var\left(\hat{Y}_0 - Y_0\right) = \frac{\sigma^2 \sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n}(X_i - \bar{X})^2} + \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} X_0^2 + \sigma^2 - 2\frac{\sigma^2 \bar{X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2} X_0$

$$= \sigma^2 \left[\frac{\sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n}(X_i - \bar{X})^2} + \frac{n}{n \sum_{i=1}^{n}(X_i - \bar{X})^2} X_0^2 + 1 - 2\frac{n\bar{X}}{n \sum_{i=1}^{n}(X_i - \bar{X})^2} X_0\right]$$

$$= \sigma^2 \left[\frac{\sum_{i=1}^{n} X_i^2 + nX_0^2 - 2n\bar{X}X_0}{n \sum_{i=1}^{n}(X_i - \bar{X})^2} + 1\right] = \sigma^2 \left[\frac{\sum_{i=1}^{n} X_i^2 + nX_0^2 - 2n\bar{X}X_0 + n\bar{X}^2 - n\bar{X}^2}{n \sum_{i=1}^{n}(X_i - \bar{X})^2} + 1\right]$$

$$= \sigma^2 \left[\frac{\sum_{i=1}^{n}(X_i^2 - \bar{X}^2)}{n \sum_{i=1}^{n}(X_i - \bar{X})^2} + \frac{n(X_0 - \bar{X})^2}{n \sum_{i=1}^{n}(X_i - \bar{X})^2} + 1\right] = \sigma^2 \left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n \sum_{i=1}^{n}(X_i - \bar{X})^2} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} + 1\right]$$

$$= \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} + 1\right]$$

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

Διακύμανση σφάλματος πρόβλεψης

(συνέχεια)

$$var(\hat{Y}_0 - Y_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} + 1\right]$$

✓ $var(\hat{Y}_0 - Y_0)$ αυξάνεται όταν

- μειώνεται το n
- μεγαλώνει η απόκλιση της X_0 και της $\bar{X}$ $(X_0 - \bar{X})$
- αυξάνεται η διακύμανση του διαταρακτικού όρου σ^2

- Όταν το δείγμα είναι αρκετά μεγάλο τότε η πρόβλεψη υπολογίζεται στο σημείο του μέσου της X_i, $X_0 = \bar{X}$ και $var(\hat{Y}_0 - Y_0)$ λαμβάνει την ελάχιστη τιμή της, $var(\hat{Y}_0 - Y_0) = \sigma^2$

Διάστημα εμπιστοσύνης της πρόβλεψης

- Διάστημα εμπιστοσύνης της πρόβλεψης ή διάστημα πρόβλεψης: διάστημα εντός του οποίου θα βρίσκεται η αληθινή τιμή της εξαρτημένης μεταβλητής στο σημείο 0, Y_0 με κάποια πιθανότητα $1 - \alpha$, όπου α προκαθορισμένο επίπεδο σημαντικότητας

Διάστημα εμπιστοσύνης της πρόβλεψης

- Αν ισχύουν οι κλασικές υποθέσεις και ο διαταρακτικός όρος κατανέμεται κανονικά, $u_i \sim N(0,\sigma^2)$, $\forall i$ τότε το σφάλμα πρόβλεψης

$$\hat{Y}_0 - Y_0 = (\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)X_0 - u_0$$

θα ακολουθεί την κανονική κατανομή καθώς οι εκτιμητές ΕΤ $\hat{\beta}_0$ και $\hat{\beta}_1$ είναι κανονικά κατανεμημένοι συνεπώς το σφάλμα της πρόβλεψης είναι το άθροισμα κανονικά κατανεμημένων τιμών.

- Η κατανομή δίνεται ως εξής

$$\hat{Y}_0 - Y_0 \sim \mathrm{N}\left[0, var\left(\hat{Y}_0 - Y_0\right)\right]$$

Διάστημα εμπιστοσύνης της πρόβλεψης

- Αν διαιρέσουμε το σφάλμα πρόβλεψης με την τυπική του απόκλιση $\sqrt{var(\hat{Y}_0 - Y_0)}$ θα πάρουμε την τυποποιημένη μεταβλητή του σφάλματος πρόβλεψης $\frac{\hat{Y}_0 - Y_0}{\sqrt{var(\hat{Y}_0 - Y_0)}}$
- Όταν η διακύμανση του διαταρακτικού όρου σ^2 θεωρείται γνωστή, η μεταβλητή ακολουθεί την τυποποιημένη κανονική κατανομή

$$\frac{\hat{Y}_0 - Y_0}{\sqrt{var(\hat{Y}_0 - Y_0)}} \sim \mathrm{N}(0,1)$$

Διάστημα εμπιστοσύνης της πρόβλεψης

- Όταν η διακύμανση του διαταρακτικού όρου σ^2 δεν είναι γνωστή, η τυποποιημένη μεταβλητή ακολουθεί την t-student κατανομή

$$\frac{\hat{Y}_0 - Y_0}{\sqrt{var(\hat{Y}_0 - Y_0)}} \sim t_{(n-2)}$$

- Το διάστημα εμπιστοσύνης της αληθινής τιμής Y_0 ικανοποιεί την παρακάτω σχέση

$$\Pr\left(\hat{Y}_0 - t_c \cdot \sqrt{var(\hat{Y}_0 - Y_0)} \leq Y_0 \leq \hat{Y}_0 + t_c \cdot \sqrt{var(\hat{Y}_0 - Y_0)}\right) = 1 - \alpha$$

Και ορίζεται ως

$$\hat{Y}_0 - t_c \cdot \sqrt{var(\hat{Y}_0 - Y_0)} \leq Y_0 \leq \hat{Y}_0 + t_c \cdot \sqrt{var(\hat{Y}_0 - Y_0)}$$

όπου t_c η κριτική τιμή της κατανομής t-student με $n-2$ βαθμούς ελευθερίας

Άσκηση

Σας δίνεται το παρακάτω εκτιμημένο υπόδειγμα

$$\widehat{RATE_i} = 3.204 + 0.076\ EXP_i$$
$$(se) \quad (0.709) \quad (0.044)$$

Όπου $n = 24$, το RATE αναφέρεται στην αξιολόγηση εργαζομένων και παίρνει τιμές 1-7 (1=poor, 7=excellent) και το EXP είναι τα χρόνια απασχόλησης στη συγκεκριμένη εργασία (Hill, Griffiths & Lim, Principle of Econometrics, 4th Ed.)

i. Ερμηνεύστε τον εκτιμημένο συντελεστή κλίσης $\hat{\beta}_1$.
ii. Υπολογίστε το 95% διάστημα εμπιστοσύνης του συντελεστή κλίσης β_1.
iii. Προβείτε σε έλεγχο στατιστικής σημαντικότητας για το $\hat{\beta}_1$ σε επίπεδο σημαντικότητας, $a = 5\%$.
iv. Για το ερώτημα iii σας δίνεται ότι η p-τιμή είναι 0.0982. Αν επιλέξουμε την πιθανότητα του σφάλματος τύπου I να είναι $a = 0.05$, μπορούμε να απορρίψουμε την μηδενική υπόθεση ή όχι?
v. Ελέγξτε την μηδενική υπόθεση ότι $\beta_1 = 0$ έναντι της εναλλακτικής ότι λαμβάνει θετική τιμή σε επίπεδο σημαντικότητας $a = 5\%$.

Άσκηση

$$\widehat{RATE}_i = 3.204 + 0.076 \; EXP_i$$
$$(se) \quad (0.709) \quad (0.044)$$

$n = 24$

RATE αξιολόγηση εργαζομένων και παίρνει τιμές 1-7 (1=poor, 7=excellent)
EXP τα χρόνια απασχόλησης στη συγκεκριμένη

i. Ερμηνεύστε τον εκτιμημένο συντελεστή κλίσης $\hat{\beta}_1$

- Για κάθε επιπλέον χρόνο εργασιακής εμπειρίας, η αξιολόγηση των εργαζομένων αυξάνεται κατά μέσο όρο με 0.076 μονάδες

Άσκηση

$$\widehat{RATE}_i = 3.204 + 0.076\ EXP_i$$
$$(se) \quad (0.709) \quad (0.044)$$

$n = 24$

RATE αξιολόγηση εργαζομένων και παίρνει τιμές 1-7 (1=poor, 7=excellent)

EXP τα χρόνια απασχόλησης στη συγκεκριμένη

ii. Υπολογίστε το 95% διάστημα εμπιστοσύνης του συντελεστή κλίσης β_1

Το διάστημα εμπιστοσύνης δίνεται ως εξής

$$\hat{\beta}_1 \pm t_c \cdot se(\hat{\beta}_1)$$

$\hat{\beta}_1^L = \hat{\beta}_1 - t_c \cdot se(\hat{\beta}_1)$ και $\hat{\beta}_1^U = \hat{\beta}_1 + t_c \cdot se(\hat{\beta}_1)$ αποτελούν αντίστοιχα το κάτω και το άνω όριο του διαστήματος εμπιστοσύνης β_1

Η κριτική τιμή δίνεται ως εξής $t_c = t_{a/2,n-2} = t_{0.025,22} = 2.074$

Άρα $\hat{\beta}_1^L = \hat{\beta}_1 - t_c \cdot se(\hat{\beta}_1) = 0.076 - 2.074 * 0.044 = -0.015$ και $\hat{\beta}_1^U = \hat{\beta}_1 + t_c \cdot se(\hat{\beta}_1) = 0.076 + 2.074 * 0.044 = 0.167$

Άσκηση

$$\widehat{RATE}_i = 3.204 + 0.076\ EXP_i$$
$$(se) \quad (0.709) \quad (0.044)$$

$n = 24$

RATE αξιολόγηση εργαζομένων και παίρνει τιμές 1-7 (1=poor, 7=excellent)

EXP τα χρόνια απασχόλησης στη συγκεκριμένη

iii. Προβείτε σε έλεγχο σημαντικότητας για το $\hat{\beta}_1$ σε επίπεδο σημαντικότητας, $a = 5\%$.

Ο έλεγχος σημαντικότητας αντιστοιχεί στον έλεγχο της παρακάτω υπόθεσης

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

Το στατιστικό κριτήριο για τον έλεγχο της υπόθεσης είναι $t = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{0.076-0}{0.044} = 1.727$

Για τον δικατάληκτο έλεγχο η H_0 απορρίπτεται αν η απόλυτη τιμή του στατιστικού κριτηρίου είναι μεγαλύτερη ή ίση της κριτικής τιμής της κατανομής του $|t| > t_c$, καθώς $|1.727| < 2.074$ η H_0 δεν απορρίπτεται

Άσκηση

$$\widehat{RATE}_i = 3.204 + 0.076\ EXP_i$$
$$(se) \quad (0.709) \quad (0.044)$$

$n = 24$

RATE αξιολόγηση εργαζομένων και παίρνει τιμές 1-7 (1=poor, 7=excellent)

EXP τα χρόνια απασχόλησης στη συγκεκριμένη

iv. Για το ερώτημα iii σας δίνεται ότι η p-τιμή είναι 0.0982. Αν επιλέξουμε την πιθανότητα του σφάλματος τύπου I να είναι $a = 0.05$, μπορούμε να απορρίψουμε την μηδενική υπόθεση ή όχι?

Για να απορρίψουμε τη μηδενική υπόθεση χρησιμοποιώντας την p-τιμή θέλουμε $\mathrm{p-value} < \alpha$

Για τον έλεγχο σημαντικότητας στο ερώτημα iii έχουμε ότι p-τιμή=0.0982 και $a = 0.05$ οπότε 0.0982>0.05 και επομένως η H_0 δεν απορρίπτεται

Άσκηση

$$\widehat{RATE}_i = 3.204 + 0.076\ EXP_i$$
$$(se)\quad (0.709)\quad (0.044)$$

$n = 24$

RATE αξιολόγηση εργαζομένων και παίρνει τιμές 1-7 (1=poor, 7=excellent)

EXP τα χρόνια απασχόλησης στη συγκεκριμένη

v. Ελέγξτε την μηδενική υπόθεση ότι $\beta_1 = 0$ έναντι της εναλλακτικής ότι λαμβάνει θετική τιμή σε επίπεδο σημαντικότητας, $a = 5\%$.

Η υπόθεση δίνεται ως $H_0\colon \beta_1 = 0$ και $H_1\colon \beta_1 > 0$

Το στατιστικό κριτήριο για τον έλεγχο της υπόθεσης είναι $t \equiv \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{0.076-0}{0.044} = 1.727$

Για τον μονοκατάληκτο έλεγχο προς τα δεξιά η H_0 απορρίπτεται όταν t> t_c

Η κριτική τιμή σε αυτή την περίπτωση είναι $t_c = t_{a,n-2} = t_{0.05,22} = 1.717$

Επομένως $1.727 > 1.717$ και η H_0 απορρίπτεται (τα χρόνια απασχόλησης έχουν θετική επίδραση στην αξιολόγηση των εργαζομένων.)

Περίληψη

✓ Πρόβλεψη

- Σφάλμα πρόβλεψης
- Διάστημα εμπιστοσύνης για την πρόβλεψη

Βιβλιογραφία

- ✓ Τζαβαλής, κεφ. 5
- ✓ Gujarati & Porter, κεφ. 5

Ευχαριστώ!

Ερωτήσεις?