Εισαγωγή στην οικονομετρία (1404)

Διάλεξη 15

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

Περίγραμμα Διάλεξης

- Πολλαπλό υπόδειγμα παλινδρόμησης
  - Υποθέσεις πολλαπλού υποδείγματος
- Εκτίμηση πολλαπλού υποδείγματος
  - Μέθοδος ΕΤ
- Στατιστικές ιδιότητες εκτιμητή ΕΤ
- Εκτίμηση της διακύμανσης του διαταρακτικού όρου
- Στατιστική επαγωγή
  - Έλεγχοι ατομικών υποθέσεων
  - Διαστήματα εμπιστοσύνης

Πολλαπλό υπόδειγμα με 2 ανεξάρτητες μεταβλητές

- Έστω ένα υπόδειγμα με 2 ανεξάρτητες μεταβλητές

$$Y_i = \beta_0 X_{0i} + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, i = 1, \ldots, n$$

ή

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

όπου

Y_i είναι η εξαρτημένη μεταβλητή,

$X_{0i} = 1$ (αντιπροσωπεύει τη μεταβλητή του σταθερού όρου, σταθερή ερμηνευτική μεταβλητή),

X_{1i}, X_{2i} οι ανεξάρτητες μεταβλητές,

β_0 ο σταθερός όρος και β_1, β_2 οι συντελεστές της γραμμικής παλινδρόμησης

u_i ο διαταρακτικός όρος

Πολλαπλό υπόδειγμα με 2 ανεξάρτητες μεταβλητές

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

- Οι β_1, β_2 λέγονται **συντελεστές μερικής παλινδρόμησης** (partial regression coefficients) ή **συντελεστές μερικής κλίσης** (partial slope coefficients) επειδή μετρούν μερικές μεταβολές του $E(Y_i)$

$$E(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$$

π.χ. β_1 μετράει τη μεταβολή στην μέση τιμή της Y_i, $E(Y_i)$, όταν μεταβάλλεται η X_{1i} κατά μία μονάδα, κρατώντας σταθερή την τιμή της X_{2i} (ceteris paribus)

$$\beta_1 = \left.\frac{\Delta E(Y_i)}{\Delta X_{1i}}\right|_{X_{2i}\ \text{held constant}} = \frac{\partial E(Y_i)}{\partial X_{1i}}$$

Πολλαπλό υπόδειγμα με k ανεξάρτητες μεταβλητές

- Η εξαρτημένη μεταβλητή Y_i επηρεάζεται από ένα σύνολο ερμηνευτικών (ανεξάρτητων) μεταβλητών $X_{1i}, X_{2i}, \dots, X_{ki}$

$$Y_i = \beta_0 X_{0i} + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, i = 1, \dots, n$$

ή

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$$

όπου $X_{0i} = 1$ (αντιπροσωπεύει τη μεταβλητή του σταθερού όρου, σταθερή ερμηνευτική μεταβλητή), $X_{1i}, X_{2i}, \dots, X_{ki}$ οι ανεξάρτητες μεταβλητές του υποδείγματος και β_0 ο σταθερός όρος και $\beta_1, \beta_2, \dots, \beta_k$ οι k-συντελεστές που αντιστοιχούν στις ανεξάρτητες μεταβλητές $\rightarrow k + 1$

- Η μέθοδος εκτίμησης, ΕΤ του πολλαπλού γραμμικού υποδείγματος είναι μια γενίκευση της απλής περίπτωσης για $k = 1$

Υποθέσεις πολλαπλού υποδείγματος

1. Το υπόδειγμα είναι γραμμικό ως προς τις παραμέτρους

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, i = 1, \ldots, n$$

2. $E(u_i) = 0, \forall i$

3. $var(u_i) = E(u_i^2) = \sigma^2, \forall i$

4. $cov(u_i, u_j) = 0, \forall i \neq j$ (απουσία αυτοσυσχέτισης/ανεξαρτησία)

5. Απουσία πολυσυγγραμμικότητας μεταξύ των ερμηνευτικών μεταβλητών Δεν υπάρχουν «τέλειες» γραμμικές σχέσεις μεταξύ των ερμηνευτικών μεταβλητών (no perfect collinearity)

6. Ο διαταρακτικός όρος ακολουθεί την κανονική κατανομή $u_i \sim N(0, \sigma^2), \forall i$

Σημείωση: $X_{1i}, X_{2i}, \ldots, X_{ki}$ είναι σταθερές σε επαναλαμβανόμενα δείγματα

Υποθέσεις πολλαπλού υποδείγματος

1. Το υπόδειγμα είναι γραμμικό ως προς τις παραμέτρους

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \, i = 1, \ldots, n$$

2. $E(u_i | X_{1i}, X_{2i}, \ldots, X_{ki}) = 0, \forall i$

3. $var(u_i | X_{1i}, X_{2i}, \ldots, X_{ki}) = E\left(u_i^2 \middle| X_{1i}, X_{2i}, \ldots, X_{ki}\right) = \sigma^2, \forall i$

4. $cov\left(u_i, u_j \middle| X_{1i}, X_{2i}, \ldots, X_{ki}\right) = 0, \forall i \neq j$

5. Απουσία «τέλειας» πολυσυγγραμμικότητας μεταξύ των ερμηνευτικών μεταβλητών Δεν υπάρχουν «τέλειες» γραμμικές σχέσεις μεταξύ των ερμηνευτικών μεταβλητών (no perfect multicollinearity)

6. Ο διαταρακτικός όρος ακολουθεί την κανονική κατανομή $u_i | X_{1i}, X_{2i}, \ldots, X_{ki} \sim N(0, \sigma^2), \forall i$

Σημείωση: αν οι $X_{1i}, X_{2i}, \ldots, X_{ki}$ στοχαστικές (τυχαίες μεταβλητές)

Εξωγένεια

$$E(u_i) = 0, \forall i \text{ ή } E(u_i | X_{1i}, X_{2i}, \ldots, X_{ki}) = 0, \forall i$$

- Η υπόθεση του μηδενικού δεσμευμένου μέσου συνεπάγεται $cov(X_{ki}, u_i) = E(X_{ki} u_i) = 0, k = 1, \ldots, k$
- Αν η υπόθεση ικανοποιείται τότε έχουμε εξωγενείς ερμηνευτικές μεταβλητές
  - Αν δεν ικανοποιείται έχουμε ενδογενείς ερμηνευτικές μεταβλητές → η μέθοδος ΕΤ δεν είναι κατάλληλη
- Παράλειψη μιας «σημαντικής» μεταβλητής η οποία σχετίζεται με κάποια από τις $X_{1i}, X_{2i}, \ldots, X_{ki}$ → παραβίαση της υπόθεσης
  - $cons = \beta_0 + \beta_1 inc + u$ vs. $cons = \beta_0 + \beta_1 inc + \beta_2 inc^2 + u$
- Λάθος εξειδίκευση του υποδείγματος

Πολυσυγγραμμικότητα

- Τέλεια ή πλήρης πολυσυγγραμικότητα: αν μια ανεξάρτητη μεταβλητή είναι «τέλειος» γραμμικός συνδυασμός με μια άλλη ανεξάρτητη μεταβλητή
  - Σε αυτή την περίπτωση το υπόδειγμα δεν μπορεί να εκτιμηθεί με τη μέθοδο ΕΤ
  - π.χ. $voteA = \beta_0 + \beta_1 expendA + \beta_2 expendB + \beta_3 totexpend + u$
  - $voteA$: ποσοστό ψήφων για τον υποψήφιο Α, $expendA$ και $expendB$: δαπάνες για την προεκλογική εκστρατεία του υποψηφίου Α και Β και $totexpend$: συνολικές δαπάνες
  - $totexpend = expendA + expendB$

Πολυσυγγραμμικότητα

(συνέχεια)

➢Οι ανεξάρτητες μεταβλητές μπορούν να συσχετίζονται, αλλά όχι τέλεια

- π.χ. $avgscore = \beta_0 + \beta_1 expend + \beta_2 avginc + u$
- $avgscore$: μέσο test score, $expend$: δαπάνες για εκπαίδευση και $avginc$: μέσο οικογενειακό εισόδημα
- $expend$ και $avginc$ έχουν κάποια συσχέτιση αλλά ότι τέλεια
- Οι οικογένειες με υψηλότερο μέσο εισόδημα θα ξοδεύουν περισσότερα χρήματα στην εκπαίδευση

Η Μέθοδος Ελαχίστων Τετραγώνων

✓ Έστω το υπόδειγμα

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

✓ <u>Μέθοδος ελαχίστων τετραγώνων (ΕΤ, ordinary least squares, OLS ή LS):</u> εκτιμητές (εκτιμήσεις) για τους συντελεστές β_0 , β_1 και β_2 που ελαχιστοποιούν την συνάρτηση του αθροίσματος των τετραγώνων των καταλοίπων (residual sum of squares, RSS):

$$\mathrm{RSS}\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\right) \equiv \sum_{i=1}^{n} \hat{u}_i^2$$

Η Μέθοδος Ελαχίστων Τετραγώνων

✓Το **προβλημα της μεθόδου ΕΤ** λύνεται ως εξής:

$$\underset{\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_2}{\text{ελαχ}} \ \mathrm{RSS}\left(\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_2\right) \equiv \underset{\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_2}{\text{ελαχ}} \ \sum_{i=1}^{n}\hat{u}_i^2 = \underset{\hat{\beta}_0,\hat{\beta}_1}{\text{ελαχ}} \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

$$= \underset{\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_2}{\text{ελαχ}} \ \sum_{i=1}^{n}\left(Y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}\right)\right)^2$$

$$= \underset{\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_2}{\text{ελαχ}} \ \sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}\right)^2 \qquad (1)$$

- Οι λύσεις που προκύπτουν από το πρόβλημα (1) είναι οι εκτιμητές ελαχίστων τετραγώνων

Η Μέθοδος Ελαχίστων Τετραγώνων

Για τη λύση της (1) απαιτείται ικανοποίηση των συνθηκών πρώτης και δεύτερης τάξης:

Συνθήκες πρώτης τάξης (first order conditions - FOC):

$$\frac{\partial}{\partial \hat{\beta}_0}\sum_{i=1}^{n}\hat{u}_i^2 = 0,\ \frac{\partial}{\partial \hat{\beta}_1}\sum_{i=1}^{n}\hat{u}_i^2 = 0 \text{ και } \frac{\partial}{\partial \hat{\beta}_2}\sum_{i=1}^{n}\hat{u}_i^2 = 0 \qquad (2α)$$

Συνθήκες δεύτερης τάξης (second order conditions-SOC):

$$\frac{\partial^2}{\partial \hat{\beta}_0^2}\sum_{i=1}^{n}\hat{u}_i^2 > 0,\ \frac{\partial^2}{\partial \hat{\beta}_1^2}\sum_{i=1}^{n}\hat{u}_i^2 > 0 \text{ και } \frac{\partial^2}{\partial \hat{\beta}_2^2}\sum_{i=1}^{n}\hat{u}_i^2 > 0 \qquad (2β)$$

- από τις συνθήκες πρώτης τάξης προκύπτουν οι εκτιμητές ΕΤ
- οι συνθήκες δεύτερης τάξης είναι οι ικανές συνθήκες για την επίτευξη του ελάχιστου στην συνάρτηση (1)

Η Μέθοδος Ελαχίστων Τετραγώνων

Από τις συνθήκες πρώτης τάξης προκύπτουν οι εκτιμητές ελαχίστων τετραγώνων $\hat{\beta}_0, \hat{\beta}_1$ και $\hat{\beta}_2$ ως

$$\frac{\partial}{\partial \hat{\beta}_0} \sum_{i=1}^{n} \hat{u}_i^2 = \cdots = -2 \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \right) = 0$$

$$\Rightarrow \sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_{1i} + \hat{\beta}_1 \sum_{i=1}^{n} X_{2i}$$

$$\frac{\partial}{\partial \hat{\beta}_1} \sum_{i=1}^{n} \hat{u}_i^2 = \ldots = -2 \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \right) X_{1i} = 0$$

$$\Rightarrow \sum_{i=1}^{n} Y_i X_{1i} = \hat{\beta}_0 \sum_{i=1}^{n} X_{1i} + \hat{\beta}_1 \sum_{i=1}^{n} X_{1i}^2 + \hat{\beta}_2 \sum_{i=1}^{n} X_{1i} X_{2i}$$

$$\frac{\partial}{\partial \hat{\beta}_2} \sum_{i=1}^{n} \hat{u}_i^2 = \ldots = -2 \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \right) X_{2i} = 0$$

$$\Rightarrow \sum_{i=1}^{n} Y_i X_{2i} = \hat{\beta}_0 \sum_{i=1}^{n} X_{2i} + \hat{\beta}_1 \sum_{i=1}^{n} X_{1i} X_{2i} + \hat{\beta}_2 \sum_{i=1}^{n} X_{2i}^2$$

Η Μέθοδος Ελαχίστων Τετραγώνων

- Λύνοντάς το σύστημα εξισώσεων προκύπτουν οι εκτιμητές των συντελεστών της γραμμής παλινδρόμησης του δείγματος

$$\hat{\beta}_1 = \frac{(\sum_{i=1}^{n} y_i x_{1i})(\sum_{i=1}^{n} x_{2i}^2) - (\sum_{i=1}^{n} y_i x_{2i})(\sum_{i=1}^{n} x_{1i} x_{2i})}{(\sum_{i=1}^{n} x_{1i}^2)(\sum_{i=1}^{n} x_{2i}^2) - (\sum_{i=1}^{n} x_{1i} x_{2i})^2}$$

$$= \frac{cov(y_i, x_{1i}) var(x_{2i}) - cov(y_i, x_{2i}) cov(x_{1i}, x_{2i})}{var(x_{1i}) var(x_{2i}) - cov(x_{1i}, x_{2i})^2}$$

$$\hat{\beta}_2 = \frac{(\sum_{i=1}^{n} y_i x_{2i})(\sum_{i=1}^{n} x_{1i}^2) - (\sum_{i=1}^{n} y_i x_{1i})(\sum_{i=1}^{n} x_{1i} x_{2i})}{(\sum_{i=1}^{n} x_{1i}^2)(\sum_{i=1}^{n} x_{2i}^2) - (\sum_{i=1}^{n} x_{1i} x_{2i})^2}$$

$$= \frac{cov(y_i, x_{2i}) var(x_{1i}) - cov(y_i, x_{1i}) cov(x_{1i}, x_{2i})}{var(x_{1i}) var(x_{2i}) - cov(x_{1i}, x_{2i})^2}$$

$$\hat{\beta}_0 = \bar{\mathrm{Y}} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$$

όπου $y_i = \mathrm{Y}_i - \bar{\mathrm{Y}}$, $x_{1i} = X_{1i} - \bar{X}_1$, $x_{2i} = X_{2i} - \bar{X}_2$ και $var(x_{1i}) var(x_{2i}) \neq cov(x_{1i}, x_{2i})^2$

Μέθοδος ΕΤ για k ανεξάρτητες μεταβλητές

- Από τη λύση των εξισώσεων προέκυψαν οι εκτιμητές ΕΤ για το πολλαπλό υπόδειγμα με 2 ερμηνευτικές μεταβλητές, που αποτελούν γενίκευση των εκτιμητών ΕΤ του απλού υποδείγματος

- Αντίστοιχα μπορούμε να υπολογίσουμε τους εκτιμητές ΕΤ για το πολλαπλό υπόδειγμα με k ερμηνευτικές μεταβλητές
  - Αρκετά περίπλοκες συναρτήσεις
  - Μπορούμε να χρησιμοποιήσουμε κάποιο software για τον υπολογισμό των εκτιμητών
  - Άλγεβρα μητρών (Οικονομετρία Ι)

Στατιστικές ιδιότητες εκτιμητή ΕΤ

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, i = 1, \dots, n$$

- Ο εκτιμητής ΕΤ $\hat{\beta}_j$, $j = 0, \dots, k$ αποτελεί γραμμική συνάρτηση των τιμών της εξαρτημένης μεταβλητής του υποδείγματος Y_i
- Ο εκτιμητής ΕΤ $\hat{\beta}_j$, $j = 0, \dots, k$ είναι αμερόληπτος, δηλ. η μέση τιμή του εκτιμητή $\hat{\beta}_j$ ισούται με την αληθινή (θεωρητική) τιμή αυτού στον πληθυσμό β_j, $E(\hat{\beta}_j) = \beta_j, j = 0, \dots, k$
- Ο εκτιμητής ΕΤ $\hat{\beta}_j$, $j = 0, \dots, k$ είναι αποτελεσματικός, δηλ. έχει μικρότερη διακύμανση από κάποιον άλλον εκτιμητή έστω $\hat{\beta}_j^*$, $var(\hat{\beta}_j) < var(\hat{\beta}_j^*)$

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

Στατιστικές ιδιότητες εκτιμητή ΕΤ

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

$$\hat{\beta}_j \sim N\left[E\left(\hat{\beta}_j\right), var\left(\hat{\beta}_j\right)\right] \sim N\left[\beta_j, var\left(\hat{\beta}_j\right)\right], j = 0, \dots, k$$

Η διακύμανση του εκτιμητή ΕΤ $\hat{\beta}_j$

$$var\left(\hat{\beta}_j\right) = \frac{\sigma^2}{\sum_{i=1}^{n}\left(X_{ji}-\bar{X}_j\right)^2}\left(\frac{1}{1-R_j^2}\right), j = 1, \dots, k$$

όπου R_j^2 είναι το R^2 από την παλινδρόμηση της X_{ji} με τις υπόλοιπες ερμηνευτικές μεταβλητές και την σταθερά

- R_j^2 το ποσοστό της μεταβλητότητας της X_{ji} που μπορεί να εξηγηθεί από τις υπόλοιπες ερμηνευτικές μεταβλητές
- Τυπικό σφάλμα του εκτιμητή ΕΤ $\hat{\beta}_j$: $se\left(\hat{\beta}_j\right) = \sqrt{var\left(\hat{\beta}_j\right)}$

Στατιστικές ιδιότητες εκτιμητή ΕΤ

$$var(\hat{\beta}_j) = \frac{\sigma^2}{\sum_{i=1}^{n}(X_{ij}-\bar{X}_j)^2}\left(\frac{1}{1-R_j^2}\right)$$

Η $var(\hat{\beta}_j)$:

- μειώνεται όταν μειώνεται σ^2
- μειώνεται όταν αυξάνεται η συνολική μεταβλητότητά του X_{ij}, $\sum_{i=1}^{n}(X_{ij} - \bar{X}_j)^2$
- μειώνεται όταν μειώνεται R_j^2
  - $var(\hat{\beta}_j)$ μικρότερη δυνατή τιμή όταν $R_j^2 = 0$
  - $var(\hat{\beta}_j) \rightarrow \infty$ όταν $R_j^2 = 1$
    - Πιθανή πολυσυγγραμμικότητα (θα το εξετάσουμε αργότερα)

Εκτίμηση της διακύμανσης του διαταρακτικού όρου σ²

Ένας αμερόληπτος εκτιμητής της διακύμανσης του διαταρακτικού όρου είναι ο

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

- $\hat{u}_i = Y_i - \hat{\beta}_0 - \beta_1 X_{1i} - \beta_2 X_{2i} + \cdots - \beta_k X_{ki}$
- n είναι ο αριθμός των παρατηρήσεων του δείγματος και
- k είναι ο αριθμός των συντελεστών του υποδείγματος που εκτιμώνται (δεν συμπεριλαμβάνεται ο σταθερός όρος)
- $n - k - 1$ **βαθμοί ελευθερίας** που διαθέτουμε στην εκτίμηση της σ^2
- $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ τυπικό σφάλμα της παλινδρόμησης

Στατιστική επαγωγή στο πολλαπλό υπόδειγμα

- Διάστημα εμπιστοσύνης

- Στο πολλαπλό υπόδειγμα μπορούμε να κατασκευάσουμε διαστήματα εμπιστοσύνης για τους συντελεστές (παραμέτρους) του πληθυσμού με βάση τις ιδιότητες των εκτιμητών ΕΤ (όπως και στο απλό γραμμικό υπόδειγμα)

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

όπου $-t_c$ (ή $-t_{a/2,n-k-1}$) και t_c (ή $t_{a/2,n-k-1}$) οι κριτικές τιμές της t-student με $n-k-1$ βαθμούς ελευθερίας

Στατιστική επαγωγή στο πολλαπλό υπόδειγμα

- Ατομικοί έλεγχοι στατιστικών υποθέσεων – κριτήριο t
- $H_0: \beta_j = \beta_j^0$ vs. $H_1: \beta_j \neq \beta_j^0$ ή $H_1: \beta_j > \beta_j^0$ ή $H_1: \beta_j < \beta_j^0$
  - Έλεγχος σημαντικότητας, $\beta_j^0 = 0$ ή οποιαδήποτε άλλη τιμή

$$t = \frac{\hat{\beta}_j - \beta_j^0}{se(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta_j^0}{\sqrt{var(\hat{\beta}_j)}} \sim t_{(n-k-1)}$$

- Απόρριψη H_0
  - Δίπλευρος $|t| > t_c$ $(t_c = t_{a/2,n-k-1})$

    Μονόπλευρος προς τα δεξιά t$> t_c$ $(t_c = t_{a,n-k-1})$
  - Μονόπλευρος προς τα αριστερά $t < -t_c$ $(t_c = t_{a,n-k-1})$
  - $\text{p} - \text{value} < \alpha$

Ψευδομεταβλητές

Όταν ένα μέγεθος είναι αδύνατο να ποσοτικοποιηθεί αλλά πρέπει οπωσδήποτε να χρησιμοποιηθεί σε ένα υπόδειγμα προσεγγίζεται συνήθως με μια μεταβλητή η οποία ονομάζεται ποιοτική μεταβλητή ή ψευδομεταβλητή.

Π.χ. το φύλο ενός ατόμου, το επάγγελμά του, η εθνικότητά του κ.λ.π.

Θα εξεταστούν μόνο οι περιπτώσεις των ψευδομεταβλητών που χρησιμοποιούνται σαν ανεξάρτητες μεταβλητές.

Παράδειγμα:

Θέλουμε να διερευνήσουμε την σχέση εισοδήματος και κατανάλωσης τροφίμων. Για τον σκοπό αυτό ρωτήθηκαν 20 άτομα τα οποία δήλωσαν τα παρακάτω στοιχεία.

Τρόφιμα (*FOOD*)	Εισόδημα (*INC*)	Φύλο (*GEN*)	Φύλο (*GEN*)
28	224	Α	1
32	231	Α	1
45	261	Α	1
47	270	Α	1
62	304	Α	1
58	295	Α	1
59	338	Α	1
70	354	Α	1
74	334	Α	1
70	336	Α	1
82	380	Α	1
93	422	Γ	0
68	353	Γ	0
74	351	Γ	0
92	371	Γ	0
100	422	Γ	0
110	444	Γ	0
111	473	Γ	0
118	473	Γ	0
115	464	Γ	0

Ερώτημα: Υπάρχει διαφορά στην κατανάλωση τροφίμων μεταξύ ανδρών και γυναικών; Δηλαδή, υπάρχει διαφορά στην κατανάλωση τροφίμων μεταξύ ανδρών και γυναικών που έχουν το ίδιο εισόδημα;

Αν θέλουμε να διερευνήσουμε τον ρόλο του φύλου στην σχέση αυτή θα πρέπει να μετατρέψουμε την ποιοτική μεταβλητή *GEN* σε ποσοτική.

$$GEN = \begin{cases} 1 & \text{Αν είναι άντρας} \\ 0 & \text{Αν είναι γυναίκα} \end{cases}$$

Περίληψη

- Πολλαπλό υπόδειγμα παλινδρόμησης
  - Υποθέσεις πολλαπλού υποδείγματος
- Εκτίμηση πολλαπλού υποδείγματος
  - Μέθοδος ΕΤ
- Στατιστικές ιδιότητες εκτιμητή ΕΤ
- Εκτίμηση της διακύμανσης του διαταρακτικού όρου
- Στατιστική επαγωγή
  - Έλεγχοι ατομικών υποθέσεων
  - Διαστήματα εμπιστοσύνης

Έτσι $F\hat{O}OD_i = \hat{\beta}_0 + \hat{\beta}_1 INC_i$

Η συνάρτηση που υποθέτει ότι δεν υπάρχει διαφορά μεταξύ ανδρών και γυναικών.

Ενώ $F\hat{O}OD_i = \hat{\beta}_0 + \hat{\beta}_1 INC_i + \hat{\beta}_2 GEN$

Η συνάρτηση που επιτρέπει τον έλεγχο αυτής της διαφοράς.

Όταν $GEN = 0$ $\quad F\hat{O}OD_i = \hat{\beta}_0 + \hat{\beta}_1 INC_i$

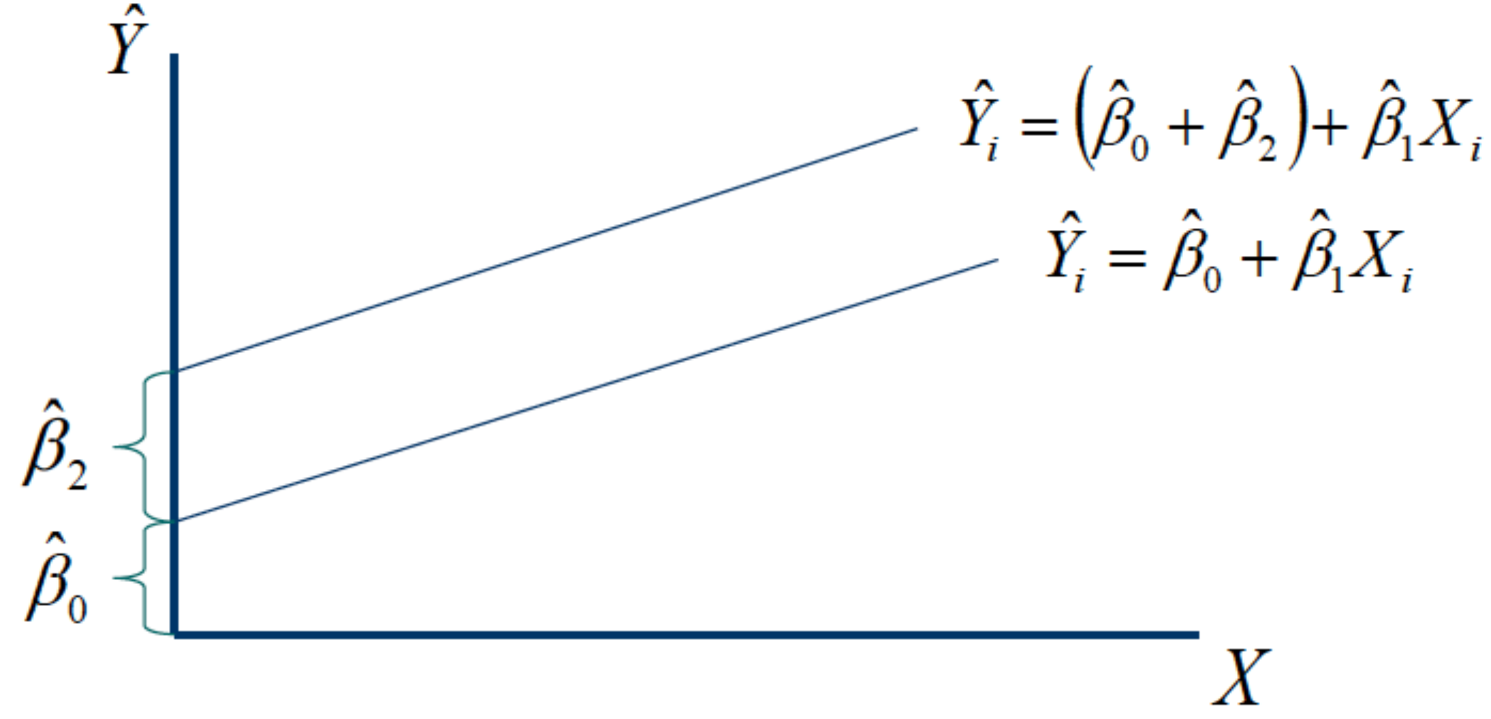
Όταν $GEN = 1$ $\quad F\hat{O}OD_i = \left(\hat{\beta}_0 + \hat{\beta}_2\right) + \hat{\beta}_1 INC_i$

Στην γενική περίπτωση του απλού υποδείγματος παλινδρόμηση

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 D$$

Όταν $D = 1$ $\quad \hat{Y}_i = \left(\hat{\beta}_0 + \hat{\beta}_2\right) + \hat{\beta}_1 X_i$

Όταν $D = 0$ $\quad \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$



Ένας συνηθισμένος έλεγχος είναι $\quad H_0 : \beta_2 = 0 \quad H_1 : \beta_2 \neq 0$

Βιβλιογραφία

- ✓ Stock & Watson, κεφ. 6
- ✓ Gujarati & Porter, κεφ. 7

Ευχαριστώ!

Ερωτήσεις?