ΕΦΑΡΜΟΣΜΕΝΗ ΟΙΚΟΝΟΜΕΤΡΙΑ

1. ΓΕΝΙΚΕΣ ΕΝΝΟΙΕΣ

Τυχαίες μεταβλητές

Τυχαία μεταβλητή είναι μια μεταβλητή της οποίας δεν γνωρίζουμε την αξία μέχρι να την παρατηρήσουμε. Αυτή είναι μια συνάρτηση του δειγματικού χώρου.

♦ Η τυχαία μεταβλητή καλείται διακριτή εάν μπορεί να πάρει ως τιμές διακριτούς αριθμούς.

♦ Η τυχαία μεταβλητή καλείται συνεχής εάν μπορεί να πάρει τιμές μέσα από ένα σύνολο πραγματικών αριθμών.

Κάθε τυχαία μεταβλητή μπορεί να περιγραφεί χρησιμοποιώντας την συνάρτηση πυκνότητας πιθανότητας (pdf) ή την συνάρτηση αθροιστικής πιθανότητας (cdf).

Για μια διακριτή τυχαία μεταβλητή X η τιμή της συνάρτηση πυκνότητας πιθανότητας f(x) είναι η πιθανότητα η τυχαία μεταβλητή X να πάρει την τιμή x , f(x) = P(X = x). H f(x) παίρνει πάντα τιμές μεταξύ 0 και 1.

Για μια συνεχή τυχαία μεταβλητή Y η συνάρτηση πυκνότητας πιθανότητας f(y) μπορεί να υπολογισθεί από μια εξίσωση.

Αναμενόμενη τιμή τυχαίας μεταβλητής, διακύμανση, τυπική απόκλιση τυχαίας μεταβλητής

Η αναμενόμενη τιμή της τυχαίας μεταβλητής X είναι ο μέσος όρος των τιμών που παίρνουμε από το εξεταζόμενο δείγμα. Και ορίζεται από την σχέση:

E[X] = $\sum_{i=1}^{n} xf(x)$ αν η τυχαία μεταβλητή είναι διακριτή

E[X] = $\int_{-\infty}^{+\infty} xf(x)$ αν η τυχαία μεταβλητή είναι συνεχής

Η διακύμανση της τυχαίας μεταβλητής X δείχνει την μεταβλητότητα αυτής και ορίζεται από την σχέση:

$$V[X] = E[X - E[X]]^2 = E[X^2] - [E[X]]^2 = \sigma^2$$

$$= \sum_{i=1}^{n} (x_i - \mu)^2 f(x_i) \text{ εαν X είναι διακριτή}$$

$$= {}_{XY} \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx \text{ εαν X είναι συνεχής}$$

Συνδιακύμανση δύο τυχαίων μεταβλητών X, Y

Έστω f(x,y) είναι η από κοινού συνάρτηση πυκνότητας πιθανότητας pdf των **τυχαίων μεταβλητών X, Y**. Τότε η συνδιακύμανση ορίζεται από την σχέση:

$$Cov(X, Y) = E[(X - \mu_X)(X - \mu_Y)] = \sigma_{XY}$$

$= \sum_x \sum_y [(x - E(X)][y - E(Y)]\, f(x,y)$ όταν οι X και Y είναι διακριτές τυχαίες μεταβλητές

$= \iint_{xy} (x - E(X))(y - E(Y))\, f(x,y)dxdy$ όταν οι X και Y είναι συνεχείς τυχαίες μεταβλητές

Αν $\sigma_{XY} > 0$, X και Y κινούνται στην ίδια κατεύθυνση
Αν $\sigma_{XY} < 0$, X και Y κινούνται στην αντίθετη κατεύθυνση
Αν $\sigma_{XY} = 0$, X και Y είναι ανεξάρτητες

Συντελεστής συσχέτισης των τυχαίων μεταβλητών X, Y

Ο συντελεστής συσχέτισης των τυχαίων μεταβλητών X, Y μετρά την γραμμική συσχέτιση μεταξύ των X,Y και ορίζεται από την σχέση:

$$\text{Corr}(X, Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \rho$$

Αν ρ = 1, X και Y έχουν μια δυνατή θετική συσχέτιση.
Αν ρ = -1, X και Y έχουν μια δυνατή αρνητική συσχέτιση.

Χρήσιμες Κατανομές

• Κανονική Κατανομή

Αν η τυχαία μεταβλητή έχει μια κανονική κατανομή, $X \sim N(\mu, \sigma^2)$, τότε η συνάρτηση κατανομής πιθανότητας είναι της μορφής:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(\chi - \mu)^2\right)$$

• Τυπική κανονική κατανομή

Όταν η τυχαία μεταβλητή Z έχει μια τυπική κανονική κατανομή, $Z \sim N(0,1)$, με μέσο όρο το 0 και διακύμανση το 1 η συνάρτηση κατανομής πιθανότητας είναι της μορφής:

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(z)^2\right)$$

• X^2 – κατανομή

Η μεταβλητή Y, που παράγεται από το άθροισμα n - τυχαίων μεταβλητών Z που ακολουθούν τυπική κανονική κατανομή, ακολουθεί την X^2 - **κατανομή** με βαθμούς ελευθερίας (d.o.f.) m.

$$Y = \sum_{i=1}^{n} z_i^2 = X^2\ (m)$$

Οι βαθμοί ελευθερίας μπορεί να θεωρηθούν ως ο αριθμός των ανεξαρτήτων πληροφοριών.

• T – κατανομή

Θεωρούμε μια τυχαία μεταβλητή Z τυπικής κανονικής κατανομής και μια τυχαία μεταβλητή Y που ακολουθεί την X^2 κατανομή. Ο λόγος της Z προς την τετραγωνική ρίζα του λόγου της Y προς τους βαθμούς ελευθερίας της παράγουν μια μεταβλητή Φ που ακολουθεί την T – κατανομή με m βαθμούς ελευθερίας. Οι Z και Y θεωρούνται ανεξάρτητες.

$$\Phi = \frac{N(0,1)}{\sqrt{\frac{X^2}{m}}} = \frac{Z}{\sqrt{\frac{Y}{m}}} \sim t(m)$$

• F- κατανομή

Έστω Y και W δύο ανεξάρτητες X^2 - κατανομές **με βαθμούς ελευθερίας αντίστοιχα** m και k respectively, τότε η μεταβλητή U προκύπτει από τον λόγο $\frac{Y/m}{W/k}$ ακολουθεί μια **F- κατανομή με βαθμούς ελευθερίας** m και k.

$$U = \frac{Y/m}{W/k} \sim F(m, k)$$

2. Απλή Παλινδρόμηση (Διμεταβλητό Υπόδειγμα)

Οικονομική Θεωρία:

Η δαπάνη για φαγητό (Y) εξαρτάται από το εισόδημα των νοικοκυριών(X).

Οικονομικό Υπόδειγμα: $\mathbf{Y = \beta_1 + \beta_2 X}$ (Προσδιοριστική ή Θεωρητική σχέση)

Y: δαπάνη

X: εισόδημα,

β_1, β_2 άγνωστες παράμετροι προς εκτίμηση

Στατιστικό ή Στοχαστικό Υπόδειγμα:

Είναι γεγονός ότι η εξαρτημένη μεταβλητή Y δεν μπορεί να καθορισθεί ακριβώς για κάθε επίπεδο εισοδήματος, δηλαδή οι δαπάνες που αντιστοιχούν στο ίδιο εισόδημα δεν είναι ίδιες. Και αυτό γιατί υπάρχουν και άλλοι παράγοντες που καθορίζουν τις δαπάνες και όχι μόνο το εισόδημα (π.χ οι καταναλωτικές συνήθειες των ατόμων που χαρακτηρίζουν την συμπεριφορά αυτών). Η ευθεία που μπορεί να προσδιορισθεί από την θεωρητική σχέση $\mathbf{Y = \beta_1 + \beta_2 X}$ δεν περνά από όλα τα σημεία (Y, X) των παρατηρήσεων και οι αποκλίσεις (θετικές ή αρνητικές) από την ευθεία λαμβάνονται υπ' όψιν με την προσθήκη μια τυχαίας μεταβλητής **e.**

Για τον λόγο αυτό μπαίνει η μια τυχαία μεταβλητή **e** και προκύπτει η σχέση:

$Y = \beta_1 + \beta_2 X + e$ (Στατιστικό ή στοχαστικό υπόδειγμα)

Y εξαρτημένη μεταβλητή

X : Ανεξάρτητη μεταβλητή

β_1, β_2 : Άγνωστες παράμετροι προς εκτίμηση

e μια τυχαία μεταβλητή που ονομάζεται και **τυχαίος όρος σφάλματος**

Ο ρόλος του τυχαίου όρου σφάλματος :

(1) Περιλαμβάνει την επίδραση άλλων παραγόντων (εκτός των ανεξάρτητων μεταβλητών)

(2) Ανθρώπινη συμπεριφορά (απρόβλεπτος παράγων)

(3) Πιθανό σφάλμα λόγω λανθασμένης εξειδίκευσης του υποδείγματος (λανθασμένη συναρτησιακή σχέση)

(4) Σφάλμα μέτρησης του Y

Βασικές Υποθέσεις του υποδείγματος $\mathbf{Y = \beta_1 + \beta_2 X + e}$

Στο υπόδειγμα $\mathbf{Y = \beta_1 + \beta_2 X + e}$ εφόσον η **e** θεωρείται μια στοχαστική μεταβλητή και η εξαρτημένη μεταβλητή Y θεωρείται στοχαστική μεταβλητή. Αν θεωρήσουμε ότι η **e** ακολουθεί τυπική κανονική κατανομή και ότι η Y ακολουθεί τυπική κανονική κατανομή τότε η μέση αναμενόμενη τιμή της Y για δεδομένη τιμή της X, $E(Y/X = x_i$, δηλαδή η X να παίρνει δεδομένη τιμή x_i), βρίσκεται από την σχέση:

$E(Y/X= x_i) = \beta_1 + \beta_2\, x_i$ (παλινδρόμηση στον πληθυσμό)

Έτσι για κάθε τιμή της $X = x_i$ αντιστοιχεί ένα πλήθος τιμών της Y που κατανέμονται κανονικά και μια αναμενόμενη μέση τιμή της Y, η $E(Y/X = x_i)$.

Οι αποκλίσεις του πλήθους τιμών της Y γύρω από την αναμενόμενη μέση τιμή της δίνονται από την στοχαστική μεταβλητή:

$e_i = Y - E(Y/X = x_i)$.

Βασικές Υποθέσεις

(1) Υπόδειγμα γραμμικό ως προς τις παραμέτρους με μία μόνο ανεξάρτητη μεταβλητή:

$$Y = \beta_1 + \beta_2 X + e$$

- ✓ Στο υπόδειγμα που αναφέρθηκε η μέση δαπάνη δεν είναι σταθερά αλλά συνάρτηση μιας μεταβλητής του εισοδήματος.

(2) $e_i \approx (0, \sigma^2)$

Ο τυχαίος όρος σφάλματος είναι μία στοχαστική τυχαία μεταβλητή (ανήκει σε μία κατανομή πιθανότητας) και επιπλέον ισχύουν τα εξής:

$$E(e_i/X= x_i) = 0, \;\; Var(e_i/X= x_i) = \sigma^2 \text{ για κάθε τιμή } x_i$$

- ✓ Η μεταβλητότητα $Var(e_i/X= x_i)$ του τυχαίου όρου σφάλματος δεν επηρεάζεται από το μέγεθος της ανεξάρτητης μεταβλητής. Ο τυχαίος όρος σφάλματος είναι **ομοσκεδαστικός** δηλ. έχει **σταθερή διακύμανση.**

(3) $Cov(e_i, e_j) = 0$

Έλλειψη συσχέτισης μεταξύ **των τιμών του τυχαίου στοχαστικού όρου** που αντιστοιχούν σε **διαφορετικές παρατηρήσεις**.

(4) Η μεταβλητή Χ είναι μη στοχαστική και επομένως οι τιμές της υπόκεινται στον έλεγχο του ερευνητή.

Ισχύει:

$E(X_i) = X_i$ (δεδομένες τιμές)

$Cov(X_i, e_i) = 0$

γιατί οι τιμές της ανεξάρτητης μεταβλητής δεν επηρεάζονται από τον τυχαίο όρο σφάλματος.

Γραμμή παλινδρόμησης στο δείγμα

Συλλέγουμε ένα δείγμα 40 νοικοκυριών , $T = 40$

$(X_1, Y_1), (X_2, Y_2), \ldots, (X_T, Y_T)$

Στόχος: Η εκτίμηση των β1 και β2.

Έστω ότι b_1, b_2 είναι οι εκτιμήσεις των παραμέτρων.

Ισχύει:

$$\hat{Y}_i = E(Y/X = x_i) = b_1 + b_2 x_i$$

Όπου $i = 1,2,\ldots,40$ το πλήθος των παρατηρήσεων

$\hat{Y}_i$: η **εκτιμηθείσα i - τιμή** της **εξαρτημένης μεταβλητής**

Εφόσον έχουμε Y_i την αντίστοιχη τιμή της Y στην τιμή της $X = x_i$ η διαφορά $\hat{e}_i = Y_i - \hat{Y}_i$ μεταξύ της i-**εκτιμηθείσας τιμής από την i - παρατηρούμενη τιμή** χαρακτηρίζεται ως **κατάλοιπο ή απόκλιση** και εκτιμά το **σφάλμα** e_i

Το μέγεθος του καταλοίπου δείχνει κατά πόσο η απόκλιση είναι καλή ή κακή.

- ✓ Αν το κατάλοιπο είναι μεγάλο τότε η εκτίμηση είναι κακή

- ✓ Αν το κατάλοιπο είναι μικρό τότε η εκτίμηση είναι καλή.
- ✓

Η εκτίμηση της γραμμής παλινδρόμησης βασίζεται στην αρχή των ελαχίστων τετραγώνων.

Για κάθε παρατήρηση του δείγματος έχουμε και ένα κατάλοιπο. Κάποια είναι θετικά, κάποια είναι αρνητικά. Για να μην υπάρξει πρόβλημα λόγω της ύπαρξης πολλών λύσεων κάνουμε χρήση των τετραγώνων των καταλοίπων. Στόχος είναι η ελαχιστοποίηση του αθροίσματος των τετραγώνων των καταλοίπων.

Η ΘΕΩΡΙΑ ΤΩΝ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ

(ΕΦΑΡΜΟΓΗ ΣΤΟ ΥΠΟΔΕΙΓΜΑ: $\mathbf{Y = \beta_1 + \beta_2 X + e}$)

Η θεωρία των ελαχίστων τετραγώνων στηρίζεται στην βασική αρχή της εκτίμησης των συντελεστών παλινδρόμησης β_1, β_2, των $\mathbf{b_1, b_2}$, έτσι ώστε να **ελαχιστοποιείται το άθροισμα των τετραγώνων των καταλοίπων**.

Αυτό οδηγεί στις ακόλουθες εξισώσεις που ονομάζονται **κανονικές** και μας επιτρέπουν να υπολογίσουμε τα b_1, b_2.

Κανονικές εξισώσεις:

$$\sum_{i=1}^{T} Y_i = \mathbf{N}\, b_1 + b_2 \sum_{i=1}^{T} X_i$$

$$\sum_{i=1}^{T} X_i Y_i = b_1 \sum_{i=1}^{T} X_i + b_2 \sum_{i=1}^{T} X_i^2$$

Από την επίλυση του συστήματος αυτού έχουμε:

$$b_1 = \overline{Y} - b_2 \overline{X}$$

$$b_2 = \frac{\sum_{i=1}^{T}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{T}(X_i - \overline{X})^2}$$

Επίσης το b_2 μπορεί να υπολογισθεί από τον τύπο:

$$b_2 = \frac{T\sum_{i}^{T} X_i Y_i - \sum_{i}^{T} X_i \sum_{i}^{T} Y_i}{T\sum_{i}^{T} X_i^2 - (\sum_{i}^{T} X_i)^2}$$

Η διακύμανση του διαταρακτικού όρου $Var(e_i/X = x_i) = \sigma^2$,για κάθε τιμή x_i, εκτιμάται από την σχέση:

$$\hat{\sigma}^2 = \frac{\sum_{i}^{T} \hat{e}_i^2}{T-2}$$

- **Ο αριθμός 2 εκφράζει το πλήθος των προς εκτίμηση παραμέτρων**

Όπου Τ-2 είναι οι **βαθμοί ελευθερίας** του υποδείγματος $Y = \beta_1 + \beta_2 X + e$

Βαθμοί ελευθερίας: Είναι ο αριθμός των ανεξάρτητων παρατηρήσεων (X,Y). Πιο συγκεκριμένα οι βαθμοί ελευθερίας είναι ίσοι με το συνολικό αριθμό παρατηρήσεων αν αφαιρέσουμε το αριθμό των γραμμικών περιορισμών στους οποίους υπόκεινται οι παρατηρήσεις X,Y.

Στο **διμεταβλητό** υπόδειγμα οι παρατηρήσεις της εξαρτημένης μεταβλητής υπόκεινται σε **δύο περιορισμούς** που δίνονται από τις σχέσεις που ακολουθούν:

$$b_1 = \overline{Y} - b_2 X$$

$$b_2 = \frac{\sum_{i=1}^{T}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{T}(X_i - \overline{X})^2}$$

Αυτό σημαίνει πως οι **βαθμοί ελευθερίας είναι T-2.**

- Οι **δύο εκτιμητές b_1, b_2 είναι τυχαίες μεταβλητές** γιατί μπορούν να πάρουν πολλές διαφορετικές τιμές μια και μπορούν να προκύψουν από πολλά διαφορετικά δείγματα του πληθυσμού.
-

Οι διακυμάνσεις των δύο εκτιμητών b_1, b_2 είναι:

$$Var(b_1) = \hat{\sigma}^2 \left[\frac{\sum_{i}^{T} X_i^2}{T\sum_{i=1}^{T}(X_i - \overline{X})^2} \right] \qquad \text{και} \qquad Var(b_2) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{T}\left(X_i - \overline{X}\right)^2}$$

Οι διακυμάνσεις των δύο εκτιμητών είναι θετική συνάρτηση της διακύμανσης.

Επιπλέον, οι διακυμάνσεις είναι αντίστροφα ανάλογες από το

$$\sum_{i=1}^{T}(X_i - \overline{X})^2$$

Οι διακυμάνσεις είναι αντίστροφα ανάλογες του μεγέθους του δείγματος Τ.

- Για την επίλυση του προβλήματος μπορούμε να κάνουμε επίσης χρήση και της **γραμμικής άλγεβρας**.

Πιο συγκεκριμένα θα κάνουμε χρήση των μητρών - πινάκων.

Από τις σχέσεις:

$Y_1 = \beta_1 + \beta_2 X_1 + e_1$

$Y_2 = \beta_1 + \beta_2 X_2 + e_2$

$Y_3 = \beta_1 + \beta_2 X_3 + e_3$

………………………..

………………………..

$Y_T = \beta_1 + \beta_2 X_T + e_T$

Όπου Τ το πλήθος των παρατηρήσεων

Οδηγούμαστε στην σχέση πινάκων:

$$\mathbf{Y = Xb + e}$$

Όπου $\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_T \end{bmatrix}$, πίνακας στήλη Τ x 1,

$\mathbf{X} = \begin{bmatrix} 1..X_1 \\ 1..X_2 \\ . \\ . \\ 1..X_T \end{bmatrix}$, πίνακας Τ x 2 $\quad$ $\mathbf{b} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$, πίνακας στήλη 2 x 1

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ e_T \end{bmatrix}$$, πίνακας στήλη T x 1

Δηλαδή έχουμε:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_T \end{bmatrix} = \begin{bmatrix} 1..X_1 \\ 1..X_2 \\ . \\ . \\ 1..X_T \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ e_T \end{bmatrix}$$

Επειδή έχουμε ότι:

$$e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ .... \\ e_T \end{bmatrix}$$, ο πίνακας στήλη των εκτιμήσεων των σφαλμάτων θα είναι ο

$$\hat{e} = \begin{bmatrix} \hat{e}_1 \\ \hat{e}_2 \\ \hat{e}_3 \\ .... \\ \hat{e}_T \end{bmatrix}$$ ο ανάστροφός του οποίου δίνεται από το διάνυσμα:

$$\hat{e}' = [\hat{e}_1\, \hat{e}_2 \ldots\ldots \hat{e}_T]$$

Το άθροισμα των τετραγώνων των σφαλμάτων $\sum_{i=1}^{T} e^2$ στην μορφή των πινάκων δίνεται από την σχέση:

$$\sum_{i=1}^{T} e^2 = [\mathrm{e}_1 \ \mathrm{e}_2 \ldots\ldots.\mathrm{e}_T]\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \ldots. \\ e_T \end{bmatrix}$$

Επίσης το άθροισμα των τετραγώνων των εκτιμήσεων των σφαλμάτων $\sum_{i=1}^{T} \hat{e}^2$ στην μορφή των πινάκων δίνεται από την σχέση:

$$\hat{e}'\hat{e} = \begin{bmatrix} \hat{e}_1 & \hat{e}_2 & \ldots & \hat{e}_T \end{bmatrix}\begin{bmatrix} \hat{e}_1 \\ \hat{e}_2 \\ \ldots \\ \hat{e}_T \end{bmatrix} = \hat{e}_1^2 + \hat{e}_2^2 + \ldots\ldots + \hat{e}_T^2 = \sum_{i=1}^{T} \hat{e}_i^2$$

Συνεπώς η θεωρία των ελαχίστων τετραγώνων που χρησιμοποιεί την ελαχιστοποίηση του αθροίσματος των τετραγώνων των καταλοίπων για να εκτιμηθούν οι συντελεστές β_1 και β_2 μας οδηγεί στην σχέση:

$$\begin{aligned} \min \sum_{i=1}^{T} \hat{e}_i^{\,2} &= \min(Y - Xb)'(Y - Xb) = \min(Y' - b'X')(Y - Xb) = \\ &= \min(Y'Y - Y'Xb - b'X'Y + b'X'Xb) \end{aligned} \qquad (1)$$

Οι διαστάσεις των ανάστροφων μητρών είναι:

Y΄: 1 x T

b΄: 1 x 2

X΄: 2 x T

Τα γινόμενα $Y'Xb$ και $b'X'Y$ ορίζονται ως **βαθμωτά μεγέθη** αφού έχουν διαστάσεις 1x1. Ο $Y'Xb$ είναι ο ανάστροφος του $b'X'Y$.

Πιο συγκεκριμένα ισχύει:

$$(b'X'Y)' = Y'(X')'(b')' = Y'Xb$$

Επειδή και οι δύο όροι είναι βαθμωτά μεγέθη και $(b'X'Y)' = Y'Xb$ δίνουν άθροισμα $2Y'Xb = 2b'X'Y$

Άρα η σχέση (1) γίνεται: $\min\left(Y'Y - 2b^{'}X^{'}Y + b^{'}X'Xb\right)$

Για την ελαχιστοποίηση της σχέσης (1) απαιτείται:

$$\frac{\partial(\hat{e}'\hat{e})}{\partial b} = 0 \Leftrightarrow -2X'Y + 2X'Xb = 0$$

$$\Leftrightarrow X'Xb = X'Y \Leftrightarrow (X'X)^{-1}(X'X)b = (X'X)^{-1}X'Y \Leftrightarrow b_{OLS} = (X'X)^{-1}X'Y$$

και οδηγούμαστε στην σχέση:

$$b_{OLS} = (X'X)^{-1}X'Y$$

Η συνθήκη αυτή πρώτης τάξης δίνει δύο εξισώσεις εφόσον το b είναι ένα διάνυσμα. Οι εξισώσεις αυτές ονομάζονται **κανονικές εξισώσεις.**

❖ Εφαρμογή της Παλινδρόμησης

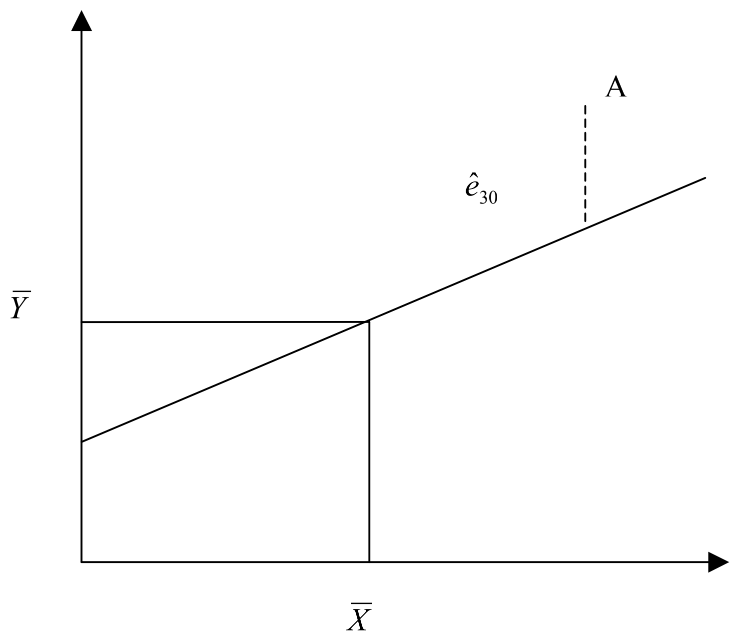
T = 40 νοικοκυριά

$$b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 7{,}383 \\ 0{,}232 \end{bmatrix}$$

Η γραμμή παλινδρόμησης με βάση το δείγμα είναι:

$$\hat{Y}_i = 7{,}383 + 0{,}232 X_i$$

Θεωρούμε ένα σημείο A της γραμμής παλινδρόμησης. Συγκεκριμένα, αντιπροσωπεύει το εισόδημα και τη δαπάνη του 30^{ου} νοικοκυριού του δείγματος.



Ερμηνεία των συντελεστών της παλινδρόμησης

$b_2 = 0{,}232$: Εάν το εβδομαδιαίο εισόδημα αυξηθεί κατά 100€, η μέση δαπάνη για φαγητό θα αυξηθεί κατά 23€

$b_1 = 7{,}383$: Είναι η μέση εβδομαδιαία δαπάνη για φαγητό όταν το εισόδημα είναι μηδενικό (η τιμή της εξαρτημένης μεταβλητής).

Πρόβλεψη: Ποια είναι η μέση εβδομαδιαία δαπάνη για φαγητό ενός νοικοκυριού με μέσο εισόδημα 100€

$$\hat{Y} = 7{,}383 + 0{,}232 * 100 = 30{,}61\,€$$

Οι προβλέψεις αντιστοιχούν σε σημεία πάνω στη γραμμή της εκτιμηθείσας παλινδρόμησης.

Διακύμανση – Συνδιακύμανση του τυχαίου σφάλματος

Στη μορφή της **γραμμικής άλγεβρας** ισχύουν τα εξής:

$$\text{Var} - \text{Cov (e)} = \text{E}\left[(e - E(e))(e - E(e))'\right] = \text{E}\left[(e-0)(e-0)'\right] = \text{E}\left[e(e)'\right] =$$

$$= E\begin{bmatrix} e_1 \\ e_2 \\ .... \\ e_T \end{bmatrix}\begin{bmatrix} e_1 & e_2 & ... & e_T \end{bmatrix} = E\begin{bmatrix} e_1^2 & e_1e_2 & ... & e_1e_T \\ e_2e_1 & e_2^2 & ... & e_2e_T \\ ..... & ... & ... & ... \\ e_1e_T & e_2e_T & ... & e_T^2 \end{bmatrix} =$$

$$\begin{bmatrix} E(e_1)^2 & E(e_1e_2) & ... & E(e_1e_T) \\ E(e_2e)_1 & E(e_2^2) & ... & E(e_2e_T) \\ ..... & .... & .... & ..... \\ E(e_1e_T) & E(e_2e_T) & ... & E(e_T^2) \end{bmatrix}$$

Αυτό σημαίνει πως όλα τα διαγώνια στοιχεία της μήτρας είναι οι διακυμάνσεις του τυχαίου όρου ενώ τα μη διαγώνια στοιχεία είναι οι συνδιακυμάνσεις

Λόγω των υποθέσεων που ισχύουν προκύπτουν τα εξής:

Var(e_i/X= x_i) = E(e_i^2) = σ^2 $= E(e_1^2) = E(e_2^2) = E(e_3^2) = ... = E(e_T^2)$, για κάθε τιμή x_i

και E(e_i, e_j)= $E\big((e_1 - E(e_1))(e_2 - E(e_2))\big) = Cov(e_1, e_2) = 0$

Τελικά η μήτρα διακυμάνσεων – συνδιακυμάνσεων των τυχαίων όρων σφάλματος θα πάρει την εξής μορφή:

$$Var - Cov(e'e) = \begin{bmatrix} \sigma^2 & 0 & ... & 0 \\ 0 & \sigma^2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & .. & \sigma^2 \end{bmatrix} = \sigma^2 I$$

Η μήτρα διακυμάνσεων – συνδιακυμάνσεων αποδίδει το σύνολο των υποθέσεων που αφορούν τους τυχαίους όρους σφάλματος.

Ο εκτιμητής bols δίνεται από την σχέση

$$b_{OLS} = (X'X)^{-1} X'Y$$

Η ύπαρξη της αντιστρόφου μήτρας εξασφαλίζεται όταν η μήτρα Χ΄Χ είναι μη ιδιάζουσα. Πράγματι η αντίστροφη μήτρα δίνεται από την σχέση:

$$(X'X)^{-1} = \frac{adj(X'X)}{|X'X|}$$

- Οι **εκτιμητές b_{OLS} είναι τυχαίες μεταβλητές** γιατί μπορούν να πάρουν πολλές διαφορετικές τιμές μια και είναι δυνατόν να προκύψουν από πολλά διαφορετικά δείγματα του πληθυσμού.

Ιδιότητες του εκτιμητή b_{OLS}

1. Ο εκτιμητής b_{OLS} είναι **αμερόληπτος**, δηλαδή έχουμε ότι: **$E[b_{OLS}] = \beta$.**

2. Ο εκτιμητής b_{OLS} είναι **αποτελεσματικός**, δηλαδή έχει την **μικρότερη δυνατή διακύμανση**: Αυτό σημαίνει ότι αν έχουμε δύο αμερόληπτους εκτιμητές $b_{1\ OLS}$ και $b_{2\ OLS}$ με διακυμάνσεις $Var(b_{1\ OLS})$, $Var(b_{2\ OLS})$ για τις οποίες γνωρίζουμε ότι: $Var(b_{1\ OLS}) < Var(b_{2\ OLS})$ αυτός που είναι αποτελεσματικός είναι ο εκτιμητής $b_{1\ OLS}$.

3. Ο εκτιμητής b_{OLS} είναι **συνεπής,** αυτό σημαίνει ότι όσον **αυξάνουν οι παρατηρήσεις ο εκτιμητής b_{OLS} πλησιάζει το β.**
Η σχέση που ισχύει είναι:

$$\lim_{n \to \infty} P\left(|b_{OLS} - \beta| < c\right) = 1, \text{ για κάθε } c > 0$$

Ο εκτιμητής ελαχίστων τετραγώνων b_{OLS} καλείται **blue** δηλαδή είναι **άριστος, γραμμικός, αμερόληπτος εκτιμητής**.

Η **μήτρα διακυμάνσεων – συνδιακυμάνσεων του b_{OLS}** δίνεται από την σχέση:

$$Var - Cov(b_{OLS}) = \hat{\sigma}^2 (X'X)^{-1} \quad (\alpha)$$

Όπου $\hat{\sigma}^2$ είναι ένας **αμερόληπτος εκτιμητής** της διακύμανσης $Var(e_i / X = x_i) = E(e_i^2) = \sigma^2$ και δίνεται από την σχέση:

$$\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{T-2} = \frac{\hat{e}'\hat{e}}{T-2}$$

- **το 2 είναι ο αριθμός των προς εκτίμηση παραμέτρων β_0 και β_1**

Όπου T-2 είναι οι **βαθμοί ελευθερίας** του υποδείγματος $Y = \beta_1 + \beta_2 X + e$
Βαθμοί ελευθερίας: Είναι ο αριθμός των ανεξάρτητων παρατηρήσεων (X,Y). Πιο συγκεκριμένα οι βαθμοί ελευθερίας είναι ίσοι με το συνολικό αριθμό παρατηρήσεων αν αφαιρέσουμε το αριθμό των γραμμικών περιορισμών στους οποίους υπόκεινται οι παρατηρήσεις X,Y.
Στο **διμεταβλητό** υπόδειγμα οι παρατηρήσεις της εξαρτημένης μεταβλητής υπόκεινται σε **δύο περιορισμούς** που δίνονται από τις σχέσεις που ακολουθούν:

$$b_1 = \overline{Y} - b_2 X$$
$$b_2 = \frac{\sum_{i=1}^{T} (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{T} (X_i - \overline{X})^2}$$

Αυτό σημαίνει πως οι **βαθμοί ελευθερίας είναι T-2.**

Υπολογισμός της σχέσης (α)

Για τον υπολογισμό της μήτρας των διακυμάνσεων – συνδιακυμάνσεων του εκτιμητή των ελαχίστων τετραγώνων.

$$Var-Cov(b_{OLS})=E\left([b_{OLS}-E(b_{OLS})][b_{OLS}-E(b_{OLS})]'\right)\Leftrightarrow Var-Cov(b_{OLS})=$$
$$=E\left[(b_{OLS}-\beta)(b_{OLS}-\beta)'\right]$$

Με δεδομένο ότι ο εκτιμητής των ελαχίστων τετραγώνων είναι διάνυσμα (2x1) θα έχουμε:

$$Var-Cov(b_{OLS})=E\begin{bmatrix} b_1-\beta_1 & b_1-\beta_1 \\ b_2-\beta_2 & (b_2-\beta_2)' \end{bmatrix}$$

$$\Leftrightarrow Var-Cov(b_{OLS})=E\left[\begin{pmatrix} b_1-\beta_1 \\ b_2-\beta_2 \end{pmatrix}\begin{pmatrix} b_1-\beta_1 & b_2-\beta_2 \end{pmatrix}\right]\Leftrightarrow Var-Cov(b_{OLS})=$$
$$=E\begin{bmatrix} (b_1-\beta_1)^2 & (b_1-\beta_1)(b_2-\beta_2) \\ (b_2-\beta_2)(b_1-\beta_1) & (b_2-\beta_2)^2 \end{bmatrix}=\begin{bmatrix} E(b_1-\beta_1)^2 & E[(b_1-\beta_1)(b_2-\beta_2)] \\ E(b_1-\beta_1)(b_2-\beta_2) & E\left[(b_2-\beta_2)^2\right] \end{bmatrix}\Leftrightarrow$$
$$\Leftrightarrow Var-Cov(b_{OLS})=\begin{bmatrix} \mathrm{var}(b_1) & \mathrm{cov}(b_1,b_2) \\ \mathrm{cov}(b_1,b_2) & \mathrm{var}(b_2) \end{bmatrix}$$

Επιπλέον έχει δειχθεί ότι:

$$b_{OLS}=\beta+(X'X)^{-1}X'e\Leftrightarrow b_{OLS}-\beta=(X'X)^{-1}X'e$$

$$Var-Cov(b_{OLS})=E\left[(b_{OLS}-\beta)(b_{OLS}-\beta)'\right]=E\left[\left((X'X)^{-1}X'e\right)\left((X'X)^{-1}X'e\right)'\right]=$$
$$=E\left[(X'X)^{-1}X'ee'X(X'X)^{-1}\right]=(X'X)^{-1}X'E(e'e)X(X'X)^{-1}=\sigma^2(X'X)^{-1}$$

Η υπόθεση της κανονικότητας του Τυχαίου όρου

Η μέθοδος των ελαχίστων τετραγώνων δεν απαιτεί την ικανοποίηση της υπόθεσης που αφορά την κανονικότητα των καταλοίπων.

Ο εκτιμητής ελαχίστων τετραγώνων είναι γραμμική συνάρτηση της εξαρτημένης μεταβλητής και συνεπώς των καταλοίπων.

$$b_{OLS}=(X'X)^{-1}X'Y=AY=A(X\beta+e)$$

Η **κατανομή πιθανότητας του εκτιμητή** των ελαχίστων τετραγώνων **εξαρτάται από αυτήν των τυχαίων όρων σφάλματος**. Αυτό σημαίνει πως και **ο εκτιμητής ελαχίστων τετραγώνων** ακολουθεί την **κανονική κατανομή.**

Έλεγχος Υποθέσεων

Έστω **το διμεταβλητό υπόδειγμα**:

$$Y_i = \beta_1 + \beta_2 X_i + e_i$$

Θέλουμε να ελέγξουμε αν ο συντελεστής κλίσης είναι ίσος με μηδέν. Δηλαδή: $\beta_2 = 0$

Ορίζουμε την μηδενική και την εναλλακτική υπόθεση.

Πιο συγκεκριμένα:

$$H_o : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$

Είναι γνωστό από την θεωρία ότι:

$$t = \frac{b_2 - \beta_2}{\sqrt{\operatorname{var} b_2}} \sim t_{T-2}$$

Όταν ισχύει η μηδενική υπόθεση τότε έχουμε ότι:

$$t = \frac{b_2}{\sqrt{\operatorname{var} b_2}} \sim t_{T-2}$$

t: η στατιστική ελέγχου της μηδενικής υπόθεσης

Η διαδικασία ελέγχου υποθέσεων περιλαμβάνει τα εξής βήματα:

1. Επιλέγουμε το επίπεδο σημαντικότητας α (σφάλμα τύπου Ι). Το α δηλώνει την πιθανότητα να απορρίψουμε τη μηδενική υπόθεση ενώ αυτή είναι σωστή

2. Βρίσκουμε την κριτική τιμή βάσει του πίνακα έτσι ώστε:

$$P[|t_{T-2}| > t_c] = a$$

3. Υπολογίζουμε την στατιστική: $t = \dfrac{b_2 - \beta_2}{\sqrt{\operatorname{var} b_2}}$, ***η οποία στην μηδενική υπόθεση β*** $_2$ ***= 0 παίρνει την μορφή*** $t = \dfrac{b_2}{\sqrt{\operatorname{var} b_2}}$

4. Αποδεχόμαστε την μηδενική υπόθεση όταν: $|t| < t_c$

Το μέγεθος του διαστήματος εμπιστοσύνης καθώς και το αποτέλεσμα του ελέγχου υποθέσεων είναι συνάρτηση:

1. Του μεγέθους του δείγματος
2. Του αριθμού των παραμέτρων προς εκτίμηση.

Αυτό ισχύει επειδή οι δύο παραπάνω παράγοντες καθορίζουν τους βαθμούς ελευθερίας και συνεπώς την κριτική τιμή t_c, όπως αυτή ορίζεται από τους πίνακες κατανομής.

Ο λόγος (t-ratio) του $\mathbf{b}_{i\ OLS}$

Είναι ο λόγος του **εκτιμηθέντος συντελεστή προς το τυπικό του σφάλμα**.

__**Εμπειρικός κανόνας:**__ Αν η απόλυτη τιμή της εκτιμηθείσας παραμέτρου είναι μεγαλύτερη του διπλάσιου του εκτιμηθέντος τυπικού σφάλματος σε ένα δίπλευρο έλεγχο τότε η μηδενική υπόθεση απορρίπτεται και συνεπώς η υπό εξέταση παράμετρος είναι στατιστικά σημαντική.
Σε μαθηματικούς όρους το παραπάνω συνοψίζεται στα εξής:

Αν $|b_i| > 2se(\hat{b}_i)$ ή αλλιώς: $\left|\dfrac{b_i}{s.e.(\hat{b}_i)} > 2\right|$

τότε η προς εξέταση παράμετρος β_i είναι **στατιστικά σημαντική.**

Ο λόγος $\left|\dfrac{b_i}{s.e.(\hat{b}_i)}\right|$ είναι γνωστός ως **t-ratio**(b_i)**.**

$$\textbf{t-ratio}\ (b_i) = \left|\frac{b_i}{s.e.(\hat{b}_i)}\right|$$

Ο συντελεστής προσδιορισμού $\mathbf{r^2}$

Ένα υπόδειγμα μπορεί να χρησιμοποιηθεί για

α) εκτίμηση και

β) για πρόβλεψη.

Για την πραγματοποίηση καλών προβλέψεων πρέπει το υπόδειγμα να μπορεί να εξηγήσει όσο το δυνατόν μεγαλύτερο ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής.

Έστω ότι θέλουμε να συγκρίνουμε δύο υποδείγματα.

1. $Y_i = \beta_1 + \beta_2 X_i + e_i$
2. $Y_i = \beta + e_i$

Ποιο από τα δύο υποδείγματα ερμηνεύει καλύτερα τις μεταβολές της εξαρτημένης μεταβλητής. Για το δεύτερο υπόδειγμα η καλύτερη πρόβλεψη και για οποιαδήποτε νοικοκυριό είναι ο μέσος όρος $\bar{Y} = b_{OLS}$. Αυτή όμως είναι μια πολύ περιοριστική πρόβλεψη αφού δεν μπορεί να εξηγήσει τις αποκλίσεις που υπερβαίνουν ή που υπολείπονται του μέσου όρου.

Συνεπώς το πρώτο υπόδειγμα μπορεί να εξηγήσει καλύτερα γιατί τα νοικοκυριά δαπανούν περισσότερο ή λιγότερο από το μέσο νοικοκυριό.

Η συνολική απόκλιση από το μέσο όρο δίνεται από την σχέση:

$$Y - \bar{Y} = \left(Y - \hat{Y} \right) + \left(\hat{Y} - \bar{Y} \right) \quad (\alpha)$$

($Y - \bar{Y}$): Συνολική απόκλιση

Η διαφορά $\left(Y - \hat{Y} \right) = \hat{e}$ ονομάζεται **κατάλοιπο** και είναι το τμήμα εκείνο της μεταβλητότητας της εξαρτημένης μεταβλητής που δεν ερμηνεύεται από την μεταβλητότητα της ανεξάρτητης μεταβλητής. Αποτελεί την ανεξήγητη απόκλιση από τον μέσο της εξαρτημένης μεταβλητής. Οφείλεται σε παράγοντες που το υπόδειγμα δεν έλαβε υπόψη.

Η διαφορά $\left(Y - \hat{Y}\right)$ είναι το μέρος της απόκλισης που εξηγείται από την μεταβλητότητα της ανεξάρτητης μεταβλητής, αφού ισχύει:

$$\left(Y - \hat{Y}\right) = b_2\left(X - \bar{X}\right)$$

Στόχος ενός υποδείγματος είναι να περιορίζεται το απρόβλεπτο τμήμα ώστε οι προβλέψεις να είναι πιο ακριβείς.
Ο συντελεστής προσδιορισμού ορίζεται στο [0,1]. Συνεπώς οι δύο ακραίες τιμές είναι το 0 και το 1. Όσο αυξάνει το ποσοστό και τείνει προς τη μονάδα το υπόδειγμα προβλέπει καλύτερα τις τιμές του δείγματος.

Προσδιορισμός του μέτρου του συντελεστή προσδιορισμού.

Η διαδικασία αυτή λαμβάνει χώρα σε τρία στάδια:

1. Υψώνουμε στο τετράγωνο την σχέση (α)
2. Αθροίζουμε όλες τις παρατηρήσεις του δείγματος
3. Διαιρούμε με το συνολικό άθροισμα των τετραγώνων των καταλοίπων

Τετραγωνίζουμε την εξίσωση (α):

$$\left(Y_i - \bar{Y}\right)^2 = \left(\hat{Y}_i - \bar{Y}\right)^2 + \hat{e}_i^2 + 2\left(Y_i - \bar{Y}\right)\hat{e}_i$$

Αθροίζοντας όλες τις παρατηρήσεις προκύπτει η εξής σχέση:

$$\sum_{i=1}^{T} \left(Y_i - \bar{Y}\right)^2 = \sum_{i=1}^{T} \left(\hat{Y}_i - \bar{Y}\right)^2 + \sum_{i=1}^{T} \hat{e}_i^2 \quad (\beta)$$

Αυτή η σχέση προκύπτει με δεδομένο ότι: $\sum_{i=1}^{T} 2\left(Y_i - \bar{Y}\right)\hat{e}_i = 0$

Η εν λόγω σχέση προκύπτει ως εξής:

$$\sum_{i=1}^{T} \left(\hat{Y}_i - \bar{Y}\right)\hat{e}_i = \sum_{i=1}^{T} \hat{Y}_i \hat{e}_i - \sum_{t=1}^{T} \bar{Y}\hat{e}_i = \sum_{i=1}^{T} \left(b_1 + b_2 X_i\right)\hat{e}_i - \bar{Y}\sum_{i=1}^{T} \hat{e}_i = \sum_{i=1}^{T} \left(b_1 \hat{e}_i + b_2 X_i \hat{e}_i\right) - \bar{Y}\sum \hat{e}_i =$$
$$= b_1 \sum \hat{e}_i + b_2 \sum X_i e_i - \bar{Y}\sum e_i = 0$$

Αντικαθιστούμε στην σχέση (β) τα εξής:

$\sum_{i=1}^{T} \left(\hat{Y}_i - \bar{Y}\right)^2$ = ΑΤΠ - Άθροισμα τετραγώνων της παλινδρόμησης (SSR)

$\sum_{i=1}^{T} \hat{e}_i^2$ = ΑΤΚ - Άθροισμα τετραγώνων των καταλοίπων (RSS)

$\sum_{i=1}^{T} \left(Y_i - \bar{Y}\right)^2$ = ΣΑΤ - Συνολικό άθροισμα τετραγώνων (TSS)

Αυτό σημαίνει πως η σχέση (β) μπορεί να γραφεί ως εξής:

ΣΑΤ = ΑΤΠ + ΑΤΚ (γ)

ή TSS = SSR + RSS

Διαιρώντας την (γ) με την τιμή ΣΑΤ λαμβάνεται η σχέση (δ).
Δηλαδή:

$$1 = \frac{\text{ΑΤΠ}}{\text{ΣΑΤ}} + \frac{\text{ΑΤΚ}}{\text{ΣΑΤ}} = r^2 + \frac{\text{ΑΤΚ}}{\text{ΣΑΤ}}, r^2 = \frac{\sum_{i=1}^{T}\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum_{i=1}^{T}\left(Y_i - \bar{Y}\right)^2} \qquad (\delta)$$

Ο συντελεστής προσδιορισμού μπορεί να υπολογισθεί και από τους τύπους:

$$r^2 = \frac{\sum_{i=1}^{T}\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum_{i=1}^{T}\left(Y_i - \bar{Y}\right)^2} = \frac{\sum_{i}^{T}\hat{y}_i^2}{\sum_{i}^{T} y_i^2} = \frac{b_2^2 \sum_{i}^{T} x_i^2}{\sum_{i}^{T} x_i^2} = \frac{(\sum_{i}^{T} x_i y_i)^2}{(\sum_{i}^{T} x_i^2)(\sum_{i}^{T} y_i^2)}$$

Όπου $x_i = X_i - \bar{X}$ και $y_i = Y_i - \bar{Y}$.

Ο συντελεστής προσδιορισμού δίνει το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής που ερμηνεύεται από τη μεταβλητότητα της ανεξάρτητης μεταβλητής.

Παράδειγμα:

r^2 = 0,317. Αυτό σημαίνει πως το 31,7% της μεταβλητότητας της εξαρτημένης μεταβλητής (απόκλισης από τον μέσο όρο) του δείγματος ερμηνεύεται από την μεταβλητότητα του ανεξάρτητης μεταβλητής. Το υπόλοιπο 68,3% δεν ερμηνεύεται από το υπόδειγμα.

Χαμηλός συντελεστής προσδιορισμού συνεπάγεται **μικρή ικανότητα πρόβλεψης.**

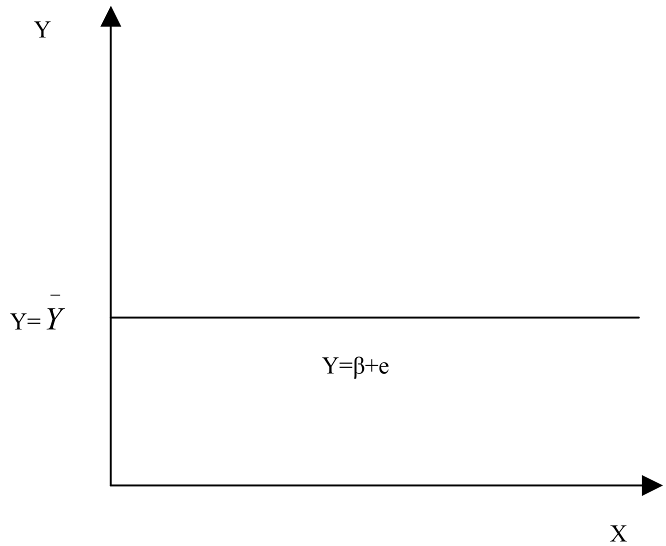
Υψηλός συντελεστής προσδιορισμού συνεπάγεται **υψηλή ικανότητα πρόβλεψης**

Ο **συντελεστής προσδιορισμού** είναι ένα **μέτρο της γραμμικής σχέσης** μεταξύ της **εξαρτημένης** και της **ανεξάρτητης μεταβλητής**.

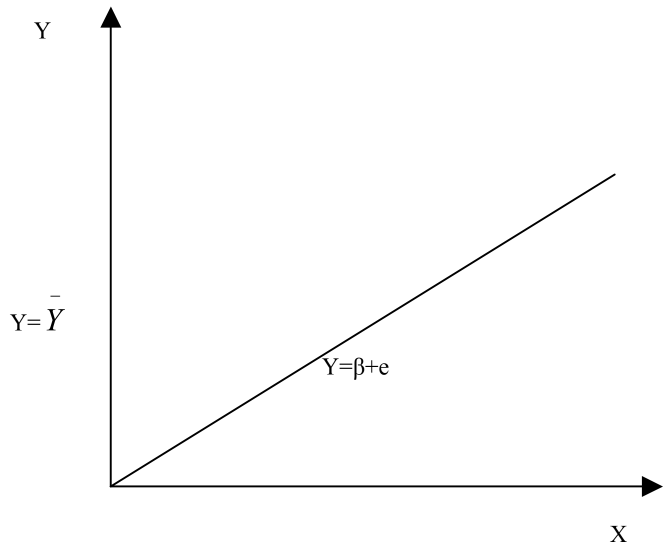
Όσο μεγαλύτερος είναι ο συντελεστής προσδιορισμού τόσο ισχυρότερη είναι η σχέση μεταξύ του Y και του $\hat{Y}$.

Όπως έχει προαναφερθεί ο συντελεστής προσδιορισμού παίρνει ως ακραίες τιμές το 0 και το 1.

r^2=0. Η παλινδρόμηση δεν ερμηνεύει καθόλου την μεταβλητότητα της εξαρτημένης μεταβλητής. Χαρακτηριστικό παράδειγμα μιας τέτοιας περίπτωσης είναι το μονομεταβλητό υπόδειγμα για το οποίο ισχύει: TSS = RSS, SSR=0 Δηλαδή: $\hat{Y}_t = \bar{Y}$. Σε αυτή την περίπτωση η παλινδρόμηση είναι μία ευθεία γραμμή.



r^2=1: Η παλινδρόμηση ερμηνεύει όλη την μεταβλητότητα της εξαρτημένης μεταβλητής. Δηλαδή ισχύει RSS =0. Σε αυτή την περίπτωση το σύνολο των παρατηρήσεων του δείγματος βρίσκεται στην γραμμή της παλινδρόμησης. Πιο συγκεκριμένα:



Το ενδεχόμενο αυτό είναι αδύνατο να συμβεί στην πράξη.

Υψηλή τιμή του συντελεστή προσδιορισμού δεν συνεπάγεται κακό υπόδειγμα. Αυτή είναι συνάρτηση του είδους της οικονομετρικής ανάλυσης.

1) **Χρονολογικές σειρές**: Αναμένεται υψηλή τιμή του συντελεστή προσδιορισμού λόγω της κοινής τάσης στις μεταβλητές.
2) **Διαστρωματικά στοιχεία**: Εμφανίζουν χαμηλότερη τιμή του συντελεστή προσδιορισμού.

Την μεγαλύτερη σημασία έχουν τα χαρακτηριστικά των εκτιμήσεων από τα οποία τα σημαντικότερα είναι η στατιστική σημαντικότητα των εκτιμήσεων καθώς και η συνέπεια αυτών σύμφωνα με την οικονομική θεωρία.

Η σχέση ΣΑΤ = ΑΤΠ + ΑΤΚ ισχύει στις ακόλουθες περιπτώσεις:

1) Η **μέθοδος εκτίμησης** είναι αυτή των **ελαχίστων τετραγώνων**
2) Το **υπόδειγμα είναι γραμμικό**

3) Το **υπόδειγμα** περιλαμβάνει **σταθερό όρο**.

Αν κάποιες από τις παραπάνω υποθέσεις δεν ισχύει τότε ο συντελεστής προσδιορισμού ορίζεται με διαφορετικό τρόπο.

Ο συντελεστής συσχέτισης r

Δίνεται από τον τύπο:

$$r = \frac{\sum_{i}^{T} x_i y_i}{\sqrt{(\sum_{i}^{T} x_i^2)(\sum_{i}^{T} y_i^2)}}$$

Εκφράζει την συσχέτιση των μεταβλητών X και Y . Παίρνει τιμές μεταξύ του -1 και 1. Όταν παίρνει αρνητικές τιμές έχουμε αρνητική συσχέτιση, ενώ όταν παίρνει θετικές τιμές έχουμε θετική συσχέτιση μεταξύ των μεταβλητών.

Αρνητική συσχέτιση των X και Y σημαίνει ότι αυξανομένου του X μειώνεται το Y και αντιστρόφως.

Θετική συσχέτιση των X και Y σημαίνει ότι αυξανομένου του X αυξάνεται το Y και αντιστρόφως.

Όταν r = 0 δεν έχουμε συσχέτιση των X και Y.

Ο συντελεστής συσχέτισης r μετρά την διασπορά των τιμών του δείγματος γύρω από την ευθεία παλινδρόμησης. Όσο η απόλυτη τιμή του συντελεστή συσχέτισης είναι πολύ κοντά στο 1 τόσο η διασπορά των τιμών του δείγματος γύρω από την ευθεία παλινδρόμησης είναι μικρή και αντίστροφα.

Αξιολόγηση του συνόλου του υποδείγματος

Η σημαντικότητα του συνόλου του υποδείγματος μπορεί να ελεγχθεί με την τιμή F – Στατιστικής η οποία στο **διμεταβλητό υπόδειγμα** ισούται με t^2 ($F = t^2$).
Η F – Στατιστική δίνεται από την σχέση:

$$F = \frac{(\text{ΑΤΠ})/(\kappa - 1)}{(\text{ΑΤΚ})/(T - \kappa)} = \frac{r^2/(k-1)}{(1-r^2)/(T-k)}$$

Ο καλύτερος τρόπος για να ελεγχθεί το συνολικό υπόδειγμα είναι να κατασκευασθεί ο **πίνακας ανάλυσης της μεταβλητότητας** της **εξαρτημένης μεταβλητής**.

Μεταβλητότητα της εξαρτημένης μεταβλητής

Πηγή μεταβλητότητας	Αθροίσματα Τετραγώνων	Βαθμοί ελευθερίας	Μέσοι	F
Παλινδρόμηση Χ	ΑΤΠ	κ-1	ΑΤΠ/(κ-1)	F
Κατάλοιπα e	ΑΤΚ	Τ-κ	ΑΤΚ/(Τ-κ)	
Σύνολο	ΣΑΤ	Τ-1	ΣΑΤ/(Τ-1)	

Όπου κ ο αριθμός των προς εκτίμηση παραμέτρων (στο διμεταβλητό υπόδειγμα έχουμε κ = 2)

Με την F – Στατιστική ελέγχεται η μηδενική υπόθεση:

$$H_0: \beta_1 = \beta_2 = 0$$
$$H_1: \text{Ένα εκ των } \beta_i \neq 0$$

σε επίπεδο σημαντικότητας α

Αν F – Στατιστική < $F_{(\alpha/2, \kappa-1, T-\kappa)}$ η H_0 γίνεται δεκτή

Αν F – Στατιστική > $F_{(\alpha/2, \kappa-1, T-\kappa)}$ η H_0 γίνεται απορρίπτεται

- ✓ Η **αποδοχή της H_0** σημαίνει ότι η **ανεξάρτητη μεταβλητή X δεν επηρεάζει σημαντικά την εξαρτημένη μεταβλητή.**

Σε περίπτωση που εκτιμηθούν οι συντελεστές του υποδείγματος με την θεωρία των πινάκων και στον πίνακα μεταβλητότητας της εξαρτημένης μεταβλητής μπορεί να χρησιμοποιηθούν πίνακες για να περιγράψουν τα μεγέθη ΣΑΤ, ΑΤΠ, ΑΤΚ, F.

Αυτό θα παρουσιασθεί στην εξέταση του γενικευμένου κλασικού υποδείγματος $Y_i = \beta_0 X_{0i} + \beta_1 X_{1i} + \ldots\ldots\ldots\ldots + \beta_k X_{ki} + \varepsilon_i$ που ακολουθεί.

Το γενικευμένο κλασικό υπόδειγμα

$Y_i = \beta_0 X_{0i} + \beta_1 X_{1i} + \ldots\ldots\ldots\ldots + \beta_k X_{ki} + \varepsilon_i$

(με την μέθοδο των πινάκων -μητρών).

Παρουσίαση του υποδείγματος

$Y_i = \beta_0 X_{0i} + \beta_1 X_{1i} + \ldots\ldots\ldots\ldots + \beta_k X_{ki} + \varepsilon_i$

Ο πρώτος δείκτης 0,1, 2,…, κ αναφέρεται στην ερμηνευτική – ανεξάρτητη μεταβλητή και ο δεύτερος δείκτης στην i - παρατήρηση του δείγματος.

Επομένως για Τ παρατηρήσεις θα προκύψει το σύστημα:

$Y_1 = \beta_0 X_{01} + \beta_1 X_{11} + \ldots\ldots\ldots\ldots.. + \beta_k X_{k1} + \varepsilon_1$

$Y_2 = \beta_0 X_{02} + \beta_1 X_{12} + \ldots\ldots\ldots\ldots.. + \beta_k X_{k2} + \varepsilon_2$

…………………….

……………………

$Y_T = \beta_0 X_{0T} + \beta_1 X_{1T} + \ldots\ldots\ldots\ldots.. + \beta_k X_{kT} + \varepsilon_T$

Υπό μορφή πινάκων το παραπάνω σύστημα παίρνει την μορφή:

$$\mathbf{Y = Xb + e}$$

Όπου

Όπου $\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_T \end{bmatrix}$, πίνακας στήλη Τ x 1, των τιμών της εξαρτημένης μεταβλητής

$\mathbf{X} = \begin{bmatrix} X_{01}..X_{11}...........X_{K1} \\ X_{02}..X_{12}..........X_{K2} \\ . \\ . \\ X_{0T}..X_{1T}..........X_{KT} \end{bmatrix}$, πίνακας Tx(κ+1), των τιμών των ανεξαρτήτων μεταβλητών με τα στοιχεία της πρώτης μεταβλητής ίσα με τη μονάδα.

$$\mathbf{b} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_\kappa \end{bmatrix}$$, πίνακας στήλη (κ+1) x 1 των συντελεστών παλινδρόμησης

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ e_T \end{bmatrix}$$, πίνακας στήλη T x 1, των τιμών του διαταρακτικού όρου.

Όπως αναφέρθηκε οι κανονικές εξισώσεις λαμβάνονται από την ελαχιστοποίηση των τετραγώνων των καταλοίπων.

Τα κατάλοιπα υπό την μορφή των πινάκων υπολογίζονται από την παλινδρόμηση του δείγματος.

$$\mathbf{Y = Xb + e} \text{ άρα } \mathbf{e = Y - Xb}$$

Έχουμε ότι:

$$e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \cdots \\ e_T \end{bmatrix} \quad \text{και} \quad \hat{e} = \begin{bmatrix} \hat{e}_1 \\ \hat{e}_2 \\ \hat{e}_3 \\ \cdots \\ \hat{e}_T \end{bmatrix}$$, ο ανάστροφός του πίνακα στήλη δίνεται από το διάνυσμα:

$$\hat{e}' = [\hat{e}_1 \, \hat{e}_2 \ldots\ldots\ldots \hat{e}_T]$$

Επομένως η συνάρτηση που ελαχιστοποιείται είναι η:

$$\hat{e}'\hat{e} = \begin{bmatrix} \hat{e}_1 & \hat{e}_2 & \ldots & \hat{e}_T \end{bmatrix} \begin{bmatrix} \hat{e}_1 \\ \hat{e}_2 \\ \ldots \\ \hat{e}_T \end{bmatrix} = \hat{e}_1^2 + \hat{e}_2^2 + \ldots\ldots + \hat{e}_T^2 = \sum_{i=1}^{T} \hat{e}_i^2$$

Συνεπώς :

$$\begin{aligned} &\min \sum_{i=1}^{T} \hat{e}_i^{2} = \min(Y - Xb)'(Y - Xb) = \min(Y' - b'X')(Y - Xb) = \\ &= \min(Y'Y - Y'Xb - b'X'Y + b'X'Xb) \end{aligned} \qquad \textbf{(I)}$$

Οι διαστάσεις των αντιστρόφων πινάκων δίνονται ως ακολούθως:

Y΄: 1 x T

b΄: 1 x (κ+1)

X΄: (κ+1) x T

Τα γινόμενα $Y'Xb$ και $b'X'Y$ ορίζονται ως βαθμωτά μεγέθη αφού έχουν διαστάσεις 1x1. Ο $Y'Xb$ είναι ο ανάστροφος του $b'X'Y$

Πιο συγκεκριμένα ισχύει:

$$(b'X'Y)' = Y'(X')'(b')' = Y'Xb$$

Επειδή και οι δύο όροι είναι βαθμωτά μεγέθη και $(b'X'Y)' = Y'Xb$ δίνουν άθροισμα $2Y'Xb = 2b'X'Y$

Άρα η σχέση (I) γίνεται: $\min(Y'Y - 2b'X'Y + b'X'Xb)$

Για την ελαχιστοποίηση απαιτείται:

$$\frac{\partial(\hat{e}'\hat{e})}{\partial b} = 0 \Leftrightarrow -2X'Y + 2X'Xb = 0$$

Η συνθήκη πρώτης τάξης δίνει δύο εξισώσεις αφού το b είναι ένα διάνυσμα. Οι εξισώσεις αυτές ονομάζονται **κανονικές εξισώσεις και είναι:**

$$\mathbf{b}_{OLS} = \mathbf{(X'X)^{-1}X'Y}$$

Όπου

$$\mathrm{b}_{OLS} = \begin{bmatrix} b_o \\ b_1 \\ b_2 \\ .... \\ b_k \end{bmatrix}, \quad X'X = \begin{bmatrix} T..............\sum X_1................\sum X_k \\ \sum X_1..........\sum X_1^2................\sum X_k X_1 \\ .... \\ .... \\ \sum X_k...\sum X_1 X_k.........................\sum X_k^2 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} \sum Y \\ \sum X_1 Y \\ .... \\ .... \\ \sum X_k Y \end{bmatrix}, \quad \mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ .... \\ .... \\ 0 \end{bmatrix}$$ μηδενικός πίνακας – στήλη (κ+1) x 1

Στη μορφή της **γραμμικής άλγεβρας** ισχύουν τα εξής:

$$\text{Var – Cov (e)} = \mathrm{E}\left[(e - E(e))(e - E(e))'\right]$$

όπου

$$e=\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ .... \\ e_T \end{bmatrix}$$, ενώ εξ υποθέσεως ισχύουν τα εξής:

$$Var-Cov(e)=E(ee')=E\begin{bmatrix} e_1 \\ e_2 \\ .... \\ e_T \end{bmatrix}\begin{bmatrix} e_1 & e_2 & \dots & e_T \end{bmatrix}=E\begin{bmatrix} e_1^2 & e_1e_2 & \dots & e_1e_T \\ e_2e_1 & e_2^2 & \dots & e_2e_T \\ ..... & \dots & \dots & \dots \\ e_1e_T & e_2e_T & \dots & e_T^2 \end{bmatrix}=$$

$$\begin{bmatrix} E(e_1)^2 & E(e_1e_2) & \dots & E(e_1e_T) \\ E(e_2e)_1 & E(e_2^2) & \dots & E(e_2e_T) \\ ..... & .... & .... & ..... \\ E(e_1e_T) & E(e_2e_T) & \dots & E(e_T^2) \end{bmatrix}$$

Τελικά με βάση τις υποθέσεις που αναφέρθηκαν και διέπουν τα σφάλματα καταλήγουμε στον ίδιο τύπο.

$$Var-Cov(e'e)=\begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & .. & \sigma^2 \end{bmatrix}=\sigma^2 I$$

Εκτιμητές των διακυμάνσεων – συνδιακυμάνσεων των συντελεστών

$$Var-Cov(b_{OLS})=E(b_{OLS}b_{OLS}')=E\{[b_{OLS}-E(b_{OLS})][b_{OLS}-E(b_{OLS})]'\}=$$

$$E\begin{bmatrix} b_0 \\ b_1 \\ .... \\ b_k \end{bmatrix}\begin{bmatrix} b_0 & b_1 & ...... & b_k \end{bmatrix}=E\begin{bmatrix} b_0^2 & b_0b_1 & \dots & b_0b_k \\ b_1b_0 & b_1^2 & \dots & b_1b_k \\ ..... & \dots & \dots & \dots \\ b_kb_0 & b_kb_1 & \dots & b_k^2 \end{bmatrix}=$$

$$\begin{bmatrix} E(b_0)^2 & E(b_0b_1) & \dots & E(b_0b_k) \\ E(b_1b_0) & E(b_1^2) & \dots & E(b_1b_k) \\ ..... & .... & .... & ..... \\ E(b_kb_0) & E(b_kb_1) & \dots & E(b_k^2) \end{bmatrix}$$

$$= \begin{bmatrix} Var(b_0) & Cov(b_0 b_1) & \ldots & Cov(b_0 b_k) \\ Cov(b_1 b_0) & Var(b_1) & \ldots & Cov(b_1 b_k) \\ \ldots\ldots & \ldots & \ldots & \ldots\ldots \\ Cov(b_k b_0) & Cov(b_k b_1) & \ldots & Var(b_k) \end{bmatrix} = \hat{\sigma}^2 (X'X)^{-1}$$

Τελικά έχουμε: Var – Cov(b_{OLS}) = $\hat{\sigma}^2 (X'X)^{-1}$

Όπου η $\hat{\sigma}^2$ η εκτίμηση της σ^2 που δίνεται από την σχέση:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{T} e_i^2}{T-k-1} = \frac{Y'Y - b_{OLS}{}' X'Y}{T-k-1}$$

Όπου κ+1 ο αριθμός των προς εκτίμηση παραμέτρων – συντελεστών του υποδείγματος.

Ο Συντελεστής πολλαπλού προσδιορισμού $\mathbf{R}^2_{y.x_1x_2\ldots x_k}$

Ο συντελεστής $R^2_{y.x_1x_2\ldots x_k}$ μετράει το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής που ερμηνεύεται από τις ανεξάρτητες μεταβλητές και δίνεται από τον τύπο:

$$\mathbf{R}^2_{y.x_1x_2\ldots x_k} = \frac{\text{ΑΤΠ}}{\text{ΣΑΤ}} = \frac{b'_{OLS} X'Y - T\bar{Y}^2}{Y'Y - T\bar{Y}^2}$$

Αθροίσματα Τετραγώνων σε μορφή πινάκων

ΣΑΤ = $Y'Y - T\bar{Y}^2$

ΑΤΚ = $Y'Y - b'_{OLS} X'Y = (1 - \mathbf{R}^2_{y.x_1x_2\ldots x_k})(Y'Y - T\bar{Y}^2)$

ΑΤΠ = $b'_{OLS} X'Y - T\bar{Y}^2 = \mathbf{R}^2_{y.x_1x_2...x_k} (Y'Y - T\bar{Y}^2)$

Αξιολόγηση του υποδείγματος

1. Αξιολόγηση των ατομικών συντελεστών

Ελέγχεται η μηδενική υπόθεση:

$H_0: \beta_i = 0$ με $H_1: \beta_i \neq 0$
Σε επίπεδο σημαντικότητας α

Χρησιμοποιούμε το t - ratio$= \dfrac{b_i}{\sqrt{\text{var}\, b_i}}$, όπου $\sqrt{Var(b_i)}$ το τυπικό σφάλμα του b_i που δίνεται από την τετραγωνική ρίζα του αντίστοιχου διαγώνιου στοιχείου του πίνακα $\hat{\sigma}^2 (X'X)^{-1}$.

Επίσης μπορούμε να εκτιμήσουμε το διάστημα εμπιστοσύνης για τον συντελεστή β_i από την σχέση:

$$b_i - \sqrt{Var(b_i)}\ t_{\alpha/2, T-\kappa-1} \leq \beta_i \leq b_i + \sqrt{Var(b_i)}\ t_{\alpha/2, T-\kappa-1}$$

2. Αξιολόγηση ολοκλήρου του υποδείγματος

Όπως αναφέρθηκε η σημαντικότητα του συνόλου του υποδείγματος μπορεί να ελεγχθεί με την τιμή F – Στατιστικής .

Η F – Στατιστική δίνεται από την σχέση:

$$F = \frac{(\text{ΑΤΠ})/(\kappa+1-1)}{(\text{ΑΤΚ})/(T-\kappa-1)} = \frac{R^2{}_{y.x_1x_2...x_k} / k}{(1-R^2{}_{y.x_1x_2...x_K})/(T-k-1)} =$$

$$(\frac{b'_{OLS}X'Y - T\bar{Y}^2}{k})/(\frac{Y'Y - b'_{OLS}X'Y}{T-k-1})$$

Ο καλύτερος τρόπος για να ελεγχθεί το συνολικό υπόδειγμα είναι να κατασκευασθεί ο πίνακας ανάλυσης της μεταβλητότητας της εξαρτημένης μεταβλητής

Μεταβλητότητα της εξαρτημένης μεταβλητής

Πηγή μεταβλητότητας	Αθροίσματα Τετραγώνων	Βαθμοί ελευθερίας	Μέσοι	F
Παλινδρόμηση Χ	$ΑΤΠ = b'_{OLS}X'Y - T\bar{Y}^2$ $= R^2_{y.x_1x_2...x_k}(Y'Y - T\bar{Y}^2)$	κ	ΑΤΠ/κ	F
Κατάλοιπα e	$ΑΤΚ = Y'Y - b'_{OLS}X'Y =$ $(1 - R^2_{y.x_1x_2...x_k})(Y'Y - T\bar{Y}^2)$	Τ-κ-1	ΑΤΚ/(Τ-κ-1)	
Σύνολο	$ΣΑΤ = Y'Y - T\bar{Y}^2$	Τ-1	ΣΑΤ/(Τ-1)	

(Όπου κ+1 ο αριθμός των προς εκτίμηση παραμέτρων)

Με την F – Στατιστική ελέγχεται η μηδενική υπόθεση:

$$H_0: \beta_1 = \beta_2 = \ldots\ldots = \beta_k = 0$$

$$H_1: \text{Ένα εκ των } \beta_i \neq 0$$

σε επίπεδο σημαντικότητας α

Αν F – Στατιστική < $F_{(\alpha,\kappa,T-\kappa-1)}$ η H_0 γίνεται δεκτή

Αν F – Στατιστική > $F_{(\alpha,\kappa,T-\kappa-1)}$ η H_0 γίνεται απορρίπτεται

- ✓ **Η αποδοχή της H_0** σημαίνει ότι **η ανεξάρτητη μεταβλητή X δεν επηρεάζει σημαντικά την εξαρτημένη μεταβλητή.**

Παραδείγματα:

1. Παράδειγμα (χρησιμοποιούνται αθροίσματα)

Έστω τα ακόλουθα δεδομένα που αφορούν δαπάνες για διατροφή (Y) και διαθέσιμο εισόδημα (X), ανά μήνα, οκτώ νοικοκυριών. (σε εκατοντάδες ευρω)

Έτος	**Δαπάνες (Y)**	**Εισόδημα (X)**	**XY**
2000	60	110	6600
2001	62	113	7006
2002	65	116	7540
2003	67	119	7973
2004	70	120	8400
2005	73	121	8833
2006	74	125	9250
2007	78	130	10140

Να εφαρμοσθεί το γραμμικό μοντέλο

$Y_i = \beta_1 + \beta_2 X_i + e_i$ στα δεδομένα.

Από τις κανονικές εξισώσεις έχουμε ότι:

$$b_1 = \overline{Y} - b_2 \overline{X}$$

$$b_2 = \frac{\sum_{i=1}^{T}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{T}(X_i - \overline{X})^2}$$

Επίσης το b_2 μπορεί να υπολογισθεί από τον τύπο:

$$b_2 = \frac{T\sum_{i}^{T} X_i Y_i - \sum_{i}^{T} X_i \sum_{i}^{T} Y_i}{T\sum_{i}^{T} X_i^2 - (\sum_{i}^{T} X_i)^2} = \frac{8*65742 - 954*549}{8*114052 - (954)^2} = 0{,}95$$

Όπου T = 8, $\sum_{i=1}^{8} Y_i$ = 549, $\sum_{i=1}^{8} X_i = 954$, $\sum_{i=1}^{8} X_i^2$ = 114.052, $\sum_{i=1}^{8} X_i Y_i$= 65.742

Το b_1 = $\overline{Y} - b_2 \overline{X}$ = 68,63 – 0,95* 119,25= - 44,66

Όπου $\overline{Y} = 68{,}63$ και $\overline{X}$ = 119,25.

Το μοντέλο που εκτιμάται είναι:

$$\hat{Y}_i = -44{,}66 + 0{,}95 X_i$$

Η εκτίμηση της διακύμανσης του διαταρακτικού όρου είναι ίση με:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{T} \hat{e}_i^2}{T-k} = \frac{11,194}{8-2} = 1,87$$

Το άθροισμα των τετραγώνων των σφαλμάτων είναι:

$$\sum_{i=1}^{8} \hat{e}_i^2 = 11,194$$

Οι εκτιμήσεις των διακυμάνσεων των δύο εκτιμητών b_1, b_2 είναι:

$$Var(b_1) = \hat{\sigma}^2 \left[\frac{\sum_{i}^{8} X_i^2}{8\sum_{i=1}^{8}(X_i - \bar{X})^2} \right] = 1,87^* \frac{114.052}{8 * 287,48} = 92,73$$

Όπου $\sum_{i=1}^{8}(X_i - \bar{X})^2 = \sum_{i=1}^{8} x_i^2 = 287,48$

και $$Var(b_2) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{8}\left(X_i - \bar{X}\right)^2} = \frac{1,87}{287,48} = 0,006$$

Τα τυπικά σφάλματα των b_1, b_2 είναι αντίστοιχα $s.e(b_1) = \sqrt{92,73} = 9,63$ και $s.e(b_2) = \sqrt{0,006} = 0,08$.

Αξιολόγηση των συντελεστών του υποδείγματος

Προσδιορίζω τον λόγο **t-ratio**(b_i) = $\left| \dfrac{b_i}{s.e.(\hat{b}_i)} \right|$ για κάθε συντελεστή του υποδείγματος.

Έχουμε:

t-ratio(b_1) = $\left| \dfrac{-44{,}66}{9{,}63} \right| = 4{,}64$

t-ratio(b_2) = $\left| \dfrac{0{,}95}{0{,}08} \right| = 11{,}88$

Η μηδενική υπόθεση είναι:

$$H_0: \beta_i = 0 \ \text{ με}$$

$$H_0: \beta_i \neq 0$$

Επειδή $t_{a/2;T-k} = t_{0,025;6} = 2{,}447 <$ t-ratio(b_i) η μηδενική υπόθεση H_0 δεν γίνεται **δεκτή** που σημαίνει ότι **η ανεξάρτητη – ερμηνευτική μεταβλητή επηρεάζει σημαντικά την εξαρτημένη μεταβλητή.**

Διάστημα εμπιστοσύνης για τους συντελεστές $\boldsymbol{\beta}_i$

$$\mathbf{b}_i - s.e(b_i) * t_{a/2;T-k} \leq \boldsymbol{\beta}_i \leq \mathbf{b}_i + s.e(b_i) * t_{a/2;T-k}$$

Για τον συντελεστή β_1 έχουμε:

$$-44{,}66 - 9{,}63 * 2{,}447 \leq \boldsymbol{\beta}_1 \leq -44{,}66 + 9{,}63 * 2{,}447$$

$$\mathbf{-68{,}22 \leq \boldsymbol{\beta}_1 \leq 21{,}11}$$

Για τον συντελεστή β_2 έχουμε:

$$0{,}95 - 0{,}08 * 2{,}447 \leq \boldsymbol{\beta}_2 \leq 0{,}95 + 0{,}08 * 2{,}447$$

$$\mathbf{0{,}754 \leq \boldsymbol{\beta}_2 \leq 1{,}146}$$

Ο **συντελεστής προσδιορισμού** μπορεί να υπολογισθεί και από τους τύπους:

$$r^2 = \frac{\sum_{i=1}^{T}\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum_{i=1}^{T}\left(Y_i - \bar{Y}\right)^2} = \frac{\sum_{i}^{T}\hat{y}_i^2}{\sum_{i}^{T} y_i^2} = \frac{b_2^2\sum_{i}^{T} x_i^2}{\sum_{i}^{T} y_i^2} = \frac{b_2\sum_{i}^{T} x_i y_i}{\sum_{i}^{T} y_i^2} = \frac{(\sum_{i}^{T} x_i y_i)^2}{(\sum_{i}^{T} x_i^2)(\sum_{i}^{T} y_i^2)}$$

Επιλέγουμε τους τύπους:

$$r^2 = \frac{b_2^2\sum_{i}^{8} x_i^2}{\sum_{i}^{8} y_i^2} = \frac{0{,}95^2 * 287{,}48}{271{,}90} = 0{,}95$$

ή

$$r^2 = \frac{b_2\sum_{i}^{8} x_i\, y_i}{\sum_{i}^{8} y_i^2} = \frac{0{,}95 * 273{,}76}{271{,}90} = 0{,}95$$

Όπου $\sum_{i=1}^{8} y_i^2 = 271{,}90$ και $\sum_{i=1}^{8} x_i\, y_i = 273{,}76$

Άρα το 95% της μεταβλητότητας του Υ ερμηνεύεται από την μεταβλητότητα του Χ.

Ο συντελεστής συσχέτισης r

Δίνεται από τον τύπο:

$$r = \frac{\sum_{i}^{8} x_i y_i}{\sqrt{(\sum_{i}^{8} x_i^2)(\sum_{i}^{8} y_i^2)}} = \frac{273,76}{\sqrt{287,48 * 271,90}} = 0,979$$

Έχουμε ισχυρή θετική συσχέτιση μεταξύ των μεταβλητών Χ   και Υ

Αξιολόγηση του συνόλου του υποδείγματος

Η σημαντικότητα του συνόλου του υποδείγματος μπορεί να ελεγχθεί με την τιμή F – Στατιστικής η οποία στο **διμεταβλητό υπόδειγμα** ισούται με t^2 ($F = t^2$).

Η F – Στατιστική δίνεται από την σχέση:

$$\mathrm{F} = \frac{(\mathrm{AT\Pi})/(\kappa - 1)}{(\mathrm{ATK})/(T - \kappa)} = \frac{r^2/(k-1)}{(1-r^2)/(T-k)}$$

Όπου κ =2 οι προς εκτίμηση συντελεστές

Ο καλύτερος τρόπος για να ελεγχθεί το συνολικό υπόδειγμα είναι να κατασκευασθεί ο **πίνακας ανάλυσης της μεταβλητότητας** της **εξαρτημένης μεταβλητής**.

Μεταβλητότητα της εξαρτημένης μεταβλητής

Πηγή μεταβλητότητας	Αθροίσματα Τετραγώνων	Βαθμοί ελευθερίας	Μέσοι	F
Παλινδρόμηση X	ΑΤΠ=260,71	κ-1= 1	ΑΤΠ/(κ-1) = 260,71	F= 139,42
Κατάλοιπα e	ΑΤΚ= 11,19	Τ-κ = 6	ΑΤΚ/(Τ-κ) = 1,87	
Σύνολο	ΣΤΑ = 271,90	Τ-1= 7	ΣΑΤ/(Τ-1) = 38,84	

Όπου κ ο αριθμός των προς εκτίμηση παραμέτρων (στο διμεταβλητό υπόδειγμα έχουμε κ = 2)

Με την F – Στατιστική ελέγχεται η μηδενική υπόθεση:

$$H_0: \beta_1 = \beta_2 = 0$$

$$H_1: \text{Ένα εκ των } \beta_i \neq 0$$

σε επίπεδο σημαντικότητας α = 5% και 10%

Έχουμε $F_{(\alpha,\kappa-1,T-\kappa)} = F_{(0,05;1,6)} = 5{,}99$ και $F_{(\alpha,\kappa-1,T-\kappa)} = F_{(0,10;1,6)} = 3{,}78$

Εφόσον F – Στατιστική = 139,42 > $F_{(0,05;1,6)} = 5{,}99$ και F – Στατιστική = 139,42 > $F_{(0,10;1,6)} = 3{,}78$ η H_0 **δεν γίνεται δεκτή** άρα η **ανεξάρτητη μεταβλητή X επηρεάζει στατιστικά σημαντικά την εξαρτημένη μεταβλητή.**

- Έχουμε επίσης ότι στο μονομεταβλητό υπόδειγμα $Y_i = \beta_1 + \beta_2 X_i + e_i$

$F = t^2$, ($\sqrt{F} = \sqrt{139{,}42} \approx$ t-ratio(b_2) = 11,88).

2.Παράδειγμα (χρησιμοποιείται άλγεβρα πινάκων)

Έστω τα παρακάτω ετήσια στοιχεία για τις μεταβλητές Y, X_0, X_1, X_2

Παρατηρήσεις	**Y**	**X_0**	**X_1**	**X_2**
1	7	1	4	1
2	12	1	7	2
3	17	1	9	5
4	20	1	12	8
Σύνολο	**56**	**4**	**32**	**16**

Να εκτιμηθεί το πολυμεταβλητό υπόδειγμα

$$\mathbf{Y}_i = \boldsymbol{\beta}_0 \mathbf{X}_{0i} + \boldsymbol{\beta}_1 \mathbf{X}_{1i} + \boldsymbol{\beta}_2 \mathbf{X}_{2i} + \mathbf{e}_i$$

(πηγή: Α. Ανδρικόπουλος, Οικονομετρία, τόμος Α, 2003)

Υπό μορφή πινάκων το υπόδειγμα έχει την ακόλουθη μορφή:

$$Y = X\,b + e$$

Όπου

$$Y = \begin{bmatrix} 7 \\ 12 \\ 17 \\ 20 \end{bmatrix}, \quad X = \begin{bmatrix} 1.....4......1 \\ 1......7......2 \\ 1.......9......5 \\ 1......12.....8 \end{bmatrix}, \quad b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}, \quad e = \begin{bmatrix} e_0 \\ e_1 \\ e_2 \end{bmatrix}$$

Κανονικές εξισώσεις:

$$\mathbf{b}_{OLS} = (\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{X}^{'}\mathbf{Y}$$

όπου $\mathbf{X}^{'} = \begin{bmatrix} 1.....1......1.........1 \\ 4......7......9........12 \\ 1.......2......5.........8 \\ \\ \end{bmatrix}$

Άρα έχουμε:

$$\mathrm{b}_{OLS} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \\ \end{bmatrix} = [\begin{bmatrix} 1.....1......1.........1 \\ 4......7......9........12 \\ 1.......2......5.........8 \\ \\ \end{bmatrix} \begin{bmatrix} 1.....4......1 \\ 1......7......2 \\ 1.......9......5 \\ 1......12.....8 \\ \end{bmatrix}]^{-1} \begin{bmatrix} 1.....1......1.........1 \\ 4......7......9........12 \\ 1.......2......5.........8 \\ \\ \end{bmatrix} \begin{bmatrix} 7 \\ 12 \\ 17 \\ 20 \\ \end{bmatrix}$$

Όπου $\mathbf{X}^{'}X = \begin{bmatrix} 1.....1......1.........1 \\ 4......7......9........12 \\ 1.......2......5.........8 \\ \\ \end{bmatrix} \begin{bmatrix} 1.....4......1 \\ 1......7......2 \\ 1.......9......5 \\ 1......12.....8 \\ \end{bmatrix} = \begin{bmatrix} 4.......32......16 \\ 32.....290.....159 \\ 16.......159......94 \\ \\ \end{bmatrix}$ (*)

με $\left|X^{'}X\right| = 236$

(*) Ο πίνακας $\mathrm{X}^{'}X$ μπορεί να υπολογισθεί και από την σχέση:

$$\mathbf{X}'X = \begin{bmatrix} T \ldots\ldots\ldots\ldots \sum X_1 \ldots\ldots\ldots\ldots\ldots \sum X_k \\ \sum X_1 \ldots\ldots\ldots \sum X_1^2 \ldots\ldots\ldots\ldots\ldots \sum X_k X_1 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \sum X_k \ldots \sum X_1 X_k \ldots\ldots\ldots\ldots\ldots\ldots\ldots \sum X_k^2 \end{bmatrix} = \begin{bmatrix} 4 \ldots\ldots 32 \ldots\ldots 16 \\ 32 \ldots\ldots 290 \ldots\ldots 159 \\ 16 \ldots\ldots 159 \ldots\ldots 94 \\ \\ \end{bmatrix}$$

Όπου T = 4, $\Sigma X_1 = 32$, $\Sigma X_2 = 16$, $\Sigma X_1^2 = 290$, $\Sigma X_2^2 = 94$, $\Sigma X_1 X_2 = 159$

$$(\mathbf{X}'\mathbf{X})^{-1} = [\begin{bmatrix} 4 \ldots\ldots 32 \ldots\ldots 16 \\ 32 \ldots\ldots 290 \ldots\ldots 159 \\ 16 \ldots\ldots 159 \ldots\ldots 94 \\ \\ \end{bmatrix}]^{-1} = \frac{1}{236} \begin{bmatrix} 1979 \ldots\ldots -464 \ldots\ldots\ldots 448 \\ -464 \ldots\ldots\ldots 120 \ldots\ldots -124 \\ 448 \ldots\ldots -124 \ldots\ldots\ldots\ldots 136 \\ \\ \end{bmatrix}$$

και

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1 \ldots 1 \ldots 1 \ldots\ldots 1 \\ 4 \ldots\ldots 7 \ldots\ldots 9 \ldots\ldots 12 \\ 1 \ldots\ldots 2 \ldots\ldots 5 \ldots\ldots 8 \\ \\ \end{bmatrix} \begin{bmatrix} 7 \\ 12 \\ 17 \\ 20 \\ \end{bmatrix} = \begin{bmatrix} 56 \\ 505 \\ 276 \\ \\ \end{bmatrix} (**)$$

(**) Ο πίνακας X΄Y μπορεί να προκύψει και από την σχέση:

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} \sum Y \\ \sum X_1 Y \\ \ldots\ldots \\ \ldots\ldots\ldots \\ \sum X_k Y \end{bmatrix} = \begin{bmatrix} 56 \\ 505 \\ 276 \end{bmatrix}$$

Όπου $\Sigma Y = 56$, $\Sigma X_1 Y = 505$, $\Sigma X_2 Y = 276$

Τελικώς προκύπτει:

$b_0 = 0,644$

$b_1 = 1,661$

$b_2 = 0,017$

Η εκτιμώμενη συνάρτηση παλινδρόμησης είναι:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 = 0,644 + 1,661 X_1 + 0,017 X_2$$

Εκτίμηση της διακύμανσης του διαταρακτικού όρου

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{T} \hat{e}_i^2}{T-k} = \frac{Y'Y - b_{OLS}{}' X' Y}{T-k}$$

Όπου κ = 3 ο αριθμός των προς εκτίμηση συντελεστών

Έχουμε

$$Y'Y = \begin{bmatrix} 7 & 12 & 17 & 20 \end{bmatrix} \begin{bmatrix} 7 \\ 12 \\ 17 \\ 20 \end{bmatrix} = 882$$

$$b'_{OLS}X'Y = [0{,}644 \quad 1{,}661 \quad 0{,}017]\begin{bmatrix} 56 \\ 505 \\ 276 \\ \\ \end{bmatrix} = 879{,}56$$

Συνεπώς έχουμε:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{T} \hat{e}_i^2}{T-k} = \frac{Y'Y - b_{OLS}'X'Y}{T-k} = \frac{882 - 879{,}56}{4-3} = 2{,}44$$

Εκτιμητής διακυμάνσεων – συνδιακυμάνσεων των $\mathbf{b}_i$

Var – Cov ($\mathbf{b}_{OLS}$) = $\hat{\sigma}^2$ $(\mathbf{X}'\mathbf{X})^{-1}$ =

$$= \frac{2{,}44}{236}\begin{bmatrix} 1979.......-464.........448 \\ -464..........120......-124 \\ 448.......-124...........136 \\ \\ \end{bmatrix} = \begin{bmatrix} 20{,}44.......-4{,}80.........4{,}63 \\ -4{,}80..........1{,}24......-1{,}28 \\ 4{,}63.........-1{,}28...........1{,}41 \\ \\ \end{bmatrix}$$

Άρα έχουμε:

$\mathrm{Var}(b_0) = 20{,}44$, (τυπικό σφάλμα) $\mathrm{s.e}(b_0) = \sqrt{20{,}44} = 4{,}52$

$\mathrm{Cov}(b_0, b_1) = -4{,}80$

$\mathrm{Var}(b_1) = 1{,}24$, (τυπικό σφάλμα) $\mathrm{s.e}(b_1) = \sqrt{1{,}24} = 1{,}10$

$\mathrm{Cov}(b_0, b_2) = -4{,}63$

$\mathrm{Var}(b_2) = 1{,}41$, (τυπικό σφάλμα) $\mathrm{s.e}(b_2) = \sqrt{1{,}41} = 1{,}18$

$\text{Cov}(b_1, b_2) = -1{,}28$

Συντελεστής πολλαπλού προσδιορισμού

$$\mathbf{R}^2_{y.x_1x_2} = \frac{\text{ΑΤΠ}}{\text{ΣΑΤ}} = \frac{b'_{OLS}X'Y - T\bar{Y}^2}{Y'Y - T\bar{Y}^2} = \frac{897{,}56 - 4*(14)^2}{882 - 4*(14)^2} = 0{,}975$$

Αξιολόγηση των ατομικών συντελεστών του υποδείγματος

Προσδιορίζω τον λόγο **t-ratio**(b_i) = $\left|\frac{b_i}{s.e.(\hat{b_i})}\right|$ για κάθε συντελεστή του υποδείγματος.

Έχουμε:

t-ratio(b_0) = $\left|\frac{0{,}644}{4{,}52}\right| = 0{,}142$

t-ratio(b_1) = $\left|\frac{1{,}661}{1{,}10}\right| = 0{,}151$

t-ratio(b_2) = $\left|\frac{0{,}017}{1{,}18}\right| = 0{,}014$

Η μηδενική υπόθεση είναι:

$$H_0: \beta_i = 0 \ \text{ με}$$
$$H_0: \beta_i \neq 0$$

Επειδή $t_{a/2;T-k} = t_{0{,}025;1} = 12{,}706 >$ t-ratio(b_i) η μηδενική υπόθεση H_0

γίνεται **δεκτή** που σημαίνει ότι **μεμονωμένα οι ανεξάρτητες – ερμηνευτικές μεταβλητές $\mathbf{X}_i$ δεν επηρεάζουν σημαντικά την εξαρτημένη μεταβλητή.**

Διάστημα εμπιστοσύνης για τον συντελεστή $\boldsymbol{\beta}_i$

$$\mathbf{b}_i - \text{s.e}(\text{b}_i) * \text{t}_{a/2;T-k} \leq \boldsymbol{\beta}_i \leq \mathbf{b}_i + \text{s.e}(\text{b}_i) * \text{t}_{a/2;T-k}$$

Για τον συντελεστή b_2 έχουμε:

$0{,}017 – 1{,}18 * 12{,}706 \leq \boldsymbol{\beta}_2 \leq 0{,}017 + 1{,}18 * 12{,}706$

$$\mathbf{-\ 14{,}97 \leq \boldsymbol{\beta}_2 \leq 15}$$

Αξιολόγηση του συνόλου του υποδείγματος

Μεταβλητότητα της εξαρτημένης μεταβλητής

Πηγή μεταβλητότητας	Αθροίσματα Τετραγώνων	Βαθμοί ελευθερίας	Μέσοι	F
Παλινδρόμηση X	ΑΤΠ = $b'_{OLS} X'Y - T\bar{Y}^2$ = $R^2_{y.x_1x_2...x_k}(Y'Y - T\bar{Y}^2)$ = 95,561	κ-1 = 3-1= 2	ΑΤΠ/(κ-1) = 95,561/2 = 47,781	**F**
Κατάλοιπα e	ΑΤΚ = $Y'Y - b'_{OLS} X'Y$ = $(1- R^2_{y.x_1x_2...x_k})(Y'Y - T\bar{Y}^2)$ = 2,44	Τ-κ = 4 – 3 =1	ΑΤΚ/(Τ-κ) = 2,44/1 = 2,44	
Σύνολο	ΣΑΤ = $Y'Y - T\bar{Y}^2$ = 98	Ν-1 = 4 – 1 =3	ΣΑΤ/(Τ-1) =98/3 =32,667	

(Όπου κ ο αριθμός των προς εκτίμηση παραμέτρων - συντελεστών)

$$F = \frac{\frac{\text{ΑΤΠ}}{\kappa-1}}{\frac{\text{ΑΤΚ}}{\text{N}-\kappa}} = \frac{47{,}781}{2{,}44} = 19{,}61$$

Με την F – Στατιστική ελέγχεται η μηδενική υπόθεση:

$$H_0: \beta_1 = \beta_2 = \ldots\ldots = \beta_k = 0$$

$$H_1: \text{Ένα εκ των } \beta_i \neq 0$$

σε επίπεδο σημαντικότητας α = 5%

Αν F – Στατιστική < $F_{(\alpha,\kappa-1,T-\kappa)}$ η H_0 γίνεται δεκτή

Αν F – Στατιστική > $F_{(\alpha,\kappa-1,T-\kappa)}$ η H_0 γίνεται απορρίπτεται

Επειδή

$$\mathbf{F} = \frac{\frac{\text{ΑΤΠ}}{\kappa-1}}{\frac{\text{ΑΤΚ}}{T-\kappa}} = \frac{47{,}781}{2{,}44} = 19{,}61 < F_{0,05;2;1} = 200$$

η H_0 γίνεται **δεκτή** που σημαίνει ότι **συνολικά οι ανεξάρτητες μεταβλητές $\mathbf{X}_i$ δεν επηρεάζουν στατιστικά σημαντικά την εξαρτημένη μεταβλητή**.

Στην περίπτωση του τριμεταβλητού υποδείγματος

$$\mathbf{Y}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{X}_{1i} + \boldsymbol{\beta}_2 \mathbf{X}_{2i} + \boldsymbol{\varepsilon}_i$$

Μπορούμε να χρησιμοποιήσουμε και την μέθοδο επίλυσης χωρίς την βοήθεια των πινάκων.

Χρησιμοποιούμε τις κανονικές εξισώσεις που προκύπτουν από την θεωρία των ελαχίστων τετραγώνων με την βοήθεια των τιμών x_1, x_2, και y (των αποκλίσεων από τους μέσους όρους των τιμών των μεταβλητών X_1 , X_2 και Y).

Από τις εξισώσεις αυτές προκύπτουν οι εκτιμήσεις b_0, b_1 , b_2 των β_0 , β_1 , β_2 .

$$b_1 = \frac{(\sum_i^T x_1 y)(\sum_i^T x_2^2) - (\sum_i^T x_2 y)(\sum_i^T x_1 x_2)}{(\sum_i^T x_1^2)(\sum_i^T x_2^2) - (\sum_i^T x_1 x_2)^2}$$

$$b_2 = \frac{(\sum_i^T x_1^2)(\sum_i^T x_2 y) - (\sum_i^T x_1 y)(\sum_i^T x_1 x_2)}{(\sum_i^T x_1^2)(\sum_i^T x_2^2) - (\sum_i^T x_1 x_2)^2}$$

$$b_0 = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2$$

Ο Συντελεστής πολλαπλού προσδιορισμού $\mathbf{R}^2_{y.x_1x_2}$

Ο συντελεστής $\mathbf{R}^2_{y.x_1x_2}$ μετράει το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής που ερμηνεύεται από τις ανεξάρτητες μεταβλητές και δίνεται από τον τύπο:

$$\mathbf{R}^2_{y.x_1x_2} = \frac{\text{ΑΤΠ}}{\text{ΣΑΤ}} = \frac{b_1\sum_i^T x_1 y + b_2 \sum_i^T x_2 y}{\sum_i^T y^2}$$

Ο Διορθωμένος Συντελεστής πολλαπλού προσδιορισμού $\bar{R}^2_{y.x_1x_2}$

$$\bar{R}^2_{y.x_1x_2} = 1 - \frac{\sum_i^T e_i^2/(T-k)}{\sum_i^T y_i^2/(T-1)} = 1 - (1 - \mathrm{R}^2_{y.x_1x_2})\frac{T-1}{T-k}$$

Υπολογίζεται για απαλοιφθεί ο επηρεασμός του μεγέθους του δείγματος και του αριθμού των ανεξάρτητων μεταβλητών επί του συντελεστή πολλαπλού προσδιορισμού $\mathrm{R}^2_{y.x_1x_2}$.

Απλοί και Μερικοί συντελεστές Προσδιορισμού

Α. Απλοί συντελεστές Προσδιορισμού

Αν παλινδρομήσουμε το Y επί του X_1 τότε υπολογίζουμε τον απλό συντελεστή προσδιορισμού

$$\mathrm{r}^2_{yx_1} = \frac{(\sum_i^4 yx_1)^2}{(\sum_i^4 x_1^2)(\sum_i^4 y^2)}$$

Όμοια υπολογίζονται οι απλοί συντελεστές Προσδιορισμού

$$\mathrm{r}^2_{yx_2} = \frac{(\sum_i^4 yx_2)^2}{(\sum_i^4 x_2^2)(\sum_i^4 y^2)}$$ Αν παλινδρομήσουμε το Y επί του X 2 και

$$\mathrm{r}^2_{x_1x_2} = \frac{(\sum_i^4 x_1x_2)^2}{(\sum_i^4 x_1^2)(\sum_i^4 x_2^2)}$$ Αν παλινδρομήσουμε το X_1 επί του X 2

B. Μερικοί συντελεστές Προσδιορισμού

Με τον μερικό συντελεστή προσδιορισμού υπολογίζουμε το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής Y, που εξαρτάται από την ανεξάρτητη μεταβλητή X_i, αν αφαιρέσουμε την επίδραση των άλλων μεταβλητών επί της Y στο υπόδειγμα παλινδρόμησης.

Έτσι μπορούμε να ορίσουμε:

1.Τον μερικό συντελεστή προσδιορισμού που υπολογίζει το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής Y, που εξαρτάται από την ανεξάρτητη μεταβλητή X_1, με την υπόθεση ότι η X_2 παραμένει σταθερή.

$$\mathrm{r}^2_{yx_1.x_2} = \frac{b_1^2 \sum_{i=1}^4 x_1^2}{\sum_{i=1}^4 y^2}$$

2. Τον μερικό συντελεστή προσδιορισμού που υπολογίζει το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής Υ, που εξαρτάται από την ανεξάρτητη μεταβλητή X_2, με την υπόθεση ότι η X_1 παραμένει σταθερή.

$$r^2_{yx_2.x_1} = \frac{b_2^2 \sum_{i=1}^{4} x_2^2}{\sum_{i=1}^{4} y^2}$$

3. Τον μερικό συντελεστή προσδιορισμού που υπολογίζει το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής X_2, που εξαρτάται από την ανεξάρτητη μεταβλητή X_1, με την υπόθεση ότι η Y παραμένει σταθερή.

$$r^2_{x_2 x_1 . y} = \frac{\sum_{i=1}^{4} (x_1 x_2)^2}{(\sum_{i=1}^{4} x_1^2)(\sum_{i=1}^{4} x_2^2)}$$

Εκτίμηση της διακύμανσης του διαταρακτικού όρου

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

Όπου κ ο αριθμός των προς εκτίμηση συντελεστών

Οι διακυμάνσεις των b_1 , b_2

$$Var(b_1) = \hat{\sigma}^2 \left[\frac{1}{(1 - r^2_{x_1 x_2}) \sum_{i=1}^{T} x_1^2} \right]$$, τυπικό σφάλμα $(b_1) = \sqrt{b_1}$

$$Var(b_2) = \frac{\sigma^2}{(1 - r^2_{x_1 x_2})\sum_i^T x_2^2} \;, \quad \text{τυπικό σφάλμα } (b_2) = \sqrt{b_2}$$

- ✓ Η αξιολόγηση των συντελεστών όπως και ολόκληρου του υποδείγματος γίνεται όπως και στο απλό μονομεταβλητό υπόδειγμα.
- ✓ Το F στατιστικό δίνεται από την σχέση:

$$\text{F} = \frac{(\text{ΑΤΠ})/(\kappa - 1)}{(\text{ΑΤΚ})/(T - \kappa)} = \frac{R^2_{y.x_1 x_2} \;/(k-1)}{(1 - R^2_{y.x_1 x_2}\;)/(T-k)}$$

Παράδειγμα:

Δίνονται τα δεδομένα:

Έτος	**Y**	$\mathbf{X}_1$	$\mathbf{X}_2$
2003	7	3	1
2004	11	5	2
2005	16	6	4
2006	18	10	5

Να προσαρμοσθεί το μοντέλο $\mathbf{Y}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{X}_{1i} + \boldsymbol{\beta}_2 \mathbf{X}_{2i} + \boldsymbol{\varepsilon}_i$

Υπολογίζουμε τα αθροίσματα:

Από τις εξισώσεις αυτές προκύπτουν οι εκτιμήσεις b_0, b_1 , b_2 των β_0 , β_1 , β_2.

$$b_1 = \frac{(\sum_i^4 x_1 y)(\sum_i^4 x_2^2) - (\sum_i^4 x_2 y)(\sum_i^4 x_1 x_2)}{(\sum_i^4 x_1^2)(\sum_i^4 x_2^2) - (\sum_i^4 x_1 x_2)^2} = \frac{40*11 - 27*15}{26*11 - (15)^2} = 0,57$$

$$b_2 = \frac{(\sum_i^T x_1^2)(\sum_i^T x_2 y) - (\sum_i^T x_1 y)(\sum_i^T x_1 x_2)}{(\sum_i^T x_1^2)(\sum_i^T x_2^2) - (\sum_i^T x_1 x_2)^2} = \frac{26*27 - 40*15}{26*11 - (15)^2} = 1,67$$

$$b_0 = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 = 13 - 0,57*6 - 1,67*3 = 4,57$$

όπου $\sum_{i=1}^4 x_1 y = 40$, $\sum_{i=1}^4 x_2 y = 27$, $\sum_{i=1}^4 x_2 x_1 = 15$, $\sum_{i=1}^4 x_1^2 = 26$

$\sum_{i=1}^4 x_2^2 = 11$, $\sum_{i=1}^4 y^2 = 74$

Άρα το υπόδειγμα είναι: $\hat{Y}_i = 4,57 + 0,57X_{1i} + 1,67X_{2i}$

Ο Συντελεστής πολλαπλού προσδιορισμού $\mathbf{R}^2_{y.x_1x_2}$

Ο συντελεστής $\mathbf{R}^2_{y.x_1x_2}$ μετράει το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής που ερμηνεύεται από τις ανεξάρτητες μεταβλητές και δίνεται από τον τύπο:

$$\mathbf{R}^2_{y.x_1x_2} = \frac{\text{ΑΤΠ}}{\text{ΣΑΤ}} = \frac{b_1 \sum_i^T x_1 y + b_2 \sum_i^T x_2 y}{\sum_i^T y^2} = \frac{0,57*40 + 1,67*27}{74} = 0,93$$

Ο Διορθωμένος Συντελεστής πολλαπλού προσδιορισμού

$$\bar{R}^2_{y.x_1x_2} = 1 - \frac{\sum_i^T \hat{e}_i^2/(T-k)}{\sum_i^T y_i^2/(T-1)} = 1 - (1 - \mathrm{R}^2_{y.x_1x_2})\frac{T-1}{T-k} = 1 - (1\text{-}0{,}93)\frac{4-1}{4-3} = 0{,}79$$

Υπολογίζεται για απαλοιφθεί ο επηρεασμός του μεγέθους του δείγματος και του αριθμού των ανεξάρτητων μεταβλητών επί του συντελεστή πολλαπλού προσδιορισμού $\mathrm{R}^2_{y.x_1x_2}$.

Απλοί και Μερικοί συντελεστές Προσδιορισμού

Α. Απλοί συντελεστές Προσδιορισμού

Αν παλινδρομήσουμε το Y επί του X_1 τότε υπολογίζουμε τον απλό συντελεστή προσδιορισμού $\mathrm{r}^2_{yx_1}$:

$$\mathrm{r}^2_{yx_1} = \frac{(\sum_i^4 yx_1)^2}{(\sum_i^4 x_1^2)(\sum_i^4 y^2)} = \frac{(40)^2}{26*74} = 0{,}83\ , \quad \mathrm{r}_{yx_1} = 0{,}91$$

Όμοια υπολογίζονται οι απλοί συντελεστές Προσδιορισμού

$$\mathrm{r}^2_{yx_2} = \frac{(\sum_i^4 yx_2)^2}{(\sum_i^4 x_2^2)(\sum_i^4 y^2)} = \frac{27^2}{11*74} = 0{,}90\ , \quad \mathrm{r}_{yx_2} = 0{,}95$$

Αν παλινδρομήσουμε το Y επί του X_2 και

$$r^2_{x_1x_2} = \frac{(\sum_i^4 x_1x_2)^2}{(\sum_i^4 x_1^2)(\sum_i^4 x_2^2)} = \frac{15^2}{26*11} = 0,79 \text{ , } r_{x_1x_2} = 0,89$$

Αν παλινδρομήσουμε το X_1 επί του X_2

B. Μερικοί συντελεστές Προσδιορισμού

1. Μερικός συντελεστής προσδιορισμού που υπολογίζει το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής Y, που εξαρτάται από την ανεξάρτητη μεταβλητή X_1, με την υπόθεση ότι η X_2 παραμένει σταθερή.

$$r^2_{yx_1.x_2} = \frac{b_1^2 \sum_{i=1}^4 x_1^2}{\sum_{i=1}^4 y^2} = \frac{0,57^2 * 26}{74} = 0,11$$

2. Μερικός συντελεστής προσδιορισμούς που υπολογίζει το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής Y, που εξαρτάται από την ανεξάρτητη μεταβλητή X_2, με την υπόθεση ότι η X_1 παραμένει σταθερή.

$$r^2_{yx_2.x_1} = \frac{b_2^2 \sum_{i=1}^4 x_2^2}{\sum_{i=1}^4 y^2} = \frac{1,67^2 * 11}{74} = 0,41$$

3. Μερικός συντελεστής προσδιορισμού που υπολογίζει το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής X_2, που εξαρτάται από την ανεξάρτητη μεταβλητή X_1, με την υπόθεση ότι η Y παραμένει σταθερή.

$$r^2_{x_2\,x_1.y} = \frac{\sum_{i=1}^{4}(x_1x_2)^2}{(\sum_{i=1}^{4}x_1^2)(\sum_{i=1}^{4}x_2^2)} = \frac{15^2}{26*11} = 0,79$$

Εκτίμηση της διακύμανσης του διαταρακτικού όρου

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{T}\hat{e}_i^2}{T-k} = \frac{3,11}{4-3} = 3,11$$

Οι διακυμάνσεις των b_1 , b_2

$$Var(b_1) = \hat{\sigma}^2\left[\frac{1}{(1-r^2_{x_1x_2})\sum_{i=1}^{4}x_1^2}\right] = 3,11*\frac{1}{(1-0,79)*26} = 0,57$$

s.e (b_1) = $\sqrt{b_1}$ = $\sqrt{0,57}$ = 0,75

$$Var(b_2) = \frac{\hat{\sigma}^2}{(1-r^2_{x_1x_2})\sum_{i}^{4}x_2^2} = \frac{3,11}{(1-0,79)*11} = 1,35$$

s.e (b_2) == $\sqrt{b_2}$ = $\sqrt{1,35}$ = 1,16

Αξιολόγηση των συντελεστών και ολοκλήρου του υποδείγματος

1. Αξιολόγηση των συντελεστών του υποδείγματος

Προσδιορίζω τον λόγο **t-ratio**(b_i) **=** $\left|\dfrac{b_i}{s.e.(\hat{b_i})}\right|$ για κάθε συντελεστή του υποδείγματος.

Έχουμε:

$$\textbf{t-ratio}(b_1) = \left|\frac{0,57}{0,75}\right| = 0,76$$

$$\textbf{t-ratio}(b_2) = \left|\frac{1,67}{1,16}\right| = 1,44$$

Η μηδενική υπόθεση είναι:

$$H_0: \beta_i = 0 \ \text{ με}$$

$$H_0: \beta_i \neq 0$$

Επειδή $t_{a/2;T-k} = t_{0,025;1} = 12,706 > \text{t-ratio}(b_i)$ η μηδενική υπόθεση H_0 γίνεται **δεκτή** που σημαίνει ότι **μεμονωμένα οι ανεξάρτητες – ερμηνευτικές μεταβλητές X_i δεν επηρεάζουν σημαντικά την εξαρτημένη μεταβλητή.**

Διάστημα εμπιστοσύνης για τους συντελεστές $\boldsymbol{\beta}_i$

$$\mathbf{b}_i - \text{s.e}(\text{b}_i) * \text{t}_{a/2;T-k} \leq \boldsymbol{\beta}_i \leq \mathbf{b}_i + \text{s.e}(\text{b}_i) * \text{t}_{a/2;T-k}$$

Για τον συντελεστή β_1 έχουμε:

$0,57 - 0,75 * 12,706 \leq \boldsymbol{\beta}_1 \leq 0,57 + 0,75 * 12,706$

$$\mathbf{-8,96} \leq \boldsymbol{\beta}_1 \leq \mathbf{10,10}$$

Για τον συντελεστή β_2 έχουμε:

$1,67 - 1,16 * 12,706 \leq \boldsymbol{\beta}_2 \leq 1,67 + 1,16 * 12,706$

$$\mathbf{-13,07} \leq \boldsymbol{\beta}_2 \leq \mathbf{16,41}$$

2. Αξιολόγηση ολοκλήρου του υποδείγματος

Η σημαντικότητα του υποδείγματος μπορεί να ελεγχθεί με την τιμή F – Στατιστικής η οποία δίνεται από την σχέση:

$$\text{F} = \frac{(\text{ΑΤΠ})/(\kappa-1)}{(\text{ΑΤΚ})/(T-\kappa)} = \frac{R^2_{y.x_1x_2}\ /(k-1)}{(1-R^2_{y.x_1x_2}\ )/(T-k)} = \frac{0,93/(3-1)}{(1-0,93\ )/(4-3)} = 6,64$$

Με την F – Στατιστική ελέγχεται η μηδενική υπόθεση:

$$\text{H}_0: \beta_1 = \beta_2 = \ldots\ldots = \beta_k = 0$$
$$\text{H}_1: \text{Ένα εκ των } \beta_i \neq 0$$

σε επίπεδο σημαντικότητας α = 5%

Αν F – Στατιστική < $F_{(\alpha,\kappa-1,T-\kappa)}$ η H_0 γίνεται δεκτή

Αν F – Στατιστική > $F_{(\alpha,\kappa-1,T-\kappa)}$ η H_0 γίνεται απορρίπτεται

Επειδή από τους πίνακες για α=5% με βαθμούς ελευθερίας 2, 1 αριθμητή και παρανομαστή έχουμε $F_{5\%;2,1}$ = 199 > F = 6,64

- Η H_0 γίνεται **δεκτή** που σημαίνει ότι **συνολικά οι ανεξάρτητες μεταβλητές X_i δεν επηρεάζουν στατιστικά σημαντικά την εξαρτημένη μεταβλητή**.

Έλεγχος Γραμμικών περιορισμών των μερικών συντελεστών παλινδρόμησης

A) Έστω η λογαριθμική μορφή της συνάρτησης ζήτησης:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + e_t \quad (1)$$

Y_t : Η ζητούμενη ποσότητα
X_2 : Η τιμή
X_3 : Το εισόδημα
X_4 : Ο πλούτος

Όλες οι μεταβλητές είναι σε λογαριθμική μορφή

Ο συντελεστής $\beta_2 = \dfrac{\partial Y_t}{\partial X_{2t}}$ εκφράζει τη μεταβολή της ζητούμενης ποσότητας σε μία μεταβολή της τιμής. Όμως επειδή οι δύο μεταβλητές είναι εκφρασμένες σε λογαριθμική μορφή ο αντίστοιχος συντελεστής εκφράζει ελαστικότητες. Πιο συγκεκριμένα:

$$\beta_2 = \frac{\partial \log(Y_t)}{\partial \log(X_{2t})} = \frac{\frac{\partial Y_t}{Y_t}}{\frac{\partial X_{2t}}{X_{2t}}} = \varepsilon_{Y_t, X_{2t}} \quad (2)$$

Συνεπώς όλοι οι προς εκτίμηση παράμετροι είναι οι αντίστοιχες ελαστικότητες.

Έστω ότι θέλουμε να ελέγξουμε αν οι ελαστικότητες του εισοδήματος και του πλούτου είναι ίσες.

Για την πραγματοποίηση του ελέγχου αυτού ορίζουμε **την μηδενική και την εναλλακτική υπόθεση:**

$$H_o : \beta_3 = \beta_4$$
$$H_1 : \beta_3 \neq \beta_4$$

Όταν ισχύει η μηδενική υπόθεση τότε η στατιστική t που δίνεται από την σχέση:

$$t = \frac{(b_3 - b_4) - (\beta_3 - \beta_4)}{\hat{s.e}.(b_3 - b_4)} = \frac{(b_3 - b_4) - 0}{\hat{s.e}.(b_3 - b_4)} \sim t_{T-4} \quad (3)$$

- ✓ Ο αριθμός 4 είναι το πλήθος των προς εκτίμηση συντελεστών.

Τ: το μέγεθος του δείγματος

k=4

Εάν $|t| > t_{T-4}$ τότε η μηδενική υπόθεση απορρίπτεται.

Ο τυχαίος όρος σφάλματος της διαφοράς των δύο μεταβλητών δίνεται από την σχέση:

$$s.e(b_3 - b_4) = \sqrt{\hat{Var}(b_3 - b_4)} = \sqrt{\hat{Var}(b_3)} + \sqrt{\hat{Var}(b_4)} - \sqrt{2\hat{\text{cov}}(b_{3,} b_4)} \qquad (4)$$

Παράδειγμα:

Έστω το υπόδειγμα:

$$A_t = \beta_1 + \beta_2 Y_t + \beta_3 P_t + \beta_4 W_t + e_t \quad (5)$$

Y_t Το εισόδημα

P_t Το επίπεδο των τιμών

W_t Ο πλούτος των ατόμων

A_t Η ζήτηση του προϊόντος

Χρησιμοποιούνται ετήσια στοιχεία για την περίοδο 1962 – 2002.
Να εξεταστεί η ισότητα των συντελεστών του εισοδήματος και του πλούτου.
Να ερμηνευθούν οι συντελεστές του εκτιμημένου υποδείγματος καθώς και ο συντελεστής προσδιορισμού. Να εξεταστεί η στατιστική σημαντικότητα της παλινδρόμησης.

Ο αριθμός των παρατηρήσεων ανέρχεται σε T = 44.
Η παλινδρόμηση μετά την εκτίμησή της δίνεται από την σχέση:

$$\hat{A}_t = 1{,}90 + 0{,}90Y_t - 0{,}80P_t + 0{,}95W_t$$
$$(\hat{s.e})\ (0{,}90)\ (0{,}15)\ (0{,}70)\quad (0{,}30)$$

Δίνεται ότι $Cov(b_2, b_4) = 0{,}03$, $R^2 = 0{,}76$.

Με εφαρμογή του ελέγχου που περιγράφηκε παραπάνω αναλυτικά προκύπτουν τα εξής:

$$H_o : \beta_2 = \beta_4$$
$$H_1 : \beta_2 \neq \beta_4$$

Όταν ισχύει η μηδενική υπόθεση τότε η στατιστική t που δίνεται από την σχέση:

$$t = \frac{(b_2 - b_4) - (\beta_2 - \beta_4)}{\hat{s.e.}(b_2 - b_4)} = \frac{(b_2 - b_4) - 0}{\hat{s.e.}(b_2 - b_4)} = \frac{0,90 - 0,95}{\sqrt{\hat{Var}(b_2) + \hat{Var}(b_4) - 2\hat{\mathrm{cov}}(b_2, b_4)}} = \frac{0.90 - 0.95}{0.225} = -0.218$$

Συγκρίνουμε την απόλυτη τιμή με την αντίστοιχη κριτική τιμή των πινάκων η οποία για β.ε. = T-k = 44-4 = 40 και επίπεδο σημαντικότητας α=5% είναι $t_c = 2{,}021$ (κ=4, 0 αριθμός των προς εκτίμηση συντελεστών)

Επειδή $|t| < t_c$ αποδεχόμαστε τη μηδενική υπόθεση.

Ερμηνεία του συντελεστή προσδιορισμού:

Το 76% της συνολικής μεταβλητότητας της ζητούμενης ποσότητας ερμηνεύεται από μεταβολές στο εισόδημα, τον πλούτο και την τιμή ενώ το 24% ερμηνεύεται από τον τυχαίο όρο σφάλματος.

Ερμηνεία του σταθερού όρου $b_1 = 1{,}90$:

εκφράζει την ζητούμενη ποσότητα όταν το σύνολο των μεταβλητών είναι μηδενικές.

Ερμηνεία του συντελεστή του εισοδήματος $b_2 = 0{,}90$:

εκφράζει τη μεταβολή της ζητούμενης ποσότητας όταν το εισόδημα μεταβληθεί κατά μία μονάδα

Ερμηνεία του συντελεστή της τιμής $b_3 = -0{,}80$:

εκφράζει την μείωση της ζητούμενης ποσότητας όταν η τιμή μεταβληθεί κατά μία μονάδα.

Ερμηνεία του συντελεστή του πλούτου $b_4 = 0{,}95$:

εκφράζει την μείωση της ζητούμενης ποσότητας όταν ο πλούτος μεταβληθεί κατά μία μονάδα.

- Για τον έλεγχο της **στατιστικής σημαντικότητας του συντελεστή προσδιορισμού** χρησιμοποιείται η στατιστική F.

$$H_o : R^2 = 0$$
$$H_1 : R^2 \neq 0$$

Η στατιστική δίνεται από την σχέση:

$$F = \frac{\frac{R^2}{k-1}}{\frac{(1-R^2)}{T-k}} = \frac{\frac{0{,}76}{3}}{\frac{0{,}24}{40}} = 42{,}22$$

$$F_{a;\,k-1,T-k} = F_{0{,}05;3{,}40} = 2{,}84$$

- ✓ **Επειδή η τιμή της στατιστικής υπερβαίνει την κριτική τιμή απορρίπτουμε την μηδενική υπόθεση και αποδεχόμαστε την εναλλακτική σύμφωνα με την οποία η παλινδρόμηση είναι στατιστικά σημαντική.**

B) Έστω η **συνάρτηση παραγωγής Cobb – Douglas**

$$Y_t = \beta_1 X_2^{\beta_2} X_3^{\beta_3} e^{e_t} \qquad (6)$$

Για τον γραμμικό μετασχηματισμό της συνάρτησης απαιτείται η λογαρίθμηση της συνάρτησης

Y_t προϊόν

X_2 Εργασία

X_3 Κεφάλαιο

$$\ln(Y_t) = \ln(\beta_1) + \beta_2 \ln(X_2) + \beta_3 \ln(X_3) + e_t \qquad \textbf{(7)*}$$

*** Το υπόδειγμα (7) είναι χωρίς περιορισμούς**

Θέλουμε να εξετάσουμε αν έχουμε **σταθερές οικονομίες κλίμακας**, δηλαδή αν ισχύει

$$\beta_2 + \beta_3 = 1$$

Για την πραγματοποίηση του δίπλευρου ελέγχου ορίζουμε την μηδενική και εναλλακτική υπόθεση και συνεπώς:

$$H_0 : \beta_2 + \beta_3 = 1$$
$$H_1 : \beta_2 + \beta_3 \neq 1$$

Ο έλεγχος μπορεί να γίνει τόσο με την t – student στατιστική όσο και με την στατιστική F.

- **Ο έλεγχος με την t – student στατιστική**

Αφού εκτιμήσουμε την εξίσωση με τη βοήθεια του εκτιμητή ελαχίστων τετραγώνων τότε εφαρμόζουμε το έλεγχο με την t – student στατιστική.

$$t = \frac{(b_2 + b_3) - (\beta_2 + \beta_3)}{\hat{s.e}.(b_2 + b_3)} = \frac{(b_2 + b_4) - 1}{\hat{s.e}.(b_2 + b_4)} \sim t_{T-k}$$

k=3

Εάν $|t| > t_{t-k}$ απορρίπτουμε την μηδενική υπόθεση για σταθερές αποδόσεις κλίμακας.

- **Η προσέγγιση με τον έλεγχο F**

Εφαρμόζουμε την μέθοδο ελαχίστων τετραγώνων τόσο με περιορισμούς όσο και χωρίς περιορισμούς.

Πιο συγκεκριμένα:

Έστω ότι οι αποδόσεις κλίμακας είναι σταθερές δηλαδή ότι ισχύει

$$\beta_2 + \beta_3 = 1 \Leftrightarrow \beta_2 = 1 - \beta_3 \Leftrightarrow \ln(Y_t) = \ln\beta_1 + (1 - \beta_3)\ln(X_{2t}) + \beta_3 \ln(X_{3t}) + e_t \Leftrightarrow$$
$$\Leftrightarrow \ln(Y_t) - \ln(X_{2t}) = \ln\beta_1 - \left(\beta_3 \ln(X_{2t}) + \beta_3(\ln(X_{3t})\right) + e_t \Leftrightarrow$$
$$\ln\left(\frac{Y_t}{X_{2t}}\right) = \ln(\beta_1) + \beta_3\left(\frac{\ln(X_{3t})}{\ln(X_{2t})}\right) + e_t \quad (8)$$

**** Το υπόδειγμα (8) είναι με περιορισμούς**

Με αυτόν τον τρόπο προκύπτει ένα υπόδειγμα που είναι συνάρτηση μιας και όχι δύο ερμηνευτικών μεταβλητών. Το παραπάνω υπόδειγμα είναι **υπόδειγμα με περιορισμούς.**

Από την εκτίμηση του παραπάνω υποδείγματος (8) δίνονται και το άθροισμα των τετραγώνων των καταλοίπων (RSS_1, **υπό περιορισμό**)

Από την εκτίμηση της (7) προκύπτει το άθροισμα των τετραγώνων των καταλοίπων χωρίς περιορισμό (RSS_2, **χωρίς περιορισμό**).

Η στατιστική F σε αυτήν την περίπτωση δίνεται από την σχέση:

$$F = \frac{(RSS_1 - RSS_2)/\lambda}{RSS_2/(T-k)} \sim F_{\lambda,T-k} \quad (9)$$

*** Όπου λ ο αριθμός των περιορισμών**

Αν το υπόδειγμα ικανοποιεί τους περιορισμούς θα ισχύει $RSS_2 = RSS_1$, άρα η τιμή της στατιστικής θα υπολείπεται της κριτικής τιμής που προκύπτει από τους πίνακες.

Στην αντίθετη περίπτωση $F > F_{\lambda,T-k}$

Η στατιστική F μπορεί να πάρει και την εξής μορφή:

$$F = \frac{(R^2{}_2 - R^2{}_1)/\lambda}{(1-R^2{}_2)/(T-k)} \quad (10)$$

*** Ο δείκτης 1 αναφέρεται στο υπόδειγμα με περιορισμούς και ο δείκτης 2 στο υπόδειγμα χωρίς περιορισμούς**

Η σχέση αυτή μπορεί να προκύψει με δεδομένη την ισότητα των συνολικών αθροισμάτων των τετραγώνων $TSS_2 = TSS_1$. Αυτό ισχύει επειδή τα δύο υποδείγματα έχουν την ίδια ανεξάρτητη μεταβλητή.

Εφαρμογή

Η ζήτηση για κοτόπουλα στις Η.Π.Α. 1960 – 1982 δίνεται από το υπόδειγμα

$$\ln(Y_t) = \ln(\beta_1) + \beta_2 \ln(X_2) + \beta_3 \ln(X_3) + \beta_4 \ln(X_4) + \beta_5 \ln(X_5) + e_t \quad (11)$$

Y_t κατά κεφαλήν κατανάλωση κοτόπουλου σε λίμπρα

X_2 κατά κεφαλήν πραγματικό διαθέσιμο εισόδημα

X_3 τιμή χοιρινού σε λίμπρα

X_4 τιμή βόειου κρέατος σε λίμπρα

X_5 τιμή κοτόπουλου σε λίμπρα

Οι τιμές των συντελεστών ερμηνεύονται ως εξής:

β_2 ελαστικότητα ως προς το διαθέσιμο εισόδημα

β_3 ελαστικότητα ως προς την τιμή του υποκατάστατου

β_4 ελαστικότητα ως προς την τιμή του υποκατάστατου

β_5 ελαστικότητα ως προς την τιμή του κοτόπουλου

Τα αναμενόμενα πρόσημα των συντελεστών είναι θετικά εκτός από αυτό της τιμής. Ο ρόλος της τιμής των υποκατάστατων προϊόντων είναι ασήμαντος όταν ισχύει $\beta_4 = \beta_3 = 0$

Ορίζουμε την μηδενική και την εναλλακτική υπόθεση και έτσι έχουμε:

$$H_0 : \beta_3 = \beta_4 = 0$$
$$H_1 : \beta_3 \neq 0 \quad ή / και \quad \beta_4 \neq 0$$

Εκτιμούμε το υπόδειγμα με περιορισμό και χωρίς περιορισμό και λαμβάνουμε τα δύο ακόλουθα αποτελέσματα:

$$\ln Y_t = 2{,}1898 + 0{,}3425 \ln X_{2t} - 0{,}5046 \ln X_{3t} + 0{,}1495 \ln X_{4t} + 0{,}09 \ln X_{5t}$$
$$R_2^2 = 0{,}9823$$

$$\ln Y_t = 2{,}038 + 0{,}415 \ln X_{2t} + 0{,}372 \ln X_{5t}$$
$$R_1^2 = 0{,}9801$$

Η τιμή του συντελεστή προσδιορισμού υπό περιορισμό είναι μικρότερη σε σχέση με αυτή χωρίς προσδιορισμό αποτέλεσμα που συνδέεται με τον αριθμό των ερμηνευτικών μεταβλητών.

Πόσο όμως αυτή είναι στατιστικά σημαντική, αυτό ελέγχεται από την στατιστική F.

$$F = \frac{(R^2{}_2 - R^2{}_1)/\lambda}{\left(1 - R^2{}_2\right)/\left(T - k\right)} = 2{,}57$$

Η κριτική τιμή που προκύπτει από τον πίνακα λαμβάνεται για λ=2, T=23 και k = 5 έτσι προκύπτει: $F_{a;2,18} = 3{,}55$

Ισχύει F $< F_{a;2,18} = 3{,}55$ άρα η μηδενική υπόθεση δεν μπορεί να απορριφθεί για επίπεδο σημαντικότητας α=5%.

- ✓ **Συνεπώς ο συντελεστής προσδιορισμού δεν είναι στατιστικά σημαντικός**

Παραδείγματα (Πολυμεταβλητού υποδείγματος παλινδρόμησης)

Α. Έστω το οικονομικό Υπόδειγμα:

$$TR = \beta_1 + \beta_2 (P) + \beta_3 (AD)$$

TR: Συνολικό Έσοδο
P: Τιμή
AD: Διαφημιστική Δαπάνη

β_2, β_3 : Συντελεστές μερικής παλινδρόμησης

Για παράδειγμα ο συντελεστής β_2 εξετάζει πόσο θα μεταβληθεί το συνολικό έσοδο αν οι δαπάνες για διαφήμιση μεταβληθούν κατά μία μονάδα.

Στατιστικό Υπόδειγμα:

$$TR = \beta_1 + \beta_2 (P) + \beta_3 (AD) + e_t$$

Ο τυχαίος όρος σφάλματος θα μπορούσε να περιλαμβάνει την τιμή ενός ανταγωνιστικού προϊόντος, τη φορολογική πολιτική της κυβέρνησης κ.τ.λ.

Η γενική μορφή του πολυμεταβλητού υποδείγματος είναι η εξής:

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + e_t$$

Θεωρείται ότι X_{1t}=1 για κάθε παρατήρηση ώστε να είναι δυνατή η γραφή του υποδείγματος με τη γραμμική άλγεβρα.

Με τη μορφή της γραμμικής άλγεβρας το υπόδειγμα γράφεται ως εξής:

$$Y=Xb+e$$

Y : Tx1
X : Tx3
b : 3x1
e: Tx1

Ο αριθμός των γραμμών T δηλώνει τον αριθμό των παρατηρήσεων του δείγματος
Ο αριθμός των στηλών του πίνακα X δηλώνει τις ερμηνευτικές μεταβλητές
Οι υποθέσεις του υποδείγματος είναι ίδιες με αυτές του διμεταβλητού υποδείγματος

Επιπλέον υπόθεση:

Μία ανεξάρτητη μεταβλητή δεν είναι ένας γραμμικός συνδυασμός μιας άλλης ανεξάρτητης μεταβλητής. Το υπόδειγμα δηλαδή δεν υπόκειντα σε πολυσυγγραμικότητα. Δεν είναι δυνατό π.χ. να ισχύει $X_{2t} = c_1 + c_2 X_{1t}$. Αν σχετίζονται οι μεταβλητές δεν είναι δυνατός ο διαχωρισμός της επίδρασης της κάθε μεταβλητής στην εξαρτημένη μεταβλητή.

Η υπόθεση αυτή εκφράζεται και ως εξής:

Ο βαθμός της μήτρας Χ είναι πλήρης. Πιο συγκεκριμένα: Βαθμός της Χ, r(X) = 3. Αυτό σημαίνει πως υπάρχει γραμμική ανεξαρτησία μεταξύ των στηλών.

- ✓ **Ως βαθμός της μήτρας ορίζεται ο μέγιστος αριθμός των γραμμικά ανεξάρτητων στηλών της μήτρας.**

Στο υπόδειγμα

$$TR = \beta_1 + \beta_2 (P) + \beta_3 (AD) + e_t \ ,$$

Μπορεί να δειχθεί ότι: $r(X'X) = r(X)$

Εάν $r(X'X) = 3 \Leftrightarrow |X'X| \neq 0$.

Σε αυτή την περίπτωση θα υπάρχει η

$$(X'X)^{-1} = \frac{adj(X'X)}{|X'X|} \Leftrightarrow b_{OLS} = (X'X)^{-1} X'Y$$

Αν υπάρχει γραμμική σχέση μεταξύ των ερμηνευτικών μεταβλητών τότε δεν υπάρχει η $|X'X|$.

<u>Έστω το εκτιμημένο υπόδειγμα:</u>

$$\hat{TR}_t = 104790 - 6{,}6P_t + 2{,}98(AD_t) \text{ για } T = 43$$

Αυτό σημαίνει πως

$$b_1 = 104790$$
$$b_2 = -6{,}6$$
$$b_3 = 2{,}98$$

Ερμηνεία των αποτελεσμάτων:

Η τιμή $b_2 = -6{,}6$.Το πρόσημο είναι συνεπές με την οικονομική θεωρία.

Η ζήτηση του προϊόντος είναι ελαστική. Το αποτέλεσμα αυτό προκύπτει επειδή μείωση των τιμών αυξάνει τη ζήτηση κατά μεγαλύτερο ποσοστό και συνεπώς έχουμε αύξηση των συνολικών εσόδων.

Δηλαδή εδώ θα συμπεράνουμε:

$b_2 > 0$ τότε η ζήτηση είναι ανελαστική

$b_2 < 0$ τότε η ζήτηση είναι ελαστική

$b_3 > 0$: Η αύξηση της δαπάνης για διαφήμιση συνεπάγεται αύξηση των συνολικών εσόδων.

Πρόβλεψη

Έστω ότι η τιμή του αγαθού είναι € 2 και οι συνολικές διαφημιστικές δαπάνες είναι € 10000 Τότε τα συνολικά έσοδα της επιχείρησης θα είναι τα εξής:

$$\hat{TR}_t = 104790 - 6{,}6 * (2) + 2{,}98 * (10000) = 121340 €$$

Αν βρεθεί ο συντελεστής προσδιορισμού ίσος με 0,58 τότε το 58% της συνολικής μεταβλητότητας των συνολικών εσόδων ερμηνεύεται από την μεταβλητότητα του επιπέδου των τιμών και αυτή της διαφημιστικής δαπάνης, ενώ το 42% ερμηνεύεται από την μεταβλητότητα του τυχαίου όρου σφάλματος.

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum \hat{e}_t^2}{\sum \left(Y_t - \bar{Y}\right)^2}$$

Ο προσαρμοσμένος συντελεστής προσδιορισμού

Έστω ότι εκτιμούμε δύο διαφορετικά υποδείγματα το ένα που περιλαμβάνει μόνο την τιμή και το άλλο τόσο την τιμή όσο και τη διαφημιστική δαπάνη.

Πιο συγκεκριμένα:

(α) $TR = \beta_1 + \beta_2(P) + \beta_3(AD) + e_t$

(β) $TR = \beta_1^* + \beta_2^*(P) + e_t$

Αν εκτιμήσουμε τα δύο υποδείγματα τότε προφανώς θα ισχύει: $\beta_1 \neq \beta_1^*$ καθώς και $\beta_2 \neq \beta_2^*$.

Ποια είναι η διαφορά των συντελεστών προσδιορισμού στα δύο διαφορετικά υποδείγματα. Είναι γνωστό πως **ο συντελεστής προσδιορισμού αυξάνεται όταν ο αριθμός των ερμηνευτικών μεταβλητών αυξάνεται οπότε μειώνεται και ο τυχαίος όρος σφάλματος**:

$$e_t \downarrow \Leftrightarrow \sum e_t^2 \downarrow \Leftrightarrow R^2 \uparrow$$

Άρα το υπόδειγμα με τις περισσότερες ερμηνευτικές μεταβλητές χαρακτηρίζεται από μεγαλύτερο συντελεστή προσδιορισμού.

- Για να συγκρίνουμε δύο υποδείγματα με **διαφορετικό αριθμό ερμηνευτικών μεταβλητών** χρησιμοποιούμε τον **προσαρμοσμένο συντελεστή προσδιορισμού**.

Αυτός δίνεται από την σχέση:

$$\bar{R}^2 = 1 - \frac{\dfrac{\sum \hat{e}_t^2}{T-k}}{\dfrac{\sum (Y_t - Y)^2}{T-1}} \Leftrightarrow \bar{R}^2 = 1 - \frac{\sum \hat{e}_t^2}{\sum \left(Y - \bar{Y}\right)^2} \frac{(T-1)}{(T-k)} \Leftrightarrow \bar{R}^2 = 1 - (1 - R^2)\frac{T-1}{T-k}$$

Το τελευταίο τμήμα δηλώνει την σχέση μεταξύ του συντελεστή προσδιορισμού και το προσαρμοσμένου συντελεστή προσδιορισμού.

Για **διαφορετικούς βαθμούς ελευθερίας** προκύπτουν **διαφορετικές σχέσεις μεταξύ του συντελεστή προσδιορισμού και του προσαρμοσμένου συντελεστή προσδιορισμού.**

Πιο συγκεκριμένα:

Όταν k = 1 $\Leftrightarrow R^2 = \bar{R}^2$

Όταν $R^2 = 1 \Leftrightarrow \bar{R}^2 = R^2 = 1$

Όταν k > 1 $\Leftrightarrow R^2 < \bar{R}^2$

Όταν $k \uparrow$ ώστε k > 1 τότε έχουμε δύο συγκρουόμενες επιδράσεις:

$$\downarrow \sum \hat{e}_t^2 \Leftrightarrow \uparrow R^2 \Leftrightarrow \uparrow \bar{R}^2$$
$$\downarrow (T-k) \Leftrightarrow \downarrow \bar{R}^2$$

Συνεπώς η **μία επίδραση αυξάνει τον προσαρμοσμένο συντελεστή προσδιορισμού και η άλλη επίδραση τον μειώνει.**

Ο **προσαρμοσμένος συντελεστής προσδιορισμού** μπορεί να πάρει και **αρνητικές τιμές** αν η μείωση του προσαρμοσμένου συντελεστή προσδιορισμού υπερτερεί της αύξησης του προσαρμοσμένου συντελεστή προσδιορισμού.

Ο προσαρμοσμένος συντελεστής προσδιορισμού είναι χρήσιμος όταν **συγκρίνουμε δύο υποδείγματα ως προς την ερμηνευτικότητά τους.**

Επιπλέον, η απώλεια των βαθμών ελευθερίας επηρεάζει τη διακύμανση των εκτιμητών και συνεπώς την ακρίβεια των εκτιμήσεων.

Οι διακυμάνσεις του σφάλματος και των εκτιμητών δίνονται από τις ακόλουθη σχέσεις:

$$\hat{\sigma}^2 = \frac{\sum \hat{e}_t^2}{T-k}, \quad Varb_{OLS} = \hat{\sigma}^2 \left(X'X\right)^{-1}$$

Αν έχουμε δύο υποδείγματα που διαφέρουν ως προς την ερμηνευτική μεταβλητή και ως προς την εξαρτημένη μεταβλητή τότε δεν μπορούμε να χρησιμοποιήσουμε ούτε τον συντελεστή προσδιορισμού ούτε τον προσαρμοσμένο συντελεστή προσδιορισμού

- ✓ Αυτό σημαίνει πως **όταν συγκρίνουμε δύο υποδείγματα με βάση τον συντελεστή προσδιορισμού τα υποδείγματα θα πρέπει να έχουν την ίδια εξαρτημένη μεταβλητή.**

Έλεγχος υπόθεσης για τους συντελεστές παλινδρόμησης

Η t - στατιστική που χρησιμοποιείται δίνεται από την σχέση:

$t = \dfrac{b_i}{s.e(\hat{b}_i)} \sim t_{T\text{-}\kappa}$, όταν ισχύει η μηδενική υπόθεση.

- ✓ Γενικά για **ένα πολυμεταβλητό υπόδειγμα με k προς εκτίμηση συντελεστές οι βαθμοί ελευθερίας της t-statistic είναι Τ-κ**

$$\hat{TR}_t = 104790 - \underset{(3{,}191)}{6{,}6}P_t + \underset{(0{,}1669)}{2{,}98}(AD_t)$$

Για τον έλεγχο της στατιστικής σημαντικότητας της τιμής του προϊόντος ακολουθούμε την εξής διαδικασία:

$$H_o : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$

Με αντικατάσταση των τιμών του προηγούμενου υποδείγματος

$$t = \frac{b_2 - \beta_2}{s.e(\hat{b_2})} = \frac{-6{,}642 - 0}{3{,}191} = -2{,}08$$

Για επίπεδο σημαντικότητας α=5% και για βαθμούς ελευθερίας T-3=43-3=40, η κριτική τιμή που προκύπτει είναι t_c=2,021

Επειδή σε απόλυτες τιμές η τιμή της στατιστικής υπερβαίνει της κριτικής τιμής $(|t| > t_{cr})$ έχουμε απόρριψη της μηδενικής υπόθεσης με συνέπεια ο **εκτιμηθείς συντελεστής της τιμής επηρεάζει το συνολικό έσοδο**.

Ελέγχουμε επίσης την μηδενική υπόθεση:

$$H_o : \beta_3 = 0$$
$$H_1 : \beta_3 \neq 0$$

Με αντικατάσταση των τιμών του προηγούμενου υποδείγματος

$$t = \frac{b_3 - \beta_3}{s.e(\hat{b_3})} = \frac{2{,}984 - 0}{0{,}1669} = 17{,}88$$

Για επίπεδο σημαντικότητας α=5% και για βαθμούς ελευθερίας Τ-3=43-3 =40, η κριτική τιμή που προκύπτει είναι t_c=2,021

Επειδή σε απόλυτες τιμές η τιμή της στατιστικής υπερβαίνει της κριτικής τιμής $(|t| > t_{cr})$ έχουμε απόρριψη της μηδενικής υπόθεσης με συνέπεια ο **εκτιμηθείς συντελεστής της διαφημιστικής δαπάνης επηρεάζει το συνολικό έσοδο.**

Έλεγχος υπόθεσης για το σύνολο των συντελεστών μερικής παλινδρόμησης

$$H_o : \beta_2 = \beta_3 = 0$$
$$H_1 : \beta_2 \neq 0 \ \eta \ \beta_3 \neq 0 \ \eta \ \beta_2 \neq 0 \ \kappa\alpha\iota \ \beta_3 \neq 0$$

Ο έλεγχος αυτός είναι ο έλεγχος στατιστικής σημαντικότητας της παλινδρόμησης. Αν **δεν είναι δυνατή η απόρριψη της μηδενικής υπόθεσης τότε η παλινδρόμηση θα έχει τη μορφή:**

$TR = \beta_1 + e_t$ (Μονομεταβλητό υπόδειγμα)

Σε αυτή την περίπτωση η παλινδρόμηση δεν είναι στατιστικά σημαντική

- ✓ Είναι γνωστό ότι:

  TSS = SSR + RSS

  TSS: T-1 βαθμούς ελευθερίας

  SSR: 2 βαθμούς ελευθερίας

  RSS: T-3 βαθμούς ελευθερίας

Προφανώς οι βαθμοί ελευθερίας θα πρέπει να είναι ίσοι και από τις δύο πλευρές.

Εμείς γνωρίζουμε μόνο τους βαθμούς ελευθερίας του TSS και του RSS. Συνεπώς με αυτό τον τρόπο είναι δυνατός ο προσδιορισμός των βαθμών ελευθερίας του SSR.

Για τον συγκεκριμένο έλεγχο χρησιμοποιούμε την στατιστική F.

Για την πραγματοποίηση του συγκεκριμένου ελέγχου υποθέτουμε τα εξής:

Υποθέτουμε $e_t \sim N(0,\sigma^2) \Leftrightarrow b_{OLS} \sim N(\beta_{i,}\sigma_{b_i}{}^2)$

$$H_o : \beta_2 = \beta_3 = 0$$
$$H_1 : \beta_2 \neq 0 \ \eta \ \beta_3 \neq 0 \ \eta \ \beta_2 \neq 0 \ \kappa\alpha\iota \ \beta_3 \neq 0$$

Τότε η μεταβλητή F δίνεται από την σχέση:

$$F = \frac{\frac{SSR}{k-1}}{\frac{RSS}{T-k}} \sim F_{\kappa-1,T-k}$$

Συνεπώς για k = 3 και T = 43 η παραπάνω σχέση διαμορφώνεται ως εξής:

$$F = \frac{\frac{SSR}{2}}{\frac{RSS}{43-3}} \sim F_{2,40}$$

Αν οι δύο ερμηνευτικές μεταβλητές έχουν μεγάλη επίδραση στην εξαρτημένη μεταβλητή τότε SSR>>RSS. Δηλαδή στην περίπτωση αυτή οι ερμηνευτικές μεταβλητές εξηγούν μεγάλο μέρος της μεταβλητότητας της εξαρτημένης μεταβλητής ενώ ο τυχαίος όρος όχι. Μεγάλη τιμή της στατιστικής F $\Leftrightarrow F > F_{cr}$. Άρα **η πιθανότητα να απορριφθεί η μηδενική υπόθεση είναι μεγάλη.**

Παράδειγμα:

SSR = 11776,18

RSS = 1805,17

T = 52

Β.ε RSS = 52 – 3 = 49

Με εφαρμογή του τύπου προκύπτει:

$$F = \frac{\frac{11776,18}{2}}{\frac{1805,17}{49}} = 159,83$$

Η αντίστοιχη κριτική τιμή είναι η εξής όπως προκύπτει από τον πίνακα:

$F_{0,05;2,49} \approx 3,19$ για επίπεδο σημαντικότητας α = 0,05

Επειδή η τιμή της στατιστικής υπερβαίνει την κριτική τιμή η παλινδρόμηση είναι στατιστικά σημαντική ($F > F_{0,05;2,49}$)

Σχέση μεταξύ της στατιστικής F και του συντελεστή προσδιορισμού R^2

Η στατιστική F δίνεται από την σχέση: $F = \frac{\frac{SSR}{k-1}}{\frac{RSS}{T-k}} \sim F_{\kappa-1,T-k}$

$$\Leftrightarrow \frac{T-k}{k-1}\left(\frac{SSR}{RSS}\right) = \frac{T-k}{k-1}\left(\frac{SSR}{TSS-SSR}\right) = \left(\frac{T-k}{k-1}\right)\left(\frac{\frac{SSR}{TSS}}{\frac{TSS-SSR}{TSS}}\right)$$

$$= \left(\frac{T-k}{k-1}\right)\frac{R^2}{1-R^2} \Leftrightarrow F = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{T-k}}$$

$A\nu \quad R^2 = 0 \Leftrightarrow F = 0$

$R^2 >> \Leftrightarrow F >>$

$R^2 \to 1 \Leftrightarrow F \to \infty$

Συνεπώς διαπιστώνεται η ύπαρξη θετικής σχέσης μεταξύ των δύο συντελεστών.

Ο έλεγχος δηλαδή F είναι ένας έλεγχος στατιστικής σημαντικότητας του συντελεστή προσδιορισμού.

Πιο συγκεκριμένα:

$H_o : \beta_2 = \beta_3 = 0$ είναι ισοδύναμο με $R^2 = 0$

$H_1 : \beta_2 \neq 0 \;\; \eta \;\beta_3 \neq 0 \;\eta\; \beta_2 \neq 0 \;\kappa\alpha\iota\; \beta_3 \neq 0$ είναι ισοδύναμο με $R^2 \neq 0$

Παραβιάσεις του γραμμικού υποδείγματος

Α. Πολυσυγγραμμικότητα

Μια από τις βασικές υποθέσεις του κλασσικού γραμμικού υποδείγματος είναι ότι δεν πρέπει οι ερμηνευτικές μεταβλητές να εξαρτώνται μεταξύ τους.

Αν υπάρχει **πλήρης εξάρτηση** ανάμεσα σε δύο ερμηνευτικές μεταβλητές (η μια από τις δύο μεταβλητές διαφέρει από την άλλη κατά σταθερό όρο π.χ $X_1 = X_2 + c$ ή κατά πολλαπλασιαστικό παράγοντα $X_1 = cX_2$ ή συνδυασμό αυτών $X_1 = c_1X_2 + c_2$) τότε το υπόδειγμα δεν μπορεί να εκτιμηθεί.
Η ορίζουσα του πίνακα $(\mathbf{X}'\mathbf{X})^{-1}$ **δεν υπάρχει, εφόσον ο βαθμός του πίνακα δεν είναι πλήρης, υπάρχει συσχέτιση μεταξύ των γραμμών και των στηλών του πίνακα.**

Έτσι από την σχέση

$$X'Xb = X'Y$$

δεν μπορεί να προσδιορισθεί το διάνυσμα b.

Αν συσχέτιση ανάμεσα σε δύο ερμηνευτικές μεταβλητές δεν είναι ακριβής και είναι μερική (π.χ στις χρονολογικές σειρές) τότε το υπόδειγμα μπορεί να εκτιμηθεί αλλά τα αποτελέσματα της εκτίμησης δεν ερμηνεύουν ακριβώς την οικονομική ή άλλης μορφής σχέση που υπάρχει.

ΔΙΑΠΙΣΤΩΣΗ ΤΗΣ ΠΟΛΥΣΥΓΓΡΑΜΜΙΚΟΤΗΤΑΣ

1. Συντελεστής πολλαπλού προσδιορισμού και t στατιστικές

Όταν ο συντελεστής πολλαπλού προσδιορισμού έχει υψηλή τιμή που αυτό συνεπάγεται ότι η F – στατιστική έχει υψηλή τιμή και παράλληλα οι ατομικοί συντελεστές δεν είναι στατιστικά σημαντικοί αυτό φανερώνει ύπαρξη πολυσυγγραμμικότητας στο υπόδειγμα.

2. Συντελεστές συσχέτισης

Κριτήριο Frisch

Έστω θέλουμε να ελέγξουμε αν το υπόδειγμα

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + e_t$$

παρουσιάζει πολυσυγγραμμικότητα.

Εκτιμούμε διαδοχικά τα υποδείγματα

$\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_2 X_{2t}$, με r_1^2, se($\hat{\beta}_2$)

$\hat{Y}_t = \hat{\gamma}_1 + \hat{\gamma}_2 X_{3t}$, με r_2^2, se($\hat{\gamma}_2$)

$\hat{Y}_t = \hat{\delta}_1 + \hat{\delta}_2 X_{4t}$, με r_2^2, se($\hat{\delta}_2$)

Από τα τρία αυτά υποδείγματα επιλέγουμε το καλύτερο, αυτό που ικανοποιεί τα θεωρητικά και στατιστικά κριτήρια.
Κατόπιν εισάγουμε διαδοχικά τις άλλες ερμηνευτικές μεταβλητές και εξετάζουμε την επίδραση που θα έχουν στους συντελεστές παλινδρόμησης στα τυπικά σφάλματά τους και στον συντελεστή πολλαπλού προσδιορισμού.

Αν η εισαγωγή της νέα μεταβλητής στο υπόδειγμα **βελτιώνει την συνολική ερμηνευτικότητα** του υποδείγματος και **οι συντελεστές παλινδρόμησης δεν μεταβάλλονται σημαντικά** η νέα μεταβλητή είναι **χρήσιμη και παραμένει** στο υπόδειγμα.

Αν η εισαγωγή της νέα μεταβλητής στο υπόδειγμα **μεταβάλλει σημαντικά τα πρόσημα ή το μέγεθος των συντελεστών παλινδρόμησης** η νέα μεταβλητή **παραμένει** στο υπόδειγμα.

Κριτήριο Klein

Αν προσδιορισθεί ο μερικός συντελεστής συσχέτισης $r_{x_i xj}$ μεταξύ των δύο μεταβλητών των X_i και X_j, που θεωρούμε ότι συσχετίζονται γραμμικά, και έχουμε ότι

$$r^2_{x_i xj} \geq R^2_{y.x_1 x_2 .....x_k}$$

τότε η πολυσυγγραμμικότητα στο υπόδειγμα είναι επιβλαβής.

Κριτήριο Theil

Αν μια ανεξάρτητη μεταβλητή X_i συσχετίζεται με τις υπόλοιπες και έχουμε ότι

$$R^2_{x_i .x_1 x_2 .....x_k} \geq R^2_{y.x_1 x_2 .....x_k}$$

τότε η πολυσυγγραμμικότητα στο υπόδειγμα είναι επιβλαβής.

ΜΕΘΟΔΟΙ ΕΚΤΙΜΗΣΗΣ ΥΠΟΔΕΙΓΜΑΤΩΝ ΜΕ ΠΟΛΥΣΥΓΓΡΑΜΜΙΚΟΤΗΤΑ

1. Όταν γνωρίζουμε την **ακριβή τιμή ενός ή περισσοτέρων συντελεστών** ή την **ακριβή σχέση μεταξύ συντελεστών** του υπό εκτίμηση υποδείγματος

Ι. Έστω ότι στο υπόδειγμα

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + e_t$$

γνωρίζουμε ότι το $\beta_3 = 0{,}47$.
Τότε αντικαθιστούμε στο υπόδειγμα το $\boldsymbol{\beta}_3 = 0{,}47$ και εκτιμούμε το υπόδειγμα:

$$Y_t - 0{,}47 X_{3t} = \beta_1 + \beta_2 X_{2t} + \beta_4 X_{4t} + e_t$$

Δηλαδή εκτιμούμε τα $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_4$.

II. Έστω ότι στο υπόδειγμα

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + e_t$$

γνωρίζουμε το $\beta_3 = 0{,}4\beta_2$.
Τότε αντικαθιστούμε στο υπόδειγμα το $\beta_3 = 0{,}4\beta_2$ και εκτιμούμε το υπόδειγμα:

$$Y_t = \beta_1 + \beta_2 X_{2t} + 0{,}4\beta_2 X_{3t} + \beta_4 X_{4t} + e_t$$

ή

$$Y_t = \beta_1 + \beta_2 (X_{2t} + 0{,}4X_{3t}) + \beta_4 X_{4t} + e_t$$

ή

$$Y_t = \beta_1 + \beta_2 X^* + \beta_4 X_{4t} + e_t \text{ όπου } X_t^* = (X_{2t} + 0{,}4X_{3t})$$

Δηλαδή εκτιμούμε τα $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_4$. Από την σχέση $\beta_3 = 0{,}4\beta_2$ εκτιμούμε και το $\mathbf{b}_3$.

III. Στις οικονομικές σχέσεις αρκετές φορές οι ερμηνευτικές μεταβλητές συσχετίζονται εκ της φύσεως και υπάρχει φανερό πρόβλημα πολυσυγγραμμικότητας.
Στην περίπτωση του μοντέλου παραγωγής $Y = AL^{\beta_1}K^{\beta_2}e$ η εργασία L και το κεφάλαιο K δεν είναι ανεξάρτητες μεταβλητές.
Με την προϋπόθεση της σχέσης $\beta_1 + \beta_2 = 1$ (σταθερές αναλογίες κλίμακος στην παραγωγική διαδικασία) εκτιμάται τελικά το υπόδειγμα.

$$\text{Ln}\left(\frac{Y}{K}\right) = \text{LnA} + \beta_1 \text{ Ln}\left(\frac{L}{K}\right) + \text{Lne}$$

Εκτιμώνται τα $\hat{A}$, b_1 και από την σχέση $\beta_1 + \beta_2 = 1$ έχουμε $b_2 = 1 - b_1$

Παραδείγματα

1. Έστω το υπόδειγμα $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + e_t$ με $X_3 = 2X_2 - 1$
Το υπόδειγμα αυτό παρουσιάζει πολυσυγγραμμικότητα και δεν μπορεί να εκτιμηθεί όμως μπορούμε να εκτιμήσουμε το μοντέλο που προκύπτει αν αντικατασταθεί το X_3 με το $2X_2 - 1$.

Έτσι έχουμε $Y = \beta_1 + \beta_2 X_2 + \beta_3(2X_2 - 1) + \beta_4 X_4 + e$

ή $Y = \beta_1 - \beta_3 + (\beta_2 + 2\beta_3)X_2 + \beta_4 X_4 + e$

ή $Y = \gamma_0 + \gamma_1 X_2 + \beta_4 X_4 + e$ με $\gamma_0 = \beta_1 - \beta_3$ και $\gamma_1 = (\beta_2 + 2\beta_3)$

το οποίο εκτιμάται και συγκεκριμμένα εκτιμώνται τα $\hat{\gamma}_0$, $\hat{\gamma}_1$ και $\hat{\beta}_4$.
Από τους αρχικούς συντελεστές εκτιμάται μόνο ο $\hat{\beta}_4$.

2. Δίνονται οι τιμές για τις μεταβλητές X_1 και X_2 και X_3 στο υπόδειγμα

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + e_t$$

Μπορεί να εκτιμηθεί το υπόδειγμα;

$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$
2	3	7
3	5	9
5	7	13
7	9	17
9	12	21
11	15	25
13	17	29

Αν παρατηρήσουμε τις τιμές των μεταβλητών $\mathbf{X}_1$ και $\mathbf{X}_3$ θα δούμε ότι υπάρχει μια σχέση γραμμική αυτών των μεταβλητών της μορφής:

$\mathbf{X}_3 = 2\,\mathbf{X}_1 + \mathbf{3}$, πράγμα που σημαίνει ότι το υπόδειγμα παρουσιάζει πολυσυγγραμμικότητα και πρέπει να εκτιμηθεί πλέον το υπόδειγμα που θα προκύψει αν αντικαταστήσουμε το $\mathbf{X}_3$ με το $2\mathbf{X}_1 + \mathbf{3}$ όπως στο παράδειγμα 1.

Β. Αυτοσυσχέτιση

Μία από τις βασικές υποθέσεις του γραμμικού υποδείγματος είναι πως η **συνδιακύμανση των διαταρακτικών όρων είναι μηδενική.**
Αυτό σημαίνει πως:

$$Cov(e_t, e_s) = 0 \quad t \neq s$$

Η παραπάνω υπόθεση σημαίνει πως οι διάφορες τιμές του διαταρακτικού όρου δεν συσχετίζονται μεταξύ τους.
Η συνηθέστερη μορφή αυτοσυσχέτισης είναι η **αυτοσυσχέτιση πρώτης τάξης**.

Σε αυτή την μορφή ισχύει το εξής:

$$e_t = \rho e_{t-1} + \varepsilon_t$$

ρ: ο συντελεστής αυτοσυσχέτισης
$\varepsilon_t \sim (0, \sigma_t^2)$, πρόκειται για μία τυχαία μεταβλητή

Στην περίπτωση που η τιμή του διαταρακτικού όρου της παρούσας περιόδου συσχετίζεται με την τιμή του διαταρακτικού όρου περισσότερων παρελθουσών τιμών αυτού η μορφή της αυτοσυσχέτισης δίνεται από την σχέση:

$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \rho_3 e_{t-3} + \varepsilon_t$$

Σε αυτή την περίπτωση η μορφή της αυτοσυσχέτισης λέγεται αυτοπαλίνδρομο σχήμα **τρίτης τάξης.**

Αυτοσυσχέτιση πρώτης τάξης

Έστω το γραμμικό υπόδειγμα $Y_t = \alpha + \beta X_{1t} + \gamma X_{2t} + u_t$

Τα κατάλοιπα χαρακτηρίζονται από το αυτοπαλίνδρομο σχήμα πρώτης τάξης. Πιο συγκεκριμένα:

$$e_t = \rho e_{t-1} + \varepsilon_t$$

Ο διαταρακτικός όρος της παρούσας περιόδου είναι δυνατόν να προσδιοριστεί ως συνάρτηση των παρελθουσών περιόδων.

Ισχύει:

$$e_t = \rho e_{t-1} + \varepsilon_t = \rho(\rho e_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \rho^2 e_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t = \rho^2(\rho e_{t-3} + \varepsilon_{t-2}) + \rho\varepsilon_{t-1} + \varepsilon_t$$
$$= \rho^3 e_{t-3} + \rho^2 \varepsilon_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t = \dots\dots\dots\dots\dots\dots\dots\dots\dots = \rho^t e_o + \sum_{s=0}^{t-1} \rho^{t-(s+1)} \varepsilon_{s+1}$$

Επειδή ο συντελεστής αυτοσυσχέτισης ρ < 1 για πολύ μεγάλη τιμή του t η για μηδενική τιμή του e_o η παραπάνω σχέση γράφεται ως εξής:

$$e_t = \sum_{s=0}^{\infty} \rho^s \varepsilon_{t-s}$$

Σε αυτή την περίπτωση ο μέσος, η διακύμανση και η συνδιακύμανση δίνονται από την σχέση:

$$E(e_t)=0$$
$$Var(e_t)=\sigma^2=\sigma_\varepsilon^2\frac{1}{1-\rho^2}$$
$$\operatorname{cov}(e_t,e_{t-s})=\rho^s\sigma^2$$

Συνέπειες της αυτοσυσχετίσεως

Ο **εκτιμητής** ελαχίστων τετραγώνων που προκύπτει από ένα υπόδειγμα που χαρακτηρίζεται από **αυτοσυσχέτιση** είναι **αναποτελεσματικός.**
Η **εκτίμηση της διακύμανσης** είναι **μεροληπτική.**
Η θετική αυτοσυσχέτιση μεταξύ της ερμηνευτικής μεταβλητής και του διαταρακτικού όρου οδηγεί σε σοβαρή υποεκτίμηση της διακύμανσης του συντελεστή β.
Συνεπώς **τα συμπεράσματα που προκύπτουν θα είναι αναξιόπιστα και οι διακυμάνσεις θα είναι μεροληπτικές.**
Επιπλέον, **οι προβλέψεις είναι αναποτελεσματικές** με δεδομένο ότι οι διακυμάνσεις των συντελεστών και του διαταρακτικού όρου δεν είναι ελάχιστες.
Ένα υπόδειγμα που συμπεριλαμβάνει κ – ερμηνευτικές μεταβλητές η διακύμανση των συντελεστών των ελαχίστων τετραγώνων δίνεται από την σχέση:

$$V(\hat{\beta})=\sigma^2\left(X'X\right)^{-1}X'\Omega X\left(X'X\right)^{-1}$$

Για την αυτοσυσχέτιση πρώτης τάξεως έχουμε:

$$\Omega=\begin{bmatrix}1 & \rho & \dots & \rho^{T-1}\\ \rho & 1 & \dots & \rho^{T-2}\\ \dots & \dots & \dots & \dots\\ \rho^{T-1} & \rho^{T-2} & \dots & 1\end{bmatrix}\text{ και } V(e)=\sigma^2\Omega$$

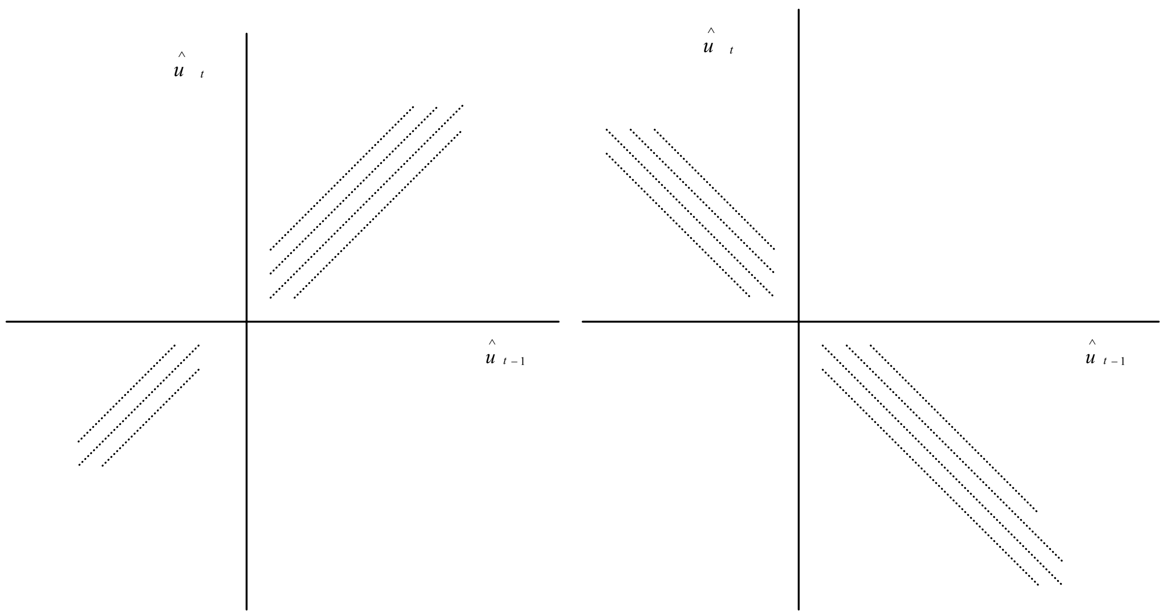
Έλεγχος για την ύπαρξη της αυτοσυσχέτισης

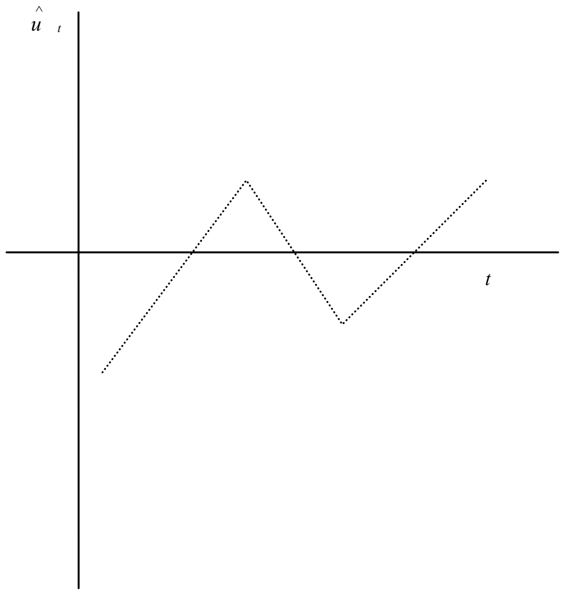
1. Γραφική ανάλυση των καταλοίπων

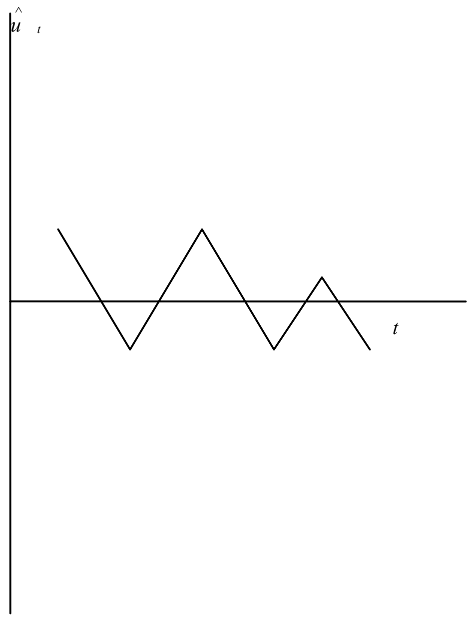
Η μέθοδος αυτή περιλαμβάνει την κατασκευή του διαγράμματος διασποράς των καταλοίπων $\hat{e}_t$ και $\hat{e}_{t-1}$.

Σύμφωνα με τον έλεγχο αυτόν υπάρχει ένδειξη **θετικής αυτοσυσχέτισης** αν οι παρατηρήσεις των καταλοίπων συγκεντρώνονται στο **πρώτο** και **τρίτο τεταρτημόριο**, ενώ αν η συγκέντρωση των παρατηρήσεων των καταλοίπων γίνεται στο **δεύτερο και τέταρτο τεταρτημόριο** έχουμε την ύπαρξη **αρνητικής αυτοσυσχέτισης** (Σχήμα 1 και 2).

Εναλλακτική μέθοδος γραφικής απεικόνισης είναι **η διαχρονική απεικόνιση των καταλοίπων**. Αν τα **σημεία των διαδοχικών καταλοίπων παραμένουν αμετάβλητα** αυτό αποτελεί ένδειξη **θετικής αυτοσυσχέτισης,** ενώ η **συχνή εναλλαγή των σημείων** δηλώνει **αρνητική αυτοσυσχέτιση** (Σχήμα 3 και 4).







Βασικό μειονέκτημα αυτής της μεθόδου είναι η **χαμηλή αξιοπιστία** της σε σχέση με τους υπόλοιπους αυστηρούς στατιστικούς ελέγχους.

2. Έλεγχος Durbin – Watson

Το εν λόγω κριτήριο στηρίζεται στην στατιστική που δίνεται από τον τύπο:

$$d = \frac{\sum_{t=2} \left(\hat{e}_t - \hat{e}_{t-1} \right)^2}{\sum_{t=1} e_t^2}$$

Η στατιστική αυτή είναι γνωστή ως **στατιστική Durbin – Watson.**

Αναπτύσσοντας την ταυτότητα προκύπτει:

$$d = \frac{\sum_{t=2}\left(\hat{e}_t^2 + \hat{e}^2_{t-1} + 2\hat{e}_t\,\hat{e}_{t-1}\right)}{\sum_{t=1}\hat{e}^2_t} = \frac{\sum_{t=2}\hat{e}_t^2 + \sum_{t=2}\hat{e}_{t-1}^2 + 2\sum_{t=2}\hat{e}_t\,\hat{e}_{t-1}}{\sum_{t=1}\hat{e}_t^2} = 2\left(1 - \frac{\sum_{t=1}\hat{e}_t\,\hat{e}_{t-1}}{\sum_{t=1}e_t^2}\right) \approx 2(1-\hat{\rho})$$

$$\hat{\rho} = \frac{\sum \hat{e}_t\,\hat{e}_{t-1}}{\sum \hat{e}_t^2}$$

Με δεδομένο ότι ο συντελεστής συσχέτισης ανήκει στο διάστημα $[-1,1]$ η εν λόγω στατιστική παίρνει τιμές στο διάστημα $[0,4]$.

Πιο συγκεκριμένα:

- Αν δεν υπάρχει αυτοσυσχέτιση $\hat{\rho} = 0$ έχουμε d = 2
- Αν υπάρχει τέλεια θετική αυτοσυσχέτιση $\hat{\rho} = 1$ έχουμε d = 0
- Αν υπάρχει τέλεια αρνητική αυτοσυσχέτιση $\hat{\rho} = -1$ έχουμε d = 4

Ορίζουμε την μηδενική και την εναλλακτική υπόθεση περί ύπαρξης αυτοσυσχέτισης μεταξύ των καταλοίπων:

$$H_o : \rho = 0$$
$$H_1 : \rho \neq 0$$

Για την πραγματοποίηση του ελέγχου χρησιμοποιούνται πίνακες οι οποίοι περιέχουν ανώτερα και κατώτερα όρια τα οποία συγκρίνονται με την τιμή της στατιστικής που προκύπτει από την αντικατάσταση των δεδομένων του υποδείγματος.

Τα **δυνητικά αποτελέσματα** του ελέγχου για την ύπαρξη **θετικής αυτοσυσχέτισης** είναι τα εξής:

$d < d_L$, υπάρχει **θετική αυτοσυσχέτιση** και **απορρίπτεται η μηδενική υπόθεση για μη ύπαρξη αυτοσυσχέτισης**

$d > d_U$ **δεν υπάρχει αυτοσυσχέτιση**

$d_L < d < d_U$ **το αποτέλεσμα του ελέγχου είναι αβέβαιο**.

Για την ύπαρξη της **αρνητικής αυτοσυσχέτισης** τα **δυνητικά αποτελέσματα** είναι τα εξής:

$4 - d < d_L$, υπάρχει αρνητική **αυτοσυσχέτιση και απορρίπτεται η μηδενική υπόθεση για μη ύπαρξη αυτοσυσχέτισης**

$4 - d > d_U$ **δεν υπάρχει αυτοσυσχέτιση**

$d_L < d < d_U$ το **αποτέλεσμα του ελέγχου είναι αβέβαιο.**

Το εν λόγω κριτήριο είναι κατάλληλο για έλεγχο αυτοσυσχέτισης πρώτης τάξης.

Σημαντικό μειονέκτημα του ελέγχου είναι η αβέβαιη περιοχή ενώ βασική αδυναμία είναι η μη εφαρμογή του ελέγχου αν το υπόδειγμα περιλαμβάνει χρονικές υστερήσεις της εξαρτημένης μεταβλητής.

Παράδειγμα

Εκτιμήθηκε από 20 παρατηρήσεις το υπόδειγμα Y = -2,461 + 0,78 X

(0,50) (0,01)

που αφορούσε την σχέση παλινδρόμησης των εισαγωγών Y επί του ΑΕΠ (X). Ευρέθηκαν $\sum \hat{e}_t^2$ =573,069 και $\sum (e_t - e_{t-1})^2$ = 537,192.

Να ελεγχθεί με τον δείκτη D –W αν υπάρχει αυτοσυσχέτιση των τιμών του διαταρακτικού όρου.(Ανδρικόπουλος, τομ.Α,2003).

Ο δείκτης D –W = $\sum (e_t - e_{t-1})^2 / \sum \hat{e}_t^2 = \frac{537{,}192}{573{,}069} = 0{,}937$

Από τους πίνακες του D –W σε επίπεδο σημαντικότητας α=5% με 20 παρατηρήσεις και μια ανεξάρτητη μεταβλητή (κ=1) βρίσκουμε τις τιμές d_L= 1,20 και d_U= 1,41.

Επειδή ο δείκτης D –W = 0,937 < d_L= 1,20 η υπόθεση H_0: ρ>0 δεν απορρίπτεται. Συνεπώς υπάρχει σημαντική θετική συσχέτιση.

Αυτό επιβεβαιώνεται από την προσεγγιστική τιμή του εκτιμητή

$$\hat{\rho} \approx 1 - \frac{d}{2} = 1 - \frac{0{,}937}{2} = 0{,}5315.$$

3. Έλεγχος με την t - στατιστική

Ο έλεγχος αυτός είναι δυνατόν να πραγματοποιηθεί όταν τα κατάλοιπα ενός υποδείγματος είναι παρατηρούμενη μεταβλητή η οποία μπορεί να εκτιμηθεί με την μέθοδο των ελαχίστων τετραγώνων. Με δεδομένη την ισχύ των βασικών υποθέσεων του γραμμικού υποδείγματος ακολουθείται η εξής διαδικασία:

Υπολογίζονται τα κατάλοιπα του υποδείγματος από την εφαρμογή της μεθόδου των ελαχίστων τετραγώνων:

Πρώτο στάδιο:

$$\hat{e}_t = Y_t - \hat{\beta}_o - \hat{\beta}_1 X_{1t} - .... - \hat{\beta}_k X_{kt}$$

Δεύτερο στάδιο:

Εκτιμάται το υπόδειγμα: $e_t = \rho \hat{e}_{t-1} + \varepsilon_t$

Ορίζουμε την μηδενική και την εναλλακτική υπόθεση:

$$H_o : \rho = 0$$
$$H_1 : \rho \neq 0$$

Η στατιστική t δίνεται από την σχέση: $t = \dfrac{\hat{\rho} - 0}{s.e.(\hat{\rho})}$

Με σύγκριση των τιμών της στατιστικής με αυτήν των πινάκων προκύπτει το αποτέλεσμα του ελέγχου.

4. Κριτήριο h του Durbin

Βελτίωση του ελέγχου του Durbin – Watson αποτελεί ο έλεγχος με το κριτήριο h του Durbin. Ο έλεγχος αυτός μπορεί να χρησιμοποιηθεί ακόμη και αν το υπόδειγμα περιλαμβάνει την χρονική υστέρηση της εξαρτημένης μεταβλητής.
Έστω το υπόδειγμα:

$$\mathrm{Y}_t = \beta_0 + \beta_1 \mathrm{X}_t + \beta_2 \mathrm{Y}_{t-1} + \mathrm{e}_t \qquad \text{με} \qquad \mathrm{e}_t = \rho \mathrm{e}_{t-1} + \nu_t$$

Η στατιστική που χρησιμοποιείται είναι η εξής:

$$h = \hat{\rho} \sqrt{\frac{T}{1 - TVar(b_2)}}$$

T: το μέγεθος του δείγματος

$\hat{\rho}$: η εκτίμηση του συντελεστή αυτοσυσχέτισης πρώτης τάξεως που προκύπτει από τα κατάλοιπα των ελαχίστων τετραγώνων.
Η θεωρητική τιμή της h – στατιστικής ακολουθεί την Z ή την t κατανομή ανάλογα το μέγεθος του δείγματος.
Αν ορίσουμε την μηδενική και την εναλλακτική υπόθεση ως ακολούθως:

$$H_o : \rho = 0$$
$$H_1 : \rho \neq 0$$

Συγκρίνεται η τιμή h που υπολογίσθηκε με την z κριτήρια τιμή σε επίπεδο σημαντικότητας α%.

Αν η τιμή h > z ή h < -z τότε η H_0 απορρίπτεται που σημαίνει έχουμε αυτοσυσχέτιση των τιμών του διαταρακτικού όρου (ανάλογα έχουμε και στις άλλες εναλλακτικές υποθέσεις).

Παράδειγμα

Ας υποθέσουμε ότι έχουμε 49 παρατηρήσεις για τις δαπάνες διατροφής Y και το διαθέσιμο εισόδημα X μιας κατηγορίας οικογενειών. Το υπόδειγμα που προέκυψε είναι:

$$Y_t = 49 + 0{,}40\, X_t + 0{,}15\, Y_{t-1}$$

με $Var(b_2) = 0{,}018$, $R^2 = 0{,}85$ και $D - W = 1{,}9$

Να ελεγχθεί με το **Κριτήριο h του Durbin** αν υπάρχει αυτοσυσχέτιση των τιμών του διαταρακτικού όρου.

Υπολογίζουμε τον εκτιμητή $\hat{\rho} \approx 1 - \frac{d}{2} = 1 - \frac{1{,}9}{2} = 0{,}05.$

Επομένως το κριτήριο h του Durbin είναι:

$$h = \hat{\rho}\sqrt{\frac{T}{1 - TVar(b_2)}} = 0{,}05\sqrt{\frac{49}{1 - 49 * 0{,}018}} = 1{,}02$$

Εφόσον το δείγμα είναι μεγάλο η h – στατιστική ακολουθεί την Z κατανομή.

Σε επίπεδο σημαντικότητας α=5% έχουμε κριτήρια τιμή z = 1,96 και - z = -1,96.

Έχουμε -1,96 < h – στατιστική = 1,02 < 1,96 που σημαίνει ότι δεν απορρίπτεται η μηδενική υπόθεση, δηλαδή δεν έχουμε αυτοσυσχέτιση.

Έλεγχος αυτοσυσχέτισης μεγαλύτερης τάξης

Για τον έλεγχο της αυτοσυσχέτισης μεγαλύτερης τάξης μπορούν να χρησιμοποιηθούν δύο έλεγχοι με την κατανομή F και την κατανομή X^2. Το αυτοπαλίνδρομο σχήμα p - τάξεως δίνεται από την σχέση:

$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + ...\rho_p e_{t-p}$$

Ορίζουμε την μηδενική και την εναλλακτική υπόθεση:

$$H_o : \rho_1 = \rho_2 = ... = \rho_k = 0$$
$$H_1 : \rho_1 \neq 0 \; ή / και \; \rho_2 \neq 0 \; ή / και ....\rho_k \neq 0$$

Εκτιμούμε το υπόδειγμα με τη μέθοδο των ελαχίστων τετραγώνων και υπολογίζουμε τα κατάλοιπα $\hat{e}_t$.
Εκτιμάται η βοηθητική παλινδρόμηση μεταξύ των καταλοίπων και των χρονικών υστερήσεων αυτών.
Συνεπώς εκτιμάται η παλινδρόμηση:

$$\hat{e}_t = \alpha_o + \alpha_1 X_{1t} + \alpha_2 X_{2t} + ... + \rho_1 e_{t-1} + \rho_2 e_{t-1} + ... + \rho_k e_{t-k} + \varepsilon_t$$

Έλεγχος Breusch – Godfrey

Οι παρατηρήσεις του δείγματος είναι Τ
Η στατιστική που χρησιμοποιείται είναι η:

$$(T-p)R^2 \sim \mathrm{X}^2_{T-p}$$

Αυτό σημαίνει πως ο έλεγχος πραγματοποιείται με σύγκριση της στατιστικής με την X^2 κατανομή ενώ οι βαθμοί ελευθερίας είναι T-p.

Η μηδενική υπόθεση απορρίπτεται όταν $(T-p)R^2 > \mathrm{X}^2_{T-p}$.

Ενώ η μηδενική υπόθεση γίνεται δεκτή όταν $(T-p)R^2 < \mathrm{X}^2_{T-p}$.

Έλεγχος με την F - κατανομή

Σε αυτόν τον έλεγχο η στατιστική που χρησιμοποιείται είναι η εξής:

$$F = \frac{(RSSr - RSSur)/\, p}{RSSur/(T-k-p-1)} \sim F_{p,T-k-1;a}$$

ΜΕΘΟΔΟΙ ΕΚΤΙΜΗΣΗΣ ΥΠΟΔΕΙΓΜΑΤΩΝ ΜΕ ΑΥΤΟΣΥΣΧΕΤΙΣΗ

Εξετάζουμε την περίπτωση να έχουμε αυτοσυσχέτιση α τάξης **AR(1).** Χρησιμοποιείται η γενικευμένη μέθοδος των ελαχίστων τετραγώνων. Συγκεκριμένα μετατρέπεται το αρχικό υπόδειγμα που έχει αυτοσυσχέτιση α τάξης σε υπόδειγμα που να μην παρουσιάζει αυτοσυσχέτιση.

Α. Εξετάζουμε την περίπτωση όταν είναι **γνωστός ο συντελεστής αυτοσυσχέτιασης** $\hat{\rho}$**.**

Έστω το υπόδειγμα Τ παρατηρήσεων:

Y = Xβ + ε (υπό μορφή πινάκων) το οποίο παρουσιάζει αυτοσυσχέτιση α τάξης.

Μετατρέπουμε το παραπάνω υπόδειγμα στο υπόδειγμα:

$$\begin{bmatrix} Y_2-\hat{\rho}Y_1 \\ Y_3-\hat{\rho}Y_2 \\ \ldots\ldots\ldots \\ \ldots\ldots\ldots \\ \ldots\ldots\ldots \\ Y_T-\hat{\rho}Y_{T-1} \end{bmatrix} = \begin{bmatrix} 1.....X_2-\hat{\rho}X_1 \\ 1.....X_3-\hat{\rho}X_2 \\ 1.......... \\ .... \\ .... \\ 1.....X_T-\hat{\rho}X_{T-1} \end{bmatrix} \begin{bmatrix} \beta_0(1-\hat{\rho}) \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1^* \\ \varepsilon_2^* \\ \ldots \\ \ldots \\ .. \\ \varepsilon_{T-1}^* \end{bmatrix} \quad (\alpha)$$

Το οποίο μπορεί να εκτιμηθεί και το οποίο είναι ομοσκεδεαστικό.

B. Σε περίπτωση που **δεν είναι γνωστός** ο **συντελεστής αυτοσυσχέτισης** μπορούμε να χρησιμοποιήσουμε την διαδικασία των Hildreth – Lu.

Συγκεκριμένα επειδή γνωρίζουμε ότι ο συντελεστής συτοσυσχέτισης παίρνει τιμές από το -1 έως το 1, δίνουμε τιμές σ’ αυτόν τιμές 0 , - 0,1, 0,1, -0,2, 0,2,………, -0,9, 0,9, -1, 1.

Για κάθε τιμή το μετασχηματισμένο υπόδειγμα (α) εκτιμάται με την μέθοδο των ελαχίστων τετραγώνων και υπολογίζεται το άθροισμα των τετραγώνων των καταλοίπων. Το υπόδειγμα που θα δώσει το ελάχιστο άθροισμα των τετραγώνων των καταλοίπων θα είναι το ζητούμενο.

Γ. Ετεροσκεδαστικότητα

Μία από τις σημαντικότερες παραβιάσεις του γραμμικού υποδείγματος είναι αυτή της υπόθεσης:

$$E(e_t)^2 = \sigma^2$$

Αυτό σημαίνει πως η διακύμανση του τυχαίου όρου σφάλματος μεταβάλλεται από παρατήρηση σε παρατήρηση. Δηλαδή ισχύει:

$$E(e_t)^2 = \sigma_t^{\,2}$$

Βασικές συνέπειες της ύπαρξης ετεροσκεδαστικότητας σε ένα υπόδειγμα είναι η αναποτελεσματικότητα των εκτιμητών b_{OLS} ενώ αυτοί διατηρούν την αμεροληψία τους.

Αν γραφεί το υπόδειγμα με τη μορφή μητρών τότε το διμεταβλητό υπόδειγμα που δίνεται από την σχέση:

$$Y_t = \beta_0 + \beta_1 X_t + e_t$$

γράφεται με την ακόλουθη μορφή:

$$Y = X\mathrm{B} + e$$

Για το υπόδειγμα ισχύει:

$$
\begin{aligned}
&E(e) = 0 \\
&V(e) = E(e'e) = \sigma^2 \Omega \\
&E(e'e) = \begin{bmatrix} \sigma_1^2 & 0 & .... & 0 \\ 0 & \sigma_2^2 & .... & 0 \\ 0 & 0 & .... & 0 \\ 0 & 0 & 0 & \sigma_T^2 \end{bmatrix}
\end{aligned}
$$

Όταν ισχύουν οι εν λόγω υποθέσεις η διακύμανση του εκτιμητή b_{OLS} θα δίνεται από την σχέση:

$$V(b_{OLS}) = \sigma^2 (X'X)^{-1} X'\Omega X (X'X)^{-1}$$

Αυτό σημαίνει πως η διακύμανση εμφανίζεται μεγαλύτερη σε ένα υπόδειγμα που πάσχει από ετεροσκεδαστικότητα με αποτέλεσμα να μην είναι δυνατή η χρήση αυτού για τον έλεγχο στατιστικής σημαντικότητας του υποδείγματος.

Στην περίπτωση που αυτά χρησιμοποιηθούν τότε τα συμπεράσματα για τις παραμέτρους στον πληθυσμό θα είναι αναξιόπιστα. Για τον λόγο αυτό οι εκτιμητές παύουν να είναι άριστοι διατηρώντας όμως την ιδιότητα της συνέπειας.

Κριτήρια ελέγχου για ετεροσκεδαστικότητα

Ο έλεγχος για ύπαρξη ετεροσκεδαστικότητας μπορεί να γίνει με διάφορα κριτήρια μεταξύ των οποίων τα σημαντικότερα είναι τα εξής:

1. **Συντελεστής συσχετίσεως Spearman**
2. **Κριτήριο Goldfeld – Quand**
3. **Κριτήριο White**

1. Συντελεστής συσχετίσεως Spearman

Ο συντελεστής συσχετίσεως κατά τάξεις του **Spearman** δίνεται από την σχέση:

$$r_s = 1 - \frac{6\sum_{i=1}^{T} d_1^2}{T(T^2 - 1)} \qquad (\alpha)$$

d_i: οι διαφορές των τάξεων των καταλοίπων και των αντίστοιχων τιμών της ερμηνευτικής μεταβλητής με την οποία συνδέεται ο τυχαίος όρος με την διακύμανση.

Ο έλεγχος που γίνεται αφορά την εξής υπόθεση:

$$Ho : \rho_s = 0$$
$$H_1 : \rho_s \neq 0$$

Η υπόθεση ότι **δεν υπάρχει ετεροσκεδαστικότητα** ισοδυναμεί με τον έλεγχο της υπόθεσης H_0 ότι ο **συντελεστής συσχέτισης είναι μηδέν**.

Η στατιστική που χρησιμοποιείται για τον έλεγχο είναι η σχέση (β) και η κριτική τιμή προέρχεται από την t – κατανομή με δεδομένο ότι για T > 8 η:

$$t = r_s \frac{\sqrt{N-2}}{\sqrt{1-r_{6_e}^2}} \quad (\beta)$$

ακολουθεί t – κατανομή με N-2 βαθμούς ελευθερίας.

Άρα τα αποτελέσματα του ελέγχου είναι τα εξής:

Αποδοχή της μηδενικής υπόθεσης αν $|t| < t_{a/2,N-2}$ άρα **δεν υπάρχει ετεροσκεδαστικότητα.**

Απόρριψη της μηδενικής υπόθεσης αν $|t| \succ t_{a/2,N-2}$ άρα **υπάρχει ετεροσκεδαστικότητα.**

Παράδειγμα

Εξετάσθηκαν οι δαπάνες για διατροφή Y και το διαθέσιμο εισόδημα X σε 25 οικογένειες. Το υπόδειγμα που εκτιμήθηκε είναι;

Y = 0,745 + 0,823 X

(1,10) (32,12)

Με $R^2 = 0,975$.

Δίνεται $\sum_{i=1}^{25} d_i^2 =$ 39. Να εξετασθεί αν το υπόδειγμα παρουσιάζει ετροσκεδαστικότητα χρησιμοποιώντας τον συντελεστή συσχέτισης Spearman.

Έχουμε :

$$r_s = 1 - \frac{6\sum_{i=1}^{T} d_1^2}{T(T^2-1)} = 1- \frac{6(39)}{25(25^2-1)} = 1- 0,015 = 0,985$$

Άρα $t = r_s \frac{\sqrt{N-2}}{\sqrt{1-r_{6e}^2}} = 0,985 \frac{\sqrt{25-2}}{\sqrt{1-(0,985)^2}} = 27,33.$

Εφόσον t =27,33 > $t_{a/2,N-2}$ = $t_{0,025,23}$ = 2,069 η υπόθεση H_0 απορρίπτεται που σημαίνει ~~δεν~~ **υπάρχει ετεροσκεδαστικότητα**.

2. Κριτήριο Goldfeld – Quand

Ο έλεγχος αυτός πραγματοποιείται σε επί μέρους στάδια. Τα στάδια του ελέγχου αυτού είναι τα εξής:

- Κατατάσσουμε τις παρατηρήσεις της ερμηνευτικής μεταβλητής με την οποία υποθέτουμε ότι συνδέεται η διακύμανση του διαταρακτικού όρου κατά φθίνουσα σειρά.

- Στην συνέχεια παραλείπουμε c - κεντρικές παρατηρήσεις. Ο αριθμός των παρατηρήσεως που παραλείπεται δεν πρέπει να ξεπερνά **το ένα τέταρτο του συνολικού αριθμού των παρατηρήσεων.**

- Στην συνέχεια χωρίζουμε το εναπομείναν δείγμα σε δύο επί μέρους δείγματα. Η πρώτη ομάδα (**ομάδα 1**) περιλαμβάνει (T-c)/2 χαμηλές τιμές της ερμηνευτικής μεταβλητής ενώ η δεύτερη ομάδα (**ομάδα 2**) τις (T-c)/2 υψηλές τιμές της ερμηνευτικής μεταβλητής.

- Εφαρμόζουμε την μέθοδο των ελαχίστων τετραγώνων και στις δύο ομάδες.

- Με την εφαρμογή των δύο μεθόδων εκτιμάμε και τα αντίστοιχα αθροίσματα τετραγώνων των καταλοίπων οπότε και υπολογίζουμε την στατιστική που θα χρησιμοποιήσουμε ως κριτήριο ελέγχου ($SSR_{1,}$ SSR_2 τα αντίστοιχα αθροίσματα τετραγώνων των καταλοίπων στις δύο ομάδες)

- Ο έλεγχος που θα πραγματοποιηθεί αφορά την ακόλουθη μηδενική και εναλλακτική υπόθεση:

$$
\begin{aligned}
Ho &: \sigma_1^2 = \sigma_2^2 \\
H_1 &: \sigma_1^2 \neq \sigma_2^2
\end{aligned}
$$

Η H_o σημαίνει ότι δεν υπάρχει ετεροσκεδαστικότητα εφόσον θεωρούμε ότι $\sigma_1^2 = \sigma_1^2 = \sigma^2 =$ σταθερή , δηλαδή η διακυμάνσεις των καταλοίπων είναι ίσες στις δύο ομάδες 1, 2 και σταθερή η τιμή τους.

- Η στατιστική που χρησιμοποιείται είναι η εξής:

$$\lambda = \frac{SSR_2 / v_2}{SSR_1 / v_1} = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} \sim F_{a,v_2,v_1}$$

Όπου v_1, v_2 οι βαθμοί ελευθερίας των ομάδων 1 και 2

Τα αθροίσματα των τετραγώνων των καταλοίπων χρησιμοποιούνται προσεγγιστικά για να αποδώσουν την διακύμανση των καταλοίπων.

- Τα δυνητικά αποτελέσματα του ελέγχου είναι τα εξής:

λ < F_{a,v_2,v_1} αποδοχή της μηδενικής υπόθεσης και άρα **δεν υπάρχει ετεροσκεδαστικότητα**

λ > F_{a,v_2,v_1} απόρριψη της μηδενικής υπόθεσης και άρα **υπάρχει ετεροσκεδαστικότητα**

Παράδειγμα

Έχουμε 24 παρατηρήσεις που αφορούν 24 νοικοκυριά και στα οποία εξετάσθηκαν οι καταναλωτικές δαπάνες Y σε σχέση με το διαθέσιμο εισόδημά τους X.

Κατατάσσουμε τις τιμές X κατ' αύξουσα τάξη μεγέθους και αφαιρούμε τις τέσσερις κεντρικές τιμές. Έτσι δημιουργούμε δύο ομάδες παρατηρήσεων την ομάδα (A) που οι τιμές X είναι κάτω από τις κεντρικές τιμές και την ομάδα (B) που είναι πάνω από τις κεντρικές τιμές.

Εκτιμούμε τα δύο υποδείγματα που αντιστοιχούν στις δύο ομάδες.

Ομάδα A:

$$Y_A = 1{,}03 + 0{,}91X_A,\ R_A^2 = 0{,}978,\ \hat{\sigma}_A^2 = 0{,}435$$
$$(0{,}567)\ (0{,}056)$$

$$Y_B = 3{,}56 + 0{,}78X_B,\ R_B^2 = 0{,}934,\ \hat{\sigma}_B^2 = 3{,}437$$
$$(0{,}345)\ (0{,}087)$$

Άρα $\lambda = \frac{SSR_2 / \nu_2}{SSR_1 / \nu_1} = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_A^2} = \frac{3{,}437}{0{,}435} = 7{,}90$

Επειδή $F_{(5\%,8,8)} = 3{,}44 < \lambda = 7{,}90$ η υπόθεση της ετεροσκεδαστικότητας απορρίπτεται.

(Όπου $v_1 = v_2 = [(T\text{-}c\text{-}2k)/2]$ οι βαθμοί ελευθερίας, c ο αριθμός των κεντρικών τιμών, κ 0 αριθμός των παραμέτρων του υποδείγματος.

Στην περίπτωσή μας έχουμε: $v_1 = v_2 = [(T\text{-}c\text{-}2k)/2] = (24\text{-}4\text{-}2{*}2)/2 = 8)$.

3. Κριτήριο White

Έστω το υπόδειγμα:

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + e_t$$

Το κριτήριο αυτό δεν προϋποθέτει τον προσδιορισμό της ερμηνευτικής μεταβλητής από την οποία εξαρτάται η διακύμανση καθώς και την κανονικότητα των καταλοίπων. Ο έλεγχος βασίζεται στον συντελεστή προσδιορισμού της παλινδρόμησης των καταλοίπων στις ερμηνευτικές μεταβλητές του υποδείγματος. Και η εφαρμογή του κριτηρίου αυτού πραγματοποιείται σε στάδια.

Αυτά είναι τα εξής:

- Εφαρμόζεται η μέθοδος των ελαχίστων τετραγώνων στο αρχικό υπόδειγμα με αποτέλεσμα να εκτιμώνται και τα κατάλοιπα του υποδείγματος.

➢ Στην συνέχεια εκτιμάται η βοηθητική παλινδρόμηση:

$$\hat{e}_t^2 = \alpha_0 + \alpha_1 X_{t1} + \alpha_2 X_{t2} + \alpha_3 X_{t1}^2 + \alpha_4 X_{t2}^2 + \alpha_5 X_{t1} X_{t2} + \varepsilon_t$$

Στην συνέχεια υπολογίζεται ο συντελεστής Προσδιορισμού.

Υπό την ισχύ της μηδενικής υπόθεσης η στατιστική

$$TR^2 \sim \mathrm{X}_{a,p}^2 \text{ κατανομή}$$

όπου p = ο αριθμός των ερμηνευτικών μεταβλητών και
T = το σύνολο των παρατηρήσεων

➢ Ορίζω την μηδενική και την εναλλακτική υπόθεση.

$$H_0 : \alpha_0 = \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$$
$$H_1 : \alpha_0 \neq 0 \text{ ή/και } \alpha_1 \neq 0 \text{ ή/και } \alpha_2 \neq 0 \text{ ή/και } \alpha_3 \neq 0 \text{ ή/και } \alpha_4 \neq 0 \text{ ή/και } \alpha_5 \neq 0$$

Για τον έλεγχο της υπόθεσης χρησιμοποιείται η παραπάνω στατιστική.

➢ Τα αποτελέσματα του ελέγχου υποθέσεων είναι τα εξής:

- Αποδοχή της μηδενικής υπόθεσης όταν $TR^2 < \mathrm{X}_{a,p}^2$

  (**δεν υπάρχει ετεροσκεδαστικότητα**)

- Απόρριψη της μηδενικής υπόθεσης όταν $TR^2 > \mathrm{X}_{a,p}^2$

  (**υπάρχει ετεροσκεδαστικότητα**)

Παράδειγμα

Από 40 δεδομένα που αφορούν την σχέση Κατανάλωσης Y και διαθέσιμου εισοδήματος X προέκυψε η βοηθητική παλινδρόμηση:

$$\hat{e}_t^2 = -45{,}2 + 3{,}14X_{t1} - 2{,}54X_{t2} - 0{,}021X_{t1}^2 + 0{,}32X_{t2}^2 + 0{,}085X_{t1}X_{t2} + \varepsilon_t$$

Με $R^2 = 0{,}353$, $T R^2 = 14{,}12$

Έχουμε p =5, $X^2_{0,05} = 11{,}07 < T R^2 = 14{,}12$

Η μηδενική υπόθεση ότι **δεν υπάρχει ετεροσκεδαστικότητα δεν γίνεται δεκτή.**

Αν πάρουμε α=1% τότε $X^2_{0,01} = 15{,}08 > T R^2 = 14{,}12$ οπότε γίνεται **δεκτή** η μηδενική υπόθεση ότι **δεν υπάρχει ετεροσκεδαστικότητα**

ΜΕΘΟΔΟΙ ΕΚΤΙΜΗΣΗΣ ΥΠΟΔΕΙΓΜΑΤΩΝ ΜΕ ΕΤΕΡΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑ

Όταν υποθέσουμε την μορφή της ετεροσκεδαστικότητας μπορούμε να μετασχηματίσουμε το υπόδειγμα που πάσχει από αυτήν.

Συγκεκριμένα υποθέτουμε ότι:

$$\sigma_t^2 = \sigma^2 X_t^2$$

όπου X_t η ερμηνευτική μεταβλητή για την οποία θεωρούμε ότι η διακύμανση του διαταρακτικού όρου μεταβάλλεται ανάλογα με το τετράγωνο αυτής.

Μετασχηματίζουμε το υπόδειγμα:

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

διαιρώντας κατά μέλη με την X_t

Οπότε προκύπτει το υπόδειγμα:

$$Y^*_t = \beta_0 X^*_t + \beta_1 + \varepsilon^*_t$$

με $Y^*_t = \frac{Y_t}{X_t}$, $X^*_t = \frac{1}{X_t}$, $\varepsilon^*_t = \frac{\varepsilon_t}{X_t}$

Το υπόδειγμα αυτό δεν παρουσιάζει ετεροσκεδαστικότητα και μπορεί να εκτιμηθεί.

- **Γενικά αν θεωρήσουμε ότι $\sigma_t^2 = \sigma^2 f(X_t)$ τότε το υπόδειγμα γίνεται ομοσκεδαστικό αν διαιρέσουμε την αρχική σχέση με $\sqrt{f(X_t)}$.**

  - Αν θεωρήσουμε την σχέση $\sigma_t^2 = \sigma^2 X_t$ διαιρούμε με $\sqrt{X_t}$
  - Αν θεωρήσουμε την σχέση $\sigma_t^2 = \sigma^2 E(Y_t)$ διαιρούμε με $\sqrt{\hat{Y}_t}$

Παράδειγμα

Η σχέση μεταξύ αποταμίευσης Y και διαθέσιμου εισοδήματος X που προέκυψε από 18 οικογένειες είναι η ακόλουθη: (Γ. Χρήστου, τομ.Α, 2002).

$$\underset{(3,63)}{Y_t = -12{,}45} + \underset{(0{,}014)}{0{,}203X_t}, \quad R^2 = 0{,}90$$

Υποθέτουμε ότι $\sigma_t^2 = \sigma^2 X_t^2$ οπότε το υπόδειγμα μετασχηματίζεται στο:

$$Y^*_t = 0{,}207 - 13{,}21X^*_t$$

με $Y^*_t = \dfrac{Y_t}{X_t}$, $X^*_t = \dfrac{1}{X_t}$, $\varepsilon^*_t = \dfrac{\varepsilon_t}{X_t}$

Άρα η συνάρτηση αποταμίευσης που προκύπτει όταν μετασχηματίσουμε τις μεταβλητές είναι:

$$\hat{Y}_t = -13{,}21 + 0{,}207X_t$$

$$(2{,}15) \quad (0{,}01)$$

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Ανδρέας Ανδρικόπουλος (2003): Οικονομετρία, Βασική θεωρία και εφαρμογές, Τόμος Α, εκδόσεις Μπένου, Αθήνα, 2003.

2. Κολυβά - Μαχαίρα Φ., Μπόρα – Σέντα Ε. (1999): Στατιστική, εκδόσεις Ζήτη, Θεσσαλονίκη, 1999.

3. Μπόρα – Σέντα Ε., Μωυσιάδης Χ. (1997): Εφαρμοσμένη Στατιστική, Β' έκδοση, εκδόσεις Ζήτη, Θεσσαλονίκη, 1997.

4. Γεώργιος Χρήστου (2002): Εισαγωγή στην Οικονομετρία, Τόμος Α, εκδόσεις Gutenberg, Αθήνα, 2002.

5. Norman Draper – Harry Smith (1997): Εφαρμοσμένη ανάλυση παλινδρόμησης, Μετάφραση – Επιμέλεια Ε. Χατζηκωνσταντινίδης, Α. Καλαματιανού, εκδόσεις Παπαζήση, Αθήνα, 1997.