Γραμμικά Μοντέλα
Το Γραμμικό Μοντέλο σε Μορφή Πινάκων

Διδάσκουσα: Λουκία Μελιγκοτσίδου
Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών
Τμήμα Μαθηματικών

March 27, 2020

Το Απλό Γραμμικό Μοντέλο με Πίνακες

Έχουμε γράψει το απλό γραμμικό μοντέλο για κάθε μεταβλητή στο δείγμα μας ως εξής
$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, i = 1, \ldots, n.$$

Δηλαδή έχουμε
$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_1 + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_2 + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_n + \varepsilon_n \end{aligned}$$

<u>Ορίζουμε</u>
$$\underset{(n\times 1)}{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \underset{(n\times 2)}{X} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}, \underset{(2\times 1)}{B} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \underset{(n\times 1)}{\mathcal{E}} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$
Το μοντέλο μπορεί να γραφτεί σε μορφή πινάκων ως
$$\underset{(n\times 1)}{Y} = \underset{(n\times 1)}{XB} + \underset{(n\times 1)}{\mathcal{E}}$$

Θέλουμε να εκτιμήσουμε συνολικά το διάνυσμα B με τη μέθοδο ελαχίστων τετραγώνων.

<u>Θα χρειαστούμε:</u>

- $Y'Y = (Y_1, Y_2, \ldots, Y_n) \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \sum Y_i^2$

  $$X'X = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ X_1 & X_2 & \ldots & X_n \end{pmatrix} \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} = \begin{pmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \sum Y_i \\ \sum X_i Y_i \end{pmatrix}$$

- Αντίστροφος τετραγωνικού πίνακα (2×2)
  $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ και έστω η ορίζουσα $D = ad - cb \neq 0$
  Τότε μπορούμε να βρούμε τον αντίστροφο: $A^{-1} = \begin{pmatrix} d/D & -b/D \\ -c/D & a/D \end{pmatrix}$
  $\Rightarrow$ O $X'X = \begin{pmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{pmatrix}$ έχει ορίζουσα
  $D = n \sum X_i^2 - (\sum X_i)^2 = n\Big[\sum X_i^2 - \dfrac{(\sum X_i)^2}{n}\Big] = n\sum(X_i - \bar{X})^2 > 0$
  Άρα $(X'X)^{-1} = \begin{pmatrix} \dfrac{\sum X_i^2}{n\sum(X_i - \bar{X})^2} & -\dfrac{\bar{X}}{\sum(X_i - \bar{X})^2} \\ -\dfrac{\bar{X}}{\sum(X_i - \bar{X})^2} & \dfrac{1}{\sum(X_i - \bar{X})^2} \end{pmatrix}$

- Έστω Y τυχαίο διάνυσμα $n \times 1$, δηλαδή Y διάνυσμα του οποίου τα στοιχεία είναι τυχαίες μεταβλητές. Είναι
  $E(Y) = \begin{pmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{pmatrix}$, $\underbrace{D(Y) = \Sigma}_{\text{Πίνακας Συνδιακύμανσης}}$ $n \times n$ πίνακας, συμμετρικός και θετικά ημιορισμένος με ij στοιχείο $\Sigma_{ij} = Cov(Y_i, Y_j)$.
  Στην κύρια διαγώνιο του Σ βρίσκονται οι διασπορές $V(Y_i)$, ενώ τα μη διαγώνια στοιχεία του είναι οι ανά δύο συνδιακυμάνσεις. Στο γραμμικό μοντέλο υπό μορφή πινάκων, $Y = XB + \mathcal{E}$, οι κλασικές υποθέσεις για τους τυχαίους όρους γράφονται ως $E(\mathcal{E}) = 0$ και $D(\mathcal{E}) = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} = \sigma^2 I_{(n\times n)}$.

- Εάν Y, W τυχαία διανύσματα και A πίνακας σταθερών τέτοια ώστε $W = AY$, τότε
  $E(A) = A$
  $E(W) = E(AY) = AE(Y)$
  $D(W) = D(AY) = AD(Y)A'$

Επομένως, στο γραμμικό μοντέλο το Y είναι τυχαίο διάνυσμα με

$$E(Y) = E(XB + \mathcal{E}) = E(XB) + E(\mathcal{E}) = XB + 0 = XB$$
$$D(Y) = D(XB + \mathcal{E}) = D(\mathcal{E}) = \sigma^2 I_{(n\times n)}.$$

Εκτιμήτρια ελαχίστων τετραγώνων του B

Κανονικές εξισώσεις:
$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i$$
$$\hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_i$$

Υπό μορφή πινάκων:

$$X'X\hat{B} = X'Y \qquad (1)$$

γιατί $X'X\hat{B} = \begin{pmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i \\ \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2 \end{pmatrix}$
και $X'Y = \begin{pmatrix} \sum Y_i \\ \sum X_i Y_i \end{pmatrix}$

$$(1) \overset{\text{Λύνουμε ως προς } \hat{B}}{\Longrightarrow} \underbrace{(X'X)^{-1}(X'X)}_{I}\hat{B} = (X'X)^{-1}X'Y \Rightarrow \boxed{\hat{B} = (X'X)^{-1}X'Y}$$

Ο εκτιμητής ελαχίστων τετραγώνων του διανύσματος B, ως στατιστική συνάρτηση, είναι τυχαίο διάνυσμα με αναμενόμενη τιμή $E(\hat{B}) = E\Big((X'X)^{-1}X'Y\Big) = (X'X)^{-1}X'E(Y)$
$$= (X'X)^{-1}X'[E(XB + \mathcal{E})]$$
$$= (X'X)^{-1}X'XB = B,$$

και πίνακα συνδιακύμανσης
$$D(\hat{B}) = D\Big((X'X)^{-1}X'Y\Big) = (X'X)^{-1}X'D(Y)\Big((X'X)^{-1}X'\Big)'$$
$$= (X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1}$$
$$= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}.$$

Μετασχηματισμοί Συνάρτησης Παλινδρόμησης ώστε να γίνει Γραμμική

Ένδειξη για το αν η σχέση ανάμεσα στην απαντητική ή εξαρτημένη μεταβλητή, Y, και την επεξηγηματική ή ανεξάρτητη μεταβλητή, X, είναι γραμμική αποτελεί το γράφημα (σημειόγραμμα) των ζευγών x_i, y_i του δείγματος. Σε κάποιες περιπτώσεις, που υπάρχει σχέση ανάμεσα στις μεταβλητές αλλά αυτή δεν είναι γραμμική, είναι εφικτό, με κατάλληλους μετασχηματισμούς να προκύψει ένα απλό γραμμικό μοντέλο και να αξιοποιηθούν οι γνωστοί τρόποι εκτίμησης και συμπερασματολογίας.

- Πολλαπλασιαστικό Μοντέλο

  $Y_i = \gamma_0 \gamma_1^{X_i} \varepsilon_i$, όπου γ_0, γ_1 παράμετροι και ε_i τ.μ. με $E(\varepsilon_i) = 1$.

  Παίρνοντας το μετασχηματισμό $Y_i' = \log_{10} Y_i$ έχουμε

  $$Y_i' = \log_{10} Y_i = \log_{10} \gamma_0 + X_i \log_{10} \gamma_1 + \log_{10} \varepsilon_i$$

  ή $Y_i' = \beta_0 + \beta_1 X_i + \varepsilon_i'$, όπου $\beta_0 = \log_{10} \gamma_0, \beta_1 = \log_{10} \gamma_1$ παράμετροι και ε_i' τυχαία σφάλματα με $E(\varepsilon_i) = 0$.

- Μοντέλο $Y_i = \beta_0 + \dfrac{\beta_1}{X_i} + \varepsilon_i$

  Χρησιμοποιούμε το μετασχηματισμό $X_i' = \dfrac{1}{X_i}$, οπότε έχουμε

  $$Y_i = \beta_0 + \beta_1 X_i' + \varepsilon_i$$

- Μοντέλο $Y_i = \dfrac{1}{1 + exp\{\beta_0 + \beta_1 X_i + \varepsilon_i\}}$

  Παίρνουμε $Y_i' = \dfrac{1}{Y_i}$, οπότε

  $$Y_i' = 1 + exp\{\beta_0 + \beta_1 X_i + \varepsilon_i\}$$

  Συνεχιζουμε θέτοντας $Y_i'' = \ln(Y_i' - 1)$,οπότε

  $$Y_i'' = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Πολλαπλή Παλινδρόμηση

Το γενικό γραμμικό μοντέλο με p ανεξάρτητες μεταβλητές, ή αλλιώς το μοντέλο πολλαπλής παλινδρόμησης, γράφεται ως

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + \varepsilon_i, \quad i = 1, \ldots, n.$$

Μετασχηματισμοί

- Μοντέλο $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i$
  Θέτουμε $X_{i1} = X_i$ και $X_{i2} = X_i^2$, οπότε παίρνουμε
  $$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$
- Μοντέλο $Y_i = \dfrac{1}{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i}$
  Θέτουμε $Y_i' = \dfrac{1}{Y_i}$, οπότε έχουμε
  $$Y_i' = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$
- Μοντέλο $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i$
  θέτουμε $X_{i3} = X_{i1} X_{i2}$, οπότε έχουμε
  $$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i.$$

Το Μοντέλο με Δύο Επεξηγηματικές Μεταβλητές

Το μοντέλο
$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i,$$
σε μορφή πινάκων γράφεται ως
$$\underset{(n\times 1)}{Y} = \underset{(n\times 1)}{XB} + \underset{(n\times 1)}{\mathcal{E}},$$
όπου
$$X = \begin{pmatrix} 1 & X_{11} & X_{12} \\ 1 & X_{21} & X_{22} \\ \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} \end{pmatrix}, \quad B = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \quad Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

Οι κανονικές εξισώσεις είναι και πάλι

$$\underbrace{X'X\hat{B} = X'Y}_{\text{εξισώσεις γενικά για το γραμμικό μοντέλο}}$$

Για το συγκεκριμένο μοντέλο, οι κανονικές εξισώσεις παίρνουν τη μορφή

$$X'X\hat{B} = \begin{pmatrix} n & \sum X_{i1} & \sum X_{i2} \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1}X_{i2} \\ \sum X_{i2} & \sum X_{i2}X_{i1} & \sum X_{i2}^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \sum Y_i \\ \sum X_{i1}Y_i \\ \sum X_{i2}Y_i \end{pmatrix} = X'Y$$

Λύνοντας το σύστημα των κανονικών εξισώσεων προκύπτουν οι εκτιμητές ελαχίστων τετραγώνων των συντελεστών του μοντέλου.

Το Γενικό Γραμμικό Μοντέλο σε Μορφή Πινάκων

Το γενικό γραμμικό μοντέλο, με p ανεξάρτητες μεταβλητές γράφεται ως

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + \varepsilon_i, \quad i = 1, \ldots, n.$$

Σε μορφή πινάκων, οποιοδήποτε γραμμικό μοντέλο (απλή παλινδρόμηση, πολλαπλή παλινδρόμηση, μετασχηματισμένο μοντέλο ώστε να γίνει γραμμικό) γράφεται ως

$$\underset{(n\times 1)}{Y} = \underset{(n\times 1)}{XB} + \underset{(n\times 1)}{\mathcal{E}},$$

όπου X ο πίνακας σχεδιασμού του μοντέλου και B το διάνυσμα των παραμέτρων που έχει αντίστοιχη διάσταση με τον πίνακα σχεδιασμού, δηλαδή ανάλογα με το πόσες στήλες έχει ο πίνακας X τόσοι είναι και οι προς εκτίμηση συντελεστές του γραμμικού μοντέλου (σταθερά και συντελεστές κλίσης).

Στη γενική περίπτωση έχουμε

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & X_{11} & \ldots & X_{1p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \ldots & X_{np} \end{pmatrix}, \quad B = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \mathcal{E} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Κανονικές εξισώσεις

$$X'X\hat{B} = X'Y \qquad \Rightarrow \qquad \hat{B} = (X'X)^{-1}X'Y$$

Ο εκτιμητής ελαχίστων τετραγώνων του διανύσματος B έχει αναμενόμενη τιμή

$$\begin{aligned} E(\hat{B}) = E\Big((X'X)^{-1}X'Y\Big) &= (X'X)^{-1}X'E(Y) \\ &= (X'X)^{-1}X'[E(XB + \mathcal{E})] \\ &= (X'X)^{-1}X'XB = B, \end{aligned}$$

και πίνακα συνδιακύμανσης

$$\begin{aligned} D(\hat{B}) = D\Big((X'X)^{-1}X'Y\Big) &= (X'X)^{-1}X'D(Y)\Big((X'X)^{-1}X'\Big)' \\ &= (X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}. \end{aligned}$$

Αν επιπλέον έχει γίνει η υπόθεση της κανονικότητας για τους τυχαίους όρους, τότε η κατανομή του διανύσματος $\hat{B}$ είναι η πολυδιάστατη κανονική με διάνυσμα μέσων $E(\hat{B}) = B$ και πίνακα συνδιακύμανσης $D(\hat{B}) = \sigma^2(X'X)^{-1}$. Τα διαγώνια στοιχεία του πίνακα $D(\hat{B})$ είναι οι διακυμάνσεις των επί μέρους εκτιμητών των αντίστοιχων παραμέτρων (των αντίστοιχων στοιχείων του διανύσματος $\hat{B}$), ενώ τα μη διαγώνια στοιχεία αυτού του πίνακα είναι οι ανα δύο συνδιακυμάνσεις των αντίστοιχων εκτιμητών.

Εκτίμηση του σ^2 στο γενικό γραμμικό μοντέλο

- Υπενθυμίζουμε ότι, αν $Y_1, Y_2, \ldots, Y_n$ τυχαίο δείγμα από κατανομή με γνωστό μέσο μ και άγνωστη διασπορά σ^2, τότε η εκτιμήτρια του σ^2 είναι η $\hat{\sigma}^2 = \frac{1}{n}\sum(Y_i - \mu)^2$. Αν ο μέσος μ είναι άγνωστος θα εκτιμηθεί από το $\bar{Y}$ και το σ^2 εκτιμάται από το $S^2 = \frac{1}{n-1}\sum(Y_i - \bar{Y})^2$ (όπου έχουμε χρησιμοποιήσει το άθροισμα των τετραγωνικών αποκλίσεων των Y_i από τον κοινό τους μέσο). Είναι $E(S^2) = \sigma^2$, δηλαδή η S^2 είναι αμερόληπτη εκτιμήτρια της διασποράς. Διαιρούμε με τους βαθμούς ελευθερίας $n - 1$ γιατί έχουμε εκτιμήσει μία άγνωστη παράμετρο, το μ, μέσω του $\bar{X}$.

- Υπενθυμίζουμε, επίσης, ότι στα γραμμικά μοντέλα τα Y_i έχουν διαφορετικές κατανομές που εξαρτώνται από τα X_i. Επομένως, η απόκλιση κάθε παρατήρησης πρέπει να υπολογιστεί από το μέσο της. Στο απλό γραμμικό μοντέλο είναι $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$. Άρα, αν συμβολίσουμε με $\hat{\varepsilon}_i$ τις εκτιμήσεις των σφαλμάτων (κατάλοιπα-*residuals*), υπολογίζουμε το άθροισμα

$$\sum \hat{\varepsilon_i}^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

error sum of squares or residuals sum of squares
άθροισμα τετραγώνων καταλοίπων

Μια αμερόληπτη εκτιμήτρια του σ^2 είναι η

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

mean square error
μέσο τετραγωνικό σφάλμα

Διαιρούμε με $n-2$ (β.ε.) καθώς έχουν χαθεί δύο βαθμοί ελευθερίας επειδή πριν φτάσουμε στην εκτίμηση της διασποράς έχουν εκτιμηθεί δύο παράμετροι.

- Στο γενικό γραμμικό μοντέλο (πολλαπλή παλινδρόμηση) είναι
  $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \ldots + \hat{\beta}_p X_{pi}$.
  Άρα, μια αμερόληπτη εκτιμήτρια του σ^2 είναι τώρα η

$$\hat{\sigma}^2 = \frac{1}{n-p-1} \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \ldots - \hat{\beta}_p X_{pi})^2$$

mean square error
μέσο τετραγωνικό σφάλμα

Διαιρούμε με $n-p-1$ (β.ε.) καθώς έχουν χαθεί $p+1$ βαθμοί ελευθερίας επειδή πριν φτάσουμε στην εκτίμηση της διασποράς έχουν εκτιμηθεί $p+1$ παράμετροι.

Συμπερασματολογία για τους συντελεστές του γενικού γραμμικού μοντέλου

Αν με $\hat{\beta}_j$, $j = 0, 1, \ldots, p$, συμβολίσουμε τους εκτιμητές των συντελεστών της πολλαπλής παλινδρόμησης, με $S^2(\hat{B}) = \hat{\sigma}^2 (X'X)^{-1}$ την εκτιμήτρια του πίνακα συνδιακύμανσης $D(\hat{B})$ και με $s^2(\hat{\beta}_j)$ το jj διαγώνιο στοιχείο του πίνακα $S^2(\hat{B})$, τότε η στατιστική συνάρτηση $\dfrac{\hat{\beta}_j - \beta_j}{s(\hat{\beta}_j)} \sim t_{(n-p-1)}$. Το αποτέλεσμα αποδεικνύεται με τον ίδιο τρόπο που αποδείχθηκε το αντίστοιχο αποτέλεσμα για τους συντελεστές της απλής γραμμικής παλινδρόμησης. Κάτω από την υπόθεση της κανονικότητας για τους τυχαίους όρους, $\varepsilon_i \sim N(0, \sigma^2)$, ισχύουν $\hat{\beta}_j \sim N(\beta_j, \sigma^2(\beta_j)) \leftarrow \frac{\hat{\beta}_j - \beta_j}{\sigma(\hat{\beta}_j)} \sim N(0, 1)$ και $\frac{\sum \hat{\varepsilon}_j^2}{\sigma^2} \sim \chi^2(n-p-1)$.

Προσοχή, οι βαθμοί ελευθερίας της $t-Student$ κατανομής είναι πάντα οι βαθμοί ελευθερίας του μέσου τετραγωνικού σφάλματος, δηλαδή της εκτιμήτριας του σ^2. Στο αποτέλεσμα αυτό στηρίζεται η κατασκευή διαστημάτων εμπιστοσύνης και η διενέργεια ελέγχων υποθέσεων για τους επί μέρους συντελεστές του γενικού γραμμικού μοντέλου.

Έτσι έχουμε $P\Big(-t_{\frac{\alpha}{2}}(n-p-1) \leq \dfrac{\hat{\beta}_j - \beta_j}{s(\hat{\beta}_j)} \leq t_{\frac{\alpha}{2}}(n-p-1)\Big) = 1-\alpha.$

Άρα το δ.ε. συντελεστή $1-\alpha$ για το β_j είναι

$$\hat{\beta}_j \pm t_{\frac{\alpha}{2}}(n-p-1)s(\hat{\beta}_j).$$

Αντιστοίχως, για τον έλεγχο στατιστικής σημαντικότητας του β_j χρησιμοποιείται η στατιστική συνάρτηση

$$T = \frac{\hat{\beta}_j}{s(\hat{\beta}_j)},$$

η οποία, κάτω από την ισχύ της H_0, ακολουθεί $t(n-p-1)$ κατανομή.

Για τα δεδομένα του δείγματός μας υπολογίζουμε την παρατηρούμενη τιμή της ελεγχοσυνάρτησης στο δείγμα, έστω t^*. Απορρίπτουμε την H_0, σε επίπεδο στατιστικής σημαντικότητας α, αν $|t^*| > t_{\alpha/2}(n-p-1)$ ή ισοδύναμα αν $p-value < \alpha$. Δηλαδή, απορρίπτουμε την $H_0 : \beta_j = 0$ έναντι της αμφίπλευρης εναλλακτικής, αν η παρατηρούμενη τιμή της ελεγχοσυνάρτησης υπερβαίνει το $\alpha/2$ άνω ποσοστιαίο σημείο της $t(n-p-1)$ κατανομής.

Για τον έλεγχο της $H_0 : \beta_j = \beta_{j0}$ έναντι της μονόπλευρης εναλλακτικής $H_1 : \beta_j > \beta_{j0}$ η ελεγχοσυνάρτηση που χρησιμοποιείται είναι η

$$T = \frac{\hat{\beta}_j - \beta_{j0}}{s(\hat{\beta}_j)},$$

η οποία, κάτω από την ισχύ της H_0 ακολουθεί $t(n-p-1)$ κατανομή. Απορρίπτουμε την H_0, σε επίπεδο στατιστικής σημαντικότητας α, έναντι της συγκεκριμένης εναλλακτικής, αν $t^* > t_\alpha(n-p-1)$ ή ισοδύναμα αν $p-value < \alpha$.

Άσκηση

Έστω το γραμμικό μοντέλο

$$Y_i = \beta_0 + \beta_1 X_{1i}^2 + \frac{\beta_2}{X_{2i}} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \quad \text{ανεξάρτητα.}$$

(a) Να βρεθεί ο εκτιμητής ελαχίστων τετραγώνων, $\hat{B}$, του διανύσματος $B = (\beta_0, \beta_1, \beta_2)'$, καθώς και η αναμενόμενη τιμή και ο πίνακας συνδιακύμανσης του $\hat{B}$.

(b) Θεωρήστε το παραπάνω μοντέλο σε μορφή πινάκων, $Y = XB + \mathcal{E}$. Ένα δείγμα $n = 30$ παρατηρήσεων έδωσε τα εξής αποτελέσματα

$$(X'X)^{-1} = \begin{pmatrix} 1.7 & -0.06 & -0.08 \\ -0.06 & 0.92 & 0.55 \\ -0.08 & 0.55 & 0.54 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 1.6 \\ -30 \\ 47 \end{pmatrix}$$

και $\sum_{i=1}^{n} \hat{\epsilon}_i^2 = 6$. Να γίνουν, σε ε.σ.σ $\alpha = 0.05$, οι παρακάτω έλεγχοι υποθέσεων:
(i) ο έλεγχος στατιστικής σημαντικότητας του β_1, (ii) ο έλεγχος της $H_0 : \beta_2 = 7$ έναντι της εναλλακτικής $H_1 : \beta_2 > 7$ και (iii) ο έλεγχος της $H_0 : \beta_1 + \beta_2 = 8$ έναντι της εναλλακτικής $H_1 : \beta_1 + \beta_2 < 8$.
Δίνονται $t_{0.025}(27) = 2.056$ και $t_{0.05}(27) = 1.703$.

Υποδείξεις

(a) Πρόκειται για ένα μοντέλο πολλαπλής παλινδρόμησης, γραμμικό ως προς τις παραμέτρους, για το οποίο πρέπει απλά να οριστεί σωστά ο πίνακας σχεδιασμού και μετά η εκτίμηση ελαχίστων τετραγώνων και η στατιστική συμπερασματολογία γίνεται κατά τα γνωστά.

(b) Για τον έλεγχο (iii) πρέπει να βρεθεί η κατανομή του αθροίσματος $\hat{\beta}_1 + \hat{\beta}_2$ και μετά η κατανομή της στατιστικής συνάρτησης $T = \frac{\hat{\beta}_1 + \hat{\beta}_2 - 8}{s(\hat{\beta}_1 + \hat{\beta}_2)}$. Προσοχή τα $\hat{\beta}_1$ και $\hat{\beta}_2$ δεν είναι ασυσχέτιστα, όπως φαίνεται από τον πίνακα συνδιακύμανσης του $\hat{B}$, και αυτό θα πρέπει να ληφθεί υπόψη στον υπολογισμό της διασποράς $s^2(\hat{\beta}_1 + \hat{\beta}_2)$.