Η από κοινού πιθανοθεωρητική συμπεριφορά 2 τ.μ., X και Y, εκφράζεται πλήρως από την από κοινού κατανομή τους, $F_{X,Y}(x,y)$. Η πιθανοθεωρητική συμπεριφορά της Y για συγκεκριμένη τιμή x της X $(X = x)$ εκφραζεται πλήρως από τη δεσμευμένη κατανομή, $F_{Y|X}(y|x)$

$$(Y|X = x) \sim F_{Y|X=x}$$

Στο μάθημα αυτό θα ασχοληθούμε με την παλινδρόμηση, που αναφέρεται στη συμπεριφορά της μέσης τιμής και της διασποράς της ΄ απόκρισης΄, Y, σαν συνάρτηση της ΄ συμμεταβλητής΄ X. Θα μελετήσουμε το μοντέλο

$$(Y|X = x) \sim (\mu(x), \sigma^2(x)),$$

όπου <u>καταχρηστικά</u> θα χρησιμοποιούμε το σύμβολο $\sim$. Στη παραπάνω έκφραση απλά αναφέρουμε πως $\mu(x) = E(Y|X = x)$ και $\sigma^2(x) = Var(Y|X = x)$. Θα επικεντρωθούμε στην ειδική περίπτωση, όπου η διασπορά δεν εξαρτάται από τη τιμή της X, δηλαδή έχουμε

$$(Y|X = x) \sim (\mu(x), \sigma^2).$$

Η παλινδρόμηση λέγεται γραμμική όταν $\mu(x)$ είναι μια οποιαδήποτε γραμμική συνάρτηση των παραμέτρων $\beta_0, \beta_1, ..., \beta_{p-1}$ (Δηλαδή είναι γραμμική ως προς ΄ τα β,΄ όχι ως προς ΄ τα x΄).

Τα παρακάτω μοντέλα είναι όλα παραδείγματα γραμμικων μοντέλων

$$\begin{aligned}
(Y|X = x) &\sim (\mu(x) = \beta_0 + \beta_1 x, \sigma^2) \\
(Y|X = x) &\sim (\mu(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3, \sigma^2) \\
(Y|X = x) &\sim (\mu(x) = \beta_0 + \beta_1 e^x, \sigma^2) \\
(Y|X = x) &\sim (\mu(x) = \beta_0 + \beta_1 \ln x, \sigma^2),\ x > 0. \\
(Y|X = x) &\sim (\mu(x) = \beta_0 + \beta_1 \frac{1}{x}, \sigma^2),\ x > 0.
\end{aligned}$$

Το παρακάτω μοντέλο δεν είναι γραμμικο (είναι ΄μη γραμμικό΄)

$$(Y|X = x) \sim \left(\mu(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}, \sigma^2\right)$$

Μία ειδική περίπτωση των γραμμικών μοντέλων (που θα μελετήσουμε εκτενώς) υποθέτει κανονική κατανομή για την απόκριση για κάθε x, δηλαδή

$$(Y|X = x) \sim N(\mu(x), \sigma^2),$$

όπου $\mu(x)$ είναι μια γραμμική συνάρτηση των παραμέτρων $\beta_0, \beta_1, ..., \beta_{p-1}$. Σε αυτή τη περίπτωση το σύμβολο $\sim$ σημαίνει ΄ ακολουθεί την κατανομή΄. Ένας ισοδύναμος τρόπος ορισμού ενός μοντέλου παλιδρόμησης προκύπτει ορίζοντας το ΄ σφάλμα΄ $\epsilon(x) = Y - \mu(x)$, και γράφοντας

$$Y = \mu(x) + \epsilon(x), \ \epsilon(x) \sim (0, \sigma^2).$$

1 Εκτιμητής Ελαχίστων Τετραγώνων στην απλή γραμμική παλινδρόμηση

Στην απλή γραμμική παλινδρόμηση έχουμε $p = 2$, κι έτσι $\mu(x) = \beta_0 + \beta_1 x$. Δοθέντος ενός δείγματος $(Y_1, x_1), (Y_2, x_2), \ldots, (Y_n, x_n)$ το απλό γραμμικό μοντέλο (απλή γραμμική παλιδρόμηση) για τα δεδομένα μας γράφεται ως:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \ i = 1, \ldots, n, \tag{1.1}$$

Y_i είναι η απόκριση της i-οστής παρατήρησης, x_i η τιμή της συμμεταβλητής της i-οστής παρατήρησης και ϵ_i είναι το σφάλμα (δηλαδή η απόκλιση της παρατήρησης Y_i από την μέση τιμή της, $E(Y_i) = \mu_i = \beta_0 + \beta_1 x_i$) . Θεωρούμε τα x ώς σταθερές τιμές κι όχι ως τ.μ.. Για τα σφάλματα υποθέτουμε πως είναι ομοσκεδαστικά ($var(\epsilon_i) = \sigma^2$) και ασυσχέτιστα ($cov(\epsilon_i, \epsilon_j) = 0$ για $i \neq j$).

Εναλλακτικά το μοντέλο μπορεί να γραφτεί ώς:

$$\begin{aligned} Y_i | x_i &\sim (\mu_i, \sigma^2), \ cov(Y_i, Y_j | x_i, x_j) = 0, \ i \neq j \\ \mu_i &= \beta_0 + \beta_1 x_i, \ i = 1, \ldots, n, \end{aligned} \tag{1.2}$$

Οι εκτιμητές ελαχίστων τετραγώνων είναι:

$$\begin{aligned} \hat{\beta}_1 &= \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2} \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \end{aligned} \tag{1.3}$$

Σημείωση: Από τα παραπάνω προκύπτει πως $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$, κι άρα η ευθεία ελαχίστων τετραγώνων περνάει από το σημείο $(\bar{x}, \bar{y})$.

Οι εκτιμητές ελαχίστων τετραγώνων ελαχιστοποιούν το άθροισμα των τετραγωνικών σφαλμάτων, ήτοι είναι οι λύσεις του μαθηματικού προβλήματος

$$\min_{\beta_0, \beta_1} \sum \epsilon_i^2(\beta_0, \beta_1) \Leftrightarrow \min_{\beta_0, \beta_1} \sum (y_i - \beta_0 - \beta_1 x_i)^2 \tag{1.4}$$

Παραγωγίζοντας καταλήγουμε πως οι εκτιμητές ελαχίστων τετραγώνων προκύπτουν ως λύσεις των εκτιμητικών εξισώσεων

$$\begin{aligned}\sum \epsilon_i(\beta_0,\beta_1) &= \sum (y_i-\beta_0-\beta_1 x_i)=0\\ \sum x_i\epsilon_i(\beta_0,\beta_1) &= \sum x_i(y_i-\beta_0-\beta_1 x_i)=0 \end{aligned}\qquad (1.5)$$

(Απόδειξη:

$$\begin{aligned}\frac{d}{d\beta_0}\sum (y_i-\beta_0-\beta_1x_i)^2 &= -2\sum (y_i-\beta_0-\beta_1x_i)=0\\ \frac{d}{d\beta_1}\sum (y_i-\beta_0-\beta_1x_i)^2 &= -2\sum x_i(y_i-\beta_0-\beta_1x_i)=0,\end{aligned}$$

από όπου προκύπτουν ως λύσεις οι εκτιμητές $\hat\beta_0,\hat\beta_1$. Για να δείξουμε ότι αυτές οι τιμές ελαχιστοποιούν τη συνάρτηση που έχουμε παίρνουμε τις 'δεύτερες' παραγώγους:

$$\begin{aligned}\frac{d^2}{d\beta_0^2}\sum (y_i-\beta_0-\beta_1x_i)^2 &= 2n\\ \frac{d^2}{d\beta_1^2}\sum (y_i-\beta_0-\beta_1x_i)^2 &= 2\sum x_i^2\\ \frac{d^2}{d\beta_0 d\beta_1}\sum (y_i-\beta_0-\beta_1x_i)^2 &= 2\sum x_i\end{aligned}$$

Πρέπει να δείξω πως ο πίνακας με τις 'δεύτερες' παραγώγους, όταν εκτιμάται στα $\beta_0=\hat\beta_0$ και $\beta_1=\hat\beta_1$ είναι θετικά όρισμένος. Στη περίπτωσή μας έχουμε τον πίνακα

$$2\begin{bmatrix} n & \sum x_i\\ \sum x_i & \sum x_i^2.\end{bmatrix}$$

Αρκεί να έχω $n>0,\sum x_i^2>0$ και η ορίζουσα >0. Η ορίζουσα ισούται με $2(n\sum x_i^2-(\sum x_i)^2)=2n(\sum x_i^2-\frac{1}{n}(\sum x_i)^2)=2n\sum (x_i-\bar x)^2>0$. Άρα έχουμε τοπικό ελάχιστο στο $\beta_0=\hat\beta_0$ και $\beta_1=\hat\beta_1$.)

Οι εκτιμητές ελαχίστων τετραγώνων μπορούν να εκφραστούν ως γραμμικοί συνδυασμοί των Y:

$$\begin{aligned}\hat{\beta}_1 &= \sum c_i Y_i, \text{ με } c_i = \frac{x_i - \bar{x}}{S_{xx}} \\ \hat{\beta}_0 &= \sum d_i Y_i, \text{ με } d_i = \frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{S_{xx}},\end{aligned} \tag{1.6}$$

όπου

$$S_{xx} = \sum (x_i - \bar{x})^2$$

Για τους συντελεστες c_i, d_i ισχύουν τα παρακάτω:

$$\begin{gathered}\sum c_i = 0, \ \sum c_i^2 = \frac{1}{S_{xx}}, \ \sum c_i x_i = 1 \\ \sum d_i = 1, \ \sum d_i^2 = \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}, \ \sum d_i x_i = 0 \\ \sum c_i d_i = -\frac{\bar{x}}{S_{xx}}\end{gathered} \tag{1.7}$$

(

$$\begin{aligned}\sum c_i d_i &= \sum (\frac{x_i - \bar{x}}{S_{xx}})(\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{S_{xx}}) \\ &= \frac{1}{n}\sum (\frac{x_i - \bar{x}}{S_{xx}}) - \sum (\frac{x_i - \bar{x}}{S_{xx}})\frac{(x_i - \bar{x})\bar{x}}{S_{xx}} \\ &= 0 - \frac{\bar{x}}{S_{xx}^2}\sum (x_i - \bar{x})^2 = -\frac{\bar{x}}{S_{xx}}\end{aligned}$$

)

Καθότι $Var(Y_i) = cov(Y_i, Y_i) = \sigma^2 \ \forall i$ και $Cov(Y_i, Y_j) = 0 \ \forall i \neq j$ έχουμε

$$\begin{aligned}E(\sum \alpha_r Y_r) &= \sum \alpha_r E(Y_r) \\ Cov(\sum \alpha_r Y_r, \sum \gamma_r Y_r) &= \sum_r \sum_s \alpha_r \gamma_s cov(Y_r, Y_s) \\ &= \sigma^2 \sum \alpha_r \gamma_r\end{aligned} \tag{1.8}$$

μπορούμε να αποδείξουμε πως

$$\begin{aligned} E(\hat{\beta}_0) = \beta_0, \;\; Var(\hat{\beta}_0) = \sigma^2[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}] \\ E(\hat{\beta}_1) = \beta_1, \;\; Var(\hat{\beta}_1) = \sigma^2 \frac{1}{S_{xx}} \\ Cov(\hat{\beta}_0, \hat{\beta}_1) = -\sigma^2 \frac{\bar{x}}{S_{xx}}. \end{aligned} \tag{1.9}$$

Οι προσαρμοσμένες τιμές (fits) και τα υπόλοιπα ή κατάλοιπα (residuals) ισούνται με

$$\begin{aligned} \hat{Y}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i \\ \hat{\epsilon}_i &= Y_i - \hat{Y}_i. \end{aligned} \tag{1.10}$$

Η προσαρμοσμένη τιμή $\hat{Y}_i$ είναι απλά η εκτίμηση της μέσης τιμής $\mu_i = \beta_0 + \beta_1 x_i$. Σύγχρόνως είναι και πρόβλεψη της τιμής της απόκρισης Y για μια νέα παρατήρηση με τιμή της συμμεταβλητής ίση με x_i. Καθοτι η ευθεία ελαχίστων τετραγώνων περνάει από το σημείο $(\bar{x}, \bar{y})$, ισχύει πως $\sum \hat{Y}_i = \sum Y_i$. Επίσης από τις εκτιμητικές εξισώσεις (1.5) προκύπτει πως $\sum \hat{\epsilon}_i = 0$ και $\sum x_i \hat{\epsilon}_i = 0$

Οι προσαρμοσμένες τιμές και τα υπόλοιπα είναι γραμμικοί συνδυασμοί των παρατηρήσεων $Y_1, \ldots, Y_n$:

$$\begin{aligned} \hat{Y}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i = \sum_{j=1}^{n} (d_j + x_i c_j) Y_j \\ \hat{\epsilon}_i &= Y_i - \hat{Y}_i = \sum_{j=1}^{n} [\delta_{[j=i]} - (d_j + x_i c_j)] Y_j, \end{aligned} \tag{1.11}$$

όπου $\delta_{[j=i]} = 1$ όταν $j = i$ και ισούται με 0 όταν $j \neq i$.

Χρησιμοποιώντας τα αποτελέσματα στις εξισώσεις (1.8) μπορούμε να δείξουμε πώς

$$\begin{aligned}
E(\hat{Y}_i) = \mu_i = \beta_0 + \beta_1 x_i \quad &, \quad var(\hat{Y}_i) = \sigma^2(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}) \\
E(\hat{\epsilon}_i) = 0 \quad &, \quad var(\hat{\epsilon}_i) = \sigma^2(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}) \\
cov(\hat{\epsilon}_i, \hat{\epsilon}_j) = -\sigma^2(\frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{S_{xx}}) \quad &, \quad cov(\hat{\epsilon}_i, \hat{\beta}_0) = 0 \\
cov(\hat{\epsilon}_i, \hat{\beta}_1) = 0 \quad &, \quad cov(\hat{\epsilon}_i, \hat{Y}_j) = 0 \ \forall i, j \\
cov(\hat{\epsilon}_i, \bar{Y}) = 0 \quad &, \quad cov(\hat{\beta}_i, \bar{Y}) = 0 \qquad (1.12)
\end{aligned}$$

Μια απερόληπτη εκτίμηση της διασποράς των σφαλμάτων δίνεται από την $s^2 = \frac{\sum \hat{\epsilon}_i^2}{n-2} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}$. Δηλαδή, ισχύει πως

$$E(s^2) = E(\frac{\sum \hat{\epsilon}_i^2}{n-2}) = E(\frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}) = \sigma^2. \qquad (1.13)$$

(<u>Απόδειξη:</u> Καταρχάς $E(\hat{\epsilon}_i) = E(Y_i - \hat{Y}_i) = E(Y_i) - E(\hat{Y}_i) = \mu_i - \mu_i = 0$

$$\begin{aligned}
E(\sum \hat{\epsilon}_i^2) &= \sum E(\hat{\epsilon}_i^2) \\
&= \sum (Var(\hat{\epsilon}_i) + (E(\hat{\epsilon}_i))^2) \\
&= \sum Var(\hat{\epsilon}_i) \\
&= \sigma^2 \left(n - 1 - \sum \frac{(x_i - \bar{x})^2}{S_{xx}} \right) \\
&= \sigma^2 (n - 2),
\end{aligned}$$

και το αποτέλεσμα προκύπτει άμεσα.)

2 Κανονική Απλή Γραμμική Παλινδρόμηση

Το κανονικό απλό γραμμικό μοντέλο γράφεται ως

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_i + \epsilon_i,\ i = 1, \ldots, n, \\ \epsilon_1, \epsilon_2, \ldots, \epsilon_n &\sim^{iid} N(0, \sigma^2), \end{aligned} \tag{2.1}$$

ή εναλλακτικά

$$\begin{aligned} Y_i | x_i &\sim^{ind} N(\mu_i, \sigma^2),\ i = 1, \ldots, n, \\ \mu_i &= \beta_0 + \beta_1 x_i \end{aligned} \tag{2.2}$$

Η υπόθεση κανονικότητας μαζι με τα αποτελέσματα στο προηγούμενο κεφάλαιο συνεπάγονται τα εξής:

$$\begin{aligned} &\hat{\boldsymbol{\beta}} \sim MVN(\boldsymbol{\beta}, \Sigma) \quad , \\ &\hat{\beta}_0 \sim N(\beta_0, \sigma^2[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}]) \quad , \quad \hat{\beta}_1 \sim N(\beta_1, \sigma^2 \frac{1}{S_{xx}}), \end{aligned} \tag{2.3}$$

όπου η 1η πρόταση παραπάνω απλά αναφέρει πως το 2×1 διάνυσμα των εκτιμητριών

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

ακολουθεί πολυμεταβλητή κανονική κατανομή με μέση τιμή

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

και πίνακα διακυμάνσεων-συνδιακυμάνσεων $\mathbf{\Sigma}$, όπου

$$\mathbf{\Sigma} = \begin{bmatrix} var(\hat{\beta}_0) & cov(\hat{\beta}_0, \hat{\beta}_1) \\ cov(\hat{\beta}_0, \hat{\beta}_1) & var(\hat{\beta}_1) \end{bmatrix} = \begin{bmatrix} \sigma^2_{\hat{\beta}_0} & \sigma_{\hat{\beta}_0, \hat{\beta}_1} \\ \sigma_{\hat{\beta}_0, \hat{\beta}_1} & \sigma^2_{\hat{\beta}_1} \end{bmatrix} = \begin{bmatrix} \sigma^2[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}] & -\sigma^2 \frac{\bar{x}}{S_{xx}} \\ -\sigma^2 \frac{\bar{x}}{S_{xx}} & \sigma^2 \frac{1}{S_{xx}} \end{bmatrix}$$

Μία άμεση συνέπεια του αποτελέσματος στην (2.3) είναι πως κάθε γραμμικός συνδυασμός των εκτιμητριών $\hat{\beta}_0$ και $\hat{\beta}_1$ ακολουθούν κανονική κατανομή. Συγκεκριμένα, έστω l_1, l_2 πραγματικοί αριθμοί (όχι και τα 2 ίσα μηδέν). Τότε

$$l_1 \hat{\beta}_0 + l_2 \hat{\beta}_1 \sim N(l_1 \beta_0 + l_2 \beta_1, l_1^2 \sigma^2_{\hat{\beta}_0} + l_2^2 \sigma^2_{\hat{\beta}_1} + 2 l_1 l_2 \sigma_{\hat{\beta}_0, \hat{\beta}_1}), \tag{2.4}$$

από όπου προκύπτει πως

$$\frac{l_1 \hat{\beta}_0 + l_2 \hat{\beta}_1 - (l_1 \beta_0 + l_2 \beta_1)}{\sigma \sqrt{l_1^2 [\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}] + l_2^2 \frac{1}{S_{xx}} - 2 l_1 l_2 \frac{\bar{x}}{S_{xx}}}} \sim N(0, 1). \tag{2.5}$$

Από τα αποτελέσματα στη (1.12) σε συνδυασμό με την υπόθεση κανονικότητας προκύπτει πως τα υπόλοιπα, $\hat{\epsilon}_i$, είναι στατιστικά ανεξάρτητα των εκτιμητριών, $\hat{\beta}_0$ και $\hat{\beta}_1$. Άρα τα υπόλοιπα, $\hat{\epsilon}_i$, είναι ανεξάρτητα και του γραμμικού συνδυασμού $l_1\hat{\beta}_0 + l_2\hat{\beta}_1$. Τέλος, καθότι $s^2 = \frac{\sum \hat{\epsilon}_i^2}{n-2}$ είναι συνάρτηση των υπολοίπων, $\hat{\epsilon}_i$, έχουμε πως $l_1\hat{\beta}_0 + l_2\hat{\beta}_1$ και s^2 είναι στατιστικά ανεξάρτητα.

Αργότερα θα δείξουμε πως, κάτω από την υπόθεση κανονικότητας των σφαλμάτων, έχουμε πως

$$(n-2)\frac{s^2}{\sigma^2} \sim \chi^2_{n-2}. \tag{2.6}$$

Από τα αποτελέσματα στις (2.5) και (2.6) έχουμε

$$\begin{aligned} &\frac{\frac{l_1\hat{\beta}_0 + l_2\hat{\beta}_1 - (l_1\beta_0 + l_2\beta_1)}{\sigma\sqrt{l_1^2[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}] + l_2^2\frac{1}{S_{xx}} - 2l_1l_2\frac{\bar{x}}{S_{xx}}}}}{\sqrt{(n-2)\frac{s^2}{\sigma^2}/(n-2)}} \\ &= \frac{l_1\hat{\beta}_0 + l_2\hat{\beta}_1 - (l_1\beta_0 + l_2\beta_1)}{s\sqrt{l_1^2[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}] + l_2^2\frac{1}{S_{xx}} - 2l_1l_2\frac{\bar{x}}{S_{xx}}}} \sim t_{n-2}. \end{aligned} \tag{2.7}$$

2.1 Συμπερασματολογία για παραμέτρους, Διάστημα εμπιστοσύνης, Διάστημα Πρόβλεψης

Βάσει του αποτελέσματος στην (2.7) ένα $100(1-a)\%$ διάστημα εμπιστοσύνης για τον γραμμικό συνδυασμό $l_1\beta_0 + l_2\beta_1$ δίνεται από:

$$l_1\hat{\beta}_0 + l_2\hat{\beta}_1 \pm t_{n-2;\frac{\alpha}{2}} s\sqrt{l_1^2[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}] + l_2^2\frac{1}{S_{xx}} - 2l_1l_2\frac{\bar{x}}{S_{xx}}},$$

όπου $t_{n-2;\alpha/2}$ είναι τό <u>άνω</u> $\alpha/2$ ποσοστιαίο σημείο της t_{n-2} κατανομής.

Επίσης ο έλεγχος της μηδενικής υπόθεσης $H_o: \ l_1\beta_0 + l_2\beta_1 = r$ έναντι της δίπλευρης εναλλακτικής $H_a: \ l_1\beta_0 + l_2\beta_1 \neq r$, σε επίπεδο σημαντικότητας α απορρίπτει την μηδενική όταν

$$\frac{|l_1\hat{\beta}_0 + l_2\hat{\beta}_1 - r|}{s\sqrt{l_1^2[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}] + l_2^2\frac{1}{S_{xx}} - 2l_1l_2\frac{\bar{x}}{S_{xx}}}} > t_{n-2;\frac{\alpha}{2}}.$$

Ειδικές περιπτώσεις ενδιαφέροντος:

1) Για την κλίση της ευθείας παλινδρόμησης παίρνουμε $l_1 = 0$ και $l_2 = 1$. Ειδικά, για να ελέγξουμε τη στατιστική σημαντικότητα της παλινδρόμησης κάνουμε τον έλεγχο $H_0 : \beta_1 = 0$ έναντι της δίπλευρης εναλλακτικής. Η μηδενική υπόθεση απορρίπτεται όταν

$$\frac{|\hat{\beta}_1|}{s\sqrt{\frac{1}{S_{xx}}}} > t_{n-2;\frac{\alpha}{2}}. \tag{2.8}$$

2) Για την μέση τιμή της απόκρισης σε καθορισμένη τιμή της συμμεταβλητής x_0, δηλαδή για την $\mu_0 = \beta_0 + \beta_1 x_0$, παίρνουμε $l_1 = 1$ και $l_2 = x_0$. Έτσι για παράδειγμα, ένα 95% <u>Διάστημα Εμπιστοσύνης για την μέση τιμή</u> $\mu_0 = \beta_0 + \beta_1 x_0$ δίνεται από

$$\begin{aligned} \hat{\beta}_0 + \hat{\beta}_1 x_0 \quad &\pm \quad t_{n-2;0.025} s\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{Sxx}} \\ \hat{Y}_0 \quad &\pm \quad t_{n-2;0.025} s\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{Sxx}} \end{aligned}$$

<u>Διάστημα Πρόβλεψης μιας καινούργιας τιμής</u> της απόκρισης για συγκεκριμένη τιμή, x_0, της συμμεταβλητής.

Έστω πως θέλουμε να προβλέψουμε μια νέα τιμη Y_0 δοθέντος συγκεκριμένης τιμης της συμμεταβλητής, x_0. Π.χ. έστω πως θέλουμε να προβλεψουμε την τιμή της FEV γιά ένα παιδί με ύψος 65 ίντσες (εννοείται πως δεν αναφερόμαστε σε παιδί που ήδη παρατηρήσαμε την τιμή του, υπάρχει δηλαδή η παρατήρησή του στην βάση δεδομένων). Θέλουμε δηλαδή να προβλέψουμε την τυχαία μεταβλητή $Y_0 = \beta_0 + \beta_1 x_0 + \epsilon_0$. Μπορεί να αποδειχθεί πως η βέλτιστη γραμμική πρόβλεψη της Y_0 (με κριτήριο το μέσο τετραγωνικό σφάλμα πρόβλεψης) είναι η $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$. Η πρόβλεψη $\hat{Y}_0$ είναι προφανώς ανεξάρτητη της τ.μ. που θέλουμε που θέλουμε να προβλέψουμε, Y_0 (γιατί;). Παρατηρούμε πως η πρόβλεψη της Y_0 είναι ίδια με την αμερόληπτη εκτίμηση της μέσης τιμής $\mu_0 = \beta_0 + \beta_1 x_0$.

Για το σφάλμα της της πρόβλεψης $Y_0 - \hat{Y}_0$ ισχύουν

$$\begin{aligned} E(Y_0 - \hat{Y}_0) \quad &= \quad 0 \\ Var(Y_0 - \hat{Y}_0) \quad &= \quad Var(Y_0) + Var(\hat{Y}_0) = \sigma^2(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{Sxx}) \end{aligned} \tag{2.9}$$

που μαζί με την υπόθεση κανονικότητας μας δίνει

$$\frac{Y_0 - \hat{Y}_0}{\sigma\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{Sxx}}} \sim N(0, 1). \tag{2.10}$$

Ισχύει πως $Y_0 - \hat{Y}_0$ και s^2 είναι ανεξάρτητα (καθώς $Y_0, \hat{\beta}_0, \hat{\beta}_1$ είναι ανεξάρτητα της s^2). Άρα, το αποτέλεσμα στην (2.10) σε συνδυασμό με το αποτέλεσμα στην (2.6) μας δίνει

$$\frac{Y_0 - \hat{Y}_0}{s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{Sxx}}} \sim t_{n-2}, \tag{2.11}$$

από όπου προκύπτει πως ισχύει ότι

$$P(\hat{Y}_0 - t_{n-2;\frac{\alpha}{2}} s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{Sxx}} < Y_0 < \hat{Y}_0 + t_{n-2;\frac{\alpha}{2}} s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{Sxx}}) = 1 - \alpha.$$

Το διάστημα

$$\hat{Y}_0 \quad \pm \quad t_{n-2;\frac{\alpha}{2}} s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{Sxx}}$$

λέγεται ένα $100(1-\alpha)\%$ Διάστημα Πρόβλεψης της Y_0.

2.2 Διάσπαση χ^2, ANOVA

Λήμμα 1. *Έστω $U = U_1 + U_2$. Άν $U \sim \chi^2_k$, $U_1 \sim \chi^2_l$, με $k > l$ και U_1 είναι ανεξάρτητη της U_2, τότε $U_2 \sim \chi^2_{k-l}$*

Απόδειξη. Για της ροπογεννήτριες των τ.μ. ισχύει, λόγω ανεξαρτησίας των U_1 και U_2

$$M_U(t) = M_{U_1}(t)M_{U_2}(t)$$

Η ροπογεννήτρια μιας τ.μ. με χ^2_ν κατανομή ισούται με

$$M(t) = \frac{1}{(1-2t)^{\nu/2}},\ t, 1/2$$

και συνεπώς ισχύει

$$\frac{1}{(1-2t)^{k/2}} = \frac{1}{(1-2t)^{l/2}}(t)M_{U_2}(t),$$

κι άρα

$$M_{U_2}(t) = \frac{1}{(1-2t)^{(k-l)/2}}(t).$$

Καθώς η ροπογεννήτρια (όταν υπάρχει) καθορίζει την κατανομή, συμπεραίνουμε πως $U_2 \sim \chi^2_{k-l}$. □

Θεώρημα 2. *Σε μιά απλή γραμμική παλινδρόμηση με ομοσκεδαστικά και ασυσχέτιστα σφάλματα που ακολουθούν κανονική κατανομή ισχύουν:*

$$
\begin{aligned}
(n-2)\frac{s^2}{\sigma^2} \sim \chi^2_{n-2} \quad &\Leftrightarrow \quad \frac{SSE}{\sigma^2} \sim \chi^2_{n-2} \\
\frac{\hat\beta_1^2 S_{XX}}{\sigma^2} \sim^{H_0} \chi^2_1 \quad &\Leftrightarrow \quad \frac{SSR}{\sigma^2} \sim^{H_0} \chi^2_1 \\
SSE \perp\!\!\!\perp SSR,&
\end{aligned}
$$

όπου $H_0 : \beta_1 = 0$.

Απόδειξη. Γράφουμε

$$
\begin{aligned}
\epsilon_i &= \hat\epsilon_i + (\epsilon_i - \hat\epsilon_i) \\
&= \hat\epsilon_i + (Y_i - \beta_0 - \beta_1 x_i - Y_i + \hat\beta_0 + \hat\beta_1 x_i) \\
&= \hat\epsilon_i + (\hat\beta_0 - \beta_0) + (\hat\beta_1 - \beta_1)x_i. \qquad (2.12)
\end{aligned}
$$

Γνωρίζουμε πως

$$
\begin{aligned}
\bar Y &= \beta_0 + \beta_1 \bar x + \bar\epsilon \\
\bar Y &= \hat\beta_0 + \hat\beta_1 \bar x \\
&\Longrightarrow (\hat\beta_0 - \beta_0) = -(\hat\beta_1 - \beta_1)\bar x - \bar\epsilon. \qquad (2.13)
\end{aligned}
$$

Από τις (2.13) και (;;) προκύπτει

$$
\begin{aligned}
\epsilon_i &= \hat\epsilon_i + (\hat\beta_1 - \beta_1)(x_i - \bar x) - \bar\epsilon \Longrightarrow \\
\epsilon_i^2 &= \hat\epsilon_i^2 + (\hat\beta_1 - \beta_1)^2 (x_i - \bar x)^2 + \bar\epsilon^2 \\
&+ 2(\hat\epsilon_i(\hat\beta_1 - \beta_1)(x_i - \bar x) - \hat\epsilon_i\bar\epsilon - (\hat\beta_1 - \beta_1)(x_i - \bar x)\bar\epsilon) \qquad (2.14)
\end{aligned}
$$

. Από τις εκτιμητικές εξισώσεις στη (1.5) έχουμε πως

$$
\begin{aligned}
\sum \hat\epsilon_i &= 0 \\
\sum x_i \hat\epsilon_i &= 0 \\
\Longrightarrow \sum (x_i - \bar x)\hat\epsilon_i &= 0 \qquad (2.15)
\end{aligned}
$$

Από τις (2.14) και (2.15) παίρνουμε

$$
\begin{aligned}
\sum \epsilon_i^2 &= \sum \hat\epsilon_i^2 + (\hat\beta_1 - \beta_1)^2 \sum (x_i - \bar x)^2 + n\bar\epsilon^2 \Longrightarrow \\
\frac{\sum \epsilon_i^2}{\sigma^2} &= \frac{\sum \hat\epsilon_i^2}{\sigma^2} + \frac{(\hat\beta_1 - \beta_1)^2 S_{xx}}{\sigma^2} + \frac{n\bar\epsilon^2}{\sigma^2} \qquad (2.16)
\end{aligned}
$$

. Καθότι τα ϵ_i είναι ένα τυχαίο δείγμα από την $N(0,\sigma^2)$ κατανομή έχουμε

$$\begin{aligned} \frac{1}{\sigma^2}\sum \epsilon_i^2 &\sim \chi_n^2 \\ \bar{\epsilon} &\sim N(0,\frac{\sigma^2}{n}) \implies \frac{1}{\sigma^2}n\bar{\epsilon}^2 \sim \chi_1^2. \end{aligned} \tag{2.17}$$

. Από την (2.3) έχουμε

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2\frac{1}{S_{xx}}) \implies \frac{(\hat{\beta}_1-\beta_1)^2 S_{xx}}{\sigma^2} \sim \chi_1^2. \tag{2.18}$$

Τα αποτελέσματα στη (1.12) μαζί με την υπόθεση κανονικότητας συνεπάγονται πως

Α) $\frac{\sum \hat{\epsilon}_i^2}{\sigma^2} \perp\!\!\!\perp \frac{(\hat{\beta}_1-\beta_1)^2 S_{xx}}{\sigma^2}$

Β) $\frac{\sum \hat{\epsilon}_i^2}{\sigma^2} \perp\!\!\!\perp \frac{n\bar{\epsilon}^2}{\sigma^2}$ (καθότι $\bar{\epsilon} = \bar{Y} - \beta_0 - \beta_1\bar{x}$)

Γ) $\frac{(\hat{\beta}_1-\beta_1)^2 S_{xx}}{\sigma^2} \perp\!\!\!\perp \frac{n\bar{\epsilon}^2}{\sigma^2}$.

Από τις τρεις παραπάνω προτάσεις προκύπτει πως $\frac{(\hat{\beta}_1-\beta_1)^2 S_{xx}}{\sigma^2} + \frac{n\bar{\epsilon}^2}{\sigma^2} \sim \chi_2^2$ και χρησιμοποιώντας το Λήμμα παραπάνω παίρνουμε το πρώτο αποτέλεσμα του θεωρήματος. Το δεύτερο αποτέλεσμα προκύπτει από την (2.18) και το γεγονός πως η μηδενική υπόθεση λέει πως $\beta_1 = 0$. □

Από το παραπάνω αποτέλεσμα προκύπτει ο F έλεγχος της στατιστικής σημαντικότητας της παλινδρόμησης, δηλαδή της $H_0 : \beta_1 = 0$ έναντι της δίπλευρης εναλλακτικής. Καθότι $\frac{SSR}{\sigma^2} \sim^{H_0} \chi_1^2$, $\frac{SSE}{\sigma^2} \sim \chi_{n-2}^2$, και $SSR \perp\!\!\!\perp SSE$ έχουμε πως

$$\frac{MSR}{MSE} = \frac{SSR/1}{SSE/n-2} \sim^{H_0} F_{1,n-2}, \tag{2.19}$$

με αποτέλεσμα η $H_0 : \beta_1 = 0$ να απορρίπτεται σε επίπεδο σημαντικότητας α όταν η παρατηρηθείσα τιμή της σ.σ.ε. $F = \frac{MSR}{MSE} > F_{1,n-2;\alpha}$. Μπορεί να δειχθεί πως ο παραπάνω έλεγχος είναι ισοδύναμος του ελέγχου στην (2.8).

3 Πολλαπλή Γραμμική Παλινδρόμηση

Θεωρούμε το πολλαπλό γραμμικό μοντέλο (πολλαπλή γραμμική παλιδρόμηση) με $p-1$ συμμεταβλητές :

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_{p-1} x_{i,p-1} + \epsilon_i,\ i = 1,\ldots,n, \tag{3.1}$$

Y_i είναι η απόκριση της i-οστής παρατήρησης, $x_{i,k}$ η τιμή της k συμμεταβλητής της i-οστής παρατήρησης και ϵ_i είναι το σφάλμα (δηλαδή η απόκλιση της παρατήρησης Y_i από την μέση τιμή της . Θεωρούμε τα x ώς σταθερές τιμές κι όχι ως τ.μ.. Για τα σφάλματα υποθέτουμε πως είναι ομοσκεδαστικά ($var(\epsilon_i) = \sigma^2$) και ασυσχέτιστα ($cov(\epsilon_i, \epsilon_j) = 0$ για $i \neq j$).

Εναλλακτικά το μοντέλο μπορεί να γραφτεί ώς:

$$\begin{aligned} Y_i|x_{i1}, x_{i2}, \ldots, x_{i,p-1} &\sim (\mu_i, \sigma^2),\ cov(Y_i, Y_j|x_{i1}, x_{i2}, \ldots, x_{i,p-1}, x_{j1}, x_{j2}, \ldots, x_{j,p-1}) = 0,\ i \neq j \\ \mu_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_{p-1} x_{i,p-1},\ i = 1, \ldots, n, \end{aligned} \tag{3.2}$$

Μπορούμε να γράψουμε το μοντέλο με χρήση πινάκων και διανυσμάτων, ορίζοντας

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1,p-1} \\ 1 & x_{21} & x_{22} & \ldots & x_{2,p-1} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{n,p-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \\ \vdots \\ \boldsymbol{x}_n^T \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-1} \end{bmatrix}, \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Στις παραπάνω εκφράσεις το δίανυσμα

$$\boldsymbol{x}_i = \begin{bmatrix} 1 \\ x_{i1} \\ x_{i2} \\ \vdots \\ x_{i,p-1} \end{bmatrix}$$

είναι το διάνυσμα με τις τιμές των συμμεταβλητών για την i παρατήρηση.
Το μοντέλο για την i απόκριση γράφεται ως

$$Y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \epsilon_i,\ i = 1, \ldots, n, \tag{3.3}$$

ή εναλλακτικά

$$\begin{aligned} Y_i|\boldsymbol{x}_i &\sim (\mu_i, \sigma^2),\ cov(Y_i, Y_j|\boldsymbol{x}_i, \boldsymbol{x}_j) = 0,\ i \neq j \\ \mu_i &= \boldsymbol{x}_i^T \boldsymbol{\beta},\ i = 1, \ldots, n. \end{aligned} \tag{3.4}$$

Το μοντέλο για όλες τις παρατηρήσεις γράφεται ως

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon} &\sim (\boldsymbol{0}_n, \sigma^2 \boldsymbol{I}_n), \end{aligned} \tag{3.5}$$

όπου για τον $n \times p$ πίνακα συμμεταβλητών $\boldsymbol{X}$ ισχύει $rank(X) = p$, $\mathbf{0}_n = E(\boldsymbol{\epsilon})$ είναι ένα $n \times 1$ διάνυσμα με μηδενικά, $\boldsymbol{I}_n$ είναι ένας $n \times n$ ταυτοτικός πίνακας με μοναδες στην διαγώνιο και μηδενικά εκτος διαγωνίου. Καθότι υποθέτουμε ασυσχέτιστα και ομοσκεδαστικά σφάλματα ισχύει

$$Var(\boldsymbol{\epsilon}) = \begin{bmatrix} var(\epsilon_1) & cov(\epsilon_1,\epsilon_2) & \dots & cov(\epsilon_1,\epsilon_{n-1}) & cov(\epsilon_1,\epsilon_n) \\ cov(\epsilon_2,\epsilon_1) & var(\epsilon_2) & \dots & cov(\epsilon_2,\epsilon_{n-1}) & cov(\epsilon_2,\epsilon_n) \\ \vdots & \vdots & \dots & \vdots & \vdots \\ cov(\epsilon_{n-1},\epsilon_1) & cov(\epsilon_{n-1},\epsilon_2) & \dots & var(\epsilon_{n-1}) & cov(\epsilon_{n-1},\epsilon_n) \\ cov(\epsilon_n,\epsilon_1) & cov(\epsilon_n,\epsilon_2) & \dots & cov(\epsilon_n,\epsilon_{n-1}) & var(\epsilon_n) \end{bmatrix} = \sigma^2 \boldsymbol{I}_n.$$

Ο εκτιμητής ελαχίστων τετραγώνων του διανύσματος συντελεστών $\boldsymbol{\beta}$ ελαχιστοποιεί το άθροισμα των τετραγωνικών σφαλμάτων, ήτοι είναι οι λύσεις του μαθηματικού προβλήματος

$$\min_{\boldsymbol{\beta}} \sum \epsilon_i^2(\boldsymbol{\beta}) \Leftrightarrow \min_{\boldsymbol{\beta}} \boldsymbol{\epsilon}^T(\boldsymbol{\beta})\boldsymbol{\epsilon}(\boldsymbol{\beta}) \Leftrightarrow \min_{\boldsymbol{\beta}} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \tag{3.6}$$

Παραγωγίζοντας καταλήγουμε πως οι εκτιμητές ελαχίστων τετραγώνων προκύπτουν ως λύσεις των εκτιμητικών εξισώσεων

$$\boldsymbol{X}^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = \mathbf{0}_p \tag{3.7}$$

Οι εκτιμητής ελαχίστων τετραγώνων προκύπτει ως λύση του παραπάνω γραμμικού συστήματος p εξισώσεων:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}. \tag{3.8}$$

Προφανώς για τα υπόλοιπα $\hat{\boldsymbol{\epsilon}} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ της προσαρμογής ισχύει

$$\boldsymbol{X}^T\hat{\boldsymbol{\epsilon}} = \mathbf{0}_p, \tag{3.9}$$

δηλαδή τα υπόλοιπα υπόκεινται σε p γραμμικούς περιορισμούς (και άρα δεν μπορεί να είναι στατιστικά συσχέτιστα). Προκύπτει άμεσα πως

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T E(\boldsymbol{Y}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\beta} \\ Var(\hat{\boldsymbol{\beta}}) &= Var((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T Var(\boldsymbol{Y})\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1} \\ &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\sigma^2\boldsymbol{I}_n\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}. \end{aligned} \tag{3.10}$$

Στα παραπάνω αποτελέσματα καταλήξαμε χρησιμοποιώντας το γεγονός πως αν $\boldsymbol{U}$ είναι ένα $m \times 1$ τυχαίο διάνυσμα και $\boldsymbol{C}$ ένας $k \times m$ πίνακας σταθερών, τότε $\boldsymbol{CU}$ είναι ένα $k \times 1$ τυχαίο διάνυσμα με $E(\boldsymbol{CU}) = \boldsymbol{C}E(\boldsymbol{U})$ και $Var(\boldsymbol{CU}) = \boldsymbol{C}Var(\boldsymbol{U})\boldsymbol{C}^T$, ένας $k \times k$ πίνακας.

Ο $p \times p$ πίνακας διακυμάνσεων-συνδιακυμάνσεων του $\hat{\boldsymbol{\beta}}$

$$Var(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \begin{bmatrix} var(\hat{\beta}_0) & cov(\hat{\beta}_0,\hat{\beta}_1) & cov(\hat{\beta}_0,\hat{\beta}_2) & \dots & cov(\hat{\beta}_0,\hat{\beta}_{p-1}) \\ cov(\hat{\beta}_1,\hat{\beta}_0) & var(\hat{\beta}_1) & cov(\hat{\beta}_1,\hat{\beta}_2) & \dots & cov(\hat{\beta}_1,\hat{\beta}_{p-1}) \\ cov(\hat{\beta}_2,\hat{\beta}_0) & cov(\hat{\beta}_2,\hat{\beta}_1) & var(\hat{\beta}_2) & \dots & cov(\hat{\beta}_2,\hat{\beta}_{p-1}) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ cov(\hat{\beta}_{p-1},\hat{\beta}_0) & cov(\hat{\beta}_{p-1},\hat{\beta}_1) & cov(\hat{\beta}_{p-1},\hat{\beta}_2) & \dots & var(\hat{\beta}_{p-1}) \end{bmatrix}$$

έχει στην διαγώνιο τις διασπορές των εκτιμητριών ($var(\hat{\beta}_i) = \sigma^2_{\hat{\beta}_i}$) και εκτός διαγωνίου τις συνδιασπορές μεταξύ των εκτιμητριών ($cov(\hat{\beta}_i, \hat{\beta}_j) = \sigma_{\hat{\beta}_i,\hat{\beta}_j}$) .

Οι προσαρμοσμένες τιμές είναι

$$\begin{aligned} \hat{\boldsymbol{Y}} &= \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \\ &= \boldsymbol{H}\boldsymbol{Y}. \end{aligned} \tag{3.11}$$

Ο πίνακας $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ είναι συμμετρικός και ταυτοδύναμος, δηλαδή ισχύει $\boldsymbol{H}^2 = \boldsymbol{H}$ (δείξτε το) και λέγεται Hat Matrix. Ισχύει πως

$$\begin{aligned} \boldsymbol{H}\boldsymbol{X} &= \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{X} \\ (I_n - \boldsymbol{H})\boldsymbol{X} &= \boldsymbol{0} \end{aligned} \tag{3.12}$$

Τα υπόλοιπα είναι

$$\begin{aligned} \hat{\boldsymbol{\epsilon}} &= \boldsymbol{Y} - \hat{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \\ &= (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{Y} = (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{\epsilon} = (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{\epsilon} \end{aligned} \tag{3.13}$$

Ο πίνακας $(\boldsymbol{I}_n - \boldsymbol{H})$ είναι συμμετρικός και ταυτοδύναμος , δηλαδή ισχύει $(\boldsymbol{I}_n - \boldsymbol{H})^2 = (\boldsymbol{I}_n - \boldsymbol{H})$ (δείξτε το).

Μια αμερόληπτη εκτιμήτρια της διασποράς των σφαλμάτων σ^2 δίνεται από

$$\begin{aligned} s^2 &= \frac{1}{n-p}\sum(Y_i - \hat{Y}_i)^2 = \frac{1}{n-p}\sum\hat{\epsilon}_i^2 \\ &= \frac{1}{n-p}\hat{\boldsymbol{\epsilon}}^T\hat{\boldsymbol{\epsilon}} = \frac{1}{n-p}\boldsymbol{Y}^T(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{Y} \ (= \frac{1}{n-p}\boldsymbol{\epsilon}^T(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{\epsilon} \end{aligned} \tag{3.14}$$

Αντικαθιστώντας το σ^2 με την εκτιμήτρια s^2 στον $p \times p$ πίνακα διακυμάνσεων-συνδιακυμάνσεων του $\hat{\boldsymbol{\beta}}$ παίρνουμε

$$\hat{Var}(\hat{\boldsymbol{\beta}}) = s^2(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \begin{bmatrix} v\hat{a}r(\hat{\beta}_0) & c\hat{o}v(\hat{\beta}_0,\hat{\beta}_1) & c\hat{o}v(\hat{\beta}_0,\hat{\beta}_2) & \dots & c\hat{o}v(\hat{\beta}_0,\hat{\beta}_{p-1}) \\ c\hat{o}v(\hat{\beta}_1,\hat{\beta}_0) & v\hat{a}r(\hat{\beta}_1) & c\hat{o}v(\hat{\beta}_1,\hat{\beta}_2) & \dots & c\hat{o}v(\hat{\beta}_1,\hat{\beta}_{p-1}) \\ c\hat{o}v(\hat{\beta}_2,\hat{\beta}_0) & c\hat{o}v(\hat{\beta}_2,\hat{\beta}_1) & v\hat{a}r(\hat{\beta}_2) & \dots & c\hat{o}v(\hat{\beta}_2,\hat{\beta}_{p-1}) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ c\hat{o}v(\hat{\beta}_{p-1},\hat{\beta}_0) & c\hat{o}v(\hat{\beta}_{p-1},\hat{\beta}_1) & c\hat{o}v(\hat{\beta}_{p-1},\hat{\beta}_2) & \dots & v\hat{a}r(\hat{\beta}_{p-1}) \end{bmatrix}$$

, που είναι η εκτίμηση του πίνακα διακυμάνσεων-συνδιακυμάνσεων του $\hat{\boldsymbol{\beta}}$ κι έχει στην διαγώνιο τις εκτιμήσεις των διασπορών των εκτιμητριών ($v\hat{a}r(\hat{\beta}_i) = s^2_{\hat{\beta}_i}$) και εκτός διαγωνίου τις εκτιμήσεις των συνδιασπορών μεταξύ των εκτιμητριών $(c\hat{o}v(\hat{\beta}_i, \hat{\beta}_j) = s_{\hat{\beta}_i,\hat{\beta}_j})$.

3.1 Πολλαπλή Κανονική Γραμμική Παλινδρόμηση

Το μοντέλο για όλες τις παρατηρήσεις γράφεται ως

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{X\beta} + \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon} &\sim N(\boldsymbol{0}_n, \sigma^2 \boldsymbol{I}_n), \end{aligned} \tag{3.15}$$

όπου για τον $n \times p$ πίνακα συμμεταβλητών $\boldsymbol{X}$ ισχύει $rank(X) = p$, $\boldsymbol{0}_n = E(\boldsymbol{\epsilon})$ είναι ένα $n \times 1$ διάνυσμα με μηδενικά, $\boldsymbol{I}_n$ είναι ένας $n \times n$ ταυτοτικός πίνακας με μοναδες στην διαγώνιο και μηδενικά εκτος διαγωνίου.

Έχουμε πως

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, Var(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}), \tag{3.16}$$

από όπου προκύπτει πως κάθε γραμμικός συνδυασμός των εκτιμήσεων των συντελεστών στην παλινδρόμηση ακολουθεί κανονική κατανομή. Πιο συγκεκριμένα, αν πάρουμε ένα $p \times 1$ διάνυσμα $\boldsymbol{l} \neq \boldsymbol{0}_p$ (δηλαδή $\boldsymbol{l}^T = (l_0, l_1, l_2, \ldots, l_{p-1}) \neq (0, 0, 0, \ldots, 0)$)τότε

$$\boldsymbol{l}^T\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{l}^T\boldsymbol{\beta}, Var(\boldsymbol{l}^T\hat{\boldsymbol{\beta}}) = \boldsymbol{l}^T Var(\hat{\boldsymbol{\beta}})\boldsymbol{l} = \sigma^2\boldsymbol{l}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{l}), \tag{3.17}$$

από όπου προκύπτει πως

$$\frac{\boldsymbol{l}^T\hat{\boldsymbol{\beta}} - \boldsymbol{l}^T\boldsymbol{\beta}}{\sigma\sqrt{\boldsymbol{l}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{l}}} \sim N(0, 1), \tag{3.18}$$

Μπορεί να δειχθεί πως

$$\begin{aligned} (n-p)\frac{s^2}{\sigma^2} &\sim \chi^2_{n-p} \\ s^2 \perp\!\!\!\perp \hat{\boldsymbol{\beta}}. \end{aligned} \tag{3.19}$$

Άρα, εύκολα προκύπτει

$$\frac{\boldsymbol{l}^T\hat{\boldsymbol{\beta}} - \boldsymbol{l}^T\boldsymbol{\beta}}{s\sqrt{\boldsymbol{l}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{l}}} \sim t_{n-p}. \tag{3.20}$$

3.2 Συμπερασματολογία για παραμέτρους, Διάστημα εμπιστοσύνης, Διάστημα Πρόβλεψης

Από την (3.19) προκύπτει πως ένα $100(1-\alpha)\%$ Δ.Ε. για τον γραμμικό συνδυασμό $\boldsymbol{l}^T\boldsymbol{\beta}$ δίνεται από

$$\boldsymbol{l}^T\hat{\boldsymbol{\beta}} \pm t_{n-p;\frac{\alpha}{2}}\sqrt{\boldsymbol{l}^T\hat{Var}(\hat{\boldsymbol{\beta}})\boldsymbol{l}},$$

όπου $t_{n-p;\alpha/2}$ είναι τό <u>άνω</u> $\alpha/2$ ποσοστιαίο σημείο της t_{n-p} κατανομής και $\hat{Var}(\hat{\boldsymbol{\beta}}) = s^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.

Αντίστοιχα, ο έλεγχος της μηδενικής υπόθεση H_o : $\boldsymbol{l}^T\boldsymbol{\beta} = r$ έναντι της δίπλευρης εναλλακτικής απορρίπτει την H_o σε επίπεδο σημαντικότητας α όταν

$$\frac{|\boldsymbol{l}^T\hat{\boldsymbol{\beta}} - r|}{\sqrt{\boldsymbol{l}^T\hat{Var}(\hat{\boldsymbol{\beta}})\boldsymbol{l}}} > t_{n-p;\frac{\alpha}{2}},$$

Ως παράδειγμα, αν πάρουμε $\boldsymbol{l}$ νά έχει 1 στην $i+1$ θέση και 0 αλλού παίρνουμε ένα $100(1-\alpha)\%$ Δ.Ε. για τον συντελεστή β_i

$$\hat{\beta}_i \pm t_{n-p;\frac{\alpha}{2}} s_{\hat{\beta}_i} \tag{3.21}$$

Αν πάρουμε $\boldsymbol{l} = \boldsymbol{x}_0$, όπου

$$\boldsymbol{x}_0 = \begin{bmatrix} 1 \\ x_{01} \\ x_{02} \\ \vdots \\ x_{0,p-1} \end{bmatrix}$$

παίρνουμε ένα $100(1-\alpha)\%$ Δ.Ε. για την μέση απόκριση $\mu_0 = \boldsymbol{x}_0^T\boldsymbol{\beta} = \beta_0 + \beta_1 x_{01} + \beta_2 x_{02} + \ldots + \beta_{p-1}x_{0,p-1}$

$$\begin{aligned} \hat{Y}_0 \pm t_{n-p;\frac{\alpha}{2}} s_{\hat{Y}_0} &: \\ \boldsymbol{x}_0^T\hat{\boldsymbol{\beta}} \pm t_{n-p;\frac{\alpha}{2}}\sqrt{\boldsymbol{x}_0^T\hat{Var}(\hat{\boldsymbol{\beta}})\boldsymbol{x}_0} &: \\ \boldsymbol{x}_0^T\hat{\boldsymbol{\beta}} \pm t_{n-p;\frac{\alpha}{2}} s\sqrt{\boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0} & \end{aligned} \tag{3.22}$$

Χρησιμοποιώντας το γεγονός πως για μια καινούρια παρατήρηση Y_0 με τιμές συμμεταβλητών $(x_{01}, x_{02}, \ldots, x_{0,p-1})$ ισχύει $var(Y_0 - \hat{Y}_0) = \sigma^2 + \sigma^2_{\hat{Y}_0}$ καταλήγουμε σε ένα $100(1-\alpha)\%$ Διάστημα Πρόβλεψης της Y_0 :

$$\begin{aligned} \boldsymbol{x}_0^T\hat{\boldsymbol{\beta}} \pm t_{n-p;\frac{\alpha}{2}}\sqrt{s^2 + \boldsymbol{x}_0^T\hat{Var}(\hat{\boldsymbol{\beta}})\boldsymbol{x}_0} &: \\ \boldsymbol{x}_0^T\hat{\boldsymbol{\beta}} \pm t_{n-p;\frac{\alpha}{2}} s\sqrt{1 + \boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0} & \end{aligned} \tag{3.23}$$

Έστω πως θέλουμε να ελέγξουμε την γενική γραμμική υπόθεση,

$$\begin{aligned} H_0 &: \boldsymbol{R\beta} = \boldsymbol{r} \\ H_1 &: \boldsymbol{R\beta} \neq \boldsymbol{r}, \end{aligned} \tag{3.24}$$

όπου $\boldsymbol{R}$ είναι ένας $q \times p$ πίνακας με $rank(R) = q$, $q \leq p$, και $\mathbf{r}$ ένα $q \times 1$ διάνυσμα πραγματικών τιμών. Η σχέση $\boldsymbol{R\beta} = \boldsymbol{r}$ επιβάλλει q περιορισμούς στις παραμέτρους στο $\boldsymbol{\beta}$.

Για παράδειγμα αν $p = 4$ και

$$\boldsymbol{R} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{r} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

, τότε η $H_0 : \ \beta_1 = \beta_2 = \beta_3 = 0$, και ο έλεγχος στην (3.24) είναι ο έλεγχος στατιστικής σημαντικότητας της πολλαπλής παλινδρόμησης .

Αν

$$\boldsymbol{R} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{r} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

τότε η $H_0 : \ \beta_2 = \beta_3 = 0$, και ο έλεγχος στην (3.24) είναι συνηθίζεται να λέγεται έλεγχος στατιστικής σημαντικότητας των συμμεταβλητών x_2, x_3 σε ένα μοντέλο όπου ήδη υπάρχει (έχει περιληφθεί) η x_1.

Αν

$$\boldsymbol{R} = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{r} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

τότε η $H_0 : \ \beta_1 = \beta_2, \beta_3 = 0$, και ο έλεγχος στην (3.24) είναι ένας έλεγχος υπόθεσης αν μπορούμε να χρησιμοποιήσουμε το πιο περιορισμένο Reduced μοντέλο $\mu_i = \beta_0 + \beta_1(x_{i1} + x_{i2})$, έναντι του πλήρους Full μοντέλου, $\mu_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}$.

Πριν δώσουμε την στατιστική συνάρτηση ελέγχου για τον έλεγχο στην (3.24) θα αναφέρουμε ένα θεώρημα.

Θεώρημα 3. *Έστω πως το $k \times 1$ τυχαίο διάνυσμα, $\boldsymbol{Y}$, έχει πολυμεταβλητή κανονική κατανομή $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, όπου ο $k \times k$ πίνακας $\boldsymbol{\Sigma}$ είναι θετικά ορισμένος. Τότε*

$$(\boldsymbol{Y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{Y} - \boldsymbol{\mu}) \sim \chi^2_k.$$

Καθότι $\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r} \sim N(\boldsymbol{R\beta} - \boldsymbol{r}, \sigma^2 \boldsymbol{R}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{R}^T)$, απλή εφαρμογή του παραπάνω θεωρήματος μας δίνει το παρακάτω

$$\frac{1}{\sigma^2}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^T (\boldsymbol{R}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{R}^T)^{-1} (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}) \sim^{H_0} \chi^2_q,$$

που σε συνδυασμό με τα αποτελέσματα στη (3.19) μας δίνει πως

$$
\begin{aligned}
F &= \frac{\frac{(\boldsymbol{R}\hat{\boldsymbol{\beta}}-\boldsymbol{r})^T(\boldsymbol{R}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{R}^T)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}}-\boldsymbol{r})}{q}}{s^2} \\
&= \frac{\frac{(\boldsymbol{R}\hat{\boldsymbol{\beta}}-\boldsymbol{r})^T(\boldsymbol{R}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{R}^T)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}}-\boldsymbol{r})}{q}}{\frac{SSE_{Full}}{n-p}} \sim^{H_0} F_{q,n-p}. \qquad (3.25)
\end{aligned}
$$

Μπορεί να αποδειχθεί πως ισοδύναμα

$$
\begin{aligned}
F &= \frac{\frac{SSR_{Full}-SSR_{Red}}{q}}{\frac{SSE_{Full}}{n-p}} \\
&= \frac{\frac{SSE_{Red}-SSE_{Full}}{q}}{\frac{SSE_{Full}}{n-p}} \sim^{H_0} F_{q,n-p}. \qquad (3.26)
\end{aligned}
$$

Ως παράδειγμα θεωρήστε την πολλαπλή γραμμική παλινδρόμηση με $n = 100$, $p = 6$ και θέλουμε να ελέγξουμε την υπόθεση $H_0: \ \beta_3 = \beta_4 = \beta_5 = 0$. Η σ.σ.ε. τότε ισούται με

$$
F = \frac{\frac{SSR(X_3,X_4,X_5|X_1,X_2)}{3}}{\frac{SSE_{Full}}{100-6}}, \qquad (3.27)
$$

όπου $SSR(X_3, X_4, X_5|X_1, X_2) = SSR(X_3, X_4, X_5, X_1, X_2) - SSR(X_1, X_2)$ είναι το άθροισμα τετραγώνων που ¨εξηγούν' οι μεταβλητές X_3, X_4, X_5 πέρα των X_1, X_2, και μπορεί να υπολογιστεί με τα Sequential Sum of Squares:

$$
\begin{aligned}
SSR(X_3, X_4, X_5|X_1, X_2) &= SSR(X_3|X_1, X_2) + SSR(X_4|X_1, X_2, X_3) + SSR(X_5|X_1, X_2, X_3, X_4) \\
&= SSR(X_4|X_1, X_2) + SSR(X_3|X_1, X_2, X_4) + SSR(X_5|X_1, X_2, X_3, X_4) \\
&= SSR(X_5|X_1, X_2) + SSR(X_4|X_1, X_2, X_5) + SSR(X_3|X_1, X_2, X_4, X_5) \\
&= etc.. \qquad (3.28)
\end{aligned}
$$