<u>Πολλαπλή Γραμμική Παλινδρόμηση</u>

Θεωρώ εξαρτημένη μεταβλητή Y και $\mathrm{X}_1,...,\mathrm{X}_p$, $p \geq 2$, ανεξάρτητες μεταβλητές.

Το μοντέλο της πολλαπλής γραμμικής παλινδρόμησης έχει τη μορφή $\mathrm{Y} = \beta_0 + \beta_1 \cdot \mathrm{X}_1 + \beta_2 \cdot \mathrm{X}_2 + ... + \beta_p \cdot \mathrm{X}_p + \varepsilon$, όπου β_i, $i = 0,1,...,p$, άγνωστοι παράμετροι και ε, σφάλμα.

Τα δεδομένα που διαθέτουμε έχουν την παρακάτω μορφή:

$$
\begin{array}{ccccc}
\mathrm{Y}_1 & \mathrm{X}_{11} & \mathrm{X}_{12} & ... & \mathrm{X}_{1p} \\
\mathrm{Y}_2 & \mathrm{X}_{21} & \mathrm{X}_{22} & ... & \mathrm{X}_{2p} \\
. & . & . & & . \\
. & . & . & & . \\
. & . & . & & . \\
\mathrm{Y}_n & \mathrm{X}_{n1} & \mathrm{X}_{n2} & ... & \mathrm{X}_{np}
\end{array}
$$

Άρα $\mathrm{Y}_i = \beta_0 + \beta_1 \cdot \mathrm{X}_{i1} + \beta_2 \cdot \mathrm{X}_{i2} + ... + \beta_p \cdot \mathrm{X}_{ip} + \varepsilon_i$, $i = 1,2,...n$.

Για να εκτιμήσω τις άγνωστες παραμέτρους β_i, θεωρώ την ελαχιστοποίηση του αθροίσματος των τετραγώνων των σφαλμάτων,

$$\sum_{i=1}^{n} \varepsilon_i^{\,2} = \sum_{i=1}^{n} \left(\mathrm{Y}_i - \beta_0 - \beta_1 \cdot \mathrm{X}_{i1} - \beta_2 \cdot \mathrm{X}_{i2} - ... - \beta_p \cdot \mathrm{X}_{ip}\right)^2 = \Phi\left(\beta_0, \beta_1, ..., \beta_p\right).$$

Παραγωγίζοντας έχουμε:

$$\frac{\partial \Phi}{\partial \beta_0} = -2 \cdot \sum_{i=1}^{n} \left(\mathrm{Y}_i - \beta_0 - \beta_1 \cdot \mathrm{X}_{i1} - \beta_2 \cdot \mathrm{X}_{i2} - ... - \beta_p \cdot \mathrm{X}_{ip}\right) = 0 \Rightarrow$$
$$\Rightarrow \hat{\beta}_0 + \hat{\beta}_1 \cdot \overline{\mathrm{X}}_1 + \hat{\beta}_2 \cdot \overline{\mathrm{X}}_2 + ... + \hat{\beta}_p \cdot \overline{\mathrm{X}}_p = \overline{\mathrm{Y}} \quad (*)$$

$$\frac{\partial \Phi}{\partial \beta_1} = -2 \cdot \sum_{i=1}^{n} \mathrm{X}_{i1} \cdot \left(\mathrm{Y}_i - \beta_0 - \beta_1 \cdot \mathrm{X}_{i1} - \beta_2 \cdot \mathrm{X}_{i2} - ... - \beta_p \cdot \mathrm{X}_{ip}\right) = 0 \Rightarrow$$
$$\Rightarrow \hat{\beta}_0 \cdot \sum_{i=1}^{n} \mathrm{X}_{i1} + \hat{\beta}_1 \cdot \sum_{i=1}^{n} \mathrm{X}_{i1}^{\,2} + \hat{\beta}_2 \cdot \sum_{i=1}^{n} \mathrm{X}_{i1} \cdot \mathrm{X}_{i2} + ... + \hat{\beta}_p \cdot \sum_{i=1}^{n} \mathrm{X}_{i1} \cdot \mathrm{X}_{ip} = \sum_{i=1}^{n} \mathrm{X}_{i1} \cdot \mathrm{Y}_i$$

Γενικότερα, $\dfrac{\partial \Phi}{\partial \beta_j} = -2 \cdot \sum_{i=1}^{n} \mathrm{X}_{ij} \cdot \left(\mathrm{Y}_i - \beta_0 - \beta_1 \cdot \mathrm{X}_{i1} - \beta_2 \cdot \mathrm{X}_{i2} - ... - \beta_p \cdot \mathrm{X}_{ip}\right) = 0 \Rightarrow$

$$\Rightarrow \hat{\beta}_0 \cdot \sum_{i=1}^{n} \mathrm{X}_{ij} + \hat{\beta}_1 \cdot \sum_{i=1}^{n} \mathrm{X}_{ij} \cdot \mathrm{X}_{i1} + ... + \hat{\beta}_j \cdot \sum_{i=1}^{n} \mathrm{X}_{ij}^{\,2} + ... + \hat{\beta}_p \cdot \sum_{i=1}^{n} \mathrm{X}_{ij} \cdot \mathrm{X}_{ip} = \sum_{i=1}^{n} \mathrm{X}_{ij} \cdot \mathrm{Y}_i \quad (**)$$

Από $(*)$, $(**)$ έχουμε σύστημα με $(p+1)$ αγνώστους και αν υπάρχει η λύση του, τη συμβολίζω με $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_p$, οπότε $\hat{\mathrm{Y}} = \hat{\beta}_0 + \hat{\beta}_1 \cdot \mathrm{X}_1 + \hat{\beta}_2 \cdot \mathrm{X}_2 + ... + \hat{\beta}_p \cdot \mathrm{X}_p$.

Τα υπόλοιπα ορίζονται από την $e = \mathrm{Y} - \hat{\mathrm{Y}}$.

Το πολλαπλό μοντέλο γραμμικής παλινδρόμησης, μπορεί να εκφραστεί υπό μορφή υπό μορφή πινάκων, με τον εξής τρόπο:

$$\underset{n\times 1}{\begin{pmatrix} \mathrm{Y}_1 \\ \mathrm{Y}_2 \\ . \\ . \\ . \\ \mathrm{Y}_n \end{pmatrix}} = \underset{n\times(p+1)}{\begin{bmatrix} 1 & \mathrm{X}_{11} & \mathrm{X}_{12} & . & . & . & \mathrm{X}_{1p} \\ 1 & \mathrm{X}_{21} & \mathrm{X}_{22} & . & . & . & \mathrm{X}_{2p} \\ . & . & . & & & & . \\ . & . & . & & & & . \\ . & . & . & & & & . \\ 1 & \mathrm{X}_{n1} & \mathrm{X}_{n2} & . & . & . & \mathrm{X}_{np} \end{bmatrix}} \cdot \underset{(p+1)\times 1}{\begin{pmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_p \end{pmatrix}} + \underset{n\times 1}{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{pmatrix}} \Rightarrow$$

$$\Rightarrow \underline{\mathrm{Y}} = \mathbf{X}\cdot\underline{\beta} + \underline{\varepsilon}$$

Οπότε, $\sum_{i=1}^{n}\varepsilon_i^{\,2} = \underline{\varepsilon}^{\mathrm{T}}\cdot\underline{\varepsilon} = \left(\underline{\mathrm{Y}} - \mathbf{X}\cdot\underline{\beta}\right)^{\mathrm{T}}\cdot\left(\underline{\mathrm{Y}} - \mathbf{X}\cdot\underline{\beta}\right) = \left(\underline{\mathrm{Y}}^{\mathrm{T}} - \underline{\beta}^{\mathrm{T}}\cdot\mathbf{X}^{\mathrm{T}}\right)\cdot\left(\underline{\mathrm{Y}} - \mathbf{X}\cdot\underline{\beta}\right) =$

$= \underline{\mathrm{Y}}^{\mathrm{T}}\cdot\underline{\mathrm{Y}} - \underline{\beta}^{\mathrm{T}}\cdot\mathbf{X}^{\mathrm{T}}\cdot\underline{\mathrm{Y}} - \underline{\mathrm{Y}}^{\mathrm{T}}\cdot\mathbf{X}\cdot\underline{\beta} + \underline{\beta}^{\mathrm{T}}\cdot\mathbf{X}^{\mathrm{T}}\cdot\mathbf{X}\cdot\underline{\beta}$.

Αλλά η διάσταση του $\underline{\beta}^{\mathrm{T}}\cdot\mathbf{X}^{\mathrm{T}}\cdot\underline{\mathrm{Y}}$ είναι 1, οπότε ισχύει ότι:

$\underline{\beta}^{\mathrm{T}}\cdot\mathbf{X}^{\mathrm{T}}\cdot\underline{\mathrm{Y}} = \left(\underline{\beta}^{\mathrm{T}}\cdot\mathbf{X}^{\mathrm{T}}\cdot\underline{\mathrm{Y}}\right)^{\mathrm{T}} = \underline{\mathrm{Y}}^{\mathrm{T}}\cdot\mathbf{X}\cdot\underline{\beta}$

Συνεπώς, $\sum_{i=1}^{n}\varepsilon_i^{\,2} = \underline{\mathrm{Y}}^{\mathrm{T}}\cdot\underline{\mathrm{Y}} - 2\cdot\underline{\beta}^{\mathrm{T}}\cdot\mathbf{X}^{\mathrm{T}}\cdot\underline{\mathrm{Y}} + \underline{\beta}^{\mathrm{T}}\cdot\mathbf{X}^{\mathrm{T}}\cdot\mathbf{X}\cdot\underline{\beta}$.

<u>Σημείωση:</u>
Ισχύουν τα παρακάτω:

(α) $\dfrac{\partial \mathbf{A}(\mathrm{x})\cdot\mathbf{B}(\mathrm{x})}{\partial \mathrm{X}} = \mathbf{B}(\mathrm{x})\cdot\dfrac{\partial \mathbf{A}(\mathrm{x})}{\partial \mathrm{X}} + \mathbf{A}(\mathrm{x})\cdot\dfrac{\partial \mathbf{B}(\mathrm{x})}{\partial \mathrm{X}}$

(β) $\dfrac{\partial tr\left[\mathbf{A}(\mathrm{x})\right]}{\partial \mathrm{X}} = tr\left[\dfrac{\partial \mathbf{A}(\mathrm{x})}{\partial \mathrm{X}}\right]$

(γ) $\dfrac{\partial \underline{\mathrm{X}}^{\mathrm{T}}\cdot\underline{\alpha}}{\partial \mathrm{X}} = \underline{\alpha}$

(δ) $\dfrac{\partial \underline{\mathrm{X}}^{\mathrm{T}}\cdot\mathbf{A}\cdot\underline{\mathrm{X}}}{\partial \mathrm{X}} = 2\cdot\mathbf{A}\cdot\underline{\mathrm{X}}$

Οπότε, παραγωγίζοντας την $\sum_{i=1}^{n}\varepsilon_i^{\,2} = \underline{\mathrm{Y}}^{\mathrm{T}}\cdot\underline{\mathrm{Y}} - 2\cdot\underline{\beta}^{\mathrm{T}}\cdot\mathbf{X}^{\mathrm{T}}\cdot\underline{\mathrm{Y}} + \underline{\beta}^{\mathrm{T}}\cdot\mathbf{X}^{\mathrm{T}}\cdot\mathbf{X}\cdot\underline{\beta}$,

έχουμε:

$$\frac{\partial}{\partial\underline{\beta}}\left(\underline{\mathrm{Y}}^{\mathrm{T}}\cdot\underline{\mathrm{Y}} - 2\cdot\underline{\beta}^{\mathrm{T}}\cdot\mathbf{X}^{\mathrm{T}}\cdot\underline{\mathrm{Y}} + \underline{\beta}^{\mathrm{T}}\cdot\mathbf{X}^{\mathrm{T}}\cdot\mathbf{X}\cdot\underline{\beta}\right) = -2\cdot\mathbf{X}^{\mathrm{T}}\cdot\underline{\mathrm{Y}} + 2\cdot\mathbf{X}^{\mathrm{T}}\cdot\mathbf{X}\cdot\underline{\beta} = 0 \Rightarrow$$

$$\Rightarrow \mathbf{X}^{\mathrm{T}}\cdot\mathbf{X}\cdot\underline{\beta} = \mathbf{X}^{\mathrm{T}}\cdot\underline{\mathrm{Y}} \Rightarrow \underline{\hat{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\cdot\mathbf{X}\right)^{-1}\cdot\mathbf{X}^{\mathrm{T}}\cdot\underline{\mathrm{Y}} \qquad \text{(κανονικές εξισώσεις)}$$

Η τελευταία συνεπαγωγή ισχύει αν και μόνο αν υπάρχει ο $\left(\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X}\right)^{-1}$. Για να υπάρχει ο αντίστροφος, αρκεί $r(\mathbf{X}) = p+1$.

Από την παραπάνω σχέση παρατηρούμε ότι $\underline{\hat{\beta}}$ είναι γραμμικός συνδυασμός των παρατηρήσεων.

Υπόλοιπα:

$$\underline{e} = \begin{pmatrix} \underline{e}_1 \\ \underline{e}_2 \\ \cdot \\ \cdot \\ \cdot \\ \underline{e}_n \end{pmatrix} = \underline{\mathrm{Y}} - \underline{\hat{\mathrm{Y}}} = \underline{\mathrm{Y}} - \mathbf{X} \cdot \underline{\hat{\beta}} = \underline{\mathrm{Y}} - \mathbf{X} \cdot \left(\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X}\right)^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \underline{\mathrm{Y}} = (\mathbf{I} - \mathbf{P}) \cdot \underline{\mathrm{Y}},$$

όπου $\mathbf{P} = \mathbf{X} \cdot \left(\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X}\right)^{-1} \cdot \mathbf{X}^{\mathrm{T}}$.

Ο πίνακας $\mathbf{P}$ είναι $(n \times n)$, συμμετρικός και ταυτοδύναμος.

$$\mathbf{P}^{\mathrm{T}} = \mathbf{P}$$

$$\mathbf{P}^2 = \left[\mathbf{X} \cdot \left(\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X}\right)^{-1} \cdot \mathbf{X}^{\mathrm{T}}\right] \cdot \left[\mathbf{X} \cdot \left(\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X}\right)^{-1} \cdot \mathbf{X}^{\mathrm{T}}\right] = \mathbf{X} \cdot \left(\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X}\right)^{-1} \cdot \mathbf{X}^{\mathrm{T}} = \mathbf{P}$$

Με αυτές τις ιδιότητες, ο $\mathbf{P}$ ονομάζεται πίνακας προβολής.

Σημείωση:

Λόγω των πιο πάνω ιδιοτήτων του $\mathbf{P}$, ισχύει $(\mathbf{I} - \mathbf{P})^{\mathrm{T}} = (\mathbf{I} - \mathbf{P})$, $(\mathbf{I} - \mathbf{P})^2 = (\mathbf{I} - \mathbf{P})$.

Ισχύει επίσης (και αποδυκνύεται με τον ίδιο τρόπο, όπως και στην απλή γραμμική παλινδρόμηση) ότι:

$$\begin{array}{ccccc} \sum_{i=1}^{n} \left(\mathrm{Y}_i - \overline{\mathrm{Y}}\right)^2 & = & \sum_{i=1}^{n} \left(\hat{\mathrm{Y}}_i - \overline{\mathrm{Y}}\right)^2 & + & \sum_{i=1}^{n} \left(\mathrm{Y}_i - \hat{\mathrm{Y}}_i\right)^2 \\ SS_{tot} & = & SS_{reg} & + & SS_{res} \\ n-1 & = & p & + & n-(p+1) \end{array} \qquad \text{(βαθμοί ελευθερίας)}$$

Ο συντελεστής προσδιορισμού είναι $R^2 = \dfrac{SS_{reg}}{SS_{tot}}$.

Στατιστικές Ιδιότητες
Εκτιμήτριας Ελαχίστων Τετραγώνων

Υποθέσεις για τα σφάλματα:

(1) $\mathrm{E}(\varepsilon_i) = 0$, $i = 1,2,\ldots,n$

(2) $Cov(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$, $Var(\varepsilon_i) = \sigma^2$, $i = 1,2,\ldots,n$

(3) $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$, $i = 1,2,\ldots,n$

Υπό μορφή πινάκων, έχω τις ισοδύναμες συνθήκες αν ορίσω $\underline{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^{\mathrm{T}}$:

(1)' $\mathrm{E}(\underline{\varepsilon}) = \underline{0}$

(2)' $Cov(\underline{\varepsilon}) = \sigma^2 \cdot \mathbf{I}$

(3)' $\underline{\varepsilon} \sim \mathrm{N}_n(\underline{0}, \sigma^2 \cdot \mathbf{I})$

Επειδή $\underline{\mathrm{Y}} = \mathbf{X} \cdot \underline{\beta} + \underline{\varepsilon}$, οι παραπάνω συθνήκες δίνουν:

(1)'' $\mathrm{E}(\underline{\mathrm{Y}}) = \mathbf{X} \cdot \underline{\beta}$

(2)'' $Cov(\underline{\mathrm{Y}}) = \sigma^2 \cdot \mathbf{I}$

(3)'' $\underline{\mathrm{Y}} \sim \mathrm{N}_n(\mathbf{X} \cdot \underline{\beta}, \sigma^2 \cdot \mathbf{I})$

Πρόταση:

Έστω $\underline{\hat{\beta}}$ η εκτιμήτρια ελαχίστων τετραγώνων για το πολλαπλό μοντέλο παλινδρόμησης $\underline{\mathrm{Y}} = \mathbf{X} \cdot \underline{\beta} + \underline{\varepsilon}$.

(α) Όταν ισχύουν οι (1)'' και (2)'', έχω ότι $\mathrm{E}(\underline{\hat{\beta}}) = \underline{\beta}$ και $Cov(\underline{\hat{\beta}}) = \sigma^2 \cdot (\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X})^{-1}$.

(β) Αν ισχύει επιπλέον η (3)'', τότε $\underline{\hat{\beta}} \sim \mathrm{N}_{p+1}\left(\underline{\beta}, \sigma^2 \cdot (\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X})^{-1}\right)$.

Απόδειξη:

(α) Έχω ότι $\underline{\hat{\beta}} = (\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \underline{\mathrm{Y}}$. Άρα:

$$\mathrm{E}(\underline{\hat{\beta}}) = \mathrm{E}\left[(\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \underline{\mathrm{Y}}\right] = (\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \mathrm{E}(\underline{\mathrm{Y}}) = (\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \mathbf{X} \cdot \underline{\beta} = \underline{\beta}$$

$$Cov(\underline{\hat{\beta}}) = Cov\left[(\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \underline{\mathrm{Y}}\right] = (\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot Cov(\underline{\mathrm{Y}}) \cdot \left[(\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^{\mathrm{T}}\right]^{\mathrm{T}} =$$

$$= (\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \sigma^2 \cdot \mathbf{I} \cdot \mathbf{X} \cdot (\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X})^{-1} = \sigma^2 \cdot (\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \mathbf{X} \cdot (\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X})^{-1} =$$

$$= \sigma^2 \cdot (\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X})^{-1}$$

(β) Αφού $\underline{\hat{\beta}}$ είναι γραμμικός μετασχηματισμός του $\underline{\mathrm{Y}}$, έχω ότι:

$$\underline{\hat{\beta}} \sim \mathrm{N}_{p+1}\left(\underline{\beta}, \sigma^2 \cdot (\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X})^{-1}\right)$$

Σημείωση:

Γενικά, αν $\underline{X} \sim N_p(\underline{\mu}, \Sigma) \Rightarrow E(\underline{X}^T \cdot A \cdot \underline{X}) = \underline{\mu}^T \cdot A \cdot \underline{\mu} + tr(A \cdot \Sigma)$, αφού $\underline{X}^T \cdot A \cdot \underline{X}$ είναι μονοδιάστατος αριθμός και άρα $\underline{X}^T \cdot A \cdot \underline{X} = tr(\underline{X}^T \cdot A \cdot \underline{X}) \Rightarrow$

$$\Rightarrow E(\underline{X}^T \cdot A \cdot \underline{X}) = E[tr(\underline{X}^T \cdot A \cdot \underline{X})] = E[tr(A \cdot \underline{X} \cdot \underline{X}^T)] = tr[E(A \cdot \underline{X} \cdot \underline{X}^T)] =$$
$$= tr(A \cdot \underline{\mu} \cdot \underline{\mu}^T + A \cdot \Sigma) = tr(A \cdot \underline{\mu} \cdot \underline{\mu}^T) + tr(A \cdot \Sigma) = tr(\underline{\mu}^T \cdot A \cdot \underline{\mu}) + tr(A \cdot \Sigma) =$$
$$= \underline{\mu}^T \cdot A \cdot \underline{\mu} + tr(A \cdot \Sigma)$$

Πρόταση:

Έστω $\hat{\sigma}^2 = \frac{1}{n-(p+1)} \cdot \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$. Τότε $E(\hat{\sigma}^2) = \sigma^2$.

Απόδειξη:

$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} e_i^2 = \underline{e}^T \cdot \underline{e} = [(I-P) \cdot \underline{Y}]^T \cdot [(I-P) \cdot \underline{Y}] = \underline{Y}^T \cdot (I-P)^T \cdot (I-P) \cdot \underline{Y} =$$
$$= \underline{Y}^T \cdot (I-P) \cdot (I-P) \cdot \underline{Y} = \underline{Y}^T \cdot (I-P)^2 \cdot \underline{Y} = \underline{Y}^T \cdot (I-P) \cdot \underline{Y}$$

Άρα, $E(\underline{Y}^T \cdot (I-P) \cdot \underline{Y}) = E^T(\underline{Y}) \cdot (I-P) \cdot E(\underline{Y}) + tr[(I-P) \cdot Cov(\underline{Y})] =$

$$= (X \cdot \underline{\beta})^T \cdot (I-P) \cdot (X \cdot \underline{\beta}) + tr[(I-P) \cdot \sigma^2 \cdot I] = \underline{\beta}^T \cdot X^T \cdot (I-P) \cdot X \cdot \underline{\beta} + \sigma^2 \cdot tr(I-P) =$$
$$= 0 + \sigma^2 \cdot tr(I-P) = \sigma^2 \cdot tr(I-P)$$

Αλλά, $tr(I) = n$ και

$$tr(P) = tr\left[X \cdot (X^T \cdot X)^{-1} \cdot X^T\right] = tr\left[(X^T \cdot X)^{-1} \cdot X^T \cdot X\right] = tr\left(I_{(p+1)\times(p+1)}\right) = p+1.$$

Άρα, $E\left[\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2\right] = \sigma^2 \cdot [n-(p+1)] \Rightarrow \hat{\sigma}^2$ είναι αμερόληπτη για σ^2.

Θεώρημα:

Αν $\underline{X} \sim N(\underline{0}, \Sigma)$, $\Sigma > 0$ και A συμμετρικός, τότε:

$\underline{X}^T \cdot A \cdot \underline{X} \sim X_r^2 \Leftrightarrow A \cdot \Sigma$ ταυτοδύναμος και $r(A \cdot \Sigma) = r$

Απόδειξη:

Παραλείπεται.

Πρόταση:

Έστω το μοντέλο πολλαπλής γραμμικής παλινδρόμησης $\underline{Y} = \mathbf{X} \cdot \underline{\beta} + \underline{\varepsilon}$.

Όταν ισχύει η ιδιότητα (3)'', ισχύουν:

(α) $\underline{\hat{\beta}}$ και $\hat{\sigma}^2$ είναι ανεξάρτητα

(β) $\dfrac{SS_{res}}{\sigma^2} \sim X^2_{n-(p+1)}$

(γ) $\underline{\hat{\beta}}$ και $\hat{\sigma}^2$ είναι επαρκείς και πλήρεις στατιστικές συναρτήσεις για $\underline{\beta}$ και σ^2 αντίστοιχα

(δ) $\underline{\hat{\beta}}$ και $\hat{\sigma}^2$ είναι ΑΟΕΔ των $\underline{\beta}$ και σ^2 αντίστοιχα

(ε) SS_{reg} και SS_{res} είναι ανεξάρτητα

Απόδειξη:

(α) Έχω ότι $\underline{\hat{\beta}} = \left(\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X}\right)^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \underline{Y} = \mathbf{A} \cdot \underline{Y}$, όπου $\mathbf{A} = \left(\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X}\right)^{-1} \cdot \mathbf{X}^{\mathrm{T}}$ και $\hat{\sigma}^2 = \dfrac{SS_{res}}{n-p-1} = \dfrac{\underline{e}^{\mathrm{T}} \cdot \underline{e}}{n-p-1} = \dfrac{\left[(\mathbf{I}-\mathbf{P}) \cdot \underline{Y}\right]^{\mathrm{T}} \cdot \left[(\mathbf{I}-\mathbf{P}) \cdot \underline{Y}\right]}{n-p-1}$.

Αν δείξω ότι $\underline{\hat{\beta}}$ και $(\mathbf{I}-\mathbf{P}) \cdot \underline{Y}$ είναι ασυσχέτιστα, τότε εξασφαλίζω και την ανεξαρτησία, λόγω κανονικότητας.

$$Cov\left(\mathbf{A} \cdot \underline{Y}, (\mathbf{I}-\mathbf{P}) \cdot \underline{Y}\right) = \mathbf{A} \cdot Cov(\underline{Y}, \underline{Y}) \cdot (\mathbf{I}-\mathbf{P})^{\mathrm{T}} = \mathbf{A} \cdot Cov(\underline{Y}) \cdot (\mathbf{I}-\mathbf{P})^{\mathrm{T}} =$$
$$= \mathbf{A} \cdot \sigma^2 \cdot \mathbf{I} \cdot (\mathbf{I}-\mathbf{P})^{\mathrm{T}} = \sigma^2 \cdot \mathbf{A} \cdot (\mathbf{I}-\mathbf{P}) = 0$$

(β) Έχω ότι $\underline{e} = (\mathbf{I}-\mathbf{P}) \cdot \underline{Y}$.

Άρα $\mathrm{E}(\underline{e}) = (\mathbf{I}-\mathbf{P}) \cdot \mathrm{E}(\underline{Y}) = (\mathbf{I}-\mathbf{P}) \cdot \mathbf{X} \cdot \underline{\beta} = \mathbf{X} \cdot \underline{\beta} - \mathbf{P} \cdot \mathbf{X} \cdot \underline{\beta} = \mathbf{X} \cdot \underline{\beta} - \mathbf{X} \cdot \underline{\beta} = \underline{0}$ και

$$Cov(\underline{e}) = (\mathbf{I}-\mathbf{P}) \cdot Cov(\underline{Y}) \cdot (\mathbf{I}-\mathbf{P})^{\mathrm{T}} = (\mathbf{I}-\mathbf{P}) \cdot \sigma^2 \cdot \mathbf{I} \cdot (\mathbf{I}-\mathbf{P}) = \sigma^2 \cdot (\mathbf{I}-\mathbf{P})^2 = \sigma^2 \cdot (\mathbf{I}-\mathbf{P})$$

Άρα, $\underline{e} \sim \mathrm{N}_n\left(\underline{0}, \sigma^2 \cdot (\mathbf{I}-\mathbf{P})\right)$.

Αλλά, $\dfrac{SS_{res}}{\sigma^2} = \dfrac{\underline{e}^{\mathrm{T}} \cdot \underline{e}}{\sigma^2} = \underline{e}^{\mathrm{T}} \cdot \left(\dfrac{\mathbf{I}}{\sigma^2}\right) \cdot \underline{e}$ και $\left(\dfrac{\mathbf{I}}{\sigma^2}\right) \cdot \sigma^2 \cdot (\mathbf{I}-\mathbf{P}) = (\mathbf{I}-\mathbf{P})$ ταυτοδύναμος με βαθμίδα $n-(p+1)$. Άρα $\dfrac{\underline{e}^{\mathrm{T}} \cdot \underline{e}}{\sigma^2} \sim X^2_{n-(p+1)}$.

(γ) Χωρίς απόδειξη.

(δ) Προκύπτει από το (γ).

(ε) Το εκτιμώμενο μοντέλο είναι $\hat{\mathrm{Y}}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot \mathrm{X}_{i1} + \hat{\beta}_2 \cdot \mathrm{X}_{i2} + \ldots + \hat{\beta}_p \cdot \mathrm{X}_{ip}$, το οποίο, λόγω της εξίσωσης $\overline{\mathrm{Y}} = \hat{\beta}_0 + \hat{\beta}_1 \cdot \overline{\mathrm{X}}_1 + \hat{\beta}_2 \cdot \overline{\mathrm{X}}_2 + \ldots + \hat{\beta}_p \cdot \overline{\mathrm{X}}_p$, μπορεί να γραφτεί σαν $\hat{\mathrm{Y}}_i = \hat{\beta}_1 \cdot \left(\mathrm{X}_{i1} - \overline{\mathrm{X}}_1\right) + \hat{\beta}_2 \cdot \left(\mathrm{X}_{i2} - \overline{\mathrm{X}}_2\right) + \ldots + \hat{\beta}_p \cdot \left(\mathrm{X}_{ip} - \overline{\mathrm{X}}_p\right) + \overline{\mathrm{Y}}$.

Άρα,

$$SS_{reg} = \sum_{i=1}^{n} \left(\hat{Y}_i - \overline{Y}\right)^2 = \sum_{i=1}^{n} \left[\hat{\beta}_1 \cdot \left(X_{i1} - \overline{X}_1\right) + \hat{\beta}_2 \cdot \left(X_{i2} - \overline{X}_2\right) + \ldots + \hat{\beta}_p \cdot \left(X_{ip} - \overline{X}_p\right) + \overline{Y}\right]^2$$

και συνεπώς είναι συνάρτηση των $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$, οπότε από το (α) προκύπτει ότι $SS_{res} \perp SS_{reg}$.

<u>Έλεγχοι υποθέσεων για τα β_i (καθένα ξεχωριστά):</u>

Έστω ο έλεγχος $H_0 : \beta_i = b_i$ (b_i δεδομένη τιμή) προς $H_1 : \beta_i \neq b_i$.

Επειδή $\underline{\hat{\beta}} \sim N_{p+1}\left(\underline{\beta}, \sigma^2 \cdot \left(\mathbf{X}^T \cdot \mathbf{X}\right)^{-1}\right)$, έχω ότι η i - συνιστώσα του διανύσματος είναι κανονική με μέση τιμή β_i και διακύμανση $\sigma^2 \cdot c_{i+1,i+1}$, $i = 0,1,\ldots,p$, όπου $\left(\mathbf{X}^T \cdot \mathbf{X}\right)^{-1} = \left(c_{i+1,j+1}\right)$, $i = 0,1,\ldots,p$, $j = 0,1,\ldots,p$.

Άρα, $\hat{\beta}_i \sim N\left(\beta_i, \sigma^2 \cdot c_{i+1,i+1}\right)$.

Επομένως, $t_i = \dfrac{\hat{\beta}_i - \beta_i}{\hat{\sigma} \cdot \sqrt{c_{i+1,i+1}}} = \dfrac{\dfrac{\hat{\beta}_i - \beta_i}{\sigma \cdot \sqrt{c_{i+1,i+1}}}}{\sqrt{\dfrac{\hat{\sigma}^2}{\sigma^2}}} \sim t_{n-(p+1)}$.

Άρα για τον έλεγχο της υπόθεσης $H_0 : \beta_i = b_i$ προς $H_1 : \beta_i \neq b_i$, σε επίπεδο σημαντικότητας α, απορρίπτω την H_0 αν $|t_i| > t_{n-(p+1),\alpha/2}$.

<u>Έλεγχος ύπαρξης παλινδρόμησης:</u>

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0$ προς $H_1 : \beta_\kappa \neq 0$, για κάποιο $\kappa \in \{1,2,\ldots,p\}$

Θεωρώ την ελεγχοσυνάρτηση $F = \dfrac{MS_{reg}}{MS_{res}} \sim F_{p,n-(p+1)}$ και απορρίπτω την H_0, σε επίπεδο σημαντικότητας α, αν $F > F_{p,n-(p+1),\alpha}$.

ΠΙΝΑΚΑΣ ΑΝΑΔΙΑ

Πηγή Μεταβλητότητας	Άθροισμα Τετραγώνων	Βαθμοί Ελευθερίας	Μέσα Τετράγωνα	F - test
Παλινδρόμηση	SS_{reg}	p	$MS_{reg} = \frac{SS_{reg}}{p}$	$F = \frac{MS_{reg}}{MS_{res}}$
Σφάλμα	SS_{res}	$n-(p+1)$	$MS_{res} = \frac{SS_{res}}{n-(p+1)}$	
Ολική Μεταβλητότητα	SS_{tot}	$n-1$		

Επαναληπτικές Μετρήσεις:

Οι επαναληπτικές μετρήσεις, στην πολλαπλή γραμμική παλινδρόμηση, χρησιμεύουν ως εξής:

(α) Παρέχουν μέθοδο ελέγχου της ορθότητας του μοντέλου.

(β) Παρέχουν εκτιμήτρια για την σ^2, η οποία δεν εξαρτάται από το μοντέλο.

Έλεγχοι υποθέσεων για περισσότερα από ένα β_i συγχρόνως:

(Μέθοδος του επιπλέον αθροίσματος τετραγώνων – extra sum of squares)

Υποθέτω μοντέλο της μορφής $\mathrm{Y}_i = \beta_0 + \beta_1 \cdot \mathrm{X}_{i1} + \beta_2 \cdot \mathrm{X}_{i2} + \ldots + \beta_p \cdot \mathrm{X}_{ip} + \varepsilon_i$.

Εκτελούμε έλεγχο της υπόθεσης $\mathrm{H}_0 : \beta_\kappa = \beta_{\kappa+1} = \ldots = \beta_p = 0$, $\kappa \le p$ προς $\mathrm{H}_1 : \beta_\nu \ne 0$, για κάποιο $\nu \in \{\kappa, \kappa+1, \ldots, p\}$.

Η μέθοδος αυτή αποτελείται απο τα εξής βήματα:

(1) Όταν ισχύει η H_0, το μοντέλο είναι της μορφής

$\mathrm{Y}_i = \beta_0 + \beta_1 \cdot \mathrm{X}_{i1} + \beta_2 \cdot \mathrm{X}_{i2} + \ldots + \beta_{\kappa-1} \cdot \mathrm{X}_{i(\kappa-1)} + \varepsilon_i$.

Εκτιμούμε τις παραμέτρους για αυτό το μοντέλο και καλουμε τις εκτιμήτριες $\hat{\beta}_0^*, \hat{\beta}_1^*, \ldots, \hat{\beta}_{\kappa-1}^*$.

Το αντίστοιχο εκτιμώμενο μοντέλο είναι

$\hat{\mathrm{Y}}_i^* = \hat{\beta}_0^* + \hat{\beta}_1^* \cdot \mathrm{X}_{i1} + \hat{\beta}_2^* \cdot \mathrm{X}_{i2} + \ldots + \hat{\beta}_{\kappa-1}^* \cdot \mathrm{X}_{i(\kappa-1)}$.

Μπορώ να υπολογίσω το $SS_{res}(\mathrm{H_0 M}) = \sum_{i=1}^{n} \left(\mathrm{Y}_i - \hat{\mathrm{Y}}_i^*\right)^2$.

Ισχύει ότι $\frac{SS_{res}(\mathrm{H_0 M})}{\sigma^2} \sim X^2_{n-\kappa}$.

(2) Όταν δεν ισχύει η H_0, έχουμε το πλήρες μοντέλο, δηλαδή

$\mathrm{Y}_i = \beta_0 + \beta_1 \cdot \mathrm{X}_{i1} + \beta_2 \cdot \mathrm{X}_{i2} + \ldots + \beta_p \cdot \mathrm{X}_{ip} + \varepsilon_i$.

Το εκτιμώμενο πλήρες μοντέλο είναι $\hat{\mathrm{Y}}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot \mathrm{X}_{i1} + \hat{\beta}_2 \cdot \mathrm{X}_{i2} + \ldots + \hat{\beta}_p \cdot \mathrm{X}_{ip}$.

Εκτιμούμε τις παραμέτρους με τις εκτιμήτριες $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_p$.

Οπότε, $SS_{res}(\Pi\mathrm{M}) = \sum_{i=1}^{n} (\mathrm{Y}_i - \hat{\mathrm{Y}}_i)^2$ και $\frac{SS_{res}(\Pi\mathrm{M})}{\sigma^2} \sim X^2_{n-(p+1)}$.

(3) Ισχύει ότι $SS_{res}(\mathrm{H}_0\mathrm{M}) \geq SS_{res}(\Pi\mathrm{M})$.

Όσο αυξάνεται ο αριθμός των παραμέτρων στο μοντέλο, τόσο μειώνεται το άθροισμα μετραγώνων των υπολοίπων.

Άρα αν η μείωση αυτή δεν είναι σημαντική, τότε δεν πρέπει να εισάγουμε περισσότερους όρους.

Άρα απορρίπτω την H_0, αν η διαφορά είναι σημαντική.

Με προϋπόθεση την ανεξαρτησία, έχουμε ότι $\frac{SS_{res}(\mathrm{H}_0\mathrm{M}) - SS_{res}(\Pi\mathrm{M})}{\sigma^2} \sim X^2_{p+1-\kappa}$.

Επειδή όμως το σ^2 είναι άγνωστο, έχω $\hat{\sigma}^2 = \frac{SS_{res}(\Pi\mathrm{M})}{n-(p+1)}$.

Άρα $F = \frac{\frac{SS_{res}(\mathrm{H}_0\mathrm{M}) - SS_{res}(\Pi\mathrm{M})}{p+1-\kappa}}{\frac{SS_{res}(\Pi\mathrm{M})}{n-(p+1)}} \sim F_{p+1-\kappa, n-(p+1)}$.

Απορρίπτω την H_0, σε επίπεδο σημαντικότητας α, αν $F > F_{p+1-\kappa, n-(p+1), \alpha}$.

Η διαφορά $SS_{res}(\mathrm{H}_0\mathrm{M}) - SS_{res}(\Pi\mathrm{M})$ λέγεται επιπλέον άθροισμα τετραγώνων.

<u>Παρατηρήσεις:</u>

(1) Όταν $\mathrm{H}_0 : \beta_i = 0$, τότε το F - test που προκύπτει ονομάζεται μετικό και ισχύει ότι $t_i^{\,2} = F_i$.

(2) Όταν $\mathrm{H}_0 : \beta_1 = \beta_2 = ... = \beta_p = 0$, τότε το F - test συμπίπτει με τον έλεγχο του πίνακα ΑΝΑΔΙΑ.