1 Πίνακες

1.1 Φασματική ανάλυση ενός πραγματικού συμμετρικού πίνακα

Θεώρημα 1. *Έστω $\boldsymbol{A}$ ένας πραγματικός $n \times n$ συμμετρικός πίνακας. Ισχύει πως $\boldsymbol{A} = \boldsymbol{P\Lambda P}^T$, όπου $\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ είναι ένας $n \times n$ διαγώνιος πίνακας που περιέχει τις (πραγματικές) ιδιοτιμές του πίνακα $\boldsymbol{A}$, ενώ $\boldsymbol{P} = [\boldsymbol{p}_1\ \boldsymbol{p}_2\ \ldots \boldsymbol{p}_n]$ είναι ένας $n \times n$ ορθοκανονικός πίνακας (δηλαδή $\boldsymbol{PP}^T = \boldsymbol{P}^T\boldsymbol{P} = \boldsymbol{I}_n$) που περιέχει τα n (πραγματικά) ιδιοδιανύσματα του $\boldsymbol{A}$.*

Θεώρημα 2. *Έστω $\boldsymbol{\Sigma}$ ένας $n \times n$ θετικά ορισμένος συμμετρικός πίνακας (positive definite, pd) . Ισχύει πως $\lambda_i > 0$ για κάθε $i = 1, 2, \ldots, n$.*

Απόδειξη. Καθότι $\boldsymbol{\Sigma}$ είναι pd ισχύει εξ ορισμού πως $\forall \boldsymbol{a} \neq \boldsymbol{0}_n$ έχουμε $\boldsymbol{a}^T\boldsymbol{\Sigma a} > 0$. Επίσης έχουμε $\boldsymbol{\Sigma} = \boldsymbol{P\Lambda P}^T \implies \boldsymbol{P}^T\boldsymbol{\Sigma P} = \boldsymbol{\Lambda}$.

Έστω $\boldsymbol{b} \neq \boldsymbol{0}$. Έχουμε $\boldsymbol{b}^T\boldsymbol{\Lambda b} = \boldsymbol{b}^T\boldsymbol{P}^T\boldsymbol{\Sigma P b} = (\boldsymbol{Pb})^T\boldsymbol{\Sigma}(\boldsymbol{Pb}) = \boldsymbol{a}^T\boldsymbol{\Sigma a} > 0$, όπου $\boldsymbol{a} = \boldsymbol{Pb} \neq \boldsymbol{0}$ (καθότι $\boldsymbol{P}$ έχει τάξη n). Παίρνοντας $\boldsymbol{b}^T = [1\ 0 \ldots 0]$ έχουμε $\lambda_1 > 0$. Παίρνοντας $\boldsymbol{b}^T = [0\ 1 \ldots 0]$ έχουμε $\lambda_2 > 0$. κ.ο.κ.. □

Θεώρημα 3. *Έστω $\boldsymbol{\Sigma}$ ένας $n \times n$ θετικά ορισμένος συμμετρικός πίνακας. Ισχύει πως υπάρχει $n \times n$ (pd) συμμετρικός πίνακας $\boldsymbol{\Sigma}^{1/2}$, ώστε $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}$. Ο πίνακας $\boldsymbol{\Sigma}^{1/2}$ λέγεται τετραγωνική ρίζα του $\boldsymbol{\Sigma}$*

Απόδειξη. $\boldsymbol{\Sigma} = \boldsymbol{P\Lambda P}^T = \boldsymbol{P\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{P}^T = \boldsymbol{P\Lambda}^{1/2}\boldsymbol{P}^T\boldsymbol{P\Lambda}^{1/2}\boldsymbol{P}^T = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}$, όπου $\boldsymbol{\Sigma}^{1/2} = \boldsymbol{P\Lambda}^{1/2}\boldsymbol{P}^T$ είναι $n \times n$ pd πίνακας. □

Θεώρημα 4. *Έστω $\boldsymbol{A}$ ένας συμμετρικός $n \times n$ ταυτοδύναμος πίνακας. Ο $\boldsymbol{A}$ έχει ιδιοτιμές ίσες με 0 ή 1. Ο αριθμός των ιδιοτιμών που ισούνται με 1 είναι ίσος με $tr(\boldsymbol{A}) = rank(\boldsymbol{A})$.*

Απόδειξη. $\boldsymbol{A}^2 = \boldsymbol{A} \implies \boldsymbol{P\Lambda P}^T\boldsymbol{P\Lambda P}^T = \boldsymbol{P\Lambda P}^T \implies \boldsymbol{P\Lambda}^2\boldsymbol{P}^T = \boldsymbol{P\Lambda P}^T \implies \boldsymbol{\Lambda}^2 = \boldsymbol{\Lambda} \implies \lambda_i = 0$ ή $1, i = 1, \ldots, n$. Ο αριθμός των μονάδων ισούται (προφανώς) με το άθροισμα των στοιχείων της διαγωνίου του $\boldsymbol{\Lambda}$, δηλαδή ισούται με το $tr(\boldsymbol{\Lambda}) = rank(\boldsymbol{\Lambda})$. Επίσης ισχύει $tr(\boldsymbol{A}) = tr(\boldsymbol{P\Lambda P}^T) = tr(\boldsymbol{P}^T\boldsymbol{P\Lambda}) = tr(\boldsymbol{\Lambda})$ και $rank(\boldsymbol{A}) = rank(\boldsymbol{\Lambda})$. □

2 Εφαρμογές Στατιστική

Λέμε πως το τυχαίο $n \times 1$ διάνυσμα $\boldsymbol{Y}$ ακολουθεί την πολυμεταβλητή κανονική κατανομή με μέσο $\boldsymbol{\mu}$ και πίνακα διασπορών συνδιασπορών $\boldsymbol{\Sigma}$, $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ αν η σ.π.π. του $\boldsymbol{Y}$ είναι:

$$f_{\boldsymbol{Y}}(y_1, y_2, \dots, y_n) = f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}}$$

Στο μάθημα αυτό θα υποθέσουμε πως $\boldsymbol{\Sigma}$ είναι $n \times n$ (pd) πίνακας. Ισχύουν τα ακόλουθα.

1. Η ροπογεννήτρια του $\boldsymbol{Y}$ είναι
   $M_{\boldsymbol{Y}}(\boldsymbol{t}) = E(e^{\boldsymbol{t}^T \boldsymbol{y}}) = e^{\boldsymbol{t}^T \boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^T \boldsymbol{\Sigma} \boldsymbol{t}}$.
2. Έστω $\boldsymbol{K}$ ένας σταθερός $k \times n$ πίνακας. Τότε $\boldsymbol{K}\boldsymbol{Y} \sim N(\boldsymbol{K}\boldsymbol{\mu}, \boldsymbol{K}\boldsymbol{\Sigma}\boldsymbol{K}^T)$.
3. Γράφοντας
   $$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{bmatrix} \sim N(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix})$$
   Έχουμε $\boldsymbol{Y}_1 \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$.
4. Έχουμε
   $\boldsymbol{Y}_1 | \boldsymbol{Y}_2 = \boldsymbol{y}_2 \sim N(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{y}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$

Απλό παράδειγμα:

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim N(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 & 2 \\ 2 & 9 \end{bmatrix})$$

Έχουμε $|\boldsymbol{\Sigma}| = 32$ και

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{32} \begin{bmatrix} 9 & -2 \\ -2 & 4 \end{bmatrix}$$

$$-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) = -\frac{1}{2}\frac{1}{32} \begin{bmatrix} (y_1 - 1) & (y_2 - 2) \end{bmatrix} \begin{bmatrix} 9 & -2 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} (y_1 - 1) \\ (y_2 - 2) \end{bmatrix}$$

Άρα $-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) = -\frac{1}{64}\{9(y_1-1)^2-4(y_1-1)(y_2-2)+4(y_2-2)^2\}$

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{e^{-\frac{9}{64}(y_1-1)^2+\frac{4}{64}(y_1-1)(y_2-2)-\frac{4}{64}(y_2-2)^2}}{(2\pi)\sqrt{32}}$$

Η τιμή της σ.π.π. στο σημείο $(y_1,y_2)=(2,3)$ ισούται με

$$\begin{aligned} f_{Y_1,Y_2}(2,3) &= \frac{e^{-\frac{9}{64}(2-1)^2+\frac{4}{64}(2-1)(3-2)-\frac{4}{64}(3-2)^2}}{(2\pi)\sqrt{32}} \\ &= 0.024444 \end{aligned}$$

Επίσης για τον υπολογισμό της ροπογεννήτριας έχουμε

$$\boldsymbol{t}^T\boldsymbol{\mu} = t_1\mu_1+t_2\mu_2 = t_1+2t_2, \;\; \boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t} = \begin{bmatrix} t_1 & t_2 \end{bmatrix}\begin{bmatrix} 4 & 2 \\ 2 & 9 \end{bmatrix}\begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = 4t_1^2+9t_2^2+4t_1t_2$$

Η ροπογεννήτρια ισούται με $M_{\boldsymbol{Y}}(\boldsymbol{t}) = M_{Y_1,Y_2}(t_1,t_2) = e^{t_1+2t_2+\frac{1}{2}(4t_1^2+9t_2^2+4t_1t_2)}$

$Y_1|Y_2=y_2 \sim N(1+\frac{2}{9}(y_2-2), 4-\frac{4}{9})$, δηλαδή
$Y_1|Y_2=y_2 \sim N(1+\frac{2}{9}(y_2-2), \frac{32}{9})$.
$Y_1|Y_2=3 \sim N(\frac{11}{9}, \frac{32}{9})$.

$$f_{Y_1|Y_2=3}(2) = \frac{1}{\sqrt{\frac{32}{9}}\sqrt{2\pi}}e^{-\frac{1}{2\frac{32}{9}}(2-\frac{11}{9})^2} = 0.1943$$

Θεώρημα 5. *Έστω $\boldsymbol{Y} \sim (\boldsymbol{0}_n, \boldsymbol{I}_n)$, δηλαδή $E(\boldsymbol{Y}) = \boldsymbol{0}_n$ και $Var(\boldsymbol{Y}) = \boldsymbol{I}_n$. Έστω $\boldsymbol{P}$ ένας $n\times n$ ορθογώνιος πίνακας (δηλαδή $\boldsymbol{P}\boldsymbol{P}^T = \boldsymbol{P}^T\boldsymbol{P} = \boldsymbol{I}_n$) . Τότε ισχύει πως $\boldsymbol{Z} = \boldsymbol{P}\boldsymbol{Y} \sim (\boldsymbol{0}_n, \boldsymbol{I}_n)$*

Απόδειξη. $E(\boldsymbol{Z}) = E(\boldsymbol{P}\boldsymbol{Y}) = \boldsymbol{P}E(\boldsymbol{Y}) = \boldsymbol{0}_n$, $Var(\boldsymbol{Z}) = Var(\boldsymbol{P}\boldsymbol{Y}) = \boldsymbol{P}Var(\boldsymbol{Y})\boldsymbol{P}^T = \boldsymbol{P}\boldsymbol{P}^T = \boldsymbol{I}_n$, □

Ένα επακόλουθο του παραπάνω είναι πως $\boldsymbol{Y} \sim N(\boldsymbol{0}_n, \boldsymbol{I}_n) \implies \boldsymbol{Z} = \boldsymbol{P}\boldsymbol{Y} \sim N(\boldsymbol{0}_n, \boldsymbol{I}_n)$.

Θεώρημα 6. *Έστω $\boldsymbol{Y} \sim N(\boldsymbol{0}_n, \boldsymbol{\Sigma})$, όπου $\boldsymbol{\Sigma}$ είναι $n\times n$ θετικά ορισμένος πίνακας. Τότε ισχύει πως $\boldsymbol{Y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} \sim \chi^2_n$.*

Απόδειξη. $\boldsymbol{Y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} = \boldsymbol{Y}^T\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y} = \boldsymbol{Z}^T\boldsymbol{Z}$, με $\boldsymbol{Z} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y}$, όπου $\boldsymbol{\Sigma}^{-1/2} = (\boldsymbol{\Sigma}^{1/2})^{-1} = \boldsymbol{P}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{P}^T$. Καθότι $E(\boldsymbol{Z}) = \boldsymbol{0}, Var(\boldsymbol{Z}) = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{I}$, έχουμε πως $\boldsymbol{Z} \sim N(\boldsymbol{0}_n, \boldsymbol{I}_n)$, κι άρα $\boldsymbol{Y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} = \boldsymbol{Z}^T\boldsymbol{Z} = \sum Z_i^2 \sim \chi_n^2$. □

Θεώρημα 7. *Έστω $\boldsymbol{Y} \sim N(\boldsymbol{0}_n, \boldsymbol{I}_n)$ και A είναι $n \times n$ ταυτοδύναμος πίνακας. Τότε ισχύει πως $\boldsymbol{Y}^T A\boldsymbol{Y} \sim \chi^2_{rank(\boldsymbol{A})}$.*

Απόδειξη. $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T$. Καθότι $\boldsymbol{A}$ είναι ταυτοδύναμος και χωρίς βλάβη τις γενικότητας υποθέτουμε πως τα πρώτα $rank(\boldsymbol{A})$ στοιχεία της διαγωνίου του $\boldsymbol{\Lambda}$ είναι μονάδες με τα υπόλοιπα 0.Έχουμε $\boldsymbol{Y}^T\boldsymbol{A}\boldsymbol{Y} = \boldsymbol{Y}^T\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T\boldsymbol{Y} = \boldsymbol{Z}^T\boldsymbol{\Lambda}\boldsymbol{Z} = \sum_{i=1}^{rank(\boldsymbol{A})} Z_i^2 \sim \chi^2_{rank(\boldsymbol{A})}$. □

Θεώρημα 8. *Έστω $\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I})$, με $\boldsymbol{X}$ έναν $n\times p$ πίνακα με $rank(X) = p$. Οι εκτιμητές μεγίστης πιθανοφάνειας των παραμέτρων είναι*

1. *$\hat{\boldsymbol{\beta}}_{MLE} = \hat{\boldsymbol{\beta}}_{LS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ (ο εκτιμητής ελαχίστων τετραγώνων).*
2. *$\hat{\sigma}^2 = \frac{(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{LS})^T(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{LS})}{n} = \frac{1}{n}\boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y}$, όπου $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ είναι ο γνωστός hat matrix.*
3. *Η μέγιστη τιμή του λογάριθμου της πιθανοφάνειας ισούται με*

$$-\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\hat{\sigma}^2) - \frac{n}{2}$$

Απόδειξη. Εδώ έχουμε $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}$, $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$. Η πιθανοφάνεια $L(\boldsymbol{\beta}, \sigma^2)$ ισούται με

$$L(\boldsymbol{\beta}, \sigma^2) = \frac{e^{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})}}{(2\pi)^{n/2}|\sigma^2\boldsymbol{I}|^{1/2}} = \frac{e^{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})}}{(2\pi)^{n/2}(\sigma^2)^{n/2}}$$

Ο λογάριθμος της πιθανοφάνειας ισούται με

$$l(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) \tag{2.1}$$

Η μεγιστοποίηση της (2.1) είναι εύκολη. Παίρνουμε την παράγωγο ως προς σ^2 και την θέτουμε ίση με το 0.

$$\begin{aligned}\frac{\partial l(\boldsymbol{\beta},\sigma^2)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2}+\frac{1}{2(\sigma^2)^2}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta})=0\\ \Longrightarrow \tilde{\sigma}^2 &= \frac{(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta})}{n}\end{aligned}$$

Η δεύτερη παράγωγος της (2.1) ως προς σ^2 ισούται με

$$\begin{aligned}\frac{\partial^2 l(\boldsymbol{\beta},\sigma^2)}{\partial (\sigma^2)^2}\|_{\tilde{\sigma}^2} &= +\frac{n}{2(\sigma^2)^2}\|_{\tilde{\sigma}^2}-\frac{1}{(\sigma^2)^3}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta})\|_{\tilde{\sigma}^2}\\ &= -\frac{n}{2(\tilde{\sigma}^2)^2}<0\end{aligned}$$

και άρα $\tilde{\sigma}^2$ μεγιστοποιεί την πιθανοφάνεια για κάθε τιμή του $\boldsymbol{\beta}$.

Όσον αφορά την μεγιστοποίηση της (2.1) ως προς $\boldsymbol{\beta}$, αρκεί να ελαχιστοποιήσουμε το άθροισμα των τετραγωνικών σφαλμάτων, δηλαδή έχουμε το πρόβλημα ελαχιστοποίησης:

$$\min_{\boldsymbol{\beta}}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta}) \tag{2.2}$$

Η λύση είναι εξ ορισμού ο εκτιμητής ελαχίστων τετραγώνων $\hat{\boldsymbol{\beta}}_{LS}$. Αμέσως συνεπάγεται πως ο εκτιμητής μεγίστης πιθανοφάνειας της διασποράς ισούται με $\hat{\sigma}^2=\frac{(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{LS})^T(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{LS})}{n}$.

Για να βρούμε την μορφή του $\hat{\boldsymbol{\beta}}_{LS}$, δηλαδή την λύση του προβλήματος (2.2), παραγωγίζουμε την ποσότητα $(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta})$ ως προς $\boldsymbol{\beta}$:

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\beta}}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta}) &= \frac{\partial}{\partial\boldsymbol{\beta}}(-\boldsymbol{y}^T\boldsymbol{X\beta}-\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{y}+\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X\beta})\\ &= \frac{\partial}{\partial\boldsymbol{\beta}}(-2\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{y}+\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X\beta})\\ &= -2\boldsymbol{X}^T\boldsymbol{y}+2\boldsymbol{X}^T\boldsymbol{X\beta}.\end{aligned}$$

Η εκτιμητική εξίσωση που προκύπτει είναι

$$\boldsymbol{X}^T\boldsymbol{X\beta}=\boldsymbol{X}^T\boldsymbol{y}\implies\hat{\boldsymbol{\beta}}=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

Η δεύτερη παράγωγος είναι

$$\begin{aligned}\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta}) &= \frac{\partial}{\partial\boldsymbol{\beta}^T}(-2\boldsymbol{X}^T\boldsymbol{y}+2\boldsymbol{X}^T\boldsymbol{X\beta})\\ &= 2(\boldsymbol{X}^T\boldsymbol{X})^T=2(\boldsymbol{X}^T\boldsymbol{X})\end{aligned}$$

Καθότι $rank(\boldsymbol{X}) = p$ ο πίνακας $2(\boldsymbol{X}^T\boldsymbol{X})$ είναι θετικά ορισμένος και άρα το $\hat{\boldsymbol{\beta}}$ δίνει τοπικό ελάχιστο. Σε συνδυασμό με το γεγονός πως $(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \to \infty$, όταν $\boldsymbol{\beta} \to \pm\infty$ έχουμε πως $(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$ ελαχιστοποιείται στο $\hat{\boldsymbol{\beta}}$.

Τέλος, Η μέγιστη τιμη του λογαριθμου της πιθανοφάνειας προκύπτει θέτοντας τις τιμές των εκτιμητών μεγίστης πιθανοφάνεις στη (2.1), δηλαδή έχουμε

$$\begin{aligned}
\max_{\boldsymbol{\beta},\sigma^2} l(\boldsymbol{\beta}, \sigma^2) &= -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \\
&= -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\hat{\sigma}^2) - \frac{n}{2}
\end{aligned}$$

□

Θεώρημα 9. *Έστω $\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I})$, με $\boldsymbol{X}$ έναν $n \times p$ πίνακα με $rank(X) = p$. Για τους εκτιμητές μεγίστης πιθανοφάνειας των παραμέτρων ισχύει:*

1. $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1})$
2. $n\frac{\hat{\sigma}^2}{\sigma^2} = (n-p)\frac{s^2}{\sigma^2} \sim \chi^2_{n-p}$, *όπου* $s^2 = \frac{n}{n-p}\hat{\sigma}^2$ *είναι ο αμερόληπτος εκτιμητής της διασποράς.*
3. $\hat{\boldsymbol{\beta}}$ *και* s^2 *είναι ανεξάρτητα.*
4. *Αν* $\boldsymbol{R}$ *είναι ένας* $q \times p$ *πίνακας με* $rank(\boldsymbol{R}) = q \leq p$, *τότε*
$$\frac{(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{R}\boldsymbol{\beta})^T(\boldsymbol{R}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{R}^T)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{R}\boldsymbol{\beta})}{qs^2} \sim F_{q,n-p}$$

Απόδειξη.

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

Άρα χρησιμοποιώντας την ιδιότητα (2.) της Κανονικής κατανομής παραπάνω έχουμε $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1})$,

Έχουμε $(n-p)\frac{s^2}{\sigma^2} = \frac{1}{\sigma^2}\boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y} = \frac{1}{\sigma^2}\boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{H})(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$, καθότι $(\boldsymbol{I} - \boldsymbol{H})$ είναι ταυτοδύναμος με $rank(\boldsymbol{I} - \boldsymbol{H}) = tr(\boldsymbol{I} - \boldsymbol{H}) = n -$

$tr(\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T) = n - tr(\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}) = n - tr(\boldsymbol{I}_p) = n - p$. Παρατηρούμε πως $(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y} = (\boldsymbol{I}-\boldsymbol{H})(\boldsymbol{Y}-\boldsymbol{X\beta})$ καθότι $(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{X} = \boldsymbol{0}$. Άρα

$$\begin{aligned}(n-p)\frac{s^2}{\sigma^2} &= \frac{1}{\sigma^2}(\boldsymbol{Y}-\boldsymbol{X\beta})^T(\boldsymbol{I}-\boldsymbol{H})(\boldsymbol{I}-\boldsymbol{H})(\boldsymbol{Y}-\boldsymbol{X\beta}) \\ &= (\frac{1}{\sigma}(\boldsymbol{Y}-\boldsymbol{X\beta}))^T(\boldsymbol{I}-\boldsymbol{H})(\frac{1}{\sigma}(\boldsymbol{Y}-\boldsymbol{X\beta})) \\ &= \boldsymbol{Z}^T(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Z}, \ \ (\text{όπου } \boldsymbol{Z} \sim N(\boldsymbol{0},\boldsymbol{I})) \\ &\sim \chi^2_{n-p}.\end{aligned}$$

Έχουμε

$$\begin{aligned}Cov((\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y},\hat{\boldsymbol{\beta}}) &= Cov((\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y},(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}) \\ &= (\boldsymbol{I}-\boldsymbol{H})Var(\boldsymbol{Y})((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T)^T \\ &= \sigma^2(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \boldsymbol{0},\end{aligned}$$

και, λόγω κανονικότητας των $(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y},\hat{\boldsymbol{\beta}}$ συμπεραίνουμε πως $(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y},\hat{\boldsymbol{\beta}}$ είναι ανεξάρτητα. Τέλος καθότι s^2 είναι συνάρτηση του $(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y}$, ππροκύπτει πως $\hat{\boldsymbol{\beta}}$ και s^2 είναι ανεξάρτητα.

Έχουμε $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta},\sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}) \implies \boldsymbol{R}\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{R\beta},\sigma^2\boldsymbol{R}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{R}^T) \implies (\boldsymbol{R}\hat{\boldsymbol{\beta}}-\boldsymbol{R\beta}) \sim N(\boldsymbol{0},\sigma^2\boldsymbol{R}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{R}^T)$. Από το Θεώρημα 6 προκύπτει πως $\frac{1}{\sigma^2}(\boldsymbol{R}\hat{\boldsymbol{\beta}}-\boldsymbol{R\beta})^T(\boldsymbol{R}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{R}^T)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}}-\boldsymbol{R\beta}) \sim \chi^2_q$. Εξετάζουμε το κλάσμα

$$\frac{\frac{1}{\sigma^2}(\boldsymbol{R}\hat{\boldsymbol{\beta}}-\boldsymbol{R\beta})^T(\boldsymbol{R}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{R}^T)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}}-\boldsymbol{R\beta})/q}{(n-p)\frac{s^2}{\sigma^2}/(n-p)}.$$

Ο αριθμητής είναι ανεξάρτητος του παρανομαστή (καθότι $\hat{\boldsymbol{\beta}}$ και s^2 είναι ανεξάρτητα) και από τον ορισμό της F κατανομής προκύπτει το ζητούμενο. □

Το τελευταίο αποτέλεσμα αποτελεί την βάση για τον έλεγχο της γενικής γραμμικής υπόθεσης $\boldsymbol{R\beta} = \boldsymbol{r}$, όπου $\boldsymbol{R}$ είναι $q \times p$ πίνακας και ισχύει $rank(\boldsymbol{R}) = q$.

Εφαρμογή: Θεωρήστε μια έρευνα που επιδιώκει να εξετάσει την αποτελεσματικότητα μιας διαφημιστικής εκστρατείας όσον αφορά τις πωλήσεις ενός προϊόντος σε καταστήματα τροφίμων. Στην έρευνα επιλέχθηκε ένα απλό τυχαίο δείγμα από $n = 100$ καταστήματα τροφίμων της πρωτεύουσας. Η μεταβλητές της έρευνας ήταν:

1. Σχετική αύξηση στον όγκο πωλήσεων του προϊόντος: (Πωλήσεις ένα μήνα μετά την εκστρατεία - Πωλήσεις ένα μήνα πριν την εκστρατεία)/ Πωλήσεις ένα μήνα πριν την εκστρατεία (Y).

2. Μέσο οικογενειακό εισόδημα της περιοχής που ανήκει το κατάστημα (x)

3. Μέγεθος του καταστήματος: μικρό, μεσαίο, μεγάλο. (δ_1, δ_2 είναι οι δείκτες για μεσαίο και μεγάλο κατάστημα αντιστοιχα)

Το γραμμικό μοντέλο που χρησιμοποιήθηκε είναι:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_i + \beta_2 \delta_{1i} + \beta_3 \delta_{2i} + \beta_4 \delta_{1i} x_i + \beta_5 \delta_{2i} x_i + \epsilon_i \\ \epsilon_1, \ldots, \epsilon_{100} &\sim_{iid} N(0, \sigma^2) \end{aligned}$$

Βάσει του παραπάνω μοντέλου έχω 3 μοντέλα, ένα για κάθε μέγεθος καταστήματος.

- Για τα μικρά καταστήματα: $\delta_{1i} = 0$ και $\delta_{2i} = 0$ κι έχουμε

  $$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i.$$

  Στα μικρά καταστήματα, η επίδραση του μέσου οικογενειακού εισοδήματος στην μέση αύξηση των πωλήσεων εκφράζεται από τη παράμετρο β_1.

- Για τα μεσαία καταστήματα: $\delta_{1i} = 1$ και $\delta_{2i} = 0$ κι έχουμε

  $$\begin{aligned} Y_i &= \beta_0 + \beta_2 + \beta_1 x_i + \beta_4 x_i + \epsilon_i \\ &= (\beta_0 + \beta_2) + (\beta_1 + \beta_4) x_i + \epsilon_i. \end{aligned}$$

  Στα μεσαία καταστήματα, η επίδραση του μέσου οικογενειακού εισοδήματος στην μέση αύξηση των πωλήσεων είναι $\beta_1 + \beta_4$.

- Για τα μεγάλα καταστήματα: $\delta_{1i} = 0$ και $\delta_{2i} = 1$ κι έχουμε

  $$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_i + \beta_3 + \beta_5 x_i + \epsilon_i \\ &= (\beta_0 + \beta_3) + (\beta_1 + \beta_5) x_i + \epsilon_i. \end{aligned}$$

  Στα μεγάλα καταστήματα, η επίδραση του μέσου οικογενειακού εισοδήματος στην μέση αύξηση των πωλήσεων είναι $\beta_1 + \beta_5$.

Γράψτε τις παρακάτω μηδενικές υποθέσεις στην μορφή $H_0 : \boldsymbol{R\beta} = \boldsymbol{r}$, όπου $\boldsymbol{R}$ έχει q γραμμές και ισχύει $rank(\boldsymbol{R}) = q$. Επίσης γράψτε την σ.σ.ε και την κατανομή της κάτω από την μηδενική υπόθεση.

1. H_0 : η εκστρατεία δεν ήταν αποτελεματική.

   Απάντηση: Η μηδενική υπόθεση αναφέρει πως η μέση σχετική αύξηση στον όγκο πωλήσεων ήταν 0, ανεξάρτητα μεγέθους καταστήματος και περιοχής, δηλαδή $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$. Δηλαδή $H_0 : \boldsymbol{\beta} = \boldsymbol{0}_6$, κι έχουμε $\boldsymbol{R} = \boldsymbol{I}_6$. Βάσει του τελευταίου αποτελεσματος στο Θεώρημα (9) έχουμε

   $$\frac{\hat{\boldsymbol{\beta}}^T (\boldsymbol{X}^T \boldsymbol{X}) \hat{\boldsymbol{\beta}}}{6 s^2} \overset{H_0}{\sim} F_{6,94}$$

2. H_0 : η εκστρατεία ήταν το ίδιο αποτελεσματική ανεξάρτητα του μεγέθους του καταστήματος.

   Απάντηση: Ισοδύναμα, $H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$. Δηλαδή $H_0 : \boldsymbol{R\beta} = \mathbf{0}_4$, όπου

$$\boldsymbol{R} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{r} = \mathbf{0}_4$$

   Βάσει του τελευταίου αποτελεσματος στο Θεώρημα (9) έχουμε

$$\frac{(\boldsymbol{R}\hat{\boldsymbol{\beta}})^T(\boldsymbol{R}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{R}^T)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}})}{4s^2} \overset{H_0}{\sim} F_{4,94}$$

   Παρατηρήστε εδώ πως $\boldsymbol{R}\hat{\boldsymbol{\beta}}$ είναι απλά το διάνυσμα $(\hat{\beta}_2\ \hat{\beta}_3\ \hat{\beta}_4\ \hat{\beta}_5)^T$, ενώ $\boldsymbol{R}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{R}^T$ είναι απλά ο 4×4 κάτω υποπίνακας του $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$, (δηλαδή συνδυάστε τις 4 τελευταίες γραμμές με τις 4 τελευταίες στήλες).

3. H_0 : η επίδραση του μέσου οικογενειακού εισοδήματος στην αποτελεσματικότητα της εκστρατείας ήταν ίδια για κάθε μεγέθος καταστήματος.

   Απάντηση: Η μηδενική υπόθεση λέει πως $\beta_1 = \beta_1 + \beta_4 = \beta_1 + \beta_5$. Ισοδύναμα, $H_0 : \beta_4 = \beta_5 = 0$. Άρα

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} = \mathbf{0}_2.$$

   Δηλαδή $H_0 : \boldsymbol{R\beta} = \mathbf{0}_2$, όπου

$$\boldsymbol{R} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{r} = \mathbf{0}_2$$

   Βάσει του τελευταίου αποτελεσματος στο Θεώρημα (9) έχουμε

$$\frac{(\boldsymbol{R}\hat{\boldsymbol{\beta}})^T(\boldsymbol{R}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{R}^T)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}})}{2s^2} \overset{H_0}{\sim} F_{2,94}$$

   Παρατηρήστε εδώ πως $\boldsymbol{R}\hat{\boldsymbol{\beta}}$ είναι απλά το διάνυσμα $(\hat{\beta}_4\ \hat{\beta}_5)^T$, ενώ $\boldsymbol{R}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{R}^T$ είναι απλά ο 2×2 κάτω υποπίνακας του $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$, (δηλαδή συνδυάστε τις 2 τελευταίες γραμμές με τις 2 τελευταίες στήλες).

4. H_0 : η επίδραση του μέσου οικογενειακού εισοδήματος στην αποτελεσματικότητα της εκστρατείας ήταν ίδια για τα μεσαίου και μεγάλου μεγέθους καταστήματα.

   Η μηδενική υπόθεση λέει πως $\beta_1 + \beta_4 = \beta_1 + \beta_5$, ή ισοδύναμα $\beta_4 = \beta_5$, δηλαδή $\beta_4 - \beta_5 = 0$. Ήτοι

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} = 0$$

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

   Βάσει του τελευταίου αποτελεσματος στο Θεώρημα (9) έχουμε

$$\frac{(\boldsymbol{R}\hat{\boldsymbol{\beta}})^T(\boldsymbol{R}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{R}^T)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}})}{s^2} \overset{H_0}{\sim} F_{1,94}$$

Το επόμενο Θεώρημα λέει πως σε ένα γραμμικό μοντέλο με ομοσκεδατικά κι ασυσχέτιστα σφάλματα ο εκτιμητης ελαχίστων τετραγώνων είναι βέλτιστος μέσα στην οικογένεια των γραμμικών κι αμερόληπτων εκτιμητών (Best Linear Unbiased Estimator, BLUE).

Θεώρημα 10. *Έστω $\boldsymbol{Y} \sim (\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I})$, με $\boldsymbol{X}$ έναν $n\times p$ πίνακα με $rank(X) = p$. Με άλλα λογια $E(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta}$, και $Var(\boldsymbol{Y}) = \sigma^2\boldsymbol{I}$. Δεν υποθέτουμε Κανονικότητα. Ο εκτιμητης ελαχίστων τετραγώνων είναι ο εκτιμητής με τον μικρότερο πίνακα διασπορών-συνδιασπορών μέσα στην οικογένεια των γραμμικών κι αμερόληπτων εκτιμητών.*

Απόδειξη. Προφανώς ο εκτιμητής $\hat{\boldsymbol{\beta}}_{LS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ είναι γραμμικός (ως προς $\boldsymbol{Y}$) κι έχουμε δείξει πως είναι αμερόληπτος. Επίσης έχουμε δείξει πως $Var(\hat{\boldsymbol{\beta}}_{LS}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.

Έστω $\tilde{\boldsymbol{\beta}} = \boldsymbol{C}\boldsymbol{Y}$ ένας άλλος αμερόληπτος γραμμικός εκτιμητής. Καθότι ο $\tilde{\boldsymbol{\beta}}$ είναι αμερόληπτος έχουμε $E(\tilde{\boldsymbol{\beta}}) = E(\boldsymbol{C}\boldsymbol{Y}) = \boldsymbol{C}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\beta}$, $\forall \boldsymbol{\beta}$, κι άρα ισχύει $\boldsymbol{C}\boldsymbol{X} = \boldsymbol{I}$. Προφανώς $Var(\tilde{\boldsymbol{\beta}}) = \sigma^2 \boldsymbol{C}\boldsymbol{C}^T$.

Θεωρούμε την διαφορά $\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{LS}$. Καθότι η διαφορά είναι τ.δ. ισχύει $Var(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{LS}) \geq \boldsymbol{0}$, δηλαδή ο πίνακας $Var(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{LS})$ είναι μη αρνητικά ορισμένος. Άρα

$$\begin{aligned}
Var(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{LS}) &= Var((\boldsymbol{C} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T)\boldsymbol{Y}) \\
&= \sigma^2(\boldsymbol{C} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T)(\boldsymbol{C} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T)^T \\
&= \sigma^2(\boldsymbol{C}\boldsymbol{C}^T - \boldsymbol{C}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{C}^T + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}) \\
&= \sigma^2\boldsymbol{C}\boldsymbol{C}^T - \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1} \text{ (καθότι } \boldsymbol{C}\boldsymbol{X} = \boldsymbol{I}) \\
&= Var(\tilde{\boldsymbol{\beta}}) - Var(\hat{\boldsymbol{\beta}}_{LS}) \geq \boldsymbol{0} \implies Var(\tilde{\boldsymbol{\beta}}) \geq Var(\hat{\boldsymbol{\beta}}_{LS})
\end{aligned}$$

□

2.1 Εκτιμητής Γενικευμένων Ελαχίστων Τετραγώνων (Generalized Least Squares (GLS) estimator)

Θεωρήστε το γραμμικό μοντέλο (γραμμική παλιδρόμηση) με πιθανή έλλειψη ομοσκεδαστικότητας ή/και ύπαρξη αυτοσυσχέτισης των σφαλμάτων (και άρα και των παρατηρήσεων):

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim (\boldsymbol{0}, \boldsymbol{\Sigma}), \tag{2.3}$$

όπου $\mathbf{Y}$ είναι ένα $n \times 1$ διάνυσμα αποκρίσεων, $\mathbf{X}$ ο $n \times p$ πίνακας σχεδιασμού ή συμμεταβλητών (η πρώτη στήλη είναι ένα διάνυσμα μονάδων όταν έχουμε σταθερά (intercept) στην παλινδρόμηση), $\boldsymbol{\beta}$ είναι το $p \times 1$ διάνυσμα παραμέτρων, και $\boldsymbol{\epsilon}$ είναι ένα $n \times 1$ διάνυσμα σφαλμάτων. Επίσης υποθέτουμε πως $rank(\boldsymbol{X}) = p$, δηλαδή ο $\boldsymbol{X}$ έχει γραμμικά ανεξάρτητες στήλες. Ισοδύναμα μπορούμε να γράψουμε το μοντέλο (2.3) ως

$$\boldsymbol{Y} \sim (\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}).$$

Ο Πίνακας $\boldsymbol{\Sigma}$ είναι ο $n \times n$ πίνακας διακυμάνσεων-συνδυακυμάνσεων των σφαλμάτων, με i, j στοιχείο $\sigma_{i,j} = cov(Y_i, Y_j)$, όπου $i = 1, \ldots, n$ και $j = 1, \ldots, n$. Υποθέτουμε πως ο $\boldsymbol{\Sigma}$ είναι θετικά ορισμένος (p.d.). Σημείωση: όταν $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n$ έχουμε την κλασσική παλινδρόμηση (δηλ. έχουμε ομοσκεδαστικά και ασυσχέτιστα σφάλματα).

Έχουμε δείξει πως ο εκτιμητής ελαχίστων τεραγώνων $\hat{\boldsymbol{\beta}}_{LS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ είναι Βέλτιστος Γραμμικός Αμεροληπτος Εκτιμητής (BLUE) του $\boldsymbol{\beta}$ όταν $\boldsymbol{\Sigma} =$

$\sigma^2 \boldsymbol{I}_n$. Το ερώτημα που θα απαντήσουμε τώρα είναι: ποιός είναι ο (BLUE) εκτιμητής του $\boldsymbol{\beta}$ στην γενική περίπτωση.

Θεώρημα 11. *Ο BLUE του $\boldsymbol{\beta}$ ισούται με* $\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y}$

Απόδειξη. Έχουμε δείξει πως η τεραγωνική ρίζα του p.d. πίνακα $\boldsymbol{\Sigma}$ (με φασματική ανάλυση $\boldsymbol{\Sigma} = \boldsymbol{P} \boldsymbol{\Lambda} \boldsymbol{P}^T$) υπάρχει, είναι μοναδικός. και δίνεται από τον τύπο $\boldsymbol{\Sigma}^{1/2} = \boldsymbol{P} \boldsymbol{\Lambda}^{1/2} \boldsymbol{P}^T$. Ο αντίστροφος του ισούται με $\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{P} \boldsymbol{\Lambda}^{-1/2} \boldsymbol{P}^T$. Πολλαπλασιάζουμε από δεξιά τα δύο μέλη της εξίσωσης του μοντέλου στο (2.3) με τον πίνακα $\boldsymbol{\Sigma}^{-1/2}$ κι έχουμε :

$$\mathbf{Y}^* = \mathbf{X}^* \boldsymbol{\beta} + \boldsymbol{\epsilon}^*, \ \boldsymbol{\epsilon}^* \sim (\mathbf{0}, I_n), \tag{2.4}$$

όπου

$$\begin{aligned} \boldsymbol{Y}^* &= \boldsymbol{\Sigma}^{-1/2} \boldsymbol{Y} \\ \boldsymbol{X}^* &= \boldsymbol{\Sigma}^{-1/2} \boldsymbol{X} \\ \boldsymbol{\epsilon}^* &= \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\epsilon}. \end{aligned}$$

Είναι προφανές πως $Var(\boldsymbol{\epsilon}^*) = Var(\boldsymbol{\Sigma}^{-1/2} \boldsymbol{\epsilon}) = \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1/2} = \boldsymbol{I}_n$. Με αυτόν τον μετασχηματισμό καταλήγουμε από το μοντέλο στο (2.3) σε ένα μοντέλο κλασσικής παλινδρόμησης στο (2.4). Ο BLUE του $\boldsymbol{\beta}$ προφανώς είναι ο εκτιμητής

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{GLS} &= (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \mathbf{X}^{*T} \mathbf{Y}^* \\ &= (\mathbf{X}^T \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Sigma}^{-1/2} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Sigma}^{-1/2} \mathbf{Y} \\ &= (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y}. \end{aligned} \tag{2.5}$$

□

Συνήθως γράφουμε τον πίνακα διακυμάνσεων-συνδιακυμάνσεων ως $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{V}$, όπου $\sigma^2 > 0$. Τότε έχουμε

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^T \boldsymbol{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{V}^{-1} \mathbf{Y} \tag{2.6}$$

2.1.1 Ιδιότητες του Εκτιμητή Γενικευμένων Ελαχίστων Τετραγώνων (GLS) - Σύγκριση με τον Εκτιμητή Ελαχίστων Τετραγώνων (LS)

Για τον $\hat{\boldsymbol{\beta}}_{GLS}$ έχουμε

$$
\begin{aligned}
E(\hat{\boldsymbol{\beta}}_{GLS}) &= E\left((\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}\right) \\
&= (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}E(\boldsymbol{Y}) \\
&= (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\beta} \\
Var(\hat{\boldsymbol{\beta}}_{GLS}) &= Var((\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}) \\
&= (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}Var(\boldsymbol{Y})\left((\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\right)^T \\
&= (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\left((\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\right)^T \\
&= (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \\
&= (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \\
&= (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{X})^{-1} \qquad (2.7)
\end{aligned}
$$

Από την άλλη, για τον $\hat{\boldsymbol{\beta}}_{LS}$ έχουμε

$$
\begin{aligned}
E(\hat{\boldsymbol{\beta}}_{LS}) &= \boldsymbol{\beta} \\
Var(\hat{\boldsymbol{\beta}}_{LS}) &= Var((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}) = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\boldsymbol{\Sigma}\mathbf{X})(\mathbf{X}^T\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\boldsymbol{V}\mathbf{X})(\mathbf{X}^T\mathbf{X})^{-1} \qquad (2.8)
\end{aligned}
$$

Στην περίπτωση κανονικού γραμμικού μοντέλου, δηλαδή αν $\boldsymbol{Y} \sim N(\boldsymbol{X\beta}, \boldsymbol{\Sigma})$, έχουμε πως

$$
\begin{gathered}
\hat{\boldsymbol{\beta}}_{GLS} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{X})^{-1}) \\
\hat{\boldsymbol{\beta}}_{LS} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\boldsymbol{V}\mathbf{X})(\mathbf{X}^T\mathbf{X})^{-1}). \qquad (2.9)
\end{gathered}
$$

Άσκηση (για το σπίτι).

Θεωρήστε το κανονικό γραμμικό μοντέλο $\boldsymbol{Y} \sim N(\boldsymbol{X\beta}, \sigma^2\boldsymbol{V})$, με $\boldsymbol{V}$ γνωστό p.d. πίνακα. Βρείτε τους εκτιμητές μεγίστης πιθανοφάνειας των $\boldsymbol{\beta}$ και σ^2.

Ένδειξη: χρησιμοποιήσετε τη φασματική ανάλυση του $\boldsymbol{V}$ για να ορίσετε το $\boldsymbol{V}^{1/2}$ και το Θεώρημα (8).

2.1.2 Παραδείγματα

Παράδειγμα 1: Y_i είναι δειγματικοί μέσοι που προκύπτουν από k_i ανεξάρτητες παρατηρήσεις στο i-οστό εργαστήριο ($i = 1, \ldots, n$).

$$\begin{aligned} Y_i &= \boldsymbol{x}_i^T \boldsymbol{\beta} + \epsilon_i \\ \epsilon_i &\sim^{ind} N(0, \sigma^2 \frac{1}{k_i}) \end{aligned}$$

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \\ \vdots \\ \boldsymbol{x}_n^T \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1,p-1} \\ 1 & x_{21} & x_{22} & \ldots & x_{2,p-1} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{n,p-1} \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \sigma^2 \begin{bmatrix} \frac{1}{k_1} & 0 & \ldots & 0 \\ 0 & \frac{1}{k_2} & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & \frac{1}{k_n} \end{bmatrix}$$

Εδώ ισχύει πως $\boldsymbol{V}$ είναι γνωστός (πράγμα σπάνιο).

Παράδειγμα 2: Παλινδρόμηση με AR(1) σφάλματα.

$$\begin{aligned} Y_t &= \boldsymbol{x}_t^T \boldsymbol{\beta} + \epsilon_t \\ \epsilon_t &= \phi \epsilon_{t-1} + \nu_t, \ \nu_t \sim^{iid} N(0, \sigma_\nu^2) \\ & -1 < \phi < 1 \end{aligned}$$

Εύκολα μπορεί να δειχθεί πως

$$\begin{aligned} \epsilon_t &\sim N(0, \frac{\sigma_\nu^2}{1-\phi^2}) \\ cov(\epsilon_t, \epsilon_{t-k}) &= \phi^k \frac{\sigma_\nu^2}{1-\phi^2} \end{aligned}$$

$$\boldsymbol{\Sigma} = \sigma_\nu^2 \begin{bmatrix} \frac{1}{1-\phi^2} & \frac{\phi}{1-\phi^2} & \frac{\phi^2}{1-\phi^2} & \ldots & \frac{\phi^{n-1}}{1-\phi^2} \\ \frac{\phi}{1-\phi^2} & \frac{1}{1-\phi^2} & \frac{\phi}{1-\phi^2} & \ldots & \frac{\phi^{n-2}}{1-\phi^2} \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ \frac{\phi^{n-1}}{1-\phi^2} & \frac{\phi^{n-2}}{1-\phi^2} & \frac{\phi^{n-3}}{1-\phi^2} & \ldots & \frac{1}{1-\phi^2} \end{bmatrix}$$

Παράδειγμα 3: Απλό γραμμικό μοντέλο τυχαίων παραγόντων (one way random effects model).

$$\begin{aligned} Y_{ij} &= \mu_i + \nu_{ij},\ \nu_{ij} \sim^{iid} N(0,\sigma^2),\ j=1,\ldots,n_i \\ \mu_i &\sim^{iid} N(\mu,\sigma_\pi^2),\ i=1,\ldots,k \\ \nu_{ij} &\perp \nu_{i'j'},\ i\neq i' \\ \nu_{ij} &\perp \mu_i,\ \forall(i,j) \end{aligned} \tag{2.10}$$

Ισοδύναμα, ορίζοντας $\pi_i = \mu_i - \mu$, γράφουμε το μοντέλο ως

$$\begin{aligned} Y_{ij} &= \mu + \pi_i + \nu_{ij},\ \nu_{ij} \sim^{iid} N(0,\sigma^2),\ j=1,\ldots,n_i \\ \pi_i &\sim^{iid} N(0,\sigma_\pi^2),\ i=1,\ldots,k \\ \nu_{ij} &\perp \nu_{i'j},\ i\neq i' \\ \nu_{ij} &\perp \pi_i,\ \forall(i,j) \end{aligned} \tag{2.11}$$

Το μοντέλο μπορεί να γραφεί ως Γενικό, Γραμμικό μοντέλο, δηλαδή στην μορφή (2.3). Ορίζω $\epsilon_{ij} = \pi_i + \nu_{ij}$. Έχουμε

$$\begin{aligned} var(\epsilon_{ij}) &= \sigma_\pi^2 + \sigma^2 \\ cov(\epsilon_{ij},\epsilon_{ij'}) &= \sigma_\pi^2,\ j\neq j' \\ cov(\epsilon_{ij},\epsilon_{i'j'}) &= 0,\ i\neq i' \end{aligned} \tag{2.12}$$

Εδώ $\boldsymbol{Y} = (Y_{11},\ldots,Y_{1n_1},Y_{21},\ldots,Y_{2n_2},\ldots,\ldots,Y_{k1},\ldots,Y_{kn_k})^T$, $\boldsymbol{\beta} = \mu$, $\boldsymbol{X} = \mathbf{1}_N$, όπου $\mathbf{1}_N$ είναι ένα διάνυσμα με $N = n_1 + \ldots + n_k$ μονάδες, $\boldsymbol{\epsilon} = (\epsilon_{11},\ldots,\epsilon_{1n_1},\epsilon_{21},\ldots,\epsilon_{2n_2},\ldots,\ldots,\epsilon_{k1},\ldots,\epsilon_{kn_k})^T$ και $var(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}^*$ δίνεται από

$$\boldsymbol{\Sigma}^* = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \ldots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{\Sigma}_k \end{bmatrix} = \sigma^2 \begin{bmatrix} \boldsymbol{V}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{V}_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \ldots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{V}_k \end{bmatrix} = \sigma^2 \mathscr{V}$$

όπου $\boldsymbol{\Sigma}_i = var(\boldsymbol{\epsilon}_i)$ είναι ένας $n_i \times n_i$ πίνακας συνδιακυμάνσεων των παρατηρήσεων από το i-οστό εργαστήριο και δίνεται από

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \sigma_\pi^2+\sigma^2 & \sigma_\pi^2 & \ldots & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2+\sigma^2 & \ldots & \sigma_\pi^2 \\ \vdots & \vdots & \ldots & \vdots \\ \sigma_\pi^2 & \sigma_\pi^2 & \ldots & \sigma_\pi^2+\sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1+\phi & \phi & \ldots & \phi \\ \phi & 1+\phi & \ldots & \phi \\ \vdots & \vdots & \ldots & \vdots \\ \phi & \phi & \ldots & 1+\phi \end{bmatrix}$$

όπου $\phi = \frac{\sigma_\pi^2}{\sigma^2} > 0$

Παράδειγμα 4: Απλό γραμμικό μοντέλο με ένα σταθερό παράγοντα κι εναν τυχαίο παράγοντα (two way mixed effects model).

$$\begin{aligned} Y_{ij} &= \mu + \alpha_i + \beta_j + \nu_{ij},\ \nu_{ij} \sim^{iid} N(0,\sigma^2),\ j = 1,\ldots,n \\ \alpha_i &\sim^{iid} N(0,\sigma_\alpha^2),\ i = 1,\ldots,m \\ \nu_{ij} &\perp \nu_{i'j},\ i \neq i' \\ \nu_{ij} &\perp \alpha_i,\ \forall(i,j) \end{aligned} \tag{2.13}$$

Μπορώ να γράψω το μοντέλο ώς εξής:

$$\begin{aligned} Y_{ij} &= (\mu + \beta_j) + \alpha_i + \nu_{ij},\ \nu_{ij} \sim^{iid} N(0,\sigma^2),\ j = 1,\ldots,n \\ \alpha_i &\sim^{iid} N(0,\sigma_\alpha^2),\ i = 1,\ldots,m \\ \nu_{ij} &\perp \nu_{i'j},\ i \neq i' \\ \nu_{ij} &\perp \alpha_i,\ \forall(i,j), \end{aligned} \tag{2.14}$$

όπου ο όρος $(\mu + \beta_j)$ είναι σταθερός.

Το μοντέλο το γράφω σαν κανονικό Γενικό γραμμικό μοντέλο ($\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$),

$$\boldsymbol{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n} \\ - \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2n} \\ - \\ \vdots \\ \vdots \\ - \\ Y_{m1} \\ Y_{m2} \\ \vdots \\ Y_{mn} \end{bmatrix} = \begin{bmatrix} \mu+\beta_1 \\ \mu+\beta_2 \\ \vdots \\ \mu+\beta_n \\ - \\ \mu+\beta_1 \\ \mu+\beta_2 \\ \vdots \\ \mu+\beta_n \\ - \\ \vdots \\ \vdots \\ - \\ \mu+\beta_1 \\ \mu+\beta_2 \\ \vdots \\ \mu+\beta_n \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_1 \\ \vdots \\ \alpha_1 \\ - \\ \alpha_2 \\ \alpha_2 \\ \vdots \\ \alpha_2 \\ - \\ \vdots \\ \vdots \\ - \\ \alpha_m \\ \alpha_m \\ \vdots \\ \alpha_m \end{bmatrix} + \begin{bmatrix} \nu_{11} \\ \nu_{12} \\ \vdots \\ \nu_{1n} \\ - \\ \nu_{21} \\ \nu_{22} \\ \vdots \\ \nu_{2n} \\ - \\ \vdots \\ \vdots \\ - \\ \nu_{m1} \\ \nu_{m2} \\ \vdots \\ \nu_{mn} \end{bmatrix}$$

Εδώ $\boldsymbol{\beta} = (\mu, \beta_1, \ldots, \beta_n)^T$. Άρα ο $mn \times (n+1)$ Πίνακας σχεδιασμού είναι

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 & 0 & \ldots & 0 \\ 1 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & 1 \\ - & - & - & - & - \\ 1 & 1 & 0 & \ldots & 0 \\ 1 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & 1 \\ - & - & - & - & - \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ - & - & - & - & - \\ 1 & 1 & 0 & \ldots & 0 \\ 1 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & 1 \end{bmatrix}$$

και ο πίνακας διασποράς είναι ο $mn \times mn$ πίνακας

$$\boldsymbol{\Sigma}^* = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ldots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{\Sigma} \end{bmatrix} = \sigma^2 \begin{bmatrix} \boldsymbol{V} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{V} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ldots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{V} \end{bmatrix} = \sigma^2 \mathscr{V}$$

όπου $\boldsymbol{\Sigma}$ είναι ένας $n \times n$ πίνακας συνδιακυμάνσεων των παρατηρήσεων από το i-οστό εργαστήριο και δίνεται από

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \sigma_\alpha^2 + \sigma^2 & \sigma_\alpha^2 & \ldots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma^2 & \ldots & \sigma_\alpha^2 \\ \vdots & \vdots & \ldots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \ldots & \sigma_\alpha^2 + \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1+\phi & \phi & \ldots & \phi \\ \phi & 1+\phi & \ldots & \phi \\ \vdots & \vdots & \ldots & \vdots \\ \phi & \phi & \ldots & 1+\phi \end{bmatrix}$$

όπου $\phi = \frac{\sigma_\alpha^2}{\sigma^2} > 0$

Προσοχή εδώ $rank(\boldsymbol{X}) = n < n+1$. Κανένα πρόβλημα! (;). Παραμετροποιώ το μοντέλο ορίζοντας $\mu_j = \mu + \beta_j$, κι έχω ως $\boldsymbol{\beta} = (\mu_1, \ldots, \mu_n)^T$. Δηλαδή δουλεύω με το ισοδύναμο μοντέλο:

$$
\begin{array}{rcl}
Y_{ij} & = & \mu_j + \alpha_i + \nu_{ij},\ \nu_{ij} \sim^{iid} N(0,\sigma^2),\ j = 1,\ldots,n \\
\alpha_i & \sim^{iid} & N(0,\sigma^2_\alpha),\ i = 1,\ldots,m \\
\nu_{ij} & \perp & \nu_{i'j},\ i \neq i' \\
\nu_{ij} & \perp & \alpha_i,\ \forall(i,j)
\end{array} \tag{2.15}
$$

και

$$
\boldsymbol{X} = \begin{bmatrix}
1 & 0 & \ldots & 0 & \\
0 & 1 & \ldots & 0 & \\
\vdots & \vdots & \vdots & \vdots & \\
0 & 0 & \ldots & 1 & \\
- & - & - & - & \\
1 & 0 & \ldots & 0 & \\
0 & 1 & \ldots & 0 & \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \ldots & 1 & \\
- & - & - & - & \\
\vdots & \vdots & \vdots & \vdots & \\
\vdots & \vdots & \vdots & \vdots & \\
- & - & - & - & \\
1 & 0 & \ldots & 0 & \\
0 & 1 & \ldots & 0 & \\
\vdots & \vdots & \vdots & \vdots & \\
0 & 0 & \ldots & 1 &
\end{bmatrix}
$$

Παράδειγμα 5: Κανονικό γραμμικό μοντέλο με τυχαίους συντελεστές (random coefficients model).

Θεωρήστε επαναλαμβανόμενες μετρήσεις σε άτομα που συμμετέχουν σε μια κλινική δοκιμή. Υποθέτουμε πως μετά την έναρξη θεραπείας η απόκριση Y (ενδεχομένως) ακολουθεί μια γραμμική τάση (αυξάνεται ή μειώνεται) μέσα σε ένα δεδομένο χρονικό διάστημα. Η τάση αυτή ΄ποικίλει΄από άτομο σε άτομο.

$$\begin{aligned} Y_{ij} &= b_{0i} + b_{1i}t_{ij} + \nu_{ij},\ i=1,\ldots,k,\ j=1,\ldots,n_i \\ b_{0i} &\sim^{iid} N(\beta_0, \sigma_0^2) \\ b_{1i} &\sim^{iid} N(\beta_1, \sigma_1^2) \\ cov(b_{0i}, b_{1i}) &= \sigma_{01} \\ \nu_{ij} &\sim^{iid} N(0, \sigma^2) \\ \nu_{ij} &\perp (b_{0i}, b_{1i})^T \end{aligned} \tag{2.16}$$

Είναι λογικό πως υπάρχει ανεξαρτησία των τυχαίων συντελεστών, σφαλμάτων (κι άρα και των παρατηρήσεων) μεταξύ των ατόμων. Ισοδύναμα, παίρνοντας $b_{0i} = \beta_0 + (b_{0i} - \beta_0) = \beta_0 + \pi_{0i}$ και $b_{1i} = \beta_1 + (b_{1i} - \beta_0) = \beta_1 + \pi_{1i}$ γράφουμε:

$$\begin{aligned} Y_{ij} &= \beta_0 + \beta_1 t_{ij} + \pi_{0i} + \pi_{1i}t_{ij} + \nu_{ij}\ i=1,\ldots,k,\ j=1,\ldots,n_i \\ \pi_{0i} &\sim^{iid} N(0, \sigma_0^2) \\ \pi_{1i} &\sim^{iid} N(0, \sigma_1^2) \\ cov(\pi_{0i}, \pi_{1i}) &= \sigma_{01} \\ \nu_{ij} &\sim^{iid} N(0, \sigma^2) \\ \nu_{ij} &\perp (\pi_{0i}, \pi_{1i})^T \end{aligned} \tag{2.17}$$

Το μοντέλο γράφεται εύκολα ως μικτό μοντέλο σε μορφή πινάκων ως εξής. Γραφουμε το διάνυσμα αποκρίσεων του ι-οστού ατόμου, ως $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})^T$. Έχουμε

$$\begin{aligned} \boldsymbol{Y}_i &= \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\pi}_i + \boldsymbol{\nu}_i,\ i=1,\ldots,k, \\ \boldsymbol{\pi}_i &\sim^{iid} \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{B}),\ \boldsymbol{\nu}_i \sim^{iid} N(0, \sigma^2\boldsymbol{I}_{n_i}) \\ \boldsymbol{\nu}_i &\perp \boldsymbol{\pi}_{i'}, \forall(i, i'), \end{aligned} \tag{2.18}$$

όπου

$$\boldsymbol{X}_i = \begin{bmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \boldsymbol{\pi}_i = \begin{bmatrix} \pi_{i0} \\ \pi_{i1} \end{bmatrix}, \boldsymbol{\nu}_i = \begin{bmatrix} \nu_{i1} \\ \nu_{i2} \\ \vdots \\ \nu_{in_i} \end{bmatrix}.$$

Εδώ έχουμε $\boldsymbol{X}_i = \boldsymbol{Z}_i$. Ισοδύναμα γράφουμε

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{X\beta} + \boldsymbol{Z\pi} + \boldsymbol{\nu} && (2.19) \\ \boldsymbol{\pi} &\sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{I}_k \otimes \boldsymbol{B}),\ \boldsymbol{\nu} \sim \boldsymbol{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_N) && (2.20) \end{aligned}$$

όπου

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \vdots \\ \boldsymbol{Y}_k \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_k \end{bmatrix}, \boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 & \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_2 & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{Z}_k \end{bmatrix}, \boldsymbol{\pi} = \begin{bmatrix} \boldsymbol{\pi}_1 \\ \boldsymbol{\pi}_2 \\ \vdots \\ \boldsymbol{\pi}_k \end{bmatrix}, \boldsymbol{\nu} = \begin{bmatrix} \boldsymbol{\nu}_1 \\ \boldsymbol{\nu}_2 \\ \vdots \\ \boldsymbol{\nu}_k \end{bmatrix}.$$

, από όπου προκύπτει πως

$$\boldsymbol{Y} \sim \boldsymbol{N}(\boldsymbol{X\beta}, \boldsymbol{Z}(\boldsymbol{I}_k \otimes \boldsymbol{B})\boldsymbol{Z}^T + \sigma^2 \boldsymbol{I}_N) \quad (2.21)$$

Στήν ειδική περίπτωση που $\boldsymbol{Z}_1 = \ldots = \boldsymbol{Z}_k = \mathscr{Z}$ έχουμε $\boldsymbol{Z} = \boldsymbol{I}_k \otimes \mathscr{Z}$.

Παράδειγμα 6: Κανονικό γραμμικό μοντέλο με τυχαίους συντελεστές και 2 ομάδες.

Το μοντέλο επεκτείνεται εύκολα να συμπεριλάβει την επίδραση μιας θεραπείας (φαρμάκου) σε σύγκριση με ένα (placebo). Έστω $d_{ij} = 1$ αν το άτομο παίρνει το φάρμακο (θεραπεία) και $d_{ij} = 0$ αν το άτομο παίρνει το placebo. Τότε μία πιθανή γενίκευση του μοντέλου είναι:

$$\begin{aligned} Y_{ij} &= \beta_0 + \beta_1 t_{ij} + \beta_2 d_{ij} + \pi_{0i} + \pi_{1i} t_{ij} + \nu_{ij}\ i = 1, \ldots, k,\ j = 1, \ldots, n_i \\ \pi_{0i} &\sim^{iid} N(0, \sigma_0^2) \\ \pi_{1i} &\sim^{iid} N(0, \sigma_1^2) \\ cov(\pi_{0i}, \pi_{1i}) &= \sigma_{01} \\ \nu_{ij} &\sim^{iid} N(0, \sigma^2) \\ \nu_{ij} &\perp (\pi_{0i}, \pi_{1i})^T \end{aligned} \quad (2.22)$$

Το παραπάνω μοντέλο λέει πως το φάρμακο ασκεί επίρροια μόνο στο (intercept) αλλά όχι στην κλίση, δηλαδή οι μέσες πληθυσμιακές ευθείες είναι παράλληλες!!! Άρα οι κλισεις είναι ίδιες. Για να εξετάσω αν υπάρχεια επίρροια του φαρμάκου στην κλίση εξετάζω το μοντέλο

$$
\begin{aligned}
Y_{ij} &= \beta_0 + \beta_1 t_{ij} + \beta_2 d_{ij} + \beta_3 d_{ij} t_{ij} + \pi_{0i} + \pi_{1i} t_{ij} + \nu_{ij} \; i = 1, \ldots, k, \; j = 1, \ldots, n_i \\
\pi_{0i} &\sim^{iid} N(0, \sigma_0^2) \\
\pi_{1i} &\sim^{iid} N(0, \sigma_1^2) \\
cov(\pi_{0i}, \pi_{1i}) &= \sigma_{01} \\
\nu_{ij} &\sim^{iid} N(0, \sigma^2) \\
\nu_{ij} &\perp (\pi_{0i}, \pi_{1i})^T
\end{aligned}
\tag{2.23}
$$

Εδώ βλέπουμε πως για την ομάδα placebo $E(Y_t) = \beta_0 + \beta_1 t$ ενώ για την ομάδα φαρμάκου $E(Y_t) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)t$.

Στο μοντέλο αυτό ισχύει ότι ισχύει και στο παράδειγμα 5, με τις εξής επισημάνσεις (διαφορές):

$$
\boldsymbol{X}_i = \begin{bmatrix} 1 & t_{i1} & d_{i1} & d_{i1}t_{i1} \\ 1 & t_{i2} & d_{i2} & d_{i2}t_{i2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_{in_i} & d_{in_i} & d_{in_i}t_{in_i} \end{bmatrix}, \boldsymbol{Z}_i = \begin{bmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{bmatrix} \neq \boldsymbol{X}_i, \; \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}
$$

2.2 Εκτίμηση με την μέθοδο της μέγιστης Πιθανοφάνειας (Maximum Likelihood Estimation (MLE))

Με την υπόθεση κανονικότητας έχουμε

$$
\boldsymbol{Y}|\boldsymbol{X} \sim N(\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\sigma^2, \boldsymbol{\phi}) = \sigma^2 \boldsymbol{V}(\boldsymbol{\phi})). \tag{2.24}
$$

Η loglikelihood είναι:

$$
l(\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}) = -\frac{n}{2} ln 2\pi - \frac{n}{2} ln \sigma^2 - \frac{1}{2} ln |\boldsymbol{V}(\boldsymbol{\phi})| - \frac{1}{2\sigma^2} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{V}^{-1}(\boldsymbol{\phi}) (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \tag{2.25}
$$

Δοθέντος $\boldsymbol{\phi}$ (και σ^2), η loglikelihood μεγιστοποιείται ως προς $\boldsymbol{\beta}$ στην τιμή $\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\phi}) = (\mathbf{X}^T\boldsymbol{V}^{-1}(\boldsymbol{\phi})\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{V}^{-1}(\boldsymbol{\phi})\mathbf{Y}$. (γιατί;;)

Άρα αρκεί να μεγιστοποιήσουμε 'συγκεντρωμένη' πιθανοφάνεια

$$l_1(\sigma^2, \boldsymbol{\phi}) = -\frac{n}{2}ln2\pi - \frac{n}{2}ln\sigma^2 - \frac{1}{2}ln|\boldsymbol{V}(\boldsymbol{\phi})| - \frac{1}{2\sigma^2}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\phi}))^T\boldsymbol{V}^{-1}(\boldsymbol{\phi})(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\phi})) \tag{2.26}$$

Δοθέντος $\boldsymbol{\phi}$ η $l_1(\sigma^2, \boldsymbol{\phi})$ μεγιστοποιείται ως προς σ^2 στην τιμή

$$\hat{\sigma}^2(\boldsymbol{\phi}) = \frac{(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\phi}))^T\boldsymbol{V}^{-1}(\boldsymbol{\phi})(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\phi}))}{n}, \tag{2.27}$$

(γιατί;;).

Καθότι

$$\begin{aligned}
(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS})^T\boldsymbol{V}^{-1}(\boldsymbol{\phi})(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}) = \\
(\boldsymbol{Y} - \boldsymbol{X}(\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{Y})^T\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}(\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{Y}) \\
= \boldsymbol{Y}^T(\boldsymbol{I} - (\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{V}^{-1})^T\boldsymbol{V}^{-1}(\boldsymbol{I} - \boldsymbol{X}(\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{V}^{-1})\mathbf{Y} \\
= \boldsymbol{Y}^T(\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\mathbf{X}^T\boldsymbol{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{V}^{-1})\boldsymbol{Y}
\end{aligned} \tag{2.28}$$

έχουμε

$$\hat{\sigma}^2(\boldsymbol{\phi}) = \frac{1}{n}\boldsymbol{Y}^T(\boldsymbol{V}^{-1}(\boldsymbol{\phi}) - \boldsymbol{V}^{-1}(\boldsymbol{\phi})\boldsymbol{X}(\mathbf{X}^T\boldsymbol{V}^{-1}(\boldsymbol{\phi})\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{V}^{-1}(\boldsymbol{\phi}))\boldsymbol{Y} \tag{2.29}$$

Άρα αρκεί να μεγιστοποιήσουμε 'συγκεντρωμένη'-'συγκεντρωμένη' πιθανοφάνεια.

$$\begin{aligned}
l_1(\boldsymbol{\phi}) &= -\frac{n}{2}ln2\pi - \frac{n}{2}ln\hat{\sigma}^2(\boldsymbol{\phi}) - \frac{1}{2}ln|\boldsymbol{V}(\boldsymbol{\phi})| - \frac{1}{2\hat{\sigma}^2(\boldsymbol{\phi})}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\phi}))^T\boldsymbol{V}^{-1}(\boldsymbol{\phi})(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\phi})) \\
&= -\frac{n}{2}ln2\pi - \frac{n}{2}ln\hat{\sigma}^2(\boldsymbol{\phi}) - \frac{1}{2}ln|\boldsymbol{V}(\boldsymbol{\phi})| - \frac{n}{2}
\end{aligned} \tag{2.30}$$

Δηλαδή <u>ελαχιστοποιούμε</u> την ποσότητα

$$nln\hat{\sigma}^2(\boldsymbol{\phi}) + ln|\boldsymbol{V}(\boldsymbol{\phi})| \tag{2.31}$$

ως προς $\boldsymbol{\phi}$ και παίρνουμε $\hat{\boldsymbol{\phi}}_{MLE}$. κι έχουμε

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{MLE} &= \hat{\boldsymbol{\beta}}_{GLS}(\hat{\boldsymbol{\phi}}_{MLE}) \\
\hat{\sigma}^2_{MLE} &= \hat{\sigma}^2(\hat{\boldsymbol{\phi}}_{MLE}) = \frac{(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\hat{\boldsymbol{\phi}}_{MLE}))^T\boldsymbol{V}^{-1}(\hat{\boldsymbol{\phi}}_{MLE})(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\hat{\boldsymbol{\phi}}_{\boldsymbol{MLE}}))}{n}
\end{aligned} \tag{2.32}$$

Εναλλάκτικά μπορώ να δουλέψω χωρίς να ξεχωρίζω το σ^2, δηλαδή με το $\boldsymbol{\Sigma}(\boldsymbol{\psi})$. Εδώ $\boldsymbol{\psi}$ είναι όλες οι παράμετροι στον πίνακα διασποράς συνδιασποράς.

Η loglikelihood είναι:

$$l(\boldsymbol{Y},\boldsymbol{X};\boldsymbol{\beta},\boldsymbol{\psi}) = -\frac{n}{2}ln2\pi - \frac{1}{2}ln|\boldsymbol{\Sigma}(\boldsymbol{\psi})| - \frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\psi})(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}) \tag{2.33}$$

και η συγκεντρωμένη πιθανοφάνεια ισούται με

$$l_1(\boldsymbol{\psi}) = -\frac{n}{2}ln2\pi - \frac{1}{2}ln|\boldsymbol{\Sigma}(\boldsymbol{\psi})| - \frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\psi}))^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\psi})(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\psi})) \tag{2.34}$$

όπου $\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\psi}) = (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\psi})\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\psi})\mathbf{Y}$,

και ελαχιστοποιώ

$$l_1(\boldsymbol{\psi}) = ln|\boldsymbol{\Sigma}(\boldsymbol{\psi})| + (\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\psi}))^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\psi})(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}(\boldsymbol{\psi})) \tag{2.35}$$

2.3 Γενικευμένα Γραμμικά Μοντέλα (Generalized Linear Models)

Θεωρήστε τα παρακάτω μοντέλα με μια συμμεταβλητή:

- Απλή κλασσική Κανονική Γραμμική Παλινδρόμηση

$$\begin{aligned} Y_i &\sim^{ind} N(\mu_i, \sigma^2) \\ \mu_i &= \beta_0 + \beta_1 x_i, \;\; i = 1, \ldots, n. \end{aligned} \tag{2.36}$$

- Απλή Λογιστική Παλινδρόμηση

$$\begin{aligned} Y_i &\sim^{ind} Bernoulli(\pi_i) \\ \ln(\frac{\pi_i}{1-\pi_i}) &= \beta_0 + \beta_1 x_i, \;\; i = 1, \ldots, n. \end{aligned} \tag{2.37}$$

Ως παράδειγμα, θεωρήστε ένα πείραμα σε ποντίκια εργαστηρίου που αφορά στην εξεύρεση 'κατάλληλης' δοσολογίας ενός ποντικοφάρμακου. Η βιολόγος κάνει ένεση ποντικοφάρμακου σε επίπεδο δόσης x_i, στο τυχαία επιλεγμένο ποντίκι i και παρατηρεί αν το ποντίκι πέθανε σε χρονικό διάστημα 10 λεπτών. Η απόκριση, Y_i, παίρνει τη τιμή 1 αν το ποντίκι πεθάνει, αλλιώς παίρνει τη τιμή 0. Η παράμετρος $\pi_i = \pi(x_i)$ είναι η πιθανότητα θανάτου μέσα σε δέκα λεπτά όταν η δόση ισούται με x_i. Ο λόγος $\frac{\pi_i}{1-\pi_i}$ είναι τα *odds* θανάτου του ποντικιού σε δόση x_i του δηλητηρίου. Σημειώστε πως $\ln(\frac{\pi_i}{1-\pi_i})$ παίρνει τιμές στο $(-\infty, +\infty)$.

Ισοδύναμα το μοντέλο μπορεί να γραφεί ως

$$\begin{aligned} Y_i &\sim^{ind} Bernoulli(\pi_i) \\ \pi_i &= \frac{exp(\beta_0 + \beta_1 x_i)}{1 + exp(\beta_0 + \beta_1 x_i)}, \;\; i = 1, \ldots, n. \end{aligned}$$

Θυμηθείτε πως για μια *Bernoulli* τ.μ. ισχύει $E(Y_i) = \mu_i = \pi_i$ (επίσης $Var(Y_i) = \mu_i(1-\mu_i) = \pi_i(1-\pi_i)$) κι άρα μπορούμε να γράψουμε το μοντέλο κι ως

$$\begin{aligned} Y_i &\sim^{ind} Bernoulli(\mu_i) \\ \ln(\frac{\mu_i}{1-\mu_i}) &= \beta_0 + \beta_1 x_i, \;\; i = 1, \ldots, n. \end{aligned} \tag{2.38}$$

- Απλή Poisson Παλινδρόμηση

$$\begin{aligned} Y_i &\sim^{ind} Poisson(\mu_i) \\ \ln \mu_i &= \beta_0 + \beta_1 x_i, \;\; i = 1, \ldots, n. \end{aligned} \tag{2.39}$$

Για παράδειγμα Y_i μπορεί να είναι ο αριθμός εισαγωγών στο ΑΧΕΠΑ λόγω κορονοϊού κατά την ημέρα i και x_i μια μέτρηση του επιπέδου του covid19 στα αστικά λύμματα της Θεσσαλονίκης 7 μέρες πριν.

Η Απλή Poisson Παλινδρόμηση μπορεί να γραφεί ισοδύναμα ως

$$\begin{aligned} Y_i &\sim^{ind} Poisson(\mu_i) \\ \mu_i &= exp(\beta_0 + \beta_1 x_i), \;\; i = 1, \ldots, n. \end{aligned}$$

Τι κοινό έχουν αυτά τα 3 μοντέλα;

Παρατηρούμε πως και στα 3 ορίζουμε πρώτα την κατανομή της απόκρισης και μετά 'συνδεόυμε' τον μέσο της κατανομής με την συμμεταβλητή. Με άλλα λόγια έχουν τη γενική μορφή

$$\begin{aligned} Y_i &\sim^{ind} F_{Y_i}(y_i) \\ g(\mu_i) &= \beta_0 + \beta_1 x_i, \;\; i = 1, \ldots, n, \end{aligned} \tag{2.40}$$

όπου μ_i είναι ο μέσος της τ.μ. Y_i (ισοδύναμα της κατανομής F_i). Στην απλή κανονική γραμμική παλινδρόμηση έχουμε $g(t) = t$, στην λογιστική παλινδρόμηση έχουμε $g(t) = ln(\frac{t}{1-t})$, και στην Poisson παλινδρόμηση έχουμε $g(t) = ln(t)$.

Στην γενική περίπτωση, και γράφοντας τη σ.π.π. της κατανομής $F_{Y_i}(y_i)$ ως $f_{Y_i}(y_i)$, υποθέτουμε πως η κατανομή ανήκει στην εκθετική οικογένεια κατανομών , δηλαδή η σ.π.π είναι της μορφής

$$f_{Y_i}(y_i) = exp\left[\frac{\gamma_i y_i - b(\gamma_i)}{\tau^2} - c(y_i, \tau)\right], \tag{2.41}$$

κι έχουμε

$$\begin{aligned} Y_i &\sim^{ind} F_{Y_i}(y_i) \\ g(\mu_i) &= \boldsymbol{x}_i^T \boldsymbol{\beta}, \;\; i = 1, \ldots, n, \end{aligned} \tag{2.42}$$

όπου το διάνυσμα $\boldsymbol{x}_i^T = (1 \;\; x_{i1} \;\; x_{i2} \ldots x_{i,p-1})$ περιέχει τις τιμές των συμμεταβλητών για την παρατήρηση i. Η συνάρτηση $g()$ λέγεται συνάρτηση σύνδεσης (link function).

Μπορεί να αποδειχθεί πως

$$
\begin{aligned}
E(Y_i) = \mu_i &= \frac{\partial b(\gamma_i)}{\partial \gamma_i} \\
Var(Y_i) = \tau^2 \frac{\partial^2 b(\gamma_i)}{\partial \gamma_i^2} &= \tau^2 v(\mu_i).
\end{aligned} \tag{2.43}
$$

Η συνάρτηση $v(\mu_i) = \frac{\partial^2 b(\gamma_i)}{\partial \gamma_i^2}$ καλείται συνάρτηση διασποράς (variance function).

Άσκηση

1. Δείξτε πως οι σ.π.π. σε κάθε ένα από τα 3 παραδείγματα ανήκουν στη εκθετική οικογένεια.

   Για τη λογιστική παλινδρόμηση θα δείξουμε ότι η $Bernoulli$ ανήκει στην εκθετική οικογένεια.

$$
\begin{aligned}
f_Y(y) &= \pi^y (1-\pi)^{1-y}, \ y = 0, 1 \\
&= (\frac{\pi}{1-\pi})^y (1-\pi) \\
&= exp[\ln(\frac{\pi}{1-\pi})^y + \ln(1-\pi)] \\
&= exp[y \ln(\frac{\pi}{1-\pi}) + \ln(1-\pi)]
\end{aligned}
$$

   ανήκει στη εκθετική οικογένεια, με $\tau = 1$, $c(y_i, \tau) = 0$, $\gamma = \ln(\frac{\pi}{1-\pi})$, $b(\gamma) = -\ln(1-\pi)$.

2. Για κάθε παράδειγμα βρείτε την variance function.

   Για τη λογιστική παλινδρόμηση ξέρουμε $Var(Y) = \pi(1-\pi) = \mu(1-\mu)$, κι άρα $v(\mu) = \mu(1-\mu)$.

2.3.1 Εκτίμηση των παραμέτρων με την μέθοδο Μεγίστης πιθανοφάνειας

Η πιθανοφάνεια δίνεται από

$$\begin{aligned} L(\boldsymbol{\beta}, \tau) &= \prod_{i=1}^{n} f_{Y_i}(y_i; \boldsymbol{\beta}, \tau) \\ &= \prod_{i=1}^{n} exp\left[\frac{\gamma_i y_i - b(\gamma_i)}{\tau^2} - c(y_i, \tau)\right]. \end{aligned}$$

Ο λογάριθμος της πιθανοφάνεις ισούται με

$$\begin{aligned} l(\boldsymbol{\beta}, \tau) &= \ln L(\boldsymbol{\beta}, \tau) \\ &= \sum_{i=1}^{n} \frac{\gamma_i y_i - b(\gamma_i)}{\tau^2} - \sum_{i=1}^{n} c(y_i, \tau). \end{aligned}$$

Για την ώρα ας θεωρήσουμε το τ γνωστό. Αυτό συμβαίνει στη λογιστική παλινδρόμηση και στην Poisson παλινδρόμηση. Ο MLE εκτιμητής του $\boldsymbol{\beta}$ είναι η λύση του παρακάτω προβλήματος:

$$\max_{\boldsymbol{\beta}} \frac{1}{\tau^2} \sum_{i=1}^{n} \left(\gamma_i y_i - b(\gamma_i)\right), \tag{2.44}$$

ή ισοδύναμα

$$\max_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(\gamma_i y_i - b(\gamma_i)\right), \tag{2.45}$$

Οι εκτιμητικές εξισώσεις για το $\boldsymbol{\beta}$ προκύπτουν θέτοντας την παράγωγο, ως προς $\boldsymbol{\beta}$, του $\sum_{i=1}^{n} \left(\gamma_i y_i - b(\gamma_i)\right)$ ίσο με $\mathbf{0}_p$. Έχουμε

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} \sum_{i=1}^{n} \left(\gamma_i y_i - b(\gamma_i)\right) &= \sum_{i=1}^{n} y_i \frac{\partial \gamma_i}{\partial \boldsymbol{\beta}} - \frac{\partial b(\gamma_i)}{\partial \boldsymbol{\beta}} \\ &= \sum_{i=1}^{n} y_i \frac{\partial \gamma_i}{\partial \boldsymbol{\beta}} - \frac{\partial b(\gamma_i)}{\partial \gamma_i} \frac{\partial \gamma_i}{\partial \boldsymbol{\beta}} \\ &= \sum_{i=1}^{n} (y_i - \mu_i) \frac{\partial \gamma_i}{\partial \boldsymbol{\beta}} \\ &= \sum_{i=1}^{n} (y_i - \mu_i) \frac{\partial \gamma_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \end{aligned} \tag{2.46}$$

Για να υπολογίσουμε τις παραγώγους στην τελευταία έκφραση της (2.46) έχουμε

$$\begin{aligned}
\frac{\partial \mu_i}{\partial \gamma_i} &= \frac{\partial^2 b(\gamma_i)}{\partial \gamma_i^2} = v(\mu_i) \\
&\Longrightarrow \frac{\partial \gamma_i}{\partial \mu_i} = (\frac{\partial \mu_i}{\partial \gamma_i})^{-1} = \frac{1}{v(\mu_i)} \\
\frac{\partial \mu_i}{\partial \boldsymbol{\beta}} &= \frac{\partial \mu_i}{\partial g(\mu_i)} \frac{\partial g(\mu_i)}{\partial \boldsymbol{\beta}} \\
&= (\frac{\partial g(\mu_i)}{\partial \mu_i})^{-1} \frac{\partial \boldsymbol{x}_i^T \boldsymbol{\beta}}{\partial \boldsymbol{\beta}} \\
&= \frac{1}{g_\mu(\mu_i)} \boldsymbol{x}_i,
\end{aligned}$$

όπου οριζουμε $g_\mu(\mu_i) = g^{'}(\mu_i)$, και $\boldsymbol{x}_i$ είναι το $p \times 1$ διάνυσμα με τις τιμές των συμμεταβλητών για την i-οστή παρατήρηση. Άρα έχουμε

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta}, \tau) &= \frac{1}{\tau^2} \sum_{i=1}^{n} (y_i - \mu_i) \frac{\partial \gamma_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \\
&= \frac{1}{\tau^2} \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{v(\mu_i) g_\mu(\mu_i)} \boldsymbol{x}_i \\
&= \frac{1}{\tau^2} \sum_{i=1}^{n} (y_i - \mu_i) w_i g_\mu(\mu_i) \boldsymbol{x}_i, \qquad (2.47)
\end{aligned}$$

όπου ορίσαμε

$$w_i = \frac{1}{v(\mu_i) g_\mu^2(\mu_i)}.$$

Ορίζουμε τους δύο $n \times n$ διαγώνιους πίνακες $\boldsymbol{W}, \boldsymbol{\Delta}$, με $\boldsymbol{W}$ να έχει ως i-οστό διαγώνιο στοιχείο το w_i, και $\boldsymbol{\Delta}$ να έχει ως i-οστό διαγώνιο στοιχείο το $g_\mu(\mu_i)$. Επίσης ορίζουμε $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\beta})$ το $n \times 1$ διάνυσμα $(\mu_1 \ \mu_2 \ldots \mu_n)^T$, με i-οστό στοιχείο το $\mu_i = g^{-1}(\boldsymbol{x}_i^T \boldsymbol{\beta})$. Η εξίσωση (2.47) γράφεται πιο συνοπτικά ως

$$\frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta}, \tau) = \frac{1}{\tau^2} \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{\Delta} (\boldsymbol{y} - \boldsymbol{\mu}) \qquad (2.48)$$

και ο εκτιμητής μεγίστης πιθανοφάνειας προκύπτει ως λύση των p εκτιμητικών εξισώσεων

$$\begin{aligned}
\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{\Delta} (\boldsymbol{y} - \boldsymbol{\mu}) &= \mathbf{0}_p \\
&\Longleftrightarrow \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{\Delta} \boldsymbol{y} = \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{\Delta} \boldsymbol{\mu} \qquad (2.49)
\end{aligned}$$

Άσκηση

1. Γράψτε αναλυτικά τις εκτιμητικές εξισώσεις για την απλή λογιστική παλινδρόμηση που πρέπει να λυθούν ως προς β_0, β_1.

Για απλή λογιστική παλινδρόμηση έχουμε $\boldsymbol{x}_i = (1\ x_i)^T$, $v(\mu_i) = \mu_i(1-\mu_i)$, $g(\mu_i) = \ln \frac{\mu_i}{1-\mu_i} \implies g_\mu(\mu_i) = \frac{1-\mu_i}{\mu_i}\frac{1}{(1-\mu_i)^2} = \frac{1}{\mu_i(1-\mu_i)}$. Άρα

$$w_i = \frac{1}{\mu_i(1-\mu_i)\frac{1}{(\mu_i(1-\mu_i)})^2} = \mu_i(1-\mu_i)$$

και βάσει της (2.47) οι εκτιμητικές εξισώσεις γράφονται ως

$$\begin{aligned}
\sum_{i=1}^{n}(y_i-\mu_i)w_i g_\mu(\mu_i)\boldsymbol{x}_i = \mathbf{0}_2 &\implies \\
\sum_{i=1}^{n}(y_i-\mu_i)\mu_i(1-\mu_i)\frac{1}{\mu_i(1-\mu_i)}\boldsymbol{x}_i = \mathbf{0}_2 &\implies \\
\sum_{i=1}^{n}(y_i-\mu_i)\boldsymbol{x}_i = \mathbf{0}_2 &\implies \\
\sum_{i=1}^{n}(y_i-\mu_i)\begin{bmatrix}1\\ x_i\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}&.
\end{aligned}$$

Οι 2 εκτιμητικές εξισώσεις είναι

$$\begin{aligned}
\sum_{i=1}^{n}(y_i-\mu_i) &= 0\\
\sum_{i=1}^{n}x_i(y_i-\mu_i) &= 0.
\end{aligned}$$

Βλέπουμε πως η μορφή των δύο εκτιμητικών εξισώσεων στην απλή λογιστική παλινδρόμηση είναι ίδια με τη μορφή που έχουμε στην κλασσική απλή κανονική γραμμική παλινδρόμηση, δηλαδή το άθροισμα των υπολοίπων ισούται με 0 και το άθροισμα των γινομένων των υπολοίπων με τη συμμεταβλητή ισούται με 0. Όμως σε αντίθεση με τη κλασσική απλή κανονική γραμμική παλινδρόμηση το σύστημα των 2 εξισώσεων δεν είναι γραμμικό ως προς τις παραμέτρους. Αυτό οφείλεται στο ότι τα μ_i δεν είναι γραμμικές συναρτήσεις των β_0, β_1. Η πλήρης μορφή των 2 εξισώσεων είναι:

$$\sum_{i=1}^{n}\left(y_i - \frac{exp(\beta_0+\beta_1 x_i)}{1+exp(\beta_0+\beta_1 x_i)}\right) = 0$$
$$\sum_{i=1}^{n} x_i\left(y_i - \frac{exp(\beta_0+\beta_1 x_i)}{1+exp(\beta_0+\beta_1 x_i)}\right) = 0.$$

Φαίνεται καθαρά πως το παραπάνω σύστημα είναι μη γραμμικό ως προς τις παραμέτρους και λύνεται αποκλειστικά με αριθμητικές μεθόδους.