Άσκηση (Ανάλυση διακύμανσης κατά 2 παράγοντες)

Σε ένα πείραμα θέλαμε να εξετάσουμε την επίδραση 4 διαφορετικών καυσίμων και διαφορετικών τρόπων εκτόξευσης πάνω στο βεληνεκές κάποιου τύπου ρουκέτας. Τα δεδομένα που έχουμε στη διάθεσή μας είναι τα ακόλουθα όπου το βεληνεκές μετριέται σε km.

Τρόπος εκτόξευσης	Καύσιμο 1	2	3	4
1	45,9	57,6	52,2	41,7
2	46	51	50,1	38,8
3	45,7	56,9	55,3	48,1

ΛΥΣΗ

Y : βεληνεκές → Παράγοντας B : Καύσιμο → 1ο καυσ., 2ο, 3ο, 4ο

Y : βεληνεκές → Παράγοντας A : Τρόπος εκτόξευσης → 1ος τρόπος, 2ος, 3ος

Μοντέλο Ανάλυσης Διακύμανσης κατά 2 παράγοντες

$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$, $i=1,\dots,3$ $(I=3)$, $j=1,\dots,4$ $(J=4)$

Υποθέτουμε ότι ισχύουν οι υποθέσεις για τα σφάλματα και οι πλευρικές συνθήκες $\sum_{i=1}^{3} \alpha_i = \sum_{j=1}^{4} b_j = 0$

ΠΙΝΑΚΑΣ ΑΝΑΔΙΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

Πηγή Μεταβλητότητας	SS	β.ε	MS	F-πηλίκα
Παράγοντας A	50,582	2	25,426	$F_A = \frac{MS_A}{MS_{res}} = 4,428$
Παράγοντας B	293,702	3	97,6	
Υπόλοιπα	34,455	6	5,743	$F_B = \frac{MS_B}{MS_{res}} = 17,048$
Ολική Μεταβλητότητα	379,009	11		

$H_0^A : \alpha_1 = \alpha_2 = \alpha_3 = 0$ & $H_0^B : b_1 = b_2 = b_3 = b_4$

Για την $1^η$ Υπόθεση:

Αφού $F_A = 4,428 < \begin{cases} F_{2,6,0.01} = 10,9 \\ F_{2,6,0.05} = 5,14 \end{cases}$ τότε η H_0^A

δεν απορρίπτεται σε επίπεδο σημαντικότητας 0.01 ή 0.05

Άρα, οι τρόποι επιτύγχανσης μπορούν να θεωρηθούν ισοδύναμοι

Δηλαδή, δεν ασκούν σημαντική επίδραση στην Y.

Για την $2^η$ Υπόθεση

Αφού $F_B = 17,048 > \begin{cases} F_{3,6,0.01} = 9,78 \\ F_{3,6,0.05} = 4,76 \end{cases}$ τότε η H_0^B

απορρίπτεται σε επίπεδο σημαντικότητας 0.01 ή 0.05

Άρα, υπάρχει κάποιο ή κάποια καύσιμα που ασκούν σημαντικότερη επίδραση στο βελτινευές

Αυτό θα το διαπιστώσαμε μέσω πολ/λων συγκρίσεων και συγκεκριμένα με τη μέθοδο της ΕΣΔ.

$E\Sigma\Delta = \sqrt{\frac{2 \cdot MSres}{I}} \cdot t_{\frac{a}{2},(I-1)(J-1)} \Longrightarrow E\Sigma\Delta = 4,793$

$\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2} = -9,3$	$\bar{Y}_{\cdot 2} - \bar{Y}_{\cdot 3} = 2,634$
$\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 3} = -6,66$	$\bar{Y}_{\cdot 2} - \bar{Y}_{\cdot 4} = 12,3$
$\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 4} = 3$	$\bar{Y}_{\cdot 3} - \bar{Y}_{\cdot 4} = 9,666$

Επειδή, $|\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}| = 93 > 4,793 \Rightarrow$ η $H_0: b_1 = b_2$ απορ. κάποιο από τα δύο επίπεδα είναι πιο καλό και αφου $\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2} = -9,3 < 0 \Rightarrow \bar{Y}_{\cdot 1} < \bar{Y}_{\cdot 2} \Rightarrow$ το 2^o καυσιμο ασκεί μεγαλύτερη επίδραση στο βελτιωτες απο το 1^o καυσιμο

Επειδή, $|\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 3}| = 6,66 > 4,793 \Rightarrow$ η $H_0: b_1 = b_3$ απορ. και αφου $\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 3} < 0 \Rightarrow$ Το 3^o καύσιμο ασκεί σημαντικότερη επίδραση απο το 1^o καυσιμο

Επειδή, $|\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 4}| = 3 < 4,793 \rightarrow \beta_1 \equiv \beta_4$

Επειδή, $|\bar{Y}_{\cdot 2} - \bar{Y}_{\cdot 3}| = 2,634 > 4,793 \rightarrow \beta_2 > \beta_4$

Επειδή, $|\bar{Y}_{\cdot 2} - \bar{Y}_{\cdot 4}| = 12,3 > 4,793 \rightarrow \beta_2 > \beta_4$

Επειδή, $|\bar{Y}_{\cdot 3} - \bar{Y}_{\cdot 4}| = 9,666 > 4,793 \rightarrow \beta_3 > \beta_4$

$$\left.\begin{array}{l} 1 \equiv 4 < 2 \text{ και } 3 \\ 2 \equiv 3 > 1 \text{ και } 4 \end{array}\right\} \quad 1 \equiv 4 < 2 \equiv 3$$

Άρα, θα επιλέξαμε το 2 ή 3 τα οποία είναι ισοδύναμα και καλύτερα απο τα άλλα.

Άσκηση 7 (Φυλλάδιο)

Y_1 ανεξ. της Y_2

$E(Y_1) = \theta$ και $E(Y_2) = 2\theta$, $\theta \in \mathbb{R}$

Να βρεθεί ΕΕΤ για τη θ και να υπολογιστεί το SSres.

ΛΥΣΗ

$$\left.\begin{array}{l} E(Y_1) = \theta \\ E(Y_2) = 2\theta \end{array}\right\} \quad E\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} E(Y_1) \\ E(Y_2) \end{pmatrix} = \begin{pmatrix} \theta \\ 2\theta \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \theta$$

Θεωρώ το γραμμικό μοντέλο $\underset{\sim}{Y} = X \underset{\sim}{b} + \underset{\sim}{\varepsilon}$ τέτοιο ώστε

$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$ και $X = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $b = \theta$, $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$ με $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$

Οι ΕΕΤ για την θ είναι ο εξής:

$$\hat{\theta} = \left((1,2)\begin{pmatrix}1\\2\end{pmatrix} \right)^{-1} \cdot (1,2) \cdot \begin{pmatrix}Y_1\\Y_2\end{pmatrix} = \frac{1}{5} \cdot (Y_1 + 2Y_2)$$

$$SS_{res} = \underset{\sim}{Y}^T \cdot \underset{\sim}{Y} - \underset{\sim}{\hat{b}}^T \cdot X^T \cdot Y =$$

$$= (Y_1, Y_2)\begin{pmatrix}Y_1\\Y_2\end{pmatrix} - \frac{1}{5}(Y_1+2Y_2)(1,2)\begin{pmatrix}Y_1\\Y_2\end{pmatrix} =$$

$$= Y_1^2 + Y_2^2 - \frac{1}{5}(Y_1+2Y_2)^2.$$

Άσκηση 9 (φυλλάδιο)

Αν σε ένα γραμμικό μοντέλο υπάρχει σταθερός όρος τότε το $SS_{res} = 0$

ΛΥΣΗ.

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (Y_i - \hat{Y}_i) \overset{\text{Ο.Δ.}}{=} 0$$

Κανονικές εξισώσεις : $X^T \cdot X \cdot \underset{\sim}{b} = X^T \cdot \underset{\sim}{Y}$ και ικανοποιούνται στο $\hat{\underset{\sim}{b}}$

Τότε $X^T \cdot X \cdot \hat{b} = X^T \cdot Y \Rightarrow X^T Y - X^T \cdot X \cdot \hat{\underset{\sim}{b}} = 0 \Rightarrow X^T(\underset{\sim}{Y} - X\hat{\underset{\sim}{b}}) = \underset{\sim}{0} \Rightarrow$

$$\Rightarrow (Y - X\hat{\underset{\sim}{b}})^T \cdot X = \underset{\sim}{0} \Rightarrow (Y - \hat{Y})^T \cdot X = 0 \Rightarrow$$

$$\Rightarrow (Y_1 - \hat{Y}_1, \dots, Y_n - \hat{Y}_n) \cdot \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix} = \underset{\sim}{0} \Rightarrow$$

$$\Rightarrow \sum_{i=1}^{n} (Y_i - \hat{Y}_i) = 0 \quad \text{(από το γινόμενο με την } 1^{\text{η}} \text{ στήλη)}$$

Άσκηση 10 (φυλλάδιο)

Έστω το μ.π.γ.π $\underset{\sim}{Y} = X\underset{\sim}{b} + \underset{\sim}{\varepsilon}$

Νδο:

a) $\sum_{i=1}^{n} \hat{Y}_i (Y_i - \hat{Y}_i) = 0$ και β) $\sum_{i=1}^{n} Var(\hat{Y}_i) = \sigma^2 \cdot (p+1)$

ΛΥΣΗ

α) Όπως και στην άσκηση 9

$$(\underset{\sim}{y} - X\hat{\underset{\sim}{b}})^T \cdot X = \underset{\sim}{0} \Rightarrow (\underset{\sim}{y} - X\hat{\underset{\sim}{b}})^T X \cdot \hat{\underset{\sim}{b}} = \underset{\sim}{0} \Rightarrow$$

$\hat{\underset{\sim}{y}} = X \cdot \hat{\underset{\sim}{b}}$

$$\Rightarrow (\underset{\sim}{y} - \hat{\underset{\sim}{y}}) \cdot \hat{\underset{\sim}{y}} = \underset{\sim}{0} \Rightarrow \sum_{i=1}^{n} (y_i - \hat{y}_i) \cdot \hat{y}_i = 0.$$

Άρα, το $\hat{\underset{\sim}{y}} \perp \underset{\sim}{e}$.

β) Το $\sum_{i=1}^{n} Var(\hat{Y}_i)$ είναι το ίχνος του $Var(\hat{\underset{\sim}{y}})$ γιατί

$$Var(\hat{\underset{\sim}{y}}) = \begin{pmatrix} Var(\hat{y}_1) & & Cov(\hat{y}_i, \hat{y}_j) \\ & Var(\hat{y}_2) & \\ Cov(\hat{y}_j, \hat{y}_i) & & \ddots \\ & & & Var(\hat{y}_n) \end{pmatrix}$$

$\hat{\underset{\sim}{y}} = X\hat{\underset{\sim}{b}}$

$$Var(\hat{\underset{\sim}{y}}) = Var(X\hat{\underset{\sim}{b}}) = X \cdot Var(\hat{\underset{\sim}{b}}) X^T = X(\sigma^2 (X^T X)^{-1}) X^T =$$
$$= \sigma^2 \cdot X(X^T \cdot X)^{-1} X^T$$

$$\sum_{i=1}^{n} Var(\hat{Y}_i) = trace[Var(\hat{\underset{\sim}{y}})] = trace[\sigma^2 X(X^T X)^{-1} \cdot X^T] =$$

$$= \sigma^2 \cdot trace(\underbrace{X}_{A}\underbrace{(X^T X)^{-1} X^T}_{B}) \quad \underline{trace(AB^T) = trace(A^T B)}$$

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

$$= \sigma^2 \cdot trace(X^T \cdot X^{TT} ((X^T X)^{-1})^T) =$$

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

$$= \sigma^2 \, trace(I_{p+1}) = \sigma^2 \cdot (p+1)$$

Άσκηση 11 ($\$o\$$)

Για το μοντέλο $Y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + \varepsilon_i$, $\forall i = 1, \ldots, n$ με $\varepsilon_i \sim N(0, \sigma^2)$ και ανεξάρτητα, οι κανονικές εξισώσεις είναι οι εξής:

$$\begin{cases} 10\hat{b}_0 + 2\hat{b}_1 - 6\hat{b}_2 = 4 \\ 2\hat{b}_0 + 2\cdot\hat{b}_1 = 6 \\ 6\hat{b}_0 + 5\cdot\hat{b}_2 = 7 \end{cases}$$

α) Αν $n=10$ και $\sum_{i=1}^{n} Y_i^2 = 107$ να υπολογιστούν οι ΕΕΤ $\hat{\underset{\sim}{b}}$ και το MSres

β) Να ελεγχθεί με F τεστ ότι $H_0: b_1 = 2b_2$

γ) Να κατασκευαστεί ένα t τεστ για τον έλεγχο $H_0: b_1 = 2b_2$

ΛΥΣΗ

α) $\underset{\sim}{Y} = X\underset{\sim}{b} + \underset{\sim}{\varepsilon}$

Οι κανονικές εξισώσεις είναι: $X^T X \underset{\sim}{b} = X^T \underset{\sim}{Y}$ ή $X^T X \hat{\underset{\sim}{b}} = X^T \underset{\sim}{Y}$.

Συνεπώς από την υπόθεση:

$$X^T X = \begin{bmatrix} 10 & 2 & -6 \\ 2 & 2 & 0 \\ -6 & 0 & 5 \end{bmatrix} \quad \text{και} \quad X\underset{\sim}{Y} = \begin{bmatrix} 4 \\ 6 \\ 7 \end{bmatrix}$$

και γνωστό από τη θεωρία ότι $\hat{\underset{\sim}{b}} = (X^T X)^{-1} X^T Y$ ①

$$(X^T X)^{-1} = \frac{1}{8}\begin{bmatrix} 10 & -10 & 12 \\ -10 & 14 & -12 \\ 12 & -12 & 16 \end{bmatrix}$$

Έτσι, στην ① βρίσκουμε $\hat{\underset{\sim}{b}} = \begin{pmatrix} 8 \\ -5 \\ 11 \end{pmatrix}$

$$MS_{res} = \frac{SS_{res}}{n-p-1} = \frac{1}{10-2-1} \cdot (\underset{\sim}{Y}^T \underset{\sim}{Y} - \hat{\underset{\sim}{b}}^T \cdot X^T \cdot \underset{\sim}{Y}) =$$

$$= \frac{1}{7}\left(107 - (8 \; -5 \; 11)\begin{pmatrix} 4 \\ 6 \\ 7 \end{pmatrix}\right) = 4.$$

β) F τεστ → Γενική γραμ. υπόθεση $(A\underset{\sim}{b} = \underset{\sim}{c})$

$$A \cdot \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} = \underset{\sim}{c}$$

Θεωρώ $A = (0, 1, -2)$, $\underset{\sim}{c} = 0$

Τότε το $A\underset{\sim}{b} = c \Leftrightarrow b_1 - 2b_2 = 0 \Leftrightarrow b_1 = 2b_2$

Άρα, $H_0: b_1 = 2b_2 \Leftrightarrow H_0: A\underset{\sim}{b} = \underset{\sim}{c}$ με $A = (0,1,-2)_{1\times 3}$, $\underset{\sim}{c} = 0$

F-τεστ για τον έλεγχο της $H_0: A\underset{\sim}{b} = \underset{\sim}{c}$:

$$F = \frac{(A\hat{\underset{\sim}{b}}-\underset{\sim}{c})^T (A\cdot(X^TX)^{-1}\cdot A^T)^{-1}\cdot(A\hat{\underset{\sim}{b}}-\underset{\sim}{c})}{SSres}\cdot\frac{n-p-1}{q}$$

με $F \sim F_{q,\,n-p-1}$ υπό την υπόθεση Ho με κ.π

$F \geq F_{q,\,n-p-1,\,\frac{\alpha}{2}}$

Εφαρμογή : για $n=10$, $p=2$, $q=1$ ← πλήθος γραμμών του A.

γ) κατασκευή t-τεστ

$Ho: b_1 = 2b_2 \Leftrightarrow Ho: \underset{\sim}{\alpha}^T\cdot\underset{\sim}{b} = 0$ όπου $\underset{\sim}{\alpha} = \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix}$

Θα στηριχτώ στον εκτιμητή των παρακάτω που υποστηρίζουν στην Ho. Δηλ. στους εκτιμητές $\hat{\underset{\sim}{b}}$.

Γνωστό ότι: $\hat{\underset{\sim}{b}} \sim N_3\left(\begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}, \sigma^2\cdot(X^TX)^{-1}\right) \Rightarrow$

$\Rightarrow \underset{\sim}{\alpha}^T\cdot\hat{\underset{\sim}{b}} \sim N(\underset{\sim}{a}^T\underset{\sim}{b},\, a^T\cdot(\sigma^2(X^TX)^{-1})\underset{\sim}{a}) \Rightarrow$

$\Rightarrow \hat{b}_1 - 2\hat{b}_2 \sim N(b_1 - 2b_2,\, 10.25\,\sigma^2)$

και υπό την Ho : $b_1 = 2b_2$

$\hat{b}_1 - 2\hat{b}_2 \sim N(0,\, 10.25\,\sigma^2)$ υπό την Ho

$\frac{\hat{b}_1 - 2b_2}{\sigma\cdot\sqrt{10,25}} \sim N(0,1)$ υπό την Ho

Το σ άγνωστο άρα διαιρώ με το $\frac{SSres}{\sigma^2} \sim \chi^2_{n-p-1}$

Θεωρώ λοιπόν

$$t = \frac{\frac{\hat{b}_1 - 2\hat{b}_2}{\sigma\sqrt{10,25}}}{\sqrt{\frac{SSres}{\sigma^2}/(n-p-1)}} \overset{Ho}{\sim} t_{n-p-1} \quad \text{ή} \quad t = \frac{\hat{b}_1 - 2\hat{b}_2}{\sqrt{10,25\ MSres}} \sim t_7 \text{ υπό την Ho}$$

Κρίσιμη περιοχή: Μεγάλες τιμές του t, $|t| > c$

$\alpha = P(\text{Απορ. Ho} \mid \text{Ho αληθής}) \Rightarrow \ldots \Rightarrow c = t_{\alpha/2,\,7}$

Άρα, κ.π $|t| \geq t_{\frac{\alpha}{2},7}$.