

ΕΞΕΤΑΣΕΙΣ

Μάθημα : ΑΝΑΛΥΣΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΕΠΙΣΤΗΜΟΝΟΛΟΓΙΚΗ ΕΞΕΤΑΣΗ Ακαδημ. Έτους 2003-04

***** Διάρκεια Εξέτασης : 2.30 ώρες *****

ΖΗΤΗΜΑ 1

A) Εστω σημειακή πρόβλεψη $\hat{Y}_{x_0} = \mathbf{x}_0' \hat{\beta}$, να βρεθεί η μέση τιμή $E(Y_{x_0} - \hat{Y}_{x_0})$ και η διασπορά

$V(Y_{x_0} - \hat{Y}_{x_0})$ με Y_{x_0} παρατήρηση στο σημείο \mathbf{x}_0 .

B) Θεωρείται ότι ο χρόνος Y που πρέπει πωλητής να διανύσει προκειμένου να παραδώσει παγωτά σε καταστήματα χωριών συνδέεται γραμμικά με τον αριθμ. κουτιών (X_1), την απόσταση (X_2), τον αριθμ. καταστημάτων (X_3) και τον αριθμ. χωριών (X_4).

I) Με βάση τον παρακάτω πίνακα και για μέγεθος δείγματος $n=25$ να βρεθεί το καταλληλότερο μοντέλο.

Πλήθος μεταβλητών στο μοντέλο	R^2	C_p	S	x	x	x	x
				1	2	3	4
1	96.4	50.5	3.0934	X			
1	69.8	577.1	8.9494			X	
2	98.3	14.6	2.1703	X			X
2	98.1	18.4	2.2880	X	X		
3	99.0	3.6	1.7377	X	X		X
3	98.5	12.4	2.0762	X	X	X	
4	99.0	5.0	1.7550	X	X	X	X

$$(S = (e'e / (n-k-1))^{1/2})$$

II) (i) Προσαρμόζεται το μοντέλο $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ στα δεδομένα. Να βρεθεί η σημειακή πρόβλεψη της παρατήρησης Y_{x_0} για $\mathbf{x}_0 = (1, X_1, X_2) = (1, 4, 200)$.

(ii) Πιστεύετε ότι η εισαγωγή της μεταβλητής X_4 στο μοντέλο βελτιώνει την προσαρμογή του ;

(iii) Να κατασκευαστεί ένα 0.99 διάστημα εμπιστοσύνης της β_1 του μοντέλου που επιλέξατε.

[Δίνονται (χωρίς την X_4): $\hat{\beta}_1 = 2.744$, $\hat{\beta}_2 = 0.013$, $c_{11} = 0.050$, $se(\hat{\beta}_2) = 0.0028$ και (με την X_4): $\hat{\beta}_1 = 2.535$, $\hat{\beta}_2 = 0.009$, $\hat{\beta}_3 = 2.599$, $c_{11} = 0.041$, $se(\hat{\beta}_2) = 0.0023$, $V\hat{a}r(\hat{\beta}_4) = 0.394$]. (Βαθμ.3.5)

ΖΗΤΗΜΑ 2

Εστω το γενικό γραμμικό μοντέλο $y = X\beta + \varepsilon$

A) (i) Να βρεθεί η διασπορά του υπολοίπου e_i και η συνδιασπορά μεταξύ των υπολοίπων

e_i και e_j , όπου X ο πίνακας σχεδιασμού και $H = X(X'X)^{-1}X'$.

(ii) Αν $\hat{\beta}$ είναι η εκτιμήτρια ελαχίστων τετραγώνων της β δείξτε ότι $\hat{\beta} = \beta + R\varepsilon$, όπου

$R = (X'X)^{-1}X'$. Είναι η $\hat{\beta}$ αμερόληπτη εκτιμήτρια της β ;

B) Εστω k ο αριθμός των επεξηγηματικών μεταβλητών. Δείξτε ότι η ελεγχοσυνάρτηση για την

προσαρμογή του μοντέλου γράφεται ως $F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$. Αν $n=20$, $k=4$ και $R^2=0.90$ και

$\alpha=0.05$ τι συμπεραίνουμε για τη σχέση μεταξύ y και τις 4 επεξηγηματικές μεταβλητές;

(Βαθμ.3.0)

ΖΗΤΗΜΑ 3

A) Πειραματιστής ισχυρίζεται ότι η αντοχή συνθετικών ινών σχετίζεται με την επί τοις % περιεκτικότητα τους σε βαμβάκι. Δίνονται 4 επίπεδα ποσοστών βαμβακιού με μετρήσεις της αντοχής σε 5 τεμάχια ινών ανά επίπεδο.

Ποσοστό βαμβακιού			
(15%)	(20%)	(25%)	(30%)
7	12	19	7
7	17	25	10
15	12	22	11
11	18	19	15
9	18	23	11

- (i) Με βάση τη μέθοδο της ανάλυσης διασποράς να εξετάσετε αν το ποσοστό του βαμβακιού στα τεμάχια ινών επδρά στην αντοχή τους.
- (ii) Περιγράψτε σύντομα τη διαδικασία που ακολουθείται προκειμένου να απαντηθεί το προηγούμενο ερώτημα με την προσαρμογή ενός μοντέλου γραμμικής παλινδρόμησης. Θα έχετε τα ίδια αποτελέσματα.

B) Δείξτε ότι στο μοντέλο $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, $i=1, \dots, k$, $j=1, \dots, r$, γενικώς ισχύει

$$E(SSA) = (k-1)\sigma^2 + r \sum_{i=1}^k \alpha_i^2.$$

$$(\text{Δίνεται ότι } SSA = \sum_{i=1}^k r(\bar{y}_i - \bar{y})^2)$$

(Βαθμ. 3.5)