Μάθημα : ΑΝΑΛΥΣΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
Εξετάσεις Φεβρουαρίου 2009-10

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

ΖΗΤΗΜΑ 1 (Βαθμ. 3)

(i) Δείξτε ότι στο απλό γραμμικό μοντέλο $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $i = 1, \ldots, n$ το διαγώνιο στοιχείο του πίνακα προβολής $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ είναι $h_{ii} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i = \dfrac{1}{n} + \dfrac{(x_i - \bar{x})^2}{\sum_{j=1}^{n}(x_j - \bar{x})^2}$. Πώς μας βοηθούν τα h_{ii} στην αναγνώριση σημείων επιρροής σε μια ανάλυση παλινδρόμησης;

[Δίνεται $\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$]

(ii) Δείξτε ότι στο γενικό γραμμικό μοντέλο $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ ισχύει ότι $\mathbf{X}'\mathbf{e} = \mathbf{0}$ και από αυτό βρείτε ότι στο απλό γραμμικό μοντέλο ισχύει $\sum_{i=1}^{n} e_i = 0$ καθώς και ότι $\sum_{i=1}^{n} x_i e_i = 0$.

ΖΗΤΗΜΑ 2 (Βαθμ. 3.5)

Πειραματιστής θέλει να εξετάσει τη σχέση μεταξύ της μεταβλητής y (χμ/λίτρο) και έξι επεξηγηματικών μεταβλητών $X_1, X_2, \ldots, X_6$ που αφορούν χαρακτηριστικά 29 οχημάτων.

(i) Αξιοποιώντας τον έλεγχο-F καθώς και τα κριτήρια R^2, C_p, να βρείτε το καλύτερο μοντέλο

Μεταβλητές	R^2	Mallows C_p	S_{Y_x}	x 1	x 2	x 3	x 4	x 5	x 6
1	76.0	2.4	3.1220	X					
1	72.7	6.4	3.3309						X
2	76.5	3.9	3.1491	X		X			
2	77.1	3.1	3.1040					X	X
3	80.3	1.3	2.9323			X		X	X
3	78.2	3.9	3.0902	X				X	X
4	80.6	3.0	2.9721		X	X		X	X
4	80.4	3.3	2.9895			X	X	X	X
5	80.6	5.0	3.0324	X	X	X		X	X
5	80.6	5.0	3.0332		X	X	X	X	X
6	80.6	7.0	3.0976	X	X	X	X	X	X

[**Δίνεται** $S_{Y_x} = \left(e'e / (n-k-1) \right)^{1/2}$]

(ii) Στη συνέχεια θεωρώντας το απλό γραμμικό μοντέλο με τη σταθερά και τη μεταβλητή X_1 μόνο, να κατασκευαστεί ένα 0.99 - διάστημα εμπιστοσύνης για την παράμετρό της [$\hat{\beta}_1 = -0.047056$] και
(iii) με βάση το 17-οστό διαγώνιο στοιχείο του πίνακα **H** ελέγξτε αν η 17^{η} παρατήρηση της μεταβλητής X_1 αποτελεί σημείο επιρροής στο μοντέλο.

[Δίνονται οι τιμές της X_1:

350	350	250	351	225	440	231	262	89.7	96.9	350	85.3
171	258	140	302	**500**	440	350	318	231	360	96.9	460

133.6 318 351 351 360, $\sum_{i=1}^{29} x_i^2 = 2716901.36$]

ΖΗΤΗΜΑ 3 (Βαθμ. 3.5)

Ι) Τυχαία επιλεγμένο δείγμα 20 τούβλων από την ημερήσια παραγωγή ενός εργοστασίου κατανεμήθηκε τυχαία σε τέσσερις διαφορετικές συνθήκες αποθήκευσης. Μετά από μια εβδομάδα μετρήθηκε η επί τοις % περιεκτικότητα τους σε νερό:

Συνθήκες			
1	**2**	**3**	**4**
7.1	7.0	8.1	7.9
8.3	5.4	6.4	9.6
7.6	7.4	7.1	10.0
8.3	6.4	6.8	7.1
8.2	5.9	7.7	8.5

(i) Μέσω ενός μοντέλου παλινδρόμησης να εξετάσετε αν υπάρχουν διαφοροποιήσεις μεταξύ των τεσσάρων συνθηκών αποθήκευσης (SSR=13.26).

(ii) Υιοθετώντας την κωδικοποίηση

$$x_1 = \begin{cases} 1, \text{ αν συνθήκη 1} \\ 0, \text{ αλλιώς} \end{cases} \quad x_2 = \begin{cases} 1, \text{ αν συνθήκη 2} \\ 0, \text{ αλλιώς} \end{cases} \quad x_3 = \begin{cases} 1, \text{ αν συνθήκη 3} \\ 0, \text{ αλλιώς} \end{cases}$$

να συμπληρωθεί ο παρακάτω πίνακας

Μεταβλητές	$\hat{\beta}$	se($\hat{\beta}$)	t	p-τιμή
Σταθερά	8.6200	0.3762	22.91	<0.001
x1	-0.7200	0.5321		
x2	-2.2000	0.5321		
x3	-1.4000	0.5321		

Με βάση αυτά τα αποτελέσματα τι συμπεραίνετε για τις επί μέρους διαφορές μεταξύ των συνθηκών αποθήκευσης;

ΙΙ) Δώστε τον ορισμό ενός μοντέλου παλινδρόμησης Poisson με δύο επεξηγηματικές μεταβλητές και γράψτε τη συνάρτηση πιθανοφάνειάς του.