Μάθημα : ΑΝΑΛΥΣΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΕΞΕΤΑΣΕΙΣ ΧΕΙΜΕΡΙΝΟΥ ΕΞΑΜΗΝΟΥ 2013-14

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

ΖΗΤΗΜΑ 1

Έστω γενικό γραμμικό μοντέλο $\mathbf{y=X\beta+\varepsilon}$, με $\varepsilon \sim N_n(\mathbf{0},\sigma^2 I_n)$, όπου X ο πίνακας σχεδιασμού με k επεξηγηματικές μεταβλητές.

(i) Να βρεθεί η διασπορά $V(e_i)$ και συνδιακύμανση $\operatorname{cov}(e_i,e_j)$, $i \neq j$, των υπολοίπων e_i, $i=1, \ldots, n$.

(ii) Κάνοντας χρήση της ιδιότητας $E(\mathbf{y'Ay}) = \operatorname{tr}(\mathbf{AV}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$, όπου $\boldsymbol{\mu} = E(\mathbf{y})$, $V=V(\mathbf{y})$ και $\mathbf{A=(I-H)}$, $\mathbf{H=X(X'X)^{-1}X'}$, δείξτε ότι $S^2=\dfrac{SSE}{n-k-1}$ είναι μια αμερόληπτη εκτιμήτρια της σ^2 και ότι για το απλό γραμμικό μοντέλο $E(y)=\beta_0+\beta_1 x$ αυτή ανάγεται στην $S^2 = \dfrac{1}{n-2}\left\{S_{yy} - \dfrac{S_{xy}^2}{S_{xx}}\right\} = \dfrac{1}{n-2}S_{yy}\left\{1-r_{xy}^2\right\}$, όπου SSE το άθροισμα τετραγώνων λόγω σφάλματος από την προσαρμογή του μοντέλου.

(iii) Έστω το απλό γραμμικό μοντέλο $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ με $V(\varepsilon_i) = \sigma^2 x_i^2$, $i=1, \ldots, n$. Πώς θα μετασχηματίσουμε το μοντέλο έτσι ώστε να σταθεροποιηθεί η διασπορά του τυχαίου σφάλματος;

(Βαθμ. 3.0)

ΖΗΤΗΜΑ 2

(i) Να προσαρμοστεί το μοντέλο $E(y_x)=\beta_0+\beta_1 x$ στα ακόλουθα δεδομένα και να βρεθεί ο συντελεστής προσδιορισμού R^2.

X	1.6	1.8	1.4	2.0	1.2	2.2	1.0	2.4
Y	8.5	6.0	7.0	5.0	10.0	2.0	19	2.0

(ii) Να κατασκευαστεί ένα 0.99-διάστημα εμπιστοσύνης πρόβλεψης της $\mu_{\mathbf{x}_0} = E(y_{\mathbf{x}_0})$, όπου $y_{\mathbf{x}_0}$ μια νέα παρατήρηση, όταν $\mathbf{x}_0=(x_{00},x_{01})'=(1, 1.5)'$. [Δίνεται $\mathbf{x}'_0(\mathbf{X'X})^{-1}\mathbf{x}_0 = 1/n + (x_{01}-\bar{x})^2/S_{xx}$].

(iii) Με βάση ένα στατιστικό έλεγχο υποθέσεων και τους συντελεστές προσδιορισμού R^2, θεωρείτε ότι βελτιώνεται η προσαρμογή αν στο μοντέλο εισαχθεί και η μεταβλητή x^2 ;

[Δίνεται $\sum_{i=1}^{n} x_i y_i = 84.4$, $\sum_{i=1}^{n} x_i^2 = 24.8$, $\sum_{i=1}^{n} y_i^2 = 651.25$. Δεδομένου ότι στο μοντέλο υπάρχει ήδη η μεταβλητή x, δίνεται ότι $\hat{\beta}_2=7.515$, $\sqrt{c_{22}}=1.93$ και SSE=26.537]. **(Βαθμ. 2.5)**

$\hat{y}_{x_0} \pm t_{n-2,a/2} \cdot s\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\Sigma(x_i-\bar{x})^2}}$

ΖΗΤΗΜΑ 3

A) Δώστε μια σύντομη περιγραφή **δύο** διαδικασιών με βήματα για την επιλογή ενός μοντέλου παλινδρόμησης. Θα καταλήξουν στο ίδιο μοντέλο;

B) Εξετάζεται η γραμμική παλινδρόμηση της Y σε σχέση με τις επεξηγηματικές μεταβλητές X_1, X_2, X_3 ,X_4. Με βάση τον παρακάτω πίνακα και για μέγεθος δείγματος n=13, να βρεθεί το καταλληλότερο μοντέλο, κάνοντας χρήση και ενός ελέγχου F, όπου κρίνετε ότι είναι απαραίτητος.

Πλήθος μεταβλητών στο μοντέλο	R^2	C_p (Mallows)	S	X_1	X_2	X_3	X_4
1	67.5	138.7	8.9639				X
1	66.6	142.5	9.0771		X		
2	97.9	2.7	2.4063	X	X		
2	97.2	5.5	2.7343	X			X
3	98.2	3.0	2.3087	X	X		X
3	98.2	3.0	2.3121	X	X	X	
4	98.2	5.0	2.4460	X	X	X	X

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

(Βαθμ. 2.5)

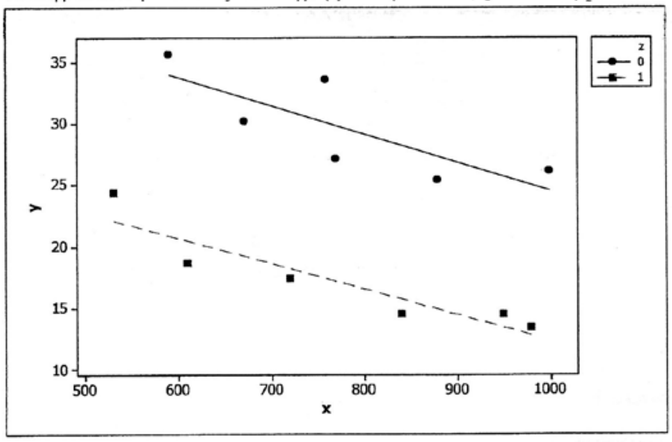
$\hat{b}_1$ $\underline{S_{xy}}$ $\hat{b}_0 = y - \hat{b}_1 x$

ΖΗΤΗΜΑ 4 (Επιλέξτε 1 από τα ακόλουθα 3 ερωτήματα) **(Βαθμ. 2.0)**

Α) Σε δεδομένα που αφορούν την ταχύτητα x δύο τύπων εργαλείων και την αντίστοιχη διάρκεια ζωής τους y (ώρες), προσαρμόζεται το μοντέλο $E(y_x)=\beta_0+\beta_1 x+\beta_2 z+\beta_3 w$, όπου z μια δείκτρια μεταβλητή (z=1, όταν το εργαλείο είναι του τύπου Α και z=0, όταν είναι του τύπου Β) και w=xz εκφράζει την αλληλεπίδραση μεταξύ των δύο τύπων εργαλείων και την ταχύτητά τους x.

(i) Πιστεύετε ότι διαφοροποιούνται οι κλίσεις των μοντέλων για τους δύο τύπους εργαλείων με την ταχύτητα x; [Δίνονται $\hat{\beta}_3=0.002214$, $se(\hat{\beta}_3)=0.009137$, SSE(x,z,w)= 43.59].

(ii) Θεωρώντας ότι δεν υπάρχει αλληλεπίδραση μεταξύ των δύο τύπων εργαλείων και την ταχύτητά τους x, υπάρχει διαφοροποίηση μεταξύ των δύο τύπων εργαλείων ως προς τη διάρκεια ζωής τους με δεδομένη τη συμμετοχή της x στο μοντέλο; Ερμηνεύστε το $\hat{\beta}_2$. Πιστεύετε ότι τα δεδομένα μπορούν να εκφραστούν μέσω ενός απλού γραμμικού μοντέλου; [Δίνονται $\hat{\beta}_2=-12.643$ και SSE(x,z)= 43.91].



Β) Τυχαία επιλεγμένο δείγμα 20 τούβλων από την ημερήσια παραγωγή ενός εργοστασίου κατανεμήθηκε τυχαία σε τέσσερις διαφορετικές συνθήκες αποθήκευσης. Μετά από ένα χρονικό διάστημα μετρήθηκε η επί τοις % περιεκτικότητά τους σε νερό

Συνθήκες αποθήκευσης			
1	**2**	**3**	**4**
7.4	7.0	7.9	8.1
8.3	5.4	9.6	6.4
7.6	7.2	10.0	7.1
8.3	6.4	9.1	6.7
8.2	5.9	8.5	7.7

39,8
31,9
45,7
36
} 152,8

2337,84
7485
27904

$\frac{SSR/k}{SSE/(n-p)}$, $\frac{SSE}{n-p}=s^2$

$\frac{MSR}{MSE} \sim F_{k,(n-p)}$

Υιοθετώντας την κωδικοποίηση

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

Στα δεδομένα προσαρμόστηκε το ακόλουθο μοντέλο παλινδρόμησης $E(y)=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3$,

(i) Να γίνει ο έλεγχος $H_0:\beta_1=\beta_2=\beta_3=0$ με εναλλακτική H_1: τουλάχιστον ένα $\beta_j \neq 0$

[Δίνονται $\sum_{i=1}^{c}\sum_{j=1}^{n_i} y_{ij}^2 = 1194.1$, SSR=18.94].

(ii) Να συμπληρωθεί και να ερμηνευτεί ο παρακάτω πίνακας

Μεταβλητές	$\hat{\beta}$	$se(\hat{\beta})$	**t**	**p-τιμή**
Σταθερά	7.20	0.3116	23,107	
X_1	0.76	0.4407	1,725	
X_2	-0.82	0.4407	-1,867	
X_3	1.82	0.4407	4,072	

Γ) Έστω μοντέλο της παλινδρόμησης Poisson $f(y)=\frac{\exp(-\mu_x)\mu_x^y}{y!}$, y=0,1,2,..., με συνάρτηση σύνδεσης $g(\mu_x)=\ln\mu_x=\boldsymbol{\beta}'\mathbf{x}$. (i) Δείξτε πώς προσαρμόζεται αυτό το μοντέλο στην περίπτωση που υπάρχουν 2 συμμεταβλητές x_1, x_2. (ii) Δώστε τον ορισμό των υπολοίπων Pearson για ένα μοντέλο παλινδρόμησης Poisson. Σε τι μας χρησιμεύουν;