

Μάθημα : ΑΝΑΛΥΣΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ - ΕΞΕΤΑΣΕΙΣ ΧΕΙΜΕΡΙΝΟΥ ΕΞΑΜΗΝΟΥ 2016-17
 ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
 ***** Διάρκεια Εξέτασης : 2.30 ώρες *****

!! Επιλέξτε 4 Ζητήματα από τα 7 !!

ΖΗΤΗΜΑ 1 (Βαθμ. 2.5)

(Α) Έστω γενικό γραμμικό μοντέλο $y = X\beta + \epsilon$, με $\epsilon \sim N(0, \sigma^2 I_n)$, όπου X ο πίνακας σχεδιασμού με k επεξηγηματικές μεταβλητές. Να βρεθεί η $E(\epsilon)$ και ο πίνακας διασποράς-συνδιασποράς $V(\epsilon)$ των υπολοίπων ϵ . Να βρεθεί η $E(\epsilon_i)$, η $V(\epsilon_i)$, i, \dots, n , καθώς και η $\text{cov}(\epsilon_i, \epsilon_j)$, $i \neq j$.

(Β) Έστω $y = \beta_0 + \epsilon \sim N(\beta_0, \sigma^2)$, χωρίς επεξηγηματικές μεταβλητές, τότε με βάση τ.δ. y_1, y_2, \dots, y_n , ανεξαρτήτων παρατηρήσεων από την κατανομή αυτή, δείξτε με τη μέθοδο μέγιστης πιθανοφάνειας ότι $\hat{\beta}_0 = \bar{y}$ και κατά συνέπεια ότι $SST = SSE$.

ΖΗΤΗΜΑ 2 (Βαθμ. 2.5)

(Α) Έστω Μοντέλο Α: $E(y) = X_1\beta_1$ με k επεξηγηματικές μεταβλητές και με σταθερά β_0 και

Μοντέλο Β: $E(y) = X_1\beta_1 + X_2\beta_2$, με επιπλέον q μεταβλητές. Δείξτε ότι $R_A^2 \leq R_B^2$.

(Β) Γράψτε το συντελεστή συσχέτισης $r_{y, \hat{y}}$ μεταξύ των παρατηρήσεων y και \hat{y} ενός γενικού γραμμικού μοντέλου. Όταν η σταθερά περιλαμβάνεται στο μοντέλο παλινδρόμησης ισχύει ότι $\sum_{i=1}^n \epsilon_i = 0$. Δείξτε ότι σε αυτή την περίπτωση $r_{y, \hat{y}}^2 = R^2$.

ΖΗΤΗΜΑ 3 (Βαθμ. 2.5)

(Α) Έστω γενικό γραμμικό μοντέλο $y = X\beta + \epsilon$, με $\epsilon \sim N_n(0, \sigma^2 I_n)$ και ε.ε.τ. $\hat{\beta} = (X'X)^{-1} X'y$. Να βρεθεί ο πίνακας διασποράς-συνδιασποράς της $\hat{\beta}$. Έστω σημειακή πρόβλεψη $\hat{y}_{x_0} = x_0' \hat{\beta}$, να βρεθεί η διασπορά $V(\hat{y}_{x_0})$, όπου x_0 μία νέα παρατήρηση διάνυσμα.

(Β) Περιγράψτε πώς μέσω μιας ψευδομεταβλητής Z ($=1$, αν τα δεδομένα ανήκουν στην κατηγορία I και $=0$, αν ανήκουν στην II) στο μοντέλο $E(Y) = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ$, μπορούμε να εξετάσουμε αν (Α) δύο διαφορετικές ευθείες ή (Β) δύο παράλληλες ευθείες ή (Γ) μια ευθεία ταιριάζουν στα δεδομένα μας.

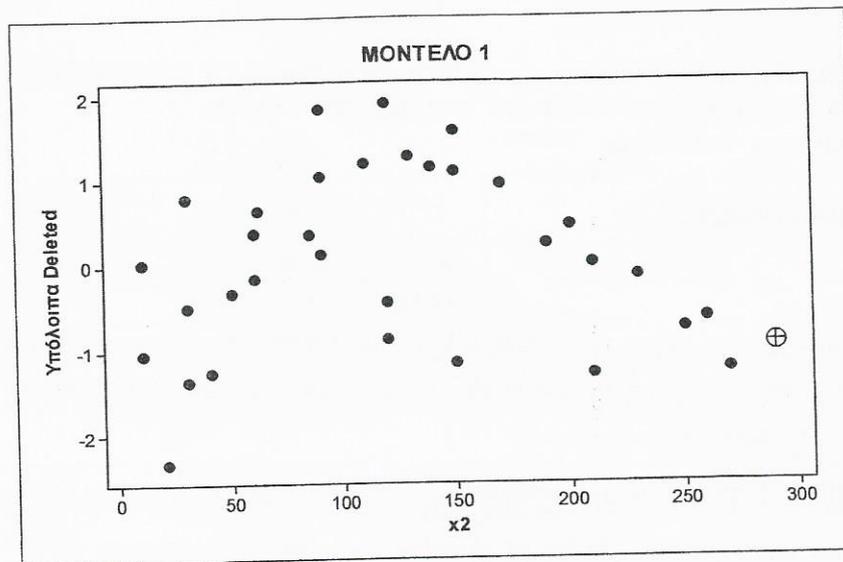
ΖΗΤΗΜΑ 4 (Βαθμ. 2.5)

(Α) Δώστε τον ορισμό των "deleted" υπολοίπων. Πώς μας βοηθάνε;

(Β) Μία χημική αντίδραση Y σχετίζεται γραμμικά με τις μεταβλητές X_1 (ποσότητα δραστικής ουσίας) και X_2 (χρόνος σε λεπτά) (Μοντέλο 1) (i) Συμπληρώστε τον παρακάτω πίνακα ($n=36$) και κατασκευάστε ένα 0.99-διάστημα εμπιστοσύνης για το συντελεστή β_2 . Σχολιάστε τα αποτελέσματά σας.

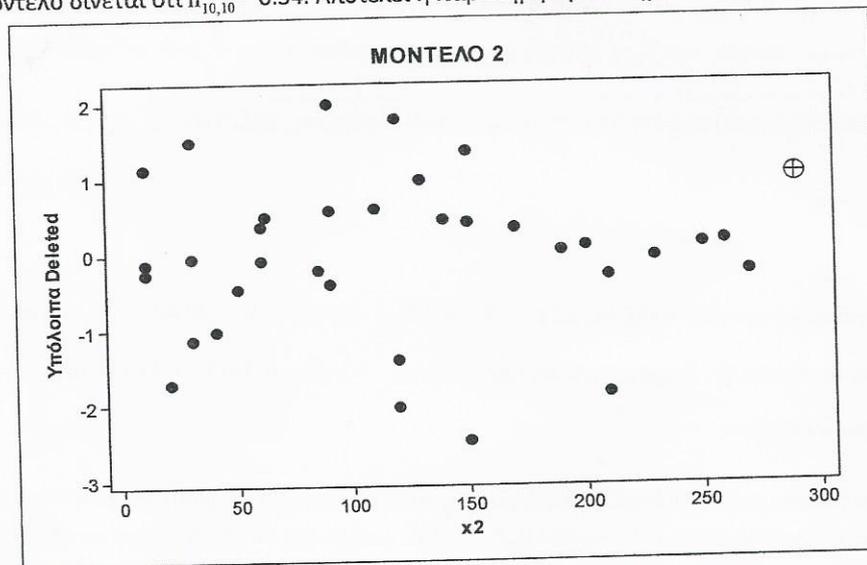
[Δίνεται: $r_{x_1, x_2} = -0.13$, $SST = 4761.8$, $S = \left(\frac{SSE}{(n-k-1)} \right)^{1/2} = 4.78$]

Μεταβλητές	$\hat{\beta}$	$se(\hat{\beta})$	t	p-τιμή	VIF
Σταθερά	11.09	1.67	6.64	<0.001	-
X_1	350.12	39.68			
X_2	0.11	0.01			
$R^2 = \quad \%$					



(ii) Δεδομένου ότι στο μοντέλο υπάρχουν οι μεταβλητές X_1 και X_2 εξετάστε αν το μοντέλο βελτιώνεται με την προσθήκη της X_2^2 (Μοντέλο 2), αξιοποιώντας έναν στατιστικό έλεγχο, το κριτήριο $R^2 = \text{---} \%$, καθώς και τις γραφικές παραστάσεις. [Δίνεται: $SSE = 459.8$]

(iii) Για το τελικό μοντέλο δίνεται ότι $h_{10,10} = 0.34$. Αποτελεί η παρατήρηση 10 σημείο επιρροής;



ΖΗΤΗΜΑ 5 (Βαθμ. 2.5)

(Α) Δώστε τον ορισμό του διορθωμένου συντελεστή προσδιορισμού \bar{R}^2 . Ποιά είναι η χρήση του ;
 (Β) Εξετάζεται η γραμμική παλινδρόμηση της Y σε σχέση με 5 επεξηγηματικές μεταβλητές X_1, X_2, X_3, X_4, X_5 . Με βάση τον παρακάτω πίνακα και για μέγεθος δείγματος $n=29$, επιλέξτε τα δύο καλύτερα μοντέλα και στη συνέχεια συμπληρώνοντας με έναν έλεγχο F , να βρεθεί το βέλτιστο από τα δύο.

	R^2	\bar{R}^2	AIC	C_p	S	X_1	X_2	X_3	X_4	X_5
1	72.1	71.0	147.62	38.5	12.3280				X	
1	39.4	37.1	170.06	112.7	18.1540	X				
2	85.9	84.8	129.06	9.1	8.9321			X	X	
2	82.0	80.6	136.82	17.8	10.0760				X	X
3	87.4	85.9	128.48	7.6	8.5978		X	X	X	
3	86.5	84.9	130.56	9.7	8.9110	X		X	X	
4	89.1	87.3	126.34	5.8	8.1698	X	X	X	X	
4	88.0	86.0	129.01	8.2	8.5550	X		X	X	X
5	89.9	87.7	126.17	6.0	8.0390	X	X	X	X	X

ΖΗΤΗΜΑ 6 (Βαθμ. 2.5)

Κατασκευαστής θέλει να συγκρίνει 3 τύπους Η/Υ ως προς την ευκολία χρήσης τους. Για το σκοπό αυτό επιλέχθηκαν τυχαία 12 απόφοιτοι Μαθηματικών του ίδιου επιπέδου, τέσσερις σε κάθε τύπο Η/Υ. Ο πίνακας δείχνει τη βαθμολόγηση (στα 100) της επίδοσης του κάθε αποφοίτου σε μια σειρά υπολογισμών.

Τύποι Η/Υ		
I	II	III
81	94	76
74	85	67
78	82	62
62	88	67

Θεωρώντας τις δείκτριες μεταβλητές $x_1 = \begin{cases} 1, & \text{αν τύπου I} \\ 0, & \text{αλλιώς} \end{cases}$, $x_2 = \begin{cases} 1, & \text{αν τύπου II} \\ 0, & \text{αλλιώς} \end{cases}$

προσαρμόζεται στα δεδομένα το μοντέλο παλινδρόμησης $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.

(i) Μέσω αυτού του μοντέλου να εξετάσετε αν υπάρχουν διαφοροποιήσεις μεταξύ των τριών τύπων Η/Υ ως προς τις επιδόσεις των αποφοίτων ($SSE=389.5$).

(ii) Να συμπληρωθεί και να ερμηνευτεί ο παρακάτω πίνακας

Μεταβλητές	$\hat{\beta}$	$se(\hat{\beta})$	t	p-τιμή
Σταθερά	68.00	3.29	20.67	<0.001
x1	5.75	4.65		
x2	19.25	4.65		

ΖΗΤΗΜΑ 7 (Βαθμ. 2.5)

(Α) Έστω μοντέλο παλινδρόμησης Poisson $f(y) = \frac{\exp(-\mu_x) \mu_x^y}{y!}$, $y=0,1,2, \dots$, με συνάρτηση σύνδεσης $g(\mu_x) = \ln \mu_x = \beta'x$ και

ελεγχουσυνάρτηση Deviance $= -2(\hat{\ell}_M - \hat{\ell}_{\text{κορ}}) = 2 \sum_{i=1}^n [y_i \ln(y_i / \hat{\mu}_i) - (y_i - \hat{\mu}_i)]$, όπου $\hat{\ell}_M$ η μεγιστοποιημένη λογαριθμοποιημένη

συνάρτηση πιθανοφάνειας του μοντέλου που μας ενδιαφέρει και κριτήριο $AIC = -2\hat{\ell}_M + 2d$, όπου d ο συνολικός αριθμός παραμέτρων στο μοντέλο. Δώστε τον ορισμό των υπολοίπων Pearson για το μοντέλο αυτό. Πώς τα χρησιμοποιούμε;

(Β) Προσαρμόζονται μοντέλα της παλινδρόμησης Poisson σε n=30 αεροσκάφη δύο τύπων A και B και εξετάζεται η σχέση του αριθμού ζημιών (Y) ανά αεροσκάφος, με τις συμμεταβλητές X_1 (=1, τύπος A και =0, τύπος B) και X_2 (βάρος βομβών σε τόνους) καθώς και με τη X_3 (μήνες εμπειρίας του πληρώματος). (i) Να συμπληρωθούν οι παρακάτω πίνακες.

(ii) Συγκρίνετε τα τρία μοντέλα με την ελεγχουσυνάρτηση Deviance (με διαδοχική αφαίρεση) και με το κριτήριο AIC και γράψτε το προσαρμοσμένο τελικό μοντέλο.

(iii) Ενισχύστε τα συμπεράσματά σας με την πιο κάτω γραφική παράσταση των υπολοίπων Pearson του τελικού μοντέλου.

(iv) Υπολογίστε και ερμηνεύστε τις εκτιμημένες ποσότητες $\exp(\hat{\beta}_j)$ του **τελικού μοντέλου**.

ΜΟΝΤΕΛΟ: 3 Μεταβλητές	$\hat{\beta}_j$	$se(\hat{\beta}_j)$	z_j	p-τιμή
Σταθερά	-0.406	0.877	-0.463	0.644
X_1	0.569	0.504		
X_2	0.165	0.068		
X_3	-0.014	0.008		

Ελεγχουσυνάρτηση deviance δίνεται ως $D_3 = 25.95$ και η τιμή του κριτηρίου $AIC_3 = 87.65$

ΜΟΝΤΕΛΟ: 2 Μεταβλητές	$\hat{\beta}_j$	$se(\hat{\beta}_j)$	z_j	p-τιμή
Σταθερά	-0.699	0.853	-0.819	0.413
X_2	0.222	0.046		
X_3	-0.012	0.008		

Ελεγχουσυνάρτηση deviance δίνεται ως $D_2 = 27.22$ και η τιμή του κριτηρίου $AIC_2 = 86.92$

ΜΟΝΤΕΛΟ: 1 Μεταβλητές	$\hat{\beta}_j$	$se(\hat{\beta}_j)$	z_j	p-τιμή
Σταθερά	-1.70	0.507	-3.356	<0.001
X_2	0.231	0.047		
<p>Ελεγχουσυνάρτηση deviance δίνεται ως $D_1 = 29.21$ με αντίστοιχη τιμή $\hat{\ell}_1 = \underline{\hspace{2cm}}$ και τιμή του κριτηρίου $AIC_1 = \underline{\hspace{2cm}}$</p>				
<p>D_0 (για το μοντέλο χωρίς καμία συμμεταβλητή) = 53.88 με αντίστοιχη τιμή $\hat{\ell}_0 = -53.79$ και $AIC_0 = \underline{\hspace{2cm}}$</p>				

Γραφική παράσταση των υπολοίπων Pearson του τελικού μοντέλου

