

ΣΕΙΡΑ 2

A) Δείξτε ότι

1) $\mathbf{I} - \mathbf{H}$, $\mathbf{H} - \frac{1}{n}\mathbf{J}_n$ και $\mathbf{I} - \frac{1}{n}\mathbf{J}_n$ είναι ταυτοδύναμοι πίνακες

2) $SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{y} - n\bar{y}^2$ και $SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n y_i^2 - \hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{y}$

B) (I) Με βάση τα δεδομένα

α/α	y	x_1	x_2	x_3
1	105	5	10	96
2	290	10	22	107
3	98	6	11	97
4	85	4	9	83
5	168	7	15	107
6	106	4	12	66
7	98	5	8	105
8	191	8	18	103
9	151	7	15	107
10	171	6	15	82
11	95	3	9	58
12	337	8	23	65
13	236	8	18	108
14	110	3	8	68
15	110	2	11	35

1) Να συμπληρωθεί ο πίνακας $\mathbf{X}'\mathbf{X}$:

$$\mathbf{X}'\mathbf{X} = \begin{array}{cccc} & -- & 86 & -- & 1287 \\ 86 & 566 & 1305 & -- & \\ -- & 1305 & 3112 & 17944 & \\ 1287 & -- & 17944 & 117677 & \end{array}$$

Στη συνέχεια αφού βεβαιωθεί ότι ο παρακάτω πίνακας είναι ο $(\mathbf{X}'\mathbf{X})^{-1}$ (αποκλίσεις μπορεί να οφείλονται σε στρογγυλοποίηση)

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{array}{cccc} 12.0424 & 3.33286 & -1.13986 & -0.181609 \\ 3.3329 & 1.03479 & -0.34553 & -0.053222 \\ -1.1399 & -0.34553 & 0.11859 & 0.017576 \\ -0.1816 & -0.05322 & 0.01758 & 0.002887 \end{array}$$

να βρεθεί το $\mathbf{X}'\mathbf{y}$ και ακολούθως να προσαρμοστεί το μοντέλο

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3.$$

- 2) Να κατασκευαστεί ο πίνακας ανάλυσης διασποράς.
- 3) Είναι η συνολική παλινδρόμηση στατιστικά σημαντική (με βάση την p -τιμή);
- 4) Να γίνουν οι έλεγχοι $H_0: \beta_i = 0$ με εναλλακτική την $H_0: \beta_i \neq 0$, $i = 1, 2, 3$ (με βάση την p -τιμή).
- 5) Να βρεθεί ο συντελεστής προσδιορισμού R^2 .

- 6) Να βρεθούν οι συντελεστές συσχέτισης $r_{x_i x_j}$, $i \neq j$, $i, j = 1, 2, 3$. Να σχολιάσετε τη στατιστική σημαντικότητα των εκτιμήσεων (από το (4)) σε σχέση με τις τιμές αυτές του συντελεστή συσχέτισης. Τι συμπεραίνετε από τις τιμές των $VIF=75.5, 40.0$ και 20.9 που αντιστοιχούν στους συντελεστές $\beta_1, \beta_2, \beta_3$;

II) Η προσαρμογή του μοντέλου επαναλαμβάνεται χωρίς τη μεταβλητή x_1 . Ο παρακάτω πίνακας δίνει τις εκτιμήσεις των συντελεστών και την ανάλυση διασποράς:

The regression equation is
 $y = -34.6 + 15.3 x_2 - 0.201 x_3$

Predictor	Coef	SE Coef	T	P
Constant	-34.55	26.86	-1.29	0.223
x2	15.331	1.332	11.51	0.000
x3	-0.2007	0.2874	-0.70	0.498

S = 23.4845 R-Sq = 92.1% R-Sq(adj) = 90.8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	76933	38466	69.75	0.000
Residual Error	12	6618	552		
Total	14	83551			

- 1) Να προσδιοριστούν οι τιμές των VIF για τους συντελεστές β_2, β_3 , τι συμπεραίνετε; $\chi^2_{\alpha, 1}, \chi^2_{\alpha, 2}, S(\cdot)$
- 2) Να συγκρίνετε τη στατιστική σημαντικότητα των εκτιμήσεων των συντελεστών με τα αποτελέσματα του II(4).
- 3) Να συγκρίνετε τις δύο τιμές του R^2 καθώς και τα αποτελέσματα των δύο αναλύσεων διασποράς (I και II).
- 4) Να κατασκευαστούν 0.95-δ.ε. των συντελεστών β_2, β_3 .
- 5) Να κατασκευαστεί ένα 0.99-δ.ε μιας νέας παρατήρησης y_{x_0} όταν ένα νέο $x_0 = (1, x_2, x_3)' = (1, 20, 90)'$.

III) Να επαναληφθεί η προσαρμογή του μοντέλου αφαιρώντας τη μεταβλητή x_3 (δηλαδή ζητείται ο πίνακας αποτελεσμάτων με τις εκτιμήσεις των συντελεστών και την ανάλυση διασποράς, όπως στο ερώτημα (II)), ο υπολογισμός του R^2 και να γίνουν οι ακόλουθες γραφικές παραστάσεις με χρήση H/Y.

- e με \hat{y}
- e με x_2

IV) Να επαναληφθεί η προσαρμογή του μοντέλου με χρήση H/Y με τις μεταβλητές x_2 και x_2^2 (πάλι ο πίνακας αποτελεσμάτων με τις εκτιμήσεις των συντελεστών και την ανάλυση διασποράς), ο υπολογισμός του R^2 και να γίνουν οι ακόλουθες γραφικές παραστάσεις

- e με \hat{y}
- έλεγχος κανονικότητας των e

V) Τελικά συμπεράσματα