

Στατιστικά Μοντέλα

Σειρά 1

A) Δείξτε ότι για το απλό γραμμικό μοντέλο $E(y_x) = \beta_0 + \beta_1 x$ ισχύουν τα ακόλουθα :

1) $R^2 = r_{xy}^2$, R^2 ο συντελεστής προσδιορισμού, r_{xy} ο δειγματικός συντελεστής συσχέτισης (Pearson) των x και y παρατηρήσεων,

2) $\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}_i$, **3)** $\sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$, **4)** $\frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{r_{xy} \sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$.

B) Τα δεδομένα στο αρχείο cholesterol.txt αφορούν επίπεδα ολικής χοληστερόλης (mg/ml) 24 ασθενών (y) και την ηλικία τους (x).

(i) Να κατασκευαστεί ένα διάγραμμα διασποράς μεταξύ των δύο μεταβλητών y και x και να προσαρμοστεί το μοντέλο $E(y) = \beta_0 + \beta_1 x$ στα δεδομένα.

(ii) Να γίνει ο έλεγχος $H_0: \beta_1 = 0$ έναντι της $H_1: \beta_1 \neq 0$ και επιπλέον να προσδιοριστεί ένα 95% διάστημα εμπιστοσύνης (δ.ε.) για το συντελεστή της x στο μοντέλο που προσαρμόστηκε. Πώς ερμηνεύουμε το $\hat{\beta}_1$;

(iii) Να κατασκευαστεί ένα 99% δ.ε. πρόβλεψης για το επίπεδο χοληστερόλης y ενός ασθενή ηλικίας 35 ετών, καθώς και για την αναμενόμενη τιμή της, $E(y)$.

(iv) Να γίνει ο γραφικός έλεγχος της Κανονικής κατανομής και η γραφική παράσταση e_i με \hat{y}_i , για τα υπόλοιπα e_i . Τι συμπεραίνετε;

Γ) Συμμετοχή Y (%) ιδιοκτητών κατοικιών σε δειγματοληπτική έρευνα, σε σχέση με την αξία της κατοικίας (X) :

x	3	7	12	17	25	35	45	55	70	120
y	86	79	76	69	65	62	52	51	51	48

(i) Να κατασκευαστεί ένα διάγραμμα διασποράς μεταξύ των δύο μεταβλητών y και x και να προσαρμοστεί το μοντέλο $\frac{100}{100-y} = \alpha + \frac{\beta}{x}$.

(ii) Να εκτιμηθεί σημειακά η άγνωστη παρατήρηση y και να κατασκευαστεί ένα 95% διάστημα εμπιστοσύνης (δ.ε.) για την πρόβλεψη της παρατήρησης y , καθώς και ένα προσεγγιστικό 95% δ.ε. για τη μέση τιμή της, $E(y)$, όταν η αξία της κατοικίας είναι $x = 20$.

Χ. Καρώνη
Χειμερινό εξάμηνο 2022-23
ΣΕΜΦΕ