

Στατιστική II, Σεπτέμβρης 2019

Θέμα 1: Σε μία στατιστική ανάλυση με το λογισμικό R πήραμε τα εξής αποτελέσματα για την εξήγηση της μέγιστης συγκέντρωσης του όζοντος O_3 σε μία μέρα ως συνάρτηση της ταχύτητας του ανέμου Vx σε συγκεκριμένη κατεύθυνση στις 12 το μεσημέρι.

	Coefficients:			
	Estimate	Std. Error	t value	$Pr(> t)$
(Intercept)	87.073	?	29.979	$< 2e-16$
Vx	0.930	0.215	?	?

Residual standard error: 20.5 on ? degrees of freedom

Multiple R-Squared: ?, Adjusted R-squared: ?

F-statistic: ? on ? and 48 DF, p-value: $7.932e-05$

- (1) Με βάση τα παραπάνω αποτελέσματα γράψτε την εξίσωση της ευθείας των ελαχίστων τετραγώνων που αντιστοιχεί στα σημεία (Vx_i, O_3_i) , $i = 1, \dots, n$, όπου n είναι το πλήθος των παρατηρήσεων και βρείτε την τιμή του n .
- (2) Υπολογίστε το άθροισμα των τετραγώνων των σφαλμάτων SSE .
- (3) Κάντε πρόβλεψη της μέγιστης συγκέντρωσης του όζοντος σε ταχύτητα ανέμου $Vx = 20$.
- (4) Συμπληρώστε όλες τις άγνωστες ποσότητες που εμφανίζονται ως ερωτηματικά στα αποτελέσματα της παλινδρόμησης. Αιτιολογήστε τις απαντήσεις σας.

Θέμα 2: Θεωρούμε το κανονικό πολλαπλό γραμμικό μοντέλο

$$y_i = \beta_1 + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \epsilon_i, \quad 1 \leq i \leq n.$$

Αν X είναι ο πίνακας σχεδιασμού και Y το διάνυσμα των παρατηρήσεων, τότε παρατηρήθηκε ότι

$$X^T X = \begin{pmatrix} 30 & 20 & 0 \\ 20 & 20 & 0 \\ 0 & 0 & 10 \end{pmatrix} \quad X^T Y = \begin{pmatrix} 15 \\ 20 \\ 10 \end{pmatrix} \quad Y^T Y = 59.5.$$

- (1) Καθορίστε την τιμή του n , το μέσο \bar{x}_3 των $x_{i,3}$ και το δειγματικό συντελεστή συσχέτισης $r_{2,3}$ των $x_{i,2}$ με τα $x_{i,3}$.
- (2) Εκτιμήστε τα β_1 , β_2 , β_3 και σ^2 με τη μέθοδο των ελαχίστων τετραγώνων.
- (3) Βρείτε ένα 95% διάστημα εμπιστοσύνης για την παράμετρο β_2 .
- (4) Να γίνει ο έλεγχος της υπόθεσης $\beta_3 = 0.8$ έναντι $\beta_3 \neq 0.8$ σε ε.σ.σ. 0.1.
- (5) Να γίνει ο έλεγχος $\beta_2 + \beta_3 = 3$ έναντι $\beta_2 + \beta_3 \neq 3$ σε ε.σ.σ. 0.05.
- (6) Να βρεθεί η τιμή R_a^2 του προσαρμοσμένου R^2 .

Δίνονται:

$$t_{27}(0.1) = 1.31, \quad t_{27}(0.05) = 1.70, \quad t_{27}(0.025) = 2.05, \quad t_{27}(0.01) = 2.47$$

$t_{17}(0.1) = 1.33, \quad t_{17}(0.05) = 1.74, \quad t_{17}(0.025) = 2.11, \quad t_{17}(0.01) = 2.57$ όπου $t_d(a)$ είναι το a -άνω ποσοστημόριο της κατανομής Student με d βαθμούς ελευθερίας.

Θέμα 3: Ένας φοιτητής θέλει να προσομοιώσει 10 τιμές X_i από την κατανομή του Pareto με συνάρτηση πυκνότητας $f(x) = \alpha x^{-\alpha-1}$, $x \geq 1$. Διαθέτει ένα νόμισμα (που υποθέτουμε αμερόληπτο), με τη βοήθεια του οποίου πραγματοποιεί 100 ανεξάρτητες ρίψεις.

- (1) Προτείνεται έναν τρόπο προσομοίωσης 10 τιμών U_i από τη συνεχή ομοιόμορφη στο $[0, 1]$ με προσέγγιση τουλάχιστον 3 δεκαδικών ψηφίων.
- (2) Με τη βοήθεια των παραπάνω τιμών U_i , προσομοιώστε τις 10 τιμές των X_i .

Θέμα 4: Μετρήθηκαν οι ανοσοσφαιρίνες M του ορού 9 ενήλικων ατόμων πριν και 15 μέρες μετά τον αντιχολερικό εμβολιασμό. Οι τιμές που βρέθηκαν (σε κατάλληλη μονάδα μέτρησης) δίνονται στις στήλες 2 και 3 του παρακάτω πίνακα. Ζητείται η εφαρμογή του ελέγχου Wilcoxon κατά ζεύγη με τη βοήθεια του στατιστικού T^+ . Υπενθυμίζεται ότι μία διαφορά είναι στατιστικά σημαντική αν $T^+ \leq c_{n,\alpha}$ ή $T^+ \geq d_{n,\alpha}$ για κατάλληλες σταθερές που προσδιορίζονται από την κατανομή του T^+ .

ΠΙΝΑΚΑΣ 1. Ανοσοσφαιρίνες πριν και μετά τον εμβολιασμό.

Ατομο	Πριν	Μετά
1	158	168
2	140	168
3	165	161
4	100	120
5	110	122
6	142	134
7	104	116
8	254	270
9	180	180

- (1) Αν δίνεται η τιμή $c_{n,\alpha}$, να προσδιοριστεί το $d_{n,\alpha}$.
- (2) Να αξιολογηθεί στατιστικά η διαφορά των τιμών των ανοσοσφαιρινών αυτών πριν και μετά τον εμβολιασμό με τον έλεγχο Wilcoxon κατά ζεύγη σε $\alpha = 0.01$ και $\alpha = 0.05$. Δίνονται οι τιμές $c_{8,0.01} = 0$, $c_{8,0.05} = 3$, $c_{9,0.01} = 1$, $c_{9,0.05} = 5$.

Να απαντήσετε σε όλα τα θέματα.

ΚΑΛΗ ΕΠΙΤΥΧΙΑ
ΔΙΑΡΚΕΙΑ ΕΞΕΤΑΣΗΣ: 3 ΩΡΕΣ