Στατιστική II, Ιούνιος 2019

Θέμα 1: Σε μία στατιστική ανάλυση με το λογισμικό R πήραμε τα εξής αποτελέσματα για την εξήγηση της μέγιστης συγκέντρωσης του όζοντος $O3$ σε μία μέρα ως συνάρτηση της θερμοκρασίας $T15$ στις 15 το μεσημέρι.

	Coefficients: Estimate	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	26.262	?	2.13	0.038
T12	2.807	0.562	?	?

Residual standard error: 19.59 on ? degrees of freedom
Multiple R-Squared: ?, Adjusted R-squared: ?
F-statistic: ? on ? and 48 DF, p-value: 8.17e-06

(1) Με βάση τα παραπάνω αποτελέσματα γράψτε την εξίσωση της ευθείας των ελαχίστων τετραγώνων που αντιστοιχεί στα σημεία $(T15_i, O3_i)$, $i = 1, \ldots, n$, όπου n είναι το πλήθος των παρατηρήσεων και βρείτε την τιμή του n.
(2) Υπολογίστε το άθροισμα των τετραγώνων των σφαλμάτων SSE.
(3) Κάντε μία πρόβλεψη της μέγιστης συγκέντρωσης του όζοντος σε θερμοκρασία $T15 = 15$ και εκτιμήστε την τυπική απόκλιση του σφάλματος πρόβλεψης.
(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 $Y = \mathbf{X}\beta + \epsilon$. Με τη βοήθεια 21 παρατηρήσεων πήραμε τα εξής αποτελέσματα :

$$\begin{aligned} \hat{Y} &= 6.683_{(2.67)} + 0.44_{(2.32)}X_2 + 0.425_{(2.47)}X_3 + 0.171_{(2.09)}X_4 + 0.009_{(2.24)}X_5 \\ R^2 &= 0.54, \end{aligned}$$

όπου για κάθε συντελεστή, ο αριθμός μέσα στις παρενθέσεις αναπαριστά την απόλυτη τιμή του στατιστικού του ελέγχου Student για τη μηδενικότητα της αντίστοιχης παραμέτρου.

(1) Ελέγξτε τη μηδενικότητα του β_2 σε επίπεδο σημαντικότητας $a = 0.05$.
(2) Βρείτε ένα 95% διάστημα εμπιστοσύνης για την παράμετρο β_2. Πώς μπορούμε να επιβεβαιώσουμε το αποτέλεσμα της προηγούμενης ερώτησης ;
$t_{20}(0.05) = 1.73$, $t_{20}(0.025) = 2.09$, $t_{16}(0.05) = 1.75$, $t_{16}(0.025) = 2.12$, όπου $t_d(a)$ είναι το a-άνω ποσοστημόριο της κατανομής Student με d βαθμούς ελευθερίας.
(3) Κάντε τον έλεγχο της $\mathcal{H}_0$: '$\beta_4 = 1$' έναντι της $\mathcal{H}_1$: '$\beta_4 \neq 1$' σε επίπεδο σημαντικότητας $a = 0.05$.
(4) Περιγράψτε τον καθολικό έλεγχο του Fisher (υποθέσεις, στατιστικό ελέγχου, κατανομή και περιοχή απόρριψης). Υπολογίστε το στατιστικό του ελέγχου F σε αυτό το παράδειγμα και πραγματοποιήστε τον έλεγχο σε επίπεδο σημαντικότητας $a = 0.05$.
$f_{4,16}(0.05) = 3.01$, $f_{5,16}(0.05) = 2.85$, $f_{4,21}(0.05) = 2.84$, $f_{5,21}(0.05) = 2.68$, όπου $f_{d_1,d_2}(a)$ είναι το a-άνω ποσοστημόριο της κατανομής του Fisher $\mathcal{F}_{d_1,d_2}$.

Θέμα 3: Την καλοκαιρινή περίοδο ένας νεαρός έχει αποφασίσει να πηγαίνει καθημερινά αποκλειστικά είτε για μπάνιο στη θάλασσα, είτε για βόλτα στο βουνό. Του αρέσει βέβαια να το αποφασίζει τυχαία (ισοπίθανα) κάθε μέρα και ανεξάρτητα από το τί έχει συμβεί τις προηγούμενες μέρες. Υποθέτουμε ότι αν πάει θάλασσα ο χρόνος παραμονής του εκεί X_θ ακολουθεί την εκθετική κατανομή $Exp(2)$, ενώ αν πάει στο βουνό $X_\beta \sim Exp(3)$, όπου η παράμετρος της εκθετικής εκφράζει μέση τιμή και όχι ρυθμό.

(1) Να βρεθεί η συνάρτηση κατανομής F της τυχαίας μεταβλητής X που εκφράζει το χρόνο παραμονής του νεαρού καθημερινά στη θάλασσα ή στο βουνό.
(2) Αν έχετε στη διάθεσή σας ένα τυχαίο δείγμα από 50 ομοιόμορφες στο $(0,1)$ τυχαίες μεταβλητές U_i, τότε καθορίστε έναν τρόπο προσομοίωσης των χρόνων παραμονής $\{X_i\}_{i=1}^{50}$ στην καθημερινή δραστηριότητα, χωρίς να μας ενδιαφέρει να διακρίνουμε τον τύπο της δραστηριότητας.
(3) Υποδείξτε έναν τρόπο προσομοίωσης των χρόνων $\{X_i\}_{i=1}^{50}$, με το ίδιο τυχαίο δείγμα ομοιόμορφων, αλλά με την απαίτηση να διακρίνουμε ταυτόχρονα τον τύπο δραστηριότητας [υπόδειξη: χρησιμοποιήστε τις ομοιόμορφες για να προσομοιώσετε κατάλληλα προσημασμένες X_i].

Θέμα 4: Έστω (X_n) ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών με άγνωστη συνάρτηση κατανομής F.

(1) Να οριστεί η εμπειρική συνάρτηση κατανομής $\widehat{F}_n$ που αντιστοιχεί στο τυχαίο δείγμα $\{X_i\}_{i=1}^{n}$. Τί εκφράζει το $\widehat{F}_n(t)$ για κάθε $t \in \mathbb{R}$;
(2) Έστω $t_j,\ j = 1, 2, \ldots, k$ διατεταγμένοι πραγματικοί αριθμοί σε αύξουσα σειρά. Να δείξετε ότι η ακολουθία των τυχαίων διανυσμάτων $(\widehat{F}_n(t_1), \ldots, \widehat{F}_n(t_k))$ συγκλίνει με πιθανότητα 1 στο τυχαίο διάνυσμα $(F(t_1), \ldots, F(t_k))$. Τί συμπεραίνουμε για την εκτιμήτρια $(\widehat{F}_n(t_1), \ldots, \widehat{F}_n(t_k))$ του $(F(t_1), \ldots, F(t_k))$;
(3) Να δείξετε ότι η παραπάνω εκτιμήτρια είναι και ασυμπτωτικά κανονική, εξετάζοντας την ασυμπτωτική συμπεριφορά της ακολουθίας $\sqrt{n}(\widehat{F}_n(t_1) - F(t_1), \ldots, \widehat{F}_n(t_k) - F(t_k))$. Να υπολογίσετε και τον ασυμπτωτικό πίνακα συνδιακύμανσης.

Να απαντήσετε σε όλα τα θέματα.

ΚΑΛΗ ΕΠΙΤΥΧΙΑ
ΔΙΑΡΚΕΙΑ ΕΞΕΤΑΣΗΣ: 3 ΩΡΕΣ