**ΓΛΩΣΣΑ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ R**

**ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΟΝ**

**ΑΝΑΛΟΓΙΣΜΟ**



ΑΚΑΔΗΜΑΪΚΟ ΕΤΟΣ 2019-2020

ΕΡΓΑΣΙΑ 1

**Οδηγίες**

* Η καταληκτική ημερομηνία παράδοσης της εργασίας είναι η Δευτέρα 13/01/2019. Μπορείτε να αφήσετε την **τυπωμένη** εργασία σας είτε σε εμένα προσωπικά, είτε στο γραμματοκιβώτιό μου στον 5ο όροφο.

**ΠΡΟΣΟΧΗ**: H τυπωμένη εργασία, πέρα από τις απαντήσεις/σχολιασμούς των ερωτημάτων, θα πρέπει να περιλαμβάνει απαραιτήτως, σε κάθε ερώτημα, **το output της R κονσόλας** από το οποίο θα προκύπτει η απάντηση στο ερώτημα **συμπεριλαμβανομένου** και του κώδικα (εντολές του R).

* Όλες οι εντολές του R (κώδικας) που χρησιμοποιήσατε στην τυπωμένη εργασία σας θα πρέπει να υπάρχουν σε ένα αρχείο με όνομα **MAE190XX.R** (όπου MAE190ΧΧ ο αριθμός μητρώου σας) το οποίο θα πρέπει να στείλετε στην ηλεκτρονική διεύθυνση [dantz@unipi.gr](mailto:dantz@unipi.gr) μέχρι τη Δευτέρα 13/01/2019.

**ΠΡΟΣΟΧΗ**: Το αρχείο **MAE190XX.R** είναι **βοηθητικό** και δεν λαμβάνεται υπόψη στη διαμόρφωση της βαθμολογίας σας (βαθμολογείται μόνο η τυπωμένη εργασία σας).

* Στους ελέγχους υποθέσεων, όταν δεν δίνεται το επίπεδο σημαντικότητας , να θεωρείται ότι 

**Άσκηση 1**

Το πλαίσιο που ακολουθεί περιέχει τους χρόνους (σε μήνες) που μεσολάβησαν μεταξύ 38 διαδοχικών εκρήξεων του ηφαιστείου Kilauea της Hawaii από το 1830 έως το 1950.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 73 | 6 | 77 | 81 | 91 | 101 | 135 | 61 | 65 | 68 |
| 18 | 20 | 23 | 12 | 14 | 18 | 23 | 26 | 26 | 27 |
| 2 | 3 | 3 | 40 | 41 | 41 | 6 | 8 | 11 | 12 |
| 37 | 38 | 38 | 6 | 73 | 6 | 51 |  |  |  |

1. Υπολογίστε τη δειγματική μέση τιµή, τη δειγματική διάμεσο, τη δειγματική έκταση, τη δειγματική διασπορά, το δειγματικό συντελεστή ασυμμετρίας[[1]](#footnote-1) και το πρώτο και τρίτο τεταρτημόριο των δεδομένων.
2. Κατασκευάστε θηκόγραμμα και κανονικό Q-Q διάγραμμα για τα δεδομένα. Μπορείτε να συμπεράνετε ότι τα δεδομένα προέρχονται από κανονικό πληθυσμό;
3. Να γίνουν έλεγχοι κανονικότητας των δεδομένων, σε επίπεδο σημαντικότητας , χρησιμοποιώντας μόνο τις συναρτήσεις lillie.test και shapiro.test. Μπορούμε να υποστηρίξουμε ότι τα δεδομένα προέρχονται από κανονικό πληθυσμό;
4. Να κατασκευαστεί ιστόγραμμα συχνοτήτων των δεδομένων (τάξεις: 0-20, 20-40, ..., 120-160) στο οποίο να προσαρμοστεί «συνάρτηση πυκνότητας πιθανότητας» κατάλληλης Γάμμα κατανομής (οι παράμετροι της κατανομής Γάμμα να εκτιμηθούν με τη μέθοδο των ροπών[[2]](#footnote-2)).
5. Να εξετάσετε με Q-Q διάγραμμα και με έλεγχο Kolmogorov-Smirnov, αν τα δεδομένα προέρχονται από εκθετική κατανομή με παράμετρο .
6. Να εξετάσετε, σε επίπεδο σημαντικότητας , αν η διάμεσος του πληθυσμού είναι ίση με 20 ή όχι χρησιμοποιώντας προσημικό έλεγχο και έλεγχο Wilcoxon Signed-Rank. Ποια είναι τα συμπεράσματά σας;

**Άσκηση 2**

**A:** Για τη γεωμετρική κατανομή με παράμετρο  (πιθανότητα επιτυχίας) να βρεθεί η συνάρτηση πιθανότητας της αντίστοιχης

1. αποκομμένης στο 0 κατανομής, και
2. τροποποιημένης στο 0 κατανομής με μάζα πιθανότητας στο 0 ίση με 0.05.

Να δοθεί γραφική παράσταση και πίνακας των συναρτήσεων πιθανότητας  των δύο παραπάνω κατανομών καθώς και της γεωμετρικής κατανομής με παράμετρο  για .

Η γραφική παράσταση και ο πίνακας θα πρέπει να δοθούν στην ακόλουθη μορφή που αναφέρεται στην περίπτωση .

k Truncated Geometric Modified Geometric Geometric

0 0.0000000 0.0500000 0.6000000

1 0.6000000 0.5700000 0.2400000

2 0.2400000 0.2280000 0.0960000

3 0.0960000 0.0912000 0.0384000

4 0.0384000 0.0364800 0.0153600

5 0.0153600 0.0145920 0.0061440

6 0.0061440 0.0058368 0.0024576

7 0.0024576 0.0023347 0.0009830

8 0.0009830 0.0009339 0.0003932

9 0.0003932 0.0003736 0.0001573

10 0.0001573 0.0001494 0.0000629

11 0.0000629 0.0000598 0.0000252

12 0.0000252 0.0000239 0.0000101

13 0.0000101 0.0000096 0.0000040

14 0.0000040 0.0000038 0.0000016

15 0.0000016 0.0000015 0.0000006

16 0.0000006 0.0000006 0.0000003

17 0.0000003 0.0000002 0.0000001

18 0.0000001 0.0000001 0.0000000

19 0.0000000 0.0000000 0.0000000

20 0.0000000 0.0000000 0.0000000



**B:** Στη συνέχεια θεωρήστε ένα τυχαίο δείγμα , , …,  (μεγέθους 6) από την παραπάνω τροποποιημένη γεωμετρική κατανομή και βρείτε τη συνάρτηση πιθανότητας  της τυχαίας μεταβλητής



για  με ακρίβεια 7 δεκαδικών ψηφίων.

**Άσκηση 3**

Κατά την εκπόνηση μιας μελέτης για την επίδραση της υγρασίας (*X*1), της θερμοκρασίας (*X*2) και της βαρομετρικής πίεσης (*X*3) στις εκπομπές διοξειδίου του αζώτου (*Υ*) πετρελαιοκίνητων φορτηγών, συλλέχθηκαν 20 μετρήσεις εκπομπών σε διάφορες χρονικές στιγμές υπό διαφορετικές πειραματικές συνθήκες. Οι μετρήσεις της υγρασίας (*X*1) είναι σε GPP (H2O grains per pounds of dry air), της θερμοκρασίας (*X*2) σε oF (βαθμούς Φαρενάιτ) και της βαρομετρικής πίεσης (*X*3) σε inHg (ίντσες υδραργύρου), ενώ οι μετρήσεις στις εκπομπές διοξειδίου του αζώτου (*Υ*) είναι σε g/km (γραμμάρια NO2 ανά χιλιόμετρο). Τα δεδομένα παρατίθενται στον παρακάτω πίνακα (αρχείο Askisi3.txt)

|  |  |  |  |
| --- | --- | --- | --- |
| *Y* | *X*1 | *X*2 | *X*3 |
| 0.90 | 72.4 | 76.3 | 29.18 |
| 0.91 | 41.6 | 70.3 | 29.35 |
| 0.96 | 34.3 | 77.1 | 29.24 |
| 0.89 | 35.1 | 68.0 | 29.27 |
| 1.00 | 10.7 | 79.0 | 29.78 |
| 1.10 | 12.9 | 67.4 | 29.39 |
| 1.15 | 8.3 | 66.8 | 29.69 |
| 1.03 | 20.1 | 76.9 | 29.48 |
| 0.77 | 72.2 | 77.7 | 29.09 |
| 1.07 | 24.0 | 67.7 | 29.60 |
| 1.07 | 23.2 | 76.8 | 29.38 |
| 0.94 | 47.4 | 86.6 | 29.35 |
| 1.10 | 31.5 | 76.9 | 29.63 |
| 1.10 | 10.6 | 86.3 | 29.56 |
| 1.10 | 11.2 | 86.0 | 29.48 |
| 0.91 | 73.3 | 76.3 | 29.40 |
| 0.87 | 75.4 | 77.9 | 29.28 |
| 0.78 | 96.6 | 78.7 | 29.29 |
| 0.82 | 107.4 | 86.8 | 29.03 |
| 0.95 | 54.9 | 70.9 | 29.37 |

1. Nα προσαρμοστεί στα δεδομένα ένα πρότυπο πολλαπλής γραμμικής παλινδρόμησης με εξαρτημένη μεταβλητή τις εκπομπές διοξειδίου του αζώτου και ανεξάρτητες μεταβλητές όλες τις περιβαλλοντικές μετρήσεις. Ποια είναι η προσαρμοσμένη εξίσωση παλινδρόμησης;
2. Είναι στατιστικά σημαντική η παλινδρόμηση σε επίπεδο σημαντικότητας ;
3. Ποιες ανεξάρτητες μεταβλητές του προτύπου του ερωτήματος (i) είναι στατιστικά σημαντικές στην εκπομπή διοξειδίου του αζώτου, σε επίπεδο σημαντικότητας ;
4. Να βρεθεί το 99% διάστημα εμπιστοσύνης για κάθε έναν από τους τρεις συντελεστές παλινδρόμησης.
5. Ποιες είναι οι τιμές των ποσοτήτων SSR, SSE και SST;
6. Ποια είναι η τιμή της εκτίμησης της διασποράς των σφαλμάτων;
7. Ποιο είναι το 90% διάστημα εμπιστοσύνης για τη μέση τιμή της εκπομπής διοξειδίου του αζώτου όταν η υγρασίας είναι 75 GPP, η θερμοκρασίας είναι 70 oF και η βαρομετρική πίεσης είναι 29.5 inHg;
8. Χρησιμοποιώντας τη συνάρτηση stepAIC του πακέτου MASS να βρεθεί το καταλληλότερο γραμμικό μοντέλο για την περιγραφή των εκπομπών διοξειδίου του αζώτου.

**ΚΑΛΗ ΕΠΙΤΥΧΙΑ!!!**

1. Ο συντελεστής ασυμμετρίας μιας τυχαίας μεταβλητής *Χ* δίνεται από τον τύπο  . [↑](#footnote-ref-1)
2. Η συνάρτηση πυκνότητας πιθανότητας της κατανομής Γάμμα με παραμέτρους *α* και *b* δίνεται από τον τύπο

   .

   Οι πρώτες δύο ροπές της τ.μ.  είναι: , .

   Οι εκτιμητές των παραμέτρων *a*, *b*, με τη μέθοδο των ροπών, από δείγμα μεγέθους *n* προκύπτουν από τη λύση του συστήματος

    όπου  [↑](#footnote-ref-2)