Πρώτη Εργασία Γραμμικών και Γενικευμένων Γραμμικών Μοντέλων, ΠΜΣ, ΣΑΧΜ, Πανεμιστήμιο Αιγαίου.
Καταληκτική ημερομηνία: **8 Μάη 2024**

Σημείωση: Η εργασία καλό είναι να γίνει σε latex. Μη γράφετε εξισώσεις σε Word.

1. (25 μονάδες) Θεωρήστε τη τετραγωνική μορφή $q(y_1, y_2, y_3) = 9y_1^2 + 7y_2^2 + 3y_3^2 - 2y_1y_2 + 4y_1y_3 - 6y_2y_3$

   (α΄) Γράψτε τη $q(y_1, y_2, y_3)$ ως $\boldsymbol{y}^T\boldsymbol{A}\boldsymbol{y}$, με $\boldsymbol{A}$ να είναι συμμετρικός πίνακας.

   (β΄) Χρησιμοποιώντας την απάντησή σας στο προηγούμενο ερώτημα δείξτε πως $q(y_1, y_2, y_3) > 0$, εκτός αν $y_1 = y_2 = y_3 = 0$. Να μη γίνει χρήση R εδώ.

   (γ΄) Έστω πως $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{A})$, όπου $\boldsymbol{\mu}^T = (3, 2, 1)$ και $\boldsymbol{A}$ ο πίνακας που βρήκατε παραπάνω.

   i. Γράψτε τη κατανομή της $(Y_1|Y_2 = y_2, Y_3 = y_3)$.

   ii. Υπολογίσετε την τιμή της σ.π.π. $f_{Y_1|Y_2=1,Y_3=2}(2)$. Επιβεβαιώσετε την απάντηση με την R.

   iii. Γράψτε τη κατανομή της $\left((Y_1, Y_2)^T|Y_3 = y_3\right)$.

   iv. Υπολογίσετε την τιμή της σ.π.π. $f_{Y_1,Y_2|Y_3=2}(1, 2)$. Επιβεβαιώσετε την απάντηση με την R.

   (δ΄) Έστω πως $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, όπου $\boldsymbol{\Sigma}$ είναι ένας 3×3 θετικά ορισμένος πίνακας.

   i. Βρείτε τη κατανομή της τετραγωνικής μορφής $Q = (\boldsymbol{Y} - \boldsymbol{\mu})^T\boldsymbol{B}(\boldsymbol{Y} - \boldsymbol{\mu})$, με $\boldsymbol{B} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{A}\boldsymbol{\Sigma}^{-1/2}$.

   ii. Προσεγγίστε με προσωμοίωση ($n_{sim} = 100000$) την $P(Q > 0.5)$ με χρήση της R.

   iii. Προσεγγίστε με προσωμοίωση ($n_{sim} = 100000$), με χρήση της R, το άνω 5% ποσοστιαίο σημείο της κατανομής της Q.

2. (10 μονάδες) Θεωρήστε το κανονικό γραμμικό μοντέλο $\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma} = \sigma^2\boldsymbol{V})$, όπου $\boldsymbol{X}$ είναι ένας $n \times p$ πίνακας συμμεταβλητών με $rank(X) = p$ και ο πίνακας $\boldsymbol{V}$ είναι <u>γνωστός</u> p.d. πίνακας.

   (α΄) Γράψτε τη πιθανοφάνεια που προκύπτει από το μοντέλο.

   (β΄) Βρείτε τους εκτιμητές μεγίστης πιθανοφάνειας των $\boldsymbol{\beta}$ και σ^2. (Μπορείτε να κάνετε χρήση αποτελεσμάτων του Θεωρήματος 8 στις σημειώσεις μου)

   (γ΄) Βρείτε τη μέγιστη τιμή που παίρνει ο λογάριθμος της πιθανοφάνειας.

   (δ΄) Βρείτε τον πίνακα $cov(\hat{\boldsymbol{\beta}}_{LS}, \hat{\boldsymbol{\beta}}_{MLE})$.

   (ε΄) Βρείτε τον πίνακα $cov(\hat{\boldsymbol{\beta}}_{LS}, \hat{\boldsymbol{\epsilon}}_{LS})$, όπου $\hat{\boldsymbol{\epsilon}}_{LS} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{LS}$.

3. (20 μονάδες) Θεωρήστε το κανονικό γραμμικό μοντέλο $\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, όπου $\boldsymbol{X}$ είναι ένας $n \times p$ πίνακας συμμεταβλητών με $rank(X) = p$. Έστω πως υπάρχει αντιστρέψιμος συμμετρικός πίνακας $\boldsymbol{F}$ για τον οποίο $\boldsymbol{\Sigma}\boldsymbol{X} = \boldsymbol{X}\boldsymbol{F}$.

   (α΄) Δείξτε πως κάτω από αυτή την συνθήκη ισχύει $\hat{\boldsymbol{\beta}}_{LS} = \hat{\boldsymbol{\beta}}_{MLE}$.

   (β΄) Βρείτε την σ.π.π. του $\hat{\boldsymbol{\beta}}_{MLE}$. Η σ.π.π. να γραφεί σα συνάρτηση του $\boldsymbol{F}$.

4. (10 μονάδες) Θεωρήστε 200 παρατηρήσεις από ένα γραμμικό μοντέλο πολλαπλής παλινδρόμησης με μια ποιοτική συμμεταβλητή με 5 στάθμες (κατηγορίες) και μια ποσοτική μεταβλητή.
   (α΄) Γράψτε το μοντέλο σαν Γραμμικό μοντέλο που περιλαμβάνει τους κύριους παράγοντες παραπάνω και τις αλληλεπιδράσεις μεταξύ της ποιοτικής και της ποσοτικής μεταβλητής.
   (β΄) Γράψτε την σ.σ.ε. για τον έλεγχο της μηδενικής υπόθεσης που λέει πως δεν χρειάζεται να συμπεριληφθεί καμμιά αλληλεπίδραση.
   (γ΄) Γράψτε την σ.σ.ε. για τον έλεγχο της μηδενικής υπόθεσης που λέει πως μπορούμε να αγνοήσουμε την ποιοτική μεταβλητή.
   (δ΄) Βρείτε με χρήση της R την κρίσιμη περιοχή του ελέγχου όταν $\alpha = 0.01$.

5. (15 μονάδες) 5.2 και 5.5 από το βιβλίο του *Searle*, σελίδες 154-155.

6. (15 μονάδες) Θεωρήστε ένα πείραμα σε ποντίκια εργαστηρίου που αφορά στην εξεύρεση 'κατάλληλης' δοσολογίας ενός ποντικοφάρμακου. Η βιολόγος κάνει ένεση ποντικοφάρμακου σε επίπεδο δόσης x_{i1}, στο τυχαία επιλεγμένο ποντίκι i $(i = 1, \ldots, 18)$ που είναι υπέρβαρο ή όχι ($x_{i2} = 1$, αν είναι υπέρβαρο) και παρατηρεί αν το ποντίκι πέθανε σε χρονικό διάστημα 10 λεπτών. Η απόκριση, Y_i, παίρνει τη τιμή 1 αν το ποντίκι πεθάνει, αλλιώς παίρνει τη τιμή 0.

   (α΄) Γράψτε ένα γενικευμένο γραμμικό μοντέλο για τα δεδομένα σας.
   (β΄) Γράψτε την πιθανοφάνεια του μοντέλου.
   (γ΄) Γράψτε τις εκτιμητικές εξισώσεις για την ML εκτίμηση των παραμέτρων, χρησιμοποιώντας τα σφάλματα.
   (δ΄) Στο αρχείο 'data for logistic regression.docx' θα βρείτε τα βασικά αποτελέσματα της ανάλυσης του πειράματος.
      i. Επιβεβαιώσετε πως οι εκτιμήσεις των παραμέτρων είναι όντως ML εκτιμήσεις.
      ii. Δώστε εκτίμηση του λόγου των odds θανάτου δύο ποντικιών που παίρνουν την ίδια δόση και όπου το ένα ποντίκι είναι υπέρβαρο (το άλλο όχι).

7. (15 μονάδες) Βρείτε τις εκτιμητικές εξισώσεις μεγίστης πιθανοφάνειας για τη *Poisson* παλινδρόμηση με συμμεταβλητές $x_1, x_2 = x_1^2, x_3$ (να υπάρχει και σταθερός όρος). (Π.χ., στο παράδειγμα με το ΑΧΕΠΑ να λαμβάνεται υπόψη και η θερμοκρασία κατά την ημέρα της μέτρησης στο εργαστήριο).

<u>Σημείωση:</u> Να εξηγείτε κάθε βήμα και να δίνετε τις εντολές της R όταν το ερώτημα ζητεί χρήση της.