Πρώτη Εργασία Γραμμικών και Γενικευμένων Γραμμικών Μοντέλων, ΠΜΣ, ΣΑΧΜ, Πανεμιστήμιο Αιγαίου.
Καταληκτική ημερομηνία: **15/05/2026**

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

1. (30 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 $\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y}$, 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 $\boldsymbol{\mu}^\top = (2, 2, 1)$ και $\boldsymbol{A}$ ο πίνακας που βρήκατε παραπάνω.
      i . Γράψτε τη κατανομή της $(Y_2|Y_1 = y_1, Y_3 = y_3)$.
      ii . Βρείτε το 1ο τεταρτημόριο, τη διάμεσο και το 3ο τεταρτημόριο της κατανομής της $(Y_2|Y_1 = 1, Y_3 = 2)$. Να μη γίνει χρήση R εδώ.
      iii . Υπολογίστε την τιμή της σ.π.π. $f_{Y_2|Y_1=1,Y_3=2}(1.25)$. Επιβεβαιώστε την απάντηση με την R.
      iv . Υπολογίστε την τιμή της α.σ.κ. $F_{Y_2|Y_1=1,Y_3=2}(1.25)$. Επιβεβαιώστε την απάντηση με την R.
      v . Γράψτε τη κατανομή της $\left((Y_1, Y_2)^\top|Y_3 = y_3\right)$.
      vi . Υπολογίστε την τιμή της σ.π.π. $f_{Y_1,Y_2|Y_3=2}(2, 2)$. Επιβεβαιώστε την απάντηση με την R.
      vii . Με χρήση της R υπολογίστε την $P(Y_1 > 2, Y_2 > 2)$.
      viii . Με χρήση της R υπολογίστε την $P(Y_1 > 2, Y_2 > 2|Y_3 = 1)$.

   (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 $Q = (\boldsymbol{Y} - \boldsymbol{\mu})^\top \boldsymbol{B}(\boldsymbol{Y} - \boldsymbol{\mu})$, με $\boldsymbol{B} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{A}\boldsymbol{\Sigma}^{-1/2}$.
      ii . Προσεγγίστε με προσωμοίωση ($n_{sim} = 100000$) την $P(Q > 10)$ με χρήση της R.
      iii . Προσεγγίστε με προσωμοίωση ($n_{sim} = 100000$), με χρήση της R, το 25% ποσοστιαίο σημείο της κατανομής της Q.

2. (8 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 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. (25 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 σαν συνάρτηση του $\boldsymbol{F}$.
   (γ′) Θεωρήστε το απλό γραμμικό μοντέλο τυχαίων παραγόντων με $n_i = n$. Δείξτε πως στο μοντέλο αυτό ισχύει η παραπάνω συνθήκη, $\boldsymbol{\Sigma}\boldsymbol{X} = \boldsymbol{X}\boldsymbol{F}$. Βρείτε τον πίνακα $\boldsymbol{F}$ και την κατανομή του $\hat{\boldsymbol{\beta}}_{MLE}$ σε αυτό το μοντέλο.

(δ΄) Βρείτε τον πίνακα πληροφορίας του $Fisher$ που αφορά <u>όλες</u> της παραμέτρους στο απλό γραμμικό μοντέλο τυχαίων παραγόντων με $n_i = n$.

4. (10 μονάδες) Θεωρήστε 400 παρατηρήσεις από ένα κλασσικό γραμμικό μοντέλο πολλαπλής παλινδρόμησης με μια ποιοτική συμμεταβλητή με 5 στάθμες (κατηγορίες) και δυο ποσοτικές μεταβλητές. Τα σφάλματα είναι ασυσχέτιστα κι ομοσκεδαστικά κι ακολουθούν κανονική κατανομή.
   (α΄) Γράψτε ένα Γραμμικό μοντέλο που περιλαμβάνει τους κύριους παράγοντες παραπάνω και τις αλληλεπιδράσεις μεταξύ της ποιοτικής συμμεταβλητής και των δυο ποσοτικών μεταβλητών.
   (β΄) Γράψτε την σ.σ.ε. για τον έλεγχο της μηδενικής υπόθεσης που λέει πως δεν χρειάζεται να συμπεριληφθεί καμμιά αλληλεπίδραση. Βρείτε με χρήση της R την κρίσιμη περιοχή του ελέγχου όταν $\alpha = 0.01$.
   (γ΄) Γράψτε την σ.σ.ε. για τον έλεγχο της μηδενικής υπόθεσης που λέει πως μπορούμε να αγνοήσουμε την ποιοτική μεταβλητή. Βρείτε με χρήση της R την κρίσιμη περιοχή του ελέγχου όταν $\alpha = 0.01$.

5. (12 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).
      iii . Δώστε εκτίμηση του λόγου των odds θανάτου δύο μη υπέρβαρων ποντικιών όπου το ένα παίρνει δόση κατά 5 μονάδες παραπάνω από το άλλο.

7. (15 μονάδες) Η $Poisson$ παλινδρόμηση με $offset$ μεταβλητή c ορίζεται ως:

$$\begin{aligned} Y_i &\sim^{ind} Poisson(\mu_i) \\ \mu_i &= c_i exp(\beta_0 + \beta_1 x_i), \;\; i = 1, \ldots, n. \end{aligned}$$

   Π.χ., στο παράδειγμα με το ΑΧΕΠΑ το $offset$ θα μπορούσε να είναι ο όγκος του δείγματος των λυμμάτων όπου γίνεται η καταμέτρηση.
   (α΄) Γράψτε τον λογάριθμο της πιθανοφάνειας.
   (β΄) Βρείτε τις εκτιμητικές εξισώσεις μεγίστης πιθανοφάνειας για τη $Poisson$ παλινδρόμηση με $offset$.
   (γ΄) Βρείτε αναλυτικά την ασυμπτωτική κατανομή του $\hat{\boldsymbol{\beta}}_{MLE}$.

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