

Δεύτερη Εργασία Γραμμικών και Γενικευμένων Γραμμικών Μοντέλων, ΠΜΣ, ΣΑΧΜ,
Πανεπιστήμιο Αιγαίου.

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

1. Χρησιμοποιώντας την R , προσομοιώστε δεδομένα από ένα απλό μοντέλο τυχαίων συντελεστών (τυχαίων ευθειών) για 50 άτομα και δέκα συνεχείς ημέρες. Το μέσο (πληθυσμιακό) intercept να είναι ίσο με $\beta_0 = 50$ και η (πληθυσμιακή) μέση κλίση να είναι ίση με $\beta_1 = 1$. Οι διακυμάνσεις να είναι: για τα (subject-specific) intercepts $\sigma_0^2 = 1$, για τις (subject-specific) κλίσεις $\sigma_1^2 = 1$, και για το σφάλμα παρατήρησης $\sigma^2 = 16$. Η συν-διακύμανση μεταξύ των intercepts και των κλίσεων να είναι ίση με $\sigma_{01} = -0.05$.
 - (α') Δώστε κατάλληλα γραφήματα για τα δεδομένα που γεννήσατε.
 - (β') Χρησιμοποιώντας τις κατάλληλες εντολές στην R , προσαρμόσετε το μοντέλο και εκτιμήσετε τις παραμέτρους.
 - (γ') Δώστε ένα 95% δ.ε. για το μέσο (πληθυσμιακό) intercept.
 - (δ') Δώστε ένα 95% δ.ε. για τη (πληθυσμιακή) μέση κλίση.
 - (ε') Κάνετε ελέγχο της υπόθεση έλλειψης χρονικής επίδρασης στην μέση απόκριση σε ε.σ. 5%.
 - (ς') Δώστε ένα 95% δ.ε. για τη μέση (πληθυσμιακή) τιμή της απόκρισης την 8η ημέρα.

2. (Συνέχεια της προηγούμενης) Προσομοιώστε δεδομένα από ένα μοντέλο τυχαίων συντελεστών για 100 άτομα και δέκα ημέρες. Τα 50 άτομα να είναι στο placebo group και τα υπόλοιπα 50 άτομα στο drug group. Το μέσο intercept για το placebo group να είναι 50 και η μέση κλίση για το placebo group να είναι 1. Το μέσο intercept για το drug group να είναι 50 και η μέση κλίση για το drug group να είναι 1.50. Οι διακυμάνσεις να είναι: για τα intercept 1, για τις κλίσεις 1, για το σφάλμα παρατήρησης 16 και η συνδιακύμανση μεταξύ intercept και κλίσης να είναι -0.05.
 - (α') Δώστε κατάλληλα γραφήματα για τα δεδομένα που γεννήσατε.
 - (β') Χρησιμοποιώντας τις κατάλληλες εντολές στην R , προσαρμόσετε το μοντέλο και εκτιμήσετε τις παραμέτρους.
 - (γ') Προσαρμόσετε ένα μοντέλο που να μη λαμβάνει καθόλου υπόψη το φάρμακο (θεραπεία).
 - (δ') Ελέγξτε την υπόθεση που αναφέρει πως το φάρμακο δεν επηρεάζει την μέση απόκριση.
 - (ε') Ελέγξτε την υπόθεση που αναφέρει πως το φάρμακο δεν επηρεάζει την χρονική πορεία της μέσης απόκρισης. Ο έλεγχος να γίνει με χρήση του Likelihood ratio test statistic.
 - (ς') Δώστε 95% δ.ε. για την διαφορά των κλίσεων στις 2 ομάδες.
 - (ζ') Δώστε 95% δ.ε. για την διαφορά της μέσης απόκρισης μεταξύ των 2 ομάδων κατά την 8η ημέρα.

3. Θεωρήστε πως η μέση τιμή ημερήσιων θανάτων σε μια χώρα από μία πανδημία ακολουθεί το μοντέλο:

$$\mu_t = \frac{9000\sqrt{2\pi}}{45} \phi\left(\frac{t-100}{45}\right), t = 1, 2, \dots$$

όπου t αναφέρεται στην ημέρα, και η συνάρτηση $\phi(\cdot)$ είναι η συνάρτηση πυκνότητας πιθανότητας μιας τυπικής κανονικής κατανομής. Για τους παρατηρούμενους ημερήσιους αριθμούς θανάτων, Y_t , υποθέσετε πως είναι ανεξάρτητοι μεταξύ τους και ακολουθούν την $Poisson(\mu_t)$ κατανομή.

- (α') Δώστε κατάλληλα γραφήματα για τα δεδομένα που γεννήσατε.
- (β') Δείξτε πως το παραπάνω μοντέλο είναι ένα γενικευμένο γραμμικό μοντέλο. Τι ιδιαίτερο έχει;
- (γ') Σε ποια ημέρα αναμένουμε τους περισσότερους θανάτους; Ποια η αναμενόμενη τιμή των θανάτων σε αυτή την μέρα;
- (δ') Χρησιμοποιώντας την R , προσομοιώστε τιμές ημερησίων θανάτων, για $t = 1, 2, \dots, 120$.
- (ε') Προσαρμόστε το κατάλληλο γενικευμένο γραμμικό μοντέλο στα δεδομένα. Εκτιμήστε την αναμενόμενη τιμή ημερησίων θανάτων για την 121η ημέρα και δώστε ένα 90% διάστημα εμπιστοσύνης για αυτήν. Σχολιάστε!

4. Θεωρήστε πως η τιμή μιας αιματολογικής εξέτασης, x , είναι ενδεικτική μιας ασθένειας (μεγαλύτερη τιμή σημαίνει 'περισσότερη' ένδειξη ασθένειας). Έστω x_i η τιμή της αιματολογικής εξέτασης για το i -οστό άτομο και Y_i η δείκτρια ασθένειας (=1 αν το i -οστό άτομο βρέθηκε, μετά από βιοψία, ότι είχε τελικά ασθένεια). Εδώ Y_i είναι η απόκριση, ενώ x_i είναι η συμμεταβλητή.

- (α') Γράψτε ένα απλό γενικευμένο γραμμικό μοντέλο για τα δεδομένα.
- (β') Ποιες είναι οι τιμές των παραμέτρων του μοντέλου σας αν δίνεται πως 1) τα *odds* ασθένειας στην τιμή $x = 10$ είναι ίση με $e^{0.4}$, 2) τα *odds* ασθένειας στην τιμή $x = 20$ είναι ίση με $e^{0.8}$;
- (γ') Δώστε το γράφημα της συνάρτησης της πιθανότητας ασθένειας με τις παραμέτρους που βρήκατε.
- (δ') Χρησιμοποιώντας την R , και χρησιμοποιώντας τιμές $x_1, x_2, \dots, x_{200} \sim Unif(0, 30)$ προσομοιώστε 200 τιμές των Y_i , με τις παραμέτρους που βρήκατε.
- (ε') Χρησιμοποιώντας τα δεδομένα που γεννήσατε, εκτιμήστε, με χρήση R , τις παραμέτρους κι ελέγξτε τη στατιστική σημαντικότητα της παλινδρόμησης σε ε.σ. 5%.
- (ε') Δώστε το γράφημα της εκτίμησης της συνάρτησης της πιθανότητας ασθένειας με τις εκτιμήσεις που βρήκατε. Το γράφημα αυτό να δοθεί μαζί με το γράφημα στο (γ) σε κοινό plot.

Σημείωση: Στις προσομοιώσεις σας ξεκινάτε με `seed=` ημερομηνία γέννησής σας και προσθέτετε +1 για την επόμενη.

BONUS: Στα Ερωτήματα 1 και 2 βρείτε τις εκτιμήσεις των παραμέτρων μεγιστοποιώντας απευθείας τις πιθανοφάνειες (αφού πρώτα τις ορίσετε στη R , σύμφωνα με τις σημειώσεις). Σχολιάστε.