Auf diesem Aufgabenblatt beschäftigen wir uns zum einen mit marginalen Modellen für nicht normalverteilte longitudinale Daten – insbesondere mit der Schätzung mittels generalisierte Schätzgleichungen (GEEs) – und zum anderen mit fehlenden Werten. Die zu bearbeitenden Aufgaben beziehen sich auf die Inhalte der zehnten und elften Vorlesungsfolien.

Aufgabe 1: GEE

Im Datensatz `leprosylong.txt` sind die Anzahl von Lepra-Bazillen vor und nach der Behandlung mit Antibiotika enthalten. Es existieren für jeden Patienten folgende Variablen: Medikament (`Drug`), Anzahl der Bazillen (`count`) und Zeit (`time`). Medikament C entspricht einer Placebo-Behandlung. `time = 0` entspricht dem Zustand vor der Behandlung, `time = 1` dem Zustand nach der Behandlung.

(a) Bestimmen Sie Mittelwert und Varianz der Anzahlen der Bazillen vor und nach der Behandlung für jede Medikamentengruppe und plotten Sie die Verläufe getrennt nach Gruppen. Für welche Einflussgrößen erwarten Sie einen signifikanten Effekt und was würden Sie an der Studie kritisieren?

(b) Fitten Sie zunächst ein Poisson-GLM mit Zeit, Behandlung, sowie deren Interaktion als Einflussgrößen.

- **(i)** Interpretieren Sie die Koeffizientenschätzer.
- **(ii)** Berechnen Sie anhand der Pearson-Residuen einen Schätzer für den Dispersionsparameter ϕ.

(c) Welche Annahmen müssen für die Schätzung eines marginalen Modells getroffen werden?

(d) Schätzen Sie ein marginales Modell mit den gleichen Einflussgrößen wie das Modell in (b) mittels GEE. Gehen Sie zunächst von unabhängigen Beobachtungen aus.
Hinweis: Verwenden Sie die Funktion `gee` im Paket `gee` mit dem Argument `corstr = 'independence'`.

- **(i)** Was verbirgt sich hinter den „robusten“ und den „naiven“ Schätzungen der Varianzen der Parameter?
- **(ii)** In welchen Fällen ist die robuste Varianzschätzung geeignet?
- **(iii)** In welchem Fall sind die beiden Varianzschätzungen gleich?

(e) Lesen Sie sich die Hilfeseite zu `?quasipoisson` durch und fitten Sie nun ein Quasipoisson-Modell mit der Funktion `glm`.
Hinweis: Verwenden Sie hierfür `family=quasipoisson(link='log')`.

(i) Was fällt Ihnen auf, wenn Sie die Ergebnisse mit denen aus (d) vergleichen?

(ii) Zeigen Sie, dass die Koeffizientenschätzer des einfachen Poissonmodells mit denen des Quasi-Poisson Modells übereinstimmen.

(f) Benutzen Sie die Pearson-Residuen des Quasipoisson-Modells, um eine vorläufige Schätzung für die Korrelation zwischen den zwei Messungen eines Patienten zu erhalten.

(g) Schätzen Sie das Modell erneut unter der realistischeren Annahme, dass Beobachtungen am selben Patienten nicht unabhängig sind. Verwenden Sie eine unstrukturierte Korrelation.

(h) Worauf bezieht sich das "generalized" im Namen der Schätzgleichungen GEE?

Aufgabe 2: Fehlende Werte

In dieser Aufgabe geht es um das Einordnen von fehlenden Werten in Missing-Mechanismen. Dies kann sehr wichtig sein um entscheiden zu können, welche Methoden sich anwenden lassen und welche nicht. Betrachten Sie das folgende LMM

$$Y_{ij} = \beta_{x:t} x_{ij} t_{ij} + b_i + \epsilon_{ij},$$

mit $i = 1, \ldots, N = 30$ Personen und den beiden Zeitpunkten $t_{i1} = 0$ und $t_{i2} = 1$, für alle $i = 1, \ldots, N$. x_{ij} sei eine binäre Einflussgröße.

Sei R_i ein Indikator für das Beobachten der zweiten Messung für die i-te Person, d.h. $R_i = 1$ wenn Y_{i2} beobachtet wurde und $R_i = 0$, wenn Y_{i2} fehlt.

(a) Wieso handelt es sich um Dropout, wenn fehlende Werte für Y_{i2} vorliegen?

Im folgenden betrachten wir verschiedene Dropoutwahrscheinlichkeiten.

(b) Betrachten Sie $P(R_i = 1 | Y_{i1}, Y_{i2}, \mathbf{X})$ für $R_i \sim \text{Bernoulli}(\pi_i)$ mit

$$\text{logit}(\pi_i) = \gamma_0 + \gamma_x x_{i1} + \gamma_1 Y_{i1} + \gamma_2 Y_{i2}.$$

Geben Sie die Restriktionen für $(\gamma_0, \gamma_1, \gamma_2)$ an wenn der Dropout

(i) MCAR

(ii) MAR

(iii) NMAR

ist. Finden Sie zudem Beispiele für alle drei Arten von fehlenden Werten.

(c) Sei stattdessen

$$\text{logit}(\pi_i) = \alpha_0 + \alpha_1 b_i,$$

mit $\alpha_0 = -0.5$ und $\alpha_1 = 3$. Um was für einen Dropout-Mechanismus handelt es sich hier? Begründen Sie Ihre Antwort.