Uogólnione modele liniowe

Kolokwium, 03.02.2014

1. Plik chapman.dat zawiera dane dotyczące dwustu dorosłych mężczyzn. W zbiorze jest siedem zmiennych:
   - (Ag) - wiek (w latach),
   - (S) - skurczowe ciśnienie krwi (w milimetrach słupka rtęci),
   - (D) - rozkurczowe ciśnienie krwi (w milimetrach słupka rtęci),
   - (Ch) - cholesterol (w miligramach na decylitr),
   - (H) - wzrost (w calach),
   - (W) - waga (w funtach),
   - (CNT) - choroba wieńcowa – 1, jeśli choroba pojawiła się w ciągu ostatnich 10 lat; 0 - jeśli się nie pojawiła.

   Celem jest modelowanie prawdopodobieństwa p pojawienia się choroby wieńcowej (pozostałe zmienne mogą występować jako zmienne objaśniające). Jeśli zastosować model logistyczny ze wszystkimi (oprócz CNT) sześcioma zmiennymi objaśniającymi, ile wynosi $\hat{p}$ dla sześćdziesięcioletniego mężczyzny z ciśnieniem 140/90, cholesterolem 200, mierzącego 69 cali i ważącego 200 funtów?

2. Zbiór birth.txt zawiera informacje o liczbie dzieci posiadanych przez każdą z badanych 141 kobiet. Poza liczbą dzieci zawarta w zbiorze jest informacja o wieku kobiet. Zmienne:
   - (age) wiek kobiety
   - (children) - liczba dzieci.

   Rozważamy model poissonowski $y \sim \text{Poiss}(\lambda)$, $g(\mu) = x\beta$, przy czym y - liczba dzieci, x - wiek kobiety, μ - średnia liczba dzieci.

   (a) Dopasować powyższy model poissonowski z kanoniczną funkcją łączącą. Ocenić jakość dopasowania. Ile wynosi $\hat{\mu}$ (jako funkcja od x)?

   (b) Dopasować powyższy model poissonowski z identycznościową funkcją łączącą. Ocenić jakość dopasowania. Ile tym razem wynosi $\hat{\mu}$?

   (c) Jak zmienia się, według wskazań modelu z kanoniczną funkcją łączącą, liczba dzieci wraz ze wzrostem wieku matki o rok? Jaka jest odpowiedź na analogiczne pytanie w sytuacji identycznościowej funkcji łączącej?

   (d) Narysować na jednym wykresie krzywe wartości dopasowanych z obydwu modeli (wykresy liczba dzieci jaku funkcja wieku kobiet).

   (e) Ocenić istotność zmiennej age w obydwu modelach.

3. Zbiór dane.txt zawiera obserwacje zmiennych $x1$, $x2$ i y.

   (a) Dopasować do zmiennej y odpowiednią kombinację liniową funkcji zmiennych objaśniających $x1$ i $x2$. Ocenić jakość dopasowania oraz istotność predyktorów.

   (b) Dopasować do zmiennej y kombinację liniową funkcji f_1, f_2:

   $$f_1(x,c) = \begin{cases} x - c, & x > c, \\ 0, & wpp. \end{cases},$$

   $$f_2(x,c) = \begin{cases} c - x, & x < c, \\ 0, & wpp. \end{cases}$$

   od zmiennych objaśniających (z odpowiednio dobranymi wartościami c). Ocenić jakość dopasowania oraz istotność predyktorów.

   (c) Porównać jakość dopasowania powyższych modeli z jakością dopasowania zwykłego modelu liniowego.