8. Dane zawierają liczbę przypadków zachorowania na AIDS w Australii w kolejnych kwartałach od 1984 do 1988 roku w zależności od daty diagnozy (lista 3.xls). W tej fazie epidemii wydaje się, że liczba zachorowań rośnie wykładniczo.

| | Kwartały | | | |
|---|---|---|---|---|
| Lata | 1 | 2 | 3 | 4 |
| 1984 | 1 | 6 | 16 | 23 |
| 1985 | 27 | 39 | 31 | 30 |
| 1986 | 43 | 51 | 63 | 70 |
| 1987 | 88 | 97 | 91 | 104 |
| 1988 | 110 | 113 | 149 | 159 |

(A) Narysuj wykres liczby zachorowań $y_i$ jako funkcji czasu $i$ $(i = 1, 2, \ldots, 20)$

(B) Jakie przekształcenie $y_i$ oraz $i$ linearyzuje ten wykres?

(C) Rozsądnym modelem dla $y$ może być rozkład Poissona. Co można zaproponować jako funkcję łączącą i co będzie zmienną objaśniającą (patrz punkt (B))?

(D) Przeprowadź estymację parametrów modelu $GLIM$ i oceń jego jakość. Spróbuj interpretować znaczenie parametrów modelu.

(E) Jaki model wyjaśniłby, czy różnica liczby zachorowań w kolejnych kwartałach jest istotna w rozwoju epidemii AIDS?

9. W tabeli podane są: liczba białych ciałek krwi ($BCK$) w momencie diagnozy oraz długość życia (w tygodniach) ($DZ$) od momentu diagnozy dla 17 chorych na białaczkę (lista 3.xls).

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $BCK$ | 2291 | 759 | 4266 | 2570 | 6026 | 10471 | 10000 | 16982 | 5370 |
| $DZ$ | 65 | 156 | 100 | 134 | 16 | 108 | 121 | 4 | 39 |
| $BCK$ | 7079 | 9333 | 32359 | 34674 | 100000 | 100000 | 52481 | 100000 | |
| $DZ$ | 143 | 56 | 26 | 22 | 1 | 1 | 5 | 65 | |

(A) Narysuj wykres $DZ$ jako funkcji czasu $BCK$. Jakie przekształcenie linearyzuje ten wykres?

(B) Rozsądnym modelem dla $DZ$ może być rozkład wykładniczy. Co można zaproponować jako funkcję łączącą i co będzie zmienną objaśniającą (patrz punkt (A))?

(C) Przeprowadź estymację parametrów modelu $GLIM$ i oceń jego jakość. Spróbuj interpretować znaczenie parametrów modelu.

(D) Porównaj wartości rzeczywiste $DZ$ i estymowane z modelu $\widehat{DZ}$. Oblicz reszty standaryzowane

$$r_i = \frac{DZ_i - \widehat{DZ}_i}{\widehat{DZ}_i}$$

Skomentuj uzyskane wyniki.

10. Dysponujemy *jedną* obserwacją $Y$ z rozkładu dwumianowego $b(n, \pi)$.

(A) Znajdź wartość statystyki Walda $(\hat{\pi} - \pi)'\mathcal{I}(\hat{\pi} - \pi)$, gdzie $\hat{\pi}$ jest estymatorem największej wiarygodności parametru $\pi$ oraz $\mathcal{I}$ jest informacją Fishera.

(B) Sprawdź, że statystyka Walda jest w tym przypadku identyczna ze statystyką punktową $U'\mathcal{I}^{-1}U$.

(C) Oblicz odchylenie

$$2(l(\hat{\pi}, y) - l(\pi, y)).$$

(D) Asymptotycznie, statystyka Walda i odchylenie mają rozkład $\chi^2(1)$. Użyj obu metod do oceny zgodności danych $n = 10$ i $y = 3$ z modelami: (i) $\pi = 0.1$, (ii) $\pi = 0.3$, (iii) $\pi = 0.5$. Czy obie metody testowania prowadzą do tego samego wniosku?