15. $Y_1, Y_2, \cdots, Y_N$ są niezależnymi zmiennymi losowymi o rozkładzie Poissona, $E(Y_i) = \mu_i$ i $log(\mu_i) = \beta_1 + \sum_{j=2}^{J} x_{ij}\beta_j \quad i = 1, 2, \cdots, N$.Pokaż, że

(A) Statystyka punktowa (*score statistics*) dla $\beta_1$ wyraża się wzorem $U_1 = \sum_{i=1}^{N}(Y_i - \mu_i)$

(B) Estymatory największej wiarygodności $\hat{\mu}_i$ spełniają warunek $\sum_{i=1}^{N} \hat{\mu}_i = \sum_{i=1}^{N} y_i$

(C) Odchylenie dla modelu Poissona, postaci

$$D = 2\sum_{i=1}^{N}[o_i log(\frac{o_i}{e_i}) - (o_i - e_i)]$$

przybiera w tym przypadku postać $D = 2\sum_{i=1}^{N} o_i log(\frac{o_i}{e_i})$

16. Dane (plik: polisy.xls) dotyczą ubezpieczeń komunikacyjnych w dużych miastach Wielkiej Brytanii: liczbę $n$ polis ubezpieczeniowych, liczbę roszczeń $y$ dla samochodów w różnych kategoriach ubezpieczeniowych $CAR$ wraz z kategorią wieku kierowców $WIEK$ oraz miejscem zamieszkania właściciela polisy $MJSC$ (1 dla Londynu, 0 - dla innych miast)

| $CAR$ | $WIEK$ | $MJSC = 0$ | | $MJSC = 1$ | |
|---|---|---|---|---|---|
| | | $y$ | $n$ | $y$ | $n$ |
| 1 | 1 | 65 | 317 | 2 | 20 |
| 1 | 2 | 65 | 476 | 5 | 33 |
| 1 | 3 | 52 | 486 | 4 | 40 |
| 1 | 4 | 310 | 3259 | 36 | 316 |
| 2 | 1 | 98 | 486 | 7 | 31 |
| 2 | 2 | 159 | 1004 | 10 | 81 |
| 2 | 3 | 175 | 1355 | 22 | 122 |
| 2 | 4 | 877 | 7660 | 102 | 724 |
| 3 | 1 | 41 | 223 | 5 | 18 |
| 3 | 2 | 117 | 539 | 7 | 39 |
| 3 | 3 | 137 | 697 | 16 | 68 |
| 3 | 4 | 477 | 3442 | 63 | 344 |
| 4 | 1 | 11 | 40 | 0 | 3 |
| 4 | 2 | 35 | 148 | 6 | 16 |
| 4 | 3 | 39 | 214 | 8 | 25 |
| 4 | 4 | 167 | 1019 | 33 | 114 |

(A) Użyj regresji Poissona do oceny związku między frakcją wystąpienia roszczeń $\frac{y}{n}$ a czynnikami $CAR$, $WIEK$ i $MJSC$, potraktowanymi jako zmienne nominalne, wraz z interakcjami

(B) Spróbuj regresji Poissona, kodując zmienną $WIEK$ porządkowo (np tworząc 3 zmienne nominalne: dla wieku 1, dla wieku nie późniejszego niż 2 i dla wieku nie późniejszego niż 3). Porównaj otrzymane wyniki z wynikami z (A).