

Ekonometria, 2014/15

Zestaw zadań nr 3

Zadanie 1. W zbiorze danych `pima` zdecyduj która ze zmiennych powinna być zmienną objaśnianą, a które zmienne powinny być objaśniające. Uzasadnij, dlaczego taki model ma sens i podaj jego interpretację.

Zadanie 2. Na podstawie posiadanej wiedzy (używając danych `bwt lwt` `race smoke` z pakietu `birthwt` z biblioteki `MASS`) dopasuj poprawny model objaśniający wagę noworodków.

Zadanie 3. W zbiorze danych `prostate` (co to za dane? zob. dokumentację) dopasuj model w którym `lpsa` jest zmienną objaśnianą, a wszystkie pozostałe zmienne są objaśniające.

- Oblicz 90% i 95% przedziału ufności dla parametru związanego z wiekiem. Jaki jest związek tych przedziałów z p -value?
- Oblicz i narysuj 95% zbiór ufności dla parametrów związanych z wiekiem i `Ibph` . Umieść środek układu współrzędnych na wykresie. Miejsce środka układu współrzędnych związane jest z pewną hipotezą - jaka to hipoteza?
- Załóżmy, że mamy nowego pacjenta z danymi `lcavol 1.44692` `lweight 3.62301` `age 65.00000` `Ibph 0.30010` `svi 0.00000` `lcp 0.79851` `gleason 7.00000` `pgg45 15.00000`. Oblicz 95% przedziału ufności dla `lpsa` .
- W modelu odczytaj istotne statystycznie współczynniki, podaj interpretacje.

Zadanie 4. Dla różnych wartości parametru $a \in (0, 1)$ przetestuj znanymi testami (np. `library{nortest}`) czy wektor `x=(1-a)*rnom(100)+a*runif(100)` ma rozkład normalny.

Zadanie 5. Utwórz zmienna `x` składająca się z około 100 liczb, a następnie niech `y=5*x+eps`, gdzie za `eps` podstawiamy kolejno rozkład normalny `eps=rnorm(100)`, rozkład wykładniczy, i kolejne znane rozkłady (najlepiej zmieniając parametry). Dopasuj model liniowy $y \sim x$ i zaobserwuj jak ze zmianą `eps` zmieniają się współczynniki w tym modelu oraz jak wygląda $Q-Q$ plot.

Zadanie 6. Na dowolnym zbiorze danych z biblioteki `faraway` przeprowadź następujące czynności:

- dopasuj model liniowy, dbając o jego sensowność w rzeczywistości (na przykład bez sensu jest próba objaśnienia dochodu rodziny za pomocą płci dziecka i jego wydatków na hazard)

- *przeprowadź poznanymi metodami diagnostykę modelu, ogranicz się przy prezentowaniu zadania do tych wykresów o których masz coś do powiedzenia.*