

Statystyka i Probabilistyka - Projekt 3

Należy przygotować model regresji liniowej wielorakiej (lub wielokrotnej) dla wybranych przez Państwa danych oraz przeprowadzić diagnostykę modelu. W sprawozdaniu powinny znaleźć się następujące rzeczy:

1. Dokładny opis wykorzystanych danych oraz ich źródło. Dane powinny posiadać przynajmniej 30 obserwacji na każdą zmienną niezależną (zalecane jest dużo więcej).
2. Krótki opis teoretyczny modelu regresji wielorakiej
3. Zbudowanie modelu i jego analiza w R. Według pliku, który podesłałem na ostatnich zajęciach. Zbudowanie modelu funkcją `lm`. Interpretacja modelu. Interpretacja wartości z `summary(model)`. Sprawdzenie założeń (normalność, korelacja, autokorelacja reszt, homoskedastyczność itd, wyznaczenie zbioru dla którego model ma sens itd.) Wykresy diagnostyczne z interpretacją. Na koniec wnioski i podsumowanie (Czy model jest ok itd). Redukcja wymiarowości poprzez różne metody usuwania zmiennych niezależnych (regresja krokowa, macierz korelacji). Metody wykrywania wartości wpływowych (Wykresy z odległością cooke'a, reszt studentyzowanych, DFBETAS, DFFITS, usunięcie obserwacji wpływowych i zbudowanie nowego modelu. Sprawdzenie czy modele poprawiły się za pomocą poznanych metryk (R^2 , RMSE itd)