

Ćwiczenie nr 8 – przykład

1. Przygotowywanie danych

- Wczytaj dane z pliku *surgical.txt*. Dane te dotyczą 54 pacjentów, którzy przeszli operację wątroby. O każdym z pacjentów zebrano następujące informacje: *blood* – wynik krzepnięcia krwi, *prognostic* – indeks prognostyczny (zawierający wiek pacjenta), *enzyme* – wynik testu enzymów, *liver* – wynik testu wątroby, *survival* – czas przeżycia po operacji.
- Narysuj wykresy poszczególnych zmiennych. Zidentyfikuj wstępnie obserwacje odstające.
- Dopasuj model pełny (ze zmienną zależną *survival*), oraz inne modele pierwszego rzędu. Narysuj wykresy diagnostyczne.
- Zbadaj (na podstawie wykresów reszt) czy warto rozważać czynniki interakcji.
- Przeprowadź niezbędne transformacje.
- Sporządź ponownie wykresy zależności, wypisz macierz korelacji oraz przeprowadź wstępną diagnostykę

2. Redukcja ilości zmiennych niezależnych

- Znajdź najlepsze podzbiory zmiennych niezależnych ze względu na różne kryteria.

3. Wybór i udoskonalanie modelu

- Dopasuj model z dodanymi wszystkimi podwójnymi czynnikami interakcji. Sprawdź na podstawie wykresu reszt, czy interakcja między wszystkimi trzema zmiennymi niezależnymi powinna być dodana do modelu.
- Dopasuj model z wszystkimi efektami interakcji (również potrójnymi). Porównaj poprzez analizę wariancji z modelem bez interakcji.
- Narysuj wykresy reszt w zależności od poszczególnych zmiennych niezależnych w modelu, zmiennych wyrzuconych z modelu oraz dopasowanych wartości.
- Poprzez wykresy regresji częściowej (partial regression plots) sprawdź czy zmienna zależna nie wykazuje zależności krzywoliniowej od którejś ze zmiennych. (funkcja **avp** w pakiecie *car*). Jeśli tak to zbadaj ten efekt.
- Zbadaj normalność reszt.
- Poprzez czynnik wzrostu wariancji (variance inflation factor – vif) zbadaj czy zmienne zależne nie są zbyt silnie skorelowane. (funkcja **vif** w pakiecie *car*).
- Wyszukaj obserwacje odstające i wpływowe. Zbadaj ich wpływ.

4. Potwierdzenie poprawności modelu

- Wczytaj dodatkowe dane do sprawdzenia modelu z pliku *validation.txt*.
- Dopasuj model do tych danych. Porównaj z poprzednim wynikiem.
- Przeprowadź predykcję nowych danych na podstawie modelu opartego na poprzednich danych. Porównaj z rzeczywistymi wartościami.