

Ćwiczenie nr 3 – regresja wieloczynnikowa

1. Wczytaj dane z pliku `caincedl.txt`.

Dane przedstawiają wysokość zarobków w Kaloiforni w zależności od płci i wykształcenia badanych osób.

Zweryfikuj hipotezę dotyczącą równości przeciętnych zarobków w Kalifornii.

Przeprowadź testy normalności zarobków i wykształcenia.

Skonstruuj model regresji liniowej prostej dla zmiennej zależnej *zarobki* i niezależnej *wykształcenie*.

2. Wczytaj dane z pliku `smsa.txt`.

Dane zawierają nast zmienne: *land* - powierzchnia w milach kwadratowych *populat* - estymowana wielkość populacji w 1977 (w tysiącach), *pcity* - procent populacji w głównych miastach w 1976, *psen* - procent populacji osób powyżej 65 roku życia w 1976, *nophy* - ilość aktywnych zawodowo lekarzy na dzień 31 grudnia 1977, *nobed* - ilość łóżek szpitalnych w 1977, *highschgr* - procent osób powyżej 25 roku życia, którzy przeszli co najmniej 12 letni kurs edukacji, *laborf* - ilość osób w wieku produkcyjnym w 1977 (w tysiącach), *pinc* - dochód brutto w 1976 (w milionach dolarów), *crime* - całkowita ilość poważnych przestępstw w 1977, *reg* - geograficzna klasyfikacja regionów (1=NE, 2=NC, 3=S, 4=W), *crime.pop* - stosunek zmiennej *crime* do zmiennej *populat*.

Wybierz cztery zmienne niezależne najlepsze do opisu przestępczości i skonstruuj odp. model regresji. Przeprowadź analizę poprawności modelu.

3. Wskazówki:

- `caincedl$zarobki[caincedl$plec==1]` – zarobki mężczyzn.
- `t.test` – test t, `wilcox.test` – test Wilcoxona, `kruskal.test` – test Kruskala-Wallisa.
- `pairs` – wykresy zależności pomiędzy zmiennymi
- `cov`, `cor` – macierze kowariancji i korelacji