

SAR 2020/2021 Laboratorium 3

20-21.10.2020

3.1 (Anscombe quartet)

Wczytać zbiór *anscombe_quartet.txt*.

- Do każdej z czterech par zmiennych dopasować prostą MNK.
- Porównać otrzymane współczynniki dopasowanych prostych MNK, współczynniki R^2 i współczynniki korelacji.
- W jednym oknie sporządzić 4 wykresy rozrzutu Y_i od X_i , $i = 1, 2, 3, 4$. W którym przypadku można mówić o przybliżonej zależności liniowej y od x ?

3.2 (rozkład QR, model regresji wielokrotnej)

Plik *realest.txt* zawiera dane na temat domów na przedmieściach Chicago. Rozważamy liniowy model regresji opisujący zależność ceny domu (zmienna *Price*) od pozostałych zmiennych w zbiorze.

- Wyznacz macierz eksperymentu X .
- Oblicz estymatory parametrów $\hat{\beta}$ z definicji (porównaj z wartościami obliczonymi przy użyciu funkcji `coef()`).
- Zastosuj rozkład QR dla macierzy $X = QR$, Q - macierz ortogonalna $n \times n$, R - macierz trójkątna górna $p \times p$. Udowodnij że $R\hat{\beta} = Q'Y$.
- Oblicz estymatory parametrów korzystając z rozkładu QR macierzy eksperymentu. Sprawdź zgodność z wynikami z punktu (b).

3.3 (model regresji wielokrotnej)

Plik *realest.txt* zawiera następujące dane na temat domów na przedmieściach Chicago: cena domu (*Price*), liczba sypialni (*Bedroom*), powierzchnia w stopach kwadratowych (*Space*), liczba pokoi (*Room*), szerokość frontu działki w stopach (*Lot*), roczny podatek od nieruchomości (*Tax*), liczba łazienek (*Bathroom*), liczba miejsc parkingowych w garażu (*Garage*) i stan domu (*Condition*, 0 - dobry, 1 - wymaga remontu). Dopasować liniowy model regresji opisujący zależność ceny domu od pozostałych zmiennych w zbiorze.

- Wyznacz macierz eksperymentu.
- Oblicz estymatory parametrów z definicji i porównaj z wartościami obliczonymi przy użyciu funkcji `coef()`. Oblicz SST, SSR, SSE i współczynnik determinacji z definicji.
- Jaki wpływ na cenę ma zwiększenie liczby sypialni o 1, kiedy wartości wszystkich pozostałych zmiennych objaśniających są ustalone? Znaleźć uzasadnienie tego pozornie błędnego wyniku. Porównać ten wynik z wynikiem otrzymanym dla modelu liniowego opisującego zależność ceny domu jedynie od liczby sypialni.
- Masz dom w tej okolicy, w dobrym stanie, z 3 sypialniami, o powierzchni 1500 stóp kwadratowych, z 8 pokojami, 40 stopami szerokości działki, 2 łazienkami, 1 miejscem w garażu i podatkiem w wysokości 1000 dolarów. Za ile spodziewasz się go sprzedać? Wykonaj predykcje korzystając z definicji oraz funkcji `predict()`.

- e) Oblicz estymator wariancji błędów korzystając z definicji oraz funkcji `summary()`.
- f) Oblicz estymatory błędów standardowych współczynników korzystając z definicji oraz funkcji `summary()`. Podaj rozkład wektora $\hat{\beta}$. Wypisz estymator macierzy kowariancji estymatora $\hat{\beta}$ i porównaj wynikiem funkcji `vcov`.
- g) Policz estymator wariancji dla predykcji z podpunktu d) ($\text{Var}(y_{pred}) = \text{Var}(\hat{y}) + \text{Var}(\varepsilon)$, $\hat{y} = x'\hat{\beta}$).