

Ćwiczenia 4

1. Używając danych `cars` narysuj wykres rozrzutu, gdzie pierwsza zmienna to prędkość, a druga to droga hamowania. Dopasuj do tego prostą regresji i narysuj ją na wykresie. Następnie dopasuj i narysuj funkcję kwadratową.
2. **(S)** Używając danych `hills` z pakietu `MASS`, zbadaj liniową zależność odległości (`dist`) od czasu (`time`) oraz wspinaczki (`climb`) od czasu. Narysuj uzyskane proste regresji wraz z danymi obok siebie na jednym wykresie.
3. Dla zbioru danych `homedata` (wartości domów w roku 1970 i 2000 w Maplewood w stanie New Jersey (USA)) z pakietu `UsingR`, zbadaj zależność pomiędzy starymi i nowymi cenami (stara cena powinna być zmienną objaśniającą). Dokonaj prognozy ceny domu w roku 2000, który w roku 1970 wart był 75 000 \$.
4. **(S)** Koszt domu zależy od liczby pokoi. Poniższe dane przedstawiają dane dla pewnego miasta.

Cena (w tys.)	300	250	400	550	317	389	425	289	389	559
Liczba pokoi	3	3	4	5	4	3	6	3	4	5

Wykonaj diagram korelacyjny oraz znajdź prostą regresji. Czy na poziomie istotności 0,05 liczba pokoi ma istotny wpływ na cenę? Jaka będzie cena mieszkania dwupokojowego według oszacowanego modelu?

5. **(S)** Zbiór `florida` z pakietu `UsingR` zawiera wyniki wyborów prezydenckich w USA w roku 2000. Narysuj wykres rozrzutu dla wyników uzyskanych przez Busha i Buchanana. Zidentyfikuj dwie najbardziej odstające obserwacje. Dopasuj model liniowy zależności do danych z usuniętymi obserwacjami odstającymi. Jak wygląda prognoza wyników w tym modelu dla Buchanana w okręgu Miami-Dade bazująca na ilości głosów oddanych tam na Busha?
6. **(S)** W zbiorze danych `emissions` (emisja CO₂ a poziom PKB (26 państw)) z pakietu `UsingR` istnieje obserwacja odstająca. Narysuj diagram korelacyjny dla zmiennych GDP oraz CO₂. Na jego podstawie wyznacz obserwację odstającą. Przyjmując zmienną CO₂ za zmienną objaśnianą znajdź prostą regresji z obserwacją odstającą jak również bez niej. Jak zmieniły się wyniki? Narysuj obie proste na jednym wykresie.
7. Zbiór danych `homeprice` z pakietu `UsingR` zawiera informacje na temat domów sprzedanych w New Jersey w roku 2001. Czy liczba toalet (zmienna `half`) miała wpływ na cenę (`sale`)? Jak zmieniają się współczynniki regresji liniowej jeśli usuniemy wyraz wolny z modelu? Czy takie działanie jest sensowne w tym przypadku?

8. (*) Firma elektroniczna okresowo importuje przesyłki pewnej dużej części używanej jako komponent w kilku swoich produktach. Wielkość przesyłki zmienia się w zależności od harmonogramów produkcji. Do obsługi i dystrybucji do zakładów montażowych, przesyłki o wielkości 250 tysięcy części lub mniejsze są wysyłane do magazynu A. Większe przesyłki są wysyłane do magazynu B, ponieważ ten magazyn posiada specjalistyczny sprzęt, który zapewnia większe oszczędności skali dla dużych przesyłek.

Koszt w tys. dolarów (y)	Wielkość przesyłki w tys. części (x)
11,95	225
14,13	350
8,93	150
10,98	200
10,03	175
10,13	180
13,75	325
13,30	290
15,00	400
7,97	125

Narysuj wykres rozrzutu dla tych danych. Na jego podstawie zaproponuj punkt zmiany. Dopasuj odpowiedni model regresji segmentowej. Dodaj go do wykresu. Jaki punkt zmiany został wyestymowany? Zapisz równania regresyjne. Jaki będzie koszt przesyłki o rozmiarze 150, a jaki o rozmiarze 300? Dodaj te predykcje do wykresu.