

## Regresja liniowa i prognozowanie

Model regresji liniowej bywa często używany do prognozowania. Jest to dość ryzykowne. Doświadczenie polega na: (1) wylosowaniu próbki punktów  $(x, y)$  według modelu:

$$y = f(x) + \varepsilon,$$

gdzie  $f$  jest znaną funkcją a  $\varepsilon$  zaburzeniem losowym o znanym rozkładzie, (2) Zbudowaniu na podstawie próbki modelu regresji liniowej, (3) Wykonaniu prognozy na podstawie tego modelu.

### Opis eksperymentu

Zadajemy kolejno: licznosc próbki  $N$ , przedział określoności funkcji  $f$ ,  $a < b$ , długość przedziału prognozy  $c$ , wzór określający funkcję  $f$ , poziom istotności,  $\alpha$ , rozkład zaburzający  $\varepsilon$  oraz metodę wyboru punktów z przedziału  $[a, b]$ . Otrzymujemy wykres na którym oprócz wylosowanych punktów: linia czarna jest wykresem funkcji  $f$ , czerwona prosta jest wyznaczona metodą najmniejszych kwadratów, krzywe zielone wyznaczają przedziały ufności dla średniej prognozy, a krzywe niebieskie przedziały prognozy na poziomie ufności  $1 - \alpha$ . Otrzymujemy także wartość współczynnika determinacji  $R^2$ .

### Uwagi

Warto znaleźć przykłady w których  $R^2$  jest bliskie 1, przedziały prognozy są w miarę dokładne, natomiast prognoza różni się istotnie od „prawdziwej” funkcji  $f$  opisującej obserwowane zjawisko.

### Więcej informacji o problemie

Amir Aczel, Statystyka w Zarządzaniu, PWN, 2004.

Lesław Gajek, Marek Kałużka, *Wnioskowanie statystyczne dla studentów*, Wydawnictwo Naukowo-Techniczne, Warszawa, 1998.

Jerzy Ombach, Marcin Mazur, *Rachunek Prawdopodobieństwa i statystyka*, <http://wazniak.mimuw.edu.pl>, 2006.