**Ćwiczenie 5. Analiza regresji**

1. **Cel ćwiczenia:**

Celem ćwiczenia jest nabycie umiejętności poprawnego wyznaczenia równania regresji wielokrotnej do modelowania zależności liniowych dla więcej niż jednej zmiennej niezależnej, oraz poprawne zinterpretowanie uzyskanych wyników a także ocena zdolności prognostycznej uzyskanego modelu.

1. **Zagadnienia do samodzielnego opracowania:**

Pojęcie modelu matematycznego. Zasada działania metody regresji wielokrotnej. Istotność statystyczna modelu. Jakość dopasowania modelu. Zdolność prognostyczna modelu. Rachunek macierzowy (mnożenie macierzy, transponowanie macierzy, odwracanie macierzy).

1. **Przebieg ćwiczenia:**
2. Pobierz plik z danymi ze strony: [www.chem.univ.gda.pl/pchs/dane/cw\_5.ods](http://www.chem.univ.gda.pl/pchs/dane/cw_5.ods), następnie wyeksportuj dane do pliku .csv i zaimportuj w programie R, w pakiecie R Commander.
3. Wyznacz równanie modelu regresji wielokrotnej zmiennej zależnej (log Kow): 
4. Wytnij macierz A, która będzie zawierać zmienne niezależne (Dipole, SAS) z zakresu od 1 do 178 **A=Dataset[1:178,1:2]**.
5. Stwórz wektor jedynek **j <- matrix(1, nrow=178, ncol=1).**
6. Połącz wektor jedynek z macierzą zmiennych niezależnych poleceniem **X=cbind(j,A)**
7. Wyznacz parametry modelu regresji wielokrotnej 

Aby odwrócić macierz, należy zamienić notację macierzy z wyciętymi zmiennymi niezależnymi na notację wektorową komendą: **B=c(X)**, a następnie połączyć powstałe wektory zmiennych komendą **C=cbind(B$j,B$Dipole,B$SAS).**

Odwracanie macierzy: **solve(nazwa\_macierzy)**

1. Wytnij macierz Z, która będzie zawierać zmienne niezależne (Dipole, SAS) z zakresu od 179 do 237 **Z=Dataset[179:237,1:2]**
2. Oszacuj zmienną zależną modelem MLR (mnożenie wektorowe w programie R: „\*”)
3. Sprawdź jakość dopasowania modelu
* Wyznaczając współczynnik determinacji



* Wyznaczając błąd RMSEC (ang. *Root Mean Square Error of Calibration*).



* Rysując wykres rozrzutu  od 
1. **Sprawozdanie:**
	* + 1. Wyniki uzyskane na zajęciach wraz z interpretacją.
2. **Literatura**
	* + 1. A. Łomnicki, „Wprowadzenie do statystyki dla przyrodników”, Wydanie trzecie uzupełnione, Wyd. Naukowe PWN, Warszawa, 2005
			2. J. Mazerski, „Podstawy chemometrii”, Wydawnictwo Politechniki Gdańskiej, Gdańsk, 2000
			3. Tomasz Puzyn, Noriyuki Suzuki and Maciej Haranczyk, „*How Do the Partitioning Properties of Polyhalogenated POPs Change When Chlorine Is Replaced with Bromine?”, Environ. Sci. Technol.,* 2008, 42 (14), pp 5189–5195