

Pracownia Chemii Teoretycznej A/A' – objaśnienie do
ćwiczenia 7 (QSAR) – wersja dla gr. 6,10,11 (2014/15)
(poprawiona 04.01.2015)

W pliku, który Państwo zapamiętali jako Output (jest to plik tekstowy), powinny się znajdować następujące dane:

1. Model REG: szukamy przebiegu procedury REG, czyli

```
The REG Procedure
Model: MODEL1
Dependent Variable: pSw
```

i dalej

```
Stepwise Selection: Step 1
```

i dalej przewijamy, aż znajdziemy tekst podobny do:

```
All Variables Left in the Model Are Significant
```

- (a) następnie poszukujemy tabeli zatytułowanej:

```
Analysis of Variance
```

Znajdujemy w niej parametr F, p. kolumna F Value, oraz Pr, a także parametr Corrected Total i odchylenie standardowe modelu Root MSE (porównaj je z typową wielkością *pSw*!)

- (b) R^2 znajdujemy pod spodem, p. R-Square
- (c) W tabeli Parameter Estimates znajdujemy
 - i. jakie deskryptory zostały użyte w modelu REG – kolumna Variable, przy czym Intercept oznacza wyraz wolny,
 - ii. jakie są współczynniki regresji liniowej – kolumna Parameter Estimate,
 - iii. oraz jakie są odchylenia standardowe dla tych współczynników – kolumna Standard Error
 - iv. wielkości VIF, czyli Variance Inflation – ostatnia kolumna

- (d) Znajdujemy tabelę `Output Statistics`, a w niej przewidywaną wartość `pSw` dla badanej cząsteczki. Pod tabelą znajduje się dodatkowo

Predicted Residual SS (PRESS)

Używając `Corrected Total` oraz `PRESS`, obliczamy Q^2 .

2. Model PLS. Poniższe czynności wykonujemy dla każdego z czterech przebiegów modelu PLS (dla jednej, dwóch, trzech i czterech zmiennych `nfac`).
 - (a) Szukamy `The PLS Procedure`. to znajdziemy tabelę, w której ostatnia kolumna będzie nosić nazwę `pred_plsn`, gdzie $n = 1, 2, 3, 4$. W tej kolumnie znajdujemy wartość `pSw` dla wybranej cząsteczki.
 - (b) W sumie w `Output` powinny być cztery wywołania procedury PLS, rozróżnialne po `Number of Factors` (równemu 4 lub 3 lub 2 lub 1).

Model ma własności predykcyjne, jeśli

1. $R^2 > 0.8$
2. $Q^2 > 0.7$
3. F duże
4. $VIF < 10$

W przypadku każdego modelu (w sumie jest ich 5) trzeba porównać przewidywany przez model `pSw` dla cząsteczki z wartością doświadczalną.