

# Metody Statystyczne w Zarządzaniu Wierzytelnościami Masowymi

## Laboratorium 8

### Wymagane biblioteki

```
library(data.table)
library(car)
library(MASS)

load("KrukUWr2018.RData")
```

### Zadanie 1

Przygotuj rankę danych `cases_loanamount` bazując na tabeli `cases` tylko z przypadkami kredytów gotówkowych.

- Sprawdź miary pozycyjne wszystkich cech w utworzonej ramce.
- Sprawdź licznosc braków danych dla wszystkich cech.

### Zadanie 2

Interesującą nas cechą będzie `LoanAmount`, której wartość będziemy modelować w celu zastąpienia NA's.

- Metodą losowania z rozkładu zastąp braki danych w cesze `Land`.
- Metodą ekspercką zastąp braki danych w cechach: `Other`, `Gender`, `MeanSalary` i `GDPPerCapita`.
- Dokonaj dyskretyzacji cechy `DPD` a wartościom NA przypisz poziom `brak danych`, cechę zapisz jako `D_DPD`

### Zadanie 3

Na podstawie `cases_loanamount` przygotuj ramki danych:

- `cases_loanamount_nas`, która zawiera wszystkie przypadki brakujących wartości zmiennej `LoanAmount`.
- `cases_loanamount_wonas`, która zawiera 10000 kompletnych przypadków zmiennej `LoanAmount`.
- Podziel zbiór `cases_loanamount_wonas` na uczący i testowy w proporcji 0.7/0.3 poprzez utworzenie indeksów `ix_trn` i `ix_tst`.

### Zadanie 4

Zbadaj rozkład `cases_loanamount_wonas$LoanAmount`. Jeżeli jest taka potrzeba zaproponuj transformację.

### **Zadanie 5**

Zbuduj model regresji liniowej `m1` gdzie zmienną modelowaną jest `LoanAmount` a zmiennymi objaśniającymi : `TOA`, `Principal`, `Interest`, `Other`, `GDPPerCapita`, `MeanSalary`, `D_DPD`, `Age`, `Gender`

### **Zadanie 6**

Sprawdź rozkład składnika resztowego wyestymowanego modelu `m1`.

### **Zadanie 7**

Korzystając ze zbioru testowego dokonaj predykcji (wyniki zapisz w obiekcie `m1_pred`), a następnie oblicz bez używania gotowych funkcji: `RSS`, `RSE`, `TSS` i  $R^2$ .

### **Zadanie 8**

Dokonaj oceny jakości predykcji za pomocą znanych Ci miar .

### **Zadanie 9**

Sprawdź jak obserwacje odstające wpływają na współczynniki modelu oraz na ocenę jakości za pomocą wybranych przez Ciebie miar w zadaniu 9.

### **Zadanie 10**

Czy w modelu `m1` istnieją cechy, które charakteryzują się współliniowością? Jeżeli tak jak zmieni się model i ocena jakości po ich wyeliminowaniu.

### **Zadanie 11**

Zaproponuj własny model regresji liniowej korzystając z wiedzy zdobytej na wykładzie (transformacje zmiennych, interakcje między cechami, dyskretyzacja cech ciągłych, itp...), który będzie cechować się lepszymi parametrami oceny jakości predykcji od wyestymowanego na ćwiczeniach.