

### Regresja logistyczna. Problem 1.

Plik „dane\_cancer.txt” zawiera obserwacje dotyczące pacjentów oddziału onkologii pewnego szpitala. Zmienna binarna *Niepowodzenia* określa, czy doszło do wznowienia się (tzw. wznowy) choroby nowotworowej, zmienna *Nowotwor* podaje stopień zaawansowania nowotworu na początku leczenia, a zmienna *VEGF* oznacza poziom białek z rodziny VEGF. Wykonać następujące polecenia:

- a) wczytać i wyświetlić dane z pliku „dane\_cancer.csv” z pierwszym wierszem jako nagłówkiem,
- b) na podstawie obserwacji z obiektu dane skonstruować model regresji logistycznej o nazwie „mlogitNV”, postaci  $mlogitNV: Niepowodzenia \sim Nowotwor + \log(VEGF)$ ,
- c) na podstawie modelu „mlogitNV”, oszacować prawdopodobieństwa wznowy nowotworu dla pacjentów o wartościach cechy *Nowotwor*, odpowiednio: 1, 2, 3 i wartości cechy *VEGF* = 1000.
- d) podaj pseudo  $R^2$

e) oblicz OR dla pacjentów dla których wartość zmiennej  $\log(VEGF)$  wzrośnie o jedną jednostkę (względem grupy bez zmiany  $\log(VEGF)$ , pozostałe parametry bez zmian)