

# Domáca úloha 9 - Moderná aplikovaná regresia

Prosím odovzdať do najbližších cvičení

## 1 (60b)

`read.table("https://raw.githubusercontent.com/SamuelHudec/Applied-Regression/master/devil_data.csv")`  
Vezmite dáta, ktoré som pre vás zostavil ja, obsahujú relevantné, ale aj irrelevantné prediktory, niektoré sú transformované niektoré nie. Vašou úlohou je nájsť dáta generujúci proces. Ten komu sa to podarí dostane okrem bodov a večnej slávy aj sladkú odmenu. Svoj záver dobre zdôvodnite. Pamätajte, že metódy majú aj svoje slabiny.

## 2 (40b)

Majme dáta `seatpos` s `hipcenter` ako odozvou,

- (a) Fitnite plný lineárny model. Okomentujte efekt dĺžky nohy na odozvu.
- (b) Spočítajte 95% predikčný interval spoľahlivosti na budúcu hodnotu priemerného šoféra.
- (c) Teraz použite AIC na najdenia najlepšieho modelu. V tomto modeli okomentujte efekt dĺžky nohy na odozvu a spočítajte interval na budúcu hodnotu.
- (d) Porovnajte závery z oboch modelov a vysvetlite.

## 3 (Bonus)

Kullback-Leiblerová vzdialenosť je daná predpisom:

$$KL(g, f) = \int g \log(g/f) dy$$

kde  $g$  a  $f$  sú pravdepodobnostné distribúcie. Spočítajte KL vzdialenosť medzi  $N(\mu_1, \sigma_1^2)$  a  $N(\mu_2, \sigma_2^2)$ .