

# Domáca úloha 2 - Moderná aplikovaná regresia

Prosím odovzdať do cvičení.

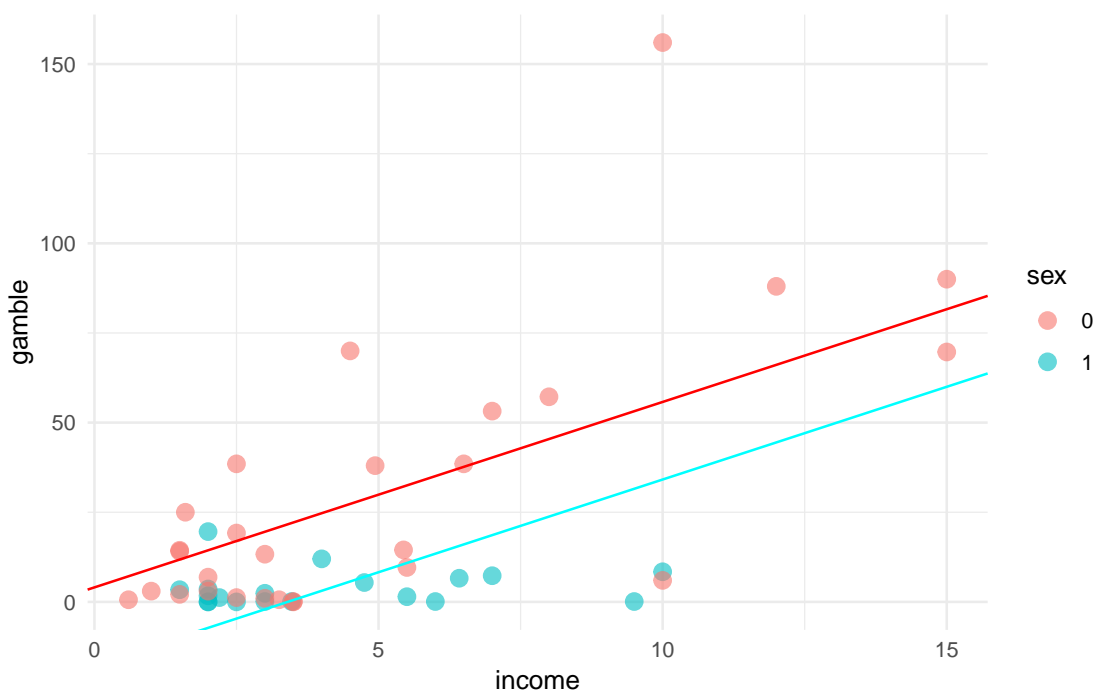
## 1 (70b)

Vráťme sa k dátam `teengamb` z knižnice `faraway` o gemblovaní tínedžerov v Británii. Pomocou lineárnej regresie vysvetlite výdavky na gemblovanie pomocou pohlavia, socioekonomického statusu, príjmu a výsledku verbálneho testu.

- Koľko percent variácie výdavkov na gemblovanie je vysvetlených prediktormi?
- Ktoré pozorovanie má najväčšiu hodnotu reziduálu ( $\hat{\epsilon}_i$ )?
- Vypočítajte strednú hodnotu a medián reziduálov. Zakreslite QQ-plot, povedali by ste o reziduáloch, že pochádzajú z normálneho rozdelenia?
- Vypočítajte koreláciu reziduálov s fitovanými hodnotami. Vysvetlite.
- Vypočítajte koreláciu výdavkov s príjmom a vysvetlite.
- Ak by sme všetky prediktory zafixovali. Aký rozdiel vo výdavkoch na gemblovanie predikuje model medzi chlapcom a dievčaťom?

## 2 (30b)

Uvažujme jednoduchší model ako v prvej úlohe, pričom gamble ostane ako odozva, ale voči prediktorm pohlavie a príjem. Do obrázku gamble vs. income vkreslite dve regresné priamky zvlášť pre mužov a zvlášť pre ženy. Použite na to odhady z výstupu zo summary modelu.



## 3 (Bonus)

Ukážte že  $\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$  teda  $TSS = RSS + ESS$ .