

Domáca úloha 4 - Moderná aplikovaná regresia

Prosím odovzdať do najbližších cvičení

1 (20b)

Vezmite `prostate` data, fitnite plný model s `lpsa` ako odozvou.

- (a) Majme nového pacienta menom *Karol*, namerali ste mu `lcavol=1.44692`, `lweight=3.62301`, `age=65`, `lbph=0.3001`, `svi=0`, `lcp=-0.79851`, `gleason=7`, `pgg45=15`. Použite model na predikciu Karolovej `lpsa` a konfidenčného intervalu (95%).
- (b) Zopakujte (a) s výnimkou veku a ten dajte na 20. Vysvetlite, prečo tento nový konfidenčný interval je širší.
- (c) Fitnite submodel na základe signifikancie prediktorov, teda vyhodte tie ktoré sú signifikantné na úrovni 5%. Teraz zopakujte (a) a (b), sú tieto nové konfidenčné intervaly širšie alebo užšie, ktoré z týchto predikcií by ste uprednostnili? A prečo?

2 (60b)

Majme dáta `happy`, ako prvé si urobte pre seba počítačnú hrubú obhliadku dát spojenú s naštudovaním problému.

- (a) Fitnite regresný model s `happy` ako odozvou. Čo nám hovorí koeficient `love`?
- (b) `love` prediktor nadobúda tri hodnoty, ale najčastejšie sa vyskytujú 2 a 3. Vytvorte novú premennú `clove` ktorá bude nadobúdať hodnoty 0 ak `love` bude 1 alebo 2. Tento nový prediktor zamente za pôvodný v regresnom modeli. Je táto zmena výrazná alebo nie? Prečo?
- (c) Fitnite model ktorý bude mať len `clove` ako prediktor. Interpretujte jeho koeficienty, ako sa dajú výsledky porovnať s modelom vyššie?
- (d) Vykreslite plot `happy` vs. `work`, zvýraznite hodnoty `clove` v plote tak ako na cvičeniach. (hint. použite funkciu `jitter` na lepšie zobrazenie množstva bodov na jednom mieste v plote).
- (e) Použite `xtabs(~ clove+work, happy)` na zobrazenie križovej tabuľky. Ak by sme chceli hľadať páry na základe `clove` s rovnakým statusom práce (prediktor `work`), aké je maximálne číslo zhôd, ktoré by sme mohli dosiahnuť?
- (f) Pre každý status práce spočítajte priemerný rozdiel (diferenciu) šťastia (`happy`) na základe `clove`. Spočítajte priemer týchto rozdielov, ktorý z koeficientou spočítaných predtým môže byť vhodný ekvivalent na tento priemer?

3 (20b)

Znova vezmite dáta z cvičenia 2. a sami vytvorte obrázok najlepšie zachytávajúci (zobrazujúci) problematiku, do ktorého zakreslite vaše regresne priamky (priamky). V kreativite sa medze nekladú, ale zároveň obrázok nemôže byť prešperkovaný.

4 (Bonus)

Ukážte že v jednoduchom gausovskom modeli $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ pre $i = 1, 2, \dots$ odhady $\hat{\beta}_0$ a $\hat{\beta}_1$ sú nezávislé ak $\bar{x} = 0$.