

Domáca úloha 6 - Moderná aplikovaná regresia

Prosím odovzdať do najbližších cvičení

1 (40b)

Vezmime dáta, ktoré už poznáme `cheddar`, fitnite plný regresný model s `taste` akou odozvou.

- (a) Je prediktor `Lactic` štatisticky významný v tomto modeli? Vypíšte p-hodnotu zo `summary` modelu.
- (b) Pridajte (zašumte) k prediktoru normálne rozdelené chyby so strednou rovnou nula a štandardnou odchylkou 0.01 a znova fitnite ten istý model. Ako sa zmenila p-hodnota prediktora `Lactic`?
- (c) Zopakujte tie isté výpočty ako v (b), ale 1000 krát (použite `for` cyklus) zakaždým uložte do vektora p-hodnotu (teda budete mať vektor p-hodnôt dĺžky 1000) a vezmite priemer z p-hodnôt. Mali tieto simulácie kvalitatívnejšiu výpovednú hodnotu p-hodnoty prediktora ako v prípade (b) ?
- (d) Zopakujte tie isté výpočty ako v (c), ale smerodajná odchylka chýb nech je 0.1. Zmenilo sa niečo na výsledku? vysvetlite.

2 (60b)

Majme dáta `longley`

- (a) Opíšte aký problém budete modelovať.
- (b) Uropte obhliadku dát spojenú s vizualizáciou.
- (c) Spočítajte čísla podmienenosti.
- (d) Spočítajte koreláciu medzi prediktormi.
- (e) Spočítajte variance inflation factors (VIF).
- (f) Fitnite plný regresný model s `Employed` akou odozvou. Diagnostikujte tento model.

Každý krok poriadne okomentujte. Napíšte krátky záver vašich zistení.

3 (Bonus)

Nech náhodná premenná pochádza z množiny pravdepodobnostných distribúcií, pre ktoré platí, že $Var(y)$ je proporciálna $(E(y))^q$. Navrhnete transformáciu $h(y)$ tak, aby variancia transformovanej premennej $Var(h(y))$ nezávisela od $E(y)$.