

Domáca úloha 3 - Moderná aplikovaná regresia

Prosím odovzdať do najbližších cvičení

1 (20b)

Uvažujme jednoduchý gausovský lineárny regresný model $y_i = a + bx_i + \epsilon$, kde $i = 1, 2, \dots, n$ a $\epsilon_i \sim N(0, \sigma^2)$. Odvoďte, čomu sa rovná odhad najmenších štvorcov pre $\hat{a}_{ols} = ?$ a $\hat{b}_{ols} = ?$

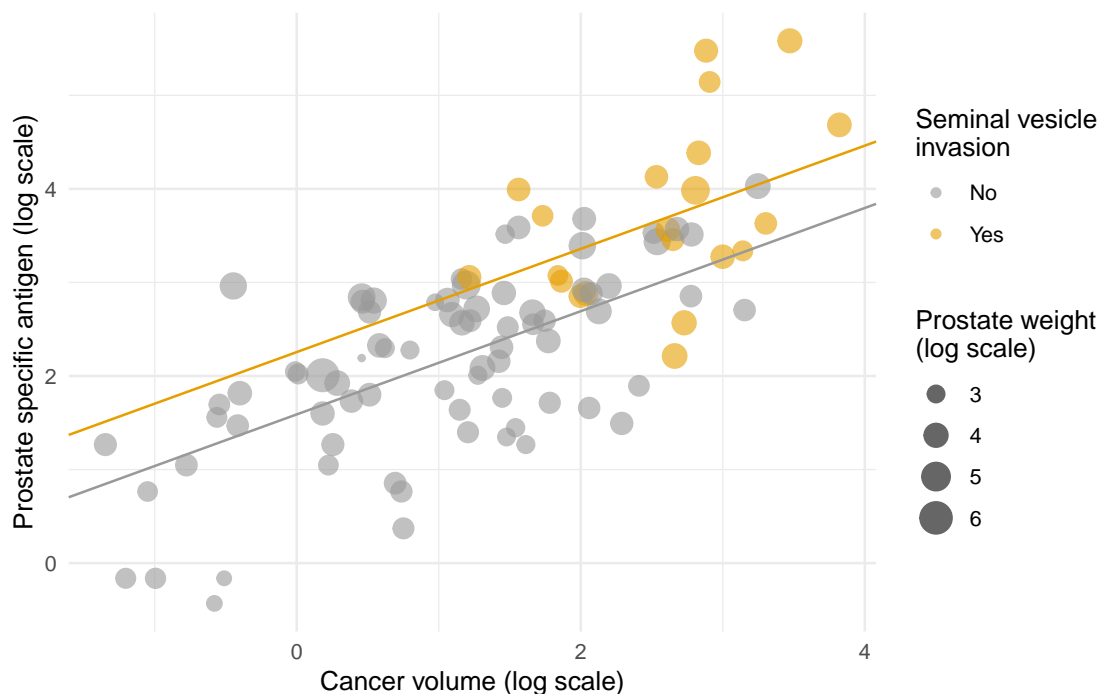
2 (60b)

Vezmite dáta `prostate` z knižnice `faraway` a fitnite model s `lpsa` ako odozvou.

- Spočítajte 90 a 95% konfidenčný interval parametra `age` len za pomoci výstupu zo `summary()`. Je podľa vás prediktor štatisticky významný?
- Napíšte funkciu, ktorá spočíta permutačný test F štatistiky a odskúšajte ju na modeli. Výsledky porovnajete s výstupom F štatistiky, ktorý nájdete v `summary()` (hint. Faraway str.40).
- Vyhodte všetky prediktory, ktoré nie sú signifikantné na 5% úrovni. Otestujte tento nový model voči modelu so všetkými prediktormi. Ktorý z týchto by ste zvolili?

3 (20b)

Na základe výstupu modelu z 2(c) zreplikujte nasledujúci obrázok `lpsa` vs. `lcavol` a zakreslite do neho regresné priamky. Nakoniec sa pokúste vysvetliť, prečo priamky z modelu neprechádzajú priemerom, ale majú výrazne vykrivenú smernicu.



4 (Bonus)

Ukážte, že priamka odhadnutá metódou najmenších štvorcov $y = \hat{\beta}_0 + \hat{\beta}_1 x$ prechádza bodom (\bar{x}, \bar{y}) .