

Tema 8 - laborator 10

Regresie liniară multiplă

Ce factori determinau prețul unui PC în vremea apariției acestora? Fișierul Prices.csv descrie un eșantion de 500 de vânzări de PC, colectat din 1993 până în 1995 în Statele Unite. Pe lângă prețul de vânzare, au fost colectate informații despre frecvența procesorului în MHz, dimensiunea hard diskului în MB, dimensiunea Ram în MB și dacă producătorul este premium (de exemplu, IBM, COMPAQ). Să presupunem că se ia în considerare modelul de regresie $y \sim \mathcal{N}(\mu, \sigma)$, unde

$$\mu = \alpha + \beta_1 x_1 + \beta_2 x_2,$$

y este prețul de vânzare, x_1 este frecvența procesorului și x_2 este logaritmul (natural) al mărimii hard diskului.

1. Folosind distribuții a priori slab informative asupra parametrilor α , β_1 , β_2 și σ , folosiți PyMC pentru a simula un eșantion suficient de mare din distribuția a posteriori.

2. Obțineți estimări de 95% HDI ale parametrilor β_1 și β_2 .

3. Pe baza rezultatelor obținute, sunt frecvența procesorului și mărimea hard diskului predictorii utili ai prețului de vânzare?

4. Să presupunem acum că un consumator este interesat de un computer cu o frecvență de 33 MHz și un hard disk de 540 MB. Simulați 5000 de extrageri din prețul de vânzare așteptat (μ) și construiți un interval de 90% HDI pentru acest preț.

5. În schimb, să presupunem că acest consumator dorește să prezică prețul de vânzare al unui computer cu această frecvență și mărime a hard disk-ului. Simulați 5000 de extrageri din distribuția predictivă posterioară și utilizați aceste extrageri simulate pentru a găsi un interval de predicție de 90% HDI.

6. Afectează în vreun fel prețul faptul că producătorul este premium? Justificați.