

## Tema 7 - laborator 9

### Regresie liniară

Ex. 1. Preziceți consumul unei mașini (în mile/galon: *mpg*) pe baza cailor putere (*CP*). Fișierul *auto-mpg.csv* conține un set de date cu observații, din care ne interesează doar cele corespunzătoare cailor putere și valorile *mpg* corespunzătoare.

- Încărcați setul de date într-un Pandas DataFrame și trasați un grafic (după eventuala curățare a datelor) pentru a vizualiza relația de dependență dintre cele două variabile: *CP* și *mpg*.
- Definiți modelul în PyMC folosind *CP* ca variabilă independentă și *mpg* ca variabilă dependentă.
- Determinați care este dreapta de regresie care se potrivește cel mai bine datelor.
- Adăugați graficului de la punctul a) regiunea 95% HDI pentru distribuția predictivă a posteriori. Ce concluzie puteți trage asupra modelului?

Ex. 2. O companie dorește să analizeze relația dintre cheltuielile pentru publicitate și vânzările unui produs. Ei au colectat date despre cheltuielile săptămânale în publicitate (în mii de dolari) și veniturile săptămânale din vânzări (în mii de dolari) pentru ultimele 20 de săptămâni.

Ipoteza companiei este că există o relație liniară între cheltuieli și venituri, dar se așteaptă ca această relație să fie influențată de incertitudini sau zgomot în date (de exemplu, alte influențe externe care afectează vânzările). Compania vrea să construiască un model bayesian de regresie liniară care:

- să estimeze coeficienții regresiei (interceptul și panta);
- să ofere intervale de credibilitate (HDI) pentru acești coeficienți;
- să prezică veniturile viitoare pentru noi niveluri ale cheltuielilor de publicitate.

Datele colectate pentru cele două atribute sunt redate în tabelul următor:

publicitate:	1.5	2.0	2.3	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	10.5	11.0
vânzare:	5.2	6.8	7.5	8.0	9.0	10.2	11.5	12.0	13.5	14.0	15.0	15.5	16.2	17.0	18.0	18.5	19.5	20.0	21.0	22.0

Folosind eventual modelul de la exercițiul precedent, rezolvați problemele puse mai sus.