

## Cerințe de proiect

1. Folosind setul de date **X** efectuați operații de statistică descriptivă pentru variabilele din acest set de date (medie, varianța, quartile, boxplot, interpretări).
2. Folosind setul de date **X** construiți două modele de regresie (o regresie simplă și una multiplă) alegând după cum considerați potrivite variabila răspuns și respectiv variabilele predictor.

Adăugați la setul de date inițial una sau mai multe variabile pe care să le considerați potrivite a fi incluse în cel puțin un model de regresie. **Generați** datele aferente variabilei nou adăugate conform unei repartiții potrivite (folosiți funcțiile din **R** care încep cu r: ex. pentru repartiția normală **rnorm**).

**Justificați** alegerile făcute și interpretați rezultatele obținute în urma evaluării celor două modele de regresie. Care din cele două modele construite considerați că este mai potrivit pentru setul vostru de date? Dați cel puțin două argumente pentru alegerea făcută.

3. **Alegeți** o repartiție **diferită** de cele studiate la laborator sau la cursul de Probabilități și Statistică și construiți în două reprezentări alăturate funcția de masă/densitatea de probabilitate (după cum e o repartiție a unei variabile aleatoare discretă sau continuă) și respectiv funcția de repartiție. Indicați proprietățile pe care le identificați la cele două funcții și precizați la ce este folosită repartiția respectivă în practică (adică ce fel de fenomene poate modela).

### BONUS:

1. Fie două variabile aleatoare discrete **X** și **Y** cu repartițiile:

$$X : \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix} \text{ și respectiv } Y : \begin{pmatrix} y_1 & y_2 & \dots & y_m \\ q_1 & q_2 & \dots & q_m \end{pmatrix}$$

- a) Construiți o funcție **frepcomgen** care primește ca parametri **m** și **n** și care generează un tabel cu repartiția comună a v.a. **X** și **Y** incompletă, dar într-o formă în care poate fi completată ulterior.

**Observație:** Se cere la a) să generați valorile lui **X**, valorile lui **Y** și suficient de multe valori pentru  $p_i$ ,  $q_j$  și respectiv  $\pi_{ij}$  astfel încât să poată fi determinată repartiția comună a celor două v.a.

**Nota:** În construirea algoritmului puteți începe de la cazul particular  $m=3$  și  $n=2$ . Dacă reușiți să oferiți soluția doar pentru acest caz particular, dar nu și pentru cazul general veți primi punctaj parțial.

- b) Construiți o funcție ***f<sub>complepcom</sub>*** care completează repartiția comună generată la punctul anterior (pentru cazul particular sau pentru cazul general).  
**Nota:** În cazul în care nu știți să rezolvați punctul a) puteți construi o funcție care să determine repartiția comună pornind de la un exemplu discutat la seminar.
- c) Având la dispoziție repartiția comună a v.a.  $X$  și  $Y$  de la punctul b) calculați:
- 1)  $\text{Cov}(5X, -3Y)$
  - 2)  $P(0 < X < 3/Y > 2)$
  - 3)  $P(X > 6, Y < 7)$
- d) Pentru exemplul obținut la punctul b) construiți două funcții ***f<sub>verind</sub>*** și respectiv ***f<sub>vernecor</sub>*** cu ajutorul cărora să verificați dacă variabilele  $X$  și  $Y$  sunt:
- 1) independente
  - 2) necorelate

### **Observații:**

- 1) Fiecare echipă identifică setul de date care i-a fost asignat după documentul excel trimis sefului de grupa.
- 2) Trimiteți pe adresa [simona.cojocea@live.com](mailto:simona.cojocea@live.com) o singură arhivă care să conțină: scripturile .R ***comentate*** cu rezolvările cerințelor de proiect și un document .docx sau .pdf care să explice pe larg cum au fost soluționate cerințele împreună cu justificările și interpretările aferente. Documentul trebuie să conțină și numele fiecărui membru al echipei și grupa din care face parte.
- 3) Punctajul este următorul:
  1. 5p    2. 10p (câte 2.5p pentru fiecare model de regresie și 5p pentru generarea datelor, interpretări și justificări)    3. 5 p
  2. **BONUS:** 10 p

**Succes!**