Universitatea din Bucureşti
Facultatea de Matematică şi Informatică

Nume și prenume: ____________________ Nota: __________

Examen

Timp de lucru 2h30. Toate documentele scrise și/sau calculatoarele electronice de mână sunt autorizate. Computerele personale, telefoanele mobile/smartwatch-urile precum și orice modalitate de comunicare între voi sunt **strict interzise**. Mult succes !

Exercițiul 1

20p

Să presupunem că observațiile (x_i, Y_i), $i = 1, \dots, n$ sunt făcute după modelul $Y_i = \alpha + \beta x_i + \varepsilon_i$, unde $x_1, \dots, x_n$ sunt constante iar erorile ε_i sunt variabile aleatoare centrate, necorelate și de varianță σ^2.

1. Presupunem că β este cunoscut iar α este necunoscut.

a) Determinați estimatorul $\tilde{\alpha}$ al lui α obținut prin metoda celor mai mici pătrate.

b) Calculați varianța lui $\tilde{\alpha}$ și arătați că este mai mică decât cea a lui $\hat{\alpha}$ (estimatorul lui α obținut prin metoda celor mai mici pătrate atunci când și α și β sunt considerate necunoscute).

c) Repetați punctele a) și b) de mai sus pentru situația în care α este cunoscut și β este necunoscut.

2. Presupunem că $\varepsilon_1, \dots, \varepsilon_n$ sunt i.i.d. repartizate $\mathcal{N}(0, \sigma^2)$ și că modelul este apoi reparametrizat astfel

$$Y_i = \alpha' + \beta'(x_i - \bar{x}) + \varepsilon_i.$$

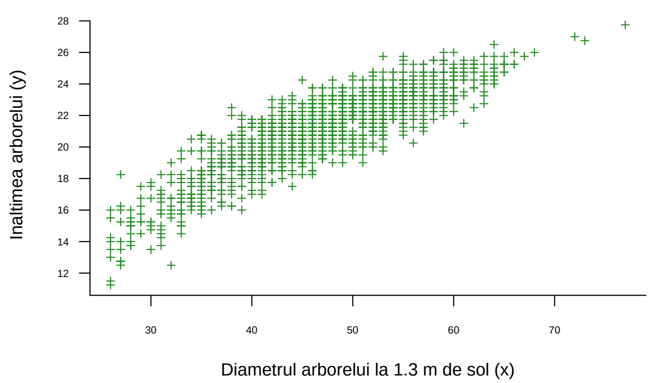
a) Arătați că $\hat{\beta}' = \hat{\beta}$ și că $\hat{\alpha}' \neq \hat{\alpha}$ unde $\hat{\alpha}$ și $\hat{\beta}$ estimatorii de verosimilate maximă a lui α și β, iar $\hat{\alpha}'$ și $\hat{\beta}'$ estimatorii de verosimilitate maximă a lui α' și β'.

b) Arătați că $\hat{\alpha}'$ și $\hat{\beta}'$ sunt necorelate, prin urmare sub ipoteza de normalitate sunt independente.

Exercițiul 2

30p

Dorim să explicăm înălțimea unui arbore y (măsurată în metrii) în funcție de diametrul său x (măsurat în centimetrii) la 1.3 m de sol și de radicalul acestuia. Dispunem de $n = 1429$ de perechi (x_i, y_i) de măsurători (a se vedea figura de mai jos).



Considerăm modelul de regresie următor:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 \sqrt{x_i} + \varepsilon_i, \quad 1 \leq i \leq n$$

unde ε_i sunt variabile aleatoare independente, repartizate normal de medie 0 și dispersie σ^2. Considerând

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 & \sqrt{x_1} \\ \vdots & \vdots & \vdots \\ 1 & x_n & \sqrt{x_n} \end{pmatrix} \quad \text{și} \quad \boldsymbol{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

am observat

$$\boldsymbol{X}^\intercal \boldsymbol{X} = \begin{pmatrix} ? & ? & 9791.6 \\ ? & 3306476 & ? \\ ? & 471237.9 & 67660 \end{pmatrix}, \quad \boldsymbol{X}^\intercal \boldsymbol{Y} = \begin{pmatrix} 30312.5 \\ 1461695.8 \\ 209685.6 \end{pmatrix}, \quad \boldsymbol{Y}^\intercal \boldsymbol{Y} = 651857.9.$$

1. Determinați valorile necunoscute, ?, din matricea $\boldsymbol{X}^\intercal \boldsymbol{X}$.
2. Care este valoarea diametrului mediu $\bar{x}$ și care este înălțimea medie $\bar{y}$?
3. În urma calculului obținem (cu aproximație)

$$(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} = \begin{pmatrix} 5.295 & 0.116 & -1.577 \\ 0.116 & 0.002 & -0.035 \\ -1.577 & -0.035 & 0.471 \end{pmatrix}.$$

Calculați estimatorii $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ prin metoda celor mai mici pătrate și reprezentați pe graficul de mai sus curba de regresie.

4. Calculați estimatorul lui σ^2 și construiți un interval de încredere de nivel de încredere de 95% pentru β_2.
5. Testați ipoteza $\beta_1 = 0$ la un nivel de semnificație de 10%.
6. Construiți cîte un interval de predicție de nivel 95% pentru y_{n+1} știind că $x_{n+1} = 49$ și respectiv $x_{n+1} = 25$. Care dintre acestea este mai mare ? Ne așteptam la așa ceva ?