

Exerciții de seminar 2

Regresie

Obiectivul acestui seminar este de a prezenta câteva exerciții de regresie liniară.

1 Regresie liniară simplă

1.1 Metoda celor mai mici pătrate

Ex. 1.1



Arătați că estimatorii $\hat{\beta}_0$ și $\hat{\beta}_1$ obținuți prin metoda celor mai mici pătrate, adică valorile coeficienților β_0 și β_1 care minimizează funcția

$$S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

sunt dați de expresiile

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad \text{și} \quad \hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Trebuie să determinăm

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg \min_{(\beta_0, \beta_1) \in \mathbb{R} \times \mathbb{R}} S(\beta_0, \beta_1)$$

și observând că funcția $S(\beta_0, \beta_1)$ este convexă ea admite un punct de minim. Acesta se obține ca soluție a sistemului $\nabla S = 0$ de ecuații normale,

$$\begin{cases} \frac{\partial S}{\partial \beta_0} = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0 \\ \frac{\partial S}{\partial \beta_1} = -2 \sum_{i=1}^n x_i (y_i - \beta_0 - \beta_1 x_i) = 0 \end{cases}$$

Din prima ecuație obținem prin sumare $n\beta_0 + \beta_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$ ceea ce conduce la $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$.

A doua ecuație conduce la

$$\beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i$$

și înlocuind β_0 cu expresia obținută anterior, obținem soluția

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i (y_i - \bar{y})}{\sum_{i=1}^n x_i (x_i - \bar{x})} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

De asemenea, se poate verifica că $S(\beta_0, \beta_1)$ se scrie sub forma

$$S(\beta_0, \beta_1) = n [\beta_0 - (\bar{y} - \beta_1 \bar{x})]^2 + \left[\sum_{i=1}^n (x_i - \bar{x})^2 \right] \left[\beta_1 - \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]^2 + \left[\sum_{i=1}^n (y_i - \bar{y})^2 \right] \left[1 - \frac{(\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}))^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2} \right]$$

care justifică în egală măsură soluția obținută anterior.

Ex. 1.2



Arătați că estimatorii obținuți prin metoda celor mai mici pătrate, $\hat{\beta}_0$ și $\hat{\beta}_1$, sunt estimatori nedeplasați.

Coeficienții $\hat{\beta}_0$ și $\hat{\beta}_1$ obținuți prin metoda celor mai mici pătrate sunt dați de $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ și $\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$ (aceștia sunt variabile aleatoare deoarece sunt funcții de Y_i care sunt variabile aleatoare). Înlocuind în expresia lui $\hat{\beta}_1$ pe y_i cu $\beta_0 + \beta_1 x_i + \varepsilon_i$ avem

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i + \varepsilon_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \frac{\underbrace{\sum_{i=1}^n (x_i - \bar{x})\beta_0}_{=0} + \beta_1 \sum_{i=1}^n (x_i - \bar{x})x_i + \sum_{i=1}^n (x_i - \bar{x})\varepsilon_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\beta_1 \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (x_i - \bar{x})\varepsilon_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})\varepsilon_i}{\sum_{i=1}^n (x_i - \bar{x})^2}. \end{aligned}$$

Conform ipotezei modelului de regresie liniară simplă, $\mathbb{E}[\varepsilon_i] = 0$, prin urmare $\mathbb{E}[\hat{\beta}_1] = \beta_1$ ceea ce arată că $\hat{\beta}_1$ este un estimator nedeplasat pentru β_1 .

În mod similar,

$$\mathbb{E}[\hat{\beta}_0] = \mathbb{E}[\bar{y} - \bar{x}\hat{\beta}_1] = \beta_0 + \bar{x}\beta_1 - \bar{x}\beta_1 = \beta_0$$

ceea ce arată că $\hat{\beta}_0$ este un estimator nedeplasat pentru β_0 .

Ex. 1.3



Calculați matricea de varianță-covarianță a estimatorilor $\hat{\beta}_0$ și $\hat{\beta}_1$.

Notăm cu $W = \begin{pmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_1) \end{pmatrix}$ matricea de varianță-covarianță a estimatorilor $\hat{\beta}_0$ și $\hat{\beta}_1$.

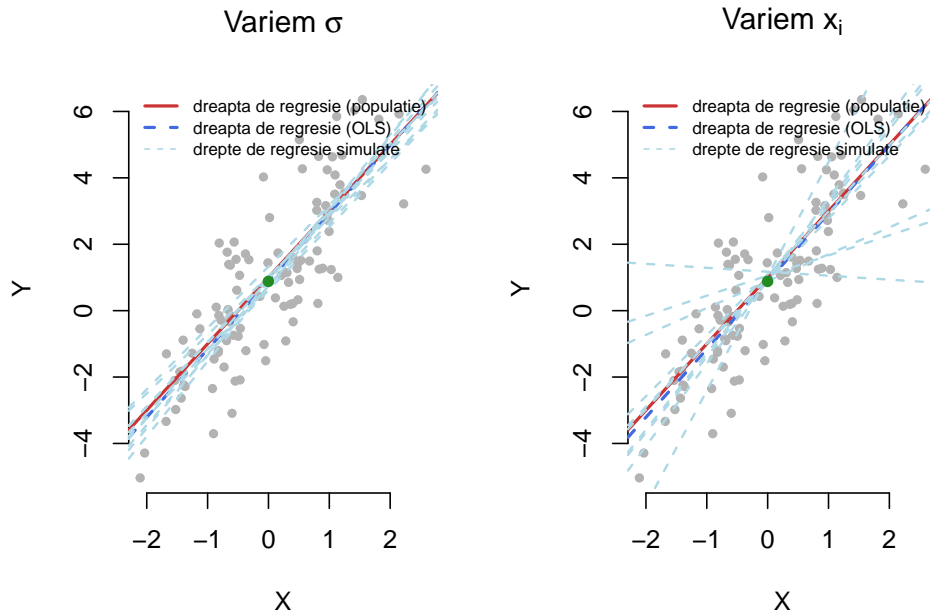
Avem, folosind expresia lui $\hat{\beta}_1$ determinată la punctul anterior și homoscedasticitatea și necorelarea erorilor $Cov(\varepsilon_i, \varepsilon_j) = \delta_{ij}\sigma^2$, că

$$\begin{aligned} Var(\hat{\beta}_1) &= Var\left(\beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})\varepsilon_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) = Var\left(\frac{\sum_{i=1}^n (x_i - \bar{x})\varepsilon_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) \\ &= \frac{Var(\sum_{i=1}^n (x_i - \bar{x})\varepsilon_i)}{[\sum_{i=1}^n (x_i - \bar{x})^2]^2} = \frac{\sum_{i,j} (x_i - \bar{x})(x_j - \bar{x})Cov(\varepsilon_i, \varepsilon_j)}{[\sum_{i=1}^n (x_i - \bar{x})^2]^2} \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2}{[\sum_{i=1}^n (x_i - \bar{x})^2]^2} = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}. \end{aligned}$$

Din expresia $Var(\hat{\beta}_1)$ observăm că dacă σ^2 este mică (cu alte cuvinte y_i sunt aproape de dreapta de regresie) atunci estimarea este mai precisă. De asemenea, se constată că pe măsură ce valorile x_i sunt mai dispersate în jurul valorii medii \bar{x} estimarea coeficientului $\hat{\beta}_1$ este mai precisă ($Var(\hat{\beta}_1)$ este mai mică). Acest fenomen se poate observa și în figura de mai jos în care am generat 100 de valori aleatoare X și 100 de valori pentru Y după modelul

$$y = 1 + 2x + \varepsilon$$

cu $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Dreapta roșie descrie adevărata relație $f(x) = 1 + 2x$ în populație iar dreapta albastră reprezintă dreapta de regresie calculată cu ajutorul metodei celor mai mici pătrate (OLS). Dreptele albastre deschise au fost generate tot cu ajutorul metodei celor mai mici pătrate atunci când variem σ^2 (în figura din stânga) și respectiv pe x_i în jurul lui \bar{x} (în figura din dreapta).



Pentru a determina $Var(\hat{\beta}_0)$, vom folosi relația $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ ceea ce conduce la

$$\begin{aligned} Var(\hat{\beta}_0) &= Var(\bar{y} - \hat{\beta}_1 \bar{x}) = Var(\bar{y}) - 2Cov(\bar{y}, \hat{\beta}_1 \bar{x}) + Var(\hat{\beta}_1 \bar{x}) \\ &= Var\left(\frac{1}{n} \sum_{i=1}^n y_i\right) - 2\bar{x}Cov(\bar{y}, \hat{\beta}_1) + \bar{x}^2 Var(\hat{\beta}_1) \\ &= \frac{\sigma^2}{n} + \bar{x}^2 \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} - 2\bar{x}Cov(\bar{y}, \hat{\beta}_1). \end{aligned}$$

Pentru $Cov(\bar{y}, \hat{\beta}_1)$ avem (ținând cont de faptul că β_0 , β_1 și x_i sunt constante)

$$\begin{aligned} Cov(\bar{y}, \hat{\beta}_1) &= Cov\left(\frac{1}{n} \sum_{i=1}^n y_i, \beta_1 + \frac{\sum_{j=1}^n (x_j - \bar{x}) \varepsilon_j}{\sum_{j=1}^n (x_j - \bar{x})^2}\right) = \frac{1}{n} \sum_{i=1}^n Cov\left(\beta_0 + \beta_1 x_i + \varepsilon_i, \beta_1 + \frac{\sum_{j=1}^n (x_j - \bar{x}) \varepsilon_j}{\sum_{j=1}^n (x_j - \bar{x})^2}\right) \\ &= \frac{1}{n} \sum_{i=1}^n Cov\left(\varepsilon_i, \frac{\sum_{j=1}^n (x_j - \bar{x}) \varepsilon_j}{\sum_{j=1}^n (x_j - \bar{x})^2}\right) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\sum_{j=1}^n (x_j - \bar{x})^2} Cov\left(\varepsilon_i, \sum_{j=1}^n (x_j - \bar{x}) \varepsilon_j\right) \\ &= \frac{1}{\sum_{j=1}^n (x_j - \bar{x})^2} \sum_{i=1}^n \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x}) Cov(\varepsilon_i, \varepsilon_j) = \frac{1}{\sum_{j=1}^n (x_j - \bar{x})^2} \sum_{i=1}^n \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x}) \delta_{ij} \sigma^2 \\ &= \frac{\sigma^2}{\sum_{j=1}^n (x_j - \bar{x})^2} \underbrace{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})}_{=0} = 0 \end{aligned}$$

prin urmare

$$Var(\hat{\beta}_0) = \frac{\sigma^2}{n} + \bar{x}^2 \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Calculul covarianței dintre $\hat{\beta}_0$ și $\hat{\beta}_1$ rezultă aplicând relațiile de mai sus

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = Cov(\bar{y} - \hat{\beta}_1 \bar{x}, \hat{\beta}_1) = Cov(\bar{y}, \hat{\beta}_1) - \bar{x} Var(\hat{\beta}_1) = -\frac{\sigma^2 \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Observăm că $Cov(\hat{\beta}_0, \hat{\beta}_1) \leq 0$ iar intuitiv, cum dreapta de regresie (bazată pe estimatorii obținuți prin metoda celor mai mici pătrate) $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$ trece prin centrul de greutate al datelor (\bar{x}, \bar{y}) , dacă presupunem $\bar{x} > 0$ remarcăm că atunci când creștem panta (creștem $\hat{\beta}_1$) ordonata la origine scade (scade $\hat{\beta}_0$) și reciproc.

Matricea de varianță-covarianță a estimatorilor $\hat{\beta}_0$ și $\hat{\beta}_1$ devine

$$W = \begin{pmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_1) \end{pmatrix} = \begin{pmatrix} \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} & -\frac{\sigma^2 \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ -\frac{\sigma^2 \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} & \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{pmatrix}.$$

Ex. 1.4



Arătați că în cadrul modelului de regresie liniară simplă, suma valorilor reziduale este nulă.

Observăm, folosind definiția $\hat{\varepsilon}_i = y_i - \hat{y}_i$, că

$$\begin{aligned} \sum_{i=1}^n \hat{\varepsilon}_i &= \sum_{i=1}^n (y_i - \hat{y}_i) = \sum_{i=1}^n (y_i - \hat{\beta}_0 - x_i \hat{\beta}_1) \\ &= \sum_{i=1}^n \left[y_i - \underbrace{(\bar{y} - \bar{x} \hat{\beta}_1)}_{=\hat{\beta}_0} - x_i \hat{\beta}_1 \right] = \sum_{i=1}^n (y_i - \bar{y}) - \hat{\beta}_1 \sum_{i=1}^n (x_i - \bar{x}) = 0 \end{aligned}$$

Ex. 1.5



Arătați că în modelul de regresie liniară simplă statistica $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{\varepsilon}_i^2$ este un estimator nedeplasat pentru σ^2 .

Ținând cont de faptul că $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ și $\bar{y} = \beta_0 + \beta_1 \bar{x} + \bar{\varepsilon}$ (prin însumarea după i a relațiilor $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$) găsim că

$$\begin{aligned}\hat{\varepsilon}_i &= y_i - \hat{y}_i = (\beta_0 + \beta_1 x_i + \varepsilon_i) - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \\ &= (\underbrace{\bar{y} - \beta_1 \bar{x} - \bar{\varepsilon}}_{=\beta_0} + \beta_1 x_i + \varepsilon_i) - (\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i) \\ &= (\beta_1 - \hat{\beta}_1)(x_i - \bar{x}) + (\varepsilon_i - \bar{\varepsilon})\end{aligned}$$

și prin dezvoltarea binomului și utilizând relația $\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})\varepsilon_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$ găsim

$$\begin{aligned}\sum_{i=1}^n \hat{\varepsilon}_i^2 &= (\beta_1 - \hat{\beta}_1)^2 \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 + 2(\beta_1 - \hat{\beta}_1) \sum_{i=1}^n (x_i - \bar{x})(\varepsilon_i - \bar{\varepsilon}) \\ &= (\beta_1 - \hat{\beta}_1)^2 \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 + 2(\beta_1 - \hat{\beta}_1) \sum_{i=1}^n (x_i - \bar{x})\varepsilon_i - 2(\beta_1 - \hat{\beta}_1)\bar{\varepsilon} \sum_{i=1}^n (x_i - \bar{x}) \\ &= (\beta_1 - \hat{\beta}_1)^2 \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 - 2(\beta_1 - \hat{\beta}_1)^2 \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 - (\beta_1 - \hat{\beta}_1)^2 \sum_{i=1}^n (x_i - \bar{x})^2.\end{aligned}$$

Luând media găsim că

$$\mathbb{E} \left(\sum_{i=1}^n \hat{\varepsilon}_i^2 \right) = \mathbb{E} \left(\sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 \right) - \sum_{i=1}^n (x_i - \bar{x})^2 \text{Var}(\hat{\beta}_1) = (n-1)\sigma^2 - \sigma^2 = (n-1)\sigma^2$$

unde am folosit că $\mathbb{E} \left(\frac{1}{n-1} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 \right) = \sigma^2$ (deoarece $\text{Var}(\varepsilon_i) = \sigma^2$).

Concluzionăm că $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{\varepsilon}_i^2$ este un estimator nedeplasat pentru σ^2 .

Ex. 1.6



Fie x_{n+1} o nouă valoare pentru variabila X și ne propunem să prezicem valoarea y_{n+1} conform modelului

$$y_{n+1} = \beta_0 + \beta_1 x_{n+1} + \varepsilon_{n+1}$$

cu $\mathbb{E}[\varepsilon_{n+1}] = 0$, $\text{Var}(\varepsilon_{n+1}) = \sigma^2$ și $\text{Cov}(\varepsilon_{n+1}, \varepsilon_i) = 0$ pentru $i = 1, \dots, n$.

Arătați că varianța răspunsului mediu prezis este

$$\text{Var}(\hat{y}_{n+1}) = \sigma^2 \left[\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]$$

iar varianța erorii de predicție $\hat{\varepsilon}_{n+1}$ satisface $\mathbb{E}[\hat{\varepsilon}_{n+1}] = 0$ și

$$Var(\hat{\varepsilon}_{n+1}) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right].$$

Cum $\hat{y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{n+1}$ avem

$$\begin{aligned} Var(\hat{y}_{n+1}) &= Var(\hat{\beta}_0 + \hat{\beta}_1 x_{n+1}) = Var(\hat{\beta}_0) + 2Cov(\hat{\beta}_0, \hat{\beta}_1) + x_{n+1}^2 Var(\hat{\beta}_1) \\ &= \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} - 2 \frac{\sigma^2 \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} + \frac{\sigma^2 x_{n+1}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \left[\frac{1}{n} \sum_{i=1}^n x_i^2 - 2x_{n+1} \bar{x} + x_{n+1}^2 \right] \\ &= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 + \bar{x}^2 - 2x_{n+1} \bar{x} + x_{n+1}^2 \right] \\ &= \sigma^2 \left[\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]. \end{aligned}$$

Constatăm că atunci când x_{n+1} este departe de valoarea medie \bar{x} răspunsul mediu are o variabilitate mai mare.

Pentru a obține varianța erorii de predicție $\hat{\varepsilon}_{n+1} = y_{n+1} - \hat{y}_{n+1}$ să observăm că y_{n+1} depinde doar de ε_{n+1} pe când \hat{y}_{n+1} depinde de ε_i , $i \in \{1, 2, \dots, n\}$. Din necorelarea erorilor deducem că

$$Var(\hat{\varepsilon}_{n+1}) = Var(y_{n+1} - \hat{y}_{n+1}) = Var(y_{n+1}) + Var(\hat{y}_{n+1}) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right].$$

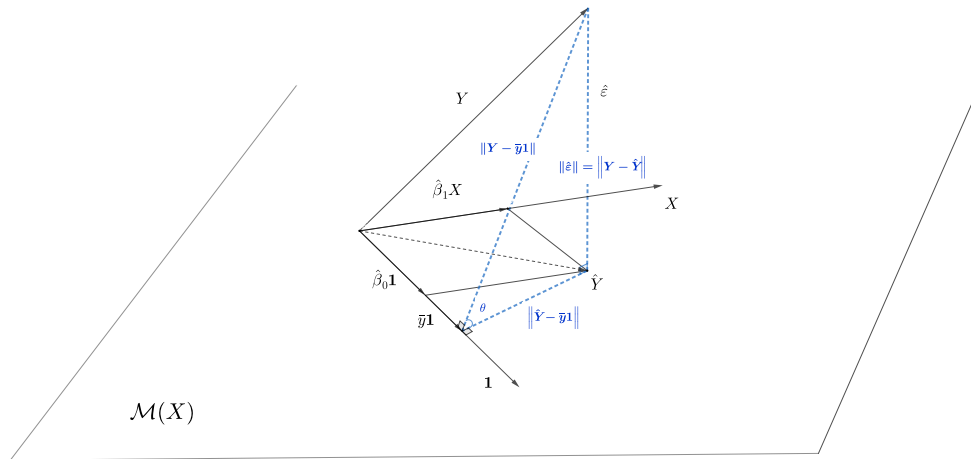
1.2 Coeficientul de determinare R^2 și coeficientul de corelație

În această secțiune încercăm să abordăm problema de regresie liniară simplă într-un context geometric. Din punct de vedere vectorial dispunem de doi vectori: vectorul $X = (x_1, x_2, \dots, x_n)^\top$ a celor n observații ale variabilei explicative și vectorul $Y = (y_1, y_2, \dots, y_n)^\top$ compus din cele n observații ale variabilei răspuns, pe care vrem să o explicăm. Cei doi vectori aparțin spațiului \mathbb{R}^n .

Fie $\mathbf{1} = (1, 1, \dots, 1)^\top \in \mathbb{R}^n$ și $\mathcal{M}(X)$ subspațiul liniar din \mathbb{R}^n de dimensiune 2 generat de vectorii $\{\mathbf{1}, X\}$ (acești vectori nu sunt coliniari deoarece X conține cel puțin două elemente distincte). Notăm cu \hat{Y} proiecția ortogonală a lui Y pe subspațiul $\mathcal{M}(X)$ și cum $\{\mathbf{1}, X\}$ formează o bază în $\mathcal{M}(X)$ deducem că există $\hat{\beta}_0, \hat{\beta}_1 \in \mathbb{R}$ astfel ca $\hat{Y} = \hat{\beta}_0 \mathbf{1} + \hat{\beta}_1 X$. Cum, din definiția proiecției ortogonale, \hat{Y} este unicul vector din $\mathcal{M}(X)$ care minimizează distanța euclidiană (deci și pătratul ei)

$$\|Y - \hat{Y}\| = \sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$$

deducem că $\hat{\beta}_0, \hat{\beta}_1$ coincid cu valorile obținute prin metoda celor mai mici pătrate. Astfel coeficienții $\hat{\beta}_0$ și $\hat{\beta}_1$ se reprezintă coordonatele proiecției ortogonale a lui Y pe subspațiul generat de vectorii $\{\mathbf{1}, X\}$ (a se vedea figura de mai jos).



Observăm că, în general, vectorii $\{\mathbf{1}, X\}$ nu formează o bază ortogonală în $\mathcal{M}(X)$ (cu excepția cazului în care $\langle \mathbf{1}, X \rangle = n\bar{x} = 0$) prin urmare $\hat{\beta}_0 \mathbf{1}$ nu este proiecția ortogonală a lui Y pe $\mathbf{1}$ (aceasta este $\frac{\langle Y, \mathbf{1} \rangle}{\|\mathbf{1}\|^2} \mathbf{1} = \bar{y} \mathbf{1}$) iar $\hat{\beta}_1 X$ nu este proiecția ortogonală a lui Y pe X (aceasta fiind $\frac{\langle Y, X \rangle}{\|X\|^2} X$).

Fie $\hat{\epsilon} = Y - \hat{Y} = (\hat{\epsilon}_1, \hat{\epsilon}_2, \dots, \hat{\epsilon}_n)^\top$ vectorul valorilor reziduale. Aplicând Teorema lui Pitagora (în triunghiul albastru) rezultă (descompunerea ANOVA pentru regresie) că

$$\begin{aligned} \|Y - \bar{y}\mathbf{1}\|^2 &= \|\hat{Y} - \bar{y}\mathbf{1}\|^2 + \left\| \underbrace{\hat{\epsilon}}_{Y - \hat{Y}} \right\|^2 \\ \sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (\underbrace{\hat{\epsilon}_i}_{y_i - \hat{y}_i})^2 \\ SS_T &= SS_{reg} + RSS \end{aligned}$$

Din definiția coeficientului de determinare R^2 avem că

$$R^2 = \frac{SS_{reg}}{SS_T} = \frac{\|\hat{Y} - \bar{y}\mathbf{1}\|^2}{\|Y - \bar{y}\mathbf{1}\|^2} = 1 - \frac{\|\hat{\epsilon}\|^2}{\|Y - \bar{y}\mathbf{1}\|^2}$$

și conform figurii de mai sus $R^2 = \cos^2(\theta)$. Prin urmare dacă $R^2 = 1$, atunci $\theta = 0$ și $Y \in \mathcal{M}(X)$, deci $y_i = \beta_0 + \beta_1 x_i$, $i \in \{1, 2, \dots, n\}$ (punctele eșantionului sunt perfect aliniate) iar dacă $R^2 = 0$, deducem că $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = 0$, deci $\hat{y}_i = \bar{y}$ (modelul liniar nu este adaptat în acest caz, nu putem explica mai bine decât media).

Ex. 1.7



Arătați că

$$R^2 = r_{xy}^2 = r_{y\hat{y}}^2$$

unde r_{xy} este coeficientul de corelație empiric dintre x și y .

Din definiția coeficientului de determinare și folosind coeficienții $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ și $\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$ obținuți prin metoda celor mai mici pătrate avem

$$\begin{aligned} R^2 &= \frac{\|\hat{Y} - \bar{y}\mathbf{1}\|^2}{\|Y - \bar{y}\mathbf{1}\|^2} = \frac{\sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\sum_{i=1}^n (\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\ &= \frac{\hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \left(\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)^2 \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\ &= \frac{[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})]^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2} = r_{xy}^2. \end{aligned}$$

Pentru a verifica a doua parte, $R^2 = r_{y\hat{y}}^2$, să observăm că

$$r_{y\hat{y}}^2 = \frac{[\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})]^2}{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2 \sum_{i=1}^n (y_i - \bar{y})^2}$$

iar $\bar{\hat{y}} = \frac{\sum_{i=1}^n \hat{y}_i}{n} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y}$, prin urmare

$$r_{y\hat{y}}^2 = \frac{[\sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \bar{y})]^2}{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \sum_{i=1}^n (y_i - \bar{y})^2}.$$

De asemenea

$$\begin{aligned} \sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \bar{y}) &= \sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \hat{y}_i + \hat{y}_i - \bar{y}) \\ &= \sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \end{aligned}$$

și cum

$$\begin{aligned} \sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) &= \sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y})(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \\ &= \sum_{i=1}^n (\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i - \bar{y})[(y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x})] \\ &= \hat{\beta}_1 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) - \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \underbrace{\frac{S_{xy}}{S_{xx}}}_{\hat{\beta}_1} S_{xy} - \frac{S_{xy}^2}{S_{xx}^2} S_{xx} = 0 \end{aligned}$$

deducem că $r_{y\hat{y}}^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = R^2$.

1.3 Aplicații numerice

Ex. 1.8



Tabelul de mai prezintă o serie de date privind greutatea taților și respectiv a fiului lor cel mare

Tata : 65 63 67 64 68 62 70 66 68 67 69 71
Fiu : 68 66 68 65 69 66 68 65 71 67 68 70

Obținem următoarele rezultate numerice

$$\sum_{i=1}^{12} t_i = 800 \quad \sum_{i=1}^{12} t_i^2 = 53418 \quad \sum_{i=1}^{12} t_i f_i = 54107 \quad \sum_{i=1}^{12} f_i = 811 \quad \sum_{i=1}^{12} f_i^2 = 54849.$$

1. Determinați dreapta obținută prin metoda celor mai mici pătrate a greutății fiilor în funcție de greutatea taților.
2. Determinați dreapta obținută prin metoda celor mai mici pătrate a greutății taților în funcție de greutatea fiilor.
3. Arătați că produsul pantelor celor două drepte este egal cu pătratul coeficientului de corelație empirică dintre t_i și f_i (sau coeficientul de determinare).

1. Dreapta de regresie a greutății fiilor în funcție de greutatea taților este $f = \hat{\alpha}_0 + \hat{\alpha}_1 t$ unde coeficienții sunt dați de

$$\hat{\alpha}_0 = \bar{f} - \hat{\alpha}_1 \bar{t}, \quad \hat{\alpha}_1 = \frac{\sum_{i=1}^{12} (t_i - \bar{t})(f_i - \bar{f})}{\sum_{i=1}^{12} (t_i - \bar{t})^2}$$

Pentru datele din problema noastră coeficienții sunt $\hat{\alpha}_0 = 35.8$ și $\hat{\alpha}_1 = 0.48$ iar dreapta de regresie este $f = 35.8 + 0.48t$ (a se vedea figura din stânga).

2. Dreapta de regresie a greutății taților în funcție de greutatea fiilor este $t = \hat{\beta}_0 + \hat{\beta}_1 f$ unde coeficienții sunt dați de

$$\hat{\beta}_0 = \bar{t} - \hat{\beta}_1 \bar{f}, \quad \hat{\beta}_1 = \frac{\sum_{i=1}^{12} (f_i - \bar{f})(t_i - \bar{t})}{\sum_{i=1}^{12} (f_i - \bar{f})^2}$$

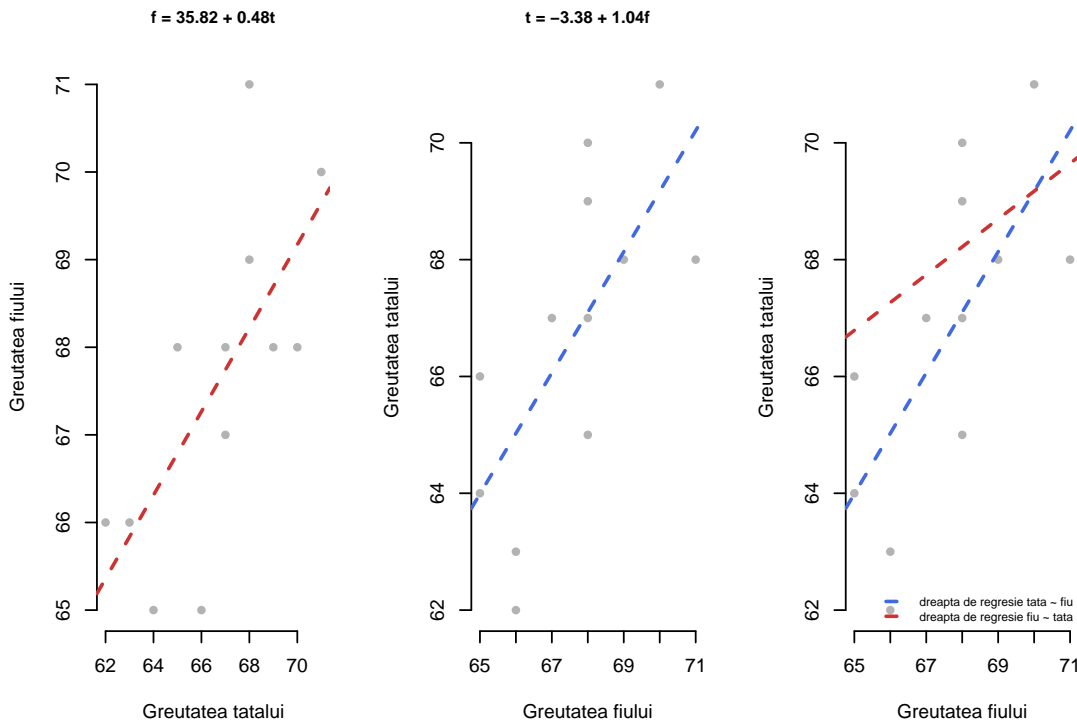
În cazul problemei, coeficienții sunt $\hat{\beta}_0 = -3.38$ și $\hat{\beta}_1 = 1.03$ iar dreapta de regresie este $t = -3.38 + 1.03f$ (a se vedea figura din mijloc).

3. Produsul pantelor celor două drepte este

$$\hat{\alpha}_1 \hat{\beta}_1 = \frac{\left[\sum_{i=1}^{12} (f_i - \bar{f})(t_i - \bar{t}) \right]^2}{\sum_{i=1}^{12} (f_i - \bar{f})^2 \sum_{i=1}^{12} (t_i - \bar{t})^2}$$

și conform [exercițiului 2](#) și a definiției coeficientului de determinare avem

$$\hat{\alpha}_1 \hat{\beta}_1 = r_{f,t}^2 = R^2.$$



Ex. 1.9



Dorim să exprimăm înălțimea y (măsurată în picioare) a unui arbore în funcție de diametrul său x (exprimat în centimetri) la înălțimea de 1m30 de la sol. Pentru aceasta dispunem de 20 de măsurători $(x_i, y_i) = (\text{diametru}, \text{înălțime})$ și în urma calculelor am obținut rezultatele următoare: $\bar{x} = 4.53$, $\bar{y} = 8.65$ și

$$\frac{1}{20} \sum_{i=1}^{20} (x_i - \bar{x})^2 = 10.97 \quad \frac{1}{20} \sum_{i=1}^{20} (y_i - \bar{y})^2 = 2.24 \quad \frac{1}{20} \sum_{i=1}^{20} (x_i - \bar{x})(y_i - \bar{y}) = 3.77.$$

1. Notăm cu $y = \hat{\beta}_0 + \hat{\beta}_1 x$ dreapta de regresie. Calculați coeficienții $\hat{\beta}_0$ și $\hat{\beta}_1$.
2. Dați și calculați o măsură care descrie calitatea concordanței datelor cu modelul propus.
3. Să presupunem că abaterile standard pentru estimatorii $\hat{\beta}_0$ și $\hat{\beta}_1$ sunt $\hat{\sigma}_0 = 1.62$ și respectiv $\hat{\sigma}_1 = 0.05$. Presupunem că erorile ε_i sunt variabile aleatoare independente repartizate normal de medie 0 și varianțe egale. Vrem să testăm ipotezele $H_0 : \beta_j = 0$ versus $H_1 : \beta_j \neq 0$ pentru $j = 0, 1$. De ce acest test este interesant în contextul problemei noastre?

1. Estimatorii coeficienților drepte de regresie $y = \hat{\beta}_0 + \hat{\beta}_1 x$ sunt dați de

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \approx 0.344$$

și respectiv

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \approx 7.09$$

2. Pentru a măsura calitatea concordanței datelor la modelul de regresie vom folosi coeficientul de determinare R^2 . Am văzut că acesta corespunde pătratului coeficientului de corelație empirică:

$$R^2 = r_{x,y}^2 = \left(\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \right)^2 \approx 0.58.$$

Observăm că modelul de regresie liniară simplă explică un pic mai mult de jumătate din variabilitatea datelor.

3. Sub ipoteza modelului condiționat normal (erorile ε_i sunt variabile aleatoare independente repartizate normal de medie 0 și varianțe egale) avem că $\hat{\beta}_j \sim \mathcal{N}(\beta_j, \sigma_{\hat{\beta}_j}^2)$ și înlocuind varianțele $\sigma_{\hat{\beta}_j}^2$ cu estimatorii

$$\hat{\sigma}_j^2, \text{ deducem că } \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_j} \sim t_{n-2}.$$

Prin urmare, sub H_0 avem că

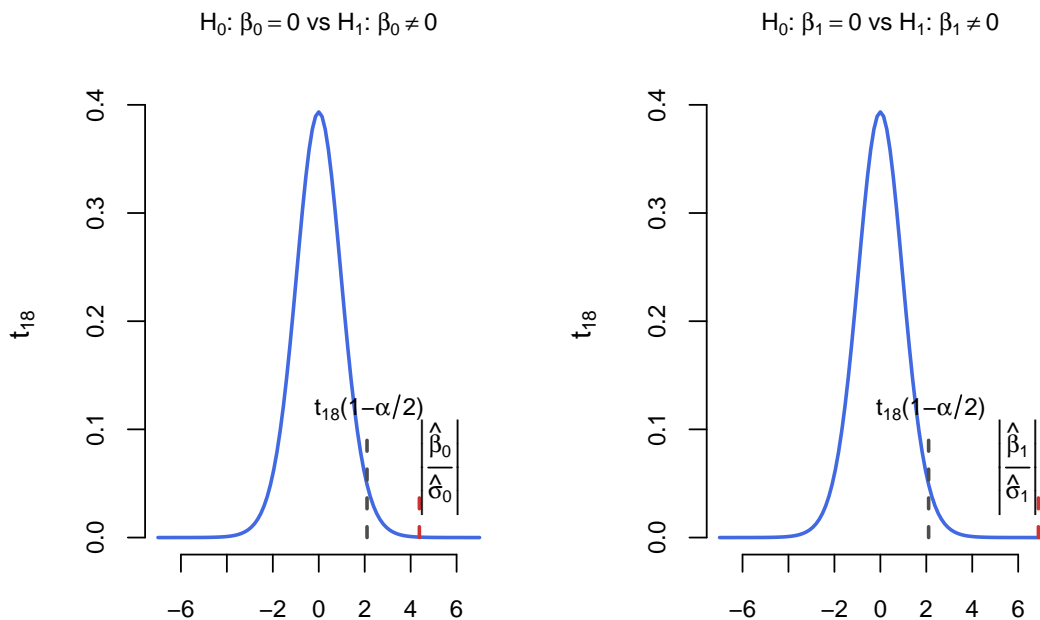
$$\frac{\hat{\beta}_0}{\hat{\sigma}_0} \sim t_{18},$$

iar pentru un prag de semnificație $1 - \alpha = 95\%$, ținând seama că $\left| \frac{\hat{\beta}_0}{\hat{\sigma}_0} \right| \approx 4.38$ și că $t_{18}(1 - \alpha/2) \approx 2.1$, concluzionăm că respingem ipoteza nulă.

În mod similar, pentru $\hat{\beta}_1$ găsim că

$$\left| \frac{\hat{\beta}_1}{\hat{\sigma}_1} \right| \approx 6.88 > 2.1$$

de unde respingem ipoteza nulă $H_0 : \beta_1 = 0$ în acest caz de asemenea.



2 Regresie liniară multiplă

Principiul problemei de regresie este de a modela o variabilă y , numită variabilă răspuns sau variabilă dependentă (cea pe care vrem să o explicăm), cu ajutorul unei funcții care depinde de un anumit număr de variabile $\mathbf{x} = (x_1, \dots, x_p)^\top$, numite variabile explicative sau predictor sau independente (vom folosi în general primele două denumiri)

$$y \approx g(\mathbf{x}) = g(x_1, \dots, x_p).$$

Astfel având dat un eșantion de talie n de $(p+1)$ -upluri (\mathbf{x}, y) , $n > p$, ne propunem să determinăm g . Aproximarea, \approx , din relația anterioară se poate descrie matematic sub forma

$$g = \arg \min_{f \in \mathcal{G}} \sum_{i=1}^n L(y_i - f(\mathbf{x}_i))$$

unde $L(\cdot)$ se numește funcție de cost sau de pierdere iar \mathcal{G} este o clasă de funcții dată.

În general funcția de cost poate lua multe forme, e.g. **Loss function**, dar de cele mai multe ori vom întâlni două: funcția de cost absolut $L(x) = |x|$ sau funcția de cost pătratic $L(x) = x^2$.

În ceea ce privește clasa de funcții \mathcal{G} vom considera funcțiile liniare

$$\mathcal{G} = \{f : \mathbb{R}^p \rightarrow \mathbb{R} \mid f(\mathbf{x}) = \begin{pmatrix} 1 & \mathbf{x} \end{pmatrix} \boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p\}.$$

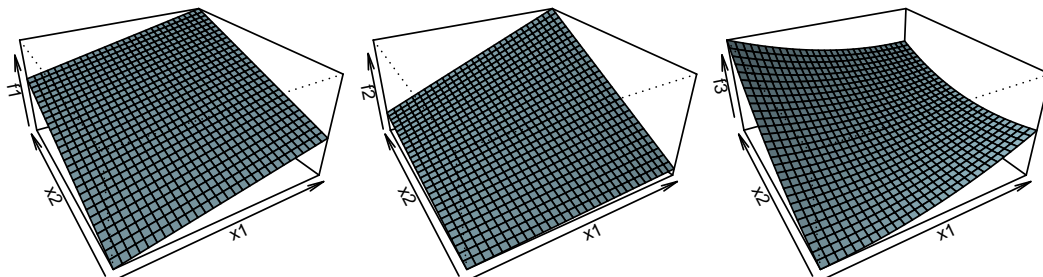
Atunci când vorbim de regresie liniară ne referim la liniaritatea în parametrii β_j și nu în variabilele explicative, e.g. $f_1(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, $f_2(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \underbrace{x_1 x_2}_{x_3}$ sau $f_3(x_1, x_2) = \beta_0 + \beta_1 x_1 +$

$\beta_2 x_2 + \beta_3 \underbrace{x_1 x_2}_{x_3} + \beta_4 \underbrace{x_1^2}_{x_4} + \beta_5 \underbrace{x_2^3}_{x_5}$ sunt toate funcții liniare în $\boldsymbol{\beta}$.

$$f_1(x_1, x_2) = -1 + 2x_1 + 3x_2$$

$$f_2(x_1, x_2) = x_1 + 3x_2 + 2x_1 x_2$$

$$f_3(x_1, x_2) = 3 + x_1 + 8x_2 - 4x_1 x_2 + 2x_1^2 + 3x_2^2$$



2.1 Modelare

Modelul de regresie liniară multiplă reprezintă o extensie a modelului de regresie liniară simplă atunci când numărul variabilelor explicative este finit (pentru mai multe detalii privind analiza de regresie se pot consulta monografiile [Seber and Lee, 2003], [Weisberg, 2014] sau [Faraway, 2015]). Presupunem că datele verifică modelul

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, 2, \dots, n$$

unde

- x_{ij} sunt valori cunoscute și nu sunt aleatoare
- parametrii β_j sunt necunoscuți și nu sunt aleatori
- ε_i sunt variabile aleatoare necunoscute

Scris sub formă compactă modelul devine

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

unde

- \mathbf{X} se numește *matricea de design*

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}_{n \times (p+1)}$$

- \mathbf{Y} este *vectorul răspuns* și este un vector aleator, $\boldsymbol{\beta}$ este *vectorul parametrilor* sau coeficienților și este necunoscut iar $\boldsymbol{\varepsilon}$ este *vectorul erorilor* și este un vector aleator

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}_{n \times 1}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}_{(p+1) \times 1} \quad \text{și} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}_{n \times 1}.$$

Ipotezele modelului de regresie liniară multiplă sunt:

$$\begin{aligned} \mathcal{H}_1 &: \text{rang}(\mathbf{X}) = p + 1 \\ \mathcal{H}_2 &: \mathbb{E}[\boldsymbol{\varepsilon}] = 0, \text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 I_n \end{aligned}$$

Prima ipoteză ne spune că matricea de design \mathbf{X} are coloanele liniar independente iar a doua ipoteză se referă la centralitatea erorilor (medie nulă), homoscedasticitatea (aceeași varianță) și necorelarea acestora.

2.2 Metoda celor mai mici pătrate

În această secțiune vom prezenta estimatorul lui $\boldsymbol{\beta}$ obținut prin metoda celor mai mici pătrate, metodă care folosește ca funcție de cost L funcția $L(x) = x^2$. Estimatorul $\hat{\boldsymbol{\beta}}$ obținut prin metoda celor mai mici pătrate este definit prin

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})]^2 = \arg \min_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^n \left(y_i - \sum_{j=0}^p \beta_j x_{ij} \right)^2 = \arg \min_{\beta \in \mathbb{R}^{p+1}} \|Y - X\beta\|^2$$

unde am folosit convenția $x_{i0} = 1$, $i = 1, 2, \dots, n$.

Ex. 2.1



Arătați că dacă ipoteza \mathcal{H}_1 este adevărată atunci estimatorul $\hat{\beta}$ obținut prin metoda celor mai mici pătrate este

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Ne propunem să prezentăm mai multe metode de calcul pentru estimatorul $\hat{\beta}$.

a) Metoda geometrică

Din punct de vedere geometric, ne plasăm în spațiul variabilelor \mathbb{R}^n (am presupus că $n > p$). Vectorul variabilelor răspuns $Y \in \mathbb{R}^n$ iar matricea de design X poate fi văzută ca fiind formată din $p + 1$ vectori

$$\text{coloană, } X = \begin{bmatrix} \underbrace{X_0}_{\mathbf{1}=(1, \dots, 1)} & |X_1| \cdots |X_p| \end{bmatrix}.$$

Începem prin a reaminti câteva noțiuni de algebră liniară, pentru mai multe detalii se poate consulta monografia [Ornea and Turtoi, 2000]. Fiind dată matricea (de design) $X \in \mathcal{M}_{n,p+1}(\mathbb{R})$ putem defini nucleul matricei

$$\ker(X) = \{u \in \mathbb{R}^{p+1} \mid Xu = 0\}$$

ca fiind subspațiul lui \mathbb{R}^{p+1} care conține vectorii ortogonali pe liniile din matricea X . De asemenea, imaginea matricei X este subspațiul vectorilor din \mathbb{R}^n care se pot scrie ca o combinație liniară de coloanele matricei X și este definit prin

$$\text{Im}(X) = \{v \in \mathbb{R}^n \mid \exists u \in \mathbb{R}^{p+1} \text{ a.î. } Xu = v\}.$$

Se poate arăta (a se vedea [Ornea and Turtoi, 2000, Capitolul 1]) că $\dim \ker(X) = p + 1 - \text{rang}(X)$ și că $\dim \text{Im}(X) = \text{rang}(X)$ ceea ce conduce la $\dim \ker(X) + \dim \text{Im}(X) = p + 1$ (caz particular al *teoremei lui Grassman*).

Reamintim că doi vectori u și v sunt ortogonali dacă $\langle u, v \rangle = 0$, iar în acest caz scriem $u \perp v$, și că două subspații V și W sunt ortogonale dacă $\forall v \in V$ și respectiv $\forall w \in W$ avem $v \perp w$ și notăm $V \perp W$. De asemenea, spațiul ortogonal al lui V este spațiul V^\perp care conține toți vectorii ortogonali pe V și are ca proprietăți: $V \cap V^\perp = \{0\}$ și $(V^\perp)^\perp = V$. Se poate arăta cu ușurință că dacă V este un subspațiu a lui $W \subset \mathbb{R}^n$ atunci pentru orice vector $w \in W$ avem că $w = v + v^\perp$ unde $v \in V$ și $v^\perp \in V^\perp$ iar descompunerea se face în mod unic (acest rezultat se mai scrie și sub forma $W = V \oplus V^\perp$). Se numește *proiecție ortogonală*¹ a lui w pe V de-a lungul lui V^\perp endomorfismul $p_V : W \rightarrow W$ definit prin $p_V(w) = v$. Se poate arăta cu ușurință că dacă V este un subspațiu vectorial al lui W atunci au loc următoarele proprietăți:

- i) pentru orice $w \in W$ avem: $w = p_V(w) + p_{V^\perp}(w)$, $\|p_V(w)\| \leq \|w\|$ iar $w - p_V(w) \perp v$, $\forall v \in V$
- ii) pentru orice $w \in W$ are loc $\|w - p_V(w)\| = \inf_{v \in V} \|w - v\|$

¹În general, numim *proiecție* un endomorfism $p : W \rightarrow W$ cu proprietatea că există o descompunere în sumă directă $W = V_1 \oplus V_2$ (și spunem *proiecție pe V_1 de-a lungul lui V_2*) astfel încât $p(w) = v_1$, $\forall w \in W$ cu $w = v_1 + v_2$, $v_1 \in V_1$ și $v_2 \in V_2$.

- iii) dacă p este o proiecție ortogonală atunci $p \circ p = p$
- iv) dacă p este un endomorfism care verifică $p \circ p = p$ și în plus $\text{Im}(p) \perp \ker(p)$ atunci p este proiecția ortogonală pe $\text{Im}(p)$ de-a lungul lui $\ker(p)$
- v) dacă $\{e_1, e_2, \dots, e_k\}$ este o bază ortonormală în V atunci $p_V(w) = \sum_{i=1}^k \langle w, e_i \rangle e_i$

Spunem că o matrice pătratică $P \in \mathcal{M}_n(\mathbb{R})$ este o matrice de proiecție dacă este idempotentă $P^2 = P$ (numele vine de la faptul că pentru $x \in \mathbb{R}^n$ aplicația liniară Px este proiecția lui x pe $\text{Im}(P)$ de-a lungul lui $\ker(P)$ - a se vedea [Ornea and Turtoi, 2000, Capitolul 1, Secțiunea 5.2] sau [Yanai et al., 2011, Capitolul 2]). Dacă în plus matricea P este simetrică, i.e. $P = P^\top$, atunci Px este proiecția ortogonală a lui x pe $V = \text{Im}(P)$ de-a lungul lui $V^\perp = \ker(P)$, cu alte cuvinte în descompunerea $x = Px + (I - P)x$ vectorii Px și respectiv $(I - P)x$ sunt ortogonali [Yanai et al., 2011, Capitolul 2, Secțiunea 2.2]. Prin urmare matricea P este o matrice de proiecție ortogonală dacă are loc relația $v \perp v - Pv$ adică $\langle v, v - Pv \rangle = 0$ pentru toți v .

Dacă $X \in \mathcal{M}_{m,k}(\mathbb{R})$ este o matrice de rang $(X) = k$ ($m \geq k$), deci $X^\top X$ este inversabilă, atunci pentru a determina matricea de proiecție ortogonală P_X pe subspațiul imagine $\text{Im}(X)$ să observăm că dacă $v \in \text{Im}(X)$ atunci $v = X\alpha$ și cum $P_X v = v$ deducem că $P_X v = X \underbrace{(X^\top X)^{-1} X^\top v}_{=\alpha}$. În plus dacă $v^\perp \in \text{Im}(X)^\perp =$

$\ker(X^\top)$ atunci $X^\top v^\perp = 0$ prin urmare $X (X^\top X)^{-1} X^\top v^\perp = 0$ și astfel, pentru $w = v + v^\perp$ arbitrar, avem $P_X w = X (X^\top X)^{-1} X^\top w$ de unde găsim că

$$P_X = X (X^\top X)^{-1} X^\top.$$

În mod similar se arată că matricea de proiecție ortogonală pe $\ker(X)$ este $P_{X^\perp} = I - P_X$. Nu este dificil de văzut că cele două matrice, P_X și respectiv P_{X^\perp} , sunt idempotente. De asemenea, ținând seama că valorile proprii ale unei matrice idempotente sunt 0 sau 1 concluzionăm că $\text{rang}(P_X) = \text{Tr}(P_X)$.

În cazul particular în care $X = v$ avem

$$P_v = v (v^\top v)^{-1} v^\top = \frac{vv^\top}{\|v\|^2}$$

prin urmare proiecția vectorului u pe v este $P_v u = \frac{vv^\top}{\|v\|^2} u = \frac{\langle u, v \rangle}{\|v\|^2} v$.

Pentru problema noastră, notăm cu $\mathcal{M}(X) = \text{Im}(X)$ subspațiul imagine și conform ipotezei \mathcal{H}_1 avem că $\text{rang}(X) = p + 1$, deci $\dim \mathcal{M}(X) = p + 1$. Din definiția spațiului imagine avem că toți vectorii din $\mathcal{M}(X)$ sunt de forma $X\alpha$, cu $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_p) \in \mathbb{R}^{p+1}$

$$X\alpha = \sum_{i=0}^p \alpha_i X_i.$$

Conform modelului de regresie, $Y = X\beta + \varepsilon$, vectorul răspuns Y este suma dintre un element din $\mathcal{M}(X)$ și un element din \mathbb{R}^n care nu are niciun motiv să aparțină tot lui $\mathcal{M}(X)$. Astfel, problema minimizării funcției $S(\beta) = \|Y - X\beta\|^2$ revine la a găsi acel vector din $\mathcal{M}(X)$ care este cel mai aproape de Y în sensul distanței euclidiene, cu alte cuvinte de a determina vectorul proiecției ortogonale a lui Y pe $\mathcal{M}(X)$ (a se vedea figura de mai jos).

Proiecția ortogonală a lui Y pe $\mathcal{M}(X)$ se notează cu $\hat{Y} = P_X Y$, unde P_X este matricea de proiecție ortogonală pe $\mathcal{M}(X)$, iar spațiul ortogonal $\mathcal{M}(X)^\perp$ se mai numește și spațiul reziduurilor și are dimensiunea $\dim \mathcal{M}(X)^\perp = n - (p + 1)$ (teorema lui Grassman). Să remarcăm că $\hat{Y} \in \mathcal{M}(X)$ deci putem scrie $\hat{Y} = X\hat{\beta}$ unde $\hat{\beta}$ reprezintă estimatorul obținut prin metoda celor mai mici pătrate iar elementele lui $\hat{\beta}$ sunt coordonatele vectorului \hat{Y} în baza $\{X_0, X_1, \dots, X_p\}$ a spațiului $\mathcal{M}(X)$. Cum reperul $\{X_0, X_1, \dots, X_p\}$ nu este neapărat ortogonal, elementele $\hat{\beta}_j$ nu sunt neapărat coordonatele proiecției lui Y pe X_j , aceasta din urmă fiind dată de

$$\begin{aligned} P_{X_j} \mathbf{Y} &= P_{X_j} P_X \mathbf{Y} = P_{X_j} \sum_{i=0}^p \hat{\beta}_i \mathbf{X}_i \\ &= \sum_{i=0}^p \hat{\beta}_i P_{X_j} \mathbf{X}_i = \hat{\beta}_j \mathbf{X}_j + \sum_{i \neq j} \hat{\beta}_i P_{X_j} \mathbf{X}_i \end{aligned}$$

În cazul în care \mathbf{X}_j este ortogonal pe \mathbf{X}_i , $i \neq j$, atunci $P_{X_j} \mathbf{Y} = \hat{\beta}_j \mathbf{X}_j$ iar dacă $\langle \mathbf{X}_i, \mathbf{X}_j \rangle = 0$ pentru toți $i \neq j$ atunci matricea $\mathbf{X}^\top \mathbf{X} = \text{diag}(\|\mathbf{X}_0\|^2, \|\mathbf{X}_1\|^2, \dots, \|\mathbf{X}_p\|^2)$.

O altă metodă (tot prin proiecții) de a arăta că $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$ se bazează pe observația că proiecția ortogonală $\hat{\mathbf{Y}} = \mathbf{X} \hat{\beta}$ este definită ca fiind unicul vector pentru care $\mathbf{Y} - \hat{\mathbf{Y}}$ este ortogonal pe $\mathcal{M}(X)$. Cum $\mathcal{M}(X)$ este generat de $\{\mathbf{X}_0, \mathbf{X}_1, \dots, \mathbf{X}_p\}$ putem spune că $\mathbf{Y} - \hat{\mathbf{Y}}$ este ortogonal pe fiecare \mathbf{X}_i :

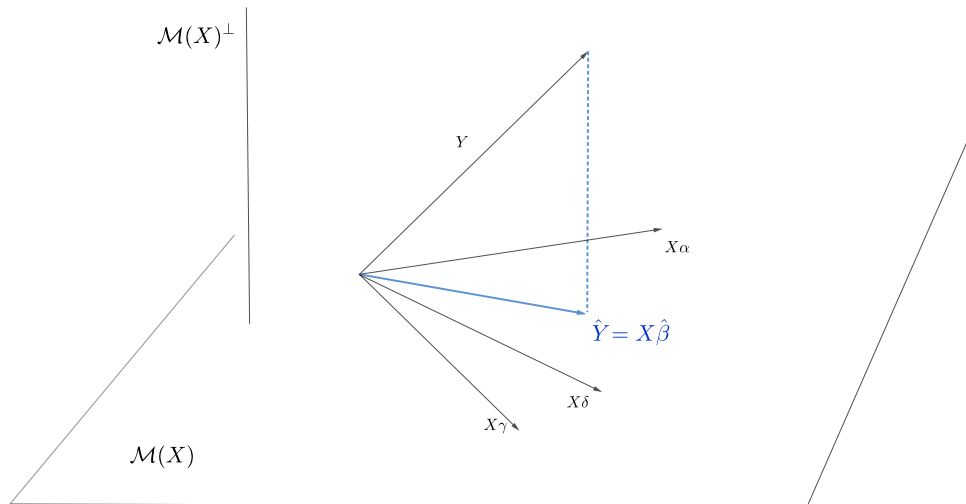
$$\begin{cases} \langle \mathbf{X}_0, \mathbf{Y} - \hat{\mathbf{Y}} \rangle = 0 \\ \langle \mathbf{X}_1, \mathbf{Y} - \hat{\mathbf{Y}} \rangle = 0 \\ \vdots \\ \langle \mathbf{X}_p, \mathbf{Y} - \hat{\mathbf{Y}} \rangle = 0 \end{cases} \iff \begin{cases} \langle \mathbf{X}_0, \mathbf{Y} - \mathbf{X} \hat{\beta} \rangle = 0 \\ \langle \mathbf{X}_1, \mathbf{Y} - \mathbf{X} \hat{\beta} \rangle = 0 \\ \vdots \\ \langle \mathbf{X}_p, \mathbf{Y} - \mathbf{X} \hat{\beta} \rangle = 0 \end{cases} \iff \mathbf{X}^\top (\mathbf{Y} - \mathbf{X} \hat{\beta}) = 0$$

de unde găsim *sistemul de ecuații normale*

$$\mathbf{X}^\top \mathbf{X} \hat{\beta} = \mathbf{X}^\top \mathbf{Y}$$

care, atunci când ipoteza \mathcal{H}_1 este adevărată - matricea $\mathbf{X}^\top \mathbf{X}$ este inversabilă, revine la

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}.$$



Notând cu $P_{X^\perp} = I_n - P_X$ matricea de proiecție ortogonală pe $\mathcal{M}^\perp(X)$, unde $P_X = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ este matricea de proiecție ortogonală pe $\mathcal{M}(X)$, observăm că descompunerea

$$\mathbf{Y} = \hat{\mathbf{Y}} + (\mathbf{Y} - \hat{\mathbf{Y}}) = P_X \mathbf{Y} + (I - P_X) \mathbf{Y} = P_X \mathbf{Y} + P_{X^\perp} \mathbf{Y}$$

nu este altceva decât descompunerea ortogonală a lui \mathbf{Y} pe $\mathcal{M}(X)$ și respectiv $\mathcal{M}^\perp(X)$. De asemenea este de remarcat faptul că în ceea ce privește notația folosită în literatura de statistică de specialitate, matricea de proiecție ortogonală P_X se mai notează și cu H (care vine de la *hat*, $\hat{\mathbf{Y}} = H\mathbf{Y}$).

b) Metoda analitică

O a doua metodă este metoda analitică. Problema noastră cere să căutăm vectorul $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$ care minimizează funcția

$$\begin{aligned} S(\boldsymbol{\beta}) &= \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} \\ &= \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} - 2\mathbf{Y}^\top \mathbf{X}\boldsymbol{\beta} + \|\mathbf{Y}\|^2 \end{aligned}$$

unde am folosit faptul că $\mathbf{Y}^\top \mathbf{X}\boldsymbol{\beta} = (\mathbf{Y}^\top \mathbf{X}\boldsymbol{\beta})^\top = \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y}$ (sunt elemente de dimensiune 1×1). Pentru aceasta trebuie să rezolvăm ecuația $\nabla S(\boldsymbol{\beta}) = 0$ și să verificăm că soluția este într-adevăr punct de minim.

Reamintim că dacă $f : \mathbb{R}^k \rightarrow \mathbb{R}$ atunci

$$\nabla f = \frac{\partial f}{\partial \mathbf{x}^\top} = \left(\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \cdots \quad \frac{\partial f}{\partial x_k} \right).$$

În particular, pentru $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x}$ (formă liniară) avem

$$\nabla f = \frac{\partial \mathbf{a}^\top \mathbf{x}}{\partial \mathbf{x}^\top} = \left(\frac{\partial \sum_{i=1}^k a_i x_i}{\partial x_1} \quad \frac{\partial \sum_{i=1}^k a_i x_i}{\partial x_2} \quad \cdots \quad \frac{\partial \sum_{i=1}^k a_i x_i}{\partial x_k} \right) = \mathbf{a}^\top$$

iar $\frac{\partial \mathbf{a}^\top \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$.

În cazul în care f este o aplicație liniară, $f(\mathbf{x}) = A\mathbf{x}$ unde $A \in \mathcal{M}_{m,k}(\mathbb{R})$, atunci

$$A\mathbf{x} = \begin{pmatrix} \sum_{j=1}^k a_{1j}x_j \\ \sum_{j=1}^k a_{2j}x_j \\ \vdots \\ \sum_{j=1}^k a_{mj}x_j \end{pmatrix}, \quad \frac{\partial A\mathbf{x}}{\partial x_j} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

și

$$\frac{\partial A\mathbf{x}}{\partial \mathbf{x}^\top} = \left(\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \dots, \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}, \dots, \begin{pmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{pmatrix} \right) = \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{ik} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mk} \end{pmatrix} = A.$$

În mod similar se poate verifica și relația $\frac{\partial \mathbf{x}^\top A^\top}{\partial \mathbf{x}} = A^\top$.

Dacă f este o formă pătratică $f(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x}$ cu $A \in \mathcal{M}_k(\mathbb{R})$ (matrice pătratică de ordin k), atunci

$$\mathbf{x}^\top A\mathbf{x} = \sum_{i=1}^k \sum_{j=1}^k a_{ij}x_i x_j = \sum_{i=1}^k a_{ii}x_i^2 + \sum_{i=1}^k \sum_{\substack{j=1 \\ j \neq i}}^k a_{ij}x_i x_j$$

de unde

$$\frac{\partial \mathbf{x}^\top A \mathbf{x}}{\partial x_r} = 2a_{rr}x_r + \sum_{j \neq r} a_{rj}x_j + \sum_{i \neq r} a_{ir}x_i = \sum_{j=1}^k a_{rj}x_j + \sum_{i=1}^k a_{ir}x_i$$

prin urmare

$$\frac{\partial \mathbf{x}^\top A \mathbf{x}}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial \mathbf{x}^\top A \mathbf{x}}{\partial x_1} \\ \vdots \\ \frac{\partial \mathbf{x}^\top A \mathbf{x}}{\partial x_k} \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^k a_{1j}x_j + \sum_{i=1}^k a_{i1}x_i \\ \vdots \\ \sum_{j=1}^k a_{kj}x_j + \sum_{i=1}^k a_{ik}x_i \end{pmatrix} = A\mathbf{x} + A^\top \mathbf{x}.$$

De asemenea putem observa că $\frac{\partial \mathbf{x}^\top A \mathbf{x}}{\partial \mathbf{x}^\top} = (A\mathbf{x} + A^\top \mathbf{x})^\top = \mathbf{x}^\top A^\top + \mathbf{x}^\top A$.

În plus, dacă $A \in \mathcal{M}_k(\mathbb{R})$ este o matrice simetrică ($A^\top = A$) atunci

$$\frac{\partial \mathbf{x}^\top A \mathbf{x}}{\partial \mathbf{x}} = 2A\mathbf{x}, \quad \frac{\partial \mathbf{x}^\top A \mathbf{x}}{\partial \mathbf{x}^\top} = 2\mathbf{x}^\top A^\top.$$

Revenind la problema noastră observăm că $S(\beta)$ este pătratică în β iar matricea $\mathbf{X}^\top \mathbf{X}$ este simetrică, prin urmare

$$\nabla S(\beta) = \frac{\partial S(\beta)}{\partial \beta^\top} = \frac{\partial}{\partial \beta^\top} (\beta^\top \mathbf{X}^\top \mathbf{X} \beta - 2\mathbf{Y}^\top \mathbf{X} \beta + \|\mathbf{Y}\|^2) = 2\beta^\top (\mathbf{X}^\top \mathbf{X}) - 2\mathbf{Y}^\top \mathbf{X} = 0$$

este echivalentă cu *sistem de ecuații normale* (prin trecerea la transpusă)

$$(\mathbf{X}^\top \mathbf{X})\beta = \mathbf{X}^\top \mathbf{Y}$$

care atunci când ipoteza \mathcal{H}_1 este adevărată, ceea ce conduce la inversabilitatea matricei $\mathbf{X}^\top \mathbf{X}$ (are valori proprii nenule), revine la

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}.$$

Pentru a arăta că $\hat{\beta}$ este într-adevăr un punct de minim pentru $S(\beta)$ trebuie să arătăm că matricea hessiană este pozitiv definită. Matricea hessiană, ținând cont de simetria matricii $\mathbf{X}^\top \mathbf{X}$, este

$$\frac{\partial^2 S(\beta)}{\partial \beta \partial \beta^\top} = \frac{\partial}{\partial \beta} \left(\frac{\partial S(\beta)}{\partial \beta^\top} \right) = \frac{\partial}{\partial \beta} (2\beta^\top (\mathbf{X}^\top \mathbf{X}) - 2\mathbf{Y}^\top \mathbf{X}) = 2\mathbf{X}^\top \mathbf{X}$$

iar pentru $\mathbf{u} \in \mathbb{R}^{p+1} \setminus \{0\}$ avem

$$\mathbf{u}^\top \frac{\partial^2 S(\beta)}{\partial \beta \partial \beta^\top} \mathbf{u} = \mathbf{u}^\top (2\mathbf{X}^\top \mathbf{X}) \mathbf{u} = \langle \mathbf{X} \mathbf{u}, \mathbf{X} \mathbf{u} \rangle = 2\|\mathbf{X} \mathbf{u}\|^2 > 0$$

deci $\mathbf{X}^\top \mathbf{X}$ este pozitiv definită prin urmare și $\frac{\partial^2 S(\beta)}{\partial \beta \partial \beta^\top}$ este pozitiv definită.

Ex. 2.2



Arătați că dacă ipotezele \mathcal{H}_1 și \mathcal{H}_2 sunt adevărate atunci estimatorul $\hat{\beta}$ obținut prin metoda celor mai mici pătrate este nedeplasat, i.e. $\mathbb{E}[\hat{\beta}] = \beta$ iar matricea de varianță-covarianță $Var(\hat{\beta})$ este

$$Var(\hat{\beta}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}.$$

Pentru început, deoarece vorbim de operații cu vectori aleatori sau matrice aleatoare, vom reaminti câteva proprietăți de calcul a acestora (acestea generalizează noțiunile corespunzătoare din cazul variabilelor aleatoare). Reamintim că fiind dată matricea $\mathbf{Z} = (Z_{ij})_{i,j}$ unde Z_{ij} , $1 \leq i \leq m$, $1 \leq j \leq k$ sunt variabile aleatoare de medie $\mathbb{E}[Z_{ij}]$, media matricei $\mathbb{E}[\mathbf{Z}]$ este definită ca matricea mediilor $(\mathbb{E}[Z_{ij}])_{i,j}$. În plus dacă $\mathbf{A} \in \mathcal{M}_{l,m}(\mathbb{R})$, $\mathbf{B} \in \mathcal{M}_{k,q}(\mathbb{R})$ și $\mathbf{C} \in \mathcal{M}_{l,q}(\mathbb{R})$ sunt trei matrice cu coeficienți constanți atunci

$$\mathbb{E}[\mathbf{AZB} + \mathbf{C}] = \mathbf{A}\mathbb{E}[\mathbf{Z}]\mathbf{B} + \mathbf{C}.$$

În mod similar, dacă \mathbf{Z} și \mathbf{T} sunt doi vectori aleatori de dimensiune $m \times 1$ și respectiv $k \times 1$ atunci covarianța acestora este definită prin $Cov(\mathbf{Z}, \mathbf{T}) = (Cov(Z_i, T_j))_{i,j}$ și se poate verifica relația

$$Cov(\mathbf{Z}, \mathbf{T}) = \mathbb{E}[(\mathbf{Z} - \mathbb{E}[\mathbf{Z}])(\mathbf{T} - \mathbb{E}[\mathbf{T}])^\top] = \mathbb{E}[\mathbf{ZT}^\top] - \mathbb{E}[\mathbf{Z}]\mathbb{E}[\mathbf{T}]^\top = Cov(\mathbf{T}, \mathbf{Z})^\top.$$

În particular pentru $\mathbf{T} = \mathbf{Z}$ avem matricea de varianță-covarianță $Var(\mathbf{Z}) = Cov(\mathbf{Z}, \mathbf{Z})$ care, în conformitate cu relația de mai sus, este egală cu

$$Var(\mathbf{Z}) = \mathbb{E}[(\mathbf{Z} - \mathbb{E}[\mathbf{Z}])(\mathbf{Z} - \mathbb{E}[\mathbf{Z}])^\top] = \mathbb{E}[\mathbf{ZZ}^\top] - \mathbb{E}[\mathbf{Z}]\mathbb{E}[\mathbf{Z}]^\top.$$

Un calcul direct arată că dacă $\mathbf{A} \in \mathcal{M}_{l,m}(\mathbb{R})$ și $\mathbf{B} \in \mathcal{M}_{q,k}(\mathbb{R})$ sunt două matrice cu coeficienți constanți atunci

$$Cov(\mathbf{AZ}, \mathbf{BT}) = \mathbf{A}Cov(\mathbf{Z}, \mathbf{T})\mathbf{B}^\top$$

iar atunci când $\mathbf{T} = \mathbf{Z}$ și $\mathbf{A} = \mathbf{B}$ găsim că $Var(\mathbf{AZ}) = \mathbf{A}Var(\mathbf{Z})\mathbf{A}^\top$. De asemenea, dacă \mathbf{Z} , \mathbf{T} , \mathbf{U} și \mathbf{V} sunt vectori aleatori de dimensiune $m \times 1$ iar $a, b, c, d \in \mathbb{R}$ sunt constante reale atunci

$$Cov(a\mathbf{Z} + b\mathbf{T}, c\mathbf{U} + d\mathbf{V}) = acCov(\mathbf{Z}, \mathbf{U}) + adCov(\mathbf{Z}, \mathbf{V}) + bcCov(\mathbf{T}, \mathbf{U}) + bdCov(\mathbf{T}, \mathbf{V})$$

și respectiv

$$Var(a\mathbf{Z} + b\mathbf{T}) = a^2Var(\mathbf{Z}) + ab(Cov(\mathbf{Z}, \mathbf{T}) + Cov(\mathbf{T}, \mathbf{Z})) + b^2Var(\mathbf{T}).$$

Pentru a verifica nedeplasarea estimatorului $\hat{\beta}$ obținut prin metoda celor mai mici pătrate să notăm că, în contextul ipotezei \mathcal{H}_1 , acesta este $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$. Astfel putem scrie

$$\mathbb{E}[\hat{\beta}] = \mathbb{E}[(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}] = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbb{E}[\mathbf{Y}] = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbb{E}[\mathbf{X}\beta + \varepsilon]$$

și cum $\mathbb{E}[\varepsilon] = 0$ conform \mathcal{H}_2 deducem că

$$\mathbb{E}[\hat{\beta}] = \underbrace{(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}}_{=I_{p+1}} \beta + \underbrace{\mathbb{E}[\varepsilon]}_{=0} = \beta.$$

Pentru matricea de varianță-covarianță avem

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \text{Var}((\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \text{Var}(\mathbf{Y}) ((\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top)^\top \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \text{Var}(\mathbf{Y}) \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \end{aligned}$$

unde în ultima egalitate am ținut cont de simetria matricei $\mathbf{X}^\top \mathbf{X}$: $((\mathbf{X}^\top \mathbf{X})^{-1})^\top = (\mathbf{X}^\top \mathbf{X})^{-1}$. Din modelul de regresie avem că $\text{Var}(\mathbf{Y}) = \text{Var}(\mathbf{X}\beta + \varepsilon) = \text{Var}(\varepsilon)$ și din ipoteza \mathcal{H}_2 avem $\text{Var}(\varepsilon) = \sigma^2 I_n$ prin urmare

$$\text{Var}(\hat{\beta}) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \underbrace{\text{Var}(\mathbf{Y})}_{=\sigma^2 I_n} \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} = \sigma^2 \underbrace{(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1}}_{=I_{p+1}} = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}.$$

Ex. 2.3



Estimatorul $\hat{\beta}$ obținut prin metoda celor mai mici pătrate este estimatorul liniar, nedeplasat de varianță minimală (BLUE).

Rezultatul acestui exercițiu este cunoscut și sub denumirea de **Teorema Gauss-Markov** (a se vedea [Faraway, 2015, Capitolul 2]). Pentru a putea soluționa acest exercițiu trebuie pentru început să remarcăm câteva aspecte: în primul rând, *liniaritatea* estimatorului se referă la *liniaritatea în raport cu Y*, adică vorbim de clasa estimatorilor de forma $\mathbf{A}\mathbf{Y}$ unde $\mathbf{A} \in \mathcal{M}_{p+1,n}(\mathbb{R})$; în al doilea rând pentru a determina *varianța minimală* avem nevoie de o relație parțială de ordine pe mulțimea matricelor simetrice, ori o asemenea relație există și spune că $\mathbf{S}_1 \leq \mathbf{S}_2$ atunci când $\mathbf{S} = \mathbf{S}_2 - \mathbf{S}_1$ este pozitiv definită (valorile proprii ale lui \mathbf{S} sunt ≥ 0), cu alte cuvinte $\mathbf{S}_1 \leq \mathbf{S}_2$ dacă $\mathbf{x}^\top \mathbf{S}_1 \mathbf{x} \leq \mathbf{x}^\top \mathbf{S}_2 \mathbf{x}$ pentru orice \mathbf{x} .

În acest context, cum $\hat{\beta}$ este un estimator nedeplasat (conform exercițiului anterior) și liniar deoarece $\hat{\beta} = \underbrace{(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top}_{\mathbf{A}} \mathbf{Y}$, considerând $\tilde{\beta}$ un alt estimator liniar și nedeplasat pentru β vrem să arătăm că

$\text{Var}(\tilde{\beta}) \geq \text{Var}(\hat{\beta})$. Avem, conform proprietăților matricei de varianță-covarianță de mai sus, descompunerea

$$\text{Var}(\tilde{\beta}) = \text{Var}(\tilde{\beta} - \hat{\beta} + \hat{\beta}) = \text{Var}(\tilde{\beta} - \hat{\beta}) + \text{Var}(\hat{\beta}) + \text{Cov}(\tilde{\beta} - \hat{\beta}, \hat{\beta}) + \text{Cov}(\hat{\beta}, \tilde{\beta} - \hat{\beta}).$$

Cum matricele de varianță-covarianță sunt pozitiv semidefinite (pentru $\mathbf{a} \in \mathbb{R}^{p+1}$ avem $\mathbf{a}^\top \text{Var}(\tilde{\beta}) \mathbf{a} = \text{Var}(\mathbf{a}^\top \tilde{\beta}) \geq 0$, deoarece varianța este pozitivă) este suficient să arătăm că $\text{Cov}(\tilde{\beta} - \hat{\beta}, \hat{\beta}) = 0$. Liniaritatea estimatorului $\tilde{\beta}$ implică $\tilde{\beta} = \mathbf{A}\mathbf{Y}$ iar nedeplasarea acestuia, $\mathbb{E}[\tilde{\beta}] = \beta$ pentru orice β , conduce la $\mathbb{E}[\mathbf{A}\mathbf{Y}] = \mathbf{A}\mathbf{X}\beta = \beta$ pentru orice β de unde $\mathbf{A}\mathbf{X} = I_{p+1}$.

În final avem

$$\begin{aligned} \text{Cov}(\tilde{\beta} - \hat{\beta}, \hat{\beta}) &= \text{Cov}(\tilde{\beta}, \hat{\beta}) - \text{Var}(\hat{\beta}) = \text{Cov}(\mathbf{A}\mathbf{Y}, (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}) - \text{Var}(\hat{\beta}) \\ &= \mathbf{A} \text{Cov}(\mathbf{Y}, \mathbf{Y}) ((\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top)^\top - \text{Var}(\hat{\beta}) = \mathbf{A} \underbrace{\text{Var}(\mathbf{Y})}_{=\sigma^2 I_n} \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} - \underbrace{\text{Var}(\hat{\beta})}_{=\sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}} \\ &= \sigma^2 \underbrace{\mathbf{A}\mathbf{X}}_{=I_{p+1}} (\mathbf{X}^\top \mathbf{X})^{-1} - \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} = 0. \end{aligned}$$

Ex. 2.4



Fie $\hat{\varepsilon} = \mathbf{Y} - \hat{\mathbf{Y}}$ vectorul valorilor reziduale. Arătați că sub ipotezele \mathcal{H}_1 și \mathcal{H}_2 au loc relațiile

1. $\mathbb{E}[\hat{\varepsilon}] = 0$ și $\text{Var}(\hat{\varepsilon}) = \sigma^2 P_{X^\perp}$
2. $\mathbb{E}[\hat{\mathbf{Y}}] = \mathbf{X}\beta$ și $\text{Var}(\hat{\mathbf{Y}}) = \sigma^2 P_X$
3. $\text{Cov}(\hat{\varepsilon}, \hat{\mathbf{Y}}) = 0$

1. Am văzut că $\hat{\mathbf{Y}} = P_X \mathbf{Y}$ prin urmare

$$\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n)^\top = \mathbf{Y} - \hat{\mathbf{Y}} = (I - P_X) \mathbf{Y} = P_{X^\perp} \mathbf{Y} = P_{X^\perp} (\mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}) = P_{X^\perp} \boldsymbol{\varepsilon}$$

deoarece $\mathbf{X} \boldsymbol{\beta} \in \mathcal{M}(X)$, deci $P_{X^\perp} \mathbf{X} \boldsymbol{\beta} = 0$. Astfel

$$\mathbb{E}[\hat{\boldsymbol{\varepsilon}}] = \mathbb{E}[P_{X^\perp} \boldsymbol{\varepsilon}] = P_{X^\perp} \mathbb{E}[\boldsymbol{\varepsilon}] = 0$$

iar, ținând seama de proprietatea de simetrie $P_{X^\perp}^\top = P_{X^\perp}$ și idempotență $P_{X^\perp}^2 = P_{X^\perp}$ a matricei de proiecție,

$$\text{Var}(\hat{\boldsymbol{\varepsilon}}) = \text{Var}(P_{X^\perp} \boldsymbol{\varepsilon}) = P_{X^\perp} \text{Var}(\boldsymbol{\varepsilon}) P_{X^\perp}^\top = P_{X^\perp} \underbrace{\text{Var}(\boldsymbol{\varepsilon})}_{=\sigma^2 I_n} P_{X^\perp} = \sigma^2 P_{X^\perp}^2 = \sigma^2 P_{X^\perp}.$$

2. Din nedeplasarea lui $\hat{\boldsymbol{\beta}}$ și definiția lui $\hat{\mathbf{Y}} = \mathbf{X} \hat{\boldsymbol{\beta}}$ avem

$$\mathbb{E}[\hat{\mathbf{Y}}] = \mathbb{E}[\mathbf{X} \hat{\boldsymbol{\beta}}] = \mathbf{X} \mathbb{E}[\hat{\boldsymbol{\beta}}] = \mathbf{X} \boldsymbol{\beta}.$$

În mod similar

$$\text{Var}(\hat{\mathbf{Y}}) = \text{Var}(\mathbf{X} \hat{\boldsymbol{\beta}}) = \mathbf{X} \underbrace{\text{Var}(\hat{\boldsymbol{\beta}})}_{=\sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}} \mathbf{X}^\top = \sigma^2 \underbrace{\mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top}_{=P_X} = \sigma^2 P_X.$$

3. Pentru a calcula $\text{Cov}(\hat{\boldsymbol{\varepsilon}}, \hat{\mathbf{Y}})$ folosim $\text{Var}(\hat{\boldsymbol{\varepsilon}}) = \sigma^2 P_{X^\perp}$ și $\text{Var}(\mathbf{Y}) = \text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 I_n$ și avem

$$\begin{aligned} \text{Cov}(\hat{\boldsymbol{\varepsilon}}, \hat{\mathbf{Y}}) &= \text{Cov}(\hat{\boldsymbol{\varepsilon}}, \mathbf{Y} - \hat{\boldsymbol{\varepsilon}}) = \text{Cov}(\hat{\boldsymbol{\varepsilon}}, \mathbf{Y}) - \text{Var}(\hat{\boldsymbol{\varepsilon}}) = \text{Cov}(P_{X^\perp} \mathbf{Y}, \mathbf{Y}) - \sigma^2 P_{X^\perp} \\ &= P_{X^\perp} \text{Var}(\mathbf{Y}) P_{X^\perp}^\top - \sigma^2 P_{X^\perp} = \sigma^2 P_{X^\perp} - \sigma^2 P_{X^\perp} = 0. \end{aligned}$$

Ex. 2.5



Statistica $\hat{\sigma}^2 = \frac{\|\hat{\boldsymbol{\varepsilon}}\|^2}{n-(p+1)}$ este un estimator nedeplasat pentru σ^2 .

Ca să arătăm că $\hat{\sigma}^2 = \frac{\|\hat{\boldsymbol{\varepsilon}}\|^2}{n-(p+1)}$ este un estimator nedeplasat pentru σ^2 trebuie să calculăm $\mathbb{E}[\|\hat{\boldsymbol{\varepsilon}}\|^2]$. Cum $\|\hat{\boldsymbol{\varepsilon}}\|^2$ este un scalar atunci $\|\hat{\boldsymbol{\varepsilon}}\|^2 = \text{Tr}(\|\hat{\boldsymbol{\varepsilon}}\|^2)$ (este egal cu urma sa) prin urmare

$$\mathbb{E}[\|\hat{\boldsymbol{\varepsilon}}\|^2] = \mathbb{E}[\text{Tr}(\|\hat{\boldsymbol{\varepsilon}}\|^2)] = \mathbb{E}[\text{Tr}(\hat{\boldsymbol{\varepsilon}} \hat{\boldsymbol{\varepsilon}}^\top)]$$

și cum pentru orice matrice \mathbf{A} urma verifică $\text{Tr}(\mathbf{A}^\top \mathbf{A}) = \text{Tr}(\mathbf{A} \mathbf{A}^\top) = \sum_{i,j} a_{ij}^2$ avem ($\mathbb{E}[\hat{\boldsymbol{\varepsilon}}] = 0$)

$$\begin{aligned} \mathbb{E}[\|\hat{\boldsymbol{\varepsilon}}\|^2] &= \mathbb{E}[\text{Tr}(\hat{\boldsymbol{\varepsilon}} \hat{\boldsymbol{\varepsilon}}^\top)] = \mathbb{E}[\text{Tr}(\hat{\boldsymbol{\varepsilon}} \hat{\boldsymbol{\varepsilon}}^\top)] = \text{Tr}(\mathbb{E}[\hat{\boldsymbol{\varepsilon}} \hat{\boldsymbol{\varepsilon}}^\top]) = \text{Tr}(\mathbb{E}[\hat{\boldsymbol{\varepsilon}} \hat{\boldsymbol{\varepsilon}}^\top] - \mathbb{E}[\hat{\boldsymbol{\varepsilon}}] \mathbb{E}[\hat{\boldsymbol{\varepsilon}}]^\top) \\ &= \text{Tr}(\text{Var}(\hat{\boldsymbol{\varepsilon}})) = \text{Tr}(\sigma^2 P_{X^\perp}) = \sigma^2 \text{Tr}(P_{X^\perp}). \end{aligned}$$

Cum $P_{X^\perp} = I - P_X$ este matrice de proiecție avem că $\text{Tr}(P_{X^\perp}) = \text{rang}(P_{X^\perp}) = n - (p + 1)$ (urma matricei de proiecție este egală cu dimensiunea spațiului pe care proiectăm) prin urmare

$$\mathbb{E}[\|\hat{\boldsymbol{\varepsilon}}\|^2] = \sigma^2 [n - (p + 1)]$$

de unde $\mathbb{E}[\hat{\sigma}^2] = \mathbb{E}\left[\frac{\|\hat{\boldsymbol{\varepsilon}}\|^2}{n-(p+1)}\right] = \sigma^2$.

Folosind estimatorul $\hat{\sigma}^2$ pentru σ^2 putem construi un estimator $\hat{\sigma}_{\hat{\beta}}^2$ pentru varianța $Var(\hat{\beta}) = \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1}$:

$$\hat{\sigma}_{\hat{\beta}}^2 = \hat{\sigma}^2(\mathbf{X}^\top \mathbf{X})^{-1} = \frac{\|\hat{\varepsilon}\|^2}{n - (p + 1)}(\mathbf{X}^\top \mathbf{X})^{-1}.$$

În particular, un estimator pentru abaterea standard a estimatorului coeficientului β_j , $\hat{\beta}_j$, este dat de elementul $j + 1$ de pe diagonala matricii $(\mathbf{X}^\top \mathbf{X})^{-1}$:

$$\hat{\sigma}_{\hat{\beta}_j} = \hat{\sigma} \sqrt{[(\mathbf{X}^\top \mathbf{X})^{-1}]_{j+1, j+1}}, \quad j = 0, 1, \dots, p.$$

Ex. 2.6



Fie $(x_{n+1,1}, \dots, x_{n+1,p})$ o nouă observație și considerăm $\mathbf{x}_{n+1}^\top = (1, x_{n+1,1}, \dots, x_{n+1,p})$. Ne propunem să prezicem valoarea y_{n+1} conform modelului

$$y_{n+1} = \mathbf{x}_{n+1}^\top \beta + \varepsilon_{n+1}$$

cu $\mathbb{E}[\varepsilon_{n+1}] = 0$, $Var(\varepsilon_{n+1}) = \sigma^2$ și $Cov(\varepsilon_{n+1}, \varepsilon_i) = 0$ pentru $i = 1, \dots, n$.

Arătați că eroare de predicție $\hat{\varepsilon}_{n+1} = y_{n+1} - \hat{y}_{n+1}$ verifică proprietățile

1. $\mathbb{E}[\hat{\varepsilon}_{n+1}] = 0$
2. $Var(\hat{\varepsilon}_{n+1}) = \sigma^2 (1 + \mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{n+1})$

1. Cum $\mathbb{E}[\varepsilon_{n+1}] = 0$ și ținând cont de nedeplasarea estimatorului $\hat{\beta}$ avem

$$\mathbb{E}[\hat{\varepsilon}_{n+1}] = \mathbb{E}[y_{n+1} - \hat{y}_{n+1}] = \mathbb{E} \left[\underbrace{\mathbf{x}_{n+1}^\top \beta + \varepsilon_{n+1}}_{y_{n+1}} - \underbrace{\mathbf{x}_{n+1}^\top \hat{\beta}}_{\hat{y}_{n+1}} \right]$$

de unde $\mathbb{E}[\hat{\varepsilon}_{n+1}] = \mathbf{x}_{n+1}^\top (\beta - \mathbb{E}[\hat{\beta}]) + \mathbb{E}[\varepsilon_{n+1}] = 0$.

Cum $\hat{\beta}$ depinde doar de variabilele $\{\varepsilon_1, \dots, \varepsilon_n\}$ iar $Cov(\varepsilon_{n+1}, \varepsilon_i) = 0$ pentru $i = 1, \dots, n$ deducem că

$$\begin{aligned} Var(\hat{\varepsilon}_{n+1}) &= Var(y_{n+1} - \hat{y}_{n+1}) = Var(\mathbf{x}_{n+1}^\top (\beta - \hat{\beta}) + \varepsilon_{n+1}) \\ &= Var(\mathbf{x}_{n+1}^\top (\beta - \hat{\beta})) + Var(\varepsilon_{n+1}) = \mathbf{x}_{n+1}^\top Var(\beta - \hat{\beta}) \mathbf{x}_{n+1} + \sigma^2 \\ &= \mathbf{x}_{n+1}^\top Var(\hat{\beta}) \mathbf{x}_{n+1} + \sigma^2 = \mathbf{x}_{n+1}^\top \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{n+1} + \sigma^2 \\ &= \sigma^2 (1 + \mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{n+1}). \end{aligned}$$

Ex. 2.7



Fie modelul de regresie $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$ unde matricea de design se scrie

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}_{n \times (p+1)} = \begin{bmatrix} \underbrace{\mathbf{X}_0}_{\mathbf{1}=(1, \dots, 1)} & |\mathbf{X}_1| & \cdots & |\mathbf{X}_p| \end{bmatrix} = [\mathbf{1} | \mathbf{Z}]$$

cu \mathbf{Z} matricea de dimensiune $n \times p$ formată din coloanele $\{\mathbf{X}_1, \dots, \mathbf{X}_p\}$. În contextul exercițiului anterior, fie $\mathbf{z}_{n+1}^\top = (x_{n+1,1}, \dots, x_{n+1,p})$ o nouă observație și considerăm $\mathbf{x}_{n+1}^\top = (1, \mathbf{z}_{n+1}^\top)$. Arătați că varianța erorii de predicție este

$$\text{Var}(\hat{\varepsilon}_{n+1}) = \sigma^2 \left[1 + \frac{1}{n} + \frac{1}{n} (\mathbf{z}_{n+1} - \bar{\mathbf{x}})^\top \mathbf{\Gamma}^{-1} (\mathbf{z}_{n+1} - \bar{\mathbf{x}}) \right]$$

unde $\mathbf{\Gamma} = \frac{1}{n} \mathbf{Z}^\top \mathbf{Z} - \bar{\mathbf{x}} \bar{\mathbf{x}}^\top$ este o matrice simetrică și pozitiv definită atunci când $\mathbf{X}^\top \mathbf{X}$ este inversabilă iar $\bar{\mathbf{x}}^\top = (\bar{x}_1, \dots, \bar{x}_p)$ cu $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$.

Pentru început să observăm că matricea $\mathbf{X}^\top \mathbf{X}$ se scrie sub forma

$$\begin{aligned} \mathbf{X}^\top \mathbf{X} &= \begin{bmatrix} \mathbf{1}^\top \\ \mathbf{X}_1^\top \\ \vdots \\ \mathbf{X}_p^\top \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{X}_1 & \cdots & \mathbf{X}_p \end{bmatrix} = \begin{pmatrix} \underbrace{\mathbf{1}^\top \mathbf{1}}_{=n} & \underbrace{\mathbf{1}^\top \mathbf{X}_1}_{=n\bar{x}_1} & \cdots & \underbrace{\mathbf{1}^\top \mathbf{X}_p}_{=n\bar{x}_p} \\ \mathbf{X}_1^\top \mathbf{1} & \mathbf{X}_1^\top \mathbf{X}_1 & \cdots & \mathbf{X}_1^\top \mathbf{X}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_p^\top \mathbf{1} & \mathbf{X}_p^\top \mathbf{X}_1 & \cdots & \mathbf{X}_p^\top \mathbf{X}_p \end{pmatrix} = \begin{pmatrix} n & n\bar{x}_1 & \cdots & n\bar{x}_p \\ n\bar{x}_1 & \mathbf{X}_1^\top \mathbf{X}_1 & \cdots & \mathbf{X}_1^\top \mathbf{X}_p \\ \vdots & \vdots & \ddots & \vdots \\ n\bar{x}_p & \mathbf{X}_p^\top \mathbf{X}_1 & \cdots & \mathbf{X}_p^\top \mathbf{X}_p \end{pmatrix} \\ &= \begin{pmatrix} n & n\bar{\mathbf{x}}^\top \\ n\bar{\mathbf{x}} & \mathbf{Z}^\top \mathbf{Z} \end{pmatrix} = n \begin{pmatrix} 1 & \bar{\mathbf{x}}^\top \\ \bar{\mathbf{x}} & \frac{1}{n} \mathbf{Z}^\top \mathbf{Z} \end{pmatrix}. \end{aligned}$$

Reamintim, e.g. [Henderson and Searle, 1981], că dacă \mathbf{F} este o matrice pătrată inversabilă care se scrie sub formă de bloc de patru submatrice

$$\mathbf{F} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}$$

cu \mathbf{A} inversabilă atunci matricea $\mathbf{Q} = \mathbf{D} - \mathbf{C} \mathbf{A}^{-1} \mathbf{B}$ este inversabilă iar inversa matricii \mathbf{F} este

$$\mathbf{F}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1} \mathbf{B} \mathbf{Q} \mathbf{C} \mathbf{A}^{-1} & -\mathbf{A}^{-1} \mathbf{B} \mathbf{Q} \\ -\mathbf{Q} \mathbf{C} \mathbf{A}^{-1} & \mathbf{Q} \end{pmatrix}.$$

În cazul problemei noastre avem $\mathbf{A} = 1$, $\mathbf{B} = \bar{\mathbf{x}}^\top$, $\mathbf{C} = \bar{\mathbf{x}}$ și $\mathbf{D} = \frac{1}{n} \mathbf{Z}^\top \mathbf{Z}$ prin urmare

$$\mathbf{Q} = \mathbf{D} - \mathbf{C} \mathbf{A}^{-1} \mathbf{B} = \frac{1}{n} \mathbf{Z}^\top \mathbf{Z} - \bar{\mathbf{x}} \bar{\mathbf{x}}^\top = \mathbf{\Gamma}$$

iar

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \frac{1}{n} \begin{pmatrix} 1 & \bar{\mathbf{x}}^\top \\ \bar{\mathbf{x}} & \frac{1}{n} \mathbf{Z}^\top \mathbf{Z} \end{pmatrix}^{-1} = \frac{1}{n} \begin{pmatrix} 1 + \bar{\mathbf{x}}^\top \mathbf{\Gamma}^{-1} \bar{\mathbf{x}} & -\bar{\mathbf{x}}^\top \mathbf{\Gamma}^{-1} \\ -\mathbf{\Gamma}^{-1} \bar{\mathbf{x}} & \mathbf{\Gamma}^{-1} \end{pmatrix}.$$

Dat fiind $\mathbf{x}_{n+1}^\top = (1, \mathbf{z}_{n+1}^\top)$ am văzut în exercițiul precedent că varianța erorii de predicție este dată de formula

$$\text{Var}(\hat{\varepsilon}_{n+1}) = \sigma^2 (1 + \mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{n+1}).$$

Ținând cont de scrierea sub formă de blocuri a matricii $(\mathbf{X}^\top \mathbf{X})^{-1}$ găsim că

$$\begin{aligned}\mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{n+1} &= (1 \quad \mathbf{z}_{n+1}^\top) \frac{1}{n} \begin{pmatrix} 1 + \bar{\mathbf{x}}^\top \mathbf{\Gamma}^{-1} \bar{\mathbf{x}} & -\bar{\mathbf{x}}^\top \mathbf{\Gamma}^{-1} \\ -\mathbf{\Gamma}^{-1} \bar{\mathbf{x}} & \mathbf{\Gamma}^{-1} \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{z}_{n+1} \end{pmatrix} \\ &= \frac{1}{n} (1 + \bar{\mathbf{x}}^\top \mathbf{\Gamma}^{-1} \bar{\mathbf{x}} - \mathbf{z}_{n+1}^\top \mathbf{\Gamma}^{-1} \bar{\mathbf{x}} - \bar{\mathbf{x}}^\top \mathbf{\Gamma}^{-1} \mathbf{z}_{n+1} + \mathbf{z}_{n+1}^\top \mathbf{\Gamma}^{-1} \mathbf{z}_{n+1}).\end{aligned}$$

Notând cu \mathbf{Z}_c matricea (centrată) cu coloanele $\mathbf{X}_j - \bar{\mathbf{x}}_j \mathbf{1}$, $j = 1, \dots, p$ putem observa că

$$\mathbf{\Gamma} = \frac{1}{n} \mathbf{Z}^\top \mathbf{Z} - \bar{\mathbf{x}} \bar{\mathbf{x}}^\top = \frac{1}{n} \mathbf{Z}_c^\top \mathbf{Z}_c$$

prin urmare $\mathbf{\Gamma} = \mathbf{\Gamma}^\top$, deci $\mathbf{\Gamma}$ este simetrică. Astfel, folosind faptul că un scalar este egal cu transpusul său, avem

$$\mathbf{z}_{n+1}^\top \mathbf{\Gamma}^{-1} \bar{\mathbf{x}} = (\mathbf{z}_{n+1}^\top \mathbf{\Gamma}^{-1} \bar{\mathbf{x}})^\top = \bar{\mathbf{x}}^\top \mathbf{\Gamma}^{-1} \mathbf{z}_{n+1}$$

ceea ce conduce la

$$\begin{aligned}\mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{n+1} &= \frac{1}{n} (1 + \bar{\mathbf{x}}^\top \mathbf{\Gamma}^{-1} \bar{\mathbf{x}} - 2\bar{\mathbf{x}}^\top \mathbf{\Gamma}^{-1} \mathbf{z}_{n+1} + \mathbf{z}_{n+1}^\top \mathbf{\Gamma}^{-1} \mathbf{z}_{n+1}) \\ &= \frac{1}{n} [1 + (\mathbf{z}_{n+1} - \bar{\mathbf{x}})^\top \mathbf{\Gamma}^{-1} (\mathbf{z}_{n+1} - \bar{\mathbf{x}})]\end{aligned}$$

de unde găsim

$$Var(\hat{\varepsilon}_{n+1}) = \sigma^2 \left[1 + \frac{1}{n} + \frac{1}{n} (\mathbf{z}_{n+1} - \bar{\mathbf{x}})^\top \mathbf{\Gamma}^{-1} (\mathbf{z}_{n+1} - \bar{\mathbf{x}}) \right].$$

Mai mult, să remarcăm că dacă $\mathbf{X}^\top \mathbf{X}$ este inversabilă atunci matricea $\mathbf{\Gamma}$ este pozitiv definită. Într-adevăr am văzut că $\mathbf{\Gamma} = \frac{1}{n} \mathbf{Z}_c^\top \mathbf{Z}_c$ iar pentru $\mathbf{u} \in \mathbb{R}^p$ putem scrie

$$\mathbf{u}^\top \mathbf{\Gamma} \mathbf{u} = \frac{1}{n} \mathbf{u}^\top \mathbf{Z}_c^\top \mathbf{Z}_c \mathbf{u} = \frac{1}{n} \|\mathbf{Z}_c \mathbf{u}\|^2 \geq 0$$

deci $\mathbf{\Gamma}$ este pozitiv semidefinită. Avem egalitate $\frac{1}{n} \|\mathbf{Z}_c \mathbf{u}\|^2 = 0$ atunci când $\mathbf{Z}_c \mathbf{u} = 0$ sau, altfel spus, atunci când

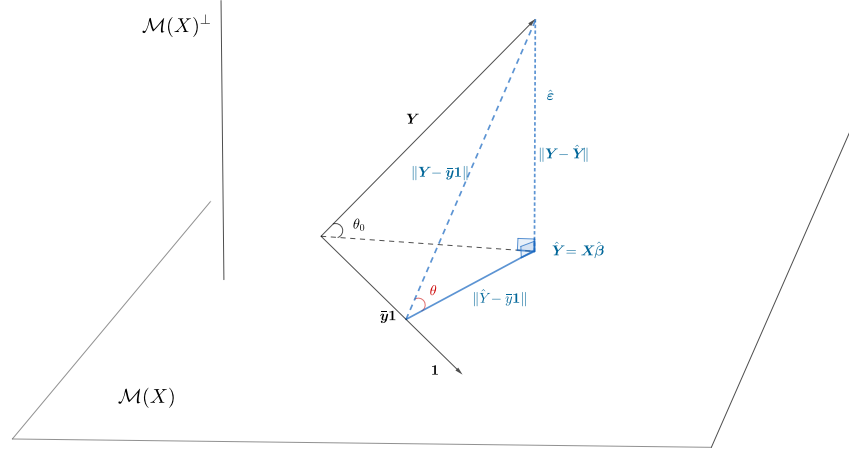
$$u_1(\mathbf{X}_1 - \bar{\mathbf{x}}_1 \mathbf{1}) + \dots + u_p(\mathbf{X}_p - \bar{\mathbf{x}}_p \mathbf{1}) = 0 \iff \sum_{j=1}^n u_j \mathbf{X}_j = \left(\sum_{j=1}^n u_j \bar{\mathbf{x}}_j \right) \mathbf{1}.$$

Această ultimă relație ne spune că prima coloană a lui \mathbf{X} se scrie ca o combinație liniară de celelalte, ceea ce înseamnă că $\text{rang}(\mathbf{X}) \leq p$ de unde concluzionăm că matricea $\mathbf{X}^\top \mathbf{X}$ nu este inversabilă.

Cum $\mathbf{\Gamma}$ este simetrică și pozitiv definită deducem că $(\mathbf{z}_{n+1} - \bar{\mathbf{x}})^\top \mathbf{\Gamma}^{-1} (\mathbf{z}_{n+1} - \bar{\mathbf{x}}) \geq 0$, cu egalitate dacă și numai dacă $\mathbf{z}_{n+1} = \bar{\mathbf{x}}$, adică $\mathbf{x}_{n+1}^\top = (1, \mathbf{z}_{n+1}^\top) = (1, \bar{\mathbf{x}}^\top)$. În acest caz $Var(\hat{\varepsilon}_{n+1}) = \sigma^2 (1 + \frac{1}{n})$, rezultat care generalizează rezultatul corespunzător de la regresia liniară simplă: trebuie să ne plasăm în centrul norului de puncte ale variabilelor explicative pentru a prezice variabila răspuns cu mai multă precizie.

2.3 Interpretări geometrice - coeficientul de determinare

Fie modelul de regresie $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ în care $\mathbf{Y} \in \mathbb{R}^n$ este vectorul răspuns, $\mathbf{X} = [\mathbf{1} | \mathbf{X}_1 | \dots | \mathbf{X}_p] \in \mathcal{M}_{n,p+1}(\mathbb{R})$ este matricea de design ($\text{rang}(\mathbf{X}) = p + 1$), $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$ este vectorul parametrilor și $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ este vectorul erorilor. Notăm cu $\mathcal{M}(\mathbf{X})$ subspațiul generat de coloanele matricii \mathbf{X} , cu $\mathcal{M}(\mathbf{X})^\perp$ subspațiul ortogonal, cu $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ proiecția ortogonală a lui \mathbf{Y} pe $\mathcal{M}(\mathbf{X})$ și cu $\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} \in \mathcal{M}(\mathbf{X})^\perp$ vectorul reziduurilor.



Conform figurii de mai sus observăm că putem aplica *Teorema lui Pitagora* în triunghiul albastru și obținem descompunerea

$$\|\mathbf{Y} - \bar{\mathbf{y}}\mathbf{1}\|^2 = \|\hat{\mathbf{Y}} - \bar{\mathbf{y}}\mathbf{1}\|^2 + \|\hat{\boldsymbol{\varepsilon}}\|^2$$

unde

- $SS_T = \|\mathbf{Y} - \bar{\mathbf{y}}\mathbf{1}\|^2$ se numește **suma abaterilor pătratice totale** (variația totală)
- $SS_{reg} = \|\hat{\mathbf{Y}} - \bar{\mathbf{y}}\mathbf{1}\|^2$ se numește **suma abaterilor pătratice de regresie** (variația explicată de model)
- $RSS = \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2 = \|\hat{\boldsymbol{\varepsilon}}\|^2$ se numește **suma abaterilor pătratice reziduale** (variația reziduală)

Coeficientul de determinare R^2 (în R are denumirea de **Multiple R-Squared**) este definit prin

$$R^2 = \cos^2(\theta) = \frac{\text{Variația explicată de model}}{\text{Variația totală}} = \frac{\|\hat{\mathbf{Y}} - \bar{\mathbf{y}}\mathbf{1}\|^2}{\|\mathbf{Y} - \bar{\mathbf{y}}\mathbf{1}\|^2} = 1 - \frac{\|\hat{\boldsymbol{\varepsilon}}\|^2}{\|\mathbf{Y} - \bar{\mathbf{y}}\mathbf{1}\|^2}$$

și ne dă proporția din variabilitatea lui \mathbf{Y} explicată prin modelul de regresie. Cum R^2 nu ține cont de dimensiunea spațiului $\mathcal{M}(\mathbf{X})$ se definește *coeficientul de determinare ajustat* R_a^2 (în R are denumirea de **Adjusted R-Squared**) prin

$$R_a^2 = 1 - \frac{\frac{\|\hat{\boldsymbol{\varepsilon}}\|^2}{n-(p+1)}}{\frac{\|\mathbf{Y} - \bar{\mathbf{y}}\mathbf{1}\|^2}{n-1}} = 1 - \frac{n-1}{n-(p+1)} \frac{\|\hat{\boldsymbol{\varepsilon}}\|^2}{\|\mathbf{Y} - \bar{\mathbf{y}}\mathbf{1}\|^2}.$$

Ajustarea corespunde la împărțirea normelor la pătrat cu dimensiunea subspațiului din care fac parte vectorii.

Ex. 2.8



Ne propunem să determinăm evoluția unei variabile răspuns y_i în funcție de două variabile explicative x_i și z_i . Fie $\mathbf{X} = (\mathbf{1}, \mathbf{x}, \mathbf{z})$ matricea de design.

1. Am obținut rezultatele următoare

$$\mathbf{X}^T \mathbf{X} = \begin{pmatrix} 25 & 0 & 0 \\ ? & 9.3 & 5.4 \\ ? & ? & 12.7 \end{pmatrix}$$

- Determinați valorile care lipsesc (?).
 - Cât este n ?
 - Calculați coeficientul de corelație dintre \mathbf{x} și \mathbf{z} .
2. Modelul de regresie liniară a lui \mathbf{Y} în funcție de $\mathbf{1}, \mathbf{x}, \mathbf{z}$ este

$$\mathbf{Y} = -1.6\mathbf{1} + 0.61\mathbf{x} + 0.46\mathbf{z} + \hat{\boldsymbol{\varepsilon}}, \quad \|\hat{\boldsymbol{\varepsilon}}\|^2 = 0.3$$

- Determinați $\bar{\mathbf{y}}$.
- Calculați suma abaterilor pătratice de regresie SS_{reg} , suma abaterilor pătratice totale SS_T , coeficientul de determinare și coeficientul de determinare ajustat.

- a) Pentru a determina cele trei valori care lipsesc să notăm că matricea $\mathbf{X}^T \mathbf{X}$ este simetrică, prin urmare $\mathbf{X}^T \mathbf{X} = (\mathbf{X}^T \mathbf{X})^T$ de unde

$$\begin{pmatrix} 25 & 0 & 0 \\ ? & 9.3 & 5.4 \\ ? & ? & 12.7 \end{pmatrix} = \begin{pmatrix} 25 & ? & ? \\ 0 & 9.3 & ? \\ 0 & 5.4 & 12.7 \end{pmatrix}$$

ceea ce conduce la valorile 0, 0, 5.4.

- Pentru a determina valoarea lui n să observăm că

$$\mathbf{X}^T \mathbf{X} = \begin{pmatrix} \mathbf{1}^T \\ \mathbf{x}^T \\ \mathbf{z}^T \end{pmatrix} (\mathbf{1} \quad \mathbf{x} \quad \mathbf{z}) = \begin{pmatrix} \mathbf{1}^T \mathbf{1} & \mathbf{1}^T \mathbf{x} & \mathbf{1}^T \mathbf{z} \\ \mathbf{x}^T \mathbf{1} & \mathbf{x}^T \mathbf{x} & \mathbf{x}^T \mathbf{z} \\ \mathbf{z}^T \mathbf{1} & \mathbf{z}^T \mathbf{x} & \mathbf{z}^T \mathbf{z} \end{pmatrix} = \begin{pmatrix} n & n\bar{x} & n\bar{z} \\ n\bar{x} & \mathbf{x}^T \mathbf{x} & \mathbf{x}^T \mathbf{z} \\ n\bar{z} & \mathbf{z}^T \mathbf{x} & \mathbf{z}^T \mathbf{z} \end{pmatrix}$$

ceea ce arată că $n = (\mathbf{X}^T \mathbf{X})_{1,1} = 25$.

- Coeficientul de corelație dintre \mathbf{x} și \mathbf{z} este dat de

$$r_{\mathbf{x}, \mathbf{z}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(z_i - \bar{z})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (z_i - \bar{z})^2}}$$

și cum, din punctul anterior, $\bar{x} = \frac{(\mathbf{X}^T \mathbf{X})_{1,2}}{n} = 0$ și $\bar{z} = \frac{(\mathbf{X}^T \mathbf{X})_{1,3}}{n} = 0$ deci

$$r_{\mathbf{x}, \mathbf{z}} = \frac{\sum_{i=1}^n x_i z_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n z_i^2}} = \frac{\mathbf{x}^T \mathbf{z}}{\sqrt{\mathbf{x}^T \mathbf{x}} \sqrt{\mathbf{z}^T \mathbf{z}}} = \frac{(\mathbf{X}^T \mathbf{X})_{2,3}}{\sqrt{(\mathbf{X}^T \mathbf{X})_{2,2}} \sqrt{(\mathbf{X}^T \mathbf{X})_{3,3}}} = \frac{5.4}{\sqrt{9.3} \sqrt{12.7}} \approx 0.5$$

- a) Avem modelul

$$y_i = -1.6 + 0.61x_i + 0.46z_i + \hat{\varepsilon}_i$$

care prin sumare devine

$$\bar{y} = -1.6 + 0.61\bar{x} + 0.46\bar{z} + \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i.$$

Cum $\mathbf{1}$ aparține modelului, deci $\mathbf{1} \in \mathcal{M}(\mathbf{X})$ și ținând cont de faptul că $\hat{\varepsilon} \in \mathcal{M}(\mathbf{X})^\perp$ deducem că $\langle \mathbf{1}, \hat{\varepsilon} \rangle = 0$ sau $\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i = 0$. Astfel

$$\bar{y} = -1.6 + 0.61 \underbrace{\bar{x}}_{=0} + 0.46 \underbrace{\bar{z}}_{=0} + \underbrace{\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i}_{=0} = -1.6.$$

b) Avem că suma abaterilor pătratice explicate de model este dată de

$$SS_{reg} = \|\hat{\mathbf{Y}} - \bar{y}\mathbf{1}\|^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^n (0.61x_i + 0.46z_i)^2$$

adică

$$\begin{aligned} SS_{reg} &= \|\hat{\mathbf{Y}} - \bar{y}\mathbf{1}\|^2 = 0.61^2 \sum_{i=1}^n x_i^2 + 2 \times 0.61 \times 0.46 \sum_{i=1}^n x_i z_i + 0.46^2 \sum_{i=1}^n z_i^2 \\ &= 0.61^2 (\mathbf{X}^\top \mathbf{X})_{2,2} + 2 \times 0.61 \times 0.46 (\mathbf{X}^\top \mathbf{X})_{2,3} + 0.46^2 (\mathbf{X}^\top \mathbf{X})_{3,3} = 9.18. \end{aligned}$$

Suma abaterilor pătratice totale se determină folosind formula de descompunere a varianței:

$$\underbrace{SS_T}_{\|\mathbf{Y} - \bar{y}\mathbf{1}\|^2} = \underbrace{SS_{reg}}_{\|\hat{\mathbf{Y}} - \bar{y}\mathbf{1}\|^2} + \underbrace{RSS}_{\|\hat{\varepsilon}\|^2} = 9.18 + 0.3 = 9.48.$$

Coeficientul de determinare R^2 este

$$R^2 = \frac{SS_{reg}}{SS_T} = \frac{9.18}{9.48} \approx 0.968$$

ceea ce arată că aproximativ 97% din variabilitatea datelor este explicată de modelul de regresie iar coeficientul de determinare ajustat este

$$R_1^2 = 1 - \frac{n-1}{n-(p+1)} \frac{\|\hat{\varepsilon}\|^2}{\|\mathbf{Y} - \bar{y}\mathbf{1}\|^2} = 1 - \frac{n-1}{n-(p+1)} (1 - R^2) = 1 - \frac{25-1}{25-(2+1)} (1 - 0.968) \approx 0.965$$

ceea ce verifică relația generală $R_a^2 < R^2$.

Ex. 2.9



Fie $Z \in \mathcal{M}_{n,q}(\mathbb{R})$ o matrice de rang q și $X \in \mathcal{M}_{n,p}(\mathbb{R})$ o matrice de rang p , $p > q$, compusă din cei q vectori coloană ai lui Z și din alți $p - q$ vectori liniar independenți. Considerăm modelele următoare:

$$\begin{aligned} Y &= Z\beta + \varepsilon, \\ Y &= X\gamma + \eta \end{aligned}$$

Presupunem, pentru simplitate, că niciun model nu conține termenul liber (constanta). Notăm cu P_X și P_Z matricele de proiecție ortogonală pe subspațiile $\mathcal{M}(X)$ și respectiv $\mathcal{M}(Z)$, generate de coloanele lui X și Z . De asemenea notăm cu $P_{X \cap Z^\perp}$ matricea de proiecție ortogonală pe $\mathcal{M}(X) \cap \mathcal{M}(Z)^\perp$, subspațiul ortogonal al lui $\mathcal{M}(Z)$ din $\mathcal{M}(X)$, astfel

$$\mathbb{R}^n = \mathcal{M}(X) \oplus \mathcal{M}(X)^\perp = (\mathcal{M}(Z) \oplus (\mathcal{M}(X) \cap \mathcal{M}(Z)^\perp)) \oplus \mathcal{M}(X)^\perp$$

1. Exprimați $\|P_X Y\|^2$ în funcție de $\|P_Z Y\|^2$ și de $\|P_{X \cap Z^\perp} Y\|^2$.
2. Comparați coeficienții de determinare a celor două modele, R_Z^2 și R_X^2 .

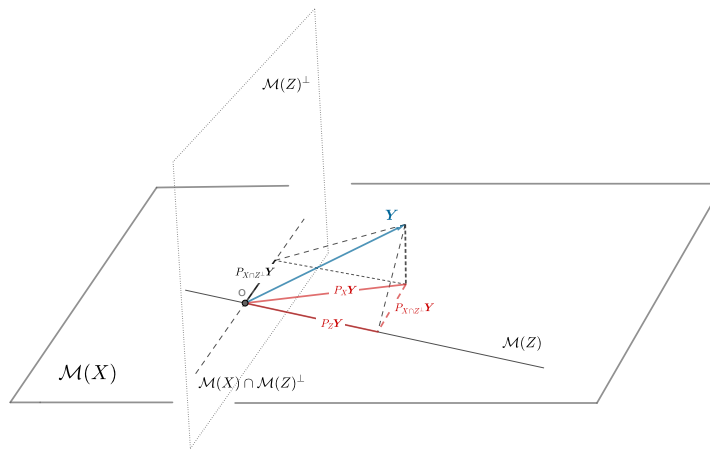
Acest exercițiu face referire la modele imbricate - *nested models*.

1. Deoarece $\mathcal{M}(Z) \subset \mathcal{M}(X)$ și aplicând *teorema celor trei perpendiculare*² avem, vezi figura de mai jos,

$$\begin{aligned} \hat{Y}_p &= P_X Y = \underbrace{(P_Z + P_{Z^\perp})}_{I} P_X Y = P_Z P_X Y + P_{Z^\perp} P_X Y \\ &= P_X Y + P_{X \cap Z^\perp} Y = \hat{Y}_q + P_{X \cap Z^\perp} Y \end{aligned}$$

și din Teorema lui Pitagora (vezi triunghiul roșu) deducem că

$$\|P_X Y\|^2 = \|P_Z Y\|^2 + \|P_{X \cap Z^\perp} Y\|^2.$$



² Teorema celor trei perpendiculare spune că dacă V și W sunt două subspații vectoriale astfel încât $V \subset W$ atunci $P_V P_W = P_W P_V = P_V$.

2. Dacă modelul nu are termenul liber (constanta) atunci coeficientul de determinare este definit prin $R_X^2 = \frac{\|P_X \mathbf{Y}\|^2}{\|\mathbf{Y}\|^2}$ prin urmare

$$R_X^2 = \frac{\|P_X \mathbf{Y}\|^2}{\|\mathbf{Y}\|^2} = \frac{\|P_Z \mathbf{Y}\|^2 + \|P_{X \cap Z^\perp} \mathbf{Y}\|^2}{\|\mathbf{Y}\|^2} = R_Z^2 + \frac{\|P_{X \cap Z^\perp} \mathbf{Y}\|^2}{\|\mathbf{Y}\|^2} \geq R_Z^2.$$

Rezultatul de mai sus se generalizează și în cazul modelelor imbricate care conțin termenul constant și ne spune arată că modelul mai mare are un coeficient de determinare superior modelului mai mic. Altfel spus, dacă la un model dat adăugăm una sau mai multe variabile explicative ameliorăm variabilitatea explicată de model, chiar dacă variabilele explicative suplimentare nu sunt pertinente! În acest caz este de preferat coeficientul de determinare ajustat.

Ex. 2.10



Considerăm modelul de regresie liniară

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

unde $\mathbf{Y} \in \mathbb{R}^n$, $\mathbf{X} \in \mathcal{M}_{n,p}(\mathbb{R})$ este o matrice compusă din p vectori ortogonali, $\boldsymbol{\beta} \in \mathbb{R}^p$ iar $\boldsymbol{\varepsilon} \in \mathbb{R}^n$. Fie \mathbf{Z} matricea formată din primele q coloane ale lui \mathbf{X} și \mathbf{U} matricea formată din ultimele $p - q$ coloane ale lui \mathbf{X} . Prin metoda celor mai mici pătrate obținem

$$\hat{\mathbf{Y}}_X = \hat{\beta}_1^X \mathbf{X}_1 + \cdots + \hat{\beta}_p^X \mathbf{X}_p$$

$$\hat{\mathbf{Y}}_Z = \hat{\beta}_1^Z \mathbf{X}_1 + \cdots + \hat{\beta}_q^Z \mathbf{X}_q$$

$$\hat{\mathbf{Y}}_U = \hat{\beta}_{q+1}^U \mathbf{X}_{q+1} + \cdots + \hat{\beta}_p^U \mathbf{X}_p$$

1. Arătați că $\|P_X \mathbf{Y}\|^2 = \|P_Z \mathbf{Y}\|^2 + \|P_U \mathbf{Y}\|^2$.
2. Pentru $i \in \{1, 2, \dots, p\}$ dat, arătați că $\hat{\beta}_i^X = \hat{\beta}_i^Z$ dacă $i \leq q$ și $\hat{\beta}_i^X = \hat{\beta}_i^U$ altfel.

1. Cum $P_Z + P_{Z^\perp} = I$ putem scrie

$$\hat{\mathbf{Y}}_X = P_X \mathbf{Y} = (P_Z + P_{Z^\perp})P_X \mathbf{Y} = P_Z P_X \mathbf{Y} + P_{Z^\perp} P_X \mathbf{Y},$$

și din $P_Z P_X = P_{Z \cap X} = P_Z$ găsim că $\hat{\mathbf{Y}}_X = P_Z \mathbf{Y} + P_{Z^\perp} P_X \mathbf{Y}$. De asemenea, să notăm că $P_{Z^\perp} P_X = P_{Z^\perp \cap X}$ și ținând cont de faptul că matricea \mathbf{X} are coloanele ortogonale avem $P_{Z^\perp \cap X} = P_U$ (în fapt ortogonalitatea coloanelor lui \mathbf{X} implică $\mathcal{M}(X) = \mathcal{M}(Z) \oplus \mathcal{M}(U)$, sumă directă de spații ortogonale). Prin urmare obținem descompunerea ortogonală

$$\hat{\mathbf{Y}}_X = P_Z \mathbf{Y} + P_U \mathbf{Y} = \hat{\mathbf{Y}}_Z + \hat{\mathbf{Y}}_U$$

și din Teorema lui Pitagora avem

$$\|P_X \mathbf{Y}\|^2 = \|P_Z \mathbf{Y}\|^2 + \|P_U \mathbf{Y}\|^2.$$

2. Vom arăta relația pentru $i \leq q$, cazul general fiind analog. Din formula generală avem că $\hat{\beta}^X$ este

$$\hat{\beta}^X = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$$

și cum coloanele lui \mathbf{X} sunt ortogonale deducem că matricea $\mathbf{X}^\top \mathbf{X}$ este o matrice diagonală, $\mathbf{X}^\top \mathbf{X} = \text{diag}(\|\mathbf{X}_1\|^2, \dots, \|\mathbf{X}_p\|^2)$. Notând de asemenea că $\mathbf{X}^\top \mathbf{Y}$ este un vector coloană cu elemente de tipul $\langle \mathbf{X}_i, \mathbf{Y} \rangle$, deducem că

$$\hat{\beta}^X = \left(\frac{\langle \mathbf{X}_1, \mathbf{Y} \rangle}{\|\mathbf{X}_1\|^2} \quad \cdots \quad \frac{\langle \mathbf{X}_p, \mathbf{Y} \rangle}{\|\mathbf{X}_p\|^2} \right)^\top,$$

prin urmare $\hat{\beta}_i^X = \frac{\langle \mathbf{X}_i, \mathbf{Y} \rangle}{\|\mathbf{X}_i\|^2}$. Pentru $i \leq q$, coloana i a matricei \mathbf{Z} este $\mathbf{Z}_i = \mathbf{X}_i$ și aplicând raționamentul anterior găsim că

$$\hat{\beta}_i^Z = \frac{\langle \mathbf{Z}_i, \mathbf{Y} \rangle}{\|\mathbf{Z}_i\|^2} = \frac{\langle \mathbf{X}_i, \mathbf{Y} \rangle}{\|\mathbf{X}_i\|^2} = \hat{\beta}_i^X.$$

Rezultatul exercițiului ne spune că în cazul în care variabilele explicative sunt ortogonale, a efectua o regresie multiplă revine la a efectua p regresii simple.

2.4 Modelul (condiționat) normal

Considerăm modelul de regresie $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$ în care $\mathbf{Y} \in \mathbb{R}^n$ este vectorul răspuns, $\mathbf{X} = [\mathbf{1} | \mathbf{X}_1 | \cdots | \mathbf{X}_p] \in \mathcal{M}_{n,p+1}(\mathbb{R})$ este matricea de design, $\beta \in \mathbb{R}^{p+1}$ este vectorul parametrilor și $\varepsilon \in \mathbb{R}^n$ este vectorul erorilor. Până acum, ipotezele modelului erau

$$\begin{aligned} \mathcal{H}_1 &: \text{rang}(\mathbf{X}) = p + 1 \\ \mathcal{H}_2 &: \mathbb{E}[\varepsilon] = 0, \text{Var}(\varepsilon) = \sigma^2 I_n \end{aligned}$$

În această secțiune ne propunem să studiem proprietățile statistice ale modelului de regresie liniară plasându-ne într-un context parametric și anume în contextul modelului gaussian:

$$\begin{aligned} \mathcal{H}_1 &: \text{rang}(\mathbf{X}) = p + 1 \\ \mathcal{H}'_2 &: \varepsilon \sim \mathcal{N}(0, \sigma^2 I_n) \end{aligned}$$

Observăm că ipoteza \mathcal{H}'_2 este un caz particular al ipotezei \mathcal{H}_2 și în plus implică faptul că reziduurile sunt independente și identic repartizate. Ipoteza de normalitate a erorilor ne permite să scriem funcția de verosimilitate, să deducem repartițiile estimatorilor propuși, să construim intervale de încredere și să testăm ipoteze statistice.

Ex. 2.11



Considerăm modelul de regresie liniară

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon$$

sub ipotezele \mathcal{H}_1 și \mathcal{H}'_2 de mai sus. Determinați estimatorii de verosimilitate maximă pentru β și ε .

Începem prin a observa că din modelul de regresie

$$y_i = \mathbf{x}_i^\top \beta + \varepsilon_i$$

și conform ipotezei \mathcal{H}'_2 avem $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ prin urmare $y_i \sim \mathcal{N}(\mathbf{x}_i^\top \beta, \sigma^2)$ și în plus y_i sunt variabile aleatoare independente deoarece ε_i sunt independente. Astfel funcția de verosimilitate se scrie

$$\begin{aligned} L(\beta, \sigma^2; \mathbf{Y}) &= \prod_{i=1}^n f_Y(y_i) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mathbf{x}_i^\top \beta)^2} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2} \|\mathbf{Y} - \mathbf{X}\beta\|^2} \end{aligned}$$

ceea ce conduce la logaritmul funcției de verosimilitate

$$l(\beta, \sigma^2; \mathbf{Y}) = \log L(\beta, \sigma^2; \mathbf{Y}) = -\frac{n}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \|\mathbf{Y} - \mathbf{X}\beta\|^2.$$

Vrem să determinăm

$$(\hat{\beta}_{VM}, \hat{\sigma}_{VM}^2) = \arg \max_{(\beta, \sigma^2) \in \mathbb{R}^{p+1} \times (0, \infty)} l(\beta, \sigma^2; \mathbf{Y})$$

și observăm că pentru σ^2 fixat, a maximiza funcția $l(\beta, \sigma^2; \mathbf{Y})$ revine la a determina valoarea lui β care minimizează $\|\mathbf{Y} - \mathbf{X}\beta\|^2$, aceasta ne fiind alta decât $\hat{\beta}$ obținut prin metoda celor mai mici pătrate. Astfel găsim că

$$\hat{\beta}_{VM} = \hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}.$$

Pentru a determina $\hat{\sigma}_{VM}^2$ să observăm că problema revine la a determina soluția ecuației

$$\frac{\partial l(\hat{\beta}_{VM}, \sigma^2; \mathbf{Y})}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \|\mathbf{Y} - \mathbf{X}\hat{\beta}_{VM}\|^2 = 0$$

care este $\hat{\sigma}_{VM}^2 = \frac{\|\mathbf{Y} - \mathbf{X}\hat{\beta}_{VM}\|^2}{n} = \frac{\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2}{n}$. Comparând acest rezultat cu cel obținut prin metoda celor mai mici pătrate observăm că

$$\hat{\sigma}_{VM}^2 = \frac{n - (p + 1)}{n} \hat{\sigma}^2,$$

deci estimatorul de verosimilitate maximă $\hat{\sigma}_{VM}^2$ a lui σ^2 est un estimator deplasat. În cele ce urmează vom considera estimatorul nedepășat $\hat{\sigma}^2 = \frac{\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2}{n - (p + 1)} = \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{n - (p + 1)}$ ca estimator al lui σ^2 .

2.4.1 Repartițiile estimatorilor, intervale și regiuni de încredere

Înainte de a investiga repartițiile estimatorilor $\hat{\beta}$ și $\hat{\sigma}^2$ pentru modelul de regresie liniară sub ipoteza de normalitate, să reamintim câteva noțiuni legate de vectorii gaussieni (pentru mai multe detalii se poate consulta [Seber and Lee, 2003, Capitolul 2] sau [Jacod and Protter, 2003, Capitolul 16]).

Spunem că un vector $\mathbf{Y} = (Y_1, \dots, Y_n) \in \mathbb{R}^n$ este un vector gaussian dacă toate combinațiile liniare $\sum_{i=1}^n a_i Y_i$ sunt repartizate normal (posibil degenerat dacă toți coeficienții sunt nuli). Acest vector admite o medie $\boldsymbol{\mu} = \mathbb{E}[\mathbf{Y}]$ și o matrice de varianță-covarianță $\Sigma_{\mathbf{Y}} = \mathbb{E}[(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})^\top]$ care caracterizează complet repartiția lui \mathbf{Y} . În acest caz notăm $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma_{\mathbf{Y}})$. Un vector gaussian \mathbf{Y} admite o densitate de repartiție $f_{\mathbf{Y}}$ pe \mathbb{R}^n dacă și numai dacă matricea sa de varianță-covarianță $\Sigma_{\mathbf{Y}}$ este inversabilă și în acest caz

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \Sigma_{\mathbf{Y}}}} e^{-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^\top \Sigma_{\mathbf{Y}}^{-1}(\mathbf{y} - \boldsymbol{\mu})}$$

În cazul în care matricea $\Sigma_{\mathbf{Y}}$ nu este inversabilă înseamnă că \mathbf{Y} ia valori într-un subspațiu de dimensiune $n_0 < n$ unde este repartizat ca un vector gaussian de dimensiune n_0 .

Figura de mai jos prezintă patru repartiții normale bivariate ($n = 2$) $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma_{\mathbf{Y}})$, cu $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top$ iar $\Sigma_{\mathbf{Y}} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$, rotite cu unghiul α . În acest caz densitatea de repartiție este

$$f_{\mathbf{Y}}(y_1, y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[\left(\frac{y_1-\mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{y_1-\mu_1}{\sigma_1} \right) \left(\frac{y_2-\mu_2}{\sigma_2} \right) + \left(\frac{y_2-\mu_2}{\sigma_2} \right)^2 \right]}$$

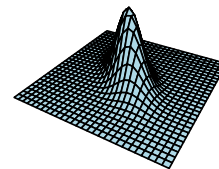
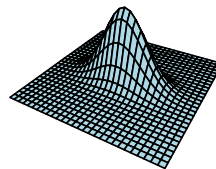
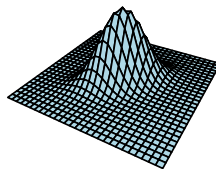
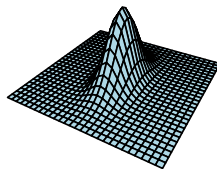
unde $\rho = \frac{\sigma_{12}}{\sigma_1\sigma_2}$ este coeficientul de corelație.

$$\alpha = 0$$

$$\alpha = \frac{\pi}{4}$$

$$\alpha = \frac{\pi}{2}$$

$$\alpha = \frac{3\pi}{2}$$



Una dintre proprietățile importante ale vectorilor gaussieni este stabilitatea prin transformări afine: dacă $\mathbf{A} \in \mathcal{M}_{m,n}(\mathbb{R})$ și $\mathbf{b} \in \mathcal{M}_{m,1}(\mathbb{R})$ sunt respectiv o matrice și un vector de scalari atunci

$$\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma_{\mathbf{Y}}) \implies \mathbf{AY} + \mathbf{b} \sim \mathcal{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\Sigma_{\mathbf{Y}}\mathbf{A}^T)$$

De asemenea, plecând de la funcția caracteristică, se poate verifica proprietatea de independență a componentelor unui vector gaussian care ne spune că acestea sunt independente dacă și numai dacă matricea de varianță-covarianță este o matrice diagonală.

Avem următorul rezultat care face legătura dintre repartiția normală și repartiția χ^2 :



Fie $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma_{\mathbf{Y}})$ un vector gaussian în \mathbb{R}^n . Dacă $\Sigma_{\mathbf{Y}}$ este o matrice inversabilă atunci

$$(\mathbf{Y} - \boldsymbol{\mu})^T \Sigma_{\mathbf{Y}}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \sim \chi_n^2$$

repartiția χ^2 cu n grade de libertate.

Deoarece matricea de varianță-covarianță $\Sigma_{\mathbf{Y}}$ este o matrice simetrică și pozitiv definită atunci din **Teorema de descompunere spectrală** ea se poate descompune sub forma $\Sigma_{\mathbf{Y}} = \mathbf{Q}\mathbf{\Delta}\mathbf{Q}^T$ unde $\mathbf{Q} = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ este o matrice ortogonală ($\mathbf{Q}^T = \mathbf{Q}^{-1}$) formată din vectorii proprii $\mathbf{v}_1, \dots, \mathbf{v}_n$ corespunzători valorilor proprii $\lambda_1, \dots, \lambda_n$, iar $\mathbf{\Delta}$ este matricea diagonală $\text{diag}(\lambda_1, \dots, \lambda_n)$. Dacă notăm cu $\mathbf{\Delta}^{-\frac{1}{2}}$ matricea diagonală de coeficienți diagonali $\frac{1}{\sqrt{\lambda_1}}, \dots, \frac{1}{\sqrt{\lambda_n}}$ (aceste fracții există deoarece pozitiv definirea matricei $\Sigma_{\mathbf{Y}}$ implică faptul că valorile proprii $\lambda_i > 0$, $i = 1, \dots, n$) atunci

$$\Sigma_Y = Q\Delta Q^\top \implies \Sigma_Y^{-1} = Q\Delta^{-1}Q^\top = (Q\Delta^{-\frac{1}{2}}Q^\top)(Q\Delta^{-\frac{1}{2}}Q^\top) = \Sigma_Y^{-\frac{1}{2}}\Sigma_Y^{-\frac{1}{2}}.$$

Prin urmare găsim că

$$(Y - \mu)^\top \Sigma_Y^{-1} (Y - \mu) = \left(\Sigma_Y^{-\frac{1}{2}} (Y - \mu) \right)^\top \left(\Sigma_Y^{-\frac{1}{2}} (Y - \mu) \right)$$

și cum vectorii gaussieni rămân gaussieni și prin aplicarea unor transformări afine (vezi rezultatul de mai sus) avem că

$$Y \sim \mathcal{N}(\mu, \Sigma_Y) \implies \Sigma_Y^{-\frac{1}{2}}(Y - \mu) \sim \mathcal{N}(0, I_n).$$

Astfel vectorul $V = [V_1, \dots, V_n]^\top = \Sigma_Y^{-\frac{1}{2}}(Y - \mu)$, care nu este altceva decât vectorul Y centrat și redus, este gaussian standard ($V_j \sim \mathcal{N}(0, 1)$ și V_i și V_j sunt independente) și

$$(Y - \mu)^\top \Sigma_Y^{-1} (Y - \mu) = \|V\|^2 = V_1^2 + \dots + V_n^2 \sim \chi_n^2.$$

Următorul rezultat, cunoscut sub denumirea de **Teorema lui Cochran**, asigură că descompunerea unui vector gaussian în componente independente pe subspații ortogonale ne dă vectori independenți a căror repartiție o putem explicita. Această teoremă poate fi văzută și ca o generalizare a Teoremei lui Pitagora:



Fie $Y \sim \mathcal{N}(\mu, \sigma^2 I_n)$, $\mathcal{M} \subset \mathbb{R}^n$ un subspațiu de dimensiune p , P matricea de proiecție ortogonală pe \mathcal{M} și $P_\perp = I_n - P$ matricea de proiecție ortogonală pe \mathcal{M}^\perp . Au loc următoarele proprietăți:

1. $PY \sim \mathcal{N}(P\mu, \sigma^2 P)$ și $P_\perp Y \sim \mathcal{N}(P_\perp \mu, \sigma^2 P_\perp)$
2. vectorii PY și $P_\perp Y = Y - PY$ sunt independenți
3. $\frac{\|P(Y-\mu)\|^2}{\sigma^2} \sim \chi_p^2$ și $\frac{\|P_\perp(Y-\mu)\|^2}{\sigma^2} \sim \chi_{n-p}^2$.

Acum avem instrumentele necesare pentru a deduce repartițiile estimatorilor în modelul gaussian de regresie liniară.

Ex. 2.12



Considerăm modelul de regresie liniară $Y = X\beta + \varepsilon$, sub ipotezele \mathcal{H}_1 și \mathcal{H}'_2 . Dacă presupunem că varianța σ^2 este cunoscută atunci au loc proprietățile

1. Vectorul $\hat{\beta}$ este un vector gaussian de medie β și matrice de varianță-covarianță $\sigma^2(X^\top X)^{-1}$, i.e. $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2(X^\top X)^{-1})$
2. $\hat{\beta}$ și $\hat{\sigma}^2$ sunt independenți
3. $[n - (p + 1)] \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-(p+1)}^2$

1. Pentru a arăta că vectorul $\hat{\beta}$ este un vector gaussian vom folosi expresia acestuia obținută prin metoda celor mai mici pătrate

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y = (X^\top X)^{-1} X^\top (X\beta + \varepsilon) = \beta + (X^\top X)^{-1} X^\top \varepsilon.$$

Conform ipotezei \mathcal{H}'_2 , $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$ este un vector gaussian prin urmare și $\hat{\beta} = \beta + \underbrace{(X^\top X)^{-1} X^\top}_{A} \varepsilon$ este tot un vector gaussian repartizat

$$\hat{\beta} = \beta + \underbrace{(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top}_{\mathbf{A}} \varepsilon = \beta + \mathbf{A} \varepsilon \sim \mathcal{N}(\beta, \mathbf{A} \sigma^2 I_n \mathbf{A}^\top) = \mathcal{N}(\beta, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1})$$

deoarece $\mathbf{A} \sigma^2 I_n \mathbf{A}^\top = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \sigma^2 I_n \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$.

2. Fie $\mathcal{M}(X)$ subspațiul lui \mathbb{R}^n generat de coloanele matricei \mathbf{X} și fie $P_X = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ matricea de proiecție ortogonală pe $\mathcal{M}(X)$. Avem

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \left(\underbrace{\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top}_{=P_X} \right) \mathbf{Y} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top P_X \mathbf{Y}$$

prin urmare $\hat{\beta}$ este un vector aleator ce depinde de $P_X \mathbf{Y}$. Observăm de asemenea că din definiția lui $\hat{\sigma}^2$,

$$\hat{\sigma}^2 = \frac{\|\hat{\varepsilon}\|^2}{n - (p + 1)} = \frac{\|\mathbf{Y} - P_X \mathbf{Y}\|^2}{n - (p + 1)}$$

acesta este o funcție de $\mathbf{Y} - P_X \mathbf{Y}$. Aplicând Teorema lui Cochran (de mai sus) observăm că vectorii $P_X \mathbf{Y}$ și $\mathbf{Y} - P_X \mathbf{Y}$ sunt independenți prin urmare și $\hat{\beta}$ și $\hat{\sigma}^2$ sunt independenți ca funcții de vectori independenți.

3. Notând cu P_{X^\perp} proiecția ortogonală pe subspațiul ortogonal $\mathcal{M}(X)^\perp$, subspațiu de dimensiune $n - (p + 1)$, avem

$$\hat{\varepsilon} = \mathbf{Y} - P_X \mathbf{Y} = P_{X^\perp} \mathbf{Y} = P_{X^\perp} \left(\underbrace{\mathbf{X} \beta}_{\in \mathcal{M}(X)} + \varepsilon \right) = P_{X^\perp} \varepsilon,$$

unde $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$.

Din Teorema lui Cochran găsim că

$$(n - (p + 1)) \frac{\hat{\sigma}^2}{\sigma^2} = \frac{\|P_{X^\perp} \varepsilon\|^2}{\sigma^2} = \frac{\|P_{X^\perp} (\varepsilon - \mathbb{E}[\varepsilon])\|^2}{\sigma^2} \sim \chi_{n-(p+1)}^2.$$

Ex. 2.13



Considerăm modelul de regresie liniară $\mathbf{Y} = \mathbf{X} \beta + \varepsilon$, sub ipotezele \mathcal{H}_1 și \mathcal{H}'_2 și presupunem că varianța σ^2 nu este cunoscută. Atunci au loc proprietățile

1. pentru $j \in \{0, 1, \dots, p\}$, notând cu $[(\mathbf{X}^\top \mathbf{X})^{-1}]_{j+1, j+1}$ elementul $j + 1$ de pe diagonala matricei $(\mathbf{X}^\top \mathbf{X})^{-1}$, avem

$$\hat{\beta}_j \sim \mathcal{N}(\beta_j, \sigma^2 [(\mathbf{X}^\top \mathbf{X})^{-1}]_{j+1, j+1}), \quad T_j = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma} \sqrt{[(\mathbf{X}^\top \mathbf{X})^{-1}]_{j+1, j+1}}} = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim t_{n-(p+1)}$$

2. fie $R \in \mathcal{M}_{q, p+1}(\mathbb{R})$ o matrice de rang q ($q \leq p + 1$) atunci

$$\frac{1}{q \hat{\sigma}^2} \left(R(\hat{\beta} - \beta) \right)^\top [R(\mathbf{X}^\top \mathbf{X})^{-1} R^\top]^{-1} R(\hat{\beta} - \beta) \sim F_{q, n-(p+1)}$$

1. Din exercițiul 2.12 am văzut că $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1})$ prin urmare

$$\hat{\beta}_j \sim \mathcal{N}(\beta_j, \sigma^2 [(\mathbf{X}^\top \mathbf{X})^{-1}]_{j+1, j+1})$$

ceea ce implică $\frac{\hat{\beta}_j - \beta_j}{\sigma \sqrt{[(\mathbf{X}^\top \mathbf{X})^{-1}]_{j+1, j+1}}} \sim \mathcal{N}(0, 1)$.

Scriind

$$T_j = \frac{\frac{\hat{\beta}_j - \beta_j}{\sigma \sqrt{[(\mathbf{X}^\top \mathbf{X})^{-1}]_{j+1, j+1}}}}{\frac{\hat{\sigma}}{\sigma}} = \frac{\frac{\hat{\beta}_j - \beta_j}{\sigma \sqrt{[(\mathbf{X}^\top \mathbf{X})^{-1}]_{j+1, j+1}}}}{\sqrt{\frac{(n-(p+1)) \frac{\hat{\sigma}^2}{\sigma^2}}{n-(p+1)}}}$$

și ținând cont de faptul că $(n-(p+1)) \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-(p+1)}$ iar $\hat{\beta}_j$ și $\hat{\sigma}^2$ sunt independente, deducem că $T_j \sim t_{n-(p+1)}$.

2. Să remarcăm că matricea $R(\mathbf{X}^\top \mathbf{X})^{-1} R^\top$ pătratică de ordin q este inversabilă deoarece matricea $(\mathbf{X}^\top \mathbf{X})^{-1}$ este de rang $p+1$, $p+1 \geq q$. Cum $\hat{\beta}$ este un vector gaussian deducem că și $R\hat{\beta}$ este un vector gaussian de medie și matrice de covarianță

$$R\hat{\beta} \sim \mathcal{N}(R\beta, \sigma^2 R(\mathbf{X}^\top \mathbf{X})^{-1} R^\top)$$

Astfel (a se vedea rezultatul privind legătura dintre repartiția normală și repartiția χ^2 de mai sus)

$$\frac{1}{\sigma^2} (R(\hat{\beta} - \beta))^\top [R(\mathbf{X}^\top \mathbf{X})^{-1} R^\top]^{-1} R(\hat{\beta} - \beta) \sim \chi_q^2$$

În expresia de mai sus, înlocuim pe σ^2 cu $\hat{\sigma}^2$ și ținând cont că $\hat{\beta}$ și $\hat{\sigma}^2$ sunt independente iar $(n-(p+1)) \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-(p+1)}$ concluzionăm că

$$\frac{1}{q \hat{\sigma}^2} (R(\hat{\beta} - \beta))^\top [R(\mathbf{X}^\top \mathbf{X})^{-1} R^\top]^{-1} R(\hat{\beta} - \beta) = \frac{\frac{1}{\sigma^2} (R(\hat{\beta} - \beta))^\top [R(\mathbf{X}^\top \mathbf{X})^{-1} R^\top]^{-1} R(\hat{\beta} - \beta)}{\frac{(n-(p+1)) \frac{\hat{\sigma}^2}{\sigma^2}}{n-(p+1)}} \sim \frac{\frac{\chi_q^2}{q}}{\frac{\chi^2_{n-(p+1)}}{n-(p+1)}} = F_{q, n-(p+1)}.$$

Rezultatele din exercițiile anterioare conduc la următoarele intervale (atunci când parametrii sunt considerați separat) și regiuni de încredere (atunci când luăm în calcul mai mulți parametrii simultan prin urmare ținem cont și de dependența dintre ei) pentru parametrii modelului de regresie liniară:

- 1) Un interval de încredere bilateral de nivel de încredere $1 - \alpha$ pentru parametrul β_j , $j \in \{0, 1, \dots, p\}$ este dat de

$$I_{\beta_j} = \left[\hat{\beta}_j - t_{n-(p+1)} \left(1 - \frac{\alpha}{2}\right) \hat{\sigma} \sqrt{[(\mathbf{X}^\top \mathbf{X})^{-1}]_{j+1, j+1}}, \hat{\beta}_j + t_{n-(p+1)} \left(1 - \frac{\alpha}{2}\right) \hat{\sigma} \sqrt{[(\mathbf{X}^\top \mathbf{X})^{-1}]_{j+1, j+1}} \right]$$

unde $t_{n-(p+1)} \left(1 - \frac{\alpha}{2}\right)$ este cuantila de ordin $1 - \frac{\alpha}{2}$ a repartiției Student $t_{n-(p+1)}$.

- 2) Un interval de încredere bilateral de nivel de încredere $1 - \alpha$ pentru σ^2 este dat de

$$I_{\sigma^2} = \left[\frac{[n-(p+1)] \hat{\sigma}^2}{\chi^2_{n-(p+1)} \left(1 - \frac{\alpha}{2}\right)}, \frac{[n-(p+1)] \hat{\sigma}^2}{\chi^2_{n-(p+1)} \left(\frac{\alpha}{2}\right)} \right]$$

unde $\chi^2_{n-(p+1)} \left(1 - \frac{\alpha}{2}\right)$ și $\chi^2_{n-(p+1)} \left(\frac{\alpha}{2}\right)$ sunt cuantilele de ordin $1 - \frac{\alpha}{2}$ și respectiv $\frac{\alpha}{2}$ a repartiției $\chi^2_{n-(p+1)}$.

3) O regiune de încredere de nivel de încredere $1 - \alpha$ pentru q ($q \leq p + 1$) parametrii β_j , notați $(\beta_{j_1}, \dots, \beta_{j_q})$, este dată de

- atunci când σ este cunoscută, de

$$RC^{1-\alpha}(R\beta) = \left\{ R\beta \in \mathbb{R}^q \mid \frac{1}{\sigma^2} \left(R(\hat{\beta} - \beta) \right)^\top [R(\mathbf{X}^\top \mathbf{X})^{-1} R^\top]^{-1} R(\hat{\beta} - \beta) \leq \chi_q^2(1 - \alpha) \right\}$$

- atunci când σ este necunoscută, de

$$RC^{1-\alpha}(R\beta) = \left\{ R\beta \in \mathbb{R}^q \mid \frac{1}{q\hat{\sigma}^2} \left(R(\hat{\beta} - \beta) \right)^\top [R(\mathbf{X}^\top \mathbf{X})^{-1} R^\top]^{-1} R(\hat{\beta} - \beta) \leq F_{q,n-(p+1)}(1 - \alpha) \right\}$$

unde R este o matrice de dimensiune $q \times (p + 1)$ cu toate elementele nule exceptând elementele R_{i,j_i} care sunt egale cu 1.

Ca ilustrare a regiunii de încredere de mai sus, să presupunem că $p \geq 2$ și $q = 2$ iar matricea R este

$$R = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \end{pmatrix}$$

ceea ce conduce la $R(\hat{\beta} - \beta) = \begin{pmatrix} \hat{\beta}_0 - \beta_0 \\ \hat{\beta}_1 - \beta_1 \end{pmatrix}$. Dacă presupunem că σ^2 este necunoscut atunci regiunea de încredere pentru (β_0, β_1) este dată de

$$RC^{1-\alpha}(\beta_0, \beta_1) = \left\{ (\beta_0, \beta_1) \in \mathbb{R}^2 \mid \frac{1}{2\hat{\sigma}^2} (\hat{\beta}_0 - \beta_0 \quad \hat{\beta}_1 - \beta_1) [R(\mathbf{X}^\top \mathbf{X})^{-1} R^\top]^{-1} \begin{pmatrix} \hat{\beta}_0 - \beta_0 \\ \hat{\beta}_1 - \beta_1 \end{pmatrix} \leq F_{2,n-(p+1)}(1 - \alpha) \right\}$$

și notând cu c_{ij} elementul de pe linia i coloana j a matricei $(\mathbf{X}^\top \mathbf{X})^{-1}$, găsim că

$$RC^{1-\alpha}(\beta_0, \beta_1) = \left\{ (\beta_0, \beta_1) \in \mathbb{R}^2 \mid \frac{c_{22}(\hat{\beta}_0 - \beta_0)^2 - 2c_{12}(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1) + c_{11}(\hat{\beta}_1 - \beta_1)^2}{2\hat{\sigma}^2(c_{11}c_{22} - c_{12}^2)} \leq F_{2,n-(p+1)}(1 - \alpha) \right\}$$

ceea ce arată că regiunea are formă de elipsă (vezi [Ornea and Turtoi, 2000, Capitolul 2, Secțiunea 16]).

Ex. 2.14



Fie $(x_{n+1,1}, \dots, x_{n+1,p})$ o nouă observație și considerăm $\mathbf{x}_{n+1}^\top = (1, x_{n+1,1}, \dots, x_{n+1,p})$. Ne propunem să prezicem valoarea y_{n+1} conform modelului

$$y_{n+1} = \mathbf{x}_{n+1}^\top \beta + \varepsilon_{n+1}$$

cu $\varepsilon_{n+1} \sim \mathcal{N}(0, \sigma^2)$ independentă de ε_i , $1 \leq i \leq n$.

Arătați că un interval de încredere de nivel de încredere $1 - \alpha$ pentru y_{n+1} este dat de

$$\left[\mathbf{x}_{n+1}^\top \hat{\beta} - t_{n-(p+1)} \left(1 - \frac{\alpha}{2} \right) \hat{\sigma} \sqrt{1 + \mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{n+1}}, \mathbf{x}_{n+1}^\top \hat{\beta} + t_{n-(p+1)} \left(1 - \frac{\alpha}{2} \right) \hat{\sigma} \sqrt{1 + \mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{n+1}} \right]$$

Plecând de la un eșantion de talie n , $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$, găsim estimatorul de verosimilitate maximă (același cu cel obținut prin metoda celor mai mici pătrate) $\hat{\beta}$ a lui β cu ajutorul căruia putem prezice valoarea \hat{y}_{n+1} după relația

$$\hat{y}_{n+1} = \mathbf{x}_{n+1}^\top \hat{\beta}.$$

Pentru a cuantifica eroarea de predicție $y_{n+1} - \hat{y}_{n+1}$ folosim următoarea descompunere

$$y_{n+1} - \hat{y}_{n+1} = \mathbf{x}_{n+1}^\top (\beta - \hat{\beta}) + \varepsilon_{n+1},$$

care este o sumă de două variabile gaussiene independente, ε_{n+1} și $\hat{\beta}$ care este un vector gaussian care depinde de ε_i , $1 \leq i \leq n$. Prin urmare $y_{n+1} - \hat{y}_{n+1}$ este un vector gaussian și ținând cont că matricea de varianță-covarianță $Var(\hat{\varepsilon}_{n+1}) = \sigma^2 (1 + \mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{n+1})$ avem

$$y_{n+1} - \hat{y}_{n+1} \sim \mathcal{N}(0, \sigma^2 (1 + \mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{n+1})).$$

Această relație se rescrie sub forma

$$\frac{y_{n+1} - \hat{y}_{n+1}}{\sigma \sqrt{1 + \mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{n+1}}} \sim \mathcal{N}(0, 1)$$

și înlocuind σ cu estimatorul său $\hat{\sigma}$ obținem

$$\frac{y_{n+1} - \hat{y}_{n+1}}{\hat{\sigma} \sqrt{1 + \mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{n+1}}} = \frac{\frac{y_{n+1} - \hat{y}_{n+1}}{\sigma \sqrt{1 + \mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{n+1}}}}{\frac{\sigma}{\hat{\sigma}}} = \frac{\frac{y_{n+1} - \hat{y}_{n+1}}{\sigma \sqrt{1 + \mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{n+1}}}}{\sqrt{\frac{[n-(p+1)]\sigma}{n-(p+1)}}} \sim \frac{\mathcal{N}(0, 1)}{\sqrt{\frac{\chi_{n-(p+1)}^2}{n-(p+1)}}}.$$

Remarcăm că numărătorul și numitorul sunt variabile aleatoare independente deoarece $y_{n+1} - \hat{y}_{n+1} = \mathbf{x}_{n+1}^\top (\beta - \hat{\beta}) + \varepsilon_{n+1}$ este independentă de $\hat{\sigma}$ pentru că $\hat{\sigma}$ este independentă de $\hat{\beta}$ și de ε_{n+1} (estimatorul $\hat{\sigma}$ nu depinde decât de ε_i , $1 \leq i \leq n$). Prin urmare găsim că

$$\frac{y_{n+1} - \hat{y}_{n+1}}{\hat{\sigma} \sqrt{1 + \mathbf{x}_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{n+1}}} \sim t_{n-(p+1)}$$

de unde rezultă intervalul de încredere dorit.

2.4.2 Testare de ipoteze statistice

Considerăm modelul de regresie liniară

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon$$

sub ipotezele $\mathcal{H}_1 : rang(\mathbf{X}) = p + 1$ și $\mathcal{H}_2' : \varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$. În particula, acest model ne spune că $\mathbb{E}[\mathbf{Y}] = \mathbf{X}\beta \in \mathcal{M}(X)$ unde $\mathcal{M}(X)$ este subspațiul de dimensiune $p + 1$ a lui \mathbb{R}^n generat de coloanele matricei de design \mathbf{X} .

Pentru început, în această secțiune ne propunem să testăm dacă ultimii q coeficienți, $q = p + 1 - p_0 \leq p + 1$, sunt nuli, altfel spus vrem să testăm ipotezele

$$H_0 : \beta_{p_0+1} = \dots = \beta_p = 0 \quad \text{versus} \quad H_1 : \exists j \in \{p_0 + 1, \dots, p\} \beta_j \neq 0$$

În termeni de model, ipoteza H_0 ne spune că modelul revine la

$$\mathbf{Y} = \mathbf{X}_0 \beta_0 + \varepsilon_0$$

sub ipotezele $\mathcal{H}_1 : \text{rang}(\mathbf{X}_0) = p_0$ și $\mathcal{H}_2 : \varepsilon_0 \sim \mathcal{N}(0, \sigma^2 I_n)$, unde matricea $\mathbf{X}_0 \in \mathcal{M}_{n, p_0}(\mathbb{R})$ este compusă din primele p_0 coloane ale lui \mathbf{X} iar $\beta_0 \in \mathcal{M}_{p_0, 1}(\mathbb{R})$. Fie $\mathcal{M}(X_0)$ subspațiul lui $\mathcal{M}(X)$ de dimensiune p_0 generat de coloanele lui \mathbf{X}_0 (deoarece $\text{rang}(\mathbf{X}) = p + 1$ deducem că $\text{rang}(\mathbf{X}_0) = p_0$). Sub ipoteza H_0 avem că $\mathbb{E}[\mathbf{Y}] = \mathbf{X}_0 \beta_0 \in \mathcal{M}(X_0)$.

Ex. 2.15



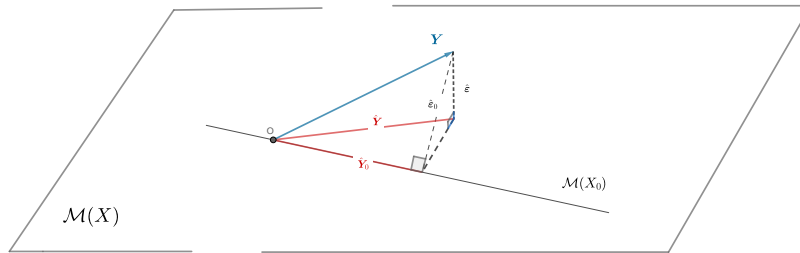
Sub ipoteza H_0 statistica de test

$$F = \frac{n - (p + 1)}{q} \frac{\|\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0\|^2}{\|\mathbf{Y} - \hat{\mathbf{Y}}\|^2} = \frac{n - (p + 1)}{q} \frac{RSS_0 - RSS}{RSS} \sim F_{q, n - (p + 1)}$$

unde $F_{q, n - (p + 1)}$ este repartiția lui Fisher cu $(q, n - (p + 1))$ grade de libertate.

a) Abordarea geometrică

Vom prezenta mai jos o abordare din punct de vedere geometric a problemei de testare. Considerăm spațiul $\mathcal{M}(X_0) \subset \mathcal{M}(X)$ și am văzut că sub ipoteza nulă avem $\mathbb{E}[\mathbf{Y}] = \mathbf{X}_0 \beta_0 \in \mathcal{M}(X_0)$. În această situație, metoda celor mai mici pătrate consistă în proiectarea lui \mathbf{Y} pe $\mathcal{M}(X_0)$ pentru a obține $\hat{\mathbf{Y}}_0$ (vezi figura de mai jos).



Ideea procedurii de testare, și prin urmare a deciziei de a respinge sau nu ipoteza nulă, este următoarea: dacă proiecția $\hat{\mathbf{Y}}_0$ a lui \mathbf{Y} pe $\mathcal{M}(X_0)$ este *aproape* de proiecția $\hat{\mathbf{Y}}$ a lui \mathbf{Y} pe $\mathcal{M}(X)$ atunci nu respingem ipoteza nulă, în caz contrar o respingem în favoarea alternativei H_1 . De fapt, dacă informația adusă de cele două modele nu diferă foarte mult atunci este recomandat să păstrăm modelul mai simplu (principiul parcimoniei).

Pentru a cuantifica *apropierea* dintre cele două proiecții putem considera distanța euclidiană $\|\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0\|^2$ dintre acestea. Această distanță variază în funcție de datele și de unitățile de măsură a acestora și pentru a evita această problemă de scară se recomandă *standardizarea* acestei distanțe prin împărțirea ei la pătratul normei erorii reziduale $\|\hat{\varepsilon}\|^2 = \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2 = [n - (p + 1)]\hat{\sigma}^2$. Cum vectorii aleatori $\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0$ și $\hat{\varepsilon}$ nu aparțin în

subspații de aceeași dimensiune trebuie să îi împărțim și prin gradele de libertate corespunzătoare, respectiv $q = p + 1 - p_0$ și $n - (p + 1)$. Obținem astfel statistica de test

$$F = \frac{\frac{\|\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0\|^2}{p+1-p_0}}{\frac{\|\mathbf{Y} - \hat{\mathbf{Y}}\|^2}{n-(p+1)}} = \frac{n - (p + 1)}{(p + 1) - p_0} \frac{\|\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0\|^2}{\|\mathbf{Y} - \hat{\mathbf{Y}}\|^2}.$$

Pentru a putea utiliza această statistică de test este necesar să determinăm repartiția sa sub ipoteza nulă. Notând cu P_X și respectiv P_{X_0} matricele de proiecție ortogonală pe $\mathcal{M}(X)$ și $\mathcal{M}(X_0)$, știm că

$$\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0 = P_X \mathbf{Y} - P_{X_0} \mathbf{Y}$$

și cum $\mathcal{M}(X_0) \subset \mathcal{M}(X)$, deci $P_{X_0} \mathbf{Y} = P_{X_0} P_X \mathbf{Y}$ și

$$\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0 = P_X \mathbf{Y} - P_{X_0} \mathbf{Y} = (I_n - P_{X_0}) P_X \mathbf{Y} = P_{X_0}^\perp P_X \mathbf{Y}.$$

Astfel am găsit că $\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0 \in \mathcal{M}(X_0)^\perp \cap \mathcal{M}(X)$ și cum $\mathbf{Y} - \hat{\mathbf{Y}} \in \mathcal{M}(X)^\perp$ găsim că $\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0 \perp \mathbf{Y} - \hat{\mathbf{Y}}$. Vectorii $\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0$ și $\mathbf{Y} - \hat{\mathbf{Y}}$ sunt elemente din spații ortogonale ceea ce implică $Cov(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0, \mathbf{Y} - \hat{\mathbf{Y}}) = 0$ și cum sunt și vectori gaussiani (conform ipotezei \mathcal{H}'_2) deducem că sunt și independenți ceea ce arată că numărătorul și numitorul lui F sunt independente.

Conform Teoremei lui Cochran avem pentru numitor

$$\frac{1}{\sigma^2} \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2 = \frac{1}{\sigma^2} \|P_{X^\perp} \mathbf{Y}\|^2 = \frac{1}{\sigma^2} \|P_{X^\perp} (\mathbf{X}\beta + \varepsilon)\|^2 = \frac{1}{\sigma^2} \|P_{X^\perp} \varepsilon\|^2 \sim \chi^2_{n-(p+1)}$$

iar pentru numărător

$$\frac{1}{\sigma^2} \|\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0\|^2 = \frac{1}{\sigma^2} \|P_{X_0}^\perp P_X \mathbf{Y}\|^2 = \frac{1}{\sigma^2} \|P_{X_0}^\perp P_X (\mathbf{Y} - \mathbf{X}\beta)\|^2 \sim \chi^2_q$$

Sub H_0 , parametrul $\|P_{X_0}^\perp P_X \mathbf{X}\beta\|^2$ este nul deoarece $\mathbf{X}\beta \in \mathcal{M}(X_0)$ ceea ce conduce la

$$F \stackrel{H_0}{\sim} F_{q, n-(p+1)}.$$

Putem observa de asemenea că aplicând Teorema lui Pitagora

$$\begin{aligned} \|\mathbf{Y} - \hat{\mathbf{Y}}_0\|^2 &= \|\mathbf{Y} - P_X \mathbf{Y} + P_X \mathbf{Y} - P_{X_0} \mathbf{Y}\|^2 = \|(I_n - P_X) \mathbf{Y} + P_{X_0} P_X \mathbf{Y} - P_{X_0} \mathbf{Y}\|^2 \\ &= \|P_{X^\perp} \mathbf{Y} + (I_n - P_{X_0}) P_X \mathbf{Y}\|^2 = \|P_{X^\perp} \mathbf{Y} + P_{X_0}^\perp P_X \mathbf{Y}\|^2 \\ &= \|P_{X^\perp} \mathbf{Y}\|^2 + \|P_{X_0}^\perp P_X \mathbf{Y}\|^2 = \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2 + \|\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0\|^2 \end{aligned}$$

altfel spus

$$\|\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0\|^2 = \|\mathbf{Y} - \hat{\mathbf{Y}}_0\|^2 - \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2 = RSS_0 - RSS.$$

Ex. 2.16



Ne propunem să testăm ipoteza nulă $H_0 : \beta_j = 0$ versus ipoteza alternativă $H_1 : \beta_j \neq 0$. Propuneți o procedură de testare pentru cele două ipoteze.

Această problemă este un caz particular al rezultatului obținut anterior pentru $q = 1$. Avem statistica de test

$$F = \frac{n - (p + 1)}{1} \frac{\|\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0\|^2}{\|\mathbf{Y} - \hat{\mathbf{Y}}\|^2} = \frac{\|\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0\|^2}{\hat{\sigma}^2}$$

cu ajutorul căreia putem construi regiunea critică

$$C = \{\omega \mid F(\omega) > F_{1, n-(p+1)}(1 - \alpha)\}.$$

Acest test este echivalent cu testul bazat pe statistica de test repartizată student cu $n - (p + 1)$ grade de libertate (mai exact $F^2 = T_j^2$)

$$T_j = \frac{\hat{\beta}_j}{\hat{\sigma}_{\hat{\beta}_j}} = \frac{\hat{\beta}_j}{\hat{\sigma} \sqrt{[(\mathbf{X}^\top \mathbf{X})^{-1}]_{j+1, j+1}}}.$$

Regiunea critică a testului este

$$C = \left\{ \omega \mid T_j(\omega) > t_{n-(p+1)} \left(1 - \frac{\alpha}{2} \right) \right\}.$$

Aceasta este forma sub care testul de semnificativitate pentru un coeficient apare în R.

Ex. 2.17



Propuneți o procedură de testare pentru ipotezele

$$H_0 : \beta_1 = \dots = \beta_p = 0 \quad \text{versus} \quad H_1 : \exists j \in \{1, \dots, p\} \beta_j \neq 0$$

Vrem să testăm dacă variabilele explicative au sau nu influență asupra variabilei răspuns, cu alte cuvinte vrem să testăm dacă toți coeficienții sunt nuli exceptînd constanta (β_0) - acesta este testul care apare ca rezultat atunci când efectuăm sumarul modelului liniar în R (comanda `summary(model)`). Suntem în cazul particular în care $\hat{\mathbf{Y}}_0 = \bar{y}\mathbf{1}$ și obținem statistica de test a lui Fisher (*testul global al lui Fisher*)

$$F = \frac{n - (p + 1)}{p} \frac{\|\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0\|^2}{\|\mathbf{Y} - \hat{\mathbf{Y}}\|^2} \sim F_{p, n-(p+1)}$$

sau exprimată în termeni de coeficient de determinare

$$F = \frac{n - (p + 1)}{p} \frac{R^2}{1 - R^2} \sim F_{p, n-(p+1)}.$$

Ex. 2.18



Vrem să testăm ipotezele

$$H_0 : \beta_{p_0+1} = \dots = \beta_p = 0 \quad \text{versus} \quad H_1 : \exists j \in \{p_0 + 1, \dots, p\} \beta_j \neq 0$$

cu ajutorul testului bazat pe raportul de verosimilități. Arătați că acest test este echivalent cu testul F al lui Fisher.

Testul bazat pe raportul de verosimilitate este

$$\Lambda(\mathbf{y}) = \frac{\sup_{\theta \in \Theta_0} L(\theta|\mathbf{y})}{\sup_{\theta \in \Theta} L(\theta|\mathbf{y})},$$

unde Θ este spațiul parametrilor modelului, Θ_0 este spațiul parametrilor corespunzător ipotezei nule iar $L(\theta|\mathbf{x})$ este funcția de verosimilitate.

Observăm că spațiul parametrilor modelului este

$$\Theta = \{(\boldsymbol{\beta}, \sigma^2) | \boldsymbol{\beta} \in \mathbb{R}^{p+1}, \sigma^2 \in (0, \infty)\},$$

cel corespunzător ipotezei nule este

$$\Theta_0 = \{(\boldsymbol{\beta}_0, \sigma^2) | \boldsymbol{\beta}_0 \in \mathbb{R}^{p_0}, \sigma^2 \in (0, \infty)\}$$

iar funcția de verosimilitate este

$$L(\boldsymbol{\beta}, \sigma^2; \mathbf{Y}) = \prod_{i=1}^n f_Y(y_i) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2} = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2}.$$

Observăm că

$$\sup_{\theta \in \Theta} L(\theta|\mathbf{y}) = \sup_{\boldsymbol{\beta}, \sigma^2} L(\boldsymbol{\beta}, \sigma^2; \mathbf{Y}) = L(\hat{\boldsymbol{\beta}}_{VM}, \hat{\sigma}_{VM}^2; \mathbf{Y})$$

unde $\hat{\boldsymbol{\beta}}_{VM} = \hat{\boldsymbol{\beta}}$ este estimatorul obținut prin metoda celor mai mici pătrate iar $\hat{\sigma}_{VM}^2 = \frac{\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2}{n} = \frac{RSS}{n}$. Astfel găsim

$$\sup_{\theta \in \Theta} L(\theta|\mathbf{y}) = \sup_{\boldsymbol{\beta}, \sigma^2} L(\boldsymbol{\beta}, \sigma^2; \mathbf{Y}) = \left(\frac{n}{2\pi \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2} \right)^{\frac{n}{2}} e^{-\frac{n}{2}} = \left(\frac{n}{2\pi RSS} \right)^{\frac{n}{2}} e^{-\frac{n}{2}}.$$

Sub H_0 avem

$$\sup_{\theta \in \Theta_0} L(\theta|\mathbf{y}) = \sup_{\boldsymbol{\beta}_0, \sigma^2 \in \Theta_0} L(\boldsymbol{\beta}_0, \sigma^2; \mathbf{Y}) = L(\hat{\boldsymbol{\beta}}_0, \hat{\sigma}_0^2; \mathbf{Y}) = \left(\frac{n}{2\pi RSS_0} \right)^{\frac{n}{2}} e^{-\frac{n}{2}}$$

unde $RSS_0 = \|\mathbf{Y} - \mathbf{X}_0 \hat{\boldsymbol{\beta}}_0\|^2$ iar $\hat{\sigma}_0^2 = \frac{\|\mathbf{Y} - \mathbf{X}_0 \hat{\boldsymbol{\beta}}_0\|^2}{n} = \frac{RSS_0}{n}$.

Prin urmare raportul de verosimilitate devine

$$\Lambda(\mathbf{y}) = \frac{\sup_{\theta \in \Theta_0} L(\theta|\mathbf{y})}{\sup_{\theta \in \Theta} L(\theta|\mathbf{y})} = \frac{\left(\frac{n}{2\pi RSS_0} \right)^{\frac{n}{2}} e^{-\frac{n}{2}}}{\left(\frac{n}{2\pi RSS} \right)^{\frac{n}{2}} e^{-\frac{n}{2}}} = \left(\frac{RSS_0}{RSS} \right)^{-\frac{n}{2}}$$

iar regiunea critică a testului se scrie (testul respinge ipoteza nulă atunci când statistica Λ este suficient de mică)

$$C = \{\mathbf{Y} \in \mathbb{R}^n | \Lambda(\mathbf{Y}) < \lambda_0\}.$$

Observăm că pentru $\lambda > 0$, funcția $h(\lambda) = \lambda^{-\frac{n}{2}} - 1$ este descrescătoare, adică $\lambda < \lambda_0$ implică $h(\lambda) > h(\lambda_0)$.
 Avem

$$h(\lambda) > h(\lambda_0) \iff \frac{RSS_0 - RSS}{RSS} > h(\lambda_0) \iff \frac{n - (p + 1)}{(p + 1) - p_0} \frac{RSS_0 - RSS}{RSS} > f_0$$

unde f_0 se determină din condiția

$$\mathbb{P}_{H_0} \left(\frac{n - (p + 1)}{(p + 1) - p_0} \frac{RSS_0 - RSS}{RSS} > f_0 \right) = \alpha$$

și cum $\frac{n - (p + 1)}{(p + 1) - p_0} \frac{RSS_0 - RSS}{RSS} \sim F_{(p + 1) - p_0, n - (p + 1)}$ conform exercițiului 2.15 deducem că $f_0 = f_{(p + 1) - p_0, n - (p + 1)}(1 - \alpha)$. Am găsit că regiunea critică a testului bazat pe raportul de verosimilitate este identică cu cea a testului F a lui Fisher (vezi exercițiul 2.15).

Referințe

- Julian Faraway. *Linear Models with R*. CRC Press, 2nd edition, 2015. ISBN 978-1-4398-8734-9. (Citat la paginile 13 și 20.)
- H. V. Henderson and S. R. Searle. On deriving the inverse of a sum of matrices. *SIAM Review*, 23(1):53–60, 1981. doi: 10.1137/1023004. URL <https://epubs.siam.org/doi/abs/10.1137/1023004>. (Citat la pagina 23.)
- Jean Jacod and Philip Protter. *Probability essentials*. Springer, 2003. (Citat la pagina 31.)
- L. Ornea and A. Turtoi. *O introducere in geometrie*. Fundatia Theta, 2000. (Citat la paginile 14, 15 și 36.)
- George Seber and Alan Lee. *Linear Regression Analysis*. Wiley, 2nd edition, 2003. ISBN 0-471-41540-5. (Citat la paginile 13 și 31.)
- Sanford Weisberg. *Applied Linear Regression*. Wiley, 4th edition, 2014. ISBN 978-1-118-38608-8. (Citat la pagina 13.)
- Haruo Yanai, Kei Takeuchi, and Yoshio Takane. *Projection matrices, generalized inverse matrices, and singular value decomposition*. Springer-Verlag New York, 2011. ISBN 978-1-4419-9886-6. doi: <https://doi.org/10.1007/978-1-4419-9887-3>. (Citat la pagina 15.)