07 Régression Linéaire

mardi 18 octobre 2022

1. Modèle de régression linéaire Gaussien
   - -1- $Y_i = a + bx_i + \varepsilon_i$, où $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$
   - -2- Y_i sont indépendantes
   - -3- Y_i ne sont pas identiquement distribuées
   - -4- Hypothèse/Conditions
     1- Linéarité
     2- Normalité (erreurs~N)
     3- Homoscédasticité (variance des erreurs sont même)
     4- Indépendance
2. Estimation de modèle de régression linéaire
   - -1- Log-vraisemblance
     1- $l(a, b, \sigma^2; y_1, \dots, y_n) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum(y_i - a - bx_i)^2$
     2- $\begin{cases} \hat{a} = \bar{Y} - \hat{b}\bar{x} \\ \hat{b} = \frac{S_{xY}}{s_x^2} \\ S_{\mathrm{res}} = \hat{\sigma}_{MV}^2 = \frac{1}{n}\sum(y_i - \hat{a} - \hat{b}x_i)^2 \end{cases}$
   - -2- Variables sans biais
     1- $\hat{\sigma}^2$ estimateur sans biais de σ^2
       1> $\hat{\sigma}^2 = \frac{n}{n-2}\hat{\sigma}_{MV}^2$
     2- $cov(\hat{a}, \hat{b}) = -\frac{\hat{x}\sigma^2}{ns_x^2}$
   - -3- Lois des estimateur et indépendance
     1- $\hat{a} \sim \mathcal{N}\left(a, \frac{\sigma^2}{n}\left(1 + \frac{\hat{x}^2}{s_x^2}\right)\right)$
     2- $\hat{b} \sim \mathcal{N}\left(b, \frac{\sigma^2}{ns_x^2}\right)$
     3- $\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-2}^2$
   - -4- Intervalles de confiance
     1- Fonctions pivotales

1> $\dfrac{(\hat{a}-a)}{\sqrt{\dfrac{\sigma^2}{ns_x^2}}} \sim \mathcal{T}_{n-2}$

2> $\dfrac{(\hat{b}-b)}{\sqrt{\dfrac{\hat{\sigma}^2}{n}\left(1+\dfrac{\bar{x}^2}{s_x^2}\right)}} \sim \mathcal{T}_{n-2}$

3> $\dfrac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}$

-5- Qualité de l'ajustement et validation des hypothèses

1- Analyse de variance

1> $\frac{1}{n}\sum\left(Y_i-\bar{Y}\right)^2=\frac{1}{n}\sum\left(\bar{Y}-\hat{Y}_i\right)^2+\frac{1}{n}\sum\left(Y_i-\hat{Y}\right)^2=\frac{1}{n}\sum\left(\bar{Y}-\hat{Y}_i\right)^2+\hat{\sigma}^2_{MV}$

a) $S_Y^2 = S_{reg} + S_{res}$

b) $Si\ \forall i, Y_i = \hat{Y}_i \Rightarrow S_Y^2 = S_{reg}$

c) $Si\ \hat{b} = 0, \hat{Y}_i = \bar{Y} \Rightarrow S_Y^2 = S_{res}$

2- Coefficient de détermination

1> $R^2 = \dfrac{S_{reg}}{S_Y^2} = \dfrac{S_{xY}^2}{s_x^2 S_Y^2} \in [0,1]$

3- Analyse des résidus $\hat{\varepsilon}_i$

1> Homoscédasticité 同方差性

a) $Var(\hat{\varepsilon}_i) = (1-h_{ii})\sigma^2, h_{ii} = \dfrac{1}{n} + \dfrac{(x_i-\bar{x})^2}{ns_x^2}$

3. Prévision

-1- Fonction pivotale pour $E(Y_0)$

-1- $\begin{cases} E(Y_0) = E(\hat{Y}_0) \\ Var(\hat{Y}_0) = \sigma^2\left(\dfrac{1}{n} + \dfrac{(\bar{x}-x_0)^2}{ns_x^2}\right) \end{cases}$

-2- σ^2 connue

1> $\dfrac{\hat{Y}_0 - E(Y_0)}{\sqrt{\sigma^2\left(\dfrac{1}{n} + \dfrac{(\bar{x}-x_0)^2}{ns_x^2}\right)}} \sim \mathcal{N}(0,1)$

2> $IC = \hat{Y}_0 \pm u_{1-\frac{\alpha}{2}}\sqrt{\sigma^2\left(\dfrac{1}{n} + \dfrac{(\bar{x}-x_0)^2}{ns_x^2}\right)}$

-3- σ^2 inconnue

1> $$\frac{\hat{Y}_0 - E(Y_0)}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{ns_x^2}\right)}} \sim \mathcal{T}_{n-2}$$

2> $$IC = \hat{Y}_0 \pm t_{n-2,1-\frac{\alpha}{2}}\sqrt{\hat{\sigma}^2\left(\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{ns_x^2}\right)}$$

-2- Fonction pivotale pour Y_0 au niveau $1-\alpha$

-1- $$\begin{cases} E(Y_0 - \hat{Y}_0) = 0 \\ Var(\hat{Y}_0 - Y_0) = \sigma^2\left(\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{ns_x^2}\right) + \sigma^2 \end{cases}$$

-2- σ^2 connue

1> $$\frac{\hat{Y}_0 - Y_0}{\sqrt{\sigma^2\left(\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{ns_x^2}+1\right)}} \sim \mathcal{N}(0,1)$$

2> $$IC = \hat{Y}_0 \pm u_{1-\frac{\alpha}{2}}\sqrt{\sigma^2\left(\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{ns_x^2}+1\right)}$$

-3- σ^2 inconnue

1> $$\frac{\hat{Y}_0 - Y_0}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{ns_x^2}+1\right)}} \sim \mathcal{T}_{n-2}$$

2> $$IC = \hat{Y}_0 \pm t_{n-2,1-\frac{\alpha}{2}}\sqrt{\hat{\sigma}^2\left(\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{ns_x^2}+1\right)}$$