

Régression linéaire multiple et méthodes de régularisation

LARIBI Yanis, SOUIDI Yassine

École des Mines de Nancy



Sommaire

- 1** Introduction et problématique
- 2** Sélection de sous modèles
- 3** Méthodes de régularisation
- 4** Application : jeu de données Life_expectancy
 - Modélisation : approche globale
 - Modélisation : approche plus fine

Contexte

On souhaite estimer, modéliser et prédire une variable à expliquer quantitative Y à partir d'un ensemble de variable explicatives quantitatives et/ou qualitatives X_1, \dots, X_p selon un modèle paramétrique de la forme :

$$Y = f_{\theta}(X_1, \dots, X_p) + \varepsilon$$

Où : $\theta \in \Theta \subset \mathbb{R}^d$, $f_{\theta} : \mathbb{R}^p \rightarrow \mathbb{R}$ et $\varepsilon \in \mathbb{R}$ erreur aléatoire.

Cas de la régression linéaire multiple :

On cherche $f_{\theta} : \mathbb{R}^p \rightarrow \mathbb{R}$ dans l'ensemble \mathcal{H} des formes affines de \mathbb{R}^p

$$\mathcal{H} := \{(x_1, \dots, x_p) \rightarrow \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \mid \beta_0, \dots, \beta_p \in \mathbb{R}\}$$



Régression linéaire multiple

Si nous disposons d'un jeu de données comprenant N observations de la variable à expliquer \mathbf{Y} associées aux valeurs des prédicteurs $\mathbf{X}_1, \dots, \mathbf{X}_p$ correspondantes, alors notre modèle peut s'écrire :

$$y_i = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_p x_{i,p} + \varepsilon_i \quad \forall i = 1, \dots, N$$

Soit encore, matriciellement :

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

Où :

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & x_{N2} & \cdots & x_{Np} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

Exemples d'applications

Marketing Mix Modeling

Y : Nombre de ventes hebdomadaires d'une enseigne

X : Budget TV, budget digital, prix, promotions

Biologie / Santé

Y : Tension artérielle systolique

X : Âge, IMC, consommation de sodium, activité physique

Finance (FF3 facteurs)

Y : Rendement excès d'un titre

X : Prime de marché, SMB, HML

Hypothèses de la régression linéaire multiple

Hypothèses structurelles

- **Linéarité** : $\mathbb{E}(y_i \mid x_{i1}, \dots, x_{ip}) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \quad \forall i = 1, \dots, N$
- **Absence de multicolinéarité** : $\text{rank}(X) = p + 1$ (*rang plein*)

Hypothèses stochastiques

- **Indépendance** : $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j$
- **Homoscedasticité** : $\text{Var}(\varepsilon_i) = \sigma^2 \quad \forall i = 1, \dots, N$
- **Normalité** : $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$

Estimation des paramètres à partir des données

Estimateur des moindres carrés ordinaires

$$\hat{\beta}^{OLS} := \arg \min_{\beta \in \mathbb{R}^{p+1}} \|Y - X\beta\|^2$$

Si la matrice $X^T X$ est inversible, alors l'estimateur des moindres carrés ordinaires est donné par :

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

- Sous les hypothèses précédentes, cet estimateur est sans biais et de variance minimale parmi les estimateurs linéaires.
- puisque $rg(X^T X) = rg(X)$ il faut que la matrice X soit de rang plein pour que $X^T X$ soit inversible.

Estimation par maximum de vraisemblance et lois des estimateurs

Autre méthode classique d'estimation paramétrique :

$$\hat{\beta}^{\text{MV}} = \arg \max_{\beta \in \mathbb{R}^{p+1}} L_N(Y; \beta)$$

Où $L_N(Y; \beta)$ désigne la vraisemblance.

NB : $\hat{\beta}^{\text{OLS}} = \hat{\beta}^{\text{MV}}$

Sous l'hypothèse du b.b.g (i.e $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_N)$) nous avons $Y \sim \mathcal{N}(X\beta, \sigma^2 I_N)$.

Grâce au **Théorème de Cochran** (cf rapport), nous pouvons déterminer la loi des estimateurs :

$$\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2(X^T X)^{-1}) \quad \text{et} \quad \hat{\sigma}^2 \sim \frac{\sigma^2}{N-p-1} \chi^2(N-p-1)$$

Problématiques

■ Multicolinéarité : Variables explicatives corrélées :

Matrice $X^T X$ mal conditionnée \Rightarrow estimation des coefficients instable.

⇒ Régularisation (Ridge,LASSO,Elastic-Net) pour stabiliser l'estimation

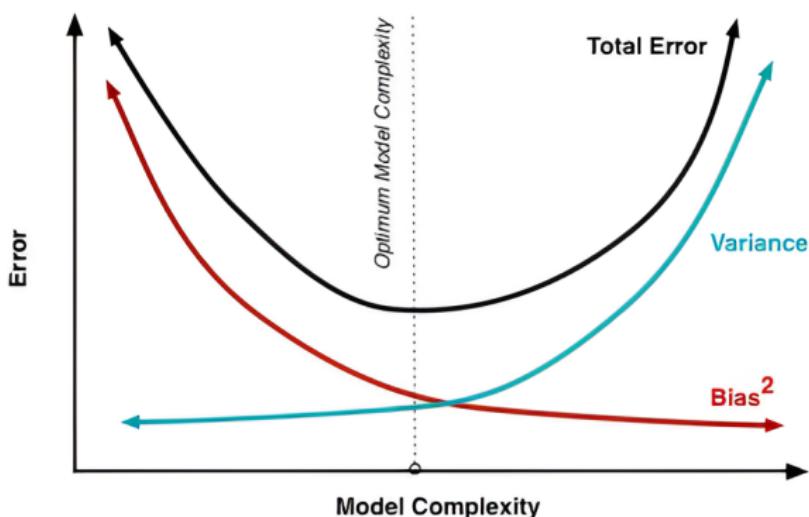
■ **Parcimonie** : Trop de variables explicatives :

⇒ modèle difficile à interpréter.

⇒ problèmes de sur-ajustement.

⇒ Sélection des variables essentielles

Compromis biais-variance



$$\text{Erreurs}(x_0) = \sigma^2 + \text{Bias}^2(\hat{f}(x_0)) + \text{Var}(\hat{f}(x_0))$$

- Ici, la complexité de notre modèle provient principalement du nombre de coefficients à estimer ($p + 1$).

Facteur d'inflation de la variance (VIF)

Formule du VIF :

$$\text{VIF}(X_i) = \frac{1}{1 - R_i^2}$$

où R_i^2 est le coefficient de détermination de la régression :

$$X_i = \beta_0 + \beta_1 X_1 + \cdots + \beta_{i-1} X_{i-1} + \beta_{i+1} X_{i+1} + \cdots + \beta_p X_p$$

R_i^2 mesure la proportion de variance de X_i expliquée par les autres variables.

- $\text{VIF}(X_i) \geq 10 \Rightarrow$ Multicolinéarité détectée

Algorithmes de sélection

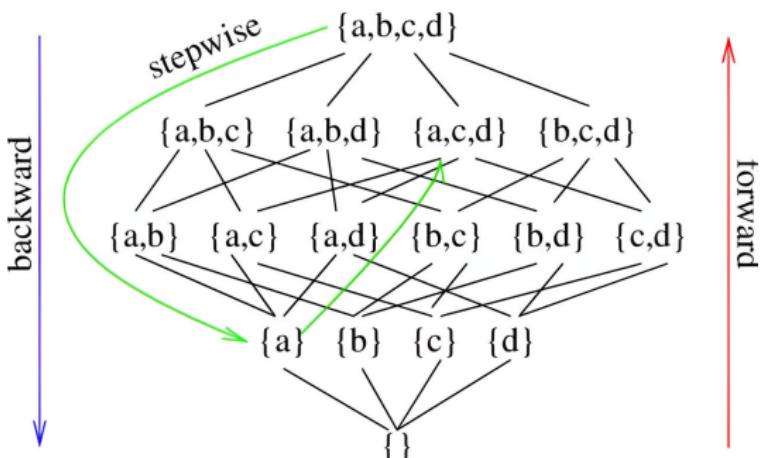


Schéma illustrant les trois procédures¹

Critères d'information

Akaike :

$$AIC = -2 \ln \mathcal{L} + 2k$$

Bayesian :

$$BIC = -2 \ln \mathcal{L} + k \ln(N)$$

Où :

$$k = p + 1$$

¹Rüping 2006

Ridge (aka Tikhonov)

Estimation des paramètres

$$\hat{\beta}_{\text{ridge}} = \arg \min_{\beta \in \mathbb{R}^{p+1}} \|Y - X\beta\|_2^2 + \lambda_2 \sum_{i=1}^p \beta_i^2$$

- Pour λ_2 assez grand on a : $\forall i = 0, \dots, p \quad |\hat{\beta}_i^{\text{ridge}}| < |\hat{\beta}_i^{\text{OLS}}|$ (réduction de la complexité, cf discussion biais-fluctuation)
- L'ajout du terme de pénalisation règle le problème de singularité rencontré lors de l'estimation par moindres carrées ordinaires.
- En effet, on montre que : $\hat{\beta}^{\text{ridge}} = (X^T X + \lambda_2 \text{Id}_{p+1})^{-1} X^T Y$ dès lors que $\lambda_2 > 0$.

LASSO : Least Absolute Shrinkage and Selection Operator

Estimation des paramètres¹

$$\hat{\beta}_{\text{lasso}} = \arg \min_{\beta \in \mathbb{R}^{p+1}} \|Y - X\beta\|_2^2 + \lambda_1 \sum_{i=1}^p |\beta_i|$$

- Pour λ_1 assez grand on a : $\forall i = 0, \dots, p$ $|\hat{\beta}_i^{\text{lasso}}| < |\hat{\beta}_i^{\text{OLS}}|$.
- Pour certains $i \in I$ on a $\hat{\beta}_i^{\text{lasso}} = 0$.
- $L_1 \Rightarrow$ non-différentiabilité en zéro \Rightarrow pas de formule explicite.
- méthodes numériques : coordinate descente, Least Angle Regression Selection (LARS).

¹R. Tibshirani 1996; Robert Tibshirani 2011

Elastic-Net

Estimation des paramètres

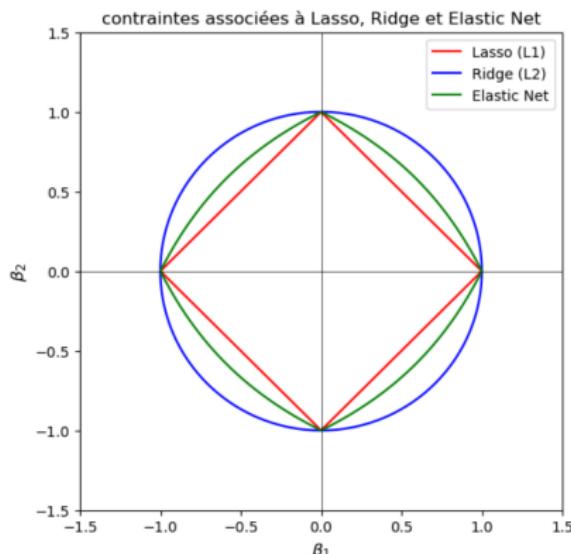
$$\hat{\beta}_{\text{elastic net}} = \arg \min_{\beta \in \mathbb{R}^{p+1}} \|Y - X\beta\|_2^2 + \lambda_2 \sum_{i=1}^p \beta_i^2 + \lambda_1 \sum_{i=1}^p |\beta_i|$$

- Combine les effets de Ridge et LASSO, i.e réduction des estimation des paramètre et sélection de sous modèle.
- plus pertinent que LASSO dans le cas $p \gg N$.
- "grouping effect" → les coefficients associées à des prédicteurs corrélés devraient être proche (en valeur absolue).
- Schéma numérique itératif : LARS-EN (adaptation de LARS).

Pénalisation et contraintes associées

Par exemple, pour Ridge :

$$\min_{\beta} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_2^2 \iff \min_{\|\beta\|_2^2 \leq t} \|Y - X\beta\|_2^2$$



Métriques de performance

Erreur quadratique moyenne (MSE)

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

R^2 et R^2 ajusté

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

$$R_{\text{adj}}^2 = 1 - \frac{N-1}{N-p-1} (1 - R^2)$$

- y_i est la valeur observée et \hat{y}_i la prédition du modèle.
- $R^2 \in [0, 1]$, un R^2 proche de 1 → le modèle explique une grande partie de la variance observée.

Comment choisir les hyper-paramètres ?

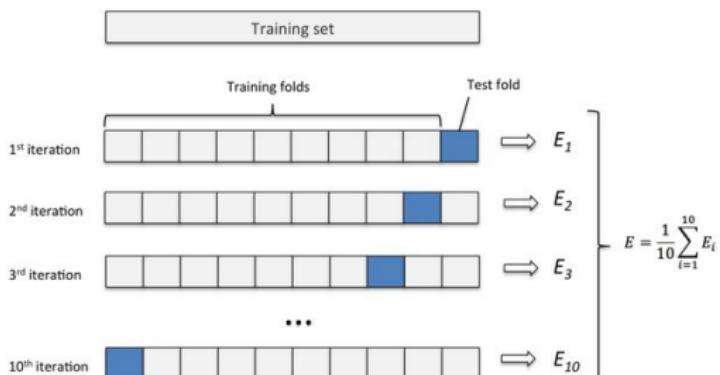


Figure: Validation croisée à **K=10** plis

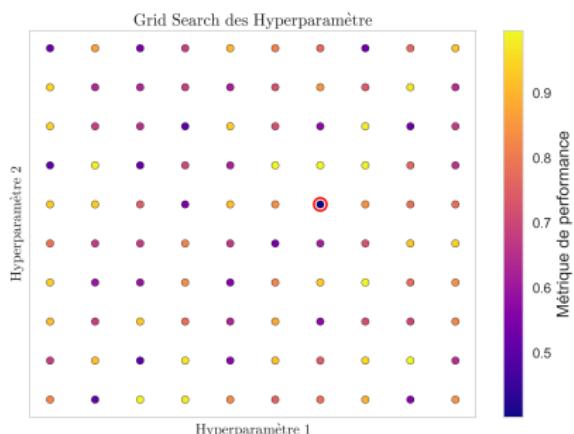


Figure: Grille de recherche des hyperparamètres (exemple illustratif)



Application : modélisation de l'espérance de vie

Y : espérance de vie à la naissance (années)

X : variables économiques, démographiques, sociales, sanitaires...

Définition : l'espérance de vie à la naissance, notée $e(0)$ désigne le nombre moyen d'années qu'un nouveau-né peut espérer vivre selon les taux de mortalité observés pour une certaine année.

Formule :

$$e(0) = \sum_{x=0}^{\infty} S(x)$$

où $S(x) = \prod_{i=0}^{x-1} (1 - q_i)$ est la probabilité de survivre jusqu'à l'âge x , et q_i est le quotient de mortalité à l'âge i .

Variables explicatives



Date et Lieux:

- Année : 2000 à 2015
- Région : Afrique, Asie, Océanie, Amérique centrale et Caraïbes, Union européenne, Moyen-Orient, Amérique du Nord, Reste de l'Europe, Amérique du Sud.

Immunisation

- Hepatite B (HepB3)
- Rougeole (MCV1)
- Polio (Pol3)
- Diphtherie (DTP3)

Santé / mortalité

- Mortalité infantile, décès chez les moins 5, Mortalité adulte.
- IMC, Maigreur (5–9 ans, 10–19 ans)
- Cas VIH pour les 15–49
- Consommation d'alcool.

Économique / social

- PIB/habitant en (US\$ courant).
- Population (millions)
- Nombre moyen d'années d'études chez les +25 ans
- Statut économique : développé Vs en développement.



Dataset Life_expectancy

Notre jeu de données comprends $N = 2864$ observations et $k = 21$ variables. On dispose, pour **179 pays**, des valeurs des différentes variables sur **16 années différentes (2000 → 2016)**.

Country <chr>	Region <chr>	Y... <dbl>	Infant_deaths <dbl>	Under_five_deaths <dbl>
Turkiye	Middle East	2015	11.1	13.0
Spain	European Union	2015	2.7	3.3
India	Asia	2007	51.5	67.9
Guyana	South America	2006	32.8	40.5
Israel	Middle East	2012	3.4	4.3
Costa Rica	Central America and Caribbean	2006	9.8	11.2
Russian Federation	Rest of Europe	2015	6.6	8.2
Hungary	European Union	2000	8.7	10.1
Jordan	Middle East	2001	22.0	26.1
Moldova	Rest of Europe	2008	15.3	17.8

1-10 of 2,864 rows | 1-5 of 21 columns

Previous 1 2 3 4 5 6 ... 287 Next

Figure: Premières lignes du jeu de données Life_expectancy

Pré-traitement des données

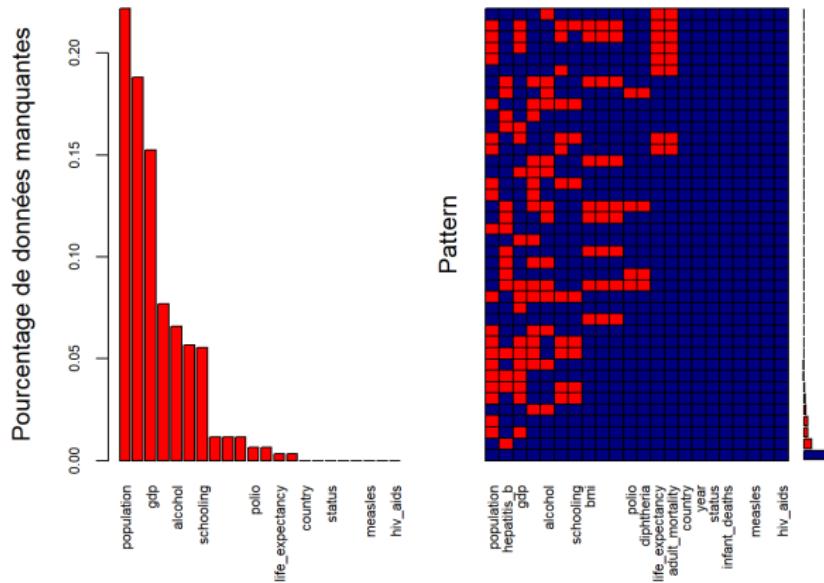


Figure: Taux de valeurs manquantes par ordre décroissant à gauche, position des valeurs manquantes (en rouge) dans le dataset à droite.

Pré-traitement des données

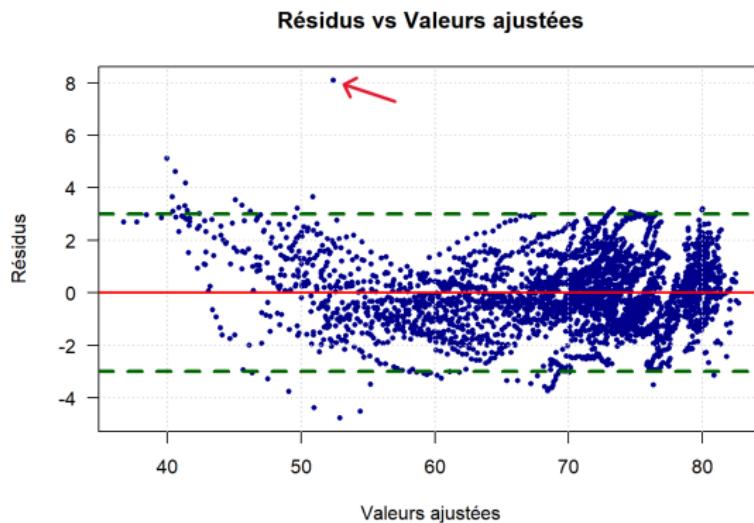
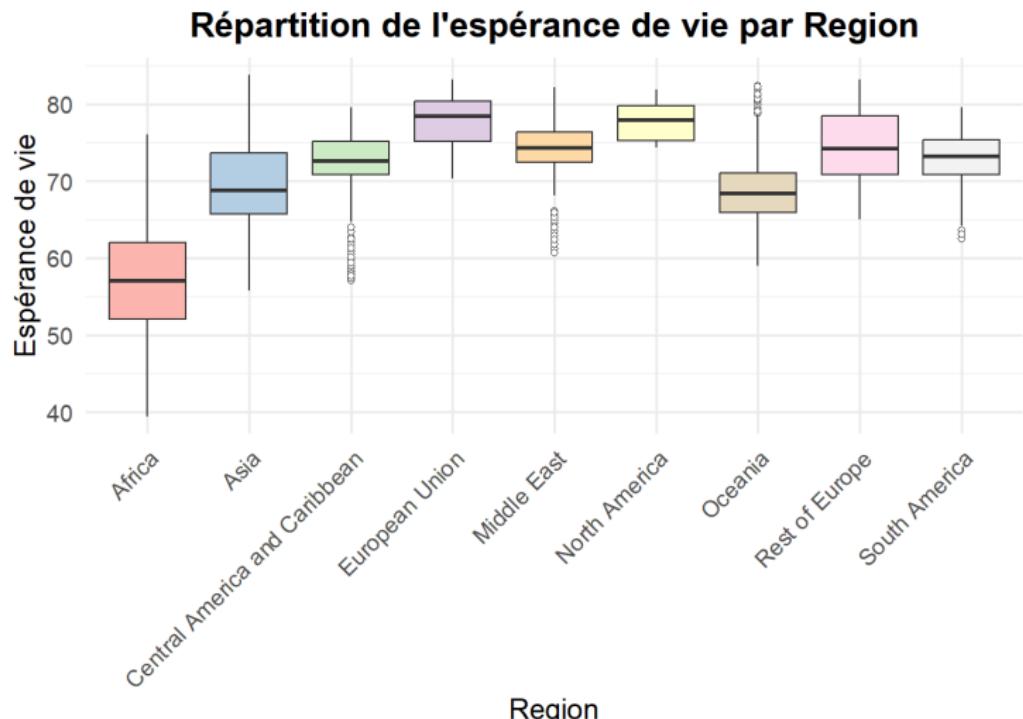


Figure: Répartition des résidus : $r_i = y_i - \hat{y}_i$ en fonction des \hat{y}_i

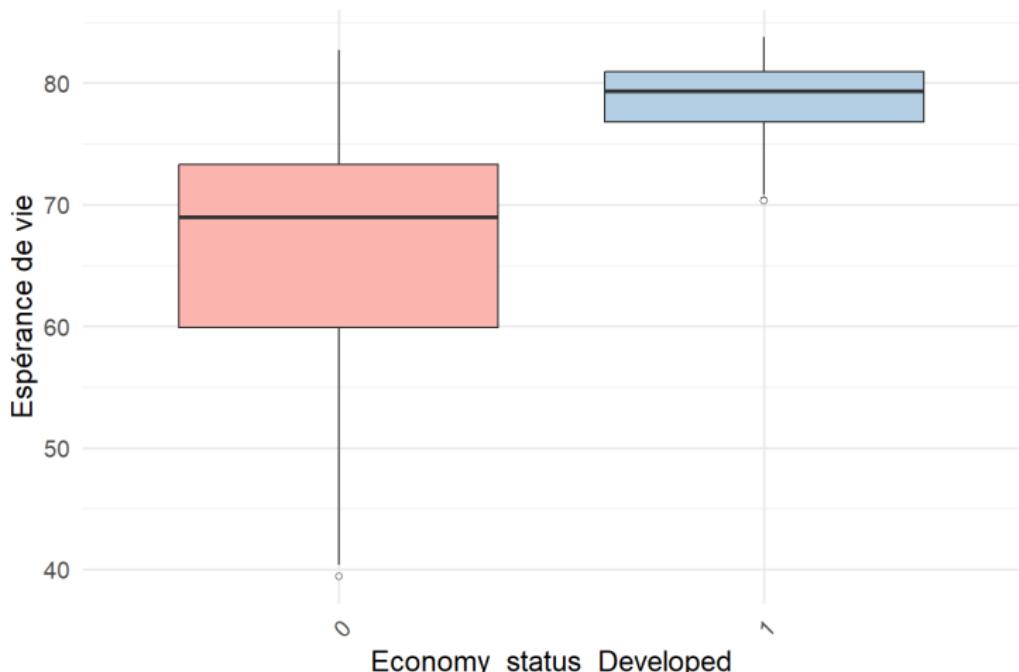
- Observation n°226 (Haïti en 2010 : séisme magnitude 7, plus de 300k morts)

Statistique descriptive sur le dataset Life_expectancy



Statistique descriptive sur le dataset Life_expectancy

Répartition de l'espérance de vie par Economy_status_Develo



Statistique descriptive sur le dataset Life_expectancy

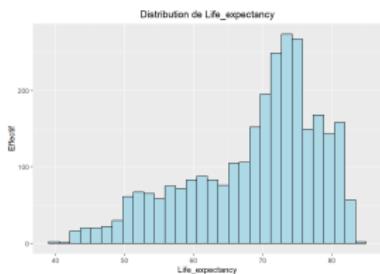


Figure: Histogramme de Life_expectancy

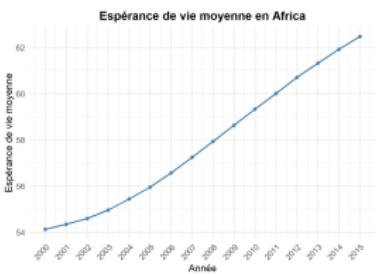


Figure: évolution de Life_expectancy au cours du temps en Afrique

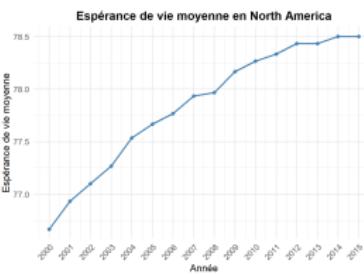
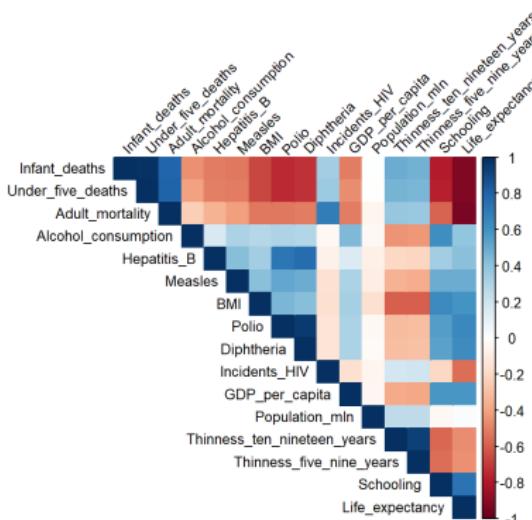


Figure: évolution de Life_expectancy au cours du temps en Amérique du Nord

Modélisation : approche globale

Approche globale

On considère toutes les années et tous les pays en même temps et on exclu les variables country, Region et Year.



Variable	Correlation
Schooling	0.732
Polio	0.641
Diphtheria	0.628
BMI	0.598
GDP_per_capita	0.583
Measles	0.490
Hepatitis_B	0.418
Alcohol_consumption	0.399
Population_min	0.329
Thinnness_five_nine_years	-0.468
Thinnness_ten_nineteen_years	-0.458
Incidents_HIV	-0.553
Infant_deaths	-0.620
Under_five_deaths	-0.920
Adult_mortality	-0.945

Figure: Coefficients de corrélation de Pearson.

Figure: Matrice de corrélation entre variables quantitatives.

Modélisation : approche globale



Sélection automatique de sous-modèles

Les trois méthodes de sélection aboutissent au même modèle où les variables Measles, Polio, Diphtheria, Population_mln et Thinness_five_nine_years ont été éliminées.

```
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 8.485e+01  5.441e-01 155.947 < 2e-16 ***
## Adult_mortality            -4.799e-02  6.123e-04 -78.373 < 2e-16 ***
## Infant_deaths              -4.463e-02  6.135e-03 -7.274 4.47e-13 ***
## Economy_status_Developed1  6.634e-01  1.070e-01   6.202 6.40e-10 ***
## Under_five_deaths          -5.970e-02  3.760e-03 -15.879 < 2e-16 ***
## GDP_per_capita              2.604e-05  2.270e-06  11.475 < 2e-16 ***
## Alcohol_consumption         6.425e-02  9.822e-03   6.542 7.18e-11 ***
## Schooling                   9.082e-02  1.652e-02   5.498 4.17e-08 ***
## BMI                         -1.488e-01  1.840e-02  -8.087 8.94e-16 ***
## Incidents_HIV               9.377e-02  1.794e-02   5.226 1.85e-07 ***
## Thinness_ten_nineteen_years -3.293e-02  7.875e-03  -4.181 2.98e-05 ***
## Hepatitis_B                  -5.789e-03  1.889e-03  -3.065  0.0022 **
## 
## Residual standard error: 1.352 on 2851 degrees of freedom
## Multiple R-squared:  0.9794, Adjusted R-squared:  0.9793
## F-statistic: 1.234e+04 on 11 and 2851 DF,  p-value: < 2.2e-16
```

Figure: sortie R du résumé du modèle stepwise

Modélisation : approche globale

Multicolinéarité : VIFs

Variable	VIF
Infant_deaths	44.62
Under_five_deaths	43.81
Adult_mortality	7.76
Schooling	4.30
Economy_status_Developed	2.94
Incidents_HIV	2.86
BMI	2.55
Alcohol_consumption	2.40
GDP_per_capita	2.31
Thinness_ten_nineteen_years	1.91
Hepatitis_B	1.43

Variable	VIF
Adult_mortality	7.71
Under_five_deaths	6.60
Schooling	4.30
Economy_status_Developed	2.91
Incidents_HIV	2.83
BMI	2.55
Alcohol_consumption	2.35
GDP_per_capita	2.27
Thinness_ten_nineteen_years	1.90
Hepatitis_B	1.40

Figure: VIFs du modèle stepwise

Figure: VIFs après avoir supprimé la variable **infant_death**

Modélisation : approche globale



Analyse des résidus

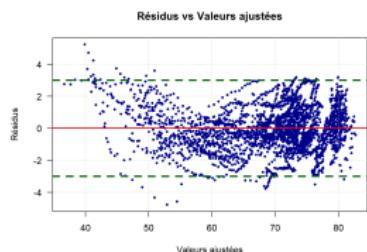


Figure: résidus en fonction des valeurs ajustées

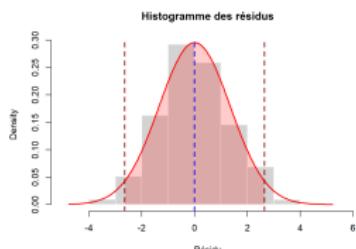


Figure: Histogramme des résidus

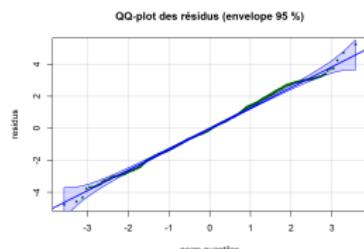


Figure: Q-Q plot des résidus

- Le test de **Jarque-Bera** passe avec une **p-value** de **23%** \Rightarrow résidus gaussiens.

Modélisation : approche globale



Méthodes de régularisation

Détermination des hyperparamètres \Rightarrow Validation croisée à 10-plis :

- **Ridge** : $\lambda_2^{\text{opt}} = 0.89$, MSE = 5.47
- **LASSO** : $\lambda_1^{\text{opt}} = 0.013$, MSE = 3.39
- **Elastic-Net** : $(\lambda^{\text{opt}}, \alpha^{\text{opt}}) = (0.011, 0.9)$, MSE = 3.41

Pour rappel :

$$\hat{\beta}^{\text{OLS}} = (X^\top X)^{-1} X^\top Y, \quad \hat{\beta}^{\text{Ridge}} = (X^\top X + \lambda_2 I_N)^{-1} X^\top Y.$$

Les conditionnement des matrices à inverser valent respectivement :

$$\kappa(X^\top X) \approx 2 \times 10^{17}, \quad \kappa(X^\top X + \lambda_2 I) \approx 204.$$

\Rightarrow Gain significatif de stabilité grâce à **Ridge** et sélection de variables via **LASSO** et **Elastic-Net**

Modélisation : approche globale

Bilan

Comparaison des coefficients des modèles

Variable	Sous-modèle	Ridge	LASSO	Elastic Net
(Intercept)	8.44e+01	6.89e+01	6.89e+01	6.89e+01
Adult_mortality	-4.83e-02	-3.86e+00	-5.42e+00	-5.43e+00
Economy_status_Developed1	7.40e-01	4.30e-01	2.83e-01	2.79e-01
Under_five_deaths	-8.49e-02	-2.10e+00	-2.51e+00	-2.52e+00
GDP_per_capita	2.83e-05	6.72e-01	4.47e-01	4.45e-01
Alcohol_consumption	7.43e-02	6.94e-02	2.45e-01	2.46e-01
Schooling	9.02e-02	3.91e-01	2.72e-01	2.76e-01
BMI	-1.50e-01	6.41e-02	-2.56e-01	-2.73e-01
Incidents_HIV	1.07e-01	-6.44e-01	1.45e-01	1.61e-01
Thinness_ten_nineteen_years	-3.77e-02	-8.55e-02	-1.07e-01	-1.15e-01
Hepatitis_B	-3.90e-03	-1.07e-01	-6.83e-02	-7.91e-02
Infant_deaths	NA	-1.92e+00	-1.37e+00	-1.37e+00
Measles	NA	9.03e-02	9.65e-03	1.29e-02
Polio	NA	1.72e-01	0.00e+00	5.98e-03
Diphtheria	NA	1.65e-01	0.00e+00	0.00e+00
Population_mln	NA	9.09e-02	0.00e+00	-2.22e-03
Thinness_five_nine_years	NA	3.29e-02	0.00e+00	0.00e+00

Figure: Récapitulatif des coefficients estimés pour les différents modèles

Modélisation : approche plus fine

Approche plus fine

Filtrage du jeu de données initial selon :

- **Statut économique** : Développé / En développement
- **Année** : 2000 / 2015

	dataset	economy_status	Year
(a)		developed	2000
(b)		developed	2015
(c)		developing	2000
(d)		developing	2015

Nombre de pays développés : 37

Nombre de pays en développement : 142

Modélisation : approche plus fine



Séparation des pays en fonction de leurs statut de développement économique

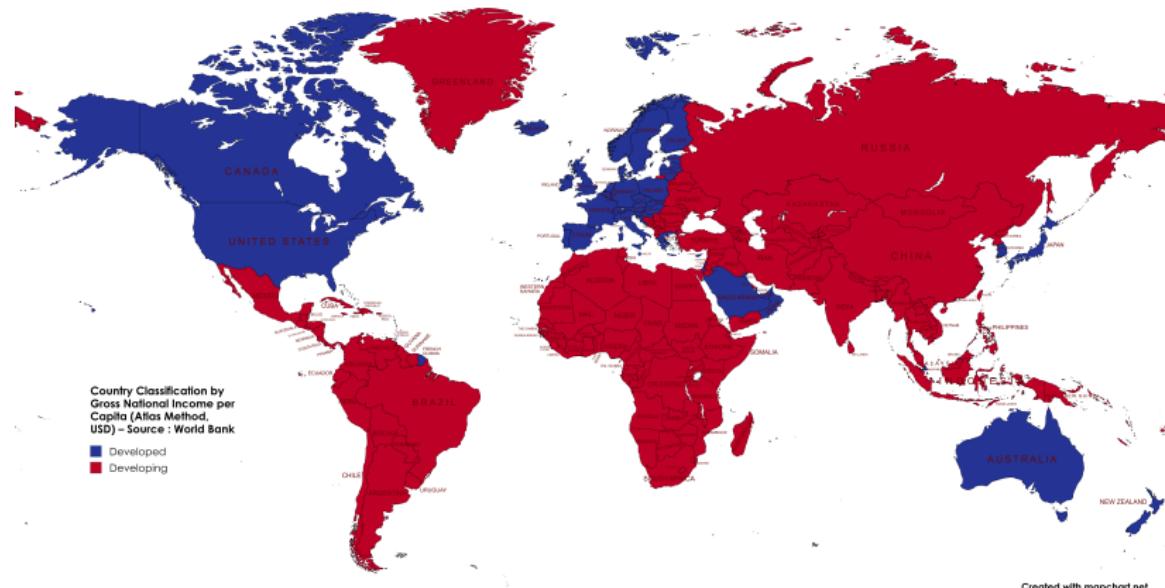


Figure: Répartitions des pays selon leurs revenu national brut par habitant moyen sur la période étudiée (seuil RNB/habitant $\approx 25000\$$)

Modélisation : approche plus fine

Estimation des coefficients pour les différents modèles

Variable	Sous-modèle	Ridge	LASSO	Elastic Net
(Intercept)	8.35e+01	7.66e+01	7.66e+01	7.66e+01
Adult_mortality	-5.43e-02	-1.44e+00	-2.20e+00	-2.07e+00
Under_five_deaths	-1.24e-01	-3.01e-01	0.00e+00	0.00e+00
GDP_per_capita	1.85e-05	3.22e-01	1.63e-01	1.26e-01
Alcohol_consumption	-1.03e-01	-4.41e-01	0.00e+00	0.00e+00
Incidents_HIV	7.69e+00	-2.91e-03	3.91e-02	0.00e+00
Infant_deaths	NA	-3.52e-01	-3.62e-01	-3.65e-01
Hepatitis_B	NA	-2.37e-01	-7.05e-03	0.00e+00
Measles	NA	-6.36e-02	0.00e+00	0.00e+00
BMI	NA	-3.06e-01	0.00e+00	0.00e+00
Polio	NA	9.33e-02	0.00e+00	0.00e+00
Diphtheria	NA	-2.01e-01	0.00e+00	0.00e+00
Population_min	NA	1.86e-01	0.00e+00	0.00e+00
Thinness_ten_nineteen_years	NA	-3.03e-01	0.00e+00	0.00e+00
Thinness_five_nine_years	NA	-1.96e-01	0.00e+00	0.00e+00
Schooling	NA	-5.21e-02	0.00e+00	0.00e+00

Variable	Sous-modèle	Ridge	LASSO	Elastic Net
(Intercept)	1.03e+02	8.02e+01	8.02e+01	8.02e+01
Adult_mortality	-6.86e-02	-1.42e+00	-2.07e+00	-2.03e+00
Under_five_deaths	-1.51e-01	-2.24e-01	0.00e+00	0.00e+00
Thinness_five_nine_years	-8.55e-01	-2.34e-01	-3.14e-01	-2.21e-02
BMI	-3.41e-01	-3.72e-01	-2.30e-01	-2.76e-01
Polio	-7.91e-02	-1.60e-02	0.00e+00	0.00e+00
Incidents_HIV	7.85e+00	1.23e-01	1.70e-01	2.04e-01
Infant_deaths	NA	-3.16e-01	-2.05e-01	-2.42e-01
Alcohol_consumption	NA	-2.15e-01	0.00e+00	-1.13e-03
Hepatitis_B	NA	-1.40e-01	-6.90e-02	-1.02e-01
Measles	NA	-6.69e-02	0.00e+00	0.00e+00
Diphtheria	NA	-6.27e-02	0.00e+00	0.00e+00
GDP_per_capita	NA	2.22e-01	1.12e-01	1.12e-01
Population_min	NA	2.06e-01	1.26e-01	1.66e-01
Thinness_ten_nineteen_years	NA	-3.98e-01	-4.41e-02	-3.98e-01
Schooling	NA	-8.43e-02	0.00e+00	0.00e+00

(a) Developed - 2000

(b) Developed - 2015

Modélisation : approche plus fine

Estimation des coefficients pour les différents modèles

Variable	Sous-modèle	Ridge	LASSO	Elastic Net
(Intercept)	7.93e+01	6.37e+01	6.37e+01	6.37e+01
Under_five_deaths	-8.77e-02	-2.55e+00	-3.91e+00	-3.84e+00
Adult_mortality	-4.07e-02	-3.98e+00	-5.13e+00	-5.13e+00
Alcohol_consumption	1.51e-01	2.32e-01	3.45e-01	3.44e-01
Infant_deaths	NA	-2.04e+00	-8.58e-01	-9.28e-01
Hepatitis_B	NA	-5.58e-02	0.00e+00	0.00e+00
Measles	NA	9.85e-02	0.00e+00	0.00e+00
BMI	NA	1.63e-01	0.00e+00	0.00e+00
Polio	NA	3.40e-01	7.25e-02	7.68e-02
Diphtheria	NA	3.38e-01	1.64e-01	1.67e-01
Incidents_HIV	NA	-7.07e-01	0.00e+00	0.00e+00
GDP_per_capita	NA	2.76e-01	7.74e-02	7.79e-02
Population_min	NA	1.46e-01	0.00e+00	0.00e+00
Thinness_ten_nineteen_years	NA	-3.06e-01	-6.18e-02	-6.55e-02
Thinness_five_nine_years	NA	2.45e-01	0.00e+00	0.00e+00
Schooling	NA	2.10e-01	0.00e+00	4.71e-04

(c) Developing - 2000

Variable	Sous-modèle	Ridge	LASSO	Elastic Net
(Intercept)	8.06e+01	6.92e+01	6.92e+01	6.92e+01
Adult_mortality	-4.62e-02	-3.04e+00	-3.94e+00	-3.91e+00
Under_five_deaths	-9.30e-02	-1.77e+00	-2.70e+00	-2.56e+00
Alcohol_consumption	1.42e-01	2.20e-01	3.12e-01	2.91e-01
GDP_per_capita	4.01e-05	4.14e-01	2.97e-01	2.88e-01
Infant_deaths	NA	-1.52e+00	-3.14e-01	-4.71e-01
Hepatitis_B	NA	2.55e-02	0.00e+00	0.00e+00
Measles	NA	2.50e-01	1.30e-01	1.26e-01
BMI	NA	4.37e-02	0.00e+00	0.00e+00
Polio	NA	-2.64e-02	0.00e+00	0.00e+00
Diphtheria	NA	6.97e-02	0.00e+00	0.00e+00
Incidents_HIV	NA	-3.14e-01	0.00e+00	0.00e+00
Population_min	NA	1.30e-02	0.00e+00	0.00e+00
Thinness_ten_nineteen_years	NA	-9.63e-02	0.00e+00	0.00e+00
Thinness_five_nine_years	NA	1.75e-01	0.00e+00	0.00e+00
Schooling	NA	2.25e-01	1.53e-02	2.23e-02

(d) Developing - 2015

Modélisation : approche plus fine

Merci pour votre attention !

Modélisation : approche plus fine



Références

-  Rüping, Stefan (Oct. 20, 2006). "Learning Interpretable Models". In.
-  Tibshirani, R. (1996). "Regression Shrinkage and Selection via the Lasso". In: *Journal of the royal statistical society* 58.1, pp. 267–288.
-  Tibshirani, Robert (2011). "Regression Shrinkage and Selection via the Lasso: A Retrospective". In: *Journal of the royal statistical society* 73.3, pp. 273–282.