



# Projet Micro-économétrie

Analyse des determinants socio-économiques de l'espérance de vie (Panel 2000-2015)

Ibrahima Caba Bah

2025-12-23

## Contents

<b>1 Présentation de la problématique Recherche</b>	<b>2</b>
<b>2 Presentation des données</b>	<b>2</b>
2.1 Variables retenues pour l'analyse . . . . .	2
<b>3 Statistiques descriptives</b>	<b>2</b>
3.1 Corrélation . . . . .	4
3.2 Visualisation . . . . .	5
<b>4 Méthodes d'Estimations</b>	<b>7</b>
4.1 Modèle Poolé . . . . .	7
4.2 Effets Fixes . . . . .	9
4.3 Modèle à effet aléatoire . . . . .	10
<b>5 Comparaison globale des modèles</b>	<b>12</b>
<b>6 Test de sélections</b>	<b>12</b>
6.1 Test de Hausman . . . . .	12
6.2 Test de Breusch-Pagan . . . . .	13
6.3 Test de Fisher . . . . .	13
<b>7 Conclusion</b>	<b>14</b>

# 1 Présentation de la problématique Recherche

Dans ce projet, nous cherchons à répondre à la question de recherche suivante : **Quels sont les déterminants socio-économiques et sanitaires de l'espérance de vie , et comment la richesse nationale, l'éducation et le statut de développement influencent-ils la longévité des populations ?**

L'objectif principal est de comprendre les facteurs majeurs qui allongent ou réduisent la durée de vie et d'évaluer

Pour répondre à cette question, nous utilisons une base de données de panel couvrant la période 2000-2015. Notre méthodologie repose sur la comparaison de trois approches économétriques complémentaires :

- Le modèle Poolé (MCO), pour observer les tendances globales entre les pays.
- Le modèle à Effets Fixes (Within), Il se concentre sur l'évolution interne de chaque pays en supprimant les facteurs qui ne changent pas, comme la culture ou la géographie.
- Le modèle à Effets Aléatoires, pour intégrer les variables invariantes comme le statut de développement.

L'arbitrage entre ces modèles, guidé par le test de Hausman, nous permettra de conclure sur les leviers les plus fiables pour améliorer la santé publique mondiale.

## 2 Presentation des données

La base de données a été obtenu via la plateforme **kaggle(Life Expectancy (WHO))**. Ce jeu de données rassemble des indicateurs produits par le Global Health Observatory(GHO) de l'organisation Mondiale de la santé (OMS) et par les Nations Unies .

L'échantillon comporte 2938 observation reparties sur 193 pays.

### 2.1 Variables retenues pour l'analyse

Afin d'étudier l'impact des facteurs économiques et sanitaires sur l'espérance de vie, nous avons retenu les variables suivantes :

- Pays : Pays concerné par l'observation
- Année : Année de l'observation (couvrant la période 2000-2015)
- esp\_vie : Esperance de vie à la naissance
- Pib : Le Produit Intérieur Brut (PIB) par habitant, représentant la richesse économique du pays.
- Educ : Le nombre moyen d'années de scolarisation de la population.
- Hiv : Taux de mortalité lié au Vih/Sida
- Imc : L'Indice de Masse Corporelle moyen de la population
- Polio :Le taux de couverture vaccinale contre la poliomylérite
- Statut : Statut de développement économique ( 0 = Pays en développement et 1 = Pays développé )

## 3 Statistiques descriptives

```
##      Pays           Année       esp_vie          Pib
##  Length:2450      Min.   :2000   Min.   :36.30  Min.   :1.681e+00
##  Class :character  1st Qu.:2004   1st Qu.:63.42  1st Qu.:4.652e+02
```

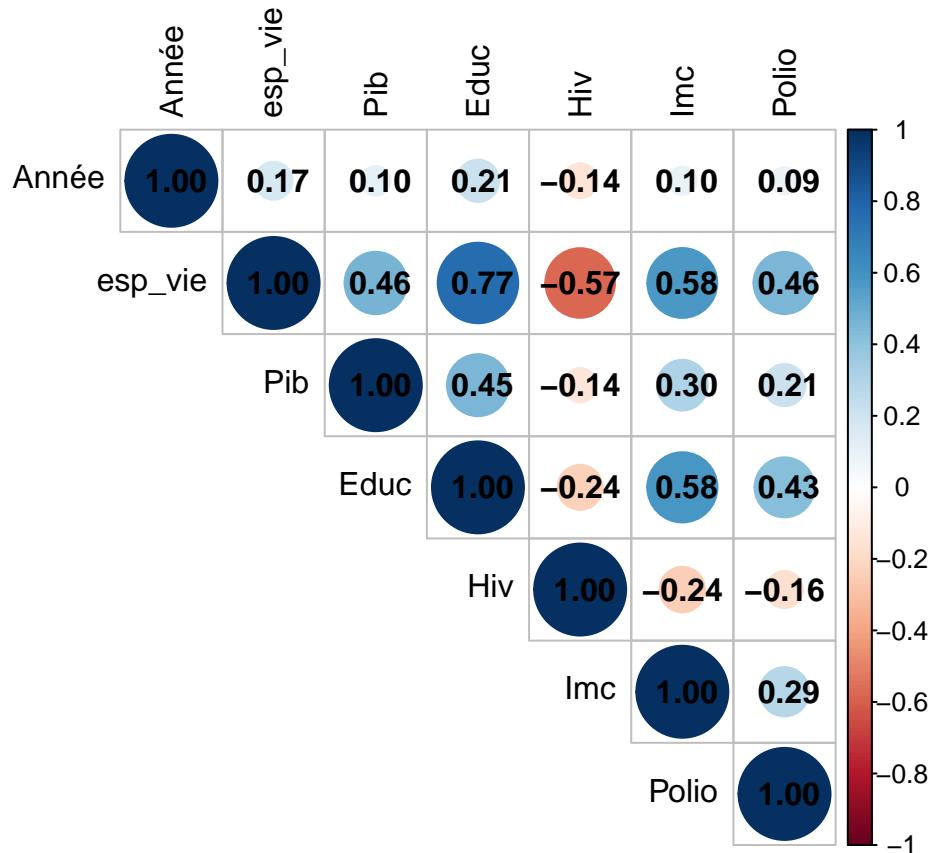
```

##   Mode :character Median :2007 Median :72.30 Median :1.815e+03
##               Mean :2007 Mean :69.48 Mean :7.586e+03
##               3rd Qu.:2011 3rd Qu.:76.00 3rd Qu.:6.266e+03
##               Max. :2015  Max. :89.00  Max. :1.192e+05
##      Educ          Hiv          Imc          Polio
## Min. : 0.00  Min. : 0.100  Min. : 1.40  Min. : 3.00
## 1st Qu.:10.20 1st Qu.: 0.100 1st Qu.:19.20 1st Qu.:78.00
## Median :12.50 Median : 0.100 Median :43.80 Median :93.00
## Mean  :12.18  Mean  : 1.901 Mean  :38.35 Mean  :82.73
## 3rd Qu.:14.50 3rd Qu.: 0.800 3rd Qu.:56.20 3rd Qu.:97.00
## Max. :20.70  Max. :50.600  Max. :77.60  Max. :99.00
##      Statut
## Length:2450
## Class :character
## Mode  :character
##
##
```

En regardant les outils de statistiques descriptives ci dessus, on remarque que :

- esp\_vie : L'espérance de vie varie de 36,3 à 89 ans, ce qui montre des inégalités sanitaires extrêmes entre les pays, avec une moyenne globale de 69,5 ans.
- Pib : On a une moyenne de 7 586, mais un écart énorme entre le minimum (1,6) et le maximum (119 200). Cette variable reflète la richesse nationale et la capacité à financer les hôpitaux.
- Educ : Varie de 0 à 20,7 ans, avec une moyenne de 12,2 ans. C'est un indicateur clé, car le niveau d'instruction permet souvent de mieux comprendre les règles d'hygiène et de prévention.
- Hiv : On observe une moyenne basse de 1,9, mais un pic effrayant à 50,6 dans le pays le plus touché. Cela montre que l'épidémie reste très localisée et dévastatrice pour certaines populations.
- Polio : Le taux de vaccination varie de 3 % à 99 %, avec une bonne moyenne de 82,7 %. Cela témoigne de l'efficacité des campagnes de santé publique mondiales.

### 3.1 Corrélation

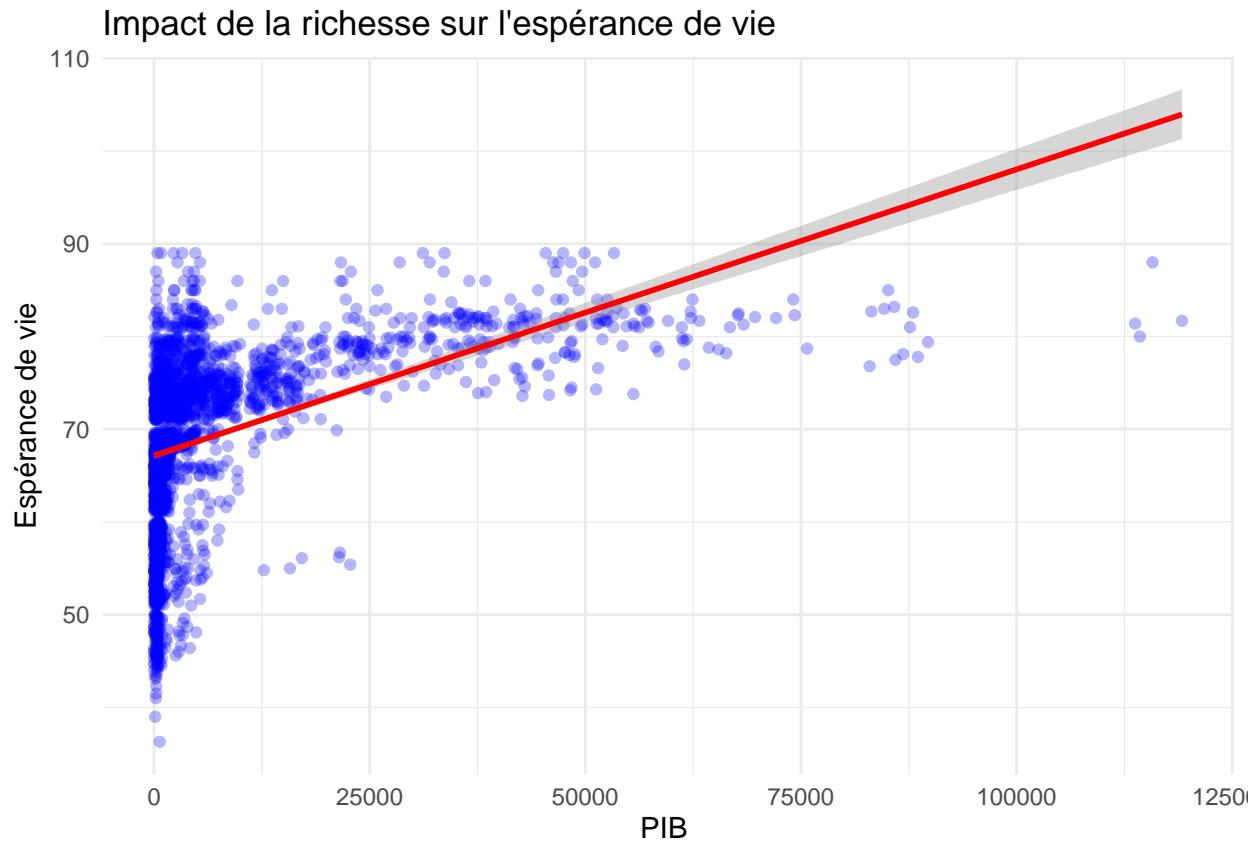


La matrice de corrélation nous montre que :

- L'espérance de vie et l'éducation (Educ) présentent la corrélation positive la plus forte du tableau (0.77). Ce lien très puissant indique que le niveau d'éducation est le facteur le plus déterminant pour augmenter la longévité, bien avant la richesse économique (Pib, 0.46).
- À l'inverse, il y a un lien négatif entre l'espérance de vie et le VIH : plus le virus est présent, moins l'espérance de vie est élevée

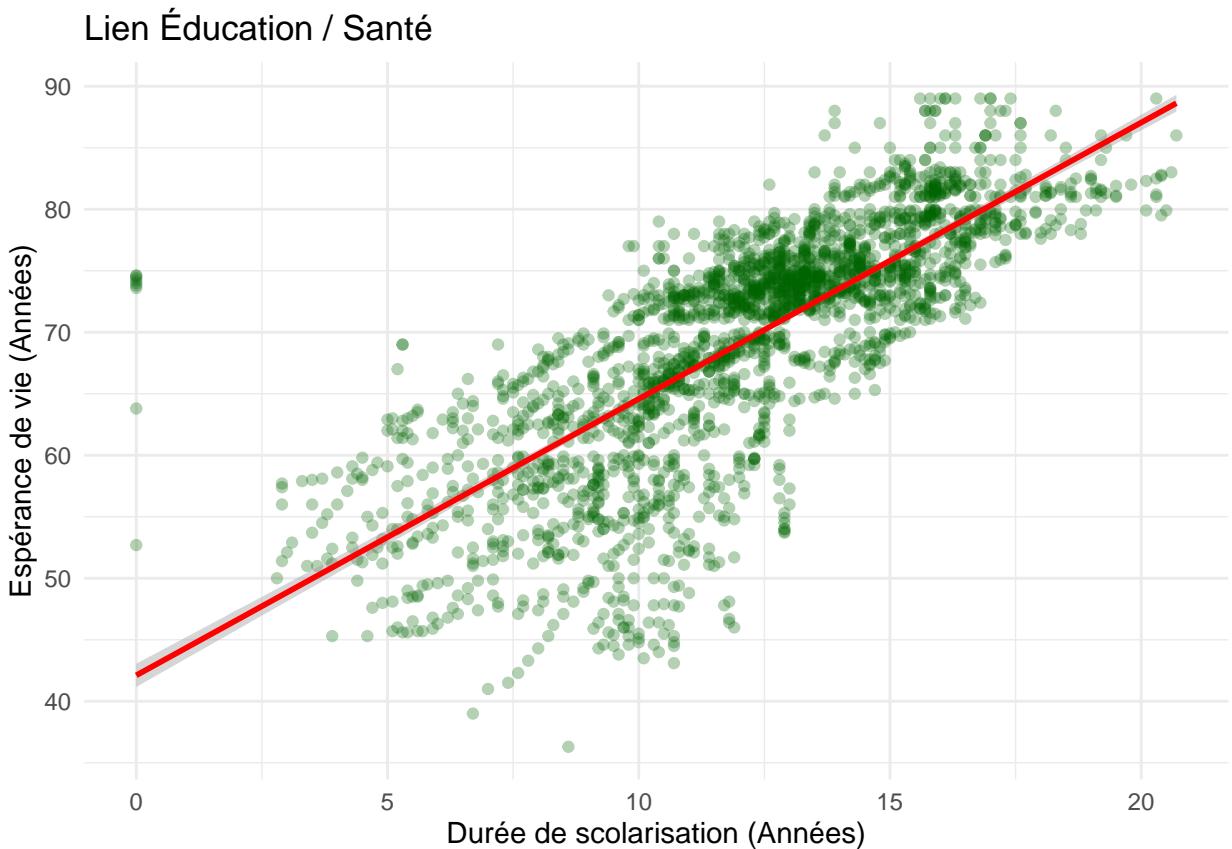
## 3.2 Visualisation

### 3.2.1 Relation entre PIB et Esperance de Vie



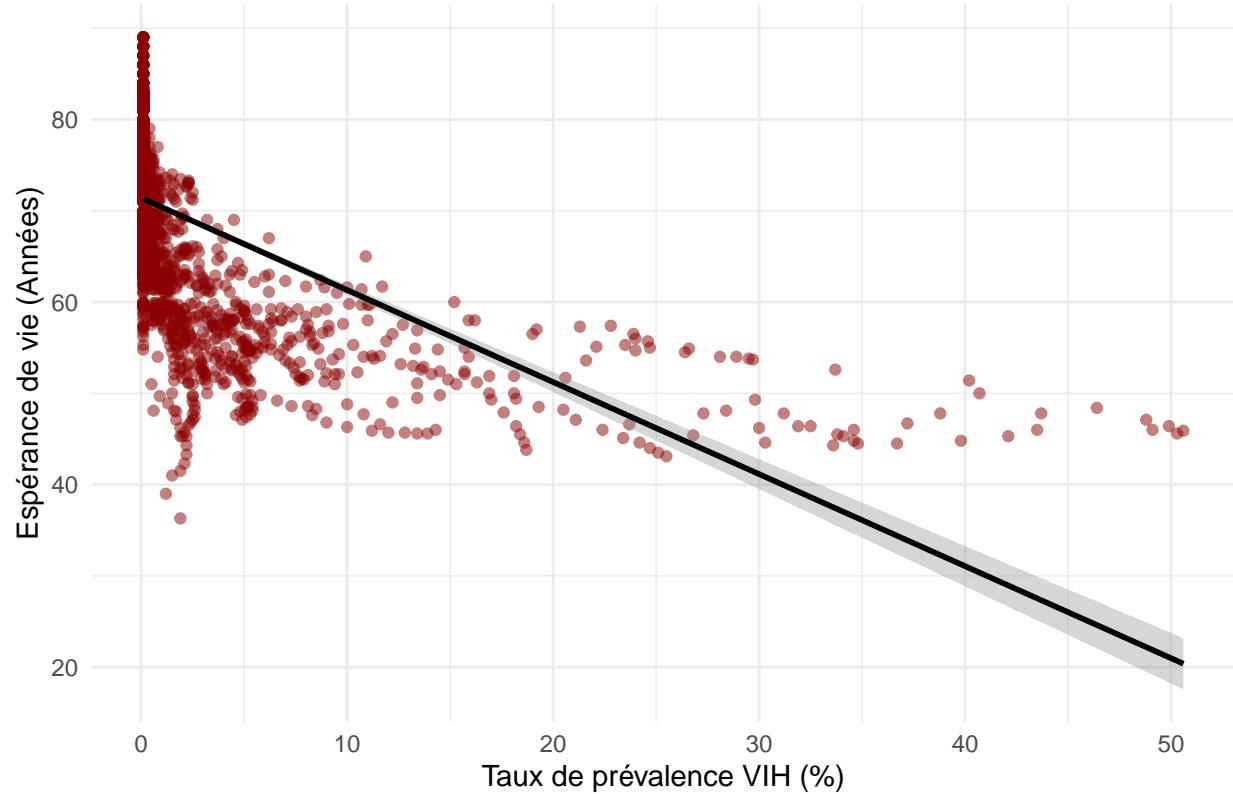
Le nuage de points montre une relation **linéaire et positive** : l'espérance de vie augmente de manière constante avec la durée de scolarisation. Cela confirme que l'accès à l'éducation est un levier stable et puissant pour améliorer la santé publique à tous les niveaux.

### 3.2.2 Nuage de points : Éducation vs Espérance de vie



Le graphique montre une corrélation positive claire : la droite de régression ascendante indique que globalement, plus un pays est riche, plus son espérance de vie est élevée. Même si les points sont dispersés, la tendance générale confirme que le développement économique reste un facteur clé de l'amélioration de la santé.

## Lien négatif : VIH et Santé



Contrairement aux deux précédents, ce graphique met en évidence une corrélation négative marquée. La droite de régression descendante illustre visuellement le coefficient de -0.57 calculé précédemment : plus le taux de prévalence du VIH est élevé dans un pays, plus l'espérance de vie chute drastiquement. Ce graphique agit comme le "miroir inversé" du développement : il montre comment une crise sanitaire majeure peut annuler les gains obtenus par l'éducation ou la richesse

## 4 Méthodes d'Estimations

### 4.1 Modèle Poolé

On applique le McO normalement sans tenir compte des données de panel et aux différentes dimensions. Ce modèle suppose que tous les pays sont identiques et ignore l'hétérogénéité inobservée qui peut être corrélée aux variables explicatives

```
modele_poolé<- plm(esp_vie ~ Pib + Educ + Hiv + Polio + Statut, data = base, model = "pooling", effect = summary(modele_poolé)
```

```
## Pooling Model
##
## Call:
## plm(formula = esp_vie ~ Pib + Educ + Hiv + Polio + Statut, data = base,
##       effect = "individual", model = "pooling", index = c("Pays",
##                 "Année"))
##
```

```

## Unbalanced Panel: n = 155, T = 8-16, N = 2450
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.    Max.
## -25.89967 -2.83325  0.14435  3.08337 23.50793
##
## Coefficients:
##                               Estimate Std. Error t-value Pr(>|t|)
## (Intercept)        4.7975e+01 6.1865e-01 77.5479 < 2.2e-16 ***
## Pib                7.1404e-05 7.6043e-06  9.3899 < 2.2e-16 ***
## Educ               1.5737e+00 3.7298e-02 42.1939 < 2.2e-16 ***
## Hiv                -7.0392e-01 1.7444e-02 -40.3524 < 2.2e-16 ***
## Polio              5.2492e-02 4.4281e-03 11.8542 < 2.2e-16 ***
## StatutDeveloping -1.4839e+00 2.9778e-01 -4.9833 6.69e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:  227600
## Residual Sum of Squares: 51146
## R-Squared: 0.77527
## Adj. R-Squared: 0.77481
## F-statistic: 1686.3 on 5 and 2444 DF, p-value: < 2.22e-16

```

Ce modèle nous donne une première estimation naïve car il mélange tous les pays et toutes les années sans distinction.

## 1. Interprétation des coefficients

**Educ** : C'est le facteur le plus déterminant. En moyenne, chaque année de scolarisation supplémentaire est associée à une hausse de l'espérance de vie de 1,57 an. Cela montre que l'accès au savoir est le principal moteur de la santé.

**Hiv** : Le coefficient est de -0,704. Cela signifie qu'une augmentation de 1 unité de cette variable (taux de mortalité ou prévalence) entraîne une diminution de l'espérance de vie d'environ 0,70 an (soit environ 8 mois et demi), toutes choses égales par ailleurs.

**Pib** : La richesse a un effet positif, mais très faible mathématiquement (coefficient de 0,00003). Cela suggère que si l'argent compte, son impact brut est bien moindre que celui des facteurs sociaux.

**Polio** : une meilleure couverture contre la polio augmente l'espérance de vie de 0,04 an par point de pourcentage.

**StatutDeveloping** : Le coefficient est de -1,48. Cela signifie que, à niveau de richesse, d'éducation et de santé égal, le simple fait d'être un "pays en développement" réduit l'espérance de vie d'environ 1 an et demi par rapport à un pays développé.

## 2. Significativité des variables

L'examen des p-values montre que nos quatre variables principales (Pib, Educ, Hiv et Polio) sont toutes significatives au seuil de 1 %. Cela confirme qu'elles ont un impact réel et statistiquement fiable sur la durée de vie. En revanche, le statut "en développement" n'apparaît pas comme un facteur explicatif majeur une fois que l'on prend déjà en compte l'éducation et la richesse.

Le coefficient de détermination ( $R^2$ ) est de 0,775, montrant que le modèle capture 77,5 % de la variance de l'espérance de vie.

F-statistic (1344) : La valeur est très élevée avec une p-value proche de zéro. Cela valide la pertinence globale du modèle : les variables choisies expliquent la variance de l'espérance de vie bien mieux que le simple hasard.

Cette première estimation montre que l'éducation et la lutte contre les épidémies (VIH) sont les piliers de la survie humaine. Si le revenu (Pib) est significatif, il ne suffit pas à garantir une vie longue s'il n'est pas accompagné de politiques éducatives et sanitaires.

## 4.2 Effets Fixes

Les modèles à effets fixes supposent que les caractéristiques inobservables propres à chaque pays sont corelées aux variables explicatives

### 4.2.1 Within

Ce modèle, corrige les limites du modèle poolé en prenant en compte l'hétérogénéité inobservée des pays ( $c_i$ )

```
modele_within <- plm(esp_vie ~ Pib + Educ + Hiv + Polio + Statut, data = base, model = "within", index = c("Pays", "Année"))
summary(modele_within)

## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = esp_vie ~ Pib + Educ + Hiv + Polio + Statut, data = base,
##       model = "within", index = c("Pays", "Année"))
##
## Unbalanced Panel: n = 155, T = 8-16, N = 2450
##
## Residuals:
##      Min.    1st Qu.     Median    3rd Qu.     Max.
## -24.16091 -0.88498 -0.15307  0.58463  10.64426
##
## Coefficients:
##             Estimate Std. Error t-value Pr(>|t|)
## Pib      1.2651e-05 4.4824e-06  2.8223 0.0048093 ***
## Educ     8.6243e-01 3.8867e-02 22.1893 < 2.2e-16 ***
## Hiv     -4.3163e-01 1.5719e-02 -27.4592 < 2.2e-16 ***
## Polio    8.0758e-03 2.4399e-03   3.3099 0.0009478 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    16282
## Residual Sum of Squares: 9824.2
## R-Squared:               0.39661
## Adj. R-Squared:          0.355
## F-statistic: 376.473 on 4 and 2291 DF, p-value: < 2.22e-16
```

Dans ce modèle, la variable Statut (pays développé ou en développement) a été automatiquement exclue de l'estimation car le modèle à effets fixes ne peut analyser que ce qui change dans le temps pour un même pays. Comme le statut d'une nation (développée ou non) est resté identique durant toute la période de l'étude (2000-2015), cette variable est considérée comme "invariante". Elle est donc absorbée par l'effet fixe propre à chaque pays.

**Educ** : Si la durée de scolarisation augmente d'un an au sein d'un pays, l'espérance de vie gagne en moyenne 0,86 an.

**Hiv** : Une hausse de 1 % de la prévalence du virus entraîne une chute de l'espérance de vie de 0,43 an.

**Polio** : Une amélioration de 1 % de la couverture vaccinale se traduit par un gain de longévité de 0,008 an.

**Pib**: Le coefficient de 0,000012 confirme que l'augmentation du revenu national a un effet positif, bien que marginal, sur la santé à court terme au sein d'un même pays.

Le modèle explique environ 39,6 % de la variance interne.

### 4.3 Modèle à effet aléatoire

Contrairement aux effets fixes qui supposent que l'effet individuel est corrélé avec les régresseurs, ce modèle suppose que ces effets sont purement aléatoires. Il permet donc d'intégrer et d'analyser des variables qui ne changent pas dans le temps comme le "Statut" de développement.

```
modele_aleatoire <- plm(esp_vie ~ Pib + Educ + Hiv + Polio + Statut, data = base, model = "random", index = c("Pays", "Année"))
summary(modele_aleatoire)

## Oneway (individual) effect Random Effect Model
##      (Swamy-Arora's transformation)
##
## Call:
## plm(formula = esp_vie ~ Pib + Educ + Hiv + Polio + Statut, data = base,
##       model = "random", index = c("Pays", "Année"))
##
## Unbalanced Panel: n = 155, T = 8-16, N = 2450
##
## Effects:
##           var std.dev share
## idiosyncratic 4.288   2.071 0.246
## individual    13.122   3.622 0.754
## theta:
##   Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.8019 0.8585 0.8585 0.8578 0.8585 0.8585
##
## Residuals:
##   Min. 1st Qu. Median Mean 3rd Qu. Max.
## -24.79820 -1.10814 -0.01642 -0.00203 0.79073 11.36133
##
## Coefficients:
##             Estimate Std. Error z-value Pr(>|z|)
## (Intercept) 6.2796e+01 9.5012e-01 66.0929 < 2.2e-16 ***
## Pib          1.5551e-05 4.5952e-06  3.3843 0.0007137 ***
## Educ         9.6775e-01 3.8192e-02 25.3391 < 2.2e-16 ***
## Hiv          -4.5343e-01 1.5766e-02 -28.7593 < 2.2e-16 ***
## Polio        1.0091e-02 2.5058e-03   4.0271 5.648e-05 ***
## StatutDeveloping -6.3278e+00 8.0988e-01  -7.8132 5.575e-15 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares: 20831
## Residual Sum of Squares: 11173
```

```

## R-Squared:      0.46364
## Adj. R-Squared: 0.46254
## Chisq: 2046.3 on 5 DF, p-value: < 2.22e-16

```

Le modèle à effets aléatoires propose une approche intermédiaire : il considère que les différences entre les pays ne sont pas fixes, mais qu'elles résultent d'un processus aléatoire. Contrairement au modèle précédent, il permet de réintroduire des variables qui ne changent pas dans le temps.

## 1. Interprétation des coefficients

**Éducation (Educ)** : Le coefficient est de 0,96. Cela signifie qu'une année d'étude supplémentaire est corrélée à un gain d'espérance de vie de près d'un an. Ce résultat est plus proche du modèle à effets fixes que du modèle poolé.

**VIH (Hiv)** : L'impact reste fortement négatif avec un coefficient de -0,45. On retrouve ici une stabilité de l'effet du virus sur la santé mondiale.

**Richesse (Pib)** : Le coefficient est de 0,00002. Bien que positif et significatif, l'impact financier demeure très faible par rapport aux variables sociales.

**Statut (StatutDeveloping)** : À l'inverse du modèle à effets fixes, cette variable est ici conservée. Elle affiche un coefficient de -6,32, indiquant que, toutes choses égales par ailleurs, vivre dans un pays en développement réduit l'espérance de vie de plus de 6 ans par rapport aux pays développés.

## 2. Significativité et Qualité du modèle

Toutes les variables explicatives sont significatives à 1 %. Le  $R^2$  global est de 0,464, ce qui signifie que le modèle explique environ 46 % de la variance totale. La statistique de Fisher est également très élevée (2046), confirmant la validité statistique de cette approche.

Le modèle à effets aléatoires semble séduisant car il combine les avantages du modèle poolé (conservation de la variable Statut) et la précision du modèle à effets fixes. Il suggère que les différences entre les pays sont des perturbations aléatoires qui ne sont pas liées à nos variables explicatives.

## 5 Comparaison globale des modèles

	<i>Dependent variable:</i>		
	esp_vie		
	(1)	(2)	(3)
Pib	0.00001*** (0.00001)	0.00001*** (0.00000)	0.000002*** (0.00000)
Educ	1.574*** (0.037)	0.862*** (0.039)	0.968*** (0.038)
Hiv	-0.704*** (0.017)	-0.432*** (0.016)	-0.453*** (0.016)
Polio	0.052*** (0.004)	0.008*** (0.002)	0.010*** (0.003)
StatutDeveloping	-1.484*** (0.298)		-6.328*** (0.810)
Constant	47.975*** (0.619)		62.796*** (0.950)
Observations	2,450	2,450	2,450
R <sup>2</sup>	0.775	0.397	0.464
Adjusted R <sup>2</sup>	0.775	0.355	0.463
F Statistic	1,686.300*** (df = 5; 2444)	376.473*** (df = 4; 2291)	2,046.303***

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Ce tableau nous permet de comparer les coefficients obtenus via les trois approches (Poolé, Effets Fixes et Effets Aléatoires) avant de procéder aux tests de sélection.

## 6 Test de sélections

Après avoir estimé les trois modèles (Poolé, Effets Fixes et Effets Aléatoires), nous constatons des différences dans les coefficients et les résultats. Pour déterminer rigoureusement quel modèle est le plus fiable et le mieux adapté à nos données, nous devons procéder à une série de tests statistiques de spécification

### 6.1 Test de Hausman

Ce test est utilisé pour décider s'il faut choisir les Effets Fixes ou les Effets Aléatoires.

H0 = les effets individuels ci ne sont pas corrélés aux variables explicatives

H1 = les effets individuels ci sont corrélés aux variables explicatives

```
hausman_test <- phtest(modele_within,modele_aléatoire)
hausman_test
```

```

## Hausman Test
##
## data: esp_vie ~ Pib + Educ + Hiv + Polio + Statut
## chisq = 65.332, df = 4, p-value = 2.19e-13
## alternative hypothesis: one model is inconsistent

```

Le test de Hausman nous donne une **p-value de  $2,19 \times 10^{-13}$** , ce qui est quasiment égal à zéro et donc bien inférieur au seuil de 5 %. Ce résultat nous force à rejeter le modèle à effets aléatoires, car il est jugé “inconsistent” ou non fiable dans ce cas précis.

## 6.2 Test de Breusch-Pagan

Ce test est utilisé pour vérifier la présence d’effets spécifiques aux pays et déterminer si le modèle Poolé est suffisant ou s’il faut passer à un modèle de panel.

```

library(lmtest)
bptest(esp_vie ~ Pib + Educ + Hiv + Polio + Statut+factor(Pays)+factor(Année), data=base)

##
## studentized Breusch-Pagan test
##
## data: esp_vie ~ Pib + Educ + Hiv + Polio + Statut + factor(Pays) +      factor(Année)
## BP = 242.45, df = 173, p-value = 0.0003887

```

H0 = pas d’effets de panel, tous les pays ont le même ci

H1 = effets de panels à prendre en compte

Le test de Breusch-Pagan affiche une **p-value de 0,0003887**, un résultat bien inférieur au seuil critique de 5 %. Cette valeur nous oblige à rejeter l’idée que tous les pays se ressemblent et à confirmer l’existence d’effets de panel significatifs.

## 6.3 Test de Fisher

Ce test permet de comparer le modèle à Effets Fixes (Within) et le modèle Poolé

```

Fisher <- pFtest(modele_within, modele_poolé)
Fisher

##
## F test for individual effects
##
## data: esp_vie ~ Pib + Educ + Hiv + Polio + Statut
## F = 62.982, df1 = 153, df2 = 2291, p-value < 2.2e-16
## alternative hypothesis: significant effects

```

H0 = pas d’effets de panel

H1 = effets de panel a prendre en compte

La **p-value < 2,2e-16**, ce qui est bien en dessous du seuil 5 %. Ce résultat nous pousse à rejeter l’hypothèse nulle et confirme la présence d’effets individuels significatifs propres à chaque pays , Cela prouve que chaque pays possède des particularités uniques qui ne peuvent pas être ignorées.

## 7 Conclusion

L'objectif central de cette étude était d'identifier le meilleur modèle économétrique pour expliquer les variations de l'espérance de vie à travers 193 pays. Nous avons mis en concurrence trois approches : le modèle poolé , le modèle à effets aléatoires et le modèle à effets fixes.

Pour arbitrer ce choix méthodologique, nous nous sommes appuyés sur différents types de tests statistiques rigoureux :

- Le Test de Fisher a d'abord disqualifié le modèle poolé. Il a démontré que l'hypothèse d'homogénéité (supposer que tous les pays fonctionnent de la même manière) était fausse. Il existe bien des "effets de panel" significatifs qu'on ne peut ignorer.
- Le Test de Breusch-Pagan a confirmé ce résultat en révélant une hétérosécédasticité marquée : la variance des erreurs n'est pas constante, ce qui valide la présence de spécificités propres à chaque pays. Le verdict du test de Hausman
- L'étape décisive a été le Test de Hausman. Ce test vérifie si les caractéristiques spécifiques des pays (comme la culture ou le climat) sont indépendantes de nos variables explicatives (comme le PIB ou l'éducation). Dans notre cas, l'hypothèse nulle a été rejetée avec une p-value très significative ( $2.19 \times 10^{-13}$ ). Cela signifie que les effets individuels sont corrélés aux variables explicatives. Autrement dit, les facteurs invisibles propres à un pays influencent directement son niveau d'éducation ou de richesse. Le modèle à effets aléatoires est donc biaisé et doit être écarté.

Le Modèle à Effets Fixes s'impose donc comme le seul estimateur convergent et fiable pour cette étude. Contrairement aux autres modèles, il neutralise les biais liés aux caractéristiques inobservées de chaque nation.

En capturant cette hétérogénéité structurelle, il nous permet d'affirmer avec robustesse que l'éducation et les politiques de santé (vaccination, lutte contre le VIH) sont des déterminants statistiquement significatifs de la longévité, indépendamment des spécificités fixes de chaque pays.