

Projet économétrie

Uriel FAGNINO et Cedric KOFFI

2023-11-29

1. Problématique et équation à estimer

1.1. Problématique

Face aux défis croissants liés à la qualité de vie et au bien-être des individus, il devient impératif de comprendre les déterminants du bien-être subjectif sous l'angle de l'économie. Ainsi, la présente étude vise à examiner à l'aide de méthodes économétriques les multiples facteurs susceptibles d'influencer le bien-être subjectif. Telle est la problématique de notre projet où nous nous focalisons sur l'année 2019.

1.2. Equation du modèle

Dans le cadre de la construction de notre modèle visant à expliquer les déterminants du bien être subjectif nous avons entrepris une revue approfondie de la littérature scientifique. Parmi les nombreuses théories et modèles disponibles, l'enquête historique sur l'état du bonheur dans le monde menées par le World Happiness Report ont attiré notre attention. En effet, l'enquête propose une approche globale qui établit des liens entre le score du bonheur d'un pays, le PIB par habitant, l'espérance de vie, l'aide sociale. Cette approche vise à décrire dans quelle mesure ces différents facteurs contribuent à l'évaluation du bonheur au sein de chaque pays. Ainsi, l'équation de notre modèle devrait se présenter comme suit:

$$Score = \beta_0 + \beta_1(PIB/habitant) + \beta_2(Aide.Sociale) + \beta_3(Esperance.de.vie) + \epsilon$$

2. Collecte des données

Pour commencer notre étude, nous avons procédé à la constitution de notre base de données, pour cela nous avons sélectionné trois variables explicatives et une variable à expliquer. Notre échantillon est constitué de données prises en coupes instantanées en 2019 pour 114 états dans le monde. Les données ont été prises sur le site <https://databank.worldbank.org>.

Nous avons combiné cette base avec la liste des pays de l'Afrique pour créer la variable PA (variable indicatrice) qui prend les valeurs :

- 0 si le pays n'est pas un pays Africain;
- 1 si le pays est un pays Africain

Nous pouvons lire ci-dessous le descriptif de la base finale que nous allons utiliser pour ce projet.

```
## # A tibble: 6 x 2
##   Code      Definition
##   <chr>      <chr>
## 1 Country    Country names
## 2 Score      Happiness Indicator
## 3 Healthy    Life expectancy at birth
## 4 Social_support Government expenditure on education Percentage of GDP
## 5 GDP_per_Capita GDP per capita constant International dollars 2019
## 6 PA        1 if African country 0 if not
```

Table 1: Dictionnaire des variables

Code	Definition
Country	Country names
Score	Happiness Indicator
Healthy	Life expectancy at birth
Social_support	Government expenditure on education Percentage of GDP
GDP_per_Capita	GDP per capita constant International dollars 2019
PA	1 if African country 0 if not

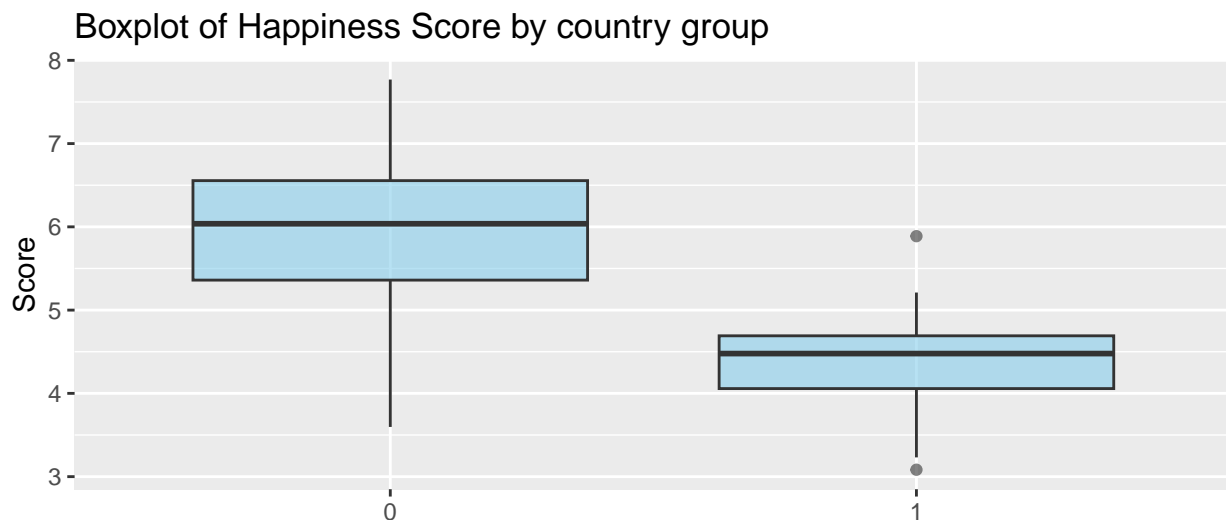
3. Analyse descriptive des variables

3.1. Analyse univariée

Pour en apprendre davantage sur notre variable d'intérêt et nos variables explicatives, nous allons procéder à une analyse statistique descriptive.

```
##
## Statistique Descriptive
## =====
## Statistic      N      Mean      St. Dev.      Min      Max
## -----
## Score          114      5.55        1.11        3.08        7.77
## Healthy        114      73.22        7.94       53.26       84.36
## Social_support 114      4.37        1.61        0.27        9.32
## GDP_per_Capita 114 16,198.92 21,302.57 270.14 106,729.00
## -----
```

```
## $`0`
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  3.597  5.361   6.037   6.010  6.555   7.769
##
## $`1`
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  3.083  4.058   4.478   4.386  4.691   5.888
```



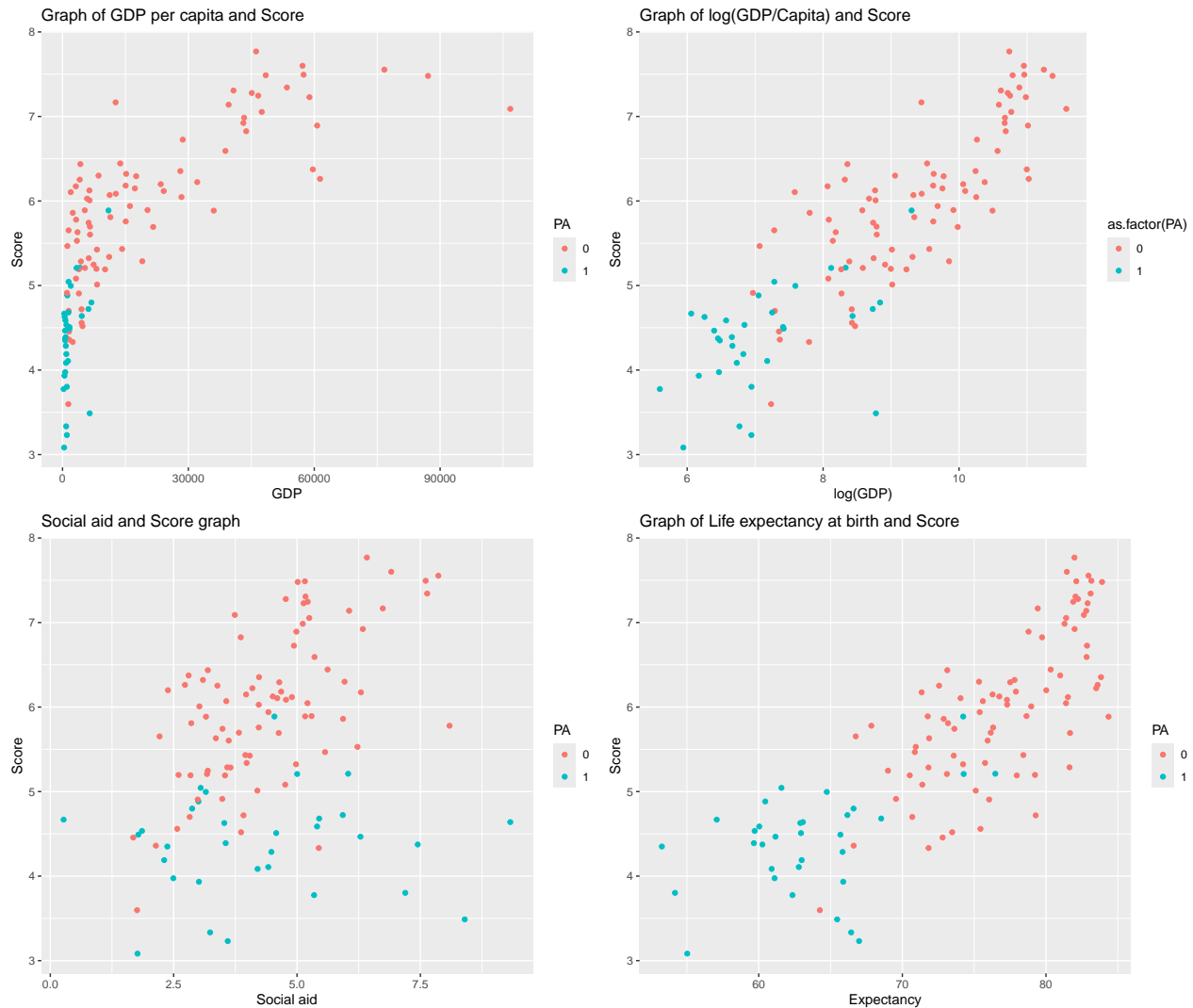
```
##
## Point atypiques
## =====
## Central African Republic
## -----
##
## Point atypiques
## =====
## Mauritius
## -----
```

L'analyse des statistiques descriptives par pays révèle que les pays non africains affichent un score de bonheur moyen de 6.010, surpassant ainsi celui des pays africains qui est de 4.386. Cette disparité est également corroborée par le boxplot, illustrant une dispersion distincte entre les deux groupes de pays.

Deux individus atypiques se démarquent au sein du groupe des pays africains, à savoir la République centrafricaine et l'île Maurice. Cette observation peut s'expliquer par le fait que la République centrafricaine présente un score de bonheur légèrement supérieur à la moyenne des pays africains de notre échantillon, avec un PIB par habitant relativement élevé. En revanche, l'île Maurice présente des caractéristiques opposées.

3.2. Analyse bivariée

Pour mieux comprendre la relation entre notre variable dépendante et nos variables explicatives, nous recourons à la représentation graphique sous forme de nuage de points. Cela nous permet d'explorer visuellement les liens et les tendances entre ces différentes variables.



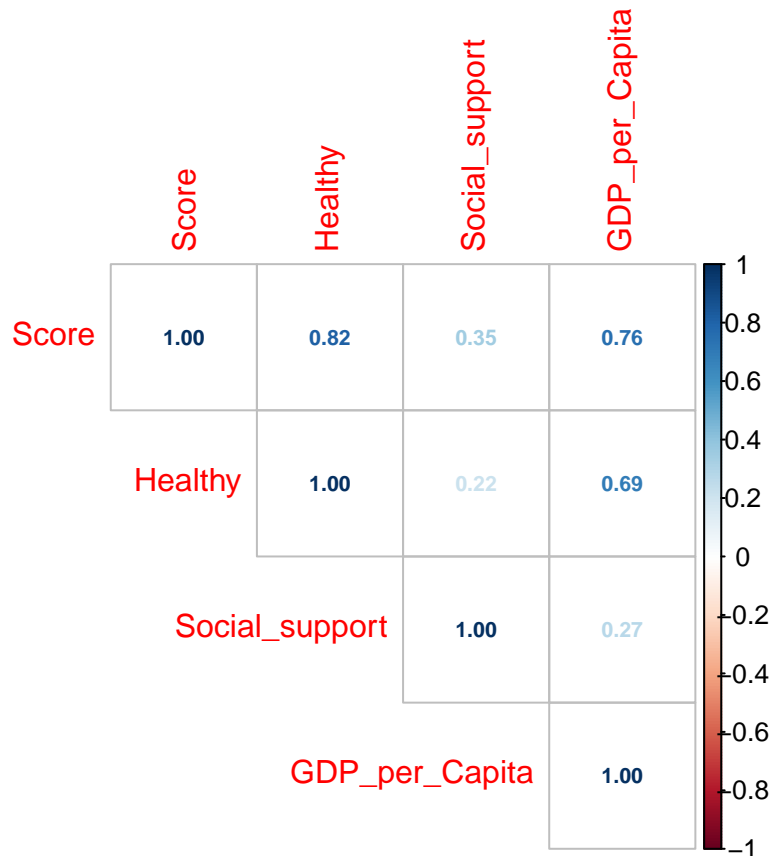
L'examen des quatre graphiques en nuage de points, illustrant la relation entre la variable à expliquer et les variables explicatives, révèle plusieurs observations significatives. Premièrement, il apparaît qu'il n'existe pas de lien linéaire évident entre le score de bonheur et le PIB. Deuxièmement, on observe une forte dispersion du score de bonheur par rapport au niveau d'aide sociale. Enfin, le quatrième graphique indique une relation linéaire entre notre variable dépendante (le score de bonheur) et l'espérance de vie à la naissance. Pour remédier à l'absence de linéarité entre le PIB et le score, nous avons pris le logarithme des deux variables, ce qui a révélé une relation linéaire dans le deuxième graphique.

Par ailleurs, il est intéressant de noter qu'il n'y a pas de distinction significative dans la dispersion des deux groupes de pays. Les pays non africains persistent en bas des graphiques, tandis que les pays africains restent en haut. Par exemple, au et à mesure que l'espérance de vie à la naissance augmente, le score de bonheur a tendance à augmenter, démontrant une tendance cohérente dans les deux groupes de pays.

3.2. Analyse de la corrélation

Afin d'explorer les relations entre deux variables et de faciliter la sélection des variables explicatives lors de l'estimation de notre modèle, nous entreprenons une analyse de la corrélation. Cette approche nous permet d'évaluer la force et la direction de l'association statistique entre les variables, contribuant ainsi à une compréhension approfondie des interdépendances au sein de notre ensemble de données. En utilisant des mesures de corrélation, nous sommes en mesure d'identifier les liens linéaires entre les variables, ce qui guide efficacement le choix des variables à inclure dans le processus d'estimation du modèle statistique.

```
##
## Matrice de corrélation
## =====
##          Score Healthy Social_support GDP_per_Capita
## -----
##      Score          1    0.816      0.345      0.758
##      Healthy      0.816    1      0.219      0.692
##      Social_support 0.345  0.219    1      0.270
##      GDP_per_Capita 0.758  0.692    0.270    1
## -----
```



Dans un premier temps, on observe une corrélation positive notable entre notre variable à expliquer (le score de bonheur) et les variables explicatives, à l'exception de la corrélation entre la variable explicative "aide sociale" et la variable de bonheur. Dans un second temps, l'analyse des corrélations entre les variables explicatives révèle une corrélation positive de 0.69 entre la variable PIB et l'espérance de vie à la naissance. La valeur absolue du coefficient de corrélation entre deux variables explicatives n'excède pas 0.8, on peut soupçonner d'une absence de colinéarité entre les variables explicatives.

Nous pouvons renforcer notre analyse en évaluant l'absence de multicolinéarité entre les variables en utilisant la règle de Klein. Si le carré du coefficient de corrélation est supérieur au R^2 , on peut soupçonner de la colinéarité.

```
##          Score Healthy Social_support GDP_per_Capita
## Score          TRUE  FALSE          FALSE          FALSE
## Healthy        FALSE  TRUE          FALSE          FALSE
## Social_support FALSE  FALSE          TRUE          FALSE
## GDP_per_Capita FALSE  FALSE          FALSE          TRUE
```

Cette règle met en avant une absence de colinéarité entre les variables.

4. Estimation du modèle

4.1. Spécification du modèle

Pour le choix finale de notre modèle, nous mettrons en route et comparons différents modèles à celui dont nous supposons par hypothèse plus adapté à notre étude dont l'équation est la suivante :

$$Score = \beta_0 + \beta_1 * \log(PIB/habitant) + \beta_2 * \log(PIB/habitant) * PA + \beta_3 * (Aide.Sociale) + \beta_4 * (Esprance.de.vie) + \beta_5 * (PA) + \epsilon$$

```
##
## Modèles de régression log linéaire
## =====
##                               Dependent variable:
##                               -----
##                               Score                               log(Score)
##                               (1)                               (2)                               (3)
## -----
## log(GDP_per_Capita)        0.498***                          0.493***                          0.083***
##                               (0.081)                          (0.077)                          (0.015)
##
## log(Healthy)                0.445                             0.147
##                               (1.237)                          (0.238)
##
## log(Social_support)         0.214*                             0.035
##                               (0.127)                          (0.024)
##
## PA                          1.860**                          -0.410**                          -0.094**
##                               (0.872)                          (0.191)                          (0.037)
##
## Healthy                     0.014
##                               (0.017)
##
## Social_support              0.100***
##                               (0.033)
##
## log(GDP_per_Capita):PA      -0.303***
##                               (0.113)
##
## Constant                   -0.198                          -0.846                          0.311
##                               (0.880)                          (4.821)                          (0.928)
## -----
## Observations                114                             114                             114
## R2                          0.781                             0.758                             0.745
## Adjusted R2                 0.771                             0.749                             0.735
## Residual Std. Error         0.529 (df = 108)                 0.554 (df = 109)                 0.107 (df = 109)
## F Statistic                  77.145*** (df = 5; 108)           85.326*** (df = 4; 109)           79.480*** (df = 4; 109)
## =====
## Note:                                                                *p<0.1; **p<0.05; ***p<0.01
```

Il semble que le premier modèle, présente le R2 ajusté le plus élevé parmi les modèles que nous avons considérés. Un R2 ajusté de 0.771 indiquant que 77,10% de la variabilité du score est expliquée par les variables indépendantes incluses dans le modèle avec effet croisé. Cependant, il est essentiel de réaliser un test de Ramsey afin de vérifier la spécification du modèle.

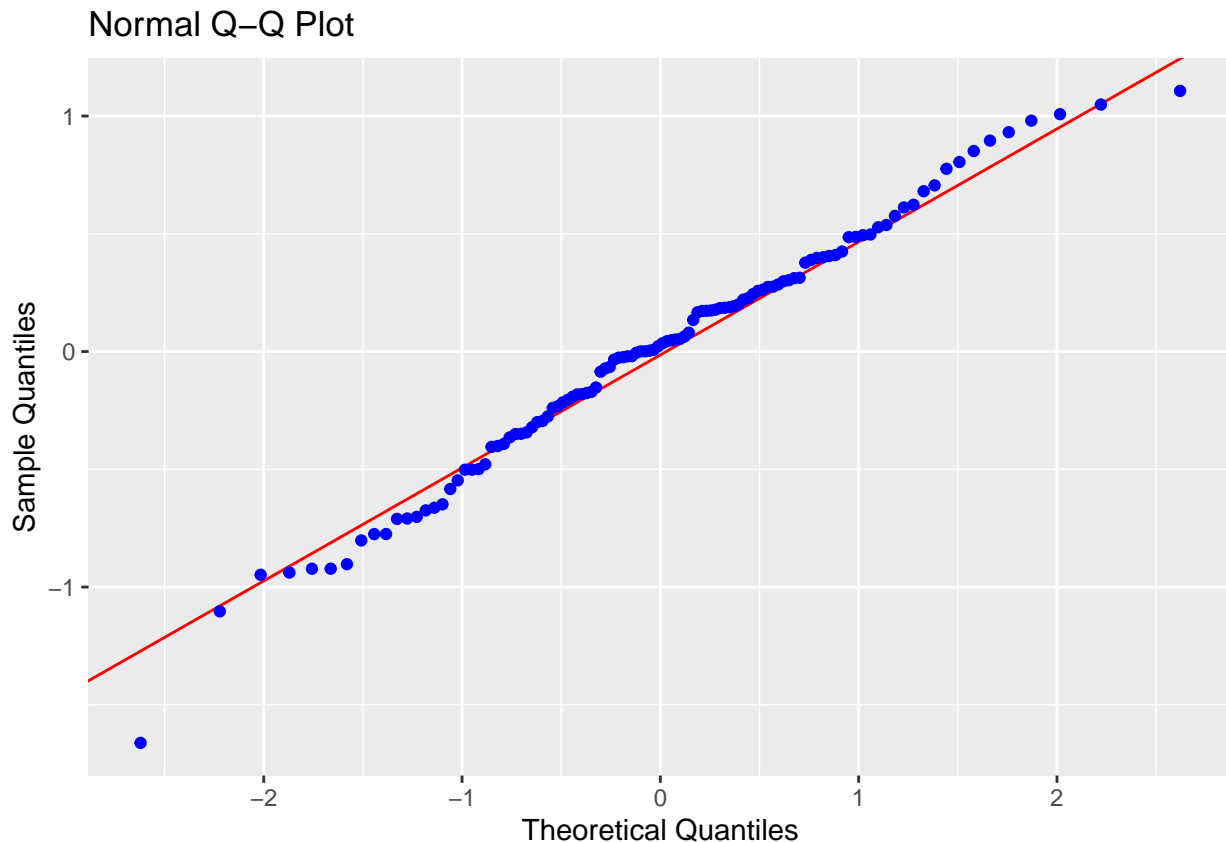
```
##
## RESET test
##
## data: modele
## RESET = 5.6537, df1 = 2, df2 = 106, p-value = 0.004645
```

La p-value du test de Ramsey est inférieure au seuil de 5% alors, le modèle est correctement spécifié. Toutefois, nous procéderons à une évaluation des critères de validité des hypothèses du modèle pour garantir une interprétation robuste des résultats.

4.1. Estimation du modèle

4.1.1. Test et validation de la méthode d'estimation

a) Normalité des résidus



```
##
## Shapiro-Wilk normality test
##
## data: modele$residuals
## W = 0.98881, p-value = 0.4724
```

Le graphique qqplot des résidus révèle que la grande majorité des observations s'alignent le long de la droite. De plus, la p-value du test de normalité des résidus dépasse le seuil de 5%, ce qui nous conduit à accepter l'hypothèse de normalité des résidus. Par conséquent, on peut affirmer que les résidus suivent approximativement une distribution normale.

b) Homoscédasticité des résidus

```
##
## studentized Breusch-Pagan test
##
## data: modele
## BP = 10.218, df = 5, p-value = 0.0693
```

La p-value est supérieure au seuil de 5% alors on accepte l'hypothèse nulle d'homoscédasticité.

4.1.2. Interpretation des coefficients, qualité de l'ajustement du modèle et des variables.

```
##
## Modèle d'estimation
## =====
##                               Dependent variable:
##                               -----
##                               Score
## -----
## log(GDP_per_Capita)          0.498***
##                               (0.081)
##
## PA                          1.860**
##                               (0.872)
##
## Healthy                      0.014
##                               (0.017)
##
## Social_support               0.100***
##                               (0.033)
##
## log(GDP_per_Capita):PA      -0.303***
##                               (0.113)
##
## Constant                    -0.198
##                               (0.880)
##
## -----
## Observations                 114
## R2                          0.781
## Adjusted R2                 0.771
## Residual Std. Error         0.529 (df = 108)
## F Statistic                  77.145*** (df = 5; 108)
## =====
## Note:                       *p<0.1; **p<0.05; ***p<0.01
```

- Test de significativité globale du modèle)

On utilise le test de Fisher : H_0 : tous les coefficients sont nuls sauf la constante, c'est à dire

$$\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

La statistique du test de Fisher Fobs = 89.61 et la p-value: < 0.05 (5%) donc on rejette l'hypothèse nulle, il y a au moins un coefficient différent de zéro, le modèle est donc **globalement satisfaisant**.

- Interpretation des effets et Significativité des coefficients

Le coefficient 1 affiche une significativité statistique robuste, observée tant au seuil de 5% qu'au seuil de 1%. Cela indique une confiance élevée dans la relation entre l'augmentation du PIB par habitant et l'accroissement du score de bonheur. En d'autres termes, si le PIB par habitant augmente d'un pourcentage, nous observerons une augmentation de 0.0049824 dans le score de bonheur.

Concernant le coefficient 4, il est également significatif au seuil de 5%, soulignant une relation statistiquement importante. Cela indique lorsque l'aide sociale (dépense du Gouvernement dans l'éducation pourcentage du PIB) augmente d'un point de pourcentage du PIB alors le score du bonheur augmentera de 0.1.

Par ailleurs, le coefficient des effets croisés des variables $\log(\text{GDP_per_Capita})$ et de la variable PA, sont significatif au seuil de 5%. En effet, pour les pays africains une augmentation du PIB par habitant d'un pourcentage est associée à une augmentation du score de bonheur de 0.001953 $((0.49824-0.30294)/100)$. En revanche, pour les pays non africains l'augmentation est de 0.0049824 dans le score de bonheur. Cette différence souligne des variations dans la relation entre le PIB par habitant et le bonheur selon la région.