

## On vous répond

Chers lecteurs, au vu des réponses au sondage proposé dans la revue de la semaine passée, vous êtes une majorité à vouloir en savoir plus sur l'espérance de vie. On constate que l'espérance de vie actuelle mondiale, soit 71.7 ans, est 1.4 fois supérieure à celle de 1960 qui était de 52.6 ans. Pour expliquer cette croissance, nous avons choisi d'introduire dans notre étude le niveau d'éducation dans chaque pays. En effet, on remarque que de plus en plus de personnes vont à l'école. Cette évolution s'accompagne du développement des secteurs pointus tels que la recherche ou la santé. Ainsi, le temps d'étude croissant a pu permettre des avancées importantes dans ces secteurs. Ces améliorations sont manifestement bénéfiques à nos vies. C'est pourquoi la question à laquelle nous allons répondre est la suivante : **le niveau d'étude d'un pays a-t-il un impact sur l'espérance de vie de sa population?**

Nous allons tout d'abord analyser chaque variable indépendamment avant de les mettre en relation. Ici, les deux qui nous intéressent sont X le niveau d'éducation par pays en 2014, et Y l'âge moyen d'espérance de vie par pays en 2014. Notre étude portera sur une sélection de 171 pays.

Commençons par analyser la variable qualitative X, qui correspond au niveau d'éducation par pays en 2014. X comprend 2 niveaux, c'est-à-dire 2 modalités qui sont « Faible » et « Élevé ». On a donc divisé les 171 pays étudiés en 2 catégories, dont la loi observée est la suivante :

	Élevé	Faible
Effectif	113	58
Fréquence	0.66	0.34

Ce tableau signifie que 113 pays ont un niveau d'éducation élevé, ce qui représente 66% des 171 pays, et que 58 pays ont un niveau d'éducation faible, soit 34% des pays.

On peut également représenter ces effectifs et fréquences sous forme d'histogramme et de diagramme circulaire :

histogramme de Y en effectifs

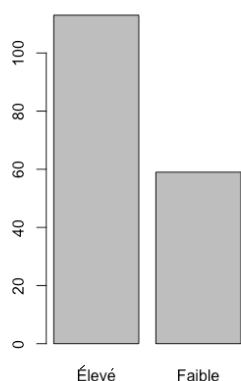
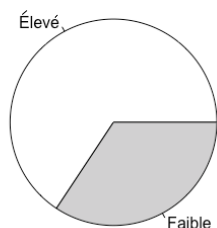


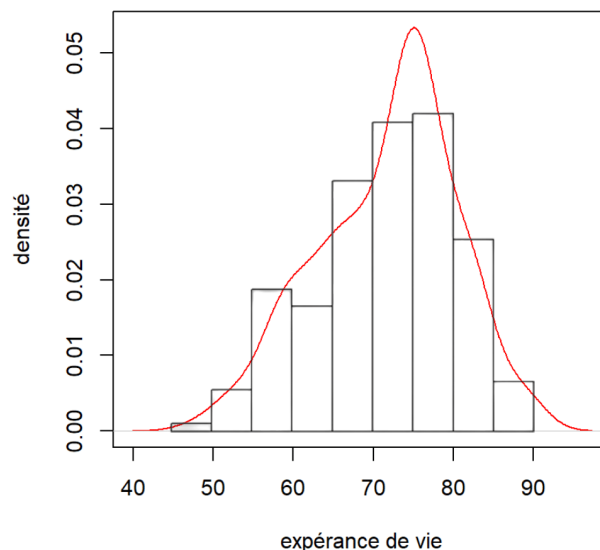
diagramme circulaire de Y



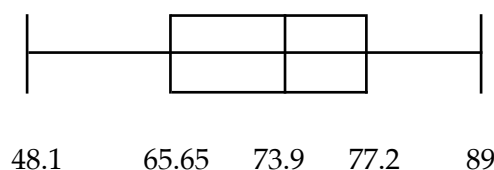
Intéressons-nous maintenant à la variable que nous tenterons d'expliquer par la suite, c'est-à-dire Y, l'âge moyen d'espérance de vie par pays en 2014.

Y est une variable quantitative qui prend ses valeurs dans l'intervalle [48.1 ; 89], il s'agit donc d'une variable continue qui possède alors une densité. Cette densité est représentée par la ligne rouge sur le graphique ci-dessous :

Histogramme de l'espérance de vie et sa densité



boîte à moustaches de Y

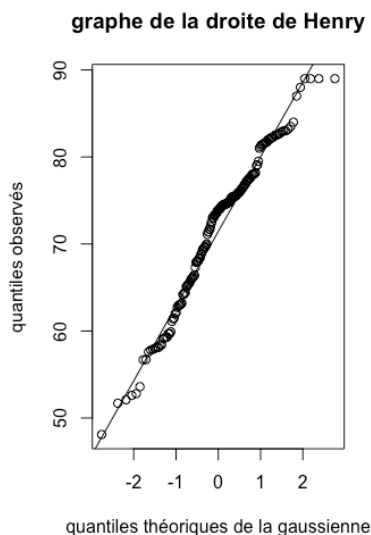


Cet histogramme et cette boîte à moustaches corroborent les données que nous avons recueillies sur Y :

- L'âge minimum moyen d'espérance de vie de la population d'un pays est de 48.1 ans.
- L'âge maximal moyen d'espérance de vie de la population d'un pays est de 89 ans.
- 50% des pays ont un âge moyen d'espérance de vie supérieur à 73.9 ans.
- 25% des pays ont un âge moyen d'espérance de vie supérieur à 77.2 ans.
- 25% des pays ont un âge moyen d'espérance de vie inférieur à 65.65 ans.

Comme vous pouvez le remarquer, la densité et l'histogramme se superposent, il n'y a pas de grandes disparités entre les deux. On peut également remarquer l'allure gaussienne que prend la densité, malgré le fait qu'elle ne soit pas parfaitement symétrique.

Pour vérifier si notre observation s'avère juste, il faut en premier lieu tracer la droite de Henry. Cette droite produit un graphe de points plus ou moins alignés si la densité est gaussienne. Si nous obtenons un alignement approximatif des points, cela appuiera notre intuition et nous pourrons alors entreprendre un test statistique qui confirmera pour de bon la gaussianité de la densité.



Avec les données récoltées de l'année 2014, nous obtenons le graphe ci-dessus. On peut noter que tous les points sont globalement alignés, mises à part quelques exceptions que l'on peut remarquer en haut à droite. Ainsi, nous pouvons modéliser la loi de Y par une gaussienne, voyons à présent ce que nous dit le test statistique que nous allons effectuer.

Le test statistique que nous allons faire ici est le test de Shapiro. Pour cela, il faut tout d'abord poser un problème de test qui est le suivant :

$$H_0 : Y \text{ suit une loi gaussienne}$$

$$H_1 : Y \text{ ne suit pas une loi gaussienne}$$

avec  $\alpha = 0.001$  comme risque d'erreur.

Si on obtient une pi-value, qui est un indicateur de vraisemblance, plus grande que notre risque d'erreur alors on choisit  $H_0$ .

Au contraire, si la pi-value est inférieure à notre risque d'erreur, alors on rejette  $H_0$  et on accepte  $H_1$ .

En effectuant ce test, on trouve une pi-value de 0.002. La pi-value est donc supérieure au risque d'erreur que nous nous sommes fixés, nous décidons alors de choisir  $H_0$  et d'accepter la gaussianité de notre variable Y.

Nous voici donc en possession de toutes les informations nécessaires sur les deux variables que nous allons étudier.

Néanmoins, dans l'étude que nous menons, nous ne pouvons pas étudier deux variables qui ne sont pas du même type : ici nous souhaitons avoir deux variables qualitatives.

Pour cela, nous allons discrétiser Y, l'âge moyen d'espérance de vie par pays, en tranches. Cette discrétisation consiste à diviser Y en différentes classes, ici en différents niveaux.

Nous créons 3 niveaux différents correspondants au niveau d'espérance de vie d'un pays :

- Faible pour les pays dont l'âge moyen d'espérance de vie est inférieur ou égal à 65.5 ans, soit 25 % des pays.
- Moyen pour les pays dont l'âge moyen d'espérance de vie est strictement compris entre 65.5 ans et 71.3 ans soit 15% des pays.
- Élevé pour les pays dont l'âge moyen d'espérance de vie est supérieur ou égal à 71.3 ans, soit 60% des pays.

Dans l'analyse commune des deux variables qui suit, nous allons considérer que Y n'est plus une variable quantitative mais bien une variable qualitative que nous avons discrétisée en tranches comme indiqué précédemment.

Ainsi Y est alors une variable qualitative qui possède 3 modalités : Élevé, Moyen et Faible.

L'analyse univariée de nos variables X et Y est donc terminée. Nous pouvons alors commencer à s'intéresser à l'essence même de notre étude et chercher le lien éventuel qui unit ces deux variables. Pour vous le rappeler, il s'agit ici de savoir si le niveau d'espérance de vie Y peut être expliquée par le niveau d'éducation X.

Pour que les données vous paraissent plus claires et lisibles, faisons un tableau de contingence de X et de Y en effectifs et en probabilité.

Ce premier tableau est la table de contingence en effectifs. Dans ce type d'étude il est important d'avoir au moins 5 observations dans chaque case pour les tests suivants. Cette condition justifie ainsi notre découpage de Y cité ci-dessus.

Niveau d'espérance de vie \ Niveau d'éducation	Faible	Moyen	Élevé	Profil moyen de X
	Élevé	5	10	98
Faible	37	16	5	58
Profil moyen de Y	42	26	103	171

Niveau d'espérance de vie \ Niveau d'éducation	Faible	Moyen	Élevé	Profil moyen de X
	Élevé	0.03	0.06	0.57
Faible	0.22	0.09	0.03	0.34
Profil moyen de Y	0.25	0.15	0.60	1

À gauche, nous pouvons voir la table de contingence mais cette fois-ci en probabilité. À titre d'exemple, la probabilité qu'un pays ait un niveau d'éducation élevé et un niveau d'espérance élevé est de 0.57.

Ce qui nous intéresse maintenant, c'est de calculer le profil ligne, c'est-à-dire la loi de Y, le niveau d'espérance de vie, sachant X, le niveau d'éducation. On obtient le tableau suivant :

Niveau d'espérance de vie \ Niveau d'éducation	Faible	Moyen	Élevé
	Élevé	0.04	0.09
Faible	0.64	0.27	0.09
profil moyen de Y	0.25	0.15	0.60

Ce tableau permet de lire les informations suivantes :

- Un niveau d'espérance de vie faible se retrouve moins dans les pays où le niveau d'éducation est élevé : -21% par rapport à tous les pays
- Un niveau d'espérance de vie faible se retrouve plus dans les pays où le niveau d'éducation est faible : +39% par rapport à tous les pays.

- Un niveau d'espérance de vie élevé se retrouve plus dans les pays où le niveau d'éducation est élevé : +27% par rapport à tous les pays
- Un niveau d'espérance de vie élevé se retrouve moins dans les pays où le niveau d'éducation est faible : -51% par rapport à tous les pays.

On voit alors se dessiner un lien entre nos deux variables.

Le deuxième outil qui va nous permettre de préciser cette relation est l'indice d'attraction-répulsion.

À travers l'élaboration des indices d'attraction et de répulsion, on peut voir si le niveau d'éducation des pays influe sur leur niveau d'espérance de vie. Si l'indice est proche de 1, c'est que le niveau d'éducation n'influe pas sur le niveau d'espérance de vie.

Si l'indice est supérieur à 1, c'est qu'il y a une certaine attraction entre le niveau d'éducation et le niveau d'espérance de vie.

Au contraire, si l'indice est inférieur à 1, c'est qu'il y a répulsion entre le niveau d'éducation et le niveau d'espérance de vie.

Voici ci-dessous le tableau des indices d'attraction-répulsion :

Niveau d'espérance de vie \ Niveau d'éducation	Faible	Moyen	Élevé
Élevé	0.18	0.58	1.43
Faible	2.59	1.81	0.14

Ces indices nous permettent de récolter les informations suivantes :

- Les pays à faible espérance de vie n'ont pas un niveau d'éducation élevé car l'indice est de 0.18, c'est donc un indice de répulsion. Au contraire, ils ont un niveau d'éducation faible puisqu'il s'agit ici d'un indice d'attraction de 2.59.
- Les pays au niveau moyen d'espérance de vie ont un niveau d'éducation faible puisque leur indice est de 1.81.
- Les pays à forte espérance de vie ont un niveau d'éducation élevé car l'indice est de 1.43, c'est donc un indice d'attraction. Au contraire, ils n'ont pas un niveau d'éducation faible puisqu'il s'agit d'un indice de répulsion de 0.14.

Grâce à ces indices, on constate que la relation entre niveau d'espérance de vie et niveau d'éducation n'est pas neutre, les deux variables sont liées.

Il reste encore un outil qui peut nous permettre de renforcer notre intuition quant au lien entre ces variables : le tableau de contribution au chi2.

Ce tableau contient des pourcentages qui nous indique une contribution plus ou moins forte.

Plus la contribution est petite, moins les variables ont un impact l'une sur l'autre.

Plus la contribution est grande, plus les variables ont un impact l'une sur l'autre.

Néanmoins, on ne sait pas si l'impact est positif ou négatif.

En conclusion, nos résultats prouvent qu'il y a bien une relation entre l'espérance de vie des populations et le niveau d'éducation du pays. Ce résultat peut être expliqué par la meilleure qualité de vie qu'apporte un niveau d'éducation élevé et à l'inverse par les conditions de vie difficiles qui peuvent résulter d'une éducation fragile. Cependant, cette analyse peut être approfondie avec d'autres facteurs pour expliquer davantage les différences d'espérance de vie entre les pays, tels que l'accès aux soins, la qualité de l'alimentation ou encore le niveau de richesse des populations.

On sait cependant que si tout contribuait de manière équitable, on aurait 16.66% (100% divisé par 6) dans toutes les cases du tableau. On fixe donc le seuil de dépendance à 16.66%.

Néanmoins, on voit bien qu'il y a des disparités entre les cases du tableau suivant.

Niveau d'espérance de vie \ Niveau d'éducation	Faible	Moyen	Élevé
Élevé	18 %	3 %	12 %
Faible	35 %	5 %	25 %

On remarque les points suivants :

- La case où se trouve 35% contribue beaucoup au chi2, tout comme celle à 25%
- Au contraire, les cases où le niveau d'espérance de vie est moyen contribuent très peu au chi2.

Pour finir notre analyse, intéressons-nous plus globalement au lien entre X et Y. Nous allons donc voir si elles sont indépendantes ou non. Pour cela, nous réalisons le test d'indépendance du chi2 qui suit le problème suivant :

$H_0$ : X et Y sont indépendantes

$H_1$ : X et Y sont dépendantes

On choisit ici un  $\alpha = 0.001$  comme risque d'erreur.

De nouveau, si on obtient une pi-value plus petite que notre risque d'erreur alors on prend  $H_1$ .

À l'inverse, si la pi-value est supérieure au risque d'erreur, on choisit  $H_0$ .

En effectuant le test, on obtient une pi-value très petite, très proche de 0. On a donc une pi-value inférieure à notre risque d'erreur.

Ainsi nous décidons de choisir  $H_1$  et d'accepter la dépendance entre les deux variables qui sont, pour vous le rappeler, Y le niveau d'espérance de vie par pays et X le niveau d'éducation par pays.