ISSN 2593-9157

An open-access journal of population research

REVUE Quetelet JOURNAL

Volume 9/2
2021

Taux de croissance annuel moyen via une formule géométrique : toujours, parfois, jamais ?

Christophe Vandeschrick
christophe.vandeschrick@uclouvain.be
Université catholique de Louvain, Louvain-la-Neuve
Centre de recherche en démographie

DEMO Centre de recherche en démographie
IACCHOS Institut d'analyse du changement dans l'histoire et les sociétés contemporaines

UCLouvain

DOI.10.14428/rqj2021.09.01.02 • ©2021 Christophe Vandeschrick

PUL PRESSES UNIVERSITAIRES DE LOUVAIN

Résumé - Abstract

Résumé

Dans le cadre de l'analyse diachronique de taux, d'indices ou de rapports, la littérature propose régulièrement de calculer la moyenne via une formule géométrique. Cet article vise à montrer que cette règle n'est pas acceptable. La réflexion est régie par le principe de base suivant : le taux de croissance annuel moyen correspond à la valeur du taux qui, appliquée durant les différentes années sous observation, conduit à la même croissance que celle effectivement observée entre le début et la fin de la période d'observation. Son application aux trois hypothèses de croissance de la population utilisées couramment en démographie démontre que le taux de croissance moyen n'est jamais obtenu via une formule géométrique de la moyenne. Par ailleurs, l'application du principe de base fait émerger une formule du taux moyen sous hypothèse de croissance linéaire en contradiction avec la formule habituellement suivie et qui doit donc être abandonnée.

Mots-clés : *taux de croissance, moyenne, moyenne arithmétique, moyenne harmonique, moyenne géométrique.*

Abstract

In the context of the diachronic analysis of rates, indices or ratios, the literature regularly proposes to calculate the average via a geometric formula. This article aims to show that this rule is not acceptable. The reflection is governed by the following basic principle: the average annual growth rate corresponds to the value of the rate which, applied during the different years under observation, leads to the same growth as that actually observed between the beginning and end of the observation period. Its application to the three population growth hypotheses commonly used in demography demonstrates that the average growth rate is never obtained via a geometric formula of the average. Moreover, the standard application of the basic principle gives rise to a formula for the average rate under the assumption of linear growth which contradicts the formula usually followed and which must therefore be abandoned.

Keywords: *growth rate, mean, arithmetic mean, harmonic mean, geometric mean.*

1. Introduction

Comment calculer, au cours d'une période pluriannuelle, le taux de croissance annuel moyen au départ des taux annuels successifs (l'adjectif « annuel » sera régulièrement omis dans la suite du texte) ? Dans la littérature (*cf.* Braverman 1978, p. 96 ; Lewis 2012, pp. 100-102 ; Py 2007, p. 104 ou Spiegel 1984, p. 60), on peut trouver une règle selon laquelle ce taux moyen s'obtiendrait systématiquement via une formule géométrique, définie comme la racine énième du produit des n taux observés. En démographie, le taux de croissance moyen se calcule généralement via des formules où n'apparaissent pas les valeurs observées du taux, mais bien les populations initiale et finale. De ce fait, chacune des moyennes obtenues ne peut pas être qualifiée d'arithmétique, de géométrique, etc. La démographie offre néanmoins un terrain de réflexion intéressant par rapport à un usage supposé automatique de la formule géométrique en cas de taux, notamment parce qu'elle prend habituellement en compte trois hypothèses ou modèles de croissance. Pour établir les conséquences de ces trois hypothèses sur la formule à utiliser, les calculs du taux moyen seront établis en faisant apparaitre explicitement les valeurs des taux annuels. Ces formules permettront de démontrer que la règle de l'utilisation systématique de la formule géométrique n'est pas acceptable.

La démonstration repose sur le principe suivant, en prenant l'exemple de l'évolution de l'effectif d'une population sur plusieurs années successives : **le taux de croissance annuel moyen sur une période est la valeur du taux qui, appliquée aux différentes années de la période, conduit à la même variation de la population que celle effectivement observée**. Ce principe de base (selon sa dénomination dans la suite du texte) est en fait l'application – restreinte au cas spécifique du taux de croissance d'une variable dans le temps – d'une définition alternative de la moyenne : « *la moyenne est la valeur de la variable qui, appliquée à toutes les unités sous observation, conserve l'effet global de la variable sur l'ensemble des unités sous observation*. » (Vandeschrick 2017, p. 6). Cette définition se traduit par le principe de base si les mots ou expressions « variable », « unités sous observation » et « effet global » sont respectivement remplacés par « taux de croissance », « années » et « variation de la population ».

Mutatis mutandis, ce principe de base sera aussi utilisé pour le coefficient multiplicateur, à savoir le nombre par lequel il faut multiplier la population pour obtenir son effectif après une certaine période. Même sans référence à la définition alternative, ce principe devrait être accepté sans réticence, et ce, d'autant plus qu'il est parfois utilisé dans la littérature, éventuellement de manière implicite (Dehon *et al.* 2015, pp. 100-101 ; Lewis 2012, pp. 100-102 ; Wonnacott and Wonnacott 1990, p. 667 ; Wunsch *et al.* 2001, pp. 24-27).

L'importance du taux de croissance en démographie - et aussi dans d'autres domaines - et surtout la suggestion d'un recours automatique à la formule géométrique pour le calcul de sa moyenne justifient le présent texte.

Le point 2 est réservé à un exposé sur la façon dont les démographes introduisent la notion de taux de croissance moyen, avec comme conclusion qu'il faut suivre une autre procédure pour être en mesure de résoudre la question posée dans le titre. Le point 3 établit les formules du taux moyen correspondant aux trois hypothèses de croissance utilisées en démographie et le point 4, celles pour le coefficient multiplicateur moyen. Le point 5 propose un tableau récapitulatif des formules introduites ainsi qu'une adaptation de la terminologie conforme au principe de base. Au point 6, des commentaires illustrent, d'une part, la nécessité de réaliser toutes les étapes d'un processus calculatoire sous la même hypothèse de croissance (ce qui n'est pas toujours fait en démographie) et, d'autre part, l'efficacité du principe de base par rapport à la méthode essai-erreur parfois utilisée en statistique, notamment pour justifier le recours à la formule géométrique de la moyenne.

Le principe de base vaut pour la croissance dans le temps de toute grandeur, et pas uniquement pour l'effectif d'une population. Une des qualités majeures du taux de croissance est d'ailleurs de permettre de comparer la variation temporelle de variables de différentes natures, par exemple, l'effectif d'une population et le produit intérieur brut (PIB). Aussi, certaines parties du texte quitteront le domaine purement démographique.

2. Présentation usuelle des taux de croissance annuels moyens en démographie

Comment le taux de croissance et sa moyenne sont-ils présentés en démographie ? Pour répondre à cette question, nous prendrons l'explication proposée par Wunsch *et al.* (2001, pp. 24-27), notamment parce que l'explication y est plus élaborée que bien souvent, même s'il y subsiste certaines formes d'imprécisions. Les auteurs soulignent l'intérêt de prendre en compte les différentes méthodes du calcul du taux moyen malgré le peu de différence entre les résultats obtenus selon l'hypothèse de croissance retenue, vu le niveau faible de la croissance actuellement observé dans bon nombre de populations humaines. Wunsch *et al.* réfutent cet argument en citant l'exemple de l'épidémie du sida au début de laquelle la population de malades doublait chaque année. La même réflexion pourrait s'appliquer à la pandémie du Covid.

Wunsch *et al.* commencent le point intitulé « Le taux de croissance » comme suit : « Pour simplifier, plaçons-nous immédiatement dans l'unité de temps standard en démographie, l'année, et supposons connus les effectifs de la population P_t au 1[er] janvier de l'année t et P_{t+1}, au 1[er] janvier de l'année $t+1$. Il ne reste plus qu'à rapporter l'accroissement P_{t+1} - P_t à l'effectif de la population. Mais quel effectif ? ». Pour le choix de l'effectif, trois options sont envisagées :

- Prendre l'effectif de départ, soit P_0. Ce choix est justifié par le fait que « c'est la population elle-même qui produit les évènements (naissances, décès, migrations) qui concourent à son accroissement ». Ce caractère autoreproductif a comme conséquence que l'accroissement est d'autant plus important en nombre absolu que l'effectif initial de la population est lui-même important. Même si Wunsch *et al.* n'ont pas employé cette appellation, les calculs prenant en compte P_0 se développent en posant l'hypothèse d'une croissance géométrique (*cf.* Tableau 1).

- Pour prendre en compte le fait que l'effectif de la population change au cours du temps, il pourrait sembler préférable de prendre la population moyenne au cours de l'année comme diviseur, cette dernière étant le plus souvent estimée par la moyenne des populations en début et fin d'année. Dans ce cas, le calcul s'opère en posant l'hypothèse d'une croissance linéaire.

- Enfin, pour mieux prendre en compte le fait que l'effectif de la population change continuellement en fonction du temps, il est intéressant d'envisager aussi l'hypothèse de croissance exponentielle. Chez Wunsch *et al.*, le caractère continu du temps est symbolisé par l'usage de la lettre h, mais pour simplifier les notations, l'indicateur h du temps est remplacé ici par t, ce que Wunsch *et al.* justifient dans une note.

Notons que les hypothèses géométrique et exponentielle sont apparentées, mais diffèrent par le fait que la première se développe dans un cadre de référence discret (avec bilan entre deux moments généralement séparés d'un an) et la deuxième, dans un cadre temporel continu. Cette opposition discret-continu a des conséquences non négligeables sur la valeur des taux de croissance annuels calculés pour une même population durant une période donnée. Par ailleurs, vu le choix de la population utilisée comme diviseur dans les différentes hypothèses, il est possible de dégager des informations qualitatives sur les taux obtenus. Par exemple, en cas de croissance positive, la population initiale est plus faible que la population en cours d'année ; en conséquence, le taux selon l'hypothèse géométrique est le plus élevé (et vice-versa en cas de croissance négative). Enfin, d'autres modèles que les trois hypothèses citées peuvent encore être utilisés (par exemple, le modèle logistique), mais nous avons décidé de nous limiter aux trois les plus connus. Le Tableau 1 reprend les équations principales de l'exposé de Wunsch *et al*.

Tableau 1. Synthèse des équations utilisées par Wunsch *et al.* (2001)

Hypothèse…	**… géométrique**	**… linéaire**	**… exponentielle**
Taux de croissance sur un an	$a = \frac{P_{t+1} - P_t}{P_t}$	$r = \frac{P_{t+1} - P_t}{(P_t + P_{t+1})/2}$	$r^* = \frac{ln(P_{t+1}/P_t)}{1}$
Évolution de la population sur *t* années, à taux constant	$P_t = P_0(1 + a)^t$	Pas de formule dans Wunsch *et al.*	$P_t = P_0 * e^{r^*t}$
Taux moyen calculé sur une période de *t* années	$a = \sqrt[t]{\frac{P_t}{P_0}} - 1$	$r = \frac{P_t - P_0}{t(P_0 + P_t)/2}$ (1)	$r^* = \frac{ln(P_t/P_0)}{t}$

Note : le 1 au dénominateur du taux sur un an dans le cas exponentiel peut surprendre. Il correspond à la présence du « *h* » dans la 2e formule de la p. 27 chez Wunsch *et al.* Cette division par 1 est sans effet numérique, mais utile à d'autres points de vue, non discutés ici.

En prenant l'exemple de l'hypothèse géométrique, *a* revêt différentes significations. Il pourrait sembler souhaitable de modifier les équations pour souligner ce fait :

• Le taux de croissance sur une année (1ère ligne) : il faudrait sans doute ajouter un indice temporel signalant l'année concernée, à savoir soit *t*, soit *t*+1. En se référant à la dernière équation de la p. 24 et son commentaire, le choix de *t* semble s'imposer vu que le taux calculé entre P_0 et P_1 se réfère à la croissance durant l'année *0* (Wunsch *et al*., p. 24) : « si l'on part de l'année 0 ») :

$$a_t = \frac{P_{t+1} - P_t}{P_t}.$$

• Le taux de croissance constant permettant de passer de la population au temps *0* à la population au temps *t* (2e ligne), avec un écart pluriannuel entre *0* et *t* : pourquoi n'envisager que le cas où le taux est constant d'année en année ? Ce choix est sans doute motivé par le fait que l'objectif est l'établissement du calcul du taux moyen au départ des populations en début et fin d'observation. Pour se libérer de la contrainte de la constance du taux, introduisons la lettre « *T* » indiquant le moment du dernier décompte de la population. Dès lors, l'équation s'écrirait :

$$P_T = P_0 \prod_{t=0}^{T-1} (1 + a_t). \qquad (2)$$

Cette équation est cruciale pour la suite du texte, car elle permettra de faire émerger, au point 4, une formule de calcul du taux moyen au départ des différentes valeurs annuelles observées.

• Le taux de croissance annuel moyen entre le moment *0* et le moment *t* (à remplacer par *T*) (3e ligne) : pourquoi ne pas indiquer le caractère moyen du taux en ajoutant, par exemple, une barre au-dessus du *a* ? Dès lors, cette équation s'écrirait :

$$\bar{a} = \sqrt[T]{\frac{P_T}{P_0}} - 1$$

Notons que les deux dernières équations se réfèrent implicitement au principe de base et l'accréditent.

Mutatis mutandis, le même type de commentaire vaut pour les deux autres hypothèses. Au Tableau 2 figurent les équations révisées suivant le raisonnement proposé pour l'hypothèse géométrique. Dans le cas de l'hypothèse linéaire, l'équation montrant l'évolution de la population sur t années était absente de l'exposé de Wunsch *et al.* Cette situation est révélatrice de ce qui pourrait être qualifié de « fragilité conceptuelle » à propos de l'hypothèse linéaire. Dans le Tableau 2, l'équation manquante a été ajoutée. Sa logique sera abordée dans le point 3. Par ailleurs, Pressat (1980, pp.155-158) propose une équation suivant cette logique, mais limitée à un calcul sur une année (*cf.* point 3, formule (4), et point 6.1). Pour cette hypothèse linéaire, l'équation (1) des Tableaux 1 et 2 est problématique, car incohérente avec l'équation de l'évolution de la population sur T années ajoutée dans le Tableau 2 (*cf.* point 4.2).

Tableau 2. Révision des équations utilisées par Wunsch *et al.* (2001)

Hypothèse…	… géométrique	… linéaire	… exponentielle
Taux de croissance sur une année	$a_t = \frac{P_{t+1} - P_t}{P_t}$	$r_t = \frac{P_{t+1} - P_t}{(P_t + P_{t+1})/2}$	$r_t^* = \frac{ln(P_{t+1}/P_t)}{1}$
Évolution de la population sur T années	$P_T = P_0 \prod_{t=0}^{T-1}(1 + a_t)$	$P_T = P_0 * \prod_{i=0}^{T-1}\left(\frac{2 + r_t}{2 - r_t}\right)$	$P_T = P_0 * e^{\sum_{t=0}^{T-1} r_t^*}$
Taux moyen calculé sur une période de T années	$\bar{a} = \sqrt[T]{\frac{P_T}{P_0}} - 1$	$\bar{r} = \frac{P_T - P_0}{T(P_0 + P_T)/2}$ (1)	$\bar{r}^* = \frac{ln(P_T/P_0)}{T}$

Au-delà de ces quelques commentaires, les explications de Wunsch *et al.* ne donnent pas les clés pour répondre à la question posée dans ce texte. En effet, les formules du calcul du taux moyen ne font pas apparaitre les valeurs observées du taux, ce qui empêche de voir si les formules directes (avec les populations initiale et finale) dissimulent ou pas une formule géométrique de la moyenne.

Comme solution, nous proposons de convoquer les formules du calcul du taux moyen pour les trois hypothèses au départ du principe de base appliqué aux taux annuels observés. Tout d'abord, cette façon de procéder raccroche le calcul du taux moyen du calcul de la moyenne en général où les valeurs de la variable dont on cherche la moyenne apparaissent. Surtout, en comparant des formules bâties sur le principe de base, nous pourrons déterminer, sans risque d'erreur, si le taux de croissance moyen doit toujours, parfois ou jamais être calculé selon une formule géométrique.

3. Le taux de croissance selon les trois hypothèses : démonstration

Supposons une population passant de 10.000 habitants au 01/01/2011 à 25.000 au 01/01/2012 (puis à 65.000 et 87.000 en 2013 et 2014) (Tableau 3). Ces données ont été choisies de façon délibérée afin d'obtenir des taux de croissance bien contrastés pour les trois hypothèses de croissance (pour d'autres exemples, *cf.* point 6.2). Par ailleurs, dans un souci d'uniformisation avec les habitudes de la statistique, la notation adoptée ne suit pas celle utilisée par Wunsch *et al*. (encadré 1).

Encadré 1. La notation

Chez Wunsch *et al.*, le temps est indiqué par la lettre t et le taux de croissance est symbolisé par différentes lettres (r, a ou r^*). Dans un souci d'uniformisation avec les habitudes de la statistique, la notation a été modifiée. Par exemple, nous allons montrer que le taux annuel moyen selon l'hypothèse exponentielle se calcule via une formule arithmétique bien connue :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}\text{ , où}$$

- x désigne la variable dont on cherche la moyenne, soit le taux de croissance annuel selon l'hypothèse exponentielle (Tableau 3, dernière colonne) ;
- $\bar{x}$ désigne la moyenne dont la valeur est recherchée ;
- i désigne une des unités sous observation pour laquelle on connait la variable, soit une année pour laquelle le taux est connu ;
- n désigne le nombre d'années pour lesquelles le taux est connu, soit 3 dans l'exemple ;
- x_i désigne la valeur de la variable pour une des unités sous observation. Si $i = 1$, x_1 correspond à la valeur du *taux de croissance selon l'hypothèse exponentielle* en 2011, soit 0,91629 (91,629 %).

Il est à noter que le Tableau 3 renseigne les populations connues pour une date précise, soit le 1er janvier de l'année indiquée dans la 2e colonne. Le calcul des 3 taux selon une hypothèse nécessite la connaissance de la population à 4 dates. Dans ce Tableau 3, i pourra varier de *1* à *4* (soit $n+1$).

Enfin, le point 4 traitera du coefficient multiplicateur. Pour symboliser cette nouvelle variable dont on cherche la moyenne, x est remplacé par y. Cette notation sera utile pour montrer aisément les liens entre le taux de croissance et le coefficient multiplicateur et le calcul de leur moyenne.

Tableau 3. Variation de la population et les taux annuels selon les trois hypothèses de croissance

i	Année	Population au 1er janvier	Croissance annuelle	Taux de croissance selon l'hypothèse…		
				… géométrique	… linéaire	… exponentielle
1	2011	10.000	15.000	1,50000	0,85714	0,91629
2	2012	25.000	40.000	1,60000	0,88889	0,95551
3	2013	65.000	22.000	0,33846	0,28947	0,29152
4	2014	87.000	-	-	-	-

Dans ce contexte, x_1 désigne le taux de croissance en 2011. L'évolution de la population entre le 1er janvier 2011 *(P_1)* et le 1er janvier 2012 *(P_2)* selon les différentes hypothèses s'exprime comme suit

- Hypothèse géométrique :

$$P_2 = P_1 * (1 + x_1). \tag{3}$$

- Hypothèse linéaire (plus de détails en annexe 1) :

$$P_2 = P_1 * \left(\frac{1 + 0{,}5 * x_1}{1 - 0{,}5 * x_1}\right) = P_1 * \left(\frac{2 + x_1}{2 - x_1}\right). \tag{4}$$

Cette formule peut surprendre. Pourtant, elle est présente dans Pressat 1980, p. 157.

- Hypothèse exponentielle :

$$P_2 = P_1 * e^{x_1}. \tag{5}$$

Encadré 2. Un risque de confusion à éviter

Il existe un risque de confusion entre les deux notions suivantes :

- l'*hypothèse géométrique* décrite par l'équation (3) montrant le passage d'une population initiale à une population finale ;
- la *formule géométrique* de la moyenne, soit la racine énième du produit des *n* observations.

Ces deux notions ne sont pas superposables, malgré la présence de l'adjectif « géométrique » dans leurs appellations respectives. Par exemple, le choix de l'hypothèse géométrique de croissance pour le calcul des taux n'impose pas la formule géométrique pour le calcul de leur moyenne.

Le taux de croissance (annuel) en 2011 pour les trois hypothèses se calcule comme suit selon des formules dérivées des équations précédentes :

- Hypothèse géométrique :

$$x_1 = \frac{P_2 - P_1}{P_1} = \frac{25.000 - 10.000}{10.000} = 1,50000$$

- Hypothèse linéaire (cf. Annexe 1) :

$$x_1 = \frac{P_2 - P_1}{0,5 * (P_1 + P_2)} = \frac{25.000 - 10.000}{0,5 * (10.000 + 25.000)} = 0,85714. \quad (6)$$

- Hypothèse exponentielle :

$$x_1 = ln\left(\frac{P_2}{P_1}\right) = ln\left(\frac{25.000}{10.000}\right) = 0,91629.$$

Ces formules reproduisent celles de la 1ère ligne du Tableau 1. Notons que, dans ce texte, les résultats sont donnés avec une précision suffisamment importante (par exemple, 5 décimales ici) pour donner à chacun·e la possibilité de vérifier les calculs sans trop craindre les effets d'arrondi.

Tout est en place maintenant pour faire émerger les formules du calcul du taux moyen selon les trois hypothèses de croissance et reprenant les valeurs observées du taux afin de déterminer si ces formules respectent ou non la règle qui voudrait que le taux de croissance moyen se calcule toujours selon la formule géométrique. À partir de ce point, l'ordre suivi pour aborder les hypothèses est modifié par rapport aux premiers tableaux. En effet, nous les envisagerons par ordre de difficulté croissante. En suivant ce principe, nous commençons ici par l'hypothèse exponentielle (avec le recours à la formule de la moyenne la plus connue) et finirons par l'hypothèse linéaire (avec le raisonnement le moins immédiat).

3.1. Taux moyen en cas d'hypothèse exponentielle

Au départ de l'équation (5), la population au 01/01/2014 peut se déduire de la population au 01/01/2011 et du taux de croissance des trois années :

$$P_4 = P_1 * (e^{x_1} * e^{x_2} * e^{x_3}) = P_1 * e^{\sum_{i=1}^{3} x_i}.$$

$$P_4 = 10.000 * e^{2,16332} = 87.000.$$

L'application du principe de base selon lequel le taux moyen est celui qui, se répétant durant les trois années d'observation, produit la même variation de population, se traduit comme suit, avec :

$$\sum_{i=1}^{3} x_i = \sum_i x_i \ :$$

$$P_4 = P_1 * e^{\sum_i x_i} = P_1 * e^{\sum_i \bar{x}}. \quad (7)$$

Si n est égal au nombre de taux observés, cette équation s'écrit d'une manière générale :

$$P_{n+1} = P_1 * e^{\sum_i x_i} = P_1 * e^{\sum_i \bar{x}} = P_1 * e^{n*\bar{x}}.$$

On en déduit que :

$$\sum x_i = n * \bar{x} \quad \Rightarrow \quad \bar{x} = \frac{\sum_i x_i}{n} = 0{,}72111. \tag{8}$$

Dans le cas de l'hypothèse exponentielle, le taux moyen est donc obtenu via une formule arithmétique. Le calcul trivial, vu la démonstration ci-dessus, montre que la moyenne obtenue permet de reconstituer la variation observée sur les trois années :

$$P_4 = P_1 * e^{3*0{,}72111} = 87.000.$$

3.2. Taux moyen en cas d'hypothèse géométrique

Au départ de l'équation (3), la population au 01/01/2014 peut se déduire de la population au 01/01/2011 et du taux de croissance des trois années :

$$P_4 = P_1 * (1 + x_1) * \ (1 + x_2) * (1 + x_3) = P_1 * \prod_i (1 + x_i). \tag{9}$$

Si n est égal au nombre de taux observés, l'application du principe de base implique que :

$$\prod_i (1 + x_i) = \prod_i (1 + \bar{x}) = (1 + \bar{x})^n \ \Rightarrow \ \bar{x} = \sqrt[n]{\prod_i (1 + x_i)} - 1 = 1{,}05671. \tag{10}$$

Dans le cas de l'hypothèse géométrique, le taux moyen est donc obtenu via une formule dont l'appellation ne nous est pas connue, mais qui **n'est pas** une formule… géométrique, malgré la présence de la racine énième d'un produit. La moyenne obtenue permet de reconstituer la variation observée de la population :

$$P_4 = \ P_1 * (1 + 1{,}05671)^3 = 87.000\,. \tag{11}$$

3.3. Taux moyen en cas d'hypothèse linéaire

Au départ de l'équation (4), la population au 01/01/2014 peut se déduire de la population au 01/01/2011 et du taux de croissance des trois années (pour une démonstration détaillée, *cf.* annexe 1) :

$$P_4 = \ P_1 * \left(\frac{2 + x_1}{2 - x_1}\right) * \left(\frac{2 + x_2}{2 - x_2}\right) * \left(\frac{2 + x_3}{2 - x_3}\right) = P_1 * \prod_i \left(\frac{2 + x_i}{2 - x_i}\right). \tag{12}$$

Si n est égal au nombre de taux observés, l'application du principe de base implique que (*cf.* annexe 1) :

$$\prod_i \left(\frac{2 + x_i}{2 - x_i}\right) = \prod_i \left(\frac{2 + \bar{x}}{2 - \bar{x}}\right) = \left(\frac{2 + \bar{x}}{2 - \bar{x}}\right)^n \ \Rightarrow \ \bar{x} = 2 * \frac{\sqrt[n]{\prod_i \left(\frac{2 + x_i}{2 - x_i}\right)} \ - \ 1}{\sqrt[n]{\prod_i \left(\frac{2 + x_i}{2 - x_i}\right)} \ + \ 1} = 0{,}69140. \tag{13}$$

L'appellation de cette formule (13) ne nous est pas connue, mais **il ne s'agit pas** non plus d'une formule géométrique. La moyenne ainsi obtenue permet de reconstituer la variation observée :

$$P_4 = P_1 * \left(\frac{2 + 0{,}69140}{2 - 0{,}69140}\right)^3 = 87.000. \quad (14)$$

Ces exemples montrent bien que le taux de croissance moyen durant une période n'est jamais calculé via une formule géométrique. Est-ce à dire que la formule géométrique n'est jamais de mise dans l'analyse de la variation temporelle d'une variable ? Le point suivant montre que non.

4. Un exemple de formule géométrique de la moyenne : le coefficient multiplicateur

Le « coefficient multiplicateur » se définit comme le nombre par lequel il faut multiplier la population initiale de la période pour laquelle il est calculé pour obtenir la population finale. Il se calcule en divisant la population finale par la population initiale. Le taux de croissance et le coefficient multiplicateur représentent deux façons d'exprimer la croissance relative d'une variable dans le temps. Vu sa définition et au contraire du taux de croissance, la valeur du coefficient multiplicateur ne varie pas selon l'hypothèse de croissance retenue. Par contre, sa relation avec le taux de croissance est, elle, différente selon l'hypothèse de croissance (point 4.1). Le coefficient multiplicateur offre des possibilités intéressantes pour faciliter et surtout accélérer le calcul du taux de croissance (point 4.2).

4.1. Le coefficient multiplicateur moyen et sa relation avec le taux de croissance

Entre le 1er janvier de deux années successives, le coefficient multiplicateur (symbolisé par *y*) se calcule en divisant la population la plus récente par la plus ancienne (pour rappel, cette définition est d'application pour les trois hypothèses de croissance). En conséquence, multiplier la valeur la plus ancienne par ce coefficient donne la valeur la plus récente. Pour suivre, calcul du coefficient multiplicateur pour la 1ère année du Tableau 3 :

$$y_1 = \frac{25.000}{10.000} = 2{,}50000 \qquad \text{et} \qquad 10.000 * 2{,}50000 = 25.000.$$

La même procédure peut s'appliquer pour les deux années suivantes (2,60000 ; 1,33846). Dès lors, la population au 01/01/2014 peut se déduire de la population au 01/01/2011 et des coefficients multiplicateurs des trois années :

$$P_4 = P_1 * y_1 * y_2 * y_3 = P_1 * \prod_i y_i. \quad (15)$$

Si n égale le nombre de coefficients observés, l'application du principe de base implique que :

$$\prod_i y_i = \prod_i \bar{y} = \bar{y}^n \Rightarrow \bar{y} = \sqrt[n]{\prod_i y_i} = 2{,}05671. \quad (16)$$

Le coefficient multiplicateur moyen est donc obtenu via une formule géométrique au sens strict (racine énième du produit des n valeurs dont on cherche la moyenne). La moyenne ainsi obtenue permet de reconstituer l'évolution observée de la population sur les trois années :

$$P_4 = P_1 * (2{,}05671)^3 = 87.000. \quad (17)$$

C'est dans le cadre de l'hypothèse géométrique que le lien entre le taux de croissance et le coefficient multiplicateur s'exprime de la façon la plus évidente : le coefficient correspond à $(1+x_i) = y_i$ (*cf.* formules (9) et (15)). Pour les deux autres hypothèses, la situation est moins évidente. C'est la raison pour laquelle l'hypothèse géométrique sera la première envisagée ici. Vu la relation ente le taux et le coefficient, le coefficient moyen peut se déterminer au départ du taux moyen :

$$\bar{y} = 1 + \bar{x}.$$

Il peut aussi se calculer selon un autre formule reprenant les taux de croissance observés :

$$\bar{y} = \sqrt[n]{\prod_i y_i} = \sqrt[n]{\prod_i 1 + x_i}. \qquad (18)$$

L'équation (18) établit le lien entre le coefficient multiplicateur et le taux moyen dans le cadre de l'hypothèse géométrique. Qu'en est-il de ce lien pour les deux autres hypothèses de croissance ? En se limitant aux équations essentielles :

- Hypothèse exponentielle :

$$y_i = e^{x_i} = \frac{P_{i+1}}{P_i}.$$

$$\bar{y} = \sqrt[n]{\prod_i y_i} = \sqrt[n]{\prod_i e^{x_i}}.$$

- Hypothèse linéaire (qui donne les écritures les moins immédiates (*cf.* annexe 1))

$$y_i = \left(\frac{2 + x_i}{2 - x_i}\right) = \frac{P_{i+1}}{P_i}.$$

$$\bar{y} = \sqrt[n]{\prod_i y_i} = \sqrt[n]{\prod_i \left(\frac{2 + x_i}{2 - x_i}\right)}.$$

À chaque fois, sans surprise, le coefficient multiplicateur moyen s'obtient via une formule géométrique de la moyenne appliquée aux coefficients multiplicateurs annuels, chacun étant une expression incluant les taux de croissance x_i, de manière simple pour les hypothèses géométrique $(1+x_i)$ ou exponentielle (e^{x_i}) ou, moins immédiate, pour l'hypothèse linéaire $((2+x_i)/(2-x_i))$.

4.2. Le coefficient multiplicateur et les formules directes

Pour le coefficient multiplicateur, une formule bien connue permet le calcul aisé de sa moyenne au départ des populations initiale et finale, ce qui justifie de la qualifier de « directe » :

$$P_{n+1} = P_1 * \prod_i y_i \;\Rightarrow\; \prod_i y_i = \frac{P_{n+1}}{P_1} \;\Rightarrow\; \bar{y} = \sqrt[n]{\frac{P_{n+1}}{P_1}} = 2{,}05671.$$

Ce type de formule directe peut aussi s'utiliser pour le **calcul du taux moyen** selon les trois hypothèses, ce qui revient à inclure dans la formule le rapport entre les populations initiale et finale :

- Hypothèse géométrique au départ de la formule (10) :

$$P_{n+1} = P_1 * \prod_i (1 + \bar{x}) \;\Rightarrow\; \frac{P_{n+1}}{P_1} = (1 + \bar{x})^n \;\Rightarrow\; \bar{x} = \sqrt[n]{P_{n+1}/P_1} - 1. \qquad (19)$$

- Hypothèse exponentielle au départ de la formule (7) :

$$P_{n+1} = P_1 * e^{n*\bar{x}} \;\Rightarrow\; \frac{P_{n+1}}{P_1} = e^{n*\bar{x}} \;\Rightarrow\; \bar{x} = \frac{ln(P_{n+1}/P_1)}{n}.$$

- Hypothèse linéaire au départ de la formule (12) (*cf.* annexe 1) :

$$P_{n+1} = P_1 * \prod_i \left(\frac{2+\bar{x}}{2-\bar{x}}\right) \;\Rightarrow\; \prod_i \left(\frac{2+\bar{x}}{2-\bar{x}}\right) = \frac{P_{n+1}}{P_1} \;\Rightarrow\; \bar{x} = 2\frac{\sqrt[n]{\frac{P_{n+1}}{P_1}} - 1}{\sqrt[n]{\frac{P_{n+1}}{P_1}} + 1}. \quad (20)$$

Sous hypothèses exponentielle et géométrique, ces formules directes sont bien connues et couramment utilisées (Tableau 1). Ce n'est pas le cas pour l'équation sous hypothèse linéaire. En fait, cette formule (20) n'est pas compatible avec la formule (1) du Tableau 1, comme va le montrer l'application de ces deux formules aux données du Tableau 3 (la vérification qui suit le calcul de la moyenne est réalisée en toute logique par la formule (14)) :

- avec la moyenne calculée selon la formule (1) :

$$\bar{x} = \frac{77.000}{3 * \frac{97.000}{2}} = 0{,}52921 \quad \text{et} \quad P_4 = 10.000 * \left(\frac{2+0{,}52921}{2-0{,}52921}\right)^3 = 50.851{,}3.$$

- avec la moyenne calculée selon la formule (20) :

$$\bar{x} = 2 * \frac{\sqrt[3]{\frac{87.000}{10.000}} - 1}{\sqrt[3]{\frac{87.000}{10.000}} + 1} = 0{,}69140 \quad \text{et} \quad P_4 = 10.000 * \left(\frac{2+0{,}69140}{2-0{,}69140}\right)^3 = 87.000\,.$$

Alors que le taux moyen selon la formule (20) fait passer en trois ans la population de 10.000 à 87.000 comme observé, la moyenne selon la formule (1) donne seulement 50.851,3 en contradiction avec le principe de base. Pour comprendre d'où vient l'erreur, il est utile d'établir l'équation du passage de la population initiale à la population finale correspondant à la logique de la formule habituelle. On peut démontrer que :

$$\bar{x} = \frac{P_{n+1} - P_1}{n * (P_1 + P_{n+1})/2} \;\Rightarrow\; P_{n+1} = P_1 * \left(\frac{2+n\bar{x}}{2-n\bar{x}}\right).$$

Cette expression doit être comparée à la formule (20) qui implique que :

$$P_{n+1} = P_1 * \left(\frac{2+\bar{x}}{2-\bar{x}}\right)^n.$$

Pour que les parenthèses de ces deux équations soient équivalentes, il faut que :

- $n = 1$, ce qui revient à dire que l'observation ne concerne qu'une année ;
- et/ou $\bar{x} = 0$, ce qui revient à considérer une croissance nulle.

On peut en conclure que l'égalité entre les deux formules (1) et (20) n'est pas acceptable. L'erreur de la formule habituelle du taux moyen (1) vient du fait qu'elle n'a pas été déduite de l'évolution de la population au départ des taux observés (principe de base), mais d'une simple adaptation de la formule du taux annuel sans en vérifier la validité. La formule (1) étant inappropriée, elle doit être définitivement remplacée par la formule (20) (pour un exemple chiffré, voir l'annexe 2).

Utiliser les formules directes permet de calculer à moindres frais le taux moyen, notamment parce que le calcul des taux annuels n'est plus un préalable à celui de leur moyenne et qu'il n'est plus nécessaire de les combiner pour obtenir la moyenne. Par contre, ces formules ne montrent pas la nature exacte du calcul du taux moyen au départ des taux observés. Cet inconvénient ne gomme pas les avantages pratiques de ces formules, particulièrement en cas de longue période d'observation.

5. Récapitulatif des formules et conséquences terminologiques du principe de base

Le Tableau 4 donne une vue synthétique du raisonnement suivi et des formules citées dans le texte dans le cadre des trois hypothèses de croissance. Le processus calculatoire suivi dans ce texte comporte les étapes suivantes : pour chacune des trois hypothèses de croissance, calcul des taux et coefficients multiplicateurs annuels et puis calcul de leur moyenne, avec en complément la mise en place de formules directes.

Tableau 4. Récapitulatif des formules selon les 3 hypothèses de croissance

Hypothèse…	… géométrique	… linéaire	… exponentielle
Taux de croissance sur une année (x_i)	$x_i = \frac{P_{i+1} - P_i}{P_i}$	$x_i = \frac{P_{i+1} - P_i}{(P_i + P_{i+1})/2}$	$x_i = ln\left(\frac{P_{i+1}}{P_i}\right)$
Évolution de la population sur un an	$P_{i+1} = P_i(1 + x_i)$	$P_{i+1} = P_i\left(\frac{2 + x_i}{2 - x_i}\right)$	$P_{i+1} = P_i e^{x_i}$
Évolution de la population sur n années	$P_{n+1} = P_1 \prod_{i=1}^{n}(1 + x_i)$	$P_{n+1} = P_1 \prod_{i=1}^{n}\left(\frac{2 + x_i}{2 - x_i}\right)$	$P_{n+1} = P_1 e^{\sum_{i=1}^{n} x_i}$
Taux moyen calculé sur une période de n années	$\bar{x} = \sqrt[n]{\prod_{i=1}^{n}(1 + x_i)} - 1$	$\bar{x} = 2\frac{\sqrt[n]{\prod_{i=1}^{n}\left(\frac{2 + x_i}{2 - x_i}\right)} - 1}{\sqrt[n]{\prod_{i=1}^{n}\left(\frac{2 + x_i}{2 - x_i}\right)} + 1}$	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
P_{n+1}/P_1 : expression de la variation relative totale avec $\bar{x}$	$\frac{P_{n+1}}{P_i} = \prod_{i=1}^{n}(1 + \bar{x})$	$\frac{P_{n+1}}{P_i} = \prod_{i=1}^{n}\left(\frac{2 + \bar{x}}{2 - \bar{x}}\right)$	$\frac{P_{n+1}}{P_1} = e^{\sum_{i=1}^{n} \bar{x}}$
Calcul direct du taux moyen sur une période de n années en utilisant P_{n+1}/P_1	$\bar{x} = \sqrt[n]{\frac{P_{n+1}}{P_1}} - 1$	$\bar{x} = 2\frac{\sqrt[n]{\frac{P_{n+1}}{P_1}} - 1}{\sqrt[n]{\frac{P_{n+1}}{P_1}} + 1}$	$\bar{x} = \frac{ln(P_{n+1}/P_1)}{n}$
Coefficient multiplicateur sur un an (y_i)	$y_i = \frac{P_{i+1}}{P_i} = 1 + x_i$	$y_i = \frac{P_{i+1}}{P_i} = \frac{2 + x_i}{2 - x_i}$	$y_i = \frac{P_{i+1}}{P_i} = e^{x_i}$
Coefficient multiplicateur moyen sur une période de n années	$\bar{y} = \sqrt[n]{\prod_{i=1}^{n} y_i}$ (même formule pour les trois hypothèses)		
Produit sous la racine exprimé via le rapport P_{n+1}/P_1	$\prod_{i=1}^{n} y_i = \frac{P_{n+1}}{P_1}$ (même formule pour les trois hypothèses)		
Calcul direct du coefficient multiplicateur moyen sur n années	$\bar{y} = \sqrt[n]{\frac{P_{n+1}}{P_1}}$ (même formule pour les trois hypothèses)		

Il faut rappeler que la case restée vide dans la colonne de l'hypothèse linéaire du Tableau 1 a été comblée. Par ailleurs, la formule du taux moyen dans cette colonne a été remplacée par une autre dont nous avons montré la logique par rapport au principe de base (*cf.* point 5). Ce dernier a ainsi montré toute son utilité pour faire progresser la réflexion théorique en ce qui concerne le taux de croissance moyen en cas d'hypothèse linéaire de croissance.

Dans ce récapitulatif, on voit apparaitre des formules de la moyenne à l'appellation bien connue. Pour les désigner dans le texte, pourquoi avons-nous utilisé, par exemple, l'expression « formule arithmétique de la moyenne » plutôt que l'expression habituelle « moyenne arithmétique » ? La raison est simple : même si le taux moyen en cas d'hypothèse exponentielle est obtenu via une formule arithmétique (8) ou si, pour le coefficient multiplicateur, il s'agit d'une formule géométrique (16), le premier résultat ne se transforme pour autant pas en une « *moyenne arithmétique* » et le second, en une « *moyenne géométrique* ». Pour ces deux résultats, il s'agit tout simplement de la « *moyenne* », soit la seule valeur qui respecte le principe de base. Il n'y a donc aucune raison d'individualiser ces moyennes en leur ajoutant un adjectif.

Les adjectifs « *arithmétique* » et « *géométrique* » ne peuvent donc pas porter sur le mot « *moyenne* ». Par contre, ils retrouvent une certaine utilité s'ils portent sur le mot « *formule* », à la réserve près que, à notre connaissance, toutes les formules de la moyenne (une infinité sur le plan théorique) ne disposent pas d'un adjectif connu, comme c'est le cas du taux de croissance moyen avec les hypothèses linéaire et géométrique.

Ces commentaires mettent en évidence qu'il est toujours légitime de poser la question : « que vaut le taux annuel moyen ? » Ajouter à cette question un adjectif comme *arithmétique* ou *géométrique* (ou autre) pourrait laisser accroire l'existence de plusieurs moyennes différentes alors que l'objectif poursuivi par ce paramètre est toujours le même : respecter le principe de base. Ainsi, l'unicité du concept est établie en parallèle avec la multiplicité des formules pour son calcul.

6. Apports méthodologiques complémentaires

6.1. Le principe de base et cohérence des calculs des indices de croissance

En se référant au Tableau 4, il appert que, dans un même processus calculatoire, toutes les opérations doivent s'effectuer sous la même hypothèse de croissance. Ce principe n'a pas toujours été respecté en démographie. Ainsi, dans plusieurs publications, Pressat (1961, 1973, 1980) propose une réflexion sur le taux de croissance. Ces essais illustrent un déficit à ce sujet.

Dans Pressat (1961, pp. 294-300 ou 1973, p. 230), le taux d'accroissement est défini sur la base de l'hypothèse linéaire de croissance : le taux d'accroissement est calculé par rapport à la population moyenne. Cependant, plus loin, les formules pour calculer le temps de doublement et la population après 100 ans en supposant un taux de croissance constant sont écrites sous hypothèse de croissance géométrique. Dans Pressat (1983, pp. 225-227 et pp. 239-240), on retrouve ce type de présentation. N'est-ce pas une erreur de mélanger deux hypothèses de croissance dans un même processus calculatoire ?

Pressat (1980, pp. 155-158) rectifie partiellement le tir en ajoutant une étape intermédiaire pour tenter justifier ce passage de calculs sous hypothèse linéaire à d'autres sous hypothèse géométrique et éviter ainsi le caractère brutal de ce changement d'hypothèses :

- Le taux d'accroissement est défini sur la base de l'évolution linéaire. Pressat ajoute l'équivalent de l'équation (4), progrès conceptuel significatif par rapport à la publication de 1973, mais dont il ne tire aucun profit pour la suite de l'exposé théorique.

- Une étape intermédiaire signale que les hypothèses linéaire et géométrique donnent des résultats où les différences sont négligeables, principe dégagé sur la base d'un seul exemple avec un taux de croissance « modéré » de 2%. (Cette étape est décrite avec plus de détail plus bas).

- Ensuite, vu le caractère négligeable des différences, les calculs du temps de doublement et de la population après 100 ans sont faits sous hypothèse de croissance géométrique avec des formules qui seraient « commodes » pour ces calculs.

Cette confusion est étonnante, car Pressat écrit toutes les équations nécessaires pour l'éviter, dont l'équation pour calculer la population à un moment donné au départ de la population un an auparavant et du taux de croissance sous hypothèse de croissance linéaire. De plus, l'exemple utilisant un taux de croissance relativement faible (2%) empêche la généralisation de la conclusion sur la possibilité de prendre une valeur issue d'un calcul fait sous une hypothèse pour l'utiliser dans d'autres calculs réalisés sous une autre hypothèse. En effet, si, en général, au vu des niveaux actuellement habituels de croissance dans certaines populations humaines, cette incohérence a peu d'incidence sur les résultats, d'autres circonstances donnent à voir des taux de croissance et de décroissance drastiquement plus importants, comme l'évolution des hospitalisations et des décès liés à la Covid à l'échelle hebdomadaire. Ce type de variation entrainerait des différences marquées selon l'hypothèse de croissance retenue, ce qui invalide la procédure suivie par Pressat. Ainsi, en cas de doublement en un an (comme l'évolution du nombre de malades en début d'épidémie du sida), le taux de croissance selon l'hypothèse géométrique est de 100 %, mais de 66,67% selon l'hypothèse linéaire (Wunsch *et al.* 2001, p. 26). Selon l'hypothèse exponentielle, ce taux est de 69,31 %.

Par ailleurs, pour sa démonstration, Pressat part d'un taux de croissance (désigné par « a ») valant 0,02 et défini selon la formule d'application en cas d'hypothèse linéaire. Ensuite, il calcule le coefficient multiplicateur (*CM*) selon les hypothèses linéaire et géométrique avec la valeur de 0,02 :

• Hypothèse linéaire : $$CM = \frac{(2+a)}{(2-a)} = 1{,}0202.$$

• Hypothèse géométrique : $$CM = (1+a) = 1{,}02.$$

Pressat obtient donc deux valeurs distinctes pour le coefficient. Le Tableau 5 permet de critiquer la démonstration de Pressat en supposant que $P_i = 99.000$ et $P_{i+1} = 101.000$. Pour chacune des deux hypothèses, le taux et le coefficient multiplicateur ont été calculés. Dans le cas de l'hypothèse linéaire, le taux vaut précisément 0,02, soit la valeur choisie par Pressat pour sa démonstration ; cette valeur a été calculée selon une formule linéaire (Pressat 1980, formule (E) p. 156). La valeur des taux diffère selon l'hypothèse, mais pas la valeur du coefficient multiplicateur, ce qui est conforme à la théorie.

La démonstration de Pressat ne repose pas sur une base acceptable car, dans un même processus calculatoire, une seule hypothèse doit régir toutes les étapes exécutées, par exemple les calculs des taux annuels, du taux moyen, du temps de doublement ou de la croissance de la population à l'horizon de 100 ans.

Tableau 5. Taux de croissance *(a)* et coefficients multiplicateurs *(CM)* sur un an

Hypothèse linéaire	$a = \frac{101.000 - 99.000}{0,5 * (101.000 + 99.000)} = 0,02000.$
	$CM = \frac{(2 + 0,02)}{(2 - 0,02)} = 1,0202\ .$
Hypothèse géométrique	$a = \frac{101.000 - 99.000}{99.000} = 0,02020.$
	$CM = (1 + 0,02020) = 1,02020.$

6.2. Méthode essai-erreur versus *principe de base : avantage à ce dernier*

En cas de calcul de la moyenne de ratios ou de facteurs d'accroissement, Wonnacott et Wonnacott (1990, p. 667) aboutissent à l'équation ((16)), mais en suivant la méthode essai-erreur (pour une autre illustration de cette procédure, *cf.* Dehon *et al*. (2015, pp. 100-101)). Dans l'exemple de Wonnacott et Wonnacott, l'effectif d'une population d'une banlieue passe de 1.000 à 2.000 puis à 16.000 respectivement en 1950, 1960 et 1970, avec un coefficient multiplicateur décennal de 2 entre 1950 et 1960 et de 8 entre 1960 et 1970. La question porte sur le *coefficient décennal moyen*. Comme 1er essai, les auteurs utilisent la formule arithmétique :

$$\bar{y} = (2 + 8)/2 = 5.$$

Ensuite, ils montrent que le résultat n'est pas adéquat. L'explication proposée recourt implicitement au principe de base : « *If a fivefold increase occurred over two successive decades, the overall increase would be twenty-fivefold, not sixteen* ». En effet, énoncer cette dernière phrase suppose d'avoir identifié l'équation (17) pour réaliser le calcul qui montre le caractère inacceptable du 1er résultat obtenu.

Suite à ce constat d'échec, les auteurs introduisent une 2e formule : « *Instead, let us try the geometric mean* », expression qui indique explicitement le recours à la méthode essai-erreur :

$$\bar{y} = \sqrt{2 * 8} = 4.$$

Pour finir, ils vérifient la cohérence de ce nouveau résultat toujours en recourant au principe de base : « *If a fourfold increase occurred over two successive decades, the overall increase would be sixteenfold that actually did occur* ».

Pourquoi utiliser, *a posteriori* (après un essai), le principe de base pour vérifier la cohérence du résultat ? Pourquoi ne pas l'utiliser *a priori*, sans idée préconçue sur la formule déduite logiquement de ce principe ? Prolongeons ce questionnement en prenant l'exemple du taux en cas d'hypothèse linéaire (équation (13)) : combien faudra-t-il d'essais-erreurs avant de trouver la formule appropriée ? Sera-t-il simplement possible d'identifier la bonne formule en suivant cette méthode ? Sur un plan simplement pratique, comment introduire, dans le développement de la méthode essai-erreur, la formule (13) vu qu'elle n'a pas d'appellation ? Au contraire, le principe de base permet d'identifier la formule adéquate sans difficulté majeure, y compris dans cette circonstance. C'est un avantage incontestable à mettre au crédit de cette procédure. Par ailleurs, la méthode essai-erreur peut déboucher sur des conclusions fausses.

Ainsi, supposons une personne ayant eu un cours de statistique où les formules arithmétique, géométrique, harmonique, contra-harmonique (ou « *self weighted mean* » ; *cf.* par exemple, Lann et Falk 2006) et quadratique ont été explicitées (y compris, donc, des formules non citées au préalable dans ce texte). En reprenant les données du Tableau 3, il lui est demandé d'utiliser la méthode essai-erreur pour déterminer la formule adéquate pour calculer **la moyenne des trois taux calculés selon l'hypothèse de croissance géométrique**. Attirée par la formule géométrique, elle commencera donc par cette formule pour passer après, si nécessaire, aux formules arithmétique, harmonique, contra-harmonique et quadratique (sans ordre particulier). Les résultats sont repris dans le Tableau 6. La précision des résultats est justifiée pour donner la possibilité de vérifier en profondeur les calculs.

Tableau 6. Taux moyen si hypothèse géométrique : inefficacité de la méthode essai-erreur

	Type de formule de la moyenne				
	Géométrique	**Arithmétique**	**Harmonique**	**Contra-harmonique**	**Quadratique**
Formule de la moyenne	$\bar{x} = \sqrt[n]{\prod x_i}$	$\bar{x} = \frac{\sum x_i}{n}$	$\bar{x} = \frac{n}{\sum 1/x_i}$	$\bar{x} = \frac{\sum (x_i)^2}{\sum x_i}$	$\bar{x} = \sqrt[2]{\frac{\sum (x_i)^2}{n}}$
$\bar{x}$	0,93305	1,14615	0,70651	1,43220	1,28122
Vérification	72.232,40402	98.851,33819	49.696,75962	143.878,71268	118.713,52071

La moyenne selon la formule géométrique ne réussit pas le test de vérification (équation (11)) avec moins de 73.000 à comparer aux 87.000 qui ont été observés (Tableau 3) :

$$P_4 = P_1 * (1 + 0{,}93305)^3 = 72.232{,}4\,.$$

Les essais suivants ne réussiront pas plus le test. Il ne semble pas possible de calculer le taux de croissance moyen sur la période. Cette conclusion est évidemment fausse. Le recours au principe de base donne un taux moyen selon l'hypothèse géométrique de croissance de 1,05671 (équation (10)), qui réussit le test :

$$P_4 = P_1 * (1 + 1{,}05671)^3 = 87.000{,}0\,.$$

Comme la formule utilisée n'a pas d'appellation connue, il est bien difficile de l'inclure dans l'exercice, à moins de la baptiser par une périphrase du genre : « formule adaptée au calcul du taux de croissance moyen en cas d'hypothèse géométrique », ce qui supposerait d'avoir établi au préalable la validité de la formule dans cette circonstance. En procédant de la sorte, on peut s'interroger sur l'intérêt même de l'exercice dans la forme décrite. Ce constat illustre une limitation importante dans le recours à la méthode essai-erreur. Le principe de base dépasse toutes les difficultés.

Par ailleurs, l'emploi de la méthode essai-erreur en cas de croissance faible peut en révéler toute la fragilité. Le Tableau 7.a montre une population ayant connu une croissance (très) faible : au départ du même effectif que dans le Tableau 3 (10.000), la croissance absolue est pratiquement 200 fois plus faible ici (400 contre 77.000). La colonne y_i du Tableau 7.a reprend **la valeur des coefficients multiplicateurs** au cours des trois années. Vu la nature de cette variable, il n'est pas nécessaire de préciser l'hypothèse de croissance retenue. Reprenons la démonstration de Wonnacott et Wonnacott avec ces données.

Le Tableau 7.b montre la valeur du coefficient moyen telle que calculée selon les cinq mêmes formules que dans le Tableau 6, ainsi qu'un calcul de vérification visant à voir si le coefficient moyen appliqué durant trois années produit bien la population finale (formule (17)) :

$$P_4 = P_1 * (\bar{y})^3.$$

Tableau 7. *Coefficient multiplicateur* en cas de croissance faible : inefficacité de la méthode essai-erreur

Tableau 7.a

A	Population	y_i
01/01/2011	10.000	1,01290
01/01/2012	10.129	1,01293
01/01/2013	10.260	1,01365
01/01/2014	10.400	-

Tableau 7.b

	Arithmétique	Géométrique	Harmonique	Contra-harmonique	Quadratique
Formule	$\bar{y} = \frac{\sum y_i}{n}$	$\bar{y} = \sqrt[n]{\prod y_i}$	$\bar{y} = \frac{n}{\sum 1/y_i}$	$\bar{y} = \frac{\sum (y_i)^2}{\sum y_i}$	$\bar{y} = \sqrt[2]{\frac{\sum (y_i)^2}{n}}$
A $\bar{y}$ (5 décimales)	1,01316	1,01316	1,01316	1,01316	1,01316
B Vérification	10.400,0	10.400,0	10.400,0	10.400,0	10.400,0
C $\bar{y}$ (7 décimales)	1,0131595	1,0131594	1,0131593	1,0131596	1,0131595
D Vérification	10.400,002	10.400,000	10.399,998	10.400,005	10.400,004

Comme fait par Wonnacott et Wonnacott, le 1er essai est réalisé selon la formule arithmétique. Si l'analyse des résultats devait se baser sur les lignes A (taux moyen avec 5 décimales, soit des millièmes de %) et la ligne B (calcul de la vérification, avec une décimale à l'effectif après 3 ans), la 1ère moyenne obtenue selon la formule arithmétique réussirait le test. Il n'y aurait donc aucune raison de procéder à d'autres essais. Toutefois, si malgré cela, les quatre autres formules de la moyenne sont testées, la conclusion pourrait être qu'indépendamment du choix de la formule, le test est réussi. Cette situation pourrait susciter un doute quant à la pertinence de la recherche de la formule adéquate de la moyenne.

Dans la ligne C, les taux moyens apparaissent avec 7 décimales, soit un degré de précision inhabituel. Dans la ligne D, le résultat de la vérification est donné avec 3 décimales pour l'effectif après 3 ans, ce qui pourrait surprendre vu la nature de cette grandeur qui s'exprime en individus. Si l'analyse des résultats se base sur ces lignes C et D, seule la formule géométrique de la moyenne produit un résultat qui réussit le test. La formule géométrique s'impose donc au détriment des quatre autres, ce qui est la conclusion correcte (*cf.* point 4.1). Cette situation souligne une faiblesse supplémentaire de la méthode essai-erreur : il faut exiger une très grande précision dans les calculs pour aboutir à une conclusion solide, si du moins le travail se fait avec des croissances annuelles plutôt faibles. C'est pour éviter ce genre de situation que nous avons opté pour des exemples avec des niveaux de croissance importants, voire irréalistes. Wonnacott et Wonnacott avaient procédé de même, mais avec des coefficients multiplicateurs pour des décades, il est vrai.

En utilisant la méthode essai-erreur de la même façon que Wonnacott et Wonnacott, Lewis se propose de démontrer une règle explicite obligeant le recours à la formule géométrique de la moyenne dans certaines circonstances : *« Arithmetic means are appropriate for calculating the average of economic variables that are measured in units such as dollars, number of persons, or weeks of unemployment. (...) However, some economic activity is described by variables defined as rates of change, indices, or ratios. Because such variables have different mathematical properties, we must use a different formula, the geometric mean, to calculate the average value* » (Lewis 2012, pp. 100-102).

Cette démonstration se limite à une comparaison entre les formules arithmétique et géométrique de la moyenne, sans que ce choix initial des formules à comparer ne soit en rien justifié. Cette option initiale est difficilement compréhensible vu qu'il existe une infinité théorique de formules du calcul de la moyenne. Tout se passe comme si Lewis connaissait la conclusion avant de commencer la réflexion. Son raisonnement est annulé par toute une série d'imprécisions, d'approximations, d'erreurs qui auraient pu être évitées à peu de frais en utilisant le principe de base. Parmi d'autres, voici une chrestomathie de points faibles dans la démonstration de Lewis :

• Lewis indique que le choix de la formule s'opère en fonction de caractéristiques mathématiques de la variable. En prenant un jeu de données à propos de la descendance, exprimée sous la forme d'un rapport (enfant(s) par individu), nous avons montré que la moyenne peut aussi bien s'obtenir via une formule arithmétique qu'harmonique ou géométrique selon la question posée (Vandeschrick et Wautelet 2003, pp. 74-76). L'affirmation de Lewis est donc sans fondement.

• Lewis écrit « *Note that we can calculate the geometric mean only when the product of all* X_i *is positive. If the product is negative or equal to zero, we cannot take the* n*th root* » (p. 100). Or, la présence de taux annuels négatifs (qui, en nombre impair, rendent le produit négatif) et/ou nuls (un seul taux nul annule le produit) correspond à des situations plausibles, voire banales, en matière de taux de croissance. Selon Lewis, dans des cas de ce type, il faudrait recourir à d'autres indices. Quel crédit accorder à une règle qui ne s'accommode pas de situations banales et obligerait à introduire d'autres indices pour traiter certaines situations alors que ce n'est pas indispensable pour qui applique le principe de base ? Par ailleurs, où est le problème de prendre la racine énième de 0 ?

• En comparant les résultats selon les formules arithmétique et géométrique de la moyenne, Lewis intervertit les deux résultats. Elle aurait donc dû conclure que la formule recherchée était l'arithmétique, soit l'inverse de sa conclusion.

• Enfin, plus loin dans son livre (pp. 196-197 et p. 204), Lewis calcule des taux moyens en utilisant des formules explicitées aux points 3 et 4 du présent texte en cas d'hypothèse de croissance géométrique ou exponentielle. Ces pages sont donc en contradiction avec la règle d'usage de la moyenne géométrique telle qu'énoncée en pages 100-102 de son livre.

En définitive, si Lewis avait suivi le principe de base sans idée préconçue sur le type de formule à utiliser, elle aurait évité de faire une démonstration illusoire de l'omniprésence de la formule géométrique en cas de variables « *as rates of change, indices, or ratios* » et aurait surtout déterminé correctement la bonne formule du calcul du taux de croissance moyen.

La méthode essai-erreur suppose d'établir/d'identifier, au préalable la liste des formules à tester (impossible à établir vu le cortège infini de formules envisageables) et la formule permettant de vérifier la cohérence du résultat avec les données observées. Par ailleurs, la précision des calculs doit être calibrée en fonction du niveau de la croissance. En somme, la méthode essai-erreur est tout simplement à proscrire. Le recours au principe de base est plus direct, plus simple à implémenter et surtout plus efficace, et libère de certaines contingences inhérentes à la méthode essai-erreur.

7. Conclusions

Les réflexions de ce texte consacré au calcul du taux de croissance moyen en cas d'analyse diachronique pluriannuelle reposent sur un principe de base : le taux de croissance annuel moyen est la valeur du taux qui, appliquée aux différentes années de la période d'observation, conduit à la variation de la population effectivement observée sur l'ensemble de cette période.

L'application de ce principe permet de vérifier la justesse de tous les développements théoriques démographiques habituels en matière de moyenne des taux de croissance ou des coefficients multiplicateurs. En cas d'hypothèse de croissance géométrique ou exponentielle, les formules directes habituelles du taux moyen reprises dans le Tableau 1 ont ainsi été retrouvées par application du principe de base. Il en va de même pour le calcul du coefficient multiplicateur moyen. La seule exception concerne le taux moyen en cas d'hypothèse linéaire. La formule reprise dans le Tableau 1 n'est pas correcte et doit être remplacée par une formule dérivée de l'application du principe de base. En outre, l'utilisation de ce principe a permis de démontrer que :

• Contrairement à une règle présente dans la littérature, la formule géométrique de la moyenne n'est pas indiquée pour trouver la valeur du taux de croissance moyen en cas d'analyse diachronique.

• Le calcul de la moyenne des taux, régi par le principe de base, se décline en différentes formules selon l'hypothèse de croissance en cause. Par ailleurs, les taux calculés selon une hypothèse donnée ne peuvent donner lieu qu'au calcul d'une seule valeur moyenne, soit la valeur en accord avec le principe de base. De ce fait, la seule question à poser face à une série de taux est « que vaut la moyenne » ? Si nécessaire, les adjectifs *arithmétique*, *géométrique*, *harmonique*… pourraient porter sur le mot « *formule* » après sa détermination, mais nullement sur le mot « *moyenne* » dans la question de savoir que vaut la moyenne.

• La méthode essai-erreur est fréquemment utilisée, notamment pour déterminer la bonne formule du calcul de la moyenne des taux ou des coefficients multiplicateurs. Cette méthode n'est pas conseillée. La principale difficulté vient du fait que certaines formules (dont les formules (10) et (13)) n'ont pas été clairement identifiées dans la typologie classique des moyennes. De ce fait, elles ne peuvent pas être reprises dans la liste des formules à tester. En conséquence, la méthode essai-erreur est inopérante et doit être évitée au profit du principe de base.

Bibliographie

Braverman J., 1978, *Fundamentals of Business Statistics,* New York, Academic Press, 597 p.

Dehon C., Droesbeke J.-J. et Vermandele C., 2015, Éléments de statistique (6e édition), Bruxelles – Paris, Éditions de l'Université de Bruxelles – Ellipses/ Édition Marketing.
https://doi.org/10.7202/029367ar

Lann A. and Falk R., 2006, Tell Me the Method, I'll Give You the Mean, The American Statistician, 20 (4), 322-327.
https://doi.org/10.1198/000313006x151460

Lewis M., 2012, *Applied Statistics for Economists*, London and New York, Routledge, 446 p.

Pressat R.,1961, *L'analyse démographique. Concepts-Méthodes-Résultats*, Paris, Presses Universitaires de France, 402 p.

Pressat R.,1973, *L'analyse démographique. Concepts-Méthodes-Résultats*, Paris, Presses Universitaires de France, 321 p.

Pressat R., 1980, *Démographie statistiqu*e, Paris, Presses Universitaires de France, 194 p.

Pressat R.,1983, *L'analyse démographique. Concepts-Méthodes-Résultats*, Paris, Presses Universitaires de France, 295 p.

Py B., 2007, *Statistique descriptive. Nouvelle méthode pour bien comprendre et réussir. 5e édition*, Paris, Economica, 353 p.

Spiegel M., 1984, *Théorie et applications de la statistique. 14e édition*, Auckland, McGraw-Hill, 358 p.

Vandeschrick C., 2017, « La moyenne : l'approche de Chisini revisitée. Exemples et enseignements », *Statistique et Enseignement* 8(1), p. 3–20.

Vandeschrick C. et Wautelet J.-M., 2003, *De la statistique descriptive aux mesures des inégalités*, Louvain-la-Neuve – Paris, Academia-Bruylant – L'Harmattan, 242 p.

Wonnacott T., Wonnacott R., 1990, *Introductory Statistics for Business and Economics (Fourth Edition)*, New York, John Wiley & Sons, 815 p.
https://doi.org/10.2307/3315129

Wunsch G., Vallin J. et Caselli G., 2001, L'accroissement de la population, *in* Caselli Graziella, Vallin Jacques et Wunsch Guillaume, *Démographie : analyse et synthèse. Volume I La dynamique des populations*, Chapitre 3, Paris, INED, pp. 23-33.
https://doi.org/10.7202/010855ar

Annexes

• Annexe 1. Taux de croissance moyen en cas d'hypothèse linéaire

L'établissement des différentes formules dans le cadre des hypothèses exponentielle et géométrique est plutôt immédiat. C'est moins vrai dans le cas de l'hypothèse linéaire. Cette annexe donne quelques explications à ce sujet en partant de l'équation du calcul du taux de croissance durant l'année *i* (formule(6)). En effet, cette dernière équation est bien établie sur le plan théorique et non mise en cause. L'évolution de la population sur an va être déduite de la formule du taux de croissance :

$$x_i = \frac{P_{i+1} - P_i}{0{,}5 * (P_i + P_{i+1})} \Rightarrow x_i * 0{,}5 * (P_i + P_{i+1}) = P_{i+1} - P_i \Rightarrow P_{i+1} * (1 - 0{,}5x_i) = P_i * (1 + 0{,}5x_i)$$

$$\Rightarrow P_{i+1} = P_i * \frac{(1 + 0{,}5x_i)}{(1 - 0{,}5x_i)}$$

$$\Rightarrow P_{i+1} = P_i * \frac{(2 + x_i)}{(2 - x_i)}.$$

Les deux dernières équations correspondent à la formule (4) du texte. Il est à noter que Pressat (1980, p. 156-158) obtient la dernière équation, mais sans en tirer profit pour la suite de son raisonnement (*cf.* point 6.1). Ce qui suit n'est donc pas repris dans la publication de Pressat. Après avoir appliqué la dernière formule à chacune des années sous observation, la variation de la population sur l'ensemble de la période de *n* années s'écrit (formule (12) dans le texte) :

$$\begin{aligned} P_{n+1} &= P_1 * \frac{(2 + x_1)}{(2 - x_1)} * \frac{(2 + x_2)}{(2 - x_2)} * \dots * \frac{(2 + x_n)}{(2 - x_n)} \\ &= P_1 * \prod_i \left(\frac{2 + x_i}{2 - x_i}\right). \end{aligned}$$

L'application de la définition de la moyenne implique de remplacer les taux observés par leur moyenne tout en conservant la variation de la population sur l'ensemble de la période (formule (13)) :

$$P_{n+1} = P_1 * \prod_i \left(\frac{2 + x_i}{2 - x_i}\right) = P_1 * \prod_i \left(\frac{2 + \bar{x}}{2 - \bar{x}}\right) = P_1 * \left(\frac{2 + \bar{x}}{2 - \bar{x}}\right)^n.$$

On peut déduire de cette équation, une égalité qui permettra, en isolant la moyenne, d'en déduire la formule au départ des taux observés (formule(13)) :

$$\prod_i \left(\frac{2 + x_i}{2 - x_i}\right) = \left(\frac{2 + \bar{x}}{2 - \bar{x}}\right)^n \Rightarrow \sqrt[n]{\prod_i \left(\frac{2 + x_i}{2 - x_i}\right)} = \left(\frac{2 + \bar{x}}{2 - \bar{x}}\right) \Rightarrow (2 - \bar{x}) * \sqrt[n]{\prod_i \left(\frac{2 + x_i}{2 - x_i}\right)} = 2 + \bar{x}$$

$$\Rightarrow 2 * \sqrt[n]{\prod_i \left(\frac{2 + x_i}{2 - x_i}\right)} - 2 = \bar{x} + \bar{x} * \sqrt[n]{\prod_i \left(\frac{2 + x_i}{2 - x_i}\right)}$$

$$\Rightarrow \bar{x} = 2 * \frac{\sqrt[n]{\prod_i \left(\frac{2 + x_i}{2 - x_i}\right)} - 1}{\sqrt[n]{\prod_i \left(\frac{2 + x_i}{2 - x_i}\right)} + 1}.$$

Finalement, la formule directe (20)) est obtenue en remplaçant le produit sous la racine par le rapport P_{n+1}/P_1 :

$$\bar{x} = 2 * \frac{\sqrt[n]{P_{n+1}/P_1} - 1}{\sqrt[n]{P_{n+1}/P_1} + 1}, \quad avec \prod_i \left(\frac{2 + x_i}{2 - x_i}\right) = \frac{P_{n+1}}{P_1}.$$

• Annexe 2. Inadéquation le la formule habituelle du taux moyen sous hypothèse linéaire

Pour compléter l'exposé théorique dans le corps du texte, l'inadéquation de la formule (1) est illustrée ici par un exemple numérique. Le Tableau A.1 reprend la valeur des taux annuels selon les trois hypothèses pour les données du Tableau 3. Dans les trois dernières colonnes, figurent respectivement les rapports entre les valeurs selon les hypothèses géométrique et exponentielle ; linéaire et exponentielle et finalement linéaire et géométrique. La ligne « $\overline{x}$ » reprend les valeurs des moyennes selon les formules démontrées dans le texte et les rapports entre ces moyennes. La dernière ligne prend en compte la valeur de la moyenne issue de la formule habituelle pour l'hypothèse linéaire (soit 0,52921 (formule (1)).

Tableau A.1. taux de croissance et rapport des taux selon les trois hypothèses

i	Année	Taux annuel selon l'hypothèse…			Rapport		
		… expo.	… géo.	… linéaire	Géo./Expo.	Linéaire/Expo.	Linéaire/Géo.
1	2011	0,91629	1,50000	0,85714	163,70%	93,54%	57,14%
2	2012	0,95551	1,60000	0,88889	167,45%	93,03%	55,56%
3	2013	0,29152	0,33846	0,28947	116,10%	99,30%	85,53%
$\overline{x}$		0,72111	1,05671	0,69140	146,54%	95,88%	65,43%
Rapport avec la moyenne selon la formule pour l'hypothèse linéaire (0,52921)					-	73,39%	50,08%

Conclusion : les rapports de l'avant-dernière ligne sont en cohérence avec les taux annuels. Par exemple, dans le cas de la comparaison entre les hypothèses linéaire et exponentielle, le rapport entre moyennes valant 95,88% est compatible avec les valeurs obtenues pour les rapports annuels (entre 93% et un peu plus de 99%). À l'inverse, le pourcentage dans la dernière ligne du tableau est incohérent : 73,39% est extérieur aux pourcentages pour les taux annuels. La situation est identique en comparant les hypothèses linéaire et géométrique. Ce constat montre que, dans le cadre de l'hypothèse linéaire, la formule habituelle de la moyenne (la formule (1)) n'est pas en cohérence avec la formule du taux annuel (formule(6)). Or, cette dernière formule n'est pas discutable : l'hypothèse de croissance linéaire oblige à considérer que le dénominateur du taux annuel x_i doit être égal à la moyenne entre les populations P_i et P_{i+1}. Dès lors, la formule habituelle doit être abandonnée et remplacée par (20).