

Pour les statisticiens, avoir la moyenne ne suffit pas

Publié le 12 juillet 2023 sur le [blog de l'Insee](#)
Temps de lecture : 8 minutes

Jean-William Angel, Insee.



FRANÇAIS	12.5
MATHÉMATIQUES	9,2
TECHNOLOGIE	17
ANGLAIS	11.5
HISTOIRE	10

La moyenne est facile à calculer et à comprendre. Mais elle résume souvent mal l'information qu'elle doit représenter. C'est le cas par exemple pour les salaires : les plus hauts salaires augmentent la moyenne sans que la situation de la majorité des salariés en soit changée. C'est aussi le cas quand les situations entre les individus sont très différentes : un indice moyen des prix peut rendre mal compte de la situation des personnes qui dépensent beaucoup dans les produits qui augmentent le plus. Pour remédier à ces difficultés, les statisticiens ont plusieurs outils. La médiane, qui n'est pas sensible aux valeurs extrêmes, vient se substituer à la moyenne. Les indicateurs de dispersion traduisent les écarts entre les moins et les mieux lotis. Et, puisqu'un indicateur ne peut convenir à tous, ils développent les cas types pour mieux cerner la réalité de chacun. Par exemple sur le chômage ou l'espérance de vie.

Dès l'école primaire, nous avons été habitués à nous comparer à la moyenne. Le calcul est simple et facile à comprendre : on additionne les notes et on divise par le nombre d'élèves. De par ces vertus, la moyenne est un indicateur statistique répandu bien au-delà de la classe : taille moyenne, salaire moyen, nombre de kilos de chocolats mangés par les Français en moyenne dans l'année, etc. On la trouve aussi lorsqu'elle n'est pas explicite. Derrière le nombre d'habitants au kilomètre carré, autrement dit la densité de population, se cache une moyenne qui ne dit pas son nom : on compte le nombre total d'habitants d'une zone, par exemple un département, et on le divise par la surface du département exprimée en kilomètres carrés. Et c'est là que les ennuis de la moyenne commencent. Car l'illusion d'uniformité que celle-ci suggère – la densité serait la même partout dans le département – est rarement pour ne pas dire jamais vérifiée. Il y aura forcément des zones denses et des zones peu peuplées. De même que dans une classe, il est rare que tout le monde ait la même note.

Résumer l'information

Le métier de statisticien consiste à rendre compte d'une réalité sociale ou économique au moyen d'indicateurs qui vont résumer l'information. Qui dit résumé dit perte d'information. Un seul indicateur ne suffira donc pas. Il en faudra plusieurs pour décrire la société avec assez de précision et de nuance et prendre le temps du commentaire.

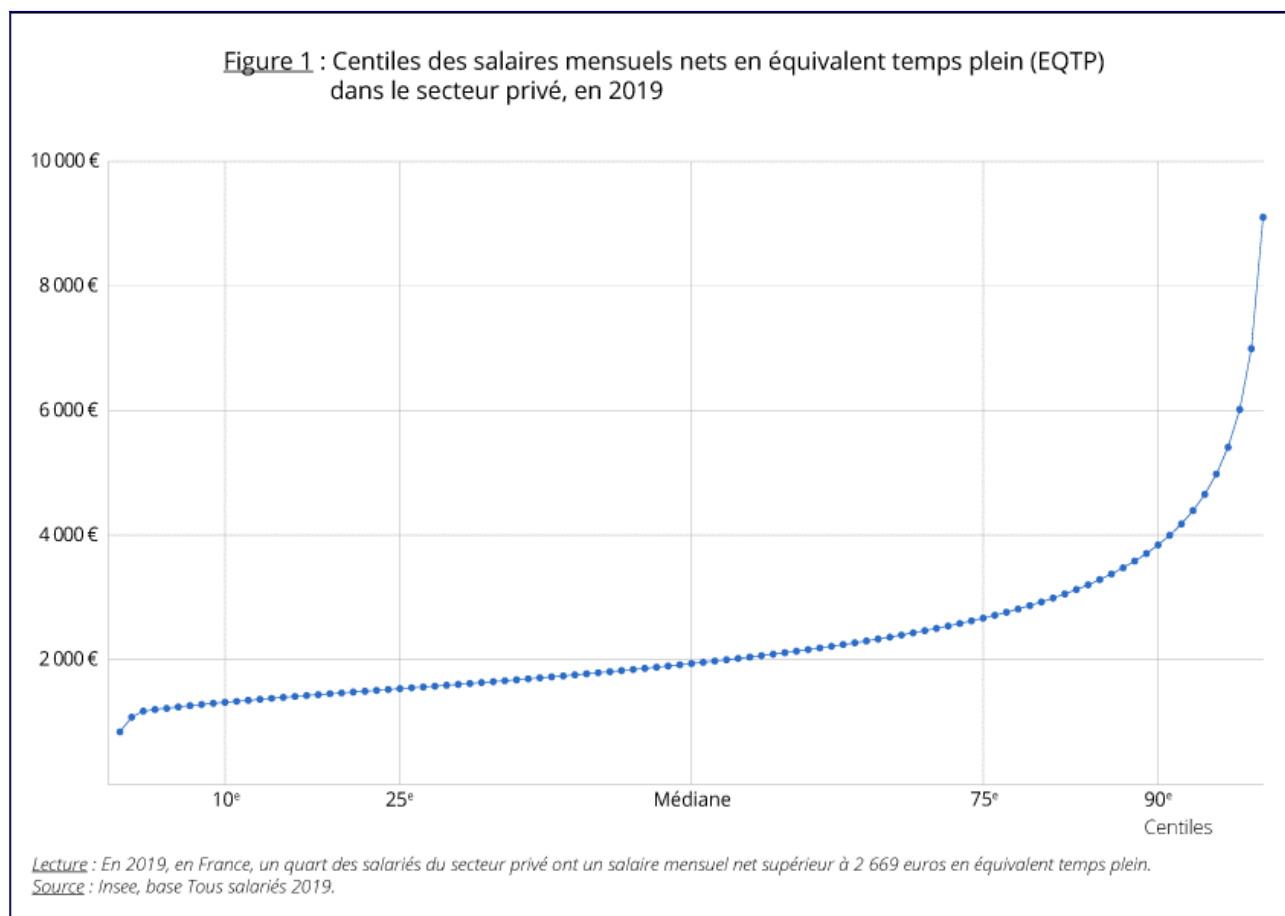
Hélas, le temps de l'information est de plus en plus court à mesure que nous en recevons de plus en plus. Un chiffre semble parfois le maximum que le public soit en capacité de recevoir, « pour ne pas le noyer ». La moyenne est alors le candidat idéal. Les statisticiens seraient ainsi confrontés à un choix pédagogique douloureux : produire des moyennes pour faire passer leur message, au risque de le compromettre, ou rendre compte de la complexité et ne pas être repris. En réalité, les statisticiens produisent beaucoup d'indicateurs pour permettre à ceux qui le souhaitent de faire pièce à la dictature de la moyenne. Petit parcours dans la forêt qui se cache derrière la moyenne.

Salaire et pauvreté, par décile et territoire

S'il est un domaine où la moyenne peut être très chahutée par les extrêmes, c'est bien celui des revenus. L'exemple du petit village dans lequel vient s'installer un milliardaire est vite compris : le revenu moyen va faire un bond puisqu'on va ajouter à tous les revenus des villageois celui du milliardaire. Mais chaque villageois n'en sera pas plus riche pour autant. Parmi les statistiques dont dispose le statisticien, la médiane a le grand avantage d'être insensible aux extrêmes. Il suffit de classer les individus du moins au plus, riche, grand, diplômé, etc. et de pointer la valeur, X, qui sépare la population en deux. Pour l'exemple des revenus, une moitié gagnera moins que X euros, et l'autre moitié gagnera plus. Dans le cas du milliardaire, ce n'est pas son revenu qui intervient dans le calcul mais le fait qu'il est un individu de plus dans ce que les statisticiens appellent la distribution, c'est-à-dire le classement des individus selon la variable concernée. Aussi, [l'Insee](#) comme la Dares, le service statistique du ministère de l'emploi, publient-ils les revenus et salaires médians en plus des moyennes.

La médiane partage donc une distribution en deux parties. Il suffit d'augmenter le nombre de parties, de même taille, pour affiner l'analyse. Quatre découpages sont couramment utilisés : les quartiles, qui découpent en quatre, les quintiles, en cinq, les déciles, en dix, et les centiles, en cent. Le premier quartile comprend donc les 25 % des individus qui gagnent le moins, pour rester sur les salaires. Le 2e quartile, les 25 % suivants. Idem pour les quintiles (tranches de 20 %), les déciles (10 %) et les centiles (1 %) (*figure 1*). Dans le cas des revenus ou des patrimoines, on peut même aller au-delà des centiles. On en trouvera une illustration dans la [datavisualisation des salaires](#) proposée par l'Insee.

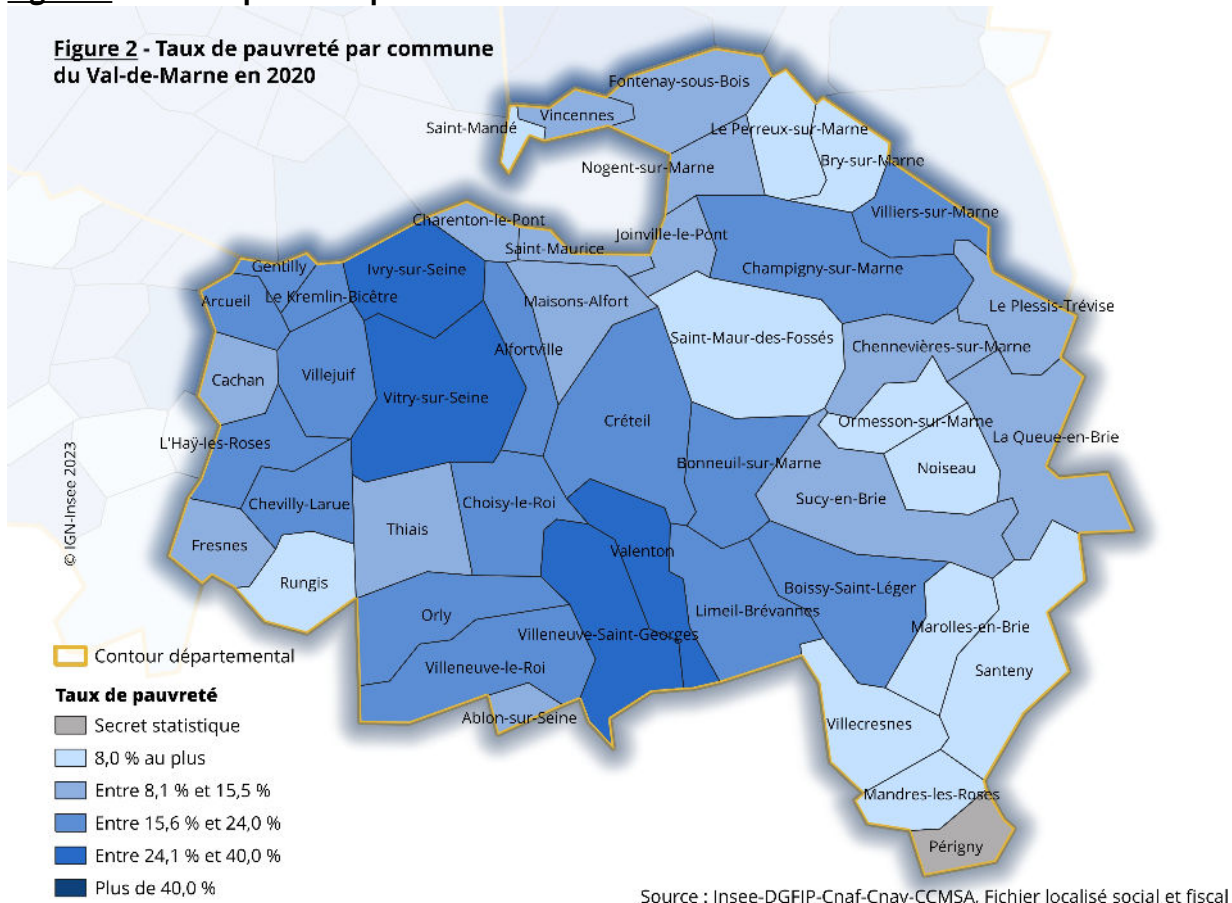
Figure 1 : Centiles des salaires mensuels nets en équivalent temps plein (EQTP) dans le secteur privé, en 2019



Lecture : En 2019, en France, un quart des salariés du secteur privé ont un salaire mensuel net supérieur à 2 669 euros en équivalent temps plein.
Source : Insee, base Tous salariés 2019.

Lorsqu'on considère les ménages, on ne raisonne plus en salaire mais en niveau de vie. Celui-ci est calculé à partir de l'ensemble des revenus : aux salaires s'ajoutent le cas échéant d'autres revenus tels que les indemnités de chômage, les retraites et pensions, d'autres prestations sociales et les revenus du patrimoine. Les niveaux de vie sont diffusés par décile et à des niveaux géographiques très fins, comme les communes ou les bassins de vie. On dispose ainsi d'une connaissance précise de leur distribution qui permet de déterminer les taux de pauvreté, y compris par tranche d'âge. Une richesse de données qui va bien au-delà du « taux moyen de pauvreté », 14,6 % en 2020, qui fait la Une des médias (figure 2).

Figure 2 : Taux de pauvreté par commune du Val-de-Marne en 2020



Source : Insee-DGFIP-Cnaf-Cnav-CCMSA, Fichier localisé social et fiscal.

Des indices des prix personnalisés

L'indice des prix illustre différemment les écarts de la moyenne. Cet indicateur est dit synthétique car il rend compte d'une grande variété de produits et de services. Chacun de ces produits et services est affecté d'un poids qui transcrit sa part dans le budget des ménages. La plupart des critiques portent sur cette pondération. Car tout le monde ne consomme pas les mêmes choses : les propriétaires ne supportent pas de loyer, les non-fumeurs ne subissent pas les augmentations de prix sur les cigarettes, les végétariens ne consomment pas de viande mais plus de légumes, ceux qui utilisent leur voiture pour aller travailler dépensent davantage en carburant que ceux qui utilisent les transports en commun, etc. Sans compter que les ménages les plus aisés consacrent une part plus faible de leur revenu à leur consommation. Disposer d'un indice synthétique est donc utile mais ne suffit pas à traduire la grande diversité des cas de figure. Pour y remédier, l'Insee calcule depuis longtemps [plusieurs indices des prix](#).

Chaque mois, l'institut publie l'indice des prix à la consommation pour les ménages urbains dont le "chef" est ouvrier ou employé. Il est détaillé en quelque [170 indices](#) portant sur l'alimentaire, les boissons, les loyers, les meubles, le gaz et l'électricité, etc.

Un second indice mensuel est [spécifique aux bas revenus](#) : celui des ménages du premier quintile (20 %) des niveaux de vie.

En outre, l'Insee calcule des indices annuels par :

- décile de niveau de vie ;
- tranche d'âge ;
- catégorie socio-professionnelle de la personne de référence du ménage ;

– type de ménage (personne seule, famille monoparentale, couple, sans ou avec enfants) ;
– et distingue les propriétaires de leur logement des accédants à la propriété et des locataires.
Enfin, l'indice des prix d'ensemble est complété par celui des [produits de grande consommation vendus dans la grande distribution](#). Il reflète au plus près les achats du quotidien. Son calcul est notamment alimenté par les données de caisse (cf. [Dupont, 2023](#)). Et pour encore plus de sur mesure, le [simulateur d'inflation](#) personnalisé permet à chacun de spécifier la répartition de ses dépenses en 20 postes.

Chômage, espérance de vie : cerner au mieux la réalité

L'Insee mobilise aussi toutes ces données fines dans ses analyses. En témoigne cet [éclairage de la note de conjoncture de mars 2023](#), qui estime les écarts d'inflation selon différentes catégories de ménages, jeunes ou âgés, habitant dans des petites communes ou de grandes agglomérations, modestes ou aisés.

Pour le chômage aussi, l'institut s'applique à retracer au mieux la variété des situations concrètes sur le marché du travail. Le chômage est ainsi calculé [chaque trimestre](#) pour les femmes et les hommes et pour trois tranches d'âge (15-24 ans, 25-49 et 50-64). [Chaque année](#), il est mesuré selon la catégorie socio-professionnelle (en sept modalités), le diplôme et la durée au chômage, croisés par le sexe et l'âge. Toutes variables qui donnent lieu à autant de taux de chômage, qui suivent la définition du Bureau international du travail (BIT), utiles aux décideurs publics pour mieux cibler les populations les plus touchées.

L'Insee mesure également chaque trimestre le chômage aux niveaux régional et départemental, ainsi que pour les 302 [zones d'emploi](#).

Concluons cette courte revue par l'espérance de vie. Sa moyenne, qui distingue les femmes et les hommes, est souvent mise en avant pour faire état des progrès de la science, comparer les systèmes de santé des pays ou alimenter les débats sur le temps consacré au travail. L'Insee propose depuis peu [un outil](#) qui fournit très simplement les espérances de vie par âge, par département, mais aussi et surtout par niveau de vie, diplôme ou catégorie sociale. Un bon moyen de vérifier l'adage que l'on doit sûrement au premier statisticien : « Derrière cette moyenne, se cache une réalité contrastée. »

*

* *

Les quelques exemples qui ont jalonné ce parcours pourraient être multipliés. Les statistiques portant sur les entreprises sont elles aussi très détaillées, notamment par secteur et par zone géographique. Sur tous les domaines qu'ils couvrent, les services statistiques ministériels produisent eux aussi des indicateurs très détaillés, en plus des moyennes.

Celles-ci restent certes utiles pour livrer une première approche d'une réalité sociale ou économique. Mais il faudrait apprendre très tôt aux élèves, et aux parents, que la médiane rend mieux compte des « distributions ». Et qu'il n'est pas inutile d'enseigner à tous l'art de la dispersion. Statistique.

Pour en savoir plus

- Dupont F., 2023, « [Quels types de sources l'Insee utilise-t-il pour construire ses statistiques ?](#) », *Blog de l'Insee*, mai
- Insee, 2023, « [Au premier trimestre 2023, l'emploi salarié augmente dans quatre régions sur cinq et il est quasi stable dans les autres – Emploi et taux de chômage localisés \(par région et département\) – premier trimestre 2023](#) », *Informations rapides*, juillet.
- Insee, 2023, « [Indices des prix à la consommation – Résultats par regroupement de produits et produits détaillés \(Coicop\)](#) », *Séries chronologiques*, juillet.
- Insee, 2023, « [Au premier trimestre 2023, le taux de chômage est stable à 7,1 %](#) », *Informations rapides* n° 124, mai

- Insee, 2023, « [En avril 2023, les prix des produits de grande consommation vendus dans la grande distribution augmentent de 15,1 % sur un an – Indice des prix dans la grande distribution – avril 2023](#) », *Informations rapides* n° 121, mai
- Insee, 2023, « [Principaux résultats sur les revenus et la pauvreté des ménages en 2020 – Principaux indicateurs sur les revenus et la pauvreté aux niveaux national et local en 2020 – Dispositif Fichier localisé social et fiscal \(Filosofi\)](#) », *Insee Résultats*, avril
- Insee, 2023, « [La croissance résiste, l'inflation aussi](#) », *Note de conjoncture*, mars
- Insee, 2023, « [Principaux résultats sur les revenus et la pauvreté des ménages en 2020 – Niveau de vie et pauvreté des territoires en 2020](#) », *Insee Résultats*, janvier
- Insee, 2023, « [Les espérances de vie](#) », *Outils interactifs*.
- Insee, 2023, « [Les salaires](#) », *Outils interactifs*.
- Insee, 2022, « [Activité, emploi et chômage en 2021 et en séries longues – Chômage et halo autour du chômage](#) », *Insee résultats*, juillet