

MESURE DE DISPERSION DES TAUX :

A- STATISTIQUE D'ORDRE ET ROBUSTESSE

La statistique des MIR indique un taux moyen pondéré qui est calculé à partir des observations de l'échantillon. Comme cette caractéristique est « centrale », elle ne donne pas d'information sur la dispersion des taux. Il convient donc en plus de la moyenne, de donner une indication de cette dispersion. A cet égard, il existe plusieurs mesures de dispersions, chacune de ces mesures ayant des avantages et des inconvénients.

Traditionnellement, on peut donner *l'écart absolu* – c'est-à-dire l'écart entre l'observation la plus petite et l'observation la plus grande –, mais celui-ci peut comporter des valeurs aberrantes, c'est à dire des valeurs très éloignées de la « masse » des valeurs et donner une image trop « agrandie » de la dispersion. Souvent *la variance* – c'est-à-dire la moyenne du carré des écarts à la moyenne – est donnée, mais celle-ci est également très influencée par les valeurs aberrantes.

Lorsque les données sont des observations de base, il convient d'utiliser un indicateur de dispersion qui ne soit pas influencé par les valeurs extrêmes et qui donne une idée proche de la réalité et qui soit, de plus, relativement intuitive. Pour ce faire, on utilise des *statistiques d'ordre*, c'est à dire des statistiques basées sur le classement des données. On y trouve la médiane qui est la donnée qui se situe au milieu, c.-à-d. en ayant 50 p.c. des observations à sa droite (« plus grandes ») comme à sa gauche (« plus petites »). Elle peut être représentée par le tableau suivant où les indices indiquent le numéro d'ordre.

Q1				Médiane					Q3									
X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅	X ₁₆	X ₁₇	X ₁₈	X ₁₉

Cette statistique est *robuste* car elle n'est pas influencée par des données des extrémités quel que soit leur éloignement. Il existe toute une gamme de caractéristiques statistiques reposant sur l'ordre des données. Une mesure de *dispersion robuste* est l'intervalle interquartile (IQ), c'est-à-dire l'écart entre le premier quartile (Q1) et le troisième quartile (Q3). Le quartile est construit sur le même principe que la médiane, il permet de séparer 25 p.c. et 75 p.c. des données, il est insensible aux données extrémales. La généralisation de ce principe consiste à l'étendre à un niveau de précision qui est de l'ordre du pour-cent, on parle alors de percentile¹.

Ces mesures robustes peuvent être pondérées par une variable auxiliaire (par exemple des volumes pour des taux qui représentent la variable d'intérêt) : le classement se fait alors selon la variable d'intérêt mais le fractile est calculé par le cumul de la variable auxiliaire. On parle de quantiles pondérés dans ce cas-là.

Ces mesures robustes sont d'autant plus justifiées que les données sont échantillonnées.

B- DISPERSION DES DONNEES DANS LES MIR

Afin de respecter une condition de lisibilité et de robustesse, la statistique de dispersion retenue est celle de l'intervalle interquartile. Toutefois, cette statistique n'est publiée que si certains requis sont respectés tels que le respect de la confidentialité – et, partant, un minimum de déclarants pour l'élaboration de cette statistique –, un minimum d'activité et une certaine cohérence entre les rubriques pour lesquelles cette statistique est publiée. De plus, cette dispersion ne sera calculée que pour les nouveaux contrats.

Comme la statistique des MIR repose sur une moyenne pondérée par les nouveaux contrats, et que l'IQ repose sur les taux, il est tout à fait possible que des taux ne figurant pas dans l'IQ aient un poids important dans la moyenne pondérée; alors des quartiles pondérés sont calculés, avec indication dans le tableau.

¹ Ou, en terme très général de fractile.