

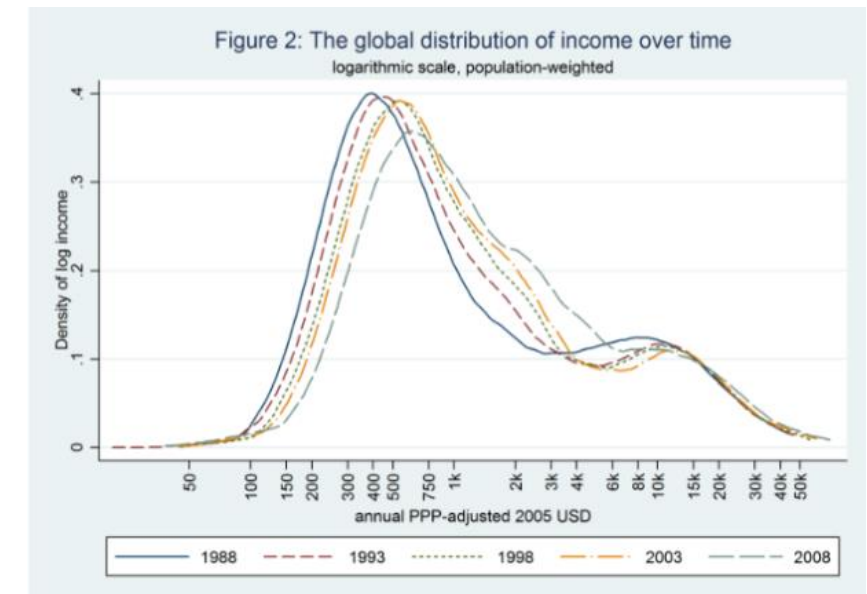
Description d'une variable quantitative.

Les indicateurs de tendance centrale

- **Mode** : valeur la plus fréquente d'une variable = valeur qui apparaît le plus souvent
 - non sensible aux valeurs extrêmes (*outliers*), mais sensible à l'amplitude des classes
 - peut être unique : distribution unimodale ou multiple : distribution bimodale (2 modes), trimodale (3 modes) ou plus généralement multimodale (plusieurs modes)
 - indicateur de population hétérogène → pour une distribution multimodale, la population peut-être divisée en plusieurs sous-groupes

⚠ peut également être calculé pour une variable **qualitative** (seul indicateur de tendance centrale qui peut être identifié pour ce type de variable)

Exemple



Source : Lakner C and Milanovic B, (2013) Global Income Distribution : From the Fall of the Berlin Wall to the Great Recession. Policy Research Working Paper; No. 6719. © World Bank, Washington, DC. <http://hdl.handle.net/10986/16935> License: [CC BY 3.0 IGO](#)

Description d'une variable quantitative.

Les indicateurs de tendance centrale

- **Moyenne** : somme de toutes les observations divisée par le nombre d'observations
 - Formule mathématique : $\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$
 - indicateur le plus simple pour résumer l'information fournie par un ensemble de données statistiques
 - très sensible aux valeurs extrêmes
 - représente mal les populations hétérogènes
- **Moyenne pondérée** : somme des observations pondérées par un coefficient, le tout divisé par la somme des coefficients
 - Formule mathématique : $\bar{X} = \frac{\sum_{i=1}^n x_i \times c_i}{\sum_{i=1}^n c_i}$
 - utile lorsqu'on travaille sur des unités de tailles différentes (ex. : unités géographiques)

Description d'une variable quantitative.

Les indicateurs de tendance centrale

- **Médiane** : valeur qui partage la distribution des valeurs, classées par ordre croissant, en deux classes d'effectifs égaux
 - ➔ 50 % des valeurs se situent sous la médiane et 50 % au-dessus
 - non sensible aux valeurs extrêmes ni à l'amplitude des classes

Exercice

On interroge au hasard 5 diplômés d'un master, et on obtient les revenus suivant, en euros mensuels :

1480, 1590, 2130, 1180, 9350

1. Calculer le revenu moyen

Revenu moyen = $(1480 + 1590 + 2130 + 1180 + 9350) / 5 = 15730 / 5 = 3146$

2. Calculer le revenu médian

On classe les valeurs par ordre croissant pour trouver la valeur centrale

1180, 1480, **1590**, 2130, 9350

3. Quel indicateur retenir...
 - a) ... pour convaincre vos parents de vous laisser faire ces études?
 - b) ... pour avoir une estimation raisonnable de vos revenus futurs?

Le revenu médian synthétise mieux la distribution

La médiane est moins sensible aux valeurs extrêmes (aux « queues » de distribution)

Description d'une variable quantitative.

Les mesures de dispersion

- Les **mesures de dispersion** principales sont les suivantes :
 - L'**étendue** (*range*) : différence entre la plus petite valeur et la plus grande (mesure sensible aux valeurs extrêmes)
 - L'**écart interquartile** (*interquartile range - IQR*)
 - La **variance**
 - L'**écart-type** (*standard deviation*)

Indicateur de tendance centrale	Mesure de dispersion
Médiane	Écart interquartiles
Moyenne	Écart-type

Description d'une variable quantitative.

Dispersion autour de la médiane

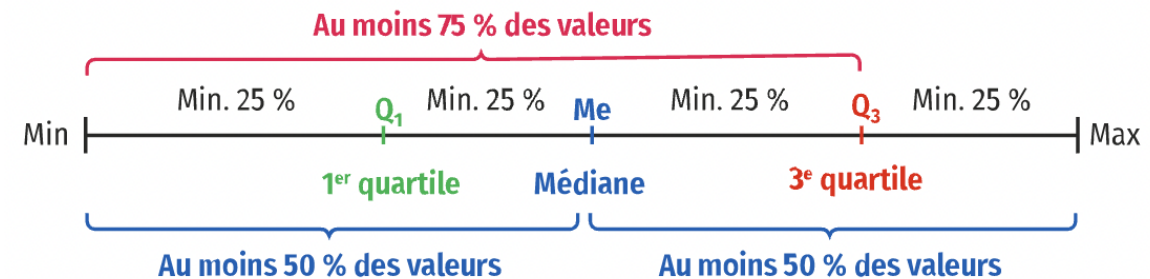
- **Quantiles** : valeurs qui coupent la distribution en groupes d'effectifs égaux

- quantiles d'ordre 4 = **quartiles**, notés **Q** : divisent la distribution en 4 groupes égaux

- Q1, le premier quartile, est la valeur au-dessous de laquelle se situent 25 % des valeurs
- Q2, le deuxième quartile, est la valeur au-dessous de laquelle se situent 50 % des valeurs = médiane
- Q3, le troisième quartile, est la valeur au-dessous de laquelle se situent 75 % des valeurs

/!\ Q4 n'existe pas à proprement parler, c'est la valeur maximale

- quantiles d'ordre 5 = **quintiles** : coupent la distribution en 5 groupes égaux
- quantiles d'ordre 10 = **déciles** (*deciles*), notés **D** : coupent la distribution en 10 groupes égaux
- quantiles d'ordre 100 = **centiles** (*percentiles*), notés **C** : coupent la distribution en 100 groupes égaux



Description d'une variable quantitative.

Dispersion autour de la médiane

- L'**écart** ou **intervalle interquartile** : différence entre Q3 et Q1 : $IQR = Q3 - Q1$
 - grande robustesse aux valeurs aberrantes
 - 50% des valeurs de la série sont comprises dans l'intervalle interquartile
- Le **rapport inter-quantile** (*inter-quantile ratio*) : division du quantile supérieur par le quantile inférieur
Exemple : pour les inégalités de revenus, on utilise le plus souvent les déciles, on calcule alors $D9/D1$

Description d'une variable quantitative.

Dispersion autour de la moyenne

- La moyenne ne dit rien quant à la dispersion de la distribution, autrement dit on ne sait rien des écarts de chaque individu à la moyenne
 - **Écart-type**
 - plus les valeurs sont dispersées autour de la moyenne, plus il est important (= moins la moyenne synthétise bien l'ensemble des observations)
⇔ plus il est faible, plus la population est homogène (si 0, cela signifie que toutes les observations sont égales)
 - **Variance**
 - il s'agit de l'écart moyen au carré par rapport à la moyenne
- !/ Comme les valeurs sont élevées au carré, la variance une unité différente (l'unité au carré), ce qui la rend difficilement interprétable, mais qui permet de raisonner en valeur absolue
- ➔ **il est conseillé de toujours utiliser l'écart-type pour décrire un échantillon, car cela facilite l'interprétation**

Description d'une variable quantitative.

Dispersion autour de la moyenne

- Notations mathématiques

- **Écart-type** = racine carrée des écarts à la moyenne

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2}$$

- **Variance** = distance de chaque individu statistique à la moyenne de la variable observée

$$V = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2$$

- **Propriétés de l'écart-type**

- Il est nécessairement positif
- Il est exprimé dans la même mesure que la variable correspondante (en année, en mètre, en points, etc.)
- Il est sensible aux valeurs extrêmes (*outliers*) et permet donc d'en identifier la présence

Description d'une variable qualitative.

Le tableau de fréquence (tri à plat)

- Le **tableau de fréquence** ou **tri à plat** : tableau de la répartition des répondants dans les différentes modalités de la variable
 - il permet **d'avoir une première idée des résultats** et constitue naturellement la base des rapports d'enquête

Exemple

Pour chaque **variable** qualitative, les **effectifs** pour chacune des **modalités** sont présentées

Fréquence de supp

supp	Effectifs	% du total	% cumulés
OJ	30	50.0%	50.0%
VC	30	50.0%	100.0%
Total	60	100.0%	