

Objectif du TD : réinvestissement des connaissances de statistiques, médiane, quartiles, écart interquartile, déciles, comparaison moyenne/médiane.

Mise en place des notions de diagrammes en boîte et d'écart type pour étudier la dispersion des séries statistiques.

**I- Caractéristiques de dispersion d'une série statistique :** étude d'un exemple

**Exemple : le niveau de vie des individus en France en 2009.**

Les données, tableaux et graphiques proviennent de l'étude de l'INSEE : « les revenus et patrimoines des ménages, éditions 2006 et 2012 ».

Les données collectées lors d'une enquête sont regroupées en vue d'une étude statistique pour être analysées et commentées.

Un graphique de base : le regroupement par classe

remarque : sur ce graphique de l'INSEE, les **effectifs** sont ramenés en pourcentage de l'**effectif total** ( c'est-à-dire du nombre total des individus ), en ordonnée sont donc représentées les **fréquences** de la série statistique. (Les données sont regroupées par ordre croissant du niveau de vie)

1) **médianes , quartiles , déciles**

Pour mettre en évidence les caractéristiques de dispersion de la série statistique, on détermine la **médiane**, les **quartiles** et les **déciles**.

Pour cela, on utilise les **effectifs cumulés** et la lecture de graphiques ou tableaux.

Compléter le tableau suivant en mettant en Evidence sur les deux graphiques, les valeurs trouvées.

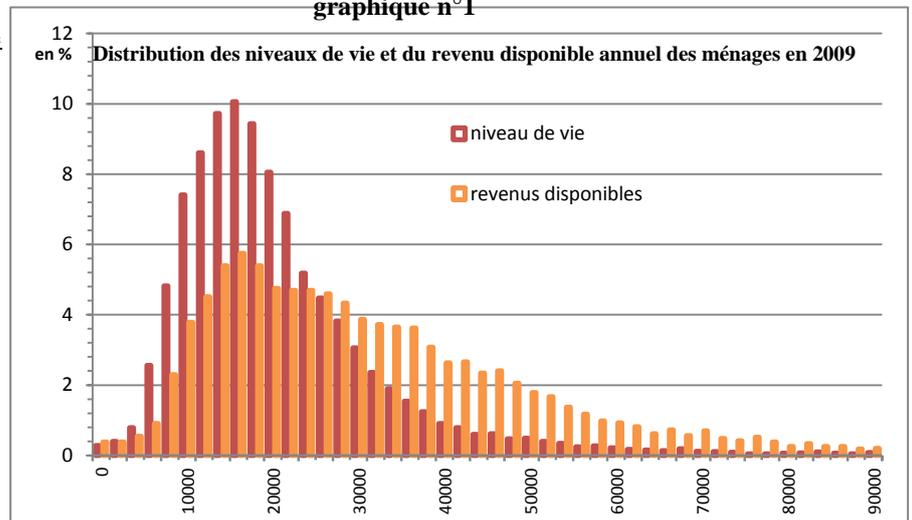
Premier décile D1	Premier quartile Q1	médiane	Troisième quartile Q3	Dernier décile D9

Attention : D1,Q1, Me, Q3 et D9 sont dans l'unité des valeurs du caractère de la série statistique.

Construire le diagramme en boîte appelé aussi « boîte à moustache » associé à cette série statistique



graphique n°1

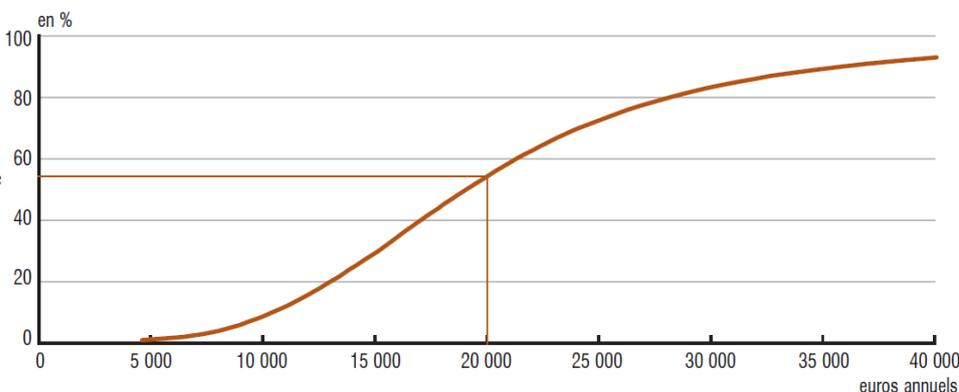


Source : d'après le graphique de la distribution des revenus et des niveaux de vie relatif à 2003 et le graphique des répartitions pour 2009, publiés par l'Insee.

Lecture : le pas de l'histogramme est de 2000 € ; la hauteur de la barre de coordonnées n en abscisse est donc égale à la proportion de revenus ou niveaux de vie compris entre n et n+2000 : ainsi 9,4% des individus ont un niveau de vie annuel en 2009 compris entre 18000€ et 20000€ et 5,4% des ménages ont un revenu disponible compris entre 18000€ et 20000€

graphique n°2

1. Répartition des niveaux de vie en 2009



Champ : France métropolitaine, personnes vivant dans un ménage dont le revenu déclaré au fisc est positif ou nul et dont la personne de référence n'est pas étudiante.  
Lecture : en 2009, 54 % des personnes ont un niveau de vie annuel inférieur à 20 000 euros.  
Sources : Insee ; DGFiP ; Cnaf ; Cnav ; CCMSA, enquête Revenus fiscaux et sociaux 2009.

En math, on demandera souvent la valeur de l'écart interquartile :  $Q3-Q1 =$   
Interprétation :

=

En SES, on demandera plutôt les rapports interdéciles : exemple  $\frac{D9}{D1} =$

Interprétation :

2) **Rappel : ne pas confondre moyenne et médiane !**

2. Quelques indicateurs sur le niveau de vie de 2001 à 2009

	montants annuels en euros constants 2009									
	2001	2002	2003	2004	2005	2006	2007	2008	2009	
Niveau de vie médian	17 410	17 880	17 830	17 740	18 030	18 300	18 690	19 000	19 080	
Niveau de vie moyen	20 230	20 730	20 610	20 590	20 820	21 340	21 690	22 130	22 140	

Que signifie le constat d'un revenu moyen supérieur au revenu médian ?

Comment expliquer cet écart ?

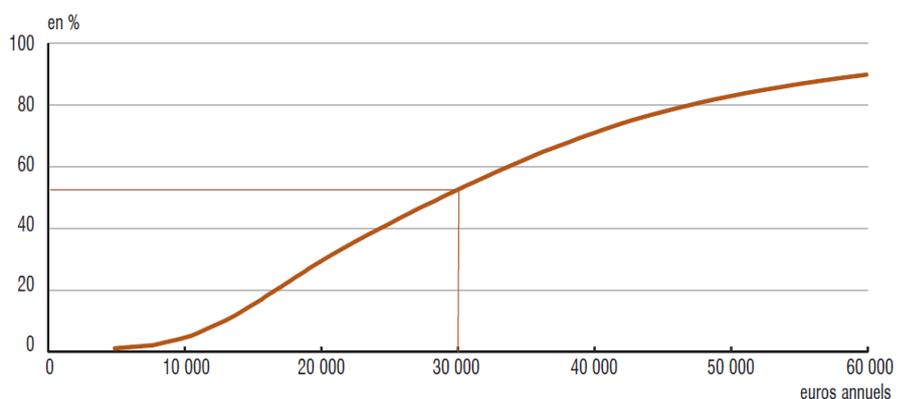
3) **Comparaison des écart-types et des diagrammes en boîte des deux séries : niveau de vie et revenu disponible des ménages**

La **moyenne** est souvent associée à l'**écart type** de la série.

Pour l'écart type de la série du niveau de vie des individus : 17000€ environ, pour l'écart type de la série des revenus des ménages : 31000€ environ. Comment peut-on interpréter cette différence sur les deux séries ?

Construire sur la page précédente, avec une couleur différente, le diagramme en boîte de la série du revenu disponible des ménages. Comment interpréter la différence des deux diagrammes en boîte ?

1. Répartition du revenu disponible des ménages en 2009



Champ : France métropolitaine, ménages dont le revenu déclaré au fisc est positif ou nul et dont la personne de référence n'est pas étudiante.  
Lecture : en 2009, 52 % des ménages ont un revenu disponible annuel inférieur à 30 000 euros.  
Sources : Insee ; DGFIP ; Cnaf ; Cnav ; CCMSA, enquête Revenus fiscaux et sociaux 2009.

**II- Définition de la variance et de l'écart-type d'une série statistique**

On considère la série statistique ci-contre :

Valeurs	$x_1$	$x_2$	...	$x_p$	Total
Effectifs	$n_1$	$n_2$	...	$n_p$	$N$

La moyenne de cette série est le réel  $\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{N} = \frac{1}{N} \sum_{i=1}^p n_i x_i$ .

**Définition** La **variance** d'une série statistique est le réel noté  $V$ , défini par :

$$V = \frac{1}{N} (n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2) \text{ ou } V = \frac{1}{N} \sum_{i=1}^p n_i(x_i - \bar{x})^2.$$

L'**écart-type** d'une série statistique est le réel noté  $\sigma$  et défini par  $\sigma = \sqrt{V}$ .

**Propriété** Si les valeurs de la série sont données avec leurs fréquences  $f_1, f_2, \dots, f_p$ , alors :

$$V = \sum_{i=1}^p f_i(x_i - \bar{x})^2 \text{ avec } \bar{x} = \sum_{i=1}^p f_i x_i.$$

**Propriété**

La variance est également donnée par la formule  $V = \frac{1}{N} \sum_{i=1}^p n_i x_i^2 - \bar{x}^2$ .

**Exemple :**

On considère la série ci-contre :

Valeurs	1	2	4	5
Effectifs	2	5	9	4

La moyenne étant  $\bar{x} = 3,4$  ; la variance est :

$$V = \frac{1}{20} (2 \times (1 - 3,4)^2 + 5 \times (2 - 3,4)^2 + 9 \times (4 - 3,4)^2 + 4 \times (5 - 3,4)^2) = \frac{34,8}{20} = 1,74.$$

Alors, l'écart-type est  $\sigma = \sqrt{1,74} \approx 1,32$  (au centième près).