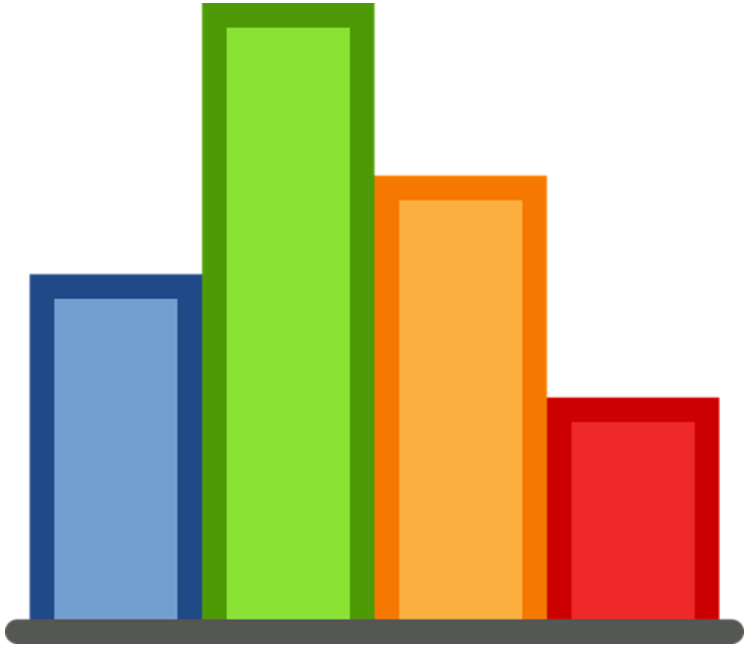


Statistique descriptive

Gérard Barmarin

Cours 2

VILLE DE WAVRE
IFOSUP
INSTITUT DE FORMATION SUPERIEURE

1 Définitions
2 Types de variables
3 Graphes

Aujourd’hui:

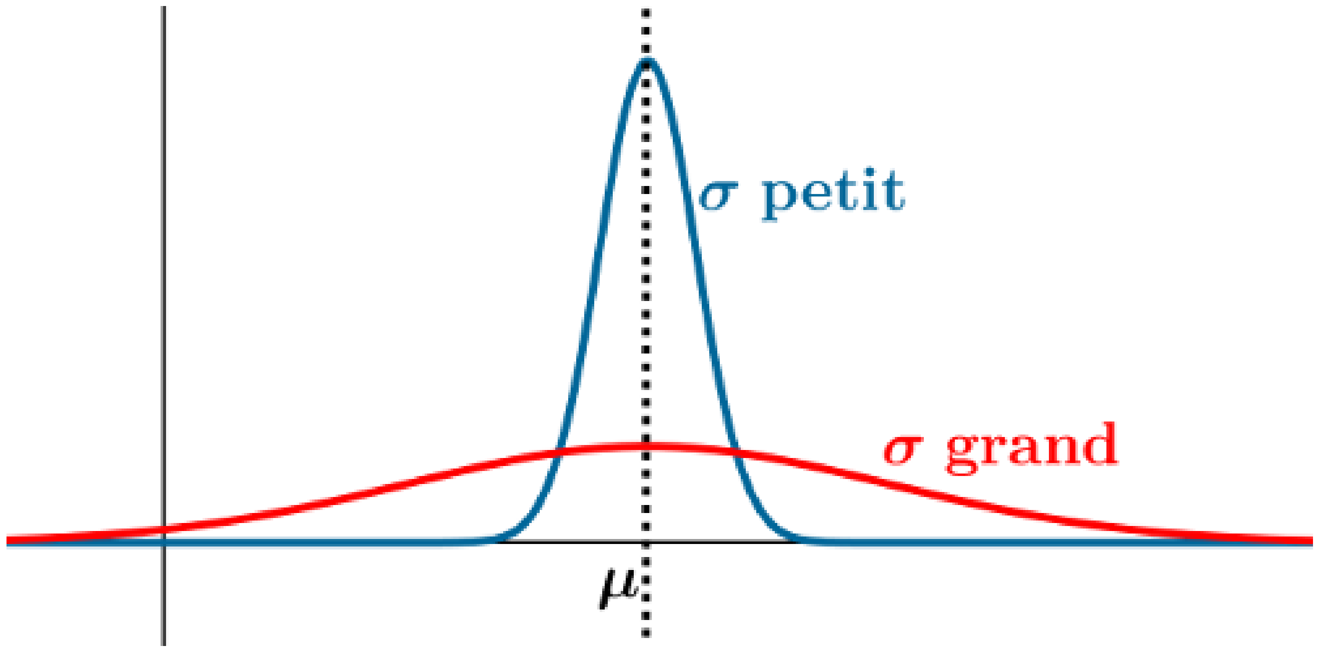
4 **Indicateurs statistiques**

Indicateurs statistiques

CARACTERISTIQUES DE DISPERSION

Définition de la dispersion

On appelle dispersion statistique, la tendance qu'ont les valeurs de la distribution d'un caractère à s'étaler, à se disperser, de part et d'autre d'une valeur centrale.



Indicateur numérique de la dispersion d’une distribution.

On souhaite qu’il soit nul si la dispersion est nulle, et qu’il soit d’autant plus grand que la dispersion est grande.

Les trois paramètres de dispersion absolue les plus courants sont l'étendue, l'intervalle interquartile, et l'écart type. Ces paramètres s’articulent autour de deux idées :

(1) la « dispersion » comme « étendue » de la distribution : ce seront l’**étendue** et l’**étendue interquartile**

(2) la « dispersion » comme « éloignement moyen entre les observations et le centre de la distribution » : ce seront l’**écart absolu moyen** et l’**écart-type**, qui se calcule en prenant la racine carrée de la **variance**

Indicateurs statistiques

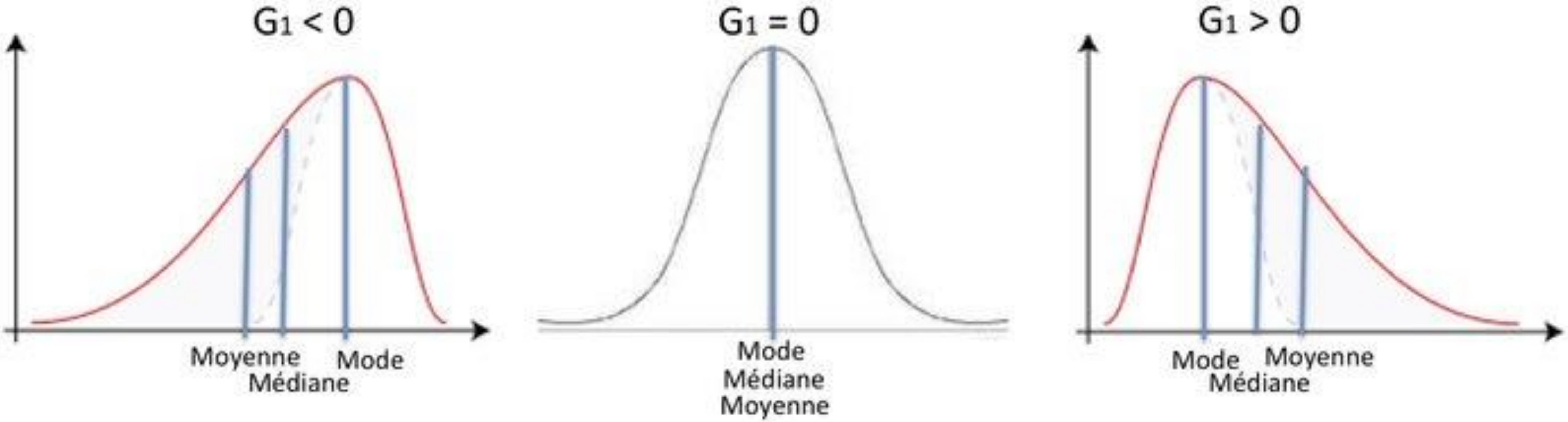
CARACTERISTIQUES DE CENTRALITE OU DE LOCALISATION

Ce sont des nombres « résumés de la collection d’observations » qui se résument à
« Globalement, où se situe la distribution sur l’axe des modalités ? »

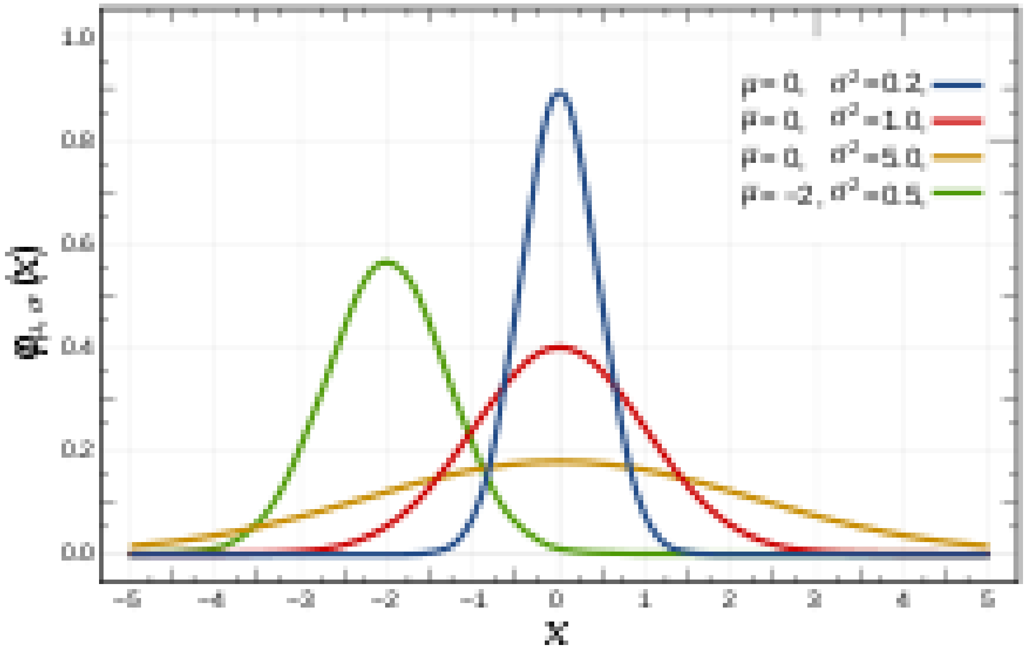
ou « Quel est le centre de cette distribution ? ».

Nous allons voir les trois caractéristiques de centralité (ou de tendance centrale) les plus utilisées, chacune ayant ses usages, ses avantages et ses inconvénients.

- **le mode** est un centre de concentration de la collection des observations
- **la médiane** est un centre de position de la collection des observations
- **la moyenne** arithmétique est un centre d’équilibre de la collection des observations



Asymétrie en j Courbe normale Asymétrie en i



Indicateurs statistiques

CARACTERISTIQUES DE DISPERSION

4.1.1 L’étendue (ou amplitude)

L’étendue (Range en Anglais) est l’écart entre la valeur de la plus grande observation et celle de la plus petite.

C’est une mesure particulièrement naïve de la dispersion d’une collection d’observations, car elle ne dépend que de deux d’entre elles (la plus petite et la plus grande, donc éventuellement des valeurs extrêmes !), et aucunement de la manière dont se distribuent les autres observations.

Exemple

La détermination de l’étendue est immédiate.

Variable observée : poids des individus d’une classe:

80, 52, 58, 80, 62, 89, 88, 60, 88, 62, 63, 87, 66, 65, 65, 65, 74, 72, 72, 70, 74, 74, 70, 80, 78, 75, 76, 75, 75, 75, 75, 75, 75, 75, 75

On met d’abord les valeurs dans l’ordre:

52, 58, 60 , 60, 62, 62, 63, 65, 65, 65, 66, 70, 70, 72, 72, 74, 74, 74, 75, 75, 75, 75, 75, 75, 75, 75, 76, 78, 80, 80, 80, 87, 88, 88, 89.

Plus petite observation (la première) : 52kg Plus grande observation (la dernière) : 89kg

Etendue : 89 – 52 = 37kg

Pourquoi est-ce un indicateur naïf?

Pourquoi est-ce un indicateur naïf?

Reprenons la distribution précédente

52, 58, 60 , 60, 62, 62, 63, 65, 65, 65, 66, 70, 70, 72, 72, 74, 74, 74, 75, 75, 75, 75, 75, 75, 75, 75, 76, 78, 80, 80, 80, 87, 88, 88, **89**.

Et comparons-la à celle-ci:

52, 52, 52, 52 , 52, 52, 52, 52, 52, 52, 52 , 52, 52, 52, 52 52, 52, 52 , 52, 52, 52, 52, 52, 52, 52 , 52, 52, 52, 52, 87, 87, 87, 87, 87, **89**.

Manifestement la répartition des poids n'est pas la même alors que l'étendue l'est

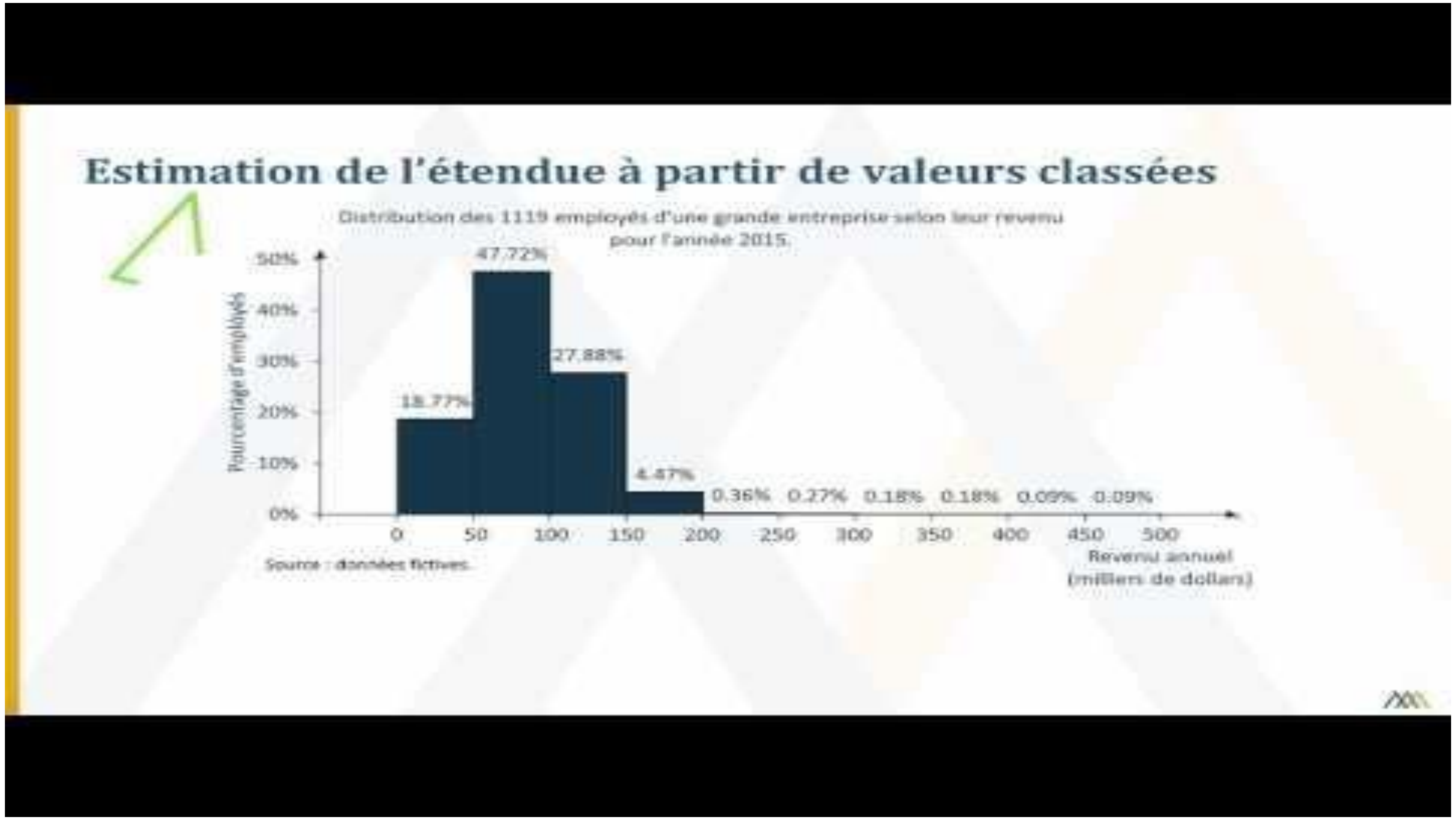
Le seul cas où l'on pourrait hésiter est celui où la distribution est donnée après un regroupement en classes et si l'on n'a pas accès à l'ensemble des données.
Dans ce cas, on fait comme si les individus d'une classe y étaient répartis uniformément :

on considère donc que la borne inférieure de la première classe est la plus petite observation, et que la borne supérieure de la dernière classe est la plus grande.

Valeur de la borne inférieure de la plus petite classe : 50kg
Valeur de la borne supérieure de la plus grande classe : 89kg

L'étendue est égale à la différence de ces deux valeurs : 89 – 50 = 39kg

classe (kg)	individus: poids en kg	Effectif
[50-54[	52	1
[55-59[	58	1
[60-64[	62 60 60 63 62	5
[65-69[	65 65 66 65	4
[70-74[	72 70 72 74 74 74 70	7
[75-79[	75 75 75 75 76 75 75 75 75 78	10
[80-84[	80 80 80	3
[85-89]	89 88 88 87	4



https://www.youtube.com/watch?v=7Re9VF2fRUU

4.1.2 L’étendue interquartile

L’étendue interquartile est l’écart entre le premier quartile et le troisième quartile.

En d’autres termes, c’est l’étendue de la « moitié centrale » de la distribution, l’étendue de la distribution dont on a exclu les 25% d’observations le plus à gauche, et les 25% d’observations le plus à droite. De cette manière, on pallie au plus gros défaut de l’étendue, la sensibilité aux extrêmes.
Il s’agit d’une mesure de dispersion absolue autour de la médiane (= second quartile).
L’intervalle interquartile relatif (IIQR) correspondant à l’étendue interquartile divisée par la médiane est une mesure de dispersion relative par rapport à la médiane (sans unité)

La détermination de l’étendue interquartile est simple : **déterminez le premier quartile Q1 et le troisième quartile Q3 et calculez la différence Q3 – Q1.**

Exemple

Reprenons notre exemple du poids des individus d'une classe :
Voici les 35 valeurs des observations dans l'ordre croissant (ne pas oublier de les mettre dans l'ordre !)

n°	poids (kg)		
1	52		
2	58		
3	60		
4	60		
5	62		
6	62		
7	63		
8	65		
9	65	1$^{\text{er}}$ quartile	
10	65		
11	66		
12	70		
13	70		
14	72		
15	72		
16	74		
17	74		
18	74	médiane	EIQ: 76-65 = 11 kg
19	75		
20	75		
21	75		
22	75		
23	75		
24	75		
25	75		
26	75		
27	76	3$^{\text{ème}}$ quartile	
28	78		
29	80		
30	80		
31	80		
32	87		
33	88		
34	88		
35	89		

Pour 35 valeurs, la médiane ou second quartile est la 18$^{\text{ème}}$ valeur. Le premier quartile est la 9$^{\text{ème}}$ (17 – 8) et le troisième est la 27$^{\text{ème}}$ valeur (19+8)

Dans notre exemple, on a donc:

1$^{\text{er}}$ quartile = 9$^{\text{ème}}$ valeur = 65 kg
2$^{\text{me}}$ quartile = 27$^{\text{ème}}$ valeur = 76 kg

Étendue interquartile (EIQ) = 76-65 = 11 kg

On peut aussi dire que 50% de l'échantillon se situe entre 65 (1$^{\text{er}}$ quartile) et 76 kg (3$^{\text{ème}}$ quartile)

Formules pour calculer plus finement les quartiles pour une variable discrète:

Q_0: valeur minimum de X Q_4: valeur maximum de X

Comment trouver ou couper pour Q_1, Q_2 et Q_3? (n est l'effectif total correspondant au nombre total d'observations)

Q_1: y = (n+3)/4
Q_2: y = (2n +2)/4
Q_3: y = (3n+1)/4

Si le chiffre y obtenu est un entier, Q_i = la y^{ème} valeur de X (soit X_y), prise dans l'ensemble des X rangés dans l'ordre

Sinon on calcule Q en interpolant comme ceci entre X_y et X_{y+1}
$Q1 = X_y*0,25 + X_{y+1}*0,75$
$Q2 = X_y*0,5 + X_{y+1}*0,5$
$Q3 = X_y*0,75 + X_{y+1}*0,25$

Exemple (35 valeurs, n = 35):

52, 58, 60 , 60, 62, 62, 63, 65, 65, **65**, **66**, 70, 70, 72, 72, 74, 74, 74, 75, 75, 75, 75, 75, 75, 75, 75, 76, 78, 80, 80, 80, 87, 88, 88, 89

Pour Q_1: y = (35+3)/4 = 9,5, on interpole entre la 9ème et la 10ème valeur: Q_1 = 65*0,25 +66*0,75 = 16,25 + 49,5 = 65,75
Pour Q_3: y = (3*35+1)/4 = 26,5, on interpole entre la 26ème et la 27ème valeur: Q_3 = 75*0,75 +76*0,25 = 75,25

4.3 L’écart absolu moyen

L’écart absolu moyen est la moyenne des écarts (**en grandeur absolue**) entre chaque observation et la moyenne des n observations.

On ne peut pas simplement calculer la moyenne des écarts des n observations à la moyenne, car celle-ci est toujours nulle par définition de la moyenne.

Exemple

Soient les 5 valeurs suivantes: 4, 6, 9, 10, 11

La moyenne vaut: $\frac{4+6+9+10+11}{5} = \frac{40}{5} = 8$

valeur	écart à la moyenne
4	4-8 = -4
6	6-8 = -2
9	9-8 = 1
10	10-8 = 2
11	11-8 = 3

la moyenne des valeurs **absolues** des écarts à la moyenne

$$\frac{4+2+1+2+3}{5} = \frac{12}{5} = 2{,}4$$

Moyenne des écarts: $\frac{-4-2+1+2+3}{5} = \frac{0}{5} = 0$

4.4 L’écart-type

L’écart-type est la racine carrée de la variance, et **la variance est la moyenne des carrés des écarts (en grandeur absolue) entre chaque observation et la moyenne des n observations**

La variance n’est rien de plus qu’une moyenne, non pas une moyenne des valeurs mais une moyenne des écarts au carré.

on lui préfère sa racine carrée appelée écart-type. La variance en elle-même est donc inutile, elle est juste un intermédiaire statistique au calcul de l’écart-type (en anglais Standard deviation).

Pour interpréter l’écart-type, on dit que grosso modo, les valeurs observées s’écartent de plus ou moins un écart type autour de la moyenne.

Prenons l'exemple d'une variable d'âge dont la moyenne vaudrait 40 ans et l'écart type 11ans.

Est-ce beaucoup ?

Tout est relatif, 11 est important par rapport à 40 car 11/40 = 27 %. Si on avait une dispersion de 11 ans sur une population d'âge moyen de 80 ans, ce serait tout à fait différent, on aurait 11/80 soit 14 % de variabilité.

Le fait de rendre relatif l'écart-type par rapport à la moyenne (= de diviser l'écart type par la moyenne) permet de définir **le coefficient de variation** : il exprime, en pourcentage, l'importance de la variabilité par rapport à la valeur centrale. Cet indice est très utile lorsque l'on a des populations à comparer dont la moyenne est différente. Le coefficient de variation est une mesure de la dispersion relative (sans unité) alors que l'écart type est une mesure de dispersion absolue autour de la moyenne et s'exprime dans l'unité de la variable observée

https://www.youtube.com/watch?v=CiFoBkipJQk

Toutefois, il y a un autre moyen très commode de calculer la variance : ce moyen est à peu près **deux fois plus rapide**, et il limite fortement les erreurs d'arrondis. Il consiste en ceci :

- élevez au carré chacune des n observations ;
- calculez la moyenne de ces n carrés ;
- finalement, retranchez de ce nombre le carré de la moyenne des n observations. Vous avez la variance !

Cette méthode de calcul de la variance, qui est toujours applicable, s'exprime en français en disant que :

« la variance est égale à la moyenne des carrés des observations, moins le carré de la moyenne des observations ».

Ce qui donne du sens à l'écart-type : le résultat de Bienaymé-Tchébychev

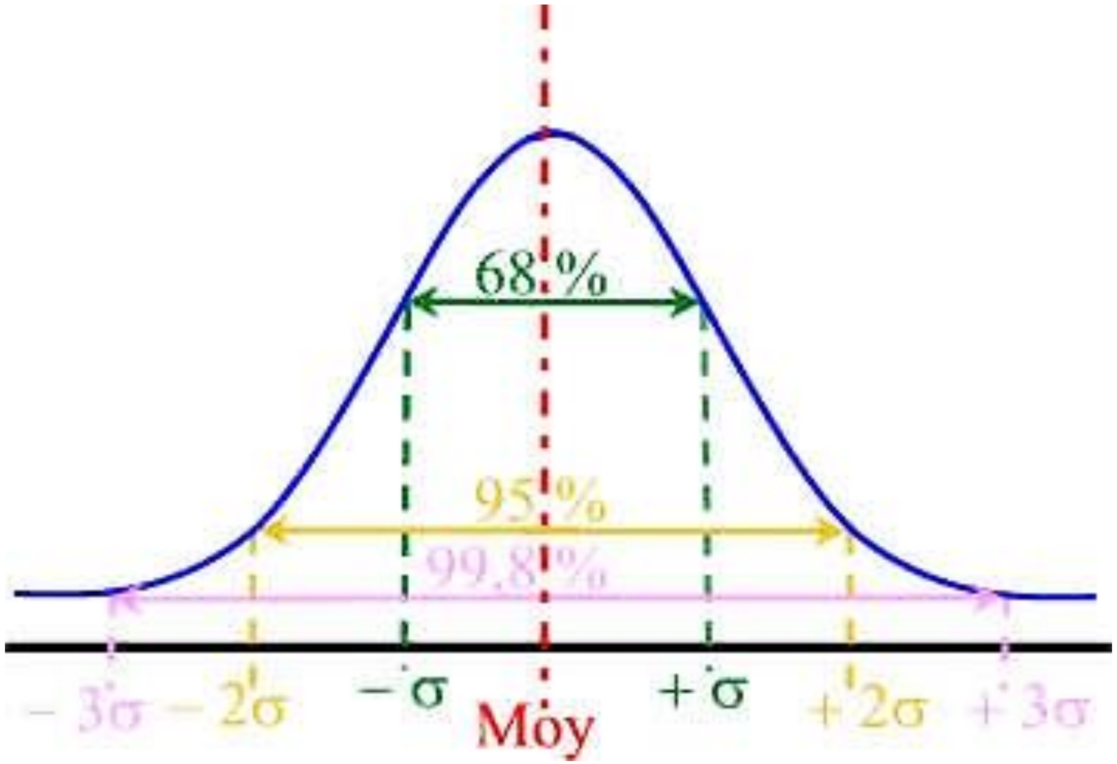
On démontre que **pour n'importe quelle collection d'observations**, la proportion des observations…

- qui sont à plus de 2 écarts-types de la moyenne (de chaque côté) est inférieure à ¼
- qui sont à plus de 3 écarts-types de la moyenne est inférieure à 1/9 (=11% donc 89% sont à moins de 3 σ)
- qui sont à plus de 4 écarts-types de la moyenne est inférieure à 1/16 (6,2%)
- qui sont à plus de 5 écarts-types de la moyenne est inférieure à 1/25 (4%)

et cetera.

En fait, la proportion des observations qui sont à plus de t écarts-types de la moyenne de chaque côté est inférieure ou égale à $1/t^2$, et cela est vrai pour toute collection d'observations et toute valeur de t.

Pour une loi normale, on peut aller beaucoup plus loin!

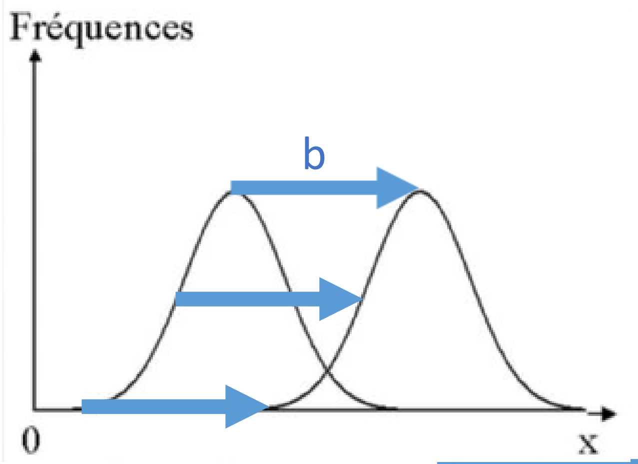


Comment une transformation linéaire de la variable observée X affecte-t-elle les caractéristiques de dispersion de la distribution sur X ?

L'idée à garder à l'esprit est la suivante : toutes les caractéristiques de dispersion se calculent à partir des écarts entre les observations, ou des écarts entre les observations et leur moyenne : il suffit donc de voir comment les transformations mentionnées affectent ces écarts !

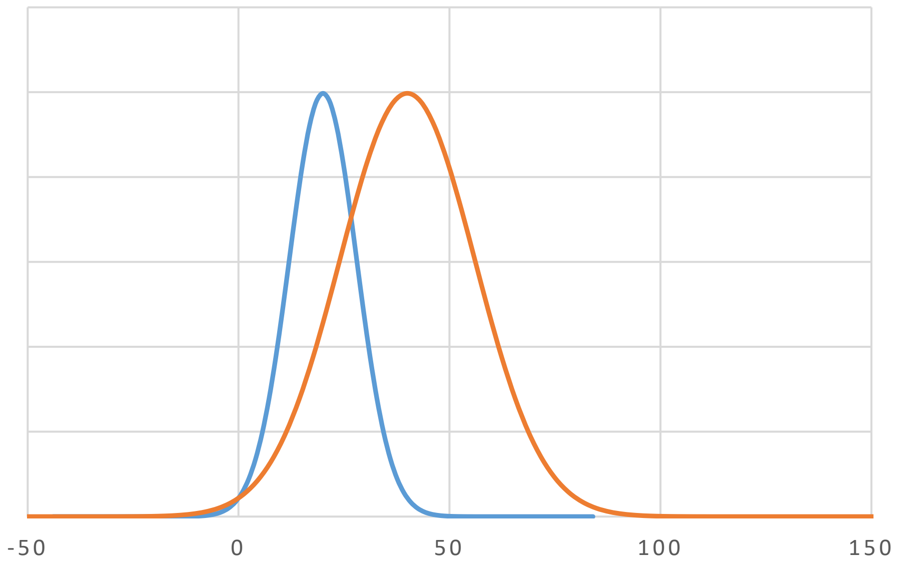
Et donc :

- **Variance (a.X + b) = a^2. Variance(X)**
- **Ecart type de (a.X + b) = a . Ecart type de X**
- **Étendue de (aX + b) = a . étendue de X**
- **Écart absolu moyen de (aX + b) = a . Écart absolu moyen de X**



b ne fait que déplacer l'ensemble des valeurs dans une direction, mais ne déforme pas la courbe, donc n'influence pas sur sa dispersion

Par contre la multiplication de X par a (ici 2) change bien l'étalement de la courbe!



5 CARACTERISTIQUES DE CENTRALITE OU DE LOCALISATION

Ce sont des nombres « résumés de la collection d'observations » à l'aide desquels on tente de répondre à ces deux questions :

« Globalement, où se situe la distribution sur l'axe des modalités ? » ou « Quel est le centre de cette distribution ? ».

Nous allons voir les trois caractéristiques de centralité (ou de tendance centrale) les plus utilisées, chacune ayant ses usages propres, avec ses avantages et ses inconvénients.

- **le mode** est un centre de concentration de la collection des observations
- **la médiane** est un centre de position de la collection des observations
- **la moyenne** arithmétique est un centre d'équilibre de la collection des observations

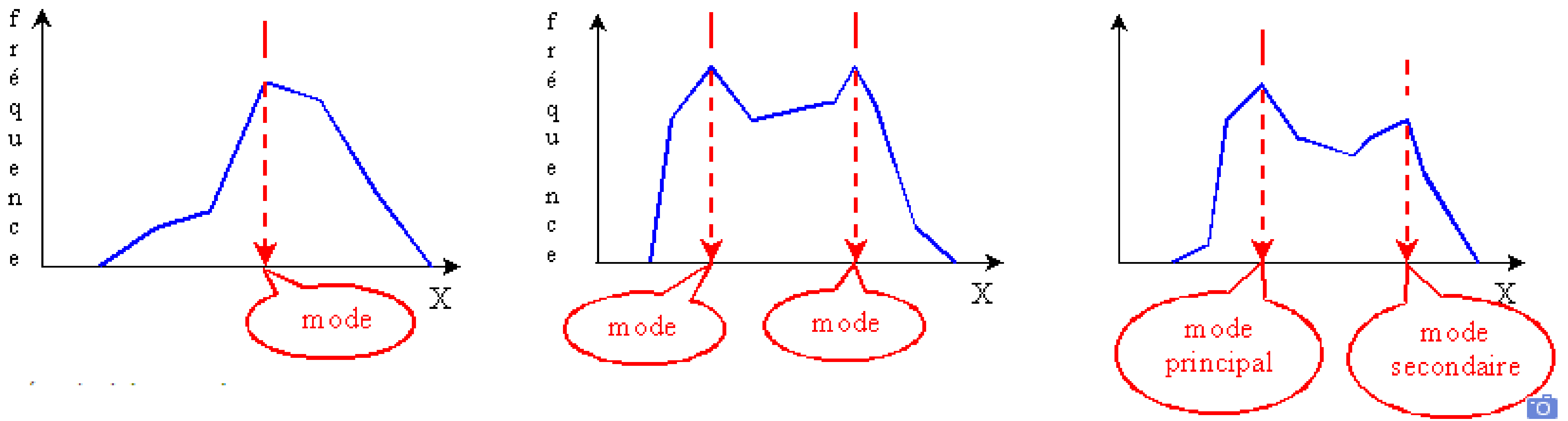
5.1 Le mode

Le mode (valeur dominante) ou la classe modale, est la valeur la plus représentée d'une variable, l’observation la plus observée ou la classe de plus forte densité.

Moyen mnémotechnique: c'est la valeur la plus « à la mode ».

En pratique : dans le cas d’une variable numérique où M est grand et où on est obligé de créer des classes, le mode est par convention le centre de la classe modale, et la classe modale est la classe de plus forte densité ;
dans tous les autres cas, le mode est la modalité (ou échelon ou valeur) qui a le plus grand effectif càd observée le plus souvent.

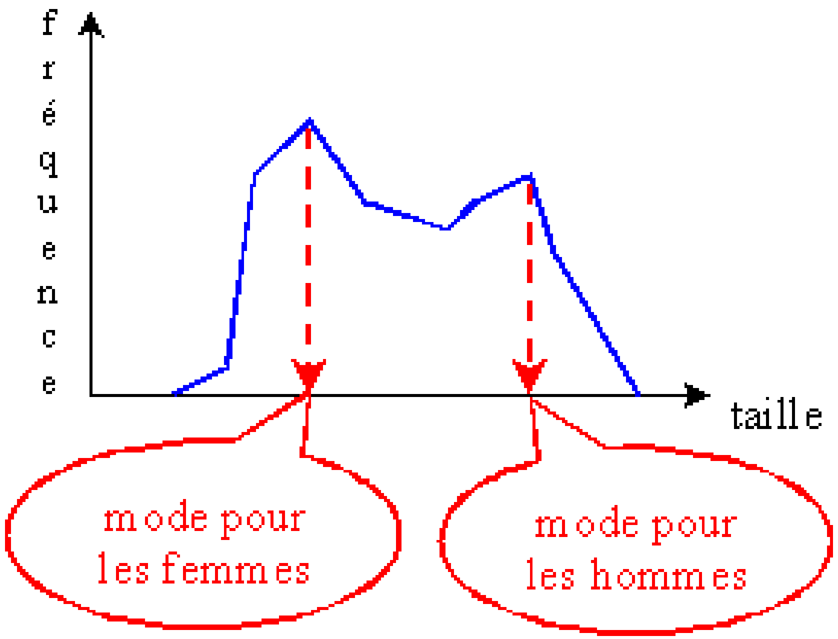
Une répartition peut être unimodale ou plurimodale (bimodale, trimodale…), si deux ou plusieurs valeurs de la variable considérée émergent également, voire sans aucun mode (distribution uniforme) si toutes les valeurs de la variable considérée sont également réparties.



Si une distribution est plurimodale,
cela pourrait-il avoir une signification?

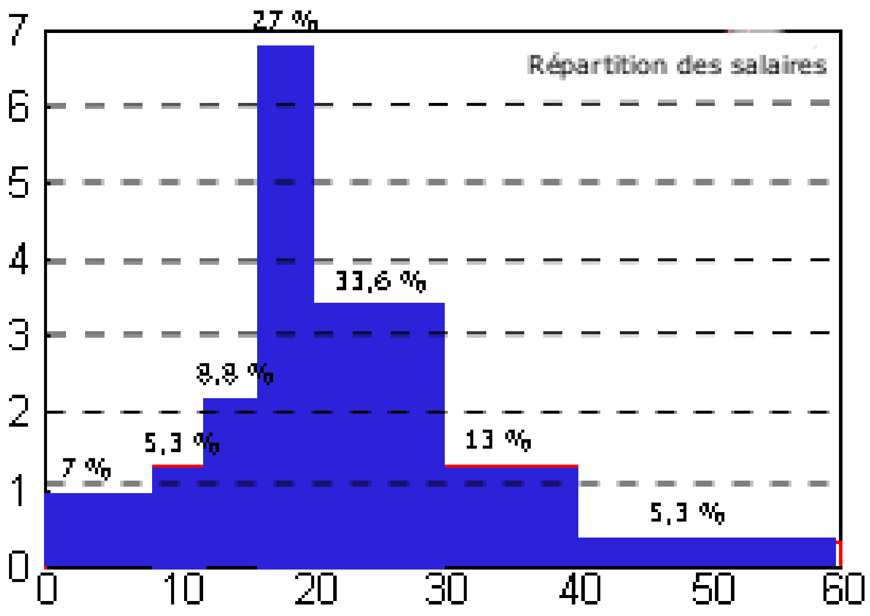
Si une distribution est plurimodale, cela pourrait-il avoir une signification?

Sans doute l'échantillon n'est-il pas homogène pour le caractère concerné…



Distribution de la taille des individus dans une population adulte

Exemple d'histogramme avec des classes de largeurs différentes :
L'observation du tableau d'observations laisserait penser que la classe modale serait la classe [20;30[puisqu'elle correspond au plus grand effectif. Mais une observation de l'histogramme corrige cette idée fausse :



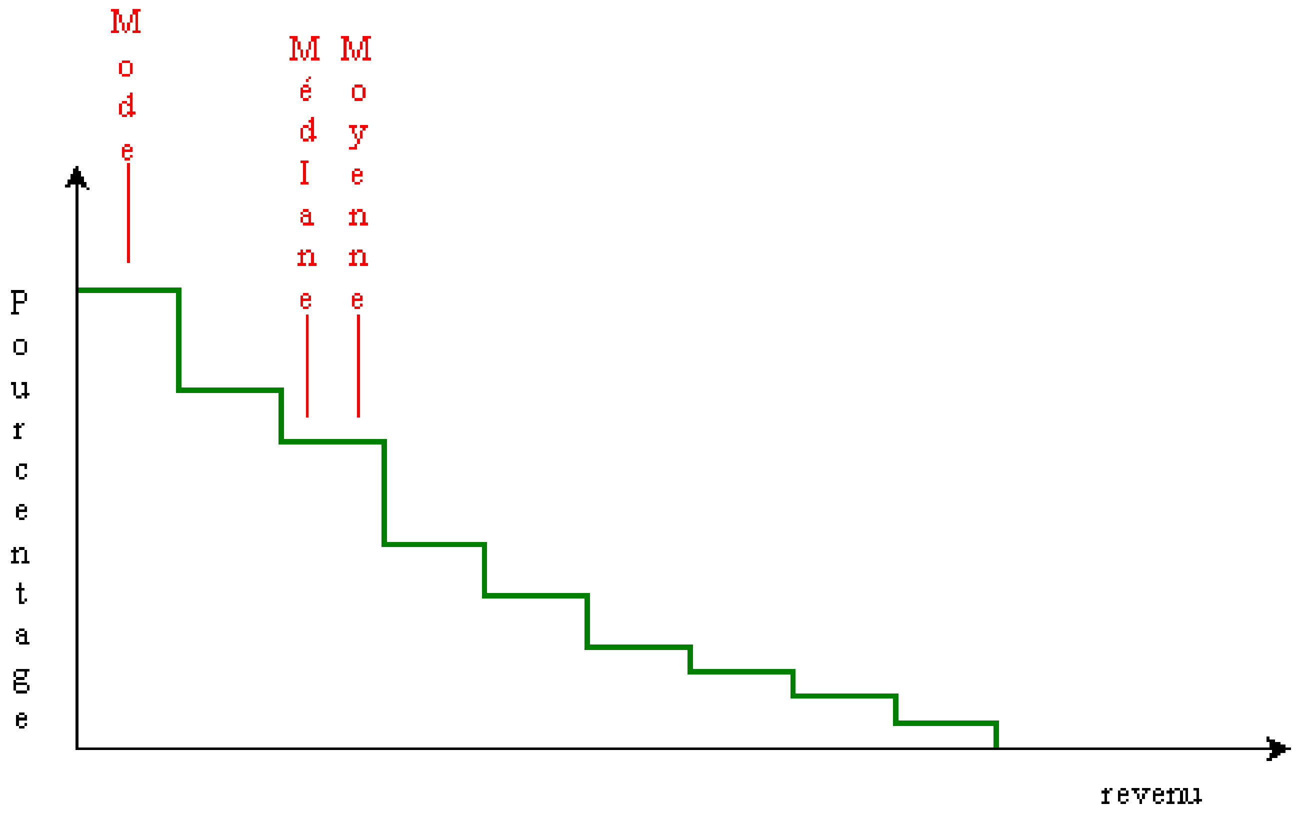
Répartition des revenus annuels en milliers d'Euros dans une population de 4370 personnes.								
Salaires	entre 0 (inclus) et 8 exclus	entre 8 (inclus) et 12 exclus	entre 12 (inclus) et 16 exclus	entre 16 (inclus) et 20 exclus	entre 20 (inclus) et 30 exclus	entre 30 (inclus) et 40 exclus	entre 40 (inclus) et 60 exclus	Total
Effectifs	306	231	385	1180	1468	568	232	4370

La classe modale est la classe [16; 20[dont la densité d'effectif est la plus élevée.
Faites donc très attention quand les classes ont des largeurs différentes !

Le mode est une mauvaise mesure du centre,
car la classe la mieux représentée n'est pas nécessairement au centre de la distribution.

Exemple :

Répartition des revenus dans une population.



5.2 MOYENNE (ou moyenne arithmétique)

Le statisticien est un homme qui prétend qu'avoir la tête dans une fournaise et les pieds dans la glace permet de bénéficier d'une température moyenne agréable.

La moyenne arithmétique est le centre d'«équilibre» de la collection d'observations.

C'est la valeur par rapport à laquelle les écarts des observations se compensent tous, c.-à-d. s'annulent.

On la représente souvent par $\overline{X}$

Comment calculer la moyenne?

Sommer toutes les observations et diviser par le nombre d'observation

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n}$$

ou

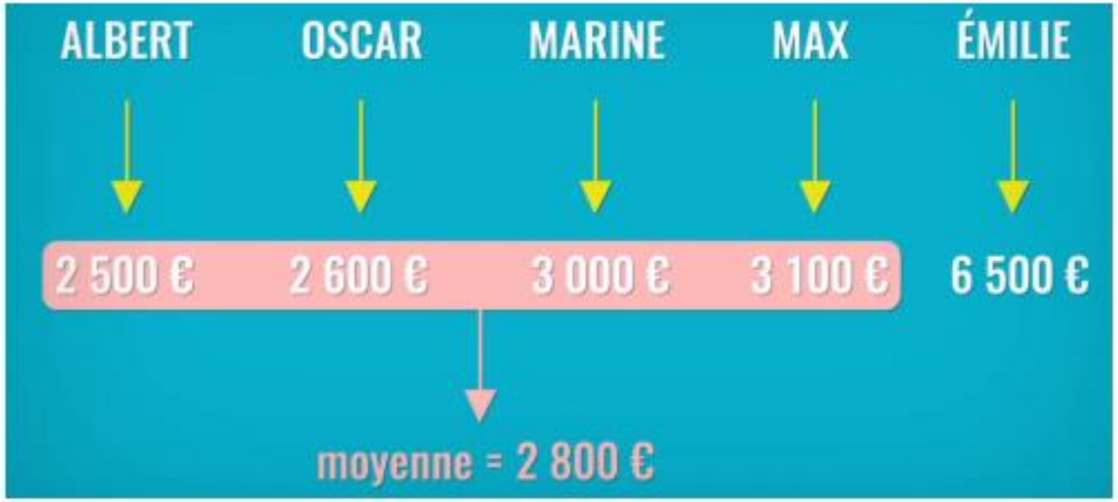
$$\overline{X} = \frac{1}{n} \sum X$$

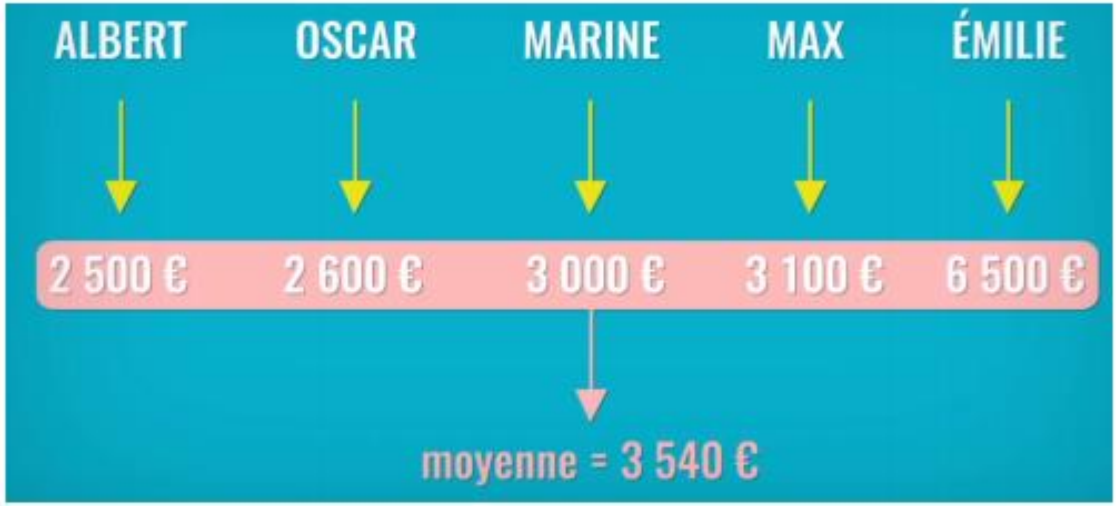
Vous avez tous, au moins une fois dans votre vie,
calculé votre moyenne dans un cours en ajoutant toutes vos notes
et en divisant par le nombre de notes.

Propriétés et observations

La moyenne est sensible aux valeurs extrêmes de la distribution :

quelques observations fort éloignées du centre peuvent fortement peser sur sa valeur (tandis qu'elles n'auront aucune influence sur la médiane ou le mode).

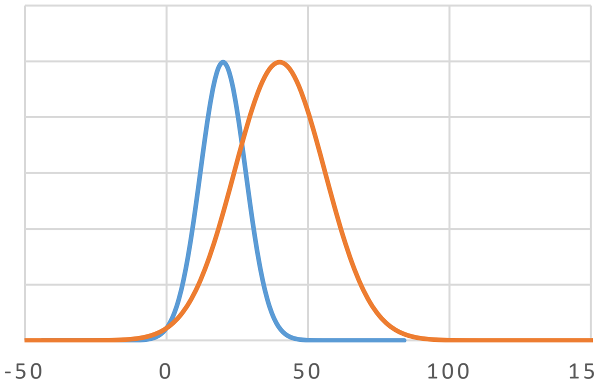




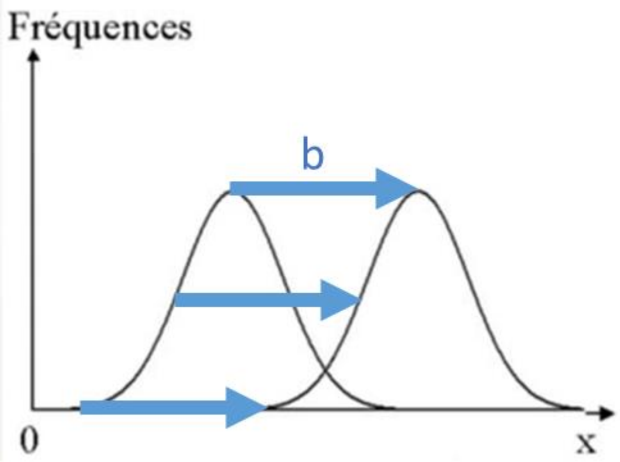
Autre propriété:

Moyenne(aX+b) = a.Moyenne(X) +b

(c'est vrai pour toutes les caractéristiques de localisation)



La moyenne bouge aussi par la déformation de la courbe due à la multiplication



La moyenne bouge avec la courbe par translation

5.3 La médiane

La médiane est la valeur observée (ou la modalité) qui, dans un rangement ordinal des observations, a la moitié des observations à sa gauche et l'autre moitié à sa droite.

ATTENTION ! C'est une notion qui n'a pas de sens dans le cas d'une variable nominale !

En effet, les modalités d'une variable nominale n'étant pas ordonnées, on ne peut pas ranger ordinalement une collection d'observations nominales.

Dans un graphique, c'est donc la position sur l'axe des modalités (valeurs observées) qui est telle que la moitié des observations sont à sa gauche, et l'autre moitié à sa droite. Malheureusement, dans les distributions concrètes, plusieurs cas embarrassants peuvent se produire …

- une telle position (ou valeur, ou échelon) n’existe pas (comme dans la collection d’observations ordinale ou numérique 1 2 2 3, où 2 laisse seulement un quart des observations à sa gauche et un quart à sa droite).

- Souvent une telle position existe, mais ce n’est pas une modalité observable de la variable (comme dans la collection numérique 1 2 3 4 où toutes les valeurs comprises entre 2 et 3 ont la moitié des observations à leur gauche et l’autre moitié à leur droite, bien que ces valeurs ne soient peut-être pas observables pour la variable étudiée par exemple le nombre d’enfants qui doit être un nombre entier).

- Parfois il y en a plusieurs (comme dans la collection numérique 1 2 3 4 où toutes les valeurs comprises entre 2 et 3 ont la moitié des observations à leur gauche et l’autre moitié à leur droite).

- En fait, le seul cas où l’on peut déterminer une unique valeur parfaitement médiane est celui d’une variable numérique avec M grand : en faisant comme si les individus de chaque classe étaient répartis de façon parfaitement uniforme au sein de leur classe, on peut trouver la valeur (observable) qui laisse la moitié des observations à sa gauche et l’autre moitié à sa droite !

Dans le cas d'une variable numérique où, $\mathcal{M}$ étant grand, les observations ont été regroupées en classes, la médiane est la valeur numérique exactement telle que la moitié des observations lui sont inférieures, et l'autre moitié supérieure.

Dans tous les autres cas, la médiane s'obtient en considérant le milieu du rangement ordinal des observations:

- Si le nombre n d'observations est impair, on considère l'observation occupant le rang (n/2 + ½). La modalité (échelon ou valeur) présentée par cette observation est dite « médiane » si les observations (strictement) à sa gauche sont exactement aussi nombreuses que les observations (strictement) à sa droite. Sinon, elle est **quasi-médiane**.
- Si n est pair, on considère les deux observations occupant le rang n/2 et (n/2 + 1). Si ces deux observations présentent la même modalité, cette modalité est dite « médiane » ou quasi-médiane selon qu'on a ou pas le même nombre d'observations à sa gauche qu'à sa droite. Si ces deux observations présentent deux modalités différentes, alors l'intervalle ouvert entre ces deux modalités est l'intervalle médian et tout ce qu'il contient est médian : dans le cas numérique, on prendra souvent le centre de cet intervalle comme valeur médiane ; dans le cas ordinal, les éventuels échelons inclus dans cet intervalle seront des échelons médians (quoique non observés) et les coupures seront des coupures médianes (il peut arriver, dans le cas ordinal, que l'intervalle médian ne soit constitué que d'une unique coupure, qui est donc la coupure médiane, et qu'il n'y ait donc pas d'échelon médian).

Positions respectives du mode, de la médiane et de la moyenne.

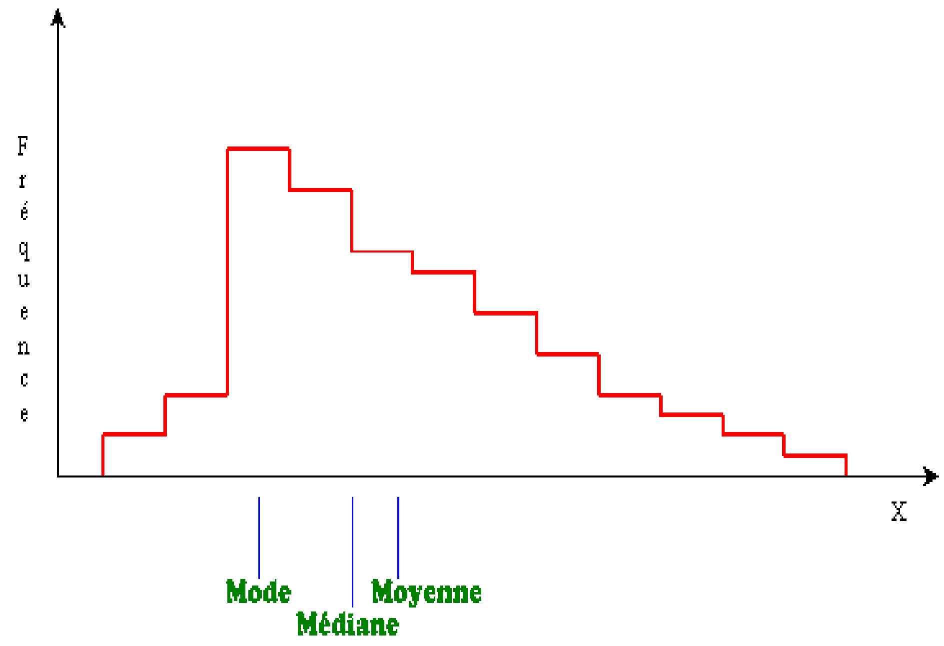
On peut montrer que toutes les dispositions sont possibles et pourquoi.

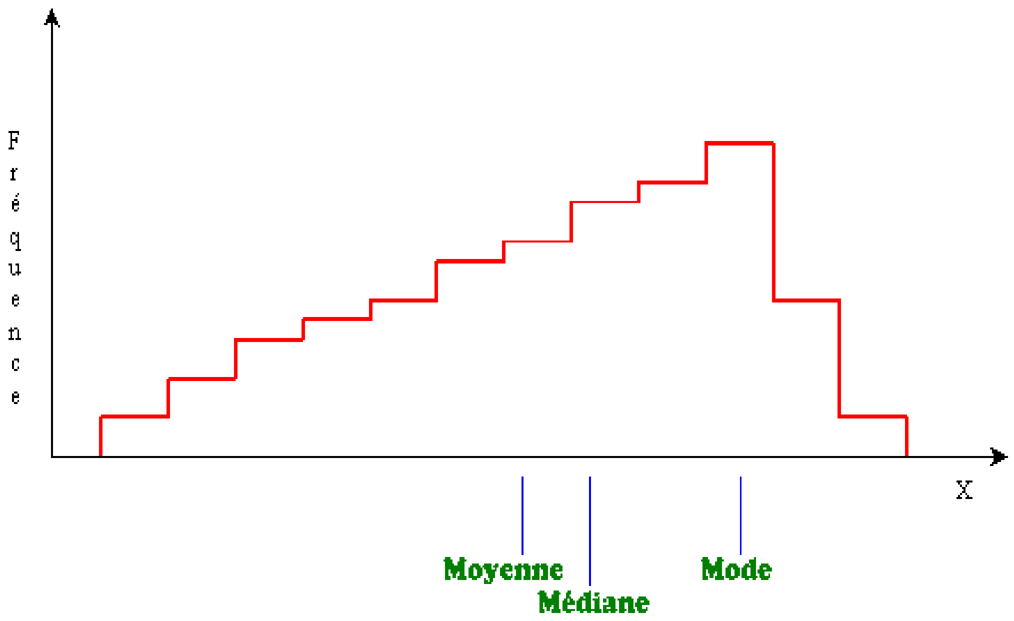
Mais on peut montrer aussi que si une distribution unimodale n'est pas très assymétrique, la moyenne est souvent plus « décalée » du mode que la médiane.

En conséquence, dans la plupart des distributions unimodales assez régulières,

la médiane se situe entre le mode et la moyenne.

Distribution asymétrique





A. Distribution étalée é droite:

B. Distribution étalée é gauche:

Si les valeurs entrêmes sont modifiées, la médiane ne change pas car elle n'est pas sensible aux valeurs extrêmes.
Par contre la moyenne change car elle tient compte de toutes les valeurs.

Exercice

Distribution sur leur âge des employés d'une entreprise

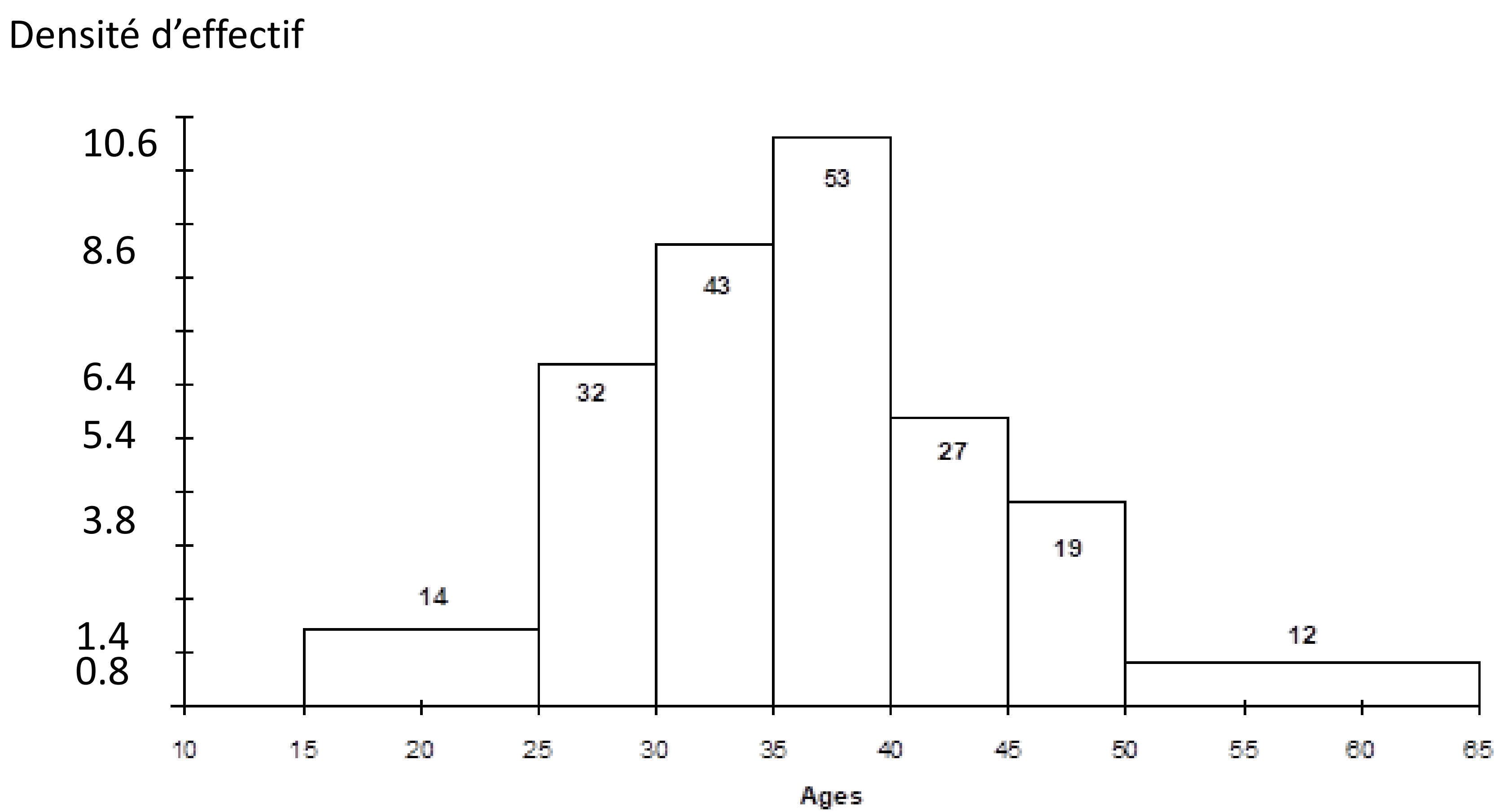
[15,25[	[25,30[	[30,35[	[35,40[	[40,45[	[45,50[	[50,65[
14	32	43	53	27	19	12

Type de variable?
Valeurs possibles?
Graphique? Type? Réalisez-le
Graphe des effectifs cumulés
Étendue?, Mode?, médiane? Moyenne? Écart type?

Type de variable: numérique continue
Valeur: entre 15 et 64 ans

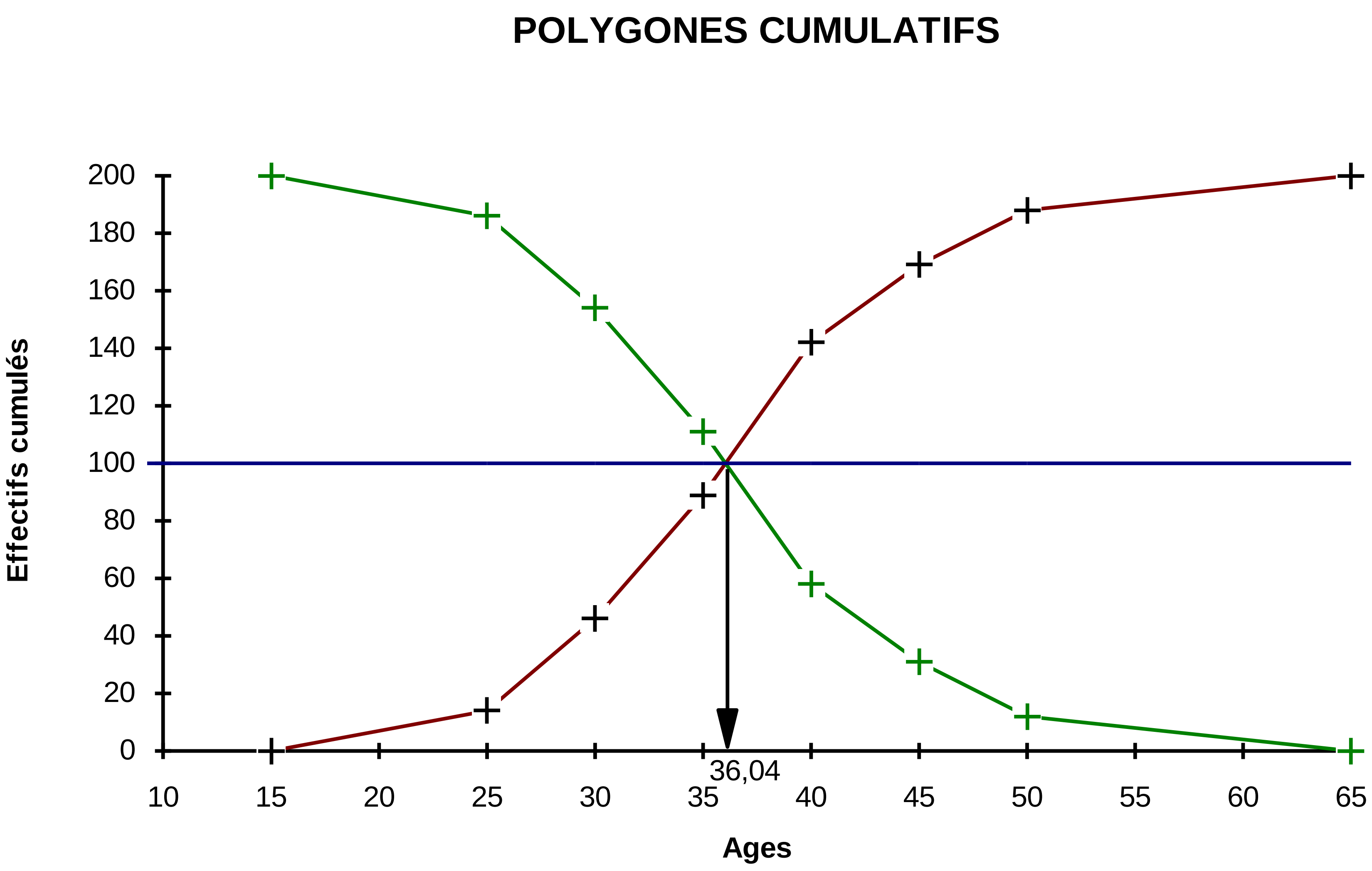
Graphique: Histogramme (variable continue et classes)

Méthode en bref : on construit des rectangles dont les aires sont proportionnelles aux effectifs des classes correspondantes (« histogramme » : représentation par les surfaces).
En fait, vous pouvez construire l'histogramme comme nous l'avons vu au cours, en calculant d'abord les densités de chaque classe. Donc les hauteurs des rectangles sont égales aux effectifs divisés par la largeur de classe (qui est ici égale à 5). On obtient ainsi la densité d'effectif pour chaque classe.



Histogramme de la distribution de l'âge des employés d'une entreprise

Effectifs cumulés



Étendue: 64 – 15 = 49 ans

Classe modale: [35-40[

Valeur modale: 37,5 ans

Médiane (à partir du graphe cumulé): +/-36 ans

Ages	Effectifs	Centres	Produits	Cum. à g.	Cum. à d.	Carrés	Produits
$[e_g ; e_d]$	ni	xi	ni fois xi	de e_d	de e_g	xi^2	ni fois xi^2
[15,25[	14	20,0	280,00	14	200	400,00	5.600,00
[25,30[	32	27,5	880,00	46	186	756,25	24.200,00
[30,35[	43	32,5	1.397,50	89	154	1.056,25	45.418,75
[35,40[	53	37,5	1.987,50	142	111	1.406,25	74.531,25
[40,45[	27	42,5	1.147,50	169	58	1.806,25	48.768,75
[45,50[	19	47,5	902,50	188	31	2.256,25	42.868,75
[50,65[	12	57,5	690,00	200	12	3.306,25	39.675,00
	200		7.285,00				281.062,50

Moyenne:

$$\bar{x} = \frac{\sum n_i x_i}{\sum n_i} = \frac{7285}{200} = 36{,}43 \text{ ans}$$

Ecart type

$$\sigma = \sqrt{\frac{281062{,}50}{200} - (36{,}43)^2}$$

$$\sigma = \sqrt{1405{,}3125 - 1327{,}1449}$$

$$\sigma = \sqrt{78{,}1676} = 8{,}841$$

Rappel des deux formules possibles:

$$\sigma = \sqrt{\frac{\sum n_i \left(x_i - \bar{x}\right)^2}{\sum n_i}} = \sqrt{\frac{\sum n_i\, x_i^2}{\sum n_i} - \bar{x}^2}$$

Médiane par le calcul

demi effectif : $\frac{200}{2} = 100$

classe de la 100e personne : $[35;40[$

rang dans cette classe : 100 - 89 = 11

amplitude de cette classe : $40 - 35 = 5$

effectif de cette classe : 53

médiane :

$$Q_2 = 35 + \frac{5 \times 11}{53} = 36{,}04$$