Chapitre 2 : Les caractéristiques de tendance centrale

Objectif général du chapitre :

Comprendre et appliquer les caractéristiques de tendance centrale

I. Le mode

Objectifs spécifiques :

- Saisir la manière de déterminer le mode.

Durée : 0,45 H.

Contenu :

1) *Définition :* noté M_0. C'est une valeur de la variable statistique qui a l'effectif (ou la fréquence) le (ou la) plus élevé (e). En d'autre terme, M_0 est la valeur de la variable la plus fréquente (ou la plus observée).

2) *Détermination :*

a/ Cas d'une variable discrète : Dans ce cas, pour trouver le mode, on cherche sur le tableau statistique la valeur x_i qui correspond à l'effectif (ou la fréquence) le (ou la) plus élevé (e).

Application à l'exemple 2 :

Nombre de retards	n_i	$f_i = n_i / n$
0	8	0,16
1	6	0,12
2	**22**	**0,44**
3	10	0,2
4 et plus	4	0,08
Total	50	1

Titre : Tableau 1 : La répartition des effectifs et des fréquences des ouvriers selon le nombre de retards.

M_0 = 2 retards.

Interprétation : le nombre de retards le plus observé est égal à 2. La majorité des ouvriers sont, alors, arrivés en retards 2 fois.

Remarque : on peut déterminer graphiquement le mode, à partir du diagramme en bâtons. L'abscisse du bâton le plus long correspond à M_0.

b/ Cas d'une variable continue :

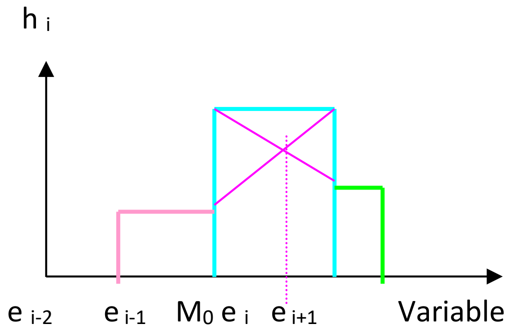
D'abord, on doit déterminer la classe modale $[e_{i-1}, e_i [$ qui contient absolument le mode. Elle correspond au h_i maximum. Ensuite, on déduit M_0 à partir de la formule suivante :

$$M_0 = \boxed{\frac{e_{i-1}(h_i - h_{i+1}) + e_i(h_i - h_{i-1})}{(h_i - h_{i+1}) + (h_i - h_{i-1})}}$$

Avec $h_i = f_i$ (ou n_i) si les classes sont d'amplitude a_i égales.

$h_i = f_i^C$ (ou n_i^C) si les classes sont d'amplitude a_i inégales.

Remarque : on peut déterminer le mode dans le cas continu, à partir du graphique (histogramme) de la manière suivante :



Titre: Graphique 1: Histogramme : La détermination du mode à partir du graphique dans le cas général.

Application à l'exemple 3 :

Salaire horaire en D	$h_i = f_i^C = \alpha f_i / a_i$
[0,8 – 1 [	0, 16
[1 – 1,2 [	**0, 28**
[1,2 – 1,6 [	0,215
[1,6 – 1,8 [	0,13
Total	-------

Titre: Tableau 2: La distribution des fréquences corrigées des ouvriers selon le salaire horaire.

α = 0,2.

La classe modale est **[1 – 1,2 [** puisqu'elle correspond au h_i maximum (0,28).

D'où, $M_0 = \frac{1(0{,}28-0{,}215)+1{,}2(0{,}28-0{,}16)}{(0{,}28-0{,}215)+(0{,}28-0{,}16)} \approx 1{,}13^D$.

Interprétation : La plupart des ouvriers touchent un salaire horaire de 1,13D.

Remarque : Une distribution statistique peut être uni-modale (possède un seul mode), bimodale (deux modes) ou tri-modale, …La distribution pluri -modale traduite l'hétérogénéité de la population.

II. La médiane

Objectifs spécifiques :

- Dénombrer les caractéristiques de la médiane et maîtriser son calcul.

Durée : 1,5 H.

Contenu :

1) <u>Définition</u> : (noté M_e) est une valeur de la variable tel que le nombre des observations de valeurs inférieur à la médiane sera égal au nombre des observations de valeurs supérieur à cette même valeur de la variable (qui est la médiane). Alors, la médiane partage la population totale en deux groupes d'effectifs égaux. Elle est définie par la fréquence (ou l'effectif) cumulée :

M_e est telle que : $F (M_e) = 1 / 2$ ou $E (M_e) = n / 2$.

2) <u>Détermination :</u>

a/ Cas de « n » observations individuelles : Soit $x_1, x_2,\dots, x_i, \dots. x_n$: rangés par ordre croissant ou décroissant. On distingue les deux cas suivants :

- « n » impaire : C'est à dire il existe un $L \in \aleph$ tel que : $n = 2 (L) + 1$.

La médiane est, alors, la valeur x_{L+1} : c'est l'observation de rang (L + 1).

Exemple : on a les valeurs suivantes : 2 – 3 – **4** – 5 – 6

$$x_1 - x_2 - \mathbf{x_3} - x_4 - x_5$$

$n = 5 = 2 (2) + 1$, d'où L = 2. Et puisque $M_e = x_{L+1} = x_3 = 4$.

- « n » paire : C'est à dire il existe un $L \in \aleph$ tel que : $n = 2 (L)$. La médiane est, dans ce cas, « un interval médian » dont les extrémités sont les observations de rang (L) et (L+1) : $[x_L - x_{L+1}]$.

Exemple : on a les valeurs suivantes : 0 – 1 – **2 – 5** – 6 - 7

$$x_1 - x_2 - \mathbf{x_3 - x_4} - x_5 - x_6$$

$n = 6 = 2 (L)$, d'où L = 3. L'intervalle médian est, alors, $[x_3 - x_4] = [2 – 5]$.

Remarque : x_L et x_{L+1} peuvent être de même valeur, dans ce cas on aura une valeur de la médiane.

b/ Cas d'une variable discrète : A partir du tableau, on cherche une valeur x_i, tel que : $F (x_i) = 0,5$ (ou $E (x_i) = n / 2$). On distingue deux cas :

☺ Il existe sur le tableau statistique une valeur x_i tel que F (x_i) = 0,5 (ou E (x_i) = n / 2). Alors, $M_e = x_i$ par convention.

☺ Il n'existe pas sur le tableau statistique une valeur x_i tel que F (x_i) = 0,5 (ou E (x_i) = n / 2). Alors, par convention, $M_e = x_i$ tel que $F (x_i) < 0,5 < F (x_{i+1})$ ou $E (x_i) < n / 2 < E (x_{i+1})$.

Application à l'exemple 2 :

Nombre de retards	$f_i = n_i / n$	FC ↑ : $F(x_i) = \sum_{j=1}^{i-1} f_j$
0	0,16	0
1	0,12	0,16
2	0,44	0,28
3	0,2	0,72
4 et plus	0,08	0,92
Total	1	

Titre : Tableau 3 : Distribution des fréquences cumulées croissantes des ouvriers selon le nombre de retatrds.

On a : F (2) = 0,28 < 0,5 < F (3) = 0,72, d'où, par convention, M_e = 2 retards.

Interprétation : le nombre d'ouvriers ayant un nombre de retards inférieur à 2 est égal au nombre d'ouvriers ayant un nombre de retards supérieur à 2.

c/ Cas d'une variable continue :

Dans ce cas, on cherche sur le tableau une valeur « e_i », tel que F (e_i) = 0,5 (ou E (e_i) = n / 2). On distingue deux cas :

☺ Il existe sur le tableau statistique une valeur « e_i », tel que F (e_i) = 0,5 (ou E (e_i) = n / 2). Alors, $M_e = e_i$ (la borne supérieure de l'intervalle qui correspond au FC ↑ = 0,5).

☺ Il n'existe pas sur le tableau statistique une valeur « e_i », tel que F (e_i) = 0,5 (ou E(e_i) = n/2). Alors, on détermine la classe médiane $[e_{i-1}, e_i[$, tel que $F (e_{i-1}) < 0,5 < F (e_i)$ ou $E (e_{i-1}) < n/2 < E(e_i)$.

M_e sera calculer, alors par interpolation linéaire, suivant la formule suivante :

$$\boxed{M_e = e_{i-1} + \frac{e_i - e_{i-1}}{F(e_i) - F(e_{i-1})}\left(0,5 - F(e_{i-1})\right)}$$

Application à l'exemple 3 :

Salaire horaire en D	f_i	FC ↑ : $F(x_i) = \sum_{j=1}^{i-1} f_j$
[0,8 – 1 [	0, 16	0,16
[1 – **1,2** [	0, 28	**0,44**
[1,2 – **1,6** [	0,43	**0,87**
[1,6 – 1,8 [	0,13	1
Total	1	

Titre : Tableau4 : Distribution des fréquences cumulées croissantes des ouvriers selon le salaire horaire.

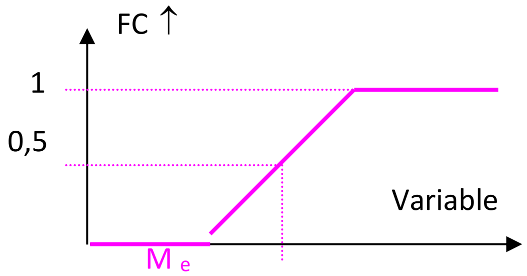
Ma classe médiane est déterminée de la manière suivante :

F (1,2) = 0,44 < 0,5 < F (1,6) = 0,87

D'où $M_e \in [\, 1,2 - 1,6 \,[$, par la suite, $M_e = 1,2 + \frac{1,6-1,2}{0,87-0,44}(0,5-0,44) = 1,256$ D.

Interprétation : La moitié du personnel touche un salaire horaire inférieur à 1,256 D, l'autre moitié touche un salaire horaire supérieur à 1,256 D.

Remarque : La médiane peut être déterminée graphiquement de la manière suivante : c'est l'abscisse du point de la courbe cumulative dont l'ordonné est égal à 0,5.



Titre : Graphique 2 : Courbe cumulative : Détermination de M_e à partir du graphique dans le cas général

III. La moyenne arithmétique

Objectifs spécifiques :

- Identifier la notion de moyenne arithmétique.

Durée : 1,5 H.

Contenu :

1) <u>Définition</u> : C'est la somme des valeurs observées divisée par le nombre total des observations.

2) ***Détermination :*** On distingue les trois cas suivants :

a/ Cas de « n » observations individuelles : Soit $x_1, x_2, \ldots, x_i, \ldots. x_n$: Rangés par ordre croissant ou décroissant. $\boxed{\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}}$: c'est la moyenne arithmétique simple.

Exemple d'application : soit les valeurs suivantes : 150, 90, 100, 100, 130.

$\overline{X} = \frac{150+90+100+100+130}{5} = 114.$

b/ Cas d'une variable discrète : Soit les modalités $x_1, x_2, \ldots, x_i, \ldots. x_k$, d'effectifs respectifs $n_1, n_2, \ldots, n_i, \ldots. n_k$; avec $\sum_{i=1}^{k} n_i = n$.

$$\boxed{\overline{X} = \frac{\sum_{i=1}^{k} n_i x_i}{n} = \sum_{i=1}^{k} f_i \times x_i}$$: c'est la moyenne arithmétique pondérée.

Application à l'exemple 2 :

Nombre de retards	n_i	$n_i \times x_i$
0	8	0
1	6	6
2	22	44
3	10	30
4 et plus	4	16
Total	50	96

Titre : Tableau 5 : Calcul du nombre de retards moyen par ouvrier.

$\overline{X} = \frac{\sum_{i=1}^{k} n_i x_i}{n} = \frac{96}{50} = 1{,}92$ retards.

Interprétation : Le nombre moyen de retards par ouvrier est de 1,92.

c/ Cas d'une variable continue :

Soit les classes $[e_{i-1}, e_i[$, avec i = 1, 2, … , k ; Et d'effectifs respectifs : $n_1, n_2, \ldots, n_i, \ldots n_k$, avec : $\sum_{i=1}^{k} n_i = n$, d'où, $\boxed{\overline{X} = \frac{\sum_{i=1}^{k} n_i c_i}{n} = \sum_{i=1}^{k} f_i c_i}$: C'est la moyenne arithmétique pondérée.

Avec $c_i = \frac{e_{i-1}+e_i}{2}$: c'est le centre de la classe i : $[e_{i-1}, e_i[$.

Application à l'exemple 3 :

Salaire horaire en D	f_i	c_i	$f_i \times c_i$
[0,8 – 1 [	0,16	0,9	0,144
[1 – **1,2** [	0,28	1,1	0,308
[1,2 – **1,6** [	0,43	1,4	0,602
[1,6 – 1,8 [	0,13	1,7	0,221
Total	1		1,275

Titre: Tableau 6: Le salaire horaire moyen par ouvrier.

$$\overline{X} = \sum_{i=1}^{k} f_i c_i = 1{,}275^{\text{ D}}.$$

Interprétation : Le salaire horaire moyen par ouvrier est de 1,275 $^{\text{D}}$.

3) <u>Propriétés</u> : on va citer deux propriétés de la moyenne arithmétique : celle de la linéarité et celle de l'union.

❖ La propriété de l'union : on a deux populations P_1 et P_2, d'effectifs respectifs n_1 et n_2, de moyennes arithmétiques respectives $\overline{X}_1$ et $\overline{X}_2$. Alors, la population totale $P = P_1 \cup P_2$, d'effectif total $n = n_1 + n_2$, et a pour moyenne arithmétique : $\boxed{\overline{X} = \frac{n_1 \overline{X}_1}{n} + \frac{n_2 \overline{X}_2}{n} =}$

$$\boxed{f_1 \times \overline{X}_1 + f_2 \overline{X}_2}$$

Cette propriété peut être généraliser pour l'union de k population, et on applique le même principe.

❖ La propriété de la linéarité : soit X et X' deux variables statistiques telles que :

$x_i = a\,x'_i + x_0$, alors, $\boxed{\overline{X} = a\,\overline{X}' + x_0}$.

Intérêt pratique de cette propriété :

♪ Elle facilite le calcul de la moyenne comme dans le cas où, par exemple, on calcul la moyenne d'une série de salaires, puis, après une augmentation (+ 5%) de ses derniers, on appliquerait directement cette propriété afin de trouver la nouvelle moyenne.

♪ Elle facilite le calcul de la moyenne comme dans le cas de changement de variable (changement d'échelle « a » ou d'origine « x_0 »).

<u>Calcul de la moyenne par changement de variable :</u> se fait en trois temps :

- 1ère étape : effectuer un changement de variable revient à remplacer la variable initiale

X par une autre variable X' dite « auxiliaire », de manière à rendre le calcul de $\overline{X}'$ plus simple que celui de $\overline{X}$, et telles que :

Va D

$x'_i = \frac{x_i - x_0}{a}$

Avec : $x_0 = \frac{x_1 + x_k}{2}$, si x_0 est une modalité de la variable ; Sinon, on prend la modalité la plus proche de x_0.

On peut aussi prendre x_0, directement, le mode.

Et, a = la plus grande valeur des ($x_i - x_0$).

Va C

$c'_i = \frac{c_i - c_0}{a}$

Avec : $c_0 = \frac{c_1 + c_k}{2}$, si c_0 est une valeur centrale de la distribution ; sinon, on prend la modalité la plus proche de c_0.

On peut aussi prendre c_0, directement, le centre de la classe modale.

Et, a = la plus petite amplitude.

- 2ème étape : calculer $\overline{X}'$ de la manière suivante :

$$\overline{X}' = \frac{\sum_{i=1}^{k} n_i x'_i}{n} = \sum_{i=1}^{k} f_i \times x'_i$$

$$\overline{X}' = \frac{\sum_{i=1}^{k} n_i c'_i}{n} = \sum_{i=1}^{k} f_i c'_i$$

- 3ème étape : utiliser la linéarité pour déduire $\overline{X}$ de $\overline{X}'$, de la manière suivante :

Puisque : $x'_i = \frac{x_i - x_0}{a}$, alors,

$x_i = a\,x'_i + x_0 \Leftrightarrow \overline{X} = a\,\overline{X}' + x_0$.

Puisque : $c'_i = \frac{c_i - c_0}{a}$, alors,

$c_i = a\,c'_i + c_0 \Leftrightarrow \overline{X} = a\,\overline{X}' + c_0$.

Exemple d'application :

J'ai pris ces données fictives (x_i et n_i) très grands, exprès pour montrer l'utilité de la méthode de changement de variable.

$x_0 = \frac{x_1 + x_k}{2} = \frac{885+925}{2} = 905$: Cette valeur n'est pas l'une des modalités de X, on choisit, alors la modalité la plus proche de 905 qui est 915. Cette valeur coïncide avec celle du mode de cette distribution.

a = 10 : Le plus grand ($x_i - x_0$).

x_i	n_i	$x_i - x_0 = x_i - 915$	$x'_i = \frac{x_i - 915}{10}$	$n_i \times x'_i$
885 915 925	124 183 93	- 30 0 10	- 3 0 1	- 372 0 93
Total	400			- 279

Titre : Tableau 7 : Calcul de la moyenne par changement de variable.

On remarque, alors, qu'il y'a une grande différence entre les valeurs initiales x_i et celles auxiliaires x'_i. Ces dernières sont plus petites (en valeur absolue) et rendent le calcul de la moyenne plus simple.

Ainsi, on a suivi les trois étapes du changement de variable :

- 1ère étape : $x'_i = \frac{x_i - 915}{10}$.
- 2ème étape : $\overline{X}' = \frac{\sum_{i=1}^{k} n_i x'_i}{n} = \frac{-279}{400} = -0,697$.
- 3ème étape : Déduire $\overline{X}$ de $\overline{X}'$, puisque : $x'_i = \frac{x_i - x_0}{a}$, alors, $x_i = a\,x'_i + x_0$

$\Leftrightarrow \overline{X} = a\,\overline{X}' + x_0 = 10 \times (-0,697) + 915 =$ **908,025**.

IV. La relation entre mode, médiane et moyenne arithmétique

Objectifs spécifiques :

- Discerner la relation qui peut exister entre les trois paramètres précédents.

Durée : 0,45 H.

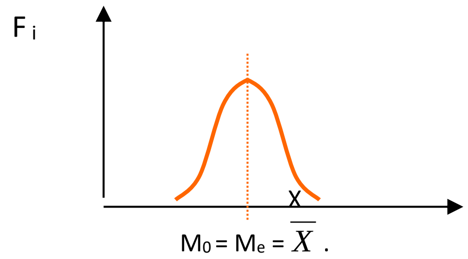
Contenu :

Pour les distributions unies-modales, il peut exister l'une des trois relations observées entre M_0, M_e et $\overline{X}$.

- Distribution symétrique : si $M_0 = M_e = \overline{X}$.

Cette situation peut être schématisée comme suit :

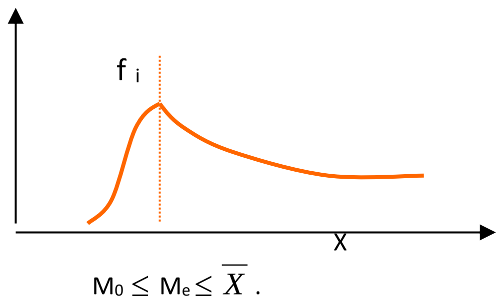
Titre : Graphique 3 : Distribution symétrique.



- Distribution asymétrique positive :

Si $M_0 \leq M_e \leq \overline{X}$: étalement à droite.

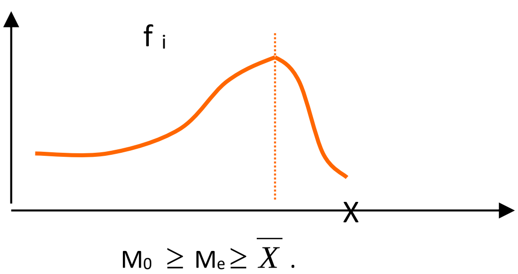
Titre : Graphique 4 : Distribution asymétrique à gauche.



- Distribution asymétrique négative :

Si $M_0 \geq M_e \geq \overline{X}$: étalement à gauche.

Titre : Graphique 5 : Distribution asymétrique à droite.



V. Les autres types de moyennes

Objectifs spécifiques :

- Maîtriser deux autres types de moyennes.

Durée : 1,5 H.

Contenu :

1) <u>Moyenne géométrique :</u>

a/ Définition : est notée G, elle se définit, selon le cas, de la manière suivante :

♠ Cas de "n" observations individuelles : $x_1, x_2, \ldots, x_i, \ldots, x_n$.

$$G = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} = \sqrt[n]{x_1 \times x_2 \times \ldots \times x_n}.$$

♠ Cas de variable discrète : avec n_i et x_i allant de 1 à k.

$$G = \left(\prod_{i=1}^{k} x_i^{n_i}\right)^{\frac{1}{n}} = \sqrt[n]{x_1^{n_1} \times x_2^{n_2} \times \ldots \times x_k^{n_k}}.$$

♠ Cas de variable continue : avec n_i et c_i allant de 1 à k.

$$G = \left(\prod_{i=1}^{k} c_i^{n_i}\right)^{\frac{1}{n}} = \sqrt[n]{c_1^{n_1} \times c_2^{n_2} \times \ldots \times c_k^{n_k}}.$$

b/ Propriétés : la moyenne géométrique présente deux propriétés :

- Soient X, Y et Z trois variables statistiques tel que : $z_i = x_i \times y_i$, quelque soit i ; alors,

  $G(Z) = G(X \times Y) = G(X) \times G(Y)$.

- Soient X, Y et Z trois variables statistiques tel que : $z_i = \frac{x_i}{y_i}$, quelque soit i ; alors,

$G(Z) = G\left(\frac{X}{Y}\right) = \frac{G(X)}{G(Y)}$.

Remarque :

La moyenne géométrique est utilisée, surtout, en économie afin de calculer des variables sous forme de taux (par exemple taux de croissance, taux de chômage).

2) <u>Moyenne harmonique :</u>

a/ Définition : est notée H et définie, suivant le cas, comme suit :

♠ Cas de « n » observations individuelles : $x_1, x_2, \ldots, x_i, \ldots, x_n$.

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

♠ Cas de variable discrète : avec n_i et x_i allant de 1 à k.

H = $\frac{n}{\sum_{i=1}^{k} n_i \left(\frac{1}{x_i} \right)}$

♠ Cas de variable continue : avec n_i et x_i allant de 1 à k.

H = $\frac{n}{\sum_{i=1}^{k} n_i \left(\frac{1}{c_i} \right)}$

Remarque 1 :

La moyenne harmonique est utilisée, surtout, quand la variable en question se présente sous la forme d'un rapport entre deux grandeurs (vitesse en km par h).

Remarque 2 : On vérifie, toujours, la relation suivante entre les trois moyennes : H $\leq$ G$\leq$ $\overline{X}$.