Chapitre 3. Les caractéristiques de dispersion

1. LES CARACTÉRISTIQUES DE DISPERSION ABSOLUES
2. LES CARACTÉRISTIQUES DE DISPERSION RELATIVES

Chapitre 3. Les caractéristiques de dispersion

1. Les caractéristiques de dispersion absolue

- L'étendue
- Les intervalles interquantiles
- L'écart absolu moyen
- La variance et l'écart type

2. Les caractéristiques de dispersion relative

- Le Minmax
- Les écarts interquantiles
- Le coefficient de variation
- Les moments

Chapitre 3. Les caractéristiques de dispersion

Certaines distributions, tout en présentant les mêmes caractéristiques de tendance centrales et de position, peuvent être très différentes les unes des autres. Considérons deux pays A et B. Les revenus moyens dans les pays A et B sont identiques. Mais, dans le pays A il y a une proportion importante d'individus qui gagnent bien plus que la moyenne et une autre proportion qui gagne bien moins que la moyenne. Dans le pays B, en revanche, les revenus sont concentrés autour de la moyenne. Les distributions des revenus dans les deux pays sont données par la figure suivante.

Chapitre 3. Les caractéristiques de dispersion

:

:

- Insérer graphique word

.

Chapitre 3. Les caractéristiques de dispersion

Dans le cas du pays B la moyenne arithmétique ne présente qu'un intérêt limité car elle n'est pas très représentative des valeurs de la distribution qui sont très dispersées autour d'elle. Le résumé est complété par d'autres caractéristiques qui donnent une idée sur l'éloignement et la dispersion des valeurs autour de la tendance centrale. Il s'agit des caractéristiques de dispersion. Comme pour les caractéristiques de tendance centrale, on peut imaginer plusieurs caractéristiques de dispersion : l'étendue, les écarts interquantiles, l'écart absolu moyen et la variance ou l'ecart-type

Chapitre 3. Les caractéristiques de dispersion

1. les caractéristiques de dispersion absolue

Ces caractéristiques s'expriment dans l'unité même du caractère pour lequel elles sont censées mesurer la dispersion. Elles ne sont pas très indiquées pour comparer la dispersion de plusieurs distributions qui présentent des unités différentes ou des dimensions très différentes.

Chapitre 3. Les caractéristiques de dispersion

1. les caractéristiques de dispersion absolue / L'étendue

L'étendue, notée *e*, représente la longueur de l'intervalle de variation de la série. C'est la différence entre la plus grande et la plus petite valeur.

$$e = xmin_{max}$$

L'étendue est très rarement utilisée seule. Certes elle est très facile à calculer mais il ne prend pas en considération toutes les observations et elle est très sensible aux valeurs extrêmes.

Chapitre 3. Les caractéristiques de dispersion

1. les caractéristiques de dispersion absolue / Les intervalles interquartiles

Si les quantiles sont des mesures de tendance centrale, les écarts entre quantiles de même type peuvent être utilisés pour apprécier la dispersion. Généralement, on définit les écarts interquartiles (q_3-q_1 qui contient 50% des observations) et l'écart inter-déciles (d_9-d_1 qui contient 80% des observations).

Bien entendu, lorsqu'on compare plusieurs distributions différentes on préférera utiliser des écarts relatifs. Les écarts interquartiles et interdéciles deviennent respectivement

$$(q_3\text{-}q_1)/q_2 \qquad \text{et} \qquad (d_9\text{-}d_1)/d_5.$$

Le plus grand reproche qui est fait à ces caractéristiques et qu'elles ne se prêtent pas facilement aux manipulations algébriques.

Chapitre 3. Les caractéristiques de dispersion

1. les caractéristiques de dispersion absolue / L'écart absolu moyen

L'écart absolu moyen, noté $\bar{e}_m$, est la moyenne arithmétique de la valeur absolue des écarts par rapport à la moyenne arithmétique

$$\bar{e}_m = \frac{1}{n}\sum_{i=1}^{k} n_i|x_i - \bar{x}| = \bar{e}_m = \sum_{i=1}^{k} f_i|x_i - \bar{x}|$$

L'écart absolu médian se définit de la même manière

$$\bar{e}_m = \frac{1}{n}\sum_{i=1}^{k} n_i\,|x_i - Mé| = \bar{e}_m = \sum_{i=1}^{k} f_i|x_i - Mé|$$

Comme pour la caractéristique précédente, le seul inconvénient de l'écart absolu moyen ou médian est la présence de la fonction valeur absolue qui la rend très difficile à traiter algébriquement.

1. les caractéristiques de dispersion absolue / La variance et l'écart type

Définition : La variance, notée $V(x)$ ou σ^2, se définit comme la moyenne arithmétique des carrés des écarts des valeurs à la moyenne.

$$V(x) = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \bar{x})^2 = \sum_{i=1}^{k} f_i(x_i - \bar{x})^2$$

Cette formule est dite formule de définition, elle conduit à quelques approximations dans les calculs. En pratique, on utilise une autre formule de la variance obtenue par décomposition de la formule précédente et appelée formule développée. Il s'agit de

$$V(x) = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \bar{x}^2 = \sum_{i=1}^{k} f_i x_i^2 - \bar{x}^2$$

La variance peut ainsi être définie comme *la moyenne des carrés moins le carré de la moyenne.*

Propriétés de la variance :

(*i*) La variance, a la différence des autres caractéristiques de dispersion se prête aisément aux calculs.

(*ii*) $V(x) \geq 0$ avec $V(x) = 0$ pour x une constante.

(*iii*) $V(ax + b) = a^2V(x)$.

(*iv*) Si et sont les variances de deux distributions de tailles respectives n_1 et n_2 alors la variance totale $V(x)$ est donnée par

$$V(x) = \frac{n_1 V(x)_1 + n_2 V(x)_2}{n_1 + n_2}$$

si les deux distributions présentent la même moyenne arithmétique.

(*v*) La variance est très sensible aux valeurs aberrantes

(*v*) La variance s'exprime dans le double de l'unité de la valeur de la variable x elle n'a pas de pouvoir descriptif.

La limite présentée dans la dernière propriété nous conduit à introduire une nouvelle caractéristique appelée écart-type, notée σ, définie comme la racine carrée de la variance. $\sigma = \sqrt{V(x)}$. Bien entendu, l'écart-type s'exprime dans la même unité que la variable *x*.

La variance, et donc l'écart-type, sous-estiment l'écart des valeurs par rapport à la moyenne notamment pour *n* petit.

Dans ce cas on peut utiliser une variance corrigée appelée quasi-variance. Elle est donnée par

$$S_n = \frac{1}{n-1}\sum_{i=1}^{k} n_i(x_i - \overline{x})^2$$

1. les caractéristiques de dispersion absolue / La variance et l'écart type

L'écart type est la caractéristique de dispersion la plus utilisée. Son seul inconvénient est qu'elle est une caractéristique absolue. Elle ne peut être utilisée pour comparer la dispersion des distributions très différentes (qui ont des moyennes différentes).

.

.

Chapitre 3. Les caractéristiques de dispersion

2. les caractéristiques de dispersion relative

Toutes les caractéristiques précédentes sont utilisées pour décrire une seule distribution. Pour comparer plusieurs distributions entre elles d'un point de vue de la dispersion, on doit utiliser des caractéristiques relatives pour lesquelles les échelles et les unités ne présentent pas d'importance. Généralement elles se présentent sous la forme d'une caractéristique de dispersion absolue et une caractéristique de tendance centrale.

.

Chapitre 3. Les caractéristiques de dispersion

2. les caractéristiques de dispersion absolue / Le minmax

Le Minmax se définit comme le rapport entre la valeur la plus élevée est la valeur la plus faible de la distribution

$$Min\,max = xmin_{max}$$

:

Comme pour l'étendue cette caractéristique n'est pas très pertinente en présence de valeurs extrêmes.

.

Chapitre 3. Les caractéristiques de dispersion

2. les caractéristiques de dispersion relative/ Le écarts interquantiles

On retiendra deux écarts. Les écarts interquartiles et interdéciles. Soient

$$IQ = (q_3 - q_1)/q_2$$
$$IK = (d_9 - d_1)/d_5$$

L'inconvénient de ce type de caractéristiques est qu'elles ne présentent pas de limites supérieures.

Chapitre 3. Les caractéristiques de dispersion

2. les caractéristiques de dispersion absolue / Le coefficient de variation

Ce coefficient de variation est la caractéristique de dispersion relative la plus utilisée. Elle est donnée par

$$CV = \frac{\sigma}{\bar{x}} \quad \text{ou} \quad CV\% = \frac{\sigma}{\bar{x}} \times 100$$

.

Sa seule limite est qu'elle est très sensible à la moyenne arithmétique et ne peut pas être utilisée lorsque celle-ci est proche de zéro.

Chapitre 3. Les caractéristiques de dispersion

2. les caractéristiques de dispersion absolue / Le coefficient de variation

Exemple. Une entreprise emploie des hommes et des femmes. Le salaire moyen des hommes est de 6000 MAD et celui des femmes est de 3000 MAD. L'écart-type des salaires pour les hommes est de 500 MAD et pour les femmes de 300 MAD. Malgré ça, il serait faux d'affirmer que les salaires des hommes présente une plus forte dispersion que celui des femmes. En effet, comme les deux distributions sont assez différentes (les salaires moyens sont très différents) il serait plus approprié de comparer les dispersions en utilisant une caractéristique de dispersion relative *i.e* le coefficient de variation. Celui des hommes est de 8,3% et ce lui des femmes est de 10%. Aussi, la dispersion des salaires plus importante chez les hommes que chez les femmes.

Les moments sont une généralisation des concepts de moyenne et de variance.

Définition : On appelle moment général d'ordre *r* par rapport à une origine *a* la quantité donnée par :

.

$$M_r = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - a)^r$$

.

Généralement on définit deux types de moments :

Les moments simples : On appelle moment simple d'ordre *r*, noté m_r , le moment d'origine 0. Les moments simples sont une généralisation de la moyenne

$$m_r = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^r = \sum_{i=1}^{k} f_i x_i^r$$

Ainsi, $m_1 = \bar{x}$ et $m_2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2$

Les moments centrés : On appelle moment centré d'ordre *r*, noté μ_r, le moment d'origine $\bar{x}$. Les moments centrés sont une généralisation de la variance. $\mu_r = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \bar{x})^r = \sum_{i=1}^{k} f_i\,(x_i - \bar{x})^r$

Pour $\mu_2 = \frac{1}{n}\sum n_i(x_i - \bar{x})^2 = \sigma^2$

Comme $\sigma^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \bar{x}^2$ il vient que $\mu_2 = m_2 - m_1^2$

- On peut mettre ainsi, tous les moments centrés en fonctions des moments simples d'ordre inférieur.
- Les moments centrés d'ordre 3 et 4 sont très utilisés. Ils permettent de caractériser la forme d'une distribution. Ils font partie des caractéristiques de forme.

Propriétés :

$$\mu_2 = m_2 - m_1^2$$

$$\mu_3 = m_3 - 3m_2m_1 + 2m_1^3$$

$$\mu_4 = m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4$$

Exemple. Calculez les moments centrés d'ordre 2, 3 et 4 de la série statistique suivante (salaires des 200 individus d'une entreprise)

xi	ni
[3000-4000[	42
[4000-5000[	53
[5000-6000[	85
[6000-7000[	12
[7000-8000[	8
Total	200

Chapitre 3. Les caractéristiques de dispersion

2. les caractéristiques de dispersion absolue / Les moments

x_i	n_i	c_i	$n_i \times c_i$	$(c_i - \overline{x})^2$	$n_i \times (c_i - \overline{x})^2$	$n_i c_i^2$	$n_i c_i^3$	$n_i \times (c_i - \overline{x})^3$	C_i^2	C_i^3	$n_i c_i^4$
[3-4[	42	3,5	147	2,11	88,91	514,5	1800,75	-129,37	12,25	42,88	6302,62
[4-5[	53	4,5	238,5	0,20	10,97	1073,25	4829,62	-4,99	20,25	91,13	21733,31
[5-6[	85	5,5	467,5	0,29	25,24	2571,25	14141,87	13,75	30,25	166,38	77780,31
[6-7[	12	6,5	78	2,38	28,64	507	3295,5	44,25	42,25	274,63	21420,75
[7-8[	8	7,5	: 60	6,47	51,81	450	3375	131,87	56,25	421,88	25312,5
Total	**200**	**27,5**	**991**	**11,485125**	**205,595**	**5 116,00**	**27 442,75**	**55,52355**	**161,25**	**996,88**	**152 549,50**

:

.

.