Probabilités et statistiques-séance 1

Licence 1-semestre 1

Allegret Audrey

Maître de Conférences - Université de Toulon, LEAD

21 novembre 2022

Coordonnées

- Adresse mail :`audrey.sallenave@univ-tln.fr`
- 1 CC d'1h30 à mi-parcours
- Partiel de 1h30h-2h

Plan

Ecart absolu moyen

Définition de L'écart absolu moyen

Pour avoir une meilleure idée de la dispersion proprement dite, il faut regarder les **écarts par rapport à la valeur centrale**.

Ecart absolu moyen

Définition de L'écart absolu moyen

Pour avoir une meilleure idée de la dispersion proprement dite, il faut regarder les **écarts par rapport à la valeur centrale**.
On va commencer par s'intéresser aux **écarts par rapport à la moyenne**.

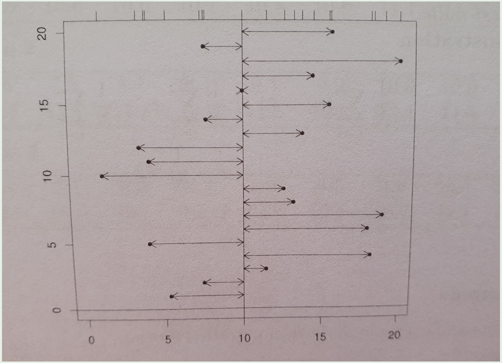
Ecart absolu moyen

Définition de L'écart absolu moyen

Pour avoir une meilleure idée de la dispersion proprement dite, il faut regarder les **écarts par rapport à la valeur centrale**.
On va commencer par s'intéresser aux **écarts par rapport à la moyenne**.

Ecart absolu moyen

Représentation graphique



Ecart absolu moyen

Définition de L'écart absolu moyen

L'écart absolu moyen est la moyenne arithmétique de la valeur absolue des écarts à la moyenne :

Ecart absolu moyen

Définition de L'écart absolu moyen

L'écart absolu moyen est la moyenne arithmétique de la valeur absolue des écarts à la moyenne :

$$eam = \frac{1}{N}\sum_{i=1}^{n}|x_i - m| = \frac{|x_1 - m| + |x_2 - m| + ... + |x_n - m|}{N}$$

Ecart absolu moyen

Définition de L'écart absolu moyen

L'écart absolu moyen est la moyenne arithmétique de la valeur absolue des écarts à la moyenne :

$$eam = \frac{1}{N}\sum_{i=1}^{n}|x_i - m| = \frac{|x_1 - m| + |x_2 - m| + ... + |x_n - m|}{N}$$

où $m = \frac{1}{N}\sum_{i=1}^{n} x_i$

Ecart absolu moyen

Définition de L'écart absolu moyen

L'écart absolu moyen est la moyenne arithmétique de la valeur absolue des écarts à la moyenne :

$$eam = \frac{1}{N}\sum_{i=1}^{n}|x_i - m| = \frac{|x_1 - m| + |x_2 - m| + \ldots + |x_n - m|}{N}$$

où $m = \frac{1}{N}\sum_{i=1}^{n} x_i$

Remarque : si on ne mettait pas les valeurs absolues, on trouverait que la moyenne des écarts est égale à 0.

Ecart absolu moyen

Définition de L'écart absolu moyen

L'écart absolu moyen est la moyenne arithmétique de la valeur absolue des écarts à la moyenne :

$$eam = \frac{1}{N}\sum_{i=1}^{n}|x_i - m| = \frac{|x_1 - m| + |x_2 - m| + \ldots + |x_n - m|}{N}$$

où $m = \dfrac{1}{N}\sum_{i=1}^{n} x_i$

Remarque : si on ne mettait pas les valeurs absolues, on trouverait que la moyenne des écarts est égale à 0.
C'est un résultat général : **la moyenne des écarts à la moyenne est nulle dont la démonstration**.

Ecart absolu moyen

Définition de L'écart absolu moyen

Si les données sont données sous cette forme :

Valeurs	x_1	x_2	...	x_k
Effectifs	n_1	n_2	...	n_k

Ecart absolu moyen

Définition de L'écart absolu moyen

Si les données sont données sous cette forme :

Valeurs	x_1	x_2	...	x_k
Effectifs	n_1	n_2	...	n_k

La formule mathématique s'écrit :

Ecart absolu moyen

Définition de L'écart absolu moyen

Si les données sont données sous cette forme :

Valeurs	x_1	x_2	...	x_k
Effectifs	n_1	n_2	...	n_k

La formule mathématique s'écrit :

$$eam = \frac{n_1 * |x_1 - m| + n_2 * |x_2 - m| + ... + n_k * |x_k - m|}{N}$$

où $N = n_1 + n_2 + ... + n_k$

r

Ecart absolu moyen

Démonstration

Si on mettait pas les valeurs absolues, on trouverait que la moyenne des écarts est égal à 0.

Ecart absolu moyen

Démonstration

Si on mettait pas les valeurs absolues, on trouverait que la moyenne des écarts est égal à 0.
Cela est un résultat général : la moyenne des écart à la moyenne est nulle :

Ecart absolu moyen

Démonstration

Si on mettait pas les valeurs absolues, on trouverait que la moyenne des écarts est égal à 0.
Cela est un résultat général : la moyenne des écart à la moyenne est nulle :

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x}) &= \frac{1}{N}\sum_{i=1}^{N}(x_i)-\frac{1}{N}\sum_{i=1}^{N}(\overline{x})\\ &= \overline{x}-\frac{1}{N}N\overline{x}\\ &= \overline{x}-\overline{x}\\ &= 0\end{aligned}$$

Ecart absolu moyen

Exemple

1.2	3.5	4	4.1	5.3	7.4	7.6	7.7	10	11.5
12.6	13.2	13.7	14.4	15.4	15.5	18	18.2	18.9	19.8

Ecart absolu moyen

Exemple

1.2	3.5	4	4.1	5.3	7.4	7.6	7.7	10	11.5
12.6	13.2	13.7	14.4	15.4	15.5	18	18.2	18.9	19.8

On commence par calculer la moyenne arithmétique des valeurs observées :

Ecart absolu moyen

Exemple

1.2	3.5	4	4.1	5.3	7.4	7.6	7.7	10	11.5
12.6	13.2	13.7	14.4	15.4	15.5	18	18.2	18.9	19.8

On commence par calculer la moyenne arithmétique des valeurs observées :

$$m = \frac{1}{20}(1.2 + 3.5 + ... + 19.8) = 11.1$$

Ecart absolu moyen

Exemple

1.2	3.5	4	4.1	5.3	7.4	7.6	7.7	10	11.5
12.6	13.2	13.7	14.4	15.4	15.5	18	18.2	18.9	19.8

On commence par calculer la moyenne arithmétique des valeurs observées :

$m = \frac{1}{20}(1.2 + 3.5 + ... + 19.8) = 11.1$

On calcule ensuite tous les écarts par rapport à la moyenne m = 11,1 : $|1.2 - 11.1| = 9.9$, $|3.5 - 11.1| = 7.6$, $|4 - 11.1| = 7.1$ etc...

On obtient les valeurs suivantes :

Ecart absolu moyen

Exemple

Valeur absolue des écarts à la moyenne									
9.9	7.6	7.1	7	5.8	3.7	3.5	3.4	1.1	0.4
1.5	2.1	2.6	3.3	4.3	4.4	6.9	7.1	7.8	8.7

Ecart absolu moyen

Exemple

Valeur absolue des écarts à la moyenne									
9.9	7.6	7.1	7	5.8	3.7	3.5	3.4	1.1	0.4
1.5	2.1	2.6	3.3	4.3	4.4	6.9	7.1	7.8	8.7

Il ne reste plus qu'à calculer la moyenne de ces écarts :

Ecart absolu moyen

Exemple

Valeur absolue des écarts à la moyenne									
9.9	7.6	7.1	7	5.8	3.7	3.5	3.4	1.1	0.4
1.5	2.1	2.6	3.3	4.3	4.4	6.9	7.1	7.8	8.7

Il ne reste plus qu'à calculer la moyenne de ces écarts :

$$eam = \frac{1}{20}(9.9 + 7.6 + ... + 8.7) = 4.91$$

Ecart absolu moyen

Exemple

Valeur absolue des écarts à la moyenne									
9.9	7.6	7.1	7	5.8	3.7	3.5	3.4	1.1	0.4
1.5	2.1	2.6	3.3	4.3	4.4	6.9	7.1	7.8	8.7

Il ne reste plus qu'à calculer la moyenne de ces écarts :

$$eam = \frac{1}{20}(9.9 + 7.6 + ... + 8.7) = 4.91$$

En moyenne, les données s'écartent d'environ 4.9 de la valeur centrale.

Ecart absolu moyen

Avantage de l'écart absolu moyen

L'écart absolu moyen est une quantité qui correspond très bien à l'intuition de ce qu'est une dispersion moyenne.

Ecart absolu moyen

Avantage de l'écart absolu moyen

L'écart absolu moyen est une quantité qui correspond très bien à l'intuition de ce qu'est une dispersion moyenne.
C'est une grandeur qui est toujours positive et qui est exprimée dans la même unité que la variable observée.

Ecart absolu moyen

Avantage de l'écart absolu moyen

L'écart absolu moyen est une quantité qui correspond très bien à l'intuition de ce qu'est une dispersion moyenne.
C'est une grandeur qui est toujours positive et qui est exprimée dans la même unité que la variable observée.
Elle est facile à calculer numériquement.

Ecart absolu moyen

Inconvénient de l'écart absolu moyen

Il a cependant le défaut d'être difficile à manipuler algébriquement.

Ecart absolu moyen

Inconvénient de l'écart absolu moyen

Il a cependant le défaut d'être difficile à manipuler algébriquement. En effet, les sommes de valeurs absolues ne se transforment pas bien dans les expressions algébriques.

Ecart absolu moyen

Inconvénient de l'écart absolu moyen

Il a cependant le défaut d'être difficile à manipuler algébriquement.
En effet, les sommes de valeurs absolues ne se transforment pas bien dans les expressions algébriques.
Par exemple, la valeur absolue d'une somme n'est pas la somme des valeurs absolues des termes de la somme.

Ecart absolu moyen

Inconvénient de l'écart absolu moyen

Il a cependant le défaut d'être difficile à manipuler algébriquement.
En effet, les sommes de valeurs absolues ne se transforment pas bien dans les expressions algébriques.
Par exemple, la valeur absolue d'une somme n'est pas la somme des valeurs absolues des termes de la somme.
Pour ces raisons, la quantité calculée est un bon indicateur de dispersion mais elle ne permet pas de développements théoriques.

Ecart absolu moyen

Inconvénient de l'écart absolu moyen

Il a cependant le défaut d'être difficile à manipuler algébriquement. En effet, les sommes de valeurs absolues ne se transforment pas bien dans les expressions algébriques.
Par exemple, la valeur absolue d'une somme n'est pas la somme des valeurs absolues des termes de la somme.
Pour ces raisons, la quantité calculée est un bon indicateur de dispersion mais elle ne permet pas de développements théoriques.
On lui préfère habituellement l'écart-type qui va être défini dans la section suivante.

Variance

Définition de la variance

Avec les valeurs absolues est d'ignorer dans quel sens se font les écarts par rapport à la valeur centrale (vers la gauche ou vers la droite, par défaut ou par excès).

Variance

Définition de la variance

Avec les valeurs absolues est d'ignorer dans quel sens se font les écarts par rapport à la valeur centrale (vers la gauche ou vers la droite, par défaut ou par excès).
Pour obtenir le même effet, on peut aussi élever ces écarts au carré.

Variance

Définition de la variance

Avec les valeurs absolues est d'ignorer dans quel sens se font les écarts par rapport à la valeur centrale (vers la gauche ou vers la droite, par défaut ou par excès).
Pour obtenir le même effet, on peut aussi élever ces écarts au carré.
On aboutit ainsi à la notion de variance d'une distribution.

Variance

Définition de la variance

Avec les valeurs absolues est d'ignorer dans quel sens se font les écarts par rapport à la valeur centrale (vers la gauche ou vers la droite, par défaut ou par excès).
Pour obtenir le même effet, on peut aussi élever ces écarts au carré.
On aboutit ainsi à la notion de variance d'une distribution.
Elle est définit comme la **moyenne des carrés des écarts à la moyenne.**

Variance

Définition de la variance

$$V(x) = \frac{1}{N}\sum_{i=1}^{N}(x_i - m)^2$$

$$V(x) = \frac{(x_1 - m)^2 + (x_2 - m)^2 + \ldots + (x_N - m)^2}{N}$$

Variance

Définition de la variance

$$V(x) = \frac{1}{N}\sum_{i=1}^{N}(x_i - m)^2$$

$$V(x) = \frac{(x_1 - m)^2 + (x_2 - m)^2 + ... + (x_N - m)^2}{N}$$

où $m = \overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ désigne la moyenne arithmétique.

Variance

Variance

Si les données sont représentées sous la forme du tableau suivant, alors la variance s'écrit :

Variance

Variance

Si les données sont représentées sous la forme du tableau suivant, alors la variance s'écrit :

Valeurs	v_1	v_2	...	v_k
Effectifs	n_1	n_2	...	n_k

Variance

Variance

Si les données sont représentées sous la forme du tableau suivant, alors la variance s'écrit :

Valeurs	v_1	v_2	...	v_k
Effectifs	n_1	n_2	...	n_k

$$V(x) = \frac{n_1 * (v_1 - m)^2 + n_2 * (v_2 - m)^2 + ... + n_k * (v_k - m)^2}{N}$$

Variance

Variance

Si les données sont représentées sous la forme du tableau suivant, alors la variance s'écrit :

Valeurs	v_1	v_2	...	v_k
Effectifs	n_1	n_2	...	n_k

$$V(x) = \frac{n_1 * (v_1 - m)^2 + n_2 * (v_2 - m)^2 + \ldots + n_k * (v_k - m)^2}{N}$$

où $m = \overline{x} = \frac{1}{N} \sum_{i=1}^{k} n_i v_i$ désigne la moyenne arithmétique.

Variance

Exemple

1.2	3.5	4	4.1	5.3	7.4	7.6	7.7	10	11.5
12.6	13.2	13.7	14.4	15.4	15.5	18	18.2	18.9	19.8

Variance

Exemple

1.2	3.5	4	4.1	5.3	7.4	7.6	7.7	10	11.5
12.6	13.2	13.7	14.4	15.4	15.5	18	18.2	18.9	19.8

La moyenne arithmétique des valeurs observées est $m = 11.1$. On calcule ensuite tous les écarts par rapport à la moyenne :
$(1.2 - 11.1)^2 = 9.9^2 = 98.01$; $(3.5 - 11.1)^2 = 7.6^2 = 57.76$ etc....
On obtient les valeurs suivantes :

Variance

Exemple

1.2	3.5	4	4.1	5.3	7.4	7.6	7.7	10	11.5
12.6	13.2	13.7	14.4	15.4	15.5	18	18.2	18.9	19.8

La moyenne arithmétique des valeurs observées est $m = 11.1$. On calcule ensuite tous les écarts par rapport à la moyenne :
$(1.2 - 11.1)^2 = 9.9^2 = 98.01$; $(3.5 - 11.1)^2 = 7.6^2 = 57.76$ etc....
On obtient les valeurs suivantes :

Carré des écarts à la moyenne									
98.0	57.7	50.4	49	33.6	13.6	12.2	11.5	1.2	0.1
2.2	4.4	6.7	10.8	18.4	19.3	47.6	50.4	60.8	75.6

Variance

Exemple

Il ne reste plus qu'à calculer la moyenne de ces carrés :

Variance

Exemple

Il ne reste plus qu'à calculer la moyenne de ces carrés :

$$\begin{aligned} V(x) &= \frac{1}{20}(98.01 + 57.76 + ... + 75.69) \\ &= 31.22 \end{aligned}$$

Variance

Variance

La variance est une quantité au carré. Cela signifie que si les valeurs x_i sont par exemple mesurées en mètres, alors la variance est en mètres carrés.

Variance

Variance

La variance est une quantité au carré. Cela signifie que si les valeurs x_i sont par exemple mesurées en mètres, alors la variance est en mètres carrés.
Or on s'attend à ce que la mesure de dispersion soit dans la même unité que les valeurs elles-mêmes.

Variance

Variance

La variance est une quantité au carré. Cela signifie que si les valeurs x_i sont par exemple mesurées en mètres, alors la variance est en mètres carrés.
Or on s'attend à ce que la mesure de dispersion soit dans la même unité que les valeurs elles-mêmes. C'est pourquoi on calcule la racine carrée de la variance comme nouvel indice de dispersion.

Variance

Définition de l'écart-type

On appelle écart-type la racine carrée de la variance.

Variance

Définition de l'écart-type

On appelle écart-type la racine carrée de la variance.
L'écart-type est souvent noté au moyen de la lettre grecque σ qui se lit sigma.

Variance

Définition de l'écart-type

On appelle écart-type la racine carrée de la variance.
L'écart-type est souvent noté au moyen de la lettre grecque σ qui se lit sigma.
On a les relations suivantes :

Variance

Définition de l'écart-type

On appelle écart-type la racine carrée de la variance.
L'écart-type est souvent noté au moyen de la lettre grecque σ qui se lit sigma.
On a les relations suivantes :

$$V(x) = \sigma(x)^2$$

Variance

Définition de l'écart-type

On appelle écart-type la racine carrée de la variance.
L'écart-type est souvent noté au moyen de la lettre grecque σ qui se lit sigma.
On a les relations suivantes :

$$V(x) = \sigma(x)^2$$

$$\sigma(x) = \sqrt{V(x)}$$

Variance

Définition de l'écart-type

On appelle écart-type la racine carrée de la variance.
L'écart-type est souvent noté au moyen de la lettre grecque σ qui se lit sigma.
On a les relations suivantes :

$$V(x) = \sigma(x)^2$$

$$\sigma(x) = \sqrt{V(x)}$$

Dans le cas précédent, on trouve $\sigma = 5.59$.

Variance

Variance dans le cas continu

Le tableau suivant donne la répartition des employés d'une entreprise selon le salaire mensuel. Définir l'écart-type.

	[1.5,2[	[2,2.5[	[2.5,3[	[3,3.5[	[3.5,4[	[4,5[	
n_i	31	34	40	26	13	6	**150**
v_i	1.75	2.25	2.75	3.25	3.75	4.5	
$n_i v_i$	54.25	76.50	110	84.50	48.75	27	**401**

Variance

Variance dans le cas continu

La moyenne est de 2.67 :

Variance

Variance dans le cas continu

La moyenne est de 2.67 :

$$\begin{aligned} m &= \frac{1}{150}(31 * 1.75 + ... + 6 * 4.5) \\ &= \frac{401}{150} \\ &= 2.67 \end{aligned}$$

Variance

Variance dans le cas continu

	]1.5,2]	]2,2.5]	]2.5,3]	]3,3.5]	]3.5,4]	]4,5]	
n_i	31	34	40	26	13	6	**150**
v_i	1.75	2.25	2.75	3.25	3.75	4.5	
$n_i v_i$	54.25	76.50	110	84.50	48.75	27	**401**
$v_i - m$	-0.92	-0.42	0.08	0.58	1.08	1.83	
$(v_i - m)^2$	0.846	0.176	0.006	0.34	1.167	3.35	
$n_i(v_i - m)^2$	26.24	5.998	0.255	8.746	15.163	20.093	**76.495**

Variance

Variance dans le cas continu

	]1.5,2]	]2,2.5]	]2.5,3]	]3,3.5]	]3.5,4]	]4,5]	
n_i	31	34	40	26	13	6	**150**
v_i	1.75	2.25	2.75	3.25	3.75	4.5	
$n_i v_i$	54.25	76.50	110	84.50	48.75	27	**401**
$v_i - m$	-0.92	-0.42	0.08	0.58	1.08	1.83	
$(v_i - m)^2$	0.846	0.176	0.006	0.34	1.167	3.35	
$n_i(v_i - m)^2$	26.24	5.998	0.255	8.746	15.163	20.093	**76.495**

La variance est :

Variance

Variance dans le cas continu

	]1.5,2]	]2,2.5]	]2.5,3]	]3,3.5]	]3.5,4]	]4,5]	
n_i	31	34	40	26	13	6	**150**
v_i	1.75	2.25	2.75	3.25	3.75	4.5	
$n_i v_i$	54.25	76.50	110	84.50	48.75	27	**401**
$v_i - m$	-0.92	-0.42	0.08	0.58	1.08	1.83	
$(v_i - m)^2$	0.846	0.176	0.006	0.34	1.167	3.35	
$n_i(v_i - m)^2$	26.24	5.998	0.255	8.746	15.163	20.093	**76.495**

La variance est :

$$V(x) = \frac{76.495}{150}$$
$$= 0.51$$

Variance

Variance dans le cas continu

	]1.5,2]	]2,2.5]	]2.5,3]	]3,3.5]	]3.5,4]	]4,5]	
n_i	31	34	40	26	13	6	**150**
v_i	1.75	2.25	2.75	3.25	3.75	4.5	
$n_i v_i$	54.25	76.50	110	84.50	48.75	27	**401**
$v_i - m$	-0.92	-0.42	0.08	0.58	1.08	1.83	
$(v_i - m)^2$	0.846	0.176	0.006	0.34	1.167	3.35	
$n_i(v_i - m)^2$	26.24	5.998	0.255	8.746	15.163	20.093	**76.495**

La variance est :

$$V(x) = \frac{76.495}{150} = 0.51$$

Et l'écart-type $=0.714$.

Variance

Propriétés de la variance I

Il existe une autre formule (dite formule développée) pour calculer la variance d'une distribution :

Variance

Propriétés de la variance I

Il existe une autre formule (dite formule développée) pour calculer la variance d'une distribution :

$$V(x) = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \overline{x}^2$$

$$V(x) = \frac{x_1^2 + x_2^2 + \ldots + x_N^2}{N} - \overline{x}^2$$

Variance

Propriétés de la variance I

la variance est égale à la moyenne des carrés diminuée du carré de la moyenne.

Variance

Propriétés de la variance I

la variance est égale à la moyenne des carrés diminuée du carré de la moyenne.

Si les données sont regroupées sous forme d'effectifs alors la formule développée devient :

Variance

Propriétés de la variance I

la variance est égale à la moyenne des carrés diminuée du carré de la moyenne.

Si les données sont regroupées sous forme d'effectifs alors la formule développée devient :

$$V(x) = \frac{n_1 v_1^2 + n_1 v_2^2 + ... + n_1 v_N^2}{N} - \overline{x}^2$$

$$V(x) = \frac{1}{N}\sum_{i=1}^{N} n_i v_i^2 - \overline{x}^2$$

Variance

Propriétés de la variance II

$$V(x+b) = V(x)$$

Variance

Propriétés de la variance II

$$V(x+b) = V(x)$$

Propriétés de la variance III

$$V(ax) = a^2 V(x)$$

Variance

Propriétés de la variance II

$$V(x+b) = V(x)$$

Propriétés de la variance III

$$V(ax) = a^2 V(x)$$

Propriétés de la variance IV

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

Déviation médiane absolue

Définition déviation médiane absolue

Une autre mesure de dispersion, qui ne tient compte que de la position des observations et non pas de leurs valeurs, consiste à remplacer la notion de moyenne par celle de médiane dans la définition de l'écart absolu moyen.

Déviation médiane absolue

Définition déviation médiane absolue

Une autre mesure de dispersion, qui ne tient compte que de la position des observations et non pas de leurs valeurs, consiste à remplacer la notion de moyenne par celle de médiane dans la définition de l'écart absolu moyen.
Au lieu de centrer les valeurs sur la moyenne, on les centre sur la médiane.

Déviation médiane absolue

Définition déviation médiane absolue

Une autre mesure de dispersion, qui ne tient compte que de la position des observations et non pas de leurs valeurs, consiste à remplacer la notion de moyenne par celle de médiane dans la définition de l'écart absolu moyen.
Au lieu de centrer les valeurs sur la moyenne, on les centre sur la médiane.
On obtient ainsi les écarts à la médiane, calculés en valeur absolue.

Déviation médiane absolue

Définition déviation médiane absolue

Une autre mesure de dispersion, qui ne tient compte que de la position des observations et non pas de leurs valeurs, consiste à remplacer la notion de moyenne par celle de médiane dans la définition de l'écart absolu moyen.
Au lieu de centrer les valeurs sur la moyenne, on les centre sur la médiane.
On obtient ainsi les écarts à la médiane, calculés en valeur absolue.
Ensuite, au lieu de faire la moyenne de ces écarts, on en prend la médiane.

Déviation médiane absolue

Définition déviation médiane absolue

Une autre mesure de dispersion, qui ne tient compte que de la position des observations et non pas de leurs valeurs, consiste à remplacer la notion de moyenne par celle de médiane dans la définition de l'écart absolu moyen.
Au lieu de centrer les valeurs sur la moyenne, on les centre sur la médiane.
On obtient ainsi les écarts à la médiane, calculés en valeur absolue.
Ensuite, au lieu de faire la moyenne de ces écarts, on en prend la médiane.
La quantité qu'on obtient de cette manière s'appelle la déviation médiane absolue.

Déviation médiane absolue

Application

1.2	3.5	4	4.1	5.3	7.4	7.6	7.7	10	11.5
12.6	13.2	13.7	14.4	15.4	15.5	18	18.2	18.9	19.8

La médiane de cette distribution est $M = (11.5 + 12.6)/2 = 12.05$.

Déviation médiane absolue

Application

1.2	3.5	4	4.1	5.3	7.4	7.6	7.7	10	11.5
12.6	13.2	13.7	14.4	15.4	15.5	18	18.2	18.9	19.8

La médiane de cette distribution est $M = (11.5 + 12.6)/2 = 12.05$.
On calcule donc les valeurs absolues des écarts à la médiane :
$|1.2 - 12.05| = 10.85$, $|3.5 - 12.05| = 8.55$, etc.

Déviation médiane absolue

Application

1.2	3.5	4	4.1	5.3	7.4	7.6	7.7	10	11.5
12.6	13.2	13.7	14.4	15.4	15.5	18	18.2	18.9	19.8

La médiane de cette distribution est $M = (11.5 + 12.6)/2 = 12.05$.
On calcule donc les valeurs absolues des écarts à la médiane :
$|1.2 - 12.05| = 10.85$, $|3.5 - 12.05| = 8.55$, etc.

les résultats :

10.85	8.55	8.05	7.95	6.75	4.65	4.45	4.35	2.05	0.55
0.55	1.15	1.65	2.35	3.35	3.45	5.95	6.15	6.85	7.75

Déviation médiane absolue

Application

1.2	3.5	4	4.1	5.3	7.4	7.6	7.7	10	11.5
12.6	13.2	13.7	14.4	15.4	15.5	18	18.2	18.9	19.8

La médiane de cette distribution est $M = (11.5 + 12.6)/2 = 12.05$.
On calcule donc les valeurs absolues des écarts à la médiane :
$|1.2 - 12.05| = 10.85$, $|3.5 - 12.05| = 8.55$, etc.

les résultats :

10.85	8.55	8.05	7.95	6.75	4.65	4.45	4.35	2.05	0.55
0.55	1.15	1.65	2.35	3.35	3.45	5.95	6.15	6.85	7.75

On ordonne ces quantités pour calculer leur médiane :

Déviation médiane absolue

Application

1.2	3.5	4	4.1	5.3	7.4	7.6	7.7	10	11.5
12.6	13.2	13.7	14.4	15.4	15.5	18	18.2	18.9	19.8

La médiane de cette distribution est $M = (11.5 + 12.6)/2 = 12.05$.
On calcule donc les valeurs absolues des écarts à la médiane :
$|1.2 - 12.05| = 10.85$, $|3.5 - 12.05| = 8.55$, etc.

les résultats :

10.85	8.55	8.05	7.95	6.75	4.65	4.45	4.35	2.05	0.55
0.55	1.15	1.65	2.35	3.35	3.45	5.95	6.15	6.85	7.75

On ordonne ces quantités pour calculer leur médiane :

0.55	0.55	1.15	1.65	2.05	2.35	3.35	3.45	4.35	4.45
4.65	5.95	6.15	6.75	6.85	7.75	7.95	8.05	8.55	10.85

Déviation médiane absolue

Application

La médiane vaut MAD $= (4.45+4.65)/2 = 4.55$. C'est la déviation médiane absolue.

Déviation médiane absolue

Application

La médiane vaut MAD $= (4.45+4.65)/2 = 4.55$. C'est la déviation médiane absolue.

La déviation médiane absolue est un excellent indicateur de dispersion pour plusieurs raisons :

Déviation médiane absolue

Application

La médiane vaut MAD $= (4.45+4.65)/2 = 4.55$. C'est la déviation médiane absolue.

La déviation médiane absolue est un excellent indicateur de dispersion pour plusieurs raisons :1. il s'agit d'une statistique robuste car elle est résistante à la présence de points aberrants (voir au paragraphe suivant) ;2. elle fournit un estimateur consistent de l'écart-type (ces questions seront étudiées dans le cours de statistique inférentielle) ;3. c'est toujours une quantité finie et on peut donc l'utiliser dans le cas de distributions qui n'ont pas de moyenne et de variance (ces questions seront étudiées dans le cours de probabilité).

Coefficient de variation

Coefficient de variation

Les principaux indicateurs d'une distribution, en particulier la moyenne arithmétique $\overline{x}$ et l'écart-type $\sigma(x)$, sont exprimés dans la même unité que la variable observée x.

Coefficient de variation

Coefficient de variation

Les principaux indicateurs d'une distribution, en particulier la moyenne arithmétique $\overline{x}$ et l'écart-type $\sigma(x)$, sont exprimés dans la même unité que la variable observée x.
Cela soulève un problème lorsqu'on veut comparer deux caractères qui sont exprimés chacun dans son unité : par exemple, les salaires en France et aux États-Unis sont respectivement exprimés en euros et en dollars.

Coefficient de variation

Coefficient de variation

Les principaux indicateurs d'une distribution, en particulier la moyenne arithmétique $\overline{x}$ et l'écart-type $\sigma(x)$, sont exprimés dans la même unité que la variable observée x.
Cela soulève un problème lorsqu'on veut comparer deux caractères qui sont exprimés chacun dans son unité : par exemple, les salaires en France et aux États-Unis sont respectivement exprimés en euros et en dollars.
Dans une même unité, des grandeurs économiques peuvent être sur des échelles très différentes : par exemple, la production d'acier en Europe et en Chine.

Coefficient de variation

Coefficient de variation

Les principaux indicateurs d'une distribution, en particulier la moyenne arithmétique $\overline{x}$ et l'écart-type $\sigma(x)$, sont exprimés dans la même unité que la variable observée x.
Cela soulève un problème lorsqu'on veut comparer deux caractères qui sont exprimés chacun dans son unité : par exemple, les salaires en France et aux États-Unis sont respectivement exprimés en euros et en dollars.
Dans une même unité, des grandeurs économiques peuvent être sur des échelles très différentes : par exemple, la production d'acier en Europe et en Chine.
Pour ces raisons, on définit le coefficient de variation comme le rapport entre l'écart-type et la moyenne :

Coefficient de variation

Coefficient de variation

$$C = \frac{\sigma(x)}{\overline{x}}$$

Coefficient de variation

Coefficient de variation

$$C = \frac{\sigma(x)}{\overline{x}}$$

Le coefficient de variation est une grandeur sans unité qu'on exprime souvent en pourcentage.

Coefficient de variation

Coefficient de variation

$$C = \frac{\sigma(x)}{\overline{x}}$$

Le coefficient de variation est une grandeur sans unité qu'on exprime souvent en pourcentage.
Il s'agit d'une mesure de la dispersion relative par rapport à une moyenne commune.

Coefficient de variation

Coefficient de variation

$$C = \frac{\sigma(x)}{\overline{x}}$$

Le coefficient de variation est une grandeur sans unité qu'on exprime souvent en pourcentage.
Il s'agit d'une mesure de la dispersion relative par rapport à une moyenne commune.
Il n'est bien sûr pas défini si la moyenne $\overline{x}$ est nulle mais on l'utilise principalement pour des grandeurs économiques positives.

Coefficient de variation

Application

On veut comparer les salaires dans deux entreprises (l'une française, l'autre américaine) pour lesquelles on a les données suivantes :

Coefficient de variation

Application

On veut comparer les salaires dans deux entreprises (l'une française, l'autre américaine) pour lesquelles on a les données suivantes :

	France	USA
moyenne	1500	2000
écart-type	45	50

Coefficient de variation

Application

On veut comparer les salaires dans deux entreprises (l'une française, l'autre américaine) pour lesquelles on a les données suivantes :

	France	USA
moyenne	1500	2000
écart-type	45	50

Les coefficients de variation sont respectivement de $Cv1 = 45/1500 = 0,03 = 3\%$ et $Cv2 = 50/2000 = 0,025 =$ 2,5%, ce qui indique une dispersion relative moindre dans le cas des États-Unis.

Robustesse des indicateurs

Robustesse des indicateurs

Il y a parfois, dans une distribution statistique, des valeurs qui s'écartent significativement des autres.

Robustesse des indicateurs

Robustesse des indicateurs

Il y a parfois, dans une distribution statistique, des valeurs qui s'écartent significativement des autres.
Ces valeurs peuvent être de plusieurs natures différentes : des observations qui résultent d'une erreur de mesure, des observations qui sont justes et dénotent un phénomène exceptionnel.

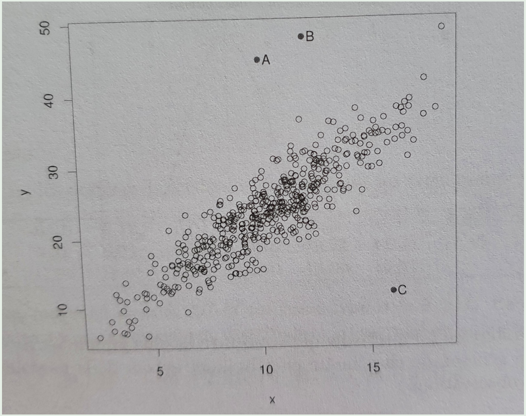
Robustesse des indicateurs

Robustesse des indicateurs

Il y a parfois, dans une distribution statistique, des valeurs qui s'écartent significativement des autres.
Ces valeurs peuvent être de plusieurs natures différentes : des observations qui résultent d'une erreur de mesure, des observations qui sont justes et dénotent un phénomène exceptionnel.
On les qualifie de points aberrants (outliers).

Robustesse des indicateurs

Robustesse des indicateurs



Intervalle de confiance

Intervalle de confiance

On observe, sur le graphique comportant des points aberrants (en dimension 1), que l'essentiel de la distribution est concentrée dans un intervalle de rayon 2σ (où σ est l'écart-type) autour de la moyenne.

Intervalle de confiance

Intervalle de confiance

On observe, sur le graphique comportant des points aberrants (en dimension 1), que l'essentiel de la distribution est concentrée dans un intervalle de rayon 2σ (où σ est l'écart-type) autour de la moyenne.
Ce phénomène n'est pas exceptionnel et correspond à la notion de distribution "normale" qui sera expliquée ultérieurement.

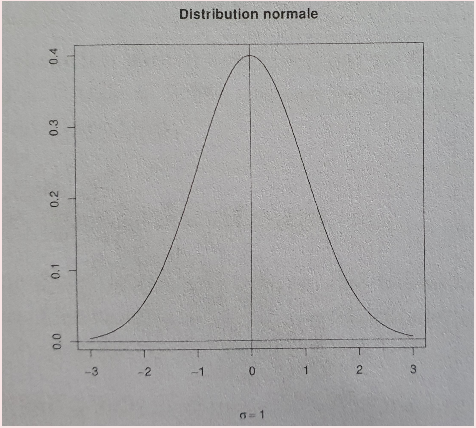
Intervalle de confiance

Intervalle de confiance

On observe, sur le graphique comportant des points aberrants (en dimension 1), que l'essentiel de la distribution est concentrée dans un intervalle de rayon 2σ (où σ est l'écart-type) autour de la moyenne.
Ce phénomène n'est pas exceptionnel et correspond à la notion de distribution "normale" qui sera expliquée ultérieurement.
De nombreuses distributions ont des densités (ou des histogrammes) qui ressemblent à une courbe en cloche (comme sur le graphique ci-dessous) :

Intervalle de confiance

Intervalle de confiance



Intervalle de confiance

Intervalle de confiance

On s'intéresse à l'intervalle qui concentre la plus grande partie des observations.

Intervalle de confiance

Intervalle de confiance

On s'intéresse à l'intervalle qui concentre la plus grande partie des observations.
En général, on cherche un encadrement autour de la moyenne contenant 95% des observations :

Intervalle de confiance

Intervalle de confiance

On s'intéresse à l'intervalle qui concentre la plus grande partie des observations.
En général, on cherche un encadrement autour de la moyenne contenant 95% des observations :

$$(m - u \leq X \leq m + u) = 0.95$$

L'intervalle $[m - u, m + u]$ s'appelle intervalle de confiance à 95%.

Intervalle de confiance

Intervalle de confiance

On s'intéresse à l'intervalle qui concentre la plus grande partie des observations.
En général, on cherche un encadrement autour de la moyenne contenant 95% des observations :

$$(m - u \leq X \leq m + u) = 0.95$$

L'intervalle $[m - u, m + u]$ s'appelle intervalle de confiance à 95%.
On montre dans le cours de probabilités que, pour une distribution issue d'une loi normale, la valeur du rayon u de l'intervalle de confiance est égale à $1,96 * \sigma$ où σ est l'écart-type de la distribution.

Intervalle de confiance

Intervalle de confiance

On s'intéresse à l'intervalle qui concentre la plus grande partie des observations.
En général, on cherche un encadrement autour de la moyenne contenant 95% des observations :

$$(m - u \leq X \leq m + u) = 0.95$$

L'intervalle $[m - u, m + u]$ s'appelle intervalle de confiance à 95%.
On montre dans le cours de probabilités que, pour une distribution issue d'une loi normale, la valeur du rayon u de l'intervalle de confiance est égale à $1,96 * \sigma$ où σ est l'écart-type de la distribution.
Parfois, ce rayon est arrondi à 2σ.

Intervalle de confiance

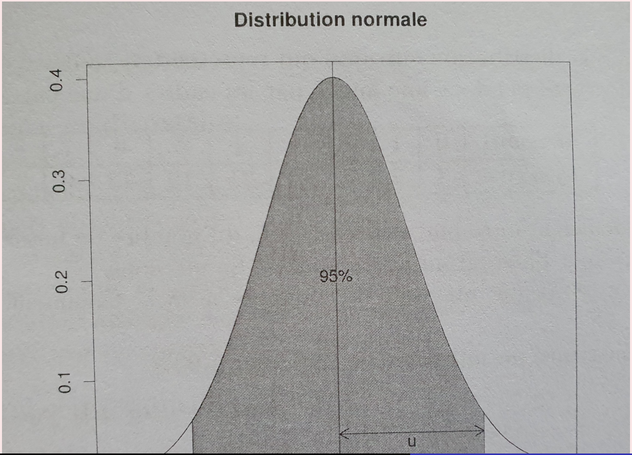
Intervalle de confiance

L'intervalle de confiance est donc

$$[m - 1,96 * \sigma ; m + 1,96 * \sigma]$$

Intervalle de confiance

Intervalle de confiance



Intervalle de confiance

Intervalle de confiance

Soit le tableau de données suivant :

Intervalle de confiance

Intervalle de confiance

Soit le tableau de données suivant :

5.46	6.01	6.12	7.70	7.78	8.12	8.39	8.87	9.17
9.29	9.31	9.71	9.73	10.75	10.93	11.30	11.46	11.51
11.58	11.77	11.95	12.04	12.80	13.64	14.54		

On a une moyenne $m \simeq 9,997$ et un écart-type $\sigma \simeq 2,314$.

Intervalle de confiance

Intervalle de confiance

Soit le tableau de données suivant :

5.46	6.01	6.12	7.70	7.78	8.12	8.39	8.87	9.17
9.29	9.31	9.71	9.73	10.75	10.93	11.30	11.46	11.51
11.58	11.77	11.95	12.04	12.80	13.64	14.54		

On a une moyenne $m \simeq 9,997$ et un écart-type $\sigma \simeq 2,314$.
En supposant que ces données sont normalement distribuées, l'intervalle de confiance est :

Intervalle de confiance

Intervalle de confiance

Soit le tableau de données suivant :

5.46	6.01	6.12	7.70	7.78	8.12	8.39	8.87	9.17
9.29	9.31	9.71	9.73	10.75	10.93	11.30	11.46	11.51
11.58	11.77	11.95	12.04	12.80	13.64	14.54		

On a une moyenne $m \simeq 9,997$ et un écart-type $\sigma \simeq 2,314$.
En supposant que ces données sont normalement distribuées, l'intervalle de confiance est :

$$[9,997 - 1,96 * 2,314; 9,997 + 1,96 * 2,314]$$

Intervalle de confiance

Intervalle de confiance

Soit le tableau de données suivant :

5.46	6.01	6.12	7.70	7.78	8.12	8.39	8.87	9.17
9.29	9.31	9.71	9.73	10.75	10.93	11.30	11.46	11.51
11.58	11.77	11.95	12.04	12.80	13.64	14.54		

On a une moyenne $m \simeq 9,997$ et un écart-type $\sigma \simeq 2,314$.
En supposant que ces données sont normalement distribuées, l'intervalle de confiance est :

$$[9,997 - 1,96 * 2,314 ; 9,997 + 1,96 * 2,314]$$

c'est-à-dire $[5,46; 14,53]$

Introduction

Introduction

Les indicateurs de forme permettent de décrire et de mesurer les caractéristiques de la courbe de fréquences d'une distribution.

Introduction

Introduction

Les indicateurs de forme permettent de décrire et de mesurer les caractéristiques de la courbe de fréquences d'une distribution. L'allure de cette courbe renseigne sur la manière dont les densités sont réparties.

Introduction

Introduction

Les indicateurs de forme permettent de décrire et de mesurer les caractéristiques de la courbe de fréquences d'une distribution.
L'allure de cette courbe renseigne sur la manière dont les densités sont réparties.
Les caractéristiques principales sont le degré d'asymétrie et le degré d'aplatissement.

Introduction

Introduction

Les indicateurs de forme permettent de décrire et de mesurer les caractéristiques de la courbe de fréquences d'une distribution.
L'allure de cette courbe renseigne sur la manière dont les densités sont réparties.
Les caractéristiques principales sont le degré d'asymétrie et le degré d'aplatissement.
Les indicateurs de concentration concernent la manière dont sont réparties les masses par rapport aux effectifs.

Indicateurs de forme : moments d'ordre p

Moments d'ordre p

Les moments sont des quantités qui étendent la notion de moyenne et celle de variance.

Indicateurs de forme : moments d'ordre p

Moments d'ordre p

Les moments sont des quantités qui étendent la notion de moyenne et celle de variance.
La moyenne est essentiellement une quantité linéaire (c'est-à-dire de degré 1) et la variance une quantité quadratique (c'est-à-dire de degré 2).

Indicateurs de forme : moments d'ordre p

Moments d'ordre p

Les moments sont des quantités qui étendent la notion de moyenne et celle de variance.
La moyenne est essentiellement une quantité linéaire (c'est-à-dire de degré 1) et la variance une quantité quadratique (c'est-à-dire de degré 2).
Les moments constituent une généralisation aux degrés supérieurs.

Indicateurs de forme : moments d'ordre p

Moments d'ordre p

Les moments sont des quantités qui étendent la notion de moyenne et celle de variance.
La moyenne est essentiellement une quantité linéaire (c'est-à-dire de degré 1) et la variance une quantité quadratique (c'est-à-dire de degré 2).
Les moments constituent une généralisation aux degrés supérieurs.
Ils ont une grande importance théorique en probabilités et fournissent des informations très utiles dans l'exploration des données statistiques.

Indicateurs de forme : moments d'ordre p

Moments d'ordre p

Les moments sont des quantités qui étendent la notion de moyenne et celle de variance.
La moyenne est essentiellement une quantité linéaire (c'est-à-dire de degré 1) et la variance une quantité quadratique (c'est-à-dire de degré 2).
Les moments constituent une généralisation aux degrés supérieurs.
Ils ont une grande importance théorique en probabilités et fournissent des informations très utiles dans l'exploration des données statistiques.
On distingue :

Indicateurs de forme : moments d'ordre p

Moments d'ordre p

Les moments sont des quantités qui étendent la notion de moyenne et celle de variance.
La moyenne est essentiellement une quantité linéaire (c'est-à-dire de degré 1) et la variance une quantité quadratique (c'est-à-dire de degré 2).
Les moments constituent une généralisation aux degrés supérieurs.
Ils ont une grande importance théorique en probabilités et fournissent des informations très utiles dans l'exploration des données statistiques.
On distingue :

1. les moments simples

Indicateurs de forme : moments d'ordre p

Moments d'ordre p

Les moments sont des quantités qui étendent la notion de moyenne et celle de variance.
La moyenne est essentiellement une quantité linéaire (c'est-à-dire de degré 1) et la variance une quantité quadratique (c'est-à-dire de degré 2).
Les moments constituent une généralisation aux degrés supérieurs.
Ils ont une grande importance théorique en probabilités et fournissent des informations très utiles dans l'exploration des données statistiques.
On distingue :

1. les moments simples
2. les moments centrés

Indicateurs de forme : moments d'ordre p

Définition : moments simples

On les définit pour un ordre particulier p où p est un nombre entier positif.

Indicateurs de forme : moments d'ordre p

Définition : moments simples

On les définit pour un ordre particulier p où p est un nombre entier positif.
Les moments simples d'ordre p correspondent à une moyenne des puissances p.

Indicateurs de forme : moments d'ordre p

Définition : moments simples

On les définit pour un ordre particulier p où p est un nombre entier positif.
Les moments simples d'ordre p correspondent à une moyenne des puissances p.
Le moment simple d'ordre p d'une variable statistique x est la moyenne (arithmétique !) des puissances p-ièmes des valeurs observées.

Indicateurs de forme : moments d'ordre p

Définition : moments simples

On les définit pour un ordre particulier p où p est un nombre entier positif.

Les moments simples d'ordre p correspondent à une moyenne des puissances p.

Le moment simple d'ordre p d'une variable statistique x est la moyenne (arithmétique !) des puissances p-ièmes des valeurs observées.

Si les données sont écrites sous forme exhaustive, la formule mathématique du moment simple d'ordre p est :

Indicateurs de forme : moments d'ordre p

Définition : moments simples

On les définit pour un ordre particulier p où p est un nombre entier positif.
Les moments simples d'ordre p correspondent à une moyenne des puissances p.
Le moment simple d'ordre p d'une variable statistique x est la moyenne (arithmétique !) des puissances p-ièmes des valeurs observées.
Si les données sont écrites sous forme exhaustive, la formule mathématique du moment simple d'ordre p est :

$$M_p = \frac{1}{N} \sum_{i=1}^{N} x_i^p$$

Indicateurs de forme : moments d'ordre p

Définition : moments simples

On les définit pour un ordre particulier p où p est un nombre entier positif.
Les moments simples d'ordre p correspondent à une moyenne des puissances p.
Le moment simple d'ordre p d'une variable statistique x est la moyenne (arithmétique !) des puissances p-ièmes des valeurs observées.
Si les données sont écrites sous forme exhaustive, la formule mathématique du moment simple d'ordre p est :

$$M_p = \frac{1}{N}\sum_{i=1}^{N} x_i^p$$

Indicateurs de forme : moments d'ordre p

Définition : moments simples

Si les données sont données sous cette forme :

Indicateurs de forme : moments d'ordre p

Définition : moments simples

Si les données sont données sous cette forme :

Valeurs	x_1	x_2	...	x_k
Effectifs	n_1	n_2	...	n_k

Indicateurs de forme : moments d'ordre p

Définition : moments simples

La formule s'écrit :

$$M_p = \frac{n_1 x_1^p + n_2 x_2^p + \ldots + n_K x_K^p}{N}$$

$$= \frac{1}{N} \sum_{i=1}^{k} n_i x_i^p$$

Indicateurs de forme : moments d'ordre p

moments simples

Les moments d'ordre p sont exprimés dans l'unité des données élevée à la puissance p : par exemple, si les x sont des quantités en mètres, le moment d'ordre 3 sera en mètres cubes.

Indicateurs de forme : moments d'ordre p

moments simples

Les moments d'ordre p sont exprimés dans l'unité des données élevée à la puissance p : par exemple, si les x sont des quantités en mètres, le moment d'ordre 3 sera en mètres cubes.
Remarque : les moments simples d'ordre p ne doivent donc pas être confondus avec des moyennes d'ordre p : dans ces dernières, on prend la puissance $1/p$ du tout pour se retrouver dans la même unité que les données.

Indicateurs de forme : moments d'ordre p

moments simples

Les moments d'ordre p sont exprimés dans l'unité des données élevée à la puissance p : par exemple, si les x sont des quantités en mètres, le moment d'ordre 3 sera en mètres cubes.
Remarque : les moments simples d'ordre p ne doivent donc pas être confondus avec des moyennes d'ordre p : dans ces dernières, on prend la puissance $1/p$ du tout pour se retrouver dans la même unité que les données.
Dans le cas particulier où $p = 1$, on retrouve la moyenne arithmétique.

Indicateurs de forme : moments d'ordre p

moments simples

Les moments d'ordre p sont exprimés dans l'unité des données élevée à la puissance p : par exemple, si les x sont des quantités en mètres, le moment d'ordre 3 sera en mètres cubes.
Remarque : les moments simples d'ordre p ne doivent donc pas être confondus avec des moyennes d'ordre p : dans ces dernières, on prend la puissance $1/p$ du tout pour se retrouver dans la même unité que les données.
Dans le cas particulier où $p = 1$, on retrouve la moyenne arithmétique.
Donc $M_1 = \overline{x}$.

Indicateurs de forme : moments d'ordre p

moments simples

Les moments d'ordre p sont exprimés dans l'unité des données élevée à la puissance p : par exemple, si les x sont des quantités en mètres, le moment d'ordre 3 sera en mètres cubes.
Remarque : les moments simples d'ordre p ne doivent donc pas être confondus avec des moyennes d'ordre p : dans ces dernières, on prend la puissance $1/p$ du tout pour se retrouver dans la même unité que les données.
Dans le cas particulier où $p = 1$, on retrouve la moyenne arithmétique.
Donc $M_1 = \overline{x}$.
Dans le cas particulier où $p = 2$, on obtient le carré de la moyenne quadratique. Donc $M_2 = m_2^2$.

Indicateurs de forme : moments d'ordre p

moments simples

Donc $M_2 = m_2^2$.

Indicateurs de forme : moments d'ordre p

moments simples

Donc $M_2 = m_2^2$.
La formule développée de la variance (moyenne des carrés moins carré de la moyenne) peut s'écrire avec les moments de la manière suivante :

Indicateurs de forme : moments d'ordre p

moments simples

Donc $M_2 = m_2^2$.
La formule développée de la variance (moyenne des carrés moins carré de la moyenne) peut s'écrire avec les moments de la manière suivante :
$V(x) = M_2 - M_1^2$

Indicateurs de forme : moments d'ordre p

moments centrées

Les moments centrés sont les moments simples appliqués aux écarts par rapport à la moyenne.

Indicateurs de forme : moments d'ordre p

moments centrées

Les moments centrés sont les moments simples appliqués aux écarts par rapport à la moyenne.
Autrement dit, on remplace les valeurs x_i par $x_i - \overline{x}$ dans les formules précédentes. On les note au moyen de la lettre grecque μ.
Les formules mathématiques sont donc (selon que les données sont exhaustives ou regroupées) :

Indicateurs de forme : moments d'ordre p

moments centrées

Les moments centrés sont les moments simples appliqués aux écarts par rapport à la moyenne.
Autrement dit, on remplace les valeurs x_i par $x_i - \overline{x}$ dans les formules précédentes. On les note au moyen de la lettre grecque μ.
Les formules mathématiques sont donc (selon que les données sont exhaustives ou regroupées) :

$$\mu_p = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^p$$

Indicateurs de forme : moments d'ordre p

moments centrées

Les moments centrés sont les moments simples appliqués aux écarts par rapport à la moyenne.
Autrement dit, on remplace les valeurs x_i par $x_i - \overline{x}$ dans les formules précédentes. On les note au moyen de la lettre grecque μ.
Les formules mathématiques sont donc (selon que les données sont exhaustives ou regroupées) :

$$\mu_p = \frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^p$$

Indicateurs de forme : moments d'ordre p

moments centrées

Ou encore selon la forme des données :

Indicateurs de forme : moments d'ordre p

moments centrées

Ou encore selon la forme des données :

$$\mu_p = \frac{1}{N}\sum_{i=1}^{k} n_i(x_i - \overline{x})^p$$

Indicateurs de forme : moments d'ordre p

démonstration moments centrés d'ordre 1

$$\mu_1 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})$$

Indicateurs de forme : moments d'ordre p

démonstration moments centrés d'ordre 1

$$\mu_1 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})$$

$$= \frac{1}{N}\sum_{i=1}^{N}x_i - \frac{1}{N}\sum_{i=1}^{N}\overline{x}$$

Indicateurs de forme : moments d'ordre p

démonstration moments centrés d'ordre 1

$$\mu_1 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})$$

$$= \frac{1}{N}\sum_{i=1}^{N}x_i - \frac{1}{N}\sum_{i=1}^{N}\overline{x}$$

$$\overline{x} - \frac{1}{N} * N * \overline{x}$$

Indicateurs de forme : moments d'ordre p

démonstration moments centrés d'ordre 1

$$\mu_1 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})$$

$$= \frac{1}{N}\sum_{i=1}^{N}x_i - \frac{1}{N}\sum_{i=1}^{N}\overline{x}$$

$$\overline{x} - \frac{1}{N} * N * \overline{x}$$

$$\overline{x} - \overline{x} = 0$$

Indicateurs de forme : moments d'ordre p

démonstration moments centrés d'ordre 1

On interprète ce résultat en disant que les écarts à gauche de la moyenne (écarts par défaut) compensent exactement les écarts à droite (écarts par excès).

Indicateurs de forme : moments d'ordre p

démonstration moments centrés d'ordre 2

$$\mu_2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2$$

Indicateurs de forme : moments d'ordre p

démonstration moments centrés d'ordre 2

$$\mu_2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2$$

On interprète ce résultat en disant que les écarts à gauche de la moyenne (écarts par défaut) compensent exactement les écarts à droite (écarts par excès).

Indicateurs de forme : moments d'ordre p

démonstration moments centrés d'ordre 2

$$\mu_2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2$$

On interprète ce résultat en disant que les écarts à gauche de la moyenne (écarts par défaut) compensent exactement les écarts à droite (écarts par excès).
Donc le moment centré d'ordre 2 n'est autre que la variance !

Indicateurs de forme : moments d'ordre p

démonstration moments centrés d'ordre 2

$$\mu_2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2$$

On interprète ce résultat en disant que les écarts à gauche de la moyenne (écarts par défaut) compensent exactement les écarts à droite (écarts par excès).
Donc le moment centré d'ordre 2 n'est autre que la variance !
On a donc : $\mu_2 = V(x)$.

Comparaison des indicateurs de tendance centrale

Comparaison des indicateurs de tendance centrale

On a vu trois indicateurs de tendance centrale : le mode, la médiane et la moyenne.

Comparaison des indicateurs de tendance centrale

Comparaison des indicateurs de tendance centrale

On a vu trois indicateurs de tendance centrale : le mode, la médiane et la moyenne.
La comparaison de ces indicateurs entre eux donne des renseignements sur la façon dont les données observées sont réparties.

Comparaison des indicateurs de tendance centrale

Comparaison des indicateurs de tendance centrale

On a vu trois indicateurs de tendance centrale : le mode, la médiane et la moyenne.
La comparaison de ces indicateurs entre eux donne des renseignements sur la façon dont les données observées sont réparties.
Dans une distribution parfaitement symétrique et concentrée autour de sa valeur centrale, les trois indicateurs coïncident :

Comparaison des indicateurs de tendance centrale

Comparaison des indicateurs de tendance centrale

On a vu trois indicateurs de tendance centrale : le mode, la médiane et la moyenne.
La comparaison de ces indicateurs entre eux donne des renseignements sur la façon dont les données observées sont réparties.
Dans une distribution parfaitement symétrique et concentrée autour de sa valeur centrale, les trois indicateurs coïncident :

$$\textbf{mode} = \textbf{médiane} = \textbf{moyenne}$$

Comparaison des indicateurs de tendance centrale

Comparaison des indicateurs de tendance centrale

Dans ce cas, tout est réparti autour du mode (c'est-à-dire de la valeur de plus forte densité ou de plus fort effectif) et, par symétrie, la médiane et la moyenne sont égales à cette quantité : il y a autant de valeurs à gauche qu'à droite et les valeurs à gauche compensent exactement les valeurs à droite.

Comparaison des indicateurs de tendance centrale

Comparaison des indicateurs de tendance centrale

Cette situation idéale sert de référence mais dans la pratique le mode, la médiane et la moyenne peuvent différer et leurs positions relatives indiquent une asymétrie dans la répartition des données.

Comparaison des indicateurs de tendance centrale

Comparaison des indicateurs de tendance centrale

Cette situation idéale sert de référence mais dans la pratique le mode, la médiane et la moyenne peuvent différer et leurs positions relatives indiquent une asymétrie dans la répartition des données. On dit que la courbe de fréquences est oblique du côté où la décroissance est la plus forte.

Comparaison des indicateurs de tendance centrale

Comparaison des indicateurs de tendance centrale

Cette situation idéale sert de référence mais dans la pratique le mode, la médiane et la moyenne peuvent différer et leurs positions relatives indiquent une asymétrie dans la répartition des données.
On dit que la courbe de fréquences est oblique du côté où la décroissance est la plus forte.
On distingue essentiellement deux situations :

Comparaison des indicateurs de tendance centrale

Comparaison des indicateurs de tendance centrale

Cette situation idéale sert de référence mais dans la pratique le mode, la médiane et la moyenne peuvent différer et leurs positions relatives indiquent une asymétrie dans la répartition des données. On dit que la courbe de fréquences est oblique du côté où la décroissance est la plus forte.
On distingue essentiellement deux situations :

1. 1. lorsque mode $<$ médiane $<$ moyenne, on dit que la distribution est oblique à gauche (ou de manière synonyme qu'elle est étalée à droite).

Comparaison des indicateurs de tendance centrale

Comparaison des indicateurs de tendance centrale

Cette situation idéale sert de référence mais dans la pratique le mode, la médiane et la moyenne peuvent différer et leurs positions relatives indiquent une asymétrie dans la répartition des données. On dit que la courbe de fréquences est oblique du côté où la décroissance est la plus forte.

On distingue essentiellement deux situations :

1. 1. lorsque mode $<$ médiane $<$ moyenne, on dit que la distribution est oblique à gauche (ou de manière synonyme qu'elle est étalée à droite).
2. 2. lorsque mode $>$ médiane $>$ moyenne, on dit que la distribution est oblique à droite (ou de manière synonyme qu'elle est étalée à gauche).

Comparaison des indicateurs de tendance centrale

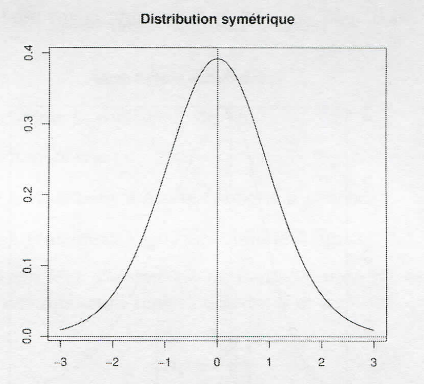
Comparaison des indicateurs de tendance centrale

Dans une distribution symétrique, le mode, la médiane et la moyenne coïncident.

Comparaison des indicateurs de tendance centrale

Comparaison des indicateurs de tendance centrale

Dans une distribution symétrique, le mode, la médiane et la moyenne coïncident.



Comparaison des indicateurs de tendance centrale

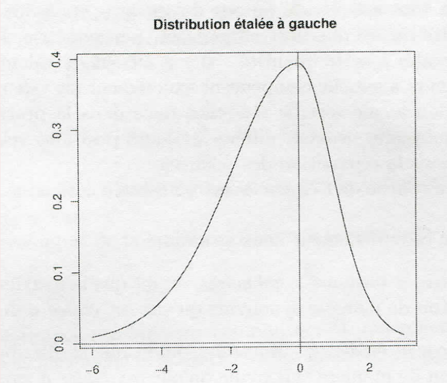
Comparaison des indicateurs de tendance centrale

La distribution ci-dessous est dite étalée vers la gauche (ou oblique à droite). On a mode $>$ médiane $>$ moyenne

Comparaison des indicateurs de tendance centrale

Comparaison des indicateurs de tendance centrale

La distribution ci-dessous est dite étalée vers la gauche (ou oblique à droite). On a mode > médiane > moyenne



Comparaison des indicateurs de tendance centrale

Comparaison des indicateurs de tendance centrale

La distribution ci-dessous est dite étalée vers la droite (ou oblique à gauche).

Comparaison des indicateurs de tendance centrale

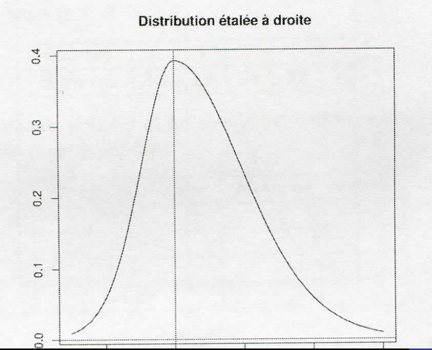
Comparaison des indicateurs de tendance centrale

La distribution ci-dessous est dite étalée vers la droite (ou oblique à gauche).
On a mode $<$ médiane $<$ moyenne

Comparaison des indicateurs de tendance centrale

Comparaison des indicateurs de tendance centrale

La distribution ci-dessous est dite étalée vers la droite (ou oblique à gauche).
On a mode $<$ médiane $<$ moyenne



Comparaison des indicateurs de tendance centrale

Comparaison des indicateurs de tendance centrale

Les cas de figure qui viennent d'être examinés ne recouvrent pas toutes les situations possibles.

Comparaison des indicateurs de tendance centrale

Comparaison des indicateurs de tendance centrale

Les cas de figure qui viennent d'être examinés ne recouvrent pas toutes les situations possibles.
Il y a des distributions qui présentent plusieurs modes...

Comparaison des indicateurs de tendance centrale

Comparaison des indicateurs de tendance centrale

Les cas de figure qui viennent d'être examinés ne recouvrent pas toutes les situations possibles.
Il y a des distributions qui présentent plusieurs modes...
Dans ce qui suit nous allons définir des indicateurs, appelés aussi coefficients, qui permettent de mesurer quantitativement le degré d'asymétrie d'une distribution et de sa courbe de fréquence.

Comparaison des indicateurs de tendance centrale

Comparaison des indicateurs de tendance centrale

Les cas de figure qui viennent d'être examinés ne recouvrent pas toutes les situations possibles.
Il y a des distributions qui présentent plusieurs modes...
Dans ce qui suit nous allons définir des indicateurs, appelés aussi coefficients, qui permettent de mesurer quantitativement le degré d'asymétrie d'une distribution et de sa courbe de fréquence.
Certains coefficients d'asymétrie sont définis à partir des quartiles, d'autres sont liés au moments d'ordre 3.

Indicateurs de forme

Coefficients d'asymétrie

Le coefficient de Yule (statisticien écossais, 1871-1951) est calculé à partir de la position des quartiles Q1, Q2 et Q3. Il s'écrit :

$$skewness = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)}$$

Indicateurs de forme

Coefficients d'asymétrie

Le coefficient de Yule (statisticien écossais, 1871-1951) est calculé à partir de la position des quartiles Q1, Q2 et Q3. Il s'écrit :

$$skewness = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)}$$

$$s = \frac{Q_1 + Q_3 - 2Q_2}{Q_3 - Q_1}$$

1. si s $=0$, il y a symétrie
2. si s >0, il y a étalement à droite (oblique à gauche)
3. si s <0, il y a étalement à gauche (oblique à droite).

Indicateurs de forme

Coefficients d'asymétrie : coefficient de Pearson

Pearson a calculé deux indicateurs de symétrie. Le premier se base sur la moyenne $\overline{x}$ et le mode M_o.

Indicateurs de forme

Coefficients d'asymétrie : coefficient de Pearson

Pearson a calculé deux indicateurs de symétrie. Le premier se base sur la moyenne $\overline{x}$ et le mode M_o.
Il est défini par :

$$s = \frac{\overline{x} - M_o}{\sigma}$$

Indicateurs de forme

Coefficients d'asymétrie : coefficient de Pearson

Pearson a calculé deux indicateurs de symétrie. Le premier se base sur la moyenne $\overline{x}$ et le mode M_o.
Il est défini par :

$$s = \frac{\overline{x} - M_o}{\sigma}$$

1. si $s = 0$, il y a symétrie
2. si $s > 0$, il y a étalement à droite (oblique à gauche)
3. si $s < 0$, il y a étalement à gauche (oblique à droite)

Indicateurs de forme

Coefficients d'asymétrie : coefficient de Pearson

Le second, noté β_1 il est plus utilisé.

Indicateurs de forme

Coefficients d'asymétrie : coefficient de Pearson

Le second, noté β_1 il est plus utilisé.
Il est défini à partir des moments centrés d'ordre 2 et 3 :

Indicateurs de forme

Coefficients d'asymétrie : coefficient de Pearson

Le second, noté β_1 il est plus utilisé.
Il est défini à partir des moments centrés d'ordre 2 et 3 :

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

Indicateurs de forme

Coefficients d'asymétrie : coefficient de Pearson

Le second, noté β_1 il est plus utilisé.
Il est défini à partir des moments centrés d'ordre 2 et 3 :

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

Le coefficient β_1 est toujours positif ou nul.

Indicateurs de forme

Coefficients d'asymétrie : coefficient de Pearson

Le second, noté β_1 il est plus utilisé.
Il est défini à partir des moments centrés d'ordre 2 et 3 :

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

Le coefficient β_1 est toujours positif ou nul.
S'il est nul, il y a symétrie.

Indicateurs de forme

Coefficients d'asymétrie : coefficient de Pearson

Le second, noté β_1 il est plus utilisé.
Il est défini à partir des moments centrés d'ordre 2 et 3 :

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

Le coefficient β_1 est toujours positif ou nul.
S'il est nul, il y a symétrie.
Sinon, la distribution est oblique et tout dépend du signe de μ_3 : par exemple, si $\mu_3 > 0$, c'est oblique à gauche.

Indicateurs de forme

Coefficients d'asymétrie : coefficient de Fisher

Noté γ_1 il est défini par :

Indicateurs de forme

Coefficients d'asymétrie : coefficient de Fisher

Noté γ_1 il est défini par :

$$\gamma_1 = \sqrt{\beta_1}$$

Indicateurs de forme

Coefficients d'asymétrie : coefficient de Fisher

Noté γ_1 il est défini par :

$$\gamma_1 = \sqrt{\beta_1}$$

C'est aussi une grandeur sans dimension.

Indicateurs de forme

Coefficients d'asymétrie : coefficient de Fisher

Noté γ_1 il est défini par :

$$\gamma_1 = \sqrt{\beta_1}$$

C'est aussi une grandeur sans dimension.
L'interprétation est toujours la même :

Indicateurs de forme

Coefficients d'asymétrie : coefficient de Fisher

Noté γ_1 il est défini par :

$$\gamma_1 = \sqrt{\beta_1}$$

C'est aussi une grandeur sans dimension.
L'interprétation est toujours la même :

1. si $\gamma_1 = 0$, il y a symétrie ;
2. si $\gamma_1 > 0$, il y a étalement à droite (oblique à gauche) ;
3. si $\gamma_1 < 0$, il y a étalement à gauche (oblique à droite).

Indicateurs de forme

Coefficients aplatissement

Les moments d'ordre 4 renseignent sur le degré d'aplatissement de la courbe de fréquences d'une distribution.

Indicateurs de forme

Coefficients aplatissement

Les moments d'ordre 4 renseignent sur le degré d'aplatissement de la courbe de fréquences d'une distribution.
L'aplatissement est jugé en se référant au modèle de la courbe de densité de la loi normale.

Indicateurs de forme

Coefficients aplatissement

Les moments d'ordre 4 renseignent sur le degré d'aplatissement de la courbe de fréquences d'une distribution.
L'aplatissement est jugé en se référant au modèle de la courbe de densité de la loi normale.
On dira qu'une courbe de fréquences est plus ou moins aplatie que le modèle de la loi normale.

Indicateurs de forme

Coefficients aplatissement

Les moments d'ordre 4 renseignent sur le degré d'aplatissement de la courbe de fréquences d'une distribution.
L'aplatissement est jugé en se référant au modèle de la courbe de densité de la loi normale.
On dira qu'une courbe de fréquences est plus ou moins aplatie que le modèle de la loi normale.
Le coefficient qui permet de mesurer quantitativement l'aplatissement s'appelle le kurtosis.

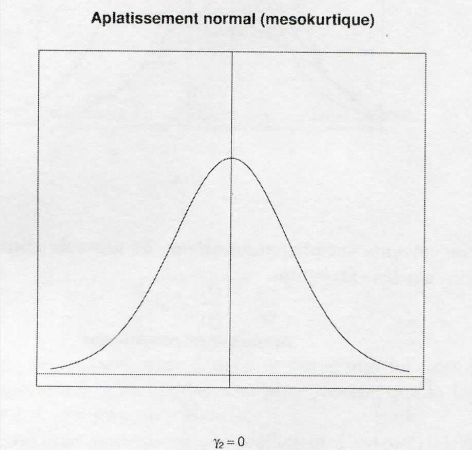
Indicateurs de forme

Coefficients aplatissement

Les moments d'ordre 4 renseignent sur le degré d'aplatissement de la courbe de fréquences d'une distribution.
L'aplatissement est jugé en se référant au modèle de la courbe de densité de la loi normale.
On dira qu'une courbe de fréquences est plus ou moins aplatie que le modèle de la loi normale.
Le coefficient qui permet de mesurer quantitativement l'aplatissement s'appelle le kurtosis.
La courbe suivante présente un aplatissement normal, comparable à celui de la densité d'une loi normale de Gauss.

Indicateurs de forme

Coefficients aplatissement



Indicateurs de forme

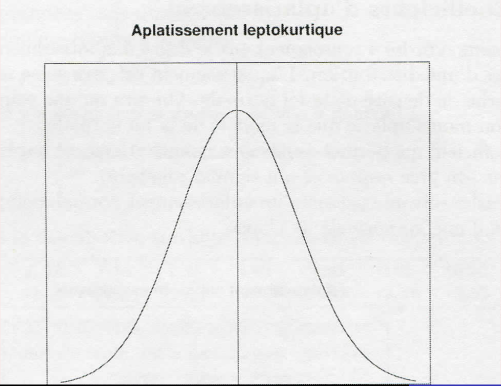
Coefficients aplatissement

La courbe suivante est plus plate qu'une loi normale. En compensation elle est plus dense sur les extrêmités.

Indicateurs de forme

Coefficients aplatissement

La courbe suivante est plus plate qu'une loi normale. En compensation elle est plus dense sur les extrêmités.



Indicateurs de forme

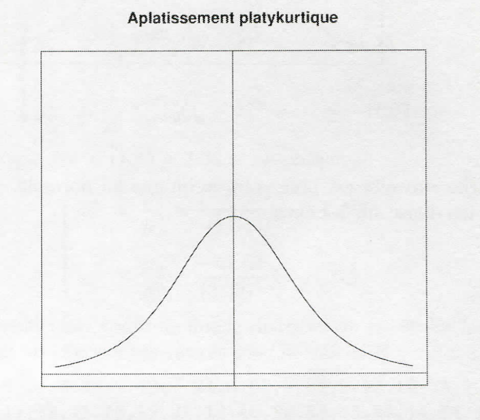
Coefficients aplatissement

Voici une comparaison des trois situations :

Indicateurs de forme

Coefficients aplatissement

Voici une comparaison des trois situations :



Indicateurs de forme

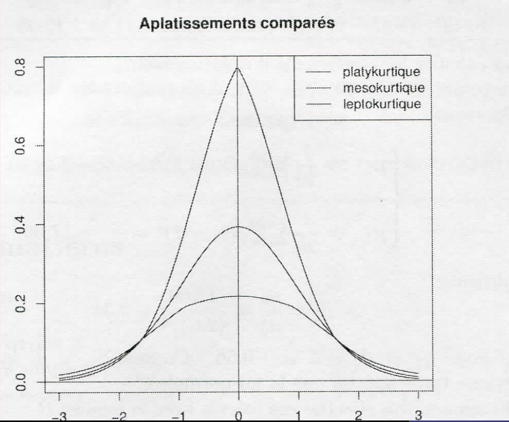
Coefficients aplatissement

Voici une comparaison des trois situations :

Indicateurs de forme

Coefficients aplatissement

Voici une comparaison des trois situations :



Indicateurs de forme

Coefficients aplatissement

Pearson a proposé d'utiliser le coefficient suivant :

Indicateurs de forme

Coefficients aplatissement

Pearson a proposé d'utiliser le coefficient suivant :

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

Indicateurs de forme

Coefficients aplatissement

Pearson a proposé d'utiliser le coefficient suivant :

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

On montre que ce rapport vaut 3 dans le cas d'une loi normale parfaite. Donc si β_2 est supérieur à 3, la courbe sera plus pointue que la loi normale et si β_2 est inférieur à 3, elle sera plus aplatie.

Indicateurs de forme

Coefficients aplatissement

Pearson a proposé d'utiliser le coefficient suivant :

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

On montre que ce rapport vaut 3 dans le cas d'une loi normale parfaite. Donc si β_2 est supérieur à 3, la courbe sera plus pointue que la loi normale et si β_2 est inférieur à 3, elle sera plus aplatie. Il est plus naturel (par analogie avec le coefficient d'asymétrie), de considérer que la valeur de référence est 0 et non pas 3.

Indicateurs de forme

Coefficients aplatissement

Pearson a proposé d'utiliser le coefficient suivant :

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

On montre que ce rapport vaut 3 dans le cas d'une loi normale parfaite. Donc si β_2 est supérieur à 3, la courbe sera plus pointue que la loi normale et si β_2 est inférieur à 3, elle sera plus aplatie.
Il est plus naturel (par analogie avec le coefficient d'asymétrie), de considérer que la valeur de référence est 0 et non pas 3.
Aussi Fisher a proposé d'adopter comme coefficient d'aplatissement la quantité :

Indicateurs de forme

Coefficients aplatissement

$$\gamma_2 = \beta_2 - 3$$

On interprète le kurtosis γ_2 de la manière suivante :

Indicateurs de forme

Coefficients aplatissement

$$\gamma_2 = \beta_2 - 3$$

On interprète le kurtosis γ_2 de la manière suivante :

1. si $\gamma_2 = 0$, la courbe de fréquences est comparable à celle de la loi normale.

Indicateurs de forme

Coefficients aplatissement

$$\gamma_2 = \beta_2 - 3$$

On interprète le kurtosis γ_2 de la manière suivante :

1. si $\gamma_2 = 0$, la courbe de fréquences est comparable à celle de la loi normale.
   On dit qu'elle est mésokurtique.
2. si $\gamma_2 > 0$, la courbe de fréquences est plus pointue que celle de la loi normale.

Indicateurs de forme

Coefficients aplatissement

$$\gamma_2 = \beta_2 - 3$$

On interprète le kurtosis γ_2 de la manière suivante :

1. si $\gamma_2 = 0$, la courbe de fréquences est comparable à celle de la loi normale.
   On dit qu'elle est mésokurtique.
2. si $\gamma_2 > 0$, la courbe de fréquences est plus pointue que celle de la loi normale.
   On dit qu'elle est leptokurtique.
3. si $\gamma_2 < 0$, la courbe de fréquences est plus aplatie que celle de la loi normale.

Indicateurs de forme

Coefficients aplatissement

$$\gamma_2 = \beta_2 - 3$$

On interprète le kurtosis γ_2 de la manière suivante :

1. si $\gamma_2 = 0$, la courbe de fréquences est comparable à celle de la loi normale.
   On dit qu'elle est mésokurtique.
2. si $\gamma_2 > 0$, la courbe de fréquences est plus pointue que celle de la loi normale.
   On dit qu'elle est leptokurtique.
3. si $\gamma_2 < 0$, la courbe de fréquences est plus aplatie que celle de la loi normale.
   On dit qu'elle est platykurtique.

Application

Application

On considère la distribution suivante comportant 20 valeurs numériques :

0.04	6.2	6.3	7.1	7.2	8.6	9.8	9.9	10.5	10.9
11.1	11.4	11.7	11.7	12.2	12.4	12.8	13.3	14.5	14.7

On va calculer les différents coefficients d'asymétrie.

Application

Application

On considère la distribution suivante comportant 20 valeurs numériques :

0.04	6.2	6.3	7.1	7.2	8.6	9.8	9.9	10.5	10.9
11.1	11.4	11.7	11.7	12.2	12.4	12.8	13.3	14.5	14.7

On va calculer les différents coefficients d'asymétrie.
Pour le coefficient de Yule, on a besoin des quartiles :

25%	50%	74%
7.2	11	12.2