

## INTERVALLES DE CONFIANCE.

### 1 Estimer la moyenne d'une série de données

#### 1.1 Estimation ponctuelle

On considère une série de données dont on souhaite estimer la moyenne notée  $m$ . On ne connaît pas toutes les valeurs de cette série, mais on dispose d'un échantillon de  $n$  valeurs  $x_1, \dots, x_n$ .

Il est naturel d'approcher  $m$  par la moyenne  $m_X$  de ces valeurs :

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

#### 1.2 Mesure de la dispersion

On définit la variance des valeurs de l'échantillon :

$$v_X = \frac{1}{n} \sum_{i=1}^n (x_i - m_X)^2$$

Ainsi que l'écart-type :

$$\sigma_X = \sqrt{v_X}$$

#### Remarques :

1. Plus les valeurs sont dispersées autour de leur moyenne, plus la variance et l'écart-type sont élevés.
2. Si toutes les valeurs sont identiques, la variance et l'écart-type sont nuls.
3. Avantage de l'écart-type : L'écart-type est homogène à la variable mesurée, alors que la variance est homogène à son carré.

#### 1.3 Intervalle de confiance

Une estimation ponctuelle donne une valeur approchée de  $m$  mais ne donne pas de renseignement sur la confiance que l'on peut accorder à cette valeur.

Par exemple, si la série est : 0,2,4,6,8,8,9,10,11,11,11,13,14,14,17,17,18,19,19,20 mais que l'échantillon considéré est 0,2,4,8,8,10,11,11,14, la moyenne obtenue sur l'échantillon est très loin de la valeur recherchée (on dit alors que l'échantillon n'est pas représentatif).

La proposition suivante donne :

- un intervalle (et non seulement une valeur) dans lequel la valeur cherchée ( $m$ ) est susceptible de se trouver,
- La probabilité que la valeur cherchée ( $m$ ) se trouve bien cet l'intervalle.

**Proposition 1** Si  $n$  est assez grand, la probabilité que :

$$m \in \left[ m_X - 1,96 \frac{\sigma}{\sqrt{n}}, m_X + 1,96 \frac{\sigma}{\sqrt{n}} \right]$$

(où  $\sigma$  désigne l'écart-type de **toute la série**) est égale à 0,95.

On dit que l'intervalle  $\left[ m_X - 1,96 \frac{\sigma}{\sqrt{n}}, m_X + 1,96 \frac{\sigma}{\sqrt{n}} \right]$  est un intervalle de confiance à 95%.

#### Remarques :

1.  $n$  assez grand signifie  $n \geq 10$  (mais c'est négociable cette année...)

2. On ne connaît pas l'écart-type  $\sigma$  de toute la série, mais on connaît l'écart-type calculé sur l'échantillon (que l'on a noté  $\sigma_X$ ).  
 On approchera  $\sigma$  par  $\sqrt{\frac{n}{n-1}}\sigma_X$  (c'est d'ailleurs la formule donnée par les feuilles de calculs (excel et OpenOffice Calc) quand on leur fait calculer l'écart-type d'une série de données).  
 Cette valeur correspond à :

$$\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - m_X)^2}$$

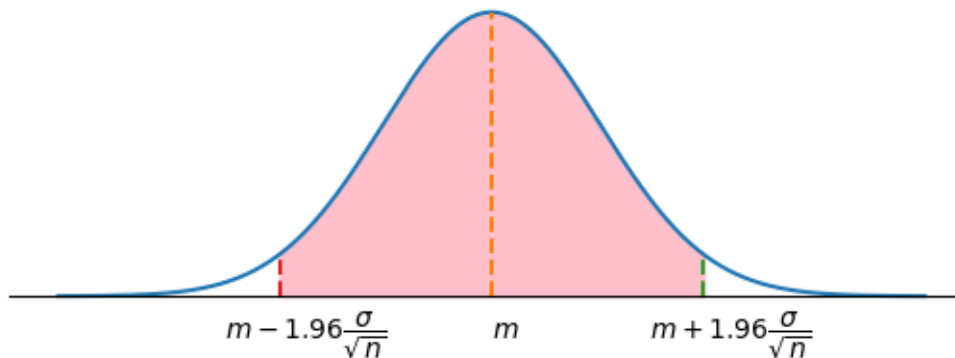
(dans le calcul de la variance, on remplace  $\frac{1}{n}$  par  $\frac{1}{n-1}$ )

3. Ces intervalles de confiance doivent apparaître sur les graphiques. Si, pour deux séries de données, ces intervalles ne se recoupent pas, on peut affirmer, avec un risque d'erreur de 5%, que les moyennes des deux séries de données sont différentes.
4. Cette proposition est un corollaire du théorème de la limite centrée, qui dit que lorsqu'on considère une suite  $(X_n)_{n \in \mathbb{N}^*}$  de variable aléatoire réelles indépendantes, de même loi, ayant pour espérance  $m$  et écart-type  $\sigma$ , la suite de variable aléatoires  $\left(\sqrt{n} \frac{(M_n - m)}{\sigma}\right)_{n \in \mathbb{N}^*}$  converge, en loi, vers une

variable aléatoire de loi normale de moyenne  $m$  et d'écart-type  $\sigma$  (on a noté  $M_n = \frac{\sum_{i=1}^n X_i}{n}$ ).

Cela signifie que pour  $n$  assez grand ( $n \geq 10$ ), on peut considérer que la variable aléatoire  $M_n$  suit une loi normale de moyenne  $m$  et d'écart-type  $\frac{\sigma}{\sqrt{n}}$ .

### Densité de probabilité - Loi normale $\mathcal{N}(m, (\frac{\sigma}{\sqrt{n}})^2)$



Sur ce graphique, l'aire coloriée vaut 0,95 (la valeur "1,96" a été choisie pour que l'aire coloriée vaille 0,95).

Donc la probabilité que  $M_n$  soit dans l'intervalle  $\left[m - 1.96 \frac{\sigma}{\sqrt{n}}, m + 1.96 \frac{\sigma}{\sqrt{n}}\right]$  est de 0,95.

Avec les notations initiales (on remplace  $M_n$  par  $m_X$ ), cela s'écrit :

$$P\left(|m_X - m| \leq 1,96 \frac{\sigma}{\sqrt{n}}\right) = 0,95$$

ou encore :

$$P\left(m \in \left[m_X - 1.96 \frac{\sigma}{\sqrt{n}}, m_X + 1.96 \frac{\sigma}{\sqrt{n}}\right]\right) = 0,95$$

## 2 Construction graphique des données statistiques

### 2.1 Quelques fonctions à connaître ou à retrouver

Pour insérer les données statistiques dans un tableau OpenOffice (ou Excel), taper dans la case et se laisser guider :

= MOYENNE( ) pour obtenir la moyenne d'une série de données

= ECARTYPE( ) pour obtenir l'écart-type

= NBVAL( ) pour obtenir l'effectif

= INTERVALLE.CONFIANCE(0,05; écart-type; effectif) pour obtenir l'intervalle de confiance à 95%

= INTERVALLE.CONFIANCE( $\alpha/100$ ; écart-type; effectif) pour obtenir l'intervalle de confiance à  $(100 - \alpha)$  %

On peut aussi cliquer sur l'onglet **Insertion** → **Fonction** → **Catégorie : statistique** pour trouver une fonction.

Si votre logiciel est configuré en anglais, vous pouvez utiliser un traducteur des fonctions sur le site :

[excel-translator.de](http://excel-translator.de)

	A	B	C	D
1				
2		Données 1	Données 2	
3		19,8	22	
4		22,1	23,9	
5		21,5	20,9	
6		20,9	23,8	
7		22	25	
8		21	24	
9		22,3	23,8	
10		21	21,7	
11		20,3	23,1	
12		20,9	22,8	
13		22	23,5	
14		22	23	
15		20,8	23	
16		21,2	23,1	
17				
18	Moyenne	21,27142857	23,11428571	
19	Écart-type	0,741545757	1,049437317	
20	Effectif	14	14	
21	Int. de conf. à 95%	0,388438285	0,549718783	
22				

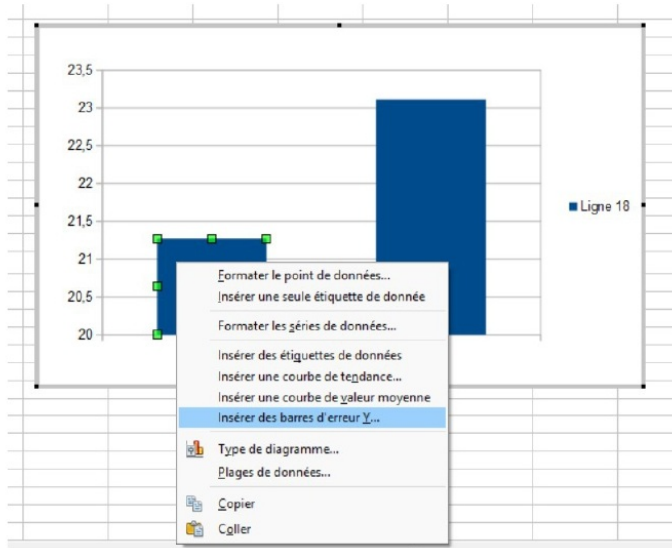
## 2.2 Insérer des diagrammes

Pour créer un diagramme :

- **Sélectionner** la ligne Moyenne (2 valeurs)
- Dans l'onglet **insertion**, sélectionner **diagramme** (type de diagramme : colonne normal)

Pour ajouter les intervalles de confiance :

- **Sélectionner** un histogramme puis cliquer droit et **sélectionner insérer des barres d'erreur Y**



Barres d'erreur pour la série de données 'Ligne 18'

**Barres d'erreur Y** Ligne

Catégorie d'erreur

Valeur constante

Pourcentage

Écart type

Plage de cellules

Indicateur d'erreur

Positif et négatif

Positif

Négatif

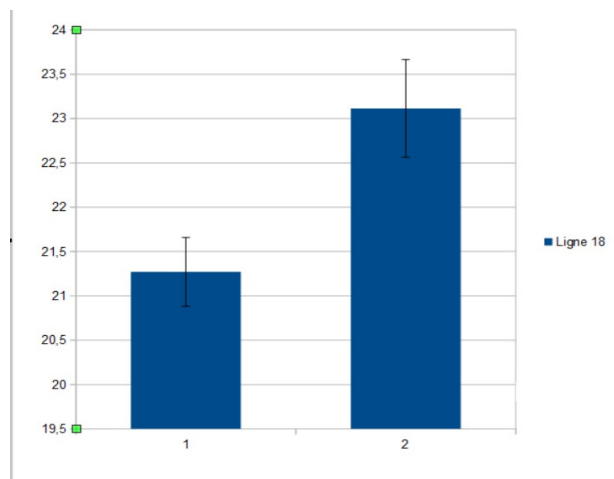
Paramètres

Positif (+) : \$Feuille1.\$B\$21:\$C\$21

Négatif (-) : \$Feuille1.\$B\$21:\$C\$21

Valeur identique pour les deux

OK Annuler Aide Rétablir



### 3 Complément : Variable aléatoire admettant une densité

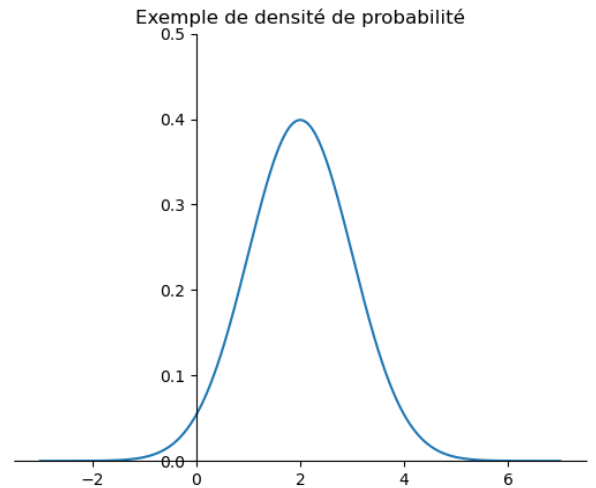
**définition 1** On appelle densité de probabilité une fonction  $f$  définie sur  $\mathbb{R}$ , positive, continue sauf éventuellement en un nombre fini de points, telle que l'intégrale  $\int_{-\infty}^{+\infty} f(t) dt$  soit convergente et que :

$$\int_{-\infty}^{+\infty} f(t) dt = 1$$

**Illustration graphique :**

La fonction  $f$  représentée par la courbe ci-contre est une densité de probabilité :

- la fonction  $f$  est définie sur  $\mathbb{R}$ ,
- la fonction  $f$  est positive, continue sur  $\mathbb{R}$ ,
- $\int_{-\infty}^{+\infty} f(t) dt = 1$ ,  
ce qui signifie que l'aire sous la courbe vaut 1.



**définition 2** Soit  $(\Omega, \mathcal{P}(\Omega), P)$  un espace probabilisé et soit  $X$  une variable aléatoire réelle définie sur  $\Omega$ . On dit que  $X$  est une variable aléatoire à densité s'il existe une fonction  $f$ , densité de probabilité, telle que pour tout  $x \in \mathbb{R}$ , on ait :

$$P(X \leq x) = \int_{-\infty}^x f(t) dt.$$

Donc, pour tous  $(a, b) \in \mathbb{R}^2$  tels que  $a \leq b$ ,

$$P(a \leq X \leq b) = \int_a^b f(t) dt$$

**Exemple :**

On considère une variable aléatoire  $X$  de densité la fonction, notée  $f$ , dont la courbe représentative est dessinée ci-contre.

Soient  $a$  et  $b$  deux réels tels que  $a \leq b$ . D'après la définition précédente :

$$P(a \leq X \leq b) = \int_a^b f(t) dt$$

Donc  $P(a \leq X \leq b)$  est l'aire du domaine colorié.

**Remarque :** Puisque  $\int_{-\infty}^{+\infty} f(t) dt = 1$ , la quantité  $\int_a^b f(t) dt$  appartient à  $[0, 1]$ .

