Les statistiques descriptives

Table des matières

1 Pourcentages – Rappels

1.1 Premières définitions

Définition 1: Pourcentage

Le symbole $\%$ désigne **une division par** 100.

Exemple 1

On a les égalités suivantes :

$$30\% = \frac{30}{100} = 30 \div 100 = 0{,}3$$
$$150\% = \frac{150}{100} = 150 \div 100 = 1{,}5$$
$$98\% = \frac{98}{100} = 98 \div 100 = 0{,}98$$
$$100\% = \frac{100}{100} = 100 \div 100 = 1$$

Définition 2: Pourcentage d'une quantité

Soit x un nombre réel. $p\%$ de x correspond à la quantité qui représente $p\%$ de x, et se calcule par :

$$p\% \text{ de } x = p\% \times x$$

Exemple 2

Si on veut calculer 30% de 50, on calcule $50 \times 30\% = 50 \times 0{,}3 = 15$. Donc 15 représente 30% de 50.

1.2 Utilisation et formules sur les variations absolues et relatives.

1.2.1 Variations absolues

Définition 3: Variations absolues

Une variation absolue est la différence entre la valeur d'une quantité entre deux variations.

Exemple 3

Un livre coûte 30 euros puis, après une réduction, il coûte maintenant $22{,}5$ euros.
La variation absolue est de $22{,}5 - 30 = -7{,}5$ euros. Le prix a donc **baissé** de $7{,}5$ euros.

Définition 4: Variation absolue

Si on considère une quantité p qui a subit une évolution et est devenu p', alors la variation absolue notée Δp se calcule par :

$$\Delta p = p' - p$$

1.2.2 Variations relatives

Définition 5: Variations relatives

Lorsque l'on regarde les évolutions d'une quantité (la température au cours du temps, la valeur d'une action au cours du temps, le prix de l'électricité au cours du temps, le salaire médian français, etc) que l'on quantifie en pourcentage, on parle de variations relatives.

Exemple 4

Si on dit «L'électricité a augmenté de 5% ce mois-ci», on veut dire que l'on paiera 5% de **plus** que le mois dernier. C'est une évolution relative car cette augmentation dépend du prix du mois dernier. Le prix a augmenté de 5% par rapport au mois dernier.
Dans cet exemple, si on note p le prix de l'électricité du mois dernier, et p' le prix du mois suivant, alors :

$$p' = \underbrace{p}_{\text{prix du mois dernier}} + \underbrace{5\% \times p}_{\text{plus cinq pourcents du prix du mois dernier}}$$

Pour aller plus vite, on peut factoriser par p :

$$p' = (1 + 5\%)p$$

Et donc $p' = 1{,}05 \times p$

Définition 6: Augmentation relative, taux, coefficient multiplicateur

Dans l'exemple précédent, on parle d'une augmentation relative de **taux relatif** 5%, et de **coefficient multiplicateur** $1{,}05$.

Proposition 1

Si on note t le taux d'évolution relative, et CM le coefficient multiplicateur, alors :

$$CM = 1 + t$$

et t peut être négatif pour désigner une réduction.

Exemple 5

Si les prix d'un live qui valait 30 € a subit une réduction de 25%, alors le nouveau prix du livre sera de $30 \times (1 - 25\%) = 30 \times 0{,}75 = 22.5$ €. Le taux d'évolution relative est ici égale à -25%.

Proposition 2

Si on considère une quantité p qui a subit une évolution et est devenu p', pour calculer le taux d'évolution t associé, on peut faire :

$$t = \frac{p' - p}{p}$$

Démonstration 1

L'égalité se démontre en quelques calculs qui se basent sur la définition de p'.

$$\begin{aligned}\frac{p'-p}{p} &= \frac{(1+t)p-p}{p} \\ &= \frac{p+tp-p}{p} \\ &= \frac{tp}{p} \\ &= t\end{aligned}$$

Ce qui est bien l'égalité que l'on voulait montrer.

Exemple 6

Notre livre est passé de $p = 30$ euros à $p' = 22.5$ euros. Le taux d'évolution t associé est donc :

$$t = \frac{22.5-30}{30} = \frac{-7.5}{30} = -0.25$$

On retrouve bien une **réduction** de 25%.

Proposition 3

Si on considère une quantité p qui a subit une évolution et est devenu p', pour calculer le coefficient multiplicateur CM associé, on peut faire :

$$CM = \frac{p'}{p}$$

Démonstration 2

Par définition on sait que :

$$p' = CM \times p$$

en divisant chaque membre de l'équation par $p \neq 0$, on obtient :

$$CM = \frac{p'}{p}$$

et c'est l'égalité que l'on voulait montrer.

Exemple 7

Notre livre est passé de $p = 30$ euros à $p' = 22.5$ euros. Le coefficient multiplicateur associé à cette évolution peut donc se calculer par :

$$CM = \frac{22.5}{30} = 0.75$$

Le coefficient multiplicateur est donc $0{,}75$. Au passage, on retrouve bien que $CM = 1 + t$ puisque $t = -0{,}25$.

Question 1

Si un pull passe de 30 euros à 40 euros, quelle est son augmentation relative ?

2 Vocabulaire

Définition 7: Série

Une série statistique est la compilation de plusieurs nombres représentant un même phénomène.

Question 2

1. Citer différentes séries statistique que vous pouvez mesurer devant vous.
2. Même question, mais cette fois-ci vous vous prenez pour l'entreprise Google : quelle série statistique cette entreprise peut-elle mesurer. À votre avis quelle est la taille de cette série ?

Les séries sont souvent gigantesques, c'est pourquoi on utilise des **indicateurs** qui permettent de les résumer.

Définition 8: Indicateur statistique

Un indicateur statistique est un calcul, à effectuer sur une série, qui permet de résumer l'information contenue dans cette série.
Il existe plusieurs types d'indicateurs : les indicateur de position, et les indicateurs de dispersion.

Exemple 8

La moyenne est un indicateur ! C'est d'ailleurs le premier que l'on va voir.

Question 3

À votre avis, la moyenne est un indicateur de position ou de dispersion ?

Définition 9: Effectif

L'effectif de la série correspond au nombre d'occurrence d'une donnée dans la série, c'est-à-dire au nombre de fois où elle est répétée.

Exemple 9

Le tableau suivant :

(x_i)	Effectifs
1	3
3	5
4	10

correspond à la série statistique :

$$x = (1; 1; 1; 3; 3; 3; 3; 3; 4; 4; 4; 4; 4; 4; 4; 4; 4; 4)$$

Cette série pourrait représenter par exemple les notes d'un contrôle noté sur 5 avec 3 élèves qui ont eu $1/5$, 5 élèves qui ont eu $3/5$ et 10 élèves qui ont eu $4/5$.

Question 4

Quel calcul permet de retrouver la taille de la série (x_i) de la question précédente ?

Question 5

À quoi ressemblerait le tableau qui correspond à la série statistique suivante ?

$$y = (3; 3; 3; 3; 4; 4; 7; 7; 7; 7; 7; 7)$$

Attention, cette fois-ci la série s'appelle (y). Vérifier que la taille de votre série est cohérente avec votre tableau.

3 Indicateur de position

3.1 Moyenne d'une série

Définition 10: Moyenne d'une série (sans regroupement par effectif)

Soit (x_i) une série statistique contenant n données. Alors, si on note $\overline{x}$ la moyenne de cette série, par définition :

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Définition 11: Moyenne d'une série (avec regroupement par effectif)

Soit (x_i) une série statistique, d'effectif (e_i) de taille n. Alors, si on note $\overline{x}$ la moyenne de cette série, on a :

$$\overline{x} = \frac{x_1 \times e_1 + x_2 \times e_2 + \ldots + x_n \times e_n}{e_1 + e_2 + \ldots e_n}$$

Exemple 10

Si on considère le tableau suivant, qui récapitule des notes d'élèves, regroupé par tranche de 5 points :

La série (x_i)	Valeur de la série retenue	Effectifs
$0 \leq x_i < 5$	2,5	12
$5 \leq x_i < 10$	7,5	23
$10 \leq x_i < 15$	12.5	8
$15 \leq x_i \leq 20$	17,5	10

Alors la moyenne de la série se calcule par :

$$\overline{x} = \frac{2{,}5 \times 12 + 7{,}5 \times 23 + 12{,}5 \times 8 + 17{,}5 \times 10}{12 + 23 + 8 + 10}$$

Question 6

Pour être sûr d'utiliser sa calculatrice correctement, montrez que la moyenne $\overline{x}$ de l'exemple précédent est :

$$\overline{x} = 9.00943396226415\ldots \approx 9$$

Question 7

Si j'ai eu trois notes (sur vingt) en dessous de 5, deux notes entre 5 et 10, et cinq notes entre 10 et 15, quelle est l'estimation de ma moyenne d'après les calculs précédents ?

Proposition 4: Linéarité de la moyenne

Quelque soit $k \in \mathbb{R}$, si on considère la série statistique $(x_i)_i$ et la série $(k \times x_i)_i$ alors :

$$\overline{k \times x} = k\overline{x}$$

Exemple 11

Si on considère la série suivante :

(x_i)	Effectifs
1	3
3	5
4	10

alors :

$$\overline{x} = \frac{1 \times 3 + 3 \times 5 + 4 \times 10}{3 + 5 + 10} = \frac{58}{18} = \frac{29}{9} \approx 3{,}2$$

Si maintenant on considère la série suivante :

(y_i)	Effectifs
10	3
30	5
40	10

Alors, on remarque que c'est la série précédente où chaque valeur est multipliée par 10. Donc

$$\overline{y} = 10\overline{x} = \frac{290}{9} \approx 32$$

Démonstration 3

Cette proposition se démontre avec l'égalité suivante, valable quelque soit $k \in \mathbb{R}$, et $(x_i)_{1 \leq i \leq n}$ une série statistique :

$$\begin{aligned} \overline{k \times x} &= \frac{k \times x_1 + k \times x_2 + \ldots + k \times x_n}{n} \\ &= \frac{k \times (x_1 + x_2 + \ldots + x_n)}{n} \\ &= k \times \frac{x_1 + x_2 + \ldots + x_n}{n} \\ &= k \times \overline{x} \end{aligned}$$

Proposition 5

Soit $a \in \mathbb{R}$ et $x = (x_i)$ une série statistique. Alors :

$$\overline{x + a} = \overline{x} + a$$

Et $x + a$ désigne la série $(x_i + a)_i$

Exemple 12

Si on considère la série :

$$x = (1; 2; 3; 4)$$

De moyenne $\overline{x} = \frac{1+2+3+4}{4} = \frac{10}{4} = 2{,}5$, alors la série

$$y = (10; 12; 13; 14) = x + 10$$

a comme moyenne $\overline{y} = \overline{x} + 10 = 12{,}5$.

Démonstration 4

Cette proposition se démontre à l'aide des égalités suivantes, vraies quelque soit la série statistique x et $a \in \mathbb{R}$:

$$\begin{aligned}
\overline{x+a} &= \frac{(x_1+a)+(x_2+a)+\ldots+(x_n+a)}{n} \\
&= \frac{(x_1+x_2+\ldots+x_n)+(a+a+\ldots+a)}{n} \\
&= \frac{x_1+x_2\ldots x_n+n\times a}{n} \\
&= \frac{x_1+x_2+\ldots x_n}{n}+\frac{n\times a}{n} \\
&= \overline{x}+a
\end{aligned}$$

3.2 Médiane d'une série

Définition 12: Médiane d'une série

La médiane m d'une série statistique (x_i) est un nombre tel qu'il y ait autant de valeurs x_i plus grande que m que de valeurs x_i plus petite. (Dit autrement «m coupe en deux la série statistique»).
Le mot «médiane» vient du latin *medianus* qui signifie «du milieu».

3.3 Quartile

3.3.1 Premier quartile

Définition 13: Premier quartile

Le premier quartile Q_1 est le nombre tel que 25% des valeurs de la série statistique soient plus petite que Q_1.

3.3.2 Troisième quartile

Définition 14: Troisième quartile

Le troisième quartile Q_3 est le nombre tel que 75% des valeurs de la série statistique soient plus petite que Q_3.

4 Indicateur de dispersion

4.1 Écart-type

Définition 15: Écart-type

Soit (x_i) une série statistique, d'effectifs (n_i) de taille p. On calcule l'écart-type de cette série par :

$$\sigma = \sqrt{\frac{n_1(x_1-\overline{x})^2+\ldots+n_p(x_p-\overline{x})^2}{n_1+\ldots+n_p}}$$

Exemple 13

Voici une série :

$$1\ ;\ 1\ ;\ 4\ ;\ 7\ ;\ 7\ ;\ 9\ ;\ 9\ ;\ 9$$

Pour calculer son écart-type, on calcule d'abord la moyenne de la série :

$$\overline{x} = \frac{2\times 1+4+2\times 7+3\times 9}{2+1+2+3} = \frac{47}{8} = 5{,}875$$

Puis, on utilise la formule vue plus haut :

$$\sigma = \sqrt{\frac{2(1-5{,}875)^2+(4-5{,}875)^2+2(7-5{,}875)^2+3(9-5{,}875)^2}{8}}$$

Sur la calculatrice, on peut d'abord calculer «sans la racine», et appliquer la racine carré à la toute fin.

4.2 Écart interquartile

Définition 16: Écart interquartile

Si Q_1 et Q_3 sont les deux quartiles d'une série, on définit $E = Q_3 - Q_1$ l'écart interquartile de la série.