Statistiques

1 Moyenne, médiane, quartiles

1.1 Moyenne

Soit une série statistique quantitative rangée dans l'ordre croissant :

variable x_i	x_1	x_2	...	x_p	Total
effectif n_i	n_1	n_2	...	n_p	n
fréquence f_i	f_1	f_2	...	f_p	1 (= 100%)

Définition :
La moyenne $\overline{x}$ d'une série statistique quantitative est le quotient de la somme des valeurs de la série par l'effectif total :

$$\overline{x} = \frac{x_1n_1 + x_2n_2 + ... + x_pn_p}{n} = \frac{\sum\limits_{k=1}^{p} x_kn_k}{n}$$

$$\overline{x} = x_1f_1 + x_2f_2 + ... + x_pf_p = \sum\limits_{k=1}^{p} x_kf_k$$

Théorème : linéarité de la moyenne
Soit une moyenne statistique $\overline{x}$ et a et b deux nombres réels.

$$\overline{ax+b} = a\overline{x} + b.$$

Indication démonstration :

Écrire autrement $\overline{ax+b} = \dfrac{(ax_1+b)n_1 + (ax_2+b)n_2 + ... + (ax_p+b)n_p}{n} = \dfrac{\sum\limits_{k=1}^{p} (ax_k+b)n_k}{n}$.

1.2 Médiane et quartile

Définition :
Soit une série statistique quantitative discrète, telle que les valeurs de la variable sont rangées en ordre croissant.

- Dans le cas d'un effectif total n pair, la médiane M est la moyenne des deux valeurs centrales de la série.
- Dans le cas d'un effectif total n impair, la médiane M est la valeur centrale de la série.

Définition :
Soit une série statistique quantitative discrète, telle que les valeurs de la variable sont rangées en ordre croissant.

- Le premier quartile Q_1 est la plus petite valeur de la série telle qu'au moins 25% des valeurs de la série ait une valeur inférieure ou égale à Q_1.
- Le troisième quartile Q_3 est la plus petite valeur de la série telle qu'au moins 75% des valeurs de la série ait une valeur inférieure ou égale à Q_3.

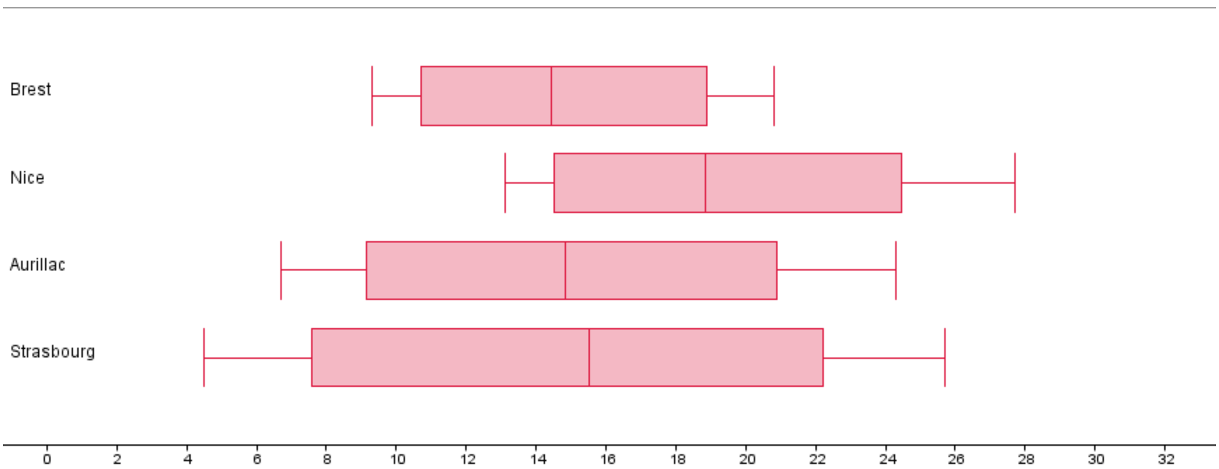
Remarque :
On peut rédiger la définition ainsi :
Le i^{eme} quartile Q_i est la plus petite valeur de la série telle qu'au moins $\left(i \times \frac{1}{4} \times 100\right)$ % des valeurs de la série ait une valeur inférieure ou égale à Q_i, avec i dans $\{1; 2; 3\}$.
Ainsi le deuxième quartile peut être assimilé à la médiane de la série.
On peut définir n'importe quel partage fractionné de la série comme par exemple :
Les déciles :
Le i^{eme} décile D_i est la plus petite valeur de la série telle qu'au moins $\left(i \times \frac{1}{10} \times 100\right)$ % des valeurs de la série ait une valeur inférieure ou égale à D_i, avec i dans $\{1; 2; 3; 4; 5; 6; 7; 8; 9\}$.
Les centiles :
Le i^{eme} centile c_i est la plus petite valeur de la série telle qu'au moins $\left(i \times \frac{1}{100} \times 100\right)$ % des valeurs de la série ait une valeur inférieure ou égale à c_i, avec i dans $\{1; 2; ...; 99\}$.

1.3 Récapitulatif, diagramme en boîtes

Exemple-exercice :
La station météo enregistre les températures maximum suivantes pour quatre villes :

Mois	Brest	Nice	Aurillac	Strasbourg
Janvier	9.3	13.1	4.5	6.7
Février	9.5	13.4	6.4	7.9
Mars	11.5	15.2	11.4	11.2
Avril	13.2	17	15.7	13.7
Mai	16.2	20.7	20.2	17.8
Juin	18.7	24.3	23.4	21.5
Juillet	20.7	27.3	25.7	24.3
Août	20.8	27.7	25.4	24
Septembre	19.1	24.6	21	20.3
Octobre	15.7	21	15.3	16
Novembre	12.2	16.6	8.8	10.4
Décembre	9.9	13.8	5.2	7.5

On donne les résultats suivants :



Statistiques

	n	Moyenne	σ	s	Min	Q1	Médiane	Q3	Max
Brest	12	14.7333	4.1812	4.3671	9.3	10.7	14.45	18.9	20.8
Nice	12	19.5583	5.2009	5.4321	13.1	14.5	18.85	24.45	27.7
Aurillac	12	15.1083	6.1871	6.4623	6.7	9.15	14.85	20.9	24.3
Strasbourg	12	15.25	7.5704	7.907	4.5	7.6	15.5	22.2	25.7

1. Retrouver les calculs de la moyenne (notée s dans les résultats), de la médiane et des quartiles de chaque ville.
2. Commenter les diagrammes.

2 Ecart-type

Soit une série statistique quantitative rangée dans l'ordre croissant dont la moyenne est $\overline{x}$:

variable x_i	x_1	x_2	...	x_p	Total
effectif n_i	n_1	n_2	...	n_p	n
fréquence f_i	f_1	f_2	...	f_p	1 (= 100%)

Définition :

- La **variance**, V_x d'une série statistique quantitative est la moyenne des carrés des écarts à la moyenne $\overline{x}$:

$$V_x = \frac{\sum_{i=0}^{n} n_i \left(x_i - \overline{x}\right)^2}{n}$$

- L'**écart-type**, σ_x d'une série statistique quantitative est la racine carrée de la variance V_x :

$$\sigma_x = \sqrt{V_x}$$

Exemple :
Soit la série statistiques dont dont on donne le tableau des valeurs et des effectifs :

x_i	2	4	5	6	9
n_i	3	7	9	4	5

1. Calculer la moyenne $\overline{x}$ de la série.
2. Compléter le tableau par les valeurs :

$(x_i - \overline{x})^2$					
n_i	3	7	9	4	5

3. En déduire la variance, puis l'écart-type de la série.

Théorème :
Avec les notations précédentes,

$$V_x = \overline{x^2} - \overline{x}^2$$

Démonstration : laissée en exercice

Exemple :
Retrouver la variance de la série de l'exemple précédent à partir du résultat $V_x = \overline{x^2} - \overline{x}^2$.

Théorème :
Soit une série statistique quantitative de variable x et un nombre réel a, note V_x la variance de la variable x.

$$V_{ax} = a^2.V_x$$

Démonstration : laissée en exercice

Exemple-exercice :
La station météo enregistre les températures maximum suivantes pour quatre villes :

Mois	Brest	Nice	Aurillac	Strasbourg
Janvier	9.3	13.1	4.5	6.7
Février	9.5	13.4	6.4	7.9
Mars	11.5	15.2	11.4	11.2
Avril	13.2	17	15.7	13.7
Mai	16.2	20.7	20.2	17.8
Juin	18.7	24.3	23.4	21.5
Juillet	20.7	27.3	25.7	24.3
Août	20.8	27.7	25.4	24
Septembre	19.1	24.6	21	20.3
Octobre	15.7	21	15.3	16
Novembre	12.2	16.6	8.8	10.4
Décembre	9.9	13.8	5.2	7.5

1. Retrouver les calculs de l'écart-type de chaque températures des villes observées (les réponses sont données dans la première partie).
2. Commenter les écarts.

Exercice : interprétation de l'écart-type :
Comparer la moyenne et l'écart-type des deux séries, interpréter :

nombre d'essais par match	0	1	2	3	4	5
nombre de matches pour l'équipe 1	4	6	2	1	2	1
nombre de matches pour l'équipe 2	7	2	1	3	2	1

3 Exemple-exercice

Exemple-exercice :
Les médias nous annoncent sans cesse des nouvelles extraordinaires et font de chaque année une année d'événements records : année la plus chaude, la plus pluvieuse, nombre record de catastrophes aériennes, etc. S'agit-il de phénomènes aléatoires, ou bien le XXe siècle était-il un siècle de catastrophes ? Nous allons donner un élément de réponse. Étudions un phénomène quantifiable (par exemple la hauteur de pluie tombée pendant une année en un endroit donné) pendant n années consécutives et écrivons les différents résultats sous la forme d'une liste de nombres présentée entre crochets. Nous dirons qu'un nombre de cette liste est un record lorsque celui-ci est strictement supérieur à tous les nombres qui le précèdent.
Remarque : un premier résultat d'une liste est par convention un résultat record. Exemple : la liste [12 ; 14 ; 11 ; 15] contient 3 records : 12, 14 et 15.

1. Combien y a-t-il de records dans la liste : [5 ; 10 ; 11 ; 4 ; 8 ; 15 ; 20 ; 12 ; 21 ; 6 ; 12 ; 5 ; 14 ; 18] ?
2. Quel est le nombre minimal de records d'une liste ? Donner un exemple de liste à 5 termes ayant un nombre minimal de records.
3. Quel est le nombre maximal de records d'une liste de n termes ? Donner un exemple de listes à 5 termes ayant un nombre maximal de records.
4. On a simulé à l'aide d'un ordinateur 40 fois de suite une liste de 100 nombres aléatoires et calculé le nombre de records de ces 40 listes. Voici les nombres de records obtenus : 9, 7, 5, 6, 3, 4, 3, 10, 4, 4, 4, 5,13, 8, 7, 2, 7, 5, 6, 14, 8, 2, 3, 11, 4, 2, 3, 2, 5, 11, 2, 4, 9, 9, 7, 6, 4, 5, 8, 10.
   (a) Quel est le nombre moyen de records par siècle que nous donne cette simulation?
   (b) Déterminer l'écart-type associé à la série. Vous semble-t-il élevé ?
   (c) Déterminer les quartiles de la série.
5. Sur une série de 10000 listes de 100 nombres aléatoires, on trouve les résultats suivants : commenter. (Réalisé avec Python)

```
Nombre de records : [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]
effecifs : [184, 703, 1680, 2214, 2074, 1583, 910, 414, 164, 56, 14, 2, 2]
moyenne :  4.727599999999999
écart-type : 1.7653889769679656
Q1 3
Q2 5
Q3 6
```

Note : une partie du programme principal, en Python, pour déterminer une liste de record pour une liste donnée :

```
#simulation de records

import random                         #importation du module du hasard

###initialisation des variables
liste=[]
tailleliste=100
listeRecord=[]
nombreRecord=0
###début du programme
for i in range(0,tailleliste,1):      #boucle Pour i allant de 0 à tailleliste avec un pas de 1
    liste.append(random.randint(1,100))   #.append permet d'ajouter un élément à une liste
                                      #fin de la boucle Pour, retour de l'indentation
listeRecord.append(liste[0])
j=1

for i in range(1,tailleliste,1):
    if liste[i]>listeRecord[j-1] :    #indentation pour la condition if
        j=j+1
        listeRecord.append(liste[i])
                                      #fin de la condition if, fin de la boucle Pour
nombreRecord=len(listeRecord)
print(liste)                          #affichage par print
print(listeRecord)
print(nombreRecord)
```