**STATISTIQUE DESCRIPTIVE ET ANALYSE DE DONNEES**

**I) Représentation graphique**

Le **diagramme circulaire** est utilisé pour représenter une série dont le caractère est qualitatif.

Exemple

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Saison préférée des élèves :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Saison | Hiver | Automne | Printemps | Eté |
| Pourcentage | 30% | 15% | 10% | 45% |

 |  |

Le **diagramme en bâtons** est utilisé pour représenter une série dont le caractère est quantitatif discret ; il met en évidence le mode de la série.

Exemple

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Résultats du lancer d’un dé :

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Faces obtenues | 1 | 2 | 3 | 4 | 5 | 6 |
| Effectif | 32 | 28 | 38 | 31 | 39 | 30 |

 |  |

L’**histogramme** est utilisé pour représenter une série dont le caractère est quantitatif continu.

Exemple

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Taille des élèves d’une classe :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Taille (m) | 1,61,65 | 1,651,7 | 1,71,75 | 1,751,8 |
| Effectif | 8 | 8 | 15 | 18 |
| Taille (m) | 1,81,85 | 1,851,9 | 1,91,95 | 1,952 |
| Effectif | 7 | 4 | 4 | 1 |
| Taille (m) | 22,05 |
| Effectif | 1 |

 |  |

Le **polygone des effectifs cumulés croissants** pour une série regroupée en classes est formé des segments reliant les points ayants pour abscisse l’extrémité droite xi de chaque classe et pour ordonnée Ni l’effectif cumulé croissant en xi.(pour les effectifs cumulés décroissants, on prend pour abscisse l’extrémité gauche xi de chaque classe.)

Exemple

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Salaires dans une entreprise :

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Salaire | 800 | 900 | 1000 | 1100 | 1200 | 1300 |
| Effectifs | 0 | 42 | 49 | 75 | 18 | 16 |
| Effectifs cumulés croissants | 0 | 42 | 91 | 166 | 184 | 200 |

 |  |

**II) Médiane, quartiles et déciles**

 1) Médiane

Définition

La **médiane** est la valeur qui se trouve au MILIEU de la série, qui la partage en deux séries d’effectif égal. Les valeurs de la série doivent être rangées par ordre croissant ! La médiane est un **paramètre de tendance centrale**.

Exemple

Cette série statistique porte sur l’âge des joueurs de l’équipe de France vice-championne du monde en 2006 :

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Âge** | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | **TOTAL** |
| **Effectif** | 0 | 1 | 2 | 0 | 1 | 0 | 4 | 0 | 5 | 2 | 1 | 2 | 2 | 0 | 1 | 1 | **22** |
| **E C C** | 0 | 1 | 3 | 3 | 4 | 4 | 8 | 8 | 13 | 15 | 16 | 18 | 20 | 20 | 21 | 22 |  |

Il y a 22 joueurs concernés par l’étude.

. La médiane est donc la demi somme des 11ème et 12ème termes.

 , la médiane de cette série statistique est de 28 ans.

Exemple

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Valeur** | 1 | 2 | 3 | 4 | 5 | 6 |
| **Effectif** | 6 | 11 | 25 | 19 | 15 | 5 |
| **Effectif cumulé** |  |  |  |  |  |  |

L’effectif total est 81.

. La médiane est donc la 41ème valeur de cette série.

 , la médiane de cette série statistique est 3.

Exemple

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Classe** | **[ 0 ; 2 [** | **[ 2 ; 4 [** | **[ 4 ; 6 [** | **[ 6 ; 8 [** |
| **Fréquence** | 10 % | 38 % | 45 % | 7 % |
| **Fréquence cumulée** | 10 % | 48 % | 93 % | 100 % |

La classe médiane est donc la classe [ 4 ; 6 [.

 2) Les quartiles

Définitions

Le **premier** **quartile** est la plus petite valeur donnée **Q1** de la liste telle qu’**au moins un quart** des données de la liste sont inférieures ou égales à Q1.

Le **troisième quartile** est la plus petite valeur donnée **Q3** de la liste telle qu’**au moins trois quarts** des données de la liste sont inférieures ou égales à Q3.

Exemple

Si on reprend le tableau « age équipe de France ».

 . On s’intéresse alors à la 6ème valeur de cette série et donc Q1 = 26.

. On s’intéresse alors à la 17ème valeur de cette série et donc Q3 = 31.

Exemple

Si on reprend le tableau « valeur ».

. On s’intéresse alors à la 21ème valeur de cette série et donc Q1 = 3.

. On s’intéresse alors à la 61ème valeur de cette série et donc Q3 = 4.

 3) Ecart interquartile

Définitions

Le nombre Q3 - Q1 s’appelle **l’écart** **interquartile**.

L’écart interquartile est un **paramètre de dispersion**.

Exemple

Si on reprend le tableau « age équipe de France ».

Q1 = 26 Q3 = 31

. Ce qui signifie que l’écart interquartile est de 5.

Exemple

Si on reprend le tableau « valeur ».

Q1 = 3 Q3 = 4

. Ce qui signifie que l’écart interquartile est de 1.

 4) Intervalle interquartile

Définition

On appelle **intervalle interquartile** l’intervalle .

Propriété

L’intervalle interquartile  contient au moins 50 % des termes de la série statistique.

Exemple

Si on reprend le tableau « age équipe de France ».

Q1 = 26 Q3 = 31

L’intervalle  contient 4 + 5 + 2 + 1+ 2 = 14 valeurs soit  des données.

 5) Déciles

Définition

Les neufs **déciles**, notés D1, D2, …, D9, sont les valeurs de la liste qui partagent l’effectif total en dix groupes de même effectif.

Exemple

Si on reprend le tableau « age équipe de France ».

 . On s’intéresse alors à la 3ème valeur de cette série et donc D1 = 22.

. On s’intéresse alors à la 20ème valeur de cette série et donc D9 = 32

 6) Ecart interdécile

Définition

Le nombre D9 - D1 s’appelle **l’écart** **interdécile**.

L’écart interdécile est un **paramètre de dispersion**.

Exemple

Si on reprend le tableau « age équipe de France ».

D1 = 22 D9 = 32

. Ce qui signifie que l’écart interdécile est de 10.

 7) Intervalle interdécile

Définition

On appelle **intervalle interdécile** l’intervalle .

Propriété

L’intervalle interdécile  contient au moins 80 % des termes de la série statistique.

Exemple

Si on reprend le tableau « age équipe de France ».

D1 = 22 D9 = 32

L’intervalle  contient 2 + 1 + 4 + 5 +2 + 1 + 2 + 2 = 19 valeurs soit  des données.

**III) Diagramme en boîte**

Pour visualiser rapidement la répartition de données autour de la médiane et aussi pour comparer entre elles plusieurs séries, on utilise souvent un **diagramme en boîte**.

Définition

Le diagramme en boîte d’une série à l’allure suivante :



Remarque

La boîte centrale représente **l’intervalle** **interquartile** et contient donc la moitié des données.

Exemple





**IV) Moyenne, variance et écart type**

 1) Moyenne

Définition

Soit *x* un caractère qui prend les valeurs *x1, x2, … xp*. Les effectifs respectifs de chaque valeur sont *n1, n2, … np*. L’effectif total est donc  *= N.* Alors la **moyenne** de cette série statistique est :



La moyenne est un **indicateur de tendance centrale**.

Exemple

Dans une classe, 7 élèves ont eu 12, 3 élèves ont eu 15, 5 élèves ont eu 9 et 1 seul élève a eu 18.

Alors la moyenne de la classe est = ×××× = = 12

 2) Ecart type et variance

Définition

**L’écart type** est . L’écart type est un **paramètre de dispersion**.

Définition

La variance d’une série statistique est . La variance est un **paramètre de dispersion**.

Remarque

L’avantage de l’écart type sur la variance est qu’il s’exprime comme la moyenne dans la même unité que les données.

**V) Résumé d’une série statistique**

On résume souvent une série statistique par un paramètre de tendance centrale associé à un paramètre de dispersion. Deux choix sont couramment proposés : **le couple « moyenne et écart type »** qui a l’inconvénient d’associer deux paramètres sensibles aux valeurs extrêmes et **le couple « médiane et écart interquartile »** qui n’a pas ce défaut mais dont la détermination est moins pratique.