**1.3   Les échelles de mesures.**

Pour les variables qualitatives, il existe deux échelles de mesure. **L’échelle nominale** qui s’adresse aux variables qualitatives nominales, elle ne sert qu’à coller une étiquette aux unités statistiques, elle ne les classe pas sur une échelle à une dimension.

**Exemple 1.3.1 :**

* X= sexe, alors X est une variable qualitative nominale et son échelle est nominale.
* Y=le numéro du dossard d’un joueur de hockey. Même si Y prend des valeurs numériques, ce n’est qu’une variable nominale et son échelle est nominale. Car on peut tout aussi bien mettre des lettres sur leur dossard ou des dessins.

L’autre **échelle est l’échelle ordinale** et s’adresse aux variables qualitatives ordinales, on l’appelle comme cela car il y a un ordre entre ses modalités.

**Exemple 1.3.2 :**

* X= le niveau scolaire d’une personne adulte, alors ses modalités peuvent être : primaire, secondaire, collégial, universitaire. Il y a un ordre chronologique entre ces modalités.
* Y= la note finale obtenue dans un cours de statistique, ses modalités seront : F, E, D, C, B, A ou A+. Il y a un ordre de mérite entre ces modalités.

Pour les variables quantitatives, il existe aussi deux types d’échelles, la première échelle est **l’échelle d’intervalle**. On l’appelle comme ça car la seule opération possible est la différence. On reconnaît une échelle d’intervalle par l’absence du zéro absolu (c'est-à-dire que si X=0, cela ne veut pas dire absence de ce qu’on mesure).

**Exemple 1.3.3 :**

* T= la température en degrés Celsius. Le jour où T=0 14â„ƒ"> , ça ne veut pas dire absence de température. Si on considère deux journées où la température  est respectivement égale à 10 et 30 degrés, ça veut seulement dire qu’il y  a un écart de 20 degrés entre ces deux journées. Si on prend deux sots d’eau où la température est respectivement égale à 35 et 45 degrés, si on les mélange, on ne va pas obtenir une eau chauffée à 80 degrés. Alors l’échelle de cette variable est une échelle d’intervalle.
* X=la date de naissance, si on est en 2010 et qu’on considère une personne née en 1950 et une autre née en 1980, tout ce qu’on peut dire est qu’il y a une différence d’âge de 30 ans entre elles. On ne peut pas dire que l’une est deux fois plus âgée que l’autre, car l’année prochaine ce ne serait plus vrai. Alors l’échelle de cette variable est une échelle d’intervalle.

L’autre échelle est **l’échelle de rapports**. C’est l’échelle la plus maniable, la plus riche. Elle admet un zéro absolu, c'est-à-dire si la variable est nulle, cela signifie l’absence de ce qu’on mesure. On peut faire toutes les opérations algébriques avec une telle échelle.

**Exemple : 1.3.4 :**

* X=le revenu familial annuel (en dollars), si X=0 cela veut dire qu’il n’y a pas eu de revenu. Si on prend deux familles dont le revenu respectif est de 30 000 et 120 000 dollars, on peut dire qu’il y a un écart de 90 000 dollars entre ces deux revenus, on peut aussi dire que la deuxième famille gagne 4 fois plus que la première. Si on additionne ces deux revenus, on aura un revenu global de 150 000 dollars. Alors l’échelle de cette variable est une échelle de rapports.
* Y=le nombre d’enfants dans un ménage. Si Y=0 cela veut dire que cette famille n’a pas d’enfant. On peut faire toutes les opérations algébriques avec les modalités de cette variable, donc son échelle est une échelle de rapports.

**1.4   Les tableaux et graphiques.**

Dans ce paragraphe on va détailler comment résumer l’information contenue dans une série de données soit par des tableaux ou des graphiques. On va commencer par les variables qualitatives.

**1.4.1    Cas de variables qualitatives.**

On va considérer deux exemples où on a des variables qualitatives observées sur un échantillon et suivre le traitement possible de ces données.

**Exemple 1.4.1.1 :** On a pris un échantillon de 50 achats de boissons non-alcoolisées achetées dans une grande surface, en notant par :

CC=Coca-Cola; S=Sprite; CL=Coke-Light; P=Perrier; PC=Pepsi-Cola. On a obtenu les résultats suivants.

1. S  PC  CL  CC  CC  PC  CL  CC  CL  CC  CC  CC  CL  PC  CC
2. P  P  S  CC  CL  PC  CL  PC  CC  PC  PC  CC  PC CC  CC  PC
3. PC  PC  S  CC  CC  CC  S  P  CL  P  PC  CC  PC  S  CC  CL

Alors ici la variable est X=Boisson non-alcoolisée, qui est une variable qualitative nominale. Pour présenter ces données sous forme de tableau, on dresse un tableau, dans la première colonne on énumère les cinq modalités de la variable, dans la seconde colonne on donne la f**réquence absolue**ou l’effectif de chacune des modalités (c'est-à-dire  le  nombre de fois que cette modalité se répète dans l’échantillon) et dans la troisième colonne, on donne **la fréquence relative** de chacune des modalités. La **fréquence relative** d’une modalité étant égale à sa fréquence absolue divisée par la taille de l’échantillon. Ce qui donne :

|  |  |  |
| --- | --- | --- |
| **Tableau des fréquences des boissons non-alcoolisées** | | |
| X=Boisson | Fréquences absolues | Fréquences relatives |
| CC | 19 | 0,38 |
| CL | 8 | 0,16 |
| PC | 13 | 0,26 |
| P | 5 | 0,10 |
| S | 5 | 0,10 |
| Total | n=50 | 1 |
| 1. : données fictives. | | |

     Ce tableau s’appelle **tableau de fréquences** de la variable.

**Remarque :** Pour une présentation complète des tableaux et graphiques, on doit mettre le titre en haut et la source des données en bas.

En ce qui concerne la représentation graphique, on va donner deux graphiques qui résument la même information contenue dans le tableau des fréquences.

* Le diagramme à barres (horizontales ou verticales). Où on met sur un axe les modalités de la variable et sur l’autre axe les fréquences absolues ou les fréquences relatives.

|  |
| --- |
| Répartition des ventes des boissons non alcoolisées selon la marque |

**Remarque :** Les largeurs des barres doivent être les mêmes pour une belle esthétique du graphique, ainsi que la distance entre les bandes. On peut aussi ajouter les fréquences absolues au dessus des bandes.

* Le deuxième graphique qu’on peut faire est **le diagramme à secteurs** (ou circulaire) qui est une sorte de tarte où chaque modalité occupe une partie qui reflète sa fréquence relative.

|  |
| --- |
| Diagramme circulaire donnant la répartition des boissons non alcoolisées selon la marque |

**Exemple 1.4.1.2 :**Lors d’une enquête de satisfaction de la clientèle, une compagnie de courtage a demandé à un échantillon de 60 clients d’indiquer leur degré de satisfaction vis-à-vis de leur conseiller financier, sur une échelle de 1 à 7, le 1 correspondant à <<pas du tout satisfait>> et le 7 correspondant à << extrêmement satisfait>>. On a obtenu les résultats suivants :

1. 7  6  6  7  5  5  7  3  6  7  7  6  6  6  5  5  6  7  7

6  6  4  4  7  6  7  6  7  6  5  7  5  7  6  4  7  5  7  6

6  5  3  7  7  6  6  6  6  5  5  6  6  7  7  5  6  6  6  6

Ici  la variable, ``degré de satisfaction`` est une variable qualitative ordinale. On peut résumer l’information contenue dans ces données sous forme d’un tableau de fréquences ce qui donne :

|  |  |  |
| --- | --- | --- |
| **Tableau des fréquences du degré de satisfaction des clients.** | | |
| Degré de satisfaction | Fréquences absolues | Fréquences relatives |
| 1 | 0 | 0,0000 |
| 2 | 0 | 0,0000 |
| 3 | 2 | 0,0333 |
| 4 | 3 | 0,0500 |
| 5 | 12 | 0,2000 |
| 6 | 25 | 0,4167 |
| 7 | 18 | 0,3000 |
| Total | n=60 | 1,0000 |
| 1. : Données fictives. | | |

En ce qui concerne la représentation graphique, les mêmes graphiques qu’on a utilisés pour une variable qualitative nominale font l’affaire. Ce qui donne :

|  |
| --- |
| Degré  de satisfaction |

|  |
| --- |
| Répartition du degré de satisfaction des clients |

|  |
| --- |
| Diagramme circulaire, donnant la répartition du degré de satisfaction des clients. |

**1.4.2 Cas de variables quantitatives.**

Le traitement des variables quantitatives discrètes étant différent de celui des variables quantitatives continues, on va donc réserver un sous paragraphe à chacune d’elles.

**1.4.2.1 : Cas des variables quantitatives discrètes.**

Soit X une variable quantitative discrète dont le nombre de modalités n’est pas trop grand. Alors on peut dresser un tableau des fréquences comme celui utilisé pour les variables qualitatives auquel on peut ajouter une colonne supplémentaire où on met les fréquences relatives cumulées au fur et à mesure qu’on ajoute une modalité de la variable. En ce qui concerne la représentation graphique, un seul graphique s’associe avec les variables quantitatives discrètes : **le diagramme à bâtons.**

**Exemple 1.4.2.1.1** : Un inspecteur en contrôle de qualité a extrait de sa base de données, un échantillon de 40 semaines où il a noté X, le nombre d’accidents de travail enregistrés par semaine. Il a obtenu les résultats suivants :

1. 0  4  2  2  1  3  2  0  5  4  3  2  4  5  6  6  4  2  0
2. 4  4  2  6  2  4  3  0  4  3  4  3  3  5  5  4  2  2  1

On peut donc dresser le tableau des fréquences suivant.

|  |  |  |  |
| --- | --- | --- | --- |
| **Tableau des fréquences du nombre d’accidents par semaine** | | | |
| Le nombre d’accidents par semaine. | Fréquences absolues | Fréquences relatives | Fréquences relatives cumulées |
| 0 | 4 | 0,100 | 0,100 |
| 1 | 2 | 0,050 | 0,150 |
| 2 | 10 | 0,250 | 0,400 |
| 3 | 7 | 0,175 | 0,575 |
| 4 | 10 | 0,250 | 0,825 |
| 5 | 4 | 0,100 | 0,925 |
| 6 | 3 | 0,075 | 1,000 |
| Total | n=40 | 1,000 |  |

Quant au diagramme à bâtons, on obtient quelque chose comme :

|  |
| --- |
| Répartition du nombre d’accidents par semaine. |

|  |
| --- |
| Le nombre d’accidents par semaine |

**Remarque :** Les bâtons ne doivent pas avoir d’épaisseur, car la variable prend exactement les valeurs 0, 1, 2,…On peut ajouter les effectifs ou les fréquences relatives sur les bâtons.

**1.4.2.2 : Cas de variables quantitatives continues.**

Considérons maintenant un échantillon de données provenant d’une variable quantitative continue ou discrète avec un grand nombre de modalités. Il est donc inconcevable de dresser un tableau où on énumère les modalités d’une telle variable, il serait non analysable. Il faut donc grouper ces données en classes de valeurs. Deux questions se posent alors :

* Combien de classes faut-il former ?
* Quelles seront les largeurs de chacune des classes ?

La réponse à la première question, dépend de la taille de l’échantillon, le nombre de classe à former est donné par la formule de Sturges suivante :

14Le nombre de classes:K=1+103Log(n)"> . Ainsi, par exemple, si n=150, il faut former14K=1+103Log150=8,2536â‰…9">  (on arrondit à l’entier immédiatement supérieur). Une fois qu’on sait combien de classes à former. On essaie de former des classes de même amplitude (largeur) et cette amplitude sera égale à

  14A=laplusgrandeobservation-lapluspetiteobservationK=Xmax-XminK"> .

On arrondit cette amplitude selon les données pour avoir des bornes de classes faciles à manipuler.

**Exemple 1.4.2.2.1 :** Soit X, les recettes  quotidiennes(en dollars) d’un petit magasin. On a sélectionné un échantillon de taille n=40 jours au hasard qui ont donné les résultats suivants :

16,00   58,50   68,20   78,00     79,45     142,20   145,3   186,70   209,05   216,75

219,70  247,75  249,10  256,00  257,15  262,35  268,60  269,60   270,15  284,45

319,00  332,00  343,29  350,75  354,90  372,60  383,20  389,20   404,55  420,20

428,50  432,40  444,60  446,80  456,10  458,10  493,95  511,95   521,05  621,35

Le nombre de classe à former est 14K=1+103log40=6,34â‰…7 classes">  d’amplitude chacune égale à 14A=621,35-167=86,48â‰…90"> . Cette amplitude est arrondie à 90. Ce qui donne le tableau des fréquences suivant, où les classes sont des intervalles fermés à gauche et ouverts à droite sauf le dernier qui est un intervalle fermé des deux côtés.

|  |
| --- |
| Répartition des 40 semaines selon les recettes hebdomadaires du dépanneur |

|  |  |  |  |
| --- | --- | --- | --- |
| X=les recettes | Fréquences absolues | Fréquences relatives | Fréquences relatives cumulées |
| [10 ; 100[ | 5 | 0,125 | 0,125 |
| [100 ;190[ | 3 | 0,075 | 0,200 |
| [190 ;280[ | 11 | 0,275 | 0,475 |
| [280 ;370[ | 6 | 0,150 | 0,625 |
| [370 ;460[ | 11 | 0,275 | 0,900 |
| [460 ;550[ | 3 | 0,075 | 0,975 |
| [550 ;640] | 1 | 0,025 | 1,000 |
| Total | n=40 | 1,000 |  |

Quand aux graphiques, on va ici préviligier trois graphiques pour les variables quantitatives continues.

* **L’histogramme**, qui est une suite de rectangles juxtaposés les uns aux autres dressés au-dessus de chacune des classes, dont la largeur est égale à l’amplitude  de la classe (prise comme unité de mesure) et dont la surface reflète la fréquence relative de la classe qu’il représente.

|  |
| --- |
| Histogramme donnant la répartition des 40 semaines en fonction des recettes hebdomadaires |

|  |
| --- |
| Fréquences relatives |

* **Le polygone des fréquences**, qui consiste à joindre le milieux des sommets des rectangles d’un histogramme par une ligne en zig-zag et cette ligne se ferme en ajoutant aux deux extrémités deux classes fictives de même amplitude que les autres, comme ça la surface délimitée par l’histogramme est identique à celle délimitée par le polygone des fréquences. Le polygone de fréquences est très utile quand on veut comparer le comportement de la même variable mesurée sur plusieurs groupes (on peut penser à comparer le revenu des hommes et des femmes)  ou la même variable mesurée sur le même échantillon à différents instants (on peut comparer le poids du même groupe à différents moments d’une diète).

|  |
| --- |
| Polygone des fréquences donnant la répartition des 40 semaines selon les recettes hebdomadaires. |

* **La courbe des fréquences cumulées (Ogive).**

Comme son nom l’indique, elle consiste à tracer le graphique des fréquences cumulées, en mettant les limites des classes sur l’axe horizontal et les fréquences cumulées sur l’axe vertical, ces dernières se cumulant à la fin de chacune des classes. Ce graphique aura l’allure d’une courbe croissante variant entre 0 et 1.

|  |
| --- |
| Ogive de la répartition des 40 semaines selon les recettes  hebdomadaires |

**Remarque :**Lorsque les classes ne sont pas de même amplitude, il faut se rappeler que la surface du rectangle d’un histogramme étant égale à sa fréquence relative à la classe associée à ce rectangle, alors si la largeur de cette classe par exemple est le double de la l’amplitude de base, la hauteur du rectangle doit être divisée par deux.

**1.5 : Les mesures de tendance centrale**

On appelle mesures de tendance centrale, des valeurs de la variable susceptibles de nous donner une idée sur la donnée qui occupe le centre d’une série statistique. On va décrire dans ce paragraphe, les trois plus importantes mesures de tendance centrale que sont **le mode, la moyenne et la médiane**.

**1.5.1.1 : Le mode**

On appelle le mode d’une variable X, la valeur de la variable qui a la plus grande fréquence et on le note Mo(X). Le mode est une importante mesure de tendance centrale pour les variables qualitatives nominales.

**Remarque :** Une distribution peut avoir un seul mode et on dit qu’elle est unimodale, ou plusieurs modes et on dit qu’elle est multimodale.

**Exemple 1.5.1.1.1 :** Si on reprend l’exemple des boissons non-alcoolisées, on avait le tableau des fréquences suivant :

|  |  |  |
| --- | --- | --- |
| **Tableau des fréquences des boissons non-alcoolisées** | | |
| X=Boisson | Fréquences absolues | Fréquences relatives |
| CC | 19 | 0,38 |
| CL | 8 | 0,16 |
| PC | 13 | 0,26 |
| P | 5 | 0,10 |
| S | 5 | 0,10 |
| Total | n=50 | 1 |

Alors, le mode de cette variable est Mo(X)=Coca-Cola (CC), cela signifie que dans cet échantillon, la boisson la plus fréquemment achetée est Coca-Cola.

**Exemple 1.5.1.1.2 :** En reprenant l’exemple des recettes quotidiennes d’un petit magasin, où la variable est quantitative continue avec des données groupèes en classes, on avait le tableau des fréquences suivant :

|  |  |  |
| --- | --- | --- |
| X=les recettes | Fréquences absolues | Fréquences relatives |
| [10 ; 100[ | 5 | 0,125 |
| [100 ;190[ | 3 | 0,075 |
| [190 ;280[ | 11 | 0,275 |
| [280 ;370[ | 6 | 0,150 |
| [370 ;460[ | 11 | 0,275 |
| [460 ;550[ | 3 | 0,075 |
| [550 ;640] | 1 | 0,025 |
| Total | n=40 | 1,000 |

Ici, on voit qu’il y a deux classes qui ont les plus hautes fréquences, on les appelle des classes modales. Alors on est en présence d’une distribution de données bimodale, et les deux modes sont les milieux des deux classes modales, à savoir Mo(X)=235 et Mo(X)=415. Cela veut dire que dans cet échantillon les recettes quotodiennes les plus fréquentes sont soit de 235$ ou de 415$. Il y a des auteurs qui font des interpolations à l’intérieur des classes modales pour trouver le mode, on estime que c’est un effort inutile, vue que dans le cas d’une variable quantitative le mode joue un rôle très marginal. On voit que le mode d’une variable est une mesure de tendance centrale facile à déterminer et s’applique à tous les types de variables, mais sa portée comme mesure d’analyse est très limitée.

**1.5.2 :  La moyenne.**

La moyenne arithmétique ou simplement la moyenne est la mesure de tendance centrale la plus connue. Elle ne s’applique qu’aux variables quantitatives. On va décrire la méthode pour calculer la moyenne d’une variable quantitative selon que les données sont en vrac, groupées par valeurs ou groupées par classes.

**1.5.2.1 : Les données en vrac.**

Soit X une variable quantitative dont les valeurs observées sur un échantillon forment une série en vrac 14x1,x2,â€¦.,xn">  alors la moyenne de cet échantillon est

14x=x1+x2+â€¦.+xnn=i=1nxin">

**Exemple 1.5.2.1.1** : Un commerçant a l’habitude de noter dans son registre le nombre de clients qui se présentent quotidiennement à son magasin. On a pris un échantillon de taille 10 de ce registre et on trouvé les valeurs suivantes :

120  105  90  201  196  65  88  163  103  116

Alors dans cet échantillon le nombre moyen des clients qui se présentent à ce magasin par jour est donné par la formule suivante :

14x=x1+x2+â€¦.+xnn=120+105+â€¦.+11610=124,7">  clients par jour.

**1.5.2.2 : Les données groupées par valeurs.**

Soit X une variable quantitative discrète dont les données se présentent sous forme d’un tableau où elles sont classées par valeurs, supposons que la taille de l’échantillon est n et qu’il y a k valeurs différentes pour cette variable. Alors la moyenne d’un tel échantillon de données est :

14x=valeur\*(safrÃ©quenceabsolue)tailledel'Ã©chantillon=i=1kxifin">

**Exemple 1.5.2.2.1 :** Reprenons les données de l’exemple 1.4.2.1.1, où X est le nombre d’accidents de travail par semaine. On avait le tableau de données suivant :

|  |  |
| --- | --- |
| **Tableau des fréquences du nombre d’accidents par semaine** | |
| X | Fréquences absolues |
| 0 | 4 |
| 1 | 2 |
| 2 | 10 |
| 3 | 7 |
| 4 | 10 |
| 5 | 4 |
| 6 | 3 |
| Total | n=40 |

Alors la moyenne de cet échantillon est égale à

14x=0\*4+1\*2+â€¦..+6(3)40=3,025">  accidents par semaine.

**1.5.2.3 : Les données groupées par classes.**

Supposons qu’on est devant un tableau où les données provenant d’un échantillon sont groupées par classes. Alors pour calculer la moyenne de cet échantillon, on va utiliser une formule approximative, où chaque classe est assimilée à son centre et on utilise la même formule que pour le cas où les données sont groupées par valeurs. Si on note par 14mi"> **,** le milieu de la ième classe et qu’on suppose que la taille de l’échantillon est n et qu’il y a k classes, alors la moyenne de l’échantillon est :

14x=i=1kmifin">

**Exemple 1.5.2.3.1 :**En reprenant l’exemple 1.4.2.2.1 où X est la recette quotidienne d’un petit magasin, on avait le tableau suivant auquel on a ajouté une colonne à gauche contenant le milieu des classes :

|  |  |  |
| --- | --- | --- |
| 14mi"> | X=les recettes | Fréquences absolues |
| 55 | [10 ; 100[ | 5 |
| 145 | [100 ;190[ | 3 |
| 235 | [190 ;280[ | 11 |
| 325 | [280 ;370[ | 6 |
| 415 | [370 ;460[ | 11 |
| 505 | [460 ;550[ | 3 |
| 595 | [550 ;640] | 1 |
|  | Total | n=40 |

Alors la moyenne de cet échantillon est :

14x=i=1kmifin=55\*5+â€¦+595\*(1)40=298dollars.">

**1.5.2.4 : Les propriétés d’une moyenne échantillonnale.**

Soit X une variable quantitative dont la moyenne échantillonnale est 14x "> et soit Y une autre variable quantitative transformée linéaire de X, c'est-à-dire que 14Y=a+b\*X">  où a et b sont des constantes réelles. Alors la moyenne échantillonnale de Y sera égale à

  14y=a+b\*x.">

On dit que la moyenne conserve la transformation linéaire entre les variables.

**Exemple 1.5.2.4.1 :**Soit X, le nombre d’heures qu’un étudiant travaille à temps partiel par semaine. Supposons qu’à partir d’un échantillon d’étudiants, on a pu trouver qu’en moyenne le nombre d’heures travaillées par ces étudiants est égale à 14x=14,5">  heures/semaine. Si le salaire horaire est de 10$ et que les patrons de ces étudiants leur offrent 30$ par semaine pour leurs déplacements, quel est le gain net moyen hebdomadaire de ces étudiants ? Posons Y, le gain net hebdomadaire de ces étudiants alors 14Y=30+10\*X"> , donc le gain moyen hebdomadaire de cet échantillon d’étudiants est égal à

  14y=30+10\*x=30+10\*14,5=175$"> .

**1.5.3 : La médiane.**

La médiane est la valeur de la variable qui divise l’échantillon en deux groupes d’égal effectif. Il y a 50% des données qui sont inférieures ou égales à la médiane et 50% des données qui sont supérieures ou égales à la médiane. La médiane se calcule pour des variables qualitatives ordinales et pour des variables quantitatives. On note la médiane d’une variable X par Med(X) ou par 14x">  . Dans ce qui suit on va décrire les façons de calculer une médiane dans les différents cas possibles.

**1.5.3.1 : Cas d’une variable qualitative ordinale**.

Puisque les modalités d’une telle variable sont déjà ordonnées par nature, alors pour déterminer la médiane, on calcule 14l=50%\*n"> , et donc

14MedX=x=x(l)+xl+12silestunentierxl+1siln'estpasunentier">

Où 14xl+1 "> signifie, l’observation occupant le rang immédiatement supérieur à 14l.">

**Exemple 1.5.3.1.1 :**Reprenons les données de l’exemple 1.4.1.2, où X est le degré de satisfaction de la clientèle, on avait le tableau suivant :

|  |  |
| --- | --- |
| **Tableau des fréquences du degré de satisfaction des clients.** | |
| Degré de satisfaction | Fréquences absolues |
| 1 | 0 |
| 2 | 0 |
| 3 | 2 |
| 4 | 3 |
| 5 | 12 |
| 6 | 25 |
| 7 | 18 |
| Total | n=60 |

Ici, n=60 et 14l=50%\*n=30estunentier"> , alors

14MedX=x=x(30)+x(31)2=6+62=6"> . Le degré de satisfaction médian de la clientèle est égal à 6. Ce qui veut dire que dans cet échantillon 50% des clients ont un degré de satisfaction de 6 ou moins et l’autre 50% un degré de satisfaction de 6 ou plus.

**1.5.3.2 : Cas de données quantitatives en vrac ou groupées par valeurs.**

On doit d’abord  ordonner les données par ordre croissant avant d’appliquer la même procédure que pour les variables qualitatives ordinales. Ci-après nous  donnerons un exemple pour chacun de ces deux cas.

**Exemple 1.5.3.2.1 :**Reprenons les données de l’exemple 1.5.2.1.1 où la variable est le nombre de clients qui se présentent quotidiennement au magasin. On  avait des données en vrac :

120  105  90  201  196  65  88  163  103  116

En les ordonnant, on aura : 65  88  90  103  105  116  120  163  196  201.

Ici, n=10 et 14l=50%\*n=5estunentier"> , alors

14MedX=x=x(5)+x(6)2=105+1162=110,5"> . Ce qui veut dire qu’à partir de cet échantillon, on peut affirmer que  dans 50% des journées, ce magasin reçoit 110 clients ou moins par jour et dans l’autre 50% des journées, il reçoit 110 clients ou plus.

**Exemple 1.5.3.2.2 : :** Reprenons les données de l’exemple 1.4.2.1.1, où X est le nombre d’accidents de travail par semaine. On avait le tableau de données où les modalités de la variable sont groupées par valeurs, qu’on va changer un peu en ajoutant une donnée supplémentaire :

|  |  |
| --- | --- |
| **Tableau des fréquences du nombre d’accidents par semaine** | |
| Nombre d’accidents par semaine | Fréquences absolues |
| 0 | 4 |
| 1 | 2 |
| 2 | 10 |
| 3 | 7 |
| 4 | 10 |
| 5 | 4 |
| 6 | 4 |
| Total | n=41 |

Ici, n=41 et14l=50%\*n=20,5n'estpasunentier"> , alors

14MedX=x=x20,5+1=x21=l'observationqui">

14occupela 21Ã¨meposition=3"> .

C’est-à-dire que dans cet échantillon, dans 50% des semaines, on observe 3 accidents ou moins par semaine et l’autre 50% des semaines, on observe 3 accidents ou plus par semaine.

**1.5.3.3 : Cas de données groupées par classes.**

Dans le cas où on dispose d’un tableau de fréquences complet (incluant les fréquences cumulées) des données groupées par classes. Il faut d’abord déterminer la classe médiane, qui est la classe où les fréquences cumulées  dépassent pour la première fois 50%. Cette classe aura la forme :

14Cm=[binf;bsup["> , alors on obtient la médiane par interpolation à l’intérieur de cette classe médiane et on obtient la formule suivante :

14MedX=x=binf+(0,5-Fm-1)fr,m\*Am">  où

14binf=borneinfÃ©rieuredelaclassemÃ©diane.">

14F(m-1)=lafrÃ©quencecumulÃ©eavantlaclassemÃ©diane.">

14fr,m=lafrÃ©quencerelativedelaclassemÃ©diane.">

14Am=l'amplitudedelaclassemÃ©diane.">

**Exemple 1.5.3.3.1 :** En reprenant les données où X donne la recette quodienne d’un petit magasin, on retrouve  le tableau des fréquences suivant :

|  |  |  |  |
| --- | --- | --- | --- |
| X=les recettes | Fréquences absolues | Fréquences relatives | Fréquences relatives cumulées |
| [10 ; 100[ | 5 | 0,125 | 0,125 |
| [100 ;190[ | 3 | 0,075 | 0,200 |
| [190 ;280[ | 11 | 0,275 | 0,475 |
| [280 ;370[ | 6 | 0,150 | 0,625 |
| [370 ;460[ | 11 | 0,275 | 0,900 |
| [460 ;550[ | 3 | 0,075 | 0,975 |
| [550 ;640] | 1 | 0,025 | 1,000 |
| Total | n=40 | 1,000 |  |

Alors ici, la classe médiane est 14Cm=[binf;bsup["> =[280 ;370[

14binf=280">          14F(m-1)=0,475">

14fr,m=0,150">        14Am=90">    ce qui donne une médiane égale à :

14x=binf+(0,5-Fm-1)fr,m\*Am=280+(0,5-0,475)0,150\*90=295$">

Ce qui veut dire qu’en se basant sur cet échantillon de données, 50% des recettes quotidiennes de ce petit magasin sont inférieures ou égales à 295$ et les autres 50% sont supérieures ou égales à 295$.

**Remarque 1 :** Le calcul de la médiane est basé sur l’ordre des observations et non sur leur valeur. Contrairement à la moyenne, la médiane est insensible aux données extrêmes. Dans le cas où les données sont très différentes, la médiane est une meilleure mesure de tendance centrale.

**Remarque 2 :** Si pour une variable X quantitative les 3 mesures de tendance centrale sont presque égales, on dit alors que la variable est symétrique et alors n’importe laquelle de ces mesures peut être utilisée comme mesure de cette tendance centrale. S’il y a un grand écart entre ces mesures alors c’est la médiane qu’on doit priviligier.

**1.6 : Les mesures de position.**

On a déjà parlé de la médiane comme mesure de tendance centrale, mais elle est aussi une mesure de position car elle permet de diviser une série d’observations en deux groupes chacun contenant 50% de données. On va définir d’autres mesures de position qui permettent d’autres découpages d’une série d’observations.

**1.6.1 : Les quartiles**. Lorsqu’on veut diviser les données en quatres groupes, chacun contenant 25% des observations, on utilise des mesures appelées quartiles.

14Q1"> =le 1er quartile, à sa gauche il y a 25% des observations, qu’on note 14Q1.">

14Q2"> =le 2ème quartile, coincide avec la médiane, qu’on note 14Q2=MedX.">

14Q3"> =le 3ème quartile, à sa gauche il y a 75% des observations, qu’on note 14Q3.">

On va décrire la façon de les calculer, dans les 3 cas possibles pour une variable quantitative.

**1.6.1.1 : Les données en vrac**.  On suit les étapes suivantes.

Étape 1 : On ordonne les données par ordre croissant.

Étape 2 : On calcule l’indice 14l=i%\*(n)">  où i est le pourcentage correspondant à la mesure voulue et n  est le nombre d’observations.

Étape 3 : (a) si 14l">  n’est pas un entier, alors le ième quartile est égal à l’observation occupant la position immédiatement supérieure à 14l"> .

(b) si 14l "> est un entier, alors le ième quartile est la moyenne des observations occupant les positions 14l">  et 14l+1.">

**Exemple 1.6.1.1.1** : n=12 et les observations sont :

-2  -3  10  12  120  11  4  8  6  13  130  200.

Étape 1 : -3  -2  4  6  8  10  11  12  13  120  130  200.

Étape 2 : Si on veut déterminer 14Q1"> , on calcule 14l1=25%\*n=3"> .

Si on veut déterminer 14Q2"> , on calcule 14l2=50%\*n=6"> .

Si on veut déterminer 14Q3"> , on calcule 14l3=75%\*n=9"> .

Étape 3 : Puisque 14l1">  est un entier alors 14Q1=la 3Ã¨meobs+la 4Ã¨meobs2=4+62=5.">

Puisque 14l2">  est un entier alors 14Q2=la 6Ã¨meobs+la 7Ã¨meobs2=10+112=10,5.">

Puisque 14l3">  est un entier alors 14Q3=la 9Ã¨meobs+la 10Ã¨meobs2=13+1202=66,5.">

**Exemple 1.6.1.1.2** : n=10 et les observations sont :

1. 10  12  8  6  100  15  6  3  14.

Étape 1 :  3  3  6  6  8  10  12  14  15  100

Étape 2 : Si on veut déterminer 14Q1"> , on calcule 14l1=25%\*n=2,5"> .

Si on veut déterminer 14Q2"> , on calcule 14l2=50%\*n=5"> .

Si on veut déterminer 14Q3"> , on calcule 14l3=75%\*n=7,5"> .

Étape 3 : Puisque 14l1">  n’est pas un entier alors 14Q1=la 3Ã¨meobservation=6.">

Puisque 14l2">  est un entier alors&nbs