

Le programme extrait du [Bulletin officiel spécial n° 6 du 28 août 2008](#)

Programme du cours de 5^{ème} pour le traitement des données

Capacités : Effectifs. *Fréquences. Classes. Tableau de données, représentations graphiques de données.

Connaissances : Calculer des effectifs, * Calculer des fréquences. Regrouper des données en classes d'égale amplitude. Lire et interpréter des informations à partir d'un tableau ou d'une représentation graphique (diagrammes divers, histogramme). Présenter des données sous la forme d'un tableau, les représenter sous la forme d'un diagramme ou d'un histogramme (dans ce cas les classes sont toujours de même amplitude).

Commentaires : Les élèves sont entraînés à lire, interpréter et représenter des données en utilisant un vocabulaire adéquat dans des contextes qui leur sont familiers. Le calcul d'effectifs cumulés n'est pas un attendu. * Les écritures $\frac{4}{10}$, $\frac{2}{5}$, 0,4 40 % sont utilisées pour désigner une fréquence: elles permettent d'insister sur les diverses représentations d'un même nombre. Le choix de la représentation est lié à la nature de la situation étudiée. L'utilisation d'un tableur permet d'enrichir ce travail en le prolongeant à des situations plus complexes que celles qui peuvent être traitées « à la main ».

Rappel sur le programme du cours de 6^{ème} pour le traitement des données

Capacités : Représentations usuelles : tableaux. Repérage sur un axe. Représentations usuelles : diagrammes en bâtons, * diagrammes circulaires ou demi-circulaires, graphiques cartésiens.

Connaissances : Lire, utiliser et interpréter des données à partir d'un tableau. Lire interpréter et compléter un tableau à double entrée.

* Organiser des données en choisissant un mode de présentation adapté : tableaux en deux ou plusieurs colonnes, tableaux à double entrée. Lire et compléter une graduation sur une demi-droite graduée, à l'aide d'entiers naturels, de décimaux, de fractions simples $\frac{1}{2}$, $\frac{1}{10}$, $\frac{1}{4}$, $\frac{1}{5}$

* ou de quotients (placement exact ou approché). Lire, utiliser et interpréter des informations à partir d'une représentation graphique simple.

Commentaires : Il s'agit d'un premier pas vers la capacité à recueillir des données et à les présenter sous forme de tableau. Ce travail doit être l'occasion de manier les instruments de tracé et de mesure. La capacité visée concerne l'aptitude à faire une interprétation globale et qualitative de la représentation étudiée (évolution d'une grandeur en fonction d'une autre). Dès la classe de 6^e, l'utilisation de calculatrices et de logiciels permet de familiariser les élèves avec le passage d'un type d'organisation, d'un type de présentation à un autre.

I] Traitement des données simples

a) Effectifs et fréquences

En statistiques (analyse et représentation des données à des fins descriptives ou prévisionnelles), les données sont parfois fournies de façon brute, par la liste de toutes les valeurs obtenues. Voici l'exemple d'une série de 40 mesures des diamètres de tubes PVC fabriqués par une usine :

12,5 – 12,4 – 12,5 – 12,4 – 12,5 – 12,5 – 12,6 – 12,4 – 12,5 – 12,5
 12,5 – 12,6 – 12,7 – 12,5 – 12,5 – 12,6 – 12,6 – 12,5 – 12,4 – 12,4
 12,5 – 12,5 – 12,5 – 12,6 – 12,5 – 12,4 – 12,6 – 12,4 – 12,5 – 12,4
 12,5 – 12,4 – 12,5 – 12,5 – 12,5 – 12,5 – 12,4 – 12,5 – 12,4 – 12,5.

Évidemment ici, le premier travail sera de construire le tableau des effectifs, en comptant méthodiquement les valeurs identiques.

Valeurs des diamètres en cm	12,4	12,5	12,6	12,7	total
Effectifs (nombre de valeurs)	11	23	6	1	40

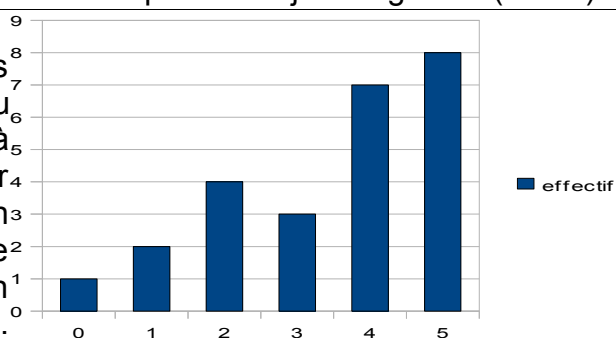
Définitions : L'*effectif* d'une valeur est le nombre d'individus (objets, personnes) qui présentent cette valeur dans la population considérée. L'*effectif total* est le nombre d'individus composant cette population. Le quotient $f = \frac{\text{effectif}}{\text{effectif total}}$ est appelé *fréquence* de la valeur considérée.

Dans notre exemple on peut ainsi calculer les fréquences relatives à chaque valeur. Le nombre obtenu est compris entre 0 et 1 (car l'effectif est toujours inférieur ou égal à l'effectif total) et il est souvent exprimé sous la forme d'un pourcentage (en multipliant la fréquence par 100).

Valeurs des diamètres en cm	12,4	12,5	12,6	12,7	total
Fréquences (en %)	$11 \div 40 \times 100 = 27,5$	57,5	15	2,5	100

Propriété : le total des fréquences des valeurs d'une série statistique est toujours égal à 1 (100%).

En ce qui concerne les données, elles sont parfois fournies par un graphique comme un *histogramme* ou un *diagramme circulaire* qui correspondent finalement à des tableaux d'effectifs ou de fréquences déguisés par la représentation en rectangles (histogrammes) ou en secteurs angulaires. Voici donc un exemple où le tableau des effectifs se déduit de la lecture d'un graphique : il s'agit des notes obtenues à un DM (sur 5).

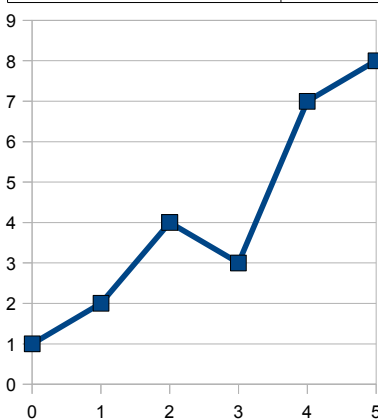


La lecture de ce graphique conduit au tableau des effectifs ou des fréquences suivant :

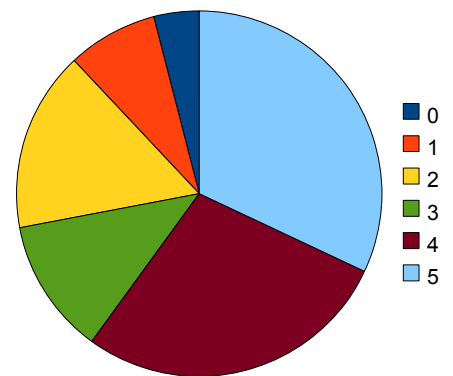
Note au DM (sur 5)	0	1	2	3	4	5	total
Effectifs	1	2	4	3	7	8	25
Fréquences (en %)	4	8	16	12	28	20	100

La représentation graphique des données est une étape importante en statistiques, car elle permet de **communiquer** des informations chiffrées d'une façon visuelle et globale. On aurait pu présenter autrement les informations concernant cette série de notes, par exemple avec un *diagramme circulaire* (un secteur représente une valeur) ou avec *une courbe* (une ligne montre l'évolution des effectifs). La confection de tels diagrammes nécessite parfois l'emploi d'un coefficient de proportionnalité, comme pour le calcul de l'angle du secteur angulaire qui est proportionnel à l'effectif (ou la fréquence) de la valeur à représenter : l'effectif total correspond alors à 360° (un tour).

Note au DM (sur 5)	0	1	2	3	4	5	total
Angle du secteur	$0,04 \times 360 = 14^\circ$	29°	58°	43°	101°	72°	$\approx 360^\circ$

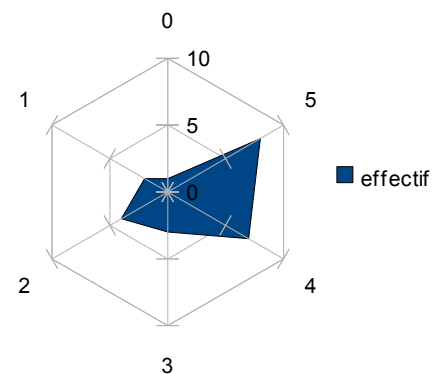


Remarque : du fait des arrondis, la somme des angles peut ne pas faire toujours 360°, mais elle doit se situer aux alentours de 360 : 358 ou 361 par exemple. De la même manière, la somme des fréquences peut ne pas toujours être égale à 100%. Par contre, les effectifs étant des nombres entiers qui ne sont pas arrondis, le total des effectifs devrait être connu exactement.



b) Utilisation d'un tableur

Les graphiques sont très faciles à faire avec un tableur. Le tableur peut servir à beaucoup d'autres choses que la confection des graphiques (sa vocation première est de calculer) mais accessoirement il est très utile pour ce travail de mise en forme. Pour information, nous utilisons le tableur de OpenOffice (scalc qui enregistre des fichiers appelés « classeur » avec l'extension .odt) qui est libre et gratuit, téléchargeable sur internet en même temps qu'un le traitement de texte et autres logiciels de bureautique, très performant et régulièrement mis à jour au même titre que d'autres ensemble logiciels mieux connus mais payants...



Rappelons comment taper une formule pour ne pas changer une valeur lors du recopiage de la formule (en tirant sur la poignée en bas à droite de la cellule sélectionnée) : il faut insérer un \$ devant la référence de ligne (un chiffre) ou de colonne (une lettre) qui ne doit pas être incrémentée. Par exemple ici, nous avons tapé B2/\$H2*360 pour le calcul de l'angle en B3. La référence \$H2 indique qu'il ne faut pas modifier la colonne en recopiant la formule : en C3 on trouvera C2/\$H2*360, etc.

B3		Σ	=	$=B2/\$H2*360$				
	A	B	C	D	E	F	G	H
1	note	0	1	2	3	4	5	total
2	effectif	1	2	4	3	7	8	25
3	angle	14	29	58	43	101	115	360

NB : si on utilise un tableur, la ligne pour le calcul de l'angle est évidemment inutile (le tableur calcule l'angle sans qu'on ait besoin de lui dire).

II] Traitement des données regroupée par classe

a) Regroupement en classes

On peut être amené à regrouper les valeurs à l'intérieur de *classes* (groupe de valeurs proches) comme dans l'exemple suivant qui donne les effectifs et salaires des employés d'une PME (Petite et Moyenne Entreprises) :

Catégorie	Ouvrier simple	Ouvrier qualifié	Cadre moyen	Cadre supérieur	Dirigeant
Effectif	50	25	15	10	2
Salaire	de 800 à 1100 €	de 1100 à 1500 €	de 1500 à 2500 €	de 2500 à 5000 €	de 5000 à 11000 €

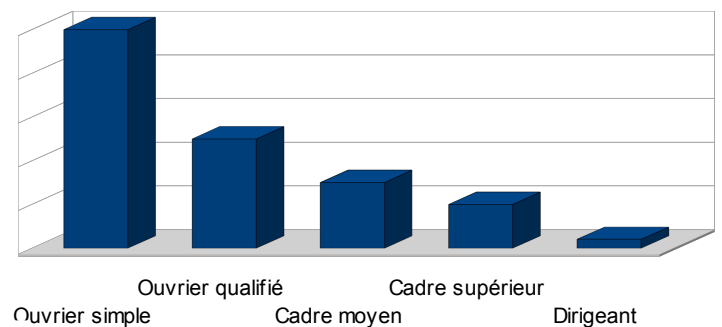
La valeur statistique étudiée ici est le salaire des employés. Chaque catégorie nominale (désignée par un nom) comme « Ouvrier simple » ou « Dirigeant » correspond à une fourchette de salaires, un intervalle à l'intérieur duquel sont situés les salaires des différents individus composant la classe considérée.

On peut se demander à quelle catégorie nominale appartient un employé qui gagnerait 1100 €, car cette valeur se situe comme limite supérieure de la catégorie « Ouvrier simple » et comme limite inférieure de la catégorie « Ouvrier qualifié ». En réalité, pour éviter ce genre de problème de localisation, on dira que la classe « Ouvrier simple » concerne les employés ayant un salaire allant de 800 € inclus à 1100 € exclu. On notera aussi cela $800 \leq \text{salaire} < 1100$ ou avec une notation spéciale $[800 ; 1100[$ qui signifie la même chose (le crochet tourné vers l'intérieur de l'intervalle pour une valeur incluse, tourné vers l'extérieur pour une valeur exclue).

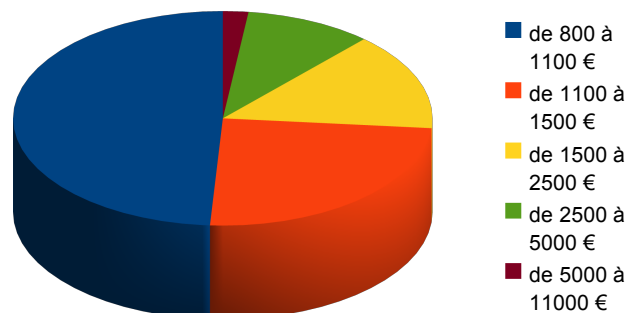
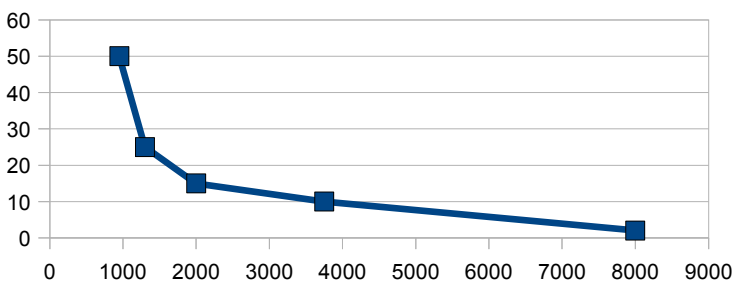
Ici encore, on va vouloir présenter les données sous la forme graphique. Que ce soit un histogramme ou un diagramme circulaire, on doit être en mesure de tirer l'information utile concernant les valeurs extrêmes des classes ainsi que les effectifs. La légende et le titre prennent toute leur importance du fait que l'on doit comprendre sans ambiguïté ce qui a été représenté.

Répartition des employés d'une PME

Tableau des effectifs



Salaire par catégorie



Remarque : la courbe ci-dessus apporte une information sur le salaire moyen touché par les employés d'une catégorie, ce qui permet de mieux se rendre compte de la répartition des salaires dans la PME. Un « Ouvrier simple » gagnant entre 800 et 1100 €, on peut considérer qu'il gagne en moyenne 950 € (on calcule la moitié de la somme de 800 et 1100). Car le regroupement sous forme de classes fait perdre l'information des valeurs individuelles. On doit pouvoir se faire pourtant une idée du salaire d'un individu d'une classe (pour calculer la moyenne des salaires ou la part des salaires par catégorie).

b) Utilisation des fréquences

Les fréquences sont, en fait, des fractions calculées à partir des effectifs. Avec la formule $f = \frac{\text{effectif}}{\text{effectif total}} \times 100$, on utilise des fractions de dénominateur 100 (des pourcentages). Mais on pourrait tout aussi bien utiliser des fractions simplifiées (dont le dénominateur est le plus petit entier possible) comme dans l'exemple suivant qui donne la répartition des élèves d'une classe de 5^{ème} de collège selon leur âge en juin. Les données brutes sont celles-ci : 13-13-12-13-13-12-12-12-12-12-13-13-12-13-12-14-13-12-13-12-12-13-12-12-12-12-13-13-12-12-12-12-13-12.

F37 =NB.SI(F\$3:F\$36;12)

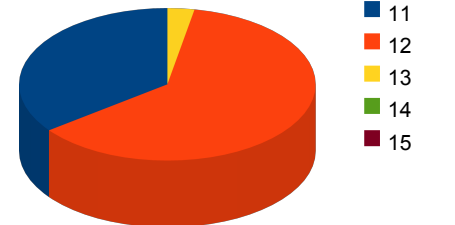
	Age
3	13
4	13
5	12
6	13
7	13
8	12
9	12
10	12
11	12
12	12
13	13
14	13
15	12
16	13
17	12
18	14
19	13
20	12
21	13
22	12
23	12
24	13
25	12
26	12
27	12
28	12
29	13
30	13
31	12
32	12
33	12
34	12
35	13
36	12
37	20
38	13
39	1

On obtient le tableau des effectifs suivants où les fréquences ont été exprimées sous la forme d'une fraction irréductible. On peut dire que 10 élèves sur 17 ont 12 ans. Cela revient au même de dire 20 élèves sur 34 (mais les nombres sont plus importants) ou que les élèves de 12 ans représentent un peu moins de 60% de la classe. La fréquence décimale plus précise de 0,5882352941 apporte une information moins lisible et presque ridicule avec sa très grande précision.

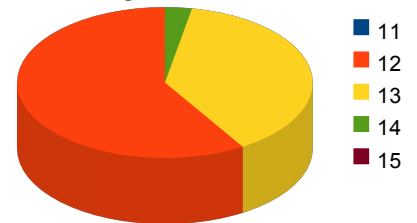
âge fin 5ème	11	12	13	14	15	Total
effectif	0	20	13	1	0	34
fréquence	0	10/17	13/34	1/34	0	1

Remarque sur l'utilisation du tableur : les effectifs par classe d'âge de cette population sont aisés à calculer. Mais le tableur peut être une aide précieuse pour effectuer ces décomptes laborieux (lorsqu'en effectuant l'opération manuellement on n'obtient pas le total attendu, il faut tout recompter...). Ici, nous avons sommé dans le tableau des données brutes, en utilisant la fonction NB.SI() qui requiert dans les parenthèses une plage de données suivie de ; et de la valeur concernée. Par exemple, dans la cellule F37 nous avons calculé l'effectif des élèves ayant un âge égal à 12 ans. Pour cela, nous avons tapé la formule =NB.SI(F\$3:F\$36;12) pour dire que nous voulons le Nombre de valeurs dans la plage allant de F3 à F36, lorsque les valeurs sont égales à 12. L'utilisation de \$ avant le numéro de ligne permet de recopier la formule en ne changeant pas ces références de la plage, dans une autre ligne. Pour l'âge 13 par exemple, nous avons juste recopié la formule en tirant sur la poignée, et en changeant le 12 en 13 dans la cellule F38 : =NB.SI(F\$3:F\$36;13).

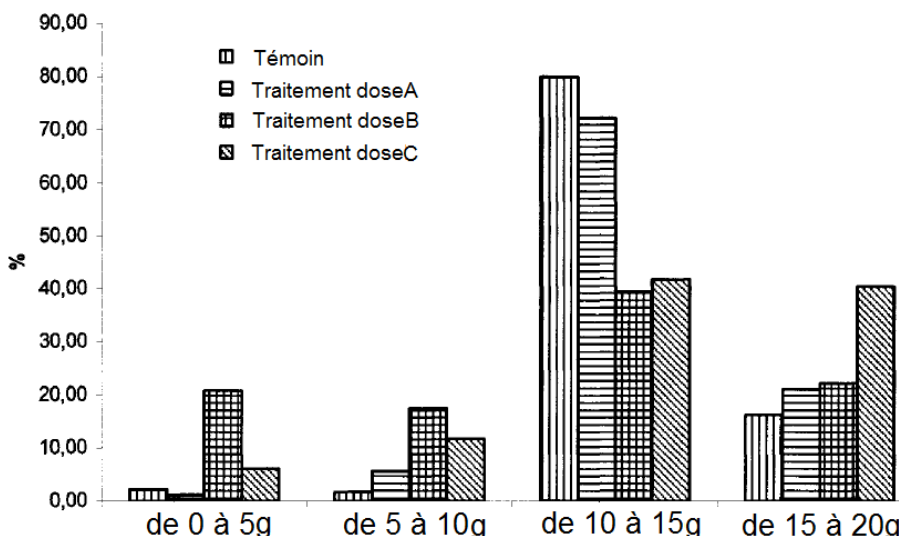
Âges en début de 5ème
classe de 5ème3 de H4 2011-2012



Âges en fin de 5ème



L'avantage des fréquences exprimée en pourcentages est de pouvoir les comparer à d'autres populations, ou la même population à d'autres moments ou en d'autres lieux. Comparons par exemple 4 groupes de rats de laboratoire, après une expérience stressante, avec une préparation particulière pour chaque population (injection d'un traitement contre l'anxiété à plus ou moins forte dose). La variable statistique étudiée est la perte de poids des animaux après l'expérience. Le graphique ci-dessous donne les résultats pour les 4 groupes. On peut comparer directement les résultats sur un graphique car ils sont exprimés en pourcentages.



On voit ainsi que le traitement à la dose A n'apporte pas un grand changement dans la réponse au stress : la perte de poids consécutive à cette expérience est principalement de 10 à 15g (près de 80% pour le groupe témoin, des animaux qui n'ont pas été traité, et juste 75% pour le groupe traité à la dose A) alors qu'elle se répartit différemment pour les animaux traités aux doses B et C (davantage d'animaux indifférents au stress et aussi, curieusement, davantage de réponses excessives pour certains).