

### Solutions des exercices de la série n°1

*N.B. les calculs et les diagrammes sont faits à l'aide de l'outil Excel de Microsoft*

**Exercice 1 :** Une enquête par sondage est menée auprès d'une population d'adultes de plus de 21 ans ; la population est constituée de  $N = 3120$  individus et le taux de sondage est de 15%. Toutes les personnes sélectionnées sont contactées par téléphone pour leur poser la question suivante : "quelle description pouvez-vous donner à votre santé physique actuelle ?" On donne dans le tableau suivant les réponses observées

Excellente      Bonne      Correcte      Mauvaise      Pas d'opinion

1. Quelle est la taille  $n$  de l'échantillon de l'enquête ?
2. Quel est le caractère étudié et quel est son type ?
3. Peut-on calculer pour ce caractère des moyennes ou qu'il est préférable d'utiliser des pourcentages ?
4. Dans l'étude on a eut 30.96% qui ont dit que leur santé était excellente. Combien de personnes ont fourni cette réponse.

#### Solution de l'exercice1

1. Le taux de sondage est

$$\frac{n}{N} = 15\% = 0.15$$

Ce qui donne  $n = 3120 * 0.15 = 468$ .

On a sélectionné 468 personnes de la population pour constituer notre échantillon.

2. Le caractère étudié est *l'état de la santé physique* de la population ciblé.  
Le type de ce caractère est qualitatif ; étant donné que les modalités observées à l'exception de la modalité " Pas d'opinion" acceptent un certain ordre, par exemple Excellente est mieux que Bonne, on peut considérer que le caractère est qualitatif ordinal.
3. On préférera toujours pour un caractère qualitatif le calcul des pourcentages même si le caractère est ordinal. Mais dans certaines études où on veut donner une note pour se repérer par rapport à d'autres études (on parle aussi de cote ou d'appréciation) les calculs de moyennes sur les modalités de la variable qualitative ordinale sont permis, dans ce cas, la modalité est codée d'une manière à refléter son positionnement dans l'ensemble des modalités.

Par exemple on peut coder ainsi

|            |       |          |          |               |
|------------|-------|----------|----------|---------------|
| 4          | 3     | 2        | 1        | 99            |
| Excellente | Bonne | Correcte | Mauvaise | Pas d'opinion |

Mais pas de cette manière :

|            |       |          |          |               |
|------------|-------|----------|----------|---------------|
| 1          | 2     | 3        | 4        | 5             |
| Excellente | Bonne | Correcte | Mauvaise | Pas d'opinion |

4. Parmi les 468 qui ont été contacté, 30.96% ont déclaré que leur santé était excellente. Si  $n_1$  est le nombre de personnes qui ont répondu de cette manière, alors

$$n_1 = 30.96\% * 468 = 144.89$$

Il y a 145 personnes qui déclarent que leur santé est excellente.

**Exercice 2 :** Le gérant d'une agence bancaire s'inquiète du nombre de comptes courants dont le solde est débiteur. Comme son système de classement lui permet d'accéder aisément à des groupes de cinq comptes, il relève, dans cent groupes distincts, le nombre de comptes débiteurs et obtient les données suivantes :

```

1 0 0 0 2 2 1 0 0 0 0 4 0 0 2 1 1 1 0 2
0 0 0 1 3 1 5 0 2 0 0 1 1 2 0 3 0 0 1 1
3 1 0 0 0 3 2 1 0 1 1 0 0 1 2 0 1 3 1 0
0 2 0 4 0 3 0 0 0 2 2 1 1 0 0 1 0 1 0 1
1 2 1 0 0 0 0 0 0 1 3 0 1 1 0 1 2 1 0 0

```

- Déterminer la population statistique, l'unité statistique, le caractère et l'effectif total.
- Quel est le type de la série statistique ? Représenter la série statistique dans un tableau statistique.
- Compléter le tableau en ajoutant la ligne des fréquences et des fréquences cumulés.
- Quelle est la fréquence des groupes ayant au plus un compte débiteur ? Quelle est la fréquence des groupes ayant plus de trois comptes débiteurs ?
- Tracer le diagramme en bâtons et la courbe cumulative de la série statistique.

### **Solution de l'exercice2**

- Dans cet exercice on comprend qu'on s'intéresse aux comptes débiteurs qui présentent un risque potentiel pour la banque, la population cible est l'ensemble des comptes de l'agence bancaire, le caractère à observer est l'état d'un compte, ce caractère peut prendre deux modalités : débiteur ou non débiteur.

La difficulté ou peut être la confidentialité pour accéder à la base des données de tous les comptes nous oblige à considérer une autre population celle des groupes de cinq comptes, l'unité statistique est le groupe de cinq comptes, le caractère à observer est le nombre de comptes débiteurs dans le groupe observé, il peut prendre les valeurs 0, 1, 2, 3, 4 et 5 ; l'effectif total est le nombre total de groupes observés, c.à.d. qu'il est égal à 100.

- Le type de la série statistique est quantitatif discret, la représentation de la série dans le tableau-synthèse est :

*Distribution des groupes selon le nombre de comptes débiteurs*

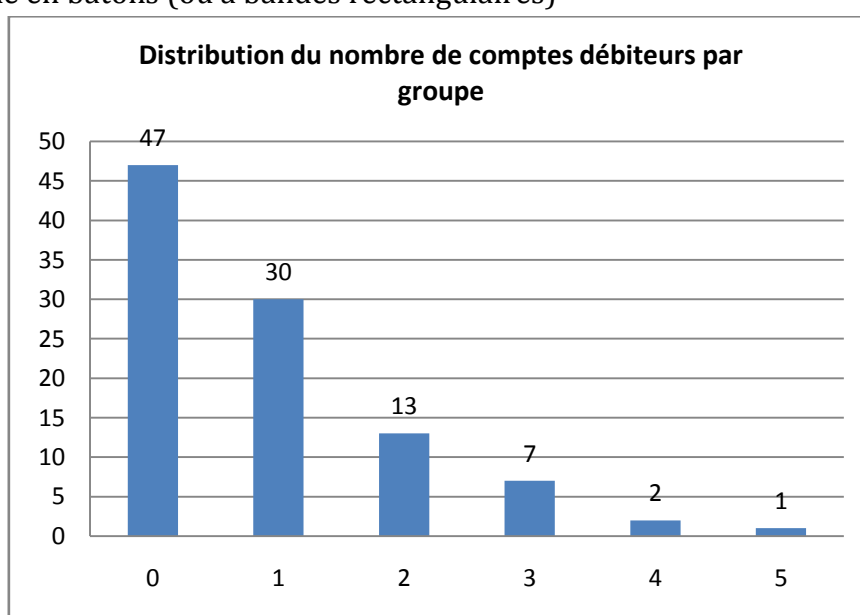
| Nombre de comptes débiteurs | Effectif $n_i$ |
|-----------------------------|----------------|
| 0                           | 47             |
| 1                           | 30             |
| 2                           | 13             |
| 3                           | 7              |
| 4                           | 2              |
| 5                           | 1              |
| Total                       | 100            |

- Ajout des fréquences et des fréquences cumulés :

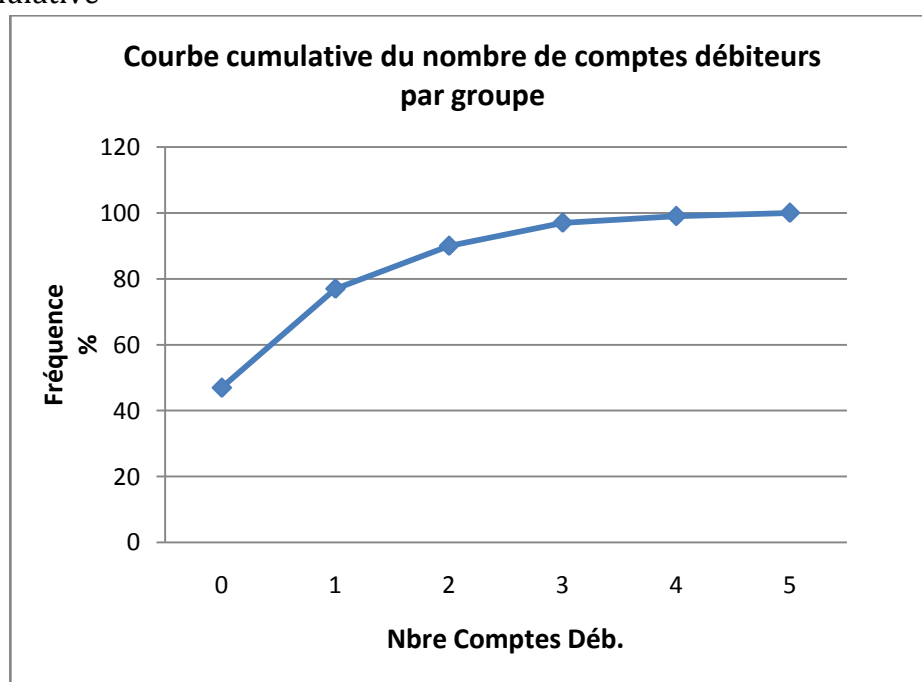
*Distribution des groupes selon le nombre de comptes débiteurs*

| Nombre de comptes débiteurs dans un groupe | Effectif<br>$n_i$ | Fréquence<br>$f_i$ | Fréquence cumulée<br>$F_i$ |
|--------------------------------------------|-------------------|--------------------|----------------------------|
| 0                                          | 47                | 0.47               | 0.47                       |
| 1                                          | 30                | 0.30               | 0.77                       |
| 2                                          | 13                | 0.13               | 0.90                       |
| 3                                          | 7                 | 0.07               | 0.97                       |
| 4                                          | 2                 | 0.02               | 0.99                       |
| 5                                          | 1                 | 0.01               | 1.00                       |
| Total                                      | 100               | 1.00               |                            |

4. La fréquence des groupes ayant au plus un compte débiteur est égale à  $0.47 + 0.30 = 0.77$ .  
 La fréquence des groupes ayant plus de trois comptes débiteurs est égale à  $0.07 + 0.02 + 0.01 = 0.10$
5. Le diagramme en bâtons (ou à bandes rectangulaires)



La courbe cumulative



**Exercice 3 :** On mesure les tailles de cent étudiants d'une faculté universitaire. Les résultats obtenus, rangés par ordre croissant et exprimés en centimètres, sont les suivants :

162 165 167 167 168 168 169 170 171 171 171 171 171 171 172 172  
 172 172 172 172 172 173 173 173 173 173 173 173 174 174 174 174  
 174 174 174 174 174 174 174 174 174 174 174 175 175 175 175 175  
 175 175 175 175 175 175 175 175 176 176 176 176 176 176 176 176  
 176 176 176 176 177 177 177 177 177 177 177 177 178 178 178  
 178 178 178 178 178 178 179 179 179 179 179 179 179 180 181 181  
 181 182 183 185

1. Déterminer la population statistique, l'unité stat, le caractère et l'effectif total.
2. Grouper ces observations en un nombre adéquat de classes de même amplitude.
3. Tracer l'histogramme des fréquences cumulées et la courbe cumulative.
4. Déterminer graphiquement (et par calcul) le pourcentage des étudiants dont la taille est entre 175 cm et 178 cm.
5. Déterminer graphiquement la taille  $T_m$  telle que 50% de la population ait des tailles supérieures strictement à  $T_m$ .

### **Solution de l'exercice3**

1. La population statistique : les étudiants de la faculté universitaire.

L'unité statistique : un étudiant inscrit dans cette faculté.

Le caractère : taille d'un étudiant.

L'effectif total : 100.

2. On va utiliser la formule de Sturges pour déterminer le nombre de classes :

$$1 + 3.322 * \log_{10}(100) = 7.644$$

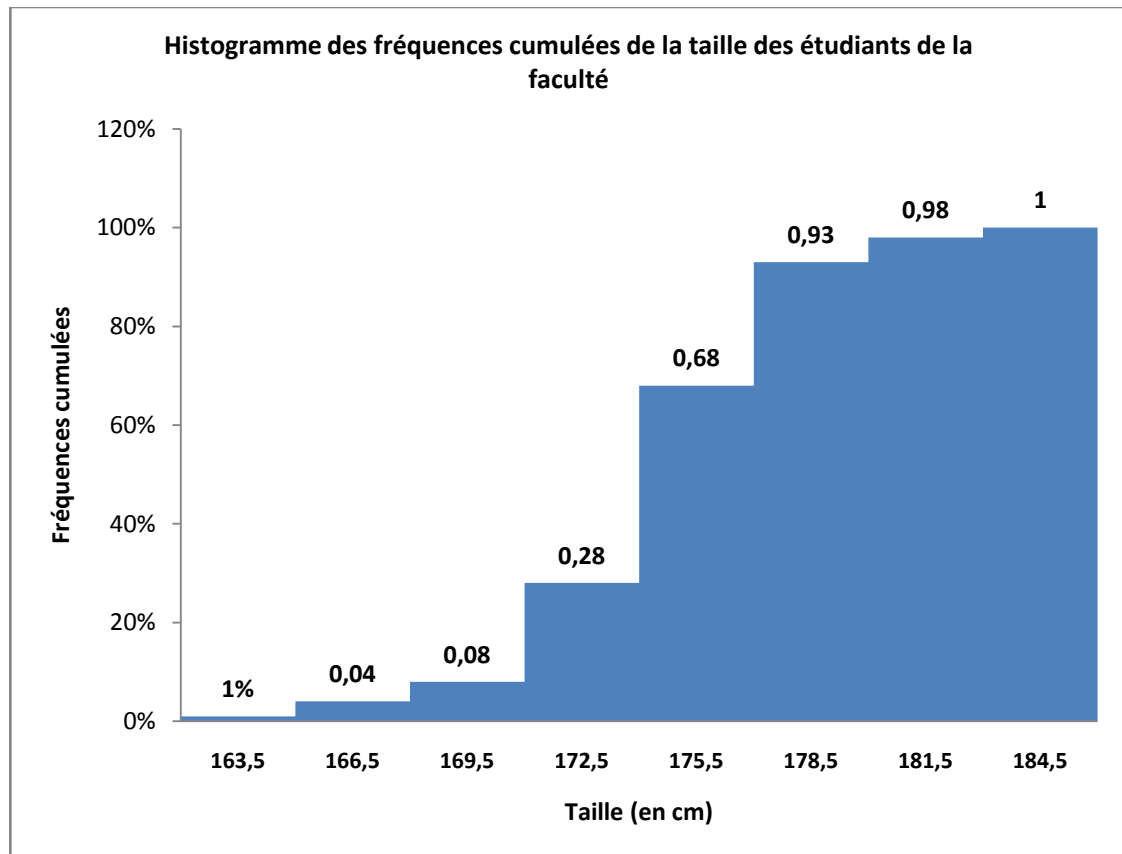
L'entier le plus proche est 8. On va distribuer les données sur 8 classes, on a :

$k = 8$ ,  $n = 100$ , l'étendu  $e = 185 - 162 = 23$ , l'amplitude  $e/k = 23/8 = 2.875$ , on corrige cette amplitude et on prendra l'amplitude  $c = 3$  (les données ne présentent pas de décimales après la virgule). Les classes seront ainsi,  $[162, 162+3[ = [162, 165[$  ;  $[165, 168[$  ;  $[168, 171[$  ; ...

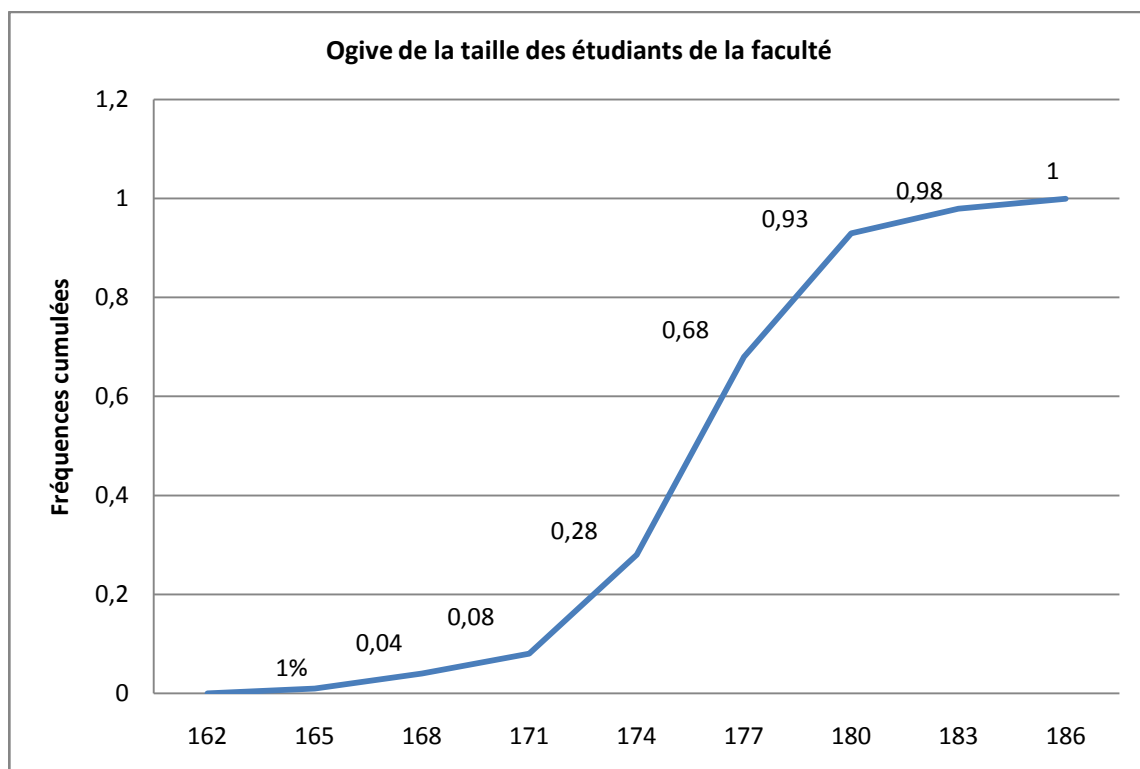
*Distribution des étudiants de la faculté selon leur taille*

| Taille d'un étudiant<br>en cm | Effectif<br>$n_i$ | Fréquence<br>$f_i$ | Fréquence cumulée<br>$F_i$ |
|-------------------------------|-------------------|--------------------|----------------------------|
| [162 , 165[                   | 1                 | 0.01               | 0.01                       |
| [165 , 168[                   | 3                 | 0.03               | 0.04                       |
| [ 168, 171[                   | 4                 | 0.04               | 0.08                       |
| [ 171, 174[                   | 20                | 0.20               | 0.28                       |
| [ 174, 177[                   | 40                | 0.40               | 0.68                       |
| [ 177, 180[                   | 25                | 0.25               | 0.93                       |
| [ 180, 183[                   | 5                 | 0.05               | 0.98                       |
| [ 183, 186[                   | 2                 | 0.02               | 1.00                       |
| Total                         | 100               | 1                  |                            |

3. Histogramme des fréquences cumulées:

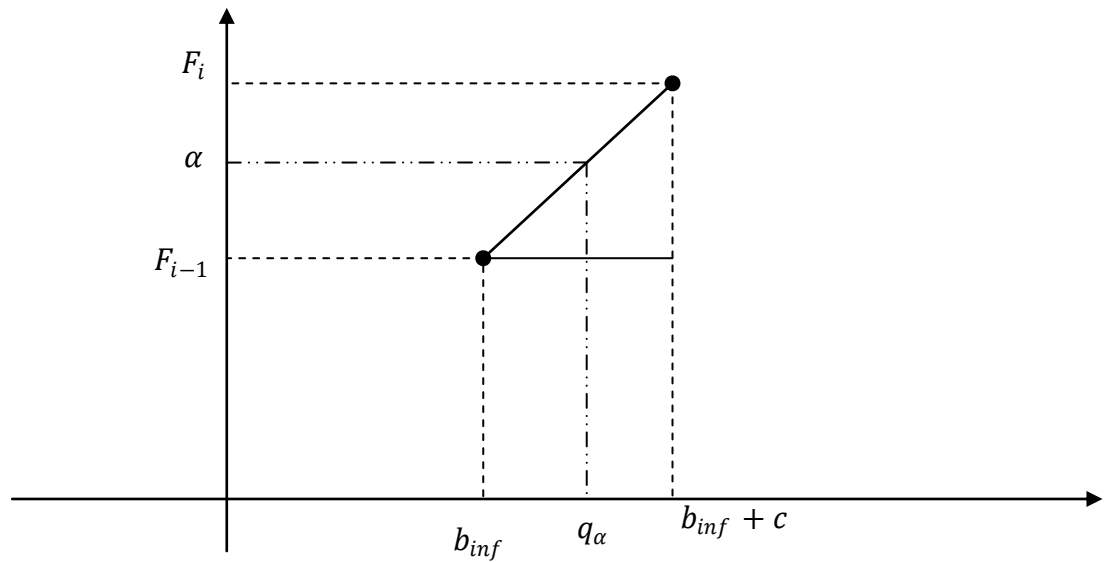


Courbe cumulative des fréquences (ou polygone des fréquences cumulées)



4. Par calcul direct sur les données on doit chercher l'effectif de la classe  $[175, 178[$  qui est égal à 34, ce qui donne un pourcentage égal 34%.

Par utilisation du graphe on doit se baser sur la courbe cumulative (ogive) en utilisant le principe de l'interpolation linéaire :



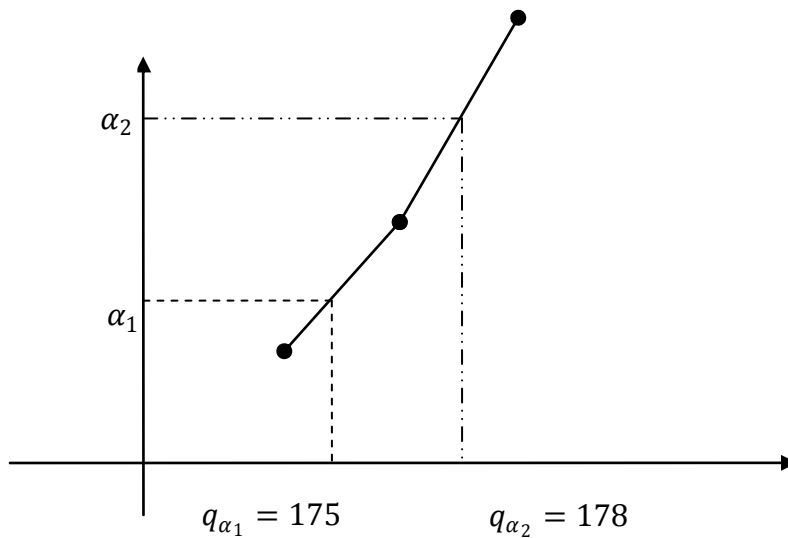
Ce qui donne par interpolation linéaire que

$$\frac{q_{\alpha} - b_{inf}}{\alpha - F_{i-1}} = \frac{c}{F_i - F_{i-1}} = \frac{c}{f_i}$$

On aura la formule du cours (calcul des quantiles)

$$q_{\alpha} = b_{inf} + \left( \frac{\alpha - F_{i-1}}{f_i} \right) * c$$

Si on reprend l'ogive précédent pour les deux intervalles [ 174, 177[ et [ 177, 180[



Si on écrit  $\alpha$  en fonction de  $q_{\alpha}$  on aura :

$$\alpha = \left( \frac{q_{\alpha} - b_{inf}}{c} \right) * f_i + F_{i-1}$$

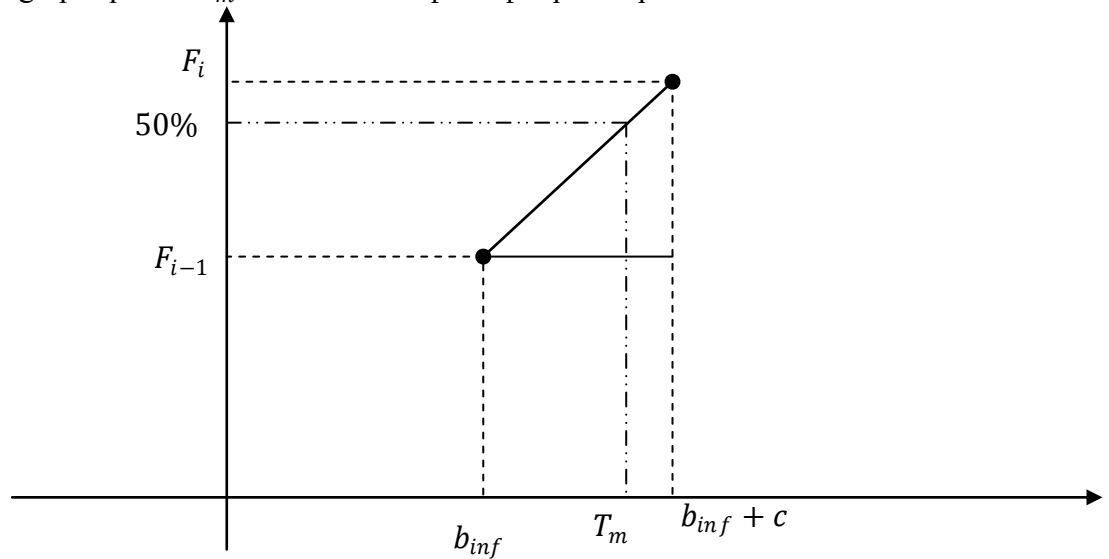
|             | $n_i$ | $f_i$ | $F_i$ |
|-------------|-------|-------|-------|
| [ 174, 177[ | 40    | 0.40  | 0.68  |
| [ 177, 180[ | 25    | 0.25  | 0.93  |

On calcule  $\alpha_1$  et  $\alpha_2$ ,

$$\alpha_1 = \left( \frac{175 - 174}{3} \right) * 0.40 + 0.68 = 0.4133 \text{ et } \alpha_2 = \left( \frac{178 - 177}{3} \right) * 0.25 + 0.93 = 0.7633$$

Le pourcentage demandé est égal  $0.7633 - 0.4133 = 35\%$

Pour l'observation graphique de  $T_m$  c'est le même principe que la question 4 :



On a l'égalité suivante

$$\frac{T_m - b_{inf}}{50\% - F_{i-1}} = \frac{c}{F_i - F_{i-1}} = \frac{c}{f_i}$$

Ce qui donne la formule du cours pour la médiane

$$T_m = b_{inf} + \left( \frac{0.5 - F_{i-1}}{f_i} \right) * c$$

Le calcul donne le résultat suivant

$$T_m = 174 + \left( \frac{0.5 - 0.28}{0.40} \right) * 3 = 175.65$$

Conclusion : 50% des étudiants ont une taille inférieure à 175.65cm et 50% des étudiants ont une taille supérieure à 175.65cm.

**Exercice 4 :** Le tableau ci-dessous donne la répartition de la population active occupée (ayant effectivement un emploi) selon la CSP (catégorie socioprofessionnelle), en France, en mars 1988 (tableaux de l'Economie Française, INSEE, 1989, p. 59).

| CSP                                             | Effectifs en milliers | Fréquences (%) |
|-------------------------------------------------|-----------------------|----------------|
| agriculteurs exploitants                        | 1312                  | 06.1           |
| artisans, commerçants, chefs d'entreprises      | 1739                  | 08.1           |
| cadres, professions intellectuelles supérieures | 2267                  | 10.6           |
| professions intermédiaires                      | 4327                  | 20.1           |
| employés                                        | 5815                  | 27.0           |
| ouvriers                                        | 6049                  | 28.1           |

1. Déterminer la population statistique, l'unité statistique, le caractère et l'effectif total.
2. Quel est le type de la série statistique ? quelle la modalité la plus fréquente ?
3. Tracer deux graphiques représentant le caractère.

#### Solution de l'exercice 4 :

1. La population statistique : l'ensemble des personnes actives occupant un emploi en France dans le mois de mars 1988.

L'unité statistique : une personne active exerçant un emploi en France le mois de mars 1988.

Le caractère : CSP ou Catégorie Socioprofessionnelle.

L'effectif total : est la somme des effectifs observés c.à.d.  $N = 21\,509\,000$ .

$(1312 + 1739 + 2267 + 4327 + 5815 + 6049 = 21509)$

2. Le type de la série est : *qualitatif nominal*.

La modalité la plus fréquente est "ouvriers"

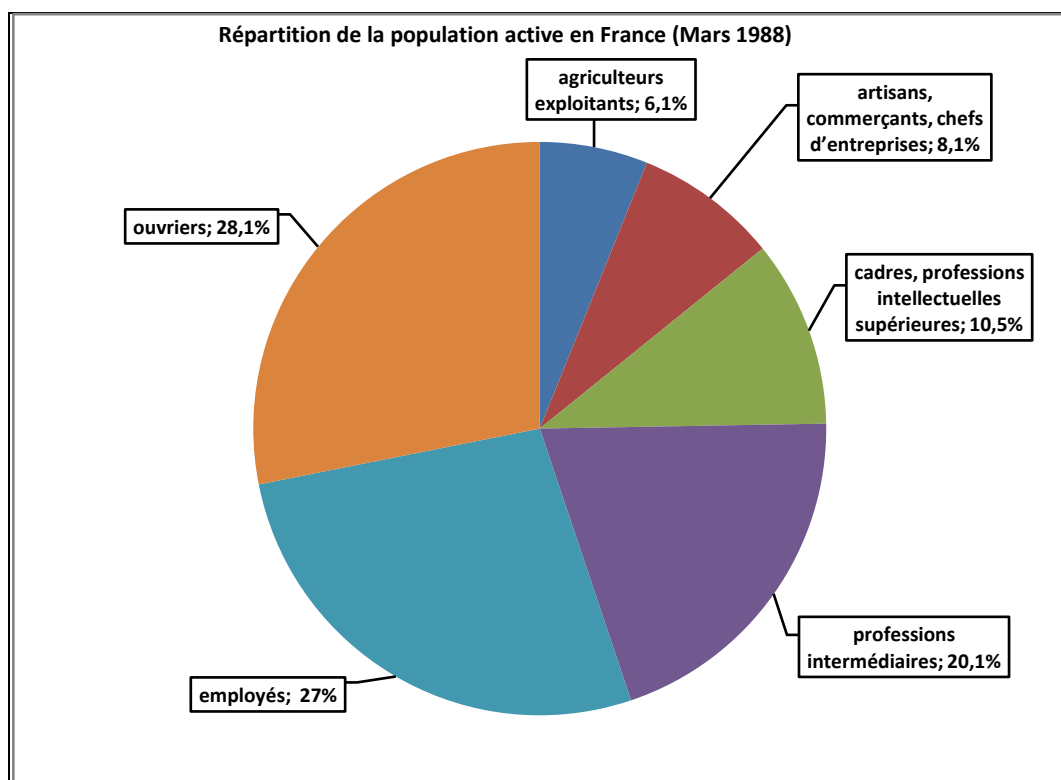
3. Diagrammes

a) Diagramme sectoriel circulaire

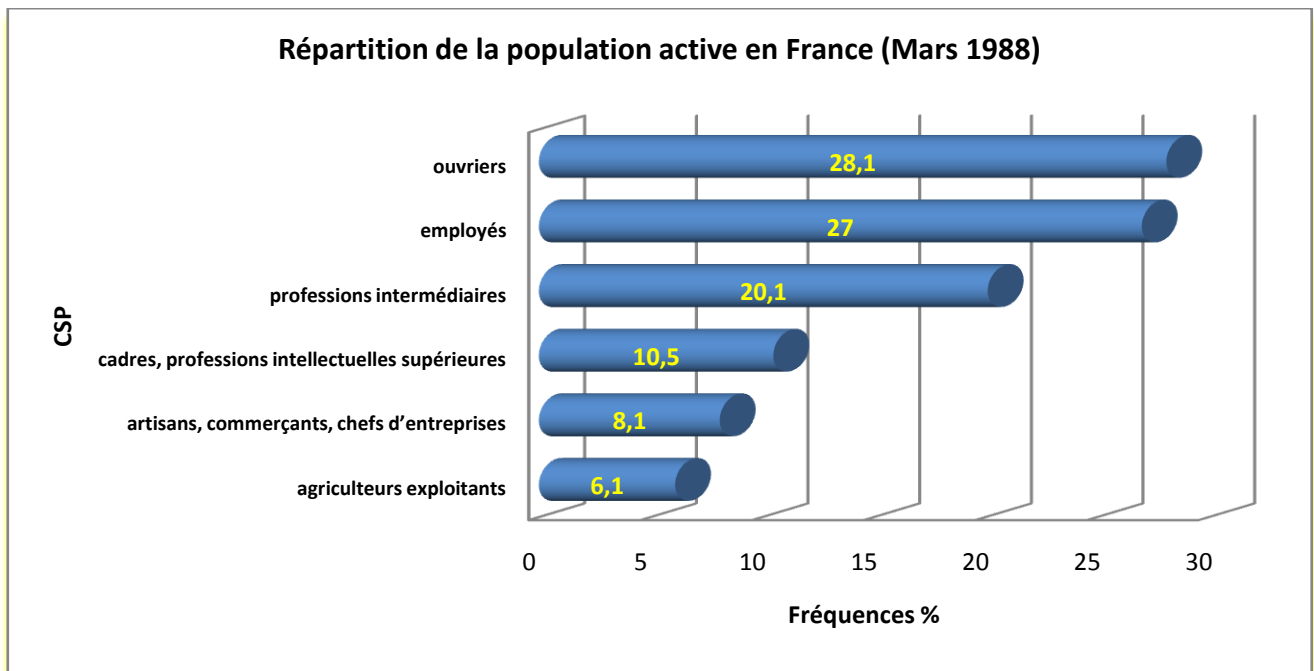
On détermine tout d'abord l'angle de chaque secteur, puis on trace sur papier :

| CSP                                             | Effectifs en milliers | Fréquences (%) | Angle au centre du secteur(°) |
|-------------------------------------------------|-----------------------|----------------|-------------------------------|
| agriculteurs exploitants                        | 1312                  | 6.1            | 21.96                         |
| artisans, commerçants, chefs d'entreprises      | 1739                  | 8.1            | 29.16                         |
| cadres, professions intellectuelles supérieures | 2267                  | 10.5           | 38.16                         |
| professions intermédiaires                      | 4327                  | 20.1           | 72.36                         |
| employés                                        | 5815                  | 27             | 97.2                          |
| ouvriers                                        | 6049                  | 28.1           | 101.16                        |
| Total                                           | 21509                 | 100            | 360                           |

Remarque : l'utilisation du logiciel Excel est commode pour ce type de graphe.



b) Diagramme à bâtons horizontales



**Exercice 5 :** Un groupe d'étudiants d'une promotion de première année de DEUG d'Economie ont passé l'épreuve de rattrapage du cours de statistique descriptive. Ils ont obtenu les notes suivantes :

|                 |     |     |     |     |    |     |     |     |     |    |
|-----------------|-----|-----|-----|-----|----|-----|-----|-----|-----|----|
| Numéro d'examen | 123 | 122 | 334 | 554 | 16 | 121 | 446 | 323 | 100 | 86 |
| Note            | 18  | 3   | 12  | 0   | 12 | 8   | 2   | 5   | 12  | 8  |

Le professeur souhaite faire une étude statistique de la distribution de ces notes.

1. Quel est l'effectif total ? Combien de valeurs a pris le caractère étudié ?
2. Calculez le mode des notes, la note moyenne et la note médiane.

Solution de l'exercice 5 :

1. L'effectif total est égal à 10.

Le nombre de valeurs pris par le caractère étudié est égal à 7.

2. On va remplir tout d'abord le tableau de synthèse : distribution des notes de statistique descriptive au rattrapage.

| Note  | Effectif<br>$n_i$ | Fréquence<br>$f_i$ |
|-------|-------------------|--------------------|
| 0     | 1                 | 0.1                |
| 2     | 1                 | 0.1                |
| 3     | 1                 | 0.1                |
| 5     | 1                 | 0.1                |
| 8     | 2                 | 0.2                |
| 12    | 3                 | 0.3                |
| 18    | 1                 | 0.1                |
| Total | 10                | 1                  |

Le mode des notes  $Mo$  est égale à 12 ; on dira que la note la plus fréquence dans la série est 12.

La note moyenne est égale à

$$\mu = \frac{1 * 0 + 1 * 2 + 1 * 3 + 1 * 5 + 2 * 8 + 3 * 12 + 1 * 18}{10} = \frac{80}{10} = 8$$

La note moyenne résultat du rattrapage de statistique descriptive est 08.

La note médiane :

$n = 10$  est pair, donc la médiane se situe entre la 5<sup>ème</sup> et la 6<sup>ème</sup> valeurs de la série ordonnée dans le sens croissant ; la cinquième valeur et la sixième valeur coïncident avec 8 ; c.à.d. qu'on a :

$$Me = \frac{8 + 8}{2} = 8$$

On dira qu'il y a 50% des étudiants qui ont eut une note supérieure à 08 et que 50% ont eut une note inférieure à 08.

### **Exercice 6 :**

1. Sur une population de 100 unités une variable statistique  $X$  est de moyenne 5. Sachant que sur trois autres unités, la variable  $X$  prend les valeurs 5.02, 4.92 et 5.08, calculer la moyenne de  $X$  sur les 103 unités.
2. Une entreprise a commandé les achats de matières premières au cours des six mois précédents, on donne le tableau des achats avec le coût par kilogramme acheté et le nombre de kilogrammes par achat :

| Achat | Coût par kg (DH) | Nombre de kg |
|-------|------------------|--------------|
| 1     | 34.5             | 540          |
| 2     | 31.8             | 680          |
| 3     | 38.1             | 460          |
| 4     | 26.6             | 1400         |
| 5     | 28.5             | 880          |
| 6     | 22.8             | 1860         |

Un responsable dans l'entreprise veut s'informer sur le coût moyen par kg de matières premières. Vous devez calculer ce coût moyen.

### **Solution de l'exercice 6 :**

1. On utilisera la moyenne pondérée :

$$\bar{x}_p = \frac{\sum_{i=1}^4 p_i x_i}{\sum_{i=1}^4 p_i} = \frac{100 * 5 + 1 * 5.02 + 1 * 4.92 + 1 * 5.08}{100 + 1 + 1 + 1} = \frac{515.02}{103} = 5.0002$$

2. Puisque les quantités commandées varient entre 460kg et 1860kg, on devrait calculer une moyenne du coût d'achat qui prend en compte les différences entre les quantités achetées. On utilisera la moyenne pondérée, les poids seront les quantités commandées :

$$\bar{x}_p = \frac{\sum_{i=1}^6 p_i x_i}{\sum_{i=1}^6 p_i} = \frac{540 * 34.5 + 680 * 31.8 + 460 * 38.1 + 1400 * 26.6 + 880 * 28.5 + 1860 * 22.8}{540 + 680 + 460 + 1400 + 880 + 1860}$$

$$\bar{x}_p = \frac{162508}{5820} = 27.92 \text{ DH}$$

Le coût moyen par kg de matières premières est égal à 27.92 DH