

FICHE D'ANIMATION

Matière :

MATHEMATHIQUES

Thème :

STATISTIQUE 1^{ère} A & 2^{ème} A

Introduction

Ce travail prétend être un support scientifique et pédagogique du contenu du programme de la 1^{ère} et la 2^{ème} Année (toutes les sections) de l'enseignement secondaire en statistique:

- ✚ les nouvelles notions seront définies et illustrées à travers des exemples
- ✚ L'usage de la calculatrice sera expliqué et valorisé
- ✚ des exemples de simulation des expériences aléatoires seront exposés par le biais de l'outil informatique.

Résumé du contenu disciplinaire et des compétences visées (1^{ère} et 2^{ème} Années)

(Source: programme officiel, pages: 13, 24, 36, 46 et 52)

- Série statistique à une variable
- Collecte des données discrètes ou continues
- Organisation et représentation des données dans un tableau, un diagramme(à barres, circulaire, boîte à moustache, histogramme ou une courbe graphique:
polygone des E.C ↗, E.C ↘, F.C ↗ et F.C ↘
- Représenter graphiquement une série chronologique
- Paramètres de position: médianes , quartiles, moyenne et mode.
- Paramètres de dispersion: l'étendue, la variance et l'écart type.
- Placer la médiane et la moyenne d'une série statistique dans une représentation
- Exploiter une représentation graphique:
 - lire la médiane et la moyenne
 - faire des interprétations et des prédictions sur la fréquence d'apparition des phénomènes aléatoires
- Simuler des expériences aléatoires

D) Série statistique à un caractère

1) Activité préliminaire (Vocabulaire)

Un statisticien a réalisé des études portant sur:

- 1) une série de notes données par un professeur
- 2) la répartition des ménages selon le nombre des pièces dont ils disposent
- 3) la répartition des ménages selon le revenu total du ménage
- 4) la répartition de la population tunisienne par tranche d'âge
- 5) la répartition des touristes qui ont visité la Tunisie en 2006 par nationalité

Pour chacune de ces études, préciser:

- a) la population
- b) l'unité statistique
- c) le caractère
- d) la nature du caractère
- e) quelques modalités du caractère

L'essentiel du cours

- ◆ **La population** est l'ensemble des éléments à observer. (touchés ou concernés par l'étude)
- ◆ Chaque élément de la population est appelé **un individu ou une unité statistique**.
- ◆ **Un caractère (ou une variable) statistique** est toute propriété observée ou mesurée sur les individus d'une population.
- ◆ Un caractère statistique peut être:
 - **quantitatif**: lorsque ses valeurs sont des nombres.
 - **qualitatif**: lorsque ses valeurs (ou modalités) ne sont pas mesurables: ne peuvent pas être exprimés par des nombres.
- ◆ Un caractère statistique quantitatif peut être:
 - **discret (ou à valeurs isolées)**: lorsque le nombre de ses valeurs est fini.
 - **continu**: lorsque ses valeurs décrivent des intervalles de IR.
- ◆ **Une série statistique** est l'ensemble de toutes les modalités, valeurs ou classes d'une variable statistique affectées de leurs effectifs.

2) Caractère qualitatif (Différentes représentations graphiques)

Activité 1

Le tableau ci-contre représente la répartition des élèves d'un lycée par sexe.

- a) Recopier puis compléter ce tableau
- b) Représenter cette série par un diagramme à barres.

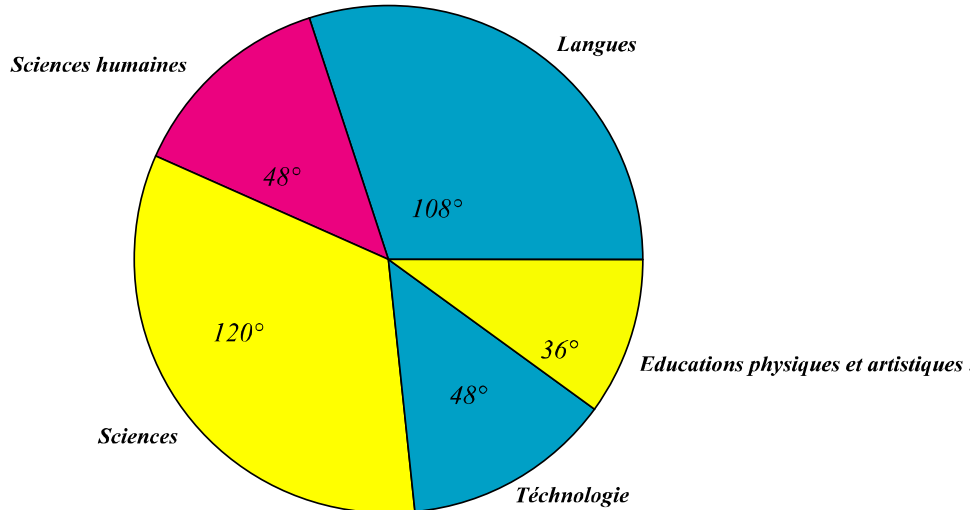
| Sexe | Masculin | Féminin |
|----------------------|----------|---------|
| Effectifs (n_i) | 190 | 313 |
| Fréquences (f_i) | | |

Activité 2

Le diagramme circulaire ci-contre représente la répartition des professeurs d'un lycée par groupe de matières.

Déterminer l'effectif de chaque groupe de matières sachant que l'effectif total est 60.

répartition des professeurs d'un lycée par groupe de matières.



L'essentiel du cours

◆ Dans un diagramme circulaire, à chaque modalité du caractère on associe un angle au centre dont la mesure α en degré est proportionnelle à son effectif n_i (ou sa fréquence f_i) suivant la règle:

$$\alpha = \frac{360}{N} \times n_i = 360 \times f_i$$

Auto évaluation

Le tableau ci- dessous représente la répartition de la population active dans la région du sud-est de la Tunisie par niveau d'enseignement :

| Niveau d'enseignement | Sans niveau | 1er cycle de l'enseignement de base | 2ème cycle de l'enseig. de base et enseig. secondaire | Enseignement supérieur |
|-----------------------|-------------|-------------------------------------|---|------------------------|
| Effectif (f_i) | 25356 | 95763 | 79799 | 25286 |

Source: I.N.S 2004

Représenter cette série par un diagramme circulaire.

2) Caractère quantitatif

a) Caractère quantitatif discret

Activité 1

La liste ci - dessous représente les nombres des patients consultés chaque jour du mois de Janvier 2006 par Dr B. Salah médecin de libre pratique exerçant à une ville A:

0, 10, 5, 6, 9, 4, 4, 0, 15, 12, 3, 8, 5, 10, 0, 13, 7, 9, 5, 4, 2, 0, 6, 10, 12, 8, 3, 8, 0, 5 et 3.

1) Recopier et compléter le tableau ci – dessous:

| | | | | | | | | | | | | | |
|-----------------------------------|---|---|---|---|---|---|---|---|---|----|----|----|----|
| Nombre des patients (x_i) | 0 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 13 | 15 |
| Nombre des jours (n_i) | | | | | | | | | | | | | |
| Effectifs cumulés croissants | | | | | | | | | | | | | |
| Effectifs cumulés décroissants | | | | | | | | | | | | | |

2) Combien de patients Dr B. Salah consulte – t – il en moyenne par jour pendant ce mois?

3) La série ainsi obtenue et qu'on désignera par: série A est – elle plurimodale? Justifier.

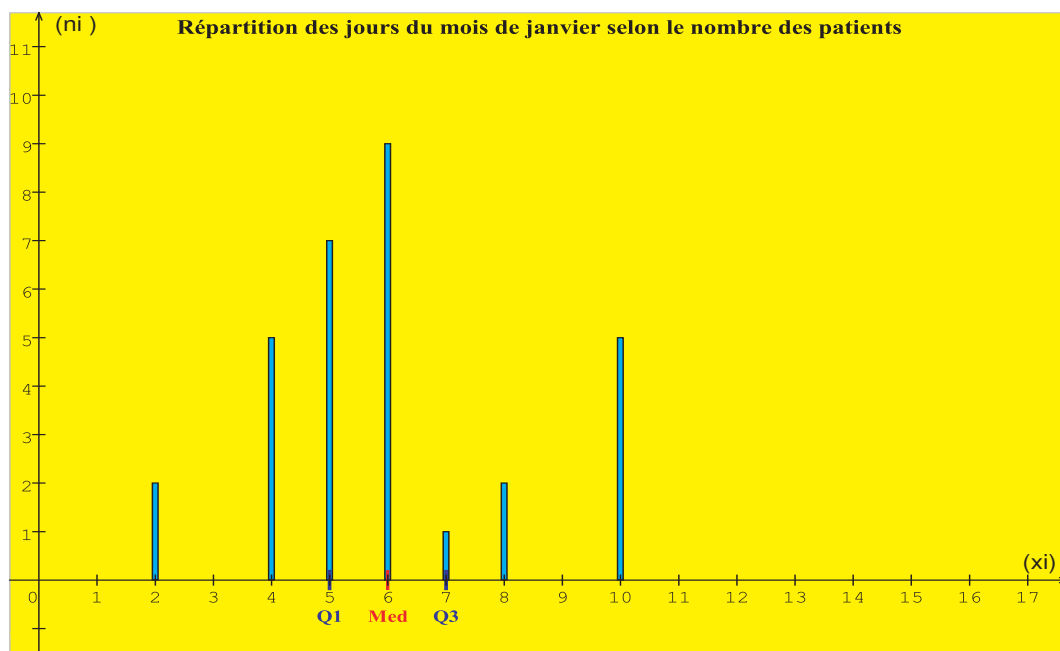
4) Déterminer les quartiles de cette série.

5) Représenter cette série par un diagramme adéquat (en bâtons).

6) Retrouver graphiquement la valeur de Me.

Activité 2

L'étude du même caractère chez Dr B. Abdallah qui exerce à un village B a conduit à l'obtention de la série B représentée par le diagramme en bâtons ci – dessous:



1) Recopier et compléter le tableau ci – dessous:

| | | | | | | | |
|------------------------------------|--|--|--|--|--|--|--|
| Nombre des patients (x_i) | | | | | | | |
| Nombre des jours (n_i) | | | | | | | |
| Fréquences (f_i) | | | | | | | |
| Fréquences cumulées croissantes | | | | | | | |

2) a) Recopier le diagramme en bâtons ci – dessus puis compléter le par la ligne des fréquences cumulées croissantes.

b) Déduire la médiane de cette série (graphiquement)

3) Représenter le diagramme en boîte de cette série.

Activité 3

On admettra que Dr B. Salah et Dr B. Abdallah perçoivent les mêmes honoraires: S dinars pour chaque consultation.

1) Déterminer le montant moyen de la recette quotidienne de chacun de ces deux médecins.

2) Déterminer en fonction de S les recettes maximales et minimales de ces deux médecins. Déduire l'étendue de chacune de deux séries A et B.

3) Pour chacune de deux séries A et B, déterminer le pourcentage de la population ayant une valeur dans l'intervalle $[\bar{x} - \sigma, \bar{x} + \sigma]$. Que peut-on dire de ces deux séries?

L'essentiel du cours

Paramètres de position usuels

◆ **Le mode** d'une série statistique à valeurs isolées notée: **Mo** est la valeur du caractère qui lui correspond le plus grand effectif: c'est la valeur la plus fréquente de la série.

Remarque: une série statistique peut être multimodale.

◆ **La médiane** d'une série statistique à valeurs isolées notée: **Me** est la valeur du caractère qui partage la population en en deux séries de même effectif.

Point méthode:

• Si l'effectif total N est impair ($N = 2p + 1$) alors Me est la valeur du caractère prise par le (p+1)-ème individu:

$$\text{Me} = x_{p+1}$$

• Si l'effectif total N est pair ($N = 2p$) alors Me est la moyenne des valeurs du caractère prises par le p-ème et le (p+1)-ème individus:

$$\text{Me} = \frac{x_p + x_{p+1}}{2}$$

◆ **La moyenne** d'une série statistique à valeurs isolées notée: \bar{x} est le réel défini par:

$$\begin{aligned}\bar{x} &= \frac{1}{N} (n_1x_1 + n_2x_2 + n_3x_3 + \dots + n_px_p) \\ &= f_1x_1 + f_2x_2 + f_3x_3 + \dots + f_px_p\end{aligned}$$

Paramètres de dispersion usuels

◆ **L'étendue** d'une série statistique à valeurs isolées est la différence entre la valeur maximale et la valeur minimale du caractère

◆ On appelle **quartiles** les trois valeurs Q_1 , Q_2 et Q_3 du caractère qui partagent la population en quatre séries de même effectif.

Remarque: $Q_2 = Me$.

Point méthode:

• Q_1 est l'abscisse du point d'intersection de la ligne des fréquences cumulées croissantes et la droite d'équation: $f_i = 0.25$

• Q_2 est l'abscisse du point d'intersection de la ligne des fréquences cumulées croissantes et la droite d'équation: $f_i = 0.5$

• Q_3 est l'abscisse du point d'intersection de la ligne des fréquences cumulées croissantes et la droite d'équation: $f_i = 0.75$

◆ **L'intervalle interquartile** est l'intervalle $[Q_1, Q_3]$.

Son amplitude est appelée **l'écart interquartile**.

◆ **La variance** qu'on note: V est la moyenne des carrés des écarts des valeurs du caractère à sa moyenne \bar{x}

$$\begin{aligned}V &= \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + n_3(x_3 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2}{n_1 + n_2 + n_3 + \dots + n_p} \\ &= \frac{(n_1x_1^2 + n_2x_2^2 + n_3x_3^2 + \dots + n_px_p^2)}{n_1 + n_2 + n_3 + \dots + n_p} - \bar{x}^2 \\ &= f_1(x_1 - \bar{x})^2 + f_2(x_2 - \bar{x})^2 + f_3(x_3 - \bar{x})^2 + \dots + f_px_p^2\end{aligned}$$

◆ **L'écart type** qu'on note: σ est la racine carrée de la variance. Il permet de mesurer la plus ou moins grande dispersion des valeurs de part et d'autre de la moyenne.

$$\sigma = \sqrt{V}$$

Auto évaluation

Activité 1

Ci- dessous la liste des notes obtenues par les 30 élèves d'une classe lors d'un devoir de contrôle:

11, 11, 10, 9, 6, 7, 8, 10, 14, 10, 6, 14, 17, 10, 11, 12, 13, 16, 15, 6, 7, 10, 15, 13, 9, 9, 8, 11, 5, 12.

1) Recopier puis compléter le tableau ci-dessous:

| Note (x_i) | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|--------------------------------|---|---|---|---|---|----|----|----|----|----|----|----|----|
| Effectif (n_i) | 1 | 3 | 2 | 2 | 3 | 5 | 4 | 2 | 2 | 2 | 2 | 1 | 1 |
| Effectifs cumulés croissants | | | | | | | | | | | | | |
| Effectifs cumulés décroissants | | | | | | | | | | | | | |

2) Pour la série statistique obtenue déterminer:

a) Le mode et l'étendue

b) La médiane: Me

c) le premier quartile: Q_1

d) le troisième quartile: Q_3

3) Représenter cette série par diagramme adéquat (diagramme en bâtons)

4) Retrouver graphiquement la valeur de Me

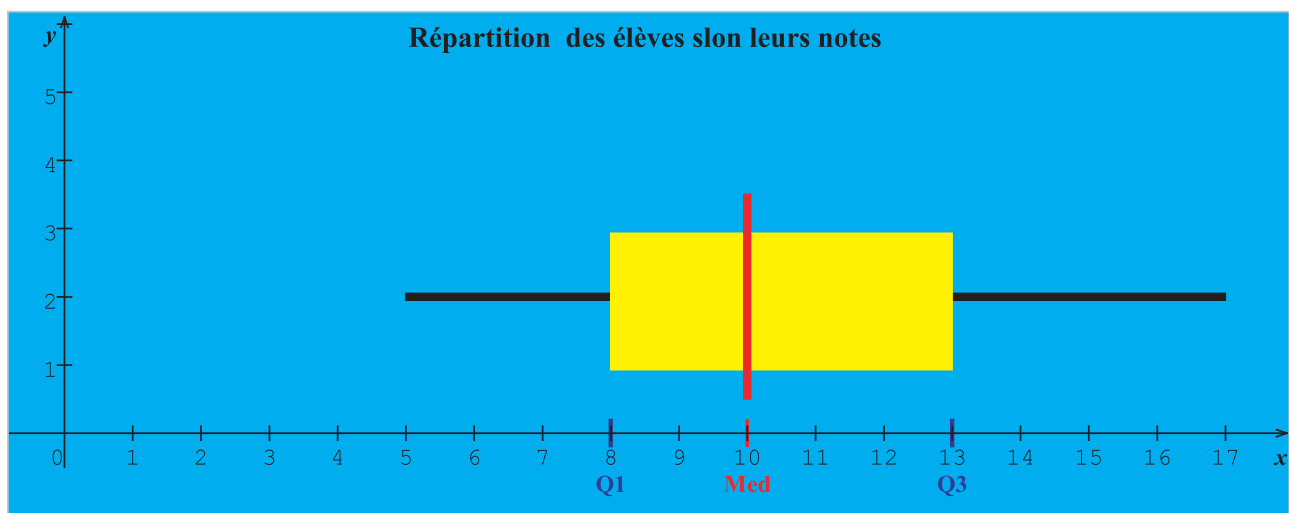
5) Représenter le diagramme en boîte de cette série

2')Ci- dessous le diagramme en boîte de la série statistique obtenue. En déduire:

a) L'étendue

b) La médiane: Me

c) Si un élève a eu 12. Sera-t-il classé parmi le 1/4 de la classe formé des meilleurs élèves?



Activité 2

Le diagramme en bâtons des effectifs ci-dessous représente la répartition des appels téléphoniques reçus quotidiennement par Walid durant 3 mois.

1°) Recopier et compléter le tableau ci-dessous:

| | | | | | | | |
|---------------------------------------|--|--|--|--|--|--|--|
| Modalités (xi) | | | | | | | |
| Effectifs (ni) | | | | | | | |
| Fréquences (fi) | | | | | | | |
| Fréquences cumulées croissantes | | | | | | | |

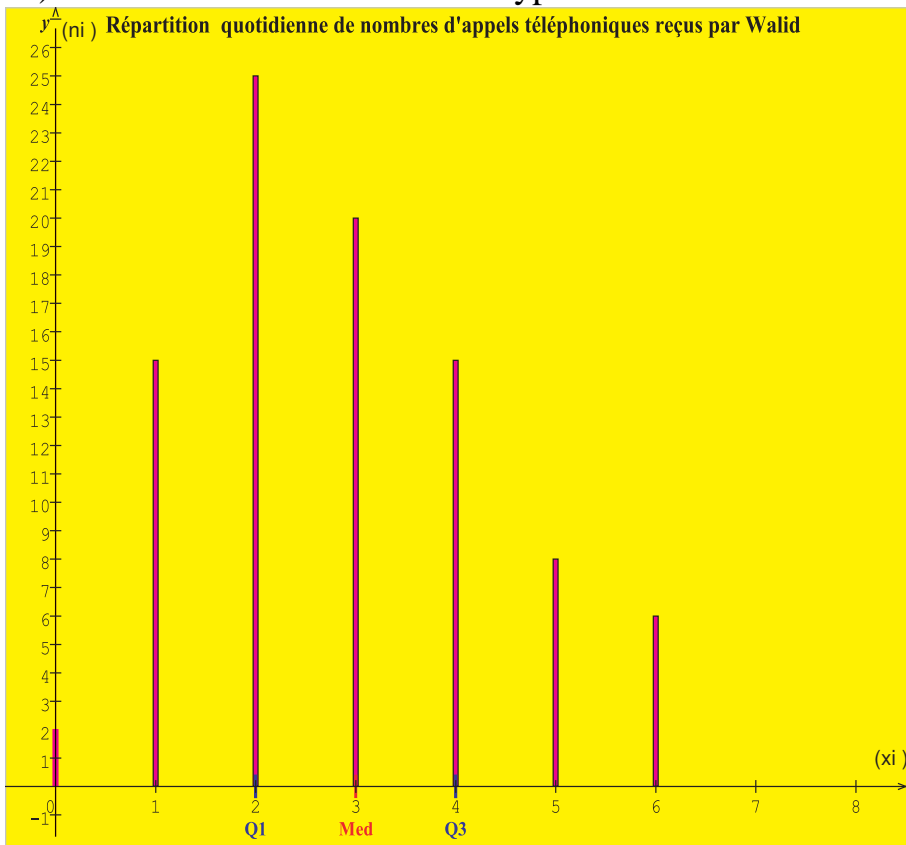
2°) Tracer sur le même graphique la ligne des fréquences cumulées croissantes.

Déduire la médiane de la série.

3°) Déterminer le nombre d'appels reçus quotidiennement en moyenne par Walid.

4°) Donner le diagramme en boîte de cette série.

5°) Calculer la variance et l'écart type de cette série.



b) Caractère quantitatif continu

Activité 1

Les mesures des tailles de soixante personnes sont consignées dans le tableau ci – dessous:

| Tailles en cm | Moins de 160 | [160 , 165[| [165, 170[| [170 , 175[| 175 et plus |
|------------------|--------------|-------------|------------|-------------|-------------|
| Effectif (ni) | 1 | 5 | 21 | 29 | 4 |

- 1) Représenter cette série par un diagramme adéquat.
- 2) Déterminer sa classe modale, sa médiane et sa moyenne \bar{x} .
- 3) Déterminer la fréquence en pourcentage des individus dont la taille appartient à $[\bar{x} - \sigma, \bar{x} + \sigma[$ où σ st l'écart type de cette série.

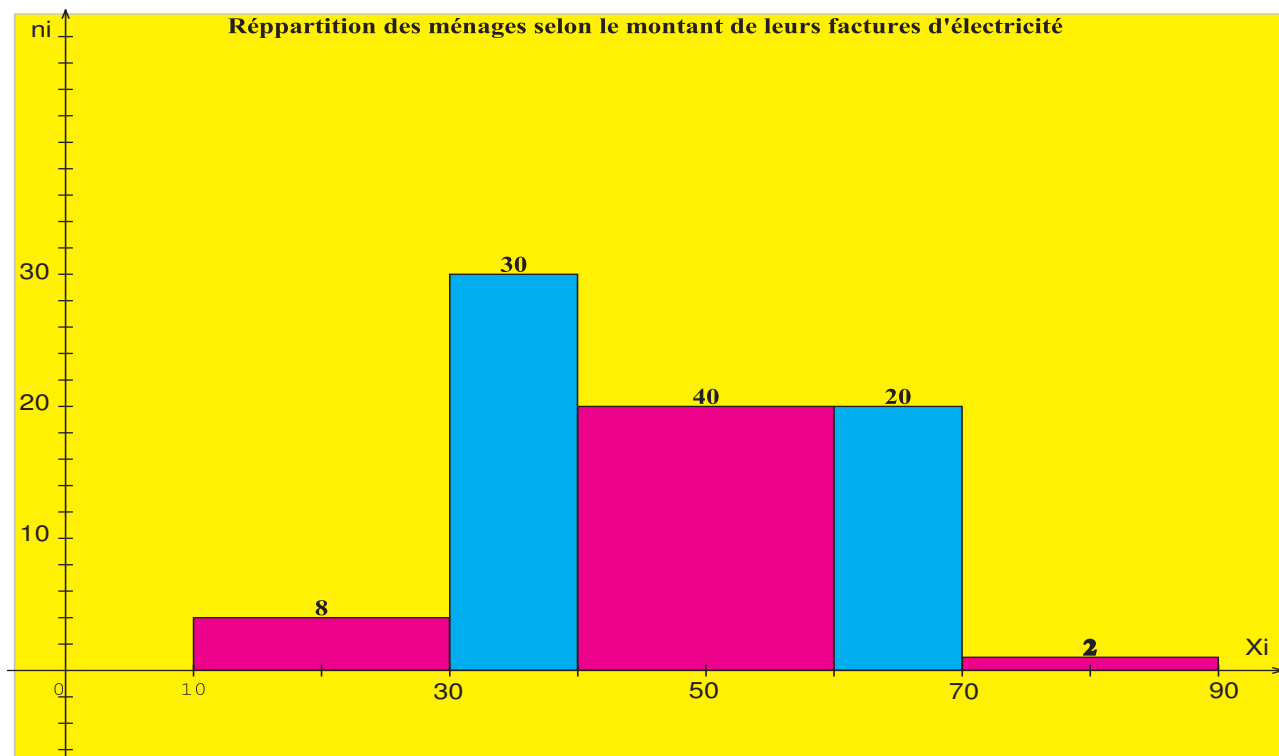
Activité 2

L'histogramme ci- dessous est une représentation graphique de la répartition de cent ménages d'un village selon le montant de leurs factures d'électricité au cours du premier trimestre de 2006.

- 1) Compléter le tableau ci – dessous:

| Montant en dinars | Moins de 30 | [30 , 40 [| [40 , 60 [| [60 , 70 [| 70 et plus |
|---------------------------------|-------------|------------|------------|------------|------------|
| Effectif (n_i) | | | | | |
| Fréquence (n_i) | | | | | |
| Fréquences Cumulées croissantes | | | | | |

- 2) Déterminer la classe modale ainsi que la moyenne \bar{x} .
- 3) Déterminer le montant M en dinars tel que 75 % des ménages payent une facture inférieure à M.
- 4) Déterminer la fréquence des factures dont le montant n'appartient pas à $[\bar{x} - \sigma , \bar{x} + \sigma]$ où σ st l'écart type de cette série



L'essentiel du cours

Paramètres de position usuels

◆ **La classe modale** d'une série statistique continue est:

- la classe du caractère qui lui correspond la plus grande fréquence (ou le plus grand effectif) : si toutes les classes ont la même amplitude.
- la classe du caractère pour laquelle la fréquence (ou l'effectif) par unité d'amplitude est la plus élevée.

Remarque: une série statistique peut être multimodale.

◆ **La médiane** d'une série statistique continue notée: **Me** est la valeur du caractère qui partage la population en deux séries de même effectif.

Point méthode: pour déterminer Me:

- on détermine la classe à laquelle appartient Me: c'est la classe au bout de laquelle l'effectif cumulé croissant atteint ou dépasse la moitié de l'effectif total (ou la fréquence cumulée croissante atteint ou dépasse 50 %)
- on suppose qu'à l'intérieur de cette classe, la répartition est uniforme et par suite on déterminera Me par interpolation linéaire.

◆ **La moyenne** d'une série statistique continue notée: \bar{x} est le réel défini par:

$$\begin{aligned}\bar{x} &= \frac{1}{N} (n_1 c_1 + n_2 c_2 + n_3 c_3 + \dots + n_p c_p) \\ &= f_1 c_1 + f_2 c_2 + f_3 c_3 + \dots + f_p c_p\end{aligned}$$

tel que pour tout i de l'ensemble $\{1, 2, 3, \dots, p\}$ c_i désigne le centre de la i -ème classe.

Paramètres de dispersion usuels

◆ **L'étendue** d'une série statistique continue est la différence entre la borne supérieure de la dernière classe (contenant les plus fortes valeurs du caractère) et la borne inférieure de la première classe (contenant les plus faibles valeurs du caractère)

◆ On appelle **quartiles** les trois valeurs Q_1 , Q_2 et Q_3 du caractère qui partagent la population en quatre séries de même effectif.

Remarque: $Q_2 = Me$.

Point méthode:

• Q_1 est l'abscisse du point d'intersection de la ligne des fréquences cumulées croissantes et la droite d'équation: $f_i = 0.25$

- Q_2 est l'abscisse du point d'intersection de la ligne des fréquences cumulées croissantes et la droite d'équation: $f_i = 0.5$
- Q_3 est l'abscisse du point d'intersection de la ligne des fréquences cumulées croissantes et la droite d'équation: $f_i = 0.75$

♦ **L'intervalle interquartile** est l'intervalle $[Q1, Q3]$.
Son amplitude est appelée **l'écart interquartile**.

♦ **La variance** qu'on note: V est la moyenne des carrés des écarts des centres des classes du caractère à sa moyenne \bar{x}

$$\begin{aligned}
 V &= \frac{n_1(c_1 - \bar{x})^2 + n_2(c_2 - \bar{x})^2 + n_3(c_3 - \bar{x})^2 + \dots + n_p(c_p - \bar{x})^2}{n_1 + n_2 + n_3 + \dots + n_p} \\
 &= \frac{(n_1c_1^2 + n_2c_2^2 + n_3c_3^2 + \dots + n_pc_p^2)}{n_1 + n_2 + n_3 + \dots + n_p} - \bar{x}^2 \\
 &= f_1(c_1 - \bar{x})^2 + f_2(c_2 - \bar{x})^2 + f_3(c_3 - \bar{x})^2 + \dots + f_p(c_p - \bar{x})^2
 \end{aligned}$$

♦ **L'écart type** qu'on note: σ est la racine carrée de la variance. Il permet de mesurer la plus ou moins grande dispersion des valeurs de part et d'autre de la moyenne.

$$\sigma = \sqrt{V}$$

Auto évaluation

Activité

La série ci – dessous est la répartition des ouvriers d'une entreprise selon leurs salaires mensuels:

| classe de salaires en dinars | 280 et moins | [280,320[| [320,360[| [360,450[| [450,600[| 600et plus |
|------------------------------|--------------|-----------|-----------|-----------|-----------|------------|
| Effectifs (n_j) | 20 | 190 | 240 | 80 | 220 | 50 |

- 1) Représenter cette série par un histogramme
- 2) Déterminer la médiane de cette série
- 3) Déterminer l'intervalle interquartile de cette série puis représenter le diagramme en boîte de cette série.
- 4) Déterminer le pourcentage des ouvriers de cette entreprise dont le salaire appartient à l'intervalle $[\bar{x} - \sigma, \bar{x} + \sigma]$.

Simuler le jet d'une pièce avec Excel :

Dans la case A1 : écrire la formule : =SI(ALEA() $<$ 0,5;"PILE";"FACE").

ALEA() est un nombre aléatoire compris entre 0 et 0,99999999 que l'ordinateur tire à notre place.

Il y a donc 1 chance sur 2 que ce nombre soit plus petit que 0,5.

Dans ce cas, la réponse affichée dans la case est PILE, et dans le cas contraire FACE.

Etendre la simulation à 10 jets consécutifs de la même pièce :

Edition Copier

Cliquer en maintenant appuyé sur la zone de A2 à A10 : elle passe en surbrillance

Edition Coller

Les 10 premières cases de la colonne prennent des valeurs aléatoires équiprobables parmi PILE ou FACE.

Retirer 10 nouvelles valeurs au hasard :

Cliquer en haut dans la barre de formule

Touche Entrée

Les 10 nouvelles valeurs tirées sont affichées à la place des anciennes.

Préparation du tableau des effectifs :

Ecrire PILE en case D1, FACE en case E1, et Effectifs en case C2

Dans la case D2 : écrire la formule : =NB.SI(\$A\$1:\$A\$1000;"=PILE")

Dans la case E2 : écrire la formule : =NB.SI(\$A\$1:\$A\$1000;"=FACE")

La fonction NB.SI parcourt toutes les 1000 cases désignées, y compris les cases vides.

Elle compte le nombre des cases dont le résultat est exactement PILE (resp. FACE) et l'affiche.

Préparation du graphique :

Passer la zone rectangulaire de D1 à E2 en surbrillance

Cliquer en haut l'icône de l'Assistant graphique

Type de graphique : Secteurs

Suivant ... Titre du graphique : Résultats des tirages ... Terminer

Le graphique en forme de camembert apparaît.

Il traduit bien visuellement les proportions relatives de PILE et FACE

Observation des fluctuations des fréquences à taille d'échantillon constante :

Retirer 10 nouvelles valeurs et observer la nouvelle répartition de PILE et FACE : que constate t'on ?

Observation de l'influence de la taille de l'échantillon :

Créer un tirage de N = 20 valeurs de PILE ou FACE en copiant la formule en A10 dans la zone de A11 à A20

Observer alors les fluctuations comme précédemment.

Puis refaire la même chose avec N = 50 ; N = 100 ; jusqu'à N = 200.

Que peut-on dire des variations de fréquences quand la taille N de l'échantillon augmente ?