CHAPITRE 1 : DISTRIBUTIONS STATISTIQUES À UNE SEULE VARIABLE

1- CONCEPTS DE BASE

1.1- La statistique

Le terme statistique désigne la méthode scientifique qui consiste à collecter des données chiffrées sur des ensembles nombreux, puis à synthétiser ces chiffres en vue de déduire des analyses et des commentaires des résultats obtenus.

1.2- Population et unité statistique

La population désigne l'ensemble d'objets, d'individus ou d'animaux qui partagent certaines caractéristiques communes et sur laquelle on cherche à étudier des informations quantifiables. Ainsi, on peut parler de la population des étudiants, la population des entreprises, la population du parc autos, etc.

L'ensemble des éléments qui constituent une population sont appelés des ***individus*** ou des ***unités statistiques***. Par exemple, dans la population des lampes électriques produites par une usine, ***une lampe*** représente *l'individu* ou *l'unité statistique*.

Remarque :

Si l'étude se limité à une partie de la population, alors cette partie est appelée ***un échantillon***.

1.3- Le caractère

L'étude de la population statistique choisie se fait selon un ou plusieurs ***caractères*** appelés également ***variables statistiques***. Par exemple, on étudie la population des salariés d'une entreprise selon l'ancienneté, le salaire, diplôme obtenu, les congés, etc.

Le choix des caractères dépendra de l'objectif de l'étude menée par le statisticien.

1.4- La modalité

Le caractère étudié peut varier d'un individu à un autre, il prend ainsi des différentes modalités. Par exemple l'étude d'une population des voitures selon le caractère « couleur ». Alors, ce caractère peut prendre les modalités : noir, gris, bleu, vert, rouge, etc.

Remarques :

❶ Si les modalités d'un caractère étudié sont des ***chiffres***, alors on parle d'un caractère ***quantitatif***. Par exemple la distance de freinage, la durée de vie d'un appareil, le volume horaire d'un cours…

❷ Si les modalités sont des ***noms***, le caractère est dit ***qualitatif.*** Par exemple la nationalité des touristes, les parcours d'une licence, la marque d'automobile, etc.

2- TABLEAUX STATISTIQUES

Après la collecte des données relatives à une variable statistique, il convient de les présenter sous forme d'un tableau statistique. Ce tableau constitue une représentation chiffrée qui dépend de la nature de la variable étudiée (quantitative ou qualitative)

2.1- Cas d'une variable quantitative

Une variable quantitative est une variable mesurable et chiffrée. Elle peut être discrète ou continue.

2.1.1- Variable discrète

Soit une variable statistique X ayant k modalités, tel que $X=\{x_1, x_2, \ldots, x_i, \ldots, x_k\}$ et d'effectif respectivement $n_1, n_2, \ldots, n_i, \ldots, n_k$, alors son tableau statistique se présente comme suit :

Modalité x_i	**Effectif n_i**
x_1	n_1
x_2	n_2
…	…
x_i	n_i
…	…
x_k	n_k

Remarques :

❶ La somme des effectifs est égale à la taille de la population étudiée, alors on peut écrire :

$$\sum_{i=1}^{k} n_i = N$$

❷ Si on divise l'effectif n_i par N, on obtient alors la fréquence de la modalité x_i, notée $\boldsymbol{f_i}$, telle que :

$$f_i = \frac{n_i}{N}$$

Exemple n°1:

Une enquête effectuée sur un échantillon des abonnés d'un opérateur de télécommunication visant à étudier le nombre d'appels reçus au bout d'une heure donnée. Les résultats se présentent dans le tableau suivant :

Nombre d'appels x_i	**Nombre de ménages n_i**	**Fréquence f_i**
0	25	0,040
1	268	0,424
2	188	0,297
3	97	**0,153**
4	38	0,060
5 et plus	16	**0,025**
Totaux	**632**	**1,000**

D'après ce tableau, on remarque que 15.3% des ménages reçoivent3 appels et 2.5% reçoivent plus que de 4 appels téléphoniques au cours d'une heure donnée.

2.1.2- Variable continue

Une variable statistique est dite continue si ses modalités se présentent sous forme d'intervalle, ou classe, de forme $[x_i, x_{i+1}[$,

Classes	**Effectif n_i**	**Fréquence f_i**
$[x_1, x_2[$	n_1	f_1
$[x_2, x_3[$	n_2	f_2
…	…	…
$[x_i, x_{i+1}[$	n_i	f_i
…	…	…
$[x_{k-1}, x_k[$	n_k	f_k
Totaux	**N**	**1**

Exemple n°2:

La répartition de la population tunisienne par groupe d'âge se présente ainsi en 2014 :

Classes d'âge x_i	Effectifs n_i (en milliers)	Fréquence f_i
Moins de 10	1 821 179	0,166
[10, 20[	1 617 132	**0,147**
[20, 30[	1 860 027	0,169
[30, 40[	1 800 766	0,164
[40, 50[	1 410 804	0,128
[50, 60[	1 185 774	0,108
[60, 70[	687 726	0,063
70 et plus	599 348	**0,055**
Totaux	**10 982 756**	**1,000**

Source : Institut National des Statistiques, 2014

On déduit de ce tableau que 14.7% de la population tunisienne âgée de moins 20 ans et seulement 5.5% des tunisiens qui dépassent 70 ans. D'où notre population est jeune.

2.2- Cas d'une variable qualitative

2.2.1- Définition

Une variable statistique est dite qualitative dont ses modalités ne peuvent pas être mesurées par des chiffres.

Exemple n°3:

La distribution du personnel de la santé publique se présente dans le tableau suivant :

Catégorie	Effectif n_i	Fréquence f_i
Médecins Généralistes	2361	0,052
Médecins Spécialistes	2347	**0,051**
Personnel Paramédical	27 098	0,593
Administratif et Technique	2 139	0,047
Personnel Ouvrier	11 749	0,257
Totaux	**45 694**	**1,000**

Source : Ministère de la Santé Publique, 2016

On déduit de ce tableau que 5.1% du personnel de la santé publique sont des médecins spécialistes.

3- REPRESENTATIONS GRAPHIQUES

3.1- Représentation d'un caractère quantitatif

On rappelle qu'un caractère est dit quantitatif, lorsqu'il fera l'objet d'une mesure chiffrée. Si cette mesure peut prendre uniquement des chiffres entiers, alors le caractère est dit quantitatif discret. Inversement, il est dit quantitatif continu

3.1.1- Représentation d'un caractère discret

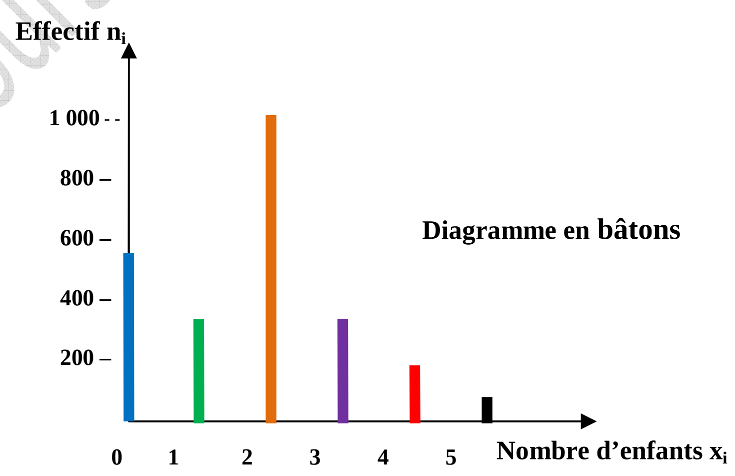
Les variables statistiques discrètes sont représentées par un diagramme appelé diagramme en bâtons, où chaque point de coordonnées $(x_i, n_i\ (f_i))_{i=1, \dots, k}$ est représenté par un segment vertical de longueur proportionnel à l'effectif (fréquence) correspondant (e).

Exemple n°4:

Les résultats d'une enquête sur la répartition de 3000 familles selon le nombre des enfants à charges se présentent dans le tableau suivant :

Nombre d'enfants x_i	Nombre de ménages n_i	Fréquence f_i
0	590	0,197
1	380	0,127
2	1 020	0,340
3	560	0,187
4	370	0,123
5	80	0,027
Totaux	**3000**	**1,000**

Le caractère étudié dans cet exemple est de type quantitatif discret, qui peut se représenter par le diagramme en bâtons suivant :



3.1.2- Représentation d'un caractère continu

Une variable quantitative continue se représente graphiquement par ***un histogramme***. L'histogramme est une représentation des rectangles juxtaposés dans un repère où chaque classe $[x_i, x_{i+1}[$, $_{i=1,\ldots,k}$, est représentée par un rectangle de longueur proportionnelle à l'effectif correspondant n_i ou à la fréquence f_i.

Exemple n°5:

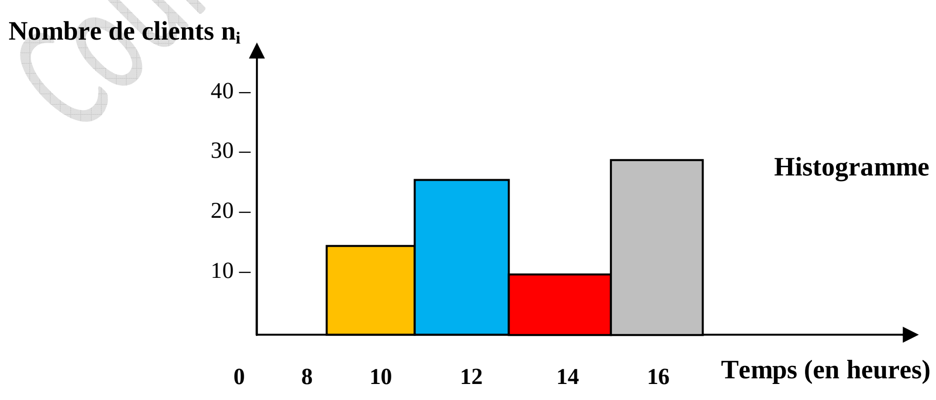
On se propose d'étudier la fréquentation des clients d'une agence d'assure en fonction des temps lors d'une journée ouvrable. Le tableau suivant présente le résultat de cette étude :

Temps en heures	**Nombre des clients n_i**	**Fréquence f_i**
[08 , 10[	15	0,187
[10 , 12[	26	0,325
[12 , 14[	9	0,112
[14 , 16[	30	0,376
Totaux	**80**	**1,000**

1- Déterminer la population statistique, le caractère étudié et préciser sa nature
2- Représenter graphiquement ce caractère

Solution

1- La population étudiée est la clientèle d'une agence d'assurance. Le caractère étudié est la fréquentation des clients, il s'agit d'un caractère quantitatif continu.
2- Ce caractère se représente graphiquement par un histogramme



Remarques

❶ La différence entre la borne supérieure et la borne inférieure d'une classe $[x_i , x_{i+1}[$ est appelée ***l'amplitude*** de cette classe, notée a_i, telle que : $\boxed{a_i = x_{i+1} - x_i}$

❷ Si la variable statistique est rangée dans des classes d'amplitudes inégales, alors on doit corriger l'effectif observé n_i, par l'amplitude de la classe correspondante a_i. On obtient ainsi l'effectif corrigé, noté n_{ci} :

$$\boxed{n_{ci} = \frac{n_i}{a_i} \times a_0}$$

Avec a_0 : la plus petite amplitude observée.

❸ Si on relie par une courbe les milieux des sommets des rectangles qui constituent l'histogramme, on obtient donc le ***polygone des fréquences***.

Exemple n°6:

Soit le tableau statistique suivant :

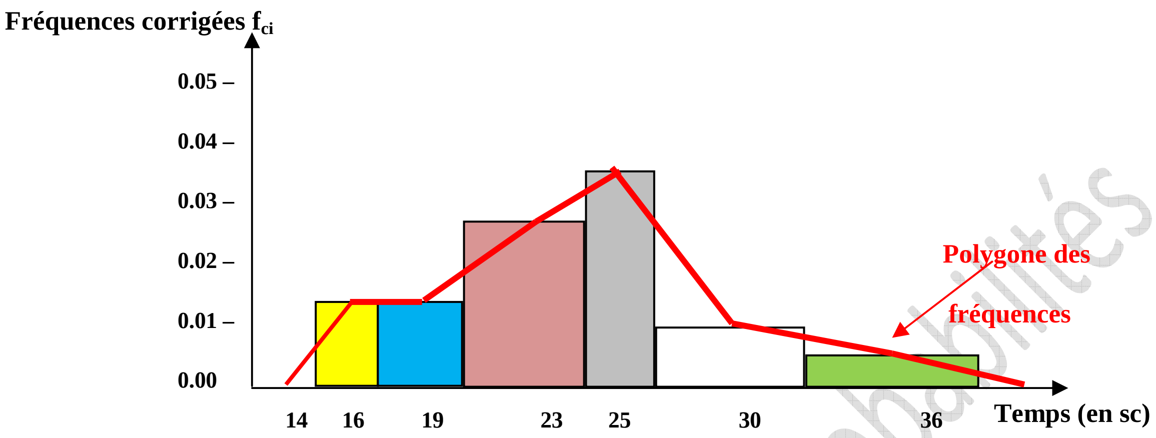
Temps en secondes	**Nombres des coureurs ni**	**Amplitude a_i**	**Effectif corrigé n_{ci}**	**Fréquence corrigée f_{ci}**
[14 , 16[	4	2	4	0,138
[16 , 19[	6	3	4	0,138
[19 , 23[	16	4	8	0,276
[23 , 25[	10	2	10	0,345
[25 , 30[	5	5	2	0,069
[30 , 36[	3	6	1	0,034
Totaux	**44**	**-----**	**29**	**1,000**

Vu que les amplitudes sont inégales, il est nécessaire de déterminer les **effectifs corrigés**. Sachant que la plus faible amplitude est : $a_0 = 2$; par exemple, pour la classe [16 , 19[, d'amplitude 3, son effectif corrigé est :

$$n_c = \frac{6}{3} \times 2 = 4$$

.

L'histogramme et le polygone des fréquences sont représentés sur le graphique suivant :



3.2- Représentation d'un caractère qualitatif

3.2.1- Diagramme en tuyaux d'orgue

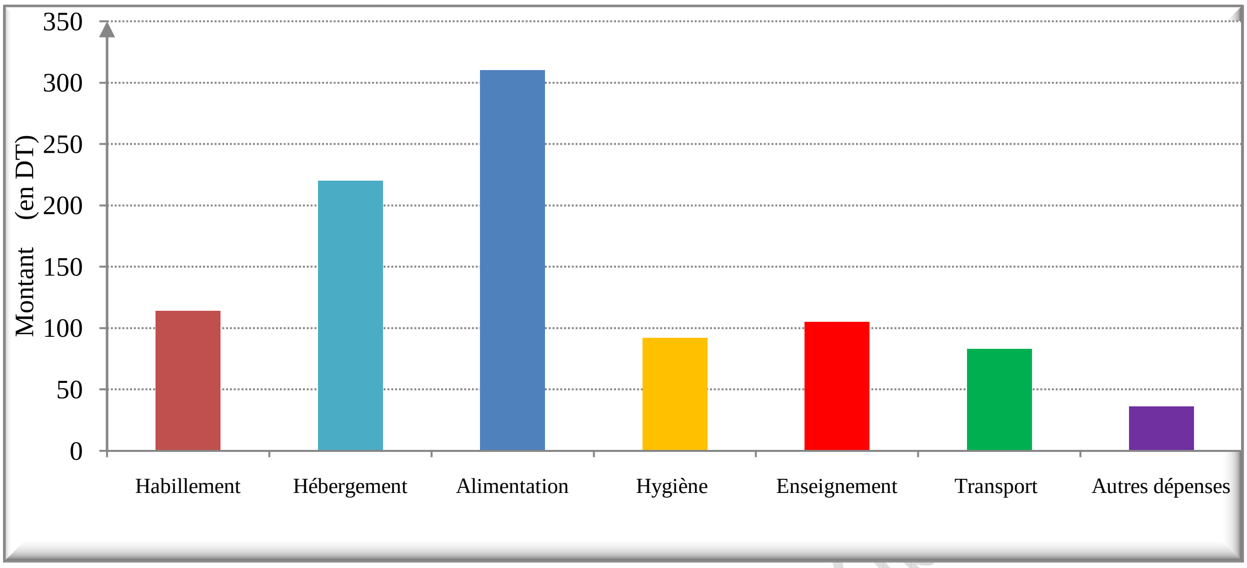
Ce diagramme consiste à représenter chacune des modalités de la variable par un rectangle proportionnelle à sa fréquence (ou son effectif). Les rectangles ainsi obtenus doivent être ***équidistants*** (séparés de la même distance) et de bases ***égales.***

Exemple n°7:

La structure des dépenses mensuelles (en dinars) d'un cadre moyen est la suivante :

Nature des dépenses	Montant en dinars	Fréquence f_i
Habillement	114	0,119
Hébergement	220	0,229
Alimentation	310	0,323
Hygiène & soins	92	0,096
Enseignement, culture & loisir	105	0,109
Transport & communication	83	0,086
Autres dépenses	36	0,038
Totaux	**960**	**1,000**

La représentation graphique de la nature des dépenses mensuelles se fait par le diagramme de **tuyaux d'orgue** :



3.2.2- Diagramme circulaire

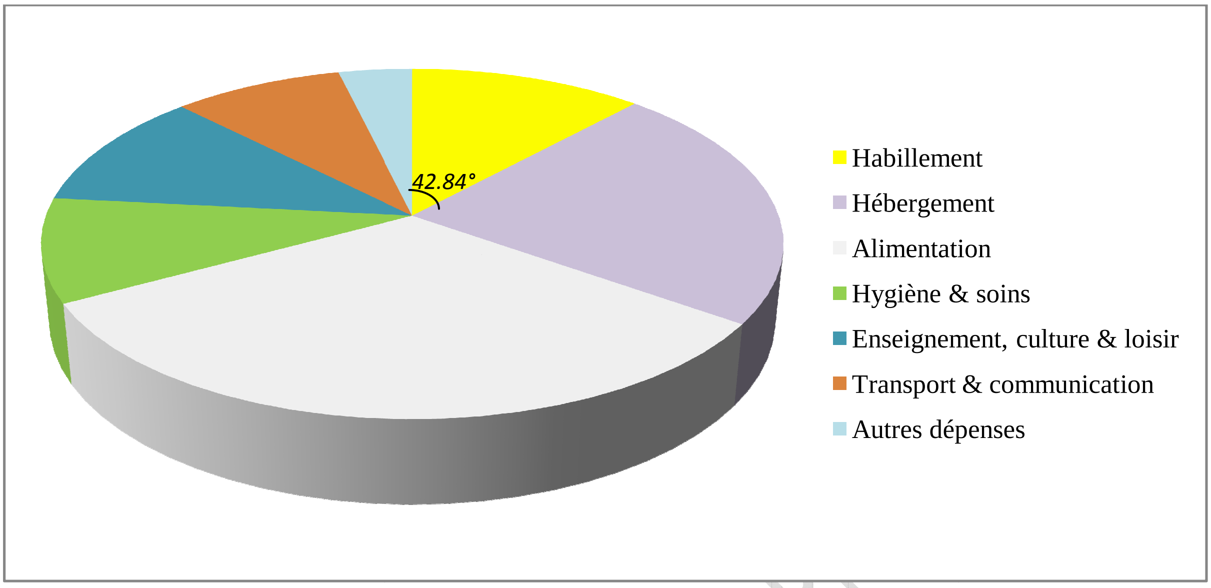
Dans ce diagramme, on représente chaque modalité par un secteur d'angle α est proportionnel à sa fréquence (ou à son effectif). Ainsi, le secteur d'angle α_i de la modalité x_i de manière que :

$$\boxed{\alpha_i = 360° \times f_i}$$

Si on reprend l'exemple précédent, les angles α_i des secteurs qui constituent le diagramme circulaire correspondant se présentent dans le tableau suivant :

Nature des dépenses	Fréquence f_i	Angle α_i (en degré)
Habillement	0,119	42,84
Hébergement	0,229	82,44
Alimentation	0,323	116,28
Hygiène & soins	0,096	34,56
Enseignement, culture & loisir	0,109	39,24
Transport & communication	0,086	30,96
Autres dépenses	0,038	13,68
Totaux	**1,000**	**360,00**

Par conséquent, ce diagramme se représente comme suit :



4- FONCTION CUMULATIVE

4.1- Fonction cumulative d'une variable discrète

On se propose de déterminer les valeurs de la fonction cumulative aux différentes modalités prises par la variable. On obtient ainsi la ***fonction de répartition***.

4.1.1- Détermination de la fonction de répartition

Considérons une variable discrète X présentée dans le tableau suivant :

Modalité x_i	**Fréquence f_i**	**$F(x_i)$ = Prop $(X \leq x_i)$**
x_1	f_1	f_1
x_2	f_2	$f_1 + f_2$
…	…	…
x_i	f_i	$f_1 + f_2 + \ldots + f_i$
…	…	…
x_k	f_k	1

D'après ce tableau, on remarque que : $\mathbf{F(x_1) = f_1}$ **et** $\mathbf{F(x_k) = 1}$.

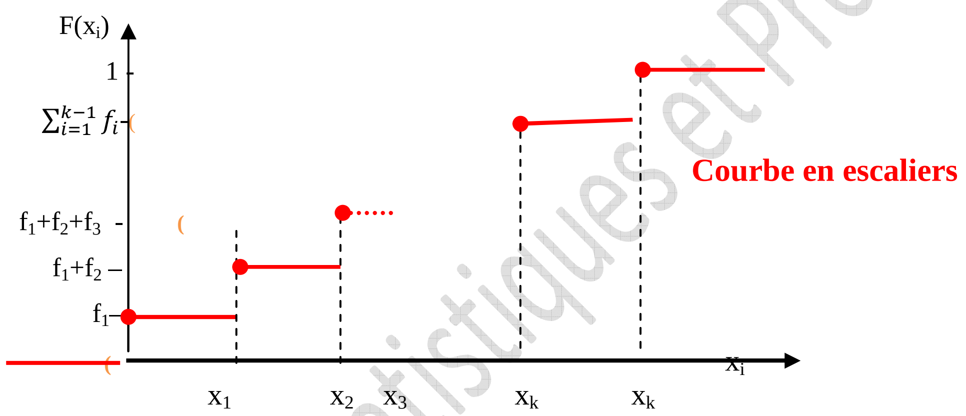
Exemple n°8:

Le service de scolarité de l'ISET propose d'étudier la distribution des absences d'un groupe de 30 étudiants à la fin d'un semestre donné. Les résultats de cette étude se présentent ainsi :

Nombre d'absences x_i	Nombre des étudiants n_i	Fréquence f_i	Fréquence cumulative (x_i)
0	11	0,367	0,367
1	7	0,233	0,600
2	5	0,167	0,767
3	4	0,133	0,900
4	2	0,067	0,967
5	1	0,033	1,000
Totaux	**30**	**1,000**	**-----**

4.1.2- Présentation de la fonction de répartition

La représentation graphique de la fonction cumulative d'une variable discrète est une ***fonction en escaliers***, comme le montre le graphique suivant :



Exemple n°9:

Une entreprise spécialisée dans la production des appareils électriques, propose d'étudier le nombre des articles défectueux qui présentent un défaut de fabrication pour les 20 derniers jours. Les résultats sont présentés dans le tableau suivant :

Nombre défectueux	10	11	12	13	14	15	16
Jours	2	4	2	6	3	2	1

1- Préciser le caractère étudié et déterminer sa nature
2- Calculer et représenter graphiquement les courbes cumulatives croissante et décroissante

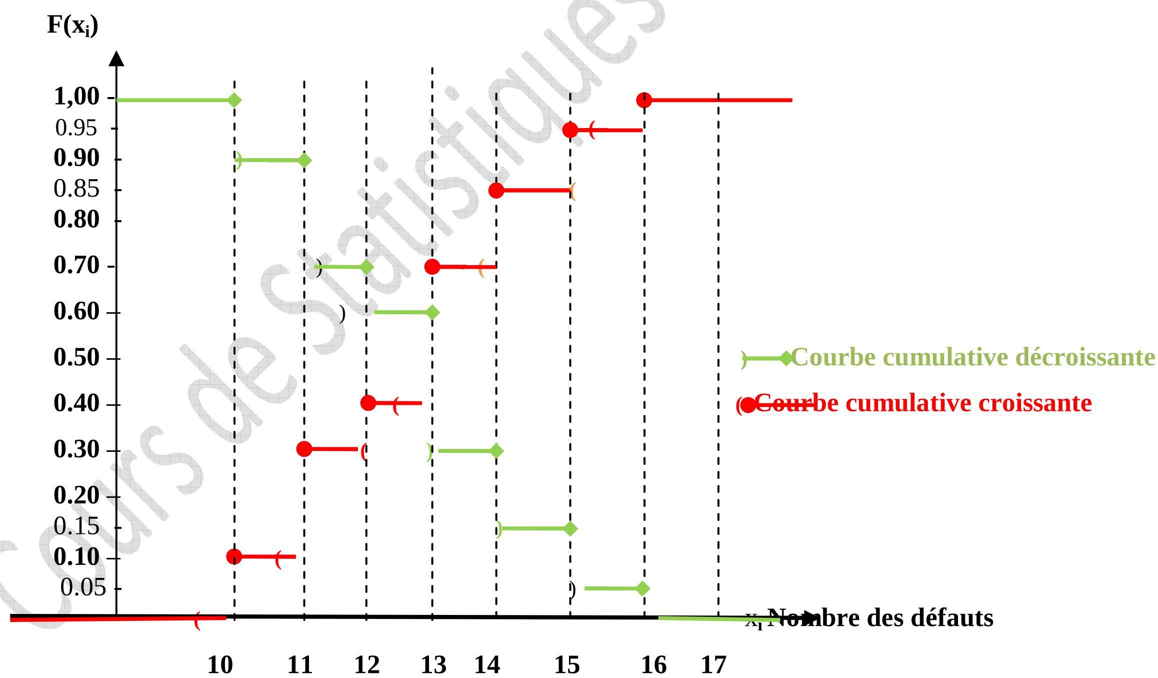
Solution

1- Le caractère étudié X est le nombre des articles défectueux par jour de production. Il s'agit d'un caractère quantitatif discret.

2- Calcul et représentation de la fonction de répartition (croissante et décroissante)

Nombre des défectueux x_i	**Effectif n_i**	**Fréquence f_i**	**F (x_i) Cumulée croissante**	**F (x_i) Cumulée décroissante**
10	2	0,10	0,10	0,90
11	4	0,20	0,30	0,70
12	2	0,10	0,40	0,60
13	6	0,30	0,70	0,30
14	3	0,15	0,85	0,15
15	2	0,10	0,95	0,05
16	1	0,05	1,000	0,000
Totaux	**20**	**1**	**-----**	**-----**

- La courbe des fréquences cumulatives croissantes passe de 0 à 1.
- La courbe des fréquences cumulatives décroissantes décroit de 1 à 0.



4.2- Fonction cumulative d'une variable continue

4.2.1- Détermination de la fonction de répartition

La fonction de répartition d'une variable statistique, X, se détermine de la façon suivante :

Classes	**Effectifs n_i**	**Fréquences f_i**	**$F(x_i)$ = Prop ($X \leq x$)**
$[x_1, x_2[$	n_1	f_1	f_1
$[x_2, x_3[$	n_2	f_2	$f_1 + f_2$
…	…	…	…
$[x_i, x_{i+1}[$	n_i	f_i	$f_1 + f_2 + \ldots + f_i$
$[x_{k-1}, x_k[$	n_k	f_k	1
Totaux	**N**	**1**	**----**

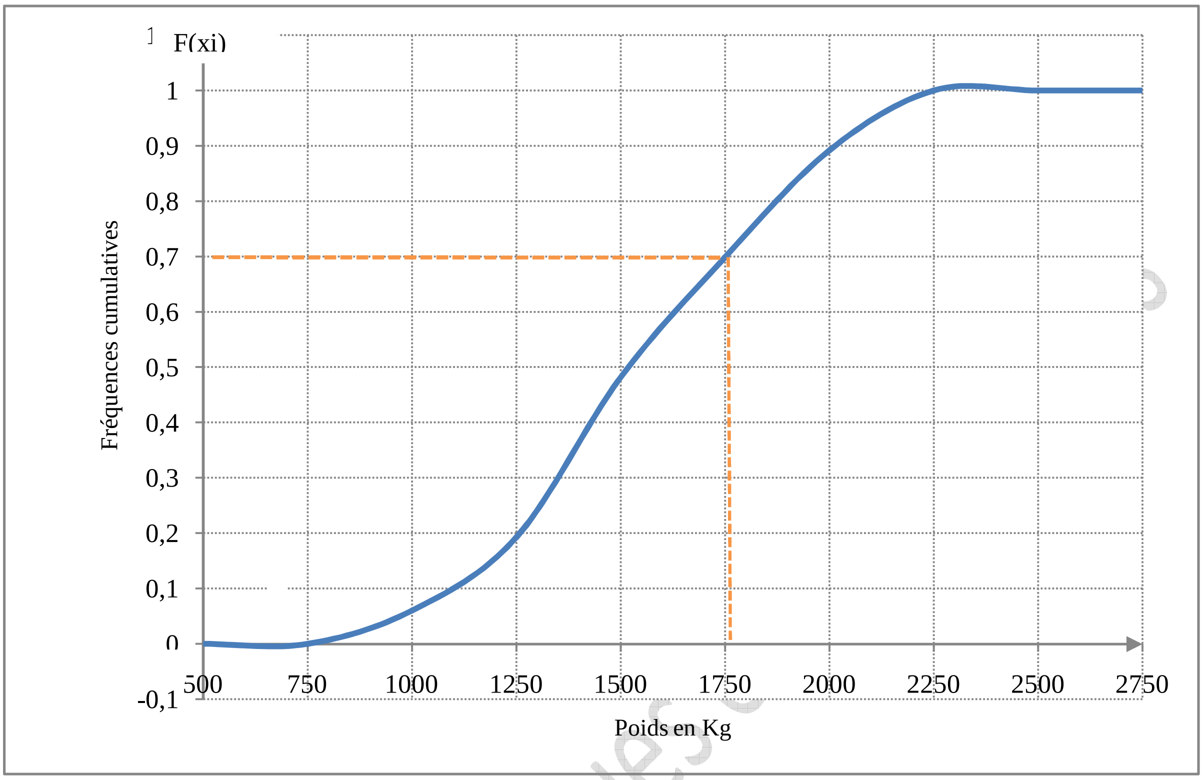
Exemple n°10:

Le poids des poissons pêchés mensuellement pat des bateaux du port de Bizerte se présentent dans le tableau suivant :

Poids en Kg : x_i	**Nombre des bateaux n_i**	**Fréquence f_i**	**Fréquence Cumulée. $F(x_i)$**
[750 , 1000[	5	0,060	0,060
[1000 , 1250[	11	0,133	0,193
[1250 , 1500[	**24**	0,289	**0,482**
[1500 , 1750[	18	0,217	0,699
[1750 , 2000[	16	0,193	0,892
[2000 , 2250[	9	0,108	1,000
Totaux	**83**	**1,000**	**----**

D'après ce tableau, on constate que 24 bateaux pêchent mensuellement un poids compris entre 1250 kg et 1500 kg. La fréquence cumulée croissante jusqu'à cette classe de est égale à : 0,060 + 0,133 + 0,289 = 0,482

4.2.2- Représentation graphique de la fonction de répartition



D'après cette courbe, on déduit que 70 % des bateaux pêchent en moyenne moins de 1 750 Kg de poissons par mois.