Cours Math310

Première partie : statistique descriptive (Suite)

Représentations graphiques

M. BEZOUI

29 novembre 2011

Plan de travail

Plan de travail

1. Quelques definitions

2. I.1.2. Représentations graphiques
   - Cas de variables qualitative
   - Cas de Série statistique Discrète
   - Cas de Série statistique Continue
   - Cas de d'amplitude de classes inégales
   - Fonction de répartition et courbes cumulative

Quelques définitions

Soient une population à "n" individus ayant les modalités $x_1, x_2, ..., x_k$, $k \leq n$

Définition (Effectif (fréquence absolue))

noté n_i : **nombre** d'individus ayant pris la modalité x_i

Définition (fréquence (fréquence relative))

noté f_i : **proportion** d'individus ayant pris la modalité x_i

$f_i = \frac{n_i}{n}$ en pourcentage : $f_i\% = \frac{n_i}{n} * 100$

$\sum_{i=1}^{k} n_i = n_1 + n_2 + ... + n_k = n$, $\sum_{i=1}^{k} f_i = f_1 + f_2 + ... + f_k = 1$

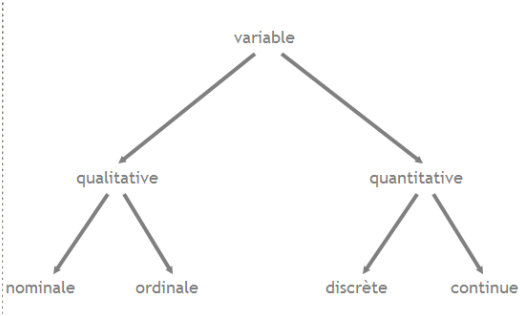


Figure: Schéma récapitulatif des variables

Quelques définitions

Définition (Série statistique)

Donnée de toutes les observations

Définition (distribution statistique)

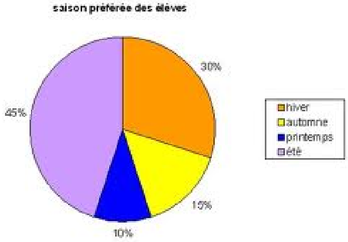
Donnée du couple $\{x_i, n_i\}_{i=\overline{1,k}}$ ou $\{x_i, f_i\}_{i=\overline{1,k}}$

Plan de travail

1. Quelques definitions
2. I.1.2. Représentations graphiques
   - Cas de variables qualitative
   - Cas de Série statistique Discrète
   - Cas de Série statistique Continue
   - Cas de d'amplitude de classes inégales
   - Fonction de répartition et courbes cumulative

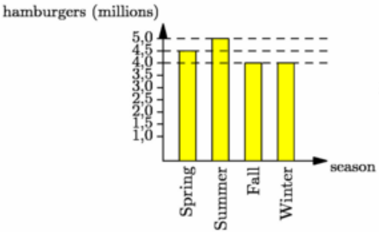
Diagramme circulaire (Cas qualitatif)

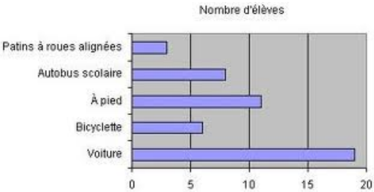
$$\alpha_i = f_i \times 360$$



Par exemple, sur la figure, pour trouver l'angle correspondant à la modalité "Hiver", on doit procéder de cette manière :
$\alpha_{hiver} = 0.3 * 360 = 108$

Tuyaux d'orgue ou Diagramme à bondes (Cas qualitatif)

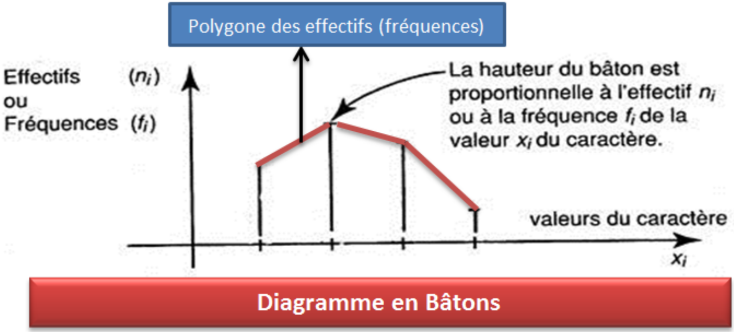




Le graphe à gauche est le plus utilisé, il suffit de trouver les fréquences ensuite de fixer l'echelle du graphe et de faire correspondre la fréquence des variables à la hauteur, (l'axe des ordonées), tandis qu'à l'axe des abscisses on attribut les modalités.

Diagramme en Bâtons (V. Quantitative Disrètes)

Pour chaques modalité x_i on représente un ségment de droite proportionnel aux effectifs (n_i) ou aux fréquences (f_i)



Remarques

Remarque

Le graphe a la même allure, si on représente les fréquences au lieu des effectifs en ordonnées.

Remarque

Les polygones d'effectifs (ou de fréquence) s'obtient on liant la somme des bâtonnets.

La variable statistique continue (Rappel)

Dans le cas d'un caractère quantitatif continu, l'établissement du tableau de fréquences implique d'effectuer au préalable une répartition en classes des données. Cela nécessite de définir le nombre de classes attendu et donc l'amplitude associée à chaque classe ou intervalle de classe.

En règle générale, on choisit des classes de même amplitude. Diverses formules empiriques permettent d'établir le nombre de classes pour un échantillon de taille n.

- La règle de **STURGE** : k=Nombre de classes $= 1 + (3,3log(n))$,
- La règle de **YULE** : k=Nombre de classes $= 2,5.n^{\frac{1}{4}}$.

La variable statistique continue (Rappel)

Dans le cas d'un caractère quantitatif continu, l'établissement du tableau de fréquences implique d'effectuer au préalable une répartition en classes des données. Cela nécessite de définir le nombre de classes attendu et donc l'amplitude associée à chaque classe ou intervalle de classe.

En règle générale, on choisit des classes de même amplitude.

Diverses formules empiriques permettent d'établir le nombre de classes pour un échantillon de taille n.

- La règle de **STURGE** : k=Nombre de classes $= 1 + (3,3log(n))$,
- La règle de **YULE** : k=Nombre de classes $= 2,5.n^{\frac{1}{4}}$.

La variable statistique continue (Rappel)

Dans le cas d'un caractère quantitatif continu, l'établissement du tableau de fréquences implique d'effectuer au préalable une répartition en classes des données. Cela nécessite de définir le nombre de classes attendu et donc l'amplitude associée à chaque classe ou intervalle de classe.
En règle générale, on choisit des classes de même amplitude. Diverses formules empiriques permettent d'établir le nombre de classes pour un échantillon de taille n.

- La règle de **STURGE** : k=Nombre de classes $= 1 + (3,3 log(n))$,
- La règle de **YULE** : k=Nombre de classes $= 2,5.n^{\frac{1}{4}}$.

La variable statistique continue (Rappel)

Dans le cas d'un caractère quantitatif continu, l'établissement du tableau de fréquences implique d'effectuer au préalable une répartition en classes des données. Cela nécessite de définir le nombre de classes attendu et donc l'amplitude associée à chaque classe ou intervalle de classe.
En règle générale, on choisit des classes de même amplitude.
Diverses formules empiriques permettent d'établir le nombre de classes pour un échantillon de taille n.

- La règle de **STURGE** : k=Nombre de classes $= 1 + (3,3log(n))$,
- La règle de **YULE** : k=Nombre de classes $= 2,5.n^{\frac{1}{4}}$.

V.S Continu (suite)

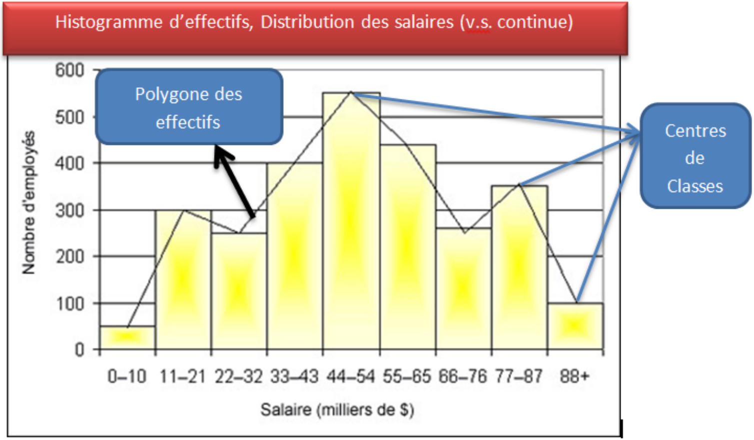
Si on a k classe $[e_0, e_1[, [e_1, e_2[, ..., [e_{k-1}, e_k[$, le centre x_i de la classe $c_i = [e_{i-1}, e_i[, i = \overline{1, k}$ est :

$$x_i = \frac{e_{i-1} + e_i}{2} = e_{i-1} + \frac{a_i}{2}$$

tel que

- e_{i-1} : extrémité inférieure (borne gauche de la classe c_i).
- e_i : extrémité superieure (borne droite de la classe c_i).
- a_i est l'amplitude de la i^{eme} classe, elle est égale à $e_i - e_{i-1}$.

Représentation graphique : Histogramme



La hauteur des barre $h_i = n_i$.

Représentation graphique : Histogramme (suite : Remarques)

Remarque

L'aire de l'histogramme est égale à la somme des aires des rectangles

$$Aire = \sum_{i=1}^{k}(a.f_i) = a.\sum_{i=1}^{k}(f_i) = a$$

Représentation graphique : Histogramme (suite)

L'histogramme est la représentation graphique d'une variable continue. A chaque classe de la variable, correspond la surface d'un rectangle qui a pour base l'amplitude de classe. Comme c'est la surface des rectangles qui représente les phénomènes étudiés, on remarque que :

- si les amplitudes sont égales, alors les hauteurs des rectangles sont proportionnelles aux effectifs ou aux fréquences.
- si les amplitudes sont inégales, il faudra corriger la hauteur des rectangles de manière à ce que leur surface corresponde bien à n_i (les effectifs) ou f_i (les fréquences).

Exemple : Cas de classes d'amplitudes inégales

Exemple de l'étude de la taille en cm dans une classe de terminale :

Classes x_i	Effectifs n_i	Fréquences f_i %	Amplitudes a_i
[150 – 160[	3	8,57 %	10
[160 – 170[	8	22,86 %	10
[170 – 175[	**13**	**37,14 %**	**5**
[175 – 180[	**7**	**20 %**	**5**
[180 – 200[	4	11,43 %	20
total	**35**	**100 %**	

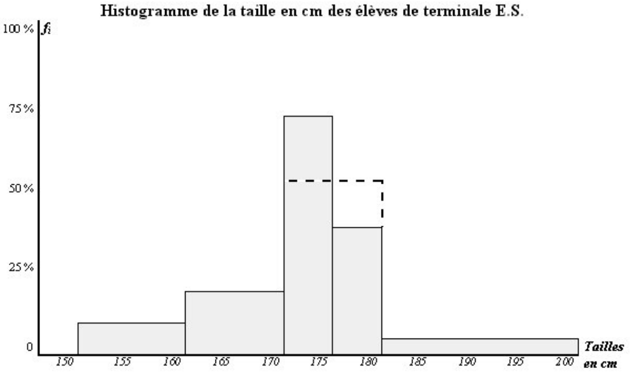
L'amplitude de la classe [170 - 175[est de 5. Si l'amplitude de référence est 10, il faudra multiplier la hauteur par deux puisque la base de cette clase est égale à la moitié de l'amplitude de référence.

Pour simplifier le calcul, on peut ajouter deux colonnes au tableau de départ (Amplitude de référence = 10) :

Classes x_i	Effectifs n_i	Fréquences f_i %	Amplitudes a_i	c = coefficient correcteur = 10 /a_i	Hauteur corrigée $h_i = f_i . c$
[150 – 160[	3	8,57 %	10	1	8,57
[160 – 170[	8	22,86 %	10	1	22,86
[170 – 175[	13	37,14 %	5	2	74,28
[175 – 180[	7	20 %	5	2	40
[180 – 200[	4	11,43 %	20	0,5	5,57
total	35	100 %			

Histogramme : Cas de classes d'amplitudes inégales

Pratiquement, on prend $h_{corige} = \frac{74.28+40}{2} = 57.14$. comme sur cette figure :



L'aire de l'histogramme est conservé après correction des

Fonction de Répartition (Cas Discret)

Soit une variable statistique discrète ayant k modalités $x_1, x_2, ..., x_k$ données par ordre croissant. La fonction de répartition est définie de R dans [0,1], comme suit :

$$F(x) = \begin{cases} 0 & x < x_1; \\ f_1 & x_1 \leq x < x_2; \\ f_1 + f_2 & x_2 \leq x < x_3; \\ \vdots & \vdots \\ f_1 + f_2 + ... + f_i & x_2 \leq x < x_3; \\ \vdots & \vdots \\ f_1 + f_2 + ... + f_k = 1 & x \geq x_k. \end{cases}$$

Fonction de répartition (suite)

On remarque que cette la fonction F est :

- Positive : $0 \leq F(x) \leq 1$,
- Monotone (croissante), Discontinue aux points $x_1, x_2, ..., x_k$.
- $\lim_{x \to -\infty} = 0$, $\lim_{x \to +\infty} = 1$

Fonction de répartition (suite)

On remarque que cette la fonction F est :

- Positive : $0 \leq F(x) \leq 1$,
- Monotone (croissante), Discontinue aux points $x_1, x_2, ..., x_k$.
- $\lim_{x \to -\infty} = 0$, $\lim_{x \to +\infty} = 1$

Fonction de répartition (suite)

On remarque que cette la fonction F est :

- Positive : $0 \leq F(x) \leq 1$,
- Monotone (croissante), Discontinue aux points $x_1, x_2, ..., x_k$.
- $\lim_{x \to -\infty} = 0$, $\lim_{x \to +\infty} = 1$

Fonction de répartition (suite)

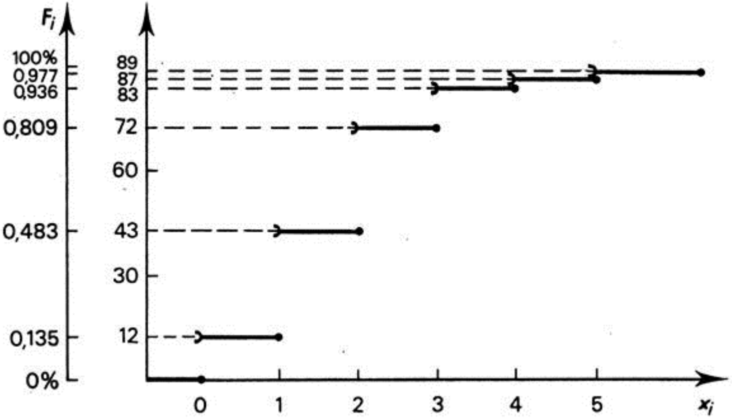
On remarque que cette la fonction F est :

- Positive : $0 \leq F(x) \leq 1$,
- Monotone (croissante), Discontinue aux points $x_1, x_2, ..., x_k$.
- $\lim_{x \to -\infty} = 0$, $\lim_{x \to +\infty} = 1$

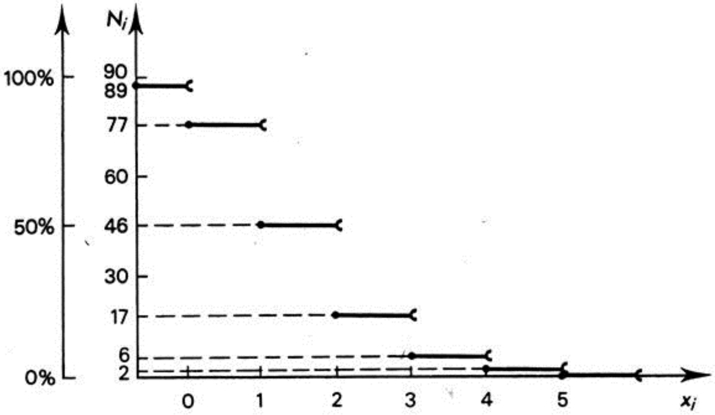
Courbe comulative (Pour le cas discret)

C'est une courbe en secalier comportant des palliers horizontaux. Il y a similitude entre la courbe comulative et la courbe de la fonction de répartition.

Courbe Comulative (Pour le cas discret)



Courbe Comulative (Pour le cas discret)

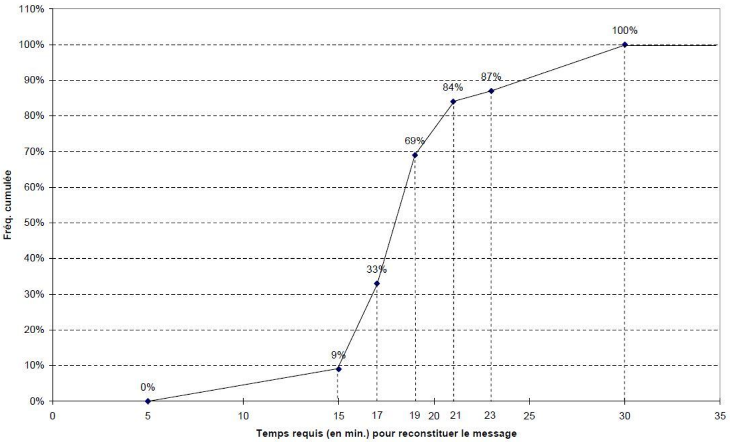


Fonction de répartition (Cas continu)

$$F(x) = \begin{cases} 0 & x < e_0; \\ F(e_1) = f_1 & ; \\ F(e_2) = f_1 + f_2 & ; \\ \vdots & \\ F(e_3) = f_1 + f_2 + ... + f_i & ; \\ \vdots & \vdots \\ F(e_k) = f_1 + f_2 + ... + f_k = 1 & \\ F(x) = 1 & six \geq e_k \end{cases}$$

Courbe comulative (pour le cas Continu)

on a présenté un message mutilé à 60 personnes et mesuré le temps nécessaire à la reconstitution exacte du message. Les résultats (compris entre 5 et 30 minutes) de ces 60 personnes ont donné lieu à la courbe cumulative des fréquences suivantes :


110%
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Fréq. cumulée
0%
9%
33%
69%
84%
87%
100%
0
5
10
15
17
19
20
21
23
25
30
35
Temps requis (en min.) pour reconstituer le message

Le pourcentage de personnes qui ont mis strictement plus de 21 min. est donné par 100% -84%=16%.

En passe au prochain cours ! ! !