Série Statistique continue

Enseignants :Aloui.M/Bacha.I/Debbech.A **AU :2019/2020**

Classe :1Ing-Info **.**

Rappelons qu'une variable est dite continue lorsqu'elle peut prendre une infinité de valeurs. En pratique, il ne s'agit pas d'une infinité puis que les statistiques sont établies sur une variable mesurée et qu'aucune mesure n'a la précision de l'infiniment petit.

On considère donc une variable comme continue dès lors qu'on lui attribue un grand nombre de valeurs possibles.

Pour pouvoir analyser la série statistique, on est obligé de définir de classes de valeurs.

Ce sont des intervalles qui se suivent et qui englobent toutes les valeurs de la série.

Ces intervalles sont semi ouverts.La borne basse est comprise et la borne haute est exculue.

Exemple $[a, b[$. La différence entre les deux bornes est appelée **l'amplitude**.

Exemple

Salaire	[800, 1000[	[1000, 1200[	[1200, 1400[	[1400, 1600[	[1600, 1800[	[1800, 2200[
Effectif	30	45	60	85	20	10

Sur la série ci-dessus, l'amplitude de chaque classe est de 200 sauf la dernière classe où elle est de 400.En effet, il n'est pas obligatoire de toujours prendre la même.Pour calculer une moyenne, il faut retenir le centre de chaque classe :[800, 1000[; [1000, 1200[; [1200, 1400[; [1400, 1600[; [1600, 1800[; [1800, 2200[Le centre d'un intervalle $[a, b[= \frac{a+b}{2}$

On obtient alors le tableau suivant :

Centre	900	1100	1300	1500	1700	2000
Effectif	30	45	60	85	20	10

La moyenne de cette série statistique est :

$$M = \frac{900 \times 30 + 1100 \times 45 + 1300 \times 60 + 1500 \times 85 + 1700 \times 20 + 2000 \times 10}{30 + 45 + 60 + 85 + 20 + 10} = 1344$$

La médiane et les quartiles se déterminent grâce aux fréquences cumulées

Salaire	[800, 1000[	[1000, 1200[	[1200, 1400[	[1400, 1600[	[1600, 1800[	[1800, 2200[
Centre	900	1100	1300	1500	1700	2000
Effectifs	30	45	60	85	20	10
Fréquence	$\frac{30}{250} = 0.12$	$\frac{45}{250} = 0.18$	$\frac{60}{250} = 0.24$	$\frac{85}{250} = 0.34$	$\frac{20}{250} = 0.08$	$\frac{10}{250} = 0.04$
Fréquence cumulée $\nearrow$	$0.12 \nearrow +$	$0.3 \nearrow +$	$0.54 \nearrow +$	$0.88 \nearrow +$	$0.96 \nearrow +$	1

Pour connaitre la médiane, il faut sélectionner la valeur qui correspond à la fréquence cumulée 0.5.

Pour connaitre le premier quartile Q_1 ,il faut sélectionner la valeur qui correspond à la fréquence cumulée 0.25.

Pour connaitre le troisième quartile Q_3, il faut sélectionner la valeur qui correspond à la fréquence cumulée 0.75.

Revenons à notre série

Salaire	[800, 1000[	[1000, 1200[	[1200, 1400[	[1400, 1600[	[1600, 1800[	[1800, 2200[
Centre	900	1100	1300	1500	1700	2000
Effectifs	30	45	60	85	20	10
Fréquence	$\frac{30}{250} = 0.12$	$\frac{45}{250} = 0.18$	$\frac{60}{250} = 0.24$	$\frac{85}{250} = 0.34$	$\frac{20}{250} = 0.08$	$\frac{10}{250} = 0.04$
Fréquence cumulée $\nearrow$	$0.12 \nearrow +$	$0.3 \nearrow +$	$0.54 \nearrow +$	$0.88 \nearrow +$	$0.96 \nearrow +$	1

La médiane correspond à la fréquence cumulée $f_c = 0.5$ on a alors la classe médiane $[1200, 1400[$ c'est à dire $M_e \in [1200, 1400[$.

L'estimation de M_e peut être déterminer par le calcul. ou bien à partir du polygône des fréquences cumulées.

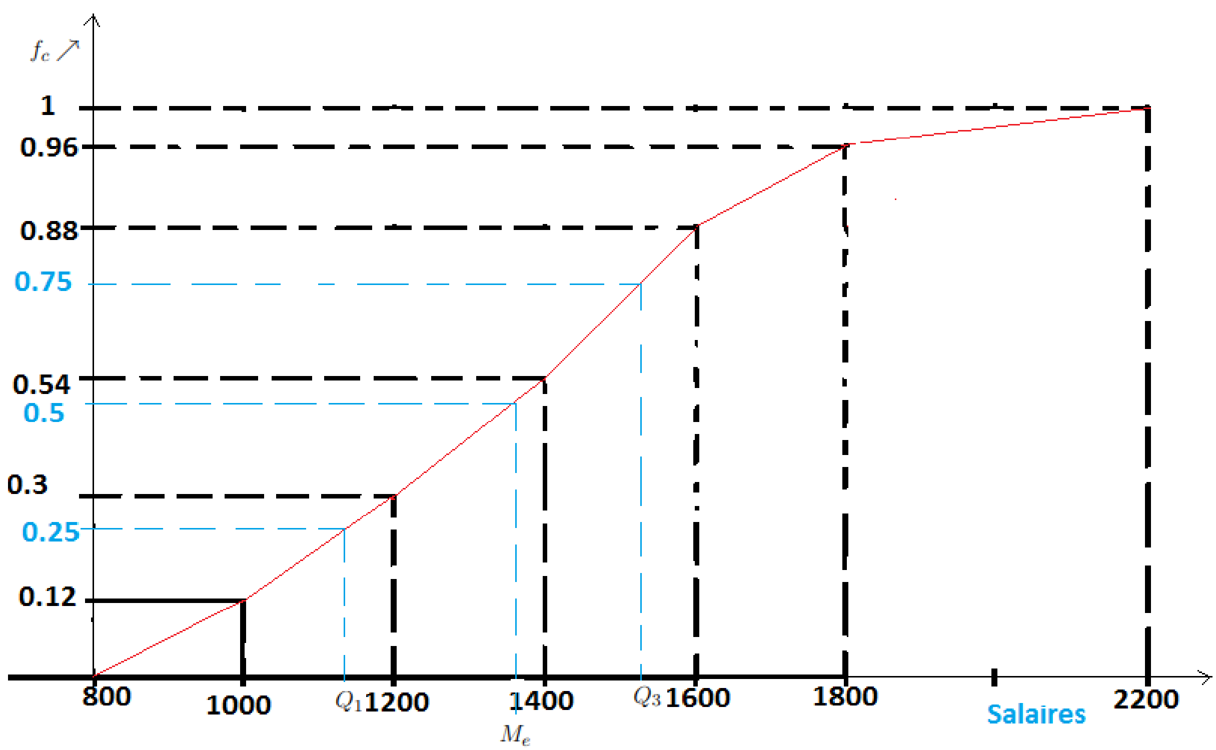
Estimation de la valeur de la médiane à partir du polygône des fréquences cumulées

On utilise un repère dont l'axe vertical est celui des fréquences cumulées et l'axe horizontal celui des valeurs du caractères.

Il faut alors indiquer par des points les fréquences cumulées qui correspondent aux extrêmités de classe puis relier ces points par des segments de droites.

Le tableau des valeurs qui permet de tracer la courbe est le suivant :les extremités des intervalles

x	800	1000	1200	1400	1600	1800	2200
y	0	0.12 :la première valeur de $f_c \nearrow$	0.3	0.54	0.88	0.96	1



$$M_e \in [1200, 1400[$$

$$Q_1 \in [1000, 1200[$$

$$Q_3 \in [1400, 1600[$$

Pour trouver une valeur approchée de M_e par exemple il faut utiliser l'échelle choisie et faire la règle de 3.

Par le calcul : $M_e \in [x_A, x_B[$

donc M_e : $x_A + (x_B - x_A)\dfrac{y_{M_e}}{y_B} = 1200 + 200 + \dfrac{0.5}{0.54} = 1385$

Graphes

La distribution d'une série continue se présente graphiquement par un histogramme.

Dans un repère orthogonale on porte en abscisse les valeurs des bornes des intervalles (selon l'unité choisie) puis pour chaque intervalle on trace un rectangle dont l'aire est proportionnelle à l'effectif (ou bien à la fréquence)suivant l'unité choisi

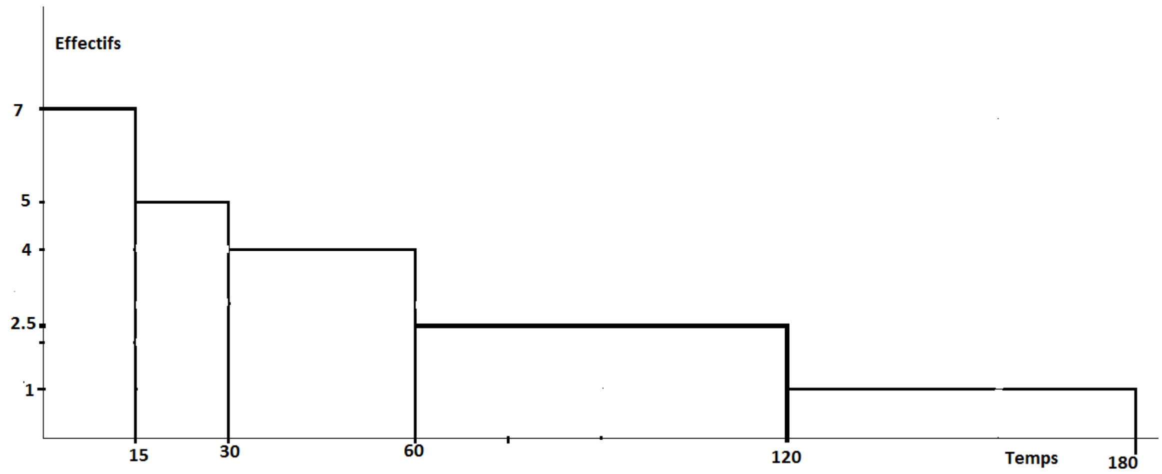
Remarque : En pratique, il est conseillé de commencer par construire un tableau donnant la largeur et l'aire de chaque rectangle (selon les unités choisies).On peut alors facilement en déduire la hauteur de chaque rectangle ce qui facilite la construction graphique de l'histogramme :

Exemple :

Le temps passé devant la télévision par 34 élèves pendant une certaine journée :

Temps en minute	[0,15[	[15,30[	[30,60[	[60,120[	[120,180[
Nombre d'élèves	7	5	8	10	4
L'aire du rectangle en cm^2	7	5	8	10	4
Largeur du rectangle en cm	1	1	2	4	4
Hauteur $= \frac{aire}{Largeur}$ **en cm**	7	5	4	2.5	1

1cm $\longrightarrow$ 15 cm



Echantillons

Lorsqu'on travaille sur une population de grande taille, il est rarement possible d'avoir accés aux données relatives à l'ensemble de population.

On utilise alors un échantillon de cette population.

Définition :

Un échantillon de taille n est une séléction de n individus choisis "au hasard" dans une population.

Intervalle de fluctuation :

Si l'on effectue plusieurs échantillage de même taille sur une même population, on obtiendra en général des fréquences légérement differentes pour un caractère donnée donnée.

Voici par les résultats que l'on pourrait obtenir en prélevant l'échantillons de 100 personnes (dans une éléction).

échantillon	n1	n2	n3	n4	n5
Fréquence en %	52 %	55 %	42%	50%	48%

Ce phénomène s'appelle fluctuation d'échantillonage.

Théorème et Définition

On note p la proportion d'un caractère dans une population donnée.On prélève un échantillon de taille n de cette population et on note f la fréquence du caractère dans l'échantillon.

Si $0.2 \leq p \leq 0.8$ et si $n \geq 25$ alors dans au moins 95% des cas $f \in I = [p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}]$

I est appelé intervalle de fluctuation au seuil 95%

Remarque :

Bien retenir la signification de chacune des variables :

p proportion du caractère dans l'ensemble de la population :

f la fréquence du caractère dans l'échantillon :

n Taille de l'échantillon :

Exemple

Entreprise A	**Entreprise B**
100 employés dont 43 femmes	2500 employés dont 1150 femmes

Quelle entreprise respecte la mesure de parité H/F ?

Proportion Théorique p= 50%	Proportion Théorique p= 50%
$f_A = \frac{43}{100} = 0.43$	$f_B = \frac{1150}{2500} = 0.46$
Intervalle de fructuation	
$I_{f_A} = [0.5 - \frac{1}{10}, 0.5 + \frac{1}{10}]$	$I_2 = [0.5 - \frac{1}{50}, 0.5 + \frac{1}{50}]$
$= [0.4, 0.6]$	$= [0.48, 0.52]$
$f_A \in I_{f_A}$	$f_B \notin I_2$

L'entreprise A repecte la parité homme et femme