



**CORRECTION : Statistique descriptive**  
**IUT STID, 1<sup>ère</sup> année,**  
**Devoir 1**  
 Vendredi 27 octobre 2006

## 1 Questions de cours

---

- Pour représenter les effectifs cumulés d'une variable quantitative continue regroupée en classes, on peut utiliser :
  - soit un histogramme cumulatif : sur un axe horizontal gradué, on place les classes ; chaque classe est représentée par un rectangle dont l'aire est proportionnelle à son effectif cumulé ;
  - soit un polygone cumulatif : sur un axe horizontal gradué, on place les classes et sur un axe vertical gradué, les effectifs cumulés. On représente ensuite le graphe de la fonction qui, à tout réel  $x$  associe son effectif cumulé (il s'agit d'une ligne brisée).
- Dans l'histogramme cumulatif, il peut arriver que la hauteur d'un rectangle soit inférieure à la hauteur du rectangle qui le précède : en effet, la hauteur est proportionnelle aux effectifs cumulés divisés par les amplitudes des classes. Lorsque les classes en sont pas toutes de même amplitude, les hauteurs peuvent donc ne pas être croissantes, ce qui peut paraître surprenant pour un diagramme cumulatif.

## 2 Exercice

---

- Comme  $X_2 = 2X_1$ , on sait, par une propriété vue en cours, que  $\bar{X}_2 = 2\bar{X}_1$  et  $\sigma_{X_2} = 2\sigma_{X_1}$ .
- La variable  $X_2$  est répartie comme suit :

Salaires	[2000 ; 3000[	[6000 ; 10000[
Effectifs	50	10

- Calculons les coordonnées de la courbe de Lorenz de la variable  $X_1$  :

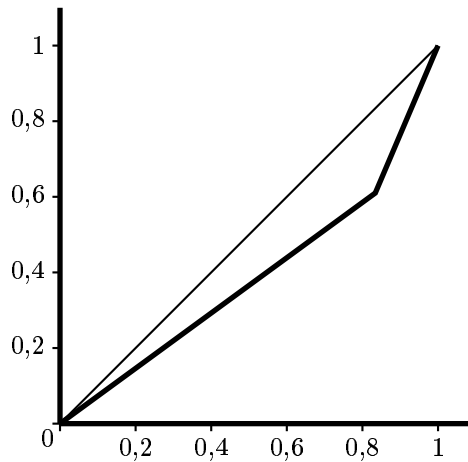
$X_1$	[1000 ; 1500[	[3000 ; 5000[	
Effectifs	50	10	60
$f_i^*$	$\frac{5}{6}$	1	
$c_i$	1250	4000	
$n_i c_i$	62 500	40 000	102 500
$v_i$	$\frac{62500}{102500} \simeq 0,610$	1	

ainsi que celles de la variable  $X_2$  :

$X_1$	[2000 ; 3000[	[6000 ; 10000[	
Effectifs	50	10	60
$f_i^*$	$\frac{5}{6}$	1	
$c_i$	2500	8000	
$n_i c_i$	125 000	80 000	205 000
$v_i$	$\frac{125000}{205000} \simeq 0,610$	1	

ce qui nous permet de constater que les deux courbes de Lorenz sont identiques. On obtient donc le graphique suivant :

**Courbe de Lorenz des variables  $X_1$  et  $X_2$**



Le calcul des indices de Gini (identique pour les deux variables) se fait en prolongeant le tableau de calculs précédent :

$X_1$	[1000 ; 1500[	[3000 ; 5000[	
Effectifs	50	10	60
$v_i$	$\frac{62500}{102500} \simeq 0,610$	1	
$v_{i-1} + v_i$	0,610	1,610	
$f_i$	$\frac{5}{6}$	$\frac{1}{6}$	
$f_i(v_{i-1} + v_i)$	0,508	0,268	0,776

ce qui donne  $G = 1 - 0,776 = 0,224$ . Ces deux traitements statistiques (courbes de Lorenz et indice de Gini) montrent, pour les deux distributions, une concentration plutôt faible de la masse salariale (même si 17 % des plus hauts revenus concentrent 39 % de la masse salariale).

- Le fait de doubler les salaires de tous les employés au mois de décembre a pour conséquence une augmentation globale des salaires (le salaire moyen est doublé) mais aussi de la dispersion de ceux-ci (l'écart type est doublé). En conséquence, cela ne réduit pas les inégalités présentes dans l'entreprise puisque l'indice de Gini constant est le signe d'une concentration identique de la masse salariale.

### 3 Problème

- La population est l'ensemble des chefs d'exploitation de la Corrèze (en 1998). Sa taille est égale à  $n = 0 + 149 + 396 + 1665 + 1857 + 1421 + 2053 + 1400 + 1756 = 10\,697$  chefs d'exploitation.

La variable étudiée est l'âge, de type quantitatif continu.

- Pour la suite du problème, nous construisons un tableau de calculs :

$C_i$	$n_i$	$c_i$	$n_i c_i$	$n_i c_i^2$	$n_i^*$
[10; 15[	0	12,5	0	0	0
[15; 25[	149	20	2 980	59 600	149
[25; 30[	396	27,5	10 890	299 475	545
[30; 40[	1 665	35	58 275	2 039 625	2 210
[40; 50[	1 857	45	83 565	3 760 425	4 067
[50; 55[	1 421	52,6	74 602,5	3 916 631,25	5 488
[55; 60[	2 053	57,5	118 047,5	6 787 731,25	7 541
[60; 65[	1 400	62,5	87 500	5 468 750	8 941
[65; 80[	1 756	72,5	127 310	9 229 975	10 697
TOTAL	10 697		563 170	31 562 212,5	

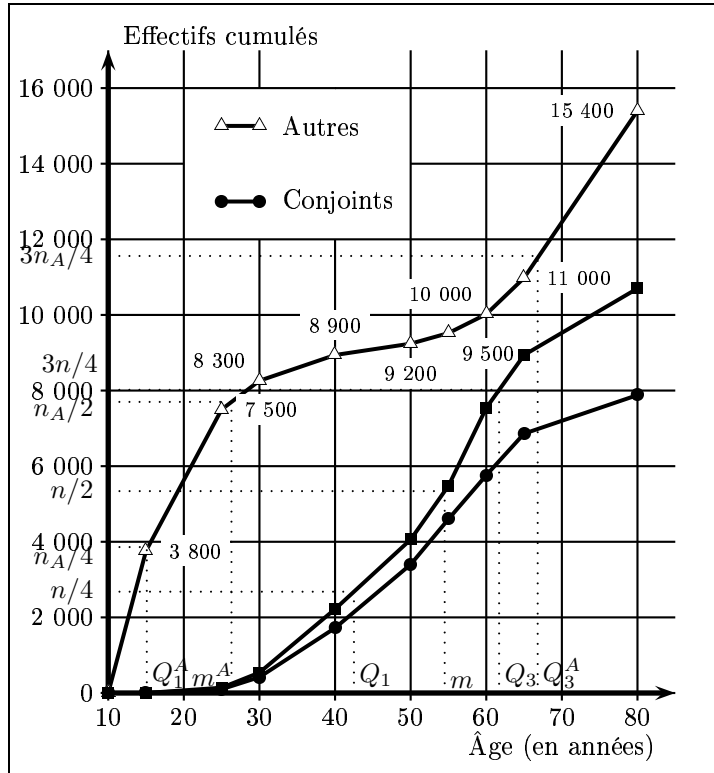
d'où nous déduisons :

- $\bar{X} = \frac{563170}{10697} \simeq 52,65$  ans ;
  - $Var(X) = \frac{31562212,5}{10697} - \bar{X}^2 \simeq 178,5442$  et donc  $\sigma_X = \sqrt{Var(X)} \simeq 13,36$  ans.
3.  $\frac{n}{2} = \frac{10697}{2} = 5348,5$  donc la classe médiane est la classe dont le premier effectif cumulé dépasse 5 348,5, c'est-à-dire, la classe  $[50 ; 55[$ . Ainsi,

$$m = 50 + (55 - 50) \times \frac{5348,5 - 4067}{5488 - 4067} \simeq 54,51 \text{ ans.}$$

La comparaison moyenne / médiane indique une légère dissymétrie de la série statistique vers la gauche, c'est à dire, un étirement vers les âges faibles.

4. (a) Complétons le graphique fourni en énoncé :

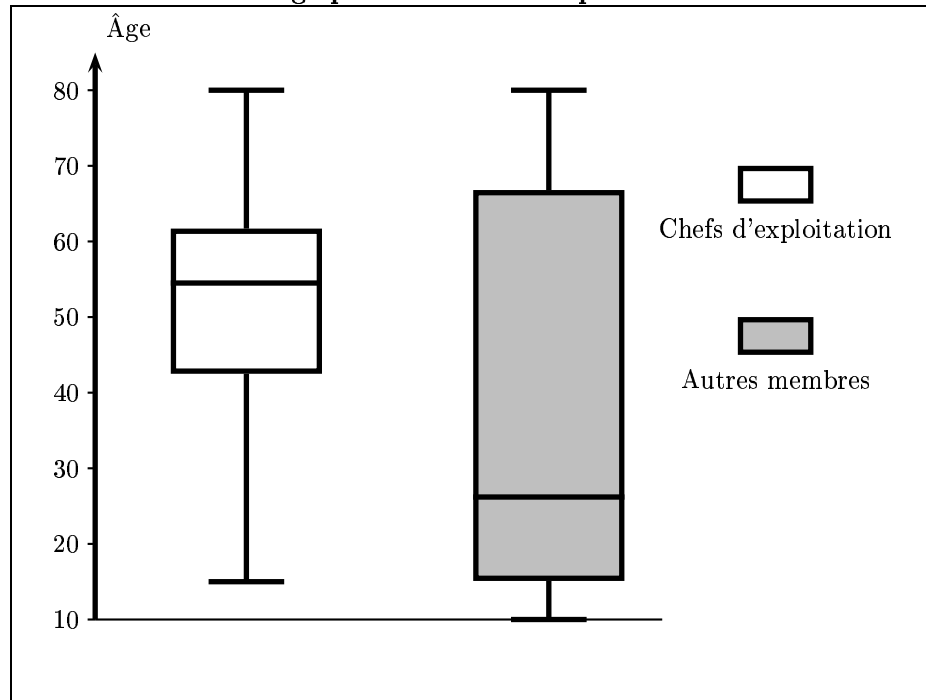


- (b)  $\frac{n}{4} = 2674,25$  et  $\frac{3n}{4} = 8022,75$  d'où, graphiquement, on lit,  $m \simeq 55$  ans,  $Q_1 \simeq 43$  ans,  $Q_3 \simeq 62$  ans.

- (c) La taille des autres membres des exploitations est  $n_A = 15400$  d'où  $\frac{n}{4} = 3850$ ,  $\frac{n}{2} = 7700$  et  $\frac{3n}{4} = 11550$ . On lit donc les trois quartiles de l'âge des autres membres de l'exploitation :  $Q_1^A \simeq 15$  ans,  $m^A \simeq 26$  ans et  $Q_3^A \simeq 67$  ans.

On en déduit les boîtes à moustaches suivantes :

**Boîtes à moustaches de l'âge pour les chefs d'exploitation et les autres membres**



(d) L'âge des chefs d'exploitation est réparti de manière relativement homogène avec un léger étirement vers les âges les plus jeunes (peu de chefs d'exploitation très jeunes). Au contraire, les autres membres de l'exploitation sont soit très jeunes (l'âge médian est 26 ans), soit assez âgés (le troisième quartile est 67 ans) mais il y a une faible concentration de ces personnes dans la période 30/60 ans.

5. (a) Tableau d'effectifs pour les autres membres de l'exploitation :

Âge (en années)	$n_i^*$	$n_i$
[10 ; 15[	3 800	3 800
[15 ; 25[	7 500	3 700
[25 ; 30[	8 300	800
[30 ; 40[	8 900	600
[40 ; 50[	9 200	300
[50 ; 55[	9 500	300
[55 ; 60[	10 000	500
[60 ; 65[	11 000	1 000
[65 ; 80[	15 400	4 400

(b) L'âge moyen est globalement inférieur chez les autres membres de l'exploitation, ce qui confirme la part importante des jeunes chez ceux-ci. L'écart type chez les autres membres de l'exploitation est beaucoup plus élevé : cela tend à montrer que la dispersion est plus importante dans l'âge de ceux-ci que chez les chefs d'exploitation et est due à la part importante des jeunes et des séniors.