



Énoncé Statistique descriptive

IUT STID, 1<sup>ère</sup> année

Devoir 1 : *Correction*

25 octobre 2007

Nom : Nathalie Villa-Vialaneix

**Attention !** Toutes les questions doivent être effectuées sur la feuille d'énoncé! **Aucune autre copie ne sera acceptée.**

Par ailleurs, il sera **tenu compte des justifications et de la rédaction** des réponses dans la notation.

## 1 Quantiles

On répondra sur la place laissée libre en expliquant la méthode utilisée (calculs effectués et/ou traits de construction).

Déterminer, dans les situations suivantes, le 2<sup>ème</sup> quantile d'ordre 3 que l'on notera  $q_2^{(3)}$  :  
(Les situations sont indépendantes les unes des autres.)

**Situation 1 :** Les données sont fournies sous forme brutes :

12 ; 14 ; 11 ; 17 ; 14 ; 15 ; 10 ; 12 ; 17 ; 13 ; 14 ; 15 ; 15 ; 10 ; 11 ; 11 ; 12

Réponse : Pour les données ci-dessus, on a  $N = 17$ .  $q_2^{(3)}$  est donc la  $(\lfloor \frac{2 \times 17}{3} \rfloor + 1) = 12^{\text{ème}}$  valeur dans l'ordre croissant des valeurs. Or, les valeurs rangées dans l'ordre croissant sont :

10 ; 10 ; 11 ; 11 ; 11 ; 12 ; 12 ; 12 ; 13 ; 14 ; 14 ; 14 ...

soit  $q_2^{(3)} = 14$ .

**Situation 2 :** Les données sont fournies sous la forme d'un tableau d'effectifs :

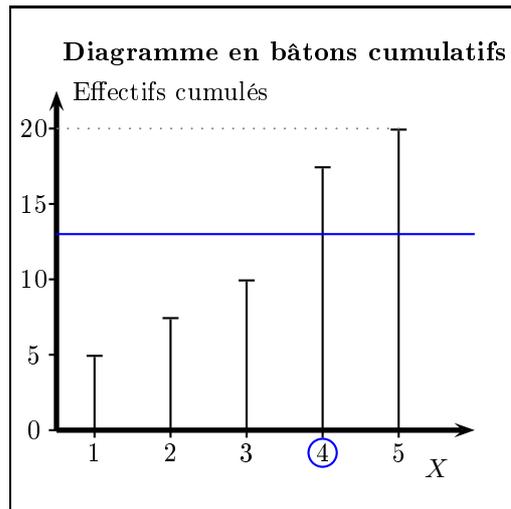
$x_i$	A	B	C	D
$n_i$	4	3	6	2

Réponse : Pour les données ci-dessus,  $N = 4 + 3 + 6 + 2 = 15$ . Ainsi,  $\frac{15 \times 2}{3} = 10$ . Or, le tableau d'effectifs cumulés est :

$x_i$	A	B	C	D
$n_i$	4	3	6	2
$n_i^*$	4	7	13	15

donc le premier effectif cumulé dépassant 10 est 13 et donc  $q_2^{(3)} = C$ .

**Situation 3 :** Les données sont fournies d'un diagramme en bâtons :



Réponse : On lit  $N = 20$  donc  $\frac{20 \times 2}{3} \simeq 13,33$ . On lit alors sur le graphique  $q_2^{(3)} = 4$ .

**Situation 4** Les données sont regroupées en classes et on connaît le tableau d'effectifs cumulés :

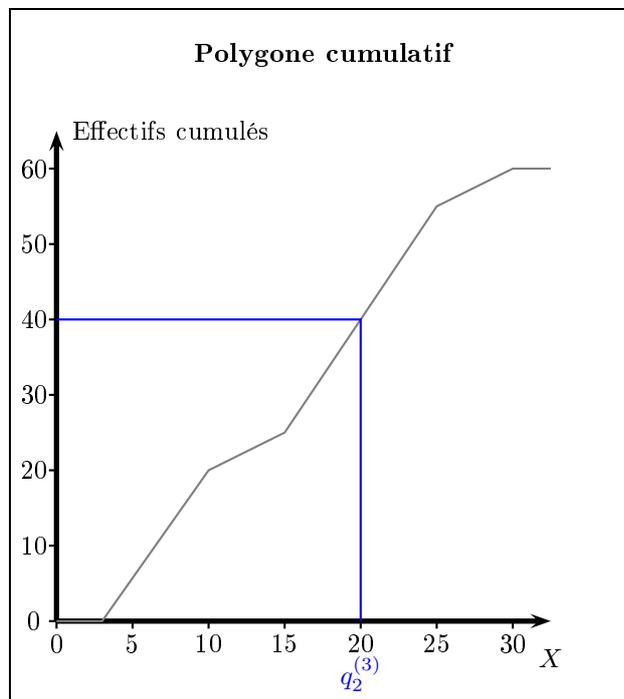
$\mathcal{C}_i$	$[0; 10[$	$[10; 12[$	$[12; 15[$	$[15; 20[$	$[20; 30[$	$[30; 50[$
$n_i^*$	5	10	20	25	50	60

Réponse :  $N = 60$  donc  $\frac{60 \times 20}{3} = 40$  ce qui prouve que  $q_2^{(3)}$  appartient à la classe  $[20; 30[$ . Sa valeur exacte est donnée par le calcul :

$$q_2^{(3)} = 20 + (30 - 20) \times \frac{40 - 25}{50 - 26}$$

soit  $q_2^{(3)} = 22$ .

**Situation 5** Les données sont connues par le biais d'un polygone cumulatif :

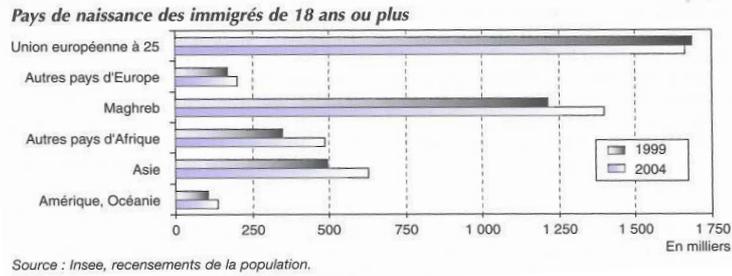


Réponse : On lit  $N = 60$  soit  $\frac{60 \times 2}{3} = 40$  et on en déduit, par lecture graphique,  $q_2^{(3)} \simeq 20$ .

## 2 Représentations graphiques

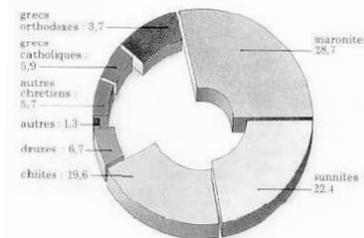
Pour chacun des graphiques ci-dessous, répondre aux questions posées :

**Graphique 1** : Le graphique ci-dessous concerne la situation française. (Source : France, portrait social, INSEE Références, 2006)



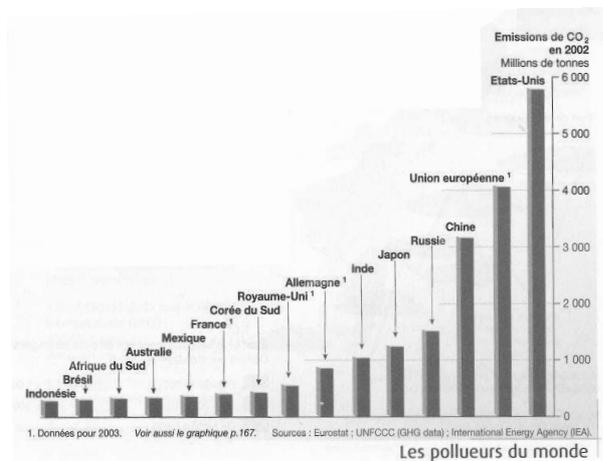
1. Quelle est (quelles sont) la (les) population(s) étudiée(s) ? *On a deux populations : la population des étrangers vivant en France en 2004 et la population des étrangers vivant en France en 1999.*
2. Quelle est (quelles sont) la (les) variable(s) étudiée(s) ? Son (leurs) type(s) ? *On étudie la variable Pays de naissance qui est de type qualitative nominale.*
3. Quel est le nom de ce graphique ? *Ce graphique est un diagramme en colonnes.*

**Graphique 2 :** Le graphique ci-dessous donne la répartition des différentes confessions au Liban. (Source : *Théma Encyclopédie, Le Monde Aujourd'hui (1991), Larousse, France*)



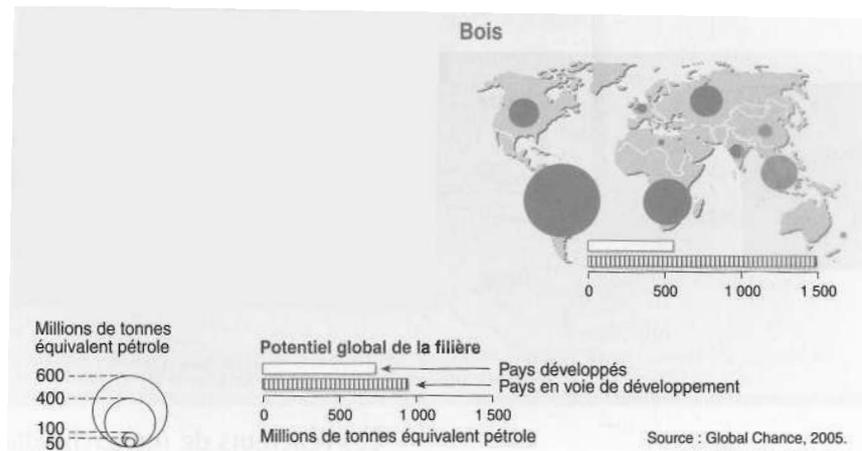
1. Quelle est (quelles sont) la (les) population(s) étudiée(s) ? *La population étudiée est la population libanaise.*
2. Quelle est (quelles sont) la (les) variable(s) étudiée(s) ? Son (leurs) type(s) ? *La variable étudiée est la Confession, de type qualitative nominale.*
3. Expliquer en quoi le graphique donne une représentation fautive des données. *Dans ce graphique, à la fois l'angle des arcs de portion cercle mais aussi leur largeur sont proportionnels à l'effectif. Ainsi, le volume des disques, qui est la partie visible, est proportionnelle à l'effectif au carré.*

**Graphique 3 :** (Source : *L'atlas du Monde Diplomatique, Édition 2006*)



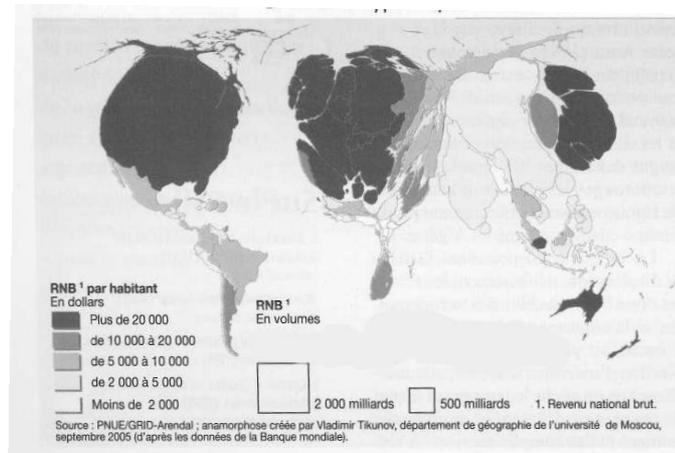
1. Quelle est (quelles sont) la (les) population(s) étudiée(s) ? Quelle est la taille de la (des) population(s) ? *La population étudiée est un ensemble de pays du monde. Sa taille est N = 15.*
2. Quelle est (quelles sont) la (les) variable(s) étudiée(s) ? Son (leurs) type(s) ? *La variable étudiée est le Volume d'émission de CO<sub>2</sub>, de type quantitative continue.*

**Graphique 4 :** Le graphique ci-dessous concerne le **potentiel mondial en énergie renouvelable** (ici, le bois). (Source : *L'atlas du Monde Diplomatique*, Édition 2006)



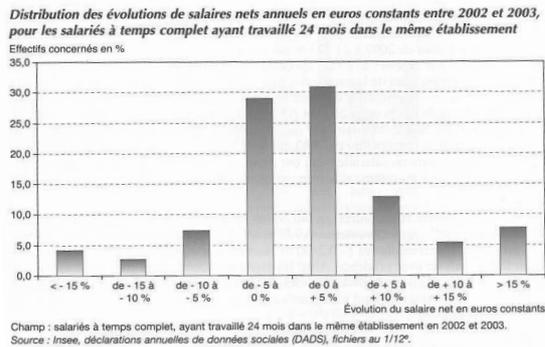
1. Quelle est (quelles sont) la (les) population(s) étudiée(s)? *La population étudiée est un ensemble de régions géographiques.*
2. Quelle est (quelles sont) la (les) variable(s) étudiée(s)? Son (leurs) type(s)? *La variable étudiée est le Potentiel énergétique en bois, de type quantitative continue.*
3. Quelle est la méthode de construction des disques? *Les disques ont des aires proportionnelles à la valeur du potentiel énergétique en bois de la région correspondante.*

**Graphique 5 :** Le graphique ci-dessous a pour titre « **Le monde hypertrophié des riches en 2003** ». (Source : *L'atlas du Monde Diplomatique*, Édition 2006)



1. Quelle est (quelles sont) la (les) population(s) étudiée(s)? *La population étudiée est l'ensemble des pays du monde.*
2. Quelle est (quelles sont) la (les) variable(s) étudiée(s)? Son (leurs) type(s)? *Les variables étudiées sont le Revenu National Brut et le Revenu National Brut par habitant, toutes deux de type quantitative continue.*
3. Comment chaque pays est-il représenté? *Chaque pays a une aire proportionnelle à son RNB et est colorié par un niveau de couleur correspondant à son RNB par habitant.*

**Graphique 6 :** (Source : *France, portrait social, INSEE Références*, 2006)



1. Quelle est (quelles sont) la (les) population(s) étudiée(s)? *La population étudiée est les salariés français à temps plein ayant travaillé au moins 24 mois dans le même établissement.*
2. Quelle est (quelles sont) la (les) variable(s) étudiée(s)? Son (leurs) type(s)? *La variable étudiée est l'évolution du salaire en euros constants entre 2002 et 2003 qui est de type quantitative continue.*
3. Critiquer ce graphique. Avec quel type de graphique aurait-il fallu représenter ces données? *La variable étudiée est quantitative continue et le graphique représente les effectifs de diverses classes de cette variable. Or, l'axe des abscisses, sur lequel est placée la variable n'est pas gradué; il existe des espaces sans sens entre deux classes consécutives, les classes ne sont pas toutes de même amplitude et sont représentées par des colonnes de même largeur, etc. Le diagramme qui aurait convenu pour cette représentation est un histogramme.*

### 3 Problème

Voici la répartition des différentes tranches d'âges en Limousin et en Île de France en 2004 :  
(Source : « La France et ses régions », Publication INSEE-Références, Édition 2006)

Limousin :

Âge	[0; 10[	[10; 25[	[25; 60[	[60; 75[	[75; 100]
$f_i$ (%)	9,7	16,1	46,3	16,0	11,9

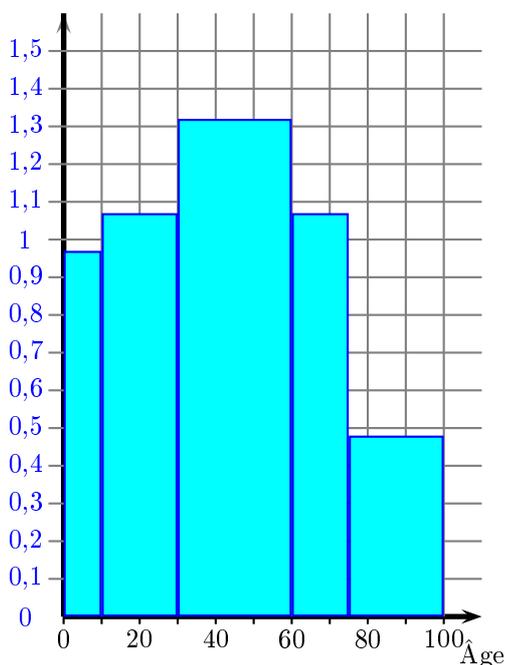
Île de France :

Âge	[0; 10[	[10; 25[	[25; 60[	[60; 75[	[75; 100]
$f_i$ (%)	13,4	19,9	50,4	10,4	5,9

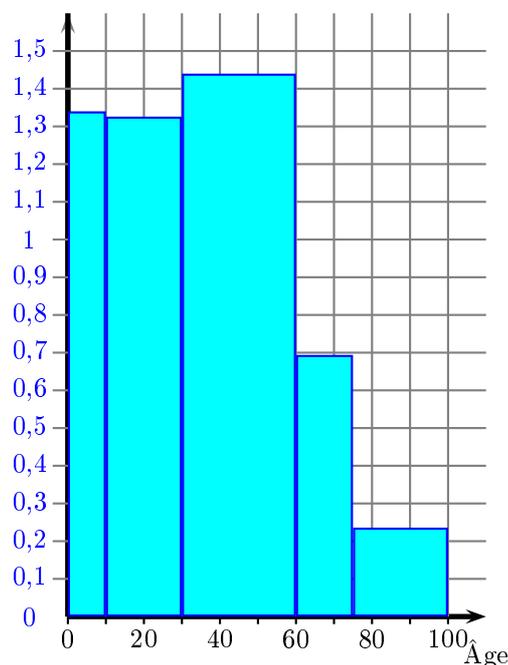
1. Compléter les deux histogrammes ci-dessous en choisissant sur l'axe des ordonnées une graduation appropriée (qui soit la même pour les deux histogrammes afin de permettre une comparaison).  
Réponse :

<i>Limousin</i>					
Âge	[0; 10[	[10; 25[	[25; 60[	[60; 75[	[75; 100]
$d_i$	$\frac{9,7}{10} = 0,97$	$\frac{16,1}{15} = 1,07$	$\frac{46,3}{35} = 1,32$	$\frac{16,0}{15} = 1,07$	$\frac{11,9}{25} = 0,48$
<i>Île de France</i>					
Âge	[0; 10[	[10; 25[	[25; 60[	[60; 75[	[75; 100]
$d_i$	$\frac{13,4}{10} = 1,34$	$\frac{19,9}{15} = 1,33$	$\frac{50,4}{35} = 1,44$	$\frac{10,4}{15} = 0,69$	$\frac{5,9}{25} = 0,24$

Âge de la population en Limousin  
(histogramme)



Âge de la population en Île de France  
(histogramme)



Commenter la différence entre ces deux graphiques.

Réponse : La population d'Île de France est jeune (forte concentration d'individus sur les 3 premières classes) ; la population du Limousin est beaucoup plus âgée, avec moins de jeunes (première et seconde classes) et plus de personnes âgées (plus de 60 ans), même si la classe majoritaire reste la classe des actifs (25-60 ans).

2. Déterminer la moyenne et l'écart type,  $\bar{X}_L$  et  $\sigma_{X_L}$ , de l'âge en Limousin.

Réponse :

$$\bar{X}_L = (5 \times 0,097 + 17,5 \times 0,161 + 42,5 \times 0,463 + 67,5 \times 0,16 + 87,5 \times 0,119) \simeq 44,19 \text{ ans ;}$$

et

$$\text{Var}(X_L) = (5^2 \times 0,097 + 17,5^2 \times 0,161 + 42,5^2 \times 0,463 + 67,5^2 \times 0,16 + 87,5^2 \times 0,119) - 44,19^2 \simeq 575,14$$

d'où  $\sigma_{X_L} = \sqrt{\text{Var}(X_L)} \simeq 23,98 \text{ ans}$ .

3. On donne, pour l'âge en Île de France,  $\bar{X}_{IdF} \simeq 37,76$  et  $\sigma_{X_{IdF}} \simeq 21,79$ . Commenter la différence entre ces valeurs et celles trouvées à la question précédente.

Réponse : L'âge moyen en Limousin est sensiblement plus élevé qu'en Île de France, ce qui confirme le fait que la population limousine est plus âgée que la population parisienne. Par ailleurs, l'écart type de l'âge en Limousin est légèrement plus élevé que l'écart type de l'âge en Île de France, ce qui s'explique par le fait que la population d'Île de France est essentiellement composée de moins de 60 ans alors que la population limousine est plus répartie sur l'ensemble des classes d'âge.

4. Comparativement à la population régionale, est-on plus « jeune » à 25 ans en Île de France ou à 30 ans en Limousin ?

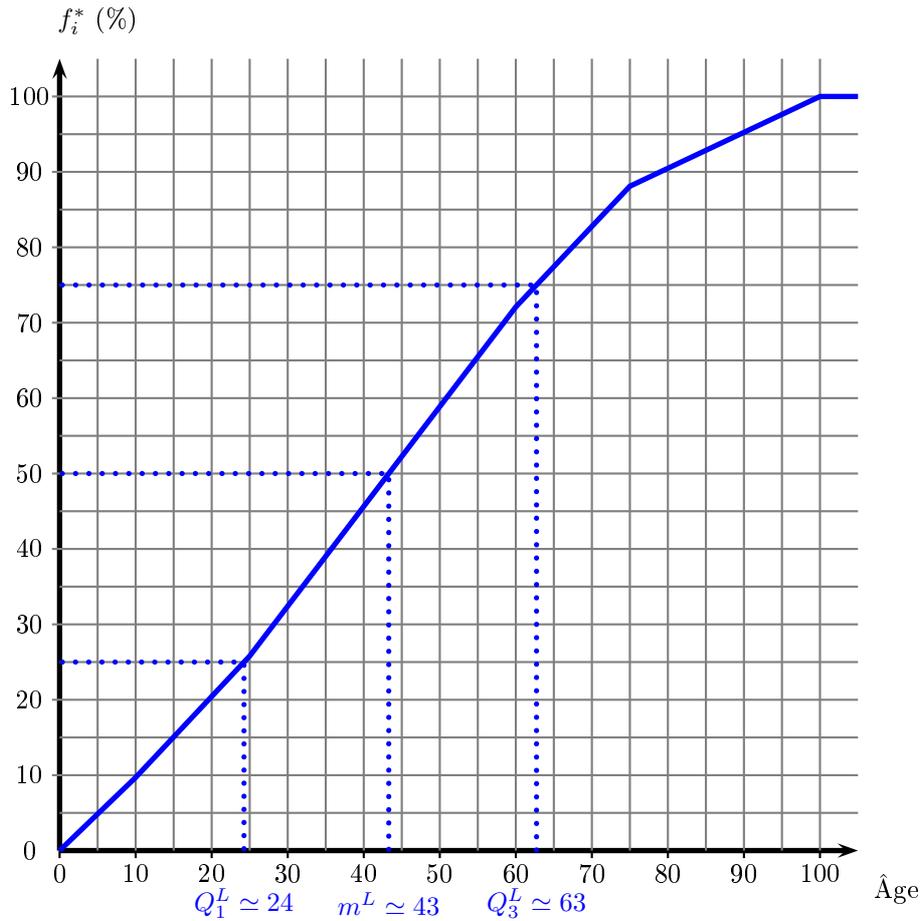
Réponse :  $\tilde{X}_L(30) = \frac{30-44,19}{23,98} \simeq -0,592$  et  $\tilde{X}_{IdF}(25) = \frac{25-37,76}{21,79} \simeq -0,586$ . On peut donc dire qu'on est légèrement moins âgé à 30 ans en Limousin qu'à 25 en Île de France comparativement à la population régionale.

5. Construire le polygone cumulatif de la variable « Âge » en Limousin sur le repère ci-dessous :

Réponse : Tableau des fréquences cumulées de l'âge en Limousin :

Âge	[0; 10[	[10; 25[	[25; 60[	[60; 75[	[75; 100]
$f_i^*$ (%)	9,7	25,8	72,1	88,1	100

**Polygone cumulatif de l'âge de la population limousine**



6. Déterminer graphiquement les trois quartiles,  $Q_1^L$ ,  $m^L$  et  $Q_3^L$ , de la variable « Âge » pour la population limousine.

Réponse : Détermination de  $Q_1^L$  :  $N/4 = 25$ ; on lit :  $Q_1^L \simeq 24$  ans.

Détermination de  $m^L$  :  $N/2 = 50$ ; on lit :  $m^L \simeq 43$  ans.

Détermination de  $Q_3^L$  :  $(3N)/4 = 75$ ; on lit :  $Q_3^L \simeq 63$  ans.

7. Déterminer numériquement l'âge médian de la population limousine.

Réponse : D'après le tableau d'effectifs cumulés de la question 5.,  $m^L$  se situe dans la classe  $[25; 60[$ ; sa valeur est donc égale à :

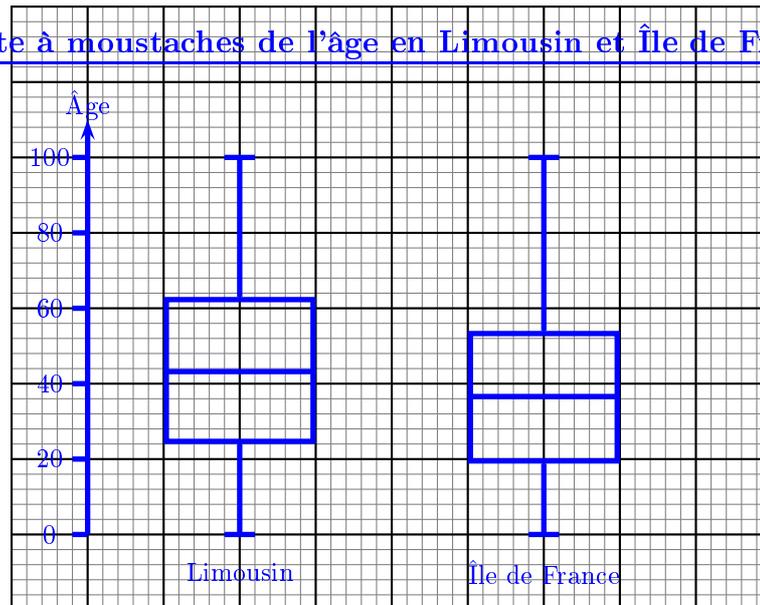
$$m^L = 25 + (60 - 25) \times \frac{50 - 25,8}{72,1 - 25,8} \simeq 43,29$$

ce qui est cohérent avec la valeur trouvée graphiquement.

8. On donne, pour la population d'Île de France,  $Q_1^{IdF} \simeq 18,7$ ,  $m^{IdF} \simeq 36,6$  et  $Q_2^{IdF} \simeq 54,0$ . Effectuer côte à côte sur un même repère les boîtes à moustaches de la répartition des âges en Île de France et en Limousin.

Réponse :

## Boîte à moustaches de l'âge en Limousin et Île de France



Commenter la différence entre les deux boîtes à moustaches. Retrouve-t-on des commentaires similaires à ceux effectués précédemment (histogrammes et moyennes/écart type) ?

Réponse : *On retrouve, quoique de manière moins visible, les commentaires faits précédemment : la population d'Île de France est composée d'un plus grand de jeunes (« boîte » située plus bas) et d'un nombre moins important de personnes âgées (« moustache supérieure » plus grande).*