

Corrigé

Exercice 1 [10 pts]. L'évolution de l'âge des mères à la naissance des enfants en France métropolitaine (1975-2015)

- 1) [3 pts] Calculez pour l'année 2015, l'âge médian des mères à la naissance de leur enfant ainsi que les 1^{er} (Q1) et 3^e quartiles (Q3) et les 1^{er} et 9^e déciles (D1 et D9). Calculez également les 1^{er} et 9^e déciles pour l'année 1975.
- 2) [4 pts] Calculez pour l'année 2015, l'âge moyen des mères à la naissance de leur enfant ainsi que la dispersion moyenne autour de cette valeur (variance et écart-type).
- 3) [3 pts] Sur un même graphique, représentez les « boîtes à moustaches » pour les années 1975 et 2015. Puis synthétisez dans un commentaire de 10 à 20 lignes maximum les principales évolutions de la distribution des naissances selon l'âge de la mère entre 1975 et 2015.

Tableau 1 : Distribution des naissances de l'année 2015 selon l'âge révolu de la mère

Classe d'âges de la mère	Naissances 2015
15-19 ans	14 713
20-24 ans	94 006
25-29 ans	240 486
30-34 ans	253 586
35-39 ans	125 856
40-44 ans	29 907
45 ans et +	1 867
Total	760 421

Source : Insee, Etat-civil.

Champ : France métropolitaine

Tableau 2 : Cumul (en proportion) des naissances de l'année 1975 selon l'âge exact de la mère

Age exact de la mère	% cumulé des naissances 1975
< 15 ans	0%
< 20 ans	10%
< 25 ans*	47%
< 30 ans	81%
< 35 ans	93%
< 40 ans	98%
< 50 ans	100%

Source : Insee, Etat-civil

Champ : France métropolitaine

*Note de lecture : les mères de 47 % des 745 065 enfants nés en 1975 avaient moins de 25 ans.

Tableau 3 : Indicateurs de tendance centrale et de dispersion de l'âge des mères à la naissance des enfants nés en 1975

Indicateur	Valeur 1975
Médiane	25,4 ans
1 ^{er} quartile	22,1 ans
3 ^e quartile	29,1 ans
Moyenne	26,1 ans
Ecart-type	5,4 ans

Source : Insee, Etat-civil

Corrigé

1) Pour calculer les quantiles, il faut commencer par déterminer la répartition des naissances selon la classe d'âges de la mère. Pour chaque catégorie d'âges de la mère ($x, x+4$), on détermine donc une fréquence relative ($f_{x, x+4}$) en rapportant les naissances ($N_{x, x+4}$) à l'ensemble des naissances (N) de l'année 2015 :

$$f_{x, x+4} = \frac{N_{x, x+4}}{N}$$

Exemple. La part des naissances d'enfants dont les mères sont âgées de 15-19 ans lors de l'accouchement est :

$$f_{15-19} = \frac{N_{15-19}}{N} = \frac{14\,713}{760\,421} = 0,019 = 1,9\%$$

Puis on cumule l'ensemble de ces fréquences relatives pour déterminer la série des fréquences cumulées croissantes (f_c). Par exemple, la part de naissances issues de femmes dont l'âge est inférieur à 25 ans est la somme de la proportion de naissances à 15-19 ans et 20-24 ans :

$$fc_{<25} = f_{15-19} + f_{20-24} = 1,9 + 12,4 = 14,3\%$$

Tableau de synthèse des résultats

Classe d'âges de la mère	Naissances 2015	Répartition des naissances 2015 (%)	Bornes inférieures des classes d'âges	Cumul des naissances en %
15-19 ans	14 713	1,9%	moins de 15 ans	0,0%
20-24 ans	94 006	12,4%	moins de 20 ans	1,9%
25-29 ans	240 486	31,6%	moins de 25 ans	14,3%
30-34 ans	253 586	33,3%	moins de 30 ans	45,9%
35-39 ans	125 856	16,6%	moins de 35 ans	79,3%
40-44 ans	29 907	3,9%	moins de 40 ans	95,8%
45 ans et +	1 867	0,2%	moins de 45 ans	99,8%
Total	760 421	100,0%	moins de 50 ans	100,0%

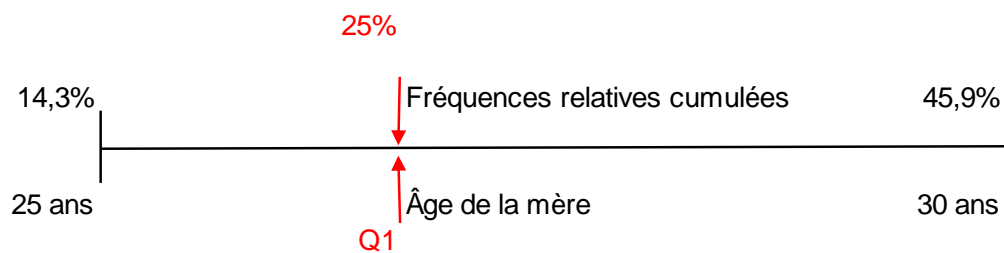
Il s'agit ensuite de déterminer l'âge correspondant à chacun des quantiles demandés.

- D1 (1^{er} décile) sépare la série statistique en deux parties inégales : l'âge des mères qui concentrent 10% des naissances est inférieur à D1, tandis que 90% des naissances sont issues de femmes ayant un âge supérieur à ce premier décile ;
- Q1 (1^{er} quartile) sépare la série statistique en deux parties inégales : l'âge des mères qui concentrent 25% des naissances est inférieur à Q1, tandis que 75% des naissances sont issues de femmes ayant un âge supérieur à ce premier quartile ;
- Q2 (2^e quartile) qui correspond à la médiane sépare la série statistique en deux parties égales : l'âge des mères qui concentrent la moitié des naissances est inférieur à Q2, tandis que l'autre moitié des naissances sont issues de femmes ayant un âge supérieur à la médiane ;
- Q3 (3^e quartile) sépare la série statistique en deux parties inégales : l'âge des mères qui concentrent 75% des naissances est inférieur à Q3, tandis que 25% des naissances sont issues de femmes ayant un âge supérieur à ce troisième quartile ;
- D9 (9^e décile) sépare la série statistique en deux parties inégales : l'âge des mères qui concentrent 90% des naissances est inférieur à D9, tandis que 10% des naissances sont issues de femmes ayant un âge supérieur à ce neuvième décile ;

Une fois que l'on a identifié l'intervalle d'âges au sein duquel se trouve le quantile, on fait l'hypothèse que sur cet intervalle d'âges les naissances se répartissent de manière régulière selon l'âge de la mère. De ce fait, on peut établir un rapport de proportionnalité entre :

- d'une part, la position relative de la fréquence cumulée correspondant au quantile considéré au sein de l'intervalle des fréquences relatives cumulées correspondant aux bornes de la tranche d'âges au sein de laquelle se trouve le quantile ;
- d'autre part, la position relative du quantile au sein de la tranche d'âges dans laquelle il se trouve.

Exemple pour le premier quartile :



On peut transcrire ce schéma de la manière suivante :

$$\frac{25\% - 14,3\%}{45,9\% - 14,3\%} = \frac{Q1 - 25}{30 - 25}$$

$$\frac{25\% - 14,3\%}{45,9\% - 14,3\%} \times (30 - 25) = Q1 - 25$$

$$Q1 = 25 + \frac{25\% - 14,3\%}{45,9\% - 14,3\%} \times (30 - 25) = 26,7 \text{ ans}$$

En adoptant la même démarche, on détermine chacun des autres quantiles :

$$D1 = 20 + \frac{10\% - 1,9\%}{14,3\% - 1,9\%} \times (25 - 20) = 23,3 \text{ ans}$$

$$Q2 = 30 + \frac{50\% - 45,9\%}{79,3\% - 45,9\%} \times (35 - 30) = 30,6 \text{ ans}$$

$$Q3 = 30 + \frac{75\% - 45,9\%}{79,3\% - 45,9\%} \times (35 - 30) = 34,4 \text{ ans}$$

$$D9 = 35 + \frac{90\% - 79,3\%}{95,8\% - 79,3\%} \times (40 - 35) = 38,2 \text{ ans}$$

Pour l'année 1975, on connaît déjà le premier décile à partir du tableau 2 de l'énoncé : 20 ans. Le 9^e décile est de son côté :

$$D9 = 30 + \frac{90\% - 81\%}{93\% - 81\%} \times (35 - 30) = 33,8 \text{ ans}$$

2) Le calcul de l'âge moyen des mères qui ont donné naissance à un enfant en 2015 nécessite de déterminer les centres de classes d'âges, qui correspondent à la moyenne simple des bornes inférieures de deux groupes d'âges consécutifs. On fait dans ce cas l'hypothèse d'une distribution régulière des femmes qui ont eu un enfant sur l'intervalles d'âge considéré.

L'âge moyen est la moyenne des centres de classes d'âges pondérés par le poids de chaque classe d'âges dans la distribution des naissances en 2015 :

$$a = 17,5 \times 0,019 + 22,5 \times 0,124 + \dots + 42,5 \times 0,039 + 47,5 \times 0,002 = 30,6 \text{ ans}$$

Puis, le calcul de la variance et de l'écart type nécessite de calculer pour chaque tranche d'âges le carré de l'écart à l'âge moyen à la maternité en 2015, puis d'effectuer la moyenne de ces carrés des écarts à la moyenne pondérés par le poids de chaque tranche d'âges dans la distribution des naissances en 2015 :

$$\begin{aligned} Var = & 0,019 \times (17,5 - 30,6)^2 + 0,124 \times (22,5 - 30,6)^2 \dots + 0,039 \times (42,5 - 30,6)^2 \\ & + 0,002 \times (47,5 - 30,6)^2 = 29,8 \text{ ans}^2 \end{aligned}$$

$$\text{Ecart - type} = \sqrt{\text{Variance}} = \sqrt{29,8} = 5,5 \text{ ans}$$

Tableau de synthèse des calculs

Classe d'âges de la mère	Centre de classe d'âges [1]	Répartition des naissances 2015 (%) [2]	Moyenne % Naissances x Centre de classe d'âges [1]x[2]	(Centre de classe d'âges - moyenne) au carré [3]	Variance [2]x[3]
15-19 ans	17,5	1,9%	0,3	172,9	3,3
20-24 ans	22,5	12,4%	2,8	66,4	8,2
25-29 ans	27,5	31,6%	8,7	9,9	3,1
30-34 ans	32,5	33,3%	10,8	3,4	1,1
35-39 ans	37,5	16,6%	6,2	46,9	7,8
40-44 ans	42,5	3,9%	1,7	140,49	5,5
45 ans et +	47,5	0,2%	0,1	283,9	0,7
Total		100,0%	30,6		29,8

3) Synthèse des résultats

Comparaison de la distribution des naissances en 1975 et 2015 selon l'âge de la mère (boîtes à moustaches)

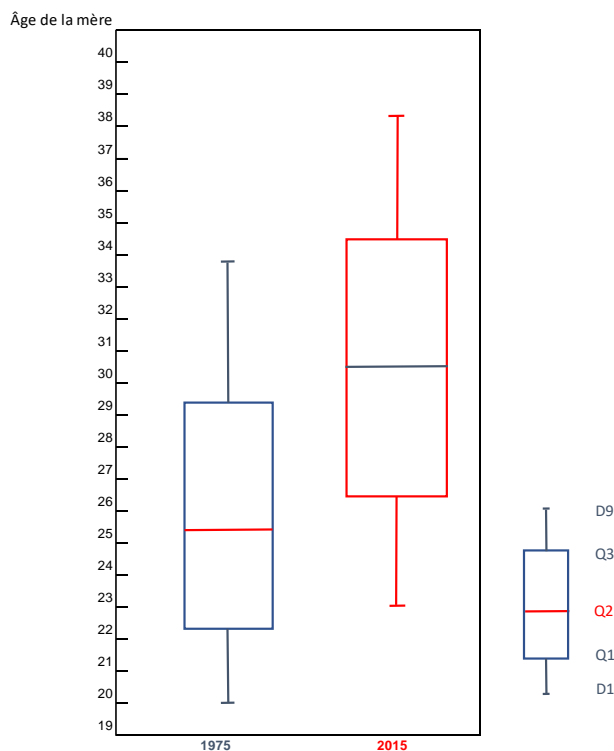


Tableau de synthèse des résultats

Indicateur	Valeur 1975	Valeur 2015
1er décile	20,0 ans	23,3 ans
1 ^{er} quartile	22,1 ans	26,7 ans
Médiane	25,4 ans	30,6 ans
3 ^e quartile	29,1 ans	34,4 ans
9e décile	33,8 ans	38,2 ans
Moyenne	26,1 ans	30,6 ans
Ecart-type	5,4 ans	5,5 ans

Commentaires :

En quarante ans, l'âge moyen des femmes qui ont eu des enfants dans l'année a augmenté de 4,5 ans : il était de 26,1 ans en 1975 et de 30,6 ans en 2015, des valeurs très proches voire identiques à celles des médianes (respectivement 25,4 ans et 30,6 ans). En fait, tout semble s'être passé comme si toutes les femmes en 2015 avaient eu leur enfant 4 à 5 ans plus tard que leurs aînées. La comparaison des quartiles à quarante ans d'intervalle est éloquente : le premier quartile a augmenté de 4,6 ans, la médiane de 5,2 ans et le troisième quartile de 5,3 ans. Alors que 50 % des femmes qui ont eu un enfant en 1975 avaient entre 22,1 et 29,1 ans (un intervalle de 7 années), la moitié de celles qui ont donné naissance à un enfant en 2015 ont cette fois entre 26,7 et 34,4 ans (un intervalle de 7,7 ans décalé de 5 ans environ par rapport à la situation de l'année 1975). Cette translation de 5 ans environ de l'âge de quasiment toutes les mères entre 1975 et 2015 explique que la dispersion des âges autour de la valeur moyenne n'ait pas varié : elle est toujours d'environ 5,5 ans. La comparaison des boîtes à moustaches résume bien, de manière graphique, ce vieillissement généralisé de l'âge des mères à la naissance de leur enfant en quarante ans.

Exercice 2 [10 pts]. La contribution du sexe et de la catégorie professionnelle sur la variance de l'espérance de vie à 35 ans en 2003

À partir d'une décomposition de la variance totale, indiquez laquelle des deux caractéristiques, le sexe ou la catégorie professionnelle, avait le plus d'effet sur la dispersion des espérances de vie à 35 ans de la population active française en 2003 (tableau 1).

- [2 pts] Complétez le tableau 1.
- [6 pts] Calculez la contribution du sexe à la variance totale à partir de la variance inter-groupe, puis la contribution de la catégorie professionnelle à partir de la variance intra-groupe.
- [2 pts] Commentez les données du tableau 1 et vos résultats afin de dégager en quelques lignes la ou les principales informations relatives à l'inégale espérance de vie à 35 ans des actifs français en 2003.

Tableau 1. Espérance de vie à 35 ans (e_{35}) de la population active selon le sexe et la catégorie professionnelle en France en 2003

Catégorie professionnelle	Sexe				Ensemble	
	Hommes		Femmes			
	%	e_{35}	%	e_{35}	%	e_{35}
Cadres supérieurs	9,1	46,6	4,5	50,9		
Professions intermédiaires	12,0	44,8	8,0	49,8		
Agriculteurs exploitants	3,4	45,3	2,8	50,1		
Indépendants	5,7	44,4	2,8	50,1		
Employés	5,7	42,1	19,9	49,4		
Ouvriers	19,3	40,9	6,8	48,6		
Ensemble	55,2	43,4	44,8	49,6		

Source : d'après E. Cambois, C. Laborde, J.-M. Robine (2008). La « double peine » des ouvriers : plus d'années d'incapacité au sein d'une vie plus courte, *Population et Sociétés*, n° 441.

Corrigé

1) Les fréquences relatives qui figurent dans le tableau 1 sont des fréquences élémentaires. Leur somme est égale à 1 ou 100 %.

De ce fait, pour un même sexe, la somme des fréquence relatives élémentaires permet de déterminer la fréquence marginale « en colonne », qui permet ainsi de connaître au sein de la population étudiée la part des hommes (55,2 %) et celle des femmes (44,8 %).

Pour une même catégorie professionnelle, la somme des fréquences relatives élémentaires correspond à la fréquence marginale « en ligne », soit la part de cette catégorie professionnelle au sein de la population étudiée. Par exemple pour les cadres supérieurs, comme les cadres supérieurs masculins ($f_{H,CS}$) et féminins ($f_{F,CS}$) représentent respectivement 9,1 % et 4,5 % de la population totale, les cadres supérieurs représentent 13,6 % de la population étudiée.

$$f_{CS,H} + f_{CS,F} = \frac{n_{CS,H}}{N} + \frac{n_{CS,F}}{N} = \frac{n_{CS}}{N} = f_{CS}.$$

Pour chaque catégorie professionnelle, l'espérance de vie à 35 ans moyenne est la moyenne des espérances de vie à 35 ans de chaque sexe pondérée par la proportion de chacun des sexes au sein de la catégorie professionnelle considérée. Par exemple, pour les cadres supérieurs :

$$\begin{aligned}
 e_{35,CS} &= \frac{e_{35,CS,H} \times f_{CS,H} + e_{35,CS,F} \times f_{CS,F}}{f_{CS,H} + f_{CS,F}} = \frac{e_{35,CS,H} \times f_{CS,H}}{f_{CS,H} + f_{CS,F}} + \frac{e_{35,CS,F} \times f_{CS,F}}{f_{CS,H} + f_{CS,F}} \\
 &= \frac{e_{35,CS,H} \times f_{CS,H}}{f_{CS}} + \frac{e_{35,CS,F} \times f_{CS,F}}{f_{CS}} = e_{35,CS,H} \times f_{H/CS} + e_{35,CS,F} \times f_{F/CS} \\
 &= \frac{46,6 \times 9,1 + 50,9 \times 4,5}{9,1 + 4,5} = 48,0 \text{ ans}
 \end{aligned}$$

Tableau de synthèse :

Catégorie professionnelle	Sexe				Ensemble	
	Hommes		Femmes			
	%	e ₃₅	%	e ₃₅	%	e ₃₅
Cadres supérieurs	9,1	46,6	4,5	50,9	13,6	48,0
Professions intermédiaires	12,0	44,8	8,0	49,8	20,0	46,8
Agriculteurs exploitants	3,4	45,3	2,8	50,1	6,2	47,5
Indépendants	5,7	44,4	2,8	50,1	8,5	46,3
Employés	5,7	42,1	19,9	49,4	25,6	47,8
Ouvriers	19,3	40,9	6,8	48,6	26,1	42,9
Ensemble	55,2	43,4	44,8	49,6	100,0	46,2

2) Variance inter-groupe : variances des (espérances de vie) moyennes selon le sexe

$$Var(e_{35-S}) = f_H \times (e_{35_H} - e_{35})^2 + f_F \times (e_{35_F} - e_{35})^2$$

$$Var(e_{35-S}) = \frac{55,2}{100,0} \times (43,4 - 46,2)^2 + \frac{44,8}{100,0} \times (49,6 - 46,2)^2 = 9,51 \text{ ans}^2$$

3) Variance intra-groupe : moyenne des variances (des espérances de vie) selon la PCS

On calcule la variance selon la PCS au sein de chacun des sexes, puis on effectue la moyenne de ces deux variances pondérée par le poids de chaque sexe dans la population totale.

- Variance selon la PCS parmi les hommes :

$$Var(e_{35_PCS/H})$$

$$\begin{aligned}
 &= f_{CS/H} \times (e_{35_{CS,H}} - e_{35_H})^2 + f_{PI/H} \times (e_{35_{PI,H}} - e_{35_H})^2 \\
 &+ f_{AE/H} \times (e_{35_{AE,H}} - e_{35_H})^2 + f_{I/H} \times (e_{35_{I,H}} - e_{35_H})^2 \\
 &+ f_{E/H} \times (e_{35_{E,H}} - e_{35_H})^2 + f_{O/H} \times (e_{35_{O,H}} - e_{35_H})^2
 \end{aligned}$$

$Var(e_{35_PCS/H})$

$$= \frac{9,1}{55,2} \times (46,6 - 43,4)^2 + \frac{12,0}{55,2} \times (44,8 - 43,4)^2 + \frac{3,4}{55,2} \times (45,3 - 43,4)^2 \\ + \frac{5,7}{55,2} \times (44,4 - 43,4)^2 + \frac{5,7}{55,2} \times (42,1 - 43,4)^2 + \frac{19,3}{55,2} \times (40,9 - 43,4)^2 \\ = 4,80 \text{ ans}^2$$

- Variance selon la PCS parmi les femmes :

$Var(e_{35_PCS/F})$

$$= f_{CS/F} \times (e_{35_{CS,F}} - e_{35_F})^2 + f_{PI/F} \times (e_{35_{PI,F}} - e_{35_F})^2 \\ + f_{AE/F} \times (e_{35_{AE,F}} - e_{35_F})^2 + f_{I/F} \times (e_{35_{I,F}} - e_{35_F})^2 + f_{E/F} \times (e_{35_{E,F}} - e_{35_F})^2 \\ + f_{O/F} \times (e_{35_{O,F}} - e_{35_F})^2$$

$Var(e_{35_PCS/F})$

$$= \frac{4,5}{44,8} \times (50,9 - 49,6)^2 + \frac{8,0}{44,8} \times (49,8 - 49,6)^2 + \frac{2,8}{44,8} \times (50,1 - 49,6)^2 \\ + \frac{2,8}{44,8} \times (50,1 - 49,6)^2 + \frac{19,9}{44,8} \times (49,4 - 49,6)^2 + \frac{6,8}{44,8} \times (48,6 - 49,6)^2 \\ = 0,38 \text{ ans}^2$$

- Moyenne des deux variances selon la PCS :

$$Var(e_{35_PCS}) = f_H \times Var(e_{35_PCS/H}) + f_F \times Var(e_{35_PCS/F})$$

$$Var(e_{35_PCS}) = \frac{55,2}{100,0} \times 4,80 + \frac{44,8}{100,0} \times 0,38 = 2,82 \text{ ans}^2$$

4) Variance totale de l'espérance de vie à 35 ans : somme de la variance de l'espérance de vie à 35 ans selon le sexe et de la variance de l'espérance de vie à 35 ans selon la PCS

La variance totale est la somme de la variance inter-groupe et de la variance intra-groupe.

$$Var(e_{35}) = Var(e_{35_S}) + Var(e_{35_PCS})$$

$$Var(e_{35}) = 9,51 + 2,82 = 12,33 \text{ ans}^2$$

5) Contributions respectives du sexe et de la PCS à la dispersion des espérances de vie à 35 ans

$$\text{Contribution du sexe : } \frac{Var(e_{35_S})}{Var(e_{35})} = \frac{9,51}{12,33} = 77 \%$$

$$\text{Contribution de la PCS : } \frac{Var(e_{35_PCS})}{Var(e_{35})} = \frac{2,82}{12,33} = 23 \%$$

Commentaire :

L'espérance de vie à 35 ans des actifs français en 2003 est de 49,6 ans. L'écart entre hommes et femmes est important : à 35 ans, les actives peuvent espérer vivre en moyenne 6,2 ans de plus que les actifs (49,6 ans contre 43,4 ans). Il existe aussi des inégalités selon la catégorie professionnelle, surtout chez les hommes. Ainsi, parmi ces derniers, à partir de 35 ans, les cadres supérieurs peuvent espérer vivre 5,7 ans de plus que les ouvriers (46,6 ans contre 40,9 ans). L'écart entre ces deux catégories n'est que de 2,3 ans parmi les femmes (50,9 ans contre 48,6 ans). Les écarts entre sexe sont plus importants que ceux entre catégorie

professionnelle. Ainsi, l'espérance de vie à 35 ans la plus élevée chez les hommes (celle des cadres) est moins importante que la plus faible des espérances de vie des femmes : à 35 ans, les cadres supérieurs masculins peuvent espérer vivre deux ans de moins que les femmes occupant une fonction d'ouvrier.

La décomposition de la variance précise ces observations : les trois quarts de la variance des espérances de vie selon le sexe et la catégorie professionnelle s'expliquent par les écarts entre sexe et un quart seulement par les inégalités selon la catégorie professionnelle.