I EXERCICE-1

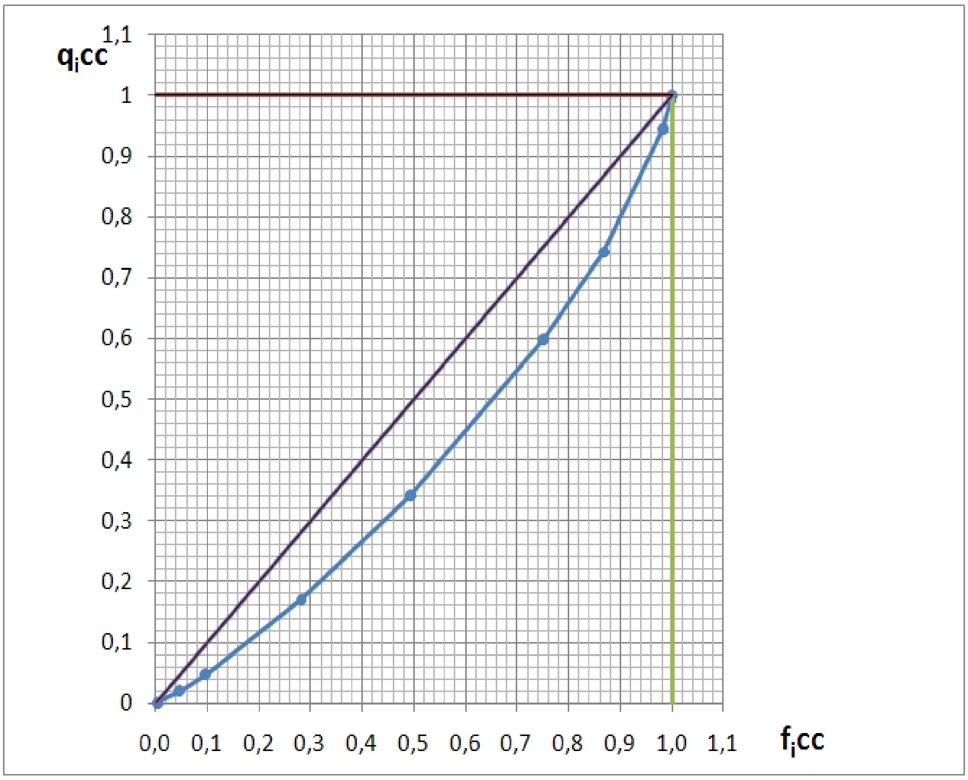
1.

a_i	b_i	x_i	n_i	n_i*x_i	f_i	f_icc	$n_i*x_i^2$	q_i	q_icc	Ai	di	S_i
0	50	25	9246	231150	0.0036	0.0036	5778750	0.0007	0.0007	50	184.92	0.0000
50	70	60	111107	6666420	0.0431	0.0467	399985200	0.0191	0.0198	20	5555.35	0.0004
70	80	75	129457	9709275	0.0502	0.0969	728195625	0.0278	0.0476	10	12945.70	0.0017
80	100	90	476716	42904440	0.1849	0.2817	3861399600	0.1229	0.1705	20	23835.80	0.0202
100	120	110	543670	59803700	0.2108	0.4926	6578407000	0.1713	0.3418	20	27183.50	0.0540
120	150	135	666398	89963730	0.2584	0.7510	12145103550	0.2577	0.5994	30	22213.27	0.1216
150	180	165	301428	49735620	0.1169	0.8679	8206377300	0.1425	0.7419	30	10047.60	0.0784
180	300	240	294687	70724880	0.1143	0.9821	16973971200	0.2026	0.9445	120	2455.73	0.0964
300	500	400	42172	16868800	0.0164	0.9985	6747520000	0.0483	0.9928	200	210.86	0.0158
500	800	650	3871	2516150	0.0015	1.0000	1635497500	0.0072	1.0000	300	12.90	0.0015
			2578752	349124165	1		57282235725					0.3900

La moyenne est donnée par : $\overline{x} = \dfrac{\sum n_i x_i}{N} = \dfrac{349124165}{2578752} \simeq \boxed{135.38}$

L'écart-type est : $\sigma(X) = \sqrt{\dfrac{1}{N}\sum n_i x_i^2 - \overline{x}^2} = \sqrt{\dfrac{57283235725}{2578752} - 135.38^2} = \boxed{62.34}$

2. cf tableau.

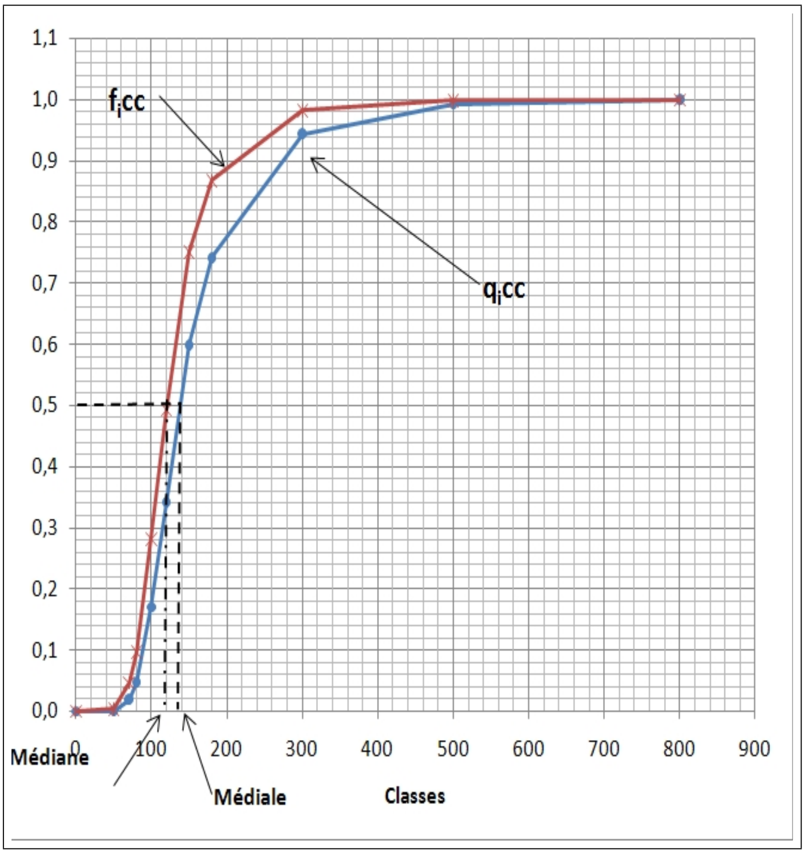


On calcule la surface de concentration en calculant l'aire des trapèzes situés sous la courbe de Lorentz et on trouve une aire de 0.39 pour la somme des trapèzes, donc : $0.5 - 0.39 = 0.11$, pour l'aire de concentration, donc unindice de Gini qui vaut : $I = 2 * 0.11 = 0.22$. On a donc un indice plus proche de 0 que de 1 ; la concentration est donc assez faible.

3. On calcule la médiale Ml comme on calcule une médiane ; on localise la médiale avec les valeurs relatives cumulées croissantes, puis on opère une interpolation linéaire ; on se retrouve dans la classe $[120; 150[$ et on obtient : $\dfrac{59.94 - 34.18}{150 - 120} = \dfrac{50 - 34.18}{Ml - 120}$ soit

$Ml = 120 + 30\dfrac{50 - 34.18}{59.94 - 34.18} = \boxed{138.42}$

On peut l'obtenir graphiquement à l'aide du polygone des valeurs relatives cumulées croissantes que l'on coupe par la droite $y = 0.50$.



Les agents de l'état qui gagnent moins de 138.37kF par an se partagent la moitié de la masse salariale, donc autant d'argent que ceux qui gagnent plus que $138.42F$.

4. Il est clair que la médiane est inférieure strictement à la médiale car les individus qui ont un salaire inférieur ou égal à la médiane constituent 50% des agents de l'état et leur masse salariale est nécessairement inférieure à 50% de la masse salariale, puisqu'elle est inférieure à celle des salariés gagnant plus que la médiane.

II EXERCICE-1

	x_i	y_i	x_i^2	y_i^2	x_iy_i
1	25	62,9	625	3956,41	1572,5
2	44	6,2	1936	38,44	272,8
3	33	34	1089	1156	1122
4	74	23,5	5476	552,25	1739
5	46	42,9	2116	1840,41	1973,4
6	47	51,2	2209	2621,44	2406,4
7	30	63,8	900	4070,44	1914
8	39	48,9	1521	2391,21	1907,1
9	18	71,1	324	5055,21	1279,8
10	21	67,2	441	4515,84	1411,2
11	17	55,3	289	3058,09	940,1
12	50	29,7	2500	882,09	1485
	444	556,7	19426	30137,83	18023,3

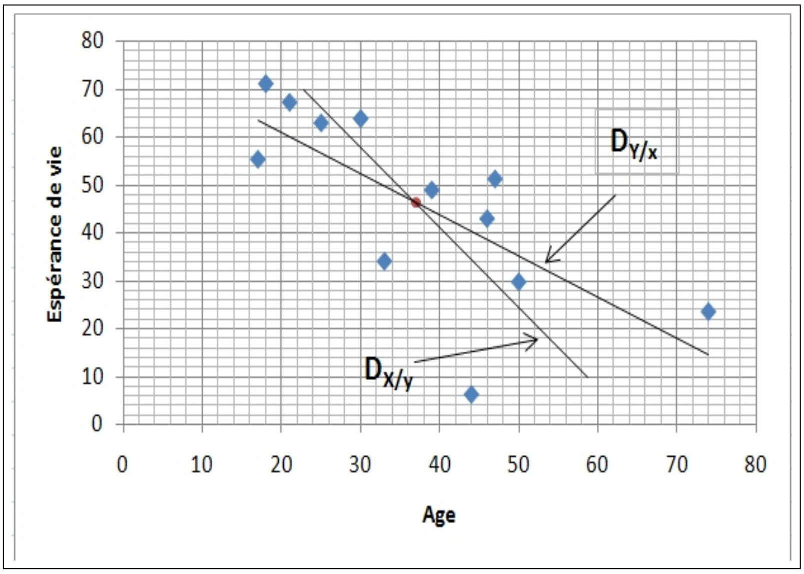
	x	y
Moyenne	37,0000	46,3917
V	249,8333	359,2991
σ	15,8061	18,9552

1. Le point moyen est le point G de coordonnées $\overline{x}$ et $\overline{y}$, avec d'après le tableau : $\begin{cases} \overline{x} = 37 \\ \overline{y} = 46.39 \end{cases}$

CORRIGE DEVOIR DE NÖEL

2. $\begin{cases} D_{y/x} : y = \widehat{a}x + \widehat{b} \\ \overline{y} = \widehat{a}\overline{x} + \widehat{b} \\ Cov\,(x;y) = \dfrac{18023.3}{12} - 37 * 46.39 \simeq -214.55 \\ \widehat{a} = \dfrac{Cov\,(x;y)}{V\,(x)} = \dfrac{-214.55}{249.8333} \simeq -0.858\,8 \end{cases}$: ce qui donne : $\widehat{b} = \overline{y} - a\overline{x} = 46.39 - (-0.8588)\,(37) = 78.1656$

   soit : $\boxed{y = -0.858\,8x + 78.1656}$



3. $r = \dfrac{Cov\,(x;y)}{\sigma\,(x)\,\sigma\,(y)} = \dfrac{-214.488\,3}{15.8061 * 18.9552} = \boxed{-0.715\,9}$

   On a un coefficient de corrélation proche de -1, mais pas très fort donc on peut conjecturer un lien linéaire moyen entre les deux variables, la corrélation étant négative, ce qui traduit que x et y varient en sens contraires.

4. Il est clair que la droite de régression de X en y n'a pas grand sens ici, X étant la variable explicative et Y la variable expliquée, mais on peut toujours rechercher son équation.

   $\begin{cases} D_{x/y} : x = \widehat{a'}y + \widehat{b'} \\ \overline{x} = \widehat{a'}\overline{y} + \widehat{b'} \\ \widehat{a'} = \dfrac{Cov\,(x;y)}{V\,(y)} = \dfrac{-214.55}{359.30} \simeq -0.597\,1 \end{cases}$: ce qui donne : $\widehat{b'} = \overline{x} - \widehat{a'}\overline{y} = 37 - (-0.597\,1)\,(46.39) = 64.6995$

   soit : $\boxed{x = -0.5971y + 64.6995}$, soit $y = -\dfrac{1}{0.5971}x + \dfrac{64.6995}{0.5971} = -1.6748x + 108.356\,2$

5. Prévision : $\widehat{y}\,(60) = -0.858\,8 * 60 + 78.1656 \simeq 26.64$ ans