1.

a_i	b_i	A_i	n_i	x_i	n_ix_i	$n_ix_i^2$	n_icc	d_i
0	1000	1000	129	500	64500	32250000	129	0.129
1000	2000	1000	165	1500	247500	371250000	294	0.165
2000	3000	1000	408	2500	1020000	2550000000	702	0.408
3000	4000	1000	108	3500	378000	1323000000	810	0.108
4000	5000	1000	56	4500	252000	1134000000	866	0.056
5000	10000	5000	90	7500	675000	5062500000	956	0.018
10000	50000	40000	44	30000	1320000	39600000000	1000	0.0011
			1000		3957000	50073000000		

a. Le tableau statistique permet de calculer $\overline{x} = \frac{1}{n}\sum n_i x_i = \frac{3957000}{1000} = 3957$

b. La classe modale est celle de plus grande densité, c'est-à dire la classe $[2000; 3000[$ et le mode est calculé en considérant les classes encadrant la classe modale, ce qui donne avec les notations du cours : $\left\{\begin{array}{l} x_1 = 2000 \\ x_2 = 3000 \end{array}\right.$, $\left\{\begin{array}{l} h = 0.408 \\ h_1 = 0.165 \text{ et } h_2 = 0.108 \end{array}\right.$

$\left\{\begin{array}{l} k_1 = h - h_1 = 0.408 - 0.165 \simeq 0.243 \\ k_2 = h - h_2 = 0.408 - 0.108 = 0.300 \end{array}\right.$: et pour conclure :

$M_o = \frac{k_2 x_1 + k_1 x_2}{k_2 + k_1} = \frac{0.300 * 2000 + 0.243 * 3000}{0.543} = \boxed{2447.5}$; comme prévu, le mode est plus petit que 2500, le centre de la classe modale, car il est attiré par la classe de gauche, de densité plus importante.

c. Coût médian : on localise la médiane grâce aux effectifs cumulés croissants et on effectue une interpolation linéaire ; la classe médiane est la classe $[2000; 3000[$, et on obtient : $\frac{702-294}{3000-2000} = \frac{500-294}{M_e - 2000}$, soit $M_e = \frac{(500-294)*1000}{702-294} + 2000 = \boxed{2504.9}$

d. $Q_1 \simeq 1733.33$ et $Q_3 \simeq 3744.44$. Celà signifie que 25% des sinistres ont un coût inférieur ou égal à 1733.33€ et 75% un coût inférieur ou égal à 3744.44€.

e. Le coefficient de Yule est donné par : $C_Y = \frac{Q_1 + Q_3 - 2M_e}{Q_3 - Q_1} = \frac{3744.44+1733.33-2*2504.9}{3744.44-1733.33} = \boxed{0.232\,7}$; ce coefficient est positif et traduit un étalement de la série à droite.

f. $V(x) = \frac{50073000000}{1000} - 3957^2 = \boxed{3.441\,5 \times 10^7}$ et l'écart-type, $\sigma(x) = \sqrt{3.441\,5 \times 10^7} = \boxed{5866.4}$

g. $\overline{x} - 0.5\sigma(x) = 3957 - 0.5 * 5866.4 = 1023.8$ et $\overline{x} + 0.5\sigma(x) = 3957 + 0.5 * 5866.4 = 6890.2$; on doit donc ajouter les effectifs des classes de 2000 à 5000, puis estimer les effectifs des intervalles $[1023.8; 2000[$ et $[5000\,; 6890.2[$, ce qui se fait en utilisant la densité :

Intervalle	Amplitude	Densité	Effectif
$[1023.8\,; 2000[$	976.2	0.165	$0.165 * 976.2 = 161.07$
$[5000\,; 6890.2[$	1890.2	0.018	$0.018 * 1890.2 = 34.02$
$[2000; 5000\,[$			$408 + 108 + 56 = 572$

, ce qui donne un total de : $572 + 34.02 + 161.07 = 767.09$, soit une proportion de : $\frac{767.09}{1000} = 0.7671$ soit $\boxed{76.71\%}$.

2. On suppose que la série est appoximativement distribuée suivant une loi normale de moyenne 3950 € et d'écart-type 5850 €.

a. On calcule cette probabilité en standardisant la variable aléatoire, c'est-à dire en se ramenant à la loi normale centrée réduite. On pose $Z = \frac{X-m}{\sigma} = \frac{X-3950}{5850}$, où m et σ désignent respectivement la moyenne et l'écart-type de X, alors Z suit la loi $\mathcal{N}(0; 1)$. On obtient : $P(3950 - 0.5 * 5850 < X < 3950 + 0.5 * 5850) = P(-0.5 \leq Z \leq 0.5) = 2F(0.5) - 1$, où F est la fonction de répartition de la loi normale centrée réduite, dont les valeurs sont lues dans la table ; on lit $F(0.5) \simeq 0.6915$, ce qui donne : $2 * 0.6915 - 1 = \boxed{0.383}$, ce qui montre que l'hypothèse de normalité ne peut être retenue.

b. La loi normale est symétrique, on aurait : $M_e = m = \boxed{3950}$, alors qu'ici la médiane est : 2504.9.

I EXERCICE-2

1.

	x	y
moy	35.4	342.8
var	203.31	27796.16
ecart-type	14.26	166.72

CORRIGE SUJET A

x	y	x*y
22	492	10824
50	186	9300
44	180	7920
32	384	12288
55	120	6600
60	120	7200
38	276	10488
22	480	10560
21	510	10710
45	252	11340
52	126	6552
33	360	11880
19	570	10830
17	588	9996
21	498	10458
		146946

On en déduit la covariance : $Cov\,(x;y) = \dfrac{1}{n}\sum x_i y_i - \overline{x} * \overline{y} = \dfrac{146946}{15} - 35.4 * 342.8 == \boxed{-2338.72}$

2. $r \simeq \boxed{-0.9838}$

3. $r = \dfrac{Cov\,(x;y)}{\sigma\,(x)\,\sigma\,(y)}$; ce coefficient est toujours compris entre -1 et 1 ; c'est un indicateur qui contribue à valider ou à refuser le modèle linéaire. Son signe, ici négatif traduit que les variables x et y varient en sens contraires ; par ailleurs, plus sa valeur absolue se rapproche de 1 (r proche de 1 ou de -1) plus le modèle linéaire est adapté aux données ; si r est proche de 0, le modèle linéaire est rejeté.

4. On trouve : $\widehat{y} = \widehat{a}x + \widehat{b} = \boxed{-11.5034x + 750.0207}$; avec $\widehat{a} = \dfrac{Cov(x\ ;\ y)}{V(x)}$ et $\widehat{b} = \overline{y} - \widehat{\mathbf{a}}\overline{x}$

5. L'équation de l'analyse de la variance est : $\mathbf{SCT} = \mathbf{SCE} + \mathbf{SCR}$

   $SCT = nV\,(y) = 15 * 27796.16 = 416942.4$ et $SCE = R^2 * SCT \simeq (-0.9838)^2 * 416942.4 \simeq 403542.89$ ou $V(\widehat{y}) = V(\widehat{a}x + \widehat{b}) = \widehat{a}^2 V(x) = (-11.5034)^2 * 203.31 = 26903.65$ et $SCE = nV(\widehat{y}) = 15 * 26903.65 = 403554.75$

   On trouve $SCR = SCT - SCE = 416942.4 - 403542.89 = \boxed{13399.51}$

   Méthode TGV : $SCR = SCT - SCE = SCT - R^2 * SCT = \left(1 - R^2\right) SCT = \left(1 - (-0.9838)^2\right) * 416942.4 \simeq 13399.51$

6. $R^2 = \dfrac{SCE}{SCT} = r^2 \simeq (-0.9838)^2 = \boxed{0.9679}$; ce coefficient donne la part de la variation totale expliquée par le modèle, ici 96.79%.

   vérification : $\dfrac{SCE}{SCT} = \dfrac{403542.89}{416942.4} \simeq 0.9679$

7. $\widehat{y}\,(25) = -11.5034 * 25 + 750.0207 \simeq \boxed{462.44}$

II EXERCICE-3 (2.5 pts)

Age (X)	F	H	n_{i+}	f_{i+}	x_i	$f_{i+}xi$
[20 ; 25[	60	55	115	0.1211	22.5	2.72
[25 ; 35[	32	275	307	0.3232	30	9.69
[35 ; 40[	120	260	380	0.4000	37.5	15.00
[40 ; 55[	23	60	83	0.0874	47.5	4.15
[55 ; 60[	20	45	65	0.0684	57.5	3.93
	255	695	950	1		35.50

1. Les fréquences relatives marginales, f_{i+}, sont calculés dans le tableau.
2. La fréquence des hommes de plus de 55 ans est une fréquence partielle, elle est donnée par : $f_{5,2} = \dfrac{45}{950} = \boxed{4.74 \times 10^{-2}}$.
3. la fréquence des $[25; 35[$ parmi les femmes est une fréquence conditionnelle : $f_{i=2/j=1} = \dfrac{32}{255} = \boxed{0.1255}$.
4. L'âge moyen est calculé dans le tableau : $\overline{x} = \sum f_{i+}x_i = \boxed{35.5}$ ans.