I

a_i	b_i	A_i	n_i	x_i	n_ix_i	$n_ix_i^2$	n_icc	d_i
0	1500	1500	102	750	76500	57375000	102	0.068
1500	2000	500	250	1750	437500	765625000	352	0.5
2000	3000	1000	350	2500	875000	2187500000	702	0.35
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	0.001
			1000		4014000	50130000000		

1. a. Le tableau statistique permet de calculer $\overline{x} = \dfrac{1}{n}\sum n_ix_i = \dfrac{4014000}{1000} = 4014.$

   b. La classe modale est celle de plus grande densité, c'est-à dire la classe $[3000; 4000[$ et le mode est calculé en considérant les classes encadrant la classe modale, ce qui donne avec les notations du cours : $\left\{\begin{array}{l} x_1 = 1500 \\ x_2 = 2000 \end{array}\right.$, $\left\{\begin{array}{l} h = 0.50 \\ h_1 = 0.068 \text{ et } h_2 = 0.35 \end{array}\right.$

   $\left\{\begin{array}{l} k_1 = h - h_1 = 0.50 - 0.068 \simeq 0.432 \\ k_2 = h - h_2 = 0.50 - 0.35 = 0.15 \end{array}\right.$: et pour conclure :

   $M_o = \dfrac{k_2x_1 + k_1x_2}{k_2 + k_1} = \dfrac{0.15 * 1500 + 0.432 * 2000}{0.432 + 0.15} = \boxed{1871.13}$; comme prévu, le mode est plus grand que le centre de la classe modale, car il est attiré par la classe de droite, de densité plus importante.

   c. Q_1: on le localise grâce aux effectifs cumulés croissants et on effectue une interpolation linéaire ; la classe médiane est la classe $[1500; 2000[$, et on obtient : $\frac{352-102}{2000-1500} = \frac{250-102}{Q-1500}$, soit $Q_1 = \frac{(250-102)*500}{352-102} + 1500 == \boxed{1796.0}$

   d. $Q_2 \simeq 2422.86$ et $Q_3 \simeq 3444.44$. Celà signifie que 50% des sinistres ont un coût inférieur ou égal à 2422.86€ et 75% un coût inférieur ou égal à 3444.44€.

   e. Le coefficient de Yule est donné par : $C_Y = \dfrac{Q_1 + Q_3 - 2M_e}{Q_3 - Q_1} = \frac{1796+3444.44-2*2422.86}{3444.44-1796} = \boxed{0.2395}$; ce coefficient est positif et traduit un étalement de la série à droite.

   f. $V(x) = \frac{50130000000}{1000} - 4014^2 = \boxed{34017804}$ et l'écart-type, $\sigma(x) = \sqrt{34017804} = \boxed{5832.48}$

   g. $\overline{x} - 0.75\sigma(x) = 4014 - 0.75 * 5832.48 = -360.36$ et $\overline{x} + 0.75\sigma(x) = 4014 + 0.75 * 5832.48 = 8388.36$; on doit donc ajouter les effectifs des classes de 0 à 5000, puis estimer les effectifs des intervalles $[5000; 8388.36[$, ce qui se fait en utilisant la densité :

Intervalle	Amplitude	Densité	Effectif
$[5000; 8388.36[$	3388.36	0.018	$3388.36 * 0.018 = 60.99$
$[0; 5000[$			866

   , ce qui donne un total de : $866 + 61 = 927.$, soit une proportion de : 0.927 soit $\boxed{92.7\%}$.

2. On suppose que la série est appoximativement distribuée suivant une loi normale de moyenne 4015 € et d'écart-type 5830 €.

   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 $Z = \frac{X-m}{\sigma} = \frac{X-4015}{5830}$, 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 : $P(4015 - 0.75 * 5830 < X < 4015 + 0.75 * 5830) = P(-0.75 \leq Z \leq 0.75) = 2F(0.75) - 1$, où F est la fonction de répartition de la loi normale centrée réduite, dont les valeurs sont lues dans la table ; on lit $F(0.75) \simeq 0.7734$, ce qui donne : $2 * 0.7734 - 1 = \boxed{0.546\,8}$, ce qui montre que l'hypothèse de normalité ne peut être retenue.

   b. La loi normale est symétrique, on aurait : $M_e = m = \boxed{4014}$, alors qu'ici la médiane est : 2422.86.

II EXERCICE-2

1.

	x	y
moy	35.53	339.13
v	201.32	27800.92
ecart-type	14.19	166.74

x	y	xy
23	482	11086
50	186	9300
44	180	7920
32	384	12288
55	120	6600
60	110	6600
38	256	9728
22	470	10340
21	510	10710
45	252	11340
52	126	6552
34	360	12240
19	568	10792
17	588	9996
21	495	10395
		145887

On en déduit la covariance : $Cov\left(x;y\right) = \dfrac{1}{n}\sum x_i y_i - \overline{x} * \overline{y} = \dfrac{145887}{15} - 35.53 * 339.13 = \boxed{-2323.49}$

2. $r \simeq \boxed{-0.9826}$

3. $r = \dfrac{Cov\left(x;y\right)}{\sigma\left(x\right)\sigma\left(y\right)}$; 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 : $\widehat{y} = \widehat{a}x + \widehat{b} = \boxed{-11.5477x + 749.4627}$; 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 : $\mathbf{SCT} = \mathbf{SCE} + \mathbf{SCR}$

   $SCT = nV\left(y\right) = 15 * 27800.92 = 417013.8$ et $SCE = R^2 * SCT \simeq \left(-0.9826\right)^2 * 417013.8 \simeq 402627.97$ ou $V(\widehat{y}) = V(\widehat{a}x + \widehat{b}) = \widehat{a}^2 V(x) = \left(-11.5477\right)^2 * 201.32 = 26845.90$ et $SCE = nV(\widehat{y}) = 15 * 26845.90 = 402688.5$

   On trouve $SCR = SCT - SCE = 417013.8 - 402627.97 = \boxed{14385.83}$

   Méthode TGV : $SCR = SCT - SCE = SCT - R^2 * SCT = \left(1 - R^2\right) SCT = \left(1 - \left(-0.9826\right)^2\right) * 417013.8 \simeq 14385.83$

6. $R^2 = \dfrac{SCE}{SCT} = r^2 \simeq \left(-0.9826\right)^2 = \boxed{0.9655}$; ce coefficient donne la part de la variation totale expliquée par le modèle, ici 96.55%.

   vérification : $\dfrac{SCE}{SCT} = \dfrac{402688.5}{417013.8} \simeq 0.9657$

7. $\widehat{y}\left(25\right) = -11.5477 * 25 + 749.4627 \simeq \boxed{460.77}$

III EXERCICE-3 (2.5 pts)

SEXE	F	H	Ensemble			
Age	n_{i1}	n_{i2}	n_{i+}	f_{i+}	x_i	$f_i x_i$
[20;25[	40	55	95	0.1044	23	2.35
[25;35[	32	245	277	0.3044	30	9.13
[35;40[	120	260	380	0.4176	38	15.66
[40;55[	33	60	93	0.1022	48	4.85
[55;60[	23	42	65	0.0714	58	4.11
n_{+j}	248	662	910	1		36.10
f_{+j}	0.2725	0.7275				

CORRIGE PARTIEL B

1. Les fréquences relatives marginales, f_{+j}, sont calculés dans le tableau.
2. La fréquence des femmes de moins de 25 ans est une fréquence partielle, elle est donnée par : $f_{5,2} = \frac{40}{910} \simeq \boxed{4.40 \times 10^{-2}}$
3. la fréquence des $[35; 40[$ parmi les hommess est une fréquence conditionnelle : $f_{i=3/j=2} = \frac{260}{662}. = 0.392\,7$
4. L'âge moyen est calculé dans le tableau : $\overline{x} = \sum f_{i+}x_i = \boxed{36.10}$ ans.