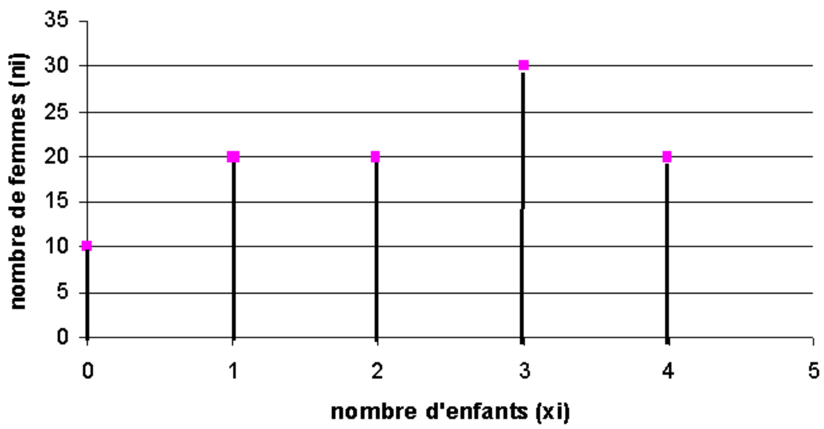
Master2 Informatique : Intelligence Artificielle (IA)
Module : ***Fouille de données (FD)*** Année universitaire :

SERIE TD N°3 (corrigé)

Exercice ①

1. population : 100 femmes de 40 ans ; unité statistique : une femme ; caractère : le nombre d'enfants (caractère quantitatif discret) ; modalités : au nombre de 5 (0,1,2,3,4).
2. La distribution statistique étant discrète, le diagramme différentiel est un diagramme en bâtons. Dans ce diagramme, on porte en abscisse les différentes modalités du caractère c'est à dire les différentes valeurs prises par la variable (0,1,2,3,4) ; en ordonnée seront indiqués soit les effectifs soit les fréquences relatives afférentes à chaque modalité.

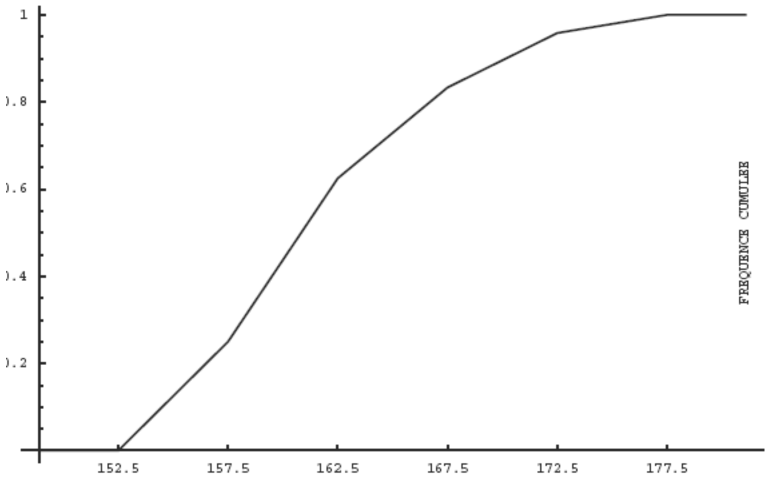
x_i (Nombre d'enfants)	n_i (effectif)	f_i (fréquence relative)
0	10	0,1
1	20	0,2
2	20	0,2
3	30	0,3
4	20	0,2
Total	100	1



Exercice ②

La variable statistique est continue.

Bornes des classes b_i	Effectifs cumulés N_i	Fréquence cumulée F_i
152.5	0	0
157.5	6	$\frac{1}{4}$
162.5	15	$\frac{5}{8}$
167.5	20	$\frac{5}{6}$
172.5	23	$\frac{23}{24}$
177.5	24	1



$$m = \frac{1}{24}\ (155 \times 6 + 160 \times 9 + 165 \times 5 + 170 \times 3 + 175) = \frac{485}{3} \simeq 161.667$$

$$\frac{\frac{15}{24} - \frac{6}{24}}{162.5 - 157.5} = \frac{\frac{1}{2} - \frac{6}{24}}{M_e - 157.5}$$

$$\frac{15 - 6}{162.5 - 157.5} = \frac{12 - 6}{M_e - 157.5}$$

$$\frac{9}{5} = \frac{6}{M_e - 157.5}$$

$$M_e = 157.5 + 5\ \ \frac{6}{9} \simeq 160.833$$

$$s = \sqrt{\left(\frac{1}{24}\ \left((155 - 161.667)^2\, 6 + (160 - 161.667)^2\, 9 + (165 - 161.667)^2\, 5 + (170 - 161.667)^2\, 3 + (175 - 161.667)^2\, 1\right)\right)} \simeq 5.52771$$

Exercice ③

1. Identifier Population, individu, caractères et modalités

Population Les 12 mois de l'année

Individu Un mois parmi ces mois

Caractère Il y a deux caractères : Pluviométrie (mesurée en mm et noté P) et la Température (mesurée en °C et noté T)

Modalités Pour le caractère Pluviométrie : Les modalités varient de 5 à 79 mm

Pour le caractère Température : Les modalités varient de 10 à 28°C

2. Calculer la moyenne et l'écart type de P et T

Les moyennes et écart types de P et T vont être directement calculés à partir de la série statistique

Individu	P (mm) (1)	T (°C) (2)	$(P_i - \overline{P})^2$ (3)	$(T_i - \overline{T})^2$ (4)
1	13	23	544,4444444	32,11111111
2	23	17	177,7777778	0,111111111
3	49	14	160,4444444	11,11111111
4	49	10	160,4444444	53,77777778
5	50	10	186,7777778	53,77777778
6	64	11	765,4444444	40,11111111
7	79	13	1820,444444	18,77777778
8	48	15	136,1111111	5,444444444
9	40	17	13,44444444	0,111111111
10	10	23	693,4444444	32,11111111
11	5	27	981,7777778	93,44444444
12	6	28	920,1111111	113,7777778
	$\overline{P}$ =36,33 mm	$\overline{T}$ =17,33 °C	546,7222222 (mm)²	37,88888889 (°C)²
			σ_P = 23,38209191 mm	σ_T = 6,155395104 °C

3.

Individu	P (mm) (1)	T (°C) (2)	$(P_i - \overline{P})(T_i - \overline{T})$ (3)	$P_i T_i$ (4)
1	13	23	-132,22	299
2	23	17	4,44	391
3	49	14	-42,22	686
4	49	10	-92,89	490
5	50	10	-100,22	500
6	64	11	-175,22	704
7	79	13	-184,89	1027
8	48	15	-27,22	720
9	40	17	-1,22	680
10	10	23	-149,22	230
11	5	27	-302,89	135
12	6	28	-323,56	168
	$\overline{P}$ =36,33 mm	$\overline{T}$ =17,33 °C	$\sum_{i=1}^{n}(P_i - \overline{P})(T_i - \overline{T})$=-1527,33	$\sum_{i=1}^{n} P_i.T_i$ =60305
			$\mathrm{cov}(P,T)$ = -1527,33/12 $\mathrm{cov}(P,T)$ = -127,28 mm.°C	$\mathrm{cov}(P,T)$ = (1/12)*6030-17.33*36.33 $\mathrm{cov}(P,T)$ = -127,28 mm.°C

La covariance $\mathrm{cov}(P,T)$ peut être calculée soit par

L'Equation $\mathrm{cov}(P,T) = \frac{1}{n}\sum_{i=1}^{n}(P_i - \overline{P})(T_i - \overline{T})$ (Voir Colonne N°3 du tableau ci-dessus)

Ou bien par

L'Equation $\mathrm{cov}(P,T) = \frac{1}{n}\sum_{i=1}^{n} P_i.T_i - \overline{T}.\overline{P}$ (Voir Colonne N°4 du tableau ci-dessus)

$$T = \overline{T} - \frac{\mathrm{cov}(P,T)}{\sigma_P^2}\overline{P} + \frac{\mathrm{cov}(P,T)}{\sigma_P^2}P = 17.33 - \frac{(-127.28)}{546,73}36.33 + \frac{(-127.28)}{546,73}P$$

Soit : $T = 25.78 - 0.23P$

En particulier lorsque pour $P = \overline{P} = 36.33$ mm on trouve P=17.42°C qui est très proche de 17.33°C