Statistiques descriptives univariées

Exercice 1. Nombre inverventions pour panne.

Nombre inverventions	0	1	2	3	4	5	6
Nombre de photocopieurs	2	10	25	40	20	15	8

1. *Population :* les photocopieurs à gros tirage du modèle en question de cette entreprise.
   Échantillon : 120 individus de cette population.
   Caractère : la fiabilité et plus précisément le nombre d'interventions pour panne.
   Nature du caractère : Quantitatif discret.
2. Tableau statistique complet :

Nombre inverventions	0	1	2	3	4	5	6
n_i effectifs	2	10	25	40	20	15	8
$f_i = n_i/N$ fréquence relative	$\frac{2}{120}$	$\frac{10}{120}$	$\frac{25}{120}$	$\frac{40}{120}$	$\frac{20}{120}$	$\frac{15}{120}$	$\frac{8}{120}$
fréquence relative (décimaux)	$0,167$	$0,833$	$0,208$	$0,333$	$0,167$	$0,125$	$0,067$
N_i Effectifs cumulés	2	12	37	77	97	112	120
$F_i = N_i/N$ Effectifs cumulés relatifs	$\frac{2}{120}$	$\frac{12}{120}$	$\frac{37}{120}$	$\frac{77}{120}$	$\frac{97}{120}$	$\frac{112}{120}$	$\frac{120}{120} = 1$

   Effectifs cumulés relatifs

3. La fonction de répartition est $P(X \leq t)$ c"est-à-dire que la valeur de cette fonction $F(t)$ sur un réel t est le pourcentage que le nombre d'intervention soit au plus t. On a donc :

$t =$ inverventions	0	1	2	3	4	5	6
$P(X \leq t)$	$\frac{2}{120} = 0,167$	$\frac{12}{120} = 0,1$	$\frac{37}{120} = 0,308$	$\frac{77}{120}$=0,642	$\frac{97}{120} = 0,808$	$\frac{112}{120} = 0,933$	$\frac{120}{120} = 1$

4. On calcule comme partie sur le total :
   Au moins 3 : c'est $f_3 + f_4 + f_5 + f_6 = 83/120 = 0,692$, ou alors $1 - F_2 = 1 - 37/120$.
   Au plus 5 : on trouve $F_5 = 112/120 = 0,933$.
   Entre 1 et 4 : c'est $f_1 + f_2 + f_3 + f_4 = 95/120$=0,792. Ou $F_4 - F_0 = (97 - 2)/120$.
5. Tendance centrale :
   Mode : c'est le plus fréquent i. e. $mo = 3$.
   Médiane : c'est $me = 3$ car $P(X \leq 3) = 0,642$ et $P(X \geq 3) = 1 - 0,308 = 0,692$.
   Moyenne : on a $\bar{X} = 1/120(10 + 2 \times 25 + \cdots + 6 \times 8) = 383/120 = 3,192$.
   Quartiles : on a $q_{25} = 2$ et $q_{75} = 4$.
   Le mode et la médiane coïncident, ce n'est pas très surprenant pour une variable discrète. La moyenne est légèrement supérieure, il y a une certaine concentration de valeurs très supérieures à la médiane.
6. Dispersion.
   Étendue : c'est la différence entre la plus grande et la plus petite modalité, $6 - 0 = 6$.
   Étendue interquartile : c'est la différence entre q_{75} et q_{25}, ici $4 - 2 = 2$.
   Variance : c'est $V[X] = 1/N(n_1(x_1 - \bar{X})^2 + \cdots + n_r(x_r - \bar{X})^2)$, ici :
$$V[X] = \frac{1}{120}(2 \times (3,192 - 0)^2 + \cdots + 8 \times (3,192 - 6)^2) = 1,921.$$
   Écart-type : c'est $\sigma(X) = \sqrt{V[X]} = \sqrt{1,921} = 1,386$.
   Coefficient de variation : c'est $c_V = \sigma/\bar{X} = 1,386/3,192 = 0,434$.
   Écart absolu de variation : c'est $EAM = \frac{1}{120}(2 \times |3,192| + \cdots + 8 \times |3,192 - 6|) = 1,096$.
   Interprétation : on sait que l'écart-type est entre 0 (échantillon uniforme) et la moitié de l'étendue (échantillon on ne peut plus hétérogène). Ici on est à environs 1/6 de l'étendue, ce qui fait une hétérogénéité normale.

Exercice 2. Intervalles de la forme $[x_i, x_{i+1}[$. Résultats :

i	1	2	3	4	5	6	7	8
x_i	200	300	400	450	500	520	600	700
x_{i+1}	300	400	450	500	520	600	700	1000
n_i	4	7	9	7	6	10	5	4
a_i	100	100	50	50	20	80	100	300

1. *Population :* semaines de vente d'un produit chocolatier.
   Échantillon : les semaines d'une année.
   Caractère : la quantité de produit vendu, exprimée en Kilogrammes.
   Nature du caractère : Quantitatif continu.
2. Tableau statistique complet avec $A = 100$ et $N = 52$:

Classe i	1	2	3	4	5	6	7	8
$[x_i, x_{i+1}[$	$[200, 300[$	$[300, 400[$	$[400, 450[$	$[450, 500[$	$[500, 520[$	$[520, 600[$	$[600, 700[$	$[700, 1000[$
x_i	200	300	400	450	500	520	600	700
x_{i+1}	300	400	450	500	520	600	700	1000
n_i effectifs	4	7	9	7	6	10	5	4
a_i	100	100	50	50	20	80	100	300
a_i^r	1	1	0,5	0,5	0,2	0,8	1	3
$n_i^r = n_i/a_i^r$	4	7	18	14	30	12,5	5	1,33
$f_i = n_i/N$	0,076	0,134	0,173	0,134	0,115	0,192	0,096	0,077
$f_i^r = n_i/a_i^r$	0,076	0,134	0,346	0,268	0,575	0,24	0,096	0,026
N_i	4	11	20	27	33	43	48	52
F_i	0,076	0,211	0,384	0,519	$0,634$	0,827	0,923	1
c_i	250	350	425	475	510	560	650	850

   Histogramme : hauteur sur la classe C_i égale à la fréquence relative f_i^r.
3. Fonction de répartition. Nulle pour $x \leq x_1$, linéaire entre $(x_1, 0)$ et (x_2, F_1) puis entre (x_i, F_{i-1}) et (x_{i+1}, F_i).
4. On a $F(x) = P(X \leq x)$ fonction de répartition, représentée par le polygône des fréquences.
   Thalès : on interpole $(x, F(x))$ sur le segment reliant (x_i, F_{i-1}) à (x_{i+1}, F_i) donc :
$$\frac{F(x) - F_{i-1}}{F_i - F_{i-1}} = \frac{x - x_i}{x_{i+1} - x_i}.$$
   Calcul $F(x)$: on trouve
$$F(x) = \frac{F_i - F_{i-1}}{x_{i+1} - x_i}(x - x_i) + F_i = \frac{f_i}{a_i}(x - x_i) + F_{i-1}.$$
   (a) $P(X \leq 500) = P(X < 500)$ donc $f(500) = 0$;
   (b) $P(X < 450) = P(X \leq 450) = F(450) = 0,384$;
   (c) $P(X \leq 450) = F(450) = 0,384$;
   (d) $P(X \geq 520) = 1 - P(X < 520) = 1 - P(X \leq 520) = 1 - F(520) = 0,366$;
   (e) $P(475 \leq X \leq 680) = P(X \leq 680) - P(X \geq 475) = F(680) - (1 - F(475)) = F(680) + F(475) - 1$.
   On calcule les deux valeurs par interpolation linéaire. On voit que 680 appartient à C_7 et 475 à C_4
$$F(680) = \frac{0,096}{100}(680 - 600) + 0,923 = 0,999,$$
$$F(475) = \frac{0,134}{50}(475 - 450) + 0,519 = 0,586.$$
   Donc finalement :
$$P(475 \leq X \leq 680) = 0,586 + 0,999 - 1 = 0,585.$$

5. Tendence centrale.
Moyenne : on calcule $\bar{X} = 1/N \sum_{i=1}^{R} n_i c_i$ donc :
$$\bar{X} = \frac{1}{52}(4 \times 250 + \cdots + 4 \times 850) = 498,269.$$
Médiane : on voit qu'elle se trouve dans C_4. On interpole :
$$\frac{me - x_i}{x_{i+1} - x_i} = \frac{1/2 - F(x_i)}{F(x_{i+1}) - F(x_i)},$$
donc :
$$me = \frac{x_{i+1} - x_i}{F(x_{i+1}) - F(x_i)} \left(\frac{1}{2} - F(x_i) \right) + x_i = \frac{a_i}{f_i} \left(\frac{1}{2} - F(x_i) \right) + x_i$$
Ici :
$$me = \frac{a_4}{f_4} \left(\frac{1}{2} - F(x_4) \right) + x_4 = \frac{50}{0,134} (0,5 - 0,384) + 450 = 493,283$$
6. Dispersion.
Variance : on calcule $V[X] = 1/N \sum_{i=1}^{R} n_i (c_i - \bar{X})^2$ donc :
$$V[X] = \frac{1}{52}(4 \times (250 - 492,283)^2 + \cdots + 4 \times (850 - 492,283)^2) = 21217,455.$$
Écart-type : c'est $\sigma = \sqrt{V[X]}$ donc :
$$\sigma = \sqrt{21217,455} = 145,662.$$
Coefficient de variation : c'est $c = \sigma / \bar{X}$ donc :
$$c = \frac{145,662}{493,283} = 0,296.$$
Écart absolu moyen EAM : c'est $EAM = 1/N \sum_{i=1}^{R} n_i |c_i - \bar{X}|$, donc :
$$EAM = \frac{1}{52}(4 \times |250 - 492,283| + \cdots + 4 \times |850 - 492,283|) = 109,51.$$
7. La distribution n'est pas très homogène, avec deux modalités prédominantes, un mode à 6 Kilogrammes (10 individus) et un mode secondaire à 3 Kilogrammes (9 individus).

Exercice 3. 8 données en milliers d'Euros :
$$15, 16, 18, 19, 19, 21, 75, 95.$$
1. *Moyenne arithmétique*, exprimée en milliers d'euros :
$$\bar{X} = \frac{1}{8}(15 + 16 + 18 + 19 + 19 + 21 + 75 + 95) = 34,76.$$
Moyenne harmonique, exprimée en milliers d'euros,
$$\bar{X} = \frac{8}{\frac{1}{15} + \frac{1}{16} + \frac{1}{18} + \frac{1}{19} + \frac{1}{19} + \frac{1}{21} + \frac{1}{75} + \frac{1}{95}} = 22,132.$$
Moyenne géométrique, exprimée en milliers d'euros, $34,76$
$$\bar{X} = \sqrt[8]{15 \times 16 \times 18 \times 19 \times 19 \times 21 \times 75 \times 95} = 26,362.$$
Rappelons que $\sqrt[B]{A} = A^{\frac{1}{B}} = \exp(\frac{1}{B} \ln(A))$.
2. La moyenne harmonique, plus proche des valeurs autour de 19, semble tenir moins compte des grands écarts du centre représentés par 75 et 95.

Exercice 4. Notons q_i la quantité sur la période i et m_i le montant depensé sur la période i. Bien entendu, $p_i = m_i / q_i$. Notons Q quantité totale et M le montant depensé au total. On sait que le prix moyen au litre P est $P = M/Q$. Bien sûr $Q = q_1 + q_2 + q_3 + q_4$.

1. Si $q_i = q$ pour tout i alors $Q = 4q$. Ainsi :
$$P = \frac{M}{Q} = \frac{m_1 + m_2 + m_3 + m_4}{4q} = \frac{1}{4}\left(\frac{m_1}{q} + \frac{m_2}{q} + \frac{m_3}{q} + \frac{m_4}{q}\right) = \frac{1}{4}(p_1 + p_2 + p_3 + p_4).$$
C'est la moyenne arithmétique des p_i. Pour les valeurs numériques, on trouve $P = 0,375$.
2. Si $m_i = m$ pour tout i, on a $M = 4m$. De plus, $q_i = m/p_i$, donc :
$$P = \frac{M}{Q} = \frac{4m}{q_1 + q_2 + q_3 + q_4} = 4\frac{m}{\frac{m}{p_1} + \frac{m}{p_2} + \frac{m}{p_3} + \frac{m}{p_4}} = \frac{4}{\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} + \frac{1}{p_4}}.$$
C'est la moyenne harmonique.

Exercice 5. On sait que la vitesse v_i du trajet i-ième ($i = 1$ pour l'allée, $i = 2$ pour le retour) est :
$$v_i = \frac{d}{t_i},$$
où d est la distance (qui ne change au retoure par rapport à l'allée) et t_i est le temps écoulé sur le trajet i. Donc la vitesse moyenne est
$$v = \frac{2d}{t_1 + t_2}.$$
Nous connaissons v_i et non t_i donc nous écrivons $t_i = \frac{d}{v_i}$. Donc :
$$v = \frac{2d}{\frac{d}{v_2} + \frac{d}{v_2}} = \frac{2}{\frac{1}{v_2} + \frac{1}{v_2}}.$$
C'est bien la moyenne harmonique. On trouve $v = 144Km/h$.

Exercice 6. Le coefficient est $\alpha_i = 1 + \tau_i$ la variation sur l'année i. On sait que $p_{i+1} = \alpha_i p_i$. On a donc $p_4 = \alpha_3\alpha_2\alpha_1\alpha_0 p_0$. Pour la moyenne on imagine $\alpha_i = \alpha$ pour tout i. Donc $p_4 = \alpha^4 p_0$. Si le prix du produit triple en 4 ans, on impose $p_4 = 3p_0$, donc $3p_0 = \alpha^4 p_0$. De cette façon on trouve $\alpha^4 = 3$, i. e. $\alpha = \sqrt[4]{3} = 1,316$.