Analyse de données

Contrôle continu 21 octobre 2019

Temps disponible : 1 heure

Partie I

Le pin noir (*Pinus nigra*), très résistant à la sécheresse, fait l'objet d'une étude préalable à la valorisation massive de cet espèce dans un contexte de réchauffement climatique. Dans ce but, la résistance à un champignon et la taille des arbres on été étudiées. Les chercheurs ont examiné un petit nombre d'individus de Carinthie, en Autriche.

D'abord, la taille des arbres a été mesurée et confiée au tableau suivant (expression en mètres).

21	30	50	46	45	38	28	22
27	42	37	35	48	46	26	26
34	34	51	48	32	44	51	38

Ensuite, la résistance à une maladie des arbres a été prise en compte, pour les individus du même échantillon. Sphaeropsis sapinea est une espèce de champignons ascomycètes, responsables d'une maladie du pin noir, qui entraîne la perte de vigueur des pousses du pin.

Pour chacun des exemplaires, les chercheurs ont vérifié si l'arbre était atteint ou non de la maladie. Les résultats ont été confiés au tableau suivant (malade = M, sain = S ; les individus étant présentés dans le même ordre que pour la taille) :

$$\begin{array}{cccccccc} M & M & S & S & S & M & M & M \\ M & S & M & M & S & S & M & S \\ M & S & S & S & M & S & S & S \end{array}$$

1. Définir l'individu statistique, la population statistique à laquelle correspond cet échantillon et le type de données pour la résistance à Sphaeropsis sapinea et pour la taille.
2. Réaliser un diagramme représentatif de résistance à Sphaeropsis sapinea. Quel est le mode de ce caractère ?
3. Calculer la médiane de la taille des arbres examinés, puis des arbres sains.
4. Calculer la moyenne et l'écart-type de la taille des arbres sains, puis des arbres malades.
5. Dire si les arbres sains sont plus ou moins grands, puis si ils présentent une plus ou moins grande homogénéité de taille par rapport aux arbres malades.

Partie II

On souhaite comparer la taille des pins noirs autrichiens aux pins noirs de Turquie. Une population de ces derniers a été recensée. Les données on été distribuées en huit classes de 4m d'étendue, de 29m jusqu'à 61m. On a confié les résultats au tableau suivant :

Classe	1	2	3	4	5	6	7	8
Nombre individus	4	12	77	61	23	11	12	5

6. Donner le tableau statistique puis dresser un histogramme des données collectées.
7. Calculer la moyenne $\bar{X}$ et l'écart-type s de l'échantillon.
8. Fournir une représentation graphique des fréquences cumulées.
9. Calculer la médiane de l'échantillon et en fournir une représentation graphique.
10. Dire lesquels parmi les pins noirs d'Autriche et de Turquie sont plus grands. Lequel des deux échantillons présente une taille plus homogène ?

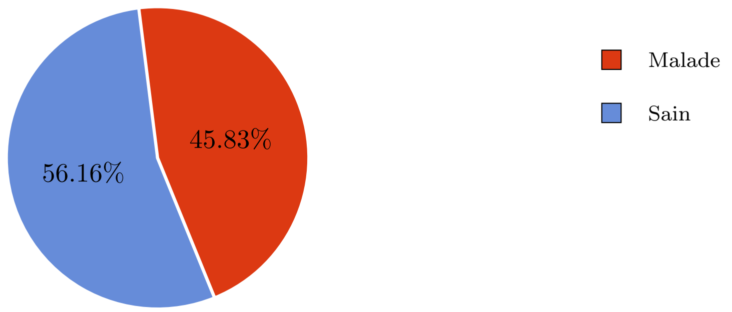
Analyse de données

Contrôle continu du 21 octobre 2019

Correction rapide

1. Individu : un pin noir de Carinthie. Population statistique : les pins noirs d'Autriche, ou de Carinthie. Type de données : était sanitaire de l'arbre (malade ou pas), donnée qualitative nominale ; puis taille de l'arbre, donnée quantitative continue.
2. Diagramme circulaire ou en bâtons. L'effectif M étant égal à 11, on a $P(M) = 45,83\%$, puis $P(S) = 56,16\%$, total 24 individus. Mode, S.

Diagramme circulaire représentatif de la résistence à la maladie



3. Étant donnée un nombre impair $2k+1$ de valeurs croissantes $x_1, \ldots, x_{2k+1}$, leur médiane est x_{k+1}. Si on a un nombre pair $2k$ de valeurs, la médiane est la moyenne entre x_k et x_{k+1}.
   On a 13 arbres sains, ainsi $13 = 2k + 1$ avec $k = 6$, dont voici les valeurs de la taille (en mètres) classées par ordre croissant :
   $$26, 34, 38, 42, 44, 45, 46, 46, 48, 48, 50, 51, 51.$$
   Donc la taille médiane des arbres sains est $x_{k+1} = x_7 = 46m$. Ainsi :
   $$me = 46m, \qquad \text{(abres sains)}.$$
   On a 24 arbres au total, donc $24 = 2k$ avec $k = 12$, dont les tailles classées sont :
   $$21, 22, 26, 26, 27, 28, 30, 32, 34, 34, 35, 37,$$
   $$38, 38, 42, 44, 45, 46, 46, 48, 48, 50, 51, 51.$$
   Ainsi la taille médiane des arbres est le centre entre $x_{12} = 37m$ et $x_{13} = 38m$, donc
   $$me = 37,5m, \qquad \text{(tous les abres)}.$$
4. La formule de la moyenne de N valeurs $x_1, \ldots, x_N$ est
   $$\bar{X} = (x_1 + \cdots + x_N)/N \quad .$$
   Pour les arbres sains on a $N = 13$ donc :
   $$\bar{X} = \frac{1}{13}(26 + \cdots + 51)m = 43,76m, \qquad \text{(abres sains)}.$$
   Pour les arbres malades on a $N = 11$. Les tailles (classées) des arbres malades sont (en mètres) :
   $$21, 22, 26, 27, 28, 30, 32, 34, 35, 37, 38.$$

Ainsi :

$$\bar{X} = \frac{1}{11}(21 + \cdots + 30)m = 30m, \qquad \text{(abres malades).}$$

La formule de l'écart type d'un échantillon est (données brutes) :

$$s = \sqrt{\frac{(x_1 - \bar{X})^2 + \cdots + (x_N - \bar{X})^2}{N - 1}}.$$

Ici, s est exprimé en mètres.
Écart-type des arbres sains :

$$s = \sqrt{\frac{(26 - 43,76)^2 + \cdots + (51 - 43,76)^2}{13 - 1}}m = 7,29m.$$

Écart-type des arbres malades :

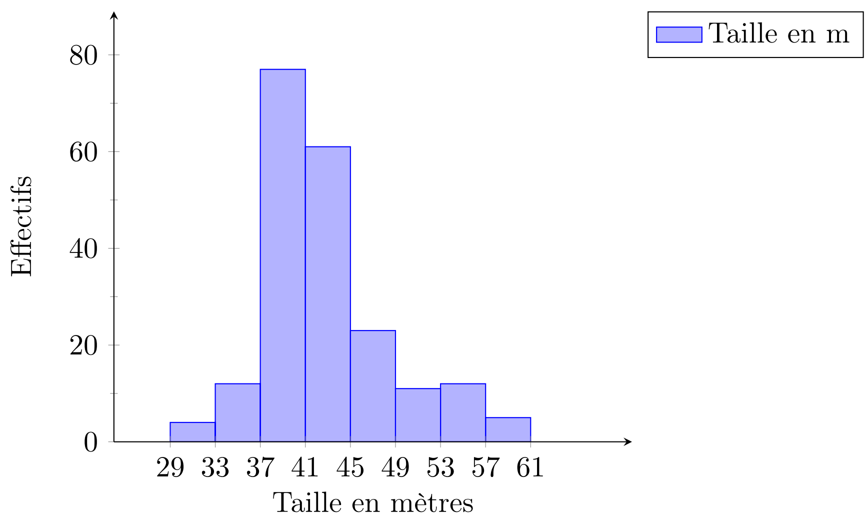
$$s = \sqrt{\frac{(21 - 30)^2 + \cdots + (38 - 30)^2}{11 - 1}}m = 5,76m.$$

5. Comme la moyenne de la taille pour les arbres sains vaut $46,76m$ et celle des arbres malades vaut $30m$, on a que en moyenne les abres sains sont plus grands que les arbres malades.
   Pour la variabilité, il faut utiliser le coefficient de variation $c = s/\bar{X}$. Pour les sains, il vaut $0,16$; pour les malades, il vaut $0,18$, donc l'échantillon d'arbres sains est plus homogène.
6. Voici le tableau.

Classe	1	2	3	4	5	6	7	8
Borne inférieure	29	33	37	41	45	49	53	57
Borne supérieure	33	37	41	45	59	53	57	61
Centre de classe	31	35	39	43	47	51	55	59
Effectifs	4	12	77	61	23	11	12	5
Fréquence	0,020	0,059	0,376	0,298	0,112	0,054	0,059	0,024
Effectifs cumulés	4	16	93	154	177	188	200	205
Fréquences cumulées	0,02	0,08	0,45	0,75	0,86	0,92	0,98	1,00

Voici l'histogramme.

Histogramme représentatif de la taille



7. La moyenne de r valeurs $(x_1, \ldots, x_r)$ apparaissant avec effectifs $(n_1, \ldots, n_r)$ est donnée par $\bar{X} = (n_1 x_1 + \cdots + n_r x_r)/N$, où $N = n_1 + \cdots + n_r$. Ici, x_i est le centre de la i-ième classe.
   On a $N = 205$. Le calcul donne :

$$\bar{X} = \frac{1}{205}(4 \times 31 + \cdots + 5 \times 59)m = 42,77m.$$

L'écart-type se calcule par :

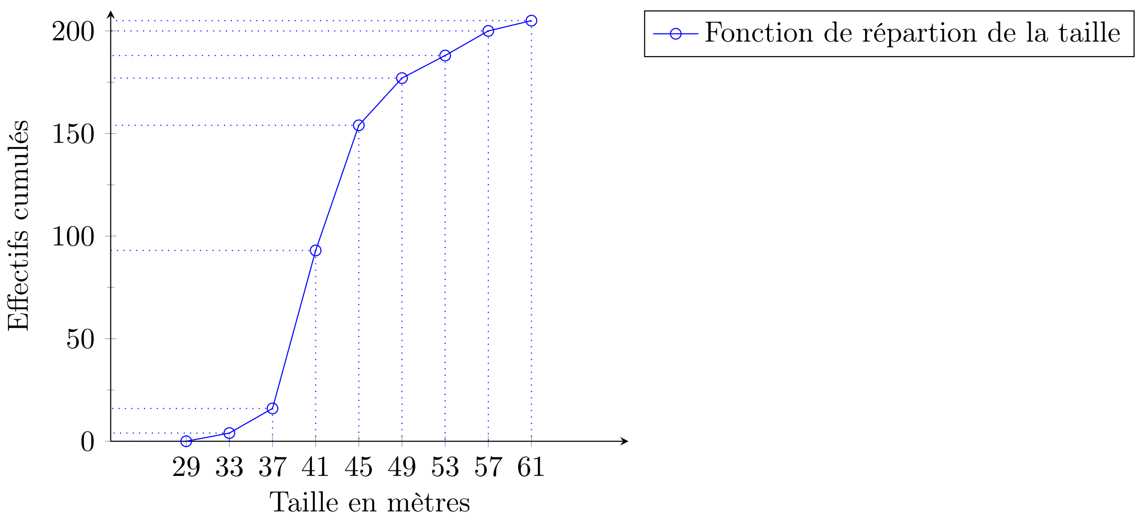
$$s = \sqrt{\frac{n_1(x_1 - \bar{X})^2 + \cdots + n_r(x_r - \bar{X})^2}{N-1}},$$

ce qui donne :

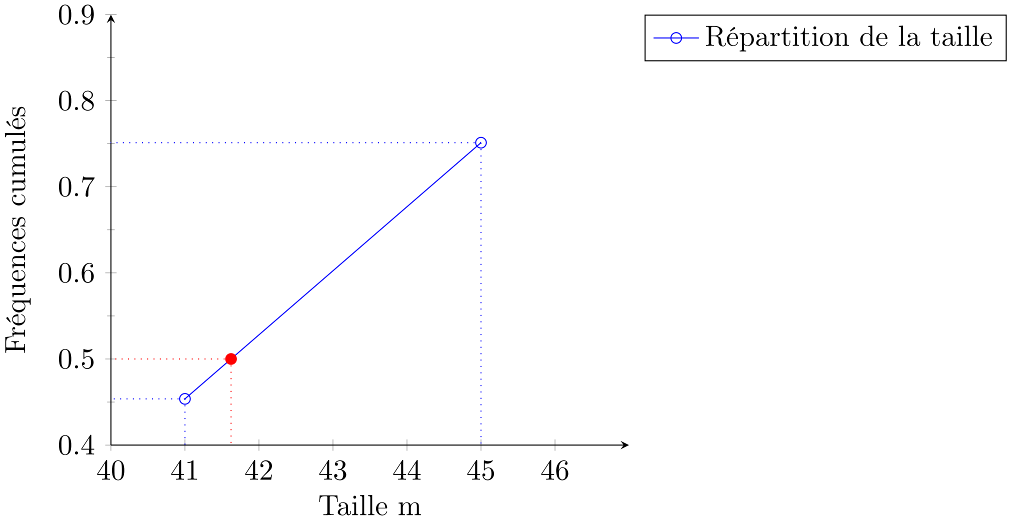
$$s = \sqrt{\frac{4 \times (31 - 42,77)^2 + \cdots + 5 \times (59 - 42,77)^2}{205 - 1}} = 5,71m.$$

8. On forme une ligne cassée reliant, pour toute classe C_i, les points b_{i-1} et b_i où l'abscisse de b_j est l'extrême supérieur b_j de la classe C_j et l'ordonnée de b_j est la fréquence cumulée F_j de C_j. On place les points sur un graphique avec titre et axes bien identifiés.

Fonction de répartition de la distribution de la taille



9. On identifie la classe médiane C_j, ici $j = 4$. Cette classe correspond à l'intervalle $[41, 45]m$. On trouve graphiquement le point qui correspond à la fréquence cumulée $0,5$ grâce au théorème de Thalès selon la figure suivante :



On calcule al médiane m, exprimée en mètres par la formule suivante

$$\begin{aligned} m &= b_{j-1} + \frac{0,5 - F_{j-1}}{F_j - F_{j-1}}(b_j - b_{j-1}) \\ &= 41 + \frac{0,5 - 0,45}{0,75 - 0,45}(45 - 41) = 41,62 \end{aligned}$$

en on obtient la valeur $m = 41,62$ exprimée en mètres.

10. On a $\bar{X} = 37,45m$ pour les pins autrichiens et $\bar{X} = 42,77m$ pour les pins turques, qui sont donc plus grands en moyenne.

    Quant à l'homogénéité, nous avons $s = 9,55m$ et $c = 0,255$ pour l'échantillon autrichien, tandis que pour les pins de Turquie on a $s = 5,70$ et $c = 0,133$, donc l'échantillon de pins noirs de Turquie est plus homogène.