

STT-1920 MÉTHODES STATISTIQUES

SOLUTIONS DES EXERCICES DU CHAPITRE 1

NUMÉRO 1. Déterminez la nature (quantitative continue, quantitative discrète, qualitative ordinale ou qualitative nominale) de chacune des variables suivantes.

- (a) La moyenne cumulative d'un étudiant de l'Université Laval.
- (b) La cote (A+, A, A-, B+,...) qu'un étudiant reçoit dans un cours.
- (c) La section (A, B, S) du cours STT-1000 à laquelle un étudiant de génie électrique est inscrit.
- (d) Le nombre de fois qu'un étudiant du baccalauréat en agronomie échoue le cours STT-1920 durant son baccalauréat.
- (e) Le nombre de louveteaux dans la portée d'une louve.
- (f) Le poids d'un loup à la naissance.
- (g) La couleur (blanche, jaune, orange, verte, bleue, marron, noire) de la ceinture d'un adepte de karaté Shotokan.
- (h) La couleur du feu de circulation en arrivant à l'intersection du Chemin Ste-Foy lorsque vous roulez sur la rue Myrand en direction nord.
- (i) Le salaire annuel d'un professeur de cégep.
- (j) Le nombre de tomates produites par un plant de tomate durant le mois d'août.
- (k) La nature d'une variable statistique.

RÉPONSES :

- (a) Quantitative continue.
- (b) Qualitative ordinale.
- (c) Qualitative nominale.
- (d) Quantitative discrète.
- (e) Quantitative discrète.
- (f) Quantitative continue.
- (g) Qualitative ordinale.
- (h) Qualitative ordinale.
- (i) Quantitative continue.
- (j) Quantitative discrète.
- (k) Qualitative nominale.

NUMÉRO 2. On a mesuré les diamètres de 34 prunes. Voici les résultats, en centimètres.

4.03	4.05	3.96	4.09	4.28	4.04	4.18	4.23	4.14
4.12	4.03	3.94	4.02	4.08	4.13	4.04	3.93	4.08
4.37	4.07	4.11	4.03	4.00	3.97	4.01	4.09	4.06
3.92	4.19	3.96	4.48	4.24	4.06	3.98		

- (a) Calculez la moyenne.
- (b) Calculez l'écart-type.
- (c) Calculez le coefficient de variation.
- (d) Dessinez un histogramme approprié.
- (e) Calculez le résumé à cinq nombres.
- (f) Dessinez le diagramme en boîte.
- (g) Qu'arrive-t-il à la moyenne si les valeurs 4.18, 4.23, 4.14 et 4.12 sont remplacées par les valeurs 4.68, 4.73, 4.64 et 4.62 ?
- (h) Qu'arrive-t-il à la médiane si les valeurs 4.18, 4.23, 4.14 et 4.12 sont remplacées par les valeurs 4.68, 4.73, 4.64 et 4.62 ?
- (i) Calculez le 30^e centile.

SOLUTIONS.

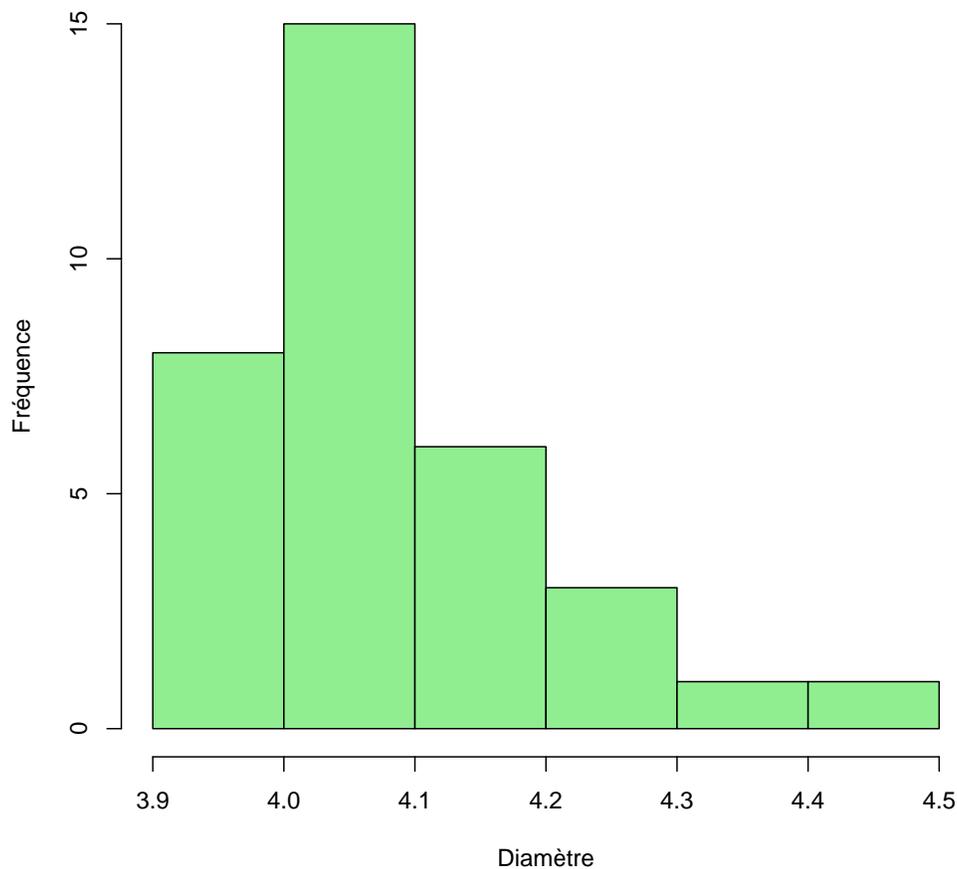
- (a) Avec la formule $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, j'obtiens

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{4.03 + 4.05 + 3.96 + \cdots + 4.06 + 3.98}{34} = 4.0856 \text{ cm.}$$

- (b) Avec la formule $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$, j'obtiens

$$\begin{aligned} s &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \sqrt{\frac{1}{33} \{(4.03 - 4.0856)^2 + \cdots + (3.98 - 4.0856)^2\}} \\ &= 0.1241 \text{ cm.} \end{aligned}$$

- (c) J'obtiens $s/\bar{x} = (0.1241 \text{ cm})/(4.0856 \text{ cm}) = 0.03$.
- (d) Avec l'aide du logiciel R, j'obtiens l'histogramme suivant. Vous devriez être capable d'obtenir cet histogramme aussi bien à la main qu'avec le logiciel R.



- (e) J'ordonne mes 34 diamètres, du plus petit au plus grand. Le plus petit diamètre est 3.92 cm et le plus grand est 4.48 cm. J'obtiens les quartiles $Q1 = q_{0.25}$, $Q2 = q_{0.50}$ et $Q3 = q_{0.75}$ à l'aide de la formule générale

$$q_\gamma = x_{(1+(n-1)\gamma)}.$$

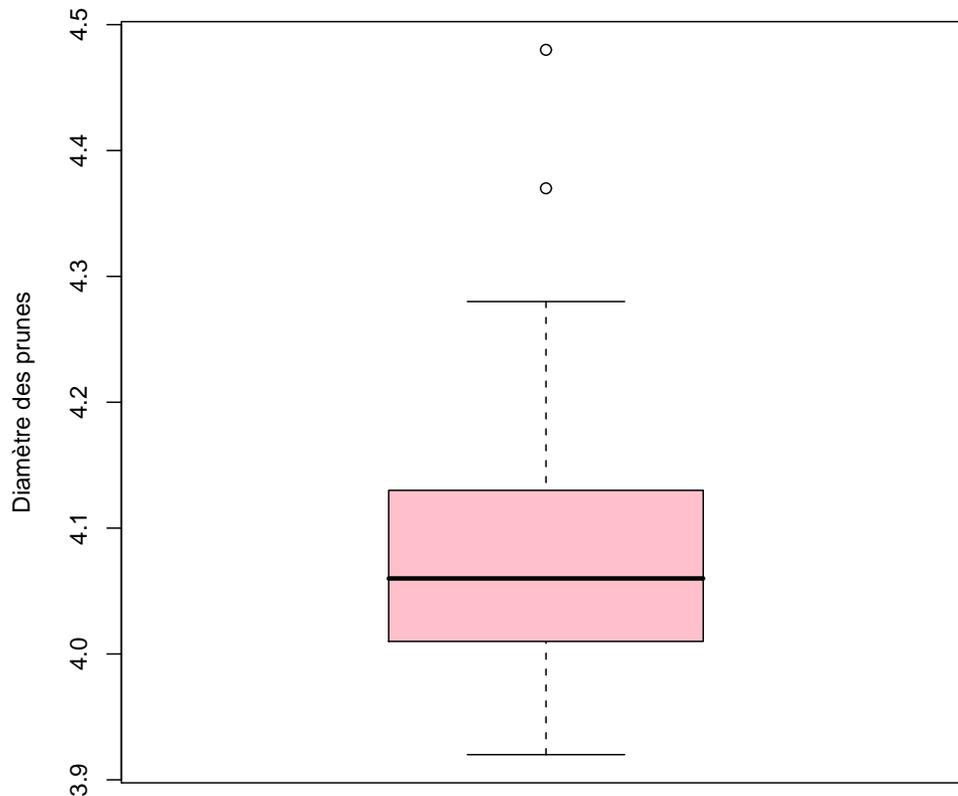
Par exemple, pour obtenir $Q1$, je fais

$$Q1 = q_{0.25} = x_{(1+(34-1)/4)} = x_{(9.25)} = x_{(9)} + \{0.25 \times (x_{(10)} - x_{(9)})\}.$$

Comme $x_{(9)} = 4.01$ et $x_{(10)} = 4.02$, j'obtiens $Q1 = 4.0125$. De la même façon, j'obtiens $Q2 = 4.0600$ et $Q3 = 4.1275$. Voici donc mon résumé à cinq nombres :

min	$Q1$	$Q2$	$Q3$	max
3.92	4.01	4.06	4.13	4.48

- (f) Avec l'aide de R, j'obtiens le diagramme en boîte suivant :



Vous devriez être capable de tracer ce diagramme en boîte à la main en utilisant la méthode décrite dans les notes de cours.

- (g) Si on remplace les valeurs 4.18, 4.23, 4.14 et 4.12 par les valeurs 4.68, 4.73, 4.64 et 4.62, la nouvelle moyenne est 4.144.
- (h) Si on remplace les valeurs 4.18, 4.23, 4.14 et 4.12 par les valeurs 4.68, 4.73, 4.64 et 4.62, la médiane ne change pas.
- (i)

$$q_{0.30} = x_{1+(34-1)\times 0.3} = x_{10.9} = x_{10} + \{0.9 \times (x_{11} - x_{10})\} = 4.02 + \{0.9 \times (4.03 - 4.02)\} = 4.029.$$

NUMÉRO 3. Un biologiste a mesuré les longueurs de 310 perchaudes attrapées dans le Chenal du Moine en août 2006. Le tableau suivant résume ces 310 mesures :

Longueur	Nombre de perchaudes
entre 21.0 et 22.0 cm	50
entre 22.0 et 23.0 cm	90
entre 23.0 et 24.0 cm	70
entre 24.0 et 26.0 cm	60
entre 26.0 et 30.0 cm	40

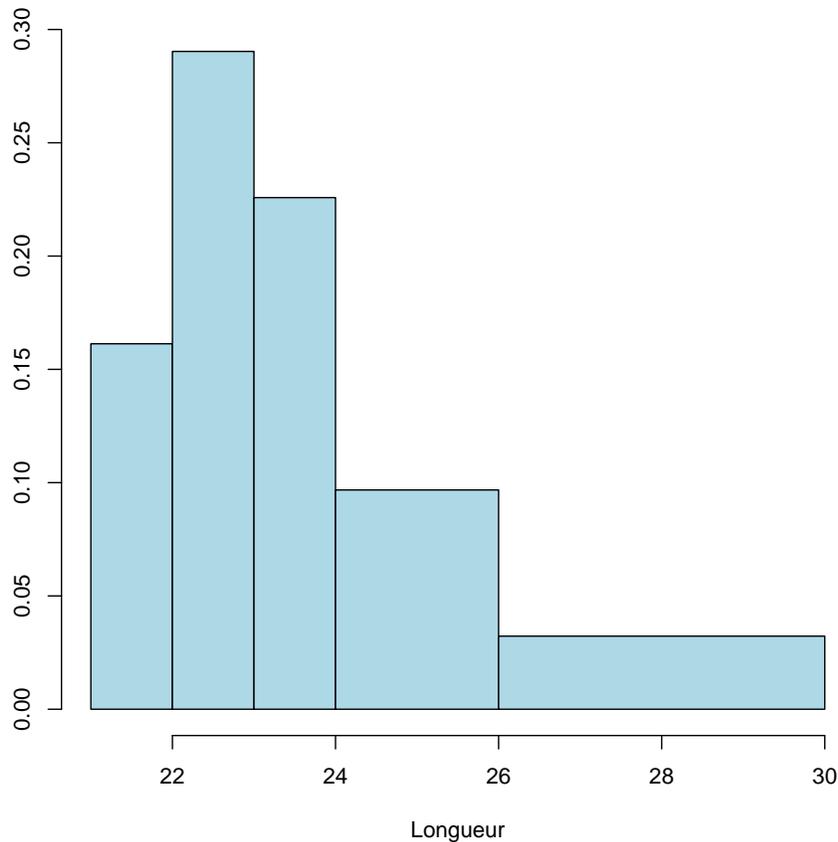
- (a) Représentez ces données à l'aide d'un histogramme.
- (b) Calculez une approximation pour la moyenne.
- (c) Calculez une approximation pour l'écart-type.
- (d) Calculez une approximation pour la médiane.
- (e) Calculez une approximation pour la proportion de perchaudes dont la longueur est située entre 25 et 27 cm.

SOLUTION.

- (a) On peut faire l'histogramme à la main. On complète d'abord notre tableau des données groupées de la façon suivante :

Longueur des perchaudes	Nombre de perchaudes	Fréquence relative	Hauteur de la boîte
entre 21.0 et 22.0 cm	50	50/310	$h_1 = \frac{50}{310 \times 1} = 0.161$
entre 22.0 et 23.0 cm	90	90/310	$h_2 = \frac{90}{310 \times 1} = 0.290$
entre 23.0 et 24.0 cm	70	70/310	$h_3 = \frac{70}{310 \times 1} = 0.226$
entre 24.0 et 26.0 cm	60	60/310	$h_4 = \frac{60}{310 \times 2} = 0.097$
entre 26.0 et 30.0 cm	40	40/310	$h_5 = \frac{40}{310 \times 4} = 0.031$

On obtient alors l'histogramme suivant :



(b) Avec la formule $\bar{x}_* = \frac{1}{n} \sum_{j=1}^{\ell} (f_j \times m_j)$ j'obtiens

$$\begin{aligned} \bar{x}_* &= \frac{1}{n} \sum_{j=1}^{\ell} (f_j \times m_j) = \frac{1}{310} \sum_{j=1}^5 (f_j \times m_j) \\ &= \frac{(50 \times 21.5) + (90 \times 22.5) + (70 \times 23.5) + (60 \times 25.0) + (40 \times 28.0)}{310} \\ &= 23.75806 \text{ cm.} \end{aligned}$$

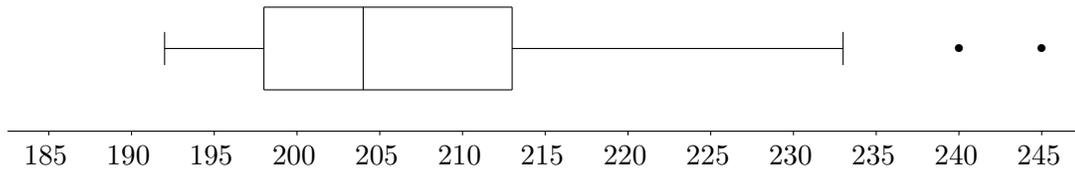
(c) Avec la formule $s_* = \sqrt{\frac{1}{n-1} \sum_{j=1}^{\ell} (f_j \times (m_j - \bar{x}_*)^2)}$ j'obtiens

$$\begin{aligned} s_* &= \sqrt{\frac{1}{n-1} \sum_{j=1}^{\ell} (f_j \times (m_j - \bar{x}_*)^2)} \\ &= \sqrt{\frac{1}{309} \sum_{j=1}^5 (f_j \times (m_j - \bar{x}_*)^2)} \\ &= \sqrt{\frac{1}{309} \{(50 \times (21.5 - 23.75806)^2) + \dots + (40 \times (28 - 23.75806)^2)\}} \\ &= 1.98241 \text{ cm.} \end{aligned}$$

(d) On utilise l'histogramme. On veut trouver, sur l'axe horizontal, la valeur qui est telle que 50% de la surface en est à gauche et 50% en est à droite. Cette valeur est $23 + \frac{15}{70} = 23.2143$. On peut également faire le raisonnement à partir de notre tableau des données groupées. On veut la valeur qui est telle que 155 observations lui sont inférieures et 155 lui sont supérieures. On note que cette valeur est quelque part entre 23 et 24 puisqu'on a $50 + 90 = 140$ observations inférieures à 23 et $60 + 40 = 100$ observations supérieures à 24. On a 70 observations entre 23 et 24. On voudrait couper ça en un groupe de 15 et un groupe de 55 pour en avoir en tout $50 + 90 + 15 = 155$ du côté gauche et $55 + 60 + 40 = 155$ du côté droit. On coupe donc notre intervalle $(23, 24]$ en un morceau de longueur $15/70$ et un morceau de longueur $55/70$. Notre approximation de la médiane est donc $23 + \frac{15}{70}$.

(e) On fait comme si la moitié des 60 observations de l'intervalle $(24, 26]$ étaient entre 25 et 26 et le quart des 40 observations de l'intervalle $(26, 30]$ étaient entre 26 et 27. Notre approximation pour le nombre d'observations entre 25 et 27 est donc $30 + 10 = 40$. Notre approximation pour la proportion des observations qui sont situées entre 25 et 27 est donc $40/310$.

NUMÉRO 4. Le diagramme en boîte suivant représente la distribution des poids (exprimés en livres) des 92 joueurs qui se sont présentés au camp d'entraînement de l'équipe de football du Rouge et Or en juin dernier.



Les énoncés suivants sont-ils vrais ou faux ?

- (a) La somme des poids des 3 joueurs les plus pesants est inférieure à 720 livres.
- (b) La moyenne de ces 92 poids est 204 livres.
- (c) L'écart inter-quartile de ces 92 poids est de 15 livres.
- (d) [Pour les étudiants qui sont familiers avec la loi normale] La loi normale est un bon modèle pour décrire cette distribution de poids.

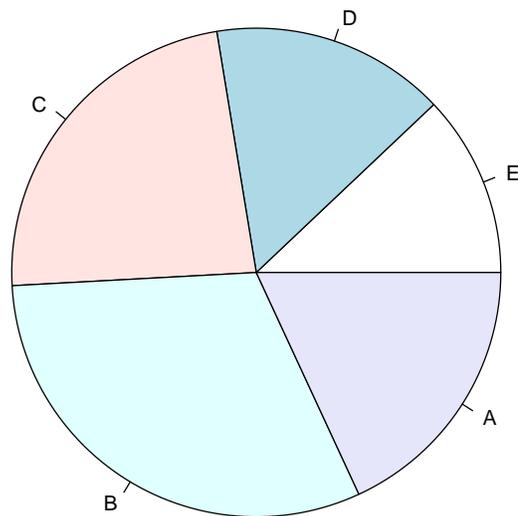
SOLUTION.

- (a) Vrai! Les trois plus grands poids sont 233, 240 et 245 livres. La somme de ces 3 poids est donc 718 livres.
- (b) Faux! C'est la médiane qui est égale à 204 livres. Et comme la distribution est très asymétrique et étirée vers la droite, la moyenne est supérieure à 204 livres.
- (c) Vrai! l'écart interquartile est $EIQ = Q3 - Q1 = 213 - 198 = 15$ livres..
- (d) Faux! La distribution est très asymétrique. La loi normale n'est donc pas un bon modèle pour décrire la distribution de ces poids.

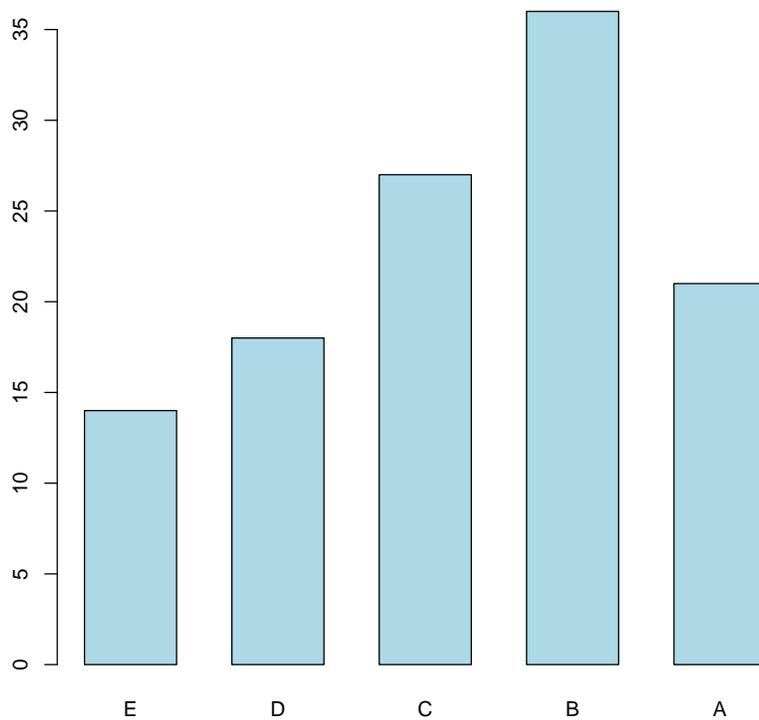
NUMÉRO 5. Parmi les 116 étudiants inscrits au cours STT-20694 à l'automne 2006, 21 ont reçu un A, 36 ont reçu un B, 27 ont reçu un C, 18 ont reçu un D et 14 ont reçu un E.

- (a) Représentez ces données à l'aide d'un diagramme en pointes de tarte.
- (b) Représentez ces données à l'aide d'un diagramme en bâtons.

SOLUTION. Avec l'aide de R, j'obtiens le diagramme en pointes de tarte que voici :



Avec l'aide de R, j'obtiens le diagramme en bâtons que voici :



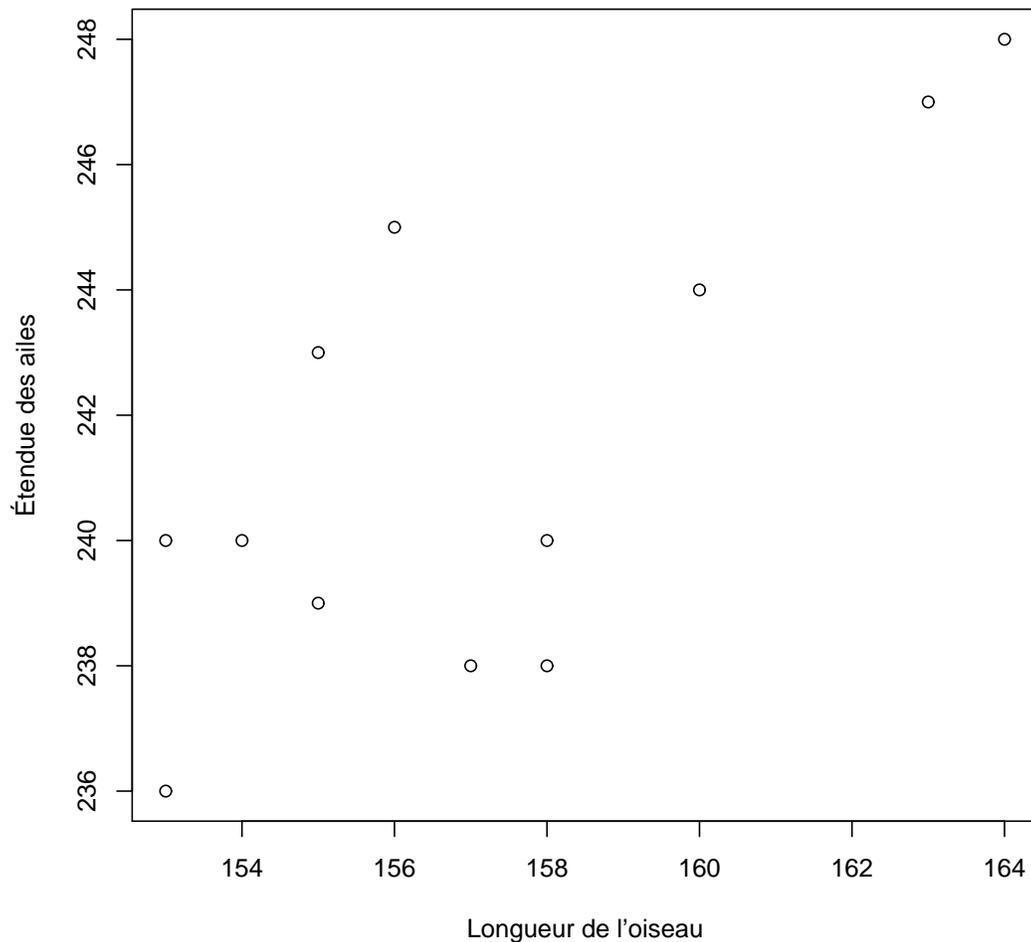
NUMÉRO 6. On a mesuré la variable $X = \text{longueur}$ (de la tête à la queue) et la variable $Y = \text{étendue}$ (du bout d'une aile à l'autre) de 12 oiseaux d'une certaine espèce. Voici les résultats en mm.

Longueur	156	154	153	153	155	163	157	155	164	158	158	160
Étendue	245	240	240	236	243	247	238	239	248	238	240	244

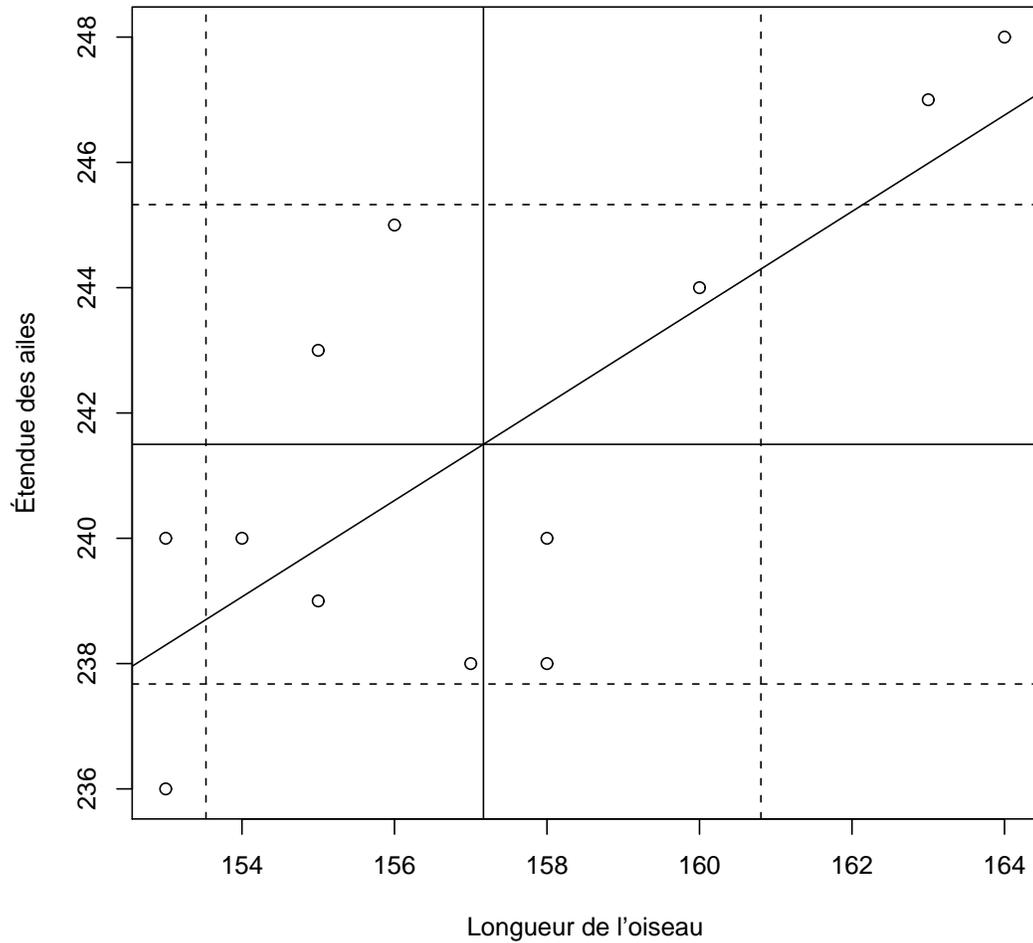
- Tracez le diagramme de dispersion.
- Calculez les moyennes \bar{x} et \bar{y} .
- Calculez les écarts-types s_x et s_y .
- Calculez le coefficient de corrélation r .
- Sur le diagramme de dispersion, interprétez les résultats obtenus en (b), (c) et (d).

SOLUTION.

- Avec l'aide de R, j'obtiens le diagramme de dispersion suivant :

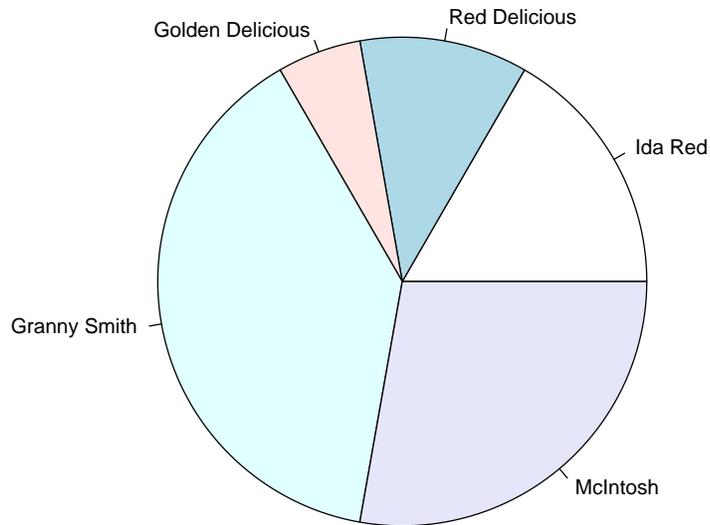


- (b) $\bar{x} = 157.17$ et $\bar{y} = 241.50$.
- (c) $s_x = 3.64$ et $s_y = 3.83$.
- (d) $r = 0.73$.
- (e) Le centre du nuage de points est situé au point $(\bar{x}, \bar{y}) = (157.17, 241.50)$. L'écart-type $s_x = 3.64$ mesure la dispersion en X alors que l'écart-type $s_y = 3.83$ mesure la dispersion en Y . Dans le graphe ci-dessous, la droite verticale pleine est située à la moyenne \bar{x} et les droites verticales en pointillés sont situées à $\bar{x} - s_x$ et $\bar{x} + s_x$. La droite horizontale pleine est située à la moyenne \bar{y} et les droites horizontales en pointillés sont situées à $\bar{y} - s_y$ et $\bar{y} + s_y$. Le coefficient de corrélation $r = 0.73$ mesure le degré d'association linéaire entre les deux variables. On voit sur le graphe que plus x est grand, plus y a tendance à être grand. Le coefficient de corrélation r est une mesure de la force de cette association.



NUMÉROS 7, 8, 9. Voir les solutions vidéo-clips sur le site web du cours.

NUMÉRO 10. Un verger compte 1512 pommiers. Le diagramme suivant représente les proportions de chacune des 5 espèces de pommiers de ce verger :



(a) Complétez le tableau suivant :

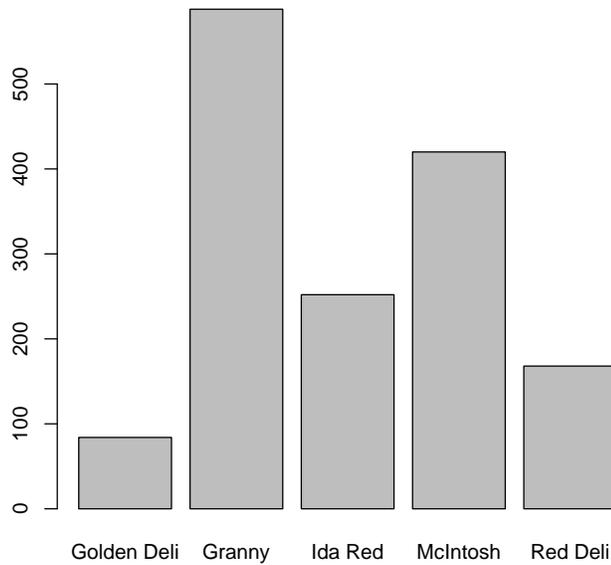
Espèce de pommes	Fréquence relative	Nombre de pommiers
Ida Red		
Red Delicious		
Golden Delicious		
Granny Smith		
McIntosh		

(b) Représentez ces données à l'aide d'un diagramme en bâtons.

SOLUTION.

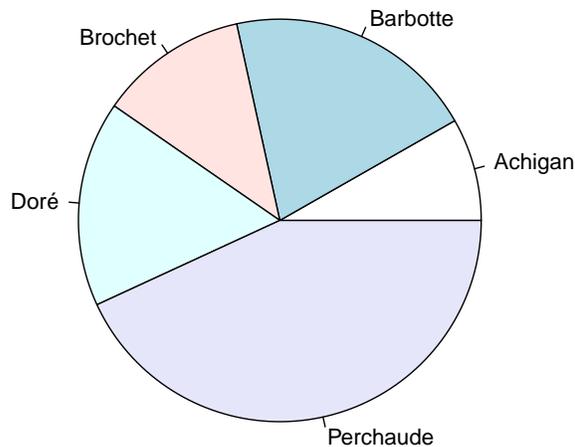
Espèce de pommes	Fréquence relative	Nombre de pommiers
Ida Red	$60/360$	252
Red Delicious	$40/360$	168
Golden Delicious	$20/360$	84
Granny Smith	$140/360$	588
McIntosh	$100/360$	420

Voici le diagramme en bâton produit avec R :



NUMÉRO 11. Lors d'une journée de pêche sur le lac St-Pierre, un groupe de pêcheurs ont attrapé 9 achigans, 22 barbottes, 13 brochets, 18 dorés et 47 perchaudes. Représentez ces données à l'aide d'un diagramme en pointes de tarte.

SOLUTION.



NUMÉRO 12. Voici les longueurs, en cm, des 22 barbottes dont il est question au numéro précédent :

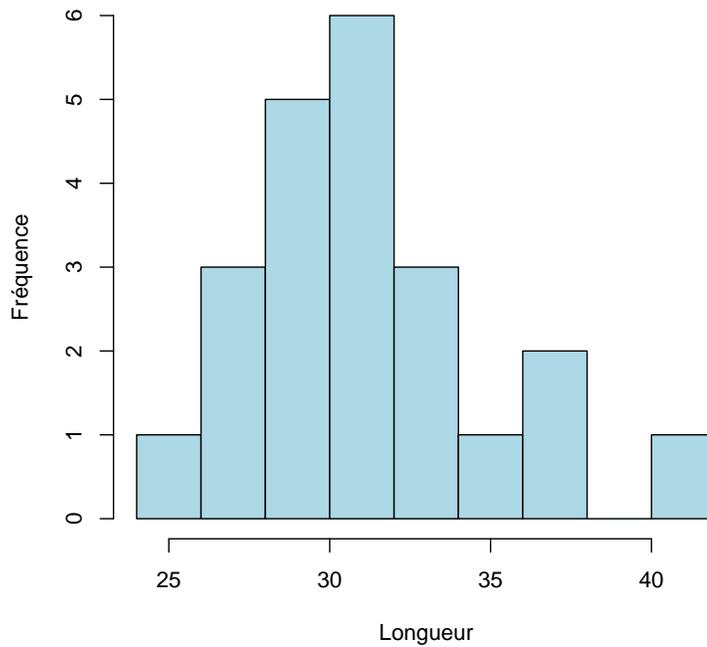
32.2 29.5 29.3 31.9 36.9 30.8 40.6 31.9 26.7 37.9 34.4
30.3 32.9 24.4 32.2 28.3 29.2 31.0 30.6 27.3 30.0 27.6

- (a) Calculez la moyenne.
- (b) Calculez l'écart-type.
- (c) Calculez le coefficient de variation.
- (d) Dessinez un histogramme approprié.
- (e) Calculez le résumé à cinq nombres.
- (f) Dessinez le diagramme en boîte.

SOLUTION.

(a) (b) et (c) $\bar{x} = 31.18$ cm, $s = 3.77$ cm, $s/\bar{x} = 0.121$

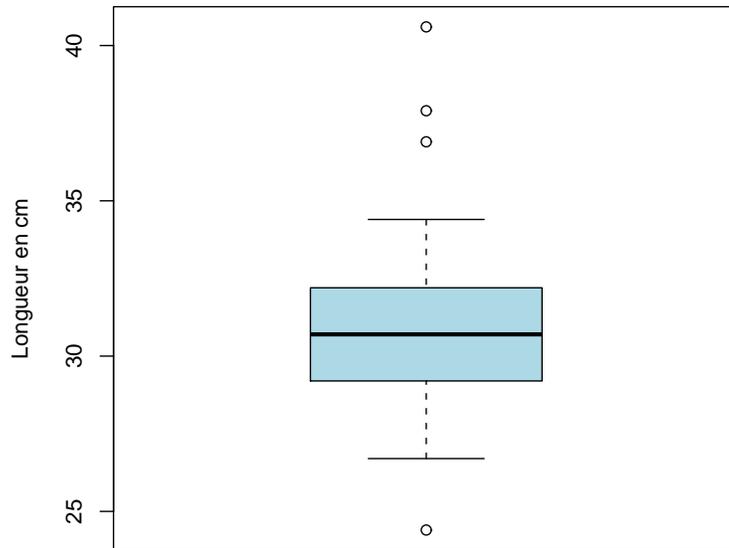
(d) Voici l'historgramme :



(e) Voici le résumé à cinq nombres :

min	Q1	Q2	Q3	max
24.400	29.225	30.700	32.200	40.600

(f) Voici le diagramme en boîte. Notez la présence d'observations aberrantes.



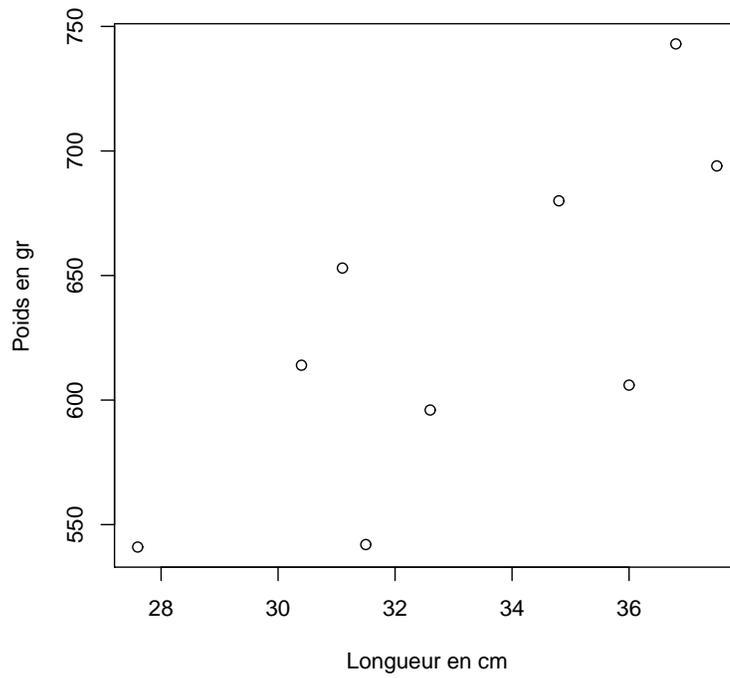
NUMÉRO 13. Voici les longueurs, en cm, et les poids, en gr, des 9 achigans dont il est question au numéro 11 :

Longueur	31.5	34.8	31.1	37.5	30.4	32.6	36.0	27.6	36.8
Poids	542	680	653	694	614	596	606	541	743

- Tracez le diagramme de dispersion.
- Calculez les moyennes \bar{x} et \bar{y} .
- Calculez les écarts-types s_x et s_y .
- Calculez le coefficient de corrélation r .
- Sur le diagramme de dispersion, interprétez les résultats obtenus en (b), (c) et (d).

SOLUTION.

- Voici le diagramme de dispersion.



- (b) $\bar{x} = 33.14$ cm et $\bar{y} = 629.9$ gr.
- (c) $s_x = 3.33$ cm et $s_y = 68.51$ gr.
- (d) $r = 0.745$.
- (e) Comme au numéro 6 (e).