

Entrée en matière

En contexte

Manuel • p. 50

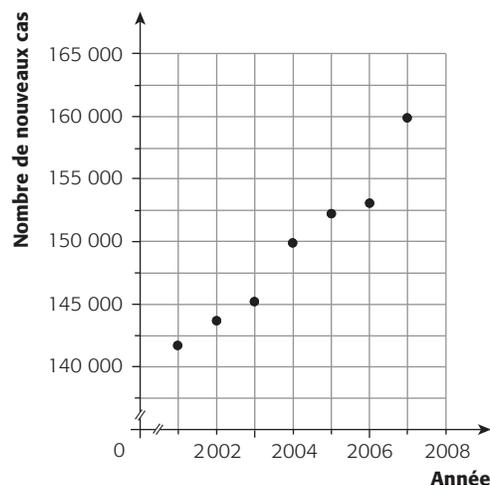
1. a) La population visée par cette étude est formée des jeunes Québécois de 12 à 19 ans.
- b) Le caractère étudié est le nombre de portions de fruits et de légumes consommés dans la journée d'hier. Ce caractère est de type quantitatif discret.
- c) La moyenne est de 3,6 portions de fruits et légumes consommés par jour et la médiane est de 5 portions.
- d) Puisque la médiane est 5, on peut en déduire que 50 % des jeunes interrogés ont atteint l'objectif. Cependant, puisque la moyenne est de 3,6 portions de fruits et légumes, de façon générale, les jeunes n'auraient pas atteint l'objectif du Défi Santé 5/30 pour la consommation de fruits et de légumes.
- e) Non, cette conclusion ne peut pas être généralisée à l'ensemble des jeunes Québécois de 12 à 19 ans puisque l'échantillon n'est pas représentatif de la population. Les jeunes proviennent tous de la même école. L'échantillon est trop petit et on ne sait pas si on a respecté l'âge de la population (12 à 19 ans). En effet, dans une école secondaire, il y a très peu de jeunes âgés de 18 ou 19 ans.

Manuel • p. 51

2. a) Vrai
- b) Faux
- c) Impossible à déterminer. On ne connaît pas le nombre de garçons et de filles qui ont participé au sondage.
- d) On voit que 50 % des filles ont fait au plus 4 séances et que 25 % des garçons ont fait au plus 4 séances. Le pourcentage chez les filles est donc le double, mais comme on ne connaît pas le nombre de filles et le nombre de garçons interrogés, on ne peut affirmer que le nombre de filles ayant fait plus de 4 séances est le double de celui des garçons.
- e) Vrai

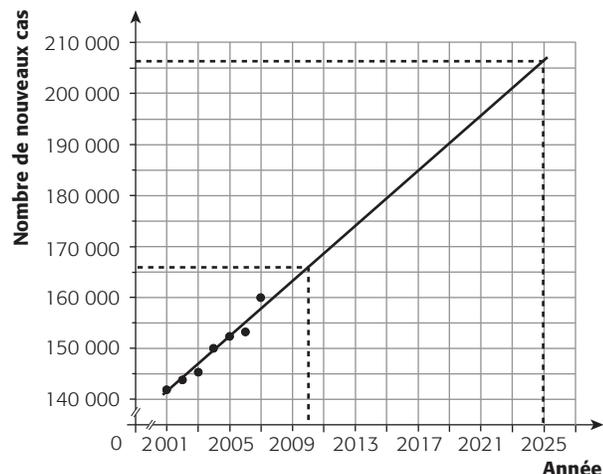
3. a)

Le cancer au Canada



b)

Le cancer au Canada



- c) 1) En 2010, le nombre de nouveaux cas déclarés sera d'environ 167 000 (voir la droite ajustée au nuage de points ci-dessus).
- 2) En 2025, le nombre de nouveaux cas déclarés sera d'environ 206 500 (voir la droite ajustée au nuage de points ci-dessus).
- d) L'estimation 1), faite pour 2010, est la plus fiable, parce qu'elle se trouve plus près du nuage de points. S'il devait y avoir une erreur d'évaluation dans la droite ajustée au nuage de points, cette erreur serait accentuée au fil des ans. De plus, d'ici 2025, des changements dans le mode de vie et des avancées technologiques sont susceptibles de modifier les résultats.

En bref

Manuel • p. 52

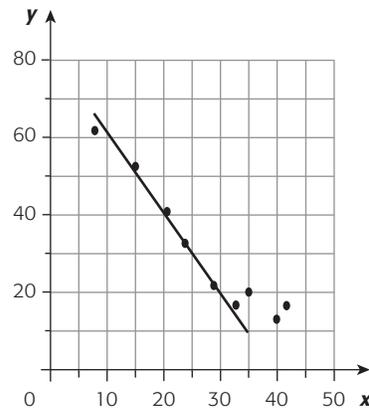
- Situation ①

 - Population : Les écoles d'une commission scolaire
 - Individu : Chaque école
 - Caractère étudié : Le nombre d'élèves par école

Situation ②

 - Population : Les étudiants d'une université
 - Individu : Chacun des étudiants
 - Caractère étudié : L'âge des étudiants
- La moyenne est d'environ 161,1.
 - $Q_1 = 153$, $Q_2 = 156$ et $Q_3 = 161$
 - (voir au bas de la page)
 - L'étendue est de 77.
L'étendue interquartile est de 8.
 - Le deuxième quart
 - Le quatrième quart
 - Oui, cette distribution comporte des données aberrantes. Toutes les données plus grandes que 173 sont aberrantes, soit 180, 192 et 225.
- $y_1 = \frac{-x}{2} + 8$ $y_2 = 4x - 7$ $y_3 = \frac{2x}{3} + \frac{16}{3}$

4. a) et b)



- c) 1) 42 2) 15

5. L'estimation par interpolation est généralement plus fiable que l'estimation par extrapolation.

Section 1 Les mesures de dispersion et de position



Les courbes de croissance

Manuel • p. 53

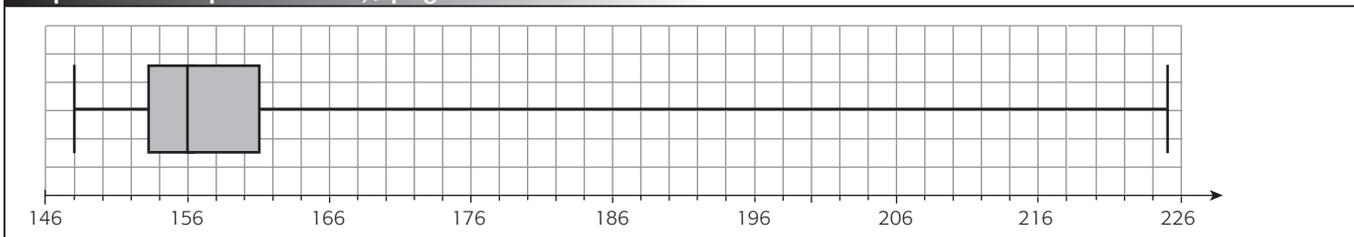
On calcule l'indice de masse corporelle de chacun des garçons. (voir au bas de la page)

Graphique de la croissance (voir à la page 27)

Graphique de l'IMC (voir à la page 28)

Seuls Simon, Laurent et Lucas ont un indice de masse corporelle se situant sur le «chemin de la santé».

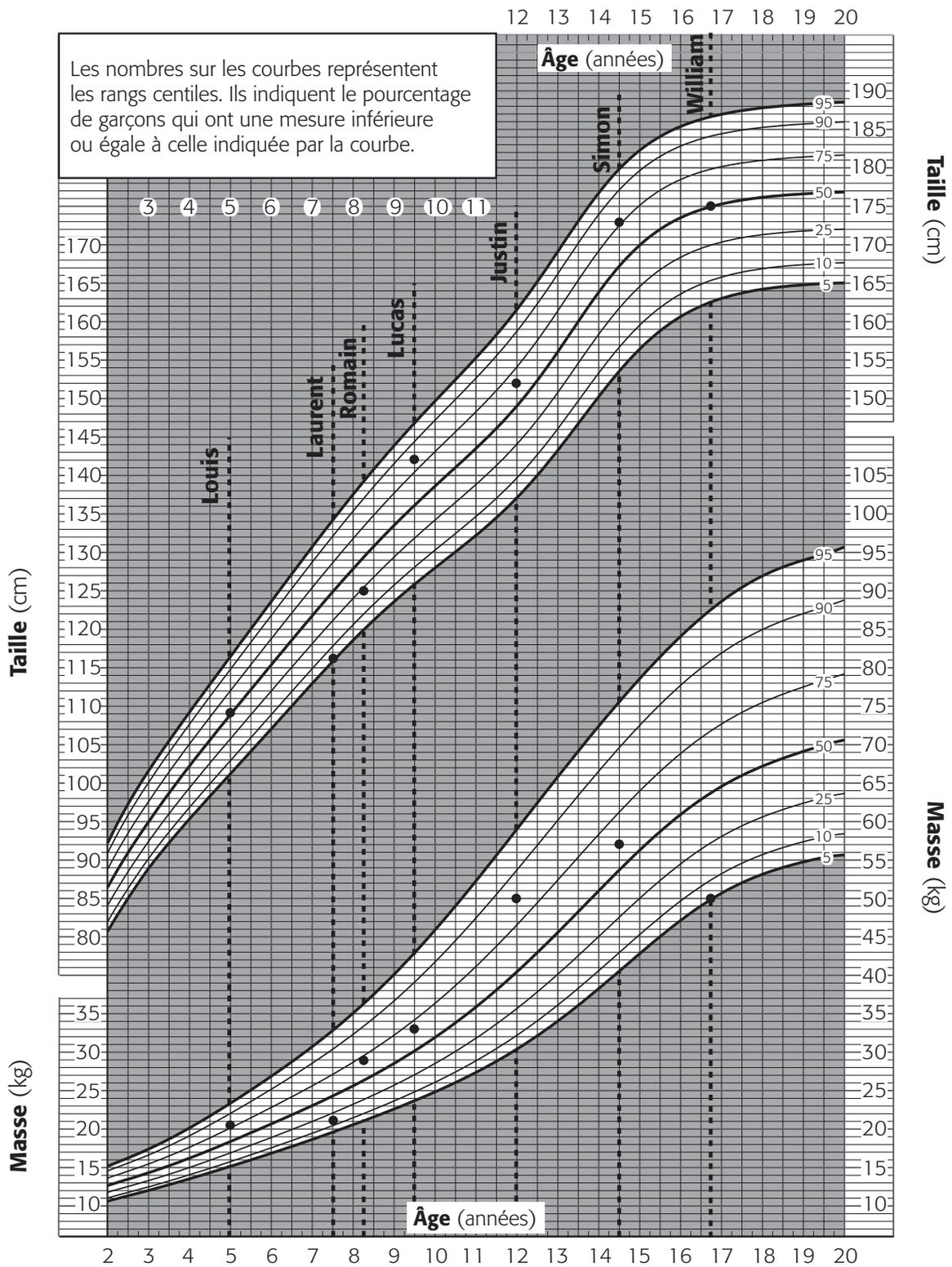
Réponse à la question 2 c), page 52



Réponse à la question page 53

| Nom | Âge | Taille (rang centile) | Taille (m) | Masse (rang centile) | Masse (kg) | IMC | Appréciation |
|---------|-------|-----------------------|------------|----------------------|------------|------|--------------|
| Simon | 14,5 | 75 | 1,73 | 65 | 57 | 19,0 | Correcte |
| Laurent | 7,5 | 5 | 1,16 | 10 | 21 | 15,6 | Correcte |
| Justin | 12 | 60 | 1,52 | 80 | 50 | 21,6 | Attention |
| William | 16,75 | 50 | 1,75 | 5 | 50 | 16,3 | Attention |
| Romain | 8,25 | 25 | 1,25 | 75 | 29 | 18,6 | Attention |
| Louis | 5 | 50 | 1,09 | 80 | 20,5 | 17,3 | Attention |
| Lucas | 9,5 | 80 | 1,42 | 65 | 33 | 16,4 | Correcte |

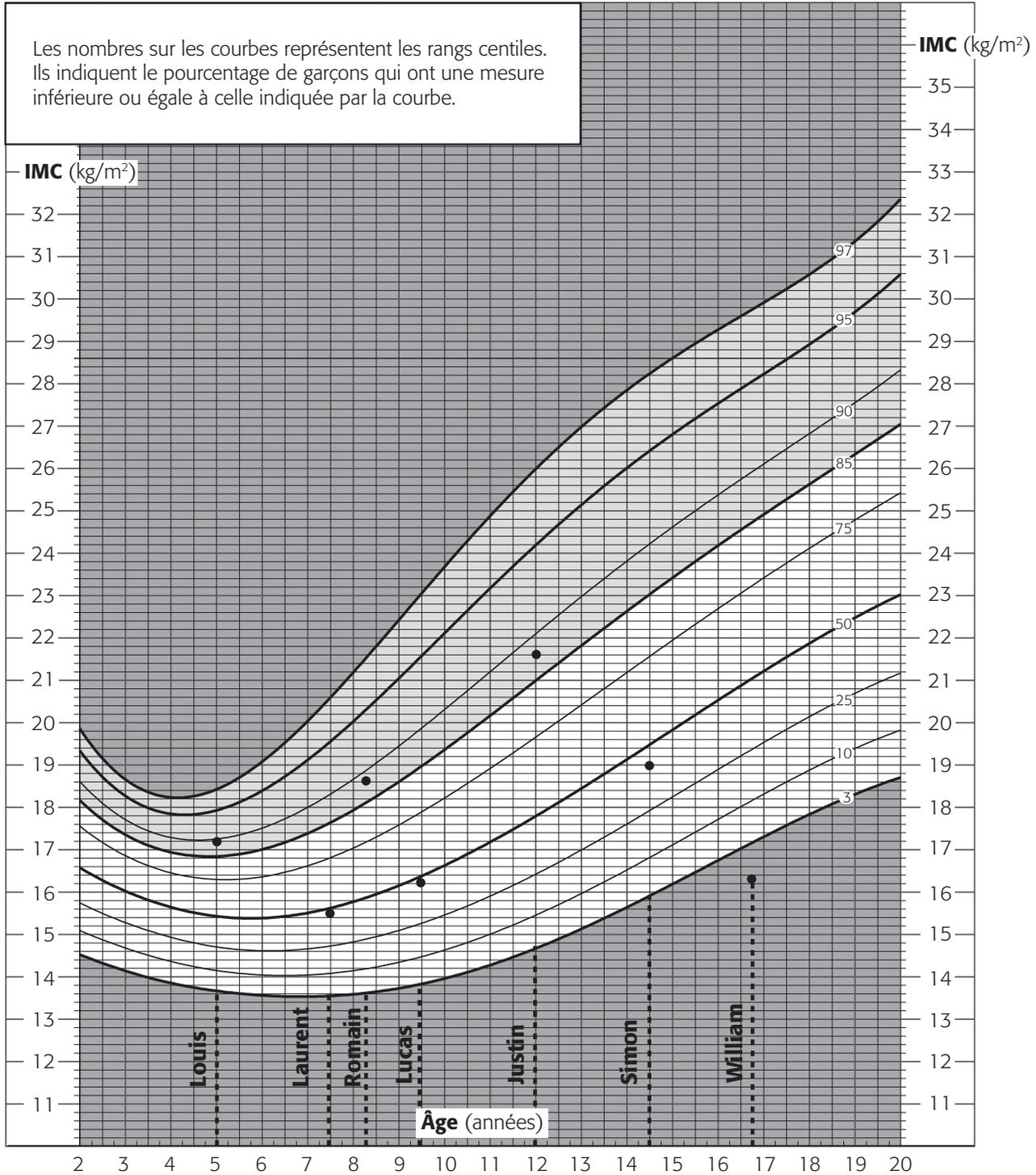
Courbes de croissance pour les garçons de 2 à 20 ans



Adapté de : *Courbes de croissance des CDC américains pour les garçons de 2 à 20 ans* publié par le National Center for Health Statistics en collaboration avec le National Center for Chronic Disease Prevention and Health Promotion, 2000.

**Courbes de croissance pour les garçons de 2 à 20 ans :
indice de masse corporelle (IMC)**

Les nombres sur les courbes représentent les rangs centiles. Ils indiquent le pourcentage de garçons qui ont une mesure inférieure ou égale à celle indiquée par la courbe.



Adapté de : *Courbes de croissance des CDC américains pour les garçons de 2 à 20 ans : indice de masse corporelle (IMC)* publié par le National Center for Health Statistics en collaboration avec le National Center for Chronic Disease Prevention and Health Promotion, 2000.

ACTIVITÉ

D'EXPLORATION

① Des joueurs de taille

Manuel • p. 54

A Le diagramme à tige et à feuilles se présente comme un histogramme couché sur le côté. Les deux modes de représentation ont ainsi la même allure. À la différence de l'histogramme, le diagramme à tige et à feuilles permet de voir toutes les données de la distribution.

- B** 1) Aucune modification
2) Pour la tige 19, les feuilles (les chiffres des unités) 0-2-3-3-5 seraient remplacées par 2-3-3-5-7.

Manuel • p. 55

C La taille, en cm, des joueurs des Knicks

| | | |
|----|--|-------------|
| 17 | | 2 |
| 18 | | 3 4 5 |
| 19 | | 3 5 8 8 |
| 20 | | 0 2 3 5 7 8 |
| 21 | | 3 |

D Plusieurs réponses sont possibles. *Exemple :*

- La taille minimale chez les Warriors est de 1,85 m alors qu'elle est de 1,72 m chez les Knicks.
- La taille maximale chez les Warriors est de 2,07 m alors qu'elle est de 2,13 m chez les Knicks.
- L'étendue de la taille des joueurs des Knicks (41) est presque égale au double de celle des joueurs des Warriors (22).
- Les tailles des joueurs des Warriors sont plus proches les unes des autres que celles des joueurs des Knicks.

E Les tailles de l'équipe des Warriors sont le plus homogènes.

F Plusieurs réponses sont possibles. *Exemples :*

La moyenne des tailles est presque la même dans les deux équipes. Toutefois, l'une des équipes est constituée de joueurs de tailles très variées dont un joueur de très grande taille, alors que l'autre est constituée de joueurs ayant tous des grandes tailles, peu variées. Quelle est celle qui est favorisée? On ne peut y répondre d'une façon catégorique; selon certains, c'est la première, selon d'autres, c'est la seconde!

G On peut conserver une même tige pour les deux distributions et présenter les feuilles de chaque distribution de part et d'autre de cette tige :

La taille, en cm,
des joueurs des Warriors

| | |
|--|-----------------|
| | 8 8 5 |
| | 5 3 3 2 0 |
| | 7 5 5 4 3 2 2 0 |

La taille, en cm,
des joueurs des Knicks

| | | |
|----|--|-------------|
| 17 | | 2 |
| 18 | | 3 4 5 |
| 19 | | 3 5 8 8 |
| 20 | | 0 2 3 5 7 8 |
| 21 | | 3 |

Ai-je bien compris?

1. a) ①

| | | |
|---|--|---------------------|
| 1 | | 7 |
| 2 | | 1 2 3 4 5 5 5 7 8 9 |
| 3 | | 0 1 3 4 6 8 9 |
| 4 | | 3 |

②

| | | |
|----|--|-----------|
| 12 | | 1 5 8 9 |
| 13 | | 5 |
| 14 | | 0 0 4 8 9 |
| 15 | | 3 4 |
| 16 | | 0 6 |
| 17 | | |
| 18 | | 3 |

b) Les données sont plus dispersées dans la deuxième distribution que dans la première.

2. Plusieurs réponses sont possibles. *Exemple :*

- a) L'âge des élèves d'une classe de mathématiques de la séquence CST de la 2^e année du 2^e cycle.
b) L'âge des patients dans la salle d'attente d'un hôpital.

ACTIVITÉ

D'EXPLORATION

② Dansez dans les rues

Manuel • p. 56

A Plusieurs réponses sont possibles. *Exemples :*

- Les hommes ont le plus apprécié le film, car ils ont attribué plus de cotes 9 et 10 que les femmes.
- Les femmes ont le plus apprécié le film, car aucune d'elles n'a attribué une cote inférieure à 7.

B Dans le groupe des femmes, les appréciations sont le plus homogènes.

C La moyenne des cotes d'appréciation dans chacun des deux groupes est de 8. Cette mesure ne révèle donc pas de différence entre l'appréciation des hommes et celles des femmes.

D (voir au bas de la page suivante)

- E** Plus les valeurs des écarts à la moyenne dans un groupe sont grandes, plus les données sont dispersées (plus les cotes d'appréciation sont hétérogènes).
- F** Non, car la moyenne des écarts à la moyenne d'une distribution est toujours nulle.
- G** (voir au bas de la page)
- H** (voir au bas de la page)

- I** L'écart moyen des cotes attribuées par les hommes est de 1,7. Cela signifie que la cote attribuée par un homme s'écarte, en moyenne, de 1,7 de la moyenne des cotes attribuées par les hommes (qui est de 8).
- L'écart moyen des cotes attribuées par les femmes est de 0,7. Cela signifie que la cote attribuée par une femme s'écarte, en moyenne, de 0,7 de la moyenne des cotes attribuées par les femmes (qui est de 8).

Réponses aux questions **D**, **G**, et **H**, page 56

D

| Cotes attribuées par les hommes | Écart à la moyenne | Cotes attribuées par les femmes | Écart à la moyenne |
|---------------------------------|--------------------|---------------------------------|--------------------|
| 10 | 2 | 10 | 2 |
| 10 | 2 | 9 | 1 |
| 10 | 2 | 9 | 1 |
| 9 | 1 | 8 | 0 |
| 9 | 1 | 8 | 0 |
| 9 | 1 | 8 | 0 |
| 9 | 1 | 8 | 0 |
| 8 | 0 | 8 | 0 |
| 8 | 0 | 7 | -1 |
| 6 | -2 | 7 | -1 |
| 5 | -3 | 7 | -1 |
| 3 | -5 | 7 | -1 |
| Moyenne : 8 | | Moyenne : 8 | |

G

| Cotes attribuées par les hommes | Valeur absolue de l'écart à la moyenne | Cotes attribuées par les femmes | Valeur absolue de l'écart à la moyenne |
|---------------------------------|--|---------------------------------|--|
| 10 | 2 | 10 | 2 |
| 10 | 2 | 9 | 1 |
| 10 | 2 | 9 | 1 |
| 9 | 1 | 8 | 0 |
| 9 | 1 | 8 | 0 |
| 9 | 1 | 8 | 0 |
| 9 | 1 | 8 | 0 |
| 8 | 0 | 8 | 0 |
| 8 | 0 | 7 | 1 |
| 6 | 2 | 7 | 1 |
| 5 | 3 | 7 | 1 |
| 3 | 5 | 7 | 1 |

H

| | Cotes attribuées par les hommes | Cotes attribuées par les femmes |
|-------------|---------------------------------|---------------------------------|
| Écart moyen | $\frac{5}{3} \approx 1,6$ | $\frac{2}{3} \approx 0,6$ |

- J** Plus l'écart moyen est grand, plus la distribution est hétérogène. Et, inversement, plus l'écart moyen est petit, plus la distribution est homogène.

Manuel • p. 57

- K** Dans le groupe des hommes, pour que l'écart moyen passe de $\frac{20}{12}$ à 1,25 ou $\frac{20}{16}$ après l'ajout de 4 nouvelles cotes, il faut que la somme des valeurs absolues des écarts à la moyenne des quatre nouvelles cotes soit égale à 0. Autrement dit, il faut que chacune de ces quatre nouvelles cotes soit égale à la moyenne, c'est-à-dire, 8. Les cotes attribuées par les quatre hommes sont donc 8, 8, 8 et 8.

Dans le groupe des femmes, pour que l'écart moyen passe de $\frac{8}{12}$ à 1,25 ou $\frac{30}{24}$ après l'ajout des 12 nouvelles cotes, il faut que la somme des valeurs absolues des écarts à la moyenne des 12 nouvelles cotes soit égale à 22.

Plusieurs réponses sont possibles. *Exemple :*

Les cotes attribuées par les 12 femmes sont 5, 5, 5, 6, 6, 7, 7, 9, 9, 9, 10, 10.

Ai-je bien compris?

- a) ① $\bar{x} = 9$ ② $\bar{x} = 12,9$ ③ $\bar{x} = 43,5$ ④ $\bar{x} = 1,3$
 b) ① ÉM = 2,4 ② ÉM = 3,08
 ③ ÉM = 9,75 ④ ÉM \approx 0,71
 c) ④ - ① - ② - ③

ACTIVITÉ

D'EXPLORATION ③ Le test de mémoire

Manuel • p. 58

- A** Gabrielle est classée 67^e sur 176 participants, cela veut dire, que sur 100, elle serait 38^e.

Sarah-Maude est classée 78^e sur 204 participants; cela veut dire que sur 100, elle serait 38^e.

Gabrielle et Sarah-Maude sont au même rang dans leur distribution respective. Cela ne veut toutefois pas dire que les deux ont obtenu le même score.

- B** $\frac{204 - 77}{204} = \frac{127}{204}$
 $\frac{127}{204} \cdot 100 \approx 62,3 \%$
- C** 63

- D** Il faut calculer $R_{100}(193)$, soit le pourcentage, arrondi à l'unité supérieure, des scores moins bons ou égaux à 193 :

$$R_{100}(193) = \frac{204 - 72}{204} \cdot 100 \approx 64,7$$

$$R_{100}(193) = 65$$

- E** Nombre de données inférieures ou égales au score recherché : $\frac{51}{100} \cdot 204 = 104,04$

Le score recherché occupe la 104^e position de la distribution à partir du moins bon score, c'est-à-dire la 101^e meilleure position. Sa valeur est 167.

Manuel • p. 59

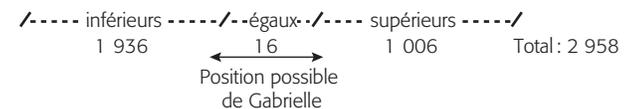
- F** Nombre total de participants: 2 958
 Nombre de scores inférieurs à celui de Kevin : 1 247
 Nombre de scores supérieurs à celui de Kevin : 1 632
 On en déduit que le nombre de personnes ayant le score de Kevin est :

$$2\,958 - (1\,247 + 1\,632) = 79.$$

$$\text{Le rang centile de Kevin} = \frac{(1\,247 + 79)}{2\,958} \cdot 100 \approx 44,8$$

Le rang centile de Kevin est 45.

- G** Sur 2 958 participants, le rang centile de Gabrielle est 66. Cela veut dire que 66 % des participants, soit environ 1 952, ont un score inférieur ou égal au sien. Autrement dit, environ 1 006 participants ont un meilleur score qu'elle.



La position de Gabrielle dans la distribution peut se situer n'importe où entre la 1007^e et la 1022^e inclusivement.

Ai-je bien compris?

- a) ① 63 ② 64 ③ 59
 b) ① $X = 103$ ② $X = 65$ ③ $X = 27$
- Le rang centile de Francis est 68.

Mise en pratique

Manuel • p. 63

- Niveau de difficulté : faible**

a) L'étendue est de 52 ans.
 b) L'âge médian est de 31 ans.
 c) Diagramme de quartiles $Q_1 = 23$, $Q_2 = 31$ et $Q_3 = 53,5$
 (voir la représentation graphique au bas de la page suivante)

2. Niveau de difficulté : faible

a)

| | |
|---|-----------|
| 2 | 5 8 |
| 3 | 2 3 |
| 4 | 3 3 9 |
| 5 | 0 2 2 4 8 |
| 6 | 3 5 |
| 7 | 3 7 |
| 8 | 1 7 |

b)

| | |
|----|-------|
| 11 | 0 2 |
| 12 | 1 7 |
| 13 | 8 |
| 14 | 0 5 |
| 15 | 3 9 |
| 16 | 1 2 3 |
| 17 | |
| 18 | 8 |

c)

| | |
|-----|-------|
| 121 | 8 |
| 122 | 2 |
| 123 | 5 7 9 |
| 124 | |
| 125 | 1 5 |
| 126 | 4 |
| 127 | 2 |
| 128 | 8 |
| 129 | 0 5 |
| 130 | |
| 131 | 6 |

3. Niveau de difficulté : faible

a) Note (%) à l'examen théorique

| | |
|---|---------------------------|
| 5 | 3 6 |
| 6 | 3 5 5 |
| 7 | 1 2 6 6 8 |
| 8 | 0 0 0 1 2 3 3 5 6 6 8 8 9 |
| 9 | 0 0 2 6 9 |

Note (%) à l'examen pratique

| | |
|---|-----------------------------|
| 6 | 9 |
| 7 | 2 2 4 6 6 6 6 8 8 9 |
| 8 | 0 2 3 3 3 5 5 6 6 6 8 8 9 9 |
| 9 | 1 3 6 |

b) Examen théorique:
Moyenne : 79,75 %
Médiane : 81,5 %
Étendue : 46 %

Examen pratique:
Moyenne : 82,11 %
Médiane : 83 %
Étendue : 27 %

c) La distribution des notes est plus homogène pour l'examen pratique.

Manuel • p. 64

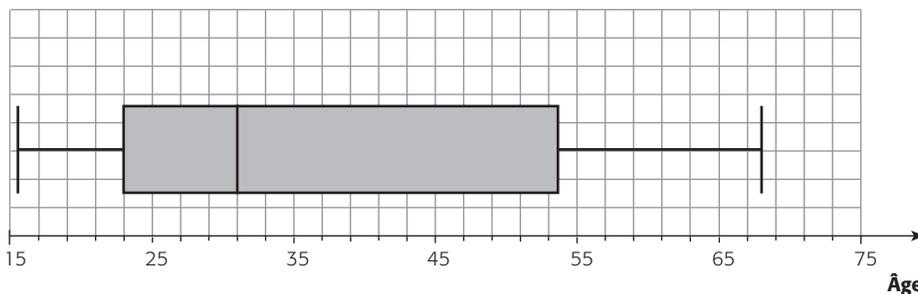
4. Niveau de difficulté : faible

- a) Le diagramme à tige et à feuilles ① correspond au diagramme de quartiles (A).
Le diagramme à tige et à feuilles ② correspond au diagramme de quartiles (D).
Le diagramme à tige et à feuilles ③ correspond au diagramme de quartiles (B).
Le diagramme à tige et à feuilles ④ correspond au diagramme de quartiles (C).

b) Équipe Bouffard, équipe Hébert, équipe Lavigne et équipe Navarrete.

Réponse à la question 1 c), page 63

Âge des employés du club de golf Le Verdoyant



5. Niveau de difficulté : faible

a) $\bar{x} = 27,4$

b)

| Donnée | Écart à la moyenne | Valeur absolue de l'écart à la moyenne |
|--------|--------------------|--|
| 30 | 2,6 | 2,6 |
| 27 | -0,4 | 0,4 |
| 18 | -9,4 | 9,4 |
| 41 | 13,6 | 13,6 |
| 23 | -4,4 | 4,4 |
| 16 | -11,4 | 11,4 |
| 38 | 10,6 | 10,6 |
| 27 | -0,4 | 0,4 |
| 24 | -3,4 | 3,4 |
| 30 | 2,6 | 2,6 |

c) 5,88

6. Niveau de difficulté : faible

a) $\bar{x} = 9$ $ÉM = 2,67$

b) $\bar{x} = 31$ $ÉM = 4,89$

c) $\bar{x} = 21,2$ $ÉM = 3,84$

Manuel • p. 65

7. Niveau de difficulté : faible

- a) Équipe Bouffard $ÉM = 4,75$
 Équipe Hébert $ÉM = 5,25$
 Équipe Lavigne $ÉM = 9,25$
 Équipe Navarrete $ÉM = 7,75$

- b) Famille Bouffard – Famille Hébert –
 Famille Navarrete – Famille Lavigne

8. Niveau de difficulté : faible

- a) La distribution ② devrait avoir le plus grand écart moyen.
 b) La distribution ② devrait avoir le plus grand écart moyen.
 c) La distribution ② devrait avoir le plus grand écart moyen.
 d) La distribution ② devrait avoir le plus grand écart moyen.

9. Niveau de difficulté : moyen

Plusieurs réponses sont possibles. Il faut toutefois porter une attention particulière au respect de tous les critères. Exemple :

- a) 4–6–8–10–10–10
 b) 14–14–16–16–20–20–24–25–25–26
 c) 1–9–9–9–10–12–13–17

10. Niveau de difficulté : moyen

La moyenne de cette distribution est de 15 et l'écart moyen est de 3.

En ajoutant 12 et 18, la moyenne et l'écart moyen resteront les mêmes.

11. Niveau de difficulté : moyen

- a) Cet énoncé est faux. Voici une distribution de données : 2–3–5–5–7–8.

La moyenne est de 5 et l'écart moyen est de 1,6. Si j'ajoute une donnée égale à la moyenne soit 5, la moyenne reste 5 mais l'écart moyen diminue et devient 1,4.

- b) Cet énoncé est vrai.
 c) Cet énoncé est vrai.

Manuel • p. 66

12. Niveau de difficulté : faible

- a) Il y a 5 raisins en moyenne par biscuit.
 b) L'écart moyen est d'environ 0,67 raisin par biscuit.

13. Niveau de difficulté : faible

- a) 33
 b) 50
 c) 69

14. Niveau de difficulté : faible

- a) ① 25
 ② 66
 ③ 89
 b) ① La donnée 79 occupe le rang centile 78.
 ② La donnée 29 occupe le rang centile 78.
 ③ La donnée 56 occupe le rang centile 78.

15. Niveau de difficulté : moyen

1^{re} course : rang centile 83
 2^e course : rang centile 79
 3^e course : rang centile 89
 4^e course : rang centile 86
 Classement de Mélanie du moins bon classement au meilleur classement : 2^e course, 1^{re} course, 4^e course et 3^e course

16. Niveau de difficulté : moyen

- a) 8 personnes ont obtenu une note égale à celle de Patrick et 98 ont obtenu une note inférieure. Donc, le rang centile de Patrick est 66.

- b) On détermine le nombre de personnes qui seront refusées :
- $$\frac{40}{100} \cdot 163 = 65,2$$
- 98 personnes seront donc admises ($163 - 65 = 98$).

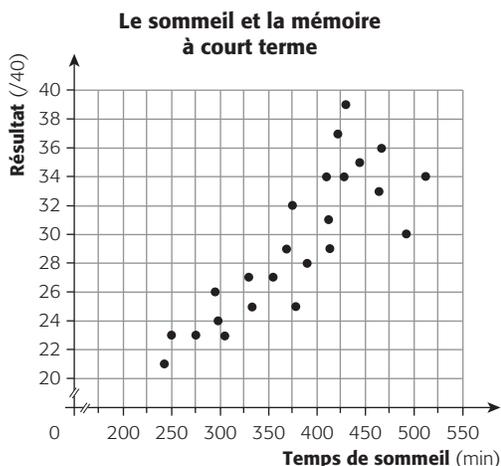
Section 2 L'appréciation qualitative d'une corrélation

Les conséquences du manque de sommeil



Manuel • p. 67

Voici la représentation graphique des résultats de cette étude :



D'après ce diagramme, on peut observer les éléments suivants.

- Il y a un lien entre le temps de sommeil et la mémoire à court terme.
- En général, plus les personnes dorment, plus elles obtiennent de bons résultats au test.
- Les meilleurs résultats au test de mémoire ont été obtenus après une période de sommeil de 420 à 470 minutes, soit après 7 à 8 heures de sommeil. Après 8 heures de sommeil, les résultats semblent tendre à diminuer. Il faut toutefois être prudent dans cette analyse puisqu'on n'a aucun résultat pour une période de sommeil de plus de 512 minutes, soit environ 8,5 heures, et qu'on n'a que deux résultats pour la période allant de 8 à 8,5 heures de sommeil.
- Les individus ayant obtenu les résultats moindres ont dormi généralement moins de 300 minutes, soit moins de 6 heures. C'est dans cet intervalle de temps que le lien entre la durée de la période de sommeil et la mémoire semble être le plus fort.

Les résultats de l'expérience permettent d'émettre les recommandations suivantes.

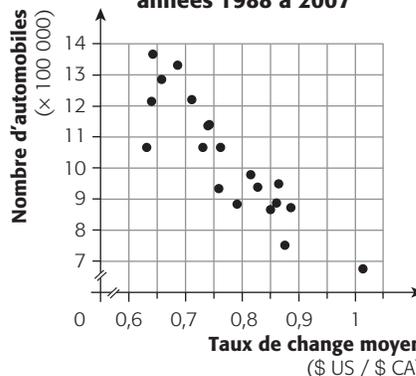
- On recommande une période de sommeil d'au moins 300 minutes. Les résultats ne permettent cependant pas de déterminer une durée de sommeil optimale.
- Puisque le manque de sommeil semble avoir un impact sur la mémoire à court terme, passer une nuit blanche la veille d'un examen pourrait réduire la capacité à se rappeler les notions récemment apprises. Il serait donc préférable de l'éviter.

ACTIVITÉ D'EXPLORATION ① La force du huard

Manuel • p. 68

- A** C'est un lien entre deux éléments.
- B** 1) Chaque dernier trimestre des années 1988 à 2007.
2) Les caractères étudiés sont le taux de change moyen (\$ US/\$ CA) et le nombre d'automobiles entrées au Canada en provenance des États-Unis.

C **Nombre d'automobiles entrées au Canada en provenance des États-Unis au cours du dernier trimestre des années 1988 à 2007**



Pour la période étudiée, lorsque le taux de change (\$ US/\$ CA) augmente, le nombre d'automobiles entrées au Canada en provenance des États-Unis diminue.

- D** Environ 750 000 automobiles en provenance des États-Unis entreraient au Canada.

Manuel • p. 69

- E** La corrélation tracée en C est négative.
- F** Plusieurs réponses sont possibles. *Exemple :*
- 1) Les variables ④ et ⑦ et les variables ⑤ et ⑥.
 - 2) Les variables ② et ⑤ et les variables ② et ⑧.
- G** Plusieurs réponses sont possibles. *Exemple :*

Pour les variables ④ et ⑦ : chaque touriste risque de dépenser de l'argent lorsqu'il vient en visite au Canada ; ainsi, plus le nombre de touristes augmente, plus les revenus du tourisme augmentent.

Pour les variables ⑤ et ⑥ : les gens ont plus envie de se rafraîchir et de se baigner lorsqu'il fait chaud ; ainsi, plus la température augmente et plus les ventes de piscines risquent d'augmenter.

Pour les variables ② et ⑤ : les gens qui ont perdu leur travail ont moins d'argent à dépenser pour des produits récréatifs ou de luxe ; ainsi, plus le taux de chômage augmente, plus les ventes de piscines diminuent.

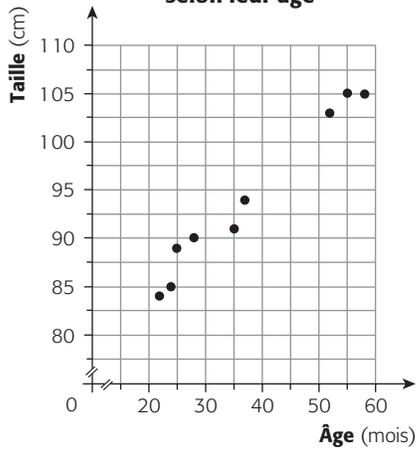
Pour les variables ② et ⑧ : les gens qui ont perdu leur travail ont moins d'argent à dépenser pour des produits récréatifs ou de luxe ; ainsi, plus le taux de chômage augmente, plus les ventes de roulottes diminuent.

H Plusieurs réponses sont possibles. *Exemple :*

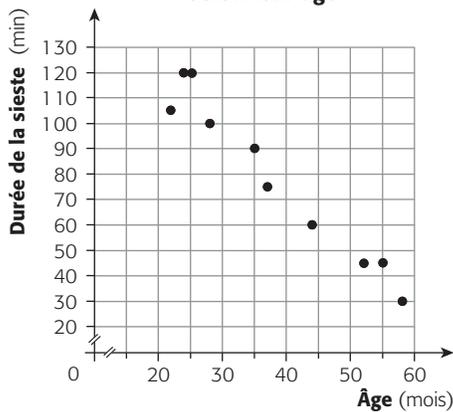
Les variables ③ et ⑥.

Ai-je bien compris?

a) **La taille de dix enfants d'une garderie selon leur âge**



Durée de la sieste de dix enfants d'une garderie selon leur âge



Nombre d'enfants dans la famille de dix enfants d'une garderie selon leur âge



- b) La corrélation entre l'âge et la taille est positive. La corrélation entre l'âge et la durée de la sieste est négative. La corrélation entre l'âge et le nombre d'enfants dans la famille est nulle.

ACTIVITÉ

D'EXPLORATION

② Skier dans les nuages

Manuel • p. 70

- A 1)** Nuage ① : Forme ovale
 Nuage ② : Forme très près d'une droite
 Nuage ③ : Forme plutôt ronde
 Nuage ④ : Forme d'une ligne droite
 Nuage ⑤ : Forme allongée ovale
 Nuage ⑥ : Forme ovale
- 2)** Nuage ① : Moyenne
 Nuage ② : Cette corrélation n'est pas linéaire.
 Nuage ③ : Nulle
 Nuage ④ : Parfaite
 Nuage ⑤ : Forte
 Nuage ⑥ : Moyenne
- 3)** Nuage ① : Positive
 Nuage ② : Positive
 Nuage ③ : Ne s'applique pas (corrélation linéaire nulle)
 Nuage ④ : Négative
 Nuage ⑤ : Négative
 Nuage ⑥ : Négative

B Non, il est possible qu'une corrélation négative soit plus forte qu'une corrélation positive, comme c'est le cas pour les nuages de points ④ et ①. Le sens d'une corrélation n'influe pas sur son intensité.

- C** Nuage ①: Lien causal
 Nuage ②: Lien fortuit
 Nuage ③: Ne s'applique pas (absence de corrélation linéaire)
 Nuage ④: Lien causal ou troisième facteur: conditions d'enneigement
 Nuage ⑤: Lien causal
 Nuage ⑥: Lien causal

Manuel • p. 71

- D** Nuage ①: Précipitations de neige
 Nuage ④: Nombre de journées où la station de ski est fermée
 Nuage ⑤: Nombre de planchistes membres ou nombre de skieurs membres
 Nuage ⑥: Précipitations de neige

Ai-je bien compris?

- ⑥-①-③-⑤-②-④
- (voir au bas de la page)

Mise en pratique

Manuel • p. 75

1. Niveau de difficulté: moyen

- a) On doit déterminer d'abord la population et l'individu statistique pour pouvoir ensuite identifier plus facilement les caractères étudiés.

Population: Les familles des élèves.
 Individu: Chacune des familles.
 Caractère étudié: Nombre d'enfants.
 Un seul caractère.

- b) On doit déterminer d'abord la population et l'individu statistique pour pouvoir ensuite identifier plus facilement les caractères étudiés.

Population: Les courses d'Adrian
 Individu: Chacune des courses
 Caractère étudié: Le temps d'une course
 Un seul caractère

- c) On doit déterminer d'abord la population et l'individu statistique pour pouvoir ensuite identifier plus facilement les caractères étudiés.

Population: Les moments de la journée.
 Individu: Chacune des mesures de temps
 Caractère étudié: La température
 Un seul caractère

2. Niveau de difficulté: faible

- ① a) Unité statistique: Chacun des petits-enfants de grand-maman Rachel

Caractères étudiés: La taille et l'âge

- b) Corrélation positive

- ② a) Unité statistique: Chacune des années de 1992 à 2007

Caractères étudiés: Nombre d'accidents mortels sur les routes de Jacksonville et prix de l'essence

- b) Corrélation nulle

- ③ a) Unité statistique: Chacun des essais routiers d'une voiture prototype

Caractères étudiés: La consommation de carburant et la distance parcourue

- b) Corrélation positive

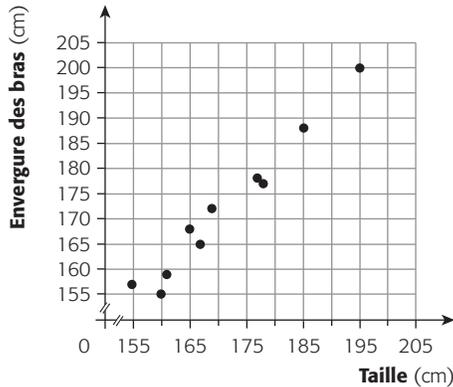
Réponse à la question 2, page 71

| Situation | a) Sens | b) Intensité | c) Nature du lien | d) Variable qui cause les variations de l'autre variable |
|-----------|----------------------|------------------|---|--|
| ① | Corrélation positive | Forte | Causal | Nombre de téléspectateurs |
| ② | Corrélation positive | Moyenne ou forte | Fortuit | — |
| ③ | Corrélation positive | Moyenne | 3 ^e facteur (nombre de personnes sur la plage) | — |
| ④ | Corrélation négative | Moyenne | Causal | Âge des enfants |

- ④ a) Unité statistique : Chacun des étudiants employés d'un restaurant à service rapide
Caractères étudiés : La moyenne générale à l'école et le nombre d'heures travaillées
- b) Corrélation négative

3. Niveau de difficulté : faible

a) Les joueurs du club Astéris

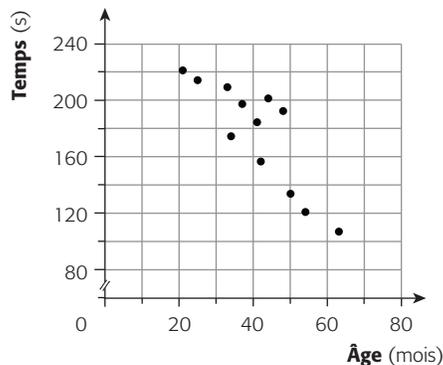


- b) La corrélation est positive et forte. Le lien est causal.

Manuel • p. 76

4. Niveau de difficulté : faible

a) Relation entre l'âge et le temps requis pour assembler un casse-tête



- b) La corrélation est négative, d'intensité forte. Le lien est causal. Plus les enfants sont âgés, moins ils mettent de temps à assembler le casse-tête.
- c) Environ 168 secondes.

5. Niveau de difficulté : faible

- a) ① et ③, ① et ⑥, ③ et ⑧, ⑥ et ⑧, ① et ⑤
- b) ① et ⑧, ③ et ⑥, ③ et ⑦, ④ et ⑧, ⑤ et ⑥, ⑤ et ⑦, ⑦ et ⑧, ④ et ⑦
- c) ① et ②, ① et ④, ① et ⑦, ② et ③, ② et ④, ② et ⑤, ② et ⑥, ② et ⑦, ② et ⑧, ③ et ④, ③ et ⑤, ④ et ⑤, ④ et ⑥, ⑤ et ⑥, ⑤ et ⑧, ⑥ et ⑦

6. Niveau de difficulté : faible

- a) Corrélation négative f) Corrélation positive
b) Corrélation positive g) Corrélation négative
c) Corrélation positive h) Corrélation positive
d) Corrélation négative i) Corrélation positive
e) Corrélation positive

Manuel • p. 77

7. Niveau de difficulté : moyen

(voir au haut de la page suivante)

8. Niveau de difficulté : faible

Plusieurs réponses sont possibles. Exemple :

- a) La taille et la masse des élèves d'une école.
b) L'altitude et la température ambiante.
c) La taille des élèves et leurs résultats scolaires.

9. Niveau de difficulté : faible

- ②, ④, ⑥, ⑤, ①, ③

10. Niveau de difficulté : faible

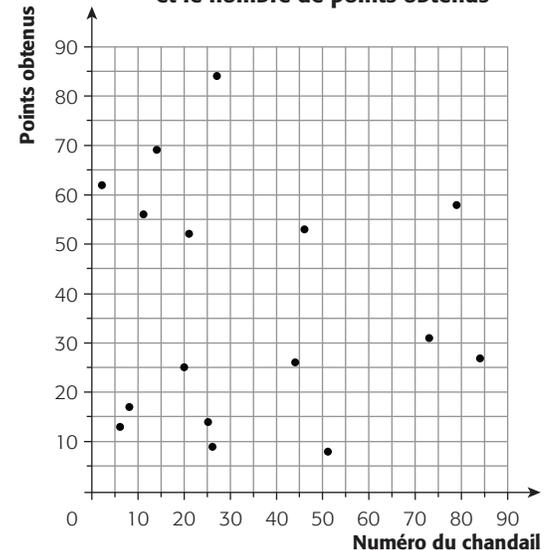
- a) Corrélation moyenne, négative
b) Corrélation forte, positive
c) Corrélation parfaite, négative
d) Corrélation faible, positive

Manuel • p. 78

11. Niveau de difficulté : faible

- a) et b)

Relation entre le numéro du chandail des joueurs et le nombre de points obtenus



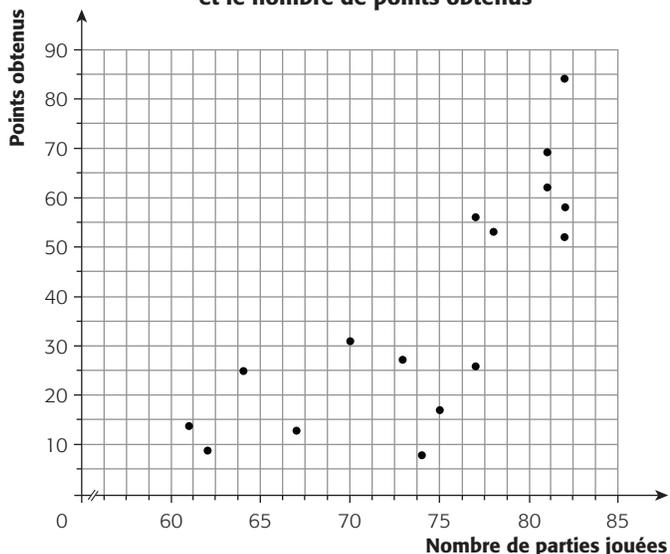
Corrélation nulle.

Réponses à la question 7, page 77

Plusieurs réponses sont possibles. *Exemple :*

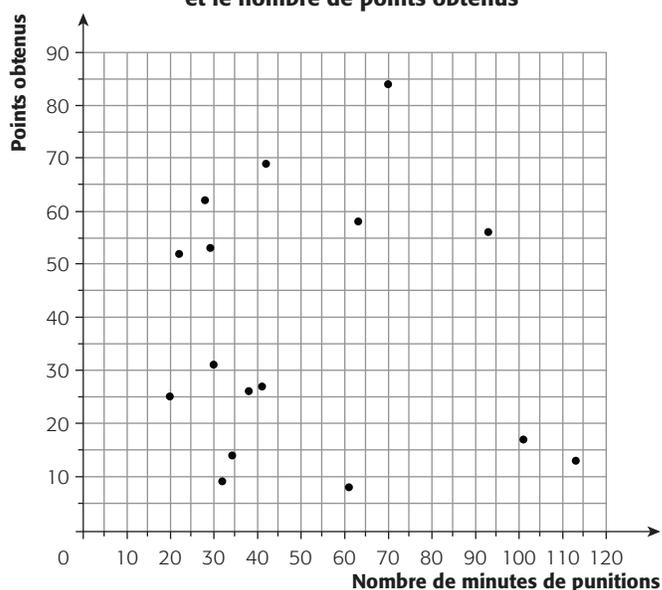
| | Intensité | Nature du lien |
|----|---|--|
| a) | Faible ou moyenne. Le temps passé à regarder la télévision n'est certes plus disponible pour la pratique d'un sport, mais le sport peut aussi être pratiqué à d'autres moments. | Lien causal. Plusieurs élèves ne prennent pas le temps de faire du sport parce qu'ils passent une grande partie de leur temps devant le téléviseur. |
| b) | Forte. Un prix plus élevé devrait correspondre à une meilleure qualité de fabrication, et donc à une plus grande durée de vie. | Lien dû à un 3 ^e facteur, soit la qualité des montres. |
| c) | Forte. De façon générale, les villes avec beaucoup de chiens ont aussi beaucoup de bornes fontaines et vice-versa. | Lien dû à un 3 ^e facteur, soit la taille de la population de la ville. |
| d) | Moyenne. Lorsque les taux d'intérêt sont plus bas, les gens peuvent se permettre de payer plus pour leur maison. Il y a cependant bien d'autres facteurs qui influent sur le prix des maisons. | Lien causal. Les baisses des taux d'intérêt font généralement monter les prix de vente des maisons, parce que les gens peuvent se permettre de payer un prix plus élevé. |
| e) | Forte. Dans la grande majorité des cas, les petits enfants ont peu de dents alors que les grands en ont plus. | Lien dû à un 3 ^e facteur, soit l'âge, qui influe à la fois sur la taille et le nombre de dents. |
| f) | Faible ou moyenne. Par le passé, les divorces ont eu tendance à augmenter au Québec, tout comme les prix, dont celui du pain. | Lien fortuit. Il n'y a pas de lien évident entre le nombre de divorces et le prix du pain. |
| g) | Forte. Même si le revenu augmente, le montant alloué au loyer augmente peu, ce qui fait diminuer son importance relative en pourcentage. | Lien causal. Lorsque les gens ont peu d'argent, une grande partie de leurs revenus doit être consacrée au paiement du loyer. |
| h) | Faible. De façon générale, les joueurs qui font beaucoup de circuits reçoivent un gros salaire. Cependant, certains bons joueurs font peu de circuits (par exemple, les lanceurs) et reçoivent aussi un gros salaire. | Lien causal. Les frappeurs de circuits sont très convoités par les équipes de la LNB, c'est pourquoi elles sont prêtes à offrir de gros salaires à ces joueurs. |
| i) | Forte. La majorité des jeunes du secondaire ont des amis et des amies du même niveau. De façon générale, les jeunes d'un même niveau sont de taille semblable. | Lien dû à un 3 ^e facteur, soit l'âge, qui influe sur la taille et le niveau des jeunes au secondaire. |

Relation entre le nombre de parties jouées et le nombre de points obtenus

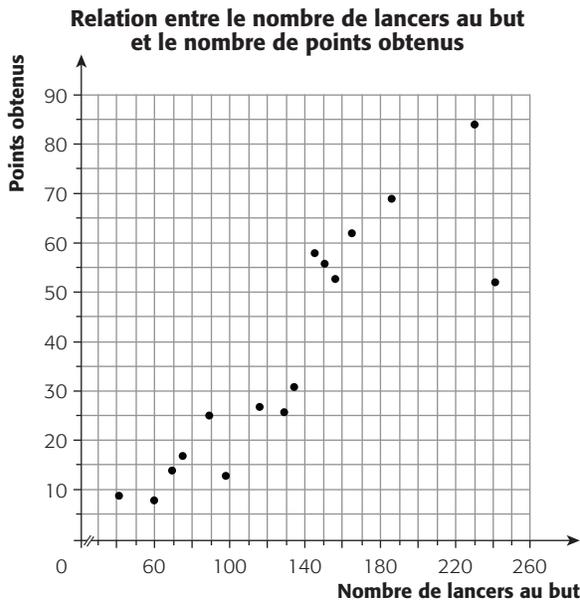


Corrélation positive et moyenne.

Relation entre le nombre de minutes de punitions et le nombre de points obtenus



Corrélation nulle.



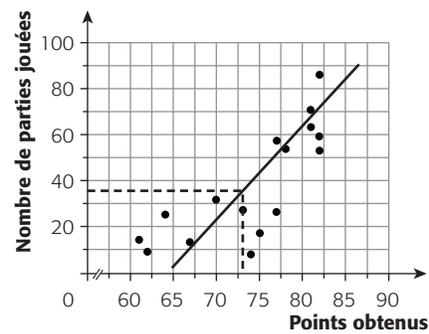
Corrélation positive et forte.

- c) 1) Le nombre de lancers au but
- 2) Comme deux relations présentent une corrélation nulle, il est impossible de déterminer laquelle de ces deux variables, le numéro du chandail ou le nombre de minutes de punition, a la plus faible corrélation avec le nombre de points obtenus.

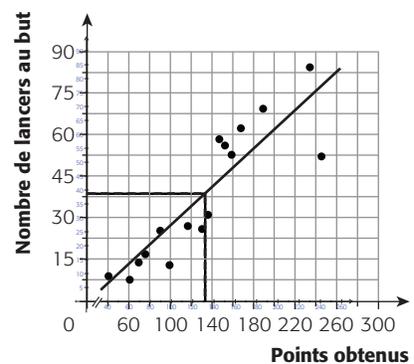
- d) Relation entre le nombre de parties jouées et le nombre de points obtenus: La variable «nombre de parties jouées» cause les variations de l'autre variable.
- Relation entre le nombre de lancers au but et le nombre de points obtenus: La variable «nombre de lancers au but» cause les variations de l'autre variable.

- e) On peut tracer la droite la mieux ajustée aux nuages de points faits en a). On se déplace ensuite le long de cette droite pour atteindre la valeur connue d'une variable, par exemple le nombre de parties jouées, et on détermine le nombre de points obtenus.

- 1) Comme la corrélation du premier nuage de points est nulle, il est impossible de déterminer le nombre de points d'après le numéro du chandail.
- 2) On trace la droite la mieux ajustée au nuage de points représentant la relation entre le nombre de parties jouées et le nombre de points obtenus. On trouve ainsi que, pour 73 parties jouées, le nombre de points obtenus par Matthieu Darce est environ 35.



- 3) On trace la droite la mieux ajustée au nuage de points représentant la relation entre le nombre de lancers au but et le nombre de points obtenus. On trouve ainsi que, pour 132 lancers au but, le nombre de points obtenus par Michel Ouellet est environ 39.



- 4) Comme la corrélation du troisième nuage de points est nulle, il est impossible de déterminer le nombre de points d'après le nombre de minutes de punition.
- f) Ce serait l'estimation 3) puisqu'il s'agit de la relation dont la corrélation est la plus forte.

Section 3 Le coefficient de corrélation linéaire

Le bonheur



Manuel • p. 79

En observant les trois premiers nuages de points représentés dans l'énoncé de la situation, on remarque que le sens de la corrélation linéaire est positif et que le coefficient de corrélation qui leur est associé est, lui aussi, positif. Dans le quatrième nuage de points, le sens de la corrélation est négatif et il en est de même pour le coefficient de corrélation qui lui est associé.

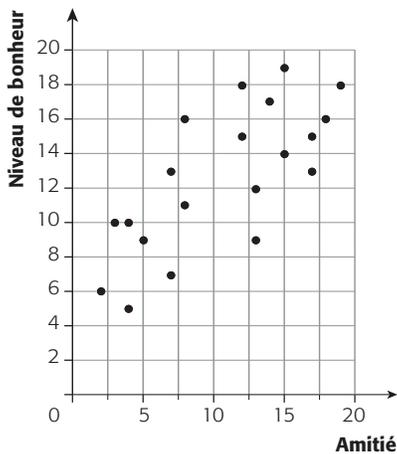
De plus, les trois premiers nuages de points sont disposés de telle sorte que l'intensité de la corrélation linéaire passe graduellement de faible à forte. Les coefficients de

corrélation qui leur sont associés passent de 0,46 à 0,92. On remarque que l'augmentation de l'intensité fait aussi augmenter le coefficient de corrélation.

Si on ordonnait les nuages de points en partant de celui dont l'intensité de la corrélation linéaire est la plus faible jusqu'à celui dont l'intensité de la corrélation linéaire est la plus forte, on arriverait à l'ordre suivant: le nuage 1, le nuage 4, le nuage 2 et le nuage 3. L'ordre des coefficients de corrélation qui leur sont respectivement associés est le suivant: 0,46, -0,64, 0,79 et 0,92. On peut conclure qu'un nuage de points dont l'intensité de la corrélation est nulle a un coefficient de corrélation se rapprochant de zéro.

On trace maintenant le nuage de points représentant le cinquième facteur susceptible d'influer sur le bonheur. Ce nuage de points a été tracé en se basant sur le tableau présenté dans l'énoncé de la situation.

L'étude de Mya sur le bonheur



En observant ce nuage de points, on peut conclure que le coefficient de corrélation est positif et qu'il se situe entre 0,46 et 0,79 (en le comparant aux nuages 1 et 2). De plus, si on le compare avec celui ayant un coefficient de corrélation de -0,64 (nuage 4), on remarque que l'intensité de ce cinquième nuage est plus forte que celle du quatrième nuage. On peut donc conclure que le coefficient de corrélation se situe entre 0,64 et 0,79. On estime donc que le coefficient de corrélation associé à cette distribution est d'environ 0,73.

En ce qui concerne la description des résultats obtenus par cette étude, on remarque que le facteur «santé» semble être celui qui a le plus d'influence sur le bonheur. À l'inverse, le facteur «satisfaction au travail» semble avoir peu d'influence sur le bonheur.

De plus, on peut critiquer les résultats de l'étude de Mya par rapport aux points suivants:

- L'échantillon ne comporte que 20 individus, et on ne sait pas comment ils ont été choisis.
- Il est difficile d'attribuer une «cote» au niveau de bonheur.

- D'autres facteurs peuvent influencer les résultats. Par exemple, l'âge des répondants. Un jeune adulte sans enfant peut répondre que les relations familiales sont peu importantes tandis qu'une jeune maman peut répondre le contraire.

ACTIVITÉ D'EXPLORATION ① Encadrement judicieux

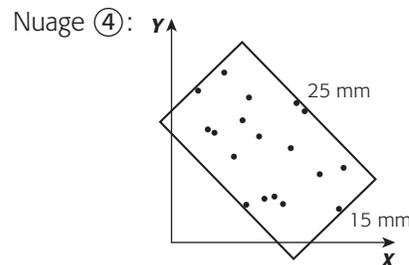
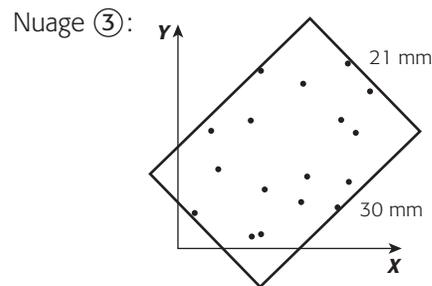
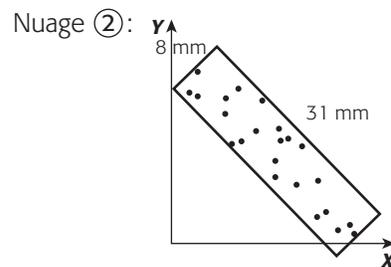
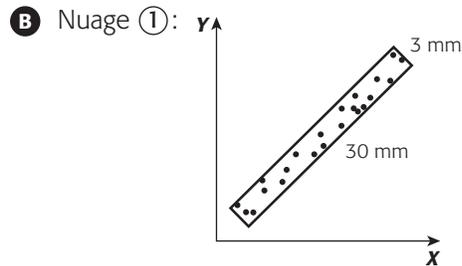
Manuel • p. 80

A Nuage ①: L'intensité est forte et le sens est positif.

Nuage ②: L'intensité est moyenne et le sens est négatif.

Nuage ③: L'intensité est très faible et le sens est positif.

Nuage ④: L'intensité est faible et le sens est négatif.

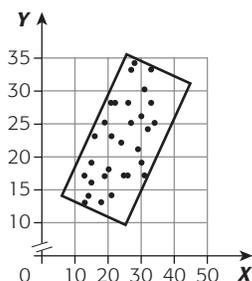


- C** Le coefficient r varie dans l'intervalle des valeurs $[-1, 1]$.
- D** Nuage ①: $r \approx 0,9$
- Nuage ②: $r \approx -0,8$
- Nuage ③: $r \approx 0,3$
- Nuage ④: $r \approx -0,4$
- E** Plus l'intensité de la corrélation linéaire est faible, plus la valeur de r se rapproche de 0. À l'inverse, plus l'intensité de la corrélation linéaire est forte, plus la valeur de r se rapproche de 1 ou de -1.

Manuel • p. 81

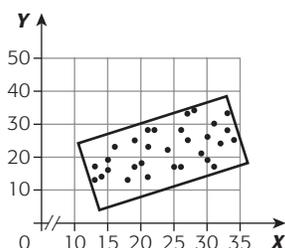
- F** Nuage ①:

$$r \approx +\left(1 - \frac{0,9}{2,1}\right) \approx 0,57$$



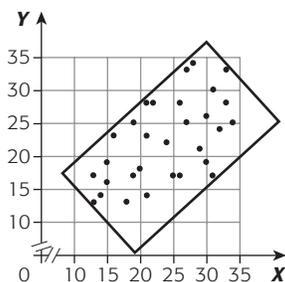
- Nuage ②:

$$r \approx +\left(1 - \frac{0,9}{2,1}\right) \approx 0,57$$



- Nuage ③:

$$r \approx +\left(1 - \frac{1,5}{2,6}\right) \approx 0,42$$



- G** Bien que les trois nuages de points représentent une même distribution, les coefficients de corrélation estimés sont légèrement différents! Cela est dû au fait qu'on n'a pas utilisé la même graduation des axes pour les trois représentations. L'estimation la plus fiable des trois est celle obtenue à partir du nuage ③. Il s'agit du seul nuage pour lequel l'étendue des points à l'horizontale et à la verticale est la même.

Ai-je bien compris?

- a) ⑤ d) ⑥

b) ④ e) ③

c) ② f) ①
- a) $r \approx 0,47$ b) $r \approx -0,64$ c) $r \approx -0,78$
- ②, ⑤, ④, ③, ①, ⑥

ACTIVITÉ

D'EXPLORATION ② Se méfier des apparences

Manuel • p. 82

- A** Situation ①: La corrélation entre les deux caractères est faible et positive.
- Situation ②: La corrélation entre les deux caractères est moyenne et positive.
- Situation ③: La corrélation entre les deux caractères est moyenne et positive.
- Situation ④: La corrélation entre les deux caractères est moyenne et positive.
- B** Les coefficients de corrélation pour chacune des situations sont respectivement 0,28, 0,57, 0,57 et 0,59.

Manuel • p. 83

- C** Situation ①: Non. Ce nuage de points présente deux points aberrants. Il s'agit probablement de joueurs blessés qui n'ont pas joué au cours de la dernière semaine. Si on enlève ces deux points aberrants, le lien entre les deux caractères est moyen.
- Situation ②: Non. On remarque que la relation entre l'âge et la moyenne au bâton n'est pas linéaire. À partir d'un certain âge, plus un joueur vieillit, plus sa moyenne au bâton diminue.
- Situation ③: Non. En supposant qu'il y ait des jeunes de tous les niveaux du primaire qui participent au défi, un troisième facteur influe alors sur les deux autres, c'est celui de l'âge. La corrélation entre la mesure des pieds et les paniers réussis est très grande chez les personnes qui ont les plus petits pieds (qui sont donc théoriquement plus jeunes), mais on remarque que la corrélation s'affaiblit pour les personnes qui ont de plus grands pieds (qui sont donc théoriquement plus âgées).
- Situation ④: Non. Ce nuage de points présente un point aberrant. Ce point semble représenter un joueur de l'équipe qui a un salaire beaucoup plus élevé que les autres et qui consacre beaucoup plus de temps que les autres à une ou des causes humanitaires. Le point aberrant influe sur la corrélation, mais puisqu'il ne s'agit pas d'une anomalie, on ne peut pas l'exclure.

D Les nuages de points ① et ④ présentent des points aberrants et on devra les exclure s'il s'agit d'anomalies. Pour le nuage ①, il semble s'agir de deux joueurs blessés, donc, de cas d'exception. En ce qui concerne le nuage ④, on peut prétendre qu'il s'agit d'un joueur vedette qui est très en demande. Il s'agit ici d'un cas exceptionnel, mais pas d'une anomalie.

E Plusieurs réponses sont possibles. *Exemple :*

Situation ① : Il existe un lien assez fort entre le salaire d'un joueur de hockey et son temps de glace au cours d'une semaine.

Situation ② : Au-dessous d'un certain âge, plus un joueur de baseball vieillit, plus sa moyenne au bâton augmente. À partir d'un certain âge, plus un joueur vieillit, plus sa moyenne au bâton diminue.

Situation ③ : Pour les élèves d'une école primaire, la corrélation entre la mesure des pieds et le pourcentage de paniers réussis est très grande chez les personnes qui ont les plus petits pieds. Par contre, la corrélation s'affaiblit pour les personnes qui ont de plus grands pieds.

Situation ④ : À l'exception d'un joueur, il n'y a pas de lien évident entre le salaire des joueurs de soccer et le temps consacré à une cause humanitaire.

F L'affirmation ② est la plus juste. En effet, pour bien qualifier la corrélation linéaire entre deux caractères, il est nécessaire d'observer le nuage de points, car celui-ci permet, entre autres, de révéler la présence de points aberrants (s'il y a lieu) ou de voir quelle est la forme du nuage de points.

Ai-je bien compris?

- a) $r \approx 0,6$
 b) La corrélation est positive et d'intensité moyenne. Cependant, le nuage de points révèle la présence de deux sous-populations, les enfants et les adultes, à l'intérieur desquelles la corrélation est nulle.

Mise en pratique

Manuel • p. 86

1. Niveau de difficulté : faible

- a) ④ d) ⑨ g) ①
 b) ⑤ e) ③ h) ⑥
 c) ⑦ f) ⑧ i) ②

2. Niveau de difficulté : faible

- ④, ③, ①, ②, ⑤, ⑥

3. Niveau de difficulté : faible

- a) $r \approx -0,57$
 b) $r \approx 0,45$
 c) $r \approx 0,86$

Manuel • p. 87

4. Niveau de difficulté : moyen

Plusieurs réponses sont possibles. *Exemples :*

- a) On détermine d'abord les dimensions possibles d'un nuage de points qui possède un coefficient de corrélation de $-0,95$, par exemple.

$$r \approx \pm \left(1 - \frac{\text{mesure du petit côté}}{\text{mesure du grand côté}} \right)$$

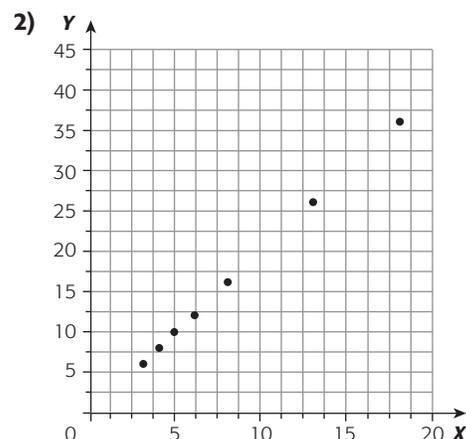
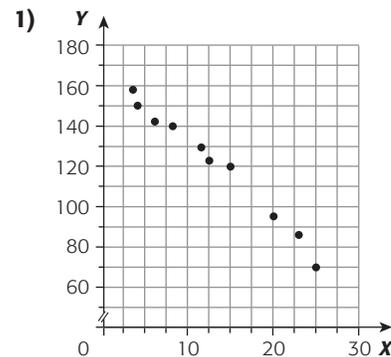
$$0,95 = 1 - \frac{\text{mesure du petit côté } (x)}{\text{mesure du grand côté } (y)}$$

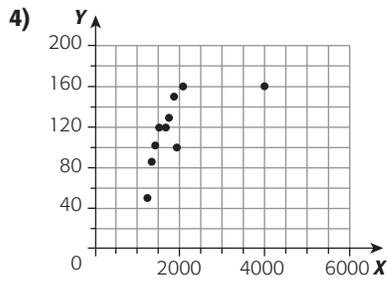
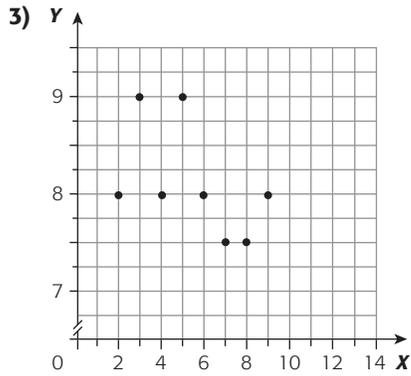
$$-0,05 = \frac{-x}{y}$$

$$0,05 = \frac{x}{y}$$

Si x égale 5 mm, alors y vaut 100 mm.

On trace le nuage de points de façon à respecter les dimensions établies, puis on répète la même démarche pour chaque coefficient de corrélation linéaire.





b) Plusieurs réponses sont possibles. *Exemple :*

1) **Le prêt sans intérêts de 300 \$**

| Le nombre de remboursements de 20 \$ en moyenne effectués | Le montant qu'il reste à rembourser |
|---|-------------------------------------|
| 5 | 200 |
| 10 | 110 |
| 6 | 180 |
| 2 | 260 |
| 13 | 35 |
| 4 | 220 |
| 8 | 145 |
| 9 | 120 |
| 11 | 75 |
| 3 | 240 |

2) **Achat de reliures à anneaux à la papeterie**

| Nombre de reliures à anneaux | Prix payé (\$) |
|------------------------------|----------------|
| 3 | 6 |
| 4 | 8 |
| 6 | 12 |
| 5 | 10 |
| 8 | 16 |
| 13 | 26 |
| 18 | 36 |

3) **Les chiens du quartier**

| Âge du chien | Âge de son propriétaire |
|--------------|-------------------------|
| 2 | 45 |
| 6 | 36 |
| 9 | 42 |
| 1 | 23 |
| 4 | 63 |
| 2 | 55 |
| 7 | 53 |
| 7 | 70 |
| 3 | 18 |
| 5 | 21 |

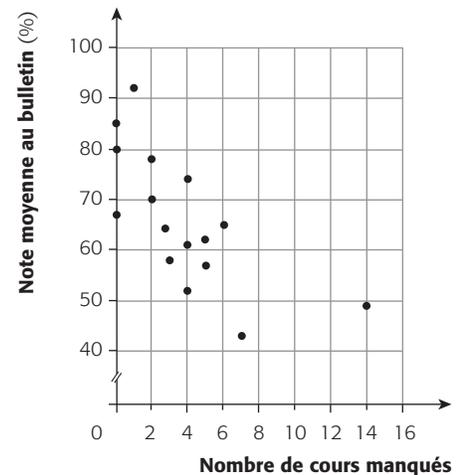
4) Plusieurs réponses sont possibles. *Exemple :*

L'accumulation de déchets par mois dans plusieurs municipalités

| Nombre d'habitants | Nombre de tonnes de déchets |
|--------------------|-----------------------------|
| 1 200 | 50 |
| 1 320 | 86 |
| 1 400 | 102 |
| 1 500 | 120 |
| 1 650 | 120 |
| 1 726 | 130 |
| 1 845 | 150 |
| 1 900 | 100 |
| 2 045 | 160 |
| 4 000 | 160 |

5. Niveau de difficulté : faible

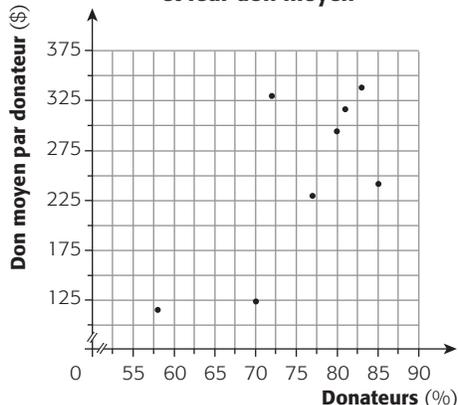
a) La relation entre le nombre de cours manqués et la note moyenne au bulletin



- b) La corrélation est négative et d'intensité moyenne.
- c) $r \approx 0,58$
- d) Si un étudiant a manqué 8 cours, sa note devrait se situer entre 38 % et 70 %.

6. Niveau de difficulté : faible

a) La relation entre le pourcentage de donateurs et leur don moyen



- b) $r \approx 0,55$
- c) Plus le pourcentage de donateurs est élevé, plus le don moyen par donateur augmente.
- d) Étant donné que le coefficient de corrélation aurait baissé à environ 0,37, il aurait été difficile de tirer une conclusion valable.
- e) Il vaut mieux en tenir compte, puisqu'il ne s'agit pas d'une anomalie.

Manuel • p. 88

7. Niveau de difficulté : moyen

- a) La relation n'est pas linéaire.
- b) Le prix du billet n'est pas une variable mais une valeur fixe.
- c) Le nombre de points est insuffisant.
- d) La valeur du coefficient de corrélation sera très affectée par le point aberrant.

8. Niveau de difficulté : moyen

- a) 1) 0,16 2) 0,64
- b) Plusieurs réponses sont possibles. *Exemples :*
 Au point A, la température était fraîche, mais la crèmerie a reçu la visite de trois autobus d'écoliers.
 Au point B, la crèmerie était fermée ce jour-là.
- c) La corrélation est positive et d'intensité moyenne.

9. Niveau de difficulté : faible

- a) La corrélation entre l'obésité et la consommation de fruits et de légumes est plus forte qu'entre l'obésité et l'activité physique.
- b) Les Québécois consomment plus de fruits et de légumes que les autres Canadiens mais, en ce qui concerne l'activité physique, ils en font un peu moins que la moyenne canadienne.
- c) Si on tient compte uniquement des deux facteurs considérés, on peut dire qu'ils ont de meilleures habitudes de vie que les habitants des autres provinces. Leur pourcentage d'obésité est aussi beaucoup plus bas que dans toutes les autres provinces canadiennes.
- d) Bien que les données de la Colombie-Britannique soient à l'écart des autres, on doit en tenir compte, car c'est une province au même titre que les autres.

Section 4 La droite de régression



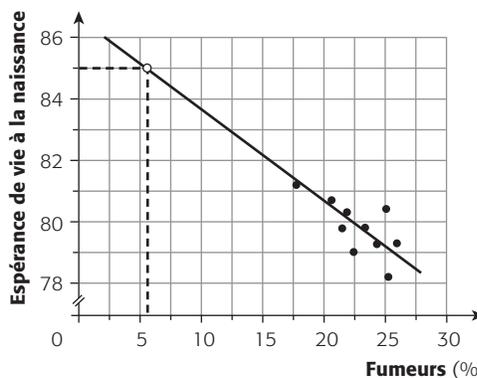
Les années s'envolent en fumée

Manuel • p. 89

Plusieurs réponses sont possibles. *Exemple :*

D'abord, on trace le nuage de points et, ensuite, la droite la mieux ajustée à ce nuage de points.

La relation entre l'espérance de vie et le pourcentage de fumeurs dans les provinces canadiennes



Par extrapolation, on estime que l'espérance de vie sera de 85 ans lorsqu'approximativement 5,8 % de la population âgée de plus de 12 ans sera constituée de fumeurs réguliers.

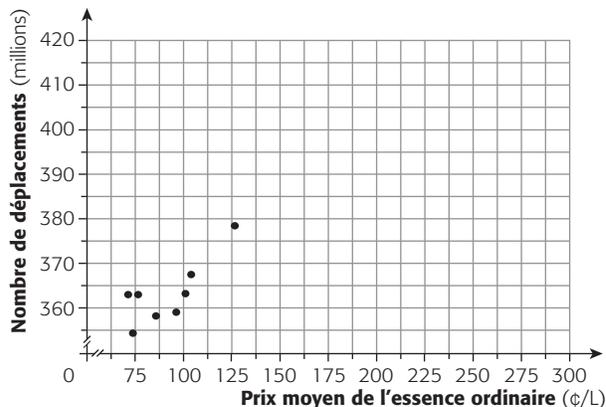
Au Québec, les fumeurs réguliers constituent 25,1 % de la population âgée de plus de 12 ans. Le pourcentage de fumeurs au Québec devrait donc diminuer d'environ 19,3 % pour que l'espérance de vie soit de 85 ans.

ACTIVITÉ D'EXPLORATION

1 Se déplacer autrement

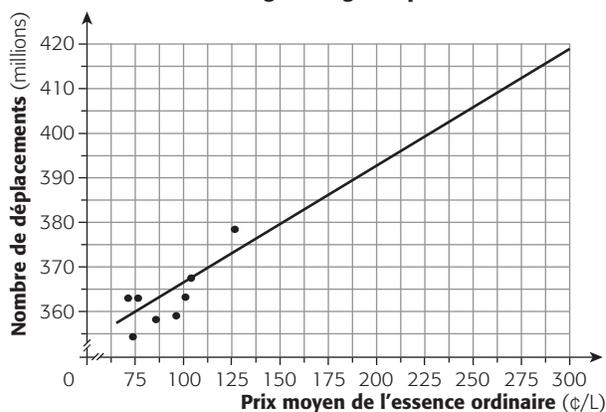
Manuel • p. 90

A L'achalandage enregistré par la STM



B Oui

C L'achalandage enregistré par la STM



D Si le prix de l'essence est 3 \$ le litre, cela veut dire 300 ¢ le litre. Sur le graphique, on regarde sur la droite de régression ce que vaut Y lorsque X vaut 300.

Manuel • p. 91

E Groupe 1 :

Moyenne des abscisses :

$$\frac{71,4 + 73,8 + 76,7 + 85,8}{4} \approx 76,9$$

Moyenne des ordonnées :

$$\frac{363,2 + 354,9 + 363,2 + 358,4}{4} \approx 359,9$$

Groupe 2 :

Moyenne des abscisses :

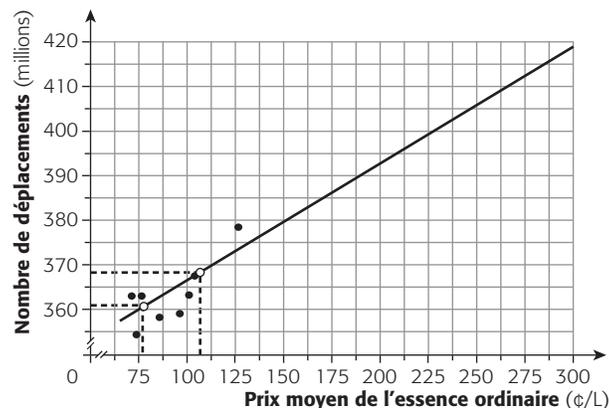
$$\frac{96,4 + 100,8 + 104,3 + 126,5}{4} = 107$$

Moyenne des ordonnées :

$$\frac{359,3 + 363,3 + 367,5 + 378,5}{4} \approx 367,2$$

Les points P_1 et P_2 sont respectivement (76,9, 359,9) et (107, 367,2). Voici leur représentation dans le nuage de points ainsi que la droite passant par ces deux points :

L'achalandage enregistré par la STM



F On détermine le taux de variation :

$$\frac{367,2 - 359,9}{107 - 76,9} \approx \frac{7,3}{30,1} \approx 0,24$$

Le taux de variation est d'environ 0,24. Dans cette situation, cela signifie qu'il y a une augmentation d'environ 24 000 déplacements à la STM pour chaque augmentation de un cent du prix de l'essence.

G Le taux de variation de la droite de Mayer est d'environ 0,24. On détermine la valeur de l'ordonnée à l'origine à l'aide des coordonnées du point $M_2(107, 367,2)$:

$$Y \approx 0,24X + b$$

$$367,2 \approx 0,24(107) + b$$

$$341,5 \approx b$$

L'équation de la droite de Mayer

$$\text{est } Y \approx 0,24X + 341,5.$$

H 1) Si le prix de l'essence est de 1,15 \$ le litre, cela veut dire 115 ¢ le litre.

$$Y \approx 0,24X + 341,5$$

$$Y \approx 0,24(115) + 341,5$$

$$Y \approx 369,1$$

On estime que l'achalandage annuel de la STM sera d'environ 369,1 millions de déplacements si le prix moyen d'un litre d'essence ordinaire est de 1,15 \$.

2) Si le prix de l'essence est 3 \$ le litre, cela veut dire 300 ¢ le litre.

$$Y \approx 0,24X + 341,5$$

$$Y \approx 0,24(300) + 341,5$$

$$Y \approx 413,5$$

On estime que l'achalandage annuel de la STM sera d'environ 413,5 millions de déplacements si le prix moyen d'un litre d'essence ordinaire est de 3 \$.

- I** C'est la prédiction par interpolation qui est la plus fiable, car rien ne garantit que la situation continuera de progresser de façon linéaire en dehors de l'intervalle étudié.

Ai-je bien compris?

a) $Y = \frac{3}{4}X + 6$

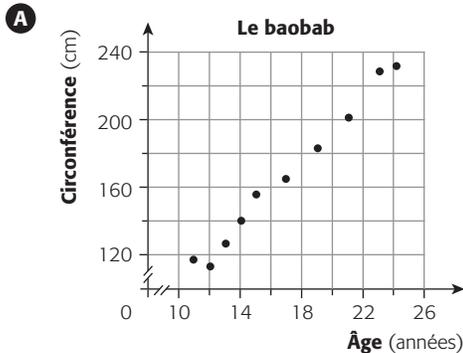
b) 1) $Y = 21$ 2) $X = 8$

ACTIVITÉ

D'EXPLORATION

2 Faire le tour du baobab

Manuel • p. 92



B $M_1(12, 118)$; $M_2(16, 161)$; $M_3(23, 228)$.

C (voir la représentation ci-dessous, en F)

D Moyenne des abscisses: $\frac{12 + 16 + 23}{3} = 17$

Moyenne des ordonnées: $\frac{118 + 161 + 228}{3} = 169$

Les coordonnées du point P sont (17, 169).
(voir la représentation ci-dessous, en F)

Manuel • p. 93

E Le taux de variation est $\frac{228 - 118}{23 - 12} = \frac{110}{11} = 10$.

Le taux de variation de la droite médiane-médiane est 10 et cette droite passe par P (17, 169):

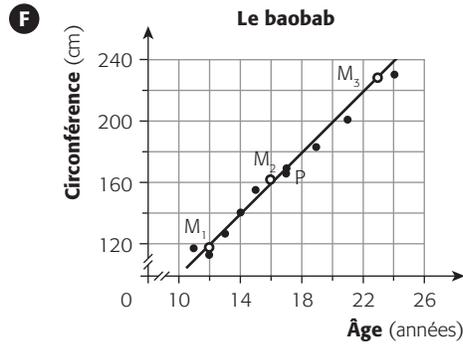
$$Y = 10X + b$$

$$169 = 10(17) + b$$

$$169 = 170 + b$$

$$-1 = b$$

L'équation de la droite médiane-médiane est $Y = 10X - 1$.



G 1) $Y = 10X - 1$

$$175 = 10X - 1$$

$$176 = 10X$$

$$17,6 = X$$

On estime qu'un baobab dont la circonférence du tronc est de 175 cm est âgé de près de 18 ans.

2) $Y = 10X - 1$

$$Y = 10(40) - 1$$

$$Y = 400 - 1$$

$$Y = 399$$

On estime qu'un baobab âgé de 40 ans a une circonférence d'environ 400 cm.

- H** Non. On estime que le plus gros baobab du monde est âgé d'environ 6 000 ans. Ce n'est pas un arbre «jeune». Ici, l'étude est basée sur 10 «jeunes» baobabs. Rien ne garantit que la situation continuera de progresser de façon linéaire en dehors de l'intervalle étudié. D'ailleurs, si la droite médiane-médiane était valable pour tous les baobabs, voici le résultat qu'on obtiendrait en vérifiant l'âge d'un baobab dont la circonférence est de 40 m (ou 4 000 cm):

$$Y = 10X - 1$$

$$4\ 000 = 10X - 1$$

$$4\ 001 = 10X$$

$$400,1 = X$$

On arriverait à environ 400 ans. Cette prédiction est loin de représenter la réalité, soit 6 000 ans.

Ai-je bien compris?

1. a) 1^{er} groupe: 6 données

2^e groupe: 7 données

3^e groupe: 6 données

b) 3 groupes de 12 données chacun

c) 1^{er} groupe: 14 données

2^e groupe: 13 données

3^e groupe: 14 données

2. a) $Y = -\frac{1}{2}X + \frac{99}{2}$ ou $Y = -0,5X + 49,5$

b) 1) $Y = 39,5$

2) $X = 75$

Mise en pratique

Manuel • p. 97

1. Niveau de difficulté : moyen

a) Situation ①

À partir des points moyens, on trouve le taux de variation :

$$\frac{28 - 13}{32 - 12} = \frac{15}{20} = \frac{3}{4}$$

On trouve ensuite l'ordonnée à l'origine :

$$Y = \frac{3}{4}X + b$$

$$13 = \frac{3}{4} \cdot 12 + b$$

$$13 = 9 + b$$

$$4 = b$$

L'équation de la droite de Mayer

$$\text{est } Y = \frac{3}{4}X + 4.$$

Situation ②

À partir des points moyens, on trouve le taux de variation :

$$\frac{27 - 14}{9 - 29} = \frac{13}{-20} = -\frac{13}{20}$$

On trouve ensuite l'ordonnée à l'origine :

$$Y = -\frac{13}{20}X + b$$

$$27 = -\frac{13 \cdot 9}{20} + b$$

$$27 = -\frac{117}{20} + b$$

$$\frac{540}{20} = -\frac{117}{20} + b$$

$$\frac{657}{20} = b$$

L'équation de la droite de Mayer

$$\text{est } Y = -\frac{13}{20}X + \frac{657}{20}.$$

Situation ③

À partir des points moyens, on trouve le taux de variation :

$$\frac{40,1 - 13,1}{46 - 16} = \frac{27}{30} = \frac{9}{10}$$

On trouve ensuite l'ordonnée à l'origine :

$$Y = \frac{9}{10}X + b$$

$$13,1 = \frac{9 \cdot 16}{10} + b$$

$$13,1 = \frac{144}{10} + b$$

$$13,1 = 14,4 + b$$

$$-1,3 = b$$

L'équation de la droite de Mayer

$$\text{est } Y = \frac{9}{10}X - \frac{13}{10} \text{ ou } Y = 0,9X - 1,3.$$

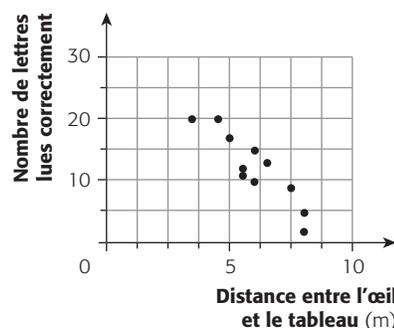
- b) 1) Situation ① : 19
 Situation ② : 19,85
 Situation ③ : 16,7
- 2) Situation ① : 37,75
 Situation ② : 3,6
 Situation ③ : 39,2

- c) L'estimation de Y lorsque $X = 20$ dans la situation ② est celle qui est la plus fiable car c'est une prédiction par interpolation dans un nuage de points où l'intensité de la corrélation entre les deux caractères est forte.

2. Niveau de difficulté : moyen

a)

Les résultats d'examens de la vue de 11 personnes



- b) On doit d'abord ordonner les données de façon croissante selon les valeurs de la variable indépendante, soit la distance.

| Distance entre l'œil et le tableau (m) | Nombre de lettres lues correctement |
|--|-------------------------------------|
| 3,5 | 20 |
| 4,5 | 20 |
| 5 | 17 |
| 5,5 | 11 |
| 5,5 | 12 |
| 6 | 10 |
| 6 | 15 |
| 6,5 | 13 |
| 7,5 | 9 |
| 8 | 2 |
| 8 | 5 |

On constitue ensuite deux groupes de données. Étant donné qu'il y a 11 couples, on prend les 6 premiers dont on trouve le point moyen et on trouve ensuite le point moyen des 6 derniers.

Premier point moyen :

$$\text{Valeur de } X: \frac{3,5 + 4,5 + 5 + 5,5 + 5,5 + 6}{6} = 5$$

$$\text{Valeur de } Y: \frac{20 + 20 + 17 + 11 + 12 + 10}{6} = 15$$

Donc, le premier point moyen est (5, 15).

Deuxième point moyen :

$$\text{Valeur de } X: \frac{6 + 6 + 6,5 + 7,5 + 8 + 8}{6} = 7$$

$$\text{Valeur de } Y: \frac{10 + 15 + 13 + 9 + 2 + 5}{6} = 9$$

Donc, le deuxième point moyen est (7, 9).

Il faut ensuite trouver le taux de variation entre ces deux points.

$$\frac{9 - 15}{7 - 5} = \frac{-6}{2} = -3$$

On détermine l'ordonnée à l'origine.

$$Y = -3 \cdot X + b$$

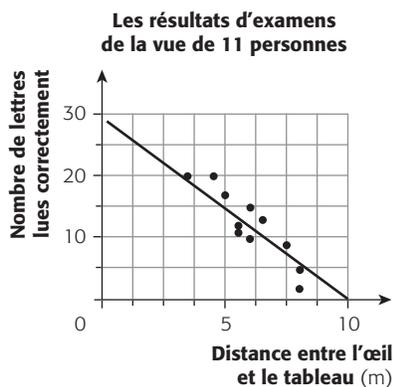
$$15 = -3 \cdot 5 + b$$

$$15 = -15 + b$$

$$30 = b$$

L'équation est donc $Y = -3X + 30$, où X représente la distance entre l'œil et le tableau, en mètres, et Y , le nombre de lettres lues correctement.

c)



d) 1) 18 2) 0

Manuel • p. 98

3. Niveau de difficulté : moyen

a) On doit d'abord ordonner les données de façon croissante selon les valeurs de la variable indépendante.

| Circonférence du poignet (cm) | Circonférence du cou (cm) |
|-------------------------------|---------------------------|
| 13 | 34 |
| 15 | 32,5 |
| 16,5 | 33 |
| 17 | 32 |
| 17 | 33,5 |
| 18 | 35 |
| 18,5 | 34,5 |
| 19 | 37 |
| 20,5 | 38 |
| 21,5 | 40,5 |
| 22,5 | 43 |

On constitue ensuite deux groupes de données. Étant donné qu'il y a 11 couples, on prend les 6 premiers dont on trouve le point moyen, puis on trouve le point moyen des 6 derniers.

Premier point moyen :

$$\text{Valeur de } X: \frac{13 + 15 + 16,5 + 17 + 17 + 18}{6} \approx 16,8$$

$$\text{Valeur de } Y: \frac{34 + 32,5 + 33 + 33,5 + 32 + 35}{6} \approx 33,33$$

Le premier point moyen est (16,08, 33,33).

Deuxième point moyen :

$$\text{Valeur de } X: \frac{18 + 18,5 + 19 + 20,5 + 21,5 + 22,5}{6} = 20$$

$$\text{Valeur de } Y: \frac{35 + 34,5 + 37 + 38 + 40,5 + 43}{6} = 38$$

Le deuxième point moyen est (20, 38).

On détermine le taux de variation entre ces deux points :

$$\frac{38 - 33,33}{20 - 16,08} \approx 1,19$$

On détermine la valeur de l'ordonnée à l'origine :

$$Y \approx 1,19X + b$$

$$38 \approx 1,19 \cdot 20 + b$$

$$b \approx 14,17$$

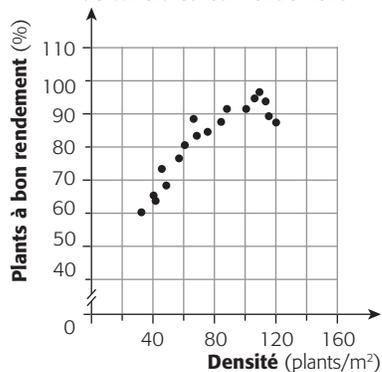
L'équation est donc $Y \approx 1,19X + 14,17$, où X représente la circonférence du poignet (en cm) et Y , la circonférence du cou (en cm).

b) 1) La circonférence du cou est d'environ 38 cm.

2) La circonférence du poignet est d'environ 13,3 cm.

4. Niveau de difficulté : moyen

a) Le lien entre la densité des plants de canola et leur rendement



- b) La corrélation est forte et positive.
- c) $Y = 0,34X + 56,4$, où X représente la densité (plants/m²) et Y , le pourcentage de plants à bon rendement.
- d) 1) Environ 83,6 % 2) Environ 100,6 %
- e) Un rendement supérieur à 100 % est impossible. Ce qui signifie que lorsqu'on atteint un rendement maximal pour une certaine densité, continuer à augmenter la densité ne fera pas augmenter davantage le rendement. La première estimation, celle de 83,6 % est la plus fiable, parce qu'elle se trouve à l'intérieur de l'intervalle des données fournies (interpolation).
- f) Non, le modèle linéaire n'est pas le plus approprié parce qu'à partir d'une certaine densité, environ 110, le rendement commence à descendre. Si on disposait de données pour des densités situées entre 120 et 160 plants/m², la forme du nuage ne serait plus linéaire.

5. Niveau de difficulté : faible

- a) Oui, à la condition que la valeur considérée se situe environ au centre des données, là où la corrélation semble plus forte.
- b) Non, parce que le nombre de données est insuffisant.
- c) Non, parce que le modèle est non linéaire.
- d) Non. La droite utilisée est très affectée par les deux points aberrants. Il s'agit probablement de la droite de Mayer. L'utilisation de cette droite de régression n'est pas appropriée pour une situation comportant un ou des points aberrants.

Manuel • p. 99

6. Niveau de difficulté : moyen

- a) Il faut d'abord trouver le taux de variation de la droite passant par les points M_1 et M_3 .

$$\frac{13 - 29}{32 - 7} = \frac{-16}{25}$$

À l'aide du point P , on trouve l'ordonnée à l'origine.

$$Y = \frac{-16}{25}X + b$$

$$20 = \frac{-16}{25} \cdot 19,5 + b$$

$$20 = -12,48 + b$$

$$32,48 = b$$

Donc, l'équation de la droite médiane-médiane

$$\text{est } Y = \frac{-16}{25}X + \frac{812}{25} \text{ ou } Y = -0,64X + 32,48.$$

- b) 1) $Y = 22,88$ 2) $Y = 3,68$

7. Niveau de difficulté : faible

- a) 1^{er} groupe: 6 données
2^e groupe: 5 données
3^e groupe: 6 données
- b) 1^{er} groupe: 8 données
2^e groupe: 8 données
3^e groupe: 8 données
- c) 1^{er} groupe: 13 données
2^e groupe: 14 données
3^e groupe: 13 données

8. Niveau de difficulté : moyen

On place en premier les couples en ordre croissant selon la valeur de X .

On constitue ensuite 3 groupes avec les 10 données, soit un premier groupe de 3 données, un deuxième groupe de 4 données et un troisième groupe de 3 données.

| X | Y |
|-----|-----|
| 21 | 52 |
| 25 | 58 |
| 27 | 57 |

| | |
|----|----|
| 32 | 61 |
| 38 | 69 |
| 40 | 66 |
| 44 | 75 |

| | |
|----|----|
| 45 | 80 |
| 50 | 82 |
| 51 | 88 |

On trouve ensuite les points médians de ces trois groupes qu'on nommera M_1 , M_2 et M_3 .

Médiane des abscisses :

1^{er} groupe: 25

2^e groupe: 39

3^e groupe: 50

Médiane des ordonnées :

1^{er} groupe : 57

2^e groupe : 67,5

3^e groupe : 82

On obtient donc $M_1(25, 57)$, $M_2(39, 67,5)$, $M_3(50, 82)$.

On trouve le point moyen des 3 points M_1 , M_2 et M_3 .

$$\text{Abscisse : } \frac{25 + 39 + 50}{3} = 38$$

$$\text{Ordonnée : } \frac{57 + 67,5 + 82}{3} = 68,8\bar{3}$$

Le point moyen est donc $P(38, 68,8\bar{3})$.

À partir des points médians M_1 et M_3 , on trouve le taux de variation :

$$\frac{57 - 82}{25 - 50} = \frac{-25}{-25} = 1.$$

On détermine ensuite l'ordonnée à l'origine à l'aide de ce taux et du point moyen $P(38, 68,8\bar{3})$.

$$68,8\bar{3} = 1 \cdot 38 + b$$

$$68,8\bar{3} = 38 + b$$

$$30,8\bar{3} = b$$

L'équation de la droite médiane-médiane est donc $Y = X + 30,8\bar{3}$.

9. Niveau de difficulté : faible

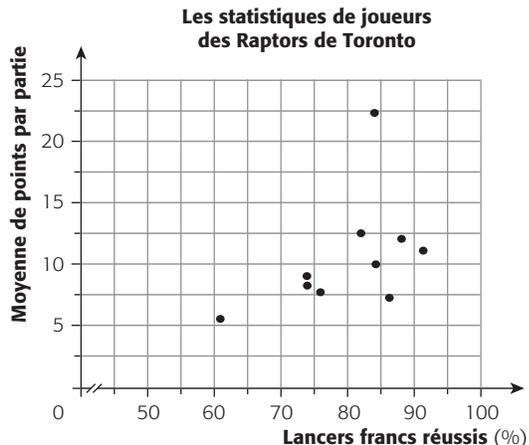
a) $Y \approx -0,005X + 18,57$, où Y représente la température (°C) et X , l'altitude (m).

b) 1) $Y \approx 16$ °C 2) $Y \approx 3,57$ °C

c) Probablement pas. La valeur considérée (10 000 m ou 10 km) est de beaucoup supérieure aux données de cette étude. Il y a peu de chances que la tendance observée entre 400 m et 2 100 m tienne aussi pour des distances beaucoup plus grandes.

10. Niveau de difficulté : moyen

a)



b) $r \approx 0,43$

c) Pour tracer la droite de Mayer, il faut ordonner les données en ordre croissant, selon la première variable, puis partager la distribution en deux groupes équipotents.

| X | Y |
|----|------|
| 61 | 5,7 |
| 74 | 8,5 |
| 74 | 9 |
| 76 | 7,8 |
| 82 | 12,5 |

| | |
|----|------|
| 84 | 22,3 |
| 84 | 10,2 |
| 86 | 7,2 |
| 88 | 12,1 |
| 91 | 11,2 |

On trouve ensuite les points moyens de ces deux groupes : P_1 et P_2 .

Premier point moyen (P_1)

En abscisse :

$$\frac{61 + 74 + 74 + 76 + 82}{5} = 73,4$$

En ordonnée :

$$\frac{5,7 + 9 + 8,5 + 7,8 + 12,5}{5} = 8,7$$

Le premier point moyen est donc $P_1(73,4, 8,7)$.

Second point moyen (P_2)

En abscisse :

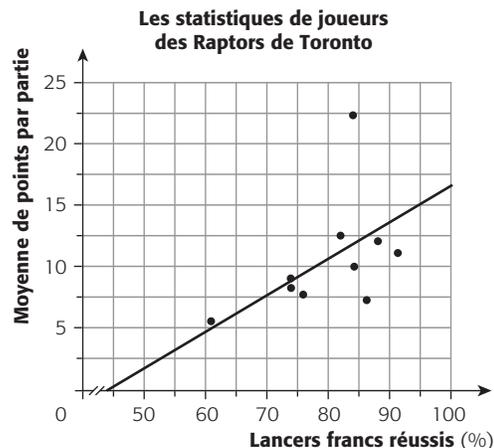
$$\frac{84 + 84 + 86 + 88 + 91}{5} = 86,6$$

En ordonnée :

$$\frac{22,3 + 10,2 + 7,2 + 12,1 + 11,2}{5} = 12,6$$

Le second point moyen est donc $P_2(86,6, 12,6)$.

On place les deux points dans le plan et on les relie pour former la droite de Mayer.



Pour tracer la droite médiane-médiane, il faut trouver les 3 points médians : M_1 , M_2 et M_3 . On ordonne les données en ordre croissant selon la variable première.

| X | Y |
|----|-----|
| 61 | 5,7 |
| 74 | 8,5 |
| 74 | 9 |

| | |
|----|------|
| 76 | 7,8 |
| 82 | 12,5 |
| 84 | 10,2 |
| 84 | 22,3 |

| | |
|----|------|
| 86 | 7,2 |
| 88 | 12,1 |
| 91 | 11,2 |

On trouve les médianes des abscisses :

1^{er} groupe : 74

2^e groupe : 83

3^e groupe : 88

On trouve les médianes des ordonnées :

1^{er} groupe : 8,5

2^e groupe : 11,35

3^e groupe : 11,2

Les trois points médians sont donc $M_1(74, 8,5)$, $M_2(83, 11,35)$ et $M_3(88, 11,2)$.

On trouve ensuite le point moyen de ces 3 points M_1 , M_2 et M_3 .

En abscisse : $\frac{74 + 83 + 88}{3} = 81,6$

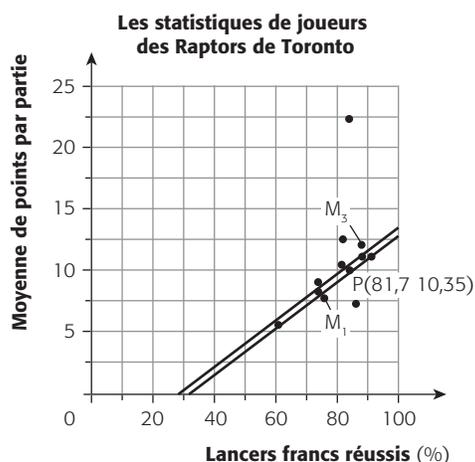
En ordonnée : $\frac{8,5 + 11,35 + 11,2}{3} = 10,35$

On obtient le point $P(81,6, 10,35)$.

On place les points M_1 , M_3 et P dans le plan cartésien.

On trace une droite pour relier M_1 et M_3 .

On trace ensuite une autre droite parallèle mais qui passe par le point P .



d) La droite de Mayer est la plus fortement influencée par le point (84, 22,3) parce qu'elle tient compte de la moyenne et non de la médiane. La médiane n'est généralement pas influencée par les valeurs extrêmes. En effet, la médiane tient compte de la donnée centrale alors que la moyenne tient compte de l'ensemble des données.

e) On privilégierait la droite médiane-médiane parce qu'elle est moins influencée par des données extrêmes, comme les données de Chris Bosh.

Consolidation

Manuel • p. 100

1. Mesure de dispersion : écart moyen, mesure de

position : rang centile

Niveau de difficulté : faible

a) ① $\bar{x} = 61$

② $\bar{x} = 58,8$

③ $\bar{x} = 94,2$

b) ① $\text{ÉM} \approx 12,93$

② $\text{ÉM} \approx 10,98$

③ $\text{ÉM} \approx 12,08$

c) Distribution ②, distribution ③ et distribution ①

d) 1) ① 74^e rang centile

② 90^e rang centile

③ 14^e rang centile

2) ① La donnée 58 occupe le 40^e rang centile.

② La donnée 53 occupe le 40^e rang centile.

③ La donnée 88 occupe le 40^e rang centile.

2. Mesure de dispersion : écart moyen

Niveau de difficulté : moyen

a) La moyenne est de 20 et l'écart moyen est de 5.

b) Plusieurs réponses sont possibles. *Exemple :* On peut changer la donnée 11 pour 21 et la donnée 25 pour 35. La moyenne augmente ainsi à 22 et l'écart moyen reste 5.

3. Mesure de position : rang centile

Niveau de difficulté : faible

a) ① 37^e rang centile ② 46^e rang centile

b) ① La donnée 58 occupe le 78^e rang centile.

② La donnée 43 occupe le 78^e rang centile.

4. **Mesure de position : rang centile**

Niveau de difficulté : moyen

Il y a 1 324 scores égaux ou inférieurs au sien.

$$\frac{1\,324 \cdot 100}{3\,156} \approx 41,95$$

Kheegan occupe le 42^e rang centile.

Manuel • p. 101

5. **Sens de la corrélation, intensité de la corrélation**

Niveau de difficulté : faible

- a) Corrélation forte, négative
- b) Corrélation parfaite, positive
- c) Corrélation nulle
- d) Corrélation faible, positive

6. **Sens de la corrélation, intensité de la corrélation,**

nature du lien entre deux variables

Niveau de difficulté : faible

a) Plusieurs réponses sont possibles. *Exemple :*

Situation ① Positif. De façon générale, les familles à plus haut revenu ont une maison de plus grande valeur.

Situation ② Négatif. Plus une équipe gagne, moins elle perd.

Situation ③ Positive. De façon générale, les villes au sud des provinces, au climat plus chaud, sont plus peuplées et ont plus de policiers

Situation ④ Aucune corrélation. Pas de lien apparent entre l'âge du directeur et le succès des élèves.

Situation ⑤ Négatif. De façon générale, une secrétaire ayant plus d'expérience prendra moins de temps pour dactylographier.

Situation ⑥ Positif. Plus la distance est grande, plus le voyage prend du temps.

Situation ⑦ Positif. De façon générale, plus l'année de fabrication est récente, plus la valeur d'une voiture est grande.

Situation ⑧ Positif. On peut présumer que, plus il y a d'habitants dans une ville, plus il y aura de feux de circulation et de restaurants.

b) Situation ① Forte. Le logement est une part importante des dépenses d'un ménage.

Situation ② Parfaite. Si chaque équipe joue n parties, tous les points seront sur la droite $v + d = n$, où v représente le nombre de victoires et d , le nombre de défaites.

Situation ③ Moyenne.

Situation ④ Aucune corrélation.

Situation ⑤ Moyenne. Il peut y avoir des exceptions. La vitesse n'augmente pas sans cesse avec l'expérience.

Situation ⑥ Moyenne. Le temps requis augmente souvent avec la distance à parcourir, mais dans les villes, à cause de la circulation, ce n'est pas nécessairement proportionnel.

Situation ⑦ Moyenne. L'usure de la voiture peut aussi jouer un rôle.

Situation ⑧ Forte. Il y a un plus grand nombre de restaurants et de feux de circulation dans les grandes villes.

c) Situation ① S'il y a un lien, il est fortuit.

Situation ② Causal. Plus une équipe a de victoires, moins elle a de défaites et vice-versa.

Situation ③ 3^e facteur d'influence, soit le nombre d'habitants.

Situation ④ S'il y a un lien, il est fortuit.

Situation ⑤ Causal. De façon générale, l'expérience augmente les performances (la vitesse).

Situation ⑥ Causal. Une grande distance prendra plus de temps à être parcourue.

Situation ⑦ Causal. Les acheteurs veulent une voiture plus récente et acceptent de payer plus.

Situation ⑧ 3^e facteur d'influence, soit le nombre d'habitants.

7. **Approximation et interprétation du coefficient de corrélation linéaire**

Niveau de difficulté : faible

- a) ② b) ⑤ c) ① d) ④ e) ③

Manuel • p. 102

8. **Distribution à deux caractères, nuage de points, sens de la corrélation, intensité de la corrélation**

Niveau de difficulté : faible

- a) 1) Situation ① 3) Situation ②
2) Situation ③ 4) Situation ④

b) Graphique 1) La corrélation est positive, d'intensité faible lorsqu'il y a un faible taux d'alcoolémie dans le sang, mais devient forte lorsque ce taux augmente.

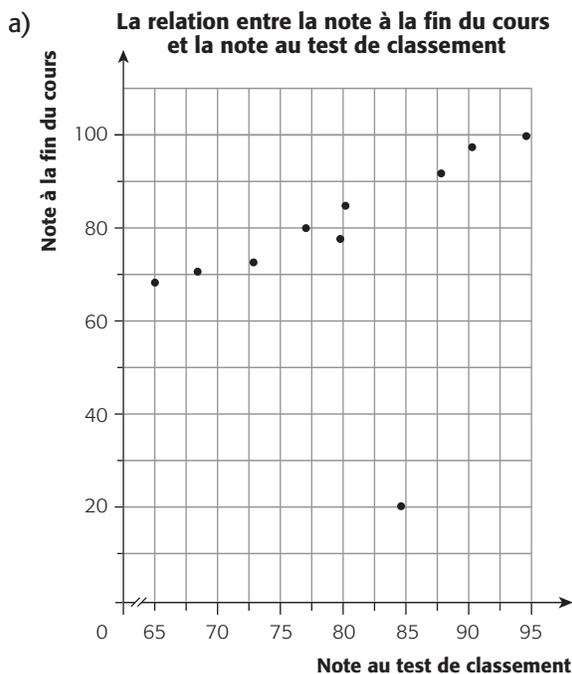
Graphique **2**) La corrélation est positive, d'intensité moyenne. Il y a quelques exceptions. Il s'agit de personnes qui ont un bon salaire malgré un petit nombre d'années de scolarité.

Graphique **3**) Corrélation positive. La corrélation est forte pour les enfants en bas âge au primaire mais devient faible lorsqu'ils vieillissent.

Graphique **4**) Hormis deux données, la corrélation est forte et négative. Plus les ventes de tuques sont élevées, moins il se vend de maillots de bain.

9. **Distribution à deux caractères, nuage de points, interprétation et approximation du coefficient de corrélation linéaire, limites de l'interprétation**

Niveau de difficulté : faible



- b) Oui, même si une donnée est à l'écart des autres.
 c) 1) 0,34 2) 0,98
 d) Plusieurs réponses sont possibles. *Exemple* : Il peut s'agir d'un étudiant qui a été malade et a manqué plusieurs cours. Il a conséquemment obtenu une mauvaise note.

10. **Droite de régression : droite de Mayer, droite médiane-médiane, prédictions**

Niveau de difficulté : faible

- a) Situation ① : $Y \approx -1,11X + 25,67$
 Situation ② : $Y \approx 0,45X + 3,02$
 b) Situation ① : $Y \approx -1,04X + 25,12$
 Situation ② : $Y \approx 0,48X + 2,83$

- c) Situation ① : selon la droite de Mayer : $Y = -7,67$
 Situation ① : selon la droite médiane-médiane : $Y \approx -6,08$
 Situation ② : selon la droite de Mayer : $Y \approx 16,52$
 Situation ② : selon la droite médiane-médiane : $Y \approx 17,23$

Manuel • p. 103

11. **Faire des liens**

Sens de la corrélation, intensité de la corrélation

Niveau de difficulté : faible

Plusieurs réponses sont possibles. *Exemples* :

- a) L'âge du conducteur et l'âge de la voiture. Les personnes plus âgées sont sans doute plus économes et veulent conserver leur voiture plus longtemps. Mais il y a sûrement beaucoup d'exceptions.
 b) La valeur de la voiture et l'âge de la voiture. De façon générale, plus la voiture est âgée, moins elle a de valeur.
 c) L'âge de la voiture et le kilométrage de la voiture. Une voiture plus âgée devrait avoir plus roulé, de façon générale.
 d) Le kilométrage de la voiture et le prix de l'assurance automobile. Le kilométrage augmente alors que les primes d'assurance restent les mêmes ou varient pour d'autres raisons.

Le prix de l'assurance automobile et l'âge du conducteur. Plus le conducteur est jeune, plus les primes d'assurance sont élevées. Mais ensuite, quand il devient très vieux, le conducteur paie plus cher pour ses assurances.

Le dossier de conduite du conducteur et le prix de l'assurance automobile. Même si le dossier s'améliore, le prix de l'assurance ne descendra pas au-dessous d'un certain minimum.

12. **Ce n'est pas pour rien**

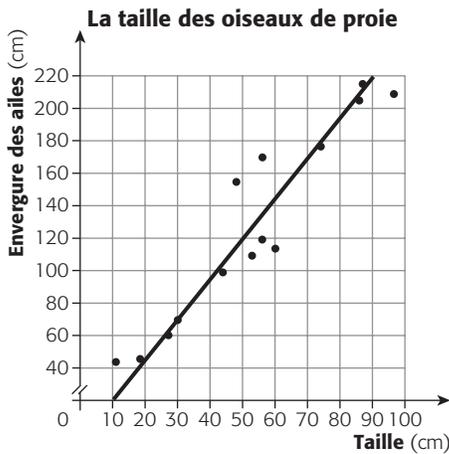
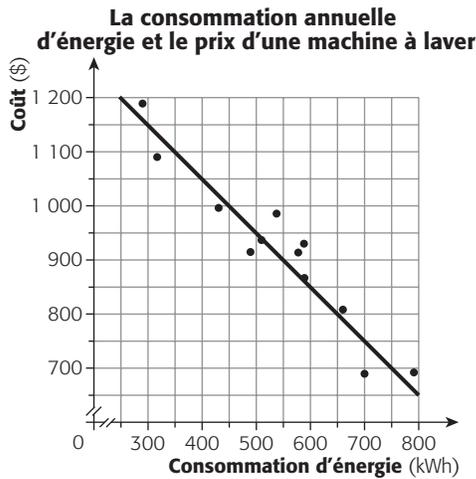
Approximation et interprétation du coefficient de corrélation linéaire, sens de la corrélation, intensité de la corrélation, droite de régression : prédictions

Niveau de difficulté : faible

- a) 1) Première situation : $-0,86$
 Deuxième situation : $0,8$
 2) Première situation : corrélation forte, négative. Les machines à laver de meilleure qualité coûtent plus cher, mais consomment moins d'électricité.

Deuxième situation : corrélation forte, positive. Il est normal qu'un oiseau de grande taille possède des ailes de grande envergure.

3) Plusieurs réponses sont possibles. Exemple :



4) Première situation :
 $Y = -X + 1\,450$

Deuxième situation :
 $Y = 2,5X - 5$

b) 1) $1\,050 = -X + 1\,450$
 $X = 400$

La consommation annuelle d'énergie d'une machine à laver qui coûte 1 050 \$ est d'environ 400 kWh.

2) $Y = 2,5X - 5$
 $Y = 2,5(40) - 5$
 $Y = 100 - 5$
 $Y = 95$

L'envergure des ailes d'un oiseau de proie dont la taille est de 40 cm est d'environ 95 cm.

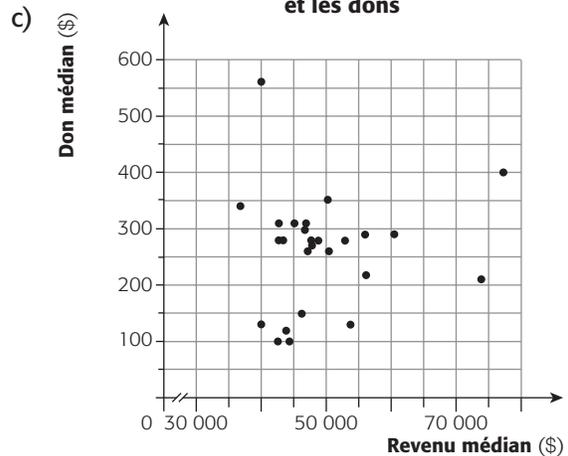
13. Les dons par région

Distribution à deux caractères, nuage de points

Niveau de difficulté : faible

- a) Si on ordonne en ordre croissant toutes les données sur le revenu des personnes d'une région, le revenu médian est celui qui se trouve au centre de la distribution. Ainsi, 50 % des gens gagnent plus que le revenu médian et 50 % gagnent moins que le revenu médian. Il en va de même pour le don médian.
- b) Plusieurs réponses sont possibles. Exemple : On peut supposer que les gens ayant un revenu plus élevé sont en mesure de faire des dons plus élevés.

La relation entre les revenus et les dons



- d) Non
- e) Il semble que ce soit en Colombie-Britannique que les gens sont le plus généreux. Il manque toutefois de données pour tirer une conclusion valable.

14. L'éparpillement

Mesure de dispersion : écart moyen

Niveau de difficulté : faible

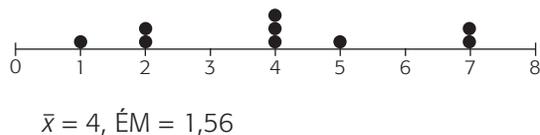
- a) La distribution ② a le plus grand écart moyen.
- b) La distribution ① a le plus grand écart moyen.
- c) La distribution ② a le plus grand écart moyen.
- d) La distribution ② a le plus grand écart moyen.

15. Remaniement

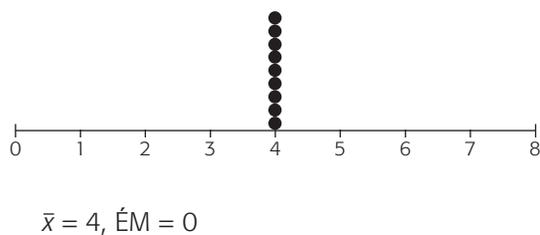
Mesure de dispersion : écart moyen

Niveau de difficulté : moyen

- a) La moyenne est de 4.
- b) 1) Plusieurs réponses sont possibles. Exemple :



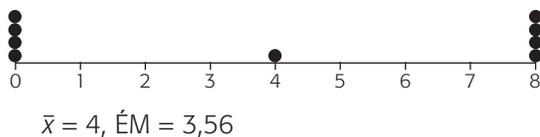
2)



- 3) Plusieurs réponses sont possibles. Exemple :



4)



16. Postes à combler

Mesure de position : rang centile

Niveau de difficulté : moyen

- a) 1) Il y a 240 personnes qui ont obtenu une note inférieure ou égale à celle de Camille. La note de Camille occupe le 72^e rang centile, alors elle sera invitée en entrevue.
- 2) Il y a 221 personnes qui ont obtenu une note inférieure ou égale à celle de William. La note de William occupe le 66^e rang centile, alors il ne sera pas invité en entrevue.
- b) La note de Catherine est la 243^e donnée de la distribution ordonnée des notes.

17. Comment se fait-il que... ?

Sens de la corrélation, intensité de la corrélation, nature du lien entre deux variables

Niveau de difficulté : faible

Plusieurs réponses sont possibles. Exemple :

- a) Les filles sont les principales responsables de l'approvisionnement en eau potable de la famille. Si la famille a accès à de l'eau facilement, les filles ont plus de temps pour aller à l'école.
- b) Les cigognes nichent principalement en territoire rural, là où il y a généralement des familles plus nombreuses.
- c) Le salaire augmentant généralement avec l'âge, il est normal de trouver moins de cheveux sur la tête de personnes plus âgées.
- d) Dans les pays en développement, les gens n'ont pas les moyens financiers de fumer et l'espérance de vie y est moins élevée que dans les pays développés. À l'intérieur d'un même pays toutefois, il y a une corrélation négative entre le tabagisme et l'espérance de vie.

18. Mise en forme

Sens de la corrélation, intensité de la corrélation

Niveau de difficulté : moyen

- a) Corrélation positive, forte.
- b) 1) À partir d'une droite de régression tracée sur le plan, on cherche la donnée qui s'en écarte le plus vers le bas. Réponse : H
- 2) À partir d'une droite de régression tracée sur le plan, on cherche la donnée qui s'en écarte le plus vers le haut. Réponse : J

19. Sauter aux conclusions

Approximation et interprétation du coefficient de corrélation linéaire

Niveau de difficulté : moyen

- a) Faux. Le coefficient de 0,6 signifie seulement que, de façon générale, les grands joueurs récupèrent plus de rebonds.
- b) Faux. La relation démontre seulement que les petits joueurs en récupèrent moins.

- c) Faux. On ne peut pas conclure que le pourcentage de grands joueurs qui récupèrent plus de rebonds que la moyenne soit de 60 %.
- d) Vrai. Mais il s'agit d'un énoncé général qui peut comporter des exceptions.

20. Il faut être Thomas

Approximation et interprétation du coefficient de corrélation linéaire

Niveau de difficulté : faible

Faux. Il faut absolument voir le nuage de points pour se prononcer. On ne peut être certain que si on obtient exactement 1 ou -1.

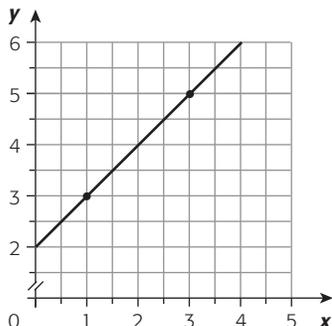
21. La seule, l'unique!

Approximation et interprétation du coefficient de corrélation linéaire

Niveau de difficulté : faible

En reliant deux points par une droite, étant donné que cette droite touche nécessairement les deux points, le coefficient est obligatoirement de 1 ou de -1. Si la droite qui relie les deux points est horizontale ou verticale, alors le coefficient de corrélation est non défini.

Exemple :



Manuel • p. 107

22. Pas si différent

Approximation et interprétation du coefficient de corrélation linéaire, coefficient de corrélation linéaire : limites de l'interprétation

Niveau de difficulté : moyen

- a) 1) Situation ① : $r \approx 0,45$
Situation ② : $r \approx 0,23$
- 2) Situation ① : $r \approx 0,81$
Situation ② : $r \approx 0,8$

- b) Plusieurs réponses sont possibles. *Exemple :*

1) Situation ① : Il s'agit d'un modèle de bougie faite d'une cire plus résistante.

Situation ② : C'est le cas d'une maison ayant un système de chauffage au bois plutôt qu'à l'électricité.

2) Situation ① : Une bougie est placée à un endroit protégé des courants d'air.

Situation ② : La maison n'est pas habitée.

23. Qui dit mieux?

Droite de régression : droite de Mayer, droite médiane-médiane et prédictions

Niveau de difficulté : faible

- a) Situation ① : $Y \approx 0,72X + 1,01$
Situation ② : $Y = 0,6X + 1,52$
- b) Situation ① : environ 29,81
Situation ② : 25,52
- c) Environ 4,29
- d) Il est préférable de choisir la droite médiane-médiane, car elle est moins influencée par les données extrêmes.

Manuel • p. 108

24. Les objectifs du millénaire

Approximation et interprétation du coefficient de corrélation linéaire, sens de la corrélation, intensité de la corrélation, nature du lien entre deux variables

Niveau de difficulté : faible

- a) $r \approx -0,52$
- b) Corrélation moyenne, négative. Plus les accouchements se font en présence de personnel qualifié, moins il y a de mortalité infantile.
- c) Plusieurs réponses sont possibles. *Exemple :* Un troisième facteur peut être en cause, soit la proximité d'un hôpital, la disponibilité de médicaments ou la facilité d'accès à de l'eau de qualité, tous ces facteurs pouvant être associés à la présence de personnel qualifié.
- d) Plusieurs réponses sont possibles. *Exemple :* Des aliments en quantité et en qualité, la qualité de l'eau, la disponibilité de médicaments ou de vaccins, etc.

25. Les gardiens de la LNH

Droite de régression : droite de Mayer et prédictions**Niveau de difficulté : faible**

- a) $Y = -6,36X + 50,18$, où Y représente le nombre de victoires et X , la moyenne de buts contre (MBC).
 b) Environ 34 victoires

26. Les bébés

Mesure de position : le rang centile**Niveau de difficulté : faible**

- a) 1) Le bébé qui avait une masse de 3 002 g à la naissance occupe le 35^e rang centile.
 2) Le bébé qui avait une masse de 3 156 g à la naissance occupe le 48^e rang centile.
 b) Le 72^e rang centile est occupé par la 151^e donnée, ce qui correspond à une masse de 3 482 g.

27. De graves conséquences

Droite de régression : droite de Mayer, droite**médiane-médiane et prédictions****Niveau de difficulté : moyen**

Il faut d'abord trouver l'équation de la droite de Mayer :

La moyenne des abscisses du Groupe 1 est d'environ 25,87.

La moyenne des ordonnées du Groupe 1 est d'environ 81,48.

La moyenne des abscisses du Groupe 2 est d'environ 28,63.

La moyenne des ordonnées du Groupe 2 est de 104,2.

Les coordonnées des deux points qui seront utilisés pour déterminer la règle sont (25,87, 81,48) et (28,63, 104,2).

$$a = \frac{104,2 - 81,48}{28,63 - 25,87}$$

$$a = \frac{568}{69}$$

$$Y = \frac{568}{69}X + b$$

$$104,2 = \frac{568}{69}(28,63) + b$$

$$b \approx -131,48$$

$$Y = \frac{568}{69}X - 131,48 \text{ ou } Y = 8,23X - 131,48, \text{ où}$$

Y représente le taux de décès liés au diabète, pour 100 000 décès, et X , le pourcentage d'enfants ayant un excès de poids.

On trouve ensuite le pourcentage d'enfants ayant un excès de poids en supposant que le taux de décès est de 100.

$$100 = 8,23X - 131,48$$

$$231,48 = 8,23X$$

$$X \approx 28,13$$

Le pourcentage d'enfants ayant un excès de poids dans le district 11 est présentement de 28,8 %. Il faudrait le diminuer de 0,67 % pour que le nombre de décès liés au diabète passe sous la barre des 100 pour 100 000 décès.

28. Plus près de chez nous

Droite de régression : droite de Mayer, droite**médiane-médiane et prédictions****Niveau de difficulté : moyen**

$$Y = 8,23X - 131,48$$

$$Y = 8,23(23) - 131,48$$

$$Y = 57,81$$

Avec les données pour la Californie, à 23 %, le taux de décès devrait être d'environ 57,81. La situation au Québec semble être très différente de celle de la Californie. Ceci peut s'expliquer en partie par le fait que les données du Québec sont pour l'année 2004 alors que celles de la Californie sont pour 1996 à 2000, ou peut-être par la présence d'un autre facteur, comme le climat ou le mode d'alimentation. Au Québec, le problème d'obésité chez les jeunes est relativement récent. Peut-être que les conséquences ne se sont pas encore fait sentir autant qu'en Californie.

29. Une différence prévisible

Droite de régression : prédictions**Niveau de difficulté : moyen**

Dans l'équation $Y = 2,51X + 45,2$, le taux de variation est de 2,51 cm de taille par cm de longueur de fémur. Avec une différence de fémur de 2 mm, on obtient 0,2 cm.

$$2,51 \cdot 0,2 = 0,50$$

Les jumelles auront une différence de taille d'environ 0,50 cm.

30. Mieux vaut prévenir que guérir...

Droite de régression : droite de Mayer, droite**médiane-médiane et prédictions****Niveau de difficulté : élevé**

On doit d'abord trouver l'équation de la droite de régression à partir d'une des deux méthodes ou à l'aide de la calculatrice.

$Y = 0,11X - 1,53$, où X représente la consommation de viande (g/personne par jour) et Y , le taux de cancer du côlon (/100 000 personnes).

Pour le Canada, l'application de cette équation donne un taux de :

$$Y = 0,11X - 1,53$$

$$Y = 0,11 \cdot 240 - 1,53$$

$$Y = 26,4 - 1,53$$

$$Y = 24,87$$

Ce taux signifie qu'environ 24,87 cancers du côlon surviendront pour chaque tranche de 100 000 habitants.

On a donc la proportion suivante :
24,87 cancers du côlon pour 100 000

Pour 33 500 000 habitants, on aura :

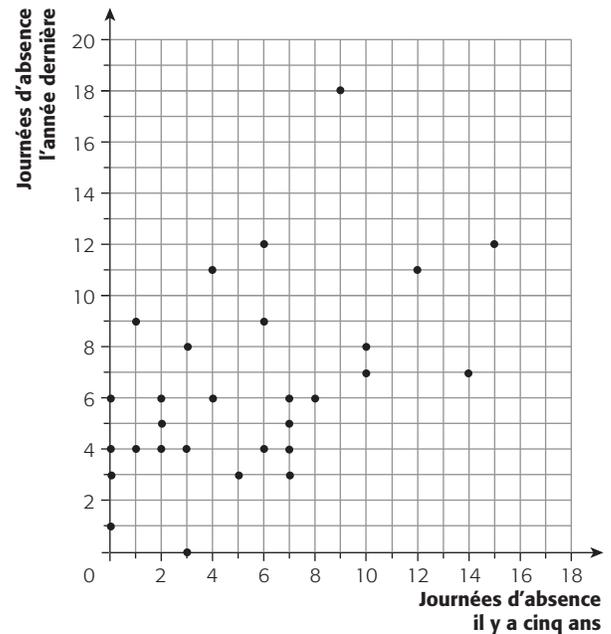
$$\frac{24,87 \cdot 33\,500\,000}{100\,000} = 8\,331,45 \text{ cas de cancer du côlon.}$$

Si 90 % des cas de cancer apparaissent chez les gens de plus de 50 ans, 10 % des cas de cancer surviennent chez les gens de moins de 50 ans.
 $10 \cdot 8\,331,45 = 833,1$

Environ 833 Canadiens de moins de 50 ans sont atteints du cancer du côlon.

de lien important entre l'absentéisme d'il y a cinq ans et l'absentéisme de l'année dernière.

La relation entre le nombre de jours d'absence il y a cinq ans et l'année dernière



3. Recommandations

Étant donné que l'absentéisme a augmenté au sein de l'entreprise au cours des dernières années, il faudra privilégier certaines mesures pour réduire ce phénomène.

Puisque le stress est un des principaux facteurs qui expliquent l'absentéisme, nous recommandons des mesures préventives pour diminuer le stress dans le milieu de travail.

Voici quelques exemples de mesures que les employeurs peuvent prendre pour procurer aux employés un environnement de travail plus agréable :

- Adapter le travail demandé aux capacités et aux ressources des employés (en tenant compte de leur âge, par exemple).
- Organiser le travail pour le rendre stimulant et permettre aux employés de faire appel à leurs compétences.
- Définir clairement les rôles et les responsabilités de chacun.
- Donner la possibilité aux employés de participer aux décisions et aux interventions qui affecteront leur travail.
- Améliorer la communication interne et réduire les incertitudes.
- Faciliter les échanges et le dialogue entre tous les acteurs de l'entreprise.

31. Le coût caché de la maladie

Approximation et interprétation du coefficient de corrélation linéaire, sens de la corrélation, intensité de la corrélation, mesure de dispersion : écart moyen

Niveau de difficulté : moyen

1. On calcule la moyenne et l'écart moyen des journées d'absence d'il y a cinq ans et de l'année dernière.
 - 5 ans plus tôt : Les gens se sont absentés en moyenne 5,5 jours et l'écart moyen est de 3,5.
 - L'année dernière : Les gens se sont absentés en moyenne 6,3 jours et l'écart moyen est de 2,7.

Les données étaient plus dispersées il y a cinq ans, tandis que pour l'année dernière l'écart moyen est plus petit, ce qui indique que les données sont moins dispersées. De façon générale, on peut constater que la moyenne des journées d'absence a augmenté et que le phénomène semble plus généralisé puisque l'écart moyen est plus petit.

2. On trace un nuage de points et on estime le coefficient de corrélation linéaire.

Le coefficient de corrélation linéaire est d'environ 0,40. La corrélation entre les deux variables est faible et positive. Il semble qu'il n'existe pas