Unité F : Statistique

Demi-cours V
Guide de l'élève

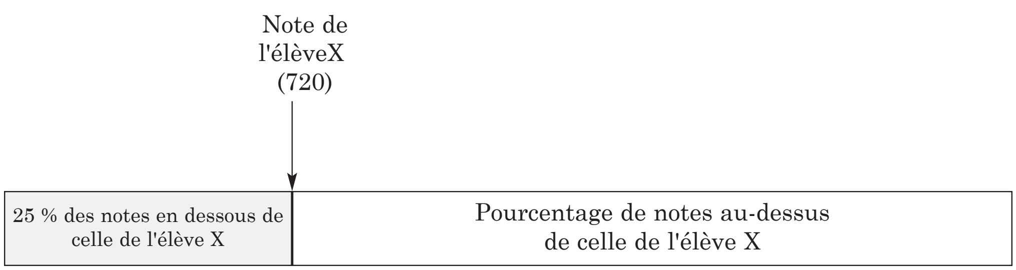
Leçon 1 : Centiles

Dans le manuel Secondaire 3 – Statistique, nous avons vu comment la moyenne, la médiane, le mode et l'étendue servent à analyser la tendance centrale (moyenne) de données et de quelle façon ces données sont dispersées. Par contre, il nous arrive souvent de vouloir savoir la position d'un terme dans un ensemble de données, c'est-à-dire comment une donnée se compare aux autres.

Par exemple, supposons que l'élève X passe un examen d'admission à l'université et qu'il obtient une note brute de 720 sur 900. Cette note est équivalente à 80 %. Habituellement, les gens considèrent cette note comme un « A », qui est lié à l'excellence. Cependant, sans savoir comment la note de l'élève se situe par rapport à celle des autres élèves, la note brute de 720 n'a pas de véritable signification.

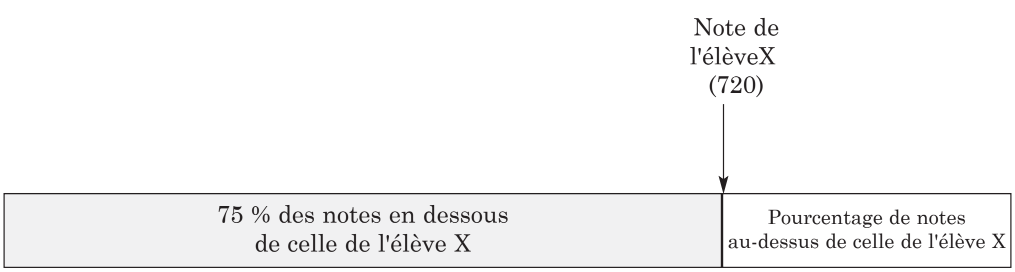
Si, par exemple, 1 000 élèves passent le même examen et que seulement 250 d'entre eux (25 %) obtiennent une note en dessous de celle de l'élève X, la note de 720 n'est pas aussi impressionnante. (Suppose que n élèves ont obtenu la même note que l'élève X.)

Exemple 1



Si, par contre, 750 (75 %) élèves obtiennent une note en dessous de 720, cette dernière est considérée très bonne. (Suppose qu'aucun autre élève n'a obtenu la même note que l'élève X.)

Exemple 2



Comme tu peux le voir, la note brute de 720 sur 900 a une plus grande signification si elle est comparée aux autres notes brutes.

Il est possible de comparer les notes en leur attribuant un **rang centile**, ce qui nous indique le pourcentage de notes se situant **en dessous** d'une note en particulier.

À l'exemple 1 de la page précédente, où 25 % des élèves ont obtenu une note en dessous de 720, le rang centile de l'élève X est de 25. Ce résultat indique aussi, bien sûr, que l'élève X a obtenu une note **plus haute que celle de seulement 25 %** de tous les élèves qui ont passé l'examen.

À l'exemple 2 de la page précédente, où 75 % des élèves ont obtenu une note en dessous de 720, le rang centile de l'élève X est de 75. Ce résultat indique aussi, bien sûr, que l'élève X a obtenu une note **plus haute que celle de 75 %** de tous les élèves qui ont passé l'examen.

Plus le rang centile est élevé, plus le nombre de notes en dessous d'une note en particulier est élevé ou plus cette note est élevée par rapport à toutes les autres notes.

Il est possible de calculer le rang centile à l'aide de la formule suivante :

$$\text{Rang centile} = \frac{\text{D} + 0{,}5\text{E}}{\text{n}} \times 100$$

D étant le nombre de notes en **D**essous d'une note en particulier.

E étant le nombre de notes brutes **É**gales à la note, comprenant la note donnée. Par contre, s'il y a *n* autres notes égales à la note donnée, *E* = 0.

n étant le **n**ombre total de notes brutes.

Cette formule permet d'ajouter toutes les notes moins élevées que la note donnée (*D*) à la moitié des notes égales à la note donnée (*E*). La somme est ensuite convertie à un rang centile en la divisant par le nombre total de notes (*n*), puis en multipliant par 100.

Exemple 1

D = 250 (25 % x 1000), *E* = 0*, *n* = 1000

* aucune note égale à la note donnée

$$\text{rang centile} = \frac{250}{100} \times 100 = 25 \text{ (le 25}^{\text{e}}\text{ centile)}$$

Exemple 2

D = 750 (75 % x 1000), E = 0*, n = 1000
*   aucune note égale à la note donnée

$$\text{rang centile} = \frac{750+0}{1000} \times 100 = 75 \text{ (le } 75^{\text{e}} \text{ centile)}$$

Exemple 3

960 élèves passent un examen d'admission à l'université. 750 de ces élèves obtiennent une note plus basse que celle de l'élève X, qui obtient 74 %. 47 autres élèves obtiennent aussi 74 %. Quel est le rang centile de l'élève X?

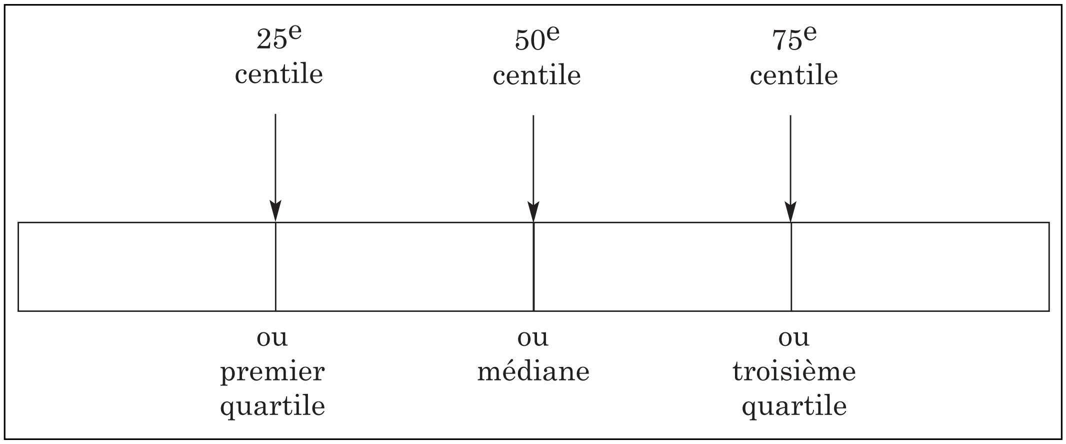
D = 750, E = 48 (47 + 1), n = 960

$$\begin{aligned}\text{rang centile} &= \frac{\text{D} + 0{,}5\text{E}}{\text{n}} \times 100 \\ &= \frac{750 + 0{,}5(48)}{960} \times 100 \\ &= \frac{750 + 24}{960} \times 100 \\ &= \frac{774}{960} \times 100 \\ &= (0{,}806\ 25) \times 100 \\ \text{rang centile} &= 80{,}625 \text{ ou } 81\ *\end{aligned}$$

* les rangs centiles sont habituellement arrondis au **prochain** nombre entier

Dans cet exemple, le rang centile de l'élève X est de 81. L'élève X a obtenu une aussi bonne note ou une meilleure note que 81 % des élèves qui ont passé l'examen d'admission.

On utilise souvent les centiles pour décrire les résultats d'examen et le classement des personnes ayant fait ces examens. Cela est très pratique dans le cadre de demandes d'emploi au gouvernement ou d'admission à des facultés universitaires ou collèges communautaires. S'il y a plus de demandeurs que de postes offerts, les candidats sont classés selon les centiles. Comme les centiles sont utilisés assez souvent, on donne un nom particulier aux 25^e^ et 75^e^ centiles. Bien sûr, le 50^e^ centile représente la médiane.



Les divers centiles sont représentés par la lettre C ayant l'indice approprié. Par exemple C_{20} représente le 20e centile et C_{88} représente le 88e centile.

Exercice : Centiles

1. Un groupe d'élèves de secondaire 1 doivent passer un test différentiel d'aptitude afin d'évaluer leurs points forts et leurs points faibles en français et en mathématiques. Le test est composé de huit sous-catégories. Le rang centile est déterminé pour chaque élève et chaque sous-catégorie.

   Utilise le rang centile des élèves A et B pour répondre aux questions qui suivent :

	Résolution de problèmes	Calculs	Habiletés math.	Raisonnement abstrait	Compréhension écrite	Raisonnement mécanique	Relations graphiques	Orthographe
ÉlèveA Rang centile	55	65	62	40	61	94	83	39
ÉlèveB Rang centile	80	96	84	77	78	52	67	62

   a) Quelle semble être la meilleure aptitude de l'élève A?

   b) Quel est le pourcentage d'élèves ayant obtenu une note moins haute que celle de l'élève A en habiletés mathématiques?

c) Pour la sous-catégorie compréhension écrite, comment l'élève A et l'élève B se comparent-ils aux autres élèves ayant passé le test?

d) En faisant référence à ce test d'aptitude en particulier, estime le domaine d'étude dans lequel l'élève B a une plus grande possibilité de bien réussir.

e) Quelle semble être la moins bonne aptitude de l'élève A?

2. Un total de 3 286 élèves passent un examen d'admission à l'université. L'élève X et 432 autres élèves obtiennent une note de 891 sur 1 200. En tout, 2 279 élèves ont obtenu moins de 891.

   Pour être admis, l'élève doit avoir un rang centile de 70 ou mieux.

   a) Quel est le rang centile de l'élève X?

   b) L'élève X a-t-il obtenu une note assez haute pour être admis à l'université?

3. Dans le cours de musique de Lucie composé de 50 élèves, 26 d'entre eux jouent un plus petit nombre d'instruments que Lucie et 4 élèves peuvent en jouer autant.

   a) Quel est son rang centile? Que signifie ce centile?

   b) Quel pourcentage d'élèves jouent plus d'instruments que Lucie?

4. Les statistiques qui suivent indiquent le revenu familial au sein de la collectivité de Manwintoba, qui compte 2 200 familles :

   C_{25} = 15 500 $ C_{50} = 28 475 $

   C_{75} = 42 750 $ C_{85} = 64 250 $

   Environ quel pourcentage des familles gagnent :

   a) moins de 28 475 $?

   b) plus de 64 250 $?

   c) moins de 42 750 $?

   d) plus de 15 500 $?

   e) entre 15 500 $ et 64 250 $?

   f) Quel est le revenu familial médian de cette collectivité?

5. L'ensemble qui suit représente 40 notes obtenues par des élèves lors d'un examen :

16	29	40	51	65	65	77	86
18	29	41	53	65	70	79	86
21	29	41	55	65	71	81	91
22	38	41	56	65	73	83	93
25	39	41	59	65	75	85	97

   Détermine le rang centile pour chacune des notes suivantes. (*Nota* : Arrondis les rangs au prochain nombre entier.)

   a) 41

   b) 93

   c) 29

   d) 65

6. Robert obtient une moyenne finale de 88 % en secondaire 4. Le collège auquel il veut être admis n'accepte aucun candidat ayant un rang centile de moins de 85. Peut-il être certain d'être admis à ce collège ou est-il possible qu'il soit refusé? Explique ta réponse.

7. Un élève obtient 38 % lors d'un examen de mathématiques. Par contre, le rang centile de l'élève sur ce test est de 82. Que peux-tu déduire du taux de réussite de la plupart des autres élèves? Qu'est-ce qui aurait pu produire de tels résultats?

Leçon 2 : Mesures de variation - Écart type

Les mesures des tendances centrales ne donnent pas toutes les informations, par elles-mêmes, à la personne qui veut connaître un ensemble de chiffres. Souvent, il est également important de savoir comment les chiffres varient, ou ce que nous appelons la **mesure de variation.**

Nous avons déjà examiné une telle mesure, soit **l'étendue**. L'étendue d'un ensemble représente la différence entre le plus grand et le plus petit chiffre de l'ensemble.

Une autre mesure de variation est **l'écart type**.

Examinons un exemple pour voir comment calculer l'écart type et ce que cet écart indique sur la variation des données.

Exemple 1

Voici une liste des résultats d'examen, sur 100, pour un groupe d'élèves de secondaire 4 de l'École secondaire Mordudesmaths.

42 53 59 66 68 68 71 76 83 94

L'étendue de l'ensemble est : $94 - 42 = 52$.

Pour déterminer l'écart type de cet ensemble, fais les étapes suivantes (les symboles sont présentés pour représenter les différentes valeurs de la formule).

Étape 1 : Détermine la moyenne $(\bar{x})$.

$$\text{Moyenne} = \frac{42+53+59+66+68+68+71+76+83+94}{10}$$

$$\text{Moyenne } (\bar{x}) = 68$$

***Nota* :** Le nombre de données est égal à 10. Donc, $n = 10$.

Étape 2 : Détermine la différence entre chaque note (x) et la moyenne $(\bar{x})$.

***Nota* :** On obtient une différence négative si la note est moins élevée que la moyenne.

Résultats d'examen (x)	Différence de la moyenne $(x - \bar{x})$
42	–26
53	–15
59	–9
66	–2
68	0
68	0
71	3
76	8
83	15
94	26

Étape 3 : Détermine le carré de chacune de ces différences.

***Nota* :** Le carré des différences produit uniquement des valeurs positives.

Résultats d'examen (x)	Différence de la moyenne $(x - \bar{x})$	Carré de la différence $(x - \bar{x})^2$
42	–26	676
53	–15	225
59	–9	81
66	–2	4
68	0	0
68	0	0
71	3	9
76	8	64
83	15	225
94	26	676

Étape 4 : Détermine la somme des carrés.

***Nota* :** Le symbole $\sum$ (la lettre grecque sigma) représente la somme d'un certain nombre de valeurs.

Somme des carrés $\left(\sum(x-\bar{x})^2\right) = 1960$

Résultats d'examen (x)	Différence de la moyenne $(x - \bar{x})$	Carré de la différence $(x - \bar{x})^2$
42	–26	676
53	–15	225
59	–9	81
66	–2	4
68	0	0
68	0	0
71	3	9
76	8	64
83	15	225
94	26	676
Somme des carrés	$\sum(x - \bar{x})^2$	1960

Étape 5 : Divise la somme des carrés par n – 1.

$$\text{Moyenne des carrés} = \frac{1960}{9}$$

Moyenne des carrés = 217,78

Étape 6 : L'écart type (S_x) est égal à la racine carrée de ce chiffre.

$$s_x = \sqrt{217{,}78}$$
$$s_x = 14{,}76$$

Bref, on peut déterminer l'écart type d'un groupe de valeurs à l'aide de ces étapes :

1. Calcule la moyenne des chiffres. $(\bar{x})$
2. Détermine la différence entre chaque chiffre et la moyenne. $(x-\bar{x})$
3. Détermine le carré de ces différences. $(x-\bar{x})^2$
4. Détermine la somme de ces carrés. $\left(\sum(x-\bar{x})^2\right)$
5. Divise la somme des carrés par n – 1. $\dfrac{\left(\sum(x-\bar{x})^2\right)}{n-1}$
6. Calcule la racine carrée de ce nombre. $\sqrt{\dfrac{\left(\sum(x-\bar{x})^2\right)}{n-1}}$

On peut représenter ces étapes à l'aide de la formule suivante :

$$\text{écart type } (s_x) = \sqrt{\frac{\left(\sum (x-\bar{x})^2\right)}{n-1}}$$

Que signifie la valeur de l'écart type par rapport à un ensemble de chiffres? Pour trouver la réponse, examinons un ensemble de résultats d'examen semblable provenant d'une autre école secondaire.

Exemple 2

Voici une liste des résultats d'examen, sur 100, pour un groupe d'élèves de secondaire 4 de l'École secondaire Einstein.

42 62 66 66 68 68 70 70 74 94

1. Détermine l'étendue de cet ensemble de notes.
2. Détermine la moyenne de cet ensemble de notes.

Comme tu peux le voir, l'étendue et la moyenne des deux écoles sont identiques. Selon ces deux calculs seulement, les résultats d'examen des deux écoles semblent identiques. Par contre, il est clair que les notes de chaque école ne le sont pas. Quelle est la différence entre les écarts-types? Utilise le tableau qui suit pour t'aider à calculer l'écart type de l'École secondaire Einstein.

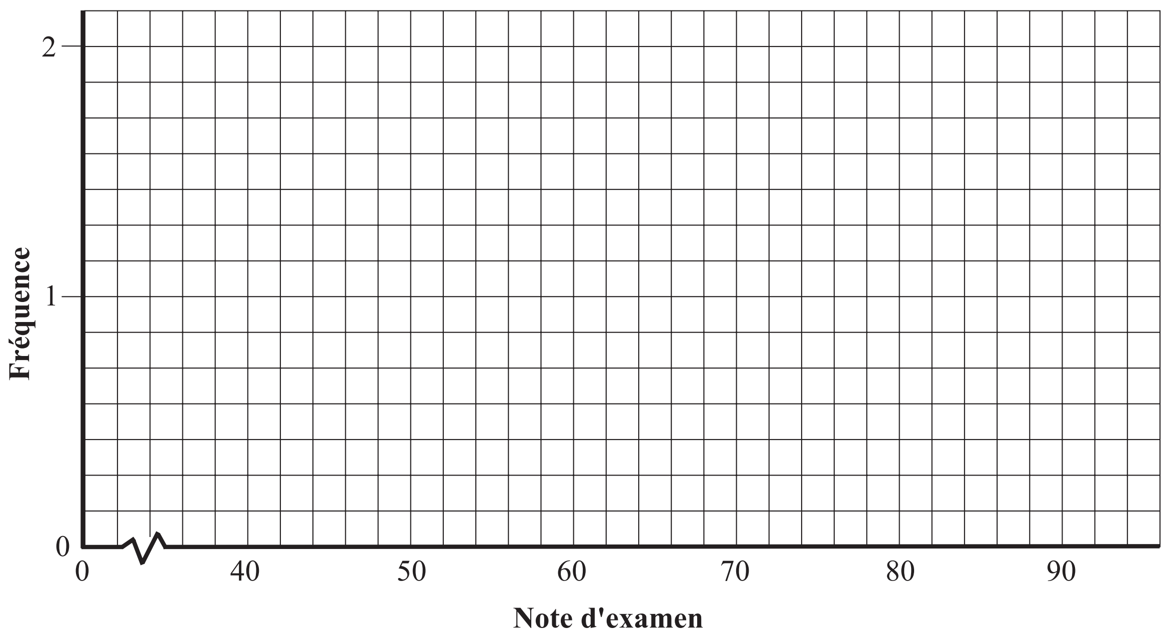
1. Détermine la différence entre chaque note et la moyenne. $(x-\bar{x})$
2. Trouve le carré de ces différences. $(x-\bar{x})^2$
3. Détermine la somme de ces carrés.
4. Divise la somme par $n - 1$.
5. Détermine la racine carrée de ce nombre.

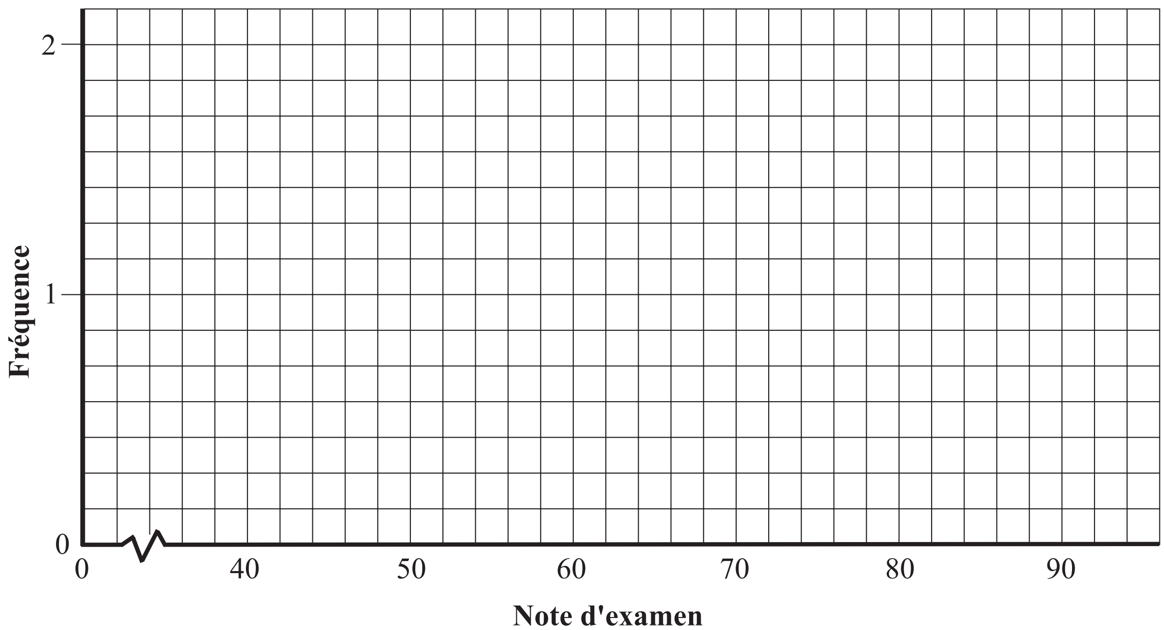
Résultats d'examen (x)	Différence de la moyenne $(x - \bar{x})$	Carré de la différence $(x - \bar{x})^2$
42		
62		
66		
66		
68		
68		
70		
70		
74		
94		
Somme des carrés	$\sum (x - \bar{x})^2$	

Selon les deux exemples, les écoles ont exactement la même moyenne et la même étendue, mais l'écart type diffère. Que représente l'écart type par rapport à cet ensemble de résultats d'examen?

En examinant attentivement les chiffres, nous pouvons voir que les résultats sont beaucoup moins répartis à l'école Mordudesmaths qu'à l'école Einstein. L'écart type reflète l'étendue des résultats à chacune des écoles. Il nous indique la variation des chiffres d'un ensemble. Plus l'écart type est élevé, plus les chiffres sont variés. Moins l'écart type est élevé, moins les chiffres de l'ensemble sont variés. L'écart type, comme l'étendue, est une mesure de variation.

Pour t'aider à visualiser l'étendue des notes à chacune des écoles, trace un histogramme sur la fréquence de chaque école. Tu utiliseras ces données dans la prochaine section.





Exercice 2 : Mesures de variation - écart type

1. Le temps requis en secondes pour que 10 coureurs terminent une course de 400 mètres se lit comme suit : 57, 58, 58, 59, 60, 60, 61, 61, 62, 64

   a) Détermine l'étendue de cet ensemble de temps.

   b) Détermine la moyenne des temps.

   c) Remplis le tableau ci-contre pour déterminer la somme des carrés.

   d) Divise cette somme par $n - 1$.

   e) Calcule la racine carrée de ce nombre.

Temps en secondes (x)	Différence de la moyenne $(x - \bar{x})$	Carré de la différence $(x - \bar{x})^2$
Somme des carrés	$\sum(x - \bar{x})^2$	

2. Le nombre de membres de neuf familles est donné ci-dessous :

   3 8 4 5 1 4 6 3 2

   a) Détermine l'étendue.
   b) Détermine la moyenne.
   c) Remplis le tableau et trouve l'écart type pour cet ensemble de données.

Membres des familles (x)	Différence de la moyenne $(x - \bar{x})$	Carré de la différence $(x - \bar{x})^2$
Somme des carrés	$\sum (x - \bar{x})^2$	

3. L'entraîneur d'une équipe de volley-ball mesure la taille (cm) de ses joueurs et détermine qu'ils mesurent : 172 176 176 178 180 181 181 182 184 184 187

   a) Détermine l'étendue.
   b) Détermine la moyenne.
   c) Remplis le tableau et trouve l'écart type pour cet ensemble de données.

Taille en cm (x)	Différence de la moyenne $(x - \bar{x})$	Carré de la différence $(x - \bar{x})^2$
Somme des carrés	$\sum (x - \bar{x})^2$	

4. Détermine l'écart type de chacun des ensembles de données suivantes. Utilise une feuille de calcul ou une calculatrice graphique si tu en as une. Tu peux aussi utiliser les tableaux vides ci-dessous.

   a) 5 7 9 7 8 6 4 3 14

   b) 6 7 8 8 10 11 11 11 13 14 14 15 16 17 19

   c) 10,8 10,3 10,4 17,3 10,5 10,6 10,7 10,4 11,1

a)

Ensemble de données (a) x	Différence de la moyenne $x - \bar{x}$	Carré de la différence $(x - \bar{x})^2$
Somme des carrés $\sum(x - \bar{x})^2$		

b)

Ensemble de données (b) x	Différence de la moyenne $x - \bar{x}$	Carré de la différence $(x - \bar{x})^2$
Somme des carrés $\sum(x - \bar{x})^2$		

c)

Ensemble des données (c) x	Différence de la moyenne $x - \bar{x}$	Carré de la différence $(x - \bar{x})^2$
Somme des carrés $\sum(x - \bar{x})^2$		

(tableau supplémentaire)

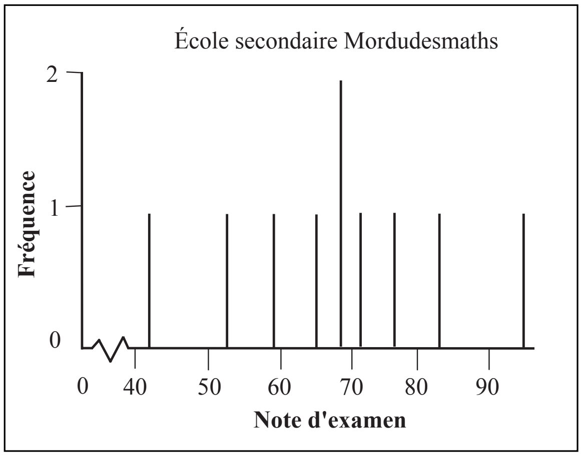
Ensemble de données x	Différence de la moyenne $x - \overline{x}$	Carré de la différence $(x - \overline{x})^2$
Somme des carrés $\sum (x - \overline{x})^2$		

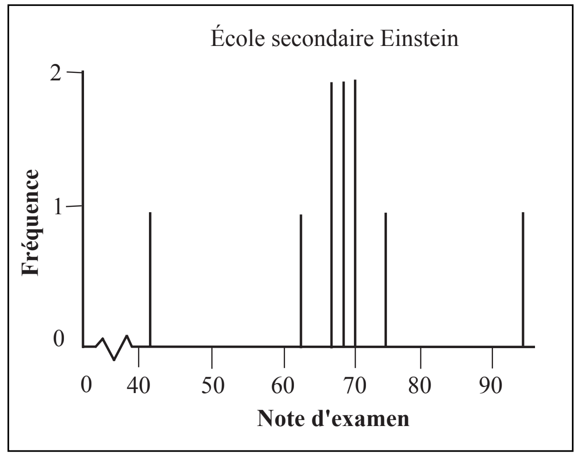
5. Le salaire moyen à une petite usine de fabrication est de 32 000 $ et l'écart type (s_x) de 1 500 $. Le propriétaire veut partager les profits de l'année précédente avec ses employés. Chacun reçoit un boni de 1 000 $. Décris l'effet du boni sur la moyenne et l'écart type.

6. Les élèves d'une classe de Mathématiques du consommateur en secondaire 4 passent un examen et obtiennent une moyenne de 74 et un écart type de 5. Si l'enseignant donne cinq points additionnels à chaque élève, décris l'effet des points additionnels sur la moyenne et l'écart type.

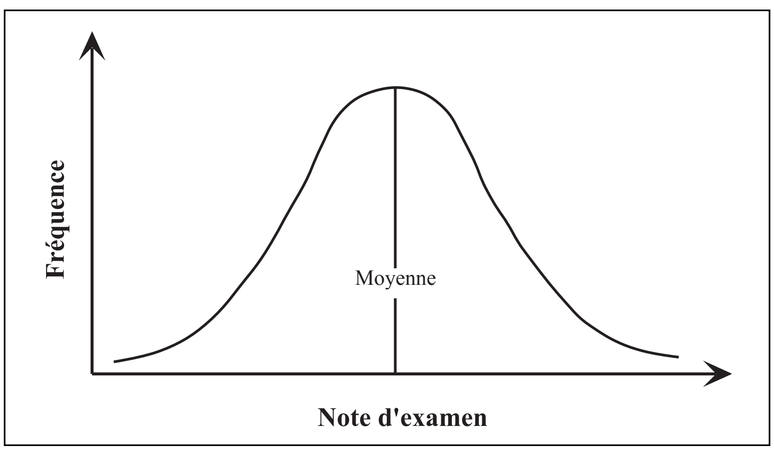
Leçon 3 : Mesures de variation - Distribution normale

Les exemples donnés dans la section qui précède sont fondés sur les résultats d'examen des élèves de l'École secondaire Mordudesmaths et de l'École secondaire Einstein. Tu trouveras ci-dessous des histogrammes de ces résultats. Compare les histogrammes à ceux que tu as tracés à la Leçon 2.





Relie les bâtons en traçant une courbe lisse de manière à joindre le haut des bâtons. La courbe représente la distribution des notes d'examen aux deux écoles. Si tu utilises les résultats d'examen pour des centaines ou des milliers d'élèves, au lieu de seulement 10 élèves, la courbe de l'histogramme ressemblerait à celle qui suit :

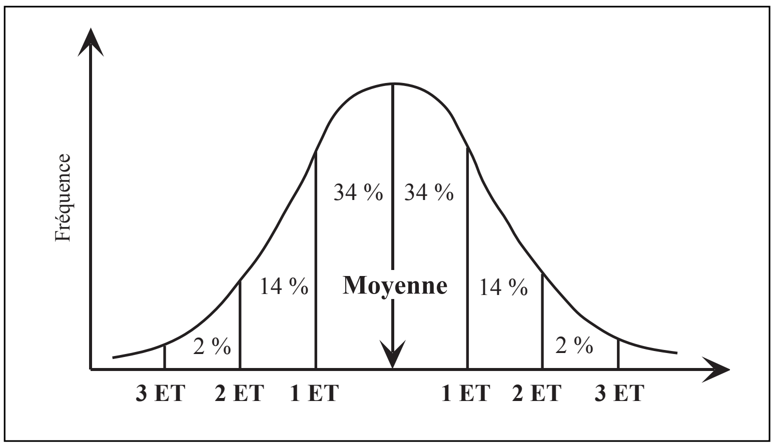


Cette courbe en « forme de cloche » s'appelle une courbe normale. La taille et le poids de personnes, les tests de QI et même la durée de vie de véhicules, pour ne nommer que ceux-là, peuvent représenter des distributions sur un histogramme ayant une courbe normale si les échantillons visent de grands groupes de population. Lorsqu'un histogramme sur la fréquence ressemble à la forme de cloche ci-dessus, le graphe est dit avoir une courbe normale et la distribution de la fréquence s'appelle une distribution normale.

Les caractéristiques d'une courbe normale sont les suivantes :

1. La courbe prend la forme de cloche étendue aux deux extrémités.
2. La moyenne se trouve au centre de la courbe, et la courbe est symétrique des deux côtés de la moyenne, ce qui signifie qu'on peut replier la courbe le long de la ligne indiquant la moyenne et que le côté gauche est égal au côté droit.
3. La moyenne est égale à la médiane, ce qui signifie qu'il existe un nombre égal de données au-dessus et en dessous de la moyenne.
4. Les notes qui forment la distribution normale ont tendance à se regrouper près du centre, et très peu de valeurs se trouvent à plus de trois écarts-types de la moyenne de chaque côté.

Le graphe ci-dessous présente la distribution normale de données. Les pourcentages ont été arrondis pour simplifier les calculs.

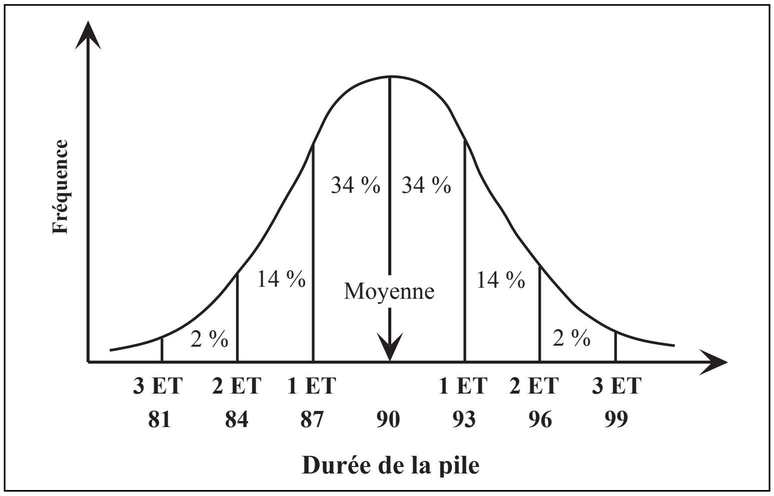


Le graphe démontre que 68 % (34 % + 34 %) de toutes les données se trouvent parmi un écart type de la moyenne. 96 % de toutes les données se trouvent parmi deux écarts-types de la moyenne. Environ 4 % de toutes les données se trouvent parmi trois écarts-types de la moyenne.

Les distributions normales se présentent sous différentes tailles et formes. Certaines sont hautes et minces et d'autres peuvent être plates et étendues. Cependant, toute distribution qui suit une courbe normale aura ces pourcentages pour chaque écart type. Cela nous permet d'examiner une distribution de fréquence normale en connaissant uniquement la moyenne et l'écart type. Prenons l'exemple qui suit.

Exemple 1

La compagnie Piles-à-vie fabrique des piles ayant une durée de distribution normale. Les piles ont une durée moyenne de 90 heures et un écart type de trois heures. La courbe normale représentant ces données ressemble à celle qui suit :



Nous pouvons déterminer certaines informations à partir de cette courbe, notamment :

1. Quel pourcentage des piles ont une durée de moins de 87 heures?

   Réponse : 16 % des piles ont une durée de moins de 87 heures.

2. Quel pourcentage des piles ont une durée de plus de 87 heures et moins de 96 heures?

   Réponse : 82 % des piles ont une durée de plus de 87 heures et moins de 96 heures.

3. Si la compagnie Piles-à-vie fabrique 5 000 piles par jour, combien de piles ont une durée de plus de 93 heures?

   Réponse : 5 000 x 16 % = 800. Donc, 800 piles ont une durée de plus de 93 heures.

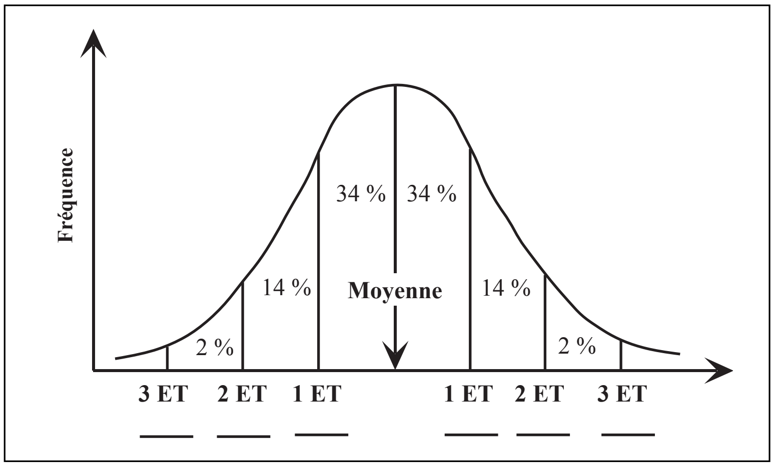
4. Si ton école achète 200 piles, combien d'entre elles ont une durée de moins de 84 heures?

   Réponse : 200 x 2 % = 4. Donc, 4 piles ont une durée de moins de 84 heures.

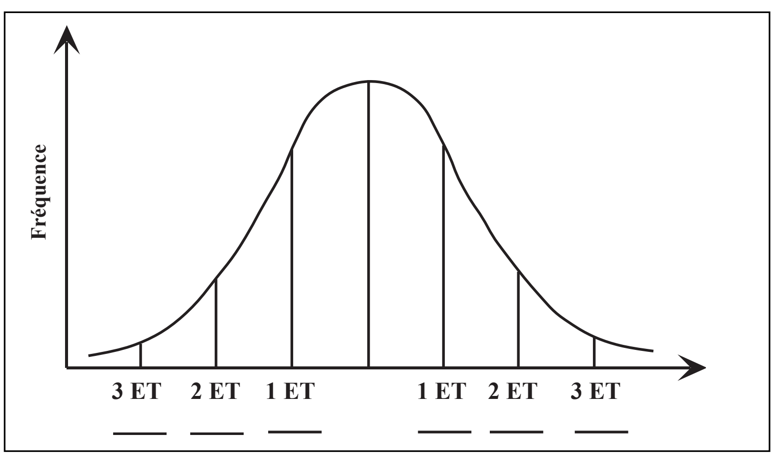
Exercice 3 : Distributions normales

À l'aide d'un graphe à courbe normale, trace la moyenne et quelques écarts-types sur chaque côté de la moyenne. Utilise le graphe pour résoudre chacun des problèmes suivants.

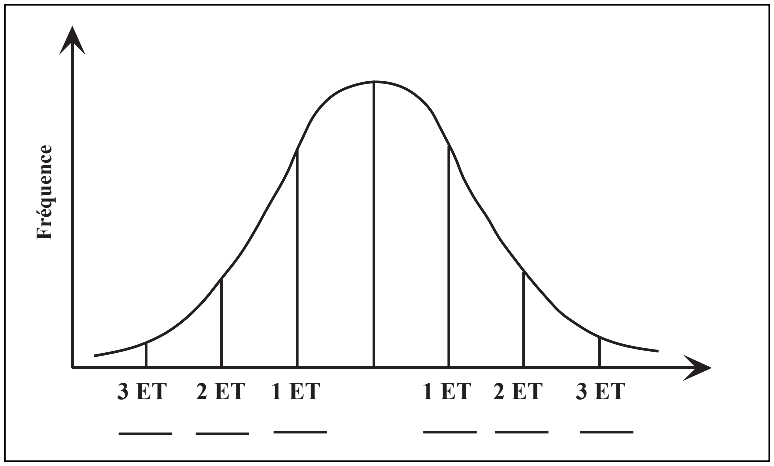
1. Un entraîneur de football dessine un graphe du poids de chacun de ses joueurs et détermine que les données sont distribuées de façon normale. La moyenne est de 100 kg et l'écart type est de 10 kg.
   a) Quel pourcentage des joueurs pèsent entre 90 kg et 110 kg?
   b) Quel pourcentage des joueurs pèsent plus de 120 kg ou moins de 80 kg?
   c) Si l'équipe est composée de 45 joueurs, combien d'entre eux pèsent plus de 110 kg?



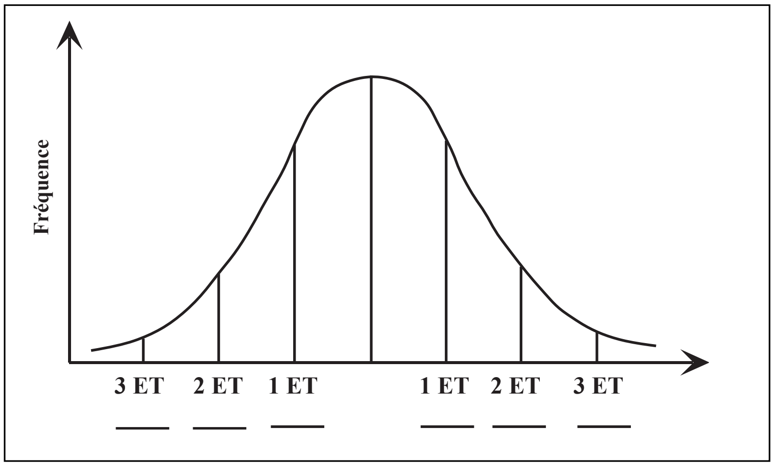
2. Le gestionnaire du personnel d'une entreprise ayant 2 000 employés détermine que la durée d'emploi moyenne est de 22 ans et que l'écart type est de 7 ans. Si les données sont distribuées de façon normale :
   a) Quel pourcentage des employés travailleront pendant plus de 15 ans?
   b) Combien d'employés travailleront entre 15 et 29 ans?
   c) Combien d'employés travailleront pendant plus de 29 ans?



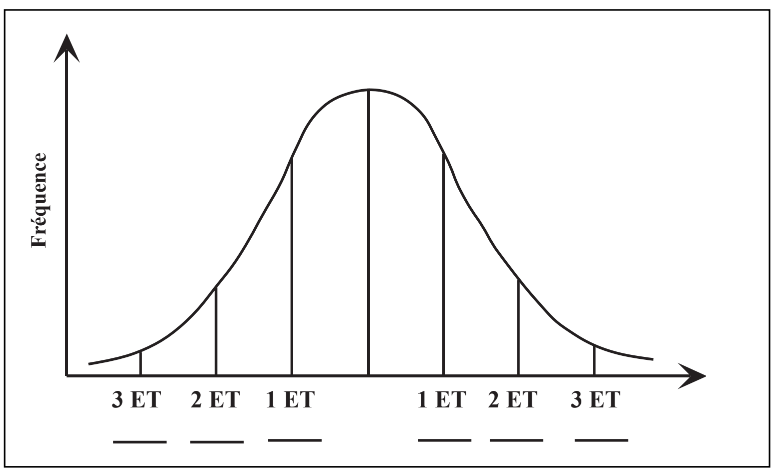
3. La société Croustilles délicieuses produit des sacs de croustilles ayant une masse de 200 g. L'entreprise sait que la quantité de croustilles par sac a une distribution normale ayant une moyenne de 210 g et un écart type de 5 g.
   a) Quel pourcentage des sacs ont une masse moins que 200 g?
   b) Si 30 000 sacs de croustilles sont produits chaque jour, combien de sacs ont une masse de plus de 215 g?
   c) Comment peux-tu expliquer qu'un sac peut avoir une masse de 400 g?



4. La durée moyenne de gestation humaine (nombre de jours entre la conception et la naissance) est estimée à environ 266 jours et a un écart type de 16 jours.
   a) Quel pourcentage des naissances ont lieu au-delà de 282 jours?
   b) Quel pourcentage des naissances ont lieu en moins que 250 jours?
   c) Quel pourcentage des naissances ont lieu à 1 écart type de chaque côté de la moyenne?
   d) Sur 2 500 naissances, combien d'entre elles sont produites d'une grossesse d'une durée de plus de 282 jours?
   e) Sur 2 500 naissances, combien d'entre elles sont produites d'une grossesse d'une durée de moins de 250 jours?

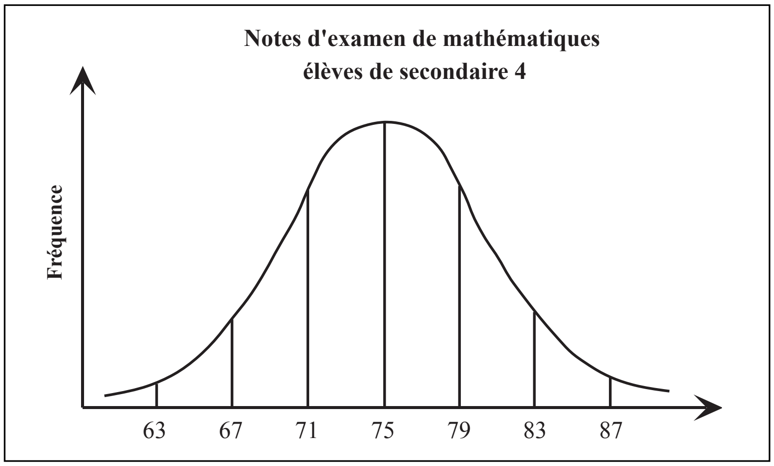


5. Sur l'emballage d'une barre de chocolat populaire, on indique que la masse est de 75 g. Naturellement, la masse véritable varie jusqu'à un certain point, mais elle varie selon une distribution normale ayant une moyenne de 75 g et un écart type de 2 g.
   a) Quelle proportion des barres pèsent moins que la masse indiquée sur l'emballage?
   b) Quelle proportion des barres pèsent plus que 79 g?
   c) Pendant une journée de production, 50 000 barres sont produites. Combien d'entre elles ont une masse variant entre 73 g et 77 g?
   d) Au cours d'une année, si tu achètes 50 de ces barres de chocolat, combien d'entre elles auront une masse moins que 71 g? plus que 79 g?

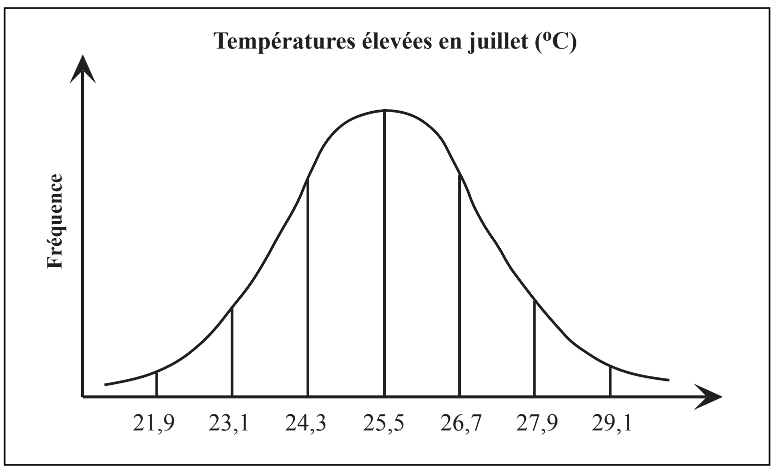


6. Pour chacune des courbes normales suivantes, détermine la moyenne et l'écart type de la distribution. Décris chacune des distributions en quelques lignes.

a)



b)



Leçon 4 : Coefficient de corrélation

Dans cette leçon, tu apprendras comment interpréter un chiffre qui mesure la relation (ou absence de relation) entre deux ensembles de valeurs. Ce chiffre s'appelle le **coefficient de corrélation**, ou « r », et peut être calculé à l'aide d'une formule dont nous traiterons plus loin dans la leçon.

Le coefficient de corrélation, *r*, est utile pour déterminer la force d'une ***relation linéaire*** pouvant exister entre deux ensembles de données. Il est important de noter que les corrélations ne sont pas toutes linéaires, mais pour les besoins de ce cours, nous examinerons uniquement celles qui le sont. Il est possible d'exprimer les données utilisées pour calculer un coefficient de corrélation de deux façons, soit à l'aide d'un tableau, soit à l'aide d'un graphique à points bidimensionnel. Aux fins de cette leçon, nous examinerons trois façons de décrire ces données sur un graphique à points, notamment :

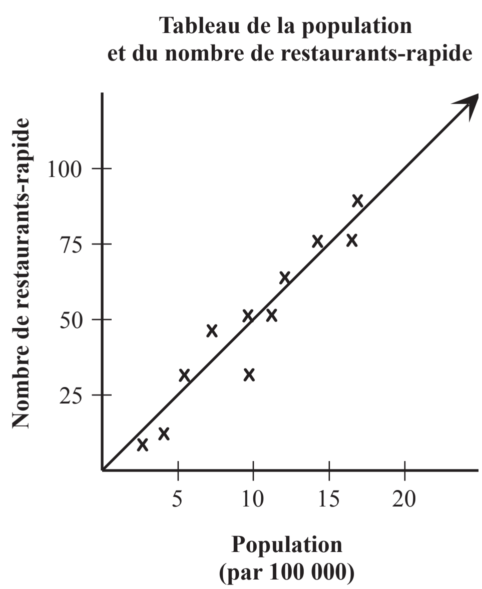
1. **corrélation positive** : si une variable *augmente*, l'autre variable *augmente* aussi, ou si une variable *baisse*, l'autre *baisse* aussi. La valeur *r* servant à décrire une corrélation positive se situe entre les valeurs 0 et +1. Une corrélation positive plus forte est représentée par une valeur *r* plus près de +1. Une corrélation positive plus faible est représentée par une valeur *r* plus près de 0. Cette tendance est exprimée dans l'exemple qui suit.

Exemple 1

La population de diverses villes (sur l'axe *x*) est tracée par rapport au nombre de restaurants-rapide dans la ville (sur l'axe *y*). Généralement, plus il y a d'habitants dans une ville, plus il y a de restaurants-rapide.

Comme l'indique le schéma ci-contre, la plupart des points se situent sur une ligne droite qui s'élève de gauche à droite.

Ce graphique représente une corrélation positive assez forte. En d'autres mots, nous pouvons déduire que plus la population est grande, plus il y a de restaurants-rapide.



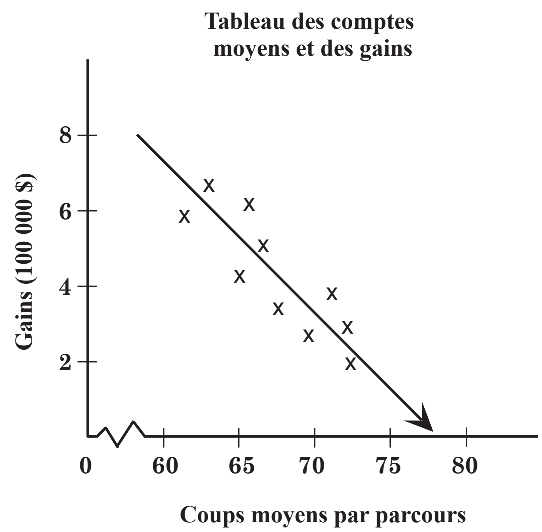
2. **corrélation négative** : si une variable *augmente*, l'autre variable *baisse*, ou si une variable *baisse*, l'autre *augmente*. La valeur *r* servant à décrire une corrélation négative se situe entre les valeurs 0 et –1. Une corrélation négative plus forte a une valeur *r* qui se situe plus près de –1, tandis qu'une corrélation négative plus faible a une valeur *r* qui se situe plus près de 0. Cette tendance est exprimée dans l'exemple qui suit.

Exemple 2

Les coups moyens d'un golfeur sont indiqués sur un graphique par rapport aux prix qu'il a gagnés lors de tournois. Généralement au golf, plus le compte est bas, plus les gains sont élevés et plus le compte est élevé, plus les gains sont moindres.

Comme l'indique le schéma ci-contre, la plupart des points se situent sur une ligne droite qui baisse vers la droite.

Ce graphique représente une corrélation négative assez forte. En d'autres mots, nous pouvons déduire que moins le compte est élevé, plus les gains augmentent.

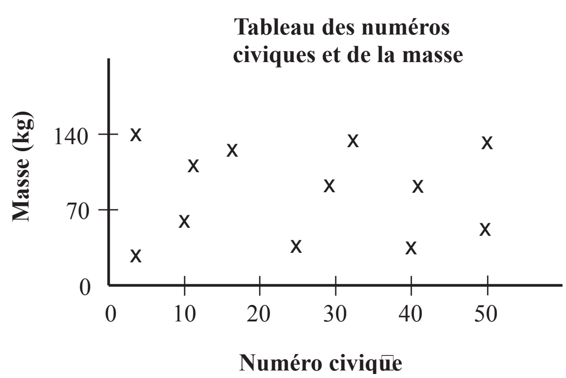


3. **corrélation nulle :** les variables n'ont essentiellement aucun lien commun, donc on ne peut pas dire qu'elles sont en relation. La valeur *r* qui représente une corrélation nulle est très près de zéro ou y est égale. Cette tendance est exprimée dans l'exemple qui suit.

Exemple 3

L'adresse postale (numéro civique) d'une personne est tracée sur un graphique par rapport à la masse en kg de cette personne.

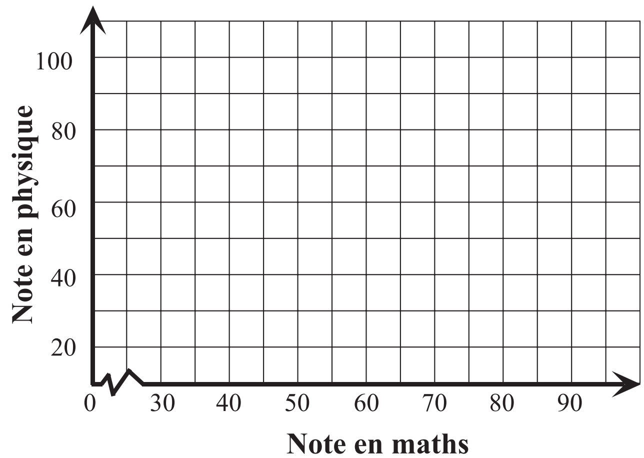
Comme tu peux le voir, il n'y a aucune corrélation positive ou négative entre le numéro civique et la masse.



Exemple 4

L'exemple qui suit représente les résultats d'examen en mathématiques et en physique obtenus par un petit groupe d'élèves en secondaire 4. Semble-t-il y avoir un lien entre le résultat des examens de maths et ceux de physique? Pour trouver cette réponse, représente les données à l'aide d'un graphique et détermine s'il y a une corrélation.

	Note en maths	Note en physique
Élève	(x)	(y)
a	63	56
b	52	54
c	83	86
d	71	75
e	53	58
f	95	87
g	46	52
h	86	90
I	68	66
j	30	38



Comme tu peux le voir, il existe une assez grande corrélation positive entre les deux types de notes. Généralement, plus la note en mathématiques est élevée, plus la note en physique est élevée.

Maintenant, calcule le coefficient de corrélation (r). La formule à utiliser est très complexe. Ton enseignant peut te la montrer. Pour les besoins de cette leçon par contre, nous emploierons la technologie pour calculer la valeur. Ton enseignant te dira comment utiliser une feuille de calculs et un logiciel de statistiques comme WinStat, le mode STATS sur une calculatrice scientifique ou une calculatrice graphique comme la TI-83.

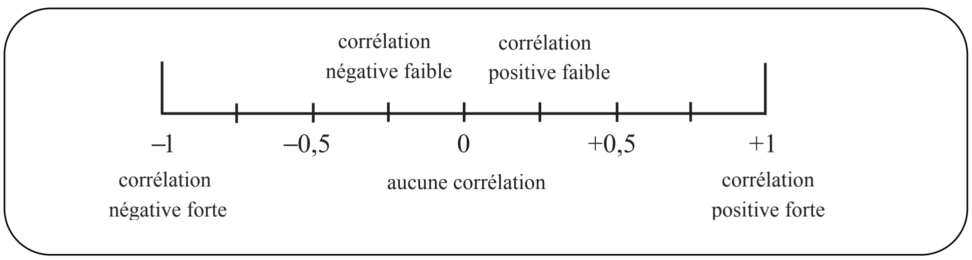
Peu importe la méthode utilisée, le coefficient de corrélation qui relie ces deux variables est égal à : r = +0,97. Il s'agit d'une corrélation positive très forte.

Ces résultats suggèrent que le fait de connaître le niveau de réussite d'un élève sur un test de mathématiques peut prédire leur résultat possible sur un test de physique. Il faut toutefois faire attention en soumettant ce type de théorie, puisque d'autres facteurs pourraient influer sur le résultat d'un élève au test.

Résumé du coefficient de corrélation (*r*)

1. L'étendue des valeurs pour tous les coefficients de corrélation, r, se situe entre –1 et +1.
2. La valeur r de +1 indique une corrélation positive très forte. C'est-à-dire que si une valeur (x) augmente, la valeur (y) augmente aussi. Ou, si une valeur baisse, l'autre baisse aussi.
3. La valeur r de –1 indique une corrélation négative très forte. C'est-à-dire que si une valeur (x) augmente, l'autre valeur (y) baisse. Ou, si une valeur baisse, l'autre valeur augmente.
4. La valeur r de 0 indique une corrélation nulle; c'est-à-dire qu'il n'y a aucune relation entre les deux ensembles de données.
5. La valeur r peut compter un nombre infini de valeurs entre –1 et +1. Les corrélations positives les plus fortes se situent près de +1, tandis que les corrélations négatives les plus fortes se situent plus près de –1. Les corrélations faibles, qu'elles soient positives ou négatives, se situent plus près de zéro.

Le graphique qui suit représente l'étendue possible de la corrélation.



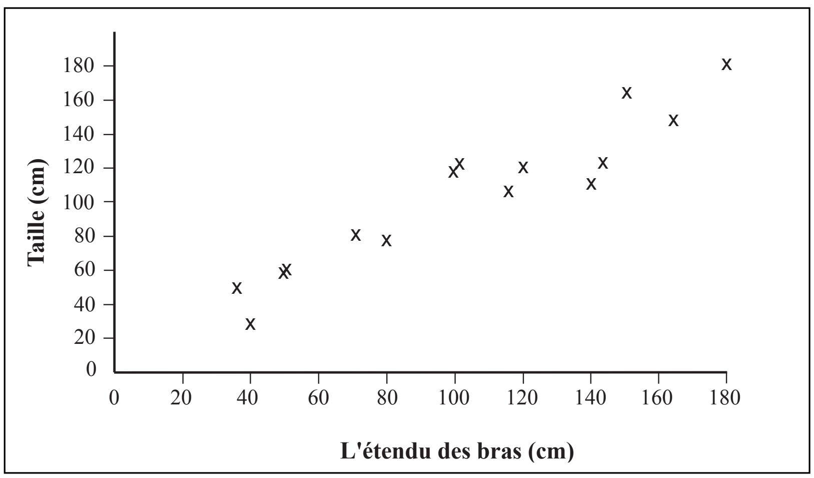
Comme nous l'avons mentionné plus tôt, il faut faire attention lorsque nous attribuons une signification à un coefficient de corrélation calculé. Même si une corrélation existe, cela ne veut pas dire qu'une différence dans une variable entraînera un changement dans l'autre. D'autres facteurs peuvent être absents de l'ensemble de données, ce qui pourrait influer sur la relation entre les variables.

Exercice 4 : Coefficient de corrélation

1. Prédis si chacune des relations suivantes a une corrélation positive, négative ou nulle. De plus, écris une phrase qui décrit cette relation.
   a) population d'un pays et nombre de maisons

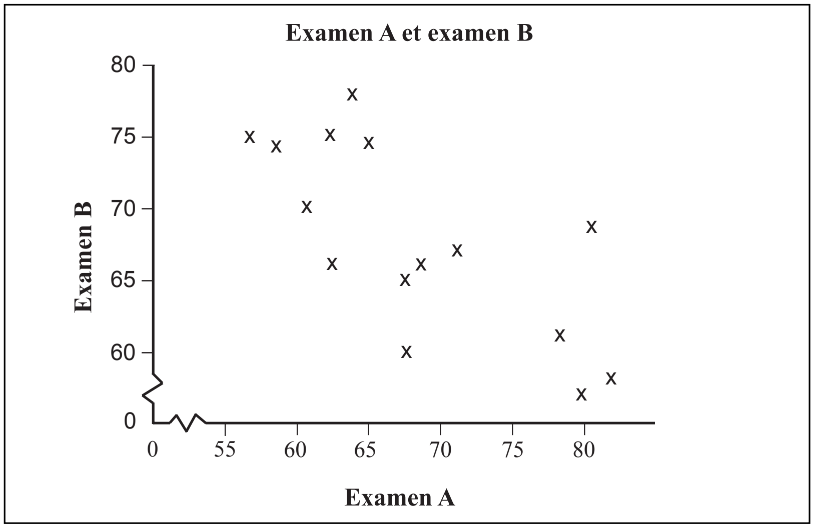
      Réponse : corrélation positive très forte. Plus la population augmente, plus le nombre de maisons augmente.
   b) nombre d'élèves dans une classe et jour de la semaine
   c) âge d'une personne et pointure de chaussures qu'elle porte
   d) prix de vente du bois et quantité de bois achetée
   e) nombre de kilomètres sur l'odomètre d'une voiture et l'âge du véhicule
   f) nombre d'heures passées à étudier pour un examen et note obtenue à l'examen
   g) nombre de bonbons M & M rouge dans 10 sacs et nombre de pages dans 10 livres
   h) nombre d'agrafes dans une agrafeuse et numéro de la pièce où se trouve l'agrafeuse

2. Pour chacun des graphiques ci-dessous, indique le type de corrélation (positive forte, négative faible, etc.) et un énoncé décrivant la relation entre les deux variables, puis choisis un chiffre qui décrit le mieux cette corrélation.
   a) Longueur d'un bras d'une personne (cm) et la taille de cette personne (cm).



choix de la valeur *r*
0,97
0,12
–0,12
–0,97

b) Résultats à l'examen A et résultats à l'examen B



choix de la valeur r
0,75
0,05
– 0,05
– 0,75

3. Dans les situations suivantes, fais les démarches indiquées :
   - crée un graphique à points et indique les valeurs approximatives sur les axes x et y;
   - à partir du nuage des points, détermine le type de corrélation linéaire (+, –, 0);
   - calcule la valeur r à l'aide d'un outil technologique;
   - décris la relation entre les deux variables sous forme d'énoncé.

   a) Nombre d'heures passées à étudier pour un examen et notes des élèves

Élève	Heures d'étude	Note à l'examen
1	8,5	88
2	2	44
3	3,5	63
4	4	58
5	7	89
6	6,5	92
7	3	50
8	8	81
9	5	57
10	6	72

b) Fred (x) et Arthur (y) classent leurs émissions télévisées préférées en ordre de 1 à 10.

Émission	Classement selon Fred	Classement selon Arthur
Hockey de la LNH	1	10
Entre amis	4	7
Millennium	8	4
Aux frontières du réel	10	3
Henri pis sa gang	7	2
Basketball de la NBA	9	1
Third Rock	3	6
Frasier	5	8
Salle d'urgence	2	9
Dateline	6	5

c) Résultats de 10 plongeurs après leur premier plongeon et leur classement final.

Plongeur	Résultat (x)	Classement (y)
1	5,5	1
2	6,6	9
3	6,7	10
4	6,8	2
5	7,1	6
6	7,8	5
7	8	8
8	9,4	4
9	9,8	7
10	9,8	3

d) L'ensemble de données qui suit est tiré du Traffic Collision Statistics Report - 1994 de Voirie et Services gouvernementaux Manitoba. Il représente le groupe d'âge du conducteur par rapport au nombre d'accidents dans lesquels ont été impliquées les personnes de ce groupe d'âge.

Détermine l'âge moyen de chaque groupe d'âge et produis un graphique représentant l'âge moyen par rapport au nombre d'accidents

Implication relative / 1000 conducteurs selon le groupe d'âge - 1994									
Groupe d'âge	16–19	20–24	25–34	35–44	45–54	55–64	65–74	75–84	> 84
Âge moyen (x)									
Nombre d'accidents (y)	145	117	80	63	54	47	41	43	21

4. Une enquête menée dans 25 pays démontre le rapport entre l'espérance de vie dans chaque pays et le nombre de personnes par téléviseur dans chaque pays. Par exemple :

	Espérance de vie	**Nombre de personnes/téléviseur**
Canada	76,4	1,6
Ouganda	51	190

Le coefficient de corrélation entre l'espérance de vie et le nombre de personnes par téléviseur est de –0,81. Comme la corrélation négative est si forte, nous pouvons déduire qu'en envoyant plus de téléviseurs à l'Ouganda, les habitants auraient une espérance de vie plus longue.

Es-tu d'accord avec cette affirmation? Justifie ta réponse.