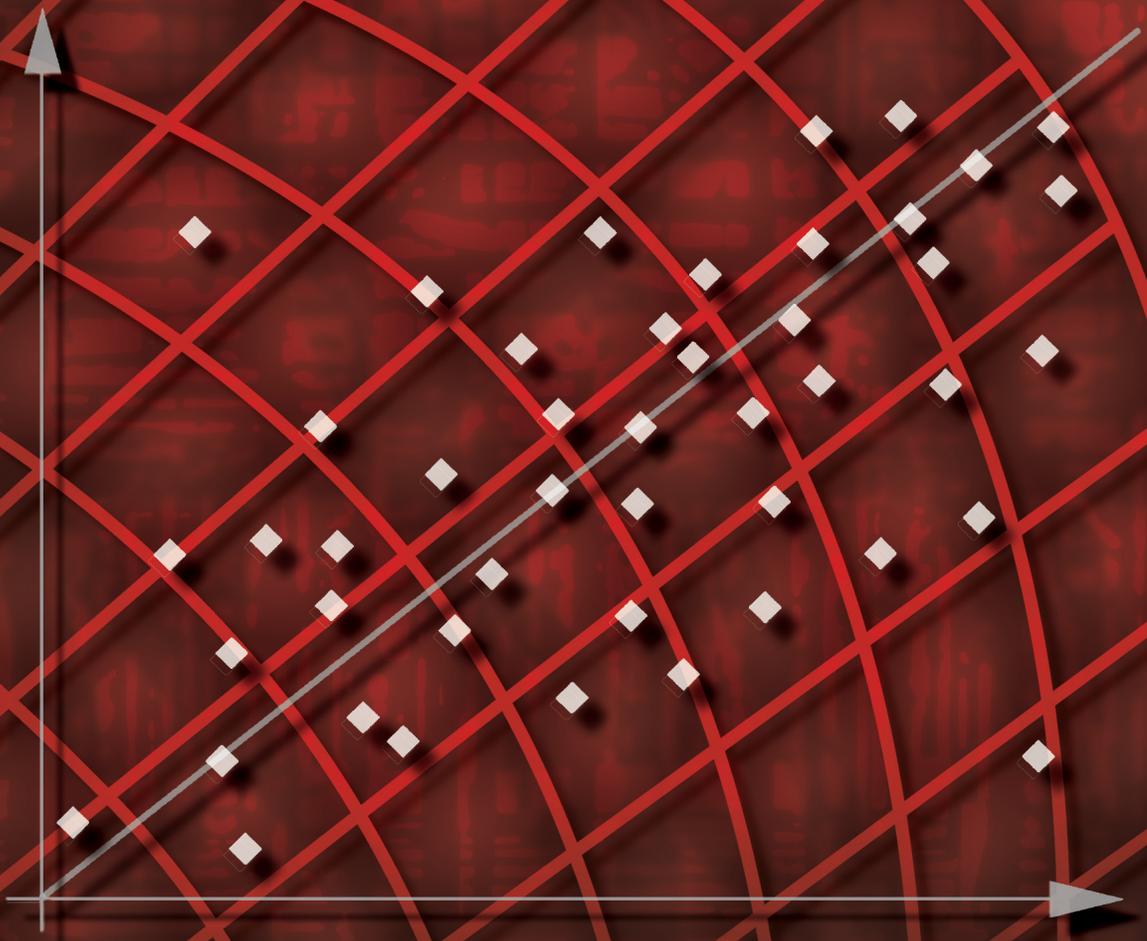


STATISTIQUES III



sofad

MAT-5102-1

STATISTIQUES III

sofad

Coordonnateur des mathématiques : Jean-Paul Groleau

Rédacteur : Alain Malouin

Réviseurs du contenu : Jean-Paul Groleau

Nadine Martin

Révisseuse linguistique : Johanne St-Martin

Édition électronique : P.P.I. inc.

Page couverture : Daniel Rémy

Première édition : 2006

© Société de formation à distance des commissions scolaires du Québec

Tous droits de traduction et d'adaptation, en totalité ou en partie, réservés pour tous pays. Toute reproduction, par procédé mécanique ou électronique, y compris la micro-reproduction, est interdite sans l'autorisation écrite d'un représentant dûment autorisé de la Société de formation à distance des commissions scolaires du Québec (SOFAD).

Dépôt légal — 2006

Bibliothèque et Archives nationales du Québec

Bibliothèque nationale et Archives Canada

ISBN 978-2-89493-301-5

TABLE DES MATIÈRES

Présentation de l'ordinogramme	0.4
Ordinogramme du programme	0.5
Comment utiliser ce guide	0.6
Introduction générale	0.9
Objectifs intermédiaires et terminaux du module	0.10
Épreuve diagnostique sur les préalables	0.15
Corrigé de l'épreuve diagnostique sur les préalables	0.21
Analyse des résultats de l'épreuve diagnostique	0.23
Suivez-vous ce cours en formation à distance?	0.25

SOUS-MODULES

1. Calcul de l'écart moyen, de la variance et de l'écart type d'une distribution de données	1.1
2. Détermination de la cote standard (cote Z) d'une donnée dans une distribution de données	2.1
3. Modes de représentation d'une distribution statistique à deux caractères : le tableau de distribution conjointe et le nuage de points	3.1
4. Coefficient de corrélation linéaire	4.1
5. Droite de régression	5.1
Synthèse finale	6.1
Corrigé de la synthèse finale	6.11
Objectifs terminaux	6.17
Épreuve d'autoévaluation	6.19
Corrigé de l'épreuve d'autoévaluation	6.29
Analyse des résultats de l'épreuve d'autoévaluation	6.33
Évaluation finale	6.35
Corrigé des exercices	6.37
Glossaire	6.85
Liste des symboles	6.90
Bibliographie	6.91
Activités de révision	7.1

PRÉSENTATION DE L'ORDINOGRAMME

BIENVENUE AU ROYAUME DES MATHÉMATIQUES!

Ce programme de mathématiques a été élaboré pour la clientèle adulte des Services d'éducation des adultes des commissions scolaires et de la formation à distance. Les activités d'apprentissage qu'il contient ont été conçues pour être réalisées en apprentissage individualisé. Toutefois, si vous éprouvez des difficultés, n'hésitez pas à consulter votre formatrice ou votre formateur ou à téléphoner à la personne-ressource qui vous a été assignée. Le tableau qui suit situe dans le programme le module que vous avez entre les mains. Il vous permet de visualiser le chemin parcouru ou qui vous reste à parcourir selon l'objectif professionnel que vous poursuivez. Suivant les exigences de votre objectif professionnel, plusieurs voies de sortie du royaume des mathématiques sont prévues.

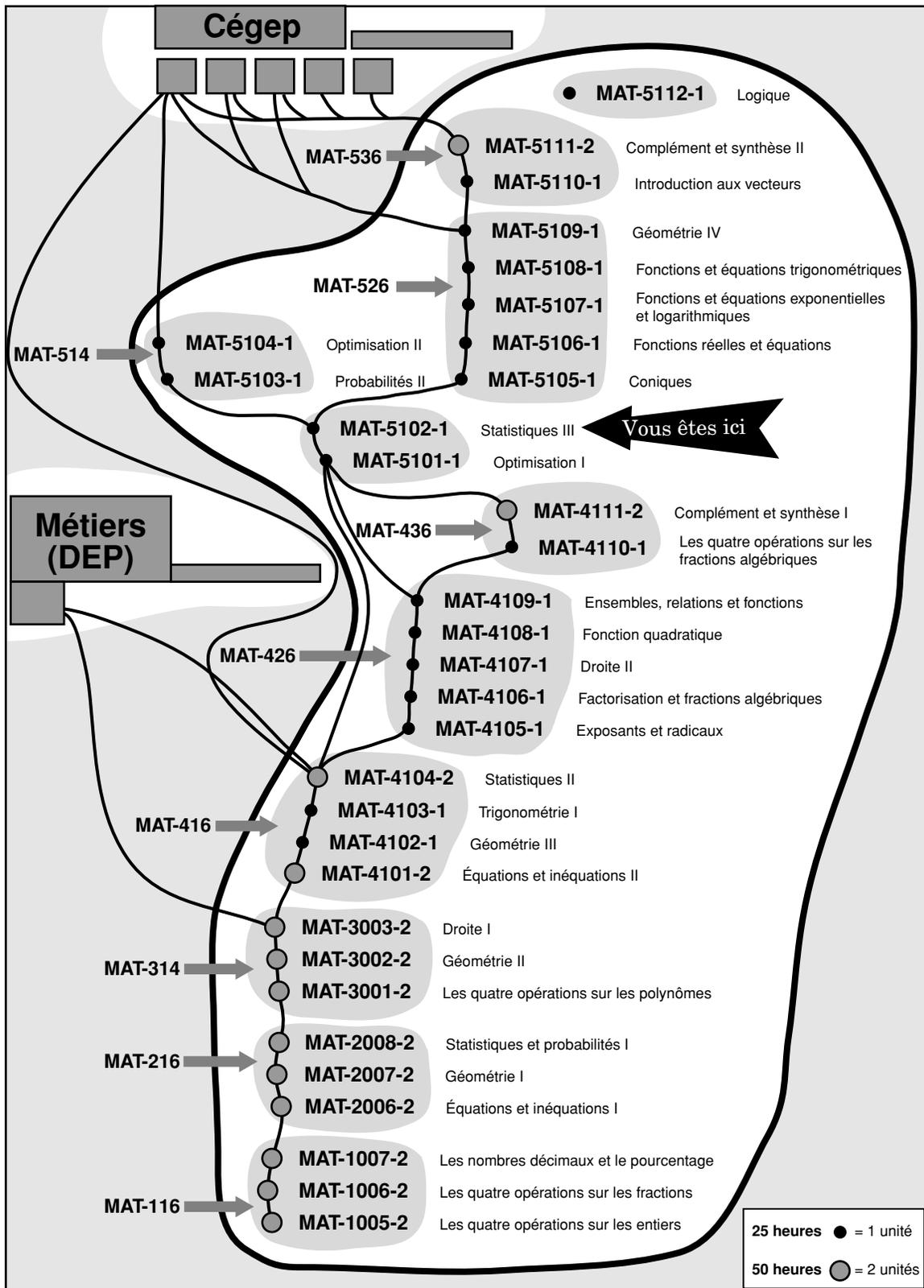
Les premières voies, les routes MAT-3003-2 (MAT-314) et MAT-4104-2 (MAT-416), vous permettent d'entreprendre des études menant à un diplôme d'études professionnelles (DEP) et certains programmes de niveau collégial (cégep) pour la route MAT-4104-2.

Les routes MAT-4109-1 (MAT-426), MAT-4111-2 (MAT-436) et MAT-5104-1 (MAT-514), vous permettent d'obtenir un diplôme d'études secondaires (DES) qui donne accès à certains programmes d'études collégiales (cégep) n'exigeant pas de compétences particulières en mathématiques avancées.

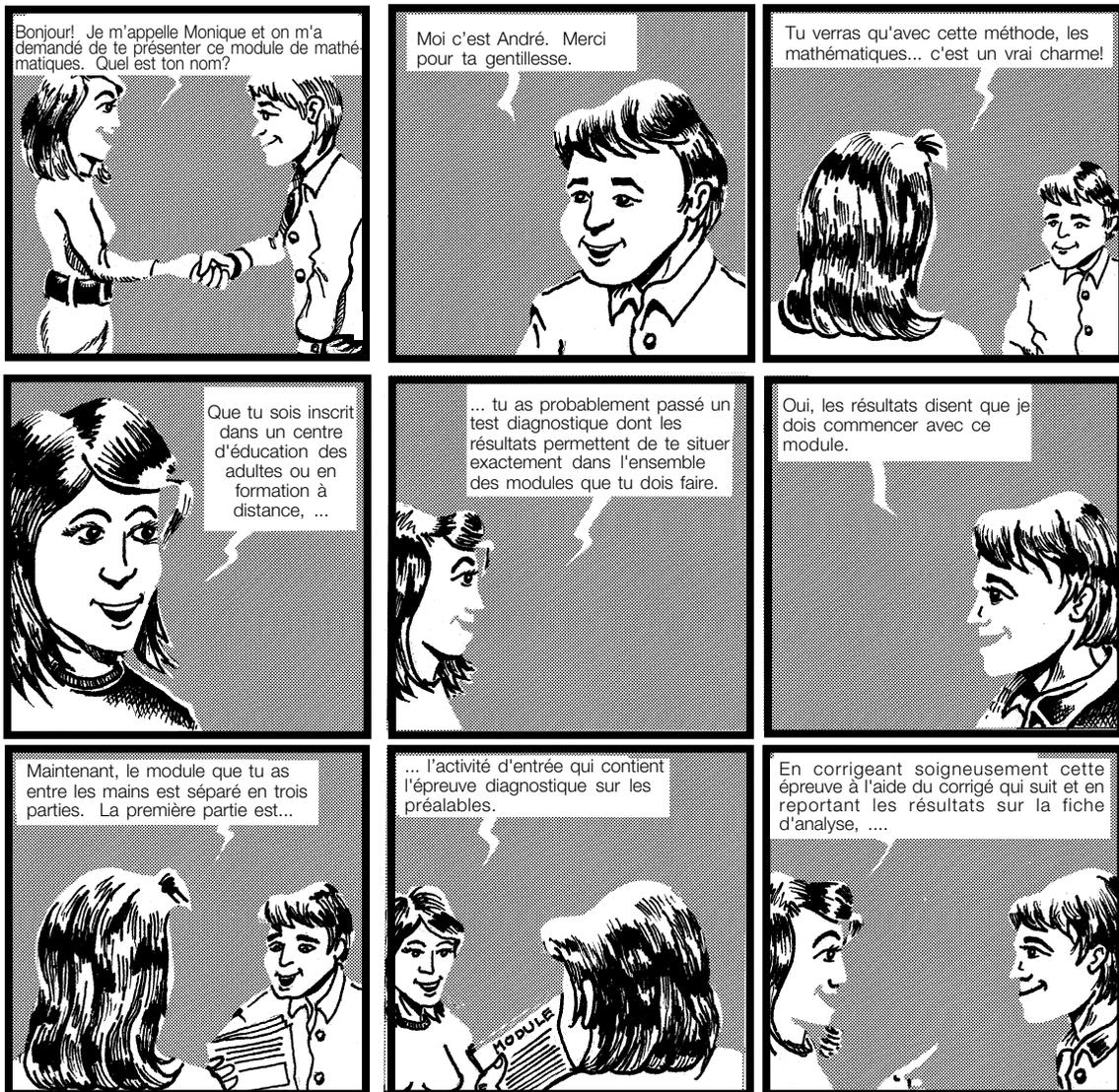
Finalement, les routes MAT-5109-1 (MAT-526) et MAT-5111-2 (MAT-536) vous permettent d'accéder au niveau collégial (cégep) dans des programmes qui exigent de solides connaissances en mathématiques et où d'autres défis vous attendent. Bonne route!

S'il s'agit de votre premier contact avec ce programme de mathématiques, après avoir examiné l'ordinogramme du programme, lisez la section intitulée « Comment utiliser ce guide »; sinon, passez directement à la section intitulée « Introduction générale ». Bon travail!

ORDINOGRAMME DU PROGRAMME



COMMENT UTILISER CE GUIDE





La ligne de départ montre le **début** de l'apprentissage.

? Le petit point d'interrogation blanc identifie les **questions** dont les réponses sont à l'intérieur du texte.

La cible signale l'**objectif** à atteindre.

Le bloc-notes indique un **rappel** des notions que tu as étudiées auparavant.

Le point d'interrogation en gras identifie les **exercices** de consolidation qui te permettront de mettre en pratique ce que tu viens d'apprendre.

La **calculatrice** te rappelle à quel moment t'en servir.

La gerbe de blé identifie une **synthèse** qui te permet de faire le point sur ce que tu viens d'apprendre. Ce logo répété plusieurs fois signifie que tu approches de la fin du module. C'est la synthèse finale qui te permet de faire le lien entre tous les apprentissages du module.

Enfin, la ligne d'arrivée indique qu'il est temps de passer à l'**autoévaluation** pour vérifier si tu as bien assimilé les apprentissages réalisés.



INTRODUCTION GÉNÉRALE

DISTRIBUTIONS STATISTIQUES À UN OU À DEUX CARACTÈRES

Le module comporte deux grandes parties. La première aborde l'étude des distributions statistiques à un caractère ou à une variable. Pour ce faire, nous allons apprendre à calculer certaines mesures de dispersion telles que l'étendue, l'intervalle semi-interquartile, l'écart moyen, la variance, l'écart type, ainsi qu'une mesure de position très importante, soit la cote standard ou la cote Z . Puisqu'il s'agit de la suite du cours *Statistiques II*, nous allons aussi revoir comment calculer les mesures de tendance centrale (moyenne, mode et médiane) et l'étendue interquartile. Enfin, nous allons revoir comment extraire les renseignements pertinents à partir de diagrammes à tiges et à feuilles et de diagrammes de quartiles.

La deuxième partie du cours porte sur les distributions statistiques à deux caractères ou variables. Nous y traiterons du tableau de distribution conjointe, du nuage de points, du coefficient de corrélation linéaire ainsi que de la droite de régression.

Ce nouveau module suppose implicitement d'illustrer au fur et à mesure les concepts présentés à l'aide de situations concrètes. L'accent est mis sur l'analyse et l'interprétation des mesures statistiques plutôt que sur leur calcul. À cet effet, l'emploi d'une calculatrice avec fonctions statistiques à 1 et à 2 caractères, la calculatrice graphique et l'ordinateur sont privilégiés.



OBJECTIFS INTERMÉDIAIRES ET TERMINAUX DU MODULE

Le module MAT-5102-1 comporte 17 objectifs et prévoit une durée d'apprentissage de 25 heures réparties comme dans le tableau ci-dessous. Les objectifs terminaux sont en caractères gras.

Objectifs	Nombres d'heures*	% (évaluation)
1 à 5	5	20 %
6	3	15 %
7 à 12	6	25 %
13 à 16	6	25 %
17	3	15 %

* Deux heures sont réservées à l'évaluation finale.

1. Calcul de l'écart moyen d'une distribution

À partir d'une distribution statistique à un caractère (variable), calculer l'écart moyen en appliquant la formule appropriée.

2. Calcul de la variance et de l'écart type d'une distribution

À partir des données recueillies par le biais d'un sondage ou d'une enquête portant sur un seul caractère, calculer la variance ainsi que l'écart type de la distribution en appliquant la formule appropriée.

3. Comparaison des différentes mesures de dispersion

À partir d'une distribution de données statistiques à un caractère, comparer l'information donnée par différentes mesures de dispersion : étendue, intervalle semi-interquartile, écart moyen et écart type.

4. Cote standard (cote Z)

Déterminer la cote standard (cote Z) d'une donnée dans une distribution statistique à un caractère en appliquant la formule appropriée.

5. Comparaison des données à l'aide de la cote standard

Étant donné une ou plusieurs distributions statistiques à un caractère, comparer des données en interprétant la cote standard (cote Z) de chaque donnée.

6. Résolution de problèmes sur les mesures de tendance centrale et de dispersion

Résoudre des problèmes faisant intervenir une ou plusieurs distributions de données à un caractère. La résolution du problème nécessite l'analyse de l'écart type de la cote standard ou de la moyenne des données concernées.

7. Construction d'un tableau de distribution conjointe

Construire le tableau de distribution conjointe correspondant à une distribution de données statistiques à deux caractères statistiques quantitatifs.

8. Construction d'un nuage de points

Construire un nuage de points correspondant à une distribution de données statistiques à deux caractères statistiques quantitatifs.

9. Description d'une corrélation linéaire

À partir du nuage de points représentant une distribution à deux caractères, décrire la corrélation linéaire qui existe entre ces deux caractères :

- par l'orientation et la distribution des points, déterminer le sens de la corrélation (nulle, positive ou négative) ainsi que son intensité (forte, moyenne, faible ou nulle)
- associer le nuage de points à une valeur approximative du coefficient de corrélation.

10. Estimation du coefficient de corrélation

Par la méthode du rectangle ou de l'ellipse, estimer le coefficient de corrélation linéaire existant entre les deux caractères d'une distribution statistique, présentée à l'aide d'un tableau de données ou d'un nuage de points.

11. Calcul du coefficient de corrélation

À partir des données représentant une distribution statistique à deux caractères, calculer le coefficient de corrélation linéaire existant entre ces deux caractères par une méthode numérique.

12. Interprétation du coefficient de corrélation

En fonction d'un contexte donné, estimer ou calculer, puis interpréter le coefficient de corrélation linéaire existant entre les deux caractères statistiques quantitatifs d'une distribution de données.

13. Estimation de la droite de régression

Estimer la position de la droite de régression à partir du nuage de points représentant une distribution statistique à deux caractères statistiques quantitatifs.

14. Détermination de l'équation de la droite de régression

Par une méthode numérique, déterminer de façon approximative l'équation de la droite de régression.

15. Détermination de l'équation de la droite de régression et calcul du coefficient de corrélation

À partir des données représentant une distribution statistique à deux caractères, déterminer l'équation de la droite de régression et calculer le coefficient de corrélation à l'aide d'un outil technologique.

16. Interprétation de la droite de régression

À l'aide de la droite de régression, prédire les données correspondant à un couple non représenté dans le nuage de points ou dans le tableau d'une distribution statistique à deux caractères.

17. Résolution de problèmes à partir de distributions statistiques à deux caractères

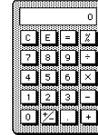
Résoudre des problèmes issus de situations faisant intervenir une ou plusieurs distributions statistiques à deux caractères statistiques quantitatifs.

Ce module comportant 17 objectifs, nous avons regroupé leur étude tel qu'il est indiqué dans le tableau ci-dessous.

Sous-module	Objectif(s)	
1	Calcul de l'écart moyen d'une distribution Calcul de la variance et de l'écart type d'une distribution Comparaison des différentes mesures de dispersion	1 2 3
2	Cote standard (cote Z) Comparaison des données à l'aide de la cote standard Résolution de problèmes sur les mesures de tendance centrale et de dispersion	4 5 6
3	Construction d'un tableau de distribution conjointe Construction d'un nuage de points	7 8
4	Description d'une corrélation linéaire Estimation du coefficient de corrélation Calcul du coefficient de corrélation Interprétation du coefficient de corrélation	9 10 11 12
5	Estimation de la droite de régression Détermination de l'équation de la droite de régression Détermination de l'équation de la droite de régression et calcul du coefficient de corrélation Interprétation de la droite de régression Résolution de problèmes à partir de distributions statistiques à deux caractères	13 14 15 16 17

ÉPREUVE DIAGNOSTIQUE SUR LES PRÉALABLES**Consignes**

- 1° Répondez autant que possible à toutes les questions.
- 2° L'utilisation de la calculatrice est permise.
- 3° Ayez en main une règle graduée en centimètres et en millimètres.
- 4° Inscrivez vos réponses directement sur la feuille.
- 5° Ne perdez pas de temps. Si vous ne pouvez répondre à une question, passez immédiatement à la suivante.
- 6° Dès que vous aurez répondu à toutes les questions auxquelles il vous est possible de répondre, corrigez les réponses à l'aide du corrigé qui suit l'épreuve diagnostique.
- 7° Vos représentations graphiques doivent être équivalentes à celles qui sont proposées dans le corrigé. Les échelles que vous avez choisies peuvent cependant être différentes de celles du corrigé. L'important est que vous fassiez un usage optimal de l'espace qui vous est alloué et que vos données soient placées de façon adéquate. En cas de doute, consultez une personne-ressource.
- 8° Transcrivez vos résultats sur la fiche d'analyse des résultats de l'épreuve diagnostique qui suit le corrigé.
- 9° Prenez connaissance des activités de révision proposées pour chacune des réponses incorrectes.
- 10° Si toutes vos réponses sont exactes, vous possédez les préalables nécessaires pour entreprendre l'étude de ce module.



1. Parmi les distributions de données décrites ci-dessous, déterminez si elles représentent un caractère statistique qualitatif ou un caractère statistique quantitatif. Cochez (✓) la case correspondant à votre choix.

	Qualitatif	Quantitatif
a) L'âge de chacune des personnes étudiant à votre centre d'éducation des adultes.	<input type="checkbox"/>	<input type="checkbox"/>
b) Les points accordés à chacune des candidates lors d'un concours de beauté.	<input type="checkbox"/>	<input type="checkbox"/>
c) La masse en kilogrammes de chacun des joueurs d'une équipe de la LNH.	<input type="checkbox"/>	<input type="checkbox"/>
d) La qualité de l'eau des lacs du Québec pour la baignade pendant l'été 2005 évaluée selon une échelle de 1 à 10.	<input type="checkbox"/>	<input type="checkbox"/>
e) Le nombre de participants du Tour de France de l'année 2005.	<input type="checkbox"/>	<input type="checkbox"/>

2. Voici les résultats obtenus par 19 élèves inscrits au cours de Mathématiques immobilières 1.

63 %, 59 %, 67 %, 84 %, 54 %, 62 %, 58 %, 65 %, 55 %, 61 %, 65 %, 70 %, 77 %, 65 %, 61 %, 57 %, 68 %, 60 %, 64 %.

a) Calculez la moyenne arithmétique de cette distribution.

.....

.....

b) Calculez la médiane de cette distribution.

.....

.....

.....

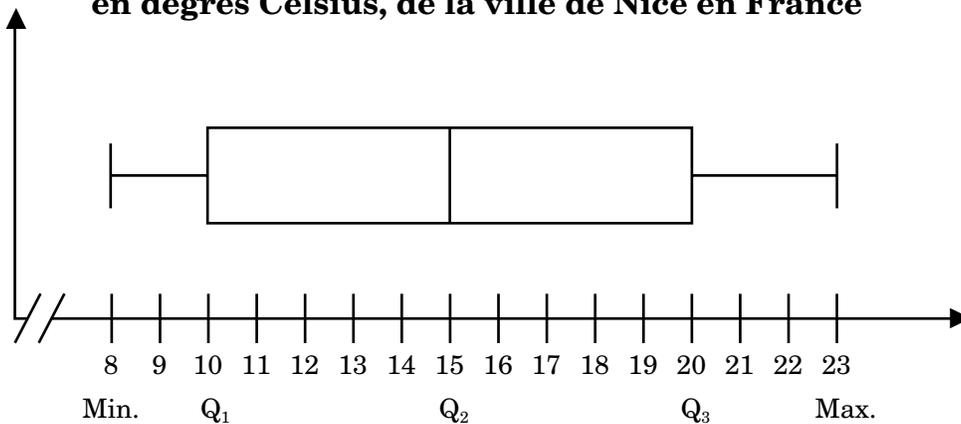
.....

c) Calculez le mode de cette distribution.

.....

3. Le diagramme de quartiles qui suit représente le relevé de températures mensuelles moyennes, en degrés Celsius, de la ville de Nice en France.

Relevé de températures mensuelles moyennes, en degrés Celsius, de la ville de Nice en France



a) Déterminez l'étendue de cette distribution.

.....

b) Déterminez l'étendue interquartile de cette distribution.

.....

c) Déterminez la médiane de cette distribution.

.....

4. Trouvez la pente des droites suivantes.

a) $2x - 3y - 2 = 0$

b) $y = -\frac{x}{4} + 13$

.....

5. Quelle est la pente d'une droite qui passe par les points suivants?

a) $(4, -2)$ et $(-1, 0)$.

.....

b) $\left(\frac{1}{2}, \frac{1}{3}\right)$ et $\left(0, \frac{2}{3}\right)$.

.....

6. Placez sur le plan cartésien ci-dessous les points dont les coordonnées sont fournies.

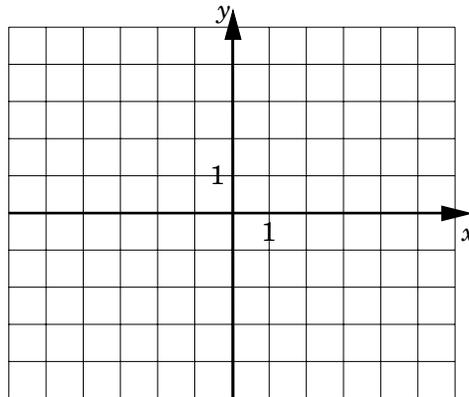
$$A(2, 4)$$

$$B\left(3, -\frac{5}{2}\right)$$

$$C(-2, 0)$$

$$D(0, 0)$$

$$E\left(-\frac{1}{2}, -4\right)$$



CORRIGÉ DE L'ÉPREUVE DIAGNOSTIQUE SUR LES PRÉALABLES

1. a) Caractère statistique quantitatif.
 b) Caractère statistique qualitatif.
 c) Caractère statistique quantitatif.
 d) Caractère statistique qualitatif.
 e) Caractère statistique quantitatif.

2. a) $63\% + 59\% + 67\% + 84\% + 54\% + 62\% + 58\% + 65\% + 55\% + 61\% + 65\% + 70\% + 77\% + 65\% + 61\% + 57\% + 68\% + 60\% + 64\% = 1\,215\%$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1\,215\%}{19} = 63,95\%$$

 b) Pour trouver la médiane, nous devons d'abord mettre les données en ordre.
 $54\%, 55\%, 57\%, 58\%, 59\%, 60\%, 61\%, 61\%, 62\%, \mathbf{63\%}, 64\%, 65\%, 65\%, 65\%, 67\%, 68\%, 70\%, 77\%, 84\%$.

$$Md = \frac{(19 + 1)^e}{2} \text{ donnée} = 10^e \text{ donnée} = 63\%$$

 c) $Mo = 65\%$, car c'est la donnée la plus fréquente.

3. a) $E = 23\text{ °C} - 8\text{ °C} = 15\text{ °C}$
 b) $EI = 20\text{ °C} - 10\text{ °C} = 10\text{ °C}$
 c) $Md = Q_2 = 15\text{ °C}$

4. a) $2x - 3y - 2 = 0$
 $-3y = -2x + 2$
 $y = \frac{2}{3}x - \frac{2}{3}$

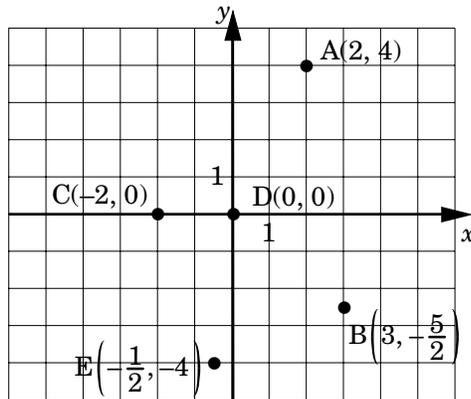
 b) $y = -\frac{x}{4} + 13$
 $m = -\frac{1}{4}$

$$\therefore m = \frac{2}{3}$$

5. a) $m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{0 - (-2)}{-1 - 4} = \frac{0 + 2}{-1 - 4} = \frac{2}{-5} = -\frac{2}{5}$

b) $m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{\frac{2}{3} - \frac{1}{3}}{0 - \frac{1}{2}} = \frac{\frac{1}{3}}{-\frac{1}{2}} = \frac{1}{3} \times \left(\frac{-2}{1}\right) = -\frac{2}{3}$

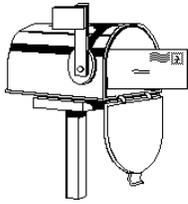
6.



ANALYSE DES RÉSULTATS DE L'ÉPREUVE DIAGNOSTIQUE

Questions	Réponses		Révision		À faire avant
	Correctes	Incorrectes	Section	Page	
1.			7.1	7.4	Sous-module 1
2. a)			7.2	7.6	Sous-module 1
b)			7.2	7.6	Sous-module 1
c)			7.2	7.6	Sous-module 1
3. a)			7.3	7.11	Sous-module 1
b)			7.3	7.11	Sous-module 1
c)			7.3	7.11	Sous-module 1
4. a)			7.4	7.14	Sous-module 3
b)			7.4	7.14	Sous-module 3
5. a)			7.5	7.14	Sous-module 3
b)			7.5	7.14	Sous-module 3
6.			7.5	7.21	Sous-module 3

- Si toutes vos réponses sont **correctes**, vous possédez les préalables nécessaires pour entreprendre l'étude de ce module.
- Pour chaque réponse **incorrecte**, référez-vous aux activités proposées dans la colonne « **Révision** ». Effectuez les activités de révision avant d'entreprendre l'étude de chaque sous-module proposée dans la colonne de droite « **À faire avant** ».



SUIVEZ-VOUS CE COURS EN FORMATION À DISTANCE ?

Vous avez présentement entre les mains le matériel didactique du cours MAT-5102-1 ainsi que les devoirs qui s'y rattachent. À ce matériel est jointe une lettre de votre tutrice ou de votre tuteur. Cette lettre vous indique les différents canaux par lesquels vous pourrez communiquer avec elle ou lui (lettre, téléphone, etc.) ainsi que les heures réservées à ces prises de contact. En plus de corriger vos travaux, la tutrice ou le tuteur est la personne-ressource qui vous aidera dans votre apprentissage. Donc, n'hésitez pas à faire appel à ses services si vous éprouvez quelque difficulté.

UNE MÉTHODE GÉNÉRALE DE TRAVAIL

L'enseignement à distance est un processus d'apprentissage d'une grande souplesse, mais il exige de votre part un engagement actif. Il requiert en effet de la régularité dans l'étude et un effort soutenu. Une méthode efficace de travail vous facilitera la tâche. Un cheminement d'apprentissage constant et productif ne peut échapper aux règles suivantes.

- Fixez-vous un horaire qui vous permet d'étudier selon vos possibilités tout en tenant compte de vos loisirs et de vos activités.
- Astreignez-vous à une étude régulière et assidue.

Pour vous aider à réussir ce cours de mathématiques, voici quelques règles à suivre concernant la théorie, les exemples, les exercices et les devoirs.

La théorie

Pour assimiler correctement les notions théoriques, portez attention aux points suivants.

- 1° Lisez attentivement le texte et surlignez les points importants.
- 2° Mémorisez les définitions, les formules et les marches à suivre pour résoudre un problème donné; cela facilitera la compréhension du texte.
- 3° Notez, à la fin du devoir, les points que vous ne comprenez pas. Votre tutrice ou votre tuteur vous donnera alors des explications pertinentes.
- 4° Essayez de poursuivre votre étude même si vous butez sur un obstacle particulier. Cependant, si une difficulté importante vous empêche de poursuivre la démarche d'apprentissage, n'attendez pas d'envoyer votre devoir pour demander des explications : adressez-vous à la tutrice ou au tuteur selon les modalités prévues dans sa lettre.

Les exemples

Les exemples sont des applications de la théorie. Ils illustrent le cheminement à suivre pour résoudre les exercices. Aussi, étudiez attentivement les solutions proposées dans les exemples et refaites-les pour vous-même avant d'entreprendre les exercices.

Les exercices

Les exercices d'un sous-module respectent généralement le modèle des exemples donnés. Voici quelques suggestions pour réussir ces exercices.

- 1° Rédigez les solutions en prenant pour modèle les exemples présentés dans le texte. Il est important de ne pas consulter le corrigé qui se trouve à la fin du texte sur des feuilles de couleur avant d'avoir terminé les exercices.
- 2° Évaluez vos solutions à l'aide du corrigé uniquement après avoir fait tous les exercices. **Attention!** Vérifiez attentivement les étapes de votre solution, même si votre réponse est exacte.
- 3° Si vous relevez une erreur dans votre réponse ou votre solution, revoyez les notions que vous n'avez pas comprises ainsi que les exemples qui s'y rattachent. Ensuite, recommencez l'exercice.
- 4° Assurez-vous d'avoir réussi tous les exercices d'un sous-module avant de passer au suivant.

Les devoirs

Le cours MAT-5102-1 comprend trois devoirs. La première page de chaque devoir indique à quels sous-modules se rapportent les questions posées. Les devoirs servent à évaluer votre degré de compréhension de la matière étudiée. Ils sont également un moyen de communication avec votre tutrice ou votre tuteur.

Quand vous aurez assimilé la matière et réussi les exercices qui s'y rattachent, rédigez sans délai le devoir correspondant.

- 1° Faites d'abord un brouillon. Apportez à vos solutions toutes les modifications nécessaires avant de mettre au propre la réponse finale.

- 2° Transcrivez au crayon à la mine, de préférence, les réponses ou les solutions dans les espaces en blanc du document à retourner.
- 3° Accompagnez chaque réponse d'une solution claire et détaillée s'il s'agit d'une question qui exige un développement.
- 4° Ne postez que un devoir à la fois; nous vous le retournerons après correction.

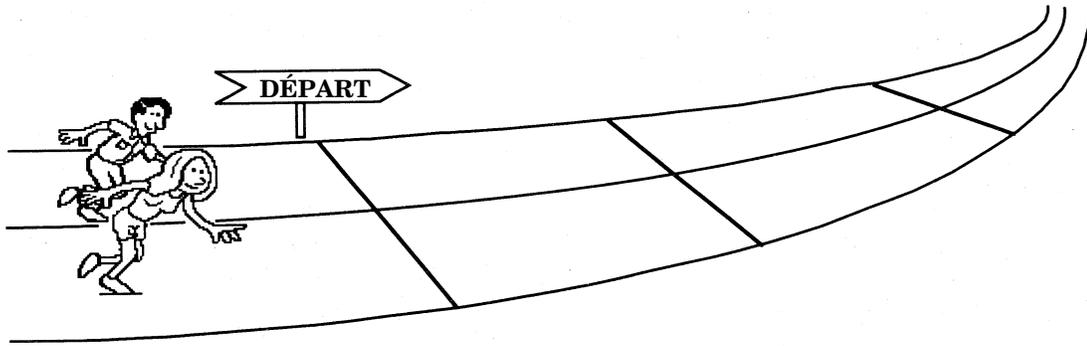
Écrivez, dans la section « Questions de l'élève », les questions que vous désirez poser à la tutrice ou au tuteur. Cette dernière ou ce dernier vous prodiguera des conseils. Elle ou il pourra vous guider dans vos études et vous orienter, s'il y a lieu.

Dans ce cours

- Le devoir 1 porte sur les sous-modules 1 et 2.
Le devoir 2 porte sur les sous-modules 3 à 5.
Le devoir 3 porte sur les sous-modules 1 à 5.

SANCTION

Lorsque vous aurez complété tous les travaux et si vous avez maintenu une moyenne d'au moins 60 %, vous serez autorisé à passer l'examen.



SOUS-MODULE 1

CALCUL DE L'ÉCART MOYEN, DE LA VARIANCE ET DE L'ÉCART TYPE D'UNE DISTRIBUTION DE DONNÉES

1.1 ACTIVITÉ D'ACQUISITION

Quel groupe choisiriez-vous?

Pour amorcer l'étude des statistiques en 5^e secondaire, Claudelle demande à ses élèves de calculer la *moyenne arithmétique*, la *médiane* et le *mode* de deux distributions de données représentant les résultats de ses élèves de l'an dernier dans les deux groupes de mathématiques 536.

Comme ces notions ont été étudiées en 4^e secondaire, les élèves de Claudelle font de grands efforts pour trouver les résultats le plus rapidement possible. Cependant, ils ne se doutent pas que les deux distributions de Claudelle donnent les mêmes résultats, car à première vue, nous ne pouvons prévoir que deux distributions si différentes peuvent produire les mêmes résultats pour les *mesures de tendance centrale*.

À partir de ces résultats, Claudelle fait remarquer à ses élèves qu'il est utile de considérer les différentes mesures de tendance centrale (moyenne arithmétique, médiane et mode), mais que très souvent il est indispensable qu'ils aient aussi de l'information quant à la **dispersion** des données afin d'établir quel groupe peut-être considéré comme le meilleur.

Les mesures de dispersion qui permettent d'évaluer l'éparpillement des données d'une distribution constitueront l'objet d'étude de ce sous-module.

Pour atteindre l'objectif de ce sous-module, vous devrez être capable de calculer l'écart moyen, l'intervalle semi-interquartile, la variance ainsi que l'écart type d'une distribution de données. Vous devrez également être capable d'interpréter la signification de ces mesures de dispersion dans le contexte décrit.



Voici donc les résultats, exprimés en pourcentage, de chacun des élèves des deux groupes de mathématiques 536 de Claudelle.

Groupe A : 52, 59, 60, 62, 64, 65, 68, 72, 80, 82, 82, 84, 86, 87, 88, 90, 91, 95, 96.

Groupe B : 44, 49, 55, 57, 61, 66, 71, 74, 81, 82, 82, 83, 85, 90, 92, 94, 98, 99, 100.

Pour les deux groupes, la moyenne arithmétique ou $\bar{x} = 77\%$, $Md = 82\%$ et $Mo = 82\%$.

Avant de poursuivre votre lecture, assurez-vous d'avoir en mémoire la définition de chacune des mesures de tendance centrale.



- *La moyenne arithmétique est égale à la somme des valeurs des données d'une distribution divisée par le nombre de ces données.*
- *La médiane est la valeur de la donnée qui occupe la position centrale lorsque les données de la distribution sont placées par ordre croissant ou décroissant.*

- *Le mode est la valeur de la donnée la plus fréquente dans une distribution de données.*

Les mathématiques ont développé un langage précis et succinct formé de différents symboles. Grâce à ceux-ci, nous pouvons par exemple traduire un énoncé tel que « la somme des résultats du groupe A est égale à 1 463 » sous la forme suivante.

$$52 + 59 + 60 + 62 + 64 + 65 + 68 + 72 + 80 + 82 + 82 + 84 + 86 + 87 + 88 + 90 + 91 + 95 + 96 = 1\,463$$

Considérons maintenant les variables $x_1, x_2, x_3, \dots, x_{18}$ et x_{19} .

Remarque

Le chiffre situé à droite et en bas de la variable se nomme « indice ». Ainsi, les variables précédentes se lisent comme suit : x un, x deux, x trois, ..., x dix-huit et x dix-neuf.

Attribuons à chacune des variables un des résultats du groupe A.

$$x_1 = 52, x_2 = 59, x_3 = 60, x_4 = 62, x_5 = 64, \dots, x_{17} = 91, x_{18} = 95 \text{ et } x_{19} = 96.$$

Nous pouvons maintenant traduire cet énoncé sous une forme encore plus concise, soit :

$$\sum_{i=1}^{19} x_i = 1\,463.$$

Le symbole de **sommation** Σ qui correspond à la lettre grecque sigma, signifie en langage mathématique « **la somme des...** ».

Ainsi, l'équation $\sum_{i=1}^{19} x_i = 1\,463$ se lit : « la somme des x_i , i variant de 1 à 19, est égale à 1 463. »

En voyant $\sum_{i=1}^{19} x_i$ nous avons :

$$\sum_{i=1}^{19} x_i = x_1 + x_2 + x_3 + x_4 + \dots + x_{18} + x_{19}.$$

Si nous substituons les valeurs numériques correspondant aux différentes variables, nous obtenons ce qui suit.

$$\sum_{i=1}^{19} x_i = 52 + 59 + 60 + 62 + 64 + \dots(11 \text{ données})\dots + 91 + 95 + 96$$

$$\sum_{i=1}^{19} x_i = 1\,463$$

Complétez maintenant l'exemple qui suit afin de vérifier si vous savez déchiffrer un énoncé mathématique comportant le symbole de sommation Σ .

Exemple 1

Soit les données du groupe B : 44, 49, 55, 57, 61, 66, 71, 74, 81, 82, 82, 83, 85, 90, 92, 94, 98, 99 et 100. En considérant que ces nombres correspondent successivement aux valeurs de x_1 à x_{19} , effectuez la sommation ci-dessous.

? $\sum_{i=1}^{19} x_i = \dots\dots\dots$

Solution

Si vous avez trouvé **1 463**, vous avez compris que ce qui peut sembler être une équation mathématique très complexe n'est en fait qu'une banale addition. Si vous avez obtenu un résultat différent, peut-être avez-vous commis une erreur de calcul?

$$\sum_{i=1}^{19} x_i = x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} + x_{11} + x_{12} + x_{13} + x_{14} + x_{15} + x_{16} + x_{17} + x_{18} + x_{19}$$

$$\sum_{i=1}^{19} x_i = 44 + 49 + 55 + 57 + 61 + 66 + 71 + 74 + 81 + 82 + 82 + 83 + 85 + 90 + 92 + 94 + 98 + 99 + 100$$

$$\sum_{i=1}^{19} x_i = 1\,463$$

L'exemple ci-après vous donnera l'occasion de mettre en pratique vos notions d'algèbre. Vous y découvrirez notamment une règle à respecter lorsqu'il s'agit d'évaluer des expressions mathématiques comportant le symbole Σ .

Exemple 2

Croyez-vous que $\sum_{i=1}^n (x_i - 5) = \sum_{i=1}^n x_i - 5$? Avant de répondre, songez à l'importance des parenthèses en algèbre. Étant donné l'absence de parenthèses dans le membre de droite de cette équation, nous en déduisons que $\sum_{i=1}^n x_i - 5 = \left(\sum_{i=1}^n x_i\right) - 5$.

Vérifiez maintenant si les deux expressions sont équivalentes en attribuant à la variable x les valeurs de 6, 7, 8, 9 et 10.

? $\sum_{i=1}^5 (x_i - 5) = \dots\dots\dots$

? $\sum_{i=1}^5 x_i - 5 = \left(\sum_{i=1}^5 x_i\right) - 5 = \dots\dots\dots$

Solution

Après avoir effectué le calcul des expressions présentées plus haut, vous devriez avoir obtenu les résultats suivants.

$$\sum_{i=1}^5 (x_i - 5) = (x_1 - 5) + (x_2 - 5) + (x_3 - 5) + (x_4 - 5) + (x_5 - 5)$$

$$\sum_{i=1}^5 (x_i - 5) = (6 - 5) + (7 - 5) + (8 - 5) + (9 - 5) + (10 - 5)$$

$$\sum_{i=1}^5 (x_i - 5) = 1 + 2 + 3 + 4 + 5$$

$$\sum_{i=1}^5 (x_i - 5) = 15$$

$$\sum_{i=1}^5 x_i - 5 = (x_1 + x_2 + x_3 + x_4 + x_5) - 5$$

$$\sum_{i=1}^5 x_i - 5 = (6 + 7 + 8 + 9 + 10) - 5$$

$$\sum_{i=1}^5 x_i - 5 = 40 - 5$$

$$\sum_{i=1}^5 x_i - 5 = 35$$

Les expressions $\sum_{i=1}^n (x_i - 5)$ et $\sum_{i=1}^n x_i - 5$ ne sont donc pas équivalentes en raison de la présence des parenthèses.

Maintenant que ce type de notation vous est familier, revenons au concept de moyenne arithmétique.

Pour calculer la moyenne arithmétique d'une distribution de données, il faut calculer la somme des valeurs de toutes les données et diviser le résultat obtenu par le nombre de données comprises dans la distribution. En termes mathématiques, nous pouvons écrire ce qui suit.

Pour calculer la moyenne arithmétique d'une distribution, nous devons appliquer la formule suivante.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

où

\bar{x} est le symbole utilisé pour désigner la moyenne arithmétique;

Σ signifie « La somme des... »;

n correspond au nombre de données incluses dans la distribution;

x_i représente chacune des valeurs des données de la distribution;

i est un indice qui peut prendre toutes les valeurs entières de 1 à n .

Exemple 3

Soit les données du groupe A : 52, 59, 60, 62, 64, 65, 68, 72, 80, 82, 82, 84, 86, 87, 88, 90, 91, 95, 96.

? Quelle valeur attribueriez-vous à n dans la formule $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$? Pourquoi?

.....

? Calculez la moyenne en pourcentage des résultats des élèves du groupe A de Claudelle.

.....

Solution

La valeur assignée à n dans la formule $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ doit être 19 puisqu'il y a 19 données dans la distribution. Si nous effectuons le calcul de la moyenne nous obtenons ce qui suit.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{\sum_{i=1}^{19} x_i}{19}$$

$$\bar{x} = \frac{1463}{19} = 77 \%$$

Cette partie théorique vous a permis d'entrevoir une facette de la rigueur du langage mathématique. Sachez que nous nous sommes limités strictement aux types d'expressions qui vous seront utiles pour l'étude de l'écart moyen, de la variance et de l'écart type.

Avant de passer à l'étude de ces mesures de dispersion, il y a deux autres exemples que nous désirons examiner avec vous. Ceux-ci serviront à mieux vous préparer pour entreprendre l'étude de la variance et de l'écart type.

Exemple 4

Il s'agit de vérifier si $\sum_{i=1}^n x_i^2 \neq \left(\sum_{i=1}^n x_i\right)^2$. Pour ce faire, nous utiliserons les 4 données suivantes : 1, 3, 5 et 6.

$$\sum_{i=1}^4 x_i^2 \neq \left(\sum_{i=1}^4 x_i\right)^2$$

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 \neq (x_1 + x_2 + x_3 + x_4)^2$$

$$1^2 + 3^2 + 5^2 + 6^2 \neq (1 + 3 + 5 + 6)^2$$

$$1 + 9 + 25 + 36 \neq (15)^2$$

$$71 \neq 225$$

Exemple 5

Nous savons que $\sqrt{x^2} = x$. Mais pouvons-nous affirmer que $\sqrt{\sum_{i=1}^n x_i^2} = \sum_{i=1}^n x_i$?

Afin de vérifier si ces deux expressions sont équivalentes, nous procéderons comme nous l'avons fait à l'exemple 4, c'est-à-dire que nous évaluerons chacun des membres de l'égalité pour les mêmes valeurs de x . Nous choisissons ici d'attribuer à x les valeurs suivantes : $-5, -4, -3, 3, 4$ et 5 .

$$\begin{aligned} \sqrt{\sum_{i=1}^n x_i^2} &\stackrel{?}{=} \sum_{i=1}^n x_i \\ \sqrt{x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2} &\stackrel{?}{=} x_1 + x_2 + x_3 + x_4 + x_5 + x_6 \\ \sqrt{(-5)^2 + (-4)^2 + (-3)^2 + (3)^2 + (4)^2 + (5)^2} &\stackrel{?}{=} -5 - 4 - 3 + 3 + 4 + 5 \\ \sqrt{25 + 16 + 9 + 9 + 16 + 25} &\stackrel{?}{=} 0 \\ \sqrt{100} &\stackrel{?}{=} 0 \\ 10 &\neq 0 \end{aligned}$$

Nous constatons que le calcul de ces expressions, pour les mêmes valeurs de la variable, mène à des résultats différents. Par conséquent, nous en concluons que $\sqrt{\sum_{i=1}^n x_i^2} \neq \sum_{i=1}^n x_i$.

Sachez encore une fois que nous nous sommes limités strictement aux types d'expressions qui vous seront utiles pour l'étude des différentes mesures de dispersion qui vont suivre. Nous récapitulons les règles et les mises en garde sur lesquelles nous avons voulu attirer votre attention. N'hésitez pas à revenir les consulter chaque fois que vous serez dans le doute.

- La valeur de n dans une expression du type $\sum_{i=1}^n x_i$ correspond au nombre de données comprises dans la distribution.
- $\sum_{i=1}^n (x_i \pm k) \neq \sum_{i=1}^n x_i \pm k$ où k est un nombre quelconque.
- $\sum_{i=1}^n x_i^2 \neq \left(\sum_{i=1}^n x_i\right)^2$
- $\sqrt{\sum_{i=1}^n x_i^2} \neq \sum_{i=1}^n x_i$

Avant de continuer à exercer votre habileté à calculer le résultat d'une expression mathématique comportant le symbole de sommation Σ , offrez-vous un moment de répit.



Saviez-vous que...

... grâce au truc mathématique qui vous est proposé ici, vous pourrez impressionner beaucoup de gens?

A	B	C	D
2	8	1	4
10	11	15	6
3	15	9	13
11	12	3	7
6	9	11	5
15	13	13	14
7	10	5	15
14	14	7	12

- Demandez à la personne désignée de choisir un nombre de 1 à 15. Présentez-lui ensuite la grille ci-dessous et dites-lui de vous indiquer la lettre qui correspond à chacune des colonnes dans laquelle le nombre choisi se retrouve.
- Additionnez les premiers chiffres des colonnes indiquées et vous aurez déterminé le « nombre mystère ».

Exercice 1.1

1. Calculez chacune des expressions qui suivent pour les valeurs assignées à la variable x .

a) $\sum_{i=1}^{12} \frac{x_i}{2}$ si les valeurs de x sont $-8, -6, -4, -2, 0, 2, 4, 6, 8, 10, 12$ et 14 .

.....

.....

.....

b) $\sum_{i=1}^8 (x_i - 3)$ si les valeurs de x sont $5, 7, 9, 11, 12, 13, 15$ et 20 .

.....

.....

.....

c) $\sum_{i=1}^6 x_i^2 + 7$ si les valeurs de x sont $-4, -3, -2, 2, 3$ et 4 .

.....

.....

.....

.....

d) $\sqrt{\sum_{i=1}^n (x_i - 1)^2}$ si les valeurs de x sont $-5, -4, -3, -2, 2, 3$ et 4 .

.....

.....

.....

.....

.....

2. Est-ce que $\sum_{i=1}^n (x_i^2 + x_i) = \sum_{i=1}^n x_i^2 + \sum_{i=1}^n x_i$? Pour effectuer votre vérification, donnez à x les valeurs $2, 4, 6$ et 8 .

.....

.....

.....

.....

.....

.....

.....

.....

3. François vient de recevoir son relevé de notes des examens du ministère de l'Éducation et il en est très fier. Voici ses résultats.

- Sciences physiques 436 : 97 %.
- Sciences physiques 416 : 96 %.
- Mathématiques 436 : 96 %.
- Histoire du Québec et du Canada : 93 %.
- Informatique : 87 %.
- Français : 81 %.
- Anglais, langue seconde : 94 %.

Calculez la moyenne de François en utilisant la formule $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ où chacun des x_i correspond au résultat dans l'une des matières.

.....

En statistiques, nous mesurons l'éparpillement d'un ensemble de données grâce à la **variance**. Cette mesure de dispersion nous permet de déterminer à quel point les données d'une distribution s'éloignent de la moyenne arithmétique.

Plus la variance est **élevée**, plus les données d'une distribution sont **dispersées**. Plus la variance est **faible**, plus les données d'une distribution sont **regroupées** autour de la moyenne arithmétique.

Vous êtes sans doute impatient de découvrir comment procéder pour déterminer la valeur de la variance associée à une distribution de données. La chose est relativement simple. Il s'agit d'appliquer la formule mathématique qui suit.

Pour calculer la variance d'une distribution, nous devons appliquer la formule suivante.

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n - 1}$$

où

s^2 est l'expression utilisée pour désigner la variance;

\bar{x} est le symbole utilisé pour désigner la moyenne arithmétique;

Σ signifie « La somme des... »;

n correspond au nombre de données incluses dans la distribution;

x_i représente chacune des valeurs des données de la distribution;

i est un indice qui peut prendre toutes les valeurs entières de 1 à n .

Revenons à la liste des résultats des élèves de Claudelle pour les deux groupes de mathématique 536. À l'aide de la formule qui vous a été présentée, nous effectuerons le calcul de la variance pour chacun des groupes.

Groupe A : 52, 59, 60, 62, 64, 65, 68, 72, 80, 82, 82, 84, 86, 87, 88, 90, 91, 95, 96

Groupe B : 44, 49, 55, 57, 61, 66, 71, 74, 81, 82, 82, 83, 85, 90, 92, 94, 98, 99, 100.

Puisque la formule pour calculer la variance est assez complexe, nous vous suggérons de procéder par étapes. La procédure à suivre est résumée dans l'encadré qui suit.

Pour calculer la variance d'une distribution de données, nous devons :

- 1° calculer la moyenne arithmétique \bar{x} de la distribution;
- 2° calculer chacun des $(x_i - \bar{x})$ et en élever le résultat au carré;
- 3° calculer la somme de tous les $(\bar{x}_i - x)^2$;
- 4° diviser le résultat obtenu par $n - 1$.

Exemple 6

Calculons la variance des résultats du groupe A en suivant l'algorithme proposé plus haut.

Groupe A : 52, 59, 60, 62, 64, 65, 68, 72, 80, 82, 82, 84, 86, 87, 88, 90, 91, 95, 96.

1° La moyenne du groupe A a été calculée à l'exemple 3.

$$\bar{x} = 77 \%$$

2° Calculons ensuite l'écart par rapport à la moyenne de chacune des données, c'est-à-dire la valeur de $(x_i - 77)$ pour chacun des résultats de la liste, et élevons-en le résultat au carré. Le tableau ci-dessous vous aidera à vous y retrouver.

x_i	$(x_i - 77)$	$(x_i - 77)^2$
52	$(52 - 77) = -25$	625
59	$(59 - 77) = -18$	324
60	$(60 - 77) = -17$	289
62	$(62 - 77) = -15$	225
64	$(64 - 77) = -13$	169
65	$(65 - 77) = -12$	144
68	$(68 - 77) = -9$	81
72	$(72 - 77) = -5$	25
80	$(80 - 77) = 3$	9
82	$(82 - 77) = 5$	25
82	$(82 - 77) = 5$	25
84	$(84 - 77) = 7$	49
86	$(86 - 77) = 9$	81
87	$(87 - 77) = 10$	100
88	$(88 - 77) = 11$	121
90	$(90 - 77) = 13$	169
91	$(91 - 77) = 14$	196
95	$(95 - 77) = 18$	324
96	$(96 - 77) = 19$	361

3° Calculons maintenant la somme de tous les $(x_i - \bar{x})^2$, c'est-à-dire de toutes les valeurs obtenues dans la dernière colonne du tableau précédent.

$$\sum_{i=1}^{19} (x_i - 77)^2 = 625 + 324 + 289 + 225 + 169 + 144 + 81 + 25 + 9 + 25 + 25 + 49 + 81 + 100 + 121 + 169 + 196 + 324 + 361 = 3\,342$$

4° Divisons enfin le tout par $n - 1$ pour obtenir la valeur de la variance.

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n - 1} = \sum_{i=1}^{19} \frac{(x_i - 77)^2}{19 - 1} = \frac{3\,342}{18} = 185,67$$

À vous maintenant d'effectuer le calcul de la variance du groupe B en complétant l'exemple 7.

Exemple 7

Nous avons reproduit pour vous les résultats du groupe B de Claudelle. Calculez la variance de cette distribution de données en suivant les étapes décrites ci-après.

Groupe B : 44, 49, 55, 57, 61, 66, 71, 74, 81, 82, 82, 83, 85, 90, 92, 94, 98, 99, 100.

? 1° Calculez la moyenne arithmétique \bar{x} de cet ensemble de résultats.

.....

.....

.....

.....

.....

? 2° Calculez chacun des $(x_i - \bar{x})$ et élevez-en le résultat au carré.

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
44		
49		
55		
57		
61		
66		
71		
74		
81		
82		
82		
83		
85		
90		
92		
94		
98		
99		
100		

? 3° Calculez la somme de tous les $(x_i - \bar{x})^2$.

.....

? 4° Divisez le résultat par $n - 1$ afin d'obtenir la valeur de la variance.

.....

Solution

Vous devriez avoir trouvé une variance de 304,56. Si tel n'est pas le cas, révisez vos calculs à l'aide du corrigé qui suit.

$$1^\circ \bar{x} = (44 + 49 + 55 + 57 + 61 + 66 + 71 + 74 + 81 + 82 + 82 + 83 + 85 + 90 + 92 + 94 + 98 + 99 + 100) \div 19$$

$$\bar{x} = \frac{1463}{19}$$

$$\bar{x} = 77\%$$

2°

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
44	$(44 - 77) = -33$	1 089
49	$(49 - 77) = -28$	784
55	$(55 - 77) = -22$	484
57	$(57 - 77) = -20$	400
61	$(61 - 77) = -16$	256
66	$(66 - 77) = -11$	121
71	$(71 - 77) = -6$	36
74	$(74 - 77) = -3$	9
81	$(81 - 77) = 4$	16
82	$(82 - 77) = 5$	25
82	$(82 - 77) = 5$	25
83	$(83 - 77) = 6$	36
85	$(85 - 77) = 8$	64
90	$(90 - 77) = 13$	169
92	$(92 - 77) = 15$	225
94	$(94 - 77) = 17$	289
98	$(98 - 77) = 21$	441
99	$(99 - 77) = 22$	484
100	$(100 - 77) = 23$	529

$$3^\circ \sum_{i=1}^{19} (x_i - 77)^2 = 1\,089 + 784 + 484 + 400 + 256 + 121 + 36 + 9 + 16 + 25 + 25 + 36 + 64 + 169 + 225 + 289 + 441 + 484 + 529 = 5\,482$$

$$4^\circ s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1} = \sum_{i=1}^{19} \frac{(x_i - 77)^2}{19-1} = \frac{5\,482}{18} = 304,56$$

En comparant les résultats du groupe B à ceux du groupe A, nous constatons que ceux-ci sont plus éparpillés. Par conséquent, le carré de leur écart par rapport à la moyenne $(x_i - \bar{x})^2$ est plus grand. La variance, qui est une mesure de la dispersion des données, a donc une valeur numérique plus élevée.

Au lieu d'éparpillement des données, nous aurions pu parler de l'*étendue* de la distribution.



L'étendue (E) d'une distribution de données est la différence entre la plus grande et la plus petite valeur numérique de cette distribution.

La variance a le désavantage de s'exprimer en unités carrées. Pour remédier à cette situation, nous en extrayons la racine carrée et nous obtenons alors la mesure de dispersion très utilisée en statistique qui est l'**écart type**. Son symbole est s et sa valeur est égale à la racine carrée de la variance, notée s^2 .

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

L'écart type, contrairement à la variance, s'exprime dans les **mêmes unités de mesure que les données** : si les données représentent des longueurs mesurées en cm, l'écart type est également exprimé en **cm**, tandis que la variance est donnée en **cm²**. Le fait qu'une mesure de dispersion soit exprimée dans les mêmes unités que les données de la distribution constitue un réel avantage lors de la vulgarisation des résultats d'un sondage ou d'une enquête.

Exemple 8

Le propriétaire de deux stations-service a décidé de compiler le nombre de litres d'essence qu'il faut pour faire le plein de chacune des automobiles qui s'arrêtent à l'une et l'autre de ses stations. L'une se situe en bordure de l'autoroute, alors que l'autre se trouve au cœur de la ville.

Voici un relevé du nombre de litres d'essence vendus à 15 automobilistes qui ont fait le plein à chacune des stations.

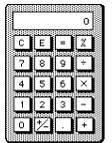
Station-service (autoroute)

30, 22, 21, 28, 25, 26, 26, 24, 29, 23, 20, 27, 25, 24, 25.

Station-service (ville)

25, 23, 30, 19, 35, 27, 15, 25, 17, 31, 14, 20, 33, 25, 36.

- ? 1. À l'aide des fonctions statistiques de votre calculatrice, calculez le nombre moyen de litres d'essence vendus aux clients qui ont fait le plein à la station-service située en bordure de l'autoroute.



$\bar{x} = \dots\dots\dots$

Si vous avez obtenu une moyenne de 25 litres d'essence par plein, vous maîtrisez le fonctionnement de votre calculatrice. Si toutefois vous ne parvenez pas à ce résultat, nous vous suggérons fortement de réviser la façon de procéder pour effectuer le calcul à l'aide des fonctions statistiques de votre calculatrice.

Remarque

Généralement, les calculatrices scientifiques possèdent aussi la fonction « STAT ». Dans ce cas-ci, il suffit de mettre la calculatrice en mode « STAT 0 », de vider la mémoire et d'entrer chacune des données à l'aide de la touche M+ ou DATA .

Ainsi, dans notre cas, nous devons faire comme suit.

Vidons d'abord la mémoire.

$\boxed{2\text{ndF}}$	$\boxed{\text{CA}}$	(ou DEL)
$\boxed{30}$	$\boxed{\text{M+}}$	(ou DATA)
\boxed{n}	$\boxed{1}$	← devrait apparaître à l'écran
$\boxed{22}$	$\boxed{\text{M+}}$	(ou DATA)
\cdot	\cdot	
\cdot	\cdot	
\cdot	\cdot	
$\boxed{25}$	$\boxed{\text{M+}}$	(ou DATA)
n	15	

Pour connaître la moyenne ou \bar{x} , nous devons peser sur les touches $\boxed{\text{RCL}}$ $\boxed{\bar{x}}$ ou $\boxed{2\text{ndF}}$ $\boxed{\bar{x}}$. La calculatrice devrait afficher 25,000... Il est à noter qu'il existe des calculatrices où nous obtenons la moyenne en pesant directement sur la touche $\boxed{\bar{x}}$.

2. En utilisant les données recueillies à la station-service située en bordure de l'autoroute, effectuez les calculs demandés à chacune des étapes ci-après. Ceux-ci ont pour but de mettre en évidence la relation qui existe entre l'écart type et la variance d'un ensemble de données.

? 1° Complétez d'abord le tableau suivant en substituant à \bar{x} la valeur de la moyenne obtenue en 1.

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
30	$(30 l - 25 l) = 5 l$	$25 l^2$
22	$(22 l - 25 l) = -3 l$	$9 l^2$
21	$(21 l - 25 l) = -4 l$	$16 l^2$
28	$(28 l - 25 l) = 3 l$	$9 l^2$
25	$(25 l - 25 l) = 0 l$	$0 l^2$
26		
26		
24		
29		
23		
20		
27		
25		
24		
25		

? 2° Calculez ensuite la somme de tous les $(x_i - \bar{x})^2$.

$$\sum_{i=1}^{15} (x_i - \bar{x})^2 = \dots\dots\dots$$

? 3° Divisez le résultat obtenu par $n - 1$ afin d'obtenir la valeur de la variance.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} = \dots\dots\dots$$

Si tous vos calculs sont exacts, vous devriez avoir trouvé une variance de $8 l^2$.

? 4° Pour trouver l'écart type de la distribution, calculez simplement la racine carrée de ce résultat.

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} = \dots\dots\dots$$

Ainsi, la valeur de l'écart type est $\sqrt{8 l^2} = 2,83 l$.

Tous ces calculs sont très longs à effectuer! Au numéro 3, vous aurez cependant à calculer l'écart type de cette même distribution à l'aide de votre calculatrice. Vous pourrez constater à quel point celle-ci vous permet d'épargner du temps.

- ? 3. À l'aide de votre calculatrice, calculez la valeur de l'écart type des données recueillies à la station-service se trouvant en bordure de l'autoroute. Pour ce faire, vous devez entrer les données exactement de la même façon que pour calculer la moyenne. La touche sur laquelle vous devrez appuyer pour obtenir la valeur de l'écart type est habituellement identifiée par la lettre *s* ou *sx*.



N.B. – Généralement, la touche *s* en est une de seconde fonction. Il faudrait donc vérifier si tel est le cas de votre calculatrice.

Si vous en ressentez le besoin, n'hésitez pas à demander l'aide d'une personne-ressource, car il est important que vous maîtrisiez cette fonction statistique.

s pour la station-service (autoroute) :

Avez-vous trouvé la même valeur qu'en 2? Si tel n'est pas le cas, révisez les instructions qui vous ont été fournies et reprenez vos manipulations.

Après avoir effectué le calcul de l'écart type selon les deux méthodes proposées en 2 et en 3, laquelle préférez-vous?

- ? 4. À l'aide de votre calculatrice, évaluez maintenant le nombre moyen de litres d'essence vendus à chacun des clients qui ont fait le plein à la station-service au cœur de la ville. Calculez également l'écart type associé à l'ensemble de ces données et arrondissez le résultat au dixième près.



\bar{x} =

s =

La calculatrice nous permet de trouver $\bar{x} = 25 \text{ l}$ et $s = 7,1 \text{ l}$. Rappelez-vous que l'écart type indique à quel point l'ensemble des données s'écarte de la valeur moyenne.

Dans cet exemple, nous constatons que l'écart type se rapportant aux données recueillies à la station-service située au cœur de la ville est supérieur à celui de la distribution de données compilées à la station-service se trouvant en bordure de l'autoroute. Cela signifie donc que les données recueillies à la station-service située au cœur de la ville sont plus dispersées. Si vous observez attentivement les données qui vous ont été présentées, vous constaterez qu'il en est ainsi.

Pour vous en convaincre, comparez les **diagrammes de quartiles** de ces deux distributions représentées dans la figure 1.1.

Nombre de litres d'essence vendus

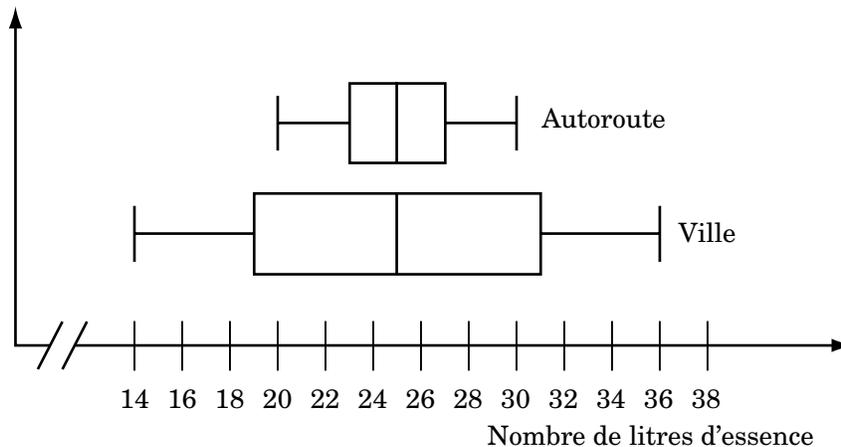


Fig. 1.1 Diagrammes de quartiles illustrant le nombre de litres d'essence vendus à 15 automobilistes qui ont fait le plein à chacune des stations



- *Les éléments importants dans la construction d'un diagramme de quartiles sont le minimum, le maximum et les trois **quartiles** Q_1 , Q_2 et Q_3 . Une fois le diagramme construit, nous sommes en mesure de bien voir l'étendue de la distribution, l'étendue de chacun des quarts et l'**étendue interquartile**.*



- L'étendue interquartile, notée EI , est la différence entre Q_3 et Q_1 . Elle est représentée par la boîte dans le diagramme de quartiles.

L'intervalle semi-interquartile, noté Q , est obtenu en divisant l'étendue interquartile par deux : $Q = \frac{EI}{2}$.

L'intervalle semi-interquartile représente la longueur moyenne des intervalles contenant chacun à peu près 25 % des données de part et d'autre de la médiane où $Md = Q_2$.

Nous considérons qu'environ 50 % des données les plus centrales sont comprises dans l'intervalle $Md - Q$ et $Md + Q$.

Exemple 9

Soit le diagramme de quartiles de la figure 1.1 représentant le nombre de litres d'essence vendus à 15 automobilistes qui ont fait le plein à la station située au cœur de la ville.

? 1° Déterminez les valeurs de Q_1 , Q_2 et Q_3 .

$Q_1 = \dots\dots\dots$, $Q_2 = Md = \dots\dots\dots$ et $Q_3 = \dots\dots\dots$.

? 2° Déterminez la valeur de l'étendue interquartile.

$EI = \dots\dots\dots$

? 3° Déterminez la valeur de l'intervalle semi-interquartile.

$Q = \dots\dots\dots$

? 4° Calculez la longueur moyenne des intervalles contenant chacun à peu près 25 % des données de part et d'autre de la médiane.

$$Md - Q = \dots\dots\dots \text{ et } Md + Q = \dots\dots\dots$$

? 5° Existe-t-il un lien entre les valeurs calculées en 4° et les valeurs de Q_1 et Q_3 ?

.....

Solution

1° Les valeurs de Q_1 , Q_2 et Q_3 sont prises à partir du diagramme de quartiles.

$$Q_1 = 19, Q_2 = Md = 25 \text{ et } Q_3 = 31.$$

$$2^\circ EI = Q_3 - Q_1 = 31 - 19 = 12$$

$$3^\circ Q = \frac{EI}{2} = \frac{12}{2} = 6$$

$$4^\circ Md - Q = 25 - 6 = 19 \text{ et } Md + Q = 25 + 6 = 31.$$

5° Oui, il existe un lien entre les valeurs calculées en 4° et les valeurs de Q_1 et Q_3 . En effet, comme l'intervalle semi-interquartile représente la longueur moyenne des intervalles contenant chacun à peu près 25 % des données de part et d'autre de la médiane, cela correspond à l'étendue interquartile, mais seulement lorsque la médiane est située exactement au centre de l'étendue interquartile.

Dans cet exemple, nous constatons qu'il est possible d'examiner l'étalement par rapport au centre de la distribution, sans tenir compte des valeurs extrêmes.

Avant de mettre vos nouvelles connaissances à l'épreuve dans les exercices qui suivent, offrez-vous un petit moment de détente.



Saviez-vous que...

...un mot de trois lettres complète tous les mots suivants?

___ ERIE

MAIL___

ROU___TE

PI___E

LOT
Solution

Exercice 1.2

1. M. Leblanc a fait installer du tapis dans chacun des 10 logements de ses deux immeubles situés sur la rue St-Pierre. Le tableau suivant indique la surface qui a été couverte dans chacun des logements.

Immeuble n° 1 (m ²)	Immeuble n° 2 (m ²)
48	36
38	42
12	37
22	28
35	45
24	36
42	24
36	46
52	39
15	48

- a) Calculez l'écart type associé aux données recueillies dans chacun des immeubles.

Immeuble n° 1 : Immeuble n° 2 :

- b) En vous référant à la valeur de l'écart type obtenue pour chacun des immeubles, déterminez dans lequel des bâtiments la surface de tapis diffère le plus d'un logement à l'autre. Motivez votre réponse.

.....

.....

.....

- c) Le calcul de la variance vous aurait-il permis de tirer les mêmes conclusions? Expliquez brièvement pourquoi.

.....

.....

.....

2. Nous avons compilé le prix de divers livres dans une vingtaine de librairies. Nous avons remarqué notamment que le prix du dernier roman de Harry Potter avait un écart type de 3 \$, alors que celui du Guide de l'auto 2004 avait un écart type de 12 \$. À la lumière de ces renseignements, quelles conclusions pouvez-vous tirer au sujet de l'uniformité des prix de ces deux livres dans les librairies ayant fait l'objet de cette étude?

.....

.....

.....

.....

3. Deux écoles primaires, situées dans des quartiers différents, ont mené un sondage auprès des parents afin de déterminer la quantité de lait que boit chacun de leurs enfants quotidiennement. Après avoir compilé les résultats, nous en sommes venu aux conclusions suivantes.

- Les enfants de l'école Terre-Soleil prennent en moyenne 450 ml de lait par jour. De plus, tous les enfants boivent sensiblement la même quantité de lait.

- Les enfants de l'école Arthur-Vaillancourt prennent en moyenne 600 ml de lait par jour. Cependant, nous avons pu constater que les quantités rapportées par les parents variaient beaucoup d'un enfant à l'autre.

Après avoir pris connaissance de ce rapport, déterminez lequel de ces ensembles de résultats donne lieu à la variance la plus élevée et dites pourquoi.

.....

.....

.....

4. Le lendemain de la victoire de la Coupe Stanley par le Lightning de Tampa Bay, les journalistes comparaient les 10 meilleurs marqueurs de chacune des Associations. Dans le tableau ci-dessous, nous avons indiqué le nombre de points accumulés par les joueurs, selon l'Association.

Meilleurs marqueurs pendant les séries de la Coupe Stanley, 2003-2004

ASSOCIATION DE L'EST		ASSOCIATION DE L'OUEST	
Joueur	Points	Joueur	Points
Brad Richards	26	Jarome Iginla	22
Martin St-Louis	24	Craig Conroy	17
Fredrik Modin	19	Martin Gélinas	15
Keith Primeau	16	Vincent Damphousse	14
Vincent Lecavalier	16	Patrick Marleau	12
Ruslan Fedotenko	14	Joe Sakic	12
Alexei Zhamnov	14	Marcus Nilson	11
Dave Andreychuk	14	Peter Forsberg	11
Jeremy Roenick	13	Shean Donovan	10
Saku Koivu	11	Jonathan Cheechoo	10

Calculez la moyenne et l'écart type de chacune des distributions de données, puis comparez les deux distributions à l'aide de ces valeurs afin de déterminer à quelle Association appartient l'équipe gagnante.

.....

.....

.....

.....

.....

5. Voici les résultats obtenus par les élèves des deux groupes de Véronica lors d'un test de mathématiques corrigé sur 25.

- Groupe 51

20	17	12	15	18
22	24	19	21	15
21	19	20	23	22
21	17	16	14	15

- Groupe 52

15	19	18	23	25
24	25	14	19	22
24	21	8	19	17
17	19	21	24	22

Calculez l'écart type des résultats de chacun de ces groupes et dites ce que vous pouvez en conclure si vous comparez ces deux résultats.

Groupe 51 : Groupe 52 :

Conclusion

.....
.....
.....

Il existe une autre mesure de dispersion utilisée en statistiques, il s'agit de l'**écart moyen**. Cette mesure de dispersion correspond à la moyenne des **valeurs absolues** des écarts à la moyenne. Pour calculer l'écart moyen, il suffit d'appliquer la formule mathématique qui suit.

Pour calculer l'écart moyen d'une distribution, nous devons appliquer la formule suivante.

$$EM = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

où

EM est l'expression utilisée pour désigner l'écart moyen ;

\bar{x} est le symbole utilisé pour désigner la moyenne arithmétique;

Σ signifie « La somme des... »;

n correspond au nombre de données incluses dans la distribution;

x_i représente chacune des valeurs des données de la distribution;

i est un indice qui peut prendre toutes les valeurs entières de 1 à n .

Puisque la formule pour calculer l'écart moyen d'une distribution est assez complexe, nous vous suggérons de procéder par étapes. La procédure à suivre est résumée dans l'encadré qui suit.

Pour calculer l'écart moyen d'une distribution de données, nous devons :

- 1° calculer la moyenne arithmétique \bar{x} de la distribution;
- 2° calculer chacun des $(x_i - \bar{x})$ et en prendre la valeur absolue;
- 3° calculer la somme de tous les $|x_i - \bar{x}|$;
- 4° diviser le résultat obtenu par n .

La valeur absolue de la différence entre deux valeurs données permet de transformer des résultats négatifs en résultats positifs. Le symbole est « | | » qui signifie « **la valeur absolue de** ». Ainsi, si la valeur de x_i est plus petite que la valeur de \bar{x} , le résultat est négatif, tandis que la valeur absolue de ce résultat est positive.

Exemple 10

Calculons l'écart moyen des résultats de Julie en suivant l'algorithme proposé plus haut.

Julie	70	72	70	72	67	70	69
--------------	----	----	----	----	----	----	----

1° Calculons d'abord la moyenne arithmétique de cet ensemble de données.

$$\bar{x} = \frac{\sum_{i=1}^7 x_i}{7} = \frac{70 + 72 + 70 + 72 + 67 + 70 + 69}{7} = \frac{490}{7} = 70$$

2° Calculons ensuite l'écart par rapport à la moyenne de chacune des données, c'est-à-dire de $(x_i - 70)$ pour chacun des résultats de la liste et prenons la valeur absolue de ce résultat. Le tableau ci-dessous vous aidera à vous y retrouver.

x_i	$(x_i - 70)$	$ x_i - 70 $
70	$(70 - 70) = 0$	0
72	$(72 - 70) = 2$	2
70	$(70 - 70) = 0$	0
72	$(72 - 70) = 2$	2
67	$(67 - 70) = -3$	3
70	$(70 - 70) = 0$	0
69	$(69 - 70) = -1$	1

3° Calculons maintenant la somme de tous les $|x_i - 70|$, c'est-à-dire de toutes les valeurs obtenues dans la dernière colonne du tableau précédent.

$$\sum_{i=1}^7 |x_i - 70| = 0 + 2 + 0 + 2 + 3 + 0 + 1 = 8$$

4° Divisons enfin le tout par n pour obtenir la valeur de l'écart moyen.

$$EM = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} = \frac{\sum_{i=1}^7 |x_i - 70|}{7} = \frac{8}{7} = 1,14$$

À vous maintenant de faire le calcul de l'écart moyen des résultats de Dominic, le copain de Julie, en complétant l'exemple 11.

Exemple 11

Nous avons reproduit pour vous le relevé de notes de Dominic. Calculez l'écart moyen de cette distribution de données en suivant les étapes décrites ci-après.

Dominic	69	75	62	83	59	66	76
----------------	----	----	----	----	----	----	----

? 1° Calculez la moyenne arithmétique \bar{x} de cet ensemble de résultats.

.....

? 2° Calculez chacun des $(x_i - \bar{x})$ et prenez la valeur absolue des résultats.

x_i	$(x_i - \bar{x})$	$ x_i - \bar{x} $
69		
75		
62		
83		
59		
66		
76		

? 3° Calculez la somme de tous les $|x_i - \bar{x}|$.

.....

? 4° Divisez le résultat obtenu par n afin d'obtenir la valeur de l'écart moyen.

.....

Solution

Vous devriez aussi avoir trouvé un écart moyen de 6,86. Si tel n'est pas le cas, réviser vos calculs à l'aide du corrigé qui suit.

$$1^\circ \bar{x} = \frac{69 + 75 + 62 + 83 + 59 + 66 + 76}{7} = \frac{490}{7} = 70$$

2°

x_i	$(x_i - 70)$	$ x_i - 70 $
69	$(69 - 70) = -1$	1
75	$(75 - 70) = 5$	5
62	$(62 - 70) = -8$	8
83	$(83 - 70) = 13$	13
59	$(59 - 70) = -11$	11
66	$(66 - 70) = -4$	4
76	$(76 - 70) = 6$	6

3° $\sum_{i=1}^7 |x_i - \bar{x}| = 1 + 5 + 8 + 13 + 11 + 4 + 6 = 48$

4° $EM = \frac{\sum_{i=1}^7 |x_i - \bar{x}|}{n} = \frac{48}{7} = 6,86$

En comparant les résultats de Dominic à ceux de Julie, nous constatons que ceux-ci sont plus éparpillés. Par conséquent, l'écart moyen, qui est une mesure de la dispersion des données, a donc une valeur numérique plus élevée. Il en aurait été de même si nous avions calculé la variance ou l'écart type.

Effectuez les exercices suivants. Assurez-vous de bien comprendre chacun de ceux-ci. N'hésitez pas à consulter une personne-ressource si vous en ressentez le besoin.

Exercice 1.3

1. Nous avons calculé l'écart moyen des pointages de deux groupes de golfeurs : l'écart moyen des pointages du groupe A est 10, alors que celui du groupe B est 36. Sans même connaître les pointages associés aux différents joueurs de chacun des groupes, que pouvez-vous dire au sujet du calibre des golfeurs des groupes A et B?

.....

.....

2. La taille, en centimètres, des joueurs de deux équipes de volley-ball est indiquée dans le tableau suivant.

Équipe des Jaunes	Équipe des Bleus
180	178
175	184
184	178
176	178
172	179
181	172
179	178
177	177

a) Comparez les deux distributions de données et, sans effectuer aucun calcul, déterminez laquelle des deux comporte les données les plus dispersées.

.....

b) Peut-être avez-vous eu du mal à répondre à la question précédente? Effectuez le calcul de l'écart moyen de ces deux distributions de données et déterminez avec certitude laquelle des deux présente les valeurs les plus étalées.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

3. Un sondage a été mené auprès des employés de deux entreprises afin de déterminer à combien de kilomètres de leur domicile se situe leur lieu de travail. Voici les données recueillies pour chacune des entreprises.

Entreprise A : 40, 38, 40, 60, 41, 40, 20, 42, 39.
Entreprise B : 14, 31, 25, 41, 19, 42, 8, 36, 9.

a) Calculez l'écart moyen de chacune de ces distributions et déterminez laquelle comporte les données les plus étalées?

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

b) Auriez-vous pu vous appuyer sur l'étendue de ces distributions pour déterminer laquelle est la plus dispersée? Pourquoi?

.....

.....

.....

.....

.....



1.2 EXERCICES DE CONSOLIDATION

1. Voici le relevé du nombre de minutes pendant lesquelles Normand et Sylvie ont navigué sur Internet au cours des neuf derniers jours.

Normand			Sylvie		
80	62	48	75	50	112
90	73	132	120	110	40
132	56	77	85	72	45

- a) Calculez, au centième près, l'écart type et la variance de ces deux distributions.

	Normand	Sylvie
Écart type		
Variance		

- b) Que pouvez-vous déduire de ces résultats?

.....

.....

.....

2. Voici les résultats obtenus par les élèves de M^{me} Duquette et de M. Mondou au dernier examen de mathématiques de l'école de la Réussite.

- Groupe de M^{me} Duquette

75	82	73	72	80	57	78
65	68	86	77	70	83	90

- Groupe de M. Mondou

78	85	70	98	81	75	77
73	79	59	72	79	81	71

Calculez l'écart type des résultats de chacun de ces groupes et dites ce que vous pouvez en conclure si vous comparez ces deux résultats.

- Groupe de M^{me} Duquette :

- Groupe de M. Mondou :

- **Conclusion**

3. Voici les résultats obtenus par les élèves de Martin et de Rosanne lors d'un tournoi de calcul mental.

Élèves de Martin			Élèves de Rosanne		
6	12	5	15	4	13
8	9	20	12	7	20
7	8	13	18	15	7
12	9	21	10	15	9

a) Calculez, au centième près, l'écart moyen de ces deux distributions.

- **Élèves de Martin**

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

- **Élèves de Rosanne**

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

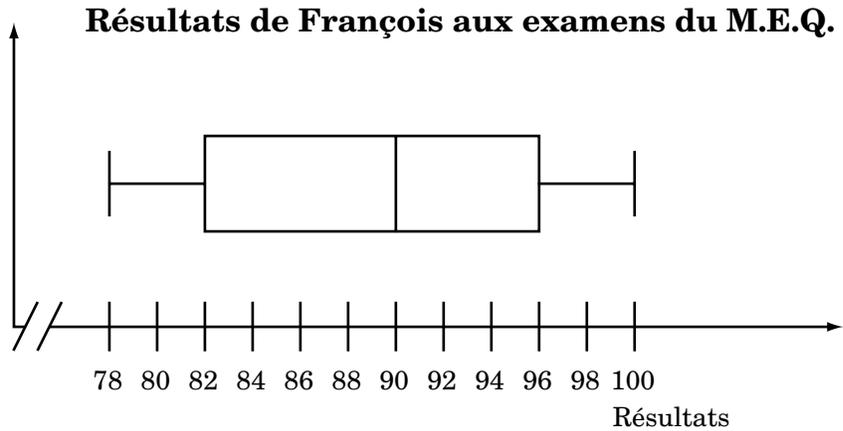
b) Que pouvez-vous déduire de ces résultats?

.....

.....

.....

4. Le présent diagramme de quartiles représente les résultats de François aux examens du ministère de l'Éducation, passés à la fin de son quatrième secondaire, en juin 2004.



- a) Calculez l'étendue de cette distribution.
-
- b) Calculez l'étendue interquartile de cette distribution.
-
- c) Calculez l'intervalle semi-interquartile à partir du diagramme de quartiles.
-
- d) Déterminez dans quel intervalle sont comprises 50 % des données les plus centrales.
-
-



1.3 ACTIVITÉ DE SYNTHÈSE

1. a) Qu'est-ce qui est mesuré par la variance, l'écart type et l'écart moyen?

.....

- b) Quel nom donnons-nous à ce genre de mesure?

.....

2. a) Complétez la phrase ci-dessous en y écrivant le terme approprié dans chacun des espaces prévus à cet effet avant de répondre aux questions suivantes.

Plus la variance est, plus les données d'une distribution sont dispersées. Plus la variance est, plus les données d'une distribution sont regroupées autour de la moyenne arithmétique.

- b) En est-il de même pour l'écart type?

.....

- c) En est-il de même pour l'écart moyen?

.....

3. Quelle est la mesure de dispersion qui ne s'exprime pas dans les mêmes unités que les données de la distribution?

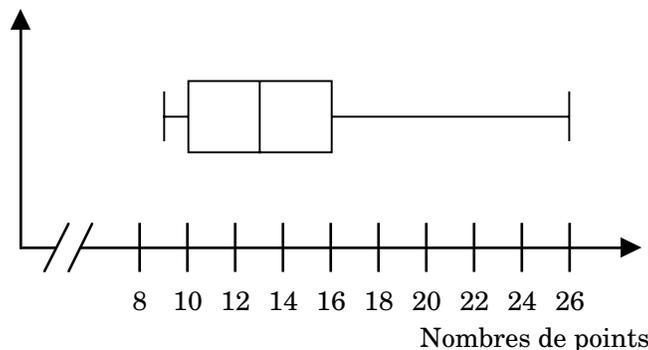
4. Voici le nombre de chambres occupées à un hôtel de Ste-Adèle lors des deux semaines de vacances des travailleurs de la construction pour l'été 2004.

90	45	139	87	110	75	100
108	115	75	112	121	114	130

- a) Quelle est l'étendue de cette distribution?
- b) Quelle en est la moyenne?
.....
- c) Quel est l'écart type de cette distribution?
.....
- d) Quel est l'écart moyen de cette distribution?
.....
.....
.....
.....
.....

5. Le présent diagramme de quartiles représente le nombre de points récoltés par les 26 meilleurs marqueurs de la LNH lors de la conquête de la Coupe Stanley pour la saison 2003-2004.

Meilleurs marqueurs en série, saison 2003-2004



a) Quelle est l'étendue de cette distribution?

.....

b) Quelle est l'étendue interquartile de cette distribution?

.....

c) Quel est l'intervalle semi-interquartile de cette distribution?

.....

1.4 LA PAGE DES MATHOPHILES

L'énigme du facteur $(n - 1)$ au dénominateur des formules de la variance et de l'écart type

Probablement que le fait de voir apparaître le facteur $(n - 1)$ au dénominateur des formules de la variance et de l'écart type en a perturbé plusieurs parmi vous. En effet, pourquoi utiliser $(n - 1)$ plutôt que n ? Évidemment, il y a une bonne raison.

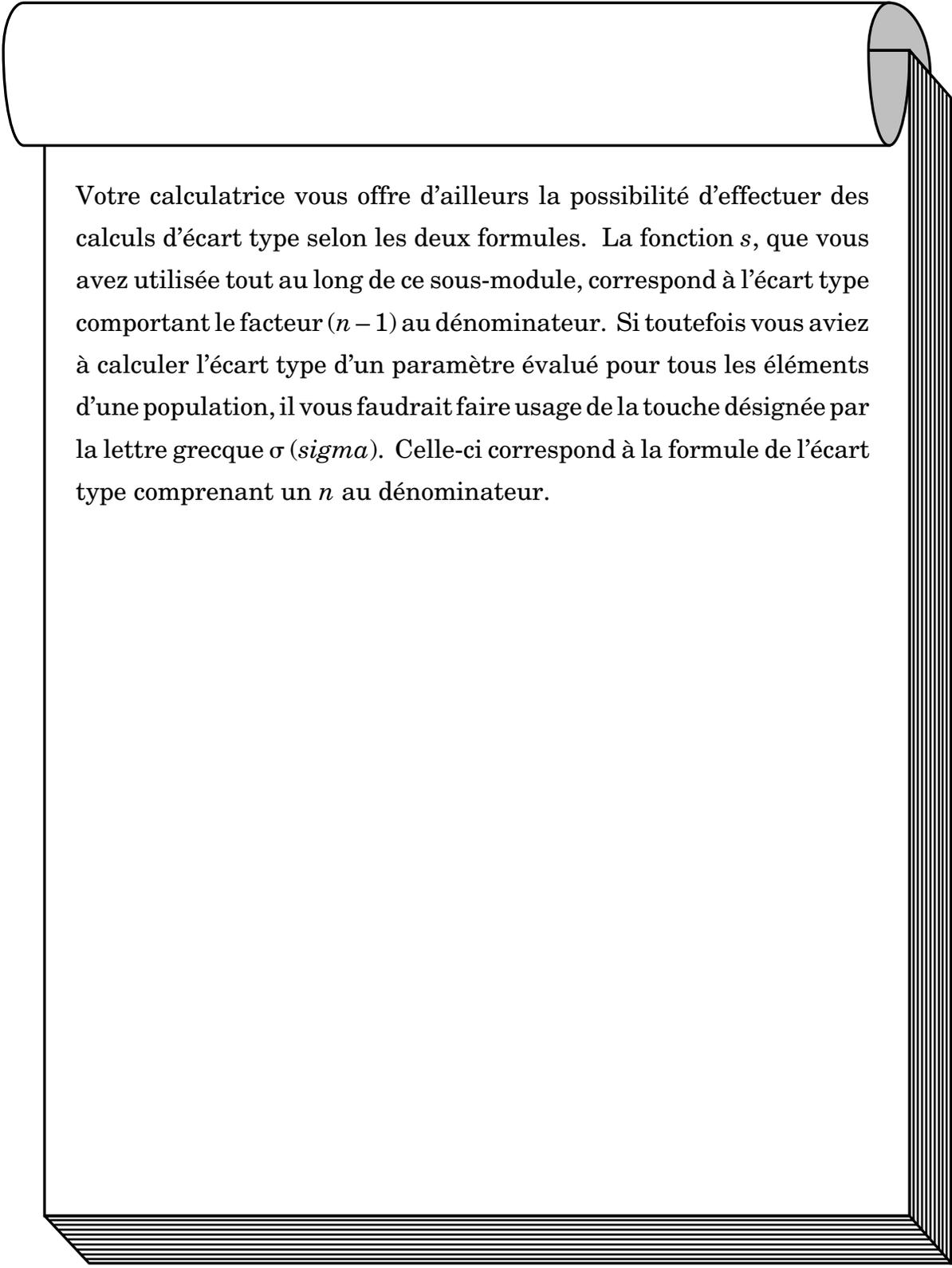
Sans entrer dans les détails, soulignons que des recherches poussées en mathématiques ont démontré que, lorsque des statistiques étaient établies à partir de **tous** les éléments d'une population, la formule ci-dessous était exacte pour calculer la variance du paramètre étudié.

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Cependant, lorsque la variance du paramètre devait être évaluée à partir d'un petit échantillon de population, il était plus juste d'utiliser le facteur $(n - 1)$ au dénominateur de cette formule. La variance ainsi obtenue se rapprochait davantage de la variance de la population.

Dans ce sous-module, puisque les distributions se limitaient à un nombre restreint de données, il est plus approprié d'effectuer des calculs de variance à l'aide de la formule :

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$



Votre calculatrice vous offre d'ailleurs la possibilité d'effectuer des calculs d'écart type selon les deux formules. La fonction s , que vous avez utilisée tout au long de ce sous-module, correspond à l'écart type comportant le facteur $(n - 1)$ au dénominateur. Si toutefois vous aviez à calculer l'écart type d'un paramètre évalué pour tous les éléments d'une population, il vous faudrait faire usage de la touche désignée par la lettre grecque σ (*sigma*). Celle-ci correspond à la formule de l'écart type comprenant un n au dénominateur.

