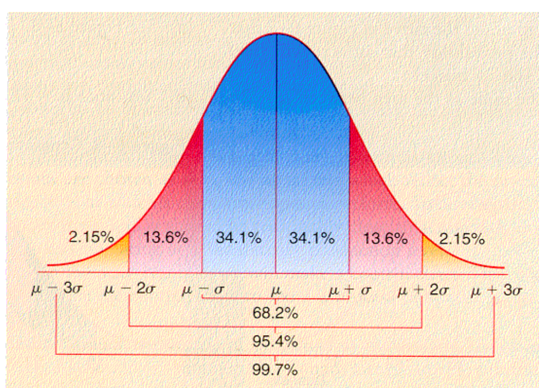
Programme local
Méthodes quantitatives

NOTES DE COURS et EXERCICES

Mathématique CST5
Collège Regina Assumpta
2025 – 2026



Nom : ________________________________

Groupe : _____

NOTES DE COURS

Les notes de cours sont tirées du livre « *Méthodes quantitatives – 4e édition* » de Christiane Simard, paru en 2008 aux éditions Modulo.
Certaines définitions sont tirées de « *Visions mathématique 3* » aux éditions CEC.

1. Définitions importantes

A. Population

L'ensemble des personnes ou des objets sur lesquels porte une étude statistique.

Exemple : L'ensemble des élèves du Collège Regina Assumpta.

B. Échantillon

Sous-ensemble de la population.

Exemples : 1)____

2)____

C. Taille d'une population ou d'un échantillon

Nombre d'éléments qui composent la population ou l'échantillon. On représente par N la taille d'une population et par n la taille d'un échantillon.

Exemple : ____

D. Variable ou caractère d'une étude statistique et modalités

La variable ou le caractère d'une étude statistique est l'élément(s) sur lequel (lesquels) porte l'étude statistique.

Exemples : ____

____

____

Les modalités d'une variable sont les valeurs possibles de la variable.

E. Les types de variables ou de caractères

On distingue 2 types de variables : **qualitatif** et **quantitatif.**

Variable ________________: Les données recueillies sont des **mots** ou des **codes**.

Variable ________________: Les données recueillies sont des **nombres** qui représentent une quantité (compter, dénombrer).

Il est possible de distinguer deux types de variables quantitatives : **discrètes** et **continues**.

Les données recueillies pour une **variable quantitative continue** sont susceptibles de prendre n'importe quelles valeurs réelles dans un intervalle.

Les données recueillies pour une **variable quantitative discrète** sont des valeurs isolées les unes des autres et qui sont généralement entières.

Exemple : Dans les situations suivantes, donne des valeurs possibles et identifie le type de caractère.

Caractère	Valeurs possibles	Type de caractère
a) L'argent que l'on a en poche.		
b) La taille d'un individu.		
c) Le nombre de cheveux sur la tête d'une personne.		
d) La couleur préférée des élèves.		
e) La superficie d'un terrain.		
f) Le nombre de parties gagnées par une équipe sportive durant la saison.		

Bref, pour vérifier si un caractère est quantitatif discret ou continu, on doit se poser la question suivante : **« Est-ce que je peux fractionner les valeurs du caractère autant de fois que je le veux ? »**

F. Le sondage, l'enquête et le recensement

Le **recensement** est une étude statistique menée sur l'ensemble de la population concernée par l'étude.

Le **sondage** est une étude statistique menée sur un échantillon.

G. Distribution statistique

Une distribution statistique à une variable est présentée par un tableau, appelé tableau de distribution, qui donne la répartition des n données de l'échantillon selon les modalités de la variable.

Exemple : On a relevé le nombre de buts marqués au hockey lors des 20 (n) premières parties de hockey de la saison par les Canadiens de Montréal.

2	3	4	3	1	0	2	3	2	4
3	4	2	1	3	2	4	3	3	3

La variable « nombre de buts » prend les valeurs (modalités) 0, 1, 2, 3 et 4. Le tableau ci-contre indique la répartition des 20 parties de hockey selon la variable « nombre de buts ». Complétez le tableau.

Distribution des parties selon le nombre de buts

Nombre de buts (x_i)	Nombre de parties (n_i)
0	
1	
2	
3	
4	
Total	

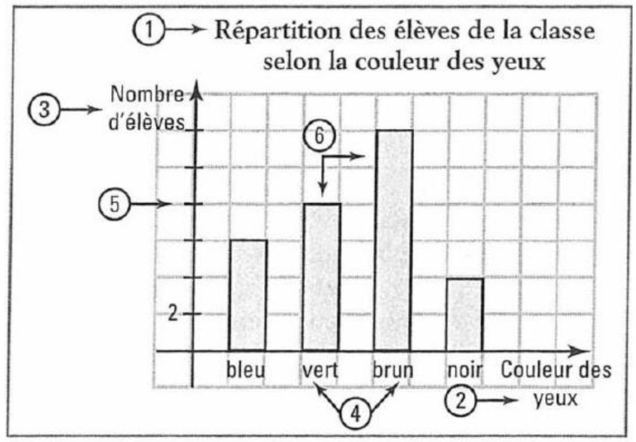
H. Diagramme à bandes

Un diagramme à bandes permet d'illustrer une variable **qualitative** ou **quantitative discrète**.

Exemple : Complétez le tableau de distribution représenté par le diagramme à bandes.

Répartition des élèves de la classe selon la couleur des yeux

Couleur des yeux	Nombre d'élèves
Total	



Les principaux éléments sont :

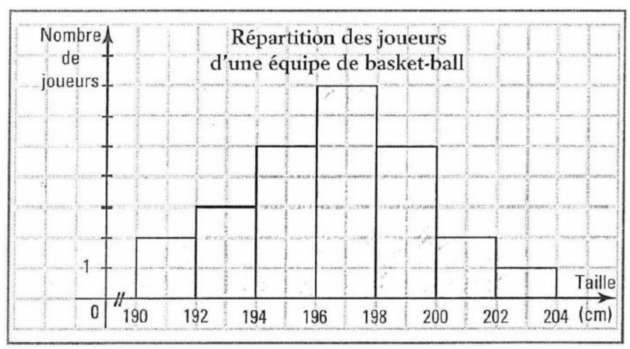
1- le titre ;
2- l'identification de l'axe horizontal : la variable « couleur des yeux » ;
3- l'identification de l'axe vertical : les fréquences « nombre d'élèves » ;
4- l'identification des bandes (les modalités de la variable) : bleu, vert, … ;
5- la graduation de l'axe vertical : l'échelle utilisée tient compte des fréquences ;
6- les bandes ont toutes la même largeur et sont également espacées. La hauteur de chaque bande est proportionnelle à la fréquence.

Dans un diagramme à bandes, les bandes peuvent être représentées **verticalement** ou **horizontalement**.

I. Histogramme

- L'histogramme représente une distribution où les données sont principalement quantitatives continues.
- Diagramme formé de bandes rectangulaires collées les unes sur les autres.
- La largeur des bandes est toujours identique dans l'histogramme.
- Les données sont regroupées par classe ou par intervalles de même longueur et placées sur l'axe horizontal.
- La hauteur des bandes, lue sur l'axe vertical, est en fonction de l'effectif, c'est-à-dire en fonction du nombre d'éléments de chaque classe.
- Un titre présente toujours l'histogramme.
- Les deux axes sont bien identifiés et gradués, c'est-à-dire que les bonds sont réguliers sur chacun des axes.
- Il est possible de faire une coupure d'axe seulement sur l'axe horizontal.

Exemple : L'histogramme ci-contre illustre la répartition des joueurs d'une équipe de basket-ball selon la taille (en cm).



On observe que :

- les joueurs ont une taille variant entre 190 cm et 204 cm.
- La classe [196, 198[renferme le plus grand nombre de joueurs.

2. Les mesures de tendance centrale

Lorsque vient le moment d'analyser un ensemble de données recueillies lors d'une étude statistique, il est naturel de s'intéresser aux données qui représentent le mieux le *centre* de la distribution. Par contre, ce centre peut être vu de différentes manières. Nous verrons trois mesures de tendance centrale : la moyenne, le mode et la médiane.

A. Moyenne arithmétique

Lorsqu'on désire représenter une série de données par un seul nombre, comme les notes d'un examen, la première mesure à laquelle on pense est la moyenne des données. La moyenne est la mesure de tendance centrale la plus connue et la plus utilisée comme représentante des données d'une série statistique. Avec la mise en situation suivante, nous apprendrons à représenter graphiquement une moyenne et à la calculer de trois façons différentes : avec les données brutes, avec les effectifs du tableau de distribution et avec les pourcentages du tableau de distribution.

Prenons la série statistique donnant le nombre de programmes préuniversitaires offerts dans chacun des 48 collèges publics du Québec en 2006 :

4	3	5	5	3	4	5	7	5	5	6	6
6	3	3	3	5	5	6	4	5	5	4	6
4	5	3	5	4	4	4	4	7	4	3	4
6	7	4	3	7	4	3	6	4	7	3	5

Source : Ministère de l'Éducation, Direction de la recherche, des statistiques et des indicateurs, Système prévisionnel SIPEEC, 2006.

i. Calcul de la moyenne avec des données brutes

Comme nous le savons, le calcul de la moyenne avec des données brutes consiste à additionner l'ensemble de ces données, puis à diviser la somme obtenue par le nombre total de données.

Moyenne avec des données brutes
$\text{Moyenne} = \dfrac{\text{somme des données}}{\text{nombre total de données}}$

Écriture symbolique :

Nous noterons la moyenne par le symbole μ, qui se lit « mu » (*m* dans l'alphabet grec), et le nombre de données par la lettre *N.*

En représentant chaque donnée de la série statistique par les symboles : x_1, x_2, x_3, x_4 et ainsi de suite, on obtient la formule suivante pour décrire le calcul d'une moyenne avec les données brutes :

$$\mu = \frac{x_1 + x_2 + x_3 + \cdots + x_N}{N}$$

On peut simplifier cette formule à l'aide de la notation sigma, symbolisée par « Σ » (S dans l'alphabet grec). Ce symbole indique que l'on doit faire la somme de tous les termes de forme x_i, l'indice i variant de 1 à *N*.

$$\mu = \frac{\Sigma x_i}{N}$$

Dans le cadre d'un sondage, on emploie le symbole $\overline{x}$ pour désigner la moyenne des données de l'échantillon et μ pour la moyenne des données de la population. De même, on utilise le symbole n pour désigner le nombre de données de l'échantillon et N pour le nombre de données de la population. La formule pour obtenir la moyenne des données de l'échantillon s'écrit donc :

$$\overline{x} = \frac{\Sigma x_i}{n}$$

Pour la moyenne des données de la mise en situation, on a :

$$\mu = \frac{\Sigma x_i}{N} = \frac{4+3+5+\cdots+3+5}{48} = \frac{223}{48} \approx 4{,}6 \text{ programmes préuniversitaires par collège}$$

<u>Interprétation :</u> En 2006, si tous les collèges publics du Québec avaient offert le même nombre de cours, il y aurait eu 4,6 programmes préuniversitaires dans chaque collège.

Attention!!: Le résultat du calcul d'une moyenne brute ne doit pas être arrondi à l'entier sous prétexte que les données brutes sont entières : la moyenne est un nombre théorique. <u>Nous conviendrons de conserver une décimale après la virgule.</u>

ii. Calcul de la moyenne avec les effectifs du tableau de distribution

Répartition des 48 collèges publics du Québec selon le nombre de programmes préuniversitaires offerts, 2006

Nombre de programmes préuniversitaires	Nombre de collèges	Pourcentage de collèges
3	10	20,8 %
4	14	29,2 %
5	12	25,0 %
6	7	14,6 %
7	5	10,4 %
Total	48	100,0 %

Avec les effectifs du tableau de distribution, on calcule ainsi la moyenne du nombre de programmes préuniversitaires offerts par les collèges publics du Québec :

Moyenne avec des effectifs
$\text{Moyenne} = \dfrac{\text{somme des produits de chaque valeur de la variable par son effectif}}{\text{nombre total de données}}$

Écriture symbolique :

En notant par la lettre n_i l'effectif correspondant à la valeur x_i, on obtient la formule suivante :
$\mu = \frac{x_1 n_1 + x_2 n_2 + x_3 n_3 + \cdots + x_k n_k}{N}$, où k représente le nombre de valeurs de la variable.

En utilisant la notation sigma, on obtient :

$$\mu = \frac{\Sigma x_i n_i}{N}, \text{ pour } i \text{ variant de 1 à } k.$$

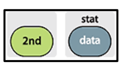
Dans l'exemple, on obtient donc :

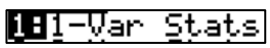
$$\mu = \frac{\Sigma x_i n_i}{N} = \frac{3 \times 10 + 4 \times 14 + 5 \times 12 + 6 \times 7 + 7 \times 5}{48} = \frac{223}{48} \approx 4{,}6$$

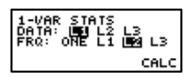
Entrer les données dans la calculatrice TI-30XS MultiView

1- 

: Entrer les modalités dans la première colonne et les effectifs dans la deuxième colonne

2- 

→ 

→ 

iii. Calcul de la moyenne avec les pourcentages du tableau de distribution

Pour trouver une formule qui permettrait de calculer une moyenne avec les pourcentages, reprenons le calcul de la moyenne avec les effectifs et apportons les modifications suivantes :

$$\mu = \frac{3\times10 + 4\times14 + 5\times12 + 6\times7 + 7\times5}{48} = 4{,}6$$

$$\mu = \frac{3\times10}{48} + \frac{4\times14}{48} + \frac{5\times12}{48} + \frac{6\times7}{48} + \frac{7\times5}{48} = 4{,}6$$

$$\mu = 3\times\frac{10}{48} + 4\times\frac{14}{48} + 5\times\frac{12}{48} + 6\times\frac{7}{48} + 7\times\frac{5}{48} = 4{,}6$$

$$\mu = 3\times20{,}8\,\% + 4\times29{,}2\,\% + 5\times25\,\% + 6\times14{,}6\,\% + 7\times10{,}4\,\% = 4{,}6$$

Moyenne avec des pourcentages
$\text{Moyenne} = \text{somme des produits de chaque valeur de la variable par son pourcentage}$

<u>Attention!!:</u> Il est à remarquer que nous n'avons pas à diviser, dans ce cas-ci, par le total des données, car cette division a déjà été faite dans le calcul du pourcentage.

Écriture symbolique

En notant f_i la fréquence relative correspondant à la valeur x_i, on obtient la formule suivante :

$$\mu = x_1 f_1 + x_2 f_2 + x_3 f_3 + \cdots + x_k f_k = \Sigma x_i f_i$$

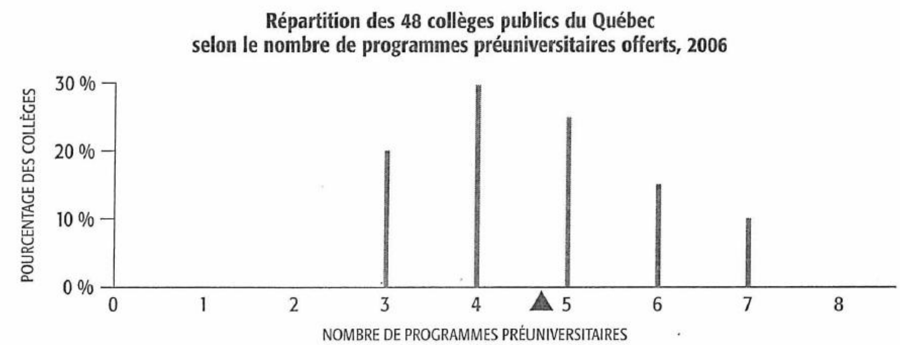
pour i variant de 1 à k, où k est le nombre de valeurs de la variable.

Dans l'exemple, on obtient donc :

$$\mu = \Sigma x_i f_i = 3 \times 20{,}8\ \% + 4 \times 29{,}2\ \% + 5 \times 25\ \% + 6 \times 14{,}6\ \% + 7 \times 10{,}4\ \% \approx 4{,}6$$

iv. Représentation graphique de la moyenne

Voici le diagramme en bâtons de la distribution du nombre de programmes préuniversitaires :



En plaçant un pivot sous l'axe horizontal du diagramme en bâtons à l'endroit où se situe la moyenne, on remarque que la moyenne correspond graphiquement au centre d'Équilibre du diagramme. On applique ici le principe des balançoires à bascule : on doit s'imaginer que l'axe horizontal est une planche sous laquelle il faut placer un pivot pour que les bâtons du diagramme se trouvent en position d'équilibre sur cette planche.

Exemple 1 :

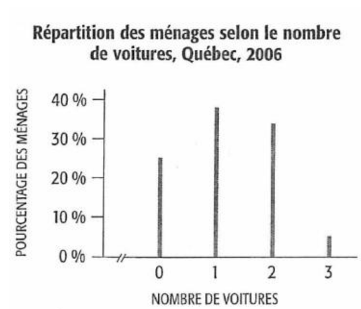
a) Calculez et interprétez la moyenne en remplaçant la catégorie « 3 et plus » par le nombre 3.

Répartition des ménages selon le nombre de voitures, Québec, 2006

Nombre de voitures	Pourcentage
0	25 %
1	38 %
2	32 %
3 et plus	5 %
Total	100 %

Source : *L'actualité*, mai 2007.

b) Représentez graphiquement la moyenne en plaçant le pivot au bon endroit.

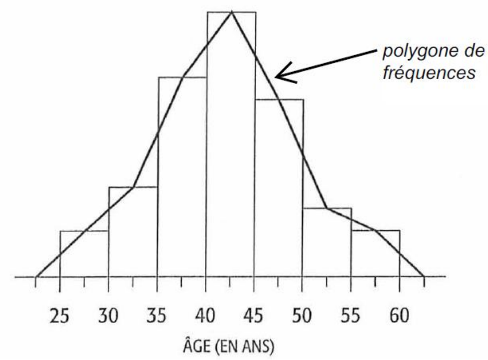


Exemple 2 : À la suite de l'étude menée auprès des 40 professeurs d'une école, nous avons construit le tableau de distribution et l'histogramme suivant.

Répartition des professeurs selon l'âge

Âge (en ans)	Nombre de professeurs
$25 \leq X < 30$	2
$30 \leq X < 35$	4
$35 \leq X < 40$	9
$40 \leq X < 45$	12
$45 \leq X < 50$	8
$50 \leq X < 55$	3
$55 \leq X < 60$	2
Total	40

Répartition des professeurs selon l'âge



a) Calculer et interpréter la moyenne de la distribution.

b) Placer un pivot sous l'axe horizontal du graphique afin d'estimer la moyenne de la distribution.

Attention!!: Avec un histogramme ou un tableau de données groupées, la moyenne n'est pas exacte, mais c'est le mieux que l'on peut faire.

v. Moyenne pondérée

Lorsque le calcul d'une moyenne se fait en multipliant chaque valeur d'une série par sa pondération, on dit que l'on calcule la moyenne pondérée de la série. La pondération est déterminée selon l'importance qui est accordée à chaque valeur par rapport aux autres valeurs de la série. La somme des pondérations doit toujours être égale à 100 %.

Exemple 1 : Supposons qu'un étudiant ait obtenu les notes suivantes en français :
Examen 1 : 65 % Examen 2 :70 % Travail : 75 %

Trouvez sa moyenne pour la session si la pondération de chaque évaluation est :
Examen 1 : 25 % Examen 2 : 35 % Travail : 40 %

Exemple 2 :

Revenu moyen des familles monoparentales par province, Canada, 2005

Terre-Neuve-et-Labrador	22 300 $	Ontario	40 300 $
Île-du-Prince-Édouard	23 100 $	Manitoba	30 300 $
Nouvelle-Écosse	28 300 $	Saskatchewan	22 800 $
Nouveau-Brunswick	22 100 $	Alberta	41 000 $
Québec	35 300 $	Colombie-Britannique	32 400 $

Source : Statistique Canada, CANSIM, tableau 202-0202.

a) Calcule la moyenne de ces 10 revenus.

b) Peut-on dire que la moyenne trouvée ci-haut est le revenu moyen des familles monoparentales vivant dans les provinces canadiennes? Justifier.

c) Utilisons le tableau suivant pour pondérer le calcul de la moyenne.

Répartition des familles monoparentales par province, Canada, 2005

Terre-Neuve-et-Labrador	1,8 %	Ontario	39,3 %
Île-du-Prince-Édouard	0,5 %	Manitoba	3,9 %
Nouvelle-Écosse	3,5 %	Saskatchewan	3,5 %
Nouveau-Brunswick	2,6 %	Alberta	9,7 %
Québec	23,0 %	Colombie-Britannique	12,2 %

Source : Statistique Canada, CANSIM, tableau 111-0009.

Bref, le fait que 39,3 % des familles monoparentales vivent en Ontario et que ces dernières ont, après l'Alberta, le revenu moyen le plus élevé (40 300 $) fera augmenter la valeur trouvée pour μ à la question a).

d) Interpréter la moyenne pondérée.

Exemple 3 : Calculer le salaire moyen des 200 employés d'une usine sachant que les 10 cadres gagnent en moyenne 45 200 $, les 50 techniciens 35 250 $ et que les 140 ouvriers 28 400 $. Interpréter la moyenne et dire si elle est représentative.

B. Mode et classe modale

Le mode est la valeur ou la catégorie qui revient le plus souvent dans une série statistique. La classe modale est la classe qui regroupe le plus de données d'une série statistique. On considère le centre de cette classe comme une approximation du mode de la distribution.

Attention!!: Pour être significatif, le mode doit avoir une fréquence qui se démarque de celle des autres. Il en est de même pour la classe modale.

Exemple 1 : Donner et interpréter le mode de la distribution ci-dessous.

Répartition des 48 collèges publics du Québec selon le nombre de programmes préuniversitaires offerts, 2006

Nombre de programmes préuniversitaires	3	4	5	6	7	Total
Nombre de collèges	10	14	12	7	5	48
Pourcentage de collèges	20,8 %	29,2 %	25,0 %	14,6 %	10,4 %	100,0 %

Le mode est : _________

Interprétation
En 2006, une pluralité de collèges publics du Québec (______ %) offraient _____ programmes pré-universitaires.

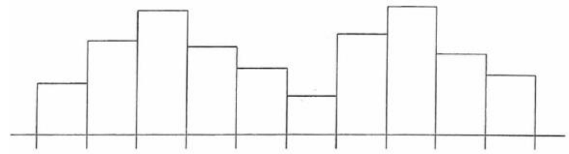
Attention!!: Le mot **pluralité** signifie « le plus grand nombre »; il indique que le pourcentage entre parenthèses est le plus haut pourcentage de la distribution. Quand le pourcentage est supérieur à 50 %, on peut utiliser le mot **majorité**.

Exemple 2 : Donner et interpréter la classe modale de la distribution suivante.

Répartition des professeurs selon l'âge

Âge (en ans)	Nombre de professeurs	Pourcentage
$25 \leq X < 30$	2	5,0 %
$30 \leq X < 35$	4	10,0 %
$35 \leq X < 40$	9	22,5 %
$40 \leq X < 45$	12	30,0 %
$45 \leq X < 50$	8	20,0 %
$50 \leq X < 55$	3	7,5 %
$55 \leq X < 60$	2	5,0 %
Total	40	100,0 %

Exemple 3 : Un club mixte de conditionnement physique a fait une étude sur le poids des personnes membres. Une fois les données compilées, on obtient l'histogramme ci-contre.



a) On dit d'une telle distribution qu'elle est bimodale, car deux classes se démarquent des autres avec approximativement la même fréquence. Pouvez-vous indiquer la cause de cette bimodalité?

Attention!!: Une bimodalité peut indiquer la présence de deux sous-populations plus homogènes que la population globale.

b) Serait-il judicieux, dans le cadre d'une campagne de promotion du club, d'utiliser la moyenne de cette série de données comme mesure de tendance centrale pour représenter le poids de ses membres?

Exemple 4 : Choisir et interpréter la meilleure mesure de tendance centrale, entre la moyenne et le mode, pour les distributions suivantes.

a) Les résultats d'un sondage publié dans la revue L'actualité en 2007 :

Répartition des répondants selon leur perception de l'influence des États-Unis sur le Québec, 2007

Influence	Positive	Négative	Neutre	Pas d'opinion	Total
Pourcentage des répondants	23 %	64 %	5 %	8 %	100 %

Source : *L'actualité*, 1er mai 2007, sondage Crop – L'actualité – Radio-Canada.

b) Les salaires de cinq ingénieurs travaillant pour une entreprise :

41 500 $ 42 250 $ 58 550 $ 64 750 $ 120 800 $

C. Médiane

La médiane, que l'on note « Md », est la valeur qui partage une série de données ordonnées en deux parties égales, chacune comprenant le même nombre de données. En d'autres termes, une valeur est la médiane d'une série de données (ordonnées) s'il y a autant de données à gauche qu'à droite de cette valeur.

i. Données non groupées en classe

Pour calculer la médiane, il faut :

Pour un nombre **impair** de données :	Pour un nombre **pair** de données :
1) Placer les données en ordre croissant. 2) Trouver la donnée centrale dans la liste.	1) Placer les données en ordre croissant. 2) Effectuer la moyenne entre les deux données du centre.

Exemple 1: Trouver et interpréter la médiane des séries statistiques suivantes.

a) Les salaires de cinq ingénieurs travaillant pour une entreprise :

41 500 $ 42 250 $ 58 550 $ 64 750 $ 120 800 $

La médiane est __________. C'est la meilleure mesure de tendance centrale de la série, car elle n'est pas influencée par les valeurs extrêmes contrairement à la moyenne.

Interprétation

Dans cette entreprise, au moins 50 % des ingénieurs gagnent ___________ $ ou moins.

Attention!!: On dit « au moins » 50 %, car trois ingénieurs sur cinq gagnent la médiane ou moins.

b) Poids à la naissance de 10 nouveau-nés :

2 350 g	3 150 g	3 252 g	3 334 g	3 552 g
3 843 g	3 926 g	4 125 g	4 650 g	3 684 g

<u>Interprétation</u>

c)

Répartition des 48 collèges publics du Québec selon le nombre de programmes préuniversitaires offerts, 2006

Nombre de programmes préuniversitaires	3	4	5	6	7	Total
Nombre de collèges	10	14	12	7	5	48

<u>Interprétation</u>

Exemple 2: Vrai ou faux? Si la médiane d'un examen sur 100 points est 68, on peut alors dire qu'au moins 50 % des étudiants ont une note de 68 % à cet examen.

ii. Données groupées en classes

Lorsque les données d'une série statistique sont groupées en classes, la médiane est égale à la valeur sur l'axe horizontal qui divise la surface de l'histogramme (donc les données, en vertu du principe de la proportionnalité) en deux parties égales.

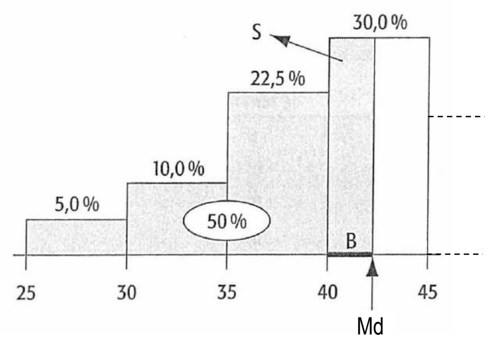
Exemple 1: Trouver la médiane de la distribution suivante.

Notre objectif est de trouver un âge sur l'axe horizontal tel que 50 % de la surface de l'histogramme (donc 50 % des données) se situe à gauche de cet âge. Pour trouver 50 % de la surface, il faut additionner les surfaces du 1er rectangle, du 2e rectangle, du 3e rectangle et un partie S de la surface du 4e rectangle.

Répartition des professeurs selon l'âge

Âge (en ans)	Pourcentage
$25 \leq X < 30$	5,0 %
$30 \leq X < 35$	10,0 %
$35 \leq X < 40$	22,5 %
$40 \leq X < 45$	30,0 %
$45 \leq X < 50$	20,0 %
$50 \leq X < 55$	7,5 %
$55 \leq X < 60$	5,0 %
Total	100,0 %

Esquisse de l'histogramme



1) Surface (50 %)

50 % = 5 % + 10 % + 22,5 % + S
50 % = 37,5 % + S
S = 50 % - 37,5 %
S = 12,5 %

2) Médiane
Md = 40 ans + B ans, où B est la longueur de la base du rectangle de surface S.

3) Valeur manquante
30 % → 5 ans
12,5 % → B

$$B = \frac{12{,}5\% \times 5}{30\%} \approx 2{,}1 \ ans$$

4) Médiane
Md = 40 + 2,1 = 42,1 ans

Interprétation
On peut estimer que 50 % des professeurs ont moins de 42,1 ans.

Exemple 2: Le tableau suivant donne la distribution de l'âge des arbres recensés sur un terrain boisé. Trouver et interpréter la médiane de cette distribution.

Répartition des arbres selon l'âge

Âge (en ans)	Pourcentage
[0 ; 10[	8 %
[10 ; 20[	28 %
[20 ; 30[	32 %
[30 ; 40[	20 %
[40 ; 50[	12 %
Total	100 %

Interprétation

3. Les mesures de position

Les mesures de position permettent de situer une donnée par rapport aux autres données d'une série statistique. Dans cette section, nous étudierons les quantiles.

A. Les différentes mesures de position

Les quantiles sont les valeurs qui partagent une distribution en un certain nombre de parties égales. Les plus utilisés sont :

- ★ les **quartiles**(Q_1, Q_2, Q_3), qui partagent une distribution en quatre parties comprenant 25 % des données ;
- ★ les **quintiles**(V_1, V_2, V_3, V_4), qui partagent une distribution en cinq parties comprenant 20 % des données ;
- ★ les **déciles**$(D_1, D_2, \dots, D_9)$, qui partagent une distribution en dix parties comprenant 10 % des données ;
- ★ les **centiles**$(C_1, C_2, \dots, C_{99})$, qui partagent une distribution en 100 parties comprenant 1% des données.

B. Données groupées en classes

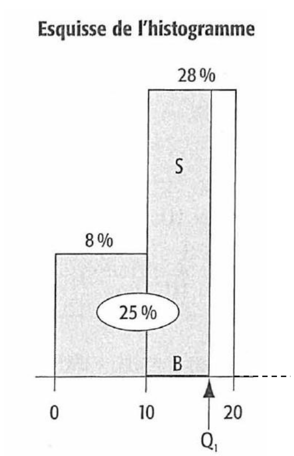
Exemple 1: Reprenons la distribution de l'âge des arbres recensés sur un terrain boisé.

Nous avons déjà estimé que 50% des arbres avaient moins de 24,4 ans. Supposons qu'un voisin curieux pose la question suivante au propriétaire du boisé : « Le quart de vos arbres ont moins de quel âge? »

Répartition des arbres selon l'âge

Âge (en ans)	Pourcentage
[0 ; 10[	8 %
[10 ; 20[	28 %
[20 ; 30[	32 %
[30 ; 40[	20 %
[40 ; 50[	12 %
Total	100 %

Pour répondre à cette question, il faut trouver un âge, disons Q_1, tel que 25 % des arbres aient un âge inférieur à Q_1. La démarche qui permettra de déterminer l'âge Q_1 est analogue à celle que nous avons suivie pour trouver la médiane.



1) Surface

25 % = 8 % + S
S = 25 % - 8 %
S = 17 %

2) Quartile 1

Q_1 = 10 + B

3) Valeur manquante

28 % → 10 ans
17 % → B

$$B = \frac{17\ \% \times 10\ ans}{28\ \%} \approx 6{,}1\ ans$$

4) Quartile 1

$$Q_1 = 10 + 6{,}1 \approx 16{,}1\ ans$$

Interprétation

On peut estimer que 25 % des arbres ont moins de 16,1 ans.

Exemple 2: Quelle réponse le propriétaire du boisé de la mise en situation devrait-il donner à son voisin si celui-ci, de plus en plus indiscret, pose la question suivante : « 10 % de vos arbres ont moins de quel âge? »

Exemple 3: Trouver et interpréter le 36e centile.

Exemple 4: Pour la distribution de l'âge des arbres, 30 ans correspond à quel centile?

Attention!!: Lorsqu'on veut déterminer le pourcentage de données d'une distribution qui sont inférieurs à une certaine valeur, on dit que l'on cherche le rang centile de cette valeur (ici le rang centile de 30 est 68).

C. Données non groupées en classes

Pour trouver les quantiles de données non groupées en classes, nous allons appliquer une procédure analogue à celle qui a été retenue pour déterminer la médiane.

- ★ Si $(i\% \times N)$ est un entier, le centile C_i est la moyenne de la $(i\% \times N)^e$ donnée et de la donnée suivante.
- ★ Si $(i\% \times N)$ n'est pas un entier, le centile C_i est la donnée dont le rang est l'entier qui suit $(i\% \times N)$.

Exemple 1: Poids à la naissance de 10 nouveau-nés :

2 350 g	3 150 g	3 252 g	3 334 g	3 552 g
3 843 g	3 926 g	4 125 g	4 650 g	3 684 g

a) Trouver et interpréter le deuxième décile de la série statistique.

b) Trouver et interpréter le troisième quartile.

Exemple 2: Trouver et interpréter le 65^{e} centile de la distribution suivante.

Répartition des répondants selon le nombre de films loués au cours du dernier mois

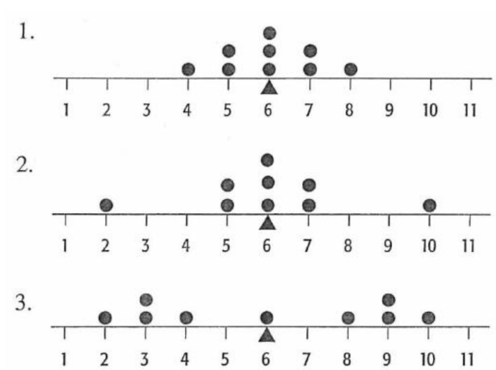
Nombre de films loués en un mois	Nombre de répondants
4	26
5	28
6	21
7	19
8 et plus	21
Total	115

4. Les mesures de dispersion

Les mesures de dispersion permettent de mesurer la dispersion des données d'une série statistique. Nous étudierons plus particulièrement l'étendue et l'écart-type.

A. Étendue

Analysons les trois séries statistiques suivantes, représentées par un pictogramme (il s'agit du nom de ce type de graphique).



Ces trois séries statistiques présentent des moyennes identiques, mais la dispersion des données autour de ces moyennes diffère d'une série à l'autre. Les données de la série 1 sont assez concentrées autour de la moyenne, alors que la série 2 comporte deux données qui sont passablement éloignées de la moyenne. Les données de la série 3 sont encore plus dispersées par rapport à la moyenne que celles de la série 2. Comment peut-on mesurer mathématiquement cette dispersion?

L'étendue de la série pourrait peut-être nous permettre de la mesurer. Calculons l'étendue de chacune des séries :

	Série 1	Série 2	Série 3
Étendue	4	8	8

Comme l'étendue ne tient compte que de la plus grande et de la plus petite valeur de la série, on ne peut mesurer la différence de dispersion entre les données de la série 2 et de la série 3.

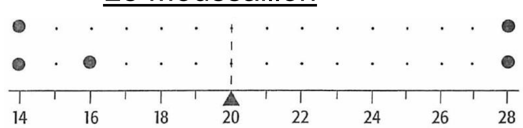
En fait, l'étendue sert très peu comme mesure de dispersion. On l'utilise surtout en contrôle de la qualité où les échantillons sont souvent de petites tailles (4 ou 5 éléments); dans ce cas, l'étendue donne une bonne idée de la dispersion des résultats de l'échantillon.

B. Variance et écart type

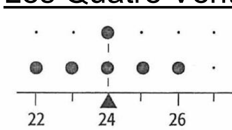
L'écart type est une mesure de dispersion qui, contrairement à l'étendue, tient compte de toutes les valeurs de la série de données. La variance est une donnée intermédiaire qui permet de calculer l'écart type.

Exemple 1: Supposons qu'un étudiant de cégep offre ses services à titre de matelot pendant les vacances. Il a reçu deux offres d'emploi avec des conditions de travail semblables. Il décide de faire son choix en se basant sur l'âge moyen de ses compagnons de voyage. Il a donc préféré l'offre d'emploi du voilier *Le Moussaillon*, où l'âge moyen est de 20 ans, à celui du voilier *Les Quatre Vents*, où il est de 24 ans. Les pictogrammes ci-dessous indiquent la distribution de l'âge pour chaque voilier. Notre cégépien a-t-il fait le bon choix?

Le Moussaillon

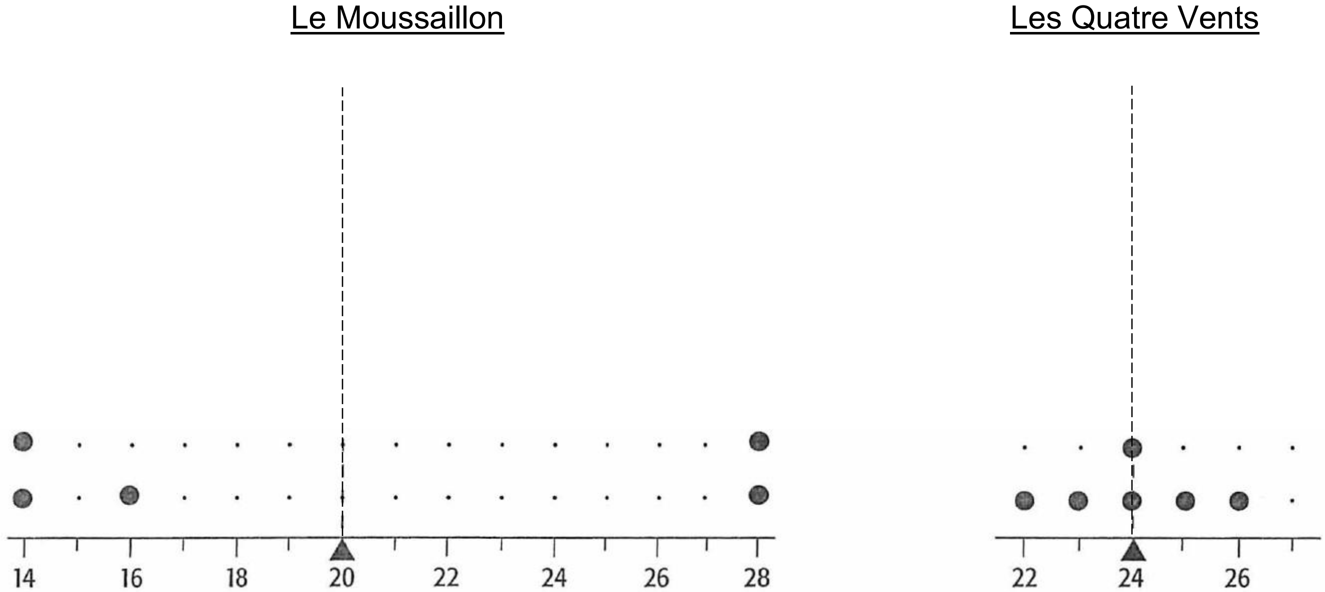


Les Quatre Vents



Bien sûr que non : il devra passer ses vacances avec M. et Mme Tremblay et leurs trois neveux, alors que l'autre voilier semble accueillir une équipe jeune et dynamique. La différence principale entre les deux distributions est la dispersion des données par rapport à la moyenne. Comment s'y prendre pour mesurer cette dispersion?

★ Pour chacune des données des pictogrammes, tracer un segment de droite représentant l'écart entre la donnée et la moyenne, et inscrire la valeur de cet écart sur le segment de droite.



a) Pour le voilier *Le Moussaillon*, calculer la moyenne des écarts d'âge par rapport à la moyenne.

b) Que remarques-tu?

La moyenne des écarts donnera toujours 0. En effet, la moyenne étant le centre d'équilibre du pictogramme, la somme des écarts située à gauche de la moyenne sera toujours égale, mais de signe négatif, à la somme des écarts située à droite de la moyenne.

c) Pour le voilier *Le Moussaillon*, calculer la moyenne des écarts d'âge <u>en valeur absolue</u> par rapport à la moyenne.

Comme personne n'aime travailler avec des valeurs absolues, nous allons transformer cette formule pour qu'elle soit plus facile à exploiter. Un moyen approprié d'éliminer les valeurs absolues consiste à appliquer le principe suivant : la valeur absolue d'un nombre élevée au carré donne le même résultat que donnerait ce nombre élevé au carré.

Par exemple, $|-4|^2 = (-4)^2 = 16$. Ce principe se nomme « variance ».

d) Calculer la variance pour le voilier *Le Moussaillon*.

Comme la variance nous donne la moyenne des carrés des écarts, il faut extraire la racine de la variance pour obtenir un écart typique de tous les écarts de la série. On donne le nom d'écart type à cet écart que l'on note σ, ou sigma (s dans l'alphabet grec); la variance quant à elle est notée σ^2.

e) Calculer l'écart type de l'âge de l'équipage du voilier *Le Moussaillon* et le représenter sur le pictogramme de la page précédente par un segment de droite.

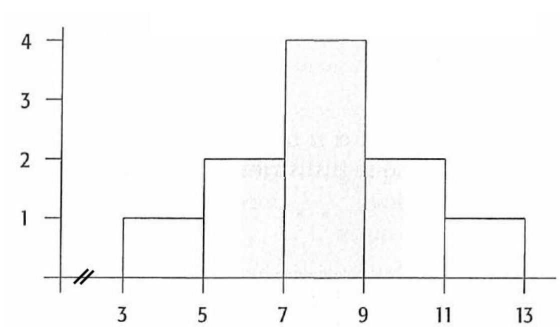
Généralement, on trouve **la plupart** des données d'une distribution entre la moyenne moins un écart type et la moyenne plus un écart type, soit entre $\mu - \sigma$ et $\mu + \sigma$. Lorsque la distribution prend la forme d'une cloche (modèle normal), environ les deux tiers des données sont comprises dans cet intervalle.

f) Donner l'interprétation de l'écart-type pour le voilier *Le Moussaillon*.

g) Calculer l'écart type de l'âge de l'équipage du voilier *Les Quatre Vents* et le représenter sur le pictogramme de la page précédente par un segment de droite. Donner l'interprétation de cet écart type.

Écart type et variance
$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\text{somme des carrés des écarts à la moyenne}}{\text{nombre total de données}}} = \sqrt{\frac{\Sigma(x_i-\mu)^2 n_i}{N}} = \sqrt{\Sigma(x_i-\mu)^2 f_i}$

Exemple 2: Calculer, représenter graphiquement et interpréter l'écart type de la distribution suivante.

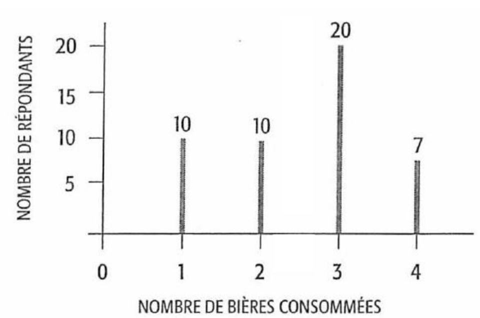


C. Écart type corrigé d'un échantillon

Dans le cadre d'une étude par sondage, l'écart type de l'échantillon est retenu pour estimer celui de la population. Toutefois, les statisticiens ont démontré que l'estimation est meilleure si l'on divise le numérateur de la formule de l'écart type par le nombre de données de l'échantillon, moins 1. On obtient ainsi l'écart type corrigé, qui prend la notation s. En se rappelant que la moyenne de l'échantillon se note $\overline{x}$ et le nombre de données n, on obtient la formule suivante pour calculer l'écart type corrigé :

Écart type corrigé
$s = \sqrt{\text{variance corrigée}} = \sqrt{\frac{\text{somme des carrés des écarts à la moyenne}}{\text{nombre total de données}-1}} = \sqrt{\frac{\Sigma(x_i - \overline{x})^2 n_i}{n-1}}$

Exemple 1: À la suite d'un 5 à 7 organisé par une entreprise pour souligner le départ d'un employé, on a demandé à un échantillon de 47 personnes combien de bières elles avaient consommées. Le diagramme suivant donne la répartition des répondants selon le nombre de bières consommées. Calculer, représenter graphiquement et interpréter l'écart-type corrigé de cette distribution.



5. **La cote Z**

La cote Z permet de situer une donnée par rapport aux autres données d'une série statistique.

Exemple 1: Un employeur désire engager un étudiant pour l'été afin de l'aider à terminer une étude de marché. Il examine donc le dossier de quatre étudiants qui viennent de terminer un cours de statistique et voudrait bien embaucher le meilleur d'entre eux. Voici les résultats :

	Note	Moyenne du groupe	Écart type du groupe
Jean	85	75	10
Lise	76	70	3
Chantale	70	60	4
Luc	75	80	5

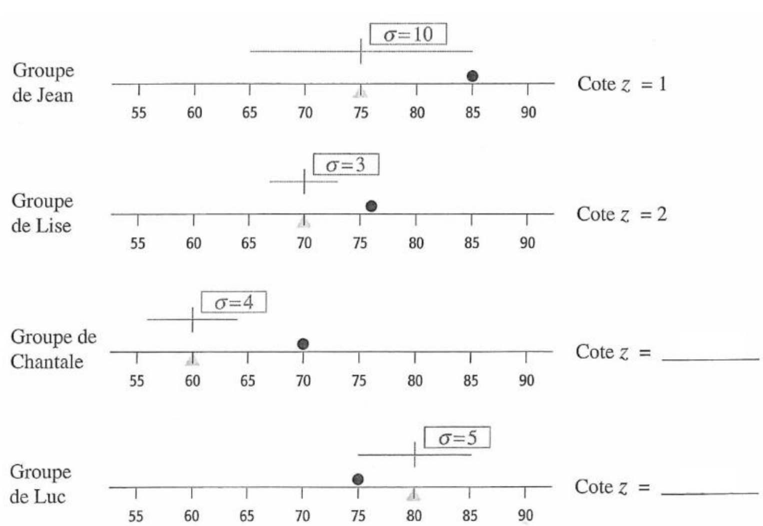
a) Si l'employeur se fie à la note seulement, quel devrait être son choix?

b) S'il considère en plus la moyenne du groupe?

c) Et s'il tient compte de la note, de la moyenne et de l'écart type, qui devrait-il embaucher?

d) Est-ce le bon choix? Pourquoi?

Pour mieux comprendre, représentons sur les diagrammes suivants la moyenne du groupe par un pivot et l'écart type par un segment de droite. La note de l'étudiant est indiquée par un point.



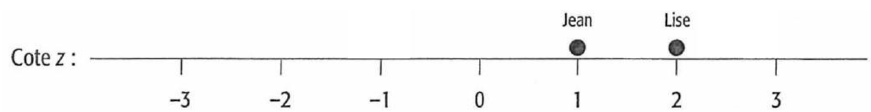
Position de chaque note par rapport aux autres notes du groupe :

- Comme la note de Jean se situe à 1 écart type au-dessus de son groupe, nous dirons que sa cote Z est 1.
- Comme la note de lise se situe à 2 écarts types au-dessus de la moyenne de son groupe, nous dirons que sa cote Z est 2.

La cote Z donne la mesure, en nombre d'écarts types, de l'écart entre une valeur et la moyenne.	$\text{cote Z} = \dfrac{\text{valeur} - \text{moyenne}}{\text{écart type}}$

e) Trouver la cote Z pour les notes de Chantale et de Luc.

f) La cote Z permet de comparer les notes des quatre étudiants en les ramenant sur une même échelle, celle des cotes Z. Comparez les cotes Z.



A. Valeurs possibles pour une cote Z

La cote Z est particulièrement utile pour comparer des résultats de nature différente. Il est important de savoir qu'une cote Z plus grande que 2 ou plus petite que -2 est assez rare, et qu'une cote Z plus grande que 3 ou plus petite que -3 est rare. C'est pour cette raison que les valeurs indiquées sur une échelle de cote Z sont généralement comprises entre -3 et +3. En fait, on a établi que, dans une série de données, <u>au maximum</u> :

12,5 % des données ont une cote $Z \geq 2$	12,5 % des données ont une cote $Z \leq -2$
8 % des données ont une cote $Z \geq 2{,}5$	8 % des données ont une cote $Z \leq -2{,}5$
5,5 % des données ont une cote $Z \geq 3$	5,5 % des données ont une cote $Z \leq -3$
4,1 % des données ont une cote $Z \geq 3{,}5$	4,1 % des données ont une cote $Z \leq -3{,}5$

Exemple 2: Chaque semaine, le géant de l'alimentation Métro publie une circulaire annonçant les soldes du jeudi pour toutes les épiceries. Le gérant de l'un de ces Métro décide un jour d'en faire un peu plus en plaçant une annonce dans le journal local. Le jeudi suivant la parution de l'annonce, il reçoit 2 280 clients alors qu'habituellement le jeudi la moyenne est de 2 000 clients avec un écart type de 80 clients. Peut-il en conclure que son annonce dans le journal local a eu un effet? Un écart de 280 clients par rapport à la moyenne est-il significatif?

Avec 2 280 clients, on obtient une cote Z de _________; cet écart de 280 clients par rapport à la moyenne est _________________________.

Conclusion :

B. Cote de rendement au collégial (Cote R)

La cote de rendement au collégial (CRC, ou « cote R » pour les intimes) est un instrument dont se servent les universités pour gérer les admissions dans les programmes contingentés. Cette cote combine, pour chaque cours suivi par un étudiant, trois informations : un indicateur de la position de cet étudiant en fonction de la note obtenue dans son groupe (la cote Z au collégial), un indicateur de la force de ce groupe et un indicateur de la dispersion de ce groupe.

i. Effet de l'utilisation de la cote Z au collégial

Dans le tableau 1, on retrouve les résultats des étudiants de trois classes pour un même cours. On suppose que tous ces étudiants veulent être admis dans un même programme d'études à l'université et que l'on a de la place que pour six d'entre eux.

TABLEAU 1
DISTRIBUTION DES NOTES

Classe A		Classe B		Classe C	
Nom de l'étudiant	**Note**	**Nom de l'étudiant**	**Note**	**Nom de l'étudiant**	**Note**
Florence	89	William	79	Annie	91
Olivier	88	Camille	78	Alexis	87
Jade	87	Vincent	77	Catherine	83
Étienne	86	Olivia	76	Émilie	79
Gabrielle	85	Francis	75	Rosalie	75
Guillaume	84	Antoine	74	Xavier	71
Marie	83	Emma	73	Jacob	67
Samuel	82	Nathan	72	Félix	63
Chloé	81	Audrey	71	Sarah	59
Sommes des notes	765		675		675
Nombre de notes	9		9		9
MOYENNE	85		75		75

► En considérant seulement les notes, qui sont les six « meilleurs » étudiants?

Si les différences observées dans les notes de ces trois groupes d'étudiants dépendent uniquement de la sévérité ou de la générosité de la notation des étudiants par leur professeur respectif, il s'ensuit que certains étudiants sont avantagés alors que d'autres sont pénalisés.

Cependant, si l'on tient compte de l'écart à la moyenne plutôt que de la note elle-même, il est possible d'éliminer des différences artificielles. Voici, dans le tableau 2, l'illustration d'une première correction apportée par la technique de la cote Z.

TABLEAU 2
ÉCARTS À LA MOYENNE

Classe A		Classe B		Classe C	
Nom de l'étudiant	**Écart**	**Nom de l'étudiant**	**Écart**	**Nom de l'étudiant**	**Écart**
Florence	4	William	4	Annie	16
Olivier	3	Camille	3	Alexis	12
Jade	2	Vincent	2	Catherine	8
Étienne	1	Olivia	1	Émilie	4
Gabrielle	0	Francis	0	Rosalie	0
Guillaume	-1	Antoine	-1	Xavier	-4
Marie	-2	Emma	-2	Jacob	-8
Samuel	-3	Nathan	-3	Félix	-12
Chloé	-4	Audrey	-4	Sarah	-16

► En considérant les écarts à la moyenne, qui sont les six « meilleurs » étudiants?

Mais en plus de tenir compte de l'écart à la moyenne, il faut aussi tenir compte de l'étalement (ou dispersion) des notes si l'on veut corriger les distorsions occasionnées par le système de notation. Pour prendre en compte l'étalement des notes, il faut calculer l'écart type des notes de chaque classe, donc :

1) élever au carré tous les écarts à la moyenne de cette classe,
2) faire la somme des écarts élevés au carré,
3) diviser cette somme par le nombre de notes,
4) extraire la racine carrée.

Le tableau 3 fournit les résultats du calcul.

TABLEAU 3
CARRÉ DES ÉCARTS À LA MOYENNE

Classe A		Classe B		Classe C	
Nom de l'étudiant	**Écart au carré**	**Nom de l'étudiant**	**Écart au carré**	**Nom de l'étudiant**	**Écart au carré**
Florence	16	William	16	Annie	256
Olivier	9	Camille	9	Alexis	144
Jade	4	Vincent	4	Catherine	64
Étienne	1	Olivia	1	Émilie	16
Gabrielle	0	Francis	0	Rosalie	0
Guillaume	1	Antoine	1	Xavier	16
Marie	4	Emma	4	Jacob	64
Samuel	9	Nathan	9	Félix	144
Chloé	16	Audrey	16	Sarah	256
Sommes des écarts²	60		60		960
Nombre de notes	9		9		9
MOYENNE	6,67		6,67		106,67
ÉCART TYPE	2,58		2,58		10,33

Avec l'ensemble de ces données, on peut finalement calculer la cote Z de chaque étudiant.

TABLEAU 4
COTES Z

Classe A		Classe B		Classe C	
Nom de l'étudiant	**Cote Z**	**Nom de l'étudiant**	**Cote Z**	**Nom de l'étudiant**	**Cote Z**
Florence	1,55	William	1,55	Annie	1,55
Olivier	1,16	Camille	1,16	Alexis	1,16
Jade	0,77	Vincent	0,77	Catherine	0,77
Étienne	0,39	Olivia	0,39	Émilie	0,39
Gabrielle	0,00	Francis	0,00	Rosalie	0,00
Guillaume	-0,39	Antoine	-0,39	Xavier	-0,39
Marie	-0,77	Emma	-0,77	Jacob	-0,77
Samuel	-1,16	Nathan	-1,16	Félix	-1,16
Chloé	-1,55	Audrey	-1,55	Sarah	-1,55

► En considérant les cotes Z, qui sont les six « meilleurs » étudiants?

Maintenant, peu importe la sévérité ou la générosité d'un professeur quant à sa façon d'attribuer des notes, celles-ci, une fois transformées en cotes Z, deviennent comparables d'une classe à l'autre. En somme, la cote Z permet de rétablir l'équité pour tous les étudiants.

ii. Effet de l'utilisation de deux indicateurs : force de groupe et dispersion de groupe

Les classements réalisés plus haut avec la cote Z sont équitables pour tous les étudiants concernés si et seulement si les classes comparées sont équivalentes (de même calibre). Or, il arrive fréquemment que les groupes ne soient pas directement comparables. Pour une même discipline, posons l'hypothèse que :

- la classe A est constituée uniquement d'étudiants « faibles »,
- la classe B est constituée uniquement d'étudiants « forts »,
- la classe C est constituée d'étudiants « forts », « moyens » et « faibles ».

Pour définir la force et la dispersion du groupe, on utilise les résultats obtenus aux épreuves ministérielles de 4^{e} et 5^{e} secondaire par tous les étudiants qui appartiennent à un même groupe au collège.

Afin de connaître la correction à apporter à la cote Z au collégial (Zcol), il faut appliquer la formule suivante :

Zcol corrigée = (Zcol × IDGZ) + IFGZ

où :

Zcol = cote Z au collégial

IDGZ = indicateur de dispersion de groupe correspondant à **l'écart type des cotes Z au secondaire** des étudiants qui composent le groupe au collégial (valeur minimale de +0,5 et valeur maximale de +1,5)

IFGZ = indicateur de force de groupe correspondant à **la moyenne des cotes Z au secondaire** des étudiants qui composent le groupe au collégial (valeur minimale de –2 et valeur maximale de +2)

Par convention, il a été établi que (Zcol × IDGZ) aura toujours une valeur comprise entre –3 et +3.

Puis pour obtenir la fameuse cote de rendement au collégial (ou cote R), il faut additionner 5 à Zcol corrigée et multiplier cette somme par 5 afin de situer les résultats sur une nouvelle échelle ayant une amplitude fixe se situant entre 0 et 50.

Cote R = (Zcol corrigée + 5) × 5

donc

Cote R = ((Zcol × IDGZ) + IFGZ + 5) × 5

À titre d'illustration, prenons les étudiants de la classe B. En supposant que l'IDGZ est de 0,60 et l'IFGZ est de 1,12 et en appliquant la formule, on obtient la cote Z au collégial corrigée, puis la cote R de chaque étudiant (voir tableau 5).

Il est à noter que, de manière générale, la plupart des cotes de rendement se situent entre 15 et 35.

TABLEAU 5
MÉTHODE DE CALCUL DE LA COTE DE RENDEMENT

Classe B						
Nom de l'étudiant	**Note**	**Zcol**	**Correction avec IDGZ**	**Correction avec IFGZ**	**Zcol corrigée**	**Cote de rendement**
William	79	1,55	0,60	1,12	2,05	35,25
Camille	78	1,16	0,60	1,12	1,82	34,08
Vincent	77	0,77	0,60	1,12	1,58	32,91
Olivia	76	0,39	0,60	1,12	1,35	31,77
Francis	75	0,00	0,60	1,12	1,12	30,60
Antoine	74	-0,39	0,60	1,12	0,89	29,43
Emma	73	-0,77	0,60	1,12	0,66	28,29
Nathan	72	-1,16	0,60	1,12	0,42	27,12
Audrey	71	-1,55	0,60	1,12	0,19	25,95

Si l'on pose maintenant l'hypothèse que, pour le groupe A, l'IDGZ est de 0,80 et l'IFGZ est de 0,55 et, pour le groupe C, l'IDGZ est de 1,00 et l'IFGZ est de 0,72 et qu'on effectue les mêmes calculs qu'au tableau 5, on obtient au tableau 6 la cote R de chacun des étudiants des trois groupes.

Tableau 6
Cotes de rendement au collégial

Classe A		Classe B		Classe C	
Nom de l'étudiant	**Cote**	**Nom de l'étudiant**	**Cote**	**Nom de l'étudiant**	**Cote**
Florence	33,95	William	35,25	Annie	36,35
Olivier	32,39	Camille	34,08	Alexis	34,40
Jade	30,83	Vincent	32,91	Catherine	32,45
Étienne	29,31	Olivia	31,77	Émilie	30,55
Gabrielle	27,75	Francis	30,60	Rosalie	28,60
Guillaume	26,19	Antoine	29,43	Xavier	26,65
Marie	24,67	Emma	28,29	Jacob	24,75
Samuel	23,11	Nathan	27,12	Félix	22,80
Chloé	21,55	Audrey	25,95	Sarah	20,85

► En considérant les cotes R, qui sont les six « meilleurs » étudiants?

Ainsi, l'appartenance à un groupe plus faible (groupe A) ne procure aucun avantage alors que l'appartenance à un groupe plus homogène ou plus fort (groupe B) ne pénalise aucunement les meilleurs étudiants de ce groupe.

iii. Conclusion

La cote R permet un ajustement adéquat à la situation de chacun des étudiants. Autrement dit, qu'un étudiant change de collège, de programme d'études ou de groupe, la cote Z au collégial obtenue à chaque cours est de ce fait corrigée par deux indicateurs (IFGZ et IDGZ) du groupe dans lequel l'étudiant a été évalué.

La cote R est, en définitive, un instrument de classement juste et équitable. Elle permet de s'assurer que le dossier scolaire des diplômés du collégial faisant une demande d'admission à l'université soit évalué le plus équitablement possible, peu importe le collège d'origine. Elle donne ainsi aux meilleurs étudiants de tous les collèges des chances égales d'accès aux programmes universitaires les plus contingentés.

Section 5B : Informations tirées de « *La cote de rendement au collégial : ce qu'elle est, ce qu'elle fait* », Bureau de coopération interuniversitaire, 12 juillet 2017. Pour en savoir plus sur la cote R, consultez ce document en ligne. Tous les détails du calcul y sont précisés.

Exercices

La première partie des exercices est tirée du livre « *Méthodes quantitatives – 4e édition »* de Christiane Simard, paru en 2008 aux éditions Modulo.
La deuxième partie des exercices est tirée du livre « *Introduction à la statistique appliquée* » de ALALOUF, LABELLE, MÉNARD, paru en 2002 aux éditions Loze-Dion

PREMIÈRE PARTIE

1. Un sondage auprès d'un échantillon de 115 membres d'un club vidéo a permis de construire le tableau suivant :

Répartition des répondants par sexe, selon le nombre de films loués au cours du dernier mois

Nombre de films loués en un mois	Nombre d'hommes	Nombre de femmes
4	18	8
5	12	16
6	11	10
7	9	10
8 et plus	10	11
Total	60	55

a) Déterminer et interpréter la médiane de la distribution pour les hommes.

b) Déterminer et interpréter la médiane de la distribution pour les femmes.

c) Donner et interpréter le mode pour la distribution des hommes et des femmes.

2. Voici un tableau représentant la répartition des immigrants accueillis au Québec en 2006, selon leur âge.

Répartition des immigrants accueillis au Québec en 2006, selon leur âge

Âge	Pourcentage
Moins de 15 ans	20,0 %
[15 ans ; 30 ans[	32,0 %
[30 ans ; 45 ans[	38,0 %
45 ans et plus	10,0 %
Total	100,0 %

Source : Institut de la statistique du Québec.

a) En 2006, l'âge médian des Québécois était estimé à 41 ans. Qu'en est-il de l'âge médian des immigrants que le Québec a accueillis cette année-là?

b) Interpréter la médiane trouvée en a).

c) Calculer et interpréter la moyenne de cette distribution.

3. Pour les séries statistiques suivantes, donner et interpréter chacune des mesures de tendance centrale, puis indiquer laquelle serait la plus représentative de la série statistique.

a) Le nombre de calendriers vendus en une journée par sept personnes.

7 8 6 9 6 36 10

b) Le sexe des sept enfants d'une même famille en partant du plus âgé jusqu'au plus jeune.

Féminin Féminin Masculin Féminin Masculin Féminin Féminin

c) Le nombre de spectateurs à chacune des six représentations d'une pièce de théâtre.

724 802 715 825 650 790

4. Le tableau suivant est tiré d'un sondage publié dans la revue *L'actualité* en 2007.

Répartition des répondants selon le nombre de films vus au cours du dernier mois, Québec, 2007

Nombre de films vus au cours du dernier mois	0	1	2	3 et plus
Nombre de répondants	378	102	66	54

Source : *L'actualité*, 1er mai 2007, sondage Crop – L'actualité – Radio-Canada.

Pour les calculs, remplacer la catégorie « 3 et plus » par le nombre 3.

a) Donner le nom, le type et les valeurs de la variable.

b) Combien de personnes ont répondu à cette question du sondage?

c) Trouver et interpréter la moyenne, le mode et la médiane de cette distribution.

d) Trouver et interpréter le 3e quartile.

5. Selon Statistique Canada, le revenu personnel médian des Québécois en 2005 était de 23 100 \$, alors que le revenu personnel moyen était de 30 100 \$. Qu'est-ce qui peut expliquer un tel écart entre ces deux mesures? Quelle mesure est-il préférable d'utiliser pour cette variable?

6. À l'aide de quelques mesures statistiques, nous allons comparer les distributions de l'âge de la mère à la naissance de son enfant pour les années 1980 et 2006.

Répartition des naissances selon l'âge de la mère, Québec, 1980 et 2006

Âge de la mère (en ans)	Pourcentage des naissances	
	1980	2006
Moins de 20	5,2 %	2,7 %
[20 ; 25[	30,2 %	15,4 %
[25 ; 30[	40,5 %	35,9 %
[30 ; 35[	19,2 %	31,0 %
[35 ; 40[	4,4 %	12,6 %
40 ans et plus	0,5 %	2,4 %
Total	100,0 %	100,0 %

Source : Institut de la statistique du Québec, 2007.

a) Compléter.
On peut estimer que la moyenne d'âge des mères qui ont donné naissance à un enfant était de 26,9 ans en 1980 et de _______ ans en 2006, une augmentation de _____ ans en 26 ans.

b) Compléter.
L'âge médian des mères qui ont donné naissance à un enfant peut être estimé à 26,8 ans en 1980 et à ________ ans en 2006, une augmentation de _______ ans en 26 ans.

c) Calculer et interpréter le premier décile de l'âge des mères qui ont donné naissance en 2006.

d) Le quatrième quintile de la distribution de l'âge des mères pour l'année 2006 est 34,2 ans. Interpréter cette mesure.

7. Quel âge ont les femmes qui subissent un avortement? Observez le tableau suivant :

Répartition des avortements thérapeutiques selon l'âge de la femme, Québec, 2004

Âge de la femme	Pourcentage
Moins de 20 ans	17,4 %
[20 ans ; 25 ans[	31,0 %
[25 ans ; 30 ans[	22,2 %
[30 ans ; 35 ans[	15,2 %
35 ans et plus	14,2 %
Total	100,0 %

Source : Statistique Canada.

a) Calculer et interpréter la moyenne de la distribution.

b) Donner et interpréter la médiane de la distribution.

c) Donner et interpréter la classe modale de la distribution.

d) Calculer et interpréter le premier quartile (Q_1) de la distribution.

8. On vous demande de calculer la moyenne des cours suivants en tenant compte du nombre de crédits pour chaque cours. Quel nom donne-t-on à ce type de moyenne?

Cours	Mathé-matiques	Histoire	Géo-graphie	Français
Notes	60 %	70 %	65 %	80 %
Nombre de crédits	3	2	2	3

9. Dans une entreprise comptant 3 départements, les 10 employés du département A gagnent en moyenne 35 000 $, les 25 employés du département B, 30 000 $ et les 5 employés du département C, 28 000 $. Quel est le salaire annuel moyen de l'ensemble des employés de cette entreprise?

10. À partir de l'enquête sur les dépenses des ménages, on a construit le tableau suivant :

Répartition des ménages propriétaires de leur logement selon la taille du ménage, Québec, 2005

Taille du ménage	Nombre
1 personne	302 036
2 personnes	722 260
3 personnes	317 044
4 personnes	371 448
5 personnes et plus	163 212
Total	1 876 000

Source : Institut de la statistique du Québec, Données sociodémographiques en bref, octobre 2007.

a) Que représente le total?

b) Donner le nom et le type de variable étudié.

c) En 2005, quel pourcentage des ménages propriétaires de leur logement sont composés de 4 personnes?

d) Donner et interpréter le mode de cette distribution.

e) Donner et interpréter la médiane de cette distribution.

f) Calculer et interpréter la moyenne en remplaçant la catégorie « 5 et plus » par « 5 ».

11. Le poids moyen des personnes du groupe A est de 66,3 kg et celui du groupe B est de 60,2 kg. Supposons que l'on réunisse les deux groupes. Peut-on, dans les conditions suivantes, calculer le poids moyen du nouveau groupe? Si oui, donner la moyenne.

a) Les deux groupes contiennent le même nombre de personnes.

b) On ne connaît pas le nombre de personnes de chaque groupe.

c) Le groupe A comprend 10 personnes et le groupe B, 40 personnes.

12. Combien de temps les Québécois consacrent-ils pour se rendre au travail? Le tableau suivant répond à cette question.

Répartition des travailleurs selon la durée de l'aller-retour entre la maison et le travail, Québec, 2005

Durée (en min)	[0; 30[	[30; 60[	[60; 90[	90 et plus	Total
Pourcentage	23 %	30 %	21 %	26 %	100 %

Source : Statistique Canada, Enquête sociale générale sur l'emploi du temps.

a) Calculer et interpréter la moyenne.

b) Donner et interpréter la classe modale.

c) Trouver la médiane et interpréter le résultat.

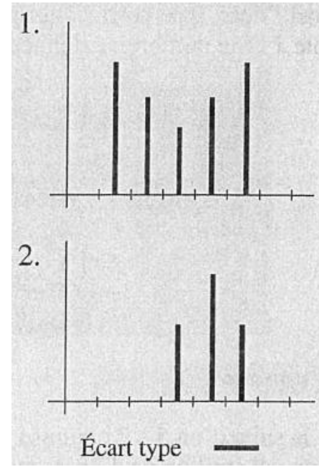
d) Indiquer et interpréter le 1er décile (D_1).

e) Indiquer et interpréter le 1er quartile (Q_1).

f) 30 minutes correspond à quel centile? Interpréter ce centile.

g) Compléter : $Md = C_? = D_? = Q_?$.

13. Sans faire de calculs, indiquer le numéro du diagramme en bâtons ci-dessous.



a) qui a la plus petite moyenne :

b) qui a le plus petit écart type :

c) dont la valeur de l'écart type est égale à la longueur du petit segment de droite qui figure sous le graphique 2 :

14.

a) Une série A représente l'âge de cinq membres d'une famille et une série B l'âge des étudiants d'une classe de cégep. Laquelle des deux séries aura le plus grand écart type?

b) Un professeur fait passer un test de classement en mathématiques dans ses deux groupes. Les deux groupes ont obtenu la même moyenne, mais l'écart type du groupe A est plus grand que celui du groupe B. Dans quel groupe peut-on dire que les étudiants sont à peu près tous au même niveau en mathématique?

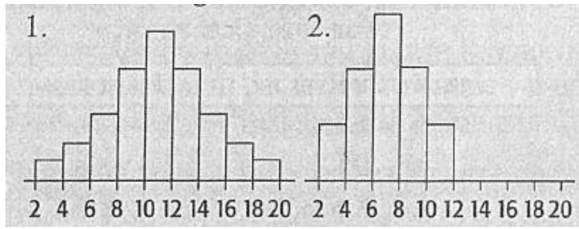
c) Dans une région aride du globe, on enregistre la précipitation quotidienne (en mm) durant 60 jours consécutifs. La moyenne des 60 données est égale à 0. Que vaut l'écart type?

15. Vrai ou faux? Toutes les données d'une distribution dont la moyenne est 70 et l'écart type 10 sont comprises entre 60 et 80?

16. Nous avons vu, dans la mise en situation de la section 4B (variance et écart type), que la moyenne d'âge de l'équipage du voilier *Le Moussaillon* était de 20 ans avec un écart type de 6,6 ans. Si l'équipage est toujours le même dans 6 ans, quels seront alors la moyenne et l'écart type de l'âge de l'équipage?

17.

a) Estimer graphiquement la moyenne de chacun des histogrammes suivants :



b) Lequel de ces histogrammes aura le plus grand écart type?

c) Utiliser un rapport de surfaces pour trouver le pourcentage de données qu'il y a dans la classe [6, 8[de l'histogramme 2.

18. Combien d'années de scolarité ont les professeurs qui vous enseignent? Les statistiques suivantes vous en donnent une idée.

Répartition des professeurs[1] de cégep selon la scolarité, Québec, 2003-2004

Scolarité (en années)	17	18	19	20	Total
Nombre de professeurs	2 577	1 935	4 071	686	9 269

1. Ces données ne s'appliquent qu'aux professeurs permanents.

Source : Ministère de l'Éducation, Système d'information sur le personnel des organismes collégiaux.

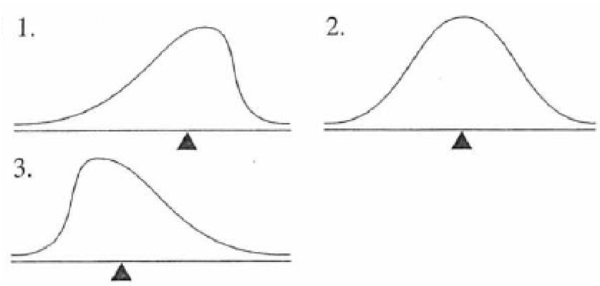
a) Quel est le pourcentage de professeurs de cégep ayant au moins 19 ans de scolarité?

b) Calculer la moyenne et l'écart type de la distribution et interpréter ces deux mesures.

c) Donner le mode et la médiane de la distribution et interpréter ces deux mesures.

d) Donner et interpréter le 3^e décile.

19. Voici trois séries statistiques et leur moyenne représentée par un pivot.



Laquelle des affirmations suivantes est vraie? L'écart type du graphique 1 est :

a) plus petit que celui du graphique 2.
b) égal à celui du graphique 3.
c) plus grand que celui du graphique 3.

20. Soit A, la série des 365 températures quotidiennes à Montréal en 2007, et B, celle des 365 températures quotidiennes à Miami la même année. D'après vous, laquelle des deux séries présente le plus grand écart type?

21. Dans chacun des cas suivants, indiquer si la mesure de tendance centrale donnée est possible compte tenu des restrictions imposées. Si non, expliquer pourquoi. Si oui, donner un exemple de résultats satisfaisant à toutes ces conditions. La représentation de la situation à l'aide d'un pictogramme facilite la recherche de solution.

a) Il y a cinq données ; la plus petite donnée de la série est 4; l'étendue est 10; $\mu = 14$.

b) Il y a cinq données : la plus petite donnée de la série est 4; l'étendue est 10; la médiane égale 14.

c) Il y a cinq données : la plus petite donnée de la série est 50; la plus grande est 100; $\mu = 55$.

22. Pour un échantillon de 10 années (de 1998 à 2007), on a enregistré la température moyenne en octobre, à Québec. Voici les températures (en °C) de l'échantillon :

7,2   5,5   6,6   7,3   4,8   6,5   8,0   8,5   6,7   7,0

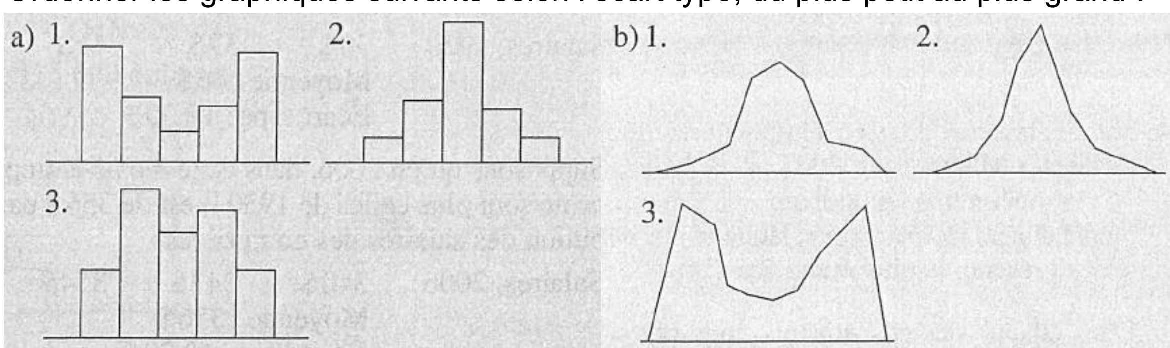
a) Quelle est l'étendue des observations?

b) Calculer et interpréter la moyenne et l'écart type corrigé de l'échantillon.

c) Indiquer et interpréter le mode et la médiane.

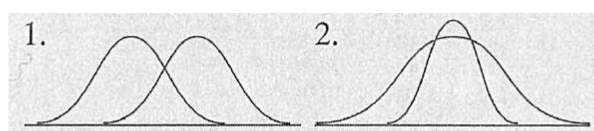
d) Indiquer et interpréter le 3^e quartile.

23. Ordonner les graphiques suivants selon l'écart type, du plus petit au plus grand :



24.

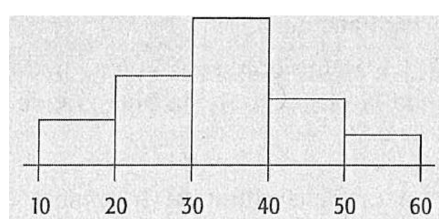
a) Parmi les deux graphiques suivants :



- **i.** Lequel présente deux distributions ayant la même moyenne?

- **ii.** Lequel présente deux distributions ayant le même écart type?

b) Quels nombres, parmi $\boxed{11,2}$, $\boxed{28,1}$ et $\boxed{1,2}$, ne peuvent être l'écart type de l'histogramme suivant?



25. Il y a une vingtaine d'années, c'était surtout des jeunes qui achetaient des motocyclettes. Aujourd'hui, en raison des coûts d'achat et d'assurances de ces véhicules, les choses ont changé comme en témoignent les statistiques suivantes :

Répartition des propriétaires d'une motocyclette selon l'âge, Québec, 2005

Âge du propriétaire	Pourcentage
Moins de 25 ans	3,8 %
[25 ans ; 35 ans[	15,9 %
[35 ans ; 45 ans[	29,6 %
[45 ans ; 55 ans[	34,5 %
55 ans et plus	16,2 %
Total	100,0 %

Source : Société de l'assurance automobile du Québec, Bilan 2006.

a) Indiquer et interpréter la classe modale.

b) Estimer l'âge moyen des propriétaires d'une motocyclette en 2005.

c) Calculer et interpréter l'écart type de la distribution.

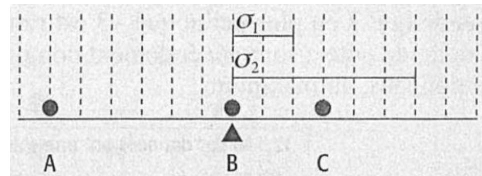
d) Les résultats obtenus en b) et c) sont-ils les valeurs exactes ou approximatives des mesures demandées?

e) Compléter. En 2005, au Québec, 50 % des propriétaires d'une motocyclette avaient moins de ______ ans, alors que 10 % des propriétaires avaient moins de _______ ans.

26. En considérant les deux écarts types suggérés, trouver la cote Z de chacun des points du pictogramme.

Avec l'écart type σ_1
Cote Z de A = ______
Cote Z de B = ______
Cote Z de C = ______

Avec l'écart type σ_2
Cote Z de A = ______
Cote Z de B = ______
Cote Z de C = ______



27. Vrai ou faux? Si votre cote Z à un examen est 2, alors :

a) vous avez deux points au-dessus de la moyenne.

b) vous avez une note égale à deux fois l'écart type du groupe.

c) vous avez une note se situant à deux écarts types au-dessus de la moyenne.

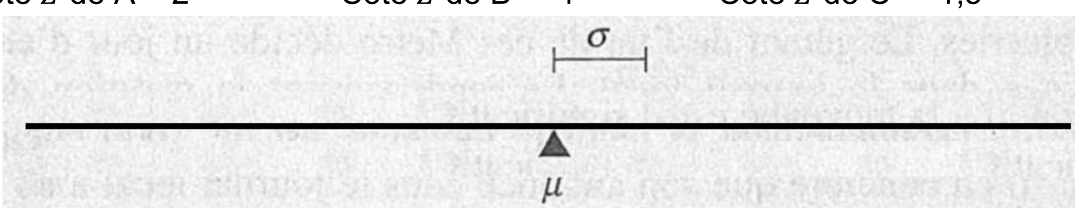
d) votre note est égale à la moyenne de l'examen plus deux fois l'écart type.

28. Un cote Z de 2 à un examen peut être considérée comme faible, moyenne, très bonne ou exceptionnelle?

29.

a) En considérant les cotes Z des points A, B et C, situer ces points sur le pictogramme.

Cote Z de A = 2 Cote Z de B = -1 Cote Z de C = -1,5



b) Quel est l'écart entre chaque point et la moyenne si l'écart type est de 10?

Écart entre A et μ = ______

Écart entre B et μ = ______

Écart entre C et μ = ______

c) Considérant les écarts ci-dessus, donner la valeur de A, B et C si la moyenne est 50.

30. Comme nous l'avons mentionné, la cote Z permet de comparer des valeurs même si celles-ci proviennent de domaines bien différents. En voici un exemple.

On veut trouver le meilleur vendeur du mois. Cet honneur sera décerné à la personne s'étant le plus distinguée dans son domaine. Trois concurrents sont en lice ; voici la description de la performance de chacun, en un mois :

- Lise a vendu 85 barres de chocolat pour les activités sportives de son école, alors que la moyenne de ventes a été de 52 barres par étudiant avec un écart type de 13 barres.
- Paul a vendu 25 préarrangements funéraires, alors que la moyenne de ventes est de 12 préarrangements funéraires avec un écart type de 6.
- Lucie a vendu 75 abonnements au *Journal de Québec*, alors que la moyenne de ventes est de 47 abonnements avec un écart type de 10.

Qui sera déclaré « vendeur du mois »? Justifier la réponse.

31. Une entreprise a retenu les services d'un laboratoire spécialisé en contrôle de la qualité afin qu'il évalue la qualité d'un mélange bitumineux provenant des deux usines qui ont présenté une soumission en réponse à son appel d'offre. Il a été convenu de procéder au contrôle de la qualité en prélevant dans la production de chaque usine un échantillon de 50 cylindres de béton et que, sur chacun des cylindres, serait mesurée la résistance à la compression. Les résultats apparaissent dans les tableaux suivants :

Usine A

Résistance à la compression (en kg/cm²)	Nombre de cylindres
[70 ; 75[	2
[75 ; 80[	4
[80 ; 85[	7
[85 ; 90[	12
[90 ; 95[	11
[95 ; 100[	11
[100 ; 105[	3
Total	50

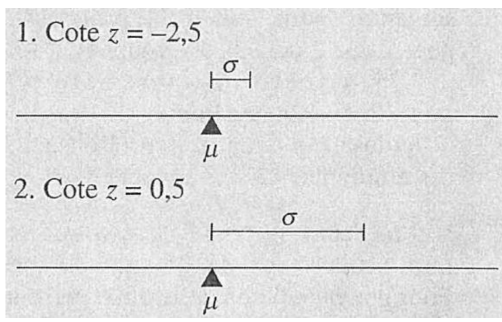
Usine B

Résistance à la compression (en kg/cm²)	Nombre de cylindres
[70 ; 80[	4
[80 ; 90[	7
[90 ; 100[	19
[100 ; 110[	12
[110 ; 120[	6
[120 ; 130[	2
Total	50

Calculer la moyenne et l'écart type corrigé de l'échantillon de chaque usine.

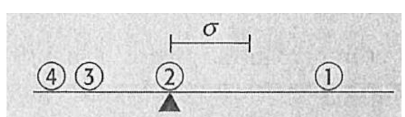
32.

a) Sur chacun des graphiques ci-dessous, placer un point correspondant à la cote Z indiquée.



b) Dans chaque cas, donner l'écart entre le point et la moyenne si l'écart type est de 8 dans le graphique 1 et de 20 dans le graphique 2.

33. On a représenté ci-dessous la note à un examen (sur 100) de quatre étudiants d'un groupe.



a) Quelle cote Z chacun de ces étudiants a-t-il à son examen?

b) Quel est l'écart entre chaque note et la moyenne du groupe si l'écart type est de 10 %?

c) Donner la note obtenue dans chaque cas si la moyenne de l'examen est de 65 %.

34. Deux étudiantes de groupes différents ont eu la même note à un examen, pourtant la cote Z de Lise est plus grande que celle de Marie. Elles ont toutes deux une note au-dessus de la moyenne.

a) Si la moyenne des examens est la même dans les deux groupes, dans quel groupe l'écart type de l'examen est-il le plus petit?

b) Si l'écart type des examens est le même dans les deux groupes, dans quel groupe la moyenne de l'examen est-elle la plus faible?

<u>DEUXIÈME PARTIE</u>

35. Calculer la moyenne et la médiane des données suivantes. Déterminer aussi le mode, s'il existe.

a) 2 2 3 3 3 4 4 4 4 4 4 5 5 6 7 8

b) 7,1 8,2 9,4 11,2 14,5 18,3 12,5

c) 2,8 2,7 3,9 4,7 2,8 1,9 7,8 8,4

36. Calculer la variance et l'écart-type de chacune des séries données au numéro 35.

37. Laquelle des deux séries suivantes vous semble la plus dispersée? Confirmer votre réponse en calculant l'écart type de chacune.

A : 40 49 50 51 60 B : 48 49 50 51 52

38. Déterminer la moyenne arithmétique, le mode, la médiane, la variance et l'écart type de la distribution suivante.

Valeur	1	2	3	4	5	TOTAL
Fréquence	0,1	0,4	0,2	0,2	0,1	1

39. Votre note est de 68 dans une classe où la moyenne est de 54 et l'écart type est de 14. Quelle est votre cote Z?

40. Calculer la cote Z de chaque membre de la série 5 7 8 9 11, puis calculez la moyenne et la variance des 5 cotes Z.

41. Un médecin vous dit que votre pression intra-occulaire est de 23. Pour une population de 100 000 personnes de votre âge, la pression moyenne est de 16,75 avec un écart type de 2,5. Combien, au maximum, y a-t-il de personnes dans la population qui ont une pression au moins aussi éloignée de la moyenne que la vôtre?

42. Deux cent cinquante étudiants répartis en six groupes ont suivi un cours de statistique. Le nombre d'étudiants et la note moyenne de chaque groupe sont indiqués dans le tableau suivant :

Groupe	Nombre d'étudiants	Moyenne du groupe
1	47	63
2	38	61
3	30	68
4	55	54
5	40	72
6	40	73

Calculez la moyenne des 250 étudiants.

43. Calculer l'écart type des températures en janvier et l'écart type des températures en juillet à Montréal à partir des données suivantes :

Températures moyennes à Montréal - janvier et juillet, 1965-1976 (en °Celsius)

Année	Janvier	Juillet
1965	–10,0	20,0
1966	–9,4	21,7
1967	–5,6	22,2
1968	–12,2	22,2
1969	–7,2	21,7
1970	–13,3	23,3
1971	–11,1	21,7
1972	–6,7	21,7
1973	–6,1	21,7
1974	–9,8	23,3
1975	–6,3	21,0
1976	–11,9	23,9

SOURCE : *Annuaire du Québec*, 1971, 1980

44. Voici la distribution du nombre de familles par logement pour la région métropolitaine de Montréal en 1981 :

Nombre de familles	Effectif
0	293 390
1	724 975
2 et plus	8 560
TOTAL	1 026 925

a) Quel est le mode de cette variable? Interpréter.

b) Quelle est la médiane? Interpréter.

45. Deux supermarchés, A et B, reçoivent en moyenne le même nombre de clients par jour. Cependant, l'écart type est beaucoup plus élevé au supermarché A. D'après vous, lequel des deux supermarchés aura des dépenses en personnel plus élevées? Justifier la réponse.

46. Un test de dextérité manuelle donne un score moyen de 60 pour la population. Un score de 65 est donc supérieur à la moyenne. Dans lequel des deux cas suivants un score de 65 est-il plus spectaculaire?

a) L'écart type de la population est égal à 1.
b) L'écart type de la population est égal à 20.

47. Deux étudiants terminent un cours de comptabilité. L'étudiant A, qui a suivi le cours avec le chargé de cours X a obtenu la note de 69 à l'examen final; l'étudiant B, avec le chargé de cours Y, a obtenu la note 75. Pour le groupe X, la moyenne est de 60 et l'écart type de 3; pour le groupe Y, la moyenne est également 60 et l'écart type de 10. Lequel des deux étudiants s'est davantage démarqué de son groupe?

48. Un marchand se plaint à la ville du fait que certains travaux effectués par la municipalité ont causé une diminution de la circulation sur la rue du marchand et donc une baisse dans ses recettes. Pour appuyer sa plainte, il signale que ses recettes sont en moyenne de 20 000 \$ par jour, et que le jour des travaux, elles n'étaient que de 19 500 \$. La ville réplique qu'un écart de 500 \$, pour des recettes moyennes de 20 000 \$, est trop petit et donc ne démontre rien. Le marchand calcule alors l'écart type de ses recettes quotidiennes. Il trouve $\sigma = 100\ \$$. Qui a raison?

Réponses des exercices

1. a) Md = 5,5 films Interprétation : 50 % des hommes ont loué 5 films ou moins au cours du dernier mois.
   b) Md = 6 films Interprétation : Au moins 50 % des femmes ont loué 6 films ou moins au cours du dernier mois.
   c) $Mo_{(hommes)}$ = 4 films $Mo_{(femmes)}$ = 5 films
   Interprétation : Au cours du dernier mois, une pluralité d'hommes (30 %) ont loué 4 films et une pluralité de femmes (29 %) en ont loué 5.
2. a) Md ≈ 29,1 ans
   b) En 2006, on peut estimer que 50 % des immigrants accueillis par le Québec avaient moins de 29,1 ans.
   c) $\mu \approx 28{,}2\ ans$ Interprétation : En 2006, si tous les immigrants accueillis par le Québec avaient eu le même âge, ils auraient eu 28,2 ans.
3. a) $\mu \approx 11{,}7$ calendriers Interprétation : Si les sept personnes avaient vendu le même nombre de calendriers en une journée, ils en auraient vendu 11,7 chacun.
   Mo = 6 calendriers Interprétation: Une pluralité de personnes (28,6 %) ont vendu 6 calendriers en une journée.
   Md = 8 calendriers Interprétation : Au moins 50 % des personnes ont vendu 8 calendriers ou moins en une journée.
   La médiane est la meilleure mesure de tendance centrale pour cette série, car la moyenne est faussée par une donnée beaucoup plus grande que les autres et l'effectif du mode ne se démarque pas assez pour être significatif.
   b) La seule mesure de tendance centrale possible est le mode.
   Mo = féminin Interprétation : Une majorité d'enfants de la famille (71,4 %) sont des filles.
   c) La moyenne et la médiane sont toutes deux assez représentatives.
   $\mu = 751$ spectateurs Interprétation : S'il y avait eu le même nombre de spectateurs à chacun des spectacles, il y aurait eu 751 spectateurs.
   Mo = aucun
   Md = 757 spectateurs Interprétation : La moitié des représentations ont été jouées devant moins de 757 spectateurs.
4. a) Nom : nombre de films vus dans le dernier mois Type : quantitatif discret Valeurs : 0, 1, 2, 3 et plus
   b) 600 personnes
   c) $\mu \approx 0{,}7$ (0,66) Interprétation : Si tous les répondants avaient vu le même nombre de films, ils auraient vu 0,7 film chacun.
   Mo = 0 film Interprétation : Une majorité de répondants (63 %) n'ont vu aucun film dans le dernier mois.
   Md = 0 film Interprétation : Au moins 50 % des répondants n'ont vu aucun film dans le dernier mois.
   d) $Q_3 = 1$ film Interprétation : Au moins 75 % des répondants ont vu un film ou moins dans le dernier mois.
5. Il y a des Québécois qui ont un revenu personnel nettement plus élevé que les autres, ce qui fait augmenter la moyenne et la rend moins représentative des données. Pour les revenus des contribuables, la médiane est toujours la meilleure mesure de tendance centrale à utiliser.
6. a) 29,6 ans; 2,7 ans
   b) 29,4 ans; 2,6 ans
   c) $D_1 \approx 22{,}4\ ans$ Interprétation : On peut estimer que 10 % des femmes qui ont donné naissance en 2006 avaient moins de 22,4 ans.
   d) On peut estimer que 80 % des femmes ayant donné naissance en 2006 avaient moins de 34,2 ans.
7. a) $\mu \approx 26{,}4\ ans$ Interprétation : Si toutes les femmes subissant un avortement au Québec en 2004 avaient eu le même âge, on peut estimer qu'elles auraient eu 26,4 ans.
   b) Md ≈ 25,4 ans Interprétation : On peut estimer que 50 % des Québécoises qui ont subi un avortement thérapeutique en 2004 avaient moins de 25,4 ans.

c) Classe modale = [20, 25[Interprétation : Une pluralité de Québécoises (31 %) qui ont subi un avortement thérapeutique en 2004 avaient entre 20 et 25 ans.

d) $Q_1 \approx 21{,}2\ ans$ Interprétation : On peut estimer que 25 % des Québécoises qui ont subi un avortement thérapeutique en 2004 avaient moins de 21,2 ans.

8. $\mu = \frac{\Sigma x_i n_i}{N} = 69\ \%$ Il s'agit de la *moyenne pondérée*.

9. $\mu = \frac{\Sigma x_i n_i}{N} = 31\ 000\ \$$

10. a) Le nombre total de ménages propriétaires de leur logement au Québec en 2005.

b) Nom : Taille du ménage Type : quantitative discrète

c) 19,8 %

d) Mo = 2 Interprétation : Au Québec, en 2005, une pluralité de ménages propriétaires de leur logement (39 %) étaient composés de 2 personnes.

e) Md = 2 Interprétation : Au moins 50 % des ménages propriétaires de leur logement étaient composés de 2 personnes ou moins au Québec en 2005.

f) $\mu = \frac{\Sigma x_i n_i}{N} \approx 2{,}7$ Interprétation : Si tous les ménages propriétaires de leurs logements avaient été composés du même nombre de personnes au Québec en 2005, on peut estimer qu'ils auraient été 2,7 personnes par ménage.

11. a) $\mu = 66{,}3 \times 0{,}5 + 60{,}2 \times 0{,}5 \approx 63{,}3\ kg$

b) Non, on ne connaît pas le poids (la pondération) de chacun des sous-groupes.

c) $\mu = 66{,}3 \times \frac{10}{50} + 60{,}2 \times \frac{40}{50} \approx 61{,}4\ kg$

12. a) $\mu = 60\ min$ Interprétation : Si tous les travailleurs prenaient le même temps pour faire l'aller-retour entre la maison et le travail, on peut estimer qu'ils prendraient 60 minutes chacun.

b) [30; 60[Interprétation : Une pluralité de travailleurs québécois (30%) mettaient entre 30 et 60 minutes pour faire l'aller-retour entre la maison et le travail en 2005.

c) Md = 57 min Interprétation : On peut estimer que 50 % des travailleurs québécois mettaient moins de 57 minutes pour faire l'aller-retour entre la maison et le travail en 2005.

d) $D_1 \approx 13\ min$ Interprétation : On peut estimer que 10 % des travailleurs québécois mettaient moins de 13 minutes pour faire l'aller-retour entre la maison et le travail en 2005.

e) $Q_1 = 32\ min$ Interprétation : On peut estimer que 25 % des travailleurs québécois mettaient moins de 32 minutes pour faire l'aller-retour entre la maison et le travail en 2005.

f) Au 23^{e} centile (c'est le rang centile de 30). Interprétation : En 2005, 23 % de travailleurs québécois mettaient moins de 30 minutes pour faire l'aller-retour entre la maison et le travail.

g) $Md = C_{50} = D_5 = Q_2$

13. a) 1 b) 2 c) 1

14. a) série A b) groupe B c) $\sigma = 0$

15. Faux, ce ne sont pas <u>toutes</u> les données, mais <u>la plupart</u> des données.

16. $\mu = 26\ ans$ et $\sigma = 6{,}6\ ans$

17. a) $\mu_1 = 11$ et $\mu_2 = 7$ b) histogramme 1 c) surface du rectangle ÷ surface de l'histogramme ≈ 3/9 ≈ 33%

18. a) 51,3 % (attention, « au moins » signifie 19 années ou plus, donc 19 ou 20 années dans ce cas)

b) $\mu \approx 18{,}3$ années et $\sigma \approx 1$ (0,96) année

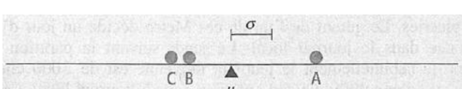
Interprétation : En 2003-2004, si les professeurs de cégep avaient eu le même nombre d'années de scolarité, ils auraient eu 18,3 années chacun. La plupart d'entre eux avaient 18 ou 19 années de scolarité (18 et 19 sont les entiers compris dans l'intervalle [17,3 ; 19,3]).

c) Mo = 19 années et Md = 19 années (c'est la valeur de la 4 635^{e} donnée).

Interprétation : En 2003-2004, une pluralité de professeurs de cégep (44 %) avaient 19 années de scolarité. Au moins 50 % des professeurs de cégep avaient 19 années de scolarité ou moins en 2003-2004.

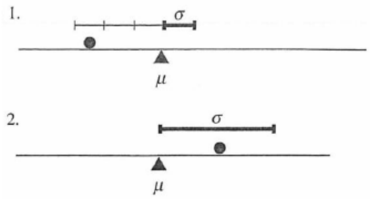
d) D_3 = 18 années (c'est la valeur de la 2 781^e donnée).
Interprétation : Au moins 30 % des professeurs de cégep avaient 18 années de scolarité ou moins en 2003-2004.

19. L'affirmation b) est vraie.
20. La série des 365 températures quotidiennes à Montréal.
21. a) Non, μ ne peut pas être 14, car le plus grand résultat (d'après l'étendue) sera 14 et les trois autres seront compris entre 4 et 14 inclus, donc 14 ne peut être le centre d'équilibre de la série.
    b) Oui, par exemple : 4 6 14 14 14
    c) Non, on ne peut trouver une série qui donnerait une moyenne de 55 dans ces conditions. Si l'on veut que 55 soit le centre d'équilibre de la série, il faudrait trouver trois valeurs, entre 50 et 55 inclus, telles que la somme, en valeur absolue, de leurs écarts à la moyenne additionnée à l'écart de la plus petite valeur (50 – 55 = -5) soit égale à 45 (l'écart entre 55 et 100). Même en donnant la plus petite valeur possible à ces trois nombres, on aurait : 50 50 50 50 100 et $\mu = 60$.
22. a) $E = 3{,}7°C$
    b) $\overline{x} \approx 6{,}8°C$ et $s \approx 1{,}1°C$ Interprétation : De 1998 à 2007, à Québec, si la température moyenne en octobre avaient été la même pour chacune de ces années, elle aurait été de 6,8°C. Pour la plupart de ces années, la température moyenne en octobre se situait entre 5,7°C et 7,9°C.
    c) Mo = aucun et Md ≈ 6,9°C Interprétation :Pour 50 % des années considérées, la température moyenne en octobre était inférieure à 6,9°C.
    d) $Q_3 = 7{,}3°C$ Interprétation : Pour au moins 75 % des années considérées, la température moyenne en octobre était de 7,3°C ou moins.
23. a) 2 < 3 < 1 b) 2 < 1 < 3
24. a) i) Graphique 2, les courbes ont la même moyenne
    ii) Graphique 1, les courbes ont le même écart type
    b) $\boxed{1{,}2}$ est trop petit, car il n'y a pas environ les deux tiers de la surface de l'histogramme dans l'intervalle $[\mu - 1{,}2\,;\mu + 1{,}2]$. Ensuite, $\boxed{28{,}1}$ est trop grand, même qu'il est plus grand que le plus grand des écarts. Donc, par élimination, l'écart type de cette distribution est $\boxed{11{,}2}$.
25. a) [45 ans; 55 ans[. Interprétation : Au Québec, en 2005, une pluralité de propriétaires (35 %) de motocyclette ont entre 45 et 55 ans.
    b) $\mu \approx 44{,}3\ ans$
    c) $\sigma \approx 10{,}6\ ans$ Interprétation : On peut estimer que la plupart des propriétaires d'une motocyclette en 2005 avaient entre 33,7 ans et 54,9 ans.
    d) Approximatives, car les données sont groupées en classes.
    e) 45,2 ; 28,9
26. -3 -1
    0 0
    1,5 0,5
27. a) Faux b) Faux c) Vrai d) Vrai
28. Très bonne
29. a)



    b) 20 -10 -15
    c) A = 70 B = 40 C = 35
30. $Cote\ Z_{Lise} = 2{,}5$ $Cote\ Z_{Paul} = 2{,}2$ $Cote\ Z_{Lucie} = 2{,}8$ Lucie sera déclarée “vendeuse du mois”.
31. Usine A: $\overline{x} = 89{,}6$ et $s \approx 7{,}6$ Usine B : $\overline{x} = 98$ et $s \approx 12{,}2$

32. a)



b) 1. Écart entre le point et μ = -20   2. Écart entre le point et μ = 10

33. a) 1. Cote Z = 2   2. Cote Z = 0   3. Cote Z = -1   4. Cote Z = -1,5

b) 1. 20 %   2. 0 %   3. -10 %   4. -15 %

c) 1. 85 %   2. 65 %   3. 55 %   4. 50 %

34. a) Celui de Lise. La distance entre la moyenne et la note étant la même dans les deux cas pour que Lise obtienne une plus grand cote Z, il faut que l'écart type de son groupe soit compris un plus grand nombre de fois dans cette distance que celui du groupe de Marie; il doit donc être plus petit.

b) Celui de Lise. Lise a une cote Z plus élevée que celle de Marie, sa note est donc plus éloignée de la moyenne de son groupe que ne l'est la note de Marie, mais comme elles ont la même note, on doit donc en déduire que la moyenne de l'examen est plus faible dans le groupe de Lise.

35.a) $\mu = 4{,}25 \approx 4{,}3$ Md = 4   Mo = 4

b) $\mu = 11{,}6$ Md = 11,2   Aucun mode

c) $\mu = 4{,}375 \approx 4{,}4$ Md = 3,35 ≈ 3,4   Mo = 2,8

36. a) $\sigma^2 \approx 2{,}563$ et $\sigma \approx 1{,}601$

b) $\sigma^2 \approx 12{,}989$ et $\sigma \approx 3{,}604$

c) $\sigma^2 \approx 5{,}269$ et $\sigma \approx 2{,}296$

37. A est la plus dispersée. $\sigma_A \approx 6{,}356$ et $\sigma_B \approx 1{,}414$

38.$\mu = 2{,}8$ Mo = 2 Md = 2,5 $\sigma^2 = 1{,}36$ $\sigma \approx 1{,}166$

39. Z = 1

40. $\mu = 8$, $\sigma = 2$. Les cinq cotes Z sont donc -1,5; -0,5; 0; 0,5 et 1,5. Ces cinq nombres ont, comme il se doit, une moyenne de 0 et une variance de 1.

41. Cote Z = 2,5. La proportion de la population dont la cote Z est, en valeur absolue, supérieur ou égale à 2,5 est au maximum 16 % ce qui, dans une population de 100 000 représente 16 000 personnes.

42. La moyenne est d'environ 64,4 %.

43. Pour janvier, $\sigma \approx 2{,}565$ et pour juillet, $\sigma \approx 1{,}021$

44. Mo = 1   Interprétation : Une majorité (70,6 %) de logements de la région métropolitaine de Montréal en 1981 avaient 1 famille.

Md = 1   Interprétation : Au moins 50 % des logements de la région métropolitaine de Montréal en 1981 avaient 1 famille.

45. Le supermarché A a probablement des dépenses plus élevées, car le nombre de clients varie davantage.

46. Si l'écart-type de la population est 1, un score de 65 est impressionnant car il y a au plus 4 % de la population avec un score aussi éloigné de la moyenne. Si l'écart-type de la population est 20, un score de 65 est assez banal.

47. A est mieux situé par rapport à son groupe que B par rapport au sien. $Z_A = 3$ et $Z_B = 1{,}5$

48. Il serait raisonnable d'attribuer la baisse des recettes aux travaux municipaux car une cote Z de -5 est très significative.