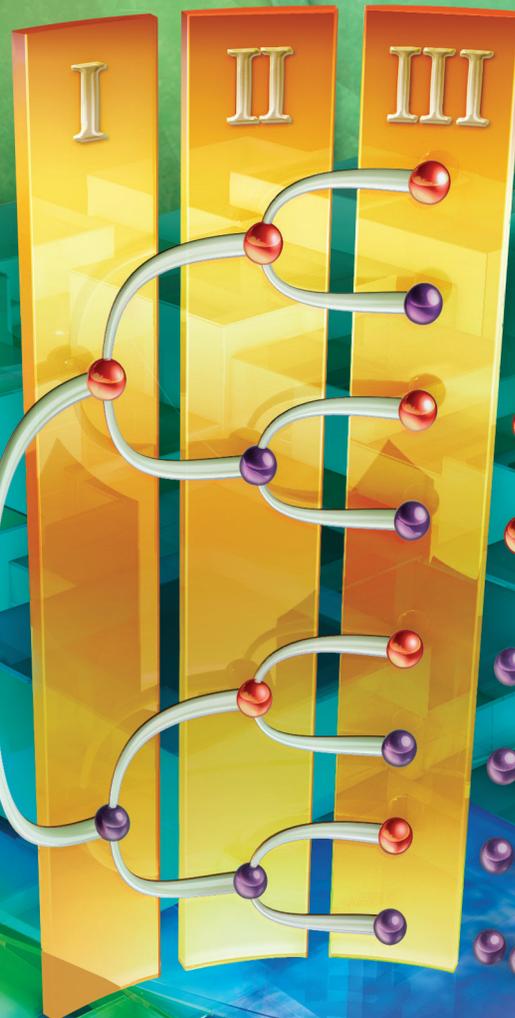


MAT-3052-2

COLLECTE DE DONNÉES



$$P(\bar{E}) = 1 - P(E)$$

sofad

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

COLLECTE DE
DONNÉES

MAT-3052-2

Guide d'apprentissage

The logo for 'sofad' is a black square with the word 'sofad' written in white lowercase letters.

Programme de la formation de base diversifiée

Domaine de la mathématique, de la science et de la technologie

Programme d'études : Mathématique

Cours de la 3^e secondaire

MAT-3051-2 Modélisation algébrique et graphique

MAT-3052-2 Collecte de données

MAT-3053-2 Représentation géométrique

Collecte de données

Ce guide a été réalisé par la SOFAD.

Chargés de projets :	Robert Longpré (SOFAD) Nancy Mayrand (SOFAD)
Rédaction :	Jean-Claude Hamel
Révision de contenu :	Steeve Lemay Isabelle Hachez
Révision linguistique :	Michelle Côté Marie-Pierre Gazaille
Révision docimologique :	Julie Gravel
Illustrations :	Serge Mercier
Montage infographique :	Serge Mercier
Page couverture :	Marc Tellier
Lecture d'épreuve :	Marie-Pierre Gazaille
Première édition :	Septembre 2014

Dans cette production, le masculin est utilisé sans aucune discrimination et uniquement dans le but d'alléger le texte.

Nonobstant l'énoncé suivant, la SOFAD autorise tout centre d'éducation aux adultes qui utilise ce guide d'apprentissage à reproduire les activités notées accessibles au <http://cours1.sofad.qc.ca/ressources>.

© SOFAD

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 microreproduction, est interdite sans l'autorisation écrite d'un représentant dûment autorisé de la SOFAD.

Dépôt légal – 2016

Bibliothèque et Archives nationales du Québec

Bibliothèque et Archives Canada

ISBN : 978-2-89493-455-5 (Imprimé)

978-2-89493-544-6 (PDF)

Janvier 2016

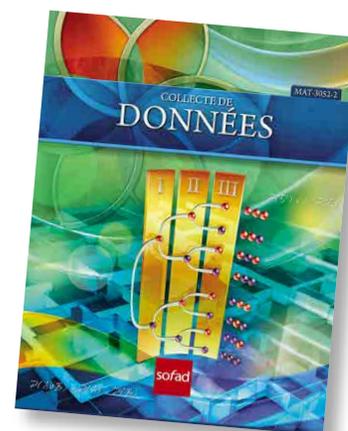
Table des matières

Introduction	7
Présentation	7
Structure du guide et consignes d'utilisation	8
Matériel complémentaire	11
Soutien à l'apprentissage	11
Information complémentaire concernant la formation à distance	12
Évaluation aux fins de sanction	12
Savoirs essentiels	13
Situation ❶ – Réaliser un sondage	15
Présentation	15
Exploration	17
Activité 1.1 – Campagne électorale	22
Activité 1.2 – Sur la scène municipale	31
Activité 1.3 – Quelques centimètres de plus	42
Un peu plus loin - Méthodes d'échantillonnage non probabiliste	51
Exercices d'intégration	54
Activité synthèse – Questionner sur l'utilisation d'Internet	59
Situation ❷ – Comparer des données	65
Présentation	65
Exploration	68
Activité 2.1 – Une négociation de salaire	72
Activité 2.2 – Le prix d'un film	80
Activité 2.3 – Rendement et risque	96
Un peu plus loin – Une tendance centrale estimée	109
Exercices d'intégration	113
Activité synthèse – Analyser la variation du prix de l'essence	117
Consignes pour la réalisation de l'activité notée 1	122
Situation ❸ – Les jeux de hasard	123
Présentation	123
Exploration	126
Activité 3.1 – Rien ne va plus!	130
Activité 3.2 – Un jeu équitable?	144
Activité 3.3 – Qui va brasser les cartes?	159
Activité 3.4 – Jeu de lettres	169
Un peu plus loin - Les factorielles!	178
Exercices d'intégration	179
Activité synthèse - Conceptions erronées en probabilités	181



Situation ④ – Utiliser les probabilités dans la vie courante	189
Présentation	189
Exploration	192
Activité 4.1 – Un couple d’aubergistes	196
Activité 4.2 – Une assiette sur le plancher	208
Activité 4.3 – Du hasard au supermarché	217
Un peu plus loin - Une étrange coïncidence	224
Exercices d’intégration	226
Activité synthèse – Serais-je en retard?	229
Consignes pour la réalisation de l’activité notée 2	235
Autoévaluation	237
Conclusion.	255
Corrigé.	257
① – Réaliser un sondage	258
② – Comparer des données	268
③ – Les jeux de hasard	283
④ – Utiliser les probabilités dans la vie courante	293
Autoévaluation	304
Lexique	311
Sources iconographiques	323
Fiche de commentaires.	325

Introduction



Présentation

Nous vous souhaitons la bienvenue dans le guide d'apprentissage du cours *Collecte de données*. Ce cours de mathématique est le deuxième que vous devez suivre en 3^e secondaire. Il a pour but de développer votre habileté à traiter des situations qui requièrent la collecte ou le traitement de données.

Dans les pages qui suivent, vous découvrirez différentes façons de construire un échantillon représentatif d'une population pour réaliser un sondage. Vous apprendrez aussi à représenter les données recueillies à l'aide de tableaux ou de diagrammes et à déterminer des mesures associées à leurs distributions pour les analyser et les comparer. De plus, vous traiterez des données issues d'expériences aléatoires. Vous utiliserez différents outils mathématiques pour dénombrer les résultats possibles et pour calculer des probabilités. Cela vous permettra de tirer des conclusions, de valider des résultats et de prendre des décisions éclairées dans une multitude de situations.

Dans ce cours, vous serez amené à exercer votre compétence à raisonner. Vous ferez des hypothèses et devrez chercher à les valider. Vous devrez justifier vos réponses à l'aide d'arguments mathématiques, en prenant soin de présenter des démarches claires et structurées. Vous serez également amené à exercer votre compétence à résoudre des situations-problèmes. Vous devrez utiliser différentes stratégies et faire preuve de persévérance pour surmonter les obstacles auxquels vous serez confronté.

Vous aurez aussi quelques occasions de développer vos compétences transversales, notamment celles d'exercer votre jugement critique et de communiquer de façon appropriée.

Vous êtes maintenant invité à parcourir les quatre situations d'apprentissage de ce guide qui vous permettront d'enrichir vos connaissances et de développer vos compétences en statistique et en probabilité.

Structure du guide et consignes d'utilisation

Le présent guide a été conçu pour permettre un apprentissage en mode individualisé, en établissement ou à distance. Il s'appuie sur l'un des principaux objectifs de l'apprentissage de la mathématique qui consiste à vous amener à résoudre divers types de situations-problèmes.

Cette orientation rendra votre cheminement des plus agréables, puisqu'elle favorisera chez vous :

- la plus grande participation possible;
- la prise en charge de votre cheminement;
- le respect de votre rythme;
- l'esprit réflexif;
- la mise à profit de votre expérience et de vos connaissances.

Tout au long de votre formation, vous trouverez dans ce guide des outils pour mesurer vos succès et déterminer les moyens à prendre afin de surmonter les aspects qui vous sembleront plus ardues. Vous pourrez ainsi progresser continuellement dans votre apprentissage.

Si vous n'y arrivez pas, une personne-ressource pourra aussi vous venir en aide. Les élèves étudiant à distance pourront obtenir le soutien d'un tuteur, alors qu'un formateur prêtera main forte aux élèves fréquentant un centre de formation. Renseignez-vous auprès de cette personne pour connaître les détails de cet accompagnement.

Les situations d'apprentissage

Le présent guide renferme des situations d'apprentissage dans lesquelles vous serez amené à découvrir de nouveaux concepts ainsi qu'à développer vos compétences. Chaque situation d'apprentissage est construite sur le même modèle. Elle comporte d'abord une *Présentation* dans laquelle une tâche vous est présentée. Normalement, vous n'arriverez pas à résoudre le problème dès le début. Ce n'est qu'à la fin de la situation d'apprentissage que vous aurez toutes les connaissances et les compétences pour y arriver.

Ensuite vient une section *Exploration* qui fait un retour sur certains savoirs vus dans les cours précédents. Suivent un certain nombre d'activités d'apprentissage. Dans chacune de ces activités, un sujet est traité et des questions vous sont posées. Il se pourrait que vous ne soyez pas en mesure de répondre à toutes les questions, car elles portent sur de nouveaux concepts. Efforcez-vous tout de même d'y trouver des réponses satisfaisantes; les explications se trouvent immédiatement après. Il est essentiel que vous cherchiez à comprendre tous les nouveaux concepts qui sont expliqués. À la fin des explications, vous trouverez des exercices. Les réponses de ces exercices se trouvent à la fin du guide.

La situation d'apprentissage est accompagnée d'une section *Un peu plus loin*. Cette partie est facultative, elle permettra d'approfondir vos connaissances. Suit la section *Exercices d'intégration* portant sur des concepts abordés dans la situation d'apprentissage. Les réponses à ces exercices se trouvent aussi à la fin du guide. Une fois ces exercices complétés, il est temps pour vous de réaliser la tâche du début dans la section *Activité synthèse*. Finalement, les situations d'apprentissage se terminent avec une *Liste des nouveaux savoirs*. C'est une section fort utile pour l'élève qui aime bien trouver les savoirs importants en un clin d'œil.

Les repères visuels

Tout au long du texte, différents signes et pictogrammes vous guideront dans votre apprentissage. Le tableau ci-dessous vous indique la signification de chacun de ces pictogrammes.

<p>Soulignement</p>	<p>Les mots et les expressions soulignés en pointillés sont définis dans le lexique à la fin du guide.</p>
<p>Le saviez-vous </p>	<p>Un texte coiffé d'une loupe ajoute un complément d'information : il ne fait pas directement partie de l'apprentissage et aucune question de l'évaluation finale ne portera sur son contenu.</p>
<p>Astuce </p>	<p>Une ampoule coiffe les encadrés qui présentent une astuce permettant de simplifier le travail ou d'offrir une autre façon de faire.</p>
<p>Rappel </p>	<p>Les encadrés coiffés d'une punaise de babillard contiennent des rappels de notions ou de concepts préalables abordés dans des cours précédents.</p>
<p>À retenir </p>	<p>Le trombone fixé sur une feuille de papier au coin relevé détermine des éléments à retenir.</p>
<p>Liste des nouveaux savoirs </p>	<p>La pince à papier accompagne les dernières pages de chaque situation d'apprentissage. Ces pages présentent un résumé des savoirs essentiels qui y ont été abordés.</p>
<p>Attention </p>	<p>Le point d'exclamation accompagne les paragraphes auxquels vous devez porter une attention particulière.</p>

Les activités notées

Ce guide est accompagné de deux activités notées présentées dans des cahiers séparés. Le but de ces activités est de vérifier votre progression véritable. Il est important de les compléter sérieusement et sans l'aide de personne. Consultez la table des matières pour connaître à quel moment faire chaque cahier. Si ces cahiers ne vous ont pas été fournis, vous pouvez les télécharger sur le site des ressources pour les apprenants au : <http://cours1.sofad.qc.ca/ressources>, sous la rubrique « Formation de base diversifiée ».

Le tableau qui suit vous indique les thèmes qui sont évalués par chacune des activités et à quel moment vous devez les réaliser.

SITUATION D'ÉVALUATION	THÈMES TOUCHÉS	MOMENT DE RÉALISATION
Activité notée 1	Distributions statistiques à un caractère (Situations d'apprentissage 1 et 2)	Après la deuxième situation d'apprentissage
Activité notée 2	Probabilité (Situations d'apprentissage 3 et 4)	Après la quatrième situation d'apprentissage

Une fois que vous avez terminé une activité notée, il est important de la remettre à votre formateur ou de la faire parvenir à votre tuteur pour fins de correction. Généralement, on ne remet qu'une seule activité notée à la fois et on doit attendre qu'elle soit corrigée avant de remettre la suivante. Informez-vous auprès de votre centre ou de votre commission scolaire.

L'autoévaluation

La dernière activité du guide consiste en une épreuve d'autoévaluation dont le but est d'évaluer les connaissances acquises, de même que les compétences développées. Une grille d'autoévaluation accompagne cette épreuve. Elle vous servira à déterminer les notions que vous maîtrisez et celles pour lesquelles une révision s'impose avant de vous présenter à l'épreuve de sanction. Des indications sur les notions à réviser sont fournies à même cette grille.

Avant de vous y attaquer, préparez-vous. Révisez les notions à l'aide des sections *Liste des nouveaux savoirs* et assurez-vous d'avoir bien fait les exercices. Il est recommandé de faire cette activité sans consulter le texte du guide ni le corrigé. Une fois l'autoévaluation terminée, comparez vos réponses avec celles du corrigé et complétez votre étude au besoin.

Le corrigé

Dans l'avant-dernière partie du guide, vous trouverez les corrigés de tous les exercices. Reportez-vous à ce corrigé pour vous assurer que vous avez bien compris tous les concepts, et ce, avant de passer à l'activité ou à la situation d'apprentissage suivante. À la fin de cette section se trouve le corrigé de l'épreuve d'autoévaluation.

Notez qu'il n'y a pas de corrigé pour les questions liées aux explications. Seuls les problèmes numérotés se trouvent dans le corrigé.

Le lexique

Dernière partie du guide, le lexique est un outil qui vous permet de consulter les définitions, classé en ordre alphabétique, des mots soulignés en pointillés dans les situations d'apprentissage. N'hésitez pas à le consulter au fil de vos lectures afin de bien comprendre les termes et expressions qui s'y trouvent.

Matériel complémentaire

Ayez sous la main tout le matériel dont vous aurez besoin :

- ce guide d'apprentissage;
- un cahier de notes où vous consignerez toutes les informations que vous jugerez pertinentes;
- une calculatrice scientifique (recommandée);
- un crayon à mine pour inscrire vos réponses et vos notes dans votre guide, un stylo-bille de couleur pour corriger vos réponses, un surligneur pour mettre en évidence les idées importantes, une gomme à effacer, une règle graduée, etc.;
- un dictionnaire, comme ouvrage de référence.
- des instruments de géométrie tels qu'une règle graduée, un rapporteur et un compas;
- des dés réguliers à 6 faces et des pièces de monnaie.

Soutien à l'apprentissage

Que vous suiviez le cours en établissement ou à distance, votre démarche ne se fera pas en solitaire. En classe, vous aurez le soutien de votre formateur; tandis qu'en formation à distance, vous pourrez compter sur le soutien d'un tuteur. Ils répondront à vos questions portant sur le contenu, vous guideront dans la réalisation des activités d'apprentissage, corrigeront vos activités notées et vous transmettront vos notes.

Information complémentaire concernant la formation à distance

Voici quelques suggestions qui vous aideront à organiser votre temps d'étude. La durée de la formation de ce cours est évaluée à environ 50 heures de travail.

Dès que vous recevez le matériel, établissez un horaire d'étude en fonction de vos obligations familiales et professionnelles, de même que des exigences du cours. Essayez de consacrer quelques heures par semaine à l'étude, de préférence en blocs de deux heures, et veillez à respecter votre horaire.

Le tuteur est la personne-ressource à qui vous ferez appel en cas de besoin. N'hésitez pas à l'interroger si vous éprouvez des difficultés avec la théorie ou les exercices, ou encore si vous avez besoin d'encouragement pour poursuivre votre étude. Notez vos questions par écrit au fur et à mesure qu'elles surgissent et communiquez avec votre tuteur pendant ses heures de disponibilités. Assurez-vous que le centre vous ait bien fourni ces informations.

Évaluation aux fins de sanction

Si vous désirez acquérir les unités rattachées à ce cours, vous devez obtenir une note d'au moins 60 % à l'évaluation finale qui aura lieu dans un centre d'éducation des adultes. Pour vous présenter à cette épreuve, il est souhaitable que vous ayez également obtenu une moyenne d'au moins 60 % aux activités notées accompagnant le présent guide. D'ailleurs, certains centres d'éducation aux adultes exigent ce résultat de 60 % aux activités notées pour vous admettre à l'épreuve officielle.

Informez-vous auprès de votre tuteur ou de votre formateur pour connaître le lieu, les horaires et les modalités de passations de l'épreuve finale. Informez-vous aussi du matériel autorisé que vous pouvez apporter.

Savoirs essentiels

Le présent cours vise l'appropriation des savoirs mathématiques suivants.

SITUATION D'APPRENTISSAGE	SAVOIRS MATHÉMATIQUES
1. Réaliser un sondage et communiquer les résultats	<ul style="list-style-type: none"> • Organisation et interprétation de données statistiques • Construction et interprétation de tableaux de distribution • Représentation et interprétation de graphiques (histogramme)
2. Comparer des données	<ul style="list-style-type: none"> • Calcul de mesures de tendance centrale et de dispersion • Représentation et interprétation de graphiques (diagrammes de quartiles)
3. Comprendre les jeux de hasard	<ul style="list-style-type: none"> • Dénombrement et calcul de probabilités • Représentation d'événements (à l'aide de tableaux et de diagrammes en arbre)
4. Utiliser les probabilités dans la vie courante	<ul style="list-style-type: none"> • Dénombrement et calcul de probabilités • Représentation d'événements (à l'aide de diagrammes, d'arbres et de figures géométriques)

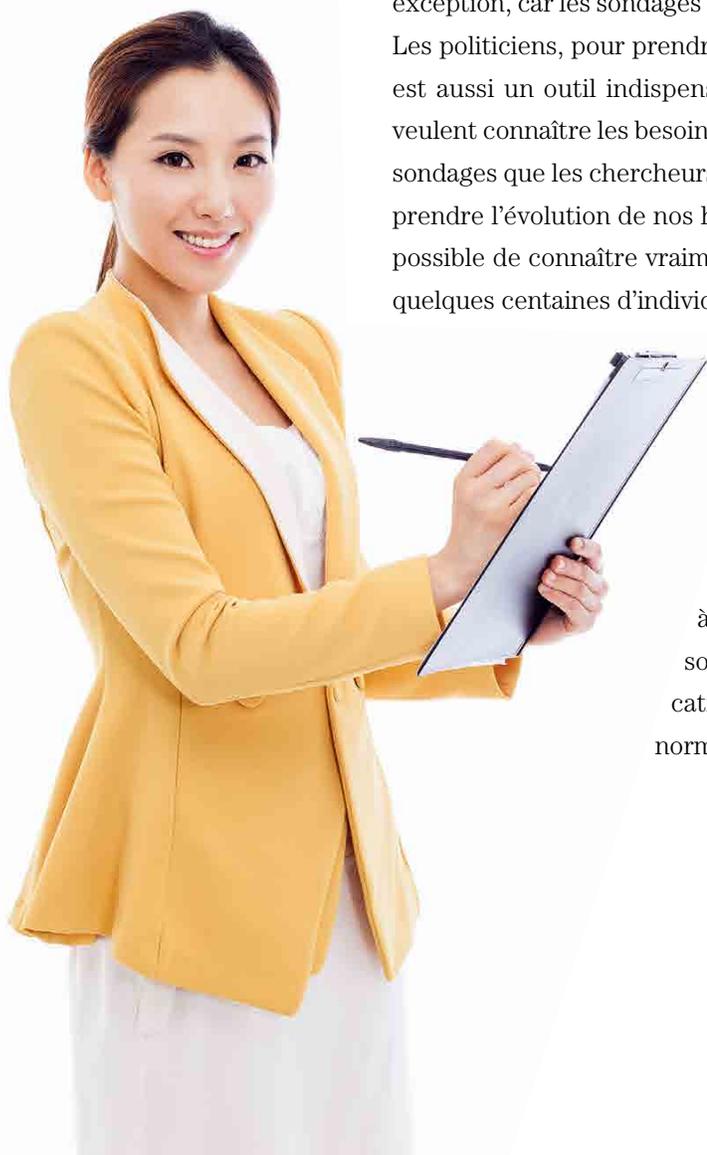
1

Réaliser un sondage

Présentation

Par téléphone, sur Internet ou en marchant dans un centre commercial, combien de fois vous êtes-vous fait demander : « Pouvez-vous prendre quelques minutes pour répondre à un sondage ? ». Si ça ne vous est jamais arrivé, vous êtes sûrement une exception, car les sondages sont omniprésents dans plusieurs domaines de la vie en société. Les politiciens, pour prendre le pouls des électeurs, ne pourraient s'en passer. Le sondage est aussi un outil indispensable des services de marketing de plusieurs entreprises qui veulent connaître les besoins et les préférences d'éventuels clients. C'est également par des sondages que les chercheurs et les analystes en sciences humaines essaient de mieux comprendre l'évolution de nos habitudes de vie. Mais tous ces sondages sont-ils fiables ? Est-il possible de connaître vraiment l'opinion de toute une population à partir des réponses de quelques centaines d'individus ? Si oui, comment s'en assurer ?

La présente situation d'apprentissage vise à vous faire découvrir différentes étapes de la réalisation d'un sondage, du choix de l'échantillon jusqu'à la communication des résultats. Vous apprendrez à utiliser de nouvelles méthodes d'échantillonnage et vous approfondirez vos connaissances sur la collecte de données et sur la présentation des résultats à l'aide de tableaux et de diagrammes. Vous aurez à réaliser un sondage. Ce faisant, vous devrez vous assurer que la communication des résultats ne soit pas ambiguë et qu'elle réponde aux normes et aux conventions en vigueur.



Questionner sur l'utilisation d'Internet

Si vous faites partie des deux milliards d'utilisateurs d'Internet dans le monde, ou encore des quelques 20 millions de Canadiens qui sont abonnés à un réseau social, il est probable que vous passiez beaucoup de temps devant votre écran, que ce soit pour communiquer, visionner des films, regarder des photos, écouter de la musique, chercher des informations, jouer, ou même faire des achats. Cet outil a changé notre façon de voir le monde au cours des dernières décennies et il est probable qu'il participe encore à bien d'autres changements à venir!



Pour la plupart des gens, Internet fait maintenant partie du quotidien, à des fins personnelles, pour le travail ou pour les études. Cependant, certaines personnes critiquent l'omniprésence du virtuel au détriment de la réalité. Si l'on passe plusieurs heures devant son ordinateur, il n'en reste plus beaucoup pour passer du temps de qualité avec nos proches.

Posez-vous maintenant la question suivante : Internet prend-il trop de place dans ma vie? Précisez l'utilisation que vous en faites et le temps que vous y consacrez : _____

Il serait intéressant de poser cette question aux gens qui vous entourent, puis de comparer leurs réponses. Notez que si l'on posait la même question dans un forum en ligne, les réponses seraient probablement différentes de celles de votre entourage. Ce serait sûrement aussi le cas si l'on s'adressait à une plus grande population qui regroupe des individus n'ayant pas nécessairement les mêmes habitudes que vous.

Imaginez maintenant la situation suivante. La directrice d'un centre de formation voudrait connaître à quelle fréquence et de quelle façon les adultes qui étudient dans son centre utilisent Internet. Elle propose deux questions :

- 1) Pour quel type d'activités utilisez-vous principalement Internet?
- 2) Combien d'heures par semaine passez-vous sur Internet à des fins personnelles (autres que pour votre travail ou vos études)?

La directrice vous demande de réaliser un sondage pour répondre à ces questions.

Votre tâche ➡

À partir des données qui vous seront fournies dans l'Activité synthèse et qui concernent une population d'élèves adultes, vous devrez d'abord choisir un échantillon représentatif, en tenant compte de la répartition des individus selon leur âge et leur sexe, puis collecter les données en réponse aux deux questions proposées par la directrice. Il vous faudra ensuite représenter ces données à l'aide de tableaux et de diagrammes appropriés. Vous devrez enfin interpréter ces diagrammes afin de dresser un portrait de la situation.

Exploration

Les questions suivantes vous permettront de réviser quelques concepts et processus concernant les sondages. Au besoin, référez-vous au lexique pour vous remémorer le sens de certains mots.

- 1.1 Complétez le texte ci-dessous en plaçant chacun des mots suivants au bon endroit : caractère, échantillon, population, recensement, sondage.

Une étude statistique porte sur un ensemble d'individus (personnes, êtres vivants, objets) que l'on désigne par le terme _____. L'étude cherche à décrire au moins un _____ particulier des éléments de cet ensemble (par exemple, le revenu annuel des personnes). Si l'on recherche l'information chez tous les éléments de l'ensemble, il s'agit d'un _____. Toutefois, la plupart du temps, il est plus économique et plus rapide de réaliser un _____ en recherchant l'information à partir d'un _____, qui est un sous-ensemble de la population.

Rappel



Pour qu'un sondage ne soit pas biaisé, l'échantillon doit être représentatif de la population étudiée. Il est donc important de porter une attention particulière à différentes sources de biais :

- une population mal définie
- un échantillon trop petit
- des éléments de l'échantillon liés entre eux
- un échantillon non représentatif de la population

- 1.2 Les questions de sondage ci-dessous contiennent une ou plusieurs erreurs. Ce sont autant de sources de biais qui peuvent invalider les résultats d'un sondage. Identifiez ces erreurs et suggérez une façon de les corriger.

a)

Allez-vous souvent au théâtre?

Oui Non

Erreur : _____

Correction possible : _____

b)

Dans quel groupe d'âge vous situez-vous ?

Moins de 20 ans

Moins de 30 ans

Plus de 30 ans

Erreurs : _____

Correction possible : _____

c)

Avez-vous apprécié l'excellent reportage de notre journaliste sur le monde interlope diffusé hier soir ?

Erreurs : _____

Correction possible : _____

Rappel



Voici deux méthodes d'échantillonnage probabiliste que vous avez déjà étudiées:

- **Échantillonnage aléatoire simple** : chacun des éléments de l'échantillon est tiré au hasard dans la population. La probabilité d'être choisi est la même pour tous les éléments de la population.
- **Échantillonnage systématique** : tous les éléments de la population sont inscrits dans une liste. On choisit un élément au hasard parmi les n premiers éléments de la liste, puis on ajoute chaque $n^{\text{ième}}$ élément qui suit dans la liste.

1.3

Une chaîne de montage robotisée fabrique des ampoules électriques, sans arrêt de 8 h à minuit. Chaque jour, une cinquantaine d'ampoules sont prélevées pour subir différents tests le lendemain afin de vérifier leur qualité. Voici trois échantillons possibles. Lequel est préférable selon vous ? Entourez la bonne réponse et nommez cette méthode d'échantillonnage.

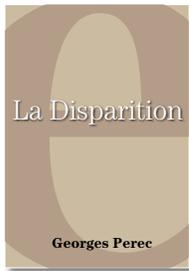
a) Les 50 premières ampoules produites de la journée.

b) Les 50 premières ampoules produites à partir d'une certaine heure choisie au hasard.

c) Une ampoule prise à chaque 20 minutes environ sur la chaîne de montage.

Méthode : _____

Le saviez-vous ?



En 1968, l'écrivain français Georges Perec a écrit un roman d'environ 300 pages, intitulé *La Disparition*. Ce roman est très particulier, car il ne contient aucun *e*.

Cet écrivain original était membre de l'Oulipo, un groupe international de mathématiciens et littéraires. Il a fondé la plupart de ses œuvres sur la notion de contraintes.

1.4 Voici un extrait du roman *La Disparition* de Georges Perec.

Anton Voyl n'arrivait pas à dormir. Il alluma. Son Jaz marquait minuit vingt. Il poussa un profond soupir, s'assit dans son lit, s'appuyant sur son polochon. Il prit un roman, il l'ouvrit, il lut; mais il n'y saisissait qu'un imbroglio confus, il butait à tout instant sur un mot dont il ignorait la signification. Il abandonna son roman sur son lit. Il alla à son lavabo; il mouilla un gant qu'il passa sur son front, sur son cou. Son pouls battait trop fort. Il avait chaud. Il ouvrit son vasistas, scruta la nuit. Il faisait doux. Un bruit indistinct montait du faubourg. Un carillon, plus lourd qu'un glas, plus sourd qu'un tocsin, plus profond qu'un bourdon, non loin, sonna trois coups.

Source : Éditions Denoël, 1969, p. 17

Voici, de nouveau, le même texte, sans les espaces, les apostrophes et la ponctuation. Les lettres ont été regroupées en paquet de 5 pour qu'il soit plus facile de les compter.

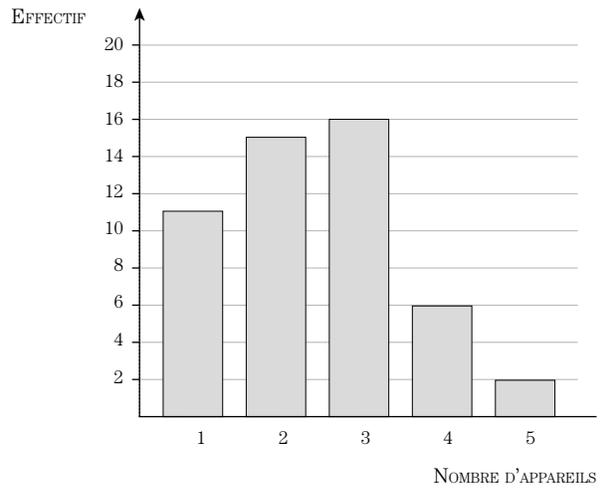
Anton	Voyln	arriv	aitpa	sàdor	mirll	allum	aSonJ	azmar	quait
minui	tving	tllpo	ussau	nprof	ondso	upirs	assit	danss	onlit
sappu	yants	urson	poloc	honll	pritu	nroma	nillo	uvrit	illut
maisi	lnysa	isiss	aitqu	unimb	rogli	oconf	usilb	utait	àtout
insta	ntsur	unmot	donti	ligno	raitl	assign	ifca	tionl	laban
donna	sonro	mansu	rsonl	itlla	llaàs	onlav	aboil	mouil	laung
antqu	ilpas	sasur	sonfr	ontsu	rsonc	ouSon	pouls	batta	ittro
pfort	llava	itcha	udllo	uvrit	sonva	sista	sscru	talan	uitll
faisa	itdou	xUnbr	uitin	disti	nctmo	ntait	dufau	bourg	Uncar
illon	plusl	ourdq	uungl	asplu	ssour	dquun	tocsi	nplus	profo
ndquu	nbour	donno	nloin	sonna	trois	coups			

Il n'y a aucun *e* dans cet extrait composé de 535 lettres, mais on y retrouve d'autres voyelles. Quel pourcentage des lettres ces voyelles représentent-elles ?

1.5 Le diagramme à bandes ci-dessous représente les résultats d'un sondage réalisé auprès d'un échantillon de ménages pour connaître le nombre d'appareils téléphoniques que chacun possède.

a) Selon vous, que signifie le mot effectif inscrit le long de l'axe vertical ?

NOMBRE D'APPAREILS TÉLÉPHONIQUES PAR MÉNAGE SELON UN SONDAGE



b) Parmi les ménages interrogés, combien possèdent

2 appareils téléphoniques ? _____

plus de 3 appareils téléphoniques ? _____

c) Quelle est la taille de l'échantillon ? _____

d) Déterminez la fréquence (%) associée à chaque nombre d'appareils.

1 appareil : _____ 2 appareils : _____ 3 appareils : _____

4 appareils : _____ 5 appareils : _____

Attention !

Dans ce guide, comme c'est généralement le cas en statistique, le mot « fréquence » désignera le rapport de l'effectif au nombre total de données. Ce rapport sera donné en pourcentage.

Sachez cependant que, dans certains livres, le mot « fréquence » est considéré comme un synonyme d'« effectif » et c'est alors l'expression « fréquence relative » qui est utilisée pour parler de ce rapport.

Activité 1.1 Campagne électorale

But ➡

- Organiser et interpréter des statistiques à l'aide des méthodes d'échantillonnage stratifié et d'échantillonnage par grappes.

Une élection partielle est annoncée dans votre circonscription pour remplacer un député provincial qui a démissionné de ses fonctions. Vous décidez de vous engager politiquement en faisant du bénévolat pour l'un des candidats.

En début de campagne, le comité électoral discute de la possibilité de réaliser un sondage pour connaître les principales préoccupations des électeurs et estimer les appuis au parti.

« Il faudrait s'assurer, dit l'une des bénévoles, que toutes les catégories d'âge soient bien représentées dans l'échantillon, car les personnes âgées n'ont pas les mêmes préoccupations que les jeunes. »

À la fin de cette réunion, on vous nomme responsable du sondage. Votre première tâche est de construire, à partir de la liste électorale, un échantillon représentatif de 500 électeurs, en tenant compte de leur âge.

Comment pourriez-vous vous assurer que votre échantillon soit conforme à la répartition des âges des électeurs dans la circonscription? Proposez une démarche.



Le tableau ci-dessous contient quelques informations concernant votre circonscription.

TABLEAU 1.1

CAMPAGNE ÉLECTORALE

NOMBRE D'ÉLECTEURS INSCRITS	51 880
RÉPARTITION DES ÉLECTEURS SELON LEUR ÂGE	
De 18 à 29 ans :	6425
De 30 à 44 ans :	12 220
De 45 à 59 ans :	17 575
60 ans et plus :	15 660
N.B. Les électeurs de votre circonscription sont répartis à peu près également entre 196 sections de vote.	

À l'aide des données du tableau ci-contre, déterminez le nombre d'électeurs de chaque groupe d'âge que devrait idéalement contenir votre échantillon.

Le saviez-vous

Au Québec, lors des élections provinciales, chaque circonscription est divisée en sections de vote contenant généralement entre 200 et 300 électeurs chacune. Lors d'une campagne électorale, les candidats et leur parti reçoivent, pour chaque section de vote, une liste comprenant les noms et prénoms des électeurs, leur date de naissance et l'adresse de leur domicile. Les candidats se servent de ces listes pour joindre les électeurs afin de sonder leur opinion ou de les convaincre de voter pour eux.

L'échantillonnage stratifié

Un échantillon a plus de chance d'être représentatif d'une population si celle-ci est homogène, c'est-à-dire si les individus qui la composent sont semblables en ce qui concerne le sujet de l'étude. Malheureusement, les populations parfaitement homogènes sont plutôt rares. Par contre, il arrive parfois qu'une population puisse être partagée en groupes distincts, souvent plus homogènes que la population elle-même. On dira alors que la population est fragmentée en strates.

Votre échantillon de 500 électeurs sera plus représentatif de la population si chaque strate, c'est-à-dire chaque groupe d'âge, y est correctement représentée. Idéalement, il faudrait que la fréquence de chaque strate de l'échantillon soit la même que dans la population.

Selon le tableau 1.1, les personnes de 18 à 29 ans représentent 6425 des 51 880 électeurs de la circonscription. Dans la population, cela correspond à une fréquence de $\frac{6425}{51\,880}$, soit environ 12,4 %.

Cette proportion doit être conservée dans l'échantillon : 12,4 % de 500 est égal à 62.

L'échantillon devra donc contenir environ 62 électeurs de 18 à 29 ans.

Déterminez la taille idéale de chaque strate en complétant le tableau suivant.

TABLEAU 1.2**RÉPARTITION DES ÉLECTEURS SELON LEUR GROUPE D'ÂGE**

GROUPE D'ÂGE	EFFECTIF	FRÉQUENCE	TAILLE IDÉALE DE LA STRATE DANS L'ÉCHANTILLON
De 18 à 29 ans	6425	$\frac{6\,425}{51\,880} \approx 12,4 \%$	$12,4 \% \times 500 \approx 62$
De 30 à 44 ans	12 220	$\frac{12\,220}{51\,880} \approx$	≈ 118
De 45 à 59 ans	17 575		
60 ans et plus	15 660		
TOTAL	51 880	$\approx 100 \%$	500

La méthode d'échantillonnage stratifié est un procédé qui permet de s'assurer que l'échantillon est partagé en strates selon les mêmes proportions que dans la population.

Pour construire l'échantillon de 500 personnes dans la présente situation, il faudrait donc choisir, au hasard :

- 62 personnes dans l'ensemble des 6425 électeurs de 18 à 29 ans;
- 118 personnes dans l'ensemble des 12 220 électeurs de 30 à 44 ans;
- 169 personnes dans l'ensemble des 17 575 électeurs de 45 à 59 ans;
- 151 personnes dans l'ensemble des 15 660 électeurs de 60 ans et plus.

Cette méthode d'échantillonnage proposée est-elle probabiliste ? Expliquez votre raisonnement.

Résumons la **méthode d'échantillonnage stratifié**. Elle consiste à :

- fragmenter la population en strates;
- déterminer la taille de chaque strate dans l'échantillon selon la proportion suivante :

$$\frac{\text{Taille de la strate dans l'échantillon}}{\text{Taille de l'échantillon}} = \frac{\text{Taille de la strate dans la population}}{\text{Taille de la population}}$$

- appliquer la méthode d'échantillonnage aléatoire simple, à chaque strate de la population, comme si chacune était une population en soi.

FIGURE 1.1



Il s'agit d'une méthode probabiliste, car elle est basée sur le hasard. Dans le cas de cette campagne électorale, si vous utilisez la méthode d'échantillonnage stratifié, chaque électeur, quel que soit son âge, aura approximativement la même probabilité d'être choisi.

L'échantillonnage par grappes

Un membre du comité électoral souligne qu'il peut être long et compliqué de joindre 500 personnes réparties dans toute la circonscription, car celle-ci couvre un très grand territoire. Il propose alors une autre façon de procéder.

« Il y a en moyenne 265 électeurs par section de vote et tous habitent dans le même quartier. Ne serait-il pas plus simple de choisir deux sections de vote au hasard, parmi les 196 sections possibles, et d'interroger tous les électeurs qui s'y trouvent? De cette façon, on limiterait considérablement les déplacements. »



Selon vous, cette méthode proposée peut-elle donner un échantillon représentatif? Si oui, à quelles conditions?

L'une des conditions, pour que la méthode proposée soit valable, consiste à s'assurer que les caractéristiques socio-économiques des électeurs, pour l'ensemble des sections de vote, soient semblables.

Ce ne serait pas le cas, par exemple, si certaines sections de vote étaient constituées majoritairement de retraités, alors que d'autres seraient plutôt constituées d'étudiants. Le revenu des électeurs pourrait aussi être un facteur à considérer. Certains quartiers dans une circonscription peuvent être plus riches que d'autres, ce qui pourrait avoir un impact sur les opinions politiques. Pour que la méthode soit valable, il faut que chaque section de vote ressemble à la population complète.

Lorsqu'une population est partagée en groupes ayant des caractéristiques semblables à celles de la population au complet, ces groupes sont appelés des grappes. Ainsi, dans la présente situation, chaque section de vote peut être considérée comme une grappe, dans la mesure où les conditions décrites ci-dessus sont respectées.

Attention !

Les grappes et les strates sont deux façons bien différentes de diviser une population. Il importe de bien les distinguer.

On parle de **grappes** si tous les groupes se ressemblent; chaque groupe possède alors toutes les caractéristiques de la population.

On parle de **strates** si tous les groupes sont différents; chacun d'eux étant associé à une caractéristique particulière de la population.

La méthode d'échantillonnage par grappes, lorsqu'elle est possible, est une façon économique de construire un échantillon représentatif. Par contre, l'un des défauts de cette méthode est qu'elle ne permet pas de prévoir exactement la taille de l'échantillon, puisque les grappes ne contiennent pas nécessairement toutes le même nombre d'éléments.

Résumons la **méthode d'échantillonnage par grappes**. Elle consiste à :

- Partager la population en grappes.
- Choisir au hasard un certain nombre de grappes.
- Constituer l'échantillon avec tous les éléments des grappes choisies.

FIGURE 1.2



Selon vous, quelles sont les sources de biais possibles de la méthode d'échantillonnage par grappes ?

Vous remarquez qu'il existe des sources de biais spécifiques à cette méthode. Par exemple, les grappes sont souvent constituées d'éléments qui sont liés d'une façon ou d'une autre et qui ne sont donc pas tout à fait indépendants les uns des autres. C'est le cas des sections de vote, puisque tous les électeurs d'une même section de vote habitent le même quartier. Ces électeurs sont plus susceptibles de s'influencer les uns les autres puisqu'ils se rencontrent beaucoup plus souvent que ne le font les électeurs de toute la circonscription.

Vous voilà en mesure de mettre en pratique ces nouvelles connaissances.

À retenir

Méthodes d'échantillonnage probabiliste

Certaines études statistiques s'intéressent à des populations qui ne sont pas homogènes. Dans ce cas, les éléments de la population présentent différents traits qui les distinguent les uns des autres et qui peuvent avoir un effet sur le caractère étudié.

Par exemple : dans une population formée de personnes, les individus peuvent se distinguer par leur âge, leur sexe, leur langue maternelle, leur niveau d'éducation, etc.

Pour construire un échantillon représentatif d'une population, il est possible de partager celle-ci en sous-ensembles, appelés **strates** ou **grappes**, selon la nature des groupes formés.

Échantillonnage stratifié

On parlera de strates, si chaque sous-ensemble est associé à une caractéristique particulière de la population.

Exemple : Construire un échantillon de 50 personnes parmi les 450 adultes d'un centre de formation, sachant que, parmi ces élèves, il y a 207 hommes et 243 femmes.

DESCRIPTION	EXEMPLE
<p>Après avoir fragmenté la population en strates, on détermine le pourcentage de l'échantillon qui sera choisi dans chaque strate selon le rapport suivant :</p> $\frac{\text{Taille de la strate dans la population}}{\text{Taille de la population}}$ <p>Les éléments de l'échantillon sont choisis au hasard dans chaque strate selon la méthode d'échantillonnage aléatoire simple.</p>	<p>Les hommes et les femmes constituent des strates de la population dans les rapports suivants :</p> <p>Hommes : $\frac{207}{450} = 46 \%$</p> <p>Femmes : $\frac{243}{450} = 54 \%$</p> <p>L'échantillon sera constitué de 23 hommes, soit 46 % de 50, et de 27 femmes, soit 54 % de 50, choisis au hasard dans leur strate respective.</p>

Échantillonnage par grappes

On parlera de grappes, si chaque sous-ensemble ressemble plutôt à la population complète avec toutes ses caractéristiques.

Exemple : Construire un échantillon de 50 personnes parmi les 450 adultes d'un centre de formation, sachant qu'il y a environ 18 classes.

DESCRIPTION	EXEMPLE
<p>La population est partagée en grappes. On choisit un certain nombre de grappes au hasard en tenant compte de la taille voulue de l'échantillon. Tous les éléments des grappes choisies font partie de l'échantillon.</p>	<p>Tenant compte du fait qu'il y a environ 25 élèves par classe, on choisit deux classes au hasard parmi toutes celles du centre. L'échantillon est constitué de tous les élèves se trouvant dans les deux classes choisies.</p>

Exercices de l'activité 1.1

1.6 Associez une méthode d'échantillonnage probabiliste à chacune des situations décrites.

a) Pour mesurer l'appréciation des spectateurs à un spectacle de variétés, on interroge la 8e personne qui sort de la salle, puis la 28e, la 48e, la 68e... jusqu'à ce qu'il ne reste plus personne dans la salle.

Aléatoire simple

b) Pour étudier les habitudes de recyclage des ménages dans une ville, on choisit cinq quartiers au hasard et l'on demande à tous les ménages de ces quartiers de répondre à un questionnaire.

Systematique

c) Pour déterminer les cotes d'écoute des émissions de télévision, un sondeur envoie un cahier d'écoute à des ménages choisis au hasard dans une liste contenant tous les numéros de téléphone.

Stratifié

d) Dans une polyvalente, pour déterminer le point de vue des élèves concernant l'intimidation à l'école, on interroge 10 % des élèves de chaque niveau, choisis au hasard dans les listes d'élèves.

Par grappes

1.7 Une organisation non gouvernementale (ONG) voudrait évaluer rapidement le nombre d'enfants de 0 à 5 ans souffrant de malnutrition dans une région d'Afrique frappée par la sécheresse. On estime qu'il y a environ 60 000 enfants de cet âge dans cette région qui regroupe 178 villages ayant des populations à peu près semblables. Les ressources médicales de l'ONG lui permettent d'examiner 2000 enfants dans un délai raisonnable.

a) Selon vous, pourquoi la méthode d'échantillonnage aléatoire simple n'est-elle pas appropriée dans cette situation?

b) Suggérez une façon efficace de former l'échantillon.

1.8) Voici la répartition de la population du Québec selon différentes régions.

POPULATION DE 15 ANS ET PLUS AU QUÉBEC

RÉGION	POPULATION
Côte-Nord et Saguenay-Lac-Saint-Jean	313 200
Bas-Saint-Laurent et Gaspésie-Îles-de-la-Madeleine	253 700
Capitale-Nationale et Chaudière-Appalaches	948 000
Mauricie et Lanaudière	620 200
Estrie et Centre-du-Québec	460 000
Montréal	1 666 600
Laval et Laurentides	800 000
Montérégie	1 216 100
Outaouais	307 700
Abitibi-Témiscamingue et Nord-du-Québec	152 400

Source : Institut de la statistique du Québec (année 2011)

a) La plupart des sondages sur la population du Québec utilisent un échantillon d'environ 1000 personnes. Si l'on tient à ce que cet échantillon représente adéquatement toutes les régions du Québec, combien de personnes devrait-on choisir dans chaque région ? Complétez le tableau suivant.

RÉPARTITION DE L'ÉCHANTILLON PAR RÉGION

RÉGION	ÉCHANTILLON
Côte-Nord et Saguenay-Lac-Saint-Jean	
Bas-Saint-Laurent et Gaspésie-Îles-de-la-Madeleine	
Capitale-Nationale et Chaudière-Appalaches	
Mauricie et Lanaudière	
Estrie et Centre-du-Québec	
Montréal	
Laval et Laurentides	
Montérégie	
Outaouais	
Abitibi-Témiscamingue et Nord-du-Québec	

b) Quelle est la probabilité que vous soyez l'une des personnes choisies dans votre région? Cette probabilité demeurerait-elle la même si vous habitiez une autre région ?

.....

.....

.....

.....

.....

.....

c) Comment nomme-t-on cette méthode d'échantillonnage? _____

Activité 1.2 Sur la scène municipale

But ➡

- Construire et interpréter des tableaux à données condensées.
- Reconnaître les sources de biais dans la représentation et l'interprétation des données.

À deux mois d'une élection municipale, le journal local où vous travaillez a réalisé un sondage maison, à partir d'un échantillon aléatoire de 160 personnes. Voici la question posée :



Voici les réponses obtenues.

S	S	T	N	A	G	N	N	N	T	G	R	N	T	G	G	T	A	S	G
T	G	R	T	N	N	A	G	G	T	A	G	A	G	T	S	N	T	T	S
G	A	S	S	G	N	N	S	N	G	S	A	A	A	G	G	G	N	T	R
N	N	A	A	S	R	A	N	G	R	N	S	N	S	R	S	T	G	N	R
S	T	T	G	T	T	T	T	N	S	N	N	G	G	S	N	S	A	S	G
T	R	G	N	S	T	N	A	A	G	T	T	R	G	G	N	G	A	G	A
N	N	N	N	G	N	T	N	N	R	G	T	S	G	A	G	T	G	R	N
A	G	N	A	T	N	R	S	G	T	G	T	N	S	N	N	A	N	N	T

Signification des lettres :

G = Gagnon;

S = Smith;

T = Tremblay;

A = aucune de ces personnes;

N = ne sait pas ou n'a pas d'opinion;

R = refus de répondre

Le directeur du journal vous demande de rédiger un court article pour décrire et interpréter les résultats de ce sondage.

Compilez les données recueillies, puis complétez le tableau de distribution ci-dessous.

TABLEAU 1.3

LA PERSONNE LA PLUS APTE À COMBATTRE LA CORRUPTION

RÉPONSE POSSIBLE	EFFECTIF	FRÉQUENCE (%)
Mme Ariane Gagnon		
M. John Smith		
M. Arthur Tremblay		
Aucune d'entre elles		
Ne sait pas ou n'a pas d'opinion		
Refus de répondre		
TOTAL		



Astuce



Pour compiler les résultats d'un sondage, surtout lorsqu'il y a plusieurs données, il peut s'avérer pratique d'utiliser un tableau de compilation comme celui ci-contre.

Une fois les résultats compilés, on peut les transposer dans la colonne des effectifs.

G	### ### ### ### ### ### ### = 36
S	### ### ### ### = 22
T	### ### ### ### ### = 29
A	### ### ### ### = 21
N	### ### ### ### ### ### ### ### = 40
R	### ### = 12

Tableau à données condensées

Le tableau 1.3, qui contient le détail des effectifs et des fréquences, porte le nom de tableau à données condensées. Ce type de tableau de distribution statistique est très utile pour mettre en évidence les résultats d'un sondage. Ils peuvent à la fois contenir une colonne des effectifs et/ou une colonne des fréquences. Si le caractère étudié est qualitatif, on énumère à la première colonne toutes les modalités.

Le total des effectifs doit toujours être égal à la taille de l'échantillon, soit 160 dans ce cas-ci. Le rapport de chaque effectif à ce total permet de calculer la fréquence correspondante que l'on exprime généralement en pourcentage. Par exemple, puisque M. Tremblay a récolté 29 intentions de vote, alors :

$$\frac{29}{160} = 0,18125 \approx 18 \% \text{ des répondants.}$$

Voici ce que vous auriez dû obtenir plus tôt.

TABLEAU 1.4

LA PERSONNE LA PLUS APTE À COMBATTRE LA CORRUPTION

MODALITÉ	EFFECTIF	FRÉQUENCE (%)
Mme Ariane Gagnon	36	23
M. John Smith	22	14
M. Arthur Tremblay	29	18
Aucune d'entre elles	21	13
Ne sait pas ou n'a pas d'opinion	40	25
Refus de répondre	12	8
TOTAL	160	≈ 100

Attention !



Le total des pourcentages inscrits dans le tableau peut être légèrement inférieur ou supérieur à 100 % en raison des arrondis de chacune des fréquences. On indiquera alors que la fréquence totale est ≈ 100 %.

Pendant la rédaction de votre article, un collègue vous suggère d'exclure toutes les réponses qui ne désignent pas l'un des candidats (c'est-à-dire les données A, N et R) afin de mieux comparer l'appui que chacun d'eux a reçu.

Comment le pourcentage d'appui pour chaque candidat serait-il modifié dans ce cas ?

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



Que pensez-vous de la suggestion de votre collègue ? Selon vous, est-ce une bonne idée ?

.....

.....

Si l'on suit la suggestion du collègue de ne considérer que les réponses qui désignent un candidat, le total des effectifs ne sera plus de 160, mais bien de 87, soit $29 + 36 + 22$. Le pourcentage associé à chacune de réponse sera donc différent, comme le montre le tableau 1.5. On obtient alors un tout autre portrait de la situation.

TABLEAU 1.5

LA PERSONNE LA PLUS APTE À COMBATTRE LA CORRUPTION APRÈS RÉPARTITION

MODALITÉ	EFFECTIF	FRÉQUENCE (%) APRÈS RÉPARTITION
Mme Ariane Gagnon	36	41
M. John Smith	22	25
M. Arthur Tremblay	29	33
TOTAL	87	≈100

Attention !

Ne pas considérer les refus de répondre et les indécis dans la communication des résultats d'un sondage a le même effet que de répartir ces réponses entre les candidats proportionnellement à leur appui respectif. C'est pourquoi, lorsqu'on procède ainsi, on spécifie généralement que les fréquences calculées ont été obtenues « après répartition ».

Suite à l'analyse de ces nouvelles fréquences, pensez-vous toujours la même chose de la suggestion de votre collègue ?

.....

.....

Le procédé suggéré par votre collègue peut être discutable, car il néglige une information importante : près de la moitié des personnes interrogées, soit 73 sur 160, n'ont pas indiqué de préférence pour l'un des candidats. Plusieurs personnes ont même spécifié qu'ils ne faisaient confiance à aucun d'eux pour combattre la corruption. Ne pas tenir compte de ces réponses dans la présentation des résultats est une source de biais évidente. Par exemple, il est faux de conclure qu'une personne sur 3 (environ 33%) considère M. Tremblay comme le plus apte à combattre la corruption. En effet, en observant le tableau 1.4, on voit que c'est plutôt 18%, soit moins d'une personne sur cinq.

La façon de considérer les indécis et ceux qui refusent de répondre est un problème délicat pour les sondeurs. Faire le choix de les exclure du calcul des pourcentages, c'est faire l'hypothèse que ces personnes n'auraient aucun impact sur les résultats s'ils devaient finalement se prononcer. Or, cette hypothèse n'est pas toujours valide. Il peut arriver que la majorité de ces gens aient des raisons particulières de ne pas exprimer leur opinion. Ces raisons mises au jour peuvent changer complètement le portrait de la situation.

Le saviez-vous



Selon plusieurs firmes de sondage, tant en Europe qu'en Amérique, il est de plus en plus difficile de réaliser des sondages d'opinion fiables à l'aide d'un échantillon aléatoire, puisque le nombre de gens qui refusent de répondre augmente continuellement chaque année. Des études statistiques aux États-Unis ont montré que, actuellement, plus de la moitié des Américains, voire 80 % dans certains cas, refusent de répondre à des sondages téléphoniques. Le taux de refus est moins élevé au Québec, mais il est également en hausse.

Dans votre article, vous devez mettre en évidence les résultats de ce sondage qui vous paraissent les plus pertinents. Écrivez deux ou trois conclusions qu'il est possible de tirer des données recueillies.



Quel titre donneriez-vous à votre article ?

En ayant en tête toutes les sources de biais possibles, il est tout de même possible d'écrire un article sur ce sondage. On pourrait, par exemple, en tirer les conclusions suivantes :

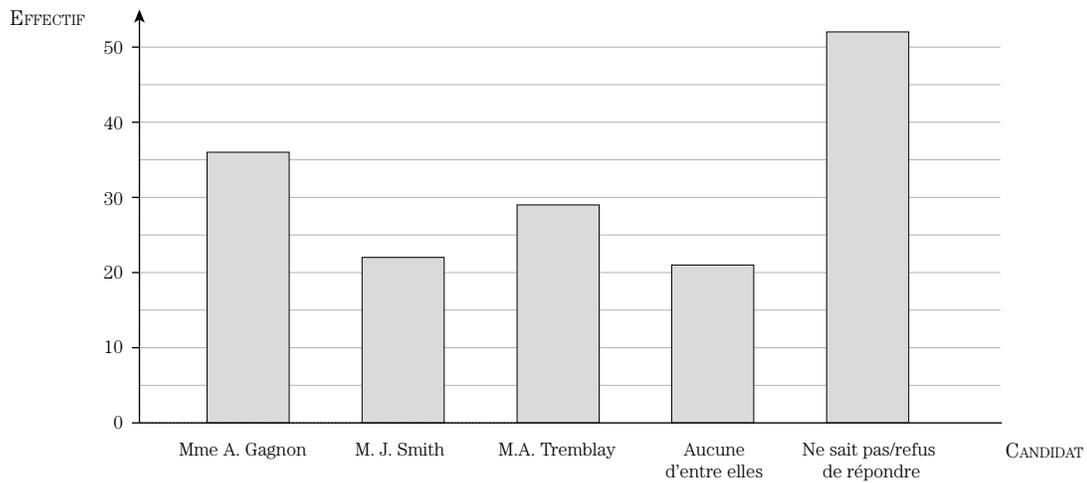
- Parmi les trois candidats, c'est en Mme Gagnon que les gens ont le plus confiance pour combattre la corruption.
- M. Smith est celui qui reçoit le moins d'appui sur cette question.
- 13 % des personnes interrogées ne font confiance à aucun des trois candidats et 33 % ne savent pas ou ont refusé de répondre à la question. Il s'agit là d'un taux élevé de non-répondants qui permet de douter de la crédibilité du sondage.

En jetant un regard rapide sur le diagramme précédent, on a l'impression que Mme Gagnon reçoit un appui quatre fois plus grand que M. Smith. Ce qui est faux, évidemment. Le problème est la graduation de l'axe vertical. Le minimum de cet axe devrait correspondre à 0 % et non à 20 %. De plus, le titre du diagramme devrait mentionner qu'il s'agit de fréquences après répartition, afin que l'électeur en soit informé.

Le diagramme ci-dessous inclut, cette fois, toutes les informations. Bien que ce diagramme soit moins spectaculaire, il correspond davantage à la réalité. C'est un modèle de diagramme que vous auriez pu construire à la page précédente.

DIAGRAMME 1.2

**LA PERSONNE LA PLUS APTE À COMBATTRE LA CORRUPTION
SELON UN SONDAGE (TAILLE DE L'ÉCHANTILLON : 160)**

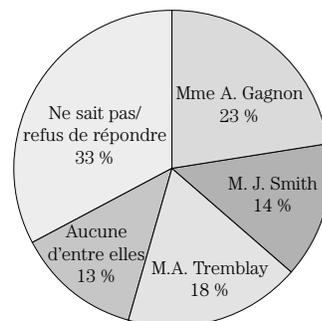


Notez que le titre du diagramme est plus précis, puisqu'il indique qu'il s'agit d'un sondage et révèle la taille de l'échantillon. Notez également que les modalités « ne sait pas » et « refus de répondre » ont été regroupées en une seule classe; ce qui est généralement le cas dans la présentation d'un sondage, puisqu'il est parfois difficile de distinguer ces catégories. Enfin, l'axe vertical représente, dans ce cas-ci, les effectifs, plutôt que les fréquences.

Attention !

Le diagramme à bandes peut servir à représenter autant les effectifs que les fréquences. Par contre, si l'on veut mettre l'accent sur la répartition des différentes modalités en pourcentage, le diagramme circulaire est souvent la meilleure option. Le choix du diagramme repose habituellement sur l'intention de l'auteur.

**LA PERSONNE LA PLUS APTE À COMBATTRE
LA CORRUPTION SELON UN SONDAGE**



À retenir

Les tableaux à données condensées

Lorsqu'une étude statistique porte sur un **caractère qualitatif** ou **quantitatif discret**, il est possible de représenter les données recueillies à l'aide d'un tableau à données condensées, à condition que le nombre de modalités ou de valeurs que peut prendre le caractère étudié soit suffisamment restreint.

Ce type de tableau indique l'effectif (et, au besoin, la fréquence) associé à chaque modalité ou valeur. Son titre doit être révélateur du caractère étudié.

Exemple : Une étude porte sur l'âge des enfants dans un camp de vacances. Le tableau de distribution ci-contre présente les données recueillies.

Ce tableau permet de constater, entre autres choses, qu'il y a six enfants de 8 ans. La somme des effectifs étant de 75, la fréquence de cet âge dans le camp de vacances est de 8 %, puisque $\frac{6}{75} = 0,08$.

ÂGE DES ENFANTS DANS UN CAMP DE VACANCES

ÂGE (a)	EFFECTIF	FRÉQUENCE (%)
8	6	8
9	16	21
10	25	33
11	20	27
12	8	11
TOTAL	75	100

La représentation graphique des données condensées

Deux types de diagramme sont utilisés pour représenter graphiquement les données recueillies dans une étude statistique portant sur un caractère qualitatif ou quantitatif discret. Le diagramme à bandes vise surtout à comparer les effectifs des différentes modalités ou valeurs alors que le diagramme circulaire vise surtout à comparer leurs fréquences.

DIAGRAMME À BANDES

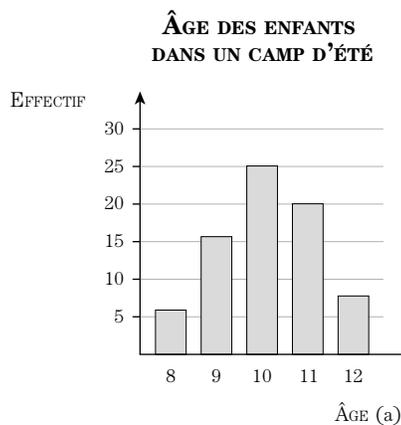
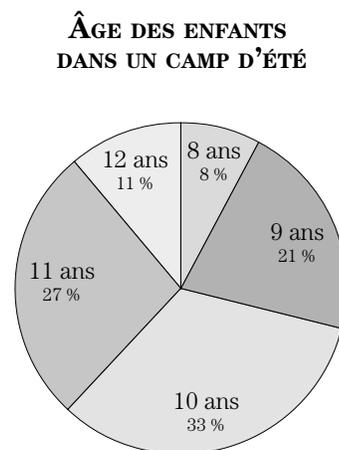


DIAGRAMME CIRCULAIRE



Des sources de biais

Pour que les résultats d'un sondage puissent être interprétés correctement, il ne suffit pas que l'échantillon choisi soit représentatif de la population visée. Des sources de biais, qui peuvent survenir à différentes phases du sondage, doivent être considérées. En voici quelques exemples :

Lors de la collecte et du traitement des données

SOURCE DE BIAIS	EXEMPLE
Une question mal formulée	Une question qui ne serait pas comprise de la même façon par tous ou une question à laquelle certaines personnes ne pourraient pas répondre étant donné les choix proposés.
Un taux de participation trop faible	Un sondage où 40 % des personnes interrogées n'ont pas d'opinion ou refusent de répondre.
Un traitement subjectif des données	Certaines données sont rejetées parce qu'elles ne correspondent pas à ce qu'on voudrait démontrer.
Des erreurs de compilation ou de transposition	Oubli de certaines données lors de la compilation.

Lors de la communication des résultats

SOURCE DE BIAIS	EXEMPLE
Un diagramme qui ne respecte pas les conventions mathématiques	Un diagramme à bandes sans titre ou avec un titre incomplet, dont les axes sont mal gradués ou mal définis, ou encore avec des bandes de largeurs non uniformes, est susceptible d'être interprété de différentes façons.
Un diagramme qui tend à favoriser un point de vue	Des artifices graphiques (choix de couleur ou effet 3D) ou une graduation particulière peuvent donner l'impression qu'une certaine modalité est beaucoup plus fréquente que d'autres, alors que ce n'est pas le cas.
Un traitement inapproprié des indécis ou des non-répondants dans un sondage d'opinion	Dans la présentation des sondages, il est courant d'exclure les personnes qui refusent de répondre ou disent ne pas avoir d'opinion. Cette façon de procéder est légitime, à condition que la répartition des opinions de ces répondants, nommée « discrets », soit semblable à celle des autres personnes de l'échantillon. Dans le cas contraire, l'exclusion de ces modalités peut entraîner un biais important.

Exercices de l'activité 1.2

1.10 Mathilde a posé une question dans un forum sur Internet, fréquenté surtout par des adolescentes.

Combien d'enfants ?	
<p>Primevère Québec, Canada</p> 	<p>Hier, nous avons parlé de ça mes deux copines et moi. Toutes les deux ne veulent pas du tout avoir d'enfants, alors que moi j'en voudrais au moins quatre. C'est vrai que ça ne marche pas très bien entre leurs parents..., mais je me demande tout de même si je suis anormale.</p> <p>Et vous, les filles, combien d'enfants aimeriez-vous avoir plus tard ?</p>

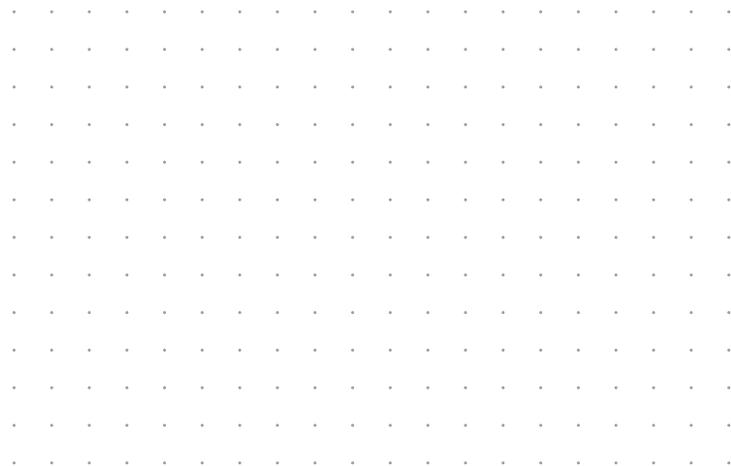
Elle a reçu près de 50 commentaires, dont 28 réponses quant au nombre d'enfants désirés.

1	3	2	4	5	0	2	3	4	4	2	3	2	2
4	3	2	1	4	2	3	2	2	3	3	1	2	2

a) Construisez un tableau à données condensées décrivant les effectifs de cette distribution, puis représentez ces effectifs au moyen d'un diagramme approprié.

Titre : _____

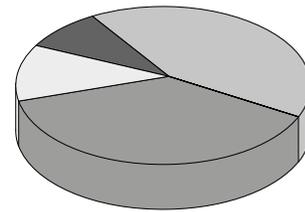
Votre diagramme



b) Indiquez au moins une source de biais dans la collecte de données de ce sondage.

1.11 Sur un site Internet portant sur l'alimentation, on peut calculer son indice de masse corporelle (IMC) en donnant sa masse (kg) et sa taille (m). Un diagramme apparaît ensuite pour décrire la répartition de toutes les personnes qui ont donné leurs mesures.

RÉPARTITION DE L'IMC DE NOS RÉPONDANTS



■ Maigreur □ Poids normal ■ Embonpoint □ Obésité

a) Quelle est la principale source de biais de ce diagramme ?

b) Pourquoi ne peut-on pas se fier à ce diagramme pour décrire l'IMC de la population en général ?

Le saviez-vous ?



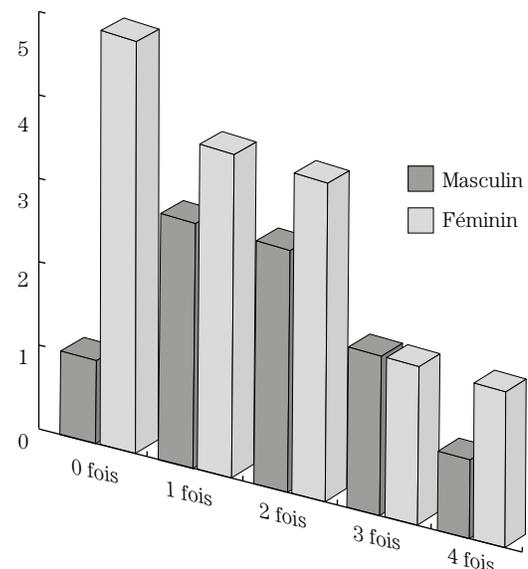
L'IMC est calculé à l'aide de la formule suivante :

$$IMC = \frac{M}{T^2}, \text{ où } M \text{ est la masse (kg) et } T, \text{ la taille (m).}$$

Un poids normal correspond à un IMC se situant entre 18,5 et 25.



1.12 Sylvio a demandé à chacun des adultes de son cours de mathématiques combien de fois, durant la dernière semaine, ils avaient fait au moins une demi-heure d'activité physique intense dans la journée. Par la suite, à l'aide d'un tableur, il a construit le diagramme à bandes multiples ci-contre.



a) Faites une critique de son diagramme.

b) Selon les données recueillies par Sylvio, les élèves de sexe féminin sont plus actives physiquement que ceux de sexe masculin. Vrai ou faux ? Justifiez votre réponse.



Activité 1.3 Quelques centimètres de plus

But ➔

- Construire et interpréter des tableaux à données groupées.
- Construire et interpréter des histogrammes.

Dans un magazine, un article présente quelques cas de couples célèbres où la femme est plus grande que l'homme et l'on s'interroge sur le vécu de ces grandes femmes. Dans un encadré qui accompagne l'article, des statistiques sont données sur la taille des femmes et des hommes au Québec.

TABLEAU 1.6

RÉPARTITION EN POURCENTAGE DES FEMMES ET DES HOMMES ADULTES SELON LEUR TAILLE AU QUÉBEC

TAILLE (cm)	FEMMES	HOMMES
[140, 150[3 %	–
[150, 160[31 %	1 %
[160, 170[50 %	15 %
[170, 180[15 %	52 %
[180, 190[1 %	30 %
[190, 200[–	2 %
TOTAL	100 %	100 %

Source : Institut de la statistique du Québec



Expliquez en vos mots ce que signifie l'expression [140, 150[dans la première colonne du tableau ci-dessus. _____

Les six intervalles inscrits dans la première colonne représentent autant de classes dans lesquelles la distribution des tailles est partagée.

Dans quelle classe votre propre taille se situe-t-elle? _____

Dans quelle classe se situe la taille d'une personne qui mesure 170 cm? _____

Que pouvez-vous dire de la taille de la majorité des femmes? Et qu'en est-il pour les hommes?

Attention !

Cette étude ne tient pas compte des personnes qui mesurent plus de 2 m et moins de 1,4 m, puisqu'elles représentent un pourcentage négligeable par rapport à l'ensemble de la population du Québec. En statistique, il est parfois nécessaire d'exclure certaines données exceptionnelles, trop éloignées de la norme, afin de mieux faire ressortir des tendances. Malheureusement, cela peut poser un problème pour certaines personnes qui se voient ainsi tenues à l'écart des statistiques.

Les données groupées en classes

Dans cette étude statistique sur la taille des femmes et des hommes, le caractère étudié est du type quantitatif continu. En effet, la taille d'un individu est une mesure qui peut prendre une infinité de valeurs, pas nécessairement entières.

Même si l'on arrondissait les données au centimètre près, le nombre de valeurs possibles resterait très grand. Il est donc inutile, voire impossible, de les représenter dans un tableau à données condensées. Pour avoir un meilleur portrait de la situation, il est préférable de regrouper ces données à l'intérieur d'intervalles qu'on appelle des classes.

Rappel

Un intervalle est un ensemble formé de tous les nombres qui sont situés entre deux bornes. Les bornes peuvent faire partie ou non de l'intervalle.

Exemple : La notation abrégée $[140, 150[$ représente l'intervalle formé de tous les nombres entre 140 et 150. Le crochet tourné vers l'intérieur devant la borne inférieure 140 indique que ce nombre fait partie de l'intervalle. Au contraire, le crochet tourné vers l'extérieur après la borne supérieure 150 indique que ce nombre ne fait pas partie de l'intervalle.

Dans le tableau 1.6 de la page précédente, la classe définie par l'intervalle $[140, 150[$ contient toutes les tailles qui sont supérieures ou égales à 140 cm et inférieures à 150 cm. Insistons sur ce point : la borne inférieure est incluse, alors que la borne supérieure est exclue. Le même principe s'applique aux autres classes. Il en est généralement ainsi dans tous les tableaux à données groupées en classes.

On peut facilement déterminer dans quelle classe se situe une donnée quelconque. Par exemple, une taille de 162 cm fait partie de la classe $[160, 170[$, car $160 \leq 162 < 170$. Par contre, une taille de 170 cm est plutôt située dans la classe suivante $[170, 180[$.

Par ailleurs, le tableau 1.6 permet d'affirmer que la majorité des femmes mesure de 160 à 170 cm. Tandis que pour les hommes, cette majorité se situe de 170 à 180 cm.

Observons quelques-unes des caractéristiques des **tableaux à données groupées en classes**.

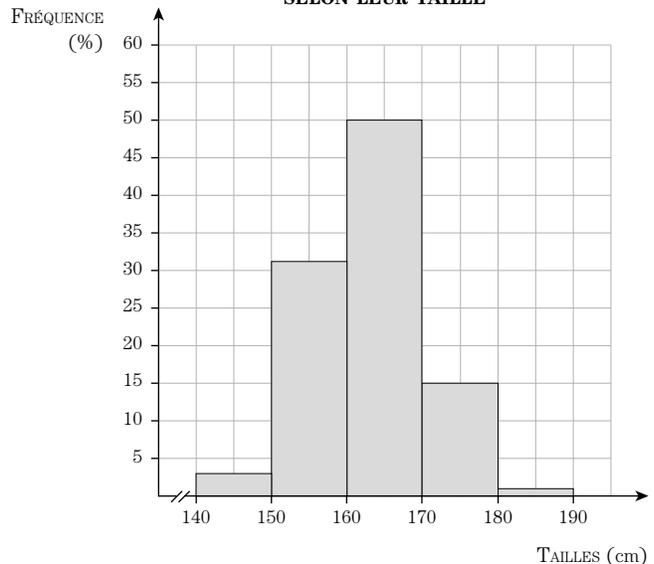
- L'amplitude des classes, soit la différence entre la borne supérieure et la borne inférieure, est constante. Dans l'exemple précédent, cette amplitude est de 10 cm. Lorsque l'on partage un ensemble de données en classes, il est préférable d'utiliser des classes de même amplitude. Cela permet de mieux les comparer entre elles.
- Les classes inscrites dans le tableau sont mutuellement exclusives. Une donnée ne peut faire partie que d'une seule classe.
- Enfin, les classes couvrent toutes les valeurs possibles.

La représentation graphique des données groupées en classes

Il existe une façon particulière de représenter graphiquement les effectifs ou les fréquences des données groupées en classes. Par convention, on utilise toujours un histogramme. Voici l'histogramme représentant la répartition des femmes du Québec selon leur taille.

DIAGRAMME 1.3

RÉPARTITION DES FEMMES DU QUÉBEC SELON LEUR TAILLE



Cet histogramme ressemble à un diagramme à bandes, mais il est différent sur deux points essentiels. Quels sont ces deux points ?

1 : _____

2 : _____

Complétez cet histogramme de la répartition des hommes du Québec selon leur taille.

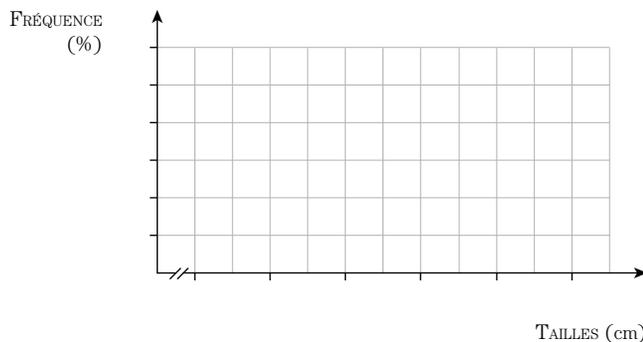
TABLEAU 1.7

RÉPARTITION DES HOMMES DU QUÉBEC SELON LEUR TAILLE

TAILLE (cm)	HOMMES
[150, 160[1 %
[160, 170[15 %
[170, 180[52 %
[180, 190[30 %
[190, 200[2 %

DIAGRAMME 1.4

TITRE _____



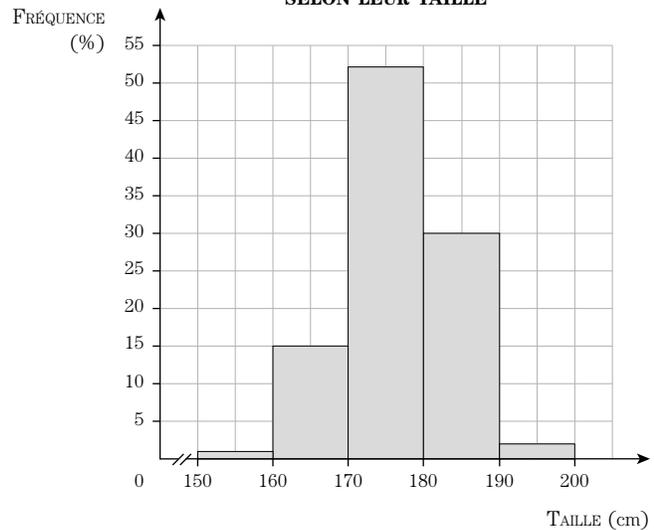
L'**histogramme** ressemble au diagramme à bandes de par ses bandes toutes de même largeur. Cependant, il possède deux différences essentielles.

- Dans un histogramme, les bandes rectangulaires dont les hauteurs indiquent les effectifs ou les fréquences sont **juxtaposées**. Il n'y a pas d'espace entre elles, tandis que dans le diagramme à bandes, elles sont séparées les unes des autres par un espace équidistant.
- Dans un histogramme, les données sont de nature quantitative continue. Ainsi, la graduation de l'axe horizontal contient les bornes de toutes les classes. La graduation va de la borne inférieure de la première classe jusqu'à la borne supérieure de la dernière classe. La base du rectangle associé à chaque classe commence à la borne inférieure de la classe et se termine à sa borne supérieure.

Attention !

Si la première classe ne débute pas à 0, il est important de faire une coupure de l'axe horizontal. Aussi, puisque toutes les classes devraient avoir la même amplitude, chacune des bandes de l'histogramme devrait être de la même largeur.

Voici l'histogramme que vous deviez construire quant à la répartition des hommes.

DIAGRAMME 1.5**RÉPARTITION DES HOMMES DU QUÉBEC
SELON LEUR TAILLE****Le choix du nombre de classes et de leur amplitude**

Lorsque vous devez construire un tableau à données groupées en classes, il vous faut d'abord déterminer le nombre de classes et leur amplitude.

Pour le nombre de classes, on utilise généralement de 5 à 10 classes. Parfois, ça peut aller jusqu'à une quinzaine de classes, s'il y a un très grand nombre de données, mais c'est plutôt exceptionnel.

Pour l'amplitude, on doit tenir compte du nombre de classes souhaité et de l'étendue de la distribution.

Rappel

L'étendue d'une distribution correspond à la différence entre la valeur maximale et la valeur minimale des données recueillies.

Pour que toutes les valeurs possibles de la distribution soient couvertes, il faut appliquer cette règle :

$$\frac{\text{Étendue}}{\text{Nombre de classes}} < \text{Amplitude}$$

De plus, par convention, on utilise généralement des amplitudes qui sont des nombres entiers arrondis. Ainsi, on préférera une amplitude de 20 plutôt que de 16, de 300 plutôt que de 265.

Voici un exemple qui vous aidera à comprendre l'organisation des classes.

Dans un hôpital, on a noté la masse à la naissance des 25 derniers nouveau-nés. Les données recueillies, exprimées en grammes, sont les suivantes :

3155, 3185, 2150, 2255, 3010, 3030, 2770, 2890, 2895, 3035, 3560, 3740, 2895,
3220, 3230, 3275, 3405, 3885, 3105, 3130, 2315, 2580, 2590, 2745, 3530.

Réorganisez ces données de façon à ce quelles soient toutes ordonnées.

Quelle est la valeur minimale? _____

Quelle est la valeur maximale? _____

Quelle est l'étendue de cette distribution? _____

Vous avez sans doute obtenu les mesures suivantes :

La masse minimale : 2150 g La masse maximale : 3885 g L'étendue : 1735 g



À partir des données fournies par l'hôpital, proposez une façon de subdiviser cette distribution en 5 classes de même amplitude.

.....

.....

.....

.....

.....

Il existe plusieurs façons de répartir cette distribution en classes. En voici un exemple:

- En divisant l'étendue par 5 (le nombre souhaité de classes), on obtient : $1735 \text{ g} \div 5 = 347 \text{ g}$. Ce qui requiert que l'amplitude des classes doit être supérieure à 347 g.
- Puisqu'il est toujours préférable d'avoir des classes avec des valeurs arrondis, on peut, par exemple, arrondir cette amplitude minimale de 347 g à la centaine supérieure et obtenir ainsi une amplitude de 400 g pour chacune des classes.

Attention !

Si l'on arrondissait l'amplitude vers le bas à 300 g, il faudrait six classes et non cinq.

Car $5 \times 300 \text{ g} < 1735 \text{ g}$, et en conséquence, cinq classes ne couvriraient pas toute l'étendue de la distribution.

Le choix de la première borne

Pour compléter la définition des classes dans cet exemple, il ne reste plus qu'à déterminer la borne inférieure de la première classe.

Sachant qu'une amplitude de 400 g est suggérée, proposez les intervalles possibles pour ces 5 classes :

MASSE (g)
[_____ , _____ [
[_____ , _____ [
[_____ , _____ [
[_____ , _____ [
[_____ , _____ [



Plusieurs valeurs sont possibles pour la première borne, mais on privilégiera un nombre entier arrondi de la même façon que pour l'amplitude.

Dans ce cas, 2000 semble une bonne valeur pour la borne inférieure du premier intervalle. D'autres suggestions auraient pu être retenues, mais nous fixerons cette première borne pour la suite des choses.

Le tableau ci-contre montre toutes les classes qui en résultent.

Complétez ce tableau à données groupées en déterminant l'effectif de chaque classe.

Représentez les données de ce tableau par un histogramme.

TABLEAU 1.8

RÉPARTITION DE 25 NOUVEAU-NÉS SELON LEUR MASSE

MASSE (g)	EFFECTIF
[2000, 2400[
[2400, 2800[
[2800, 3200[
[3200, 3600[
[3600, 4000[
TOTAL	



Comparez maintenant vos représentations avec le tableau et l'histogramme de la section À retenir suivante.

Les tableaux à données groupées en classes

Dans une étude statistique, il est souvent approprié de regrouper les données en classes, notamment quand l'étude porte sur un **caractère quantitatif continu**, ou lorsque les données sont du type quantitatif discret, mais qu'elles n'ont pas tendance à se répéter et prennent plusieurs valeurs différentes.

Les données recueillies sont compilées dans un tableau à données groupées en classes qui indique l'effectif ou la fréquence de chaque classe.

Les classes présentent généralement les propriétés suivantes :

- Chaque classe est définie par un intervalle de la forme $[a, b[$.
- L'amplitude de chaque classe (soit la valeur $b - a$) est la même pour toutes les classes.
- Les classes sont mutuellement exclusives et couvrent toutes les valeurs possibles du caractère étudié. La borne supérieure d'une classe est égale à la borne inférieure de la classe suivante.
- Le nombre de classes se situe généralement entre 5 et 10.

**RÉPARTITION DE 25 NOUVEAU-NÉS
SELON LEUR MASSE**

MASSE (g)	EFFECTIF
[2000, 2400[3
[2400, 2800[4
[2800, 3200[10
[3200, 3600[6
[3600, 4000[2
TOTAL	25

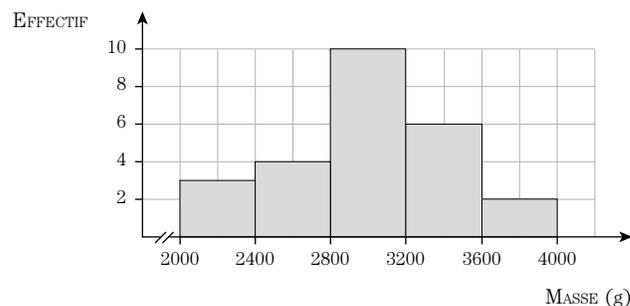
Les histogrammes

On utilise l'histogramme pour représenter graphiquement les effectifs ou les fréquences lorsque les données sont groupées en classes.

Dans un histogramme :

- Les boîtes rectangulaires, dont la hauteur indique l'effectif ou la fréquence, sont toujours juxtaposées.
- La graduation de l'axe horizontal marque le début et la fin de chaque classe.

**RÉPARTITION DE 25 NOUVEAUX-NÉS
SELON LEUR MASSE**



Exercices de l'activité 1.3

- 1.13) On a demandé à 40 personnes de déterminer, à la minute près, la durée du trajet en automobile qu'elles font pour se rendre au travail le matin. Voici les résultats en minutes.

49	16	27	57	15	12	39	46	55	29
25	17	45	58	37	35	44	64	45	25
28	40	35	53	49	65	30	46	32	15
46	52	30	38	42	28	25	20	46	36

- a) Quelle est l'étendue de cette distribution ? _____
- b) Pour grouper ces données en 6 classes, quelle devrait être l'amplitude des classes ? _____
- c) Complétez le tableau ci-dessous.

RÉPARTITION DES RÉPONDANTS SELON LA DURÉE DU TRAJET EN AUTOMOBILE POUR SE RENDRE AU TRAVAIL

DURÉE (min)	EFFECTIF	FRÉQUENCE (%)
[10, _____ [
TOTAL		



- d) Représentez la fréquence associée à chacune des classes à l'aide d'un histogramme.



1.14) Voici le résultat de la qualification au saut en longueur aux Jeux olympiques de Londres 2012.

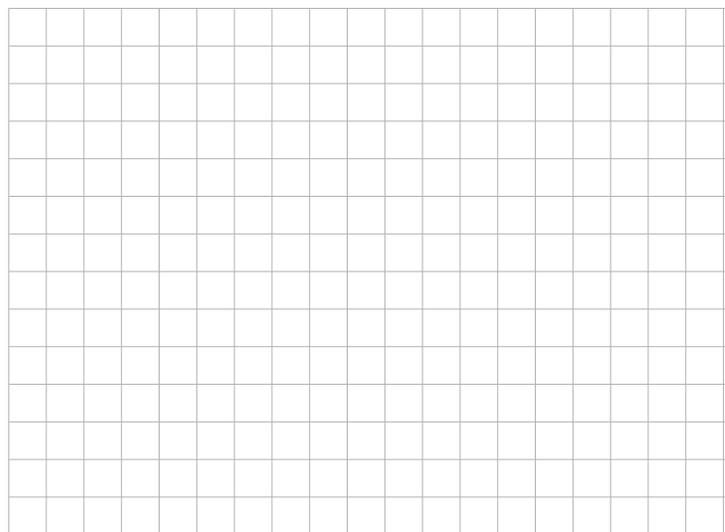
RANG	ATHLÈTES	SAUT (m)	RANG	ATHLÈTES	S AUT (m)
1	Da Silva	8,11	21	Higgs	7,76
1	Goodwin	8,11	22	Camara	7,72
3	Menkov	8,09	23	Sdiri	7,71
4	Rutherford	8,08	24	Badji	7,66
5	Tomlinson	8,06	25	Sargsyan	7,62
6	Torneus	8,03	26	Mykolaitis	7,61
7	Mokoena	8,02	27	Gbagbeke	7,59
8	Claye	7,99	28	Chuva	7,55
8	Watt	7,99	29	Tsatoumas	7,53
10	Smith	7,97	30	Wagner	7,50
11	Frayne	7,95	30	Kuznetsov	7,50
12	Bayer	7,92	32	Rivera	7,42
12	Reif	7,92	33	Lin	7,38
12	Caceres	7,92	33	Sukhasvasti	7,38
15	Petrov	7,89	35	Skhirtladze	7,26
16	Morgunov	7,87	36	Zhang	7,25
17	Arzandeh	7,84	37	Difallah	7,08
18	Gaisah	7,79	38	Novotny	6,96
18	Forbes	7,79	39	Kitchens	6,84
20	Li	7,77	40	Pahlevanyan	6,55

a) Complétez le tableau ci-dessous.

b) Tracez l'histogramme qui correspond à cette répartition.

**RÉPARTITION DES ATHLÈTES
SELON LA LONGUEUR DE LEUR
SAUT AUX JEUX OLYMPIQUES
DE LONDRES 2012**

SAUT (m)	EFFECTIF
[6,5; _____ [
[_____ ; 8,3[
TOTAL	



Un peu plus loin – Méthodes d'échantillonnage non probabiliste

Les méthodes d'échantillonnage probabiliste sont très contraignantes. Il n'est pas toujours possible d'obéir parfaitement à ces contraintes parce que les données ne sont pas disponibles, parce que ça coûterait trop cher ou que ça prendrait trop de temps. D'autres méthodes d'échantillonnage sont alors utilisées.

L'échantillonnage de commodité

Lorsque la population a une homogénéité très grande, presque parfaite, n'importe quel échantillon a de fortes chances d'être représentatif, même s'il n'est pas choisi au hasard.

Voici un exemple :

Les diabétiques doivent mesurer régulièrement le taux de sucre dans leur sang à l'aide d'un lecteur de glycémie. Pour ce faire, ils prennent un échantillon, au bout de leur doigt, qui est constitué d'une gouttelette de sang. Le diabétique n'a pas à se demander si l'échantillon est représentatif de tout son sang, car le taux de sucre a tendance à être le même dans tout le corps.

Prendre une gouttelette de sang au bout du doigt est facile et rapide à réaliser. C'est pourquoi on parle ici d'un échantillonnage de commodité. On dit aussi « échantillonnage accidentel ». On utilise alors le terme accidentel, dans le sens de fortuit, de ce qui arrive par le hasard de la vie.

C'est ce type d'échantillonnage qu'un journaliste de la télévision utilise lorsqu'il va chercher l'opinion de la rue en interrogeant quelques passants. Le journaliste sélectionne des piétons comme le diabétique sélectionne la gouttelette de son sang, par commodité, sans se préoccuper de faire un choix au hasard.

Il y a un certain hasard dans le fait qu'une personne puisse se trouver à passer justement à cet endroit, au moment où le journaliste fait son sondage. Mais c'est un hasard accidentel et difficilement mesurable. L'échantillon sélectionné ainsi n'est sûrement pas représentatif de l'ensemble de la population, car la plupart des personnes ne peuvent pas être choisies dans cet échantillon : parce qu'elles ne passent jamais à cet endroit, parce qu'elles travaillent à ce moment-là, parce qu'elles sont malades, ou encore parce qu'elles se déplacent en auto, en bicyclette, etc. Ces gens exclus auraient peut-être une opinion très différente de celle des passants interrogés.

Le risque de biais est donc très grand, mais c'est une méthode d'échantillonnage économique et rapide, ce qui explique que plusieurs recherches la préfèrent à d'autres méthodes.



1.15) Donnez au moins un autre exemple de situation de la vie où l'on utilise la méthode d'échantillonnage par commodité. Indiquez si, dans ce cas, l'échantillon est représentatif et expliquez pourquoi.

L'échantillonnage volontaire

- Vous regardez une émission de variétés à la télévision. On vous invite à voter pour votre artiste préféré en appelant à tel numéro.
- Vous consultez le site Internet d'un journal. Pourquoi ne pas répondre à la question du jour?
- Au centre commercial, on vous offre un cadeau. Vous devez répondre à un sondage sur un nouveau produit.

Toutes ces situations sont caractérisées par le fait que ce n'est pas le sondeur qui construit l'échantillon, c'est plutôt l'échantillon qui se construit de lui-même. En effet, ce sont les individus de la population qui décident volontairement de faire partie de l'échantillon.

Notez que cette méthode peut être très utile dans certains cas, par exemple pour des tests pharmaceutiques.

L'échantillonnage volontaire est probablement la méthode la plus simple à appliquer lors de sondages et sûrement l'une des plus rapides, mais elle donne des résultats qui présentent un risque élevé de biais.



1.16) Donnez au moins deux risques de biais qu'entraîne cette méthode d'échantillonnage.

- _____

- _____

L'échantillonnage par quotas

Cette méthode d'échantillonnage est très souvent utilisée par les sondeurs. Elle ressemble à l'échantillonnage stratifié, mais elle en diffère sur un point essentiel.

Voici un exemple :

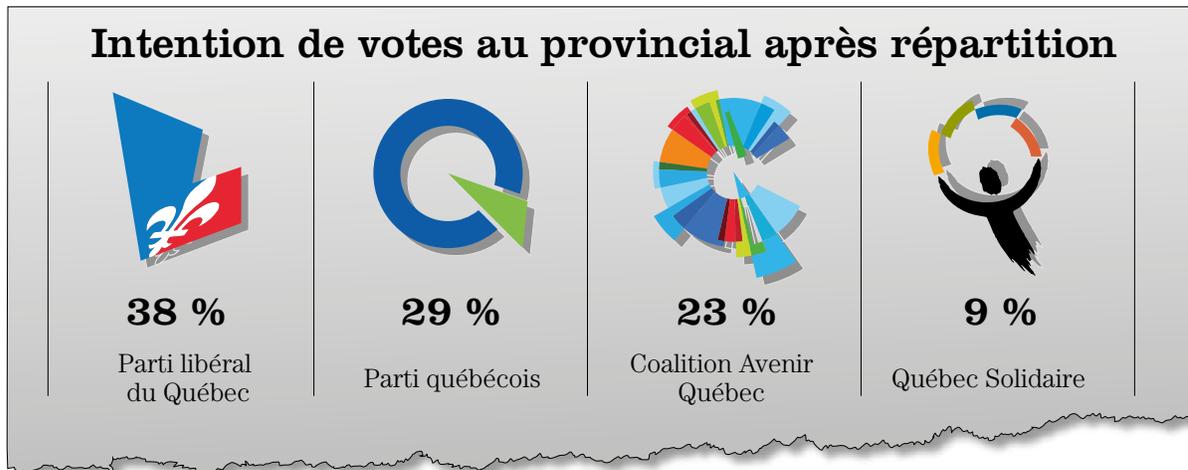
Dans un certain cégep, 349 des 415 personnes inscrites au programme de formation en soins infirmiers sont de sexe féminin. On veut faire un sondage pour déterminer quelle proportion de cette population a l'intention de poursuivre sa formation à l'université. Pour ce faire, on demande à l'enquêteur d'interroger exactement 84 élèves, de sexe féminin, et 16 élèves, de sexe masculin, en se promenant dans le pavillon réservé à ce programme. Une fois le quota atteint pour l'une des catégories, l'enquêteur devra poursuivre son travail en interrogeant seulement des élèves de l'autre catégorie jusqu'à ce qu'il atteigne le quota pour celle-ci également.



1.17) En quoi l'échantillonnage par quotas diffère-t-il de l'échantillonnage stratifié ?

1.18) Donnez un exemple de biais que peut entraîner cette méthode d'échantillonnage.

1.21 Un sondage Internet publié dans un journal quelques jours avant l'élection provinciale de 2014 donnait les résultats suivants après répartition proportionnelle des indécis.



Source : Sondage Leger marketing - Journal de Montréal, 5 août 2014

Le sondage avait été réalisé auprès de 1220 personnes du grand public dans toutes les régions du Québec et tenait compte de l'âge, du sexe et de la langue maternelle des répondants. Le sondage précisait qu'environ 158 personnes de ces gens interrogés souhaitaient voter pour un autre parti (2 %), annuleraient leur vote (3 %), ne voteraient pas (2 %), ne savaient pas quoi répondre (2 %) ou encore refusaient de répondre au sondage (4 %). De plus, le sondage mentionnait qu'un électeur sur quatre disait pouvoir encore changer d'idée d'ici au jour du scrutin.

Les deux tableaux suivants décrivent les résultats officiels de cette élection.

**RÉPARTITION DES VOTES
ÉLECTION PROVINCIALE 2014**

PARTI	NOMBRE DE VOTES	POURCENTAGE DES VOTES
Parti libéral du Québec	1 756 373	
Parti Québécois	1 074 129	
Coalition Avenir Québec	975 742	
Québec solidaire	323 263	
autres partis	102 638	
TOTAL DES VOTES EXPRIMÉS		

**RÉPARTITION DES SIÈGES
ÉLECTION PROVINCIALE 2014**

PARTI	NOMBRE DE DÉPUTÉS ÉLUS	POURCENTAGE DE REPRÉSENTATION QUANT AU NOMBRE DE SIÈGES
Parti libéral du Québec	70	
Parti Québécois	30	
Coalition Avenir Québec	22	
Québec solidaire	3	
autres partis	0	
TOTAL		

Source : Le directeur général des élections du Québec

a) Complétez ces deux tableaux de distribution.

- b) Construisez un diagramme à bandes multiples qui permettrait de comparer aisément les résultats du sondage, ceux de la répartition des votes et ceux de la répartition des sièges.



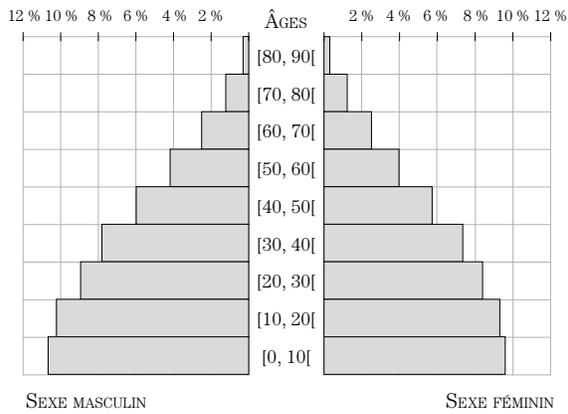
- c) Selon vous, qu'est-ce qui pourrait expliquer les différences entre les résultats du sondage et ceux de l'élection ?

- d) Selon vous, qu'est-ce qui pourrait expliquer les différences entre les résultats quant à la répartition des votes et ceux quant aux nombres de circonscriptions gagnées par chacun des partis ?

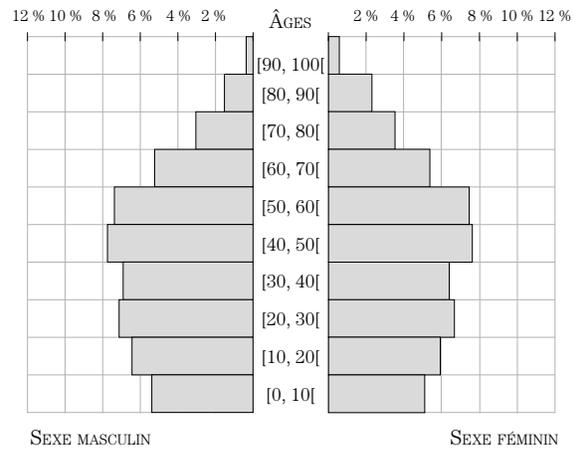
1.22 Le diagramme appelé «pyramide des âges» permet de décrire la répartition d’une population en fonction des sexes et des âges. Il est composé de deux histogrammes dont les âges sont portés horizontalement. Celui de gauche représentant, par convention, la population de sexe masculin et celui de droite, celle de sexe féminin. Les classes ont généralement une amplitude de 5 ou 10 ans.

Voici deux exemples de pyramide des âges : celle de l’Inde et celle du Canada

RÉPARTITION DE LA POPULATION EN INDE (2010)



RÉPARTITION DE LA POPULATION EN CANADA (2010)



Comparez ces deux pays en ce qui concerne :

a) la répartition de leur population selon les âges.

b) la répartition de leur population selon les sexes.

Le saviez-vous ?



En Inde, la naissance d’une fille peut être vue comme une catastrophe pour une famille, puisque ce sont habituellement les garçons qui prennent soin de leurs parents lorsque ces derniers sont âgés. De plus, la culture traditionnelle indienne exige qu’une jeune fille fournisse une dot pour se marier. Cette situation explique la présence d’avortements sélectifs en Inde, et ce, malgré une loi qui interdit aux médecins indiens de dévoiler aux parents le sexe du futur bébé. C’est pourquoi l’avortement sélectif des fœtus de sexe féminin est malheureusement très fréquent.



Activité synthèse – Questionner sur l'utilisation d'Internet

Vous êtes maintenant en mesure d'accomplir la tâche décrite au début de cette situation d'apprentissage. Rappelons d'abord le contexte.

La directrice d'un centre de formation voudrait connaître à quelle fréquence et de quelle façon les adultes, qui étudient dans son centre, utilisent Internet. Elle vous demande de réaliser un sondage pour répondre à ces deux questions.

- 1) Pour quel type d'activités utilisez-vous principalement Internet ?
- 2) Combien d'heures par semaine naviguez-vous sur Internet à des fins personnelles (autres que pour votre travail ou vos études) ?

Votre tâche ➔

À partir des données qui vous seront fournies et qui concernent une population d'élèves adultes, vous devrez d'abord choisir un échantillon représentatif, en tenant compte de la répartition des individus selon leur âge et leur sexe, puis collecter les données en réponse aux deux questions proposées par la directrice. Il vous faudra ensuite représenter ces données à l'aide de tableaux et de diagrammes appropriés. Vous devrez enfin interpréter ces diagrammes afin de dresser un portrait de la situation.

La page suivante contient les réponses aux deux questions que chacun des 480 élèves du centre donnerait si on les interrogeait. Ces données sont constituées d'une lettre et d'un nombre.

La lettre indique l'utilisation principale d'Internet selon le code suivant :

- | | |
|--|---|
| C : communication (courriel, clavardage) | D : divertissement (jeux, musique, vidéo) |
| F : opérations financières (achat, banque) | R : recherche d'informations |
| X : n'utilise pas Internet | |

Le nombre indique le temps, en heures, passé par semaine sur Internet à des fins personnelles.

Vous n'avez pas à compiler les 480 réponses. Il est plutôt suggéré de vous limiter à un échantillon de 50 personnes. Comme la méthode d'échantillonnage stratifié est appropriée dans ce cas, vous devrez tenir compte du nombre d'élèves dans chaque strate, tel que décrit dans le tableau suivant :

TABLEAU 1.9

RÉPARTITION DES ADULTES DU CENTRE SELON L'ÂGE ET LE SEXE

ÂGE (a)	FEMMES	HOMMES
19 ans et moins	93	109
20 à 24 ans	57	64
25 à 29 ans	34	30
30 à 34 ans	28	22
35 ans et plus	23	20

TABLEAU 1.10



RÉPONSES QUE DONNERAIT CHAQUE ADULTE INTERROGÉ

HOMME

19 ans ou moins	R 6	C 28	R 4	D 6	D 10	D 35	R 6	C 19	C 15	D 8	C 21	R 4	C 8	D 14	C 21
	D 15	D 30	C 6	C 10	R 15	F 3	R 7	C 13	C 20	F 3	C 4	C 8	C 14	D 28	C 6
	R 3	C 30	R 7	R 14	C 5	D 32	C 3	D 21	R 2	C 3	D 21	C 5	C 3	C 7	D 9
	D 21	D 32	R 5	C 20	R 3	R 7	C 9	R 12	R 15	C 24	R 9	R 12	D 14	R 14	C 15
	D 20	D 32	D 35	C 7	R 13	F 2	R 3	C 4	R 7	C 14	D 28	R 2	C 3	D 7	R 10
	C 11	C 14	D 14	C 1	D 8	D 30	F 1	R 8	C 2	R 2	R 4	R 5	C 8	R 14	R 1
	C 14	R 1	C 2	D 32	R 3	D 7	D 9	D 12	D 16	D 21	D 20	F 1	C 2	D 7	C 9
	C 13	D 20	C 21	C 4											
20 à 24 ans	R 4	C 5	C 4	R 6	C 15	C 21	R 3	R 21	D 21	F 4	R 15	D 30	R 1	R 8	D 35
	C 2	D 9	D 8	R 7	R 15	D 25	C 3	R 3	R 12	R 4	C 7	D 14	R 7	D 32	C 2
	R 7	D 14	R 5	C 1	C 8	D 18	R 4	C 5	C 18	R 6	C 9	R 14	D 6	D 30	C 10
	D 14	F 2	R 9	D 20	R 2	R 12	C 1	C 8	C 14	C 7	R 15	R 1	R 9	C 14	R 3
D 20	C 3	D 15	F 3												
25 à 29 ans	C 7	R 8	R 7	C 6	F 3	D 20	R 2	X 0	R 21	D 25	R 14	F 4	R 15	C 3	R 7
	D 12	C 14	R 1	R 2	D 15	D 35	D 5	R 8	D 25	C 7	D 10	C 1	C 5	R 18	R 4
30 à 34 ans	R 7	C 6	R 1	F 3	R 8	D 10	R 12	C 5	C 14	D 18	R 7	C 4	R 28	C 3	D 30
	X 0	R 2	D 20	X 0	D 8	F 3	R 4								
35 ans et plus	X 0	D 30	D 21	C 2	R 7	C 5	C 4	F 1	F 3	C 6	R 14	D 10	R 2	D 6	X 0
	R 7	R 1	X 0	R 8	X 0										

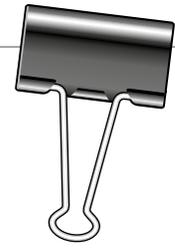
FEMME

19 ans ou moins	C 7	R 7	C 7	D 15	R 2	R 3	C 12	R 7	C 8	C 3	C 14	F 3	C 10	C 5	C 1
	C 21	R 4	C 8	R 6	D 6	C 7	R 2	R 14	C 14	R 7	D 8	C 20	R 1	C 2	D 32
	C 30	C 9	R 15	C 4	F 1	C 3	C 20	D 21	C 6	F 3	F 2	R 3	D 14	D 9	R 9
	R 5	C 6	C 5	C 28	C 4	C 4	R 3	D 8	C 4	D 30	R 12	D 28	R 3	C 30	C 15
	D 21	C 14	C 28	R 4	R 14	R 4	R 2	C 3	R 10	C 11	C 32	R 6	R 8	F 3	C 2
	R 2	D 7	D 12	C 2	R 4	R 4	R 5	R 14	R 1	F 1	C 2	D 7	C 9	C 3	D 20
C 15	C 21	C 4													
20 à 24 ans	D 14	C 10	C 3	R 9	R 15	C 14	C 7	C 12	C 14	C 15	D 30	C 1	C 5	D 9	C 22
	R 2	C 2	D 8	D 30	R 3	C 7	D 6	D 25	R 12	D 20	R 4	C 2	C 3	F 2	R 2
	C 4	R 8	R 28	F 2	R 8	R 6	R 6	R 4	D 21	C 2	R 21	C 1	R 2	C 5	R 14
	D 18	C 24	C 1	C 7	C 5	C 3	R 18	F 3	C 9	R 3	C 3	D 32			
25 à 29 ans	F 2	C 6	F 2	R 10	R 1	C 5	C 7	R 4	C 7	R 5	C 1	R 3	D 25	D 28	C 10
	C 2	C 4	C 2	R 7	R 14	C 24	C 9	C 8	R 7	R 4	C 3	D 21	C 1	D 30	C 14
	X 0	F 3	C 20	F 2											
30 à 34 ans	C 21	D 30	R 10	X 0	C 10	R 4	C 5	C 3	R 4	C 4	C 7	C 8	F 2	R 14	R 3
	R 9	D 25	F 1	C 5	C 1	C 16	F 2	X 0	R 7	R 6	C 14	R 7	R 4		
35 ans et plus	D 28	R 1	C 7	F 2	C 8	C 7	R 8	R 7	X 0	C 3	F 1	C 1	C 21		
	C 2	R 5	R 10	C 14	X 0	R 15	D 30	C 3	R 5	X 0					

Les diagrammes et votre analyse

- 1.26 Tracez les diagrammes pertinents et faites une analyse sommaire des données recueillies en un court texte qui fait ressortir les informations que vous jugez les plus intéressantes.





Liste des nouveaux savoirs

Des méthodes d'échantillonnage probabiliste

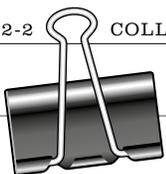
L'une des façons d'éviter les sources de biais dans le choix de l'échantillon est d'utiliser une méthode d'échantillonnage basée sur le hasard.

MÉTHODE	DESCRIPTION
Échantillonnage stratifié	Après avoir fragmenté la population en strates, on détermine le pourcentage de l'échantillon qui sera choisi dans chaque strate selon le rapport suivant : $\frac{\text{Taille de la strate dans la population}}{\text{Taille de la population}}$ Les éléments de l'échantillon sont choisis au hasard dans chaque strate par la méthode d'échantillonnage aléatoire simple.
Échantillonnage par grappes	La population est partagée en grappes. On choisit un certain nombre de grappes au hasard en tenant compte de la taille voulue de l'échantillon. Tous les éléments des grappes choisies font partie de l'échantillon.

Des sources de biais

Pour que les résultats d'un sondage puissent être interprétés correctement, il faut éviter plusieurs sources de biais, qui peuvent survenir dans différentes phases du sondage.

LORS DE LA CONSTRUCTION DE L'ÉCHANTILLON	LORS DE LA COLLECTE ET DU TRAITEMENT DES DONNÉES	LORS DE LA COMMUNICATION DES RÉSULTATS
<ul style="list-style-type: none"> • Une population mal définie • Un échantillon trop petit • Des éléments de l'échantillon qui sont liés entre eux • Un échantillon qui ne représenterait qu'une partie de la population 	<ul style="list-style-type: none"> • Une question mal formulée • Un taux de participation trop faible • Un traitement subjectif des données • Des erreurs de compilation ou de transposition • Un traitement inapproprié des indécis ou des non-répondants dans un sondage d'opinion 	<ul style="list-style-type: none"> • Un diagramme qui n'est pas conforme aux conventions mathématiques • Un diagramme qui tend à favoriser un point de vue



La représentation des données à l'aide de tableaux

TABLEAU À DONNÉES CONDENSÉES

À utiliser lorsqu'une étude statistique porte sur un caractère qualitatif ou quantitatif discret avec un nombre restreint de modalités ou de valeurs.

Exemple :

**ÂGE DES ENFANTS
DANS UN CAMP D'ÉTÉ**

ÂGE (a)	EFFECTIF	FRÉQUENCE (%)
8	6	8
9	16	21
10	25	33
11	20	27
12	8	11

TABLEAU À DONNÉES GROUPEES EN CLASSES

À utiliser lorsqu'une étude statistique porte sur un caractère quantitatif continu, ou encore pour un caractère quantitatif discret si celui-ci peut prendre plusieurs valeurs différentes.

Exemple :

**RÉPARTITION DE 25 NOUVEAU-NÉS
SELON LEUR MASSE**

MASSE (g)	EFFECTIF
[2000, 2400[3
[2400, 2800[4
[2800, 3200[10
[3200, 3600[6
[3600, 4000[2

La représentation de données à l'aide de diagrammes

Pour représenter les effectifs ou les fréquences liés à un tableau à données condensées, on utilise un diagramme à bandes ou un diagramme circulaire.

Pour représenter les effectifs ou les fréquences liés à un tableau à données groupées en classes, on utilise un histogramme.

DIAGRAMME À BANDES

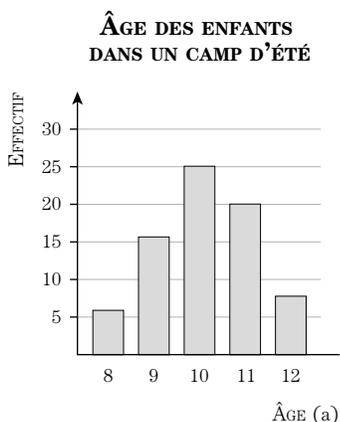
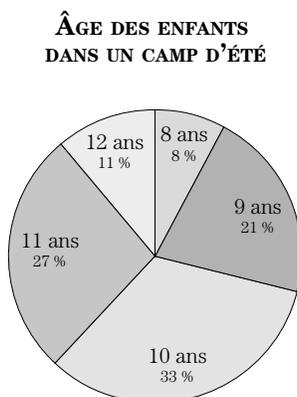


DIAGRAMME CIRCULAIRE



HISTOGRAMME

