Annexe I-1

## Introduction

Les élèves pourront aborder cette unité dans leurs temps libres. Pour les aider, l'enseignant devrait connaître le domaine de la statistique et les fonctions de la calculatrice, de même que le contenu des leçons. On leur recommande de faire l'unité avant de la remettre aux élèves.

Plusieurs exercices exigent la collecte de données dans la classe. L'enseignant peut diriger cet exercice sans y consacrer trop de temps en classe. Les données recueillies peuvent être copiées ou affichées. On trouve dans les notes à l'intention de l'enseignant des exercices pratiques supplémentaires liés à certaines leçons, ainsi que les solutions (annexe I-3).

Les premières quatre leçons abordent des sujets avec lesquels beaucoup d'élèves sont familiers, en plus de leur présenter la calculatrice TI-83.

À compter de la leçon 5, le matériel présenté est nouveau et exigera probablement plus de temps.

## Leçon 1 : Variables nominales

### Contexte

Statistique Canada rapporte que le taux de chômage a atteint 9,7 % en mai après la désaisonnalisation.

Que signifie ce pourcentage? Comment Statistique Canada a recueilli cette information? Quel est le degré de précision?

### Points importants

Tu dois voir que :

- les données sont des nombres qui appartiennent à un contexte qui intéresse les statisticiens;
- les graphiques à barres nous donnent une représentation visuelle des distributions;
- les calculatrices à affichage graphique peuvent effectuer des opérations sur les variables avant que nous analysions les données;
- nous pouvons commencer à apprécier le type de questions auquel la réflexion statistique peut répondre.

### Activités de mise en marche

1. Écris une phrase pour décrire ce que le mot « statistique » signifie pour toi. Utilise des phrases complètes, bien construites et correctes sur le plan grammatical. Ta capacité à communiquer tes idées constitue une partie aussi importante de ton travail que ta capacité à manipuler des données.
2. Inscris tes réponses aux quatre questions suivantes sur la page « Enquête numéro un », que ton enseignant distribuera.
   - Quel est ton sexe?
   - Lequel parmi les trois mots suivants décrit le mieux ton allégeance politique : libérale, modérée ou conservatrice?
   - Le Canada devrait-il abolir la pièce de 1 cent?
   - Où places-tu la statistique quant à sa valeur pour la société sur une échelle de 1 (totalement inutile) à 9 (extrêmement importante)?
3. Combien de mots contient la phrase que tu as écrite en réponse à la question 1?
4. Inscris le nombre de lettres que contient chacun des mots de ta réponse à la question 1.

Tu as recueilli des **données**. Les données sont toujours entourées d'un contexte particulier. Sans ce contexte, ce ne sont que des chiffres. Tu dois toujours te reporter au contexte dans lequel sont inscrites les données pour décrire un problème ou une expérience.

Les caractéristiques sur lesquelles tu as recueilli des données sont des **variables**. Certaines variables peuvent avoir toute une gamme de valeurs numériques, et sont à ce titre appelées des **variables de mesure**. Parallèlement, les variables nominales servent à enregistrer des désignations de catégorie. Les variables nominales **binaires** peuvent appartenir à deux catégories seulement. Différentes méthodes statistiques sont utilisées selon le type de variable mesuré.

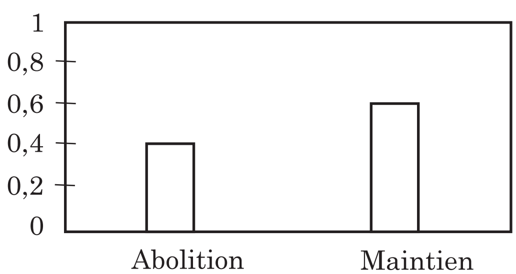
## Leçon 1 : Variables nominales (suite)

### Développement

1. Indique s'il s'agit d'une variable de mesure ou d'une variable nominale dans chacun des cas ci-dessous. S'il s'agit d'une variable nominale, indique si elle est binaire.
   a) le poids d'une personne
   b) le sexe d'une personne
   c) l'allégeance politique d'un électeur
   d) la réponse à la question sur la pièce de monnaie (page I-65)
   e) le nombre de provinces qu'une personne a visitées
   f) le fait qu'une personne a déjà visité le Mexique ou non
   g) le volume d'une tasse
   h) le nombre de mots dans une phrase
2. Dans la question 4 de la page I-65, la variable était le nombre de lettres dans chaque mot. Dans quel type de variable classes-tu la taille d'un mot?
3. Suppose que tu as classé chacun des mots selon les critères suivants :
   - 1 à 3 lettres — petit mot
   - 4 à 6 lettres — mot moyen
   - 7 à 9 lettres — mot long
   - 10 lettres et plus — mot très long

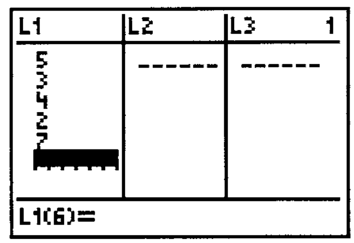
   Quel type de la variable correspond à la taille d'un mot maintenant?

   Les valeurs prises par une variable varient selon le cas. Les données affichent la dispersion. Le modèle de cette dispersion est appelé la distribution de la variable. Une bonne partie du travail du statisticien consiste à analyser des distributions de variables, pour les afficher, les résumer et les décrire.
4. Combien d'élèves ont répondu à la question sur l'abolition de la pièce de 1 cent?
5. Combien d'élèves veulent qu'elle soit abolie?
6. À quelle proportion du groupe cela correspond-il?
7. Trace un graphique à barres comme le suivant, avec deux barres de largeur égale.

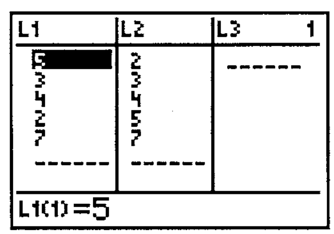


## Leçon 1 : Variables nominales (suite)

8. Entrée de données dans la calculatrice TI-83 - Méthode 1.
   - Mets la calculatrice en marche.
   - Appuie sur le bouton STAT.
   - Appuie sur ENTER ou sur 1 pour afficher l'écran de modification des données statistiques. Tu devrais voir trois colonnes avec des en-têtes L1, L2 et L3. Dans le cas contraire, appuie sur le bouton STAT de nouveau et sur 5 ou sur SetUpEditor, puis sur ENTER. L'écran affichera le mot DONE. Appuie sur STAT et sur ENTER de nouveau pour afficher les trois colonnes.
   - Si des données sont déjà inscrites dans L1, appuie sur le bouton de défilement vers le haut jusqu'à superposer la fenêtre de sélection ombragée sur l'en-tête; appuie sur CLEAR et sur ENTER pour effacer le contenu de la liste.



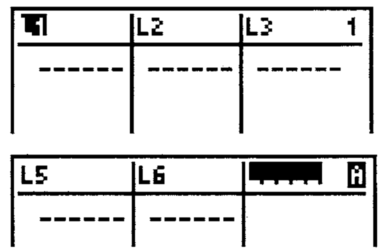

   - Tu peux maintenant entrer des données dans L1. Entre les nombres nécessaires, puis appuie sur ENTER. Essaie d'inscrire les nombres suivants : 5, 3, 4, 2, 7.
   - Quitte l'éditeur de statistiques en appuyant sur 2nd, puis sur QUIT.
9. Tri de données
   - Copie des données de la colonne L1 à la colonne L2. Pour ce faire, place le curseur sur l'en-tête L2, appuie sur 2nd L1 puis sur ENTER. Les données seront affichées à la fois dans L1 et dans L2.
   - Appuie sur le bouton STAT, puis sélectionne SORTA ( , puis appuie sur 2nd et sur L2 and), puis enfin sur ENTER pour trier la liste 2 en ordre croissant. L'écran d'édition des statistiques aura l'apparence ci-contre. Tu peux aussi faire un tri en ordre décroissant.



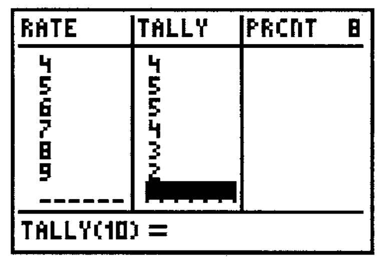
## Leçon 1 : Variables nominales (suite)

10. Nous utiliserons la calculatrice pour examiner les cotes accordées à la valeur de la statistique dans la société.
    - Premièrement, il faut calculer les réponses dans un tableau comme celui ci-dessous.

| Cote | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Calcul | | | | | | | | | |




    - On peut maintenant placer les données dans une nouvelle liste appelée COTE. Appuie sur le bouton STAT, sélectionne SetUpEditor puis appuie sur ENTER. Appuie sur le bouton STAT de nouveau, sélectionne EDIT et sur la flèche de défilement vers le haut. L'écran ci-contre devrait être affiché. (Ne tiens pas compte des données affichées dans les listes.)
    - Appuie sur la flèche vers la droite jusqu'à ce que l'écran suivant s'affiche. Entre le mot COTE dans l'espace réservé à l'en-tête de la liste.




    - Déplace le curseur à droite une autre fois pour entrer le mot CALCUL.
    - Déplace-toi à droite de nouveau pour entrer le mot POURC.
    - Inscris maintenant les données dans les deux listes. Ton écran aura cette apparence.

```
100*ʟTALLY/30
```

11. Tu peux trouver le pourcentage obtenu pour chaque cote accordée. Déplace ton curseur sur le titre POURC, entre le calcul affiché, puis appuie sur ENTER.

    ***Remarque :*** Utilise le nombre d'élèves de ta classe au lieu du nombre 30 affiché.

### Résumé

La statistique est la science des ***données***. Peut-être ces activités t'ont-elles donné une meilleure vision de ce que sont des données. Il ne s'agit pas simplement de nombres, mais de nombres qui ont une certaine signification dans un contexte donné. Il faut voir les données dans leur contexte.

Tu as découvert qu'il existait des variables nominales et des variables de mesure, et qu'on pouvait analyser leur distribution de façon visuelle. Tu as représenté des variables nominales à l'aide d'un graphique à barres. La prochaine leçon présente une technique de traçage du graphique des variables de mesure.

## Leçon 2 : Variables de mesure

### Contexte

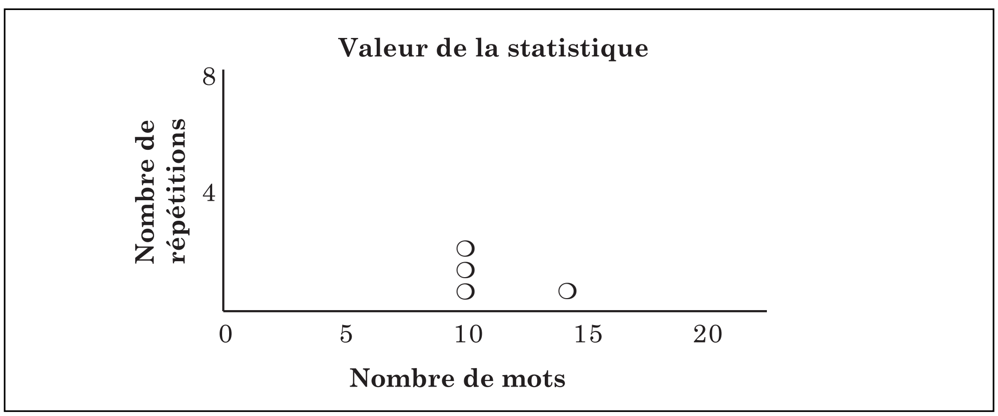
Nous avons déjà vu qu'un ensemble de données avait une ***distribution*** et tu as construis des ***graphiques à barres*** pour représenter la distribution d'ensembles des données qui comprennent des variables nominales. Nous allons maintenant étudier la distribution d'ensembles de données à l'aide de variables de ***mesure***. Nous étudierons plus particulièrement une méthode utilisée très communément pour afficher ces données. De plus, nous verrons certains termes utilisés pour décrire les ensembles de données et leur distribution.

## Points importants

- Élaborer une liste de termes pour décrire la distribution d'un ensemble de données. (Tu as déjà vu ces termes auparavant.)
- Étudier les **graphiques par points** pour afficher une distribution d'un ensemble de données numériques.
- Reconnaître les sources de **biais** dans les ensembles de données, selon la source.

## Activités de mise en marche

1. Dans la dernière leçon, tu as écris une phrase pour décrire ce que la statistique représentait pour toi. Combien de mots comprenait ta phrase? Construis un ***graphique par points*** à partir des réponses des élèves de ta classe. Il s'agit d'un graphique simple, comme celui qui apparaît ci-dessous, avec un point pour chaque réponse. L'exemple ci-dessous comprend trois phrases de dix mots et une phrase de quatorze mots.




   a) Combien y a-t-il de réponses?
   b) Quel est l'***étendue*** des données?
   c) Quelle est la valeur ***médiane***?
   d) Quel est le ***mode***?
   e) Quelle est la longueur ***moyenne***?
   f) Écris une phrase qui décrit la distribution de la longueur de la phrase.

   Nous avons fait un ***recensement***. Nous avons recueilli des données sur **la population** entière de la classe. La plupart du temps, les statisticiens ne peuvent pas recueillir des données sur toute une population, parce qu'elle est trop large ou inaccessible. Les frais d'un recensement deviennent vite astronomiques.

## Leçon 2 : Variables de mesure (suite)

Étant donné qu'il est souvent impossible de faire un recensement, nous apprendrons comment utiliser des données tirées d'un ***échantillon*** de la population. La prochaine question présente le concept de l'échantillonnage.

2. Je me demande combien de provinces et d'États un étudiant type de notre école a visités? Combien de provinces et d'États as-tu visités?
   a) Trace un graphique par points des données recueillies dans ta classe.
   b) Crois-tu que cette distribution nous donne une bonne idée de l'expérience de voyage d'un élève type de l'école? Écris une phrase ou deux pour expliquer ta réponse.

Nous avons travaillé sur un ***échantillon***. Puisque nous avons recueilli des ***statistiques*** sur une partie de la ***population*** de l'école.

Si notre classe n'est pas formée d'un groupe type d'élèves, on peut avoir une idée biaisée de l'expérience de voyage de l'élève type d'une école. Le ***biais***, ou la distorsion, résulte de données recueillies auprès d'un échantillon qui n'est pas représentatif de la population. Par exemple, si nous avions calculé l'âge moyen des élèves dans cette classe, nous pourrions nous attendre à ce que notre résultat produise une estimation biaisée de l'âge moyen des élèves de l'ensemble de l'école.

Quand on recueille des ***statistiques*** auprès d'un échantillon de la ***population***, il faut prendre garde à ne pas biaiser l'information si le but est de trouver les caractéristiques de la population.

3. Une épicerie compose un plateau de tacos et de diverses sauces et les fait goûter aux clients. On a confié ce travail à une étudiante qui travaille à temps partiel. Elle inscrit le type que chaque client préfère. Décris les sources possibles de biais dans ce contexte si la direction de l'épicerie a pour objectif de connaître les préférences de sa clientèle.
4. Un député a créé un site Internet où il veut recevoir des réponses par courriel à la question suivante : « Le gouvernement devrait-il adopter la Loi sur le contrôle des armes? ». Il a reçu 342 réponses, dont 311 sont négatives.
   a) Quelle population est en cause?
   b) Quelle variable est mesurée?
   c) Quel échantillon est utilisé?
   d) Discute des sources de biais inhérentes à ce type d'enquête.

### Résumé

- Le centre d'une distribution est en règle générale l'aspect le plus important. La moyenne, le point médian et le mode donnent de l'information sur la position des données.
- Les graphiques par points nous montrent comment les données sont distribuées, ce qui nous permet de déterminer de façon visuelle le centre et la dispersion de la distribution.
- Il faut savoir si les données portent sur une population ou sur un échantillon, sans quoi les données sont de peu d'utilité. Nous appelons une caractéristique de la population un ***paramètre***. Nous appelons une caractéristique d'un échantillon une ***statistique***.
- Les biais portent atteinte à leur valeur des données. Il faut les éviter autant que possible.

## Leçon 3 : Échantillon aléatoire simple

### Contexte

Tu as déjà vu ce qu'étaient les biais d'échantillonnage. Tu peux les supprimer au moyen d'un ***échantillon aléatoire simple***. Il s'agit d'un échantillon sélectionné en affectant à chacun des membres d'une population un nombre, puis en choisissant un ensemble de nombres à partir d'un tableau de nombres aléatoires ou d'une liste générée par le générateur de nombres aléatoires d'une calculatrice. On peut ainsi éliminer les biais.

### Points importants

- Effectuer une expérience avec un échantillon aléatoire simple (ÉAS).
- Déterminer des facteurs qui ont une incidence sur la dispersion de la distribution d'un échantillon.

### Activités de mise en marche

1. Comment éviter les biais d'échantillonnage? En utilisant un ***échantillon aléatoire simple***. Pour ce faire, il faut tout d'abord numéroter les lignes à droite de 1 à 30.
2. Mets ta calculatrice en marche et efface l'écran.
3. Appuie sur le bouton MATH et sélectionne le menu PRB.
4. Sélectionne randlnt, puis entre 1, une virgule et 30.
5. Ferme les parenthèses; ton écran aura cette apparence :

   ```
   randInt(1,30)
   ```

   appuie sur ENTER.
6. Un entier entre 1 et 30 est affiché. Encercle ce nombre à droite pour sélectionner une ligne.
7. Appuie sur ENTER de nouveau pour obtenir un autre nombre généré aléatoirement. Si le nombre est le même que le précédent, n'en tiens pas compte et appuie sur ENTER de nouveau jusqu'à ce qu'un nouveau nombre soit affiché.
8. Choisis deux autres nombres de cette façon, de sorte à sélectionner quatre lignes.
9. Mesure ces lignes au millimètre près.
10. Calcule la longueur moyenne.
11. Fabrique un graphique par points des longueurs moyennes calculées par tous les élèves de la classe. Discute de la distribution des longueurs moyennes.

## Leçon 3 : Échantillon aléatoire simple (suite)

### Déroulement

- Répète les étapes de la page I-10, en générant cette fois un échantillon de 10 nombres choisis aléatoirement entre 1 et 30 et en trouvant la longueur moyenne des lignes. Fabrique un graphique par points de la distribution des longueurs moyennes. Est-il différent du premier? Si oui, de quelle façon?
- La ***taille de l'échantillon*** a-t-elle une incidence sur la dispersion des ***statistiques*** que tu as recueillies?
- La forme de la distribution est-elle prévisible? Aurais-tu pu prédire la dispersion des longueurs moyennes si tu avais choisi 15 lignes?
- Crois-tu être en mesure de prédire de façon raisonnable la longueur moyenne des 30 lignes? Fais un estimé. Marque cette valeur estimée sur ton graphique par points.
- Mesure toutes les lignes au millimètre près et enregistre leur longueur. Trouve la longueur moyenne. (Cette valeur est un ***paramètre*** de la ***population*** des lignes).
- Marque cette longueur moyenne sur tes graphiques par points. Est-elle proche de la valeur que tu as estimée?

**Résumé**

Tu as étudié des ***échantillons aléatoires simples*** pour une variable de mesure (longueur de ligne). Tu as utilisé des graphiques par points pour illustrer la distribution de l'échantillon des longueurs.

Tu as probablement remarqué que, quand la taille de l'échantillon est plus grande, les moyennes varient moins.

## Leçon 4 : Description de distributions

### Contexte

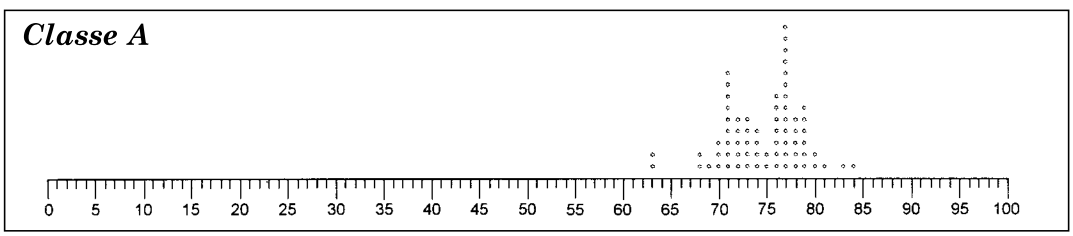
Tu sais que les données varient d'un échantillon à l'autre. Dans cette leçon, tu verras de nombreuses distributions de fréquence qui te permettront de connaître des caractéristiques que les statisticiens jugent importantes quand ils décrivent les distributions de fréquence.

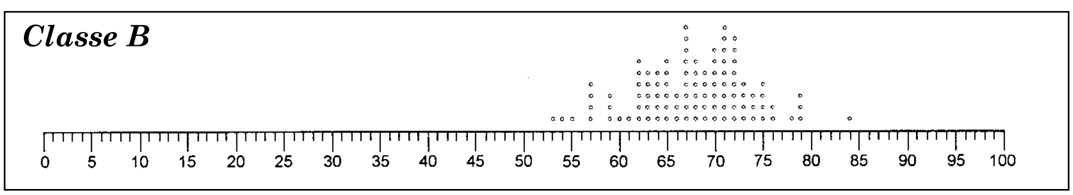
### Points importants

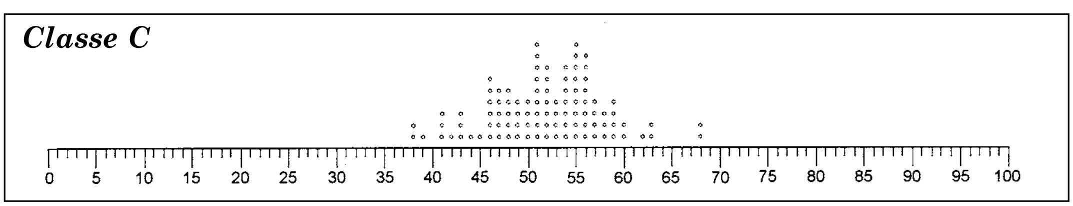
- Décrire la ***dispersion*** d'une distribution.
- Apprendre des termes qui décrivent la forme d'une distribution.
- Apprendre la signification des termes suivants : ***sommets***, ***grappe ou groupe***, valeur aberrante et ***granularité***.
- Comparer différentes façons de décrire le centre d'une distribution.
- Fabriquer un ***diagramme à tiges*** pour illustrer les variables de mesure d'une distribution.
- Utiliser le résumé en cinq points de la ***distribution*** d'une variable.
- Fabriquer un diagramme en boîte pour illustrer une distribution avec son centre et sa dispersion.
- Fabriquer un histogramme pour illustrer la distribution d'une variable.

### Activités de mise en marche

1. Nos problèmes d'introduction sont fondés sur ces graphiques par points, qui illustrent les notes obtenues par les élèves de 12 classes à un examen.



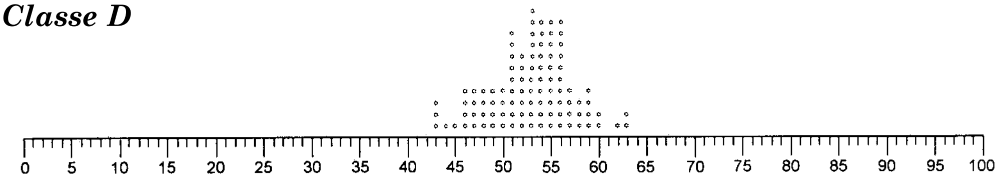



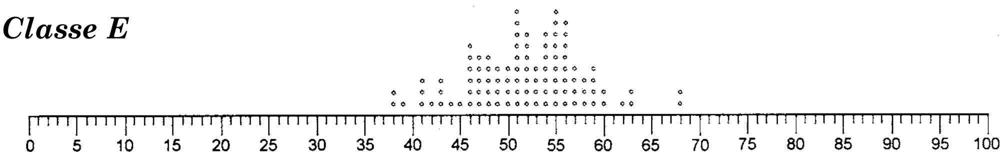

   a) Quelle est la différence la plus notoire entre les distributions des notes des classes A, B et C?

## Leçon 4 : Description de distributions (suite)

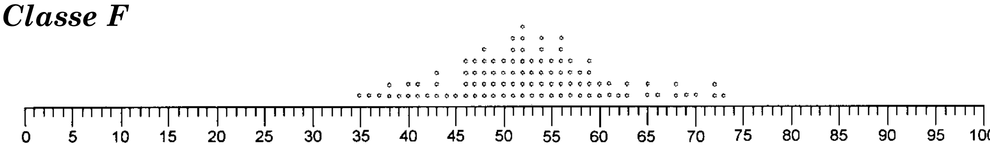
***Classe D***




***Classe E***

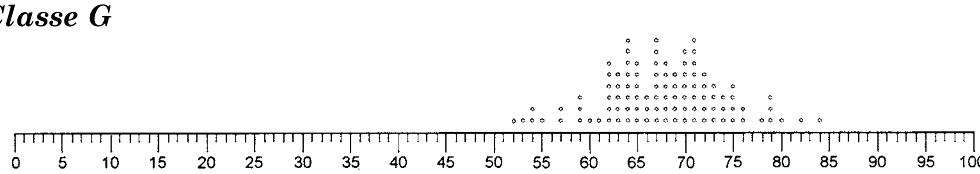



***Classe F***

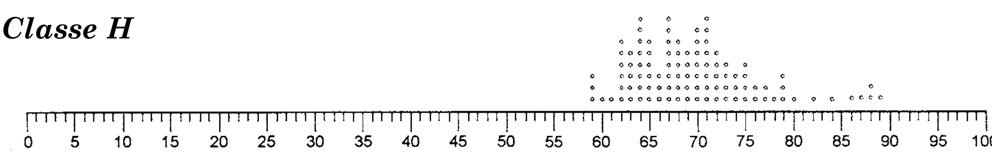



b) Quelle est la différence la plus notoire entre les distributions des notes des classes D, E et F?

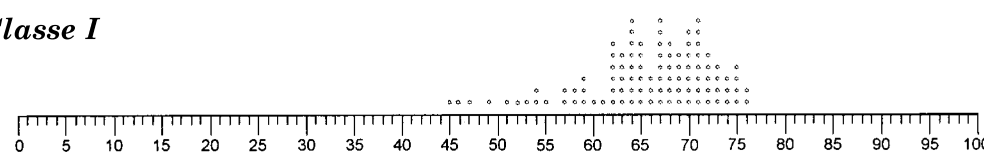
***Classe G***




***Classe H***



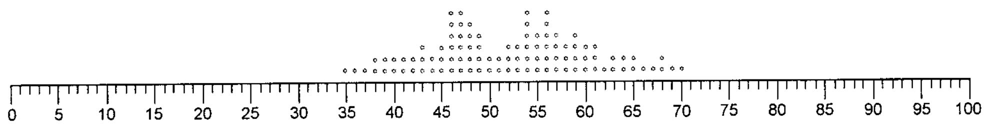

***Classe I***



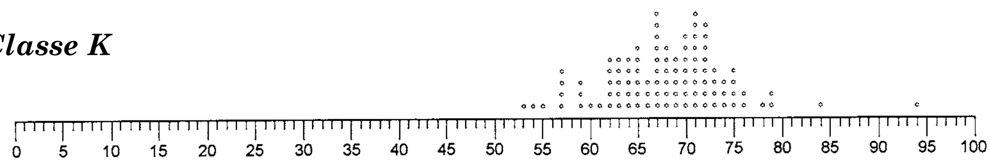

c) Quelle est la différence la plus notoire entre les distributions des notes des classes G, H et I?

## Leçon 4 : Description de distributions (suite)

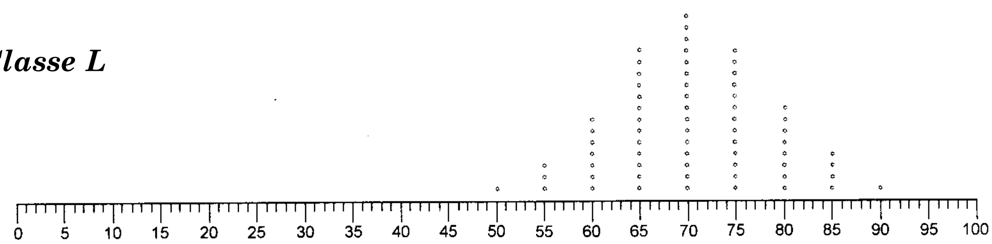
***Classe J***




***Classe K***




***Classe L***




d) Quelle est la caractéristique la plus notoire entre les distributions des notes de la classe J?

e) Quelle est la caractéristique la plus notoire entre les distributions des notes de la classe K?

f) Quelle est la caractéristique la plus notoire entre les distributions des notes des classes L?

Ces notes hypothétiques illustrent six caractéristiques qui sont souvent importantes dans l'analyse d'une distribution de données :

1. Le centre d'une distribution est en règle générale l'aspect le plus important à remarquer et à décrire. Autour de quelle valeur les données sont-elles centrées?
2. La dispersion d'une distribution représente la deuxième caractéristique en importance. À quel point les données sont-elles dispersées?
3. La forme de la distribution peut aussi nous révéler des choses. Certaines formes ont été baptisées. Une distribution est dite ***symétrique*** si une moitié est plus ou moins le reflet de l'autre - la courbe ressemble à une cloche. Une distribution est dite ***asymétrique*** si une sorte de queue se prolonge dans une direction. Elle est asymétrique vers la gauche si la queue est dirigée vers la gauche. Elle est asymétrique vers la droite si la queue est dirigée vers la droite.
4. Les valeurs ***aberrantes*** sont des points de données qui sont éloignés du groupe de points. Il faut les examiner de près.
5. Si on constate la présence de valeurs à des intervalles fixes, on parle de ***granularité*** de la distribution.

## Leçon 4 : Description de distributions (suite)

Il ne s'agit pas de règles rigides. Elles sont là simplement pour rappeler les caractéristiques qui méritent qu'on s'y attarde. Elles ne sont pas toujours pertinentes, et d'autres caractéristiques d'une distribution sont aussi dignes d'intérêt.

### Déroulement

1. On t'a déjà présenté les ***graphiques à tiges*** dans d'autres cours de mathématiques. Nous allons fabriquer un graphique à tiges avec les données ci-dessous, qui indiquent le temps en années du règne de différents souverains de l'Angleterre.

| Souverain | Années | Souverain | Années | Souverain | Années | Souverain | Années |
|---|---|---|---|---|---|---|---|
| Guillaume I | 21 | Édouard III | 50 | Édouard VI | 6 | Georges I | 13 |
| Guillaume II | 13 | Richard II | 22 | Marie I | 5 | Georges II | 33 |
| Henri I | 35 | Henri IV | 13 | Élizabeth I | 44 | Georges III | 59 |
| Étienne | 19 | Henri V | 9 | Jacques I | 22 | Georges IV | 10 |
| Henri II | 35 | Henri VI | 39 | Charles I | 24 | Guillaume IV | 7 |
| Richard I | 10 | Édouard IV | 22 | Charles II | 25 | Victoria | 63 |
| Jean | 17 | Édouard V | 0 | Jacques II | 3 | Édouard VII | 9 |
| Henri III | 56 | Richard III | 2 | Guillaume III | 13 | Georges V | 25 |
| Édouard I | 35 | Henri VII | 24 | Marie II | 6 | Édouard VIII | 1 |
| Édouard II | 20 | Henri VIII | 38 | Anne | 12 | Georges VI | 15 |

a) Combien de temps a duré le règne le plus long? Qui a régné le plus longtemps?

b) Quel a été le plus court règne? Qui était souverain durant cette période? Quelle est la signification réelle de cette valeur à ton avis?

c) Écris une phrase ou deux qui décrivent la distribution des longueurs de règne de ces souverains. Que constates-tu au sujet du centre de la distribution? Qu'en est-il de sa dispersion? Est-elle asymétrique?

d) Trouve une longueur ***médiane*** de règne. Pour les points de données impairs, ce sera facile. Le point milieu dans l'ensemble ordonné est le point médian. Pour les nombres pairs, comme c'est le cas ici, le milieu se trouve entre deux valeurs. Nous pourrions choisir une valeur quelconque entre ces deux valeurs, et faire la moyenne pour trouver le point médian. Cette valeur ne peut pas représenter la longueur réelle du règne d'un souverain, mais la moitié des souverains auront régné pendant une période plus courte que la longueur moyenne.

e) Trouve une valeur telle que le quart des souverains auront eu des règnes plus courts que celle-ci. (Trouve la médiane des règnes plus courts que le règne médian.) Cette valeur est appelée le premier ***quartile*** (Q1).

f) Trouve une valeur telle que le quart des souverains auront régné plus longtemps que cette dernière. (Trouve la médiane des règnes plus longs que le règne médian) Cette valeur est appelée le ***troisième quartile*** (Q3).

g) Tu as maintenant 5 nombres (minimum, Q1, médiane, Q3, maximum) qui te procurent un résumé utile de la distribution. Il s'agit du ***sommaire de cinq nombres*** de la distribution; ces nombres sont tout ce dont nous avons besoin pour produire une autre présentation visuelle de la distribution, le ***graphique en boîte***.

## Leçon 4 : Description de distributions (suite)

2. Utilise le sommaire de cinq nombres pour fabriquer un graphique en boîte de la distribution.
   a) Premièrement, trace une droite horizontale numérotée. Tu dois inscrire l'ensemble des valeurs correspondant à la longueur des règnes.
   b) À plusieurs espaces au-dessus de la droite, trace des droites verticales de un espace de long et, sur la même droite, inscris au-dessus de la droite numérotée les valeurs du Q1, de la médiane et du Q3.
   c) Trace une droite horizontale à partir du sommet de la droite Q1 jusqu'au sommet de la droite Q3, puis une autre droite entre la base de la droite Q1 jusqu'à la base de la droite Q3. Tu devrais obtenir une boîte, avec une droite verticale au-dessus de la valeur médiane.
   d) Trace une droite horizontale à partir du milieu de la droite Q1 jusqu'à la gauche du point au-dessus de la position de la droite numérotée qui représente la longueur minimale des règnes.
   e) Trace une droite horizontale entre le milieu de la droite Q3 vers la droite, jusqu'au point au-dessus de la position de la droite numérotée qui représente la longueur maximale des règnes.
   f) Dessine une petite étoile aux extrémités de ces droites, que nous appelons ***des moustaches***.
3. À l'aide de la calculatrice TI-83, tu peux obtenir le sommaire en cinq points et tracer un diagramme en boîte de ces données.
   a) Entre 40 nombres représentant les longueurs des règnes dans la colonne L1.
   b) Appuie sur Y= pour obtenir le menu des équations. Assure-toi qu'aucune équation n'est enregistrée et, si une équation est enregistrée, efface-la ou place le curseur sur le symbole = et appuie sur ENTER, pour que le signe ne soit pas ombragé. Cette équation sera désactivée, mais non supprimée.
   c) Appuie sur 2nd et sur STAT PLOT. Le menu de traçage des courbes des statistiques est affiché. Si des tracés sont déjà activés, sélectionne l'élément 4 du menu et appuie sur ENTER. Tous les tracés sont désactivés.
   d) Sélectionne l'élément 1 du menu puis appuie sur ENTER. Appuie de nouveau sur ENTER pour activer le tracé 1 Plot 1, puis sélectionne le graphique par boîte avec le curseur (la deuxième icône de la deuxième rangée) puis appuie sur ENTER.
   e) Déplace le curseur vers le bas pour sélectionner la Xlist. Une fois le curseur en position, appuie sur 2nd, sur L1, puis sur ENTER.
   f) Appuie maintenant sur ZOOM et 9 (ou sélectionne ZoomStat dans le menu). Un graphique en boîte est tracé.
   g) Appuie sur TRACE et sur les flèches gauche et droite pour afficher les cinq points du sommaire en cinq points.
4. Utilise ta calculatrice pour obtenir le sommaire de la distribution.
   a) Appuie sur STAT et choisis CALC, puis sélectionne l'élément 1 du menu (1-Var Stats), et appuie sur ENTER.
   b) Tu vois maintenant « 1-Var Stats » et le curseur clignote. Appuie sur 2nd et sur L1, puis sur ENTER. Ta calculatrice affiche un écran de statistique qui commence par $\overline{x}$, soit la valeur moyenne de nos données, et qui se termine par le nombre de points de données (40) dans notre distribution. Appuie sur la flèche pour faire défiler l'écran vers le bas et voir les cinq points du sommaire en cinq points.

## Leçon 4 : Description de distributions (suite)

3. Utilise ta calculatrice pour afficher un histogramme de la distribution.

   L'image des données est divisée en intervalles d'égale longueur dans l'histogramme. Une barre est affichée pour chaque intervalle. Le nombre d'observations (fréquence) appartenant à chaque intervalle détermine la hauteur de la barre dans cet intervalle. Pour les variables nominales, la hauteur de la barre correspond à la proportion des observations (fréquence relative) dans l'intervalle. Étant donné que les barres sont de largeur égale, on peut dire aussi que l'aire de chacune correspond à la fréquence relative dans cet intervalle. Aucune règle fixe n'indique la largeur ou le nombre des barres d'un histogramme. La distribution des données devrait être évidente quand on voit le diagramme - avec un peu d'expérience et de tâtonnement, le statisticien arrive à tirer beaucoup d'information de ces diagrammes.

   a) Commence de la même façon que pour le diagramme en boîte. Sélectionne l'icône histogramme (la troisième dans la rangée du haut). Utilise la même méthode de zoom pour afficher l'histogramme.

   b) Utilise la touche Trace pour voir les limites de tes barres et la valeur centrale de chacune.

   c) Appuie sur la touche WINDOW et déplace le curseur jusqu'à la rangée. Change l'échelle et appuie sur GRAPH. L'histogramme est redessiné avec des barres de la largeur inscrite dans Xscl. Fais des essais avec différentes échelles. Laquelle à ton avis résume le mieux la distribution des règnes des souverains?

**Sommaire**

Tu connais maintenant la terminologie utilisée pour décrire la forme d'une distribution. Tu sais aussi de quelle façon la taille d'un échantillon peut influer sur la dispersion des moyennes, que la variable en soit une de mesure ou qu'il s'agisse d'une proportion calculée à partir de variables nominales. Tu as utilisé le sommaire en cinq points d'une distribution pour tracer son diagramme en boîte. Tu connais certaines caractéristiques de la médiane et de la moyenne, et tu as utilisé les histogrammes pour afficher des distributions.

| |
|---|
| ***Avant de passer à la leçon suivante, discutez du contenu avec un groupe d'élèves.*** |

# Leçon 5 : La règle empirique

### Contexte

Les mesures du centre et de l'étendue que nous avons utilisées, ainsi que les affichages que nous avons obtenus sont des outils très précieux en statistique. Cependant, certaines failles nous obligent à utiliser une autre mesure de l'étendue. Nous illustrerons ce besoin dans les activités de mise en marche.

## Points importants

- Utiliser l'***écart type*** comme mesure de l'image d'une distribution.
- Trouver la règle ***empirique relative*** aux distributions en forme de cloche (symétriques).
- Utiliser l'écart type pour calculer les **cotes *z*** des données d'une distribution.

### Activités de mise en marche

Les élèves du cours de Mathématiques pré-calcul de secondaire 4 du Manitoba ont fait un examen provincial l'an dernier. Nous avons regroupé les notes obtenues à l'examen par les élèves de trois classes en vue de les analyser. Les voici :

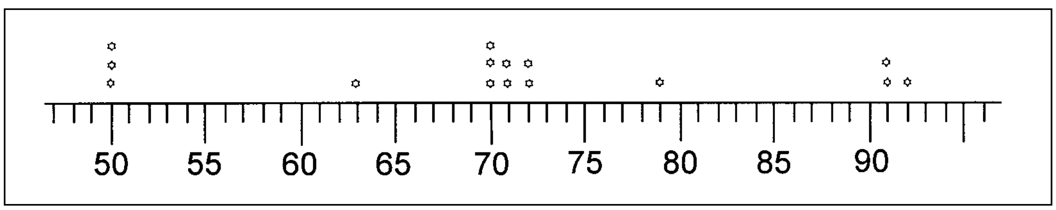
> ***Remarque :*** La taille des échantillons est petite afin de rendre la résolution du problème plus facile. Une étude réelle exigerait de considérer des échantillons plus grands.

Classe 1 : 50 50 50 63 70 70 70 71 71 72 72 79 91 91 92

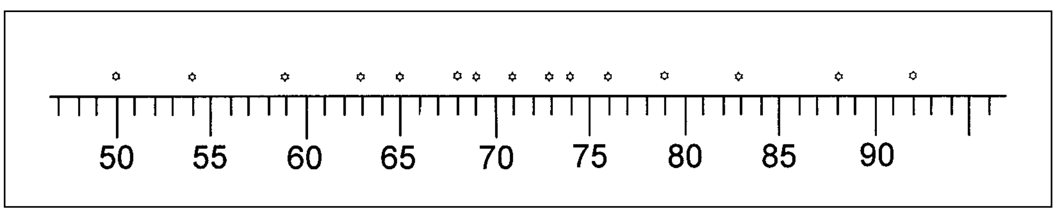
Classe 2 : 50 54 59 63 65 68 69 71 73 74 76 79 83 88 92

Classe 3 : 50 61 62 63 63 64 66 71 77 77 77 79 80 80 92

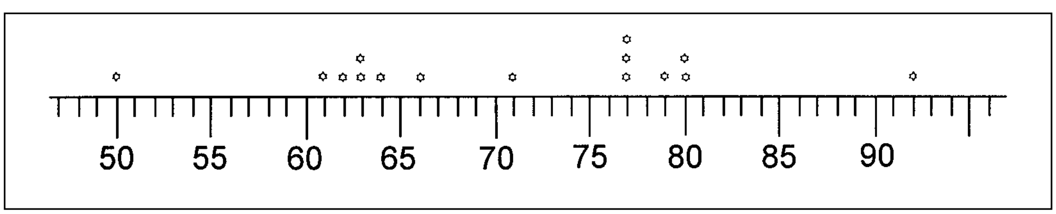
***Classe 1***




***Classe 2***




***Classe 3***



## Leçon 5 : La règle empirique (suite)

a) Ces distributions sont-elles différentes? Explique pourquoi.

b) Fabrique des diagrammes en boîte des trois distributions.

c) Si tu avais vu seulement les diagrammes en boîte, aurais-tu pu conclure que les trois distributions étaient différentes?

Nous allons maintenant laisser de côté les diagrammes en boîte et le sommaire en cinq points pour utiliser des ***histogrammes***. Nous utiliserons la ***moyenne*** comme mesure du centre et l'***écart type*** comme mesure de l'étendue. L'écart type nous permet de mesurer la distance des observations par rapport à la moyenne. Étant donné que la distance n'est pas fonction de la direction, nous mettons les valeurs au carré pour éliminer les signes dans le calcul de la distance, $x_1 - \overline{x}$ pour chaque point de donnée. Ensuite, il faut additionner toutes les valeurs $(x_1 - \overline{x})^2$. Intuitivement, on constate que la division par le nombre $n$ des points de donnée nous permettrait d'obtenir la moyenne des carrés des distances. Cette méthode convient pour l'analyse d'une population, mais les statisticiens utilisent plutôt la division par $n - 1$ pour les échantillons. Dans les deux cas, le quotient est appelé la ***variance*** ($s^2$). L'écart type ($s_x$) est sa racine carrée.

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n}(x_1 - \overline{x})^2}{n}} \text{ pour la population} \qquad s_x = \sqrt{\frac{\sum_{i=1}^{n}(x_1 - \overline{x})^2}{n-1}} \text{ pour un échantillon}$$

### Déroulement

Le nombre de calories utilisées chaque jour par une personne est appelé son taux métabolique. L'étude du taux métabolique est particulièrement importante dans les études sur l'alimentation et l'exercice. La population étudiée dans l'expérience suivante est l'ensemble des Canadiens adultes. Une étude sur la diète a été effectuée auprès d'un échantillon de sept hommes. Leurs taux métaboliques respectifs vont comme suit :

1792  1666  1362  1614  1460  1867  1439

La moyenne est 1600 calories ($\overline{x} = 1600$).

## Leçon 5 : La règle empirique (suite)

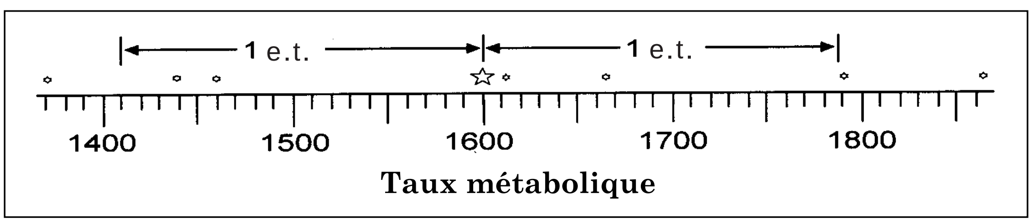
Nous allons calculer la variance ($s^2$) et l'écart type ($s_x$) en calculant l'écart de chaque observation à la moyenne, puis en mettant le résultat au carré.

| **Observation** $x$ | **Écarts** $x-\overline{x}$ | **Écarts au carré** $(x-\overline{x})^2$ |
|---|---|---|
| 1792 | 1792 – 1600 = 192 | $(192)^2$ = 36 864 |
| 1666 | 1666 – 1600 = 66 | $(66)^2$ = 4356 |
| 1362 | 1362 – 1600 = –238 | $(-238)^2$ = 56 644 |
| 1614 | 1614 – 1600 = 14 | $(14)^2$ = 196 |
| 1460 | 1460 – 1600 = –140 | $(-140)^2$ = 19 600 |
| 1867 | 1867 – 1600 = 267 | $(267)^2$ = 71 289 |
| 1439 | 1439 – 1600 = –161 | $(-161)^2$ = 25 921 |
| | | somme = 214 870 |

L'écart type $\sqrt{\dfrac{214\,870}{6}} = 189{,}24$ calories.

***Remarque :*** Nous avons divisé le résultat par $n - 1$ parce qu'il s'agit d'un échantillon.

Le diagramme ci-dessous illustre les données correspondant aux taux métaboliques des sept hommes. La valeur moyenne des taux est marquée par une étoile. Les flèches indiquent une distance de un écart type de chaque côté de la moyenne. À première vue, on pourrait penser qu'il s'agit de l'amorce d'un diagramme en boîte, mais n'oublie pas qu'on n'utilise pas les diagrammes en boîte pour l'étude de la moyenne et de l'écart type.




1. Utilise le menu des statistiques de la calculatrice pour trouver la moyenne et l'écart type dans les trois classes dont il était question dans les exercices de mise en marche. Les moyennes sont-elles différentes? Qu'en est-il des écarts types? (***Remarque :*** Ta calculatrice affiche deux symboles, $s_x$ et $\sigma_x$, qui donnent deux valeurs similaires. Nous utiliserons la première, soit l'écart type pour un échantillon.)

2. Effectue le problème sur le taux métabolique à l'aide de ta calculatrice. Trouve l'intervalle interquartile et l'écart type. Remplace maintenant l'entrée 1362 par 362, comme si les données avaient été mal lues. Quel élément est le plus touché par cette opération : l'étendue interquartile ou l'écart type?

# Leçon 5 : La règle empirique (suite)

## La règle empirique

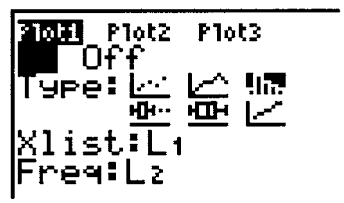
Voici les résultats d'un examen à choix multiples comportant 20 questions qui a été distribué à 213 élèves. $\overline{x}$ = 10,221 et $s_x$ = 3,859. (Les élèves représentent un échantillon de l'ensemble des élèves qui ont fait l'examen.)

| Note | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Nombre | 1 | 1 | 5 | 7 | 12 | 13 | 16 | 15 | 17 | 32 |

| Note | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| Nombre | 17 | 21 | 12 | 16 | 8 | 4 | 7 | 5 | 4 | 0 |

3. Efface la liste dans l'éditeur des statistiques. Entre les notes dans la liste 1 et les nombres dans la liste 2. Règle à Plot 1, comme l'illustre la figure.

   Règle la fenêtre de graphique comme illustré. Appuie ensuite sur la touche GRAPH.





4. La distribution semble-t-elle avoir plus ou moins une forme de cloche et être symétrique?
5. Combien de notes se trouvent à 1 écart type de la moyenne (entre $\overline{x} - s$ et $\overline{x} + s$)? Quelle proportion des 213 notes cela représente-t-il?
6. Combien de notes se trouvent à 2 écarts types de la moyenne (entre $\overline{x} - 2s$ et $\overline{x} + 2s$)? Quelle proportion de 213 notes cela représente-t-il?
7. Si nous regardons les données qui se trouvent à 3 écarts types de la moyenne, nous obtiendrions soit –1,356 ou 21,789. Quelle proportion des 213 notes se trouve entre ces valeurs?

Nous avons trouvé la ***règle empirique***. Dans les distributions en forme de cloche, environ 68 % des observations se trouvent à 1 écart type de la moyenne, environ 95 % se trouvent à 2 écarts types de la moyenne et plus de 99 % se trouvent à 3 écarts types de la moyenne. Étant donné que beaucoup de distributions ont une forme de cloche, cette règle rend l'écart type utile pour l'interprétation des données.

## Leçon 5 : La règle empirique (suite)

### Exercice

1. Les notes obtenues au Scholastic Assessment Test (SAT) sont très largement utilisées par les universités pour évaluer les demandes d'admission. Supposons que la note moyenne au SAT est 896 et que l'écart type de ces notes est 174. D'autres universités utilisent les notes au American College Test (ACT). La moyenne est de 20,6 et l'écart type de 5,2.

   a) Si Robert a obtenu 1080 au SAT, à combien de points au-dessus de la moyenne sa note se trouve-t-elle?

   b) Si Kathy a obtenu 28 à l'ACT, à combien de points au-dessus de la moyenne sa note se trouve-t-elle?

   c) Serait-il sensé de conclure que, étant donné que la réponse en (a) est supérieure à celle obtenue en (b), Robert a mieux réussi que Kathy? Explique pourquoi.

   d) À combien d'écarts types au-dessus de la moyenne la note de Robert au test se trouve-t-elle? (Divise le nombre de points qu'il a obtenus au-dessus de la moyenne par l'écart type). C'est la cote $z$ de Robert.

   e) À combien d'écarts types au-dessus de la moyenne la note de Kathy se trouve-t-elle? (Divise le nombre de points obtenus au-dessus de la moyenne par l'écart type.) C'est la cote $z$ de Kathy.

   Tu as trouvé un moyen d'utiliser les écarts types pour comparer des valeurs distinctes provenant de distributions différentes. Tu as trouvé la ***cote z*** ou la ***cote standardisée*** pour chaque personne en soustrayant la moyenne de la distribution du score brut qui nous intéresse, puis en divisant le résultat par l'écart type. Les cotes $z$ nous indiquent à combien d'écarts types au-dessus (ou au-dessous) de la moyenne une valeur se trouve. Cependant, les cotes $z$ peuvent être utilisées seulement pour les distributions en cloche.

   f) Laquelle parmi les deux personnes visées a obtenu la cote $z$ la plus élevée à son test d'admission?

   g) Explique en tes propres mots laquelle a le mieux réussi.

   h) Calcule la cote $z$ de Mike, qui a obtenu 740 au SAT et de Karine, qui a obtenu 19 à l'ACT.

   i) Laquelle des ces deux personnes a obtenu la note la plus élevée?

   j) Dans quelles conditions une cote $z$ est-elle négative?

### Sommaire

Tu as découvert qu'il fallait tenir compte d'une nouvelle mesure de l'étendue, et tu as étudié l'écart type. Tu as découvert la règle empirique qui nous donne une façon d'utiliser l'écart type de façon intuitive. Finalement, tu as utilisé les cotes $z$ pour comparer des valeurs provenant de distributions différentes. Les concepts abordés sont à la base du reste du cours.

## Leçon 6 : La signification statistique

### Contexte

Tu sais qu'on peut rassembler un échantillon représentatif de la population étudiée au moyen d'un ***échantillon aléatoire simple***. Tu sais aussi que les statistiques qui portent sur un échantillon peuvent varier et qu'elles sont influencées par la ***taille de l'échantillon***. Tu as trouvé une ***règle empirique*** qui établit qu'environ 68 % des proportions de l'échantillon de la variable mesurée se trouvent à l'intérieur de l'écart type de la moyenne, qu'environ 95 % se trouvent à l'intérieur de 2 écarts types de la moyenne et que la presque totalité se trouvent à l'intérieur de 3 écarts types de la moyenne.

Nous allons maintenant pousser plus loin l'étude des variations à l'intérieur des échantillons à l'aide d'une expérience initiale, puis nous ferons une simulation sur calculatrice pour augmenter le nombre d'échantillons.

### Points importants

- Reconnaître que les statistiques sur des échantillons varient de façon prévisible autour d'une moyenne centrale et qu'elles ont une distribution symétrique.
- Reconnaître que cette distribution est caractérisée par un paramètre de population qui s'approche de l'ensemble des statistiques.
- Reconnaître que cette distribution suit une règle empirique.
- Accepter qu'une formule permet de trouver l'écart type de la distribution d'échantillonnage d'une proportion quand la moyenne de la population et la taille de l'échantillon sont connues.
- Élaborer sur le concept de la ***signification statistique*** en analysant des simulations.

### Activités de mise en marche

Une ***population*** est l'ensemble d'un groupe d'événements visés. Il peut s'agir d'un groupe de personnes, d'objets ou d'événements. Un ***paramètre*** est la caractéristique numérique de la population alors que la ***statistique*** est la même caractéristique telle qu'elle se rapporte à un ***échantillon***.

1. La compagnie Hershey fabrique des bonbons Reese de différentes couleurs. Soit la population de ces bonbons - t'es-tu déjà questionné sur la distribution des couleurs? Comme tu ne veux pas étudier des millions de bonbons Reese, tu décides de prendre un échantillon de 25 bonbons.

   a) Soit un échantillon aléatoire de 25 bonbons; enregistre le nombre et la proportion de chaque couleur dans ton échantillon.

   | | Orange | Jaune | Brun |
   |---|---|---|---|
   | Nombre | | | |
   | Proportion | | | |

   La proportion de bonbons orange dans ton échantillon constitue-t-elle un paramètre ou une statistique?

   b) La proportion de bonbons orange fabriqués par Hershey constitue-t-elle un paramètre ou une statistique?

   c) Connais-tu la proportion de bonbons orange fabriqués par Hershey?

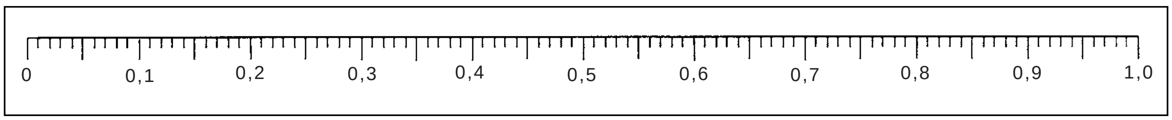
## Leçon 6 : La signification statistique (suite)

d) Quelle est la proportion de bonbons orange dans ton échantillon? Enregistre ta réponse dans le tableau que ton enseignant t'a fourni.

Ces questions nous permettent de voir que, s'il est facile de trouver la valeur d'une statistique issue d'un échantillon, il n'est pas aussi facile de trouver la valeur du paramètre de population. L'un de nos objectifs est d'estimer la valeur du paramètre à partir des statistiques.

e) Crois-tu que tous ceux qui rassemblent un échantillon de 25 bonbons obtiennent la même proportion de bonbons orange que toi?

f) Fabrique un diagramme par points représentant la proportion dans l'échantillon de bonbons orange que chaque élève de la classe a obtenue. Utilise une échelle comme la suivante :




g) Si tous les élèves estiment la proportion de bonbons orange dans la population selon la proportion se rapportant à son échantillon, arriverez-vous tous à la même estimation?

h) Selon ce que tu connais de l'échantillonnage aléatoire, écris un énoncé sur la proportion de bonbons orange dans la population.

i) Si on suppose que chaque élève connaissait seulement la statistique concernant son échantillon, la plupart des estimations sur les paramètres de la population seraient-elles raisonnablement proche du paramètre réel? Certaines seraient-elles très loin de la réalité? Explique pourquoi.

j) Si l'échantillon avait comporté 10 bonbons au lieu de 25, comment le graphique par points aurait-il été différent?

k) Si l'échantillon avait comporté 75 bonbons au lieu de 25, comment le graphique par points aurait-il été différent?

Tes résultats suggèrent que, même si les valeurs d'échantillonnage varient, la variation tend à suivre un modèle. Nous utiliserons une simulation issue de la calculatrice pour étudier ce processus plus en profondeur.

**Déroulement**

Pour effectuer une simulation, il faut tenir pour acquis que notre mélange ressemble à celui contenu dans les sacs de bonbons Reese. Pour y arriver, il faut connaître la proportion de bonbons orange dans la population. De fait, Hershey fabrique 45 % de bonbons orange. Le programme utilise cette donnée, que nous ne connaissions pas auparavant. Le programme utilise 50 échantillons aléatoires de 25 bonbons et trace le graphique des résultats. Appelle ce programme « BONBON ». Tu pourras télécharger une copie à partir de la calculatrice de ton enseignant. On trouve aussi ce programme dans la section Notes à l'intention de l'enseignant (annexe I-3), à la fin de ce document.

a) Utilise ta calculatrice pour rassembler 50 échantillons de 25 bonbons chacun (le traitement est un peu long, alors fait preuve de patience.)

b) Décris tout modèle que tu constates dans la variation des proportions dans les 50 échantillons.

c) Calcule la moyenne et l'écart type de ces proportions à l'aide de la calculatrice.

## Leçon 6 : La signification statistique (suite)

d) Les proportions dans l'échantillon sont-elles regroupées autour de la proportion dans la population, soit 0,45?

e) Il faut être plus précis dans notre réponse à cette dernière question. Ainsi, combien des 50 proportions dans l'échantillon se trouvent entre 0,35 et 0,55 (0,45 ± 0,10)? Combien se trouvent entre 0,25 et 0,65 (0,45 ± 0,20)? Combien se trouvent entre 0,30 et 0,45? (Tu peux utiliser la touche TRACE pour connaître ces valeurs)

***Remarque :*** L'écart type des données générées par ta calculatrice sera près de 0,10; dans la question, il faudra donc mettre à la place (0,45 ± 0,10), etc.

| | **Nombre de proportions dans l'échantillon** | **Pourcentage de proportions dans l'échantillon** |
|---|---|---|
| À l'intérieur de 0,45 ± 0,10 | | |
| À l'intérieur de 0,45 ± 0,20 | | |
| À l'intérieur de 0,45 ± 0,30 | | |

f) Au début, comme nous ne connaissions pas le paramètre de la population, nous aurions dû tenter d'estimer cette valeur à partir de notre statistique. Imagine que tu ne sais pas que la proportion de bonbons orange dans la population est 0,45. Si chacun des 50 élèves imaginaires représentés dans ta simulation estime la proportion de bonbons orange dans la population en choisissant une étendue entre 0,20 au-dessous et 0,20 au-dessus de la proportion dans l'échantillon, quel pourcentage d'entre eux auraient obtenu la valeur réelle de 0,45 dans cet intervalle?

g) Si tu étais l'un de ces élèves imaginaires qui étudient ta proportion dans l'échantillon, serais-tu ***sûr*** que la valeur pour ton échantillon se trouverait à l'intérieur d'un intervalle de 0,20 à la proportion dans la population? Serais-tu raisonnablement sûr que c'est le cas? Explique ce que tu veux dire.

Même si la proportion de bonbons orange dans l'échantillon varie d'un échantillon à l'autre, on reconnaît un modèle à long terme dans la variation. Il s'agit de la distribution d'échantillonnage de la proportion dans l'échantillon. Nous ne pouvons utiliser la proportion dans l'échantillon pour trouver la proportion dans la population réelle, mais nous pouvons être raisonnablement certains que la proportion dans la population se trouve à une certaine distance de la proportion dans l'échantillon. Cette distance dépend avant tout de notre degré souhaité de certitude et de la taille de l'échantillon.

h) Refais la même simulation, en utilisant cette fois-ci des échantillons de 75 bonbons. Il faudra pour ce faire changer l'inscription « 25 » dans le programme par « 75 ». Calcule la moyenne et l'écart type pour les proportions dans l'échantillon.

i) Écris tous les changements constatés dans la distribution d'échantillonnage par rapport à l'échantillon de 25 bonbons.

j) Compte le nombre (à l'aide de la touche TRACE) de proportions dans l'échantillon qui se trouvent dans un intervalle de 0,10, 0,20 et 0,30 par rapport à 0,45. Remplis le tableau de la page suivante, comme tu l'as fait auparavant.

## Leçon 6 : La signification statistique (suite)

| | Nombre de proportions dans l'échantillon | Pourcentage de proportions dans l'échantillon |
|---|---|---|
| À l'intérieur de $0,45 \pm 0,10$ | | |
| À l'intérieur de $0,45 \pm 0,20$ | | |
| À l'intérieur de $0,45 \pm 0,30$ | | |

Comment se comparent les pourcentages de proportions dans l'échantillon qui se trouvent dans l'intervalle de 0,10 par rapport à 0,45 entre les échantillons comprenant 25 et 75 éléments?

k) En règle générale, une proportion dans l'échantillon est-elle plus susceptible de se trouver à proximité de la proportion dans la population si l'échantillon est plus grand ou s'il est plus petit?

Étant donné que nos proportions dans l'échantillon suivent une distribution symétrique - en forme de cloche - , la règle empirique établit qu'environ 95 % des proportions dans l'échantillon se trouvent à l'intérieur de 2 écarts types de la moyenne.

l) Trouve la valeur qui est située à deux écarts types au-dessous de la moyenne que tu as trouvée en réponse à la question (h). Trouve la valeur qui se trouve à deux écarts types au-dessus de la moyenne.

m) Combien parmi les 50 proportions dans l'échantillon se trouvent à l'intérieur de l'intervalle que tu viens de trouver? Cette valeur est-elle proche de 95 %?

n) Si chacun des 50 élèves imaginaires devait faire le même travail que tu viens de faire avec sa propre moyenne, quel pourcentage de leurs intervalles comprendrait à ton avis la proportion dans la population réelle (0,45)?

Cela nous montre que si je veux être à environ 95 % sûr d'obtenir une proportion dans la population à l'intérieur d'une certaine distance de la proportion dans mon échantillon, cette distance devrait être environ le double de l'écart type de la distribution d'échantillonnage des proportions dans l'échantillon.

**Exercice**

1. Le tiers des objets produits par un fabricant sont défectueux. Si tu inspectes des lots aléatoires de 15 objets qui sortent de la chaîne de montage, combien d'objets défectueux pourrais-tu t'attendre à trouver dans un lot?
2. Si les inspecteurs prennent des échantillons différents de 15 objets au sortir de la chaîne de montage, t'attendrais-tu à ce que chacun d'eux trouve le même nombre d'objets défectueux?
3. Serais-tu très surpris de trouver quatre objets défectueux ou moins sur un lot de quinze si une chaîne de montage produit un tiers d'objets défectueux à long terme?
4. Serais-tu très surpris de constater que deux objets défectueux ou moins dans un lot de quinze si une chaîne de montage produit un tiers d'objets défectueux à long terme?
5. Serais-tu très surpris de retrouver aucun objet défectueux dans un lot de quinze si une chaîne de montage produit à long terme un tiers d'objets défectueux?

## Leçon 6 : La signification statistique (suite)

### Déroulement

Supposons que nous acceptons qu'un tiers de la production de gadgets d'un fabricant soit défectueuse. Supposons maintenant que les ingénieurs ont trouvé une façon de modifier le processus de fabrication qui, selon eux, devrait diminuer la proportion d'objets défectueux. Ils font des essais sur un lot de quinze objets.

a) La donnée « un tiers » est-elle un paramètre ou une statistique? Explique pourquoi.

b) Suppose que les ingénieurs échantillonnent un lot de quinze objets et qu'ils en trouvent quatre défectueux. Quelle est la proportion dans l'échantillon d'objets défectueux? Quel symbole utilise-t-on en règle générale pour représenter ce nombre?

c) Est-il possible que les inspecteurs aient trouvé quatre objets défectueux ou moins dans la population même si la modification n'a eu aucun effet sur le processus? Autrement dit, même si la proportion dans la population d'objets défectueux est encore de un tiers? Explique pourquoi.

d) Si les ingénieurs échantillonnaient 100 lots de 15 objets, devraient-ils s'attendre à trouver le même nombre d'objets défectueux dans chaque lot? Quel terme décrit ce phénomène?

Le temps est venu d'étudier le concept de la ***signification statistique***. Nous explorerons la distribution d'échantillonnage d'une proportion d'objets défectueux, et nous verrons à quelle fréquence un résultat d'échantillon observé peut se produire au hasard. Un résultat d'échantillon est ***statistiquement significatif*** s'il est peu susceptible d'être dû seulement à la dispersion d'échantillonnage.

e) Supposons que la modification n'a eu ***aucun effet***. Cela signifie que la proportion d'objets défectueux dans la population est toujours de un tiers. Utilise un dé ordinaire à six faces pour simuler la sélection aléatoire d'un lot d'objets. Lance le dé quinze fois. Si tu obtiens un 1 ou un 2, enregistre un objet défectueux (D). Si tu obtiens un autre chiffre, enregistre l'objet comme étant correct (C).

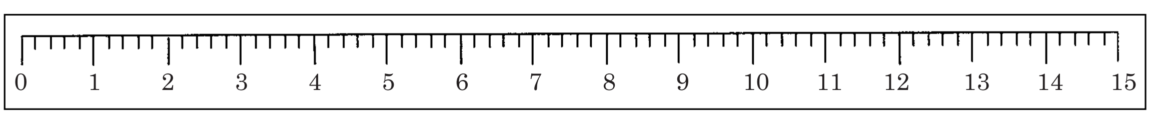
| Gadget | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N$^o$ sur le dé | | | | | | | | | | | | | | | |
| D ou C | | | | | | | | | | | | | | | |

Combien d'objets dans le lot simulé sont défectueux? Quelle proportion d'objets dans le lot sont défectueux?

f) Répète la procédure jusqu'à ce que tu aies simulé cinq lots de quinze objets. Dans chaque lot, enregistre le nombre et la proportion d'objets défectueux.

g) As-tu obtenu le même nombre de gadgets défectueux dans chaque échantillon?

h) Combine tes résultats de simulation avec ceux obtenus par les autres élèves de la classe dans un graphique par points comme celui ci-dessous.



## Leçon 6 : La signification statistique (suite)

i) Combien de lots simulés produisent quatre gadgets défectueux ou moins et dans quelle proportion?

j) Selon cette simulation, crois-tu qu'il serait très peu possible que le processus produise un lot contenant quatre gadgets défectueux ou moins alors que la proportion dans la population d'objets défectueux est de un sur trois?

k) Le texte Workshop Statistics de Allan J. Rossman fait référence au programme WIDGT de la calculatrice TI-83 (voir la page I-153, annexe I-3). La calculatrice peut simuler une sélection aléatoire allant jusqu'à 1000 lots de gadgets en tenant pour acquis que le tiers de la population est défectueuse. Si tu as ce programme, lance-le sur la calculatrice. Il te donne un histogramme des résultats, que tu peux transcrire dans le tableau ci-dessous :

| Nbre défectueux | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| Nbre de lots | | | | | | | | |

| Nbre défectueux | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|
| Nbre de lots | | | | | | | | |

1) Combien parmi les 1000 lots simulés contiennent 4 objets défectueux ou moins, et dans quelle proportion?

m) Selon cette simulation étendue, dirais-tu qu'il est très peu possible que le processus produise un lot contenant quatre objets défectueux ou moins alors que la proportion d'objets défectueux dans la population est de un sur trois?

n) Suppose que les ingénieurs ne savent pas si la modification apportée a amélioré le processus de production. Ils échantillonnent un lot de quinze objets et y trouvent quatre objets défectueux. Cela leur donne-t-il une preuve très concluante qu'ils ont amélioré le processus? Explique pourquoi.

o) Suppose maintenant qu'ils trouvent seulement deux objets défectueux dans l'échantillon. Dans quelle fréquence à long terme un résultat si extrême se reproduirait-il si la modification n'a pas amélioré le processus? Fonde ta réponse sur les 1000 lots simulés que tu as générés ci-dessus.

p) Le fait de trouver deux gadgets défectueux dans l'échantillon nous donne-t-il une preuve évidente que la modification a en effet amélioré le processus en diminuant la proportion de gadgets défectueux produits? Explique pourquoi.

q) Réponds aux questions (o) et (p) en tenant pour acquis qu'aucun gadget défectueux n'a été trouvé dans le lot.

## Leçon 6 : La signification statistique (suite)

La ***signification statistique*** nous indique s'il est assez peu probable que le comportement d'un échantillon soit dû à la chance pour qu'on cherche une autre explication. Dans cet exemple, il est très très peu probable qu'un lot ne contienne aucun objet défectueux si la modification n'avait pas amélioré le processus. Si les ingénieurs apportent les modifications puis ne trouvent aucun objet défectueux dans un échantillon, ils croiront soit :

a) que le processus n'a pas été amélioré et que l'échantillon choisi était très peu représentatif, ***ou***

b) que la modification a amélioré le processus et qu'elle a réduit la proportion d'objets défectueux dans la population à moins de un sur trois.

Ces deux possibilités peuvent être vraies, mais le fait d'obtenir un échantillon ne contenant aucun gadget défectueux est extrêmement improbable si on considère la possibilité (a). Ce fait donne beaucoup de poids à la possibilité (b), même si on ne peut éliminer l'autre possibilité complètement.

Le fait de trouver quatre gadgets défectueux est très peu utile pour démontrer si la modification a amélioré le processus parce qu'il n'est pas très inhabituel de trouver quatre objets défectueux ou moins dans un lot alors que le tiers des gadgets produits sont défectueux.

Plus loin, tu étudieras les tests statistiques sur la signification, qui font appel au type de raisonnement illustré ici.

### Sommaire

Tu as vu que les proportions dans les échantillons aléatoires simples forment une distribution symétrique en forme de cloche, distribuée autour de la moyenne d'une population selon la règle empirique. La taille de l'échantillon détermine l'écart type des proportions dans l'échantillon, et de ce fait l'image de la distribution.

> ***Avant de passer à la leçon suivante, discutez du contenu avec un groupe d'élèves.***

## Théorie nº 1 : Distributions binomiales

### Contexte

Dans la leçon 6, nous avons échantillonné une population de bonbons Reese dans laquelle 45 % étaient orange. Nous avons dit que, si nous prenions un nombre infini d'échantillons de taille $n$, alors $\bar{x} = \boldsymbol{\mu}$. (Autrement dit, la moyenne des proportions dans l'échantillon serait égale à la moyenne dans la population.) Il faut étoffer cette affirmation.

Nous utiliserons comme exemple une expérience qui consiste à tirer cinq bonbons au hasard dans une boîte de bonbons Reese.

**Un bonbon est soit orange, soit non orange. Cette expérience met donc en cause une distribution binomiale. Nous appliquerons le développement binomial.**

Étant donné que le mot ou est équivalent au signe + et que cinq bonbons sont choisis, qui sont soit orange, soit non orange, nous développerons l'expression (Orange + Non) 5 comme suit :

$$(Orange + Non)^5$$

$$= {}_5C_0\,(Orn)^5(Non)^0 + {}_5C_1\,(Orn)^4(Non)^1 + {}_5C_2(Orn)^3(Non)^2 + {}_5C_3\,(Orn)^2(Non)^3 + {}_5C_4\,(Orn)^1(Non)^4 + {}_5C_5\,(Orn)^0(Non)5$$

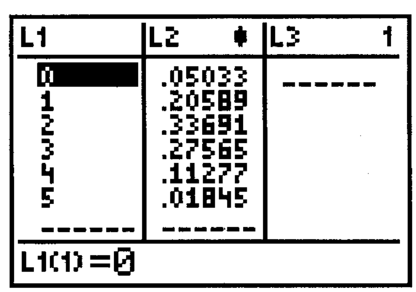
$$= {}_5C_0\,(0{,}45)^5(0{,}55)^0 + {}_5C_1\,(0{,}45)^4(0{,}55)^1 + {}_5C_2\,(0{,}45)^3(0{,}55)^2 + {}_5C_3\,(0{,}45)^2(0{,}55)^3 + {}_5C_4\,(0{,}45)^1(0{,}55)^4 + {}_5C_5\,(0{,}45)^0(0{,}55)^5$$

Cette somme comprend tous les nombres possibles de bonbons orange dans un échantillon de cinq. Elle démontre la distribution de probabilités. Ainsi, la probabilité qu'un échantillon contienne deux bonbons orange est ${}_5C_3\,(Orn)^2(Non)^3$ ou ${}_5C_3\,(0{,}45)^2(0{,}55)^3$, soit environ 0,33 69. La somme du développement est la somme de toutes les probabilités, et devrait être 1.

Nous pouvons utiliser la calculatrice TI-83 pour faire les calculs. Voici un moyen d'y arriver :

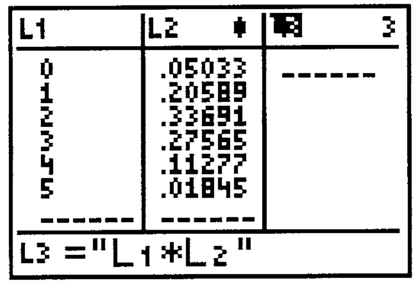
- Efface toutes les listes de statistiques (2nd MEM C1rAllLists ENTER).
- Entre les valeurs 0, 1, 2, 3, 4, 5 dans L1. Il s'agit du nombre de bonbons orange que l'on peut choisir.
- Place le curseur dans le haut de la colonne L2, pour noircir le titre.
- Entre "${}_5C_3 \times 0{,}55^{5\text{-}L1} \times 0{,}45^{L1}$ " dans l'équation L2 puis appuie sur ENTER. L'écran ci-contre s'affichera. La liste 2 affiche la probabilité d'obtenir le nombre de bonbons orange inscrit dans la liste 1.

  ***Remarque :*** Les guillemets indiquent à la calculatrice d'enregistrer la formule. L'astérisque dans le haut indique que cela a été fait.



## Théorie nº 1 : Distributions binomiales

- Quitte l'éditeur de Stats et efface l'écran.
- Appuie sur 2nd LIST et passe au menu MATH.
- Sélectionne l'élément 5 (pour trouver la somme de la liste 2), puis appuie sur 2nd L2 et sur ENTER. La somme devrait être 1.
- Va maintenant à l'en-tête de la liste 3. Inscris l'équation de la liste 3 telle qu'illustré et appuie sur ENTER.




- Trouve la somme de la liste 3. C'est le nombre attendu de résultats favorables de notre expérience, ou le nombre moyen de bonbons orange dans nos échantillons, si nous considérons un nombre infini d'échantillons. Tu devrais obtenir $\mu$ = 2,25.

Les prochaines étapes nous donnent la variance de la distribution.

- Dans la liste 4, inscris les valeurs L1 - 2,25. (Trouve les écarts à la moyenne.)
- Dans la liste 5, inscris les valeurs $L4^2$. (Élève au carré les écarts à la moyenne.)
- Dans la liste 6, inscris les L5 x L2. (Tiens compte de la probabilité pour chaque rangée.)
- Trouve la somme de la liste 6. Il s'agit de la variance de notre distribution de probabilités binomiale. $\sigma$ = 1,2375. L'écart type est la racine carrée de la variance (environ 1,11).

**Une formule nous permet de faire ce travail plus rapidement. L'exemple suivant illustre comment élaborer cette formule.**

Notre exemple consiste à choisir ***un*** bonbon dans une population dans laquelle les bonbons orange ont une ***proportion*** *p*. Nous pouvons appeler ce ***nombre*** de bonbons orange choisi *x*. (Même si on peut choisir un bonbon orange ou zéro bonbon orange, notre ***espérance mathématique*** peut être différente. Par exemple, dans le travail ci-dessus, nous avons trouvé que le nombre moyen de bonbons orange était de 2,25.) Nous dirons que la ***probabilité*** de choisir *x* bonbons orange est $f(x)$.

| $x$ | $f(x)$ | $x \cdot f(x)$ | $x - \mu$ | $(x - \mu)^2$ | $(x - \mu)^2 \cdot f(x)$ |
|---|---|---|---|---|---|
| 0 | $1 - p$ | 0 | $-p$ | $p^2$ | $p^2 \cdot (1 - p)$ |
| 1 | $p$ | $p$ | $1 - p$ | $(1 - p)^2$ | $(1 - p)^2 \cdot p$ |

| | |
|---|---|
| $\sum (x \cdot f(x)) = p$<br>$= 0 + p$<br>pour un bonbon $\mu = p$ | $\sum (x - \mu) \cdot f(x) = p^2 \cdot (1 - p) + (1 - p)^2 \cdot p$<br>$= p(1 - p)(p + 1 - p)$<br>$= p(1 - p)$<br>ainsi $\sigma^2 = p(1 - p)$ |

## Théorie nº 1 : Distributions binomiales (suite)

- La boîte à l'extrême gauche nous indique que le nombre de bonbons orange que nous devrions espérer mathématiquement est équivalent à zéro fois la probabilité d'obtenir une couleur autre que orange plus une fois la probabilité d'obtenir un bonbon orange. Cette ***espérance mathématique*** est la même que la valeur ***moyenne*** des deux résultats, en fonction de la proportion de bonbons orange dans la population.
- La boîte à l'extrême droite utilise la même technique pour nous indiquer la ***variance*** de notre distribution très simple. Certains statisticiens utilisent la lettre $q$ pour remplacer l'expression $(1 - p)$.
- Nous disposons maintenant d'expressions pour la moyenne et l'écart type d'une distribution de probabilités binomiales comprenant un seul événement. Si chacun des événements est indépendant d'autres événements similaires, alors pour une expérience mettant en cause $n$ bonbons, nous pouvons dire que :

$$\boxed{\begin{aligned} \mu &= np \\ \text{and } \sigma^2 &= np(1-p) \\ \text{or } \sigma^2 &= npq \end{aligned}}$$

Nos calculs nous indiquent que la distribution de probabilités binomiales de bonbons orange dans les échantillons de 5 éléments a une moyenne de 2,25 et un écart type de $\sqrt{1{,}2375}$.

Nos formules utilisent le fait que 0,45 de la population de bonbons est orange, et que nos échantillons comprennent 5 bonbons qui nous disent que

$$\begin{aligned} \mu &= np \\ &= 5 \times 0{,}45 \\ &= 2{,}25 \end{aligned}$$

$$\begin{aligned} \text{et } \sigma^2 &= mpq \\ &= 5 \times 0{,}45 \times 0{,}55 \\ &= 1{,}2375 \\ \text{ainsi } \sigma &= \sqrt{1{,}2375} \end{aligned}$$

En utilisant ces formules, nous pouvons trouver la moyenne et l'écart type des distributions de probabilités binomiales plus rapidement. Nous pourrons nous exercer avec des échantillons plus grands sans augmenter le nombre de calculs.

## Théorie nº 1 : Distributions binomiales (suite)

Dans la leçon 6, nous avons travaillé avec des échantillons de 25 bonbons. Nous savons maintenant que pour les échantillons de cette taille, le nombre moyen de bonbons orange sera proche de 11,25 et que l'écart type sera proche de 2,487 (tu peux vérifier cela au moyen de formules). Nous disons la même chose en affirmant que la proportion moyenne de bonbons orange est $\frac{11,25}{25} = 0,45$ et l'écart moyen de la proportion de bonbons orange est $\frac{2,487}{25} = 0,0995.$ En effet, le nombre de bonbons orange dans un échantillon est 25 et la proportion est tout simplement le nombre de bonbons orange divisé par le nombre total.

Voyons maintenant si nous pouvons traduire nos deux formules en proportions.

$$\boxed{\begin{aligned} &\mu = np \\ &\text{alors } \frac{\mu}{n} = \frac{np}{n} \\ &\text{alors } \frac{\mu}{n} = p \end{aligned}}$$

Étant donné que $\frac{\mu}{n}$ est la proportion moyenne de bonbons orange, nous savons que la proportion moyenne sera 0,45.

$$\begin{aligned} \sigma &= \sqrt{npq} \\ \text{alors } \frac{\sigma}{n} &= \frac{\sqrt{npq}}{n} \\ &= \sqrt{\frac{npq}{n^2}} \\ &= \sqrt{\frac{pq}{n}} \end{aligned}$$

Étant donné que $\frac{\sigma}{n}$ est l'écart type de la proportion de bonbons orange, nous pouvons simplifier comme suit :

$\sqrt{\frac{pq}{n}}.$

## Théorie nº 1 : Distributions binomiales (suite)

- Trouve la moyenne et l'écart type pour le ***nombre*** de bonbons jaune choisi si 25 % sont jaunes et si la taille de l'échantillon est 25. Trouve ensuite la moyenne et l'écart type de la ***proportion*** de bonbons jaunes dans les mêmes conditions.

  *Solution 1*

  $$\boxed{\begin{aligned}\mu &= np\\ \sigma &= \sqrt{npq}\end{aligned}}$$

  Nous savons que $n$ = 25, $p$ = 0,25, $q$ = 0,75.

  $$\mu = 6{,}25$$
  $$\text{Ainsi } \sigma = 2{,}165$$

  *Solution 2*

  $$\boxed{\begin{aligned}\mu &= p\\ \sigma &= \sqrt{\frac{pq}{n}}\end{aligned}}$$

  Nous savons que $n$ = 25; $p$ = 0,25.

  $$\mu = 0{,}25$$
  $$\text{Ainsi } \sigma \cong 0{,}0866$$

  Ces résultats nous donnent la même information, parce que 25 × 0,25 = 6,25 et 0,0866 × 25 = 2,165.

## Leçon 7 : Aire sous la courbe normale

### Contexte

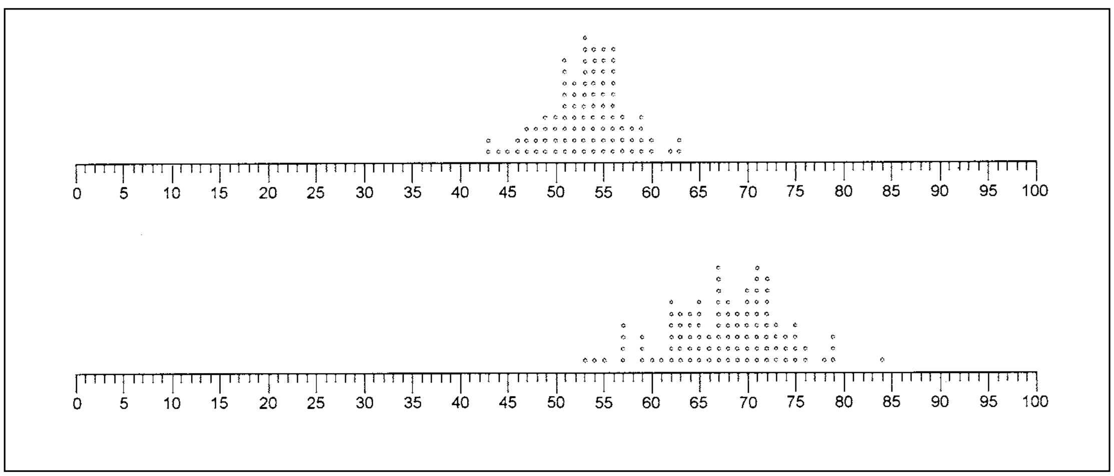
Supposons que les élèves qui ont eu des notes de 90 % ou plus obtiennent un A et que ceux qui ont 60 ou moins obtiennent un F. Dans la classe de Mme Miller, la note moyenne est 78 et l'écart type est 7, alors que dans celle de M. Sapinsky, la moyenne est 74 et l'écart type 18. Dans quelle classe compte-t-on le plus de A?

### Points importants

- Trouver l'aire sous la courbe normale.
- Reconnaître la courbe normale standard et ses caractéristiques.
- Déterminer les cotes $z$ des statistiques d'un échantillon.
- Trouver la probabilité qu'une statistique générée par un échantillon aléatoire simple se trouve dans une étendue donnée pour une caractéristique donnée.

### Activité de mise en marche

Voici deux graphiques liés à des exercices que nous avons faits auparavant. Ils nous indiquent des notes obtenues dans deux classes.




a) De quelle façon les formes de ces distributions sont-elles similaires?

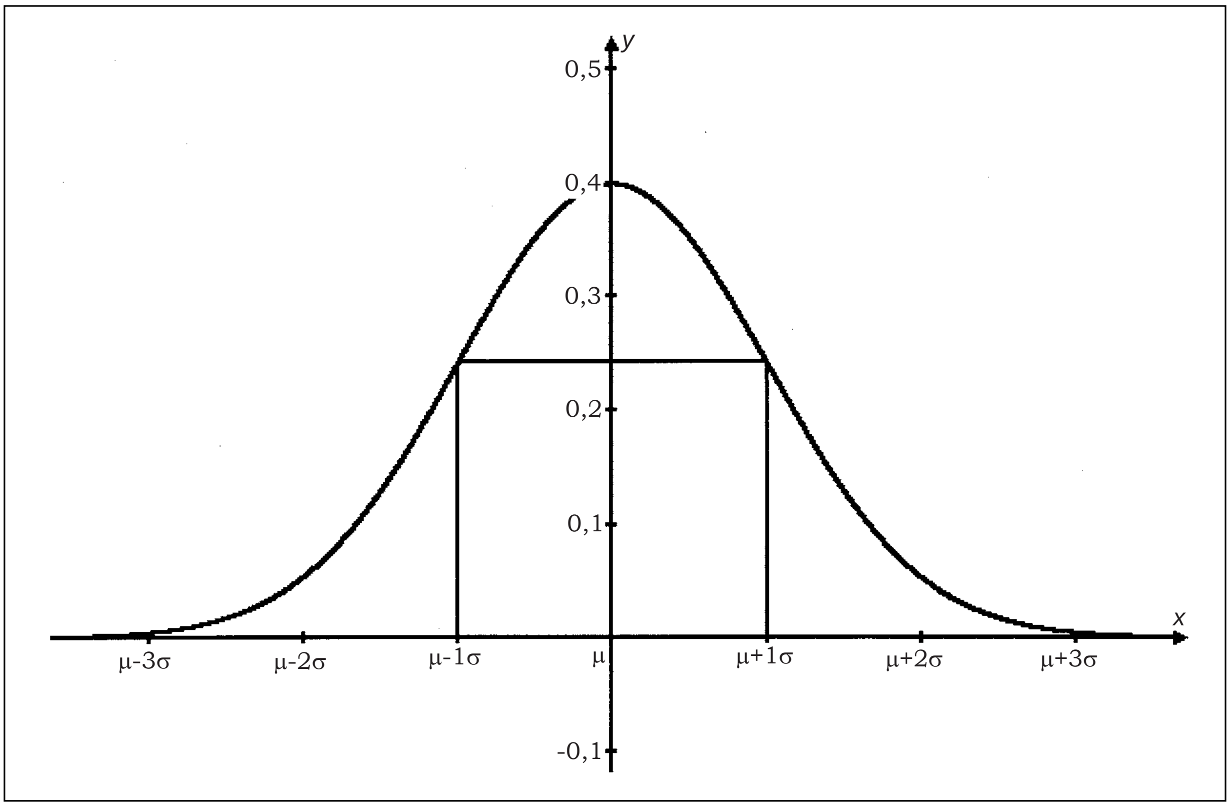
b) Trace une courbe lisse qui illustre la forme générale des deux graphiques par points.

La forme générale que nous avons vue dans ces exemples est fréquente - si fréquente en fait qu'on la considère comme étant normale. Les modèles mathématiques utilisés pour représenter des distributions qui ont cette forme sont appelés des distributions normales. Il en existe toute une famille, tout comme il existe toute une famille de paraboles définies par des fonctions quadratiques. Toutes les distributions normales ont en commun les trois caractéristiques distinctives suivantes :

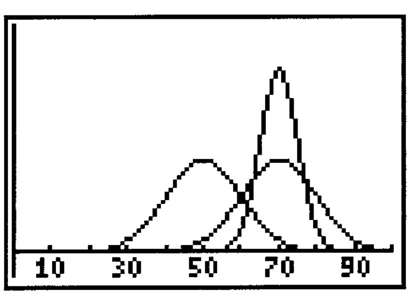
- Toutes les distributions normales sont symétriques.
- Toutes les distributions normales ont une courbe en forme de cloche.
- Toutes les distributions normales ont un seul sommet au centre.

## Leçon 7 : Aire sous la courbe normale (suite)

Toute distribution normale a une ***moyenne*** ($\mu$) et un ***écart type*** ($\sigma$) qui la distinguent. La moyenne nous indique où se trouve le centre, c'est-à-dire au sommet (qui marque le centre de la symétrie). L'écart type indique l'étendue de la distribution. Vis-à-vis de la moyenne, la courbe est concave vers le bas; elle devient concave vers le haut à une distance de un écart type de la moyenne. (Si tu conduisais une auto le long de la courbe à partir de la gauche, tu tournerais le volant vers la gauche jusqu'à ce que tu arrives au point au-dessus de $\mu - \sigma$. Tu tournerais ensuite le volant vers la droite et tu le garderais dans cette position jusqu'au point au-dessus de $\mu + \sigma$, où tu tournerais le volant vers la gauche.) Le croquis ci-dessous illustre l'apparence de toutes les courbes normales.



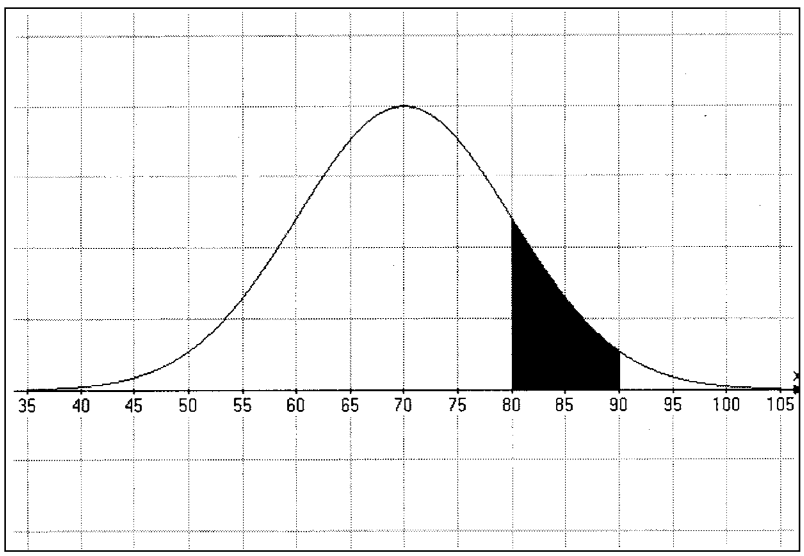

c) Voici un affichage de trois courbes normales. La courbe A a une moyenne de 70 et un écart type de 5. La courbe B a une moyenne de 70 et un écart type de 10. La courbe C a une moyenne de 50 et un écart type de 10. Marque chacune à l'aide de la lettre appropriée.



## Leçon 7 : Aire sous la courbe normale (suite)

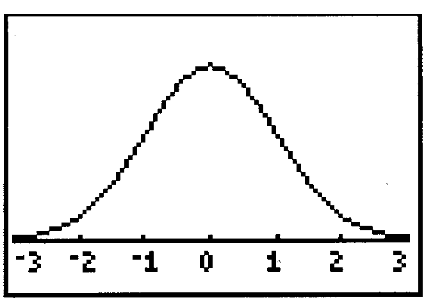
Nous pouvons faire une bonne approximation de la proportion des membres d'une population qui se trouvent à l'intérieur d'un certain intervalle de valeurs. Nous appelons cette proportion la probabilité qu'un certain membre de la population ait une valeur à l'intérieur de cet intervalle  quant à la caractéristique à l'étude. Le fait de trouver cette probabilité est assimilable au fait de trouver l'aire sous la courbe normale dans un intervalle donné (l'aire totale sous toutes les courbes normales est toujours 1.)

En théorie, si les notes à un examen sont distribuées normalement, avec une moyenne de 70 et un écart type de 10, la proportion des notes qui se trouvent entre 80 et 90 devrait être équivalente à l'aire sous la courbe normale, tel qu'illustré ci-dessous. Cette aire pourrait être trouvée à l'aide de calculs mais la calculatrice TI-83 le fait pour nous.




Toutes les distributions normales ont leur propre moyenne et leur propre écart type. Cependant, une distribution normale particulière, appelée la distribution normale standard (où $\mu = 0$ et $\sigma = 1$), a été étudiée plus étroitement. On lui a donné le nom $z$. Nous allons étudier la distribution $z$, trouver les aires sous certaines régions de sa courbe et utiliser des transformations pour décrire des aires sous d'autres distributions normales en termes de la courbe $z$.

Voici un croquis d'une distribution normale standard, la courbe $z$.




La moyenne de la distribution est 0. L'échelle de l'axe des $x$ correspond au nombre d'écarts types par rapport à la moyenne.  Plus de trois écarts types de chaque côté de la moyenne sont affichés, de sorte que le croquis comprend plus de 99,9 % de l'aire sous la courbe, et plus de 99,9 % de la population à l'étude.

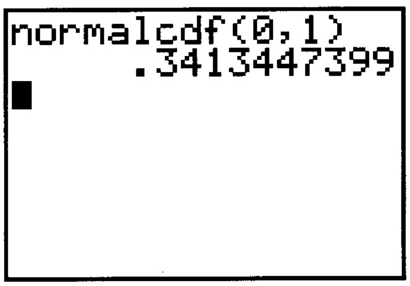
## Leçon 7 : Aire sous la courbe normale (suite)

### Déroulement

Pour trouver l'aire sous des régions de la courbe normale standard, les statisticiens ont jusqu'à tout récemment utilisé des tables. Ces tables donnent la liste des ***cotes z*** qui indiquent la distance en termes d'écarts types d'une valeur par rapport à la moyenne de la population. (On trouve une table de ce genre à la fin du document). Une cote $z$ de 1,0 indique qu'une valeur se trouve à l'écart type au-dessus de la moyenne, et la table démontre que la proportion d'une proportion distribuée normalement entre la moyenne et la cote $z$ de 1,0 est 0,3413 ($Pr(0 < z < 1) = 0{,}3413$). Autrement dit, environ 34 % de la population se trouvent entre la moyenne de la population et un écart type au-dessus.

À l'aide de ta calculatrice TI-83, calcule l'aire sous la courbe normale standard entre la moyenne et une cote $z$, comme suit :

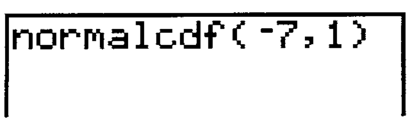
- Appuie sur 2nd et DISTR, puis sélectionne normalcdf dans le menu et inscris 0, 1 et appuie sur ENTER. L'écran suivant devrait être affiché.




Cela signifie que l'aire sous la courbe entre la moyenne et 1 écart type au-dessus de la moyenne est 0,3413447399. Essaie quelques autres valeurs à l'aide de la table et de la calculatrice, pour t'assurer que tu obtiens des résultats équivalents.

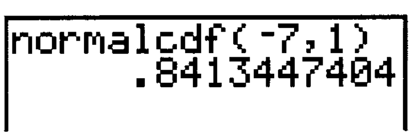
***cdf*** signifie ***cumulative density function*** (fonction de densité cumulative).

***Remarque :*** Cette calculatrice peut simplifier certaines tâches. Par exemple, si tu veux trouver l'aire totale sous la courbe à la gauche d'un point, entre les données suivantes (l'exemple utilise le point se trouvant à un écart type au-dessus de la moyenne) :




Tu calculeras ainsi l'aire entre un point se trouvant à sept écarts types à gauche de la moyenne et notre point situé à un écart type à la droite de cette moyenne.

Le chiffre –7 représente le point situé à 7 écarts types à la gauche de la moyenne. La calculatrice TI-83 ne peut trouver aucune aire à la gauche de ce point. Tout chiffre inférieur à –7 produit le même résultat. En fait, on pourrait utiliser –5 avec 6 décimales. La courbe normale standard est ***très*** près de l'axe des $x$ à une si grande distance de la moyenne.



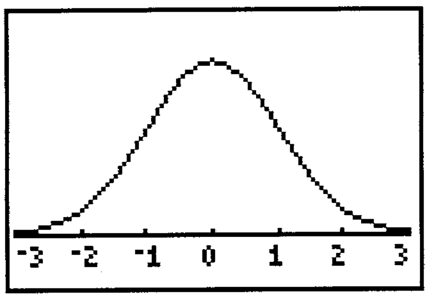

C'est la somme de la moitié de l'aire vers la gauche de la moyenne et du point 0,3413 de l'aire entre la moyenne et le point situé à un écart type à sa droite.

Cela nous indique que $Pr(z < 1) = 0{,}84134$. (La proportion de la distribution normale standard qui est située à gauche de un écart type à droite de la moyenne est 0,84134).

## Leçon 7 : Aire sous la courbe normale (suite)

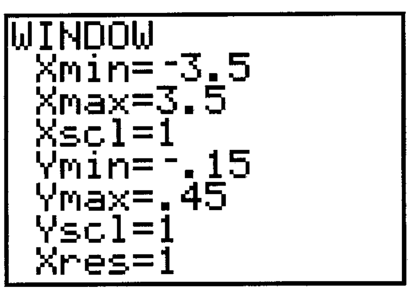
Quand on calcule des probabilités associées à la courbe normale standard, il vaut mieux tracer la courbe sur papier. Tu te souviendras ainsi que les probabilités sont liées à des aires, et tu peux vérifier si ta réponse est correcte.

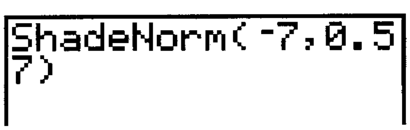
Trace des croquis comme celui ci-dessous et ombrage l'aire appropriée définie par les questions suivantes. Tu peux estimer quelle proportion de l'aire se trouvant sous la courbe a été ombragée, puis tu peux utiliser ta calculatrice (ou la table) pour vérifier si tu étais prêt de la réponse.

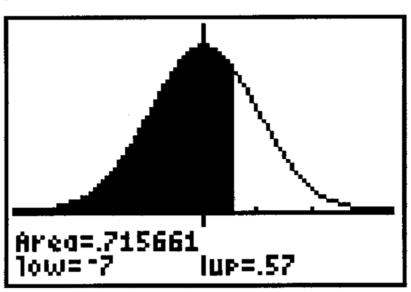



a) Pr($z$ < 0,57)
b) Pr($z \le$ 0,57) (Cette aire est-elle différente de celle trouvée en (a)?)
c) Pr($z$ > 0,57)
d) Pr($z$ < –0,57) (Y a-t-il plus d'une façon d'y arriver?)
e) Pr(–057 < $z$ < 057)
f) Pr($z$ < –3,97) (Peux-tu trouver la réponse exacte?)

La calculatrice TI-83 peut tracer la courbe et ombrager les aires au-dessous. Quand tu as tracé le croquis pour tes réponses à la question ci-dessus, utilise ta calculatrice pour comparer tes croquis. Voici comment tu peux répondre à la question (a) avec la calculatrice.



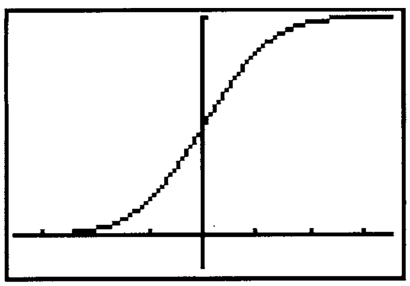

Règle la fenêtre de graphique comme ci-contre et assure-toi que tous les dessins, points et équations sont effacés ou désactivés. Ensuite, à partir de l'écran d'accueil, appuie sur 2nd et DISTR, puis sur la flèche vers la droite pour choisir le menu DRAW. Sélectionne l'élément 1 puis appuie sur ENTER. Après la parenthèse ouvrante, entre -7,0.57, ce qui indique à la calculatrice d'ombrager toute la région à gauche de la moyenne et à 0,57 écart type à la droite de la moyenne. Appuie sur ENTER pour afficher le graphique. La valeur estimée auparavant était-elle près de 0,7? La question (b) aura-t-elle une réponse différente? Explique pourquoi.

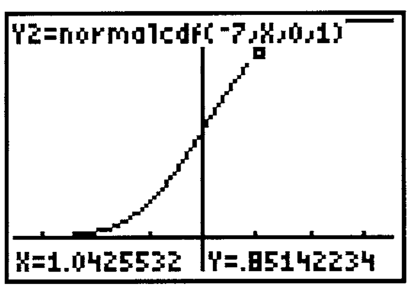
(L'option normalcdf(–7, 0.57) donne-t-elle le même résultat?)

## Leçon 7 : Aire sous la courbe normale (suite)

Dans ces questions, tu as trouvé la proportion de la population en tenant compte de valeurs $z$ limitatives. Si on inversait la question? Comment pourrais-tu trouver $k$, la valeur $z$ ou $\Pr(z < k) = 0{,}8546$? À quel point sur la courbe l'aire au-dessous de la courbe et à sa gauche est-elle équivalente à 0,8546?

- Recommence de nouveau en effaçant tous les dessins, équations et points. Utilise la même fenêtre qu'auparavant, ***mais remplace Ymax par 1***.
- Dans une rangée vide de l'éditeur d'équations, appuie sur 2nd et DISTR, sélectionne normalcdf(, puis appuie sur ENTER. Entre –7, x, 0, 1) dans l'équation pour indiquer à la calculatrice de tracer le graphique de la fonction de probabilité cumulative normale pour toutes les valeurs de $x$ telles que la moyenne est 0 et l'écart type est 1. Appuie sur la touche GRAPH pour afficher le graphique.






Cette courbe atteint un plateau à une hauteur de 1. Si tu appuies sur la touche TRACE et que tu te déplaces le long de la courbe de gauche à droite, les hauteurs indiquent l'aire totale sous la courbe normale standard vers la gauche des valeurs $z$ des points de tracé. L'écran illustré ci-contre indique que l'aire de 0,85 se trouve à la gauche de la valeur $x$ (valeur $z$) de 1,04. Si tu fais un zoom sur le point de tracé, tu peux arriver très près d'une hauteur de 0,8546 sur la courbe, et lire la valeur de $x$ correspondante (valeur z) à l'écran.

- Utilise cette méthode pour répondre aux questions suivantes :

  g) Trouve la valeur de $k$ telle que $\Pr(z < k) = 0{,}7258$.

  h) Trouve la valeur de $k$ telle que $\Pr(z > k) = 0{,}0625$.

### Normalisation de la distribution normale

Toute courbe normale peut être normalisée à l'aide d'une formule de transformation simple. La voici :

$$z = \frac{x - \mu}{\sigma}.$$

Si tu inverses les termes pour obtenir $x = z\sigma + \mu$, tu obtiens une équation similaire à l'équation de la droite $y = mx + b$.

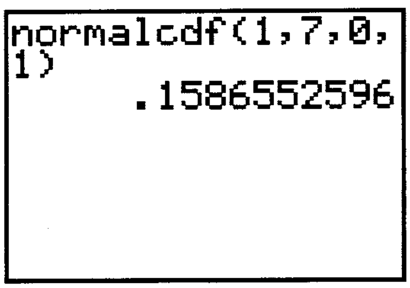
## Leçon 7 : Aire sous la courbe normale (suite)

Dans cette formule, le $x$ représente toute distribution normale, $\mu$ représente sa moyenne et $\sigma$ l'écart type. Tu peux trouver la valeur de $z$ (la distribution normale standard) en effectuant des transformations sur les valeurs de $x$. Autrement dit, pour chacune des valeurs de $x$ dans la distribution originale, une valeur de $z$ correspondante peut être trouvée dans la distribution normale standard au moyen de la formule.

À l'origine, les tests de QI étaient évalués en posant $\mu = 100$ et $\sigma = 15$. Quelle proportion de la population aurait obtenu un QI supérieur à 115?

$$z = \frac{115-100}{15} = 1$$

Par conséquent, la note 115 se trouve à un écart type au-dessus de la moyenne. L'écran ci-dessous nous indique que l'aire entre ce point et un point bien au-dessus de la moyenne est environ 0,159, ou environ 16 % de l'aire sous la courbe normale standard globale. Ainsi, environ 16 % de la population aurait obtenu un QI de plus de 115.




a) Trouve la proportion de la population dont le QI se situe entre 110 et 125.

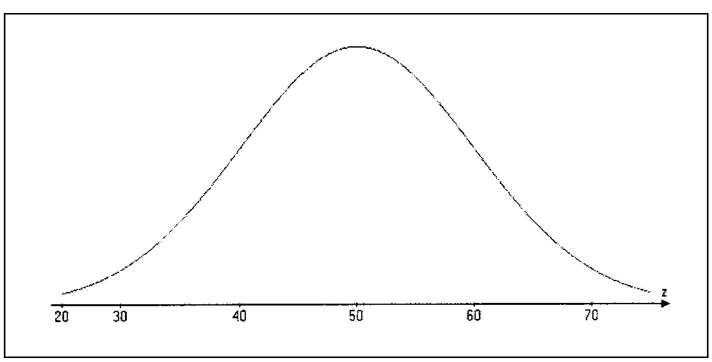
b) Forrest Gump avait un QI de 75. Quel pourcentage de la population aurait un QI inférieur à lui?

c) Quel QI devrait avoir une personne pour appartenir au 1 % supérieur de la population?

### Exercice

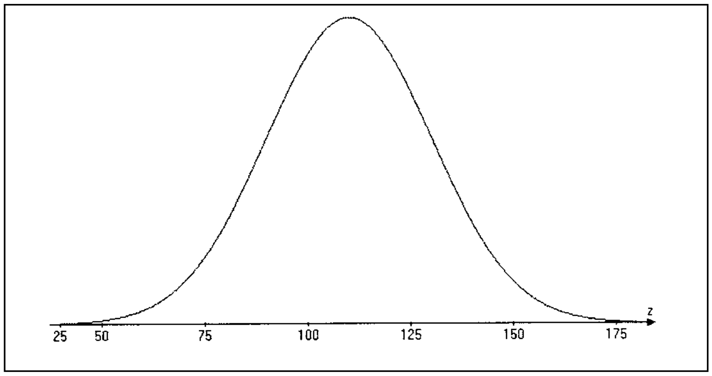
1. Pour chacune des courbes normales suivantes, trouve (aussi précisément que possible à partir du graphique) la moyenne $\mu$ et l'écart type $\sigma$ de la distribution.

   a)

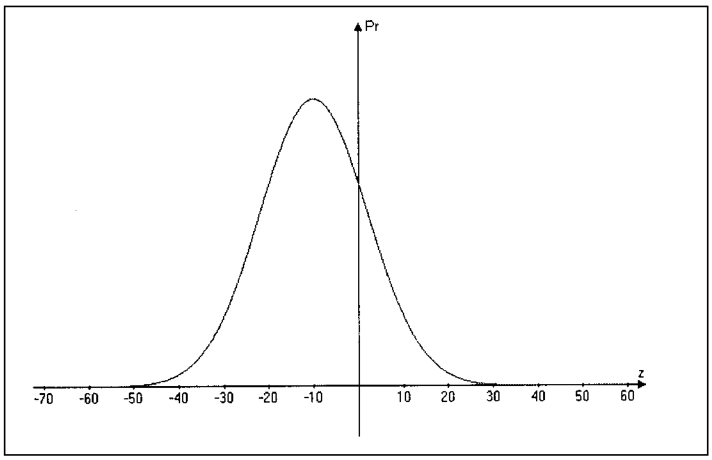


## Leçon 7 : Aire sous la courbe normale (suite)

b)




c)




2. La distribution de la durée de la grossesse chez les humaines est relativement normale, avec une moyenne $\mu$ = 266 et un écart type $\sigma$ = 16 jours. Détermine la proportion de toutes les grossesses qui durent :
   a) moins de 244 jours (8 mois)
   b) plus de 275 jours (ou 9 mois)
   c) plus de 300 jours (ou 10 mois)
   d) entre 260 et 280 jours
3. Au début de la leçon, tu as lu des données sur les classes de Mme Miller et de M. Sapinsky. Tu connais la note moyenne et l'écart type pour chacune.
   a) Trace un graphique de la distribution des notes pour chaque classe, en utilisant la même échelle.
   b) Quelle classe a obtenu la plus grande proportion de A? Indique les calculs effectués pour trouver ta réponse.
   c) Quelle classe a obtenu la plus grande proportion de F? Indique les calculs effectués pour trouver la réponse.

## Leçon 7 : Aire sous la courbe normale (suite)

### Sommaire

Dans cette leçon, tu as exploré les distributions normales et leurs caractéristiques communes. Tu as aussi étudié l'équation de transformation qui permet de définir chaque distribution par rapport à la distribution normale standard, $z$. Tu as appris à utiliser ta calculatrice (ou la table fournie dans le manuel) pour calculer la proportion d'une aire sous la courbe dans une image donnée de cotes $z$, et à résoudre des problèmes sur les proportions de populations en fonction de certaines caractéristiques.

***Avant de passer à la leçon suivante, discutez du contenu avec un groupe d'élèves.***

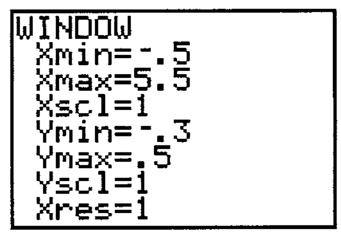
## Théorie 2 : De la fonction binomiale à la fonction normale

### Contexte

Nous pouvons composer des formules qui nous permettront de résoudre des fonctions de distribution de probabilité binomiale. Nous avons par la suite étudié le fonctionnement des fonctions de distribution de probabilité. Dans la présente section, nous allons examiner les liens entre ces deux types de fdp.

Nous reprendrons le « problème des bonbons » que nous avons traité dans la section Théorie des distributions binomiales. Nous avons résolu la distribution de probabilité binomiale par rapport au nombre de bonbons orange dans des échantillons comprenant 5 éléments, en supposant que 45 % des bonbons sont orange. Nous passerons maintenant en revue une partie du problème brièvement à l'aide de la calculatrice, puis nous aborderons le problème comme s'il s'agissait d'une fonction de distribution de probabilité normale. Suis les étapes suivantes :

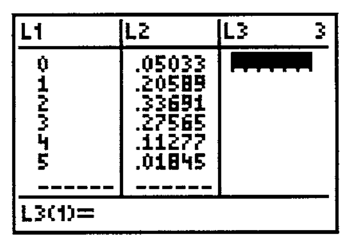
Efface toutes les listes dans ta calculatrice TI-83. Règle la fenêtre tel qu'illustré.



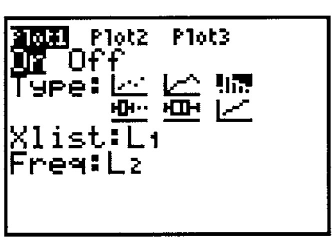

Entre les entiers 0, 1, 2, 3, 4, 5 dans la liste 1.

Entre la formule suivante dans la liste 2.

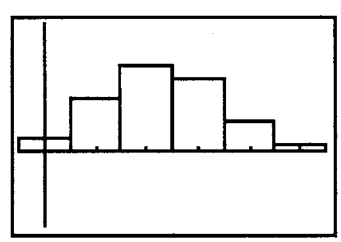
L2 = ${}_5C_{L1}$ x $0{,}55^{(5-L1)}$ x $0{,}45^{L1}$. Appuie sur ENTER pour obtenir l'écran suivant.




Règle maintenant STAT PLOT 1 comme illustré ci-dessous.

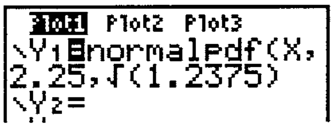



Quand tu appuies sur la touche GRAPH, tu obtiens l'écran suivant. Il illustre les probabilités associées au nombre de bonbons orange (de 0 à 5) dans un échantillon.

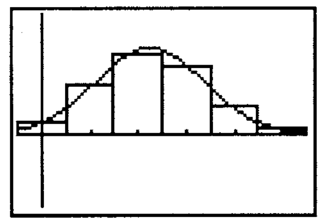


## Théorie 2 : De la fonction binomiale à la fonction normale (suite)

Entre maintenant l'équation indiquée ici dans l'éditeur d'équation. Tu obtiens le graphique de la fonction de distribution de probabilité normale pour toutes les valeurs de $x$ (ce qui représente le « nombre » de bonbons orange dans un sens continu), avec une moyenne de 2,25 et un écart type de $\sqrt{1{,}2375}$.




Quand tu appuies sur la touche GRAPH, tu obtiens cette image.



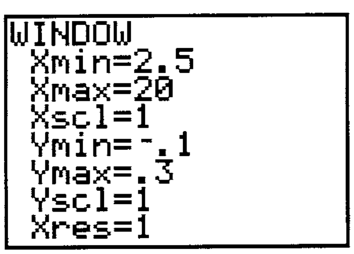
Il semble y avoir une similarité entre les deux graphiques. Étant donné que la fonction binomiale est discrète, elle est composée de barres, et la fonction normale étant continue, la courbe est lisse.

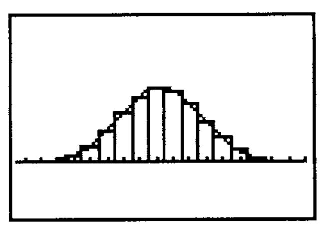
Examinons maintenant des échantillons de 25 bonbons. Pour le graphique de la fonction binomiale, inscris dans la liste 1 les entiers de 0 à 25. Dans la liste 2, inscris la formule ${}_{25}C_{\mathrm{L1}} \times 0{,}55^{(25-\mathrm{L1})} \times 0{,}45^{\mathrm{L1}}$. Étant donné que nous savons que

$$\boxed{\begin{aligned}\mu &= 25 \times 0{,}45\\ &= 11{,}25\end{aligned}} \quad \text{et} \quad \boxed{\sigma = \sqrt{25 \times 0{,}45 \times 0{,}55}}$$

l'équation du graphique normal ira comme suit :


Si la fenêtre est réglée comme suit, notre graphique prendra l'apparence suivante :



Remarque que le graphique de la fonction binomiale contient plus de barres. Les deux graphiques se ressemblent beaucoup.

Applique la même procédure avec un échantillon de 75. Il faudra faire certaines modifications aux calculs et aux réglages de la fenêtre. Il te sera peut-être difficile de voir la différence entre la courbe normale et le tracé du graphique de la fonction binomiale. Vois par toi-même!

## Théorie 2 : De la fonction binomiale à la fonction normale (suite)

### Établir des liens

Dans chacun des trois problèmes énoncés à la page I-106, l'aire totale des barres dans le graphique de la fonction binomiale est 1. En effet, la somme de toutes les probabilités est 1. De même, l'aire sous la courbe normale est 1 et, plus la taille de l'échantillon est importante, plus la courbe normale se rapproche du tracé des barres du graphique binomial. Parce qu'elle est plus facile à utiliser, les statisticiens ont utilisé la fdp normale comme approximation de la fdp binomiale. Même actuellement, bien que nous ayons des calculatrices qui calculent les probabilités binomiales beaucoup plus rapidement que la plupart des gens ne l'auraient cru possible un jour, il est beaucoup plus facile d'utiliser les probabilités normales avec un grand échantillon. Elles sont calculées par intégration et sont beaucoup moins difficiles que les développements binomiaux de grande envergure.

### Détails importants

- Dans le graphique à barres des fonctions binomiales, chacune des barres est de largeur 1. L'échelle commence à 0,5 ou 2,5 ($n$,5 en règle générale). De ce fait, le centre de chaque barre se trouve sur une valeur entière.
- L'aire de chaque barre est équivalente à la probabilité d'obtenir ce nombre entier de bonbons orange dans un échantillon. Par exemple, si un échantillon est de taille 25, la barre qui se trouve au-dessus de l'espace 11,5 à 12,5 représente la possibilité de trouver 12 bonbons orange dans un échantillon. L'aire est 0,151110. (Étant donné que la largeur de la barre est de 1 unité, la longueur est aussi 0,151110.)
- Étant donné que la courbe normale est lisse, sa hauteur n'est pas exactement égale à la probabilité souhaitée.
- De fait, il faut trouver l'aire sous la courbe entre 11,5 et 12,5. L'écran suivant montre les calculs nécessaires pour trouver l'aire sous la courbe entre 11,5 et 12,5, avec une moyenne de 11,5 et un écart type qui correspond à la racine carrée du produit affiché.

```
normalcdf(11.5,1
2.5,11.25,√(25*.
45*.55)
      .1523208618
```

Bien que la valeur obtenue ne soit pas exactement égale à la valeur binomiale, la différence apparaît seulement à la troisième décimale. Dans les échantillons plus larges, cette différence diminue.

### En conclusion

Nous pouvons utiliser la fdc pour trouver les probabilités dans les problèmes mettant en cause des binômes et un grand échantillon (≥ 5) et une distribution relativement normale (symétrique et en forme de cloche, avec des queues courtes). La résolution exige ainsi moins de calculs. De fait, les limites de la calculatrice TI-83 peuvent être atteintes et même dépassées avec des échantillons de taille courante. (Par exemple, si on essaie de trouver la fdc binomiale pour un échantillon de 1000 éléments).

## Théorie 2 : De la fonction binomiale à la fonction normale (suite)

## Exercice

- Le problème des bonbons

  $n = 5$

  $\bar{x} = 2{,}25$

  $x \approx 1{,}1124$

  Trouve Pr(3 orange)

  Fonction binomiale : $\Pr(3) = {}_5C_3(0{,}45)^3(0{,}55)^2 \approx 0{,}27565$

  ***ou*** $\Pr(3) =$ binompdf(5,0.45,3) $\approx 0{,}27565$

  Fonction normale : $\Pr(3) =$ normalcdf(2.5,3.5,2.25, $\sqrt{5\times 0{,}45\times 0{,}55}$) $\approx 0{,}28052$

  Remarque que pour la ***fdp*** binomiale, l'aire de la barre est centrée à 3. Elle est discrète. La courbe normale étant continue, il faut trouver l'aire cumulative sous la courbe entre 2,5 et 3,5 en utilisant la ***fdc*** normale. Dans les 2 cas, il est prédit qu'environ 28 % des échantillons de 5 bonbons contiendront 3 bonbons orange.

- Le problème des 25 bonbons

  $n = 25$

  $\bar{x} = 11{,}25$

  $s \approx 2{,}4875$

  Trouve Pr(10 orange)

  Fonction binomiale : P(10) = binompdf(25,0.45,10) $\approx 0{,}141889$

  Fonction normale : P(10) = normalcdf(95,105,11.25, $\sqrt{5\times 0.45\times 0.55}$) $\approx 0{,}140649$

  Les 2 méthodes nous montrent que nous devons nous attendre à ce qu'environ 14 % des échantillons de 25 bonbons contiennent 10 bonbons orange.

- Le problème des 75 bonbons : Vérifie si les deux méthodes aboutissent à des probabilités qui sont très près l'une de l'autre.
- Le problème des 1000 bonbons : Peux-tu trouver la probabilité d'obtenir 450 bonbons orange dans un échantillon de 1000? Essaie les deux méthodes.

## Leçon 8 : Le Théorème de la limite centrale

### Contexte

Nous avons vu diverses distributions, nous avons effectué des simulations de certaines à l'aide de nombreux échantillons et nous avons découvert que la courbe normale peut être utilisée pour décrire la distribution des proportions dans l'échantillon. Nous allons maintenant concrétiser certaines techniques que nous avons commencé à utiliser en présentant le ***Théorème de la limite centrale*** appliqué à une proportion dans l'échantillon.

Si un échantillon aléatoire simple de taille $n$ est tiré d'une grande population dans laquelle l'attribut étudié a une proportion $p$, alors si l'échantillon de taille $n$ est suffisamment important (30 ou plus), la distribution des proportions dans l'échantillon, $\hat{p}$ est approximativement normale, avec une moyenne $p$ et un écart type $\sqrt{\frac{pq}{n}}$.

## Points importants

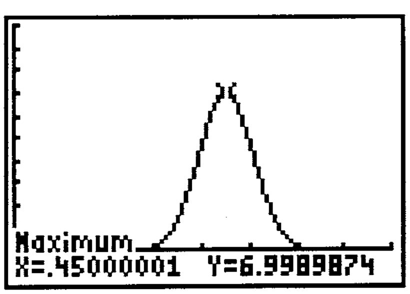
- Utiliser le Théorème de la limite centrale pour répondre à des questions déjà résolues de façon expérimentale.
- Utiliser le Théorème de la limite centrale pour répondre à des questions dont nous ne connaissons pas encore la réponse.

## Activités de mise en marche

Revenons à l'expérience sur les bonbons Reese et appliquons ce théorème. Nous utiliserons un échantillon de taille $n = 75$ et nous tiendrons pour acquis que 45 % des bonbons Hershey sont orange.

a) Selon le Théorème de la limite centrale (TLC), de quelle façon la proportion dans l'échantillon de bonbons orange varierait-elle d'un échantillon à l'autre? Décris non seulement la forme de la distribution, mais aussi la moyenne et l'écart type. Trace un croquis de la distribution.

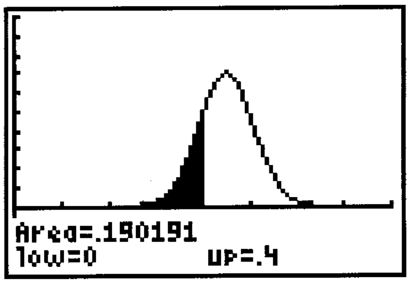
Selon le TLC, la moyenne de la distribution des moyennes de l'échantillon serait 0,45 et l'écart type $\sqrt{\frac{(0,45)(0,55)}{75}} \approx 0,057$. La distribution serait normale. La moyenne et l'écart type sont probablement très près de ta réponse.




Voici un croquis. L'échelle verticale comporte des marques de pointage à des intervalles de 0,1 unité. La fenêtre affiche Xmin = 0, Xmax = 0.8. Ymin = –3 et Ymax = 10.

## Leçon 8 : Le Théorème de la limite centrale (suite)

b) Ombre l'aire sous la courbe que tu as tracée pour montrer la probabilité que la proportion de bonbons orange dans l'échantillon soit inférieure à 0,4. Estime la valeur de la probabilité.



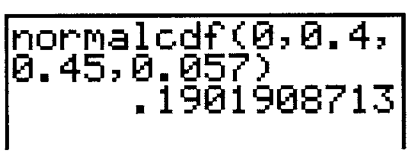

As-tu tracé cette courbe et ombré l'aire avant d'utiliser ta calculatrice pour obtenir cette image? Tu peux utiliser la commande ShadeNorm(0, 0,–1, 0.45, 0.057).

c) Utilise ta calculatrice pour trouver la probabilité que la proportion de bonbons orange dans l'échantillon soit inférieure à 0,4. Utilise une formule de normalisation et soit la table des probabilités normales standard, soit ta calculatrice. Une autre méthode consiste à utiliser la calculatrice et la moyenne ainsi que l'écart type obtenus dans la partie (a).

La formule de normalisation nous donne ce calcul :

$$\Pr(\hat{p} < 0{,}40) = \Pr\left(z < \frac{0{,}40 - 0{,}45}{0{,}057}\right).$$

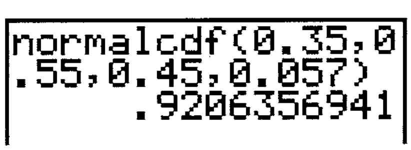
La question porte maintenant sur la courbe standard normale. Nous pouvons maintenant trouver l'aire sous la courbe et à la gauche de la cote $z$ –0,87. Pr($z$ < 0,87) = 0,19. Nous pouvons aussi utiliser l'écran suivant de la calculatrice pour accélérer le travail.




Ainsi, que nous utilisions la transformation pour trouver l'aire appropriée sous la courbe standard normale ou l'information que nous donne cette courbe normale particulière, nous arrivons au même résultat : environ 19 % des échantillons comporteront moins de 40 % de bonbons orange.

d) Calcule la probabilité qu'une proportion de bonbons orange dans l'échantillon se trouve entre 0,35 et 0,55.

Le fait d'utiliser ces deux méthodes pour résoudre des problèmes démontrera encore mieux qu'elles sont équivalentes. Voici l'écran de la solution :

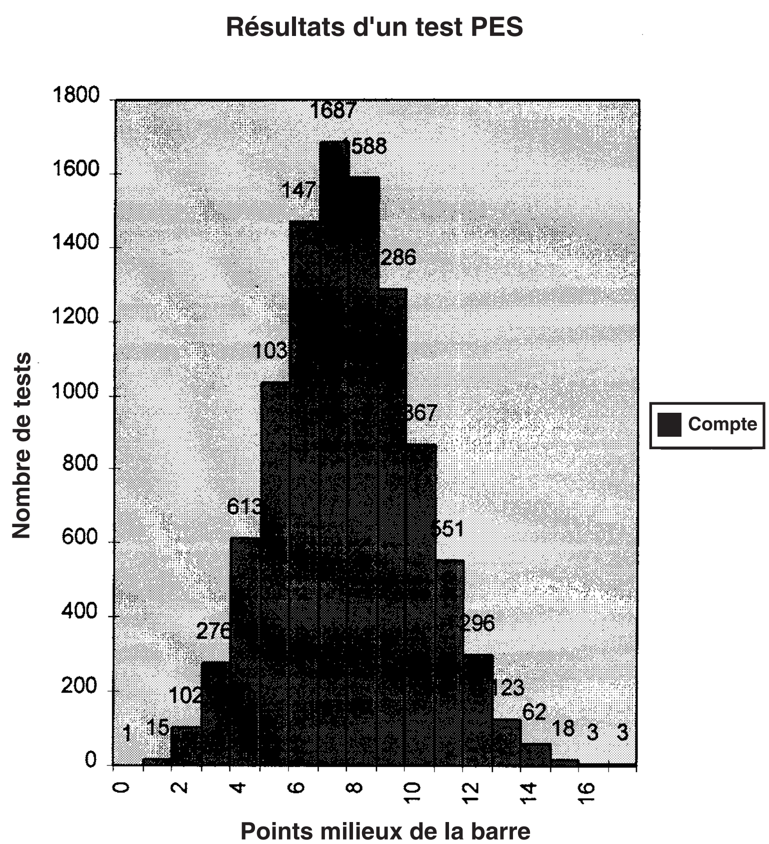


## Leçon 8 : Le Théorème de la limite centrale (suite)

e) Compare cette probabilité avec le résultat que tu as obtenu à la question j) dans la section Déroulement de la leçon 6. Dans cette question, on te demandait de trouver la même information au moyen d'une expérience mettant en cause un échantillon de 10 000 éléments.

### Déroulement

Quand on mesure la perception extrasensorielle (PES), on demande aux sujets de reconnaître laquelle parmi quatre formes (cercle, carré, étoile ou vague) figure sur une carte qu'ils ne peuvent pas voir. Posons un test qui exige la reconnaissance de 30 cartes. Si le sujet se contente de deviner, on s'attend à ce qu'il devine 25 % des cartes sur l'ensemble de l'expérience ($p$ = 0,25). Voici un graphique à barres illustrant les résultats de 10 000 tests simulés dans lesquels les sujets ont deviné les cartes au hasard.




a) Si un sujet devine, que nous révèle le TLC en ce qui a trait à la distribution d'échantillonnage de la proportion de réponses correctes dans l'échantillon? Donne la moyenne et l'écart type de la distribution, et décris sa forme.

b) Trace un croquis à main levée de cette distribution d'échantillonnage. Sur ton dessin, ombre l'aire qui correspond à la probabilité qu'un sujet qui devine obtienne 38 % ou plus de réponses correctes. À partir de cette aire, estime la valeur de la probabilité.

## Leçon 8 : Le Théorème de la limite centrale (suite)

c) Utilise la formule de normalisation et la table des probabilités normales standard pour calculer la probabilité qu'un sujet qui devine obtienne 38 % ou plus de réponses correctes. Utilise la notation suggérée dans les exemples des activités de mise en marche.

d) Utilise ta calculatrice TI-83 pour obtenir la même information.

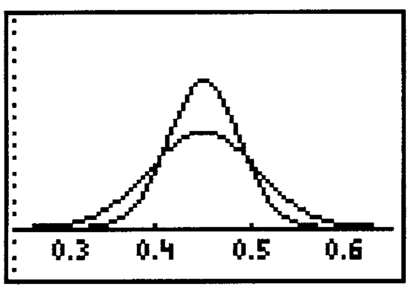
e) Étudie les 10 000 tests simulés de PES. Compare la probabilité obtenue en (d) avec la proportion des 10 000 tests simulés, dans lesquels un sujet qui devine a obtenu 38 % ou plus de réponses correctes. Les résultats sont-ils raisonnablement proches?

f) Serait-il surprenant qu'un sujet obtienne 38 % ou plus de réponses correctes s'il se contentait de deviner? Un tel événement surviendrait-il moins de 10 % du temps à long terme? Serait-il possible moins de 1 % du temps à long terme?

## Exercice

1. Les questions suivantes portent de nouveau sur les bonbons Reese. Cette fois-ci, nous tiendrons pour acquis que les échantillons ont une taille de $n = 175$ et que les bonbons orange représentent encore 45 % de la population.

   a) Selon le TLC, quelle serait la dispersion de la proportion de bonbons orange d'un échantillon à l'autre? Décris non seulement la forme de la distribution, mais aussi la moyenne et l'écart type. Qu'est-ce qui diffère par rapport à un échantillon de 75?

   b) 


   Ce croquis illustre les distributions d'échantillonnage des échantillons de 75 et de 175 éléments. Lequel est lequel? Ombre l'aire correspondant à la probabilité qu'une proportion de bonbons orange dans l'échantillon soit inférieure à 0,4 (pour un échantillon de 175 éléments). Estime la valeur de cette probabilité.

   c) Utilise la technique de la normalisation pour trouver la cote $z$ de la proportion 0,4 de bonbons orange. Utilise ensuite soit la table, soit ta calculatrice pour calculer la probabilité qu'une proportion de bonbons orange dans l'échantillon soit inférieure à 0,4 (pour un échantillon de 175 éléments). De quelle façon cette probabilité est-elle différente de celle obtenue pour un échantillon de 75 éléments?

   d) Utilise ta calculatrice sans normaliser les données pour trouver les réponses aux questions de la partie (c).

   e) Calcule la probabilité qu'une proportion de bonbons orange dans l'échantillon (échantillon de 175 éléments) se trouve entre 0,35 et 0,55. De quelle façon le résultat obtenu est-il différent de celui obtenu pour un échantillon de 75 éléments?

## Leçon 8 : Le Théorème de la limite centrale (suite)

2. Les questions suivantes portent sur l'expérience de PES.
   a) Si un sujet devine, obtiendra-t-il toujours 25 réponses correctes? Explique pourquoi.
   b) Selon le théorème de la limite centrale, de quelle façon la proportion de réponses correctes varierait-elle d'un sujet à l'autre si tous les sujets se contentaient de deviner? Précise la forme de la distribution de même que la moyenne et l'écart type. Trace le graphique de la distribution.
   c) Utilise le TLC pour trouver la probabilité qu'un sujet qui devine obtienne 27 % ou plus de réponses correctes sur 100 cartes. Trace le graphique de la distribution et ombre l'aire correspondant à cette probabilité.
   d) Selon la réponse que tu as obtenue en (c), dirais-tu qu'il serait très surprenant, plutôt surprenant ou pas tellement surprenant qu'un sujet obtienne 27 % ou plus de réponses correctes même s'il se contente de deviner à l'aveugle ce que représente chaque carte?
   e) Répète les parties (c) et (d) en supposant un pourcentage de 31 % de réponses correctes au lieu de 27 %.
   f) Répète les mêmes étapes en utilisant un pourcentage de 35 %.
   g) Combien de cartes un sujet devrait-il reconnaître correctement pour que la probabilité d'avoir réussi aussi bien ou mieux en se contentant de deviner soit de 0,025 seulement?

**Sommaire**

Tu as utilisé le Théorème de la limite centrale pour déterminer la moyenne et l'écart type pour plusieurs distributions normales.

Tu as utilisé la formule de la normalisation pour décrire l'information obtenue par rapport à la distribution standard normale, la distribution $z$. Tu as utilisé l'information normalisée de la table de distribution centrale normale pour calculer des probabilités.

Tu as calculé les mêmes probabilités sans normaliser l'information et sans la table, au moyen de la calculatrice TI-83.

***Il est recommandé de discuter du contenu de cette leçon en classe pour s'assurer que tous les élèves comprennent les concepts importants.***

Annexe I-2

## Leçon 1 : Corrigé - Variables nominales

### Déroulement

1. a) mesure
   b) nominale (binaire)
   c) nominale
   d) nominale (binaire)
   e) mesure
   f) nominale (binaire)
   g) mesure
   h) mesure
2. Il s'agit d'une variable de mesure; seules les valeurs entières sont acceptées.
3. Il s'agit d'une variable nominale.

4.-11. Ces questions utilisent les données de la classe enregistrées par les élèves. Ce type de mathématiques se passe d'explication et les étapes à suivre avec la calculatrice sont montrées. Ces questions visent à présenter la calculatrice et les distributions de variables types. Les élèves connaissent maintenant le graphique à barres simples.

## Leçon 2 : Corrigé - Variables de mesure

### Activités de mise en marche

1. Le graphique par points varie selon la classe. Il est relativement facile à dessiner et illustre la distribution des données.
   a) Variables
   b) L'***étendue*** est la distance entre les valeurs supérieures et inférieures des données. Elle nous indique dans quelle mesure les données sont étalées.
   c) La ***médiane*** est soit le nombre milieu dans un ensemble de données ordonné quand le nombre est impair, soit une valeur arbitraire entre les deux points de données entourant le centre d'un ensemble de données ordonné. En règle générale, la moyenne de ces deux points est choisie comme médiane.

      ***Exemple 1 :*** 1, 2, 3, 4, 5 médiane : 3

      ***Exemple 2 :*** 1, 2, 3, 4 médiane : 2,5 (ou toute valeur entre 2 et 3.)
   d) Le mode correspond aux valeurs qui reviennent le plus souvent.

      ***Exemple 3 :*** 1, 2, 2, 3, 4 mode : 2

      ***Exemple 4 :*** 1, 2, 2, 3, 4, 4 mode : 2, 4
   e) Pour trouver la ***moyenne***, il faut additionner des données et les diviser par leur nombre.
   f) La phrase devrait faire référence à l'étendue des données et au centre. La discussion devrait aider à préciser les trois façons de décrire le centre.
2. a) Une fois de plus, ce graphique par points varie d'une classe à l'autre.
   b) Si les résultats relatifs à l'expérience de voyage des élèves du secondaire sont biaisés, c'est peut-être parce qu'ils ont eu plus d'occasions de voyager que les plus jeunes. Discutez d'autres sources de biais.
3. Un élève sera disponible seulement après les heures de travail. Peut-être certains clients magasinent-ils seulement le matin. Ils sont différents sur cet aspect au moins de ceux qui magasinent l'après-midi, et leurs goûts aussi peuvent être différents. L'élève ne le saurait probablement pas. En plus, une élève peut hésiter à demander à certains clients d'essayer des tacos. Leur apparence et leur allure peuvent jouer dans cette réticence, ce qui rend l'étude non aléatoire. D'autres sources de biais sont possibles. Discutez-en.
4. a) La population à l'étude était probablement l'ensemble des électeurs de la région.
   b) La variable était l'ensemble binaire des réponses à la question.
   c) L'échantillon était l'ensemble des personnes qui ont envoyé un message électronique.
   d) Premièrement, des non-électeurs ont pu téléphoner. Ils ne font pas partie de la population visée. Deuxièmement, les électeurs qui n'ont pas une opinion marquée sur la question n'ont probablement pas appelé, et ils représentent peut-être le gros de l'électorat. Troisièmement, les électeurs qui se sentent émotivement concernés par la question auront tendance à envoyer des messages à répétition et à encourager leur entourage à faire de même. Si on constate une tendance connue parmi ces électeurs à privilégier l'une des deux réponses, le résultat du sondage par courrier électronique devrait être connu avant qu'il ne soit effectué. La décision de mener ce sondage a peut-être été prise en sachant que le point de vue du député serait appuyé.

## Leçon 3 : Corrigé - Échantillon aléatoire simple

### Activités de mise en marche

Les échantillons aléatoires simples en cause sont : {2, 26, 10, 15}, {11, 3, 16, 15}, {12, 23, 10, 18}. Les longueurs moyennes (au millimètre près) sont 21,75 mm, 22,75 mm et 18,5 mm. La distribution dans la classe aura une étendue relativement grande. La longueur moyenne résultant des moyennes calculées par chacun sera d'environ 20 mm.

### Déroulement

Nous pouvons prédire que, quand la taille de l'échantillon est 10, la longueur moyenne obtenue par la classe sera près de celle obtenue auparavant, mais que l'étendue de la distribution des longueurs moyennes sera moins large qu'elle ne l'était pour un échantillon de quatre éléments. Cette prédiction est-elle correcte? Pourquoi? Si nous mesurons toutes les lignes, nous obtiendrons une longueur moyenne d'un peu plus de 20 mm.

# Leçon 4 : Corrigé - Description de distributions

## Activités de mise en marche

1. a) Ces distributions sont toutes relativement compactes et symétriques, mais les valeurs moyenne diffèrent. Que cela signifie-t-il?

   b) Ces distributions ont toutes des moyennes similaires et elles sont relativement symétriques, mais leur étendue est différente. Que cela signifie-t-il?

   c) Elles ont toutes des moyennes similaires mais, si la distribution dans G est relativement symétrique, celle de H est asymétrique vers la droite et celle de I asymétrique vers la gauche. Que cela signifie-t-il?

   d) La classe J semble similaire à la classe F; mais il semble y avoir une césure dans les résultats. On dira que la distribution est ***bimodale***. Peut-être la classe comporte-t-elle deux sous-groupes. Peux-tu expliquer pourquoi?

   e) La courbe de la classe C ressemble à celle de la classe G, sauf pour les points 42 et 94. Il s'agit de valeurs ***aberrantes***. Peux-tu essayer d'expliquer ce que cela signifie? Crois-tu que, dans un certain sens, ces valeurs sont liées à des populations différentes que la majorité des points?

   f) On constate de la ***granularité*** dans la classe L. À ton avis, que nous indique ce résultat à propos des notes ou des méthodes de notation?

## Déroulement

1. a) Victoria a régné le plus longtemps.

   b) Édouard V a régné le moins longtemps.

   c) Victoria a régné pendant 63 ans et Édouard V pendant un peu moins d'un an. Le graphique à bâtons a cette apparence quand on passe d'une colonne à l'autre.

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

Quand les départs sont ordonnés, le graphique à bâtons ordonné a cette apparence.

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

Il semble que la distribution est asymétrique vers la droite (valeurs supérieures) et que la plupart des valeurs se situent entre 6 et 39 années (on a arbitrairement supprimé les 5 valeurs apparaissant à chaque extrémité).

## Leçon 4 : Corrigé - Description de distributions (suite)

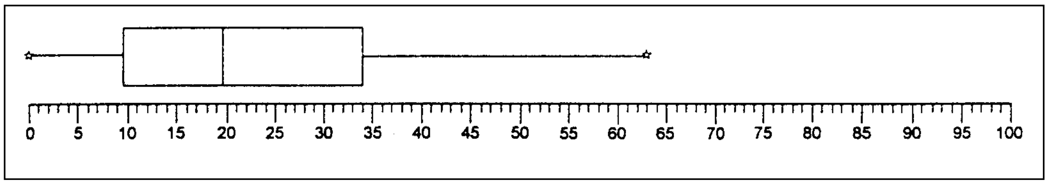
d) Si on compte 20 entrées à une extrémité ou l'autre, on voit que la valeur médiane se trouve entre 19 et 20 années. Disons 19,5 années.

e) Si on fait le compte dans les 10 entrées supérieures, nous voyons que le Q1 se trouve entre 9 et 10 années. Disons 9,5 années.

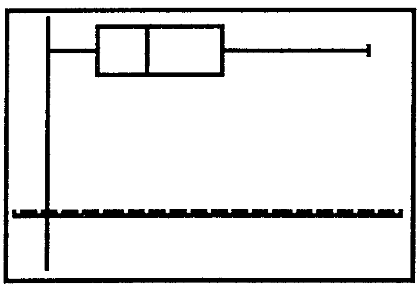
f) De même, si on compte les entrées inférieures (droite à gauche... grand à petit), nous constatons que Q3 se trouve entre 33 et 35 années. Disons donc 34 ans.

g) Ainsi, Min = 0, Q1 = 9,5, Md = 19,5, Q3 = 34, et Max = 63.

2. Le graphique en boîte devrait avoir cette apparence.

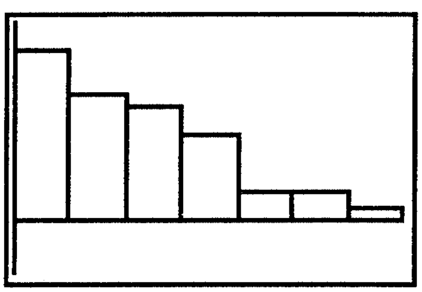



3. Le graphique en boîte affiché sur ta calculatrice aura cette apparence. Utilise la touche Trace pour afficher les valeurs en cinq points.



4. Les cinq points utilisés pour tracer le graphique en boîte, que nous avons trouvé auparavant, devraient être affichés sur ta calculatrice.

5. Le premier histogramme affiché a cette apparence. La Xscl ou la largeur de la barre est de 10,5 unités. Fais des essais en changeant cette valeur.

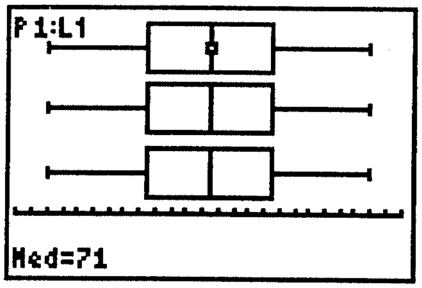


# Leçon 5 : Corrigé - La règle empirique

## Activités de mise en marche

a) Oui, ces distributions sont différentes. Les points de données ne sont pas les mêmes. Pour la classe 1, on obtient trois groupes de points avec deux groupes intermédiaires; pour la classe 2, la distribution des points est relativement égale dans toute l'étendue; pour la classe 3, on voit deux groupes avec trois points aux extrémités et au centre.

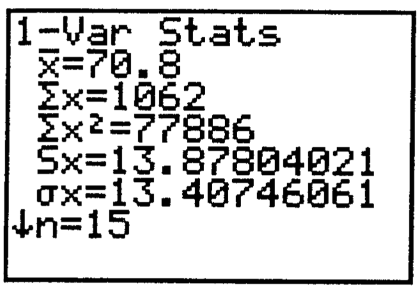
b) Les trois graphiques en boî te sont illustrés ci-dessous, avec la médiane dans le premier. Les listes utilisées étaient L1, L2 et L3.




c) Étant donné que les trois graphiques en boî te sont identiques, rien ne nous poussent à supposer que les distributions sont différentes.

## Déroulement

1. Pour obtenir les données de la classe 1, appuie sur Stat, sélectionne Calc, puis 1-Var Stats, avant d'appuyer sur 2nd, sur L1 et sur ENTER. L'écran aura cette apparence.




Classe 2 : $\bar{x}$ = 70,93 et $s_x$ = 11,8410125.

Classe 3 : $\bar{x}$ = 70,8 et $s_x$ = 10,68510311.

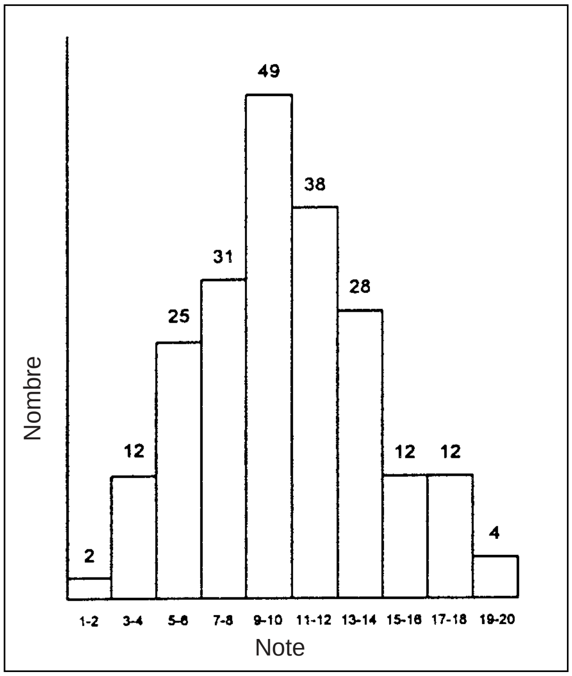
Nous n'avons pas besoin des autres valeurs affichées pour l'instant.

Les moyennes dans les deux classes sont identiques et la moyenne de l'autre classe est proche. Cependant, l'écart type diffère dans chaque cas. Cela nous indique que les distributions ne sont pas les mêmes.

2. Quand on inscrit les 7 points de données dans la colonne L4 et qu'on utilise le menu StatCalc, nous obtenons $\bar{x}$ = 1 600, médiane = 1 614, $s_x$ = 189,2397 et écart interquartile de 1 792 – 439 = 353. Si on remplace 1 362 par 362, la médiane et l'écart interquartile restent les mêmes, mais la moyenne devient 1457,1429 alors que l'écart type devient 507,9391. Les valeurs aberrantes influent sur les changements dans les deux dernières. La médiane et l'écart interquartile restent les mêmes.

## Leçon 5 : Corrigé - La règle empirique (suite)

3. Voici l'histogramme demandé.




4. Il semble avoir une forme de cloche et être symétrique.

5. Étant donné que 10,221 – 3,859 = 6,362 et que 0,221 + 3,859 = 14,08, les notes entre 6,362 et 14,08 se trouvent à l'intérieur d'un écart type de la moyenne. Ainsi, 146 notes, ou 146 ÷ 213 = 0,685 des notes se trouvent dans cette étendue.

6. De même, toutes les notes entre 2,5 et 17,9 sont à l'intérieur de 2 écarts types de la moyenne. Cela comprend toutes les notes sauf six. Par conséquent, 207 ÷ 213 = 0,972 des notes se trouvent à l'intérieur de 2 écarts types de la moyenne.

7. Il est très clair que toutes les notes se trouvent à l'intérieur de trois écarts types de la moyenne.

### Exercices

1. a) Robert a obtenu 184 points au-dessus de la moyenne au SAT.
   b) Kathy a obtenu 7,4 points au-dessus de la moyenne à l'ACT.
   c) Les deux échelles sont différentes. Il n'est pas facile de les comparer, encore moins de déterminer quelle note est la meilleure.
   d) La note de Robert se trouve à environ 1,06 écart type au-dessus de la moyenne.
   e) La note de Kathy se trouve à environ 1,42 écart type au-dessus de la moyenne.
   f) La cote *z* de Kathy, soit 1,42, est plus élevée que la cote *z* de Robert, soit 1,06.
   g) Étant donné que la cote *z* de Kathy est plus élevée que celle de Robert, elle a mieux réussi qu'une plus grande proportion des élèves qui ont participé au SAT que Robert par rapport aux élèves qui ont fait l'ACT. Par conséquent, la note de Kathy est meilleure.
   h) La cote *z* de Mike est $\frac{740-896}{174}$ soit environ –0,90. La cote *z* de Karine est $\frac{19-20{,}6}{5{,}2}$ ou environ –0,31.
   i) Étant donné que ces valeurs sont négatives, la cote *z* de Karine est plus élevée.
   j) Les cotes *z* négatives indiquent que la note brute est inférieure à la moyenne.

# Leçon 6 : Corrigé - La signification statistique

## Activités de mise en marche

a) Un échantillon comprend douze bonbons orange, cinq jaunes et huit bruns. Les réponses varieront.

b) La proportion de bonbons orange est une statistique.

c) Non.

d) La proportion de bonbons orange dans un échantillon est 0,48. Les réponses varient.

e) Non, tous n'auront pas la même proportion.

f) Le graphique par points sera relativement symétrique, avec un centre entre 0,4 et 0,5. S'il n'y a pas beaucoup de points (moins de 30), la symétrie ne sera peut-être pas aussi évidente.

g) Non, certaines personnes peuvent suggérer que la proportion dans la population est près de 0,4, et d'autres que cette valeur est plus près de 0,5. Quelques-uns peuvent suggérer des valeurs autour de 0,3 ou 0,6.

h) Une estimation de la proportion de bonbons orange dans la population se trouve entre 0,4 et 0,5. Les réponses varient.

i) La plupart des estimations seraient relativement proches. Certaines pourront se trouver assez éloignées des autres. Il est probable que Hershey mélange les couleurs assez bien et que seulement une petite proportion des emballages contiennent un mélange de couleurs qui varie de façon significative par rapport à l'ensemble des emballages.

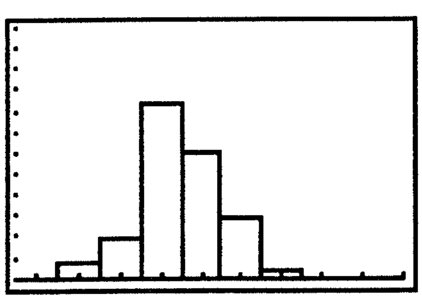
j) Un plus petit échantillon aurait donné un étendue plus large des proportions dans l'échantillon. Cela est dû au fait que chacun des nombres de bonbons dans une plus grande partie de l'échantillon et que le décompte des bonbons de couleur différente changerait la proportion de façon plus importante que ce ne serait le cas dans un échantillon plus grand.

k) Un échantillon plus grand aurait résulté en un regroupement plus étroit des proportions dans l'échantillon. Cela est dû au fait que chaque bonbon représente une proportion plus petite dans l'échantillon, ce qui a une influence moins grande sur la distribution que ce ne serait le cas pour un échantillon plus petit.

### Déroulement

a) Cela peut prendre dix ou quinze minutes.

b) Un résultat produit un graphique qui a une apparence relativement symétrique et en forme de cloche, dans lequel la barre la plus haute comprend les valeurs de $x$ telles que $0,35 \le x \le 0,45$. La première barre comprend les valeurs entre 0,15 et 1,25. La dernière comprend les valeurs entre 0,65 et 0,75. Voici le graphique :



Les réponses varient.

## Leçon 6 : Corrigé - La signification statistique (suite)

c) Un calcul nous permet d'obtenir une moyenne de l'échantillon de 0,449. L'écart type pour 500 échantillons était 0,0965 ou environ 0,1. Cinquante échantillons devraient-ils donner le même résultat?

d) Oui, les proportions dans l'échantillon sont regroupées autour de 0,45.

e) Voici un ensemble de résultats pour les trois questions. Tes résultats (si on considère 50 échantillons) peuvent différer un peu.

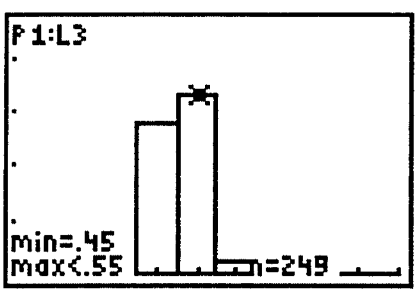
| | **Nombre de proportions dans l'échantillon** | **Pourcentage de proportions dans l'échantillon** |
|---|---|---|
| À l'intérieur de $0{,}45 \pm 0{,}10$ | 356 | 71,2% |
| À l'intérieur de $0{,}45 \pm 0{,}20$ | 475 | 95% |
| À l'intérieur de $0{,}45 \pm 0{,}30$ | 500 | 100% |

f) Il semble que 95 % des élèves auraient obtenu un paramètre de la population réelle à 0,20 de leurs statistiques (c'est-à-dire à l'intérieur de 2 écarts types de la moyenne.) Cela concorde avec la règle empirique.

g) Il est impossible de dire avec certitude si la valeur d'un échantillon se rapproche du paramètre de la population. On peut être raisonnablement certains que les résultats apparaissant dans le tableau ci-dessus sont sûrs, si on suppose que c'était le cas dans 95 % des échantillons.

h) Dans une expérience, on a obtenu une moyenne de 0,450027 et un écart type de 0,058.

i) Il y avait seulement quatre barres au lieu de six. Les notes étaient regroupées de la même façon, mais l'étendue semblait plus petite. Les barres du centre étaient plus grandes et les barres extérieures plus courtes. Voici le graphique produit par la calculatrice, où la barre la plus élevée est tracée. Une barre courte est cachée par les valeurs min-max.



## Leçon 6 : Corrigé - La signification statistique (suite)

j) Voici un ensemble de résultats pour 500 échantillons. Les tiens ne seront peut-être pas exactement pareils.

| | **Nombre de proportions dans l'échantillon** | **Pourcentage de proportions dans l'échantillon** |
|---|---|---|
| À l'intérieur de $0,45 \pm 0,10$ | 460 | 82% |
| À l'intérieur de $0,45 \pm 0,20$ | 500 | 100% |
| À l'intérieur de $0,45 \pm 0,30$ | 500 | 100% |

Ce ne sont plus 72 % mais 92 % des proportions dans l'échantillon qui se trouvent à 0,10 de 0,45.

k) Un échantillon plus grand est plus susceptible de résulter en une proportion dans l'échantillon proche de la proportion dans la population.

l) En doublant l'écart type pour obtenir 0,116, nous obtenons une étendue de 0,450 - 0,116 à 0,450 + 0,116 dans l'expérience, soit entre 0,334 et 0,566.

m) Si on fait le tri dans la liste des proportions, seulement 28 des 500 échantillons ont une proportion de bonbons orange qui ne fait pas partie de l'étendue donnée. Ainsi, dans 472 échantillons, ou 94,4 % des échantillons, les proportions se trouvaient dans cette étendue.

n) La réponse à cette question est la même que la réponse à la question (g). Environ 95 % des échantillons auront une proportion de la population à l'intérieur de 2 écarts types par rapport à la proportion dans l'échantillon.

### Exercice

1. Il est probable que cinq ou un nombre proche de cinq objets défectueux soient produits.
2. Certains peuvent obtenir quatre, d'autres six, et à l'occasion trois ou sept gadgets défectueux dans un lot. Il y en aura rarement moins que trois ou plus de sept.
3. Non. Il ne serait pas étonnant d'en obtenir quatre ou même trois.
4. Oui, cela serait inhabituel.
5. Oui, c'est vrai. Cela risque fort peu de se produire.

### Déroulement

a) La donnée un tiers dans l'énoncé du problème est un paramètre parce qu'elle décrit la population d'objets produits.

b) La proportion dans l'échantillon est $\frac{4}{15}$ et son symbole est $\hat{p}$ (p chapeau).

c) Oui, il ne serait pas étonnant de trouver quatre gadgets défectueux si le tiers de la population étaient encore défectueux. Nous savons que la distribution d'échantillonnage serait centrée à cinq objets défectueux, et que l'ensemble de points formerait une cloche, y compris quatre et six objets défectueux dans un graphique par points.

## Leçon 6 : Corrigé - La signification statistique (suite)

d) Non, la distribution d'échantillonnage comprendrait naturellement des nombres variés de gadgets défectueux, tel qu'il est décrit en (c).

e) Un essai a donné les résultats figurant dans le tableau ci-dessous.

| Gadget | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N° sur le dé | 6 | 4 | 5 | 5 | 1 | 5 | 3 | 1 | 5 | 4 | 3 | 6 | 1 | 2 | 2 |
| D ou C | C | C | C | C | D | C | C | D | C | C | C | C | D | D | D |

**Le premier lot comprenait 5 objets défectueux, soit le tiers (0,33).**

f) Voici le résultat de quatre autres essais :

| Gadget | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N° sur le dé | 1 | 1 | 2 | 6 | 3 | 2 | 6 | 3 | 2 | 4 | 6 | 3 | 5 | 4 | 4 |
| D ou C | D | D | D | C | C | D | C | C | D | C | D | C | C | C | C |

**Le deuxième lot comprenait de nouveau 5 objets défectueux, soit 1 sur 3 (0,33).**

| Gadget | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N° sur le dé | 3 | 1 | 1 | 5 | 4 | 1 | 3 | 6 | 6 | 6 | 6 | 3 | 3 | 4 | 3 |
| D ou C | C | D | D | C | C | D | C | C | C | C | C | C | C | C | C |

**Le troisième lot comprenait 3 objets défectueux, soit 1 sur 5 (0,20).**

| Gadget | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N° sur le dé | 2 | 6 | 6 | 1 | 1 | 2 | 6 | 6 | 3 | 5 | 5 | 3 | 5 | 6 | 2 |
| D ou C | D | C | C | D | D | D | C | C | C | C | C | C | C | C | D |

**Le quatrième lot comprenait 5 objets défectueux, soit 1 sur 3 (0,33).**

| Gadget | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N° sur le dé | 5 | 6 | 6 | 6 | 6 | 6 | 1 | 2 | 1 | 1 | 6 | 2 | 4 | 2 | 1 |
| D ou C | C | C | C | C | C | C | D | D | D | D | C | D | C | D | D |

**Le cinquième lot comprenait 6 objets défectueux, soit 2 sur 5 (0,40).**

g) Non.

h) Le graphique par points sera centré à 5 et relativement symétrique. Cependant, étant donné que l'échelle va plus loin sur la droite de 5 unités que sur la gauche, il est possible qu'il sera un peu asymétrique vers la droite. Notre connaissance des distributions d'échantillonnage nous indique que, si cette asymétrie existe, elle devrait être négligeable.

i) Seulement un des lots visés ci-dessus comprenait quatre gadgets défectueux ou moins. La proportion est 0,2.

j) Il est très probable que le processus produirait un lot comprenant quatre gadgets défectueux ou moins.

## Leçon 6 : Corrigé - La signification statistique (suite)

k) À l'aide de ce programme, une expérience a donné les résultats figurant dans le tableau ci-dessous.

| N^bre de gadgets défectueux | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| N^bre de lots | 1 | 16 | 62 | 127 | 183 | 218 | 182 | 115 |

| N^bre de gadgets défectueux | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|
| N^bre de lots | 69 | 21 | 3 | 3 | 0 | 0 | 0 | 0 |

l) 369, ou 0,369 des échantillons, comprenaient 4 objets défectueux ou moins dans l'expérience ci-dessus.

m) Il est relativement possible qu'un lot contient 4 objets défectueux ou moins.

n) Ce résultat ne nous donne pas beaucoup de preuves d'amélioration parce qu'il aurait très bien pu se produire sans qu'aucune amélioration ne soit apportée.

o) 79 des échantillons (ou 0,079), comportaient 2 objets défectueux ou moins. C'est moins de 10 % mais, si 1000 lots sont produits chaque jour, on pourrait s'attendre à obtenir ce même nombre de gadgets défectueux dans un lot de 79.

p) Peut-être cela n'est-il pas si inhabituel. C'est une preuve bien mince qui ne nous permet pas d'affirmer que la modification a amélioré le processus.

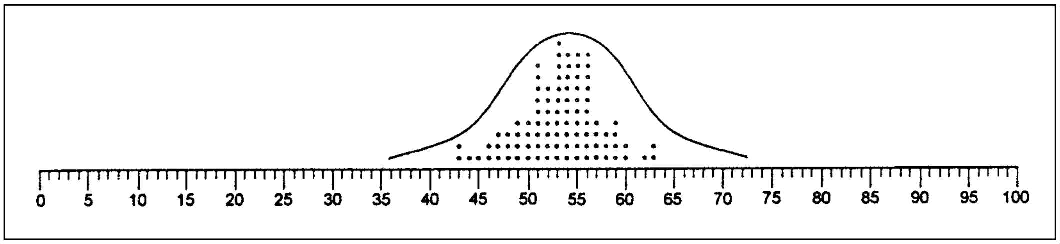
q) Seulement 1 lot sur 1000 ne comprenait aucun objet défectueux. Cela signifie qu'il est très improbable que les ingénieurs trouvent un tel lot à moins que la modification n'ait de fait amélioré le processus. Ce n'est pas une certitude cependant. Le programme dont il est question en (k) a produit un lot de ce type en supposant que le tiers de la production était toujours défectueux.

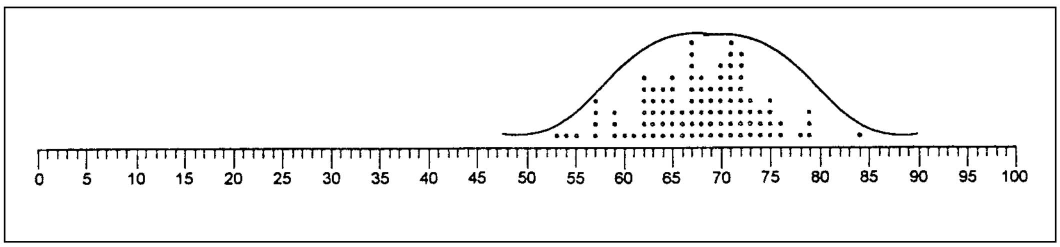
## Leçon 7 : Corrigé - Aire sous la courbe normale

### Activités de mise en marche

a) Les deux formes sont relativement symétriques autour d'un point sommet central, et décroissent à partir de ce point sommet dans chaque direction.

b)




c) La courbe la plus élevée est A, la plus à gauche est C et l'autre est la courbe B.

### Déroulement

a) Voir les exemples donnés.

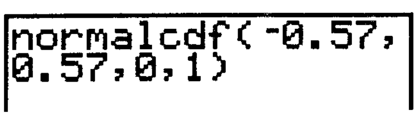
b) Ce sera la même chose. L'aire ne change pas.

c) Ce sera l'aire qui n'est pas comprise en (a) et en (b), soit 1 – 0,715661 = 0,284339.

d) Étant donné que la distribution est symétrique, la réponse sera la même qu'en (c).

e) Une façon d'y arriver est de noter que l'aire à gauche de –0,57 est la même que l'aire à droite de 0,57. Donc, la réponse est aussi 0,284339.

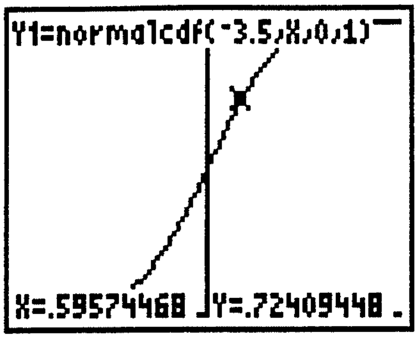
Tu peux faire cet exercice sans ta calculatrice, en considérant qu'il s'agit de l'aire entière sous la courbe, à l'exception de l'aire se trouvant entre le point 0,284339 à la gauche de –0,57 et de 0,284339 à la droite de 0,57. Ce devrait donc être 0,431322. Tu peux obtenir ce résultat avec ta calculatrice, comme il est illustré ci-dessous (si tu appuies sur ENTER).




Tu obtiendras 0,4313223844.

f) Si tu utilises –7 comme limite à gauche, la calculatrice te dira que la probabilité est d'environ $3{,}5952 \times 10^{-5}$. C'est très faible. Peut-être est-il impossible d'être précis.

## Leçon 7 : Corrigé - Aire sous la courbe normale (suite)

g) 


C'est l'écran qu'on obtient quand on applique la méthode expliquée à la leçon 7. J'ai fait un zoom d'un facteur 4 à quatre ou cinq reprises chaque fois pour tracer le point quand la valeur de *y* était 0,7258 et la valeur de *x* est 0,6010 (pour quatre chiffres significatifs). Cela signifie que $\Pr(z < k) = 0{,}7258$ quand $k \approx 0{,}6010$.

h) L'une des méthodes est la soustraction : $1 - 0{,}0625 = 0{,}9375$. Cela signifie qu'il faut trouver la valeur de *x* qui correspond à une valeur de *y* de 0,9375 sur la courbe de la fdc. Avec la même technique que celle appliquée en (g), nous trouvons que la valeur de *x* est d'environ 1,5359. Cela signifie que $k \approx 1{,}5359$.

### Normalisation des distributions normales

a) Deux méthodes peuvent être utilisées pour résoudre ce problème. Premièrement, nous appliquons la formule permettant de trouver les cotes *z* pour les QI de 110 et de 125. Si QI = 110, alors la cote *z* est $z_{110} = \dfrac{110-100}{15} = 0{,}6667.$

De même $z_{125} = \dfrac{125-100}{15} = 1{,}6667.$

Nous calculerons ensuite la valeur de la fdc (fonction de densité cumulative) pour trouver l'aire sous la courbe standard normale entre $z = 0{,}6667$ et $z = 1{,}6667$. Les résultats sont affichés dans l'écran ci-dessous.

```
normalcdf(0.6667
,1.6667,0,1)
      .2046948043
```

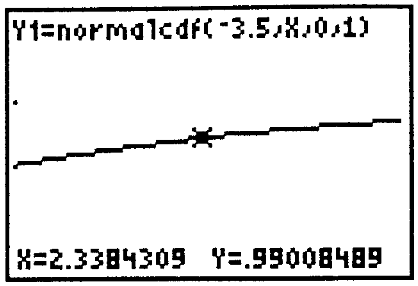
La deuxième méthode consiste à utiliser les données présentées. Le résultat est affiché dans l'écran suivant.

```
normalcdf(110,12
5,100,15)
      .2047021366
```

b) Assure-toi que tu obtiennes la même réponse sans égard à la méthode utilisée. La réponse est environ 4,8 %.

## Leçon 7 : Corrigé - Aire sous la courbe normale (suite)

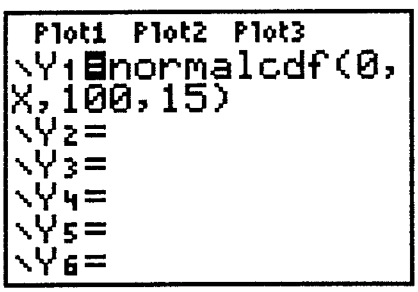
c) Tout d'abord, fais un zoom sur la courbe de la fdc pour la distribution normale standard. Tu vois que 99 % de la population se trouvent au-dessous d'un point se trouvant à environ 2,34 écarts types au-dessus de la moyenne. L'écran ci-dessous affiche le résultat.

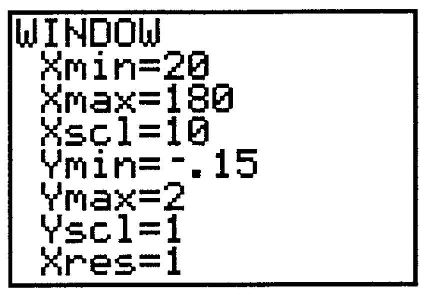



Utilise ensuite la formule de la transformation pour obtenir $2{,}338 = \dfrac{x-100}{15} \Rightarrow x = 135{,}07.$

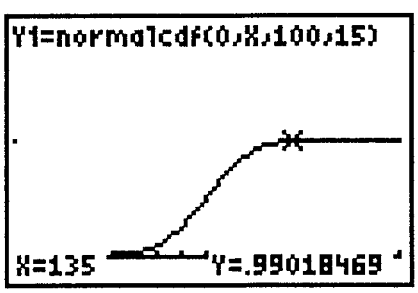
Ainsi, on peut dire que toute personne qui a un QI supérieur à 135 se trouve dans le 1 % supérieur de la population.

- Voici une autre méthode : Utilise la courbe de la fdc, avec une moyenne de 100 et un écart type de 15. Entre l'équation, comme ci-dessous.



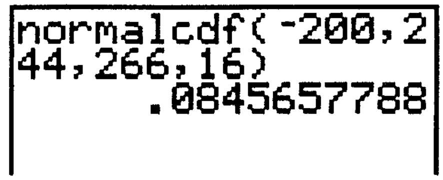

- Règle la fenêtre comme illustré.
- Appuie sur la touche GRAPH pour afficher la courbe, puis sur TRACE pour trouver qu'une personne dont le QI est 135 se trouve dans le 1 % supérieur de la population.



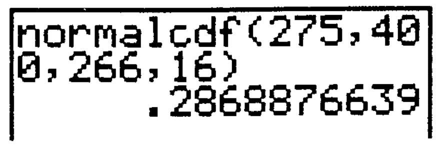
## Leçon 7 : Corrigé - Aire sous la courbe normale (suite)

### Exercice

1. a) Il semble que la moyenne est environ 50 et que l'écart type est d'environ 10.
   b) La moyenne est de 110 et l'écart type d'environ 25.
   c) Ici, la moyenne est d'environ –10 et l'écart type d'environ 10.
2. a) Montre qu'environ 8 % seulement des naissances surviennent avant 244 jours de grossesse. Remarque la limite inférieure utilisée. Est-elle raisonnable? En as-tu essayé d'autre? Pourquoi pas 0?

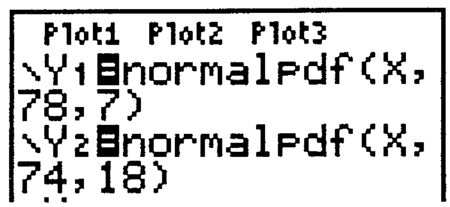


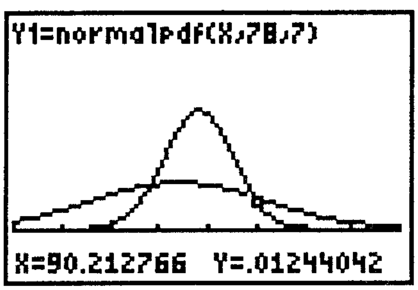

   b) Apparemment, environ 29 % des naissances surviennent après 275 jours de grossesse.
   c) C'est environ 17 %.
   d) Environ 46 % des naissances ont lieu à l'intérieur de ces limites.




3. a) La première équation s'applique à la classe de Mme Miller. La deuxième à celle de M. Sapinsky.

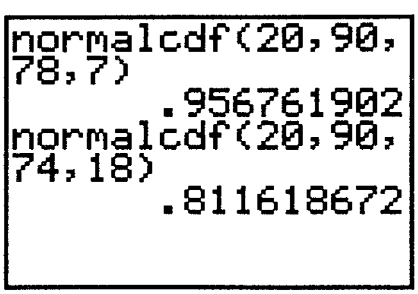
   La courbe la plus élevée correspond à la classe de Mme Miller.




Remarque que les courbes se croisent à environ 90 %. Semble-t-il y avoir une aire plus importante sous le graphique de la classe de M. Sapinsky à la droite de 90 %? Probablement qu'une plus grande proportion de ses élèves ont obtenu un A. Est-ce qu'une plus grande proportion de ses élèves ont obtenu un F?

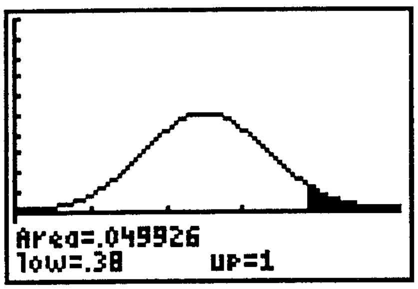
   b) Cet écran affiche les calculs. Il semble que moins de 5 % des élèves de Mme Miller ont obtenu une note au-dessus de 90, alors que 19 % des élèves de M. Sapinsky ont eu une telle note.
   c) Les calculs pour la note F seront similaires.



## Leçon 8 : Corrigé - Le Théorème de la limite central

### Déroulement

a) Selon le TLC, la moyenne de la proportion de réponses correctes dans l'échantillon (si les sujets devinent) est 0,25 et l'écart type est $\sqrt{\dfrac{(0{,}25)(0{,}75)}{30}}$,ce qui donne environ 0,079. Le graphique a une forme de cloche et il est symétrique.

b) 


Ton dessin devrait avoir cette apparence. Il semble que l'aire soit d'environ 5 % seulement par rapport à l'aire totale sous la courbe. Nous obtenons cet affichage avec les réglages suivants : Xmin = 0, Xmax = 0.5, Ymin = –3 et Ymax = 10, puis en utilisant l'écran du menu DISTR DRAW.

```
ShadeNorm(.38,1,
.25,.079)
```

c) La formule de normalisation associe à la valeur 0,38 de cette distribution une cote *z* d'environ 1,65, soit $z = \dfrac{0{,}38 - 0{,}25}{0{,}079} \approx$ 1,65. La table affecte une valeur de 0,4505 à ce score. Cela signifie que 45 % des sujets obtiennent entre 25 % et 38 % de réponses correctes. Ainsi, environ 5 % seulement obtiendront 38 % de réponses correctes ou plus (50 % obtiennent moins que 25 réponses correctes).

d) Normalcdf(0, 0.38, 0.25, 0.079) = 0,949; donc environ 95 % des sujets obtiendraient 38 % ou moins de réponses correctes. Environ 5 %, comme auparavant, obtiendraient 38 % de réponses correctes ou plus.

e) 38 % de 30 équivaut à 11,4 : nous voulons donc connaître quelle proportion dans les 10 000 résultats ont deviné la bonne réponse 12 fois ou plus. La réponse est 605, soit 6,05. Cela semble raisonnablement proche.

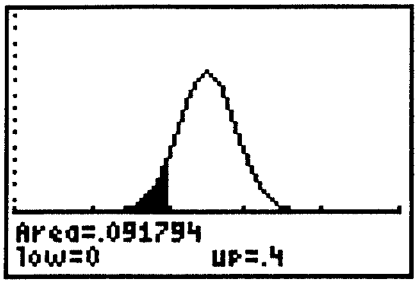
f) À long terme, un sujet obtiendrait 38 % de réponses correctes ou plus, environ 5 % seulement du temps s'il se contentait de deviner.

### Exercice

1. a) Selon le TLC, dans les 2 cas la moyenne des proportions dans l'échantillon est 0,45 mais, bien que l'écart type pour les échantillons de 175 éléments soit $\sqrt{\dfrac{(0{,}45)(0{,}55)}{175}} = 0{,}0376$, il est de $\sqrt{\dfrac{(0{,}45)(0{,}55)}{75}} = 0{,}0574$ pour les échantillons de 75 éléments. Dans chacun des cas, la distribution est symétrique et en forme de cloche (autrement dit, c'est une courbe normale).

## Leçon 8 : Corrigé - Le Théorème de la limite centrale (suite)

b) La courbe la plus haute et la plus étroite représente la distribution des échantillons comportant 175 éléments. L'écran ci-dessous illustre l'aire correspondant à la probabilité que la proportion de bonbons oranges dans un échantillon soit inférieure à 0,4; il semble que la probabilité que la proportion dans un échantillon se trouve dans cette région soit d'environ 0,09.



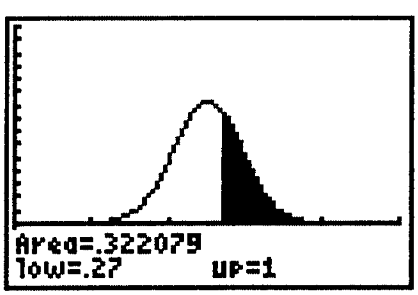

c) La cote $z$ de 0,4 pour cette distribution est $\dfrac{0,4-0,45}{0,0376} \doteq -1,33$. La table de la distribution normale réduite indique que 0,4082 de l'aire sous la courbe standard normale se trouve entre cette cote $z$ et la moyenne. Autrement dit, 0,5 – 0,4082 = 0,0918 de l'aire se trouve à la gauche de la cote $z$ –1,33.

Pour ce qui est de la distribution originale des proportions de bonbons orange dans l'échantillon, ce résultat nous indique que la probabilité qu'un échantillon comportant 175 éléments ait une proportion de bonbons orange inférieure à 0,4 est d'environ 0,09. ***Environ 9 % des échantillons comporteront moins de 40 % de bonbons oranges.***

Si nous utilisons un échantillon de 75 éléments au lieu, nous constatons que l'objet d'étude a une cote $z$ de –0,8710. La table indique 0,3078, ce qui nous dit que 0,5 – 0,3078 = 0,1922 de l'aire se trouve à la gauche de cette cote $z$. ***Environ 19 % des échantillons comporteront moins de 40 % de bonbons oranges.***

d) Avec la commande normalcdf(0, 0.4, 0.45, 0.0376), nous obtenons environ 0,0917942541 ou 0,0918, comme c'était le cas ci-dessus pour l'échantillon de 75 éléments; si nous utilisons la commande normalcdf(0, 0.4, 0.45, 0.0574), nous obtenons 0,1918551544 ou environ 0,1919. C'est très proche de la valeur ci-dessus, et probablement plus proche encore de la valeur théorique correcte que ne l'était la réponse obtenue à l'aide de la table. (Nous aurions pu trouver une valeur plus exacte en interpolant à partir de la table.)

e) L'écran ci-dessous affiche les deux réponses. La première correspond aux échantillons de 175 éléments et la deuxième aux échantillons de 75 éléments.

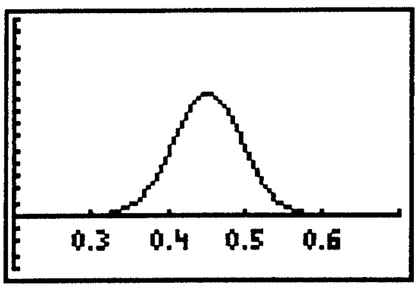



Il semble que plus de 99 % des échantillons comportant 175 éléments comprennent entre 35 % et 55 % de bonbons oranges, alors qu'environ 92 % seulement des échantillons de 75 éléments comportent ce pourcentage de bonbons oranges.

## Leçon 8 : Corrigé - Le Théorème de la limite centrale (suite)

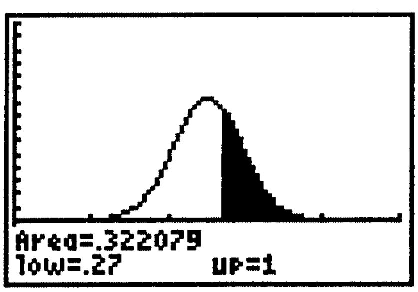
2. a) Si le sujet se contente de deviner, il pourrait obtenir 25 réponses correctes, mais il pourrait aussi en obtenir 24, 26, 23 ou 27 ou, à mesure que les probabilités diminuent, des nombres plus éloignés de 25. La proportion de réponses correctes devrait varier normalement d'un échantillon à l'autre, et à long terme le nombre moyen de réponses correctes sera 25, avec une distribution symétrique en forme de cloche autour de ce nombre.

   b) Selon le TLC, la distribution est normale et la moyenne est 0,25, avec un écart type de $\sqrt{\frac{(0{,}25)(0{,}75)}{100}} \doteq 0{,}0433$. Voici l'écran de cette distribution.




   c) La cote $z$ associé à 27 % de réponses correctes devinées est $\frac{0{,}27-0{,}25}{0{,}0433} \doteq 0{,}4169$.

   La table de valeurs sera d'environ 0,1628. Nous savons que la moyenne 0,5 + 0,1628 = 0,6626 est la proportion d'échantillons dans lesquels moins de 27 % des cartes sont devinées correctement. Ainsi, 1– 0,6626 = 0,3374 ou environ 34 %, des échantillons comprennent au moins 27 % de réponses correctes. Utilise la commande ShadeNorm(0.27, 1, 0.25, 0.0433) pour obtenir l'écran suivant de la solution :

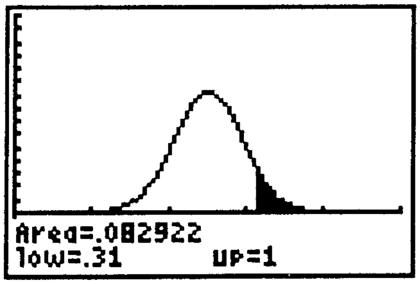



   La proportion montrée ici indique qu'environ 32 % des échantillons comportent au moins 27 % de cartes devinées correctement. C'est la seconde fois que la calculatrice donne un résultat qui ne concorde pas avec le résultat obtenu à l'aide de la table. Cela est dû au fait que les valeurs de la table augmentent à intervalle discret de 0,01 unité. Étant donné que la calculatrice donne des résultats moins éloignés, la solution est plus précise.

   d) Il ne serait pas surprenant du tout qu'un sujet devine 27 % de réponses correctes au plus. Cela surviendra une fois sur trois.

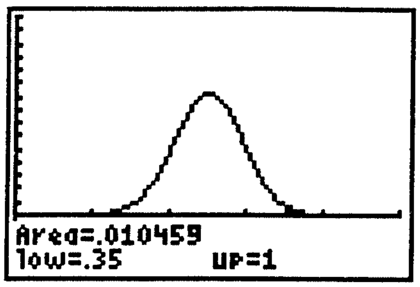
## Leçon 8 : Corrigé - Le Théorème de la limite centrale (suite)

e) Voici le résultat d'une probabilité que le sujet obtienne au moins 31 % de réponses correctes.



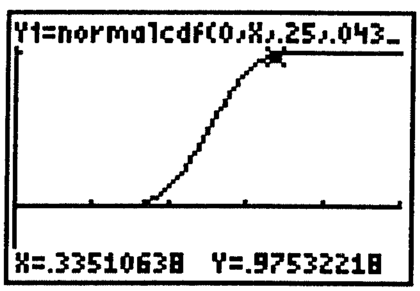

Cette fois-ci, la probabilité a chuté à environ 8 %. Il serait plus surprenant d'obtenir ce résultat que le précédent mais, encore une fois, on devrait s'attendre à ce que ce résultat se produise à quelques reprises à l'intérieur de 30 échantillons. (Les statisticiens conviennent que si moins de 30 échantillons sont étudiés, la distribution s'approche très approximativement de la normale.)

f)




Il semble que dans seulement environ 1 % des échantillons, un sujet devinerait 35 % ou plus des cartes correctement après un grand nombre d'essais. Ainsi, si un sujet y arrive vraiment, il vaut la peine de vérifier.

g) Pour que l'échantillon d'une personne se trouve au-dessous de 0,025 de la courbe, cette personne devrait avoir deviné presque toutes les cartes correctement, comme c'était le cas dans la partie (f) ci-dessus. Voici l'écran de la fonction de densité cumulative, avec le curseur positionné sur un point qui se trouve à 97,5 % de l'aire à sa gauche, ce qui laisse seulement 2,5 % à la droite. Autrement dit, la valeur de *x* de ce point est notre réponse à la question.




Des données plus significatives sur le plan de la précision peuvent être obtenues si on fait un zoom autour du point.

Annexe I-3

# Leçon 1 : Variables nominales

## Activités de mise en marche

Le module de statistique impose une lourde charge aux élèves. Ceux et celles qui acceptent de relever le défi et qui sont en mesure de composer avec la charge imposée ont besoin de toute l'aide possible. L'enseignant doit très bien connaître les activités du module, l'utilisation de la calculatrice, ainsi que la terminologie et les notions du domaine de la statistique. Il peut contribuer au bon déroulement de l'apprentissage en prévoyant d'avoir à portée de main des tableaux de collecte de données et en informant les élèves des échéances pour la présentation des données et des devoirs.

Le contenu des leçons 1 à 4 ne devrait pas être nouveau pour la plupart des élèves et sera vu plus rapidement si c'est le cas. S'il ne s'agit pas d'une simple révision, les élèves devront consacrer plus de temps à ces leçons et faire les exercices joints aux présentes notes. Dans tous les cas, des discussions en classe portant sur les notions principales aideront les élèves à comprendre la matière.

L'enseignant peut faire circuler la feuille intitulée **Enquête numéro 1** pendant que les élèves répondent aux questions. Soulignez l'importance d'écrire de manière lisible. Étant donné que les données recueillies lors de cette enquête seront utilisées dans des exercices subséquents, il faut faire circuler la feuille en question.

Les feuilles de réponses pour les questions 3 et 4 doivent également être remplies. Ces renseignements seront utilisés ultérieurement. Il est probablement utile de garder des exemplaires de ces feuilles de données pour compléter les données dans le cas de classes de petite taille.

Si les élèves travaillent de manière autonome, les feuilles de données peuvent être envoyées par la poste ou par télécopieur. Les élèves doivent avoir accès à un bon nombre de points de données pour chacune des questions. Dans le cas d'un groupe composé de moins d'une trentaine d'élèves, les données de deux classes peuvent être réunies.

### Déroulement

Les questions 1 à 6 sont simples. Lorsque les élèves tracent le graphique à barres, ils peuvent laisser un espace entre les barres. Il se peut que certains élèves sachent utiliser la calculatrice TI-83; vous pouvez alors les encourager à réaliser un graphique à barres avec leurs données, mais ils n'ont pas à le faire maintenant. La représentation des données sous forme de graphiques est décrite dans une leçon ultérieure.

## Leçon 1 : Variables nominales (suite)

**Enquête nº 1**

| Élève | Sexe | Politique | 1 cent | Valeur | Élève | Sexe | Politique | 1 cent | Valeur |
|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | 19 | | | | |
| 2 | | | | | 20 | | | | |
| 3 | | | | | 21 | | | | |
| 4 | | | | | 22 | | | | |
| 5 | | | | | 23 | | | | |
| 6 | | | | | 24 | | | | |
| 7 | | | | | 25 | | | | |
| 8 | | | | | 26 | | | | |
| 9 | | | | | 27 | | | | |
| 10 | | | | | 28 | | | | |
| 11 | | | | | 29 | | | | |
| 12 | | | | | 30 | | | | |
| 13 | | | | | 31 | | | | |
| 14 | | | | | 32 | | | | |
| 15 | | | | | 33 | | | | |
| 16 | | | | | 34 | | | | |
| 17 | | | | | 35 | | | | |
| 18 | | | | | 36 | | | | |

- Quel est votre sexe?
- Lequel des trois mots suivants décrit le mieux votre allégeance politique : libérale, modérée ou conservatrice?
- Le Canada devrait-il conserver ou abolir la pièce de 1 cent comme pièce de monnaie légale?
- Quelle est selon vous l'utilité de la statistique pour la société sur une échelle de 1 (utilité nulle) à 9 (très grande utilité)?

## Leçon 1 : Variables nominales (suite)

## Exercice

1. Dans le cas de chacune des variables mentionnées ci-après, détermine s'il s'agit d'une variable nominale (catégorie) ou d'une variable de mesure. S'il s'agit d'une variable nominale, indique si elle est binaire ou non.

   Ta taille, ton poids, la couleur de tes cheveux, ton mois de naissance, les cours que tu suis, la possession d'une carte de crédit (oui ou non).

2. Considère les réponses à la question sur l'allégeance politique.
   a) Combien d'élèves (pourcentage) ont indiqué être « libérale »? - « conservatrice »? - « modérée »?
   b) Représente les données sous forme d'un graphique à barres.
   c) Rédige une phrase ou deux sur ce que montrent tes calculs et ton graphique au sujet de la distribution des tendances politiques.

3. Le tableau ci-dessous indique le nombre de blessures liées à la pratique de sports qui ont été traitées aux États-Unis au cours d'une année récente, ainsi que les estimations du nombre de personnes pratiquant les diverses activités sportives.

| Sport | Blessures | Participants | Sport | Blessures | Participants |
|---|---|---|---|---|---|
| Basket-ball | 646 678 | 26 200 000 | Pêche | 84 115 | 47 000 000 |
| Cyclisme | 600 649 | 54 000 000 | Équitation | 71 490 | 10 000 000 |
| Baseball | 459 542 | 36 100 000 | Planche à roulettes | 56 435 | 8 000 000 |
| Football | 453 684 | 13 300 000 | Hockey | 54 601 | 1 800 000 |
| Soccer | 150 449 | 10 000 000 | Golf | 38 626 | 24 700 000 |
| Natation | 130 362 | 66 200 000 | Tennis | 29 936 | 16 700 000 |
| Volley-ball | 129 839 | 22 600 000 | Patinage | 29 047 | 7 900 000 |
| Patin à roulettes | 113 150 | 26 500 000 | Ski nautique | 26 633 | 9 000 000 |
| Haltérophilie | 86 398 | 39 200 000 | Quilles | 25 417 | 40 400 000 |

   a) Si l'on utilise le nombre de blessures comme mesure du danger d'un sport, quel est le sport le plus dangereux? Le cyclisme ou le football? Le soccer ou le hockey? La natation ou la planche à roulettes?
   b) Entre les données du tableau dans la calculatrice et calcule le taux de blessures par mille participants, dans le cas de chaque sport. D'après ce taux, quel est le sport le plus dangereux? Le cyclisme ou le football? Le soccer ou le hockey? La natation ou la planche à roulettes?
   c) Commente et compare tes réponses aux questions a) et c).
   d) Indique les trois sports les plus dangereux et les trois sports les moins dangereux, d'après le taux de blessures par mille participants.
   e) Identifie certains facteurs liés au danger d'un sport. (Quels sont les renseignements que tu pourrais utiliser pour produire une estimation plus précise du danger d'un sport?)

## Leçon 1 : Variables nominales (suite)
## Solutions

1. La taille et le poids sont des variables de mesure. Le mois de naissance, la couleur des cheveux et les cours que tu suis sont des variables nominales, mais non binaires, bien que l'on puisse affirmer que les mois peuvent être dénombrés. Quant à la possession d'une carte de crédit, il s'agit d'une variable nominale binaire.

2. Les réponses varient.

3. Le nombre de blessures par mille participants a été arrondi à une décimale près.

| Sport | Blessures/ 1000 | Sport | Blessures/ 1000 |
|---|---|---|---|
| Basket-ball | 24,7 | Pêche | 1,8 |
| Cyclisme | 11,1 | Équitation | 7,1 |
| Baseball | 12,7 | Planche à roulettes | 7,1 |
| Football | 34,1 | Hockey | 30,3 |
| Soccer | 15,0 | Golf | 1,6 |
| Natation | 2,0 | Tennis | 1,8 |
| Volley-ball | 5,7 | Patinage | 3,7 |
| Patin à roulettes | 4,3 | Ski nautique | 3,0 |
| Haltérophilie | 2,2 | Quilles | 0,6 |

a) Le cyclisme semble être plus dangereux que le football. Le soccer semble être plus dangereux que le hockey. La natation semble être plus dangereuse que la planche à roulettes.

b) Le tableau ci-dessus montre les divers taux de blessures par mille participants.

c) Toutes nos observations en (a) sont en ordre ascendant.

Une des observations pourrait être que six fois plus de personnes jouent au soccer qu'au hockey et que, même si le nombre de blessures est plus élevé au soccer, ce taux est loin d'être six fois plus élevé. Le tableau des taux montre que la probabilité d'une blessure est deux fois plus élevée au hockey qu'au soccer dans le cas de groupes composés du même nombre de participants.

d) Le football, le hockey et le basket-ball sont les sports les plus dangereux. Les quilles, le golf et la pêche ou le tennis sont les moins dangereux. On pourrait aussi classer la pêche et le tennis dans un ordre donné en utilisant plus de décimales. Devrions-nous faire ce calcul? Si oui, pourquoi?

e) Mon choix est le contact physique. Ce facteur est lié aux « contacts » entre les adversaires. Cependant, il y a certainement d'autres facteurs.

## Leçon 2 : Variables de mesure

### Activités de mise en marche

1. Les graphiques par points sont relativement faciles à réaliser à la main. Une fois terminés, ils ressemblent aux graphiques à barres. En effet, on peut tracer des lignes autour des points pour obtenir un graphique à barres. Nous utilisons des graphiques à points en raison de la facilité avec laquelle ils peuvent être réalisés à partir des données brutes recueillies. La calculatrice TI-83 ne permet pas d'effectuer facilement des graphiques à points, mais elle permet de tracer un graphique à barres correspondant à un graphique à points si l'on établit la largeur des barres (échelle *x*) à 1 unité. Lorsqu'on trace le graphique, la calculatrice indique la hauteur de chaque barre. Elle indique également les valeurs minimum et maximum de *x* pour chacune des barres. Si tu veux utiliser une échelle statistique classique, utilise 0,5 comme valeur minimum pour *x*. Chaque barre aura alors la valeur entière en son centre.

   Tu peux réaliser un graphique à points à l'aide de la calculatrice TI-83, mais il faut faire quelques opérations.  Voici un exemple.  Commence par entrer dans L1 les données du tableau.  (Ces données datent de 1996 et ont trait à la hauteur (en pieds) de grands immeubles de Philadelphie (Pennsylvanie)).

| | | | | | | |
|---|---|---|---|---|---|---|
| 548 | 405 | 375 | 400 | 475 | 450 | 412 |
| 375 | 364 | 492 | 482 | 384 | 490 | 492 |
| 490 | 435 | 390 | 500 | 400 | 491 | 945 |
| 435 | 848 | 792 | 700 | 572 | 739 | 572 |

Regroupe les immeubles dans des catégories à intervalles de dix pieds, de la façon suivante : déplace le curseur de manière à mettre le nom de la colonne L2 en surbrillance et appuie sur  ENTER.  Au bas de l'écran, le curseur devrait clignoter à côté de L2 =. Appuie sur MATH, choisis NUM, puis round (et ENTER).  Appuie maintenant sur L1/10,0) pour afficher un écran comme celui qui suit (figure 1).  Appuie sur ENTER pour voir que dans L2, les hauteurs sont arrondies au plus près multiple de 10, comme suit (figure 2).

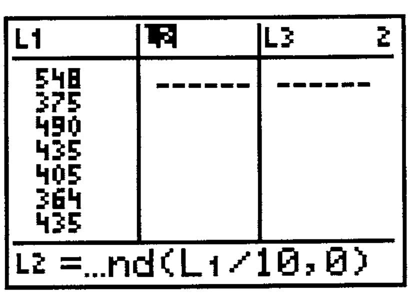



**Figure 1**

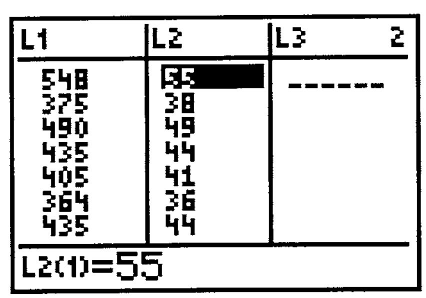



**Figure 2**

## Leçon 2 : Variables de mesure (suite)

Effectue maintenant le tri dans L2 (choix 2 du menu STAT) pour faire apparaître les listes reproduites ci-après.

Dans L3, entre le nombre d'occurrences pour chaque valeur. Par exemple, entre 1 à côté de la valeur 36. Entre 1 à côté de la première valeur 38, 2 à côté de la deuxième et 3 à côté de la troisième. Entre 1 à côté de la valeur 39, 1 à côté de la première valeur 40, 2 à côté de la deuxième, etc.

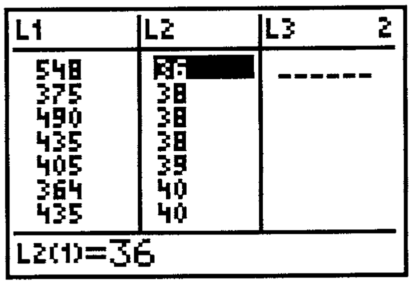



**Figure 3**

Appuie maintenant sur 2nd, puis sur STAT PLOT; choisis une représentation graphique et en particulier un graphique à points (premier choix), avec L2 comme liste X et L3 comme liste Y. Utilise la touche ZOOMSTAT pour obtenir un graphique à points comme celui-ci :

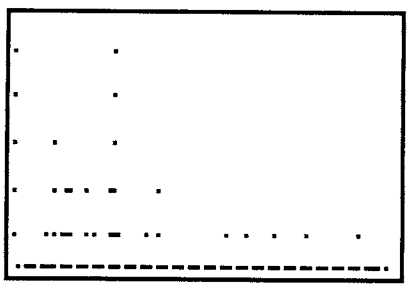


**Figure 4**

Il ne vaut peut-être pas la peine d'utiliser la calculatrice pour tracer des graphiques à points, mais tu préféreras peut-être cette méthode.

1. Si les élèves travaillent de manière autonome, il faut recueillir les données pour cette question. Un tableau ou une page qui peut être copié ferait l'affaire. Si l'enseignant établit la méthode de collecte de données et en informe les élèves, il doit également donner une date limite pour l'entrée de toutes les données. Cette façon de procéder permet de s'assurer que les élèves qui travaillent plus rapidement ont accès aux données lorsqu'ils en ont besoin.

## Leçon 2 : Variables de mesure (suite)

### Exercice

1. Décris la population, ainsi qu'un échantillon raisonnable, dans les cas énumérés ci-après.
   a) Un sondage d'opinions portant sur le contrôle des armes à feu.
   b) Une enquête sur l'emploi effectuée mensuellement par le gouvernement.
   c) Une pâtisserie a recours aux services d'un statisticien pour déterminer quel genre de publicité permettrait de faire augmenter les ventes de petits pains Kaiser.
   d) Un éditeur de musique essaie de déterminer si des stations de radio diffusent suffisamment ses titres.
2. Certaines stations de radio ont créé des numéros de téléphone « 900 » qui permettent d'enregistrer le nombre d'appels qu'elles reçoivent. Les auditeurs sont invités à composer tel numéro s'ils veulent répondre « OUI » à une question, ou tel autre numéro s'ils veulent répondre « NON ». Ils n'ont pas à dire un mot. Or, ces sondages téléphoniques font parfois des prédictions qui sont tout à fait différentes des faits une fois qu'ils sont rendus publiques. Quelles sont certaines des raisons qui peuvent expliquer cela?
3. Étudie les données recueillies précédemment auprès de la classe sur l'utilité de la statistique. Utilise le graphique à points que tu as réalisé pour répondre à ces questions.
   a) Combien d'élèves ont estimé l'utilité de la statistique à 5 sur l'échelle de 1 à 9? Quel est le pourcentage de ces élèves?
   b) Quelle proportion des élèves a estimé l'utilité de la statistique à une valeur supérieure à 5?
   c) Quelle proportion des élèves a estimé l'utilité de la statistique à une valeur inférieure à 5?
   d) Résume en quelques phrases ce que révèle le graphique à points sur la distribution des évaluations des élèves relatives à l'utilité de la statistique. Commente en particulier au degré d'acquiescement des élèves sur la question.

## Leçon 2 : Variables de mesure (suite)

### Réponses

1. a) La population pourrait être le groupe de personnes qui ont droit de vote dans la région en question. Un échantillon pourrait être composé d'un groupe restreint de personnes sélectionnées de façon aléatoire sur une liste de votants.

   b) La population pourrait être un groupe de personnes d'une circonscription donnée qui ont l'âge de voter et qui ne sont pas inscrites à une école ou à un programme de formation, ou un groupe de personnes qui ont occupé un emploi au cours de la dernière année, ou un quelconque autre groupe. Cette question montre la nécessité de définir clairement la population, afin d'éviter toute confusion.

   c) La population d'une enquête était composée d'acheteurs qui dépensent chaque semaine un montant supérieur à une certaine limite en produits d'épiceries. L'échantillon était composé de personnes sortant du magasin Safeway de Polo Park au cours d'une semaine donnée.

   d) Une telle population pourrait être composée de stations de radio AM qui diffusent de la musique pop. Ces stations prétendent peut-être qu'elles diffusent les 40 premières chansons du palmarès un certain nombre de fois par semaine. Un échantillon pourrait être composé d'un groupe de stations choisies au hasard à l'échelle du pays.

2. Un appelant ayant une opinion bien arrêtée sur la question pourrait composer à plusieurs reprises le même numéro sans révéler son identité. Ces appels seraient alors considérés comme des appels provenant de nombreuses personnes et les résultats obtenus seraient trompeurs.

3. Ces questions aboutiront probablement à un résumé indiquant un éventail de taux proches d'une valeur centrale. Les pourcentages enregistrés peuvent varier d'un groupe à l'autre.

## Leçon 3 : Échantillon aléatoire simple

### Activités de mise en marche

La meilleure façon de recueillir ces données consiste peut-être à afficher un tableau au mur où les élèves peuvent inscrire les longueurs moyennes. Ces données devraient être recueillies pour des échantillons de 4 et de 10 droites respectivement. Il faudrait également fixer une date limite.

**Exercice**

Si vous voulez effectuer une simulation de l'expérience relative à la proportion de filles dans votre classe, vous pouvez utiliser la calculatrice TI-83. Voici un exemple de simulation. Supposons qu'il y a un nombre d'élèves **É** dans votre classe, dont un nombre de filles **F**. Chaque élève choisit cinq noms dans un échantillon aléatoire simple. Les noms des filles sont identifiés par des chiffres allant de **1** à **F**. Les nombres **F+1** jusqu'à **É** sont attribués aux noms des garçons.

Le programme **Sample** utilisé ici simule ce processus. Les chiffres compris entre **1** et **É** sont choisis au hasard, sans double compte, dans des ensembles de nombres **T**. (Des ensembles de noms **T** différents peuvent avoir les mêmes nombres.) Ces ensembles de nombres **T** sont ordonnés par ordre ascendant et chaînés pour former la liste L3. Les élèves peuvent consulter cette liste pour déterminer combien de filles ont été choisies dans chaque échantillon.

```
Entre « Nbre D'ÉLÈVES? », É
Entre « TAILLE DE L'ÉCHANTILLON? »,T
Entre « Nbre d'ÉCHANTILLONS? »,U
ClrList L1,L2,L3
U+1→ Z
Lbl C
Z-1→ Z
Lbl A
seq(randInt(1,S),X,1,T,1)  L1
SortA(L1)
List(L1) → L2,
If prod(L2)=0
Goto A
If Z=U
Goto B
augment(L1,L2) → L3
While Z>1
Goto C
Else
End
Stop
Lbl B
L1 → L3
Goto C
End
```

## Leçon 3 : Échantillon aléatoire simple (suite)

**Longueurs moyennes des segments de droite**

| | **Échantillon de 4 noms** | **Échantillon de 10 noms** |
|---|---|---|
| 1 | | |
| 2 | | |
| 3 | | |
| 4 | | |
| 5 | | |
| 6 | | |
| 7 | | |
| 8 | | |
| 9 | | |
| 10 | | |
| 11 | | |
| 12 | | |
| 13 | | |
| 14 | | |
| 15 | | |
| 16 | | |
| 17 | | |
| 18 | | |

| | **Échantillon de 4 noms** | **Échantillon de 10 noms** |
|---|---|---|
| 19 | | |
| 20 | | |
| 21 | | |
| 22 | | |
| 23 | | |
| 24 | | |
| 25 | | |
| 26 | | |
| 27 | | |
| 28 | | |
| 29 | | |
| 30 | | |
| 31 | | |
| 32 | | |
| 33 | | |
| 34 | | |
| 35 | | |
| 36 | | |

## Leçon 3 : Échantillon aléatoire simple (suite)

### Exercice

Utilise la liste de ta classe pour effectuer cette expérience. Celle-ci a trait à des données nominales plutôt qu'à des données de mesure.

- Choisis cinq noms au hasard (à l'aide du générateur de nombres aléatoires de ta calculatrice, ou à l'aide d'une table de nombres aléatoires).
- Consigne le nombre de noms de filles choisies.
- Recueille et enregistre les données recueillies par la classe, et enregistre également le nombre de statistiques qui a été recueilli.
- Entre les statistiques dans une liste de l'éditeur de statistiques de ta calculatrice TI-83.
- Place le curseur sur l'entête d'une colonne vide. L'affichage ressemblera à celui-ci.

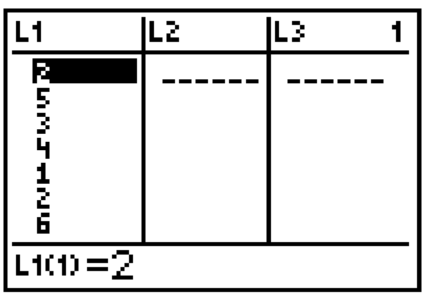



**Figure 1**

- Appuie maintenant sur ENTER. Le curseur descend au bas de l'écran et clignote à côté du symbole « = ». (Figure 2).

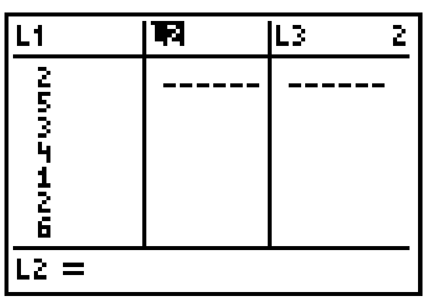



**Figure 2**

- Appuie sur 2nd puis sur L1 et ensuite sur / 5 . Appuie maintenant sur ENTER. La nouvelle liste indiquera la proportion de filles dans tes échantillons.
- Tu peux maintenant réaliser un graphique à points qui montre la distribution d'échantillon de ces statistiques.
- Refais le même exercice, mais en utilisant cette fois-ci un échantillon de dix noms.
- La dispersion de la proportion de filles choisies est-elle acceptable? Y a-t-il un biais? Quelle est la proportion moyenne choisie? Cette moyenne correspond-elle au nombre prévu? La taille de l'échantillon a-t-elle une incidence sur la dispersion de la statistique par rapport au paramètre?

## Leçon 4 : Description de distributions

### Activités de mise en marche

Les élèves devraient être en mesure de répondre aux questions a), b) et c). La question d) pourrait être difficile parce que certaines des distributions précédentes sont bimodales ou trimodales. Le point bas autour de 50 indique que la population présumée est en fait composée de 2 groupes distincts. La définition de la population était peut-être inadéquate.

**Déroulement**

Les élèves voudront peut-être entrer les données dans une liste de la calculatrice TI-83 et faire le tri de cette liste. Le tracé de base ordonné peut alors être réalisé immédiatement. La calculatrice TI-83 peut être utilisée pour obtenir le sommaire de cinq nombres. Après avoir entré les données dans L1 (ou dans toute autre liste), utilise les fonctions STAT, CALC, 1-Var Stats et tape 2nd L1. Tu devras utiliser la flèche descendante pour parcourir les statistiques à la recherche de la médiane et des quartiles.

La calculatrice produit également un tracé en boîte. Appuie sur 2nd Y= pour afficher l'écran STAT PLOT. Choisis Plot 1, place le curseur sur le mot On et appuie sur ENTER. De manière analogue, sélectionne le type de tracé en boîte qui a une ligne médiane dans la boîte et choisis L1 comme liste X et fréquence. Il se peut que tu doives modifier les dimensions de la fenêtre de graphique. La fonction ZOOM 9 est un moyen rapide d'afficher la fenêtre pour des tracés statistiques. En appuyant sur TRACE et en utilisant les flèches « droite » et « gauche », tu peux obtenir les valeurs du sommaire de cinq nombres.

## Leçon 4 : Description de distributions

### Exercice

1. Le tableau ci-dessous contient un échantillon de temps de réaction d'un sujet à un certain stimulus. Il comprend 90 entrées. Le temps a été mesuré en millisecondes.

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 14 | 11 | 15 | 07 | 07 | 20 | 10 | 14 | 09 | 08 | 06 | 12 | 12 | 10 | 14 | 11 | 13 |
| 13 | 11 | 12 | 10 | 08 | 09 | 14 | 18 | 12 | 10 | 10 | 11 | 07 | 17 | 12 | 09 | 09 | 11 |
| 14 | 12 | 12 | 10 | 09 | 07 | 11 | 09 | 18 | 06 | 12 | 12 | 10 | 08 | 14 | 15 | 12 | 11 |
| 11 | 08 | 11 | 10 | 13 | 08 | 11 | 11 | 13 | 20 | 06 | 13 | 13 | 08 | 09 | 16 | 15 | 11 |
| 20 | 08 | 17 | 12 | 19 | 14 | 17 | 12 | 18 | 16 | 15 | 16 | 10 | 20 | 11 | 19 | 20 | 13 |

Produis un ***tracé à tiges*** de ces données et utilise-le pour trouver le sommaire de cinq nombres relatif à la distribution des données.

a) Fais un tracé en boîte à la main, puis à l'aide de la calculatrice. Compare les deux tracés.

b) Utilise ta calculatrice pour trouver le sommaire de cinq nombres relatifs à la distribution des données. Compare tes réponses avec celles obtenues en (a).

c) Utilise la calculatrice pour produire un histogramme indiquant cette distribution des données.

d) Tu as étudié les affichages et tu connais les valeurs de la moyenne et de la médiane relatives à cette distribution des données. Explique pourquoi la moyenne et la médiane tombent aux points où elles se trouvent. Ces valeurs sont-elles proches ou éloignées l'une de l'autre?

2. Trace un graphique qui montre la distribution de la consommation d'essence sur l’autoroute par les voitures 1996 mentionnées ci-après. Explique ton choix de graphique.

| MODÈLE | Milles/gallon | MODÈLE | Milles/gallon |
|---|---|---|---|
| Buick Century | 31 | Lexus GS300 | 24 |
| Buick Regal | 29 | Lexus LS400 | 24 |
| Cadillac Eldorado | 26 | Lincoln Mark VIII | 24 |
| Chrysler Cirrus | 29 | Mazda 626 | 24 |
| Dodge Stratus | 36 | Nissan Maxima | 24 |
| Ford Taurus | 29 | Oldsmobile Aurora | 24 |
| Ford Thunderbird | 26 | Oldsmobile 88 | 24 |
| Hyundai Sonata | 29 | Rolls-Royce Silver Spur | 24 |
| Infiniti I30 | 28 | Toyota Camry | 24 |
| Infiniti Q45 | 22 | Volkswagon Passat | 24 |
| Jaguar XJ12 | 16 | Volvo 850 | 24 |

## Leçon 4 : Description de distributions (suite)

3. En 1995, les salaires moyen et médian versés aux joueurs de baseball des ligues majeures étaient de 275 000 dollars et de 1 089 000 dollars. Lequel des montants représente le salaire moyen et quel est le salaire médian? Justifie ta réponse.

4. Voici une série qui représente les périodes d'affectation des professeurs du département de statistique : {22, 19, 13, 8, 6, 4, 3, 1, 0}.

   a) Trouve les valeurs de la moyenne et de la médiane relatives à cette série.

   b) Supposons maintenant qu'on ait entré par erreur le nombre 222 comme première valeur. Calcule de nouveau les valeurs de la moyenne et de la médiane avec cette valeur au lieu du nombre 22.

   ***Remarque :*** Une mesure dont la valeur demeure relativement inchangée en présence de valeurs aberrantes (c'est-à-dire des points de données éloignés de la majorité des points) est dite ***résistante***.

   c) Laquelle des deux valeurs est résistante : la médiane ou la moyenne? Rédige quelques phrases pour expliquer pourquoi chacune des valeurs est, ou n'est pas, résistante.

## Leçon 4 : Description de distributions (suite)

### Réponses

1. Tu n'as peut-être pas besoin d'une réponse dans ce cas. Est-ce que d'autres élèves ont obtenu les mêmes résultats?

2. Cette question, par contre, pourrait nécessiter une discussion. Tu pourrais avoir choisi un graphique à tiges. Un graphique à points comporterait peu de points à chaque valeur « milles/gallon » et aurait en somme peu d'intérêt. Des tracés à tiges ou des graphiques à points pourraient être plus intéressants et montrer comment les données sont réparties. La distribution est-elle asymétrique? Quelle est sa portée? Quelle est sa valeur centrale? Ces aspects devraient être visibles (du moins approximativement).

   Un tracé en boîte serait également un choix intéressant et devrait révéler le même genre d'information si tu sais comment la chercher. Pourquoi ne pas essayer plusieurs types de tracé? Compare les diverses caractéristiques.

3. Nous savons que la majeure partie des équipes est composée de joueurs qui ne sont pas des vedettes. Plus de joueurs gagnent des salaires relativement faibles par rapport aux joueurs qui touchent des montants faramineux. Étant donné que la médiane est la valeur au-dessus de laquelle on trouve la moitié des salaires, cette valeur pourrait être assez faible, comparativement à la moyenne, qui est la valeur moyenne de tous les salaires. En effet, les montants astronomiques contribuent beaucoup à cette moyenne mais ne changent rien au fait que de nombreux joueurs gagnent relativement peu.

   Je pense que la valeur médiane est 275 000 dollars... ce qui n'est pas si mal! Il se peut fort bien que la moyenne soit propulsée à 1 089 000 dollars par quelques salaires très élevés.

4. a) Dans ce cas, la valeur médiane est clairement 8. La moyenne est environ 8,4.

   b) La valeur médiane est toujours 8. La moyenne est environ 30,7, ce qui correspond à une augmentation importante.

   c) La valeur médiane est résistante, la valeur moyenne non. Les questions auxquelles tu as répondu jusqu'ici devraient t'avoir permis de bien comprendre les raisons pour lesquelles il en est ainsi. Discute de ces raisons avec d'autres élèves.

## Leçon 5 : La règle empirique

### Déroulement

Les élèves peuvent utiliser leur calculatrice pour trouver la moyenne et l'écart type dans ce cas. Efface les listes de statistiques et appuie sur STAT et EDIT. Déplace le curseur au-dessus des noms des colonnes, puis vers la droite, jusqu'à ce que s'affiche une zone de nom libre. Tape le mot **COTE** dans cette zone, et dans la colonne correspondante, entre les chiffres 1 à 20. Procède de la même manière pour créer les listes **COMPTE** et **PROD**.

Après avoir entré les fréquences dans la liste **COMPTE**, place le curseur sur l'en-tête de la colonne **PROD** et appuie sur ENTER. Appuie sur 2nd et LIST, choisis **SCORE**, appuie sur * et enfin choisis la liste **COMPTE** et appuie sur ENTER. Les produits des colonnes **COTE** et **COMPTE** sont affichés dans la colonne PROD. Par la suite, passe à l'écran initial de la calculatrice, appuie sur 2nd et LIST, sélectionne le menu MATH, choisis SUM et appuie sur ENTER. Appuie sur 2nd et LIST et choisis **PROD** avant d'appuyer de nouveau sur ENTER. Le résultat, qui est la somme de toutes les cotes, est 2 177. Étant donné qu'il *y* avait 213 cotes, la cote moyenne est 2 177 ÷ 213 = 10,22 .

Comme dans le cas du problème sur le taux de métabolisme, pour trouver l'écart type des scores, nous devons créer des colonnes désignées par les noms **ÉCART** (écart des scores par rapport à la valeur moyenne) et **CARÉC** (carrés des déviations par rapport à la moyenne). Les contenus de **ÉCART** sont obtenus par le calcul **COTE** – 10,22, dans le cas de chaque entrée. La calculatrice effectue cette opération pour nous. Les contenus de **CARÉC** sont simplement les carrés des valeurs de la colonne **ÉCART**. Étant donné qu'il y a plusieurs cas de scores particuliers, il faut créer une autre colonne, désignée par le nom **TOTAL**, dont les contenus sont obtenus en multipliant **COMPTE** par **CARÉC**. Lorsqu'on veut connaître la somme de la colonne TOTAL, la calculatrice donne ***3156,6292***. Ainsi, d'après la formule mentionnée à la page I-80,

$$s_x = \sqrt{\frac{3156{,}6292}{212}}.$$

On obtient 3,8587. Nous avons donc $\overline{x}$ = 10,221 et $S_x$ = 3,859.

## Leçon 6 : La signification statistique

### Activités de mise en marche

Mettez un tableau à la disposition des élèves pour qu'ils y inscrivent les réponses à la question 1d). Fixez une date limite pour l'inscription de toutes les réponses. Les élèves qui travaillent de manière autonome ont tendance à prendre du retard s'il n'y a pas de rappel.

### Déroulement

Vous pouvez charger ce programme-ci, ainsi que d'autres programmes, dans un ordinateur et les récupérer à l'aide du programme TI-Graph Link. Il s'agit d'un programme fourni par Texas Instruments avec un câble qui se branche dans un port série. Les programmes peuvent être transférés d'une calculatrice à une autre au moyen de câbles courts fournis avec les appareils et de la touche LINK. (Appelez ce programme ***BONBONS***.)

```
For(N,1,50,1)
ClrHome
Output(5,5,"SAMPLE")
Output(5,12,N)
25 → dim(L1)
50 → dim(L2)
seq(iPart(rand+0,45),X,1,25,1)
sum(L1) → L2 (N)
L2(N)/25 → L3(N)
End
PlotsOff
0,05 → Xmin
1 → Xmax
0,1 → Xscl
–1 → Ymin
40 → Ymax
25 → Yscl
1 → Xres
Plot1(Histogram, L3)
DispGraph
```

## Leçon 6 : La signification statistique (suite)

### Exercice

Ces questions pourraient faire l'objet d'une brève discussion en classe. Dans le cas d'élèves qui travaillent de manière autonome, vous pourriez les informer quelques jours à l'avance de la tenue de cette activité. Cela encouragera peut-être les plus lents à progresser.

**Déroulement**

Mettez un tableau à la disposition des élèves pour qu'ils y inscrivent les résultats de la simulation. Certains élèves n'ont probablement pas de dés. Voici un programme (DICE) qui simule le jet d'un dé ou de deux dés.

```
AxesOff
FnOff :PlotsOff
0 → Xmin:94 → Xmax
0 → Ymin:57 → Ymax
Menu("***OPTIONS***","ONE DIE",1,"TWO DICE",2)
Lbl 1
1 → Z
Goto 0
Lbl 2
2 → Z
Lbl 0
0 → C
If Z=1
45 → H:25 → V
If Z=2
25 → H:25 V
ClrDraw
Lbl J
iPart((rand*6+1) → A
58-V → Q
H → P
Line(P-10,Q-12,P+13,Q-12)
Line(P+13,Q-12,P+13,Q+9)
Line(P+13,Q+9,P-10,Q+9)
Line(P-10,Q+9,P-10,Q-12)
If A=1
Text(V,H,"°")
If A=2
Then
Text(V-6,H-6,"°"):Text(V+6,H+6," ")
End
```

## Leçon 6 : La signification statistique (suite)

```
If A=3
Then
Text(V,H,"°"):Text(V-6,H-6,"°"):Text(V+6,H+6,"°")
End
If A=4 or A=5 or A=6
Then
Text(V-6,H-6,"°"):Text(V-6,H+6,"°"):Text(V+6,H-6,"°"):Text(V+6,H+6,"°")
End
If A=5
Text(V,H,"°")
If A=6
Then
Text(V,H-6,"°"):Text(V,H+6,"°")
End
If Z=2 and C=1
Then
Pause
Goto 0
End
If Z=2
Then
H+40 → H
C+1 → C
Goto J
End
Pause
Goto 0
Stop
```

## Leçon 6 : La signification statistique (suite)

Ce programme (WIDGT) a été conçu par Allan J. Rossman et son exécution prend environ dix minutes.

```
ClrHome
AxesOn
Disp "HOW MANY"
Input "BATCHES?",S
Output(8,3,"PLEASE WAIT")
PlotsOff
ClrDraw
FnOff
ClrList LMIDPT
ClrList LWIDGT
0 → dim( LWIDGT)
For(I,1,16)
0 → LWIDGT(I)
End
For(I,1,S)
max(randBin(15,1/3,1)) → W
LWIDGT(W+1)+1 → LWIDGT(W+1)
Output(5,8,I)
End
ClrHome
Disp « »
Pause LWIDGT
For(I,0,15)
I → MIDPT(I+1)
End
0 → Xmin
15 → Xmax
0 → Ymin
max( LWIDGT)+5 Ymax
1 → Xscl
Ymax/10 → Yscl
Plot1(Histogram, LMIDPT,WIDGT)
PlotsOn 1
DispGraph
```

## Leçon 7 : Aire sous la courbe normale

### Déroulement

Jusqu'à tout récemment, les statisticiens utilisaient des tables pour trouver la valeur de zones situées au-dessous de la courbe standard normale. Ces tables contiennent des cotes $z$ qui indiquent la distance, en écarts types, à laquelle se trouve une valeur donnée par rapport à la moyenne de la population. (Une telle table est reproduite ci-dessous.) Une ***cote z*** de 1,0 indique une valeur située à un écart type au-dessus de la moyenne, et la table indique que la proportion d'une population à distribution normale située entre la valeur moyenne et la cote $z$ de 1,0 est de 0,3413. Autrement dit, environ 34 % de la population est située entre la valeur moyenne et un écart type au-dessus de cette moyenne.

Trois faits importants

1. La valeur indiquée par la table représente la région au-dessous de la courbe standard normale située entre la moyenne (ou centre) de la distribution et le point auquel on s'intéresse.

   Par exemple, la valeur de la zone située entre la moyenne et un écart type vers la droite est de 0,3413. On peut donc dire que $\Pr(0 < z < 1) = 0{,}3413$.

2. La valeur de l'aire totale au-dessous de la courbe est 1; la valeur de chaque moitié de cette aire est donc 0,5. Nous pouvons donc calculer la valeur de la zone située à la droite ou à la gauche d'un point donné.

   Pour connaître la proportion de la population située à la gauche de $z = 1$, on ajoute cette valeur à celle de la zone située à la gauche de la moyenne. C'est-à-dire :
   $\Pr(-\infty < z < 1) = 0{,}5 + 0{,}3413 = 0{,}8413$.

3. La courbe est symétrique autour de la moyenne; la zone à droite d'un point donné est donc égale à la zone à gauche de ce point.

   Ainsi, si l'on veut connaître la proportion de la population située à la droite de $z = 1$, on peut effectuer la soustraction 0,5 – 0,3413, ce qui donne 0,1587.

## Leçon 7 : Aire sous la courbe normale (suite)

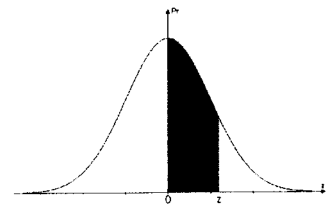


**Table de distribution standard normale**
**Pr(0 < $z$ < $z$)**

| $z$ | 0,00 | 0,01 | 0,02 | 0,03 | 0,04 | 0,05 | 0,06 | 0,07 | 0,08 | 0,09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0,0 | 0,0000 | 0,0040 | 0,0080 | 0,0120 | 0,0160 | 0,0199 | 0,0239 | 0,0279 | 0,0319 | 0,0359 |
| 0,1 | 0,0398 | 0,0438 | 0,0478 | 0,0517 | 0,0557 | 0,0596 | 0,0636 | 0,0675 | 0,0714 | 0,0753 |
| 0,2 | 0,0793 | 0,0832 | 0,0871 | 0,0910 | 0,0948 | 0,0987 | 0,1026 | 0,1064 | 0,1103 | 0,1141 |
| 0,3 | 0,1179 | 0,1217 | 0,1255 | 0,1293 | 0,1331 | 0,1368 | 0,1406 | 0,1443 | 0,1480 | 0,1517 |
| 0,4 | 0,1554 | 0,1591 | 0,1628 | 0,1664 | 0,1700 | 0,1736 | 0,1772 | 0,1808 | 0,1844 | 0,1879 |
| 0,5 | 0,1915 | 0,1950 | 0,1985 | 0,2019 | 0,2054 | 0,2088 | 0,2123 | 0,2157 | 0,2190 | 0,2224 |
| 0,6 | 0,2257 | 0,2291 | 0,2324 | 0,2357 | 0,2389 | 0,2422 | 0,2454 | 0,2486 | 0,2517 | 0,2549 |
| 0,7 | 0,2580 | 0,2611 | 0,2642 | 0,2673 | 0,2704 | 0,2734 | 0,2764 | 0,2794 | 0,2823 | 0,2852 |
| 0,8 | 0,2881 | 0,2910 | 0,2939 | 0,2967 | 0,2995 | 0,3023 | 0,3051 | 0,3078 | 0,3106 | 0,3133 |
| 0,9 | 0,3159 | 0,3186 | 0,3212 | 0,3238 | 0,3264 | 0,3289 | 0,3315 | 0,3340 | 0,3365 | 0,3389 |
| 1,0 | 0,3413 | 0,3438 | 0,3461 | 0,3485 | 0,3508 | 0,3531 | 0,3554 | 0,3577 | 0,3599 | 0,3621 |
| 1,1 | 0,3643 | 0,3665 | 0,3686 | 0,3708 | 0,3729 | 0,3749 | 0,3770 | 0,3790 | 0,3810 | 0,3830 |
| 1,2 | 0,3849 | 0,3869 | 0,3888 | 0,3907 | 0,3925 | 0,3944 | 0,3962 | 0,3980 | 0,3997 | 0,4015 |
| 1,3 | 0,4032 | 0,4049 | 0,4066 | 0,4082 | 0,4099 | 0,4115 | 0,4131 | 0,4147 | 0,4162 | 0,4177 |
| 1,4 | 0,4192 | 0,4207 | 0,4222 | 0,4236 | 0,4251 | 0,4265 | 0,4279 | 0,4292 | 0,4306 | 0,4319 |
| 1,5 | 0,4332 | 0,4345 | 0,4357 | 0,4370 | 0,4382 | 0,4394 | 0,4406 | 0,4418 | 0,4429 | 0,4441 |
| 1,6 | 0,4452 | 0,4463 | 0,4474 | 0,4484 | 0,4495 | 0,4505 | 0,4515 | 0,4525 | 0,4535 | 0,4545 |
| 1,7 | 0,4554 | 0,4564 | 0,4573 | 0,4582 | 0,4591 | 0,4599 | 0,4608 | 0,4616 | 0,4625 | 0,4633 |
| 1,8 | 0,4641 | 0,4649 | 0,4656 | 0,4664 | 0,4671 | 0,4678 | 0,4686 | 0,4693 | 0,4699 | 0,4706 |
| 1,9 | 0,4713 | 0,4719 | 0,4726 | 0,4732 | 0,4738 | 0,4744 | 0,4750 | 0,4756 | 0,4761 | 0,4767 |
| 2,0 | 0,4772 | 0,4778 | 0,4783 | 0,4788 | 0,4793 | 0,4798 | 0,4803 | 0,4808 | 0,4812 | 0,4817 |
| 2,1 | 0,4821 | 0,4826 | 0,4830 | 0,4834 | 0,4838 | 0,4842 | 0,4846 | 0,4850 | 0,4854 | 0,4857 |
| 2,2 | 0,4861 | 0,4864 | 0,4868 | 0,4871 | 0,4875 | 0,4878 | 0,4881 | 0,4884 | 0,4887 | 0,4890 |
| 2,3 | 0,4893 | 0,4896 | 0,4898 | 0,4901 | 0,4904 | 0,4906 | 0,4909 | 0,4911 | 0,4913 | 0,4916 |
| 2,4 | 0,4918 | 0,4920 | 0,4922 | 0,4925 | 0,4927 | 0,4929 | 0,4931 | 0,4932 | 0,4934 | 0,4936 |
| 2,5 | 0,4938 | 0,4940 | 0,4941 | 0,4943 | 0,4945 | 0,4946 | 0,4948 | 0,4949 | 0,4951 | 0,4952 |
| 2,6 | 0,4953 | 0,4955 | 0,4956 | 0,4957 | 0,4959 | 0,4960 | 0,4961 | 0,4962 | 0,4963 | 0,4964 |
| 2,7 | 0,4965 | 0,4966 | 0,4967 | 0,4968 | 0,4969 | 0,4970 | 0,4971 | 0,4972 | 0,4973 | 0,4974 |
| 2,8 | 0,4974 | 0,4975 | 0,4976 | 0,4977 | 0,4977 | 0,4978 | 0,4979 | 0,4979 | 0,4980 | 0,4981 |
| 2,9 | 0,4981 | 0,4982 | 0,4982 | 0,4983 | 0,4984 | 0,4984 | 0,4985 | 0,4985 | 0,4986 | 0,4986 |
| 3,0 | 0,4987 | 0,4987 | 0,4987 | 0,4988 | 0,4988 | 0,4989 | 0,4989 | 0,4989 | 0,4990 | 0,4990 |
| 3,1 | 0,4990 | 0,4991 | 0,4991 | 0,4991 | 0,4992 | 0,4992 | 0,4992 | 0,4992 | 0,4993 | 0,4993 |
| 3,2 | 0,4993 | 0,4993 | 0,4994 | 0,4994 | 0,4994 | 0,4994 | 0,4994 | 0,4995 | 0,4995 | 0,4995 |
| 3,3 | 0,4995 | 0,4995 | 0,4995 | 0,4996 | 0,4996 | 0,4996 | 0,4996 | 0,4996 | 0,4996 | 0,4997 |
| 3,4 | 0,4997 | 0,4997 | 0,4997 | 0,4997 | 0,4997 | 0,4997 | 0,4997 | 0,4997 | 0,4997 | 0,4998 |
| 3,5 | 0,4998 | 0,4998 | 0,4998 | 0,4998 | 0,4998 | 0,4998 | 0,4998 | 0,4998 | 0,4998 | 0,4998 |
| 3,6 | 0,4998 | 0,4998 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 |
| 3,7 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 |