Unité E
Variation et analyse statistique

VARIATION ET ANALYSE STATISTIQUE

Introduction

Ce module présente aux élèves deux méthodes d'utilisation des statistiques pour décrire des données et tirer des conclusions dans des situations de la vie de tous les jours. On y met l'accent sur les distributions normale et quasi normale, sur les inférences qui peuvent être établies à partir d'échantillons et sur les échantillons des populations ciblées. Certains concepts de l'Unité D, Probabilité, sont requis avant que l'élève puisse entreprendre l'Unité E.

Pratiques d'enseignement

Les élèves utiliseront la technologie pour effectuer des calculs à propos d'ensembles de données. Certaines théories doivent être enseignées, mais, à ce niveau, l'enseignement théorique rigoureux n'est ni prévu ni requis. L'interprétation des résultats est importante, mais l'analyse approfondie de ces données sera effectuée lors de cours subséquents.

Matériel d'enseignement

- calculatrices graphiques
- tableurs informatiques, *WinStat (gratuiciel, version française)*

Le matériel d'apprentissage de l'élève fourni dans le cours *Mathématiques appliquées S4 – Exercices* a été classé dans des **activités** ou **exercices d'apprentissage**. Dans le cadre des **activités d'apprentissage**, l'élève doit utiliser la technologie pour résoudre un problème plus important, plus global. Les **exercices** contiennent habituellement un plus grand nombre de questions que les élèves peuvent choisir de résoudre à l'aide de la technologie ou de méthodes plus traditionnelles. Les élèves doivent effectuer des **activités** et des **exercices d'apprentissage**. Le matériel de l'élève a été réparti comme il est illustré ci-dessous.

Résultat	Activité	Exercice
E-1	1, 2, 3, 4	1
E-2/3	5	2, 3
E-3	6	4, 5, 6

Durée

14 heures

RÉSULTATS D'APPRENTISSAGE

Résultat général

Utiliser les distributions normale et binomiale pour résoudre des problèmes comprenant des incertitudes.

Résultats spécifiques

E-1 Déterminer l'écart type de la population à partir d'un ensemble de données et à l'aide de la technologie.

STRATÉGIES PÉDAGOGIQUES

Dans cette section, les élèves apprendront de quelle manière des conclusions sont tirées de données obtenues à partir d'enquêtes, d'échantillons et de populations primaires. Le processus de résolution des problèmes doit comprendre des discussions et évaluations à propos de différents procédés.

Les élèves doivent se familiariser avec les termes techniques associés aux enquêtes et aux échantillons. Veuillez consulter l'annexe E-1, page E-40, pour obtenir la liste des termes et définitions, ainsi que l'annexe E-2, page E-42, pour obtenir des informations sur l'utilisation d'une calculatrice graphique.

Cette unité met l'accent sur les concepts généraux élaborés à partir des exemples et applications.

Une étude théorique de l'utilisation des distributions de la probabilité sera effectuée une fois que les concepts auront été élaborés et une fois que les élèves auront acquis une compréhension intuitive des possibilités et des limites des méthodes statistiques grâce à une enquête sur les intervalles de confiance.

Pour certains problèmes, des données réelles devront être obtenues de sources telles que Statistique Canada. Le site web de Statistique Canada permet à l'élève d'avoir accès à une vaste gamme de sources de données au Canada et partout dans le monde (voir l'annexe E-4, page E-45).

- **Déterminer l'écart type de la population à partir d'un ensemble de données et à l'aide de la technologie.**

  Examinez les termes suivants : population, échantillon, moyenne, médiane, mode, données discrètes et données continues. Veuillez consulter l'annexe E-1 pour obtenir les définitions de ces termes.

  Exemple 1

  Le vendeur d'une boutique de vêtements tient à jour un registre du nombre de pantalons vendus pendant la semaine. Vous trouverez ci-dessous son registre pour deux semaines. Vous devez calculer la moyenne, la médiane et le mode pour ces données en utilisant une calculatrice graphique pour confirmer la moyenne et la médiane de cet ensemble de données.

	lundi	mardi	mercredi	jeudi	vendredi	samedi
semaine 1	34	40	36	36	38	38
semaine 2	32	36	36	42	34	34

— suite

STRATÉGIES D'ÉVALUATION

NOTES

Voir l'annexe E-4, page E-45, pour obtenir la liste des ressources Internet. Ces ressources peuvent être utilisées pour effectuer l'activité 1.

Ressources imprimées

Mathématiques appliquées, Secondaire 4 – Exercices – Supplément au programme d'études, Éducation et Formation professionnelle Manitoba, 2000.

Mathématiques appliquées, Secondaire 4 – Cours destiné à l'enseignement à distance, Éducation et Formation professionnelle Manitoba, 2000.
— Module 8, Leçons 1, 2 et 3

RÉSULTATS D'APPRENTISSAGE SPÉCIFIQUES

E-1 Déterminer l'écart type de la population à partir d'un ensemble de données et à l'aide de la technologie.
– suite

STRATÉGIES PÉDAGOGIQUES

- **Déterminer l'écart type de la population à partir d'un ensemble de données et à l'aide de la technologie. (suite)**

Exemple 1 — suite

Solution

Moyenne : $\bar{x} = \dfrac{436}{12} = 36,33$

Puisque $\bar{x}$ correspond à l'estimation la plus exacte de μ, la moyenne de la population, le matériel utilise parfois μ plutôt que $\bar{x}$

Médiane : Ces données, présentées en ordre croissant sont :

32, 34, 34, 34, 36, 36, 36, 36, 38, 38, 40, 42

Médiane : $\dfrac{(36 + 36)}{2} = 36$

Mode : Le mode correspond à 36.

Nous vous encourageons à utiliser un tableur ou une calculatrice graphique. Veuillez consulter l'annexe E-2(a) pour connaître les étapes requises pour déterminer la moyenne et la médiane à l'aide de la calculatrice graphique TI-83.

La liste de droite ci-dessous fournit les renseignements indiqués à propos de vos données. (Vous pouvez faire défiler la liste en utilisant les touches fléchées vers le haut et vers le bas.)

Moyenne	$\bar{x} = 36,33$
Somme	$\sum x = 436$
Somme des carrés	$\sum x^2 = 15\ 928$
Écart type de l'échantillon	$s_x = 2,81$
Écart type de la population	$\sigma_x = 2,69$
Nombre d'éléments de données	$n = 12$
Minimum	min $x = 32$
Premier quartile	$Q_1 = 34$
Médiane	Méd = 36
Troisième quartile	$Q_3 = 38$
Maximum	max $x = 42$

Vous remarquerez que deux écarts types différents sont indiqués. La différence entre ces deux écarts types est expliquée ci-dessous.

STRATÉGIES D'ÉVALUATION

NOTES

RÉSULTATS D'APPRENTISSAGE SPÉCIFIQUES

E-1 Déterminer l'écart type de la population à partir d'un ensemble de données et à l'aide de la technologie.

– suite

STRATÉGIES PÉDAGOGIQUES

- **Déterminer l'écart type de données.**

  L'écart type d'un ensemble de scores indique la manière dont ces scores sont répartis par rapport à la moyenne.

  La formule ci-dessous est utilisée pour calculer σ, l'écart type d'une population de taille n.

$$\sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n}}$$

Exemple

Supposons que la population correspond à 2, 5, 8, 8, 9, 12, 13, 15 et 18. Déterminez l'écart type.

Solution

La moyenne de ces scores est de $\frac{\sum x}{n} = \frac{90}{9} = 10$.

La somme des carrés des écarts est la suivante :
$8^2 + 5^2 + 2^2 + 2^2 + 1^2 + 2^2 + 3^2 + 5^2 + 8^2 = 200$

Donc, $\sigma = \sqrt{\frac{200}{9}} \approx 4{,}7$.

Bien que certaines formules permettent une réduction des opérations arithmétiques qui doivent être effectuées, les populations plus grandes et les moyennes qui ne sont pas exprimées en nombres entiers entraîneront des calculs arithmétiques considérables. Les logiciels informatiques et les calculatrices comportant des touches σ sont maintenant fréquemment utilisés pour déterminer les écarts types.

- **Estimer l'écart type d'une population à partir d'un échantillon.**

  Parfois, nous disposons d'un ***échantillon*** d'une population, et nous désirons estimer l'écart type de cette population. Si, pour ce faire, nous utilisons la formule ci-dessus, l'estimation que nous obtenons est petite. Mais si nous utilisons $(n - 1)$ plutôt que n dans la formule, nous obtenons la meilleure estimation possible de l'écart type de la population.

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

> La plupart des calculatrices comportent des touches σ et s (comme dans l'exemple fourni ci-dessus). Sauf avis contraire, l'écart type correspond à σ.

STRATÉGIES D'ÉVALUATION

Problèmes

1. Les 30 élèves d'une classe reçoivent les notes ci-dessous à un examen de mathématiques. Utilisez la technologie pour déterminer la moyenne et l'écart type de cet ensemble de notes.

78	92	62	52	65	59
53	63	68	73	71	63
69	74	73	81	55	71
75	81	84	77	80	75
41	57	91	62	65	49

2. Les supermarchés E et S vendent des sacs de 5 kg de pommes de terre. Dix sacs sont choisis au hasard pour obtenir une estimation de la moyenne et de l'écart type de tous les sacs vendus dans ces supermarchés. Les poids des dix sacs choisis étaient les suivants (en kg) : 5,4, 5,4, 5,3, 5,2, 5,3, 5,3, 5,1, 5,0, 4,9 et 5,1. Donnez des estimations de la moyenne et de l'écart type obtenus.

Solutions

1. Moyenne = 69; σ_x = 12
2. Moyenne = 5,2 kg

Estimation de l'écart type de tous les sacs = S_x = 0,17 kg

NOTES

Ressources technologiques

Calculatrice graphique, telle que la calculatrice TI-83

Guide pour la calculatrice graphique

WinStat. Ce logiciel exécutera les mêmes opérations statistiques que la calculatrice graphique. Le logiciel est gratuit et peut être téléchargé à partir de l'adresse suivante : http://meth.exeter.edu/rparris/. Une version française sera disponible sous peu.

Tableur tel que *Excel* ou *Works*

L'écart-type de données est représenté par la lettre grecque minuscule σ, qui se vit sigma. La lettre majuscule sigma (Σ) désigne une sommation.

RÉSULTATS D'APPRENTISSAGE SPÉCIFIQUES

E-1 Déterminer l'écart type de la population à partir d'un ensemble de données et à l'aide de la technologie.
– suite

STRATÉGIES PÉDAGOGIQUES

- **Utiliser des données groupées.**

Nous pouvons déterminer la moyenne, la médiane et le mode de données groupées en faisant des calculs manuels ou en utilisant la technologie. (Voir l'annexe E-2.)

Exemple 1

On a demandé à 400 personnes combien de cassettes vidéo elles avaient louées au cours du dernier mois. Utilisez une calculatrice graphique pour déterminer la moyenne et la médiane de la distribution statistique indiquée ci-dessous et tracez un histogramme. Aussi, déterminez le mode à l'aide de la distribution statistique. (La calculatrice graphique peut ne pas afficher le mode.)

N^{bre} de cassettes vidéo louées	**N^{bre} de personnes**
1	28
2	102
3	160
4	70
5	25
6	13
7	0
8	2

Solution

Calculs manuels :

Nombre de cassettes vidéo louées :

(1 x 28) + (2 x 102) + (3 x 160) + (4 x 70) + (5 x 25) + (6 x 13) +

(7 x 0) + (8 x 2) = 1 211

Le nombre moyen de cassettes louées par 400 personnes est de

$$\frac{1\,211}{400} = 3{,}027\,5\,.$$

La médiane se situera dans l'intervalle contenant la 200^{e} personne (en commençant par le début ou par la fin). Puisque 28 + 102 = 130 et que 28 + 102 + 160 = 290, la médiane correspond à 3.

Le tableau indique que le mode correspond à 3.

– suite

STRATÉGIES D'ÉVALUATION

Problème

Une expérience est effectuée pour déterminer la masse approximative d'une pièce d'un cent. Pour ce faire, on pèse 300 pièces et on enregistre les masses dans le tableau de la distribution statistique ci-dessous.

Masse (grammes)	2,7	2,8	2,9	3,0	3,1	3,2	3,3	3,4
Fréquence	2	4	34	71	94	74	17	4

Utilisez la technologie pour déterminer la moyenne, la médiane, le mode et l'écart type des données. Créez un histogramme qui illustre les fréquences des différentes masses de cet ensemble de pièces d'un cent.

Solution

Enregistrez les valeurs ci-dessous.

X_{min} = 2,5

X_{max} = 3,5

X_{scl}= 0,1

Y_{min} = –10

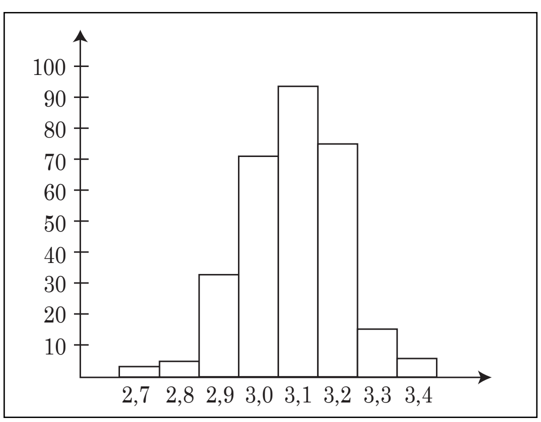
Y_{max} = 100

Moyenne = μ = 3,087 g

Médiane = 3,1 g

Mode = 3,1 g

Écart type = 0,12 g



NOTES

RÉSULTATS D'APPRENTISSAGE SPÉCIFIQUES

E-1 Déterminer l'écart type de la population à partir d'un ensemble de données et à l'aide de la technologie.

– suite

STRATÉGIES PÉDAGOGIQUES

- **Utiliser des données groupées. (suite)**

Exemple 1 — suite

Solution — suite

À l'aide d'une calculatrice graphique :

Consultez l'annexe E-2 pour connaître les étapes requises pour enregistrer les données d'une distribution statistique.

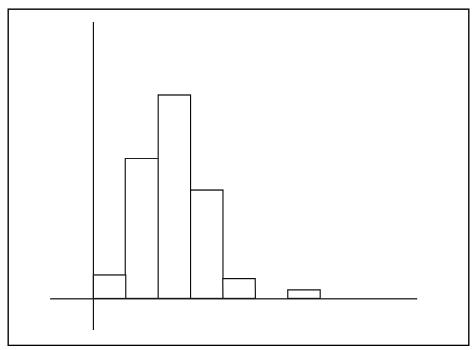
Sélectionnez $\bar{x}$ et la médiane à partir de la liste de résultats.

a) $\bar{x}$ = 3,027 5

b) Médiane = 3

c) Mode = 3

L'histogramme produit par la calculatrice ressemble à celui ci-dessous. Consultez l'annexe E-2(c) pour connaître les étapes requises pour produire un histogramme à l'aide de la calculatrice graphique TI-83.



À l'aide d'un tableur :

La moyenne peut aussi être calculée à l'aide d'un tableur. Enregistrez les formules ci-dessous dans le tableur.

	A	B	C
1	N^{bre} de vidéos = x	N^{bre} de personnes = f	Total = x*f
2	1		=A2*B2
3	2		=A3*B3
4	3		=A4*B4
9	8		=A9*B9
10		=Somme(B2:B9)	=Somme(C2:C9)
11		Moyenne =	=C10/B10

Solution

La moyenne correspond à 3,027 5.

L'option de création de graphiques du tableur peut être utilisée pour produire l'histogramme.

– suite

STRATÉGIES D'ÉVALUATION

NOTES

RÉSULTATS D'APPRENTISSAGE SPÉCIFIQUES

E-1 Déterminer l'écart type de la population à partir d'un ensemble de données et à l'aide de la technologie.

– suite

STRATÉGIES PÉDAGOGIQUES

- **Utiliser des données groupées. (suite)**

Exemple 2

On croit qu'une machine à emballer des bonbons est défectueuse. Les paquets doivent être de 90 grammes, mais en choisissant dix paquets au hasard, on s'est rendu compte que les masses réelles, en grammes, étaient les suivantes :

86, 91, 89, 92, 90, 93, 90, 90, 91, 88

Si l'écart des masses est trop grand, on considère que la machine est défectueuse. L'écart type est utilisé pour juger si la machine est défectueuse sur le plan statistique. Ainsi, si l'écart type de l'ensemble de scores est supérieur à 1,3, on considère que la machine est défectueuse et qu'elle requiert un ajustement ou une réparation. La machine est-elle défectueuse?

Solution

Le tableau ci-dessous illustre les différentes étapes à effectuer pour le calcul de la moyenne et de l'écart type.

n	**Masses**	**Écart avec la moyenne**	**Carré de l'écart**
1	86	–4	16
2	88	–2	4
3	89	–1	1
4	90	0	0
5	90	0	0
6	90	0	0
7	91	1	1
8	91	1	1
9	92	2	4
10	93	3	9
	$\Sigma x = 900$		$\Sigma(x - x)^2 = 36$

$$\text{Moyenne} = \bar{x} = \frac{\sum x}{n} = \frac{900}{10} = 90$$

$$\text{Écart type} = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} = \sqrt{\frac{36}{10}} \approx 1{,}9$$

Cet écart est supérieur à 1,3, ce qui signifie que la machine est défectueuse.

STRATÉGIES D'ÉVALUATION

NOTES

Résultats d'apprentissage spécifiques	**Stratégies pédagogiques**

E-2 Utiliser les cotes z et les tableaux de cotes z pour résoudre des problèmes.

- **Présenter les scores standardisés.**

Lorsque vous devez comparer deux distributions ou plus ayant des moyennes ou des écarts types différents, il peut être utile de ***standardiser*** les scores. Le score standardisé se nomme ***cote z*** ou et est calculé à l'aide de la formule ci-dessous :

$$z = \frac{(x - \overline{x})}{\sigma}$$

Dans cette formule, $\overline{x}$ correspond à la moyenne et σ correspond à l'écart type. Cette formule indique à combien d'écarts types un score, x, se situe de la moyenne et la position d'un score individuel dans la distribution.

Exemple 1

Si la moyenne = 21 et les écarts type = 4, calculez les cotes z des scores suivants : 25, 17, 26,5 et 16,5.

Solution

a) $z = \dfrac{(x - \overline{x})}{\sigma} = \dfrac{(25 - 21)}{4} = 1$

b) $z = \dfrac{(17 - 21)}{4} = -1$

c) $\dfrac{(26{,}5 - 21)}{4} = 1{,}375$

d) $\dfrac{(16{,}5 - 21)}{4} = -1{,}125$

Exemple 2

Si deux candidats qui ont présenté une demande de bourse ont effectué des examens différents mais aussi valables l'un que l'autre et ayant des moyennes et des écarts types différents, leurs résultats peuvent être comparés à l'aide des cotes z. Jean a obtenu une note de 206 à un examen dont la moyenne est de 190 et dont l'écart type est de 8. Jacques a obtenu une note de 91 à un examen dont la moyenne est de 81 et dont l'écart type est de 4. Qui a obtenu la meilleure note?

Solution

La cote z pour Jean est de : $(206 - 190) \div 8 = 16 \div 8 = 2$

La cote z pour Jacques est de : $(91 - 81) \div 4 = 10 \div 4 = 2{,}5$

Donc, on peut considérer que la note de Jacques est meilleure.

STRATÉGIES D'ÉVALUATION

Problèmes

1. Deux élèves ont obtenu les notes suivantes dans des cours de mathématiques comparables. Joanne a obtenu une note de 82 % à un examen dont la moyenne est de 78 % et dont l'écart type est de 5 %. Hélène a obtenu une note de 73 % à un examen dont la moyenne est de 62 % et dont l'écart type est de 8 %. Utilisez les cotes *z* pour comparer ces notes.
2. Le directeur et les enseignants d'une école décident d'utiliser les cotes *z* pour indiquer les notes obtenues à des examens à l'aide de lettres. Ils conviennent de ce qui suit :

   F = moins de –1

   D = de –1 à moins de 0,5

   C = de –0,5 à moins de 0,5

   B = de 0,5 à moins de 1

   A = de 1 et plus

   Les notes d'examens pour une classe de 36 élèves sont les suivantes :

   23, 34, 36, 39, 42, 44, 48, 50, 52, 54, 54, 55, 62, 62, 63, 64, 64, 65, 66, 67, 70, 71, 71, 75, 80, 81, 83, 85, 87, 88, 89, 94, 96, 98, 100, 100

   Déterminez les secteurs des notes qui s'appliquent aux lettres utilisées.

Solutions

1. $z_J = (82 - 78)/5 = 0{,}8$

   $z_H = (73 - 62)/8 = 1{,}38$

   Donc, la note d'Hélène est relativement plus élevée.

2. $\overline{x} = \dfrac{2412}{36} = 67, \ \sigma = 20$

   Les secteurs applicables aux lettres utilisées sont indiqués ci-dessous.

Écart réduit	Note réelle
-1	67 - 20 = 47
-0,5	67 - 10 = 57
0,5	67 + 10 = 77
1	67 + 20 = 87

Note en lettre	Secteur
F	0 - 46
D	47 - 56
C	57 - 76
B	77 - 86
A	87 -

NOTES

Ressources imprimées

Mathématiques appliquées, Secondaire 4 – Cours destiné à l'enseignement à distance, Winnipeg, MB : Éducation et Formation professionnelle Manitoba, 2000.

— Module 8, Leçon 4

Nota : Les termes ***écart réduit*** et ***cote z*** peuvent être utilisés de façon interchangeable.

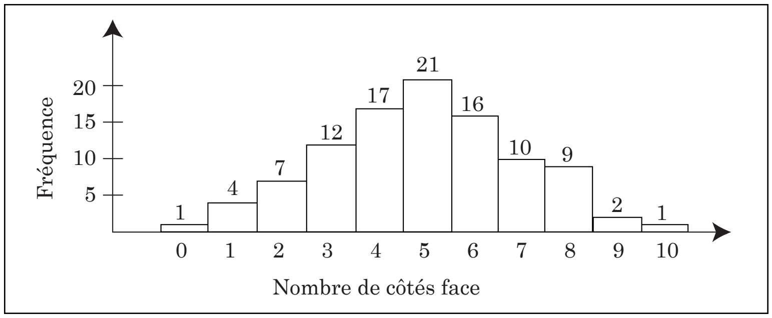
RÉSULTATS D'APPRENTISSAGE SPÉCIFIQUES

E-3 Utiliser la distribution normale et l'approximation normale de la distribution binomiale pour résoudre des problèmes reliés à des intervalles de confiance d'échantillons de grande taille.

STRATÉGIES PÉDAGOGIQUES

- **Présenter les distributions binomiales.**

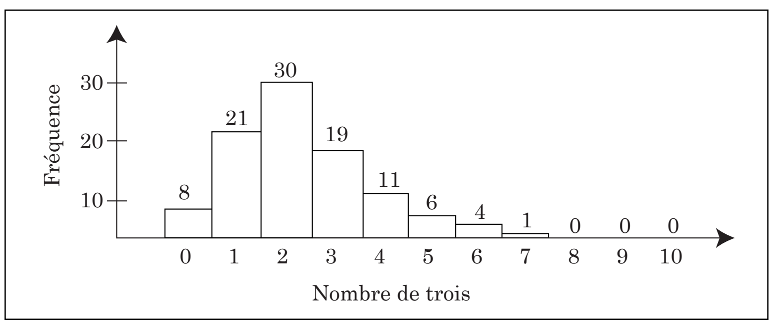
  Si nous lancions dix pièces de dix sous 100 fois dans les airs et si nous inscrivions le nombre de côtés face obtenu à chaque lancer, notre graphique de données pourrait ressembler au graphique ci-dessous.



Nous obtiendrions souvent quatre, cinq ou six côtés face et rarement zéro ou dix côtés face.

Puisque chaque pièce offre deux résultats possibles, la distribution globale se nomme une ***distribution binomiale***. Lorsque les deux résultats sont aussi probable l'un que l'autre, comme dans l'exemple ci-dessus, la distribution binomiale sera ***symétrique***.

Parfois, un des résultats n'est pas aussi probable que l'autre. Par exemple, si nous devions inscrire le nombre de trois obtenus lorsque nous lançons dix dés 100 fois, la distribution pourrait ressembler à celle illustrée ci-dessous.



Dans ce cas, la distribution binomiale n'est pas symétrique.

La moyenne, la médiane et le mode des données de toute distribution binomiale peuvent être déterminés à l'aide de calculs manuels ou à l'aide de la technologie, à l'instar de toutes les données illustrées dans un tableau.

STRATÉGIES D'ÉVALUATION

NOTES

Ressources imprimées

Mathématiques appliquées, Secondaire 4 – Cours destiné à l'enseignement à distance, Winnipeg, MB : Éducation et Formation professionnelle Manitoba, 2000.

— Module 8, Leçons 5, 6, 7 et 8

RÉSULTATS D'APPRENTISSAGE SPÉCIFIQUES

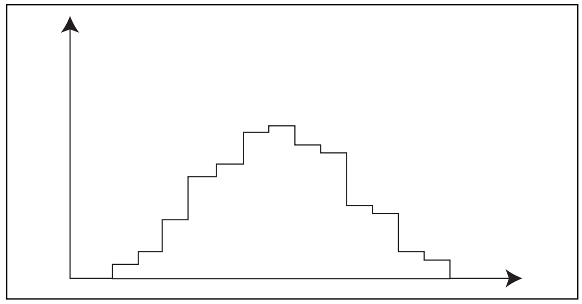
STRATÉGIES PÉDAGOGIQUES

E-3 Utiliser la distribution normale et l'approximation normale de la distribution binomiale pour résoudre des problèmes reliés à des intervalles de confiance d'échantillons de grande taille.

– suite

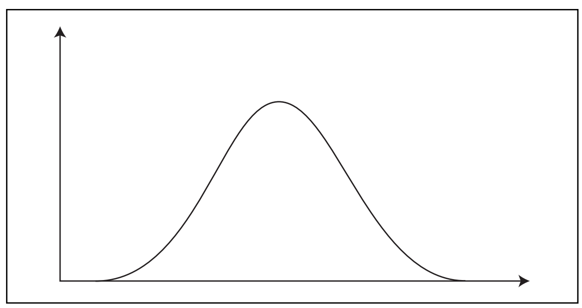
- **Présenter la distribution normale.**

  Lorsque les événements sont aussi probables l'un que l'autre, la distribution binomiale ressemble au graphique ci-dessous.



On remarque une élévation au centre et deux secteurs peu élevés aux limites.

Plus le nombre d'événements et d'essais est grand, plus le contour du graphique de la distribution est adouci. Par exemple, si nous lançons 1 000 pièces d'argent 1 000 000 de fois dans les airs, le contour du graphique de la distribution ressemblerait au contour ci-dessous.



Ainsi, notre distribution s'approcherait de la distribution théorique que l'on nomme la ***distribution normale***.

Lorsque nous effectuons des mesures, nous accumulons des erreurs mineures, et ces erreurs sont souvent réparties également d'un côté et de l'autre. C'est pourquoi les résultats des mesures effectuées ressemblent souvent à des distributions normales.

Vous trouverez ci-dessous des exemples de mesures qui produisent des distributions quasi normales.

- Si chaque élève d'une classe mesure la longueur de sa classe en utilisant une règle de 10 cm et s'il enregistre la longueur mesurée à un demi-centimètre près, le contour du graphique des mesures devrait se rapprocher d'une distribution normale puisque les erreurs mineures sont inévitables lorsqu'une règle est utilisée et qu'il est probable que les erreurs faites d'un côté et de l'autre soient réparties de manière égale.

– suite

STRATÉGIES D'ÉVALUATION

NOTES

RÉSULTATS D'APPRENTISSAGE SPÉCIFIQUES

E-3 Utiliser la distribution normale et l'approximation normale de la distribution binomiale pour résoudre des problèmes reliés à des intervalles de confiance d'échantillons de grande taille.
– *suite*

STRATÉGIES PÉDAGOGIQUES

- **Présenter la distribution normale. (suite)**
  - Les tests de QI (quotient intellectuel) pour tous les âges comprennent habituellement des questions auxquelles la personne moyenne de cet âge a 50 % des chances de répondre correctement. Les scores de QI obtenus dans la population générale à l'aide de ces tests sont habituellement répartis selon une distribution normale.

    Cela ***ne signifie pas*** que l'intelligence (peu importe comment elle est définie) est normalement distribuée. Cela signifie plutôt que le test a été prévu de cette façon. Par exemple, si nous utilisions des questions auxquelles les personnes répondraient correctement dans 80 % des cas, ou si nous combinions des questions auxquelles des réponses exactes sont fournies dans 80 % et dans 20 % des cas, nous pourrions obtenir presque toutes les distributions possibles.
  - Votre grandeur dépend de plusieurs facteurs, dont l'alimentation, mais le principal facteur demeure la combinaison des gènes dont vous héritez de vos parents. Certains gènes favorisent votre croissance, d'autres non. Vous avez environ 50 % des chances d'hériter des gènes de chacun de vos parents.

    Évidemment, les parents qui sont grands ont plus de chances d'avoir de grands enfants et les parents qui sont courts ont plus de chances d'avoir des enfants courts, mais pour les raisons expliquées ci-dessus, la distribution des grandeurs dans la population semble normale.

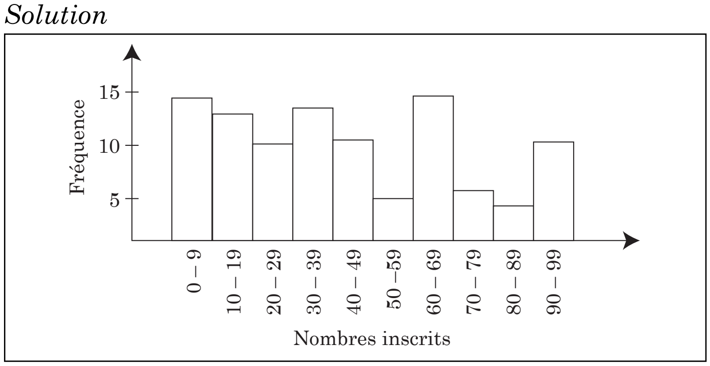
Distributions anormales

Les distributions normales obtenues lorsque des mesures sont effectuées ne sont pas le fruit de la magie. Il existe toujours une raison. Lorsque ces raisons ne s'appliquent pas, nous obtenons d'autres types de distribution.

Exemple

Si chaque élève d'une classe inscrit un nombre de 0 à 99 sur un bout de papier et si les nombres sont représentés dans un graphique (probablement en groupes de 10, soit de 0 à 9, de 10 à 19, de 20 à 29, et ainsi de suite), le graphique sera probablement de forme rectangulaire.

Solution

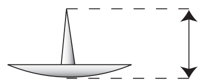


STRATÉGIES D'ÉVALUATION

NOTES

Problèmes

1 a) Laissez chaque élève mesurer les hauteurs de dix punaises à 0,1 mm près à l'aide d'un pied à coulisse ou d'un micromètre.



b) Combinez les résultats de tous les élèves. Sélectionnez des intervalles et établissez la distribution des fréquences.

c) Construisez le graphique des résultats et donnez des commentaires sur la forme du graphique.

2. a) Recueillez des données sur les grandeurs des hommes et des femmes. Établissez la distribution des fréquences et construisez le graphique des résultats obtenus.

b) Séparez les scores obtenus pour les hommes et les femmes. Établissez des distributions de fréquences distinctes et construisez le graphique des résultats.

c) Donnez des commentaires sur la forme des trois graphiques.

3. Créez 500 nombres aléatoires de deux chiffres en utilisant la méthode ci-dessous.

Ouvrez l'annuaire téléphonique à une page quelconque. Dans chaque numéro de téléphone paraissant l'un à la suite de l'autre, choisissez les deux chiffres situés à la position suivante : 487-0***93***5. Dans cet exemple, les chiffres choisis seraient 93.

Remplissez le tableau ci-dessous et tracez le graphique des résultats. Donnez des commentaires sur les résultats.

Nombres inscrits	**Fréquence**	**Nombres inscrits**	**Fréquence**
00 – 09		50 – 59	
10 – 19		60 – 69	
20 – 29		70 – 79	
30 – 39		80 – 89	
40 – 49		90 – 99	

Commentaires possibles sur les résultats :

a) Le graphique a une forme environ normale parce que les erreurs de mesurage sont habituellement réparties de manière égale au-dessus et en dessous des mesures réelles.

b) La distribution complète sera bimodale. Chaque distribution sera environ normale.

c) Le graphique est donc habituellement rectangulaire.

RÉSULTATS D’APPRENTISSAGE SPÉCIFIQUES

E-3 Utiliser la distribution normale et l'approximation normale de la distribution binomiale pour résoudre des problèmes reliés à des intervalles de confiance d'échantillons de grande taille.
– suite

STRATÉGIES PÉDAGOGIQUES

- **Examiner les écarts réduits ou cotes *z* dans les distributions normales.**

  Examinons maintenant les cotes *z* dans les distributions normales. Vous trouverez ci-dessous la reproduction d'une partie de deux des lignes d'un graphique d’une cote *z*. (Voir l'annexe E-3, page E-44, pour obtenir un graphique de cotes *z*.)

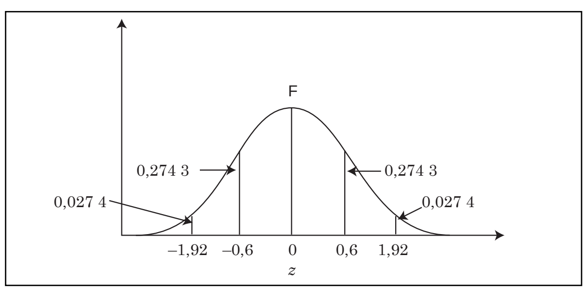
z	0,00	0,01	0,02
0,6	0,225 7	0,229 1	0,232 4
1,9	0,471 3	0,471 9	0,472 6

La première entrée, 0,225 7, nous indique que, dans une distribution normale, 0,225 7 ou 22,57 % des scores se situent entre $z = 0$ et $z = 0,6$.

Cela signifie aussi que 0,225 7 ou 22,57 % des scores se situent entre $z = 0$ et $z = –0,6$. Nous pouvons calculer que 0,5 – 0,225 7 = 0,274 3 ou 27,43 % des scores se situent à la limite et ont des cotes *z* supérieures à 0,6. De même, 27,43 % des scores se situent à la limite gauche de $z = –0,6$ et ont des cotes *z* inférieures à –0,6.

L'entrée encerclée, 0,472 6, nous indique que 47,26 % des scores d'une distribution normale ont des cotes *z* situées entre $z = 0$ et $z = 1,92$, de sorte que 50 % = 47,26 % = 2,74 % des scores se situent à la limite droite de $z = 1,92$. Si nous choisissions un score au hasard, nous aurions 2,74 % des chances que la cote *z* soit supérieure à 1,92.

Veuillez prendre note que les calculatrices graphiques comportent des fonctions qui vous permettent de calculer les probabilités à partir des cotes *z* dans une distribution normale.



– suite

STRATÉGIES D'ÉVALUATION

Problèmes

1. Dans une distribution normale, déterminez le pourcentage des scores qui se situent :
   a) au-dessus de la moyenne à moins de 1,5 écart type;
   b) au-dessus de la moyenne à plus de 1,5 écart type (secteur unilatéral);
   c) sous la moyenne à plus de 1,5 écart type;
   d) à plus de 1,5 écart type de la moyenne (secteur bilatéral);
   e) à l'intérieur de 1,5 écart type de la moyenne.
2. a) Le secteur unilatéral se situant sous la moyenne contient 1,79 % des scores dans une distribution normale. Déterminez la cote z qui délimite ce secteur.
   b) Le secteur bilatéral contient 15 % des scores dans une distribution normale. Déterminez les cotes z qui délimitent les deux parties de ce secteur.

Solutions

1. a) z = 1,5; 43,32 % des scores
   b) 50 % – 43,32 % = 6,68 % des scores
   c) symétrie; donc, 6,68 % des scores
   d) total des deux parties = 2(6,68 %) = 13,36 % des scores
   e) total entre z = –1,5 et z = 1,5 = 2(43,32 %) = 86,64 %
2. a) 50 % – 1,79 % = 48,21 % des scores se situent à z écart type ou moins de la moyenne. Déterminez où se situe l'entrée 0,482 1, soit à une cote z de 2,1. Toutefois, le secteur en question se situe sous la moyenne. Donc, la cote z exacte est de –2,1.
   b) Il s'agit d'un secteur bilatéral puisque 7,5 % des scores se situent à chaque limite. Déterminez où se situe l'entrée 0,425, soit à une cote z de 1,44. Il s'agit de la limite droite, tandis que la cote z de –1,44 délimite la limite gauche.

NOTES

La lettre grecque mu (μ) symbolise la moyenne d'une population.

RÉSULTATS D'APPRENTISSAGE SPÉCIFIQUES	STRATÉGIES PÉDAGOGIQUES

E-3 Utiliser la distribution normale et l'approximation normale de la distribution binomiale pour résoudre des problèmes reliés à des intervalles de confiance d'échantillons de grande taille.
– suite

- **Examiner les cotes *z* dans les distributions normales. (suite)**

  Exemple 1

  Utilisez un tableau de cotes z ou une calculatrice graphique pour déterminer le pourcentage des scores se situant sous la courbe de chacune des valeurs ci-dessous.

  a) $0 < z < 1$ b) $0 < z < 2$

  c) $-2 < z < 1$ d) $z < 1{,}5$

  Solution

  a) 34,13 % b) 47,72 %

  c) 81,85 % d) 93,32 %

  Exemple 2

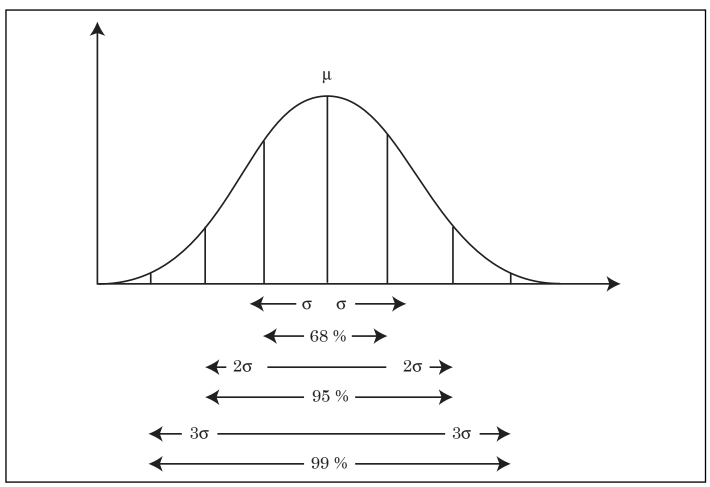
  Déterminez la cote z qui délimite la tranche supérieure de 10 % des scores.

  Solution

  $z = 1{,}28$

- **Reconnaître des faits importants.**

  Utilisez un tableau de cotes z pour confirmer que dans une distribution normale, 68 % des scores se situent à 1σ ou moins de μ, que 95 % des scores se situent à 2σ ou moins de μ et que près de 99 % des scores de situent à 3σ ou moins de μ.



  Ces faits sont importants et vous devez ne pas les oublier puisqu'ils sont utilisés très souvent.

– suite

Stratégies d'évaluation

Notes

RÉSULTATS D'APPRENTISSAGE SPÉCIFIQUES	STRATÉGIES PÉDAGOGIQUES

E-3 Utiliser la distribution normale et l'approximation normale de la distribution binomiale pour résoudre des problèmes reliés à des intervalles de confiance d'échantillons de grande taille.
– suite

- **Reconnaître des faits importants. (suite)**

Exemple 1

Supposons que les grandeurs des femmes en Amérique du Nord sont distribuées normalement. Pour être admises dans le « Club des grandes dames », les femmes doivent avoir une grandeur minimale de 5 pi 10 po. En Amérique du Nord, les femmes ont une grandeur moyenne de 5 pi 5,5 po, et l'écart type est de 2,5 po. Quel pourcentage de femmes sera admis dans ce club?

Solution

Calculs en pouces :

$$z = \frac{(5 \text{ pi } 10 \text{ po } - 5 \text{ pi } 5{,}5 \text{ po})}{2{,}5 \text{ po}} = 1{,}80$$

Le secteur situé sous la courbe normale des grandeurs supérieures à 5 pi 10 po = 0,5 – 0,464 1 = 0,035 9. Donc, 3,59 % des femmes peuvent être admises dans ce club.

Calculs en pieds :

$$z = \frac{(5{,}833\,3 - 5{,}453\,3)}{0{,}208\,3} = 1{,}80$$

Le secteur situé sous la courbe normale des grandeurs supérieures à 5 pi 10 po = 0,5 – 0,464 1 = 0,035 9. Donc, 3,59 % des femmes peuvent être admises dans ce club.

Exemple 2

Vous trouverez ci-dessous les grandeurs des pantalons (tours de taille en pouces) vendus en une seule matinée à la Mercerie Mercier :

34, 38, 36, 38, 38, 34, 36, 36, 34, 34, 36, 40
32, 36, 36, 34, 36, 34, 40, 38 ,32, 36, 34, 34

La distribution de ces grandeurs est habituellement normale. Vous devez effectuer le calculs ci-dessous et indiquer si les résultats obtenus confirment l'énoncé.

a) Utilisez une calculatrice graphique ou un tableur pour déterminer la moyenne et l'écart type de cet ensemble de scores.

b) Quelle grandeur se situe à un écart type au-dessus de la moyenne?

Quelle grandeur se situe à un écart type sous la moyenne?

c) Déterminez le nombre de scores qui se situent entre les deux valeurs de (b). Exprimez votre réponse sous forme de pourcentage par rapport au nombre total de scores.

– suite

STRATÉGIES D'ÉVALUATION

NOTES

Problème

Les registres à propos d'un avion de type airbus indiquent que les vols effectués entre Frankfurt et Montréal arrivent, en moyenne, 5,4 minutes en retard et que l'écart type est de 1,4 minute. Utilisez le tableau de la distribution normale pour estimer le pourcentage des vols qui arrivent :

a) plus de 8,2 minutes en retard;

b) moins de 4,00 minutes en retard;

c) entre 2,6 et 6,8 minutes en retard.

Solution

a) Cote z pour 8,20 min = $\dfrac{(8{,}20 - 5{,}40)}{1{,}4} = \dfrac{2{,}8}{1{,}4} = 2$

Secteur à la droite de $z = 2$ dans un tableau de distribution normale = 0,5 – 0,477 2 = 0,022 8

Pourcentage requis de 2,28 %

b) Cote z pour 4,00 min = $\dfrac{(4 - 5{,}4)}{1{,}4} = \dfrac{-1{,}4}{1{,}4} = -1$

Secteur à la gauche de $z = -1$ = 0,5 – 0,341 3 = 0,158 7

Pourcentage requis de 1,59 %

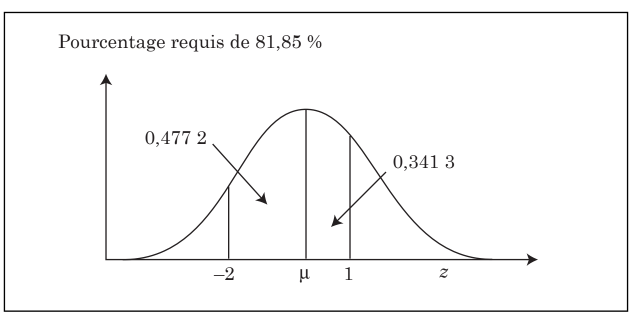
c) Cote z pour 2,6 min = $\dfrac{(2{,}6 - 5{,}4)}{1{,}4} = \dfrac{-2{,}8}{1{,}4} = -2$

Cote z pour 6,8 min = $\dfrac{(6{,}8 - 5{,}4)}{1{,}4} = \dfrac{1{,}4}{1{,}4} = 1$

Le secteur entre $z = -2$ et $z = 1$ correspond au secteur entre $z = -2$ et $z = 0$, ***plus*** le secteur entre $z = 0$ et $z = 1$.

Secteur = 0,477 2 + 0,341 3 = 0,818 5

Pourcentage requis de 81,85 %



RÉSULTATS D'APPRENTISSAGE SPÉCIFIQUES

E-3 Utiliser la distribution normale et l'approximation normale de la distribution binomiale pour résoudre des problèmes reliés à des intervalles de confiance d'échantillons de grande taille.
– suite

STRATÉGIES PÉDAGOGIQUES

- **Reconnaître des faits importants. (suite)**

  Exemple 2 — suite

  Solution

  a) Moyenne = 35,67.

  b) Moyenne + σ = 35,67 + 2,18 = 37,85.

  Moyenne – σ = 35,67 – 2,18 = 33,49.

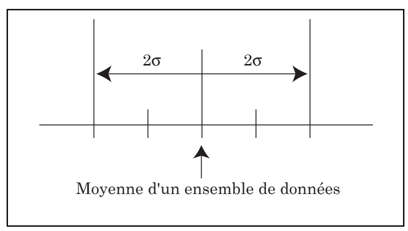
  c) Nombre de scores situés entre 33,44 et 37,80 = 16

  Pourcentage du total = 67 %

  Ces résultats se rapprochent de ceux d'une distribution normale. Nous pouvons confirmer que cet échantillon provient probablement d'une distribution normale.

- **Utiliser la distribution normale pour résoudre des problèmes reliés à des intervalles de confiance.**

  L'***intervalle de confiance*** pour la moyenne de la ***population*** peut être déterminé à l'aide des valeurs de la moyenne et de l'écart type de la population. Examinez l'intervalle illustré dans le diagramme ci-dessous.



Nous pouvons prouver que si nous étudions un très grand nombre d'ensembles de données de la même taille provenant de la population, les moyennes de 95 % de ces ensembles de données se situeraient dans l'intervalle (moyenne des données – 1,96σ, moyenne des données + 1,96σ).

Cet intervalle correspond à un ***intervalle de confiance de 95 % pour la moyenne de la population.***

Puisque 1,96 se rapproche de 2, nous utilisons souvent une approximation pour la règle ci-dessus en indiquant que l'intervalle de confiance de 95 % pour la moyenne de la population correspond à la moyenne des données de – 2σ jusqu'a la moyenne des données de + 2σ. Toutefois, les calculatrices graphiques utilisent habituellement 1,96σ et les élèves peuvent remarquer une légère différence entre une estimation calculée à l'aide de 2σ et le résultat de la calculatrice graphique.

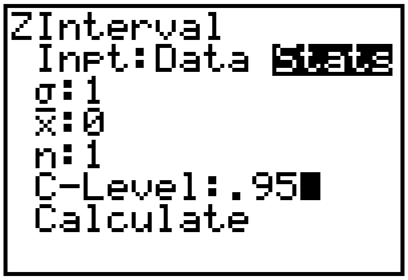
– suite

STRATÉGIES D'ÉVALUATION

NOTES

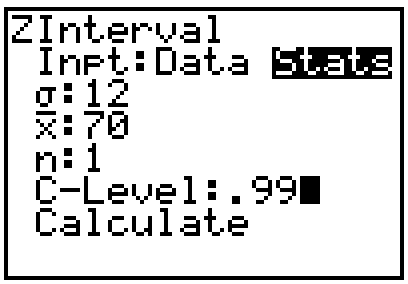
Problèmes

1. a) Interprétez l'écran ci-dessous de la calculatrice TI-83 :



b) Quel résultat obtenez-vous lorsque vous appuyez sur Calculate?

2. a) Interprétez l'écran ci-dessous de la calculatrice TI-83 :



b) Quel résultat obtenez-vous lorsque vous appuyez sur Calculate?

Solutions

1. a) Cet écran illustre les données que vous devez enregistrer dans la calculatrice pour obtenir un intervalle de confiance de 95 % pour la moyenne de la population lorsque vous savez ou que vous estimez que la valeur de la moyenne est de 0 et que l'écart type est de 1. Vous devez appuyer sur STAT, sélectionner TESTS, puis 7:ZInterval. Vous devez ensuite sélectionner STATS pour avoir accès à cet écran.

b) (–1,96 , 1,96)

2. a) Cet écran illustre les données que vous devez enregistrer dans la calculatrice pour obtenir un intervalle de confiance de 99 % pour la moyenne de la population lorsque vous savez ou que vous estimez que la valeur de la moyenne est de 70 et que l'écart type est de 12. Vous devez appuyer sur STAT, sélectionner TESTS, puis 7:ZInterval. Vous devez ensuite sélectionner STATS pour avoir accès à cet écran.

b) (39, 101)

RÉSULTATS D'APPRENTISSAGE SPÉCIFIQUES

E-3 Utiliser la distribution normale et l'approximation normale de la distribution binomiale pour résoudre des problèmes reliés à des intervalles de confiance d'échantillons de grande taille.
– suite

STRATÉGIES PÉDAGOGIQUES

- **Utiliser la distribution normale pour résoudre des problèmes reliés à des intervalles de confiance. (suite)**

Lorsque l'intervalle de confiance est de 95 %, nous pouvons dire qu'il existe 19 chances sur 20 que la moyenne se situe dans cet intervalle.

Si la moyenne d'une population = 101 et que l'écart type = σ = 10, les moyennes d'environ 95 % d'ensembles supplémentaires de données de la même taille se situeraient dans les limites approximatives $101 \pm 2(\sigma) = 101 \pm 20$ ou (81, 121). Vous remarquerez que l'intervalle de confiance de 95 % fourni par la calculatrice en utilisant $101 \pm 1{,}96\sigma$ serait (81,4, 120,6)

Nous n'avons aucune information à propos de l'endroit où se situe la moyenne de la population dans cet intervalle.

De même, un intervalle de confiance de 99 % pour la moyenne de la population est fourni en utilisant la moyenne et l'écart type (σ) d'un ensemble de données et en construisant l'intervalle (la moyenne de l'ensemble de données – $2{,}57\sigma$, la moyenne de l'ensemble de données + $2{,}57\sigma$). Les limites de l'intervalle de confiance de 99 % de la moyenne pour la moyenne de la population sont $101 \pm 2{,}575(10)$ ce qui produit l'intervalle (75, 127)

Une estimation approximative de l'intervalle dont les limites sont $101 \pm 3(10)$ est souvent utilisée, ce qui produit l'intervalle (71, 131). Cet intervalle contient plus que 99 % des scores.

Exemple 1

La distribution du temps requis pour se rendre de Angleton à Bertrand est normale, la moyenne est de 1 h 32 min et l'écart type est de 14 min. Vous devez établir un intervalle de confiance symétrique de 95 % pour la moyenne du temps requis pour ce voyage.

Solution

Pour résoudre ce problème à l'aide de la calculatrice graphique TI-83, vous devez suivre les étapes suivantes :

- Appuyez sur STAT et sélectionnez TESTS au haut de l'écran, puis sélectionnez 7:ZInterval.
- Choisissez STATS en mettant cette option en surbrillance et en appuyant sur ENTER.
- Entrez les données pour que l'écran ressemble à l'écran ci-dessous. (Vous remarquerez que la valeur $n = 1$ fait en sorte que la calculatrice traite l'information en tant que population et non en tant qu'échantillon.)

STRATÉGIES D'ÉVALUATION

NOTES

RÉSULTATS D'APPRENTISSAGE SPÉCIFIQUES

E-3 Utiliser la distribution normale et l'approximation normale de la distribution binomiale pour résoudre des problèmes reliés à des intervalles de confiance d'échantillons de grande taille.
– suite

STRATÉGIES PÉDAGOGIQUES

- **Utiliser la distribution normale pour résoudre des problèmes reliés à des intervalles de confiance. (suite)**
  - Enregistrez l'écart type donné pour *s*.

```
ZInterval
 Inpt:Data Stats
 σ:14
 x̄:92
 n:1
 C-Level:.95
 Calculate
```

  - Mettez en surbrillance le mot Calculate et appuyez sur ENTER.
  - L'écran ci-dessous devrait s'afficher et vous indiquer l'intervalle de confiance de 64,561 à 119,44.

```
ZInterval
 (64.561,119.44)
 x̄=92
 n=1
```

La réponse pourrait être présentée comme suit :
64,56 minutes $< \bar{x} <$ 119,44 minutes.

Ce résultat est interprété comme suit : Si nous devions choisir des échantillons d'une population donnée (pour mesurer le temps requis) et déterminer les moyennes correspondantes, 95 % de ces moyennes se situeraient dans l'intervalle.

Exemple 2

Un élève enregistre 30 fois le temps qu'il met pour se rendre à l'école. Les temps enregistrés sont les suivants (en minutes) :

21, 18, 18, 16, 15, 21, 17, 15, 20, 14, 17, 18, 20, 19, 16
19, 18, 16, 21, 17, 15, 20, 14, 18, 17, 15, 16, 21, 18, 20

a) Utilisez la technologie pour déterminer la moyenne et l'écart type de cet ensemble de données (qui doit être traité en tant que population).

b) Si nous supposons que les scores obtenus proviennent d'une distribution normale, utilisez ces résultats pour produire un intervalle de confiance dans lequel se situent les moyennes d'ensembles supplémentaires de données contenant chacun 30 temps, 19 fois sur 20. Déterminez un intervalle de confiance de 95 % pour la moyenne de cette distribution.

— suite

STRATÉGIES D'ÉVALUATION

NOTES

Problèmes

1. Une usine de matériel électronique produit des cartes de circuits imprimés dont les durées de vie sont normalement distribuées et dont l'écart-type est de 80 heures. Deux cent cinquante cartes de circuits imprimés sont choisies au hasard et vérifiées, et on constate que la durée de vie moyenne est de 360 heures.
   a) Déterminez un intervalle de confiance de 95 % pour la durée de vie des cartes de circuits imprimés.
   b) Déterminez un intervalle de confiance de 99 % pour la durée de vie des cartes de circuits imprimés.

2. Le revenu hebdomadaire moyen des élèves de secondaire 4 qui travaillent à temps partiel doit être estimé. Un échantillon de 35 élèves est questionné, et on constate que le revenu moyen est de 75 $ par semaine et que l'écart-type est de 15 $. Si la distribution est normale, déterminez un intervalle de confiance de 95 % pour le revenu hebdomadaire moyen des élèves de secondaire 4.

Solutions

1 a) 203,2 < $\bar{x}$ < 516,8 heures

Ainsi, la durée de vie d'une carte de circuits imprimés se situera dans l'intervalle indiqué dans 95 % des cas.

b) 153,93 < $\bar{x}$ < 566,07 heures

2. 45,60 $ < $\bar{x}$ < 104,40 $

RÉSULTATS D'APPRENTISSAGE SPÉCIFIQUES

E-3 Utiliser la distribution normale et l'approximation normale de la distribution binomiale pour résoudre des problèmes reliés à des intervalles de confiance d'échantillons de grande taille.
– suite

STRATÉGIES PÉDAGOGIQUES

- **Utiliser la distribution normale pour résoudre des problèmes reliés à des intervalles de confiance. (suite)**

Solution

a) $\bar{x} \approx 17{,}67$ minutes
$\sigma = 2{,}15$ minutes

b) En utilisant la méthode expliquée plus haut ou un tableur, vous déterminez que l'intervalle requis est de 13,46 minutes à 21,88 minutes.

L'intervalle de confiance de 95 % pour la moyenne est de (13,46, 21,88). Cet intervalle peut aussi être présenté comme suit : $13{,}46 < \bar{x} < 21{,}88$ minutes.

- **Utiliser l'approximation normale de la distribution binomiale pour résoudre des problèmes reliés à des intervalles de confiance.**

L'étude théorique des distributions binomiales n'est pas requise dans ce cours. Les élèves devraient toutefois pouvoir reconnaître les situations binomiales et mettre en pratique les formules requises pour obtenir la moyenne et l'écart type afin de prédire les résultats des situations binomiales. Les calculs à l'aide des formules ne font pas partie de ce cours.

Exemple 1

Une compagnie de vêtements sait que 80 % de tous les clients utilisent une carte de crédit pour faire leurs achats. La compagnie désire estimer le nombre de clients sur les 400 prochains clients qui utiliseront une carte de crédit. Déterminez l'intervalle de confiance de 95 % pour le nombre de clients sur les 400 prochains clients qui utiliseront une carte de crédit.

Solution

Il s'agit d'une situation binomiale puisque chaque client peut utiliser ou non une carte de crédit. Il n'existe que deux possibilités pour chaque personne.

La probabilité qu'une personne utilise une carte de crédit = $p = 0{,}8$.
La probabilité qu'une personne n'utilise pas une carte de crédit = $q = 1 - 0{,}8 = 0{,}2$.
Le nombre de clients = $n = 400$.

Le nombre moyen de clients qui utilisent une carte de crédit sur 400 clients est obtenu par la formule suivante : $\bar{x} = np$.

Il s'agit du nombre de clients qui devraient utiliser une carte de crédit.

L'écart type est obtenu par la formule suivante : $\sigma = \sqrt{npq}$.

Moyenne = $\bar{x} = 400(0{,}8) = 320$.

Écart type = $\sigma = \sqrt{400(0{,}8)(0{,}2)} = 8$.

— suite

STRATÉGIES D'ÉVALUATION

NOTES

Problèmes

1. Fred décide de faire un voyage de chasse. La probabilité qu'il atteigne le canard sur lequel il tire est de 0,4. Il pense faire 30 tirs. Déterminez l'intervalle de confiance pour le nombre de canards qu'il devrait atteindre et qui serait valide 19 fois sur 20.

2. Si 15 % de toutes les puces informatiques fabriquées sont défectueuses, déterminez un intervalle de confiance de 90 % pour le nombre moyen de puces défectueuses en paquets de 100 puces.

Solutions

1. Nombre de tirs = n = 30.

   Probabilité d'atteindre un canard = p = 0,4.

   Probabilité de ne pas atteindre un canard = q = 0,6.

   Moyenne = $\bar{x}$ = 30(0,4) = 12.

   Écart type = $\sigma = \sqrt{30(0{,}4)(0{,}6)} \approx 2{,}68$.

   Puisque nous devons obtenir un intervalle valide 19 fois sur 20, nous devons utiliser un intervalle de confiance de 95 % pour la moyenne. À l'aide de la technologie, nous obtenons (6,74, 17,25).

2. Moyenne = $\bar{x}$ = 100(0,15) = 15.

   Écart type = $\sigma = \sqrt{100(0{,}15)(0{,}85)} \approx 3{,}57$.

   Intervalle de confiance de 90 % pour la moyenne = (9,13, 20,87).

   Vous pourriez vous attendre à obtenir de 9 à 21 puces informatiques défectueuses dans chaque paquet de 100 puces informatiques.

RÉSULTATS D'APPRENTISSAGE SPÉCIFIQUES

E-3 Utiliser la distribution normale et l'approximation normale de la distribution binomiale pour résoudre des problèmes reliés à des intervalles de confiance d'échantillons de grande taille.
– suite

STRATÉGIES PÉDAGOGIQUES

- **Utiliser l'approximation normale de la distribution binomiale pour résoudre des problèmes reliés à des intervalles de confiance. (suite)**

Pour obtenir un intervalle de confiance pour les moyennes de groupes de 400 clients utilisant des cartes de crédit, il faut utiliser une distribution normale ayant une moyenne de 320 et un écart type de 8 et utiliser la technologie comme dans les cas précédents.

Cette procédure produit l'intervalle (304, 336).

Ainsi, sur 400 clients, le nombre de clients utilisant des cartes de crédit se situera entre 304 personnes et 336 personnes, 19 fois sur 20.

Exemple 2

Le directeur médical d'un camp de vacances établit qu'environ 3 % des enfants sont gravement malades pendant la saison. Si 600 enfants se sont inscrits au camp de vacances cet été :

a) déterminez, à un niveau de confiance de 95 %, combien d'enfants seront probablement gravement malades pendant la saison;

b) déterminez, à un niveau de confiance de 99 %, combien d'enfants seront probablement gravement malades pendant la saison.

Solution

Il s'agit d'une situation binomiale typique. Les deux résultats possibles sont « malades » et « non malades ».

a) $\bar{x} = np$

$\bar{x} = 600(3/100)$

$\bar{x} = 18$ enfants

$$\sigma = \sqrt{npq}$$

$$\sigma = \sqrt{600\left(\frac{3}{100}\right)\left(\frac{97}{100}\right)} = \sqrt{600(0{,}03)(0{,}97)}$$

$\sigma = 4{,}18$

Le niveau de confiance de 95 % (utilisez la calculatrice comme dans la question précédente) est de 9,81 à 26,19. Cela signifie que 19 fois sur 20, le nombre d'enfants malades sera supérieur à 9,81 mais inférieur à 26,19.

b) Le niveau de confiance de 99 % (utilisez la calculatrice comme dans la question précédente) est de 7,23 à 28,77 et peut être exprimé de la manière suivante :
$7{,}23 < \bar{x} < 28{,}77$.

Pour de plus amples renseignements, veuillez consulter l'annexe E-5 à la page E-46.

STRATÉGIES D'ÉVALUATION

NOTES

Problème

Les policiers patrouillant une autoroute déterminent qu'à un endroit en particulier, la probabilité qu'un automobiliste excède la limite de vitesse est de 60 %. Si 3 500 automobiles empruntent cette route chaque jour, déterminez l'intervalle de confiance de 90 % pour la moyenne du nombre d'automobilistes qui excèdent la limite de vitesse chaque jour.

Solution

$\bar{x} = np = (3\ 500)(0{,}60) = 2\ 100$

$\sigma = \sqrt{(3\ 500)(0{,}60)(0{,}40)} = 28{,}98$

Intervalle de confiance de 90 % : $2\ 052{,}3 < \bar{x} < 2\ 147{,}7$ ou

$2\ 052 < \bar{x} < 2\ 148$ automobiles

Ce résultat peut aussi être exprimé de la manière suivante : à chaque jour, le nombre d'automobiles excédant la limite de vitesse se situe entre 2 052 et 2 148, neuf fois sur dix.

Annexe E-1

Définition des termes fréquemment utilisés en statistique

Courbe normale : Courbe en forme de cloche (élévation au centre et limites symétriques). Il s'agit du contour de la distribution des fréquences d'une distribution binomiale ayant une grande quantité de données et dont la probabilité de chaque événement est de 0,5.

Distribution : Voir « Distribution des fréquences ».

Distribution binomiale : Données créées par des événements qui ne comportent que deux résultats possibles. Lorsque la probabilité de chaque résultat est de 0,5, on parle de courbe normale.

Distribution des fréquences : Lorsque les données sont recueillies en classes, le nombre de scores dans chaque classe se nomme la distribution des fréquences.

Distribution normale : Voir « Courbe normale ».

Données : Scores ou mesures relatives à une population.

Données continues : Scores ou mesures qui pourraient, du moins en théorie, correspondre à une valeur quelconque. Les scores obtenus aux tests correspondent habituellement à des données continues, même s'ils sont tous exprimés en nombres entiers. Un score de 65,5 pourrait être obtenu et avoir une signification.

Données discrètes : Données pour lesquelles les scores situés entre les scores obtenus n'ont aucune signification. Si on établissait une liste des températures maximales quotidiennes, les nombres entre les points de données n'auraient aucune signification.

Écart type : Mesure de la répartition des scores d'une population ou d'un échantillon par rapport à la moyenne.

Échantillon : Tout ensemble de données choisi à partir d'une population.

Échantillon aléatoire : Échantillon d'une population choisi de sorte que (a) chaque élément de données de la population a une chance égale d'être choisi et (b) qu'aucune sélection n'ait un effet sur toute autre sélection.

Erreur quadratique moyenne **($\sigma_{\overline{X}}$) :** Mesure de la répartition des moyennes d'échantillons d'une population.

Intervalle de confiance : Intervalle de confiance pour la moyenne d'une population. Cet intervalle est établi à partir d'un échantillon d'une population, de sa moyenne et de son écart type, et il correspond à l'intervalle entourant la moyenne d'un échantillon dans lequel se situe μ, la moyenne de la population. Un intervalle de confiance de 95 % comprendra μ dans 95 % des cas (19 fois sur 20).

Médiane : Lorsqu'un ensemble de scores est présenté dans un ordre précis, il s'agit du score du centre.

Mode : Lorsque les scores sont recueillis en classes, le mode correspond au score du centre de la classe contenant le nombre le plus élevé de scores.

Moyenne : Somme d'un ensemble de scores divisée par le nombre de scores.

Population : Ensemble de tous les scores ou de toutes les mesures qui sont étudiés.

s : Voir « Écart type ». Symbole qui représente une estimation de l'écart type d'une population calculée à partir des données d'un échantillon de cette population.

Sigma **(σ) :** Symbole qui représente l'écart type d'une population.

Erreur quadratique moyenne d'une proportion : Mesure de la répartition des proportions des échantillons d'une population.

Cote z ou ***Écart réduit :*** Distance entre le score d'une distribution et la moyenne, mesurée en écarts types.

Symboles utilisés dans l'unité sur les statistiques

x correspond à un élément d'une population.

$\bar{x}$ correspond à la moyenne d'un ensemble de données.

n correspond au nombre de scores d'un ensemble de données.

μ correspond à la moyenne d'une population.

Σ est un symbole de sommation. Σx correspond à la somme de tous les éléments indiqués en tant que x.

σ ou « sigma » correspond à l'écart type d'un ensemble de données.

La formule utilisée pour calculer σ est la suivante : $\sigma = \sqrt{\dfrac{\Sigma\left(x-\bar{x}\right)^2}{n}}$

Nota : La formule $s = \sqrt{\dfrac{\sum\left(x-\bar{x}\right)^2}{n-1}}$ est utilisée lorsque nous voulons estimer l'écart type d'une population à l'aide d'un échantillon de cette population.

Annexe E-2

Opérations avec la calculatrice graphique TI-83

a) Enregistrement de données dans la liste de la TI-83
- Appuyez sur le bouton STAT de la calculatrice et vérifiez si l'option EDIT au haut de l'écran est en surbrillance. (Utilisez les touches fléchées vers la gauche et vers la droite pour déplacer la barre de surbrillance.)
- À partir de la liste sur l'écran, sélectionnez l'option 1:Edit, en tapant 1 ou en mettant cette option en surbrillance en utilisant les touches fléchées vers le haut et vers le bas et en appuyant sur ENTER.
- Vous devez ensuite placer le curseur à la colonne intitulée L1. Si cette liste n'est pas vide, appuyez sur la touche fléchée vers le haut pour mettre L1 en surbrillance, sur CLEAR, puis sur ENTER.
- Enregistrez les données de la liste L1 en appuyant sur ENTER après chaque enregistrement.
- Une fois tous les renseignements enregistrés, la calculatrice peut déterminer la moyenne et la médiane.
- Appuyez sur STAT et sélectionnez l'option CALC au haut de l'écran.
- Sélectionnez l'option 1:1-Var Stats. La ligne 1-Var Stats s'affichera à l'écran suivie d'un curseur clignotant. La calculatrice attend que vous lui disiez où sont sauvegardées les données que vous désirez utiliser.
- Appuyez sur 2nd et sur 1. L1 s'affichera à la fin de la ligne. (N'oubliez pas que vous avez enregistré vos données dans la liste L1.)
- Appuyez sur ENTER et attendez.

b) Enregistrement de données dans la liste de la TI-83 à partir d'une liste des fréquences
- Appuyez sur le bouton STAT de la calculatrice et vérifiez si l'option EDIT au haut de l'écran est en surbrillance. (Utilisez les touches fléchées vers la gauche et vers la droite pour déplacer la barre de surbrillance.)
- À partir de la liste sur l'écran, sélectionnez l'option1:Edit, en tapant 1 ou en mettant cette option en surbrillance en utilisant les touches fléchées vers le haut et vers le bas et en appuyant sur ENTER.
- Vous devez ensuite placer le curseur à la colonne intitulée L1. Si cette liste n'est pas vide, appuyez sur la touche fléchée vers le haut pour mettre L1 en surbrillance, sur CLEAR, puis sur ENTER.
- Enregistrez les scores de la première colonne L1 en appuyant sur ENTER après chaque enregistrement.
- Placez le curseur dans la colonne intitulée L2. Si cette liste n'est pas vide, appuyez sur la touche fléchée vers le haut pour mettre L2 en surbrillance, sur CLEAR, puis sur ENTER.
- Enregistrez les nombres de la colonne des fréquences dans la liste L2 en appuyant sur ENTER après chaque enregistrement. Ces nombres doivent être alignés avec les nombres correspondants de L1.
- Une fois tous les renseignements enregistrés, la calculatrice peut déterminer la moyenne et la médiane.

Calcul de la moyenne et de la médiane
- Appuyez sur le bouton STAT et sélectionnez l'option CALC au haut de l'écran.
- Sélectionnez l'option 1.1-Var Stats. Une ligne affichant 1-Var Stats s'affichera à l'écran et sera suivie par un curseur clignotant. La calculatrice attend que vous lui disiez où sont sauvegardées les données que vous désirez utiliser.
- Appuyez sur 2nd et sur 1. Appuyez sur « , », sur 2nd et sur 2. La ligne 1-Var Stats L1,L2 devrait s'afficher à l'écran.
- Appuyez sur ENTER pour obtenir l'information.

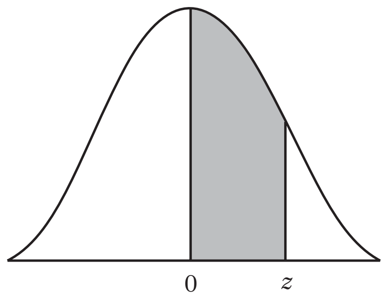
c) Comment tracer un histogramme représentant les données d'une distribution des fréquences.

- Enregistrez les données dans L1 et L2 de la manière indiquée à l'annexe E-2(b).
- Pour sélectionner STAT PLOT, appuyez sur 2nd et sur Y=.
- Sélectionnez 1:Plot1, et appuyez sur ENTER.
- Sélectionnez On et ENTER.
- Utilisez les touches fléchées pour sélectionner le troisième type, celui qui ressemble à un histogramme, et appuyez sur ENTER.
- Utilisez les touches fléchées pour mettre Xlist en surbrillance et sélectionnez L1 en appuyant sur 2nd et sur 1.
- Utilisez les touches fléchées pour mettre Freq: en surbrillance et sélectionnez L2 en appuyant sur 2nd et sur 2.
- Maintenant, enregistrez les valeurs ci-dessous.
- $X_{min} = 0$
- $X_{max} = 10$
- $X_{scl} = 1$
- $Y_{min} = -10$
- $Y_{max} = 180$
- $Y_{scl} = 20$

***Nota* :** Si vous utilisez ZOOM et 9:ZoomStat pour enregistrer les valeurs de manière automatique, la valeur 0,7 sera automatiquement enregistrée pour X_{scl}, et le graphique ne sera pas exact. Modifiez la valeur de X_{scl} à 1.

Annexe E-3

Distribution normale centrée réduite



Remarques :

1. Lorsque la valeur de *z* est supérieure à 3,09, utilisez 0,499 9 pour l'aire.
2. Utilisez ces valeurs communes qui résultent de l'interpolation :

Cotes z	Aire
1,645	0,450 0
2,575	0,495 0

Tableau des cotes z

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,000 0	0,004 0	0,008 0	0,012 0	0,016 0	0,019 9	0,023 9	0,027 9	0,031 9	0,035 9
0,1	0,039 8	0,043 8	0,047 8	0,051 7	0,055 7	0,059 6	0,063 6	0,067 5	0,071 4	0,075 3
0,2	0,079 3	0,083 2	0,087 1	0,091 0	0,094 8	0,098 7	0,102 6	0,106 4	0,110 3	0,114 1
0,3	0,117 9	0,121 7	0,125 5	0,129 3	0,133 1	0,136 8	0,140 6	0,144 3	0,148 0	0,151 7
0,4	0,155 4	0,159 1	0,162 8	0,166 4	0,170 0	0,173 6	0,177 2	0,180 8	0,184 4	0,187 9
0,5	0,191 5	0,195 0	0,198 5	0,201 9	0,205 4	0,208 8	0,212 3	0,215 7	0,219 0	0,222 4
0,6	0,225 7	0,229 1	0,232 4	0,235 7	0,238 9	0,242 2	0,245 4	0,248 6	0,251 7	0,254 9
0,7	0,258 0	0,261 1	0,264 2	0,267 3	0,270 4	0,273 4	0,276 4	0,279 4	0,282 3	0,285 2
0,8	0,288 1	0,291 0	0,293 9	0,296 7	0,299 5	0,302 3	0,305 1	0,307 8	0,310 6	0,313 3
0,9	0,315 9	0,318 6	0,321 2	0,323 8	0,326 4	0,328 9	0,331 5	0,334 0	0,336 5	0,338 9
1,0	0,341 3	0,343 8	0,346 1	0,348 5	0,350 8	0,353 1	0,355 4	0,357 7	0,359 9	0,362 1
1,1	0,364 3	0,366 5	0,368 6	0,370 8	0,372 9	0,374 9	0,377 0	0,379 0	0,381 0	0,383 0
1,2	0,384 9	0,386 9	0,388 8	0,390 7	0,392 5	0,394 4	0,396 2	0,398 0	0,399 7	0,401 5
1,3	0,403 2	0,404 9	0,406 6	0,408 2	0,409 9	0,411 5	0,413 1	0,414 7	0,416 2	0,417 7
1,4	0,419 2	0,420 7	0,422 2	0,423 6	0,425 1	0,426 5	0,427 9	0,429 2	0,430 6	0,431 9
1,5	0,433 2	0,434 5	0,435 7	0,437 0	0,438 2	0,439 4	0,440 6	0,441 8	0,442 9	0,444 1
1,6	0,445 2	0,446 3	0,447 4	0,448 4	0,449 5	0,450 5	0,451 5	0,452 5	0,453 5	0,454 5
1,7	0,455 4	0,456 4	0,457 3	0,458 2	0,459 1	0,459 9	0,460 8	0,461 6	0,462 5	0,463 3
1,8	0,464 1	0,464 9	0,465 6	0,466 4	0,467 1	0,467 8	0,468 6	0,469 3	0,469 9	0,470 6
1,9	0,471 3	0,471 9	0,472 6	0,473 2	0,473 8	0,474 4	0,475 0	0,475 6	0,476 1	0,476 7
2,0	0,477 2	0,477 8	0,478 3	0,478 8	0,479 3	0,479 8	0,480 3	0,480 8	0,481 2	0,481 7
2,1	0,482 1	0,482 6	0,483 0	0,483 4	0,483 8	0,484 2	0,484 6	0,485 0	0,485 4	0,485 7
2,2	0,486 1	0,486 4	0,486 8	0,487 1	0,487 5	0,487 8	0,488 1	0,488 4	0,488 7	0,489 0
2,3	0,489 3	0,489 6	0,489 8	0,490 1	0,490 4	0,490 6	0,490 9	0,491 1	0,491 3	0,491 6
2,4	0,491 8	0,492 0	0,492 2	0,492 5	0,492 7	0,492 9	0,493 1	0,493 2	0,493 4	0,493 6
2,5	0,493 8	0,494 0	0,494 1	0,494 3	0,494 5	0,494 6	0,494 8	0,494 9	0,495 1	0,495 2
2,6	0,495 3	0,495 5	0,495 6	0,495 7	0,495 9	0,496 0	0,496 1	0,496 2	0,496 3	0,496 4
2,7	0,496 5	0,496 6	0,496 7	0,496 8	0,496 9	0,497 0	0,497 1	0,497 2	0,497 3	0,497 4
2,8	0,497 4	0,497 5	0,497 6	0,497 7	0,497 7	0,497 8	0,497 9	0,497 9	0,498 0	0,498 1
2,9	0,498 1	0,498 2	0,498 2	0,498 3	0,498 4	0,498 4	0,498 5	0,498 5	0,498 6	0,498 6
3,0	0,498 7	0,498 7	0,498 7	0,498 8	0,498 8	0,498 9	0,498 9	0,498 9	0,499 0	0,499 0

Annexe E-4

Ressources Internet

Adresses Internet pratiques :

Statistique Canada

http://www.statcan.ca

Le site de Statistique Canada indique les adresses des sites connexes pour toutes les provinces et de nombreux sites web du monde.

Gouvernement du Manitoba

http://www.gov.mb.ca

Institut canadien d'information sur la santé

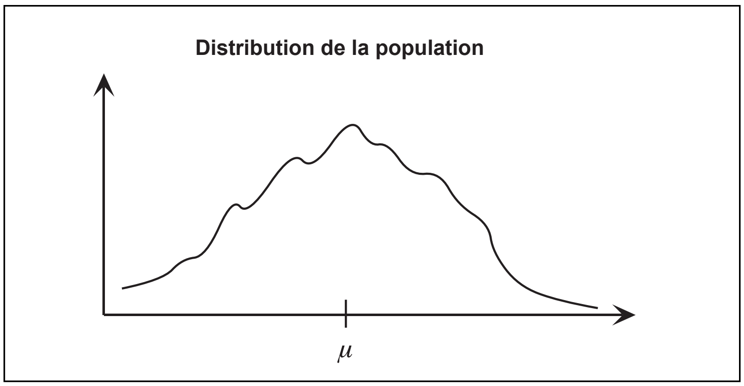
http://www.cihi.ca

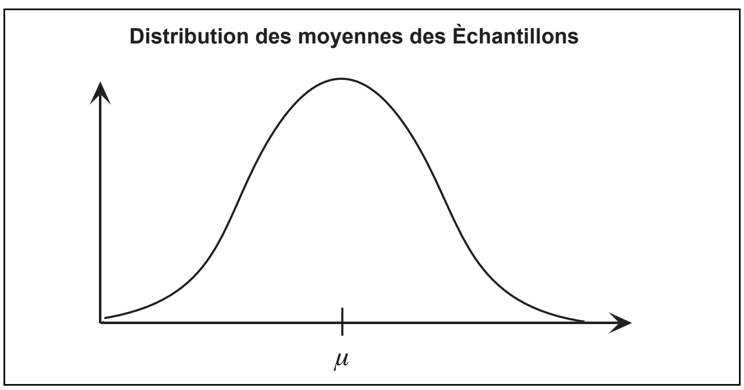
Annexe E-5

Autres éléments théoriques

- **Connaître l'erreur quadratique moyenne ($\sigma_{\overline{X}}$).**
  Supposons que la moyenne d'une population est μ et que l'écart type est σ.

  Nous pouvons prouver que si nous prenons un grand nombre d'échantillons de taille n d'une population ayant une distribution raisonnable quelconque, les moyennes des échantillons seront regroupées autour de μ dans une distribution normale.





De plus, la distribution des moyennes des échantillons aura son propre écart type, que l'on nomme l'***erreur quadratique moyenne* ($\sigma_{\overline{X}}$).**

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

Si dans la distribution, la moyenne = μ et l'écart type = σ, l'ensemble de tous les échantillons de taille n de la distribution forme une distribution normale dont la moyenne = μ et l'écart type = σ.

Exemple 1

Si les valeurs d'une distribution sont les suivantes : $\mu = 75$ et $\sigma = 20$,

a) où devraient se situer 95 % des moyennes des échantillons aléatoires de taille 25?

b) où devraient se situer 95 % des moyennes des échantillons aléatoires de taille 100?

Solutions

a) $\sigma_{\overline{X}} = \dfrac{20}{\sqrt{25}} = \dfrac{20}{5} = 4$

Nous devons nous attendre à ce que 95 % des moyennes se situent à 2 $\sigma_{\overline{X}}$ ou moins de 75. L'intervalle est de 75 – 2 x 4 à 75 + 2 x 4 ou (67, 83).

b) $\sigma_{\overline{X}} = \dfrac{20}{\sqrt{100}} = \dfrac{20}{10} = 2$

Nous devons nous attendre à ce que 95 % des moyennes se situent à $2\sigma_{\overline{X}}$ ou moins de 75. L'intervalle est de 75 – 2 x 2 à 75 + 2 x 2 ou (71, 79).

- **Construire des intervalles de confiance.**

Dans la plupart des cas, nous ne disposons que d'un échantillon, et nous aimerions savoir où se situe μ.

Si nous avons un échantillon, sa moyenne et l'écart type de la population, nous pouvons construire un intervalle de confiance autour de la moyenne de l'échantillon dans lequel μ, la moyenne de la population, se situera.

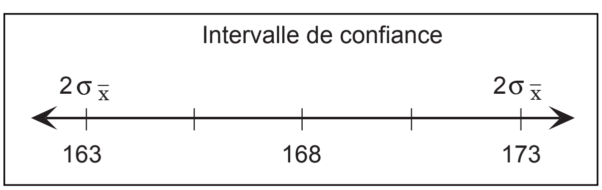
Puisque nous désirerons utiliser une estimation de l'écart type de la population, nous utiliserons la formule relative à l'écart type de la population (s). Sur la calculatrice, nous utiliserons la touche s plutôt que la touche σ.

Si un échantillon aléatoire a les valeurs suivantes : taille = 36, $\overline{x}$ = 168 et s = 15, nous pouvons calculer l'erreur quadratique moyenne comme nous l'avons fait ci-dessus.

$$\sigma_{\overline{X}} = \frac{15}{\sqrt{36}} = \frac{15}{6} = 2{,}5$$

Nous ne savons pas où se situe μ, mais peu importe son emplacement, il existe une probabilité de 95 % que la moyenne de tout échantillon, y compris la nôtre, se situe à (2 x 2,5 =) 5 de μ. Ainsi, il existe une probabilité de 95 % que μ et que la moyenne de notre échantillon, 168, se situent à une distance maximale de 5.

Par conséquent, nous pouvons construire un intervalle de confiance de 95 % autour de 168.



Il existe une probabilité de 95 % que μ se situe quelque part dans cet intervalle.

Nota : μ se situe probablement dans cet intervalle, mais cela ne nous indique pas où se situeront les moyennes des autres échantillons. Il existe une probabilité de 95 % qu'elles se situeront à une distance maximale de 5 de μ, mais nous ne savons pas où se situe μ. Si μ se situe près de l'une des limites de l'intervalle de confiance, les moyennes de nombreux échantillons futurs seraient situées loin de 168.

L'intervalle de confiance nous indique où se situe probablement μ. Il ne nous indique pas où se situeront probablement les moyennes des échantillons futurs.

Exemple 1

Les vitesses d'un échantillon aléatoire de 100 automobiles circulant devant un parc sont enregistrées par un radar. Dans cet échantillon, $\bar{x} = 39$ et $s = 6$.

a) Déterminez l'intervalle de confiance de 95 % dans lequel sera probablement située la vitesse moyenne de toutes les automobiles circulant devant le parc.

b) Pour les données ci-dessus, déterminez l'intervalle de confiance de 99 %.

Solution

a) $\sigma_{\bar{x}} = \dfrac{6}{\sqrt{100}} = 0{,}6$

L'intervalle de confiance de 95 % se situe de (39 – 2 x 0,6) à (39 + 2 x 0,6) = (37,8, 40,2).

b) L'intervalle de confiance de 99 % se situe de (39 – 3 x 0,6) à (39 + 3 x 0,6) = (37,2, 40,8).

Exemple 2

Un échantillon de 40 élèves de 6e année dans une ville effectue un examen de 20 questions de multiplications de base. Pour cet échantillon, $\bar{x} = 15$ et $s = 3$.

a) Déterminez l'intervalle de confiance de 95 % pour μ, le score moyen de la ville.

b) Quel serait le risque couru en faisant ce calcul si l'échantillon était en fait une classe complète de 40 élèves?

c) S'il était important de calculer μ le mieux possible, que pourriez-vous faire?

Solution

a) $\sigma_{\bar{x}} = \dfrac{3}{\sqrt{40}} = 0{,}474 = 0{,}5$

L'intervalle de confiance est 15 ± 1 = (14, 16).

b) Il ne s'agirait probablement pas d'un échantillon aléatoire.

c) Les réponses peuvent varier. Voici quelques réponses possibles :

- Utiliser un échantillon plus grand.
- Combiner plusieurs classes.
- Faire une sélection aléatoire parmi tous les élèves de la ville.

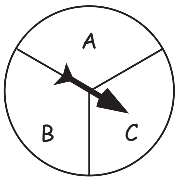
- **Déterminer les niveaux de confiance et les proportions.**

Nous utilisons l'erreur quadratique moyenne des scores de deux façons :

a) Si nous connaissons la moyenne d'une population, nous pouvons estimer où les moyennes des échantillons de taille n se situeront probablement.

b) Si nous ne disposons que d'un seul échantillon, nous pouvons construire un intervalle de confiance dans lequel la moyenne de la population se situera probablement.

Nous pouvons faire ces deux mêmes choses avec les ***proportions***. S'il existe un tiers des chances que la flèche ci-dessous s'arrête sur A, nous pouvons nous attendre à ce qu'elle s'arrête sur A dans un tiers des cas si elle est tournée un grand nombre de fois.



Pour en savoir plus, nous devons avoir l'erreur quadratique moyenne d'une proportion. Si p correspond à la probabilité qui nous intéresse et si nous pensons enregistrer N événements, donc

$$\sigma_{\overline{X}} = \sqrt{\frac{p\,(1 - p)}{N}}$$

Cette formule devient suspecte lorsque p est trop éloigné de 0,5. Il est préférable de ne pas l'utiliser lorsque $p < 0{,}2$ ou $p > 0{,}8$.

Exemple 1

Lorsque nous connaissons la proportion d'une population.

On sait que 30 % des personnes d'une population auront probablement un rhume en novembre ou en décembre. Les données seront recueillies dans un échantillon aléatoire de 100 personnes. Déterminez l'intervalle dans lequel il existe 95 % des chances que les rhumes de l'échantillon se situent.

Solution

$$\sigma_{\overline{X}} = \sqrt{\frac{(0{,}3)\,(0{,}7)}{100}} = 0{,}045$$

Nous pouvons nous attendre à ce que la proportion des rhumes de l'échantillon se situe dans l'intervalle de 0,3 – (2 x 0,045) à 0,3 + (2 x 0,045) ou (0,21, 0,39) dans 95 % des cas. Nous pouvons affirmer que dans un échantillon de 100 personnes, nous pouvons nous attendre à ce que les personnes qui ont un rhume soient dans l'intervalle (21, 39) 19 fois sur 20.

Exemple 2

Lorsque nous ne disposons que d'une proportion de l'échantillon.

Le directeur d'un supermarché détermine qu'un samedi en particulier, 25 % des 200 adultes qui sont entrés dans le magasin étaient accompagnés par des enfants. Déterminez l'intervalle de confiance de 95 % pour le pourcentage des adultes qui seront accompagnés par des enfants sur une longue période.

Solution

$$\sigma_{\overline{X}} = \sqrt{\frac{(0{,}25)\,(0{,}75)}{200}} = 0{,}03$$

L'intervalle de confiance de 95 % se situe de 0,25 – (2 x 0,03) à 0,25 + (2 x 0,03) ou à (0,19, 0,31) La proportion de la population se situe probablement dans cet intervalle.

Nous pouvons affirmer que, d'après cet échantillon de 200 adultes, la proportion accompagnée par des enfants à long terme se situe entre 19 % et 31 %, 19 fois sur 20.

- **Comprendre et interpréter les sondages d'opinions.**

  Les politiciens, les gouvernements et les journalistes veulent souvent savoir ce que les électeurs pensent d'une question ou pour qui ils voteront.

  Par exemple, une maison de sondage peut demander à un échantillon aléatoire de 100° personnes si elles voteraient ou non pour le parti R si une élection avait lieu aujourd'hui.

  Si 47 % des personnes répondaient le parti R, $\sigma_{\overline{X}} = \sqrt{\dfrac{0{,}47 \times 0{,}53}{1000}} = 0{,}016$.

  L'intervalle de confiance de 95 % pour la proportion votant pour le parti R est de $0{,}47 \pm (2 \times 0{,}016) = (0{,}438,\ 0{,}502)$. Puisque 0,032 est près de 3 %, on pourrait indiquer que le parti R est appuyé par 47 % de la population. Ce pourcentage serait exact à 3 % près, dans 95 % des cas (ou 19 fois sur 20).

  Il faut noter que :

  a) les sondages n'offrent aucune certitude, non en raison des statistiques (s'ils sont correctement effectués), mais pour les raisons suivantes :

     — il est difficile d'obtenir de bons échantillons aléatoires d'électeurs;

     — certaines personnes ne disent pas la vérité afin d'envoyer un message aux candidats;

     — certaines personnes ne disent pas la vérité parce qu'elles n'aiment pas les sondages;

     — les opinions des gens peuvent changer rapidement.

  b) la ***population*** peut être d'une taille quelconque. Si l'échantillon est aléatoire, l'intervalle de confiance est le même peu importe si la population est de 5 000 ou de 20 000 000;

  c) la largeur de l'intervalle de confiance dépend de la taille de l'échantillon. L'étendue de l'échantillon ci-dessus est de 6 %. L'intervalle de confiance pour un échantillon de 10 000 personnes (si la maison de sondage peut se le permettre) aurait une erreur quadratique moyenne $\sigma_{\overline{X}} \cong 0{,}005$ et l'intervalle de confiance serait (0,46, 0,48).

  Exemple 1

  Une maison de sondage questionne 200 personnes, choisies au hasard, et détermine que 130 de ces personnes préfèrent la marque A à la marque B. Calculez l'intervalle de confiance de 95 % pour ce résultat et rédigez un communiqué de presse à ce sujet.

  Solution

  Probabilité de choisir la marque A = 130 ÷ 200 = 0,65.

  $$\sigma_{\overline{X}} = \sqrt{\frac{(0{,}65)(1 - 0{,}65)}{200}} \approx 0{,}034$$

  Intervalle requis = $0{,}65 \pm 2(0{,}034) = 0{,}65 \pm 0{,}068 = (0{,}582,\ 0{,}718)$.

  Exemple de communiqué de presse : Le pourcentage de personnes ayant choisi la marque A est exact à 65 % à 7 points de pourcentage près, 19 fois sur 20.

  Autre exemple de communiqué de presse : Entre 58 % et 72 % de la population choisissent la marque A.

  Exemple 2

  Une maison de sondage signale que 36 % des clients d'un supermarché préfèrent la marque A, et elle affirme que cette estimation est exacte à 7 % près, dans 95 % des cas. Quelle est la taille de l'échantillon?

Solution

L'erreur quadratique moyenne ($\sigma_{\overline{X}}$) doit être de $\dfrac{0,07}{2} = 0,035$

Ce qui correspond à $\sqrt{\dfrac{0,36 \times 0,64}{n}} = 0,035.$

Au carré, $\dfrac{0,36 \times 0,64}{n} = 0,001\,2$

$$n = \frac{0,36 \times 0,64}{0,001\,2} = 192$$

La maison a probablement utilisé un échantillon de 200 personnes.