

Mathématique Appliquée 30S

Note :

Statistiques

Nom : _____

Table Des Matières

Leçon 1 : Exploration des données	p. 3
Leçon 2 : Tableaux des fréquences, histogrammes et polygones des fréquences	p. 7
Leçon 3 : L'Écart type	p. 15
Leçon 4 : Distribution normale	p. 23
Leçon 5 : Cotes Z	p. 31
Leçon 6 : Intervalles de confiance	p. 41

Leçon 1 : Exploration des données

Les statistiques sont tout partout...

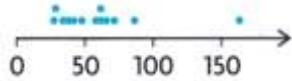
- 3 spécialistes sur 4 recommandent...
- 83% des personnes interrogées...

A) Vocabulaire

La tendance centrale : C'est les valeurs de la moyenne, la médiane et le mode

Valeur aberrante : Dans un ensemble de données, valeur qui est très différente des autres.

Tracé linéaire : Graphique dans lequel la valeur de chaque donnée d'un ensemble est marquée sous la forme d'un point au-dessus d'une droite numérique. (C'est un graphique qui montre la fréquence de données sur une droite numérique.)



Moyenne (arithmétique) (\bar{x}) : La somme des données divisées par le nombre de données.

Moyenne coupée : La moyenne des données sans la/les valeurs aberrantes.

Moyenne pondérée : La moyenne des données où chaque donnée a un certain pourcentage de la moyenne totale.

Médiane : La donnée qui se trouve dans le milieu. (Si c'est deux nombres dans le milieu, tu trouves la moyenne de ces deux nombres.)

Mode : La donnée qui répète le plus.

Dispersion : Mesure qui varie en fonction de l'étalement (l'étendue) des données dans un ensemble. La valeur de la dispersion est zéro si toutes les données d'un ensemble sont identiques; sa valeur augmente à mesure que les données s'étalent.

B) Application de vocabulaire

Exemple 1 : La tendance centrale

Suppose que Lisa a obtenu les notes suivantes à neuf tests de mathématiques :

76, 77, 89, 92, 95, 71, 76, 88, 83

a) Placez les données en ordre croissant :

b) Quel est la mode?

c) Quel est la médiane?

d) Calcule l'étendue (dispersion) :

e) Calcule la moyenne des notes obtenues :

Exemple 2 : La tendance centrale, la valeur aberrante et la moyenne coupée

Suppose que Jeanne a obtenu les notes suivantes à sept tests de mathématiques : 72, 84, 63, 77, 91, 54 et 77, 35. Le nombre maximal de points attribués pour chaque test est 100.

- a) Placez les données en ordre croissant :
- b) Quel est la mode?
- c) Quel est la médiane?
- d) Calcule l'étendue :
- e) Calcule la moyenne des notes obtenues :
- f) Détermine la valeur aberrante.
- g) Détermine la moyenne coupée.

Exemple 3 : Trouve un résultat qui manque si la moyenne est donnée

Après une absence de plusieurs jours de l'école, Laurent Hudson n'a obtenu que 41% pour son premier test de mathématiques. Par la suite, il a obtenu les notes suivantes : 68, 74, 55 et 86. Mme Couture a autorisé Laurent à reprendre le premier test. Quelle est la note minimale qu'il peut obtenir pour s'assurer une 80% comme moyenne?

Exemple 4 :

Lois est membre de la Kechewin Cree First Nation d'Alberta. Elle est bien connue pour les combinaisons de couleurs, les modèles et la finesse de ses perlage quand elle fait de la décoration pour les objets utilitaires ou sacrés. Elle note le nombre d'heures passées sur chacun des articles qu'elle produit. Cela l'aide à fixer le prix de chaque objet. Voici les heures qu'elle a passées à fabriquer ses 10 derniers sacs à main; 6; 8; 37; 3; 9; 6; 8,5; 5; 7,5 et 9.

- a) Combien d'heures Lois a-t-elle passées en moyenne à fabriquer chaque sac à main?

- b) Quelle est la moyenne coupée des heures passés à fabriquer chaque sac à main?

- c) Lequel de ces deux résultats semble constituer la meilleure moyenne du nombre d'heures qu'elle a passées à fabriquer chaque sac à main? Pourquoi?

- d) Qu'est-ce qui pourrait expliquer la présence de ces valeurs aberrantes?

Exemple 5 : La moyenne pondérée

Nicole calcule sa note finale à un cours de mathématiques au quotidien. Les projets valent 45%, les tests valent 35% et l'examen final vaut 20%.

Nicole a obtenu :

40% aux projets;
60% aux tests;
75% à l'examen final.

Calcule, en utilisant une moyenne pondérée, sa note finale.

Exemple 6 : la tendance centrale

Lors d'un recensement dans une école, on demande aux élèves combien de téléphones ils ont à la maison s'ils incluent les téléphones cellulaires. Voici le résultat :

Nombre de téléphones dans la famille	
Nombre de téléphones	Nombre d'élèves
1	4
2	130
3	200
4	160
5	48
6	55
7	11

Quelle est la moyenne, le mode, la médiane et l'étendue.

Leçon 2 : Tableaux des fréquences, histogrammes et polygones des fréquences

On peut démontrer les données d'un sondage/questionnaire sous forme d'histogramme. Ceci veut dire que les réponses numériques peuvent être collectionnées et ensuite démontrées en utilisant les barres verticales.

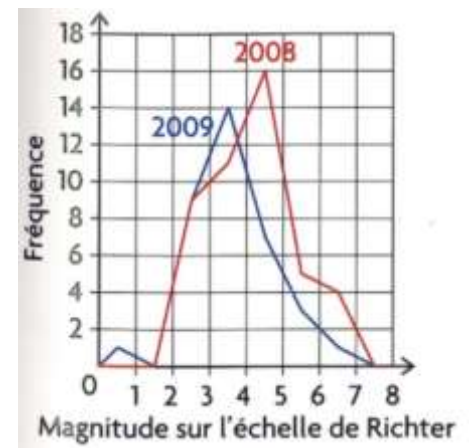
Distribution des fréquences : Ensemble d'intervalles (tableau ou diagramme), habituellement de grandeur égale, à l'intérieur desquels les données sont organisées.

Débit (m ³ /s)	Marque de pointage	Fréquence (nombre d'années)
0-500		6
500-1000		11
1000-1500		9
1500-2000		14
2000-2500		5
2500-3000		1
3000-3500		3
3500-4000		0
4000-4500		0
4500-5000		1



Histogramme : Diagramme d'une distribution des fréquences où les intervalles égaux des valeurs sont inscrits sur un axe horizontal, et les fréquences associées à ces intervalles sont indiquées par l'aire des rectangles tracés pour représenter ces intervalles.

Polygone des Fréquences : Graphique d'une distribution des fréquences, obtenu en reliant les points milieux des intervalles par des lignes droites.



Exemple 1 :

Le bassin de la rivière Rouge subit des inondations de manière récurrente. Pendant la deuxième moitié du XXe siècle, il y a eu neuf inondations notables, dont quatre graves, en 1950, 1979, 1996 et 1997. L'inondation de 1997 au Manitoba et dans le Dakota du Nord a été surnommée « l'inondation du siècle ». Les données suivantes représentent les débits de la rivière Rouge de 1950 à 1999, telles qu'elles ont été enregistrées sous le pont Redwood, à Winnipeg.

Débits d'eau maximums de la rivière Rouge, de 1950 à 1999, mesurés sous le pont Redwood*									
Année	Débit (m ³ /s)	Année	Débit (m ³ /s)	Année	Débit (m ³ /s)	Année	Débit (m ³ /s)	Année	Débit (m ³ /s)
1950	3 058	1960	1 965	1970	2 280	1980	881	1990	396
1951	1 065	1961	481	1971	1 526	1981	159	1991	280
1952	1 008	1962	1 688	1972	1 589	1982	1 458	1992	1 399
1953	357	1963	660	1973	530	1983	1 393	1993	946
1954	524	1964	1 002	1974	2 718	1984	1 048	1994	1 121
1955	1 521	1965	1 809	1975	1 671	1985	991	1995	1 877
1956	1 974	1966	2 498	1976	1 807	1986	1 812	1996	3 058
1957	654	1967	1 727	1977	187	1987	2 339	1997	4 587
1958	524	1968	510	1978	1 750	1988	564	1998	1 557
1959	991	1969	2 209	1979	3 030	1989	1 390	1999	2 180

Conseil national de recherches du Canada

Pour déterminer le débit d'eau associé à une forte inondation nous pouvons concevoir une **distribution des fréquences, un histogramme ou un polygone de fréquences.**

1. Détermine l'étendue des données : (plus grand moins plus petit)

Plus haut débit d'eau : 4 587 m³/s, en 1997

Plus bas débit d'eau : 159 m³/s, en 1981

L'étendu : 4 587 – 159 = 4 428

2. Choisir un intervalle convenable (5-10 tranches): (Divise l'étendue par le nombre d'intervalle)

$$\frac{4\,428}{10} = 442,8$$

diviser par 5 sera des valeurs trop grandes dans ce cas

3. Choisir une valeur pour l'intervalle qui est facile à utiliser. (Arrondi la réponse)

Si la grandeur de l'intervalle est 500, les intervalles commenceront à 0, 500, 1000, 1500, ... 4500.

A) Dresser une distribution des fréquences sous forme de tableau

1. Utilise les données et crée votre tableau de distribution des fréquences.

Noté des valeurs qui sont plus importants; par exemple dans ce cas des débits qui ont causé des grandes inondations et leurs années. (1950, 1979, 1996, 1997)

Débit (m^3/s)	Marque de pointage	Fréquence (nombre d'années)
0-500		6
500-1 000		11
1 000-1 500		9
1 500-2 000		14
2 000-2 500		5
2 500-3 000		1
3 000-3 500		3 inondations
3 500-4 000		0
4 000-4 500		0
4 500-5 000		1 inondation

Les débits d'eau qui causent une forte inondation sont supérieurs à 2000 m^3/s .

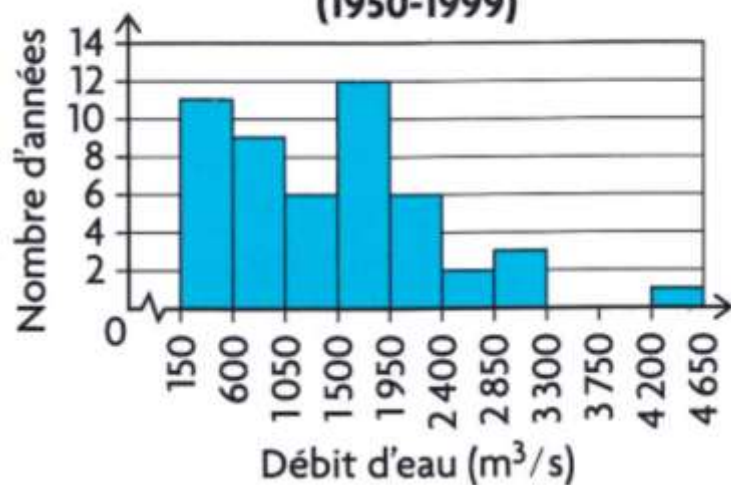
Un débit de 500 sera dans l'intervalle de 500 – 1000, alors un débit de 1500 sera dans quelle intervalle?

B) Construire un histogramme

1. Utilise les données et construis un histogramme afin de représenter les données de la distribution des fréquences puisque les données sont groupées en intervalles.

Débit (m ³ /s)	Marque de pointage	Fréquence (nombre d'années)
150-600	III III	11
600-1050	III IIII	9
1050-1500	III	6
1500-1950	III III II	12
1950-2400	III	6
2400-2850	II	2
2850-3300	III	3
3300-3750		0
3750-4200		0
4200-4650		1

Débits d'eau de la rivière Rouge à Winnipeg (1950-1999)



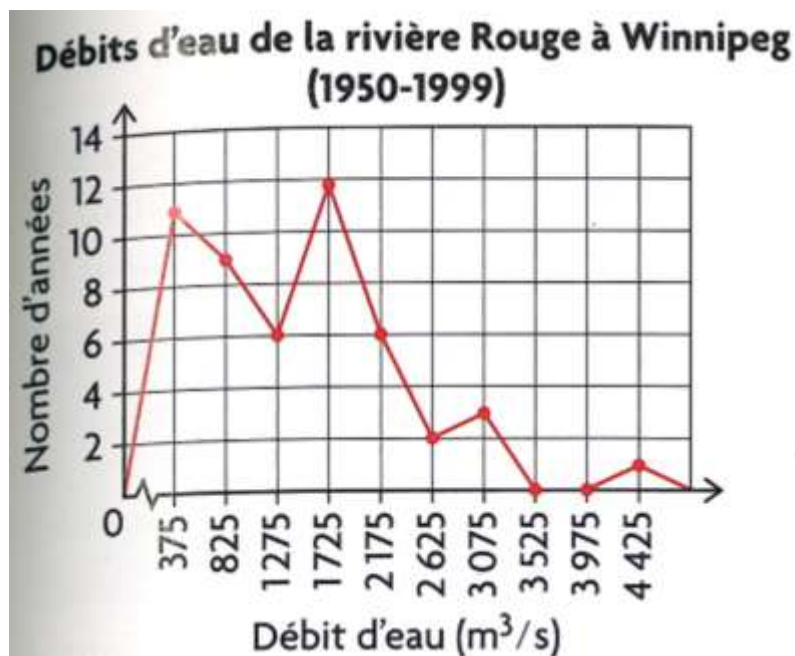
Décris les débits d'eau.

En règle générale le nombre de points de données dans chaque intervalle décroît à mesure que le débit d'eau maximum s'élève. Les débits d'eau maximums bas ont été plus courants que les débits maximums hauts. Selon l'histogramme, le débit n'a dépassé 1950 m³/s qu'en 12 occasions. Les inondations doivent avoir eu lieu pendant ces années-là. Je prédis que des inondations se produisent quand le débit est supérieur à 1950 m³/s.

C) Tracer un polygone des fréquences

1. Utilise les données et crée un polygone des fréquences.

Débit (m ³ /s)	Point milieu	Fréquence (nombre d'années)
150–600	375	11
600–1 050	825	9
1 050–1 500	1 275	6
1 500–1 950	1 725	12
1 950–2 400	2 175	6
2 400–2 850	2 625	2
2 850–3 300	3 075	3
3 300–3 750	3 525	0
3 750–4 200	3 975	0
4 200–4 650	4 425	1



La majorité des données se situent dans les quatre premiers intervalles, et le débit le plus courant est entre 1 500 et 2 000 m³/s. Après cela, les fréquences chutent brutalement. Le débit a été d'environ 2 625, 3 075 ou 4 425 m³/s lors de six années. Des inondations doivent avoir eu lieu pendant ces années-là. Les trois autres inondations ont dû se produire quand le débit était d'environ 2 175 m³/s. Selon le polygone des fréquences, les débits de six années se situaient autour de ce point milieu. Si je suppose que le débit lors de trois de ces années était supérieur ou égale à 2 175 m³/s, les inondations ont probablement lieu quand le débit est de 2 175 m³/s ou plus.

D) La Comparaison des Méthodes

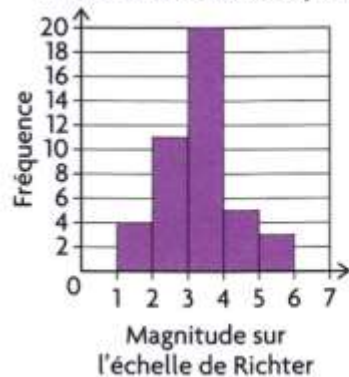
Exemple 2 :

La magnitude d'un tremblement de terre se mesure à l'aide de l'échelle de Richter. Examine les histogrammes de fréquences des magnitudes pour ces séismes au Canada de 2005 à 2009. En quelle année les tremblements de terre auraient-ils pu causer le plus de dommages ?

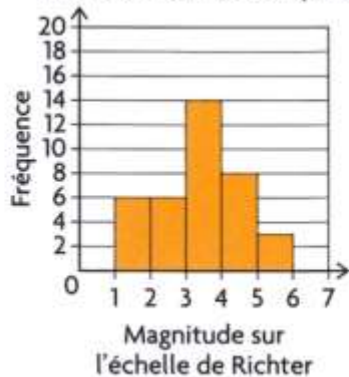
Comprendre l'échelle de Richter*	
Magnitude	Conséquences
Moins de 3,0	non ressenti; détecté par les sismographes
3,0–3,9	ressenti comme le passage d'un camion; aucun dommage
4,0–4,9	ressenti par presque tout le monde; mouvement des objets instables
5,0–5,9	ressenti par tous; dommages considérables aux bâtiments faibles
6,0–6,9	difficulté à se tenir debout; destruction partielle des bâtiments ordinaires
7,0–7,9	pertes de vie; destruction des bâtiments ordinaires
Plus de 7,9	pertes de vie nombreuses et destruction

Les histogrammes

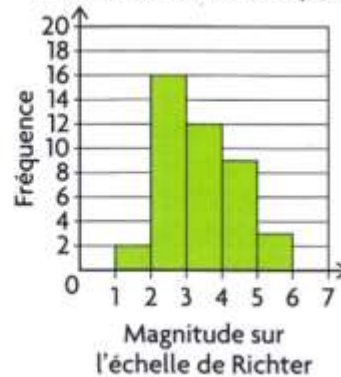
Tremblements de terre, 2005



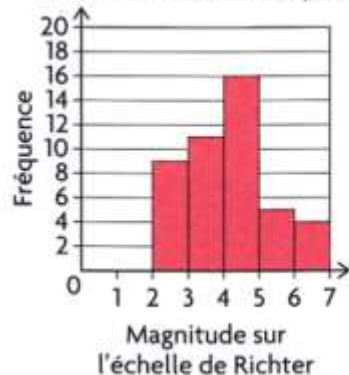
Tremblements de terre, 2006



Tremblements de terre, 2007



Tremblements de terre, 2008



Tremblements de terre, 2009

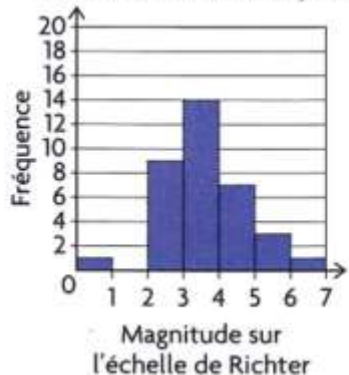
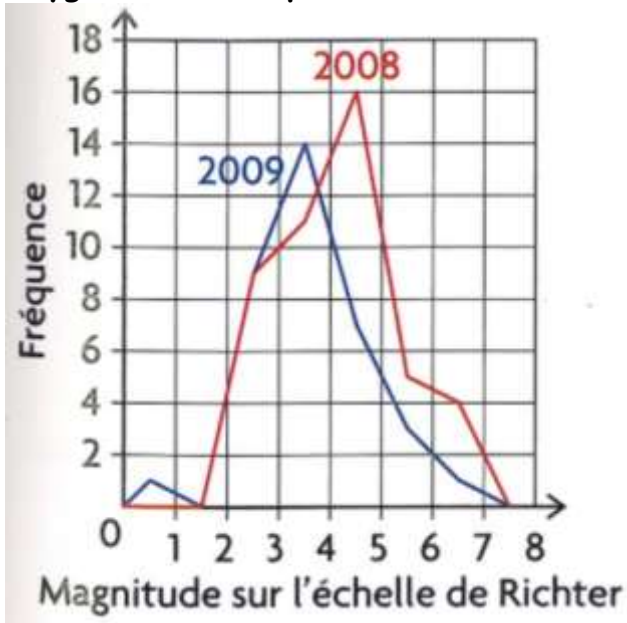


Tableau des fréquences

Année	2005	2006	2007	2008	2009
Fréquence des tremblements de terre entre 4,0 et 4,9	5	8	9	16	7

Année	Magnitude sur l'échelle de Richter			Total
	4,0–4,9	5,0–5,9	6,0–6,9	
2008	16	5	4	25
2009	7	3	1	11

Polygones des fréquences



Exemple 3 :

Dans un cours de mathé, les notes de bulletins sont les suivantes :

Notes	90	80	70	60	50	40
Fréquence	4	6	9	4	2	2

Leçon 3 : L'Écart type

Écart : La différence entre la valeur d'une donnée et la moyenne de l'ensemble de données dont elle fait partie. $(x - \bar{x})$

Écart type (σ) (standard deviation): Mesure de la dispersion (l'étendue) de la valeur des données par rapport à la moyenne. Un **petit écart type** indique que la majorité des données sont **proches** de la moyenne ; par contre, un **grand écart type** indique que la majorité des données sont **dispersées loin** de la moyenne. Le symbole σ (qui se dit « sigma ») représente l'écart type des données. (n représente combien de données il y a dans votre échantillon.)

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

A) Calcul de l'Écart-type à la main

Exploration :

Exemple 1 :

Une enseignante a deux classes d'élèves en chimie. Elle leur donne les mêmes examens. Examine la note moyenne de chacun des cinq premiers examens donnés aux deux classes. Compare les résultats des deux classes.

Examen	Classe A (%)	Classe B (%)
1	94	84
2	56	77
3	89	76
4	67	81
5	84	74

- a) Détermine les moyennes arithmétiques, \bar{x} , pour les deux classes.

- b) Remplis le tableau.

Examen	Classe A $(x - \bar{x})$	Classe A $(x - \bar{x})^2$	Classe B $(x - \bar{x})$	Classe B $(x - \bar{x})^2$
$\bar{x} \text{ A} =$ $\bar{x} \text{ B} =$				
1				
2				
3				
4				
5				
(Somme) $\sum(x - \bar{x})^2$				
$\frac{\sum(x - \bar{x})^2}{n}$ ($n = \#$ de données)				

- c) Détermine l'écart-type pour les deux classes. Explique ce que cette valeur représente en termes des données.

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

Exemple 2 :

Trouve la moyenne et l'écart type. Explique ce que représente l'écart type.

- a) Les données suivantes représentent les prix (en dollars) des téléphones mobiles.

35, 50, 60, 60, 75, 65, 80

- b) Les données suivantes donnent le nombre de «home runs» pour les 10 meilleurs batteurs pour la saison régulière du Major League Baseball en 2005.

51, 48, 47, 46, 45, 43, 41, 40, 40, 39

B) Calcul de l'Écart-type avec Excel avec une variété de donnée (qui ne répète pas)

Exemple 3 :

Daniel travaille à temps partiel à la cantine du centre communautaire local. L'une de ses tâches consiste à décharger des camions de livraison. Il s'est interrogé sur la justesse des masses inscrites sur deux cartons contenant des sacs de graines de tournesol. Il a mesuré les masses des 20 sacs contenus dans les deux cartons. L'un contenait des sacs de 227 g et l'autre, des sacs de 454 g.

Masses des sacs de 227 g (g)			
228	220	233	227
230	227	221	229
224	235	224	231
226	232	218	218
229	232	236	223

Masses des sacs de 454 g (g)			
458	445	457	458
452	457	445	452
463	455	451	460
455	453	456	459
451	455	456	450

Comment des mesures de dispersion peuvent-elles servir à déterminer si la justesse du mesurage est la même pour les deux tailles de sacs ?

a) Détermine l'étendue des données

b) Détermine l'écart-type des données.

(Utilise le programme de excel.)

Masses des sacs de 227 g				Masses des sacs de 454 g			
228	220	233	227	458	445	457	458
230	227	221	229	452	457	445	452
224	235	224	231	463	455	451	460
226	232	218	218	455	453	456	459
229	232	236	223	451	455	456	450
STDEV.P(? :?) =				$\sigma = \sqrt{\frac{\sum(x - \bar{x})^2}{n}}$			
STDEV.P(A2:D6) = 5.227571138							

C) Calcul de l'Écart-type avec la Calculatrice Graphique

STAT : EDIT : Edit

L1 : les données

L2 : la fréquence que les données répètent

L1	L2
228	1
230	1
224	1
226	1
229	1
220	1
227	1
235	1
232	1
232	1
233	1
221	1
224	1
218	1
236	1
227	1
229	1
231	1
218	1
223	1

STAT : CALC : 1- Var Stats L1, L2 entrer

c) Quelles données sont le moins constant (pas aussi proche à la moyenne ? Pourquoi ?

D) Calcul de l'Écart-type avec desmos avec une variété de donnée

<https://www.desmos.com/calculator/>

The image shows a screenshot of the Desmos calculator interface. It features two data tables and a calculation window. The first table has a header with x_1 and y_1 , and contains 12 rows of data. The second table also has a header with x_1 and y_1 , and contains 8 rows of data. Below the tables, a calculation window shows the expression $\text{stdevp}(x_1)$ and its result, $= 5.22757113773$.

x_1	y_1
228	1
230	1
224	1
226	1
229	1
220	1
227	1
235	1
232	1
233	1
221	1
224	1

x_1	y_1
218	1
236	1
227	1
229	1
231	1
218	1
223	1
232	1

$\text{stdevp}(x_1)$
 $= 5.22757113773$

$\text{stdevp}(x_1) =$

Exemple 4 :

Trouve la moyenne et l'écart type. Explique ce que représente l'écart type.

- a) Les données suivantes représentent les prix (en dollars) des téléphones mobiles.

35, 50, 60, 60, 75, 65, 80

- b) Les données suivantes donnent le nombre de «home runs» pour les 10 meilleurs batteurs pour la saison régulière du Major League Baseball en 2005.

51, 48, 47, 46, 45, 43, 41, 40, 40, 39

E) Écart-type avec les fréquences autre que 1 (donnée répétée)

Exemple 4 :

Angèle fait un sondage pour déterminer le nombre d'heures par semaine que passent les garçons de 11^e année de son école à jouer à des jeux vidéo. Elle a déterminé que la **moyenne était de 12,84 h**, avec **un écart type de 2,16 h**. Janessa a fait un sondage semblable auprès des filles de 11^e année de son école. Elle a organisé ses résultats dans ce tableau des fréquences. Compare les résultats des deux sondages.

Heures de jeu par semaine pour les filles de 11 ^e année	
Heures	Fréquence
3-5	7
5-7	11
7-9	16
9-11	19
11-13	12
13-15	5

À LA MAIN ET AVEC EXCEL

Document Excel Leçon 3 exemple 3

	A	B	C	D	E	F	G
1	Heures de jeu par semaine pour les filles de 11 ^e année						
2	Heures	Point milieu de l'intervalle (x)	Fréquence	f * x	(x- \bar{x})	(x- \bar{x}) ²	f*(x- \bar{x}) ²
3	3 - 5 heures	4	7	28	(4-8,942857)	24.43184	171.0229
4	5 - 7 heures	6	11	66	(6-8,942857)	8.660408	95.26449
5	7 - 9 heures	8	16	128	(8-8,942857)	0.88898	14.22367
6	9 - 11 heures	10	19	190	(10-8,942857)	1.117551	21.23347
7	11 - 13 heures	12	12	144	(12-8,942857)	9.346122	112.1535
8	13 - 15 heures	14	5	70	(14-8,942857)	25.57469	127.8735
9	somme total = sum(:)		70	626			541.7714
10							
11	$\bar{x} = \frac{\sum(f)(x)}{n}$	moyenne =	626/70=	8.942857			
12							
13							
14							
15							
16							
17							
18							
19							
20							
21	Avec Excel:						
22							
23	$\sigma = \text{SQRT}(F9/C9) =$	sqrt(541.7714.../70) =		2.782012192			
24							
25		OU moyenne = SUMPRODUCT(B3:B8,C3:C8)/SUM(C3:C8)		8.942857			
26							
27							
28							
29							
30							

$\sigma = 2,782 \dots$ heures

Garçons: $\bar{x} = 12,84$ h
 $\sigma = 2,16$ h

Filles: $\bar{x} = 8,942\dots$ h
 $\sigma = 2,782\dots$ h

En moyenne, les garçons jouent près de 4 heures de plus par semaine que les filles.

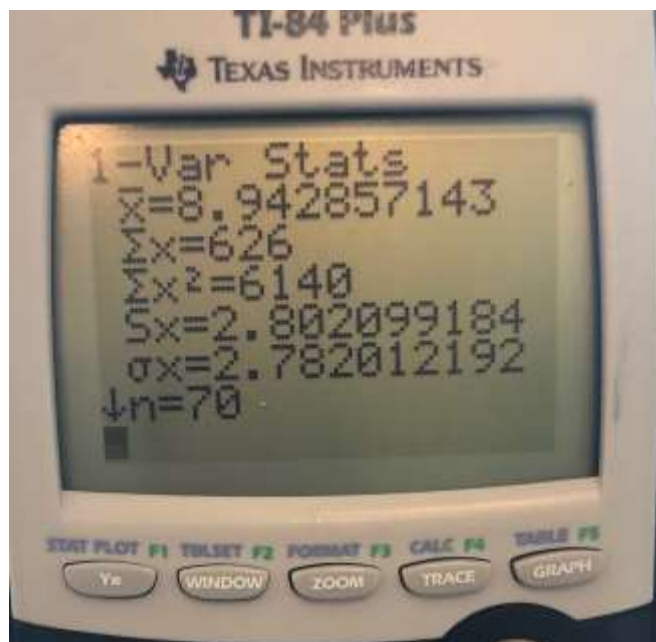
L'écart type des garçons est plus faible que celui des filles. Par conséquent, les temps de jeu des garçons sont plus proches de leur moyenne (presque 13 h) et ne varient pas autant.

AVEC CALCULATRICE GRAPHIQUE

STAT : EDIT : Edit

L1	L2
4	7
6	11
8	16
10	19
12	12
14	5

STAT : CALC : 1- Var Stats L1, L2 entrer



Heures de jeu par semaine pour les filles de 11^e année:

$\bar{x} = 8,942\dots$ h $\sigma = 2,782\dots$ h

Heures de jeu par semaine pour les garçons de 11^e année:

$\bar{x} = 12,84$ h $\sigma = 2,16$ h

- L'écart-type des filles est plus élevé que celui des garçons. Par conséquent, les temps des filles varient plus par rapport à leur moyenne de 9 h.
- L'écart-type des garçons est plus faible. Par conséquent, leurs données sont plus constantes, malgré une moyenne plus élevée.

Vidéo Excel pour trouver les tendances centrales et l'écart type

<https://www.google.ca/search?q=is+there+a+way+to+find+the+standard+deviation+of+values+with+fr%C3%A9quence+in+excel&ie=&oe=#kpvalbx= I hWXePGN4Xu Qa8pLeYAO24>

Exemple 5 :

Voici les salaires annuels des employés de MTS

Détermine l'écart-type des données à la main.

Personnes	Salaire annuel
5	36 500\$
8	48 200\$
6	61 500\$
3	75 000\$
1	110 000\$

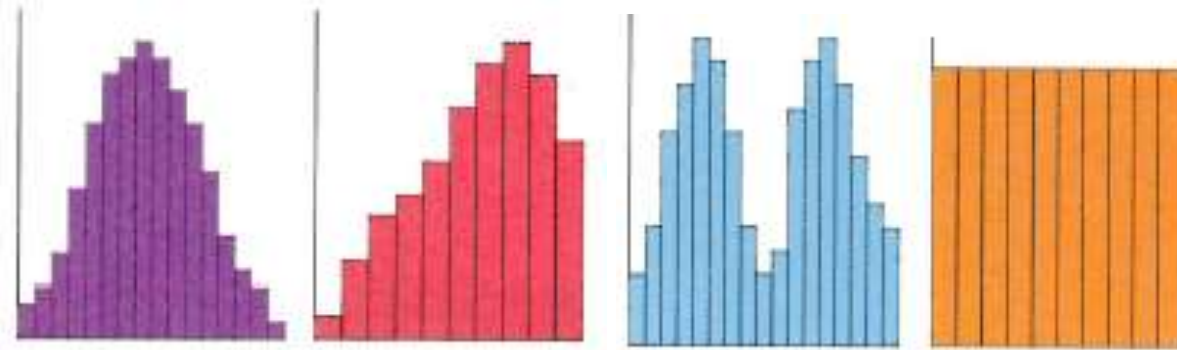
Leçon 4 : Distribution normale

Lorsqu'on considère les réponses à une question non-biaisé, à laquelle les réponses sont numériques, une distribution normale arrive si ces réponses peuvent se placer dans un graphique qui ressemble à une cloche (bell curve).

Courbe normale : Courbe symétrique qui représente la distribution normale. On l'appelle aussi **courbe en cloche**.

Distribution normale : Données qui, une fois tracées sous forme d'histogramme ou de polygone des fréquences, produisent une **distribution symétrique** de part et d'autre de la moyenne.

Une distribution des données prend parfois une forme particulière. Par exemple, le premier diagramme ci-dessous n'a qu'un seul sommet, la forme a un seul mode.



Exemple 1 :

Heidi inaugure un magasin de planches à neige près d'une station de ski de la région. Elle sait que la longueur recommandée d'une planche est associée à la taille du ou de la planchiste. Ses recherches montrent que la plupart des planchistes qui visitent la station sont des hommes âgés de 20 à 39 ans. Dans le but d'offrir les longueurs de planche les plus populaires, elle collecte des données sur la taille de 1000 hommes canadiens âgés de 20 à 39 ans.

Comment Heidi peut-elle utiliser ces données pour avoir en magasin des planches de longueur adéquate ?

$$\bar{x} = 69,52$$

$$\sigma = 2,984225193$$

Médiane : 69,5

Mode : 69,5

N = 1000 (DONNÉES)

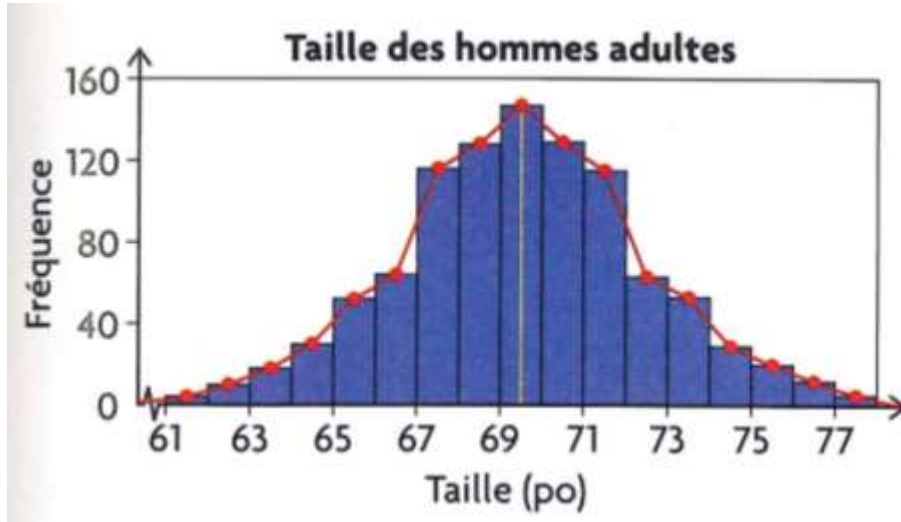
Alors les données $\frac{500 + 501}{2} = \text{médiane}$

69,5 est dans le milieu



Taille (po)	Fréquence
moins que 61	3
61-62	4
62-63	10
63-64	18
64-65	30
65-66	52
66-67	64
67-68	116
68-69	128
69-70	147
70-71	129
71-72	115
72-73	63
73-74	53
74-75	29
75-76	20
76-77	12
77-78	5
78 ou plus	2

Trace un histogramme pour montrer la distribution des tailles. Trace une ligne verticale sur l'histogramme pour représenter la moyenne. Trace ensuite un polygone des fréquences en reliant les points milieux au sommet de chaque rectangle du diagramme. Les données sont presque symétriques de part et d'autre de la moyenne et elles décroissent graduellement de chaque côté. Le polygone des fréquences ressemble à une cloche.



Pour déterminer si les données sont **normalement distribuées** on doit :

- vérifier le pourcentage des données à un écart type plus ou moins la moyenne
- vérifier le pourcentage des données à deux écarts type de la moyenne

Les données sont normalement distribuées.

Tailles à l'intérieur de un écart type de la moyenne :

$69,520 - 2,984\dots$, soit 66,5 po

$69,520 + 2,984\dots$, soit 72,5 po

Tailles à l'intérieur de deux écarts types de la moyenne :

$69,520 - 2(2,984\dots)$, soit 63,6 po

$69,520 + 2(2,984\dots)$, soit 75,5 po

Étendue	Étendue des tailles
De $\bar{x} - 1\sigma$ à $\bar{x} + 1\sigma$	D'environ 66,5 po à 72,5 po
De $\bar{x} - 2\sigma$ à $\bar{x} + 2\sigma$	D'environ 63,6 po à 75,5 po

Taille (po)	Fréquence
moins que 61	3
61-62	4
62-63	10
63-64	18
64-65	30
65-66	52
66-67	64
67-68	116
68-69	128
69-70	147
70-71	129
71-72	115
72-73	63
73-74	53
74-75	29
75-76	20
76-77	12
77-78	5
78 ou plus	2

Nombre d'hommes à l'intérieur de un écart type de la moyenne de 67 po à 73 po :
 $116 + 128 + 147 + 129 + 115 + 63$, soit 698

Nombre d'hommes à l'intérieur de deux écarts types de la moyenne de 64 po à 76 po :
 $30 + 52 + 64 + 116 + 128 + 147 + 129 + 115 + 63 + 53 + 29 + 20$, soit 944

Étendue	Étendue des tailles	Pourcentage des données
De $\bar{x} - 1\sigma$ à $\bar{x} + 1\sigma$	De 66,5 po à 72,5 po	$\frac{698}{1000}$, soit 69,8 %
De $\bar{x} - 2\sigma$ à $\bar{x} + 2\sigma$	De 63,5 po à 75,5 po	$\frac{946}{1000}$, soit 94,6 %

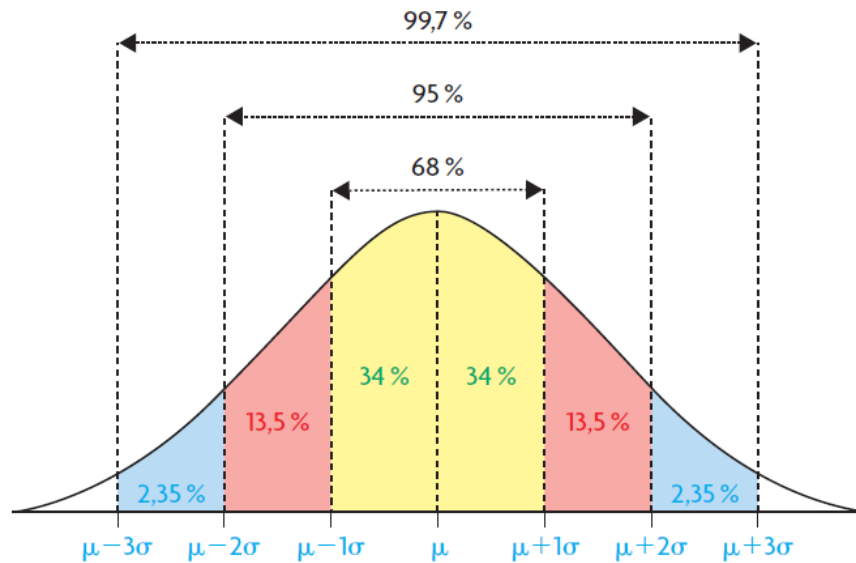
Environ 70 % des tailles sont à l'intérieur de un écart type de la moyenne.

Environ 95 % des tailles sont à l'intérieur de deux écarts types de la moyenne.

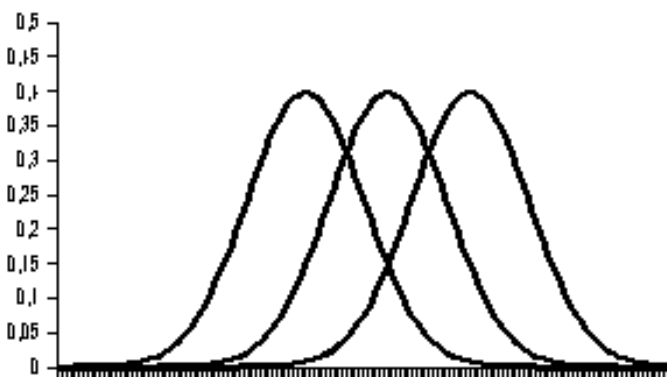
Je prédis qu'environ 70 % de mes clients masculins auront besoin de planches à neige convenant à des tailles allant de 66,5 à 72,5 po.

On peut résumer comme suit les propriétés d'une distribution normale :

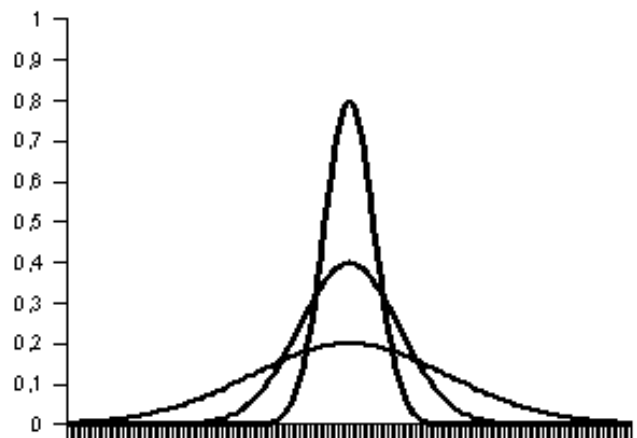
- Le graphique est symétrique. La moyenne, la médiane et le mode sont égaux (ou presque) et ils se situent sur l'axe de symétrie.
- La courbe normale a la forme d'une cloche : elle culmine au milieu, descend en pente vers les côtés et s'approche de zéro aux extrémités.
- Environ 68 % des données sont à l'intérieur de un écart type de la moyenne.
- Environ 95 % des données sont à l'intérieur de deux écarts types de la moyenne.
- Environ 99,7 % des données sont à l'intérieur de trois écarts types de la moyenne.
- L'aire sous la courbe normale est égale à 1 et elle représente 100 % des données.



Graphiquement, si trois distributions normales ne diffèrent que par leur moyenne:



Alors que si elles ne diffèrent que par leur écart type:



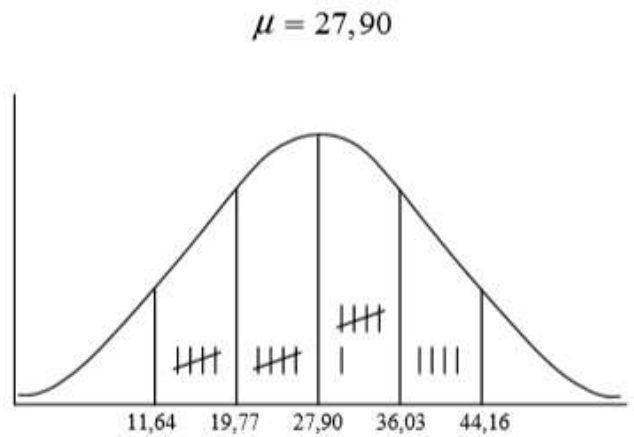
Exemple 2 :

Les données ci-dessous représentent l'âge (en années) de 20 arbres dans une zone de végétation naturelle.

15 17 18 19 19 22 24 25 26 27
 28 28 29 32 34 35 37 38 42 43

a) Détermine la moyenne et l'écart type de la population pour ces données (à 2 décimales près).

b) Détermine si ces données représentent une distribution normale. Donne 2 raisons pour justifier ta réponse. Montre ton travail en utilisant les statistiques.



-1σ à 1σ } $\frac{11}{20}$ ou 55 % des nombres se situent dans cette étendue
 l'étendue est 19,77 à 36,03 } (normale est 68 %)

-2σ à 2σ } $\frac{20}{20}$ ou 100 % des nombres se situent dans cette étendue
 l'étendue est 11,64 à 44,16 } (normale est 95 %)

Ceci ne représente pas une distribution normale car elle ne suit pas la règle de 68-95-99.

Exemple 3 :

Dans son chenil, Jim élève des huskys de Sibérie. Les données qu'il a accumulées au fil des ans lui ont appris que le poids des chiens adultes mâles est normalement distribué, avec une moyenne de 52,5 lb et un écart type de 2,4 lb. Jim s'est servi de cette information pour tracer une courbe normale avec

- 68 % des données à l'intérieur de un écart type de la moyenne ;
- 95 % des données à l'intérieur de deux écarts types de la moyenne ;
- 99,7 % des données à l'intérieur de trois écarts types de la moyenne.

a) Trace un histogramme qui représente la distribution normale.

b) Quel pourcentage des chiens mâles adultes du chenil de Jim t'attends-tu à voir peser entre 47,7 lb et 54,9 lb ?

c) Quel pourcentage des chiens mâles adultes du chenil de Jim t'attends-tu à voir peser moins que 47,7 lb ?

d) Quel pourcentage des chiens mâles adultes du chenil de Jim t'attends-tu à voir peser plus que 54,9 lb ?

Exemple 4 :

Deux équipes de baseball se rendent en avion aux Jeux autochtones de l'Amérique du Nord. Les membres de chaque équipe avaient des bagages à main pour leur matériel sportif. Les masses de ces bagages étaient normalement distribuées, selon les caractéristiques présentées ci-contre.

Équipe	μ (kg)	σ (kg)
Hommes	6,35	1,04
Femmes	6,35	0,59

- a) Trace un graphique pour montrer la distribution des masses des bagages de chaque équipe.
- b) L'équipe des femmes a gagné le championnat. Chacune a reçu en souvenir une médaille et une balle de baseball dont la masse combinée atteint 1,18 kg et qu'elles ont mis dans leurs bagages à main. Trace un graphique pour montrer le changement dans la distribution des masses de leurs bagages à main pour le voyage de retour.

Exemple 5 :

Un groupe de personnes ont été demandé la question suivante : « Combien de sports avez-vous déjà joué sur une équipe? » Voici les résultats :

Nombre de sports	Fréquence
0-4	6
5-9	6
10-14	5
15-19	1
20-24	0

a) Détermine la moyenne et l'écart type pour ces données (à 2 décimales près).

b) Détermine si ces données représentent une distribution normale. Donne 2 raisons pour justifier ta réponse. Montre ton travail en utilisant les statistiques.

Leçon 5 : Cotes Z

Distribution standard normal : Distribution normale dont la moyenne est de 0 et l'écart type est de 1.

Pour tout score donnée x à partir d'une distribution normale

$$x = \mu + z \sigma$$

Où z représente le nombre d'écart types du score par rapport à la moyenne.

Cote Z : valeur normalisée qui indique le nombre d'écart types de la valeur d'une donnée au-dessus ou au-dessous de la moyenne.

$$Z = \frac{x - \mu}{\sigma} = \frac{\text{Note} - \bar{x}}{\sigma} = \frac{\text{Note actuelle} - \text{note moyenne}}{\text{l'écart type}}$$

- Les cotes Z sont une façon de comparer deux éléments d'une façon objective.
- Les cotes Z nous indique comment loin nous sommes de la normale ou de l'écart type.

μ (*mu*): la moyenne d'une population complète (et non d'un échantillon)

Table de cotes Z : Table qui présent la fraction des données ayant une cote Z inférieur à la valeur de toute donnée dans une distribution standard normale.

A) Les Calculs et comparaison de cote z

Exemple 1 :

Ella et Serge font partie d'un club de course de Vancouver. Une partie de leur entraînement consiste en un sprint de 200 m. Le tableau ci-dessous présente les temps normalement distribués pour le sprint de 200 m à Vancouver et lors d'un récent voyage à Lake Louise. À de grandes altitudes, les temps de course s'améliorent.

Endroit	Altitude (m)	Temps moyen μ du club pour 200 m (s)	Écart type σ du club (s)	Temps de course d'Ella (s)	Temps de course de Serge (s)
Vancouver	4	25,75	0,62	24,95	25,45
Lake Louise	1 661	25,57	0,60	24,77	26,24

Détermine l'endroit le temps de course d'Ella a été le meilleur par rapport aux résultats du club ?

Vancouver:

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

$$z = \frac{24,95 - 25,75}{0,62}$$

$$z = -1,290 \dots$$

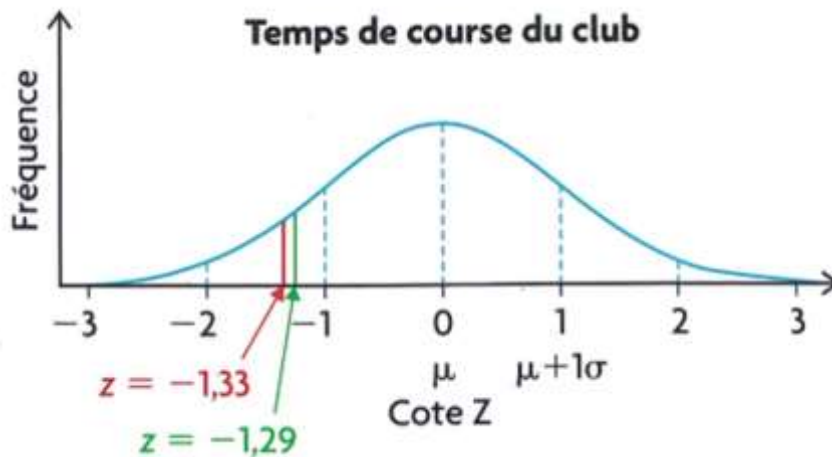
Lake Louise:

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

$$z = \frac{24,77 - 25,57}{0,60}$$

$$z = -1,333 \dots$$

La courbe standard normale (dont la moyenne est de 0 et l'écart type es de 1).



Le temps d'Ella pour la course de 200 m était meilleur que la moyenne du club aux deux endroits. Cependant, comme la cote Z d'Ella pour Lake Louise était plus basse que pour Vancouver, son temps a été meilleur à Lake Louise.

Exemple 2 :

Le tableau suivant représente les résultats du lancer de javelot pour Laurentia, Cokolin et Bricelet :

Concours	Laurentia	Cokolin	Bricelet
St. Malo	47 m	54 m	58 m
Notre Dame de Lourdes	40 m	50 m	60 m

La moyenne des lancers de tous les lanceurs à St.Malo fut 52 m avec un écart type de 2.

La moyenne des lancer de tous les lanceurs à Notre Dame de Lourdes fut 50 avec un écart type de 5.

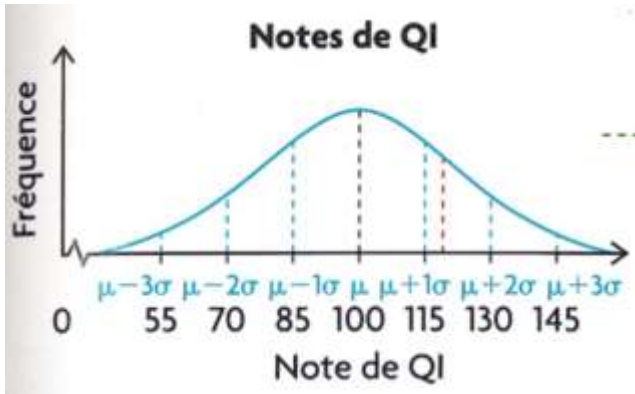
Où le lancer a –t-il été le meilleur pour les trois concurrents par rapport à l’ensemble des concurrents ?

B) Le pourcentage des résultats

Exemple 3 :

On se sert parfois de tests de quotient intellectuel (QI) pour mesurer la capacité intellectuelle d'un individu à un moment précis. Les notes de QI sont normalement distribuées, avec une moyenne de 100 et un écart type de 15. Si un individu obtient 119 lors d'un test de QI, comment cette note se compare-t-elle avec les notes de la population en général ?

Méthode 1 : calcul à la main

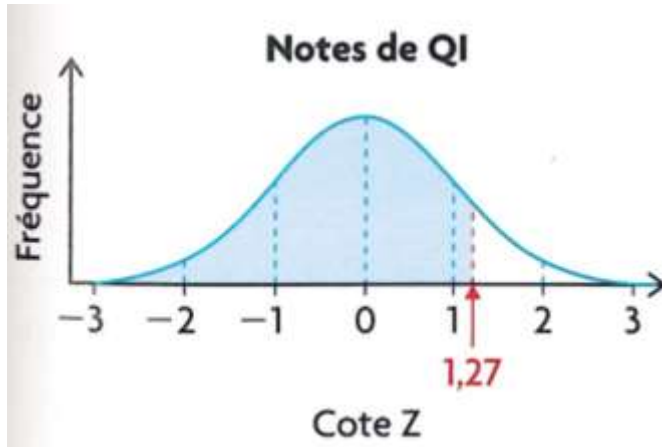


$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{119 - 100}{15}$$

$$z = 1,2666\dots$$

Courbe standard normale :



Une note de 119 correspond à environ 1,27 écart type au-dessus de la moyenne.

D'après notre table de cotes Z nous obtenons :

$$1,27 = 1,2 + 0,07$$

(Rangée de 1,2 et colonne de 0,07)

La valeur de 0,8980, dans la table, représente l'aire **sous la courbe à gauche** de la cote Z.

z	0,0	0,01	0,06	0,07
0,0	0,5000	0,5040	0,5239	0,5279
0,1	0,5398	0,5438	0,5636	0,5675
1,1	0,8643	0,8665	0,8770	0,8790
1,2	0,8849	0,8869	0,8962	0,8980
1,3	0,9032	0,9049	0,9131	0,9147

Cela signifie qu'une note de QI de 119 est supérieure à 89,80% des notes de QI dans la population général.

Alors $100\% - 89,80\% = 10,2\%$ de la population on un QI supérieur de 119.

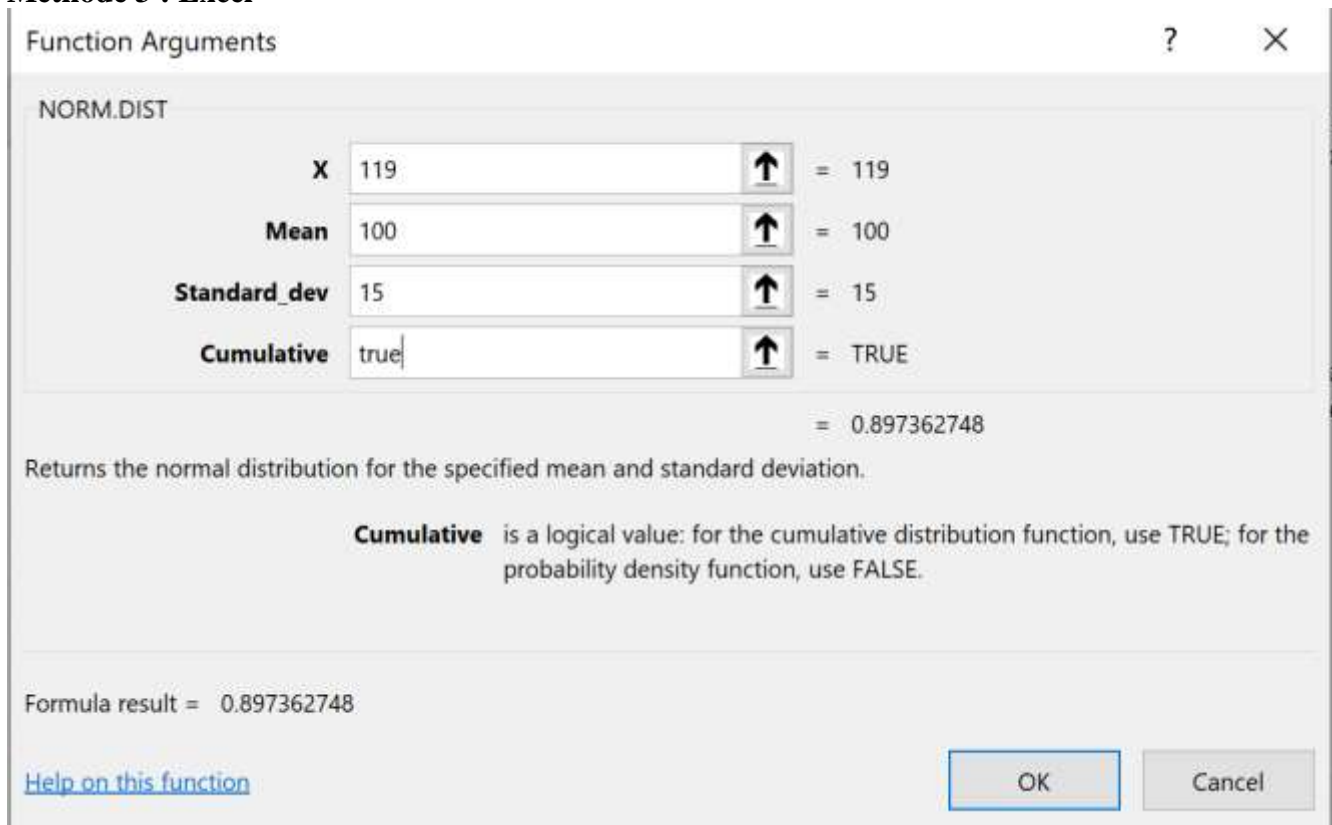
Méthode 2 : Calculatrice Graphique



2nd VARS (DISTR) :
2 : normal cdf (0,119,100,15) entrer
(moyenne de 0, donnée, moyenne
arithmétique, écart type)

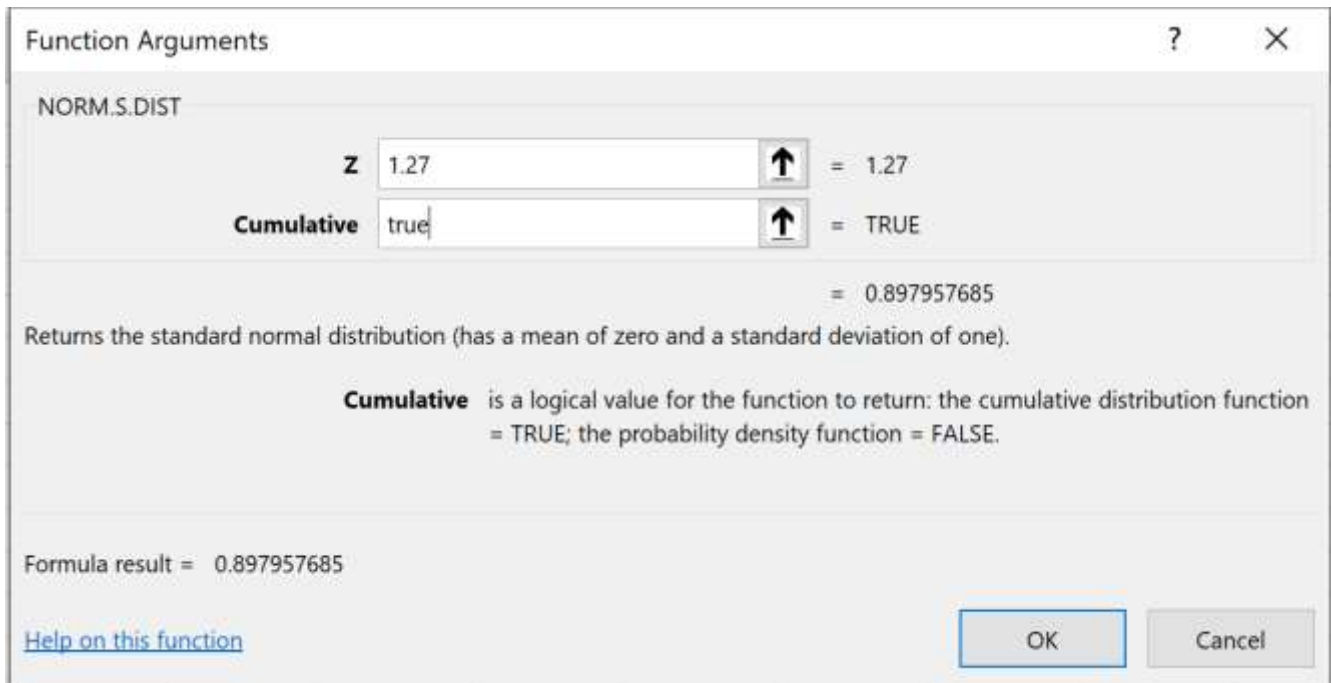
Une note de 119 est supérieure à
89,74 % de toute les notes de QI.

Méthode 3 : Excel



=NORM.DIST(119,100, 15, TRUE)
= 0,897362... = 89,74 %

Ou



=NORM.S.DIST(1.27,TRUE)

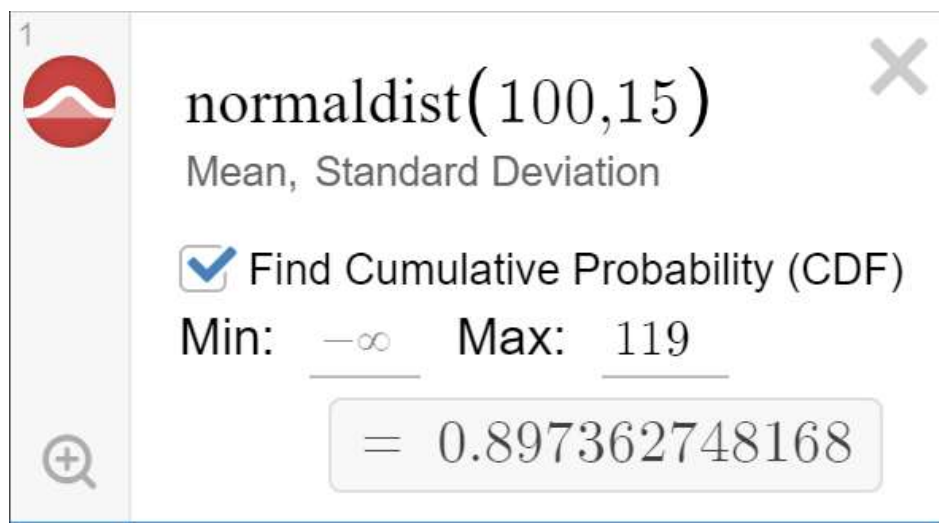
= 0,897958...

= 89,80%

Méthode 4 : Avec Desmos

normaldist(\bar{x} , σ)

Min. : _____ Max : _____



a) Détermine le % de personne qui ont un IQ plus grand que 121.

b) Détermine le % de personne qui ont un IQ plus petit que 95.

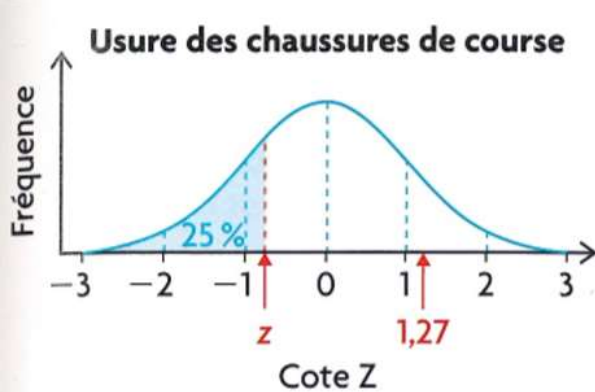
c) Détermine le % de personne qui ont un IQ entre 80 et 105.

Exemple 4 :

Les athlètes doivent remplacer leurs chaussures de course avant qu'elles perdent leur capacité d'absorption des chocs.

Les chaussures de course perdent leur capacité d'absorption après une distance moyenne de 640 km, avec un écart type de 160 km. Zack est un coureur d'élite et il veut remplacer ses chaussures après 25 % de leur durée normale. Après combien de kilomètres devrait-il remplacer ses chaussures ?

Méthode 1 : Calcul à la main



Z	0,09	0,08	0,07	0,06	0,05
-0,7	0,2148	0,2177	0,2206	0,2236	0,2266
-0,6	0,2451	0,2483	0,2514	0,2546	0,2578
-0,5	0,2776	0,2810	0,2843	0,2877	0,2912

Tu regardes dans la table de cotes Z pour une valeur qui est proche de 0,25.

La cote Z qui représente une aire de 0,25 se trouve presque à mi-chemin entre -0,67 et -0,68, soit environ -0,675.

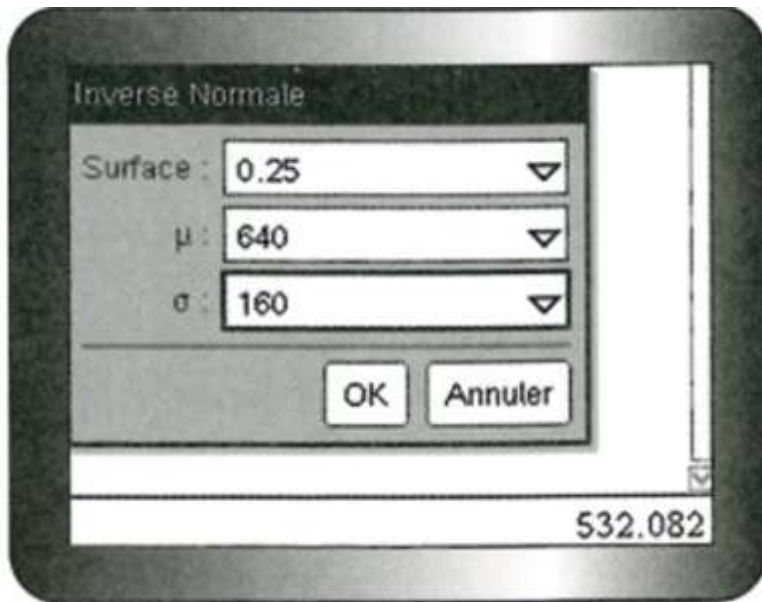
$$z = \frac{x - \mu}{\sigma}$$

$$(-0,675) = \frac{x - (640)}{(160)}$$

$$-108 = x - 640$$

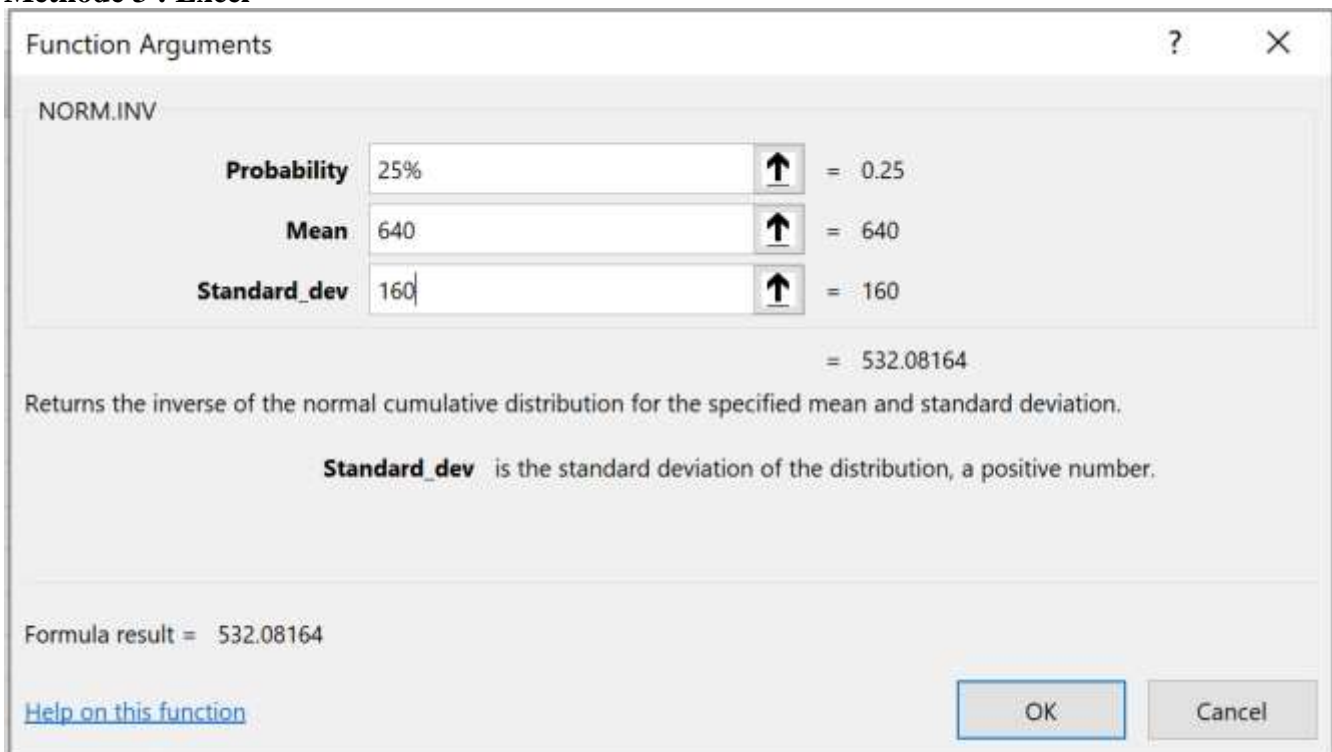
$$532 = x$$

Méthode 2 : Calculatrice Graphique



2ND VARS (DISTR) : invNorm(0.25,640,160)
invNorm(probabilité %, moyenne, écart
type)
= 532.0816401
=532

Méthode 3 : Excel



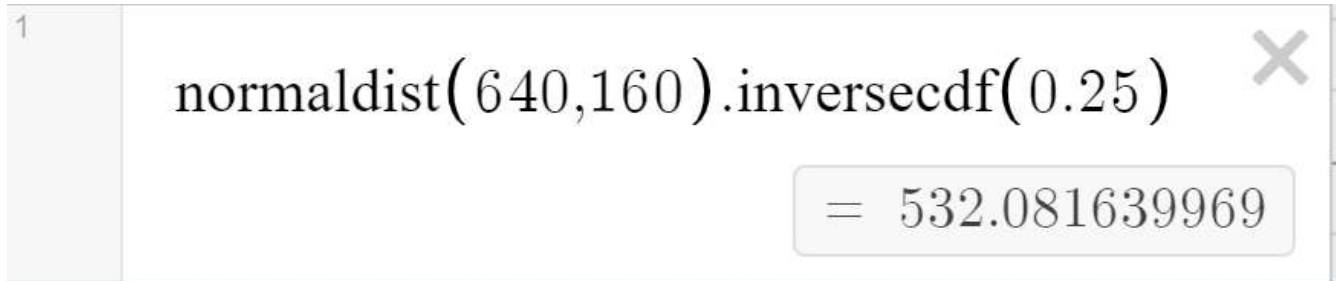
=NORM.INV(25%,640,160)
= 532.08164
= 532

Zack devrait remplacer ses chaussures de course après 532 km.

Méthode 4 : Avec Desmos

normaldist(\bar{x} , σ).inversecdf(# décimal)

Le nombre décimal que tu insères c'est le numéro qui représente les données à la gauche.



1

normaldist(640,160).inversecdf(0.25)

= 532.081639969

Exemple 5 :

Le fabricant ABC produit des tendeurs. Quand la ligne de montage fonctionne bien, la longueur des tendeurs fabriqués est distribuée normalement, avec une moyenne de 45,2 cm et un écart type de 1,3 cm. Les tendeurs qui mesurent moins de 42,0 cm ou plus de 48,0 cm sont rejetés par le service de contrôle de la qualité.

- a) Si 20 000 tendeurs sont fabriqués par jour, combien t'attendrais-tu à ce que le service de contrôle de la qualité en rejette ?
- b) Quelle décision pourrait prendre le fabricant à la lecture des résultats ?

Solution :

- i) **Détermine les cotes Z**
- ii) **Choisis une méthode et détermine la probabilité (%) des aires qui seront rejetés.**

Longueur minimum = 42 cm

Longueur maximum = 48 cm

$$z_{\min} = \frac{x - \mu}{\sigma}$$

$$z_{\max} = \frac{x - \mu}{\sigma}$$

Exemple 6 :

Un manufacturier de baladeurs a déterminé que ses appareils ont une durée moyenne de bon fonctionnement de 32,4 mois, avec un écart type de 6,3 mois.

Quelle durée (en mois) de garantie doit être offerte si le manufacturier souhaite restreindre les réparations à moins de 1,5 % de tous les baladeurs vendus ?

Exemple 7 :

Lois a eu 88% sur son examen de bio. La moyenne de la classe est de 72% et l'écart type est 9,7. Clark a eu 85% sur son examen de chimie. La moyenne de la classe est de 70% et l'écart type est de 8,4. Qui a mieux fait selon sa classe?

On calcul les cotes Z.

Résumé :

• Une cote Z indique le nombre d'écart types qui séparent la valeur d'une donnée de la moyenne. On la calcule à l'aide de la formule :

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

• Une cote Z positive indique que la valeur de la donnée se situe au-dessus de la moyenne. Une cote Z négative indique que la valeur de la donnée se situe au-dessous de la moyenne.

• On peut trouver l'aire sous la courbe standard normale, à la gauche d'une cote Z particulière, dans une table de cotes Z, avec la calculatrice graphique ou avec le programme d'Excel.

Leçon 6 : Intervalles de confiance

Marge d'erreur :

Différence possible entre l'estimation de la valeur qu'on tente de déterminer, telle qu'elle est déterminée par un échantillon aléatoire, et la véritable valeur pour la population. La marge d'erreur est généralement exprimée sous la forme d'un pourcentage approximatif tel que $\pm 5\%$.

Intervalle de confiance :

Intervalle qui contiendrait la vraie valeur qu'on tente de déterminer avec un degré de probabilité défini; l'intervalle de confiance peut s'exprimer de deux façons :

- par une écriture comme « 54,0 % \pm 3,5 % », ou
- par un intervalle comme « de 50,5 % à 57,5 % ».

L'intervalle de confiance utilise des données d'un échantillon pour déterminer comment certain que nous pouvons être de nos résultats pour la population générale.

On utilise des intervalles de confiance dans:

- Les élections
- Le contrôle de qualité dans une entreprise
- Les sondages
- N'importe quand qu'on veut faire une estimation à partir d'un échantillon

Niveau de confiance :

Probabilité que le résultat pour la « vraie » population soit contenu dans l'étendue de l'intervalle de confiance. Les sondages et les autres études se servent souvent d'un niveau de confiance de 95 %, quoique des pourcentages de 90 % ou de 99 % soient parfois employés.

Exemple 1 :

On a effectué un sondage téléphonique auprès de 600 individus choisis au hasard en milieu urbain. Selon les résultats, 76 % des individus interrogés, âgés de 18 à 34 ans, ont un compte de réseautage social. Les résultats sont justes à plus ou moins 4 points de pourcentage, 19 fois sur 20.

Calcule l'étendue des gens qui ont un compte de réseautage social et détermine la certitude des résultats.

Marge d'erreur : $\pm 4 \%$

L'intervalle de confiance 76 % $\pm 4 \%$

Alors $(76 - 4) = 72 \%$ et $(76 + 4) = 80 \%$, donc 72 % à 80 % c'est un étendu de 8 %.

Niveau de confiance c'est $19/20 = 95 \%$

La probabilité d'erreur pour ce résultat est de 5 %.

Alors si le sondage était effectué 100 fois, 95 fois sur 100, le pourcentage des individus qui ont un compte de réseautage social serait de 72 % à 80 %.

$$92\,500 \times 0,76 = 70\,300$$

$$92\,500 \times 0,04 = 3\,700$$

L'intervalle de confiance pour la population est $70\,300 \pm 3\,700$.

$$70\,300 - 3\,700 = 66\,600$$

et

$$70\,300 + 3\,700 = 74\,000$$

On peut dire, avec une confiance de 95 %, que de 66 600 à 74 000 individus, dans une population de 92 500 individus âgés de 18 à 34 ans, ont un compte de réseautage social.

Exemple 2 :

Les firmes canadiennes de sondage font fréquemment des enquêtes afin d'évaluer avec justesse les intentions de vote avant des élections. Pour ce faire, on demande à des échantillons de la population :

1. « Si une élection avait lieu aujourd'hui, pour quel parti voteriez-vous ? » Si la réponse est « Je ne sais pas », on pose alors la question suivante :

2. « Pour quel parti auriez-vous tendance à voter ? »

Le tableau suivant contient les résultats de trois sondages différents menés pendant la première semaine de novembre 2010. On considère que les résultats de ces sondages sont justes 19 fois sur 20.

Firme de sondage et données	Parti conservateur (%)	Parti libéral (%)	NPD (%)	Bloc québécois (%)	Parti vert (%)	Ne sait pas (%)
Ekos	29	29	19	9	11	12,6
taille de l'échantillon, 1 815 marge d'erreur, $\pm 2,3$ %						
Nanos	37	32	15	11	5	19,2
taille de l'échantillon, 844 marge d'erreur, $\pm 3,4$ %						
Ipsos	35	29	12	11	12	sans objet
taille de l'échantillon, 1 000 marge d'erreur, $\pm 3,1$ %						

Comment la taille de l'échantillon utilisé dans un sondage influence-t-elle :

a) la marge d'erreur des résultats publiés ?

b) l'intervalle de confiance des résultats publiés ?

Exemple 3 :

Pour respecter les normes réglementaires, la masse des balles de baseball doit se situer entre 142,0 g et 149,0 g. Une entreprise manufacturière a calibré son matériel de production pour fabriquer des balles dont la masse moyenne est de 145,0 g. Pour s'assurer que le matériel de production continue de fonctionner comme il se doit, le technicien responsable du contrôle de la qualité mesure chaque jour la masse d'un échantillon aléatoire de balles afin d'en déterminer la moyenne. Si la masse moyenne se situe entre 144,7 g et 145,3 g, alors le matériel de production fonctionne correctement. Si le niveau de la masse moyenne de l'échantillon n'est pas acceptable, on arrête les machines et on les règle. Le technicien responsable du contrôle de la qualité se réfère au tableau de la page suivante quand il fait son échantillonnage aléatoire.

Niveau de confiance	Taille de l'échantillon nécessaire
99 %	110
95 %	65
90 %	45

- a) Quel intervalle de confiance et quelle marge d'erreur le technicien utilise-t-il pour ses tests de contrôle de la qualité ?
- b) Interprète le tableau.
- c) Quelle est la relation entre le niveau de confiance et la taille de l'échantillon ?

Exemple 4 :

On a posé cette question à des électeurs et électrices : « Si une élection se tenait aujourd'hui, pour qui voteriez-vous ? » Voici les résultats du sondage :

- 53 % voteraient pour Sirois et
- 47 % voteraient pour Junot.

On a indiqué que les résultats étaient justes jusqu'à 3,8 points de pourcentage, 19 fois sur 20.

Qui remportera l'élection ?

Exemple 5 :

À une compagnie de construction, on trouve que la longueur moyenne d'une planche est de 3,05m, avec un écart type de 0,3m. Quelle est l'intervalle de confiance de 95% pour 100 planches?

Idées principales

- C'est souvent peu pratique, sinon impossible, d'obtenir des données pour une population complète. Au lieu de cela, on prend des échantillons aléatoires et on détermine la moyenne et l'écart type des données. Ces informations servent ensuite à faire des prédictions sur la population.
- Quand des données se rapprochent d'une distribution normale, un intervalle de confiance indique l'étendue dans laquelle la moyenne de tout échantillon de données d'une taille précise devrait être contenue, d'après un certain niveau de confiance. Cet intervalle de confiance peut ensuite servir à estimer l'étendue de la moyenne pour la population.
- La taille de l'échantillon, le niveau de confiance et la taille de la population déterminent l'étendue de l'intervalle de confiance d'un niveau de confiance donné.

Ce qu'il faut savoir

- Un intervalle de confiance s'exprime sous la forme d'un résultat d'enquête ou de sondage, plus ou moins la marge d'erreur.
- La marge d'erreur croît à mesure que croît le niveau de confiance (si la taille de l'échantillon est constante). La taille de l'échantillon nécessaire croît aussi à mesure que croît le niveau de confiance (si la marge d'erreur est constante).
- La taille de l'échantillon influence la marge d'erreur. Un grand échantillon mène à une marge d'erreur plus petite, en supposant que le même niveau de confiance soit requis.

Par exemple :

- On considère qu'un échantillon de 1 000 unités est juste à 6 3,1 %, 19 fois sur 20.
- On considère qu'un échantillon de 2 000 unités est juste à 6 2,2 %, 19 fois sur 20.
- On considère qu'un échantillon de 3 000 unités est juste à 6 1,8 %, 19 fois sur 20.