Index

OBJECTIF
Utiliser les fonctionnalités des TICE pour simuler des lancers d'un dé.

1. Simuler le lancer d'un dé cubique équilibré.
Il s'agit de produire aléatoirement un nombre entier compris entre 1 et 6.
Rentrer les instructions suivantes :

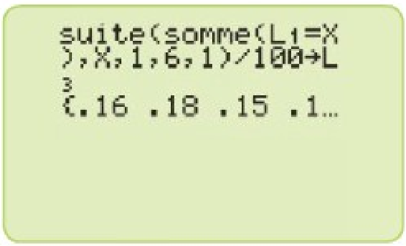
Casio	TI	Tableur
Int (6×Ran#)+1 x! nPr nCr Ran#	Deux possibilités : ent(6*NbrAléat)+1 entAléat(1,6)	=ENT(6*ALEA()) +1

2. Simuler 10 lancers d'un dé cubique équilibré et sauvegarder les résultats.
Rentrer les instructions suivantes :

Casio	TI	Tableur
Seq(Int (6×Ran#)+1,X,1,10,1)→List 1 Done List L→M Dim Fill Seq	Deux possibilités : suite(entAléat(1,6),K,1,10,1)→L1 entAléat(1,6,10)→L1	Recopier la formule =ENT(6*ALEA()) +1 dans une plage de 10 cases.

3. a. Simuler 100 lancers d'un dé cubique équilibré et déterminer la distribution des fréquences. On pourra utiliser l'une des procédures ci-dessous :

<table>
<tr><th colspan="2">Calculatrices</th><th>Tableur</th></tr>
<tr><td colspan="2">• Produire la liste des 100 nombres, la stocker en L1 (ou List 1).
• Pour la représentation de la question b, créer en L2 (ou List 2) la suite des nombres 1, 2, 3, 4, 5, 6.
• Afficher en L3 (ou List 3) la fréquence des nombres 1, 2, 3, 4, 5, 6 produits en L1. Pour cela, on crée une liste de 6 places 1, 2, 3, 4, 5, 6 et, pour chacune d'elles (X), on fait le total du nombre de tirages égaux à ce nombre : somme(L1 = X).
Ces totaux sont divisés par 100, puis stockés dans L3 (ou List 3).
Ainsi L3(2) est la fréquence de 2 dans la liste L1.</td><td rowspan="3">• Recopier la formule =ENT(6*ALEA()) +1 dans la plage A1:A100.
• Sur la plage B1:B6 créer la liste des nombres de 1 à 6.
• En C1, saisir =NB.SI($A$1:$A$100;B1)/100 puis recopier la formule jusqu'en C6.

Remarque : appuyer sur F9 pour produire un nouvel échantillon.</td></tr>
<tr><th>Casio</th><th>TI</th></tr>
<tr><td></td><td></td></tr>
</table>

b. Représenter la distribution des fréquences par un diagramme en bâtons.
Observer la représentation graphique obtenue. Que peut-on constater ?

OBJECTIF

Simuler des lancers d'une pièce et visualiser un intervalle de fluctuation.

1. a. Simuler 100 échantillons de taille 50 du lancer d'une pièce équilibrée et calculer la fréquence de Pile dans chaque échantillon. Rentrer les instructions de programme suivantes :

Casio	TI
 	 

b. Ouvrir le menu STAT et expliquer ce qui est affiché.

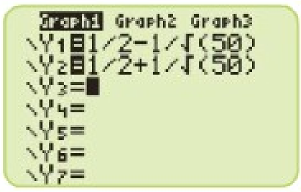
c. Construire le nuage des 100 fréquences, à l'aide des instructions suivantes :

Casio	TI
	 

d. Sur le même graphique, tracer les droites d'équation :

$$y = \frac{1}{2} - \frac{1}{\sqrt{50}} \quad \text{et} \quad y = \frac{1}{2} + \frac{1}{\sqrt{50}},$$

à l'aide des instructions ci-dessous :

Casio	TI
	

e. Dans le nuage, quel est le pourcentage de fréquences n'appartenant pas à l'intervalle :

$$\left[\frac{1}{2} - \frac{1}{\sqrt{50}} \,;\, \frac{1}{2} + \frac{1}{\sqrt{50}}\right] ?$$

2. a. Modifier le programme pour pouvoir choisir deux nombres entiers K et N, puis simuler la production de K échantillons de taille N.

b. Faire fonctionner le programme pour un choix de K ($K > 100$) et de N ($N > 50$) ; construire le nuage des K fréquences obtenues et y tracer les droites d'équation :

$$y = \frac{1}{2} - \frac{1}{\sqrt{N}} \quad \text{et} \quad y = \frac{1}{2} + \frac{1}{\sqrt{N}}.$$

c. Dans le nuage, quel est le pourcentage de fréquences n'appartenant pas à l'intervalle :

$$\left[\frac{1}{2} - \frac{1}{\sqrt{N}} \,;\, \frac{1}{2} + \frac{1}{\sqrt{N}}\right] ?$$

Rappel des notations utilisées en cours :

f est la fréquence d'un caractère mesurée lors d'un échantillonnage.
p est la proportion de ce caractère dans la population totale.

Dans les deux premiers exercices, on réalise des sondages (échantillons) et on cherche à prévoir le résultat final.

(On a besoin de l'intervalle de confiance pour ces études).

Dans le troisième exercice, on connaît p et on se demande si l'échantillon f est recevable.

(On a besoin de l'intervalle de fluctuation)

Intervalle de confiance : $\mathrm{I}_c = \left[f-\frac{1}{\sqrt{n}} ; f+\frac{1}{\sqrt{n}}\right]$

Intervalle de fluctuation : $\mathrm{I} = \left[p-\frac{1}{\sqrt{n}} ; p+\frac{1}{\sqrt{n}}\right]$

On montre : $f \in \left[p-\frac{1}{\sqrt{n}} ; p+\frac{1}{\sqrt{n}}\right]$ si et seulement si $p \in \left[f-\frac{1}{\sqrt{n}} ; f+\frac{1}{\sqrt{n}}\right]$

Commentaires :

Intervalle de confiance :

Lors d'un sondage auprès d'une partie de la population (mathématiquement, on a un échantillon), on cherche à évaluer la précision de l'estimation donnée par ce sondage.

On se donne une marge d'erreur entre les résultats du sondage et celui (inconnu) de la population totale.

Un intervalle de confiance à 95 % donnera un encadrement correct quatre-vingt-quinze fois sur cent en moyenne.

Si l'on peut répéter des estimations de même nature un grand nombre de fois, en affirmant à chaque fois que le paramètre à estimer se trouve dans cet intervalle, on se trompe en moyenne 5 fois sur cent.

f_{obs} est la fréquence observée lors du sondage.
n est la dimension de l'échantillon.
L'intervalle de confiance est : $\mathrm{I}_c = \left[f_{obs}-\frac{1}{\sqrt{n}} ; f_{obs}+\frac{1}{\sqrt{n}}\right]$.

Au risque de se tromper dans 5 % des cas, on affirme que les estimations des échantillons se trouvent dans cet intervalle I_c.

Intervalle de fluctuation :
L'intervalle de fluctuation permet de détecter un écart important par rapport à la valeur théorique p pour une estimation f établie sur un échantillon.
C'est un intervalle dans lequel la grandeur observée est censée se trouver avec une forte probabilité (souvent de l'ordre de 95 %).

48 page 280

Sondage 1 : 45 % d'opinions favorables à Mme X.

Sondage 2 : 42 % d'opinions favorables à Mme X

Les deux sondages sont effectués auprès d'un échantillon de 360 personnes.

1) I_1 est l'intervalle de confiance au seuil de 95 % correspondant au premier sondage

$\mathrm{I}_1 = [0{,}45 - \frac{1}{\sqrt{360}} ; 0{,}45 + \frac{1}{\sqrt{360}}]$ $\qquad \frac{1}{\sqrt{360}} = \frac{1}{6} \times \frac{1}{\sqrt{10}} = \frac{1}{6} \times \frac{\sqrt{10}}{10} \approx 0{,}052\,7$ par défaut

Pour donner des valeurs approchées des bornes de l'intervalle, on donne une valeur par défaut de la borne inférieure et une valeur approchée par excès de la borne supérieure afin de ne pas " perdre " des valeurs possibles.

Ici, on obtient : [0,39 ; 0,51]

2) I_2 est l'intervalle de confiance au seuil de 95 % correspondant au deuxième sondage

$I_2 = [0{,}42 - \frac{1}{\sqrt{360}} \; ; 0{,}42 + \frac{1}{\sqrt{360}}]$, soit : [0,36 ; 0,48]

3) L'intervalle I_1 est représenté par le segment [AB] et l'intervalle I_2 par l'intervalle [CD].

Leur intersection est l'intervalle $I_3 = [0{,}45 - \frac{1}{\sqrt{360}} \; ; 0{,}42 + \frac{1}{\sqrt{360}}]$ représenté par le segment [AD]

4) On ne peut pas affirmer que Mme X a une baisse de popularité ….

---

49 page 280

Au premier tour de l'élection présidentielle de 2002, L. Jospin a obtenu 16,18 % des voix et J.-M. Le Pen 16,86 %.

1) " En effet, certains des derniers sondages indiquaient 18 % pour Jospin et 14 % pour Le Pen.

Dans des conditions idéales […], on obtient sur de tels pourcentages une incertitude de plus ou moins 3 % étant donné la taille de l'échantillon "

Extrait d'un article de journal

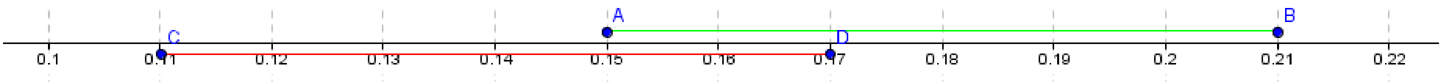
La dernière phrase donne comme intervalles de confiance :

Pour Jospin : J = [0,15 ; 0,21]

Pour Le Pen : P = [0,11 ; 0,17]

2) L'intervalle J est représenté par le segment [AB] et l'intervalle P par l'intervalle [CD].

Leur intersection est l'intervalle I = [0,15; 0,17] représenté par le segment [AD]



Il est impossible de prévoir l'ordre des candidats

3) " Pour les rares scientifiques qui savent comment sont produites les estimations, il était clair que l'écart des intentions de vote entre les candidats Le Pen et Jospin rendait tout-à-fait plausible le scénario qui s'est réalisé ".

Extrait de l'article de journal

La phrase est correcte car les résultats du vote sont dans l'intervalle I

---

51 page 280 Des simulations pour aider à une prise de décision

Données en 1979

population : garçons de moins de 15 ans Woburn	5 969
nombre de cas de leucémie sur les garçons de moins de 15 ans	9
proportion p de leucémie infantile chez les garçons aux États-Unis	0,000 52

1) simulation d'un échantillon

Soit p un nombre représentant la proportion attendue : p est un nombre entre 0 et 1

Rappel : la fonction ENT(ALEA( )+p) renvoie 0 ou 1 dans les proportions suivantes :

0 dans la proportion $1 - p$ et 1 dans la proportion p.

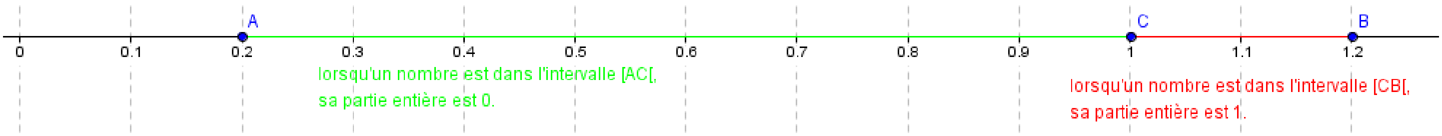
En effet :

ALEA( ) renvoie un nombre x tel que $0 \leqslant x < 1$, d'où, $p \leqslant x + p < 1 + p$.

ENT$(x + p) = 0$ lorsque $p \leqslant x + p < 1$

ENT$(x + p) = 1$ lorsque $1 \leqslant x + p < 1 + p$.

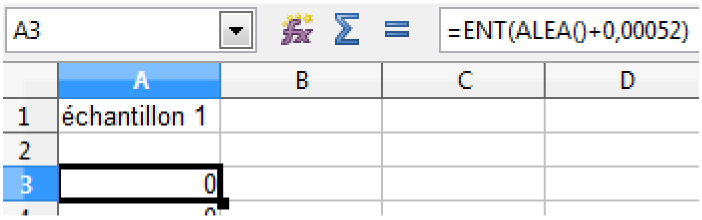
Sur cette illustration, $p = 0{,}2$



80 % des nombres ont une partie entière égale à 0,

20 % des nombres ont une partie entière égale à 1.

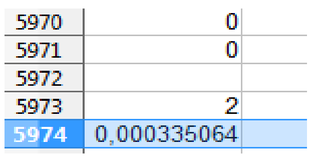
Retour à l'exercice :



Par exemple : en A3, on entre =ENT(ALEA( )+0,00052)

on sélectionne la plage A3:A5971 pour avoir 5969 cellules dans la plage, et, on recopie la formule vers le bas.

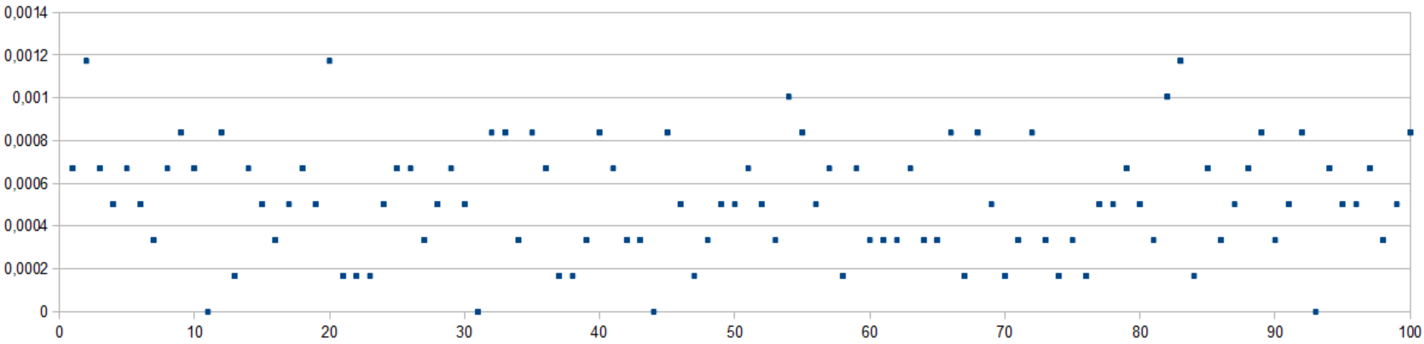
Dans la cellule A5973, on tape : =SOMME(A3:A5971) et en A5974, on tape =A5973/5969 pour avoir la fréquence



Ainsi, on voit que dans ce premier échantillon, le nombre de garçons atteint de leucémie infantile est 2 soit une fréquence égale à 0,000 33 …

2) On copie vers la droite pour reproduire les 100 échantillons.

On sélectionne la ligne (ou colonne si vous avez présenté en colonne) des 100 fréquences et on réalise le nuage de points.



3) À Woburn, la fréquence d'apparition de la leucémie infantile est $\frac{9}{5969} \approx 0{,}0015077902$

Sur la réalisation des 100 simulations précédentes, il n'y a aucune fréquence atteignant les $\frac{12}{10000}$ (cela ne veut pas dire que cela n'arrive jamais … mais la proportion sera très faible).

On peut penser que un " phénomène " lié à la ville entraîne l'augmentation de la fréquence dans cette ville … Cela peut être la qualité de l'eau.

Commentaires :

$p = 0{,}000\,52$ est connu (les services de santé notent tous les cas de maladie et la population totale est connue, donc on connaît la proportion de malades)

" Woburn " constitue un échantillon de taille $n = 5\,969$

La question qui est posée est : " cet échantillon est-il représentatif " ?

L'intervalle de fluctuation est I = $\left[0{,}00052-\frac{1}{\sqrt{5969}};0{,}00052+\frac{1}{\sqrt{5969}}\right]$ mais ***p* étant très faible,** le **modèle** proposé dans le cours **ne s'applique pas**.

La simulation de 100 échantillons de même taille permet de prévoir qu'il y a très peu de cas (aucun sur la simulation proposée mais ce n'est pas impossible pour une autre simulation) où " Woburn " est un échantillon convenable