Math93.com

TD n°1 - Seconde
Statistiques et échantillonnage

Première partie
Statistiques descriptives

Exercice 1. Comparer des séries

Lors des contrôles de mathématiques en seconde, Jean a obtenu les notes suivantes :

$$12 - 8 - 9 - 6 - 13 - 17 - 10 - 8 - 12 - 12 - 3 - 10$$

Dans le même temps, Anne a obtenu les notes suivantes :

$$9 - 13 - 10 - 8 - 8 - 10 - 11 - 10 - 11 - 9 - 11 - 10$$

1. Peut-on différencier ces deux élèves en utilisant les critères de position que sont la moyenne et la médiane ?

2. Caractéristiques de dispersion.

- **2. a.** Pour Jean, puis pour Anne, déterminer l'étendue de ses notes et l'écart interquartile.
- **2. b.** Construire les diagrammes en boites (boites à moustaches) associés aux deux séries sur un même graphique.
- **2. c.** Comparer les résultats des deux élèves à l'aide de caractéristiques de dispersion.

Réponses

(1.) $\overline{m} = M_e = 10$; *(2.a.)* $Q_1 = 8$ *et* $Q_3 = 12$, $Q'_1 = 9$ *et* $Q'_3 = 11$

Deuxième partie

Concevoir et mettre en œuvre une simulation

Génération de nombres aléatoires

Toute calculatrice et tout tableur possèdent des fonctions générant des nombres aléatoires. Il existe deux fonctions utiles : l'une donne un nombre aléatoire décimal compris entre 0 et 1, l'autre un nombre aléatoire entier compris entre deux entiers a et b.

Fonction	Calculatrice CASIO	Calculatrice TEXAS	Tableur
	OPTN, puis choisir le menu PROB, puis RAND	Dans le menu math, sélectionner PROB	Se placer dans la cellule voulue
Nombre aléatoire décimal compris entre 0 et 1	Sélectionner **Ran#** suivi de EXE	Sélectionner **1:NbrAléat**, suivi de entrer, puis entrer	Saisir =ALEA(), puis valider
Nombre aléatoire entier compris entre a et b	Sélectionner **Int** et compléter RanInt#(a,b) puis EXE	Sélectionner **5:nbrAléatEnt(1,a,b)** puis compléter entAléat(a,b) suivi de entrer	Saisir =ALEA.ENTRE.BORNES(a,b) puis valider

Simuler une expérience quand la proportion du caractère est p

On veut afficher 1 quand l'individu prélevé possède le caractère étudié, et 0 sinon.
On peut :
– soit utiliser la fonction « Partie entière » qui donne la partie entière d'un réel, combinée avec la fonction « Aléa » (la partie entière d'un nombre est le premier entier inférieur ou égal à ce nombre).

Partie entière	OPTN suivi de NUM : commande **Int**	math suivi de **NBRE**, puis commande **3:ent(**	**ENT**
Exemple : partie entière de 0,78	Int 0.78	ent(0.78)	=ENT(0,78)
Commande	Int(Ran#+p)	ent(NbrAléat+p)	=ENT(ALEA()+p)

Exemple

On cherche à modéliser un jeu de bonneteau. Un maître du jeu mélange 3 cartes, deux rois et une dame et le joueur doit trouver la dame. Il a donc 1 chance sur 3 de gagner.

La simulation d'une partie consiste à obtenir 1 (gagner) avec une probabilité $\frac{1}{3}$ et 0 avec une probabilité $\frac{2}{3}$. Il suffit d'entrer la commande :

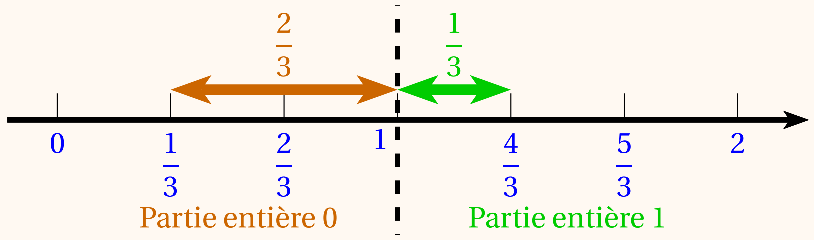
ent(NbrAléat+1/3) sur Texas ou Int(Rand#+1/3) sur Casio.

En effet :

- la fonction *NbrAléat* sur Texas ou *Rand#* sur Casio renvoie x avec $0 \leq x < 1$.
- Donc l'instruction *NbrAléat+1/3* sur Texas ou *Ran#+1/3* sur Casio renvoie $x + \frac{1}{3}$ tel que :

$$\frac{1}{3} \leq x + \frac{1}{3} < \frac{4}{3}$$

La partie entière de ce nombre est donc 0 avec une probabilité $\frac{2}{3}$ et 1 avec une probabilité $\frac{1}{3}$.



Exercice 2. Des jetons

1. Un sac contient 30% de jetons rouges et 70% de bleus. Simuler avec la calculatrice l'expérience consistant à tirer au hasard un jeton du sac et à noter sa couleur.

2. Donner un échantillon de taille 20 de cette expérience, puis la fréquence de jetons rouges de cet échantillon.

Réponses

ent(NbrAléat+0.3) sur Texas ou Int(Ran#+0.3) sur Casio

Exercice 3. Des fruits

1. Un panier contient 40% d'abricots et 60% de pêches. Simuler avec la calculatrice l'expérience consistant à tirer au hasard un fruit du panier.

2. Donner un échantillon de taille 20 de cette expérience, puis la distribution des fréquences de cet échantillon.

Réponses

ent(NbrAléat+0.4) sur Texas ou Int(Ran#+0.4) sur Casio

Exercice 4. Nombre dans un intervalle

1. Rappeler la fonction qui permet d'obtenir un nombre pseudo-aléatoire de l'intervalle $[\,0\,;\,1\,]$ sur votre calculatrice.

2. Avec cette fonction, comment peut-on obtenir un nombre pseudo-aléatoire :

2. a. de l'intervalle $[0\ ;\ 2]$?

2. b. de l'intervalle $[2\ ;\ 6]$?

Exercice 5. Un autobus

Un autobus se présente à un arrêt donné toutes les 15 minutes.

1. Simuler avec la calculatrice l'expérience consistant à noter le temps d'attente d'un usager se présentant aléatoirement à cet arrêt d'autobus.

2. Donner un échantillon de taille 20 de cette expérience (on arrondira les temps relevés à 0,1 minutes).

Réponses

NbrAléat $\times 15$ *sur Texas ou Ran#* $\times 15$ *sur Casio*

Exercice 6. Mettre en œuvre une simulation avec un tableur

On considère l'expérience aléatoire qui consiste à tirer un numéro parmi les entiers 1, 2, ... , 15 d'une loterie.

1. Simuler 100 fois cette expérience aléatoire avec un tableur.

	A	B	C	D
1	Partie	1	2	3
2	Résultat			
3	Nombre de 10			
4	Fréquence			

2. Calculer la fréquence de l'entier 10 dans cet échantillon.

Réponses

= ALEA.ENTRE.BORNES(1 ;15) et = NB.SI(B2 :CW2 ;10)

Troisième partie
Échantillonnage et intervalles de fluctuation

Propriété 1 (Intervalle de fluctuation au seuil 95% (admis))

Soit un caractère dont la proportion dans une population est p.
Pour n assez grand et p ni proche de 0, ni proche de 1, il y a environ 95% des échantillons de taille n issus de cette population qui sont tels que la fréquence f observée appartienne à l'intervalle :

$$\boxed{I = \left[p - \frac{1}{\sqrt{n}}\ ;\ p + \frac{1}{\sqrt{n}}\right]}$$

Les conditions sont : $\begin{cases} n \geq 25 \\ 0,2 \leq p \leq 0,8 \end{cases}$.

Exercice 7. Un dé et des intervalles de fluctuation

On lance n fois un dé cubique (à six faces), chaque face étant numérotée de 2 à 7. On appelle f la fréquence de sortie d'un nombre pair.

1. Calculer p, la probabilité d'obtenir un nombre pair.

2. On note I_n, l'intervalle de fluctuation asymptotique de la fréquence f au seuil 95% pour les n lancers.

2. a. Déterminer, pour $n = 50$, l'intervalle de fluctuation I_{50} (on arrondira au millième) et donner l'amplitude de l'intervalle (la différence entre les bornes) ;

2. b. Déterminer, pour $n = 500$, l'intervalle de fluctuation I_{500} (on arrondira au millième) et donner l'amplitude de l'intervalle ;

2. c. Déterminer, pour $n = 2\,500$, l'intervalle de fluctuation $I_{2\,500}$ (on arrondira au millième) et donner l'amplitude de l'intervalle ;

2. d. Déterminer n pour tel que $I_n = [0,495\ ;\ 0,505]$.

2. e. Déterminer n pour que l'amplitude de l'intervalle I_n soit de 1%.

3. On considère maintenant que $n = 2\,500$.
Sur les 2 500 lancers, on obtient :

Face	2	3	4	5	6	7	Total
Effectifs	410	425	401	425	422	417	2 500
Effectifs cumulés croissants	......	......	......	......	......	......	$\ddots$

3. a. Déterminer la médiane, la moyenne et les quartiles Q_1 et Q_3 de la série.

3. b. La fréquence f de sortie d'un nombre pair appartient-elle à l'intervalle de fluctuation $I_{2\,500}$?
Que peut-on en conclure ?

Réponses
2°) $I_{50} = [0,359\ ;\ 0,641]$, $I_{500} = [0,455\ ;\ 0,545]$, $I_{2\,500} = [0,48\ ;\ 0,52]$

Exercice 8. Surpoids

Il y a en France en 2017, environ 20 millions de personnes en surpoids (sur 67 millions de français).

1. Quelle est la proportion de personnes en surpoids en France.

2. Donner un intervalle de fluctuation au seuil de 95% de la fréquence f des personnes en surpoids dans les échantillons de 400 français.

Exercice 9. Trop de garçons ?

On sait qu'il naît en moyenne 106 garçons pour 100 filles en France.

1. Quelle est la proportion p de garçons parmi les nouveau-nés ?
2. Donner un intervalle de fluctuation au seuil de 95% de la fréquence f des garçons dans un échantillon de 100 nouveau-nés.
3. Même question avec un échantillon de 500 nouveau-nés.

Réponses
1°) $p \approx 51\%$ *2°)* $f \in [41\% ; 61\%]$ *3°)* $f \in [46\% ; 56\%]$

Exercice 10. Problème de sacs

La chaîne d'ensachage du sucre X produit des sacs de masse nominale 1 kg. Suite à un dérèglement, on observe que 26% des paquets produits ont une masse strictement inférieure à 1 kg. De tels paquets sont dits non conformes.
Un boulanger achète 50 sacs de sucre. Est-il possible, au seuil de 95%, qu'il trouve 19 sacs non conformes.

Réponses
$f \in [11,8\% ; 40,2\%]$ *et* N_c *le nombre de sacs non conformes appartient donc à l'intervalle*
$11,8\% \times 50 = 5,9 \leq N_c \leq 20,1 = 40,2\% \times 50$

Quatrième partie

Prise de décision à partir d'un échantillon

Test

On fait l'hypothèse que la proportion d'un caractère dans une population est p.
A partir d'un échantillon pour lequel la fréquence observée relative à un caractère donné est f, on regarde si cette fréquence est dans l'intervalle du fluctuation au seuil 95% : sinon, on considère que l'observation n'est pas compatible avec l'hypothèse faite.
En effet, avec cette hypothèse, ceci n'aurait lieu que dans environ 5% des échantillons de tailles n.
On prendra ainsi une décision avec un seuil de risque de 5% (ou un coefficient de confiance de 95%).

Exercice 11. Prise de décision à partir d'un échantillon

On sait que 26% de la population est allergique aux pollens de fleurs. On veut contrôler si cette proportion est valable dans un département donné. Pour cela on prélève un échantillon de 400 personnes de ce département.

1. Déterminer l'intervalle de fluctuation au seuil 95% de la fréquence des personnes allergiques aux pollens de fleurs dans les échantillons de taille 400.
2. Énoncer la règle de décision permettant d'accepter, ou non, l'hypothèse selon laquelle la proportion p de personnes allergiques dans le département est 0,26.
3. On a observé 131 personnes allergiques dans l'échantillon. Que peut-on en conclure ?

Réponses
(1.) $I = [0,24 ; 0,31]$ - *(3.)* $f = 0,3275 \notin I$

Exercice 12. Les anglais ont les yeux bleus ou verts ?

On estime que 60% des Anglais on les yeux de couleur bleue ou verte. On veut vérifier cette proportion pour une ville anglaise. On prélève un échantillon de 100 personnes de cette ville.

1. Déterminer l'intervalle de fluctuation au seuil 95% de la fréquence des Anglais ayant les yeux de couleur bleue ou verte dans les échantillons de taille 100.
2. Énoncer la règle de décision permettant d'accepter, ou non, l'hypothèse selon laquelle la proportion p de personnes ayant les yeux de couleur bleue ou verte dans cette ville est 0,6.
3. On a observé 51 personnes ayant les yeux de couleur bleue ou verte dans l'échantillon. Que conclure ?

Cinquième partie

Complément : Intervalles de confiance

Exercice 13. Problème de sondage et intervalle de confiance

Un peu d'histoire

Lors de l'élection présidentielle américaine de 1936, la revue Literary Digest procède, à partir de l'annuaire téléphonique à un « vote de paille » , auprès de 10 millions de personnes : Roosevelt est donné perdant. Au contraire, l'institut Gallup, à partir d'un échantillon représentatif, prédit l'élection de Roosevelt avec 56% *des voix. Roosevelt obtiendra 62 % des voix et la méthode Gallup est consacrée.*

Le premier sondage réalisé en France l'a été par l'entreprise IFOP en 1938 à propos des accords de Munich. À la question « Approuvez-vous les accords de Munich ? » on voit une opinion publique beaucoup plus tiède (à 57 % de oui) que ses parlementaires (87,5 % de oui, 535 voix contre 75 lors d'un vote de la chambre des députés).

1. **Quelques remarques utiles :**

   1. a. Montrer que si les conditions sont remplies, en appliquant la propriété :

$$f \in \left[p - \frac{1}{\sqrt{n}} ~;~ p + \frac{1}{\sqrt{n}}\right] \Longleftrightarrow p \in \left[f - \frac{1}{\sqrt{n}} ~;~ f + \frac{1}{\sqrt{n}}\right]$$

Remarque

- L'intervalle $\left[p - \frac{1}{\sqrt{n}} ~;~ p + \frac{1}{\sqrt{n}}\right]$ est nommé **intervalle de fluctuation** (asymptotique au seuil 95% de la fréquence f
- et l'intervalle $\left[f - \frac{1}{\sqrt{n}} ~;~ f + \frac{1}{\sqrt{n}}\right]$ est nommé **intervalle de confiance** de la proportion p (au seuil 95%).

   1. b. Donner l'amplitude de l'intervalle de fluctuation en fonction de n.

2. **Exemple : Un sondage**

   Lors d'une élection, un candidat fait réaliser un sondage pour savoir s'il obtiendra au moins 50% des voix.
   L'institut de sondage *Hypessausse* lui annonce que, sur un échantillon de 300 personnes, 52% déclarent vouloir voter pour lui.

   2. a. Montrer que l'intervalle de confiance au seuil 95% de la proportion p de personnes votant pour ce candidat, à partir des données recueillis par l'institut de sondage, est $[46\% ~;~ 58\%]$, c'est à dire que p a 95% de chance d'y appartenir. Donner l'amplitude de l'intervalle.
   Que peut-on en conclure ?

   2. b. Le candidat, peu satisfait de ce sondage, en redemande un autre. L'institut de sondage *Hypessausse* lui annonce que, sur un échantillon de 5 000 personnes cette fois, 52,5% déclarent vouloir voter pour lui.
   Montrer que l'intervalle de confiance de p est $[51{,}08\% ~;~ 53{,}92\%]$. Donner l'amplitude de l'intervalle. Que peut-on en conclure ?

   2. c. Déterminer la taille d'un échantillon permettant d'avoir un intervalle de confiance au seuil de 95% d'amplitude 1%.

Réponses
1b) *L'amplitude est de* $\frac{2}{\sqrt{n}}$ ***2c)*** $n = 40\,000$

Exercice 14. Fumer tue

On souhaite connaître la proportion p de personnes qui fument dans la population française. Pour cela, on réalise un sondage sur un échantillon représentatif de taille 1 000 personnes, et on constate que la proportion observée dans cet échantillon est $f = 40\%$.
Déterminer un intervalle de confiance au seuil 95% de la proportion de fumeurs dans la population française.

Réponses
$p \in [36{,}8\% ~;~ 43{,}2\%]$

Exercice 15. Les fameuses parts d'audience

Le Médiamat est la mesure de référence de l'audience de la télévision en France. Au 9 décembre 2013, le panel Médiamat était composé de 11 545 individus âgés de 4 ans et plus vivant dans 5 003 foyers équipés d'un audimètre à bouton-poussoir. Ce panel est représentatif des individus résidant en France métropolitaine et possédant la télévision dans leur résidence principale. Les résultats d'audience des chaînes comprennent les modes de réception de la télévision par le public en hertzien numérique (TNT), par câble analogique et numérique, ainsi que par satellite et ADSL.
Le 18 Décembre 2013, à 20h50, une chaîne de télévision française a enregistré une part d'audience de $41,1\%$.
Déterminer un intervalle de confiance au seuil 95% de la proportion p de personnes (correspondant aux critères requis) ayant regardé cette émission dans la population française.

Réponses
$p \in [40,17\% \ ; \ 42,03\%]$

Exercice 16. Les élections

Deux candidats A et B sont en concurrence lors d'une élection. Il faut obtenir au moins 50% des voix pour être élu. un sondage effectué sur un échantillon de 900 électeurs donne 51% des voix au candidats A. Celui-ci affirme :« J'ai donc 95% de chances d'être élu » .
Que peut-on dire de cette affirmation ?