Statistiques:échantillonnage

Copyright ® meilleurenmaths.com. Tous droits réservés

1. Introduction

L'échantillonnage est utilisé en statistiques lorsqu'il est impossible d'observer un caractère sur l'ensemble des individus d'une population (lorsque par exemple l'effectif total est très grand).

Pour « estimer » une proportion d'un caractère d'une population, on peut effectuer des « sondages » parmi des échantillons restreints de cette population et déterminer la fréquence du caractère de ces échantillons. Dans certains cas on peut effectuer un nombre important de simulations.

2. Expérience aléatoire

Une expérience aléatoire est une expérience admettant plusieurs résultats (ou issues) possibles, dont on ne peut pas prévoir avec certitude le résultat (ou l'issue) obtenu(e).

Exemples:

- lancer d'une pièce de monnaie, on peut obtenir: pile ou face.
- Lancer d'un dé cubique numéroté de 1 à 6, on peut obtenir: 1; 2; 3; 4; 5; 6.

3. Échantillons

Lorsque l'on répète n fois de manière indépendante une même expérience aléatoire, on obtient un échantillon de taille n de cette expérience.

4. Fluctuation d'échantillonnage

p est la proportion (inconnu) dans une population. On considère plusieurs échantillons de taille n de la population.

Pour chaque échantillon, on calcule la fréquence f du caractère étudié.

On remarque que les fréquences obtenues pour le caractère ne sont pas identiques. On dit que l'on a: « une fluctuation d'échantillonnage ».

Lorsque l'on augmente la taille des échantillons, les fréquences du caractère fluctuent toujours mais les écarts sont moindres.

5. Simulations

Lorsque l'on peut modéliser une expérience aléatoire par tirage de nombres au hasard, on peut effectuer des simulations de l'expérience en utilisant soit une calculatrice soit un tableur.

Exemples:

- pile ou face: on tire au hasard un nombre parmi les chiffres 0 et 1.
- lancé d'un dé cubique: on tire au hasard un nombre parmi les chiffres 1; 2; 3; 4; 5 et 6

6. Remarque

Lorsque l'on réalise un sondage dans la population française, on choisit un échantillon de taille (au moins) 1000 qui doit être représentatif de la population pour le caractère étudié.

7. Intervalle de fluctuation d'une fréquence au seuil de 95%

On admet le résultat suivant:

Copyright ® meilleurenmaths.com. Tous droits réservés

Si p la population du caractère est comprise entre 0,2 et 0,8 alors la fréquence f du caractère dans un échantillon de taille $n \geq 25$ appartient à l'intervalle de fluctuation $[p-\frac{1}{\sqrt{n}} ; p+\frac{1}{\sqrt{n}}]$ avec une probabilité d'au moins 0,95.

<u>Exemple:</u> jeu de pile ou face

Caractère étudié: pile

p=0,5 (exemple étudié en classe de 3ième)

n=100 $\sqrt{n}=10$

$p-\frac{1}{\sqrt{n}}=0,4$ $p+\frac{1}{\sqrt{n}}=0,6$

Pour n lancers de la pièce la fréquence f pile est comprise entre 0,4 et 0,6 avec une probabilité d'au moins 0,95.

<u>Attention:</u> Dans 100 lancers, il n'est pas impossible d'obtenir 100 fois « face ».

8. Estimation de p par f

Dire que $f \in [p-\frac{1}{\sqrt{n}} ; p+\frac{1}{\sqrt{n}}]$ c'est dire que $p-\frac{1}{\sqrt{n}} \leq f \leq p+\frac{1}{\sqrt{n}}$

$p-\frac{1}{\sqrt{n}} \leq f$ donne $p \leq f+\frac{1}{\sqrt{n}}$

$f \leq p+\frac{1}{\sqrt{n}}$ donne $f-\frac{1}{\sqrt{n}} \leq p$

On a donc: $f-\frac{1}{\sqrt{n}} \leq p \leq f+\frac{1}{\sqrt{n}}$ avec une probabilité d'au moins 95%.

(attention, l'estimation proposée n'est pas une certitude)

Lorsque cela est possible, on augmente n et on effectue plusieurs échantillon. Ceci est possible lorsque l'on peut réaliser facilement des simulations.

9. Utilisation d'un tableur

a) jeu de pile ou face

p est la probabilité de pile p=0,5

q est la probabilité de face q=0,5

On modélise cette expérience aléatoire par le choix au hasard d'un nombre parmi les deux chiffres 0 et 1. (0 représente « face » et 1 « pile »)

On réalise la simulation de 100 épreuves répétées. On détermine le nombre de « pile » (donc aussi le nombre de « face »)

f est le nombre précédent divisé par 100.

Puis on convient de refaire 9 fois les 100 épreuves et on regardera la fluctuation de la fréquence.

(On peut si l'on veut choisir n=1000 ou 10000 et réitérer 99 ou 999 fois)

Il existe une instruction dans un tableur qui permet de choisir des nombres entiers parmi une liste de nombres entiers consécutifs. Cette instruction est: « alea.entre.bornes(0;1) » (0 représente le nombre minimum et 1

Copyright ® meilleurenmaths.com. Tous droits réservés

représente le nombre maximum)

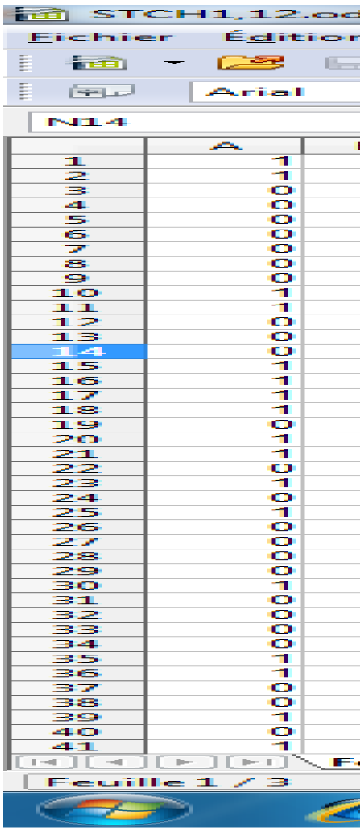
Dans le tableur,

Dans la cellule A1, on tape: « =alea.entre.bornes(0;1) », on appuie sur entrée. On étire cette formule jusqu'à J1, puis jusque J100.

On obtient un tableau 100×10 constitués de 0 et de 1.

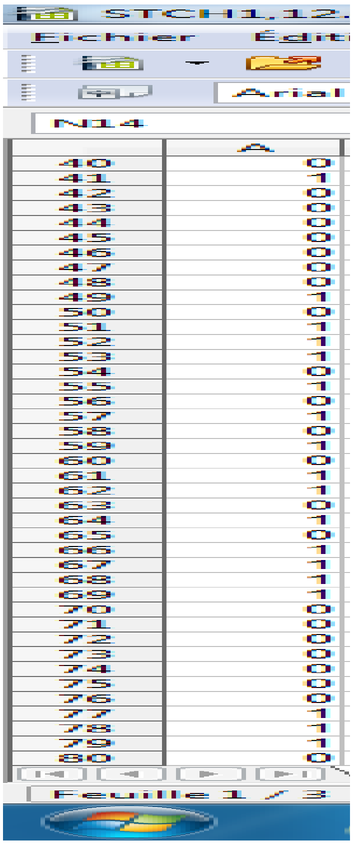
Pour obtenir le nombre de « pile » dans la 1ère colonne, dans A101, on écrit: « =somme(A1:A100), on appuie sur entrée et on étire cette formule pour les 10 colonnes c'est à dire jusque J101.

Pour obtenir le nombre de « face »dans la 1ère colonne, dans A102, on écrit: « =100-A101 », on appuie sur entrée et on étire cette formule pour les 10 colonnes c'est à dire jusque J102.


STCH1,12.od
Fichier Édition
Arial
N14

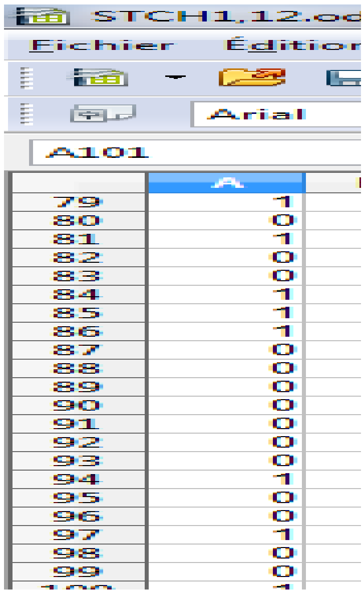
	A
1	1
2	1
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	1
11	1
12	0
13	0
14	0
15	1
16	1
17	1
18	1
19	0
20	1
21	1
22	0
23	1
24	0
25	1
26	0
27	0
28	0
29	0
30	1
31	0
32	0
33	0
34	0
35	1
36	1
37	0
38	0
39	1
40	0
41	1

Feuille 1 / 3


STCH1,12.
Fichier Édit
Arial
N14

	A
40	0
41	1
42	0
43	0
44	0
45	0
46	0
47	0
48	0
49	1
50	0
51	1
52	1
53	1
54	0
55	1
56	0
57	1
58	0
59	1
60	0
61	1
62	1
63	0
64	1
65	0
66	1
67	1
68	1
69	1
70	0
71	0
72	0
73	0
74	0
75	0
76	0
77	1
78	1
79	1
80	0

Feuille 1 / 3


STCH1,12.od
Fichier Édition
Arial
A101

	A
79	1
80	0
81	1
82	0
83	0
84	1
85	1
86	1
87	0
88	0
89	0
90	0
91	0
92	0
93	0
94	1
95	0
96	0
97	1
98	0
99	0

Les fréquences obtenues (pour « pile ») sont:

0,43; 0,43; 0,51; 0,55; 0,56; 0,45; 0,46; 0,53; 0,47; 0,56

La dernière ligne permet de calculer les fréquences de face.

(remarque: si vous effectuez un tableau, il ne sera pas identique au nôtre car c'est de l'aléatoire.

b) jeu de dé

La probabilité pour obtenir 1 est $\frac{1}{6} \approx 0,166....$

Les autres faces ont aussi la même probabilité.

On modélise l'expérience aléatoire par le choix au hasard d'un nombre parmi les six chiffres: 1; 2; 3; 4; 5; 6.

On réalise la simulation de nouveau 10 fois cent épreuves répétées.

On tape dans A1: « alea.entre.bornes(1;6) », on appuie sur entrée et on étire cette formule de A1 à J1 puis de J1 à J100.

Copyright ® meilleurenmaths.com. Tous droits réservés

On obtient un tableau 100×10 constitués de 1; 2; 3; 4; 5 et 6.

Pour compter le nombre de 1 dans la 1ière centaine d'épreuves aléatoires:

Dans A101, on écrit: « =nb.si(A1:A100;1), on appuie sur entrée et on étire cette formule jusqu'à J101.

De même, dans A102, on écrit: « =nb.si(A1:A100;2), on appuie sur entrée et on étire cette formule jusqu'à J102.

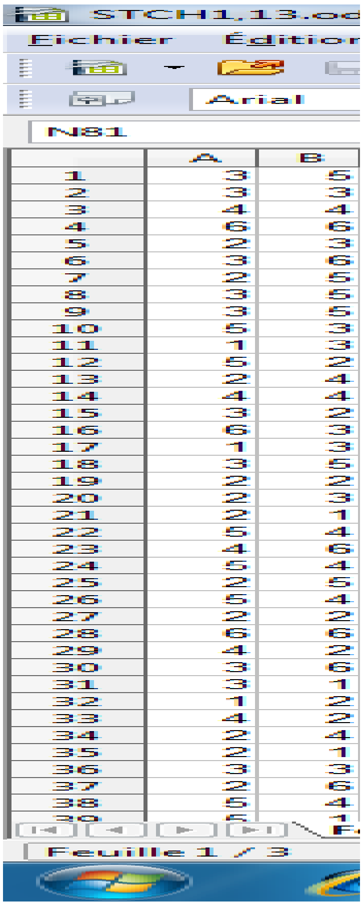
Dans A103, on écrit: « =nb.si(A1:A100;3), on appuie sur entrée et on étire cette formule jusqu'à J103.

Dans A104, on écrit: « =nb.si(A1:A100;4), on appuie sur entrée et on étire cette formule jusqu'à J104.

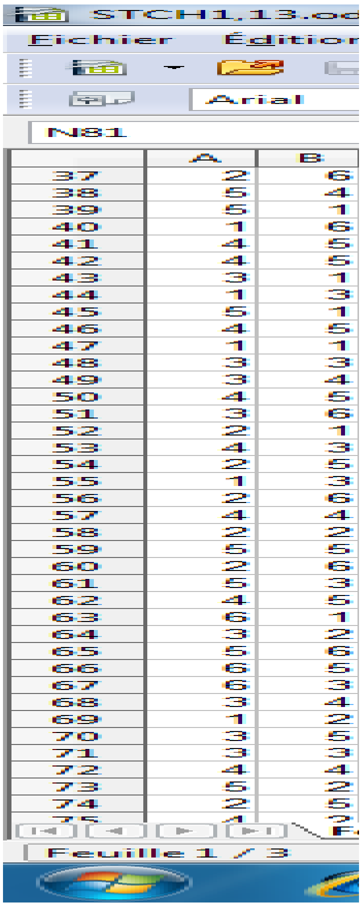
dans A105, on écrit: « =nb.si(A1:A100;5), on appuie sur entrée et on étire cette formule jusqu'à J105.

dans A106, on écrit: « =nb.si(A1:A100;6), on appuie sur entrée et on étire cette formule jusqu'à J106.

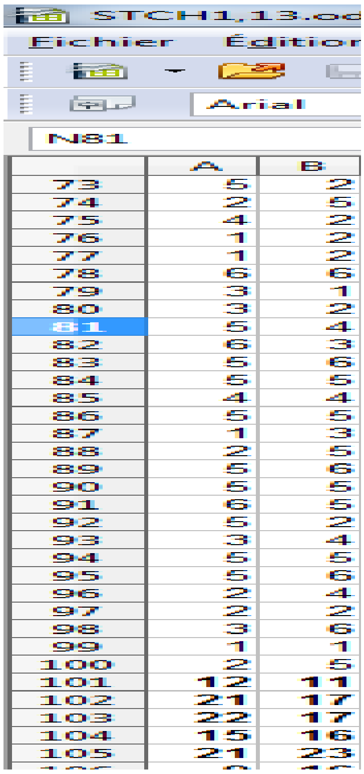
On obtient:



	A	B
1	3	5
2	3	3
3	4	4
4	6	6
5	2	3
6	3	6
7	2	5
8	3	5
9	3	5
10	5	3
11	1	3
12	5	2
13	2	4
14	4	4
15	3	2
16	6	3
17	1	3
18	3	5
19	2	2
20	2	3
21	2	1
22	5	4
23	4	6
24	5	4
25	2	5
26	5	4
27	2	2
28	6	6
29	4	2
30	3	6
31	3	1
32	1	2
33	4	2
34	2	4
35	2	1
36	3	3
37	2	6
38	5	4



	A	B
37	2	6
38	5	4
39	5	1
40	1	6
41	4	5
42	4	5
43	3	1
44	1	3
45	5	1
46	4	5
47	1	1
48	3	3
49	3	4
50	4	5
51	3	6
52	2	1
53	4	3
54	2	5
55	1	3
56	2	6
57	4	4
58	2	2
59	5	5
60	2	6
61	5	3
62	4	5
63	6	1
64	3	2
65	5	6
66	6	5
67	6	3
68	3	4
69	1	2
70	3	5
71	3	3
72	4	4
73	5	2
74	2	5



	A	B
73	5	2
74	2	5
75	4	2
76	1	2
77	1	2
78	6	6
79	3	1
80	3	2
81	5	4
82	6	3
83	5	6
84	5	5
85	4	4
86	5	5
87	1	3
88	2	5
89	5	6
90	5	5
91	6	5
92	5	2
93	3	4
94	5	5
95	5	6
96	2	4
97	2	2
98	3	6
99	1	1
100	2	5
101	12	11
102	21	17
103	22	17
104	15	16
105	21	23

Les fréquences (en 100 épreuves) obtenues pour 1 sont:

0,2; 0,19; 0,23; 0,15; 0,23; 0,11; 0,11; 0,15; 0,24; 0,13.

Les lignes suivantes permettent de calculer les fréquences de 2; de 3; de 4; de 5 et de 6.

Copyright ® meilleurenmaths.com. Tous droits réservés