

Fluctuation d'échantillonnage

Dans une urne se trouvent, 5 boules blanches, 3 boules jaunes et 2 boules noires, elles sont indiscernables au toucher. On choisit une boule de l'urne, on veut savoir si elle est jaune ou non.

Un échantillon de taille 100 de cette expérience est la liste de 100 résultats (avec 2 issues pour chacun : obtenir une boule jaune ou non) obtenus en tirant avec remise une boule 100 fois de l'urne.

Simulation d'un tel tirage

La proportion des boules jaunes dans l'urne est $p = \frac{3}{10} = 0,3$ et celle des boules non jaunes est 0,7.

$x = \text{Random} + 0,3$ correspond au tirage d'un nombre de l'intervalle $[0,3 ; 1,3[$.

$$\text{Int}(\text{Random} + 0,3) = \begin{cases} 0 & \text{si } x \in [0,3 ; 1[, \text{ l'amplitude est } 0,7 ; \text{ ce qui correspond à une boule non jaune} \\ 1 & \text{si } x \in [1 ; 1,3[, \text{ l'amplitude est } 0,3 ; \text{ ce qui correspond à une boule jaune} \end{cases}$$

Remarque : Si on détermine avec la simulation précédente, un échantillon de taille n , la fréquence de sortie de la boule jaune s'obtient en divisant la somme des résultats par n .

Activité :

A l'aide du tableur Open Office, déterminons 500 échantillons de taille 400, de cette expérience, et analysons la fréquence de sortie de la boule jaune pour chaque échantillon.

Définition :

$p = 0,3$ est la proportion des boules jaunes dans l'urne, $n = 400$ est la taille d'un échantillon,

$$p - \frac{1}{\sqrt{n}} = 0,3 - 0,05 = 0,25 ;$$

$$p + \frac{1}{\sqrt{n}} = 0,3 + 0,05 = 0,35.$$

$$\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right] = [0,25 ; 0,35] \text{ est appelé l'intervalle de fluctuation.}$$

1. Ouvrir le tableur Open Office.
2. Dans A1, taper « Pourcentage des fréquences appartenant à l'intervalle de fluctuation : ». Sélectionner la plage A1 :E1, puis cliquer sur l'icône *Fusionner et centrer les cellules*.
3. Dans A2, taper « Fréquence maximale : ». Dans A3, taper « Fréquence minimale : ».
4. Dans B5, taper « Echant. 1 ».
5. Dans A6, taper « Fréquence des jaunes », valider par *Entrée*, cliquer à nouveau sur cette cellule, puis sur *Format, Cellules, Alignement*, cocher la case *Renvoi à la ligne automatique*, puis sur Ok.
6. Dans A7, taper « Tirage 1 », valider par *Entrée*, cliquer à nouveau sur la cellule A7, utiliser la poignée de recopie jusqu'à A406, (saisir avec le bouton gauche, la croix, en bas à droite, faire glisser jusqu'à la cellule A406, puis relacher ce bouton).
7. Dans B7, taper $=ENT(ALEA()+0,3)$, valider par *Entrée*, puis cliquer sur cette cellule, utiliser la poignée de recopie pour recopier cette formule jusqu'en B406, on obtient alors un échantillon de taille 400 de l'expérience.
8. Dans B6, taper $=SOMME(B7:B406)/400$, valider par *Entrée*, c'est la fréquence de sortie de la boule jaune dans l'échantillon de taille 400.
9. Sélectionner la plage B5 :B406, cliquer sur la croix, en bas à droite de la cellule B406, et faire glisser jusqu'en SG406, on obtient alors 500 échantillons de taille 400 de l'expérience. On peut remarquer que la fréquence de sortie de la boule jaune varie d'un de ces échantillons à un autre de même taille, c'est ce qu'on appelle la fluctuation d'échantillonnage.
10. Dans C2, taper $=MAX(B6:SG6)$, c'est la plus grande fréquence de la plage B6 :SG6.
11. Dans C3, taper $=MIN(B6:SG6)$, c'est la plus petite fréquence de la plage B6 :SG6.
12. Dans F1, taper $=(NB.SI(B6:SG6;">=0,25")-NB.SI(B6:SG6;">0,35"))/5$, on obtient le pourcentage de fréquences de la plage B6 :SG6 appartenant à l'intervalle de fluctuation : $[0,25 ; 0,35]$.
A la quantité des fréquences de la plage B6 :SG6 supérieures ou égales à 0,25, on enlève les fréquences de la plage supérieures à 0,35. La commande *NB.SI(plage;critère)* indique le nombre de cellules répondant au critère indiqué dans la plage de cellules donnée.
13. Sélectionner la plage B6 :SG6, puis revenir au début de la feuille de calcul afin de pouvoir visualiser la cellule F1, dans la barre des menus, cliquer sur *Insertion*, puis sur *Diagramme*, Comme type de diagramme, choisir *Ligne*, puis *points seuls*, cliquer sur *terminer*. On obtient la répartition graphique des 500 fréquences. (Pour changer la taille des points, on pourra cliquer sur ceux-ci avec le bouton droit de la souris, puis sur *Formater les séries de données*, mettre les valeurs 0,05 pour la largeur et la hauteur.)
14. Pour reproduire la simulation, faire F9 plusieurs fois de suite et vérifier que dans chaque cas, au moins 95 % des fréquences se situent dans l'intervalle de fluctuation.