2^{nd}

Echantillonnage

__Objectifs__ *: Notion d'échantillon.*
Intervalle de fluctuation au seuil de 95% pour la proportion d'un caractère dans une population.
Réalisation d'une simulation :concevoir, mettre en œuvre et exploiter des simulations concrètes à l'aide de tableur ou calculatrice.

1) Fluctuation d'échantillonnage

Définition : Un échantillon de taille n est la série statistique formée des n résultats obtenus lorsqu'on répète n fois, de manière indépendante, une expérience dans les mêmes conditions.
Exemple : On lance un dé cubique et l'on note le numéro obtenu sur la face supérieure du dé. Si l'on répète dix fois cette expérience, en supposant qu'on ait obtenu successivement 1, 5, 2, 1, 6, 3, 2, 2, 6, 2, cette série de nombre constitue un échantillon de taille 10.

Définition : la distribution des fréquences associées à un échantillon est la liste des fréquences des issues de l'échantillon.
__Exercice 1__ : Déterminer la distribution des fréquences pour l'exemple ci-dessus.

Remarque : Les distributions de fréquences varient d'un échantillon à l'autre pour une même expérience. Ce phénomène est appelé **Fluctuation d'échantillonnage**.
Les fréquences de chaque valeur d'un caractère sont différentes si l'on simule 10 ; 20 ; 50 ;100 ; 1000 ; 10 000 ou 100 000 000 fois l'expérience.
Lorsque la taille *n* de l'échantillon augmente, l'ampleur des fluctuations des distributions de fréquence calculées sur ces échantillons de taille *n* diminue et les fréquences tendent à se stabiliser.
Pour obtenir des résultats les plus fiables possibles, il faut simuler un très grand nombre de fois l'expérience.

__Exercice 2__ : Simuler à l'aide de votre calculatrice un second échantillon de taille 10 du lancé de dé et déterminer la distribution des fréquences. Comparer vos résultats avec vos voisins.

2) Simuler une expérience, c'est remplacer une expérience par une autre (à l'aide d'une calculatrice ou d'un ordinateur) qui conduit à des résultats analogues.

a) Utilisation de la touche **RANDOM** de votre calculatrice. (Rand ou Ran# ou NbrAleat)
Cette fonction permet d'obtenir au hasard un nombre décimal X tel que $0 \leq X < 1$
La commande : intg(6 Rand + 1) permet d'obtenir au hasard un nombre entier entre 1 et 6 ; donc simule le lancer d'un dé cubique. (intg = fonction partie entière)

b) Sur le tableur, on utilisera les commandes =ALEA() ou =ENT ou =ALEA.ENTRE.BORNES(min;max).

c) Utiliser un algorithme et construire un programme avec une boucle.

__Exercice 3__ : Le lancer d'une pièce de monnaie conduit à deux issues PILE ou FACE.
Simuler un échantillon de taille 20 de cette expérience, avec votre calculatrice, ou en créant un algorithme ou en utilisant un tableur.

Remarque : Lorsqu'une expérience n'a que **deux issues possibles** (0 ou 1), on dit que cette expérience relève du **modèle de Bernoulli**.

3) Intervalle de fluctuation d'une fréquence au seuil de 95 %

Propriété admise : Soit un caractère dont la proportion dans la population donnée est *p*.
Lorsque *n* ≥ 25 et 0,2 ≤ *p* ≤ 0,8, il y a 95% des échantillons de taille *n* issus de cette population qui sont tels que la fréquence *f* du caractère dans l'échantillon appartient à un intervalle centré en *p* de la forme $\left[p-\frac{1}{\sqrt{n}}\,;p+\frac{1}{\sqrt{n}}\right]$.

Remarque : Cet intervalle $\left[p-\frac{1}{\sqrt{n}}\,;p+\frac{1}{\sqrt{n}}\right]$ s'appelle intervalle de fluctuation au seuil de 95 %, ou au risque de 5%.

Application : Prise de décision à partir d'un échantillon.

Pour apprécier si une fréquence observée *f* sur un échantillon de taille *n* est compatible ou non avec un modèle de Bernoulli de probabilité *p*, on teste l'appartenance de *f* à l'intervalle de fluctuation $\left[p-\frac{1}{\sqrt{n}}\,;p+\frac{1}{\sqrt{n}}\right]$ **au seuil de 95%.**

- Si f n'est pas dans l'intervalle de fluctuation, alors on peut rejeter l'hypothèse que l'échantillon soit compatible avec le modèle.
- Si f est dans l'intervalle de fluctuation, alors on ne peut pas rejeter l'hypothèse que l'échantillon soit compatible avec le modèle.

Quelle que soit la décision prise, il y a toujours le risque que ce ne soit pas la bonne décision dans 5% des cas.

Exercice 4 : 26% des français se déclarent allergiques aux pollens de fleurs. On étudie la fréquence *f* des personnes allergiques dans un échantillon de 400 personnes : on trouve 120 personnes allergiques.

a) Au seuil de confiance de 95 %, entre quelles valeurs varie *f* ?
b) Le nombre d'allergiques dans cet échantillon est-il anormal ?

Application : Estimer une proportion inconnue à partir d'échantillon.

$$f\in\left[p-\frac{1}{\sqrt{n}}\,;p+\frac{1}{\sqrt{n}}\right] \Leftrightarrow p-\frac{1}{\sqrt{n}}\le f\le p+\frac{1}{\sqrt{n}} \Leftrightarrow -\frac{1}{\sqrt{n}}\le f-p\le \frac{1}{\sqrt{n}}$$

$$-\frac{1}{\sqrt{n}}\le f-p\le \frac{1}{\sqrt{n}} \Leftrightarrow -\frac{1}{\sqrt{n}}\le p-f\le \frac{1}{\sqrt{n}} \Leftrightarrow f-\frac{1}{\sqrt{n}}\le p\le f+\frac{1}{\sqrt{n}}$$ On obtient $p\in\left[f-\frac{1}{\sqrt{n}}\,;f+\frac{1}{\sqrt{n}}\right]$

A partir d'une fréquence observée *f* dans un échantillon de grande taille *n*, on peut estimer la valeur d'une proportion *p* inconnue dans la population à l'aide de l'intervalle $\left[f-\frac{1}{\sqrt{n}}\,;f+\frac{1}{\sqrt{n}}\right]$

Exercice 5 : Lors d'une élection, un sondage portant sur 1000 personnes donne 400 votants pour le candidat A. Avec un risque d'erreur de 5% ; quelles informations peut-on obtenir sur la proportion de votants pour ce candidat dans la population ?

Exercice 6 : Un agriculteur désire connaître l'effectif de la population de cerfs sur son exploitation. Il ne sait pas trop comment procéder…
Il fait appel à un statisticien, celui-ci lui propose la démarche suivante :
Pendant un week-end, inviter vos amis à chasser mais avec des paint-balls (les balles ne tuent pas mais marquent les bêtes d'une couleur). Et surtout compter bien vos tirs réussis.
L'agriculteur appelle le statisticien quelque temps plus tard et lui annonce un total de 800 tirs réussis.
Le statisticien lui demande maintenant de se promener chez lui et de compter les cerfs qu'il voit et surtout de noter le nombre de cerfs colorés. La réponse arrive bientôt : 250 cerfs colorés sur 1000 vus.
Quelle est la population de cerfs sur la propriété ?