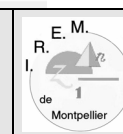
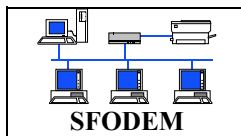


ECHANTILLONNAGES

Fiche de repérage

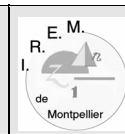


Type :	Activité d'approche de la notion de fluctuation d'échantillonnage et d'intervalle de confiance à travers quelques simulations.
Niveau :	Lycée : Seconde et Première (le fichier EXCEL pourra servir à tous les niveaux)
Mots-clés :	Probabilité théorique – Probabilité empirique – Population – Fluctuation - Echantillonnage – Tirages avec et sans remise – Schéma de Bernoulli. Intervalle de confiance.
Objectifs pédagogiques généraux :	<ul style="list-style-type: none"> • L'objectif principal est l'observation des fluctuations d'échantillonnages dans la perspective de l'approximation d'une probabilité théorique. • Un objectif plus lointain est l'approche des notions d'intervalle de confiance ou de fourchette, de taille des échantillons et de degré de confiance. • On fera pour le moins réfléchir à l'influence de la taille de l'échantillon sur les fluctuations.
Modalité :	Travail dirigé à faire en rétroprojection ou en salle d'informatique dans le cadre des modules de seconde ou de première. Le fichier echant2.xls sera installé sur chacun des postes de travail.
Dispositif technique :	<ul style="list-style-type: none"> • Salle informatique équipée Windows 98 (2000, XP) et d'EXCEL. Le classeur « sondages.xls » devra être installé sur tous les postes. • Ordinateur portable et système de vidéoprojection.
Liste et description des fichiers :	<ul style="list-style-type: none"> • Un document WORD « echant.doc » comprenant entre autres le document professeur et la fiche élève. • Un classeur EXCEL « sondages.xls »
Description activité:	<ul style="list-style-type: none"> • La population (de 2000 à 20 000 individus représentés par des cellules de deux couleurs vert-ocre) est tirée aléatoirement au départ (la probabilité théorique du vert est p). Des curseurs permettent d'agir sur la taille N de la population et sur la fréquence théorique p du caractère. • Un troisième curseur permet de choisir la taille n des échantillons (entre 100 et 1500) • Deux types d'échantillons sont prévus : <ol style="list-style-type: none"> 1. Un échantillon aléatoire de n individus (avec remise) 2. Les n individus de l'échantillon sont tirés aléatoirement sans remise (plage encadrée en rouge). • Tous les effectifs sont affichés en haut de la feuille et on renouvellera les tirages aléatoires avec la touche F9. • Des nuages dissimulent l'effectif de la population et la fréquence théorique afin de rester dans une situation expérimentale.



ECHANTILLONNAGES

Fiche Professeur



**Programme officiel
(Seconde-Première) :**

Contenus (2de) : Fréquence d'un événement. Simulation et fluctuation d'échantillonnage

Capacités attendues : Concevoir et mettre en œuvre des simulations simples.

Modalités de mise en œuvre (1èreES) : L'utilisation d'un logiciel informatique est indispensable pour accéder à une simulation sur un nombre important d'expériences. On observera dynamiquement et en temps réel, les effets des modifications des données.

Commentaires : L'enseignant pourra alors éventuellement donner les résultats de simulations préparées à l'avance et obtenues à partir de simulations sur ordinateur.

**Objectifs
pédagogiques :**

- Dans un premier temps (Activités 1 et 2) il s'agit, sur des exemples élémentaires, de faire comprendre ce que signifie simuler après modélisation de l'expérience (partie qui n'est pas prise en compte ici)
- On essaie de faire une approche de la notion de probabilité par les fréquences en multipliant les séries de simulations d'une part et en augmentant la taille de ces séries d'autre part.
- Bien visualiser les fluctuations d'échantillonnage et montrer l'influence de la taille de l'échantillon sur la fluctuation.
- L'activité 3 pose le problème de l'estimation sur la base d'échantillons dans une population dont on ignore la taille et la répartition.
- Découverte de quelques fonctionnalités du tableur.

Prérequis :

- Les connaissances de Collège.
- Le vocabulaire de Collège : population, effectif, fréquence, moyenne.

Intérêt :

- Faire comprendre qu'on ne fait que simuler un modèle dont les hypothèses servent à utiliser l'ordinateur. Les conséquences resteront à vérifier expérimentalement.
- Observation dynamique et en temps réel des modifications des données.
- Visualisation de la population et de sa répartition
- Faire sentir le lien entre probabilité et fréquence.
- Approche de la loi des grands nombres en augmentant facilement le nombre des simulations.
- On s'efforce ici de développer 3 points de vue différents (cellules pour représenter la population dans une feuille du tableur; numérique, graphique) pour essayer d'aborder la problématique des intervalles de confiance.

**Description de
l'activité
instrumentée :**

- Le fichier informatique est fait pour être utilisé en salle d'informatique mais peut être utilisé aussi en rétroprojection.
- Dans les activités 1 et 2 chaque élève effectue un certain nombre de simulations et ne fait que constater le comportement des simulations par rapport à la valeur théorique (0.5) proposée : les échantillons 1 et 2 ne sont pas utilisés et les nuages sont inutiles.
- L'activité 3 est plus riche : elle s'efforce de montrer l'intérêt de l'échantillonnage. Les nuages dissimulent les informations sur la population globale (que l'on fixera pour avoir une simulation plus crédible) et seules les informations sur les échantillons apparaissent : c'est à peu près ce qui se produit dans la réalité.

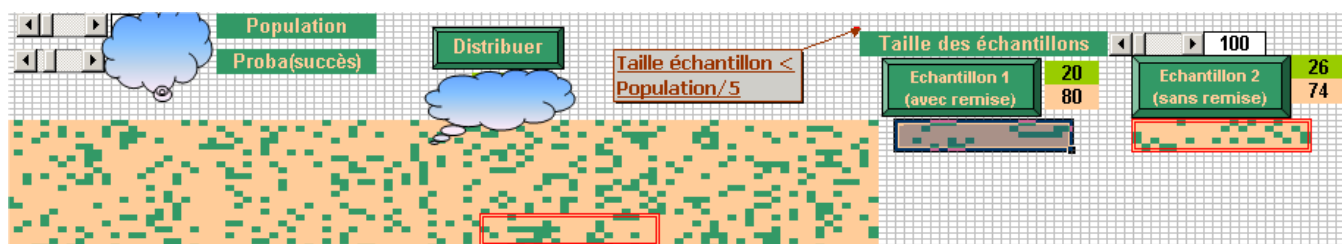
(Voir [fiche technique.](#))

Nom : [« echant.xls »](#)

Logiciel utilisé : EXCEL

Description :

- La feuille comporte 3 curseurs :
 - Un pour la population qui varie par 1000 de 2000 à 20000 (attention les délais de calcul s'allongent avec la taille de la population)
 - Un pour la fréquence théorique (probabilité) : incrémentation 0,01
 - Un pour la taille des échantillons (100 à 1500)
- Le bouton « Distribuer » lance une sélection de population avec les paramètres adoptés.
- Il y a deux types d'échantillonnages prévus :
 - Echantillon 1 : on choisit successivement et avec remise les n individus parmi les N individus de la population.
 - Echantillon 2 : on tire aléatoirement une plage de cellule de taille n dans la plage « population », il n'y a pas remise.
- Les boutons « Echantillon 1 » et « Echantillon 2 » relancent les échantillonnages mais aussi redistribuent la population (voir mode d'emploi).
- Les 2 petits nuages permettent de dissimuler des informations



Mode d'emploi :

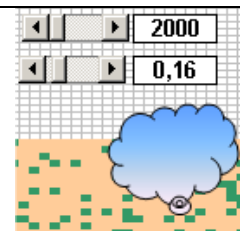
- Emploi 1 : On choisit la taille N de la population et la probabilité p avec les curseurs puis on multiplie les simulations avec le bouton « distribuer », les résultats sont affichés. On peut aussi échantillonner mais la population est redistribuée (donc les échantillons portent sur des populations différentes) et c'est long.
- Emploi 2 : Si on veut échantillonner sur une population fixe :
 - Dans la feuille choisir population et probabilité à fixer définitivement puis sélectionner la plage « **population** » dans la « zone nom » population en haut à gauche.
 - Faire alors un **Copier** » suivi dans le menu **Edition** d'un **Collage Spécial > Valeurs** qui fixe les valeurs des cellules. La population est stabilisée.
 - Cliquer sur « Echantillon 1-2 » pour multiplier les échantillons, c'est plus crédible. Mais on évitera d'enregistrer le fichier à la fin..(ou alors on travaillera sur une copie).
 - Pour donner de l'intérêt on fera choisir N et p par le voisin qui cachera ensuite ces informations.

Documentation :

- EXCEL mode d'emploi.

Objectifs : Simulation et fluctuation d'échantillonnage.

Activité 1 : simuler un lancer de pièce. (Cliquez sur **Distribuer** pour lancer une nouvelle série et déplacer les nuages pour faire apparaître les informations.)



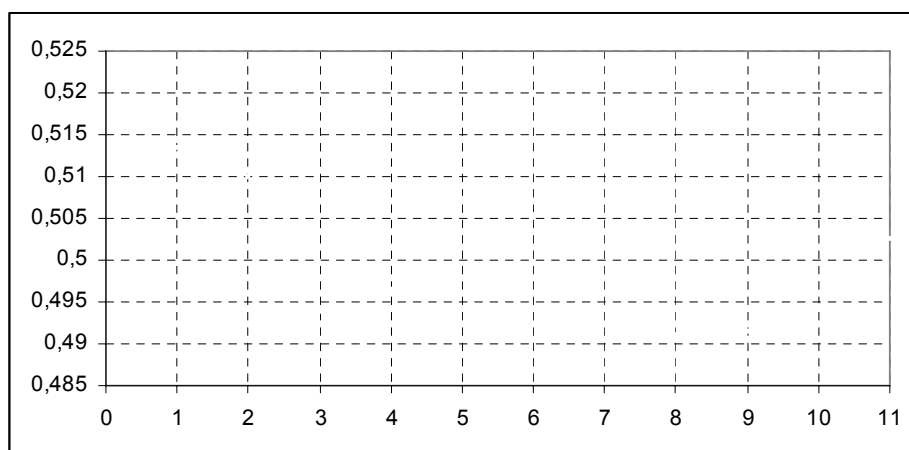
Lorsqu'on lance une pièce nous avons la conviction qu'il y a une chance sur deux pour qu'elle tombe sur « Pile ».

1. Régler la taille de la population sur $N=2000$ et la fréquence théorique sur $p=0,5$. Faire 10 séries de 2000 lancers (touche **F9**) et remplir le tableau de fréquences ci-dessous (Pile=Vert et Face=Ocre)

Séries	1	2	3	4	5	6	7	8	9	10	Moy.
Piles (fréq)											
$ f-0,5 $											

On constate que les distributions de fréquences sont différentes : c'est ce qu'on appelle **la fluctuation d'échantillonnage**.

2. Faire la moyenne **m** des fréquences **f** et comparer avec $1/2$.
3. Calculer $|f-0,5|$ et vérifier que $|f-0,5| < 1/\sqrt{2000}$ pour la plupart des séries.
4. Compléter le graphique ci-dessous par les points d'abscisses 1-2.. et d'ordonnées les fréquences. Tracer les droites d'équations $y=m$; $y=m+1/\sqrt{2000}$ et $y=m-1/\sqrt{2000}$. Commenter.



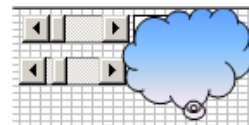
5. On va faire 10 séries de lancers en faisant varier le nombre **n** des lancers (2000-3000-4000..11000) Compléter ce tableau puis constater que les fréquences semblent se rapprocher de 0,5 en vérifiant presque toujours que $|f-0,5| < 1/\sqrt{n}$. On pourra compléter le même graphique avec des points de couleur différente.

Séries	2000	3000	4000	5000	6000	7000	8000	9000	10000	11000	Moy.
Piles (fréq.)											
$ f-0,5 $											
$1/\sqrt{n}$											

Activité 2 : simuler un lancer de dé.

Reprendre l'activité précédente avec le lancer d'un dé en vous intéressant à l'obtention du **5** : vous êtes convaincu d'avoir **1 chance sur 6** ($p=1/6 \approx 0,16$) d'obtenir 5. (vert=5 et ocre \neq 5). Faire les réglages qui s'imposent puis construire les tableaux.

Activité 3 : simuler un sondage d'opinion.



- ▶ Faire choisir la taille de la population et la fréquence de la population par votre voisin qui ensuite dissimulera les informations sur la fréquence théorique, sur la taille de la population (≥ 5000) et la distribution de ses fréquences (à l'aide des nuages)
- ▶ Le jeu va consister à deviner la distribution des fréquences à l'aide des fréquences des échantillons: c'est ce que l'on appelle un **sondage**.



1. Régler la taille des échantillons sur **n=100** puis cliquer sur « distribuer » puis fixer la population (voir la méthode sur la feuille). Faire une série de 10 sondages et compléter ce tableau pour les deux types d'échantillons : (vert=OUI et ocre=NON)

Séries de 100	1	2	3	4	5	6	7	8	9	10	Moy.
OUI (fréq) Ech.1											
OUI (fréq) Ech.2											

A-t-on une idée à présent de la distribution des fréquences des OUI et des NON dans la population ?
Les informations obtenues sont-elles précises ?

2. Faire une représentation graphique en nuage comme dans l'activité 1 (graduer l'axe des ordonnées de $m_1 - 0.3$ à $m_1 + 0.3$) et représenter les droites d'équations $y=m_1$ et $y=m_2$ (moyennes) sur le même graphique. (compléter le graphique ci-dessous).
3. Régler à présent la taille des échantillons sur **1000** et recommencer (attention placer le curseur de la population assez loin pour que sa taille soit assez grande).

Séries de 1000	1	2	3	4	5	6	7	8	9	10	Moy.
OUI (fréq) Ech.1											
OUI (fréq) Ech.2											

4. Représenter le nuage des résultats dans le même graphique par des points de couleur différente.
5. Faire une hypothèse sur la distribution des fréquences de la population (donner une fourchette pour les OUI par exemple). Déplacer le petit nuage pour vérifier votre hypothèse.

