Scénario 3

ECHANTILLONNAGE

1 Présentation générale

1.1 Objectifs :

- Citer quelques techniques de sondage.
- Réaliser un échantillonnage selon un modèle courant.

1.2 Les étapes

1. Notions de probabilité
   équiprobabilité
   estimation
   précision de l'estimation
   biais de l'estimation
2. Prise de conscience de l'intérêt de l'échantillonnage.
3. Description des 3 principales stratégies d'échantillonnage, avec leurs avantages et inconvénients.
4. (*facultative*) Application des méthodes d'échantillonnage en histologie (peut être remplacé par l'exemple de l'efficacité du vaccin contre la polio, ou un exemple d'échantillonnage en forêt).
5. Evaluation : exercice d'échantillonnage (verger de clémentiniers).

1.3 Matériel

1 dé à 6 faces.
16 transparents.
Projecteur de transparents.
Feutres à transparent.

1.4 La forme

- présentation orale et sur transparents.
- exercice par petits groupes : réalisation d'un échantillonnage.
- distribution d'un extrait du poly. : échantillonnage (de J. Vaillant).

1.5 Durée

3 heures.

1.6 Prérequis

Pas.

1.7 Evaluation

Par petits groupes pour l'exercice d'échantillonnage (étape 5).

2 Les étapes

2.1 Etape 1 : Notions

Notion de probabilité
équiprobabilité
estimation
précision de l'estimation
biais de l'estimation

Transparent 1 : notion de probabilité.
Transparent 2 : notion d'équiprobabilité.
Transparent 3 : notion d'estimation.
On dispose d'un dé à 6 faces (le dé est supposé non pipé). Le raisonnement (dé non pipé) nous conduit à dire que la probabilité d'obtenir 1, 2, 3, 4, 5 ou 6 est la meme; *elle vaut 1/6.*
On voudrait expérimenter, et estimer la probabilité d'obtenir 1, 2, 3, 4, 5 ou 6.

Pour cela, on lance le dé 6 fois. On n'obtiendra pas forcément 1 fois 1, 1 fois 2,...,1 fois 6, mais certainement un autre résultat (*effectuer le lancer, noter le résultat sur le transparent*). Le résultat obtenu est une fréquence observée (nombre de fois où j'ai obtenu 1, 2,...6). On peut estimer la probabilité d'obtenir 1, par le nombre de fois où on a effectivement obtenu 1, divisé par 6 (le nombre de *tirages*). (*Remplir la ligne correspondante du transparent*).

Cette probabilité estimée diffère de la probabilité réelle (1/6).
Transparent 4 : notion de précision de l'estimation
On a vu que la probabilité estimée diffère de la probabilité réelle. Comment peut-on améliorer notre estimation?

(Attendre les réponses) On augmente le nombre de lancers (tirages). On peut essayer 6 lancers, 60 lancers, 60000 lancers par exemple. On pourrait expérimenter. On s'apercevrait que, dans la cas de 60000 lancers, la probabilité estimée est très proche de la probabilité réelle (1/6). On "attrape" la probabilité réelle. L'estimation est précise.

De meme , l'estimation faite avec 60 lancers a une meilleure précision que l'estimation faite avec 6 lancers seulement.
A 6 lancers, on associe un premier estimateur.
à 60 lancers, on associe un deuxième estimateur.
à 60000 lancers on associe un troisième estimateur.
Transparent 5 : notion de biais de l'estimation.
On dispose d'un sac de grains.
On sait que : meme si les grains sont bien mélangés avant d'etre mis en sacs, les sacs étant ensuite manipulés, et secoués, un tri des grains se fait automatiquement : les plus légers se trouvent en haut, et les plus lourds en bas.

Problème : estimer le poids de 1000 grains
(dans la réalité, on tirera au hasard un nombre de grains donné inférieur à 1000, qu'on pèsera, on en déduira le poids de 1000 grains par une règle de 3. Pour simplifier, on dira qu'on tire 1000 grains).

Deux possibilités, au moins, nous sont offertes :

- ouvrir le sac, et tirer 1000 grains.
- percer le fond du sac, et tirer 1000 grains.

Dans le premier cas, on obtiendra des grains plutôt légers, et on sous-estimera le poids de 1000 grains; dans le deuxième cas, on obtiendra des grains plutôt lourds, et on surestimera le poids de 1000 grains.

Dans les deux cas, on obtient une **estimation biaisée**. Pourquoi cette estimation est-elle biaisée? Parce que chaque grain n'a pas la même probabilité d'être tiré.
Transparent 6 : illustration des notions de précision et de biais de l'estimation.
Voilà une illustration des notions de précision et de biais de l'estimateur.

Chacun des 4 graphiques représente un estimateur. Au centre de la **cible**, se trouve la vraie valeur (inconnue) que l'on cherche à estimer. Chaque point représente une estimation. Sur chaque cible, on trouve plusieurs estimations, qui correspondent à plusieurs expériences répétées.

Exemple : je dispose d'un dé à 6 faces. Je veux estimer la probabilité d'obtenir 1, lorsque je lance le dé. Je choisis de faire 60 tirages pour obtenir une estimation. Cela me définit un estimateur. Si je fais 10 fois 60 tirages, j'obtiens 10 estimations (donc 10 points sur la cible).

Question :
Quels sont les estimateurs précis - peu précis
biaisé - sans biais?
Réponses :

1. estimateur peu précis, biaisé.
2. estimateur peu précis, sans biais.
3. estimateur précis, biaisé.
4. estimateur précis, sans biais.

Dans le cas d'un estimateur biaisé, on a un écart (biais) systématique entre l'estimation et la vraie valeur. Cela n'est pas souhaitable. On recherche également un estimateur le plus précis possible.

2.2 Etape 2 : Prise de conscience de l'intérêt de l'échantillonnage

On pose oralement le problème suivant :
On cherche à connaitre la proportion de foyers français possédant une baignoire. Comment procéder?

Attendre les réponses des stagiaires, les amener à proposer les 2 possibilité suivantes :

- *le recensement (possibilité coûteuse en temps et argent)*
- *l'échantillonnage (sondage)*

Commentaires :

- Dans le cas du recensement, l'observateur se situe au niveau de la population donc **pas d'incertitude sur le résultat** trouvé et **détermination des valeurs exactes** recherchées.
- Dans le cas de l'échantillonnage, l'observateur obtiendra une **estimation de la valeur recherchée** (estimation de la proportion de gens ayant une baignoire).

Transparent 7 : définition de l'échantillonnage.
Insister sur la nécessité que tous les individus de la population aient une chance de se trouver dans l'échantillon pour pouvoir réaliser un échantillonnage.

On pose alors l'hypothèse qu'on dispose du fichier INSEE des foyers français qui comporte les informations suivantes :

- Numéro INSEE de chaque foyer (chacun a la même probabilité d'être tiré) ; ces numéros sont classés en fonction de l'âge du foyer (au début du fichier on a les foyer les plus anciens, en fin, les foyers les plus récents),
- la catégorie socio-professionnelle (CSP) du chef de famille.

Lancer le débat : A partir de ce fichier, comment échantillonner? *Recueillir les stratégies proposés par les formés. commenter.*
Essayer de les amener à trouver 3 stratégies principales qu'on développera par la suite (échantillonnages aléatoire simple, systématique et stratifié).

2.3 Etape 3 : Description des 3 principales stratégies d'échantillonnage avec leurs avantages et inconvénients.

2.3.1 L'échantillonnage aléatoire simple

1. Définition (**Transparent 8 :**)
   échantillonnage aléatoire simple :
   Prélèvement d'unités statistiques :
   - au hasard
   - avec la même probabilité pour toutes les unités
   - de façon indépendante
2. Application à notre cas
   Détermination de la taille de l'échantillon (nombre de foyers sondés) ou détermination du taux d'échantillonnage (ou taux de sondage)? : il s'agit d'un compromis entre le coût et la précision de l'opération.

   Exemple :
   - sondage de 5000 foyers : coût élevé mais forte précision sur le résultat
   - sondage de 500 foyers : coût plus faible mais incertitude plus forte sur le résultat cherché

   De même, on obtient une meilleure précision quand on sonde 10% de la population plutôt que 1% (notion de taux d'échantillonnage).

   Sachant la taille de l'échantillon (n), on procède ensuite au tirage au hasard de n numéros de foyer dans le fichier INSEE.
3. Avantages et inconvenients
   Avantages :
   - simplicité de conception
   - estimateurs non biaisés

   Inconvénients :
   - nécessité d'identifier chaque élément de la population (pas de problème dans notre cas, puisqu'on dispose du fichier INSEE. Cela peut poser des problèmes, par exemple si on étudie une espèce d'oiseaux : impossible de dresser la liste des individus de la population)

- protocole de sélection des unités pas toujours aisé (par exemple dans le cas où il est impossible d'identifier chaque élément de la population, donc de dresser la liste des individus de la population. Et même si on peut dresser cette liste, il peut être difficile de capturer les animaux que l'on a décidé de capturer : problèmes de visibilité, de mobilité,...).
- précision faible (on n'utilise pas l'information que l'on pourrait éventuellement avoir).

Dans notre cas :

La nécessité d'identifier chaque élément de la population n'est pas un problème (on dispose du fichier INSEE), pas de problème de protocole de sélection des unités. Le seul inconvénient : **précision faible** (on n'utilise pas la connaissance de la CSP); l'avantage : estimateur non biaisé.

2.3.2 L'échantillonnage systématique

1. Définition (**Transparent 9**)

   - soit n la taille de l'échantillon,
   - diviser la population statistique en n sous-ensembles de même taille t,
   - tirer au hasard un nombre entier entre 1 et t,
   - soit i ce nombre,
   - l'échantillon est constitué de tous les ièmes individus de chaque sous-ensemble. Les unités qui constituent l'échantillon sont régulièrement espacés.

2. Application à notre cas

   - détermination de la taille de l'échantillon (n) sur les mêmes critères que précédemment.
   - tirage au sort d'un nombre entier compris entre 1 et t $\Rightarrow$ i

     (t = N / n = pas d'échantillonnage
     = taille des sous-ensembles
     N = taille de la population
     n = taille de l'échantillon)

   - on constitue l'échantillon :
     - en prenant la ième unité, puis 1 unité toutes les t unités, ou bien (et c'est équivalent),
     - en découpant la population en n sous-ensembles de meme taille, et en prenant la ième unité de chaque sous-ensemble.

   Remarque : cette méthode d'échantillonnage n'a de sens, dans notre cas, que parce que le fichier est ordonné par ordre d'ancienneté du foyer (de cette façon, on met dans notre échantillon un peu de toutes les anciennetés du foyer).

3. Avantages et inconvénients

   Avantages :

   - commodité (préparation, exécution)
   - approprié à l'étude de répartition spéciale

- évite de prendre des individus proches (évite donc la redondance d'information)

Inconvénients :

- nécessite un découpage de la population en sous-ensembles de même taille
- problèmes si variations périodiques, et si le pas de sondage est voisin de la période — estimateur moins bon (moins efficace) que celui de l'échantillonnage aléatoire simple

Dans notre cas, meilleure prise en compte de la variabilité ancienneté du foyer (qui se trouve dans le fichier sous forme d'un **gradient**), mais pas de prise en compte de l'hétérogénéité de la variable CSP.

2.3.3 L'échantillonnage stratifié

1. Définition (**Transparent 10**)
   - Diviser la région à étudier (population) en domaines que l'on appellera strates (strates supposées homogènes).
   - Dans chaque strate, échantillonnage aléatoire simple.
2. Application à notre cas
   - On considère qu'une étude antérieure avait montré l'existence d'une relation entre la possession d'une baignoire et la catégorie socio-professionnelle du chef de famille.
   - On va donc subdiviser notre population de N foyers, qui est hétérogène, en K sous populations (strates) plus homogènes (chaque strate étant une catégorie socio-professionnelle).
   - Détermination de la taille de l'échantillon que l'on peut étudier dans chaque strate (n) réalisée sur les mêmes critères que précédemment et selon la taille de chaque strate,
   - Tirage aléatoire des numéros INSEE que l'on désire échantillonner dans chaque strate.
3. Avantages et inconvénients

   Avantages :

   - Stratification ⇒ gain en précision, surtout si la variabilité inter-strate est forte (et si la taille d'échantillon dans chaque strate est bien choisie). Donc plus efficace que l'échantillonnage simple.
   - meilleure répartition de l'effort d'échantillonnage (l'effort d'échantillonnage est déterminé à l'intérieur de chaque strate, en fonction de l'effectif de la strate, et de son importance).

   Inconvénients :
   Les informations pour la stratification peuvent etre inadéquats (le choix des strates est mauvais, les strates ne sont pas homogènes par rapport à la variable étudiée).

   - — biais de l'estimation?
   - — c'est encore pire que de faire de l'aléatoire simple (sans biais).

Transparent 11 : tableau récapitulatif des 3 types d'échantillonnage :

- aléatoire simple
- systématique
- aléatoire stratifié.

2.4 Etape 4. Application

Application des modes d'échantillonnage :

- **aléatoire**
- **systématique**
- **stratifié**

en histologie
Distribuer les fiches histologie.

Histologie : C'est la partie de l'anatomie qui traite des **tissus organiques**. En histologie, on s'intéresse à des organes, ou des parties d'organes. Selon la taille des structures à observer, l'observation est faite au microscope optique ou au microscope électronique. Pour le microscope optique, ces organes sont fixés, inclus dans de la paraffine, coupés au microtome (épaisseur de quelques microns) et colorés. En microscopie électronique, on inclut l'organe dans une résine, et l'épaisseur des coupes est d'environ 0,1 micron.

Il est en général impossible d'observer toutes les coupes d'un organe entier (à moins que l'organe soit très petit), on procède donc à un **échantillonnage**. Voici des exemples concrets :

Transparent 12 : *fig. 1* coupe d'une hypophyse en microscopie optique

(hypophyse : glande située à la base du cerveau)

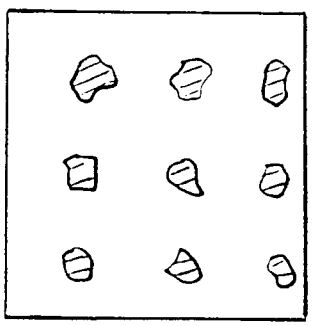
Deux types d'échantillonnage sont schématisés :
a. échantillonnage aléatoire simple
b. échantillonnage systématique

Chaque carré représente un champ oculaire d'observation. On déplace la lame sous l'objectif, de façon à la positionner dans le champ oculaire.

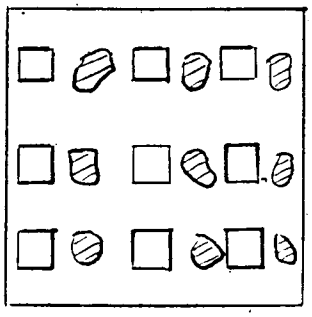
Pour des raisons pratiques, il est plus facile d'effectuer un échantillonnage systématique qu'un échantillonnage aléatoire simple : il est plus facile de déplacer la lame de façon systématique que de positionner la lame, à chaque fois, en des coordonnées tirés au sort : cette pratique serait très fastidieuse.

Donc, on pratique en général un échantillonnage systématique. L'échantillonnage systématique est adapté au contexte, puisqu'on échantillonne sur une surface (répartion spatiale); la méthode évite la redondance d'information; elle évite que les champs oculaires soient trop proches, ou se superposent.

Remarque :
Il faut faire attention à la structure du tissu : ce type d'échantillonnage risque d'être mauvais, dans le cas d'une structure périodique, et d'un pas égal à la période.
Par exemple, imaginons une structure de ce type :



et un échantillonnage systématique de ce type :



Problème!

fig. 2 échantillonnage systématique en microscopie électronique
L'échantillonnage systématique est également beaucoup pratiqué en microscopie électronique :
Les coupes de tissu sont montées sur grille. Il existe différents types de grilles dont les formes et tailles de maillage varient. Elles permettent différents types d'échantillonnage.

Ici, on observe systématiquement le coin supérieur droit de chaque maille, en déplaçant la lame dans le champ de l'objectif. On fait, en général, une photo de ce qui est vu sur l'écran du microscope à chaque emplacement. On pratique ensuite l'observation, en général un comptage, sur les photos.

L'utilisation des grilles rend très facile la pratique de l'échantillonnage systématique.

Remarque : comme pour la microscopie optique, il faut faire attention à la structure des tissus : attention aux tissus à structure périodique, et au risque de pas égal à la période.
Donc une connaissance de la structure du tissu est nécessaire.

Transparent 13 : *fig. 3* coupe dans le cartilage d'un os :

Question : Existe-t-il une structure dans ce tissu?

Réponse : Il existe une structure en couche, un gradient dans la forme et la taille des cellules, le système est ordonné.

Question : Quel type d'échantillonnage choisir?

échantillonnage aléatoire simple? Pourquoi pas, mais il y a peut être mieux. l'échantillonnage aléatoire simple donne des estimateurs de faible précision.

échantillonnage systématique? Oui, cela permettrait de s'arranger pour prendre des cellules de toutes les tailles et de toutes les formes. $\Longrightarrow$ pratiquer un échantillonnage systématique dans le sens du gradient.

échantillonnage stratifié? On n'a pas réellement de strates, c'est plutôt un gradient. Donc, on fera de l'échantillonnage systématique.

fig. 4 muscle squelettique :

Question : Existe-t-il une structure dans ce tissu?

Réponse : Oui, il existe une structure périodique. On pourrait déterminer des strates, et faire de l'échantillonnage stratifié.

Conclusion :
S'il existe une structure dans le tissu, c'est le type de structure qui doit nous conduire au type d'échantillonnage :
gradient $\Longrightarrow$ systématique
régions homogènes $\Longrightarrow$ stratifié

Des raisons pratiques peuvent aussi nous faire choisir un type d'échantillonnage plutôt qu'un autre (par exemple, l'échantillonnage systématique est plus facile à mettre en œuvre en microscopie, d'autant qu'on dispose de grilles toutes faites, avec des maillages de différentes tailles).

Donc, il faut avoir une connaissance de ce qu'on observe, pour choisir un type d'échantillonnage.

Transparent 14 : *fig. 5*
On peut également faire des études, sur un type de tissu donné, qui nous permettent de choisir le type d'échantillonnage à réaliser par la suite.

Exemple : sur un type d'organe, le ganglion du système sympathique, on veut compter le nombre de nucléoles. On fait 30 coupes par organe. On procède à quatre types d'échantillonnage : aléatoire simple, stratifié avec un premier choix de strates, stratifié avec un deuxième choix de strates, systématique.
On représente la variabilité de l'estimation (des nombres de nucléoles) en fonction du nombre de coupes étudiées (de 1 à 30), pour les quatre types d'échantillonnage.
Plus la variabilité de l'estimation est grande, plus la précision de l'estimation est faible : l'échantillonnage systématique donne des estimations de meilleure précision que les échantillonnages stratifiés, qui donnent eux-même des estimations de meilleure précision que l'échantillonnage aléatoire simple.

Remarque : *(facultative)*
On n'a pas étudié le biais (seulement la précision) : il faudrait ne pas se contenter de cette étude, et regarder la structure du tissu, pour vérifier que l'échantillonnage systématique ne risque pas de donner une estimation biaisée (en cas de structure périodique, par exemple).

Conclusion :
Le choix d'un type d'échantillonnage peut se faire en fonction :

- de raisons pratiques (de commodité)
- de l'étude de la structure du tissu (de la population)
- d'études théoriques.

Les exemples de l'étape 4 sont tirés de "Stereologicel Methods" (WEIBEL)

Exercice : *(collectif et oral)*
Transparent 15 : *fig. 6*
Voici un schéma qui représente un organe composé de quatre blocs A, B, C et D.
Dans chaque bloc, on extrait une tranche, qui est découpée en dés. On tire au hasard deux dés, parmi les dés de chaque tranche. De chaque dé tiré au hasard, on extrait des coupes régulièrement espacées. Chaque coupe est posée sur une lame et étudiée au microscope.

Question : Trois types d'échantillonnage successifs ont été utilisé ici, lesquels?

Réponse :

échantillonnage stratifié : On décompose l'organe en quatre strates A, B, C et D. Dans chaque strate, on fait un échantillonnage aléatoire simple, et on obtient une tranche par strate.

échantillonnage aléatoire simple : Les tranches sont divisés en dés. On tire au hasard deux dés parmi les dés de chaque tranche.

échantillonnage systématique : On extrait de chaque dé des coupes régulièrement espacées.

2.5 Etape 5. Evaluation : exercice d'échantillonnage (verger de clémentiniers)

1. Distribuer la feuille stagiaire et le plan stagiaire. Expliquer l'exercice. Faire plusieurs groupes, chacun faisant un type d'échantillonnage (si possible plusieurs aléatoire simple).
2. Chaque groupe fait le tirage des numéros d'arbres. Le formateur passe dans chaque groupe et indique, *sans montrer son plan*, quels sont les arbres du tirage qui ont une note 4 ou 5.
3. On discute des décision prises par chaque groupe : théoriquement, compte tenu des fréquences d'infestation précisées dans le plan du formateur, on peut s'attendre à :
   - Les tirages aléatoires simples peuvent mener à différentes conclusions (ci-joints 2 exemples de tirage à décisions différentes).
   - Par contre les tirages stratifiés doivent conduire généralement à la même décision : traiter dans la pente mais pas dans le fond de la vallée.
   - Pour un tirage aléatoire simple conduisant à un traitement partout, la décision est peu écologique et peu économique.
   - Pour un tirage aléatoire simple conduisant au non traitement, il y a visiblement risque de propagation de l'infestation.

2.6 Points stratégiques

probabilité
représentation d'un échantillon
échantillon- population

2.7 Mots-clés

variabilité
précision
biais
stratification
hétérogénéité
échantillonnage
échantillon
sondage
variabilité
population
taux de sondage
échantillonnage stratifié
échantillonnage systématique
échantillonnage aléatoire simple

2.8 Vocabulaire

population
échantillon
échantillonnage stratifié
échantillonnage systématique
échantillonnage aléatoire simple
estimation
précision d'estimation
biais
probabilité

2.9 Références bibliographiques

WEIBEL E. R. (1979) *Stereological Methods*
Practical methods for biological morphometry. Academic Press.

GOUET J.P. *Quelques tables statistiques usuelles* Document interne de l'ITCF.

VAILLANT J., DENIS J.B., ROSPARS J.P. (Octobre 89)
Pratique de l'échantillonnage. Notions de base et illustrations. Groupe échantillonnage de Versailles.

BOUHOT L. (1990) *Distribution spatio-temporelle et stratégies d'échantillonnage des attaques de Tomicus piniperdas L. et Ips sexdentatus Boern. sur le pin sylvestre en forêt d'Orléans.* Thèse de l'université Claude Bernard. Lyon 1.

ITCF

3 Annexes

3.1 Documents stagiaires

Présentation du scénario (feuilles stagiaire 1, 2)
Table de nombres au hasard (feuilles stagiaire 3, 4)
4 fiches histologie (feuilles histologie H1, H2, H3, H4)
Fiche exercice clémentiniers (feuille stagiaire 5)
Plan du verger (feuille stagiaire 6)
Fascicule extrait de *Pratique de l'échantillonnage* (numérotées 7 à 13)
Vocabulaire du scénario 3 (1 page)

3.2 Documents formateur

Plan du verger avec les arbres infestés (feuille formateur 1)
Solution de l'exercice *clémentiniers* (feuille formateur 2, 3)
2 exemples de tirages aléatoires simples dans le verger
Transparents réduits commentés.

3.3 Transparents

16 transparents numérotés de 1 à 16

3.4 Annexe 1 : Le vaccin Salk

3.5 Annexe 2 : Protocoles d'échantillonnage en forêts

3.6 Annexe 3 : Fiches ITCF