

Méthodes d'Echantillonnage et d'Enquête

Dr Ir. Codjo Emile AGBANGBA
Biostatistique, Géostatistique et Modélisation

Enseignant Chercheur/UAC/EPAC/Gen

OBJECTIFS DU COURS

- Développer les différentes méthodes d'échantillonnage
- Concevoir les questionnaire et collecter les données à l'aide des outils mobiles

Méthodes d'Echantillonnage

Principaux éléments d'une collecte de donnée

Quelle technique d'échantillonnage? Quelle taille d'échantillon?

Taille de l'échantillon

Comment déterminer la taille de l'échantillon ?

Elle résulte d'un compromis entre les ressources disponibles et le niveau de précision d'analyse voulu.

La taille de l'échantillon est très importante pour assurer la validité des résultats de recherche. Si n est insuffisante, les conclusions ne peuvent pas être validées sur la population.

Méthodes d'Echantillonnage

Taille de l'échantillon

Méthode 1

La taille de l'échantillon à enquêter peut être déterminée par une approximation normale d'une variable binomiale (Dagnelie, 1998) :

$$n = \frac{U_{1-\alpha/2}^2 \times p(1-p)}{d^2}$$

n = taille de l'échantillon ; est la valeur de la variable normale aléatoire pour une valeur de probabilité de $\alpha = 0,05$; = 1,96 ; *p* est la proportion d'individus qui connaissent l'espèce et une utilisation de celle-ci; *d* est la marge d'erreur en % ($1 \leq d \leq 15$).

Méthode 2

La taille de l'échantillon à enquêter peut être déterminée par une approximation normale d'une variable continue (Dagnelie, 1998) :

$$n = \frac{t_{1-\alpha/2}^2 \times cv^2}{d^2}$$

cv = coefficient de variation du paramètre considéré, $t_{1-\alpha/2}$ = valeur critique de la statistique de Student (tend vers une distribution normale pour de grande taille d'échantillon : $n > 30$, *d* = marge d'erreur, en %). ($1 \leq d \leq 15$).

Méthodes d'Echantillonnage

- Il a deux types d'échantillonnage : l'échantillonnage probabiliste et non probabiliste.
- Dans l'échantillonnage non probabiliste , une méthode subjective de sélection des unités est appliquée à une population.
- C'est un moyen rapide, facile et bon marché de sélectionner un échantillon.
- Cependant, s'il veut formuler des inférences au sujet de la population à partir de l'échantillon, l'analyste des données doit supposer que l'échantillon est représentatif de la population.
- Cette supposition est souvent risquée si l'échantillon est non probabiliste.

Méthodes d'Echantillonnage

- L'échantillonnage probabiliste, comprend la sélection d'unités dans une population selon le principe du choix aléatoire ou au hasard.
- L'échantillonnage probabiliste est plus complexe, demande davantage de temps et coûte habituellement plus cher que l'échantillonnage non probabiliste.
- Étant donné que les unités de la population sont sélectionnées au hasard, et que la probabilité d'inclusion de chaque unité peut être calculée, il est cependant possible de faire des estimations fiables, ainsi que des estimations de l'erreur d'échantillonnage, et de formuler des inférences au sujet de la population.

Échantillonnage non probabiliste

- L'échantillonnage non probabiliste est un moyen de sélectionner des unités d'une population à l'aide d'une méthode subjective (c.-à-d. non aléatoire).
- Il n'est pas nécessaire d'avoir une base de sondage complète pour l'échantillonnage non probabiliste qui est donc un moyen rapide, facile et bon marché d'obtenir des données.
- L'échantillonnage non probabiliste pose un problème : il n'est pas évident qu'il est possible de généraliser et d'appliquer les résultats de l'échantillon à toute la population.
- La raison de cette constatation est que la sélection d'unités dans une population pour un échantillon non probabiliste peut donner des biais d'importance.

- Étant donné le biais de sélection et (habituellement) l'absence de base de sondage, la probabilité d'inclusion d'une personne ne peut être calculée pour les échantillons non probabilistes et il est donc impossible de faire des estimations fiables ou des estimations de leur erreur d'échantillonnage.
- Il faut supposer que l'échantillon est représentatif de la population pour faire des inférences sur celle-ci.

Avantages de l'échantillonnage non probabiliste :

i. Il est rapide et pratique.

Règle générale, les échantillons non probabilistes sont obtenus en peu de temps et l'enquête est rapide : il

est très facile de simplement sortir et poser des questions à la première centaine de personnes rencontrées dans la rue.

ii. Il est relativement bon marché.

Il faut habituellement quelques heures seulement du temps d'un intervieweur pour faire ce genre d'enquête. De plus, les échantillons non probabilistes ne sont généralement pas dispersés géographiquement et les frais de déplacement des intervieweurs sont donc minimes.

iii. Une base de sondage n'est pas nécessaire.

• iv. Il peut être utile pour les études de recherche et d'élaboration d'enquête.

Inconvénients de l'échantillonnage non probabiliste :

- i. Il faut avoir des hypothèses solides sur la représentativité de l'échantillon pour formuler des inférences sur la population. Étant donné que tous les échantillons non probabilistes comportent un biais de sélection, il est souvent dangereux de formuler ces hypothèses. Il vaudrait mieux procéder à un échantillonnage probabiliste si des inférences sont nécessaires.

- ii. Il est impossible de déterminer la probabilité qu'une unité de la population soit sélectionnée pour l'échantillon, et des estimations fiables et des estimations de l'erreur d'échantillonnage ne peuvent donc être faites.

Cinq différents types de méthodes d'échantillonnage non probabilistes

- l'échantillonnage à l'aveuglette,
- l'échantillonnage à participation volontaire,
- l'échantillonnage au jugé,
- l'échantillonnage par quotas et l'échantillonnage probabiliste modifié.
- L'échantillonnage de réseaux ou boule de neige.

Echantillonnage à l'aveuglette

- Les unités sont sélectionnées de façon arbitraire, sans idée préconçue, et la planification est minime, sinon nulle. Celui qui fait l'échantillonnage à l'aveuglette présume que la population est homogène : si les unités de la population sont toutes semblables, n'importe quelle unité peut être choisie pour l'échantillon.
- L'interview de « l'homme de la rue » est un exemple d'échantillonnage à l'aveuglette parce que l'intervieweur choisit n'importe quel passant.
- Sauf si la population est vraiment homogène, les biais de l'intervieweur et du passant au moment de l'échantillonnage peuvent malheureusement avoir des répercussions sur la sélection.

Échantillonnage à participation volontaire

- Cette méthode fait appel à des répondants volontaires. Les volontaires doivent généralement faire l'objet d'un examen pour obtenir un ensemble de caractéristiques qui convient aux objectifs de l'enquête (p. ex., les personnes atteintes d'une maladie en particulier).
- Cette méthode peut être marquée d'un important biais de sélection, mais elle est parfois nécessaire. Pour des raisons de déontologie, on peut faire appel, par exemple, à des volontaires ayant des conditions médicales particulières pour procéder à certaines expériences médicales.

Échantillonnage à participation volontaire

- Voici un autre exemple d'échantillonnage à participation volontaire : au cours d'une émission radio ou télédiffusée, une question fait l'objet d'une discussion et les citoyens à l'écoute sont invités à téléphoner pour exprimer leurs opinions.
- Seuls ceux que le sujet intéresse vraiment d'une façon ou d'une autre ont tendance à répondre.
- La majorité silencieuse ne répond habituellement pas et nous avons donc un biais de sélection marqué.
- L'échantillonnage à participation volontaire sert souvent à sélectionner des particuliers pour des groupes de discussion ou des interviews approfondies (c.-à-d. une mise à l'essai qualitative qui exclut la généralisation appliquée à la population complète).

Échantillonnage au jugé

- À l'aide de cette méthode, l'échantillonnage est fait en tenant compte des idées préalables sur la composition et le comportement de la population.
- Un expert qui connaît la population décide quelles unités devraient être choisies. Autrement dit, l'expert sélectionne à dessein ce qui est considéré comme un échantillon représentatif.
- Les biais du chercheur peuvent marquer l'échantillonnage au jugé qui peut être encore plus biaisé qu'un échantillonnage à l'aveuglette.
- Étant donné que les idées préconçues du chercheur sont reflétées dans l'échantillon, des biais importants peuvent être intégrés si ces idées préconçues sont inexactes.
- Il peut cependant être utile aux études de recherche, par exemple, lors de la sélection de personnes pour des groupes de discussion ou des interviews approfondies, afin de vérifier des aspects particuliers d'un questionnaire.

Échantillonnage par quotas

- Voilà l'un des échantillonnages non probabilistes les plus communs. L'échantillonnage est fait jusqu'à ce qu'un nombre déterminé d'unités (quotas) soient sélectionnées dans diverses sous-populations.
- L'échantillonnage par quotas est un moyen d'atteindre les objectifs de taille d'échantillon pour les sous-populations.
- Les quotas peuvent être établis selon des proportions de population. S'il y a 100 hommes et 100 femmes dans la population, par exemple, et s'il faut tirer un échantillon de 20 personnes, 10 hommes et 10 femmes peuvent être interviewés.
- L'échantillonnage par quotas peut être considéré préférable à d'autres formes d'échantillonnage non probabiliste (p. ex., échantillonnage au jugé) parce qu'il faut inclure des membres de sous-populations différentes.

Échantillonnage par quotas

- L'échantillonnage par quotas ressemble à l'échantillonnage stratifié parce que des unités semblables sont regroupées.
- La méthode de sélection des unités est cependant différente.
- Les unités sont sélectionnées aléatoirement dans l'échantillonnage probabiliste, mais dans l'échantillonnage par quotas, une méthode non aléatoire est appliquée, c'est-à-dire que l'intervieweur décide habituellement qui est ajouté à l'échantillon.
- Les unités sollicitées qui ne sont pas disposées à participer sont simplement remplacées par d'autres qui le sont, et l'on ignore en fait le biais de non-réponse.

Échantillonnage probabiliste modifié

- L'échantillonnage probabiliste modifié est une combinaison d'échantillonnage probabiliste et non probabiliste. Les premières étapes sont habituellement axées sur l'échantillonnage probabiliste (voir la section suivante).
- La dernière étape est un échantillon non probabiliste, habituellement un échantillon par quotas. Des secteurs géographiques peuvent être sélectionnés, par exemple, à l'aide d'un plan d'échantillonnage probabiliste et ensuite, dans chaque région, un échantillon de personnes peut être choisi par quotas.

Échantillonnage probabiliste

- L'échantillonnage probabiliste est une méthode qui permet de formuler des inférences sur la population, compte tenu des observations tirées de l'échantillon.
- Celui-ci devrait être libre de tout biais de sélection pour formuler les inférences.
- L'échantillonnage probabiliste évite ce biais par la sélection aléatoire d'unités de la population (à l'aide d'un ordinateur ou d'un tableau de nombres aléatoires)
- Le terme aléatoire signifie que la sélection n'est pas biaisée, c'est un tirage au sort.
- L'échantillonnage probabiliste ne permet pas à l'intervieweur de décider subjectivement qui doit être choisi.

Échantillonnage probabiliste

- Voici les deux principaux critères de l'échantillonnage probabiliste : la sélection des unités est aléatoire, toutes les unités de la population de l'enquête ont une probabilité d'inclusion différente de zéro dans l'échantillon et il est possible de calculer ces probabilités.
- Il n'est pas nécessaire que toutes les unités aient la même probabilité d'inclusion et, en fait, dans les enquêtes les plus complexes, la probabilité d'inclusion varie d'une unité à l'autre.

Échantillonnage probabiliste

- Il y a de nombreux types différents de plans d'échantillonnage probabiliste.
- Le plus élémentaire est l'échantillonnage aléatoire simple et la complexité des plans s'accroît ensuite pour englober l'échantillonnage systématique, l'échantillonnage avec probabilité proportionnelle à la taille, l'échantillonnage par grappes, l'échantillonnage stratifié, l'échantillonnage à plusieurs degrés, l'échantillonnage à plusieurs phases et l'échantillonnage par répliques.

Échantillonnage probabiliste

- Chacune de ces techniques d'échantillonnage est utile dans différentes situations. Si l'objectif de l'enquête est simplement d'obtenir des estimations de la population en général, et si la stratification serait inappropriée ou impossible, l'échantillonnage aléatoire simple pourrait alors être le meilleur choix.
- Si le coût de la collecte des données de l'enquête est élevé et si les ressources sont disponibles, l'échantillonnage par grappes est souvent le choix.
- Si des estimations de sous-populations sont aussi demandées (p. ex., des estimations par province, groupe d'âge ou taille d'entreprise), l'échantillonnage stratifié est habituellement appliqué.

Échantillonnage probabiliste

- La majorité des plans plus complexes ont recours à l'information auxiliaire de la base de sondage pour améliorer l'échantillonnage.
- Si la base a été créée à partir d'un recensement précédent ou de données administratives, il peut y avoir une mine de renseignements supplémentaires qui peuvent servir à l'échantillonnage.
- Dans le cas d'une enquête sur les exploitations agricoles (fermes), par exemple, l'organisme statistique peut avoir la taille de chaque exploitation en hectares tirée du recensement agricole le plus récent.

Échantillonnage probabiliste

- S'il s'agit d'une enquête sur les citoyens, l'information (p. ex., âge, sexe, origine ethnique, etc.) peut être disponible pour chacun dans le plus récent recensement de la population.
- Lors d'une enquête sur les entreprises, l'organisme statistique peut avoir de l'information administrative, notamment, sur le genre d'industrie (p. ex., détaillant, grossiste, fabricant), le genre d'entreprise (p. ex., magasin d'aliments), le nombre d'employés, etc.
- L'information auxiliaire améliore l'échantillonnage s'il y a corrélation entre les données auxiliaires et les variables de l'enquête.

Avantages de l'échantillonnage probabiliste :

- **la sélection de chaque unité est aléatoire**, la probabilité d'inclusion de chaque unité peut être calculée, il est possible de faire des estimations fiables et d'estimer l'erreur d'échantillonnage de chaque estimation.
- On peut donc formuler des inférences sur la population. Un plan d'échantillonnage probabiliste permet en fait souvent d'utiliser un échantillon relativement petit pour formuler des inférences sur une grande population.

Inconvénients de l'échantillonnage probabiliste

- Voici les principaux : **il est plus difficile, il demande plus** de temps et il coûte habituellement plus cher que l'échantillonnage non probabiliste.
- Les frais de création et d'entretien d'une base de sondage de bonne qualité sont substantiels en général.
- Étant donné que les échantillons probabilistes ont tendance à être géographiquement répartis plus largement dans la population que les échantillons non probabilistes, les tailles d'échantillon sont habituellement plus grandes, la collecte des données coûte souvent plus chère et sa gestion est plus difficile.
- Pour un organisme statistique, la capacité de formuler des inférences à partir d'un échantillon probabiliste surpasse habituellement ses inconvénients.

Échantillonnage aléatoire simple (EAS)

- L'échantillonnage aléatoire simple (EAS) est le point de départ de tout plan d'échantillonnage probabiliste.
- L'EAS est une méthode de sélection en une étape qui garantit que chaque échantillon possible de taille n a une chance égale d'être sélectionné.
- *Chaque unité de l'échantillon a donc la même probabilité d'inclusion. Cette probabilité, p , est égale à n/N , où N est le nombre d'unités dans la population.*
- L'échantillonnage peut être fait avec ou sans remise. L'échantillonnage avec remise permet à une unité d'être sélectionnée plus d'une fois.
- L'échantillonnage sans remise signifie que lorsqu'une unité a été sélectionnée, elle ne peut l'être de nouveau.

Échantillonnage aléatoire simple (EAS)

- L'échantillonnage aléatoire simple avec remise (EASAR) et l'échantillonnage aléatoire simple sans remise (EASSR) sont pratiquement identiques si la taille de l'échantillon est une très petite fraction de la taille de la population parce que la possibilité que la même unité apparaisse plus d'une fois dans l'échantillon est minime.
- L'échantillonnage sans remise donne généralement des résultats plus précis et est plus pratique du point de vue opérationnel.
- Aux fins de ce chapitre, l'échantillonnage est supposé être sans remise, sauf avis contraire.

Avantages de l'EAS comparativement à d'autres techniques d'échantillonnage probabiliste

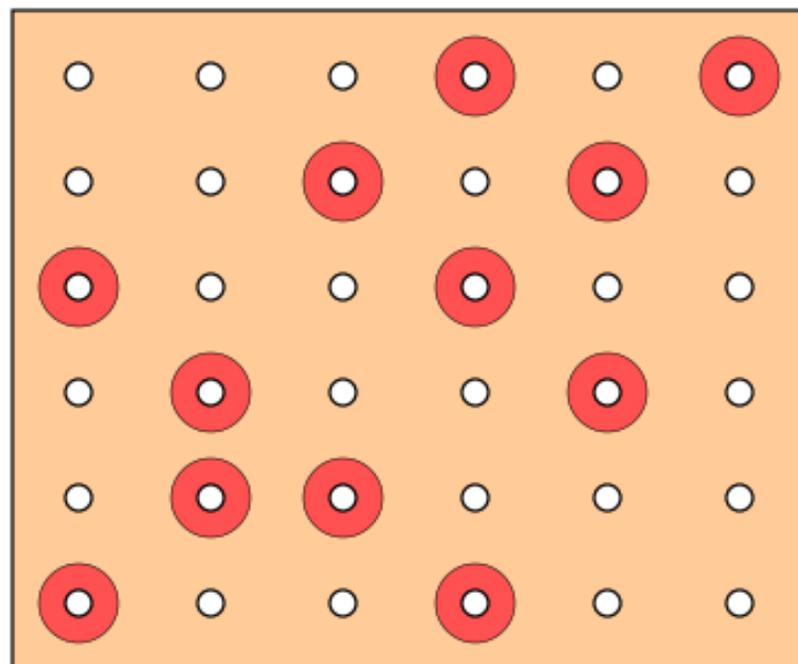
- i. C'est la technique d'échantillonnage la plus simple.
- ii. Il n'est pas nécessaire d'avoir de l'information supplémentaire (auxiliaire) dans la base de sondage pour tirer l'échantillon.
- Les seuls renseignements nécessaires sont une liste complète de la population de l'enquête et de l'information permettant d'entrer en communication avec les personnes choisies.

- iii. L'élaboration technique n'est pas nécessaire.
- La théorie sous-jacente à l'EAS est bien établie et il y a des formules standard pour déterminer la taille de l'échantillon, les estimations de la population et de la variance, et ces formules sont faciles à appliquer.

Inconvénients de l'EAS

- i. L'information auxiliaire n'est pas utilisée même si cette information existe dans la base de sondage. Les résultats peuvent donc donner des estimations statistiquement moins efficaces que celles d'un autre plan d'échantillonnage.
- ii. Il peut coûter cher s'il y a des interviews sur place parce que l'échantillon peut être largement étalé géographiquement.
- iii. L'échantillon de l'EAS peut être « mauvais ». Tous les échantillons de taille n ont une chance égale d'être ajoutés à l'échantillon et il est donc possible d'obtenir un échantillon qui n'est pas bien réparti et qui représente peu la population.

Échantillon aléatoire simple (illustré, $n=12$)



Échantillonnage systématique (SYS)

- Les unités d'un échantillonnage systématique (SYS) sont sélectionnées à intervalles réguliers dans la population. L'échantillonnage systématique sert parfois si l'organisme statistique veut utiliser un EAS, mais s'il n'y a pas de liste disponible, ou si l'ordre de la liste est approximativement aléatoire, auquel cas, le SYS est encore plus simple à faire que l'EAS. Un intervalle d'échantillonnage et une origine choisie au hasard sont nécessaires. Si une liste est utilisée et si la taille de la population, N , est un multiple de la taille de l'échantillon, n , chaque k e unité est sélectionnée lorsque l'intervalle k est égal à N/n . Un seul nombre, l'origine r , est choisi au hasard entre 1 et k inclusivement. Les unités sélectionnées sont donc : $r, r+k, r+2k, \dots, r+(n-1)k$. Chaque unité, comme dans l'EAS, a une probabilité d'inclusion, p , égale à n/N , mais, contrairement à l'EAS, chaque combinaison de n unités n'a pas une chance égale d'être sélectionnée :
- dans un SYS, nous pouvons uniquement sélectionner les échantillons dont les unités sont séparées par k .
- Seulement k échantillons possibles peuvent donc être tirés de la population à l'aide de cette méthode.

Échantillonnage systématique (SYS)

- Supposons, pour illustrer le SYS, qu'une population contienne $N=54$ unités et qu'un échantillon de taille $n=9$ unités soit sectionné. L'intervalle d'échantillonnage serait $k = N/n = 54/9 = 6$. Un nombre aléatoire entre 1 et $k = 6$, disons 2, est ensuite choisi.
- Les unités de la population sélectionnées pour l'échantillon sont ensuite numérotées : 2, 8, 14, 20, 26, 32, 38, 44 et 50.
- En présence d'un intervalle d'échantillonnage de 6 et d'une population dont la taille est de 54 unités, il y a seulement six échantillons SYS possibles, mais il y a plus de 25 millions d'échantillons aléatoires simple de taille 6 possibles.

Échantillonnage systématique (SYS)

- Un avantage de l'échantillonnage systématique est qu'il peut être utilisé lorsqu'il n'y a pas de liste disponible des unités de la population. Une base de sondage peut être établie dans ce cas en choisissant chaque *ke* *personne jusqu'à la fin de la population*.
- Le SYS pose un problème : la taille de l'échantillon, *n*, est connue seulement après la sélection de l'échantillon.
- Il peut y avoir un autre problème si l'intervalle d'échantillonnage, *k*, correspond à une certaine périodicité dans la population.
- Supposons, par exemple, qu'une enquête sur la circulation est faite dans un secteur et qu'une journée seulement de la semaine peut être échantillonnée, autrement dit, *k* est chaque 7^e jour.
- Les débits de la circulation dans l'enquête seront extrêmement différents si les jours échantillons sont toujours le dimanche au lieu d'être toujours le mardi. Bien entendu, si la période d'échantillonnage est le 5^e jour, chaque jour de la semaine peut alors être visé par l'enquête.
- Malheureusement, dans la plupart des cas, la périodicité n'est pas connue d'avance.

Avantages de L'échantillonnage SYS

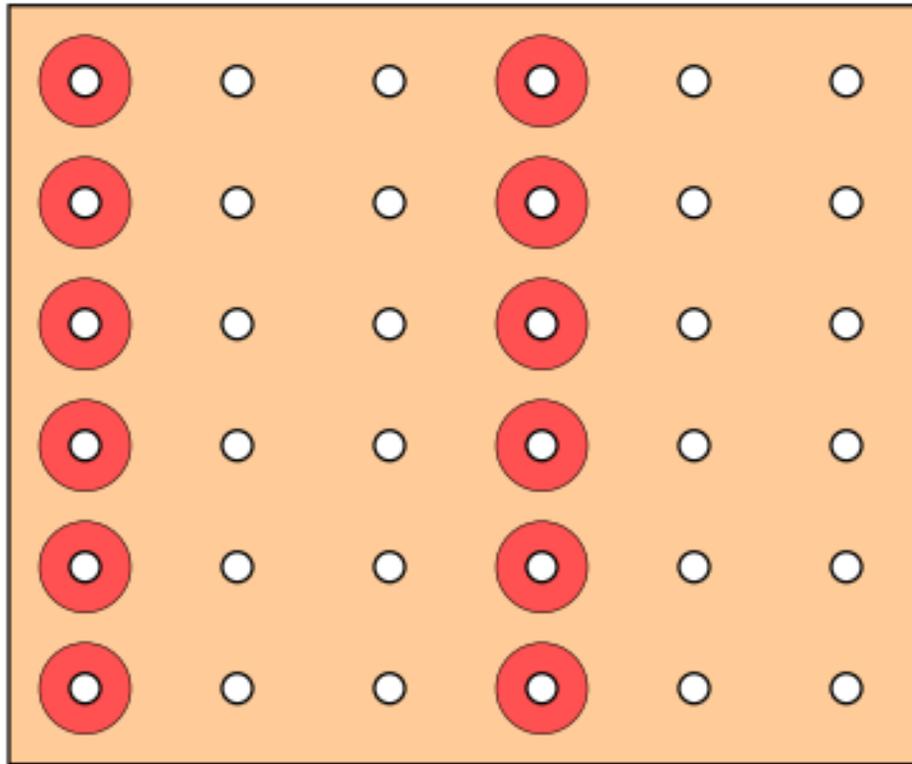
- i. C'est un substitut de l'EAS lorsqu'il n'y a pas de base de sondage.
- ii. Contrairement à l'EAS, l'information auxiliaire de la base de sondage n'est pas nécessaire.
- iii. Il peut donner un échantillon mieux réparti que celui de l'EAS (compte tenu de l'intervalle d'échantillonnage et de la méthode de tri de la liste).
- iv. C'est une théorie aussi bien établie que celle de l'EAS et les estimations sont faciles à calculer.
- v. Il est plus simple que l'EAS parce qu'un seul nombre aléatoire est nécessaire.

Inconvénients de L'échantillonnage SYS

- i. Il peut donner un « mauvais » échantillon si l'intervalle d'échantillonnage correspond à une certaine périodicité dans la population.
- ii. L'information auxiliaire qui peut être disponible dans la base de sondage n'est pas utilisée, comme dans le cas de l'EAS, et le résultat peut être une stratégie d'échantillonnage inefficace.
- iii. La taille de l'échantillon final n'est pas connue d'avance lorsqu'une base de sondage conceptuelle est utilisée.
- iv. Il n'a pas d'estimateur non biaisé de la variance d'échantillonnage. L'échantillon systématique est souvent traité comme un échantillon aléatoire simple pour faire l'estimation de variance. C'est approprié seulement lorsque la liste est triée au hasard. (Pour davantage d'information sur l'estimation de la variance pour un échantillon systématique, on consultera Cochran (1977) ou Lohr (1999).)
- v. Il peut donner une taille d'échantillon variable si la taille de la population, N , ne peut être divisée également par la taille de l'échantillon voulue, n (mais il est possible d'éviter cela en utilisant le SYS circulaire).

Inconvénients de L'échantillonnage SYS

Échantillon systématique (illustré, $n=12$, $N=36$, $k=3$)



- L'EAS et le SYS circulaire sont deux plans d'échantillonnage probabiliste à probabilité égale parce que chaque échantillon possible a exactement la même chance d'être sélectionné.
- Les techniques d'échantillonnage ne donnent pas toutes des probabilités égales.
- Les plans d'échantillonnage décrits dans les sections suivantes peuvent donner des probabilités inégales.
- On se rappellera que dans un échantillonnage probabiliste, le critère n'est pas que toutes les unités aient la même probabilité d'inclusion, mais plutôt qu'elles aient une probabilité d'inclusion connue différente de zéro. L'échantillonnage avec probabilités inégales peut souvent améliorer l'efficacité statistique de la stratégie d'échantillonnage.

Échantillonnage avec probabilité proportionnelle à la taille (PPT)

- L'échantillonnage avec probabilité proportionnelle à la taille (PPT) est une technique qui utilise des données auxiliaires et donne des probabilités d'inclusion inégales.
- Si les tailles des unités de la population varient et si ces tailles sont connues, l'information peut servir pendant l'échantillonnage pour accentuer l'efficacité statistique.
- L'échantillonnage PPT peut augmenter énormément la précision si les mesures des tailles sont précises et si les variables d'intérêt sont corrélées avec la taille de l'unité.
- Quand on dispose de mesures de tailles moins précises, il vaut mieux créer des groupements de tailles et procéder à l'échantillonnage stratifié.

Échantillonnage avec probabilité proportionnelle à la taille (PPT)

- Un bon exemple d'une variable de taille de l'échantillonnage PPT est la superficie. L'échantillonnage PPT est souvent utilisé dans les enquêtes sur les exploitations agricoles et la mesure de la taille est la taille de l'exploitation agricole (ferme) en hectares.
- La taille d'une exploitation agricole peut, bien entendu, augmenter (ou diminuer) si l'exploitant achète ou vend une terre, mais dans la majorité des cas, la taille de l'exploitation agricole est constante d'année en année.
- De plus, des questions typiques aux enquêtes sur les exploitations agricoles, notamment les revenus, les récoltes, le bétail et les dépenses, sont souvent corrélées avec la propriété foncière.
- D'autres mesures de taille pour les enquêtes sur les entreprises comprennent le nombre d'employés, les ventes annuelles et le nombre d'emplacements, mais ces variables risquent davantage de changer d'année en année.

Échantillonnage avec probabilité proportionnelle à la taille (PPT)

- Dans un échantillonnage PPT, la taille de l'unité détermine la probabilité d'inclusion. Dans le cas d'une exploitation agricole ayant une superficie de 200 hectares, par exemple, la probabilité d'être sélectionnée est donc deux fois celle d'une exploitation de 100 hectares.

Avantage de l'échantillonnage PPT

- Le principal **avantage de l'échantillonnage PPT est qu'il peut améliorer l'efficacité statistique de la** stratégie d'échantillonnage à l'aide de l'information auxiliaire.
- Le résultat peut être une diminution importante de la variance de l'échantillonnage comparativement à l'EAS ou même à l'échantillonnage stratifié.

Inconvénients de l'échantillonnage PPT

- i. Il faut avoir une base de sondage qui contient de l'information auxiliaire à jour de bonne qualité pour toutes les unités de la base qui peuvent servir de mesures de la taille.
- ii. Il est inapproprié si les mesures de la taille ne sont pas précises ou stables. Dans ces circonstances, il vaut mieux créer des groupements de tailles et faire un échantillonnage stratifié.
- iii. Il n'est pas toujours applicable parce que chaque population n'a pas nécessairement une mesure de la taille stable mise en corrélation avec les principales variables de l'enquête.
- iv. Le résultat peut être une stratégie d'échantillonnage statistiquement moins efficace que celle de l'EAS pour les variables de l'enquête qui ne sont pas corrélées avec les variables de la taille.
- v. L'estimation de la variance d'échantillonnage d'une estimation est plus complexe.
- vi. La création d'une base de sondage coûte plus cher et est plus complexe que celle de l'EAS ou du SYS parce que la taille de chaque unité dans la population doit être mesurée et sauvegardée.

Méthodes d'échantillonnage PPT

Comment obtient-on un échantillon PPT? Il y a de nombreuses méthodes d'échantillonnage PPT, mais trois techniques sont habituellement utilisées sont la méthode aléatoire, la méthode systématique et la méthode systématique aléatoire. (Il est supposé dans ce qui suit que les mesures de la taille sont des valeurs entières.)

i. Méthode aléatoire d'échantillonnage PPT :

- pour chaque unité de la population, faire le calcul cumulatif des mesures de la taille des unités jusqu'à l'unité elle-même comprise,
- déterminer l'étendue correspondant à chaque unité dans la population, c'est-à-dire à partir de la somme cumulative de l'unité précédente (mais sans l'inclure) jusqu'à la somme cumulative de l'unité courante,
- sélectionner un nombre aléatoire entre 0 (si les mesures de taille ne sont pas des nombres entiers) ou 1 (si les mesures de taille sont des nombres entiers) et la taille cumulative totale, et sélectionner l'unité dont l'étendue comprend le nombre aléatoire,
- répéter l'étape précédente jusqu'à ce que *n unités soient sélectionnées*.

i. Méthode aléatoire d'échantillonnage PPT :

Ferme	Taille	Taille cumulative	Étendue
1	50	50	1-50
2	1000	1050	51-1050
3	125	1175	1051-1175
4	300	1475	1176-1475
5	500	1975	1476-1975
6	25	2000	1976-2000

Trois nombres aléatoires entre 1 et 2000 sont sélectionnés pour obtenir un échantillon de trois unités. Supposons que ces nombres sont : 1697, 624 et 1109. Les exploitations agricoles (fermes) sélectionnées sont donc : les fermes 5, 2 et 3.

Dans le cas de la méthode aléatoire d'échantillonnage PPT sans remise, si plus d'une unité est sélectionnée, essayer de maintenir les probabilités directement proportionnelles à la taille et estimer les variances d'échantillonnage des estimations de l'enquête peuvent susciter des complications. La situation devient encore plus compliquée si plus de deux ou trois unités sont sélectionnées avec PPT sans remise et, en fait, fait l'objet d'un nombre considérable de travaux de recherche. La majeure partie de cette recherche est contenue dans les ouvrages de Horvitz et Thompson (1952), Yates et Grundy (1953), Rao, Hartley et Cochran (1962), Fellegi (1963), Brewer et Hanif (1983).

ii. Méthode systématique :

- pour chaque unité de la population, faire le calcul cumulatif des mesures de taille des unités jusqu'à l'unité elle-même comprise,
- déterminer l'étendue correspondant à chaque unité dans la population, c'est-à-dire à partir de la somme cumulative de l'unité précédente (mais sans l'inclure) jusqu'à la somme cumulative de l'unité courante,
- déterminer l'intervalle d'échantillonnage, $k = (\text{taille cumulative totale})/n$,
- déterminer une origine choisie au hasard, r , entre 0 (si les mesures de taille ne sont pas des nombres entiers) ou 1 (si les mesures de taille sont des nombres entiers) et k ,
- sélectionner les unités dont l'étendue contient les nombres aléatoires $r, r+k, r+2k, \dots, r+(n-1)k$.

iii. Méthode systématique aléatoire :

- La liste est établie au hasard dans cette méthode avant l'application de l'échantillonnage systématique.
- Si la liste est utilisée dans l'ordre original, comme dans le cas de l'échantillonnage systématique, certains échantillons possibles peuvent être éliminés.
- Lorsque la liste est établie au hasard, le nombre d'échantillons éventuels qui peuvent être tirés est à la hausse.
- On se souviendra des problèmes que posent ces méthodes. Dans le cas des méthodes systématiques
- aléatoires et systématiques, par exemple, si la taille d'une unité est plus grande que l'intervalle, elle peut être sélectionnée plus d'une fois. Ce problème peut être résolu uniquement en répartissant ces grandes
- unités en strates distinctes et en en faisant l'échantillonnage à part. La difficulté d'estimation des variances d'échantillonnage est un autre problème.

Échantillonnage par grappes

- L'échantillonnage par grappes est le processus de sélection aléatoire de groupes complets (grappes) d'unités de la population dans la base de sondage.
- C'est habituellement une stratégie d'échantillonnage statistiquement moins efficace que l'EAS et elle est appliquée pour plusieurs raisons.
- Premièrement, l'échantillonnage par grappes peut réduire énormément le coût de la collecte, surtout si la population est largement dispersée et si on a recours à des interviews sur place.
- Deuxièmement, il n'est pas toujours pratique d'échantillonner des unités distinctes de la population.
- Il est parfois plus facile de faire l'échantillonnage de groupes d'unités de la population (p. ex., ménages complets).
- Troisièmement, elle permet de faire des estimations pour les grappes elles-mêmes (p. ex., revenu moyen par ménage).

Échantillonnage par grappes

- L'échantillonnage par grappes est un processus en deux étapes. Premièrement, la population est regroupée en grappes (il peut s'agir de grappes naturelles, p. ex., ménages, écoles).
- La deuxième étape est la sélection d'un échantillon de grappes et l'interview de toutes les unités des grappes sélectionnées.
- La base de sondage peut déterminer la méthode d'échantillonnage. Jusqu'à maintenant, la cible a été l'échantillonnage d'unités individuelles de la population à partir d'une liste.
- Si les unités de la population sont naturellement regroupées, il est souvent plus facile d'établir une base de sondage pour ces groupes et d'en faire l'échantillonnage, plutôt que d'essayer d'établir une liste de toutes les unités individuelles de la

Échantillonnage par grappes

- population. Le client peut être intéressé, par exemple, à échantillonner les enseignants, mais avoir seulement une liste des écoles.
- Dans le cas des enquêtes sur les ménages ou les exploitations agricoles, de nombreux pays n'ont pas de listes complètes et à jour des gens, des ménages ou des exploitations agricoles dans aucune grande région géographique, mais ils ont des cartes des régions.
- Il est alors possible d'établir une base aréolaire et de répartir les secteurs géographiques en régions (grappes), de faire l'échantillonnage des régions et d'interviewer chacun dans la région.
- Divers plans d'échantillonnage peuvent servir pour sélectionner les grappes, notamment, l'EAS, le SYS ou le PPT. Un plan commun utilise le PPT dont l'échantillonnage est proportionnel à la taille de la grappe.

Échantillonnage par grappes

- Il ne faut pas oublier un certain nombre de considérations pour l'échantillonnage par grappes. Les estimations seront statistiquement efficaces si les unités d'une grappe sont aussi différentes que possible.
- Autrement, si les unités d'une grappe sont semblables, elles donnent toutes de l'information semblable et il suffirait d'interviewer une unité.

Avantages de l'échantillonnage par grappes

- i. Il peut réduire énormément le coût de la collecte parce que l'échantillon est moins dispersé que celui de l'EAS. C'est particulièrement important si la population est largement répartie et si l'enquête comprend des interviews sur place parce qu'il est possible d'économiser en diminuant le temps de déplacement des intervieweurs, en particulier pour les populations en milieu rural.
- ii. Il est plus facile à appliquer que l'EAS ou le SYS aux populations regroupées naturellement par grappes (p. ex., ménages, écoles) et à certaines populations conceptuelles, par exemple, les personnes qui traversent une frontière pendant une période déterminée. Il peut être difficile, coûteux ou impossible d'établir une liste de toutes les unités individuelles de ce genre de population comme l'exige l'EAS.

Avantages de l'échantillonnage par grappes

iii. Il permet de faire des estimations pour les grappes elles-mêmes. Les estimations du nombre moyen d'enseignants par école sont un exemple (lorsque les écoles sont en grappes).

iv. Il peut être statistiquement plus efficient qu'un EAS si les unités des grappes sont hétérogènes (différentes) du point de vue des variables de l'étude et si les grappes sont homogènes (semblables). Ce n'est cependant pas le cas en pratique, habituellement.

Inconvénients de l'échantillonnage par grappes

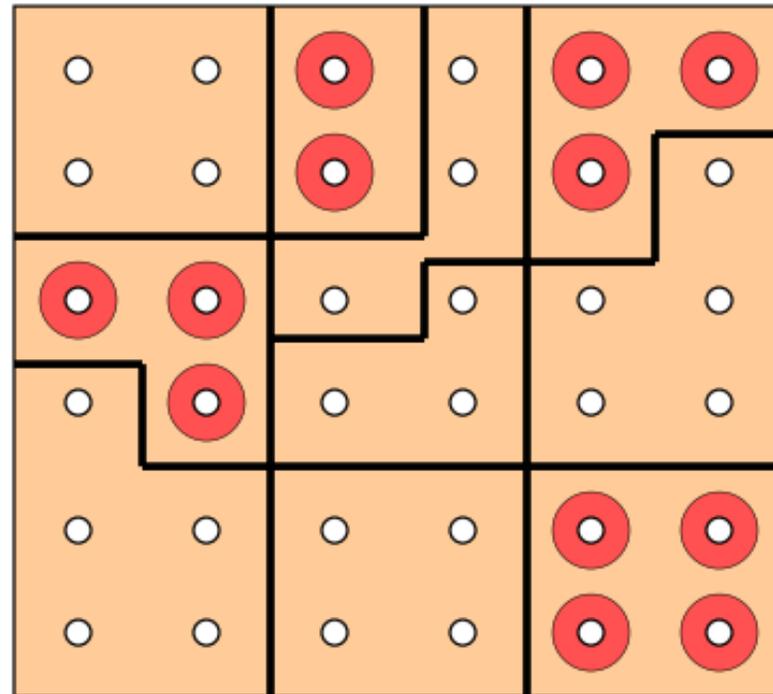
- i. Il peut être statistiquement moins efficient que l'EAS si les unités des grappes sont homogènes du point de vue des variables de l'étude. C'est souvent le cas parce que les unités d'une grappe ont tendance à avoir des caractéristiques semblables. Le nombre de grappes sélectionnées peut cependant être augmenté pour éliminer cette perte d'efficacité statistique.
- ii. La taille finale de l'échantillon n'est pas connue d'avance parce que le nombre d'unités d'une grappe est déterminé seulement à la conclusion de l'enquête.

Inconvénients de l'échantillonnage par grappes

iii. L'organisation de l'enquête peut être plus complexe que dans le cas d'autres méthodes.

iv. L'estimation de la variance peut être plus complexe que celle de l'EAS si les grappes sont échantillonnées sans remise.

Échantillon par grappes (illustré, quatre grappes sont échantillonnées)



Échantillonnage stratifié (STR)

- Au cours de l'échantillonnage stratifié, la population est répartie en groupes homogènes mutuellement exclusifs intitulés strates et des échantillons indépendants sont ensuite sélectionnés dans chaque strate.
- N'importe quel plan d'échantillonnage mentionné dans ce chapitre peut servir à l'échantillonnage d'une strate, à partir de méthodes plus simples comme l'EAS ou le SYS, jusqu'aux méthodes plus complexes comme l'échantillonnage PPT, par grappes, à plusieurs degrés ou à plusieurs phases