

## Chapitre I : L'échantillonnage en écologie.

### Introduction

La biodiversité et ses composantes sont bien complexes à appréhender de manière analytique que ce soit au niveau local, national ou régional, et leur évaluation nécessite diverses techniques d'études et de suivi faisant appel à diverses disciplines scientifiques.

En écologie, il est généralement impossible de mesurer une ou des caractéristiques sur l'ensemble des unités d'un groupe d'intérêt. Ceci peut résulter de plusieurs causes :

- Les contraintes de temps ;
- Les contraintes d'argent ;
- Manque de personnel qualifié ;
- Ou encore, il peut être impossible de mettre la main sur l'ensemble des individus d'une population.

Donc on ne mesure pas "tout" un système biologique (ce ne serait plus un échantillonnage, et ce serait, de plus, techniquement et conceptuellement impossible), mais bien un fragment de l'ensemble, prélevé pour juger de certaines propriétés de ce tout.

Selon Colin (1970) : "un échantillon est un fragment d'un ensemble prélevé pour juger de cet ensemble". De nombreuses méthodes d'observations et de mesures appliquées à de tels fragments peuvent être proposées, adaptés à chaque cas particulier en vue d'obtenir une représentation satisfaisante de l'objet étudié. Il faut exprimer explicitement de quelle propriété on veut juger avant de pouvoir concevoir un plan d'échantillonnage. L'échantillonnage doit être adapté à tester l'hypothèse que l'on a fait, à une échelle spatiale et temporelle donnée, sur la structure ou la dynamique du système biologique étudiée. Il est impératif de prendre le temps de planifier son échantillonnage.

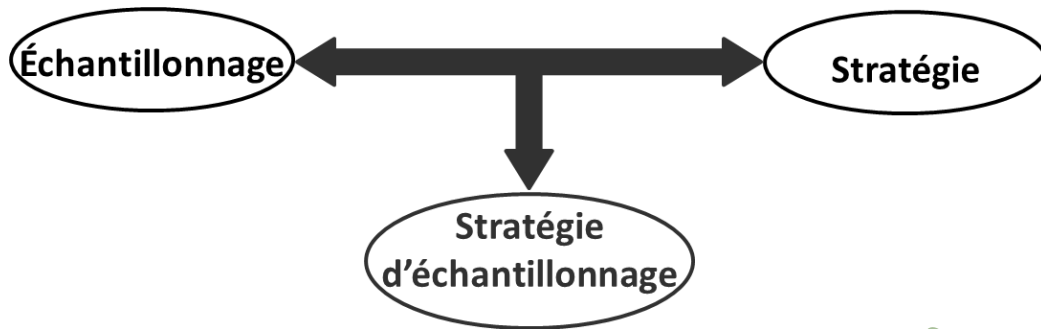
### 1. Echantillonnage

#### 1.1. Principes généraux

L'échantillonnage est, au mieux, adapté à tester l'hypothèse que l'on a fait, à une échelle spatiale et temporelle donnée, sur la structure ou la dynamique du système biologique étudiée. Il est donc important de prendre le temps de planifier son échantillonnage.

La mise en place d'un plan d'échantillonnage est conditionnée par le choix du problème et la façon de le poser :

- Le choix des variables à étudier ;
- Le choix des échelles d'observation et du découpage de l'objet (zone d'étude) ;
- Le choix des méthodes de traitements des données recueillies. (par exemple réfléchir à l'exploitation statistique des résultats avant de commencer l'étude).



- L'objectif de l'étude (question/hypothèse préalablement correctement posée) ;
- Les contraintes naturelles (hétérogénéité spatiale, variété d'échelles significantes, etc.) ;
- Les contraintes techniques (temps disponible, fiabilité des mesures, etc.) et financières ;
- Les contraintes mathématiques (qualité des données et des instruments mathématiques, etc.).

**Mode opératoire = Plan d'échantillonnage.**

## 1.2. Unité(s) et élément(s) d'échantillonnage

En statistiques descriptives, l'élément d'échantillonnage peut être une entité concrète comme un individu, un système, un objet, ... etc. ou abstraite comme une relation comportementale sur laquelle on mesure ou on observe la variable étudiée (Figure 1).

❖ **Population-cible** : ensemble des éléments visés, en principe, par l'échantillonnage.

*Quelle est la population-cible ?* Il s'agit là de la population sur laquelle on aimerait bien que les conclusions de l'étude portent.

❖ **Population statistique** : ensemble des éléments effectivement représentés par l'échantillonnage. Les éléments qui la composent se caractérisent par au moins une caractéristique commune et exclusive qui permet de les distinguer sans ambiguïté.

*Quelle est la population statistique ?* Il faut mentionner la ou les caractéristiques qui permettent de la distinguer de toute autre population statistique.

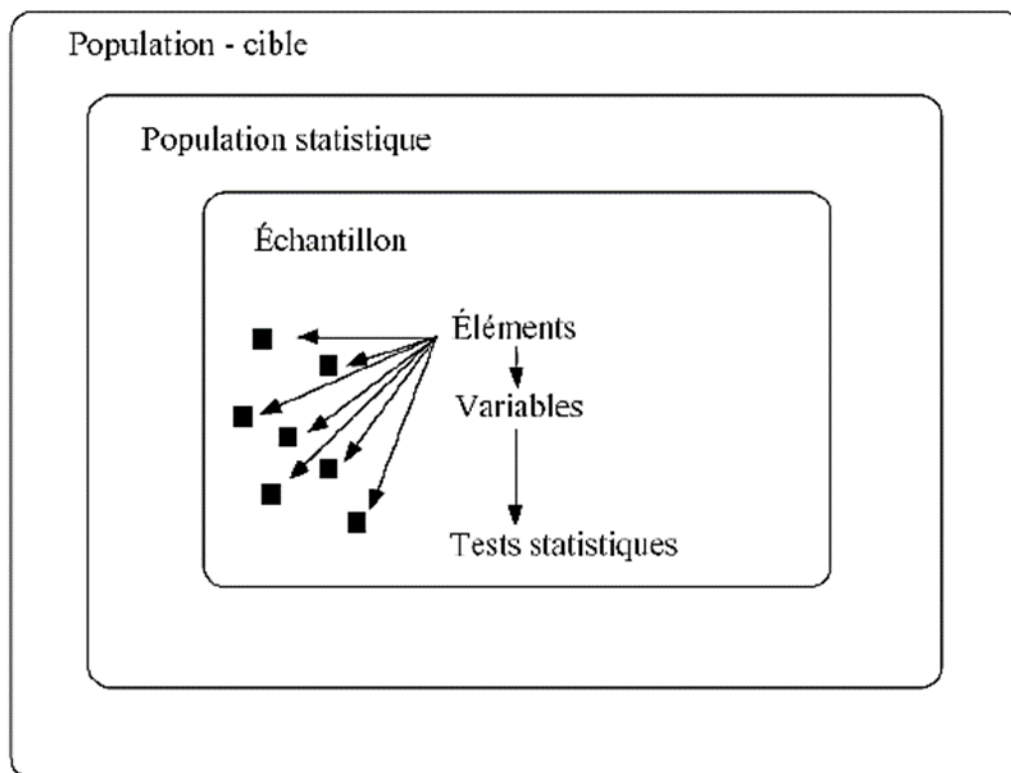
❖ **Échantillon (sample)** : fragment d'un ensemble prélevé pour juger de cet ensemble. Fraction de la population statistique sur laquelle des mesures sont faites pour connaître les propriétés de cette population.

*Quel est l'échantillon ? Quel est son effectif ?*

❖ **Élément** : membre d'une population statistique (spécimen, prélèvement d'eau, individu...)

*Quel est l'élément ?* Il faut le définir de manière à pouvoir le reconnaître sans ambiguïté.

❖ **Variable** : Toute caractéristique mesurable ou observable sur un élément d'échantillonnage (*var. propre*) ou sur son environnement (*var. associée*) = Descripteur, caractère, attribut, observation, trait, profil (en géophysique), item (dans les questionnaires de psychologie), stimulus (en étude du comportement).



**Figure 1 : Unités et éléments d'échantillonnage.**

## 2. Classification des descripteurs

Les variables pouvant intervenir dans une description de structure ou de fonctionnement d'un objet étudié sont très nombreuses. Dans chaque étude statistique il est très important de considérer la nature des données (observations, caractères, attributs) que l'on va tester ; car c'est d'elle que dépend la nature des opérations possibles et donc des statistiques utilisables dans chaque situation. Il est donc primordial de préciser la nature de chaque variable, ou caractère. Ces variables peuvent être classés en différentes catégories.

### 2.1. Descripteurs qualitatifs

Sont des catégories définies sans assignation d'une mesure ni même d'un caractère permettant de les ordonner les unes par rapport aux autres. Dans ce type de variable les modalités ne sont pas quantifiables (pas mesurables) (couleur des yeux, douleur, ...). Ce sont des noms ou ce qui revient au même des sigles ou des codes. Les différentes modalités ne sont pas ordonnables. Attention, même si les modalités sont des codes numériques, les opérations sur les modalités n'ont aucun sens.

#### Exemple :

- Type de relief avec trois modalités (plaine, montagne, plateau), ou encore taille d'une niche écologique avec quatre modalités (petite, moyenne, grande, très grande).
- Les différents taxons constituant un peuplement. Pour chaque taxon considéré, le descripteur est la présence ou absence.
- Couleur, catégorie, condition de culture ou d'élevage, espèce...

## 2.2. Descripteurs quantitatifs

Caractère auquel on peut associer un nombre c'est-à-dire que l'on peut "mesurer" (grandeur mesurable). Ce sont des quantités véritables, pour lesquelles on peut déterminer des rapports et des différences. Cette définition concerne un très grand nombre de descripteurs utilisés en écologie et qui mesure des abondances, des taux, pourcentage, volume, biomasse, mesure de poids, longueur, concentration, fréquence,... etc.

## 2.3. Descripteurs semi-quantitatifs (Ordinaux)

Descripteurs ordinaux sont définis par l'existence d'une relation d'ordre (plus petite ou plus grande ; ou bien antérieure ou postérieure, etc.) sans toutefois qu'il soit possible de mesurer une distance entre deux états distincts.

**Exemple :** Les stades de développement d'une espèce. Pour un organisme à croissance continue, un ensemble de classe d'âge ou de taille délimitées arbitrairement. Stades de succession d'un peuplement naturel le long d'un gradient spatio-temporel.

**N.B. :** Les descripteurs qualitatifs peuvent devenir semi quantitatifs, si on les classe selon leurs fréquences (classement de ces espèces par fréquences décroissantes).

En plus de ces trois catégories de descripteurs, il existe une dernière catégorie ; il s'agit de la catégorie de descripteurs complexes ou synthétiques.

## 2.4. Descripteurs synthétiques (Complexes)

Les descripteurs cités précédemment sont des descripteurs simples, c'est à dire, caractérisés, pour chaque observation, par un seul nombre ou par la spécification d'une modalité. Cependant, les descripteurs complexes, nous permettent de rendre compte de plusieurs observations simples dans le même plan d'échantillonnage.

**Exemple :** Soit un ensemble d'espèces (chacune caractérisée par son abondance relative en une station). On calcule un indice de diversité (descripteur quantitatif), et on établit la loi de décroissance des abondances des espèces rangées de la plus abondante à la plus rare (descripteur semi-quantitatif).

## 3. Choix des descripteurs

Dans les études écologiques, les descripteurs utilisés sont extrêmement divers. Le choix de ces derniers dépend du type du modèle descriptif ou explicatif attendu en fin d'analyse, c'est à dire, du pré-modèle. Ci-dessous, sont cités quelques exemples de descripteurs :

### 3.1. Descripteur d'occupation de l'espace-temps

Peuvent être :

- ☞ Qualitatifs, présence ou absence d'un taxon et indication du type d'occupation du milieu (espèce endogée vie dans le sol ou épiphyte, planctonique,... etc.) ;
- ☞ Semi-quantitatifs (échelle d'abondance/dominance) ;
- ☞ Quantitatifs (biomasses, effectifs d'organismes par unité de volume ou de surface du biotope).

### 3.2. Descripteurs biométriques et démographiques

C'est des descripteurs nécessaires à l'application des modèles dynamiques de populations (ex. : démographie des populations).

### 3.3. Descripteur structuraux

Outre les structure spatio-temporelle et démographiques, en écologie il existe aussi des structures liées à la répartition de la biomasse en espèces distinctes (distribution des individus par espèces, diversité spécifique), des structures trophiques, ...etc.

Ces descripteurs peuvent être quantitatif, semi quantitatif ou qualitatif. La structure trophique est décrite par les biomasses relatives des producteurs, des consommateurs et des décomposeurs. La structure spatio-temporelle : Stratification de la végétation, succession de végétation où chaque stade prépare l'installation du suivant.

### 3.4. Descripteurs systématiques

Les plus fréquents sont ceux qui rendent compte de la dynamique d'une biomasse, d'une espèce ou d'un élément chimique (allongement des rameaux). On retrouve les descripteurs biométriques et démographiques, s'il s'agit d'un modèle de dynamique d'une population.

## 4. Types d'échantillonnage

### 4.1. Réalisation d'un échantillon

Un échantillon est qualifié de représentatif de la population statistique lorsque chaque élément de la population a une probabilité connue et différente de zéro d'appartenir à l'échantillon. Le cas le plus connu est celui où chaque unité de la population a la même probabilité d'appartenir à l'échantillon. Un échantillon sera à choix raisonné lorsque les unités seront sélectionnées en fonction de critères préétablis choisis par l'expérimentateur.

Un plan d'échantillonnage est un protocole de sélection des éléments de la population statistique pour obtenir un échantillon représentatif (ou aléatoire). Il doit conduire à un maximum de précision avec un minimum d'effort (ou de coût).

Lorsque l'échantillon n'est pas représentatif il donne une valeur biaisée de l'estimateur. Par exemple, soit un échantillon de 200 poissons extrait d'une population statistique de 400 poissons si les 200 poissons ont tous 1 an et que l'âge moyen des 400 poissons est de 1 an, la valeur de l'estimateur ici l'âge donné par l'échantillon sera sans biais ou encore exact ou encore juste. Le biais est donc la différence (significative) entre l'espérance mathématique ou la valeur moyenne de l'estimateur et la valeur estimée par l'échantillon.

### 4.2. Méthodes d'échantillonnage en écologie

Il existe deux grandes catégories de méthodes d'échantillonnage, à savoir : les méthodes d'échantillonnage non probabilistes (empiriques) et les méthodes d'échantillonnage probabilistes.

- ☞ Méthodes d'échantillonnage non probabilistes : les éléments sont inclus dans l'échantillon sans probabilité connue. Une évaluation de l'exactitude des résultats ne peut être faite.
- ☞ Méthodes d'échantillonnage probabiliste : les éléments sélectionnés ont une probabilité connue de faire partie de l'échantillon.

Le choix d'un mode d'échantillonnage dépend de notre problématique. Chaque méthode possède ses avantages et ses inconvénients, ses propres caractéristiques techniques de mise en place et d'analyse des résultats. Donc le choix d'une méthode d'échantillonnage appropriée tient compte de la précision, du but recherché, ainsi que des contraintes (temps, ressources financières, etc.).

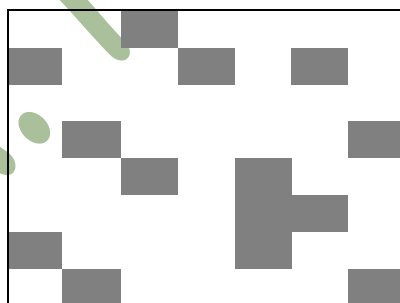
Les méthodes d'échantillonnage non probabiliste ; dont : l'échantillonnage de commodité, à l'aveuglette, volontaire, au jugé et par quotas, bien que moins coûteuse et plus pratiques, elles sont moins exactes. Beaucoup d'experts dans le domaine recommandent de s'en éloigner ; vu qu'elles supposent que les caractéristiques de la population sont distribuées de manière aléatoire, chose qui n'est pas vérifiée dans tous les cas.

Dans ce cours nous aborderons seulement l'échantillonnage probabiliste, à savoir : l'échantillonnage aléatoire, systématique, stratifié et l'échantillonnage en grappes.

#### 4.2.1. Echantillonnage aléatoire simple (au hasard)

L'échantillonnage aléatoire simple est une méthode qui consiste à prélever au hasard et de façon indépendante « n » unités d'échantillonnage d'une population de « N » éléments (Figure 3).

Les échantillons sont répartis au hasard. Chaque point dans l'espace étudié a donc une chance égale d'être échantillonné. Les données ainsi récoltées ne sont pas biaisées. A partir d'une carte ou d'une photographie aérienne ; une pratique largement utilisée consiste à utiliser une grille pour les choisir de manière plus aisée. Une méthode garantissant sécurité et représentativité consiste à dresser la liste complète et sans répétition des éléments de la population, à les numéroté, puis à tirer au sort « n » d'entre eux à l'aide d'une table de nombres aléatoires ou de tout autre système générant des chiffres aléatoires. Chaque élément sélectionné peut être remis dans la population après son tirage pour éventuellement être choisi une deuxième fois : on parle alors d'échantillonnage avec remise. Cette méthode se prête aux analyses statistiques, mais elle demande de prélever un grand nombre d'échantillons.



**Figure 3 :** échantillonnage aléatoire simple (Au-hasard).

Il est parfois difficile en écologie d'effectuer un échantillonnage aléatoire simple conforme à la définition. En effet, pour qu'aucune erreur systématique ne s'introduise dans le prélèvement de l'échantillon, il faut :

- ☞ dresser la liste complète et sans répétition des éléments de la population ;
- ☞ les numéroté de 1 à N ;
- ☞ procéder, à l'aide d'une table de nombres aléatoires, au tirage au sort de n unités différentes.



Il faut signaler que l'EAS s'avère facile à réaliser si la population n'est pas trop grande et si les éléments sont facilement identifiables et repérables. Toutefois, dans la majorité des autres cas, le processus s'avère difficile, voire irréalisable.

### Exemple :

Nous désirons tirer au hasard un échantillon de **6 individus** dans une population de **300 individus** ; On numérotés de 001 à 300 dans une base de sondage.

Pour faire on utilise une table de nombres aléatoires (Tableau I) : on lira les 3 derniers chiffres de la troisième colonne de la table, en lisant de haut en bas, en ne retenant que les nombres compris entre 001 et 300, et en rejetant tout nombre qui apparaîtrait pour la deuxième fois (pour ne pas tirer deux fois le même individu). Enfin, on choisit le point d'entrée en décidant de commencer la lecture à la troisième ligne.

**Tableau I** : La table de nombres aléatoires (Hill, 1977)

17406	39516	24449	74015	43890	55118	27902	39548	72071	59327
73855	89884	56589	48471	36709	11110	31920	17626	67691	93202
36164	92283	27244	05702	62405	73041	93132	01371	08683	21829
78994	18171	59840	62624	74518	99822	00293	88084	30716	53590
83212	53819	70820	03945	68127	93070	44870	67948	01338	97258
28619	99766	49157	09339	55573	6683	87189	89254	17652	67314
85240	11095	46806	33176	51940	23127	90561	36469	85603	83297
98448	89051	17846	85123	76892	30908	04175	77862	13924	20099
46186	14176	89969	37701	11456	20210	32542	32146	77162	18567
35549	53554	76295	16048	38190	92244	56808	39717	83105	99486
34904	46147	60505	54418	78345	50570	58052	81085	26388	10243
06953	03928	58301	79506	99089	00231	95700	55663	45833	87967
85680	11023	45747	55475	97488	98531	02054	53841	70698	41772
99513	65174	23148	88646	24301	27909	83026	50673	75087	61801
92336	80844	34686	44894	34011	82172	97959	57568	47239	35986
53784	79567	64673	68622	87745	02165	94061	09140	63912	24787
86457	26826	20293	37348	76714	26917	68221	08181	73036	39186
26903	50718	03261	22399	19519	03808	82688	93418	75681	96105
17307	67935	72471	95209	78716	21582	93147	80487	23588	97851
50990	85208	68410	36019	02200	96138	53902	91022	17194	45198

### Avantages de l'EAS

- ☞ Il est connu et accepté universellement ;
- ☞ Les estimateurs ne sont pas biaisés. Il est possible de généraliser, c'est-à-dire de déterminer jusqu'à quel point les résultats obtenus avec l'échantillon s'appliquent à la population ;
- ☞ Le calcul des estimateurs s'avère facile et la majorité des banques de programmes informatiques se prêtent à ce plan ;
- ☞ Il n'est pas nécessaire d'élaborer une planification sophistiquée de la collecte et du traitement des données, dans la mesure où l'on est assuré que le processus de sélection est réellement au hasard non biaisé par une tendance, consciente ou non.

### Inconvénients de l'EAS

- ☞ L'inventaire des éléments de la population est souvent une opération difficile et parfois même impossible.
- ☞ Le protocole de sélection des unités d'échantillonnage est peu commode comparativement à celui de l'échantillonnage systématique.
- ☞ Son efficacité s'avère souvent très médiocre car on n'utilise pas les informations apportées par des expériences antérieures, ou ne collecte pas préférentiellement des éléments privilégiés dont l'accès ou l'examen s'avère plus commode ou plus économique, on ne module pas la probabilité de sélection des éléments en fonction de leur poids ou de leur importance relative.
- ☞ Coûteux et irréalisable pour de grandes populations

### 4.2.2. Echantillonnage stratifié

Ce type d'échantillonnage est particulièrement utilisé quand l'aire étudiée est hétérogène.

Il consiste à subdiviser une population hétérogène en sous-populations ou "strates" plus homogènes, mutuellement exclusives et collectivement exhaustives. La population hétérogène d'effectif "N" est ainsi découpée en "K" strates plus homogènes d'effectif "N<sub>h</sub>" de telle sorte que  $N = N_1 + N_2 + N_3 + \dots + N_K$ . Un échantillon indépendant est par la suite prélevé au sein de chacune des strates en appliquant un plan d'échantillonnage aux choix de l'écologiste.

Son application soulève deux questions : *comment construire les strates ? Quel plan d'échantillonnage adopter dans chacune des strates ?*

- ☞ La première étape consiste à choisir un critère de stratification (quantitatif ou qualitatif). Le meilleur stratificateur est la variable étudiée lors d'un recensement antérieur ou alors un autre caractère en corrélation avec la variable étudiée ;
- ☞ La 2<sup>ème</sup> étape nombres de strates : D'une façon générale, l'augmentation du nombre de strates s'accompagne d'une amélioration de la précision (mais au-delà d'un nombre  $x$  de strates  $\approx 6$ , la précision s'avère rarement profitable).
- ☞ L'effort de l'échantillonnage peut varier d'une strate à l'autre. On peut conserver la même fraction d'échantillonnage dans chaque strate, comme on peut moduler l'effort d'échantillonnage afin de minimiser le coût total de l'opération pour une précision donnée ou maximiser la précision pour un coût total fixé.

**Exemple :** Pour évaluer l'importance des populations de grands herbivores dans le parc national de Waza (Cameroun), le Parc a été divisé en 5 strates ; correspondant aux différents types de végétation. Dans chaque strate des itinéraires échantillons ont été répartis régulièrement selon le principe de l'échantillonnage systématique. L'utilisation de l'échantillonnage stratifié s'avère dans ce cas tout à fait justifiée, puisque la densité des grands herbivores varie en fonction de la couverture végétale.

### Avantages du plan

- La stratification peut entraîner des gains de précision appréciables ;
- Les estimateurs de l'échantillonnage aléatoire simple stratifié ne présentent aucun biais ;
- Ce plan constitue une solution avantageuse aux problèmes de variations de l'effort d'échantillonnage ;

### Inconvénients du plan

- La majorité des programmes informatiques de traitement statistique des données ne sont pas conçus pour ce plan.



### 4.2.3. Echantillonnage systématique

Ce type d'échantillonnage consiste à répartir les échantillons de manière régulière. L'échantillonnage systématique est une technique qui consiste à tirer au hasard un  $i^{\text{ème}}$  élément, situé entre le premier et le  $p^{\text{ème}}$  de la population puis à prélever systématiquement le  $(i + p)^{\text{ème}}$ ,  $(i + 2p)^{\text{ème}}$ ,  $(i + 3p)^{\text{ème}}$ , ...,  $(i + (n-1)p)^{\text{ème}}$  élément de la population. Les rangs des  $n$  unités sont ainsi en progression arithmétique dont la base est un nombre aléatoire  $i$  et la raison un nombre  $p$  calculé de telle sorte que l'échantillon se répartisse uniformément sur toute population. Contrairement à l'EAS, les unités ne sont pas prélevées de façon indépendante puisque le choix du 1<sup>er</sup> élément détermine la composition de tout l'échantillon. Le protocole est très simple lorsque les éléments de la population sont facilement accessibles et en nombre connu. Il suffit :

1. de choisir l'effectif  $n$  de l'échantillon
2. de calculer la raison  $p$  ( $p = N/n$ )
3. de tirer au hasard un  $i^{\text{ème}}$  élément que l'on considère comme le premier.
4. de prélever un élément toutes les  $p$  unités.

#### ❖ Avantages

- L'échantillonnage systématique s'avère beaucoup plus commode à préparer et à exécuter que l'E.A.S.
- Si l'on est assuré que tous les éléments de la population se présentent dans un ordre aléatoire, ce plan est équivalent à l'E.A.S. Il cumule alors ses propres avantages.
- Si la population présente des phénomènes d'auto corrélation, c'est-à-dire, les éléments qui se suivent dans la série ont un comportement assez semblable au niveau des variables étudiées, ce plan se révèle plus efficace que E.A.S. qui comporte par le fait du hasard des relevés très rapprochés et d'autres très éloignés, ce qui entraîne des redondances et des défauts d'information dans la répartition de la variable étudiée.
- Si les éléments de la population présentent dans leur séquence une tendance linéaire au niveau de la variable étudiée, l'échantillonnage systématique s'avère plus efficace que l'E.A.S.

#### ❖ Inconvénients

- Mise à part la répartition spatio-temporelle des relevés, le protocole de sélection des unités d'échantillonnage impose une énumération de tous les éléments de la population afin de retenir un tous les  $p$  (raison). Ce processus s'avère souvent très laborieux comparativement à celui de l'échantillonnage par degré.
- Lorsque la série d'éléments présente au niveau du caractère étudié des variations périodiques insoupçonnées, ce plan s'avère peu efficace.
- Ce plan ne permet pas de collecter préférentiellement des éléments privilégiés dont l'examen ou l'accès s'avère plus commode, il n'ajuste pas, la probabilité de sélection des éléments en fonction de leur poids ou de leur importance relative.