



Statistique

Dr. Thérèse N. Mané
Chargée de cours
tndiaye18@gmail.com

Plan

1. Introduction : définition et objectifs
2. Méthodes de collecte de données
3. Tableaux et représentation graphique des données
4. Caractéristiques des distributions à une variable

1. Définition et objectifs

La statistique est une discipline scientifique qui permet :

- de collecter des données ;
- d'extraire les informations contenues dans un ensemble de données ;
- d'analyser et d'interpréter des données.

L'objectif est de présenter, décrire et résumer des données observées pour mieux les analyser.

2. Méthodes de collecte de données

2.1. Les méthodes d'information

- les **enquêtes exhaustives** : elles ont pour but de collecter des informations sur chacun des membres d'une population.

Avantage : on peut collecter le maximum d'informations possibles.

Inconvénients : le coût, la mise en œuvre, délai de publication des résultats long...

- les **enquêtes partielles** : elles ont pour but de collecter des informations sur une partie de la population afin de connaître les tendances c'est-à-dire les résultats proches de la réalité ou étendus à l'ensemble de la population.

Avantage : moins onéreux, plus rapide.

Inconvénients : résultats contestables.

- les **monographies** : Etude exhaustive et approfondie sur un sujet précis et limité

2.2. Les instruments de l'observation

2.2.1. Le questionnaire

C'est un outil méthodologique de collecte de données à partir d'une série de questions qui s'enchaînent de façon ordonnée.

Le questionnaire est structuré autour de l'objectif d'enquête qui est préalablement défini.

Il doit permettre de répondre aux questions :

Qu'est-ce que je veux savoir ?

Pourquoi je veux savoir cela ?

En quoi ces informations sont-elles nécessaires pour mon enquête ?

2.2.2. Structure du questionnaire : quelques impératifs

Choisir le type de questions (ouverte, fermée à choix unique, fermée à choix ouverte, numérique. Dans le cas de questions fermées à choix uniques, énumérer le plus de réponses possibles.

Choisir la forme des questions (forme interrogative, forme affirmative ; éviter la forme négative)

Regrouper les questions par thématique

Noter un nombre raisonnable de questions

Poser des questions claires et adaptées au niveau intellectuel des individus interrogés ; éviter de poser des questions sans intérêt, embarrassantes, indiscrètes et accusatrices ;

Mentionner les questions d'identification des répondants selon la pertinence (sexe, âge, situation matrimoniale, niveau de scolarité, catégorie professionnelle, lieu de résidence, etc.) et la date de remplissage du questionnaire

Au besoin, préparer une petite introduction écrite ou orale pour expliquer l'objectif de l'enquête. Exemple : Soucieux d'améliorer vos conditions de vie à l'UGB, nous avons mis en place cette enquête de satisfaction afin de mieux répondre à vos attentes. A cet effet, nous vous serions reconnaissants de bien vouloir participer à cette enquête en remplissant ce questionnaire. Nous vous garantissons que les informations seront traitées de façon confidentielle et anonyme.

...et une conclusion sous forme de remerciement (écrite ou orale). Exemple : Nous vous remercions pour le temps que vous nous avez accordé pour répondre à nos questions.

2.3. L'enquêteur

Le questionnaire peut être envoyé par voie postale ou par mail, il est aussi de plus en plus élaboré en ligne (Google Forms). Mais, dans le cadre de travaux de recherche, il est préférable que l'enquêteur aille au-devant des individus sélectionnés pour dérouler son questionnaire. Il devra alors faire preuve de patience et de psychologie.

2.4. L'échantillonnage

2.4.1. Vocabulaire

- **Population-mère** : l'ensemble de la population sur laquelle porte l'étude

Exemples : les étudiants de l'UGB, les agriculteurs de Saint-Louis, les lycéens de la région de Dakar, l'ensemble des forêts du Sénégal, l'ensemble des animaux du Parc Niokolo Koba, etc.

- **Individu ou unité statistique** : un élément de la population, par conséquent, l'ensemble des individus constitue la population

Exemple : un étudiant de l'UGB, un agriculteur de Saint-Louis, un lycéen de la région de Dakar, une forêt du Sénégal, un animal du PNNK, etc.

- **Echantillon** : sous-ensemble représentatif d'une population-mère auprès duquel sera menée l'enquête
- **Echantillonnage** : l'ensemble des méthodes qui permettent de constituer l'échantillon

2.4.2. Les méthodes d'échantillonnage

Il existe des méthodes d'échantillonnage aléatoires et non aléatoires.

❑ Les méthodes aléatoires

L'échantillonnage est dit aléatoire lorsque l'on connaît les chances que chacun des individus de la population a d'être prise, et si chacun des individus constituant l'échantillon est choisi au hasard. Les méthodes d'échantillonnage aléatoire supposent l'existence d'une liste complète des individus de la population, ou au moins une liste la plus complète possible appelée Base de Sondage.

- L'échantillonnage aléatoire simple (EAS)

Dans un EAS, chaque individu de la population a une chance égale d'être incluse dans l'échantillon. Il nécessite une liste de tous les individus de la population. D'abord, on numérote tous les individus de la population de 1 à N. Ensuite on tire des nombres au hasard entre 1 et N jusqu'à ce qu'on ait réuni le nombre d'individus désiré pour former l'échantillon. Habituellement, un numéro ne peut être utilisé plus d'une fois, même s'il apparaît plusieurs fois dans le tirage au sort.

Application : Pour une population totale de 10 000 étudiants, on veut constituer un échantillon de 250 étudiants. Pour ce faire, on attribue un numéro à chaque étudiant inscrit dans la base de données. Un ordinateur va ensuite générer au hasard 250 numéros correspondant aux 250 étudiants qui vont constituer l'échantillon.

- L'échantillonnage systématique

La méthode de l'échantillonnage systématique consiste à prendre une unité à intervalle constant. L'intervalle constant est appelé le pas du système.

On numérote les individus de la population de 1 à N.

Ensuite, on calcule le pas du système. Si le résultat ne donne pas une valeur entière, on arrondit à l'entier le plus près.

$$\text{Pas} = N/n$$

N = Taille de la population (Ex : 10 000 étudiants)

n = Taille de l'échantillon (Ex : 250 étudiants)

Enfin, on choisit un numéro au hasard parmi la population totale (origine choisie au hasard) qui sera le 1^{er} nombre inclus dans l'échantillon et on applique le pas.

Application : Pas du système = $10000/250 = 40$

Si le numéro choisit au hasard est 24 alors l'échantillon sera constitué des numéros 24 64 104 144 184
...

Si le numéro choisit au hasard est 10 alors l'échantillon sera ainsi constitué : 10 50 90 130 170 210 250
...

Population : les étudiants du M1 LEA = 40 étudiants

Echantillon = 5

Pas du système = Population/Echantillon donc pas = $40/5$ Pas=8

Choisir un numéro entre 1 et 40 au hasard

Exemple 1 : Si on choisit au hasard le numéro 2 (qui sera systématiquement pris en compte) alors l'échantillon sera constitué par les numéros : **2** 10 18 26 34 correspondant à 5 étudiants

Exemple 2 : Si on choisit au hasard le numéro 11 (qui sera systématiquement pris en compte), alors l'échantillon sera constitué par les numéros : 3 **11** 19 27 35 correspondant à 5 étudiants.

Exemple 3 : Si on choisit le numéro 1 (qui sera systématique pris en compte), alors l'échantillon sera constitué par les numéros **1** 9 17 25 33

NB : Pour ne pas dépasser la taille de la population, il faut retrancher le pas à partir du premier numéro choisi au hasard (cf. exemple 2)

- L'échantillonnage stratifié

Une strate est une subdivision de la population totale en groupes ayant une ou plusieurs caractéristiques communes. Les strates permettent de regrouper les individus ayant des caractéristiques similaires. Chaque strate devient alors une population indépendante, une taille d'échantillon est déterminée pour chacune d'elle.

Application : Pour constituer notre échantillon de 250 étudiants, on décide de regrouper les étudiants par UFR. Les UFR représentent ici les strates. Les 10000 étudiants de l'UGB sont répartis de la façon suivante :

UFR LSH = 3500 étudiants

UFR SJP = 3000

UFR SEG = 1500

UFR SAT = 1000

UFR SEFS = 500

UFR S2ATA : 500

Il faut alors déterminer pour chaque UFR le nombre d'étudiants à interroger.

Application : on veut déterminer le nombre d'étudiants de l'UFR LSH qui doivent faire partie de l'échantillon. Pour ce faire on pose les opérations suivantes :

1^e opération : $(3500 \times 100) / 10000 = 35\%$

2^e opération : $(35 \times 250) / 100 = 87,5$ étudiants de l'UFR LSH

Cela veut dire que 87,5 étudiants de l'UFR LSH doivent figurer dans l'échantillon.

$$\text{UFR SEG : } (1500 \times 100) / 10000 = 15\% \\ (15 \times 250) / 100 = 37,5 \text{ étudiants}$$

$$\text{UFR SJP : } (3000 \times 100) / 10000 = 30\% \\ (30 \times 250) / 100 = 75 \text{ étudiants}$$

$$\text{UFR SEFS : } (500 \times 100) / 10000 = 5\% \\ (5 \times 250) / 100 = 12,5 \text{ étudiants}$$

$$\text{UFR S2ATA} = (500 \times 100) / 10000 = 5 \\ (5 \times 250) / 100 = 12,5 \text{ étudiants}$$

$$\text{UFR SAT : } (500 \times 100) / 10000 = 5 \\ (5 \times 250) / 100 = 12,5 \text{ étudiants}$$

- L'échantillonnage en grappes

Une grappe est un groupe de données similaires entre elles mais qui peuvent être différentes des autres groupes. Lorsque la liste de toutes les grappes est connue, on numérote les grappes puis on prend de façon aléatoire simple un certain nombre de grappes, et toutes les unités de ces grappes constituent l'échantillon.

Application : Pour constituer l'échantillon de 250 étudiants, on sélectionne au hasard 1 UFR parmi les 6 UFR de l'UGB. Seuls les étudiants de cet UFR peuvent alors faire partie de l'échantillon.

Avantage : Moins coûteux, plus rapide

Inconvénient : pas toujours représentatif de la population, ne permet pas parfois d'avoir la taille de l'échantillon.

❑ Les méthodes non aléatoires

Contrairement aux méthodes précédentes, les individus qui composent l'échantillon ne sont pas tous pris au hasard.

- L'échantillonnage accidentel ou à l'aveuglette

C'est un concours de circonstances qui fait qu'un individu se retrouve dans l'échantillon en d'autres termes, l'individu s'est trouvé au bon endroit au bon moment. Toutes les unités qui n'avaient pas de raison d'être là ne pouvaient faire partie de cet échantillon.

- L'échantillonnage par participation volontaire

C'est l'individu qui décide de faire partie ou non de l'échantillon. Méthode surtout utilisée dans le cadre d'expériences en médecine ou en psychologie.

- L'échantillonnage par quotas

L'échantillonnage par quotas est similaire à l'échantillonnage stratifié à la seule différence que pour atteindre son quota (nombre d'individus requis) dans chaque strate, l'enquêteur choisira lui-même de façon arbitraire (à l'aveuglette) les individus qui feront partie de l'échantillon.

- L'échantillonnage au jugé

Il consiste à prendre les unités de l'échantillon parmi celles que l'on considère comme des unités typiques de la population. Ce jugement peut découler de sa propre expérience ou de l'analyse critique de la situation.

Application : Pour constituer l'échantillon de 250 étudiants, l'enquêteur décide de sélectionner uniquement des étudiants qui ont passé au moins 3 ans à l'université car il juge qu'ils sont plus à même de décrire l'évolution des conditions de vie.

- L'échantillonnage à plusieurs phases

Dans ce type d'échantillonnage, les données sont collectées auprès d'un large échantillon d'unités de la population puis de façon plus détaillée auprès d'un sous-échantillon de ces unités. Les échantillons sont tirés de la même base de sondage. La sélection d'une unité dans la deuxième phase est conditionnelle à sa sélection dans la première phase.

Application : La base de sondage contient tous les étudiants de l'UGB. Mais on veut connaître une information précise par rapport aux étudiants qui viennent de la région de Dakar. On obtient alors un premier échantillon. Une fois ce premier échantillon prélevé, on peut obtenir un autre échantillon plus petit qui regroupera les étudiants résidant dans le département de Dakar.

- En définitive, il faut toujours essayer de constituer un échantillon qui soit le plus représentatif de la population étudiée, puisque la validité des résultats et des conclusions de l'étude en dépend.
- Le choix d'une méthode d'échantillonnage repose sur le jugement du chercheur et sur des critères inhérents à l'objet de l'étude. La précision des résultats, le temps requis, l'importance de l'étude et le coût sont des facteurs qui peuvent influencer sur le choix de la méthode d'échantillonnage.
- Dans les sondages où l'on procède par échantillonnage aléatoire, il y a toujours des unités sélectionnées pour lesquelles on obtient pas de réponse pour cause d'absence, de refus de donner son opinion, etc.
- L'absence de réponse influe sur les résultats de l'étude et peut entraîner une sous représentativité de certains groupes de la population.
- Comme la taille de l'échantillon a une influence sur la marge d'erreur des estimations envisagées, il est important d'insister auprès de chacune des personnes ou individus sélectionnés afin de minimiser le nombre de non-répondants.

Application : indiquer la méthode d'échantillonnage utilisée dans chaque cas (aléatoire simple, systématique, stratifié, grappes, quota, volontaire, au jugé, accidentel)

1. Un enseignant souhaite sélectionner trois de ses élèves pour participer à un concours. Pour cela, il met les noms de tous ses élèves dans un grand chapeau et, sans regarder en tire 3.
2. Une grande entreprise mène une enquête auprès de ses 100 salariés en prélevant de façon aléatoire 10 gestionnaires et 90 employés.
3. Une assistance téléphonique demande aux appelants de rester en ligne après avoir satisfait leur demande pour répondre à une courte enquête.
4. Chaque étudiant d'une même université possède un numéro d'identification. Une société de sondage utilise un logiciel générant aléatoirement 50 numéros d'identification afin que les étudiants ayant ce numéro participent à une enquête sur la qualité des restaurants universitaires.
5. Une responsable qualité d'un atelier de confection veut vérifier la conformité des chemises avant leur expédition auprès d'un client. Pour cela, elle choisit au hasard 2 cartons parmi les 10 constituant la commande et vérifie la conformité des chemises dans les cartons.
6. Un gérant d'une pizzeria veut vérifier la qualité des pizzas confectionnées, mais il n'a pas le temps de vérifier chaque pizza. Il décide de vérifier une pizza choisie aléatoirement parmi les 20 premières confectionnées chaque jour, puis de vérifier chaque 20^{ème} pizza pour constituer son échantillon.
7. Un professeur demande aux 5 premiers élèves qui entrent dans la classe de participer à une étude de recherche sur les habitudes de sommeil.

8) Le comité de pilotage de sécurité routière met en place des opérations de sécurité routière en choisissant aléatoirement des secteurs, puis en contrôlant systématiquement tous les conducteurs passant à ces endroits.

9) Un restaurateur propose à ses clients de remplir un questionnaire de satisfaction s'ils le souhaitent.

10) Pour constituer une équipe, un enseignant choisit au hasard 02 élèves de 6ème, 03 élèves de 5ème, 03 élèves de 4ème et 04 élèves de 3ème.

11) Un enquêteur utilise un programme informatique pour sélectionner de façon aléatoire 500 numéros de téléphones enregistrés et les personnes dont le numéro de téléphone a été choisi seront enquêtées.

12) Un chercheur cible des individus spécifiques susceptibles de l'aider à tester son questionnaire.