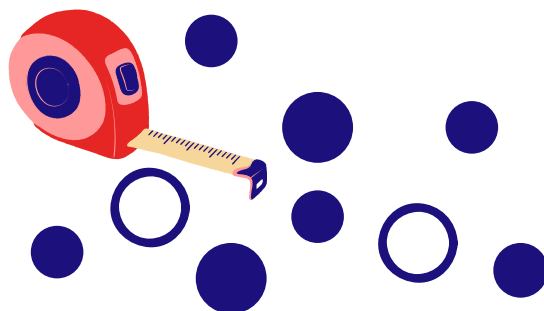


### Public visé

- Classe de 4ème
- Classe entière

### Durée

1h30 à 2h



### Objectifs et compétences

- **Objectifs de l'atelier :**
  - Réinvestir les notions de statistique au programme de la classe
  - Appréhender la notion de tirage au hasard
  - Décrire quelques méthodes de sondage
  - Aborder la notion de biais d'une méthode de sondage
  - Faire le lien avec le recensement national de la population
- **Concepts mathématiques et statistiques :** Estimation d'un paramètre (moyenne, moyenne pondérée) ; Échantillonnage ; Biais d'une méthode

### Liste du matériel

- Par élève :
  - 3 exemplaires de la feuille des 60 cercles (à imprimer en A4, contrôler que les diamètres des cercles sont 1, 2, 3, 4, 5 et 7 cm)
  - Calculatrice avec touche de génération d'entiers aléatoires
- Pour l'animateur :
  - Libre Office pour collecter et afficher les résultats en temps réel dans le classeur "Resultats"
  - Vidéo-projecteur
  - Feuille des 60 cercles à projeter
  - Procédures pour générer des entiers aléatoires avec les calculatrices des élèves (différentes selon les marques)

### Installation

- Élèves seuls
- Photocopies de la feuille de 60 cercles en quantité suffisante, 3 par élève
- Projection du classeur "Resultats" pour le recueil progressif des résultats

## Déroulement de l'atelier (résumé)

- Partie 1** Introduction générale
- Partie 2** Activité 1 : sondage empirique
- Partie 3** Activité 2 : sondage aléatoire simple
- Partie 4** Activité 3 : sondage par stratification
- Partie 5** Activité 4 : sondage à 2 degrés
- Partie 6** Synthèse et conclusion

## Partie 1 Introduction générale

On cherche à connaître la moyenne d'un caractère quantitatif dans une population. Pour cela, on peut recenser la population et mesurer le caractère sur chacun des individus mais cela peut se révéler coûteux, gourmand en temps et dans certains cas destructeur de la population. C'était le cas jusque en 2004 pour le recensement de la population française. Le recueil et l'exploitation des résultats demandaient beaucoup de temps et pouvaient ne plus être d'actualité lors de leur publication.

On peut procéder par sondage en prélevant un échantillon que l'on espère représentatif de la population.

L'objectif des activités est de décrire quelques méthodes de sondage sur un exemple simple, en sensibilisant à la notion de "tirage au hasard". Les avantages, les inconvénients et les biais des différentes méthodes de sondage seront mis en évidence.

La population à décrire est composée de 60 cercles de diamètres 2, 3, 4, 5 et 7 cm.

Chaque élève estime le diamètre moyen des disques de la feuille à partir d'un échantillon de 5 cercles en calculant, au dixième de millimètre près, le diamètre moyen des 5 cercles de son échantillon. L'ensemble des estimations du diamètre moyen des disques de la feuille est présenté à l'aide d'un logiciel.

**Projeter le tableau de saisie des résultats.**

### POINTS D'ATTENTION

- Donner les mesures des différents cercles afin de ne pas perdre de temps avec les mesures.
- **Ne pas** révéler le diamètre moyen réel avant le fin de la première activité.

## Partie 2 Activité 1 : sondage empirique - 20 min

### Distribuer la feuille.

**Consigne :** Choisir 5 cercles "au hasard" sur la feuille. Calculer le diamètre moyen des cercles choisis.

### Collecte des résultats :

- Chaque élève annonce son diamètre moyen.
- L'animateur saisit les résultats des élèves dans la feuille "Activité 1". Le diagramme en bâtons se construit.
- Observer la tendance générale : les diamètres moyens des échantillons sont généralement supérieurs à 2 cm. Révéler le diamètre moyen des cercles de la feuilles : 2 cm.

**Question :** Comment avez-vous choisi vos cercles ?

→ Biais courant : le regard est attiré par les grands cercles.

Une méthode est dite sans biais si la moyenne de tous les résultats possibles est la moyenne du caractère dans la population.

→ Analogie avec les micro-trottoirs : on interroge surtout les personnes les plus visibles/disponibles. Des individus non représentés : personnes qui sortent peu, qui n'ont pas envie de répondre.

**Conclusion :** Méthode facile mais fortement biaisée.

### POINTS D'ATTENTION

- **Ne pas** préciser ce qu'on entend par "au hasard"
- Accepter toutes les procédures de sélection
- Observer comment les élèves font leur choix

## Partie 3 Activité 2 : sondage aléatoire simple - 20 min

### Consignes :

- Numéroté les 60 cercles de la feuille de 1 à 60.
- Utiliser le générateur d'entiers aléatoires d'une calculatrice pour sélectionner 5 entiers entre 1 et 60. (*Le fonctionnement de certaines calculatrices est donné en annexe.*)
- Repérer les cercles correspondants et calculer le diamètre moyen de l'échantillon.

### Collecte des résultats :

- Saisir les résultats dans la feuille "Activité 2". Le nouveau diagramme en bâtons se construit.
- Comparer avec le premier (*utilisation de la feuille Activités 1 et 2*)

### Débriefing :

- Observer que les diamètres moyens fluctuent autour de 2 cm.
- Les résultats sont moins dispersés (plus proches les uns des autres).
- L'échantillonnage aléatoire simple est une méthode sans biais, mais elle est très coûteuse car il faut identifier tous les individus de la population.
- Lien avec le recensement de la population française.

**Transition :** Il existe d'autres méthodes de sondage, également aléatoires, qui peuvent avoir un coût moindre de mise en œuvre.

## Partie 4 Activité 3 : sondage par stratification - 25 min

**Explication du principe de stratification :** On divise la population en groupes homogènes (strates) (*voir compléments*).

### Distribuer une nouvelle feuille.

#### Consigne :

- On sépare la population en deux groupes : les 45 petits cercles dont le diamètre est inférieur ou égal à 2 cm et les 15 grands. Numéroté les petits cercles de 1 à 45 et les grands de 1 à 15.
- Observer la proportion :  $45/60 = 3/4$  et  $15/60 = 1/4$ .
- Sélectionner 4 petits cercles et 1 grand cercle en utilisant le générateur d'entiers aléatoires de la calculatrice.
- Calculer la moyenne pondérée :  $(45 \times \text{moyenne\_4\_petits} + 15 \times \text{diamètre\_1\_grand}) / 60$ .

### Collecte et analyse :

- Saisir les résultats dans la feuille "Activité 3".
- Le diagramme en bâtons se construit.
- Comparer la dispersion avec l'activité précédente.

### Débriefing :

- Analogie avec stratification par catégorie socio-professionnelle
- Avantages : Moindre dispersion à taille d'échantillon égale

**Inconvénients :** Cette méthode nécessite un recensement préalable de la population et sa connaissance selon un critère lié au caractère étudié.

**Conclusion :** Plus précis mais plus complexe à mettre en œuvre.

#### POINTS D'ATTENTION

- S'assurer que les élèves comprennent pourquoi on prend 4 petits et 1 grand.
- Vérifier le calcul de la moyenne pondérée.

## Partie 5 Activité 4 : sondage à deux degrés - 20 min

**Explication du principe du sondage :** On regroupe les individus selon un critère facilitant le recueil des résultats : géographiquement par exemple. On sélectionne d'abord un groupe géographique, puis on échantillonne à l'intérieur (*voir compléments*).

**Distribuer une nouvelle feuille.**

### Consignes :

- Faire 6 groupes numérotés de 1 à 6 de 10 cercles proches géographiquement dans la feuille.
- Utiliser le générateur d'entiers aléatoires d'une calculatrice pour sélectionner un des groupes par son numéro entre 1 et 6.
- Dans le groupe sélectionné, numéroté les cercles de 1 à 10.
- Constituer un échantillon aléatoire de 5 cercles du groupe choisi.
- Calculer le diamètre moyen de l'échantillon.

### POINTS D'ATTENTION

- *Tous les élèves ne travailleront pas sur le même paquet : c'est normal.*
- *Certains paquets peuvent être plus homogènes que d'autres.*

### Collecte et analyse :

- Saisir les résultats dans la feuille "Activité 4", le diagramme en bâtons se construit.
- Observer que la dispersion est généralement plus importante qu'avec le sondage aléatoire simple.

### Débriefing :

- Les groupes géographiques sont souvent homogènes ("qui se ressemble s'assemble").
- Le paquet choisi peut ne pas être représentatif de la population.
- Avantages : Simplicité et réduction des coûts
- Exemples concrets :
  - Choisir d'abord un établissement scolaire (élèves d'origine similaire).
  - Choisir une ligne de bus (passagers d'origine homogène)

**Conclusion :** Moins précis mais économique et pratique..

## Partie 6 Synthèse et conclusion

### Discussion :

- Comparer les 4 diagrammes côte à côte.
- Classer les méthodes par précision : Stratification > Aléatoire simple > Deux degrés > Empirique.
- Discuter du compromis coût/précision.

### Points clés à faire ressortir :

- Sondage empirique : Très biaisé, à éviter malgré sa facilité.
- Sondage aléatoire simple : Méthode de référence, sans biais mais coûteuse.
- Sondage stratifié : Meilleure précision si on connaît bien la population.
- Sondage à deux degrés : Économique mais moins précis.

## INFORMATIONS DE RÉFÉRENCE (À NE PAS COMMUNIQUER AUX ÉLÈVES)

- Diamètre moyen réel de la population : **2 cm**
  - 28 cercles de diamètre 1 cm
  - 17 cercles de diamètre 2 cm
  - 7 cercles de diamètre 3 cm
  - 5 cercles de diamètre 4 cm
  - 2 cercles de diamètre 5 cm
  - 1 cercle de diamètre 7 cm
- 45 petits cercles (diamètre  $\leq 2$  cm) : moyenne = 1,38 cm
- 15 grands cercles (diamètre  $> 2$  cm) : moyenne = 3,87 cm
- Répartition : 3/4 de petits cercles, 1/4 de grands cercles

## COMPLÉMENT 1 : LE SONDAGE PAR STRATIFICATION

Pour un sondage stratifié, on partage la population en groupes supposés relativement homogènes par rapport à l'objet d'étude (appelés strates) et on prélève dans chaque strate un échantillon aléatoire simple. Il faut donc avoir sur la population, une certaine connaissance a priori d'une caractéristique liée au problème étudié. Par exemple, pour une enquête sur les modes de consommation, on peut stratifier la population selon les différentes CSP (catégories socio-professionnelles) ou selon l'habitat (urbain/rural), etc.

Les strates peuvent être d'effectifs différents, mais il convient que les strates soient assez homogènes par rapport au caractère étudié de telle sorte que les écarts-types des variables en jeu puissent être minimisés. La moyenne estimée d'un caractère sur la population est alors la moyenne pondérée de chaque observation proportionnellement aux écarts-types du caractère dans les strates. Cette méthode est plus précise que l'échantillonnage aléatoire simple, l'erreur d'échantillonnage étant réduite dans la mesure où la dispersion globale a pu être réduite aux dispersions dans chaque strate. Pour avoir la même précision avec un sondage aléatoire simple, il faudrait donc un échantillon de la population de taille plus grande que la réunion de tous les échantillons des différentes strates. Dans notre exemple, pour ne pas compliquer inutilement la tâche des élèves, on considère deux strates contenant l'une, les petits cercles et l'autre, les grands cercles. Il est clair que la dispersion des diamètres dans chacune des strates est inférieure à la dispersion globale.

On se rapproche de l'échantillon stratifié optimal en prenant dans chaque strate un échantillon aléatoire de taille proportionnelle au poids de la strate dans la population, ce qui améliore la précision par rapport à un sondage aléatoire simple effectué à partir d'un échantillon de même taille, mais sans atteindre l'optimum théorique de l'échantillon stratifié le meilleur.

## COMPLÉMENT 2 : LE SONDAGE À DEUX DEGRÉS ET SONDAGE PAR GRAPPES

Comme pour l'échantillonnage stratifié, dans l'échantillonnage à deux degrés, on partage la population en groupes (appelés grappes ou unités primaires). Par contre cette division ne se fait pas à partir de critères liés à la problématique étudiée, mais plutôt sur une base plus pratique, géographique par exemple pour les populations humaines.

Pour un échantillonnage en grappes, une fois que la partition de la population est réalisée, on sélectionne au hasard un certain nombre de grappes pour représenter la population totale et l'ensemble de tous les individus des grappes sélectionnées constitue l'échantillon. Il n'est pas nécessaire de posséder la liste complète des individus de toute la population. Il suffit d'avoir la liste de toutes les grappes, puis la liste de tous les individus dans les grappes sélectionnées.

Pour un échantillonnage à deux degrés, on sélectionne au hasard un certain nombre d'unités primaires, puis on procède à un deuxième tirage au sort pour prélever un échantillon dans chaque unité sélectionnée, au lieu d'en interroger tous les individus. Autant pour un échantillonnage en grappes, il est préférable de sonder un grand nombre de petites grappes plutôt qu'un petit nombre de grandes grappes (pour éviter les effets de voisinage), autant pour un échantillonnage à deux degrés, il vaut mieux identifier au départ des grandes unités primaires devant contenir plus d'individus qu'il n'en faut pour l'échantillon final. Ce type d'échantillonnage permet de contrôler la taille finale de l'échantillon et offre l'avantage d'un échantillon plus concentré, donc moins coûteux. Par contre, à taille d'échantillon égale, le sondage à deux degrés est moins précis que le sondage aléatoire simple (et donc nettement moins précis que le sondage stratifié). Mais il le sera d'autant plus que les moyennes des unités primaires ne sont pas très éloignées les unes des autres.

## TABLEAU DE SYNTHÈSE DES 4 MÉTHODES DE SONDAGE

Méthode	Précision	Coût	Biais
Sondage empirique	✗ Très faible	✓ Faible	✗ Très biaisé
Sondage aléatoire simple	✓ Bonne	✗ Élevé	✓ Sans biais
Sondage stratifié	✓ ✓ Excellente	✗ Élevé	✓ Sans biais
Sondage à deux degrés	Moyenne	✓ Faible	✓ Sans biais

## COMPLÉMENT 3 : LE SONDAGE PAR LA MÉTHODE DES QUOTAS

Une autre forme d'échantillonnage non aléatoire, très courante, est le sondage par la méthode des quotas. Pour que l'échantillon soit un peu représentatif de la population, celle-ci est divisée en sous-populations (les quotas) déterminées par des proportions définies dans la population. Cette étape est semblable à celle de l'échantillonnage stratifié. Mais ensuite les individus à l'intérieur des quotas ne sont pas sélectionnés au hasard, le sondeur a toute latitude pour choisir les individus qu'il va retenir pour son échantillon. Les principaux avantages de cette méthode sont qu'elle diminue les coûts et qu'elle est facile à appliquer, tout en respectant les proportions des différents groupes identifiés dans la population ; elle ne nécessite pas non plus l'existence d'une base de sondage complète de la population étudiée. Par contre, elle peut camoufler des biais de sélection qui peuvent ne pas être négligeables.

## COMPLÉMENT 4 : SONDAGE ALÉATOIRE VS SONDAGE NON ALÉATOIRE

Devant le coût, la complexité et la quantité d'informations nécessaires à la mise en œuvre des méthodes aléatoires, sont apparues des méthodes non aléatoires.

La différence entre un échantillonnage aléatoire et un échantillonnage non aléatoire tient au fait que dans le cas de l'échantillonnage aléatoire, le hasard est une caractéristique du processus de sélection. Cependant il ne suffit pas que les individus de la population sondée aient la même probabilité d'être dans l'échantillon ; par exemple, pour une personne choisie de façon équiprobable dans une population, l'échantillon de taille  $n$  composé avec  $n - 1$  de ses voisins vérifie cette condition et pourtant n'est pas acceptable pour un sondage aléatoire. La mise en œuvre d'un tel sondage présente donc des difficultés ; pour un échantillon de taille  $n$ , il faudrait que tout sous-ensemble de  $m$  éléments ait la même probabilité d'être tiré. Dans la pratique, on limite les valeurs de  $m$  à 1, 2 ou 3.

Dans le cas de l'échantillonnage non aléatoire, on choisit d'une certaine manière les individus, afin de faciliter l'échantillonnage et de réduire son coût. Il en résulte qu'on ne peut ni estimer la variabilité du caractère étudié, ni calculer l'erreur induite dans l'estimation du paramètre considéré dans la population. Il n'existe donc aucun moyen de mesurer la précision et la fiabilité du résultat obtenu par cette méthode. Elle est malgré tout utilisée car elle est souvent plus économique et plus pratique que les méthodes aléatoires. Elle peut être pertinente pour élaborer une enquête ou pour tester un questionnaire.

Le premier sondage effectué dans l'activité (sondage empirique ou sondage à l'aveuglette) est une première forme d'échantillonnage non aléatoire, l'échantillon étant prélevé sans méthode. Comme nous l'avons vu, il a alors de grandes chances d'être biaisé.

## COMPLÉMENT 5 : MODALITÉS DE RECENSEMENT

Le recensement de la population concerne l'ensemble de la population vivant en France. Il permet de déterminer les populations de référence de la France et de ses circonscriptions administratives. Il fournit également des statistiques sur les caractéristiques des habitants (sexe, âge, profession, nationalité, emploi, etc.), des ménages (composition, taille, liens familiaux, etc) et de leurs logements (type, nombre de pièces, année d'emménagement, etc.). Les données du recensement sont diffusées chaque année.

### Historique

Le premier recensement de l'ère moderne a eu lieu en 1801. À partir de 1946, date de création de l'Insee, toutes les communes ont été recensées exhaustivement à intervalles irréguliers (tous les 6 à 9 ans). Ces recensements généraux de la population ont pris fin en 1999.

### Depuis 2004

Depuis 2004, un échantillon de la population est recensé chaque année, les années de recensement diffèrent selon les communes : la fréquence de la collecte est quinquennale pour les communes de moins de 10 000 habitants, et annuelle pour les autres communes. L'enquête de recensement est exhaustive dans le premier cas ; c'est une enquête par échantillon dans le second.

Dans celles de moins de 10 000 habitants, le recensement a lieu tous les ans par roulement sur 20 % des communes et toute la population est concernée.

Dans celles de 10 000 habitants ou plus, la collecte annuelle porte sur un échantillon d'adresses tirées au hasard et représentant environ 8 % de la population. Au bout de 5 ans, l'ensemble du territoire de chaque commune est pris en compte, et 40 % environ des habitants de ces communes ont été recensés.

Les statistiques élaborées à partir des enquêtes de recensement sont représentatives de l'ensemble de la population. Concrètement, environ 9 millions de personnes sont recensées chaque année, soit 14 % de la population vivant en France.

Les cinq premières enquêtes annuelles de recensement (EAR) ont été réalisées de 2004 à 2008. Elles ont permis de produire les résultats du recensement, millésimé 2006, date du milieu de la période. Depuis, chaque année, les résultats de recensement sont produits à partir des cinq enquêtes annuelles les plus récentes, avec abandon des informations issues de l'enquête la plus ancienne et prise en compte de l'enquête nouvelle.

## COMPLÉMENT 5 : MODALITES DE RECENSEMENT (SUITE)

Les articles 156 à 158 de la loi n°2002-276 du 27 février 2002 relative à la démocratie de proximité encadrent les opérations de recensement. Ils définissent notamment le partenariat entre l'Insee et les communes pour la réalisation du recensement.

Les communes - ou les établissements publics de coopération intercommunale (EPCI) - préparent et réalisent les enquêtes de recensement. L'Insee a pour mission d'organiser et de contrôler la collecte des informations. Il recueille ensuite l'information collectée, exploite les questionnaires et diffuse les résultats.

Près de 350 articles de lois ou de codes se réfèrent à la population de chaque circonscription administrative. Ils concernent l'organisation des communes comme notre vie quotidienne : les dotations de l'État aux communes, le nombre de conseillers municipaux, les conditions d'implantation des pharmacies, la constitution de communautés d'agglomération, les barèmes de certaines taxes (publicité, jeux, spectacles, débitants de boissons, etc.). Les résultats du recensement sont également utilisés pour prévoir les équipements collectifs nécessaires (écoles, hôpitaux, etc.).

**Annexe** **Feuille des cercles**

