# Techniques et Méthodes d’Echantillonnage

Dr. Ir. Epiphane SODJINOU, présente de façon exhaustive les concepts, techniques et méthodes liés à l’échantillonnage en statistique, particulièrement appliqués aux études quantitatives.

# 1. Clarification des concepts fondamentaux

L’**échantillonnage** est défini comme le processus de sélection d’un sous-ensemble d’individus à partir d’une population cible, dans le but de réduire les coûts, faciliter la gestion des opérations, accélérer la collecte et le traitement des données, et alléger le fardeau des informateurs. Deux option s existent lors d’une étude : réaliser un recensement (rarement faisable) ou recourir à l’échantillonnage.

**Un échantillon** est ce sous-ensemble restreint effectivement examiner par le chercheur. Il doit être représentatif : ses caractéristiques doivent correspondre à celles de la population d’origine. **La randomisation** (tirage au hasard) permet de compenser les facteurs de confusion connus ou inconnus. Il est essentiel de bien comprendre la conception et l’exécution du plan d’échantillonnage, car tout écart peut se répercuter sur la qualité des données.

Les biais possibles :

1. **Biais de sélection** : certains groupes sont sur- ou sous-représentés dans l’échantillon.
2. **Biais de non-réponse** : certains individus répondent plus que d’autres, ce qui peut fausser les résultats. Un taux élevé de non-réponse (au-delà de 20-30 %) peut entraîner un biais substantiel.

La **population** est l’ensemble des unités auxquelles s’appliquent les résultats de l’étude. L**’unité** d’échantillonnage est chaque membre de la base de sondage, qui doit être la meilleure approximation possible de la population (chaque membre y apparaît une seule fois).

# 2. Méthodes d’échantillonnage

On distingue deux grandes familles :

1. **Échantillonnage probabiliste (aléatoire)** : chaque unité a une chance quantifiable d’être sélectionnée.
2. **Échantillonnage non probabiliste** : la probabilité de sélection n’est pas égale pour tous.

# 3. Méthodes aléatoires (probabilistes)

La sélection repose sur la randomisation. Les principales méthodes sont :

**Échantillonnage aléatoire simple (EAS)** : chaque individu a une chance égale d’être sélectionné, ce qui élimine tout biais du chercheur. La sélection se fait à l’aide de tables de nombres aléatoires ou d’outils informatiques.

**Échantillonnage systématique** : on choisit un point de départ au hasard puis on sélectionne les unités à intervalles réguliers (pas de sondage). Cette méthode est simple mais peut être biaisée si la liste présente une périodicité.

**Échantillonnage stratifié** : la population est divisée en strates homogènes (groupes), et un échantillon est tiré au hasard dans chaque strate, proportionnellement à leur taille. Cela garantit la représentation de toutes les catégories et réduit la variabilité de l’échantillon.

**Échantillonnage par grappes** : la population est divisée en grappes (groupes représentatifs), puis certaines grappes sont sélectionnées au hasard, et des individus sont tirés au sein de ces grappes. Cette méthode est utile lorsque la liste complète de la population n’est pas disponible, mais elle complique l’analyse statistique.

**Échantillonnage à plusieurs degrés** : il s’agit d’un échantillonnage par grappes appliqué à plusieurs niveaux (par exemple, sélection de régions, puis de villages, puis de ménages).

**Échantillonnage à plusieurs phases** : on collecte d’abord des données de base sur un large échantillon, puis on réalise une collecte plus détaillée sur un sous-échantillon. Cette méthode est utile en cas de ressources limitées ou de manque d’informations auxiliaires.

# 4. Méthodes non aléatoires

Ces méthodes incluent l’échantillonnage par quotas, de convenance, ou par choix raisonné. Elles sont utilisées lorsque la constitution d’une base de sondage exhaustive est impossible, mais elles présentent un risque élevé de biais et limitent la généralisation des résultats.

# 5. Détermination de la taille de l’échantillon

La taille d’échantillon, dépend de la précision souhaitée, du niveau de confiance, de la variabilité de la population et des ressources disponibles.

# 6. Considérations supplémentaires

Il est crucial de bien définir la base de sondage et de veiller à sa qualité, car toute erreur à ce niveau se répercute sur la validité des résultats. La compréhension du plan d’échantillonnage est également essentielle pour l’analyse des données et l’interprétation des résultats.

# Rééchantillonnage

Le cours sur le rééchantillonnage présente les méthodes statistiques qui consistent à générer de nouveaux échantillons à partir d’un échantillon de données initial, dans le but d’estimer la précision des statistiques calculées (moyenne, variance, intervalle de confiance, etc.) ou de tester des hypothèses. Ces techniques sont particulièrement utiles lorsque la distribution théorique des estimateurs est complexe ou inconnue, ou lorsque la taille de l’échantillon est limitée.

# 1. Introduction et définition

Le rééchantillonnage désigne l’ensemble des méthodes qui créent, à partir d’un échantillon observé, de nombreux « nouveaux » échantillons, généralement par tirage avec ou sans remise. Ces méthodes permettent d’évaluer la variabilité des estimateurs sans recourir à des formules analytiques complexes. Elles sont largement utilisées en statistique moderne, notamment grâce à la puissance de calcul des ordinateurs.

# 2. Principales méthodes de rééchantillonnage

**Bootstrap** : C’est la méthode la plus connue. Elle consiste à tirer, avec remise, de nombreux échantillons de la même taille que l’échantillon initial. Pour chaque échantillon bootstrap, on recalcule la statistique d’intérêt (par exemple, la moyenne). En répétant l’opération des centaines ou des milliers de fois, on obtient une distribution empirique de la statistique, qui permet d’estimer sa variance, de construire des intervalles de confiance, ou de réaliser des tests. Le bootstrap est particulièrement utile lorsque l’on ne connaît pas la distribution de la statistique ou lorsque la taille de l’échantillon est trop petite pour appliquer des méthodes asymptotiques classiques.

**Jackknife** : Cette méthode consiste à retirer successivement une observation de l’échantillon et à recalculer la statistique à chaque fois. Elle permet d’estimer le biais et la variance de l’estimateur. Le jackknife est simple à mettre en œuvre mais moins performant que le bootstrap pour des échantillons très petits ou des statistiques non linéaires.

**Permutation (ou randomisation)** : Cette approche consiste à mélanger aléatoirement les valeurs d’un échantillon (ou de plusieurs groupes) pour tester des hypothèses, notamment en analyse de variance ou en comparaison de groupes. Elle permet de construire la distribution nulle d’un test sans hypothèse paramétrique.

# 3. Applications du rééchantillonnage

**Estimation de la précision** : Les méthodes de rééchantillonnage sont utilisées pour estimer la variance, le biais, ou les intervalles de confiance d’une statistique, même lorsque la distribution sous-jacente est inconnue.

**Tests statistiques** : Elles permettent de réaliser des tests d’hypothèses sans supposer de distribution particulière, en construisant empiriquement la distribution sous l’hypothèse nulle.

**Validation de modèles** : Le rééchantillonnage est utilisé pour la validation croisée (cross-validation) en apprentissage automatique, afin d’évaluer la performance d’un modèle sur des sous-échantillons des données.

# 4. Avantages et limites

## Avantages :

Applicabilité très large, même avec des échantillons de petite taille ou des distributions inconnues.

Facilité d’implémentation avec les outils informatiques modernes.

Permet d’obtenir des estimations robustes de la variabilité des statistiques.

## Limites :

Peut donner des résultats biaisés si l’échantillon initial n’est pas représentatif de la population.

Les méthodes comme le bootstrap supposent que l’échantillon observé reflète fidèlement la variabilité de la population.

Nécessite une puissance de calcul importante pour les grands jeux de données ou un grand nombre de répétitions.

# 5. Conclusion

Le rééchantillonnage constitue une avancée majeure en statistique appliquée, permettant d’estimer la précision des résultats et de réaliser des tests sans hypothèses fortes sur la distribution des données. Il s’impose comme un outil incontournable dans l’analyse de données modernes, notamment en biostatistique, en économie, et en science des données.

# QCM sur les techniques d’échantillonnage

1. À quoi sert l’échantillonnage dans une étude statistique ?
2. À étudier toute la population
3. À réduire les coûts et faciliter la collecte de données
4. À garantir l’absence totale de biais
5. À augmenter la taille de la population
6. Qu’est-ce qu’un échantillon représentatif ?
7. Un échantillon choisi au hasard
8. Un échantillon dont les caractéristiques reflètent celles de la population
9. Un échantillon plus petit que la population
10. Un échantillon sans non-réponse
11. Quel est le principal avantage de l’échantillonnage aléatoire simple ?
12. Il est toujours sans biais
13. Il élimine la possibilité de biais de sélection du chercheur
14. Il ne nécessite pas de liste de la population
15. Il est réservé aux grandes populations
16. Quelle méthode consiste à diviser la population en groupes homogènes avant de tirer l’échantillon ?
    1. Échantillonnage systématique
    2. Échantillonnage stratifié
    3. Échantillonnage par grappes
    4. Échantillonnage non probabiliste
17. Quel est le principal risque d’un taux élevé de non-réponse dans un échantillon ?
18. Diminution de la taille de l’échantillon
19. Introduction d’un biais dans les résultats
20. Augmentation des coûts
21. Simplification de l’analyse

# QCM sur les techniques de rééchantillonnage

1. Le rééchantillonnage permet principalement de :
2. Créer de nouveaux échantillons à partir de la population
3. Estimer la précision des statistiques à partir d’un échantillon existant
4. Remplacer l’échantillonnage aléatoire
5. Réduire la taille de l’échantillon initial
6. Quelle méthode de rééchantillonnage consiste à tirer, avec remise, de nombreux échantillons à partir de l’échantillon initial ?
7. Jackknife
8. Bootstrap
9. Stratification
10. Grappes
11. Le jackknife consiste à :
12. Tirer des échantillons avec remise
13. Retirer une observation à la fois et recalculer la statistique
14. Mélanger aléatoirement les valeurs
15. Diviser la population en strates
16. Une des limites du bootstrap est :
17. Il ne peut pas être utilisé avec des ordinateurs
18. Il suppose que l’échantillon initial est représentatif de la population
19. Il nécessite une connaissance parfaite de la population
20. Il ne permet pas d’estimer des intervalles de confiance
21. Les méthodes de permutation (randomisation) servent principalement à :
22. Calculer la moyenne d’un échantillon
23. Tester des hypothèses sans supposer de distribution particulière
24. Réduire le biais de sélection
25. Augmenter la taille de l’échantillon

# QCM complémentaire sur l’échantillonnage

1. Qu’est-ce que la base de sondage dans une étude d’échantillonnage ?
2. La liste des résultats attendus
3. La liste des unités à partir de laquelle on sélectionne l’échantillon
4. La méthode de collecte des données
5. L’ensemble des enquêteurs
6. Quelle est la différence entre unité d’échantillonnage et unité d’analyse ?
7. Il n’y en a aucune
8. L’unité d’échantillonnage est le niveau auquel on prélève, l’unité d’analyse est le niveau sur lequel porte l’analyse
9. L’unité d’échantillonnage est toujours un individu, l’unité d’analyse toujours un groupe
10. L’unité d’analyse est choisie au hasard, l’unité d’échantillonnage est fixée
11. Quel est le principal objectif de la stratification dans l’échantillonnage stratifié ?
    1. Réduire le temps de collecte
    2. Garantir la représentation de toutes les catégories de la population
    3. Simplifier le calcul de la variance
    4. Faciliter la sélection aléatoire simple
12. Dans quel cas utilise-t-on l’échantillonnage à plusieurs phases ?
13. Quand la population est homogène
14. Quand il manque des données auxiliaires pour stratifier ou que le budget est limité
15. Quand la base de sondage est identique à la population
16. Quand on ne veut pas utiliser de questionnaire
17. Qu’est-ce qu’un biais de non-réponse ?
18. Lorsque tous les individus répondent à l’enquête
19. Lorsque certains individus de l’échantillon ont une probabilité plus élevée de répondre que d’autres, ce qui peut fausser les résultats
20. Lorsque l’échantillon est trop petit
21. Lorsque la base de sondage est incomplète
22. Quelle est la meilleure approche pour limiter le biais de non-réponse ?
23. Ignorer les non-réponses
24. Utiliser un échantillonnage systématique
25. Tout mettre en œuvre pour obtenir un taux de réponse élevé
26. Augmenter la taille de la population

# QCM sur l’échantillonnage (notions avancées et spécifiques)

1. Lorsqu’on parle de la taille de l’échantillon, quel facteur n’influence PAS directement son calcul ?
   1. Le niveau de confiance souhaité
   2. Le nom de l’enquêteur
   3. La variabilité de la population
   4. La précision attendue des résultats
2. Parmi les méthodes suivantes, laquelle est une méthode d’échantillonnage non probabiliste ?
3. Échantillonnage aléatoire simple
4. Échantillonnage par quotas
5. Échantillonnage par grappes
6. Échantillonnage systématique
7. Quelle est la différence principale entre la base de sondage et la population ?
8. La base de sondage contient toujours plus d’unités que la population
9. La base de sondage est la liste à partir de laquelle on sélectionne l’échantillon, la population est l’ensemble des unités concernées par l’étude
10. La population est toujours une liste exhaustive
11. Il n’y a aucune différence
12. Dans l’échantillonnage systématique répété, pourquoi effectue-t-on plusieurs passages dans la liste avec des points de départ différents ?
13. Pour augmenter la taille de l’échantillon
14. Pour protéger contre la périodicité dans la liste et estimer la variance des sous-échantillons
15. Pour simplifier la sélection
16. Pour réduire le coût de l’enquête
17. La post-stratification consiste à :
18. Créer des strates avant le tirage de l’échantillon
19. Ajuster l’analyse après le tirage, en utilisant des informations de stratification obtenues après la sélection
20. Tirer au hasard dans chaque strate
21. Sélectionner uniquement les strates les plus grandes
22. Quelle unité correspond au niveau sur lequel porte l’analyse, et non forcément la sélection ?
23. Unité d’échantillonnage
24. Unité d’analyse
25. Unité déclarante
26. Unité de référence
27. L’échantillonnage à plusieurs phases est particulièrement utile lorsque :
28. On dispose d’un budget illimité
29. On ne dispose pas de toutes les informations auxiliaires sur la base de sondage ou le budget est limité
30. On souhaite uniquement estimer la moyenne
31. La population est très homogène

# QCM sur le rééchantillonnage (notions avancées et spécifiques)

1. Le bootstrap permet de construire des intervalles de confiance pour une statistique en :
2. Utilisant des formules analytiques classiques
3. Utilisant la distribution empirique obtenue par rééchantillonnage avec remise
4. Calculant uniquement la moyenne
5. Ignorant la variabilité de l’échantillon
6. Le jackknife est particulièrement adapté pour :
7. Les statistiques très non-linéaires
8. Estimer le biais et la variance d’une statistique en retirant une observation à chaque fois
9. Les très grands échantillons uniquement
10. Les populations infinies
11. En apprentissage automatique, quelle technique de rééchantillonnage est couramment utilisée pour valider un modèle ?
12. Bootstrap
13. Validation croisée (cross-validation)
14. Échantillonnage stratifié
15. Échantillonnage systématique
16. Une limite importante du bootstrap est :
17. Il ne peut pas être utilisé avec des données numériques
18. Il suppose que l’échantillon initial est représentatif de la population
19. Il ne permet pas d’estimer la variance
20. Il ne fonctionne qu’avec des échantillons très grands
21. Les méthodes de permutation (randomisation) sont particulièrement utiles pour :
22. Calculer l’écart-type
23. Tester des hypothèses sans supposer de distribution particulière
24. Augmenter la taille de l’échantillon
25. Réduire le biais de sélection

# Questions de cours- Vrai ou Faux

1. L’échantillonnage consiste à sélectionner un sous-ensemble d’individus dans une population afin de réduire les coûts et faciliter la collecte des données. **Vrai**
2. Un échantillon doit toujours être identique à la population en termes de taille. **Faux**
3. Un échantillon représentatif doit refléter les caractéristiques principales de la population étudiée. **Vrai**
4. La randomisation permet de compenser les effets de confusion connus et inconnus lors de la sélection de l’échantillon. **Vrai**
5. Un échantillon biaisé est un échantillon qui n’est ni représentatif ni aléatoire. **Vrai**
6. Le biais de sélection se produit lorsque certains groupes d’individus sont surreprésentés ou sous-représentés dans l’échantillon. **Vrai**
7. Le biais de non-réponse n’a aucun impact sur la qualité des résultats d’une enquête. **Faux**
8. La meilleure façon de limiter le biais de non-réponse est de tout mettre en œuvre pour obtenir un taux de réponse élevé. V**rai**
9. La base de sondage est la liste des unités à partir de laquelle on sélectionne l’échantillon. **Vrai**
10. Dans l’échantillonnage probabiliste, chaque unité a une chance quantifiable d’être sélectionnée. **Vrai**
11. Dans l’échantillonnage non probabiliste, toutes les unités ont la même probabilité d’être sélectionnées. **Faux**
12. L’échantillonnage aléatoire simple donne à chaque membre de la population une chance égale d’être sélectionné. **Vrai**
13. L’échantillonnage systématique nécessite une liste numérotée de la population et un pas de sondage. **Vrai**
14. L’échantillonnage systématique répété permet de protéger contre la périodicité dans la liste. **Vrai**
15. La stratification consiste à regrouper les unités de la base de sondage en groupes homogènes appelés strates. **Vrai**
16. L’échantillonnage stratifié garantit que toutes les catégories de la population sont représentées dans l’échantillon. **Vrai**
17. La post-stratification consiste à ajuster l’analyse après la sélection de l’échantillon, en utilisant des informations de stratification obtenues après coup. **Vrai**
18. L’échantillonnage par grappes sélectionne d’abord des groupes représentatifs, puis des individus à l’intérieur de ces groupes. **Vrai**
19. L’échantillonnage à plusieurs degrés implique au moins deux niveaux de sélection (par exemple, grappes puis individus). **Vrai**
20. L’échantillonnage à plusieurs phases consiste à collecter d’abord des données de base sur un large échantillon, puis des données détaillées sur un sous-échantillon. **Vrai**
21. La taille de l’échantillon n’a aucune influence sur la précision des résultats. **Faux**
22. La base de sondage doit idéalement être identique à la population cible ou très similaire. **Vrai**
23. Il n’existe aucune différence entre unité d’échantillonnage, unité d’analyse, unité déclarante et unité de référence. **Faux**
24. Les méthodes d’échantillonnage non probabilistes incluent l’échantillonnage par quotas et de convenance. **Vrai**
25. L'échantillonnage systématique répété protège contre une éventuelle périodicité dans la liste. **Vrai**
26. Dans l'échantillonnage stratifié, les unités doivent obligatoirement avoir les mêmes chances d'être sélectionnées. **Faux**
27. La post-stratification est utilisée lorsque les informations sur les variables de stratification ne sont pas disponibles avant la sélection de l'échantillon. **Vrai**
28. L'échantillonnage par grappes simplifie l'estimation et l'analyse statistiques par rapport à l'échantillonnage aléatoire simple. **Faux**
29. L'échantillonnage à plusieurs degrés implique au moins deux niveaux d'échantillonnage. **Vrai**
30. L'échantillonnage à plusieurs phases est utile uniquement lorsqu'on dispose d'un budget illimité. **Faux**
31. La réduction de la variabilité dans l'échantillonnage stratifié se produit lorsque les unités d'une strate sont similaires, mais qu'il existe une variation entre les strates. **Vrai**
32. L'échantillonnage à plusieurs phases consiste à collecter des données de base auprès d'un large échantillon, puis des données plus détaillées auprès d'un sous-échantillon. **Vrai**
33. L'échantillonnage par grappes nécessite que chaque grappe soit représentative de la population mère. **Vrai**
34. Dans l'échantillonnage systématique, le pas de sondage est calculé en divisant la taille de la population par la taille souhaitée de l'échantillon. **Vrai**
35. L'échantillonnage aléatoire simple permet toujours d'obtenir un échantillon parfaitement représentatif de la population. **Faux**
36. La stratification permet de garantir que toutes les catégories de la population ciblée soient représentées en nombre suffisant. **Vrai**
37. L'échantillonnage à plusieurs degrés est identique à l'échantillonnage par grappes. **Faux**
38. Le biais de non-réponse est négligeable si le taux de non-réponse est supérieur à 30%. **Faux**
39. L'échantillonnage systématique nécessite de choisir au hasard un point de départ entre 1 et le pas de sondage. **Vrai**
40. La randomisation permet d'éliminer complètement tous les biais possibles dans un échantillon. **Faux**
41. L'échantillonnage est un processus qui permet d'étudier toute la population. **Faux**
42. La base de sondage doit contenir chaque membre de la population une fois et une seule fois. **Vrai**
43. L'échantillonnage probabiliste et l'échantillonnage aléatoire sont deux termes désignant la même approche. **Vrai**
44. Dans un échantillonnage aléatoire simple, le choix se fait avec remise, permettant à un individu d'être sélectionné plusieurs fois. **Faux**
45. L'échantillonnage stratifié peut réduire la variabilité des statistiques d'échantillon par rapport à un échantillonnage aléatoire simple. **Vrai**
46. L'échantillonnage par grappes simplifie l'estimation et l'analyse statistiques par rapport à l'échantillonnage aléatoire simple. **Faux**
47. Dans l'échantillonnage à plusieurs degrés, les unités primaires sont les plus petites unités sélectionnées. **Faux**

# QCM : Méthode adéquat

* 1. Pour une population homogène et une liste complète des individus, quelle méthode est la plus simple et la plus appropriée ?

1. Échantillonnage par grappes
2. Échantillonnage aléatoire simple
3. Échantillonnage stratifié
4. Échantillonnage à plusieurs phases
   1. Si la population peut être divisée en groupes homogènes (strates) selon une variable connue et pertinente pour l’étude, quelle méthode maximise la précision et assure la représentation de chaque groupe ?
5. Échantillonnage systématique
6. Échantillonnage stratifié
7. Échantillonnage par quotas
8. Échantillonnage à plusieurs degrés
   1. L’échantillonnage par grappes est particulièrement recommandé lorsque la population est très dispersée géographiquement et qu’il serait coûteux de dresser une liste complète de tous les individus. Vrai
   2. Dans quel contexte l’échantillonnage systématique peut-il présenter un risque de biais ?
9. Lorsque la population est très hétérogène
10. Lorsque la liste de la population présente une périodicité
11. Lorsque la taille de l’échantillon est très petite
12. Lorsque les strates sont inconnues
    1. L’échantillonnage à plusieurs phases est utile lorsque l’on dispose d’un budget limité ou que certaines informations auxiliaires ne sont pas disponibles pour toute la population. Vrai
    2. Si on ne dispose pas de variables de stratification avant la sélection de l’échantillon mais qu’on peut les obtenir après, quelle technique peut compenser ce manque ?
13. Échantillonnage systématique
14. Post-stratification
15. Échantillonnage à plusieurs degrés
16. Échantillonnage par quotas
17. L’échantillonnage par quotas est une méthode probabiliste recommandée pour les études nécessitant une inférence statistique rigoureuse. Faux
18. Pour une enquête nationale sur les ménages, avec des contraintes logistiques et une population dispersée, quelle méthode est souvent privilégiée ?
19. Échantillonnage aléatoire simple
20. Échantillonnage par grappes
21. Échantillonnage stratifié
22. Échantillonnage systématique
23. L’échantillonnage stratifié nécessite de connaître la répartition de la population selon les strates avant la sélection de l’échantillon. Vrai
24. Dans le cas où l’on souhaite réduire l’erreur d’échantillonnage et la variabilité, tout en assurant la représentation de sous-groupes spécifiques, quelle méthode privilégier ?
25. Échantillonnage aléatoire simple
26. Échantillonnage stratifié
27. Échantillonnage à plusieurs phases
28. Échantillonnage par grappes

# Epreuve 2024 : Technique et méthode d’Echantillonnage et rééchantillonnage

**Partie A**

**Répondez par « vrai » ou par « faux » aux affirmations qui suivent, et donnez la meilleure réponse lorsque vous répondez par « faux ».**

1. Dans un échantillon représentatif, les caractéristiques de l'échantillon correspondent à celles de la population. Vrai
2. Lors de l'analyse des données, il n'est pas nécessaire de comprendre la conception de l'échantillon, mais plutôt la manière dont la conception a été réellement exécutée sur le terrain. Faux. Il est essentiel de comprendre la conception de l'échantillon avant d'analyser les données, car cela influence la validité des résultats.
3. L'échantillonnage non probabiliste ne peut jamais produire d'échantillon représentatif. Faux. Un échantillonnage non probabiliste peut produire un échantillon représentatif, mais cela repose sur des hypothèses et n'est pas garanti scientifiquement.
4. Si la base de sondage et la population sont très similaires, on a peu d'erreur d'échantillonnage. Vrai
5. Dans un échantillonnage probabiliste, il n'est pas nécessaire que les chances que les unités d'observation soient sélectionnées dans l'échantillon soient égales. Vrai
6. Un échantillon sélectionné avec des chances de sélection inconnues peut être lié de manière appropriée à la population à partir de laquelle l'échantillon a été tiré. Faux. Sans connaître les chances de sélection, il est difficile de faire des inférences valides sur la population.
7. Dans les méthodes non aléatoires, la seule façon de mesurer la qualité des données en résultat consiste à comparer certains des résultats de l'enquête à l'information dont on dispose au sujet de la population. Vrai
8. La précision des estimations de l'échantillon dépend beaucoup de la taille de l'échantillon mais aussi de la taille de la population. Faux. La précision dépend principalement de la taille de l'échantillon et non de celle de la population, sauf pour les populations très petites.
9. Généralement, toutes les méthodes d'échantillonnage probabilistes permettent de tirer des échantillons représentatifs de la population cible. Vrai
10. L'échantillonnage aléatoire simple et l'échantillonnage stratifié sont considérés comme supérieurs (à toutes les méthodes d'échantillonnage) en termes de représentativité. Faux. D'autres méthodes, comme l'échantillonnage en grappes ou systématique, peuvent être tout aussi efficaces selon le contexte.

**Partie B**

1. Qu'entend-on par échantillon et quel est son intérêt ?
2. Qu'est-ce qu'une base de sondage ?
3. Citez les différentes méthodes d'échantillonnage probabilistes puis faites une brève description de l'échantillonnage systématique.
4. Citez deux types de biais d'échantillonnage, et décrivez-les brièvement.
5. Quel est l'inconvénient principal de l'échantillonnage stratifié.
6. Donnez deux avantages de l'échantillonnage non probabiliste.

Partie C

1. Qu'est-ce que le rééchantillonnage ?
2. Quelle idée sous-tend les méthodes de rééchantillonnage ?
3. Citez trois types de technique de rééchantillonnage, puis décrivez brièvement les principes de ces méthodes.
4. Citez deux avantages de la méthode du bootstrap ?

# Correction

Partie B

1. Un échantillon est un sous-ensemble représentatif d'une population, utilisé pour tirer des conclusions sur cette dernière. Il permet de réduire le coût et le temps d'une enquête tout en obtenant des résultats fiables.
2. Une base de sondage est la liste exhaustive des éléments de la population cible à partir de laquelle un échantillon peut être sélectionné.
3. Méthodes d'échantillonnage probabilistes :
   * Échantillonnage aléatoire simple o Échantillonnage stratifié o Échantillonnage en grappes o Échantillonnage systématique

Échantillonnage systématique : Il consiste à sélectionner les unités à des intervalles fixes à partir d'une liste ordonnée après avoir choisi un point de départ aléatoire.

1. Deux types de biais d'échantillonnage :
   * Biais de sélection : Se produit lorsque certaines unités de la population ont une probabilité plus faible d'être incluses dans l'échantillon.
   * Biais de non-réponse : Se produit lorsque les participants choisis ne répondent pas, créant une distorsion dans les résultats.
2. Inconvénient principal de l'échantillonnage stratifié : Il peut être coûteux et complexe à mettre en œuvre, nécessitant une connaissance préalable de la population pour bien définir les strates.
3. Deux avantages de l'échantillonnage non probabiliste :
   * Moins coûteux et plus rapide à mettre en œuvre.
   * Utile dans les études exploratoires ou lorsqu'une base de sondage n'est pas disponible.

Partie C

1. Le rééchantillonnage est une méthode statistique qui consiste à générer de nouveaux échantillons à partir d'un échantillon existant, souvent en utilisant des techniques comme le bootstrap ou le jackknife.
2. Idée sous-jacente : Le rééchantillonnage repose sur la simulation pour estimer la variabilité des statistiques sans avoir à faire des hypothèses strictes sur la distribution des données.
3. Trois techniques de rééchantillonnage :
   * Bootstrap : Tirages répétés avec remise pour estimer la distribution d'une statistique.
   * Jackknife : Exclusion séquentielle d'une observation à la fois pour évaluer la variabilité des estimations.
   * Permutation : Réarrangement aléatoire des valeurs observées pour tester des hypothèses statistiques.
4. Deux avantages de la méthode du bootstrap :
   * Ne nécessite pas d'hypothèses strictes sur la distribution des données.
   * Fournit des intervalles de confiance plus précis, même pour des échantillons de petite taille.