

Non-réponse totale et Non-réponse partielle

Mamadou Mbodj, Septembre 2024

Contact : mamadou.mbodj@ansd.sn

Cours : Traitement de données

(ENSAE 2024)

Classe : ISE 3



1. Objectifs du cours

- Définition de la non-réponse totale et la non-réponse partielle
- Identification des causes et impacts de la non-réponse
- Présentation de méthodes pour traiter ces problématiques dans les enquêtes

2. Définition de la non-réponse

- **Non-réponse totale**

- Aucune information n'est obtenue d'une unité d'enquête
- Exemple : Une personne refuse de répondre à l'enquête ou refuse de répondre à une ou plusieurs sections entières. Ce dernier cas constitue une non-réponse totale dans la ou les bases des sections concernées prises à part.

- **Non-réponse partielle**

- Certaines réponses manquent pour une unité d'enquête
- Exemple : Un répondant laisse des questions sensibles incomplètes

- **Incidence** : Affecte la qualité et la fiabilité des analyses. Sous estimation ou surestimation, problème d'extrapolation, etc.

3. Causes de la non-réponse

- **Non-réponse totale**
 - Difficulté à localiser les répondants
 - Refus catégorique de participation
 - Absence au moment de la collecte
 - Refus de répondre à une ou plusieurs sections entières
- **Non-réponse partielle :**
 - Questions complexes ou mal formulées
 - Fatigue ou désintérêt des répondants
 - Sensibilité ou confidentialité des questions

4. Conséquences de la non-réponse

- **Non-réponse totale :**
 - Risque de biais de non-réponse si les répondants diffèrent des non-répondants : biais dans estimations
 - Réduction de la taille de l'échantillon effectif : manque de précision des estimations
- **Non-réponse partielle :**
 - Données incomplètes qui compliquent les analyses
 - Perte d'information et de précision statistique

5. Traitement de la non-réponse totale

5

- **Stratégies préventives**

- Améliorer la conception de l'enquête (ex. : lettres introductives, motivation, sensibilisation, communication sur l'opération)
- Utiliser des rappels ou contacts supplémentaires
- Échantillon de remplacement pour atteindre la taille minimale de l'échantillon

- **Méthodes de correction**

- Pondération par ajustement : augmenter le poids des répondants similaires aux non-répondants
- Imputation de substitution : utiliser des unités similaires pour estimer les valeurs manquantes, notamment les non-réponses sur une ou quelques sections

5. Traitement de la non-réponse totale

6

Exemple de pondération par ajustement

- **Contexte**

Une enquête sur le revenu est menée auprès de 500 ménages répartis en zones rurales et urbaines. Les taux de réponse sont les suivants :

- Zone urbaine : 80 % de réponses.
- Zone rurale : 50 % de réponses.

- **Repondération** : On admet que le poids de chaque ménage est égal à 1

- Poids pour les zones urbaines : $1/0,8 = 1,25$
- Poids pour les zones rurales : $1/0,5 = 2$

Les ménages ruraux sont sous-représentés en raison de leur faible taux de réponse. Leur poids est donc augmenté pour refléter l'ensemble de leur groupe.

5. Traitement de la non-réponse totale

Généralisation pondération par ajustement

Soit une enquête sur n ménages. On vient de préciser que l'ajustement se fait en augmentant le poids des répondants similaires aux non-répondants. Soit n_g la taille d'un groupe d'individus similaires, n_r le nombre de ménages ayant répondu dans ce groupe et p_g le poids d'un ménage i dans ce groupe. La repondération se fait en posant, pour chaque ménage ayant répondu :

$$p_{g'} = p_g + \frac{(n_g - n_r)p_g}{n_r} = \frac{p_g}{t_g}$$

avec t_g le taux de réponse dans le groupe.

5. Traitement de la non-réponse totale

Avantages

- Réduit les biais de non-réponse
- Relativement facile à mettre en œuvre si les taux de réponse sont disponibles

Limites

- Suppose que les répondants sont similaires aux non-répondants au sein d'un même groupe, ce qui n'est pas toujours vrai
- Ne corrige pas les biais si les non-répondants diffèrent fondamentalement des répondants (ex. : en termes de revenu, d'opinion)

En résumé, la pondération par ajustement permet d'assurer que les résultats d'une enquête restent représentatifs, même en présence de non-réponse totale, à condition que les groupes soient bien définis.

5. Traitement de la non-réponse partielle

- **Techniques d'imputation** (abordées en détails dans le chapitre sur les imputations)
 - **Imputation simple** : Moyenne, médiane, ou mode
 - **Imputation conditionnelle** : Utiliser des modèles statistiques (régression, arbres de décision, knn, etc.)
 - **Imputation multiple** : Générer plusieurs jeux de données imputés et combiner les résultats

6. Cas pratique
