

Projet de Techniques d'Enquêtes

Souleymane Faye Max Michel Koffi Sébastien Culié

Magistère Ingénieur Economiste 1
AMSE - Ecole d'Économie d'Aix Marseille

Avril, 2022

Introduction

Dans ce projet, nous allons travailler sur le traitement post-enquête de la non-réponse sur une variable unique. La population est un groupe d'élèves de Licence 1 à Aix-Marseille Université. La question posée est la suivante : Pensez-vous être suffisamment informé sur les mesures anti-Covid-19 prises à AMU ? Notre objet d'études est la variable 24 de notre jeu de données : Mesures stop Covid. Il y a quatre réponses possibles :

- Oui, tout à fait
- Plutôt oui
- Plutôt non
- Non, pas du tout

Pour simplifier, nous avons codé sur SAS cette variable en deux modalités possibles :

- 1 Oui
- 2 Non

- 1 Le problème de la non-réponse dans les enquêtes
- 2 Concepts théoriques utilisés pour le traitement de la non-réponse
- 3 Traitement par imputation par la régression : code SAS
- 4 Analyse empirique et résultats obtenus
- 5 Conclusion : Takeaway

Le concept de non-réponse en techniques d'enquêtes

- Il y a non-réponse vis-à-vis de la variable Y pour l'individu échantillonné i dès lors que l'on ne dispose pas de la valeur Y_i relative à cet individu, quelle qu'en soit la cause
- La non-réponse est partielle pour les individus ici
- La non-réponse peut-être expliquée par l'indiscrétion de la question ou un état d'incertitude trop important. La deuxième semble plus probable dans le cas de la variable 24
- Introduction d'un biais de non-réponse : la non-réponse de l'individu à la question de la variable 24 nous donne une information sur celui-ci
- De plus, une diminution de la précision est causée par la non-réponse
- Si nous voulons corriger la non-réponse nous devons faire des hypothèses fortes sur le comportement des répondants
- En effet, il faut que la relation entre la variable dépendante et les variables explicatives qu'on modélise sur les répondants soit encore valable sur les non-répondants

Traitement de la non-réponse : Concepts théoriques

Il est possible de traiter la non réponse par trois méthodes :

- ① Répondération par calage sur marge
- ② Imputation par la régression
- ③ Imputation par Hot Deck

La répondération a un avantage dans le sens où il n'y a aucun problème de spécification du modèle. La contrainte dans l'imputation par régression est que les non-répondants doivent bien respecter le modèle que l'on va spécifier. Nous avons tout de même préféré l'utiliser en faisant cette hypothèse. La variable d'intérêt Y_i qui mesure la probabilité de réponse de l'individu est estimée par un modèle Logit

Traitement de la non-réponse : programmation SAS

Nous avons procédé de la manière suivante :

- ❶ Importation des données et code en Oui versus Non de nos modalités
- ❷ Code des variables administratives BN à CI en 0/1. On ne conserve que les variables qui pourraient avoir un impact sur notre variable 24
- ❸ Double estimation : avec toutes nos variables puis seulement avec les variables significatives
- ❹ On fait la prédiction de réponse des non répondants à la question de notre variable 24
- ❺ Enfin, on estime la proportion des individus ayant répondu oui à la question dans la grande population

Analyse empirique et résultats obtenus

- Résultats obtenus pour les coefficients : interprétation des coefficients de la matrice β
- Test de significativité statistique des variables qui nous amène à un second modèle : Wald et le Score de Fischer sont concluants...

Conclusion - Takeaway

- L'estimation nous a donné presque exclusivement des réponses oui à la question d'intérêt
- Les résultats de l'imputation ont des limites :
 - Notre interprétation dans les variables administratives que l'on retient
 - L'erreur qu'on utilise pour l'estimation du modèle Logit : plus on prend une erreur importante, moins les coefficients retenus sont réellement significatifs
 - En conséquence, on a pris une erreur de 0,01 pour une plus grande robustesse
- L'estimateur de la moyenne donne 0,53 pour une réponse positive à la question 24 lorsque les réponses manquantes ont été imputées par régression :
 - Il est intéressant de le comparer à celui obtenu avec les réponses manquantes