

Optimisation des plans de sondages pour les enquêtes CATI: gains potentiels et discussion

Clément Chevalier, Université de Neuchâtel
Lionel Qualité, Université de Neuchâtel et Office Fédéral de la
Statistique

Journées Suisses de la statistique
Mardi 21 Novembre 2017

Outline

1 Contexte

2 Budget et variance

- Améliorer la précision à budget constant
- Diminuer le budget à variance constante
- Un compromis entre la réduction du budget et la réduction de variance

Outline

1 Contexte

2 Budget et variance

- Améliorer la précision à budget constant
- Diminuer le budget à variance constante
- Un compromis entre la réduction du budget et la réduction de variance

Contexte

Dans le cadre des enquêtes CATI, la population Suisse se divise en 2 sous-populations:

- 1 Population non-ALTEL, dont le numéro de téléphone est connu.
- 2 Population ALTEL, dont le numéro de téléphone est inconnu.

Séparation approximativement connue dans le cadre de sondage.

Taux de réponses très différents (non-ALTEL > ALTEL)

Contexte

Lors d'enquêtes au niveau fédéral, des taux de sondage, p_1 et p_2 , sont utilisés pour les populations non-ALTEL et ALTEL, et le choix $p_1 = p_2$ semble être le plus courant.

Cependant, les probabilités de réponses très inégales entre ces sous-populations donnent des poids environ 3 fois plus importants aux individus ALTEL.

Objectif de ce travail

Étudier les gains possibles en choisissant $p_2 > p_1$. Interroger davantage la sous-population ALTEL.

Outline

1 Contexte

2 Budget et variance

- Améliorer la précision à budget constant
- Diminuer le budget à variance constante
- Un compromis entre la réduction du budget et la réduction de variance

Choix de référence

On définit un choix “de référence”:

$$(p_1, p_2) = (0.3\%, 0.3\%)$$

qui produit une enquête avec en moyenne 19'500 personnes dans l'échantillon brut et 9'750 dans l'échantillon net .

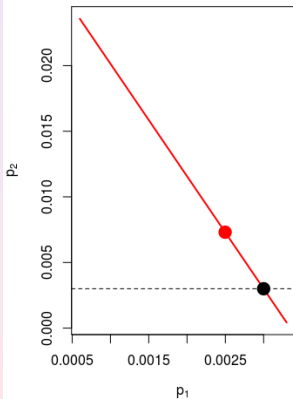
Les différents choix possibles où $p_1 \neq p_2$ seront tous comparés à ce choix de référence.

Diminuer la variance à budget constant.

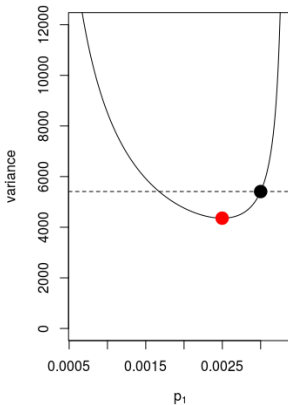
- On connaît les coûts moyens pour interroger les populations ALTEL et non-ALTEL \Rightarrow **budget de référence**.
- On peut aussi calculer une **variance de référence** de l'estimateur d'un total.
- Dans un premier temps, on étudie les couples (p_1, p_2) qui mènent au même budget que celui obtenu avec $(p_1, p_2) = (0.3\%, 0.3\%)$.

Diminuer la variance à budget constant.

inclusion proba., cste budget



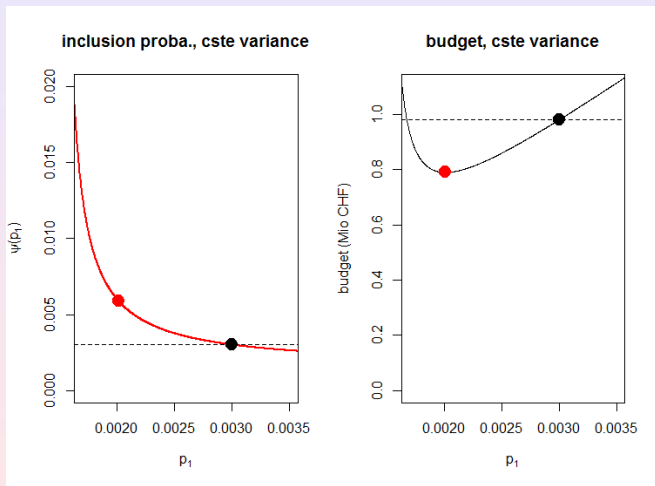
variance, cste budget



Diminuer le budget à variance constante

- Le choix de référence $(p_1, p_2) = (0.3\%, 0.3\%)$ mène à une certaine variance v_0 sur nos estimations.
- Ici on va s'intéresser à tous les autres couples $(p_1, p_2 = \psi(p_1))$ qui mènent à la même variance. Par la suite, on va regarder si des économies de budget sont possibles.

Diminuer le budget à variance constante



Robustesse à nos hypothèses sur γ

- **Résultat précédent** : Il est possible d'économiser 19.4% du budget, sans perte de qualité, en modifiant les taux de sondage, avec le choix $(p_1, p_2) = (0.201\%, 0.589\%)$.
- **Problème**: Ce résultat est sensible au ratio

$$\gamma = \frac{V_2^2}{V_1^2}$$

des variabilités des réponses entre les ALTEL et les non-ALTEL.
Si l'on s'est trompé sur la valeur de γ , ce nouveau choix ne produit plus la même variance que le choix de référence.

Robustesse à nos hypothèses sur γ

- **Résultat précédent** : Il est possible d'économiser 19.4% du budget, sans perte de qualité, en modifiant les taux de sondage, avec le choix $(p_1, p_2) = (0.201\%, 0.589\%)$.
- **Problème**: Ce résultat est sensible au ratio

$$\gamma = \frac{V_2^2}{V_1^2}$$

des variabilités des réponses entre les ALTEL et les non-ALTEL. Si l'on s'est trompé sur la valeur de γ , ce nouveau choix ne produit plus la même variance que le choix de référence.

Objectif

Proposer des choix moins ambitieux en terme d'économie de budget mais qui permettent une robustesse face à une mauvaise estimation de γ .

Résultats

tolérance souhaitée	0%	20%	40%	référence
économie	19.4%	15.0%	9.5%	0%
ρ_1	0.201%	0.212%	0.226%	0.3%
ρ_2	0.589%	0.621%	0.661%	0.3%
$\mathbb{E}(\# \text{ ech. brut})$	19'383	20'441	21'762	19'500
$\mathbb{E}(\# \text{ ech. net})$	7'801	8'226	8'758	9'750
stdev(# ech. net)	88.3	90.6	93.5	98.7

Conclusion / récapitulatif

- Un taux de sondages p_1, p_2 différent dans les sous-populations non-ALTEL et ALTEL peut conduire à des économies de budget allant jusqu'à 20% tout en maintenant la qualité de l'étude.
- Des économies moins ambitieuses, de l'ordre de 10%, permettent d'avoir en plus une robustesse face à une mésestimation de la variabilités des réponses de chaque sous population.
- Le fait d'augmenter les taux de sondage de la sous-population ALTEL qui a tendance à moins répondre n'a pas d'impact sur le biais.