



ined
INSTITUT
NATIONAL
D'ÉTUDES
DÉMOGRA
PHIQUES

Estimation de la précision d'une enquête téléphonique

Le cas de l'enquête



Nicolas RAZAFINDRATSIMA, Géraldine CHARRANCE, Gwennaëlle BRILHAULT
Institut national d'études démographiques (Ined), Service des enquêtes et sondages



13^{es} Journées de Méthodologie Statistique de l'INSEE
12 au 14 juin 2018

Plan de la présentation

0. Introduction
1. Présentation de l'enquête
2. Estimation de la variance
3. Les résultats
4. Conclusion

Introduction

- VIRAGE : une grande enquête téléphonique conduite par l'INED en 2015 sur les violences
- Quantification et analyse des déterminants de sujets sensibles
- Souhait de mettre à disposition des utilisateurs une méthode d'estimation de la précision simple à mettre en œuvre mais fiable
- Objectif: proposer une telle méthode et en tester les résultats



ined
INSTITUT
NATIONAL
D'ÉTUDES
DÉMOGRA
PHIQUES

1. Présentation de l'enquête



1.1 Objectifs de l'enquête

1.2 Méthodologie de l'enquête

1.3 Bilan de collecte

1.4 Calcul des pondérations

1.1 Objectifs de l'enquête

- Répondre au besoin de renouvellement des connaissances sur les violences à l'encontre des femmes émanant de la Mission d'évaluation de la politique de prévention et de lutte contre les violences faites aux femmes
- Actualiser les statistiques produites grâce à l'enquête Enveff réalisée en 2000 par le CRIDUP
- Produire des statistiques sur la victimation en accord avec les standards internationaux (ONU)
- Aller plus loin dans l'analyse des violences (type, contexte, auteur, conséquences...)

1.2 Méthodologie de l'enquête

Un projet de recherche, plusieurs enquêtes

- Un volet principal aléatoire par téléphone en population générale
- Trois volets complémentaires sur Internet auprès de volontaires (Virage-Universités, Virage-LGBT, Virage-Victimes)

Pourquoi une enquête par téléphone ?

- Génération aléatoire = Bonne couverture car tous les numéros ont une proba. non nulle d'appartenir à l'échantillon
- Interrogation par téléphone préférable au face-à-face pour aborder les sujets sensibles

1.2 Méthodologie de l'enquête

Sondage aléatoire à deux degrés

- 1^{er} degré : génération aléatoire de numéros de fixes & mobiles
- 2^e degré : sélection d'un individu au sein du ménage

Protocole d'appels

- Jusqu'à 40 tentatives d'appels
- Refus de niveau ménage et personnes sélectionnées rappelés
- Réponse possible en ligne si impossibilité ou réticence à répondre par téléphone

1.3 Bilan de collecte



211 235 numéros générés

Valides ?

Oui
(71%)

Non
(29%)

Contact établi ?

Oui
(87%)

Non
(13%)

Vérification éligibilité / Sélection ?

Oui
(69%)

Non
(31%)

Refus (55%)
Injoignables (36%)
Impossibles (9%)

Eligibles ?

Oui
(48%)

Non – Hors champ
(52%)

Enquêtés ?

Oui
(67%)

Non
(33%)

Refus/abandons (53%)
Injoignables (33%)
Impossibles (14%)

Taux de réponse : 49,0%
Taux de refus/abandons : 33,8%
Taux de non contact : 10,4%
Taux d'impossibles : 6,8%

1.4 Les pondérations

Les pondérations servent à corriger les biais liés au plan de sondage et à la non-participation à l'enquête

1/ Calcul des poids de sondage

$W_s = \text{nombre d'éligibles} / \text{nombre de lignes téléphoniques}$

2/ Correction du pourcentage de mobiles exclusifs

(source : baromètre du numérique, Credoc)

3/ Calage sur marges

par sexe, sur âge, lieu de naissance, CSP, type de ménage, région
(source : FD du recensement 2012, Insee)



ined
INSTITUT
NATIONAL
D'ÉTUDES
DÉMOGRA
PHIQUES

2. L'estimation de la variance

2.1 Problématique

2.2 Estimation de la « vraie » variance

2.3 La syntaxe approchée

2.4 Méthode de comparaison

2.1 Problématique

- La variance d'un plan à deux degrés est décomposable analytiquement (ex: dans Ardilly, 2006 p. 109)
 - premier terme : tient compte du sondage au premier degré (dispersion inter-UP)
 - second terme : tient compte de la dispersion intra-UP

2.1 Problématique

- La formule est impossible à mettre en œuvre ici car on n'a tiré qu'un seul individu par ménage
- Difficulté à prendre en compte la non-réponse (au 1^{er} et 2^e degré), que tente de corriger le calage sur marges

2.2 Estimation de la « vraie » variance

- Solution retenue dans ACSF, 1992 (Warszawski et al., 1997)
- Hypothèse: les individus sont tirés avec remise, directement de la population française adulte, avec des probabilités de sélection inégales
 - pas d'effet de grappe
 - la probabilité de tirer deux individus d'un même ménage est négligeable

2.2 Estimateur de variance (Hansen-Hurwitz)

- s l'échantillon, de taille n , Y la variable d'intérêt
- p_k la probabilité de sélection de l'unité k (constante d'un tirage à l'autre car tirage avec remise)
- $\hat{t}_{p,Y}$ l'estimateur du total de la variable Y

$$\text{On a : } \hat{t}_{p,Y} = \frac{1}{n} \sum_s \frac{y_k}{p_k} = \bar{z}, \text{ où } z_k = \frac{y_k}{p_k}$$

Et, l'estimateur de Hansen-Hurwitz de la variance :

$$\widehat{V}(\hat{t}_{p,Y}) = \frac{1}{n} \hat{S}_Z^2, \text{ où } \hat{S}_Z^2 = \frac{1}{n-1} \sum_s (z_i - \bar{z})^2$$

2.2 Prise en compte du calage

On note e_k le résidu de la régression de Y_k sur les variables de post-stratification X_k

$e_k = y_k - X'_k \hat{\beta}_s$, où $\hat{\beta}_s$ sont les coefficients estimés de la régression de Y sur X .

On pose \hat{t}_e le total estimé de ces résidus. La variance de l'estimateur post-stratifié du total de Y vaut alors : $\hat{V}(\hat{t}_e)$

2.2 Mise en œuvre pratique

- Calcul *a posteriori* de la probabilité de tirage
 - Estimation du nombre de téléphones éligibles (fixes et mobiles exclusifs)
 - prise en compte du nombre de téléphones par ménages et du nombre d'individus éligibles par ménage
- Estimation sans prise en compte du calage :
 - coefficient de variation CV2
- Estimation après prise en compte du calage :
 - coefficient de variation CV3

2.3 La syntaxe simplifiée

- Utilisateurs de VIRAGE non statisticiens en général
- Encourager les utilisateurs à sortir des intervalles de confiance
- Trouver une syntaxe simple à mettre en œuvre avec procédures SAS usuelles, donnant des variances plus larges que HH
- Sans mettre d'autre pondération dans le fichier diffusé, à part le poids calé

2.3 La syntaxe simplifiée

- Procédures SAS :

- *surveymeans*, option *strata*, *weight* poids après calage

```
proc surveymeans data=FIC mean sum;  
weight poids_cal;  
var VARINT;  
strata fic;
```

- Revient à postuler un sondage à 1 degré, stratifié, à probabilités inégales, avec remise

- Avantage: l'estimation du total ou de la moyenne est OK, sans biais (en principe)

- **coefficient de variation CV1**

2.4. Méthode de comparaison

- Comparaison des trois CV pour l'estimation du total et de la moyenne de 4 variables d'intérêt:
 - V1. Avoir subi une ou des violences sexuelles sans frotté collé sur la vie entière
 - V2. Avoir subi une ou des violences sexuelles avec frotté collé sur la vie entière
 - V3. Avoir souffert d'un trouble anorexique au cours de la vie (terminé ou non) avec un IMC (indice de masse corporelle) inférieur à 18,5
 - V4. Souffrir actuellement d'un épisode dépressif majeur
- 6 sous-populations (i.e domaines): sexe*âge en 3 postes (20-34, 35-49, 50-69)



ined

INSTITUT
NATIONAL
D'ÉTUDES
DÉMOGRA
PHIQUES

3. Les résultats



3. Résultats - ensemble

V.I	Total (milliers)	En %						
		CV1	CV2	CV3	Propor -tion	CV1	CV2	CV3
V1	1 485	3,8	3,4	3,4	3,8	3,8	3,3	3,3
V2	3 672	2,4	2,2	2,2	9,4	2,4	2,1	2,1
V3	742	5,3	5,1	5,0	1,9	5,3	4,8	4,8
V4	3 926	2,4	2,0	2,0	10,0	2,4	1,9	1,9

3. Résultats – hommes 20-34 ans

V.I	Total (milliers)	En %						
		CV1	CV2	CV3	Proportion	CV1	CV2	CV3
V1	76	19,3	12,3	12,3	1,4	19,2	15,9	16,0
V2	258	10,9	7,3	7,2	4,7	10,7	9,4	9,4
V3	9	46,1	38,6	39,1	0,2	46,1	50,3	50,9
V4	392	8,3	6,1	5,9	7,1	8,1	7,7	7,6

3. Résultats – synthèse

V.I	Total			Proportion		
	CV1 >= CV3	CV1 < CV3	Nb	CV1 >= CV3	CV1 < CV3	Nb
V1	5	2	7	7	0	7
V2	4	3	7	7	0	7
V3	5	2	7	5	2	7
V4	7	0	7	7	0	7
Total	21	7	28	26	2	28



ined
INSTITUT
NATIONAL
D'ÉTUDES
DÉMOGRA
PHIQUES

Conclusion



Conclusion

- L'estimateur simplifié paraît assez satisfaisant
- Pourrait être étendu à d'autres types de procédures (*surveyfreq*, *surveyreg*...)
- Le tester sur davantage de variables
- Autres méthodes d'estimation de la « vraie » variance?
- Utilisation de procédures prenant en compte le plan à encourager chez les analystes

Références

- Ardilly, Pascal. 2006. *Les techniques de sondage*. 2e édition. Paris: Technip.
- Hamel, Christelle, Alice Debauche, Elisabeth Brown, Amandine Lebugle, Tania Lejbowicz, Magali Mazuy, Amélie Charruault, Sylvie Cromer, et Justine Dupuis. et al. 2016. « Viols et agressions sexuelles en France : premiers résultats de l'enquête Virage ». *Population et Sociétés*, n° 538 (novembre).
- Hansen, M.H and Hurwitz, W.N. 1943. “On the theory of sampling from finite populations”, *Annals of Mathematical Statistics*, 14, p. 333-362
- Warszawski, Josiane, Antoine Messiah, Joseph Lellouch, Laurence Meyer, et Jean-Claude Deville. 1997. « Estimating Means and Percentages in a Complex Sampling Survey: Application to a French National Survey on Sexual Behaviour (ACSF) ». *Statistics in Medicine* 16: 397-423.



ined
INSTITUT
NATIONAL
D'ÉTUDES
DÉMOGRA
PHIQUES

Merci de votre attention

Nicolas Razafindratsima

Géraldine Charrance

Gwennaëlle Brilhaut

razafind@ined.fr

Geraldine.Charrance@ined.fr

Gwennaelle.Brilhaut@ined.fr