

## **Etude sur la non-réponse dans le recensement de la population 2010 de l'Argentine**

**Claudio Comari<sup>1</sup>**  
**Augusto E. Hoszowski<sup>2</sup>**

### **Résumé**

L'effet du nombre de questions sur la qualité d'une enquête est un sujet largement étudié. Les chercheurs savent qu'un long questionnaire peut avoir deux effets indésirables: une augmentation du taux de non-réponse et l'apparition d'un biais de mesure. Grâce à ses caractéristiques, le recensement de la population 2010 de l'Argentine a permis d'étudier les effets de la taille du questionnaire sur une question en particulier: le travail au cours de la semaine précédant le jour du recensement. Il a été possible de quantifier l'augmentation de la non réponse et le biais de mesure. Bien que les effets de la non-réponse et du biais étudiés dans ce travail puissent être considérés comme négligeables sur les estimations finales du recensement, ce travail apporte des preuves précieuses sur l'impact de l'extension du nombre de questions.

**Mots clés:** Non-réponse, taux d'activité, recensement de la population.

---

<sup>1</sup> Instituto Nacional de Estadística y Censos – Argentine — [ccoma@indec.mecon.gov.ar](mailto:ccoma@indec.mecon.gov.ar)

<sup>2</sup> Instituto Nacional de Estadística y Censos – Argentine — [ahosz@indec.mecon.gov.ar](mailto:ahosz@indec.mecon.gov.ar)

## 1. Introduction

### 1.1. Le recensement de population 2010 en Argentine.

Le recensement de population 2010 en Argentine s'est déroulé le mercredi 27 octobre 2010. Afin d'optimiser les ressources, dans le recensement de la population 2010 deux questionnaires (un questionnaire court et un questionnaire long) ont été appliqués. C'est la troisième fois que dans le recensement de la population (1980, 1991, 2010) en Argentine, dans un échantillon de logements, on recense les ménages avec un questionnaire long.

### 1.2. Plan de sondage pour le 'questionnaire long'

- Dans les agglomérations urbaines de 50,000 habitants et plus, 15% environ des ménages ont été recensés avec le questionnaire long. Les unités de sondage ont été des 'grappes' de 38 logements contigus. Certaines grappes, à cause de leurs particularités, ont été tirées avec la probabilité 1. Le reste de l'échantillon a été tiré avec un tirage systématique d'intervalle  $I=10$ .
- Dans les agglomérations urbaines de moins de 50,000 habitants, tous les ménages ont été recensés avec le questionnaire long.

### 1.3. Le bloc de questions pour déterminer la situation d'activité

Un des blocs de questions commun aux deux questionnaires est celui qui comporte les quatre questions permettant de déterminer le «statut d'activité». Ce bloc est appliqué à toutes les personnes de 14 ans et plus. Les questions de ce bloc sont

Questionnaire court	Questionnaire long
I. (Q13) La semaine dernière, avez-vous travaillé au moins une heure?	I. (Q25) La semaine dernière, avez-vous travaillé au moins une heure? (Oui →VI)
II. (Q14) La semaine dernière, avez-vous fait un petit travail, aidé un proche...?	II. (Q26) La semaine dernière, avez-vous fait un petit travail, aidé un proche...? (Oui →VI)
III. (Q15) La semaine dernière, aviez-vous un travail, mais vous étiez en vacances, en congé...?	III. (Q27) La semaine dernière, aviez-vous un travail, mais vous étiez en vacances, en congé...? (Oui →VI)
IV. (Q16) Au cours des quatre dernières semaines, avez vous cherché un travail?	V. (Q28) Au cours des quatre dernières semaines, avez vous cherché un travail?

(Q29) Quelle est l'activité de l'entreprise où vous travaillez le plus d'heures? (branche d'activité) (Question ouverte)
VI. (Q30) Quelle est votre profession? (description de l'occupation)

(Question ouverte)
VII. (Q31) Combien de personnes travaillent dans l'entreprise où vous travaillez? (Question fermée)
VIII. (Q32) Dans ce travail, êtes-vous à votre compte, salarié, indépendant?
IX. (Q33) Vous travaillez dans le secteur public, privé ou mixte?
X. (Q34) Dans ce travail, avez-vous des cotisations de retraite?
XI. (Q35) Dans ce travail, faites-vous des cotisations de retraite

Le questionnaire long comporte, outre les quatre questions de base, sept questions supplémentaires qui sont posées seulement aux *actifs occupés*. Parmi ces sept questions deux sont ouvertes (VI et VII).

Comme dans les enquêtes emploi, une personne est classée *active occupée* si elle répond *Oui* à l'une des questions I, II ou III. Cependant la plupart des actifs occupés répondent *Oui* à la première question ('La semaine dernière, avez-vous travaillé au moins une heure..').

Ce travail porte sur la non réponse et le *biais de mesure* dans la question I, dans le questionnaire long.

*La question étudiée est numérotée P13 dans le questionnaire court, et P25 dans le questionnaire long*

## 2. Domaine géographique retenu pour l'étude

Pour pouvoir étudier *l'effet questionnaire*, on a profité du fait que les questionnaires court et long comportaient le même bloc de questions de base dans les grandes agglomérations urbaines.

L'étude est restreinte à l'agglomération urbaine *Gran Buenos Aires*:

Du fait de sa taille (~12,000,000 habitants), l'échantillon de 'grappes' tiré de l'agglomération Gran Buenos Aires a une taille qui fait que les erreurs de sondage des estimations issues du questionnaire long soient négligeables.

## 3. Trois types de 'grappes'

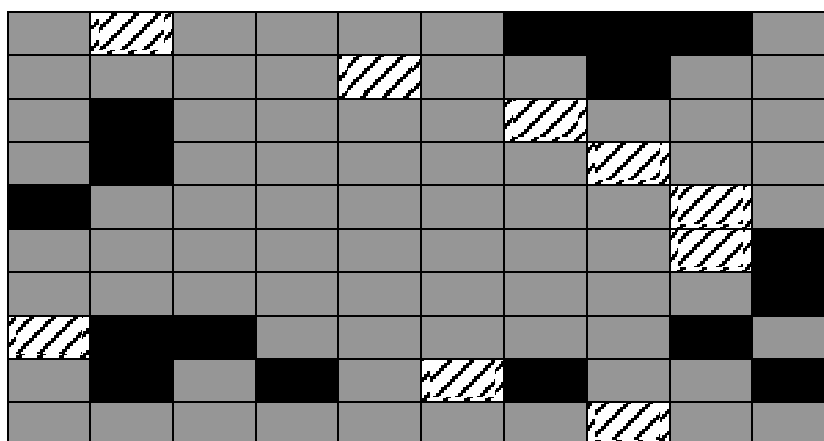
Dans les domaines où on a tiré un échantillon pour être enquêté avec le questionnaire long, l'univers de 'grappes' est divisé en trois sous-univers:

(A) Echantillon de 'grappes' tiré avec probabilité 1 (Rectangles noirs). Recensé avec questionnaire long. Ce sont des grappes situées, en général, dans des quartiers très défavorisés.

(B) Echantillon de 'grappes' tiré avec probabilité 0.1 (Rectangles rayés). Recensé avec le questionnaire long.

(C) Echantillon de 'grappes' complément de l'échantillon (B). 'Tiré' avec probabilité 0.9 (Rectangles gris). Recensé avec le questionnaire court.

Représentation du domaine GBA= (A) U (B) U (C)



- Questionnaires longs (grappes tirées avec probabilité 1)
- Questionnaires longs (grappes tirées avec probabilité 1/10)
- Questionnaires courts (grappes tirées avec probabilité 9/10)

Le tableau suivant présente le nombre de personnes et de grappes de l'agglomération *Gran Buenos Aires*, selon le type de grappe.

Prob d'inclusion des grappes	Personnes	Grappes
(A) $\pi_i = 1$	472,995	8,721
(B) $\pi_i = 0.1$	940,187	9,579
(C) $\pi_i = 0.9$	8,406,771	78,088

#### 4. Une même équipe d'agents censitaires

Pour le recensement 2010 il n'y a pas eu de formation différente pour les enquêteurs qui ont travaillé avec le questionnaire long. Ceci a éliminé la possibilité d'un *effet formation*.

#### 5. Répartition de la variable I (P13/P25)

Les tableaux suivants montrent la répartition de la variable p25/p13 (*La semaine dernière, avez-vous travaillé au moins une heure?*), selon le type de questionnaire (personnes de 14 ans et plus).

Questionnaire	P13/P25 Personnes de 14 ans et plus		
	Oui %	Non %	Valeurs manquantes %
Court ( $\pi_i = 0.9$ )	55.60	40.17	4.23
Long ( $\pi_i = 0.1$ )	51.33	42.89	5.78
Long ( $\pi_i = 1$ )	51.61	39.21	9.17
Total	55.01	40.39	4.60

Questionnaire	P13/P25. Personnes de 14 ans et plus Sans valeurs manquantes	
	Oui %	Non %
Court ( $\pi_i = 0.9$ )	58.05	41.95
Long ( $\pi_i = 0.1$ )	54.48	45.52
Long ( $\pi_i = 1$ )	56.83	43.17
Total	57.66	42.34

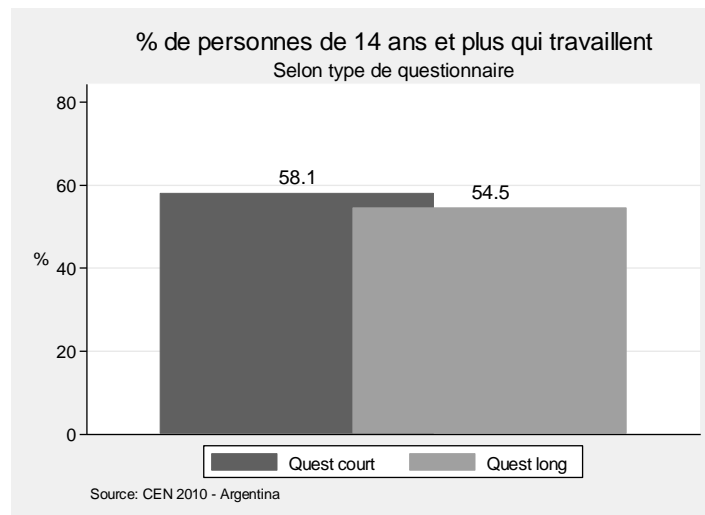
On observe:

-Le taux de non réponse dans les grappes tirées avec probabilité  $\pi_i = 1$  (9.17%) est sensiblement supérieur aux autres: Ce sont des grappes situées généralement dans des quartiers très défavorisés.

-Le taux de non réponse dans les grappes tirées avec probabilité  $\pi_i = 0.1$  (questionnaire long) est de 1.6 point supérieur à celui observé dans les grappes tirées avec probabilité  $\pi_i = 0.9$  (questionnaire court)

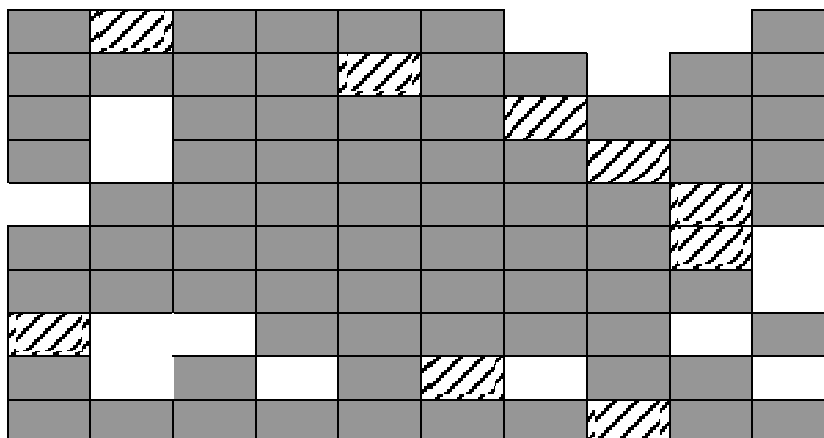
-Le questionnaire court estime à environ 4.3 points de plus que le questionnaire long le taux de personnes qui travaillent, avec les valeurs manquantes, et d'environ 3.7 points sans considérer les valeurs manquantes.

Ce qui nous intéresse ici est d'étudier la sous estimation du taux d'activité dans l'échantillon de grappes tirées avec probabilité  $\pi_i = 0.1$ , vis à vis les grappes tirées avec probabilité  $\pi_i = 0.9$ . Les grappes tirées avec probabilité  $\pi_i = 1$  sont écartées de l'étude.



## 6. Sous univers étudié

Pour pouvoir quantifier *'l'effet questionnaire'*, il faut écarter les grappes tirées avec probabilité 1. Il nous reste alors l'univers (B) U (C), représenté par deux échantillons complémentaires et disjoints: (B) (gris) et (C) (rayé):



## 7. Contrôle de l'échantillon de 'grappes'

La sous-estimation du taux de personnes qui travaillent dans le questionnaire long (par rapport à l'estimation avec le questionnaire court) peut-être due à:

- I. Un problème de tirage/implémentation de l'échantillon.
- II. Une plus forte non-réponse ciblée sur certaines catégories démographiques (e.g. hommes de 20 à 60 ans)

I. L'échantillon est tiré dans les bureaux centraux, mais les délais du recensement ne permettent pas un contrôle minutieux des différentes étapes du 'travail sur le terrain'.

II. Le questionnaire long comporte une séquence supplémentaire de sept questions, dont deux ouvertes, pour les personnes ayant répondu 'Oui' à la question p25/p13 (Avez-vous travaillé la semaine dernière...). Les agents censitaires ne sont pas des enquêteurs expérimentés →

La probabilité de laisser la question p25/p13 sans réponse ou encore de finir l'enquête avant de poser cette question est plus élevée pour certaines catégories démographiques.

Pour vérifier ou infirmer ces deux hypothèses, on présente les estimations de la structure de la population selon quelques caractéristiques corrélées avec le taux d'activité.

	Questionnaire		Différence (C-L)
	Court (C)	Long (L)	
Type de logement			
% Maison	83.3	84.0	-0.7
% Appartement	12.4	12.1	0.2
% 'Rancho/ Casilla' (logements précaires)	3.5	3.1	0.4
% Autres	0.9	0.8	0.0
Disponibilité d'eau dans le logement			
% Dans le logement	89.0	89.9	-0.9
% Hors du logement	11.0	10.1	0.9
Taille du ménage			
% Une personne	14.8	15.4	-0.5
% Deux personnes	21.9	22.4	-0.5
% Trois personnes	20.8	21.0	-0.2
% Quatre et plus	42.5	41.3	1.1
Sexe (personnes de 14 ans et plus)			
% Hommes	47.8	47.8	0.0
% Femmes	52.2	52.2	0.0
Tranche d'âge			
% 14-17	8.8	8.8	0.0
% 18-30	27.7	27.6	0.1
% 31-65	51.3	51.4	-0.1
% 66-120	12.1	12.2	-0.1
Scolarité			
Proportion de jeunes 14 à 24 ans qui suivent des études	41.7	41.9	-0.3
Lien de parenté			
% Chef de ménage	29.9	30.3	-0.4
% Conjoint	19.0	19.1	-0.1
% Fils/fille	40.3	40.1	0.2
% Autres	10.9	10.6	0.3

Les tableaux montrent que la composition des deux échantillons ne diffère pas significativement pour une série de caractéristiques corrélées au taux d'activité. Les différences sont plus élevées pour les variables portant sur le logement, à cause de l'*effet de grappe*.

La seconde possibilité c'est qu'il y ait une non réponse ciblée sur certaines sous populations. Et que ce soit *seulement* cette non réponse la cause de la sous estimation du taux d'activité dans le questionnaire long.

Nous calculons:

- La moyenne de p25/p13 (*Avez-vous travaillé la semaine dernière.. ?*) pour différentes sous populations.
- La différence, dans chaque sous population, de l'estimation à partir du questionnaire long et du questionnaire court.

Questionnaire	Sexe		Total %
	Hommes %	Femmes %	
Court	71.6	45.7	58.1
Long	68.9	41.2	54.5
Différence	2.7	4.5	3.6

Questionnaire	Tranche d'âge			
	14-17 %	18-30 %	31-65 %	66-120 %
Court	13.2	64.9	71.1	17.4
Long	9.0	61.1	67.9	13.0
Différence	4.2	3.8	3.2	4.4

Questionnaire	Lien de parenté			
	Chef de ménage %	Conjoint %	Fils/fille %	Autres %
Court	69.8	50.2	51.0	49.6
Long	67.4	45.6	46.7	45.5
Différence	2.4	4.6	4.3	4.1

Questionnaire	Taille du ménage			
	Une personne %	Deux personnes %	Trois personnes %	Quatre et plus %
Court	52.1	55.4	62.2	58.0
Long	49.3	52.1	59	54.1
Différence	2.8	3.3	3.2	3.9



Les tableaux ci-dessus indiquent:

- Dans toutes les catégories démographiques de base, le questionnaire long sous-estime systématiquement la proportion de personnes de 14 ans et plus qui ont travaillé dans la semaine de référence.
- La différence est plus élevée pour les catégories qui ont un plus faible taux d'activité: Femmes, conjoints du chef de ménage, personnes de 14 à 17 ans, personnes de 66 ans et plus.
- Pour les catégories avec un faible taux d'activité, la différence relative est très forte:
  - -9.8% pour les femmes (de 45.7% à 41.2%)
  - -31.8% pour les personnes de 14 à 17 ans (de 13.2% à 9.0%)
  - -25.3% pour les personnes de 66 ans et plus (de 17.4% à 13.0%)

Une plus forte non réponse dans ces catégories ne peut pas être responsable, à elle seule, de la sous-estimation du taux d'activité. Ceci a été vérifié en imputant avec hot-deck aléatoire les valeurs manquantes, avec les variables auxiliaires:

Sexe, tranche d'âge, taille du ménage et lien de parenté

L'imputation n'a presque pas modifié les estimations:

Estimation avec questionnaire long Valeurs imputées	Estimation avec questionnaire long Sans valeurs manquantes
57.31%	57.66%

Ce résultat suggère que la non réponse n'a pas été 'aléatoire' dans les cellules définies par les variables: tranche d'âge, taille du ménage et lien de parenté.

### 8. Probable biais de mesure dans la question p25/p13

Même si la non-réponse n'a pas été 'aléatoire', les différences entre les estimations des deux questionnaires ne peuvent être expliquées qu'en supposant un biais de mesure. Le tableau suivant montre que la différence entre les taux de non-réponse est systématiquement inférieure à la différence entre les estimations:

	Différences entre les taux de non réponses (A)	Différences entre les estimations (B)	(B)-(A)
<b>Total</b>	<b>1.6</b>	<b>3.6</b>	<b>2.0</b>
Hommes	1.7	2.7	1.0
Femmes	1.4	4.5	3.1

<b>Tranche d'âge</b>			
14-17	-0.9	4.2	5.1
18-30	1.2	3.8	2.6
31-65	1.9	3.2	1.3
66-120	3.1	4.4	1.3
<b>Relation de parenté</b>			
Chef de ménage	2.3	2.4	0.1
Conjoint	2.0	4.6	2.6
Fils/fille	0.2	4.3	4.1
Autres	1.0	4.1	3.1
Chef de ménage	2.3	2.4	0.1
<b>Taille du ménage</b>			
Une personne	3.1	2.7	-0.4
Deux personnes	2.1	3.3	1.2
Trois personnes	1.5	3.8	2.3
Quatre personnes et plus	0.8	4.1	3.3

## 9. Conclusions

Le plan de sondage de l'échantillon pour le questionnaire long dans le recensement de la population 2010 en Argentine, et l'existence d'un bloc de questions communes avec le questionnaire court a permis d'étudier deux conséquences dues au nombre de questions supplémentaires (dans le questionnaire long) posées aux personnes qui déclarent avoir travaillé pendant la semaine de référence:

- Augmentation de la non-réponse
- Biais de mesure

Ce dernier effet a été vérifié en comparant la différence entre les non réponses aux deux questionnaires, et la différence entre les estimations des deux questionnaires. Le biais de mesure peut être due à une changement délibéré d'un 'Oui' pour un 'Non' (pour éviter de poser les questions ouvertes) ou plus probablement, l'agent censitaire coche le 'Non' sans poser la question, en supposant le 'Non' comme réponse du répondant. Les fortes différences dans les catégories 'Tranche d'âge 14-17 ans' et 'Conjoint' peuvent soutenir cette hypothèse. Un point à noter c'est que dans le cas d'un recensement de la population, les agents censitaires ne sont pas d'enquêteurs entraînés, augmentant ainsi les probabilités des biais de mesure. Finalement, s'il n'y avait pas eu de questionnaire court, il aurait été impossible de détecter ce biais.