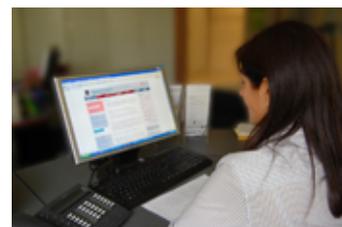


Utilisation d'une enquête légère auprès des non-répondants pour corriger la non-réponse par la méthode des groupes de réponse homogènes

Journées de Méthodologie Statistique 2015

Henri Bodet, pôle Ingénierie Statistique d'Enquête.



Deux populations aux comportements de réponse différents

1/ Exemples de situation : EMMGI

Enquête sur les modes et moyens de gestion de l'Immatériel.

Recouvre la gestion de la propriété intellectuelle, de la réputation, la notoriété, les connaissances.

Certaines entreprises ne sont pas concernées par le sujet de l'enquête – on peut penser qu'elles s'abstiendront plus facilement de répondre.

La réponse à l'une des questions de l'enquête est dans cette situation un déterminant du comportement de réponse.

Deux populations aux comportements de réponse différents

2/ Exemples de situation : FAS-GSO

Enquête sur la filière « aéronautique et spatiale » dans le grand Sud-Ouest.

On est amené à interroger des entreprises dont on ne sait pas si elles appartiennent à la filière ou non.

Là encore, on peut penser que des entreprises qui ne sont pas dans la filière seront moins portées à répondre car elles se jugeront « non-concernées » par l'enquête.

Ici aussi, la réponse à l'une des questions de l'enquête est dans cette situation un déterminant du comportement de réponse.

Une enquête auprès des non-répondants

3/ Dans les deux situations : soupçon de biais

Dans les deux situations : un taux de non-réponse qui, sans être excessif, était suffisamment fort pour qu'on y prête une attention particulière.

Dans les deux cas, on a réalisé une enquête téléphonique auprès des non-répondants pour déterminer si, oui ou non, les deux populations ont des comportements de réponse différents.

Qu'est-ce qu'une enquête légère ?

Enquête « légère » : exemple de FAS-GSO

Elle consiste à interroger l'entreprise par téléphone pour savoir si, oui ou non, elle travaille pour la filière (ou à obtenir l'information autrement).

On ne veut surtout pas remplir le questionnaire pour les non-répondants !

Le caractère « léger » (une question) de l'opération permet d'avoir l'information pour presque toutes les unités et aussi de ne pas augmenter la charge de réponse.

Conséquence : La réponse à une seule question (filiale ou non?).

Utiliser l'enquête ou pas ?

La question se pose de savoir s'il faut intégrer les résultats de l'enquête.

Et aussi de savoir comment le faire.

Utiliser l'enquête ou pas ?

Solution à exclure : n'intégrer que les unités « non-concernées » parce que « leur questionnaire est complet ».

Solution envisagée :

utiliser l'enquête pour estimer la « vraie » probabilité de réponse.

Comment utiliser l'ENR ?

La non-réponse est corrigée par la méthode des groupes de réponses homogènes.

En théorie, il faudrait construire les groupes suivant l'appartenance ou non à la filière.

Or, cette information est inconnue pour les non-répondants.

L'application directe de la méthodes des GRH n'est donc pas possible.

Comment utiliser l'ENR ?

En fait, on n'a pas besoin de savoir dans quel groupe est chaque non-répondant.

On a juste besoin de connaître leur nombre.

Or, l'ENR permet justement de l'estimer.

On calcule donc la probabilité de réponse dans chaque GRH en remplaçant le nombre de non-répondants dans le GRH par son estimation.

Utilisation sur l'enquête FAS-GSO

La population enquêtée se partageait en deux ensembles :

- une population « cœur de la filière » qui a l'habitude d'être enquêtée.
- une population « potentiellement dans la filière » qui n'a pas l'habitude d'être enquêtée.

Exemple sur l'enquête FAS-GSO

Sur la population « cœur de la filière »

- Un bon taux de réponse ;
- conformément à l'intuition, un meilleur taux de réponse « dans la filière » ;
- une faible différence de comportement entre les deux groupes (dans et hors de la filière).

Exemple sur l'enquête FAS-GSO

Sur la population « potentiellement dans la filière »

- Un plus faible taux de réponse ;
- contrairement à l'intuition, un meilleur taux de réponse « hors de la filière » ;
- une forte différence de comportement entre les deux groupes.

Sur cette population, on a utilisé l'ENR pour appliquer la méthode des GRH.

Exemple sur l'enquête FAS-GSO

Comparaison entre les méthodes de redressement pour la partie
« potentiellement dans la filière ».

	Méthode de redressement	Nombre de salariés dédiés à la filière	Chiffre d'affaires réalisé avec la filière	Écart sur le nombre de salariés	Écart sur le chiffre d'affaires
Avant calage sur marges	GRH sans prise en compte de l'ENR	23 464	3 590	3,7 %	4,0 %
	GRH avec prise en compte de l'ENR	24 325	3 733		

Exemple sur l'enquête FAS-GSO

Comparaison entre les méthodes de redressement pour la partie
« potentiellement dans la filière »

	Méthode de redressement	Nombre de salariés dédiés à la filière	Chiffre d'affaires réalisé avec la filière	Écart sur le nombre de salariés	Écart sur le chiffre d'affaires
Après calage sur marges	GRH sans prise en compte de l'ENR	27 841	4 394	2,4 %	3,9 %
	GRH avec prise en compte de l'ENR	28 502	4 566		

La méthode fonctionne-t-elle ?

La suite de l'exposé se penche sur la question de la pertinence théorique de la méthode.

Il s'agit d'essayer de quantifier l'arbitrage biais-variance qu'elle représente.

Quel est l'impact de l'utilisation de l'ENR ?

Effectivement, en procédant ainsi :

On supprime le biais dû à la différence de comportement de réponse.

On rajoute une troisième phase de sélection aléatoire et on augmente donc la variance.

Illustration par des simulations

Dans le cas d'un fort biais de non-réponse.

simulations dans le cas où $p_0 = 0.8$ et $p_1 = 0.4$

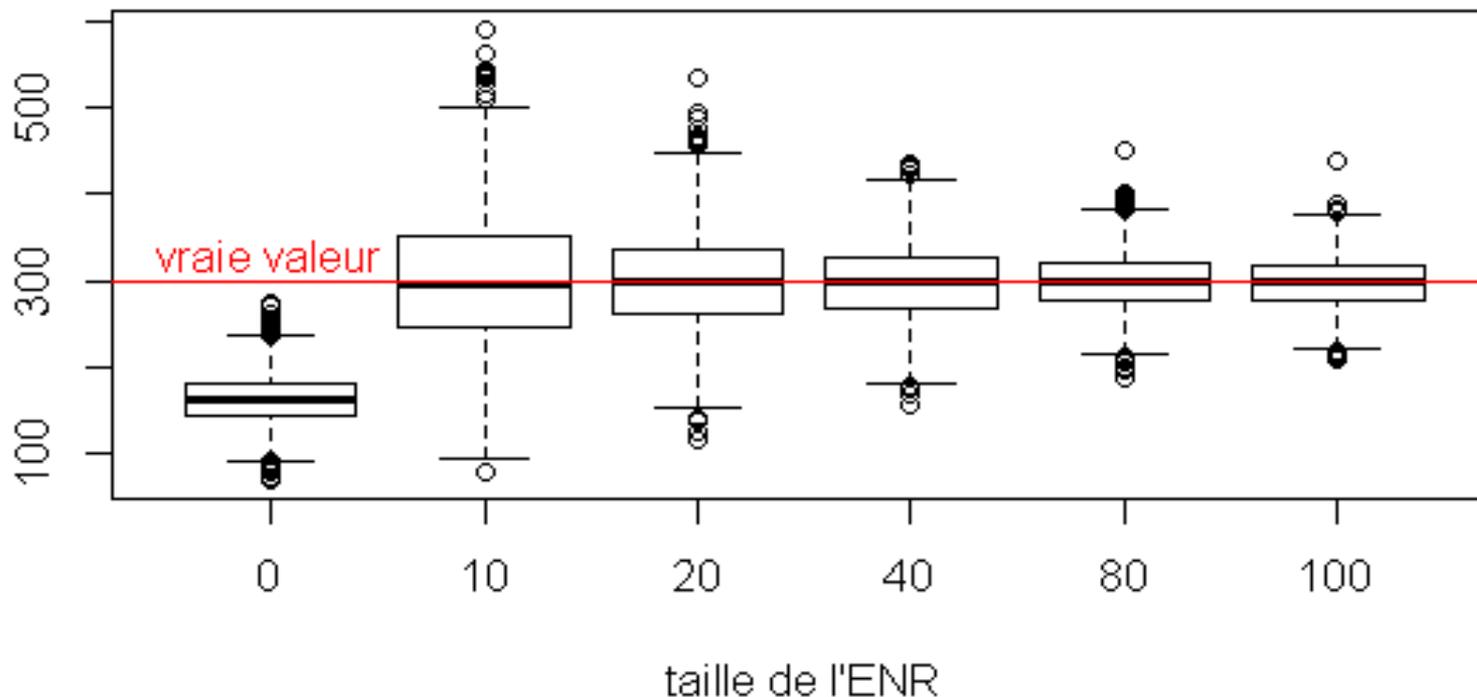


Illustration par des simulations

Dans le cas d'un faible biais de non-réponse.

simulations dans le cas où $p_0 = 0.75$ et $p_1 = 0.69$

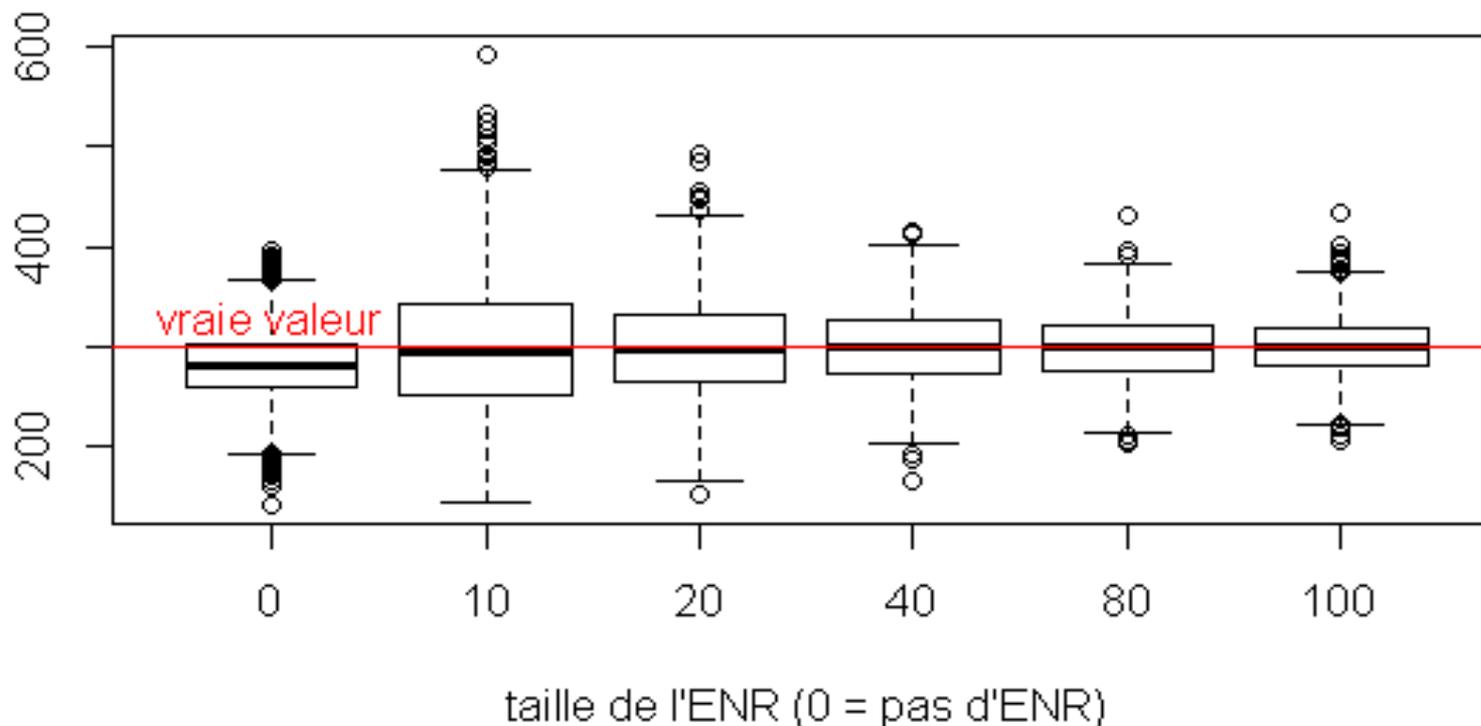
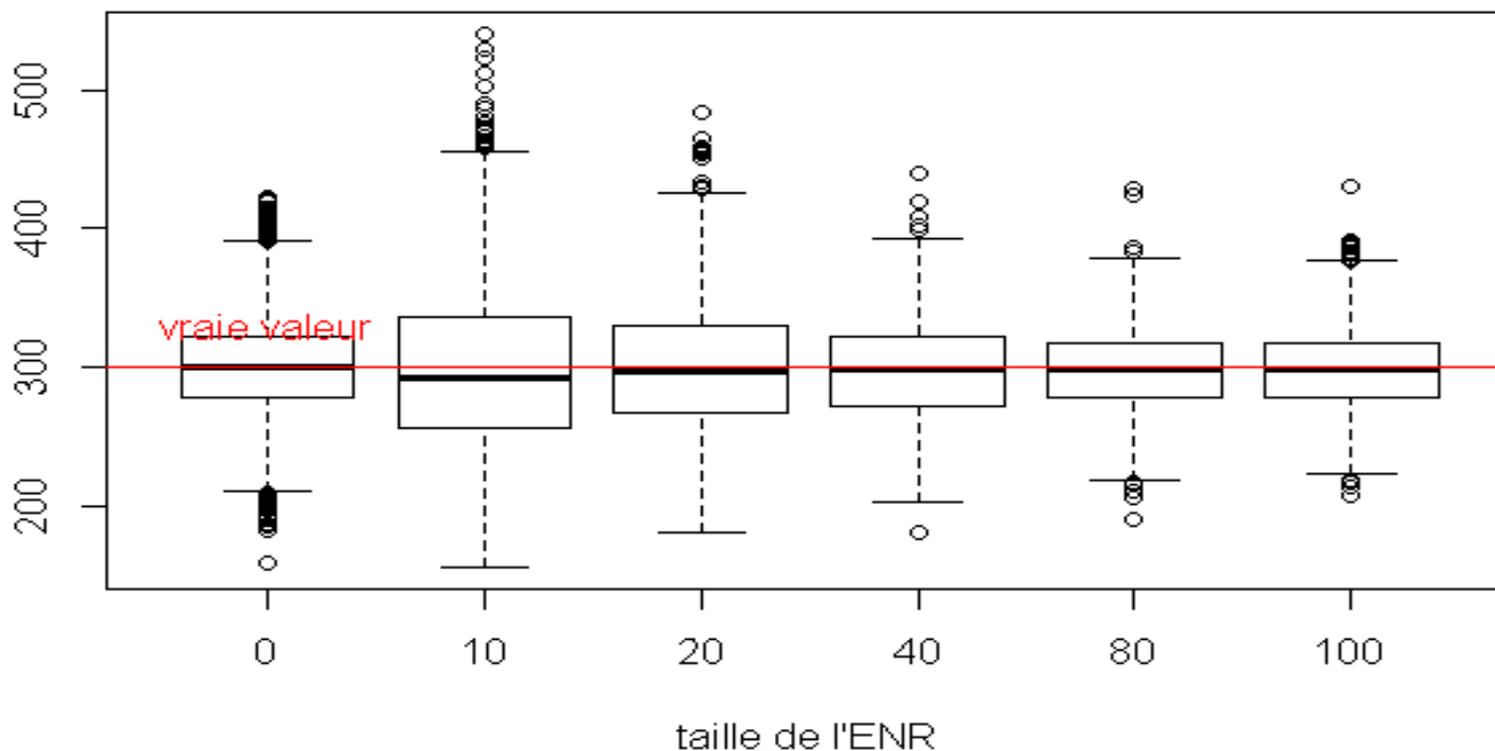


Illustration par des simulations

Dans le cas d'une absence de biais de non-réponse.

simulations dans le cas où $p_0 = 0.74$ et $p_1 = 0.74$



Étude théorique de l'estimateur

En se plaçant dans un cadre très simplifié.

On peut estimer par le calcul :

- Le biais dû à la différence de comportement de réponse.
- La variance supplémentaire due à la prise en compte de l'enquête auprès des non-répondants.

En comparant le biais et la « variance supplémentaire », on voit qu'à partir d'une certaine taille de l'ENR, la réduction du biais l'emporte sur l'augmentation de la variance.

Étude théorique de l'estimateur

On peut en fournir un ordre de grandeur :

$$k^* = \left[\frac{1}{n(f_1 q_1 + (1 - f_1) q_0)} + \frac{B^2}{f_1 q_1 (1 - f_1) q_0} \right]^{-1}$$

n	f ₁	p ₁	p ₀	B	k*
500	0,15	0,4	0,8	- 0,06	3
500	0,15	0,75	0,69	-0,02	54

Retour sur les simulations

En se plaçant dans le cas d'un faible biais de non-réponse, on voit que l'expression précédente fournit bien un ordre de grandeur de la taille minimale de l'ENR.

taille de l'enquête auprès des non-répondants	Écart-type	espérance de l'estimateur (vraie valeur :300)	Erreur moyenne	Supplément de variance constaté sur les simulations	valeur donnée par l'approximation
0 (pas d'ENR)	33	280	13%	///	///
10	68	299	23%	3538	3647
20	50	299	17%	1402	1671
40	38	298	13%	368	683
80	31	298	10%	-102	189
100	29	298	10%	-218	90

Conclusion

1 / Sur la validité théorique de la méthode.

La méthode permet effectivement de supprimer le biais dû au comportement de réponse différent.

Ceci se fait au prix d'une aggravation de la variance.

La validité dépend de l'ampleur du biais et de la taille de l'enquête auprès des non-répondants.

Dans un cas simplifié, on peut déterminer une taille minimale pour que la démarche soit valide.

Conclusion

2 / Sur l'expérimentation.

Qu'on décide de l'intégrer ou pas, l'enquête sur les non-répondant permet d'établir le sens et l'ampleur du biais dû à la différence de comportement de réponse.

La prise en compte de l'ENR pour constituer les GRH oblige à considérer des groupes de réponses homogènes plus grands.

On peut atténuer le biais en constituant les GRH sur des variables auxiliaires liées à celle dont on ne dispose pas.

La prise en compte d'un calage sur marge réduit l'écart entre les méthodes.