

ECHANTILLONNAGE ET MARGE D'ERREUR : GRANDS PRINCIPES

Origine et objectif de la fiche :

Les retours d'expérience des collectivités locales ayant réalisé ou fait réaliser des enquêtes relatives à la prévention des déchets mettent en évidence leur difficulté à définir l'échantillon † : *« Quelle est la taille optimale de mon échantillon pour garantir des résultats fiables ? Comment construire un échantillon représentatif de ma population ? Quel degré de confiance je peux accorder à mes résultats ? »*

Lorsqu'on réalise une enquête, on a rarement les moyens de contacter tous les membres de la population qui nous intéressent sans quoi on réaliserait un « recensement ». Cette population de référence est, ce que les statisticiens appellent, la « population mère ». Pour réaliser l'enquête, on doit se limiter à une population plus réduite (c'est à dire l'échantillon) qui est censée représenter la population-mère et qui doit nous permettre de généraliser les résultats observés.

L'objectif de cette fiche est d'identifier des éléments clefs qui permettent à une collectivité de penser l'échantillonnage de sa population, soit pour rédiger un CCTP avec une indication de la taille de l'échantillon, soit pour effectuer l'enquête en interne.

Sommaire :

A – La taille de l'échantillon	2
B – Le mode d'échantillonnage	4
C – La représentativité de l'échantillon	5
D – L'enquête exhaustive	6

A. La taille de l'échantillon :

Quatre paramètres sont à prendre en compte lorsque l'on amorce la réflexion sur la taille d'un échantillon :

- ❶ **le taux de sondage** : c'est-à-dire la part de la population interviewée dans la population totale.
- ❷ **la marge d'erreur** : c'est-à-dire l'estimation statistique de l'amplitude d'erreur sur les résultats (lorsqu'une modalité représente 50% des réponses avec une marge d'erreur de 5%, cela veut dire que si le même sondage est refait dans les mêmes conditions, la réponse peut varier entre 45% et 55%).
- ❸ **le nombre de « sous-population »** : c'est-à-dire des cibles plus restreintes que l'on souhaite étudier à l'intérieur même de l'échantillon... Chacune doit avoir un effectif suffisant pour que les résultats soient significatifs, plus les sous-populations sont nombreuses, plus la taille de l'échantillon global sera grande.
- ❹ **les contraintes budgétaires** : les limites du commanditaire et la recherche d'optimisation du rapport qualité / coût.

Dans les sciences humaines, on considère qu'une marge d'erreur de 5% est suffisante, seuil que l'on atteint avec un échantillon de 400 personnes. Quant au taux de sondage, il n'a aucune influence sur la marge d'erreur, c'est la taille de l'échantillon qui est importante (pour exemple, les sondages d'opinion nationaux sont effectués sur des échantillons de 1 000 personnes).

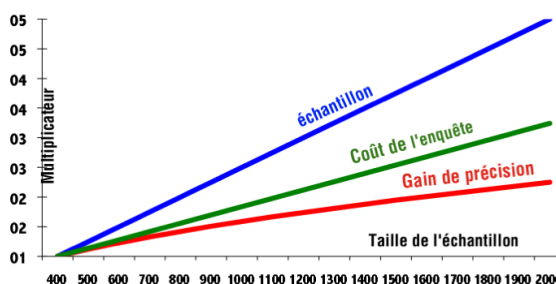
Outil de calcul de la marge d'erreur en fonction de la taille de l'échantillon : possibilité de cliquer dans le tableau pour changer la taille de l'échantillon et obtenir la marge d'erreur correspondante.

Echantillon	400
Marge d'erreur	4,90%

Exemple de calcul de marge d'erreur en fonction de différentes tailles d'échantillon :

Echantillon	100	125	150	175	200	225	250	275	300	325
Marge d'erreur	9,80%	8,77%	8,00%	7,41%	6,93%	6,53%	6,20%	5,91%	5,66%	5,44%
Echantillon	400	500	600	700	800	900	1 000	1 100	1 200	1 300
Marge d'erreur	4,90%	4,38%	4,00%	3,70%	3,46%	3,27%	3,10%	2,95%	2,83%	2,72%

La marge d'erreur diminue avec l'augmentation de la taille de l'échantillon, mais pas de manière proportionnelle : Il convient donc de trouver le rapport optimum entre la précision de l'enquête et son coût.



La taille de l'échantillon global est aussi déterminée en fonction du nombre de « sous-populations » que l'on souhaite analyser, ce qui peut amener à augmenter la taille de

l'échantillon global pour conserver des marges d'erreurs acceptables dans chaque sous-population.

C'est le cas d'un EPCI qui souhaite observer les résultats sur l'ensemble de son territoire mais aussi pour chaque collectivité adhérente ou « sous territoire ».

C'est aussi le cas d'un EPCI qui souhaite observer les résultats selon le critère de l'habitat « collectif » ou « individuel » soit 2 « sous-populations ».

Les tableaux ci-dessous illustrent le rapport entre : le nombre de « sous-territoires », leur effectif (liée à la marge d'erreur souhaitée) et la taille de l'échantillon global nécessaire au final.

Nombre de "sous-territoire"	2	3	4	5	6	7	8	9	10
Effectif pour chaque "sous-territoire"	100	100	100	100	100	100	100	100	100
dont la marge d'erreur est de	9,8%								
La taille de l'échantillon global est de	200	300	400	500	600	700	800	900	1000
marge d'erreur sur le global	6,9%	5,7%	4,9%	4,4%	4,0%	3,7%	3,5%	3,3%	3,1%

Nombre de "sous-territoire"	2	3	4	5	6	7	8	9	10
Effectif pour chaque "sous-territoire"	150	150	150	150	150	150	150	150	150
dont la marge d'erreur est de	8,0%								
La taille de l'échantillon global est de	300	450	600	750	900	1050	1200	1350	1500
marge d'erreur sur le global	5,7%	4,6%	4,0%	3,6%	3,3%	3,0%	2,8%	2,7%	2,5%

Nombre de "sous-territoire"	2	3	4	5	6	7	8	9	10
Effectif pour chaque "sous-territoire"	200	200	200	200	200	200	200	200	200
dont la marge d'erreur est de	6,9%								
La taille de l'échantillon global est de	400	600	800	1000	1200	1400	1600	1800	2000
marge d'erreur sur le global	4,9%	4,0%	3,5%	3,1%	2,8%	2,6%	2,5%	2,3%	2,2%

Recommandation :

Pour un échantillon sans découpage en « sous-territoire » nous préconisons au minimum une taille d'échantillon global de 400 questionnaires aboutis pour garantir une marge d'erreur inférieure à 5%.

Lorsque l'on a des « sous-territoire » leur effectif minimum est de 100 questionnaires.

B. Le mode d'échantillonnage :

Il existe deux modes d'échantillonnage : proportionnel ou stratifié qui permettent d'ajuster la taille de l'échantillon global à la finesse des résultats que l'on souhaite observer.

L'échantillon proportionnel dans lequel chaque critère a le même poids dans la population mère et dans l'échantillon. Que ce soit les critères socioprofessionnels (catégories sociales, âge, sexe...) mais aussi des critères d'habitat (collectif ou individuel), de territoire (urbain, semi-urbain ou rural) ou des « sous territoires ».

ECHANTILLON PROPORTIONNEL	Poids dans la population	Poids dans l'échantillon	Effectif enquêté	Marge d'erreur
Critère 1	32%	32%	192	7,1%
Critère 2	28%	28%	168	7,6%
Critère 3	22%	22%	132	8,5%
Critère 4	18%	18%	108	9,4%
Total	100%	100%	600	4,0%

C'est la méthode la plus simple. Les résultats sont directement exploitables sans redressement statistique.

Elle est privilégiée lorsque les écarts entre les sous populations sont faibles et que tous les effectifs sont suffisants pour effectuer des analyses fiables.

L'échantillon stratifié dans lequel les critères n'ont pas le même poids dans l'échantillon et dans la population mère. Cette méthode offre la possibilité d'obtenir des marges d'erreur plus raisonnables et adaptées à chaque critère afin de permettre des traitements pertinents, mais l'analyse sur l'ensemble de l'échantillon nécessitera un redressement statistique afin de rétablir le poids réel de chaque critère. Un coefficient de pondération sera attribué à chaque critère afin de lui redonner son poids réel dans la population.

ECHANTILLON STRATIFIE	Poids dans la population	Poids dans l'échantillon	Effectif enquêté	Marge d'erreur
Critère 1	32%	25%	150	8,0%
Critère 2	28%	25%	150	8,0%
Critère 3	22%	25%	150	8,0%
Critère 4	18%	25%	150	8,0%
Total	100%	100%	600	4,0%

La méthode est plus complexe. Elle est privilégiée lorsque les écarts entre les sous populations sont importants et que certains effectifs sont insuffisants pour effectuer des analyses fiables.



Pour ne pas augmenter d'une manière démesurée la taille de l'échantillon global il est important de minimiser le nombre de « sous populations » donc de regrouper les territoires...

C. La représentativité de l'échantillon :

Que l'échantillon soit construit à partir d'un échantillon proportionnel ou stratifié, **quand on connaît la structure de la population de référence**, la représentativité est assurée par la **méthode des quotas**. Un échantillon peut être représentatif suivant une, deux ou trois variables (Catégorie professionnelle, sexe, âge par exemple). Les données de base sont issues de recensement de la population (Données INSEE : <http://www.recensement-2006.insee.fr/home.action>)

Les paramètres sur lesquels la représentativité est recherchée sont choisis en fonction des objectifs de l'enquête. Dans les enquêtes sur la Prévention des déchets en direction des CITOYENS, c'est bien la structure de la population qui importe mais aussi les critères d'habitat.

Exemple illustré : construction d'un échantillon de 600 questionnaires aboutis pour un EPCI composé de 4 groupements de communes. Identification du nombre de questionnaires nécessaires à réaliser dans chaque groupement de commune selon la catégorie socioprofessionnelle de la personne de référence dans le foyer.

EPCI Source INSEE RP2008 - Population des ménages par taille du ménage et catégorie socioprofessionnelle de la personne de référence	CC 1	CC 2	CC 2	CC 4	TOTAL EPCI
Agriculteurs exploitants	845	1 149	275	928	3 197
Artisans, commerçants, chefs entreprise	1 191	2 062	912	1 683	5 848
Cadres, professions intellectuelles sup.	2 395	3 813	2 494	2 957	11 659
Professions intermédiaires	3 881	6 975	2 917	4 462	18 235
Employés	1 373	2 691	1 236	1 770	7 070
Ouvriers	4 903	10 232	2 716	6 362	24 213
Retraités	3 384	7 106	2 711	4 330	17 531
Autres sans activité professionnelle	425	570	171	312	1 478
TOTAL	18 398	34 598	13 430	22 803	89 229
Poids réel de la population					
Agriculteurs exploitants	1%	1%	0%	1%	4%
Artisans, commerçants, chefs entreprise	1%	2%	1%	2%	7%
Cadres, professions intellectuelles sup.	3%	4%	3%	3%	13%
Professions intermédiaires	4%	8%	3%	5%	20%
Employés	2%	3%	1%	2%	8%
Ouvriers	5%	11%	3%	7%	27%
Retraités	4%	8%	3%	5%	20%
Autres sans activité professionnelle	0%	1%	0%	0%	2%
TOTAL	21%	39%	15%	26%	100%
Echantillon PROPORTIONNEL					
Agriculteurs exploitants	6	8	2	6	21
Artisans, commerçants, chefs entreprise	8	14	6	11	39
Cadres, professions intellectuelles sup.	16	26	17	20	78
Professions intermédiaires	26	47	20	30	123
Employés	9	18	8	12	48
Ouvriers	33	69	18	43	163
Retraités	23	48	18	29	118
Autres sans activité professionnelle	3	4	1	2	10
TOTAL	124	233	90	153	600

Quand on ne connaît pas la structure de la population de référence la représentativité de l'échantillon est assurée par la construction d'un **échantillon aléatoire**, où chaque citoyen a la même probabilité d'être sélectionné. C'est une procédure de tirage aléatoire.

Si l'enquête est réalisée par téléphone à partir d'un fichier de propriétaires de composteur individuel par exemple. Aucune donnée sociologique ne détermine le « profil de ces propriétaires » de fait, pour effectuer un tirage aléatoire, une **règle de sélection doit être établie** préalablement : « *je prends en compte une adresse toutes les 3 adresses* » par exemple ou alors « *je prends en compte un page sur deux de mon fichier* ». La règle doit être suivie sur toute la durée de la construction de l'échantillon.

Dans les enquêtes en face à face, pour affiner la méthode, il est possible de déterminer des strates de jours et/ou de lieux d'enquête par exemple. Si l'enquête cible les « vacanciers d'un territoire ». Pour construire un échantillon représentatif aléatoire, une règle de sélection doit être aussi établie : il peut être décidé d'interviewer : « *tous les jours de la semaine, pendant les 8 semaines d'été, 10 personnes par jour, présentes sur un lieu identique (un marché, une plage...)* » soit 560 questionnaires aboutis. Mais la règle pourrait être « *tous les jours de la semaine, 1 semaine sur deux, pendant les 8 semaines d'été, 20 personnes par jour, présentes sur un lieu identique (un marché, une plage...)* » ce qui représente aussi 560 questionnaires aboutis. **L'important est de couvrir le maximum de diversité** (les vacanciers du mois de juillet ne sont pas les mêmes que ceux du mois d'août ; ceux qui se déplacent le WE diffèrent aussi des vacanciers en semaine etc.) et de **respecter la règle sur toute la durée de la construction de l'échantillon**.

Si l'enquête est effectuée auprès d'une population présente lors d'un évènement spécifique : un congrès, une animation en GMS, une réunion... La construction d'un échantillon en mode aléatoire reste d'actualité. La règle de sélection doit être adaptée à l'évènement : « *je propose d'interviewer une personne sur trois qui s'intéressent à mon animation* » ou « *dans cette salle je vais interviewer, tous les 2 rangs, la dernière personne de la rangée* » etc.

D. L'enquête exhaustive :

Dans le cas où l'enquête est réalisée sur l'ensemble (ou presque) de la population ciblée on parle d'enquête exhaustive. La représentativité est assurée de fait. Il n'y a pas d'interrogation sur la taille de l'échantillon dans la mesure où l'ensemble de la population est interrogé.

Ces situations sont plutôt exceptionnelles et correspondent à des situations précises : petite collectivité ou enquête sur un public très ciblé lors d'un évènement spécifique. Dans ce dernier cas, l'enquête reflète des résultats attachés à la population de cet évènement. Ils ne sont pas généralisables à l'ensemble de la population du territoire. (Cf : M5 : Réaliser une enquête en interne : possibilités et limites).