

# CALCUL DE LA TAILLE D'UN ÉCHANTILLON POUR UNE ENQUÊTE

## CHAPITRE 3B

Lorsque l'on mène une enquête quantitative auprès d'une large population (ménages d'une commune), il n'est ni possible, ni nécessaire, d'interroger toute la population concernée. Cette fiche indique la taille que doit avoir l'échantillon pour obtenir des résultats d'enquête représentatifs de l'ensemble de la population.

La démarche de calcul peut être divisée en cinq étapes.

- Quelle est la population cible ?
- Jusqu'où doit aller l'exactitude ?
- De quelle taille d'échantillon ai-je besoin ?
- Les personnes interrogées répondront-elles ?
- Combien de personnes dois-je contacter ?

### I. QUELLE EST LA POPULATION CIBLE ?

La première étape est de déterminer la population cible (ménages, entreprises, etc.) ainsi que sa taille. Il peut être utile d'effectuer un recensement ou de faire une estimation s'il n'y a pas d'autres solutions (notamment lorsque l'on a affaire à de très grandes populations, par exemple à l'échelle d'un pays).

### II. JUSQU'OUÙ DOIT ALLER L'EXACTITUDE ?

Dans la mesure où une enquête ne porte pas sur 100 % de la population cible, les réponses obtenues sont forcément légèrement biaisées. On devra donc déterminer l'exactitude nécessaire et suffisante pour atteindre les objectifs de l'enquête. Deux paramètres doivent être considérés : la marge d'erreur et l'intervalle de confiance.

## 1. La marge d'erreur

La marge d'erreur représente la fourchette de certitude à l'intérieur de laquelle les réponses obtenues sont exactes. Elle est généralement comprise entre 1 et 10 %. Prenons pour exemple une enquête réalisée avec une marge d'erreur de 5 %, dans laquelle 50 % des personnes interrogées ont déclaré posséder une toilette. Cela signifie que 45 à 55 % de la population dans son ensemble est réellement équipée de toilette.

Une marge d'erreur de 5 % est communément considérée comme suffisante et il est déconseillé de choisir une marge supérieure à 10 %.

## 2. L'intervalle de confiance

L'intervalle de confiance est la probabilité que l'échantillon de personnes interrogées ait une influence sur les résultats de l'enquête. Prenons pour exemple une enquête réalisée avec un échantillon de cent personnes et un intervalle de confiance de 95 %. Cela signifie que si la même enquête est réalisée une seconde fois avec cent autres personnes choisies parmi la population cible, les résultats obtenus seront, dans 95 % des cas, identiques à ceux de la première enquête. L'intervalle couramment utilisé est de 95 %. Il est déconseillé de prendre un intervalle inférieur à 90 %.

## III. DE QUELLE TAILLE D'ÉCHANTILLON AI-JE BESOIN ?

L'échantillon est calculé avec la formule suivante<sup>1</sup> :

$$n = \frac{t_p^2 \times P(1 - P) \times N}{t_p^2 \times P(1 - P) + (N-1) \times y^2}$$

Avec :

- n : taille de l'échantillon.
- N : taille de la population cible (nombre de ménages, d'usagers, etc.), réelle ou estimée.
- P : proportion attendue d'une réponse de la population ou proportion réelle. Si l'enquête ne porte que sur un seul critère (l'équipement en toilettes) et qu'une étude a déjà été menée sur le sujet, proposant un résultat chiffré (40 % d'équipement en toilettes par exemple), la proportion est donnée par le résultat précédemment obtenu (0,4 dans cet exemple), ou par le résultat attendu. Dans le cas d'une étude multicritère ou lorsque aucune autre étude n'a été réalisée, ce qui est fréquent dans le domaine de l'assainissement, elle peut être fixée à 0,5 par défaut, ce qui permet d'avoir le plus grand échantillon possible.
- $t_p$  : intervalle de confiance d'échantillonnage. Le tableau n° 1 donne les valeurs de  $t_p$  associées aux intervalles de confiance.
- y : marge d'erreur d'échantillonnage.

<sup>1</sup> REA L.M. *et al.*, 1997.

TABLEAU N° 1

Valeurs de  $t_p$  associées aux intervalles de confiance

Intervalle de confiance	$t_p$
90 %	1,65
95 %	1,96
99 %	2,69

TABLEAU N° 2

## Calcul de la taille d'échantillon en fonction de plusieurs paramètres

Population	Marge d'erreur <sup>(a)</sup>			Intervalle de confiance <sup>(b)</sup>		
	10 %	5 %	1 %	90 %	95 %	99 %
< 100	Population entière			Population entière		
100	50	80	99	74	80	88
500	81	218	476	176	218	286
1 000	88	278	906	215	278	400
10 000	96	370	4 900	264	370	623
100 000	96	383	8 763	270	383	660
1 000 000	97	384	9 513	271	384	664

<sup>(a)</sup> Avec  $t_p = 1,96$ .

<sup>(b)</sup> Avec  $y = 5 \%$ .

À partir d'une population de plus de 20 000 personnes/ménages/etc., la taille de l'échantillon varie peu pour des marges d'erreur supérieures à 5 %.

## IV. COMBIEN DE PERSONNES DOIS-JE CONTACTER ?

Les personnes contactées pour l'enquête peuvent choisir de refuser d'y participer et d'y répondre, raison pour laquelle il est préférable de contacter un nombre de personnes légèrement supérieur à celui initialement prévu pour l'échantillonnage. Le mode de contact influe sur le pourcentage de réponses : le démarchage à domicile peut atteindre un taux de réponse supérieur à 90 %, tandis qu'une enquête diffusée par courrier électronique atteindra au mieux les 30 %.

Le nombre de personnes contactées doit être adapté au taux de réponse estimé.



**POUR ALLER PLUS LOIN**

SurveyMonkey, *Taille de l'échantillon de sondage* [consulté le 21/03/2014], <https://fr.surveymonkey.com/mp/sample-size/>.