



Recommandations sur les méthodes statistiques simplifiées appliquées à des projets humanitaires

Jean Lapegue – Senior Advisor WASH ACF-France

1 - Introduction sur les méthodes robustes statistiques et l'échantillonnage

Les projets humanitaires sont de plus en plus soumis à produire des évidences des besoins ou de l'impact de leurs activités. De plus, il est de plus en plus fréquent de publier ces données dans le cadre de publications scientifiques ou de les présenter lors de colloques scientifiques.

Des procédures statistiques robustes, mais qui restent simples et applicables pratiquement sur le terrain, sont essentielles pour les acteurs comme ACF, complétant ainsi par la possibilité de valoriser son action des projets dont la qualité et la pertinence technique sont reconnus.

D'autre part, conduire une enquête robuste ou une vraie analyse statistique n'est pas forcément plus complexe ou ne prend pas plus de temps que de tenter une méthode approximative. Une analyse statistique ou une enquête non robuste rend les résultats inutilisables, impossibles à publier et inexportables, et jette potentiellement le discrédit sur les projets et l'organisation.

Pour effectuer des enquêtes robustes, il convient de sélectionner soigneusement l'échantillon d'étude. Cette étape est primordiale pour être en mesure d'interpréter les résultats de l'étude sur l'échantillon au niveau de la population totale.

Il s'agit tout d'abord d'opter pour la technique d'échantillonnage la plus adaptée compte tenu :

- a) des objectifs majeurs de l'étude menée (qui conditionnera le degré de précision),
- b) des caractéristiques de la population d'étude (taille, groupes différenciés),
- c) des contraintes de terrain.

Au-delà de cette notion de représentativité, le principe de l'échantillonnage implique que tous les individus de la population considérée doivent avoir, au mieux, la même probabilité de faire partie de l'échantillon choisi.

2 - Définitions

Population et individu: P

Ensemble exhaustif des individus de la zone concernée par l'enquête (il peut s'agir de personnes, familles, foyers, infrastructures, etc). Un élément de cet ensemble est appelé individu.

Echantillon: N

Groupe d'individus qui seront enquêtés. L'échantillon N est censé être représentatif de la population P enquêtée.

Caractère et modalités

La particularité commune que l'on étudie est appelée caractère.

Les valeurs prises par le caractère sont aussi appelées les modalités. La différence des valeurs extrêmes du caractère s'appelle l'étendue.

Effectif n_i et fréquence f_i

Le nombre d'individus (n_i) d'une modalité est appelé effectif.

Le nombre total d'individus de la population enquêtée est appelé effectif total (N).

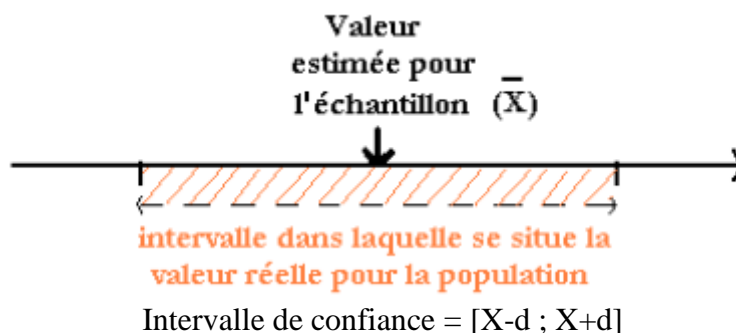
Le rapport $f_i = n_i / N$ est appelé fréquence. f_i est un nombre toujours compris entre 0 et 1 (exprimé en %). La somme des nombres f_i est toujours égale à 1.

Degré de Précision d

Le degré de précision estime l'intervalle de confiance dans lequel on va situer la valeur réelle de la population. Ce dernier est exprimé en points de pourcentage (ex : +/- 5%).

(Exemple: si le taux de diarrhée des enfants de l'échantillon est de 20% avec un degré de précision de +/- 5%, alors on peut dire que ce taux se situe entre 15% et 25% pour la population totale des enfants).

On choisira **5%** dans le cas d'enquêtes épidémiologiques (qualité biologique d'une eau, taux de diarrhées, etc.) ou **10%** dans tous les autres cas (enquête CAP, enquête sanitaire, etc.).



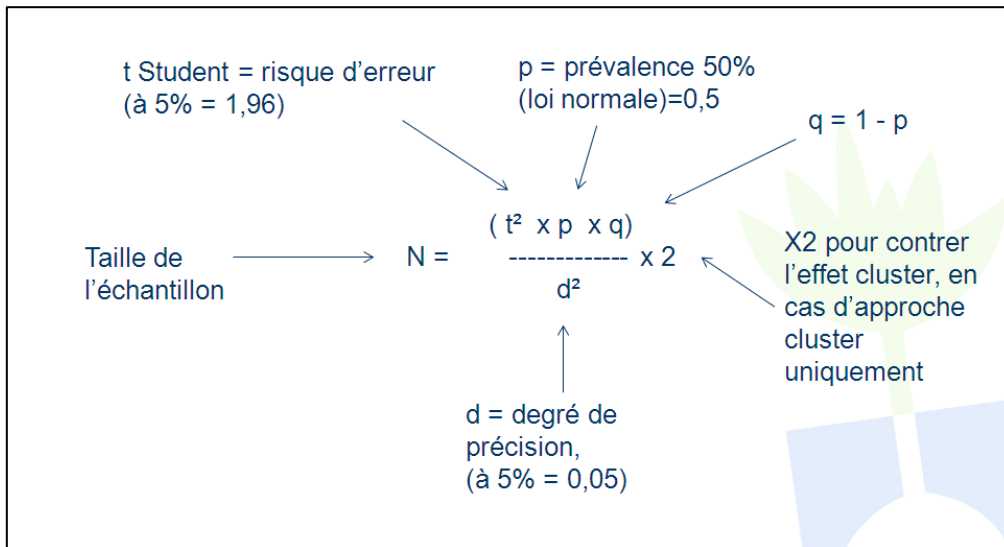
Représentativité:

Capacité de l'échantillon à représenter la population. La taille de l'échantillon ne contribue que dans une certaine mesure à la représentativité de la population¹.

¹ Par exemple, pour une méthode par grappe, avec 10% de degré de précision, et au-delà d'une population homogène de 1500 individus, la taille de l'échantillon n'est plus fonction de la population.

2 - Formule classique de calcul de la taille de l'échantillon

Il faut savoir comment est établie cette formule, mais on préférera utiliser la table (section 6) pour estimer la taille souhaitable de l'échantillon, selon la méthode choisie.



3 - Différentes méthodes d'échantillonnage usuelles

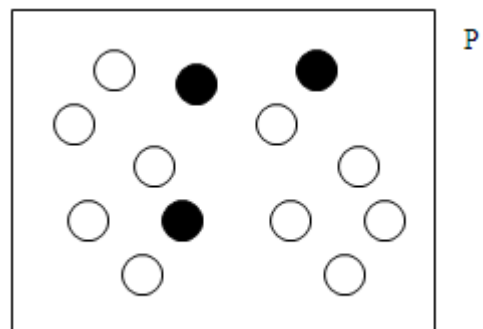
On ne présente ici que les deux méthodes qui doivent être utilisées et connues en priorité² et qui répondent à la majorité des cas pratiques du terrain:

- la méthode aléatoire simple, à utiliser seulement si l'on a la liste nominale de tous les individus, ce qui est rare.
- la méthode par grappe, fortement recommandée dans tous les cas

- Individu enquêté
- Individu non enquêté

a) Aléatoire simple

Méthode seulement possible si on a la liste des individus (par exemple noms de familles d'un camp de réfugiés, patients d'un hôpital, etc). Le choix des individus de l'échantillon doit se faire de manière aléatoire (pas en tirant les noms au hasard mais en utilisant par exemple la fonction Excel statistique **ALEA.ENTRE.BORNES(x; y)** de votre tableau Excel.



² On ne parlera volontairement pas ici des autres méthodes aléatoires (échantillonnage pondéré, échantillonnage aléatoire stratifié proportionnel) ou non-aléatoires (échantillonnage raisonné, systématique et cumulatif), ni encore des méthodes mixtes associant les méthodes par grappe et méthodes aléatoires.

b) Grappe (cluster)

Il s'agit de la méthode qui **doit être utilisée en priorité** dans le cadre des enquêtes statistiques liées aux projets humanitaires. Cette méthode est recommandée par l'OMS et UNICEF (enquêtes nutritionnelles). Elle est simple, rapide et robuste. On décompose la population en 30 « grappes » qui seront chacune représentée par 5 à 7 individus³ (pour un degré de précision de 10% qui s'applique aux enquêtes CAP ou aux enquêtes sanitaires) ou 15 à 26 individus (pour un degré de précision de 5% qui convient mieux aux enquêtes épidémiologiques, qualité biologique, etc). Donc 210 ou 780 interviews selon le degré choisi.

Note : si l'on veut comparer deux groupes d'une même population (par exemple les habitants ayant bénéficié d'un projet d'ACF et ceux qui n'ont pas bénéficié), il faudra appliquer la méthode statistique choisie pour chacun des deux groupes. Avec une méthode cluster, et un degré de précision de 10%, cela porte le nombre d'individus interviewés à 420 (2 fois 210).

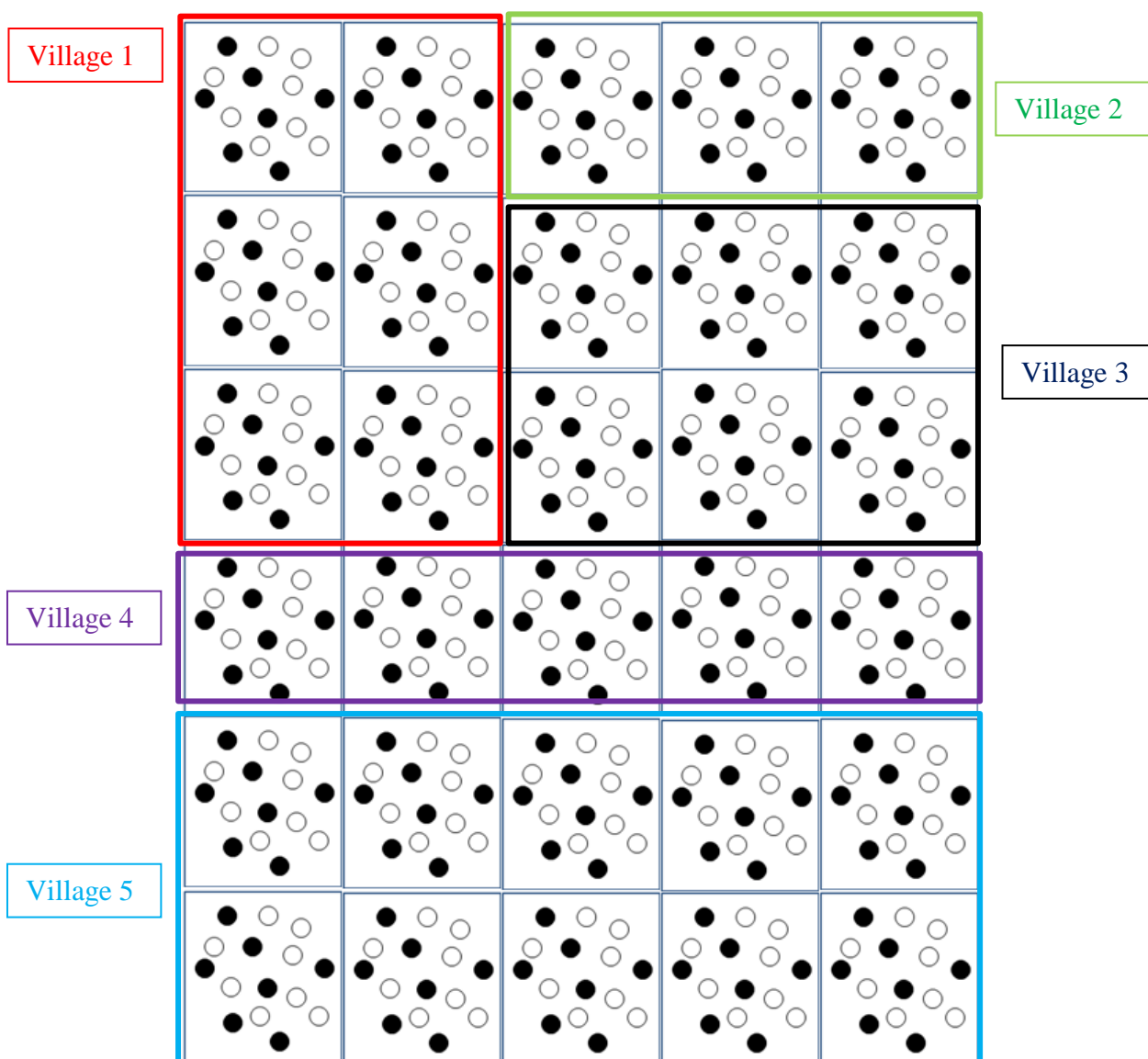


Figure : Distribution des 30 grappes contenant chacune 7 individus interrogés, représentative de la population totale de la zone (5 villages)

³ Voir table section 6 pour le nombre d'individus interviewés par grappe, qui dépend de la taille de la population

4 – Exercice pratique (approche par grappe)

ACF débute un projet WASH dans une population de 158070 personnes, dispersées entre 4 gros villages du Nord-Ouest de la RCA. Calculer le nombre d'interviews nécessaires par village pour avoir une bonne représentativité des résultats d'une enquête CAP?⁴

a) Questions que se pose l'enquêteur (*En bleu, réponse proposée*)

Q1 - Quelle méthode: aléatoire ou grappe? *Grappe (car pas la liste nominale des ménages)*

Q2 – Quel degré de précision? *10% (pas de dimension épidémiologique dans un CAP)*

Q3- Quelle est la taille de l'échantillon N? *210 individus (familles) seront interrogés*

Q4- Combien de grappes? *30 (méthode classique UNICEF-OMS)*

Q5- Combien de ménages interviewés par grappe? *7 (méthode UNICEF-OMS à 10%)*

Q7 – Combien d'interview auront lieu par village?

a) Chacune des 30 grappes va représenter $158070/30 = 5269$ personnes

b) La répartition des grappes se fait intuitivement comme cela :

Village	Population	Nb Clusters	Nb interviews
Bouforo	12000	2	14
Ana	43000	8	56
Bocaranga	62070	12	84
Baboua	41000	8	56
Total	158070	30	210

5 – Critères de choix de la méthode?

	Echantillonnage exhaustif	Echantillonnage aléatoire simple	Echantillonnage par groupement
Méthode	Chaque ménage est interrogé.	Sélection aléatoire d'unité parmi la population générale. Chaque unité a une chance égale d'être incluse dans l'enquête.	Les ensembles sont sélectionnés aléatoirement parmi la population totale. Plusieurs individus de chaque groupe sont alors interrogés.
Avantages	Bonne représentativité Bonne connaissance de la population	Idéale pour des études statistiques	Fait gagner du temps dans le transport et réduit aussi les coûts
Désavantages	Long à mettre en oeuvre Long pour l'enregistrement des données Coût élevé	Difficile à atteindre en pratique Nécessite une liste précise de toute la population Coûteuse à conduire du fait de l'éparpillement des échantillons sur la zone entière	Les unités proches les unes des autres peuvent être très similaires et donc peu représentatives de l'ensemble de la population L'erreur d'échantillonnage est plus grande que pour un échantillonnage aléatoire simple
Critères	Petite population cible	Dans tous les cas avec une liste exhaustive de la population cible	Dans tous les cas, même avec des populations dispersées

⁴ Nous ne possédons pas le détail de la liste nominale des ménages dans cette simulation

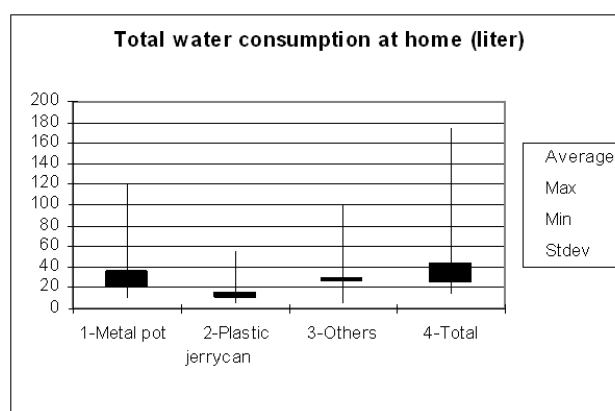
6 – Echantillonnage: table applicable pour méthodes aléatoire simple, grappe et stratifié

P	Aléatoire simple				Groupement (30 grappes)					
	Précision 5%		Précision 10%		Précision 5%			Précision 10%		
	Taille d'échantillon (N)	% du nb de ménage	Taille d'échantillon (N)	% de Ménage	Taille d'échantillon (N)	# d'échantillon par groupes	% de Ménage	Taille d'échantillon (N)	# d'échantillon par groupes	% de Ménage
<=200	132	66 %	65	32 %	Pas pertinent	Pas pertinent	Pas pertinent	150	5	75 %
<=500	217	43 %	81	16 %	450	15	90 %	180	6	36 %
<=1000	278	28 %	88	9 %	570	19	57 %	180	6	18 %
<=1500	306	20 %	96*	6 %	630	21	42 %	210	7	14 %
<=2000	322	16 %	96*	5 %	660	22	33 %	210	7	11 %
<=3000	341	11 %	96*	3 %	690	23	23 %	210	7	7 %
<=4000	350	9 %	96*	2 %	720	24	18 %	210	7	5 %
<=4500	384*	9 %	96*	2 %	780	26	9 %	210	7	5 %
<=5000	384*	8 %	96*	2 %	780	26	8 %	210	7	4 %
<=10000	384*	4 %	96*	1 %	780	26	4 %	210	7	2 %
<=100000	384*	0,4 %	96*	0,1 %	780	26	0,4 %	210	7	0,2 %

7 – Présentation des résultats, groupe test, stratégies de mesure d'impact, éthique

Lors de la présentation des résultats de l'enquête statistique :

- Toujours expliquer le choix de la méthodologie et le justifier (degré de précision, méthode choisie, etc)
- Quand on travaille avec deux groupes (un groupe projet qui bénéficie des projets d'ACF et un groupe 'contrôle' qui n'en bénéficie pas) il faut conserver une approche éthique : (par exemple on testera le même groupe avant et après un projet).
- Quand on veut comparer deux groupes (cas de zone hétérogène) chacun des groupes testés devra bénéficier d'une méthode robuste, sous peine de ne pas pouvoir les comparer.
- Les résultats de l'enquête doivent être systématiquement présentés accompagnées de leur intervalle de confiance (écart type ou 'standard déviation' en anglais). Cette fonction **ECARTYPE(x ; y)** est facile à trouver sous Excel (cf. Figure ci-dessous).



Total water consumption by day

- Lorsque l'on compare deux enquêtes (par exemple CAPs pré et post projet) il est recommandé d'interroger les mêmes individus.