CATHOLIC RELIEF SERVICES

CRS Échantillonne

GUIDE SUR LE CALCUL DE LA TAILLE DES ÉCHANTILLONS QUANTITATIFS

Ce guide a été rédigé par Stacy Prieto, PhD. Il s'inspire largement de l'ouvrage *Going beyond simple sample size calculations : a practitioner's guide de* Brendon McConnell et Marcos Vera-Hernández, qui rend ce guide plus accessible au personnel de la CRS et applicable au contexte opérationnel de la CRS.

Photo de couverture par Katie Price

Catholic Relief Service | 228 W. Lexington Street, Baltimore, MD 21201, États-Unis | crs.org | crsespanol.org
Pour plus d'informations, contactez stacy.prieto@crs.org.

Table des matières

Liste des figures

Liste des tableaux

Remerciements

L'auteur tient à remercier Benjamin Allen, James Campbell, Tony Castleman, Heather Dolphin, John Hembling et TD Jose de CRS, Stephanie Martin de TANGO International et Eric Gerber de l'Université Northeastern pour leurs contributions importantes au contenu de ce guide.

Ce guide est dédié au personnel du programme de suivi, d'évaluation, de responsabilisation et d'apprentissage (SERA) du projet CRS qui travaille sans relâche pour améliorer ses connaissances et sa capacité à mettre en œuvre des pratiques SERA mondiales de haute qualité.

Abréviations

ANJE	Alimentation du nourrisson et du jeune enfant
BHA	Bureau de l'USAID pour l'assistance humanitaire
CCI	Coefficient de corrélation intra-groupe
CI	Intervalle de confiance
EMD	Effet Minimal Détectable
FPC	Correction de la population finie
FTF	Initiative Feed the Future de l'USAID
IMC	Indice de masse corporelle
L'IPTT	Tableau de suivi des performances des indicateurs
MIRA	Mesure des indicateurs pour l'analyse de la résilience
ONG	Organisation non gouvernementale
PPT	Probabilité proportionnelle à la taille
RMA	Régime alimentaire minimum acceptable
SD	Écart type
SE	Erreur type
SERA	Suivi, évaluation, redevabilité et apprentissage
SILC	Communautés d'épargne et de prêt interne
USAID	Agence des États-Unis pour le développement international
USDA	Département de l'agriculture des États-Unis

Introduction

CRS, et l'ensemble de la communauté internationale du développement, sont devenus de plus en plus rigoureux dans leur évaluation quantitative des projets de développement. Dans ses politiques et procédures internes de suivi, d'évaluation, de redevabilité et d'apprentissage (SERA), CRS exige des études de référence et des évaluations finales pour tous les projets d'une valeur supérieure à 1 million de dollars (Catholic Relief Services 2023). Après la collecte des données de l'évaluation finale, les valeurs du tableau de suivi des performances des indicateurs (IPTT) pour les indicateurs au niveau des résultats doivent être comparées à leurs valeurs de référence respectives, afin de savoir si les moyennes ou les proportions d'indicateurs diffèrent considérablement d'une période à l'autre.

La détection d'une différence statistiquement significative entre deux moyennes ou proportions nécessite : 1) qu'une différence existe réellement et, 2) que l'échantillon représentatif à partir duquel les moyennes ou les proportions sont calculées est suffisamment grand.

En outre, les données relatives à de nombreux indicateurs annuels de suivi de la performance sont recueillies auprès d'un échantillon de bénéficiaires de projets, et cet échantillon doit être d'une taille suffisante pour obtenir une estimation raisonnablement précise de la valeur de l'indicateur parmi tous les bénéficiaires du projet.

Ce guide résume les principaux éléments à prendre en compte lors de la détermination de la taille des échantillons, avec des exemples spécifiques au travail de CRS. Il vise à combler une lacune dans les connaissances de l'organisme en ce qui concerne les calculs de la taille de l'échantillon, à réduire la confusion quant aux équations qui conviennent à l'utilisation dans quels contextes et à fournir un point de référence cohérent à l'échelle de l'organisme.

Objectif

Ce guide s'adresse aux responsables SERA de proposition, aux coordinateurs SERA de projet, aux responsables SERA des programmes nationaux et aux conseillers techniques SERA.

Ce guide devrait être utilisé au cours de la phase de conception du projet pour allouer des ressources suffisantes dans le budget du projet à la collecte de données et pour réviser les besoins en matière de collecte de données au démarrage et à la mise en œuvre du projet. Ce guide comprend des équations de taille d'échantillon qui calculent la taille d'effet minimal détectable (EMD), car les projets de CRS cherchent généralement à détecter un changement significatif entre deux points (ligne de référence et final).

L'annexe 6 comprend une analyse des guides sur la taille des échantillons de donneurs et met en évidence certaines différences clés entre ces guides et celui-ci.

Le public visé par ce guide est constitué de responsables SERA des propositions (pour la préparation du budget SERA) ; les coordinateurs SERA des projets (pour éclairer la collecte de données lors du démarrage et de la mise en œuvre) ; et les responsables SERA des programmes nationaux et les conseillers techniques SERA (pour les aider à soutenir ce qui précède).

Quand utiliser ce guide ?

Le personnel du projet doit calculer la taille de l'échantillon pour chaque base de sondage ou type de répondant qu'il interrogera à l'aide de méthodes quantitatives. Par exemple, si un projet doit sonder les agriculteurs pour suivre un indicateur, les groupes de producteurs pour en suivre

un autre et les enfants de moins de cinq ans pour en suivre un autre, ils doivent identifier l'équation appropriée et calculer la taille de l'échantillon nécessaire pour chacun.

Si les données de plusieurs indicateurs doivent être collectées à partir d'une base d'échantillonnage, calculez la taille d'échantillon nécessaire pour chaque indicateur, puis choisissez la plus grande taille calculée.

Si les données de plusieurs indicateurs doivent être recueillies à partir d'une base de sondage, la meilleure pratique consiste à calculer la taille de l'échantillon nécessaire pour chaque indicateur, puis à choisir la plus grande taille calculée, dans la limite du raisonnable. Si la plus grande taille d'échantillon est exorbitante à collecter, concentrez-vous sur 1) les indicateurs les plus importants, généralement au niveau des résultats, ou 2) les indicateurs standard des donateurs. Consultez également un conseiller technique SERA, au besoin.

Ce guide peut également être utilisé pour estimer la taille de l'échantillon nécessaire pour les indicateurs de niveau d'output. Cependant, les projets collectent généralement des données sur les indicateurs au niveau des outputs par le biais de registres de distribution ou de formation (suivi de routine), et non d'un échantillon représentatif des bénéficiaires du projet.

Comment ce guide est-il organisé ?

Ce guide fournit un arbre décisionnel pour déterminer l'équation appropriée à utiliser lors du calcul de la taille des échantillons. Il comporte ensuite une section obligatoire sur les facteurs susceptibles d'augmenter la taille de l'échantillon requis, afin de s'assurer que la taille finale de l'échantillon est adéquate. Elle est suivie d'une section facultative sur les facteurs susceptibles de réduire la taille de l'échantillon. Il est recommandé de n'utiliser cette dernière section que si la taille de l'échantillon déterminée précédemment est trop coûteuse et qu'il est nécessaire de la réduire. Le guide comprend des équations qui mesurent le changement entre deux périodes. Il comprend également des équations permettant de mener une évaluation simple (et de ne pas détecter de changement au fil du temps ou entre les groupes).

Bien que le présent document ne vise qu'à guider les calculs de la taille de l'échantillon, il comporte une brève section sur les mythes sur la taille de l'échantillon et sur la façon dont le plan d'échantillonnage affecte l'analyse des données.

Le guide comporte six annexes : 1) coefficients de corrélation intra-groupe ; 2) une feuille de calcul Excel et 3) un exemple de code R pour des calculs plus complexes ; 4) un langage passe-partout pour décrire les calculs de la taille de l'échantillon dans des documents écrits officiels (par exemple, les rapports des donateurs, les propositions de projets, etc.) ; 5) un guide de référence rapide ; et 6) un examen d'autres guides clés sur la taille des échantillons.

Informations nécessaires avant d'utiliser ce guide

Ce guide suppose que quelques étapes ont été suivies avant d'utiliser ce guide.

1) Les équipes de projet ont décidé si elles feraient des comparaisons statistiques entre les groupes avec les résultats. Par exemple, comparer les bénéficiaires du projet (groupe de traitement) à un groupe témoin ; comparaison des valeurs de référence et finales ; comparer les hommes aux femmes ; comparaison des groupes d'âge, etc.

2) L'IPTT du projet, qui contient des valeurs de référence et des valeurs cibles estimées, a déjà été élaboré. Habituellement, l'IPTT est estimé à l'étape de la conception du projet, avec un examen documentaire permettant de déterminer les estimations de la valeur de référence à partir d'autres projets ou études similaires. Il devrait s'agir d'études plus récentes portant sur des zones géographiques, des saisons et des conditions climatiques identiques ou similaires (p. ex., sécheresse ou non-sécheresse).

Après avoir examiné qui devrait utiliser ce guide, quand l'utiliser et recueilli les informations nécessaires, les équipes de projet peuvent maintenant examiner l'arbre décisionnel de la taille de l'échantillon (section 1), afin de déterminer l'équation nécessaire pour chaque indicateur.

Enregistrement et fourniture d'une aide en espèces polyvalente au Brésil. [Felippe Thomaz]

1. Arbre décisionnel de la taille de l'échantillon

La figure 1, l'arbre décisionnel de la taille de l'échantillon, est une représentation graphique de la question clé à laquelle il faut répondre pour chaque indicateur : « Quelle équation de taille d'échantillon les équipes de projet devraient-elles utiliser ? » L'arbre décisionnel guide les utilisateurs vers l'équation appropriée (il y en a huit), en fonction de la comparaison statistique des données recueillies entre les groupes ; recueillir des données pour un indicateur continu ou binaire ; et s'ils regroupent l'échantillon pendant la collecte des données.[1]

Les points (A) et (B) de la figure 1 fournissent des renseignements supplémentaires pour faciliter la prise de décision.

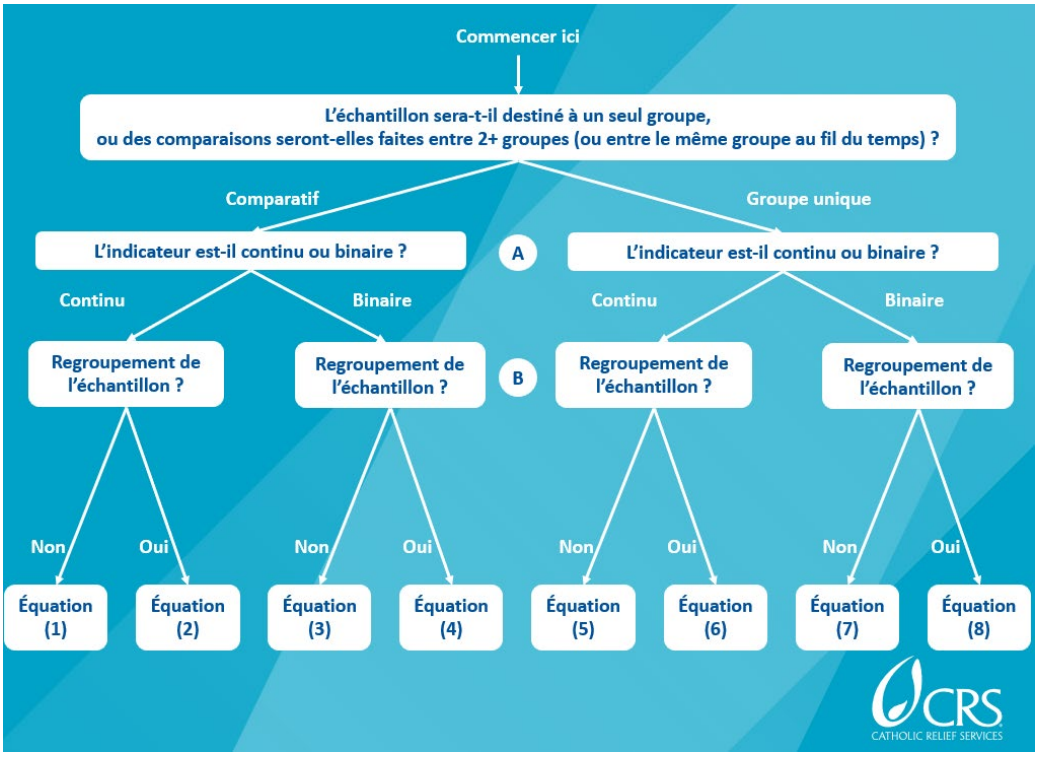

FIGURE 1. ARBRE DECISIONNEL DE LA TAILLE DE L'ECHANTILLON

Les points (A) et (B) de la figure 1 fournissent des renseignements supplémentaires pour faciliter la prise de décision. En ce qui concerne (A), les indicateurs continus et binaires suivent des distributions de probabilité différentes, ils nécessitent donc des équations de taille d'échantillon différentes :

(A) Les indicateurs continus recueillent des données qui ont un large éventail de valeurs possibles. Quelques exemples :

[1] Ce guide utilise le terme « groupe » pour décrire ce que la documentation statistique appelle habituellement « grappe ». Il s'agit d'améliorer l'accessibilité du langage pour le personnel du CRS, tout en reconnaissant qu'il n'est pas techniquement correct.

- Valeur moyenne des ventes annuelles des exploitations agricoles et des entreprises[2]
- Nombre moyen d'hectares faisant l'objet de pratiques ou de technologies de gestion améliorées
- Rendement moyen des produits agricoles ciblés
- Volume moyen des produits vendus par les exploitations agricoles et les entreprises
- Taux moyen de fréquentation en classe
- Nombre moyen de groupes d'aliments consommés par un ménage
- Indice du capital social au niveau des ménages

Les indicateurs binaires collectent des données sous la forme d'une réponse oui/non. Quelques exemples :

- Pourcentage de personnes ou d'organisations qui utilisent une pratique améliorée
- Pourcentage d'élèves qui savent lire
- Pourcentage d'enfants de moins de 5 ans souffrant d'un retard de croissance
- Pourcentage de personnes ayant accès à un financement lié à l'agriculture

(B) Le regroupement consiste à sélectionner au hasard un village, une école, une association de producteurs, etc., comme groupe, puis à sélectionner au hasard des personnes au sein de ce groupe pour l'échantillon. L'échantillonnage aléatoire simple, en revanche, signifie avoir une liste de toutes les personnes dans tous les villages, et sélectionner au hasard des personnes dans cette liste, sans d'abord sélectionner un groupe.

Souvent, les échantillons sont regroupés pour économiser de l'argent lorsque les répondants à l'enquête sont répartis sur une vaste zone géographique. Le regroupement de l'échantillon permet d'éviter d'avoir à visiter chaque village, ce qui pourrait être plus coûteux.[3]

Il convient de noter que, dans le cas du regroupement, les équations de taille des échantillons indiquent à la fois le nombre de groupes à étudier et le nombre d'individus au sein de chaque groupe. Il est important de garder le nombre d'individus par groupe aussi similaire que possible. Par exemple, si les écoles de la zone ciblée sont petites et n'ont pas toujours 15 élèves répondant aux critères de sélection, n'utilisez pas 15 individus/groupe dans les équations de taille de l'échantillon.[4]

[2] Il convient de noter que lorsque l'on utilise un échantillon pour déclarer les valeurs d'indicateurs standards tels que « la valeur des ventes, le nombre d'hectares faisant l'objet de pratiques de gestion améliorées, le nombre de personnes ayant recours à des pratiques de gestion améliorées, etc. », ces indicateurs doivent être extrapolés à partir de la valeur moyenne d'un échantillon. Par conséquent, calculez la taille d'échantillon nécessaire pour la valeur moyenne de ces indicateurs.

[3] Le regroupement, tout en réduisant les coûts de collecte de données, complique l'analyse. Voir la section 6.1.2 pour plus de détails.

[4] Les modèles d'évaluation utilisés par les praticiens du développement utilisent fréquemment des méthodes d'échantillonnage de probabilité proportionnelle à la taille (PPT), afin d'augmenter la probabilité que des groupes avec des populations plus grandes soient incluses dans l'échantillon. Veuillez consulter la section ci-dessous sur l'analyse des données pour connaître les limites de la méthodologie PPT.

2. Les huit équations principales

Les huit équations relatives à la taille de l'échantillon sont expliquées à la section 2 et supposent qu'un échantillon représentatif est choisi au hasard[5] dans la population.

2.1 Groupes d'étude multiples

Les huit équations fournissent la taille d'échantillon nécessaire pour chaque groupe comparé.

Avant de se plonger dans les équations, il est important de noter que les huit équations fournissent la taille d'échantillon nécessaire pour chaque groupe de comparaison. Des exemples de groupes de comparaison sont le groupe de référence et le groupe final, ou le traitement et le contrôle.[6] Si l'on recueille des données au départ pour les comparer aux données d'évaluation finale, le nombre final fourni doit être appliqué dans chaque cas. Ainsi, si la taille de l'échantillon est de 100, 100 observations sont nécessaires au départ et 100 au final. Soit 100 observations dans le groupe de traitement et 100 dans le groupe témoin. De plus, si l'on analyse les résultats entre les strates[7], par exemple en testant les différences statistiques entre les régions géographiques ou le sexe, le nombre final fourni doit être appliqué à chaque strate.

2.2 Les équations

Cette section guidera les utilisateurs à travers les équations (1-8), expliquera chacun des termes de chaque équation et fournira un exemple pratique pour chaque équation. Notez que l'annexe 2 est une feuille de calcul Excel avec les équations (1-8) déjà programmées, pour aider aux calculs.

2.2.1 Comparaisons entre les groupes

La référence pour les quatre premières équations est McConnell et Vera-Hernández (2015). Cette référence est un bon choix pour le contexte CRS parce que 1) elle fournit les équations exactes nécessaires, plutôt que de faire référence à des commandes empaquetées dans un logiciel statistique ; 2) Il dérive mathématiquement les équations de taille de l'échantillon et/ou fournit des références pour les dérivations ; 3) Il fournit des équations d'effet minimum détectable (EMD) pour les indicateurs binaires et continus.

Les équations (1 et 2) sont les mêmes que celles utilisées dans le guide de Feed the Future à l'intention des évaluateurs tiers (Stukel 2018a), alors que les équations (3-4) diffèrent quelque peu. Voir l'annexe 6 pour une comparaison plus approfondie de ce guide avec le guide de Feed the Future.

Notez que si vous faites des comparaisons entre les groupes :

- Pour un indicateur continu recueilli à partir d'un échantillon non groupé, utilisez l'équation (1)

[5] Les échantillons non aléatoires et les biais qui y sont associés sont abordés à la section 6.1.1.

[6] Il est à noter qu'une version ultérieure de ce guide pourrait faire référence à des équations différentes, probablement plus efficaces, à utiliser pour déterminer la taille totale de l'échantillon et la répartition entre 2 groupes de traitement ou plus (Duflo, Glennerster et Kremer, 2007).

[7] Voir la section 4.3 lors de la stratification de l'échantillon à des fins organisationnelles (et non pour analyser les différences entre les strates). Dans ce cas, il n'est pas nécessaire d'augmenter la taille de l'échantillon.

- Pour un indicateur continu d'un échantillon groupé, utilisez l'équation (2)
- Pour un indicateur binaire d'un échantillon non groupé, utilisez l'équation (3)
- Pour un indicateur binaire 'un échantillon groupé, utilisez l'équation (4)

Reportez-vous à la section 1 si vous n'êtes pas clair sur les termes « continu », « binaire » ou « groupe ».

2.2.1.1 Équation (1)

Utilisez l'équation (1) lorsque vous effectuez des comparaisons entre des groupes pour un indicateur continu recueilli à partir d'un échantillon non groupé.

Lorsque vous effectuez des comparaisons entre des groupes pour **un indicateur continu recueilli à partir d'un échantillon non groupé**, utilisez l'équation (1) :

$$n^* = \frac{2\left(t_\beta + t_{\alpha/2}\right)^2 SD^2}{\delta^2} \qquad (1)$$

Où

- t_β (bêta) est la valeur critique de la queue gauche de la distribution t inverse[8] avec (n^*-1) degrés de liberté.[9] En règle générale, la valeur critique de t_β choisie est de 80 % et représente la puissance de l'échantillon. Ainsi, il y a une probabilité de 20 % de ne pas trouver de différence par rapport à l'intervention, bien qu'il y en ait une (également connue sous le nom d'erreur de type II). La valeur critique nécessaire peut être obtenue à partir d'une table t appropriée ou à l'aide de la commande Excel LOI.STUDENT.INVERSE (telle qu'utilisée dans l'annexe 2).[10]
- $t_{\alpha/2}$ (alpha) est la valeur critique à deux extrémités de la loi t inverse. En règle générale, la valeur t_α choisie est 5 % et représente le niveau de signification. Avec cette valeur, il y a 5 % de chances de rejeter l'hypothèse nulle (généralement sans différence entre les périodes et les groupes de comparaison), alors qu'elle ne devrait pas être rejetée (également connue sous le nom d'erreur de type I). La valeur critique nécessaire peut être obtenue à partir d'une table de T appropriée ou à l'aide de la commande Excel LOI.STUDENT.INVERSE.N[11] (telle qu'utilisée dans l'annexe 2).
- SD est l'écart-type de l'indicateur. Si vous ne le savez pas, veuillez consulter l'annexe 1.4 pour obtenir une liste des valeurs possibles, ainsi qu'un guide sur la façon de calculer l'écart-type par rapport aux données existantes. Obtenir des données d'autres projets au sein du même pays, du même programme, de la même région ou à l'échelle mondiale ; Il peut également être utile d'effectuer un examen documentaire des écarts-types pour des indicateurs similaires provenant d'études antérieures.
- δ (delta) est le changement ciblé de l'indicateur dû à la programmation. En règle générale, δ est la différence entre les valeurs cibles de référence et de durée de vie du projet pour l'indicateur.

Pour prendre un exemple concret, au stade du démarrage du projet, le projet d'approvisionnement local et régional d'aide alimentaire financé par le Département de l'agriculture des États-Unis

[8] Lorsque vous utilisez un écart-type d'échantillon (et non l'écart-type réel de l'ensemble de la population, qui est inconnu), prenez les valeurs critiques de la distribution t (et non de la distribution normale ou *z*).
[9] Les degrés de liberté sont le nombre d'observations utilisées dans l'analyse. Étant donné *que n** est encore inconnu, passez par quelques itérations, en changeant l'estimation de *n**, jusqu'à ce que le *n** utilisé pour trouver la valeur critique de la distribution t soit égal au *n** recommandé par l'équation. Un exemple de ce processus itératif est présenté dans la feuille de calcul de l'Appendix 2.
[10] Si votre interface Excel est en anglais, c'est T.INV.
[11] Si votre interface Excel est en anglais, c'est T.INV.2T.

(USDA) au Honduras avait besoin d'estimer la taille de l'échantillon requis pour l'indicateur « Augmentation moyenne du rendement pour les participants au projet » cultivant des carottes. Grâce à des recherches auprès de sources externes, le personnel du programme a estimé le rendement moyen des carottes à 20 324 kg/ha, avec un écart-type de 7 848 kg/ha (Lana 2012). Le personnel du programme a ensuite examiné une gamme de valeurs cibles du projet pour voir comment la taille des échantillons résultants variait. Le projet s'attendait à une forte augmentation des rendements grâce à l'amélioration des techniques, et a donc estimé qu'il n'aurait besoin d'enquêter que sur 21 agriculteurs.

$$n^* = \frac{2(0{,}88 + 2{,}26)^2 * 7\,848^2}{(20\,324 * 0{,}35)^2}$$

où $t_\beta = 0{,}88$ (à partir de 10 degrés de liberté) ; $t_{\alpha/2} = 2{,}26$; $SD = 7\,848$; et $\delta = (20\,324 * 0{,}35)$. Cependant, pour un changement plus petit, comme une augmentation moyenne de 10 % du rendement, le projet devrait interroger 236 agriculteurs. En bref, pour détecter de plus petits changements à l'aide d'un indicateur continu, des échantillons de plus grande taille sont nécessaires. Il est souvent utile d'établir un tableau, comme c'est le cas dans les feuilles de calcul de l'annexe 2, afin d'examiner les différentes options. Le tableau 1 en est un exemple.

TABLEAU 1. EXEMPLE DE RENDEMENT DE LA CAROTTE AU HONDURAS

CHANGEMENT (δ)	35%	30%	25%	20%	15%	10%
TAILLE TOTALE DE L'ECHANTILLON (n^*)	22	29	40	61	106	237

Avec le tableau 1 en main, le personnel SERA peut alors aider à identifier des objectifs réalisables, pour lesquels des changements statistiques par rapport à la valeur de référence peuvent être détectés, dans le respect des contraintes budgétaires.

2.2.1.2 Équation (2)

Utilisez l'équation (2) lorsque vous effectuez des comparaisons entre des groupes pour un indicateur continu recueilli à partir d'un échantillon groupé.

Lorsque vous effectuez des comparaisons entre des groupes pour **un indicateur continu recueilli à partir d'un échantillon groupé**, utilisez l'équation (2), qui ajoute le terme $(1 + (m - 1)\rho)$ à l'équation (1). Ce terme augmente la taille de l'échantillon pour compenser les effets du regroupement.

$$n^* = m^* k^* = \frac{2\left(t_\beta + t_{\alpha/2}\right)^2 SD^2}{\delta^2}(1 + (m - 1)\rho) \qquad (2)$$

où

- m est le nombre de personnes échantillonnées dans chaque groupe.
- k est le nombre de groupes échantillonnés.
- ρ est la coefficient de corrélation intra-groupe (CCI) anticipée de la valeur de référence du projet. L'CCI est une mesure de la part de la variabilité de l'indicateur qui est due aux différences entre les groupes par rapport aux individus au sein des groupes. Il est expliqué plus en détail à l'annexe 1, y compris la façon dont il est corrélé avec l'effet du plan et une liste de valeurs CCI potentielles.
- t_β, $t_{\alpha/2}$, SD et δ sont définis comme ci-dessus.

Pour simplifier, l'équation (2) est présentée comme suit. Cependant, pour faciliter le calcul dans les exemples de l'annexe 2, m est résolu en fonction de k. Ainsi, les utilisateurs saisiront k (groupes) et calculeront m (individus). Notez que, dans le cas groupé, la distribution-t a 2*(k-1) degrés de liberté.

Il est utile de placer les différentes options dans un tableau et de jouer avec différents nombres de groupes pour déterminer la meilleure taille d'échantillon compte tenu des contraintes de ressources.

Membre du SILC au Burkina Faso à côté du coffre-fort du groupe. [Sam Phelps]

Dans un autre exemple, les projets McGovern-Dole Food for Education financés par l'USDA en Sierra Leone et au Burkina Faso ont voulu voir comment les dépenses d'éducation différaient entre les ménages qui étaient ou n'étaient pas membres des communautés d'épargne et de prêt interne (SILC). À l'aide d'un examen documentaire, le projet a supposé $\rho = 0{,}28$, comme l'a montré une étude ougandaise similaire sur les caractéristiques des ménages par rapport à la communauté qui contribuent à l'épargne (Chowa, Ansong, and R. Despard 2014). En utilisant les valeurs moyennes d'une étude SILC en Zambie, l'équipe a noté que les membres du SILC par rapport aux non-membres du SILC ont dépensé 152 % de plus en frais d'éducation que la moyenne de 10,47 USD (Noggle 2017). L'écart-type des données zambiennes était élevé, à 24,87 USD. Le projet a estimé que le budget pourrait permettre de détecter un changement inférieur à 152 %, au cas où une différence aussi importante n'aurait pas été observée dans l'étude Sierra Leone/Burkina Faso. Ils ont décidé d'interroger 7 membres du SILC dans chacun des 27 SILC, ce qui permettrait une augmentation de 115 % (11,95 USD) par rapport à la moyenne du groupe témoin de l'étude zambienne.

$$n^* = \frac{2(0{,}88 + 2{,}26)^2 * 24{,}87^2}{11{,}95^2}(1 + (7 - 1)0{,}28)$$

ou $t_\beta = 0{,}88$; $t_{\alpha/2} = 2{,}26$; $SD = 24{,}87$; $\delta = 11{,}95$; $k = 7$; et $\rho = 0{,}28$. Le tableau 2 en est un exemple.

TABLEAU 2. EXEMPLE DE SIERRA LEONE / BURKINA SILC

CHANGEMENT (δ)	15,91	14,13	13,09	12,04	10,47
GROUPES (k)	27	27	27	27	27
INDIVIDUS (m)	2	3	4	7	41
TAILLE TOTALE DE L'ECHANTILLON (n^*)	54	81	107	162	1 080

Comme le montre le tableau 2, si l'on maintient constant le nombre de groupes à 27 et l'écart-type à 24,87, l'équipe de projet choisit le compromis entre la taille des échantillons de changement (δ) vs. la taille individuelle (m) et, par extension, le budget. Vingt-sept groupes, avec 2 individus chacune, est la plus petite taille globale de l'échantillon, mais limite la taille du changement que l'équipe peut détecter.

2.2.1.3 Équation (3)

Utilisez l'équation (3) lorsque vous effectuez des comparaisons entre des groupes pour un indicateur binaire recueilli à partir d'un échantillon non groupé.

Lorsque vous effectuez des comparaisons entre des groupes pour **un indicateur binaire recueilli à partir d'un échantillon non groupé**, utilisez l'équation (3) :

$$n^* = \left(p_1(1-p_1) + p_0(1-p_0)\right)\frac{\left(z_\beta + z_{\alpha/2}\right)^2}{\delta^2} \qquad (3)$$

où

- p_1 est le cible du projet pour l'indicateur
- p_0 est la valeur de référence
- z_β (bêta) est la valeur critique unilatérale de la loi normale inverse. En règle générale, la valeur critique choisie de z_β est 80 % et représente la puissance de l'échantillon. La valeur critique nécessaire peut être obtenue à partir d'une table de Z appropriée, ou à l'aide de la commande Excel LOI.NORMALE.INVERSE.N[12] (telle qu'elle est utilisée dans l'annexe 2).
- $z_{\alpha/2}$ est la valeur critique bilatérale de la loi normale inverse. En règle générale, la valeur z_α choisie est 5 % et représente le niveau de signification. La valeur critique nécessaire peut être obtenue à partir d'une table de Z appropriée, ou à l'aide de la commande Excel LOI.NORMALE.INVERSE.N (telle qu'elle est utilisée dans l'annexe 2).
- δ (delta) est le changement ciblé de l'indicateur et est $p_1 - p_0$.[13]

Pour le démontrer, le projet international McGovern-Dole Food for Education, financé par l'USDA, au Burkina Faso, a voulu sonder les mères pour mesurer l'indicateur « Pourcentage de participants aux interventions nutritionnelles au niveau communautaire qui pratiquent des comportements d'alimentation promue du nourrisson et du jeune enfant (ANJE) ». Ils s'attendaient à ce que la valeur de référence soit de 40 % et ont fixé l'objectif final à 55 %. L'équipe a conclu qu'elle interrogerait 171 mères d'enfants de moins de 2 ans.

$$\mathrm{n}^* = \left(0{,}55(1-0{,}55) + 0{,}40(1-0{,}40)\right)\frac{(0{,}84+1{,}96)^2}{0{,}15^2}$$

où $p_1 = 0{,}55$; $p_0 = 0{,}40$; $z_\beta = 0{,}84$; $z_{\alpha/2} = 1{,}96$; et $\delta = 0{,}15$.

2.2.1.4 Équation (4)

Utilisez l'équation (4) lorsque vous effectuez des comparaisons entre des groupes pour un indicateur binaire recueilli à partir d'un échantillon groupé.

Lorsque vous effectuez des comparaisons entre des groupes pour **un indicateur binaire recueilli à partir d'un échantillon groupé**, utilisez l'équation (4) qui ajoute le terme $(1 + (m-1)\rho)$ à l'équation (3). Ce terme augmente la taille de l'échantillon pour compenser les effets du regroupement.

$$n^* = m^*k^* = \left(p_1(1-p_1) + p_0(1-p_0)\right)\frac{\left(z_\beta + z_{\alpha/2}\right)^2}{\delta^2}(1+(m-1)\rho) \qquad (4)$$

où tous les paramètres sont définis comme ci-dessus.

[12] Si votre interface Excel est en anglais, il c'est NORM.INV.

[13] Il convient de noter que, dans le cas binaire des groupes de traitement et de contrôle, le seul moment où une répartition égale et 50/50 entre les groupes de traitement et de contrôle est optimale est lorsque $p_0 = 1 - p_1$, par exemple lorsque la valeur de référence est prévue à 40 % et que la valeur d'évaluation finale prévue dans le groupe de traitement (valeur cible) est de 60 %. Reportez-vous à la Section 4.4 pour savoir comment déterminer la division optimale lorsque $p_0 \neq 1 - p_1$.

Pour reprendre un autre exemple de l'équipe de l'USDA du Burkina Faso, ils utilisaient une évaluation de la performance pour déterminer le « nombre d'enseignants/ éducateurs/ assistants d'enseignement qui démontrent l'utilisation de techniques ou d'outils d'enseignement nouveaux et de qualité ». Pour calculer la taille de l'échantillon nécessaire à la ligne de référence, l'équipe a utilisé les données d'évaluation finale d'une phase précédente pour calculer l'CCI pertinent de 0,44. Dans leur IPTT, ils ont estimé la valeur de référence à 49 %. Leur cible pour l'indicateur était de 65 %. L'équipe a conclu qu'elle avait besoin d'un échantillon de 2 enseignants dans chacune des 105 écoles pour détecter un changement statistique entre la valeur de référence et la valeur finale.

$$n^* = \left(0{,}65(1-0{,}65)+0{,}49(1-0{,}49)\right) * \frac{(0{,}84+1{,}96)^2}{0{,}16^2}(1+(2-1)0{,}44)$$

où $p_1 = 0{,}65$; $p_0 = 0{,}49$; $z_\beta = 0{,}84$; $z_{\alpha/2} = 1{,}96$; $\delta = 0{,}16$, $m = 2$; et $\rho = 0{,}44$. Encore une fois, en particulier avec les conceptions en groupes, il est utile de créer un tableau pour trouver la meilleure combinaison de groupes et le nombre de personnes par groupe pour s'adapter aux contraintes de ressources, comme le fait le tableau 3 :

TABLE 3. EXEMPLE D'ENSEIGNANT AU BURKINA FASO

CHANGER (δ)	16%	16%	16%	16%	16%
GROUPES (k)	146	105	92	85	81
INDIVIDUS (m)	1	2	3	4	5
TAILLE TOTALE DE L'ECHANTILLON (n^*)	147	211	276	340	405

Comme le montre le tableau 3, si l'on maintient constante la variation souhaitée de 16 %, l'équipe de projet choisit le compromis entre la taille de l'échantillon de groupes (k) et celle de l'échantillon individuel (m). Cent quarante-six groupes, avec 1 individu dans chacune, est la plus petite taille globale de l'échantillon. Cependant, le fait de se rendre dans 105 écoles correspondait mieux à d'autres indicateurs qui nécessitaient des visites dans ces mêmes écoles, ce qui réduisait les coûts globaux de collecte de données.

2.2.2 Groupes uniques

Si une équipe de projet est certaine qu'elle n'a besoin d'évaluer qu'un contexte pour un seul groupe et qu'elle ne prévoit pas utiliser les données pour détecter des changements au fil du temps, les équations (5 à 8) doivent être utilisées. Cela pourrait être approprié, par exemple, pour une évaluation des besoins dans laquelle aucune comparaison statistique entre les points temporels, les zones géographiques ou le sexe ne sera faite.

Les équations (5-8) sont très similaires aux équations (1-4), mais lorsqu'il n'y a pas de changement entre les groupes témoins, le terme δ n'est plus pertinent. Les équations n'incluent pas de terme β, car il n'y a plus de problème avec les erreurs de type II (ne pas trouver de différence, bien qu'il y en ait une).

Les équations (5-8) suivent de très près celles de Sullivan (2019), et sont les mêmes que celles utilisées dans le guide Feed the Future à l'intention des partenaires de mise en œuvre (Stukel 2018b). Voir l'annexe 6 pour une comparaison plus approfondie de ce guide avec le guide Feed the Future.

Si vous recueillez des données qui serviront à analyser les données d'un seul groupe, utilisez les équations (5-8)

- Dans le cas d'un indicateur continu recueilli à partir d'un échantillon non groupé, utiliser l'équation (5)
- Pour un indicateur continu d'un échantillon groupé, utilisez l'équation (6)
- Pour un indicateur binaire d'un échantillon non groupé, utilisez l'équation (7)
- Pour un indicateur binaire d'un échantillon groupé, utilisez l'équation (8)

2.2.2.1 Équations (5 et 6)

Il n'y a plus d'écarts-types de deux échantillons, donc le 2 dans le numérateur des équations (1-2) n'est plus nécessaire. Cependant, pour estimer la moyenne de la population à l'intérieur d'une fourchette utile, les équations comprennent une marge d'erreur (E). Par exemple, si l'on suppose que le poids moyen est de 60 kg et que E est fixé à 30 kg, l'échantillon estimera la fourchette de poids de la population adulte entre 30 et 90 kg, ce qui n'est pas utile. Au lieu de cela, définissez E sur 5 kg pour obtenir une estimation du poids réel de la population adulte de 55 à 65 kg.

Utiliser l'équation (5) pour les évaluations d'un seul groupe d'un indicateur continu recueilli auprès d'un échantillon non groupé. L'équation (6) est la version groupée correspondante.

L'équation (1) des indicateurs continus devient ainsi :

$$n^* = \frac{t_{\alpha/2}^2 SD^2}{E^2} \qquad (5)$$

Multipliez l'équation (5) par $(1 + (m - 1)\rho)$ pour l'utiliser avec des échantillons groupés.[14] Ce terme augmente la taille de l'échantillon pour compenser les effets du regroupement.

$$n^* = m^* k^* = \frac{t_{\alpha/2}^2 SD^2}{E^2}(1 + (m - 1)\rho) \qquad (6)$$

2.2.2.2 Équation (7 et 8)

Dans les équations (3-4) pour les indicateurs binaires, seule la probabilité estimée de l'indicateur dans la population évaluée est nécessaire et il n'y a donc pas de terme p_1 et p_0. Il convient de noter que, comme il est décrit à l'annexe 3, la variance avec un indicateur binaire est plus grande lorsque $p = 50\ \%$; C'est également à ce moment-là que la taille de l'échantillon pour les équations (7-8) est la plus grande. En l'absence d'autres données, calculer la taille de l'échantillon pour les équations (7-8) avec $p = 50\ \%$.

Utiliser l'équation (7) pour les évaluations d'un seul groupe d'un indicateur binaire recueilli à partir d'un échantillon non groupé. L'équation (8) est la version groupée correspondante.

L'équation (3) devient donc :

$$n^* = \big(p(1 - p)\big)\frac{z_{\alpha/2}^2}{E^2} \qquad (7)$$

Multipliez l'équation (7) par $(1 + (m - 1)\rho)$ pour l'utiliser avec des échantillons en groupes, car ce terme augmente la taille de l'échantillon pour compenser les effets du regroupement.

$$n^* = \big(p(1 - p)\big)\frac{z_{\alpha/2}^2}{E^2}(1 + (m - 1)\rho) \qquad (8)$$

[14]Comme pour l'équation (2) ci-dessus, pour simplifier l'écriture, l'équation (6) est présentée comme telle. Cependant, pour des raisons de simplicité de calcul dans l'annexe 2, m est résolu pour en fonction de k. Dans ce cas, la distribution t a k-1 degrés de liberté.

2.2.3 Résumé.

Avant de finaliser la taille des échantillons nécessaires, les équipes de projet devraient examiner la section 3 afin de déterminer si d'autres critères s'appliquent à leur situation.

La section 2.2 a fourni des équations (1 à 4) pour calculer la taille minimale de l'échantillon nécessaire pour détecter une différence statistique entre deux groupes de comparaison (p. ex., ligne de référence et finale ; traitement et contrôle, etc.). L'équation appropriée dépendait du fait que l'indicateur était a) continu ou binaire et b) recueilli à partir d'un échantillon groupé.

Si les équipes de projet effectuent des évaluations simples et n'ont pas l'intention de détecter des différences statistiques entre les groupes de comparaison (p. ex., sexe, région géographique, contrôle/traitement, ligne de référence/finale, etc.), il convient alors d'utiliser les équations (5 à 8) au lieu des équations (1 à 4). Selon la marge d'erreur ou la probabilité estimée choisie, cela peut se traduire par des tailles d'échantillon calculées plus petites.

Avant de finaliser la taille des échantillons nécessaires, les équipes de projet devraient examiner la section 3 afin de déterminer si d'autres critères s'appliquent à leur situation. Si c'est le cas, ils devront encore augmenter la taille de leur échantillon.

Après avoir finalisé les calculs des sections 2 et 3, si les équipes de projet craignent que la taille de l'échantillon calculée soit trop grande compte tenu des contraintes budgétaires, passez en revue la section 4. Il peut être possible de réduire la taille de l'échantillon sans affecter la taille du changement à détecter.

3. Examen obligatoire : Éléments qui augmenteront la taille de l'échantillon

Les équations (1-8) présentent la taille minimale nécessaire pour détecter un changement statistique au fil du temps. La section 3 donne un bref aperçu des considérations supplémentaires qu'il peut être nécessaire de prendre en compte pour déterminer la taille finale de l'échantillon que les équipes de projet devraient utiliser. Il s'agit des pertes de données, de l'attrition (ensembles de données de panel) et de sous-populations spécifiques.

3.1 Perte de données

CRS recommande de supposer une perte de données de 5 %.

Si les équipes de projet prévoient que certaines données recueillies seront perdues en raison d'une erreur d'outil, d'un recenseur ou d'une saisie de données, elles doivent recueillir des données auprès d'un échantillon plus grand que le minimum calculé. Le CRS recommande de supposer une perte de données de 5 %, mais les équipes de projet devraient consulter leurs collègues de leur programme national sur l'expérience antérieure en matière de perte de données.

3.2 Attrition

Si les équipes de projet prévoient que certaines données recueillies seront perdues en raison de l'impossibilité de suivre des individus ou des groupes de l'étude de référence à la collecte finale (attrition) ;[15] Ensuite, ils devraient collecter des données auprès d'un échantillon plus grand que le minimum calculé. Ce problème est susceptible d'être plus grave lors de la collecte d'un ensemble de données de panel (dans lequel le même individu est suivi au fil du temps) par rapport à une coupe transversale répétée (généralement lorsque le même groupe est suivi au fil du temps, mais que les individus au sein du groupe varient).

Pour les ensembles de données de panel, étant donné que les mêmes personnes sont interrogées au moins deux fois, CRS recommande d'augmenter la taille de l'échantillon de 10 % supplémentaires (au-delà des concessions de perte de données faites ci-dessus), bien qu'il y ait quelques exceptions.[16] Les équipes de projet devraient consulter leurs collègues de leur programme national au sujet de leur expérience antérieure en matière de taux d'attrition.

3.3 Sous-populations propres à un indicateur

Sachez que, pour certains indicateurs, les équipes de projet peuvent ne pas être en mesure d'identifier les personnes ayant les critères nécessaires avant la collecte des données. Par exemple, l'indicateur du Résultat Mondial de CRS « Prévalence des enfants de 6 à 23 mois recevant un régime minimum acceptable (RMA) » ne s'applique qu'aux personnes qui

[15] Le suivi des bénéficiaires au fil du temps (données de panel) peut aider à réduire la taille globale de l'échantillon nécessaire. Voir la section 4.5 pour plus de détails.

[16] L'étude MIRA a recueilli des données de panel sur les mêmes ménages pendant plus de 24 mois. Le taux d'attrition était de 3 % à 5 % ; cela est dû en grande partie au suivi fréquent (mensuel) et à l'utilisation par le SCG d'agents recenseurs intégrés qui résident dans les collectivités visées par l'enquête.

s'occupent d'enfants âgés de 6 à 23 mois. Il se peut que les équipes de projet ne disposent pas d'une liste détaillée de ces soignants avant la collecte des données (en particulier à la ligne de référence) et qu'elles doivent donc suréchantillonner la population cible dans l'espoir d'échantillonner suffisamment de soignants pour atteindre la taille d'échantillon identifiée. Les équipes de projet doivent consulter leurs collègues de leur programme national sur l'expérience antérieure avec des sous-populations spécifiques à un indicateur.

Par exemple, si les équipes de projet savent qu'environ 1 ménage sur 3 dans leurs communautés participantes ont des enfants âgés de 6 à 23 mois, elles pourraient augmenter la taille de l'échantillon de 300 % par rapport à la taille calculée. Toutefois, les recenseurs ne peuvent poser les questions d'enquête relatives à l'indicateur RMA aux ménages qu'une fois qu'ils ont confirmé qu'ils s'occupent un enfant âgé de 6 à 23 mois dans leur ménage.

3.4 Changements importants entre la ligne de référence et la ligne finale

Lors de comparaisons entre groupes, si un changement important est estimé entre les groupes, la taille de l'échantillon calculée peut être trop petite pour que chaque point de données individuel soit significatif.

Dans cet exemple, on estime que 80 % des agriculteurs adopteront une nouvelle technique (une fois qu'elle leur aura été démontrée), car la technique est facile à adopter et constitue une grande amélioration par rapport à la pratique actuelle. D'après les évaluations de la conception du projet, la technique n'est actuellement utilisée que par 10 % des agriculteurs. En raison de ce changement grande, seuls cinq agriculteurs doivent être échantillonnés de référence et à la fin de l'enquête pour détecter un changement de 70 % au fil du temps, à l'aide de l'équation (3). Cependant, en utilisant l'équation (7) et en ajustant la marge d'erreur jusqu'à ce que seulement 5 agriculteurs soient échantillonnés, nous voyons que la valeur de base aura une marge d'erreur de 27%. Ainsi, si la moyenne de l'échantillon de référence est de 10 %, nous pouvons seulement dire que la véritable valeur de référence se situe entre 0 % et 37 %. Pour cette raison, il est préférable d'utiliser l'équation (7) pour calculer la taille de l'échantillon, en estimant que la valeur de référence est de 10 % et avec une marge d'erreur de 10 %. Cela recommanderait d'échantillonner 35 agriculteurs au départ, et nous pourrions dire que la valeur de référence réelle se situe entre 0 % et 20 %.

Si les équipes de projet utilisent les équations (1-4), il est recommandé de de recouper les estimations ponctuelles avec les équations (5-8).

La taille de l'échantillon final peut être mise à jour ultérieurement en conséquence, à l'aide des paramètres des données de l'étude de référence.

3.5 Résumé

Si, à la suite d'un examen de la section 3, les équipes de projet déterminent qu'elles doivent augmenter la taille de leur échantillon pour tenir compte de la perte de données, de l'attrition ou de sous-populations particulières, cela peut être fait en augmentant le nombre d'individus ou de groupes échantillonnés. Le choix dépend de la probabilité anticipée qu'un groupe cesse de fonctionner pendant la durée du projet (il est peu probable que les écoles publiques cessent, mais le SILC ou les associations de producteurs peuvent le faire) par rapport aux individus qui quittent ce groupe. D'autres groupes ou individus « de secours » devraient être choisis au hasard de la même manière que les autres répondants au sondage.

4. Autres considérations

Cette section présente des concepts plus avancés et des outils de référence disponibles en dehors de ce guide, sur lesquels les lecteurs peuvent avoir besoin de plus d'informations. Cette section se concentre sur 1) le facteur de correction de la population finie ; 2) la relation entre les valeurs de référence des indicateurs, les cibles et la taille des échantillons ; 3) la stratification des échantillons ; 4) les indicateurs binaires dans les évaluations d'impact ; et 5) les ensembles de données de panel et les calculs de la taille de l'échantillon.

4.1 Le facteur de correction pour population finie (FPC)

Si vous craignez que la taille de l'échantillon calculée soit trop grande pour les contraintes budgétaires, vérifiez si elle peut être réduite par le facteur de correction pour population finie. Le FPC est utilisé lorsque la taille de l'échantillon est grande par rapport à la taille de la population. Utilisez le FPC si l'échantillon calculé représente plus de 5 % de la population pour laquelle l'indicateur est recueilli (qu'il soit continu ou binaire). Le FPC théorique est $1 - \left(\frac{n}{N}\right)$, bien qu'il soit parfois écrit sous la forme $\frac{(N-n)}{N}$, où *n* est la taille de l'échantillon calculée et *N* est la taille de la population. Le FPC est ensuite multiplié par l'intervalle de confiance respectif à partir duquel les équations de taille d'échantillon sont dérivées (les intervalles de confiance sont discutés plus en détail à la section 6.2.6). Nous ne montrons pas l'algèbre ici, mais l'équation (9) est utilisée pour ajuster la taille initiale de l'échantillon par le FPC ; notez qu'il suit Thompson (2012):

$$n_{\text{FPC}} = \frac{1}{\frac{1}{n^*} + \frac{1}{N}} \qquad (9)$$

*n** étant la taille initiale de l'échantillon calculée, et *N* défini comme ci-dessus.

Il n'est pas recommandé d'appliquer le FPC si l'on suppose une validité externe, par exemple dans le cadre d'une évaluation d'impact ou d'activités de recherche.

Le FPC doit être appliqué à la taille de l'échantillon calculée avant d'effectuer les ajustements pour tenir compte de la perte de données, de l'attrition ou de sous-populations particulières décrites à la section 3.

Il convient de noter que, dans la littérature économique, le FPC est généralement ignoré parce que les chercheurs supposent que la taille de l'échantillon est petite par rapport à l'ensemble de la population (Cameron and Trivedi 2005). Cela serait particulièrement vrai lorsqu'il s'agit de supposer une validité externe (généralisable au-delà d'un projet spécifique) avec des activités d'évaluation d'impact ou de recherche. Cela se produit généralement dans le cas d'un plan expérimental dans lequel un ou plusieurs groupes de traitement sont comparés à un groupe témoin. Dans de tels cas, il n'est pas recommandé d'appliquer le FPC.[17]

4.2 Fixer des objectifs réalisables et être conscient des coûts de collecte de données qui y sont associés

Il est important de comprendre les implications pour les budgets SERA à l'étape de la conception du projet. Les données relatives aux indicateurs standards des donateurs doivent être collectées et communiquées, et des ressources adéquates doivent être allouées pour détecter un changement dans ces indicateurs au fil du temps.

[17] Si l'on rapporte des intervalles de confiance autour d'une moyenne d'échantillon, et si les critères du FPC s'appliquent, il faut également utiliser le FPC pour ajuster l'intervalle de confiance. Voir la section 6.2.7.

Cependant, les indicateurs personnalisés sont quelque peu à la discrétion des équipes de projet. Par exemple, lors de la conception d'un récent projet éducatif au Togo, l'équipe a voulu collecter des indices de masse corporelle (IMC) pour les enfants d'âge scolaire. Il est difficile de trouver ces données pour n'importe quel pays, et l'équipe de conception du projet a estimé qu'elle pourrait voir un petit changement (3 à 5 %) dans cet indicateur au cours du projet de 5 ans, mais n'en était pas sûre. Compte tenu du très faible changement pour un indicateur continu, la taille de l'échantillon requise était 50 % plus grande que celle de tout autre indicateur de projet.

L'équipe de conception du projet a donc décidé de collecter les données par le biais d'une étude spéciale, tandis que les équipes SERA recueillaient d'autres données auprès des élèves, en utilisant la taille d'échantillon plus petite recommandée pour d'autres indicateurs. Bien qu'il ne soit peut-être pas possible de détecter un changement statistique au cours de la vie du projet, les données devraient s'avérer utiles pour éclairer les futurs projets au Togo et potentiellement d'autres projets d'éducation CRS dans le monde. Si, de manière inattendue, un changement plus grand que prévu est réalisé, l'équipe de projet sera en mesure de détecter une différence statistique entre la ligne de référence et la ligne finale. Ainsi, l'équipe de projet a supprimé cet indicateur de l'IPTT du projet, ce qui présentait l'avantage supplémentaire de ne pas engager l'équipe de projet dans un changement ciblé qui serait très coûteux à détecter si elle pouvait le faire.

4.3 Stratification

Si les équipes de projet souhaitent effectuer des comparaisons statistiques entre les strates, la taille de l'échantillon doit être augmentée pour en tenir compte.

La stratification consiste simplement à organiser ou à classer les données en groupes. Cela a été mentionné ci-dessus, lorsque nous avons discuté du traitement et des groupes témoins, du sexe ou de la séparation géographique. Lors du calcul de la taille des échantillons, il est important de penser à la stratification que les équipes de projet feront avec les données pour deux raisons :

1) **Comparaisons statistiques entre strates.** Si les équipes de projet souhaitent effectuer des comparaisons statistiques entre les strates, la taille de l'échantillon doit être augmentée pour en tenir compte. Comme il est décrit à la section 2.1, cela se fait généralement en multipliant la taille finale de l'échantillon par le nombre de strates à analyser. Par exemple, si vous testez les résultats scolaires des garçons par rapport à ceux des filles (2 strates) et que la taille de l'échantillon est de 5 enfants dans chacune des 50 écoles, parlez à 5 garçons et 5 filles dans chaque école.

2) **Pas besoin de comparaisons statistiques.** Si les équipes de projet ont besoin de stratifier les données, mais ne veulent pas faire de comparaisons statistiques entre les strates (cela est souvent fait pour la désagrégation des données selon le sexe), la taille de l'échantillon n'a pas besoin d'augmenter. Par exemple, si l'équipe souhaite connaître les rendements des producteurs masculins et féminins, elle peut sélectionner au hasard des producteurs masculins dans une liste, et sélectionner séparément les productrices, afin d'assurer une représentation égale des deux sexes dans l'échantillon, sans augmenter la taille globale de l'échantillon.

Évitez d'avoir à utiliser des poids d'échantillon lors de la stratification d'un échantillon en utilisant l'échantillonnage systématique à intervalles fractionnaires.

Les poids de l'échantillon, décrits à la section 6.2.1, doivent tenir compte de toute stratification de l'échantillon. Cela est vrai même lorsqu'il n'y a pas de comparaison statistique entre les strates.

- L'utilisation d'un échantillonnage systématique par intervalles fractionnés est un moyen potentiel de répondre à la nécessité de disposer de poids. On trouvera de plus amples renseignements sur cette stratégie d'échantillonnage à la section 9.4.2 du document Stukel (2018b).

Lors de comparaisons statistiques entre les strates et le regroupement de l'échantillon, il est possible d'augmenter le nombre de groupes ou d'individus au sein de chaque groupe. Pour reprendre l'exemple du point 1) ci-dessus, si vous testez les différences entre les régions

géographiques (2 strates), parlez à 5 enfants dans chacune des 50 écoles de la région A et faites de même dans la région B. Si vous testez les différences selon la région géographique et le sexe (4 strates), parlez à 5 garçons et 5 filles dans chacune des 50 écoles de la région A, et faites de même dans la région B.

4.4 Utilisation d'un traitement inégal et d'une taille de groupe témoin avec des indicateurs binaires

Lors de l'utilisation d'une évaluation d'impact pour mesurer un indicateur binaire[18] à partir d'un échantillon groupé, il est plus efficace de tenir compte d'un nombre inégal de groupes dans le groupe témoin par rapport au groupe de traitement. Avec les indicateurs binaires, le seul moment où une répartition 50/50 entre le groupe de traitement et le groupe témoin est optimale est lorsque $p_0 = 1 - p_1$, par exemple lorsque la valeur de référence est anticipée à 40 % et que la valeur finale du groupe de traitement est anticipée à 60 %.

Bien que la répartition puisse être laissée à 50/50, si l'on effectue une évaluation d'impact (ou une recherche) avec des comparaisons entre le groupe de traitement et le groupe témoin, il peut être plus rentable d'avoir une répartition inégale. Cela peut être particulièrement vrai s'il y a un coût élevé pour le traitement lui-même ; Le groupe de traitement peut être le plus petit des deux groupes.

Compte tenu de la rareté de ce type d'événement dans le contexte de CRS, le présent guide recommande simplement aux équipes de projet de se reporter aux équations 17, 18 et 21 (et à la feuille de calcul Excel qui les accompagne) dans McConnell et Vera-Hernández (2015) pour calculer la répartition optimale entre le groupe de traitement et le groupe témoin.

4.5 Utiliser des ensembles de données de panel

Pour calculer la taille des échantillons qui tient compte de la puissance statistique supplémentaire fournie par les ensembles de données de panel, voir McConnell et Vera-Hernández (2015).

Si elles suivent la même personne au fil du temps, les équipes de projet peuvent gagner en puissance statistique en utilisant la valeur de référence de chacun lors de l'analyse des données d'évaluation finales. Le calcul de la taille des échantillons qui tient explicitement compte de la puissance statistique supplémentaire fournie par les ensembles de données de panel nécessite que les équipes de projet disposent d'informations supplémentaires pour les équations de taille de l'échantillon, qui peuvent ne pas être disponibles. Les informations supplémentaires sont les suivantes :

- Pour les indicateurs continus, l'CCI à la fois au niveau individuel et au niveau du groupe. Si vous êtes intéressé par cette approche, reportez-vous à l'équation (16) de McConnell et Vera-Hernández (2015).
- Pour les indicateurs binaires, la valeur de référence de l'indicateur, ou toute autre covariable, peut réduire la taille de l'échantillon calculé. L'utilisateur devra estimer empiriquement l'effet de la covariable sur l'indicateur avant de calculer la taille de l'échantillon, et les équations de la taille de l'échantillon elles-mêmes peuvent être fatigantes pour ceux qui ne sont pas familiers avec l'algèbre linéaire. Si vous êtes intéressé par cette approche, référez-vous aux équations (24-25) de McConnell et Vera-Hernández (2015). L'annexe 3 fournit un exemple de code R pour les équations (24-25).

Au lieu d'utiliser des équations de taille d'échantillon spécifiques aux ensembles de données de panel, utilisez simplement les équations (1 à 4) de ce guide, le cas échéant. Ils ne tiennent pas

[18] Notez qu'avec des indicateurs continus, il est moins efficace d'avoir une répartition inégale entre le groupe de traitement et le groupe témoin. Voir l'équation 3 dans McConnell et Vera-Hernández (2015).

compte de la puissance statistique supplémentaire fournie par les ensembles de données de panel, mais garantiront que la taille de l'échantillon est suffisamment grande.

4.6 Résumé

Cette section comprenait des sujets spéciaux que les utilisateurs avancés, probablement les conseillers techniques, devraient connaître en ce qui concerne les calculs de la taille de l'échantillon. Il comprenait une discussion générale sur le facteur de correction population finie ; la relation entre la taille des échantillons et les cibles du projet ; stratification de l'échantillon ; des indicateurs binaires dans les évaluations d'impact ; et le calcul de la taille des échantillons pour les ensembles de données de panel.

5. Mythes sur la taille de l'échantillon

Dans le domaine du développement international, il existe deux malentendus largement répandus concernant le calcul de la taille de l'échantillon : 1) la taille des échantillons dépend de la taille de la population sous-jacente et 2) le fait que les indicateurs binaires nécessitent des échantillons plus grands que les indicateurs continus. Les deux mythes sont démystifiés dans cette section.

5.1 La taille des échantillons dépend de la taille de la population sous-jacente

Après avoir examiné ce guide, il convient de noter que la taille de la population sous-jacente n'est prise en compte que lorsque le FPC est jugé approprié (section 4.1). Les facteurs influençant le calcul de la taille de l'échantillon sont 1) l'ampleur du changement à détecter (ou marge d'erreur acceptable) et 2) le nombre de groupes à comparer. Ce n'est que lorsque la taille de l'échantillon est supérieure à 5 % de la population sous-jacente que nous devrions envisager de la réduire du facteur FPC.

5.2 Les indicateurs binaires nécessitent des échantillons de plus grande taille

Certaines données peuvent être « raisonnablement » converties de binaires en données continues, et vice-versa. Il existe un mythe souvent énoncé selon lequel l'utilisation de la version continue est préférable, car elle nécessitera une taille d'échantillon plus petite (Frost 2020). En comparant les équations de base pour les variables continues et binaires, les équations (1) et (3) respectivement, chacune a des paramètres différents et il n'y a aucune preuve mathématique que l'un d'entre eux aboutirait toujours à une taille d'échantillon plus petite ou plus grande.

$$n^* = \frac{2\left(t_\beta + t_{\alpha/2}\right)^2 SD^2}{\delta^2} \quad (1) \qquad n^* = \left(p_1(1-p_1) + p_0(1-p_0)\right)\frac{\left(z_\beta + z_{\alpha/2}\right)^2}{\delta^2} \quad (3)$$

À titre d'exemple, en utilisant les résultats des examens des élèves des écoles primaires en Sierra Leone, l'indicateur standard requis par les donateurs du projet est le pourcentage d'élèves qui réussissent l'examen de lecture, qui est un indicateur binaire. L'équipe du projet s'attendait à ce que la valeur de référence soit de 41 % et s'est fixé un objectif de 58 %, et a donc voulu mesurer un changement de 17 %. Les étudiants qui répondent correctement à 3 des 5 questions sont considérées comme ayant réussi l'examen. La valeur de référence et la cible anticipés pourraient théoriquement être convertis en score. Si 41 % des élèves ont obtenu au moins 3 points au test (et que tous les autres ont obtenu un score de zéro), le score moyen de référence serait de 3*0,41 = 1,23 par élève ; La cible serait donc de 3*0,58 = 1,74 par élève, et la variation serait de 0,51. En utilisant les données de score de test réelles de cet exemple, l'écart-type est de 1,83. Dans cet exemple, l'équation (1) de la version continue de l'indicateur calcule la plus grande taille d'échantillon.

$$\mathrm{n}^* = \frac{2(0{,}88 + 2{,}26)^2 * 1{,}83^2}{0,51^2} = 205$$

$$n^* = \left(0{,}58(1 - 0{,}58) + 0{,}41(1 - 0{,}41)\right)\frac{(0{,}84 + 1{,}96)^2}{0,17^2} = 132$$

Il est important de se rappeler que les indicateurs continus et binaires ont des distributions de probabilité et donc des variances différentes. Pour cette raison, ils nécessitent des équations différentes et il n'est pas possible de faire des déclarations générales sur les calculs qui en résultent. Les paramètres nécessaires pour chaque équation sont différents parce qu'ils mesurent des choses fondamentalement différentes.

5.3 Résumé

Utilisez toujours l'équation appropriée pour le type d'indicateur, car les tests statistiques finaux effectués à l'étape de l'analyse dépendront du type d'indicateur. En fin de compte, l'objectif des calculs de la taille de l'échantillon est d'avoir un échantillon suffisamment grand pour détecter les changements statistiques à l'étape de l'analyse, et la taille de la population sous-jacente n'est pas un facteur dans ces calculs.

6. Plan d'échantillonnage et analyse

Bien que ce guide se concentre sur les calculs de la taille de l'échantillon, il est important de comprendre comment la méthode choisie pour la sélection de l'échantillon, effectuée avant de déterminer l'équation de taille de l'échantillon à utiliser, influe sur l'analyse. Cette section donne un bref aperçu de l'échantillonnage aléatoire et du biais de sélection de l'échantillon ; de l'échantillonnage proportionnel à la taille de la population ; et l'utilisation de poids d'échantillon au stade de l'analyse.

6.1 Sélection de l'échantillon

Cette section décrit l'importance de l'utilisation d'échantillons aléatoires et les défis liés à l'utilisation de la méthode de probabilité proportionnelle à la taille (PPT).

6.1.1 Utiliser des échantillons aléatoires et documenter tout biais d'échantillonnage dû à l'échantillonnage non aléatoire

S'il manque 5 % ou réponses à une question individuelle, dans la section des limites du rapport d'étude, décrivez toute source de biais attribuable à ces omissions.

Les échantillons représentatifs doivent toujours être choisis au hasard, à partir d'une liste préremplie ou d'un recensement rapide, et la probabilité de sélectionner une personne dans l'échantillon doit être connue.[19] CRS dispose de références internes pour diverses façons de sélectionner un échantillon quantitatif aléatoire (Culligan et al. 2019). Il convient de noter que la sélection aléatoire de l'échantillon est essentielle à l'obtention d'une validité externe ; Il sera difficile de publier à l'externe les résultats d'une évaluation ou d'une recherche provenant d'un échantillon non aléatoire.

Cependant, souvent en raison de contraintes de ressources, un échantillonnage non aléatoire et donc un biais de sélection se produisent. Cela peut être dû à des contraintes de sécurité qui empêchent les équipes d'étude d'atteindre une zone interdite ou lorsque les listes à partir desquelles les individus ou les groupes sont sélectionnés au hasard sont obsolètes, et qu'il s'avérerait trop coûteux ou impossible de localiser ceux sélectionnés au hasard. S'il manque 5 % ou plus de répondants ou de réponses à une question individuelle (Cameron and Trivedi 2005), dans la section Limites du rapport d'évaluation, décrivez le mieux possible les sources de biais attribuables à ces omissions.

Par exemple, si les élèves ne sont pas présents à l'école le jour où ils sont censés faire l'objet d'une enquête, en quoi les élèves absents diffèrent-ils de ceux qui sont présents ? Est-ce qu'un test t des moyennes montre que la proportion de groupes clés (sexe, origine ethnique, région géographique)[20] dans l'échantillon est la même que celle de ceux qui n'ont pas été inclus ? Si ce n'est pas le cas, comment l'échantillon pourrait-il être biaisé ? Sinon, comment les étudiants qui ne présenteraient pas ce jour-là pourraient-ils être différents ? Ne pourraient-ils pas obtenir d'aussi bons résultats aux examens d'alphabétisation, etc., parce qu'ils pourraient souvent manquer l'école ?

[19] Pour cette raison, il n'est pas recommandé d'utiliser une « marche aléatoire » pour la sélection des ménages, à moins que le nombre total de ménages dans une communauté ne soit connu.

[20] L'analyste n'a peut-être pas beaucoup d'informations sur les élèves qui ne sont pas présents. Cependant, en fonction des noms des élèves et de l'emplacement des écoles, ils pourraient au moins avoir cette information.

Autre exemple, que se passe-t-il si l'on peut mesurer le rendement de certains agriculteurs, mais pas d'autres parce qu'ils a) n'ont pas planté la culture faisant l'objet de l'enquête ou b) ont planté la culture mais ne l'ont pas récoltée ?

Imaginez d'autres scénarios dans lesquels cela pourrait se produire et réfléchissez à ce qui peut être dit sur les différences entre ceux qui pourraient ou ne pourraient pas être étudiés.

6.1.2 L'échantillonnage de population proportionnelle à la taille (PPT) n'est peut-être pas appropriée dans le contexte de CRS

Le PPT est une méthode de sélection des groupes d'études. Il est couramment utilisé pour tenir compte de la taille des groupes lors de leur sélection lors de la première étape de la collecte de données, au cours de laquelle chaque individu de chaque groupe a une probabilité égale d'être sélectionné dans l'échantillon. Si, à la deuxième étape, un échantillon aléatoire simple ou systématique est utilisé pour sélectionner tout le monde parmi tous les individus du groupe, alors l'échantillon est « auto pondéré » et aucun poids d'échantillon n'a besoin d'être appliqué à l'étape de l'analyse.

La sélection des groupes par PPT n'est auto pondérée que si une sélection aléatoire simple ou systématique de l'échantillon est utilisée à la deuxième étape.

Les analystes des données recueillies par l'intermédiaire d'un échantillon sélectionné par l'PPT doivent comprendre quatre choses :

1) Si l'échantillon a été stratifié (tel qu'il est décrit à la section 4.3), ou si un échantillon aléatoire simple ou systématique n'a pas été utilisé à la deuxième étape, l'échantillon n'est pas auto pondérée et les poids de l'échantillon doivent être utilisés (voir la section 6.2 pour une discussion sur les poids de l'échantillon).

2) Pour utiliser l'PPT, la mesure de la taille doit être la même que l'unité d'échantillonnage utilisée à l'étape de l'analyse. Il est à noter que différentes unités (ménages, producteurs, soignants, etc.) sont souvent nécessaires pour différents indicateurs dans le cadre d'un même projet, de sorte qu'un échantillon différent devra être tiré pour chaque unité d'échantillonnage. Par exemple, il est incorrect d'utiliser le nombre total de ménages dans un village comme « taille » en PPT, puis d'utiliser les personnes qui s'occupent d'enfants âgés de 6 à 23 mois comme unité d'échantillonnage. Si l'on utilise l'PPT dans cet exemple, il est nécessaire de savoir d'abord combien de personnes s'occupent d'enfants âgés de 6 à 23 mois dans chaque village et de l'utiliser comme mesure de la « taille ».

3) Dans le cas de l'PPT, il faut utiliser les estimateurs de Hansen-Hurwitz ou de Horvitz-Thompson pour estimer la moyenne de l'échantillon. De plus, ces estimateurs doivent être utilisés lors du calcul de la variance dans les modèles de régression (Hansen and Hurwitz 1942; Horvitz and Thompson 1952). Ce point n'est généralement pas abordé dans les autres guides d'échantillonnage.

4) De plus, lors de l'utilisation de l'PPT, la mesure de la taille doit être précise, sinon elle surestime ou sous-estime la variance de l'échantillon, par rapport à une simple sélection aléatoire de groupes (Thomsen, Tesfu, and Binder 1986). Même si les mesures de référence de la taille sont exactes, si l'on utilise une coupe transversale répétée dans les mêmes groupes lors de l'évaluation finale et que la « taille » des groupes change considérablement au fil du temps, le même problème d'estimation erronée de la variance de l'échantillon se produira.

Feed the Future (FtF) recommande l'utilisation de l'PPT dans son guide d'échantillonnage destiné aux responsables de la mise en œuvre. Lors d'une enquête auprès de groupes de producteurs, il est recommandé de sonder chaque producteur du groupe sélectionné ; Cela permettrait d'éviter les problèmes décrits aux points 1) et 2) ci-dessus. En ce qui concerne le point 2), lors de l'exécution d'enquêtes auprès des ménages, il est recommandé d'utiliser une liste actualisée de tous les

participants individuels au projet comme mesure de la taille. Il utilise ensuite la plus grande taille d'échantillon calculée pour tous les indicateurs du projet afin de sonder les participants sélectionnés, tout en reconnaissant qu'il existe un risque de ne pas échantillonner un nombre adéquat de répondants par indicateur individuel (lorsque l'indicateur ne s'applique qu'à une sous-population spécifique) (Stukel 2018a, 2018b).

Le guide de suivi et d'évaluation des activités d'urgence du BHA fournit trois exemples d'utilisation de l'PPT pour sélectionner des groupes avec plusieurs indicateurs. Le ménage est toujours l'unité d'échantillonnage, et un échantillon aléatoire simple ou systématique est utilisé dans le deuxième étape. Les exemples démontrent la complexité de l'utilisation de l'PPT comme méthode d'échantillonnage et devraient être examinés par tout membre du personnel de l'CRS qui souhaite utiliser l'PPT pour la collecte de données à indicateurs multiples (*Technical Guidance for Monitoring, Evaluation, and Reporting for Emergency Activities* 2022).

Compte tenu de la complexité de l'analyse, des préoccupations concernant les mesures de la taille des groupes et de l'augmentation des coûts, le personnel de l'CRS doit utiliser l'PPT avec prudence. Au lieu de l'PPT, les groupes et les individus peuvent être sélectionnés par d'autres formes d'échantillonnage aléatoire et les poids de l'échantillon peuvent être utilisés dans l'analyse. Il convient toutefois de noter que si l'on décide de ne pas utiliser l'PPT et que les analystes effectuent des régressions sur les données collectées, les poids d'échantillonnage peuvent réduire la précision des estimations des coefficients (Lee and Solon 2011). Une discussion plus détaillée sur les poids d'échantillonnage utilisés dans les analyses de régression est présentée à la section 6.2 ci-dessous.

6.1.3 Résumé

Il est important de sélectionner des échantillons aléatoires. Chaque fois que cela n'est pas possible, assurez-vous de documenter au mieux tout biais de sélection d'échantillon. Si vous utilisez la sélection de groupes proportionnelle à la taille de la population (PPT), assurez-vous de comprendre sa complexité et ses limites avant de l'utiliser. Si l'PPT n'est pas compatible, reportez-vous à la section 6.2 sur les poids des échantillons.

6.2 Utilisation de la taille des échantillons dans l'analyse des données

Cette section décrit pourquoi, quand et comment utiliser les poids d'échantillonnage. Il donne également un aperçu des poids de non-réponse et des poids d'échantillon dans les analyses de régression. Il se termine par une section sur la façon de calculer les intervalles de confiance et l'applicabilité du facteur de correction de la population finie à l'étape de l'analyse.

6.2.1 Poids de l'échantillon - calcul

Les poids d'échantillon ne s'appliquent qu'à certains échantillons groupés ou stratifiés.

Les poids d'échantillon ne s'appliquent qu'à certains échantillons groupés ou stratifiés (comme dans le cas de l'échantillonnage aléatoire simple ou systématique, tout le monde a la même probabilité de sélection d'échantillon). Elles s'appliquent aux échantillons stratifiés dont la taille a été augmentée pour permettre l'analyse entre les strates, ou aux échantillons qui ont été stratifiés à des fins organisationnelles (voir la section 6.1.2).

Les poids d'échantillon reflètent le nombre de personnes dans la population qu'un individu représente. Par exemple, si les écoles sont choisies au hasard dans la première étape d'un plan

groupé, et dans la deuxième étape, 10 filles sont choisies au hasard dans la seule classe de 2e année de l'école[21] avec 30 filles dans la classe, chaque fille échantillonnée représente 3 autres filles.

Statistiquement, les poids d'échantillon sont la probabilité inverse d'être sélectionné dans l'échantillon et sont définis comme $w_i = 1/\pi_i$, où π_i est la probabilité que l'individu *i* soit sélectionné dans l'échantillon. Dans l'exemple ci-dessus, π_i pour chaque fille est 10/30, et donc son w_i est 3.

Un point supplémentaire, pour faciliter les calculs. Dans cet exemple, si au cours de la première étape, 50 écoles ont été sélectionnées au hasard parmi 200, alors chaque école représente 4 autres écoles. Étant donné que chaque école a le même poids d'échantillonnage, les poids de l'école peuvent être ignorés dans les calculs.

Si les 50 écoles ont été stratifiées entre deux régions géographiques, ce qui a donné lieu à des pondérations scolaires différentes, leurs pondérations doivent être incluses. Par exemple, la région A compte 110 écoles et la région B en compte 90. Vingt-cinq écoles ont été choisies au hasard dans la région A et 25 dans la région B. Ainsi, les écoles de la région A ont une pondération de 110/25 et les écoles de la région B ont une pondération de 90/25.

6.2.2 Poids de l'échantillon - utilisation

Les poids d'échantillon doivent toujours être utilisés pour fournir des statistiques descriptives univariées pour des indicateurs individuels, tels que les moyennes/proportions, les totaux, les médianes, etc.

Cependant, les résultats des analyses de régression multivariée devraient idéalement présenter des résultats non pondérés et pondérés et, lorsqu'il y a des différences, inclure une discussion des raisons sous-jacentes. Par exemple, les observations d'une école qui compte 90 élèves de deuxième année contre 30 auront 3 fois plus de poids ; S'il y a des effets de projet hétérogènes pour les grandes écoles par rapport aux petites écoles (p. ex., les grandes écoles ont un ratio enseignant/élèves plus élevé ; ce manque d'attention des élèves se traduit par de moins bons résultats scolaires, etc.), les moyennes conditionnelles peuvent être différentes pour les analyses pondérées et non pondérées (Solon, Haider, and Wooldridge 2015).

6.2.3 Moyennes/proportions pondérées et totaux

L'équation pour calculer une moyenne pondérée univariée est la suivante :

$$\bar{y} = \sum_{i}^{I} y_i * w_i \Big/ \sum_{i}^{I} w_i \qquad (10)$$

où *y* est notre indicateur d'intérêt pour l'individu *i* de poids *w*.

À titre d'exemple simple, utilisez un indicateur binaire (réussite = 1 / échec = 0) pour 5 élèves de 2 classes chacun (Table 4). Étant donné que chaque classe a un nombre différent d'élèves, ceux de la classe A (30 élèves) ont un poids de 30/5 = 6 et ceux de la classe B (40 élèves) ont un poids de 40/5 = 8.

[21] Cet exemple fait référence à un indicateur de l'éducation de l'USAID qui spécifie les élèves de 2e année. Ainsi, toutes les classes qui ne sont pas des classes de 2e année seraient exclues de l'enquête et n'auraient pas besoin d'être prises en compte dans les poids de l'échantillon.

TABLE 4. EXEMPLES DE DONNEES PONDEREES

ID d'étudiant	y_i	w_i	$y_i * w_i$
1	1	6	6
2	1	6	6
3	1	6	6
4	1	6	6
5	0	6	0
6	1	8	8
7	0	8	0
8	1	8	8
9	0	8	0
10	0	8	0
SOMME	**6**	**70**	**40**

Élèves de l'école primaire au Lesotho. [Dooshima Tsee]

Ainsi, la proportion pondérée d'élèves ayant réussi l'équation suivante (10) est la suivante :

$$\hat{y}_w = 40/70 = 57\ \%$$

alors que la proportion simple est $\hat{y} = \sum_i^I y_i/n$ = 6/10 = 60 %.[22]

Lorsque vous utilisez un échantillon représentatif pour extrapoler à une population plus large, il suffit de multiplier $\hat{y}_w$ par la population totale. Dans cet exemple, on estime que 0,57*(30+40) = 40 étudiants ont réussi l'examen, comparativement à 0,60*(30+40) = 42 étudiants qui auraient été estimés à l'aide de l'échantillon non pondéré.

6.2.4 Coefficients de pondération en cas de non-réponse

Les pondérations de non-réponse sont utilisées lorsqu'une partie de l'échantillon visé n'a pas répondu à l'enquête. Cette non-réponse peut être attribuable au fait que les participants ont refusé de participer, qu'ils n'ont pas pu être localisés ou que les données n'étaient pas mesurables (p. ex., le rendement d'une culture qui n'a jamais été récoltée).

Les échantillons sont donc pondérés en fonction des caractéristiques observables des répondants et des non-répondants (comme le sexe, l'âge, l'emplacement géographique, etc.), en plus de leur probabilité de sélection, comme il est indiqué dans la section précédente. Par exemple, les répondantes seraient pondérées pour représenter les non-répondantes ou les répondantes plus

[22] En notation statistique, le ^ au-dessus de y est utilisé pour les proportions, tandis que ¯ au-dessus de y est utilisé pour les moyennes.

Utilisez les poids de non-réponse avec prudence. Ne présumez pas que les répondants peuvent remplacer adéquatement les non-répondants.

âgées pour représenter les non-répondants plus âgés. Ce guide ne fournit pas de détails sur la façon de calculer et d'utiliser les pondérations de non-réponse, mais Raab (2009) et le National Research Council (2002) explique en détail l'application pratique.

Veuillez noter que les pondérations de non-réponse doivent être utilisées avec prudence. Toute non-réponse à l'enquête qui n'est pas aléatoire crée un échantillon biaisé, et ce biais doit être documenté (voir la section 6.1.1). Les analystes ne peuvent pas présumer que les répondants réels peuvent remplacer adéquatement les non-répondants.

Pour reprendre l'exemple de la section 6.1.1, pourquoi le rendement ne peut-il être mesuré qu'à partir de certains agriculteurs ? La raison principale est-elle que certains agriculteurs n'ont pas récolté et qu'il n'y avait donc pas de rendement à mesurer ? N'ont-ils pas récolté parce qu'ils ne pouvaient pas irriguer pendant une année de sécheresse ou utiliser des intrants agricoles résistants à la sécheresse ? Dans ce cas, l'utilisation de pondérations de non-réponse pour donner plus de poids aux agriculteurs répondants qui ont récolté (et en supposant qu'ils peuvent représenter les agriculteurs qui n'ont pas récolté) exacerberait le biais de l'échantillon.

6.2.5 Échantillons groupés ou stratifiés et analyse de régression

Lorsqu'il s'agit de déclarer des moyennes conditionnelles pondérées à partir d'analyses de régression, les valeurs pondérées doivent utiliser la contrepartie pondérée appropriée (p. ex., moindres carrés pondérés, maximum de vraisemblance pondéré, etc.).

De plus, étant donné que les observations au sein d'un groupe sont probablement corrélées, les erreurs types de coefficient doivent toujours être regroupées (Cameron and Miller 2015). Les progiciels statistiques ont des fonctions pour cela ; La fonction appropriée varie en fonction de la méthode d'analyse.

Contrôlez toute stratification ou randomisation de l'échantillon dans les analyses de régression en utilisant des variables binaires pour chaque strate ou unité de randomisation (à l'exception d'une pour éviter le piège des variables fictives).

6.2.6 Intervalles de confiance

Le membre de droite de l'équation (11) est également connu sous le nom de marge d'erreur.

À titre de référence, vous trouverez ci-dessous les équations permettant de calculer les intervalles de confiance (IC) autour d'une moyenne d'échantillon. Pour tout échantillon aléatoire, l'équation suivante s'applique :

$$\text{Intervalle de confiance} = \text{moyenne} \pm SE * t_{\alpha/2} \qquad (11)$$

Où *SE* est l'erreur-type de la moyenne de l'échantillon et est définie dans le paragraphe suivant. Notez que pour les indicateurs binaires, doit être utilisé $z_{\alpha/2}$ à la place de $t_{\alpha/2}$, et les deux sont décrits dans la section 1.2. En règle générale, en sciences sociales, lorsque l'on rapporte des intervalles de confiance, α = 0,05. Le membre de droite de l'équation (11) est également connu sous le nom de marge d'erreur.

Il convient de noter que les erreurs-types de la moyenne ou de la proportion de l'échantillon diffèrent selon qu'il s'agit d'indicateurs continus (équation (12)) ou binaires (équation (13)).

$$SE_{\text{Continu}} = \frac{SD}{\sqrt{n}} \qquad (12)$$

Voir l'annexe I si l'équation de l'écart-type (*SD*) est nécessaire. Dans les équations (12) et (13), *n* est la taille finale de l'échantillon après la collecte des données.

$$SE_{\text{Binaire}} = \sqrt{\frac{p(1-p)}{n}} \qquad (13)$$

Dans le cas d'échantillons groupés, l'*SE* doit être multiplié par l'endroit $\sqrt{\frac{1+\rho}{1-\rho}}$ où ρ c'est l'CCI (Bence 1995). Cela doit être fait avant d'insérer l'*SE* dans l'équation (11). Cela est vrai pour les indicateurs continus et binaires.

6.2.7 L'FPC

Si le facteur de correction population finie (FPC) s'applique à l'échantillon (voir la section 4.1), l'*SE* doit également être multiplié par celui-ci. Pour rappel, l'FPC est $1 - \left(\frac{n}{N}\right)$.

6.2.8 Résumé

La section 6.2 explique que les poids d'échantillonnage permettent essentiellement aux observations d'un seul groupe de représenter d'autres personnes qui n'ont pas fait l'objet d'un relevé dans le même groupe ; Les observations provenant de groupes plus grands auront ainsi plus de poids. Les pondérations de l'échantillon doivent être utilisées dans les statistiques sommaires, mais les régressions doivent rapporter les résultats des échantillons pondérés et non pondérés.

La section 6.2 donne également un aperçu des poids de non-réponse et de la façon d'utiliser les échantillons pondérés dans les analyses de régression ; Les deux sous-sections incluent des liens vers d'autres références pour plus d'informations. Il s'est terminé par les équations nécessaires au calcul des intervalles de confiance pour les indicateurs binaires et continus et a noté la modification nécessaire pour les indicateurs recueillis à partir d'échantillons groupés. S'il est utilisé dans les calculs de la taille de l'échantillon, le facteur de correction population finie doit également être utilisé dans le calcul des intervalles de confiance à l'étape de l'analyse.

Annexe 1. Coefficient de corrélation intra-groupe

Le coefficient de corrélation intra-groupe (CCI) ne s'applique qu'aux plans groupés et indique dans quelle mesure la variabilité des données est due aux différences entre les groupes par rapport aux individus au sein des groupes. Pour obtenir une image fidèle de la population, si les individus au sein des groupes se ressemblent, il est préférable d'enregistrer moins d'individus au sein de chaque groupe et un plus grand nombre de groupes. Sinon, la similitude des réponses au sein d'un groupe amplifiera les différences de réponses des individus entre les groupes, ce qui entraînera des écarts-types plus grands (Killip, Mahfoud, and Pearce 2004).

A1.1 Effet de plan vs. CCI

L'effet de plan dépend de l'CCI et du nombre d'individus par groupe.

Chaque fois qu'un effet de plan est rapporté pour une étude, ou lorsqu'un effet de plan est utilisé dans une équation, les auteurs doivent clarifier l'équation sous-jacente utilisée. Dans la plupart des cas, lorsque vous travaillez avec des équations de taille d'échantillon, l'équation sous-jacente est la suivante :

$$\text{Effect de plan} = 1 + ((m - 1)\rho) \qquad (14)$$

où ρ et m sont respectivement l'CCI et le nombre d'individus par groupe. Notez que l'effet de plan dépend des deux variables, donc les équations qui utilisent un effet de plan sans spécifier m excluent un élément d'information critique.

L'CCI, cependant, ne dépend pas du nombre de personnes interrogées, ce qui la rend plus portable d'un modèle d'enquête à l'autre (Stukel 2018a). Voir ci-dessous pour plus d'informations sur la façon de calculer un CCI à partir de données d'enquête. L'appendice 1.4 fournit une liste des valeurs CCI déjà calculées pour certains indicateurs standard destinés aux principaux donateurs des projets CRS.

Ce guide encourage l'auto-calcul des CCI à partir de données récentes pour des indicateurs similaires dans la même zone

Généralement, les valeurs de base de l'CCI sont utilisées, mais l'CCI est susceptible d'augmenter avec le temps. Cela s'explique par le fait que les activités du projet se situent souvent au niveau du groupe (par exemple, la formation des enseignants est dispensée à tous les enseignants de chaque école, à tous les agriculteurs d'un groupe de producteurs, etc.). Ainsi, les résultats individuels deviendront probablement plus corrélés au sein des groupes après avoir bénéficié des interventions du projet. Etant donné que CRS opère dans la plupart de ses programmes nationaux depuis de nombreuses années, il est probable qu'un projet soit le suivi d'un autre ou le suivi d'une autre ONG, potentiellement dans les mêmes groupes. Pour cette raison, ce guide encourage l'auto-calcul des CCIs à partir de données récentes pour des indicateurs similaires dans la même zone d'opération. De plus, Handa et al. (2018) ont montré les différences entre les CCIs pour le même indicateur d'une région géographique à l'autre, ce qui souligne l'importance d'utiliser des CCIs spécifiques à un projet ou à un pays.

A1.2 Calcul de l'CCI

Cette section montre comment calculer un CCI d'échantillon inconditionnel (sans covariables).

L'CCI de l'échantillon est le rapport entre et à l'intérieur de la variance du groupe :

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} \qquad (15)$$

Une valeur CCI de 1 signifierait que toute la variation des données pourrait s'expliquer par des différences entre les groupes, de sorte que les enquêteurs devraient visiter de nombreux groupes, mais un seul individu par groupe. Un CCI proche de 0 signifierait que toute la variation des données pourrait s'expliquer par des différences entre les individus, de sorte que les enquêteurs visiteraient moins de groupes, mais étudieraient tous les individus de chacun.

A1.2.1 Écart général.

Comme ci-dessus, les équations de calcul des CCI diffèrent entre les indicateurs continus et binaires. Notez que la variance d'un indicateur aléatoire discret (continu) est définie comme suit :

$$\sigma_{\text{Continu}} = \frac{\sum(y - \bar{y})^2}{n} \qquad (16)$$

où *y* est une observation individuelle, $\bar{y}$ est la moyenne de toutes les observations, et *n* est le nombre de toutes les observations. La variance d'un indicateur binaire est la suivante :

$$\sigma_{\text{Binaire}} = p(1 - p) \qquad (17)$$

où p est la proportion de la population pour laquelle la condition est vraie. Notez que la variance d'un indicateur binaire est plus grande lorsque l'indicateur est vrai pour exactement 50 % de la population. Pour voir cela, branchez 0,35, 0,50 et 0,70 pour p. Notez que les variances respectives sont de 0,23, 0,25 et 0,21.

A1.2.2 CCI pour les indicateurs continus

Pour **les données continues**, la variance entre les groupes est calculée comme suit :

$$\sigma_b^2 = \frac{\sum_c^C n_c(\bar{y}_c - \bar{y})^2}{C - 1} \qquad (18)$$

où $\bar{y}_c$est la moyenne des données au niveau individuel de l'*c*ième groupe. $\bar{y}$ est la moyenne de toutes les observations (DataCamp 2019).[23] *C* est le nombre total de groupes. Ainsi, le numérateur additionne la variance pour chaque groupe, en la pondérant par le nombre d'observations par groupe. Le dénominateur divise par le nombre de groupes, ce qui donne la moyenne de la variance des groupes.

De même, la variance au sein d'un groupe est calculée comme suit :

$$\sigma_w^2 = \frac{\sum_{i1}^{I1}(y_{i1} - \bar{y}_1)^2 + \cdots + \sum_{ic}^{IC}(y_{ic} - \bar{y}_c)^2}{n - C} \qquad (19)$$

où y_{i1} désigne l'observation de l'individu *i* dans le groupe 1, et y_{ic} désigne l'observation de l'individu *i* dans l'*c*ième groupe. Ainsi, le numérateur additionne la somme des variances au sein de chaque groupe. Le dénominateur divise par le nombre d'observations (réduit par le nombre d'groupes), ce qui donne la moyenne de la variance au sein du groupe.

[23] Pour la variance entre les groupes, utilisez les observations pondérées. Si les observations au sein d'un groupe ont des poids différents (p. ex., l'échantillon a été stratifié selon le sexe), utilisez les observations pondérées. Notez que, si toutes les observations au sein d'un groupe ont le même poids, l'utilisation des poids de l'échantillon n'a aucun effet sur la variance au sein du groupe.

Il est à noter qu'à l'annexe 2, l'onglet « ICC_example » contient un exemple concret d'indicateur continu (résultats en l'alphabétisme - Koinadugu, Sierra Leone), où ρ = 0,92. Ce même exemple se trouve à l'annexe 3 si les équipes de projet préfèrent utiliser le logiciel statistique R.

Pour calculer l'CCI des indicateurs binaires, utilisez un logiciel statistique plus robuste qu'Excel.

A1.2.3 CCI pour les indicateurs binaires

Pour les indicateurs binaires qui suivent la loi de Bernoulli, diverses méthodes ont été proposées (Schochet 2013; Goldstein, Browne, and Rasbash 2002). Pour les indicateurs binaires, il est recommandé d'utiliser un logiciel statistique plus robuste qu'Excel. Le logiciel statistique R dispose d'une fonction intégrée, ICCbin()[24] dans le package « aod » qui calcule les CCI pour les méthodes A à C comme décrit dans Goldstein. La méthode A utilise la distribution logistique, souvent utilisée lors de l'analyse d'indicateurs binaires, et est donc recommandée. L'annexe 3 présente un exemple de code R, en utilisant les mêmes scores d'alphabétisme que pour l'exemple continu, mais en les convertissant en un indicateur binaire (réussite/échec). Dans la version binaire, ρ= 0,60.

A1.3 Calcul de l'écart-type

À titre de référence, l'équation permettant de calculer l'écart-type d'un échantillon (qui n'est utilisée qu'avec des indicateurs continus) est la suivante :

$$SD = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n-1}} \qquad (20)$$

La fonction « ECARTYPE.STANDARD » dans Excel peut également être utilisée.[25]

Si vous calculez l'écart-type à partir d'un échantillon groupé ou stratifié, utilisez l'équation pour un écart-type pondéré (The Statistical Engineering Division 1996):

$$SD_{\text{pondéré}} = \sqrt{\frac{\sum_i^I w_i(y_i - \bar{y}_w)^2}{(n'-1)\frac{\sum_i^I w_i}{n'}}} \qquad (21)$$

où n' est le nombre de poids non nuls. En pratique, tous les poids seront ≥1, donc n'= n.

Sections 6.5.2.2 et 6.5.2.3 dans Higgins, Li et Deeks (2019) fournit des conseils sur le calcul des écarts-types à partir d'articles de revues qui ne rapportent pas d'écarts-types, mais qui rapportent des moyennes, des intervalles de confiance, des erreurs-types et des valeurs de p.

L'annexe 1.4 présente les CCI et les écarts-types pour la plupart des indicateurs types du gouvernement des États-Unis.

A1.4 Valeurs de l'CCI et de l'écart-type pour certains indicateurs

Cette section de l'annexe présente les valeurs de l'CCI et les écarts-types pour certains indicateurs couramment utilisés ou spécifiques aux donateurs dont les données sous-jacentes sont généralement recueillies par l'intermédiaire d'un échantillon représentatif. Il est séparé du guide principal, afin de permettre des mises à jour fréquentes. Veuillez consulter la page couverture de l'annexe 1.4 pour connaître la date de sa plus récente mise à jour. Les lecteurs qui souhaitent ajouter des valeurs CCI aux tableaux de l'annexe 1.4, ou qui souhaitent obtenir de l'aide pour les calculer, doivent contacter l'auteur.

[24] Assurez-vous d'utiliser la fonction ICCbin() du paquet « aod », et non le paquet « ICCbin », car ce dernier ne semble pas fonctionner.

[25] Si votre interface Excel est en anglais, c'est « STDEV.S ».

Annexe 2. Calculateur de taille d'échantillon – Excel

Cette annexe est une feuille de calcul Excel, conservée séparément du guide principal. Il contient des tableaux prêts à l'emploi pour les équations (1 à 8) de ce guide, en plus d'un exemple de calcul d'un CCI continu à partir d'échantillons de données.

Annexe 3. Calculateur de taille d'échantillon – R

Cette annexe contient un code R prêt à l'emploi pour le calcul de l'CCI d'un indicateur binaire, ainsi qu'un indicateur binaire avec une covariable.

```
## Binary Indicator ICC

#############################################################################
#Sets up file and reads in data                                             #
#############################################################################

#Clears all objects from memory
rm(list=ls())

#Sets the working directory
setwd()

library("readxl")
library("aod")

#Reads Excel file and names it
data <- read_xlsx("CRS_Samples_Annex2. Excel.xlsx", sheet = "ICC_Example", range = "A1:B475",
            col_names = T)

#############################################################################
#Binary ICC - using built-in binary methods with no covariates              #
#############################################################################
#Probability of passing - from final evaluation data for 2nd graders

#Create the binary variable from the scores. In this context, a score of 4 or better was
#considered passing
Score_pass = data$LiteracyScore
Score_pass[Score_pass<4] = 0
Score_pass[Score_pass>0] = 1
data = cbind(data,Score_pass)

#To use the ICCbin function, the data must be grouped by numerator and denominator per
#each cluster, and not as individual-level observations
numerator.SchoolPass = aggregate(data$Score_pass,list(data$SchoolCode),sum)
numerator.SchoolPass = numerator.SchoolPass[,'x']
denominator.SchoolPass = aggregate(data$Score_pass,list(data$SchoolCode),length)
denominator.SchoolPass = denominator.SchoolPass[,'x']
data.SchoolPass = as.data.frame(cbind(numerator.SchoolPass,denominator.SchoolPass))

#Now, we can compute the ICC using Method A
Pass_Fail.A = iccbin(n = denominator.SchoolPass, y = numerator.SchoolPass,
              data = data.SchoolPass, method = "A")
#end###
```

```
################################################################################
#Binary sample size with one binary covariate - McConnell Equations 24-25      #
################################################################################

#Calculate sample size with one with one covariate (gender) and treatment group only
#Convert data to binaries
data$PupilsGender[data$PupilsGender=='Boy'] = 0
data$PupilsGender[data$PupilsGender=='Girl'] = 1

#Calculate residual variance
fit = lm(Score_pass ~ PupilsGender,data = data)
resid = as.data.frame(residuals(fit))
data.cov = cbind(resid,data[,'SchoolCode'])
names(data.cov) = c('resid','SchoolCode')

#Calculate total within group variance with gender covariate
y.cov_i_bar = aggregate(data.cov[,'resid'],list(data.cov[,'SchoolCode']),mean)
data.cov = merge(data.cov,y.cov_i_bar,by.x="SchoolCode",by.y = "Group.1")
colnames(data.cov)[ncol(data.cov)] = c('SchoolMeanRes')
data.cov = cbind(data.cov,(data.cov[,'resid']-data.cov[,'SchoolMeanRes'])^2)
colnames(data.cov)[ncol(data.cov)] = c('DevianceMeanRes_Sqrd')
Sum_Var.cov_withinCluster =
aggregate(data.cov$DevianceMeanRes_Sqrd,list(data.cov$SchoolCode),sum)
Total_Var.cov_withinCluster = sum(Sum_Var.cov_withinCluster[,2])/(n-g)

#Calculate between group variance with gender covariate
y.cov_bar = mean(data.cov$resid)
Var.cov_betweenCluster = sum(n_i[,2]*((y.cov_i_bar[,2] - y.cov_bar)^2))/(n-1)

#Calculate rho
rho.gender = Var.cov_betweenCluster/(Var.cov_betweenCluster+Total_Var.cov_withinCluster)

#Calculate sample size
#Ratio of treatment to control
pi = .5

#Normal cdf values
z_alpha = qnorm(0.05/2)
z_beta = qnorm(1-0.8)

#Values analyzed
x.0 = 0
x.1 = 1

#Probability of being a boy or girl in sample
theta.0 = 253/(221+253)
theta.1 = 221/(221+253)

#Probability of passing at baseline for boys (61%)/ girls (56%)
p0.0 = 154/253
p0.1 = 124/221

#Probability of passing at endline for boys (69%)/ girls (69%)
p1.0 = 175/253
```

```
p1.1 = 152/221

#Effect size for each of the covariate's values, where delta_q = p1q - p0q
delta.0 = p1.0 - p0.0
delta.1 = p1.1 - p0.1

#Overall effect size
delta = theta.0*(delta.0) + theta.1*(delta.1)

#M matrix
m1 = (pi*theta.0*p1.0*(1-p1.0) + (1-pi)*theta.0*p0.0*(1-p0.0)) +
  (pi*theta.1*p1.1*(1-p1.1) + (1-pi)*theta.1*p0.1*(1-p0.1))
m2 = pi*theta.0*p1.0*(1-p1.0) +
  pi*theta.1*p1.1*(1-p1.1)
m3 = x.0*(pi*theta.0*p1.0*(1-p1.0) + (1-pi)*theta.0*p0.0*(1-p0.0)) +
  x.1*(pi*theta.1*p1.1*(1-p1.1) + (1-pi)*theta.1*p0.1*(1-p0.1))
m4 = x.0*(pi*theta.0*p1.0*(1-p1.0)) +
  x.1*(pi*theta.1*p1.1*(1-p1.1))
m5 = x.0^2*(pi*theta.0*p1.0*(1-p1.0) + (1-pi)*theta.0*p0.0*(1-p0.0)) +
  x.1^2*(pi*theta.1*p1.1*(1-p1.1) + (1-pi)*theta.1*p0.1*(1-p0.1))
M = matrix (c(m1, m2, m3, m2, m2, m4, m3, m4, m5), nrow = 3, ncol = 3, byrow = T)

#g vector
g1 = theta.0*(p1.0*(1-p1.0) - p0.0*(1-p0.0)) + theta.1*(p1.1*(1-p1.1) - p0.1*(1-p0.1))
g2 = theta.0*(p1.0*(1-p1.0)) + theta.1*(p1.1*(1-p1.1))
g3 = x.0*theta.0*(p1.0*(1-p1.0) - p0.0*(1-p0.0)) + x.1*theta.1*(p1.1*(1-p1.1) - p0.1*(1-p0.1))
g = matrix (c(g1, g2, g3), nrow = 1, ncol = 3, byrow = T)

N_m.5 = ((g%*%solve(M)%*%t(g))*(((z_beta + z_alpha)^2)/delta^2)*(1 + (5 - 1)*rho.gender))/2
A = (g%*%solve(M)%*%t(g))
m_range = matrix(c(5, 10, 15, 20, 15, 30))

deff.5 = (1 + (5 - 1)*rho.gender)
k.5 = ((deff.5/5)*A*(((z_beta + z_alpha)^2)/delta^2))/2

deff.10 = (1 + (10 - 1)*rho.gender)
k.10 = ((deff.10/10)*A*(((z_beta + z_alpha)^2)/delta^2))/2

deff.15 = (1 + (15 - 1)*rho.gender)
k.15 = ((deff.15/15)*A*(((z_beta + z_alpha)^2)/delta^2))/2

deff.20 = (1 + (20 - 1)*rho.gender)
k.20 = ((deff.20/20)*A*(((z_beta + z_alpha)^2)/delta^2))/2

deff.25 = (1 + (25 - 1)*rho.gender)
k.25 = ((deff.25/25)*A*(((z_beta + z_alpha)^2)/delta^2))/2

deff.30 = (1 + (30 - 1)*rho.gender)
k.30 = ((deff.30/30)*A*(((z_beta + z_alpha)^2)/delta^2))/2

k_range = matrix(c(k.5, k.10, k.15, k.20, k.25, k.30))
Final_range = cbind(m_range,round(k_range,0),round(m_range*k_range,0))
colnames(Final_range) = c('# per cluster','# clusters','Total N')
#end###
```

Annexe 4. Descriptions formelles des calculs de la taille de l'échantillon

Essentiellement, lorsqu'il fournit des estimations de la taille de l'échantillon dans un document, le lecteur doit disposer de suffisamment d'informations pour être en mesure de recréer les calculs si nécessaire. Par conséquent, notez si le projet utilisera un plan en groupes, la taille de l'échantillon sera calculée pour chaque base d'échantillonnage qui sera étudiée (c.-à-d. les élèves et les écoles, les enseignants, les producteurs, etc.). Notez également le nombre de groupes et le nombre d'individus par groupe. Référencez l'équation utilisée à partir d'un document accessible au public. Clarifiez également les valeurs de tous les paramètres utilisés (α, β, δ, ρ, p, SD, etc.) Il peut être plus facile de présenter certaines de ces informations dans un tableau. Voir l'exemple ci-dessous d'une proposition récente.

> « Une approche d'échantillonnage en groupes en deux étapes sera utilisée pour sélectionner tous les répondants aux enquêtes quantitatives (tableau A4.1). À la première étape, les écoles seront sélectionnées au hasard en tant que groupes, puis les élèves, les enseignants, les cuisiniers, les soignants et les mères (au sein des communautés respectives qui alimentent les écoles) seront sélectionnés à la deuxième étape. Dans chaque école, le directeur de l'école sera également interviewé. Les équations utilisées pour déterminer la taille de l'échantillon génèrent la taille minimale de l'échantillon nécessaire pour détecter une différence statistique dans les indicateurs clés de résultats au fil du temps. Tous les échantillons seront augmentés d'au moins 5 %, en cas d'erreurs de collecte ou de saisie.
>
> La taille des échantillons a été calculée à l'aide des équations (6), (19) et (22) pour les résultats continus en groupes, binaires non groupés et binaires en groupes, respectivement, chez McConnell et Vera-Hernandez (2015), en utilisant la puissance standard de 80 % et le niveau de signification de 5 %. Les détails propres à l'indicateur sont notés dans des notes de bas de page individuelle. Certains indicateurs ont été convertis en pourcentages afin que les changements dans les différences moyennes puissent être détectés au cours de la durée du projet.

TABLE 5. EXEMPLE DE PRESENTATION DE LA TAILLE DE L'ECHANTILLON

INDICATEUR	REPONDANT INDIVIDUEL	BASE DE REFERENCE (ESTIMATION)	CIBLE (DUREE DE VIE DU PROJET)	CORRELATION INTRA-GROUPE (CCI)	GROUPES	INDIVIDU PAR GROUPE
Taux moyen d'assiduité des étudiants[26]	Classes	93 %	97 %	0,74	1 500	5
Pourcentage d'élèves démontrant qu'ils savent lire	Étudiants	21 %	41 %	0,43 (Duflo, Glennerster, and Kremer 2007)	44	5
Pourcentage d'aidants naturels qui déclarent consacrer du temps à des activités d'alphabétisation	Aidants naturels	42 %	62 %	0,20[27]	34	5
Pourcentage d'enfants âgés de 6 à 23 mois recevant un régime alimentaire minimum acceptable	Mères d'enfants de 6 à 23 mois	67 %	79 %	0,08 (Moss et al. 2018)	56	5

[26] Il s'agit d'un nouvel indicateur pour l'USDA, et il est difficile de trouver des valeurs CCI publiées pour celui-ci. Cependant, d'après les calculs de l'CCI à partir des données officielles sur la fréquentation du projet McGovern-Dole (phase 4) mis en œuvre par CRS en Sierra Leone, la variabilité attendue est basée en grande partie sur les écoles, et non sur les salles de classe à l'intérieur des écoles, d'où l'importance de l'CCI. L'écart-type pertinent était de 0,44. Compte tenu de la très grande taille de l'échantillon nécessaire, STARS utilisera simplement un recensement de toutes les salles de classe de toutes les écoles pour cet indicateur pour l'étude de référence et réexaminera ces calculs à l'aide des données de référence après la collecte.

[27] Compte tenu de la nature personnalisée de cet indicateur, il est difficile de trouver une valeur CCI publiée pour celui-ci. Étant donné qu'il s'agira d'une pratique familiale, on suppose une valeur d'CCI de 0,20, qui se situe entre les CCI identifiés au niveau de l'école et de la mère ci-dessus.

Annexe 5. Guide de référence rapide

Il est préférable d'utiliser cette annexe après s'être familiarisé avec le guide principal.

Ce guide de référence rapide se veut un rappel rapide des étapes de base du processus de calcul de la taille de l'échantillon. Il est préférable de l'utiliser après avoir examiné le guide complet et ne couvre pas autant de scénarios possibles.

Étape 1 : Dressez une liste d'indicateurs à mesurer (le tableau 6 en donne un exemple). Pour chaque indicateur, notez s'il est continu ou binaire ; le répondant ; et le groupe du répondant (si la collecte de données doit être regroupée). *Section 1 du guide.*

Étape 2 : Ajouter à la liste, pour chaque indicateur, le changement à mesurer. Il s'agit souvent de la cible de durée de vie du projet moins la valeur de référence attendue, mais il peut également s'agir de différences attendues entre les groupes témoins et de traitement à la fin d'une étude, ou de différences entre les groupes à la fin d'un projet.

Si aucune comparaison n'est effectuée, notez la marge d'erreur à l'intérieur de laquelle l'indicateur sera mesuré. *Section 2.2.2 du guide.*

Etape 3 : Ajouter à la liste, pour chaque indicateur dont la collecte de données sera regroupée, le coefficient de corrélation intra-groupe (CCI). *Annexe 1 du guide.*

Étape 4 : Ajouter à la liste, pour chaque indicateur continu, l'écart-type correspondant. *Annexe 1.3 du guide.*

Étape 5 : Pour chaque indicateur, utilisez la figure 1 pour déterminer l'équation à utiliser. Ajoutez-le à la liste. Il est à noter que, si, à l'étape 2, il a été déterminé qu'aucun changement au fil du temps ne sera mesuré, utilisez les équations (5 à 8) au lieu des équations (1 à 4).

Étape 6 : À l'aide des renseignements ci-dessus et de la feuille de calcul Excel de l'annexe 2, calculez le nombre minimal de groupes (s'il y a lieu) et de répondants par groupe. Ajoutez-le à la liste.

Étape 7 : Si les répondants (bases de sondage) chevauchent les indicateurs, utilisez la plus grande taille d'échantillon recommandée par type de répondant.

Étape 8 : Finalisez les calculs en augmentant la taille de l'échantillon de 5 à 20 % pour tenir compte des erreurs de collecte de données, de l'attrition, etc. *Section 3 du guide.*

TABLE 6. EXEMPLE DE LISTE PREPARATOIRE POUR LES CALCULS

INDICATEUR	CONTINU OU BINAIRE	TYPE DE REPONDANT	TYPE DE GROUPE	CHANGEMENT A DETECTER	CCI	ÉCART TYPE	ÉQUATION	GROUPES NECESSAIRES	REPONDANTS NECESSAIRES PAR GROUPE	TAILLE FINALE DU GROUPE (REPONDANT)
Taux moyen d'assiduité des élèves	Continu	Classes	École	4 %	0,74	0,44	2	1,500	5	Recensement[28]
Pourcentage d'élèves démontrant qu'ils peuvent lire un texte de niveau scolaire	Binaire	Étudiants	École	20 %	0,43	N/A	4	44	5	45 (6)
Pourcentage de personnes démontrant l'utilisation de nouvelles pratiques de préparation des aliments sécuritaires	Binaire	Cuisiniers	École	35 %	0,90	N/A	4	14	2	16 (2)
Pourcentage d'soignants qui ont déclaré avoir consacré du temps à des activités d'alphabétisation avec leurs élèves au cours de la semaine précédente	Binaire	Proches aidants	École	20 %	0,20	N/A	4	34	5	35 (3)
Pourcentage d'enfants de 6 à 23 mois recevant un régime alimentaire minimum acceptable	Binaire	Soignants d'enfants de 6 à 23 mois	École	12 %	0,08	N/A	4	56	5	56 (6)

[28] Notez que le projet a desservi moins de 1 500 écoles, ils devront donc effectuer un recensement de toutes les salles de classe dans toutes les écoles.

Confidentiel - Annexe 6. Autres guides de taille d'échantillon

L'annexe 6 ne doit pas être partagée à l'extérieur de CRS. Cela comprend les organisations partenaires de CRS.

La présente annexe met l'accent sur d'autres guides sur la taille de l'échantillon, qui ne sont pas des CRS. Il s'agit d'une annexe distincte du guide principal, car il contient des renseignements qui sont la propriété de CRS.

Bibliographie

Bence, James R. 1995. "Analysis of Short Time Series: Correcting for Autocorrelation." *Ecology* 76 (2): 628-639. https://doi.org/0.2307/1941218.

Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50 (2): 317-372. https://doi.org/10.3368/jhr.50.2.317. http://jhr.uwpress.org/content/50/2/317.abstract.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: methods and applications*. Cambridge University Press.

Catholic Relief Services. 2023. "Monitoring, Evaluation, Accountability and Learning (MEAL) Policies & Procedures." https://www.crs.org/our-work-overseas/research-publications/monitoring-evaluation-accountability-and-learning-policies-procedures.

Chowa, Gina, David Ansong, and Mathieu R. Despard. 2014. "Financial Capabilities: Multilevel Modeling of the Impact of Internal and External Capabilities of Rural Households." *Social Work Research* 38 (1): 19-35. https://doi.org/10.1093/swr/svu002.

Culligan, Mike, Leslie Sherriff, Clara Hagens, Guy Sharrock, and Roger Steele. 2019. *A Guide to the MEAL DPro: Monitoring, Evaluation, Accountability and Learning for Development Professionals.* Catholic Relief Services, Humentum, and the Humanitarian Leadership Academy (Downloaded from http://mealdpro.org/ on July 25, 2019).

DataCamp. 2019. "Inferential Statistics Course." Accessed July 9, 2019. https://www.datacamp.com/community/open-courses/inferential-statistics.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2007. *Using Randomization in Development Economics Research: A Toolkit*. Vol. 6059.*Discussion Paper Series*. London: Centre for Economic Policy Research.

Frost, Jim. 2020. "Comparing Hypothesis Tests for Continuous, Binary, and Count Data." Statistics By Jim. Accessed 28 October 2020. https://statisticsbyjim.com/hypothesis-testing/comparing-hypothesis-tests-data-types/#:~:text=Additionally%2C%20the%20samples%20sizes%20are,sizes%20can%20become%20quite%20large.

Goldstein, Harvey, William Browne, and Jon Rasbash. 2002. "Partitioning Variation in Multilevel Models." *Understanding Statistics* 1 (4): 223-231. https://doi.org/10.1207/S15328031US0104_02.

Handa, Sudhanshu, Thomas de Hoop, Mitchell Morey, and David Seidenfeld. 2018. "ICC Values in International Development: Evidence across Many Domains in sub-Saharan Africa." Centre for the Study of African Economics conference, United Kingdom.

Hansen, Morris H., and Willimam N. Hurwitz. 1942. "On the Theory of Sampling from Finite Populations." *The Annals of Mathematical Statistics* 14 (4): 333-362. https://doi.org/https://www.jstor.org/stable/2235923.

Higgins, JPT, T Li, and JJ Deeks, eds. 2019. *Cochrane Handbook for Systematic Reviews of Interventions*. Edited by JPT Higgins, J Thomas, J Chandler, M Cumpston, T Li, MJ Page and VA Welch. Version 6.0 ed, *Cochrane Handbook for Systematic Reviews of Interventions*: Available from www.handbook.cochrane.org.

Horvitz, Daniel G., and Donovan J. Thompson. 1952. "A generalization of sampling without replacement from a finite universe." *Journal of the American Statistical Association* 47 (260): 663-685. https://doi.org/10.2307/2280784.

Killip, Shersten, Ziyad Mahfoud, and Kevin Pearce. 2004. "What is an intracluster correlation coefficient? Crucial concepts for primary care researchers." *Annals of Family Medicine* 2 (3): 204-208. https://doi.org/10.1370/afm.141.

Lana, Milza M. 2012. "The effects of line spacing and harvest time on processing yield and root size of carrot for Cenourete® production." *Horticultura Brasileira* 30 (2): 7. https://doi.org/10.1590/S0102-05362012000200020.

Lee, Jin Young, and Gary Solon. 2011. "The fragility of estimated effects of unilateral divorce laws on divorce rates." *The BE Journal of Economic Analysis & Policy* 11 (1). https://doi.org/10.3386/w16773.

McConnell, Brendon, and Marcos Vera-Hernandez. 2015. *Going beyond simple sample size calculations: a practitioner's guide.* Institute for Fiscal Studies. https://ifs.org.uk/publications/going-beyond-simple-sample-size-calculations-practitioners-guide.

Moss, Cami, Tesfaye Hailu Bekele, Mihretab Melesse Salasibew, Joanna Sturgess, Girmay Ayana, Desalegn Kuche, Solomon Eshetu, Andinet Abera, Elizabeth Allen, and Alan D Dangour. 2018. "Sustainable Undernutrition Reduction in Ethiopia (SURE) evaluation study: a protocol to evaluate impact, process and context of a large-scale integrated health and agriculture programme to improve complementary feeding in Ethiopia." *BMJ Open* 8 (7). https://doi.org/10.1136/bmjopen-2018-022028.

National Research Council. 2002. *Studies of Welfare Populations: Data Collection and Research Issues*. Edited by Michele Ver Ploeg, Robert A. Moffitt and Constance F. Citro. Washington, DC: The National Academies Press.

Noggle, Eric. 2017. *The SILC Financial Diaries* Catholic Relief Services (Baltimore, MD). https://www.crs.org/sites/default/files/tools-research/efi-silc-diaries_final_30-october-2017_en_final.pdf.

Raab, Gillian, and Susan Purdon. 2009. "5.2.1 How the weights are calculated using population data." *Practical Exemplars on the Analysis of Surveys*. ReStore Project, National Centre for Research Methods (NCRM). Last Modified 5 July 2009. Accessed 22 April 2020. http://www.restore.ac.uk/PEAS/nonresponse.php.

Schochet, Peter Z. 2013. "Statistical Power for School-Based RCTs With Binary Outcomes." *Journal of Research on Educational Effectiveness* 6 (3): 263-294. https://doi.org/10.1080/19345747.2012.725803.

Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge. 2015. "What Are We Weighting For?" *Journal of Human Resources* 50 (2): 301-316. https://doi.org/10.3368/jhr.50.2.301.

Stukel, Diana Maria. 2018a. *Feed the Future Population-Based Survey Sampling Guide.* Food and Nutrition Technical Assistance Project, FHI 360 (Washington, DC. Downloaded from https://www.fantaproject.org/sites/default/files/resources/FTF-PBS-Sampling%20Guide-Apr2018.pdf on July 8, 2019: FHI 360).

---. 2018b. *Participant-Based Survey Sampling Guide for Feed the Future Annual Monitoring Indicators.* Food and Nutrition Technical Assistance Project, FHI 360 (Washington, DC. Downloaded from https://www.fantaproject.org/sites/default/files/resources/Sampling-Guide-Participant-Based-Surveys-Sep2018_0.pdf on July 8, 2019: FHI 360).

Sullivan, Lisa. 2019. "Power and Sample Size Determination." Accessed July 19, 2019. http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Power/BS704_Power_print.html.

Technical Guidance for Monitoring, Evaluation, and Reporting for Emergency Activities. 2022. USAID Bureau for Humanitarian Assistance (Washington, DC. Downloaded from https://www.usaid.gov/sites/default/files/2022-05/BHA_Emergency_ME_Guidance_February_2022.pdf on December 18, 2023).

The Statistical Engineering Division. 1996. DATAPLOT Reference Manual. Downloaded from https://www.itl.nist.gov/div898/software/dataplot/refman2/ch2/weightsd.pdf on July 25, 2019: National Institutes of Standards and Technology, Information Technology Laboratory.

Thompson, Steven K. 2012. *Sampling*. Edited by Walter A. Shewhart and Samuel S. Wilks. Third ed.*Wiley Series in Probability and Statistics*. Hoboken, New Jersey: Wiley.

Thomsen, Ib, Dinke Tesfu, and David A. Binder. 1986. "Estimation of Design Effects and Intraclass Correlations When Using Outdated Measures of Size." *International Statistical Review / Revue Internationale De Statistique* 54 (3): 343-349. https://doi.org/10.2307/1403063