

# Méthodologie d'échantillonnage et Échantillonneur ASDE

Par Michel Rochon

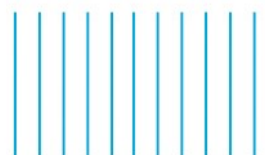
L'énoncé suivant définit de façon générale la méthodologie utilisée par Échantillonneur ASDE pour tirer des échantillons téléphoniques.

« Le logiciel Échantillonneur ASDE fournit des échantillons téléphoniques, générales, stratifiés géographiquement, à composition aléatoire. La stratification est basée sur chaque niveau géographique relié à la demande d'échantillon requise. Dans chaque géographie, les indicatifs téléphoniques et des combinaisons de préfixes sont assignés, selon la pondération accordée à une ou plusieurs géographies, ce qui permet ainsi l'inclusion tout en évitant tout biais d'exclusion qui pourrait en résulter en utilisant une règle décisionnelle de pluralité. »

Le logiciel Échantillonneur ASDE a été conçu et perfectionné au Canada depuis 1995. Aujourd'hui une grande majorité de firmes en recherche et sondage en font usage, dont plusieurs des mieux connues et raffinées.

D'après les présentes définitions aux États-Unis nos échantillons sont des échantillons « Aléatoire B ». Afin de clarifier notez les définitions suivantes:

1. Les échantillons Aléatoire A sont des échantillons de numéros aléatoires sélectionnés systématiquement avec une probabilité égale dans tous les blocs de numéro de téléphone éligibles.
2. Les échantillons Aléatoire B sont des échantillons de numéros aléatoires distribués parmi tous les blocs éligibles, en proportion de leur densité de ménages avec numéro de téléphone répertorié.



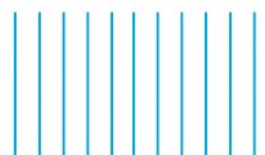
La différence entre Aléatoire A et Aléatoire B est celle-ci:

Les **échantillons Aléatoire A** utilisent chaque bloc éligible même quand les blocs n'ont pas de numéros associés à eux dans les bottins téléphoniques, tandis que les échantillons Aléatoire B n'utilisent que les blocs qui ont au moins un numéro répertorié.

Les **échantillons Aléatoire B** sont plus efficaces, offrent un facteur de projection très élevé et répondent aux besoins de la plupart des études de recherche. Étant donné que ces échantillons sont sélectionnés parmi des blocs selon leur densité de ménages avec numéro de téléphone répertorié, (non pas tirés de tous les blocs de façon égale comme avec les échantillons Aléatoire A), il existe une possibilité que des régions extrêmement non répertoriée soient sous-représentées. Par contre vous gagnez une plus grande efficacité en retour car le taux de numéros en service est beaucoup plus élevé dans les échantillons Aléatoire B.

### **Échantillons ASDE à composition aléatoire (CA)**

Le **cadre d'échantillonnage à composition aléatoire présenté par d'Échantillonneur ASDE** découle d'une analyse complète de 120 millions numéros de téléphone répertoriés dans les annuaires téléphoniques en Amérique du Nord, ainsi que du géocodage des indicatifs au niveau des centrales téléphoniques dans chaque pays (indicatif régional + 3 premiers chiffres du numéro xxx-xxx). Cette analyse se répète à tous les 6 mois avec chaque nouvelle édition des données électroniques de numéros de téléphones provenant des INFOBASES d'AXCIOM. Les annuaires électroniques sont complétés avec l'addition de la liste Telcordia d'indicatifs en fonction en Amérique du Nord.



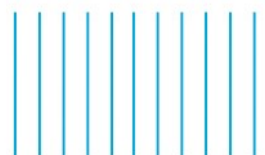
Les numéros de téléphone sont analysés au niveau du bloc à 100 numéros : xxx-xxx-xnnn ; les numéros résidentiels et d'affaires sont comptés dans chaque bloc; le plus petit et le plus grand numéro assigné dans chaque bloc sont identifiés. Les numéros d'affaires et résidentiels font l'objet d'une recherche séparée.

Pour chaque région cible sélectionnée par l'utilisateur, le logiciel Échantillonneur ASDE regroupe tous les blocs faisant partie de la région cible et cumule le total de numéros de téléphone en service. (N) Pour la taille d'échantillon requise (n) dans chaque région cible, le logiciel emploie une procédure d'échantillonnage systématique afin de choisir un bloc de départ avec une probabilité proportionnelle au nombre de numéros de téléphone répertoriés.

Le logiciel crée par la suite, de façon aléatoire, des numéros de téléphone se situant entre le plus petit et plus grand numéro connu, jusqu'à temps qu'un numéro soit précisément trouvé. Le programme procède ensuite (N/n) au prochain bloc, (ou si  $N/n <$  grandeur du bloc à l'intérieur du même bloc), et répète la procédure de sélection. Chaque numéro généré est vérifié contre une base de données accréditée de numéros de téléphone répertoriés. Le numéro est identifié comme étant répertorié (NR) ou non répertorié (NNR) et classé soit numéro d'affaires ou résidentiel.

Pour les **échantillons résidentiels** les numéros d'entreprises sont rejetés et remplacés à l'intérieur du même bloc fonctionnel. Pour les **échantillons d'affaires**, les numéros résidentiels sont rejetés et remplacés à l'intérieur du même bloc fonctionnel. L'information requise est conservée (nom, adresse et code postal) une fois que le numéro est identifié comme répertorié (NR).

Pour les **numéros non répertoriés (NNR)** un code postal approximatif (ou code zip aux États-Unis) est ajouté. Les géocodes de toutes les frontières pertinentes de recensement et électorales sont également ajoutés par la suite. L'échantillon à composition aléatoire est donc constitué des numéros NR et NNR qui se joignent ensemble. Des mesures pondérées sont appliquées à l'échantillon NNR selon la densité des numéros répertoriés dans le bloc où le numéro NNR a été généré.



## Évaluations

Le programme, avec toutes ses hypothèses sous-jacentes et résultats, a été analysé à fond par un évaluateur externe et indépendant, monsieur Alan Sunter. Monsieur Sunter est un membre de l'Association américaine des statistiques et un des premiers experts au Canada en matière de méthodologie en échantillonnage et d'évaluation de cadre d'échantillons.

Monsieur Sunter connaît les éléments spécifiques liés au fonctionnement du logiciel d'Échantillonneur ASDE, ainsi que les structures internes et algorithmes de la base de données. Il a accès aux versions opérationnels du programme et aux échantillons tirés, avec tracé de vérification du processus interne d'échantillonnage dans le programme. Il a confirmé sans réserve que l'échantillon offert par Échantillonneur ASDE est l'équivalent statistique d'un échantillon à composition aléatoire tout en insérant au procédé de sondage des avantages considérables en terme de coût.

Il note dans son évaluation en 1996;

« Vous m'avez demandé de comparer votre procédé d'échantillonnage avec la procédure d'échantillons à composition aléatoire plus typique qui est souvent spécifié comme « obligatoire » dans les requêtes de propositions gouvernementales. Votre procédé, tel que démontré dans une analyse antérieure, sélectionne les numéros répertoriés et non répertoriés avec une probabilité égale et sans remplacement à l'intérieur de chaque strate (normalement un indicatif au niveau centrale téléphonique ou un bloc d'indicatifs au même niveau).

L'échantillon de numéros répertoriés est de grandeur fixe, déterminée par l'utilisateur en choisissant l'option d'allocation de l'échantillon. L'échantillon de numéros non répertoriés démontre une variance d'échantillonnage dans la grandeur d'échantillon qui est effectivement atteinte, mais cette variance (qui est inévitable quoiqu'il arrive) n'a qu'un impact négligeable, selon moi, sur l'efficacité du modèle dans l'ensemble. Les variances d'échantillonnage des estimations de strate seront



équivalentes, ou moindres, que celles obtenues par de simples échantillons aléatoires de pareille grandeur. »

Dans son évaluation de 2001 monsieur Sunter indique:

"Comme dans mes revues précédentes je confirme que le système sélectionne des numéros parmi les catégories répertoriés/non répertoriés dans chaque strate avec une probabilité égale et sans remplacement, et que les variances d'échantillonnage des estimations de strate seront équivalentes, ou moindres, que celles obtenues par de simples échantillons aléatoires de pareille grandeur. »

### **Profil de monsieur Alan Sunter**

Alan Sunter est un statisticien et consultant avec une expérience qui comprend cinq ans à titre de directeur des méthodes de sondage d'affaires chez Statistiques Canada, deux ans comme consultant aux Agences nationales de statistiques de la Suède et du Royaume-Uni, deux ans avec l'organisation du Sondage mondial sur la fécondité, et 12 ans en consultation pour des agences privées et publiques au Canada ainsi qu'avec des agences de statistiques dans des pays moins développés. Il a conçu et dirigé plusieurs sondages d'affaires au Canada ainsi que des sondages résidentiels au Bangladesh, à la Côte d'Ivoire, au Nigeria, Bénin, à l'Éthiopie, au Trinidad et Tobago et en Colombie.

Il est spécialisé en conception de sondage, traitement des données de sondage, et la mesure des erreurs d'échantillonnage et non-échantillonnage dans les sondages. Il a également publié de nombreux exposés en théorie statistique. Il a été élu à l'Institut internationale des statistiques et devenu membre de l'Association américaine des statistiques pour ses contributions importantes à la théorie et pratique de la conception et développement des sondages.

