

Simplifions la statistique

Par Louis Blais, stat., ASSQ

ÉCHANTILLONNAGE STRATIFIÉ

En foresterie, il est très fréquent d'échantillonner sur un territoire dont tous les critères ne sont pas homogènes. La délimitation d'une unité de sondage se fait selon un ou plusieurs critères (stratification ou regroupée), mais pour d'autres critères, il est difficile d'obtenir l'homogénéité.

Voici quelques exemples de territoires non homogènes :

- un volume de bois doit être prélevé ou mesuré sur plusieurs parties disjointes de territoires de tailles différentes. Certaines parties présentent un plus grand volume de bois par hectare que d'autres;
- la régénération doit être mesurée, mais son portrait (quantité, essence, compétition)

varie sur plusieurs parties disjointes de territoires de tailles différentes.

Si l'hétérogénéité du territoire n'est pas prise en compte, il y a un risque qu'une petite strate au volume très différent des autres ait une influence telle que le volume moyen par hectare en soit altéré de façon importante.

L'échantillonnage stratifié est tout indiqué pour remédier à ce genre de problème. Il s'agit d'un procédé d'échantillonnage où la population est répartie en sous-groupes homogènes et où le tirage d'échantillons est indépendant dans chaque strate. Dans le cas qui nous concerne, le territoire forestier est divisé en plusieurs sous-territoires disjoints et homogènes, appelés strates¹, dans lesquels seront placées des unités d'échantillonnage. Ce type d'échantillonnage comporte plusieurs avantages, car il permet :

- de calculer une moyenne pondérée en fonction de l'importance de chaque strate;
- de diminuer la variabilité globale du plan de sondage, particulièrement si les moyennes ou les variances entre les strates sont différentes. En diminuant la variabilité globale, le nombre d'unités d'échantillonnage requis diminue;
- d'optimiser la répartition des unités d'échantillonnage;
- d'être utilisé lorsqu'une précision globale s'avère nécessaire, sans égard aux strates. Si l'on exige une précision par strate, l'échantillonnage stratifié n'est pas utile. Dans ce cas, il faut faire un plan de sondage par strate et les traiter individuellement et non de façon globale.

Il faut cependant connaître *a priori* une donnée quantifiable sur chaque strate pour pouvoir mesurer l'importance d'une strate par rapport à une autre et déterminer la pondération entre les strates. Dans la plupart des cas, en foresterie, la donnée quantifiable est la superficie de chaque strate. L'utilisation de l'échantillonnage stratifié nécessite au moins deux unités d'échantillonnage par strate. Le calcul de la précision statistique de ce type d'échantillonnage en dépend.

1. Le terme « strate » utilisé dans cette fiche doit être pris au sens large. Il peut désigner une strate forestière, une strate regroupée, un type d'intervention sylvicole, un territoire administratif, etc.

CONSTRUCTION D'UNE STRATE

La construction d'une strate doit être reliée à la variable d'intérêt. Par exemple, si la variable d'intérêt est le volume à récolter d'une essence précise, les différentes strates devraient être construites en fonction de la variation spatiale du volume de cette essence. Dans ce cas, il faut, par exemple, regrouper les zones à fort volume, puis les zones à faible volume et enfin celles qui se situent entre les deux.

L'utilisation de strates forestières peut être un bon critère, mais cette méthode peut aussi générer un très grand nombre de strates. Dans ce cas, le regroupement de strates en fonction de la variable d'intérêt est une bonne solution. Il est connu que la composition en essences, la densité et le drainage ont une influence sur le volume de bois sur pied.

Idéalement, l'unité de sondage devrait comporter un petit nombre de strates. Puisqu'il doit y avoir au moins deux unités d'échantillonnage par strate pour pouvoir en calculer la précision, un grand nombre de strates implique un nombre minimal d'unités d'échantillonnage représentant au moins deux fois le nombre de strates, ce qui peut faire augmenter considérablement le coût de l'inventaire.

ALLOCATION DES UNITÉS D'ÉCHANTILLONNAGE

Il existe plusieurs façons d'allouer les unités d'échantillonnage dans chaque strate, c'est-à-dire de déterminer combien d'unités seront échantillonnées dans chacune des strates.

Généralement, il y a trois critères : la taille des strates, leur variabilité et le coût associé à chacune. À partir de ces trois critères, il est possible d'optimiser l'allocation des unités dans chaque strate. La formule générale est la suivante :

$$n_h = \left(\frac{d_h}{\sum_{i=1}^H d_i} \right) n$$

où

$$\text{« } d_h \text{ »} = \frac{A_h S_h}{\sqrt{c_h}}$$

« n_h » est le nombre d'unités d'échantillonnage de la strate « h »;

« n » est le nombre total d'unités d'échantillonnage;

« A_h » est la superficie de la strate « h »;

« s_h » est l'écart-type de la strate « h »;

« c_h » est le coût de réalisation d'un inventaire de la strate « h »;

« H » est le nombre total de strates.

Cette formule tient compte des trois éléments :

- si les écarts types de même que les coûts des strates sont égaux, cette formule revient à allouer les unités d'échantillonnage de façon proportionnelle à la taille de la strate, c'est-à-dire à attribuer plus d'unités d'échantillonnage aux strates de plus grande taille;
- si les superficies de même que les coûts des strates sont égaux, cette formule permet d'allouer plus d'unités d'échantillonnage à une strate qui a une plus grande variabilité;

- si les superficies de même que les écarts types des strates sont égaux, cette formule permet de diminuer le nombre d'unités d'échantillonnage dans une strate lorsque l'inventaire à l'intérieur de cette strate est plus coûteux.

S'il est impossible de connaître l'écart type ou les coûts, le mieux est d'y aller de façon proportionnelle à la taille de la strate.

PRÉCISION DÉSIRÉE DE L'INVENTAIRE

La précision désirée de l'inventaire est un facteur déterminant lors de la planification du plan de sondage. Elle peut se faire à deux niveaux :

- l'unité de sondage totale;
- la strate.

Prenons l'enquête sur la population active au Canada. Si on demandait à 2 000 personnes si elles occupent un emploi, la précision sur le taux de chômage en général serait très bonne. Cependant, la précision sur le taux de chômage par province ou par région serait alors faible ou très faible. Comme les élus prennent certaines décisions en fonction du taux de chômage par région, cela exige une plus grande précision à ce niveau. Pour y arriver, Statistique Canada interviewe 56 000 ménages afin de rejoindre environ 100 000 personnes. C'est un inventaire plus coûteux, mais qui permet de prendre de meilleures décisions pour les localités canadiennes. Le même principe peut s'appliquer à la foresterie.

La précision peut être établie en fonction de la valeur des bois, une option rarement mentionnée que l'on utilise en Suède. En effet, si un peuplement contient des bois de grande valeur, il est possible d'exiger une plus grande précision dans cette strate par rapport à une autre dont les bois sont de moins grande valeur. C'est une autre façon d'optimiser la répartition des unités d'échantillonnage.

Pour plus de détails, consultez le *Guide d'inventaire et d'échantillonnage en milieu forestier* (Méthot et autres, 2013), *Sampling: Design and Analysis* (Lohr, 1999) ou *Sampling Strategies for Natural Resources and the Environment* (Gregoire et Valentine, 2008).

POUR EN SAVOIR PLUS...

Ministère des Forêts, de la Faune et des Parcs
Direction de l'aménagement et de l'environnement forestiers
5700, 4e Avenue Ouest
Québec (Québec) G1H 6R1
daef@mffp.gouv.qc.ca