

**Mise en application des
concepts d'échantillonnage
et de statistique**

Préparées par :



Centre collégial de transfert de technologie
en foresterie
Donald Blouin, ing.f., M.Sc.

Mars 2001

TABLE DES MATIÈRES

| | |
|--|----------|
| Table des matières | i |
| Liste des figures | i |
| Introduction | ii |
| 1. THÉORIE DE L'ÉCHANTILLONNAGE | 1 |
| 2. COMPILATIONS STATISTIQUES | 3 |
| 2.1 Paramètres de base | 3 |
| 2.2 Regroupement de strates | 7 |
| 2.3 Comparaison de deux moyennes | 8 |
| Références | 9 |
| | |
| ANNEXE 1 - Distribution normale | |
| ANNEXE 2 - Table de t de « Student » | |

Liste des figures

| | |
|---|---|
| Figure 1 - Hauteurs moyennes identiques de deux peuplements différents | 3 |
| Figure 2 - Distribution de fréquence normale avec une moyenne égale à 0 et un écart-type ou un intervalle de confiance de 1 | 4 |

INTRODUCTION

Dans le cadre de la réalisation des interventions forestières sur les forêts publiques au Québec, les bénéficiaires de CAAF doivent rendre compte de ce qu'ils font et de la qualité de leurs travaux au ministère des Ressources Naturelles. Afin de guider les bénéficiaires dans ces démarches, le MRN publie et met à jour régulièrement deux documents de référence : les « Instructions relatives à l'application du règlement sur la valeur des traitements sylvicoles admissibles en paiement des droits » (MRN, 2000a) et la « Méthode d'échantillonnage pour les suivis des interventions forestières » (MRN, 2000b) . Il s'agit de documents ayant force de loi décrivant les procédures à suivre pour l'évaluation de la qualité des interventions forestières. Pour s'entendre, les bénéficiaires et les agents du Ministère doivent d'abord connaître et comprendre les démarches suggérées et savoir interpréter les résultats des compilations réalisées.

Ce document vise à faire un rappel sur les éléments de base de la théorie de l'échantillonnage et sur la réalisation de compilations statistiques.

1. THÉORIE DE L'ÉCHANTILLONNAGE

Lors de la planification d'intervention sylvicole, l'unité d'échantillonnage est la référence utilisée par le Ministère des Ressources naturelles (MRN, 2000b). Il s'agit d'un territoire ayant une superficie maximale de 250 ha, le plus homogène possible, sur lequel un seul type d'intervention sylvicole sera pratiqué. En terme mathématique, l'unité d'échantillonnage de l'inventaire forestier devient la population pour l'inventaire d'intervention.

La population peut être définie comme étant tous les arbres présents dans l'unité d'échantillonnage. Lorsque vient le temps d'évaluer un paramètre forestier et qu'il est techniquement impossible de mesurer l'ensemble des individus de la population, c'est là qu'entre en jeu le concept d'échantillonnage.

L'échantillonnage consiste à la réalisation de la mesure de certains paramètres sur une fraction de la population. Le taux d'échantillonnage de la forêt feuillue mature normalement recommandé au Québec est de l'ordre de 1/125 de la superficie.

Afin d'être considéré comme valable, l'échantillonnage doit être établi en respectant certaines règles de base. Les règles visent à éliminer toutes subjectivités dans le processus de positionnement et d'établissement des parcelles-échantillons et de faire en sorte que les résultats obtenus soit le plus représentatifs possible de la population. Il existe plusieurs types d'échantillonnage; nous ferons ici la description de l'échantillonnage aléatoire et de l'échantillonnage systématique.

L'échantillonnage aléatoire consiste à choisir au hasard la localisation des points de mesure. Dans une unité d'échantillonnage de 250 ha avec des points de mesure de 400 m² et un taux d'échantillonnage de 1/125, il s'agira de choisir au hasard 50 points de mesure parmi les 6250 points de mesure disponibles dans la population.

L'échantillonnage aléatoire présente les principaux avantages suivants (OIFQ, 1996) :

- ✍ il élimine toute subjectivité dans le choix des unités d'échantillonnage à mesurer, car le choix des unités n'est nullement influencé par le personnel qui les sélectionne;
- ✍ il permet de calculer l'erreur d'échantillonnage.

Il comporte aussi certains inconvénients :

- ✍ il est parfois difficile de se rendre sur les parcelles choisies aléatoirement;
- ✍ les unités d'échantillonnage sont distribuées de façon irrégulière et certains secteurs peuvent donc être sur ou sous-échantillonnés;
- ✍ les distances qui séparent les unités d'échantillonnage peuvent prolonger la durée des travaux;
- ✍ cette méthode peut entraîner des coûts prohibitifs si les unités d'échantillonnage sont concentrées dans des zones difficilement accessibles.

L'échantillonnage aléatoire représente davantage un concept théorique et est très peu utilisé dans la pratique, dû à la contrainte d'application qu'il exige. On a plutôt recours à l'échantillonnage systématique.

Tel que décrit dans le guide du MRN sur les « Méthodes d'échantillonnage pour les suivis des interventions forestières » (MRN, 2000b), le plan de sondage doit respecter les principes suivants :

- ✍ couvrir l'ensemble de la superficie par des virées équidistantes;
- ✍ faire une distribution systématique des parcelles-échantillons sur chacune des virées.

Dans cette méthode d'échantillonnage systématique, la localisation de la première parcelle est choisie aléatoirement, alors que les autres parcelles sont toutes équidistantes (Rondeux, 1993).

Les principaux avantages de l'échantillonnage systématique sont les suivants :

- ✍ l'échantillon est plus facile à sélectionner, car on a besoin d'un seul nombre aléatoire;
- ✍ les unités d'échantillonnage sont généralement mieux réparties dans la population, on peut donc estimer la moyenne plus justement qu'avec un échantillon aléatoire comparable;
- ✍ les coûts d'accès aux parcelles-échantillons sont moindres, car le nombre d'unités d'échantillonnage d'une virée est fixé en fonction de ce qu'une équipe peut faire dans une journée normale de travail;
- ✍ les équipes risquent moins de se tromper dans l'emplacement des parcelles;
- ✍ la localisation des unités d'échantillonnage est plus facile.

Les principaux inconvénients de cette méthode sont :

- ✍ la non-possibilité théorique de calcul valable de l'erreur d'échantillonnage, car elle n'est pas fondée sur les lois du hasard. Seulement la première unité est tirée au hasard. Le choix de la première unité détermine automatiquement le choix des autres, qui y deviennent dépendantes. En pratique, les formules utilisées pour l'échantillon aléatoirement peuvent aussi servir à calculer l'erreur d'échantillonnage d'un sondage systématique (Rondeux, 1993; Cochran, 1997);
- ✍ la périodicité, problème qui se pose lorsque la population présente un caractère cyclique qui coïncide avec le pas de sondage. L'échantillon ainsi obtenu n'est pas représentatif de la population.

Assumant qu'en forêt naturelle, il n'y ait pas de répartition cyclique des arbres ou des conditions dans lesquelles ils se trouvent, l'application de l'échantillonnage systématique est reconnue et les compilations statistiques qui lui sont associées sont tout à fait justifiées. Dans le cas où les éléments sont cycliques, le MRN propose des solutions pour palier à la situation :

- ✍ dans le cas de plantations ou regarnis, établir des virées à angles avec les lignes de plantation ($\pm 75^\circ$);
- ✍ établir des virées perpendiculaires aux chemins de débardage dans le cas de coupes commerciales.

2. COMPILATIONS STATISTIQUES

2.1 PARAMÈTRES DE BASE

Paramètre à estimer

- À elle seule, une moyenne ne suffit pas pour qualifier correctement un paramètre (Bédard, 1996). La figure 1 illustre bien que des moyennes similaires de la hauteur de deux peuplements peuvent caractériser deux paysages tout à fait différents et ainsi des écarts-types différents. Il est donc nécessaire d'utiliser des compilations statistiques afin de qualifier les valeurs moyennes des paramètres étudiés.

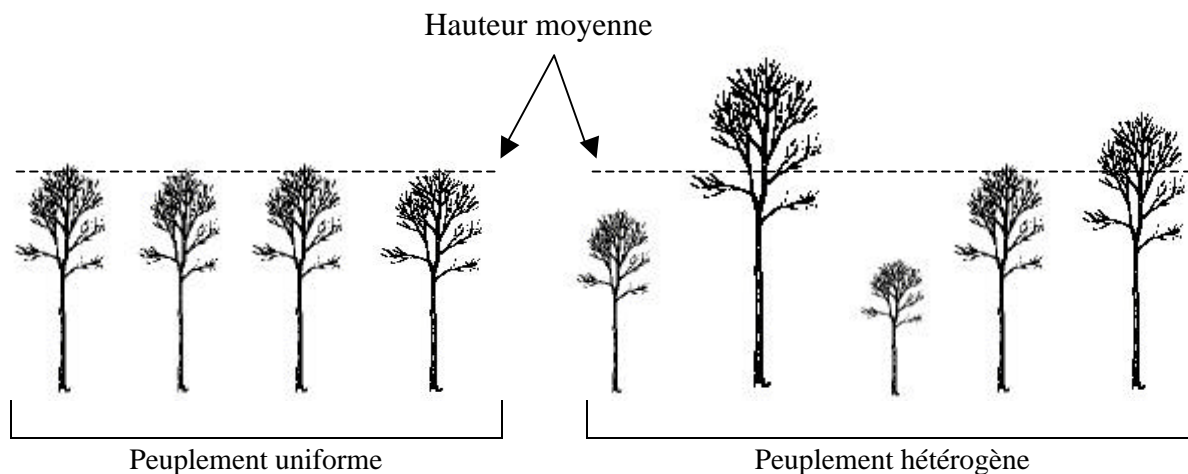


Figure 1 - Hauteurs moyennes identiques de deux peuplements différents

L'écart-type (variance de l'échantillon)

- Ce paramètre exprime la dispersion de la valeur des unités par rapport à la moyenne de l'échantillon. Tel qu'illustré sur la figure 1, le peuplement uniforme a un écart-type petit alors que le peuplement hétérogène a un écart-type plus élevé.

L'erreur standard (variance de la moyenne)

- Ce paramètre exprime la variation de la valeur moyenne de la population étudiée. Elle mesure la dispersion des moyennes d'échantillons autour de la moyenne de la population.

Le coefficient de variation

- C'est la variation relative de l'écart-type par rapport à la moyenne. Cette valeur est exprimée en pourcentage.

La distribution normale

- ☞ Ce paramètre représente la distribution de fréquence théorique et normale d'une population.
- ☞ Il est possible d'en déduire, pour une moyenne donnée, la distribution de fréquence autour de laquelle on retrouve 66 % des valeurs de l'échantillon à l'intérieur d'un écart-type, 95 % des valeurs à l'intérieur de deux écarts-types et 99 % des valeurs à l'intérieur de trois écarts-types (figure 2 et annexe 1). Cela signifie que si nous retournons sur le terrain prendre les mesures d'une parcelle additionnelle, nous avons 95 % de probabilité que le résultat de cette parcelle se retrouve à l'intérieur de l'intervalle compris entre deux écarts-types autour de la moyenne.
- ☞ Concernant l'erreur standard, si l'on échantillonnait la population un très grand nombre de fois et que l'on construisait, pour chaque échantillon, l'intervalle de deux erreurs standards autour de la moyenne, 95% de ces intervalles contiendrait la moyenne de la population.

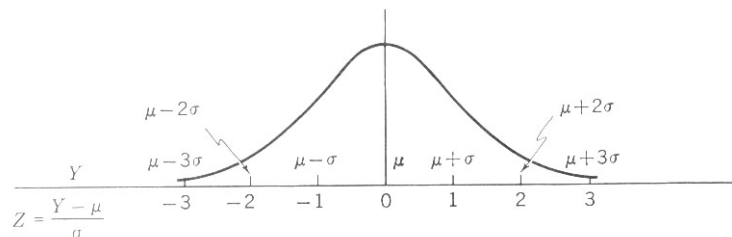


Figure 2 - Distribution de fréquence normale avec une moyenne, μ , égale à 0 et une variance, σ^2 , égale à 1.

Le t de student

- ☞ Il s'agit de l'adaptation de la distribution normale en fonction d'un échantillonnage. Plus le taux d'échantillonnage est faible, plus la courbe est aplatie et les extrémités sont longues. Plus le taux d'échantillonnage est élevé, plus la courbe s'approche de celle de la distribution normale (table de t de student, annexe 2).

L'intervalle de confiance

- ☞ Intervalle à l'intérieur duquel devrait se trouver la vraie valeur du paramètre estimé, et ce avec un certain niveau de probabilité (habituellement 95 %).

La précision

- ☞ Il s'agit du complément de la demi-largeur de l'intervalle de confiance à 95% de la moyenne, exprimée en pourcentage de σ . Plus l'erreur standard est faible, plus grande est la précision.

Les formules permettant le calcul de ces paramètres sont les suivantes :

| | |
|-----------------------------------|--|
| Paramètre à estimer | $\bar{R} = \frac{\sum y_i}{\sum x_i}$ |
| Variance de l'échantillon | $S^2 = \frac{1}{n-1} \left(\sum y_i^2 - \frac{\hat{R}^2 \sum x_i^2}{2R} - \sum x_i y_i \right)$ |
| Écart-type | $S = \sqrt{S^2}$ |
| Variance de la moyenne | $S_{\bar{R}}^2 = \frac{S^2}{n}$ |
| Erreur standard | $E = \sqrt{S_{\bar{R}}^2} = \frac{S}{\sqrt{n}}$ |
| Coefficient de variation | $CV = \frac{S}{\bar{R}} \times 100$ |
| Précision | $P = 100 - T \times \frac{E}{\bar{R}} \times 100$ |
| Intervalle de confiance | $\bar{R} \pm T \times \frac{S}{\sqrt{n}}$ |
| Dimension de l'échantillon | $n = \frac{T^2 \times S^2}{d^2}$ |

- Où :
- y_i = Volume, nombre de tiges ou surface terrière de la parcelle;
 - x_i = Superficie de la parcelle (ha);
 - \bar{y} = Volume, nombre de tiges ou surface terrière moyenne à l'hectare;
 - S^2 = Variance de l'échantillon;
 - S = Écart-type de l'échantillon, si $n=1$ on utilise 60 % du paramètre estimé;
 - $S^2_{\bar{y}}$ = Variance de la moyenne;
 - E = Erreur de la moyenne;
 - CV = Coefficient de variation (%);
 - P = Précision (%);
 - T = 97.5^{ème} percentile de la loi de t de « Student » avec $n-1$ degrés de liberté ;
si $n \geq 120$ utiliser $T = 1.96$;
 - n = Nombre de parcelles;
 - \bar{x} = Superficie moyenne des parcelles;
 - d = Erreur relative correspondant à l'erreur permise par rapport au paramètre estimé
(pour une précision de 90 %, l'erreur relative est égale à 10 % du paramètre
estimé ($d = 10 \% \bar{y}$)).

2.2 REGROUPEMENT DE STRATES

Dans le cas de regroupement de strates (ou secteurs d'intervention), les paramètres calculés doivent être pondérés en fonction de la superficie de chaque strate échantillonnée.

Paramètres pondérés

$$R = \frac{\sum_h N_h R_h}{N}$$

$$S_R^2 = \frac{\sum_h N_h^2 S_{R_h}^2}{N^2}$$

- où :
- R = Moyenne pondérée;
 - S_R^2 = Variance de la moyenne pondérée;
 - N_h = Superficie de la strate h ou nombre de placettes dans la strate h;
 - R_h = Moyenne de la strate h;
 - $S_{R_h}^2$ = Variance de la moyenne de la strate h;
 - N = Superficie totale du regroupement ou nombre de placettes dans le regroupement.

Les autres paramètres (erreur standard, intervalle de confiance et précision) sont calculés de la même façon que précédemment en utilisant la moyenne et la variance de la moyenne pondérée.

2.3 COMPARAISON DE DEUX MOYENNES

La comparaison de deux moyennes est faite par l'utilisation du test de t. En assumant que les deux variances sont différentes, on veut tester l'hypothèse :

$$H_0 : \mu_1 = \mu_2 \text{ versus}$$

$$H_1 : \mu_1 \neq \mu_2$$

En utilisant les moyennes et variances des échantillons, nous pouvons calculer la statistique t_0

$$t_0 = \frac{\bar{R}_1 - \bar{R}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

où t_0 est distribuée selon une distribution de t de Student.

On rejette donc H_0 si $|t_0| > t_{\alpha/2, n_1 + n_2 - 2}$, où $n_1 + n_2 - 2$ degré de liberté.

Où : \bar{R}_1 = Moyenne du groupe 1;

\bar{R}_2 = Moyenne du groupe 2;

S_1^2 = Variance de l'échantillon du groupe 1;

S_2^2 = Variance de l'échantillon du groupe 2;

n_1 = Nombre de parcelles du groupe 1;

n_2 = Nombre de parcelles du groupe 2.

RÉFÉRENCES

- BÉDARD, S. et A. CARLE, 1996. Guide d'inventaire d'intervention en forêt feuillue. Institut québécois d'aménagement de la forêt feuillue (IQAFF). ISBN 0-662-80714-6, 62 p.
- COCHRAN, W.G., 1977. Sampling Techniques (third edition). John Wiley & Sons inc. US, ISBN 0-471-16240-X, 428 p.
- MINISTÈRE DES RESSOURCES NATURELLES DU QUÉBEC, 2000a. Instructions relatives à l'application du règlement sur la valeur des traitements sylvicoles admissibles en paiement des droits – Exercice 2000-2001. Forêt Québec, Dir. de l'assistance technique, Div. des traitements sylvicoles. ISBN 2-550-35590-3, 67 p.
- MINISTÈRE DES RESSOURCES NATURELLES DU QUÉBEC, 2000b. Méthodes d'échantillonnage pour les suivis des interventions forestières. Forêt Québec, Dir. de l'assistance technique, Div. des traitements sylvicoles. ISBN 2-550-35720-5, 204 p.
- ORDRE DES INGÉNIEURS FORESTIERS DU QUÉBEC (OIFQ), 1996. Manuel de foresterie. Les Presses de l'Université Laval. ISBN 2-7637-7479-2, 1428 p.
- RONDEUX, J., 1993. La mesure des arbres et des peuplements forestiers. Les Presses agronomiques de Gembloux. ISBN 2-87016-041-0, 521 p.
- STEEL, R.G.D., et J.H. TORRIE, 1980. Principles and procedures of statistics – A Biometrical Approach (Second edition). McGraw-Hill inc. US, ISBN 0-07-060926-8, 633 p.

Annexe 1

Distribution normale

Annexe 2

Table de t de student
