La Méthode statistique et la Foresterie - III -Distributions statistiques - Application de la Loi Normale, discussion des exemples d'échantillonnage

Lucien Brenac

► To cite this version:

Lucien Brenac. La Méthode statistique et la Foresterie - III -Distributions statistiques - Application de la Loi Normale, discussion des exemples d'échantillonnage. *Revue forestière française*, 1953, S, pp.42. 10.4267/2042/26919 . hal-03384227

HAL Id: hal-03384227

<https://hal.science/hal-03384227>

Submitted on 18 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

APPLICATION DE LA LOI NORMALE

DISCUSSION DES EXEMPLES D'ÉCHANTILLONNAGE

Reportons-nous au chapitre I (Echantillonnage) et aux figures 1 à 4, pages 18 et 20. Il s'agit de l'inventaire du volume effectué avec une intensité d'échantillonnage de

$$\frac{32}{400} = 8 \%$$

par les trois méthodes : au hasard, stratifié, systématique.

La moyenne calculée sur l'échantillon est étendue à l'ensemble. L'erreur à craindre correspondant à une probabilité donnée (voir chap. VI, p. 61) est estimée en utilisant les propriétés de la distribution normale.

En nous fondant sur ses propriétés, nous essaierons d'abord de nous rendre compte si le peuplement étudié est homogène, ensuite nous calculerons les erreurs d'échantillonnage.

Pour cela, il convient de calculer les moyennes et écarts-type des distributions de volumes formées par les 400 petits carrés (fig. 1) d'une part, et, d'autre part, par les 16 blocs. Pour une raison d'économie de calcul, nous commencerons par les blocs.

Sont indiqués ci-dessous, dans un ordre identique à celui de la figure 1, les moyennes et variances correspondant à chacun des blocs.

Moyenne	Variance	Moyenne	Variance	Moyenne	Variance	Moyenne	Variance
7,84	14,21	16,88	28,83	20,44	29,22	18,04	9,25
21,04	26,35	19,36	20,75	18,88	22,17	14,92	34,22
14,24	8,05	16,56	10,51	23,52	36,89	23,44	20,86
6,84	14,75	18,28	26,23	23,68	31,93	20,44	7,45
49,96	63,36	71,08	86,32	86,52	120,21	76,84	71,78

Somme des moyennes $49,96 + 71,08 + 86,52 + 76,84 = 284,40$

Somme des variances $63,36 + 86,32 + 120,21 + 71,78 = 341,67$

Les variances peuvent être ici additionnées comme ayant un dénominateur commun (25 carrés dans chaque bloc).

Moyenne des 400 carrés :

$$\frac{284,40}{16} = 17,775$$

Rapportée à cette moyenne, le calcul de la somme des carrés des écarts entre moyenne des blocs $\bar{x}_b$ et moyenne générale $\bar{x}$ donne

$$(\bar{x}_b - \bar{x})^2 = 372,78$$

La variance de la distribution des 400 carrés rapportée à leur moyenne s'obtient en utilisant la propriété d'additivité des carrés des écarts. On peut écrire en effet :

$$x - \bar{x} = x - \bar{x}_b + \bar{x}_b - \bar{x}$$

En élevant au carré et en faisant la somme étendue aux 16 blocs et 400 carrés, on obtient la relation suivante :

$$\Sigma (x - \bar{x})^2 = 25 \Sigma (\bar{x}_b - \bar{x})^2 + \Sigma (x - \bar{x}_b)^2$$

Le double produit

$$2 \sum (\bar{x}_b - \bar{x}) (x - \bar{x}_b)$$

du second membre s'annule en effet car $(\bar{x}_b - \bar{x})$, constant dans chaque bloc, étant mis en facteur, la somme correspondante des 25 valeurs de $x - \bar{x}_b$ est nulle, comme conséquence de la définition de $\bar{x}_b$. On obtient donc à partir des valeurs déjà calculées :

$$\sum (x - \bar{x})^2 = 25 \times 372,78 + 25 \times 341,67 = 25 \times 714,45 = 17\,861,25$$

d'où la variance

$$\frac{\sum (x - \bar{x})^2}{N} = \frac{17\,861,25}{400}$$

et l'écart-type

$$\sigma'_x = \sqrt{\frac{17\,861,25}{400}} = 6,68$$

Quant à l'écart-type de la moyenne $\bar{x}_b$ des blocs, il est égal à

$$\sigma'_{\bar{x}_b} = \sqrt{\frac{372,78}{16}} = 4,83$$

HOMOGENÉITÉ DU PEUPLEMENT

Ayant ainsi calculé les paramètres de la distribution de x :

$$\bar{x} = 17,78 \text{ et } \sigma'_x = 6,68$$

et de la distribution de $\bar{x}_b$:

$$\bar{\bar{x}} = 17,78 \text{ et } \sigma'_{\bar{x}_b} = 4,83$$

nous sommes en mesure d'estimer le degré d'homogénéité du peuplement en comparant les effectifs ou fréquences réelles à ceux résultant de la distribution normale caractérisés par le tableau de la page 40.

	Effectifs compris à l'intérieur des limites			
	$\pm \sigma'$		$\pm 2 \sigma'$	
	Théorique	Calculé	Théorique	Calculé
	%	%	%	%
Distribution de x (400 petits carrés)	68	67	95,4	95,7 (1)
Distribution de $\bar{x}_b$ (moyenne des 16 blocs)	68	68	95,4	87,5 (2)

(1) 268 carrés entre 11,1 et 24,4 au lieu de 272.

383 entre 4,4 et 31,1 au lieu de 380.

(2) 11 moyennes entre 13 et 23 et 14 entre 8 et 27 au lieu de 15.

L'ensemble des mesures est donc homogène, les écarts constatés dans la répartition des 400 carrés étant visiblement non significatifs. Il convient de remarquer que l'étude ainsi faite sur les valeurs unitaires ne tient aucun compte de leur répartition sur la surface étudiée. Au contraire, la distribution des moyennes de blocs tient compte de cette répartition suivant le découpage établi. Il en résulte un écart sensible dû aux plus faibles valeurs : 2 moyennes de blocs, 6,84 et 7,84, traduisent un écart supérieur à $-2\sigma'$:

$$17,78 - 2 \times 4,83 = 8,12,$$

alors que l'effectif probable en dessous de cet écart est :

$$16 \times 0,0228 = 0,366 \text{ au lieu de } 2.$$

Ce résultat est la conséquence de la dépendance déjà analysée (p. 23) des volumes portés par des surfaces voisines. Comme l'indépendance des parcelles choisies est une condition d'un bon échantillonnage, on voit par là toute l'importance qui s'attache au choix de la surface élémentaire de ces parcelles et à la qualité de la stratification.

La dispersion relative des valeurs est traduite au moyen du coefficient de variation :

$$V = \frac{100 \times \sigma'}{\bar{x}} = \frac{100 \times 6,68}{17,78} = 37,5 \%$$

ERREURS D'ÉCHANTILLONNAGE

Il s'agit de calculer l'erreur à craindre, correspondant à la probabilité de 0,05, sur la moyenne $\bar{x}$ estimée.

Dans le cas de l'échantillonnage au hasard et dans celui de l'échantillonnage stratifié, nous pourrions calculer 2 valeurs de l'erreur : l'une qui peut être appelée la vraie valeur, sera déterminée à partir de la totalité de l'information dont nous disposons, l'autre, calculée comme cela a lieu en pratique, à partir d'un seul échantillon, est une valeur estimée.

Dans tous les cas, le principe du calcul repose sur les propriétés de la distribution normale. La probabilité 0,05 correspond très sensiblement à un écart égal à deux fois l'écart type. On exprime cet écart, comme le coefficient de variation, en % de la moyenne pour obtenir une erreur relative en %.

Les résultats du calcul sont consignés dans le tableau ci-dessous :

Mode d'échantillonnage	au hasard	stratifié	systématique
<i>Moyennes :</i>			
vraie	17,78	17,78	17,78
estimée	18,16	17,53	17,88
écart %	+2,0	-1,4	+0,5
<i>Coefficient de variation :</i>			
vrai %	37,5	26	
estimé %	35,1	33,7	
<i>Erreur d'échantillonnage. P = 0,05 :</i>			
vraie %	±13,3	± 9,2	
estimée %	±12,4	±11,9	

Les erreurs d'échantillonnage calculées sont bien de l'ordre de grandeur de celles que l'expérience a dégagées. Les écarts sur la moyenne sont sensiblement proportionnels aux erreurs vraies. Dans le cas de l'échantillonnage systématique, on ne peut que constater l'écart sur la moyenne, à l'exclusion de tout calcul d'écart-type et d'erreur, puisque le mode d'échantillonnage choisi définit un seul échantillon et qu'il n'y a pas de tirage au hasard.

En général, dans l'échantillonnage systématique, la situation géométrique des parcelles élémentaires dans chaque bloc est bien définie identiquement, mais la détermination de cette position commune fait l'objet d'un tirage au hasard. Dans ces conditions, cet échantillonnage peut s'analyser comme un tirage contagieux d'un chapelet de boules dans une urne unique, un seul choix déterminant tous les autres. L'erreur à craindre ne peut être estimée à partir d'un échantillon unique.

L. BRENAC.