LES DEVELOPPEMENTS METHODOLOGIQUES REALISES DANS LE CADRE DE L'ENQUÊTE TERUTI–LUCAS

Serge AMORICH, Philippe MICHEL et Benoit MIROUSE (*), Amandine MARY(**)

(*) Service de la Statistique et de la Prospective - MAAPRAT
(**) Université Toulouse I et Université Toulouse III

Introduction

L'enquête TERUTI–LUCAS est une enquête qui se déroule annuellement (exception faite pour l'année 2011 où elle n'a pas eu lieu pour des raisons budgétaires).

L'enquête TERUTI-LUCAS fournit une information régulière sur l'occupation et l'utilisation du territoire en France, elle donne par exemple des estimations sur les superficies des principales cultures annuelles et constitue une base de sondage pour d'autres enquêtes.

La note rédigée est le résultat de travaux multiples qui se sont étalés dans le temps en fonction des besoins du Service de la Statistique et de la Prospective (SSP). Deux personnes doivent être citées : Philippe FOURNIER et Amandine MARY.

Philippe FOURNIER est celui qui a élaboré et mis en place l'enquête TERUTI avant que celle-ci ne soit remplacée par l'enquête TERUTI-LUCAS. Les premières formules ont été élaborées par Philippe FOURNIER avec l'aide de François TURLOT sur la base des enseignements de Jacques DESABIE.

Amandine MARY, étudiante en Magistère à Toulouse a eu pour mission de reprendre les divers écrits (ceux de Philippe MICHEL et de Serge AMORICH) et de réaliser un document unique. Il est évident, mais il convient de le rappeler, que les réponses apportées par nos collègues de l'INSEE à nos questions (nous pensons à Jean-Claude DEVILLE, Guillaume CHAUVET et Daniel BONNERY) ont permis de finaliser les formules que le Service de la Statistique et de la Prospective (SSP) utilise aujourd'hui dans ses publications.

Ce rappel étant fait, il convient maintenant d'expliquer les modalités de construction de l'échantillon et de présenter les formules utilisées dans le calcul des estimateurs et des variances estimées. Une description des programmes de calcul élaborés par Benoît MIROUSE sera ensuite effectuée, elle permettra de montrer les diverses possibilités offertes à l'utilisateur. Enfin, un exemple dans l'utilisation des résultats obtenus sera effectué afin de mieux appréhender les interrogations et les chantiers nouveaux qui ne manqueront pas d'apparaître dans les mois à venir.

1. La construction de l'échantillon TERUTI–LUCAS

L'enquête TERUTI–LUCAS est une enquête par sondage aréolaire à deux niveaux de tirage, tout d'abord tirage des segments (un segment se définit comme une portion de territoire dans lequel sont regroupés les points observés), puis des points à l'intérieur du segment échantillonné (un point se définit comme une portion de territoire élémentaire, celle qui est observée par l'enquêteur).

Une grille systématique, régulière et carrée orientée Nord – Sud et Est – Ouest de maille 3 km sur 3 km sert de base au tirage d'échantillon. Chaque intersection détermine un segment, en fait le point 11 du dit-segment. Chaque segment est identifié par sa position en ligne et colonne (LLL-CCC) dans le repère normé dont l'origine est à l'extrême sud-ouest du territoire national.

Le tirage d'échantillon a été effectué selon les prescriptions d'EUROSTAT sur la base de la projection UTM (ellipsoïde GRS80) qui entraîne le moins de déformation, en France métropolitaine UTM31N, aux Antilles UTM20N, en Guyane UTM22N et à la Réunion UTM40S. Il est systématique et sans stratification préalable. En Martinique, Guadeloupe et Réunion la grille de base est de 2 km sur 2 km afin d'obtenir la précision voulue pour chacun des départements pris individuellement.

En France métropolitaine, les coordonnées ont été transformées en Lambert II étendu (Lambert IV en Corse) pour permettre la superposition aux supports IGN (Scan 25 et BD Ortho).

1.1. Les segments

Le segment est le premier niveau de tirage. Il s'agit d'une portion de territoire dont la taille peut varier de 1 500 mètres sur 600 mètres (si 10 points enquêtés, soit 90 hectares) à 1 500 mètres sur 1 500 mètres (si 25 points enquêtés, soit 225 hectares).

L'échantillon de segments a été découpé en 4 sous-échantillons inclus les uns dans les autres :
- "4" est l'échantillon déterminé par la grille de base la plus fine (3 ou 2 km)
- "1" est la partie de "4" correspondant à la grille LUCAS de 18 km
- "2" est la partie de "4", complément de "1" pour constituer la grille 6 km
- "3", appelé "6 km densifié", correspond au doublement du nombre de segments de "2" + "1" en quinconces.

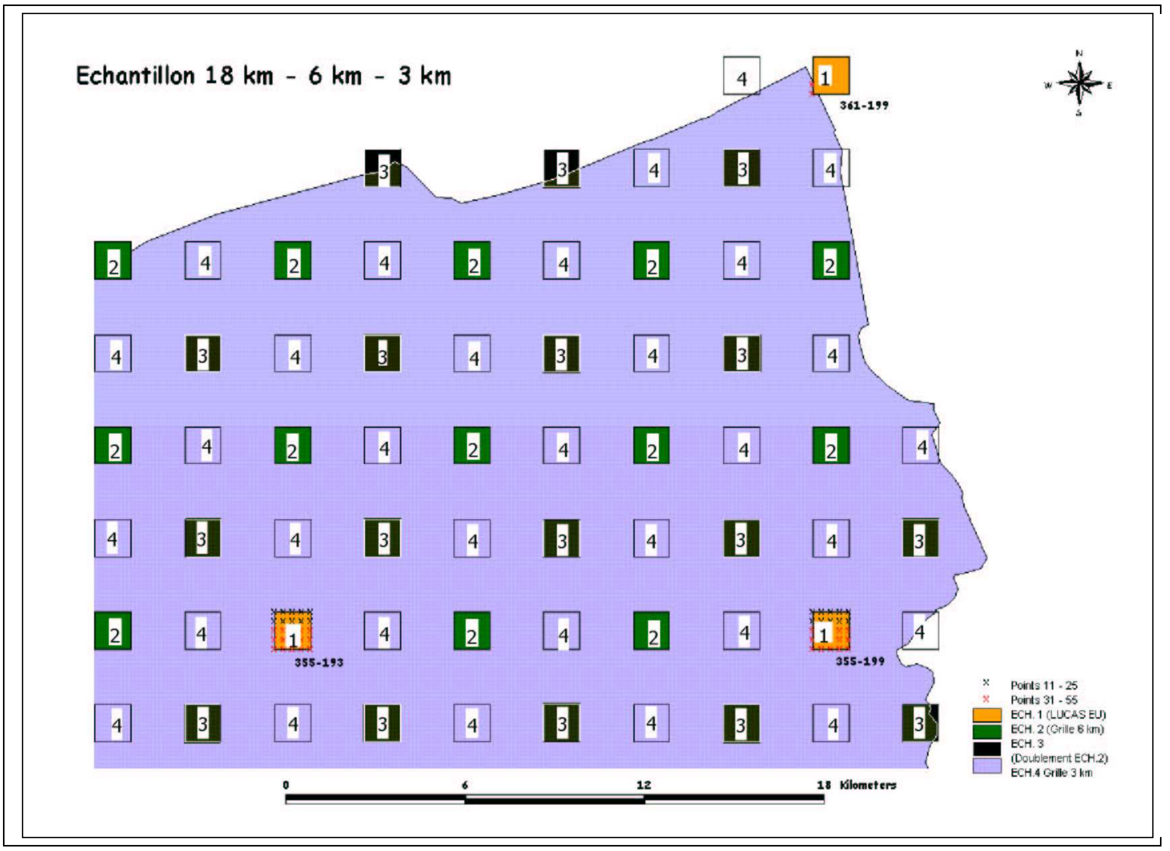
Le sous-échantillon "3" n'existe pas pour les Antilles et pour la Réunion.

Chaque segment se voit affecter le code de l'échantillon le plus "petit" auquel il appartient. Si par exemple on décide de lancer l'enquête sur le sous-échantillon "2" on devra sélectionner les segments affectés de "1" et "2", pour le sous-échantillon "3" les segments affectés de "1", "2" ou "3" et pour le sous-échantillon "4" la totalité des segments.

La ventilation des segments par sous-échantillon est la suivante :

Sous-échantillon	1 (18 km)	1+2 (6 km)	1+2+3 (6 km Doublé)	1+2+3+4 (maître 3 km)
France	1 764	15 896	31 655	64 492

La représentation graphique ci-dessous précise le positionnement des segments selon leur appartenance à tel ou tel sous-échantillon pour une zone géographique donnée.

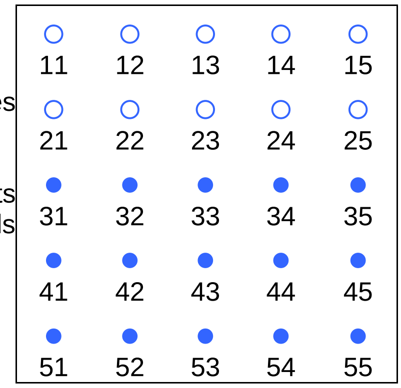


L'enquête LUCAS (version française) n'a rien à voir avec l'enquête LUCAS effectuée par EUROSTAT (enquête basée sur un tirage à un seul degré de points). C'est pour cette raison que nous appelons l'enquête LUCAS (version française) enquête TERUTI-LUCAS pour bien la distinguer de l'enquête LUCAS.

1.2. Les points

Chaque segment contient 5 lignes de 5 points orientés comme les segments Nord- Sud, Est – Ouest et espacés régulièrement de 300 m.

Les 2 premières lignes (points 11 à 25) correspondent aux points LUCAS, les 3 suivantes ont été ajoutées pour répondre à d'éventuels besoins qui pourraient apparaître.



Tout segment enquêté a obligatoirement ses deux premières lignes enquêtées.

1.3. L'échantillon national

L'objectif de l'enquête nationale annuelle est de visiter les segments de l'échantillon "6 km doublé", c'est à dire les sous-échantillons 1 + 2 + 3 pour l'ensemble des départements autres que le 75, 90, 92, 93, 94, 9A, 9B, 9D et les sous-échantillons 1 + 2 + 3 + 4 pour les départements 75, 90, 92, 93, 94, 9A, 9B et 9D, soit 33 071 segments de 10 points (11 à 25).

1.4. Les extensions locales

Une des caractéristiques essentielles de l'enquête TERUTI-LUCAS est sa modularité. L'échantillon peut être utilisé comme base de sondage pour organiser une enquête disjointe du passage annuel. Il permet aussi d'envisager des renforcements locaux ou nationaux.

Les renforcements peuvent prendre deux formes utilisables séparément ou conjointement :
- augmenter le nombre de segments,
- augmenter le nombre de points visités dans les segments.

Cette possibilité permet de répondre à des besoins ciblés émanant de demandeurs nationaux ou locaux. La zone concernée par une extension peut être déterminée en fonction des besoins du demandeur sans être limité par les frontières administratives. La seule contrainte est de sélectionner un nombre de points suffisant pour obtenir des résultats fiables.

2. Les formules de calcul

2.1. Pour une année donnée[1]

Le tirage de l'échantillon pour une zone géographique donnée[2] est un tirage à deux degrés :

- 1er degré de tirage : tirage sans remise et systématique, à probabilité égale de segments constituant les unités primaires

- 2ème degré de tirage : tirage sans remise et systématique, équiprobable à l'intérieur d'un même segment de n_i points dans chaque unité primaire (n_i pouvant variant de 1 à 10). Dans la majorité des cas, n_i sera égal à 10. Chacun des points constitue une unité secondaire.

2.1.1. Estimation de la superficie en catégorie de territoire h

La superficie en catégorie de territoire h pour une zone géographique donnée est égale à :

$$T_h = A \times \frac{\sum_{i \in U} \sum_{j \in U_i} Y_{i,j}}{\sum_{i \in U} \sum_{j \in U_i} X_{i,j}} = A \times \frac{Y}{X} = A \times R = A \times \frac{\sum_{i \in U} N_{ih}}{N} = A \times \frac{N_h}{N}$$

avec

T_h	superficie en catégorie de territoire h pour la zone géographique considérée
A	superficie de la zone géographique considérée
$Y_{i,j}$	vaut 1 si le point j du segment i est en catégorie du territoire h, 0 dans le cas contraire
$X_{i,j}$	vaut 1 quelle que soit la catégorie du territoire du point j du segment i
U	liste des segments dans la zone géographique considérée
U_i	liste des points du segment i

[1] L'élaboration des formules initiales de calcul pour une année donnée résulte des travaux de Philippe FOURNIER en collaboration avec François TURLOT. Ces travaux sont présentés dans le document suivant : Philippe FOURNIER, "Etude sur l'utilisation du territoire - Méthodologie - Résultats 1969-1970-1971", Supplément "Série Études", n°104, novembre 1972, 112 p.

[2] Par zone géographique donnée, il faut entendre une zone géographique au sens commun du terme, pour laquelle est appliquée un taux de sondage au 1er degré uniforme (référence au même sous-échantillon, à savoir le 3 ou le 4) ainsi qu'un taux de sondage au 2ème degré uniforme (référence au même nombre de lignes enquêtées, à savoir les 2 premières lignes ou les 5 lignes). Quatre situations sont théoriquement possibles.

N_{ih}	nombre de points ayant la catégorie de territoire h dans le segment i
N_h	nombre de points ayant la catégorie de territoire h pour l'ensemble des segments de la zone géographique considérée
N	nombre de points pour l'ensemble des segments de la zone géographique considérée

Dans ces conditions, un estimateur naturel de T_h est :

$$\boxed{\hat{T}_h = A \times \frac{\frac{M}{m}\sum_{i\in S}\frac{N_i}{n_i}\sum_{j\in S_i}Y_{i,j}}{\frac{M}{m}\sum_{i\in S}\frac{N_i}{n_i}\sum_{j\in S_i}X_{i,j}}}$$

avec

$\hat{T}_h$	estimateur de la superficie en catégorie de territoire h pour la zone géographique considérée
M	nombre total de segments dans la zone géographique considérée : comme la superficie d'un segment est égal à 90 hectares, M est obtenu en divisant A par 90
m	nombre de segments enquêtés dans la zone géographique considérée
S	liste des segments enquêtés dans la zone géographique considérée
S_i	liste des points enquêtés dans le segment i
N_i	nombre total de points dans le segment i : comme la superficie d'un point est égale à 7 mètres carrés, N_i est obtenu en divisant 90 hectares par 7 mètres carrés
n_i	nombre de points enquêtés dans le segment i (en général, le nombre de points enquêtés est égal à 10)

En faisant l'hypothèse que le nombre de points enquêtés dans un segment i, à savoir n_i , est le même pour chaque segment (10 dans le cas présent) et en remarquant que le nombre de points N_i dans un segment est le même pour chaque segment (chaque segment a la même superficie, à savoir 90 hectares), il est immédiat de constater que le rapport $\frac{N_i}{n_i}$ est constant.

Sous cette hypothèse, l'expression de l'estimateur se simplifie et devient :

$$\boxed{\hat{T}_h = A \times \frac{\sum_{i\in S}\sum_{j\in S_i}Y_{i,j}}{\sum_{i\in S}\sum_{j\in S_i}X_{i,j}} = A\times\hat{R} = A\times\frac{\hat{Y}}{\hat{X}} = A\times\frac{\sum_{i\in S}n_{ih}}{\sum_{i\in S}n_i} = A\times\frac{n_h}{n}}$$

avec

n_{ih}	nombre de points enquêtés dans le segment i ayant la catégorie de territoire h
n_i	nombre de points enquêtés dans le segment i (égal à 10)
n	nombre total de points enquêtés dans la zone géographique considérée (égal à $10\times m$)

2.1.2. Variance estimée de l'estimateur de la superficie

L'estimateur $\hat{T}_h$ est défini comme le produit $A\times\hat{R}$. Les termes figurant au numérateur et au dénominateur du ratio sont aléatoires.

Dans ces conditions, l'estimateur est biaisé et le calcul de la précision estimée de l'estimateur est obtenu à partir du calcul de l'écart quadratique moyen (EQM).

2.1.2.1. Le principe de la linéarisation[3]

Afin d'obtenir l'expression de l'EQM, il convient de linéariser l'expression du ratio de manière à ce que celui-ci se ramène à un total dont il suffit de prendre la variance selon le plan de sondage considéré.

Soit θ une fonction de différentes variables $\theta = f\left(Y_1, Y_2, \ldots, Y_i, \ldots, Y_p\right)$.

Un estimateur de θ est $\hat{\theta} = f(\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_i, \ldots \hat{Y}_p)$

En utilisant un développement limité au voisinage de $(Y_1, Y_2, \ldots, Y_i, \ldots, Y_p)$, nous pouvons écrire :

$\hat{\theta} = \theta + \sum_{i=1}^{p} \frac{\partial f}{\partial Y_i}\left(Y_1, Y_2, \ldots, Y_p\right)\left(\hat{Y}_i - Y_i\right) + R$ où R désigne le reste

Cette expression peut s'écrire $\hat{\theta} = \theta + \sum_{i=1}^{p} a_i\left(\hat{Y}_i - Y_i\right) + R$ en posant $a_i = \frac{\partial f}{\partial Y_i}\left(Y_1, Y_2, \ldots, Y_i, \ldots, Y_p\right)$

En prenant le pseudo-estimateur $\hat{\theta}_0 = \theta + \sum_{i=1}^{p} a_i\left(\hat{Y}_i - Y_i\right)$, il est immédiat que $E\left(\hat{\theta}_0\right) = \theta$

Si R est petit alors $\hat{\theta} \cong \hat{\theta}_0$

En conséquence, $\hat{\theta} \cong \theta + \sum_{i=1}^{p} a_i\left(\hat{Y}_i - Y_i\right) = \theta + \sum_{i=1}^{p} a_i \hat{Y}_i - \sum_{i=1}^{p} a_i Y_i$

$$EQM\left(\hat{\theta}_0\right) = E\left[\left(\hat{\theta}_0 - \theta\right)^2\right] = E\left[\left(\hat{\theta}_0 - E\left(\hat{\theta}_0\right)\right)^2\right] = V\left(\hat{\theta}_0\right)$$

$$EQM\left(\hat{\theta}\right) \cong V\left(\hat{\theta}_0\right)$$

En posant $\hat{Y}_i = \sum_{k \in S} \frac{Y_{i,k}}{\pi_k}$, nous pouvons écrire $\sum_{i=1}^{p} a_i \hat{Y}_i = \sum_{i=1}^{p} a_i \sum_{k \in S} \frac{Y_{i,k}}{\pi_k} = \sum_{k \in S} \frac{Z_k}{\pi_k}$ avec $Z_k = \sum_{i=1}^{p} a_i Y_{i,k}$

En conséquence, $V\left(\hat{\theta}_0\right) = V\left(\hat{Z}\right)$ avec $\hat{Z} = \sum_{k \in S} \frac{Z_k}{\pi_k}$

$$EQM\left(\hat{\theta}\right) \cong EQM\left(\hat{\theta}_0\right) = V\left(\hat{\theta}_0\right) = V\left(\hat{Z}\right)$$

2.1.2.2. Application au ratio

$R = f\left(Y_1, Y_2\right) = \frac{Y_1}{Y_2}$ avec $Y_1 = Y$ et $Y_2 = X$

D'où $Z_k = \frac{\partial f}{\partial Y_1} Y_{1,k} + \frac{\partial f}{\partial Y_2} Y_{2,k} = \frac{1}{Y_2} Y_{1,k} - \frac{Y_1}{Y_2^2} Y_{2,k}$

Par conséquent

$$\hat{Z} = \sum_{k \in S} \frac{Z_k}{\pi_k} = \frac{1}{Y_2} \sum_{k \in S} \frac{Y_{1,k}}{\pi_k} - \frac{Y_1}{Y_2^2} \sum_{k \in S} \frac{Y_{2,k}}{\pi_k} = \frac{1}{Y_2}\left(\sum_{k \in S} \frac{Y_{1,k}}{\pi_k} - \frac{Y_1}{Y_2} \sum_{k \in S} \frac{Y_{2,k}}{\pi_k}\right) = \frac{1}{Y_2}\left(\sum_{k \in S} \frac{Y_{1,k}}{\pi_k} - R \sum_{k \in S} \frac{Y_{2,k}}{\pi_k}\right)$$

$$\hat{Z} = \frac{1}{Y_2} \sum_{k \in S} \frac{\left(Y_{1,k} - R Y_{2,k}\right)}{\pi_k} = \frac{1}{Y_2}\left(\hat{Y}_1 - R \hat{Y}_2\right)$$

Nous avons donc : $\hat{Z} = \frac{1}{X} \hat{Y} - \frac{Y}{X^2} \hat{X}$

---

[3] Nous renvoyons le lecteur à l'ouvrage de Pascal ARDILLY :
P.ARDILLY, Les techniques de sondage, éditions Technip, Paris, 2006, p.584 - 591

$$EQM\left(\hat{R}\right) \cong V\left(\hat{Z}\right) \text{ avec } \hat{Z}=\frac{1}{X}\left(\hat{Y}-R\hat{X}\right)$$

$$V\left(\hat{Z}\right)=V\left[\frac{1}{X}\left(\hat{Y}-R\hat{X}\right)\right]=\frac{1}{X^2}V\left(\hat{Y}-R\hat{X}\right)=\frac{1}{X^2}V\left(\hat{W}\right) \text{ avec } \hat{W}=\hat{Y}-R\hat{X}$$

2.1.2.3. La prise en compte du plan de sondage

La variance vraie de l'estimateur linéarisé du ratio s'écrit :

$$V\left(\hat{Z}\right)=\frac{1}{X^2}\left[M^2\left(1-\frac{m}{M}\right)\frac{S_w^2}{m}+\frac{M}{m}\sum_{i\in U}N_i^2\left(1-\frac{n_i}{N_i}\right)\frac{S_{wi}^2}{n_i}\right]$$

où $S_w^2=\frac{\sum_{i\in U}\left(W_i-\overline{W}\right)^2}{M-1}$ avec $\overline{W}=\frac{\sum_{i\in U}W_i}{M}$

et $S_{wi}^2=\frac{\sum_{j\in U_i}\left(W_{i,j}-\overline{W_i}\right)^2}{N_i-1}$ avec $\overline{W_i}=\frac{\sum_{j\in U_i}W_{i,j}}{N_i}$

<u>Remarque</u> : écrire $\sum_{i\in U}$ revient à écrire $\sum_{i=1}^{M}$ comme écrire $\sum_{i\in U_i}$ revient à écrire $\sum_{j=1}^{N_i}$

La variance estimée de l'estimateur du ratio linéarisé se définit à partir de la formule suivante :

$$\boxed{\hat{V}\left(\hat{Z}\right)=\frac{1}{\hat{X}^2}\left[M^2\left(1-\frac{m}{M}\right)\frac{s_w^2}{m}+\frac{M}{m}\sum_{i\in S}N_i^2\left(1-\frac{n_i}{N_i}\right)\frac{s_{wi}^2}{n_i}\right]}$$

où $s_w^2=\hat{S}_w^2$ et $s_{wi}^2=\hat{S}_{wi}^2$

<u>Rappel</u> :

- s_w^2 et s_{wi}^2 doivent être construits en tenant compte du caractère systématique du tirage (tant au premier qu'au second degré de tirage)

- $W_{i,j}=Y_{i,j}-RX_{i,j}$ est estimé par $\hat{W}_{i,j}=Y_{i,j}-\hat{R}X_{i,j}$ avec $\hat{R}=\frac{\sum_{i\in S}\sum_{j\in S}Y_{i,j}}{\sum_{i\in S}\sum_{j\in S}X_{i,j}}$

- $W_i=\sum_{j\in U_i}W_{i,j}$ est estimé par $\hat{W}_i=\frac{N_i}{n_i}\sum_{j\in S_i}\hat{W}_{i,j}$

- comme , $X=\sum_{i\in U}\sum_{j\in U_i}X_{i,j}$, nous avons $\hat{X}=\frac{M}{m}\sum_{i\in S}\frac{N_i}{n_i}\sum_{j\in S_i}X_{i,j}$

2.1.2.4. Le calcul de s_w^2

$$s_w^2=\hat{S}_w^2=\frac{\sum_{i\in S}\left(\hat{W}_i-\hat{\overline{W}}\right)^2}{m-1}=\frac{\sum_{i\in S}\left(\hat{W}_i^2-2\hat{W}_i\hat{\overline{W}}+\hat{\overline{W}}^2\right)}{m-1}=\frac{\sum_{i\in S}\hat{W}_i^2-2\hat{\overline{W}}\sum_{i\in S}\hat{W}_i+m\hat{\overline{W}}^2}{m-1}$$

$$s_w^2=\frac{\sum_{i\in S}\hat{W}_i^2-2\hat{\overline{W}}m\hat{\overline{W}}+m\hat{\overline{W}}^2}{m-1}=\frac{\sum_{i\in S}\hat{W}_i^2-m\hat{\overline{W}}^2}{m-1}$$

Or $\hat{\overline{W}}=\frac{\sum_{i\in S}\hat{W}_i}{m}=\frac{\sum_{i\in S}\frac{N_i}{n_i}\sum_{j\in S_i}\hat{W}_{i,j}}{m}=\frac{1}{m}\frac{N_i}{n_i}\sum_{i\in S}\sum_{j\in S_i}\hat{W}_{i,j}=\frac{1}{m}\frac{N_i}{n_i}\sum_{i\in S}\sum_{j\in S_i}\left(Y_{i,j}-\hat{R}X_{i,j}\right)$

car $\frac{N_i}{n_i}$ est constant quel que soit i

$$\hat{\overline{W}}=\frac{1}{m}\frac{N_i}{n_i}\left[\sum_{i\in S}\sum_{j\in S_i}Y_{i,j}-\hat{R}\sum_{i\in S}\sum_{j\in S_i}X_{i,j}\right]=\frac{1}{m}\frac{N_i}{n_i}\left[\sum_{i\in S}\sum_{j\in S_i}Y_{i,j}-\frac{\left[\sum_{i\in S}\sum_{j\in S_i}Y_{i,j}\right]}{\left[\sum_{i\in S}\sum_{j\in S_i}X_{i,j}\right]}\sum_{i\in S}\sum_{j\in S_i}X_{i,j}\right]=0$$

En conséquence, $s_{\bar{w}}^2=\frac{\sum_{i\in S}\hat{W}_i^2}{m-1}$

Le fait que le tirage soit systématique, nous pouvons écrire[4] :

$$s_{\bar{w}}^2=\frac{\sum_{i\in S}\hat{W}_i^2}{m-1}=\frac{\sum_{i=1}^{nc-1}\left(\hat{W}_{i+1}-\hat{W}_i\right)^2}{2nc}$$

avec nc le nombre de comparaisons effectuées entre deux segments voisins tant en ligne qu'en colonne pour l'ensemble des segments échantillonnés dans la zone géographique considérée[5].

Comme $\hat{W}_i=\frac{N_i}{n_i}\sum_{j\in S_i}\hat{W}_{i,j}=k\sum_{j\in S_i}\hat{W}_{i,j}$, nous pouvons écrire que $s_{\bar{w}}^2=\frac{k^2}{2nc}\sum_{i=1}^{nc-1}\left(\sum_{j\in S_{i+1}}\hat{W}_{i+1,j}-\sum_{j\in S_i}\hat{W}_{i,j}\right)^2$

En conséquence, $M^2(1-\frac{m}{M})\frac{1}{m}s_{\bar{w}}^2=\frac{M}{m}k^2(M-m)\frac{\sum_{i=1}^{nc-1}\left(\sum_{j\in S_{i+1}}\hat{W}_{i+1,j}-\sum_{j\in S_i}\hat{W}_{i,j}\right)^2}{2nc}$

2.1.2.5. Le calcul de s_{wi}^2

$$s_{wi}^2=\hat{S}_{wi}^2=\frac{\sum_{j\in S_i}\left(\hat{W}_{i,j}-\hat{\overline{W}}_i\right)^2}{n_i-1}=\frac{\sum_{j\in S_i}\left(\hat{W}_{i,j}^2-2\hat{W}_{i,j}\hat{\overline{W}}_i+\hat{\overline{W}}_i^2\right)}{n_i-1}=\frac{\sum_{j\in S_i}\hat{W}_{i,j}^2-2\hat{\overline{W}}_i\sum_{j\in S_i}\hat{W}_{i,j}+n_i\hat{\overline{W}}_i^2}{n_i-1}$$

$$s_{wi}^2=\frac{1}{n_i-1}\left[\sum_{j\in S_i}\hat{W}_{i,j}^2-2n_i\hat{\overline{W}}_i^2+n_i\hat{\overline{W}}_i^2\right]=\frac{1}{n_i-1}\left[\sum_{j\in S_i}\hat{W}_{i,j}^2-n_i\hat{\overline{W}}_i^2\right]=\frac{1}{n_i-1}\left[\sum_{j\in S_i}\hat{W}_{i,j}^2-n_i\left(\frac{\sum_{j\in S_i}\hat{W}_{i,j}}{n_i}\right)^2\right]$$

$$s_{wi}^2=\frac{1}{n_i-1}\left[\sum_{j\in S_i}\hat{W}_{i,j}^2-\frac{\left(\sum_{j\in S_i}\hat{W}_{i,j}\right)^2}{n_i}\right]=\frac{1}{n_i-1}\left[\frac{n_i\sum_{j\in S_i}\hat{W}_{i,j}^2-\left(\sum_{j\in S_i}\hat{W}_{i,j}\right)^2}{n_i}\right]$$

Le fait que le tirage soit systématique, nous pouvons écrire :

---

[4] Nous renvoyons le lecteur à l'ouvrage de Pascal ARDILLY qui aborde la problématique du tirage systématique. Pascal ARDILLY, Les techniques de sondage, éditions Technip, Paris, 2006, p.79-90

[5] Un exemple de calcul est présenté dans l'annexe 01 de la présente note qui reprend l'approche présentée par Amandine MARY dans son rapport de stage et qui précise la notion de voisinage des segments.
Amandine MARY, Eléments pour une analyse de l'occupation du territoire à partir des enquêtes TERUTI et TERUTI-LUCAS, Service la Statistique et de la Prospective (SSP) - Université Toulouse I - Université Toulouse III, rapport de stage, avril-août 2009, p.28-32

$$s_{wi}^2=\frac{1}{n_i-1}\left[\frac{n_i\sum_{j\in S_i}\hat{W}_{i,j}^2-\left(\sum_{j\in S_i}\hat{W}_{i,j}\right)^2}{n_i}\right]=\frac{\sum_{j=1}^{nc_i-1}\left(\hat{W}_{i,j+1}-\hat{W}_{i,j}\right)^2}{2nc_i}$$

avec nc_i le nombre de comparaisons effectuées dans le segment i considéré entre deux points voisins tant en ligne qu'en colonne[6].

De plus

$$\frac{M}{m}\sum_{i\in S}N_i^2\left(1-\frac{n_i}{N_i}\right)\frac{s_{wi}^2}{n_i}=\frac{M}{m}\sum_{i\in S}\left[\left(\frac{N_i}{n_i}\right)^2 n_i\left(1-\frac{n_i}{N_i}\right)s_{wi}^2\right]=\frac{M}{m}\sum_{i\in S}\left[k^2n_i\left(1-\frac{1}{k}\right)s_{wi}^2\right]=\frac{M}{m}\sum_{i\in S}\left[k^2n_i\left(\frac{k-1}{k}\right)s_{wi}^2\right]$$

Comme $\frac{k-1}{k}\cong 1$, nous pouvons écrire que ;

$$\frac{M}{m}\sum_{i\in S}\left[k^2n_i\left(\frac{k-1}{k}\right)s_{wi}^2\right]=\frac{M}{m}k^2\sum_{i\in S}n_is_{wi}^2$$

En conséquence, $\frac{M}{m}\sum_{i\in S}N_i^2\left(1-\frac{n_i}{N_i}\right)\frac{s_{wi}^2}{n_i}=\frac{M}{m}k^2\sum_{i\in S}n_i\frac{\sum_{j=1}^{nc_i-1}\left(\hat{W}_{i,j+1}-\hat{W}_{i,j}\right)^2}{2nc_i}$

2.1.2.6. La formule finale de la variance estimée

La variance estimée de l'estimateur du ratio linéarisé étant défini à partir de la formule initiale suivante :

$$\hat{V}\left(\hat{Z}\right)=\frac{1}{\hat{X}^2}\left[M^2\left(1-\frac{m}{M}\right)\frac{s_W^2}{m}+\frac{M}{m}\sum_{i\in S}N_i^2\left(1-\frac{n_i}{N_i}\right)\frac{s_{wi}^2}{n_i}\right]$$

En tenant compte des résultats obtenus précédemment, nous pouvons écrire que :

$$\hat{V}\left(\hat{Z}\right)=\frac{1}{\hat{X}^2}\frac{M}{m}k^2\left[\frac{M-m}{2nc}\sum_{i=1}^{nc-1}\left(\sum_{j\in S_{i+1}}\hat{W}_{i+1,j}-\sum_{j\in S_i}\hat{W}_{i,j}\right)^2+\sum_{i\in S}\frac{n_i}{2nc}\sum_{j=1}^{nc_i-1}\left(\hat{W}_{i,j+1}-\hat{W}_{i,j}\right)^2\right]$$

Comme $\hat{X}=\frac{M}{m}\sum_{i\in S}\frac{N_i}{n_i}\sum_{j\in S_i}X_{i,j}=\frac{M}{m}k\sum_{i\in S}\sum_{j\in S_i}X_{i,j}=\frac{M}{m}kn$ avec $n=\sum_{i\in S}n_i$ et $\hat{X}^2=\left(\frac{M}{m}kn\right)^2$

nous avons alors $\frac{1}{\hat{X}^2}\frac{M}{m}k^2=\frac{m}{Mn^2}$

De plus $\hat{T}_h=A\hat{R}$ entraîne $\hat{V}\left(\hat{T}_h\right)=A^2\hat{V}\left(\hat{R}\right)=A^2\hat{V}\left(\hat{Z}\right)$ et comme $M=\frac{A}{90}$ nous pouvons écrire que $A^2\frac{1}{\hat{X}^2}\frac{M}{m}k^2=A^2\frac{m}{Mn^2}=\frac{90mA}{n^2}$

Finalement, la variance estimée de l'estimateur de la superficie en catégorie de territoire h pour une zone géographique donnée est définie par la formule suivante :

---

[6] Un exemple de calcul est présenté dans l'annexe 02 de la présente note qui reprend l'approche présentée par Amandine MARY dans son rapport de stage et qui précise la notion de voisinage entre points.
Amandine MARY, <u>Eléments pour une analyse de l'occupation du territoire à partir des enquêtes TERUTI et TERUTI-LUCAS</u>, Service la Statistique et de la Prospective (SSP) - Université Toulouse I - Université Toulouse III, rapport de stage, avril-août 2009, p.34-36

$$\hat{V}\left(\hat{T}_h\right)=\frac{90mA}{n^2}\left[\frac{M-m}{2nc}\sum_{i=1}^{nc-1}\left(\sum_{j\in S_{i+1}}\hat{W}_{i+1,j}-\sum_{j\in S_i}\hat{W}_{i,j}\right)^2+\sum_{i\in S}\frac{n_i}{2nc_i}\sum_{j=1}^{nc_i-1}\left(\hat{W}_{i,j+1}-\hat{W}_{i,j}\right)^2\right]$$

Si nous remplaçons $\hat{W}_{i,j}$ par l'expression $Y_{i,j}-\hat{R}X_{i,j}$, la variance estimée s'écrit alors de la manière suivante :

$$\hat{V}\left(\hat{T}_h\right)=\frac{90mA}{n^2}\left[\frac{M-m}{2nc}\sum_{i=1}^{nc-1}\left(\sum_{j\in S_{i+1}}\left(Y_{i+1,j}-\hat{R}X_{i+1,j}\right)-\sum_{j\in S_i}\left(Y_{i,j}-\hat{R}X_{i,j}\right)\right)^2+\sum_{i\in S}\frac{n_i}{2nc_i}\sum_{j=1}^{nc_i-1}\left(\left(Y_{i,j+1}-\hat{R}X_{i,j+1}\right)-\left(Y_{i,j}-\hat{R}X_{i,j}\right)\right)^2\right]$$

Il est possible de décomposer cette variance estimée en deux composantes : une composante résultant du premier degré du tirage dénommée COMPOSANTE INTER et une composante résultant du second degré du tirage dénommée COMPOSANTE INTRA.

La composante inter est égale à $\frac{90mA}{n^2}\left[\frac{M-m}{2nc}\sum_{i=1}^{nc-1}\left(\sum_{j\in S_{i+1}}\hat{W}_{i+1,j}-\sum_{j\in S_i}\hat{W}_{i,j}\right)^2\right]$

La composante intra est égale à $\frac{90mA}{n^2}\left[\sum_{i\in S}\frac{n_i}{2nc_i}\sum_{j=1}^{nc_i-1}\left(\hat{W}_{i,j+1}-\hat{W}_{i,j}\right)^2\right]$

Une telle approche s'inscrit dans la continuité de l'approche développée par Philippe FOURNIER.

2.2. Au cours d'une période[7]

Nous nous intéressons maintenant à l'évolution de la superficie d'une catégorie de territoire h dans une zone géographique donnée au cours d'une période P (il y a donc un début de période, l'année t_0 et une fin de période, l'année t_1). Les deux années peuvent être consécutives. Cependant, les évolutions concernant l'occupation du territoire étant faibles (pour ne pas dire très faibles) sur le court terme, il semble évident qu'étudier l'évolution n'a de sens que si nous raisonnons en moyenne ou longue période (des analyses basées sur des périodes au moins égales à quatre ans nous semblent justifiées).

2.2.1. Estimation de l'évolution de la superficie en catégorie de territoire h au cours d'une période

A partir des formules précédentes, nous avons deux estimateurs :

- un estimateur de la superficie en catégorie de territoire h pour la zone géographique considérée en début de période t_0 : $\hat{T}_h^{t_0}=A\frac{\sum_{i\in S}\sum_{j\in S_i}Y_{i,j}^{t_0}}{\sum_{i\in S}\sum_{j\in S_i}X_{i,j}^{t_0}}=A\frac{n_h^{t_0}}{n^{t_0}}$

- un estimateur de la superficie en catégorie de territoire h pour la zone géographique considérée en fin de période t_1 : $\hat{T}_h^{t_1}=A\frac{\sum_{i\in S}\sum_{j\in S_i}Y_{i,j}^{t_1}}{\sum_{i\in S}\sum_{j\in S_i}X_{i,j}^{t_1}}=A\frac{n_h^{t_1}}{n^{t_1}}$

avec

$Y_{i,j}^{t_0}$	vaut 1 si le point j du segment i est en catégorie de territoire h à la date t_0, 0 dans le cas contraire

[7] C'est la partie innovante par rapport aux travaux précurseurs de Philippe FOURNIER.

$X_{i,j}^{t_0}$	vaut 1 quelle que soit la catégorie de territoire du point j du segment i à la date t_0
$n_h^{t_0}$	nombre de points enquêtés dans la zone géographique considérée ayant la catégorie de territoire h en t_0
n^{t_0}	nombre total de points enquêtés dans la zone géographique considérée en t_0
$Y_{i,j}^{t_1}$	vaut 1 si le point j du segment i est en catégorie de territoire h à la date t_1, 0 dans le cas contraire
$X_{i,j}^{t_1}$	vaut 1 quelle que soit la catégorie de territoire du point j du segment i à la date t_1
$n_h^{t_1}$	nombre de points enquêtés dans la zone géographique considérée ayant la catégorie de territoire h en t_1
n^{t_1}	nombre total de points enquêtés dans la zone géographique considérée en t_1

L'estimateur de l'évolution de la superficie en catégorie de territoire h pour la zone géographique considérée au cours de la période P est :

$$\hat{T}_h^{t_1}-\hat{T}_h^{t_0}=\hat{\Delta}_h^P=A\left[\frac{\sum_{i\in S}\sum_{j\in S_i}Y_{i,j}^{t_1}}{\sum_{i\in S}\sum_{j\in S_i}X_{i,j}^{t_1}}-\frac{\sum_{i\in S}\sum_{j\in S_i}Y_{i,j}^{t_0}}{\sum_{i\in S}\sum_{j\in S_i}X_{i,j}^{t_0}}\right]=A\left[\frac{n_h^{t_1}}{n^{t1}}-\frac{n_h^{t_0}}{n^{t0}}\right]$$

Comme le nombre total de points enquêtés en t_0 et en t_1 est le même, l'estimateur de l'évolution de la superficie en catégorie de territoire h pour la zone géographique considérée au cours de la période P est :

$$\hat{T}_h^{t_1}-\hat{T}_h^{t_0}=\hat{\Delta}_h^P=A\left[\frac{\sum_{i\in S}\sum_{j\in S_i}Y_{i,j}^{t_1}-\sum_{i\in S}\sum_{j\in S_i}Y_{i,j}^{t_0}}{\sum_{i\in S}\sum_{j\in S_i}X_{i,j}}\right]=A\left[\frac{n_h^{t_1}-n_h^{t_0}}{n}\right]$$

2.2.2. Variance estimée de l'évolution de la superficie en catégorie de territoire h au cours d'une période

Il convient préalablement de construire une variable $Y_{i,j}^P=Y_{i,j}^{t_1}-Y_{i,j}^{t_0}$

Cette variable peut prendre les valeurs suivantes :

0	lorsqu'il n'y a pas d'évolution entre les deux années, c'est à dire que le point j du segment i a conservé le même état par rapport à la catégorie de territoire h (être ou ne pas être en catégorie de territoire h) : ce constat recouvre deux situations bien distinctes. - soit le point était en catégorie de territoire h en t_0, il l'est toujours en t_1 - soit le point n'était pas en catégorie de territoire h en t_0, il ne l'est toujours pas en t_1, peu importe la catégorie de territoire en t_0 et en t_1 (même s'il y a eu changement de catégorie de territoire au cours de la période, le principal est que la catégorie de territoire n'est pas h)
1	lorsqu'il y a eu un changement positif : le point j du segment i n'était pas en catégorie de territoire h en t_0, il l'est devenu en t_1
-1	lorsqu'il y a eu un changement négatif : le point j du segment i était en catégorie de territoire h en t_0, il ne l'est plus en t_1

Si le nombre de points enquêtés en début et en fin de période n'est pas identique, nous avons alors $n^{t_0}\neq n^{t_1}$. Il convient d'estimer la différence de deux ratios.

La résolution du problème passe par l'utilisation du principe de linéarisation qui a été présenté précédemment.

Nous avons par définition $T_h^{t1}-T_h^{t0}=\Delta_h^{t1t0}=A\left[\frac{\sum_{i\in U}\sum_{j\in S}Y_{i,j}^{t1}}{\sum_{i\in U}\sum_{j\in S}X_{i,j}^{t1}}-\frac{\sum_{i\in U}\sum_{j\in S}Y_{i,j}^{t0}}{\sum_{i\in U}\sum_{j\in S}X_{i,j}^{t0}}\right]=A\left(\frac{Y_1}{X_1}-\frac{Y_0}{X_0}\right)$

La variable d'intérêt est ici la différence de deux ratios, soit $R=R_1-R_0=\frac{Y_1}{X_1}-\frac{Y_0}{X_0}$

L'estimateur naturel est $\hat{R}=\hat{R}_1-\hat{R}_0=\frac{\hat{Y}_1}{\hat{X}_1}-\frac{\hat{Y}_0}{\hat{X}_0}$

Ici $R=f(Y_1,Y_2,Y_3,Y_4)=\frac{Y_1}{Y_2}-\frac{Y_3}{Y_4}$ avec $Y_1=Y_1$, $Y_3=Y_0$, $Y_2=X_1$ et $Y_4=X_0$

D'où $Z_k=\frac{\partial f}{\partial Y_1}Y_{1,k}+\frac{\partial f}{\partial Y_2}Y_{2,k}+\frac{\partial f}{\partial Y_3}Y_{3,k}+\frac{\partial f}{\partial Y_4}Y_{4,k}=\frac{1}{Y_2}Y_{1,k}-\frac{Y_1}{Y_2^2}Y_{2,k}-\frac{1}{Y_4}Y_{3,k}+\frac{Y_3}{Y_4^2}Y_{4,k}$

Par conséquent

$$\hat{Z}=\sum_{k\in S}\frac{Z_k}{\pi_k}=\frac{1}{Y_2}\sum_{k\in S}\frac{Y_{1,k}}{\pi_k}-\frac{Y_1}{Y_2^2}\sum_{k\in S}\frac{Y_{2,k}}{\pi_k}-\frac{1}{Y_4}\sum_{k\in S}\frac{Y_{3,k}}{\pi_k}+\frac{Y_3}{Y_4^2}\sum_{k\in S}\frac{Y_{4,k}}{\pi_k}$$

$$\hat{Z}=\frac{1}{Y_2}\left[\sum_{k\in S}\frac{Y_{1,k}}{\pi_k}-\frac{Y_1}{Y_2}\sum_{k\in S}\frac{Y_{2,k}}{\pi_k}\right]-\frac{1}{Y_4}\left[\sum_{k\in S}\frac{Y_{3,k}}{\pi_k}-\frac{Y_3}{Y_4}\sum_{k\in S}\frac{Y_{4,k}}{\pi_k}\right]$$

$$\hat{Z}=\frac{1}{Y_2}\left[\sum_{k\in S}\frac{Y_{1,k}}{\pi_k}-R_1\sum_{k\in S}\frac{Y_{2,k}}{\pi_k}\right]-\frac{1}{Y_4}\left[\sum_{k\in S}\frac{Y_{3,k}}{\pi_k}-R_0\sum_{k\in S}\frac{Y_{4,k}}{\pi_k}\right]$$

$$\hat{Z}=\frac{1}{Y_2}\left(\hat{Y}_1-R_1\hat{Y}_2\right)-\frac{1}{Y_4}\left(\hat{Y}_3-R_0\hat{Y}_4\right)$$

En conséquence, en effectuant les remplacements adéquats, $\hat{Z}=\frac{1}{X_1}\left(\hat{Y}_1-R_1\hat{X}_1\right)-\frac{1}{X_0}\left(\hat{Y}_0-R_0\hat{X}_0\right)$

$$EQM\left(\hat{R}\right)\cong V\left(\hat{Z}\right)$$

$$V\left(\hat{Z}\right)=V\left[\frac{1}{X_1}\left(\hat{Y}_1-R_1\hat{X}_1\right)-\frac{1}{X_O}\left(\hat{Y}_0-R_0\hat{X}_0\right)\right]=V\left[\frac{\hat{W}_1}{X_1}-\frac{\hat{W}_0}{X_0}\right]$$

avec $\hat{W}_1=\hat{Y}_1-R_1\hat{X}_1$ et $\hat{W}_0=\hat{Y}_0-R_0\hat{X}_0$

Comme le nombre de points enquêtés en début et en fin de période est constant[8], la démarche présentée lors de la construction de la variance estimée de l'estimateur de la superficie en catégorie de territoire h pour une zone géographique donnée va s'appliquer dans le cadre de la construction de la variance estimée de l'estimateur de l'évolution de la superficie en catégorie de territoire h pour une zone géographique donnée : il suffit pour cela de remplacer simplement dans les formules :
$Y_{i,j}$ par $Y_{i,j}^P=Y_{i,j}^{t1}-Y_{i,j}^{t0}$

En conséquence, toujours dans les mêmes formules, $\hat{W}_{i,j}$ sera égal à $\left[\left(Y_{i,j}^{t1}-Y_{i,j}^{t0}\right)-\hat{R}X_{i,j}\right]$

avec $\tilde{R}=\frac{\sum_{i\in S}\sum_{j\in S_i}Y_{i,j}^P}{\sum_{i\in S}\sum_{j\in S_i}X_{i,j}}=\frac{\sum_{i\in S}\sum_{j\in S_i}\left(Y_{i,j}^{t1}-Y_{i,j}^{t0}\right)}{\sum_{i\in S}\sum_{j\in S_i}X_{i,j}}$

La variance estimée de l'estimateur de l'évolution de la superficie en catégorie de territoire h pour une zone géographique donnée est donnée par la formule suivante :

[8] C'est le même échantillon qui est suivi au cours du temps (les segments sont les mêmes, les points sont les mêmes).

$$\hat{V}\left(\hat{\Delta}_h^P\right) = \frac{90mA}{n^2}\left[\frac{M-m}{2nc}\sum_{i=1}^{nc-1}\left(\sum_{j\in S_{i+1}}\hat{W}_{i+1,j} - \sum_{j\in S_i}\hat{W}_{i,j}\right)^2 + \sum_{i\in S}\frac{n_i}{2nc_i}\sum_{j=1}^{nc_i-1}\left(\hat{W}_{i,j+1}-\hat{W}_{i,j}\right)^2\right]$$

$$\hat{V}\left(\hat{\Delta}_h^P\right) = \frac{90mA}{n^2}\left[\frac{M-m}{2nc}\sum_{i=1}^{nc-1}\left(\sum_{j\in S_{i+1}}\left(\left(Y_{i+1,j}^{t_1}-Y_{i+1,j}^{t_0}\right)-\hat{R}X_{i+1,j}\right) - \sum_{j\in S_i}\left(\left(Y_{i,j}^{t_1}-Y_{i,j}^{t_0}\right)-\hat{R}X_{i,j}\right)\right)^2\right]$$

$$+\frac{90mA}{n^2}\left[\sum_{i\in S}\frac{n_i}{2nc_i}\sum_{j=1}^{nc_i-1}\left[\left(\left(Y_{i,j+1}^{t_1}-Y_{i,j+1}^{t_0}\right)-\hat{R}X_{i,j+1}\right)-\left(\left(Y_{i,j}^{t_1}-Y_{i,j}^{t_0}\right)-\hat{R}X_{i,j}\right)\right]^2\right]$$

3. Macro SAS

3.1. Présentation

Pour mettre en œuvre les formules développées précédemment, une macro SAS paramétrable a été programmée. Elle permet d'appliquer ces formules dans la plupart des situations, tant en estimation pour une année donnée qu'entre deux années.

Elle se veut relativement simple et souple d'utilisation. Il est ainsi possible de travailler sur une des nomenclatures construites par défaut à l'issue de l'enquête, ou bien de créer sa propre nomenclature. Tout niveau géographique supra-communal peut également être analysé : la région, le département, ou n'importe quel regroupement de communes.

En outre, la macro traite le cas où l'échantillon est plus ou moins dense selon la zone du territoire, que cela soit en terme de nombre de segments, ou bien de nombres de points par segment. C'est notamment le cas lorsqu'on augmente la taille de l'échantillon pour obtenir des résultats plus précis dans une zone donnée du territoire.

3.2. Syntaxe

```
%PRECISION_TERUTI_LUCAS (    IN0               = ,
                             IN1               = ,
                             SURFACES          = ,
                             OUT               = ,
                             NOMENCLATURE      = ,
                             CODE_OCCUPATION   = ,
                             REG               = ,
                             DEP               = ,
                             COM               =    ) ;
```

Les paramètres obligatoires sont les suivants :

- IN0 : pointe vers la table SAS contenant les données de l'enquête Teruti-Lucas d'une année donnée ;
- SURFACES : pointe vers la table SAS contenant la surface des différents niveaux géographiques (France, région, département, commune) ;
- OUT : nom de la table SAS en sortie qui va contenir les différentes estimations ;
- NOMENCLATURE : le nom de la nomenclature qu'on souhaite analyser.

Par défaut, les estimations sont effectuées sur la France entière, pour tous les codes de la nomenclature indiquée.

Les paramètres optionnels sont les suivants :

- IN1 : si on souhaite un calcul de précision en évolution, IN1 pointe vers la table SAS contenant les données de l'année la plus récente ;
- CODE_OCCUPATION : indique un code particulier de la nomenclature qu'on souhaite traiter ;
- REG : indique la région qu'on souhaite traiter. Si plusieurs régions sont renseignées, le calcul est effectué sur la zone géographique regroupant ces régions.
- DEP : indique le département qu'on souhaite traiter. Si plusieurs départements sont renseignés, le calcul est effectué sur la zone géographique regroupant ces départements.
- COM: indique la commune qu'on souhaite traiter. Si plusieurs communes sont renseignées, le calcul est effectué sur la zone géographique regroupant ces communes.

3.3. Les sorties

La table en sortie contient les informations suivantes :

- la zone considérée : France, région(s), département(s), commune(s)
- la modalité de la nomenclature
- l'année traitée (ou les deux années pour une évolution)
- l'estimateur de la surface de la modalité considérée
- la précision de cet estimateur
- les bornes inférieures et supérieures de l'intervalle de confiance à 95% de l'estimateur

Par exemple, le tableau ci-dessous présente les résultats pour l'estimation, dans la région 91 (Languedoc-Roussillon), de l'évolution entre 2006 et 2010 de la surface prise par les trois modalités d'une nomenclature agrégée (Agricole, Artificiel, Naturel) :

zone	code	annee	S_chap	precision	b_inf	b_sup	niveau
Reg 91	Agricole	2006	86761013	1.9	83453059	90068967	95 %
Reg 91	Agricole	2010	83569679	2.0	80369909	86769449	95 %
Reg 91	Agricole	2006=>2010	-3191334	19.4	-4402414	-1980254	95 %
Reg 91	Artificiel	2006	19003763	4.6	17306614	20700912	95 %
Reg 91	Artificiel	2010	20283903	4.4	18530263	22037543	95 %
Reg 91	Artificiel	2006=>2010	1280140	25.0	652903	1907377	95 %
Reg 91	Naturel	2006	171845224	1.1	168231083	175459365	95 %
Reg 91	Naturel	2010	173756418	1.0	170201767	177311069	95 %
Reg 91	Naturel	2006=>2010	1911194	32.8	681440	3140948	95 %

En 2006, pour cette région, l'estimateur de la surface en « Agricole » est de 83 761 013 ares, avec une précision de 1,9 %. La valeur vraie de la surface en « Agricole » a alors 95 % de chance de se situer entre 83 453 059 et 90 068 967 ares.

En 2010 l'estimateur de cette surface est de 83 569 679 ares, et l'estimateur de l'évolution entre 2006 et 2010 est de - 3 191 334 ares. La précision de cette évolution est de 19,4 %.

4. Construction d'un abaque

4.1. Présentation

Cette partie présente une méthode d'estimation de la précision à laquelle on peut s'attendre a priori, en fonction de la zone géographique et de la nomenclature étudiées. L'abaque ainsi constitué permet notamment de mesurer grossièrement le gain de précision qu'on peut espérer obtenir lorsqu'on augmente le nombre de segments enquêtés.

Les données permettant de réaliser la construction d'un abaque résultent du calcul de la précision des estimateurs sur un grand nombre de situations contrastées.

Les départements et les régions ont été croisés avec différentes nomenclatures disponibles à l'issue de l'enquête sur le terrain, à des niveaux de détail plus ou moins fins. Parmi tous ces croisements, seules sont conservées les modalités de la nomenclature comportant au moins 20 points, car les estimateurs sont très peu précis lorsque le nombre de points est trop faible. Au total cela représente 3 774 situations différentes.

Pour chacune de ces situations, la précision de l'estimateur de la superficie de telle ou telle modalité a été estimée.

4.2. Estimation a priori de la précision

La forme complexe de l'estimateur de la variance de ce plan de sondage ne permet pas d'expliquer directement de façon simple la précision par les différents paramètres. Pour arriver à un résultat simple et utilisable, une approximation grossière est faite afin de pouvoir estimer une fonction empirique reliant la précision à la valeur de cette estimation. L'hypothèse supposée est que la variance est – à un facteur près – celle de la loi binomiale :

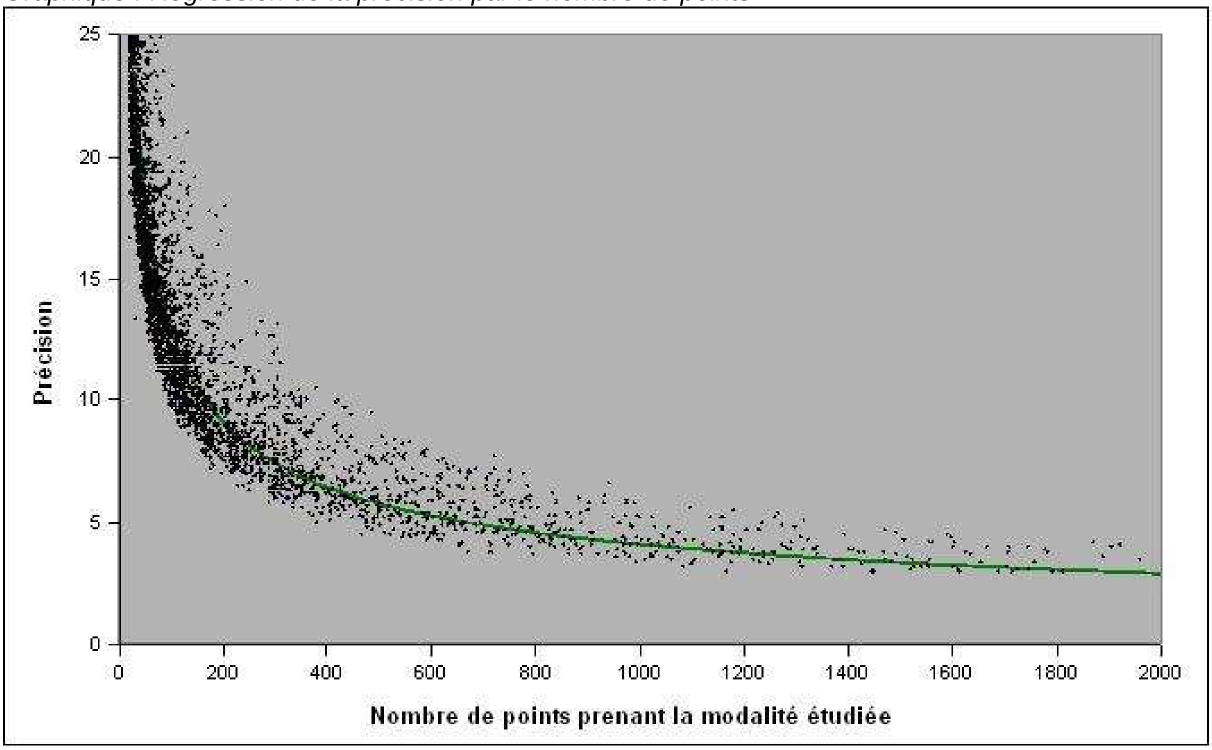
Précision (n) = $\frac{a}{\sqrt{n}}$ où n est le nombre de points prenant la modalité étudiée

La précision est régressée par le nombre de points correspondant à la modalité considérée.

L'équation est de la forme : Précision (surface) = $\frac{129}{\sqrt{n}}$

La précision de l'estimateur de la précision se calcule en tenant compte de l'hétéroscédasticité des résidus : la dispersion des résidus n'est en effet pas homogène sur toutes les valeurs prises par le nombre de points. Dans le cas d'un faible nombre de points, l'intervalle de confiance augmente alors exponentiellement.

Graphique : Régression de la précision par le nombre de points



4.3. Sensibilité de la précision au nombre de points par segment

Les mêmes 3 774 estimations sont cette fois effectuées uniquement sur les 5 premiers points de chaque segment. La précision est à nouveau régressée par le nombre de points prenant la modalité considérée.

La forme de la précision est cette fois : Précision (surface) = $\frac{114}{\sqrt{n}}$

Le fait de passer de 10 à 5 points par segment tout en doublant le nombre de segments (donc en conservant le même nombre de points enquêtés) permet ainsi un gain de précision de l'ordre de 11 %. Ce résultat est logique, car dans un plan de sondage à deux degrés, la variance s'explique en majeure partie par la variance résultant du premier degré de tirage : doubler le nombre d'unités primaires fait plus diminuer la variance que doubler le nombre d'unités secondaires par unité primaire.

4.4. L'abaque

Les estimations précédentes permettent de constituer un abaque pour un échantillon de 5 ou 10 points par segment. En fonction du nombre de points appartenant à la modalité considérée et du nombre de points enquêtés par segment, on peut donc estimer la précision attendue.

Il convient tout de même de rappeler que ces estimations ont été réalisées avec une approximation sur la forme fonctionnelle de la précision. Ces résultats doivent être maniés avec précaution. Ils permettent surtout de donner un ordre de grandeur.

Nombre de points	Précision attendue (%)	
	5 points par segment	10 points par segment
20	26	31
30	21	24
40	18	21
50	16	19
60	15	17
70	14	16
80	13	15
90	12	14
100	11	13
120	10	12
140	10	11
160	9	10
180	9	10
200	8	9
220	8	9
240	7	9
260	7	8
280	7	8
300	7	8
320	6	7
340	6	7
360	6	7
380	6	7
400	6	7
420	6	7
440	5	6
460	5	6
480	5	6
500	5	6

Lecture : dans une zone comptant 100 points dans la modalité étudiée, et où on enquête 10 points par segment, en moyenne on peut s'attendre à une précision de 13 %. Dans une zone comptant 100 points dans la modalité étudiée, et où on enquête 5 points par segment (on a donc deux fois plus de segments que dans le cas précédent), en moyenne on peut s'attendre à une précision de 11 %.

Outre le fait de pouvoir estimer rapidement la précision en fonction du nombre de points, cet abaque permet ainsi d'estimer le gain de précision qu'on peut espérer obtenir lorsqu'on augmente le nombre de segments enquêtés dans une zone géographique donnée.

Bibliographie

[1] AMORICH S., GIRALT A., PALACIO-RABAUD V., POIRET M. et WULLEMS M.-A., Précision de l'enquête utilisation du territoire, Ministère de l'Agriculture, de la Pêche et de l'Alimentation, SCEES, Paris, 1997, 73 p.

[2] ARDILLY P., Les techniques de sondage, éditions Technip, Paris, 2006

[3] COCHRAN W., Sampling techniques, éditions Wiley, New-York, 1977

[4] DESABIE J., Théorie et pratique des sondages, éditions Dunod, Paris, 1966

[5] FOURNIER P., "Etude sur l'utilisation du territoire - Méthodologie - Résultats 1969-1970-1971", Ministère de l'Agriculture, SCEES, Supplément "Série Études", n°104, novembre 1972, 112 p.

[6] MARY A., Eléments pour une analyse de l'occupation du territoire à partir des enquêtes TERUTI et TERUTI-LUCAS, Service la Statistique et de la Prospective (SSP) - Université Toulouse I - Université Toulouse III, rapport de stage, avril-août 2009, 3 tomes

ANNEXE 01

Amandine MARY, Eléments pour une analyse de l’occupation du territoire à partir des enquêtes TERUTI et TERUTI-LUCAS, Service de la Statistique et de la Prospective(SSP) - Université Toulouse I - Université Toulouse III, rapport de stage, avril-août 2009, p.28-32

La présentation d'un exemple va permettre de mieux appréhender le passage d'une somme à l'autre dans l'égalité $\frac{1}{m-1}\sum_{i\in S}(\hat{W}_i^{\,2}) = \frac{1}{2nc}\sum_{i=1}^{nc-1}(\hat{W}_{i+1} - \hat{W}_i)^2$:

<u>Exemple dans le cas de l'enquête TERUTI-LUCAS</u> :

Soit une sous zone géographique représentée par 22 segments disposés de la manière suivante (ici échantillon non renforcé c'est à dire deux segments seront dit voisins si leurs numéros sont séparés par +2 ou bien par +2000).

Cet exemple concerne un département et le code occupation qui nous intéresse est le code 272 (les vignes).

Représentation des numéros des segments :

ligne \ colonne	12	13	14	15	16	17	18	19	20	21
4		.		.		.		60208		.
5	.		.		.		61207		61209	
6		62202		.		62206		62208		62210
7	.		.		63205		63207		63209	
8		64202		64204		64206		.		.
9	65201		65203		65205		65207		65209	
10		66202		66204		.		66208		66210

Pour chaque segment i, nous connaissons les $\hat{W}_i$, le $\hat{W}_i$ est variable de segment à segment. Dans l'exemple ci-dessus, le $\hat{W}_i$ va prendre 22 valeurs.

Représentation des $\hat{W}_i$ (de l'ordre de 10^4) :

ligne \ colonne	12	13	14	15	16	17	18	19	20	21
4		.		.		.		-1,57		.
5	.		.		.		-1,57		-1,57	
6		-1,57		.		-1,57		0,97		-1,57
7	.		.		-0,29		-0,29		6,09	
8		-0,29		0,97		-1,57		.		.
9	2,25		2,25		2,25		-1,57		-1,57	
10		2,25		4,81		.		3,53		-1,57

{ <u>Explication</u> : on rappelle que $\hat{W}_i = \frac{N_i}{n_i}\sum_{j\in S_i}\hat{W}_{i,j} = \frac{N_i}{n_i}\sum_{j\in S_i}(Y_{i,j} - \hat{R}X_{i,j})$.

En prenant par exemple le segment portant le numéro 66208, les points qui ont été enquêtés

sont les 10 points 11, 12, 13, 14, 15, 21, 22, 23, 24, 25 (donc $n_i = 10$), ensuite on regarde parmi ces 10 points ceux qui possèdent l'occupation considérée. Les points qui possèdent l'occupation considérée sont les points 13, 14, 15 et 24 (donc pour chacun de ces points $Y_{i,j} = 1$), les autres ne possèdent pas l'occupation considérée. ($X_{i,j} = 1$ pour chaque point quelle que soit l'occupation).
On fait de même pour chaque segment.

On obtient donc à la fin le nombre de points enquêtés dans la sous zone possédant l'occupation h et le nombre total de points enquêtés dans la sous zone.

Et ainsi on obtient $\hat{R} = \dfrac{\text{le nombre de points enquêtés dans la sous zone possédant l'occupation h}}{\text{le nombre total de points enquêtés dans la sous zone}}$

ici $\hat{R} = 0.124$.

Donc en revenant à notre segment 66208, on peut calculer $\hat{W}_{66208}$:

$$\hat{W}_{66208} = \frac{N_{66208}}{n_{66208}} \sum_{j \in S_{66208}} (Y_{66208,j} - \hat{R}X_{66208,j}) = \frac{N_{66208}}{n_{66208}} (Y_{66208,11} - \hat{R}X_{66208,11} + Y_{66208,12} - \hat{R}X_{66208,12}$$
$$+ Y_{66208,13} - \hat{R}X_{66208,13} + Y_{66208,14} - \hat{R}X_{66208,14} + Y_{66208,15} - \hat{R}X_{66208,15} + Y_{66208,21} - \hat{R}X_{66208,21}$$
$$+ Y_{66208,22} - \hat{R}X_{66208,22} + Y_{66208,23} - \hat{R}X_{66208,23} + Y_{66208,24} - \hat{R}X_{66208,24} + Y_{66208,25} - \hat{R}X_{66208,25})$$
$$= \frac{(900000/7)}{10}(0 - \hat{R}*1 + 0 - \hat{R}*1 + 1 - \hat{R}*1 + 1 - \hat{R}*1 + 1 - \hat{R}*1 + 0 - \hat{R}*1 + 0 - \hat{R}*1 + 0 - \hat{R}*1$$
$$+ 1 - \hat{R}*1 + 0 - \hat{R}*1) = \frac{90000}{7}(-10\hat{R} + 4) = (1.28*10^4)*(-10*0.124 + 4)$$
$$= 1.28*(-10*0.124 + 4)*10^4 = 3.53*10^4$$

Et on fait de même pour tous les segments.}

$\rightarrow$Calcul de la somme $\frac{1}{m-1}\sum_{i\in S}(\hat{W}_i^2)$:

$$\sum_{i\in S}(\hat{W}_i^2) = \hat{W}^2_{60208}$$
$$+ \hat{W}^2_{61207} + \hat{W}^2_{61209}$$
$$+ \hat{W}^2_{62202} + \hat{W}^2_{62206} + \hat{W}^2_{62208} + \hat{W}^2_{62210}$$
$$+ \hat{W}^2_{63205} + \hat{W}^2_{63207} + \hat{W}^2_{63209}$$
$$+ \hat{W}^2_{64202} + \hat{W}^2_{64204} + \hat{W}^2_{64206}$$
$$+ \hat{W}^2_{65201} + \hat{W}^2_{65203} + \hat{W}^2_{65205} + \hat{W}^2_{65207} + \hat{W}^2_{65209}$$
$$+ \hat{W}^2_{66202} + \hat{W}^2_{66204} + \hat{W}^2_{66208} + \hat{W}^2_{66210}$$
$$= (-1.57*10^4)^2$$
$$+ (-1.57*10^4)^2 + (-1.57*10^4)^2$$
$$+ (-1.57*10^4)^2 + (-1.57*10^4)^2 + (0.97*10^4)^2 + (-1.57*10^4)^2$$
$$+ (-0.29*10^4)^2 + (-0.29*10^4)^2 + (6.09*10^4)^2$$
$$+ (-0.29*10^4)^2 + (0.97*10^4)^2 + (-1.57*10^4)^2$$
$$+ (2.25*10^4)^2 + (2.25*10^4)^2 + (2.25*10^4)^2 + (-1.57*10^4)^2 + (-1.57*10^4)^2$$
$$+ (2.25*10^4)^2 + (4.81*10^4)^2 + (3.53*10^4)^2 + (-1.57*10^4)^2$$
$$= (10^4)^2 *[(-1.57)^2$$
$$+ (-1.57)^2 + (-1.57)^2$$
$$+ (-1.57)^2 + (-1.57)^2 + (0.97)^2 + (-1.57)^2$$
$$+ (-0.29)^2 + (-0.29)^2 + (6.09)^2$$
$$+ (-0.29)^2 + (0.97)^2 + (-1.57)^2$$
$$+ (2.25)^2 + (2.25)^2 + (2.25)^2 + (-1.57)^2 + (-1.57)^2$$
$$+ (2.25)^2 + (4.81)^2 + (3.53)^2 + (-1.57)^2] = 120.01*10^8$$

ici m=22 donc m-1=21 d'où $\frac{1}{m-1}\sum_{i\in S}(\hat{W}_i^2) = \frac{120.01*10^8}{21} = 5.7*10^8$.

$\rightarrow$Calcul de la somme $\frac{1}{2nc}\sum_{i=1}^{nc-1}(\hat{W}_{i+1} - \hat{W}_i)^2$:

Il convient de calculer les différences entre segments voisins tant au niveau des lignes que des colonnes.

- au niveau des lignes :

Somme des carrées des différences des $\hat{W}_i$ pour les segments voisins en ligne :

$$
\begin{aligned}
S1 &= (\hat{W}_{61209} - \hat{W}_{61207})^2 \\
&+ (\hat{W}_{62208} - \hat{W}_{62206})^2 + (\hat{W}_{62210} - \hat{W}_{62208})^2 \\
&+ (\hat{W}_{63207} - \hat{W}_{63205})^2 + (\hat{W}_{63209} - \hat{W}_{63207} -)^2 \\
&+ (\hat{W}_{64204} - \hat{W}_{64202})^2 + (\hat{W}_{64206} - \hat{W}_{64204})^2 \\
&+ (\hat{W}_{65203} - \hat{W}_{65201})^2 + (\hat{W}_{65205} - \hat{W}_{65203})^2 + (\hat{W}_{65207} - \hat{W}_{65205})^2 + (\hat{W}_{65209} - \hat{W}_{65207})^2 \\
&+ (\hat{W}_{66204} - \hat{W}_{66202})^2 + (\hat{W}_{66210} - \hat{W}_{66208})^2 \\
&= (-1.57*10^4 - (-1.57*10^4))^2 \\
&+ (0.97*10^4 - (-1.57*10^4))^2 + (-1.57*10^4 - 0.97*10^4)^2 \\
&+ (-0.29*10^4 - (-0.29*10^4))^2 + (6.09*10^4 - (-0.29*10^4))^2 \\
&+ (0.97*10^4 - (-0.29*10^4))^2 + (-1.57*10^4 - 0.97*10^4)^2 \\
&+ \begin{Bmatrix} (2.25*10^4 - 2.25*10^4)^2 + (2.25*10^4 - 2.25*10^4)^2 \\ + (-1.57*10^4 - 2.25*10^4)^2 + (-1.57*10^4 - (-1.57*10^4))^2 \end{Bmatrix} \\
&+ (4.81*10^4 - 2.25*10^4)^2 + (-1.57*10^4 - 3.53*10^4)^2 \\
&= (10^4)^2 * [(-1.57 - (-1.57))^2 \\
&+ (0.97 - (-1.57))^2 + (-1.57 - 0.97)^2 \\
&+ (-0.29 - (-0.29))^2 + (6.09 - (-0.29))^2 \\
&+ (0.97 - (-0.29))^2 + (-1.57 - 0.97)^2 \\
&+ (2.25 - 2.25)^2 + (2.25 - 2.25)^2 + (-1.57 - 2.25)^2 + (-1.57 - (-1.57))^2 \\
&+ (4.81 - 2.25)^2 + (-1.57 - 3.53)^2] = 109.15*10^8
\end{aligned}
$$

Le nombre de différences lignes est égal à 13 : dl=13.

- au niveau des colonnes :

Somme des carrées des différences des $\hat{W}_i$ pour les segments voisins en colonne :

$$S2 = (\hat{W}_{66202} - \hat{W}_{64202})^2 + (\hat{W}_{64202} - \hat{W}_{62202})^2$$
$$+ (\hat{W}_{66204} - \hat{W}_{64204})^2$$
$$+ (\hat{W}_{65205} - \hat{W}_{63205})^2$$
$$+ (\hat{W}_{64206} - \hat{W}_{62206})^2$$
$$+ (\hat{W}_{65207} - \hat{W}_{63207})^2 + (\hat{W}_{63207} - \hat{W}_{61207})^2$$
$$+ (\hat{W}_{62208} - \hat{W}_{60208})^2$$
$$+ (\hat{W}_{65209} - \hat{W}_{63209})^2 + (\hat{W}_{63209} - \hat{W}_{61209})^2$$
$$= (2.25*10^4 - (-0.29*10^4))^2 + (-0.29*10^4 - (-1.57*10^4))^2$$
$$+ (4.81*10^4 - 0.97*10^4)^2$$
$$+ (2.25*10^4 - (-0.29*10^4))^2$$
$$+ (-1.57*10^4 - (-1.57*10^4))^2$$
$$+ (-1.57*10^4 - (-0.29*10^4))^2 + (-0.29*10^4 - (-1.57*10^4))^2$$
$$+ (0.97*10^4 - (-1.57*10^4))^2$$
$$+ (-1.57*10^4 - 6.09*10^4)^2 + (6.09*10^4 - (-1.57*10^4))^2$$
$$= (10^4)^2[(2.25 - (-0.29))^2 + (-0.29 - (-1.57))^2$$
$$+ (4.81 - 0.97)^2$$
$$+ (2.25 - (-0.29))^2$$
$$+ (-1.57 - (-1.57))^2$$
$$+ (-1.57 - (-0.29))^2 + (-0.29 - (-1.57))^2$$
$$+ (0.97 - (-1.57))^2$$
$$+ (-1.57 - 6.09)^2 + (6.09 - (-1.57))^2] = 156.7*10^8$$

Le nombre de différences colonnes est égal à 10 : dc=10

Ensuite il faut sommer ces résultats :
$$S1 + S2 = 109.15*10^8 + 156.7*10^8 = (109.15 + 156.7)*10^8 = 265.85*10^8$$

Puis diviser cette somme par [2*(dl+dc)]

On obtient finalement : $\frac{1}{2nc}\sum_{i=1}^{nc-1}(\hat{W}_{i+1} - \hat{W}_i)^2 = \frac{265.85*10^8}{[2*(13+10)]} = \frac{265.85}{46}*10^8 = 5.7*10^8$

On a donc $\frac{1}{m-1}\sum_{i\in S}(\hat{W}_i^2) = \frac{1}{2nc}\sum_{i=1}^{nc-1}(\hat{W}_{i+1} - \hat{W}_i)^2$

ANNEXE 02

Amandine MARY, Eléments pour une analyse de l’occupation du territoire à partir des enquêtes TERUTI et TERUTI-LUCAS, Service de la Statistique et de la Prospective(SSP) - Université Toulouse I - Université Toulouse III, rapport de stage, avril-août 2009, p.34-36

La présentation d'un exemple va permettre de mieux appréhender le passage d'une somme à l'autre dans l'égalité $\frac{1}{n_i - 1}\left(\frac{n_i \sum_{j\in S_i}\left(\hat{W}_{i,j}^{\ 2}\right) - (\sum_{j\in S_i} \hat{W}_{i,j})^2}{n_i}\right) = \frac{1}{2nc_i}\sum_{j=1}^{nc_i - 1}(\hat{W}_{i,j+1} - \hat{W}_{i,j})^2$:

Exemple : le calcul de ces expressions s'effectue pour un segment i, puisque les deux premières lignes sont enquêtées uniquement, on ne s'intéressera qu'aux $\hat{W}_{i,j}$ de ces lignes.

Représentation des numéros des points :

●	●	●	●	●
11	12	13	14	15
●	●	●	●	●
21	22	23	24	25
o	o	o	o	o
31	32	33	34	35
o	o	o	o	o
41	42	43	44	45
o	o	o	o	o
51	52	53	54	55

Représentation des $\hat{W}_{i,j}$:

●	●	●	●	●
0,876	-0,124	0,876	0,876	-0,124
●	●	●	●	●
0,876	-0,124	-0,124	0,876	0,876
o	o	o	o	o
.	.	.	.	.
o	o	o	o	o
.	.	.	.	.
o	o	o	o	o
.	.	.	.	.

{ Explication : on rappelle que $\hat{W}_{i,j} = Y_{i,j} - \hat{R}X_{i,j}$.

On choisit un segment parmi les segment présentés précédemment et on regarde les points à l'intérieur de ce segment (ici on a encore $\hat{R} = 0.124$).
Pour ce segment, les points qui possèdent l'occupation souhaitée h sont les points 11, 13, 14, 21, 24 et 25.

En prenant par exemple le point ayant le numéro 12, ici dans notre exemple il ne possède pas l'occupation h donc $Y_{i,12} = 0$ et $X_{i,12} = 1$ (vaut 1 quelle que soit l'occupation) donc on aura pour ce point $\hat{W}_{i,12} = Y_{i,12} - \hat{R}X_{i,12} = 0 - \hat{R}*1 = -0.124$.

En prenant par exemple le point ayant le numéro 24, ici dans notre exemple il possède l'occupation h donc $Y_{i,24} = 1$ et $X_{i,24} = 1$ (vaut 1 quelle que soit l'occupation) donc on aura pour ce point $\hat{W}_{i,24} = Y_{i,24} - \hat{R}X_{i,24} = 1 - \hat{R}*1 = 1 - 0.124 = 0.876$.}

→ <u>Calcul de la somme</u> $\frac{1}{n_i - 1}\left(\frac{n_i \sum_{j\in S_i}\left(\hat{W}_{i,j}^{\ 2}\right) - \left(\sum_{j\in S_i}\hat{W}_{i,j}\right)^2}{n_i}\right)$:

Ici pour ce segment les 10 points on pu être enquêtés donc $n_i = 10$.

On a

$$\sum_{j\in S_i}\left(\hat{W}_{i,j}^{\ 2}\right) = (\hat{W}_{i,11})^2 + (\hat{W}_{i,12})^2 + (\hat{W}_{i,13})^2 + (\hat{W}_{i,14})^2 + (\hat{W}_{i,15})^2$$
$$+ (\hat{W}_{i,21})^2 + (\hat{W}_{i,22})^2 + (\hat{W}_{i,23})^2 + (\hat{W}_{i,24})^2 + (\hat{W}_{i,25})^2$$
$$= (0.876)^2 + (-0.124)^2 + (0.876)^2 + (0.876)^2 + (-0.124)^2$$
$$+ (0.876)^2 + (-0.124)^2 + (-0.124)^2 + (0.876)^2 + (0.876)^2 = 4.6658$$

Et

$$\left(\sum_{j\in S_i}\hat{W}_{i,j}\right)^2 = (\hat{W}_{i,11} + \hat{W}_{i,12} + \hat{W}_{i,13} + \hat{W}_{i,14} + \hat{W}_{i,15}$$
$$+ \hat{W}_{i,21} + \hat{W}_{i,22} + \hat{W}_{i,23} + \hat{W}_{i,24} + \hat{W}_{i,25})^2$$
$$= (0.876 - 0.124 + 0.876 + 0.876 - 0.124$$
$$+ 0.876 - 0.124 - 0.124 + 0.876 + 0.876)^2 = 22.6957$$

Donc

$$\frac{1}{n_i - 1}\left(\frac{n_i \sum_{j\in S_i}\left(\hat{W}_{i,j}^{\ 2}\right) - \left(\sum_{j\in S_i}\hat{W}_{i,j}\right)^2}{n_i}\right) = \frac{1}{10-1}\left(\frac{10*4.6658 - 22.6957}{10}\right) = 0.26$$

→ <u>Calcul de la somme</u> $\frac{1}{2nc_i}\sum_{j=1}^{nc_i - 1}(\hat{W}_{i,j+1} - \hat{W}_{i,j})^2$:

Il convient de calculer les différences entre points voisins tant au niveau des lignes que des colonnes.

- au niveau des lignes :

Somme des carrés des différences des $\hat{W}_{i,j}$ pour les points voisins en ligne :

$$S1 = (\hat{W}_{i,12} - \hat{W}_{i,11})^2 + (\hat{W}_{i,13} - \hat{W}_{i,12})^2 + (\hat{W}_{i,14} - \hat{W}_{i,13})^2 + (\hat{W}_{i,15} - \hat{W}_{i,14})^2$$
$$+ (\hat{W}_{i,22} - \hat{W}_{i,21})^2 + (\hat{W}_{i,23} - \hat{W}_{i,22})^2 + (\hat{W}_{i,24} - \hat{W}_{i,23})^2 + (\hat{W}_{i,25} - \hat{W}_{i,24})^2$$
$$= (-0.124 - 0.876)^2 + (0.876 - (-0.124))^2 + (0.876 - 0.876)^2 + (-0.124 - 0.876)^2$$
$$+ (-0.124 - 0.876)^2 + (-0.124 - (-0.124))^2 + (0.876 - (-0.124))^2 + (0.876 - 0.876)^2$$
$$= 5$$

Il y a 8 différences qui ont été calculées : dl=8.

- au niveau des colonnes :

Somme des carrés des différences des $\hat{W}_{i,j}$ pour les points voisins en colonne :

$$S1 = (\hat{W}_{i,21} - \hat{W}_{i,11})^2$$
$$+ (\hat{W}_{i,22} - \hat{W}_{i,12})^2$$
$$+ (\hat{W}_{i,23} - \hat{W}_{i,13})^2$$
$$+ (\hat{W}_{i,24} - \hat{W}_{i,14})^2$$
$$+ (\hat{W}_{i,25} - \hat{W}_{i,15})^2$$
$$= (0.876 - 0.876)^2$$
$$+ (-0.124 - (-0.124))^2$$
$$+ (-0.124 - 0.876)^2$$
$$+ (0.876 - 0.876)^2$$
$$+ (0.876 - (-0.124))^2$$
$$= 2$$

Il y a 5 différences qui ont été calculées : dc=5.

Ensuite il faut sommer ces résultats : $S1 + S2 = 5 + 2 = 7$

Puis diviser cette somme par [2*(dl+dc)]

On obtient finalement : $\frac{1}{2nc_i}\sum_{j=1}^{nc_i - 1}(\hat{W}_{i,j+1} - \hat{W}_{i,j})^2 = \frac{7}{[2*(8+5)]} = \frac{7}{26} = 0.26$

On a donc

$$\frac{1}{n_i - 1}\left(\frac{n_i \sum_{j\in S_i}\left(\hat{W}_{i,j}{}^2\right) - (\sum_{j\in S_i}\hat{W}_{i,j})^2}{n_i}\right) = \frac{1}{2nc_i}\sum_{j=1}^{nc_i - 1}(\hat{W}_{i,j+1} - \hat{W}_{i,j})^2$$