DIRECTION DES RESSOURCES VIVANTES

ECHANTILLONNAGE EN CRIEE
POUR
L'ESTIMATION DES STRUCTURES DEMOGRAPHIQUES

par

R. CHEVALIER et B. MESNIL

IFREMER

DRV.86.006/RH/Nantes

INSTITUT FRANÇAIS DE RECHERCHE POUR L'EXPLOITATION DE LA MER

CENTRE DE NANTES

Rue de l'Ile d'Yeu
Boîte Postale 1049

44037 - NANTES CEDEX

DIRECTION DES RESSOURCES VIVANTES

DEPARTEMENT RESSOURCES HALIEUTIQUES

LABORATOIRE ERHAL

AUTEUR (S) : MM. R. CHEVALIER, B. MESNIL		CODE : *cf. liste des codes* N° DRV-86.006/RH NANTES
TITRE ECHANTILLONNAGE EN CRIEE POUR L'ESTIMATION DES STRUCTURES DEMOGRAPHIQUES.		date : 16.12.1986 tirage nb : 25
		Nb pages : 35 Nb figures : Nb photos :
CONTRAT *(intitulé)* N°		DIFFUSION libre ☒ restreinte ☐ confidentielle ☐

RÉSUMÉ Cette note a pour objectif une reconnaissance des problèmes rencontrés lors de l'estimation des structures démographiques des débarquements en criée.

Préalable à la définition d'une stratégie optimale d'échantillonnage, la réflexion porte sur les choix d'un plan de sondage et des estimateurs.

Nous avons utilisé comme fonction à optimiser, une valeur moyenne des taux instantanés de mortalité par pêche F, calculés par l'analyse de cohorte. Sur ce point crucial, de la définition d'une fonction objective des choix plus élaborés peuvent être envisagés.

ABSTRACT

Some problems encountered in estimating the age structure of landings from fish market sampling are reviewed in this document.

Being perequisite to the definition of on optimal sampling strategy, the selection of a sampling scheme and the choice of the estimators are discussed.

The definition of an objective function to optimize is also critical. We have used an average of instantaneous fishing mortality rates obtained from a (pseudo-) cohort analysis. More elaborate functions can be envisaged.

mots-clés : Echantillonnage - clé taille-âge .

key words : Sampling - age length key

IFREMER

SOMMAIRE

INTRODUCTION

L'estimation des structures démographiques des débarquements à partir d'opérations d'échantillonnage réalisées en criée présente quelques difficultés découlant en partie de la nature du matériel sondé et aussi des conditions de travail des enquêteurs.

En effet, les populations soumises à notre analyse sont en évolution permanente. Les échantillonnages biologiques doivent être répétitifs pour permettre l'étude de ces évolutions.

D'autre part les observations sont liées aux activités d'exploitation de la ressource par les pêcheries et aux techniques de débarquement et de vente.

Les effectifs débarqués se répartissent le plus souvent suivant un schéma de multistratification impliquant des strates spatio-temporelles, des points de débarquements, des métiers et, quand elles existent, des catégories commerciales.

Mais les poids relatifs de chacune de ces strates peuvent constamment varier sous l'influence :

- des variations naturelles des populations étudiées,
- des modifications du rapport des différents métiers,
- du changement du schéma de fréquentation des strates spatio-temporelles.

Les calculs théoriques d'optimisation ne conduisent pas toujours à des relations simples. Des approximations seront nécessaires dont l'intérêt pratique n'est toutefois pas négligeable. Cette approche classique a été suivie pour le hareng (G. BIAIS, 1985) et se poursuit pour d'autres espèces.

Notre but par cette note est de fournir quelques éclaircissements sur le type d'échantillonnage à adopter en s'appuyant essentiellement sur les critères de précisions.

Le calcul de la taille de l'échantillon requise pour atteindre un niveau de précision donné est relativement aisé quand on traite un échantillonnage aléatoire simple (E.A.S.). Mais des complications apparaissent quand d'autres types d'échantillonnage sont impliqués. C'est le cas pour le problème qui nous préoccupe.

Les déviations par rapport à un E.A.S. résultent :

- de l'usage possible d'informations supplémentaires (taille) ;
- d'une stratification plus ou moins imposée (catégories commerciales, facteurs spatio-temporels, points de débarquement) ;

- de la présence d'unités d'échantillonnage de différents types (jours de vente, caisses). Les gains ou pertes éventuels de précision qu'entraînent ces différents types d'échantillonnages seront évalués sur une population test.

- D'autre part, la précision est toujours relative à un estimateur donné. Nous examinerons dans cette étude la précision obtenue sur certains produits résultant de l'analyse de cohortes, technique couramment utilisée quand on dispose d'une évaluation des structures démographiques.

Les problémes de biais n'ont été abordés que dans la mesure où il s'agit de biais dûs aux choix des estimateurs. Nous avons en effet supposé tout au long de cette note que les unités mesurées avaient été échantillonnées aléatoirement selon un processus rigoureux.

En fait, la mise en oeuvre de telles méthodes, n'est pas toujours possible. Elle est le plus souvent irréalisable quand il s'agit de sélectionner des caisses de poisson.

On peut espérer que les biais qui peuvent en résulter sont relativement faibles dans la mesure où l'on dispose d'une stratification par catégories commerciales.

I - MODELES ET FORMULES DE BASE

A - Le double échantillonnage

L'intérêt du double échantillonnage est analysé sur un E.A.S., les avantages ou inconvénients de cette technique devant se conserver avec des types d'échantillonnage plus complexes.

Pour estimer des structures démographiques, deux procédures sont envisageables :

a) prélèvement d'un échantillon aléatoire simple et détermination directe de l'âge sur les individus sondés ;

b) la méthode du double échantillonnage ou échantillonnage en deux phases par laquelle on utilise l'information supplémentaire fournie par les distributions de longueur pour établir les structures démographiques.

Pour un âge donné, l'efficacité de la méthode dépendra de l'amplitude des variations des proportions dans les différentes strates de longueur. On constate cependant que dans beaucoup de cas la précision relative du double échantillonnage pour stratification sur l'aléatoire simple peut varier de 100 à plus de 200 % selon l'âge et la saison (MACKETT 1963). Les gains peuvent donc être apréciables et justifient l'utilisation de la méthode.

Les strates de longueur ayant été formées à partir des résultats de l'échantillonnage primaire, 3 types de sous-échantillonnage pour la détermination de l'âge peuvent être envisagés :

- prélèvement d'un nombre déterminé, constant dans chaque strate ;

- prélèvement d'un nombre proportionnel à l'effectif de la strate ;
- allocation optimale.

Cette dernière option ne paraît pas réaliste pour plusieurs raisons :

- manque d'informations fiables sur les variances ;
- difficulté d'optimiser sur plusieurs âges à la fois ;
- enfin pour un niveau donné d'échantillonnage, le gain sur l'allocation proportionnelle est faible si les proportions sont comprises entre 0,1 et 0,9 (COCHRAN, 1977) ce qui sera souvent le cas.

D'un point de vue purement statistique, le choix entre allocation fixe et allocation proportionnelle ne devrait pas se poser, cette dernière procédure étant en général plus efficace (KIMURA, 1977).

D'un point de vue technique, nous sommes enclin à penser que le choix entre l'une ou l'autre stratégie dépend de l'objectif poursuivi.

S'il s'agit d'établir une courbe de croissance, une bonne représentation de tous les âges sera mieux assurée par un prélèvement constant par strate.

S'il s'agit d'obtenir une distribution des captures par âge afin d'alimenter des modèles d'évaluation, l'allocation proportionnelle est sans doute suffisante et donc au total préférable.

Le but de ce travail étant de dégager quelques règles pouvant être appliquées aux opérations d'échantillonnage en routine, nous nous plaçons dans la deuxième perpective. Nous retiendrons donc comme modèle de base le double échantillonnage avec allocation proportionnelle.

1) Rappel des formules

Pour un âge t donné, et si on néglige les termes de correction pour population finie, on aura

$$\hat{p} = \sum_{\ell}^{L} w\ell \, \hat{p}\ell$$

$$\mathrm{Var}(\hat{p}) = \sum_{\ell}^{L} \frac{W\ell \, P\ell \, Q\ell}{\lambda \, n_1} + \sum_{\ell}^{L} \frac{W\ell (P\ell - P)^2}{n_1} \qquad \text{(Cochran, 1977)} \quad (1)$$

ℓ = indice des strates de longueur

$\frac{N\ell}{N} = W\ell$ = proportion de la population dans la strate de longueur ℓ

$\frac{n\ell}{n} = w\ell$ = proportion de l'échantillon dans la strate de longueur ℓ

P = proportion d'individus d'âge t donné dans la population

$Q = 1 - P$

$P\ell$ = proportion d'individus d'âge t donné dans la strate ℓ

$Q_\ell = 1 - P_\ell$

$\hat{p}$ = estimateur "stratifié" de P

$\hat{p}\ell$ = estimateur de $P\ell$

L = nombre de strates de longueur

n_1 = nombre d'individus mesurés (1ère phase)

λ = proportion d'individus âgés $(\lambda \leqslant 1)$

$n_2 = \lambda \, n_1$ = taille du sous-échantillon d'ageage

2) - Précision relative de l'aléatoire et du stratifié

Ecrivons l'équation de l'analyse de variance pour la population

$$(N - 1)\, S^2 = \sum_{\ell}^{L} (N\ell - 1) S_\ell^2 + \sum_{\ell}^{L} N_\ell \, (P_\ell - P)^2 \qquad (2)$$

$$S^2 = \frac{N}{N-1} \, PQ \quad = \quad \text{Variance totale}$$

$$S_\ell^2 = \frac{N\ell}{N\ell - 1} \, P_\ell \, Q_\ell \quad = \quad \text{Variance dans la strate } \ell$$

L'équation (2) peut donc s'écrire :

$$PQ = \sum_{\ell}^{L} W_\ell \, P_\ell \, Q_\ell + \sum_{\ell}^{L} W_\ell \, (P_\ell - P)^2 \qquad (3)$$

Posons $\lambda . n_1 = n_2$

$$\frac{PQ}{n_2} = \frac{\sum_{\ell}^{L} W_\ell \, P_\ell \, Q_\ell}{n_2} + \frac{\sum_{\ell}^{L} W_\ell \, (P_\ell - P)^2}{n_2} \qquad (4)$$

Soit en combinant (1) et (4) :

$$\frac{PQ}{n_2} = \text{Var}\,(\hat{p}) + \sum_{\ell}^{L} W_\ell \, (P_\ell - P)^2 \left(\frac{1}{n_2} - \frac{1}{n_1}\right) \qquad (5)$$

$\frac{PQ}{n_2}$ est l'expression de la variance de $\hat{p}$ pour un échantillonnage aléatoire simple de taille n_2.

L'équation (5) montre clairement que le mesurage préalable pour stratification sur les longueurs produit un gain de précision, sauf cas invraisemblable où l'on aurait $P_\ell = P$ pour toutes les classes de longueur.

Suivant la formulation de YATES (1981) on peut écrire la variance de l'estimateur $\hat{P}$

$$\mathrm{Var}\ (\hat{p}) = \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\ (1 - \lambda) \qquad (6)$$

Avec

$$S_1^2 = PQ$$

$$S_2^2 = \sum_{\ell}^{L} W_\ell\ P_\ell\ Q_\ell \qquad (7)$$

En posant : $K^2 = S_2^2 / S_1^2$, on obtient l'expression équivalente

$$\mathrm{Var}\ (\hat{p}) = \frac{S_1^2}{n_2}\ (\lambda + K^2\ (1 - \lambda))$$

3) Minimisation de la fonction des coûts

Prenons,

$$C = n_1\ c_1 + n_2\ c_2 \qquad (15)$$

C_1 et C_2 désignant les coûts unitaires respectifs de la première et de la deuxième phase, soit ici du mesurage et de l'ageage.

Minimiser C pour une précision donnée conduit aux relations :

$$\lambda^2 = \frac{c_1}{c_2}\ \frac{K^2}{1 - K^2}$$

$$n_2 = \frac{S_1^2}{\mathrm{Var}\ (p)} \left[\lambda + K^2\ (1 - \lambda) \right] \qquad (8)$$

La valeur optimale de λ est indépendante de la précision recherchée Elle varie par contre avec les groupes d'âge. Pour un même groupe d'âge elle varie aussi suivant la structure de la population.

Il est très probable que pour chaque groupe d'âge représenté dans l'échantillon on obtiendra des valeurs de λ différentes. On sera donc souvent contraint de choisir une valeur unique λ_o qui soit le meilleur compromis possible compte tenu de l'objectif poursuivi.

KIMURA (1977) qui étudie l'efficacité relative des sous-échantillonnages fixe et proportionnel utilise comme critère de choix la somme des variances des différents groupes d'âge. Ce critère, utile pour résoudre un problème précis, ne paraît pas répondre à nos préoccupations. En effet, la précision de l'estimation d'une structure démographique n'a qu'une valeur relative. Nous sommes plus intéressés de connaître comment la précision obtenue à ce niveau va se répercuter sur le résultat final, produit d'un modèle particulier. Nous utiliserons comme modèle l'analyse de cohortes, limitant toutefois notre réflexion à une seule cohorte et à l'examen de l'estimation du recrutement et du vecteur mortalité. Le sujet a déjà été abordé par certains auteurs, notamment J.G. POPE (1972) et Per SPARRE et al. (1977). Ces derniers utilisent l'équation des captures pour estimer les vecteurs F et un vecteur moyen $\bar{F}$. L'estimateur de la variance de $\bar{F}$ joue le rôle de fonction objectif qu'ils se proposent d'optimiser.

L'approche de J.G. POPE par l'analyse des cohortes nous paraît plus manipulable pour une étude globale du problème.

B) L'analyse de cohorte

1) Le modèle

Nous rappelons ci-dessous les deux formules de base de l'analyse de cohorte. Nous reprenons pour l'essentiel les expressions données par POPE dans son papier, avec une modification de l'indiçage en ce qui concerne l'expression des effectifs :

$$N_j = N_{j+1} \exp(M) + C_j \exp(M/2) \tag{9}$$

$$\exp(F_j) = N_j \exp(-M) / N_{j+1} \tag{10}$$

Désignant par K le nombre de groupes d'âge dans la pêcherie et par N_1 l'effectif initial du premier groupe, on peut écrire à partir de (9)

$$N_j = C_j / E \qquad j = K \tag{11}$$

$$N_j = \frac{C_K}{E} \exp\left[(K - j) M\right] + \sum_{i=0}^{K-j-1} C_{j+i} \exp[(2i+1)M/2], \quad j = 1, K-1$$

avec E = taux d'exploitation du dernier groupe d'âge.

L'intérêt de cette formulation étant que l'effectif N_j apparaît comme une fonction exclusive des captures, valeurs estimées, et de M et E supposés connus.

Notre propos étant d'évaluer l'impact des fluctuations d'échantillonnage sur les estimations des F_j et des N_j on ne s'interesse qu'à la variabilité de C_j autour de la vraie valeur débarquée. Dans l'analyse de cohorte les estimations $\hat{C}_j$ et $\hat{C}_i$ portent sur deux années différentes et sont statistiquement indépendantes.

On pourra donc écrire :

$$\mathrm{Var}\,(N_j) = \mathrm{Var}\,(C_j)\,/\,E^2 \qquad j = \kappa$$

$$\mathrm{Var}\,(N_j) = \frac{\mathrm{Var}\,(C\kappa)}{E^2} \times \exp\left[2(\kappa - j)\,M\right] + \sum_{i=0}^{\kappa-j-1} \mathrm{Var}\,(C_{j+i}) \exp\left[(2i+1)\,M\right] \qquad j = 1..,\kappa - 1 \qquad (12)$$

Pour estimer la variance des F_j nous utiliserons sans modification l'expression donnée par POPE dans son article déjà cité (1972), qui se déduit sans difficulté des formules de base :

$$\mathrm{Var}(F_j) = CV^2\,(\hat{N}_j) + CV^2\,(\hat{N}_{j+1}) - 2\exp\,(M)*CV^2\,(N_{j+1})\,\frac{N_{j+1}}{N_j} \qquad (13)$$

Le symbole $CV^2\,(\hat{N}_s)$ désigne le carré du coefficient de variation de $\hat{N}_s$.

Utilisant la relation (10) le carré du coefficient de variation de F_j aura pour expression :

$$CV^2(F_j) = \frac{CV^2\,(N_j) + CV^2\,(N_{j+1}) - 2\,CV^2\,(N_{j+1})\,e^{-F_j}}{F_j^2} \qquad (14)$$

Si les coefficients de variation de N_j et N_{j+1} sont du même ordre de grandeur on a la relation approchée :

$$CV^2\,(F_j) \simeq CV^2\,(N_j)\,2\,\frac{1 - \exp\,(-F_j)}{F_j^2} \qquad (15)$$

On peut vérifier que le coefficient de $CV^2(N_j)$ n'est inférieur à 1 que pour des valeurs de F $\geqslant 1,2$, limite, selon POPE, au delà de laquelle la validité de son approximation n'est plus assurée. On peut s'attendre à ce que dans la majorité des cas la variabilité relative des F soit plus importante que celle des N associés. Pour des motifs d'ordre statistique mais aussi en raison de l'importance du vecteur F dans les études de dynamique, l'attention devrait se porter en priorité sur ce paramètre.

2) Précision des estimations fournies par l'analyse des cohortes

a) Variance des effectifs

Si l'on désigne par Tj la capture totale l'année où la cohorte étudiée appartient au jème groupe d'âge la variance de l'effectif N_j pourra être estimée par la relation

$$\mathrm{Var}\,(\hat{N}_j) = \frac{T^2{}_\kappa \,\mathrm{Var}\,(\hat{P}_\kappa)}{E^2} \exp[\,2(\kappa-j)\,M\,] + \sum_{i=0}^{\kappa-j-1} T^2_{j+i}\,\mathrm{Var}\,(\hat{P}_{j+i}) \exp[\,(2i+1)\,M] \qquad j = 1..,\kappa-1$$

$$\mathrm{Var}\,(\hat{N}_j) = \frac{T^2{}_\kappa \,\mathrm{Var}\,(\hat{P}_\kappa)}{E^2} \qquad j = \kappa \qquad (16)$$

$$\text{avec } \mathrm{Var}\,(\hat{P}_j) = \frac{S_1^2 j}{n_1} + \frac{S_2^2 j}{n_2}\left(1 - \frac{n_2}{n_1}\right) = \frac{S_1^2 j - S_2^2 j}{n_1} + \frac{S_2^2 j}{n_2} = \frac{A_j}{n_1} + \frac{B_j}{n_2}$$

On obtient une expression de la variance de Nj de la forme

$$\mathrm{Var}\,(N_j) = \frac{Q_j}{n_1} + \frac{P_j}{n_2} \qquad (17)$$

Assimilant le coût d'échantillonnage de la cohorte au coût d'échantillonnage sur une année et considérant des coûts moyens C_1 et C_2 constants quel que soit l'âge on peut écrire :

$$C = n_1 c_1 + n_2 c_2$$

On obtient comme valeur optimale de $\lambda_j = \frac{n_2}{n_1}$

$$\lambda^2{}_j = \frac{C_1}{C_2}\,\frac{P_j}{Q_j}$$

Si l'on s'interesse plus particulièrement à la valeur du recrutement N_1 il serait logique d'utiliser la valeur λ_1. Toutefois, si l'on s'intéresse au vecteur F ou à certaines portions de ce vecteur un autre choix pourra être fait. Quoiqu'il en soit on pourra vérifier sur l'exemple qui suit que la sensibilité de la précision au choix de λ est faible.

Exemple

Nous avons reconstitué des populations fictives de captures à partir de distributions de longueur des débarquements (1977-1982) du stock nord de Merlu et d'une clé taille-âge provenant du laboratoire de La Rochelle (DECAMPS - LABASTIE, 1976).

Le tableau (1) donne les valeurs de λ, pour des rapports de coût de .02 et .01, se rapportant aux effectifs successifs de la cohorte (E terminal = .5).

Pour un rapport de coût donné on peut choisir une valeur moyenne de λ qui dans cet exemple pourrait être .13 ou .09 selon le rapport des coûts retenu. Par ailleurs, dans la mesure où l'intervalle d'estimation que nous avons retenu pour le rapport des coûts est réaliste on pourrait prendre $\lambda = .11$ comme valeur moyenne du taux de sous-échantillonnage.

Dans le tableau (2) nous avons reporté les coefficients de variation des estimations des effectifs N_j pour une valeur arbitraire $n_2 = 200$ et pour les valeurs λ_j correspondant à l'optimum. Pour l'année 1977, nous avons figuré les coefficients de variation obtenus en utilisant une valeur moyenne $\lambda = .10$. On peut vérifier que les valeurs des coefficients de variation sont très voisines des valeurs optimales. Pour l'un des cas on a une légère amélioration de la précision dûe à un surcroît de l'échantillonnage des longueurs. On obtient le phénomène inverse dans l'autre cas, la taille $n_1 = 2\ 000$ étant inférieure à la taille optimale.

Nous avons supposé plus haut la constance des coûts de mesurage et ageage quelles que soient les classes de taille. Cet exemple montre qu'à travers l'analyse de cohorte les effets de différences éventuelless peuvent être relativement faibles.

b) Variance des Fj

Dans le tableau (2) figurent les coefficients de variation des $\hat{F}_j$ estimés à partir de la formule (6).

Deux observations peuvent être faites :

- la précision des estimations augmente quand on remonte vers les premiers âges. Ceci est vrai également pour les effectifs ;

- les estimations de F_j sont plus sensibles aux aléas de l'échantillonnage que celles des N_j.

Ceci implique que les exigences sur le vecteur F seront déterminantes pour le choix du niveau d'échantillonnage.

De ce point de vue on pourrait être tenté de se donner pour les valeurs de F subterminales une précision désirée et estimer le taux d'échantillonnage à réaliser pour atteindre cet objectif.

Toutefois deux considérations doivent être prises en compte :

- les valeurs de F subterminales peuvent être assez fortement influencées (biaisées) par le choix de F terminal ;

- les incertitudes sur la détermination de l'âge des gros individus viennent se cumuler avec les effets des fluctuations de l'échantillonnage.

Dans les deux cas il est inutile de réduire drastiquement ces derniers en augmentant l'effort d'échantillonnage si les deux autres sources de variation ne peuvent être maîtrisées.

D'autre part dans la plupart des pêcheries, les vieux individus constituent une part relativement restreinte du total des débarquements. Il semble donc qu'on ait intérêt à travailler sur la valeur moyenne des F des classes les mieux représentées dans les captures.

On aura les relations (Annexe 1)

$$\bar{F}_s = \sum_{j=s}^{s+r-1} F_j / r$$

$$CV^2(\bar{F}_s) = \frac{CV^2(N_s) + CV^2(N_{s+r}) - 2\, CV^2(N_{s+r}) \exp(-r\,\bar{F}_s)}{(r\,\bar{F}_s)^2} \qquad (18)$$

Dans notre exemple et pour l'année 1977 on trouve pour les moyennes sur les âges de 2 à 6 ans inclus (rapport des coûts .02)

$\hat{\bar{F}}_{2-6} = .292$; $CV(\hat{\bar{F}}) = .163$

Cette valeur a été obtenue pour $n_2 = 200$ avec $n'_2 = \mu n_2$ on obtiendra le résultat classique $CV'(\hat{\bar{F}}) = CV(\hat{\bar{F}}) / \sqrt{\mu}$

c) Discussion-Extension à une structure stratifiée

Dans ce qui précède nous nous sommes placés au niveau d'une strate. La procédure peut être généralisée à l'ensemble de plusieurs strates. En effet, indiçant les strates par s on aura, pour un âge donné a

$\hat{C}a = N\,\hat{P}a \qquad \hat{P}a = \sum_s W_s\, \hat{P}_{as}$

$W_s = \frac{Ts}{T}$ représente le poids relatif de chaque strate.

$$\mathrm{Var}\ (\hat{P}_{as}) = \frac{{}_aS^2_{1s} - {}_aS^2_{2s}}{n_{1s}} + \frac{{}_aS^2_{2s}}{n_{2s}}$$

$$\mathrm{Var}\ (\hat{P}_a) = \Sigma\ W^2_s \quad \mathrm{Var}\ (\hat{P}_{as})$$

On obtiendra comme variance de l'effectif à l'âge t une expression de la forme

$$\mathrm{Var}\ (\hat{N}_t) = \sum_s \frac{{}_tA'_s}{n_{1s}} + \sum_s \frac{{}_tB'_s}{n_{2s}}$$

$${}_tQ'_s = W^2_s\ {}_tA_s \qquad {}_tB'_s = W^2_s\ {}_tB_s$$

Posant $C = \Sigma\ C_{os} + \sum_s n_{1s}\ C_1 + \sum_s n_{2s}\ C_2$, les procédures classiques d'optimisation conduisent aux relations :

$${}_t\lambda^2_s = \frac{{}_tB_s}{{}_tA_s}\ \frac{C_1}{C_2} \qquad s = 1.., S$$

Les taux d'échantillonnages optimaux dans chaque strate sont indépendants du poids de la strate et du coût constant C_{os} afférent à chaque strate.

On pourra comme précédemment choisir une valeur unique λ_s pour chaque strate.

L'allocation optimale peut être déterminée par les formules classiques (COCHRAN, 1977)

$$\frac{n_{2s}}{n_2} = \frac{W_s\ \sqrt{{}_tV_s\ /\ C_s}}{\Sigma\ W_s\ \sqrt{{}_tV_s\ /\ C_s}}$$

avec $${}_tV_s = \lambda_s\ {}_tA_s + {}_tB_s \qquad C_s = \frac{C_1}{\lambda_s} + C_2$$

Pour un coût fixe C

$$n_2 = \frac{(C - \sum_s C_{os})\ \sum_s W_s\ \sqrt{{}_tV_s\ /\ C_s}}{\sum_s W_s\ \sqrt{{}_tV_s\ C_s}}$$

Quelques observations peuvent être déduites des résultats qui précèdent.

Les valeurs λ optimales pour chaque strate seront les mêmes que celles obtenues par une analyse séparée sur chacune d'elle.

Les paramètres dominants en ce qui concerne la répartition de l'effort d'échantillonnage seront les poids relatifs des strates.

Il ressort de ces observations qu'une répartition proportionnelle de l'effort d'échantillonnage au poids de chaque strate devrait assurer une bonne approximation de la répartion optimale.

Rappelons à ce propos deux règles empiriques données par COCHRAN (1977) et valables quand on travaille sur des proportions : si les coûts par unité mesurée sont supposées les mêmes dans chacune des strates,

1) l'allocation optimale, la taille totale de l'échantillon étant fixée, gagne peu sur l'allocation proportionnelle si tous les P_i se situent entre .1 et .9.

2) le gain en précision de l'échantillonnage aléatoire stratifié sur l'aléatoire simple est faible à moins que les P_i varient grandement d'une strate à l'autre.

Nous avons reporté dans le tableau 3 les résultats de l'analyse portant sur des distributions considérées comme provenant de 6 strates, sous l'hypothèse d'un rapport unique ageage mesurage pour toutes les strates et d'une répartition égalitaire de l'effort d'échantillonnage entre les strates.

Sur le même tableau figurent les résultats attendus d'un échantillonnage aléatoire simple effectué sur les mêmes populations.

Les précisions obtenues sont équivalentes. On peut préciser qu'un léger gain aurait pu se manifester en faveur du stratifié si nous avions effectué une répartition proportionnelle des mesures dans les strates.

II - UNITES D'ECHANTILLONNAGE DE DIFFERENTS TYPES

Il arrivera le plus souvent que dans une même strate on soit conduit à réaliser plusieurs opérations d'échantillonnage correspondant à des jours de vente différents. Dans un premier temps on sélectionnera les jours de vente qui constitueront les unités primaires. A l'intérieur de chaque unité primaire sélectionnée on réalisera un sous-échantillonnage d'unités secondaires à partir desquelles seront bâtis les estimateurs.

Or il faut savoir que pour obtenir une précision donnée à partir d'un échantillonnage à plusieurs niveaux, il faudra en général prélever plus, et souvent beaucoup plus, d'unités secondaires qu'il ne serait nécessaire si le prélèvement était réalisé par un échantillonnage aléatoire simple. Ceci peut être aisément mis en évidence.(Annexe 2).

Le sous échantillonnage est en général moins efficace par ce que le plus souvent les grappes qui constituent les unités primaires sont formées d'individus présentant certaines affinités. Toutefois si les grappes sont construites par un processus entièrement aléatoire, il y aura équivalence avec un échantillonnage aléatoire simple de même importance.

Cependant pour que l'équivalence puisse être retenue il faudrait que le nombre de grappes soit relativement important. Ce ne sera pas le cas pour le nombre de jours de vente à l'intérieur d'une strate spatio-temporelle.

Par contre, il devrait être possible d'éliminer les caisses en tant que niveau d'échantillonnage. Ceci doit entraîner une simplification dans la formulation des estimateurs.

A) Les estimateurs classiques

Le choix d'une stratégie d'échantillonnage repose, entre autres sur deux types d'options :

- le choix d'une méthode de sélection des unités étudiées,
- le choix des estimateurs.

En ce qui concerne le premier point, nous avons implicitement admis une sélection des unités échantillonnées par une procédure aléatoire assurant à chaque unité la même probabilité d'être tirée. D'autres choix peuvent être faits pour améliorer la précision des estimateurs. Parmi ceux-ci rappelons la sélection des unités primaires avec des probalités proportionnelles à la taille ; cette dernière étant mesurée par le nombre d'unités secondaires.

Il est évident que le choix d'une telle procédure est exclu pour la sélection des jours de vente. Elle peut être envisagée dans le cas où les points de débarquements ne sont pas considérés comme des strates mais comme des niveaux et soumis à un sous-échantillonnage. Dans ce qui suit, nous conserverons le principe des probabilités égales.

Parmi les estimateurs possibles, les trois couramment rencontrés dans les problèmes de sous-échantillonnage sont :

l'estimateur rapport : $\bar{\bar{y}}_R = \dfrac{\int M_i \bar{y}_i}{\int M_i}$

l'estimateur simple : $\bar{\bar{y}}_S = \dfrac{\int \bar{y}_i}{n}$

l'estimateur pondéré : $\bar{\bar{y}}_r = \dfrac{1}{n} \dfrac{\int M_i \bar{y}_i}{\bar{M}}$

D'après JESSEN (1978) aucun de ces estimateurs n'est à préferer a priori. Il existe des cas où l'un d'entre eux sera plus précis que les autres. La qualité de ces estimateurs dépendra pour beaucoup des relations existant entre les $\bar{Y}_i$ et les M_i . Elle est mesurée par l'écart quadratique moyen (SCEM). Relativement à ce critère, le meilleur choix pourra être $\bar{\bar{y}}_s$, $\bar{\bar{y}}_p$ ou $\bar{\bar{y}}_R$ selon que la régression des $\bar{Y}_i$ sur les M_i est nulle, négative, ou positive.

1) L'estimateur rapport

Nous supposerons dans ce qui suit que le nombre total de poissons débarqués le jour i, soit T_i est connu.

On aura donc pour un âge donné et pour l'unité i

$$\hat{P}_i = \sum_{\ell} w_{i\ell} \hat{P}_{i\ell}$$

$$\text{Var } (\hat{P}_i) = \frac{S_{1i}^2 - S_{2i}^2}{n_{1i}} + \frac{S_{i2}^2}{n_{2i}}$$

formules déjà rencontrées dans la première partie.

Pour l'estimation de la proportion dans la strate on pourra utiliser l'estimateur rapport dont l'efficacité est très souvent vérifiée :

$$\hat{P}_R = \frac{\int T_i \hat{P}_i}{\int T_i} \qquad \hat{C} = T \hat{P}$$

Si l'on désigne par Q le nombre d'unités primaires (jours de vente) dans la strate et par q le nombre d'unités échantillonnées, on obtient les résultats classiques

$$\text{Var } (\hat{P}_R) \doteq \frac{1}{q \bar{T}^2} (1-f_1) \frac{\sum_i T_i^2 (P_i - P)^2}{Q - 1} + \frac{1}{q Q} \sum_i \left(\frac{T_i}{\bar{T}}\right)^2 \text{Var}(\hat{P}_i) \quad (20)$$

Note $\int$ désigne une sommation sur les échantillons
Σ une sommation sur la population

$$P = \frac{\Sigma\ T_i\ P_i}{\Sigma\ T_i}\ , \quad f_1 = \frac{q}{Q}$$

$$\mathrm{Var}\ (\hat{C}) \doteq \frac{Q^2}{q}\ (1-f_1)\ \frac{\Sigma\ T_i^{\ 2}(P_i - \bar{P})^2}{Q - 1} + \frac{Q}{q}\ \Sigma\ T_i^{\ 2}\ \mathrm{Var}\ (\hat{P}_i)$$

Les estimateurs $\hat{P}_R$ et Var $(\hat{P}_R)$ sont biaisés.

2) L'estimateur simple

$$\hat{P}_1 = \frac{1}{q}\ \int\ \hat{P}_i \tag{21}$$

On utilise donc comme estimateur de P, la moyenne simple des estimations des P_i qui auront été obtenues au cours de chacune des q opérations de mesurage-ageage.

Cet estimateur est biaisé ; en effet $\hat{P}_S$ est un estimateur de $P_S = \Sigma\ P_i\ /\ Q$ alors que la valeur exacte de P dans la population est donnée par la relation $P = \Sigma\ T_i\ P_i / \Sigma\ T_i$

$$\mathrm{Var}\ (\hat{P}_S) = \frac{1}{q}\ (1 - f_1)\ \frac{\Sigma\ (P_i - P_S)^2}{Q - 1} + \frac{1}{Qq}\ \Sigma\ \mathrm{Var}\ (\hat{P}_i) \tag{22}$$

On peut vérifier (annexe 3) que l'estimateur simple correspond à une certaine utilisation d'une clé taille-âge cumulée sur une période.

3) L'estimateur pondéré

$$\hat{P}_p = \frac{1}{q}\ \int\ T_i\ \hat{P}_i / \bar{T}$$

C'est un estimateur non biaisé de P.

$$\mathrm{Var}\ (\hat{P}_p) = \frac{1}{q}\ (1 - f_1)\ \frac{\Sigma(T_i\ P_i/\bar{T} - P)^2}{Q - 1} + \frac{1}{Qq}\ \Sigma \left(\frac{T_i}{\bar{T}}\right)^2 \mathrm{Var}\ (\hat{P}_i)$$

B) Discussion

Les estimateurs seront évalués en fonction des valeurs comparées des estimations des effectifs N et des mortalités F et des variances de ces estimations.

1) Estimation des effectifs des N et des mortalités F

L'estimateur rapport et l'estimateur simple sont biaisés mais le deuxième est "incorrect" en ce sens que le biais est constant. Pour l'estimateur rapport au contraire le biais est fonction de $\frac{1}{q}$ et diminue donc quand le taux d'échantillonnage augmente.

Les résultats de ces effets ont été évalués sur une population fictive de $Q = 6$ unités primaires, chacune d'elles étant caractérisée par le couple Pai, Ti ; Pai désignant la proportion supposée connue d'individus d'âge "a" dans l'unité i ; Ti le nombre d'individus dans l'unité i.

A partir d'un échantillon de q unités primaires, l'estimateur rapport fournira une estimation biaisée de P ; l'estimateur simple fournira une estimation de Σ Pi/Q, inexacte à moins que tous les Ti soient égaux.

On trouvera dans le tableau 4 les valeurs des proportions des différents âges calculées à partir des formules précédentes et les résultats de l'analyse de cohorte qui en résulte. Les proportions calculées par la formule pondérée étant par définition correcte, on pourra mesurer pour chaque âge la valeur moyenne du biais attaché à l'estimateur simple.

Dans le tableau 5 nous avons reporté pour quelques âges les biais relatifs attendus sur les estimations des proportions des différents âges dans les débarquements. La valeur absolue du biais est égale, pour l'estimateur simple à $P - \frac{\Sigma Pi}{Q}$

Pour l'estimateur rapport il peut être évalué par la relation

$$E(\hat{P}_R - P) = \frac{1 - f}{q\,\bar{T}^2} \left(P S^2_T - CoV(T, Ti) \right) \quad \text{(Cochran 1977)} \qquad (26)$$

où Ti représente l'effectif du groupe d'âge **i.**
la valeur de ρ reportée dans le tableau 5 mesure la corrélation entre les Pi et les Ti.

On remarque que le biais de l'estimateur simple reste toujours très élevé sauf pour l'âge 2 pour lequel la valeur de ρ est relativement faible.

Pour que le biais relatif de l'estimateur rapport tombe en dessous de 4 % il faudrait échantillonner une dizaine d'unités primaires sur une strate en comportant une soixantaine, valeur approximative des jours de vente dans un trimestre.

Les biais relatifs induits par l'estimateur simple sur l'analyse de cohorte peuvent être importants en ce qui concerne les effectifs. De l'ordre de 20 % jusqu'à l'âge 3. On observe une chute pour les âges 2 et 1. Ceci explique les différences observées sur les F. Elles restent très faibles dans la mesure où les âges impliqués (formule 10) ont des biais du même ordre de grandeur. Les variations induites peuvent être importantes dans le cas contraire (15 % pour F_2).

Il reste que le biais de l'estimateur rapport peut toujours être maîtrisé par l'intermédiaire du taux d'échantillonnage. Celui de l'estimateur simple reste soumis aux aléas des distributions relatives des P_i et des T_i.

2) Précision des estimations

Du fait de l'existence d'un biais attaché à deux des estimateurs retenus, il sera préférable de considérer l'erreur quadratique moyenne (SCEM) plutôt que la variance pour comparer la précision de ces estimateurs :

$$SCEM = Variance + (biais)^2$$

Si l'on compare les expressions (20) et (24), mis à part le fait que la première n'est qu'une approximation, elles diffèrent essentiellement par la valeur du premier terme qui représente la mesure de la variabilité entre unités primaires. Dans le tableau 6, nous avons reporté la valeur prise par ce terme pour l'estimateur rapport et pour l'estimateur non biaisé. Nous avons également transcrit les valeurs du coefficient de correlation ρ entre les P_i et les T_i. Les résultats observés concordent avec les observations de JESSEN rappelées plus haut.

En effet, pour les âges 3 à 11 pour lesquels ρ est voisin de -1 l'estimateur non biaisé paraît plus intéresant d'autant que l'estimateur de la variance de l'estimateur rapport comporte en plus un biais qui est fonction de $\frac{1}{q}$.

Par contre, dès que la correlation entre P_i et T_i devient positive (âge 1 et 2) l'estimateur rapport peut reprendre l'avantage sous réserve que q soit suffisamment grand.

Dans le tableau 7, nous avons reporté les précisions induites sur les valeurs des effectifs et des F issus de l'analyse de cohorte selon le type d'estimateurs utilisé.

a) Estimateur simple

Le biais est dans ce cas aisé à évaluer. Pour l'effectif à l'âge 1 il est (en valeur relative) égal à 1.9232 - 1.8064 = .1168 ; soit .065 en terme de coefficient de variation. La part du coefficient de variation dûe au biais représente 30 % du coefficient de variation total de N_j pour q = 1 et 45 % pour q = 4. Les résultats ne sont pas meilleurs pour les autres âges.

En ce qui concerne les mortalités par pêche, les résultats pourront être très variables selon la valeur du biais des effectifs impliqués.

b) Estimateur non biaisé et estimateur rapport

Compte tenu des réserves faites plus haut et si l'on s'en tient à notre exemple, on constate que l'estimateur non biaisé est plus efficace quand il s'applique aux individus âgés. Pour les âges 1 et 2 l'estimateur rapport est plus précis. Toutefois, les écarts s'atténuent sensiblement dès que le taux d'échantillonnage des unités primaires augmente c'est-à-dire quand l'estimateur rapport devient relativement exact. Il faut rappeler ici que le biais de l'estimateur rapport sera d'autant plus important que la correlation entre T_i et P_i sera plus fortement négative.

Si l'on s'en tient exclusivement à la notion de précision des résultats il peut être intéressant d'examiner comment évolue la mesure de cette précision quand on fait varier q et n_2 de telle sorte que le produit qn_2 (total d'individus âgés) demeure constant. Théoriquement on doit s'attendre à une précision relative plus grande pour des valeurs de q élevées ; les gains seront d'autant plus importants que la composante inter-unités primaires de la variance sera plus forte.

On trouvera ci-dessous, les coefficients de variation de la valeur moyenne des F de l'âge 2 à 6 pour quelques combinaisons de q et n_2.

Pour l'estimateur non biaisé et $C_1/C_2 = .02$, on obtient $\overline{F}_{2-6} = .255$ et :

$CV(\overline{F}) = .206$	avec $q = 1$	n_2	$= 200$
$CV(\overline{F}) = .168$	avec $q = 2$	n_2	$= 100$
$CV(\overline{F}) = .146$	avec $q = 4$	n_2	$= 50$

L'exemple que nous avons pris est intéressant sur le plan théorique en ce sens qu'il nous permet de mettre en lumière quelques propriétés connues des trois estimateurs utilisés.

Les résultats du tableau 6 montrent que les correlations entre les P_i et les T_i sont positives pour les deux premiers âges et deviennent fortement négatives pour les individus plus âgés. Ce résultat n'est pas fortuit ici. En effet, les débarquements que nous avons utilisés représentent les débarquements totaux d'années successives. Ils sont probablement influencés par le niveau de recrutement des deux ou trois premiers groupes d'âge. Il s'ensuit que la proportion de ces derniers aura tendance à augmenter avec la population totale. Pour les individus âgés les proportions observées ont tendance à varier à la baisse quand le total débarqué augmente et vice-versa.

Sauf cas spéciaux on peut s'attendre à ce que les variations du volume global des apports successifs au cours d'une même année puissent être également dûes à des apports plus importants d'individus âgés bien que les variations dûes aux jeunes classes restent potentiellement plus probables.

Pour évaluer les mérites comparés des trois estimateurs quant à leur efficacité, on devra tenir compte de ce qui précède.

III - LA STRATIFICATION EN CATEGORIES COMMERCIALES

Dans cette section nous prenons en compte le fait que pour de nombreuses espèces les individus sont, préalablement à la vente, répartis en lots de tailles plus ou moins homogènes appelés catégories commerciales.

Un échantillonnage est réalisé qui tient compte de cette stratification. Pour un âge donné i on aura l'estimateur

$$\hat{P}_i = \Sigma \frac{M_h}{M} \hat{P}_{ih}$$

$\hat{p}_{ih}$ étant l'estimateur de la proportion d'âge i dans la strate h ; il représente un des estimateurs mentionnés dans la précédente section.

La variance de $\hat{p}_i$ aura comme expression

$$\text{Var}(\hat{P}_i) = \Sigma \frac{M_h^2}{M^2} \text{Var}(\hat{P}_{ih})$$

Nous continuerons de supposer que la structure de ces débarquements reste constante d'une année sur l'autre et que les estimations des P_i et P_j correspondant à des années différentes sont indépendantes.

D'autre part, nous avons également supposé un échantillonnage constant par strate et par niveau, soit q niveaux par strate et n_2 individus âgés par niveau.

On trouvera dans le tableau 8 les résultats qui peuvent être directement comparés à ceux du tableau 7. Nous avons toutefois abandonné l'estimateur simple dont le biais, on l'a vu, est par trop élevé.

D'autre part, nous devons préciser que la stratification en catégorie commerciale ne suit pas celles habituellement pratiquées en criée pour lesquelles les recouvrements de taille peuvent être relativement importants. Sur huit catégories commerciales pour le merlu à la Rochelle, nous avons retenu cinq catégories sans recouvrement de tailles qui correspondent à peu près aux catégories locales suivantes : 1 - 2 -(3 + 4 + 5 + 6) - 7 - 8.

Cette simplification ne devrait pas modifier sensiblement les effets de la stratification.

L'examen des tableaux 7 et 8 autorise quelques remarques :

1) les différences de précision, selon les groupes d'âge, entre les deux estimateurs s'estompent nettement. Ceci est normal dans la mesure où les effets induits par les variations des effectifs sont contenus à l'intérieur de chaque catégorie.

2) les gains en précision sont relativement peu importants sur les trois premiers groupes d'âge. Ils sont de l'ordre de 25 à 150 % au delà. On remarquera (tableau 9) qu'ils se répartissent sur toutes les composantes de la variance en ce qui concerne l'estimateur rapport. Pour l'estimateur non biaisé, les gains sur la composante inter unités primaires (VN3) sont négligeables ou nuls. Comme nous l'avons souligné plus haut ceci est sans doute la conséquence du choix que nous avons fait pour construire notre population test. Quoiqu'il en soit la stratification entraîne globalement une amélioration sensible de la précision.

3) les âges 1 et 2 restent très sensibles aux variations du nombre d'unités primaires échantillonnées. Cette sensibilité décroît avec l'âge. Il y aurait donc théoriquement intérêt à choisir des stratégies différentes par catégories commerciales.

4) la stratification en catégories commerciales fait apparaître un phénomène intuitivement perceptible, à savoir que l'information apportée par les mensurations est d'autant moins efficace, à rapport de coûts constant, qu'elle s'adresse aux gros individus. Ceci apparaît au travers des valeurs de $\lambda = n_2 / n_1$ qui croissent avec l'âge.

Il s'ensuit que la stratification devrait permetttre un meilleur ajustement du rapport ageage/mesurage.

Pour notre exemple, les valeurs moyennes de λ par catégorie commerciale pourraient être approximativement les suivantes :

catégories	âges	λ
8	1 - 2	1/8
7	3 - 4 - 5	1/4
autres	$\geqslant$ 6	1/3

En supposant constant et égal à 200 le nombre global d'individus âgés et une répartition égalitaire entre catégories commerciales, le nombre de mesures optimal serait de l'ordre de 900 alors qu'il s'élèverait à 1 600 en l'absence de stratification.

Il est évident que cet exemple n'est donné qu'à titre d'illustration et que seul un calcul d'optimisation pourra donner pour chaque catégorie commerciale les valeurs optimales de q, n_2 et n_1 .

5) le coefficient de variation de l'estimation $\overline{F}_{2-6}$ = . 255 peut-être évalué pour un total d'individus âgés égal à 200 et dans le cas de l'estimateur non biaisé, à :

. 155 pour q = 1 et n_2 = 200

. 126 pour q = 2 et n_2 = 100

. 090 pour q = 4 et n_2 = 50

soit un gain de 25 à 35 % sur les mêmes estimations faites dans la précédente section.

IV - DISCUSSION

Le cheminement que nous avons suivi nous a permis de rappeler et préciser quelques résultats qui peuvent se révéler très utiles pour l'élaboration d'une stratégie d'échantillonnage en criée.

a) Les catégories commerciales

Comme on vient de le voir la prise en compte des catégories commerciales est à même d'améliorer sensiblement la précision des résultats pour un effort d'échantillonnage constant.

Mais l'existence de catégories commerciales devrait aussi permettre de simplifier la stratégie d'échantillonnage.

En premier lieu, on doit pouvoir ignorer la variabilité due aux flottilles et aux secteurs de pêche retenant uniquement comme subdivisions :

- les catégories commerciales,
- les points de débarquements,
- les strates temporelles.

En effet, on peut admettre que l'essentiel de la différence entre secteurs ou entre métiers est mesuré par l'importance relative des catégories commerciales afférentes à chacune de ces sources. Il est vrai que pour une catégorie commerciale donnée les proportions d'âge peuvent varier suivant les secteurs ou les flottilles. Mais dans la mesure où le choix des unités échantillonnées est aléatoire cette variabilité ne devrait entraîner aucun biais sur les estimations globales. Des études particulières sur les variabilités inter-flottilles ou inter-secteurs devraient faire l'objet de plans d'échantillonnage adaptés.

De même l'existence de catégories commerciales devrait permettre de supprimer le niveau "caisse" et d'utiliser directement l'individu comme unité secondaire ; les résultats doivent être en moyenne équivalents dans la mesure où le remplissage et le choix des caisses sont aléatoires (cf. annexe 2). Cette procédure doit introduire quelques simplifications dans l'expression des formules des variances.

b) La structure spatio-temporelle

Les résultats qui précèdent s'appliquent à une strate spatio-temporelle. Dans la réalité nous aurons plusieurs strates de ce type représentant plusieurs points de débarquement et un fractionnement du temps.

Comme le souligne COCHRAN (1977), il n'est pas toujours vrai qu'une stratification de l'univers échantillonné entraîne une variance plus faible. Ceci est notamment le cas quand les taux d'échantillonnage sont loin de l'allocation optimale, pour un taux global d'échantillonnage fixé. Nous avons eu un exemple de ce fait dans la première section. Mais il arrive très rarement en pratique que cette éventualité se produise, même si elle est théoriquement possible, pour une allocation proportionnelle. Or on pourra toujours réaliser pour l'ensemble des strates considérées une allocation de l'échantillonnage se rapprochant suffisamment de l'optimum pour s'assurer au minimum d'une variance au plus égale à celle qui serait obtenue sur un univers non stratifié.

Ces remarques permettent d'envisager une procédure approchée d'optimisation dont la première étape pourrait être la répartition des moyens totaux disponibles C proportionnellement au poids relatif de chaque strate spatio-temporelle. L'optimisation à l'intérieur des strates peut être traitée suivant des méthodes classiques.

Toutefois cette approche suppose que l'un des termes du couple coût-précision a été privilégié. Mais si la priorité est donnée à la précision, définie par rapport à un objectif déterminé (recrutement, production, etc...) la démarche sera quelque peu différente.

Un choix doit être fait sur ce point.

c) L'analyse de cohorte

Soulignons tout d'abord que pour nos simulations nous nous sommes placés implicitement dans une situation d'équilibre de la pêcherie en supposant que la structure des débarquements par âge se conservait d'une année sur l'autre. Cette simplification ne devrait pas altérer la nature des phénomènes.

Les produits de l'analyse de cohorte, recrutements ou F moyens , fournissent des critères objectifs à partir desquels l'optimisation de la stratégie d'échantillonnage pourra être évaluée sous réserve d'obtenir une bonne estimation des coûts des différentes phases et opérations.

Cette étude a été faite sur le rapport ageage-mesurage que nous avons désignée par λ , et pour lequel l'optimum est indépendant de la précision requise et fonction du seul rapport des coûts afférents à ces deux opérations. Nous rappellons que nous nous sommes placés sous l'hypothèse d'un sous-échantillonnage pour ageage proportionnel aux effectifs mesurés par strate de longueur.

Selon le critère utilisé on obtient des valeurs de λ différentes, mais suffisamment proches pour qu'une valeur unique, suboptimale, puisse être utilisée sans déviations importantes du coût ou de la précision, pour chaque catégorie commerciale.

d) La clé taille-âge

Les estimateurs que nous avons envisagés, et l'expression de leur variance, supposent que le nombre d'individus par âge, ou leur proposition p_i , est estimé à chaque opération d'échantillonnage.

Or, dans la pratique, cette estimation est faite périodiquement à partir d'une clé taille-âge cumul des observations sur une strate temporelle. Nous l'avons vu ceci revient implicitement à utiliser un estimateur simple et peut être la source de biais importants si la structure par âge des débarquements est liée à la quantité de poissons débarqués. Par contre les biais restent mineurs si cette structure est relativement stable. Eventualité non négligeable, dans le cadre d'une double stratification des débarquements sur le temps et les catégories commerciales.

La procédure utilisée et l'estimateur simple ne sont donc pas à rejeter définitivement d'autant que les deux autres estimateurs ne sont pas non plus sans inconvénients.

La structure de la clé taille-âge sera un facteur déterminant de la valeur de λ qui sera d'autant plus forte que le nombre de classes d'âge par classe de longueur sera plus élevé. Ceci se vérifie entre classes de taille d'une même espèce mais aussi entre espèces.

CONCLUSION

Cette note a essentiellement pour objectif d'apporter quelques informations sur les variations de précision liées au type d'échantillonnage réalisé et au choix des estimateurs. L'analyse a été faite sur des produits résultant de l'analyse de cohorte, recrutements ou F moyens. Cette approche nous a paru suffisante dans une phase descriptive.

Si l'étude doit conduire à des décisions concernant la définition de l'effort d'échantillonnage il sera nécessaire de pousser l'analyse plus avant et de mesurer les effets des choix possibles sur les produits finaux des techniques associées à l'analyse de cohorte et notamment sur les estimations des prévisions de capture.

Cette approche ne pourra se satisfaire, sans autre investigation, des simplifications que nous avons introduites dans cette note dont la plus importante est le fait de n'avoir considérée qu'une seule cohorte. L'expression des variances s'en est trouvée simplifiée. Mais dans la pratique les F moyens sont calculés sur plusieurs cohortes.

Dans la formation de ces valeurs moyennes interviendront des estimations de captures par âge portant sur la même année et donc statistiquement liées. Les formules de variance des estimateurs ne sont plus aussi simples que celles que nous avons utilisées. Il faut noter toutefois que les estimations des captures par âge pour une même année sont corrélées négativement. Aussi, il n'est pas certain que la méthode que nous avons suivie en ne considérant qu'une seule cohorte ait été avantageuse sur le plan de la précision des F moyens.

Référence

BIAIS (G.), 1985 .- Estimation de la répartition par âge des captures françaises de hareng. Cons. Int. Explor. Mer, C.M. 1985/H : 45

COCHRAN (W.G.), 1977 .- Sampling technique. Third edition. J. WILEY and Sons - New York.

JESSEN (R.J.), 1978 .- Statistical Survey Techniques - J. WILEY.

LAND (C.E.), 1971 .- Confidence intervals for linear fonctions of the normal mean and variance. Am. Math. Stat. 42, 1187-1205.

MACKETT (D.J.), 1963 .- A method of sampling the Pacific Albacore (Thunnus germo) CATCH for relative age composition. FAO Fish. Rep. (6), vol 3 : 1355-1366.

SPARRE (P.), KRUDSEN (H.) and MUNCH-PETERSEN (S.), 1977 Optimization of sampling programms for estimation of age distribution and fishing mortality Cons. Int. explor. Mer, CM 1977/F : 43

POPE (J.G.), 1972 .- An investigation of the accuracy of virtual Population Analysis Using Cohort Analysis Res. Bull. ICNAF 9, 1972

RISSER (R.) et TRAYNARD (C.E.), 1957 .- Les principes de la statistique mathématique LIVRE 1 Séries statistiques - 195 p. Gauthier-Villors.

YATES (F.), 1981 .- Sampling Methods for Censuses and Surveys. Fourth Edition - C. GRIFFIN and Co. London.

ANNEXE (1)

1) La relation $CoV(\hat{N}_j, \hat{N}_{j+1}) = Var(\hat{N}_{j+1}) * e^{M}$ (J.G. POPE, 1972) peut se généraliser à des effectifs distants de r unités de temps

$$CoV(\hat{N}_j, N_{j+r}) = Var(N_{j+r})\, e^{rM}$$

En effet,

$$CoV(\hat{N}_j, \hat{N}_{j+r}) = E\left\{\frac{(\hat{C}_\kappa - C_\kappa)}{E} \exp(\kappa - j) M + \sum_{i=o}^{\kappa-j-1} (\hat{C}_{j+i} - C_{j+i}) \exp[(2i+1) M/2]\right\}$$

$$\times \left\{\frac{\hat{C}_\kappa - C_\kappa}{E} \exp[\kappa-(j+r) M] \sum_{i=o}^{\kappa-t-1} (\hat{C}_{t+i} - C_{t+i}) \exp[(2i+1) M/2]\right\}$$

avec $t = j + r$

Si l'on retient l'indépendance des C_j l'espérance des produits $(\hat{C}_\ell - C_\ell)(C_n - C_n)$ avec $\ell \neq n$, sera nulle

On obtient donc :

$$CoV(\hat{N}_j, \hat{N}_t) =$$

$$\frac{Var(\hat{C}_\kappa)}{E^2} \exp[(2(\kappa - t) + r)M] + \sum_{i=o}^{\kappa-t-i} Var(\hat{C}_{t+i}) \exp[(2(r+i) + 1)M/2] \exp[(2i+1)M/$$

Soit :

$$CoV(\hat{N}_j, \hat{N}_t) = e^{rM} \left\{\frac{Var(\hat{C}_\kappa)}{E^2} \exp[(2(\kappa-t) M] + \sum_{i=0}^{\kappa-t-i} Var(\hat{C}_{t+i}) \exp[(2i + 1)] M\right\}$$

L'expression entre crochets représente la variance de $\hat{N}_t$.
D'où la relation annoncée.

2) D'autre part de la relation :

$$F_j = - M + L_n (N_j) - L_n (N_{j+1})$$

on déduit :

$$\sum_{j=s}^{s+r-1} F_j = - r M + L_n (N_s) - L_n(N_{s+r})$$

$$\text{Soit} \quad \bar{F} = - M + \left(L_n (N_s) - L_n (\hat{N}_{s+r}) \right) \frac{1}{r}$$

et pour les estimations :

$$\hat{\bar{F}} = - M + \left(L_n (\hat{N}_s) - L_n (\hat{N}_{s+r}) \right) \frac{1}{r}$$

On en déduit,

$$\text{Var} (\hat{\bar{F}}) = \frac{1}{r^2} \left(\frac{\text{Var}(\hat{N}_s)}{N_s^{\,2}} + \frac{\text{Var} (\hat{N}_{s+r})}{N_{s+r}^2} - \frac{2 \text{ CoV } (\hat{N}_s, \hat{N}_{s+r})}{N_s \cdot N_{s+r}} \right)$$

Soit compte tenu de la relation donnée plus haut :

$$CV^2(\hat{\bar{F}}) = \frac{CV^2(N_s) + CV^2(N_{s+r}) - 2 \; CV^2(N_{s+r}) \dfrac{N_{s+r}}{N_s} \; e^{rM}}{(r \bar{F})^2}$$

On notera que $\dfrac{N_{s+r}}{N_s}$ exp (rM) = exp $\left(- \sum_{j=s}^{s+r-1} F_j \right)$

ANNEXE (2)

Supposons une population constituée de N Unités primaires (grappes) constituées chacune de Mo éléments. On sélectionne n grappes et dans chacune on sous-échantillonne mo unités secondaires sur lesquelles on mesure une caractéristique y.

$$\bar{\bar{y}}_s = \sum_{i=1}^{n} \bar{y}_i / n \qquad \bar{y}_i = \sum_{j=1}^{mo} y_{ij}/mo$$

La variance de $\bar{\bar{y}}$ peut s'exprimer en fonction de la corrélation intra classe δ* (JESSEN, 1978) ; soit pour N suffisamment grand :

$$\text{Var}\ (\bar{\bar{y}}_s) \doteq \frac{1-f_1}{n\ Mo} S^2 \left[1 + (Mo - 1)\ \delta\right] + \frac{1-f_2}{n\ mo} S^2\ (1 - \delta)$$

$$f_1 = n/N \qquad f_2 = mo/Mo$$

S^2 représentant la variance dans la population totale,

$$S^2 = \frac{\sum_i \sum_j (y_{ij} - \bar{\bar{Y}})^2}{N - 1}$$

Pour un échantillonnage aléatoire simple (E.A.S) de taille nmo

$$\bar{\bar{y}}_a = \sum_i \sum_j y_{ij}/n\ mo$$

$$\text{Var}\ (\bar{\bar{y}}_a) = \frac{1-f_1 f_2}{n\ mo} S^2$$

* Correlation intra classe Le coefficient δ est défini par la relation :

$$\delta = \sum_{i=1}^{N} \frac{1}{N} \left\{ \frac{\sum_{j}^{Mo} \sum_{\kappa \neq j}^{Mo} (y_{ij} - \bar{Y})(y_{i\kappa} - \bar{Y})}{Mo\ (Mo-1) S^2} \right\}$$

L'efficacité relative de l'E.A.S. par rapport au sous-échantillonnage est donnée par la relation :

$$RE = \frac{V(\bar{\bar{y}}_s)}{V(\bar{\bar{y}}_a)}$$

$$RE = \frac{(1-f_1)f_2}{1-f_1f_2} \left[1 + (Mo - 1)\ \delta\right] + \frac{1-f_2}{1-f_1f_2}\ (1-\delta)$$

ou après manipulation

$$RE = 1 + \delta\ \left(\frac{Mo(1-f_1)\ f_2}{1-f_1-f_2} - 1\right)$$

1) Si $\delta = 0$, les deux stratégies sont équivalentes. Ceci arrivera si les unités primaires sont constituées par une procédure entièrement aléatoire.

2) Si f_1 et f_2 sont faibles, f_1f_2 est négligeable et ,

$$R \doteq 1 + \delta\,(mo - 1)$$

Il y a équivalence entre les deux stratégies pour mo = 1 c'est à dire, dans le cas où l'on prélève 1 unité secondaire dans chacune des n mo unités primaires sélectionnées.

Hors cette éventualité l'efficacité du sous-échantillonnage sera d'autant plus faible que δ sera proche de l'unité c'est à dire que les individus d'une même grappe seront plus ressemblants.

ANNEXE (3)

Dans l'hypothèse d'un échantillonnage en deux phases, nous avons vu que pour un âge donné et pour l'unité i l'estimation de P_i pourrait s'écrire :

$$\hat{P}_i = \sum_{\ell} w_{i\ell} \hat{P}_{i\ell}$$

avec $w_{i\ell} = \dfrac{n_{1i\ell}}{n_{1i}}$ $\qquad P_{i\ell} = \dfrac{m_{\ell i}}{n_{2\ell i}}$

$m_{\ell i}$ nombre d'individus d'âge.

$n_{2\ell i}$ nombre d'individus âgés dans la classe de longueur ℓ

Si on se place dans l'hypothèse d'un sous-échantillonnage proportionnel :

$$n_{2\ell i} = \lambda \quad n_{1\ell i}$$ soit,

$$n_{2i} = \lambda \quad n_{1i}$$ ou en faisant le rapport des deux égalités :

$$n_{2\ell i} = n_{2i} \, w_{i\ell}$$

On obtient donc comme estimation de P_i

$$\hat{P}_i = \sum_{\ell} \frac{m_{\ell i}}{n_{2i}}$$

Nous supposerons dans ce qui suit que :

1) $n_{2i} = n_2$ et $n_{1i} = n_1$ pour tout i

2) λ est constant

Si l'on travaille sur une clé taille-âge qui résulte du cumul de plusieurs observations i.

On aura comme estimation de la proportion P d'un âge donné

$$\hat{P} = \sum_{\ell} \frac{\int m_{\ell i}}{\int n_{2\ell i}} * \frac{\int n_{1\ell i}}{\int n_1}$$

En vertu des hypothèses faites :

$n_{2\ell i} = \lambda\, n_{1\ell i}$ d'où on déduit ,

$$\hat{P} = \sum_{\ell} \frac{\int m_{\ell i}}{\lambda\, q\, n_1} = \frac{1}{q} \sum_{\ell} \int \frac{m_{\ell i}}{n_2}$$

q représentant le nombre d'opérations ageage-mesurage réalisées au cours de la période de temps considérée.

Si on intervertit l'ordre des sommations,

$$\hat{P} = \frac{1}{q} \int \sum_{\ell} \frac{m_{\ell i}}{n_2} \qquad \text{Soit} \qquad \hat{P} = \frac{1}{q} \int \hat{P}_i$$

C_1/C_2	Années	1977	1978	1979	1980	1981	1982
	Age						
	1	. 107		. 117		. 133	
	2	. 114		. 124		. 134	
	3	. 128		. 137		. 139	
	4	. 130		. 143		. 142	
. 02	5	. 129		. 142		. 141	
	6	. 127		. 141		. 139	
	7	. 127		. 143		. 139	
	8	. 124		. 141		. 137	
	9	. 120		. 138		. 133	
	10	. 113		. 131		. 126	
	11	. 106		. 123		. 118	
Moyennes		. 120	. 120	. 134	. 129	. 134	. 129
	1	. 075		. 083		. 094	
	2	. 081		. 088		. 095	
	3	. 090		. 097		. 098	
	4	. 092		. 101		. 100	
. 01	5	. 091		. 101		. 099	
	6	. 090		. 100		. 098	
	7	. 090		. 101		. 099	
	8	. 088		. 099		. 097	
	9	. 085		. 097		. 094	
	10	. 080		. 092		. 089	
	11	. 075		. 087		. 083	
Moyennes		. 085	. 085	. 095	. 091	. 095	. 091

Tableau 1 : Variation de λ en fonction des coûts et des années

Rapports des Coûts : .02 200 individus âgés

Années	1977			1979			1980			1977	
AGE	nl	CV.N	CV.F	nl	CV.N	CV.F	nl	CV.N	CV.F	nl	CV.N
1	1876.6	.057	.070	1709.3	.061	.073	1498.3	.069	.106	2 000	.057
2	1749.8	.082	.118	1609.5	.081	.110	1490.2	.071	.085	-	.081
3	1564.3	.161	.239	1457.6	.141	.188	1439.0	.088	.123	-	.160
4	1543.4	.193	.347	1400.6	.188	.311	1413.1	.105	.182	-	.190
5	1553.3	.219	.428	1405.1	.221	.442	1422.8	.120	.230	-	.216
6	1575.8	.247	.388	1414.9	.244	.401	1440.0	.135	.217	-	.244
7	1572.0	.306	.555	1403.1	.289	.531	1435.2	.164	.292	-	.302
8	1613.9	.361	.645	1422.3	.335	.574	1462.6	.194	.335	-	.356
9	1668.5	.431	.931	1452.1	.406	.796	1507.0	.236	.474	-	.426
10	1768.1	.487	1.138	1529.4	.471	.967	1592.8	.271	.588	-	.482
11	1891.3	.539		1624.7	.540		1697.9	.307		-	.537
LAMBDA MOYEN		.120			.134			.134		.100	

Rapports des Coûts : .01 200 individus âgés

1	2654.0	.056	.068	2417.3	.060	.071	1119.0	.068	.096	2 000	.057
2	2474.6	.080	.115	2276.2	.079	.107	2107.5	.069	.083	-	.081
3	2212.2	.158	.234	2061.3	.138	.184	2035.0	.086	.121	-	.160
4	2182.6	.189	.340	1980.8	.185	.306	1998.5	.103	.179	-	.190
5	2196.7	.215	.420	1987.2	.217	.434	2012.2	.118	.227	-	.216
6	2228.6	.242	.380	2000.9	.240	.393	2036.5	.133	.212	-	.244
7	2223.2	.300	.545	1984.2	.284	.522	2029.6	.161	.287	-	.302
8	2282.4	.353	.633	2011.5	.328	.564	2068.4	.190	.329	-	.356
9	2359.6	.422	.916	2053.6	.398	.784	2131.2	.231	.467	-	.426
10	2500.4	.476	1.119	2162.9	.461	.952	2252.6	.266	.579	-	.482
11	2674.7	.527		2297.6	.528		2401.2	.300		-	.537
LAMBDA MOYEN		.085			.095			.095		.100	

Tableau 2 : Coefficients de variation des estimations $\hat{N}_j$ et $\hat{F}_j$

Estimateur stratifié

Rapport des coûts = .02

33 individus âgés par strate

AGE	n1	CV.N	CV.F	LAMBDA
1	283.5	.064	.075	.122
2	270.8	.077	.098	.132
3	249.6	.115	.158	.136
4	243.4	.146	.248	.135
5	245.2	.169	.316	.133
6	248.5	.193	.301	.133
7	247.5	.240	.424	.130
8	253.1	.284	.498	.127
9	260.9	.342	.714	.120
10	275.8	.390	.859	.112
11	294.3	.438		
LAMBDA MOYEN = .1264				

Estimateur aléatoire

Rapport des coûts = .02

200 individus âgés

AGE	n1	CV.N	CV.F
1	1657.8	.064	.077
2	1615.2	.071	.091
3	1515.8	.108	.148
4	1474.9	.136	.234
5	1486.3	.159	.299
6	1507.6	.181	.282
7	1501.7	.225	.397
8	1537.1	.267	.466
9	1587.0	.322	.671
10	1679.3	.367	.824
11	1793.2	.411	
LAMBDA MOYEN = .1268			

Tableau 3 : Comparaison Stratifié - Aléatoire simple

ESTIMATION PONDEREE

Capture		Analyse de cohorte	
AGE	FREQ. %	EFFECTIF relatif	F
1	27.580	.18064E+01	.185
2	38.126	.12294E+01	.420
3	13.426	.66157E+00	.254
4	6.279	.42016E+00	.181
5	3.901	.28719E+00	.163
6	4.127	.19983E+00	.259
7	2.283	.12627E+00	.223
8	1.664	.82725E-01	.251
9	.869	.52670E-01	.201
10	.574	.35255E-01	.198
11	1.184	.23675E-01	.200

ESTIMATION SIMPLE

Capture		Analyse de cohorte	
AGE	FREQ. %	EFFECTIF relatif	F
1	21.618	.19232E+01	.133
2	37.243	.13790E+01	.355
3	16.044	.79203E+00	.253
4	7.576	.50328E+00	.182
5	4.748	.34350E+00	.166
6	4.946	.23827E+00	.261
7	2.730	.15033E+00	.224
8	1.995	.98379E+00	.252
9	1.037	.62581E-01	.202
10	.684	.41853E-01	.199
11	1.404	.28080E-01	.200

M = .200 E TERMINAL = .500

Tableau 4 : Résultat de l'analyse de cohorte à partir des deux estimations des captures par âge :

- Estimation pondérée $\Sigma T_i P_i / \Sigma T_i$

- Estimation simple $\Sigma P_i / Q$

	AGE (a)	1		2		4		6		8	
	ρ	.927		.470		- .948		- .930		- .851	
	Pa	.27580		.38128		.06279		.04127		.01664	
q	Estimateur	Rapport %	Simple %	Rapport %	Simple %	Rapport %	Simple %	Rapport %	Simple %	Rapport %	Simple %
1		-.224	-.216	-.021	-.023	.208	.207	.197	.198	.181	.193
2		-.090	-.216	-.008	-.023	.083	.207	.079	.198	.072	.193
3		-.045	-.216	-.004	-.023	.042	.207	.039	.198	.036	.193
4		-.022	-.216	-.002	-.023	.021	.207	.020	.198	.018	.193
5		-.009	-.216	-.001	-.023	.007	.207	.008	.198	.007	.193

Tableau 5 : Comparaison de l'estimateur rapport et de l'estimateur simple quant aux biais relatifs sur les fréquences estimées pour quelques âges en fonction du taux d'échantillonnage du premier niveau

AGE (a)	ρ'_a	$\frac{\Sigma T^2_K (Pa_K - Pa)^2}{\bar{T}^2 (Q-1)}$	$\frac{\Sigma (T_K Pa_K - \frac{\Sigma T_K Pa_K}{Q})^2}{\bar{T}^2 (Q-1)}$
1	.927	.2781E-1	.8430E-1
2	.470	.1362E-2	.3935E-1
3	- .810	.6505E-2	.1429E-2
4	- .948	.1200E-2	.5866E-4
5	- .728	.5152E-3	.4803E-4
6	- .930	.4903E-3	.4315E-4
7	- .920	.1330E-3	.4256E-5
8	- .851	.6490E-4	.3211E-5
9	- .819	.1833E-4	.1987E-5
10	- .850	.8095E-5	.7202E-6
11	- .869	.3084E-4	.1536E-5

Tableau 6 : Mesure de la variabilité entre unités primaires pour l'estimateur non biaisé et l'estimateur rapport

Estimateur Rapport

q	1			2			4			
AGE	n1	CV.N	CV.F	n1	CV.N	CV.F	n1	CV.N	CV.F	LAMBDA
1	1718.2	.154	.522	859.1	.109	.335	429.5	.078	.180	.122
2	1641.1	.148	.195	820.5	.111	.145	410.3	.086	.111	.132
3	1512.8	.219	.534	756.4	.165	.359	378.2	.129	.225	.136
4	1475.4	.232	.552	737.7	.185	.398	368.9	.156	.292	.135
5	1485.9	.256	.613	742.9	.208	.458	371.5	.179	.356	.133
6	1506.2	.282	.562	753.1	.232	.424	376.5	.203	.335	.133
7	1499.7	.328	.634	749.9	.278	.517	374.9	.249	.447	.130
8	1533.7	.382	.686	766.8	.326	.579	383.4	.294	.518	.127
9	1581.0	.458	.882	790.5	.392	.784	395.3	.354	.730	.120
10	1671.4	.530	1.025	835.7	.450	.927	417.9	.404	.874	.112
11	1783.5	.611		891.8	.513		445.9	.456		

LAMBDA MOYEN = .1264

Estimateur simple

1	1623.6	.154	.679	811.8	.110	.433	405.9	.079	.226	.126
2	1582.7	.144	.197	791.3	.108	.144	395.7	.084	.108	.134
3	1493.3	.199	.471	746.6	.150	.318	373.3	.118	.202	.137
4	1454.7	.214	.484	727.4	.170	.353	363.7	.143	.264	.136
5	1466.5	.239	.540	733.2	.193	.409	366.6	.165	.323	.134
6	1487.7	.266	.505	743.8	.217	.384	371.9	.187	.306	.135
7	1481.8	.314	.591	740.9	.261	.479	370.5	.230	.412	.132
8	1515.0	.367	.665	757.5	.307	.549	378.7	.273	.480	.128
9	1562.3	.439	.844	781.1	.369	.735	390.6	.328	.673	.121
10	1650.9	.508	.967	825.5	.424	.863	412.7	.375	.805	.114
11	1761.1	.588		880.5	.485		440.3	.424		

LAMBDA MOYEN = .1288

Estimateur non biaisé

1	1718.2	.224	.894	859.1	.150	.569	429.5	.093	.292	.122
2	1641.1	.185	.404	820.5	.131	.266	410.3	.093	.158	.132
3	1512.8	.133	.279	756.4	.123	.215	378.2	.117	.174	.136
4	1475.4	.152	.271	737.7	.148	.257	368.9	.146	.249	.135
5	1485.9	.176	.352	742.9	.172	.330	371.5	.169	.318	.133
6	1506.2	.199	.329	753.1	.195	.311	376.5	.193	.302	.133
7	1499.7	.244	.431	749.9	.241	.426	374.6	.239	.423	.130
8	1533.7	.289	.507	766.8	.285	.500	383.4	.283	.497	.127
9	1581.0	.348	.727	790.5	.344	.717	395.3	.341	.712	.120
10	1671.4	.396	.868	835.7	.391	.860	417.9	.389	.856	.112
11	1783.5	.446		891.8	.440		445.9	.437		

LAMBDA MOYEN = .1264

Tableau 7 : Comparaison de l'estimateur rapport, de l'estimateur non biaisé et de l'estimateur simple sur la base des variances (précision)

Rapports des coûts (C_1/C_2) = . 02

Total d'individus âgés	=	200
M	=	.20
ET	=	.50
q	=	nombre de niveaux échantillonnés

q	1				2				3				
	AGE	nl	CV.N	CV.F	AGE	nl	CV.N	CV.F	AGE	nl	CV.N	CV.F	LAMBDA
ESTIMATEUR RAPPORT	1	494.6	.127	.433	1	247.3	.090	.280	1	123.6	.064	.154	.081
	2	421.3	.121	.170	2	210.7	.090	.128	2	105.3	.070	.100	.095
	3	280.0	.174	.601	3	140.0	.129	.413	3	70.0	.100	.273	.143
	4	193.1	.111	.450	4	96.5	.102	.382	4	48.3	.097	.342	.207
	5	148.4	.093	.216	5	74.2	.092	.215	5	37.1	.091	.214	.269
	6	154.2	.103	.191	6	77.1	.101	.187	6	38.5	.101	.185	.259
	7	137.2	.122	.250	7	68.6	.121	.245	7	34.3	.120	.243	.292
	8	132.2	.141	.253	8	66.1	.139	.251	8	33.1	.138	.250	.302
	9	134.0	.168	.369	9	67.0	.166	.364	9	33.5	.165	.361	.299
	10	137.7	.189	.427	10	68.9	.187	.421	10	34.4	.186	.419	.290
	11	136.5	.212		11	68.2	.209		11	34.1	.208		.293
	LAMBDA MOYEN =		.1857										
ESTIMATEUR NON BIAISE	1	494.6	.221	.894	1	247.3	.145	.569	1	123.6	.085	.291	.081
	2	421.3	.180	.402	2	210.7	.123	.263	2	105.3	.082	.152	.095
	3	280.0	.110	.309	3	140.0	.097	.252	3	70.0	.090	.218	.143
	4	193.1	.104	.343	4	96.5	.099	.334	4	48.3	.096	.329	.207
	5	148.4	.103	.265	5	74.2	.096	.235	5	37.1	.092	.219	.269
	6	154.2	.111	.227	6	77.1	.105	.203	6	38.5	.102	.189	.259
	7	137.2	.128	.257	7	68.6	.123	.248	7	34.3	.121	.244	.292
	8	132.2	.148	.271	8	66.1	.142	.259	8	33.1	.139	.252	.302
	9	134.0	.177	.390	9	67.0	.170	.372	9	33.5	.166	.363	.299
	10	137.7	.199	.442	10	68.9	.191	.428	10	34.4	.187	.420	.290
	11	136.5	.223		11	68.2	.214		11	34.1	.209		.293
	LAMBDA MOYEN =		.1857										

Tableau 8 : Comparaison de l'estimateur rapport et de l'estimateur non biaisé quant à leur variance respective. Stratification par catégorie commerciale

Rapport des coûts (C_1/C_2) = .02 Total d'individus âgés 200 q = nombre de niveaux échantillonnés
$\bar{M}$ = .20 ET = .50 Total âgé = 200

	ESTIMATEUR RAPPORT					ESTIMATEUR NON BIAISE				
	F	VARN(n1)	VARN(n2)	VN3	RC2	F	VARN(n1)	VARN(n2)	VN3	RC2
Sans Stratification	.185	.19984E+02	.13539E+02	.77271E-01	.67748E+00	.185	.19984E+02	.13539E+02	.18012E+00	.67748E+00
	.420	.12221E+02	.90754E+01	.29032E-01	.74261E+00	.420	.12221E+02	.90754E+01	.51645E-01	.74261E+00
	.254	.68887E+01	.60203E+01	.18345E-01	.87395E+00	.254	.68887E+01	.60203E+01	.24045E-02	.87395E+00
	.181	.42253E+01	.38821E+01	.69718E-02	.91877E+00	.181	.42253E+01	.38821E+01	.44181E-03	.91877E+00
	.163	.27052E+01	.24505E+01	.36912E-02	.90584E+00	.163	.27052E+01	.24505E+01	.24813E-03	.90584E+00
	.259	.17430E+01	.15367E+01	.20525E-02	.88164E+00	.259	.17430E+01	.15367E+01	.12700E-03	.88164E+00
	.223	.10632E+01	.94543E+00	.97436E-03	.88921E+00	.223	.10632E+01	.94543E+00	.49803E-04	.88921E+00
	.251	.66938E+00	.56918E+00	.54420E-03	.85031E+00	.251	.66938E+00	.56918E+00	.29899E-04	.85031E+00
	.201	.41677E+00	.33347E+00	.31165E-03	.80013E+00	.201	.41677E+00	.33347E+00	.17413E-04	.80013E+00
	.198	.26846E+00	.19220E+00	.19390E-03	.71592E+00	.198	.26846E+00	.19220E+00	.10045E-04	.71592E+00
	.200	.17249E+00	10845E+00	.12334E-03	.62873E+00	.200	.17249E+00	.10845E+00	.61440E-05	.62873E+00
Avec Stratification	.185	.51981E+01	.17000E+01	.52309E-01	.32705E+00	.185	.51981E+01	.17000E+01	.17997E+00	.32705E+00
	.420	.25288E+01	.11396E+01	.19579E-01	.45063E+00	.420	.25288E+01	.11396E+01	.51546E-01	.45063E+00
	.254	.69671E+00	.71067E+00	.11796E-01	.10200E+01	.254	.69671E+00	.71067E+00	.23379E-02	.10200E+01
	.181	.16245E+00	.34871E+00	.71355E-03	.21466E+01	.181	.16245E+00	.34871E+00	.39718E-03	.21466E+01
	.163	.41887E-01	.15206E+00	.37593E-04	.36303E+01	.163	.41887E-01	.15206E+00	.22680E-03	.36303E+01
	.259	.26612E-01	.89587E-01	.24220E-04	.33664E+01	.259	.26612E-01	.89587E-01	.11271E-03	.33664E+01
	.223	.12069E-01	.51286E-01	.12351E-04	.42492E+01	.223	.12069E-01	.51286E-01	.41148E-04	.42492E+01
	.251	.64001E-02	.29276E-01	.62509E-05	.45743E+01	.251	.64001E-02	.29276E-01	.24285E-04	.45743E+01
	.201	.37922E-02	.16908E-01	.40989E-05	.44586E+01	.201	.37922E-02	.16908E-01	.13650E-04	.44586E+01
	.198	.22751E-02	.95937E-02	.22688E-05	.42169E+01	.198	.22751E-02	.95937E-02	.75223E-05	.42169E+01
	.200	.12595E-02	.54104E-02	.12863E-05	.42958E+01	.200	.12595E-02	.54104E-02	.44550E-05	.42958E+01

M = .2 ET = .5

Tableau 9 : Résultats comparés avant et après stratification par catégories commerciales