

00000818

ETUDE DE LA QUALITÉ DES ESTIMATIONS A VUE
DU POIDS DES PRISES DÉBARQUÉES PAR LES
PECHEURS ARTISANS SÉNÉGALAIS

PAR

CATHERINE DESTANQUE

RAPPORT INTERNE

N° 64

ETUDE DE LA QUALITE DES ESTIMATIONS A VUE
DU POIDS DES PRISES DEBARQUEES PAR
LES PECHEURS ARTISANS SENEGALAIS

Catherine DESTANQUE

Le tonnage annuel des prises débarquées par la pêche artisanale sénégalaise est estimé sur la base d'un échantillonnage des captures. Le nombre des points de débarquement et l'importance de la flottille en certains points ont rendu nécessaire la création d'une procédure d'échantillonnage à plusieurs niveaux. Seuls certains sites de débarquement sont enquêtés (1er niveau) et pour un même site, l'enquêteur choisit au hasard un certain nombre de pirogues (2e niveau). Le poids débarqué par espèce doit alors être estimé pour chaque pirogue sélectionnée. Deux techniques différentes sont utilisées.

. Lorsque les poissons sont en petit nombre, l'enquêteur compte les individus présents et en mesure 5, qu'il tire au hasard. Grâce à une clé taille poids, on peut connaître le poids de chacun des cinq poissons. On peut alors parler d'un troisième niveau d'échantillonnage puisque tous les poissons de la pirogue ne sont pas mesurés.

C'est un exemple classique d'échantillonnage hiérarchique par grappe où le point de débarquement constitue une grappe primaire et la pirogue une grappe secondaire, cette dernière étant elle-même sous échantillonnée. On peut exprimer assez facilement, par espèce, un estimateur du poids débarqué, ainsi qu'un estimateur de la variance de l'estimateur du poids (Cochran, 1963).

. Lorsque le dénombrement des individus devient impossible, l'enquêteur procède alors à une estimation à vue du poids débarqué. Il n'y a plus de 3e niveau d'échantillonnage. Le poids estimé par l'enquêteur n'en reste pas moins une variable aléatoire, estimateur du poids réel du tas de poisson. En effet, on imagine très bien que deux enquêteurs placés devant le même tas proposent deux estimations différentes du même poids réel ; il en est de même pour un enquêteur

* Allocataire de recherche ORSTOM au Centre de Recherches Océanographiques de Dakar-Thiaroye, ISRA, B.P. 2241' Dakar (Sénégal).

regardant des tas de même poids, mais de formes variées ou le même tas à des moments différents. Cet estimateur suit une loi de moyenne m et de variance σ^2 . Cette variance n'est pas une variance d'échantillonnage, mais une variance d'erreur d'appréciation. Elle n'est pas calculable et ne peut être qu'estimée empiriquement.

Dans cet te optique , une expérience permettant de comparer, pour des tas de poissons de poids variés, le poids estimé par des enquêteurs au poids mesuré par une balance, a été mise en place au CRODT, fin avril 1983 .

Ce travail est une première étude de la technique d'estimation à vue très souvent utilisée, surtout pour certaines espèces de petite taille, comme la sardinelle et le pageot, pour estimer le poids débarqué par pirogue.

Il a deux objectifs :

1. Apprécier la qualité de l'information fournie par les enquêteurs en estimant, sur la base des données recueillies lors de l'expérience effectuée au CRODT,, l'exactitude (absence de biais) et la précision (variance) de l'estimateur à vue.

D'un point de vue pratique, ces résultats permettront d'envisager

- une amélioration possible des estimations

- . en sensibilisant les enquêteurs à leurs erreurs ;

- . en effectuant a posteriori des corrections sur les données recueillies ;

- la construction d'un intervalle de confiance autour de la valeur réelle du poids à partir des estimations obtenues.

2. Apprécier la validité de l'échelle de notation jusqu'ici utilisée par les enquêteurs. Théoriquement, celle-ci va de 0 à plusieurs tonnes, avec un pas de 1 kg. Dans la pratique, toutes les valeurs ne sont pas utilisées, les enquêteurs ne citant que certaines valeurs marquantes du système décimal (multiples de 5, 10, 50, 100) avec un pas qui augmente avec le poids. Ces échelles de notations , créées spontanément par chaque enquêteur sont-elles bien adaptées ? Dans quelle mesure est-il possible de les améliorer pour une meilleure qualité de l'information ?

I ETUDE DE LA QUALITE DE L' INFORMATION

1. Modélisation de l'erreur d'évaluation

1.1 Modèle de base

On considère que tout individu commet une erreur d'évaluation lorsqu'il estime à vue le poids P d'un tas de poisson. On s'intéresse à la population des enquêteurs potentiels, dont un échantillon est l'ensemble des enquêteurs du centre océanographique, et on étudie la distribution des valeurs que la variable aléatoire "erreur d'évaluation" prend sur cette population.

Soit z_{ijk} l'erreur d'évaluation commise par un enquêteur j en estimant pour la $k^{\text{ième}}$ fois le poids P_i d'un tas.

$$z_{ijk} = y_{ijk} - P_i = A_{ij} + \epsilon_{ijk}$$

y_{ijk} : évaluation du poids P_i par l'enquêteur j pour la $k^{\text{ième}}$ fois.

P_i : poids mesuré à la balance d'un tas i .

A_{ij} : biais commis par l'enquêteur j sur un tas i .

ϵ_{ijk} : erreur aléatoire d'évaluation.

$$E_k(\epsilon_{ijk}) = 0 \quad E_k(y_{ijk} - P_i) = A_{ij}$$

On suppose A_{ij} et ϵ_{ijk} indépendants.

On peut écrire :

$$E_k(y_{ijk} - P_i)^2 = \sigma_{e_{ij}}^2 + A_{ij}^2 \quad (1)$$

$$\text{avec } \sigma_{e_{ij}}^2 = \text{var}_k(\epsilon_{ijk}) = \text{var}_k(y_{ijk})$$

On retrouve ici l'égalité classique. L'espérance du carré des écarts de la valeur observée à la valeur réelle est égale à la variance des valeurs observées, plus le biais au carré.

On considère maintenant la population des enquêteurs, dont on a un échantillon, estimant le poids d'un tas i .

.../...

$$z_{ijk} = y_{ijk} - P_i = A_{ij} + \epsilon_{ijk}$$

$$E_j(A_{ij}) = A_i \quad \text{var}_j(A_{ij}) = \sigma_{A_i}^2 \quad E_{jk}(\epsilon_{ijk}) = 0$$

Le biais A_{ij} est une variable aléatoire de moyenne A_i et de variance $\sigma_{A_i}^2$.

On fait 1 'hypothèse suivante :

$$\forall j \quad \sigma_{e_{ij}}^2 = \sigma_{e_i}^2 \quad \text{donc} \quad \text{var}_{jk}(\epsilon_{ijk}) = \sigma_{e_i}^2$$

Les estimations de tous les enquêteurs ont la même précision pour un tas i , la variance de l'erreur aléatoire d'évaluation ne dépendant que du poids P_i .

On peut écrire :

$$\begin{aligned} E_{jk}(y_{ijk} - P_i)^2 &= \sigma_{A_i}^2 + \sigma_{e_i}^2 + A_i^2 \\ &= \text{var}_{jk}(y_{ijk}) + A_i^2 \end{aligned} \quad (2)$$

On retrouve ici une expression comparable à (1) avec A_i , biais global de la population d'enquêteurs et $\text{var}_{jk}(y_{ijk})$, la variance des valeurs observées.

Dans le premier cas, on se plaçait au niveau de chaque enquêteur, et dans le deuxième cas, on se place au niveau de la population des enquêteurs.

1.2 Paramètres à estimer

Pour chaque estimation d'un poids P_i , on voudrait être en mesure de corriger le biais et de construire un intervalle de confiance autour de la valeur réelle.

On peut donc, pour chaque estimation i ,

- soit se placer au niveau de la population et ne pas vouloir, ou ne pas pouvoir prendre en considération un effet enquêteur. Il est alors nécessaire de connaître :

- le biais global de la population A_i ;
- la variance totale $\text{var}_{jk}(y_{ijk}) = \sigma_{A_i}^2 + \sigma_{e_i}^2$.

- soit personnaliser l'enquêteur et il faut avoir montré l'existence d'un effet enquêteur. Il est alors nécessaire de connaître :

- le biais propre à l'enquêteur A_{ij} ;

.../...

- la variance de l'erreur aléatoire d'évaluation $\sigma_{e_i}^2$.

S'il existe un effet enquêteur, il est plus intéressant de se placer dans le deuxième cas. L'estimateur corrigé de biais sera plus juste et l'intervalle de confiance plus étroit.

Cependant, même si cet effet enquêteur a été mis en évidence, il peut ne pas être possible dans la pratique de le prendre en compte. On pourrait toujours alors se placer dans le premier cas.

On cherchera donc, pour tout poids P_i

- à estimer la variance de l'erreur aléatoire d'évaluation $\sigma_{e_i}^2$ et la variance totale de l'erreur d'évaluation égale à $A_i \sigma_i^2 + \sigma_{e_i}^2$;

- à déterminer s'il existe un effet enquêteur et à estimer A_i , A_{ij} .

1.3 Précision du modèle

On a considéré, comme c'est souvent le cas que, à la fois la justesse et la précision des estimateurs du poids étaient fonction du poids. La non homogénéité des variances $\sigma_{A_i}^2$ et $\sigma_{e_i}^2$ ne permet pas d'assimiler le modèle jusqu'ici proposé à un modèle classique d'analyse de variance.

On fait l'hypothèse que le rapport $\text{var}(y_{ijk})/P_i^2$ est constant et on effectue la transformation suivante visant à réduire les variables :

$$t_{ijk} = \frac{y_{ijk} - P_i}{P_i}$$

On obtient le modèle :

$$t_{ijk} = B_{ij} + \varepsilon'_{ijk}$$

$$\text{avec } \text{var}(\varepsilon'_{ijk}) = \sigma_e^2 \text{ et } \sigma_{e_i}^2 = P_i^2 \sigma_e^2$$

$\sigma_{e_i}^2$ variance de l'erreur aléatoire d'évaluation

$$\text{var}(B_{ij}) = \sigma_b^2 \text{ et } \sigma_{A_i}^2 = P_i^2 \sigma_b^2$$

$\sigma_{A_i}^2$ variance du biais A_{ij}

$$B_{ij} = \frac{A_{ij}}{P_i}$$

Cette hypothèse est vérifiée ultérieurement dans l'analyse des données.

• On peut décomposer B_{ij}

$$B_{ij} = \mu + B_i + C_j + D_{ij}$$

μ : moyenne générale

B_i = effet aléatoire lié au poids P_i

C_j : effet aléatoire lié à l'enquêteur j

D_{ij} : interaction poids . enquêteur

On a donc le modèle aléatoire suivant :

$$t_{ijk} = \mu + B_i + C_j + D_{ij} + \varepsilon_{ijk} \quad (E_{ijk} = \varepsilon'_{ijk} \text{ précédemment cité})$$

$$E(B_i) = 0$$

$$E(C_j) = E(D_{ij}) = E(\varepsilon_{ijk}) = 0$$

$$\text{var}(C_j) = \sigma_C^2 \quad \text{var}(D_{ij}) = \sigma_D^2 \quad \text{var}(\varepsilon_{ijk}) = \sigma_e^2 \quad \text{var}(B_i) = \sigma_B^2$$

Lors de la connection des estimations, il n'est pas possible d'utiliser tous les résultats de l'analyse d'un modèle aussi détaillé. Il paraît cependant intéressant, dans le cadre de l'expérience effectuée au CRODT, de se faire une idée des composantes de la variance de l'erreur d'évaluation.

2. Modalités pratiques de l'expérience

12 enquêteurs professionnels plus un enquêteur amateur, chercheur du centre ont bien voulu participer à cette expérience.

L'espèce de poisson choisie était le pageot rose (*Pagellus Coupei*) car c'est une des espèces pour lesquelles les enquêteurs (doivent effectuer couramment des estimations de poids à vue.

2.1 . Distribution des poids des tas étudiés

Pour des raisons de facilité de manutention, le poids des tas n'a pas dépassé 265 kg. Les conditions pratiques ont rendu difficile l'exécution d'un schéma de travail prévu à l'avance. Les tas, constitués par séries de 5, étaient faits

au jugé, en utilisant 13 disposition préexistante du poisson sur le sol, de manière à ne pas perdre trop de temps à charier des tas de poisson; pesant jusqu'à 265 kg.

La distribution obtenue, de 0 à 265 kg, présente une dissymétrie marquée, il est en effet plus facile de faire des petits tas que des gros. Elle a l'allure des distributions classiques d'abondance. Elle est donc relativement satisfaisante. Il n'y a pas de répétitions pour toutes les modalités, surtout pour les poids élevés. Cependant, la précision de la pesée, diminuant avec le poids, permet de considérer comme ayant le même poids des tas dont la différence des poids est inférieure à 3 % de leur poids.

Il y a quelques données manquantes.

2.2 Indépendance des mesures

Il était important que les mesures soient indépendantes

- pour un même enquêteur qui aurait pu être amené, en regardant plusieurs tas proches physiquement, à estimer un tas, puis à comparer les autres et à les estimer en fonction du premier. On a donc veillé à ce que, pour une même série de 5 tas, les poids soient suffisamment différents pour que les enquêteurs les estiment de façon indépendantes ;

- d'un enquêteur à l'autre. La comparaison de leurs résultats au fur et à mesure que progressait l'expérience aurait pu les influencer dans leurs estimations ultérieures. Pour cela, un carnet avait été distribué à chaque enquêteur, chacun devant passer devant les 5 tas, noter son nom, le numéro du tas et le poids estimé, sur une feuille, puis rendre ladite feuille, de manière à éviter, sinon les corrections a posteriori, du moins, l'influence des estimations des autres enquêteurs.

3. Résultats

3.1 Vérification des hypothèses

Tous les enquêteurs ont la même précision ($\forall j \sigma_{ij}^2 = \sigma_{e_1}^2$).

Sur la base des données de l'expérience, on estime, pour chaque enquêteur, $\sigma_{e_{ij}}^2$. Le test de Bartlett permet de déterminer si les variances

.../...

sont égales. Il résulte que certains enquêteurs ont des variances manifestement différentes. On travaillera donc sur les données de 10 enquêteurs seulement, dont l'homogénéité des variances est dans la limite de l'acceptable.

. Le rapport $\text{var}(y_{ijk})/P_i^2$ est constant. Pour chaque tas, 1 'écart-type des observations a été calculé. L'ajustement linéaire entre l'écart-type et le poids réel est satisfaisant ($R^2 = 0,85$). Le coefficient de la régression est 0,44.

On peut donc écrire : $\frac{\text{var}(y_{ijk})}{P_i^2} = 1,93 \quad 0,4436$

L'hypothèse est donc acceptable.

3.2 Analyse des résultats

On considère le modèle

$$t_{ijk} = \mu + B_i + C_j + D_{ij} + \epsilon_{ijk}$$

i: $1 \rightarrow I = 39$ le poids P_i varie de 1 à 265kgs

j : $1 \rightarrow J = 10$ 10 enquêteurs

k: $1 \rightarrow K(i)$ le nombre de répétitions dépend du poids considéré

$$\text{var}(\epsilon_{ijk}) = \sigma_e^2 \quad \text{var}(D_{ij}) = \sigma_D^2 \quad \text{var}(C_j) = \sigma_C^2 \quad \text{var}(B_i) = \sigma_B^2$$

Le plan d'expérience étant déséquilibré, il est difficile de construire des estimateurs, et, de plus, on ne connaît pas leur qualité. Il est donc préférable de mettre en place un plan d'expérience équilibré, sur la base des données existantes, à partir duquel on construira facilement des estimateurs optimaux pour ce plan.

En supprimant les données manquantes, on obtient un modèle équilibré possible ($i : 1 \rightarrow 9, j : 1 \rightarrow 9, k : 1 \rightarrow 2$).

La taille limite de la matrice d'incidence imposée par le logiciel utilisé ne permet pas d'étudier les interactions sur ce modèle.

En réduisant le nombre d'enquêteurs à 5, et le nombre de poids différents à 8, il devient possible de prendre en compte 1 'interaction.

On considère donc deux modèles :

$$- t_{ijk} = \mu + B_i + C_j + \varepsilon'_{ijk} \quad (1)$$

$$i : 1 \rightarrow 9 \quad j : 1 \rightarrow 9 \quad k : 1 \rightarrow 2$$

La variance résiduelle d'évaluation englobe celle de l'interaction.

$$\text{var}(\varepsilon'_{ijk}) = \sigma_{\varepsilon'}^2 = \sigma_e^2 + \sigma_D^2$$

$$- t_{ijk} = \mu + B_i + C_j + D_{ij} + \varepsilon_{ijk}$$

$$i : 1 \rightarrow 8 \quad j : 1 \rightarrow 5 \quad k : 1 \rightarrow 2$$

Il est possible de faire la part des choses entre l'interaction et l'erreur résiduelle. On peut donc tester l'interaction et estimer σ_e^2 .

Pour les 2 modèles, les tableaux d'analyse de variance sont présentés figure 1.

• Test des hypothèses d'absence d'effets des facteurs

On suppose que les erreurs sont distribuées normalement (ce qui n'est pas rigoureusement vrai), et on utilise le test de Fisher.

- L'effet lié à l'enquêteur est très marqué. Sous l'hypothèse d'absence d'effet du facteur enquêteur, la probabilité d'apparition du résultat obtenu est inférieure à 1 %.

- L'interaction enquêteur.poids est marquée. On constate que la prise en considération de l'interaction dans le modèle (2) fait diminuer de moitié la variance résiduelle par rapport au modèle (1). Les variances des effets poids et interaction enquêteur.poids sont du même ordre. Si l'on regarde les données initiales, on constate que les estimations relatives à certains poids sont davantage biaisées. On en peut cependant entrevoir aucune relation simple entre biais et poids. Ceci peut être expliqué par l'utilisation par les enquêteurs d'échelles de notations discrètes dont les modalités sont les valeurs marquantes du système décimal. L'attraction exercée par ces valeurs peut les amener à sous-estimer ou sur-estimer localement. Les échelles de notations varient avec les enquêteurs ; les biais relatifs varient donc avec le poids et avec l'enquêteur.

Ce sont des effets auxquels on pourra difficilement remédier dans la

.../...

pratique en effectuant des corrections d'estimateur et l'on devra souvent les confondre avec l'erreur aléatoire d'évaluation.

Le modèle le plus simple et utilisable dans la pratique est donc :

$$t_{ijk} = \frac{y_{ijk} - P_i}{P_i} = \mu + c_j + \epsilon''_{ijk}$$

$$\text{var}(\epsilon''_{ijk}) = \sigma_{\text{enq}}^2 = \sigma_B^2 + \sigma_D^2 + \sigma_e^2$$

• Estimation des paramètres

Il est donc intéressant d'estimer :

$$- \sigma_T^2 = \text{var}(t_{ijk}), \sigma_{\text{enq}}^2$$

$$- \mu, \forall j \ C_j$$

L'utilisation des statistiques adéquates donne les résultats suivants :

$$\hat{\sigma}_T^2 = 0,15 \quad \hat{\sigma}_T = 0,38$$

$$\hat{\sigma}_{\text{enq}}^2 = 0,07 \quad \sigma_{\text{enq}} = 0,26$$

$$\hat{\mu} = -0,026$$

Les estimations des C_j apparaissent fig. 1.

• $\hat{\mu}$ est très petit, on peut donc considérer que le biais global est nul.

y_{ij} est cependant un estimateur biaisé puisque l'enquêteur j effectue un biais C_j .

$$\text{On a } \text{var}(y_{ij}) = P_i^2 \sigma_T^2$$

$$\text{On estimera } \text{var}(y_{ij}) \text{ par } \hat{\sigma}_T^2 = (y_{ij})^2 \hat{\sigma}_T^2$$

On peut donc construire un intervalle de confiance autour de P_i tel que

$$y_{ij} - t_\alpha y_{ij} \hat{\sigma}_T < P_i < y_{ij} + t_\alpha y_{ij} \hat{\sigma}_T \text{ avec une probabilité } 1-\alpha \text{ et } t_\alpha$$

séparatrice de Student pour un risque α . (Ceci est une approximation car $\hat{\sigma}_T^2$ n'est pas distribuée selon une loi de Student. Il serait nécessaire de tabuler sa distribution).

• On pourrait construire un meilleur estimateur si l'on personnalise l'enquêteur.

.../...

ANALYSE DE VARIANCE

Modèle 1 : $t_{ijk} = \mu + B_i + C_j + \epsilon_{ijk}$

9 enquêteurs ($j : 1 \dots 9$), 9 modalités de poids ($i : 1 \dots 9$),

2 répétitions ($k : 1, 2$)

Source JC variation	SCE	DLIB	CME	F	F _{5%}
Facteur poids	0,87	8	0,11	1,71	1,94
Facteur enquêteur	13,36	8	1,67	26,02	1,94
Erreur	9,30	14,5	0,06		σ_e^2
Total	23,54	161	0,15		

Modèle 2 : $t_{ijk} = \mu + B_i + C_j + D_{ij} + \epsilon_{ijk}$

5 enquêteurs ($j : 1 \dots 5$), 8 modalités de poids ($i : 1 \dots 8$)

2 répétitions ($k : 1, 2$)

Source de variation	SCE	DLIB	CME	F	F _{5%}
Facteur poids	0,99	7	0,14	1,17	2,36
Facteur enquêteur	3,93	4	0,98	8,17	2,71
Interaction	3,29	28	0,12	4,12	1,74
Erreur	1,14	40	0,03		
Total	9,37	79	0,12		

Estimations de C_j pour les 9 enquêteurs

Enquêteur 1	- 0,16
Enquêteur 2	0,27
Enquêteur 3	0,26
Enquêteur 4	- 0,07
Enquêteur 5	0,50
Enquêteur 6	- 0,42
Enquêteur 7	0,01
Enquêteur 8	- 0,01
Enquêteur 9	- 0,37

Figure

Figure 2 : Distribution empirique des estimations
de poids à vue effectuées à Kayar en 1981

ectif

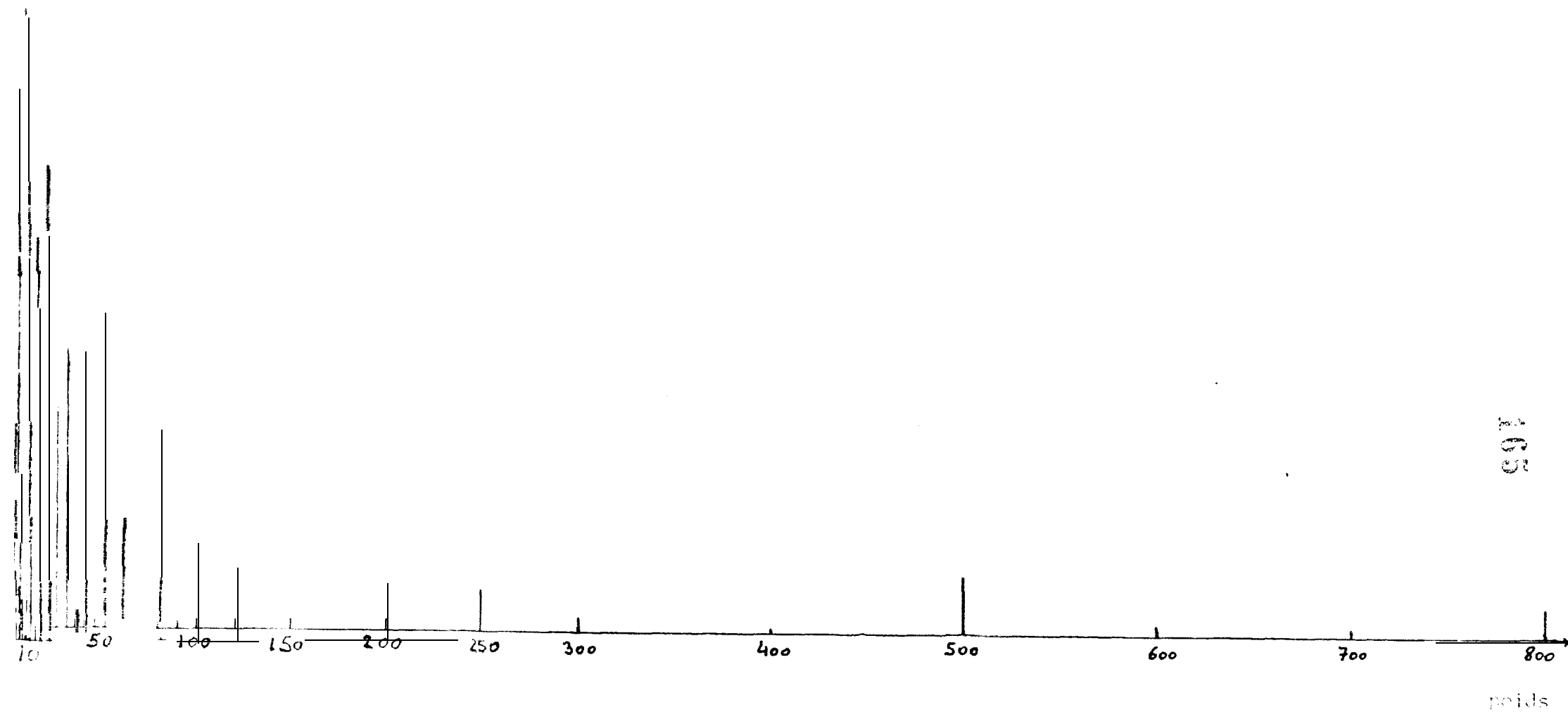
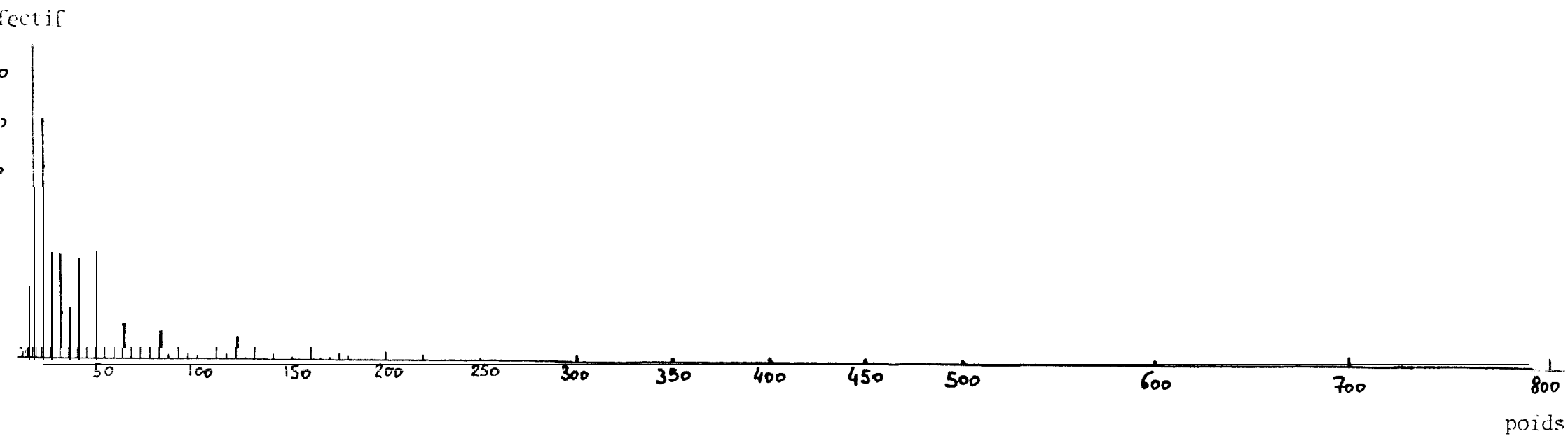
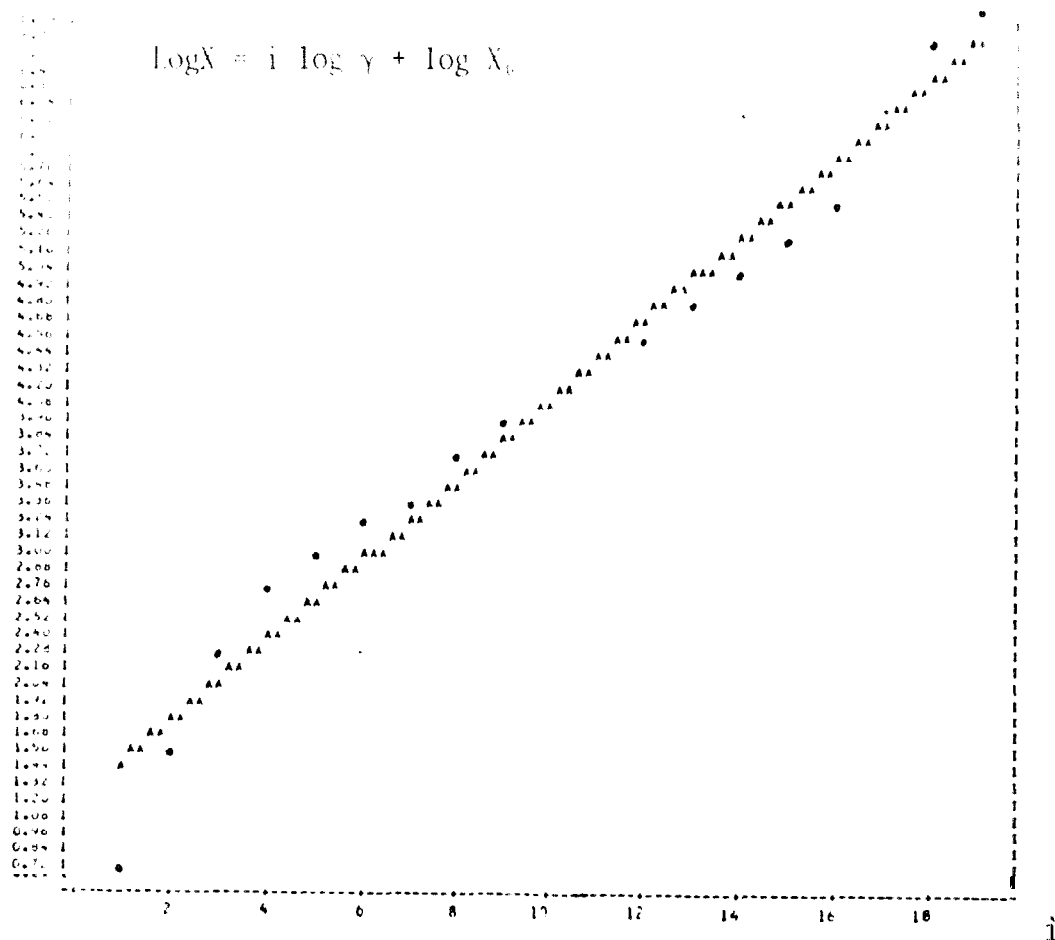


Figure 2 (suite) : Distribution empirique des estimations
de poids à vue effectuées à St-Louis en 1981



$$\log X = i \log \gamma + \log X_0$$



X

$$X = X_0 \gamma^i$$

Les valeurs ajustées de cette courbe sont les exponentielles des valeurs ajustées obtenues par la régression

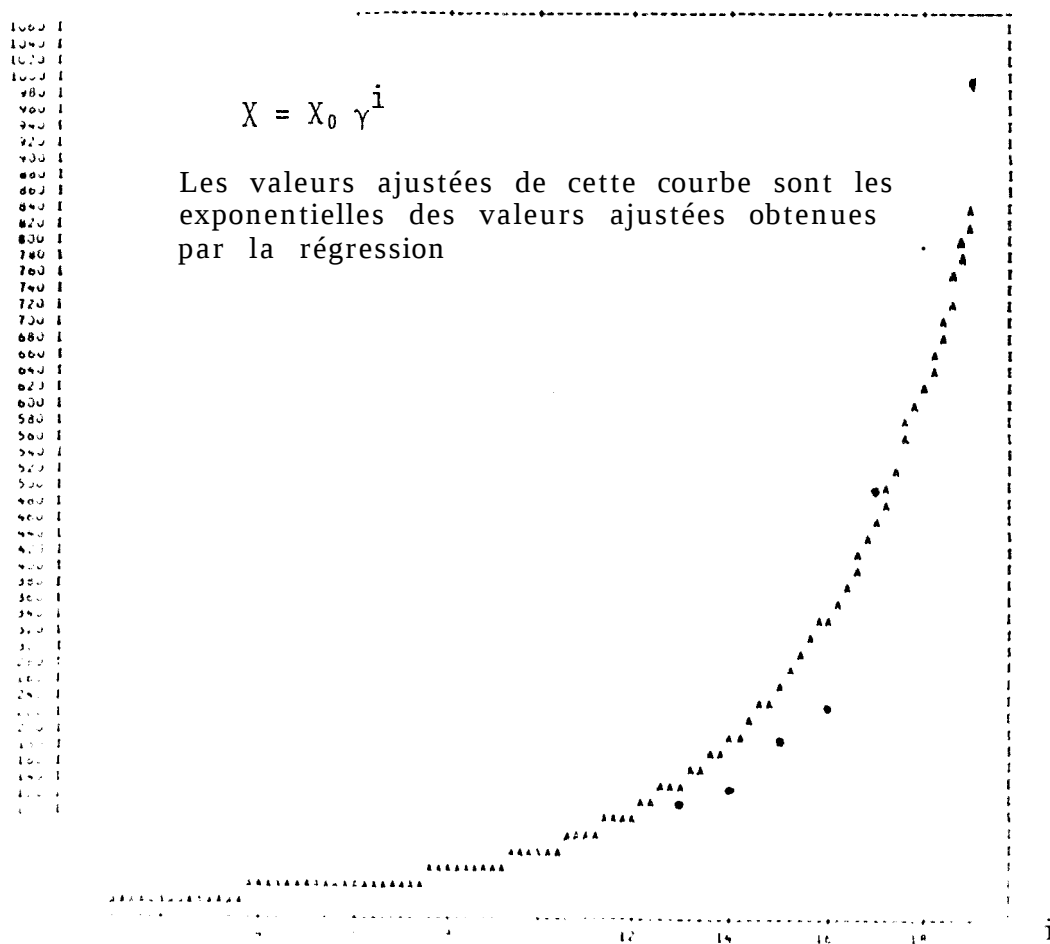
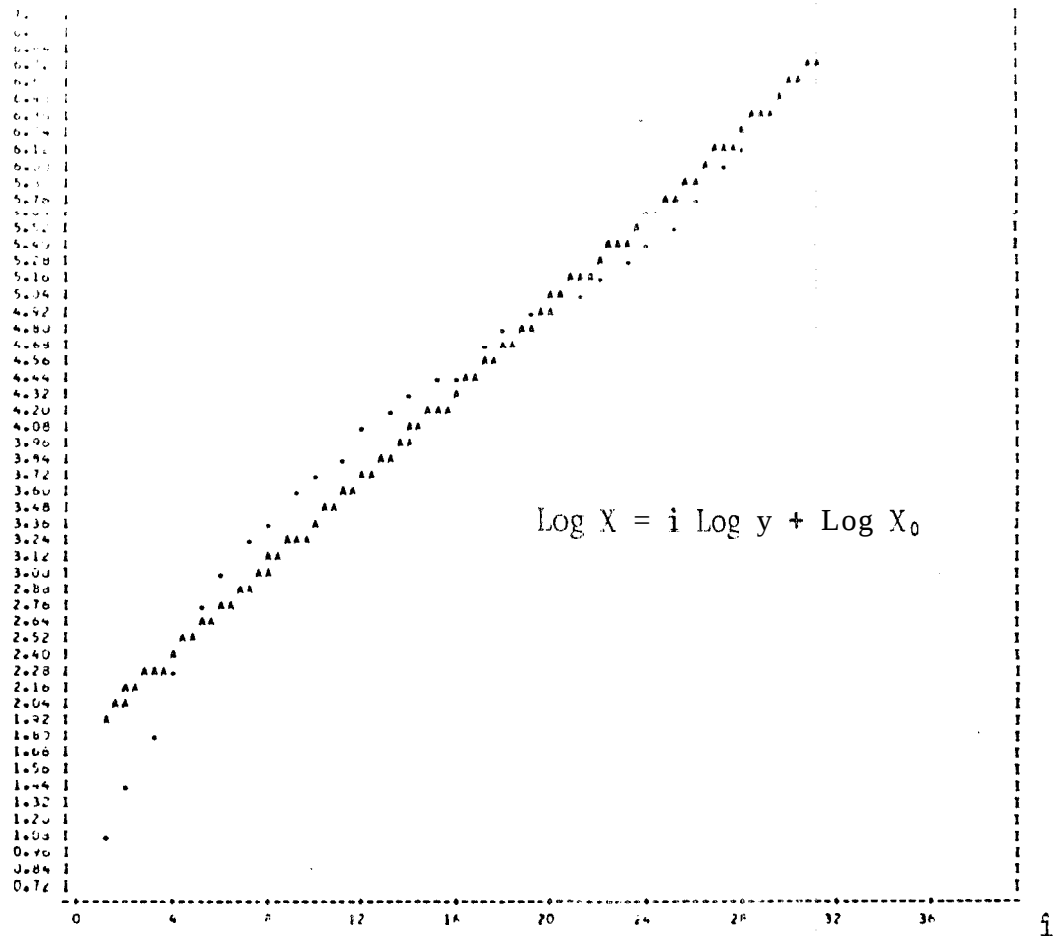


Figure n° 3 : modélisation de l'échelle de notation (Kayar)

Log X



X

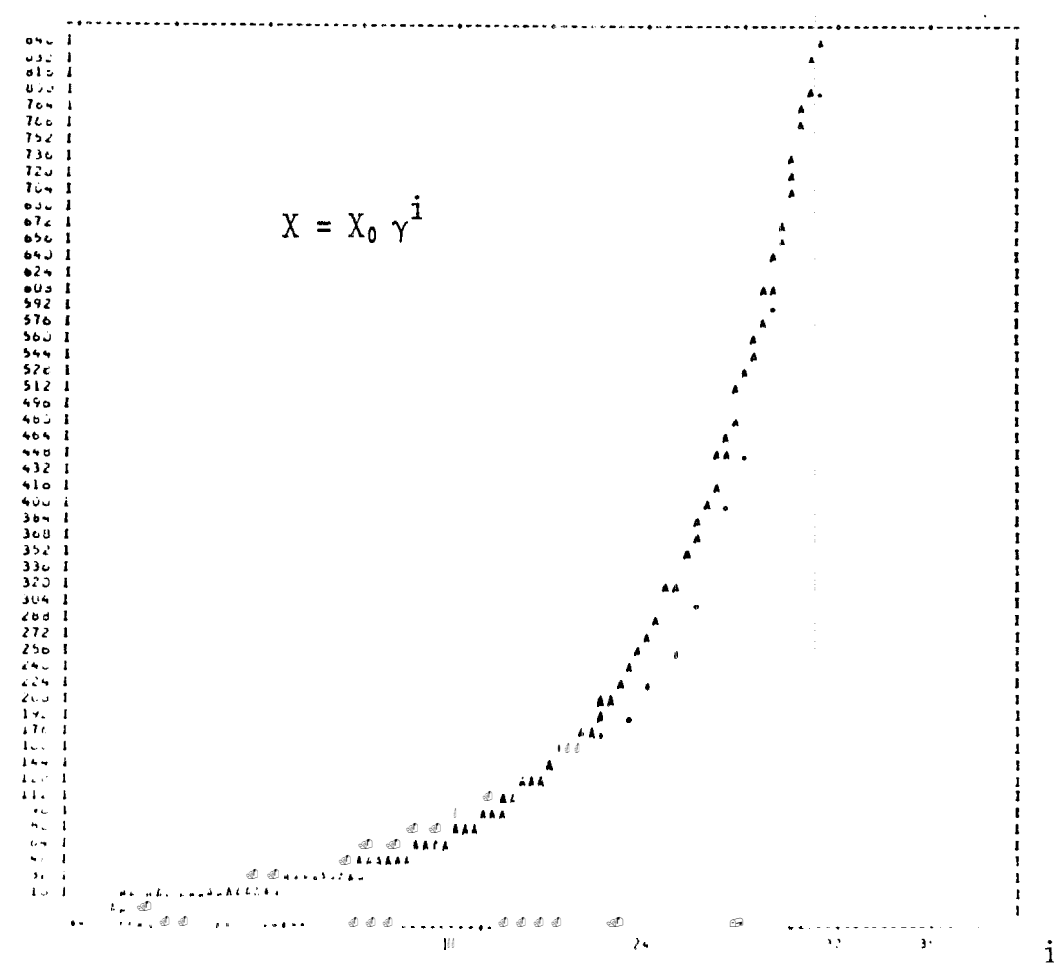


Figure n° 3 (suite) : modélisation de l'échelle de notation
(St Louis)

On le construit à partir de l'équation :

$$\frac{y_{ij} - P_i}{P_i} = C'_j + \varepsilon_{ij} \quad C'_j = \mu + C_j$$

$$\frac{y_{ij} - \hat{P}_i}{\hat{P}_i} = \hat{C}'_j$$

$$\hat{P}_i = \frac{y_{ij}}{1 + \hat{C}'_j}$$

Les propriétés de $\hat{P}_i$ dépendent de la qualité de l'estimation de C'_j . Si $\hat{C}'_j$ tend vers C'_j , l'estimateur $\hat{P}_i$ est pratiquement non biaisé. On obtient alors un estimateur plus juste et donc la variance est certainement plus petite que celle de y_{ij} puisque fonction de σ^2_{enq} et non plus de σ^2_T .

Une étude plus poussée serait nécessaire pour étudier les propriétés de cet estimateur.

. A défaut de faire des corrections a posteriori en utilisant P_i , il est possible, en s'inspirant des résultats de cette analyse, de sensibiliser les enquêteurs à leurs erreurs.

ETUDE DE L'ECHELLE DE NOTATION

1. Description de l'erreur d'évaluation

1.1. Distribution empirique des observations

Chaque année, pour chaque point de débarquement, 1 'équipe "pêche artisanale" constitue un fichier rassemblant l'ensemble des données d'enquête sur le terrain (Cury 81). Dans les fichiers "St-Louis 81" et "Kayar 81", on a sélectionné, sans distinction d'espèces, les lots de poissons dont 1 'enquêteur attaché à cette plage, avait estimé le poids à vue. Pour les valeurs allant de 1 à 1000kg, on a calculé le nombre de lots dont le poids a été estimé à chacune de ces valeurs. Les distributions empiriques des estimations de poids sont présentées fig. 2 .

Dans les deux cas, on remarque :

- la dissymétrie gauche marquée de la distribution, caractéristique de la plupart des distributions. d'abondance de poissons (débarquement au port, chalutage) , quelle que soit la méthode d'estimation de 1 'abondance utilisée ;
- la discrétisation opérée par les enquêteurs. Seules un nombre limité de valeurs entre 0 et 1000 ont des fréquences non nulles. Une distribution d'abondance étant continue, cette discrétisation est le résultat de l'utilisation par l'enquêteur d'une échelle de notation particulière qu'il s'est créé personnellement.

1.2 Echelle de notation

En première approche, on définit comme modalités de l'échelle de notation, toutes les valeurs de la distribution dont les fréquences sont non nulles.

Parmi les modalités ainsi définies, on distingue 2 catégories :

- celles qui contribuent à établir la forme de la distribution dont on sait, par expérience, qu'elle est unimodale ;
- celles qui perturbent cette forme par leur faible fréquence d'apparition.

Ce phénomène résulte du choix de l'échelle de notation qu'a fait l'enquêteur. Celui-ci utilise certaines notes de base qui expriment un ordre de grandeur du poids à estimer, c'est-à-dire un intervalle de poids à l'intérieur duquel il ne se sent pas capable de choisir une valeur plutôt qu'une autre. Il propose

alors la valeur la plus marquante de l' intervalle, en général un multiple de 10, 50 ou 100. Lorsqu' il hésite vraiment entre 2 notes de base, il propose une note intermédiaire qui correspond à une des modalités peu citées.

La notation utilisée par chaque enquêteur est donc assez subjective les enquêteurs de Kayar et de St-Louis ont des attitudes assez différentes.

Celui de Kayar utilise des modalités de base assez classiques (50, 100, 150) avec peu de modalités intermédiaires. Au contraire, celui de St-Louis évite les chiffres "ronds", indiquant 45, plutôt que 50, 160, plutôt que 150, et utilise beaucoup de modalités intermédiaires.

Dans les deux cas, on peut remarquer que la différence des valeurs que prennent 2 modalités de base successives augmente avec le poids. Ceci est assez logique, puisque la précision diminue avec le poids.

2. Modélisation de l'échelle

2.1 Construction du modèle

On considère l'ensemble des modalités de base de l'échelle. A chaque modalité, correspond un intervalle de poids regroupant les valeurs de poids parmi lesquelles l'enquêteur pense que la vraie valeur du poids se trouve.

Soient X et X' deux modalités successives de l'échelle, ΔX et $\Delta X'$, les intervalles correspondant à chaque modalité.

Cet intervalle est une estimation par l'enquêteur de l'erreur absolue qu'il commet.

L'erreur absolue étant proportionnelle au poids, on peut écrire :

$$\Delta X = aX \quad \Delta X' = aX'$$

On fait l'hypothèse que l'ensemble de ces intervalles caractérisant les différents ordres de grandeur d'estimation constituent une partition des valeurs de poids estimables.

On peut alors écrire, en supposant, pour simplifier, que le biais de l'enquêteur est faible ou nul :

$$X' = X + \frac{\Delta X'}{2} + \frac{\Delta X}{2} \quad (\text{l'intervalle est centré sur la modalité, si le biais est nul})$$

$$X' = X + \frac{aX'}{2} + \frac{aX}{2}$$

.../...

$$X' = X \frac{1 + a/2}{1 - a/2} = \gamma X$$

Les valeurs des différentes modalités suivent une progression géométrique de raison γ ::

$$X_i = X_0 \gamma^i$$

2.2 Détermination de la raison γ pour les données de Kayar⁸¹ et St-Louis⁸¹

• Sélection des modalités de base

Une modalité de base de l'échelle est une modalité dont la fréquence est plus grande que celle des modalités environnantes dans la distribution des estimations.

Il est donc nécessaire de comparer, pour chaque modalité, sa fréquence d'apparition à celle des modalités voisines.

La technique suivante a été employée :

Pour chaque modalité, on fait la moyenne des fréquences de cette modalité, des 2 modalités précédentes et suivantes. Si la fréquence de la modalité est supérieure à la moyenne ainsi calculée, on considère cette modalité comme une modalité de base de l'échelle, puisque sa fréquence est localement grande.

• Ajustement du modèle

Chaque modalité est caractérisée par sa valeur X et son rang i dans l'ensemble des modalités, ordonné selon leur valeur.

$$X_i = X_0 \gamma^i$$

On peut obtenir de manière simple des estimations de X_0 et γ en effectuant une régression linéaire de $\log X_i$ en fonction de i , sur la base du modèle :

$$\log X_i = i \log \gamma + \log X_0$$

N'ayant pas d'idées a priori sur les distributions de γ et X_0 , on estime X_0 et γ avec la méthode des moindres carrés.

On obtient :

- pour Kayar, $\gamma_k = 1,34$ $R^2 = 0,95$
- pour St-Louis, $\gamma_s = 1,17$ $R^2 = 0,95$

.../...

Les graphiques des données brutes et des valeurs ajustées sont présentés fig. 3 .

Hier; que globalement satisfaisants, ces ajustements amènent quelques remarques.

Les résidus ne sont pas distribués aléatoirement. On peut diviser l'échelle de notation en plusieurs parties dans chacune desquelles les résidus sont, soit tous positifs, soit tous négatifs. Ceci veut dire que les enquêteurs ne se créent pas une échelle homogène, sur la base d'une précision relative constante des estimations, les modalités étant en surnombre dans certaines parties et en sous-nombre, dans d'autres.

3. Construction d'un intervalle de confiance pour chaque modalité

Soit X , la valeur d'une modalité, c'est-à-dire l'estimation à vue d'un tas de poids réel P . De la première partie de ce travail, on connaît une estimation de la variance de l'erreur relative, $\hat{\sigma}_{\text{enq}}^2 \cdot \left(\frac{y_{ijk} - P_i}{P_i} = \mu + B_i + \varepsilon_{ijk}, \text{var}(\varepsilon_{ijk}) = \sigma_{\text{enq}}^2 \right)$. On estime donc la variance de l'estimateur de P par $X^2 \hat{\sigma}_{\text{enq}}^2$. P se trouve donc compris entre $X - t_\alpha X \hat{\sigma}_{\text{enq}}$ et $X + t_\alpha X \hat{\sigma}_{\text{enq}}$ avec une probabilité $1 - \alpha$, t_α séparatrice de Student pour un seuil α de confiance. On suppose que le biais est nul. ou que l'on effectue une translation de l'intervalle de confiance fonction du biais.

3.1 Comparaison de l'intervalle de confiance et de l'erreur absolue ΔX définie dans le modèle de l'échelle

On a défini l'erreur absolue $\Delta X = aX$

avec $y = \frac{1 + a/2}{1 - a/2}$ y raison de la progression

$$\text{donc } a = 2 \frac{\gamma - 1}{\gamma + 1}$$

On cherche à déterminer quel est le seuil de confiance inconsciemment choisi par l'enquêteur lorsqu'il se crée son échelle de notation.

ΔX est considéré comme un intervalle de confiance. On peut donc écrire :

$$\Delta X = 2 t_\alpha (X \hat{\sigma}_{\text{enq}})$$

.../...

$$\frac{\Delta X}{X} = 2 t_{\alpha} \hat{\sigma}_{\text{enq}} = a$$

$$t_{\alpha} = \frac{a}{2 \hat{\sigma}_{\text{enq}}}$$

On obtient donc :

pour Kayar, $t_{\alpha} = 0,6$ $\alpha > 50 \%$

pour St Louis, $t_{\alpha} = 0,36$ $\alpha > 50 \%$

Dans les deux cas, le risque de première espèce pris par l'enquêteur est très élevé. L'enquêteur surestime donc considérablement sa précision. On peut donc dire que le nombre de modalités de l'échelle est trop grand puisque, lorsque l'enquêteur note le poids à estimer par une des modalités de son échelle, une fois sur deux, une autre modalité serait plus adaptée. La précision de son échelle est donc illusoire puisqu'il n'est pas capable de faire réellement la différence entre deux modalités proches.

3. 2 Construction d'une échelle de notation adaptée à la précision de l'enquêteur

On cherche à déterminer a tel que ΔX soit un intervalle de confiance à une probabilité égale à 0,95.

$$\text{Il faut donc que } t_{0,05} = \frac{a}{2 \hat{\sigma}_{\text{enq}}}$$

$$a = 2 \hat{\sigma}_{\text{enq}} \cdot t_{0,05}$$

$$a = 1 \quad y = 3$$

Les modalités de l'échelle sont donc (1,3, 27, 81, ~~273~~, ~~719~~, 2157,...). Le nombre de modalités a considérablement diminué. Cependant, cette échelle permet d'obtenir des estimations ayant la même précision qu'avec celle actuellement utilisée par les enquêteurs.

justifier
 (1, 3, 27, 81, 2157, ...)
 .../...

Ces résultats laissent penser qu'il serait possible d'obtenir de meilleures estimations en "aidant l'enquêteur", c'est-à-dire en lui facilitant la tâche.

Un individu ne peut proposer qu'un ordre de grandeur du poids à estimer. Il serait préférable de lui demander de citer, non pas un nombre, mais un intervalle. L'enquêteur ne se poserait plus le problème de trancher entre deux nombres voisins, ce qui n'a pas de sens, étant donnée la précision avec laquelle il travaille.

Une échelle de notation, adaptée à sa précision potentielle constituée d'intervalles successives de notation représentant des ordres de grandeur, pourrait lui être proposée. Il est en effet plus facile de travailler dans un cadre précis, et il prend ainsi conscience de la précision qu'il peut avoir.

Un réétalonnage périodique des estimations, comme cela est fait pour les "appareils inertes" de mesures serait nécessaire.

On pourrait proposer une échelle de départ dont les intervalles sont les suivants : (0 - 1,5), (1,5 - 4), (5 - 15), (16 - 40), (40 - 120), (121 - 360), (361 - 1000)...

Il faudrait regarder si les résultats s'améliorent, en particulier, si ce cadre permet de fixer les idées et d'éviter des biais énormes, comme c'est le cas actuellement. On peut penser qu'une amélioration progressive de la précision des enquêteurs serait possible. Un resserrement des modalités de l'échelle serait alors envisageable.

CONCLUSION

Ce travail est une première approche de l'étude de la qualité des estimations de poids à vue. Il est maintenant possible d'avoir une idée des différentes composantes de l'erreur d'évaluation. Il faut cependant garder à l'esprit que les résultats ont été obtenus à partir d'estimations réalisées dans un cadre expérimental assez différent du cadre de travail habituel des enquêteurs, la plage, où, en particulier, les poids à estimer peuvent être beaucoup plus grands et où les poissons sont disposés de manière extrêmement variable sur le sable ou dans les pirogues.

Il est possible, dans une certaine limite, de maîtriser une des composantes de l'erreur, celle qui est due à l'enquêteur. On peut proposer de meilleurs estimateurs plus justes et plus précis. Cette première étude ne permet pas d'aller

jusque là. Les résultats permettent cependant de "réétalonner" un peu les enquêteurs. Un intervalle de confiance autour des vraies valeurs de poids peut être construit.

On pourrait enfin envisager de modifier de manière plus radicale la méthode d'estimation à vue, en encadrant davantage les enquêteurs.

- B I B L I O G R A P H I E -
=====

- . AMANCE, 1981
Manuel de la programmation statistique
I.N.R.A. - C.N.R.F.

- . RADIA, J. ; MASSON, J.P., 1977
Analyse de variance
I.N.R.A., Département de Biométrie

- . COCHRAN, W.G., 1963
Sampling technics
WILEY, New-York, 413p.

- . CURY, P., 1981
Présentation et utilisation des programmes informatiques
de la section pêche artisanale.
Arch. CRODT, 89, 105 p.

- . DÄGNELIE, P., 1973
Théorie et méthodes statistiques
2 vol. Presses Agronomiques, Gembloux, 378 + 463 p.

- . FRONTIER, S. ; VIALE, D., 1977
Utilisation d'une cotation d'abondance mise au point en
planctonologie pour l'évaluation des troupeaux de cétacés
en mer.
J. Rech. Océanogr., 2 (4), 15-22.

- . GENSTAT, 1977
A general statistical program.
Lawes Agricultural Trust (Rothamsted Experimental Station).

- . GOUET, J.P., 1974
Les comparaisons de moyennes de variance. Application à
l'agronomie.
Publication T.T.C.F.