4.1 Introduction et fondements statistiques

Ce chapitre vise à apporter une vue d'ensemble des données requises pour les évaluations et les recherches au sein de l'ICCAT.

Il propose au lecteur une explication fondamentale concernant la façon de concevoir des programmes de collecte de données, à travers l'échantillonnage des prises des bateaux, et la façon de garantir que ces derniers représentent l'ensemble de la population (section 4.2). Bien que cette question se répète tout au long de ce chapitre, nous recommandons aux lecteurs de lire d'abord cette section. Les données de taille constituent un des principaux types de données obtenues de façon routinière. La collecte de ces données et l'utilisation de celles-ci pour estimer la structure démographique des prises sont détaillées dans la section 4.3. La capture par unité d'effort (CPUE), qui est obtenue généralement des livres de bord des bateaux, représente une autre source d'information sur l'état du stock qui est utilisée dans les évaluations. Ces données doivent être standardisées dans le temps et entre les différentes zones, ainsi qu'entre les différentes catégories de navires ou d'engins de pêche afin d'assurer la cohérence des indications contenues dans ces données. Les questions relatives à l'utilisation des données de la CPUE sont détaillées dans la section 4.4.

Une série d'attributs biologiques revêtent une grande importance dans la gestion des stocks. Il s'agit notamment de l'étendue géographique et des limites du stock, de son interaction avec des sous-stocks et de ses caractéristiques de migration. Il existe une série de techniques pour examiner ces facteurs comme les méthodes génétiques (section 4.5) et le marquage (sections 4.6 et 4.7). Dans le cas des stocks définis, la connaissance des schémas de reproduction des grands pélagiques ainsi que des schémas de croissance et de mortalité définira en large mesure la capacité de régénération que possède une population. Ces attributs sont donc très importants pour la gestion et la conservation des stocks, ainsi que pour construire des modèles fiables destinés à évaluer de façon efficace les stocks. Les méthodes utilisées pour examiner ces attributs biologiques sont détaillées dans les sections 4.8 et 4.9. Il faut souligner que la plupart des approches exigent un examen détaillé du poisson. Compte tenu de la valeur élevée que possèdent la plupart des espèces de thonidés, les informations relatives à la taille sont généralement les seules données qui peuvent être recueillies sans qu'il soit nécessaire d'acheter des individus ou d'entreprendre un programme de recherche indépendant de la pêcherie.

Les programmes scientifiques d'observateurs constituent une méthode importante pour recueillir des informations sur une série de caractéristiques liées à une pêcherie comprenant notamment les schémas de prospection, la caractérisation de l'effort de pêche, les prises accessoires et la mortalité par rejet, ainsi que la collecte d'informations biologiques précises. La section 4.10 aborde différentes approches fondamentales pour optimiser la couverture des campagnes de pêche avec des observateurs, les types d'informations pouvant être collectés et l'importance d'estimer les prises accessoires à partir des données des observateurs.

4.1.1 Biostatistiques

Ce manuel contient des informations détaillées sur les méthodes statistiques utilisées dans les domaines étudiés. Pour faciliter la compréhension, cette section offre au lecteur un bref et simple rappel des concepts statistiques de base. S'ils souhaitent obtenir des informations complémentaires, les lecteurs pourront consulter d'autres documents sur les biostatistiques, notamment « Biometry » de Sokal et Rohlf (1995), qui présente à la fois la théorie soutenant les approches et des exemples utiles de leur application aux données, « Sampling techniques » de Cochran (1977) et « Sampling » de Thompson (1992). Sparre et Venema (1998) ont également rédigé un excellent manuel sur l'évaluation du stock de poissons tropicaux dont nous nous sommes largement inspirés pour rédiger cette section.

Valeur moyenne et variance

Considérons un échantillon de n poissons d'une seule espèce, pris tous lors de la même opération de pêche, x(i) étant la taille du poisson nº i, i=1, 2, …, n. La taille moyenne de l'échantillon est définie comme suit :

$$\bar{x} = \frac{\lfloor x_{(1)} + x_{(2)} + \ldots + x_{(n)} \rfloor}{n} = \frac{1}{n} * \sum_{i=1}^{n} x_{(i)}$$

Par exemple, si on a échantillonné 12 poissons ayant une taille (cm) de 176, 175, 162, 174, 161, 156, 178, 158, 195, 171, 177 et 154, la taille moyenne de cet échantillon sera :

$$\bar{x} = \frac{[176 + 175 + ... + 154]}{12} = \frac{1}{12} * 2037 = 169{,}75$$

La variance de l'échantillon, une mesure de la variabilité autour de la valeur moyenne, est définie comme suit :

$$s^2 = \frac{1}{n-1} * \left[\left(x_{(1)} - \bar{x} \right)^2 + \left(x_{(2)} - \bar{x} \right)^2 + ... + \left(x_{(n)} - \bar{x} \right)^2 \right] = \frac{1}{n-1} * \sum_{i=1}^{n} \left[x_{(i)} - \bar{x} \right]^2$$

En général, elle est calculée comme $\sum x^2 - \left(\sum x \right)^2 / n / (n-1)$ pour éviter les erreurs d'arrondi.

La variance est donc la somme des carrés des écarts à la moyenne divisée par le nombre (n) moins un. Si tous les poissons avaient la même taille, la variance serait égale à zéro. Dans l'exemple de tailles indiqué ci-dessus, la variance serait la suivante :

$$s^2 = \frac{1}{12-1} * \left[(176 - 169{,}75)^2 + (175 - 169{,}75)^2 + ... + (154 - 169{,}75)^2 \right] = \frac{1}{12-1} * 1556{,}25 = 141{,}48$$

La racine carrée de la variance, s, est l'écart-type. Dans l'exemple, s=11,89. La variance peut aussi être exprimée par rapport à la taille de la moyenne, en tant que coefficient de variation. L'écart-type est important à cet égard car il possède la même unité que la moyenne. Le coefficient de variation est le suivant :

$$\frac{s}{\bar{x}}$$

Dans l'exemple, le coefficient de variation (CV) est le suivant :

$$\frac{11{,}89}{169{,}75} = 0{,}07$$

Les statistiques se fondent en grande mesure sur la « normalité » des données. Ceci signifie essentiellement que les données (et la population d'où elles sont extraites) correspondent à une distribution normale :

$$Fc(x) = \frac{n * dL}{s * \sqrt{2\pi}} * \exp\left[-\frac{\left(x - \bar{x} \right)^2}{2s^2} \right]$$

où Fc(x) est la « fréquence calculée », n le nombre d'observations, dL la dimension de l'intervalle (de la mesure en question), s=l'écart-type, $\bar{x}$ la taille moyenne et π=3,14159…

Une distribution normale est souvent observée chez les poissons plus âgés et plus grands (les petits et jeunes poissons auraient besoin de quelques valeurs négatives pour avoir une distribution normale) lorsqu'on enregistre des fréquences de tailles de poissons issus d'une seule cohorte (c'est-à-dire du même âge), ce qui permet d'estimer la probabilité d'avoir des poissons d'une taille supérieure ou inférieure à la taille donnée dans l'échantillon. Il existe d'autres distributions de probabilité (ex. lognormale), dans lesquelles la distribution des mesures est asymétrique au lieu d'être centrée autour de la moyenne comme dans la distribution normale.

Les concepts de biais et de précision proviennent de la considération des moyennes et des distributions (**Figure 4.1.1**). Une estimation d'un échantillon est dite non biaisée si la moyenne de plusieurs estimations répétées est la même que la valeur vraie (ce qui serait le cas si tous les individus de la population totale étaient échantillonnés). Une estimation est biaisée si elle présente un écart systématique par rapport à la valeur vraie. Ce serait le cas, par exemple, si les estimations de la taille moyenne des échantillons étaient toujours supérieures à la taille vraie de la population, ce qui pourrait être dû à la sélectivité de l'engin. Dans le cas d'un échantillon non biaisé, on peut approcher davantage la valeur vraie en augmentant la taille de l'échantillon. C'est ce qu'on appelle la « cohérence » (vérifier si c'est bien ce terme qui est utilisé par les statisticiens). Dans le cas d'un échantillon biaisé, il y aura toujours une différence entre la valeur vraie et la valeur estimée.

Il faudrait prendre un échantillon aléatoire pour obtenir une estimation non biaisée. Dans ce cas, n'importe quel poisson échantillonné dans le stock (à titre d'exemple) devrait avoir exactement la même probabilité d'être échantillonné. Ceci dit, dans la pratique il est souvent difficile d'obtenir de véritables échantillons aléatoires.

La précision est une mesure consistant à déterminer si des échantillons ou des estimations sont « précis ». Dans ce cas, la variance autour de la valeur moyenne de l'échantillon ou de l'estimation est faible (**Figure 4.1.1**). Ceci ne signifie pas nécessairement que l'échantillon ou l'estimation est non biaisé : en effet, ceux-ci peuvent être précis (regroupés autour d'une valeur donnée), mais biaisés (par exemple, si cette valeur n'est pas égale à la moyenne vraie).

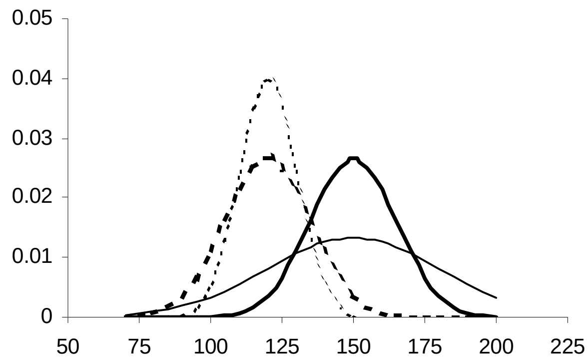

Figure 4.1.1. Démonstration du biais et de la précision. La distribution normale indiquée par le trait gras continu représente la distribution de la population. Le trait fin continu représente une distribution non biaisée, mais moins précise (la moyenne est la même que celle de la population, mais la dispersion est plus grande). La distribution représentée par le trait gras discontinu représente un échantillon biaisé : la distribution possède une variance identique par rapport à la vraie population, mais la valeur moyenne est inférieure à la valeur vraie. La distribution représentée par le trait fin discontinu est biaisée, mais plus précise.

4.1.2 Bibliographie

COCHRAN, W.G. (1977). Sampling techniques. New York, J. Wiley & Sons, Inc.

SOKAL, R.R. and F.J. Rohlf (1995). Biometry: the principles and practice of statistics in biological research. W.H. Freeman and Company, New York.

SPARRE, P. and S.C. Venema (1998). Introduction to fish stock assessment. Part 1. Manual. FAO Fish. Tech. Pap. 306(1), FAO, Rome.

THOMPSON, S. K. (1992). Sampling. John Wiley & Sons, Inc. 343p.