ВОПРОСЫ РЫБОЛОВСТВА, 2015, том 16, №4; с. 489–496
PROBLEMS OF FISHERIES, 2015, vol. 16, №4; p. 489–496

МЕТОДИЧЕСКИЕ АСПЕКТЫ РЫБОХОЗЯЙСТВЕННЫХ ИССЛЕДОВАНИЙ

УДК: 639.2; 519.22; 51-76

МАТЕМАТИЧЕСКИЕ АСПЕКТЫ СТАНДАРТИЗАЦИИ УЛОВОВ НА УСИЛИЕ

© 2015 г. А. И. Михайлов

Всероссийский научно-исследовательский институт рыбного хозяйства и океанографии, Москва, 107140
E-mail: mikhailov1984@gmail.com

Поступила в редакцию 14.10.2015 г.

Статья представляет собой обзор методов стандартизации уловов на усилие. Первоочередной целью является изложение математических основ обобщенных линейных моделей для широкого круга отраслевых специалистов. Показывается связь между обобщенными линейными моделями и колмогоровским средним. Также намечаются некоторые пути повышения робастности процедуры стандартизации.

Ключевые слова: математическое моделирование, стандартизация уловов на усилие, обобщенные линейные модели (GLM).

ВВЕДЕНИЕ

Как известно, оценка запасов, основанная на продукционных или когортных моделях, так или иначе использует индексы численности. Основным источником данных для оценки индекса численности наряду со съемками, претендующими на непосредственную оценку запаса, являются данные промысловой статистки. Улов на усилие является одним из важнейших источников статистической информации о динамике системы запас—промысел. Поскольку промысел ведется судами различных типов с различными орудиями лова, возникает задача стандартизации промысловых данных. Наблюдаемый улов на усилие (u) является функцией не только численности или биомассы запаса, но и ряда других факторов ($x_1 \ldots x_d$):

$$u = u(I, x_1 \ldots x_d). \quad (1)$$

Необходимо построить модель наблюдения (1), учитывающую все факторы влияния и разрешить ее относительно неизвестного индекса численности. Этот индекс численности и рассматривается как стандартизированный улов на усилие. Стандартизированное промысловое усилие E_{st}, т.е. такое усилие, которое обеспечивало бы данный улов, если бы все условия промысла соответствовали условиям, принятым за стандартные, вычисляется как отношение улова к индексу численности:

$$E_{\mathrm{st}} = \frac{C}{I}. \quad (2)$$

Следует отметить, что в традиционной процедуре стандартизации (Beverton, Holt, 1957) формула (2) используется для оценки индекса численности, а стандартизированное усилие вычисляется по другой формуле:

$$E_{\mathrm{st},f} = \left(\overline{C_f / E_f} \big/ \overline{C_{\mathrm{st}} / E_{\mathrm{st}}}\right) E_f, \quad (3)$$

где верхняя черта означает усреднение за весь период наблюдений, нижний индекс f — тип судна флота и тип судна, а индекс st — то же, принятое за стандартное.

Наиболее простым случаем, когда необходима стандартизация улова на усилие, является ситуация, когда доступны несколько временных рядов уловов на усилие, интерпретируемых как индексы численности. В этом случае наиболее естественной гипотезой будет мультипликативная связь между наблюдаемым уловом на усилие и индексом численности, при этом ошибка также будет мультипликативной и логнормально распределенной:

$$CPUE_{t,f} = q_f I_t \exp(\varepsilon), \qquad (4)$$

где q_f — коэффициент улавливаемости для соответствующего флота или типа судна, а ε — нормально распределенная случайная величина с нулевым математическим ожиданием.

В этом случае оценки индекса численности и коэффициента улавливаемости легко получить аналитически:

$$I_t = \exp\left(\frac{1}{m}\sum_f \ln CPUE_{t,f}\right); \qquad (5)$$

$$q_f = \exp\left(-\frac{1}{Tm}\sum_{t,f} \ln CPUE_{t,f} + \right.$$

$$\left. + \frac{1}{T}\sum_t \ln CPUE_{t,f}\right), \qquad (6)$$

где m — количество категорий (флотов или типов судов), T — период наблюдения.

Метод такого типа впервые был предложен Гаварисом (Gavaris,1980) и, хотя техника статистических вычислений значительно усложнилась, концептуальная основа стандартизации осталась, в общем-то, той же — необходимо с наименьшей ошибкой представить имеющийся большой массив промысловых данных как сочетание зависящего только от времени индекса численности и влияния иных факторов, непосредственно от времени не зависящего. К настоящему моменту в международной практике оценки запасов наиболее широкое распространение получили методы стандартизации, основанные на применении обобщенных линейных моделей (Maunder, Punt, 2004), изложению математических основ которых и посвящена эта работа.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Обобщенные линейные модели

Стандартизация уловов на усилие с использованием метода обобщенных линейных моделей (generalized linear models, далее — GLM) есть не что иное, как настройка модели наблюдения. Модель наблюдения — это статистическая гипотеза о распределении зависимой переменной, в данном случае улова на усилие u. Обобщенная линейная модель предписывает рассматривать распределения наблюдаемой величины только из экспоненциального семейства, общий вид которого задается следующим образом (McCullagh, Nelder,1989):

$$\exp(u\eta - L(u) - G(\eta)). \qquad (7)$$

Величина η называется предиктором модели, L и G — некоторые функции зависимой переменной и предиктора соответственно, которые будут описаны ниже. Все статистические моменты наблюдаемой величины будут функциями предиктора. Статистическая модель называется обобщенной аддитивной (Venables, Dichmont, 2004; Xiao et al, 2004), если предиктор как функция многих независимых переменных аддитивно факторизуема, т.е. представляет собой сумму функций одной переменной Y, $\{f_k\}$ и $\{y_k\}$:

$$\eta(t; x_1, x_2, \ldots, x_n; j_1 \ldots j_m) = Y(t) + f_1(x_1) +$$

$$+ f_2(x_2) + \ldots + f_n(x_n) + y_1(j_1) + y_2(j_2) +$$

$$+ \ldots + y_m(j_m). \qquad (8)$$

Здесь t — время, $\{x_k\}$ и $\{j_j\}$ — независимые переменные. Будучи одной из наиболее общих статистических моделей, обобщенная аддитивная модель для свой идентификации нуждается в некоторых дополнительных предположениях, поскольку каждая из функций одного переменного задается счетным числом параметров, а поскольку оценка бесконечного числа параметров невозможна, следует выбрать определенный вид функции. Множество значений независимых переменных может быть как непрерывным, так и дискретным. В последнем случае такие переменные называются категориальными. В теории GLM принято предположение о линейной зависимости предиктора от непрерывных факторов:

$$\eta(t; x_1, x_2, \ldots, x_n; j_1 \ldots j_m) = Y(t) + \beta_1 x_1 + \\ + \beta_2 x_2 \ldots + \beta_n x_n + y_1(j_1) + y_2(j_2) + \\ + \ldots + y_m(j_m). \quad (9)$$

Коэффициенты при факторах $Y(t)$, $\{\beta_k\}$ $\{y_\rho\ (j_\rho)\}$ являются неизвестными параметрами, которые необходимо определить при настройке модели. Благодаря факторизации количество наблюдений значительно превышает число оцениваемых параметров, и именно поэтому аппарат статистики может быть применен.

В обобщенных линейных и аддитивных моделях предполагается, что независимые переменные не коррелированы между собой, однако это не всегда верно. Для того чтобы учесть корреляцию переменных предиктора, используются обобщенные линейные смешанные модели (Helser et al., 2004). Формально обобщенная линейная смешанная модель означает включение в линейный предиктор дополнительных членов вида ζz, где z — случайная величина с известным (обычно нормальным, но не обязательно) законом распределения, а ζ — параметр. Обобщенные смешанные модели являются аналогом линейной фильтрации в многомерном случае: линейная комбинация независимых белых шумов позволяет получить нетривиальную автокорреляционную функцию предиктора, приближающую эмпирическую вариограмму.

В уравнении (7) $L\ (u)$ — есть взятая с обратным знаком функция правдоподобия, имеющая смысл логарифма плотности вероятности некоторого, достаточно произвольного, распределения. Установим связь между функциями $L\ (u)$ и $G\ (\eta)$. Для этого вспомним, что плотность вероятности нормирована на единицу:

$$\int \exp(\ u\eta\ \text{-}\ L(u)\ \text{-}G(\eta))\ =\ 1\ \ \forall \eta. \quad (10)$$

Тем самым мы получаем, что функция $G\ (\eta)$ не является произвольной, а однозначно восстанавливается по $L\ (u)$ преобразованием Лапласа:

$$\exp(G(\eta)) = \int \exp(\ u\eta\ \text{-}\ L(u))\ du. \quad (11)$$

Последнее есть не что иное, как производящая функция моментов распределения, чья плотность соответствует функции правдоподобия $L(u)$. Таким образом, для всякого распределения, обладающего конечной производящей функцией моментов, можно построить содержащее его экспоненциальное семейство.

Исследуем свойства случайных величин, закон распределения которых принадлежит экспоненциальному семейству. В первую очередь изучим моменты такого распределения. Для этого проще всего вычислить производящую функцию моментов распределения, по определению задаваемую следующей формулой:

$$E \exp(tu) = \exp(G(\eta + t)\ \text{-}\ G(\eta)). \quad (12)$$

Для вычисления n-го момента функции распределения достаточно соответствующее число раз продифференцировать в нуле производящую функцию:

$$M^{(n)} = \partial^n E \exp(tu)\Big|_{t=0}. \quad (13)$$

Первым моментом является само математическое ожидание, для которого использование формулы (13) дает следующий результат:

$$\mu(\eta) \equiv Eu\ =\ \frac{\partial G(\eta)}{\partial \eta} \equiv\ \phi^{-1}(\eta). \quad (14)$$

Здесь мы ввели новое обозначение — функцию связи $\varphi\ (\mu)$, которая является строго монотонной и однозначно определяемой законом распределения. У всякой монотонной функции существует обратная ей функция. Линейный предиктор есть обратная функция связи от математического ожидания. Таким образом, зависимость между математическим ожиданием и линейным предиктором однозначно определяется обратной функция связи.

Иными словами, обобщению линейную модель можно записать в следующем виде:

$$\phi(Eu) = \eta = Y(t) + \beta_1 x_1 + \beta_2 x_2 \ldots + \\ + \beta_n x_n + y_1(j_1) + y_2(j_2) + \ldots + y_m(j_m). \quad (15)$$

Функция связи задает некоторое нелинейное преобразование зависимой переменной, после которого задача становится похожа на таковую линейной регрессии. Усреднив статистику наблюдений и выполнив обратное преобразование, мы можем получить следующую формулу:

$$\phi^{-1}(\frac{1}{N}\sum_{k=1}^{N}\phi(u_k)) = \phi^{-1}(\eta) \equiv \mu(\eta) = Eu. \qquad (16)$$

Правая часть оказывается не чем иным, как средним по Колмогорову (1985), который показал, что только такие функции не меняются при замене переменных на результат усреднения, т.е. указанная процедура является наиболее общей формой усреднения. Таким образом, мы получили, что оценка математического ожидания в обобщенной модели задается колмогоровским средним, построенным с помощью функции связи.

Теперь вычислим с помощью формулы (13) второй центрированный момент — дисперсию:

$$D(u) = \partial^2_{\eta\eta}G(\eta) = \frac{\partial\mu(\eta)}{\partial\eta}. \qquad (17)$$

Дисперсия D (в общем случае — матрица ковариаций) является функцией линейного предиктора, а значит, и математического ожидания:

$$V(\mu) = \frac{1}{\phi'(\mu)}. \qquad (18)$$

Здесь было введено еще одно новое обозначение $V(\mu)$ для функциональной связи дисперсии и математического ожидания, а также получен его явный вид с использованием функции связи.

Перечисленные выше свойства наряду с нелинейной связью математического ожидания и предиктора являются ключевыми в практике применения модели GLM, поскольку позволяют моделировать данные со значительной гетероскедастичностью, т.е. с высокой неоднородностью параметров распределения, прежде всего, с различной дисперсией. Все эти свойства экспоненциального семейства распределений обусловлены тем, что функция связи однозначно задает вид распределения и все его моменты.

Варьируя функцию правдоподобия по параметрам, построим уравнение регрессии:

$$\sum_{i=1}^{N}(u_i - \frac{\partial G}{\partial\eta_i})\frac{\partial\eta_i}{\partial\beta_\alpha} = 0. \qquad (19)$$

Здесь i — номер элемента выборки, N — размер выборки; $\beta\alpha = (\beta_0, \beta_1 \dots \beta_d)$ — вектор параметров, причем нулевая компонента соответствует свободному члену, а размерность d — количеству независимых факторов.

Это уравнение в отличие от уравнения линейной регрессии нелинейно относительно параметров. Уравнение линейной регрессии является частным случаем данного уравнения, когда функция связи линейна. Стандартный метод решения нелинейных алгебраических уравнений — это итерационная процедура (O'Brien, Kell, 1997), например метод Ньютона или какой-либо другой градиентный метод (Рябенький, 2000). Выведем соответствующие формулы для вектора параметров в явном виде. Для этого введем следующее обозначение:

$$X = \frac{\partial\eta_i}{\partial\beta_\alpha}, \qquad (20)$$

где X — это матрица размером $N\times(d+1)$, состоящая из единичного столбца высотой N и столбцов независимых переменных. Теперь, воспользовавшись (14), заметим, что отклонения наблюдений от математического ожидания имеют вид:

$$\varepsilon_i = u_i - \mu(\eta_i). \qquad (21)$$

Таким образом, уравнение регрессии можно записать в виде:

$$\varepsilon^T X = 0. \qquad (22)$$

Здесь ε^T — транспонированный вектор (строка) остатков модели, размерность которого совпадает с размером выборки. Как уже отмечалось выше, этот вектор зависит нелинейным образом от вектора параметров существенно меньшей размерности. Разложим эту зависимость в ряд Тейлора в окрестности приближенного решения:

$$\varepsilon_i = \varepsilon_i(\beta^{(k)}) - \frac{\partial\mu(\beta x_i)}{\partial\beta}(\beta - \beta^{(k)}). \qquad (23)$$

Далее, используя формулу для сложной производной, а также выражения (18) и (20), получим:

$$\frac{\partial\mu(\eta_i(\beta)}{\partial\beta} = \frac{\partial\mu(\eta)}{\partial\eta}\frac{\partial\eta_i}{\partial\beta} = V(\mu_i)X_\alpha^i. \qquad (24)$$

Пусть V — диагональная матрица размером $N \times N$, состоящая из элементов V (μ_i), тогда (23) можно записать в матричном виде:

$$\varepsilon = \varepsilon(\beta^{(k)}) - VX(\beta - \beta^{(k)}). \qquad (25)$$

Таким образом, для получения итерационной процедуры оценки параметров обобщенной линейной модели необходимо найти решение системы линейных уравнений (19) или (23), выраженной в матричной форме следующим образом:

$$X^T\varepsilon(\beta^{(k)}) - X^TVX(\beta - \beta^{(k)}) = 0. \qquad (26)$$

Это уравнение легко разрешить, если матрица X^TVX размером $(d+1)\times(d+1)$ не вырождена, т.е.:

$$\det X^TVX \neq 0. \qquad (27)$$

Итак, мы получаем выражение для следующей итерации вектора параметров:

$$\beta^{(k+1)} = (X^TV(\beta^{(k)})X)^{-1}X^T\varepsilon(\beta^{(k)}) + \beta^{(k)}. \ (28)$$

Именно эта вычислительная процедура подлежит самому тщательному исследованию на устойчивость. Однако, прежде чем ставить вопрос об устойчивости, необходимо убедиться, что процедура сходится. Для этого перепишем рекуррентное соотношение (28) в виде формального ряда:

$$\beta = \beta^{(0)} + \sum_{ё=1}^{\infty}(X^TV(\beta^{(k)})X)^{-1}X^T\varepsilon(\beta^{(k)}). \ (29)$$

При получении априорной оценки снизу на член ряда можно получить минимальное значение радиуса сходимости ряда. Начальное приближение вектора параметров не должно отклоняться от истинного более чем на радиус сходимости. В качестве стартового значения рекуррентной процедуры можно использовать оценки одномерных регрессий:

$$\beta_\alpha^{(0)} = \frac{\phi^{-1}(\frac{1}{N}\sum_{k=1}^{N}\phi(u_k))}{\bar{x}_\alpha}. \qquad (30)$$

В силу выражения (16) предварительной оценкой каждого из параметров окажется отношение колмогоровского среднего зависимой переменной к среднему значению соответствующей независимой переменной.

Достаточным условием сходимости процедур (28) или (29) является сжимаемость соответствующих отображений:

$$\left\| 1 + (X^TV(\bullet)X)^{-1}X^T(u - \mu(X\bullet) \ \right\| < 1, \ (31)$$

где ● — знак операнда отображения. Разложение отображений (28) или (29) в окрестности решения до второго порядка по β преобразует условие сходимости к виду:

$$\left\| \frac{1}{2}(X^TV(\beta)X)^{-1}(\frac{d}{d\beta}(X^TV(\beta)X))\delta\beta \otimes \delta\beta \right\| < 1, (32)$$

где δβ — отклонение от истинного значения вектора параметров. Таким образом, итерационная процедура сходится квадратично:

$$\sum \frac{d\lambda_i}{d\beta_j}\delta\beta_j^2 < 1, \qquad (33)$$

где λ_i — собственные числа матрицы $\ln X^TVX$, т.е. решения следующего уравнения:

$$\det(\ln(X^TVX)^{-1} - \lambda 1) = 0. \qquad (34)$$

Несмещенность оценки предполагает выполнение следующего условия. Если каждая из компонент вектора наблюдений распределена по закону, принадлежащему экспоненциальному семейству, то математическое ожидание соответствующей оценки совпадает с истинным математическим ожиданием зависимой переменной:

$$E\mu(X\beta^{(\infty)}) = Eu, \qquad (35)$$

где β∞ — финальная оценка вектора параметров.

Требование состоятельности оценки означает, что оценка вектора параметров сходится по вероятности к своему истинному значению при увеличении объема выборки:

$$\forall \varepsilon > 0, \quad P(\left|\beta^{(\infty)} - \beta\right| < \varepsilon) \xrightarrow{N \to \infty} 0. \quad (36)$$

Для исследования устойчивости оценки параметров рассматриваются малые изменения во входных данных — значения зависимых и независимых переменных:

$$d\beta = (X^T V X)^{-1}((dX)^T u - (dX)^T V X \beta -$$
$$- (X^T V dX)\beta), \quad (37)$$

$$d\beta = (X^T V X)^{-1} X^T du. \quad (38)$$

Реакция на отклонение зависимой переменной линейна, а на отклонение в независимых переменных — нелинейна. Указанная нелинейность значительно усложняет исследование процедуры оценки параметров на робастность. Если при исследовании состоятельности оценки предполагалась истинность гипотезы о распределении, то робастность требует, чтобы оценка изменялась мало, если к исходному распределению примешивается небольшое число данных с существенно иным законом распределения, например, не имеющих математического ожидания или с бесконечной дисперсией.

$$F_\varepsilon(\beta) = (1 - \varepsilon)\, F_\varepsilon(\beta) + \varepsilon R \quad (39)$$

где $F_\varepsilon(\beta)$ — реальное распределение выборки, ε — малый параметр смешивания, $F(\beta)$ — теоретическое распределение из экспоненциального семейства, а R — некоторое распределение с бесконечными моментами.

Итак, для оценки параметров обобщенной линейной модели необходимо априори задаться некоторой функцией связи и зависимостью дисперсии от математического ожидания, т.е. принять гипотезу о функции распределения. Однако распределение не является заранее известным, поэтому прибегают к байесовским процедурам (Cao et al., 2011). Помимо байесовских методов неопределенность в исходных гипотезах может быть устранена с помощью робастных оценок параметров на основе ранговых критериев (Хьюбер, 1984). Разработка такого рода методов оценивания параметров обобщенных линейных моделей является весьма актуальной проблемой, отдельные возможные пути решения которой будут намечены ниже.

ЗАКЛЮЧЕНИЕ

В результате проведенного анализа было показано, что даже наиболее современные процедуры стандартизации уловов на усилие, основанные на обобщенных линейных и аддитивных моделях, неустойчивы по отношению к ошибкам в исходных данных и гипотезах о распределении уловов на усилие. Тем не менее использование обобщенных линейных моделей и аналогичных им методов является объективной необходимостью в силу ряда причин. Во-первых, аппарат обобщенных линейных моделей позволяет объяснять наблюдаемый улов на усилие с учетом максимального количества известных факторов. Во-вторых, обобщенные линейные модели охватывают максимально широкий класс распределений. В-третьих, для экспоненциального семейства распределений, на использовании которого основаны обобщенные линейные модели, доказано, что средние значения зависимой переменой являются минимальной достаточной статистикой для определения параметров распределения.

Для повышения робастности стандартизации уловов на усилия с помощью обобщенных линейных моделей предложено использовать следующие методы. Первый метод — использование функции псевдоправдоподобия Q (μ, u) в качестве функционала минимизации (McCullagh, Nelder, 1989):

$$Q(\mu, u) = \int_u^\mu \frac{a(\varphi)(u - t)}{V(t)} dt, \quad (40)$$

где φ — параметр дисперсии, $a(\varphi)$ — функция, определяемая видом распределения.

Функция псевдоправдоподобия обладает всеми свойствами функции правдоподобия для распределения из экспоненциального семейства — ее первая и вторая производные определяют математическое ожидание и дисперсию соответственно. Поэтому данный метод позволяет получать оценки индекса численности, не уточняя, к какому конкретно распределению из экспоненциального семейства принадлежат его наблюдаемые значения. При этом, будучи методом параметрической статистики, по отношению к ошибкам в исходных данных этот метод остается неустойчивым.

Второй метод — использование байесовской процедуры оценивания (Cao et al., 2011). Этот метод более устойчив по отношению к ошибкам исходной гипотезы о виде функции распределения, однако требует задания априорной вероятности принадлежности наблюденного улова на усилие к тому или иному распределению.

Третий метод — использование медианных или модальных оценок для индекса численности. Медианные методы наиболее робастные, однако существуют трудности в обобщении методов этого типа на распределения произвольного типа. Методы модальной регрессии более перспективны; так, ниже показана явная связь между модой распределения и линейным предиктором, определяющим индекс численности, для нескольких наиболее часто используемых распределений. Например, для нормального распределения математическое ожидание, мода, медиана и линейный предиктор совпадают. В случае обратного гауссовского распределения, часто встречающегося в приложениях, аналитическое вычисление медианы невозможно, однако мода имеет явное выражение:

$$\text{mode} = \mu\left(\sqrt{1+\frac{9\mu^2}{4\varphi^2}}-\frac{3\mu}{2\varphi}\right) = \frac{1}{\sqrt{\eta}}\left(\sqrt{1+\frac{9}{4\eta\varphi^2}}-\frac{3}{2\varphi\eta}\right). \tag{41}$$

Таким образом, дальнейшая разработка медианных и модальных методов оценивания наиболее перспективна для повышения робастности процедуры стандартизации уловов на усилие.

На практике рекомендуется применять либо метод псевдоправдоподобия, поскольку для него существует программное обеспечение, либо медианную регрессию. Хотя медианная регрессия обычно требует симметричности распределения, высокая робастность медианы в ряде случаев позволяет игнорировать асимметрию при больших отклонениях.

Работа поддержана грантом РНФ № 14-11-00687.

СПИСОК ЛИТЕРАТУРЫ

Колмогоров А.Н. Математика и механика. Избранные труды. Т. 1. / Под ред. С.М. Никольского. М.: Наука, 1985. С. 136–138.

Рябенький В.С. Введение в вычислительную математику. М.: Физматлит, 2000. 296 с.

Хьюбер Дж.П. Робастность в статистике: М.: Мир, 1984. 304 с.

Beverton R.J.H., Holt S.J. On the dynamics of exploited fish populations. London: HMSO, 1957. 533 p.

Gavaris S. Use of a multiplicative model to estimate catch rate and effort from commercial data // Can. J. Fish. Aquat. Sci. 1980. V. 37. P. 2272–2275.

Helser T.E., Punt A.E., Methot R.D. A generalized linear mixed model analysis of a multi-vessel fishery resource survey original research article // Fish. Res. 2004. V. 70. Iss. 2–3. P. 251–264.

Jie Cao, Xinjun Chen, Yong Chen et al. Generalized linear Bayesian models for standardizing CPUE: an application to a squid-jigging fishery in the northwest Pacific Ocean // Sci. Marina. 2011. V. 75. № 4. P. 679–689.

Maunder M.N., Punt A.E. Standardizing catch and effort data: a review of recent approaches // Fish. Res. 2004. V. 70. Iss. 2–3. P. 141–159.

McCulagh P., Nelder J. Generalized linear models // London: Chapman and Hall, 1989. 511 p.

O'Brien C.M., Kell L.T. The use of generalized linear models for the modelling of catch-effort series. I-Theory // Col. Vol. Sci. Pap. ICCAT. 1997. V. 46. № 4. P. 476–482.

Venables W.N., Dichmont C.M. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research // Fish. Res. 2004. V. 70. Iss. 2–3. P. 319–337.

Xiao Y., Punt A.E., Millar R.B., Quinn T.J. Models in fisheries research: GLMs, GAMS and GLMMs // Ibid. 2004. V. 70. Iss. 2–3. P. 137–139.

MATHEMATICAL ASPECTS OF CPUE STANDARDIZATION

© 2015 y. A.I. Mikhaylov

Russian Federal Research Institute of Fishery and Oceanography, Moscow, 107140

This article is proceeding of the report on the methodologically stock assessment workshop in 2014. The report was a review of methods for standardization of catch on effort. The primary purpose of this report was the presentation of the mathematical foundations of generalized linear models for a wide range of fishery researchers. In particular the relationship between generalized linear models and the Kolmogorov average is shown. Also outlines some ways to improve the robustness of the standardization process.

Keywords: mathematical modeling, CPUE standartization, generalized linear models (GLM).