

УДК 639.2.053.7:681.3

РЕГРЕССИОННЫЙ АНАЛИЗ КАК МЕТОД ПРОГНОЗИРОВАНИЯ

З. М. Аксютина, К. Г. Янковская

Рациональное управление рыбным хозяйством невозможно без количественных оценок и прогнозирования сырьевой базы.

Методы прогнозирования относительной численности промысловых рыб, применяемые в настоящее время в рыбохозяйственной науке, можно отнести к методам исследования операций. Они основаны на анализе изменений размерного и возрастного состава уловов, учета относительной численности поколений на разных этапах их жизни, изучении основных закономерностей, определяющих относительное изменение запасов (Дементьева, 1970).

Существуют они свыше 30 лет, и прогнозы их в общем оправдываются. В настоящее время поставлена задача автоматизации прогнозирования и поэтому необходимы математическое исследование и формализация этих методов в виде программ для ЭВМ.

Задача создания автоматизированной системы управления рыбным хозяйством предъявляет к прогностической информации жесткие требования оперативной выдачи долгосрочных и краткосрочных прогнозов о состоянии сырьевой базы рыболовства и ее возможном использовании.

Прогнозирование, связанное с глубинными биологическими процессами, с учетом влияния сложной и изменчивой среды обитания, на основе которого должна строиться вся хозяйственная деятельность рыбной промышленности, требует резкого повышения степени точности, а следовательно, использования современных математических и кибернетических методов исследования (Аксютина, 1970; Засосов, 1970).

В этой работе рассмотрена возможность количественного прогнозирования сырьевой базы при помощи регрессионного анализа.

Состояние рыбных запасов определяется множеством факторов, влияние которых на запасы различно и количественно мало изучено. Если из них априори может быть выделено m генеральных факторов X_1, X_2, \dots, X_m и получены количественные данные n независимых наблюдений над ними и соответствующими им значениями запаса (или величинами уловов), причем $n \geq m$, то задачу прогнозирования можно сформулировать как задачу составления по данным

$$y_k, x_{1k}, x_{2k}, \dots, x_{mk}, \quad (1)$$

где $k=1, 2, \dots, n$,
такой функции отклика

$$M(Y) = f(X_1, X_2, \dots, X_m), \quad (2)$$

по которой можно определить математическое ожидание запаса (или улова) для каждой конкретной системы значений x_1, x_2, \dots, x_m .

Величина Y предполагается случайной, так как учитываемые переменные x_i , где $i=1, 2, \dots, m$, не исчерпывают всей совокупности факторов, определяющих величину Y . В числе последних окажутся случайные величины, благодаря которым Y будет испытывать, кроме обусловленных функционально, и случайные воздействия. Поэтому переменные X_i определяют величину Y лишь в среднем как $M(Y)$, не исключая случайных флюктуаций (Смирнов, Дунин-Барковский, 1965; Френкель, 1968).

По функции отклика (2) может осуществляться краткосрочное прогнозирование, если система X_1, X_2, \dots, X_m будет системой значений гидрометеорологических данных, характеризующих перераспределение скоплений рыбы в ближайшие сроки или долгосрочное, если факторы X_i определяют величину запаса через длительный промежуток времени (например, X_1 — величина нерестового стада, X_2 — температура воды во время нереста, X_3 — температура воздуха в период созревания икры, X_4 — величина стока рек и т. д.).

Представление о неизвестной функции (2) можно получить, аппроксимируя ее одним из уравнений множественной регрессии: линейным

$$\hat{y} = a_0 + a_1 x_1 + \dots + a_m x_m,$$

линейным относительно логарифмов

$$\ln \hat{y} = a_0 + a_1 \ln x_1 + \dots + a_m \ln x_m$$

или степенным

$$\hat{y} = e^{a_0} x_1^{a_1} \dots x_m^{a_m},$$

или в общем виде

$$\hat{y} = a_0 + a_1 f(x_1) + \dots + a_m f(x_m),$$

где $f_i(x_i)$ — заранее заданные функции.

Решение задачи прогнозирования методами множественной регрессии можно разбить на следующие этапы:

- 1) отбор генеральных факторов X_1, X_2, \dots, X_m , определяющих величину запаса Y ;
- 2) сбор данных n наблюдений над системой значений $y_k, x_{1k}, x_{2k}, \dots, x_{mk}$;
- 3) выбор вида уравнения регрессии, аппроксимирующего функцию (2);
- 4) определение коэффициентов уравнения регрессии;
- 5) установление практической ценности полученного уравнения;
- 6) измерение факторов X_1, X_2, \dots, X_m для конкретного прогноза;
- 7) составление прогноза по полученному уравнению и конкретным значениями x_1, x_2, \dots, x_m ;
- 8) проверка соответствия прогноза фактической величине запаса;
- 9) уточнение уравнения прогноза по данным $n+1$ наблюдений.

Выполнение пунктов 1 и 2 требует максимальной изученности объекта прогнозирования и квалифицированного сбора количественных данных.

Мерой степени зависимости запаса Y от системы учтенных факторов X_i может служить множественное корреляционное отношение

$$\eta = \sqrt{1 - \frac{\sum_{k=1}^n (y_k - \hat{y}_k)^2}{\sum_{k=1}^n (y_k - \bar{y})^2}} \quad (3)$$

Корреляционное отношение (3) может принимать значения $0 \ll \eta \ll 1$. Близость его значений к единице говорит о значительной тесноте связи и, следовательно, об удачном отборе генеральных факторов. При линейной зависимости для оценки степени связи может быть использован и коэффициент множественной корреляции.

Вид уравнения в случае множественной регрессии приходится выбирать в основном эмпирически: предполагать его заданным, определять коэффициенты и оценивать полученное уравнение путем сопоставления с данными наблюдений и уравнениями других видов.

Коэффициенты a_i выбранного уравнения регрессии могут быть определены методом наименьших квадратов (Френкель, 1968). Для этого результаты наблюдений представляются в матричной форме:

$$X = \begin{pmatrix} x_{01}x_{11} \dots x_{m1} \\ x_{02}x_{12} \dots x_{m2} \\ \dots \\ x_{0n}x_{1n} \dots x_{mn} \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

или в случае преобразованных переменных

$$X = \begin{pmatrix} f_0(x_{01})f_1(x_{11}) \dots f_m(x_{m1}) \\ f_0(x_{02})f_1(x_{12}) \dots f_m(x_{m2}) \\ \dots \\ f_0(x_{0n})f_1(x_{1n}) \dots f_m(x_{mn}) \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

Матрица коэффициентов регрессии $A = (a_0, a_1, a_2, \dots, a_m)$ является решением системы n нормальных уравнений

$$(X'X)A' = X'Y.$$

откуда

$$A' = (X'X)^{-1}X'Y.$$

(Карпелевич, Садовский, 1968).

Критерием оценки полученного уравнения может служить близость расчетных и эмпирических данных:

$$\sum_{k=1}^n (y_k - \hat{y}_k)^2 = \min. \quad (4)$$

Для оценки полученного уравнения можно использовать критерий F Фишера, равный отношению дисперсии y_i относительно среднего арифметического значения \bar{y} к остаточной дисперсии относительно \hat{y} :

$$F = \frac{S_y^2}{S_{\text{ост}}^2}, \quad (5)$$

где

$$S_y^2 = \frac{\sum_{k=1}^n (y_k - \bar{y})^2}{n-1};$$

$$S_{\text{ост}}^2 = \frac{\sum_{k=1}^n (y_k - \hat{y}_k)^2}{n - (m+1)};$$

$m+1$ — число коэффициентов регрессии.

Критерий F можно истолковать как меру практической ценности уравнения регрессии, определяющую, во сколько раз это уравнение описывает данные наблюдения лучше, чем среднее арифметическое.

Если $F > 1,5$, то уравнение повышает точность описания. В противном случае использование уравнения не имеет смысла.

Для оценки избранного уравнения регрессии удобно использовать и третий критерий — показатель средней ошибки аппроксимации:

$$\beta = \frac{1}{n} \sum_{k=1}^n \frac{|y_k - \hat{y}_k|}{y_k} \cdot 100\%. \quad (6)$$

Составление уравнений прогноза в нескольких видах с последующим сравнением их по критериям (4), (5) или (6) дает возможность выбрать уравнение, наилучшим образом описывающее эмпирические данные.

Дать прогноз по полученному уравнению множественной регрессии и измеренным значениям факторов x_i уже не составляет труда.

Располагая данными следующего, $n+1$ -го, наблюдения можно проверить соответствие прогноза фактическим данным и уточнить уравнение регрессии, выполнив все расчеты заново для $n+1$ -ой систем:

$$y_k, x_{1k}, x_{2k}, \dots, x_{mk},$$

где $k=1, 2, \dots, n, n+1$.

Однако многократное выполнение громоздких расчетов регрессионного анализа — чрезвычайно трудоемкая работа.

В Управлении по координации вычислительных работ при Министерстве рыбного хозяйства СССР реализована программа регрессионного анализа ПРА-3 для ЭВМ «Минск-22», предусматривающая большой объем вычислений: определение числовых характеристик всех рассматриваемых случайных величин, всех парных и условных коэффициентов корреляции, коэффициентов множественной корреляции и регрессии, критерия F , теоретических значений \hat{y}_k , погрешностей $\Delta = |y_k - \hat{y}_k|$ и др. (Френкель, 1968; Программы по математической статистике для ЭВМ «Минск-22», 1969).

При наличии представительной количественной информации о факторах, определяющих запасы рыб, использование программы ПРА-3 даст возможность оперативно получать на ЭВМ количественный прогноз сырьевой базы.

Перспективно использование регрессионного анализа для оценки продуктивности водоемов в нерестово-выростных хозяйствах. Однако у рыбководов еще не сложилось единого мнения о совокупности факторов, определяющих продуктивность водоемов. Но главным препятствием для количественной оценки продуктивности оказалось отсутствие в нерестово-выростных хозяйствах достаточно полных и надежных данных об этих факторах. Примерно то же может быть сказано и о большинстве промысловых объектов.

Отсутствие систематизированной и надежной биологической информации о сырьевой базе рыболовства в настоящее время является серьезным препятствием для проведения биологических исследований на количественном уровне, для решения проблем оценки запасов рыб и прогнозирования их уловов.

Создание системы центров биоинформации рыбного хозяйства, охватывающее все бассейны, с тщательно разработанными системами хранения, первичной обработки и оперативной передачи информации, является первоочередной задачей, от решения которой зависит уровень дальнейшего развития рыбохозяйственной науки.

ЛИТЕРАТУРА

- Аксютинна З. М. Об использовании электронных вычислительных машин для анализа распределения и количественной оценки популяции рыб. Труды ВНИРО. Т. LXXI, 1970.
- Дементьева Т. Ф. Эффективность методов оценки запасов применяемых в хозяйственной науке с целью прогнозирования величины возможного вылова. Труды ВНИРО. Т. LXXI, 1970.
- Дунарский О. М., Закурдаев А. Г. Статистический анализ и обработка наблюдений на ЭВМ «Минск-22». М., изд-во «Статистика», 1971.
- Засосов А. В. Теоретические основы рыболовства. М., изд-во «Пищевая промышленность», 1970.
- Карпелевич Ф. И., Садовский Л. Е. Элементы линейной алгебры и линейного программирования. М., ФМ., 1963.
- Программы по математической статистике для ЭВМ «Минск-22». Под ред. Г. Н. Веселовой. М., ОНТИ Гиредмета, 1969.
- Смирнов Н. В., Дунин-Барковский И. В. Курс теории вероятностей и математической статистики. М., изд-во «Наука», 1965.
- Френкель А. А. Математический анализ производительности труда. М., изд-во «Экономика», 1968.

REGRESSION ANALYSIS AS A METHOD OF FORECASTING

Z. M. Aksytina and K. G. Yankovskaya

SUMMARY

Forecasting fishery resources with the help of regression analysis involves the determination (m) of principal factors that govern the state of stocks, carrying out (n) of independent observations on these factors and the related stocks ($n > m$), and the development from the empirical data of a multiple regression equation as an approximation function of the forecast. Forecasting is carried out by calculating the function of stocks from the concrete values of the factors determining it.

L'ANALYSE RÉGRESSIVE COMME UNE MÉTHODE DE LA PRÉVISION

Z. M. Aksytina et K. G. Jankovskaya

RÉSUMÉ

La prévision des ressources de la pêche au moyen de l'analyse régressive se forme du dégagement du nombre des facteurs généraux (m), qui déterminent l'état des ressources, de la réalisation des observations indépendantes (n) de ces facteurs des ressources qui leur correspondent ($n > m$) et de la construction à la base des données empiriques de l'équation de la régression multiple en tant que la fonction approximative de la prévision.

La prévision se fait comme calcul de la fonction des ressources d'après les valeurs concrètes qui déterminent leurs facteurs.

Замеченные опечатки

Стр.	Строка	Напечатано	Следует читать
15	1-я сверху	«Perscu»	«Perscy»
15	2-я снизу	10 juin	10 juillet 1941
29	13-я сверху	палочки «Рачек»	палочки «Рачок»
82	8-я сверху	Дунарский О. М.	Дукарский О. М.
100	подпись	— — — — 10 ⁻⁸ кюри/л,	— — — — 10 ⁻⁸ кюри/л
	под рис. 2	— — — — 10 ⁻⁶ кюри/л	— — — — 10 ⁻⁶ кюри/л
142	26-я сверху	в массе рыбы	в мясе рыбы
158	24-я снизу	inakiri	Inakiri
169	4-я снизу	soutre	soufre