



АНАЛИТИЧЕСКИЕ И ЧИСЛЕННЫЕ МЕТОДЫ ИССЛЕДОВАНИЯ В МАТЕМАТИКЕ

*Теория вероятностей и
математическая статистика,
теория массового обслуживания*

Н. Д. Апарович
(БГУ, Минск)

ГЕОСТАТИСТИЧЕСКИЙ ПОДХОД К АНАЛИЗУ ГИДРОЭКОЛОГИЧЕСКИХ ДАННЫХ

Работа посвящена статистическому анализу временного ряда. Исходные данные представляют собой наблюдения за температурой воды, растворимостью кислорода в воде и насыщенностью воды кислородом.

Осуществлена выборка средних значений растворимости кислорода в воде озера Баторино за август месяц с 1966 по 2012 годы на горизонте 3 м. Обработка и анализ данных выполнены с помощью языка программирования Python.

Вычислены и проанализированы описательные статистики выборки. Статистические тесты Колмогорова-Смирнова и Д'Агостино при уровне значимости $\alpha = 0.05$ не отвергают гипотезу о нормальном распределении с параметрами $N(8.83, 0.4)$ исследуемых данных, а статистический тест Дики-Фуллера показал стационарность ряда.

Построена оценка семивариограммы, и визуальным методом подобраны к ней линейная, сферическая и экспоненциальная модели. Применен метод обычного кригинга для оценки значений ряда в ненаблюдаемых точках. Проведен сравнительный анализ результатов, который показал, что модели, в которых был добавлен эффект саморodka, показали лучший результат, чем модели без него. Сравнительный анализ осуществлялся тремя методами, а именно: сравнение построенных оценок с реальными значениями ряда по выбранным метрикам, анализ дисперсий ошибок построенных прогнозных оценок и

сравнение построенных оценок методом кросс-валидации и реальных значений ряда с помощью выбранных метрик.

Исследование параметра «радиус влияния» модели семивариограммы показало, что с увеличением его значения растет точность прогноза.

К. Г. Дюндя

(ГрГУ им. Я. Купалы, Гродно)

ПРИМЕНЕНИЕ КЛАСТЕРНОГО АНАЛИЗА ДЛЯ КЛАССИФИКАЦИИ МНОГОМЕРНЫХ НАБЛЮДЕНИЙ

Кластерный анализ – процедура, предназначенная для группировки наблюдений или переменных в значимые кластеры на основе сходства между ними. Кластер – это группа объектов, внутригрупповое сходство между которыми максимально, а межгрупповое сходство сведено к минимуму по какому-то объективному критерию [1]. Большинство методов кластерного анализа производит объединение наблюдений в кластеры на основе матрицы расстояний.

В ходе проделанной работы были проанализированы данные по семи странам, отражающие значения таких признаков, как число врачей на 10 тыс. населения, смертность на 100 тыс. населения, ВВП, расходы на медицину. При выполнении работы использовались методы кластерного анализа: иерархический агломеративный метод и метод k -средних, реализованные в пакете STATISTICA. В результате проведенного исследования оптимальной была признана классификация, выделяющая два кластера, представленная в таблице 1.

Таблица 1 – Результаты классификации

Кластер 1	Кластер 2
Россия	Австралия
Беларусь	Австрия
Азербайджан	Бельгия
Армения	

Результаты дисперсионного анализа подтверждают, что классификация является обоснованной по всем признакам, за исключением одного. Наблюдается несущественная разница между кластерами лишь по числу врачей на 10 тыс. населения. Страны первого кластера