

## Колонка редколлегии

### МАТЕМАТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА ИНТЕРВАЛЬНЫХ ДАННЫХ

Перспективная и быстро развивающаяся область математических методов исследования — методы анализа интервальных данных. Результаты измерений (наблюдений, испытаний, анализов, опытов) всегда имеют некоторую неопределенность. Эту неопределенность можно описывать с помощью различных математических средств. Традиционным является вероятностно-статистический подход, согласно которому разность между реальным значением признака и зафиксированным исследователем описывается случайной величиной. Последние десятилетия для описания неопределенности всё чаще используют аппарат теории нечетких множеств. Полезным является и третий подход, при котором неопределенность описывается интервалом. Другими словами, исходные данные — не числа, а интервалы.

Длина интервала может определяться ценой деления шкалы. Например, если при измерении длины цена деления шкалы — 1 мм, то зафиксированный исследователем результат измерения 23 мм соответствует реальному значению от 22,5 до 23,5 мм, т.е. данный результат можно записать как интервал [22,5; 23,5]. Естественную неопределенность будущих значений часто описывают интервалом. Так, прогноз температуры воздуха на следующий момент может быть иметь вид [17; 20] (градусов Цельсия), т.е. от 17 до 20 °С.

Математические методы исследования, предназначенные для анализа интервальных данных, разрабатываются в нашей стране с конца 1970-х годов. Журнал «Заводская лаборатория. Диагностика материалов» всегда уделял заметное внимание этой тематике.

В 1990 г. была проведена дискуссия по анализу интервальных данных (1990/7 — здесь и далее приводим ссылки на год и номер журнала, в котором опубликована соответствующая статья). Первой была статья А. П. Вошинина, А. Ф. Бочкова и Г. Р. Сотирова «Метод анализа данных при интервальной нестатистической ошибке». В дискуссии выступили В. П. Бородюк, Е. З. Демиденко, Э. К. Лецкий, А. И. Орлов, И. Л. Легостаева, В. П. Кузнецов. В дальнейшем дискуссия была продолжена, к ней присоединились В. Г. Горский, Н. В. Скибицкий, П. В. Бронз, Е. А. Гуськова.

Есть два основных подхода к анализу интервальных данных. Александр Павлович Вошинин (1937 – 2008 гг.) и его соавторы описывали погрешности интервалами, не прибегая к использованию вероятностно-статистических составляющих. Был разработан метод анализа данных с интервальными ошибками в задачах проверки гипотез и оценивания параметров неявных и линейно параметризованных функций (2000/3). В соответствии с этим методом рассмотрено построение аналитических моделей по данным вычислительного эксперимента в задачах анализа чувствительности и оценки экономических рисков (2007/1). Интервальный подход оказался полезным для выражения неопределенности измерений и при калибровке цифровых измерительных систем (2007/11). Обсуждалось развитие интервального анализа данных и его перспективы (2002/1).

Другой подход — прикладная математическая статистика интервальных данных — развивается А. И. Орловым и его соавторами в рамках статистики объектов нечисловой природы. Речь идет о развитии методов прикладной математической статистики в ситуации, когда статистические данные — не числа, а интервалы, в частности, порожденные наложением ошибок измерения на значения случайных величин. При этом центр интервала — значение случайной величины, а длина интервала (в базовой модели) задана и равна предельной абсолютной ошибке измерения. Основополагающими являются понятия нотны [максимально возможного (по

абсолютной величине) отклонения значения статистики, вызванного погрешностями наблюдений] и рационального объема выборки. Алгоритмы и выводы статистики интервальных данных принципиально отличаются от классических. В частности, не существует состоятельных оценок, поскольку средний квадрат ошибки всегда больше квадрата нотны. Нецелесообразно увеличивать объем выборки до сверхрационального.

В рамках этого подхода разработаны алгоритмы интервальной линейной парной регрессии (2005/3). Развитие работ по прикладной математической статистике интервальных данных рассматривалось в обзорах по статистике объектов нечисловой природы (1990/3, 1995/3, 1995/5, 2009/5). При изучении устойчивости выводов, получаемых с помощью математических методов и моделей, также естественно исходить из модели интервальных данных (2010/3).

Есть принципиальное отличие между интервальной математикой и статистикой интервальных данных. Интервальная математика предназначена для отслеживания погрешностей при округлении, поэтому, например,  $[a, b] - [a, b] = [a - b, b - a]$ , а не 0, как в обычной арифметике. В то время как в прикладной математической статистике интервальных данных  $[a, b] - [a, b] = 0$ , поскольку центр интервала один и тот же, как и отклонение от него. В нашей стране интервальной математикой занимается научная школа акад. Ю. И. Шокина (З. Х. Юлдашев, С. А. Калмыков, Б. С. Добронец, С. П. Шарый, А. Н. Роголёв и др.).

Анализ интервальных данных — одна из точек роста современной прикладной статистики (1998/3), поэтому это направление подробно отражено в статьях, в которых обсуждались итоги и перспективы математических методов исследования (2002/1), нерешенные вопросы в этой области (2002/3). Одной из основных черт новой парадигмы прикладной статистики является разработка математико-статистических методов анализа интервальных данных (2012/1, часть I).

По нашему мнению, во все виды статистического программного обеспечения должны быть включены алгоритмы интервальной статистики, «параллельные» обычно используемым в настоящее время алгоритмам прикладной математической статистики. Это позволит в явном виде учесть наличие погрешностей результатов наблюдений (измерений, испытаний, анализов, опытов). В одном из ближайших номеров журнала будет помещена обобщающая статья по статистике интервальных данных.

© Член-корр. РАН Д. А. Новиков,  
докт. техн. наук, докт. экон. наук А. И. Орлов