

ВОЗМОЖНОСТИ МЕТОДОВ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ ДЛЯ ОБРАБОТКИ АРХЕОЛОГИЧЕСКОГО МАТЕРИАЛА В ХЕРСОНЕСЕ

В настоящее время весь мир переживает информационный взрыв. Стремительное развитие вычислительной техники, ориентированной при помощи различных программных средств на решение самых разных задач, коснулось всех областей науки, в том числе и археологии. Накопленная сумма знаний о прошлом человеческой цивилизации требует применения как новых носителей информации, так и новых методов ее обработки, при этом необходимо учесть, что количество археологической информации даже по устаревшим оценкам

тельности и позволяющая создать общее представление о тех или иных моделях экономики, демографии, этнографии и т.д., вполне возможно применение статистики для реконструкции различных сфер жизни древнего общества, разумеется, с определенной долей вероятности. Попытки применения математической статистики в археологии были как у нас в стране, так и за рубежом, весьма успешны, особенно зарубежные, поскольку на Западе статистические исследования всех сфер жизни давно уже превратились в целую отрасль и дают достаточно до-

Таблица 1.

Объект	Признак А	Признак В	Признак С	Признак D	Признак Е
Венчик 1	1	1	1	0	0
Венчик 2	0	1	1	1	0
Венчик 3	1	0	1	1	1
—	—	—	—	—	—
Венчик n	1	1	0	0	1

удваивается примерно каждые 8-10 лет¹. Поэтому необходимо применение ЭВМ и мощного программного обеспечения для ее хранения и обработки, так как компьютер — это единственный инструмент, позволяющий быстро обрабатывать большие массивы данных. В то же время, как массовые находки, так и единичные (при сравнении), могут при рассмотрении и анализе с разных позиций привести к самым неожиданным (или наоборот — ожидаемым) результатам. Количество же находок настолько велико, что создать целостное представление об исследуемом объекте в исследуемом временном слое можно лишь на основе статистических данных по ним. Точно так же как для анализа состояния современного общества применяется статистика, показывающая ход и уровень развития различных сфер дея-

тельные ответы на многие вопросы.

Вместе с тем, неизбежно возникает главная проблема — формализация исходных данных, иначе говоря — классификация находок по определенным правилам. Это необходимо потому, что и "ручной", и автоматизированный статистические анализы требуют либо конкретное значение некоторого оцениваемого параметра, либо диапазон (область действия) этого параметра. Разумеется, необходимо прежде четко и недвусмысленно определить, что считать и сопоставлять². Следовательно, основная трудность заключается в том, чтобы создать общие правила обращения с любым археологическим материалом³, но это уже предмет научной дискуссии, поскольку специалист А может придерживаться одних взглядов, специалист Б — других, а специалист В — отличных от первых двух.

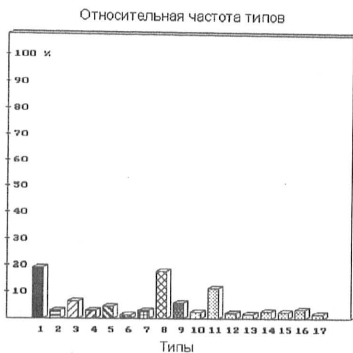


Рис.1.

Одним из примеров классификации является работа "О принципах изучения античных водосборных цистерн"⁴, где на херсонесском материале проведен анализ частоты встречаемости типов керамического материала по строго разработанной системе, а также доклад Н. Коновичи на херсонесской сентябрьской конференции 1997 г, в котором проведен анализ распространения синопских клейменных амфор на западно-понтийском побережье (тоже частотный), на основе статистических данных по временным периодам, и сделаны выводы о состоянии синопского экспорта в эти периоды⁵. В качестве аналогичного примера можно привести задачу⁶, суть которой состоит в том, что некоторый массив объектов, каждый из которых был описан по определенным правилам, необходимо рассортировать по типам и определить относительную частоту встречаемости каждого типа объекта (количество элементов каждого типа по отношению ко всему массиву в процентах). В исходном виде данные представлены в виде таблицы (Табл.1).

В данном случае каждый объект описан по системе из пяти признаков (А, В, С, D и Е), каждый из которых имеет два "состояния" (или разделения на подпризнаки), обозначенных соответственно 1 и 0. Набор из признаков А, В, С, D и Е составлял некоторый тип объекта (вещника). Программа, разработанная автором, выполняла вышеназванные расчеты и представляла результат в

виде гистограммы относительной частоты типов (Рис.1).

Как видно из графика, типы обозначены порядковыми номерами с 1 по 17 по горизонтальной оси, а процентное количество каждого типа - по вертикальной оси. Основной недостаток этой программы состоит в том, что возможности графических режимов при программировании в операционной системе DOS⁷ сильно ограничены, а значит часть информации (в данном случае - типы, каждый из которых составляет менее 1,3% от общего массива данных, причем процент отбраковки прямо пропорционально зависит от количества элементов рассматриваемого массива) отбракована как не оказывающая влияние на характеристику массива данных в целом. Тем не менее, несмотря на то, что на окончательные выводы потеря части информации существенного влияния не оказывает, для полноты информации необходимо иметь выходные результаты по всему массиву данных. Поэтому существенное преимущество имеют профессиональные пакеты для статистических расчетов, работающие под управлением графических операционных систем типа Windows, например пакет STATISTICA американской фирмы StatSoft (Рис.2).

На гистограмме, построенной программой Statistica по тем же самым данным, показано почти вдвое больше типов объектов, хотя все новые типы составляют небольшую часть массива данных, но, как сказано выше,

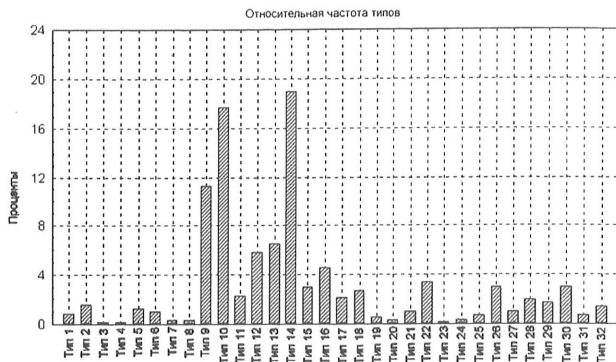


Рис.2.

для полного представления эта информация необходима. Нужно также заметить, что нумерация типов и в первом решении, и во втором является условной, возникающей в процессе "прочесывания" массива элементов; алгоритмы поиска и сортировки элементов массива в авторской программе и программе STATISTICA совершенно разные, и поэтому типы из первого решения и типы из второго не совпадают. В данном случае необходимо для приведения решения 1 к решению 2 сопоставить наборы признаков, определенные для каждого типа. Сам принцип решения задачи применим для работы с любым массивом материалом, будь то керамика, монеты, краниологические данные по серии погребений и т.д. Важно лишь объединять в массив данных объекты, одинаковые по характеру.

Недостатки и первого способа решения задачи, и второго заключаются в следующем: необходимо кодировать входные данные, придерживаясь определенной системы, которая должна быть организована на системе счисления, основание которой равно максимальному количеству подпризнаков оцениваемого признака, поскольку в общем случае их количество явно больше двух. Например для трех подпризнаков признака-троичная (0, 1, 2), четырех - четверичная (0, 1, 2, 3), восьми - восьмеричная (0, 1, 2, 3, 4, 5, 6, 7) и т.д. Полученная таким образом цифровая "сигнатура" типа является его уникальным кодовым обозначением и элемен-

том массива, с которым работает программа (в нашем случае - венчик 1 = 11100, венчик 2 = 01110, венчик 3 = 10111..., венчик n = 11001). Такой способ, приемлемый для поколения ЭВМ 70-80-х годов, не очень подходит уровню конца 90-х, поскольку необходимо четко ориентироваться в составленных цифровых кодах, что неудобно, и преимущества современных компьютеров, операционных систем и программ при этом не используются. Решением проблемы являются СУБД - системы управления базами данных, а также "электронные таблицы", причем с использованием обмена данными между ними. Главным достоинством СУБД является возможность хранения, сортировки, поиска и простых статистических расчетов по массивам данных любого типа - числовых и строковых (текстовых), а главным достоинством "электронных таблиц" - проведение достаточно сложных статистических операций с импортированными из СУБД данными. Правда, последние версии статистических программ также поддерживают возможность импорта данных из СУБД, и возможность проведения любых статистических расчетов с текстовыми строками как с данными. Например, признак 1 (форма ножки амфоры) может иметь дескриптивное (описательное) значение - "рюмкообразная", "копачковая", "конусовидная" и т.д., признак 2 (форма ручки) - "эллипсоидная", "округлая", "с 1 бороздкой", "с 2 бороздками" и т.д. Единственное требование в

Microsoft Access

Этапанная таблица: Русские камни

Код	Материал	Форма	Выпуклость	Сжатость	Сложность	Ширина камык отверстия	Высота камык отверстия
1	горный хрусталь	округлая	продольная	нет	одностороннее	узкий	нет
2	горный хрусталь	шаровидная	нет	поперечная	одностороннее	узкий	нет
3	горный хрусталь	шаровидная	нет	поперечная	одностороннее	широкий	нет
4	горный хрусталь	бочкообразная	продольная	нет	одностороннее	узкий	2 бороздки
5	горный хрусталь	бочкообразная	продольная	нет	двустороннее	широкий	нет
6	кальцит	округлая	продольная	нет	одностороннее	широкий	нет
7	кальцит	округлая	нет	поперечная	одностороннее	узкий	иногда
8	кальцит	округлая	нет	поперечная	одностороннее	широкий	нет

Рис.3.

Microsoft Access

Этапанная таблица: Русские камни

Код	Материал	Форма	Выпуклость	Сжатость	Сложность	Ширина камык отверстия	Высота камык отверстия
1	горный хрусталь	округлая	продольная	нет	одностороннее	узкий	
2	горный хрусталь	шаровидная	нет	поперечная	одностороннее	узкий	
3	горный хрусталь	шаровидная	нет	поперечная	одностороннее	широкий	
4	горный хрусталь	бочкообразная	продольная	нет	одностороннее	узкий	
5	горный хрусталь	бочкообразная	продольная	нет	двустороннее	широкий	
6	кальцит	округлая	продольная	нет	одностороннее	широкий	
7	кальцит	округлая	нет	поперечная	одностороннее	узкий	
8	кальцит	округлая	нет	поперечная	одностороннее	широкий	

Диалоговое окно: Введите значение

Код: 1

Материал: горный хрусталь

Форма: округлая

Выпуклость: продольная

Сжатость: нет

Сложность: одностороннее

Ширина камык отверстия: узкий

Высота камык отверстия: нет

Параметры и название отверстия: нет

Таблица: Русские камни

Рис.4.

этом случае - необходимость придерживаться созданных в процессе разработки системы классификации и терминологии и строго соблюдать их при создании базы данных. Таким образом, не имеет значения для современных программных средств, чем оперировать - числом или словом, что, без сомнения, весьма привлекательно. Например, в СУБД Access фирмы Microsoft таблица данных, оформленная как обычная таблица, может содержать поля, которые являются терминами, следовательно, совершенно понятными любому исследователю, пользующемуся такой базой. Так, можно создать эталонную таблицу данных в соответствии с принятой классификацией, базу данных по фактическому материалу, а затем при помощи встроенного в СУБД языка программирования разработать программный модуль, который будет определять принадлежность любого объекта из набранного фактического материала к тому или иному типу из эталонной таблицы (рис.3).

Результат выполнения запроса на определение типа (рис.4).

Аналогичным образом можно выполнить частотный анализ данных, то есть получить характеристику абсолютного или относительного соотношения типов объектов в общем массиве данных.

Разумеется, применением частотного анализа и проверки на принадлежность не ограничиваются потенциальные возможности математической статистики в археологии. Например, при сравнительном, то есть корреляционном анализе двух единичных объектов одного характера (двух бус, двух амфор, двух черепов и т.д.) можно попытаться определить степень "похожести" этих объектов, и в соответствии с датировкой построить "родословную" объектов, но в этом случае необходимо не только четко определить признаки классификации, но и их "вес" по от-

ношению к общей характеристике объекта в целом (то есть доминантные и рецессивные). Идеальность корреляционного анализа, конечно, должна быть обоснована не только интуитивными соображениями (визуальное восприятие или что-либо подобное), но и внешними факторами (наличие хотя бы одной одинаковой характеристики).

Таким образом, при помощи методов математической статистики можно попытаться реконструировать различные политико-экономические, демографические, этнографические явления и процессы в древнем обществе на некотором временном промежутке с учетом других источников данных (письменных и нумизматических источников, одиночных находок и т.п.). Статистическая обработка археологической информации требует создания огромных баз данных, поскольку количество материала очень велико и растет с каждым годом исследований, и, соответственно, больших материально-технических ресурсов, но ограничиваться созданием только баз данных по массовому археологическому материалу не следует, поскольку остается немало находок, не попадающих под определение "массовый". Поэтому следует создавать базы данных и по этим материалам, опять же для полноты информации об исследуемом объекте и временном периоде и, возможно, сопроводить сравнительный анализ поиском аналогий.

Учитывая, что зарубежные исследователи создали и используют подобные методы в своей работе, и такие методы являются обычной практикой, а контакты и совместные исследования с этими учеными в Херсонесе расширяются с каждым годом, необходимо ликвидировать отставание в этой области, чтобы разговаривать с ними на одном "языке", тем более, что в нашей практике методы математической статистики в археологии до сих пор не нашли широкого применения.

ПРИМЕЧАНИЯ

1. Каменецкий И. С., Маршак Б. И., Шер А. Я. Анализ археологических источников. - М., Наука. - 1975. - С.6
2. Археология и математика. Статистико-комбинаторные методы в археологии. - М., Наука. - 1970. - С. 5
3. Каменецкий...С.5-6
4. Золотарев М.И., Коробков Д.Ю., Ушаков С.В. О принципах изучения античных водосборных цистерн. - Севастополь, 1997.
5. Nikolae Conovichi. La diffusion des amphores Sinopeens timbres sur le littoral quest-pontique. // Херсонес в античном мире. Историко-археологический аспект. Тезисы докладов международной научной конференции. - Севастополь, 1997. - С. 140-146.
6. Автор статьи выражает признательность А. В. Седиковой за любезно предоставленные для решения задачи данные по материалам из раскопок водохранилища в Херсонесе.

7. DOS - Disk Operating System - дисковая операционная система, бывшая основной для компьютеров типа IBM PC и совместимых с ними почти до конца 1996 г. На самом деле ограниченное количество одновременно выводимых на экран элементов при низком разрешении (например, изображение размером 640x480 точек) и качество самой графики не вызывают восхищения при высоком разрешении (при размере изображения 1024x768 точек все элементы пропорционально уменьшаются, поскольку изображение строится по точкам), а применение в программах специальных методов прямого управления графическими аппаратными ресурсами компьютера является очень сложной задачей, посильной лишь системному программисту высокого класса.