

УДК 311

О СХОДИМОСТИ ОСНОВНЫХ КОНЦЕПЦИЙ ТИПОЛОГИИ ДАННЫХ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ИССЛЕДОВАНИЙ

В.В. Глинский, Е.Ю. Чemezova

Новосибирский государственный университет
экономики и управления «НИНХ»
E-mail: s444@ngs.ru

В работе представлена сравнительная характеристика двух основных концепций типологии статистических данных; обсуждаются сходства и различия, достоинства и недостатки дедуктивного и индуктивного подходов к типологии; формулируются условия и возможности их совместного применения в решении прикладных задач классификации статистической информации.

Ключевые слова: совокупностная и вариационная концепции данных, типология, портфельный анализ, группировка.

ON CONVERGENCE OF MAIN CONCEPTS OF TYPOLOGY OF SOCIAL-ECONOMIC STUDIES DATA

V.V. Glinskiy, E.Yu. Chemezova

Novosibirsk State University of Economics and Management
E-mail: s444@ngs.ru

The paper presents comparative characteristics of two main concepts of statistical data typology; similarities and differences, advantages and disadvantages of deductive and inductive approaches to typology are discussed; conditions and possibilities of their combined use in solving of application tasks of statistical information classification are formulated.

Key words: aggregate and variation concept of data, typology, portfolio analysis, grouping.

Задача выделения однородных статистических множеств решается достаточно давно, с середины XIX столетия, причем в рамках двух концепций – вариационной (индуктивной) и совокупностной (дедуктивной). У истоков первой стоял А. Кетле (1796–1876): *необходимость группировки в рамках этой концепции вызвана наличием различий индивидуальных значений признака, группировка позволяет отобрать такие единицы совокупности, вариация которых находится в определенных пределах (интервалах)*. Ее последующее развитие принадлежит в основном западным ученым, формулировка второй концепции выполнена русским статистиком Д.П. Журавским (1851). Он был первым, кто обозначил проблему [21]: *статистическая совокупность не является однородной, необходимо все-сторонне исследовать объект наблюдения, систематизировать явления одного вида (исчисление по категориям)*.

Существенный вклад в развитие дедуктивного подхода, который воплотился в методе типологической группировки, внесли представители российской академической науки: А.И. Чупров, Ю.Э. Янсон, Н.А. Каблу-

ков, А.А. Кауфман, А.А. Чупров, Р.М. Орженцкий, земские статистики А.П. Шликевич, Ф.А. Щербина, С.П. Прокопович, марксисты И.А. Гурвич, В.И. Ленин, А.И. Хряцова, В.Г. Громан, Г.И. Баскин. В дальнейшем работы в этой области были выполнены М.Н. Смит, В.С. Немчиновым. Заметный вклад в развитие теории группировок внесли Г.С. Кильдишев, Ю.М. Аболенцев, Б.Г. Миркин, О.В. Иванов. Значительное место в становлении, развитии и продвижении метода типологической группировки принадлежит ленинградской (санкт-петербургской) научной школе: теоретические исследования Л.В. Некраша, Б.Г. Плошко, И.И. Елисеевой, работы Э.К. Васильевой, С.В. Курышевой, М.А. Клупта, О.Г. Грачевой-Дивеевой, И.П. Суслова и других. История зарождения и развития основных подходов к типологии данных достаточно подробно рассмотрена О.В. Ивановым [22].

Хотелось бы отметить тем не менее, что ряд аспектов прикладного использования основных концепций типологии данных, таких как условия и ограничения, возможности одновременного использования, сходства и различия в подходах в статистических публикациях, практически не рассматривались. Данная работа является попыткой обобщения результатов авторских исследований в решении обозначенных задач [4–19, 25].

В дальнейшем своем генезисе рассматриваемые концепции вышли за пределы группировки, особенно это касается вариационной концепции, хотя содержание их не претерпело принципиальных изменений. В современной трактовке, в рамках статистической теории содержание совокупностной концепции можно определить как разбиение общей совокупности на частные (однородные типы) на основе предварительного теоретического качественного анализа, в свою очередь, вариационный подход к определению однородности статистических данных можно детерминировать как объединение единиц совокупности в однородные группы на основе сходства и различия, причем внутригрупповые различия должны быть менее существенны, чем различия между группами. В совокупностной концепции реализуется дедуктивный подход к типологии данных (от общего к частному), в основе вариационной лежит индуктивный подход (от частного к общему). Первая концепция трактует однородность как качественную (гомогенность единиц частной совокупности детерминируется единым законом развития), в вариационной концепции однородность рассматривается с позиций близости индивидуальных значений признака, упор сделан на количественный аспект.

Совокупностная концепция в прикладных исследованиях реализуется посредством следующих способов и приемов обработки данных: типологическая группировка; портфельный анализ; периодизация (в обработке временных рядов); динамическая классификация реальных совокупностей; временной принцип формирования систем обобщающих показателей, в свою очередь, основной инструментарий вариационной концепции – классический факторный анализ; метод главных компонент; дискриминантный анализ; кластерный анализ; распознавание образов.

В таблице представлена характеристика концепций по основным задачам группировки данных.

Сравнительный анализ сформировавшихся подходов свидетельствует о достаточно серьезных различиях в их содержании. Однако несмотря на это

Сравнительная оценка совокупностной и вариационной концепций выделения однородных данных

Признаки сравнения	Совокупностная концепция	Вариационная концепция
1. Необходимость предварительного исследования совокупности	Обязательный шаг: постановка цели, задач, теоретический качественный анализ позволяют наметить возможные типы в общей совокупности	Возможное условие, но не необходимое
2. Число типов	Задается число теоретически возможных типов, фактическое число типов – один из решающих результатов исследования	Три ситуации: – число типов априори задано; – число типов неизвестно; – не ставится задача определения числа типов
3. Пустые типы	Допустимы, более того наличие пустого типа на заключительной стадии исследования является одним из важнейших результатов	Отсутствуют
4. Число группировочных признаков	2–3 наиболее существенных признака	Не ограничено, признаки как существенные, так и несущественные
5. Специализация признаков	Существует возможность менять исходный набор признаков, описывающих тип	Отсутствует
6. Уровень формализации	Все основные этапы не формализованы, за редким исключением	Все принципиальные решения в ходе применения алгоритмов не формализованы, выполняются на основе «здравого смысла»
7. Элементы искусства (неформальных соображений)	Присутствуют при выборе группировочных признаков, специализации признаков, определении величины интервалов, специализации интервалов	Присутствуют при отборе признаков, стандартизации, выборе мер сходства, алгоритма, пороговых значений, оценки качества разбиения
8. Число алгоритмов	Невелико (типологическая группировка, портфолио-анализ, историческая и параллельная периодизация)	Значительное (кластерный анализ – иерархический и неиерархический, агломеративные и дивизимные процедуры, дискриминантный анализ, распознавание образов, факторный анализ, главные компоненты и т.д.)
9. Сходимость результатов	Практика	Не доказывается, либо доказательства в пределах той же концепции
10. Приложения	Прикладные (управление) и учебные задачи, исследовательские в меньшей степени	Учебные и исследовательские, прикладные – техника, внутренние бизнес-процессы, медицина...
11. В терминах классификации	Методы ручной классификации	Автоматическая классификация
12. Взаимосвязь и взаимопроникновение (возможность)	На этапе теоретического анализа совокупности	Число типов – как порог автоматической классификации

они имеют много общего. Исследуемые концепции решают одну задачу, в значительной степени пересекаются по результатам; не поддаются формализации их основные условия и ограничения (качественная и количественная однородность не заданы априори, ограничиваются целями и задачами исследования, конкретными особенностями изучаемых совокупностей); большинство алгоритмов, реализующих концепции, содержат значительный элемент искусства (являются эвристическими).

Чтобы понять, на каком подходе стоит остановиться, необходимо обратить внимание на значение вариации признаков изучаемой совокупности.

Наличие или отсутствие вариации не должно быть условием, ограничивающим решение задачи однородности. Различные условия развития отдельных единиц совокупности вполне могут привести к одним и тем же значениям изучаемого признака.

Возможны ситуации естественного разбиения общей совокупности на частные, в которых общая вариация признака полностью уходит в группы.

Ограничения. Методы автоматической классификации, в основе которых, как известно, лежит вариационная концепция, имеют более узкую сферу применения по сравнению с методами, основанными на совокупностной концепции, в частности: не допускают наличие пустых типов, в то время как в реальных исследованиях именно отсутствие теоретически возможного типа порой является лучшим из возможных результатов типологии; не дают устойчивых результатов в случае, когда приходится иметь дело с нечеткими множествами; в рамках вариационного подхода не может быть решена задача таксономии совокупности, распределенной по равномерному закону вероятностей.

Наконец, еще один момент, на который в свое время обратили внимание И.И. Елисеева, О.В. Рукавишников: «...создатели этих методов убеждены в том, что возможно точное и однозначное решение задачи классификации, убеждение, корни которого лежат в методологии точных наук...» [20, с. 45].

Прошло достаточно времени, но математический «романтизм» остался, и порой он демонстрирует свои позиции в довольно агрессивной форме. Так, например, А.И. Орлов [24] выделяет три группы методов, одной из задач которых является выделение однородных типов: дискриминантный анализ, кластерный анализ, группировка. Задачей кластер-анализа является выяснение по эмпирическим данным, насколько элементы группируются или распадаются на изолированные скопления. «Иными словами, задача – выявление естественного разбиения на классы, свободного от субъективизма исследователя, а цель – выделение групп однородных объектов, сходных между собой, при резком отличии этих групп друг от друга. «При группировке, наоборот, мы хотим разбить элементы на группы независимо от того, естественны ли границы разбиения или нет» (цит. по: [23]). Цель по-прежнему состоит в выявлении групп однородных объектов, сходных между собой, однако соседние группы могут не иметь резких отличий. Границы между группами условны, не являются естественными, зависят от субъективизма исследователя». И далее: «Пусть, к примеру, необходимо деревья, растущие в определенной местности, разбить на группы, находящиеся рядом друг с другом. Ясна интуитивная разница между несколькими отдельными рощами, далеко отстоящими друг от друга и разделенными

ми полями, и сплошным лесом, разбитым просеками на квадраты с целью лесоустройства. Однако формально определить эту разницу... сложно... Переформулируем сказанное в терминах «кластер-анализа» и «методов группировки». Выделенные с помощью первого подхода кластеры реальные, а потому могут рассматриваться как кандидаты в «естественные». Группировка дает «искусственные» классы, которые не могут быть «естественными» [24, с. 250].

В некотором смысле данная позиция, позволяет закрепить наш выбор:

1. Потому и сложно, что формально, уж лучше на основании опыта или интуитивно.

2. «Пусть, к примеру» – такая постановка задачи (естественная для математиков) к «естественному» разбиению совокупности может привести лишь случайно.

3. Если всерьез рассматривать в качестве цели исследования – близость деревьев в определенной местности, то типологическая группировка даст те же результаты.

4. Если цель – лесоустройство, тогда не исключено, что некоторые рощи придется объединять в одну группу, и это опять будет разбиением «естественным», далее с позиций лесоустройства «естественно» разбить лес на квадраты, получив в итоге частные совокупности – «делянки», с позиций изучения леса «естественно» выделить во втором случае – «опушку леса» и «остальной лес» и т.д. и т.п.

Типологическая группировка дает более сильные решения в части получения однородных данных, так как ищет эти решения не формально, не «пусть, к примеру», а путем предварительного теоретического анализа, путем постановки цели, наметки типов, которые могут теоретически быть в исследуемой совокупности, наконец, главное, путем включения опыта, интуиции, знаний специалистов отрасли, для которой эта задача решается.

Резюмируя, отметим, что проведенное сопоставление показывает возможность интегрировать сильные стороны рассмотренных подходов при решении конкретной задачи типологии: типологическая группировка, имеющая более широкую сферу применения, акцентированная на качественную однородность может рассматриваться как основа предлагаемой методики классификации; применение портфельного анализа обеспечивает наглядность представления результатов, использование методов автоматической классификации дает максимальный охват признакового пространства.

Литература

1. Айвазян С.А., Бежаева З.И., Староверов О.В. Классификация многомерных наблюдений. М.: Статистика, 1974. 240 с.
2. Андерсон Т. Введение в многомерный статистический анализ. М.: Физматгиз, 1963. 520 с.
3. Браверман Э.М., Мучник И.Б. Структурные методы обработки эмпирических данных. М.: Наука, 1983. 464 с.
4. Глинский В.В. Статистические методы поддержки управленческих решений: монография. Новосибирск: Изд-во НГУЭУ, 2008. 256 с.
5. Глинский В.В. Как измерить малый бизнес? // Вопросы статистики. 2008. № 7. С. 73–75.

6. Глинский В.В. Мифическая статистика малого бизнеса. Проблемы статистического изучения турбулентных совокупностей // ЭКО (экономика и организация промышленного производства). 2008. № 9. С. 51–62.
7. Глинский В.В., Серга Л.К. Нестабильные совокупности: концептуальные основы методологии статистического исследования // Вестник НГУЭУ. 2009. № 2. С. 137–142.
8. Глинский В.В., Серга Л.К. О государственном регулировании малого предпринимательства в России // Национальные интересы: приоритеты и безопасность. 2011. № 19. С. 2–8.
9. Глинский В.В. Периодизация демографической ситуации. Перспективы национального проекта по улучшению демографической ситуации // Вестник Самар. гос. экон. ун-та. 2008. № 10(48). С. 30–37.
10. Глинский В.В. Портфельный анализ: сущность и применение // Финансы и бизнес. 2008. № 4. С. 105–109.
11. Глинский В.В. Портфельный анализ в статистическом исследовании клиентов предприятия // Финансы и бизнес. 2009. № 1. С. 91–95.
12. Глинский В.В. О применении периодизации в портфельном анализе // Вестник Самар. гос. экон. ун-та. 2008. № 11(49). С. 14–17.
13. Глинский В.В. Типология экономического развития современной России на основе методов периодизации макроэкономических процессов // Вестник Том. гос. ун-та. 2009. № 1 (318). С. 160–166.
14. Глинский В.В. О применении портфельного анализа в оценке клиентской составляющей // Экономические и гуманитарные науки. 2009. № 1(207). С. 112–115.
15. Глинский В.В., Серга Л.К. Статистика XXI века. Вектор развития // Вестник НГУЭУ. 2011. № 1. С. 108–118.
16. Глинский В.В., Овечкина Н.И. Тенденции демографической политики: сдержанный оптимизм третьего цикла // Национальные интересы: приоритеты и безопасность. 2010. № 16. С. 14–18.
17. Глинский В.В., Макаридина Е.В. О модели жизненного цикла высшего профессионального образования России // Национальные интересы: приоритеты и безопасность. 2011. № 3. С. 12–18.
18. Глинский В.В., Гусев Ю.В., Золотаренко С.Г., Серга Л.К. Портфельный анализ в типологии данных: методология и применения в поддержке управленческих решений // Вестник НГУЭУ. 2012. № 1. С. 25–54.
19. Глинский В.В., Третьякова О.В., Скрипкина Т.Б. О типологии регионов России по уровню эффективности здравоохранения // Вопросы статистики. 2013. № 1. С. 57–69.
20. Елисеева И.И., Рукавишников В.О. Логика прикладного статистического анализа. М.: Финансы и статистика, 1982. 192 с.
21. Журавский Д.П. Об источниках и употреблении статистических сведений. М.: Госстатиздат, 1946. 120 с.
22. Иванов О.В. Развитие концепции статистической группировки // Теория статистической группировки. Ч. I. М.: Моск. экон.-стат. ин-т, 1990. 76 с.
23. Кендалл М.Д., Стьюарт А. Статистические выводы и связи. М.: Наука, 1973. 899 с.
24. Орлов А.И. Нечисловая статистика. М.: МЗ – Пресс, 2004. 516 с.
25. Чemezova Е.Ю. Типология субъектов РФ по уровню социально-экономического развития // Вестник НГУЭУ. 2010. № 1. С. 171–176.

Bibliography

1. Ajvazjan S.A., Bezhaeva Z.I., Staroverov O.V. Klassifikacija mnogomernyh nabldenij. M.: Statistika, 1974. 240 p.
2. Anderson T. Vvedenie v mnogomernyj statisticheskij analiz. M.: Fizmatgiz, 1963. 520 p.

3. *Braverman Je.M., Muchnik I.B.* Strukturnye metody obrabotki jempiricheskikh dan-nyh. M.: Nauka, 1983. 464 p.
4. *Glinskij V.V.* Statisticheskie metody podderzhki upravlencheskih reshenij: monogra-fija. Novosibirsk: Izd-vo NGUJeU, 2008. 256 p.
5. *Glinskij V.V.* Kak izmerit' malyj biznes? // Voprosy statistiki. 2008. № 7. P. 73–75.
6. *Glinskij V.V.* Mificheskaja statistika malogo biznesa. Problemy statisticheskogo izu-chenija turbulentnyh sovokupnostej // JeKO (jekonomika i organizacija promysh-lennogo proizvodstva). 2008. № 9. P. 51–62.
7. *Glinskij V.V., Serga L.K.* Nestabil'nye sovokupnosti: konceptual'nye osnovy metodo-logii statisticheskogo issledovanija // Vestnik NGUJeU. 2009. № 2. P. 137–142.
8. *Glinskij V.V., Serga L.K.* O gosudarstvennom regulirovanii malogo predprinimatel'stva v Rossii // Nacional'nye interesy: priority i bezopasnost'. 2011. № 19. P. 2–8.
9. *Glinskij V.V.* Periodizacija demograficheskoy situacii. Perspektivy nacional'nogo proekta po uluchsheniju demograficheskoy situacii // Vestnik Samar. gos. jekon. un-ta. 2008. № 10(48). P. 30–37.
10. *Glinskij V.V.* Portfel'nyj analiz: sushhnost' i primenenie // Finansy i biznes. 2008. № 4. P. 105–109.
11. *Glinskij V.V.* Portfel'nyj analiz v statisticheskom issledovanii klientov predpriyatija // Finansy i biznes. 2009. № 1. P. 91–95.
12. *Glinskij V.V.* O primenenii periodizacii v portfel'nom analize // Vestnik Samar. gos. jekon. un-ta. 2008. № 11(49). P. 14–17.
13. *Glinskij V.V.* Tipologija jekonomicheskogo razvitija sovremennoj Rossii na osnove metodov periodizacii makrojekonomicheskikh processov // Vestnik Tom. gos. un-ta. 2009. № 1 (318). P. 160–166.
14. *Glinskij V.V.* O primenenii portfel'nogo analiza v ocenke klientskoj sostavljajushhej // Jekonomicheskie i gumanitarnye nauki. 2009. № 1(207). P. 112–115.
15. *Glinskij V.V., Serga L.K.* Statistika XXI veka. Vektor razvitija // Vestnik NGUJeU. 2011. № 1. P. 108–118.
16. *Glinskij V.V., Ovechkina N.I.* Tendencii demograficheskoy politiki: sderzhannyj opti-mizm tret'ego cikla // Nacional'nye interesy: priority i bezopasnost'. 2010. № 16. P. 14–18.
17. *Glinskij V.V., Makaridina E.V.* O modeli zhiznennogo cikla vysshego professional'no-go obrazovanija Rossii // Nacional'nye interesy: priority i bezopasnost'. 2011. № 3. P. 12–18.
18. *Glinskij V.V., Gusev Ju.V., Zolotareno S.G., Serga L.K.* Portfel'nyj analiz v tipologii dannyh: metodologija i primenenija v podderzhke upravlencheskih reshenij // Vestnik NGUJeU. 2012. № 1. P. 25–54.
19. *Glinskij V.V., Treťjakova O.V., Skripkina T.B.* O tipologii regionov Rossii po urovnju jeffektivnosti zdavoohranenija // Voprosy statistiki. 2013. № 1. P. 57–69.
20. *Eliseeva I.I., Rukavishnikov V.O.* Logika prikladnogo statisticheskogo analiza. M.: Finansy i statistika, 1982. 192 p.
21. *Zhuravskij D.P.* Ob istochnikah i upotreblenii statisticheskikh svedenij. M.: Gosstat-izdat, 1946. 120 p.
22. *Ivanov O.V.* Razvitie koncepcii statisticheskoy gruppirovki // Teorija statisticheskoy gruppirovki. Ch.I. M.: Mosk. jekon.-stat. in-t, 1990. 76 p.
23. *Kendall M.D., St'juart A.* Statisticheskie vyvody i svjazi. M.: Nauka, 1973. 899 p.
24. *Orlov A.I.* Nechislovaja statistika. M.: MZ – Press, 2004. 516 p.
25. *Chemezova E.Ju.* Tipologija sub#ektov RF po urovnju social'no-jekonomicheskogo razvitija // Vestnik NGUJeU. 2010. № 1. P. 171–176.