
СТАТИСТИКА И ЭКОНОМИЧЕСКОЕ ИЗМЕРЕНИЕ

УДК 311.12

О ПОДХОДАХ К РЕШЕНИЮ ЗАДАЧИ ИДЕНТИФИКАЦИИ НЕЧЕТКИХ СОВОКУПНОСТЕЙ

Л.К. Серга

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

В статье предпринята попытка обобщения результатов авторских исследований нестабильных совокупностей; выполнена постановка задачи; проведены оценки, уточняющие пороговые значения на реальных данных для решения задачи идентификации нечетких совокупностей; в качестве информационной базы исследования использованы официальные данные государственной статистики.

Ключевые слова: нестабильные совокупности, классификация, типология, нечеткие, пороговые совокупности, функция принадлежности, распределения вероятностей.

ON APPROACHES TO SOLUTION OF THE PROBLEM OF FUZZY AGGREGATIONS

L.K. Serga

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

The article takes a shot at generalization of results of authors researches of unstable aggregations; statement of the problem was performed; assessments, refining threshold values on real data for the solution of the problem of identification of fuzzy aggregations, were made; official data of government statistics was used as data base of the research.

Key words: unstable aggregations, classification, typology, fuzzy threshold aggregations, membership function, probability distribution function.

Нестабильность становится необходимым звеном эволюции, что выражается в разработке различных эвристических моделей в тех или иных сферах человеческой деятельности. Нестабильность как феномен становится объектом исследования не только естествознания, философии, но и гуманитарных и экономических наук, давая импульс распространению концепций синергетики как общенаучной парадигмы.

Так как статистическая информация является отражением общественного бытия, перед статистической наукой также встают проблемы исследования нестабильности, т.е. проблемы учета воздействия нестабильности и неопределенности на социально-экономическое развитие общества.

Вопросами изучения нестабильности статистика занимается достаточно давно. Нестабильность динамики исследуется методами периодизации, анализа цикличности и сезонности, изучением случайной составляющей. Нестабильность пространства (территориальных образований) изучается статистикой в рамках анализа дифференциации развития территорий. Одновременно вопросы статистического исследования нестабильности совокупности (изменчивость, неопределенность состава, непостоянство структуры) до недавнего времени оставались вне пристального внимания специалистов.

Изменение экономических форм хозяйствования, существенные структурные экономические и социальные изменения общества и вызванные ими нестабильность делают достоверность данных особой проблемой статистики. Это вызывает необходимость изменения методов сбора и обобщения статистических данных, что в свою очередь требует развития основ теории статистического исследования, в частности – теории совокупностей [5, 6]. В теории совокупностей используются различные классификации, которые в условиях возрастающей нестабильности не могут в полной степени отвечать современным требованиям и решать весь спектр поставленных перед статистикой задач. Это определило необходимость исследования дополнительных классификационных разрезов реальных совокупностей, таких как динамический и пороговый [3, с. 139; 4, с. 57]. Их введение значительно расширяет аналитические возможности статистики.

Пороговый классификационный разрез учитывает четкость и точность определения границ совокупности или отдельных типов в исследовании сложных явлений. В этой связи выделяют четкие, пороговые и нечеткие совокупности [3, 5, 7, 10].

С пороговым классификационным разрезом совокупности тесно связана типология данных. Другими словами, идентификация четкой, пороговой или нечеткой совокупности фактически равнозначна типологии данных.

Система показателей, характеризующих признаки для выделения типов, может включать показатели всего спектра существующих шкал измерения: номинальные, порядковые, интервальные и относительные, от которых зависит выделяемый тип совокупности по пороговому разрезу (см. таблицу).

При проведении классификации социально-экономических явлений в качестве типобразующих используют обычно несколько информативных признаков в силу того, что система показателей в большинстве случаев точнее охарактеризует явление, чем единственная характеристика. Однако это в свою очередь приводит к нечеткости изучаемого множества.

В современное время в управлении обществом, социально-экономическими проектами приходится все больше иметь дело с нечеткими совокупностями, идентификация которых часто осуществляется непосредственно лицами, принимающими решения. Примером таких совокупностей могут служить совокупность эффективных вузов, инвестиционно привлекательных регионов, надежных заемщиков, средний класс.

Зависимость типа формируемой совокупности по пороговой классификации от количества и шкалы измерения признаков

Количество признаков	Шкала измерения признака	Уровень стабильности признака	Тип совокупности	Примечание
Один	Интервальная	×	Нечеткая, пороговая	Естественное или искусственное разбиение с установлением порога
	Отношений	×		
	Порядковая	×	Четкая	
	Номинальная	Устоявшийся	Четкая	
		Вновь введенный	Нечеткая	Применение функции принадлежности
Два и более	Интервальная и отношений	×	Нечеткая, пороговая	Расчет интегрального показателя, затем естественное или искусственное разбиение с установлением порога
	Номинальные и порядковые	Устоявшийся	Множество четких типов	
		Устоявшиеся и вновь введенные	Нечеткая	Применение функции принадлежности, методов кластеризации либо расчет интегрального показателя, затем естественное или искусственное разбиение с установлением порога
	Номинальная, порядковая, интервальная, отношений	Устоявшиеся и вновь введенные	Нечеткая	Применение функции принадлежности, методов кластеризации либо расчет интегрального показателя, затем естественное или искусственное разбиение с установлением порога

Идентификация и оценка нечетких совокупностей может осуществляться тремя способами:

- 1) с помощью методов оптимального правдоподобия (теории нечетких множеств и построения функции принадлежности);
- 2) методами автоматической классификации (кластеризации);
- 3) переводом нечеткой совокупности в категорию пороговых на основе установления порогов.

Возможность применения метода оптимального правдоподобия (в частности построения функции принадлежности) была апробирована при решении задачи идентификации среднего класса в России и Новосибирской области.

Средний класс – это совокупность домохозяйств, обладающих схожими признаками и объединяемые на основании объективных и субъективных критериев принадлежности.

Объективные критерии: уровень дохода; уровень образования; владение материальной собственностью. Субъективный критерий – самоидентификация. Субъективная идентификация со средним классом характеризуется достаточно высокой оценкой людьми своего собственного социального положения и относительной удовлетворенностью статусом [8].

Для каждого критерия выделяют границы (цензы) вхождения в средний класс. Имея цензы по всем критериям, достаточно просто построить по ним комбинационную группировку и по данным конкретного обследования определить группу людей, формирующих средний класс. В условиях турбулентности и нестабильности экономической ситуации значения критериев постоянно меняются, что в значительной степени влияет на изменение границ совокупности среднего класса. Это дает основание отнести средний класс к нечетким совокупностям и применять к ним методологию исследования подобных объектов [5, 10].

Как было отмечено выше, одним из основных критериев выделения среднего класса выступает критерий уровня доходов. Существующий, т.е. общеустановленный ценз (от 2 до 6 прожиточных минимумов) прост в расчетах, однако недостоверно определяет границу между бедными и средним классом, ввиду дискуссионности вопроса определения как размера прожиточного минимума, так и величины верхней границы доходов среднего класса. Кроме того, в условиях резкой региональной дифференциации населения по уровню жизни у населения разных регионов существуют свои представления о критериях отнесения к среднему классу.

Поскольку фактический размер дохода, являющийся границей между богатыми и средним классом, бедными и средним классом не известен и варьирует от человека к человеку, то необходимо установить некий порог в соответствии с мнением самого населения. Для установления фактических границ среднего класса по уровню доходов предлагаем использовать теорию нечетких множеств — методы анализа неопределенных данных, в которых описание неопределенностей реальных явлений и процессов проводится с помощью понятия о множествах, не имеющих четких границ. Совместимость результирующего значения (уровня дохода) с нечетким ограничением (мнением населения) символизирует функция принадлежности, которая каждому значению базовой переменной (уровню дохода) ставит в соответствие число из интервала $[0, 1]$.

С целью формирования оптимальных порогов нижнего и верхнего уровня дохода среднего класса был проведен опрос среди различных слоев населения Новосибирской области. Конкретные лица вкладывают свое содержание в понятие «средний класс», поэтому опросу подверглись работники бюджетной и коммерческой сферы, топ-менеджеры, обслуживающий персонал и представители иных категорий граждан (пенсионеры, домохозяйки). Респонденты отвечали на 2 дополняющих друг друга вопроса:

при каком месячном доходе (на одного члена семьи) Вы отнесли бы себя к среднему классу (руб.);

при каком месячном доходе (на одного члена семьи) Вы отнесли бы себя к категории богатых людей (руб.).

Для построения функции принадлежности (в основе проведенного анализа положена методика оценки «богатого человека», предложенная А.И. Орловым [8]) результаты опроса были сгруппированы по интервалам (во всех интервалах левая граница исключена, а правая включена):

0 – до 10 тыс. руб. в месяц на человека (включительно);

1 – от 10 до 20 тыс. руб.;

- 2 – от 20 до 50 тыс. руб.;
- 3 – от 50 до 100 тыс. руб.;
- 4 – от 100 до 500 тыс. руб.;
- 5 – от 500 до 1000 тыс. руб.;
- 6 – более 1000 тыс. руб.

Нечеткие множества, выражающие понятие «средний класс» и «богатый человек» в терминах его ежемесячного дохода, являются подмножеством множества из 7 интервалов, заданных выше, или множества из 7 условных номеров $\{0, 1, 2, \dots, 6\}$.

Для всех групп респондентов были рассчитаны сводные характеристики – выборочная средняя арифметическая, медиана, мода. Разброс ответов на первый вопрос – от 11 500 до 225 000 руб. в месяц на человека. Результаты опроса показывают, что у работников коммерческой сферы критерий дохода, определяющий нижнюю границу среднего класса, в несколько раз выше, чем у других участников опроса.

Построение функций принадлежности респондентов к нечетким совокупностям «средний класс» и «богатый человек» в соответствии с представлениями опрошенных, позволит определить нижнюю и верхнюю границы доходов для интересующей нас категории среднего класса.

Для этого составляется таблица с кумулятивным распределением ответов на поставленные вопросы и составляется функция принадлежности (рис. 1).

Построенные функции принадлежности позволяют определить нижнюю и верхнюю границы среднего класса по уровню доходов.

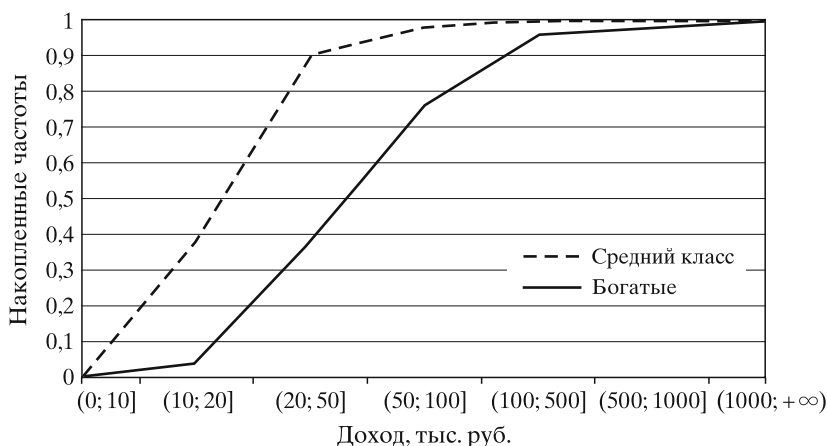


Рис. 1. Функции принадлежности к нечетким совокупностям «Средний класс» и «Богатые»

Как следует из рис. 1, нижняя граница начинается в точке с интервалом $(20; 50]$ тыс. руб., верхняя граница – в точке с интервалом $(100; 500]$ тыс. руб.

Таким образом, согласно мнению респондентов различных социальных групп, к среднему классу будет относиться население Новосибирской области с уровнем дохода от 20 000 руб. (нижняя граница у нижней группы) до 500 000 руб. (верхняя граница у верхней группы).

Для проверки полученных результатов воспользуемся существующим критерием уровня доходов для выделения среднего класса – от 2 до 6 прожиточных минимумов, т.е. представим средний класс по вышеуказанному цензу как пороговую совокупность.

В установлении единиц, обладающих единым качеством, и заключается оценка размеров пороговой совокупности. Процесс ограничения пороговой совокупности строится на статистических методах типологии данных.

Для пороговых совокупностей характерно определение границ изучаемой совокупности и отнесение к ней единиц на основе установленных критериев.

Основной из них – эволюционный критерий, способствующий развитию объекта через формирование неоднородности и, следовательно, нестабильности, созданию дополнительного импульса движения, нового витка развития. Однако установление пороговых значений имеет, с одной стороны, позитивную направленность, с другой – приводит к неоднородности совокупности, проблеме сравнимости и сопоставимости, к проблеме учета и оценки. Решение данной проблемы нам видится в установлении объективных научно обоснованных критериев и соответствующих им границ пороговой совокупности. Общая схема статистического исследования пороговых совокупностей представлена на рис. 2.

Одним из подходов к изучению пороговых совокупностей является использование методов разделения смесей вероятностных распределений – основы теории распознавания образов.



Рис. 2. Общая схема статистического исследования пороговых совокупностей

Если исходную реальную совокупность рассматривать как конечную смесь вероятностных распределений, то задачу определения оптимальных порогов-границ в типологии единиц совокупности можно решить методами разбиения реальной совокупности на частные подсовокупности (типы) со своими распределениями. В результате разделения смеси на составляющие ее компоненты получаем объективный порог. Пороговое значение будет соответствовать величине явления на пересечении кривых вероятностных распределений.

Смеси вероятностных распределений играют важную роль в статистике и теории вероятностей. Используя нормальное приближение, исследователь недооценивает большие значения результирующего распределения, считая, что их вероятности малы. Однако на практике, как правило, результирующее распределение имеет более тяжелые «хвосты» и острую вершину, по сравнению с нормальным распределением. Поэтому в сферах, где важна высокая точность, например, теории надежности и страховании, используют модели конечных смесей распределения [1, 2].

Конечной смесью распределения называют распределение вероятностей, представляющее собой линейную функцию некоторого числа компонент распределений вероятностей [12, с. 305]. Такие распределения используются для моделирования совокупностей, которые предположительно содержат отдельные группы наблюдений или отдельные группы (типы) единиц совокупности. Первоначальным примером применения такого распределения является распределение, смешанное из двух нормальных и примененное Пирсоном в 1894 г.:

$$f(x) = pN(\bar{x}_1, \sigma_1) + (1 - p)N(\bar{x}_2, \sigma_2),$$

где p – доля первой группы в совокупности;

$\bar{x}_1, \sigma_1$ – соответственно средняя и среднеквадратическое отклонение переменной в первой группе;

$\bar{x}_2, \sigma_2$ – соответствующие значения во второй группе.

Рассмотрим практическое применение методов разделения смесей вероятностных распределений для установления порога бедности на основе объективной величины прожиточного минимума (ПМ) и соответственно ценза по доходам для среднего класса [11, с. 233–259].

Объектом исследования послужила совокупность домашних хозяйств Российской Федерации (РФ). Предметом исследования выступило распределение населения РФ по размеру среднедушевого денежного дохода [9].

Предварительно была выполнена проверка гипотезы о принадлежности предмета исследования к смеси распределений и пороговым совокупностям.

Вышеуказанная совокупность является пороговой, поскольку численность бедного населения определяется с помощью величины ПМ. Прожиточный минимум, установленный государством, является своеобразным «порогом» для всей совокупности населения, введенным в целях искусственного регулирования уровня жизни его жителей. Изменяя размер прожиточного минимума, органы управления задают своеобразное «направление» развития общества.

Рассматривая распределение населения по доходам, с одной стороны, как пороговую совокупность, и с другой стороны, как смесь распределения, можно определить верхнюю границу доходов бедного населения и нижнюю границу доходов среднего класса.

В результате вычислений было выявлено, что оптимальным является декомпозиция смеси на два компонента с нормальным и логнормальным законами распределения. Причина возникновения данной ситуации – наличие систематической ошибки наблюдения, возникшая в результате объединения двух разнокачественных типов населения. Однако на практике, статистический учет единиц всей совокупности осуществляется одинаково, без учета особенностей каждой частной пороговой совокупности, что приводит к снижению качества, достоверности и точности статистической информации. Кроме того, установление границ-порогов совокупности носит субъективный характер (не всегда имеет место содержательное обоснование его величины).

По результатам исследования, выявлено, что в 2011 г., пересечение двух распределений проходит на уровне 17 500 руб. Таким образом, верхняя граница среднедушевых денежных доходов бедного населения должна быть на уровне 17 500 руб., в то время как фактический уровень прожиточного минимума составлял всего 6369 руб. Таким образом, приведенные расчеты свидетельствуют о неточности идентификации совокупности бедного населения и среднего класса: с одной стороны, занижение доли бедного населения, и завышение доли населения со средним достатком – с другой.

По официальной методологии, к среднему классу относится население со среднедушевым денежным доходом от 2 до 6 размеров прожиточного минимума. В соответствии с проведенными расчетами, нижняя граница среднего класса должна быть больше общепринятого значения в 2,7 раза (на уровне 35 000 руб.). Эти результаты согласуются с данными построенной выше функции принадлежности.

Определение верхней границы доходов среднего класса затруднено из-за малого числа интервалов, по которым выполняется распределение среднедушевых денежных доходов населения РФ. Это в очередной раз подтверждает необходимость совершенствования методологии исследования статистических совокупностей.

Таким образом, данное исследование свидетельствует об обоснованности введения дополнительных классификационных разрезов совокупностей и их использования в целях улучшения точности и качества полученных статистических данных, используемых для оперативного управления и стратегического планирования.

Литература

1. Венецкий И.Г., Венецкая В.И. Основные математико-статистические понятия и формулы в экономическом анализе. М.: Статистика, 1974. 279 с.
2. Вентцель Е.С., Овчаров Л.А. Теория вероятности и ее инженерные приложения. Учеб. пособие для вузов. 2-е изд. М.: Высшая школа, 2000. 480 с.
3. Глинский В.В. Статистические методы поддержки управленческих решений: монография. Новосибирск: Издательство НГУЭУ, 2008. С. 108–118.

4. Глинский В.В. Мифическая статистика малого бизнеса. Проблемы статистического изучения турбулентных совокупностей // ЭКО. 2008. № 9. С. 51–61.
5. Глинский В.В., Серга Л.К. Нестабильные совокупности: концептуальные основы методологии статистического исследования // Вестник НГУЭУ. 2009. № 2. С. 137–142.
6. Глинский В.В., Серга Л.К. Статистика XXI века. Вектор развития // Вестник НГУЭУ. 2011. № 1. С. 108–118.
7. Глинский В.В., Серга Л.К. О государственном регулировании малого предпринимательства в России // Национальные интересы: приоритеты и безопасность. 2011. № 19. С. 2–8.
8. Кулиджоглян К.О. Критерии принадлежности и проблема оценки среднего класса // Бизнес-статистика, финансы и банки: Теоретические и методические аспекты исследования: сб. науч. тр. Новосибирск: НГУЭУ, 2011. 220 с.
9. Орлов А.И. Нечисловая статистика. М.: МЗ-Пресс, 2004. 513 с.
10. Серга Л.К. Об одном подходе к определению пороговых значений в решении задач классификации // Вестник НГУЭУ. 2012. № 1. С. 54–60.
11. Ту Дж., Гонсалес С. Принципы распознавания образов: монография. М.: Мир, 1978. 412 с.
12. Эверитт Б.С. Большой словарь по статистике / Пер. с англ. Ф.А. Ущева, И.Ю. Чураковой; науч. ред. перевода И.И. Елисеева. 3-е изд. М.: Проспект, 2010. 736 с.
13. Официальный сайт Федеральной службы государственной статистики. URL: <http://www.gks.ru>

Bibliography

1. Veneckij I.G., Veneckaja V.I. Osnovnye matematiko-statisticheskie ponjatija i formuly v jekonomicheskom analize. M.: Statistika, 1974. 279 p.
2. Ventcel' E.S., Ovcharov L.A. Teorija verojatnosti i ee inzhenernye prilozhenija. Ucheb. posobie dlja vtuzov. 2-e izd. M.: Vysshaja shkola, 2000. 480 p.
3. Glinskij V.V. Statisticheskie metody podderzhki upravlencheskih reshenij: monografija. Novosibirsk: Izdatel'stvo NGUJeU, 2008. P. 108–118.
4. Glinskij V.V. Mificheskaja statistika malogo biznesa. Problemy statisticheskogo izuchenija turbulennyh sovkupnostej // JeKO. 2008. № 9. P. 51–61.
5. Glinskij V.V., Serga L.K. Nestabil'nye sovkupnosti: konceptual'nye osnovy metodologii statisticheskogo issledovanija // Vestnik NGUJeU. 2009. № 2. P. 137–142.
6. Glinskij V.V., Serga L.K. Statistika XXI veka. Vektor razvitija // Vestnik NGUJeU. 2011. № 1. P. 108–118.
7. 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.
8. Kulidzhoglyan K.O. Kriterii prinadlezhnosti i problema ocenki srednego klassa // Biznes-statistika, finansy i banki: Teoreticheskie i metodicheskie aspekty issledovanija: sb. nauch. tr. Novosibirsk: NGUJeU, 2011. 220 p.
9. Orlov A.I. Nechislovaja statistika. M.: MZ-Press, 2004. 513 p.
10. Serga L.K. Ob odnom podhode k opredeleniju porogovyh znachenij v reshenii zadach klassifikacii // Vestnik NGUJeU. 2012. № 1. P. 54–60.
11. Tu Dzh., Gonsales S. Principy raspoznavanija obrazov: monografija. M.: Mir, 1978. 412 p.
12. Jeveritt B.S. Bol'shoj slovar' po statistike / Per. s angl. F.A. Ushcheva, I. Ju. Churakovej; nauch. red. perevoda I.I. Eliseeva. 3-e izd. M.: Prospekt, 2010. 736 p.
13. Oficial'nyj sajт Federal'noj sluzhby gosudarstvennoj statistiki. URL: <http://www.gks.ru>