

МНОГОМЕРНАЯ КЛАССИФИКАЦИЯ ДАННЫХ: ОСОБЕННОСТИ МЕТОДИКИ, АНАЛИЗ ПРАКТИКИ И ПЕРСПЕКТИВ ПРИМЕНЕНИЯ

А.М. Ильшев, д-р экон. наук,
О.М. Шубат, канд. экон. наук,
Уральский федеральный университет

Сороковой сессией Организации Объединенных Наций 20 октября 2010 г. объявлен Всемирным днем статистики. Этот день призван обеспечить широкое освещение многочисленных достижений официальной статистики, зиждущихся на основных ценностях служения, добросовестности и профессионализма.

Авторы статьи провели анализ освещения на страницах научно-информационного журнала Росстата «Вопросы статистики» в последние годы опыта применения профессиональными статистиками достаточно сложных (но отвечающих условиям динамичного XXI века) методов многомерной классификации данных, - как в научных исследованиях, так и в статистической и управленческой практике. Использование этого опыта дает возможность научным и научно-педагогическим работникам вузов, НИИ, а также специалистам федеральных и региональных органов управления работать на качественно более высоком уровне.

Методы многомерного анализа получают все большее распространение на современном этапе развития статистической науки и практики. Это обусловлено прежде всего постоянным увеличением сложности статистических совокупностей, ростом связей между составляющими их элементами, что во многом определяется турбулентностью и непредсказуемостью общественного развития, происходящим повсеместно переходом к инновационной экономике.

Методология многомерных статистических процедур адекватна данным процессам, что объективно способствует расширению сферы их применения. Кроме того, активное использование в исследованиях и управленческой практике пакетов прикладных статистических программ позволяет минимизировать рутинную работу специалиста по проведению громоздких статистических расчетов, сконцентрировать усилия на углублении понимания и интерпретации обнаруженных закономерностей.

Понятие «многомерный анализ» включает в себя целый спектр статистических методов и процедур - множественная регрессия, кластерный, дискриминантный, факторный виды анализа и др. Все шире в исследованиях и на практике применяются методы многомерной классификации, позволяющие разделять на однородные группы изучаемые совокупности объектов, характеризующихся множеством признаков. Для решения задач их классификации применяют кластерный анализ (в случае сегментирования наблюдений по заранее неизвестным целевым группам), дискриминантный анализ или логистическую регрессию (при сегментировании по известным группам).

Кластерный анализ в социально-экономических исследованиях. Кластерный анализ - это класс математических методов, которые используются для классификации объектов или событий по однородным группам, называемым *кластерами*. При этом разбиение объектов в кластерном анализе допускается не по одному параметру, а по целому набору признаков. Кроме того, различие между зависимыми и независимыми признаками не проводится. Таким образом, кластеризация позволяет анализировать достаточно большой объем информации и сжимать эти большие массивы данных до более компактных и наглядных.

Задача кластерного анализа - разбиение множества исследуемых объектов на небольшое число однородных групп так, чтобы каждый объект принадлежал одному и только одному кластеру и чтобы объекты одного кластера были сходными, в то время как объекты, принадлежащие разным кластерам, были разнородными. Важной особенностью методов кластеризации является то, что кластеры определяются непосредственно в процессе анализа, а не заранее. В результате наличие у исследователя априорной информации о распределении генеральной совокупности (обучающих выборок) не является обязательным и необходимым условием проведения кластерного анализа.

«Cluster» в переводе с английского означает «гроздь, кисть, пучок, группа», и поэтому этот вид анализа долгое время называли «гроздевым анализом». В научной литературе подход кластерного анализа называют также численной таксономией, численной классификацией, распознаванием с самообучением. Методы кластеризации находят применение в самых разнообразных

научных направлениях - в биологии, экономике, медицине, истории, филологии и т. д. Часто этот вид анализа используется в маркетинговых исследованиях, когда создаются типологии потребителей, решается задача сегментирования рынка. В настоящее время все большую популярность приобретают методы сегментирования потребителей на основании их психографических характеристик. Кластерный анализ используют при построении типологии населенных пунктов, городов и стран.

Кластерный анализ применяется при исследовании временных рядов, иллюстрирующих динамику различных индикаторов социально-экономического развития. Кластеризация в данном случае позволяет выделять периоды, когда значения соответствующих показателей были достаточно близкими, а также определять группы временных рядов, динамика которых наиболее схожа.

О растущей популярности методов кластеризации объектов свидетельствует и число статей, ежегодно публикуемых на страницах журнала «Вопросы статистики» - статей, в которых используется аппарат кластерного анализа. К примеру, за период 2007-2010 гг. на страницах журнала было опубликовано 25 статей с применением методов многомерной классификации данных, использующих различные алгоритмы и единицы кластеризации. Определенную позитивную роль в расширении применения методов кластерного анализа в практической работе играет ежегодный конкурс аналитических записок, подготовленных территориальными органами Росстата.

Контент-анализ публикаций с применением процедуры кластеризации на страницах нашего журнала показал, что многомерный подход является универсальным, научно обоснованным и эффективным средством обработки информации, полученной в ходе экономико-статистического исследования. Практика применения многомерных методов к различным социально-экономическим совокупностям показала их результативность и возможность получения высокоаналитичных научно-практических результатов. Краткая и обобщающая характеристика этих публикаций представлена в таблице 1, а более детальное описание указанных статей показано в таблице 2.

По результатам контент-анализа были выявлены следующие особенности применения методов кластеризации авторами рассмотренных статей.

1. Процедуры многомерной классификации применяются к разнообразным объектам и явлениям социально-экономической действительности. Наиболее типичными объектами кластеризации являются территориальные (региональные) образования - 65% всех публикаций. Такому доминированию есть вполне очевидное объяснение. Дело в том, что сложившаяся в российских условиях высокая степень дифференци-

Таблица 1

Сводная характеристика публикаций с применением процедуры кластерного анализа на страницах журнала «Вопросы статистики» за 2007-2010 гг.

Объекты кластеризации		Количество статей*	
		частота	%
1	Регионы РФ, территориальные и муниципальные образования	15	65
2	Предприятия, организации	4	18
3	Виды экономической деятельности	2	9
4	Вузы	1	4
5	Ряды динамики экономических показателей	1	4
Итого		23	100
Метод кластеризации		Количество статей	
		частота	%
1	Иерархическая кластеризация	14	60
2	Метод k -средних	7	31
3	Не указан	2	9
Итого		23	100
Дополнительное применение других видов многомерного анализа		Количество статей	
		частота	%
1	Применяются	9	39
2	Не применяются	14	61
Итого		23	100

* Две статьи посвящены общеметодологическим вопросам применения многомерных методов анализа.

ции регионов и территорий по совокупности социально-экономических параметров актуализирует потребность в сегментации этих субъектов и выделении однородных групп для более корректного статистического анализа.

2. Нередко (практически в 40% случаев) исследователи практикуют комплексное использование различных методов многомерного анализа данных. К примеру, чаще всего кластерный анализ дополняется факторным анализом, реже - дисперсионным, регрессионным, а также дискриминантным анализом, что, безусловно, повышает аналитичность полученных результатов.

3. Во многих работах предлагается графическая интерпретация процесса анализа, изображающая последовательность объединения (или разделения) кластеров. Визуализация процедур иерархической кластеризации обычно проводится с помощью специального графика - дендрограммы, что дает возможность верифицировать результаты сегментации и повышает уровень доверия к полученным результатам.

4. Бесспорное достоинство отдельных статей - использование дополнительных математико-статистических процедур, свидетельствующих о строгости и корректности применения методов многомерной классификации. Так, для повышения качества кластеризации авторы прибегают к преобразованию исходных

**Описание публикаций с применением процедуры кластерного анализа
на страницах журнала «Вопросы статистики» за 2007-2010 гг.**

№ п/п	Выходные данные	Предмет исследования	Объект кластеризации	Метод кластеризации	Другие процедуры многомерного анализа
1	Ермолицкая Е.В. Методика статистического территориального анализа // ВС. 2007. № 4. С. 37-43	Способы локализации проблемных территорий	Населенные пункты Могилевской области	k -средних	Не применяются
2	Кузнецова В.Е. Методология оценки социального положения городов и районов Оренбургской области // ВС. 2007. № 4. С. 68-73	Уровень социального положения субъектов Оренбургской области	Муниципальные образования Оренбургской области	k -средних	Не применяются
3	Гаджиев Ю.А., Стыров М.М., Колечков Д.В. Дифференциация муниципальных образований Республики Коми по уровню промышленного развития // ВС. 2007. № 4. С. 74-77	Уровень промышленного развития муниципальных образований	Административные районы Республики Коми	Иерархическая кластеризация	Факторный анализ
4	Голубцова О.А. Тенденции развития лесопромышленного комплекса регионов Приволжского федерального округа // ВС. 2007. № 4. С. 81-86	Лесохозяйственная деятельность регионов и уровень развития лесоперерабатывающей промышленности	Субъекты ПФО	k -средних	Не применяются
5	Райская Н.Н., Сергиенко Я.В., Френкель А.А. Кластерный анализ регионов России по уровню инвестиционного потенциала // ВС. 2007. № 5. С. 3-9	Инвестиционный потенциал регионов РФ	Регионы РФ	Иерархическая кластеризация	Не применяются
6	Есенин М.А. Рейтинговая оценка регионов Центрального федерального округа по уровню развития малого предпринимательства // ВС. 2007. № 5. С. 49-54	Уровень развития малого бизнеса	Регионы РФ (с особым вниманием к ЦФО)	Иерархическая кластеризация	Дискриминантный и факторный анализ
7	Охрана окружающей среды Оренбургской области (аналитическая записка Оренбургстата) // ВС. 2007. № 5. С. 69-76	Состояние окружающей среды	Субъекты ПФО	k -средних	Не применяются
8	Продовольственная безопасность Республики Мордовия (аналитическая записка Мордовиястата) // ВС. 2007. № 5. С. 87-93	Условия обеспечения продовольственной безопасности в Республике Мордовия	Динамические ряды потребления продуктов питания	Иерархическая кластеризация	Факторный и регрессионный анализ
9	Заказчикова Н.А. Многомерный статистический анализ деятельности предприятий целлюлозно-бумажной промышленности // ВС. 2007. № 6. С. 54-59	Деятельность предприятий целлюлозно-бумажной промышленности	Предприятия целлюлозно-бумажной промышленности	Иерархическая кластеризация	Факторный анализ
10	Михалкин И.В., Карасева Л.А., Петухова О.В. и др. Подходы к определению пороговых значений выручки для определения сектора малого и среднего предпринимательства // ВС. 2008. № 1. С. 6-14	Уровень представленности в регионах малого и среднего бизнеса	Регионы РФ	Иерархическая кластеризация	Не применяются
11	Заиченко С.А. Использование статистических методов и показателей в оценке научной деятельности вуза // ВС. 2008. № 1. С. 31-42	Научная деятельность вузов	Вузы России	Иерархическая кластеризация	Не применяются
12	Бокун Н.Ч. Использование многомерной выборки для измерения неофициальной экономики // ВС. 2008. № 2. С. 30-37	Рассматриваются методические вопросы применения различных видов кластерного анализа при формировании многомерной выборки.			
13	Громыко Г.Л., Боноев П.С. Использование кластерного анализа в классификации сельскохозяйственных предприятий по показателям эффективности их деятельности // ВС. 2008. № 4. С. 51-54	Деятельность сельскохозяйственных предприятий	Сельскохозяйственные предприятия Республики Бурятия	k -средних	Факторный анализ
14	Ильшев А.М., Шубат О.М. Многомерный статистический анализ предпринимательской активности в региональной сфере микробизнеса // ВС. 2008. № 4. С. 42-51	Представленность в регионах индивидуального производственного предпринимательства	Регионы РФ	Иерархическая кластеризация	Регрессионный анализ
15	Чудилин Г.И., Полянская Н.В., Сиваков И.Ф. Кластерный анализ устойчивости функционирования специализированных садоводческих предприятий в Самарской области // ВС. 2008. № 6. С. 65-69	Способность предприятий к инновационной деятельности	Садоводческие предприятия Самарской области	Иерархическая кластеризация	Не применяются
16	Богаткова Л.В., Пройдакова Е.В. Математические методы в исследовании экономического развития регионов Приволжского федерального округа // ВС. 2008. № 8. С. 45-52	Экономическое положение регионов РФ	Регионы ПФО	k -средних	Факторный анализ

№ п/п	Выходные данные	Предмет исследования	Объект кластеризации	Метод кластеризации	Другие процедуры многомерного анализа
17	Зарова Е.В., Проживин Р.А. Сбалансированная система показателей развития региона: статистическое обоснование и эконометрическое моделирование // ВС. 2008. № 8. С. 59-66	Уровень социально-экономического и институционального развития регионов РФ	Муниципальные районы Самарской области	Иерархическая кластеризация	Методы эконометрического моделирования
18	Жекова С.А. Возрастные детерминанты пространственных различий в продолжительности жизни в Республике Болгария // ВС. 2008. № 9. С. 27-36	Продолжительность жизни в различных регионах Республики Болгария	Территориальные единицы Республики Болгария	Не указан	Не применяются
19	Гаджиев Ю.А., Колечков Д.В. Межрегиональные различия в экономическом развитии зоны Севера // ВС. 2008. № 10. С. 8-14	Уровень экономического развития региона	Регионы зоны Севера (16 регионов)	Иерархическая кластеризация	Не применяются
20	Сушко Е.Д. Анализ взаимозависимости эффективности деятельности предприятий и особенностей политики их менеджмента в сфере оплаты труда // ВС. 2009. № 2. С. 31-38	Эффективность функционирования предприятий	Предприятия оборонной промышленности	k -средних	Не применяются
21	Тюленева Н.А., Авхимович А.О. Сравнительный анализ затрат работодателя на рабочую силу в региональном масштабе // ВС. 2009. № 2. С. 43-48	Затраты работодателя на рабочую силу	Виды экономической деятельности	Иерархическая кластеризация	Не применяются
22	Кандилов В.П., Малышева Т.В. Мониторинг качества жизни населения муниципальных образований Республики Татарстан // ВС. 2009. № 4. С. 32-36	Уровень жизни населения	Муниципальные образования Республики Татарстан	Иерархическая кластеризация	
23	Бокун Н.Ч. Проблемы построения многомерной выборки в сфере розничной торговли // ВС. 2010. № 3. С. 52-60	Рассматриваются общеметодологические и методические вопросы построения многомерных выборок, в том числе с использованием иерархической кластеризации			
24	Жандаров А.М., Шиллер Ф.Ф. Государственное регулирование экономик: проблемы моногородов России // ВС. 2010. № 4. С. 68-75	Механизмы и способы решения проблем моногородов России	Проблемные моногорода России (27 муниципальных образований)	Не указан	Не применяются
25	Антохонова И.В., Полухина О.А. Об использовании информационных технологий и их влиянии на развитие экономики России // ВС. 2010. № 5. С. 61-67	Особенности развития информационно-коммуникационной среды в России	Виды экономической деятельности	Иерархическая кластеризация	Регрессионный анализ

данных (процедуры нормирования). Кроме того, в ряде статей проводится четкое профилирование сформированных кластеров и обоснование существенности их различий.

5. Чаще всего в представленных статьях исследователи применяют иерархические методы кластеризации. В большинстве случаев выбор метода авторами публикаций определяется спецификой исходной информации.

Методы кластерного анализа, объясняющие порядок и технику формирования кластеров, обычно делятся на две группы: иерархические и неиерархические. Каждая из этих групп представлена множеством подходов и алгоритмов, которые дают различные решения для одних и тех же данных (в принципе, неоднозначность сегментации исходного массива информации считается нормальной и типичной ситуацией для кластерного анализа).

Иерархический метод кластерного анализа позволяет формировать кластеры с помощью построения так называемой иерархической, или древовидной структуры данных. Иерархические методы в свою очередь могут быть разделены на два вида - агломеративные и

дивизивные. Агломеративная (объединительная) кластеризация предполагает последовательную группировку объектов во все более крупные сегменты. Дивизивная (разделяющая) кластеризация по своей технике противоположна агломеративной кластеризации и предполагает последовательное деление объектов на все меньшие кластеры.

Необходимо подчеркнуть, что для выполнения процедур иерархической кластеризации требуется довольно большой объем арифметических расчетов. Именно поэтому иерархические методы кластерного анализа используются при небольших объемах выборки.

Неиерархический метод кластерного анализа предполагает, что вначале определяется количество кластеров, на которые будет дробиться изучаемая совокупность. Центры этих кластеров рассчитываются, а затем все объекты в пределах заданного порогового значения объединяются в тот или иной кластер. При этом объект присоединяется к ближайшему кластеру.

Наиболее распространенным на практике методом неиерархической кластеризации является *метод k -средних (быстрый кластерный анализ)*, разбивающий совокупность объектов на k сегментов, расположенных

на возможно больших расстояниях друг от друга. Данный метод кластеризации быстрее иерархических методов; именно поэтому метод k -средних рекомендуют использовать при больших объемах выборки. Однако у этого метода есть существенные недостатки, сужающие возможности его применения.

Во-первых, метод k -средних чувствителен к «выбросам» (нетипичным значениям в исследуемой совокупности). Поскольку последние могут серьезно искажать средние значения (и следовательно, искусственно смещать кластерные центры), то сама процедура кластеризации порождает опасность некорректного распределения объектов по кластерам. Для преодоления указанного недостатка в методе могут использоваться робастные статистики - возможной альтернативой методу k -средних является метод k -медианы.

Во-вторых, результаты кластеризации рассматриваемым методом могут зависеть от порядка расположения исходных данных. Поэтому для проверки качества кластеризации приходится повторять процедуры анализа, меняя порядок следования наблюдений в массиве данных.

Следует, однако, отметить, что не во всех публикациях авторы придерживаются строгости применения процедур сегментации и оценки достоверности полученных кластеров. Часто авторы не обосновывают выбор того или иного алгоритма кластеризации и выбор мер расстояния между объектами. Кроме того, в ряде статей процедура кластеризации проводится на очень небольшом объеме выборки, в результате чего кластеры получаются слишком малого размера - один-два объекта.

Однако в целом методы кластеризации уже достаточно широко и успешно применяются (и в это вносит свой вклад журнал «Вопросы статистики») при проведении статистических исследований различных сфер социально-экономической жизни и видов деятельности: инвестиционной, промышленной и ее отдельных отраслей, агропромышленной, малого бизнеса, социальной сферы, экологической и др.

Логистическая регрессия и дискриминантный анализ. Логистическая регрессия относится к классу нелинейных регрессий и позволяет исследовать зависимость дискретных переменных от предикторов, имеющих любой вид шкалы. С помощью методов логистической регрессии рассчитывается вероятность наступления какого-либо события (например, вероятность отнесения респондента к одной из групп) в зависимости от значений независимых переменных. И в этой связи данная регрессия, наряду с дискриминантным анализом, является удобным инструментом решения задач классификации. Методы логистической регрессии активно применяются в медицинских клинических исследованиях и в банковской сфере для оценки кредитоспособности физического лица (техно-

логии скоринга), для расчета рейтинга заемщиков и управления кредитными рисками.

Дискриминантный анализ дает возможность считать вероятности попадания каждого наблюдения в ту или иную группу, а также позволяет составить уравнение дискриминантной функции. Однако сфера применения дискриминантного анализа несколько уже, чем у логистической регрессии, поскольку этот вид анализа требует выполнения строгих предположений о распределении исходных данных: предикторы должны иметь многомерное нормальное распределение с идентичными ковариационными матрицами.

Надо отметить, что аппарат логистической регрессии и ROC-анализа (ROC-кривая строится для визуализации и графического представления результатов бинарной логистической регрессии) еще недостаточно применяется отечественной наукой, рассмотрение основ этого вида регрессионного анализа крайне редко встречается в работах российских статистиков. Наиболее полно техника применения логистической регрессии представлена в работах И.И. Елисеевой, Я.Р. Магнуса, П.К. Катышева, А.А. Пересецкого, В.С. Мхитаряна. Однако чаще всего описание основ и процедур логистической регрессии и ROC-анализа все же приводится в иностранных источниках.

На страницах журнала «Вопросы статистики» за период с 2007 по 2010 г. появилось лишь восемь статей с применением аппарата логит-регрессии и дискриминантного анализа. Тем не менее частота появления таких публикаций все же демонстрирует положительную динамику. Так, в 2007 и 2008 гг. было опубликовано только по одной статье, где рассматривался аппарат логистической регрессии. В 2009 г. таких статей было уже три, и столько же публикаций вышло за неполный 2010 г. Хочется надеяться, что положительная динамика количества публикаций с применением методов многомерной классификации сохранится и в последующие периоды.

Сводная характеристика статей, в которых освещались вопросы применения логистической регрессии, дана в таблице 3. Как следует из представленных данных, этот метод классификации успешно применяется для анализа различных факторов социальной и экономической жизни, позволяя проводить многомерную группировку и классификацию в условиях большого числа разнотипных параметров.

Завершая рассмотрение практики применения многомерной классификации данных, хотелось бы, наряду с общей положительной оценкой этой практики и ее освещения на страницах журнала «Вопросы статистики», сконцентрировать внимание заинтересованных читателей (как профессиональных статистиков, так и специалистов других профилей, все чаще обращающихся к методам многомерной классификации) на трех очень важных обстоятельствах.

**Сводная характеристика публикаций с применением логистической регрессии
на страницах журнала «Вопросы статистики» за 2007-2010 гг.**

№ п/п	Выходные данные	Специфика применения логистической регрессии
1	Архипова М.Ю. Статистический анализ основных тенденций создания и использования передовых технологий // ВС. 2007. № 7. С. 67-72	Бинарная логистическая регрессия применяется для оценки интенсивности использования передовых технологий промышленными предприятиями. Проводится классификация промышленных предприятий по двум группам: 1 - предприятия с высоким уровнем использования технологий; 2 - предприятия, с низким уровнем использования технологий
2	Бурков А.В. Основные подходы к моделированию оценки вероятности трудоустройства выпускников вузов России и США // ВС. 2008. № 6. С. 47-55	В работе с помощью регрессионных логит- и пробит-моделей оценивается вероятность основания нового бизнеса выпускниками вузов в зависимости от их социально-демографических и психоэмоциональных характеристик
3	Ильшев А.М., Шубат О.М. Возможности углубления статистического изучения малого предпринимательства // ВС. 2009. № 2. С. 8-17	Оценивается влияние различных социально-экономических индикаторов на развитие в регионах России высокого уровня производственного индивидуального предпринимательства, а также вероятность отнесения региона к группе субъектов с высокой мерой представленности этого вида малого бизнеса
4	Мурзачёва Е.И.. Статистический анализ кредитного риска в экономике при финансировании малого бизнеса: оценка потенциальных потерь и факторов их возникновения // ВС. 2009. № 11. С. 19-27	Аппарат логистической регрессии применяется для оценки вероятности дефолта раннего предпринимательства
5	Образцова О.И., Гулеева Ю.А. Нарождающееся предпринимательство в различных типах поселений: выбор населением экономического поведения в условиях глобального кризиса // ВС. 2009. № 11. С. 27-39	Аппарат логит-регрессии применяется для оценки вероятности выбора индивидом варианта экономического поведения на основании различных социально-экономических индикаторов
6	Бакуменко Л.П., Сиразеева А.А. Анализ выживаемости при заболеваемости злокачественными новообразованиями в Республике Марий Эл // ВС. 2010. № 2. С. 74-77	Построена логит-регрессионная модель, которая оценивает шансы выживания онкобольных. Исследование дополняется дискриминантным анализом
7	Родионова Л.А. Демографическая политика и репродуктивное поведение женщин в России: эконометрический анализ // ВС. 2010. № 4. С. 41-43	В работе анализируется влияние двух типов пособий (по уходу за ребенком до 1,5 лет и единовременного пособия при рождении) на уровень рождаемости. Автором оценивается вероятность положительного решения женщины о рождении второго и третьего ребенка, предлагается классификационное уравнение, где предиктор - индивидуальные характеристики женщины, социально-демографические и экономические факторы
8	Власова Е.А. Анализ показателей движения трудовых ресурсов в российской экономике // ВС. 2010. № 5. С. 44-50	В работе используется аппарат множественной логистической регрессии, позволяющий классифицировать респондентов по критерию экономической активности на три группы - занятые, безработные, экономически неактивные - на основании их социально-демографических характеристик

Во-первых, столь желанный и нужный для России и ее регионов переход на инновационный путь развития, наконец-то, начинает приобретать зримые черты. В связи с этим Росстат разработал и утвердил статистический инструментарий для организации статистического наблюдения за деятельностью предприятий и организаций в сфере нанотехнологий - квартальную форму № П-1 «Сведения об отгрузке товаров, работ и услуг, связанных с нанотехнологиями». Достаточно очевидно, что для осуществления эффективного контроля и управления за их внедрением (а именно нанотехнологии являются основой шестого технологического уклада, идущего на смену четвертому и пятому) необходимо в процессе проведения статистического анализа своевременно дифференцировать весьма разнообразные регионы нашей страны по «уровню нанотехнологичности», выявить причины попадания того или иного региона в кластеры с низким и высоким уровнем развития нанотехнологий, отслеживать складывающиеся в разных кластерах тенденции.

Во-вторых, мощную научную поддержку формируемой инновационной экономике призваны оказывать создаваемые в России национальные исследовательские и федеральные университеты. Наиболее важные стороны многогранной деятельности таких университетов также должны получить полное отражение не только путем совершенствования текущего статистического наблюдения за их работой, но и посредством углубленного статистического анализа подлинных (а не виртуальных) достижений отечественной интеллектуальной элиты. При этом практически невозможно обойтись без широкого применения методов многомерного анализа - кластерного, факторного, дискриминантного, множественной регрессии.

В-третьих, властная необходимость обеспечения эффективного статистического контроля за переходом от суженного к простому, а затем и к расширенному воспроизводству человеческих ресурсов в нашей стране, потребность в более глубоком познании действующих

щих в этой сфере закономерностей требуют нестандартных подходов и к аналитическому инструментарию. Ведь здесь тесно переплетается действие экономических, социальных, психологических, физиологических, экологических, медицинских, этноконфессиональных и многих других факторов. Вполне понятно, что глубокий анализ сложнейших и недостаточно познанных процессов, которые происходят в сфере человеческой

репродукции, объективно требует комплексного применения уже известных методов и процедур многомерной классификации, а также разработки и апробации новых методов.

Нет сомнений, что имеющийся богатый опыт, высокий профессионализм и подвижничество отечественных статистиков являются залогом успешного решения стоящих перед всеми нами новых задач.

ПРИМЕНЕНИЕ КЛАСТЕРНОГО АНАЛИЗА В ИЗУЧЕНИИ КОММЕРЦИАЛИЗАЦИИ ИННОВАЦИОННЫХ ТЕХНОЛОГИЙ

Л.Л. Никитин,

Государственный университет - Высшая школа экономики

Модернизация экономики и развитие инноваций в настоящее время являются одной из ключевых задач как власти, так и предпринимательского сообщества России. Создание особых экономических зон, федеральных университетов, основание «Российской венчурной компании» и корпорации «Роснано», запуск «Сколково» и прочие подобные проекты призваны способствовать решению этой задачи. Вместе с тем перечисленные инициативы направлены в первую очередь на развитие технологий, в то время как успешные инновации, помимо этапа НИОКР, включают в себя коммерциализацию. Конечной целью последней является построение бизнес-модели, которая позволила бы либо создать на основе перспективной технологии продукт, который бы дал возможность получать прибыль и за счет этого финансировать дальнейшее развитие данной технологии, либо укрепить конкурентную позицию той или иной компании в условиях роста новой экономики.

В данной статье анализируются основные параметры бизнес-моделей успешных российских и зарубежных наукоемких компаний. Результаты анализа могут быть использованы современными российскими стартапами и инновационными подразделениями компаний для коммерциализации разрабатываемых ими технологий и выстраивания собственных бизнес-моделей.

Понятие бизнес-модели. В данной работе мы определяем бизнес-модель как *концепцию реализации стратегии компании, позволяющую ей в стратегической перспективе получать прибыль за счет создания продуктов, имеющих покупательскую ценность*. Бизнес-модель - это способ взаимодействия компании: ее материальных и нематериальных активов (в том числе интеллектуального капитала), системы бизнес-процессов и порядка принятия управленческих решений, с внешней средой, позволяющий организации созда-

вать ценность для покупателя и извлекать прибыль из его готовности за нее платить.

Цели проведения кластерного анализа. С целью выявления основных концепций построения бизнес-моделей компаниями наукоемких отраслей, типов этих бизнес-моделей нами был проведен кластерный анализ более чем 150 компаний из следующих отраслей: авиа и космос, ИТ, фармацевтика и биотехнологии. На начальных этапах также анализировались автомобильная и телекоммуникационная отрасли, однако позже они были исключены из выборки. Причиной исключения указанных отраслей из анализа стала преобладающая неприбыльность или убыточность представляющих их компаний на протяжении последних нескольких лет (с 2005 по 2010 г.), усугубившаяся с началом мирового финансово-экономического кризиса в 2007 г. Этот и ряд других фактов позволили судить о вступлении автомобильной и телекоммуникационной отраслей в фазу стратегического распада имеющихся бизнес-моделей, что в силу предсказуемости результата делало их анализ нецелесообразным в рамках нашего исследования, предполагающего в том числе выявление бизнес-моделей, еще не вступивших в эту стадию (но, возможно, подверженных этому риску).

В ходе кластерного анализа проверялись следующие гипотезы:

1. Технологическое лидерство не является достаточным условием для получения высокой прибыли;
2. В современных наукоемких отраслях существует тенденция к сегментации и дифференциации, а нишевые игроки имеют более высокую рентабельность, нежели компании, ориентированные на более широкий рынок;
3. Модель закрытых инноваций и вертикальной интеграции по цепочке «ценности» в ряде случаев менее эффективна, чем модель открытых инноваций (совме-

стных НИОКР и коммерциализации с разделением итоговой прибыли между участниками);

4. В наукоемких отраслях сетевая модель позволяет получить более высокую прибыль.

Выборка. В выборку для статистического (кластерного) анализа были включены российские и зарубежные компании, представляющие перечисленные выше отрасли: около 50 из отрасли ИКТ и по 25 компаний из аэрокосмической и фармацевтико-биотехнологической отраслей. Основными критериями включения в выборку были доступность финансовой и иной, необходимой для нашего анализа деловой информации о деятельности компании и финансовая устойчивость (компании, не получавшие прибыли на протяжении 2006-2009 гг., исключались из выборки).

Выборка также была дополнительно сбалансирована по размеру компаний таким образом, чтобы количество малых, средних и крупных предприятий было примерно равным. Как мы отмечали выше, в выборку был включен ряд российских наукоемких компаний. Следует заметить, что в России действуют лишь единицы наукоемких компаний, крупных по мировым меркам, причем все они являются представителями аэрокосмической отрасли. Ни в отрасли ИКТ (в нашем определении), ни в фармацевтической крупных российских компаний нет. Однако наше исследование сконцентрировано не на размерах и абсолютной величине выручки компаний, а на рентабельности, свидетельствующей о прибыльности, успешности бизнес-модели. По этой причине относительно малое количество крупных российских компаний в выборке не должно существенно повлиять на общий результат.

Факторы и метод анализа. Целью статистического анализа, проведенного нами, было выделить основные *работающие* бизнес-модели в наукоемкой сфере. Отметим, что в нашем исследовании рассматриваются только коммерческие наукоемкие компании. Различные некоммерческие организации (в том числе государственные), ведущие фундаментальные научные исследования и прикладные НИОКР, не являются объектом нашего исследования в первую очередь потому, что эти организации по определению не нацелены на получение прибыли и, следовательно, не в полной мере ориентированы на создание бизнес-моделей (хотя могут представлять определенный интерес как участники сетей создания ценности).

Многофакторный кластерный анализ был проведен по следующим факторам:

1. Отрасль (аэрокосмическая, ИТ и телекоммуникации, биофармацевтика). Источник информации - публичная отчетность компаний. В случае, если компания работает в нескольких сферах, ее отрасль определялась по преобладающему в выручке и прибыли направлению.

2. Размер компании (малый, средний или крупный), значения оборота за период были нормированы к отраслевым значениям (например, при равном обороте аэрокосмическая компания попадала в разряд средних, а ИТ-компания - в разряд крупных). Источник информации об оборотах компаний: для корпораций (российских АО) - публичная отчетность, для частных компаний (российских ООО) - экспертные оценки (деловая пресса, отраслевые рейтинги и т. п.). Этот фактор является вспомогательным, так как напрямую не фигурирует в предложенной нами ранее структуре бизнес-модели компании. Однако фактор размера компании достаточно важен для дальнейшего анализа, поскольку позволяет судить о ресурсах компании, необходимой для ее развития доле рынка, и главное, о возможности применения той или иной компанией модульных стратегий.

В числах данный фактор был выражен следующим образом: все компании внутри одной отрасли были проранжированы в соответствии со средним оборотом каждой из компаний в 2005-2009 гг. Затем эти компании были разделены на три группы по следующему правилу:

$$\frac{1}{3}(I_{\max} - I_{\min}) \leq I_i < \frac{2}{3}(I_{\max} - I_{\min}),$$

где I_{\max} - среднегодовая выручка самой крупной компании;
 I_{\min} - среднегодовая выручка самой маленькой по этому параметру компании.

Соответственно, если выручка i -й компании выборки I_i удовлетворяла приведенному выше неравенству, то данная компания заносилась в категорию средних. Если I_i удовлетворяла условию

$$I_i < \frac{1}{3}(I_{\max} - I_{\min}),$$

то данная компания заносилась в категорию малых, а если для I_i выполнялось условие

$$I_i \geq \frac{2}{3}(I_{\max} - I_{\min}),$$

то данная компания заносилась в категорию крупных.

Компании разных отраслей анализировались отдельно по изложенным выше причинам.

3. Позиционирование по этапам инновационного процесса:

а) интегратор (компания самостоятельно ведет НИОКР и полностью коммерциализирует разработки);

б) дирижер (компания совместно с партнерами ведет НИОКР, объединяя их результаты в процессе коммерциализации в единый продукт, который продает самостоятельно);

с) лицензиар (компания только ведет НИОКР, оформляет патенты на полученные результаты и лицензирует их, не участвуя в коммерциализации).

Источник данных - интегральная экспертная оценка по следующим параметрам:

а) количество официально оформленных патентов (источником данных для зарубежных патентов являлись официальные базы выданных (в том числе действующих) патентов США, Канады Японии и стран ЕС; для российских патентов - публикации в отраслевой и деловой прессе и экспертные оценки¹);

б) количество заключенных лицензионных договоров о передаче прав на использование интеллектуального капитала (ИК) (источник информации для зарубежных стран - собственная информация компаний, отраслевые аналитические отчеты; для России - собственная информация компаний, публикации в отраслевых и деловых СМИ, экспертные оценки);

с) количество и теснота деловых взаимосвязей с партнерами по разработке, продаже и маркетингу, а также с лицензиарами (источник данных - собственная информация компаний, публикации в отраслевой и деловой прессе).

4. Источник интеллектуального капитала:

а) секрет (организация держит результаты НИОКР в секрете);

б) патентование (результаты разработок патентуются);

с) открытие (организация-автор позволяет бесплатно использовать созданный ею ИК с целью изменить структуру отрасли и/или стимулировать развитие технологий в выгодном ей направлении).

Концептуальные различия между перечисленными моделями использования ИК представлены в таблице 1.

Таблица 1

Обзор моделей использования ИК

	Свобода ознакомления	Свобода использования
Секрет	Нет	Нет
Защита патентами	Да	Нет
Открытие	Да	Да*

* С нефинансовыми ограничениями.

Источник данных - интегральная экспертная оценка по параметрам:

а) наличие секретной (не имеющей публично доступного описания, в том числе в виде официально выданного патента) технологии (или иных видов ИК) и ее значимость для создания ценности;

б) количество официально оформленных патентов (источники информации аналогичны указанным в п. 3а);

с) активность публикации той или иной компанией (в лице своих сотрудников) статей в научных, отраслевых и деловых изданиях и других источниках информации о собственном ИК и технологиях (источник информации - базы статей научных и отраслевых изданий; автоматически составляемые дайджесты деловых СМИ);

д) наличие и значимость партнеров, использующих ИК компании (источник информации - собственные данные компаний, публикации в СМИ и экспертные оценки).

5. Базовая экономическая модель. Источник информации - экспертные оценки на основании публичной отчетности компании, статей в научных, отраслевых и деловых СМИ:

а) создание продукта, повышающего активность покупателей и поддерживающего спрос с их стороны;

б) создание системы из высокомаржинальных и защитных продуктов в более низких сегментах рынка;

с) создание системы из высокомаржинальных и массовых поддерживающих продуктов;

д) создание системы с платными транзакциями между участниками и контроль над ней;

е) циклическое создание продуктов, в течение определенного времени несущих уникальную клиентскую ценность;

ф) высокие объемы продаж небольшого количества продуктов с высокими фиксированными и низкими предельными издержками;

г) создание системы многократного использования продукта различными способами;

h) создание высокопрофессиональной, мотивированной на улучшение связи с клиентами, команды;

i) создание высококачественного нишевого продукта с высокой клиентской ценностью;

j) создание системы из взаимодополняющих продуктов с минимизацией единоразового входного платежа и обеспечение высокой прибыльности от периодических платежей;

к) создание продукта, являющегося стандартом или реализующего его и необходимого для функционирования широкого спектра смежных продуктов [1].

Все перечисленные факторы, кроме размера компании, были приведены к номинальным шкалам. Размер компании был выражен порядковой шкалой.

Многофакторный кластерный анализ проводился при помощи иерархического алгоритма методом единичных связей (метрика - минимальные евклидовы расстояния). Данная метрика была выбрана по причине низкой корреляции между перечисленными выше факторами (метод евклидовых расстояний достаточно хорошо работает для объектов в скоплениях, характерных для слабо коррелированных совокупностей).

¹ Данные источники информации были выбраны, поскольку официально патентные базы Федеральной службы по интеллектуальной собственности, патентам и товарным знакам не доступны для исследований.

Значения парных корреляций Спирмена для факторов (переменных) приведены в таблице 2. Корреляции были рассчитаны по методу Спирмена, поскольку мы используем качественные шкалы, для которых корректно рассчитывать коэффициент корреляции именно по методу Спирмена (а не Пирсона). Видно, что в основном факторы слабо коррелируют друг с другом, за исключением пары *базовая экономическая модель* - *от-*

расль. Высокая зависимость между этими факторами достаточно просто объясняется эмпирикой: действительно, внутри одной отрасли набор успешных бизнес-моделей значительно уже, нежели в целом в экономике. Тем не менее факт слабой зависимости между остальными факторами определяет наш выбор метода кластерного анализа в пользу метрики евклидовых расстояний.

Таблица 2

Коэффициенты корреляции Спирмена для факторов кластерного анализа

	Отрасль	Размер компании	Позиционирование по этапам инновационного процесса	Источник интеллектуального капитала	Базовая экономическая модель
Отрасль	-	-0,18	0,24	-0,13	-0,6
Размер компании	-0,18	-	0,1	0,05	0,03
Позиционирование по этапам инновационного процесса	0,24	0,1	-	0,19	-0,02
Источник интеллектуального капитала	-0,13	0,05	0,19	-	0,27
Базовая экономическая модель	-0,6	0,03	-0,02	0,27	-

В результате были сформированы четыре кластера, которые будут более подробно описаны далее. Заметим, что шесть компаний (6% выборки) не были отнесены ни в один из основных кластеров, что объясняется тем, что данные компании постоянно меняют свою бизнес-модель.

Итоговое количество кластеров является минимально возможным при условии однородности наблюдений внутри каждого из кластеров. В целях нашей работы мы понимаем под однородными кластерами такие, наблюдения внутри которых объединены существенными с точки зрения бизнес-модели логическими и эмпирическими сходствами. Внутри полученных нами четырех кластеров компании характеризовались близкими (в рамках рассматриваемых в нашей работе характеристик) бизнес-моделями в целом, а также сходными моделями создания ценности и, как показал дальнейший анализ, сильными сторонами и уязвимостями.

Результаты кластерного анализа. Как уже указывалось выше, каждый из четырех кластеров объединил компании, близкие не только по собственно факторам-переменным кластеризации, но и по концепциям:

- создания ценности;
- подхода к осуществлению НИОКР и коммерциализации их результатов;
- использования интеллектуального капитала.

Также внутри кластеров достаточно мало разнообразие используемых моделей прибыли и бизнес-моделей, что объяснимо с точки зрения выбранных нами факторов кластеризации.

Анализ полученных нами кластеров (которые могут отождествляться с господствующими бизнес-моделями современных наукоемких компаний) также позволяет выделить сильные и слабые стороны соответствующих бизнес-моделей. Как мы покажем далее, компании внутри выделенных нами кластеров и их бизнес-модели имеют очень сходные угрозы и соответственно похожие пути для избежания данных угроз и укрепления своей бизнес-модели и конкурентных позиций.

Описания итоговых кластеров.

Кластер 1. Представители традиционного подхода

Типичные представители: Sikorsky, Bombardier Aerospace, Robinson, Alenia Aeronautica, Eurocopter, Embraer, NXP Semiconductor, NetApp, Juniper Networks, Saab, Сухой, Yahoo!, Garmin, Acronis, Boeing, EADS, Intuit, Parallels, Xilinx, Marvell, МиГ, Dell, Adobe.

Компании, находящиеся в этой группе, характеризуются следующими особенностями:

- Позиционирование по этапам инновационного процесса - интегратор или дирижер, но никогда лицензиар.
- Источник интеллектуального капитала - в основном секрет и патентование.
- Основные экономические модели:
 - создание продукта, повышающего активность покупателей и поддерживающего спрос с их стороны;
 - создание системы из взаимодополняющих продуктов с минимизацией единоразового входного платежа и обеспечение высокой прибыльности от периодических платежей;

с) создание системы из высокомаржинальных и массовых поддерживающих продуктов.

4. Прибыльность - низкая или средняя (здесь и далее данные об уровне прибыльности даются применительно к среднеотраслевому уровню).

Подобные компании присутствуют во всех исследованных отраслях; размеры этих компаний могут быть различными. Однако следует заметить, что большая их часть достаточно давно работает на рынке; компаний моложе 10 лет в их числе нет. Именно к этой группе в основном относятся аэрокосмические компании, которые, с одной стороны, тесно связаны с военно-промышленным комплексом (или даже являются его частью). Этот факт частично объясняет выбор источников интеллектуального капитала (в данной группе это в основном секрет и патентование), поскольку для компаний, связанных с ВПК, по причинам секретности зачастую невозможно использовать открытие своего ИК. Тем не менее даже среди аэрокосмических компаний в последнее время все чаще встречается использование открытых моделей, однако в основном лишь в части, касающейся космических исследований и гражданской авиации.

Использование закрытых моделей управления ИК компаниями фармацевтической и биотехнологической отраслей, попавшими в данный кластер, также объяснимо. Многие из них достаточно давно работают в жесткой конкурентной среде, ведут все более и более дорогостоящие исследования примерно в одном направлении (список перспективных лекарств, платежеспособный спрос на которые может окупить их разработку, достаточно короткий; в нем лишь несколько сотен наименований). Поэтому наиболее естественная реакция менеджмента в данной ситуации - засекретить свои разработки и с помощью защитного патентования закрыть для конкурентов наиболее перспективные направления исследований. Однако и в этой отрасли появляются успешные открытые модели управления ИК. Один из примеров - крупная фармацевтическая компания Merck [8], несколько лет назад существенно продвинувшаяся в расшифровке определенной части генома человека, необходимой для разработки целого ряда важных лекарств, в частности от болезни Альцгеймера. В этой ситуации, чтобы защитить свои инвестиции в исследования и укрепить свою конкурентную позицию в стратегической перспективе, руководители компании приняли решение не засекретить или запатентовать полученный ИК, а поступить противоположным образом - открыть его, опубликовав значительную часть информации в отраслевых изданиях и на профессиональных форумах и конференциях. Это позволило компании не только закрепить свое авторство, но и предупредить патентование того же ИК конкурентами (а они рано или поздно также получили бы аналогичные результаты, поскольку тоже вели иссле-

дования в сходном направлении), что существенно затруднило бы Merck разработку собственно лекарств на базе полученных ею знаний о геноме. Повторим, данные модели все чаще используются в фармацевтической и биотехнологической отраслях, когда компании для проведения прикладных НИОКР (разработки лекарств, готовых к выводу на рынок) вынуждены проводить длительные фундаментальные исследования, а между получением фундаментальных научных результатов и разработкой лекарства проходит несколько лет (а иногда и десятилетий), в течение которых компании необходимо свободно использовать результаты фундаментальных исследований, а для этого - защитить их от патентования.

Что же касается ИТ-компаний, то для многих из них, как показывает подробное рассмотрение (в том числе кейс-анализ), следование закрытой модели не является оправданным. Будучи сетевой по своей сути, ИТ-отрасль (как и телекоммуникационная), как правило, выигрывает от использования открытых и сетевых моделей. Учитывая, что большинство ИТ-компаний, отнесенных к данному кластеру, являются достаточно старыми (по меркам отрасли), использование ими закрытых моделей со значительной степенью достоверности можно считать традиционным, логически вытекающим из общепринятых правил ведения бизнеса в момент создания этих компаний (когда многие из открытых моделей не существовали, не были предусмотрены юридически или не могли быть приняты сообществом партнеров).

Также целесообразно заметить, что все компании данного кластера действуют на высококонкурентных рынках, соответственно здесь же создается ценность и получается прибыль. Компании данного кластера редко являются ярко выраженными лидерами отрасли по показателям оборота (они одни из многих игроков).

Таким образом, в кластере традиционных компаний преобладают крупные технологические компании, многие из которых были основаны более 20 лет назад и последние 10 лет (или даже более) являются крупными.

Кластер 2. Технологические компании - центры экосистем

Представители: Google, Microsoft, Thales, Cisco.

Особенности компаний данной группы следующие.

1. Крупные компании.
2. Тип использования интеллектуального капитала - дирижер, другие типы не встречаются.
3. Концепция управления интеллектуальной собственностью - секрет, другие не встречаются.
4. Превалярующие бизнес-модели:
 - а) создание продукта, являющегося стандартом или реализующего его и необходимого для функционирования широкого спектра смежных продуктов;

б) создание системы из взаимодополняющих продуктов с минимизацией единовременного входного платежа и обеспечение высокой прибыльности от периодических платежей;

с) создание высококачественного нишевого (модульного) продукта с высокой клиентской ценностью.

5. Прибыльность - высокая.

Компании данного кластера являются создателями и модераторами сетей создания ценности или технологических платформ. Как правило, под их контролем находятся некоторые базовые технологии в широком смысле и системы деловых взаимосвязей. При этом последние сконфигурированы таким образом, чтобы обеспечить как можно более широкое распространение этих технологий и за счет этого получить максимальный размер и устойчивость сети создания ценности.

Целесообразно отметить, что в эту группу попало всего четыре компании (2%) из нашей выборки, однако данный кластер статистически выражен и явно отстоит от всех остальных. Малочисленность наблюдений в этом кластере объясняется тем, что, по определению, разработчиками/владельцами платформ (или экосистем) могут быть лишь немногие компании, имеющие соответствующие компетенции и ресурсы. Сам смысл платформы - в объединении целого ряда компаний вокруг компании - центра экосистемы.

С этой точки зрения объяснимо, что три из четырех компаний кластера представляют ИТ-отрасль, и лишь одна из них - Thales - является аэрокосмической. Именно в ИТ-отрасли, где наиболее сильно действуют сетевые эффекты (когда ценность сети для каждого участника возрастает с присоединением к сети нового участника), платформы могут быть наиболее выгодны как для владельцев, так и для остальных участников, а открытые модели приносят наибольшую экономическую выгоду в виде снижения издержек и роста выручки. Этим же объясняется, что из аэрокосмических компаний центром экосистемы (владельцем технологической платформы) стала именно Thales, основную долю выручки получающая от продажи электроники, то есть ИТ-продуктов: авионики, электронных систем для наземного вооружения, тепловизоров, систем наведения и управления огнем и т. п., а также программного обеспечения.

Кластер 3. Модульные лидеры

Типичные представители: Autodesk, Researchinmotion, Oracle, Cerner, Fiserv.

Компании этой группы имеют следующие особенности:

1. В основном средние компании, реже малые и крупные.

2. Превалирующее позиционирование по этапам инновационного процесса - интегратор, реже лицензиар и дирижер.

3. Источник интеллектуального капитала - защита патентами, другие в выборке не встречаются.

4. Превалирующие бизнес-модели:

а) создание продукта, являющегося стандартом или реализующего его и необходимого для функционирования широкого спектра смежных продуктов;

б) создание высококачественного нишевого (модульного) продукта с высокой клиентской ценностью.

5. Прибыльность - средняя или высокая.

Компании этого кластера предлагают наилучшее решение или технологию для определенного модуля (компонента рыночного продукта). В данной работе мы отделяем рыночные ниши от модулей. Если первые выделяются по принципу сегментации рынка на отдельные группы по характеристикам покупателей, то модули - это компоненты целостного продукта, предлагаемого на рынке. При этом модули необязательно являются комплектующими или полуфабрикатами. Это могут быть дополняющие основной продукт товары или услуги, которые, однако, *не несут в себе основную ценность* этого продукта. Модулем будет являться, например, дополняющая услуга мобильной передачи данных для ноутбука (отсутствие этой услуги не мешает использовать большинство функций ноутбука) или система управления базами данных для корпоративной информационной системы (ИС) (здесь основную клиентскую ценность несет именно ИС, настроенная под нужды клиента, а не поддерживающая ее стандартная СУБД).

Важнейшая характеристика модульных лидеров - монопродуктовость. За очень редким исключением выручка компаний этого кластера более чем на 75-80% формируется за счет одного продукта. При этом компания может иметь в своем ассортименте десятки и даже сотни позиций, однако большинство из них основано на одной и той же технологии и используется сходным образом.

В случае, если рынок или сегмент определенного модуля достаточно широк, модульный игрок может быть достаточно крупной компанией, сохраняя тем не менее высокую прибыльность. Вместе с тем большинство компаний этого кластера относится к средним, крупных среди них мало, поскольку компания, стратегия которой подразумевает постоянный рост, рано или поздно бывает вынуждена расширять свою деятельность, предлагая новые продукты помимо первоначального модуля. В свою очередь, как показал наш анализ эмпирических данных, малым компаниям достаточно сложно (в первую очередь по финансовым причинам) сохранять модульное лидерство, поскольку оно предполагает достаточно существенные вложения в НИОКР. Однако при этом целому ряду малых компаний все же удается сохранять модульное лидерство в стратегической перспективе.

С этой точки зрения объясним тот факт, что компа-

нии данного кластера редко выступают в роли лицензиара или дирижера: будучи поставщиками продукта-модуля, они сосредоточивают внутри себя полный цикл инноваций и производства. Стать лицензиарами им часто мешает невозможность в этом случае удерживать свою конкурентную позицию перед получателями лицензии, а дирижером - целостность продукта. Однако то, что и лицензиары, и дирижеры среди компаний данного кластера все же встречаются, говорит о том, что выбор в пользу интегратора не предопределен, а является фактором, требующим дополнительного анализа.

В то же время компаний, использующих какие-либо иные источники ИК, кроме защиты патентами, в выборке не оказалось. Это позволяет сделать вывод, что секрет, как и открытие, меньше подходит для модульных лидеров. Препятствием для использования секретов, по-видимому, является модульный характер продукта, встраиваемого в продукт более высокого уровня. Чтобы обеспечить совместимость между этими продуктами (оба из которых являются наукоемкими), компании вынуждены в той или иной мере публиковать интерфейсы и технические параметры своих продукта, что, как правило, приводит к невозможности обеспечить секретность технологии, особенно в долгосрочном периоде. В свою очередь полное открытие ИК несет в себе достаточно высокие риски, зачастую неприемлемые для монопродуктовой компании, каковой в большинстве случаев является модульный лидер.

Кластер 4. Технологические лидеры

Типичные представители: Apple, HP, Amdocs, Genentech, NVIDIA, Nycomed, Nokia, GlaxoSmithKline, BayerScheringPharma, Seagate, Merck&Co, Sanofi-Aventis, Лаборатория Касперского.

Особенности компаний этой группы:

1. В основном средние и крупные компании, малые встречаются редко.
 2. Основной тип позиционирования на этапах инновационного процесса - интегратор (чаще всего), реже дирижер и лицензиар.
 3. Превалирующий источник интеллектуального капитала - защита патентами, другие встречаются крайне редко.
 4. Основные экономические модели:
 - а) циклическое создание продуктов, в течение определенного времени несущих уникальную клиентскую ценность;
 - б) высокие объемы продаж небольшого количества продуктов с высокими фиксированными и низкими предельными издержками.
 5. Прибыльность различная - от низкой до высокой.
- Компании - представители данного кластера являются создателями и обладателями лучших технологий. Однако как показывает наш анализ, не всегда эти

компании имеют оптимальные бизнес-модели, что не позволяет части из них иметь высокую прибыль и рентабельность. Этот факт подтверждает поставленную нами ранее гипотезу об отсутствии прямой связи между качеством принадлежащей компании технологии и объемом ее выручки и прибыли. О том же пишет Г. Чесбро в своей статье: даже самая лучшая технология для коммерческого успеха требует верной бизнес-модели [6].

Важно заметить также, что многие из компаний данного кластера, несмотря на обладание лучшей технологией, не являются лидерами по параметрам выручки и прибыли даже на своем рынке. Так, NVIDIA на рынке графических адаптеров существенно уступает Intel, несмотря на то, что выпускает намного более производительные адаптеры. Аналогично, биофармацевтическая компания Cephalon, несмотря на обладание более эффективными лекарственными средствами в сфере наркологии и неврологии, нежели ее конкуренты, уступает большинству из них по выручке и прибыли.

Малые компании в этой группе встречаются редко (а те, что есть, получают прибыль ниже средней по отрасли), поскольку *разработка и коммерциализация* инновационной технологии в современных условиях требуют достаточно много ресурсов и зачастую недоступны малым компаниям.

Также объяснимо преимущественное позиционирование на этапах инновационного процесса как интегратора. Разработка и актуализация самых передовых по мировым меркам технологий требуют достаточно жесткого контроля над процессом разработки и коммерциализации технологии, который эффективнее всего обеспечивается в рамках модели интегратора. Однако как показывают результаты анализа, не всегда компания - разработчик самой инновационной технологии имеет столь же сильные компетенции в коммерциализации и шире - в построении эффективной бизнес-модели на базе этой технологии. Именно этим объясняется широкий диапазон прибыльностей компаний данного кластера. В ряде случаев компания - технологический лидер может повысить эффективность своего бизнеса и укрепить свою конкурентную стратегическую позицию, полностью или частично отказавшись от работы по модели *интегратор*, перейдя к моделям, в той или иной степени открытым (дирижер или лицензиар).

Защита патентами как преимущественный источник ИК объясняется, с одной стороны, тем, что полностью открытая модель несет в себе определенные риски утечек для компании - технологического лидера, однако компании этого кластера не столь консервативны, чтобы полагаться на защиту при помощи секрета (что в современных условиях все сложнее). Учитывая отраслевую специфику компаний данного

кластера, можно утверждать, что именно они в наибольшей степени могут выиграть от перехода к открытым моделям.

Достаточно узкий набор экономических моделей, используемых компаниями данного кластера, требует дополнительного анализа (который также будет проведен в рамках нашей работы), поскольку можно ожидать, что компании - технологические лидеры могли бы выиграть в прибыльности, общей эффективности своих бизнес-моделей и устойчивости стратегических позиций, используя иные экономические модели.

Основные выводы. Результаты кластерного анализа позволили подтвердить все четыре сформулированные ранее гипотезы, а также получить следующие выводы:

1. Существует тесная взаимосвязь между позиционированием компаний наукоемких отраслей на этапах инновационного процесса (интегратор, дирижер, лицензиар), источниками ИК и моделями прибыли. Данная взаимосвязь в полной мере справедлива не только для зарубежных, но и для российских компаний;

2. В наукоемких отраслях применение типовых концепций использования интеллектуальной собственности снижает потенциал прибыльности, уменьшает конкурентные преимущества и ведет к стратегическому распаду бизнес-модели;

3. Необходимое условие получения высокой прибыльности - не технологическое лидерство, а правильно сформированная бизнес-модель, формирование сетевой структуры создания ценности и использование открытых (в различных формах) моделей создания ценности для формирования экосистемы;

4. Небольшие компании - технологические лидеры

способны иметь высокую рентабельность за счет нишевой модели ценности;

5. На определенном этапе роста компаниям новой экономики для сохранения прибыльности необходимо во все большей степени использовать открытые инновации и бизнес-модели;

6. Возможность избежать стратегического распада лежит не столько в новых технологиях, сколько в новых бизнес-моделях.

Литература

1. Сливотски А., Моррисон Д. Маркетинг со скоростью мысли. М.: Эксмо, 2003. С. 35-108, 202-260.
2. Чесбро Г. Открытые инновации. Создание прибыльных технологий. М.: Поколение, 2007. - 336 с.
3. Чесбро Г. Открытые бизнес-модели. IP-менеджмент. М.: Поколение, 2008. - 352 с.
4. Чесбро Г. Логика открытых инноваций: новый подход к управлению интеллектуальной собственностью // Российский журнал менеджмента. 2004. № 2.
5. Эндрю Д., Сиркин Г. Возврат на инновации: Практик. рук. по управлению инновациями в бизнесе / Пер. с англ. Минск: ГревцовПаблшер, 2008. С. 115-189.
6. Chesbrough H. Business Model Innovation: Opportunities and Barriers//Long Range Planning, 2009, doi:10.1016/j.lrp.2009.07.010.
7. Granstrand O. The Economics and Management of Intellectual Property, Cheltenham, Edward Elgar, 1999.
8. Pisano G., Teece D. How to Capture Value from Innovation - Shaping Intellectual Property and Industry Architecture//California Management Review, Vol. 50, Issue 1, Fall 2007, pp. 278-296.
9. Shi Y., Manning T. Understanding Business Models and Business Model Risks//The Journal of Private Equity, Spring 2009, pp. 49-59.
10. Teece D.J. Managing Intellectual Capital. Organizational, Strategic, and Policy Dimensions. OxfordUniversity Press, 2000.
11. VonHippel E. Democratizing Innovation. The MIT Press, 2005, pp. 44-76.

ЭНТРОПИЙНЫЕ МОДЕЛИ В ИССЛЕДОВАНИИ СОЦИАЛЬНЫХ СИСТЕМ

В.Г. Кулаков,

Территориальный орган Росстата по Тверской области,

В.Л. Лазарев,

Санкт-Петербургский государственный университет низкотемпературных и пищевых технологий,

В.А. Федупин,

Ассоциация КСК

Положение в социальной сфере, и в частности демографическая ситуация в стране, требуют организации объективного и качественного мониторинга существующих проблем, выявления господствующих тенденций в их развитии и выработке рекомендаций по их целенаправленному управлению. Для решения та-

ких задач необходимо наличие подходов и методов, позволяющих осуществлять превентивную и высокочувствительную диагностику различных явлений на ранних стадиях их возникновения и развития. Вполне понятно, что любое прогнозирование характеризуется наличием априорной неопределенности в отношении па-

раметров исследуемых явлений. Поэтому используемые методы и модели должны быть максимально адаптированы к этой ситуации и специфике рассматриваемых явлений. В настоящее время разработке и внедрению таких методов и подходов в различных сферах уделяется значительное внимание [1-3].

Весьма перспективным с точки зрения возможностей развития и практического приложения является подход, основанный на использовании энтропийных моделей и методов теории энтропийных потенциалов. Состоятельность применения энтропийного подхода для исследования явлений и процессов в различных сферах показана, например, в работе [4]. Методы теории энтропийных потенциалов позволяют повысить качество диагностики эволюции систем и принятия решений по управлению ими в условиях априорной неопределенности. Это достигается, в частности, за счет учета дестабилизационных свойств законов распределений анализируемых параметров и базовых значений, на фоне которых рассматривается эволюция данного явления. Суть этой теории и основанных на ней методов изложена, например, в работах [5, 6]. Ниже, с использованием методов теории энтропийных потенциалов, произведено исследование рождаемости по возрастным группам матерей Тверской области, сделаны прогнозы демографических тенденций и предложены соответствующие рекомендации.

Исходные данные и постановка задач исследования. Ситуация с рождаемостью в Тверской области находится в сфере внимания региональной администрации, органов государственной статистики, в связи с чем производится сбор и систематизация статистических данных по этому вопросу. Для проведения мониторинга была использована выборка по числу родившихся детей в период с 1980 по 2008 г. включительно. Систематизация числа родившихся осуществлялась по возрасту матерей, который был представлен следующим набором возрастных групп: до 20 лет, 20-24 года, 25-29 лет, 30-34 года, 35-39 лет, 40-44 года, 45-49 лет. Отдельную небольшую группу составили матери, чей возраст не был указан по различным причинам.

Были поставлены следующие задачи:

1. Охарактеризовать тенденцию рождаемости по временному фактору;
2. Осуществить мониторинг рождаемости для различных возрастных групп матерей.

Методы и результаты исследований. Визуальный анализ исходных данных позволяет утверждать, что имеет место тенденция падения рождаемости во времени за рассматриваемый период. Для количественной характеристики этой тенденции в целом и по отдельным возрастным группам была использована модель линейной регрессии, описывающая взаимосвязь между количеством родившихся детей Y в каждом году

x (нулевой отсчет - начало временной шкалы приходится на 1980 г.) в следующем виде:

$$Y(x) = A + B \cdot x, \quad (1)$$

где A и B - коэффициенты регрессии, определяемые на основании исходных данных.

В этом случае знак при коэффициенте B будет характеризовать общую тенденцию развития явления во времени независимо от уровня адекватности модели. Значения величин коэффициентов регрессии для всех возрастных групп и в целом для всех родившихся за указанный период времени приведены в таблице 1. Так как при проведении расчетов использованы официальные данные, не содержащие случайных составляющих, то отпадает необходимость проверки значимости этих коэффициентов. Отрицательные значения коэффициентов B во всех случаях свидетельствуют о тенденции падения рождаемости во времени. Исключение составляет весьма малочисленная группа матерей, возраст которых не указан. Число родившихся детей в этой группе составляет менее 0,04% от общего числа родившихся. Такая тенденция объясняется своеобразной «модой» не указывать возраст, возникшей на рубеже веков. В советское время такое явление практически не имело места. Поэтому можно полагать, что данное исключение подтверждает общее правило.

Имея модель рождаемости (1) и численные значения коэффициентов регрессии (см. таблицу 1), можно осуществить прогнозирование рождаемости как для отдельных возрастных групп матерей, так и в целом по Тверской области на несколько лет вперед. Для получения более длительных прогнозов необходима корректировка (перерасчет) коэффициентов модели на основании дополнительных статистических данных по рождаемости за последние, не учтенные года.

Таблица 1

Коэффициенты регрессии моделей рождаемости в период 1980-2008 гг.

Возрастные группы, лет	Коэффициенты регрессии моделей рождаемости	
	A	B
Все родившиеся	22660	-442,897
До 20	2367	-30,233
20-24	9765	-221,243
25-29	6441	-135,569
30-34	3096	-53,061
35-39	825,579	-2,586
40-44	155,903	-0,688
45-49	8	-0,101
Возраст не указан	-1,903	0,584

Для решения второй из поставленных задач применялись методы статистических оценок и теории энтропийных потенциалов.

Для проведения анализа использованы следующие основные показатели.

1. Среднее значение или математическое ожидание родившихся по каждой возрастной группе в период с 1980 по 2008 г. В данном случае этот показатель позволяет оценить «вклад» каждой группы в общий результат, характеризуемый средним значением родившихся за тот же период.

2. Величина комплексного энтропийного потенциала (КЭП), характеризующая уровень неопределенности, непредсказуемости демографических процессов (в смысле прироста населения) в каждой возрастной группе, которая определяется из следующего выражения:

$$L_{\Delta} = \frac{K_e \sigma}{X_n} = \frac{\Delta_e}{X_n}, \quad (2)$$

где K_e - энтропийный коэффициент, характеризующий дестабилизирующие свойства закона распределения параметра, неопределенность в предсказуемости его значений в зависимости от вида этого закона;

σ - величина среднего квадратического отклонения параметра (СКО);

X_n - базовое значение, на фоне которого рассматривается состояние неопределенности.

Δ_e - энтропийный потенциал параметра.

Чем меньше значение величины КЭП, тем более стабильной является тенденция по рождаемости, и наоборот. Следует иметь в виду, что увеличение значения этой величины может быть обусловлено не только провалами, но и «взрывами» рождаемости, например в случае демографического роста. Такие явления и тенденции в

общем случае однозначно диагностируются с помощью соответствующих трендов. Очевидно, что в рассматриваемой ситуации говорить о «взрыве» рождаемости не приходится.

Более подробно об используемых показателях изложено в работах [5, 6]. Новизна и состоятельность применения изложенного подхода для исследования различных систем подтверждена двумя патентами на изобретение [7, 8].

На основании тех же статистических данных по рождаемости был произведен анализ с целью выявления существующих демографических тенденций. Для обеспечения идентичности условий исследований для всех возрастных групп моделирование плотностей вероятностей рождаемости осуществлялось путем разбиения соответствующих диапазонов рождаемости на шесть равных интервалов. Исключение сделано для последнего столбца таблицы, где возраст не указан. Для этой группы, в силу «слабой» представительности выборки, построение гистограммы и расчеты энтропийных характеристик оказались возможными только при разбиении на четыре интервала. При расчетах значений величин комплексного энтропийного потенциала (КЭП)- L_{Δ} в качестве базовых значений для каждой группы использовались величины математических ожиданий рождаемости - m_x . Также во всех случаях вводились поправки на смещение получаемых оценок от относительно малой представительности выборки. Результаты анализа представлены в таблице 2.

Таблица 2

Характеристики рождаемости по возрастным группам матерей Тверской области

Возрастные группы, лет	Среднее значение m_x (человек)	СКО σ_x (человек)	Энтропийный коэффициент K_e	Энтропийный потенциал Δ_e (человек)	КЭП L_{Δ}
Все родившиеся	16460	4632	1,05	4863	0,295
До 20	1944	353	1,96	690	0,355
20-24	6668	2089	0,96	1996	0,299
25-29	4543	1630	1,28	2086	0,459
30-34	2353	777	1,56	1214	0,516
35-39	790	247	1,72	426	0,539
40-44	146	31	1,86	58	0,394
45-49	6,6	3,5	1,86	6,5	0,99
Возраст не указан	6,3	9	1,32	11,8	1,88

На основании полученных результатов можно сделать следующие выводы.

Основной вклад в рождаемость по Тверской области вносит возрастная группа матерей в возрасте 20-24 года. Это видно по величине математического ожидания родившихся, которая значительно выше аналогичных показателей других возрастных групп. Эта же возрастная группа (по величине КЭП) характеризуется наибольшей стабильностью, невзирая на все кризисы и социальные потрясения. Значение энтропийного ко-

эффициента минимально и практически совпадает с базовым значением этого показателя по всему демографическому процессу рождаемости за рассматриваемый период. Также наибольший вклад в спад рождаемости в области вносит эта же возрастная группа. Это видно из данных таблицы 1: величина коэффициента B свидетельствует о том, что 50% общего спада рождаемости определяется этой возрастной группой. Образно выражаясь, можно сказать, что эта группа является «локомотивом» рождаемости в области.

Второе и третье места по рождаемости занимают группы матерей возрастов 25-29 и 30-34 лет соответственно. Можно предположить, что эти группы относительно многочисленны. Их вклад в спад рождаемости (см. таблицу 1) также соответствует второму и третьему местам. Однако для этих двух групп характерно нарастание нестабильности рождаемости (по величине КЭП). По-видимому, это обстоятельство может быть объяснено тем, что на принятие решения о рождении ребенка для женщин этих возрастных групп большее влияние оказывают социально-экономические факторы, которые существенно изменялись в рассматриваемом временном интервале.

Большой предсказуемостью рождаемости по отношению к двум вышерассмотренным группам является группа матерей возрастов до 20 лет, хотя по внесенному вкладу в рождаемость в области она находится на четвертом месте. Объяснения этого обстоятельства могут быть различными, связанными с возрастными особенностями принятия решений, спецификой данного региона и др.

Используя полученные результаты мониторинга ситуации с рождаемостью в Тверской области, можно разрабатывать адресные меры социальной поддержки возрастных групп матерей, планирования и развития со-

ответствующей инфраструктуры (образовательных учреждений, учреждений здравоохранения и др.), для контроля и управления демографическими процессами.

Литература

1. Нечеткие множества в моделях управления и искусственного интеллекта / Под ред. Д.А. Поспелова. - М.: Наука, 1986. - 312 с.
2. Прикладные нечеткие системы: Пер. с япон. / Под ред. Т. Терао, К. Асаи, М. Сугено.- М.: Мир, 1993. - 368 с.
3. Stable adaptive control and estimation for nonlinear systems: neural and fuzzy approximation techniques / Jeffrey T. Spooner. New York, NY: Wiley-Nescience, 2002. - 545 p.
4. **Прангишвили И.В.** Энтропийные и другие системные закономерности: Вопросы управления сложными системами. - М.: Наука, 2003. - 428 с.
5. **Лазарев В.Л.** Энтропийный подход к организации мониторинга и управления // Изв. РАН. Теория и системы управления. 2005. № 6. С. 61-68.
6. **Лазарев В.Л.** Исследование систем на основе энтропийных и информационных характеристик // ЖТФ. РАН, 2010. № 2. С. 1-7.
7. Патент № 2296356 RU, МПК G05B13/00. Способ контроля и управления динамической системой / Лазарев В.Л. (RU), С1, 2007.03.27.
8. Патент № 2334262 RU, МПК G05B13/00. Способ контроля и управления состоянием неопределенности системы / Лазарев В.Л. (RU), С1, 2008.09.20.

ПРИМЕНЕНИЕ НЕПАРАМЕТРИЧЕСКИХ МЕТОДОВ ИЗУЧЕНИЯ ВЗАИМОСВЯЗИ В ТЕХНОЛОГИИ ОЦЕНКИ ПЕРСОНАЛА

Е.Н. Гришина, канд. экон. наук,

И.П. Лаптева, канд. экон. наук,

Л.Н. Трусова, канд. экон. наук,

Вятская государственная сельскохозяйственная академия

В статистической практике встречаются задачи, в которых необходимо измерять связь между качественными признаками, и здесь параметрические методы в обычном виде неприменимы. В этом случае используются методы, которые не связаны с разложением совокупностей результативного и факторных признаков на отдельные элементы, с расчетами параметров распределения и обобщающих показателей. Здесь речь идет только об установлении наличия связи и измерении ее тесноты. Такие методы получили название *непараметрических*.

Рассмотрим применение непараметрических методов в технологии оценки персонала.

Существуют две мотивационные тенденции, присущие в разной степени всем людям - это стремление к успеху и стремление избежать неудачи, при-

чем у одних преобладает стремление к успеху, к высоким результатам деятельности, у других доминирует стремление избежать неудачи. Знания о мотивах личности и их иерархии помогают избежать ошибок в процессе отбора кадров, в прогнозировании активности человека на рабочем месте. Наиболее важной при этом представляется оценка мотивации к успеху.

Для оценки мотивации к успеху разработан ряд психодиагностических методик (тестов, вопросников). Одним из наиболее известных является тест «Мотивация к успеху» Т. Элерса. Тест состоит из 41 вопроса, на каждый из которых нужно ответить «Да» или «Нет».

В начале 2009 г. авторами было проведено исследование мотивов работников разных сфер деятельно-

сти и безработных. Помимо анкет, использовались также методика Элерса, тест на склонность к риску Шуберта и 16 PF Кэттелла. Всего в исследовании приняло участие 125 человек.

По методике Элерса было опрошено 112 респондентов. В группе респондентов 56% (63 человека) женщин и 44% (49 человек) мужчин. Руководители предприятий составляют 22% (25 человек), частные предприниматели - 26% (29 человек), служащие - 45% (50 человек), безработные - 7% (8 человек).

Результаты исследования мотива достижения успеха распределились следующим образом:

- от 0 до 10 баллов (низкий уровень мотивации достижения) - 6 человек;
- от 11 до 16 баллов (средний уровень) - 41 человек;
- от 17 до 20 баллов (умеренно высокий уровень) - 44 человека;
- от 21 балла и выше (слишком высокий уровень мотивации достижения) - 21 человек (см. таблицу 1).

Таблица 1

Группировка мужчин и женщин по отношению к мотивации успеха

	Уровень мотивации успеха					Частности				
	низкий	средний	умеренно высокий	слишком высокий	итого	низкий	средний	умеренно высокий	слишком высокий	итого
Женщины, человек	6	31	20	6	63	0,095	0,492	0,317	0,095	1
Мужчины, человек	0	10	24	15	49	0	0,204	0,49	0,306	1
Всего	6	41	44	21	112	0,054	0,366	0,393	0,187	1

Если бы пол не влиял на показатели мотивации успеха, то частности каждой строки были бы одинаковыми. Различия же в показателях свидетельствуют о влиянии факторного признака на распределение по результативному признаку, то есть о наличии связи между факторным и результативным признаками, причем чем больше эти различия, тем в большей мере признаки связаны между собой, тем теснее связь между ними.

Для ориентировочной оценки тесноты связей между полом и мотивацией к успеху рассчитывается коэффициент взаимной сопряженности:

$$C = \sqrt{\frac{\chi^2}{n(m_1 - 1)(m_2 - 1)}}, \quad (1)$$

где n - число единиц совокупности;

m_1 и m_2 - число групп по первому и второму признакам.

χ^2 - показатель абсолютной квадратической сопряженности Пирсона, который вычисляется по формуле:

$$\chi^2 = \sum f \sum \frac{(w_{ij} - w_j)^2}{w_j}, \quad (2)$$

где w_{ij} - частности условного распределения в i -й строке; w_j - частности безусловного распределения; j - номер столбца. Если признаки независимы, то $w_{ij} = w_j$, тогда $\chi^2 = 0$ и $C = 0$. В случае функциональной связи коэффициент взаимной сопряженности равен единице.

В нашем случае для группировки респондентов по полу $\chi^2 = 19,404$ и $C = 0,240$.

Для измерения связи между варьированием двух атрибутивных признаков, когда это варьирование образует три или более группы, используются также коэффициенты взаимной сопряженности Пирсона (C_n) и А.А. Чупрова (K_q):

$$C_n = \sqrt{\frac{\chi^2}{\chi^2 + N}} \quad \text{или} \quad C_n = \sqrt{\frac{\varphi^2}{\varphi^2 + 1}}, \quad (3)$$

$$K_q = \sqrt{\frac{\varphi^2}{\sqrt{(m_1 - 1)(m_2 - 2)}}}, \quad (4)$$

где m - число групп по каждому признаку.

φ^2 - показатель взаимной сопряженности, который вычисляется по формуле:

$$\varphi^2 = \frac{\chi^2}{N}, \quad (5)$$

где N - число единиц совокупности.

Для группировки респондентов по полу $\varphi^2 = 0,173$, $C_n = 0,384$ и $K_q = 0,316$.

Оценка значимости коэффициента взаимной сопряженности проводится при помощи критических значений χ^2 . При числе степеней свободы $K = 3$ $\chi^2 = 7,81$ при уровне значимости $\alpha = 0,05$ и $\chi^2 = 11,34$ при уровне значимости $\alpha = 0,01$. Фактическое же значение $\chi^2 = 19,404$, то есть больше критического значения при обоих уровнях значимости. Таким образом, связь между полом и мотивацией к успеху существенна. Полученные показатели свидетельствуют о том, что зависимость между полом и мотивацией к успеху средняя. Аналогичным образом рассмотрим связь между видом занятости и мотивацией к успеху (см. таблицу 2).

Для данной группировки $\chi^2 = 116,864$ и коэффициент взаимной сопряженности $C = 0,340$. При числе степеней свободы $K = 9$ $\chi^2 = 16,92$ при уровне значимости $\alpha = 0,05$ и $\chi^2 = 21,67$ при уровне значимости $\alpha = 0,01$. Связь между видом занятости и мотивацией к успеху существенна. Коэффициенты Пирсона и Чупрова для данной группировки составили: $\varphi^2 = 1,043$, $C_n = 0,715$ и $K_q = 0,59$. Полученные показатели свидетельствуют о том, что зависимость между видом занятости и мотивацией к успеху выше среднего.

Для выявления психологических особенностей респондентов с высокими и низкими показателями мотивации к успеху был проведен анализ взаимосвязи меж-

ду показателями мотивации к успеху и результатами тестов Шуберта и Кэттелла.

Таблица 2

Группировка респондентов различных видов занятости по отношению к мотивации к успеху

	Уровень мотивации к успеху					Частности				
	низкий	средний	умеренно высокий	слишком высокий	итого	низкий	средний	умеренно высокий	слишком высокий	итого
Руководители	0	2	12	11	25	0	0,08	0,48	0,44	1
Служащие	1	36	13	0	50	0	0,204	0,49	0,306	1
Частные пред- приниматели	0	0	19	10	29	0	0	0,655	0,345	1
Безработные	5	3	0	0	8	0,625	0,375	0	0	1
Всего	6	41	44	21	112	0,054	0,366	0,393	0,187	1

В нашем исследовании для изучения интенсивности связи между результатами тестов были использованы коэффициенты корреляции рангов Кендэлла (τ).

Для вычисления τ надо упорядочить ряд рангов переменной X , приведя его к ряду натуральных чисел. Затем рассматривают последовательность рангов переменной Y .

X - фактор E по тесту Кэттелла (стены)

1 2 2 3 3 3 4..... 112

Y - ранг показателей мотивации к успеху

87 21 76 18 1 20 17.....103

По данным рядам вычисляются два показателя - P и Q , характеризующих степень соответствия и несоответствия последовательности рангов переменных X и Y . Для нахождения P подсчитывается, сколько чисел, находящихся справа от каждого из элементов, имеют величину ранга большую, чем ранг рассматриваемого элемента, а для нахождения Q - сколько чисел имеют меньший ранг. При этом величины берутся со знаком минус.

Таблица 3

Коэффициенты корреляции между мотивацией к успеху и психическими особенностями респондентов

	Личностные факторы (тест Кэттелла)			Склонность к риску (тест Шуберта)
	E (доминантность/ /конформность)	Q_2 (самодостаточность/ /зависимость от группы)	O (самоуверенность/ /неуверенность)	
Мотивация к успеху	$\tau = 0,64$	$\tau = 0,62$	$\tau = 0,55$	$\tau = 0,31$

Коэффициент корреляции рангов Кендэлла рассчитывается по формуле:

$$\tau = \frac{2S}{n(n-1)}, \quad (6)$$

где $S = P + Q$, n - число единиц совокупности.

Результаты вычислений представлены в таблице 3.

Исследование показало, что высокие показатели мотивации к успеху характерны для личностей, склонных к лидированию, авторитарных, независимых, упорных в достижении цели, умеренно склонных к риску, с высокой самооценкой. Низкие показатели мотивации достижения характерны для личностей, отличающихся мягкостью характера, чертами зависимости и пессимизмом.

Таким образом, применение непараметрических методов позволило установить наличие связи и измерить ее тесноту.

Литература

1. Батракова Л.Г. Теория статистики: Учеб. пособие. - М.: КНОРУС, 2009. - 528 с.
2. Кристиани А. Мотивация успеха. 111 советов для достижения лучших результатов. - М.: Интерэксперт, 2005. - 256 с.
3. Резник А.Д. Книга для тех, кто не любит статистику, но вынужден ею пользоваться. Непараметрическая статистика в примерах, упражнениях и рисунках. - М.: РЕЧЬ, 2008. - 265 с.
4. Статистика: Учебник / Л.П. Харченко, В.Г. Ионин, В.В. Глинский и др.: Под ред. канд. экон. наук, проф. В.Г. Ионина. 3-е изд., перераб. и доп. - М.: ИНФРА-М, 2008. - 445 с.
5. <http://ru.wikipedia.org/wiki/Мотивация>.