

ЭКОНОМИКА

Бекарева Светлана Викторовна

канд. экон. наук, доцент

Индык Анастасия Андреевна

аспирант

ФГАОУ ВО «Новосибирский национальный
исследовательский государственный университет»
г. Новосибирск, Новосибирская область

ОЦЕНКА КОНКУРЕНТОСПОСОБНОСТИ ФАРМАЦЕВТИЧЕСКОЙ ОТРАСЛИ В РЕГИОНАХ СФО МЕТОДОМ СТАТИСТИЧЕСКОГО РАСПОЗНАВАНИЯ ОБРАЗОВ

***Аннотация:** в данной работе предложен алгоритм исследования фармацевтической отрасли в разрезе регионов с использованием метода статистического распознавания образов. Результатом его применения является выявление набора факторов, оптимального из имеющихся, которые определяют уровень конкурентоспособности региональной фарминдустрии. На основе предложенного показателя конкурентоспособности показана динамика состояния фармацевтической отрасли в Сибирском Федеральном округе за период 2010–2013 гг.*

***Ключевые слова:** оценка конкурентоспособности, фармацевтическая отрасль, конкурентоспособность, метод статистического распознавания.*

Введение.

Фармацевтическая отрасль Российской Федерации – одна из отраслей, развитие которой необходимо для решения текущих экономических задач, стоящих перед нашей страной. В настоящее время конкурентоспособность российской фармацевтической отрасли по сравнению с фармацевтическими отраслями ряда ведущих стран мира очень низка [1]. Поддержание национальных производителей и стимулирование инноваций в фармацевтической промышленности явля-

ется задачей государственной важности, о чем свидетельствует принятие программ и стратегий развития фармацевтики в России, однако реализация поставленных в них задач в настоящее время не очень успешна [2]. Степень развития фармацевтической промышленности и ее потенциал в Российской Федерации в региональном разрезе не однородны, поэтому, с нашей точки зрения, целесообразным является анализ отдельных регионов и выработка рекомендаций по их дальнейшему развитию.

Существует множество подходов к оценке конкурентоспособности различных экономических субъектов, а также уровней экономических объединений. Для оценки конкурентоспособности отрасли экономики на мировом рынке, как правило, используются ряд макроэкономических показателей, а также показателей, связанных со спецификой данной отрасли [1].

Описание метода и использованная статистика

Метод распознавания образов – это статистический метод классификации, в рамках которого реализована возможность оптимизации признакового пространства и решена задача минимизации вероятности ошибки разбиения через его основные параметры: размерность признакового пространства и объемы обучающих и контрольных (решающих) наблюдений. Минимизация вероятности ошибок распознавания достигается целенаправленным изменением указанных параметров [3].

В общем виде модель статистического распознавания образов решает классическую задачу классификации [4]. На входе имеется: n классифицируемых объектов, представленных данными в виде i -объектов и значений P характеризующих i -й объект признаков $x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(P)}$; обучающие выборки

$$X_{j1}, X_{j2}, \dots, X_{jn_j}, \quad j = 1, 2, \dots, k,$$

каждая j -я из которых определяет значения анализируемых признаков $X_{ji} = (x_{ji}^{(1)}, x_{ji}^{(2)}, \dots, x_{ji}^{(P)})^T$ на n_j объектах (т. е. $i = 1, 2, \dots, n_j$), о которых априори известно, что все они принадлежат j -му классу, причем число k различных выборок равно общему числу всех возможных классов (так что каждый класс представлен

своей порцией выборочных данных). С целью построения эталонных описаний обучающих выборок на первом этапе осуществляется обучение, на втором этапе с эталонными описаниями сравниваются исследуемые наблюдения и делается вывод об их состоянии [5].

В общем случае непараметрической априорной неопределенности обучение заключается в статистическом оценивании условных многомерных плотностей вероятностей (функций правдоподобия), взятых значений выбранных P показателей при условии их принадлежности к определенному классу j .

Однако на практике оценивание многомерных плотностей вероятностей (функций правдоподобия) связано с серьезными вычислительными трудностями. Это делает целесообразным сведение непараметрической априорной неопределенности к параметрической, путем расширения области применения центральной предельной теоремы (ЦПТ) теории вероятностей на конечные объемы q выборок ($q \geq 5$ вместо $q \xrightarrow{d} \infty$) [3].

При условии, что класс j идентифицируется P -мерной нормальной генеральной совокупностью с вектором средних значений a_j и ковариационной матрицей M (общей для всех классов), $j = 1, 2, \dots, k$, в качестве функций $f_j(X)$ используют P -мерные нормальные плотности $\varphi(X; a_j; M)$, где векторы средних значений $a_j = (a_j^{(1)}, a_j^{(2)}, \dots, a_j^{(P)})^T$ и ковариационная матрица $MM = (\sigma_{lq})$, $l, q = 1, 2, \dots, P$, получены с помощью расчетов по обучающим выборкам. Они имеют вид

$$a_j^{(l)} = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ji}^{(l)}, \quad l = 1, 2, \dots, p; \quad j = 1, 2, \dots, k;$$

$$\sigma_{lq} = \frac{1}{n_{o\delta} - k} \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji}^{(l)} - a_j^{(l)}) (x_{ji}^{(q)} - a_j^{(q)}),$$

$$l, q = 1, 2, \dots, p; \quad n_{o\delta} = n_1 + n_2 + \dots + n_k.$$

Метод кластеризации наблюдений имеет весьма простой вид в случае $k=2$. Таким образом, наблюдение X_v относится к классу с номером j_0 тогда и только тогда, когда

$$\frac{f_{j_0}(X_v)}{f_j(X_v)} \geq \frac{\pi_j}{\pi_{j_0}} \text{ для всех } j = 1, 2, \dots, k$$

или то же,

$$\ln \frac{f_{j_0}(X_v)}{f_j(X_v)} \geq \ln \frac{\pi_j}{\pi_{j_0}} \text{ для всех } j = 1, 2, \dots, k$$

В случае нормальных плотностей $f_j(X)$ данное соотношение эквивалентно:

$$\ln(L) = \left[X_v - \frac{1}{2}(a_{j_0} + a_j) \right]^T M^{-1}(a_{j_0} - a_j) \geq \ln \frac{\pi_j}{\pi_{j_0}} \text{ для всех } j = 1, 2, \dots, k$$

Данное соотношение задает вид логарифмической функции в задаче различения нормальных классов при постоянных значениях потерь от неправильной классификации. Данное правило кластеризации упрощается в случае двух классов ($k=2$) и одинаковых априорных вероятностей ($\pi_1 = \pi_2 = 0,5$). В этом случае наблюдение X_v следует относить к 1-му классу тогда и только тогда когда

$$\left[X_v - \frac{1}{2}(a_1 + a_2) \right]^T M^{-1}(a_1 - a_2) \geq 0,$$

и ко 2-му классу во всех остальных случаях. Легко понять, что при кластеризации на два класса одномерных ($P=1$) нормальных наблюдений решение об отношении наблюдения X_v к одному из двух классов будет определяться знаком произведения $[X_v - (a_1 + a_2)/2](a_1 - a_2)$.

Таким образом, суть метода заключается в том, что исследователь отказывается от оценивания распределений каждого из показателей и сводит задачу к случаю нормального многомерного распределения, используя ЦПТ, которая говорит, что в пределе все стандартизованные показатели с конечной дисперсией имеют нормальное распределение. Далее берутся две гипотезы: первая гипотеза состоит в том, что исследуемый регион будет принадлежать к первому кластеру (преуспевающие регионы), вторая гипотеза – исследуемый регион буде т при-

надлежать ко второму кластеру (неблагополучные регионы), при этом предполагаем, что вектор факторов имеет многомерное нормальное распределение. Затем строится функция правдоподобия и логарифмическая функция правдоподобия $\ln(L)$, значение которой отражает уровень конкурентоспособности региона с точки зрения фармацевтической отрасли.

Теперь пишем в общем виде алгоритм исследования, который мы применили на основе метода статистического распознавания образов конкретно к нашей задаче исследования.

На первом этапе (этап обучения) из общего списка регионов выбираются согласно экономическому состоянию и уровню развитию фармацевтической отрасли m_1 образующих класс S_1 (преуспевающих регионов) и m_2 образующих класс S_2 (неблагополучных регионов).

Согласно факторному анализу берутся P наиболее значимых факторов в динамике за исследуемый период. На основе этих данных вычисляются вектора a_1 и a_2 и ковариационные матрицы M_1 и M_2 показателей лучших и худших регионов соответственно. Здесь вектора a_1 и a_2 есть среднее значение каждого из факторов для m_1 преуспевающих и m_2 неблагополучных регионов соответственно. Далее вычисляются общая и обратная ковариационные матрицы M и M^{-1} где

$$M = \frac{1}{m_1 + m_2 - 2} (m_1 M_1 + m_2 M_2)$$

Для оценки достоверности диагностики $D = 1 - \alpha = 1 - \beta$ необходимо вычислить вероятность ошибок диагностики по формуле [3, с. 333]:

$$d_p^2 = (a_1 - a_2)^T M^{-1} (a_1 - a_2)$$

и параметры σ_1 и σ_2 , в свою очередь вычисляемые через объемы обучающих m_1 и m_2 и контрольной (решающей) n выборок

$$\sigma_1^2 = \frac{1}{m_1} + \frac{1}{m_2}; \quad \sigma_2^2 = \frac{1}{m_1} + \frac{1}{m_2} + \frac{4}{n}.$$

С целью минимизации ошибок диагностики $\alpha = \beta$ необходимо оптимизировать признаковое пространство (количество факторов) P путем последовательного сопоставления отношения Махаланобиса d_p^2 / d_{p-1}^2 с корнем квадратным из

последнего выбранного по нарастающей числа P признаков. Расстоянием Махаланобиса называют расстояние между центрами кластеров [3]. И как только указанное отношение станет меньше \sqrt{p}

$$d_p^2 / d_{p-1}^2 < \sqrt{P}$$

размерность признакового пространства становится оптимальным (Фомин, п. 5.3.).

При найденном таким образом оптимальном значении $p_{\text{опт}}$ вероятность ошибки всегда минимальна:

$$\alpha_p = \beta_p = F\left(-\frac{d_p}{\sigma_2}\right) + \left(\frac{\sigma_2}{\sqrt{2\pi}d_p}\right) \exp\left(-\frac{d_p^2}{2\sigma_2^2}\right) \left[\left(\frac{\sigma_2^2}{\sigma_2^2 - \sigma_1^2}\right)^{\frac{p-1}{2}} - 1\right]$$

На втором этапе выбираются основные исследуемые регионы и выписываются значения всех P показателей деятельности каждого региона, соответствующие исследуемому интервалу времени.

Согласно положениям теории статистических решений, оптимальное решающее правило основано на формировании отношения правдоподобия (или его натурального логарифма), представляющего собой отношение функций правдоподобия преуспевающих и неблагополучных регионов, где в качестве аргумента функции правдоподобия представляются данные исследуемого региона. Логарифм отношения правдоподобия представляет собой произведение полученной на этапе обучения вектора-строки $\frac{1}{2} (a_1 - a_2)^T * M^{-1}$ и вектора-столбца $[2X_{it} - (a_1 + a_2)]$ значений показателей исследуемых регионов для каждого рассматриваемого момента времени.

$$\ln(L) = \frac{1}{2} (a_1 - a_2)^T M^{-1} [2X_{it} - (a_1 + a_2)]$$

Таким образом, мы получили логарифм функции правдоподобия $\ln(L)$, который в нашем исследовании будет являться показателем уровня конкурентоспособности фармацевтической отрасли для конкретного региона. При значении данного показателя выше нуля, регион относится к категории преуспевающих,

ниже нуля – к неблагоприятным. О степени преуспевания или неблагоприятия региона свидетельствует мера удаленности значения от нуля.

Все расчеты, проведенные для практической части исследования, опираются на статистические данные по регионам с официального сайта статистики РФ Росстат, официальных сайтов Министерства финансов и Министерства здравоохранения, а также данных, полученных в Новосибирском статистическом управлении. Данные для каждого региона взяты в динамике за четыре года (2010–2013 гг.).

Для проведения данного исследования из 85 регионов России, по которым доступна официальная статистика, в конечном итоге были отобраны 32. В оставшихся регионах фармацевтическая отрасль либо развита очень слабо, либо отсутствует совсем. Также в процессе исследования из списка были исключены заведомо благополучные регионы, для получения относительно однородных объектов в выборке. Выводы исследования сконцентрированы на ситуации в регионах Сибирского федерального округа.

Для целей исследования были определены ряд факторов для оценки уровня конкурентоспособности фармацевтической отрасли в разрезе регионов. При отборе факторов мы учитывали степень полноты статистических данных, так как по ряду показателей, которые целесообразно включить в наше исследование, информация по регионам либо отсутствует, либо конфиденциальна.

Таким образом, из имеющихся данных были отобраны одиннадцать показателей по всем анализируемым регионам за весь рассматриваемый период времени:

- ВРП на душу населения (руб.);
- потребительские расходы в среднем на душу населения за месяц (руб.);
- денежные доходы в среднем на душу населения в месяц (руб.);
- число организаций, выполняющих научные исследования и разработки;
- внутренние расходы региона на научные исследования и разработки (млн. руб.);
- общее число выданных патентов (ед.);

- количество человек, занятых в фармацевтической отрасли региона (тыс. чел.);
- численность персонала, занятого научными исследованиями и разработками (чел.);
- расходы на здравоохранение (млн руб.);
- иностранные инвестиции, получаемые регионом (тыс. долл. США);
- совокупный объем производства лекарственных средств регионом (млрд руб.).

Полученные результаты.

Вспомогательным методом явился факторный анализ, в результате которого были отобраны факторы, имеющие наибольшую корреляцию с главными компонентами:

- количество человек, занятых в фармацевтической отрасли региона (тыс. чел.);
- общее число выданных патентов (ед.);
- расходы на здравоохранение (млн руб.);
- ВРП на душу населения (руб.);
- совокупный объем производства лекарственных средств регионом (млрд руб.).

Первый этап исследования – выявление оптимального числа факторов.

На данном этапе (этап обучения) из общего списка регионов на основе анализа региональной статистики мы выделили $m_1 = 4$ образующих класс S_1 (преуспевающих регионов) и $m_2 = 4$ образующих класс S_2 (неблагополучных регионов). Анализ экономического состояния регионов показал, что наиболее привлекательными регионами с точки зрения развития фарминдустрии являются следующие: Курская область, Нижегородская область, Республика Башкортостан, Республика Татарстан. Неблагополучными регионами являются: Астраханская область, Вологодская область, Республика Карелия, Смоленская область.

Пять факторов, выделенных с помощью факторного анализа, были поочередно добавлены в модель. Первые два фактора: количество человек, занятых в

фармацевтической отрасли региона, и общее число выданных патентов в регионе – не привели, согласно расчетом, к оптимальной модели. Третий добавленный фактор – расходы на здравоохранение – также не привел к итоговому виду. После добавления четвертого фактора – ВРП на душу населения, мы получили, что размерность признакового пространства, равного четырем, оптимальна.

Оптимальность подтверждается тем, что отношение Махаланобиса dd_p^2 / d_{p-1}^2 меньше, чем корень квадратный из последнего выбранного по нарастающей числа P признаков:

$$d_4^2 / d_3^2 = 0,82185 < \sqrt{P} = 2, \text{ при } P = 4.$$

Следовательно, при найденном таким образом оптимальном значении $P_{\text{опт}}$ вероятность ошибки всегда минимальна:

$$1) F\left(-\frac{d_4}{\sigma_2}\right) = 0,99992$$

$$2) \alpha_4 = \beta_4 = 0,04006$$

Второй этап – разделение на кластеры. На данном этапе мы взяли список исследуемых регионов России по Федеральным округам – всего 32 региона, предварительно исключив четыре региона, которые выделяются из общей статистики, а также регионы, использованные на этапе обучения.

Далее мы рассмотрели значения для каждого региона по четырем показателям, определенным на первом этапе исследования, которые соответствуют каждому интервалу времени за 2010–2013 гг.: количество человек, занятых в фармацевтической отрасли региона; общее число выданных патентов; расходы на здравоохранение; ВРП на душу населения. Затем был рассчитан показатель уровня конкурентоспособности фармацевтической отрасли $\ln(L)$ для каждого из рассматриваемых регионов.

Данный показатель – логарифм отношения правдоподобия – представляет собой произведение полученной на этапе обучения вектора-строки $\frac{1}{2} (a_1 - a_2)^T * M^{-1}$ и вектора-столбца $[2X_{it} - (a_1 + a_2)]$ значений показателей исследуемых регионов для каждого рассматриваемого момента времени.

Далее для каждого из 32 регионов в динамике за четыре года мы рассчитываем множитель $[2X_{it} - (a_1 + a_2)]$, где i – регион, $i = \overline{1; 32}$, $t = \overline{2010; 2013}$.

Проиллюстрируем расчеты на примере Новосибирской области (таблица 1):

Таблица 1.

Значения факторов для Новосибирской области за 2010–2013 гг.

Показатели	Ед. измерения	Новосибирская область			
		2010	2011	2012	2013
Количество человек, занятых в фармацевтической отрасли	(тыс. чел.)	44	43,6	43,3	43,2
Общее число выданных патентов	(количество единиц)	202	194	202	228
Расходы на здравоохранение	(млн руб.)	10529,3	17034,9	20052,4	18089,5
ВРП на душу населения	(млн руб.)	181732,7	223623	244441,8	301955,7

Множитель $[2X_{it} - (a_1 + a_2)]$ для данного региона рассчитывается в каждый из рассматриваемых периодов. Логарифм отношения правдоподобия $\ln(L)$, как показатель конкурентоспособности фармацевтической отрасли, представлен в таблице 2.

Таблица 2.

Значения $\ln(L)$ для Новосибирской области за 2010–2013 гг.

	Новосибирская область			
	2010	2011	2012	2013
$2X_{it} - (a_1 + a_2)$	24,4	23,6	23	22,8
	79,75	63,75	79,75	131,75
	–5917,8	7093,4	13128,4	9202,6
	–162441	–171332	–177957	–182321
$\ln(L)$	8,2333	11,2329	13,1785	13,4945

Регион Новосибирской области относится к классу преуспевающих. Также можно отметить, что с течением времени этот показатель увеличивает свое значение. Можно сделать вывод, что Новосибирская область является привлекательным регионом с точки зрения развития фармацевтической отрасли и имеет потенциал для ее развития.

Для остальных регионов расчеты выполнялись аналогичным образом.

Полученные результаты по регионам Сибирского федерального округа представлены в виде графиков на рисунке 1.

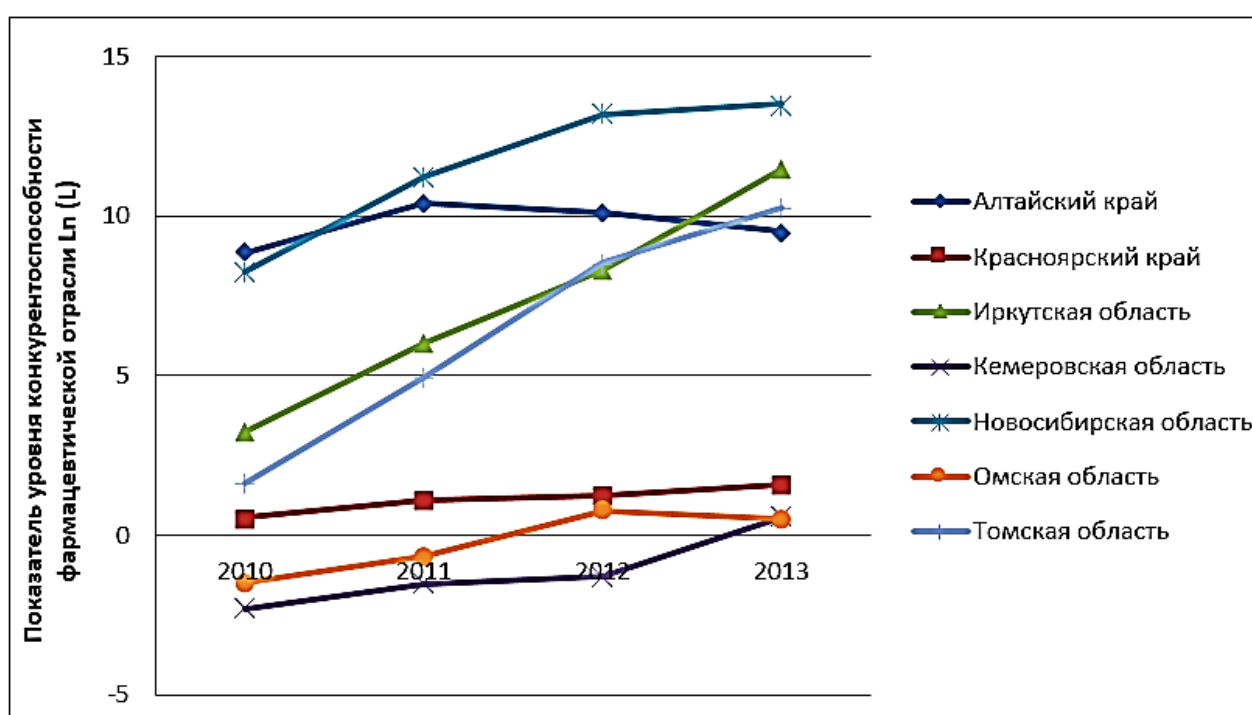


Рис. 1. Значение показателя уровня конкурентоспособности фармацевтической отрасли для регионов Сибирского федерального округа

В Сибири существует ряд фармацевтических кластеров, что обуславливает положение регионов на данном рисунке. Все регионы имеют значение показателя $\ln(L)$ больше нуля. Следует выделить Новосибирскую, Иркутскую, Томскую области и Алтайский край.

Новосибирская область – один из крупнейших научных центров страны, она удерживает лидирующие в России позиции в сфере науки и разработки новых

технологий и материалов уже более 50 лет. Здесь расположено Сибирское отделение Российской академии наук. В области действуют 53 академических института и более 30 отраслевых научно-исследовательских, 38 высших учебных заведений. Концентрация научных кадров в Новосибирске в 2,2 раза превышает общероссийский показатель [6].

Новосибирский инновационный кластер биофармацевтических технологий – это исторически сложившийся и находящийся в постоянном развитии комплекс взаимосвязанных предприятий и организаций, которые обладают передовыми научными, технологическими, образовательными и компетенциями. В кластере работают 130 предприятий и организаций [7]. В Кольцово располагается крупное Федеральное государственное учреждение науки «Государственный научный центр вирусологии и биотехнологии «Вектор».

Перечисленное выше обуславливает то, что самый высокий показатель конкурентоспособности отрасли среди регионов Сибирского федерального округа принадлежит Новосибирской области.

В Томской области также имеется фармацевтический кластер, который характеризуется сильной инновационной составляющей, опирающейся на развитую научную и образовательную базу Томских университетов и существующие проекты кооперации между вузами и компаниями.

Иркутская область имеет ярко выраженный восходящий тренд значений показателя конкурентоспособности. Действительно, сегодня в Иркутской области есть все предпосылки для создания полноценного фармацевтического кластера. Это высокий экономический потенциал, развитые химическая промышленность и энергетика, а также наличие фармацевтических предприятий, включая такого крупного производителя, как «Фармасинтез», и образовательных учреждений. Потенциальными участниками Байкальского биофармкластера могут стать около 30 предприятий и организаций. Среди них – Сибирское отделение РАН и РАМН, производители фармацевтической продукции, Иркутский государственный университет, Ростехнологии (РТ-Биотехпром), Роснано и другие [8].

Алтайский край является одним из ведущих центров развития биотехнологий в России. Особая привлекательность проекта заключается в его направленности на расширение выпуска биофармацевтической продукции высокого качества, востребованной на рынке и основанной на уникальных биоресурсах края. Развитие фармацевтической промышленности, производства биологически активных добавок на основе сырья растительного и животного происхождения обусловлено успешной реализацией в регионе кластерной политики, которая подтвердила оптимальность интеграции ресурсов, научных сил и предприятий, входящих в Алтайский биофармацевтический кластер [9]. Алтайский край имеет высокие значения рассчитанного показателя конкурентоспособности, однако его динамика в последние три года отрицательная.

Заключение.

Метод статистического распознавания образов может быть использован для анализа уровня конкурентоспособности фармацевтической отрасли страны в разрезе регионов. Данный метод позволяет дать сравнительную характеристику регионам на основе оптимального количества факторов, доступных для анализа. Для анализа состояния фармацевтической отрасли был предложен показатель уровня конкурентоспособности, который построен на основе логарифмической функции правдоподобия.

Рассмотренные территориальные единицы Сибирского федерального округа, в настоящее время могут быть охарактеризованы как регионы с положительным значением показателя конкурентоспособности, и, как правило, его возрастающей величиной. Это может свидетельствовать о наличии потенциала в развитии фармацевтической отрасли и целесообразности государственной поддержки региона в развитии данной отрасли промышленности.

Список литературы

1. Бекарев А.А. Подход к оценке конкурентоспособности национальной фармацевтической отрасли / А.А. Бекарев, С.В. Бекарева // Вестник НГУ. Серия: Социально-экономические науки. – 2014. – Т. 14. – Вып. 4. – С. 78–91.

2. Бекарев А.А. Что определяет современное состояние и конкурентоспособность российской фармацевтической отрасли / А.А. Бекарев, С.В. Бекарева // Вестник НГУ. Серия: Социально-экономические науки. – 2015. – Т. 15. – Вып. 1. – С. 23–31.
3. Фомин Я.А. Распознавание образов: теория и применения / Я.А. Фомин. – М.: Фазис, 2010. – 368 с.
4. Журавлев Ю.И. Об алгебраическом подходе к решению задач распознавания и классификации / Ю.И. Журавлев // Проблемы кибернетики. – М.: Наука, 2008. – Вып. 46. – С. 5–68.
5. Айвазян С.А. Теория вероятностей и прикладная статистика / С.А. Айвазян, В.С. Мхитарян // Прикладная статистика. Основы эконометрики: Учебник для вузов. – М.: Юнити-Дана, 2010. – 656 с.
6. Единый информационно-аналитический портал государственной поддержки инновационного развития бизнеса [Электронный ресурс]. – Режим доступа: <http://innovation.gov.ru/node/3563>
7. Российская кластерная обсерватория [Электронный ресурс]. – Режим доступа: http://cluster.hse.ru/clusters/?PAGEN_1=2&set_filter=y&arrFilter_16_1330857165=Y&arrFilter_15_4088188550=Y
8. Широкова И. Фармкластеры как сегмент региональной политики / И. Широкова // Ремедиум. – 2014. – №3. – С. 74–75.
9. Управление Алтайского края по фармацевтической промышленности и биотехнологиям [Электронный ресурс]. – Режим доступа: <http://www.ffprom22.ru/info/analitika/index.html>