

Использование рядов распределения для анализа коммерческих данных

Аннотация. В статье изложен подход к построению рядов распределения для оценки деятельности юридических лиц, которые объединены каким-либо типовым признаком, характеризующим их занятость. Описан алгоритм для анализа значений совокупности однотипных компаний при проведении мониторинга состояния отрасли или рынка, к которому они относятся, основанный на характере распределения данных по ряду предприятий, которые были получены по итогам их работы за отчетный период или являются коммерческими результатами в течение определенных временных интервалов. Предложены направления использования данного подхода при проведении анализа финансовых или коммерческих данных, которые являются однородными для совокупности однотипных организаций.

Ключевые слова: анализ данных, финансовые данные, массивы данных, финансы, прогнозирование, финансовая аналитика, скалярные значения, оценка платежеспособности, комплексная оценка, шкала оценки, оценка сделок, оценка юридических лиц, финансовый анализ, частота, относительная частота, асимметрия значений, нормальное распределение, распределение данных, формула Стерджесса, плотность распределения.

Для цитирования: Рагель Д. Использование рядов распределения для анализа коммерческих данных // Наука и инновации. 2025. №4. С. 43–46.
<https://doi.org/10.29235/1818-9857-2025-4-43-46>



Дмитрий Рагель,
доцент кафедры экономики
Белорусского государственного
университета информатики и
радиоэлектроники, кандидат
экономических наук;
agel@mail.ru

При проведении экономического анализа деятельности субъектов хозяйствования на рынке, как правило, используются экономические коэффициенты, которые помогают подробным образом охарактеризовать различные аспекты функционирования конкретного предприятия. В то же время таких показателей недостаточно для получения объективной картины, касающейся группы однотипных юридических лиц, потому что все-таки они ориентированы на оценку частных явлений и предполагают, как правило, проведение дальнейшего сопоставления полученных значений. Если же говорить о методах оценки, которые могут охарактеризовать совокупность в целом без предварительных частных оценок, то такой инструментарий в экономических аналитических практиках используется достаточно редко. В то же время при проведении анализа нередко требуются оценочные характеристики состояния совокупности субъектов рынка или рассматриваемого направления экономической

деятельности. В этом случае для рассмотрения экономических тенденций, имеющих отношение к некоторому числу организаций, возможно применение математических методов анализа данных, которые позволяют выстраивать ряды и выявлять тенденции, характеризующие совокупное состояние и динамику какой-то отрасли народного хозяйства. Данный тип методов может быть полезен при изучении рынков и отраслей для оценки деятельности совокупности юридических лиц в целях выявления общих тенденций или определения конъюнктурных ограничений, связанных с характером конкуренции, существованием возможных ограничений или влиянием различных факторов на результативность работы компаний. Коэффициентный анализ не дает такой наглядной оценки, он может быть использован как один из этапов подобной процедуры, дополняющий общую картину как частный элемент.

По результатам выборочного исследования 100 схожих по роду деятельности юридических лиц

Исходные данные объемов продаж 100 однотипных компаний по состоянию на 01.09.2023 г., млн руб.

0,86	0,77	0,54	2,22	2,7	2,26	2,996	2,29	2,25	2,98
0,97	4,14	1,77	2,31	2,42	2,07	2,41	6,55	2,29	2,44
0,74	2,11	3,61	2,99	2,64	2,32	2,28	2,25	2,2	2,93
2,97	2,51	2,94	2,1	2,61	2,45	2,44	2,19	2,37	3,44
3,55	2,92	4,1	3,12	2,93	2,55	4,64	3,34	3,12	3,22
2,34	1,97	0,91	4,34	2,92	2,91	5,55	4,44	4,65	5,15
2,71	1,86	0,95	2,9	2,65	2,98	4,36	6,41	6,1	5,16
2,94	2,94	0,76	2,99	2,44	2,39	4,32	5,33	4,33	3,34
1,89	2,14	0,61	1,33	2,5	2,42	3,37	5,59	4,11	4,92
1,45	2,89	1,12	1,56	2,38	2,94	3,02	5,91	4,69	4,2

Таблица 1. Объемы продаж предприятий на одном из рынков потребительских товаров за 8 месяцев 2023 г.

Порядковый номер интервала	Нижняя граница	Верхняя граница
1	0,147	1,326
2	1,326	2,113
3	2,113	2,899
4	2,899	3,685
5	3,685	4,471
6	4,471	5,258
7	5,258	6,044
8	6,044	6,83

Таблица 2. Распределение интервалов для классификации исходных значений

автором получены данные об объеме их продаж по состоянию на 01.09.2023 г. (табл. 1). Для того чтобы оперировать этими сведениями в рамках аналитических процедур и сделать первоначальные выводы, характеризующие динамику рассматриваемой отрасли или рынка, необходимо провести классификацию в целях дальнейшего изучения групп таких субъектов с общими количественными признаками. В качестве наиболее простого и очевидного решения можно использовать построение интервального вариационного ряда с применением формулы Стерджесса как инструмента классификации полученных значений.

Для того чтобы начать классификацию значений и затем их распределить в рамках интервалов, необходимо понять ширину выборки значений, а после этого ее размах разделить на эффективное с точки зрения дальнейшей классификации количество интервалов. Правильное распределение позволит сделать выводы на основании достоверного описания сложившейся ситуации, а также сделать выборку типичных значений для проведения дальнейшего исследования. С этой целью на основании стандартных процедур, которые доступны в любом программном обеспечении для анализа данных, необходимо установить минимальную цифру, которая позволит определить нижнюю границу для анализа и дальнейшего распределения значений из всего рассматриваемого набора. Исходя из данных табл. 1, $X_{\min} = 0,54$ млн руб. и максимальное, как верхняя граница – $X_{\max} = 6,55$ млн руб. Размах вариации при этом получится $R = 6,01$. Следующим этапом должно стать определение количества интервалов, в которые мы разместим имеющиеся у нас значения. Для этого можно воспользоваться формулой Стерджесса как наиболее распространенным и доступным инструментом [1]:

$$H = R/(1+\log 2n) = R/(1+3,322\lg n),$$

где n является размером представленного набора данных, то есть количеством рассматриваемых элементов, в нашем случае речь идет о 100 наблюдениях за характеристиками однотипных субъектов хозяйствования.

В соответствии с расчетами получается следующая оптимальная длительность интервала: $H = 0,7863$. Это говорит о том, что наши значения будут разделены на 8 групп для дальнейшего анализа частот и распределения. В то же время расчет знаменателя приведенной формулы позволит также установить количество интервалов и после

Интервал	Частота	Накопленная частота	Относительная частота (частота / n)	Накопленная относительная частота (относительная частота / n)	Плотность распределения (частота / длительность интервала)
0,147– 1,326	10	10	0,1	0,1	0,13
1,326–2,113	10	20	0,1	0,2	0,13
2,113–2,899	32	52	0,32	0,52	0,41
2,899–3,685	26	78	0,26	0,78	0,33
3,685–4,471	9	87	0,09	0,87	0,11
4,471–5,258	6	93	0,06	0,93	0,076
5,258–6,044	4	97	0,04	0,97	0,051
6,044 –6,83	3	100	0,03	1	0,038
ВСЕГО	100		1		

Таблица 3. Интервальный вариационный ряд распределения результатов продаж рассматриваемых предприятий по состоянию на 01.09.2023 г., млн руб.

округления высчитать значение каждого из них, а затем продолжить исследуемую выборку.

С учетом полученного значения для длительно-сти интервала распределим исследуемые величины (табл. 2). В целях повышения достоверности проводимого анализа и учета крайних значений выборки нижняя граница может быть смещена на половину значения интервала. Некоторые средства, которые можно использовать для анализа рядов распределения, могут не включать в интервал пограничные минимальные и максимальные цифры, что искажает конечную картину, и в таких случаях смещение границ позволяет исправить данный тип классификационной ошибки.

Рассматриваемая совокупность данных делится на интервалы до тех пор, пока в выделенном наборе начало следующего не будет равно или больше максимального значения. Некоторые аналитические инструменты в силу особенностей реализации могут не учитывать последнее, если оно совпадает с верхней границей последнего интервала, поэтому целесообразно определить верхнюю границу с таким расчетом, чтобы ее значение превышало максимум выборки.

После ранжирования величин необходимо установить, каким образом они распределились по выделенным интервалам – это является базовым действием для дальнейшего анализа участников рынка и интерпретации сложившейся ситуации. С учетом этого требуется выделить обязательные составляющие, позволяющие охарактеризовать распределение и оценить вариации рассматриваемых значений (табл. 3). Исследование рядов динамики предполагает расчет ряда показателей, на основании которых

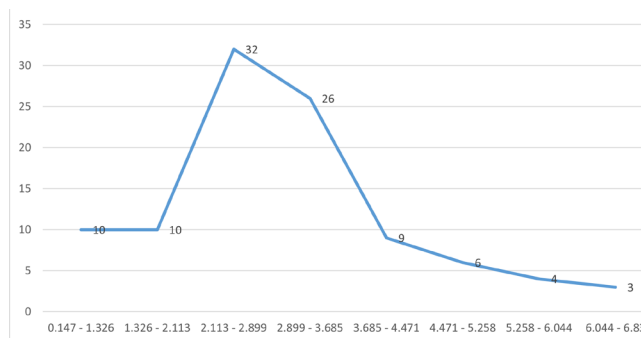


Рисунок. Итоговое распределение частот для рассматриваемого массива значений

можно определить достаточность и достоверность данных для проведения дальнейшего анализа и для формулировки выводов, характеризующих изучаемое явление.

Остановимся более подробно на показателях, входящих в состав табл. 3. Индикатор «частота» характеризует количество наблюдений, которые попали в тот или иной интервал нашего ряда распределения. Накопленная частота формирует представление об объеме данных, значения которых меньше анализируемых, и рассчитывается как сумма абсолютных значений частот предшествующих интервалов. Относительная частота представляет собой соотношение абсолютного значения частоты данного интервала и общего количества элементов рассматриваемого массива. Плотность распределения – показатель, который рассчитывается через соотношение абсолютной частоты и длительности интервала и позволяет оценить насыщенность элементами массива конкретного интервала распределения данных.

Расчет этих показателей позволяет дать базовую оценку ряда распределения на основании его модальности, характера наполнения интервалов, выбросов и прочих базовых признаков на начальной аналитической стадии.

На основании полученных данных изображение объективной картины в рассматриваемой выборке можно отобразить в виде полигона распределения значений по предприятиям исследуемого рынка (рисунки). Такое распределение требует дальнейшей интерпретации в совокупности с базовыми показателями, характеризующими рассматриваемый ряд данных, которые были рассчитаны ранее. В данном случае явно прослеживается асимметрия хвостов полигона, а также смещение центральной части графика, что требует более подробной дальнейшей интерпретации и, возможно, накопления данных по другим аналогичным периодам работы рассматриваемых субъектов хозяйствования на этом рынке.

Полученные результаты говорят о том, что наибольшее количество участников рынка относится к среднему диапазону значений, то есть 32 юридических лица находятся в диапазоне от 2,113 млн руб. до 2,899 млн руб. и 26 – от 2,899 млн руб. до 3,685 млн руб. В максимальный интервал вошли 3 субъекта хозяйствования, в минимальный – 10. При этом 48 компаний достигли значения выше среднего – 2,93 млн руб., что может говорить о предпосылках к положительной динамике или стабильной ситуации на изучаемом рынке. Более конкретные результаты могут быть получены при проведении дальнейших этапов анализа, осуществляемого с помощью коэффициентов или более специальных аналитических процедур, выбор которых зависит от задач, стоящих перед лицами, проводящими исследование.

Следует обратить внимание на характер распределения данных в нашей выборке, которому свойственна асимметрия со смещением к минимальным значениям. Таким образом, речь идет о необходимости поиска фактора, оказавшего влияние на субъекты хозяйствования. Анализ должен быть продолжен в отношении причины, вызвавшей такое положение. В этом случае можно собрать больше данных, для того чтобы получить более подробную картину, либо прибегнуть к дополнительным методам для конкретизации факторов, оказывающих влияние на объекты исследования и, как следствие, на итоговые показатели, которые мы используем для дальнейшего анализа, построения зависи-

мостей, формулировки выводов и описания явлений, имеющих место на рынке.

В дальнейшем рассматривались макроэкономические показатели и проводился коэффициентный анализ деятельности субъектов хозяйствования. Было установлено, что с учетом внешних характеристик, оказавших на них влияние, смещение значений выборки было вызвано изменениями, произошедшими в регулировании деятельности изучаемого рынка и связанными с колебаниями внешнеэкономических условий. ■

■ **Summary.** The article presents an approach to constructing distribution series for assessing the activities of legal entities that are united by some typical feature characterizing their business activities. In addition, an algorithm is described for analyzing the values of a set of companies when analyzing the state of the industry or market to which they belong. A variant of displaying the results of the analysis of the activities of legal entities similar from the point of view of the market is proposed, which is based on the analysis of the nature of the distribution of data for a number of companies that were obtained based on the results of their business activities for the reporting period or are commercial results of activities during certain time intervals. The conclusion describes the proposed directions for using this approach when analyzing financial or commercial data of a set of similar companies.

■ **Keywords:** data analysis, financial data, data sets, finance, forecasting, financial analytics, scalar values, solvency assessment, comprehensive assessment, assessment scale, transaction assessment, legal entity assessment, financial analysis, legal entity assessment, frequency, relative frequency, skewness of values, normal distribution, data distribution, Sturges formula, distribution density.

■ <https://doi.org/10.29235/1818-9857-2025-04-43-46>

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Коэн М.И. Прикладная линейная алгебра для исследователей данных / пер. с англ. А.В. Логунова. – М., 2023.
2. Мыльников Л.А. Статистические методы интеллектуального анализа данных. – СПб., 2021.
3. Помазанов М.В. Управление кредитным риском в банке: подход внутренних рейтингов. – М., 2023.
4. Практическая статистика для специалистов Data Science: пер. с англ. / П. Брюс, Э. Брюс. – СПб., 2019.

Статья поступила в редакцию
04.11.2024 г.