

DOI: 10.34680/BENEFICIUM.2024.4(53).49-57

УДК 339.37:004.9

JEL L81, M31



ОРИГИНАЛЬНАЯ СТАТЬЯ

МНОГОФАКТОРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ И ПРОГНОЗИРОВАНИЕ РОЗНИЧНЫХ ПРОДАЖ ТОВАРОВ FMCG В ОФФЛАЙН МАГАЗИНАХ С ИСПОЛЬЗОВАНИЕМ POS ДАННЫХ

Ю.В. Напольская, Проктер энд Гэмбл, Москва, Россия

Аннотация. В современном ритейле ввиду высокой конкуренции среди товаров FMCG компании-производители и поставщики используют разнообразные способы продвижения своих товаров. Параллельно уровень технологии сбора и хранения данных позволяет накапливать данные о товарах в точках продаж по широкому наименованию метрик, включая те, которые отражают усилия продавцов по увеличению продаж и результаты этих усилий. Каждая розничная точка становится источником бесценных POS данных. Использование POS данных пока не стало повсеместной практикой, однако интерес к этому источнику постоянно растет. В статье обсуждается использование POS данных для оценки уровня влияния конкретных факторов на продажи товаров FMCG в розничных точках. Проводится регрессионный анализ, используя в качестве факторных признаков такие POS данные, как полочная цена товара, суммарная дистрибуция, доля полки, доступность товара, уровень товарного запаса. Полученные уравнения регрессии анализируются с точки зрения их экономического смысла и практического применения. Делается вывод о том, что на определенном этапе развития продуктовой линейки или бренда компании-производителю следует сосредоточиться на том конкретном виде продвижения, высокий уровень влияния которого становится очевиден из уравнения регрессии. Это способствует распространению практики принятия решений на основе данных и способствует росту эффективности бизнес-процессов. Также обсуждается возможность применения регрессионного анализа с использованием POS данных в качестве факторных признаков для прогнозирования продаж, поскольку цена и уровень представленности в магазине во многом определяют уровень спроса на товары FMCG, и, значит, будут хорошими предикторами для прогноза.

Ключевые слова: корреляционный анализ, прогнозирование продаж, регрессионный анализ, розничная торговля, товары массового потребления, POS данные

Для цитирования: Напольская Ю.В. Многофакторный регрессионный анализ и прогнозирование розничных продаж товаров FMCG в оффлайн магазинах с использованием POS данных // BENEFICIUM. 2024. № 4(53). С. 49-57. DOI: 10.34680/BENEFICIUM.2024.4(53).49-57

ORIGINAL PAPER

MULTIFACTOR REGRESSION ANALYSIS AND FMCG RETAIL SALES FORECASTING IN OFFLINE STORES USING POS DATA

Y.V. Napolskaya, Procter and Gamble, Moscow, Russia

Abstract. In the modern retail trade companies-producers and suppliers use a variety of ways to promote their products due to high competition among FMCG products. At the same time, the level of data collection and storage technology allows for the accumulation of POS data across a broad range of metrics, including those that reflect sales growth efforts and the results of these efforts. Every retail outlet becomes a source of invaluable POS data. The use of POS data has not yet become a common practice, but interest in this source is constantly growing. The article examines the use of POS data to estimate the level of specific factors influence on FMCG sales in retail outlets. Regression analysis is carried out basing on such POS data as the product price, the total distribution point, the shelf share, on-shelf availability, the product stock level. The obtained regression equations are analyzed in terms of their economic sense and practical application. It is concluded that at a certain stage of development of the product line or brand the producing company should focus on that particular type of promotion, whose high level of influence becomes evident from the regression equation. This promotes the data-driven decision making and improves the efficiency of business processes. The possibility of using regression analysis with POS data as explanatory variables for sales forecasting is also discussed, since the price and the level of representation in the store largely determine the demand for FMCG products, which means they will be good predictors for the forecast.

Keywords: correlation analysis, sales forecasting, regression analysis, retail trade, FMCG, POS data

For citation: Napolskaya Y.V. Multifactor Regression Analysis and FMCG Retail Sales Forecasting in Offline Stores Using POS data // Beneficium. 2024. Vol. 4(53). Pp. 49-57. (In Russ.). DOI: 10.34680/BENEFICIUM.2024.4(53).49-57

Введение

Управление розничными продажами на рынке товаров FMCG (Fast Moving Consumer Goods, товары массового спроса) представляет собой предмет постоянного внимания со стороны как руководителей розничных сетей и отдельных магазинов, так и поставщиков. Правильный ассортимент и цена товара для конечного потребителя являются важными конкурентными преимуществами и служат для достижения целей компании по продажам и обеспечению целевого уровня прибыли. Руководствуясь правилом классического маркетинга 4P (product, price, place, promotion – продукт, цена, место, промоподдержка) [1], компании стараются обеспечить наличие правильного товара по правильной цене в как можно большем количестве точек продаж, сопровождая его соответствующей рекламной поддержкой. Для того, чтобы убедиться в том, что все эти конкурентные преимущества реализуются, компании-ритейлеры и поставщики постоянно отслеживают ряд показателей, например, цена на полке, размер скидки, цена товаров-конкурентов, уровень дистрибуции, доступность и наличие товара, доля полки, дополнительные места выкладки, собственно продажи в штуках и денежных единицах и так далее. Кроме того, каждая покупка, каждая транзакция на кассе магазина увеличивает массив доступных данных, добавляя к существующим день и время совершения покупки, цену приобретения, количество товаров в чеке, количество товарных категорий в чеке, размер чека и так далее. Эти и многие другие показатели, которые снимаются в каждой точке продаж, относятся к POS данным (point-of-sales) и предоставляют огромные возможности по анализу, выявлению тенденций и поиску ответов на множество вопросов.

POS данные используются компаниями в большей или меньшей степени, и те компании, которые первыми учатся извлекать из данных полезные инсайты, получают большое конкурентное преимущество. Например, если компания-производитель или компания-ритейлер оценит, насколько сильно влияет какой-либо фактор на продажи конкретного товара на конкретном этапе развития на рынке, то она сможет сосредоточить свои усилия именно на том компоненте, который оказывает наибольшее влияние, и использовать это знание для роста доли рынка и прибыли компании. Например, если на продажи значительно влияет размер доли полки, занимаемой продуктом, то компания может сосредоточиться на увеличении этого показателя и сфокусировать силы торгового персонала на этом компоненте. Кроме этого, зная, как влияют на уровень продаж ключевые отслеживаемые показатели, становится возможным решить также и задачу прогнозирования продаж, что является необходимым условием для обеспечения эффективной деятельности компании, формирования плана закупок, рекламной поддержки и

так далее.

Вопрос прогнозирования продаж постоянно находится в зоне внимания исследователей, которые находят и предлагают новые подходы. В работе [2] авторы систематизируют все используемые в настоящий момент методы прогнозирования объемов продаж готовой продукции и обращают внимание, что в большинстве случаев в российской и зарубежной практике используются методы трех типов: интуитивные или экспертные (основанные на субъективной оценке), методы анализа и прогнозирования временных рядов и казуальные (причинно-следственные) методы (основанные на анализе количественных показателей) [3]. Поскольку в данной работе планируется не только решить задачу прогнозирования продаж, но и оценить влияние на них различных факторов, то мы обратимся к причинно-следственным методам, которые позволят решить задачу выбора наилучших предикторов. В работе [4] также отмечается, что эвристические (экспертные) методы хорошо применимы, когда количество анализируемых данных небольшое, например, на ранних стадиях запуска нового товара, а в случаях, когда накоплено достаточное количество данных, целесообразно переходить к количественным методам. Причинно-следственные (или казуальные) методы достаточно широко представлены в научных исследованиях. Например, в работе [5] задача прогнозирования для компаний из разных отраслей решается при помощи модели авторегрессии и интегрированного скользящего среднего (ARIMA), не затрагивая вопрос влияния сторонних факторов на продажи. Используется и корреляционно-регрессионный анализ, однако подход исследователей отличается в том, какие факторы выбираются для анализа. Так, например, в работе [6] автор при прогнозировании продаж мебельной продукции анализирует такие предикторы, как показатели выручки, объема жилищного строительства, среднемесячной заработной платы и процентных ставок по потребительским кредитам коммерческих банков. В работе [7] автор также использует регрессионный анализ для прогнозирования выручки предприятия, в качестве предиктора используя один факторный признак – уровень коммерческих расходов (который включает в себя рекламные затраты). В работе [8] регрессионный анализ продаж проводится на основе трех факторных признаков: цены, уровня скидки и затрат на рекламу. В этой работе анализ временных рядов показывает лучший результат, чем регрессионный анализ, однако для однозначного вывода о преимуществе того или иного метода в работе не хватает данных о качестве построенной регрессионной модели (например, отсутствует расчет R^2), и, возможно, при доработке списка факторов качество модели могло бы быть улучшено и

сравнимо с результатом анализа временных рядов. Таким образом, в качестве предикторов для регрессионного анализа в существующих отечественных работах преобладают либо финансовые показатели предприятия, либо макроэкономические показатели, а использование POS данных ограничено. С нашей точки зрения это является хорошей возможностью для исследования, поскольку макроэкономические показатели изменяются сравнительно медленно, а ситуация в розничных точках меняется каждый день – появляются новые товары, запускаются новые механики промоактивностей, изменяются планы продаж, а с ними и привычный покупателю план расстановки товаров, и тем важнее понять, какое влияние на продажи вносят многочисленные изменения в розничных точках. Также важно, что на внешние макроэкономические факторы розничные компании и производители не имеют практического влияния, в то время как на внутренние факторы могут оказывать непосредственное влияние и управлять ситуацией с ценой, представленным товарным ассортиментом, уровнем и качеством выкладки товара, промопродвижением и т.д. Тем важнее понять, как именно внутренние факторы влияют на продажи, чтобы можно было этими продажами управлять.

Если обратиться к зарубежной практике, то в работе [9] авторы проанализировали 255 статей в научных журналах, посвященных исследованиям в области функционирования розничной торговли за период 2008-2016 гг. Они отметили, что среди часто изучаемых вопросов присутствует и вопрос прогнозирования спроса (demand forecasting), и эта задача связана с другими смежными вопросами – своевременным пополнением стока, поддержанием товарного запаса (on-shelf availability) на определенном целевом уровне, что ведет к выполнению целей компаний по продажам. Все это невозможно было бы изучать без POS данных, практически в реальном времени получаемых непосредственно в розничных магазинах. Действительно, в зарубежных работах использование POS данных при прогнозировании продаж розничных сетей используется достаточно широко. Например, в работе [10] авторы сравнивают различные методы для прогнозирования продаж, опираясь на набор данных, содержащих в том числе и POS данные, полученные из розничных магазинов и включающие информацию о номере магазина, идентификаторе проданного товара и количестве проданного товара в штуках, собранные за период с 2013 по 2018 год. Также различные методы прогнозирования продаж, в том числе и регрессионный анализ, сравнивают в работе [11], используя большой набор POS данных из 1115 розничных магазинов, включая непосредственно продажи, тип магазина, ассортимент магазина и т.д. В работе также отмечено, что прогнози-

рование продаж может быть только первым шагом, следующими шагами могут стать прогнозирование спроса (demand) и формирование соответствующего товарного запаса в магазинах, определение влияния рекламы на покупателей, гибкая работа с ценой и другие прикладные важные задачи. В работе [12] автор сравнивает несколько распространенных методов прогнозирования продаж, используя POS данные сети супермаркетов, среди которых большее влияние оказывали такие предикторы, как тип магазина, цена продукта, год открытия магазина, представленность товара на полке и другие. В работе [13] авторы также занимаются прогнозированием спроса, отмечая, что традиционные методы анализа временных рядов не учитывают факторы, влияющие на поведение покупателей в моменте, такие как цена товара, изменение цены, погодные условия и другие. В работе авторы используют как регрессионный метод, так и метод линейного программирования.

Цель настоящего исследования – оценить степень влияния различных показателей на уровень продаж многообразных типов товаров FMCG для дальнейшего выбора стратегии развития продукта, а также прогнозирование уровня продаж при заданных условиях с использованием POS данных. Для этого в данном исследовании мы будем использовать многофакторную регрессионную модель.

Результаты и их обсуждение

Перед тем, как перейти к вопросу построения регрессионной модели, необходимо определить, какие данные будут использоваться в анализе. В нашем случае мы будем работать с рядом показателей, которые отслеживаются компанией-производителем продукта в розничных точках и отвечают на вопрос о развитии и состоянии конкретного товара или бренда. Большинство этих POS данных собирается удаленно с помощью проекта по обмену данными с розничной сетью (data sharing), часть данных собирается сторонними аудиторскими агентствами при визитах в розничные точки.

В нашем распоряжении доступны следующие показатели, которые мы рассмотрим подробно:

- 1) Sales (Sales volume) – объем продаж в количественном выражении. Здесь важно, чтобы исследуемый объем продаж был выражен именно в количественном выражении, а не в денежном, поскольку мы хотим исследовать влияние цены на уровень продаж. Если продажи измерять также в денежных единицах, то эффект будет смазан, поскольку, например, при снижении цены выручка за единицу товара также снизится. Вместе с тем, здесь важно отметить, что для более точного анализа следует использовать показатель объема проданного продукта, но выраженный не в штуках,

а обратиться к приведенному объемному критерию продукта в физических единицах (мл или кг) или, например, в Statistical Unit (SU – статистическая единица потребления), который используется в крупнейших FMCG компаниях для приведения к общему виду разных размеров упаковок. Например, 1 упаковка шампуня 400 мл приравнивается к 1 SU продукта, упаковка 200 мл – 0.5 SU. Если, например, продажи составили 3 упаковки шампуня по 400 мл и 2 упаковки по 200 мл, а в следующий период продажи составили 2 упаковки по 400 мл и 3 упаковки по 200 мл, то ошибкой будет считать общие продажи, как 5 штук в первый период и 5 штук во второй период, поскольку в первом случае продажи составили 1600 мл, во втором случае 1400 мл, то есть в объеме продажи уменьшились, а не остались на прошлом уровне. Использование SU, как и мл, также устраняет эту ошибку – в первом случае продажи составили 4 SU, во втором случае 3.5 SU.

2) Price (Shelf price) – полочная цена. Здесь важно использовать именно конечную полочную цену, по которой покупатель совершил покупку. Тогда эта цена будет включать в себя возможные периоды промоподдержки в этот временной интервал, например, снижение цены, использование цены по дисконтной карте, и мы сможем исследовать влияние увеличения/ снижения цены на уровень спроса. Ошибкой будет использовать регулярную цену товара на ценнике при том, что на кассе при использовании дисконтной карты или других механик покупатель заплатит другую сумму.

3) TDP (Total Distribution Points) – суммарная дистрибуция всех SKU (Stock Keeping Unit – уникальная товарная позиция). Показатель TDP (или TD – Total Distribution) широко используется в ритейле наряду с традиционными показателями нумерической и взвешенной дистрибуции. Его преимущество по сравнению с вышеназванными показателями заключается в том, что он одновременно содержит в себе информацию о ширине дистрибуции (то есть, о количестве уникальных торговых точек, в которых присутствует продукт) и о глубине дистрибуции (количестве SKU в каждой из этих точек). Основной способ определить TDP – это использовать данные о % ACV (All Commodity Volume, общий объем продаж) – по каждой SKU и суммировать их [14]. Здесь % ACV используется как аналог взвешенной дистрибуции. Например, если Продукт А имеет показатель 60% ACV, то значит, он присутствует в магазинах, в которых продается 60% объема всей сети, а Продукт В имеет 70% ACV, то мы можем посчитать TDP как сумму, она составит 130. Максимальный показатель TDP в данном случае равен 200, так что всего один показатель TDP говорит нам о том, какой у нас есть потенциал роста (с 130 до 200).

Если объем продаж каждого магазина сети и, следовательно, % ACV неизвестны, тогда взвешенную дистрибуцию в расчете можно заменить на нумерическую. Тогда TDP это сумма всех SKU

бренда в каждом магазине присутствия. Например, если Продукт А присутствует в 10 магазинах торговой сети из 100 магазинов, а Продукт В в 20 магазинах, то показатель TDP составит 30 при максимальном значении 200, и мы также можем оценить потенциал роста. В данном исследовании мы используем показатель TDP, рассчитанный как сумма всех SKU во всех магазинах.

4) OOS (Out-of-stock) – показатель доступности товара, при OOS, равном 100%, товар полностью отсутствует на полке. Обычно целевой показатель OOS < 5%, что говорит о доступности товара OSA (on-shelf availability) на полке больше, чем в 95% случаев. При более высоких показателях OOS его влияние на продажи растет, так как покупатель не находит нужного товара на полке и не совершает покупку.

5) Stock – товарный запас продукта в количественном выражении. Мы включили этот показатель в исследование, так как исходим из предположения, что уровень товара запаса оказывает влияние на продажи. Например, при большом стоке товара его легко заметить на полке, он может быть выложен в дополнительных местах продаж, на кассах и т.п., а при стоке 1-2 штуки продукта он может быть не замечен на полке, быть заслоненным другим продуктом либо это может быть виртуальный сток (который числится на балансе торговой точки, но физически отсутствует).

6) Shelf share – доля продукта/ бренда на полках, один из показателей, описывающих представленность товара в магазине. Чем больше доля полки, тем заметнее продукт для покупателей, и тем выше ожидаются продажи. Данные о доле полки различных брендов, как правило, собираются аудиторскими агентствами.

Здесь мы ограничимся этими показателями, поскольку они почти полностью описывают ситуацию с конкретным продуктом по всем 4Р компонентам, за исключением рекламной поддержки, что не входит в сферу данного исследования. Чтобы исключить этот компонент, предположим, что уровень рекламной поддержки одинаков на протяжении всего исследуемого времени.

Для целей дальнейшего анализа мы будем работать с набором POS данных, содержащем информацию по всем вышеуказанным показателям по месяцам за достаточно длительный промежуток времени (2 года 3 месяца, 27 наблюдений). Важно упомянуть, что в общем случае при проведении анализа необходимо обеспечить такие качества исследуемых данных, как непрерывность, надежность, полнота, точность, непротиворечивость и другие, иначе в анализе могут быть допущены ошибки. Отметим также, что в данной работе мы сфокусируемся на практическом применении регрессионного анализа и его экономической трактовке, поскольку непосредственно математический аппарат подробно изложен в специальной литературе. Фрагмент исследуемых данных приведен в *табл. 1*.

Таблица 1 / Table 1

Фрагмент исходных данных для анализа / Piece of POS Data for Analysis

Наблюдения / Observation	Продажи, SU / Sales, SU	Цена, руб. / Price, rur.	Дистрибуция / TDP	Товарный запас, SU / Stock, SU	Доля полки, % / Shelf Share, %	Отсутствие товара на полке, % / OOS, %
1	117	273	331	433	10.5	6
2	155	265	302	436	10.9	5
3	160	260	327	420	10.6	4
4	140	275	290	395	10.6	4
5	130	264	302	392	10.5	6
6	120	270	320	397	10.8	6
7	139	272	309	414	10.7	5
8	156	260	340	432	10.5	2
9	126	259	312	425	10.9	5
10	102	267	280	406	10.8	3

Источник: составлено автором / Source: compiled by the author

Обозначим уровень продаж Sales как Y (результативный признак), а каждый исследуемый показатель как X_i (факторный или объясняющий признак, или предиктор). Одним из методов установления связей между случайными величинами является корреляционный анализ, который позволит нам установить уровень взаимосвязи между величинами Y и X_i . Как отмечено ранее, уровень продаж зависит от целого ряда факторов, но связь продаж с каждой из этих переменных может быть разной,

и нам нужно найти именно те переменные, которые коррелируют с нашей исследуемой величиной сильнее всего. Для выявления лучших факторных признаков и проведем корреляционный анализ. Напомним, что две переменные коррелируют между собой, то есть их значения связаны между собой тем больше, чем коэффициент корреляции ближе к ± 1 . Если коэффициент близок к 0, то корреляции нет или она незначительна. Построим таблицу коэффициентов для всех наших данных в табл. 2.

Таблица 2 / Table 2

Парные коэффициенты корреляции $r(X_1, X_2)$ / Correlation Coefficients $r(X_1, X_2)$

	$Y = \text{Продажи} / Y = \text{Sales}$	$X_1 = \text{Цена} / X_1 = \text{Price}$	$X_2 = \text{Дистрибуция} / X_2 = \text{TDP}$	$X_3 = \text{Товарный запас} / X_3 = \text{Stock}$	$X_4 = \text{Доля полки} / X_4 = \text{Shelf Share}$	$X_5 = \text{Отсутствие товара на полке} / X_5 = \text{OOS}$
$Y = \text{Sales}$	1.00	-0.76	0.29	0.16	-0.06	-0.05
$X_1 = \text{Price}$	-0.76	1.00	-0.04	-0.19	-0.09	0.05
$X_2 = \text{TDP}$	0.29	-0.04	1.00	0.35	-0.07	-0.13
$X_3 = \text{Stock}$	0.16	-0.19	0.35	1.00	0.02	-0.05
$X_4 = \text{Shelf share}$	-0.06	-0.09	-0.07	0.02	1.00	0.08
$X_5 = \text{OOS}$	-0.05	0.05	-0.13	-0.05	0.08	1.00

Источник: рассчитано автором / Source: calculated by the author

Нам нужно выбрать переменные с сильной корреляцией с переменной Sales, тогда мы сможем построить регрессию, которая хорошо будет объяснять ее поведение. Из табл. 2 видно, что переменная Sales сильнее всего коррелирует с ценой ($r(Y, X_1) = -0.76$), суммарной дистрибуцией TDP ($r(Y, X_2) = 0.29$) и товарным запасом Stock ($r(Y, X_3) = 0.16$). Связь продаж с переменными Shelf share и OOS не очень сильная, значит, их влияние на объясняемую величину Sales небольшое, и на данном этапе мы можем исключить их из рассмотрения.

Прежде чем идти дальше, необходимо исключить эффект мультиколлинеарности – то есть наличие сильной взаимосвязи между факторными признаками, которые мы собираемся выбрать для построения модели. При наличии линейной связи между факторными признаками X_i значение коэффициентов регрессионной модели будет сложно интерпретировать [15], поскольку изменение одного факторного признака приведет к изменению не только результативного признака, но и другого

факторного. Для исключения этого эффекта достаточно проанализировать взаимные коэффициенты корреляции между факторными признаками, и взять те из них, которые по абсолютной величине связаны с результативным признаком Y больше, чем между собой. В нашем примере есть два факторных признака TDP и Stock с сильным уровнем взаимосвязи ($r(X_2, X_3) = 0.35$), и каждый из них связан друг с другом сильнее, чем с Y : $0.35 > 0.29$ и $0.35 > 0.16$. Эта взаимосвязь объяснима с практической точки зрения, так как с увеличением количества точек физической дистрибуции или количества SKU в них (TDP) растет и общий уровень товарного запаса. Выбирая между двумя взаимно-коррелированными признаками необходимо понять, какой из них лучше всего включить в анализ и сделать одним из факторных признаков, объясняющих нашу модель. В нашем случае уровень взаимосвязи суммарной дистрибуции TDP с Sales ($r(Y, X_2) = 0.29$) превышает уровень взаимосвязи товарного запаса Stock с Sales ($r(Y, X_3) = 0.16$), то есть TDP лучше объясняет переменную Sales.

Кроме того, важно исходить из того, на какую переменную мы сможем повлиять впоследствии при работе с торговыми сетями. Например, мы можем влиять и на товарный запас (Stock), и на суммарную дистрибьюцию TDP, однако наше влияние на товарный запас ограничено (ввиду того, что у торговой сети существуют свои цели по товарному запасу на складах и на полках), в то время как увеличение числа магазинов или увеличение количества SKU в магазине, где продается товар, вполне точная, управляемая и хорошо отслеживаемая цифра. Поэтому в нашей модели мы рекомендуем в качестве факторного признака выбрать суммарную дистрибьюцию TDP.

После проведения корреляционного анализа мы можем приступить к поиску функционального

вида связи между результативным признаком Y и факторными признаками X_i, которые мы выбрали, то есть найти функцию вида:

$$Y_i = f(X_i), \tag{1}$$

Уравнение множественной линейной регрессии, связывающее результативный признак Y и ранее выбранные два факторных признака, TDP и Price, имеет следующий вид:

$$Y = a_0 + a_1X_1 + a_2X_2, \tag{2}$$

или

$$Sales = a_0 + a_1Price + a_2TDP, \tag{3}$$

Задача определения коэффициентов уравнения с успехом решается на практике, например, с помощью надстройки Data analysis в программе Microsoft Excel, результат работы которой представлен в табл. 3.

Таблица 3 / Table 3

Результаты поиска коэффициентов линейной регрессии в Microsoft Excel / Search Results for Linear Regression Coefficients in Microsoft Excel

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0,80430822							
R Square	0,64691171							
Adjusted R Square	0,61748768							
Standard Error	18,198664							
Observations	27							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	2	14563,04	7281,518	21,98583	3,75491E-06			
Residual	24	7948,593	331,1914					
Total	26	22511,63						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95,0%	Upper 95,0%
Intercept	369,746695	92,04126	4,017184	0,000505	179,7828695	559,7105	179,7829	559,7105
Price	-1,4502595	0,234024	-6,19706	2,1E-06	-1,93326094	-0,96726	-1,93326	-0,96726
TDP	0,45270474	0,21398	2,115639	0,044944	0,011071435	0,894338	0,011071	0,894338

Источник: рассчитано автором / Source: calculated by the author

В табл. 3 представлены наши искомые коэффициенты и ряд важных статистических показателей. Среди них важно обратить внимание на следующие:

- коэффициент детерминации R=0.64 описывает общее качество модели, чем ближе значение к 1, тем точнее модель. Данная величина говорит о том, что в нашей модели факторные признаки на 64% описывают изменение результативного признака Y;
- Significance F<0.05, следовательно, регрессионная модель является статистически значимой;
- P-value <0.05, т.е. значения коэффициентов уравнения определены с вероятностью ошибки <0.05.

В нашем случае мы получили функцию вида:
 $Sales = 369.75 - 1.45 Price + 0.45 TDP,$ (4)

где полученные коэффициенты следует трактовать следующим образом:

- a₀ = 369.75, свободный коэффициент, математический смысл которого в том, что при цене, равной 0, и TDP, равном 0, продажи составят 369.75 единиц в объеме;
- a₁ = -1.45, то есть при повышении полочной цены на 1 рубль, продажи упадут на 1.45 единиц в объеме, другими словами, в тех же величинах, что во входных данных;
- a₂ = 0.45, то есть при увеличении TDP на единицу (например, в одном магазине стало продаваться на 1 SKU больше, чем до этого), продажи за 1 период увеличатся на 0.45 единиц в объеме.

Таким образом, мы оценили влияние двух факторов – цены на полке и суммарной дистрибьюции TDP на продажи продукта.

Регрессионный анализ используется для решения двух типов задач – прогнозирование и сравнение [16]. В нашем случае полученную регрессию также можно применить на практике несколькими способами. Во-первых, ее можно использовать для прогнозирования продаж и моделирования многочисленных ситуаций при различных значениях полочной цены и TDP. Например, в случаях принятия решения касательно расширения/сужения дистрибьюции: допустим, у вас есть товар-бестселлер, который обеспечивает отличные показатели продаж и прибыли, но присутствует только в 20% розничных точек сети. Тогда эту модель можно использовать для того, чтобы спрогнозировать будущий уровень продаж при уровне дистрибьюции 30%, 40% и так далее, а потом оценить, насколько прогнозируемый рост продаж оправдывает усилия по увеличению дистрибьюции. Также модель предоставляет широкие возможности для работы с ценой. Поскольку в результате мы фактически получили модель эластичности спроса по цене, мы можем формулировать и решать разнообразные задачи: каким будет уровень продаж при заданном уровне цены, как изменится уровень продаж при предоставлении скидки покупателю при проведении промо-мероприятий и многие другие. Также возможно прогнозировать продажи при формировании календаря промо-мероприятий на квартал, полгода или год с тем, чтобы увязать целевой уровень продаж с имеющимся промо-бюджетом.

Во-вторых, подобный анализ можно провести для разных типов продуктов, разных брендов и категорий из портфеля компании, что позволит решать стратегические задачи. Уравнения регрессии для разных продуктов расскажут об индивидуальных характеристиках продуктов на их текущих этапах развития. Допустим, что уравнение регрессии для продукта А и продукта В выглядят следующим образом (5), (6):

$$Sales_A = 105 - 0.0045 Price + 3.50 TDP, \quad (5)$$

$$Sales_B = 105 - 1.45 Price + 0.05 TDP, \quad (6)$$

Основываясь на коэффициентах уравнения регрессии, мы можем сказать, что на текущем этапе дистрибьюция продукта В влияет на продажи незначительно (увеличение TDP на 1 пункт приводит к росту продаж на 0.05 единиц при той же величине цены), а для продукта А, напротив, влияние TDP на уровень продаж значительно. Таким образом, при принятии решения о расширении дистрибьюции при наличии этой возможности следует сделать фокус на расширении продукта А. Часто такая ситуация на практике возникает в случае, если продукт уже обладает очень широкой дистрибьюцией, например, шампунь ключевой линейки популярного бренда масс-маркета. Поэтому при увеличении его и так широкой дистрибьюции большого влияния на рост продаж это не оказывает. Противоположная ситуация возникает для продукта, который, напротив, недостаточно ши-

роко распространен: это может быть новый, недавно запущенный на рынок продукт или сравнительно новая категория для рынка, например, таблетки для стиральной или посудомоечной машины. В этом случае стратегия компании по развитию продукта должна обязательно включать расширение дистрибьюции.

Аналогичные рассуждения можно применить к уравнениям (5) и (6) при планировании бюджета на проведение рекламных компаний со снижением цены. При увеличении цены на 1 рубль продажи продукта В упадут на 1.45 единиц товара, а продукта А – только на 0.0045 единиц, соответственно, и при снижении цены рост продаж продукта В будет больше в объеме, чем продукта А. Это означает, что инвестировать в проведение промоакций со снижением цены следует только для продукта В, а для продукта А нужны другие способы поддержки. При проведении такого сравнения необходимо также учитывать цифры общих продаж продуктов А и В, а также их прибыльности для компании, поскольку уравнение регрессии построено на данных продаж в объеме.

Следующим этапом оценки влияния различных факторов на уровень продаж в розничных точках может стать включение и изучение большего количества факторов или объясняемых признаков, чем в нашем случае. Например, в рассматриваемом примере влияние такого фактора, как доля полки (Shelf share) минимально, и это объясняется низкой амплитудой изменения этого параметра, то есть на исследуемом периоде доля полки практически не менялась. Возможно, на более длительном отрезке времени или у другого продукта влияние доли полки будет больше, и это даст возможность ответа на вопрос – чувствителен ли продукт к доле полки, и надо инвестировать в ее расширение через фиксацию уровня доли полки с торговой сетью, либо практически не чувствителен, так как покупатель в любом случае идет за этим продуктом и находит его, как бы заметно или нет он не был представлен.

Важно подчеркнуть, что в данном анализе, как и в других случаях использования методов, основанных на количественных данных, для обеспечения его работоспособности необходимы значительные данные хорошего качества, содержащие информацию о цене продажи, о магазинах, осуществляющих продажи, о количестве этих продаж, уровне доступности товара на полке, товарном запасе и другие. Часть этих данных компания-производитель товара может получить, привлекая к сотрудничеству аудиторские и аналитические агентства. Другая часть данных может быть получена в результате налаженного процесса обмена информацией между участниками рынка: розничной точкой/ торговой сетью и поставщиком или производителем. Данные, получаемые в результате этого обмена, должны обладать определенными характеристиками, например:

- регулярность получения, заданная и обговоренная сторонами периодичность (например, ежедневно, раз в неделю, и т.д.);
- непрерывность, для обеспечения целостности картины, построенной на этих данных;
- достаточная гранулярность, то есть с возможностью разбивки данных до уровня конкретного магазина, конкретного дня, конкретного SKU. Укрупненные агрегированные данные тоже могут быть положены в основу модели, но точность будущего результата может быть ниже.
- точность, непротиворечивость и другие.

Кроме того, для анализа и будущего использования опции прогнозирования эластичности спроса по цене необходимо накопить достаточно данных с разными значениями цены в разные периоды времени. Если, допустим, цена была одинаковой в течение всего времени, то модель не сможет корректно описать зависимость продаж от цены на полке. То же справедливо и для анализа влияния других факторов.

Необходимо также упомянуть о том, что при построении регрессионной модели мы исходим из того, что другие факторы, которые могли сильно повлиять на результат продаж, в течение времени были неизменными, например, рекламная поддержка во время всего периода была одинаковой, конкурентная обстановка была равномерной и так далее. Если это не так, то результаты анализа могут быть недостоверными, а сильно влияющие факторы необходимо включить в модель, а также построить новую регрессию с учетом новых факторов.

Заключение

Таким образом, в ходе исследования была построена многофакторная регрессионная модель, в качестве факторных признаков которой были использованы POS данные из розничных точек продаж. Полученные регрессионные уравнения позволяют установить факт и уровень влияния таких факторов, как полочная цена товара и уровень суммарной дистрибуции на уровень продаж. Как следующий шаг, это позволит компаниям-производителям или ритейлерам осуществлять стратегический выбор в среднесрочной перспективе, фокусируясь на том факторе, который в данный момент имеет наибольшее влияние – например, сосредоточиться на расширении суммарной дистрибуции при минимальных изменениях цены. Кроме того, при решении ряда задач об ограниченных ресурсах, например, четко определенном уровне доступного полочного пространства, построение уравнений регрессии для разных продуктов позволит сделать выбор в пользу более выгодного для компании товара. Также регрессионный анализ на основе POS данных обладает хорошим потенциалом при прогнозировании продаж в краткосрочной и среднесрочной перспективе, со-

ответственно, может использоваться в этом качестве. Таким образом, регрессионный анализ с вовлечением POS данных позволит решать большое количество практических задач, начиная с прогнозирования продаж при заданном уровне факторов до выбора стратегии развития конкретного продукта, а также значительно расширит арсенал аналитических инструментов компаний-участников рынка FMCG.

Библиография

- [1] Котлер Ф. 300 ключевых вопросов маркетинга: отвечает Филип Котлер. М.: Олимп-Бизнес, 2006. 198 с.
- [2] Доброхотов А.В., Волынский В.Ю. Современное состояние методологии прогнозирования объемов продаж готовой продукции // Сборник научных трудов вузов России "Проблемы экономики, финансов и управления производством". 2010. № 29. С. 192-201.
- [3] Шанин И.И., Атаэзгаев Ы.Б. Методы прогнозирования объемов продаж продукции // Международный студенческий научный вестник. 2018. № 2. С. 45.
- [4] Армстронг Дж.С. Прогнозирование продаж. Маркетинг. СПб.: Питер, 2002. С. 351-368.
- [5] Мхитарян С.В., Данченко Л.А. Прогнозирование продаж с помощью адаптивных статистических методов // Фундаментальные исследования. 2014. № 9-4. С. 818-822.
- [6] Юшин А.А. Прогнозирование продаж мебельной продукции // Вестник Поволжского государственного университета сервиса. Серия: Экономика. 2015. № 1(39). С. 148-155.
- [7] Карлов А.М., Невротов Л.К. Использование метода корреляционно-регрессионного анализа при прогнозировании объема продаж на примере пищевой промышленности региона // Балтийский экономический журнал. 2019. № 2(26). С. 100-107.
- [8] Ребенок И.И., Малыгина М.П. Методы интеллектуального анализа и прогнозирования данных стационарной розничной торговой сети // Современные проблемы науки и образования. 2014. № 3. С. 151.
- [9] Mou S., Davia D.R., Nicole D. Retail Store Operations: Literature Review and Research Directions // European Journal of Operational Research. 2017. Vol. 265(2). Pp. 399-422. (На англ.). DOI: 10.1016/j.ejor.2017.07.003
- [10] Sajawal M., Usman S., Sanad A.H., Hayat A., Ashraf M.U. A Predictive Analysis of Retail Sales Forecasting using Machine Learning Techniques // Lahore Garrison University Research Journal of Computer Science and Information Technology. 2022. Vol. 6(04). Pp. 33-45. (На англ.). DOI: 10.54692/lgurjcsit.2022.0604399
- [11] Jain A., Menon M.N., Chandra S. Sales Forecasting for Retail Chains // San Diego, California: UC San Diego Jacobs School of Engineering. 2015. (На англ.).
- [12] Odegua R. Applied Machine Learning for Supermarket Sales prediction //Project: Predictive Machine Learning in Industry. 2020. (На англ.).
- [13] Anna-Lena B., Stefan M. Safety Stock Planning under Causal Demand Forecasting // International Journal of Production Economics. 2012. Vol. 140(2). Pp. 637-645. (На англ.). DOI: 10.1016/j.ijpe.2011.04.017
- [14] Paul F., Neil B., Phillip E.P., David J.R. Marketing Metrics the Manager's Guide to Measuring Marketing Performance. Third edition. Publisher: Pearson, 2016. 427 p. (На англ.).
- [15] Питер Б., Эндрю Б., Питер Г. Практическая статистика для специалистов Data Science. СПб: БХВ-Петербург, 2021. 352 с.

[16] Эндрю Г., Дженнифер Х., Аки В. Регрессия: теория и практика. С примерами на R и Stan. М.: ДМК-Пресс, 2022. 748 с.

References

- [1] Kotler Ph. According to Kotler: The World's Foremost Authority on Marketing Answers Your Questions. АМАСОМ, 2005. 168 p.
- [2] Dobrohotov A.V., Volynskiy V.U. Sovremennoe sostoyanie metodologii prognozirovaniya ob'yomov prodazh gotovoi produktsii [Current state of the methodology for forecasting sales volumes of finished products // A Collection of Scientific Works of Russian Universities "Problems of Economy, Finance and Production Management". 2010. Vol. 29. Pp. 192-201. (In Russ.).
- [3] Shanin I.I., Ataekgaev Y.B. Methods of Forecasting of Sales Volumes of Production // European Student Scientific Journal. 2018. Vol. 2. Pp. 45. (In Russ.).
- [4] Armstrong D.S. Prognozirovaniye prodazh [Sales forecasting]. Marketing. SPb.: Piter, 2002. Pp. 351-368. (In Russ.).
- [5] Mkhitaryan S.V., Danchenok L.A. Sales Forecast Using Adaptive Statistical Methods // Fundamental Research. 2014. Vol. 9-4. Pp. 818-822. (In Russ.).
- [6] Yushin A.A. Sales Forecasting for Furniture Products // Bulletin of the Volga State University of Service. Series: Economy. 2015. Vol. 1(39). Pp. 148-155. (In Russ.).
- [7] Karlov A.M., Nevrotov L.K. Using the Method of Correlation and Regression Analysis in Forecasting of Sales on the Example of the Food Industry in the Region // Baltic Economic Journal. 2019. Vol. 2(26). Pp. 100-107. (In Russ.).
- [8] Rebenok I.I., Malykhina M.P. Data Mining Methods and Forecasting Data in Fixed Retail Chain // Modern Problems of Science and Education. 2014. Vol. 3. Pp. 151. (In Russ.).
- [9] Mou S., Davia D.R., Nicole D. Retail Store Operations: Literature Review and Research Directions // European Journal of Operational Research. 2017. Vol. 265(2). Pp. 399-422. DOI: 10.1016/j.ejor.2017.07.003
- [10] Sajawal M., Usman S., Sanad A.H., Hayat A., Ashraf M.U. A Predictive Analysis of Retail Sales Forecasting using Machine Learning Techniques // Lahore Garrison University Research Journal of Computer Science and Information Technology. 2022. Vol. 6(04). Pp. 33-45. DOI: 10.54692/lgurjcsit.2022.0604399
- [11] Jain A., Menon M.N., Chandra S. Sales Forecasting for Retail Chains // San Diego, California: UC San Diego Jacobs School of Engineering. 2015.
- [12] Odegua R. Applied Machine Learning for Supermarket Sales prediction // Project: Predictive Machine Learning in Industry. 2020.
- [13] Anna-Lena B., Stefan M. Safety Stock Planning under Causal Demand Forecasting // International Journal of Production Economics. 2012. Vol. 140(2). Pp. 637-645. DOI: 10.1016/j.ijpe.2011.04.017
- [14] Paul F., Neil B., Phillip E.P., David J.R. Marketing Metrics the Manager's Guide to Measuring Marketing Performance. Third edition. Publisher: Pearson, 2016. 427 p. (In Russ.).
- [15] Peter B., Andrew B., Peter G. Practical Statistics for Data Scientists 2nd edition. Publisher: O'Reilly, 2020. 360 p.
- [16] Andrew G., Jennifer H., Aki V. Regression and Other Stories. Publisher: Cambridge University Press, 2020. 548 p. (In Russ.).

Информация об авторе / About the Author

Юлия Вячеславовна Напольская – бизнес-аналитик, Проктер энд Гэмбл, Москва, Россия / **Yulia V. Napolskaya** – Business Analyst, Procter and Gamble, Moscow, Russia
 E-mail: jnapolskaya@gmail.com
 SPIN РИНЦ 8536-2940
 ORCID 0009-0009-6072-4581
 Researcher ID KTI-2986-2024

Дата поступления статьи: 08 октября 2024
 Принято решение о публикации: 10 ноября 2024

Received: October 08, 2024
 Accepted: November 10, 2024