

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»**

Кафедра Дифференциальных уравнений и математической
экономики

Исследование продаж прессы в супермаркетах

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студентки 4 курса 441 группы

направление 09.03.03 — Прикладная информатика

механико-математического факультета

Семикиной Ольги Алексеевны

Научный руководитель
профессор, д.ф.-м.н., доцент

А.Ю. Трынин

Зав. кафедрой
зав. каф., д.ф.-м.н., профессор

С.И. Дудов

Саратов 2021

Введение. Исследования в области продаж печатной прессы за последние 5 лет говорят об общем снижении продаж продукции, но также говорят и об увеличении на 9% количества продаж в супермаркетах. В связи с этим, возникает необходимость исследования продаж, то есть выявления факторов, влияющих на объем продаж, а главное, степень их влияния. Объем продаж этого – это главный показатель результатов работы компании. Его можно выражать в фактическом числе единиц реализованной продукции или в суммарной величине средств, поступивших на счет фирмы после продажи товаров.

Актуальность исследуемой темы состоит в том, что величина продаж – часто меняются и является достаточно нестабильна. От величины продаж зависит предложение товаров и услуг на рынке и их ассортимент, количество продавцов и их конкуренция между собой.

Целью бакалаврской работы является исследование продаж прессы в супермаркетах с помощью коэффициента корреляции Пирсона, различных видов уравнений регрессий, а также их характеристик на основе данных "Пресса-Поволжья"

В соответствии с целью бакалаврской работы можно определить **задачи**. В их состав входят:

1. анализ рынка прессы;
2. выявление тенденций развития рынка прессы;
3. рассмотрение отдельных показателей, таких как общие, средние продажи объемы продаж, а также количество выпусков прессы;
4. описание и изучение статистических методов определения зависимости;
5. написание программного кода на языке Java;
6. выполнение корреляционного и регрессионного анализов;
7. вывод заключений о зависимости показателей.

Основное содержание работы. Работа состоит из введения, пяти разделов, заключения, списка использованных источников, содержащего 17 наименований. Объем работы составляет 56 страниц, включая 8 таблиц.

Введение включает в себя обоснованность, актуальность выбранной темы, основные задачи и цели, а также методы, применяемые в работе.

В первом разделе приводится аналитический материал, взятый из отраслевых докладов Роспечати за 2018, 2019 года, с помощью которого делаются выводы об актуальности темы, а также выбираются тематические группы изданий для дальнейшего исследования.

Методов распространения печатной прессы достаточно много. Это продажа в киосках, в супермаркетах, на почте, а также продажа "с рук". Но в настоящее время из-за распространения цифровых технологий увеличивается число изданий, которые частично или полностью переходят на электронный формат свободного выпуска прессы, что уменьшает издержки на печать и распространение, а также увеличивает прибыль от рекламы, размещенной на сайтах изданий.

По данным из проведенного исследования Союза предприятий печатной индустрии о продажах за 2014 - 2019 г.г. было установлено, что за данный период общие продажи прессы (розница + подписка) сократились на 45%, а среднегодовое уменьшение составило 9%. Потери в тематических сегментах изданий составили от 2,5% до 16%. Причем наименьшие потери понесли издания групп 'женские', 'развлекательные', 'кулинарные издания' - 7%. Самыми стабильными оказались группы 'сад', 'огород' - 2,5%. Не смотря на общее сокращение продаж на рынке, а так же резкое увеличение цен на данные категории, эти группы менее всего подверглись сокращению, что свидетельствует о наличие других факторов влияющих на продажи.

Сокращение продаж также объясняется резким уменьшением количества числа киосков за данный период. Так, например, в 2014 году их число составляло 41 тыс. объектов, в 2018 году - 35,2 тыс. объектов, а к концу 2019 года - 16 тыс. объектов. Но такое резкое снижение числа киосков сглаживалось увеличением числа стоек с прессой в супермаркетах. По данным Ассоциации распространителей печатной продукции, проданные в них тиражи выросли в рублевом выражении на 9%. Это говорит об актуальности и необходимости проведения исследований в данном развивающемся сегменте продаж.

В сложившейся ситуации государство решило оказать поддержку печатной индустрии. Так, в октябре 2019 года президент Владимир Путин поручил правительству подготовить предложения по поддержке печатных СМИ, среди которых стабилизация цен на бумагу и увеличение доли разрешенно-

го списания непроданных экземпляров. Также в декабре Союз предприятий печатной индустрии и Ассоциация компаний розничной торговли подписали соглашение о сотрудничестве, направленное на стабильное обеспечение потребителей печатной периодикой. На основе этого можно сказать, что в ближайшие годы рынок печатной продукции будет существовать и развиваться, не смотря на переход в электронный формат изданий.

Кроме этого, Правительство России планирует актуализировать нормативы минимальной обеспеченности населения продукцией периодической печати.

По данным за 2018 год газетный рынок разделяют на 3 группы: местные, региональные, общероссийские газеты. Можно увидеть, что на рынке преобладают издания местного и регионального уровней. Из чего следует, что изучать необходимо более локальные уровни.

Подводя общий итог ситуации на российском рынке распространения периодической печати, можно с уверенностью сказать, что реальная причина сокращения объемов реализации тиражей кроется не столько в «устаревании» печатной прессы и потере к ней интереса населения, пандемии коронавируса, сколько в резком сокращении числа мест шаговой доступности, где прессу можно купить в розницу. При этом стоимость подписки выросла настолько, что все меньше граждан могут ее себе позволить.

Во втором разделе приводится теоретический материал о корреляционном анализе.

Корреляционный анализ - установление формы связи между варьирующими признаками (линейная, нелинейная) и измерению тесноты этой связи, определение направления (положительное или отрицательное), и проверка уровня значимости полученных коэффициентов корреляции. Корреляционный анализ, на ряду с другими статистическими методами, построен на применение вероятностных моделей, отображающих поведение изучаемых признаков в данной генеральной совокупности, из которой могут быть получены экспериментальные значения x_i и y_i . Коэффициент корреляции Пирсона

расчитывается следующим способом:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{X})^2 (y_i - \bar{Y})^2}}. \quad (1)$$

Для интерпретации результатов можно воспользоваться шкалой Чеддока (Таблица 1). Эта шкала — довольно условная, однако благодаря ей становится возможным «перевести» численное (количественное) значение в качественную характеристику.

Таблица 1 — Шкала Чеддока

Коэффициент корреляции	Связь
0,1 - 0,3	слабая
0,3 - 0,5	умеренная
0,5 - 0,7	заметная
0,7 - 0,9	высокая
0,9 - 0,99	весьма высокая

Коэффициент корреляции так же, как и среднее арифметическое, изменяется от выборки к выборке при повторных исследованиях. Мерой изменчивости коэффициента корреляции служит ошибка коэффициента корреляции m , находящаяся по следующей формуле:

$$m = \sqrt{\frac{1 - r_{xy}^2}{n - 2}} \quad (2)$$

где n - число коррелирующих пар случайных величин, r_{xy} - коэффициент корреляции.

Зная ошибку коэффициента корреляции, можно вычислить критерий достоверности t - Стьюдента по следующей формуле:

$$t = \left| \frac{r_{xy}}{m} \right| = \left| r_{xy} * \sqrt{\frac{n - 2}{1 - r_{xy}^2}} \right| \quad (3)$$

Полученное значение берется по модулю, то есть всегда является положительным числом. Вычисленное значение критерия достоверности t сравнивается со значением критерия Стьюдента t_{st} , находимом в таблице соот-

ветствующих коэффициентов на пересечении строки с номером $k = n - 2$ и столбца с выбранным уровнем доверительной вероятности α . В расчетах, как правило, значение α принимается равным 0,05. Если $t > t_{st}$, делается вывод о наличии корреляционной связи между величинами. Если же $t < t_{st}$, то связь недостоверна, случайна.

В третьем разделе приводится теоретический материал о регрессионном анализе.

Регрессионный анализ - это статистический метод, устанавливающий количественно форму зависимости двух случайных величин, между которыми существует корреляционная связь. Основное понятие регрессионного анализа - это понятие регрессии. Регрессия - функция, позволяющая по значению одной переменной величины определить средние значения другой величины, связанной с первой корреляционно. Регрессию подразделяют на парную и множественную. Парной регрессией называется модель, выражающая зависимость среднего значения зависимой переменной y от одной независимой переменной :

$$y = f(x) \quad (4)$$

где – зависимая переменная (результативный признак); – независимая, объясняющая переменная (признак–фактор). Парная регрессия применяется, если имеется доминирующий фактор, обуславливающий большую долю изменения изучаемой объясняемой переменной, который и используется в качестве объясняющей переменной.

Множественной регрессией называют модель, выражающую зависимость среднего значения зависимой переменной y от нескольких независимых переменных:

$$y = f(x_1, x_2 \dots x_n) \quad (5)$$

Множественная регрессия применяется в ситуациях , когда из множества факторов, влияющих на результативный признак, нельзя выделить один доминирующий фактор и необходимо учитывать одновременное влияние нескольких факторов.

Наиболее часто применяются следующие модели регрессий:

- прямой: $\hat{y} = b_0 + b_1 * x$;

- гиперболы: $\hat{y} = b_0 + \frac{b_1}{x}$;
- параболы: $\hat{y} = b_0 + b_1 * x + b_2 * x^2$;
- показательной функции: $\hat{y} = b_0 + b_1^2$;
- степенная функция: $\hat{y} = b_0 * x^b$.

Для оценки параметров регрессий, линейных по этим параметрам, используется метод наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака от теоретических значений \hat{y} при тех же значениях фактора x минимальна, т. е.

$$\sum (y - \hat{y})^2 \rightarrow \min \quad (6)$$

Для каждой модели регрессии находятся следующие характеристики и оценки:

- коэффициент детерминации;
- средняя относительная ошибка аппроксимации;
- F - критерий Фишера;
- эластичность;
- точечное прогнозирование.

В четвертом разделе рассказывается о реализации программного кода на языке Java.

Java — это язык программирования общего назначения, который следует парадигме объектно-ориентированного программирования и подходу «Написать один раз и использовать везде». Java используется для настольных, сетевых, мобильных и корпоративных приложений. В этом языке существует большое количество готовых библиотек, которые расширяют возможности использования Java.

В ходе работы были реализованы методы вычисления коэффициента корреляции Пирсона, а также метод, позволяющий на основе характеристик моделей регрессии определять наиболее точную модуль регрессии. На основе найденной модели строится точечной прогнозирование.

Данная программа универсальна и может использоваться для нахождения различных видов зависимостей.

В пятом разделе анализируются полученные результаты после работы программы.

На основе данных по общим продажам и средним продажам с одного пункта продаж за один выпуск газет и журналов на территории Саратовской области, предоставленных ООО “Пресса Поволжья” определим корреляционную зависимость общего и среднего количества продаж газет и журналов от количества выпусков в год. Для более точного определения корреляционной зависимости заранее разделим полученные данные на следующие категории: “Гороскопы”, “Советы для женщин”, “Дачные советы”. Найдем коэффициент корреляции для первой категории продукции на основе данных приведенных в таблице 2.

Таблица 2 — Данные по продажам для категории “Гороскопы”

№	Наименование	Общие продажи	Средние продажи	Кол-во экз в год
1	Зодиакальный гороскоп №1 2018	2849	6,2	1
2	Астросовет №11 2018	1309	2,4	12
3	Астросовет №12 2018	1433	2,6	12
4	Астросовет №1 2019	1181	2,1	12
5	Астросовет №2 2019	1179	2,2	12

Найденный коэффициент корреляции с промежуточными вычислениями для общих и средних продаж приведены ниже. Найдем коэффициент корреляции для второй категории продукции на основе данных приведенных в таблице 3.

Найдем коэффициент корреляции для третьей категории продукции на основе данных приведенных в таблице 4.

Полученные результаты по всем категориям приведены в таблице 5. По вычисленным значениям можно сделать некоторые выводы. Во-первых, коэффициенты корреляции Пирсона, полученные от количества общих продаж за один экземпляр и от средних продаж с одной точки за один экземпляр для всех трех случаев, имеют небольшую разницу, что говорит о равнозначности общих и средних продаж. Во-вторых, на основе вычисленных данных, можно заметить явную зависимость между количеством выпусков в год и общими,

Таблица 3 — Данные по продажам для категории “Советы для женщин”

№	Наименование	Общие про- дажи	Средние продажи	Кол-во экз в год
1	Все для женщин №8 2019	1862	2	24
2	Все для женщин №9 2019	1839	2	24
3	Все для женщин №10 2019	1835	2	24
4	Все для женщин №11 2019	1606	1,7	24
5	Все для женщин №12 2019	1767	1,9	24
6	Добрые советы №12 2018	460	0,8	12
7	Добрые советы №1 2019	433	0,8	12
8	Добрые советы №2 2019	484	0,8	12

Таблица 4 — Данные по продажам для категории “Дачные советы”

№	Наименование	Общие продажи	Средние продажи	Кол-во экз в год
1	Мой прекрасный сад №12 2018	281	0,5	12
2	Мой прекрасный сад №1 2019	394	0,7	12
3	Мой прекрасный сад №2 2019	215	1	12
4	1000 советов дачнику №2 2019	906	1	24
5	1001 советов дачнику №3 2019	1009	1,1	24
6	1002 советов дачнику №4 2019	1041	1,2	24

Таблица 5 — коэффициенты корреляции Пирсона по всем категориям

	“Гороскопы”	“Советы для женщин”	“Дачные советы”
Общие продажи/экз	-0.989033	0.993218	0.981985
Средние продажи с точки/экз	-0.993917	0.98585	0.760886

средними продажами, ранее найденную в для общероссийских данных. Зависимость состоит в том, что чем больше количество выпусков в год отличается от 12, тем больше возрастают общие и средние продажи.

Оптимальная регрессионная модель для данных - квадратичная.

В заключении приведены результаты бакалаврской работы.

Основные результаты

- Проанализирован рынок прессы;
- Определены основные понятия корреляционного и регрессионного анализов;
- Реализован программный код на языке Java;
- На основе вычисленных данных найдена зависимость - чем больше количество выпусков в год отличается от 12, тем больше возрастают общие и средние продажи;
- С помощью программы была определена наиболее подходящая для данных модель регрессии - квадратичная.