**Введение.**

**Насколько актуальная тема исследования.**

Прогнозирование, как таковое, появилось очень давно. Ещё с самых первобытных времён людям было интересно, будет ли завтра дождь или пойдёт ли завтра снег. Конечно, первые люди не имели никакого представления о том, как устроена погода и прочее. Поэтому появились боги, благодаря которым, как думали люди, всё на свете и случается. Так зародились мифология, философия и другие, которые могли объяснить и подсказать, что же нужно делать дальше.

Время шло и люди заметили, что окружающий их мир поддаётся определённым законам и последовательностям. Например, что солнце встаёт каждое утро с одной и той же стороны, а садится с противоположной. Что происходит это всегда примерно в одно и то же время. Что привело к тому, что люди стали ещё больше изучать окружающий мир и пытаться понять его.

Когда появились первые зачатки экономики и основы торговли, стало понятно, что и количество продаваемого продукта и количество произведённого продукта можно как-то регулировать.

В какой-то момент люди осознали, что не имеет смысла производить 100 пар обуви, когда в поселении живёт всего 15 человек, ибо это не имеет смысла. 15 пар кто-то купит, а остальные либо останутся стоять и ждать своего покупателя, либо же кто-то проезжий когда-нибудь купит их, что очень навряд ли. Но это в самом простом примере, если рассматривать данный вопрос в огромных масштабах, можно понять насколько он становится актуальным. Если одна произведённая деталь стоит 5 у.е., а продаётся она за 7 у.е., но покупают её редко, появляется вопрос, а стоит ли эту делать производить и насколько много на ней можно заработать, а не наоборот, потерять.

Поэтому, управление доходами, всегда было актуальной и, с большой вероятностью, останется одной из самых актуальных прикладных областей исследования операций.

Сейчас, управление доходами – это набор стратегий и подходов, позволяющий распределить имеющиеся ресурсы так, чтобы наилучшим образом удовлетворить спрос на них и получить максимальную прибыль. Благодаря этому, прогноз спроса является очень важной составляющей многих моделей принятия решений, которые используются в области управления доходами.

В сфере услуг оптимальное управление ресурсами можно достичь с помощью контроля наличия в продаже различных продуктов. То есть, если знать, где нужно добавить товара, а где снизить, можно добиться наилучшего получения прибыли. И любая стратегия продаж зависит от спроса на товар. Очевидно, если неточно спрогнозировать продажи, это может привести к огромным убыткам.

И именно по этой причине, так востребованы подходы к моделированию спроса, не ограниченного пределами бронирования и вместимостью. Такой спрос имеет название «общий». Сложность прогнозирования такой спроса в том, что исторические данные – это уровень продаж, а не сам спрос. То есть, возможно, желание купить было у 100 человек, а деталей было всего 30. Остальные 70 человек остались без деталей. То есть спрос больше, чем сами продажи. А это отобразить на каких-то данных невозможно, тк люди, узнав, что деталей нет, обращаться не будут. Таким образом, исторические данные о продажах будут соответствовать цензурированному спросу. Процедура оценивания параметров распределения такого спроса по исходной выборке имеется название – восстановление общего неограниченного спроса, а сами методы для него имеют название – методы восстановления. Довольно трудно измерить общий неограниченный спрос, но при этом, некоторые исследователи утверждают, что использование этих методов улучшает прогноз и увеличивает прибыль. Так, было обнаружено разными учёными, что влияние восстановления спроса на доход может быть от 2 процентов до 12.

Дипломная работа посвящена теме исследования применимости метода бутстрап и других непараметрических методов восстановления распределения спроса на товары. В качестве примера взяты данные магазинов с продажами.

**Степень разработанности научной проблемы.**

Существует очень много научных трудов, которые поднимают вопрос восстановления общего неограниченного спроса, оценке и прогнозированию. В частности, в работах идёт обсуждение новых методов восстановления, сравнение точности полученных оценок и анализ влияния выбранных подходов на доход. Но в этих работах мало уделяется внимание сравнению статических свойств, полученных оценок, чувствительность методов к разнообразным ситуациям. Так же существуют методы, которые не воспринимают неполноту информации о цензурировании.

Благодаря этому всему, данные вопросы являются недостаточно раскрытыми и актуальными. Это и послужило сделать выбор такой темы дипломной работы.

Целью дипломной работой является: Исследовать применимость метода бутстрап и других непараметрических методов для восстановления распределений спроса на товары.

1. Проведён анализ недостатков, преимуществ и особенностей, в частности, метода бутстрап, и других непараметрических методов восстановления распределений спроса
2. Изучение асимптотических свойств метода бутстрап
3. Предложен и обоснован подход восстановления спроса с помощью метода бутстрап.
4. Проведены расчёты, с помощью которых можно сравнить точность оценок, полученных различными методами восстановления спроса.

**Предмет исследования:**

Восстановление спроса на товар в магазинах по цензурированным данным в рамках управления доходами.

**Объект исследования:**

Магазины где-то там в какой-то там стране с какими-то там данными

**Глава 1.**

1.1 **Управление доходами. Основные понятия.**

В последнее время, такой понятие как «управление доходами» стало не просто словосочетанием, а является прикладной областью исследования операций. Во всём мире, от малых фирм до огромных, происходит повседневных подсчёт прибыли, подсчёт произведённого и проданного товара, и прочего. В связи с техническим процессом это стало делать очень удобно, есть где и как хранить информацию. И появились средства для того, чтобы что-то с этой информацией делать. Появились целые группы аналитиков, маркетологом, менеджеров и прочих, которые хотят увеличить прибыл свою и прибыль компании, в которой они работают. Возникает вопрос, как же это сделать. Естественно, можно спрогнозировать то или иное поведение на рынке и заранее знать, что лучше и как лучше производить, продавать и покупать. Но это же будущее, которое просто так взять и спрогнозировать нельзя, естественно будет какая-то ошибка, но, чем точнее будет прогноз, тем наибольшую прибыль получат все, кто в этом участвовал. Поэтому и начали придумать различные методы прогнозирования спроса, стали их улучшать, исследовать и рассматривать, чтобы добиться наименьшей ошибки.

Существует несколько определений управления доходами. Например, это применение особых к подходов, которые предсказывают действия потребителей в какой-то из сфер и тем самым улучшить цены и тем самым максимизировать свою прибыль. Или же, управление доходами, это – умение применить знания и стратегии для того, чтобы правильно продать правильный товар, правильному человеку, в правильном месте и в правильное время.

Управление доходами включает в себя:

1. Ценовые решения. Как именно выставлять цены, как ими варьировать и прочее.
2. Структурные решения. Под каким форматом продаж выставлять продукты, будь то скидки или наоборот, завышение цен.
3. Количественные решения. Продавать или запрещать продажу. Распределение имеющихся ресурсов между различными сегментами.

Что из этого является наиболее важным и на что нужно акцентировать внимание зависит от сферы деятельности, от конкретной ситуации на рынке и от многих других факторов.

Имеется общая схема работы системы управления доходами, она заключается в том, что существует бесконечный цикл, который состоит из этапов, повторение которого происходит через определённое количество времени. Как часто этот происходит, зависит от многих факторов, таких как: скорость изменения условий бизнеса, объём данных, методы используемые для оптимизации и прогнозирования, значимость решений и другие. Из этого всего можно выделить основные шаги цикла работы системы управления доходами:

1. Во-первых, нужны данные, от которых нужно отталкиваться. Поэтому первым пунктом является сбор данных. Данные могут быть разными, это и исторические данные о спросе, причинных факторах и даже цены. Так как этот пункт очень важен и без него нельзя осуществить последующие пункты, нужно с особой серьёзностью относиться к источнику данных и наилучшим образом им воспользоваться.
2. Во-вторых, это оценка и прогнозирование. Оценка распределения спроса, оценка распределения спроса с использованием методов восстановления, прогнозирование спроса с полученными параметрами. Так же прогнозирование иных не менее важных показателей, например, возвраты, неявки и прочее.
3. В-третьих, все те данные, что получены, нужно оптимизировать. Поэтому, третьим этапом является оптимизация. Оптимизация – определение наилучших значений для параметров регулирования и контроля. То есть, рассчитываются оптимальные пределы создания продукта с учётом выбранной стратегии продаж, выбраны подходящие цены на продукцию, уровень скидок и наценок.
4. В-четвертых, последний шагом является контроль. То есть, контролирование продажность продукции с использованием тех знаний, что были получены на предыдущих шагах.

Таким образом получается, что общая схема работы системы управления доходами выглядит так: собранные данные приходят в блок прогнозирования и оценивания, полученные прогнозируемые значения становятся входными данными для определения оптимальных параметров, далее оптимальные параметры регулирования используются для контроля и обработки запросов. Затем, полученные данные отправляются в историческую базу. После чего цикл замыкается и начинается сначала. Эта процедура указана на Рисунке 1.

# 

**Рисунок - 1 Основные этапы работы системы управления доходами**

**1.2 Оценка и прогнозирование спроса.**

Оценка и прогнозирование является обширной темой, которая охватывает многие области научной деятельности, такие как статистику и экономику, технические и компьютерные науки. Некоторые методы основаны на точных статических и математических утверждениях, некоторые эвристические.

Так как под оценкой зачастую подразумевается нахождение параметров модели, которые наилучшим образом описываются заданный наблюдаемый набор данных, а прогнозированием является предугадывание будущих ненаблюдаемых величин, то в контексте системы управления доходами оценка - подборка параметров прогнозной модели, используется не часто, а вот прогнозирование – использование оцененной модели для предугадывания значений – используется часто.

Выбор подходов и методов для оценки и прогнозирования довольно обширный. Самые распространённые непараметрические методы прогнозирования – это непараметрический метод наименьших квадратов с оцениванием точности прогноза, метод Эфрона, метод Виллемейна и другие. А методы оценивания – это непараметрические методы доверительного оценивания точки наложения (встречи) двух временных рядов, непараметрических оценок плотности распределения и другие. Некоторые ученые доказали, что если улучшить хотя бы на 10 процентов точность прогноза спроса, то доход может увеличиться от 0.5 процентов до 3 процентов.

Цензурирование спроса является одним из самых сложных вопросов при прогнозировании спроса в системе управления доходами, так как имеются различная продукция, которая производится в определённом количестве и при если будет полностью выкуплен ассортимент, это не будет значить, что спрос имеет значение ровно столько, сколько было продано. Исторические данные же наоборот, должны отображать только реальный спрос, а вместо этого отражают лишь фактические продажи, что не является наилучшим показателем в прогнозировании спроса.

Использование цензурированных данных может привести как к переоценке, так и к недооценки спроса, а это в свою очередь может очень сильно повлиять на доход, то есть привести к потере дохода. Чтобы избежать этой проблемы, сначала нужно оценить параметры распределения общего спроса по цензурированным данным и только потом проводить прогнозирование. Процедура оценивания параметров распределения общего спроса по исходной цензурированной выборке имеет название восстановление общего неограниченного спроса, а методы называются – методы восстановления.

Очень сложно измерить общий неограниченный спрос, но многие исследователи утверждают, что если его использовать, можно сделать прогноз более точным и следовательно поиметь большую прибыль.

**1.3 Цензурирование спроса. Основные понятия и примеры**

Можно рассмотреть простейшую схему управления доходами на примере продавца и продуктов питания. Продукты питания имеют свойство портиться, что-то портится за день, что-то за неделю, и в этой ситуации продавцу нужно установить цену так, чтобы за период продаж продукты не испортятся, а покупатели в полной мере смогут купить все те продукты, что пожелают. Если поставить слишком низкую цену продуктам, то они будут быстро раскуплены, но доход от них будет крайне мал или вообще будет отсутствовать. При этом, так как все продукты будут раскуплены, а новые приехать не успеют, вновь пришедшие покупатели не смогут получить желаемый продукт и тем самым, возможно, больше никогда не придут в этот магазин. Если же поставить цену слишком высокую, то есть вероятность, что многие продукты вообще не будут куплены, так как какие-то покупатели решат купить аналогичный продукт за меньшую цену в другом магазине, и оставшиеся не проданные продукты будут попросту выкинуты, так как продавать испорченные продукты нельзя. Следовательно, продавец потеряет очень много прибыли. В этом и заключается суть, чтобы понять, какую именно цену нужно выставить на продукты, чтобы извлечь наибольшую выгоду для продавца.

Решение продавец будет принимать на своём опыте и своём прогнозе на спрос. Ему требуется спрогнозировать количество востребованных продуктов на разных ценовых уровнях и в последствии регулировать цены для большой эффективности. Пока время идёт, ещё не купленные продукты начинают портиться и наступает точка, когда нужно либо быстро продать, либо придётся выбрасывать. Продавец может выставить скидку на эти продукты, за частую в наших магазинах так это и происходит. Если посмотреть на продукты со скидкой и взглянуть на их срок годности, можно увидеть, что срок истечёт через короткий срок. Тем самым вроде и покупателю хорошо, он купил хороший продукт с меньшей ценой, и продавец не потерял прибыль полностью. Если же товар не продаётся так быстро, как ожидал этого продавец, он может снизить цену, чтобы извлечь наибольшую прибыль. Но намного сложнее что-либо сделать, когда товар наоборот уходит к покупателям очень быстро и, возможно, его даже не хватает. Тогда продавцу труднее сделать выбор, так как сегодня товар ушёл быстро, а завтра – неизвестно. И что было бы, если бы у него ещё больше было товара – не понятно. И в этом смысле, данные продавца являются цензурированными.

Пример продавца является простейшей иллюстрацией схемы работы управления доходами и её проблему оценивания общего спроса цензурированных данных. Таким образом, можно, да и даже нужно, воспользоваться определёнными подходами, чтобы преобразовать цензурированные данные в более точные оценки общего неограниченного спроса. Множество методов существует для восстановления данных по цензурированным данным. Есть простые, есть сложные, которые подразумевают полную оценку вероятностного распределения.

*Определение цензурированного спроса.*

Процесс сбора исторических данных может быть собран как с истинными данными, так и с цензурированными. С истинными данными происходит тогда, когда сколько людей приобрело тот или иной продукт, столько в историю данных и записалось, цензурированные данные получаются тогда, когда в какой-то момент времени продукция закончилась, а люди продолжали приходить и желали приобрести товар, но, так как у них не удавалось этого сделать, в исторические данные попадал лишь тот максимум, что был приобретён.

Исходя из этого, спрос является цензурированным, если спрос на товар больше, чем бронированное количество товара в рассматриваемый момент времени. Это можно проиллюстрировать в таблице формирования цензурированной выборки данных по спросу. В этой таблице есть три строки: Общий спрос – количество людей, желающих приобрести товар, предел бронирования – количество имеющегося товара, реальные продажи – те продажи, что получилось продать. Серым закрашены те ячейки, где реальные продажи являются цензурированными из-за того, что продать больше нет возможности, так как имеющегося товара меньше, чем общего спроса.

Таблица 1 - Пример формирования цензурированной выборки для данных спроса.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Общий спрос | 12 | 11 | 9 | 18 | 7 | 20 |
| Предел бронирования | 9 | 20 | 10 | 17 | 5 | 25 |
| Реальные продажи | 9 | 11 | 9 | 17 | 5 | 20 |

*Определение предела бронирования и уровня защиты.*

Главной задачей систем управления доходами является увеличение и максимизация прибыли за счёт правильного распределения имеющихся ресурсов. Это делается благодаря пределов бронирования и уровням защиты.

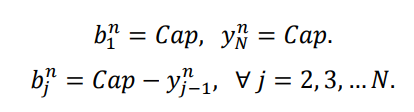
Уровень защиты является одним из самых главных параметров, который характеризует стратегию продаж по ценовым классам.

Существует два типа уровни защиты – это невложенный и вложенный. Невложенный уровень подразумевает, что каждому ценовому классу соответствует определённый сегмент в общей вместимости, который можно продать только покупателям этого класса, так как продавать высшим по цене класса становится невыгодно из-за того, что позиция будет закрыта сразу же после приобретения всех элементов выделенных под этот сегмент, несмотря на то, проданы ли места в остальных классах или нет.

Вложенная структура подразумевает, что уровень защиты Yj определяется для совокупности классов, где начиная с j и выше, и есть количество продуктов, зарезервированных для продажи определённым классам. Уровень защиты для самого нижнего класса является количеством всех продуктов, так как все продукты в сумме и будут соответствовать ей.

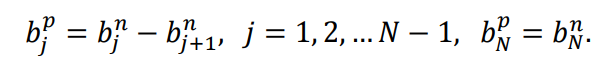
Предел бронирования так же характеризует стратегию продаж по ценовым классам. При заданной вместимости продукции в магазине (Cap), предел бронирования bj для определённого ценового класса j, является максимальная количество продукции, которые продавец готов продать за определённую цену определённому классу. После продажи этот класс закрывается. При невложенной структуре предел бронирования является таким же, как и уровень защиты. При вложенной структуре, предел бронирования и уровень защиты образуют вложенную иерархическую структуру. Предел бронирования является полной вместимостью для самого высокого класса. При такой структуре не возникает таких проблем, как когда более высокие классы уже раскуплены, а нижние ещё имеются.

Для вложенных классов предел бронирования и уровень защиты связаны соотношением:



Вложенный предел бронирования (и уровень защиты ), зная эти параметры, можно определить минимальное и максимальное количество продуктов, которые стоит зарезервировать на основе прогноза спроса в этом классе.

Предел бронирования для невложенных и вложенных ценовых классов связаны соотношением:



**1.4 Вывод**

В этой главе была рассмотрена одна из наиболее важных прикладных областей исследования операций, а именно управление доходами. Управление доходами – это особые тактики и подходы, которые позволяют предугадать поведение потребителей на определённом рынке и оптимизировать ценовую политику и доступность определённого вида товаров для потребителей, с целью максимизации прибыли.

В сфере услуг всё чаще применяются тактики и подходы управления доходами. Для этого используются современные компьютерные технологии, целые штабы аналитики и сбора данных, и прочие другие не менее важные составляющие. В этой главе разобрана общая схема работы системы управления доходами. Описаны главные шаги для успешного функционирования, в частности:

1. Сбор исторических данных
2. Оценка и прогнозирование имеющихся данных.
3. Оптимизация полученных данных
4. Контроль тех данных, что получили на предыдущих шагах.

Один из самых важных составляющих блока оценки и прогнозирования является процесс восстановления спроса по цензурированным данным.

Цензурированные данные возникают из-за того, что реальный спрос может быть намного больше, нежели реальное количество продукции. Количество продукции может быть меньше реального спроса, так как на продукцию установлены ограничения, а именно – пределы бронирования. Когда достигается этот предел, продажи прекращаются и как следствие, информация о реальном спросе теряется. Если не брать этот момент во внимание, последующие попытки спрогнозировать спрос по историческим данным окажется неверным или же будет иметь большую ошибку. Из этого вытекает, что продавец или производитель будет терять прибыль, что не есть хорошо. Эту проблему позволяет решить процесс восстановления спроса цензурированных данных. Таким образом, методы, позволяющие восстановить цензурированные данные обладают высокой значимость, так позволяют увеличить точность прогноза, а вместе с ней и максимизировать прибыль.

**Глава 2. Непараметрические методы восстановления распределений спроса.**

**2.1 Математическая постановка задачи восстановления спроса.**

Имеется случайная величина Х, её функция распределения равна Fx(a1,…,an), a1,…,an – параметры распределения. Данная переменная описывает общий спрос. Так же есть выборка из n наблюдений случайной величины

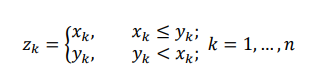
𝑥̅ = (𝑥1,… 𝑥𝑛).

Случайная величина Y является ограничением этого спроса. Функция распределения 𝐹𝑌(𝑏1, … , 𝑏𝑠).

X и Y независимые случайные величины и они могут подчиняться как непрерывным законам распределения, так и дискретным.

Так же рассмотрим случайную величину, которая описывает величину продаж: Z = min(X, Y)

Тогда z1 = min(x1, y1), …, zn = min(xn, yn), где = (z1, …, z2) – является цензурированной выборкой, которая построена на основе выборки случайной величин X. Если xk ≤ yk, то значения являются нецензурированными, если же yk < xk, тогда yk = xk и xk будут цензурированными.



То есть, если взглянуть на задачу восстановления общего неограниченного спроса по цензурированной выборке с математической точки зрения, то смысл состоит в оценке параметров распределения a1, …,an случайной величины Х по цензурированной выборке = (z1, …, z2).

В дальнейшем может помочь информация:

1. Известен закон распределения случайной величины Х
2. Для некоторых значений известно, цензурированы данные или нет. То есть имеется какой-либо индикатор или флаг, который отвечает за цензурированность данных.
   1. **Методы бутрстрап.**

Такая методика, как бутстрап, была впервые предложена американским статистиком Брэдли Эфроном в 1977 году. Вся суть метода состоит в том, чтобы построить эмпирическое распределение, то есть получить приближение теоретической функции распределения, которое построено с помощью выборки из него. Делается это с помощью многократной генерации выборок из исторических данных, которые помогают смоделировать симуляцию неизвестного распределения на основе бутстрапированных выборок.

Существует блочный бутстрап и бутстрап максимальной энтропии.

**Блочный бутстрап.**

Суть блочного бутстрапа скользящих блоков состоит в том, чтобы из выборки выбирать не отдельные значения, а группу значений. Выборка для блочного бутстрапа формируется следующим образом:

1. X = {X1,…, Xn} – выборка наблюдаемых данных.
2. Для некоторой заданной величины L∈[1,n) формируются блоки

Bi = (Xi, …,Xi+L-1), где i ∈[1,n-L+1)

1. Из распределения всех величин i выбирает b количество блоков, чтобы bL было не меньше длины реальной выборки.

**Бутстрап максимальной энтропии**

Суть бутстрапа максимальной энтропии состоит в том, чтобы построить определённое количество точек в окрестностях искомых данных так, чтобы конечная выборка максимально походила на искомую. Формируется следующим образом:

1. X = {X1,…, Xn} – выборка наблюдаемых данных.
2. X1,…, Xn упорядочены по возрастанию и на каждом интервале оценивается плотность максимальной энтропии.
3. Из равномерного распределения на интервале [0,1] случайно выбираются значения случайных величин, которые не превышают максимальное значение в выборке(квантилей).
4. Полученную выборку упорядочиваем.