

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «КубГУ»)

Кафедра прикладной математики

Отдел
В.А.

КУРСОВАЯ РАБОТА

ПОСТРОЕНИЕ МАТЕМАТИЧЕСКОЙ И ИНФОРМАЦИОННОЙ МОДЕЛЕЙ НА ОСНОВЕ ДЕРЕВЬЕВ КЛАССИФИКАЦИЙ

Работу выполнила



(подпись, дата)

Н.С. Бож

(инициалы, фамилия)

Факультет компьютерных технологий и прикладной математики курс 3

Направление 09.03.03 Прикладная информатика

Научный руководитель,

к. п. н.



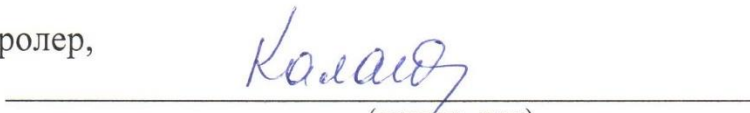
(подпись, дата)

В.А. Акиншина

(инициалы, фамилия)

Нормоконтролер,

к. ф. – м. н.



(подпись, дата)

Г.В. Калайдина

(инициалы, фамилия)

Краснодар

2017

Содержание

ВВЕДЕНИЕ	3
1.1 Статистика	4
1.2 Программа «Statistica6».....	4
1.3 Прикладная Статистика.....	6
2 Анализ Предприятия.....	9
2.1 Исходные Данные	9
2.2 Описательные Статистики	10
2.3 Корреляционная Матрица.....	10
2.4 Множественная Регрессия	14
2.5 Дисперсионный Анализ.....	16
2.6 Кластерный Анализ	16
ЗАКЛЮЧЕНИЕ	20
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	21

ВВЕДЕНИЕ

Статистика – наука, которая изучает количественную сторону массовых социально-экономических явлений в неразрывной взаимосвязи с их качественной стороной, а также количественное выражение закономерностей развития процессов в конкретных условиях места и времени. В данной работе был произведен анализ предприятия при помощи математических и экономических моделей, используя программный продукт «STATISTICA6».

В качестве задач следует выделить следующее:

- овладение методами выполнения оценок параметров больших множеств по данным выборочного наблюдения;
- приобретение навыков работы с большими массивами данных и навыков представления данных статистического наблюдения в удобном для восприятия, анализа и принятия решений виде;
- развитие аналитических навыков в ходе применения вариационного метода интерпретации полученных результатов;

Курсовая работа состоит из двух глав, введения, заключения и списка использованной информации.

В первой главе изучения программного продукта «STATISTICA6».

Во второй главе представлен анализ предприятие.

В заключении приведены основные итоги об анализе предприятия.

1.1 Статистика

Статистика – отрасль знаний, наука, в которой излагаются общие вопросы сбора, измерения и анализа массовых статистических (количественных или качественных) данных; изучение количественной стороны массовых общественных явлений в числовой форме.

Слово «статистика» происходит от латинского status – состояние дел. В науку термин «статистика» ввёл немецкий учёный Готфрид Ахенвалль в 1746 году, предложив заменить название курса «Государствоведение», преподававшегося в университетах Германии, на «Статистику», положив тем самым начало развитию статистики как науки и учебной дисциплины. Несмотря на это, статистический учёт вёлся намного раньше: проводились переписи населения в Древнем Китае, осуществлялось сравнение военного потенциала государств, вёлся учёт имущества граждан в Древнем Риме и тому подобное.

Статистика разрабатывает специальную методологию исследования и обработки материалов: массовые статистические наблюдения, метод группировок, средних величин, индексов, балансовый метод, метод графических изображений и другие методы анализа статистических данных. [1]

1.2 Программа «Statistica6»

STATISTICA – это интегрированная система анализа и управления данными. STATISTICA – это инструмент разработки пользовательских приложений в бизнесе, экономике, финансах, промышленности, медицине, страховании и других областях. STATISTICA легка в освоении и использовании.

Все аналитические инструменты, имеющиеся в системе, доступны пользователю и могут быть выбраны с помощью альтернативного пользовательского интерфейса. Пользователь может всесторонне автоматизировать свою работу, начиная с применения простых макросов для автоматизации рутинных

действий вплоть до углубленных проектов, включающих в том числе интеграцию системы с другими приложениями или Интернет. Технология автоматизации позволяет даже неопытному пользователю настроить систему на свой проект.

Процедуры системы STATISTICA имеют высокую скорость и точность вычислений.

Гибкая и мощная технология доступа к данным позволяет эффективно работать как с таблицами данных на локальном диске, так и с удаленными хранилищами данных.

Система обладает следующими общепризнанными достоинствами:

- содержит полный набор классических методов анализа данных: от основных методов статистики до продвинутых методов, что позволяет гибко организовать анализ;
- является средством построения приложений в конкретных областях;
- в комплект поставки входят специально подобранные примеры, позволяющие систематически осваивать методы анализа;
- отвечает всем стандартам Windows, что позволяет сделать анализ высокоинтерактивным;
- система может быть интегрирована в Интернет;
- поддерживает web-форматы: HTML, JPEG, PNG;
- легка в освоении, и как показывает опыт, пользователи из всех областей применения быстро осваивают систему;
- данные системы STATISTICA легко конвертировать в различные базы данных и электронные таблицы;
- поддерживает высококачественную графику, позволяющую эффективно визуализировать данные и проводить графический анализ;

— является открытой системой: содержит языки программирования, которые позволяют расширять систему, запускать ее из других Windows-приложений, например, из Excel.

STATISTICA состоит из набора модулей, в каждом из которых собраны тематически связанные группы процедур. При переключении модулей можно либо оставлять открытым только одно окно приложения STATISTICA, либо все вызванные ранее модули, поскольку каждый из них может выполняться в отдельном окне (как самостоятельное приложение Windows). При исполнении модулей STATISTICA как самостоятельных приложений в любой момент времени в любом модуле имеется прямой доступ к «общим» ресурсам (таблицам данных, языкам BASIC и SCL, графическим процедурам).[4]

1.3 Прикладная Статистика

Прикладная статистика – это наука о том, как обрабатывать данные произвольной природы. Математической основой прикладной статистики и статистических методов анализа является теория вероятностей и математическая статистика.

Описание вида данных и механизма их порождения – начало любого статистического исследования. Для описания данных применяют как детерминированные, так и вероятностные методы. С помощью детерминированных методов можно проанализировать только те данные, которые имеются в распоряжении исследователя. Например, с их помощью получены таблицы, рассчитанные органами официальной государственной статистики на основе представленных предприятиями и организациями статистических отчетов. Перенести полученные результаты на более широкую совокупность, использовать их для предсказания и управления можно лишь на основе вероятностно-статистического моделирования. Поэтому в математическую статистику часто включают лишь методы, опирающиеся на теорию вероятностей.

В простейшей ситуации статистические данные – это значения некоторого признака, свойственного изучаемым объектам. Значения могут быть количественными или представлять собой указание на категорию, к которой можно отнести объект. Во втором случае говорят о качественном признаке.

При измерении по нескольким количественным или качественным признакам в качестве статистических данных об объекте получаем вектор. Его можно рассматривать как новый вид данных. В таком случае выборка состоит из набора векторов. Есть часть координат – числа, а часть – качественные (категоризованные) данные, то говорим о векторе разнотипных данных.

Одним элементом выборки, то есть одним измерением, может быть и функция в целом. Например, описывающая динамику показателя, то есть его изменение во времени, – электрокардиограмма больного или амплитуда биений вала двигателя. Или временной ряд, описывающий динамику показателей определенной фирмы. Тогда выборка состоит из набора функций.

Элементами выборки могут быть и иные математические объекты. Например, бинарные отношения. Так, при опросах экспертов часто используют упорядочения (ранжировки) объектов экспертизы – образцов продукции, инвестиционных проектов, вариантов управленческих решений. В зависимости от регламента экспертного исследования элементами выборки могут быть различные виды бинарных отношений (упорядочения, разбиения, толерантности), множества, нечёткие множества и т. д.

Итак, математическая природа элементов выборки в различных задачах прикладной статистики может быть самой разной. Однако можно выделить два класса статистических данных – числовые и нечисловые. Соответственно прикладная статистика разбивается на две части - числовую статистику и нечисловую статистику.

Числовые статистические данные – это числа, вектора, функции. Их можно складывать, умножать на коэффициенты. Поэтому в числовой статисти-

стике большое значение имеют разнообразные суммы. Математический аппарат анализа сумм случайных элементов выборки – это (классические) законы больших чисел и центральные предельные теоремы.

Нечисловые статистические данные – это категоризованные данные, вектора разнотипных признаков, бинарные отношения, множества, нечёткие множества и др. Их нельзя складывать и умножать на коэффициенты. Поэтому не имеет смысла говорить о суммах нечисловых статистических данных. Они являются элементами нечисловых математических пространств (множеств). Математический аппарат анализа нечисловых статистических данных основан на использовании расстояний между элементами (а также мер близости, показателей различия) в таких пространствах. С помощью расстояний определяются эмпирические и теоретические средние, доказываются законы больших чисел, строятся непараметрические оценки плотности распределения вероятностей, решаются задачи диагностики и кластерного анализа, и т. д.

В прикладных исследованиях используют статистические данные различных видов. Это связано, в частности, со способами их получения. Например, если испытания некоторых технических устройств продолжаются до определённого момента времени, то получаем так называемые цензурированные данные, состоящие из набора чисел – продолжительности работы ряда устройств до отказа, и информации о том, что остальные устройства продолжали работать в момент окончания испытания. Цензурированные данные часто используются при оценке и контроле надёжности технических устройств.

[1]

2 Анализ Предприятия

2.1 Исходные Данные

D:\НАДЯ\Надя (2).xlsm : Надя											
1	2	3	4	5	6	7	8	9	10	11	12
Запчасть	Наименование	Фирма	Страна	Марка Авто	Операция	Дата	Изменение	Цена	Стоимость	Месяц	Год
1	0 суппорт тормоз	Stellox	германи	Renault	Товарный	26-Мар-14	-1	6 360	6360	март	четырна
2	0 суппорт тормоз	Stellox	германи	Renault	Заказ	7-Апр-14	1	4 050	4050	апрель	четырна
3	0 суппорт тормоз	Stellox	германи	Renault	Товарный	18-Фев-15	-1	6 890	6890	февраль	пятнадц.
4	0 суппорт тормоз	Stellox	германи	Renault	Заказ	31-Мар-15	1	3 127	3127	март	пятнадц.
5	0 суппорт тормоз	Stellox	германи	Renault	Товарный	11-Сен-15	-1	5 320	5320	сентябрь	пятнадц.
6	0 суппорт тормоз	Stellox	германи	Renault	Заказ	21-Окт-15	1	3 631	3631	октябрь	пятнадц.
7	0 суппорт тормоз	Stellox	германи	Renault	Товарный	26-Фев-16	-1	6 180	6180	февраль	шестнадц.
8	0 суппорт тормоз	Stellox	германи	Renault	Заказ	4-Мар-16	1	3 334	3334	март	шестнадц.
9	9 стартер 12V T4	Unipoint	германи	Электрика	Заказ	10-Июл-17	1	3 777	3777	июль	семнадц.
10	9 стартер 12V T4	Unipoint	германи	Электрика	Заказ	24-Июл-17	1	3 788	3788	июль	семнадц.
11	10 коленчатый вал	Ford	германи	Ford	Товарный	18-Апр-14	-2	0	0	апрель	четырна
12	10 коленчатый вал	Ford	германи	Ford	Заказ	18-Апр-14	2	110	219	апрель	четырна
13	10 коленчатый вал	Ford	германи	Ford	Заказ	8-Фев-17	1	2 000	2000	февраль	семнадц.
14	10 коленчатый вал	Ford	германи	Ford	Товарный	4-Окт-17	-1	4 010	4010	октябрь	семнадц.
15	11 решетка радиат	Tyq	китай	Ford	Товарный	22-Янв-14	-1	2 755	2755	январь	четырна

Рисунок 1 – Таблица данных

В таблице 1 представлены данные магазина автозапчастей с периода 31 января 2013 по 11 октября 2017.

Показатели:

- 1) Номер запчасти.
- 2) Наименование запчасти.
- 3) Фирма производителя.
- 4) Страна производителя.
- 5) Марка автомобиля.
- 6) Вид операции.
- 7) Дата проведения операции.
- 8) Количество(Изменение).
- 9)Цена за 1 шт. в рублях.
- 10)Стоимость.
- 11)Месяц.
- 12)Год.

2.2 Описательные Статистики

Рисунок 2 описательных статистик для итогов по магазину автозапчастей:

Переменная	Описательные статистики (Таблица)						
	N набл.	Среднее	Геометр. Среднее	Минимум	Максим.	Дисперсия	Ст.откл.
Изменение	190582	0,84		-6993,00	18775	8,521584E+03	92,3
Цена	190582	17017,55		0,00	15742371	1,132217E+10	106405,7
Стоимость	190582	25461,83		0,00	26523246	3,100322E+10	176077,3

Рисунок 2 – Описательные статистики для итогов по магазину автозапчастей

На рисунке 2 представлены следующие статистики:

- число наблюдений 190582;
- среднее: изменение :0,84; цена:17017,55; стоимость:25461,83;
- стандартное отклонение: изменение:92,3; цена:106405,7; стоимость:176077,3.
- дисперсия изменение:8,521584; цена:1,13221; стоимость:3,10032;
- минимум изменение: -6993; цена:0; стоимость:0;
- максимум изменение:18775; цена:15742371; стоимость26523246.

2.3 Корреляционная Матрица

Корреляции (Таблица_данных2_(Восстановлен))												
Отмеченные корреляции значимы на уровне $p < ,05000$												
N=190582 (Построчное удаление ПД)												
Переменная	Запчасть	Наименование	Фирма	Страна	Марка Автомобилья	Операция	Дата	Изменение	Цена	Стоимость	Месяц	Год
Запчасть	1,000000	0,107468	0,143061	0,127973	0,020813	0,043159	0,156679	0,019463	0,035336	0,016862	0,016416	0,151492
Наименование	0,107468	1,000000	0,238619	0,061927	0,635040	0,088611	0,032796	0,030080	0,063026	0,057438	0,019467	0,027638
Фирма	0,143061	0,238619	1,000000	0,508157	0,168370	0,046735	0,058095	-0,002506	0,040997	0,030880	0,007800	0,055913
Страна	0,127973	0,061927	0,508157	1,000000	0,039515	0,005730	0,049952	0,017108	-0,000869	-0,003350	-0,002141	0,050380
Марка Автомобилья	0,020813	0,635040	0,168370	0,039515	1,000000	0,046481	0,001784	0,002519	0,035843	0,036073	0,009372	-0,000534
Операция	0,043159	0,088611	0,046735	0,005730	0,046481	1,000000	0,036892	0,051154	0,070928	0,066696	0,029782	0,028792
Дата	0,156679	0,032796	0,058095	0,049952	0,001784	0,036892	1,000000	0,016014	0,014182	0,008862	0,099114	0,968678
Изменение	0,019463	0,030080	-0,002506	0,017108	0,002519	0,051154	0,016014	1,000000	-0,003916	-0,104278	-0,008039	0,017884
Цена	0,035336	0,063026	0,040997	-0,000869	0,035843	0,070928	0,014182	-0,003916	1,000000	0,690652	-0,002571	0,014637
Стоимость	0,016862	0,057438	0,030880	-0,003350	0,036073	0,066696	0,008862	-0,104278	0,690652	1,000000	0,000146	0,008656
Месяц	0,016416	0,019467	0,007800	-0,002141	0,009372	0,029782	0,099114	-0,008039	-0,002571	0,000146	1,000000	-0,150137
Год	0,151492	0,027638	0,055913	0,050380	-0,000534	0,028792	0,968678	0,017884	0,014637	0,008656	-0,150137	1,000000

Рисунок 3 – Корреляционная матрица

Переменная	Корреляции (Таблица)		
	Изменение	Цена	Стоимость
Изменение	1,000000	-0,003916	-0,104278
Цена	-0,003916	1,000000	0,690652
Стоимость	-0,104278	0,690652	1,000000

Рисунок 4 – Корреляционная матрица для определённых переменных

Между переменными (случайными величинами) может существовать функциональная связь, проявляющаяся в том, что одна из них определяется как функция от другой. Но между переменными может существовать и связь другого рода, проявляющаяся в том, что одна из них реагирует на изменение другой изменением своего закона распределения. В качестве меры зависимости между переменными используется коэффициент корреляции, который изменяется в пределах от -1 до $+1$. Если коэффициент корреляции отрицательный, это означает, что с увеличением значений одной переменной значения другой убывают. Если переменные независимы, то коэффициент корреляции равен 0 . Но если коэффициент корреляции не равен 0 (переменные называются некоррелированными), то это значит, что между переменными существует зависимость. Чем ближе значение γ к 1 , тем зависимость сильнее. [2]

Принято считать, что при $|r| < 0,25$ - корреляция слабая, $0,25 < |r| < 0,75$ - умеренная, при $|r| > 0,75$ – сильная.

Между Изменением и Стоимости, Изменении и Цена корреляция слабая, а между Стоимости и Цена корреляция умеренная.

Внутригрупповые корреляции:

Внутригрупповые корреляции (Надя)			
Группа: Операция: Товарный чек			
Отмеченные корреляции значимы на уровне p			
Переменные	Изменение	Цена	Стоимость
Изменение	1,000000	0,095552	-0,067964
Цена	0,095552	1,000000	0,911056
Стоимость	-0,067964	0,911056	1,000000

Рисунок 5 – Внутригрупповые корреляции по операции: Товарный чек

Внутригрупповые корреляции (Надя)			
Группа: Операция: Счет			
Отмеченные корреляции значимы на уровне p			
Переменные	Изменение	Цена	Стоимость
Изменение	1,000000	0,012049	-0,487223
Цена	0,012049	1,000000	0,643244
Стоимость	-0,487223	0,643244	1,000000

Рисунок 6 – Внутригрупповые корреляции по операции: Счёт

Внутригрупповые корреляции (Надя)			
Группа: Операция: Заказ			
Отмеченные корреляции значимы на уровне p			
Переменные	Изменение	Цена	Стоимость
Изменение	1,000000	-0,016407	0,369712
Цена	-0,016407	1,000000	0,594987
Стоимость	0,369712	0,594987	1,000000

Рисунок 7 – Внутригрупповые корреляции по операции: Заказ

Из Рисунков 5 – 7 видно, что в разных группах зависимости между анализируемыми величинами проявляются по-разному.

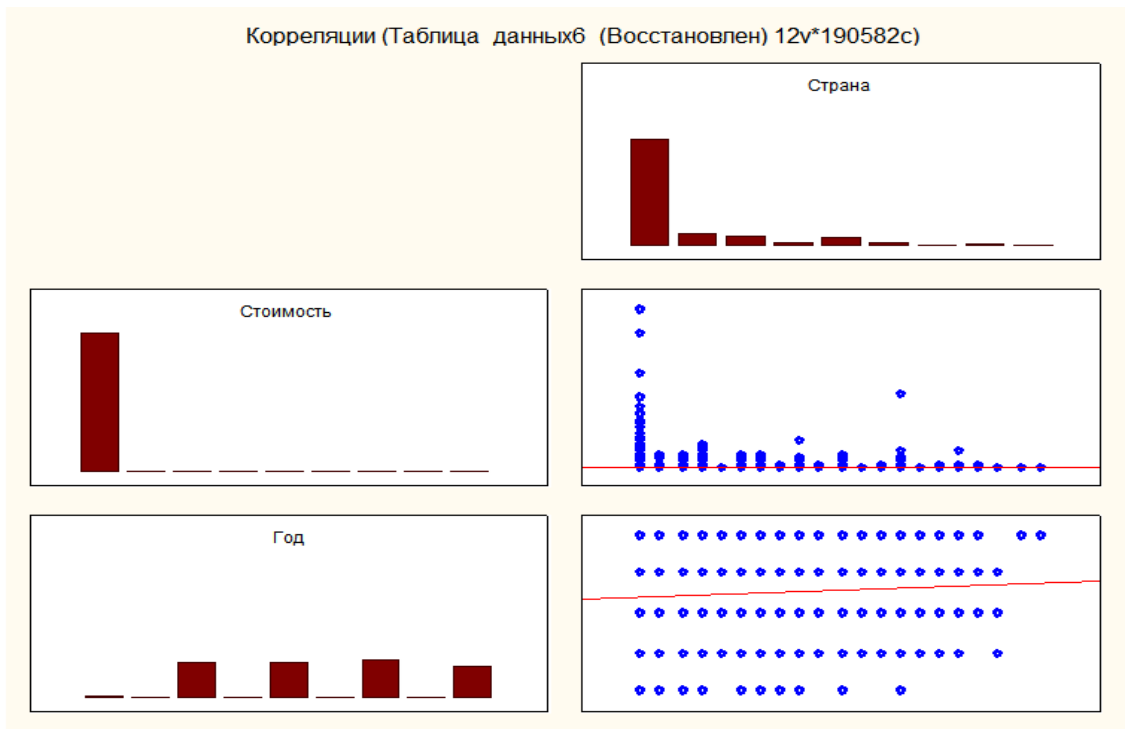


Рисунок 8 – Корреляции между Годом, Стоимостью, Страной

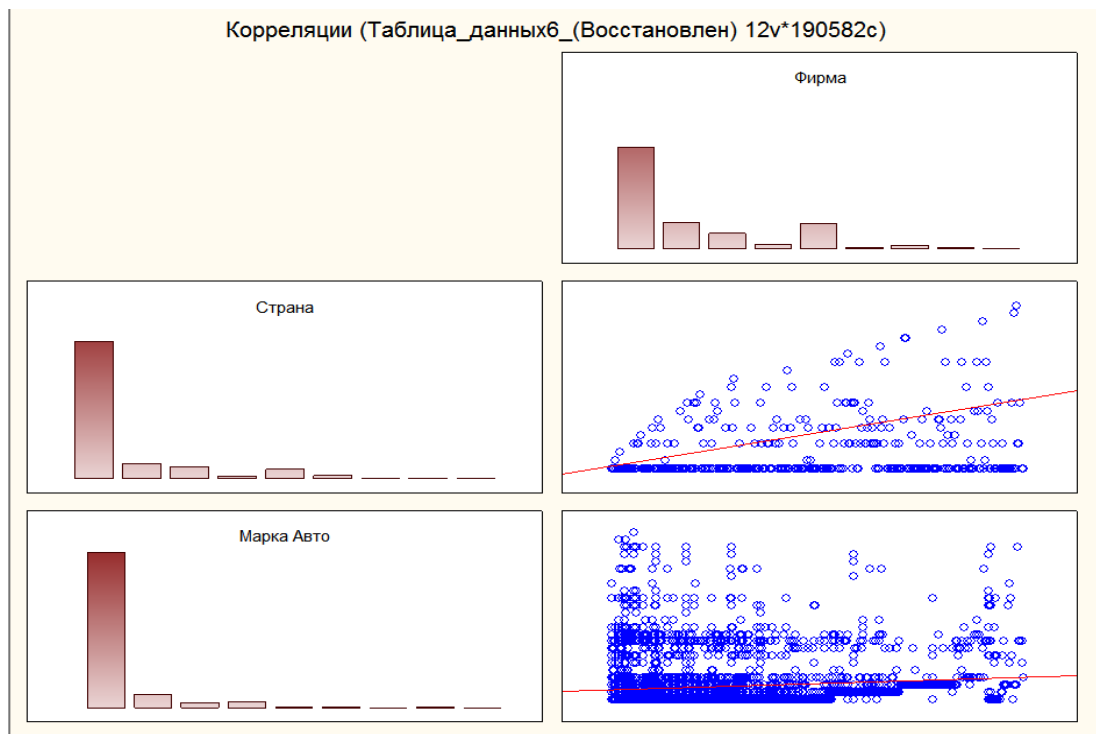


Рисунок 9 – Корреляции между Firmой, Страной, Маркой Авто

На рисунках 8-9 мы видим графики функций рассеяния и гистограммы.

На первом по зависимой переменной «Страна» и независимыми «Стоимость», «Месяц». На втором зависимой переменной «Страна» и независимыми «Стоимость», «Год».

2.4 Множественная Регрессия

Основная цель множественной регрессии – построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на исследуемый экономический показатель.

Итоговая таблица средних (Надя) N=190582 (Нет пропусков в завис. перем.)									
Операция	Изменение Среднее	Изменение N	Изменение Ст.откл.	Цена Среднее	Цена N	Цена Ст.откл.	Стоимость Среднее	Стоимость N	Стоимость Ст.откл.
Товарный чек	-1,85497	122615	2,4337	1568,1	122615	3964,7	1986,9	122615	4573,7
Счет	-3,11638	23964	60,3870	124750,5	23964	276879,3	187317,7	23964	465187,6
Заказ	10,49563	44003	186,5052	1396,2	44003	3897,4	2728,6	44003	7346,2
Всего	0,83802	190582	92,3124	17017,5	190582	106405,7	25461,8	190582	176077,3

Рисунок 10 – Итоговая таблица средних

Итоги регрессии для зависимой переменной: Стоимость (Надя) R= ,69808115 R2= ,48731729 Скоррект. R2= ,48731191 F(2, 190579)=90575, p<0,0000 Станд. ошибка оценки: 1261E2						
N=190582	БЕТА	Ст.Ош. БЕТА	B	Ст.Ош. B	z	p-знач.
Св.член			6186,511	292,4781	21,1520	0,00
Изменение	-0,101576	0,001640	-193,746	3,1285	-61,9297	0,00
Цена	0,690254	0,001640	1,142	0,0027	420,8417	0,00

Рисунок 11 – Итоги регрессии для ЗП: Стоимость

Итоговые статистики; ЗП:Стоимость (Надя)	
Статистика	Значение
Множест. R	0,698081152
Множест. R2	0,487317295
Скоррект. R2	0,487311915
F(2, 190579)	90574,9714
p	0
Стд. Ош. Оценки	126075,293

Рисунок 12 – Итоговая статистика ЗП: Стоимость

Выдвинем гипотезу о равенстве средних: $p < 0,05$, поэтому гипотезу о равенстве средних отвергается.

Максимальное значение коэффициента R^2 равно единице. Это происходит в том случае, когда линия регрессии точно соответствует всем наблюдениям, так что $\hat{y}_i = y_i$ для всех i и все остатки равны нулю. Если в выборке отсутствует видимая связь между y и x , то коэффициент R^2 будет близок к нулю. [2]

$$R^2 = 0,48731191$$

Рисунок 11 содержит стандартизованные (Beta) и нестандартизованные (jB) регрессионные коэффициенты (веса), их стандартные ошибки и уровни значимости. Коэффициенты Beta оцениваются по стандартизованным данным, имеющим выборочное среднее, равное 0 и стандартное отклонение, равное 1. Поэтому величины Beta позволяет сравнить вклады каждого предиктора в предсказание отклика. [2]

Уравнение регрессии имеет вид:

$$\hat{y} = 6186,511 - 193,746t_1 + 1,142t_2$$

Полученное уравнение следует интерпретировать следующим образом. При каждом увеличении «Цена» на 1 Р. «Стоимость» измениться на 1,142Р при прочих равных. При каждом увеличении «Изменения» на 1 ед. «Стоимость» измениться на -193,746 ед. при прочих равных. [3]

2.5 Дисперсионный Анализ

Дисперсионный анализ является наиболее общим методом сравнения средних. В дисперсионном анализе можно исследовать зависимость количественного признака (зависимой переменной) от одного или нескольких качественных признаков (факторов). Цель дискриминантного анализа состоит в том, чтобы на основе измерения различных характеристик (признаков, параметров) объекта классифицировать его, т.е. отнести к одной из нескольких групп (классов) некоторым оптимальным способом. Под оптимальным способом понимается либо минимум математического ожидания потерь, либо минимум вероятности ложной классификации. [2]

Переменная	Дисперсионный анализ (Надя)							
	Отмечены эффекты, значимые на уров. $p < ,05000$							
	Сум.кв. эффект	Ст. св. эффект	Ср.кв. эффект	Сум.кв. ошибки	Ст. св. ошибок	Ср.кв. ошибки	F	p
Изменения	5,368090E+	2	2,684045E+	1,618684E+	19057	8,493506E+	316,0	0,0
Цена	3,181397E+	2	1,590699E+	1,839651E+	19057	9,652959E+	16478,8	0,0
Стоимость	7,181031E+	2	3,590515E+	5,190521E+	19057	2,723553E+	13183,2	0,0

Рисунок 13 – Дисперсионный анализ

Из рисунка 13 следует, что переменные статистически значимы, так как уровень значимости (p) меньше чем 0,05.

2.6 Кластерный Анализ

Главное назначение кластерного анализа (от англ. cluster – гроздь, скопление) – разбиение множества исследуемых объектов и признаков на однородные в некотором смысле группы, или кластеры. Методы кластерного анализа можно применять даже тогда, когда речь идет о простой группировке, в которой все сводится к образованию групп по количественному сходству. Отличием кластерного анализа от других методов классификации является отсутствие обучающей выборки (классификация без обучения). Большое достоинство кластерного анализа в том, что он дает возможность производить разбиение объектов не по одному параметру, а по ряду признаков. Кроме того,

кластерный анализ в отличие от большинства математико-статистических методов не накладывает никаких ограничений на вид рассматриваемых объектов и позволяет исследовать множество исходных данных практически произвольной природы. [2]

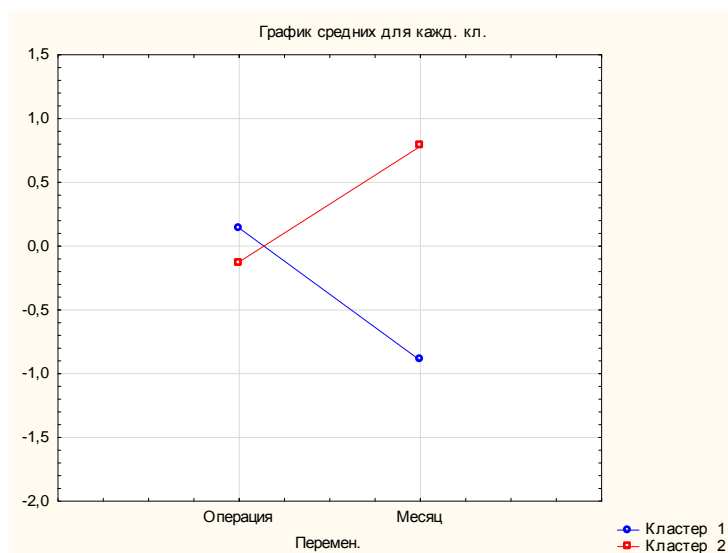


Рисунок 14 – График остатков для каждого кластера: Операции и Месяц

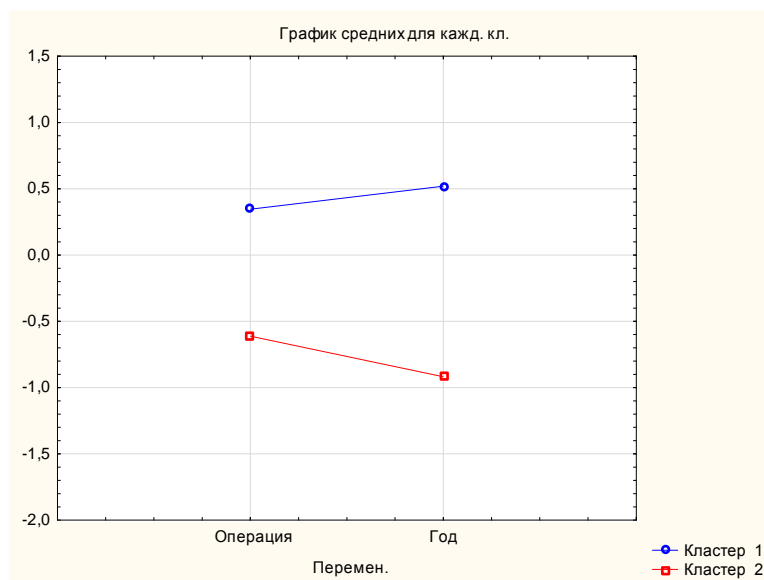


Рисунок 15 – График остатков для каждого кластера: Операции и Год

Рисунки 14 – 15 это графики остатков. На них отображены средние значения каждой переменной для каждого кластера.

перемен.	Дисперсионный анализ (кур23)					
	Между SS	сс	Внутри SS	сс	F	значим. р
Операция	3341,5	1	187239,5	190580	3401,1	0,00
Месяц	133832,9	1	56748,1	190580	449458,2	0,00

Рисунок 16 – Дисперсионный анализ между Операции и Месяцем

перемен.	Дисперсионный анализ (Таблица_данных6_(Восс)					
	Между SS	сс	Внутри SS	сс	F	значим. р
Операция	40316,11	1	150264,9	190580	51132,7	0
Год	90962,73	1	99618,3	190580	174021,1	0

Рисунок 17 – Дисперсионный анализ между Операцией и Годом

На рисунках 16 – 17 приведены значения межгрупповых и внутригрупповых дисперсий признаков. Чем меньше значение внутригрупповой и больше межгрупповой дисперсии, тем лучше признак характеризует принадлежность объектов к кластеру. Так как $p < 0.05$ для всех признаков, то вклады всех признаков в разделение объектов на группы существенные.

перемен.	Средн.класт. (кур23)	
	Кластер Но. 1	Кластер Но. 2
Операция	0,141756	-0,123686
Месяц	-0,897121	0,782762

Рисунок 18 – Среднее кластеров

Кластер Номер	Евклидовы расст. между кластерами (кур23)	
	Расстояния под диагональю	
	Квадраты расстояний над диагональю	
	Но. 1	Но. 2
Но. 1	0,000000	1,446234
Но. 2	1,202595	0,000000

Рисунок 19 – Евклидовы расстояние между кластерами

перемен.	Средн.класт. (Таблица_данных6_(Восстановлен)	
	Кластер Но. 1	Кластер Но. 2
Операция	0,345874	-0,611615
Год	0,519530	-0,918694

Рисунок 20 – Среднее кластеров

Классификация	Евклидовы расст. между кластерами (Таблица_дан)	
	Расстояния под диагональю	
Классификация	Квадраты расстояний над диагональю	
Номер	Но. 1	Но. 2
Но. 1	0,000000	1,492638
Но. 2	1,221735	0,000000

Рисунок 21 – Евклидовы расстояние между кластерами

На рисунках 18,20 приведены средние для каждого кластера, на рисунках 19,21 приведены евклидовы между кластерами.

Евклидово расстояние – наиболее общий тип расстояния. Оно попросту является геометрическим расстоянием в многомерном пространстве.

Квадрат евклидова расстояния – иногда может возникнуть желание возвести в квадрат стандартное евклидово расстояние, чтобы придать большие веса более отдаленным друг от друга объектам.

ЗАКЛЮЧЕНИЕ

В результате проделанной работы была достигнута поставленная цель, а именно был произведен анализ предприятия при помощи математических и экономических моделей, используя программу «STATISTICA6».

Была освоена и изучена программа «STATISTICA6», а именно модули линейной и нелинейной регрессии, дисперсионного анализа, непараметрической статистике.

Таким образом, используя разные способы анализа мы можем сказать, что предприятие рентабельно, развивающее и прибыльное.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Халафян А.А. «STATISTICA 6» Статистический анализ данных. 3-е изд. Учебник — М: ООО «Бином-Пресс», 2007г., 508 с.
2. Доугерти К. Введение в эконометрику: Пер. с англ. — М: ИНФРА-М, 1999 г., 416 с.
3. <http://www.hr-portal.ru/statistica/g11/g11.php#2>
4. <https://ru.wikipedia.org/wiki/Статистика>