

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Курганский государственный университет»

Кафедра «Программное обеспечение
автоматизированных систем»

Анализ данных. Часть 1. Классификация и регрессия. Статистические методы.
Методические указания к выполнению лабораторных работ
для бакалавров направлений
09.03.03 «Прикладная информатика»,
09.03.04 «Программная инженерия»

Кафедра: «Программное обеспечение автоматизированных систем».

Дисциплины: «Интеллектуальный анализ данных» (09.03.03), «Методы и алгоритмы анализа данных» (09.03.04).

Составитель: старший преподаватель Ю. В. Адаменко.

Печатается в соответствии с планом издания, утвержденным методическим советом университета «13» декабря 2024 г.

Утверждены на заседании кафедры «5» декабря 2025 г.

1 Применение классификации и регрессии

Задачи классификации и регрессии возникают практически во всех областях человеческой деятельности. Приведем несколько примеров.

В банковском деле классификация применяется для решения задач скоринга, определения кредитного рейтинга клиентов с целью минимизации рисков при выдаче кредитов, выявления мошенничества с кредитными картами.

В розничной торговле классификация может использоваться для выделения групп покупателей с определенными предпочтениями, что позволит более полно удовлетворять спрос и гибко реагировать на его изменения, прогнозировать состояние рынков.

В военной технике с помощью классификации можно распознавать тип цели, обнаруженной радиолокатором (пассажирский лайнер, бомбардировщик, истребитель и т. д.) по параметрам отраженного от цели сигнала.

В биомедицине с помощью классификации и регрессии можно диагностировать заболевания на основе наблюдаемых симптомов (температура, давление, состав крови и т. д.), оценивать ожидаемые результаты лечения.

В правоохранительной деятельности классификация используется для выявления мошеннических действий, при расследовании и предотвращении преступлений на основе анализа ранее совершенных, организации систем безопасности и т. д.

Как от классификации перейти к задаче регрессии? Например, при определении кредитного рейтинга все клиенты банка могут быть распределены по трем классам: высокому, среднему и низкому. Если эту же задачу поставить как регрессионную, то в качестве выходной переменной можно выбрать оценку вероятности возврата кредита. Как и вероятность любого события, она будет изменяться от 0 до 1. Высокая вероятность возврата (0,81) соответствует высокому кредитному рейтингу, низкая вероятность (0,1-0,3) – низкому, промежуточные значения (0,3-0,7) – среднему.

При обеих постановках задачи сохраняется главная цель минимизация рисков при кредитовании.

2 Постановка задачи классификации

В задаче классификации и регрессии требуется определить значение зависимой переменной объекта на основании значений других переменных, характеризующих данный объект.

Формально задачу классификации и регрессии можно описать следующим образом.

Имеется множество объектов: $I = \{i_1, i_2, \dots, i_j, \dots, i_n\}$,

где i_j – исследуемый объект.

Каждый объект характеризуется набором переменных:

$$I_j = \{x_1, x_2, \dots, x_h, \dots, x_m, y\},$$

где x_h – независимые переменные, значения которых известны и на основании которых определяется значение зависимой переменной y .

Каждая переменная x_h может принимать значения из некоторого множества:

$$C_h = \{c_{h1}, c_{h2}, \dots\}$$

Если значениями переменной являются элементы конечного множества, то говорят, что она имеет категориальный тип.

Если множество значений $C = \{c_1, c_2, \dots, c_r, \dots, c_k\}$ переменной y конечно, то задача называется задачей классификации. Если переменная y принимает значение на множестве действительных чисел R , то задача называется задачей регрессии.

Цель процесса классификации состоит в том, чтобы построить модель, которая использует прогнозирующие атрибуты в качестве входных параметров и получает значение зависимого атрибута.

Процесс классификации заключается в разбиении множества объектов на классы по определенному критерию.

Классификатором называется некая сущность, определяющая, какому из predetermined классов принадлежит объект по вектору признаков.

Для проведения классификации с помощью математических методов необходимо иметь формальное описание объекта, которым можно оперировать, используя математический аппарат классификации. Таким описанием в нашем случае выступает база данных. Каждый объект (запись базы данных) несет информацию о некотором свойстве объекта.

Набор исходных данных (или выборку данных) разбивают на два множества: обучающее и тестовое.

Обучающее множество (training set) – множество, которое включает данные, используемые для обучения (конструирования) модели. Такое множество содержит входные и выходные (целевые) значения примеров. Выходные значения предназначены для обучения модели.

Тестовое (test set) множество также содержит входные и выходные значения примеров. Здесь выходные значения используются для проверки работоспособности модели.

Процесс классификации состоит из двух этапов:

1 Конструирование модели (описание множества predetermined классов):

а) каждый пример набора данных относится к одному predetermined классу;

б) на этом этапе используется обучающее множество, на нем происходит конструирование модели;

в) полученная модель представлена классификационными правилами, деревом решений или математической формулой.

2 Использование модели (классификация новых или неизвестных значений):

а) оценка правильности (точности) модели:

- известные значения из тестового примера сравниваются с результатами использования полученной модели;
- уровень точности – процент правильно классифицированных примеров в тестовом множестве;
- тестовое множество, т. е. множество, на котором тестируется построенная модель, не должно зависеть от обучающего множества.

б) если точность модели допустима, возможно использование модели для классификации новых примеров, класс которых неизвестен.

В контексте бизнес-аналитики все методы можно разделить на статистические и методы машинного обучения.

Статистические методы основаны на математической статистике. К ним, в частности, относятся:

- линейная регрессия;
- логистическая регрессия;
- байесовская классификация и др.

3 Практические задания

1 OLAP-анализ. Аналитическая отчетность аптечной сети

1.1 Узел Калькулятор

Калькулятор предназначен для добавления в набор данных новых полей, которые рассчитываются по определенным правилам на основе столбцов данных и встроенных функций.

Обработчик Калькулятор находится в группе узлов «Трансформация данных» мастера обработки. Вся настройка осуществляется в окне мастера «Конструктор выражения» (рисунок 1).

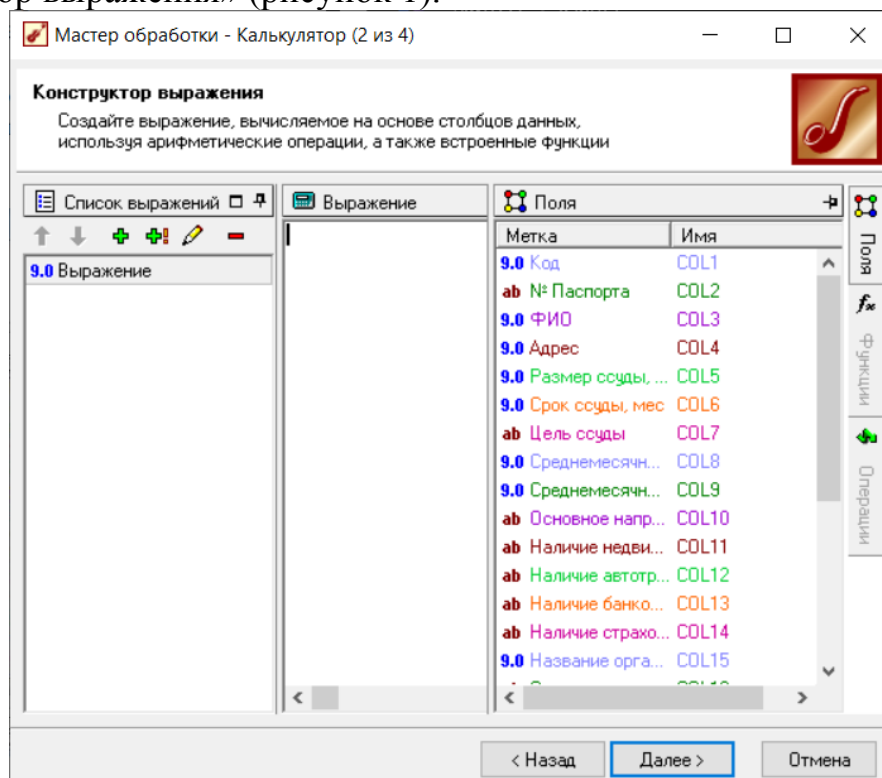


Рисунок 1 – Конструктор выражения

Создайте новый проект. Импортируйте в него текстовый файл CreditSample.txt.

Области узла Калькулятор:

- область списка вычисляемых выражений. Каждое вычисляемое выражение будет новым столбцом в результирующем наборе данных (окно «Список выражений»);
- формула, по которой будет рассчитываться выражение (окно «Выражения»);
- список всех существующих столбцов текущего набора данных, состоящих из имен и меток (вкладка «Поля»);
- вкладка «Функции» открывает вкладку со списком встроенных функций;
- вкладка «Операции» открывает вкладку со списком доступных арифметических, логических и других операций.

Область списка вычисляемых выражений изначально содержит одно пустое выражение. Двойным щелчком мыши на имени выражения в списке вызывается «Диалог редактирования» параметров выражения (рисунок 2).

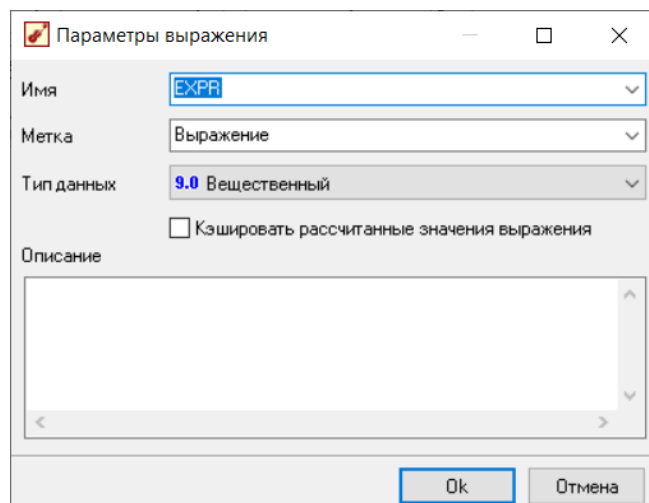


Рисунок 2 – Диалог редактирования параметров выражения

Имя – строка, которая будет служить идентификатором столбца в процедурах обработки.

Метка – метка нового столбца.

Тип данных – тип данных вычисляемого выражения.

Описание – произвольная информация, описывающая вычисляемое выражение.

Выражение можно ввести вручную с клавиатуры, однако удобнее выбирать функции, переменные и знаки операций с помощью мыши. Для добавления в формулу функций следует справа выбрать вкладку «Функции». Все функции в ней сгруппированы по видам (рисунок 3).

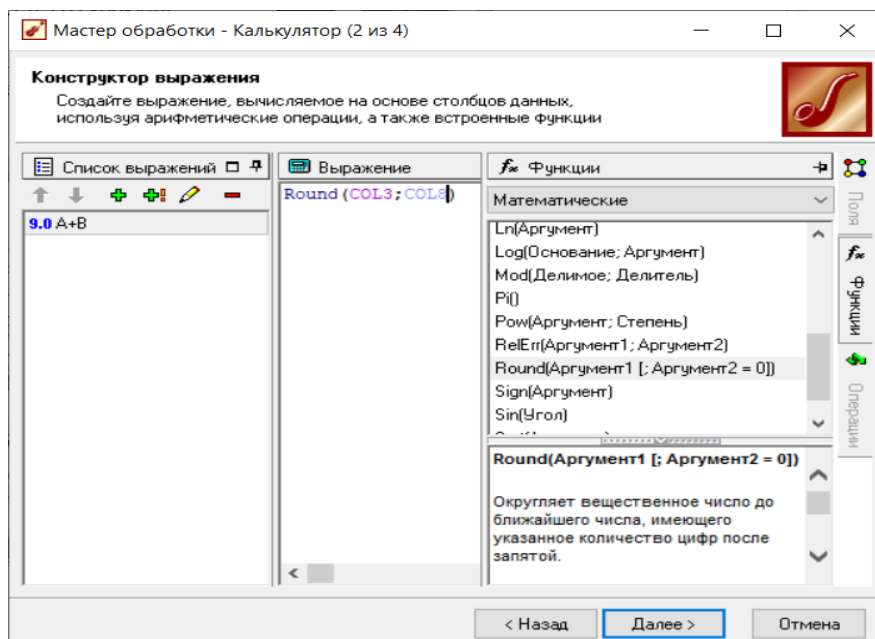


Рисунок 3 – Ввод функции в выражение

Чтобы ввести функцию в выражение, достаточно дважды щелкнуть по ее имени в списке, либо, удерживая, перетащить ее курсором мыши в нужную область формулы. Имя функции в выражении появляется вместе со скобками, куда следует ввести аргумент или аргументы.

Имена полей удобно вводить с помощью двойного щелчка в списке полей. Если в аргументе несколько полей, то их имена разделяются точкой с запятой.

Даты в формате ДД.ММ.ГГ. обязательно указываемые в кавычках. Такой способ ввода даты, хотя и допускается, но может оказаться не переносимым между разными компьютерами. По этой причине лучше использовать функцию STRTODATE().

В окне «Ввода выражения» можно вывести подсказку – комбинация клавиш Ctrl + пробел.

При создании формул при разработке сценариев очень часто используются функции IF и IFF (таблица 1).

Узел Калькулятор имеет следующее правило работы: при возникновении любой ошибки в расчете значения записи в рассчитываемое поле заносится значение null (пустое значение) и сообщение об ошибке не выдается.

Таблица 1 – Описание функций IF и IFF

Функция	Описание
IF (Условие; Значение1; Значение2)	Возвращает Значение1, если Условие истинно или Значение2, если Условие ложно. Результат функции имеет строковый тип.
IFF (Условие; Значение1; Значение2)	Возвращает Значение1, если Условие истинно или Значение2, если ложно. Результат функции может иметь любой тип.

Задание

- 1) создайте новое поле «Дата обработки», значения в котором равны текущей дате.

– вызовите «узел Калькулятор» и заполните соответствующие поля в окне «Параметры выражения» (рисунок 4).

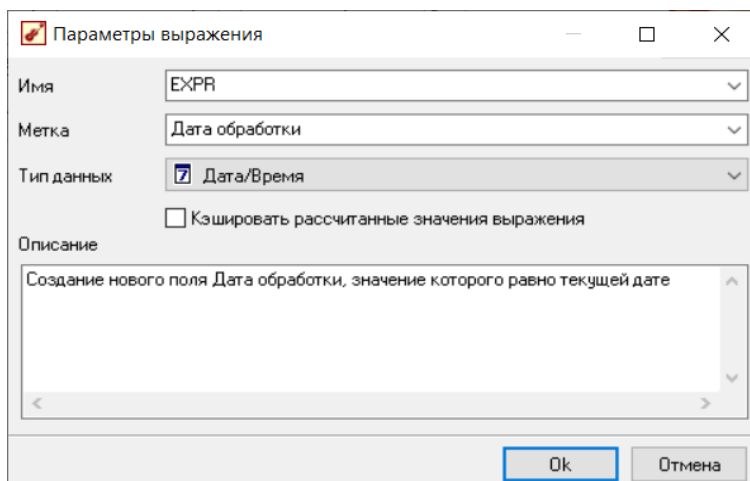


Рисунок 4 – Параметры выражения

– в окне «Выражения», используя вкладку «Функции», поместите функцию Today() – она возвращает текущую дату (рисунок 5).

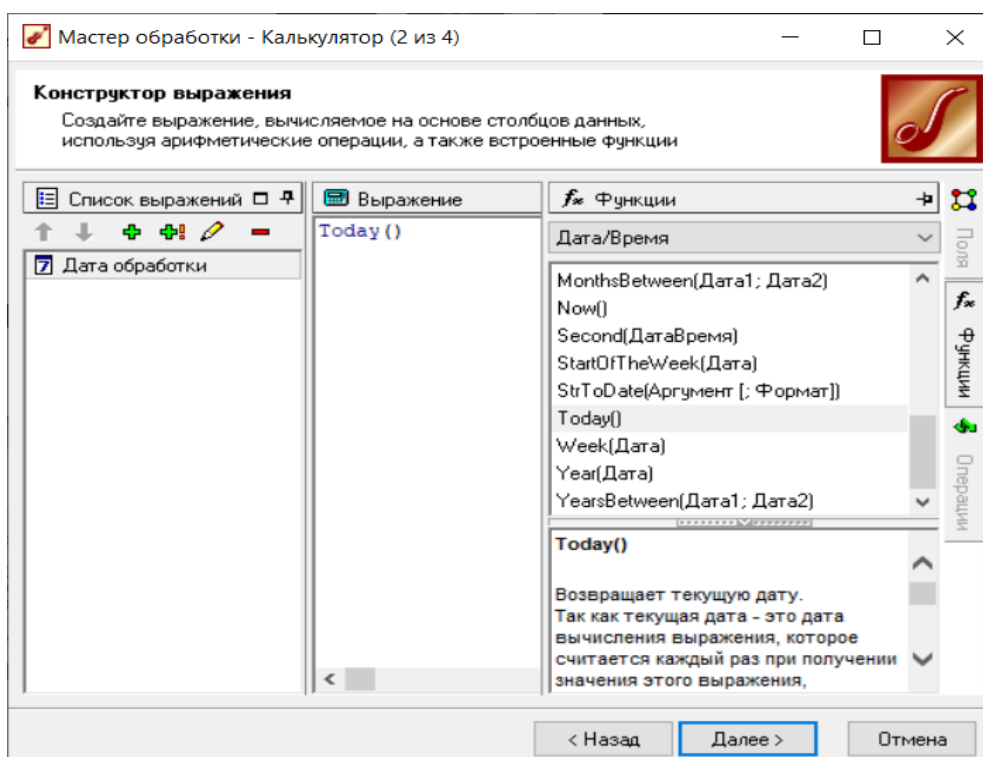


Рисунок 5 – Применение функции Today()

– поместите в таблице столбец «Дата обработки» на первое место (рисунок 6).

Дата обработки	Код	№ Паспорта	ФИО	Адрес	Размер ссуды, руб
02.08.2023	1	0936-866096	1	1	1
02.08.2023	2	8355-512943	1	1	1
02.08.2023	3	8017-098471	1	1	1
02.08.2023	4	2762-945535	1	1	1
02.08.2023	5	0459-997701	1	1	1
02.08.2023	6	6291-817248	1	1	1
02.08.2023	7	0094-883508	1	1	1
02.08.2023	8	6385-082612	1	1	1
02.08.2023	9	9290-732392	1	1	1
02.08.2023	10	7022-736193	1	1	1
02.08.2023	11	3127-709332	1	1	1
02.08.2023	12	4179-171975	1	1	1
02.08.2023	13	6099-152158	1	1	1
02.08.2023	14	3805-792726	1	1	1

Рисунок 6 – Помещение столбца «Дата обработки» на первое место в таблице

2) создайте новое поле «Размер ссуды у.е.», который рассчитывается делением на 30 поля «Размер ссуды, руб». Все значения в новом поле должны быть округлены до второго знака (рисунок 7).

Адрес	Размер ссуды, у.е.	Размер ссуды, руб	Срок ссуды, мес
1	333,33	10000	12 Покупка т
1	466,67	14000	12 Оплата за
1	833,33	25000	18 Оплата за
1	2133,33	64000	48 Покупка и
1	116,67	3500	6 Покупка т
1	283,33	8500	6 Иное
1	83,33	2500	6 Оплата ус
1	900	27000	18 Иное
1	550	16500	12 Покупка и
1	216,67	6500	6 Покупка т
1	750	22500	12 Покупка т

Рисунок 7 – Создание нового поля Размер ссуды

2) создайте новое поле «Флаг», значение в котором истинно (True), если выполняется условие: Среднемесячный доход >2000 и Наличие недвижимости = True (рисунок 8).

Размер ссуды, руб	Срок ссуды, мес	Цель ссуды	Среднемесячный доход, руб	Наличие недвижимости	Флаг	Среднемесячный расход, руб
10000	12	Покупка товара	4500	Да	✓	3000
14000	12	Оплата за образование	8500	Да	✓	3000
25000	18	Оплата за образование	7000	Да	✓	4000
64000	48	Покупка и ремонт недвижимости	7000	Да	✓	5000
3500	6	Покупка товара	3000	Нет	☐	1500
8500	6	Иное	7500	Да	✓	4000
2500	6	Оплата услуг (нед., юрид. и т.п.)	2500	Нет	☐	2000
27000	18	Иное	8000	Нет	☐	4500
16500	12	Покупка и ремонт недвижимости	7000	Да	✓	4000
6500	6	Покупка товара	5500	Да	✓	3000
22500	12	Покупка товара	15500	Да	✓	8500
25000	18	Покупка и ремонт недвижимости	7000	Да	✓	2500
39000	12	Покупка товара	16500	Да	✓	8500
16000	12	Покупка товара	7000	Да	✓	3000
10000	6	Оплата услуг (нед., юрид. и т.п.)	12500	Да	✓	7500
9500	6	Покупка и ремонт недвижимости	8000	Да	✓	4000
49500	24	Оплата услуг (нед., юрид. и т.п.)	11000	Нет	☐	4000

Рисунок 8 – Создание нового поля Флаг

4) создайте еще один столбец (рисунок 9), значение в котором равно 1, если выполняется условие: Флаг = TRUE и Давать кредит = FALSE.

Давать кредит	ФЛАГ	ФЛАГ=True и Давать кредит = False
ИСТИНА	<input checked="" type="checkbox"/>	0
ИСТИНА	<input checked="" type="checkbox"/>	0
ИСТИНА	<input checked="" type="checkbox"/>	0
ИСТИНА	<input checked="" type="checkbox"/>	0
ЛОЖЬ	<input type="checkbox"/>	0
ИСТИНА	<input checked="" type="checkbox"/>	0
ЛОЖЬ	<input type="checkbox"/>	0
ЛОЖЬ	<input type="checkbox"/>	0
ИСТИНА	<input checked="" type="checkbox"/>	0
ИСТИНА	<input checked="" type="checkbox"/>	0
ИСТИНА	<input checked="" type="checkbox"/>	0
ИСТИНА	<input checked="" type="checkbox"/>	0
ЛОЖЬ	<input checked="" type="checkbox"/>	1
ИСТИНА	<input checked="" type="checkbox"/>	0
ИСТИНА	<input checked="" type="checkbox"/>	0
ЛОЖЬ	<input checked="" type="checkbox"/>	1
ЛОЖЬ	<input type="checkbox"/>	0
ЛОЖЬ	<input checked="" type="checkbox"/>	1
ИСТИНА	<input checked="" type="checkbox"/>	0
ЛОЖЬ	<input type="checkbox"/>	0
ИСТИНА	<input type="checkbox"/>	0

Рисунок 9 – Создание еще одного столбца, значение в котором равно 1 при заданном условии

5) создайте новое поле «RATE» (рисунок 10), в котором хранится значение в поле «Срок ссуды», возведенное в степень 0,6.

Размер ссуды, руб	Срок ссуды, мес	Rate	Цель ссуды
10000	12	4128606984584	Покупка товара
14000	12	4128606984584	Оплата за образование
25000	18	6452506776941	Оплата за образование
64000	48	2033960050063	Покупка и ремонт недвижимости
3500	6	3015605158352	Покупка товара
8500	6	3015605158352	Иное
2500	6	3015605158352	Оплата услуг (мед., юрид. и т.п.)
27000	18	6452506776941	Иное
16500	12	4128606984584	Покупка и ремонт недвижимости
6500	6	3015605158352	Покупка товара
22500	12	4128606984584	Покупка товара
25000	18	6452506776941	Покупка и ремонт недвижимости
39000	12	4128606984584	Покупка товара
16000	12	4128606984584	Покупка товара
10000	6	3015605158352	Оплата услуг (мед., юрид. и т.п.)
9500	6	3015605158352	Покупка и ремонт недвижимости
49500	24	7317308726772	Оплата услуг (мед., юрид. и т.п.)
12500	6	3015605158352	Покупка и ремонт недвижимости
22000	12	4128606984584	Покупка товара

Рисунок 10 – Создание нового поля RATE

б) создайте новое поле «Сегмент» (рисунок 11), которое делит всех заемщиков на сегменты по следующим правилам:

а) ЕСЛИ Возраст ≥ 50 и Среднемесячный доход < 6000 , ТО Сегмент = Сегмент 1;

б) ЕСЛИ Возраст < 30 , ТО Сегмент = Сегмент 2;

в) Сегмент = Сегмент 3 во всех остальных случаях, не удовлетворяющим пункту а и б.

The image shows two screenshots of an Excel spreadsheet. The left screenshot shows a table with columns 'Количество лет', 'Среднемесячный доход, руб', and 'Сегмент'. The first row is highlighted in blue and contains the values 62, 4500, and 'Сегмент 1'. The right screenshot shows the same table after the 'Сегмент' field has been created, with the first row now containing 41, 4500, and 'Сегмент 3'.

Количество лет	Среднемесячный доход, руб	Сегмент
62	4500	Сегмент 1
53	5500	Сегмент 1
69	4000	Сегмент 1
26	8500	Сегмент 2
21	2500	Сегмент 2
24	8000	Сегмент 2

Количество лет	Среднемесячный доход, руб	Сегмент
41	4500	Сегмент 3
26	8500	Сегмент 2
32	7000	Сегмент 3
50	7000	Сегмент 3
30	3000	Сегмент 3
63	7500	Сегмент 3
21	2500	Сегмент 2
24	8000	Сегмент 2
50	7000	Сегмент 3

Рисунок 11 – Создание поля Сегмент

Сохраните результаты в файле L1.ded.

2 Проведение ABC-анализа

2.1 Суть метода

ABC-анализ – метод, позволяющий классифицировать ресурсы компании по степени их важности и прибыльности.

В его основе лежит Принцип Парето: 20 % всех товаров дают 80 % оборота. По отношению к ABC-анализу Принцип Парето может прозвучать так: надёжное наблюдение за 20 % позиций позволяет на 80 % контролировать систему, будь то запасы сырья и комплектующих, продаваемые товары и т. д.

Результатом ABC-анализа является группировка объектов по степени влияния на общий результат.

В процессе ABC-анализа товары делятся на три категории:

А – наиболее ценные;

В – промежуточные;

С – наименее ценные.

По сути ABC-анализ – это ранжирование ассортимента по различным параметрам. Классифицировать так можно и поставщиков, и покупателей, и складские запасы, т. е. все, что имеет достаточное количество статистических данных. Результатом является группировка объектов по степени влияния на общий результат.

ABC-анализ основан на принципе дисбаланса. При его проведении строится график зависимости совокупного эффекта от количества рассмотренных элементов. Его называют Кривой Парето, Кривой Лоренца или ABC-кривой.

Обычно объектами ABC-анализа являются поставщики, товарные группы, товарные категории и отдельные товары. Каждый из них имеет разные параметры описания и измерения: объём продаж (в денежном или

количественном измерении), доход (в денежном измерении), товарный запас, оборачиваемость и т. д.

Выбор пороговых значений параметра, по которым производится отнесение объекта к одной из категорий, зависит от конкретных особенностей решаемой задачи.

Например, можно предположить, что узкий ассортимент категории А из 10 % товаров дает 70 % дохода. Из оставшихся – 20 % (товары категории В) дают 20 % дохода, а прочие – 70 % (категория С) – всего 10 %. Из полученных результатов можно сделать вывод, что наибольшее внимание следует уделять товарам категории А, несколько меньшее – категории В, а товары категории С можно вообще рассматривать как вспомогательные.

Алгоритм ABC-анализа:

1) определить объекты анализа (клиент, поставщик, товарная группа/подгруппа, номенклатурная единица и т. п.).

2) определить параметр, по которому будет проводиться анализ объекта (средний товарный запас, объем продаж, доход, количество единиц продаж, количество заказов и т. п.).

3) сортировка объектов анализа в порядке убывания значения параметра.

4) для определения принадлежности выбранного объекта к группам А, В или С необходимо (рисунок 12):

– рассчитать долю параметра от общей суммы параметров выбранных объектов;

– рассчитать эту долю с накопительным итогом;

– присвоить значения групп выбранным объектам.

Рекомендуемое распределение (рисунок 13):

♣ группа А – объекты, сумма долей с накопительным итогом которых составляет первые 50 % от общей суммы параметров;

♣ группа В – следующие за группой А объекты, сумма долей с накопительным итогом которых составляет от 50 % до 80 % от общей суммы параметров;

♣ группа С – оставшиеся объекты, сумма долей с накопительным итогом которых составляет от 80 до 100 % от общей суммы параметров.

! Данное распределение не является универсальным; определение объектов, параметров и долей групп должно проводиться с учетом специфики бизнеса.

ABC-анализ – наиболее популярный метод для изучения ассортиментной политики, который можно с большим успехом применить к управлению запасами. ABC-анализ обычно используют с целью отслеживания объемов отгрузки определенных артикулов и частоты обращений к той или иной позиции ассортимента, а также для ранжирования клиентов по количеству или объему сделанных ими заказов.

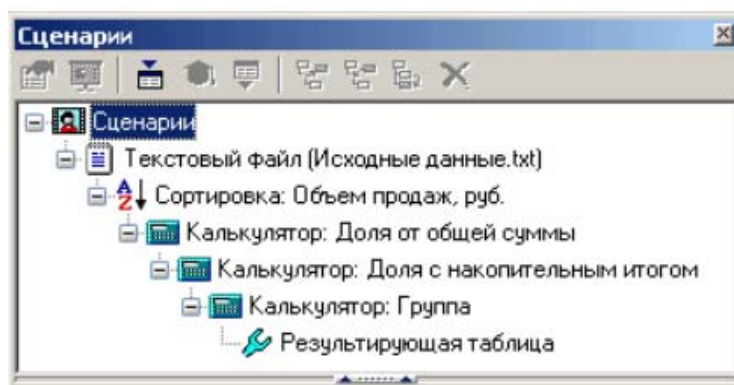


Рисунок 12 – Сценарий для определения принадлежности выбранного объекта к группам

Товар	Объем продаж, руб.	Группа
Товар 2	31003	A
Товар 6	30360	A
Товар 9	27135	A
Товар 10	25020	B
Товар 5	23100	B
Товар 8	21420	B
Товар 3	21294	C
Товар 7	18650	C
Товар 4	9035	C
Товар 1	2829	C

Рисунок 13 – Результирующая таблица по ABC-анализу

2.2 Выполнение ABC-анализа:

1) открыть новый проект в Deductor и сохранить его как L2.ded.

Скопируйте в Ваш рабочий каталог с данной лабораторной работой (№ 5) файл с ХД ФИО_Студента.Фармация (для того, чтобы не потерять данные из этого ХД).

– подключите к Вашему проекту ХД из рабочего каталога.

– используя Редактора метаданных (на вкладке «Подключения»), очистите все данные из ХД ФИО_Студента.Фармация.

Для этого сначала в процессе «Продажи» удалить все данные с помощью команды «Очистить» из всплывающего меню. А затем повторить эти действия для всех измерений.

2) импортировать в проект данные из текстовых файлов «Группы товаров.txt», «Отделы.txt», «Продажи.txt», «Товары.txt».

При помощи «Мастера экспорта» загрузить данные в ХД (учтите, что тип данных некоторых полей необходимо поменять на ЦЕЛЫЙ).

Сначала загружаем измерения, используя файлы «Группы товаров.txt», «Отделы.txt», «Товары.txt».

Потом загружаем процесс из файла «Продажи.txt».

3) импортировать данные из хранилища ФИО Студента. Фармация при помощи «Мастера импорта» вкладки «Сценарий» (рисунок 14), в качестве объекта указать процесс «Продажи». Импорт данных из ХД должен полностью соответствовать рисунку.

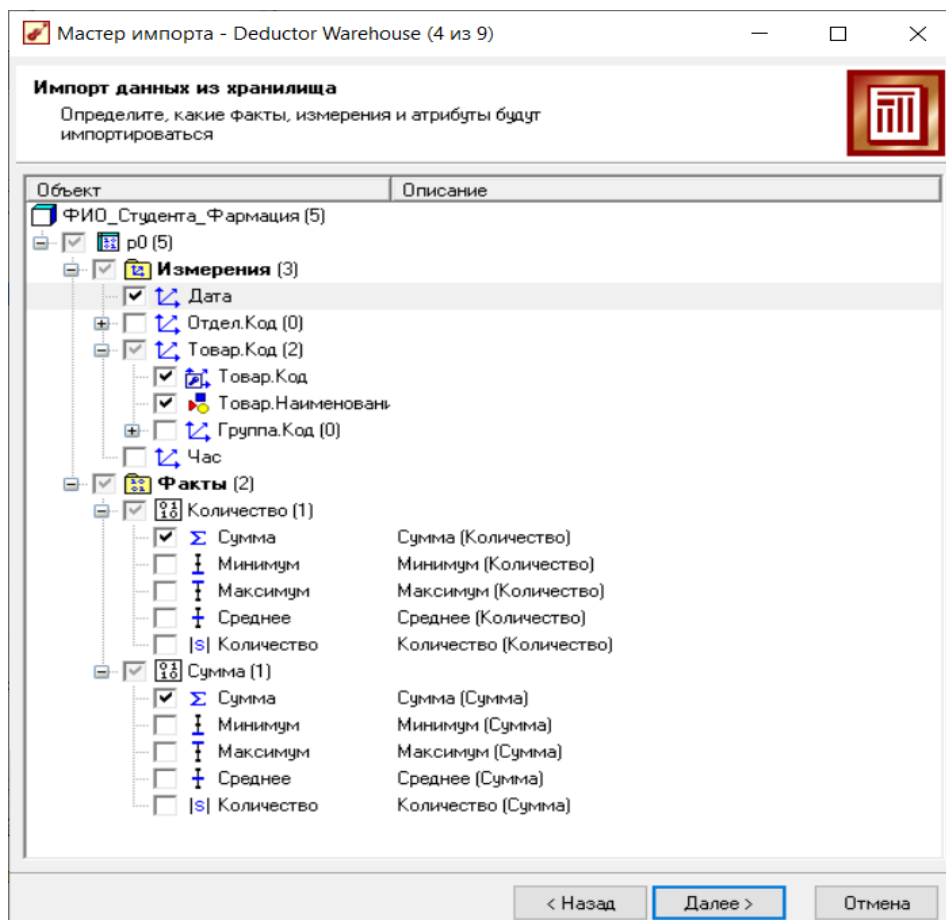


Рисунок 14 – Импорт данных из ХД

4) используя «Настройку набора данных», изменим наименования столбцов (рисунок 15).

Дата	Код товара	Наименование товара	Количество	Сумма
01.12.2008	477	Аспаркам табл. уп.контурн.б/яч. 10 Медисорб	3	13,34
01.12.2008	485	Аспаркам табл. уп.контурн.б/яч. 10 Усолье-Сибир	1	4,56
01.12.2008	1813	Дротаверин-Н.С. табл. 40 мг уп.контурн.яч. 10 пач	1	54,17
01.12.2008	2206	Йода раствор спиртовой 5% р-р спирт.наружн. 5 %	2	23,86
01.12.2008	2290	Кальций-Д_3 Никомед табл.жев. с апельс. фл. 20	1	179,38
01.12.2008	3381	Но-шпа табл. 40 мг бл. 10 кор. 2 Chinoïn	6	589,07
01.12.2008	4138	Сальбутамол аэроз.ингал.доз. 100 мкг/доза бал.	1	78,74
01.12.2008	4751	Тромбо АСС табл.п.п.о.раствор/кишечн. 50 мг бл.	1	80,98
01.12.2008	4752	Тромбо АСС табл.п.п.о.раствор/кишечн. 100 мг бл	1	85,53
01.12.2008	6100	Авиа-Море табл. уп.контурн.яч. 20 пач.картон. 1 М	2	116,73
01.12.2008	9220	Аммиака раствор р-р наружн. 10 % фл.темн.стекл	5	21,64
01.12.2008	9240	Аммиака раствор р-р наружн. 10 % фл.темн.стекл	1	6,98

Рисунок 15 – Изменение наименования столбцов

5) используя обработчик «Фильтр» (из «Мастера обработки»), оставляем все данные, кроме тех, что относятся к последнему месяцу. Это позволяет получить данные за зимние месяцы (декабрь, январь, февраль), что необходимо нам для выполнения ABC-анализа (рисунок 16).

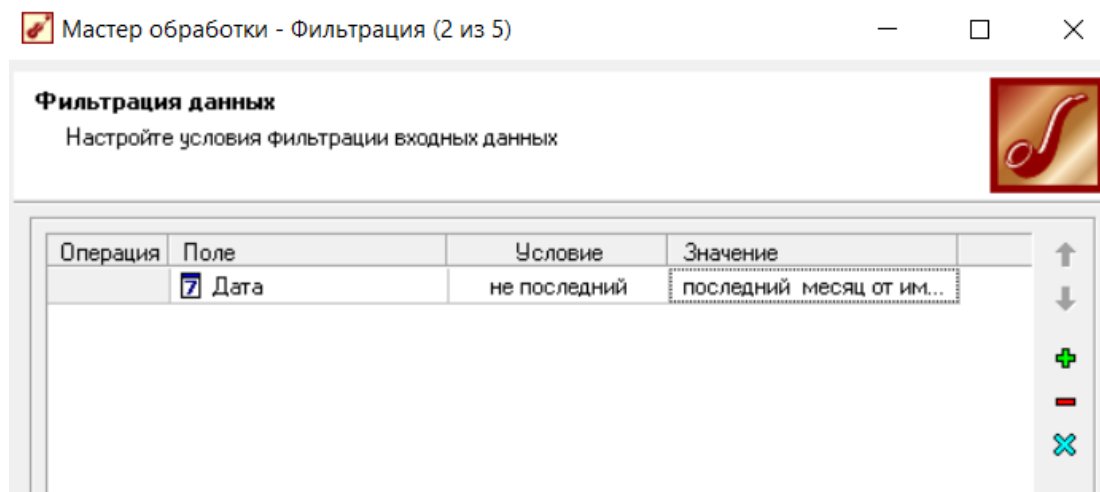


Рисунок 16 – Настройка условий фильтрации

6) получим сумму продаж за нужный период по каждому товару. Для этого используем обработку «Группировка» (рисунок 17).

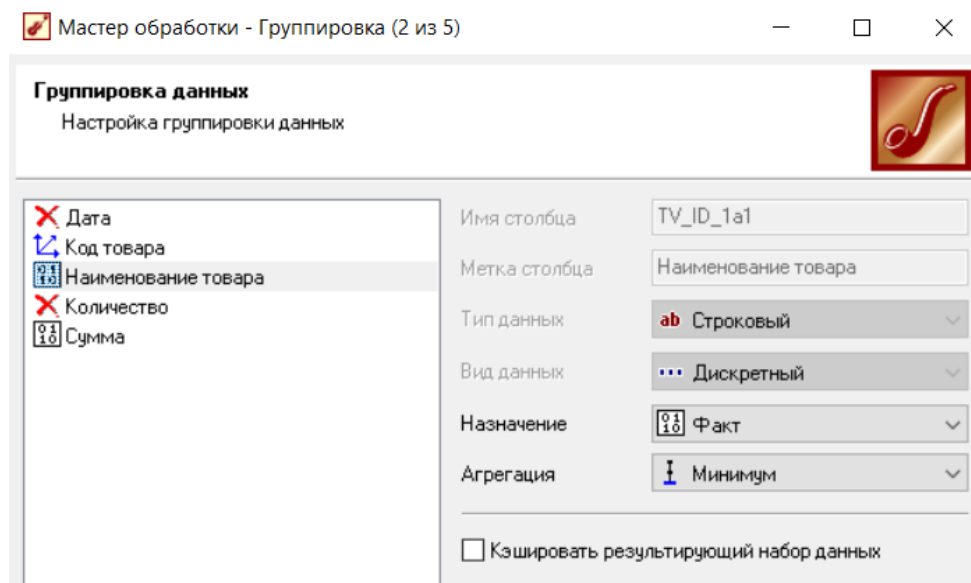


Рисунок 17 – Настройка обработки Группировка

Группировку выполняем по факту «Сумма», используя в качестве измерения столбец «Код товара». Результат обработки (рисунок 18).

Столбец «Наименование товара» устанавливаем как факт, выбрав «Минимум» в качестве метода агрегации.

Код товара	Наименование товара	Сумма
35	Адреналина гидрохлорида раствор 0,1% р-р д/ин.	66,85
354	Андрокур табл. 50 мг фл. 20 кор. 1 Schering-Plough	15673,79
477	Аспаркам табл. уп.контурн.б/яч. 10 Медисорб	233,27
485	Аспаркам табл. уп.контурн.б/яч. 10 Усолые-Сибир	4,56
487	Аспаркам табл. уп.контурн.яч. 10 Ай Си Эн Лекср	6,96
504	Аспирин табл. 500 мг бл. 10 кор. 2 Bayer С.С.	279,2
506	Аспирин табл. 100 мг бл. 10 кор. 2 Bayer С.С.	155,57
598	Ацетилсалициловая кислота табл. 0,5 г уп.контурн	530,63
682	АЦЦ 100 табл.шип. 100 мг туба 20 кор. 1 Hexal AG	3467,61
686	АЦЦ 200 табл.шип. 200 мг туба 20 кор. 1 Hexal AG	5828,74
690	АЦЦ лонг табл.шип. 600 мг туба 10 кор. 1 Hexal AG	6323,62
739	Беродуал р-р д/ингал. фл.-кап. 20 мл кор. 1 Boehrli	398,4
741	Беродуал Н аэроз.ингал.доз. 200 доз бал.аэроз. 1	4188,7
808	Бифидумбактерин пор. для приема внутрь пак.фл	1876,06
814	Бифидумбактерин пор. для приема внутрь пак.фл	1555,71
828	Бифидумбактерин сухой пор.лиоф.д/р-ра для приоб	2,34999999999
844	Бифидумбактерин сухой в свечах супп.ваг. уп.кон	235,16

Рисунок 18 – Результат обработки Группировка

7) отсортируем полученные данные в порядке убывания суммы (обработчик «Сортировка») (рисунок 19).

Код товара	Наименование товара	Сумма
16534	Боярышника настойка настойка фл. 100 мл Росби	272813,39
68036	Терпинкод табл. уп.контурн.яч. 10 Ай Си Эн Томск	115861,75
82545	Арбидол-ЛЭНС табл.п.о. 0,1 г бл. 10 кор. 1 Дальхи	98559,56
21269	Гексорал аэроз.наружн. 0,2 % бал.аэроз. 40 мл [с	47417,29
82849	ТераФлю от гриппа и простуды пор.д/р-ра для пр	38566,2
3382	Но-шпа табл. 40 мг конт. 100 кор. 1 Chinoi	36270,92
83436	Отривин спрей наз. 0,1 % фл.-кап. 10 мл кор. 1 Nc	30097,28
36259	Компливит табл.п.о. бан.полимерн. 60 пач.картон	27406,42

Рисунок 19 – Сортировка по убыванию суммы

8) рассчитаем долю выручки от продаж каждого товара в общей сумме. Для этого используем обработчик «Калькулятор»;

Доля от общей суммы вычисляется по формуле:

$$Part = \frac{Sum_i \cdot 100}{\sum_i Sum_i},$$

где Sum_i – сумма квартальных продаж i -го товара.

Для вычисления общей суммы $Sum_{\text{шт}}$ (суммарного значения) продаж следует воспользоваться формулой STAT: Stat ("ИмяСтолбца";"Тип") (рисунок 20).

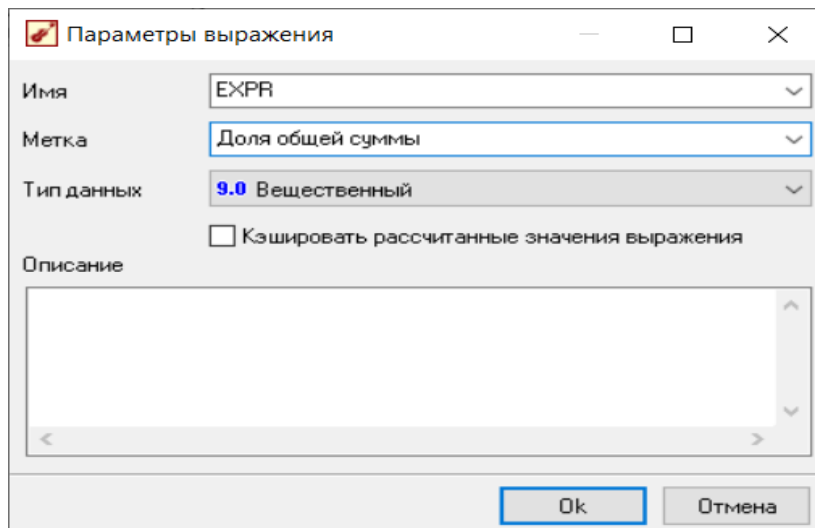


Рисунок 20 – Параметры поля «Выражение»

Функция возвращает одну из статистических характеристик данных.

Тип может принимать одно из следующих значений: Min – минимальное значение; Max – максимальное значение; Avg – среднее значение; StdDev – стандартное отклонение; Sum – сумма значений; SumSq – сумма квадратов значений; Count – количество значений; UniqueCount – количество уникальных значений; NullCount – количество пропущенных значений; Параметры: ИмяСтолбца – столбец, по которому необходимо получить элемент статистики; Тип – тип статистики, который необходимо получить.

Вычисления выполняем в поле «Выражение», предварительно задав округление значения до 2-х знаков (функция Round) (рисунок 21).

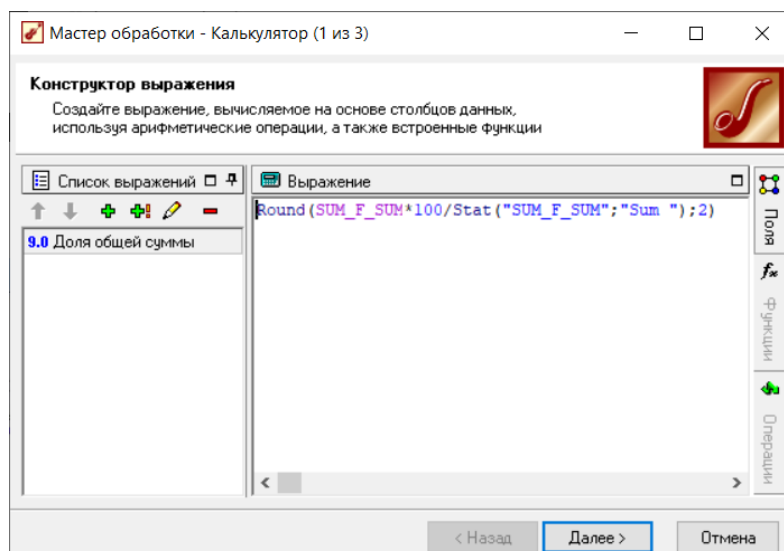


Рисунок 21 – Выполнение вычислений при помощи обработчика «Калькулятор»

Получаем в результате «Таблицу» (рисунок 22).

Код товара	Наименование товара	Сумма	Доля общей суммы
16534	Боярышника настойка настойка фл. 100 мл Росби	272813,39	10,04
68036	Терпинкод табл. уп. контурн. яч. 10 Ай Си Эн Томск	115861,75	4,26
82545	Арбидол-ПЗНС табл.п.о. 0,1 г бл. 10 кор. 1 Дальж	98559,56	3,63
21269	Гексорал аэроз. наружн. 0,2 % бал. аэроз. 40 мл [с	47417,29	1,75
82849	ТераФлю от гриппа и простуды пор.д/р-ра для пр	38566,2	1,42
3382	Но-шпа табл. 40 мг конт. 100 кор. 1 Chinoi	36270,92	1,33
83436	Отривин спрей наз. 0,1 % фл. кап. 10 мл кор. 1 Nc	30097,28	1,11
36259	Компливит табл.п.о. бан. полимерн. 60 пач. картон	27406,42	1,01
3381	Но-шпа табл. 40 мг бл. 10 кор. 2 Chinoi	27290,95	1
84190	Седальгин-Нео табл. уп. контурн. яч. 10 Balkanphar	24656,44	0,91

Рисунок 22 – Результирующая таблица

9) рассчитаем долю с накопительным итогом, воспользовавшись обработчиком «Калькулятор». Доля с накопительным итогом вычисляется по формуле:

$$ComulativeSum_i = ComulativeSum_{i-1} + Part_i,$$

где $ComulativeSum_i$ – доля с накопительным итогом для i-го товара;

$Part_i$ – доля от общей суммы для i-го товара.

В обработчике «Калькулятор» для вычисления доли с накопительным итогом есть специальная функция CUMULATIVESUM (рисунок 23):

CumulativeSum ("ИмяПоля"; ["ИмяПоляГруппы"])

Функция возвращает накапливающуюся сумму по столбцу.

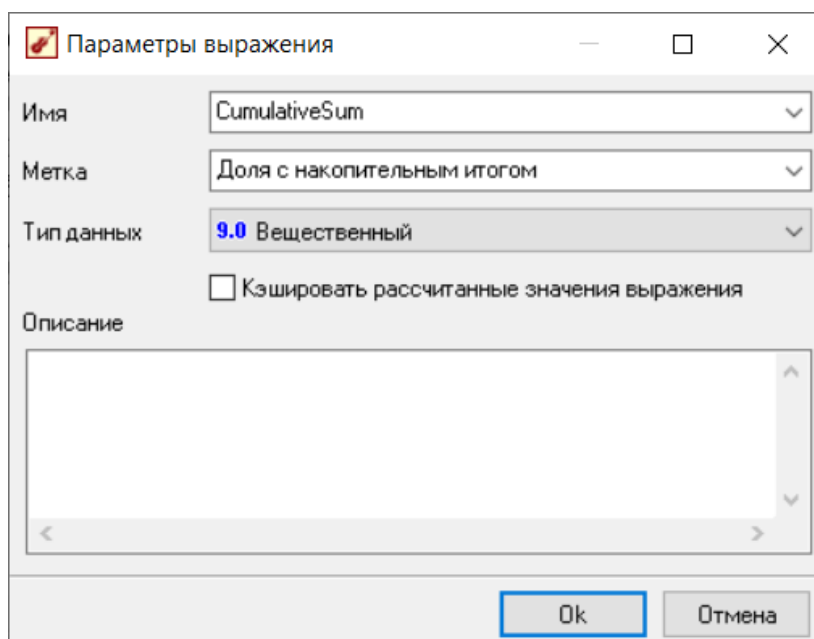


Рисунок 23 – Применение функции CUMULATIVESUM

Параметры:

Имя «Поля» – поле, по которому необходимо получить накапливающуюся сумму

Имя «Поля Группы» – поле, по которому происходит группировка данных. Этот параметр не является обязательным. Если этот параметр указан, то накапливающиеся суммы считаются в пределах групп.

Вычисления выполняем в поле «Выражение», предварительно задав параметры выражения (рисунок 24).

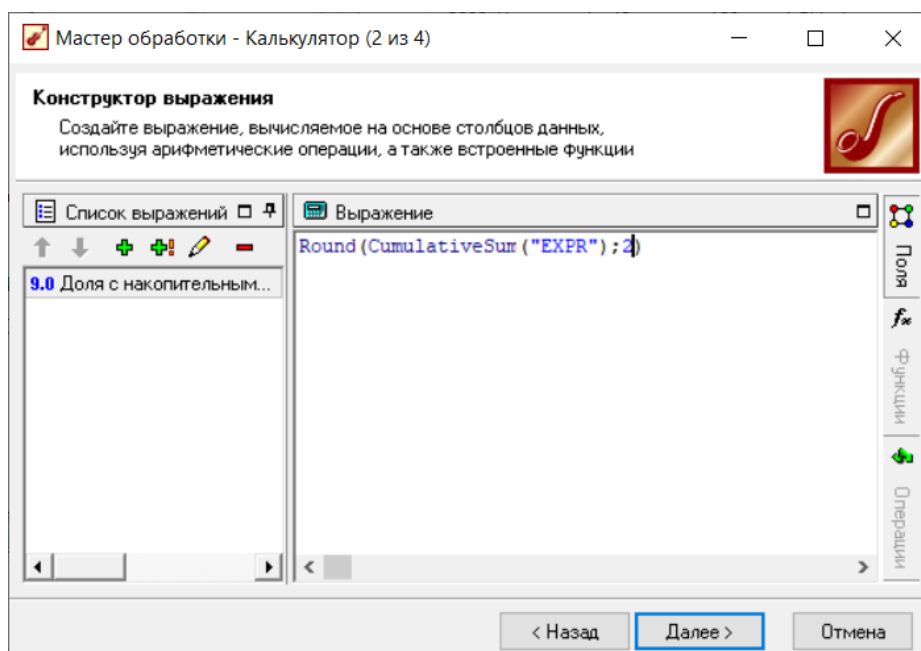


Рисунок 24 – Применение встроенных функций

Получаем результат в таблице (рисунок 25):

Код товара	Наименование товара	Сумма	Доля общей суммы	Доля с накопительным итогом
16534	Боярышника настойка настойка фл. 100 мл Росби	272813,39	10,04	10,04
68036	Терпинкод табл. уп. контурн. яч. 10 Ай Си Эн Томск	115861,75	4,26	14,3
82545	Арбидол-ЛЭНС табл.п.о. 0,1 г бл. 10 кор. 1 Дальж	98559,56	3,63	17,93
21269	Гексорал аэроз. наружн. 0,2 % бал. аэроз. 40 мл [с	47417,29	1,75	19,68
82849	ТераФлю от гриппа и простуды пор.д/р-ра для пр	38566,2	1,42	21,1
3382	Но-шпа табл. 40 мг конт. 100 кор. 1 Chinoi	36270,92	1,33	22,43
83436	Отривин спрей наз. 0,1 % фл.-кап. 10 мл кор. 1 Нс	30097,28	1,11	23,54
36259	Компливит табл.п.о. бан. полимерн. 60 пач. картон	27406,42	1,01	24,55
3381	Но-шпа табл. 40 мг бл. 10 кор. 2 Chinoi	27290,85	1	25,55
84190	Седальгин-Нео табл. уп. контурн. яч. 10 Balkanphar	24656,44	0,91	26,46
85596	Чабреца трава трава пак. бум. 50 г пач. картон. 1 Т	23139,37	0,85	27,31
35694	Коделак табл. уп. контурн. яч. 10 Ай Си Эн Томскж	21942,66	0,81	28,12
36554	Корвалол капли для приема внутрь фл.-кап. темн.	19839,78	0,73	28,85
20852	Галазолин капли наз. 0,1 % фл. 10 мл Warszawski	18120,1	0,67	29,52
20588	Вобэнзим табл.п.о. раствор./кишечн. бл. 20 кор. 1	17653,25	0,65	30,17
27289	Длянос спрей наз. 0,1 % фл.с пульт. 10 мл кор. 1	17508,74	0,64	30,81
83149	Арбидол-ЛЭНС табл.п.о. 0,05 г уп. контурн. яч. 10 п	17418,24	0,64	31,45
1151	Виферон супп. рект. 150000 МЕ уп. контурн. яч. 10 п	15925,14	0,59	32,04
354	Андрокур табл. 50 мг фл. 20 кор. 1 Schering-Plough	15673,79	0,58	32,62


Рисунок 25 – Результат в таблице

10) присвоить значение группы (рисунок 26) каждому товару по следующей схеме (рисунок 27):

если $CumulativeSum_i < 75$, то это группа А;

если $75 \leq CumulativeSum_i < 90$, то это группа В;

если $CumulativeSum_i \geq 90$, то это группа С;

Параметры калькулятора отображены ниже (для добавления выражения калькулятора в сценарий используйте кнопку ).

Для возвращения номера текущей записи источника данных воспользуйтесь функцией RowNum() (рисунок 28):

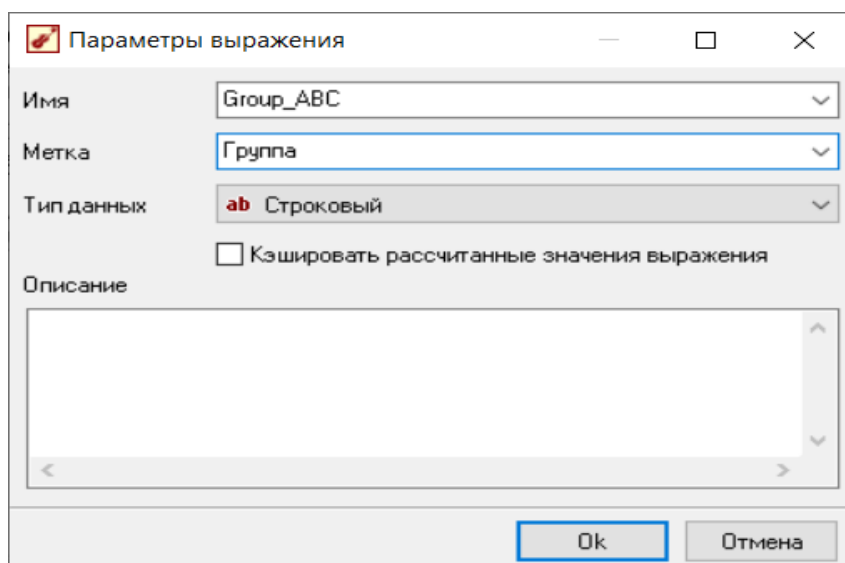


Рисунок 26 – Создание нового поля Группа

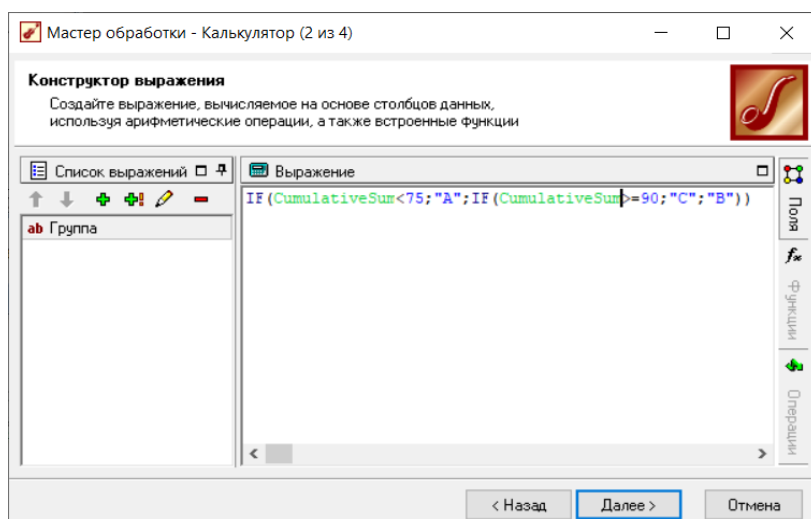


Рисунок 27 – Задание параметров выражения

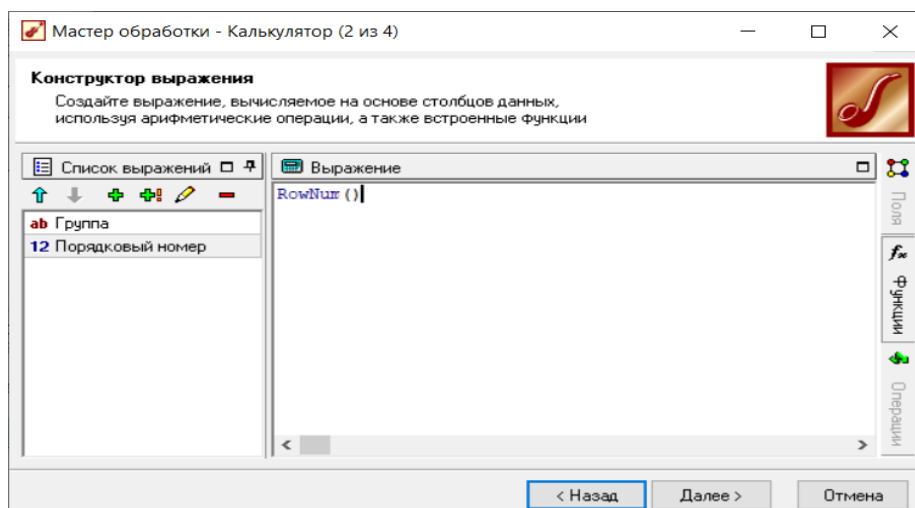


Рисунок 28 – Применение функции RowNum()

11) результат отображаем в виде «Таблицы», «Куба» и «Диаграммы» (используйте «Мастер визуализации») (рисунок 29).

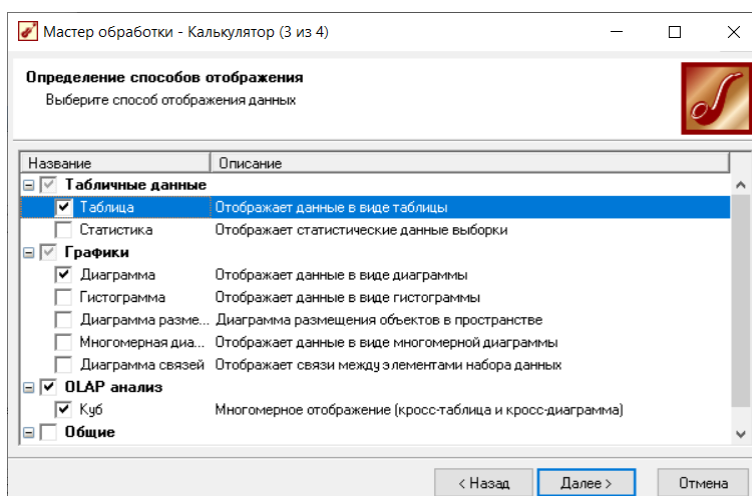


Рисунок 29 – Выбор «Мастера визуализации»

11.1 Настройка параметров «Диаграммы» (рисунок 30).

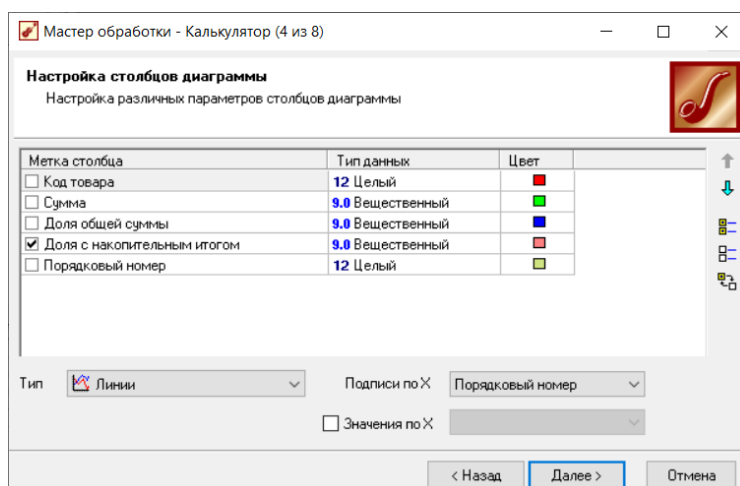


Рисунок 30 – Настройка столбцов диаграммы

11.2 Настройка параметров «Куба» (рисунок 31).

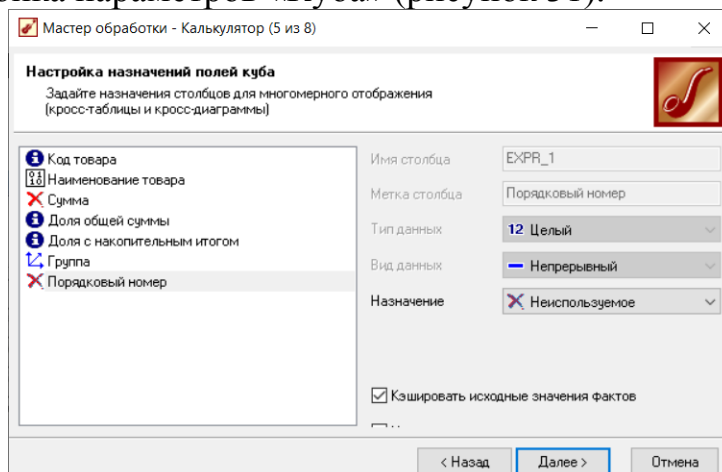


Рисунок 31 – Настройка параметров полей «Куба»

В отчете нас интересует только «Группа» и «Наименование товаров», имеющих соответствующую группу. Поэтому:

- столбец «Группа» отмечаем как измерение;
- наименование товаров отмечаем как факт;
- столбцы «Сумма» и «Порядковый номер» – неиспользуемое;
- остальные помечаем как информационные.

Группу как единственное измерение помещаем в строки кросс-таблицы (рисунок 32), а в качестве фактов (рисунок 33) используем параметр «Количество».

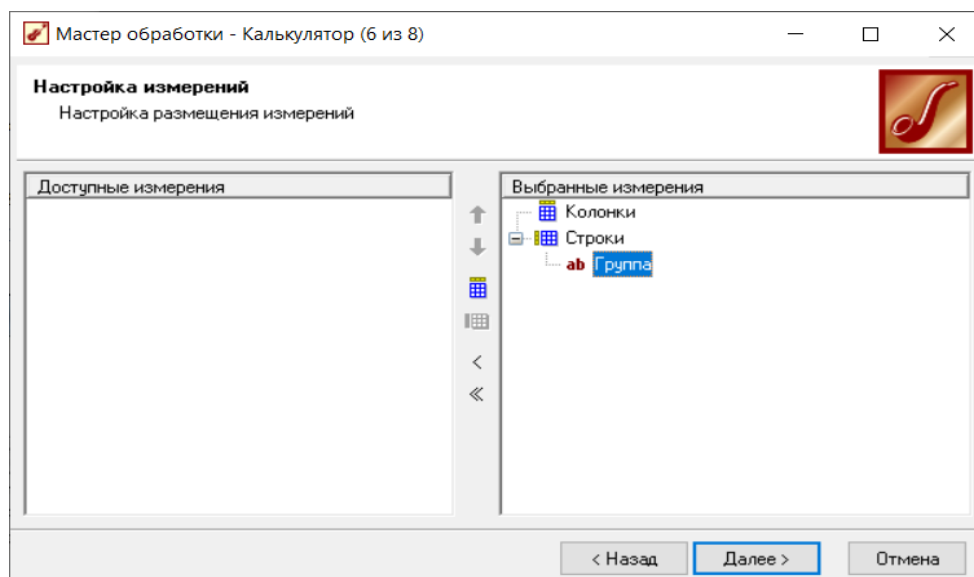


Рисунок 32 – Настройка параметров размещения измерений

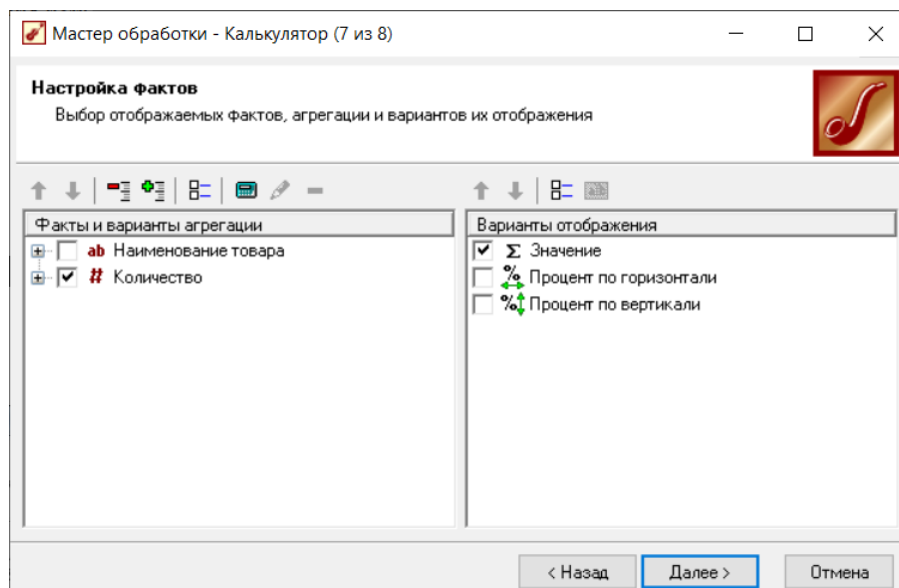


Рисунок 33 – Выбор отображения фактов

Результаты приведены на рисунках 34, 35, 36.

Код товара	Наименование товара	Сумма	Доля общей суммы	Доля с накопительным итогом	Группа	Порядковый номер
16534	Боярышника настойка настойка фл. 100 мл Росби	272813,39	10,04	10,04 A		1
68036	Терпинкод табл. уп.контурн.яч. 10 Ай Си Эн Томск	115861,75	4,26	14,3 A		2
82545	Арбидол-ЛЭНС табл.п.о. 0,1 г бл. 10 кор. 1 Дальж	98559,56	3,63	17,93 A		3
21269	Гексорал аэроз.наружн. 0,2 % бал.аэроз. 40 мл [с	47417,29	1,75	19,68 A		4
82649	ТераФлю от гриппа и простуды пор.д/р-ра для пр	38566,2	1,42	21,1 A		5
3382	Но-шпа табл. 40 мг конт. 100 кор. 1 Chinoip	36270,92	1,33	22,43 A		6
83436	Отривин спрей наз. 0,1 % фл.-кап. 10 мл кор. 1 Nc	30097,28	1,11	23,54 A		7
36259	Компливит табл.п.о. бан.полимерн. 60 пач.картон	27406,42	1,01	24,55 A		8
3381	Но-шпа табл. 40 мг бл. 10 кор. 2 Chinoip	27290,85	1	25,55 A		9
84190	Седальгин-Нео табл. уп.контурн.яч. 10 Balkanphar	24656,44	0,91	26,46 A		10
85596	Чабреца трава трава пак.бум. 50 г пач.картон. 1 1	23139,37	0,85	27,31 A		11
35694	Коделак табл. уп.контурн.яч. 10 Ай Си Эн Томский	21942,66	0,81	28,12 A		12
36554	Корвалол капли для приема внутрь фл.-кап.темн.	19839,78	0,73	28,85 A		13
20852	Галазолин капли наз. 0,1 % фл. 10 мл Warsawski	18120,1	0,67	29,52 A		14
20588	Вобэнзим табл.п.о.раствор./кишечн. бл. 20 кор. 1	17653,25	0,65	30,17 A		15
27289	Дялинос спрей наз. 0,1 % фл.с пультв. 10 мл кор. 1	17508,74	0,64	30,81 A		16
82148	Арбидол-ЛЭНС табл.п.о. 0,1 г бл. 10 кор. 1 Дальж	17418,24	0,63	31,45 A		17

Рисунок 34 – Таблица

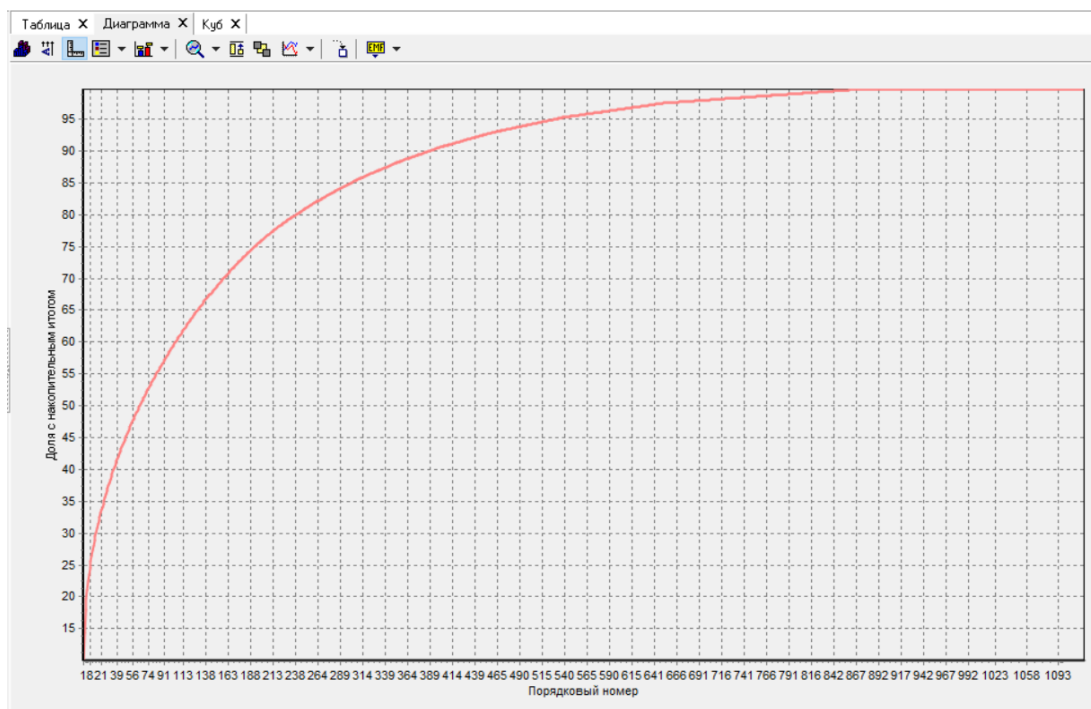


Рисунок 35 – Диаграмма

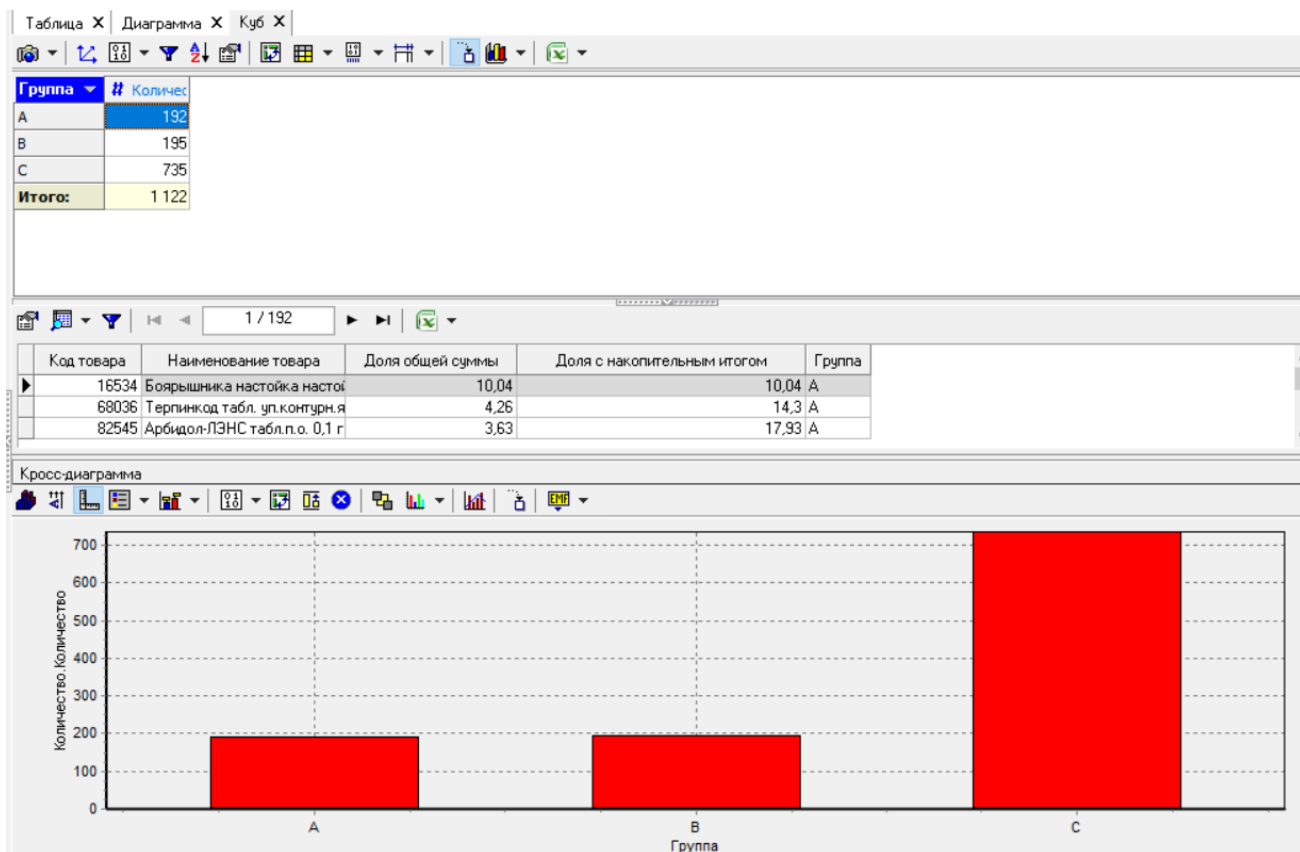


Рисунок 36 – Куб

Сохраните результаты работы в файле L2.ded.

3 Проведение XYZ-анализа.

3.1 Суть метода

Анализ XYZ – метод, позволяющий произвести классификацию ресурсов компании в зависимости от характера их потребления и точности прогнозирования изменений в их потребности.

XYZ-анализ ресурсов предполагает оценку их значимости в зависимости от частоты потребления. Если рассматривать потребление отдельных видов ресурсов в течение длительного периода, то можно установить, что в их числе есть:

- ресурсы, имеющие постоянный и стабильный спрос;
- ресурсы, расход которых подвержен определенным, например, сезонным, колебаниям;
- ресурсы, расход которых носит случайный характер.

Поэтому в пределах каждого из классов А, В и С ресурсы могут быть распределены еще и по степени прогнозируемости их расхода. Для такой классификации используются символы X, Y, Z.

К классу X относятся ресурсы, спрос на которые имеет постоянный характер или подвержен случайным незначительным колебаниям, поэтому поддается прогнозированию с высокой точностью. Удельный вес таких ресурсов в общей номенклатуре, как правило, не превышает 50–55 %.

К классу Y относятся ресурсы, потребление которых осуществляется периодически либо имеет характер падающей или восходящей тенденции. Их

прогнозирование возможно со средней степенью точности. Их удельный вес в общей номенклатуре составляет около 30 %.

К классу Z относятся ресурсы, для которых нельзя выявить какой-либо закономерности потребления. По этой причине прогнозирование их расхода невозможно (они составляют 15 % общей номенклатуры).

Соответствующий алгоритм содержит четыре этапа:

1) определение коэффициентов вариации для анализируемых ресурсов как отношения среднеквадратического отклонения к среднеарифметическому значению измеряемых параметров

$$v = \frac{\sqrt{(\sum_{i=1}^n x_i - \bar{x})/n}}{\bar{x}} \cdot 100 \%,$$

где X_i – значение параметра по оцениваемому объекту за i -й период, n – число периодов.

2) группировка ресурсов в соответствии с возрастанием коэффициента вариации параметров.

3) распределение по категориям X, Y, Z.

4) графическое представление результатов анализа (рисунок 37).

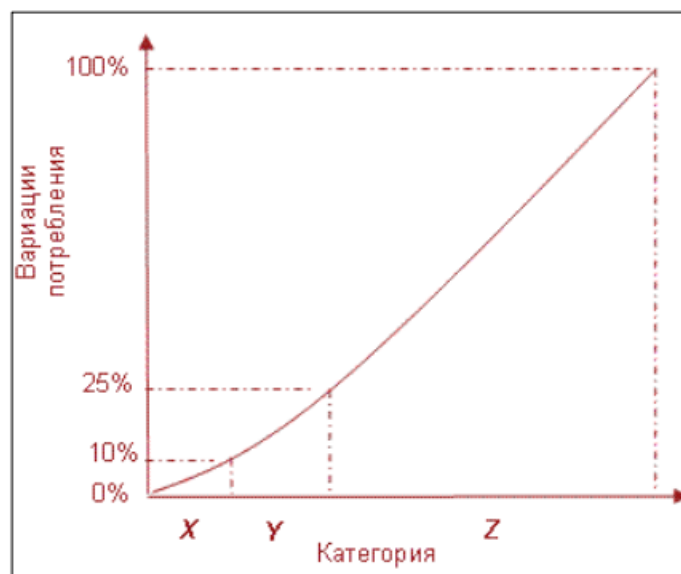


Рисунок 37 – Графические результаты анализа X, Y, Z

К категории X относят ресурсы, характеризующиеся стабильной величиной потребления, незначительными колебаниями в их расходе и высокой точностью прогноза (т. е. с вероятностью 99 % можно утверждать, что прогноз на данную группу товаров будет отклоняться от среднего значения на 10 %). Значение коэффициента вариации находится в интервале от 0 до 10 %.

Категория Y – ресурсы, для которых имеют место некоторые колебания потребности в них (например, сезонные) и средние возможности их прогнозирования. Значение коэффициента вариации – от 10 до 25 %.

Категория Z – ресурсы, потребление которых нерегулярно, точность прогнозирования невысокая. Значение коэффициента вариации – свыше 25 %.

Другим приложением XYZ-анализа является оптимизация ассортимента товаров или продукции. Тогда в категорию X попадает то, что дает стабильный и высокий доход, в категорию Y – средний, а в Z – падающий низкий и нестабильный.

В зависимости от особенностей бизнеса пороговые значения между категориями могут меняться. Например, для категории X может быть выбран диапазон 0–15 %, для категории Y – 15–50 %, а для категории Z – 50–100 %.

Алгоритм XYZ-анализа

1) определить объекты анализа (клиент, поставщик, товарная группа/подгруппа, номенклатурная единица и т. п.).

2) определить параметр, по которому будет проводиться анализ объекта (средний товарный запас, объем продаж, доход, количество единиц продаж, количество заказов и т. п.).

3) определить период и количество периодов, по которым будет проводиться анализ (неделя, месяц, квартал, полугодие, год).

4) определить коэффициент вариации для каждого объекта анализа:

– определить средние продажи за месяц (агрегация при группировке не требует дополнительного введения формул);

– определить стандартное отклонение вариационного ряда (агрегация при группировке не требует дополнительного введения формул).

5) отсортировать объекты анализа по возрастанию значения коэффициента вариации.

6) определить группы X, Y и Z.

Сценарий XYZ-анализа (рисунок 38).

Результирующая таблица по XYZ-анализу (рисунок 39):

Рекомендуемое распределение:

– группа X – объекты, коэффициент вариации по которым не превышает 10 %.

– группа Y – объекты, коэффициент вариации по которым составляет 10–25 %.

– группа Z – объекты, коэффициент вариации по которым превышает 25 %.

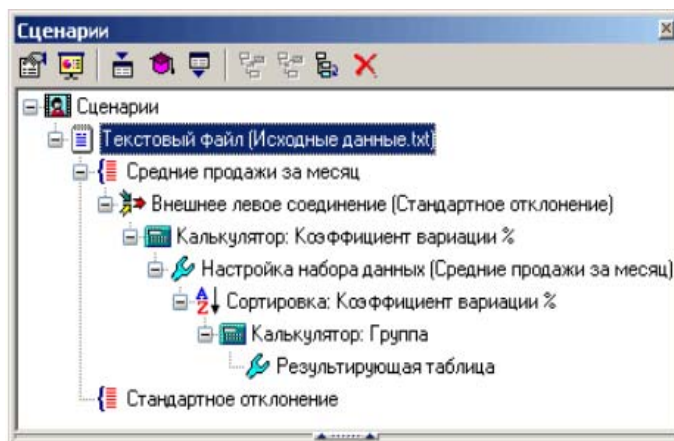


Рисунок 38 – Сценарий XYZ-анализ

Товар	Группа
Товар 2	X
Товар 4	X
Товар 9	X
Товар 3	X
Товар 5	Y
Товар 10	Z
Товар 7	Z
Товар 1	Z
Товар 6	Z
Товар 8	Z

Рисунок 39 – Результирующая таблица по XYZ-анализу

! Данный метод анализа имеет смысл, если количество анализируемых периодов больше трех; чем больше количество периодов, тем более показательными будут результаты. При этом сам период должен быть не меньше, чем горизонт планирования, принятый в компании.

XYZ-анализ представляет интерес для дистрибьюторов и производителей, имеющих свои склады. Любая закупка связана с большими издержками для компании (логистика, хранение и т. д.), а также с прямыми рисками, например, списание товара по сроку годности. Ведение точной сбалансированной закупки является приоритетной задачей как оптового, так и розничного предприятия. Применяя XYZ-анализ в отношении своих клиентов, можно строить прогноз продаж на будущие периоды, разрабатывать специальные программы для постоянных лояльных (не подверженным различным всплескам заказов) клиентов. А также проводить различные мероприятия по переводу клиентов из группы Y, Z в группу X.

3.2 Выполнение XYZ-анализа

1) переименуем узел «Настройка набора данных» в ABC-анализ.

Применим к ветке сценария ФИО_Студента.Фармация: Продажи обработчик «Настройки набора данных» для изменения наименования столбцов (дадим название этому узлу – XYZ-анализ) (рисунок 40).

Дата	Код товара	Наименование товара	Количество	Сумма
01.12.2008	477	Аспаркам табл. уп.контурн.б/	3	13,34
01.12.2008	485	Аспаркам табл. уп.контурн.б/	1	4,56
01.12.2008	1813	Дрогаверин Н.С. табл. 40 мг	1	54,17
01.12.2008	2206	Йода раствор спиртовой 5% р	2	23,86
01.12.2008	2290	Кальций-Д_3 Никомед табл.л	1	179,38
01.12.2008	3381	Ношпа табл. 40 мг бл. 10 кор	6	589,07
01.12.2008	4138	Сальбутамол аэроз.ингалядо	1	78,74
01.12.2008	4751	Тромбо АСС табл.п.о.раств	1	80,98
01.12.2008	4752	Тромбо АСС табл.п.о.раств	1	85,53
01.12.2008	6100	Авиа-Море табл. уп.контурн.я	2	116,73
01.12.2008	9220	Аммиака раствор р-р наружн	5	21,64
01.12.2008	9240	Аммиака раствор р-р наружн	1	6,98
01.12.2008	9943	Анальгин-хвиеин табл.п.о. уп.к	1	31,18
01.12.2008	10208	Андипал табл. уп.контурн.б/а	5	33,25
01.12.2008	10209	Андипал табл. уп.контурн.б/а	4	26,28

Рисунок 40 – Результат использования обработчика Фильтр

Используя обработчик «Фильтр» (из «Мастера обработки»), оставляем все данные, кроме тех, что относятся к последнему месяцу (см. аналогичные пункты при проведении ABC-анализа).

2) применяем обработчик «Дата и время». Необходимо преобразовать дату таким образом, чтобы в ней фигурировали только месяц и год (без конкретизации дня месяца). Для этого в качестве выходного параметра устанавливаем Год+Месяц (т. е. все, что продано в декабре 2008 года, будет отмечено как 01.12.2008) (рисунок 41).

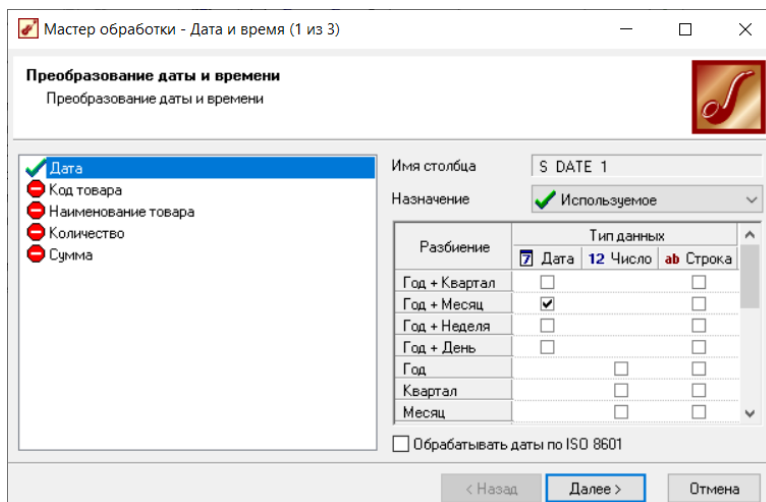


Рисунок 41 – Применение обработчика «Дата и время»

3) выполняем группировку полученных данных по месяцам (рисунок 42): Для этого оставляем два измерения «Дата» (Год + Месяц) и «Код товара», по фактам «Количество» и «Сумма» выполняем агрегацию по параметру «Сумма», а для факта «Наименование товара» – «Минимум».

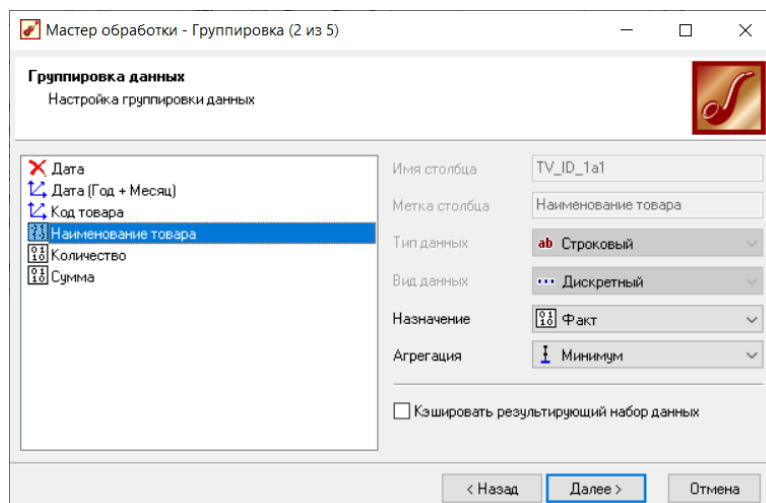


Рисунок 42 – Выполнение группировки

Видно, что таблица содержит число проданных единиц товара и сумму вырученных средств от продажи этого товара за каждый месяц необходимого периода.

4) вычисляем среднее число проданных единиц товара за период. Эту операцию выполняют последовательно две ветки сценария – «Группировка по товарам» и «Калькулятор».

Первая ветка позволяет посчитать общее число проданных единиц товара за квартал, а вторая – получить среднее значение этого параметра за три месяца. Результат представлен на рисунке 43.

Дата (Год + Месяц)	Код товара	Наименование товара	Количество	Сумма
01.12.2008	35	Адреналина гидрохлорида ра	2	66,85
01.12.2008	477	Аспаркам табл. уп. контурн. б/	46	206,3
01.12.2008	485	Аспаркам табл. уп. контурн. б/	1	4,56
01.12.2008	487	Аспаркам табл. уп. контурн. яч.	1	6,96
01.12.2008	682	АЩЦ 100 табл.шип. 100 мг туб	7	1374,22
01.12.2008	686	АЩЦ 200 табл.шип. 200 мг туб	11	2378,61
01.12.2008	690	АЩЦ лонг табл.шип. 600 мг туб	15	2714,98
01.12.2008	739	Беродуал р-р д/ингал. Фл.каг	1	398,4
01.12.2008	741	Беродуал Н аэроз.ингал.доз.	1	859,07
01.12.2008	808	Бифидумбактерин пор. для п	4	1065,88
01.12.2008	814	Бифидумбактерин пор. для п	4	372,26
01.12.2008	828	Бифидумбактерин сухой пор.	39	3227,44
01.12.2008	844	Бифидумбактерин сухой в св	2	119,96
01.12.2008	849	Бифидумбактерин форте пор.	1	122,34
01.12.2008	1137	Витачин Кальций + Витамин	2	658,31

Рисунок 43 – Результат выполнения группировки

4.1 Выполняем группировку полученных данных по товарам. В качестве фактов используем столбец «Количество» (агрегация – «Сумма») и столбец «Наименование товара» (агрегация «Минимум») (рисунок 44).

В качестве измерения – «Код товара».

Это позволит получить таблицу, содержащую данные о количестве проданных единиц товара за кварталый период.

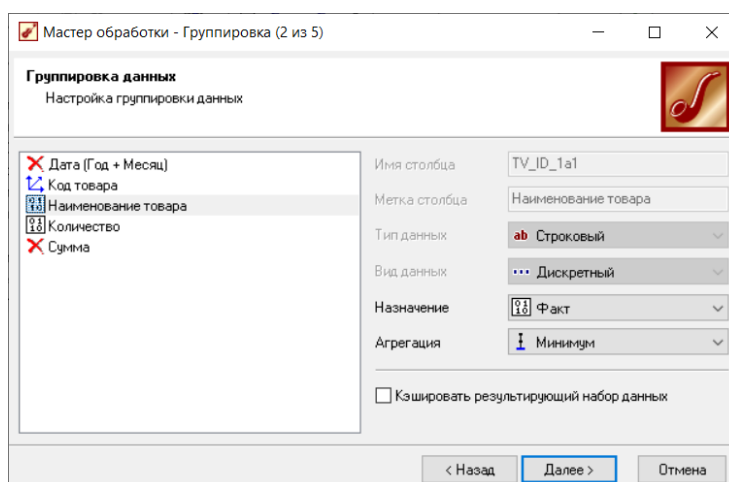


Рисунок 44 – Группировка полученных данных

4.2 Получаем среднее число продаж за квартал. Для этого используем обработчик «Калькулятор» (рисунки 45, 46).

Результат представлен на рисунке 47.

5) вычисление коэффициента вариации. Для этого необходимо вычислить:

- величину отклонения числа проданных единиц товара от средней величины продажи данного товара за период;
- квадрат отклонения числа проданных единиц товара от среднего значения.

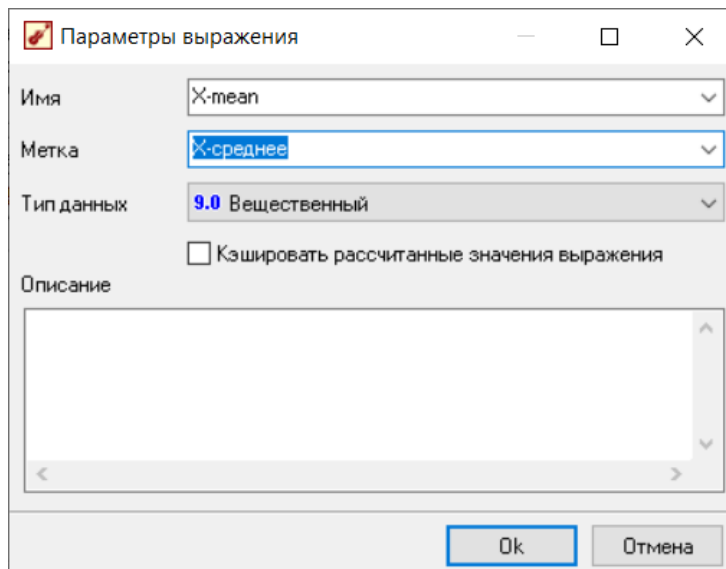


Рисунок 45 – Настройка среднего числа продаж за квартал

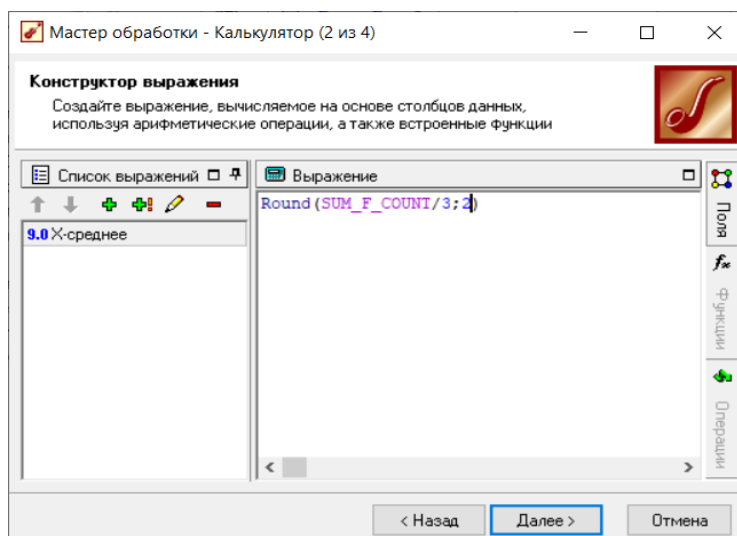


Рисунок 46 – Обработчик Калькулятор

Код товара	Наименование товара	Количество	X-среднее
35	Адреналина гидрохлорида ра	2	0,67
354	Андрокур табл. 50 мг фл. 20 к	6	2
477	Аспаркам табл. уп.конгурн.б/	52	17,33
485	Аспаркам табл. уп.конгурн.б/	1	0,33
487	Аспаркам табл. уп.конгурн.я	1	0,33
504	Аспирин табл. 500 мг бл. 10 к	3	1
506	Аспирин табл. 100 мг бл. 10 к	2	0,67
598	Ацетилсалициловая кислота	238	79,33
682	АЦЦ 100 табл.шип. 100 мг туб	18	6
686	АЦЦ 200 табл.шип. 200 мг туб	27	9
690	АЦЦ лонг табл.шип. 600 мг туб	35	11,67
739	Беродуал р-р д/ингал. фл.кап	1	0,33
741	Беродуал Н аэроз.ингалдоз.	5	1,67
808	Бифидумбактерин пор. для пр	7	2,33
814	Бифидумбактерин пор. для пр	16	5,33
828	Бифидумбактерин сухой пор.	108	36

Рисунок 47 – Среднее число проданных единиц товара за период

5.1 Вычислим величину отклонения числа проданных единиц товара от средней величины продажи данного товара за период.

Для выполнения этой операции необходимо числу продаж в каждый месяц поставить в соответствие среднее число продаж за квартал, после чего вычислить квадрат их разности.

Первое действие выполняется при помощи обработчика «Слияние с узлом» (рисунок 48), второе – при помощи обработчика «Калькулятор».

Применяем обработчик «Слияние с узлом» для узла «Группировка» по периоду «Месяц». Выбираем в качестве типа слияние – «Внутреннее соединение», а в качестве узла – «Калькулятор: X-среднее». Такой тип слияния позволит получить таблицу, состоящую только из тех строк, для которых найдено совпадение по ключевым полям каждого из узлов.

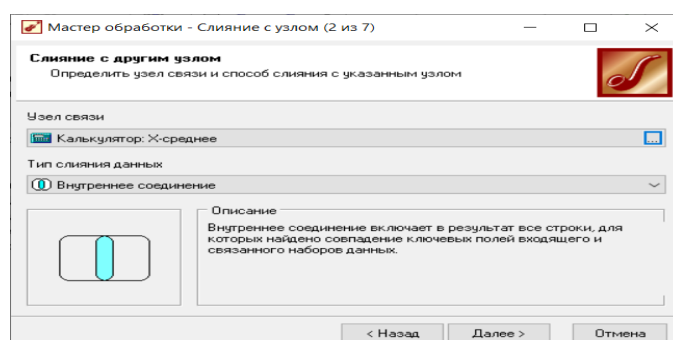


Рисунок 48 – Обработчик «Слияние с узлом»

В качестве ключевого поля, по которому будет производиться слияние, выбираем поле «Код товара» (рисунок 49).

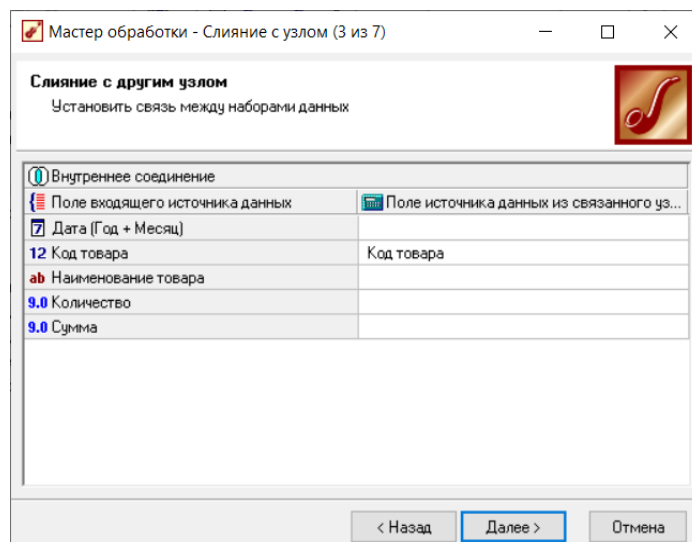


Рисунок 49 – Выбор ключевого поля

На следующем шаге обработчика выбираем те поля узлов, которые мы хотим увидеть в результирующей таблице (рисунок 50).

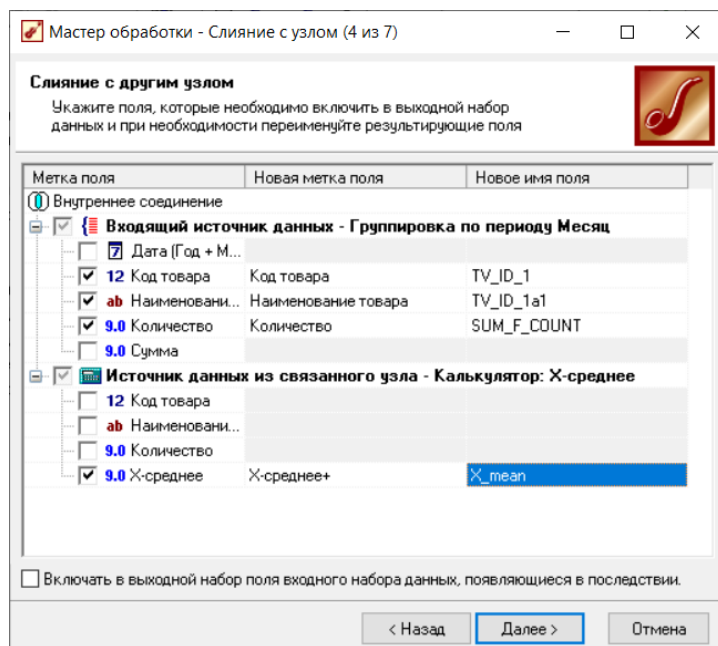


Рисунок 50 – Выбор полей необходимых в выходном наборе данных

Результирующая таблица имеет следующий вид (рисунок 51).

Код товара	Наименование товара	Количество	X-среднее+
35	Адреналина гидрохлорида ра	2	0,67
477	Аспаркам табл. уп. контурн. б/	46	17,33
485	Аспаркам табл. уп. контурн. б/	1	0,33
487	Аспаркам табл. уп. контурн. яч	1	0,33
682	АЦЦ 100 табл.шип. 100 мг туб	7	6
686	АЦЦ 200 табл.шип. 200 мг туб	11	9
690	АЦЦ лонг табл.шип. 600 мг туб	15	11,67
739	Беродуал р-р д/ингал. фл.кап	1	0,33
741	Беродуал Н аэроз.ингал.доз.	1	1,67
808	Бифидумбактерин пор. для пр	4	2,33
814	Бифидумбактерин пор. для пр	4	5,33
828	Бифидумбактерин сухой пор.	39	36
844	Бифидумбактерин сухой в св	2	1,33
849	Бифидумбактерин форте пор.	1	3
1137	Витрум Кальциум + Витамин	2	1,33
1138	Витрум Кальциум + Витамин	1	1
1145	Виферон супп. рект. 500000 М	7	7,33

Рисунок 51 – Таблица вычисления величины отклонения числа проданных единиц товара от средней величины продажи данного товара за период

5.2 Вычислим квадрат отклонения числа проданных единиц товара от среднего значения (рисунки 52, 53).

5.3 Для вычисления коэффициента вариации воспользуемся, обработчиками «Группировка» (рисунок 55) и «Калькулятор» (рисунки 56, 57).

Применяем к полученной на предыдущем шаге таблице обработчик «Группировка» по измерению «Код товара», что позволит нам получить сумму квадратов отклонений числа проданных единиц товара от средней величины.

Для факта SKO выбираем способ агрегации – «Сумма».

Для факта «Наименования товара» – «Минимум».

Для факта X_MEAN – «Первый из списка».

Вычисляем коэффициент вариации, используя «Калькулятор».

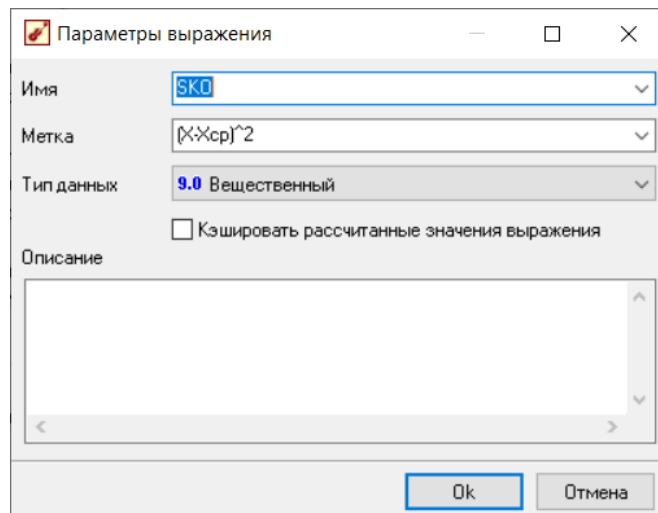


Рисунок 52 – Настройка параметров выражения

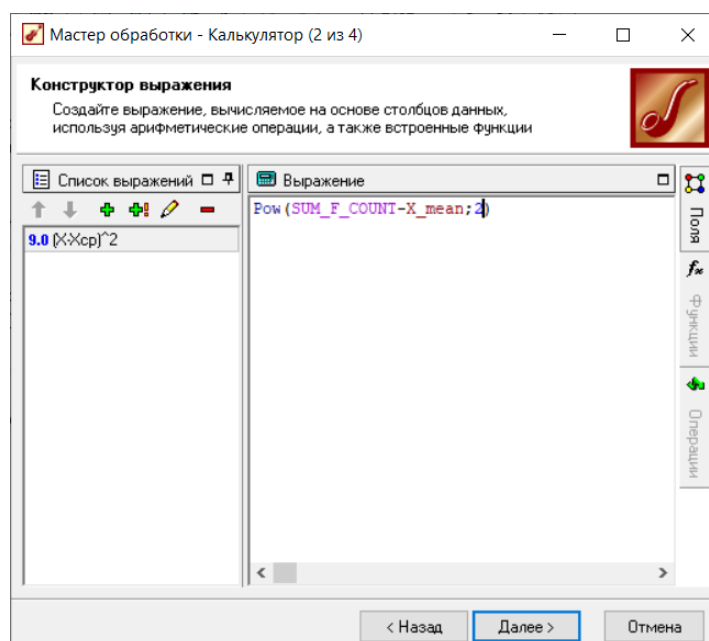


Рисунок 53 – Создание параметров выражения

Результат отражен на рисунке 54.

Код товара	Наименование товара	Количество	X-среднее+	[X:Хср]^2
35	Адреналина гидрохлорида ра	2	0,67	1,7689
477	Аспаркам табл. уп.контурн.б/	46	17,33	821,9689
485	Аспаркам табл. уп.контурн.б/	1	0,33	0,4489
487	Аспаркам табл. уп.контурн.яч	1	0,33	0,4489
682	АЦЦ 100 табл.шип. 100 мг тус	7	6	1
686	АЦЦ 200 табл.шип. 200 мг тус	11	9	4
690	АЦЦ лонг табл.шип. 600 мг тус	15	11,67	11,0889
739	Беродуал р-р д/ингал. фл.каг	1	0,33	0,4489
741	Беродуал Н аэроз.ингал.доз.	1	1,67	0,4489
808	Бифидумбактерин пор. для п/	4	2,33	2,7889
814	Бифидумбактерин пор. для п/	4	5,33	1,7689
828	Бифидумбактерин сухой пор.	39	36	9
844	Бифидумбактерин сухой в св	2	1,33	0,4489
849	Бифидумбактерин форте пор.	1	3	4
1137	Витрум Кальций + Витамин	2	1,33	0,4489
1138	Витрум Кальций + Витамин	1	1	0
1145	Виферон супп.рект. 500000 М	7	7,33	0,1089
1148	Виферон супп.рект. 1000000 I	2	2	0

Рисунок 54 – Результат вычисления квадрата отклонения числа проданных единиц товара от среднего значения

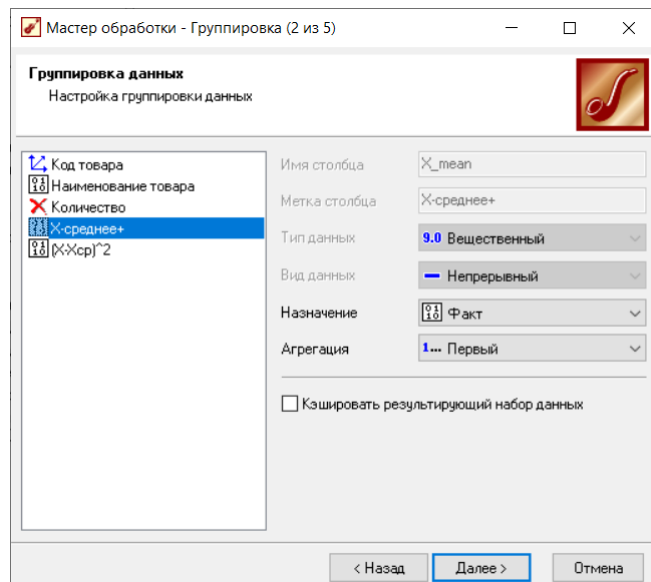


Рисунок 55 – Настройка группировки данных

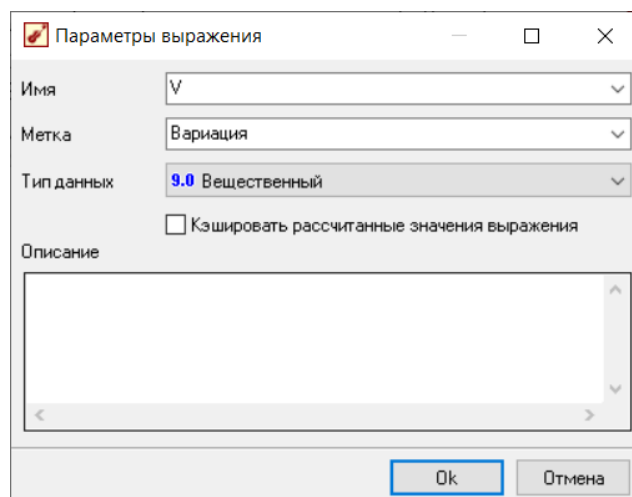


Рисунок 56 – Вычисление коэффициента вариации (шаг 1)

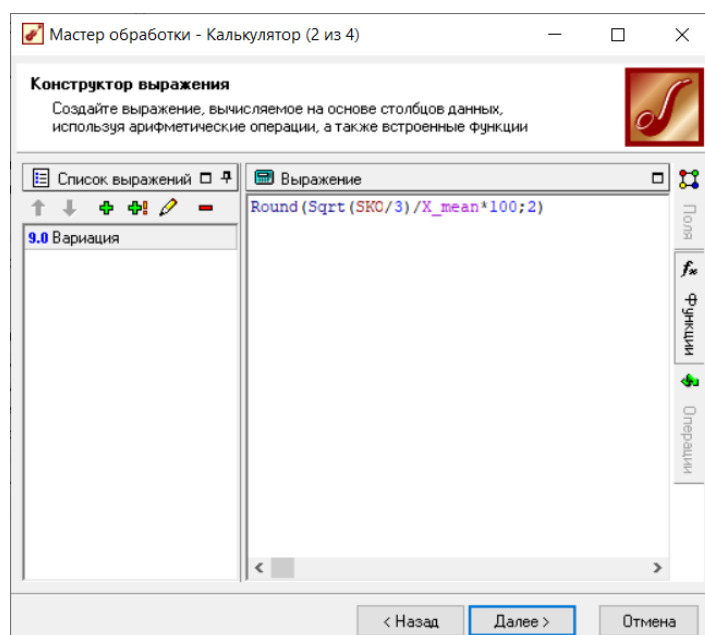


Рисунок 57 – Вычисление коэффициента вариации (шаг 2)

Результат представлен таблицей на рисунке 58.

Код товара	Наименование товара	X-среднее+	(X-Xcp) ²	Вариация
35	Адреналина гидрохлорида ра	0,67	1,7689	114,61
354	Андрокур табл. 50 мг фл. 20 к	2	16	115,47
477	Аспаркам табл. уп. контурн. бл	17,33	1232,6667	116,97
485	Аспаркам табл. уп. контурн. бл	0,33	0,4489	117,22
487	Аспаркам табл. уп. контурн. яч	0,33	0,4489	117,22
504	Аспирин табл. 500 мг бл. 10 к	1	4	115,47
506	Аспирин табл. 100 мг бл. 10 к	0,67	0,2178	40,22
598	Ацетилсалициловая кислота	79,33	3869,4178	45,27
682	АЦЦ 100 табл.шип. 100 мг туб	6	26	49,07
686	АЦЦ 200 табл.шип. 200 мг туб	9	14	24
690	АЦЦ лонг табл.шип. 600 мг туб	11,67	66,6667	40,39
739	Беродуал р-р д/ингал. фл. кап	0,33	0,4489	117,22
741	Беродуал Н аэроз. ингал.доз.	1,67	2,6667	56,46
808	Бифидумбактерин пор. для пр	2,33	4,6667	53,53
814	Бифидумбактерин пор. для пр	5,33	10,6667	35,38
828	Бифидумбактерин сухой пор.	36	18	6,8
844	Бифидумбактерин сухой в св	1,33	0,6667	35,44
848	Бифидумбактерин форте пор.	0,67	0,2178	40,22
849	Бифидумбактерин форте пор.	3	8	54,43
856	Бромокриптин Рихтер табл. 2	0,67	1,7689	114,61

Рисунок 58 – Результат суммы квадратов отклонений числа проданных единиц товара от средней величины.

б) получим разбиение товаров на группы XYZ. Для этого сначала необходимо выполнить сортировку полученной на предыдущем шаге таблицы по возрастанию коэффициента вариации, а затем разбить на основании результатов этой сортировки товары по трем группам.

Сортировка по возрастанию коэффициента вариации (рисунок 59).

Код товара	Наименование товара	X-среднее+	(X-Xcp) ²	Вариация
56778	Полудан пор.лиоф.д/ин. 200 н	0	0	0
1138	Витрум Кальций + Витамин	1	0	0
1148	Виферон супп.рект. 1000000 I	2	0	0
27487	Доктор Тайсс ревма крем кр	1	0	0
29505	Зиртек капли для приема вн	1	0	0
35334	Климен драже бл. 21 кор. 1 S	1	0	0
39850	Лив.52 табл. фл. 100 кор. 1 H	5	0	0
40531	Липоевая кислота табл.п.о. 0	1	0	0
41527	Магне В_6 табл.п.о. бл. 10 ко	7	0	0
64234	Смесь для ингаляций жидк. ф	1	0	0
64915	Спазмостенал р-р для прие	3	0	0
67419	Танакан табл.п.о. 40 мг уп.ко	4	0	0
72057	ФибС р-р д/ин.п/к амп. 1 мл	1	0	0
77139	Чабреца трава трава фильтр	1	0	0
79124	Зв.калитовое масло масло ф	2	0	0
85397	Мята перечной масло масло	1	0	0
61224	Ромашки цветки цветки пак.с	60,33	2,6667	1,56
61230	Ромашки цветки цветки филе	86	6	1,64
71254	Фалиминт драже 25 мг бл. 20	26,67	0,6667	1,77
17303	Бронхikum сироп от кашля си	19,33	0,6667	2,44
44291	Метрогил Дента гель д/десн	34,67	2,6667	2,72

Рисунок 59 – Сортировку полученной на предыдущем шаге таблицы по возрастанию коэффициента вариации

Для выполнения разбиения используем обработчик «Калькулятор» (рисунок 60, 61).

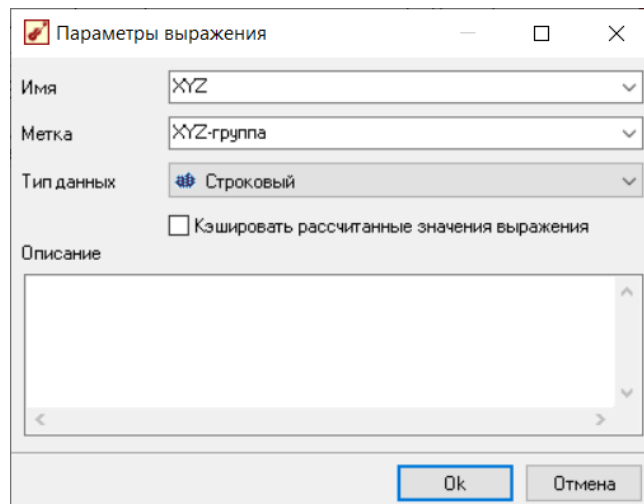


Рисунок 60 – Обработчик «Калькулятор»

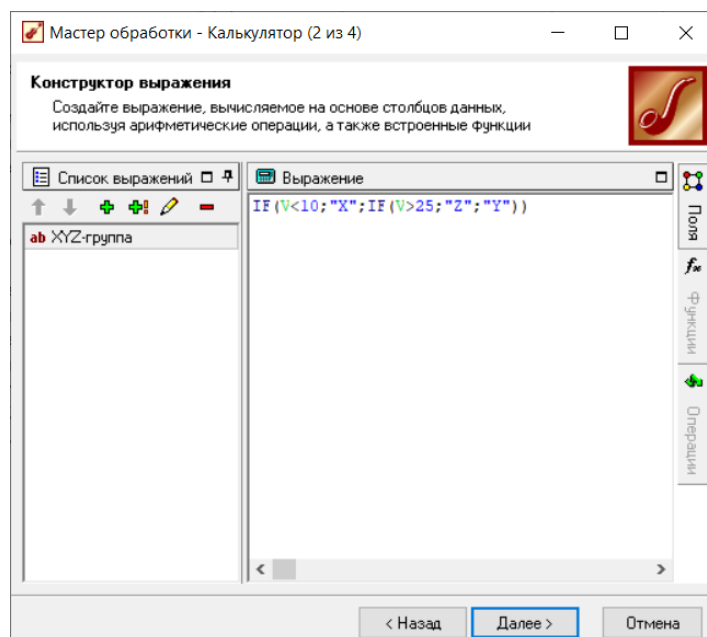


Рисунок 61 – Разбиение на основании результатов сортировки товаров по трем группам

Полученный результат приведен на рисунках 62, 63.

Сохраните результаты работы в файле L2.ded.

4 Проведение ABC-XYZ-анализ

4.1 Суть метода

ABC-XYZ-анализ позволяет разбить данные по продажам на 9 групп в зависимости от вклада в выручку компании (ABC) и регулярности покупок (XYZ).

Такая классификация упрощает работу при планировании и формировании ассортимента на различных уровнях гибких логистических систем, в производственных системах, системах снабжения и сбыта. Внедрение данного метода способствует сокращению количества упущенных продаж, уменьшению излишек товаров, минимизации суммарных затрат, связанных с запасами.

Сценарии X | Отчеты X | Таблица X | Куб X

1 / 1122

Код товара	Наименование товара	X-среднее+	(X:Хср)^2	Вариация	XYZ-группа
56778	Полудан пор.лиофд/ин. 200 г	0	0		
1138	Витрум Кальций + Витамин	1	0	0 X	
1148	Виферон супп.рект. 1000000 I	2	0	0 X	
27487	Доктор Тайсс ревма крем кр	1	0	0 X	
29505	Зиртек капли для приема вн	1	0	0 X	
35334	Климен драже бл. 21 кор. 1 S	1	0	0 X	
39850	Лив.52 табл. фл. 100 кор. 1 H	5	0	0 X	
40531	Липоевая кислота табл.п.о. 0	1	0	0 X	
41527	Магне В_6 табл.п.о. бл. 10 ко	7	0	0 X	
64234	Смесь для ингаляций жидк. ф	1	0	0 X	
64915	Спазмостенал р-р для прие	3	0	0 X	
67419	Танакан табл.п.о. 40 мг уп.ко	4	0	0 X	
72057	ФиБС р-р д/ин.п/к. амп. 1 мл	1	0	0 X	
77139	Чабреца трава трава фильтр	1	0	0 X	
79124	Эвкалиптовое масло масло ф	2	0	0 X	
85397	Мяты перечной масло масло	1	0	0 X	
61224	Ромашки цветки цветки пак. с	60,33	2,6667	1,56 X	
61230	Ромашки цветки цветки филь	86	6	1,64 X	
71254	Фалиминт драже 25 мг бл. 20	26,67	0,6667	1,77 X	
17303	Бронхикум сироп от кашля си	19,33	0,6667	2,44 X	
44291	Метрогил Дента гель д/десе	34,67	2,6667	2,72 X	
28092	Дуовит драже уп.контурн.яч.	10,67	0,6667	4,42 X	

Рисунок 62 – Таблица

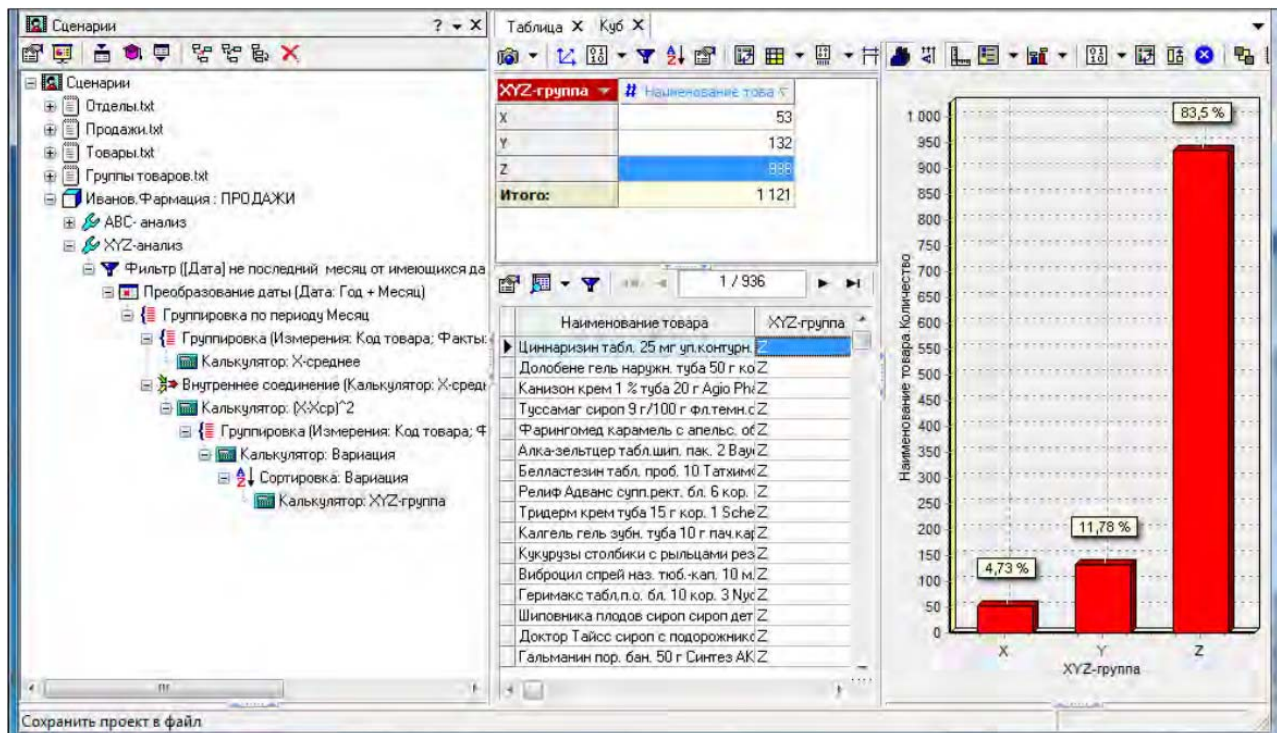


Рисунок 63 – Куб

Группа	X	Y	Z
A	Товары групп AX и VX отличаются высоким товарооборотом и стабильностью. Необходимо обеспечить их постоянное наличие, но для этого не нужно создавать избыточный страховой запас. Расход товаров этих групп стабилен и хорошо прогнозируется.	Товары групп AY и BY при высоком товарообороте имеют недостаточную стабильность расхода, и, как следствие, чтобы обеспечить постоянное наличие, нужно увеличить страховой запас.	Товары групп AZ и BZ при высоком товарообороте отличаются низкой прогнозируемостью расхода. Попытка обеспечить гарантированное наличие всех товаров данных групп только за счет избыточного страхового товарного запаса приведет к тому, что средний товарный запас компании значительно увеличится. По этим группам следует пересмотреть систему заказов.
B			
C	Для товаров группы CX можно использовать систему заказов с постоянной периодичностью и снизить страховой товарный запас.	По товарам группы CY можно использовать систему заказов с постоянной суммой (объемом) заказа, но при этом формировать страховой запас, исходя из имеющихся у компании финансовых возможностей.	В группу товаров CZ попадают все новые товары, товары непостоянного спроса, поставляемые под заказ и т. п. Часть из них можно безболезненно выводить из ассортимента, а другую часть нужно регулярно контролировать, так как именно из товаров этой группы возникают неликвидные или трудно реализуемые запасы, из-за которых компания несет потери.

Товары групп А и В составляют основной товарооборот компании. Поэтому необходимо обеспечивать постоянное их наличие. Общепринятой является практика, когда по продукции группы А создается избыточный страховой запас, а по товарам группы В – достаточный. Использование XYZ-анализа позволяет разработать более точную ассортиментную политику и за счет этого снизить суммарный товарный запас.

Алгоритм ABC-XYZ-анализа:

- 1 провести ABC-анализ;
- 2 провести XYZ-анализ;
- 3 совместить полученные результаты;
- 4 построить совмещенную матрицу.

Сценарий ABC- и XYZ-анализа (рисунок 64).

Результирующая таблица по ABC- и XYZ-анализу представлена на рисунке 65.

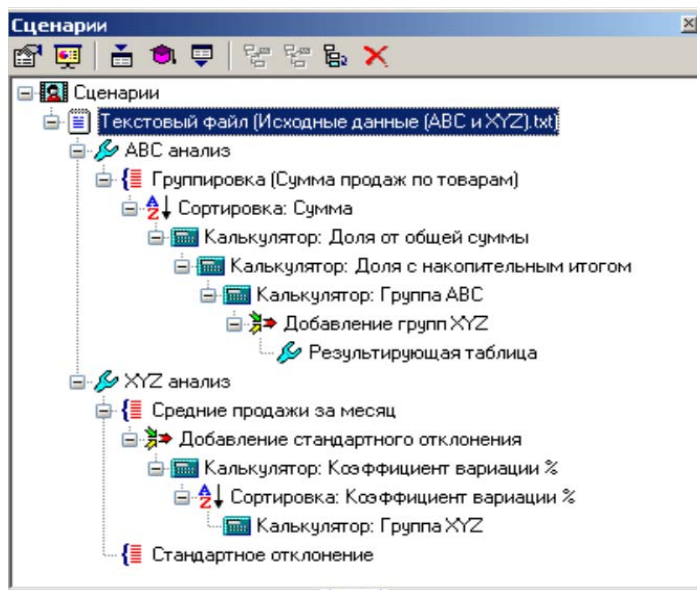


Рисунок 64 – Сценарий ABC- и XYZ-анализа

Товар	Группа ABC	Группа XYZ
Товар 9	A	X
Товар 4	B	X
Товар 7	B	Y
Товар 3	C	X
Товар 2	C	X
Товар 5	C	Y
Товар 1	C	Z
Товар 6	C	Z
Товар 8	C	Z
Товар 10	C	Z

Рисунок 65 – Результирующая таблица по ABC- и XYZ-анализу

4.2 Выполнение ABC-XYZ -анализа

1) для получения результирующей таблицы ABC-XYZ-анализа необходимо выполнить слияние узла «Калькулятор ABC-группа» и узла «Калькулятор XYZ-группа» (рисунок 66).

Для узла «Калькулятор ABC-группа» используем слияние типа «Внешнее левое соединение». Это позволит присоединить к набору данных группы ABC все строки из набора XYZ, которые совпадают с исходными по ключевым полям.

В качестве ключевого поля выбираем поле «Код товара», а в выходную таблицу передаем поля «Код товара», «Наименование товара», «Группа ABC» и «Группа XYZ» (рисунки 67, 68).

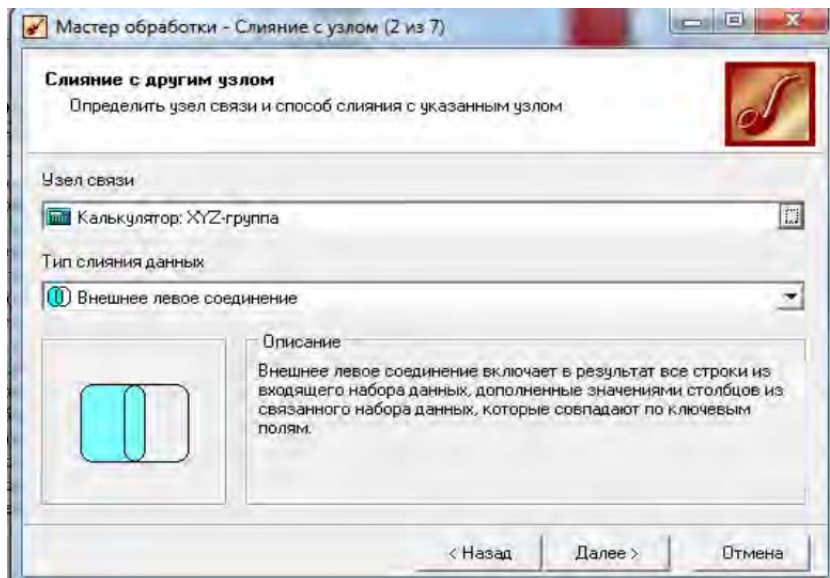


Рисунок 66 – Слияние узла «Калькулятор ABC-группа» и узла «Калькулятор XYZ-группа»

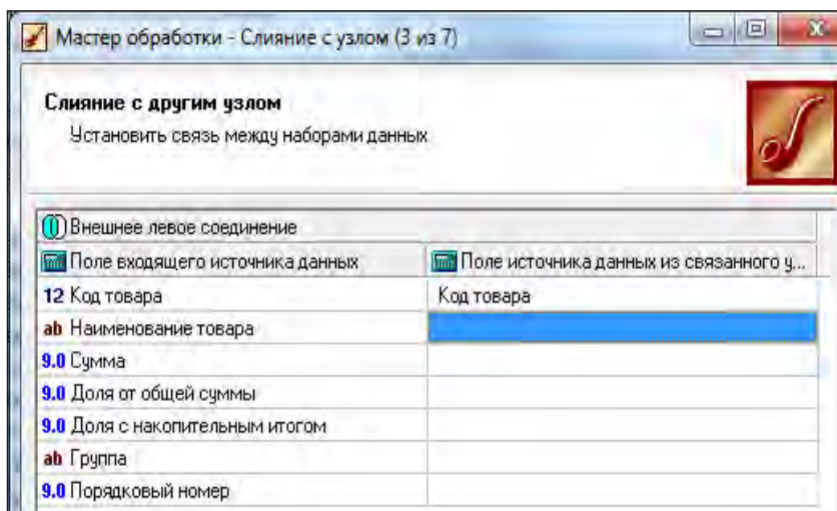


Рисунок 67 – Установление связи между наборами данных

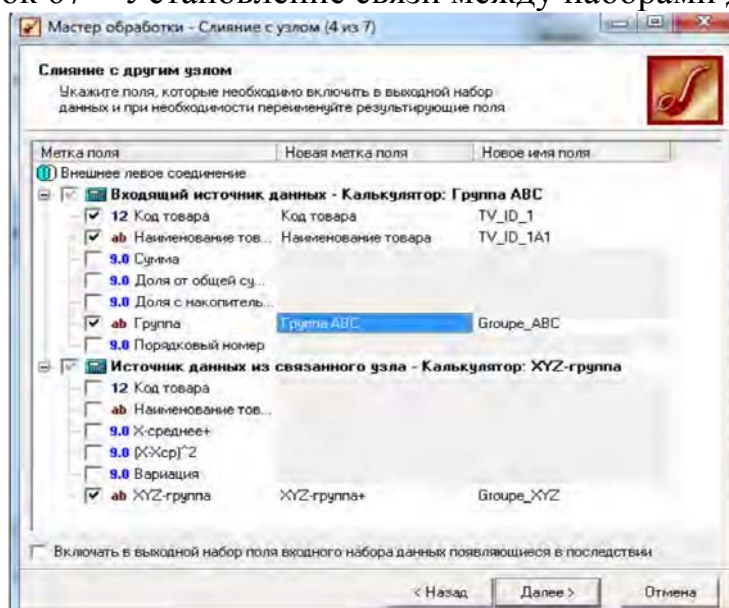


Рисунок 68 – Выбор полей, необходимых для выходной таблицы

Выбираем визуализацию «Куб». Группы ABC и XYZ будут измерениями, остальные поля – информационные (рисунок 69).

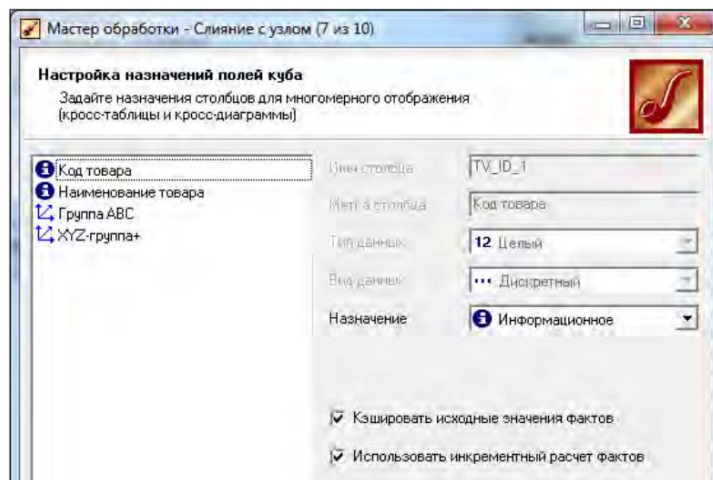


Рисунок 69 – Настройка назначений «Куба»

Распределяем измерения по строкам и колонкам и выбираем в качестве фактов число записей, соответствующих каждой группе (рисунок 70).

Факты и варианты агрегации, а также варианты отображения укажите в соответствии с рисунком 71.

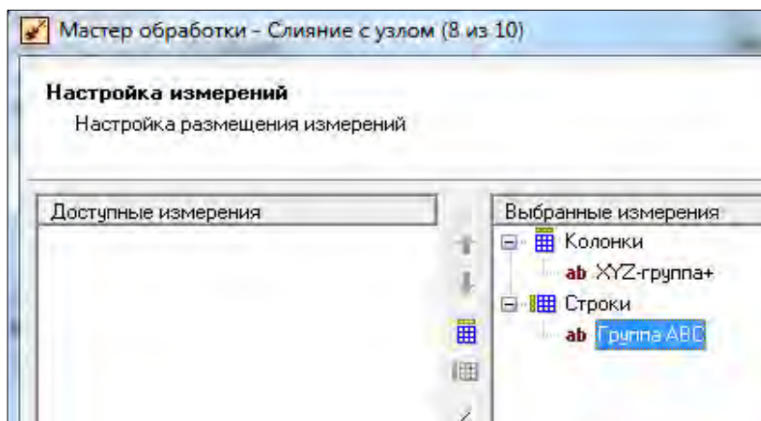


Рисунок 70 – Настройка измерений по строкам и колонкам

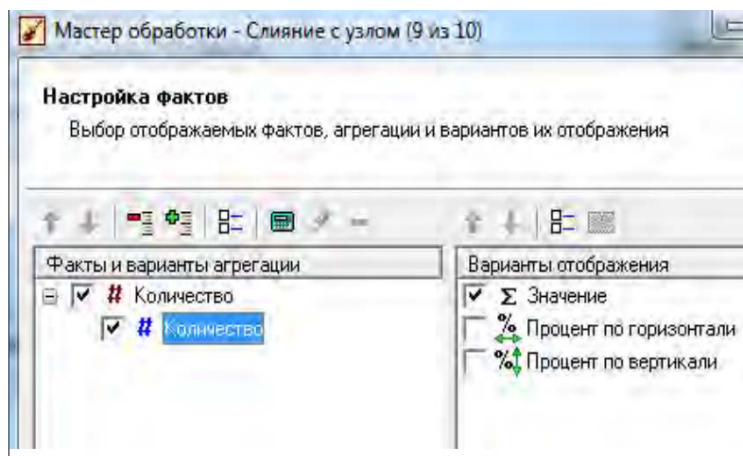


Рисунок 71 – Выбор отображаемых факт

Результат приведен на рисунках 72, 73.

Код товара	Наименование товара	Группа ABC	XYZ-группа+
	Адреналин гидрохлорида ра С		Z
354	Андрокур табл. 50 мг фл. 20 н А		Z
477	Аспаркам табл. уп.контурн.б/ С		Z
485	Аспаркам табл. уп.контурн.б/ С		Z
487	Аспаркам табл. уп.контурн.я С		Z
504	Аспирин табл. 500 мг бл. 10 н С		Z
505	Аспирин табл. 100 мг бл. 10 н С		Z
598	Ацетилсалициловая кислота С		Z
682	АЦЦ 100 табл.шип. 100 мг туг В		Z
686	АЦЦ 200 табл.шип. 200 мг туг А		Y
690	АЦЦ лонг табл.шип. 600 мг туг А		Z
739	Беродуал р-р д/ингал. фл.квн С		Z
741	Беродуал Н аэроз.ингал.доз. А		Z
808	Бифидумбактерин пор. для пв В		Z
814	Бифидумбактерин пор. для пв В		Z
828	Бифидумбактерин сухой пор. А		X
844	Бифидумбактерин сухой в свк С		Z
848	Бифидумбактерин форте пор. С		Z
849	Бифидумбактерин форте пор. С		Z
856	Бромокриптин Рихтер табл. 2 С		Z
1137	Витрум Кальцин + Витамин В		Z

Рисунок 72 – Таблица

Группа ABC	X	Y	Z	Итого:
A	32	58	102	192
B	10	43	142	195
C	11	31	692	734
Итого:	53	132	936	1121

Код товара	Наименование товара	Группа ABC	XYZ-группа+
68036	Терпинкод табл. уп.контурн.я	A	X
82545	Арбидол-ЛЭНС табл.п.о. 0,1 г	A	X
3382	Ношпа табл. 40 мг конг. 100	A	X
61230	Ромашки цветки цветки филе	A	X
1145	Виферон супп. рект. 5000000 М	A	X
83568	Алфавит табл./комплект бл.	A	X
62535	Санорин эмульс. наз. 0,1 % фл.	A	X
83215	Центрум табл.п.о. фл. 60 кон.	A	X
17303	Бронхикум сироп от кашля си	A	X
13882	Белосалик мазь туба 30 г пак	A	X
82203	Вентолин аэроз.ингал.доз. 10	A	X
41527	Магне В_6 табл.п.о. бл. 10 ко	A	X
71254	Фальсификат драже 25 мг бл. 2	A	X
828	Бифидумбактерин сухой пор.	A	X
22899	Глицин табл.сублингв. 0,1 г уп	A	X
67419	Танакан табл.п.о. 40 мг уп.ко	A	X
61224	Ромашки цветки цветки пак.т	A	X

Рисунок 73 – Куб

Сохраните результаты работы в файле L2.ded.

Проанализируйте полученные результаты и укажите в отчете по лабораторной работе выводы ABC-XYZ анализа.

5 Регрессионные модели для предсказания поведения покупателей

Теоретические сведения

Модель множественной линейной регрессии (или коротко – множественная линейная регрессия) предназначена для проверки и изучения связи (объяснения поведения) между одной зависимой переменной (эндогенной) и несколькими независимыми (экзогенными) переменными. Предполагается, что такая связь теоретически может быть описана (специфицирована) линейной зависимостью (функцией) вида:

$$Y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + U$$

где Y – зависимая (объясняемая, эндогенная) переменная – регрессанд, U – случайная составляющая модели, x_j – независимые (объясняющие,

экзогенные) переменные - регрессоры.

Мы будем рассматривать классическую модель множественной линейной регрессии. Это означает, что независимые переменные (регрессоры) предполагаются не случайными (детерминированными) величинами.

Коэффициент детерминации

Коэффициент детерминации – это доля объясненной дисперсии в общей, в случае линейной регрессии с константой определяется по формуле:

$$R^2 = \frac{ESS}{TSS}$$

где ESS – объясненная сумма квадратов отклонений;

TSS – общая дисперсия.

Коэффициент детерминации для модели с константой принимает значения от 0 до 1. Чем ближе значение коэффициента к 1, тем сильнее зависимость. При оценке регрессионных моделей это интерпретируется как соответствие модели данным. Для приемлемых моделей предполагается, что коэффициент детерминации должен быть хотя бы не меньше 50 % (в этом случае коэффициент множественной корреляции превышает по модулю 70 %). Модели с коэффициентом детерминации выше 80 % можно признать достаточно хорошими (коэффициент корреляции превышает 90 %). Равенство коэффициента детерминации единице означает, что объясняемая переменная в точности описывается рассматриваемой моделью.

R^2 – показывает насколько хорошо подогнана модель под данные.

Статистика Фишера

Статистика Фишера используется для проверки гипотезы о связи между объясняемым рядом и регрессорами. Используется нулевая гипотеза: коэффициенты при всех регрессорах равны нулю.

Статистическая значимость уравнения множественной регрессии в целом оценивается с помощью общего F-критерия Фишера:

$$F = \frac{R_{y x_1 x_2 \dots x_p}^2}{1 - R_{y x_1 x_2 \dots x_p}^2} \frac{n - m - 1}{m}$$

где m – число факторов в линейном уравнении регрессии;

n – число наблюдений, R^2 - коэффициент детерминации.

Формула для расчёта статистики Фишера для модели с константой:

$$F = \frac{R^2 / (k - 1)}{(1 - R^2) / (N - k)}$$

где: R^2 – коэффициент детерминации;

k – количество факторов, включенных в модель;

N – количество наблюдений.

Формула для расчёта статистики Фишера для модели без константы:

$$F = \frac{R^2 / k}{(1 - R^2) / (N - k)}$$

где: R^2 – коэффициент детерминации;

k – количество факторов, включенных в модель (включая константу);

N – количество наблюдений.

Для нецентрированного коэффициента детерминации может быть рассчитана соответствующая статистика Фишера.

Вероятность статистики Фишера.

Статистика Фишера имеет распределение Фишера:

- для модели с константой: $F(k - 1, N - k)$;
- для модели без константы: $F(k, N - k)$.

Нулевая гипотеза о равенстве нулю коэффициентов при всех регрессорах отклоняется, если вероятность меньше, чем уровень значимости. Рассматривают один из стандартных уровней значимости 0.1, 0.05 или 0.01.

t-статистика

Оценка значимости коэффициентов регрессии (кроме свободного члена) осуществляется сравнением t-статистики

$$t_j = \frac{a_j}{SE\sqrt{b_{jj}}}$$

с табличным значением t-статистики Стьюдента. В b_{jj} – диагональный элемент матрицы $(X^T * X)^{-1}$. SE – среднеквадратическое отклонение ошибки. Если значение превосходит табличное значение t-статистики Стьюдента, то j-й коэффициент считается значимым, в противном случае фактор, соответствующий данному коэффициенту следует исключить из модели.

Задание. Используя Excel или Calc постройте модель множественной регрессии, оцените её качество, осуществите прогнозирование возможного пополнения в семьях покупателей,

Исходные данные находятся в файле «RetailMart.xml» на листе «Обучающие данные».

В файле представлены данные покупателях магазина RetailMart и сведения о их покупках в течение года, например: пол владельца учетной записи (покупателя): мужской, женский, не указан; адрес (частный дом, квартира или абонентский ящик; недавно заказывал тест на беременность; недавно заказывал вино; недавно заказывал одежду для беременных и кормящих и другие.

В столбце S содержатся сведения о беременности покупателя (или члена семьи). Всего имеется выборка по 500 беременным и 500 небеременным. (рисунок 74).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	U	не указан	абонентский ящик	беременность	средств гигиены	женской одежды	число беремен	беремен	куд для тела	йог для подуш	имбирь от	кота ранее	заказывал	недавно для	бросаяц	регулярно	вал	беремен	Беременнос
2	M	A	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
3	M	н	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
4	M	н	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
5	U	н	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
6	F	A	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1
7	F	н	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1
8	M	н	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1
9	F	н	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
10	F	н	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1
11	F	н	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1
12	F	н	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
13	F	A	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1
14	F	н	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	1
15	U	A	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1
16	M	A	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
17	M	н	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1
18	M	P	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1

Рисунок 74 – Данные о покупателях

Необходимо создать регрессионную модель для предсказания факта беременности покупателей на основании имеющихся данных о покупках и проверить модель на тестовых данных (лист «Тестовые данные»).

Для выполнения работы выполните следующие этапы.

1 На листе «Обучающие данные с фиктивными переменными» переведите категориальные данные в числовые. Для этого введите «фиктивные переменные». Информацию о поле представьте в виде столбцов «М», «Ж» (вариант «пол не указан» будет выражаться в значениях 0 в обоих столбцах).

При использовании фиктивной переменной для кодирования категориальных данных нужно на один столбец меньше, чем количество имеющихся категорий – последняя категория может быть вычислена через остальные. Категориальная переменная «пол» имеет всего две степени свободы, т.к. степеней свободы на одну меньше, чем возможных значений переменных.

Информацию об адресе представьте в виде двух столбцов «Частный дом» и «Квартира» (вариант «а/я» будет выражаться через 0 в этих столбцах). Используйте функцию ЕСЛИ (рисунок 75).

	A	B	C	D	E	F	G
1	М	Ж	Частный дом	Квартира	Пол (М- мужской, F- женский, U - не указан)	Н-частный дом/А- квартира/Р- абонентски й ящик	тест на беременность
2	1	0	0	1	1 М	А	1
3	1	0	1	0	0 М	Н	1
4	1	0	1	0	0 М	Н	1
5	0	0	1	0	0 U	Н	0
6	0	1	0	1	1 F	А	0
7	0	1	1	0	0 F	Н	0
8	1	0	1	0	0 М	Н	0
9	0	1	1	0	0 F	Н	0
10	0	1	1	0	0 F	Н	0
11	0	1	1	0	0 F	Н	0
12	0	1	1	0	0 F	Н	0
13	0	1	0	1	1 F	А	0

Рисунок 75 – Данные о покупателях после преобразования категориальных данных в числовые

2 Создайте линейную регрессионную модель.

2.1 На листе «Модель линейной регрессии» начиная со столбца В и строки 8 вставьте данные с листа «Обучающие данные с фиктивными переменными». Вставьте строку с названиями в строку 1. Столбец U назовите «Свободный член» – в нем будет отражаться свободный член линейной регрессии, заполните его с 8 по 1007 строку единицами. Строку 2 озаглавьте «Коэффициенты модели» и запишите стартовые значения равные 1 в каждом столбце с данными (с В по U).

2.2 Проведите обучение модели

2.2.1 В столбец W добавьте (ячейка W7) добавьте название «Предсказание», а ниже для каждого покупателя поместите линейную комбинацию коэффициентов и данных покупателей (свободный член включен) (используйте функцию СУММПРОИЗВ коэффициентов на значения переменных). Отформатируйте столбец до двух знаков после запятой. Проанализируйте полученные расчетные данные.

2.2.2 Осуществите расчет отклонений в столбце X (название в ячейке X7 – «Квадрат отклонения»). Значения в столбце X – это квадрат расстояния от значения «Беременность» в столбце V до прогнозного значения в столбце W.

2.2.3 В ячейке W1 впишите название «Сумма квадратов отклонений», а в ячейке X1 найдите сумму квадратов отклонений.

2.2.4 Осуществите настройку модели оптимизации в «Поиске решений»:

- необходимо минимизировать целевую функцию (сумма квадратов отклонений);

- изменять нужно значения переменных (коэффициентов модели) в ячейках от B2 до U2;

- установите ограничения на значения коэффициентов модели « $> = -1$ » И « $< = 1$ »;

- используйте эволюционный алгоритм, так как целевая функция не линейна.

2.2.5 Представьте результаты «Поиска решений» в отчете – значения полученных коэффициентов, свободного члена и целевой функции. Сделайте вывод.

2.3 Проверьте полученную регрессионную модель на адекватность.

2.3.1 Рассчитайте коэффициент детерминации (R^2). Для этого в ячейке W2 введите название «Общая сумма квадратов», а в ячейке X2 рассчитайте дисперсию значений зависимой переменной «Беременность» (функция КВАДРОТКЛ). В ячейке W3 введите название «Сумма квадратов с выясненным происхождением» и в ячейке X3 рассчитайте объясненную дисперсию ($=X2-X1$). В ячейке W4 введите название « R^2 » и в ячейке X3 рассчитайте коэффициент детерминации (R^2) ($=X3/X2$). В отчете приведите результаты расчета и сделайте вывод.

2.3.2 Проверьте статистическую значимость уравнения множественной регрессии в целом с помощью общего F-критерия Фишера. В ячейках Y1:Z5 введите названия показателей и рассчитайте значения, представленные в таблице 1. Приведите результаты расчета, сделайте вывод.

2.3.3 Проверьте значимость отдельных переменных, проведите t-тест.

2.3.3.1 Рассчитайте среднеквадратическое отклонение прогноза (мера вариабельности отклонений прогноза модели), как квадратный корень из суммы квадрата отклонений (X1), разделенный на количество степеней свободы (Z3).

Таблица 1 – Данные для проверки

Столбец Y	Значение	Столбец Z (функции Excel)
Количество наблюдений, N	Общее количество значений в выборке	СЧЕТ
Количество коэффициентов модели, k	19 зависимых переменных и свободный член модели	СЧЕТ
Степень свободы	Количество вариантов данных за вычетом количества коэффициентов модели	Формула N-k
Значение F-статистики	Отношение «понятной» части суммы квадратов к «непонятной», умноженной на отношение количества степеней свободы к количеству зависимых переменных	Формула 2.1
Вероятность F-статистики	Вероятность выбора нулевой гипотезы	ФРАСП

2.3.3.2 Рассчитайте стандартное отклонение коэффициентов. Сложным процесс вычисления стандартного отклонения делает необходимость понимать, как обучающие данные для коэффициентов варьируются и сами по себе, и взаимодействуя с другими переменными. Первый шаг – превращение обучающей последовательности в одну большую матрицу (матрица плана) путем умножения ее на саму себя. Это произведение матрицы плана и ее же самой образует матрицу суммы квадратов и векторных произведений.

Для этого:

2.3.3.2.1 Создайте лист «Стандартное отклонение коэффициентов модели». Для формирования матрицы суммы квадратов и векторных произведений (СКВП) умножьте транспонированную матрицу плана на саму себя. Для этого вставьте строку заголовков обучающих данных на листе «Стандартное отклонение коэффициентов модели» в ячейки B1:U1 и транспонированную строку заголовков в A2:A21 (вместе с заголовком «Свободный член»). Выделите B2:U21 и далее с помощью функции МУМНОЖ и ТРАНС, умножьте транспонированную матрицу 'Модель линейной регрессии'!B8:U1007 на её саму же. Не забудьте использовать формулу для массива. Пример результата расчетов представлен на рисунке 76.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U								
1	СКВП	М	Ж	частичный дом	Квартира	беремен	заказ противоз	заказ средс	заказ препарат	заказ витамин	заказ йог для	заказ подушк	заказыва л	заказыва медово	заказыва медиста	заказыва медово	заказыва медово	заказыва и	заказыва медово	заказыва медово	заказыва свободн								
2	М		401	0	196	169	27	62	67	42	45	8	8	29	14	36	45	23	46	51	50	401							
3	Ж			0	495	239	207	37	63	61	54	71	9	8	31	14	44	47	32	69	59	71	495						
4	частичный дом					0	43	57	74	54	59	10	14	44	11	45	46	29	63	62	65	488							
5	Квартира						0	420	26	59	58	39	57	8	3	19	16	38	42	20	56	54	420						
6	беремен								27	37	43	26	75	6	5	13	19	3	2	8	5	9	5	18	17	75			
7	противоз									62	63	57	59	6	140	24	5	13	0	1	5	1	3	20	5	10	22	7	140
8	средс									67	61	74	58	5	24	141	7	14	4	4	6	3	5	19	3	12	25	17	141
9	препарат									42	54	54	39	13	5	7	106	22	3	1	11	5	14	4	12	25	4	23	106
10	витамин									45	71	59	57	19	13	14	22	128	2	4	9	10	22	9	8	24	9	22	128
11	йог для									8	9	10	8	3	0	4	3	2	18	1	2	1	0	0	1	3	1	5	18
12	подушк									8	8	14	3	2	1	4	1	4	1	18	0	0	2	0	1	5	1	4	18
13	заказыва									29	31	44	19	8	5	6	11	9	2	0	69	1	6	7	8	8	5	17	69
14	заказыва медово									14	14	11	16	5	1	3	5	10	1	0	1	30	3	3	3	3	1	5	30
15	заказыва медиста									36	44	45	38	9	3	5	14	22	0	2	6	3	92	0	10	20	6	19	92
16	заказыва медово									45	47	46	42	5	20	19	4	9	0	0	7	3	0	97	5	7	19	11	97
17	заказыва и									23	32	29	20	18	5	3	12	8	1	1	8	3	10	5	60	13	2	18	60
18	заказыва медово									46	69	63	56	17	10	12	25	24	3	5	8	3	20	7	13	130	0	22	130
19	заказыва медиста									51	59	62	51	2	22	25	4	9	1	1	5	1	6	19	2	0	123	10	123
20	заказыва свободн									50	71	65	54	17	7	17	23	22	5	4	17	5	19	11	18	22	10	131	131
21	ый член									401	495	488	420	75	140	141	106	128	18	18	69	30	92	97	60	130	123	131	1000

Рисунок 76 – Матрица СКВП

Проанализируйте значения в матрице СКВП. По диагонали считаются совпадения переменной самой с собой (то есть значения равны сумме по каждому столбцу матрицы плана), свободный член равен 1000. В ячейках, не входящих в диагональ, проанализируйте число совпадений признаков. Например, интересно, что в шести случаях совершались покупки противозачаточных средств и теста на беременность одновременно. То есть матрица СКВП дает нам представление о величине переменных и о том, насколько они пересекаются и соотносятся между собой.

2.3.3.2.2 Преобразуйте матрицу СКВП в обратную. Для этого вставьте заголовки ниже матрицы в ячейки B24:U24 и A25:A44. Выделите диапазон B24:U44 и с помощью функции МОБР как формулы массива получите обратную матрицу из матрицы СКВП.

2.3.3.2.3 Рассчитайте среднеквадратическое отклонение коэффициента как произведение среднеквадратического значения прогноза модели (ячейка X5 на листе «Модель линейной регрессии») на квадратный корень из соответствующего значения диагонального элемента обратной матрицы СКВП.

Для облегчения расчетов пронумеруйте переменные, начиная с 1 в B46 до 20 в U46. Затем используйте формулу ИНДЕКС, чтобы найти соответствующее значение диагонального элемента. Например,

ИНДЕКС(Стандартное отклонение коэффициентов модели!B25:B44; Стандартное отклонение коэффициентов модели!B46) выдает значение пересечения строки «М» со столбцом «М». Пример представлен на рисунке 77.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
25 M		0.0122	0.0097	0.0003	0.0003	0.0006	0.0000	-0.0003	-0.0002	0.0000	-0.0006	-0.0001	0.0002	-0.0005	0.0002	-0.0007	-0.0003	0.0002	0.0001	-0.0003	-0.0099
26 Жесткость		0.0097	0.0118	0.0004	0.0004	0.0006	0.0002	0.0000	-0.0001	-0.0003	-0.0005	-0.0002	0.0005	-0.0002	0.0004	-0.0005	-0.0004	0.0000	0.0001	-0.0004	-0.0100
27 дом		0.0003	0.0004	0.0133	0.0112	-0.0005	0.0013	-0.0004	0.0004	0.0001	-0.0009	-0.0008	-0.0004	0.0004	0.0001	0.0000	0.0012	-0.0001	-0.0002	0.0000	-0.0116
28 Квадратичный тест на беременность		0.0003	0.0004	0.0112	0.0136	-0.0002	0.0011	-0.0003	0.0006	-0.0001	-0.0010	0.0004	0.0003	-0.0001	0.0001	-0.0001	0.0013	-0.0002	-0.0001	0.0000	-0.0116
29 Беременность		0.0006	0.0006	-0.0005	-0.0002	0.0156	0.0002	0.0004	-0.0003	-0.0011	-0.0013	-0.0002	-0.0003	-0.0011	0.0002	0.0001	-0.0035	-0.0006	0.0006	-0.0005	-0.0010
30 Противозачаточные средства		-0.0003	0.0000	-0.0004	-0.0003	0.0004	-0.0002	0.0085	0.0005	0.0001	-0.0007	0.0004	0.0002	0.0006	-0.0004	0.0005	0.0003	-0.0004	0.0003	-0.0001	-0.0009
31 Среднее значение		-0.0002	-0.0001	0.0004	0.0006	-0.0003	0.0007	0.0005	0.0111	-0.0006	-0.0005	0.0009	-0.0004	-0.0004	-0.0001	0.0006	-0.0006	-0.0008	0.0006	-0.0006	-0.0014
32 Препараты для контрацепции		0.0000	-0.0003	0.0001	-0.0001	0.0001	0.0001	-0.0006	0.0094	0.0004	-0.0009	0.0000	-0.0009	0.0000	-0.0018	-0.0010	0.0001	0.0005	-0.0003	0.0003	-0.0002
33 Витамины		-0.0006	-0.0005	-0.0009	-0.0010	-0.0013	0.0010	-0.0007	-0.0005	0.0004	0.0576	-0.0019	-0.0005	-0.0008	0.0014	0.0012	0.0005	-0.0001	0.0005	-0.0012	0.0003
34 Бюджет		-0.0001	0.0002	-0.0008	0.0004	-0.0002	0.0006	-0.0007	0.0009	-0.0009	-0.0019	0.0578	0.0014	0.0012	0.0001	0.0010	0.0002	-0.0012	0.0005	-0.0008	-0.0008
35 Индекс		0.0002	0.0005	-0.0004	-0.0003	-0.0003	0.0005	0.0004	-0.0004	0.0000	-0.0005	0.0014	0.0160	0.0008	0.0003	-0.0002	-0.0008	0.0003	0.0004	-0.0010	-0.0014
36 Индекс		-0.0005	-0.0002	0.0004	-0.0001	-0.0011	0.0009	0.0002	-0.0004	-0.0018	-0.0008	0.0012	0.0008	0.0352	0.0003	-0.0002	-0.0004	0.0006	0.0007	-0.0001	-0.0009
37 Индекс		0.0002	0.0004	0.0001	0.0001	0.0002	0.0008	0.0006	-0.0001	-0.0010	0.0014	0.0001	0.0003	0.0003	0.0126	0.0011	-0.0007	-0.0006	0.0003	-0.0006	-0.0016
38 Индекс		-0.0007	-0.0005	0.0000	-0.0001	0.0001	-0.0004	-0.0004	0.0006	0.0001	0.0012	0.0010	-0.0002	-0.0002	0.0011	0.0118	-0.0001	0.0003	-0.0006	0.0000	-0.0006
39 Индекс		-0.0003	-0.0004	0.0012	0.0013	-0.0035	0.0002	0.0005	-0.0006	0.0005	0.0005	0.0002	-0.0008	-0.0004	-0.0007	-0.0001	0.0193	-0.0005	0.0005	-0.0013	-0.0015
40 Индекс		0.0002	0.0000	-0.0001	-0.0002	-0.0006	0.0004	0.0003	-0.0008	-0.0003	-0.0011	-0.0012	0.0003	0.0006	-0.0006	0.0003	-0.0005	0.0093	0.0011	-0.0001	-0.0012
41 Индекс		0.0001	0.0001	-0.0002	-0.0001	0.0006	-0.0001	-0.0004	0.0006	0.0003	0.0005	0.0005	0.0004	0.0007	0.0003	-0.0006	0.0005	0.0011	0.0097	0.0002	-0.0015
42 Индекс		-0.0003	-0.0004	0.0000	0.0000	-0.0005	0.0006	-0.0001	-0.0006	-0.0002	-0.0012	-0.0008	-0.0010	-0.0001	-0.0006	0.0000	-0.0013	-0.0001	0.0002	0.0092	-0.0006
43 Индекс		-0.0099	-0.0100	-0.0116	-0.0116	-0.0010	-0.0027	-0.0009	-0.0014	-0.0008	0.0003	-0.0008	-0.0014	-0.0009	-0.0016	-0.0006	-0.0015	-0.0012	-0.0015	-0.0006	0.0221
44 член																					
45																					
46		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
47		0.04114	0.04033	0.04287	0.0433	0.04649	0.03481	0.03431	0.03915	0.03604	0.08926	0.08941	0.04709	0.06982	0.04169	0.04035	0.0516	0.0359	0.03664	0.03572	0.05533

Рисунок 77 – Результаты расчета обратной матрицы СКВП и стандартного отклонения коэффициентов

2.3.3.2.4 На листе «Модель линейной регрессии» поместите в ячейку A3 заголовок «Стандартное отклонение коэффициента» и скопируйте в строку A3 значения рассчитанных отклонений из листа «Стандартное отклонение коэффициентов модели» как значения.

2.3.3.3 Рассчитайте t-статистику для каждого коэффициента по двустороннему критерию. Для этого нужно разделить значение коэффициента (по модулю) на соответствующее стандартное отклонение коэффициента. В

ячейке A4 напишите наименование «t-статистика» и рассчитайте в строке 3 значения t-статистики для каждого коэффициента.

2.3.3.4 Рассчитайте оценку распределения Стьюдента относительно значения t-статистики и количества степеней свободы (Z3). Ячейку A5 назовите «Оценка распределения Стьюдента», и далее в строке с помощью формулы СТЬЮДРАСП рассчитайте вероятность того, что коэффициент будет по меньшей мере таким, как при нулевой гипотезе. В формуле СТЬЮДРАСП используйте параметр «двусторонний критерий». Пример представлен на рисунке 78.

Рисунок 78 – Результаты проверки по t-критерию

2.3.3.5 Представьте перечень статистически незначимых критериев (значение вероятности больше либо равно 0,05). Для обучения в дальнейшем эти критерии можно будет удалить. Сделайте вывод.

3 Тестирование модели.

3.1 На вкладке «Тестовые данные» содержится тестовый набор данных о покупателях, которые не участвовали в обучении модели (контрольная выборка). Всего 1000 покупателей, 6% из которых беременны.

Ячейку V1 назовите «Линейное предсказание» и в ячейках ниже осуществите прогноз по полученной модели для всех покупателей. Представьте результат в отчете. Проанализируйте, сделайте вывод.

3.2 Оцените параметры качества модели.

В общем случае при определении точности прогноза вы должны проанализировать четыре ситуации:

- действительно положительные (отнесение беременных к беременным);
- действительно отрицательные (отнесение небеременных к небеременным);
- ложноположительные (ошибки 1 рода) – отнесение небеременных к беременным;
- ложноположительные (ошибки 1 рода) – отнесение беременных к небеременным;

3.2.1 Установите граничные значения.

Для этого создайте лист «Эффективность модели». На листе введите названия столбцов как в таблице ниже (таблица 2):

Таблица 2 – Граничные значения

А	В	С	Д	Е	Ф
Минимальное предсказание	Возможные значения границы	Точность	Избирательность модели / Доля действительно отрицательных результатов	Частота ложноположительных результатов (1-специфичность)	Истинная положительная частота / отзывчивость / Чувствительность

В ячейке А2 найдите возможную нижнюю границу отсечения беременных от небеременных (минимальное значение из всех полученных прогнозных значений на листе «Тестовые данные»). В ячейке А4 введите «Максимальное предсказание» и в А5 найдите возможную верхнюю границу отсечения беременных от небеременных (минимальное значение из всех полученных прогнозных значений на листе «Test Set»).

3.2.2 В столбце В разместите возможные значения границы в диапазоне от минимального к максимальному с шагом 0,05. Например, – 0,35; – 0,3; – 0,25 и т. д.

3.2.3 В столбце С найдите точность (прогностическую положительную ценность полученного результата) для каждого граничного значения в столбце В. Для этого нужно посчитать сколько «беременных» покупателей получили прогноз больше либо равный данному граничному значению и разделить на общее количество строк с прогнозом, больше либо равным данному значению. Проанализируйте как изменяется точность модели, сделайте вывод.

$$=СЧЁТЕСЛИМН("Test Set"!V2:V1001;">=" & B2;"Test Set"!U2:U1001;"=1")/СЧЁТЕСЛИ("Test Set"!V2:V1001;">=" & B2)$$

3.2.4 В столбце D оцените избирательность модели (доля действительно отрицательных результатов). То есть нужно рассчитать какая доля «небеременных» покупателей для данной границы были верно отнесены к общему числу небеременных. Проанализируйте как изменяется избирательность модели, сделайте вывод.

$$=СЧЁТЕСЛИМН("Test Set"!V2:V1001;"<" & B2;"Test Set"!U2:U1001;"=0")/СЧЁТЕСЛИ("Test Set"!U2:U1001;"=0")$$

3.2.5 В столбце Е рассчитайте долю ложноположительных результатов. Можно вычислить, что 1 – доля действительно отрицательных результатов. Проанализируйте, как изменяется показатель, сделайте вывод.

3.2.6 В столбце F вычислите долю действительно положительных результатов (память, чувствительность). Это доля правильно определенных беременных женщин в общем количестве беременных женщин по всему набору данных. Проанализируйте, как изменяется показатель модели, сделайте вывод (рисунок 79).

$$=СЧЁТЕСЛИМН("Test Set"!V2:V1001;">=" & B2;"Test Set"!U2:U1001;"=1")/СЧЁТЕСЛИ("Test Set"!U2:U1001;"=1")$$

	A	B	C	D	E	F
1	Минимальное предсказание	Возможные значения границы	Точность	Избирательность модели/ Доля действительно отрицательных	Доля ложноположительных результатов (1 - избирательность)	Доля действительно положительных результатов / память / Чувствительность
2		-0,35	-0,35	0,06	0,00	1,00
3			-0,3	0,06	0,00	1,00
4	Максимальное предсказание					
5			-0,25	0,06	0,01	0,99
6	1,25		-0,2	0,06	0,02	0,98
7			-0,15	0,06	0,03	0,97
8			-0,1	0,06	0,05	0,95
9			-0,05	0,06	0,08	0,92
10			0	0,07	0,11	0,89
11			0,05	0,07	0,13	0,87
12			0,1	0,07	0,18	0,82
13			0,15	0,07	0,22	0,78
14			0,2	0,09	0,34	0,66
15			0,25	0,10	0,42	0,58
16			0,3	0,11	0,49	0,51
17			0,35	0,11	0,53	0,47
18			0,4	0,16	0,69	0,31
19			0,45	0,18	0,76	0,24
			0,5	0,31	0,89	0,11

Рисунок 79 – Результаты оценки качества модели

3.3 Постройте кривую ошибок (кривая соотношения правильного и ложного сигналов – Receiver Operating Characteristic, ROC). Это график зависимости действительно положительных значений от ложноположительных (столбцы E и F). Пример представлен на рисунке 80. По данной кривой можно понять, что модель позволяет идентифицировать 40 % беременных покупателей при граничном значении прогноза 0,85 без единого ложного срабатывания. А доли действительно положительных значений 75 % можно достигнуть всего лишь при 9 ложных срабатываниях (рисунок 81).

3.4 Выберите и обоснуйте граничные значения для вашей модели.

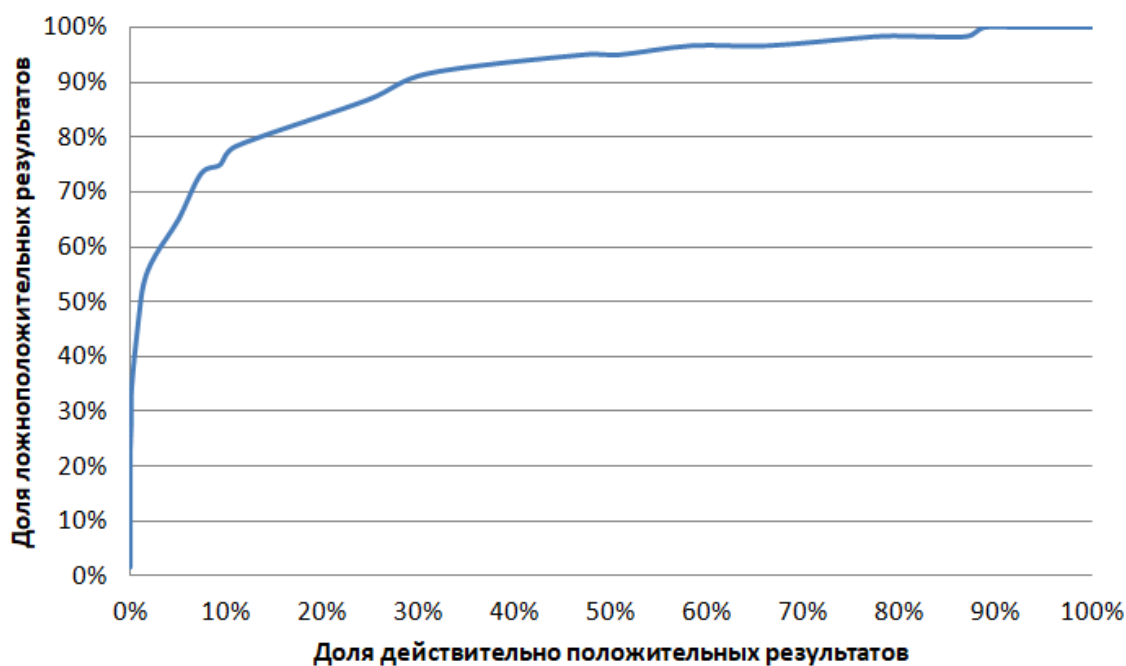


Рисунок 80 – Кривая ошибок

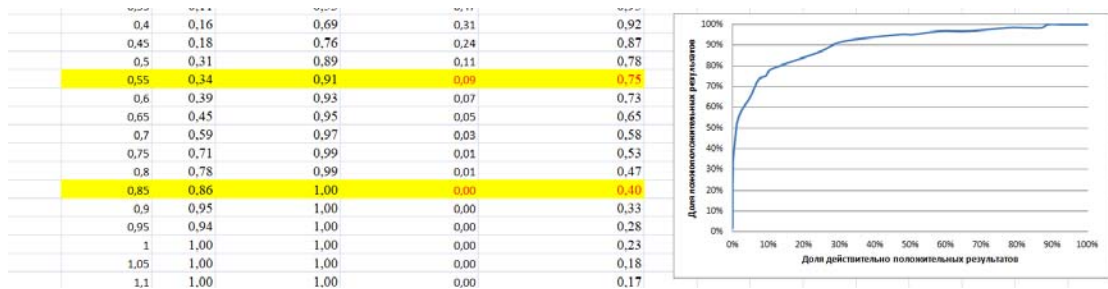


Рисунок 81 – Анализ кривой ошибок

Контрольные вопросы

- 1 Понятие множественной регрессии.
- 2 Опишите процесс обучения регрессионной модели.
- 3 Показатели значимости регрессионной модели: коэффициент детерминации, критерии Фишера и Стьюдента.
- 4 Параметры качества регрессионной модели.
- 5 Кривая ошибок.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1 Макшанов А. В. Технологии интеллектуального анализа данных : учебн. пособие / А. В. Макшанов, А. Е. Журавлев. – 2-е изд., стер. – Санкт-Петербург : Лань, 2019. – 212 с. Электронно-библиотечная система «Лань». – URL: <https://lanbook.com/catalog/informatika/tekhnologii-intellektualnogo-analiza-dannykh/?ysclid=mo2ixukxdw940403181> (дата обращения: 01.10.2025).

2 Мицель А. А. Прикладная математическая статистика : учеб. пособие / А. А. Мицель. – Томск : ТУСУР, 2019. – 113 с. – URL: <https://edu.tusur.ru/publications/9151> (дата обращения: 01.10.2025).

3 Форман Д. Много цифр : Анализ больших данных при помощи Excel / Д. Форман ; перевод А. Соколовой. – Москва : Альпина Пабlishер, 2016. – 461 с. Электронно-библиотечная система Индекс «Лань». – URL: <https://e.lanbook.com/book/87871> (дата обращения: 01.10.2025).

Адаменко Юлия Владимировна

Анализ данных. Часть 1. Классификация и регрессия. Статистические методы.
Методические указания к выполнению лабораторных работ
для бакалавров направлений
09.03.03 «Прикладная информатика»,
09.03.04 «Программная инженерия»

Редактор В. А. Лисина

БИЦ Курганского государственного университета.
640020, г. Курган, ул. Советская, 63/4.
Курганский государственный университет.