

Федеральное государственное автономное  
образовательное учреждение высшего образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт математики и фундаментальной информатики  
Базовая кафедра вычислительных и информационных технологий

**УТВЕРЖДАЮ**

/Заведующий кафедрой

 /В.В. Шайдуров


«20» июня 2016г.

## БАКАЛАВРСКАЯ РАБОТА


Направление 02.03.01 Математика и компьютерные науки

### ПРОГРАММНЫЕ СРЕДСТВА ДЛЯ РЕШЕНИЯ ЗАДАЧ СЛУЧАЙНО-МНОЖЕСТВЕННОГО РЕГРЕССИОННОГО АНАЛИЗА

Научный руководитель  
кандидат физико-математических наук,  
доцент

 /Д.В. Семенова  
20.06.2016

Выпускник

 /М.А. Атабеков  
20.06.2016

Красноярск 2016

## РЕФЕРАТ

Бакалаврская работа по теме «Программные средства для решения задач случайно-множественного регрессионного анализа» содержит 71 страницу текста, одно приложение, 20 использованных источников.

СЕТ-РЕГРЕССИЯ, СЛУЧАЙНОЕ МНОЖЕСТВО, СЕТ-КВАНТИЛЬ, СЕТ-МОДА, СЕТ-МЕДИАНА, СЕТ-ОЖИДАНИЕ, СОВМЕСТНОЕ РАСПРЕДЕЛЕНИЕ СЛУЧАЙНЫХ МНОЖЕСТВ, СУПЕРПОЗИЦИЯ СЕТ-РЕГРЕССИЙ.

Целью работы является разработка алгоритмов и программного обеспечения, основанных на теории конечных случайных множеств, и нацеленных на решение узкого перечня задач случайно-множественного регрессионного анализа, значимых для анализа нечисловых данных в статистических системах.

В результате исследований разработаны алгоритмы решения двух задач случайно-множественного регрессионного анализа: задачи сет-регрессии и суперпозиции сет-регрессий. Проведены вычислительные эксперименты, в том числе и на реальных статистических данных, и дана оценка эффективности разработанных алгоритмов. Конечным продуктом научного исследования является разработанное программное обеспечение (ПО) «SetRegression», подробно иллюстрирующее поэтапное решение задачи сет-регрессии. Для конечного пользователя ПО является универсальным инструментом получения некоторых теоретических сведений из исходных данных, представленных в соответствующем виде, в процессе научных или других исследований.

# СОДЕРЖАНИЕ

Введение . . . . .	3
1 Случайно-множественный регрессионный анализ . . . . .	6
1.1 Конечные случайные множества и их распределения . . . . .	6
1.2 Средние характеристики случайного множества . . . . .	8
1.3 Задача сет-регрессии . . . . .	10
1.4 Пример решения задачи сет-регрессии . . . . .	12
1.5 Суперпозиция сет-регрессий . . . . .	22
2 Алгоритмы и комплекс программ для сет-регрессионного анализа . .	24
2.1 Сведения о программном комплексе «SetRegression» . . . . .	24
2.2 Описание разработанных алгоритмов . . . . .	25
2.3 Анализ разработанных алгоритмов . . . . .	33
3 Пример решения задачи построения сет-регрессии для анализа ме- дицинских данных . . . . .	39
3.1 Программное решение задачи суперпозиции сет-регрессий . . .	39
3.2 Предобработка исходной статистики для анализа медицинских данных . . . . .	42
3.3 Аprobация статистики медицинского исследования в программ- ном комплексе «SetRegression» . . . . .	47
Заключение . . . . .	56
Список использованных источников . . . . .	58
Приложение А . . . . .	61

## ВВЕДЕНИЕ

Главная цель компьютерной обработки экспериментальных или статистических данных состоит в автоматическом обнаружении скрытых в них закономерностей. Эти закономерности или знания позволяют понять сущность изучаемого процесса и, опираясь на имеющиеся данные, предсказывать новые факты [8]. Использование технологий анализа данных становится необходимым при обработке результатов наблюдений, экспериментов, измерений. Существующее программное и алгоритмическое обеспечение, используемое в этой области, разнообразно по своему назначению и позволяет решать множество задач, как статистического анализа данных, так и анализа данных, в понимании этого термина как совокупности статистических методов, не предполагающих вероятностной модели изучаемого явления. Оба этих подхода можно отнести к прикладной статистике, одним из важных разделов которой является статистика объектов нечисловой природы. Данные нечисловой природы встречаются в исследованиях крайне часто, и для их обработки следует привлекать корректные методы, основанные на соответствующих вероятностных моделях [13, 14].

Нередко при обработке данных возникает ситуация, когда исходная информация об объекте исследования представлена нечисловыми признаками [3, 13]. В этом случае, адекватной математической моделью данных являются случайные множества, относящиеся к одному из объектов статистики нечисловой природы [13, 14]. Для выборочных данных, описываемых нечисловыми признаками, приходится решать те же самые задачи, что и в классических разделах математической статистики: классификация объектов без указания учителя, распознавание образов объектов, оценивание регрессионной зависимости и другие [3].

**Цели и задачи.** Целью исследования является разработка алгоритмов

и программного обеспечения, основанных на теории конечных случайных множеств, и предназначенных для решения задач сет-регрессионного анализа, имеющих широкое приложение для анализа нечисловых данных в различных статистических системах. Для достижения поставленной цели необходимо решить следующие задачи:

1. Проанализировать теоретические основы решения задач случайно-множественного регрессионного анализа по работам профессора Воробьева О.Ю., Фомина А.Ю., Семеновой Д.В., Лукьяновой Н.А., Ивановой А.И..
2. Разработать алгоритмы для программного решения задачи сет-регрессии и её модификации.
3. Оценить на практических примерах эффективность разработанных алгоритмов. Провести вычислительные эксперименты и сделать выводы по их результатам.
4. Разработать на основе полученных алгоритмов программное обеспечение для сет-регрессионного анализа бинарных данных.
5. Привести пример решения задачи на конкретном практическом примере. В качестве входных данных использовать реальные статистические данные.

**Структура работы.** Представленная работа состоит из введения, трех глав, заключения, списка использованной литературы (20 наименований) и приложений. Общий объем работы — 71 страница.

В главе 1 приведены используемые в бакалаврской работе основные определения и теоремы случайно-множественного регрессионного анализа. Рассмотрены постановки и методы решения задачи сет-регрессии и задачи суперпозиций сет-регрессий. Во второй главе описаны разработанные алгоритмы, технические требования, рекомендации по использованию программного

обеспечения, перечислены основные возможности ПО «SetRegression», приведены результаты вычислительных экспериментов и оценка эффективности полученных алгоритмов. В последней главе акцентируется внимание на практическом применении программного комплекса «SetRegression» на примере обработки статистической базы данных по результатам анкетирования двух медицинских учреждений Красноярского края ККПТД №1 и ККПТД №2 за 2012–2013 годы. Личная информация пациентов медицинских учреждений закреплена в целях соблюдения ФЗ 152 «О персональных данных». Результаты расчетов представлены в виде графика функции регрессии одного случайного множества на другое. В заключении изложены основные результаты научной работы, приведен список публикаций и конференций, на которых были представлены некоторые части данной работы, также приведен список использованных источников, использованных при подготовке бакалаврской работы.

# 1 Случайно-множественный регрессионный анализ

## 1.1 Конечные случайные множества и их распределения

Пусть  $(\Omega, \mathcal{F}, \mathbf{P})$  — вероятностное пространство. И пусть  $\mathfrak{X}$  — конечное множество событий, выбранных из  $\mathcal{F}$  ( $2^{\mathfrak{X}}, \mathcal{A}$ ) — измеримое пространство, где  $2^{\mathfrak{X}}$  — булеан некоторого конечного множества  $\mathfrak{X}$ , а  $\mathcal{A} = 2^{2^{\mathfrak{X}}}$  —  $\sigma$ -алгебра его подмножеств [15].

**Определение 1.1.** *Отображение вида  $K : \Omega \rightarrow 2^{\mathfrak{X}}$ , измеримое относительно пары алгебр  $(\mathcal{F}, 2^{2^{\mathfrak{X}}})$  такое, что  $\mathbf{P}(X) = \mathbf{P}(K^{-1}(X))$  называется случайным множеством  $K$  на конечном множестве событий  $\mathfrak{X} \subset \mathcal{F}$ .*

Любое событие  $x \in \mathfrak{X}$  разбивает пространство элементарных исходов  $\Omega$  на два непересекающихся события: событие  $x$  и его дополнение  $x^c$ , для которых верно следующее выражение:  $x^c = \Omega - x$ . Событием-терраской множества  $X$  будем называть подмножество  $\Omega$

$$\text{ter}(X) = \left( \left( \bigcap_{x \in X} x \right) \cap \left( \bigcap_{x \in X^c} x^c \right) \right),$$

где  $X \subseteq \mathfrak{X}$ ,  $X^c = \mathfrak{X} \setminus X$  — дополнение подмножества событий  $X$  до  $\mathfrak{X}$ . Заметим, что совокупность всех событий-террасок  $X \subseteq \mathfrak{X}$  образуют полную группу несовместных событий, т.е.

$$\text{ter}(X) \cap \text{ter}(Y) = \emptyset, \quad X \neq Y, \quad \Omega = \sum_{X \subseteq \mathfrak{X}} \text{ter}(X).$$

**Определение 1.2.** *Распределение случайного множества на конечном множестве событий — это специальный класс распределений, содержащий полную информацию о всех типах вероятностных зависимостей событий данного множества.*

Пусть  $K$  — случайное множество, определенное на конечном множестве событий  $\mathfrak{X}$ ,  $|\mathfrak{X}| = N$ . Распределение случайного множества представимо ше-

стью эквивалентными способами [7]. В работе используются только два из них.

— Распределение I-го рода — это набор из  $2^N$  вероятностей  $\{p(X), X \subseteq \mathfrak{X}\}$ ,

где

$$p(X) = \mathbf{P}(K = X) = \mathbf{P}(\text{ter}(X)) = \mathbf{P}\left(\bigcap_{x \in X} x \bigcap_{x \in X^c} x^c\right).$$

— Распределение II-го рода — это набор из  $2^N$  вероятностей  $\{p_X, X \subseteq \mathfrak{X}\}$ ,

где

$$p_X = \mathbf{P}(X \subseteq K) = \mathbf{P}\left(\bigcap_{x \in X} x\right).$$

Вышеописанные распределения I-го и II-го рода связаны взаимно-обратными формулами обращения Мёбиуса [7]:

$$p_X = \sum_{X \subseteq Y} p(Y), \quad (1.1)$$

$$p(X) = \sum_{X \subseteq Y} (-1)^{|Y|-|X|} p_Y. \quad (1.2)$$

Введем обозначение:  $X // \mathfrak{X}$  — подмножество  $X$  из фиксированного конечного множества  $\mathfrak{X}$ .

**Определение 1.3.** Для любых случайных множеств  $K$  и  $L$ , значения которых содержатся в непересекающихся множествах  $\mathfrak{X}, \mathfrak{Y} \subset \mathcal{F}$ ,  $\mathfrak{X} \cap \mathfrak{Y} = \emptyset$ , соответственно, совместное распределение случайных множеств есть набор из  $2^{|\mathfrak{X}|+|\mathfrak{Y}|}$  вероятностей вида

$$\{p(X, Y), X \subseteq \mathfrak{X}, Y \subseteq \mathfrak{Y}\},$$

где  $p(X, Y) = \mathbf{P}(K = X, L = Y) = \mathbf{P}(\text{ter}(X + Y // \mathfrak{X} + \mathfrak{Y}))$ .

В работе [20] предложено матричное представление совместного распределения случайных множеств в виде табл. 1.



Таблица 1 - Совместное распределение  $\mathbf{P}(K = X, L = Y)$  двух случайных множеств событий  $K$  и  $L$ , где  $K$  задано на  $\mathfrak{X}$ ,  $L$  — на  $\mathfrak{Y}$ ,  $\mathfrak{X} \cap \mathfrak{Y} = \emptyset$

	$\text{ter}(\emptyset // \mathfrak{Y})$	...	$\text{ter}(Y // \mathfrak{Y})$	...	$\text{ter}(\mathfrak{Y} // \mathfrak{Y})$	$\mathbf{P}(K = X)$
$\text{ter}(\emptyset // \mathfrak{X})$	$p(\text{ter}(\emptyset // \mathfrak{X} + \mathfrak{Y}))$	...	$p(\text{ter}(Y // \mathfrak{X} + \mathfrak{Y}))$	...	$p(\text{ter}(\mathfrak{Y} // \mathfrak{X} + \mathfrak{Y}))$	$p(\text{ter}(\emptyset // \mathfrak{X}))$
...	...	...	...	...	...	...
$\text{ter}(X // \mathfrak{X})$	$p(\text{ter}(X // \mathfrak{X} + \mathfrak{Y}))$	...	$p(\text{ter}(X + Y // \mathfrak{X} + \mathfrak{Y}))$	...	$p(\text{ter}(X + \mathfrak{Y} // \mathfrak{X} + \mathfrak{Y}))$	$p(\text{ter}(X // \mathfrak{X}))$
...	...	...	...	...	...	...
$\text{ter}(\mathfrak{X} // \mathfrak{X})$	$p(\text{ter}(\mathfrak{X} // \mathfrak{X} + \mathfrak{Y}))$	...	$p(\text{ter}(\mathfrak{X} + Y // \mathfrak{X} + \mathfrak{Y}))$	...	$p(\text{ter}(\mathfrak{X} + \mathfrak{Y} // \mathfrak{X} + \mathfrak{Y}))$	$p(\text{ter}(\mathfrak{X} // \mathfrak{X}))$
$\mathbf{P}(L = Y)$	$p(\text{ter}(\emptyset // \mathfrak{Y}))$	...	$p(\text{ter}(Y // \mathfrak{Y}))$	...	$p(\text{ter}(\mathfrak{Y} // \mathfrak{Y}))$	

**Определение 1.4.** Вероятность события  $\{K = X\}$ ,  $X \subseteq \mathfrak{X}$ , при наступлении событий  $\{L = Y\}$ ,  $Y \subseteq \mathfrak{Y}$ ,  $\mathfrak{X} \cap \mathfrak{Y} = \emptyset$ , называется **условной вероятностью** и определяется по формуле:

$$\mathbf{P}(\{K = X\} |_{\{L=Y\}}) = \mathbf{P}(\text{ter}(X) |_{\text{ter}(Y)}) = \frac{\mathbf{P}(\text{ter}(X + Y // \mathfrak{X} + \mathfrak{Y}))}{\mathbf{P}(\text{ter}(Y // \mathfrak{Y}))}. \quad (1.3)$$

Набор из  $2^{|\mathfrak{X}|}$  вероятностей (1.3) определяет условное распределение случайного множества  $K |_{L=Y}$ .

Условное распределение случайного множества  $K |_{L=Y}$  для наглядности удобно представить в виде таблицы 2.

Таблица 2 - Условное распределение случайного множества  $K |_{L=Y}$

$\{K = X\}$	$\text{ter}(\emptyset // \mathfrak{X})$	...	$\text{ter}(X // \mathfrak{X})$	...	$\text{ter}(\mathfrak{X} // \mathfrak{X})$
$\mathbf{P}(\{K = X\}  _{\{L=Y\}})$	$\frac{\mathbf{P}(\text{ter}(\emptyset + Y // \mathfrak{X} + \mathfrak{Y}))}{\mathbf{P}(\text{ter}(Y // \mathfrak{Y}))}$	...	$\frac{\mathbf{P}(\text{ter}(X + Y // \mathfrak{X} + \mathfrak{Y}))}{\mathbf{P}(\text{ter}(Y // \mathfrak{Y}))}$	...	$\frac{\mathbf{P}(\text{ter}(\mathfrak{X} + Y // \mathfrak{X} + \mathfrak{Y}))}{\mathbf{P}(\text{ter}(Y // \mathfrak{Y}))}$

## 1.2 Средние характеристики случайного множества

В [7, 6] введены средние характеристики или сет-средние случайного множества.

**Определение 1.5.** *Сет-ожиданием* случайного множества событий  $K$  называется множество

$$\mathcal{E}K = \{x : p_x \geq h\},$$

где  $h$  выбирается таким образом, чтобы мощность сет-ожидания была ближайшим целым к математическому ожиданию мощности случайного множества

$$h = \min_{h \in [0,1]} \left| |\mathcal{E}K| - \mathbf{E}|K| \right|.$$

**Определение 1.6.** *Множество*  $\text{Mod}K$ , *определенного как*

$$\text{Mod}K = \max_{X \subseteq \mathfrak{X}} \mathbf{P}(K = X)$$

называется *сет-модой* случайного множества событий  $K$ .

Если вероятность достигает максимального значения на нескольких подмножествах, то все они являются сет-модами.

**Определение 1.7.** *Множество вида*

$$\mathbf{Q}_\alpha K = \{x : p_x \geq \alpha\}.$$

будем называть *сет-квантилем* порядка  $\alpha$  случайного множества событий  $K$ , где  $\alpha$  принимает вещественные значения на интервале  $[0, 1]$ .

**Определение 1.8.** *Множество*  $\text{Med}K$ , *имеющее вид*

$$\text{Med}K = \left\{ x : p_x \geq \frac{1}{2} \right\},$$

называется *сет-медианой* случайного множества событий  $K$ .

Если все вероятности  $p_x < \frac{1}{2}$ , то сет-медиана — пустое множество. Сет-медиана эквивалентна сет-квантилю при значении порядка  $\alpha$  равном  $\frac{1}{2}$ .

Следующие теоремы об экстремальных свойствах сет-средних доказаны в [6].

**Теорема 1.1** (Экстремальность сет–квантиля). *Сет-квантиль порядка  $\alpha$  доставляет минимум среднему расстоянию до  $K$*

$$\mathbf{E}\rho_\alpha(K, \mathbf{Q}_\alpha K) = \min_{X \subseteq \mathfrak{X}} \mathbf{E}\rho_\alpha(K, X),$$

где  $\rho_\alpha$  — асимметричное расстояние на пространстве множеств событий

$$\rho_\alpha(X, Y) = |X \Delta Y|_\alpha = \alpha|X \setminus Y| + (1 - \alpha)|Y \setminus X|, \quad \alpha \in (0, 1).$$

**Теорема 1.2** (Экстремальность сет–моды). *Сет-мода доставляет минимум среднему расстоянию до  $K$*

$$\mathbf{E}\rho(K, \text{Mod}K) = \min_{X \subseteq \mathfrak{X}} \mathbf{E}\rho(K, X),$$

где  $\rho$  — дискретное расстояние на пространстве множеств событий

$$\rho(X, Y) = \begin{cases} 1, & X = Y \\ 0, & X \neq Y. \end{cases}$$

**Теорема 1.3** (Экстремальность сет–медианы). *Сет-медиана доставляет минимум среднему расстоянию до  $K$*

$$\mathbf{E}\rho(K, \text{Med}K) = \min_{X \subseteq \mathfrak{X}} \mathbf{E}\rho(K, X),$$

где  $\rho$  — расстояние на пространстве множеств событий

$$\rho(X, Y) = |X \Delta Y| = |X \setminus Y| + |Y \setminus X|.$$

### 1.3 Задача сет-регрессии

Пусть требуется построить оптимальное отображение одного случайного множества  $K$  на другое  $L$ , значения которых содержатся во множествах  $\mathfrak{X}$  и  $\mathfrak{Y}$  соответственно. Регрессией  $K$  на  $L$  называется любой оператор  $\Psi : 2^{\mathfrak{Y}} \rightarrow 2^{\mathfrak{X}}$ , приближенно представляющий статистическую зависимость  $K$  от  $L$ . Оптимальное регрессионное отображение необходимо находить из

условия минимума некоторого функционала  $\Phi$ , характеризующего близость случайного множества  $K$  и его регрессии на случайное множество  $L$ :

$$\Phi(K, \Psi(L)) \rightarrow \min_{\Psi}. \quad (1.4)$$

Это справедливо также и для других объектов нечисловой природы. В пространстве конечных множеств рассматриваемая задача получила название сет-регрессии (от англ. set - множество) [6].

Приведем постановку задачи сет-регрессии, используемую в работе [6, 9].

**Задача сет-регрессии.** Пусть заданы

- совместное распределение  $\{p(X, Y), X \subseteq \mathfrak{X}, Y \subseteq \mathfrak{Y}, \mathfrak{X} \cap \mathfrak{Y} = \emptyset\}$  случайных множеств  $K$  и  $L$ , значения которых содержатся в непересекающихся множествах  $\mathfrak{X}, \mathfrak{Y} \subset \mathcal{F}$ , соответственно;
- $\rho : 2^{\mathfrak{X}} \rightarrow 2^{\mathfrak{X}}$  — некоторая метрика в  $\mathfrak{X}$ .

Найти функцию  $\Psi : 2^{\mathfrak{Y}} \rightarrow 2^{\mathfrak{X}}$ , которая доставляет минимум функционалу  $E\rho(K, \Psi(Y))$ , где  $E$  — оператор математического ожидания.

В [6] показано, что регрессии одного случайного множества на другое в виде условной сет-моды, сет-медианы, сет-ожидания обладают экстремальным свойством минимизировать определенные характеристики близости случайного множества и его регрессии на другое случайное множество. По аналогии с многомерным статистическим подходом и используя экстремальные свойства сет-средних, регрессию одного случайного множества на другое можно искать в виде условных сет-средних [6, 7, 16]

- условного сет-квантиля порядка  $\alpha$

$$\Psi(Y) = \mathbf{Q}_{\alpha} [K |_{L=\text{ter}(Y)}] = \{x : \mathbf{P}(x |_{\text{ter}(Y)}) \geq \alpha, x \in \mathfrak{X}\}; \quad (1.5)$$

— условной сет-медианы

$$\Psi(Y) = \text{Med} [K |_{L=\text{ter}(Y)}] = \left\{ x : \mathbf{P} (x |_{\text{ter}(Y)}) \geq \frac{1}{2}, x \in \mathfrak{X} \right\}; \quad (1.6)$$

— условной сет-моды

$$\Psi(Y) = \text{Mod} [K |_{L=\text{ter}(Y)}] = \max_{X \subseteq \mathfrak{X}} \mathbf{P} (K = X |_{L=\text{ter}(Y)}); \quad (1.7)$$

— условного сет-ожидания

$$\Psi(Y) = \mathcal{E} [K |_{L=\text{ter}(Y)}] = \{x : \mathbf{P} (x |_{\text{ter}(Y)}) \geq h, x \in \mathfrak{X}\}. \quad (1.8)$$

#### 1.4 Пример решения задачи сет-регрессии

Пусть заданы

— множество событий  $\mathfrak{X} = \{x_1, x_2, x_3\}$ ,  $\mathfrak{X} \subset \mathcal{F}$ ;

— множество  $\mathfrak{Y} = \{y_1, y_2, y_3, y_4, y_5\}$ ,  $\mathfrak{Y} \subset \mathcal{F}$ ;

— совместное распределение  $\{p(X, Y), X \subseteq \mathfrak{X}, Y \subseteq \mathfrak{Y}\}$  двух случайных множеств событий  $K$  и  $L$ , значения которых содержатся в конечных множествах  $\mathfrak{X}$  и  $\mathfrak{Y}$  соответственно (таблица 3).

Необходимо найти функцию  $\Psi : 2^{\mathfrak{Y}} \rightarrow 2^{\mathfrak{X}}$ , которая доставляет минимум функционалу  $\mathbf{E}\rho(K, \Psi(Y))$ , где  $\mathbf{E}$  оператор математического ожидания, а  $\rho : 2^{\mathfrak{X}} \rightarrow 2^{\mathfrak{X}}$  — некоторая метрика в  $\mathfrak{X}$ .

Будем искать решение задачи в виде условных сет-средних по формулам (1.5) — (1.7).

На первом этапе решения из таблицы 3 совместного распределения двух случайных множеств событий  $K$  и  $L$  найдем маргинальные распределения случайных множеств  $K$  (таблица 4) и  $L$  (таблица 5), суммируя соответствующие столбцы или строки по формуле

$$p(X // \mathfrak{X}) = \sum_{Y \subseteq \mathfrak{Y}} p(X, Y).$$

Таблица 3 - Матричное представление совместного распределения двух случайных множеств событий  $K$  и  $L$ .

$\mathcal{Y} \backslash \mathcal{X}$	$\text{ter}(\emptyset)$	$\text{ter}(x_1)$	$\text{ter}(x_2)$	$\text{ter}(x_3)$	$\text{ter}(x_1x_2)$	$\text{ter}(x_1x_3)$	$\text{ter}(x_2x_3)$	$\text{ter}(x_1x_2x_3)$
$\text{ter}(\emptyset)$	0.004	0.002	0.007	0	0.001	0.005	0.011	0.002
$\text{ter}(y_1)$	0.004	0.001	0.004	0.001	0	0.0009	0.0070	0.006
$\text{ter}(y_2)$	0.004	0.004	0.003	0.0059	0.0080	0.002	0.0059	0.002
$\text{ter}(y_3)$	0.0070	0	0.002	0.001	0.0019	0.012	0.001	0.012
$\text{ter}(y_1y_2)$	0.003	0.001	0.004	0	0.019	0	0.003	0.003
$\text{ter}(y_4)$	0.0090	0.001	0.02	0.002	0.0090	0.0019	0.003	0.001
$\text{ter}(y_1y_3)$	0.005	0.002	0.003	0.004	0.002	0.001	0.002	0.0070
$\text{ter}(y_1y_4)$	0.0009	0.003	0.001	0.003	0.002	0.001	0.003	0.005
$\text{ter}(y_2y_3)$	0.004	0.001	0.002	0.01	0.002	0.005	0.003	0.011
$\text{ter}(y_5)$	0	0	0.01	0.001	0.002	0.001	0	0.005
$\text{ter}(y_2y_4)$	0.0080	0.0060	0.005	0.001	0.002	0.002	0.0070	0.0069
$\text{ter}(y_1y_2y_3)$	0.004	0.0070	0.004	0.001	0.001	0	0.002	0.001
$\text{ter}(y_1y_5)$	0.003	0.004	0	0.001	0.002	0.001	0.014	0.001
$\text{ter}(y_2y_5)$	0.004	0.002	0.0089	0.003	0.004	0.0070	0.001	0
$\text{ter}(y_3y_4)$	0.012	0.005	0.004	0.001	0.005	0.002	0.001	0.0070
$\text{ter}(y_1y_2y_4)$	0.0070	0.0069	0.003	0.004	0	0.003	0.003	0.004
$\text{ter}(y_1y_3y_4)$	0.003	0	0.004	0.003	0.001	0.003	0.001	0.0080
$\text{ter}(y_1y_2y_5)$	0.001	0.001	0.0060	0.005	0.002	0.001	0	0.003
$\text{ter}(y_3y_5)$	0	0.004	0.005	0.002	0.002	0	0.001	0.001
$\text{ter}(y_1y_3y_5)$	0.002	0.0189	0.003	0.015	0.002	0.002	0	0.001
$\text{ter}(y_2y_3y_4)$	0.003	0.023	0.0070	0.003	0.001	0.005	0.0090	0.002
$\text{ter}(y_4y_5)$	0.002	0.003	0.003	0.001	0.0070	0.0060	0	0.001
$\text{ter}(y_1y_2y_3y_4)$	0.003	0.003	0.001	0.003	0.0169	0.001	0.002	0
$\text{ter}(y_1y_4y_5)$	0.004	0.0060	0.012	0.014	0.005	0	0.002	0.002
$\text{ter}(y_2y_3y_5)$	0.0060	0.002	0.005	0.001	0.0060	0.003	0.001	0
$\text{ter}(y_1y_2y_3y_5)$	0	0.001	0.001	0.001	0.012	0.004	0.017	0.0060
$\text{ter}(y_2y_4y_5)$	0.0069	0.002	0.0060	0	0.022	0.001	0.016	0.012
$\text{ter}(y_1y_2y_4y_5)$	0.001	0.001	0.004	0.001	0.002	0.003	0.004	0.003
$\text{ter}(y_3y_4y_5)$	0.001	0.0060	0.0070	0.0070	0.005	0.005	0.002	0.001
$\text{ter}(y_1y_3y_4y_5)$	0.004	0.004	0.0009	0.001	0	0.001	0.0080	0.002
$\text{ter}(y_2y_3y_4y_5)$	0.0080	0	0.002	0.003	0.002	0.001	0.0060	0.003
$\text{ter}(y_1y_2y_3y_4y_5)$	0	0.001	0.0080	0.003	0.002	0.001	0.001	0.0090

Таблица 4 - Маргинальное распределение случайного множества событий  $K$  на  $\mathfrak{X}$

$\text{ter}(X), X \subseteq \mathfrak{X}$	$p(X // \mathfrak{X}) = \mathbf{P}(K = X)$
$\text{ter}(\emptyset) = x_1^c \cap x_2^c \cap x_3^c$	0.124
$\text{ter}(x_1) = x_1 \cap x_2^c \cap x_3^c$	0.122
$\text{ter}(x_2) = x_1^c \cap x_2 \cap x_3^c$	0.156
$\text{ter}(x_3) = x_1^c \cap x_2^c \cap x_3$	0.102
$\text{ter}(x_1x_2) = x_1 \cap x_2 \cap x_3^c$	0.149
$\text{ter}(x_1x_3) = x_1 \cap x_2^c \cap x_3$	0.082
$\text{ter}(x_2x_3) = x_1^c \cap x_2 \cap x_3$	0.137
$\text{ter}(x_1x_2x_3) = x_1 \cap x_2 \cap x_3$	0.0128

На втором этапе найдем все условные распределения случайного множества  $K$  при условии, что наступило подмножество событий  $Y \subseteq \mathfrak{Y}$ , пользуясь формулой (1.3). Например, при фиксированных  $\{K = \{x_3\}\}$  и  $\{L = \{y_3\}\}$  условная вероятность согласно (1.3) равна

$$\mathbf{P}(\{K = \{x_3\}\} | \{L = \{y_3\}\}) = \frac{0.002}{0.037} = 0.054.$$

Результаты расчетов условных распределений сведены в таблице 6.

По формуле (1.1) для каждого из условных распределений таблицы 6 найдем соответствующие вероятности событий из  $\mathfrak{X}$  (вероятности II-го рода). Результаты расчетов приведены в таблице 7.

Найдем решение задачи в виде сет-квантиля уровня  $\alpha = 1/5$ . Формула (1) примет вид

$$\mathbf{Q}_{1/5} [K |_{L=\text{ter}(Y)}] = \{x : \mathbf{P}(x |_{\text{ter}(Y)}) \geq 1/4\}. \quad (1.9)$$

Введём функцию  $I_\alpha : \mathfrak{X} \times 2^{\mathfrak{Y}} \rightarrow \{0, 1\}$ ,  $\alpha \in [0, 1]$  значения которой определя-

Таблица 5 - Маргинальное распределение случайного множества событий  $L$  на  $\mathfrak{Y}$

$\text{ter}(Y), Y \subseteq \mathfrak{Y}$	$p(Y // \mathfrak{Y}) = \mathbf{P}(L = Y)$
$\text{ter}(\emptyset) = y_1^c \cap y_2^c \cap y_3^c \cap y_4^c \cap y_5^c$	0.032
$\text{ter}(y_1) = y_1 \cap y_2^c \cap y_3^c \cap y_4^c \cap y_5^c$	0.024
$\text{ter}(y_2) = y_1^c \cap y_2 \cap y_3^c \cap y_4^c \cap y_5^c$	0.035
$\text{ter}(y_3) = y_1^c \cap y_2^c \cap y_3 \cap y_4^c \cap y_5^c$	0.037
$\text{ter}(y_1 y_2) = y_1 \cap y_2 \cap y_3^c \cap y_4^c \cap y_5^c$	0.033
$\text{ter}(y_4) = y_1^c \cap y_2^c \cap y_3^c \cap y_4 \cap y_5^c$	0.047
$\text{ter}(y_1 y_3) = y_1 \cap y_2^c \cap y_3 \cap y_4^c \cap y_5^c$	0.026
$\text{ter}(y_1 y_4) = y_1 \cap y_2^c \cap y_3^c \cap y_4 \cap y_5^c$	0.019
$\text{ter}(y_2 y_3) = y_1^c \cap y_2 \cap y_3 \cap y_4^c \cap y_5^c$	0.038
$\text{ter}(y_5) = y_1^c \cap y_2^c \cap y_3^c \cap y_4^c \cap y_5$	0.019
$\text{ter}(y_2 y_4) = y_1^c \cap y_2 \cap y_3^c \cap y_4 \cap y_5^c$	0.038
$\text{ter}(y_1 y_2 y_3) = y_1 \cap y_2 \cap y_3 \cap y_4^c \cap y_5^c$	0.02
$\text{ter}(y_1 y_5) = y_1 \cap y_2^c \cap y_3^c \cap y_4^c \cap y_5$	0.026
$\text{ter}(y_2 y_5) = y_1^c \cap y_2 \cap y_3^c \cap y_4^c \cap y_5$	0.03
$\text{ter}(y_3 y_4) = y_1^c \cap y_2^c \cap y_3 \cap y_4 \cap y_5^c$	0.037
$\text{ter}(y_1 y_2 y_4) = y_1 \cap y_2 \cap y_3^c \cap y_4 \cap y_5^c$	0.031
$\text{ter}(y_1 y_3 y_4) = y_1 \cap y_2^c \cap y_3 \cap y_4 \cap y_5^c$	0.023
$\text{ter}(y_1 y_2 y_5) = y_1 \cap y_2 \cap y_3^c \cap y_4^c \cap y_5$	0.019
$\text{ter}(y_3 y_5) = y_1^c \cap y_2^c \cap y_3 \cap y_4^c \cap y_5$	0.015
$\text{ter}(y_1 y_3 y_5) = y_1 \cap y_2^c \cap y_3 \cap y_4^c \cap y_5$	0.044
$\text{ter}(y_2 y_3 y_4) = y_1^c \cap y_2 \cap y_3 \cap y_4 \cap y_5^c$	0.053
$\text{ter}(y_4 y_5) = y_1^c \cap y_2^c \cap y_3^c \cap y_4 \cap y_5$	0.023
$\text{ter}(y_1 y_2 y_3 y_4) = y_1 \cap y_2 \cap y_3 \cap y_4 \cap y_5^c$	0.029
$\text{ter}(y_1 y_4 y_5) = y_1 \cap y_2^c \cap y_3^c \cap y_4 \cap y_5$	0.045
$\text{ter}(y_2 y_3 y_5) = y_1^c \cap y_2 \cap y_3 \cap y_4^c \cap y_5$	0.024
$\text{ter}(y_1 y_2 y_3 y_5) = y_1 \cap y_2 \cap y_3 \cap y_4^c \cap y_5$	0.042
$\text{ter}(y_2 y_4 y_5) = y_1^c \cap y_2 \cap y_3^c \cap y_4 \cap y_5$	0.066
$\text{ter}(y_1 y_2 y_4 y_5) = y_1 \cap y_2 \cap y_3^c \cap y_4 \cap y_5$	0.019
$\text{ter}(y_3 y_4 y_5) = y_1^c \cap y_2^c \cap y_3 \cap y_4 \cap y_5$	0.034
$\text{ter}(y_1 y_3 y_4 y_5) = y_1 \cap y_2^c \cap y_3 \cap y_4 \cap y_5$	0.021
$\text{ter}(y_2 y_3 y_4 y_5) = y_1^c \cap y_2 \cap y_3 \cap y_4 \cap y_5$	0.025
$\text{ter}(y_1 y_2 y_3 y_4 y_5) = y_1^c \cap y_2^c \cap y_3^c \cap y_4^c \cap y_5^c$	0.025



Таблица 6 - Условные распределения случайного множества  $K$  при условии, что наступило множество событий  $Y // \mathfrak{Y}$

$P(K L)$ \ $K$	$ter(\emptyset)$	$ter(x_1)$	$ter(x_2)$	$ter(x_3)$	$ter(x_1 x_2)$	$ter(x_1 x_3)$	$ter(x_2 x_3)$	$ter(x_1 x_2 x_3)$
$P(K _{L=ter(\emptyset)})$	0.125	0.0625	0.2187	0	0.0312	0.1562	0.3437	0.0625
$P(K _{L=ter(y_1)})$	0.1666	0.0416	0.1666	0.0416	0	0.0416	0.2916	0.25
$P(K _{L=ter(y_2)})$	0.1142	0.1142	0.0857	0.1714	0.2285	0.0571	0.1714	0.0571
$P(K _{L=ter(y_3)})$	0.1891	0	0.0540	0.0270	0.0540	0.3243	0.0270	0.3243
$P(K _{L=ter(y_1 y_2)})$	0.0909	0.0303	0.1212	0	0.5757	0	0.0909	0.0909
$P(K _{L=ter(y_4)})$	0.1914	0.0212	0.4255	0.0425	0.1914	0.0425	0.0638	0.0212
$P(K _{L=ter(y_1 y_3)})$	0.1923	0.0769	0.1153	0.1538	0.0769	0.0384	0.0769	0.2692
$P(K _{L=ter(y_1 y_4)})$	0.0526	0.1578	0.0526	0.1578	0.1052	0.0526	0.1578	0.2631
$P(K _{L=ter(y_2 y_3)})$	0.1052	0.0263	0.0526	0.2631	0.0526	0.1315	0.0789	0.2894
$P(K _{L=ter(y_5)})$	0	0	0.5263	0.0526	0.1052	0.0526	0	0.2631
$P(K _{L=ter(y_2 y_4)})$	0.2105	0.1578	0.1315	0.0263	0.0526	0.0526	0.1842	0.1842
$P(K _{L=ter(y_1 y_2 y_3)})$	0.2	0.3500	0.2	0.05	0.05	0	0.1	0.05
$P(K _{L=ter(y_1 y_5)})$	0.1153	0.1538	0	0.0384	0.0769	0.0384	0.5384	0.0384
$P(K _{L=ter(y_2 y_5)})$	0.1333	0.0666	0.2999	0.1	0.1333	0.2333	0.0333	0
$P(K _{L=ter(y_3 y_4)})$	0.3243	0.1351	0.1081	0.0270	0.1351	0.0540	0.0270	0.1891
$P(K _{L=ter(y_1 y_2 y_4)})$	0.2258	0.2258	0.0967	0.1290	0	0.0967	0.0967	0.1290
$P(K _{L=ter(y_1 y_3 y_4)})$	0.1304	0	0.1739	0.1304	0.0434	0.1304	0.0434	0.3478
$P(K _{L=ter(y_1 y_2 y_5)})$	0.0526	0.0526	0.3157	0.2631	0.1052	0.0526	0	0.1578
$P(K _{L=ter(y_3 y_5)})$	0	0.2666	0.3333	0.1333	0.1333	0	0.0666	0.0666
$P(K _{L=ter(y_1 y_3 y_5)})$	0.0454	0.4318	0.0681	0.3409	0.0454	0.0454	0	0.0227
$P(K _{L=ter(y_2 y_3 y_4)})$	0.0566	0.4339	0.1320	0.0566	0.0188	0.0943	0.1698	0.0377
$P(K _{L=ter(y_4 y_5)})$	0.0869	0.1304	0.1304	0.0434	0.3043	0.2608	0	0.0434
$P(K _{L=ter(y_1 y_2 y_3 y_4)})$	0.1	0.1	0.0333	0.1	0.5666	0.0333	0.0666	0
$P(K _{L=ter(y_1 y_4 y_5)})$	0.0888	0.1333	0.2666	0.3111	0.1111	0	0.0444	0.0444
$P(K _{L=ter(y_2 y_3 y_5)})$	0.25	0.0833	0.2083	0.0416	0.25	0.125	0.0416	0
$P(K _{L=ter(y_1 y_2 y_3 y_5)})$	0	0.0238	0.0238	0.0238	0.2857	0.0952	0.4047	0.1428
$P(K _{L=ter(y_2 y_4 y_5)})$	0.1060	0.0303	0.0909	0	0.3333	0.0151	0.2424	0.1818
$P(K _{L=ter(y_1 y_2 y_4 y_5)})$	0.0526	0.0526	0.2105	0.0526	0.1052	0.1578	0.2105	0.1578
$P(K _{L=ter(y_3 y_4 y_5)})$	0.0294	0.1764	0.2058	0.2058	0.1470	0.1470	0.0588	0.0294
$P(K _{L=ter(y_1 y_3 y_4 y_5)})$	0.1904	0.1904	0.0476	0.0476	0	0.0476	0.3809	0.0952
$P(K _{L=ter(y_2 y_3 y_4 y_5)})$	0.32	0	0.08	0.12	0.08	0.04	0.24	0.12
$P(K _{L=ter(y_1 y_2 y_3 y_4 y_5)})$	0	0.04	0.32	0.12	0.08	0.04	0.04	0.36

ются по формуле

$$I_\alpha(x, Y) = \begin{cases} 1, & \mathbf{P}(x|_{ter(Y)}) \geq \alpha \\ 0, & \text{иначе.} \end{cases} \quad (1.10)$$

Функция (1.10) фактически является индикатором множества, построенного по формуле (1.5), то есть индикатором условного сет-квантиля порядка  $\alpha$ .

Решение задачи в виде условного сет-квантиля представим в виде таблицы 8, элементы которой находятся по формуле (1.10). Каждая строка таблицы является индикаторной функцией соответствующего условного сет-квантиля.

Таблица 7 - Вероятности событий из  $\mathfrak{X}$  для условных распределений случайного множества  $K$  при условии, что наступило множество событий  $Y \subseteq \mathfrak{Y}$

$\mathfrak{Y}$	$\mathbf{P}(x_1 _{\text{ter}(Y)})$	$\mathbf{P}(x_2 _{\text{ter}(Y)})$	$\mathbf{P}(x_3 _{\text{ter}(Y)})$
ter( $\emptyset$ )	0.3125	0.6562	0.5625
ter( $y_1$ )	0.3333	0.7083	0.625
ter( $y_2$ )	0.4571	0.5428	0.4571
ter( $y_3$ )	0.7027	0.4594	0.7027
ter( $y_1y_2$ )	0.6969	0.8787	0.1818
ter( $y_4$ )	0.2765	0.7021	0.1702
ter( $y_1y_3$ )	0.4615	0.5384	0.5384
ter( $y_1y_4$ )	0.5789	0.5789	0.6315
ter( $y_2y_3$ )	0.5	0.4736	0.7631
ter( $y_5$ )	0.4210	0.8947	0.3684
ter( $y_2y_4$ )	0.4473	0.5526	0.4473
ter( $y_1y_2y_3$ )	0.4500	0.4	0.2
ter( $y_1y_5$ )	0.3076	0.6538	0.6538
ter( $y_2y_5$ )	0.4333	0.4666	0.3666
ter( $y_3y_4$ )	0.5135	0.4594	0.2972
ter( $y_1y_2y_4$ )	0.4516	0.3225	0.4516
ter( $y_1y_3y_4$ )	0.5217	0.6086	0.6521
ter( $y_1y_2y_5$ )	0.3684	0.5789	0.4736
ter( $y_3y_5$ )	0.4666	0.5999	0.2666
ter( $y_1y_3y_5$ )	0.5454	0.1363	0.4090
ter( $y_2y_3y_4$ )	0.5849	0.3584	0.3584
ter( $y_4y_5$ )	0.7391	0.4782	0.3478
ter( $y_1y_2y_3y_4$ )	0.6999	0.6666	0.2
ter( $y_1y_4y_5$ )	0.2888	0.4666	0.4
ter( $y_2y_3y_5$ )	0.4583	0.5	0.2083
ter( $y_1y_2y_3y_5$ )	0.5476	0.8571	0.6666
ter( $y_2y_4y_5$ )	0.5606	0.8484	0.4393
ter( $y_1y_2y_4y_5$ )	0.4736	0.6842	0.5789
ter( $y_3y_4y_5$ )	0.5	0.4411	0.4411
ter( $y_1y_3y_4y_5$ )	0.3333	0.5238	0.5714
ter( $y_2y_3y_4y_5$ )	0.24	0.52	0.52
ter( $y_1y_2y_3y_4y_5$ )	0.52	0.8	0.56

Таблица 8 - Матричное представление решения задачи в виде условного сет-квантиля порядка  $\alpha = 1/5$ . Каждая строка таблицы является индикаторной функцией соответствующего условного сет-квантиля

	$x_1$	$x_2$	$x_3$		$x_1$	$x_2$	$x_3$
ter( $\emptyset$ )	1	1	1	ter( $y_1y_3y_4$ )	1	1	1
ter( $y_1$ )	1	1	1	ter( $y_1y_2y_5$ )	1	1	1
ter( $y_2$ )	1	1	1	ter( $y_3y_5$ )	1	1	1
ter( $y_3$ )	1	1	1	ter( $y_1y_3y_5$ )	1	0	1
ter( $y_1y_2$ )	1	1	0	ter( $y_2y_3y_4$ )	1	1	1
ter( $y_4$ )	1	1	0	ter( $y_4y_5$ )	1	1	1
ter( $y_1y_3$ )	1	1	1	ter( $y_1y_2y_3y_4$ )	1	1	1
ter( $y_1y_4$ )	1	1	1	ter( $y_1y_4y_5$ )	1	1	1
ter( $y_2y_3$ )	1	1	1	ter( $y_2y_3y_5$ )	1	1	1
ter( $y_5$ )	1	1	1	ter( $y_1y_2y_3y_5$ )	1	1	1
ter( $y_2y_4$ )	1	1	1	ter( $y_2y_4y_5$ )	1	1	1
ter( $y_1y_2y_3$ )	1	1	1	ter( $y_1y_2y_4y_5$ )	1	1	1
ter( $y_1y_5$ )	1	1	1	ter( $y_3y_4y_5$ )	1	1	1
ter( $y_2y_5$ )	1	1	1	ter( $y_1y_3y_4y_5$ )	1	1	1
ter( $y_3y_4$ )	1	1	1	ter( $y_2y_3y_4y_5$ )	1	1	1
ter( $y_1y_2y_4$ )	1	1	1	ter( $y_1y_2y_3y_4y_5$ )	1	1	1

По таблице 8 запишем решение задачи в виде функции:

$$\Psi(Y) = \{x \in \mathfrak{X} : I_{1/5}(x, Y) = 1\},$$

для всех  $Y \subseteq \mathfrak{Y}$ .

На рис. 1 представлено решение задачи сет-регрессии в виде условного сет-квантиля порядка  $\alpha = 1/5$ .

Используя таблицу 7 и (1.10) по аналогии запишем решение задачи в виде условной сет-медианы

$$\Psi(Y) = \{x \in \mathfrak{X} : I_{1/2}(x, Y) = 1\},$$

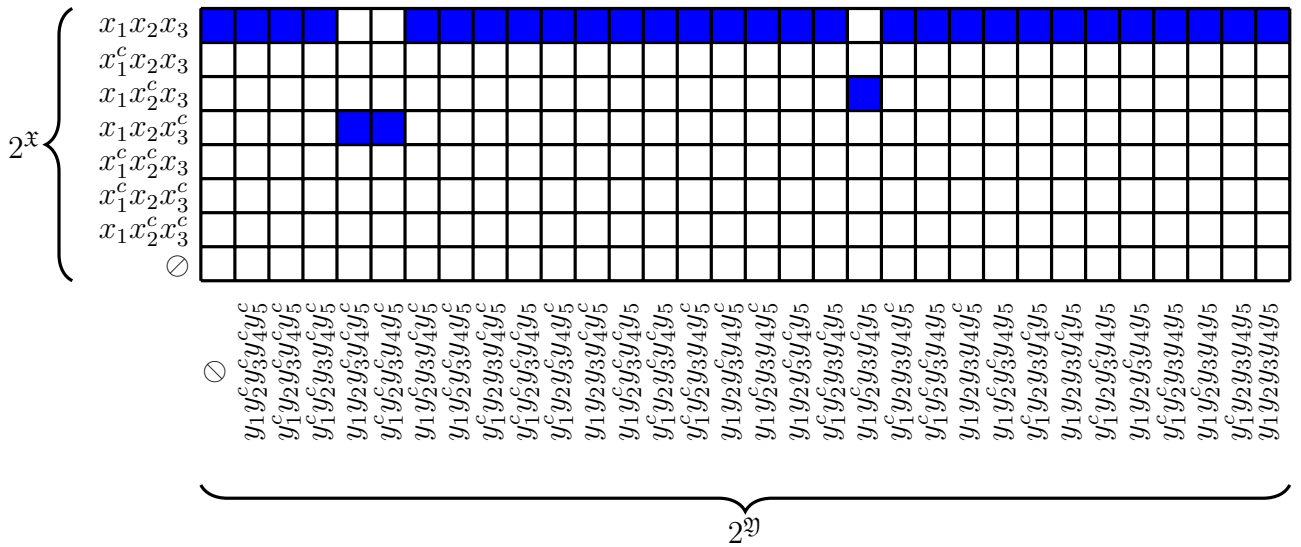


Рисунок 1 - Решение задачи сет-регрессии множества  $\mathcal{X}$  на  $\mathcal{Y}$  в виде условного сет-квантиля порядка  $\alpha = 1/5$

для всех  $Y \subseteq \mathcal{Y}$ . Матричное представление дано в таблице 9, график решения представлен на рис. 2.

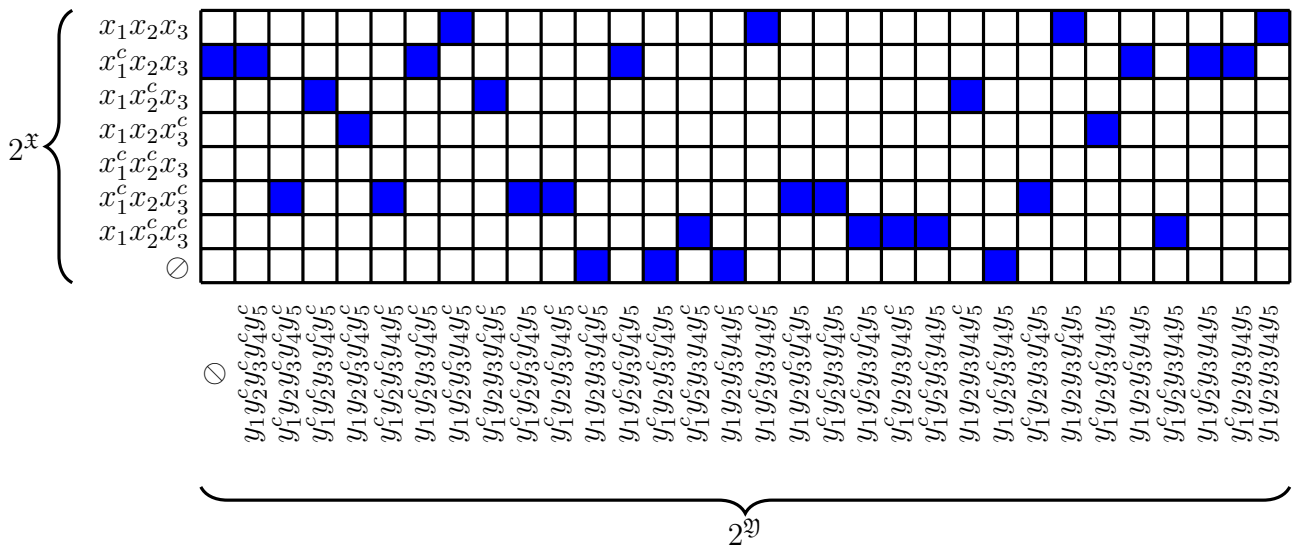


Рисунок 2 - Решение задачи сет-регрессии множества  $\mathcal{X}$  на  $\mathcal{Y}$  в виде условной сет-медианы.

Для нахождения сет-регрессии в виде сет-моды воспользуемся таблицей 6 и формулой (1.7). Введём функцию  $J : 2^{\mathcal{X}} \times 2^{\mathcal{Y}} \rightarrow \{0, 1\}$ ,  $\alpha \in [0, 1]$  значения

Таблица 9 - Матричное представление решения задачи в виде условной сет-медианы. Каждая строка таблицы является индикаторной функцией соответствующего условной сет-медианы

	$x_1$	$x_2$	$x_3$		$x_1$	$x_2$	$x_3$
ter( $\emptyset$ )	0	1	1	ter( $y_1y_3y_4$ )	1	1	1
ter( $y_1$ )	0	1	1	ter( $y_1y_2y_5$ )	0	1	0
ter( $y_2$ )	0	1	0	ter( $y_3y_5$ )	0	1	0
ter( $y_3$ )	1	0	1	ter( $y_1y_3y_5$ )	1	0	0
ter( $y_1y_2$ )	1	1	0	ter( $y_2y_3y_4$ )	1	0	0
ter( $y_4$ )	0	1	0	ter( $y_4y_5$ )	1	0	0
ter( $y_1y_3$ )	0	1	1	ter( $y_1y_2y_3y_4$ )	1	1	0
ter( $y_1y_4$ )	1	1	1	ter( $y_1y_4y_5$ )	0	0	0
ter( $y_2y_3$ )	1	0	1	ter( $y_2y_3y_5$ )	0	1	0
ter( $y_5$ )	0	1	0	ter( $y_1y_2y_3y_5$ )	1	1	1
ter( $y_2y_4$ )	0	1	0	ter( $y_2y_4y_5$ )	1	1	0
ter( $y_1y_2y_3$ )	0	0	0	ter( $y_1y_2y_4y_5$ )	0	1	1
ter( $y_1y_5$ )	0	1	1	ter( $y_3y_4y_5$ )	1	0	0
ter( $y_2y_5$ )	0	0	0	ter( $y_1y_3y_4y_5$ )	0	1	1
ter( $y_3y_4$ )	1	0	0	ter( $y_2y_3y_4y_5$ )	0	1	1
ter( $y_1y_2y_4$ )	0	0	0	ter( $y_1y_2y_3y_4y_5$ )	1	1	1

которой определяются по формуле

$$J(X, Y) = \begin{cases} 1, & \mathbf{P}(X) = \max_{Z \subseteq \mathfrak{X}} \mathbf{P}(Z | Y), \\ 0, & \text{иначе.} \end{cases} \quad (1.11)$$

Функция (1.11) фактически является индикатором множества, построенного по формуле (1.7), т.е. индикатором условной сет-моды. Решение задачи в виде условной сет-моды представим в виде таблицы 10, элементы которой находятся по формуле (1.11). График решения представлен на рис. 3.

Таблица 10 - Матричное представление решения задачи в виде условной сет-МОДЫ.

	ter( $\emptyset$ )	ter( $x_1$ )	ter( $x_2$ )	ter( $x_3$ )	ter( $x_1x_2$ )	ter( $x_1x_3$ )	ter( $x_2x_3$ )	ter( $x_1x_2x_3$ )
ter( $\emptyset$ )	0	1	0	0	0	0	0	0
ter( $y_1$ )	0	1	0	0	0	0	0	0
ter( $y_2$ )	0	0	0	1	0	0	0	0
ter( $y_3$ )	0	0	1	0	0	0	0	0
ter( $y_1y_2$ )	0	0	0	1	0	0	0	0
ter( $y_4$ )	0	0	0	0	0	1	0	0
ter( $y_1y_3$ )	1	0	0	0	0	0	0	0
ter( $y_1y_4$ )	1	0	0	0	0	0	0	0
ter( $y_2y_3$ )	1	0	0	0	0	0	0	0
ter( $y_5$ )	0	0	0	0	0	1	0	0
ter( $y_2y_4$ )	0	0	0	0	0	0	0	1
ter( $y_1y_2y_3$ )	0	0	0	0	0	0	1	0
ter( $y_1y_5$ )	0	1	0	0	0	0	0	0
ter( $y_2y_5$ )	0	0	0	0	0	1	0	0
ter( $y_1y_2y_3$ )	0	0	0	0	0	0	0	1
ter( $y_3y_4$ )	0	0	0	0	0	0	0	1
ter( $y_1y_2y_4$ )	1	0	0	0	0	0	0	0
ter( $y_1y_3y_4$ )	0	0	0	0	0	1	0	0
ter( $y_1y_2y_5$ )	0	0	0	0	0	1	0	0
ter( $y_3y_5$ )	0	0	0	0	0	0	1	0
ter( $y_1y_3y_5$ )	0	0	0	0	0	0	1	0
ter( $y_2y_3y_4$ )	0	0	0	1	0	0	0	0
ter( $y_4y_5$ )	0	0	0	1	0	0	0	0
ter( $y_1y_2y_3y_4$ )	0	0	0	0	1	0	0	0
ter( $y_1y_4y_5$ )	0	0	0	0	0	0	0	1
ter( $y_2y_3y_5$ )	0	0	0	0	0	0	0	1
ter( $y_1y_2y_3y_5$ )	0	1	0	0	0	0	0	0
ter( $y_2y_4y_5$ )	0	0	1	0	0		0	0
ter( $y_1y_2y_4y_5$ )	0	0	0	0	0	1	0	0
ter( $y_3y_4y_5$ )	0	0	0	0	1	0	0	0
ter( $y_1y_3y_4y_5$ )	0	1	0	0	0	0	0	0
ter( $y_2y_3y_4y_5$ )	0	0	0	0	0	0	0	1
ter( $y_1y_2y_3y_4y_5$ )	1	0	0	0	0	0	0	0

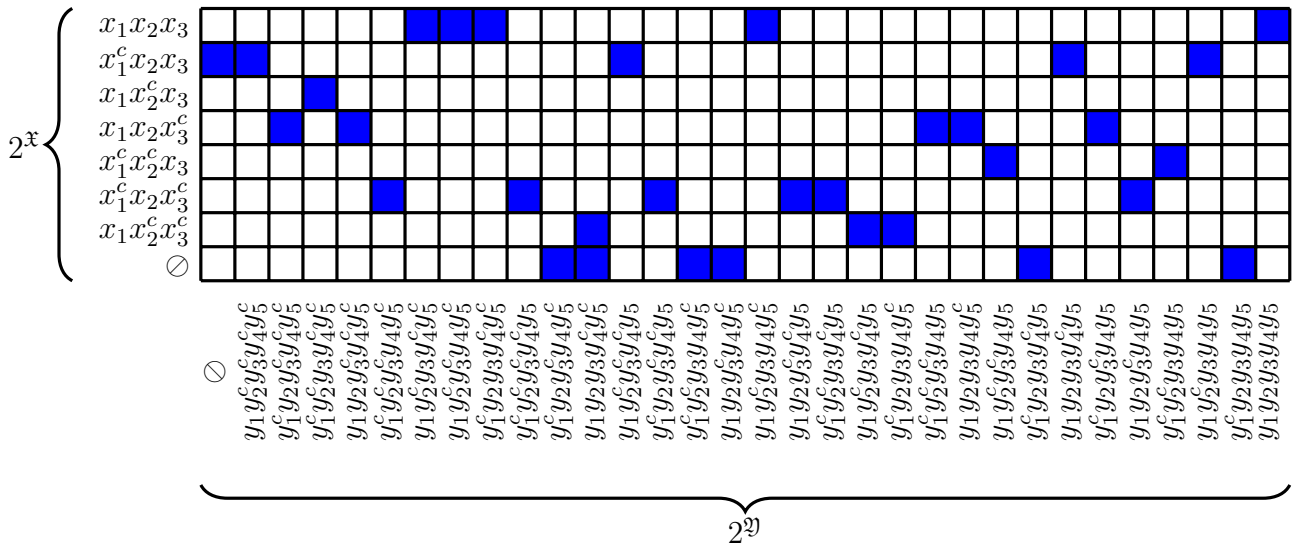


Рисунок 3 - Решение задачи сет-регрессии множества  $\mathfrak{X}$  на  $\mathfrak{Y}$  в виде условной сет-моды.

### 1.5 Суперпозиция сет-регрессий

**Задач суперпозиции сет-регрессий.** Пусть известны

- совместное распределение

$$\{p(X, Y) = \mathbf{P}(K = X, L = Y), X \subseteq \mathfrak{X}, Y \subseteq \mathfrak{Y}\}$$

двух случайных множеств  $K$  и  $L$ , значения которых содержатся в конечных множествах  $\mathfrak{X}$  и  $\mathfrak{Y}$  соответственно;

- совместное распределение  $\{p(Y, Z) = \mathbf{P}(L = Y, M = Z), Y \subseteq \mathfrak{Y}, Z \subseteq \mathfrak{Z}\}$  двух случайных множеств  $L$  и  $M$ , значения которых содержатся в конечных множествах  $\mathfrak{Y}$  и  $\mathfrak{Z}$  соответственно.

Требуется найти некоторую функцию множеств  $\Theta : 2^{\mathfrak{Z}} \rightarrow 2^{\mathfrak{X}}$ , описывающую зависимость между случайными множествами  $K$  и  $M$ .

Решение состоит из трех этапов:

1. На первом этапе необходимо найти решение задачи сет-регрессии для множеств  $\mathfrak{X}$  и  $\mathfrak{Y}$  в виде некоторого сет-среднего  $\Phi_{KL}(\{\{K = X\} |_{\{L=Y\}}\})$ ,  $X \subseteq \mathfrak{X}, Y \subseteq \mathfrak{Y}$ .

2. На втором этапе необходимо найти решение задачи сет-регрессии для множеств  $\mathfrak{Y}$  и  $\mathfrak{Z}$  в виде некоторого сет-среднего  $\Phi_{LM}(\{L = Y\} |_{\{M=Z\}})$ ,  $Y \subseteq \mathfrak{Y}$ ,  $Z \subseteq \mathfrak{Z}$ .

3. Решение задачи определяется как суперпозиция

$$\Psi(Z) = \Phi_{KM} \left( \{K = X\} \left| \{L = \Phi_{LM}(\{L=Y\} |_{\{M=Z\}}) \right. \right).$$

## Резюме

В первой главе проанализированы основополагающие определения теории случайных конечных множеств и их средних характеристик, изучены теоремы экстремальности сет-квантиля, сет-моды и сет-медианы. Приведены постановки задач сет-регрессии и суперпозиции сет-регрессий. Приведен числовой пример решения задачи сет-регрессии.



## **2 Алгоритмы и комплекс программ для сет-регрессионного анализа**

Проблемно-ориентированный программный комплекс «SetRegression» предназначен для обработки статистических данных нечисловой природы. Программа работает с базой данных совместного распределения двух множеств, записанных в файл в формате .csv, либо может автоматически генерировать псевдослучайное распределение по исходному разбиению входного множества на два подмножества на усмотрение пользователя. Мощность входного множества также указывается пользователем. Также пользователь волен самостоятельно выбрать тип сет-среднего из четырех возможных вариантов (сет-ожидание, сет-медиана, сет-квантиль, сет-мода) для получения соответствующего решения. Для первого запуска программы имеется ряд значений входных данных по умолчанию, что позволяет непросвещенному пользователю наглядно увидеть процесс работы программы и её конечный продукт — графики сет-регрессии и суперпозиции.

### **2.1 Сведения о программном комплексе «SetRegression»**

Основные особенности программы «SetRegression»:

- визуализация в реальном времени поэтапного решения задачи;
- экспорт полученных данных в документ Microsoft Excel;
- импорт исходных данных из Microsoft Excel;
- псевдослучайная генерация входных данных;
- табличное представление решения задачи на каждом этапе;
- таблично-графическое построение графика функции сет-регрессии и суперпозиции;

- ввод мощности входного множества;
- разбиение данного множества на два подмножества;
- ведение архива записей после каждой операции;
- редактирование импортированного/генерированного (псевдослучайно) значений вероятностей совместного распределения;
- возможность обработки любой нечисловой информации в виде таблицы совместного распределения.

Для эффективной работы авторы программного «SetRegression» рекомендуются иметь компьютер со следующими характеристиками:

- IBM совместимый персональный компьютер не ниже AMD A8-6410 APU with Radeon R5 Graphics (2.00 GHz);
- оперативная память не менее 8 ГБ;
- разрешение экрана  $1366 \times 768$ ;
- не менее 50 МБ дискового пространства;
- операционная система Windows 10 ( $\times 64$ ).

Для получение большей производительности следует акцентировать внимание и ресурсы на повышение значения оперативной памяти (более 8 ГБ).

В ходе вычислительных экспериментов конфликт при совместной работе программного комплекса на компьютере с другими программами в реальном времени не обнаружен.

## **2.2 Описание разработанных алгоритмов**

Для наглядности описание структуры основных модулей (алгоритмов) программы представлено на естественном языке, но частично формализован-

ном с использованием некоторых формальных конструкций и общепринятых математических обозначений. Работу программы условно можно разделить на этапы решения задачи.

1. Построение совместного распределения случайных конечных множеств (Алгоритм 1).
2. Получение набора условных распределений случайных конечных множеств (Алгоритм 2).
3. Нахождение условных вероятностей событий (Алгоритм 3).
4. Выбор вида сет-среднего случайного конечного множества (Алгоритм 4).
5. Вычисление матричного отношения и построение графика сет-функции (Алгоритм 5).
6. Построение графика суперпозиции сет-функций (Алгоритм 6).

Каждый этап решения сет-регрессионной задачи в комплексе программ «SetRegression» описаны как реакция на события нажатия на соответствующие кнопки. При запуске кнопки по умолчанию заблокированы, по мере прохождения очередного шага открывается доступ к следующей кнопке, кликнув которую пользователь запускает алгоритм для обработки ранее полученных данных. Структура алгоритмов, входные данные для обработки и результаты описаны ниже. Для удобства чтения и ясности каждая строка кода программы сопровождается подробными комментариями.

Процедура *построения совместного распределения* (алгоритм 1) предназначена для непосредственной обработки входных данных. На данном шаге по настройкам пользователя в интерфейсе программы процедура считывает выбор пользователя: мощность входного множества, правило разбиения его на два подмножества.

---

**Алгоритм 1.** Построение совместного распределения

---

**Вход:** контрольный лист CheckList с разбиением входного множества на два случайных подмножества  $K$  и  $L$ , значения которых содержатся в непересекающихся множествах  $\mathcal{X}, \mathcal{Y} \subset \mathcal{F}$ , соответственно, путем проставления галочек в соответствующих ячейках.

**Выход:** совместное распределение  $\{p(X, Y), X \subseteq \mathcal{X}, Y \subseteq \mathcal{Y}, \mathcal{X} \cap \mathcal{Y} = \emptyset\}$  случайных множеств  $K$  и  $L$ , значения которых содержатся в непересекающихся множествах  $\mathcal{X}, \mathcal{Y} \subset \mathcal{F}$ , соответственно, представленное в табличном виде.

1: Очистить mas1, mas2

▷ считать в mas1 и mas2 из контрольного листа CheckListOfSetElements

2: ReadToMass(self.mas1, self.mas2, self.CheckListOfSetElements)

3: Очистить TableOfCombinedDistribution

4: Очистить TableOfSetConditionalDistribution

▷ заполнить «шапку» таблицы из подмножеств mas1 и mas2

5: MakeSizeKL

▷ псевдослучайное заполнение таблицы совместного распределения

6: PrintRandomTbal(TableOfCombinedDistribution)

7: Разблокировать кнопку "Построить"

▷ 1 - код сообщения, История1 - блокнот для записи справа в окне программы

8: SaveMessage(1, Memo1)

---

Также при желании и наличии статистических данных в виде представленном как в таблице совместного распределения пользователь может их загрузить в программный комплекс «SetRegression», используя специальную кнопку из Блока 3 «Открыть», при нажатии ПКМ на данной кнопке откроется браузер, в котором необходимо найти адрес документа<sup>1</sup> на диске и дважды кликнуть по ней. Если таковых данных не имеется, то достаточно нажать на саму кнопку «Построить» и программа автоматически сгенерирует совместное распределение двух множеств, согласно разделению пользователя. Каждая выполненная операция фиксируется в блокноте «Мемо1», которая визуально расположена в Блоке 3. На странице 2 для нахождения функции сет-регрессии

---

<sup>1</sup>C:/ProgramFiles(×86)/.../db.csv, документ должен быть сохранен именно в формате Microsoft Exel .csv

есть аналогичные компоненты.

Для получения набора условных распределений (алгоритм 2) необходимо прежде посчитать маргинальное распределение по формуле (1.4). Далее, имея маргинальное распределение можно перейти к вычислению всех условных распределений случайного множества  $K$  при условии, что наступило подмножество событий  $Y \subseteq \mathfrak{Y}$  по формуле (1.3). Результаты вычислений записываются в таблицу набора условных распределений.

---

**Алгоритм 2.** Получение набора условных распределений

---

**Вход:** совместное распределение  $\{p(X, Y), X \subseteq \mathfrak{X}, Y \subseteq \mathfrak{Y}, \mathfrak{X} \cap \mathfrak{Y} = \emptyset\}$  случайных множеств  $K$  и  $L$ , значения которых содержатся в непересекающихся множествах  $\mathfrak{X}, \mathfrak{Y} \subset \mathcal{F}$ , соответственно, представленное в табличном виде.

**Выход:** Набор условных распределений из  $2^{\mathfrak{X}} \times 2^{\mathfrak{Y}}$  вероятностей

- 1: Добавить строку и столбец к таблице TableOfCombinedDistribution
    - ▷ считать сумму по строке и по столбцу указанной таблицы
  - 2: WriteSumColRow(TableOfCombinedDistribution)
  - 3: Очистить таблицы совместного и набора условных распределений
    - ▷ деление ячейки на сумму ячеек строки и запись в таблицу TableOfSetConditionalDistribution
  - 4: **Если** сумма элементов таблицы TableOfCombinedDistribution = 1 **тогда**
  - 5:     MakeStep2(TableOfCombinedDistribution, TableOfSetConditionalDistribution)
  - 6: **иначе**
  - 7:     ShowMessage('Сумма элементов таблицы совместного распределения '+' не равна 1')
  - 8: **Конец условия**
  - 9: SaveMessage(2, Memo1)
- 

*Нахождение условных вероятностей событий* (алгоритм 3) требует понимания классификации распределений по родам. В частности, в программе использованы распределения I-го и II-го родов. Для вычисления распределения II-го рода была использована формула обращения Мёбиуса.

---

**Алгоритм 3.** Нахождение условных вероятностей событий

---

**Вход:** Набор условных распределений из  $2^x \times 2^y$  вероятностей

**Выход:** Распределение II-го рода, состоящего из  $2^N$  вероятностей  $\{p_X, X \subseteq \mathfrak{X}\}$ .

- 1: скопировать нулевой столбец из таблицы TableOfCombinedDistribution
  - 2: TableOfConditionalProbabilitiesOfEvents.RowCount присвоить значение, равное TableOfSetConditionalDistribution.RowCount
  - 3: **Для**  $I := 0$  до TableOfSetConditionalDistribution.RowCount-1 **выполнять**
  - 4: TableOfConditionalProbabilitiesOfEvents.Cells[0,i] присвоить значение, равное TableOfSetConditionalDistribution.Cells[0,i]
  - 5: **Конец цикла**
  - 6:  $c:=0$ ; ▷ количество "негалочек" в CheckListOfSetElements
  - 7: Очистить mass
  - 8: **Для**  $I := 0$  до strtoint(CapacityOfSet.Text)-1 **выполнять**
  - 9: **Если** CheckListOfSetElements.Checked[i] **тогда** inc(c)
  - 10: **иначе**
  - 11:     setlength(mass,размерность(mass)+1)
  - 12:     mass[последний(mass)]:=i+1
  - 13: **Конец условия**
  - 14: **Конец цикла**
  - 15: TableOfConditionalProbabilitiesOfEvents.ColCount:=размерность(mass)+1;
  - 16: **Для**  $I := 1$  до  $c+1$  **выполнять** ▷ записали одноименные в шапку столбцов
  - 17: TableOfConditionalProbabilitiesOfEvents.Cells[i,0]:='(+inttostr(mass[i-1]))+'
  - 18: Очистить mass
  - 19: **Конец цикла**
  - 20: Очистить таблицу TableOfConditionalProbabilitiesOfEvents
  - 21: **Для**  $I := 1$  до TableOfConditionalProbabilitiesOfEvents.RowCount-1 **выполнять**
  - 22: **Для** от  $J := 1$  до TableOfConditionalProbabilitiesOfEvents.ColCount-1 **выполнять**
  - 23: TableOfConditionalProbabilitiesOfEvents.Cells[j,i]:=floattostr( CalcIIRoda(i,j, TableOfConditionalProbabilitiesOfEvents, TableOfSetConditionalDistribution) )
  - 24: **Конец цикла**
  - 25: **Конец цикла**
  - 26: SaveMessage(3,Мемо1)
-

В ходе работы программы пользователь может вносить корректировки или вовсе изменить входные данные, загрузив их из файла или случайно сгенерировав, кроме того программа позволяет исследовать результат в зависимости от средней характеристики, которое можно задать используя вышеописанный модуль *выбор вида сет-среднего случайного множества* (алгоритм 4).

---

**Алгоритм 4.** Выбор вида сет-среднего случайного множества

---

**Вход:** Выбор пользователем правила определения средних характеристик (сет-средние) случайного множества. По умолчанию сет-среднее задано как сет-медиана.

**Выход:** Считывание и настройка выбора типа сет-средних для дальнейшей обработки.

- 1: скопировать нулевой столбец из таблицы TableOfCombinedDistribution
  - 2: **Если** ListOfValueMiddleSet.ItemIndex = 0 **тогда** ▷ сет-ожидание
  - 3:     ComboBox3.ItemIndex:=0
  - 4: **Конец условия**
  - 5: **Если** ListOfValueMiddleSet.ItemIndex = 1 **тогда** ▷ сет-квантиль
  - 6:     ComboBox3.ItemIndex:=1 и ValueOfMiddleSet.Enabled:=Истина
  - 7:     ValueOfMiddleSet.Text:='0' и LabeledEdit2.Text:='0'
  - 8: **Конец условия**
  - 9: **Если** ListOfValueMiddleSet.ItemIndex = 2 **тогда** ▷ сет-мода
  - 10:     ComboBox3.ItemIndex:=2 и ValueOfMiddleSet.Enabled:=Ложь
  - 11:     ValueOfMiddleSet.Text:='' и LabeledEdit2.Text:=''
  - 12: **Конец условия**
  - 13: **Если** ListOfValueMiddleSet.ItemIndex = 3 **тогда** ▷ сет-медиана
  - 14:     ComboBox3.ItemIndex:=3 и ValueOfMiddleSet.Enabled:=Ложь
  - 15:     ValueOfMiddleSet.Text:='0.5' и LabeledEdit2.Text:='0.5'
  - 16: **Конец условия**
  - 17: SaveMessage(4,Memo1)
  - 18: SaveMessage(4,Memo2)
- 

Модуль *вычисления матричного отношения и построения графика сет-функции* (алгоритм 5) является одной из наиболее объемных частей программного кода. Данный алгоритм позволяет продемонстрировать пользователю фи-





*Построение графика суперпозиции* (алгоритм 6) исследовала Иванова А.И. в своих работах посвященных данной тематике, ей же принадлежит и постановка задачи. В результате продуктивной совместной работы автора и Ивановой А.И. разработаны и запрограммированы алгоритмы решения двух задач сет-регрессионного анализа: задача сет-регрессии и задача суперпозиции. Имеются совместные публикации.

---

**Алгоритм 6.** Построение графика суперпозиции

---

**Вход:** Графики функций сет-регрессий  $L$  на  $K$  и  $M$  на  $L$ .

**Выход:** График суперпозиции  $M$  на  $K$ .

1: Очистить таблицу StringGrid3

2: Скопировать «шапку» из StringGrid2 и StringGrid1 в StringGrid3

▷ обход для суперпозиции

3: **Для**  $I := 1$  до StringGrid1.ColCount-1 **выполнять**

4:     **Для**  $J := 1$  до StringGrid1.RowCount-1 **выполнять**

5:         **Если** StringGrid1.Cells[i,j]<>” **тогда**

6:              $a:=j$

7:             ВыйтиИзЦикла

8:         **Конец условия**

9:     **Конец цикла**

10:     **Для**  $K := 1$  до StringGrid2.RowCount-1 **выполнять**

11:         **Если** StringGrid2.Cells[a,k]<>” **тогда**

12:              $b:=k$

13:             ВыйтиИзЦикла

14:         **Конец условия**

15:     **Конец цикла**

16:     StringGrid3.Cells[i,b]:=’■’

17: **Конец цикла**

18: SaveMessage(9,Мемо1)

19: SaveMessage(9,Мемо2)

---

### 2.3 Анализ разработанных алгоритмов

Для анализа алгоритмов были проведены ряд вычислительных экспериментов. В ходе эксперимента варьировалась мощность входных данных и с помощью специальных программных средств производились измерения временных затрат и использованной оперативной памяти для выполнения конкретного программного модуля комплекса «SetRegression». Для чистоты (предсказуемости) эксперимента мощность входного множества задавалась числом, кратным трем, чтоб мощности трех подмножеств были равны между собой.

Исследование производительной сложности проводилось поэтапно, то есть запускался Алгоритм 1 — производились измерения по времени и занимаемой оперативной памяти, далее запускался следом Алгоритм 2, для которого выполнялась аналогичная процедура и таким образом проведен анализ разработанных алгоритмических модулей. В результате вычислительного эксперимента получена следующая статистика:

1. Алгоритм 1. Построения совместного распределения при вычислительном эксперименте показали экспоненциальную зависимость времени обработки и занимаемой памяти при этом от мощности входного множества. При мощности исходного множества равной 6 время обработки по заданному алгоритму составила 0.28 секунд, занимаемая память - практически не изменилась, но при выполнении алгоритма уже на множестве мощностью 24 элемента занимается вся свободная оперативная память (2ГБ), а время обработки составила 65.5 секунд.
2. Алгоритм 2. Получения набора условных распределений показал аналогичные Алгоритму 1 результаты с небольшим расхождением: при мощности 6 элементов — 0.32 секунды и 1.5 МБ памяти, при мощности 22 элемента — 18.13 секунд и 313.3 МБ памяти ОЗУ.

3. Алгоритм 3. Нахождения условных вероятностей событий также имеет экспоненциальный рост времени обработки относительно входных данных. Детальная оценка приведена в таблице эффективности алгоритмов (11).
4. Алгоритм 4. Задание правила для сет-среднего играет ключевую роль при построении сет-регрессии двух множеств и суперпозиции двух сет-регрессий, хотя сама процедура совсем не трудоемкая. Итого: время обработки — 0.01 секунда, на занимаемую оперативную память не влияет, так как параметры характеристики множества объявлены при запуске программного обеспечения.
5. Алгоритм 5. Вычисление матричного отношения и построение графика сет-функции можно классифицировать как алгоритм экспоненциальной сложности. При мощности 6 элементов потребовалось 0.22 секунды и 1.5 МБ ОЗУ для выполнения строго определенной последовательности действий запрограммированных данным модуль. Если увеличить мощность множества до 22, то производительность составит 56.31 секунд и 649.8 МБ оперативной памяти.
6. Алгоритм 6. Построение графика суперпозиции продемонстрировал неожиданные результаты, несмотря на сложную структуру модуля. Возможно, это связано с длиной входных данных. Сложность алгоритма 6 была оценена как полиномиальная по занимаемой памяти и квазиполиномиальную сложность по времени обработки.

Детальный отчет по проведенному вычислительному эксперименту представлен в таблице (11) в следующем параграфе.

## Таблица эффективности алгоритмов

Таблица 11 - Таблица эффективности разработанных алгоритмов.

Алгоритмы \ Мощность	6	12	18	20	22
Алгоритм 1.	0.28/1.5	0.31/1.9	1.59/22.1	5.34/80.4	18.21/313.3
Алгоритм 2.	0.32/1.5	0.25/2.0	1.93/41.8	5.62/160.7	18.13/637.2
Алгоритм 3.	0.19/1.5	0.25/2.1	1.94/42.3	8.4/161.7	33.12/639.3
Алгоритм 4.	0.01/1.5	0.01/2.1	0.01/42.3	0.01/161.7	0.01/639.3
Алгоритм 5.	0.22/1.5	0.31/2.2	2.34/44.8	13.22/166.6	56.31/649.8
Алгоритм 6.	0.19/1.5	0.28/2.2	0.53/44.8	1.24/168.5	1.88/654.6

\*каждая ячейка полученной таблицы хранит запись о времени обработки данных в секундах и о занимаемой при этом объеме оперативной памяти в МБ. Формат записи: время(сек) / объем ОЗУ(МБ).

Рассмотрим графики функций сложности алгоритмов на примере Алгоритма 2 и Алгоритма 5. Так как длина входных данных есть дискретная величина, то графики представляют с собой последовательность точек на плоскости, соединенных прямыми отрезками. Для наглядности совокупность этих отрезков представляющих некоторый маршрут на плоскости был сглажен с целью показать возможное «поведение» функции при непрерывных входных данных. Автор понимает невозможность задать длину входных данных (мощность входных множеств), представленных в программе, недискретными значениями. Исходя из полученной статистики составлены следующие графики функций, характеризующие зависимость времени обработки (ось времени обозначена «Т») и занимаемой оперативной памяти (ось оперативной памяти обозначена «V») для выполнения алгоритма от длины входа (ось длины обозначена «N»).

Имея объемный теоретический базис программное комплекс для решения задачи случайно-множественного регрессионного анализа «SetRegression» выполняет также ряд важных прикладных задач, связанных с обработкой ста-

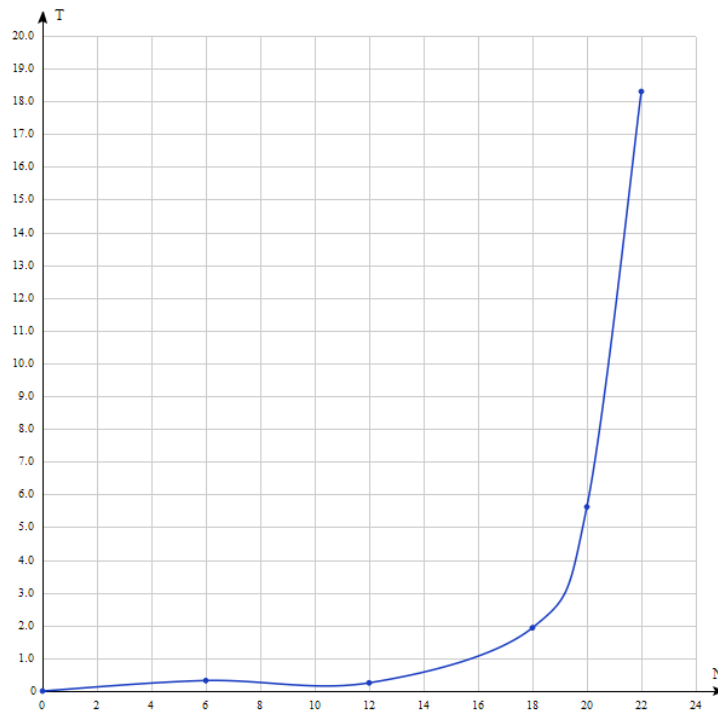


Рисунок 4 - График сложности алгоритма 2. Получение набора условных вероятностей (по времени)

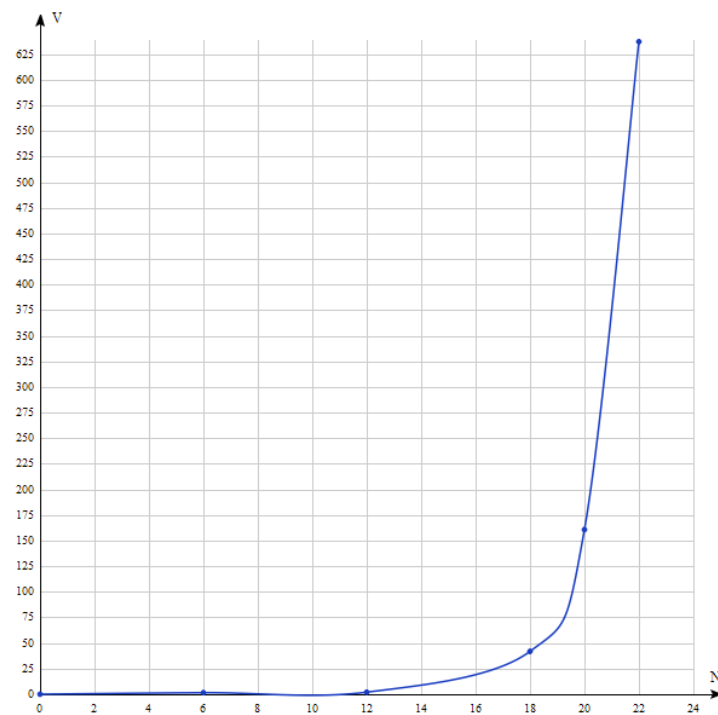


Рисунок 5 - График сложности алгоритма 2. Получение набора условных вероятностей (по памяти)

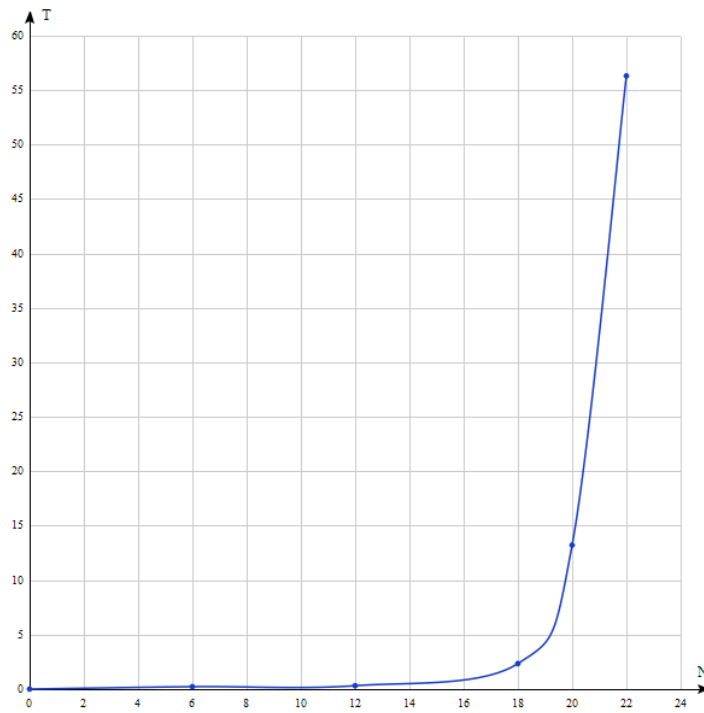


Рисунок 6 - График сложности алгоритма 5. Вычисление матричного отношения и построение графика сет-функции (по времени)

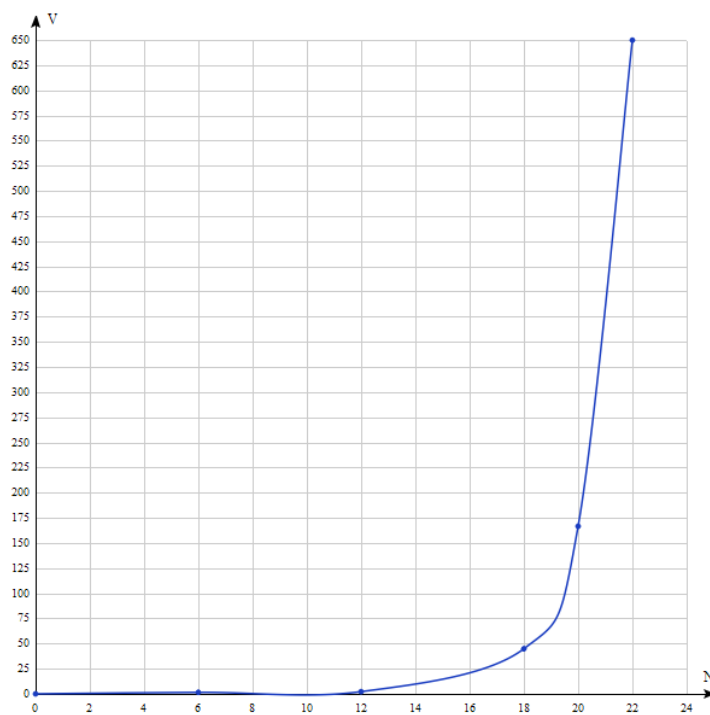


Рисунок 7 - График сложности алгоритма 5. Вычисление матричного отношения и построение графика сет-функции (по памяти)

статистических данных. Программа может эффективно работать на компьютерах меньшей производительности. Естественно, это повлияет на ряд характеристик, но существенного вреда результату не принесет. Для хранения программы на диске достаточно иметь 11.3 МБ свободной памяти. Запуск программы в состоянии покоя займет 1.5 МБ оперативной памяти вашего компьютера. Программа на суперкомпьютерах не тестировалась. Автор не исключает возможности достижения большей производительности разработанного программного продукта.

## **Резюме**

В данной главе описаны основные алгоритмы для обработки входных данных и решения поставленной задачи, также приведен список требований к ресурсам компьютера для получения аналогичных результатов. Даны рекомендации автора для наиболее эффективного использования программного комплекса «SetRegression» при обработке статистических данных нечисловой природы. Программное обеспечение для решения задачи сет-регрессионного анализа использует в качестве основного ресурса компьютера оперативную память, соответственно, чем больше объем оперативной памяти, тем больше может быть длина входных данных. На данный момент аналогов разработанной программе нет. Для оценки эффективности в главе приведена таблица эффективности запрограммированных алгоритмов. Скриншоты работы программы и пример решения задачи сет-регрессионного анализа на абстрактном примере приведены в главе 3. В Приложении А представлен исходный код программного комплекса «SetRegression».

### **3 Пример решения задачи построения сет-регрессии для анализа медицинских данных**

Теоретическое обоснование решения задачи суперпозиции сет-регрессий описано в первой главе. Во второй главе подробно рассмотрены разработанные алгоритмы и необходимые технические ресурсы для эксплуатации программного комплекса. В тексте главы 3 практически отсутствуют формулы и определения, так как они ранее были описаны. При решении практической задачи на основе статистической базы данных, полученной в процессе обработки историй болезни пациентов Красноярских краевых противотуберкулёзных диспансеров №1 и №2, акцентируется внимание именно на программное решение задачи случайно-множественного регрессионного анализа и на полученные сет-функции. При использовании тех или иных обозначений будут даны ссылки на формулы для их вычисления, приведенных в предыдущих главах. В процессе программирования неоднократно ставились вопросы: Как визуально представить работу программы? Как сделать удобный и функциональный интерфейс?

Отвечая на эти вопросы каждый раз интерфейс программного комплекса редактировался и дополнялся новыми компонентами. При разработке ПО на языке программирования Object Pascal в среде разработки приложений Delphi RAD Studio 10 Seattle был расширен список использованных библиотек [17].

#### **3.1 Программное решение задачи суперпозиции сет-регрессий**

Данный параграф приведен с целью ознакомления с интерфейсом программного комплекса «SetRegression» и основными этапами программного решения задачи суперпозиции сет-регрессий на абстрактном примере малой размерности.



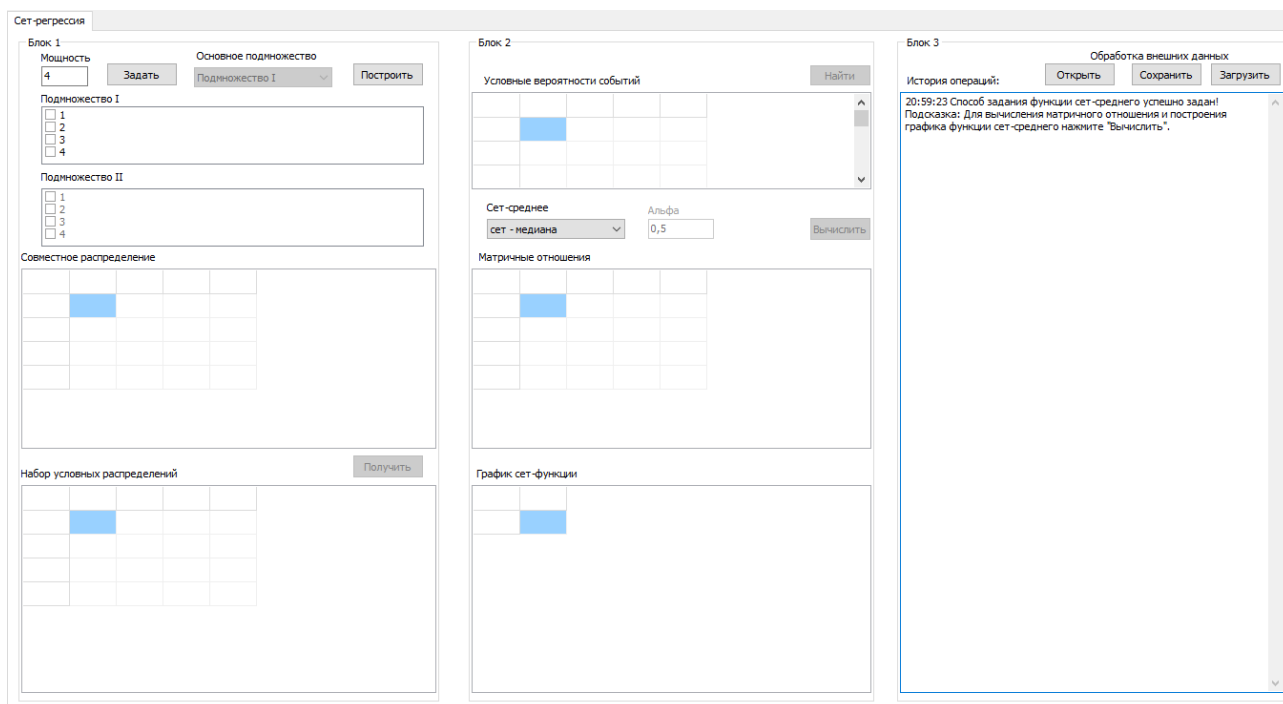


Рисунок 8 - Интерфейс программного комплекса SetRegression

В качестве примера обработки данных малой размерности на вход подадим множество  $\mathcal{M}$ ,  $|\mathcal{M}| = 4$ . Сохраним длину входа кнопкой «Задать» и разделим исходное множество на два множества  $\mathcal{M}_1$  и  $\mathcal{M}_2$ . Далее нажатием кнопки «Построить» генерируем псевдослучайные значения совместного распределения данных множеств. На третьем шаге необходимо кликнуть «Получить» для вычисления условного распределения. Шаг 4 «Найти» условные вероятности событий. На данном этапе пользователь выбирает сет-среднее из выпадающего списка, как сет-медиану, сет-квантиль, сет-моду или сет-ожидание, и кнопкой «Вычислить» составляется матрицу отношения и строится график сет-функции. Далее на появившейся новой странице пользователь проходит аналогичные шаги. На финальном этапе кнопка «Построить» определяет вид суперпозиции сет-регрессий. Результат построения модели решения задачи случайно-множественного регрессионного анализа на примере малой размерности приведен на следующей странице. Вид сет-среднего выбрана как сет-медиана.



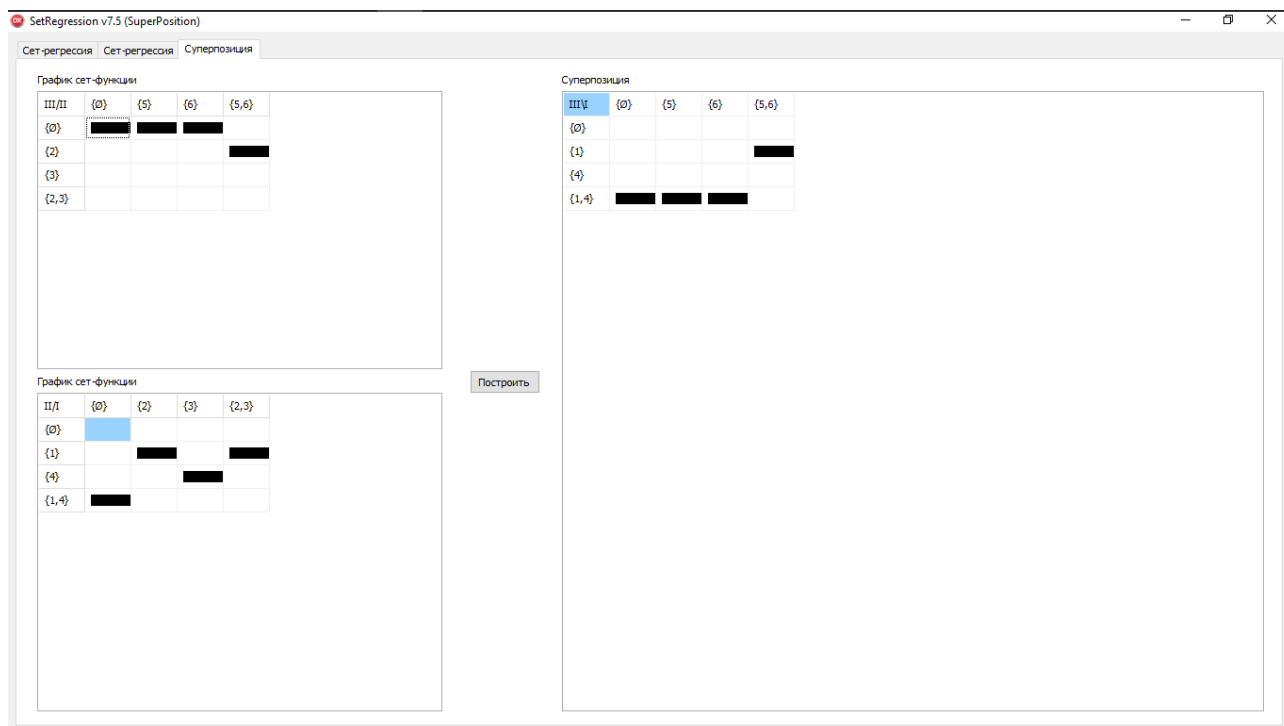


Рисунок 11 - Суперпозиции сет-регрессий

### 3.2 Предобработка исходной статистики для анализа медицинских данных

В ходе выполнения бакалаврской работы при высококвалифицированной поддержке научного руководителя совместно с кафедрой медицинской кибернетики Красноярского государственного медицинского университета им. профессора В.Ф. Войно-Ясенецкого проведены ряд мероприятий, направленных на получение прикладных результатов применения аппарата математической статистики и разработанного программного комплекса «SetRegression», выполнена обработка предоставленных данных краевым государственным бюджетным учреждением здравоохранения «Красноярского краевого противотуберкулёзного диспансера №1» и «Красноярского краевого противотуберкулёзного диспансера №2», являющихся организационно-методическими и лечебно-профилактическими центрами борьбы с туберкулёзом в Красноярском крае, для которых актуальна проблема прогнозирования и оценки возможных «очагов» тяжелого заболе-

вания, уносящего более полутора миллиона детских и взрослых жизней по всему миру. Данными учреждениями была предоставлена статистическая база данных историй болезней пациентов за 2012-2013 годы в формате электронного документа Microsoft Excel. В составленной базе имелась подробная информация о каждом пациенте, включающие анкетные как числовой («Дети в семье», «Заработная плата»), так и данные нечисловой природы («Женат/Замужем», «Причина переезда»). За указанный период учреждениями были анкетированы 728 пациентов. Автор уважает право на неприкосновенность личной жизни, в связи с этим часть данных засекречена или вовсе не учитывалась при решении задачи сет-регрессионного анализа.

По имеющейся медицинской статистике необходимо произвести вычислительные расчеты и выявить закономерности с применением разработанных алгоритмов и программного обеспечения, в качестве результата работы предоставить подробный отчет об использованной научно-теоретической основе разработанного приложения и словесные комментарии к обработанным данным и полученному графику сет-функции, имеющие прикладное значение. В качестве входных данных была использована база данных пациентов.

№ п/п	№ выд.	Статус (Болен-Умерев)	Возраст	Пол	Отчество	Фамилия	1-2 сетев.	Род	Солес. до	Раса	Вакцинозащита									
690	303	303	#####	#####	#####	Диссеминированная	0	0	1	0	1	0	0	0	0	0	0	0	0	0
691	304	304	#####	#####	#####	Диссеминированная	0	0	1	0	1	0	0	0	0	0	0	0	0	0
692	305	305	#####	#####	#####	Инфильтративная	1	0	0	1	1	0	0	0	0	0	0	0	0	0
693	306	301	#####	#####	#####	Инфильтративная	1	0	0	1	1	0	0	0	0	0	0	0	0	0
694	307	306	#####	#####	#####	Инфильтративная	1	0	0	0	0	1	0	0	0	0	0	0	0	0
695	308	313	#####	#####	#####	Инфильтративная	1	0	0	0	0	0	0	0	0	0	0	0	0	0
696	309	314	#####	#####	#####	Инфильтративная	1	0	0	0	0	0	0	0	0	0	0	0	0	0
697	310	314	#####	#####	#####	Диссеминированная	0	0	1	1	1	0	0	0	0	0	0	0	0	0
698	311	311	#####	#####	#####	Инфильтративная	1	0	0	0	0	0	0	0	0	0	0	0	0	0
699	312	306	#####	#####	#####	Инфильтративная	1	0	0	0	0	1	0	0	0	0	0	0	0	0
700	313	312	#####	#####	#####	Диссеминированная	0	0	1	1	1	0	0	0	0	0	0	0	0	0
701	314	333	#####	#####	#####	Средней	1	0	0	0	0	1	0	0	0	0	0	0	0	0
702	315	336	#####	#####	#####	Диссеминированная	0	0	1	1	1	1	0	0	0	0	0	0	0	0
703	316	332	#####	#####	#####	Диссеминированная	0	0	1	1	1	1	0	0	0	0	0	0	0	0
704	317	334	#####	#####	#####	Диссеминированная	0	0	1	1	1	1	0	0	0	0	0	0	0	0
705	318	318	#####	#####	#####	Диссеминированная	0	0	1	1	1	1	0	0	0	0	0	0	0	0
706	319	340	#####	#####	#####	Инфильтративная	1	0	0	0	0	0	0	0	0	0	0	0	0	0
707	320	341	#####	#####	#####	Инфильтративная	1	0	0	0	0	0	0	0	0	0	0	0	0	0
708	321	342	#####	#####	#####	Диссеминированная	0	0	1	1	1	1	0	0	0	0	0	0	0	0
709	322	339	#####	#####	#####	Диссеминированная	0	0	1	1	1	1	1	0	0	0	0	0	0	0
710	323	304	#####	#####	#####	Диссеминированная	0	0	1	1	1	1	0	0	0	0	0	0	0	0
711	324	303	#####	#####	#####	Инфильтративная	0	1	0	1	1	1	0	0	0	0	0	0	0	0
712	325	302	#####	#####	#####	Инфильтративная	1	0	0	0	0	0	0	0	0	0	0	0	0	0
713	326	311	#####	#####	#####	Диссеминированная	0	0	1	0	0	0	0	0	0	0	0	0	0	0
714	327	330	#####	#####	#####	Инфильтративная	1	0	0	0	0	0	0	0	0	0	0	0	0	0
715	328	331	#####	#####	#####	Диссеминированная	0	0	1	0	0	0	0	0	0	0	0	0	0	0
716	329	316	#####	#####	#####	Диссеминированная	0	0	1	1	1	1	0	0	0	0	0	0	0	0
717	330	317	#####	#####	#####	Инфильтративная	1	0	0	1	1	1	0	0	0	0	0	0	0	0
718	331	308	#####	#####	#####	Инфильтративная	0	0	0	0	0	0	0	0	0	0	0	0	0	0
719	332	327	#####	#####	#####	Диссеминированная	0	0	1	1	1	1	0	0	0	0	0	0	0	0
720	333	310	#####	#####	#####	Диссеминированная	0	0	1	1	1	1	0	0	0	0	0	0	0	0
721	334	326	#####	#####	#####	Инфильтративная	1	0	0	0	0	0	0	0	0	0	0	0	0	0
722	335	305	#####	#####	#####	Инфильтративная	1	0	0	0	0	0	0	0	0	0	0	0	0	0
723	336	337	#####	#####	#####	Инфильтративная	1	0	0	0	0	0	0	0	0	0	0	0	0	0
724	337	309	#####	#####	#####	Инфильтративная	0	0	1	1	1	1	0	0	0	0	0	0	0	0
725	338	315	#####	#####	#####	Диссеминированная	0	0	1	0	0	0	0	0	0	0	0	0	0	0
726	339	308	#####	#####	#####	Диссеминированная	0	0	1	1	1	1	0	0	0	0	0	0	0	0
727	340	319	#####	#####	#####	Инфильтративная	1	0	0	0	0	0	0	0	0	0	0	0	0	0
728	341	307	#####	#####	#####	Инфильтративная	0	1	0	1	1	1	0	0	0	0	0	0	0	0
729	342	329	#####	#####	#####	Инфильтративная	1	0	0	1	1	1	0	0	0	0	0	0	0	0

Рисунок 12 - Фрагмент исходных данные в формате Microsoft Excel.

Для выполнения поставленного задания проанализированы данные пациентов, которые удалось условно разделить в три множества данных. Множество 1 описывает *социальные характеристики*, во втором множестве собран *образ жизни*, а третье множество представляет выявленные *заболевания* у пациентов. Для удобства введем следующие обозначения множеств и информации, содержащиеся в них.

Множество 1. *Социальные характеристики:*

- Пол
- Город
- Мед.работник
- Дети в семье
- Судимость
- Не одинок.

Множество 2. *Образ жизни:*

- Частое переохлаждение
- Тяжелый физический труд
- Нервно-психологическая нагрузка
- Алкоголизм
- Наркомания
- Табакокурение.

Множество 3. *Заболевания:*

- ВИЧ

- Сахарный диабет
- ЯБЖиДПК
- ЖКТ
- Психологические заболевания
- Рак легкого.

Для кластеризации по множествам 1 - 3 разработан дополнительный модуль для считывания исходных данных из таблицы и формирования вероятностного распределения I рода. После выполнения считывания и обработки, данные сохраняются на диск в файл с расширением «.csv». Данные в файле структурированы для удобства дальнейшей обработки. Код запрограммированного модуля для предобработки приведен в приложении после кода основной программы. Обработанная статистика примет следующий вид (см. Рисунок (13)-(15)) :

```

C:\AAA\univet1_Дипломная работа\Программы\Win32\Debug\test2\Win32\Debug\test2.exe
Социальные характеристики:
Пол | Город | Мест.работы | Дети в семье | Судим | Не однок |
>> Данные успешно загружены в программу!
-----
Случаи   Пациентов   из 100%
-----
110001   (120)       17.31 %
100000   (74)        10.45 %
010001   (109)       14.97 %
000010   (3)         0.41 %
010000   (110)       15.48 %
000000   (41)        5.63 %
110011   (9)         1.24 %
110000   (66)        9.07 %
100010   (4)         0.55 %
100001   (49)        6.73 %
000001   (58)        7.97 %
110010   (21)        2.88 %
010101   (5)         0.69 %
011001   (5)         0.69 %
011000   (3)         0.41 %
001001   (3)         0.41 %
100011   (10)        1.37 %
110110   (1)         0.14 %
101011   (1)         0.14 %
100111   (1)         0.14 %
110111   (2)         0.27 %
100100   (1)         0.14 %
010011   (2)         0.27 %
100101   (1)         0.14 %
101000   (1)         0.14 %
001000   (2)         0.27 %
000011   (2)         0.27 %
010010   (5)         0.69 %
000100   (2)         0.27 %
110100   (1)         0.14 %
-----
Всего: 728 пациентов
-----
>> Расчетка данных выполнена успешно!
  
```

Рисунок 13 - Результат предобработки множества «Социальная характеристика».

```

C:\AA\univer\1_Дипломная работа\Программа\Win32\Debug\test2\Win32\Debug\test2.exe
* * *
Образ жизни:
Частое переохлаждение | Тяжелый физ. труд | Нервно-психическая нагрузка | Алкоголизм | Наркомания | Табакокурение |
>> Данные успешно загружены в программу!
=====
Случаи Пациентов из 100%
-----
110001 (120) 16.48 %
000001 (159) 21.84 %
000000 (213) 29.26 %
001000 (133) 1.79 %
101001 (2) 0.27 %
010001 (16) 2.20 %
101000 (1) 0.14 %
001001 (7) 0.96 %
010000 (16) 2.20 %
111001 (20) 2.75 %
011001 (5) 0.69 %
110011 (12) 1.65 %
100001 (50) 6.87 %
110000 (56) 7.69 %
100000 (16) 2.20 %
111011 (4) 0.55 %
100101 (2) 0.27 %
110111 (1) 0.14 %
100011 (3) 0.41 %
001011 (2) 0.27 %
111011 (1) 0.14 %
111000 (1) 0.14 %
010002 (3) 0.41 %
110002 (1) 0.14 %
000002 (1) 0.14 %
000011 (3) 0.41 %
-----
Всего: 728 пациентов
=====
>> Распечатка данных выполнена успешно!

```

Рисунок 14 - Результат предобработки множества «Образ жизни»

```

C:\AA\univer\1_Дипломная работа\Программа\Win32\Debug\test2\Win32\Debug\test2.exe
* * *
Заболевания:
ВИЧ | СД | ЯБЖидЖ | ЖТ | Псих.заболевания | Рак легкого |
>> Данные успешно загружены в программу!
=====
Случаи Пациентов из 100%
-----
000000 (469) 64.42 %
001000 (28) 3.85 %
000100 (114) 15.66 %
100100 (10) 1.37 %
010100 (15) 2.06 %
011100 (4) 0.55 %
100000 (28) 3.85 %
010000 (36) 4.95 %
001100 (19) 2.61 %
000010 (1) 0.14 %
000001 (1) 0.14 %
011000 (2) 0.27 %
101000 (1) 0.14 %
-----
Всего: 728 пациентов
=====
>> Распечатка данных выполнена успешно!

```

Рисунок 15 - Результат предобработки множества «Заболевания»

На следующем шаге составляется совместное вероятностное распределение множеств «Социальная характеристика» и «Образ жизни», множеств «Образ жизни» и «Заболевания». Используя специальный алгоритм описанный в дополнительном модуле полученное вероятностное совместное распределение данных множеств записывается в файл с расширением «.scv» для дальнейшей обработки в основном программном комплексе «SetRegression».

Социальная характеристика и образ жизни:  
 Пол | Город | Мед.работник | Дети в семье | Судим | Не одинок | Частое переохлаждение | Тяжелый физ.труд | Нервно-психическая нагрузка | Алкоголизм | Наркомания | Табачокурение |

» Данные успешно загружены в программу!

Случаи	Пациентов	из 100%
11000110001	(29)	3.98 %
11000100000	(29)	3.98 %
11000100000	(25)	3.43 %
10000010001	(20)	2.75 %
01000100100	(1)	0.14 %
00000101001	(1)	0.14 %
01000000000	(54)	7.42 %
00000000001	(12)	1.65 %
01000100000	(50)	7.07 %
11001110001	(4)	0.55 %
11000101000	(1)	0.14 %
01000010100	(1)	0.14 %
01000000001	(19)	2.61 %
11001100100	(1)	0.14 %
11000001000	(2)	0.27 %
11000101000	(4)	0.55 %
11000000000	(18)	2.47 %
11000000001	(14)	1.92 %
10001010001	(3)	0.41 %
10000111001	(5)	0.69 %
01000000100	(7)	0.96 %
01000110001	(5)	0.69 %
01000100001	(23)	3.16 %
00000110001	(9)	1.24 %
00000100000	(13)	1.79 %
10000001000	(2)	0.27 %
11000011001	(1)	0.14 %
11001011001	(1)	0.14 %
01000011000	(10)	1.37 %
11000110001	(9)	1.24 %
01010100001	(2)	0.27 %
01000110000	(9)	1.24 %
11000011000	(8)	1.10 %
10000000000	(16)	2.20 %
11000110000	(4)	0.55 %
10000000001	(15)	2.06 %
11001110000	(1)	0.14 %

Рисунок 16 - Фрагмент предобработки совместного распределения множеств «Социальная характеристика» и «Образ жизни»

Образ жизни и Заболевания:  
 Частое переохлаждение | Тяжелый физ.труд | Нервно-психическая нагрузка | Алкоголизм | Наркомания | Табачокурение | ВИЧ | СД | ЯБЖДПК | ЖТ | Псих.заболевания | Рак легкого |

» Данные успешно загружены в программу!

Случаи	Пациентов	из 100%
11000100000	(74)	10.16 %
00000100100	(7)	0.96 %
00000000000	(143)	19.64 %
00100000000	(9)	1.24 %
10100100000	(2)	0.27 %
00000100000	(101)	13.87 %
00000100010	(31)	4.26 %
11000110010	(1)	0.14 %
11000100100	(19)	2.61 %
00000000100	(39)	5.36 %
01000100000	(12)	1.65 %
10100000000	(1)	0.14 %
00100100000	(6)	0.82 %
01000000000	(12)	1.65 %
11100100000	(13)	1.79 %
00000001010	(5)	0.69 %
01100100010	(1)	0.14 %
11000101100	(2)	0.27 %
11001100000	(3)	0.41 %
00000100000	(32)	4.40 %
00000101010	(2)	0.41 %
00000001000	(15)	2.06 %
11000000000	(37)	5.08 %
11000100010	(13)	1.79 %
10000000000	(9)	1.24 %
10000100010	(5)	0.69 %
01100101000	(1)	0.14 %
11001100000	(7)	0.96 %
11001100100	(1)	0.14 %
00000001000	(4)	0.55 %
11100100100	(3)	0.41 %
00100010000	(1)	0.14 %
11010100000	(1)	0.14 %
10010100000	(1)	0.14 %
11000101000	(5)	0.69 %

Рисунок 17 - Фрагмент предобработки совместного распределения множеств «Образ жизни» и «Заболевания»

### 3.3 Апробация статистики медицинского исследования в программном комплексе «SetRegression»

Предобработанные данные хранятся на диске в виде таблицы совместного распределения удобном для программной обработки и демонстрации решения задачи случайно-множественного сет-регрессионного анализа медицинских данных в программном комплексе «SetRegression» приведены в продол-



жении данной главы. В параграфе рассмотрены этапы решение задачи.

Шаг 1. Запустить можно программу двойным нажатием на иконку на рабочем столе. На мониторе компьютера появится интерфейс программного комплекса «SetRegression».

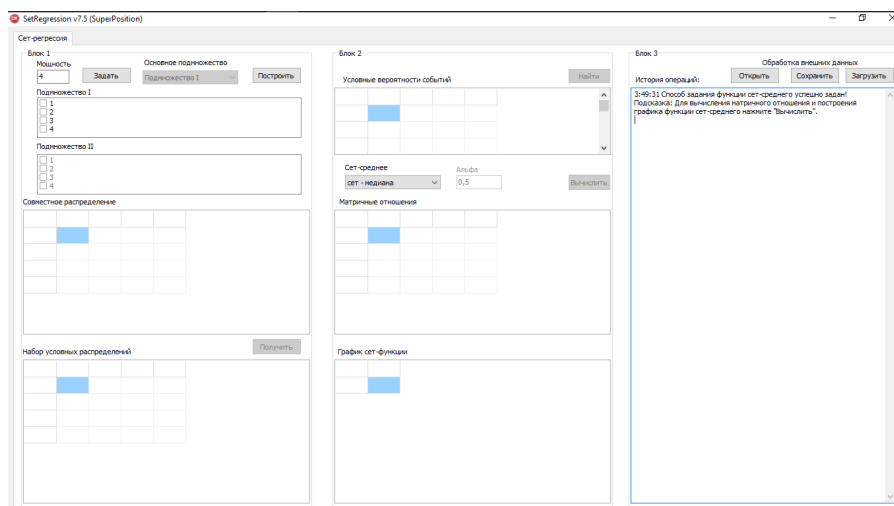


Рисунок 18 - Запуск программного комплекса SetRegression.

Шаг 2. Необходимо указать мощность множества и разбиение его на два множества. Максимальная длина входа выявленная экспериментальным путем равна 23.

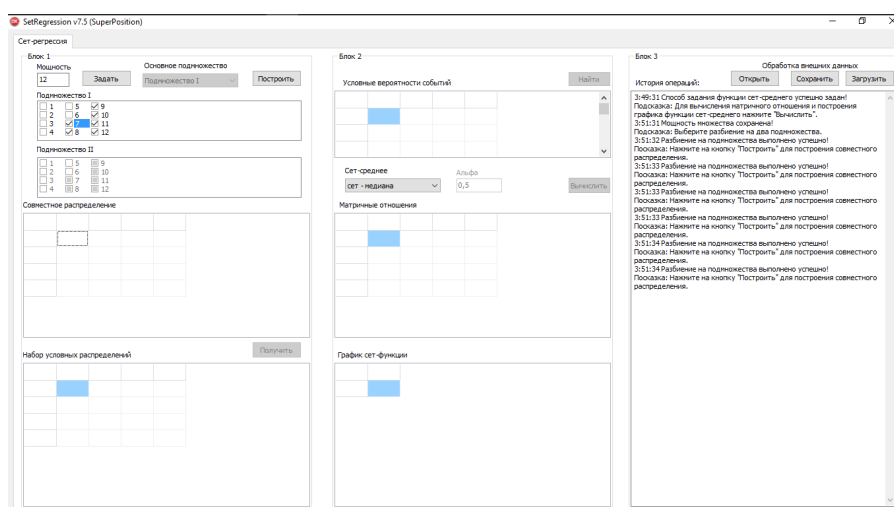


Рисунок 19 - Задание начальных параметров.

### Шаг 3. Заполнение таблицы совместного распределения.

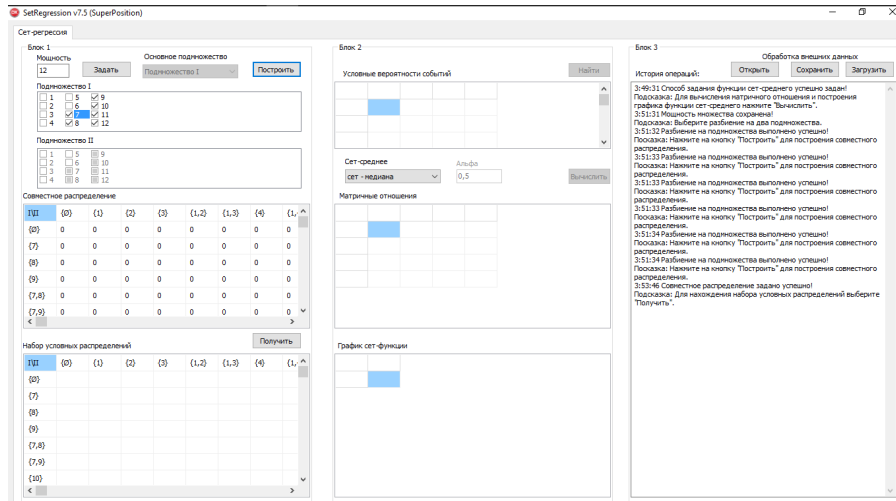


Рисунок 20 - Восстановление таблицы совместного распределения.

Шаг 4. Для загрузки данных при решении конкретной задачи в программу достаточно нажать кнопку «Открыть» и в привычном окне проводника Windows выбрать нужный документ. Для упрощения данной процедуры осуществляется просмотр документов исключительно с расширением «.csv». На рисунке представлена база данных совместного распределения множеств «Социальные характеристики» и «Образ жизни».

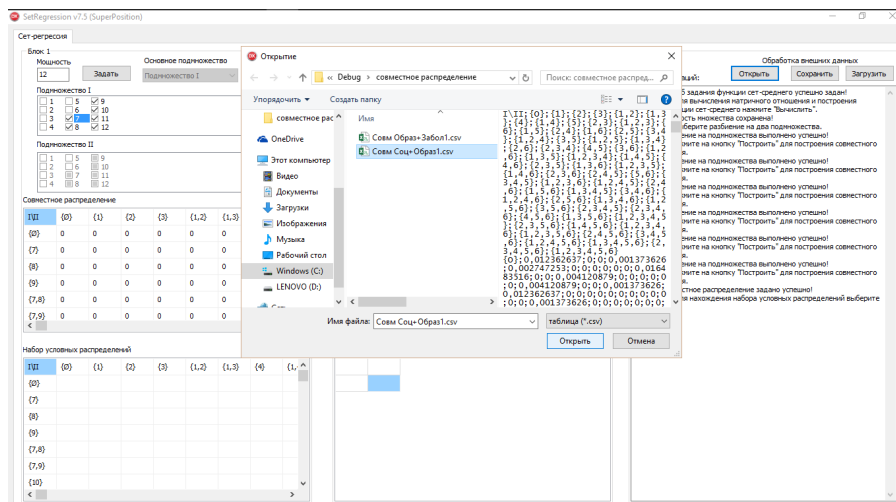


Рисунок 21 - Загрузка данных из файла в таблицу совместного распределения.



Шаг 7. График сет-функции зависит от выбора сет-среднего, который может быть определен как сет-ожидание, сет-квантиль, сет-мода или сет-медиана.

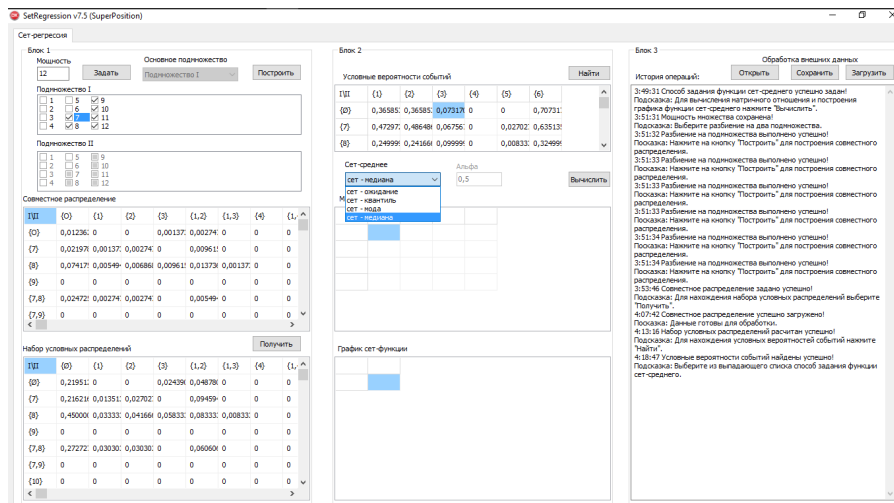


Рисунок 24 - Выбор сет-среднего из выпадающего списка.

Шаг 8. Нажатие кнопки «Вычислить» активирует последовательно два алгоритмических модуля: нахождение матричного отношения и построение графика сет-функции.

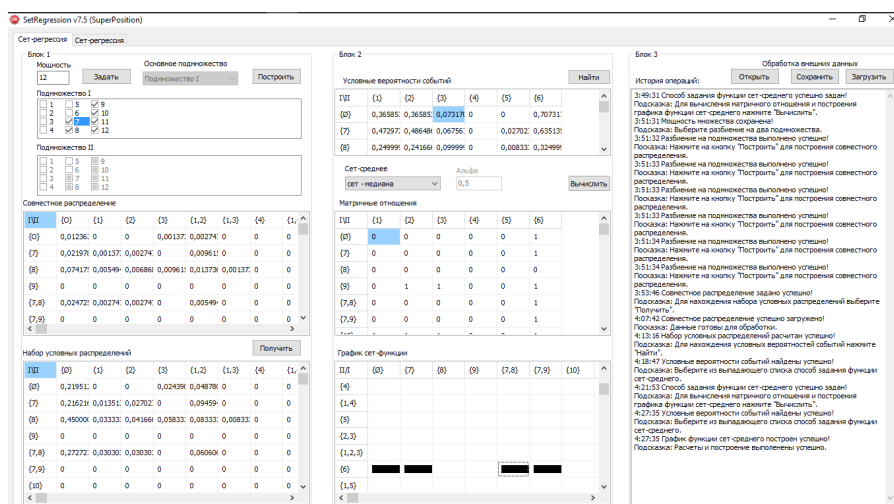


Рисунок 25 - Вычисление матричного отношения и построение графика сет-функции.

Шаг 9. Для получения новой сет-регрессии для множеств «Образ жизни» и «Заболевания» пройдены шаги 1-8.

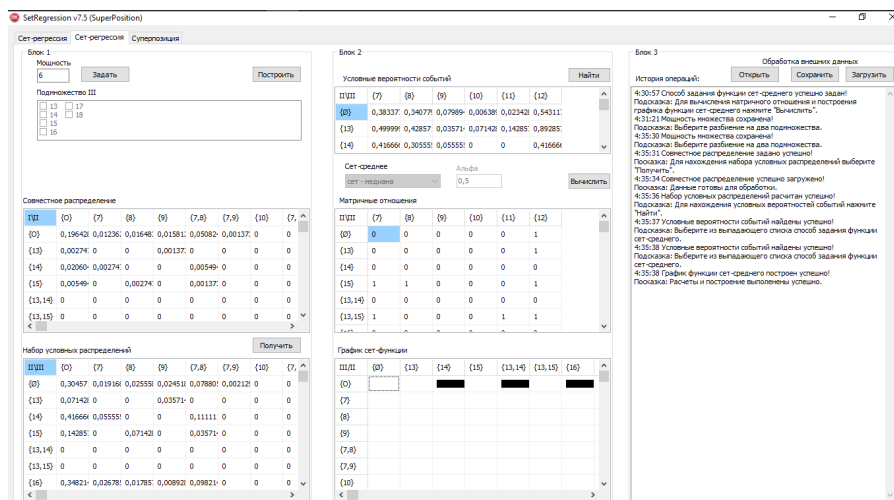


Рисунок 26 - Сет-регрессия множества «Образ жизни» и «Заболевания».

Шаг 10. На последнем шаге строится суперпозиция ранее полученных сет-регрессий. Для получения графика функции пользователь нажимает кнопку «Построить». График суперпозиции строится согласно алгоритму описанному в первой главе с учетом выбранного сет-среднего.

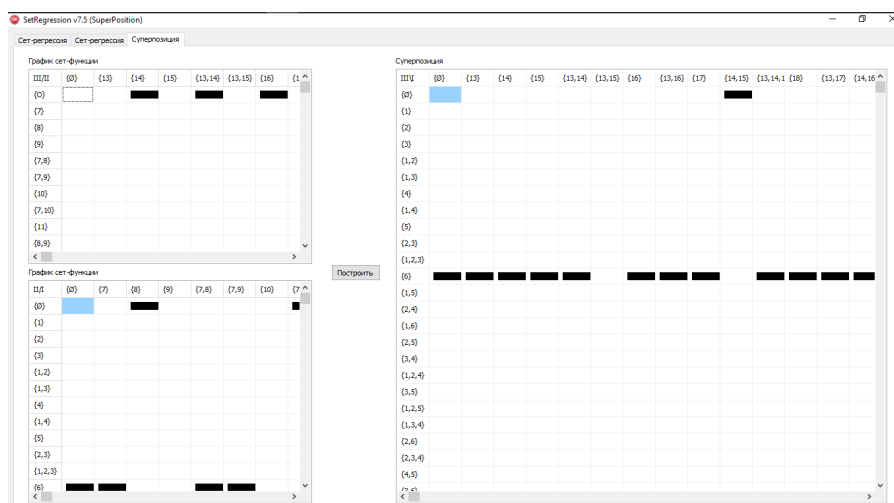


Рисунок 27 - График суперпозиции в виде сет-медианы

Вид средних характеристик в процессе выполнения программы в программном интерфейсе можно изменить только на странице 1. На странице 2 по командам пользователя обрабатываются вероятностное совместное распределение множества 2 и множества 3. На странице 3 строиться суперпозиция сет-регрессий. На следующих рисунках построены графики суперпозиции в виде различных сет-средних. Значение сет-квантиля может варьироваться от 0 до 1.

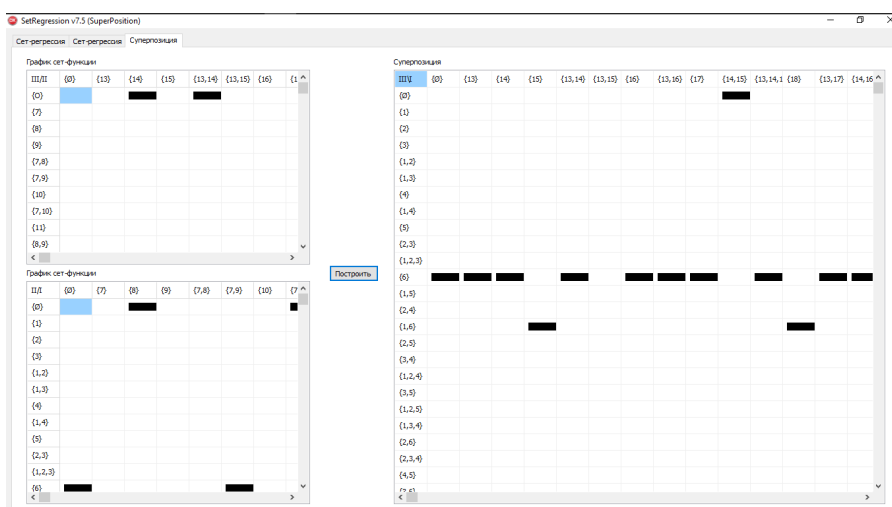


Рисунок 28 - График суперпозиции в виде сет-квантиля порядка  $\alpha = 0.45$

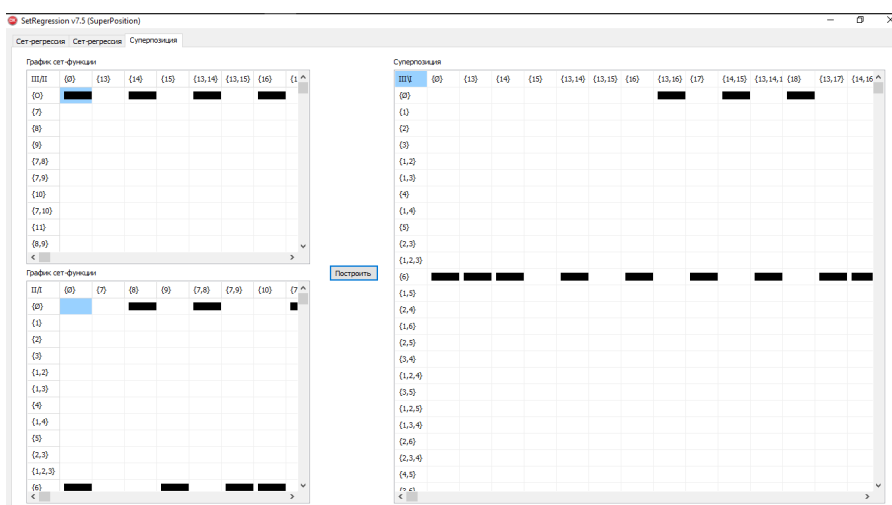


Рисунок 29 - График суперпозиции в виде сет-моды

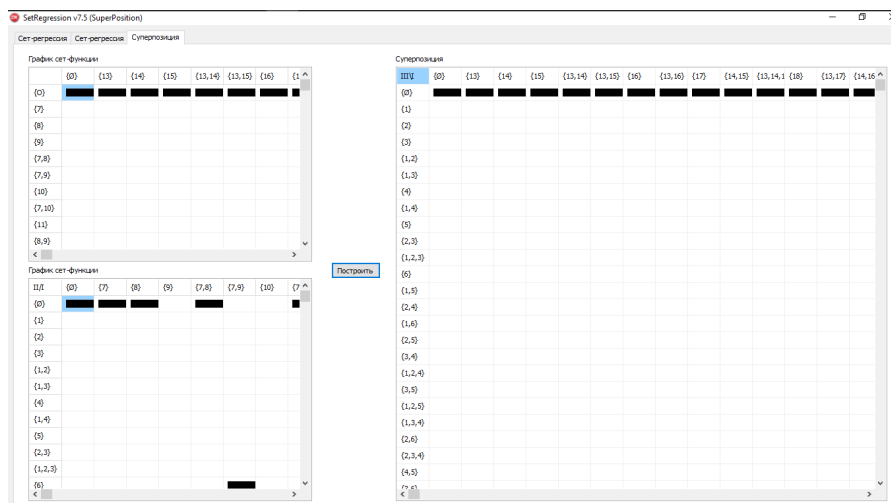


Рисунок 30 - График суперпозиции в виде сет-ожидания

Алгоритм программного решения задачи сет-регрессии имеет ступенчатый характер. Для достижения конечного результата выполнение каждого шага алгоритма обязательно. Без ущерба полученным результатам эксперимента пользователь может завершить работу программного комплекса «SetRegression» на любом этапе алгоритма. Для повторного исследования задачи пользователь может сохранить исходные данные в виде совместного распределения в файл с расширением «.csv» выбрав кнопку «Сохранить» из Блока 3 на соответствующей странице.

## Резюме

Глава посвящена программному решению задач случайно-множественного регрессионного анализа, в частности задачи сет-регрессии и суперпозиции сет-регрессий. В главе 3 проиллюстрированы и проанализированы прикладные результаты решения задачи сет-регрессии и суперпозиции для входных данных в виде документа Microsoft Excel со статистикой медицинского исследования в исходном виде, полученном от двух медицинских учреждений Красноярского края. Программный комплекс «SetRegression» является универ-

сальным инструментом для обработки любых входных данных нечислового характера представленного в виде таблицы совместного распределения, записанного в файл «\*.csv» (табличная данные записаны с разделители «;»). Пример, представления входных данных приведен в параграфе 2 данной главы. В качестве прикладного примера выполнено задание: предобработан входной файл Microsoft Excel со статистическими данными медицинского исследования за 2012-2013 годы, произведен вычислительный эксперимент и сделаны выводы по полученным результатам.



## ЗАКЛЮЧЕНИЕ

**Основные результаты.** По итогу проделанной работы достигнуты следующие результаты:

1. Изучена теоретическая основа решения задачи случайно-множественного анализа по работам Воробьёва О.Ю., Фомина А.Ю., Семеновой Д.В., Лукьяновой Н.А., Ивановой А.И..
2. Построена модель решения задачи сет-регрессии.
3. Разработаны интуитивный интерфейс ПО и алгоритмы решения данной задачи. Проведен анализ созданных алгоритмов.
4. Обработана статистика по результатам медицинского исследования. Полученная таким образом статистика в виде совместного распределения апробирована в качестве входных данных в разработанном программном обеспечении. Таким образом исследована задача сет-регрессии в виде сет-средних на практическом примере.

Разработанное программное обеспечение «SetRegression» позволяющее пользователю наблюдать не только конечный, но и промежуточный результат построения модели решения задачи случайно-множественного регрессионного анализа, является инструментом для научных исследований в области математической статистики и смежных прикладных областях, связанных с обработкой статистических данных нечисловой природы, также может быть полезна для учебных целей.

**Апробация работы.** Результаты научной деятельности и некоторые положения в рамках бакалаврского исследования, также спроектированное ПО «SetRegression» докладывались и обсуждались на научных и научно-практических конференциях:

1. XII Всероссийская научно-техническая конференция студентов, аспирантов и молодых ученых «МОЛОДЁЖЬ И НАУКА» (Красноярск, 2016);
2. XII Всероссийская научно-техническая конференция студентов, аспирантов и молодых ученых «МОЛОДЁЖЬ И НАУКА» (Красноярск, 2015);
3. XIV международная ФАМЭМС'2015 конференция (Красноярск, 2015).

**Публикации.** Имеются три публикации автора по теме данной бакалаврской работе. Публикации приведены в списке использованных источников под пунктами [1, 2, 9].

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Атабеков, М.А. Программный комплекс для расчета характеристик случайного множества событий / М. А. Атабеков // Сборник материалов Международной конференции студентов, аспирантов и молодых ученых «Перспектив Свободный-2015», посвященной 70-летию Великой Победы. — Красноярск, 2015 — С. 4-5.
2. Атабеков, М.А. Суперпозиция сет-регрессий / М. А. Атабеков // Сборник материалов Международной конференции студентов, аспирантов и молодых ученых «Перспектив Свободный-2016». — Красноярск, 2016.
3. Белов, К. А. Методы и алгоритмы случайно-множественного анализа медицинских данных: автореф. дис. на соиск. учен. степ. канд. тех. н. по специальности 05.13.01 / К. А. Белов. — Воронеж, 2005.
4. Быкова, В.В. Комбинаторные алгоритмы: множества, графы, коды: учеб. пособие / В. В. Быкова. — Красноярск: Сиб. федер. ун-т, 2015. — 152 с.
5. Быкова, В.В. Дискретная математика с использованием ЭВМ: учеб. пособие / В. В. Быкова. — Красноярск: КрасГУ, 2006. — 200с.
6. Воробьев, О.Ю. Сет-регрессионный анализ зависимостей событий в статистических системах: учебное пособие / О.Ю. Воробьев, А.Ю. Фомин. — Красноярск: ИВМ СО РАНб, 2004. — 116 с.
7. Воробьев, О.Ю. Эвентология: учебное пособие / О.Ю. Воробьев. — Красноярск: Сибирский федеральный университет, 2007. — 435 с.
8. Загоруйко, Н.Г. Прикладные методы анализа данных и знаний: учебное пособие / Н. Г. Загоруйко. — Новосибирск: Изд-во Ин-та математики, 1999. — 270 с.

9. Иванова, А.И. О задаче сет-регрессии / А.И. Иванова, М.А. Атабеков // Труды XIV конференции по финансово-актуарной математике и эвентологии многомерной статистики. — Красноярск, 2015 — С. 285-292.
10. Иванова, А.И. Нечеткая и сет-регрессионная модели распределения потребительских предпочтений между фирмами / А.И. Иванова, Д.В. Семенова // Труды XIII международной ФАМЭМС'2014 конференции (под ред. О.Ю. Воробьева) / СФУ, НИИППБ — Красноярск, 2014 — С. 164-170.
11. Кобзарь, А.И. Прикладная математическая статистика. Для инженерных и научных специальностей: учебное пособие / А.И. Кобзарь. — Москва: Физматлит, 2006.— 816 с.
12. Крянев, А.В. Математические методы обработки неопределенных данных: учебное пособие / А. В. Крянев, Г.В. Лукин ; — Москва: Физматлит, 2003. — 216 с.
13. Орлов, А. И. Нечисловая статистика: учебное пособие / А.И.Орлов. — Москва: МЗ-Пресс, 2004. — 513 с.
14. Орлов А. И. Статистика объектов нечисловой природы. Вероятность и математическая статистика: Энциклопедия. / Ю. В. Прохоров. — Москва: БРЭ, 1999. — 646 – 648 с.
15. Прохоров, Ю. В. Вероятность и математическая статистика: Энциклопедия / Ю. В. Прохоров. — Москва: Большая Российская энциклопедия, 1999. — 911 с.
16. Тарасова, О. Ю. Сеточные и регрессионные алгоритмы аппроксимации сложных систем событий: автореф. дис. на соиск. учен. степ. канд. физ.-мат. н.(151301) / О. Ю. Тарасова. — Красноярск, 2007. — 14-16 с.

17. Флёнов, М.Е. Delphi в шутки и всерьез: что умеют хакеры / М. Е. Флёнов — СПб: Питер, 2006 — 271 с.
18. Чернова, Н.И. Математическая статистка: учебное пособие / Н. И. Чернова. — Новосибирск: Новосибирский государственный университет, 2007 — 1488 с.
19. Ширяев, А.Н. Вероятность-1: учебное пособие / А. Н. Ширяев. — Москва: МЦНМО, 2004 — 520 с.
20. Lukyanova N.A. Eventological subdistributions: matrix representation of the events-terraces for a set of events / N. A. Lukyanova // Труды XV Международной ЭМ'2011 конференция. 10–11 декабря 2011. — Красноярск: КГТЭИ, СФУ, 2011. — С. 20–26.

## **Приложение А**

Страницы 61–71 изъяты по решению правообладателя.