

Краткая инструкция по работе в ISW

Оглавление

1. Главное окно программы	2
1.1. Описание всплывающих окон	3
2. Подготовка файлов для работы в ISW	4
3. «Действия»	5
3.1. Статистический анализ	5
3.2. Статистический анализ v2	8
3.3. Группирование выборки	10
3.4. Проверка на равномерность	11
3.5. «Проверка на показательность» и «Проверка на нормальность»	15
3.6. Проверка гипотезы об однородности законов двух (и более) выборок	15
3.6.1. Проверка однородности двувывборочными критериями	16
3.6.2. Проверка однородности многовыборочными критериями	20
3.7. Проверка гипотезы о случайности и об отсутствии тренда	24
3.8. Вычисление мощности	27
4. Графики	30
5. Моделирование выборок	33
5.1. Задание тренда	39
5.2. Моделирование функций от случайных величин	40
6. Форма «Параметры»	43
7. Полезные ссылки для изучения подробностей о критериях	47

1. Главное окно программы

После запуска программы пользователь первым делом встречается с главным рабочим окном программы, которое показано на рисунке 1.

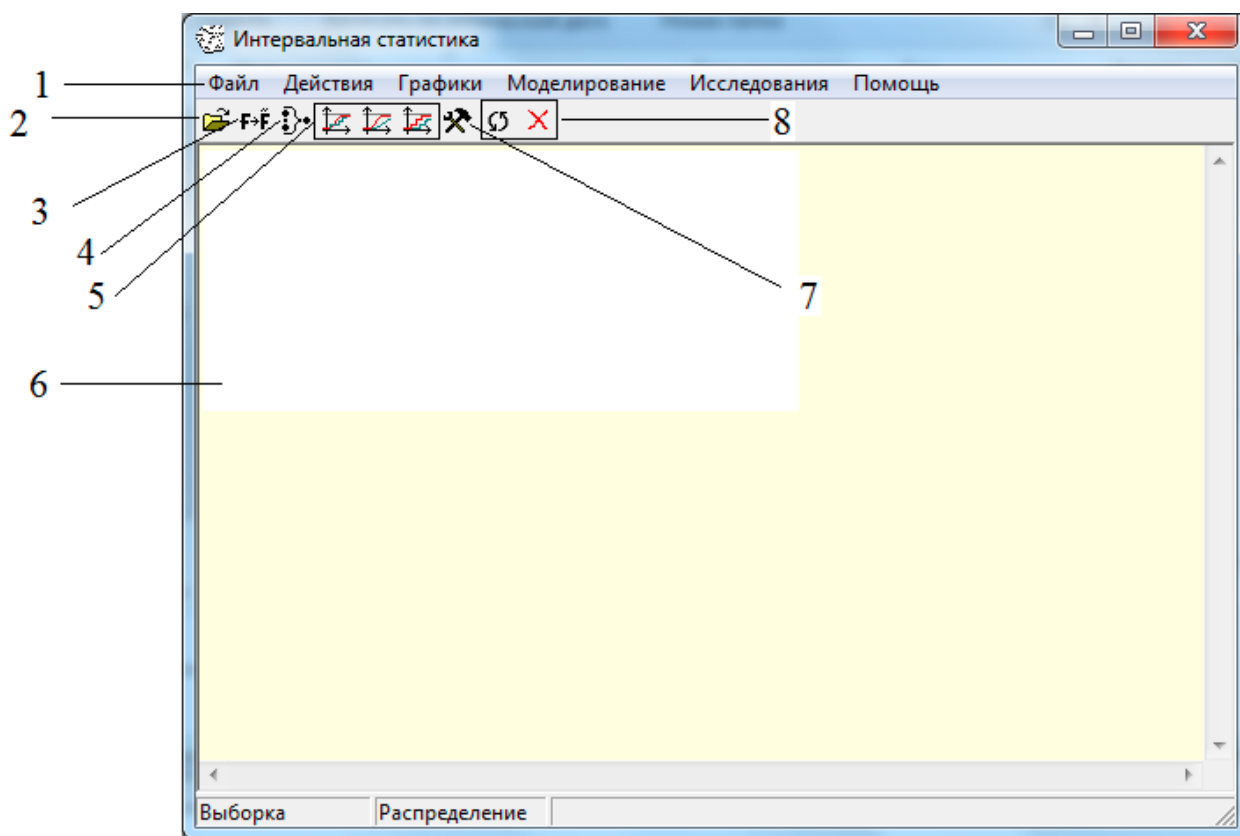


Рисунок 1 – Интерфейс заглавного окна ISW

Описание интерфейса:

1 – Всплывающие окна выбора;

2 – клавиша для добавления в систему выборок из файловой системы;

3 – клавиша для открытия окна «Оценивание параметров и проверка согласия» оно же «Статистический анализ» (описано в разделе 3.1);

4 – клавиша для открытия окна «Группирование выборки» (описано далее в 3.3);

5 – клавиши для открытия графических окон для отображения функции распределения подгруженных выборок вместе со всеми подгруженными в систему распределениями (по умолчанию подгружено стандартное нормальное), функции распределения только подгруженных

распределений и функции распределения только подгруженных выборок, соответственно.

6 – Окно уведомлений (в данном окне появляются сообщения о завершении моделирования и записи моделируемых данных в файл, а также некоторая информация о проверках согласия);

7 – клавиша для открытия окна «Параметры», в котором можно менять такие вещи как методы оценивания, методы группирования и т.д. (описано далее);

8 – клавиши инициализации и очищения окна уведомлений (6).

1.1. Описание всплывающих окон

1. Всплывающее окно «Файл» содержит только 3 опции: открытия файла из файловой системы, инициализации и выхода из программы;
2. Всплывающее окно «Действия» содержит множество операций, описание которых приводится далее в 3 главе;
3. Опции всплывающего окна «графики» соответствуют клавишам 5 главного окна (более подробное описание дается в главе 4);
4. Опции всплывающего окна «Моделирование» описаны далее в главе 5;
5. Опции всплывающего окна «Исследования» не предназначены для рядовых пользователей и содержат наработки студентов для собственных исследований;
6. Опции всплывающего окна «Помощь» предназначены для обратной связи с авторами проекта.

2. Подготовка файлов для работы в ISW

Выборка с наблюдениями хранится в текстовом формате (файл с расширением ".dat"). Структура файла зависит от типа выборки.

- Первой строкой идет комментарий, название распределения, которого будет отображаться на графиках и прочих вспомогательных инструментах.
- Далее число 0 является идентификатором для точечной выборки.
- После идентификатора идет число соответствующее объему выборок.
- Последующие числа считываются как элементы выборки. Возможные форматы подаваемых чисел 3.72264e-01 или 0.372264. Если чисел в выборке больше чем заданное число объема, то остальные числа игнорируются.

Точечная выборка объемом n наблюдений имеет следующий формат:

<название выборки>

0 n

<наблюдение 1>

<наблюдение 2>

...

<наблюдение n>

Например, в файле 1.dat содержатся данные:

Выборка 1

0 5

0.321

0.359

-0.341

1.016

0.207

3. «Действия»

3.1. Статистический анализ

Статистический анализ выборки производится в форме "Оценивание параметров и проверка согласия" (кнопка **F→F** на панели инструментов или пункт меню «Действия» → «Статистический анализ»), как показано на рисунках 2 и 3. Необходимо выбрать выборку, закон распределения, метод оценивания и критерии согласия. Для проверки простой или сложной гипотезы о согласии исходных данных с выбранным законом распределения необходимо нажать кнопку «Оценить и проверить».

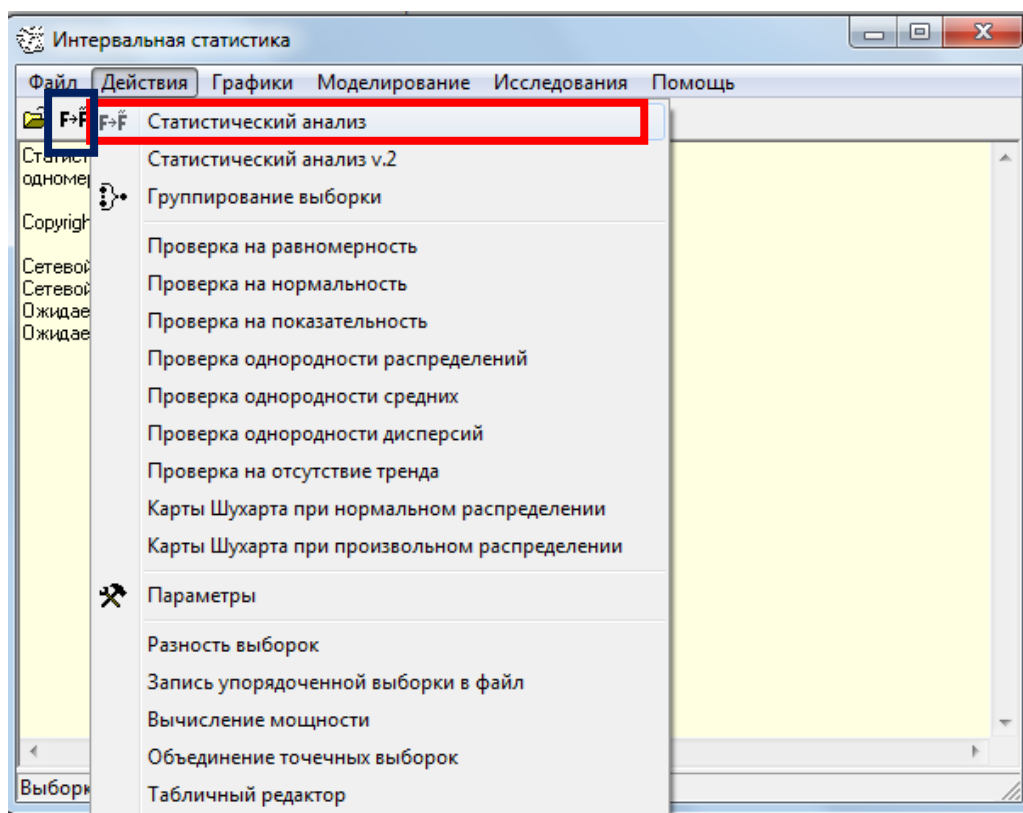


Рисунок 2 – Выбор раздела «Статистический анализ»

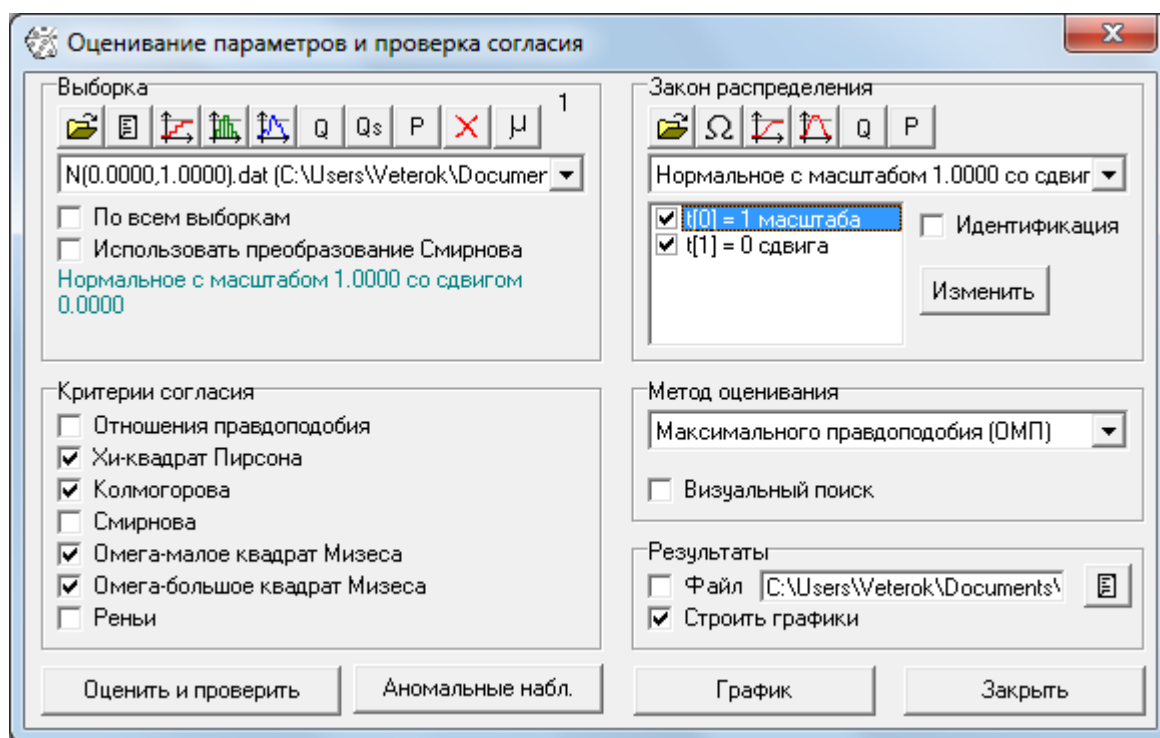






Рисунок 3– Форма «Статистический анализ»

Результат проверки гипотезы будет выведен в основном окне программы, в окне уведомлений (см. рис. 4). В окне указывается достигаемый уровень значимости – общий и для каждого выбранного критерия согласия. А также распределение с оцененными параметрами (при проверке сложной гипотезы), наилучшим образом описывающее исследуемые данные.

- Выборка

Выборку можно выбрать из списка, либо открыть файл с выборкой. В списке отображаются только те выборки, которые перечислены в разделе [Samples] в файле инициализации «is.ini». Здесь можно просмотреть саму выборку , эмпирическую функцию распределения по этой выборке , гистограмму  (если выборка группированная), ядерную оценку плотности ; а также вычислить выборочные квантили по заданным вероятностям - кнопка **Q** или по заданным точкам вычислить частоты (т.е. отношение количества наблюдений, попавших левее точки к объему выборки) – кнопка **P**.

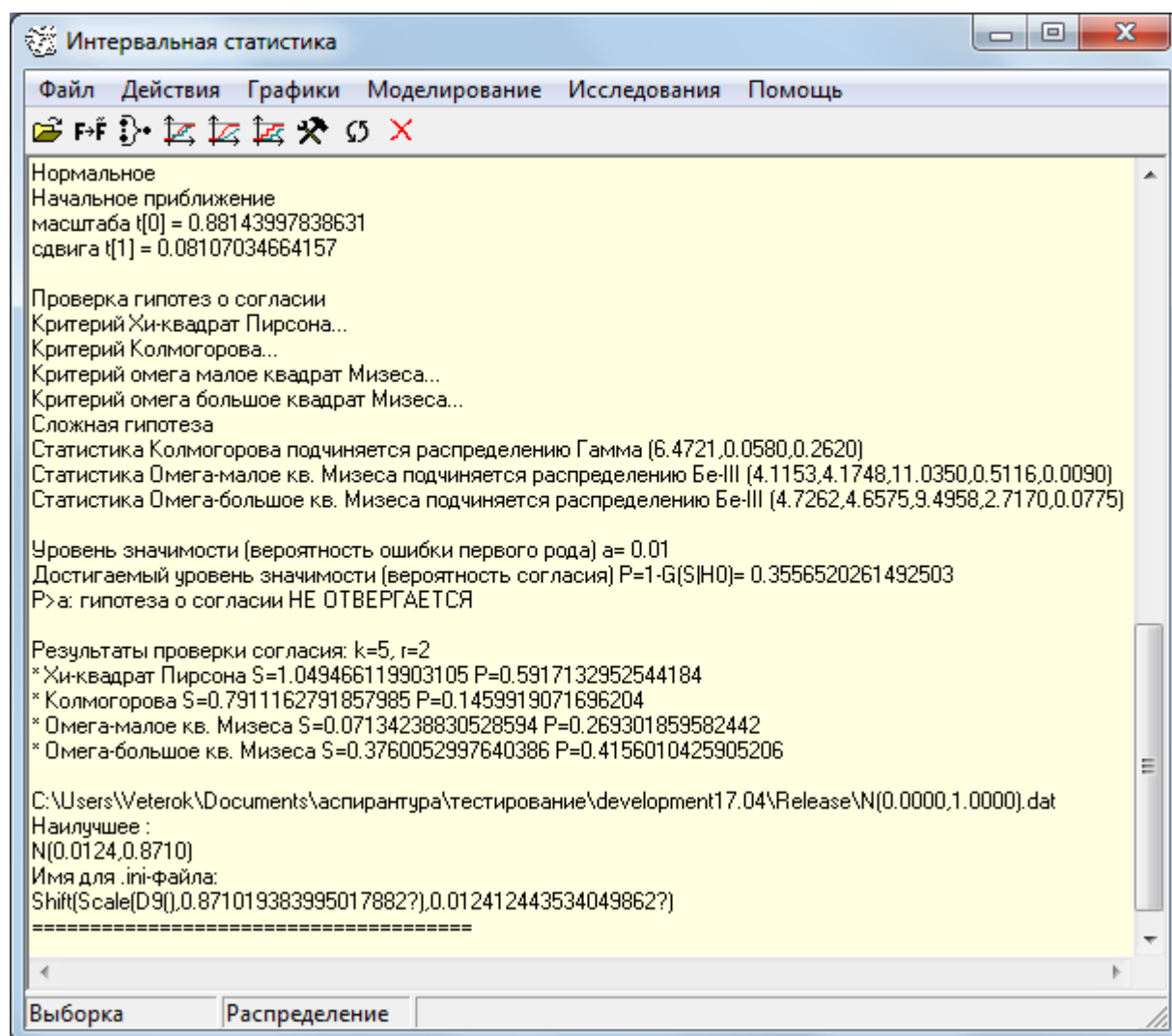






Рисунок 4 – Результат проверки гипотезы о согласии

- Закон распределения

В системе заложено более 30 стандартных распределений и возможность добавлять новые распределения, получаемые из стандартных с помощью операций сдвига, масштабирования, смеси, произведения, зеркального отображения, усечения.

В списке отображаются те распределения, которые перечислены в разделе [Distributions] в файле инициализации «is.ini». Можно открыть другой (подготовленный ранее) список распределений , он задается в файле с расширением «dst». В форме «Параметры распределений» (кнопка ) выдается информация о распределениях списка: идентификатор, наименование, тип, область определения, граница слева, граница справа,

число параметров, параметры и их значения. Здесь также предусмотрена возможность просмотра графиков функции распределения  и функции плотности , а также возможность вычисления квантилей распределения по заданным вероятностям - кнопка **Q** и по заданным точкам x вычисления вероятностей $P\{x < X\}$ – кнопка **P**.

Кнопка "График" выводит функцию распределения выбранного закона и эмпирическую функцию распределения выбранной выборки на одном рисунке.

- Оценка параметров и проверка гипотез

При нажатии кнопки "Оценить и проверить" производится поиск оценок параметров закона распределения выбранным методом оценивания и выполняется проверка согласия выбранной выборки с выбранным законом распределения. При этом вычисляются оценки тех параметров, напротив которых стоит флажок. Если не выбран ни один из критериев согласия, то производится только оценивание параметров. Проверяется простая гипотеза, если ни один из параметров не оценивается.

Если стоит флажок "По всем выборкам", то действия будут выполняться последовательно по каждой выборке. Если стоит флажок "Идентификация", то по совокупности критериев согласия будет найден наилучший закон (из тех распределений, которые представлены в списке), описывающий конкретную выборку.

- Аномальные наблюдения

При нажатии кнопки "Аномальные наблюдения" производится отбраковка аномальных наблюдений по выбранному закону распределения.

3.2. Статистический анализ v2

Расширенная версия статистического анализа приведена на рисунках 5 и 6.

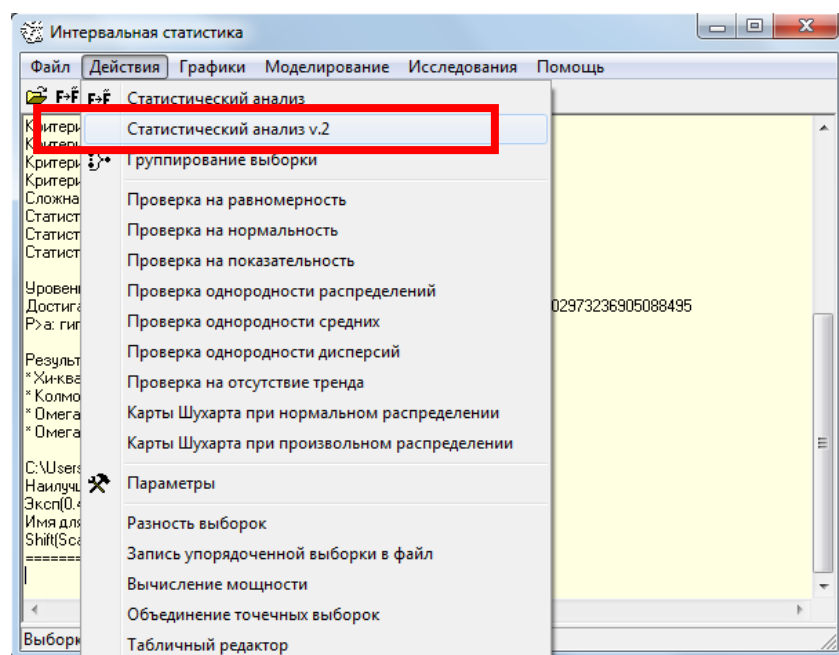


Рисунок 5 – Форма «Статистический анализ v2»

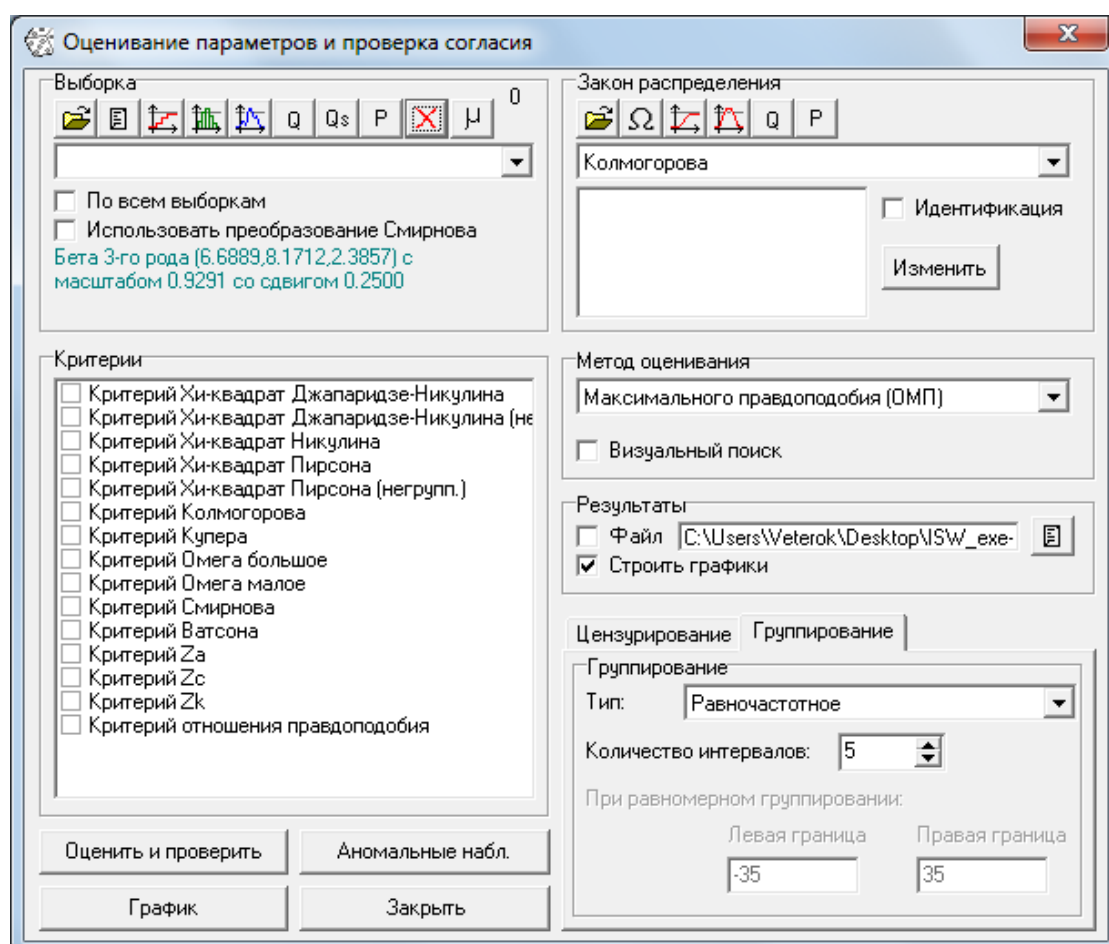



Рисунок 6 – Форма «Статистический анализ v2»

По сравнению с формой «Статистический анализ» в данную форму добавлены настройки способов группирования и критерии согласия, встроенные в данное программное обеспечение.

3.3. Группирование выборки

Группирование выборки производится в форме (переход по кнопке  на панели инструментов или пункт меню «Действия» → «Группирование выборки»), представленной на рисунке 7. Для работы в данной форме, необходимо иметь уже подгруженную выборку в систему. В форме можно изменить закон распределения, число интервалов и метод группирования. Для группирования исходной выборки необходимо нажать кнопку «Сгруппировать выборку». Результат группирования сохраняется в файл, название которого при необходимости меняется в правой нижней области формы.

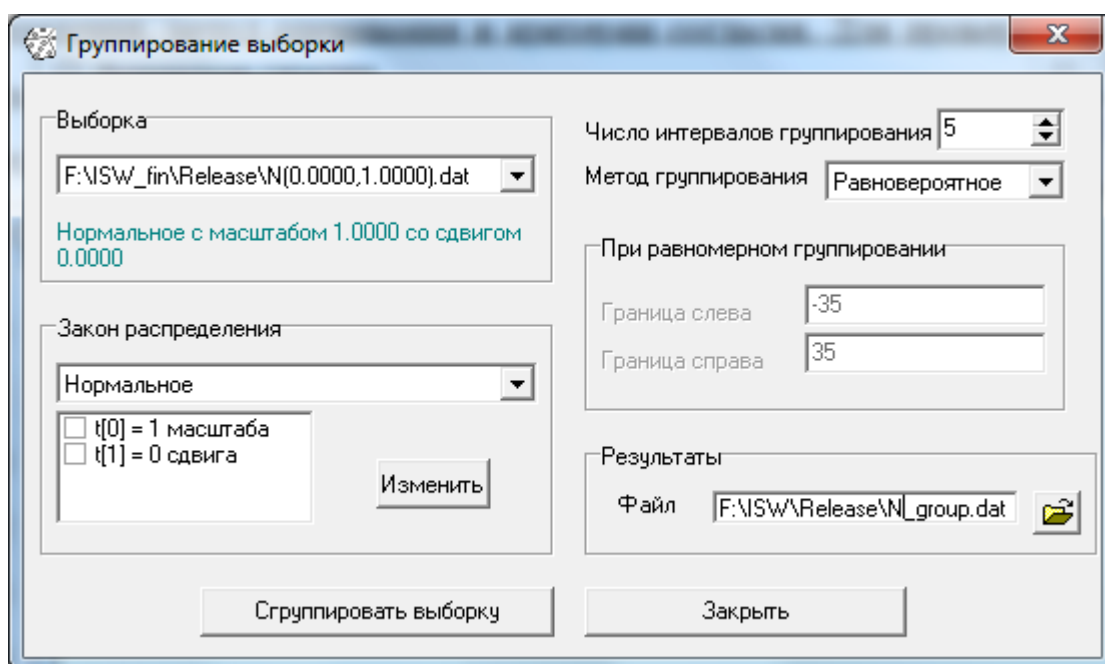


Рисунок 7 – Форма «Группирование выборки»

3.4. Проверка на равномерность

Данное окно находится в «Действия» → «Статистический анализ». Проверка на равномерность осуществляется на интервале $[0,1]$. На рисунке 8 представлен интерфейс соответствующей формы.

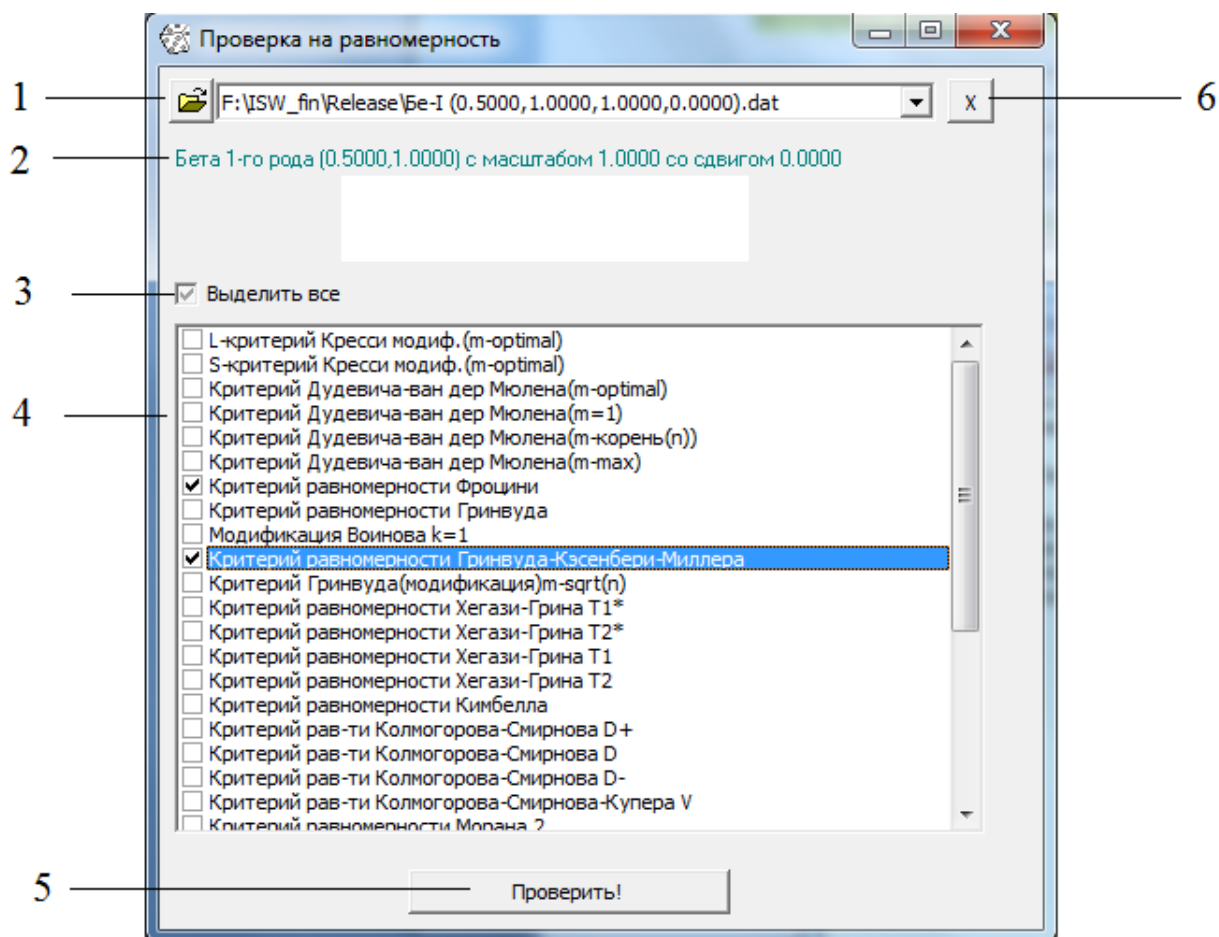


Рисунок 8 – Интерфейс «Проверки на равномерность»

- 1 – клавиша для добавления в систему выборок из файловой системы (данную опцию можно использовать и в других окнах системы);
- 2 – текстовое сообщение, содержащие комментарий выборки;
- 3 – чекбокс для выделения/снятия выделения всех критериев списка;
- 4 – список критериев равномерности, с возможными выбором группы критериев для проверки;
- 5 – клавиша запуска проверки равномерности заданной выборки выбранными критериями;
- 6 – клавиша для удаления всех выборок, подгруженных в систему.

После нажатия клавиши 5 появляется следующее окно (см. рис 9).

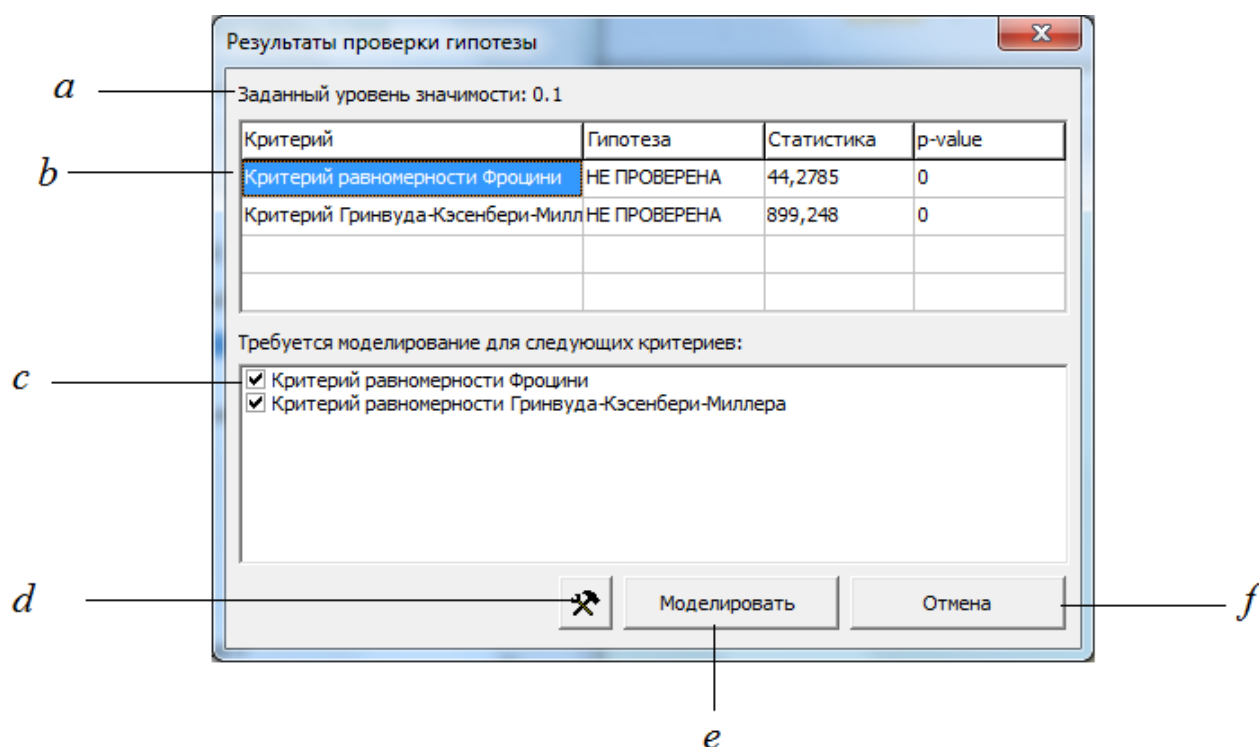



Рисунок 9 – Интерфейс «Результаты проверки гипотезы»

а – Заданный уровень значимости, по которому после моделирования в зависимости от p-value делается вывод о статусе гипотезы. Изменить данное значение можно через опцию «Параметры» (раздел «Критерии», снизу окна), которую можно попасть через кнопку  на главном окне программы или пункт меню «Действия» → «Параметры».

б – Список критериев, проверяющих гипотезу (на данном этапе уже посчитана статистика критерия, для изменения статуса «гипотезы» и получения значения p-value необходимо провести моделирование распределения статистик).

с – Список критериев с чекбоксами для выбора критериев для моделирования их распределений.

д – Клавиша открытия вспомогательного окна «Параметры моделирования» (см. рис 10). Данная форма позволяет менять количество имитационных экспериментов N при моделировании эмпирических распределений статистик, по которым будет оцениваться p-value, число используемых

потоков при распараллеливании процесса моделирования, а также дает возможность сохранить в отдельную выборку эмпирическое распределение статистики.

е – клавиша запуска моделирования распределения статистик для выбранных критериев для выборок на интервале $[0,1]$. После нажатия клавиши появится сообщение на главном окне и спустя некоторое время появляется новое окно с результатами проверок. *Моделирование распределений статистик весьма трудоемкий процесс и может занимать значительное время.*

f – клавиша для отмены/закрытия окна.

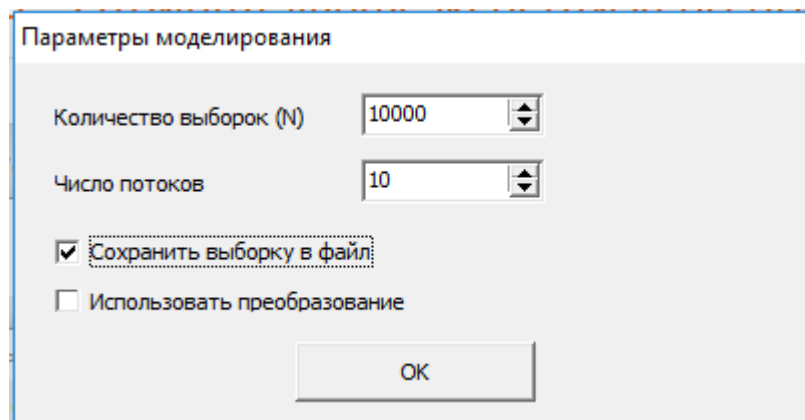



Рисунок 10 – «Параметры моделирования»

В окне «Параметры моделирования» можно задать режим «Использовать преобразование», тогда откроется следующее окно, в котором можно выбрать «Кратности» и через  задать величину Δ , с которой будут округляться моделируемые случайные величины.

Режим «Использовать преобразование» может выбираться только в том случае, если реальные измерения в анализируемой выборке представляют собой достаточно грубо округлённые значения с величиной округления Δ . Тогда есть риск, что при справедливости проверяемой гипотезы распределение статистики может существенно отличаться от того распределения, которому принадлежит статистика в случае непрерывной случайной величины. При моделировании с величиной округления Δ решение будет приниматься по реальному распределению статистики.

После нажатия клавиши «OK» запускается «Моделирование».

Параметры моделирования

Количество выборок (N)

Число потоков

☒ Сохранить выборку в файл

☒ Использовать преобразование

Кратности

Распределение	Преобразование
Рав(0.0000, 1.0000)	

В данном примере анализируемая выборка существенно отличалась от равномерного закона.

Результаты проверки гипотезы

Заданный уровень значимости: 0,1

Критерий	Гипотеза	Статистика	p-value
Критерий равномерности Фроцони	ОТКЛОНЯЕТСЯ	44,2785	0
Критерий Гринвуда-Кэсенбери-Милл	ОТКЛОНЯЕТСЯ	899,248	0

Важно! Не рекомендуется выбирать много критериев для проверки гипотезы. В связи с переполнением возможны вылеты программы.

3.5. «Проверка на показательность» и «Проверка на нормальность»

Данные формы можно выбрать во всплывающем окне «Действия» и их работа аналогична работе формы «Проверка на равномерность», описанная в предыдущем пункте. Отличием является только возможность менять интервал проверки (через сдвиг и масштаб), однако стоит заметить, что некоторые критерии в силу своих особенностей могут быть неэффективны при оценивании сдвига/масштаба.

3.6. Проверка гипотезы об однородности законов двух (и более) выборок

Для проверки гипотезы об однородности распределений по исходным (по реальным полученным) данным был создан модуль «Проверка однородности распределений» (см. рис. 11).

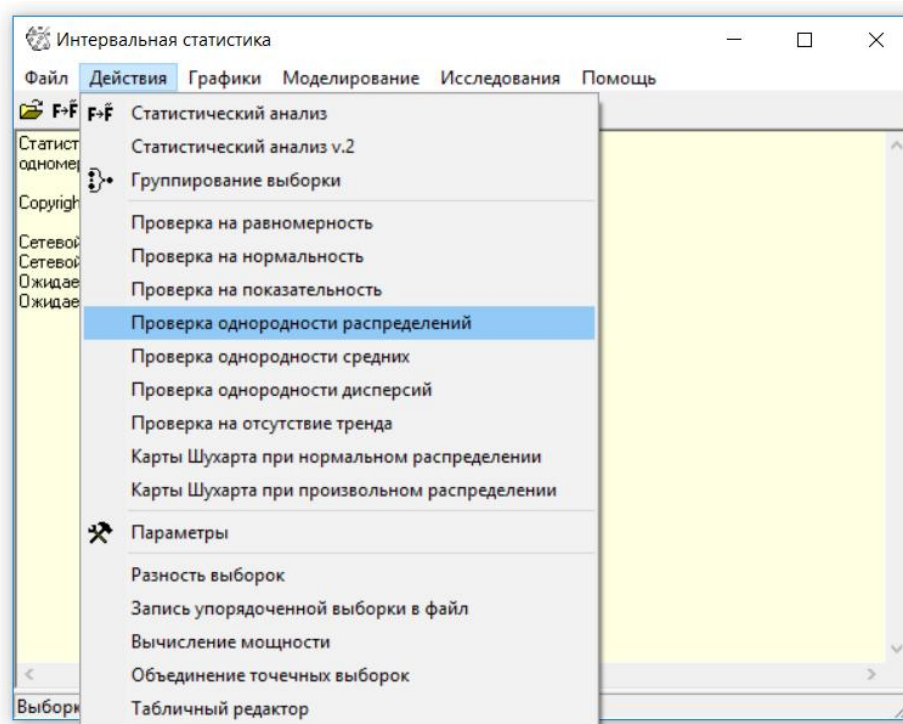



Рисунок 11 – Пункт меню «Проверка однородности распределений»

Рисунок 12 – Форма «Проверка однородности распределений»

В открывающейся форме «Проверка однородности распределений» (см. рис. 12) необходимо подгрузить несколько выборок (количество должно превышать одну), нажав на панели кнопку , выбрать критерии для проверки гипотезы однородности распределений, выбрать уровень значимости (ошибку первого рода α) и кнопку «Проверить». Поскольку критерии непараметрические, то закон распределения исходных данных не влияет на результат проверки гипотезы и достоверность статистического вывода.

3.6.1. Проверка однородности по двувывборочным критериям

Критерии однородности Андерсона-Дарлинг-Петита, Смирнова (из-за смещения распределения статистики относительно своего предельного требует моделирования распределения при проверке), Смирнова (модифицированный с поправкой, ликвидирующей смещение), Лемана-

Розенблатта предназначены для проверок гипотез об однородности законов **только двух выборок**.

Выбрав двувывборочные критерии, как показано на рисунке 13, нажимаем кнопку «Проверить». В открывшемся окне (см. рис. 14) с результатами проверки приводится список критериев, результаты проверки гипотезы, значение статистики и достигнутый уровень значимости. Для рассмотренного примера три критерия отклоняют гипотезу об однородности законов, а для критерия Смирнова требуется моделирование. Нажав кнопку



, можно установить параметры моделирования (см. рис. 15), такие как число потоков моделирования (по умолчанию выставляется максимально возможное для системы), объем моделирования (количество имитаций статистик), сохранять/не сохранять полученное распределение в файл. Данные параметры можно оставить по умолчанию. Для запуска моделирования и проверки гипотезы нажать кнопку «Моделировать».

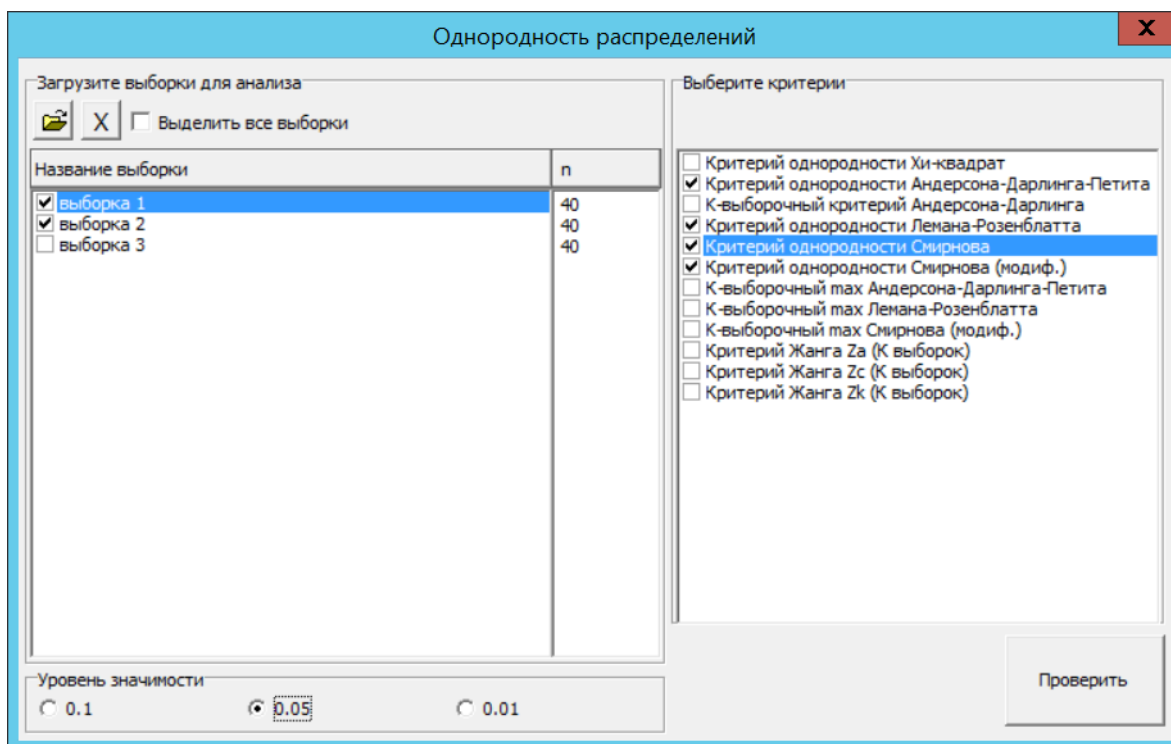


Рисунок 13 – Двувывборочные критерии однородности законов

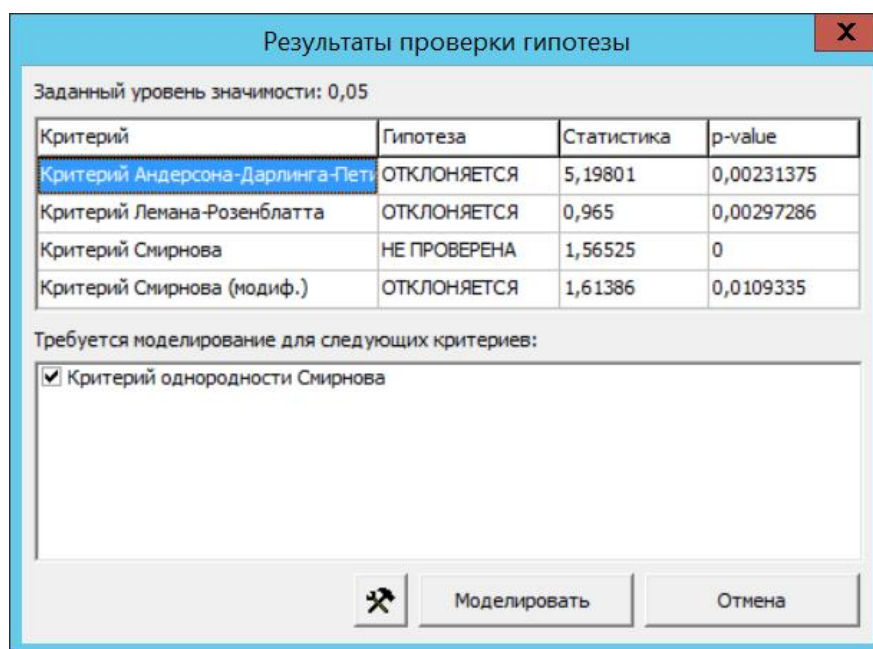


Рисунок 14 – Результаты проверки гипотезы однородности

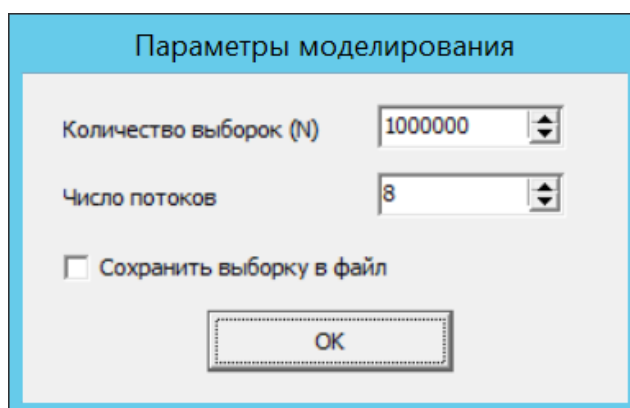


Рисунок 15 – Параметры моделирования

В основном окне отображается процент выполненного моделирования, как показано на рисунке 16.

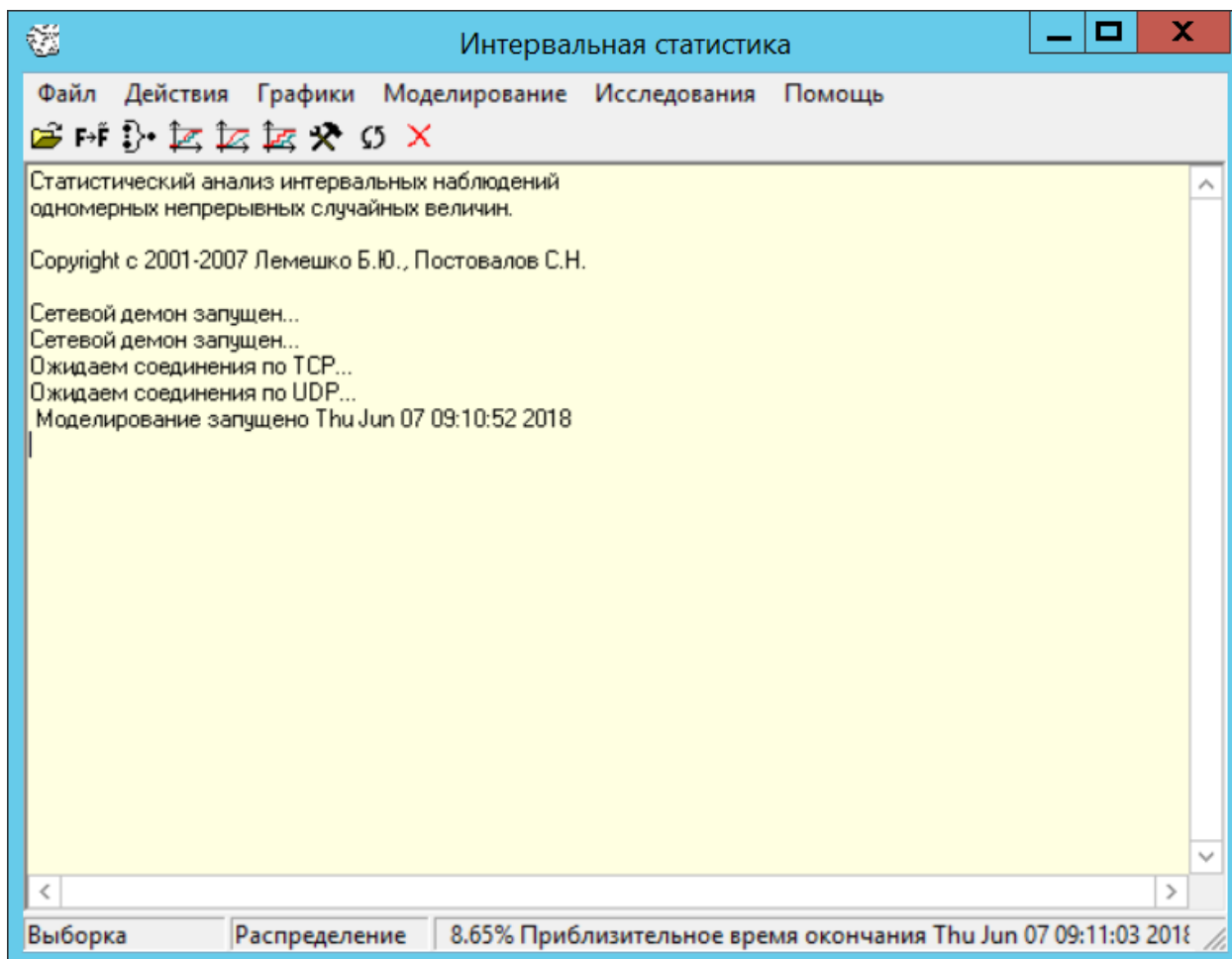


Рисунок 16 – Процент выполненного моделирования

По завершению моделирования появляется окно с результатом проверки гипотезы (см. рис. 17).

Результаты проверки гипотезы			
Заданный уровень значимости: 0,05			
Критерий	Гипотеза	Статистика	p-value
Критерий Смирнова	ОТКЛОНЯЕТСЯ	1,56525	0,014503

Рисунок 17 – Результат проверки гипотезы после моделирования

3.6.2. Проверка однородности многовыборочными критериями

Для проверки однородности законов двух и более выборок применяются следующие критерии:

- однородности χ^2 ;
- k-выборочный Андерсона–Дарлинга;
- k-выборочный максимума Смирнова;
- k-выборочный максимума Лемана–Розенблатта;
- k-выборочный максимума Андерсона–Дарлинга;
- Жанга со статистиками Z_K , Z_A , Z_C .

Критерий однородности χ^2 единственный критерий, предполагающий группирование исходных данных. Группирование организовано равночастотным методом. Необходимо обратить внимание на существенную зависимость результатов проверки по критерию однородности χ^2 от выбираемого количества интервалов r (см. рис. 18).

Критерий	Статистика	P_{value}
χ^2 , $r = 10$	25.556	0.1104
χ^2 , $r = 8$	19.200	0.1574
χ^2 , $r = 7$	21.627	0.0419

Рисунок 18 – Задание числа интервалов для критерия χ^2

Для k-выборочного критерия Андерсона–Дарлинга и критериев максимума Смирнова, максимума Лемана–Розенблатта, максимума Андерсона–Дарлинга, критерия однородности χ^2 встроены полученные предельные распределения статистик, что позволяет для вычисленного по предельному распределению значения статистики определить достигнутый уровень значимости и сделать вывод об отклонении или неотклонении гипотезы об однородности распределений (см. рис. 19).

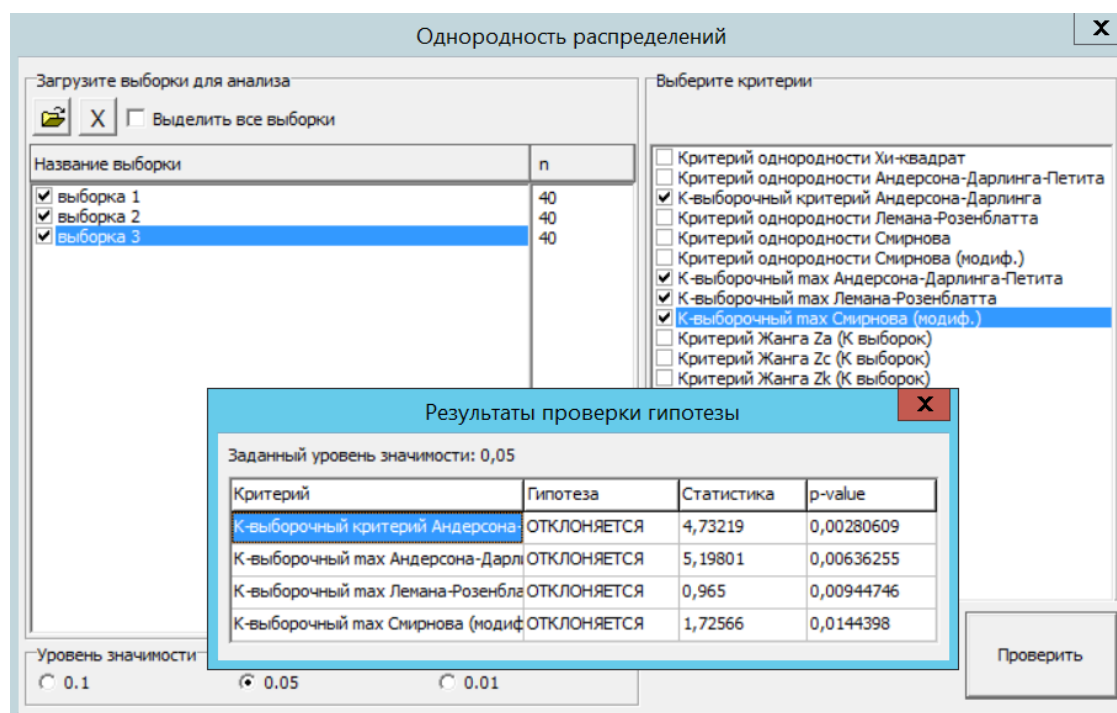


Рисунок 19 – Результат проверки гипотезы об однородности трех выборок

Следует отметить, что только для критерия Жанга со статистиками Z_K , Z_A , Z_C предельное распределение отсутствует, поскольку распределение статистик Жанга существенно зависит от количества анализируемых выборок и их объемов. Для проверки гипотезы об однородности распределений при помощи критерия Жанга со статистиками Z_K , Z_A , Z_C необходим интерактивный режим моделирования: по исходным данным вычисляется значение статистики критерия, при заданных объемах выборок моделируется распределение статистики критерия, по которому определяется

достигнутый уровень значимости, сравнивая который с ошибкой первого рода, приложение отклоняет или не отклоняет гипотезу (см. рис. 20,21,22).

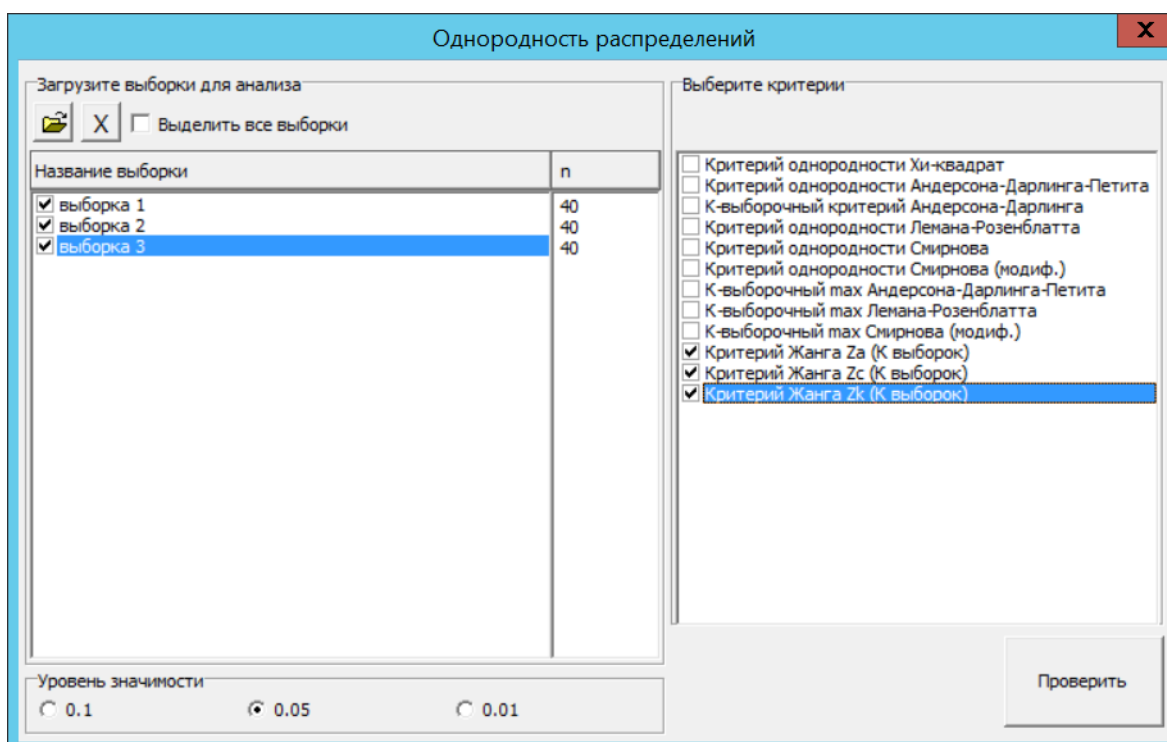


Рисунок 20 – Проверки гипотезы об однородности трех выборок критериями Жанга

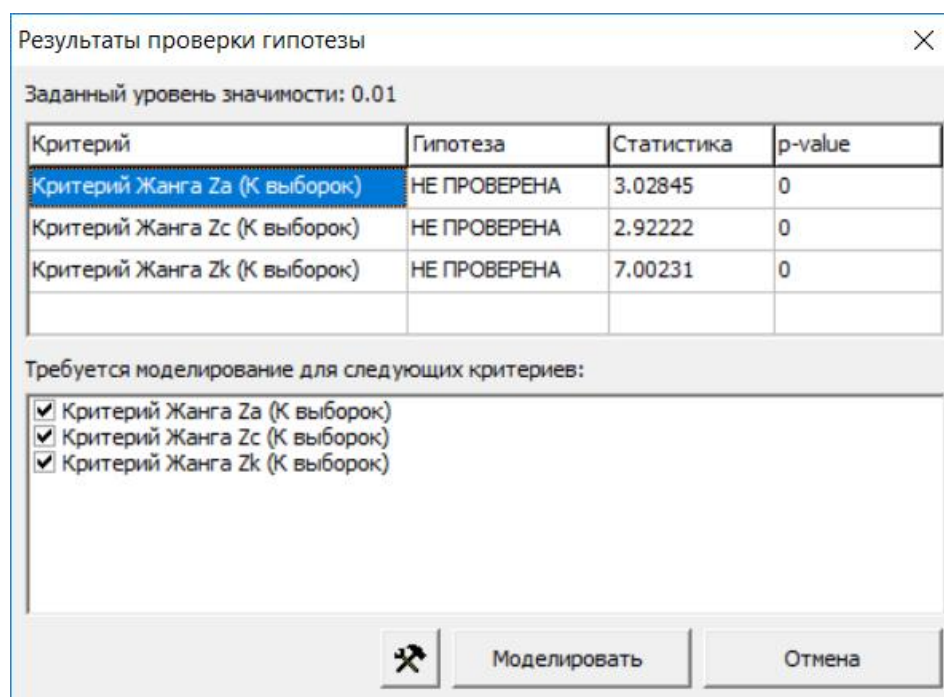


Рисунок 21 – Моделирование распределений статистик критериев Жанга

Результаты проверки гипотезы			
Заданный уровень значимости: 0.01			
Критерий	Гипотеза	Статистика	p-value
Критерий Жанга Z_a (K выборки)	ОТКЛОНЯЕТСЯ	3.02845	0.00126506
Критерий Жанга Z_c (K выборки)	ОТКЛОНЯЕТСЯ	2.92222	0.00156627
Критерий Жанга Z_k (K выборки)	НЕ ОТКЛОНЯЕТСЯ	7.00231	0.021506

Рисунок 22 – Результаты проверки гипотезы критериями Жанга

Замечание. При достаточно больших объемах выборок процедура довольно ресурсоемкая. Во избежание зависаний программы (особенно для маломощных компьютеров) рекомендуем запускать проверку гипотезы однородности законов критериями Жанга по одному.

3.7. Проверка гипотезы о случайности и об отсутствии тренда

Для проверки гипотезы о случайности и отсутствии тренда по исходным (по реальным полученным) данным был создан модуль «Проверка на отсутствие тренда» (см. рис. 23).

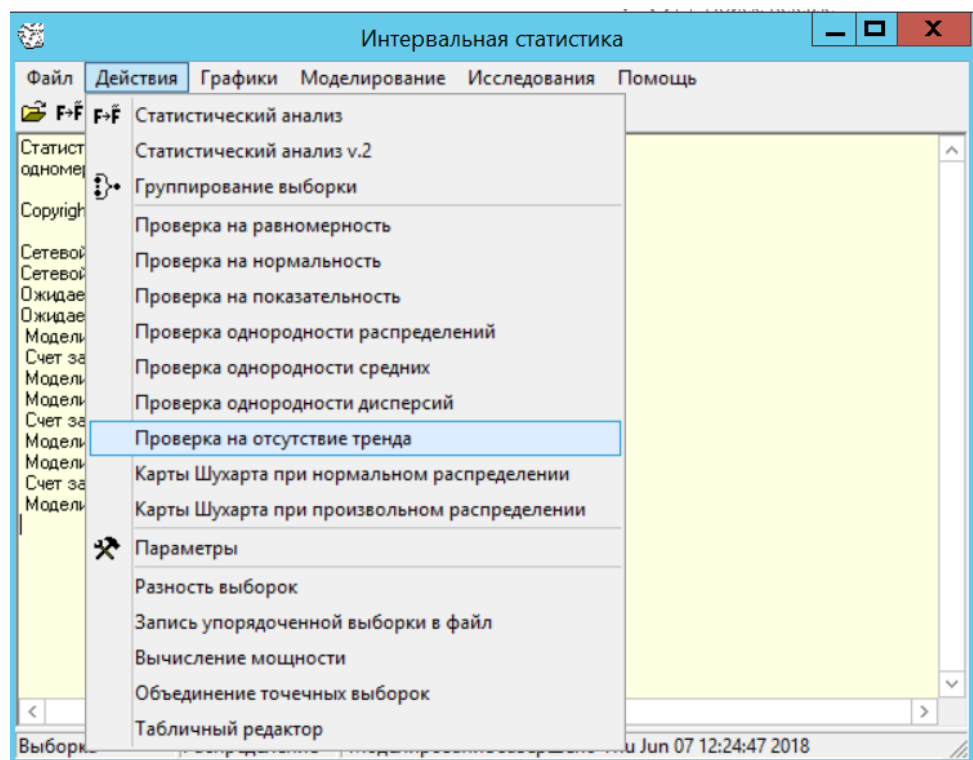


Рисунок 23 – Пункт меню «Проверка на отсутствие тренда»

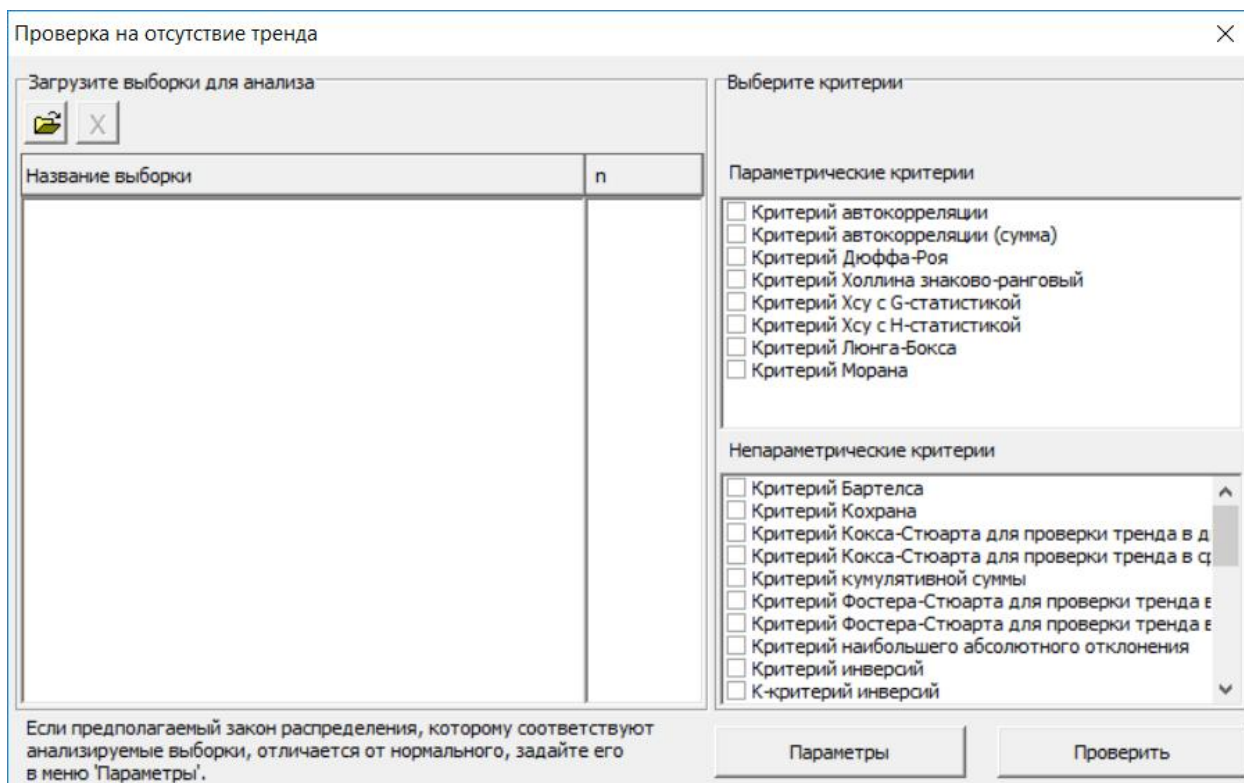



Рисунок 24 – Форма «Проверка на отсутствие тренда»

В открывающейся форме «Проверка на отсутствие тренда» (см. рис. 24) необходимо:

1. Подгрузить выборку (для проверки гипотезы должна быть отмечена только одна выборка), нажав на панели кнопку .
2. Выбрать критерии для проверки гипотезы о случайности и отсутствии тренда.
3. Настроить в разделе «Параметры» уровень значимости и закон распределения (только для параметрических критериев). Распределение статистик **параметрических критериев** зависит от закона распределения входных данных. В связи с этим для параметрических критериев в меню «Параметры» (см. рис. 25) необходимо задать закон и параметры распределения, наилучшим образом описывающие входные данные.

В системе заложено более 30 стандартных распределений и возможность добавлять новые распределения, получаемые из стандартных с помощью операций сдвига, масштабирования, смеси, произведения, зеркального отображения, усечения.

По умолчанию в списке отображаются те распределения, которые перечислены в разделе [Distributions] в файле инициализации «is.ini». В нашем примере – это стандартный нормальный закон $N(0,1)$. Для изменения параметров закона необходимо нажать кнопку «Параметры» и в открывшемся окне ввести значения параметров закона (см. рис. 16).

Параметры

Распределение

☒ Нормальное

☐ $t[0] = 1$ масштаба

☐ $t[1] = 0$ сдвига

Чтобы изменить параметры закона распределения, нажмите 'Параметры'

Параметры

Уровень значимости

☐ 0.1 ☐ 0.05 ☒ 0.01

OK

Рисунок 25 – Форма «Параметры»

Параметры закона распределения

Параметры	Значения
$t[0]$ Параметр масштаба, фиксирован	1
$t[1]$ Параметр сдвига, фиксирован	0

OK

Рисунок 26 – Форма «Параметры»

Можно открыть другой (подготовленный ранее) список распределений



, он задается в файле с расширением «dst».

4. Нажать кнопку «Проверить».

Пример проверки гипотезы о случайности и отсутствия тренда приведен на рисунках 27 и 28. На рисунке 27 критерии Дюффа-Роя, Бартелса не отклоняют гипотезу случайности и отсутствия тренда в математическом ожидании; ранговый критерий с метками Клотца не отклоняет гипотезу случайности и отсутствия тренда в дисперсии; в то время как для модифицированного критерия автокорреляции требуется моделирование распределения статистики критерия, используемого при проверке гипотезы. После завершения моделирования открывается окно с результатом проверки гипотезы, приведенное на рисунке 28. Модифицированный критерий

автокорреляции не отклоняет гипотезу о случайности и отсутствии тренда в математическом ожидании.

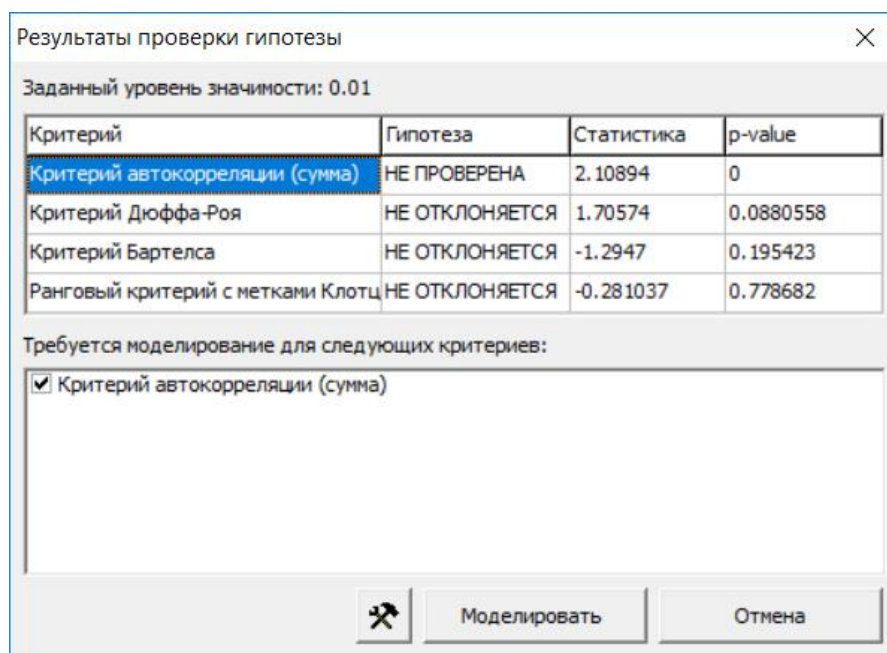


Рисунок 27 – Запуск моделирования для модификации критерия автокорреляции

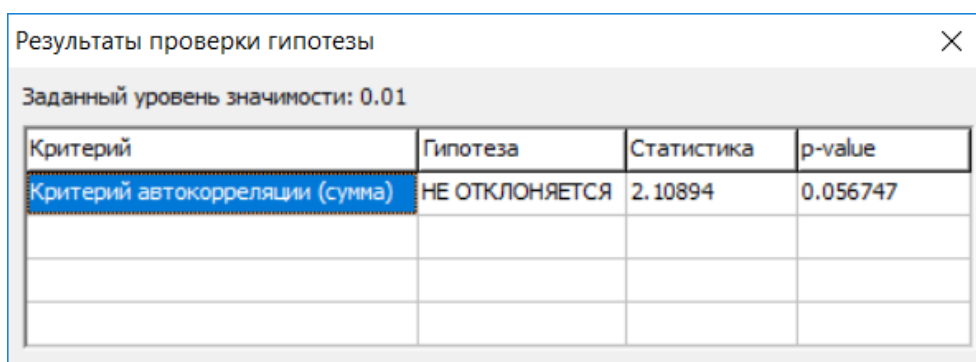


Рисунок 28 – Результаты проверки гипотезы

3.8. Вычисление мощности

Вычисление мощности производится в форме (пункт меню «Действия» → «Вычисление мощности»), представленной на рисунке 29. В форме можно выбрать выборки в качестве истинной $G(S | H_0)$ и конкурирующей $G(S | H_1)$ гипотез, менять тип критерия (правосторонний, левосторонний и двусторонний), вводить и изменять ошибки первого рода «А». Для получения соответствующих для «А» мощностей в столбце «1-В» необходимо нажать кнопку «Вычислить». В интерфейсе также добавлена

клавиша «Симметричные точки», которая при нажатии создает в столбце «А» вероятности обратные текущим ошибкам первого рода.

Вычисление мощности

G(S|H0) Критерий равномерности Гринвуда N=16600 P_{ав}(0.0000,1.0000) n=100 ГСЧ=100

G(S|H1) Критерий равномерности Гринвуда N=16600 Б-И (0.7000,1.0000,1.0000,0.0000) n=100 ГСЧ=100

А	1-В
0.15	0.465843
0.1	0.364639
0.05	0.241145
0.025	0.158193
0.01	0.093494

Критерий

☒ Правосторонний

☐ Левосторонний

☐ Двусторонний

Симметричные точки

Вычислить

>

Рисунок 29 – Форма «Вычисление мощности»

При нажатии «>» раскрывается расширенная форма «Вычисление мощности». В данной форме позволяет выбрать несколько выборок в качестве истинной гипотезы и аналогично для конкурирующей гипотезы. После нажатия клавиши «Вычислить все» происходит полный перебор пар всех выбранных $G(S|H_0)$ и $G(S|H_1)$ и в результате для каждой выбранной $G(S|H_0)$ выводится строкой для всех введенных «А» лучшая мощность. Стоит отметить, что тип критериев применяется для всех проверок.

Вычисление мощности

G(SIH0) Критерий равномерности Гринвуда N=16600 Pав(0.0000,1.0000) n=100 ГСЧ=100

G(SIH1) Критерий равномерности Гринвуда N=16600 Be-I (0.7000,1.0000,1.0000,0.0000) n=100 ГСЧ=100

A	1-B
0.15	0.465843
0.1	0.364639
0.05	0.241145
0.025	0.158193
0.01	0.093494

Критерий

☒ Правосторонний

☐ Левосторонний

☐ Двусторонний

Симметричные точки

Вычислить

>

☐ Критерий равномерности Гринвуда N=16600 Be-I (0.7000,1.0000,1.0000,0.0000) n=100 ГСЧ=100

☒ Критерий равномерности Гринвуда N=16600 Pав(0.0000,1.0000) n=100 ГСЧ=100

☐ Критерий равномерности Фроцини N=16600 Be-I (0.7000,1.0000,1.0000,0.0000) n=100 ГСЧ=100

☒ Критерий равномерности Фроцини N=16600 Pав(0.0000,1.0000) n=100 ГСЧ=100

☒ Критерий равномерности Гринвуда N=16600 Be-I (0.7000,1.0000,1.0000,0.0000) n=100 ГСЧ=100

☐ Критерий равномерности Гринвуда N=16600 Pав(0.0000,1.0000) n=100 ГСЧ=100

☒ Критерий равномерности Фроцини N=16600 Be-I (0.7000,1.0000,1.0000,0.0000) n=100 ГСЧ=100

☒ Критерий равномерности Фроцини N=16600 Pав(0.0000,1.0000) n=100 ГСЧ=100




Вычислить все

0.465843;0.364639;0.241145;0.158193;0.093494
0.936988;0.909518;0.855301;0.784096;0.683554

Рисунок 30 – Расширенная форма «Вычисление мощности»

4. Графики

На один рисунок можно вывести:

- графики эмпирических функций распределения по всем выборкам, перечисленным в разделе [Samples] инициализационного файла «is.ini» – кнопка  на панели инструментов (или в меню “Графики” выбрать “Все выборки”).
- графики всех функций распределения, перечисленных в разделе [Distributions] файла is.ini – кнопка  на панели инструментов (или в меню “Графики” выбрать “Все распределения”).
- графики эмпирических функций распределения по всем выборкам, перечисленным в разделе [Samples] и графики всех функций распределения, перечисленных в разделе [Distributions] файла is.ini – кнопка  на панели инструментов (или в меню “Графики” выбрать “Все графики”).

На рисунке 31 показан пример формирования графика при проверке гипотезы о согласии выборки объемом 50 наблюдений и нормального закона распределения с оцененными параметрами.

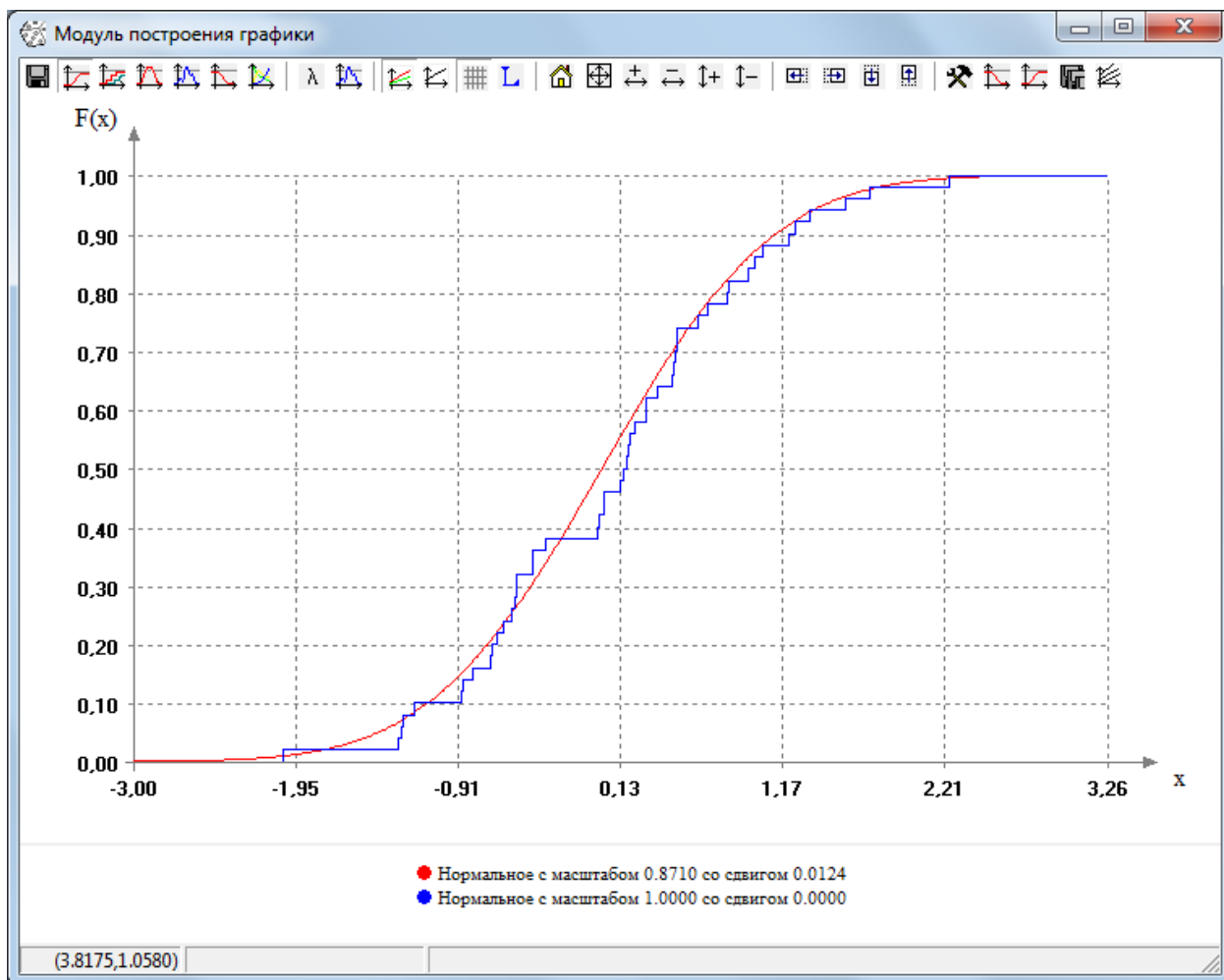






Рисунок 31 – Модуль построения графики

В окне “График” можно менять настройки графика:

-  – выводить/не выводить сетку
-   – изменить палитру
-  – задать границы по X и Y, выбрать размеры области и режим представления (см. рис. 32):

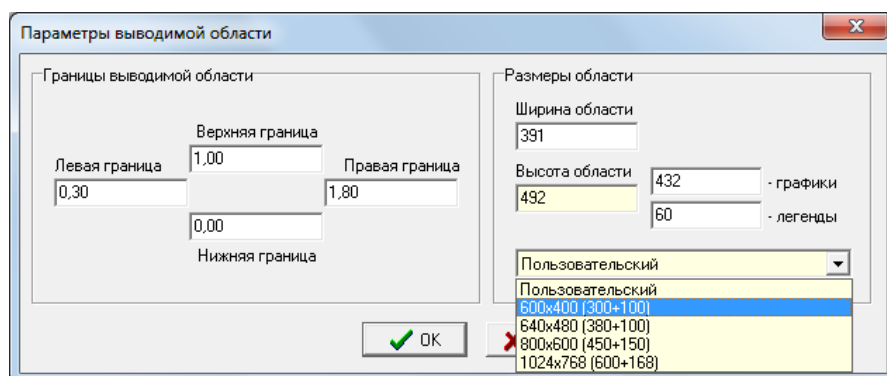












Рисунок 32 – Границы области построения

-  – сжать график по горизонтали

-  – растянуть по горизонтали
-  – растянуть по вертикали
-  – сжать по вертикали
-  – сохранить рисунок в формате bmp, jpg, emf, wmf.
-  – отобразить график(и) функции распределения
-  – отобразить график(и) функции плотности
-  временной ряд
-  отобразить эмпирическую функцию распределения
-  – ядерная оценка плотности распределения
-  – построить оценки Каплана-Мейера и Нельсона-Аалена

5. Моделирование выборок

Для моделирования выборок выбираем пункт меню «Моделирование» и в открывшемся выпадающем списке пункт «Выборка», как показано на рисунке 33.

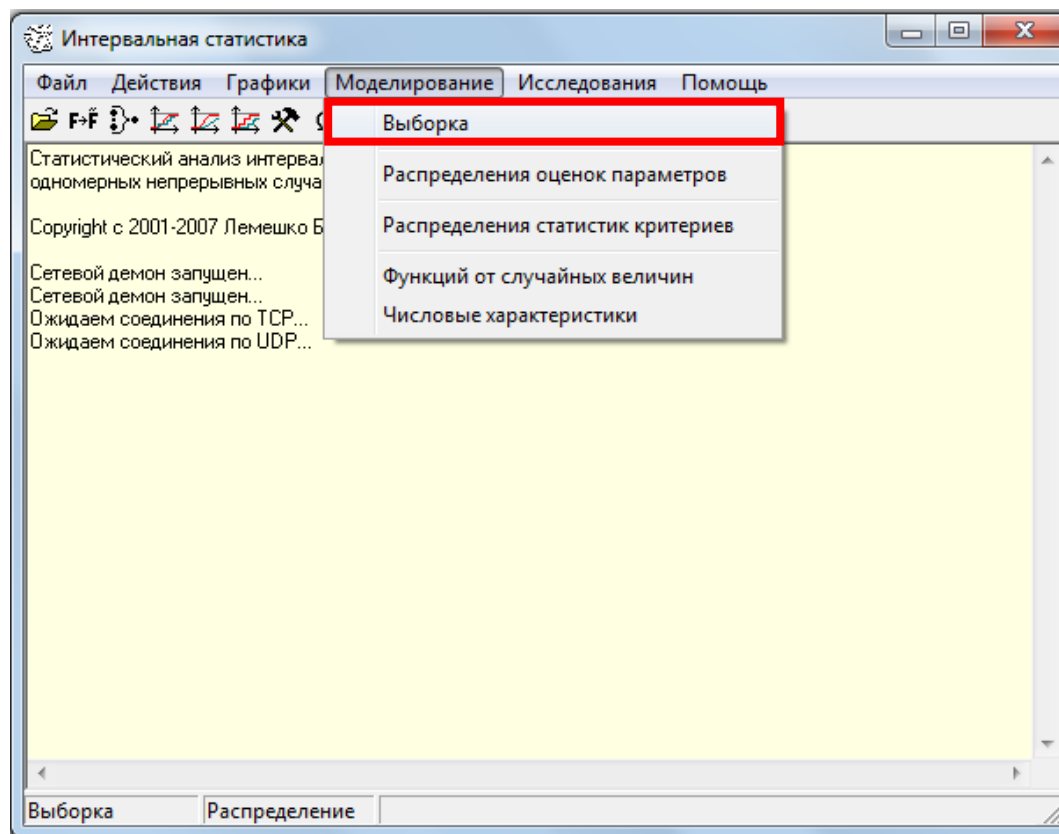


Рисунок 33 – Моделирование выборок

В открывшейся форме (см. рис. 34) тип создаваемой выборки задается выбором одной из четырех закладок: "Точечная", "Группированная", "Интервальная", "Цензурированная".

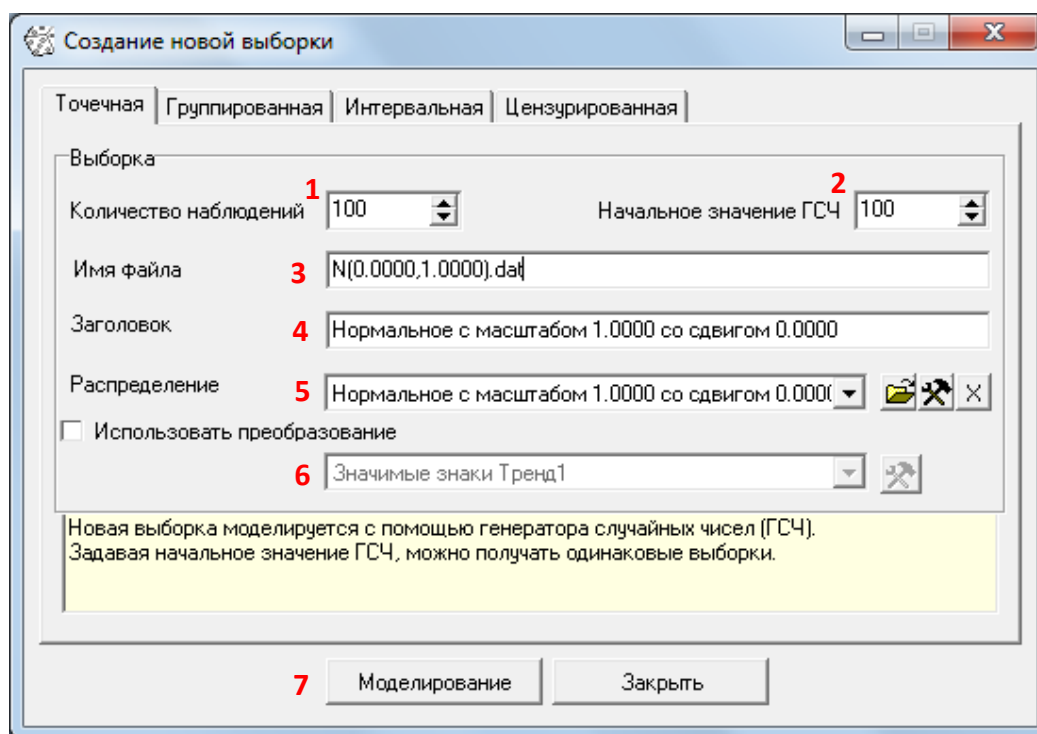




Рисунок 34 – Моделирование точечной выборки

Для **точечной выборки** выбирается:

1. объем выборки (количество наблюдений);
2. начальное значение генератора случайных чисел позволяет получать одинаковые выборки (т.е. чтобы получить ту же самую выборку второй раз, можно просто запомнить начальное значение ГСЧ);
3. имя файла сохраняемой выборки по умолчанию формируется по закону распределения случайной величины (если смоделировать по одному закону распределения с одинаковыми параметрами две выборки разных объемов, то первая выборка будет перезаписана, если не поменять название файла);
4. заголовок (первая строка в сохраняемом файле с выборкой);
5. распределение случайной величины (в системе заложено более 30 стандартных распределений и возможность добавлять новые распределения, получаемые из стандартных с помощью операций сдвига, масштабирования, смеси, произведения, зеркального отображения, усечения. По умолчанию в списке отображаются те распределения, которые перечислены в разделе [Distributions] в файле инициализации

«is.ini». В нашем примере – это стандартный нормальный закон $N(0,1)$.

Для изменения параметров закона необходимо нажать кнопку  «Параметры» и в открывшемся окне ввести значения параметров закона (см. рис. 35). Можно открыть другой (подготовленный ранее) список распределений , он задается в файле с расширением «dst» (см. рис. 36), например, Все. dst содержит все встроенные в систему распределения.

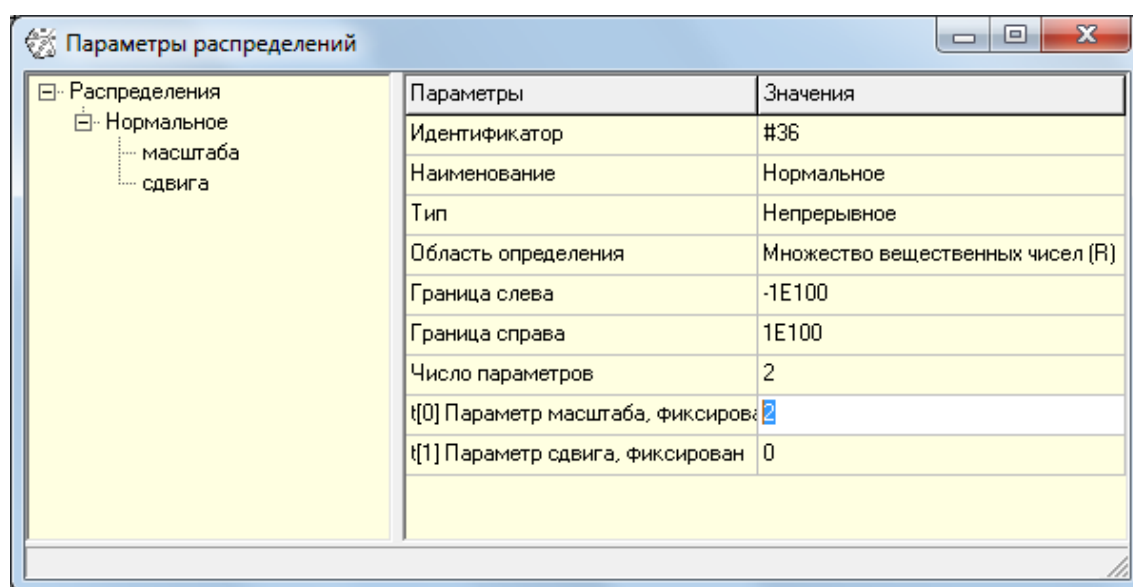


Рисунок 35 –Параметры закона

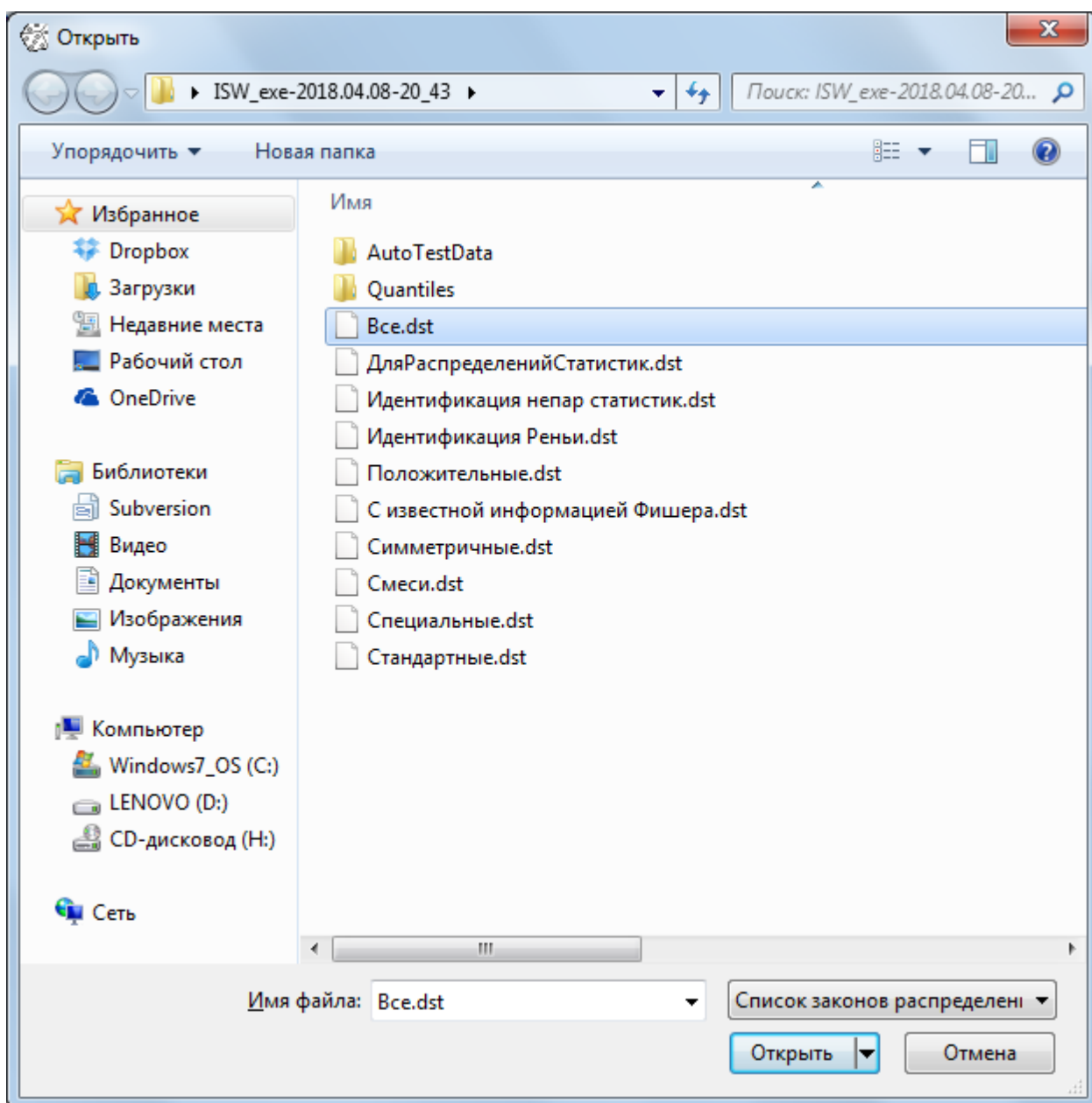



Рисунок 36 – Загрузка закона распределения из заранее подготовленных файлов

6. В случае, когда пользователь не желает добавлять в моделируемые значения тренд или округлять эти значения, галочка «Использовать преобразование» не должна быть поставлена, и пользователь переходит к пункту 7. В противном случае, поставив галочку «Использовать преобразование», пользователь выбирает в выпадающем списке то преобразование, которое хочет применить к данным (см. рис. 37), и нажимает кнопку  для настройки параметров.

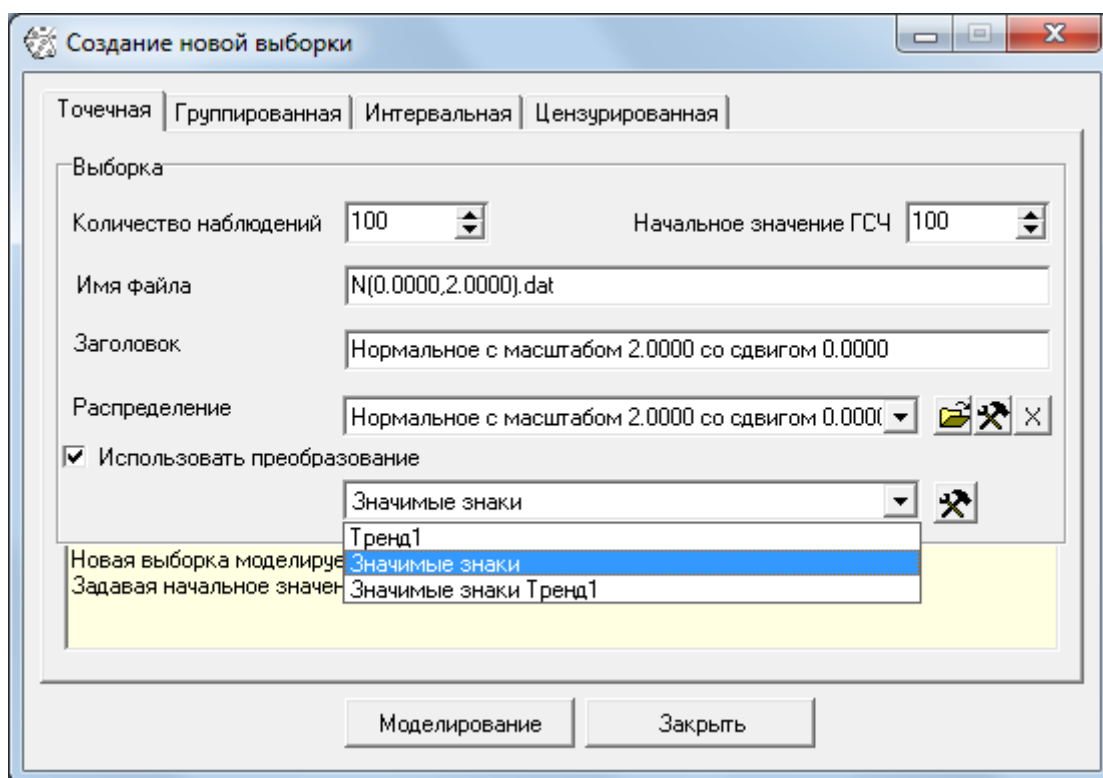


Рисунок 37 – Выбор преобразования

Задание преобразования «Тренд 1» рассматривается отдельным разделом. Для преобразования «Значимые знаки» задается параметр «порядок значимости» (см. рис. 38).

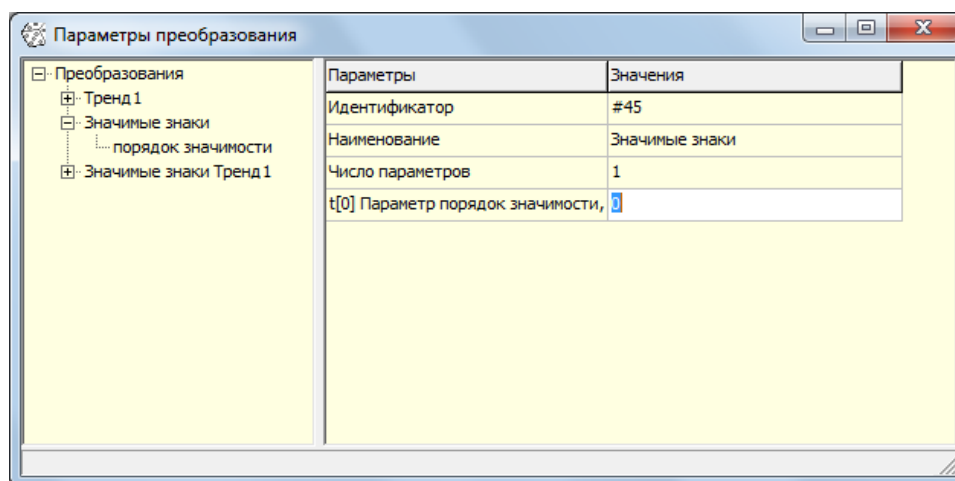


Рисунок 38 – Параметры преобразования «Значимые знаки»

7. Нажимаем кнопку «Моделирование», выборка автоматически будет сохранена, и в главном окне появится сообщение об успешном окончании моделирования (см. рис.39).

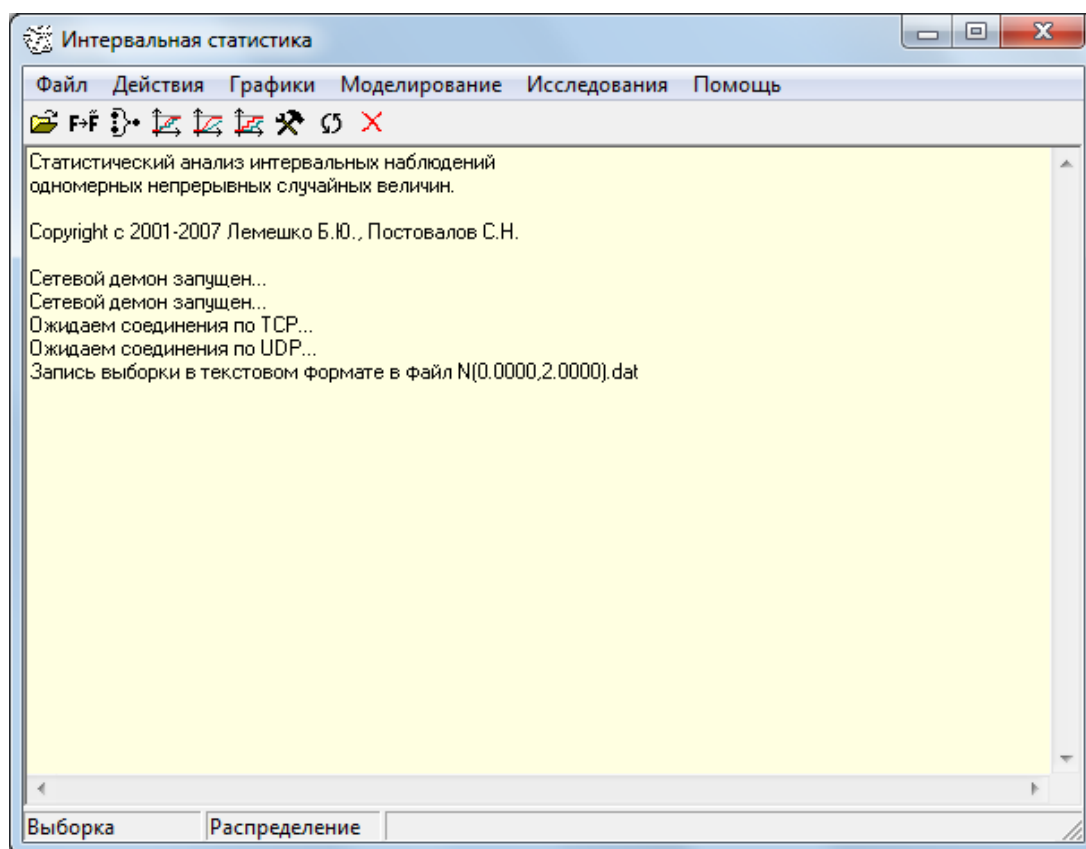


Рисунок 39 – Сообщение о смоделированной выборке

Для моделирования **группированной выборки** необходимо задать интервалы группирования. Количество и граничные точки можно вводить вручную, либо используя процедуру асимптотического группирования (кнопка "АОГ") равновероятного группирования (кнопка "РВГ") либо равномерного группирования (кнопка "РГ").

При создании **цензурированной выборки** определяется вид группирования (слева, справа или с обеих сторон), тип группирования (первый или второй) и, в зависимости от типа группирования – количество наблюдений в интервалах, или точки цензурирования.

5.1. Задание тренда

В **Тренд1** реализуется модель с параметрами, позволяющими задать скачки в математическом ожидании и дисперсии, а также линейный, периодический и смешанный тренд. Параметры сгруппированы так, чтобы можно было последовательно вводить сначала параметры для тренда в среднем, а затем для тренда в дисперсии.

В форме **Тренд1** $t[0]$, $t[1]$ – задают, соответственно, моменты начала и конца тренда в процентах %.

Модель с параметрами (в форме эти параметры соответствуют задаваемым параметрам $t[2], \dots, t[13]$) выглядит следующим образом:

$$X_i = \theta_2 + \theta_3 t_i + \theta_4 t_i^2 + \theta_5 \sin(\theta_6 2\pi t_i + \theta_7) + \xi_i(1 + \theta_8 + \theta_9 t_i + \theta_{10} t_i^2 + \theta_{11} \sin(\theta_{12} 2\pi t_i + \theta_{13}))$$

Параметр θ_2 – позволяет задать «скачок» в среднем значении, θ_8 – задать величину скачка в масштабе (в дисперсии).

Тренд в математическом ожидании		Область задания
θ_2	Параметр сдвига среднего	$(-\infty, +\infty)$
θ_3	Параметр, масштаб линейной составляющей	$(-\infty, +\infty)$
θ_4	Параметр, масштаб квадратичной составляющей	$(-\infty, +\infty)$
θ_5	Параметр, масштаб периодической составляющей	$(-\infty, +\infty)$
θ_6	Параметр, число периодов	$(0, +\infty)$
θ_7	Параметр, фаза	$(-\pi, +\pi)$
Тренд в дисперсии (ср.кв. отклонении)		Область задания
θ_8	Параметр сдвига в дисперсии	$(-1, +\infty)$
θ_9	Параметр, масштаб линейной составляющей	$(-1, +\infty)$
θ_{10}	Параметр, масштаб квадратичной составляющей	$(-1, +\infty)$
θ_{11}	Параметр, масштаб периодической составляющей	$(-1, +\infty)$
θ_{12}	Параметр, число периодов	$(0, +\infty)$
θ_{13}	Параметр, фаза	$(-\pi, +\pi)$

Примечание. Убывающую дисперсию можно задать отрицательными значениями параметров $\theta_8 \div \theta_{11}$. При этом сумма $\theta_8 + \theta_9 + \theta_{10} + \theta_{11}$ должна быть больше -1 .

Параметры преобразования в форме показаны на рисунке 40.

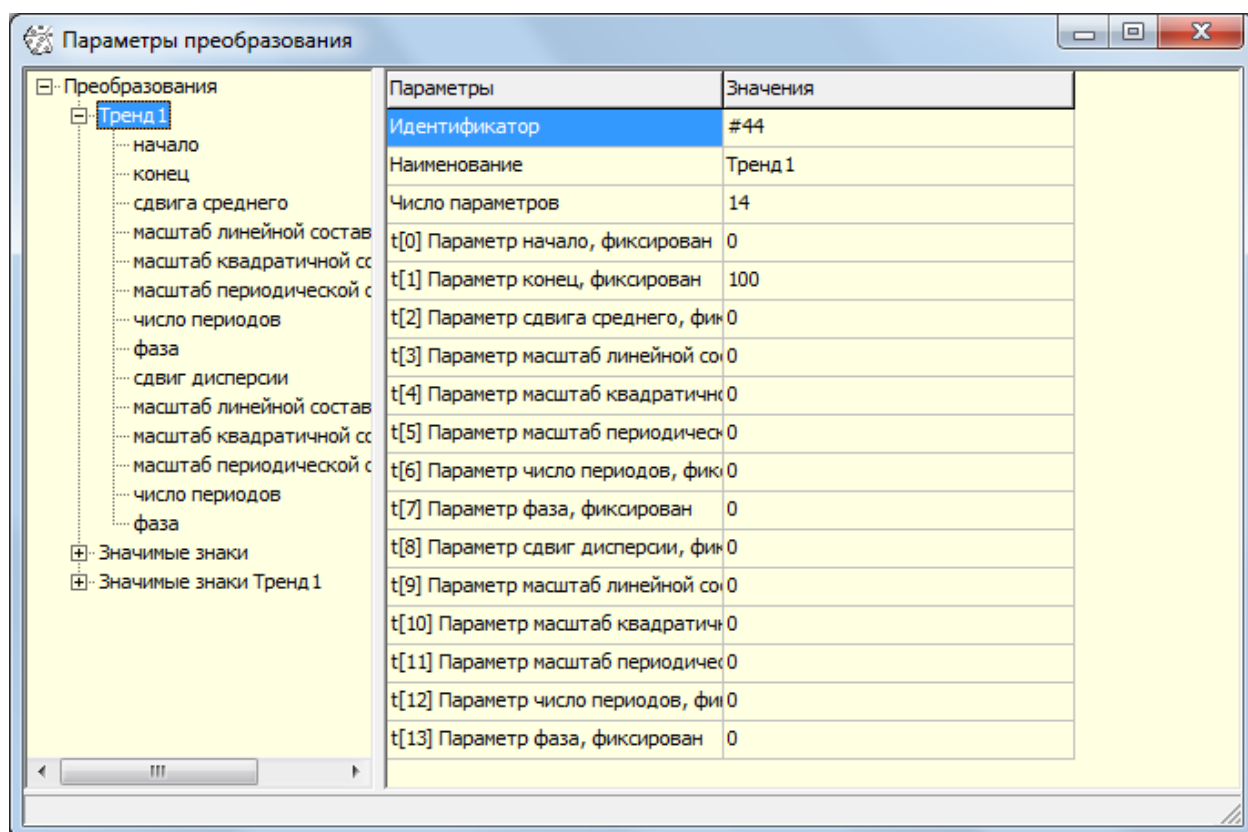


Рисунок 40 – Задание параметров модели тренда

5.2. Моделирование функций от случайных величин

Точнее, моделирование эмпирических распределений функций от случайных величин.

Перейти в данную вкладку можно следующим образом: «Моделирование» -> «Функций от случайных величин». Интерфейс позволяет задавать произвольные функции от системы независимых случайных величин, распределенных по различным одномерным законам. На рисунке 41 показана вкладка «Моделирование функций от случайных величин», отражающая описание пользовательского интерфейса:

- 1 – модуль выбора распределения из загруженных в систему законов распределения;
- 2 – модуль выбора случайной величины из ещё неиспользуемых;

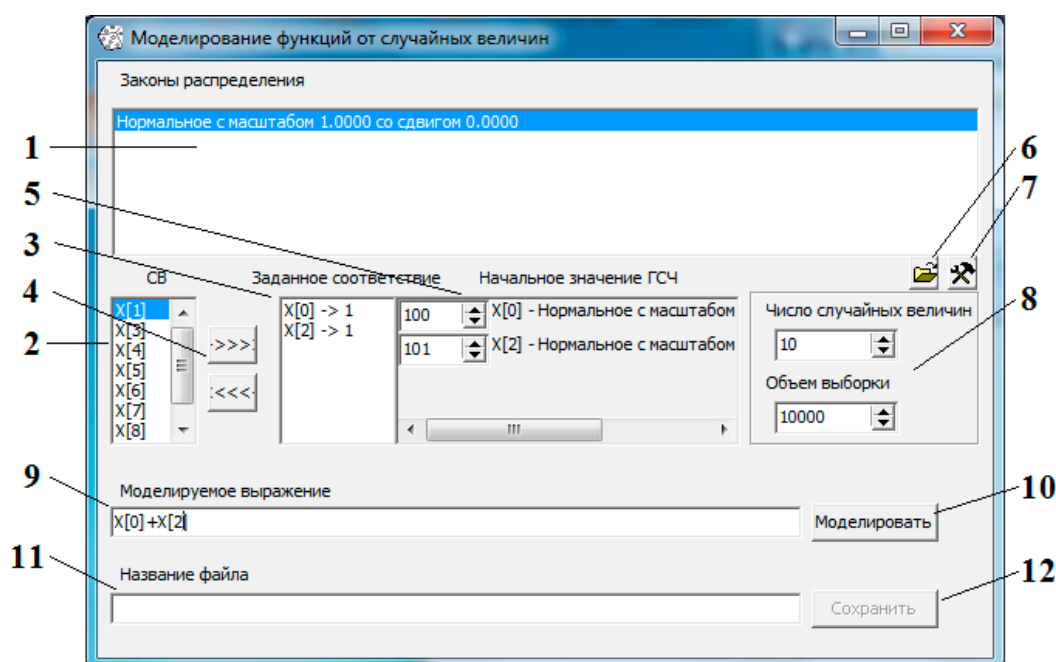


Рисунок 41 – Интерфейс «Моделирование функций от случайных величин»

- 3 – модуль, содержащий выбранные случайные величины. В модуле показано соответствие этих величин с ID распределения (не виден пользователю);
- 4 – клавиши для добавления и удаления случайных величин в модуле 3;
- 5 – модуль, содержащий информацию обо всех выбранных случайных величинах: начальном значении генератора случайных чисел, который можно менять, и о распределении случайной величины, включая параметры;
- 6 – клавиша для загрузки распределений;
- 7 – клавиша для изменения параметров распределений. После ее нажатия появляется форма, показанная на рисунке 42 (аналогично рисунку 35);

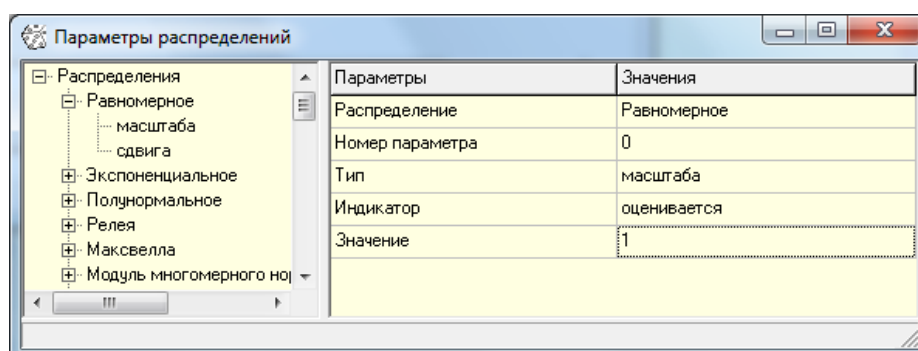


Рисунок 42 – Форма «Параметры распределений»

- 8 – Модуль для изменения параметров моделирования (максимального количества используемых случайных величин и объемов генерируемых выборок);
- 9 – модуль для записи моделируемого выражения. Операции доступные в этом модуле будут представлены далее;
- 10 – клавиша запуска моделирования функции от случайных величин. В случае некорректной записи выражения в модуль 9 будет выдана ошибка, и моделирование не запустится;
- 11 – модуль записи имени файла для сохранения результирующей выборки. После окончания моделирования имя файла будет сгенерировано по умолчанию в соответствии с видом моделируемого выражения;
- 12 – клавиша для подтверждения сохранения уже смоделированной выборки в файл с именем, указанным в 11.

Вид функций от независимых случайных величин может быть “произвольным” и задается пользователем в режиме диалога. В программной системе доступны следующие операции:


- Бинарные операции: сложение (+), вычитание (–), умножение (*), деление (/), деление по модулю (%), возведение в степень (^);
- унарные операции: унарный минус, извлечение квадратного корня (sqrt), взятие модуля (abs), тригонометрические операции (sin, cos, tg, ctg, arcsin, arccos, arctg, arcctg, sh, ch, th, cth, exp), логарифмические операции (lg, ln);
- *n*-арные операции: нахождение минимума (min), нахождение максимума (max), нахождение среднего (avg) и суммы (sum) множества случайных величин.
- в системе также доступны известные константы e (e) и π (pi);

Помимо простой моделируемой функции, задание которой показано на рисунке 41, можно привести примеры задания более сложных n -арных операций:

- $\text{avg}(X[i(0)], i(0)=0:10)$ – среднее значения 11 элементов X с индексами от 0 до 10 включительно;
- $\text{min}(X[i(0)], i(0)=0:29)$ – поиск минимума среди 30 элементов X с индексами от 0 до 29 включительно;
- синтаксис поиска максимума аналогичный;
- $\text{Sum}(X[i(0)], i(0)=0:24) + \text{Sum}(X[i(1)], i(1)=50:75)$ – нахождение суммы элементов X с индексами от 0 до 24 и от 50 до 75 включительно.

Важно! Использование элемента с индексом 30 может привести к некорректным данным. Данное значение зарезервировано системой и выборка, соответствующая данному элементу, будет состоять из нулей.

6. Форма «Параметры»

Переход к форме «Параметры» осуществляется через кнопку  на панели инструментов главного окна программы или через пункт меню «Действия» → «Параметры». Большая часть инструментария формы «Параметры» продублирована в других формах для удобства. Данная форма состоит из 7 вкладок:

- Распределения (не предназначено для рядового пользователя);
- Выборки (не предназначено для рядового пользователя);
- Задания (не предназначено для рядового пользователя);
- Оценивание (см. Рис 43). Данная форма позволяет менять методы оценивания, тип ядерных оценок;
- Критерии (см. Рис 44). Данная форма позволяет список критериев согласия при проверке гипотез, тип гипотезы о согласии (простая/сложная), уровень значимости (изменение данного параметра

нигде не реализовано кроме данной формы);

- Поиск (см. Рис 45). Данная форма позволяет менять методы поиска оптимальных значений;
- Группирование (см. Рис 46). Данная форма позволяет менять методы группирования, и число интервалов группирования;

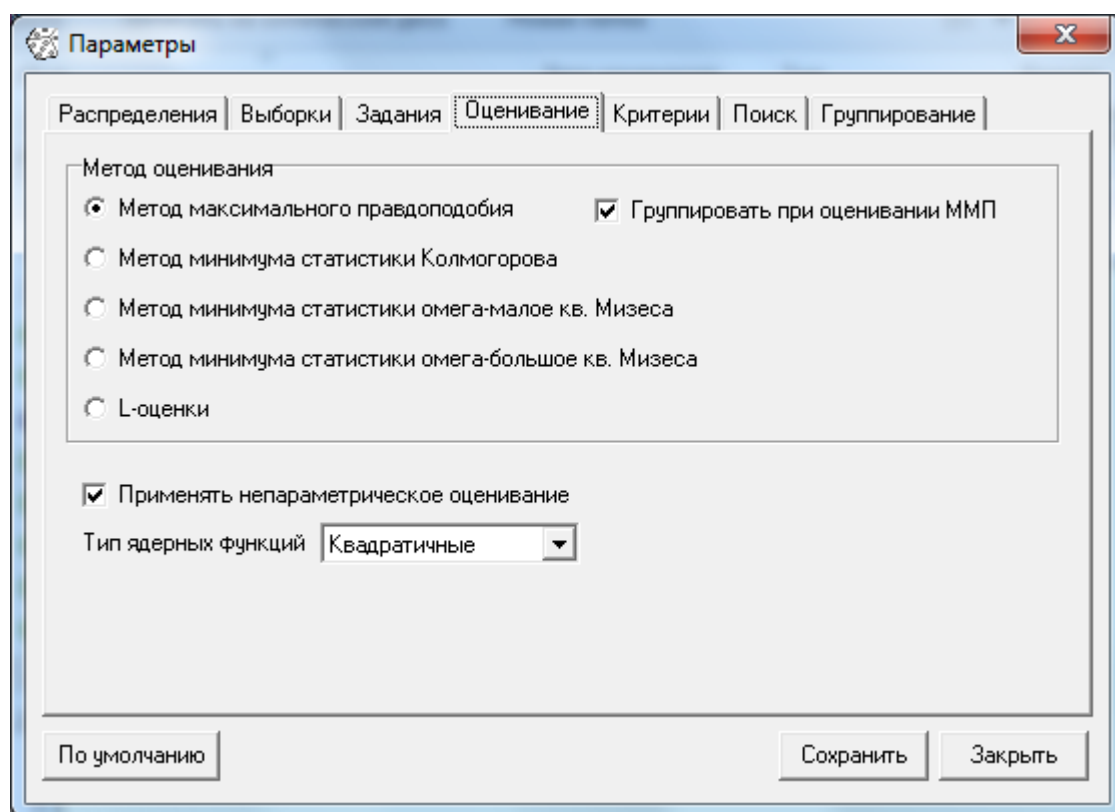


Рисунок 43 – Вкладка «Оценивание» в форме «Параметры распределений»

Примечание: «Галочка» при «Группировать при оценивании», как правило, должна быть убрана. Использовать при оценивании по большим выборкам и в предположении о возможном присутствии в выборке аномальных с позиций предполагаемого закона измерений.

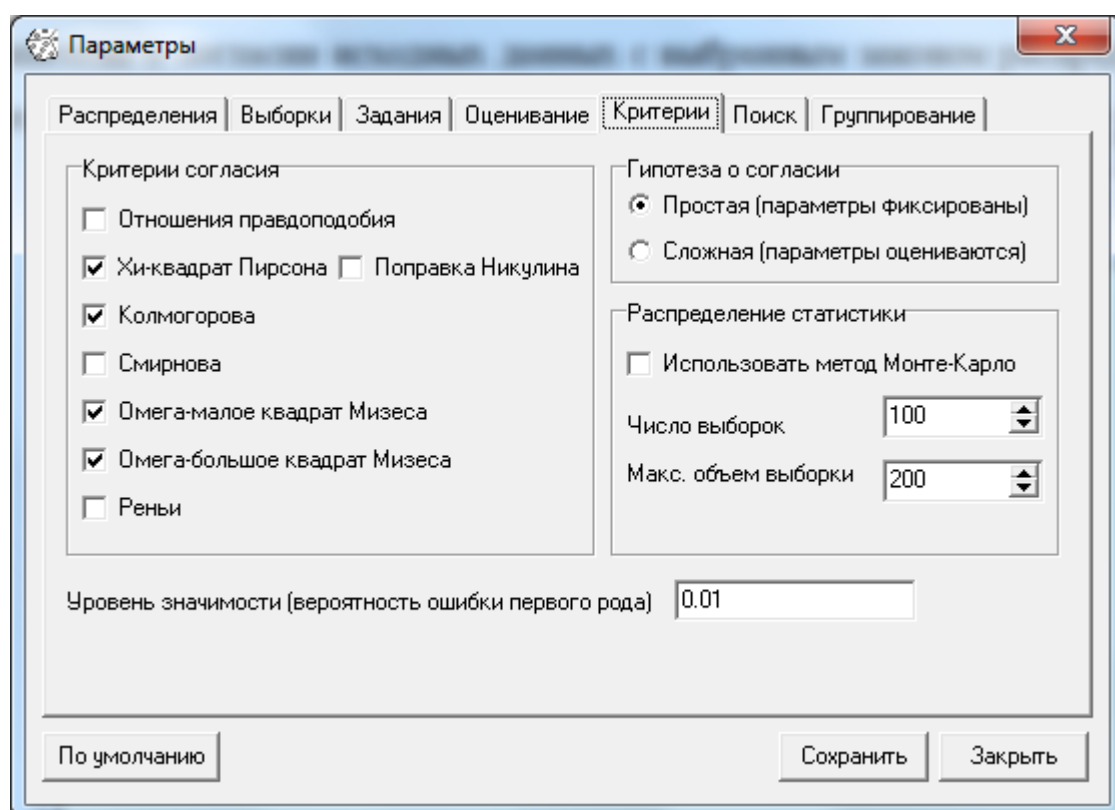


Рисунок 44 – Вкладка «Критерии» в форме «Параметры распределений»

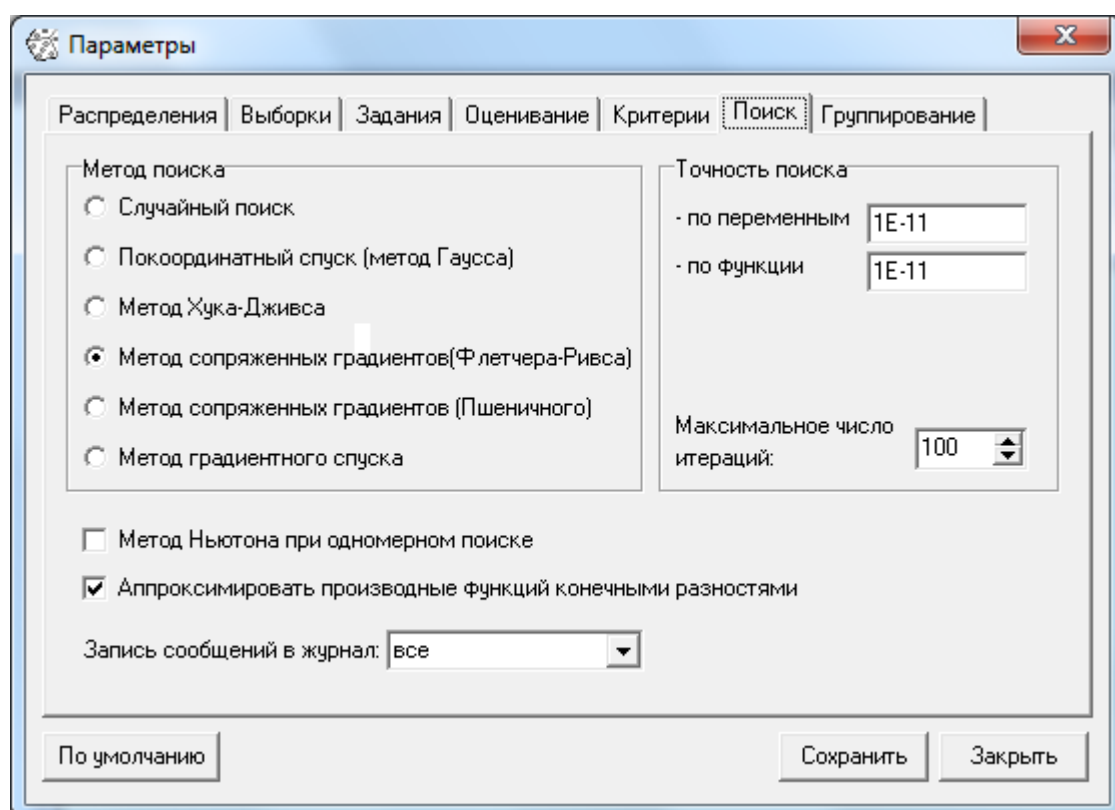


Рисунок 45 – Вкладка «Поиск» в форме «Параметры распределений»

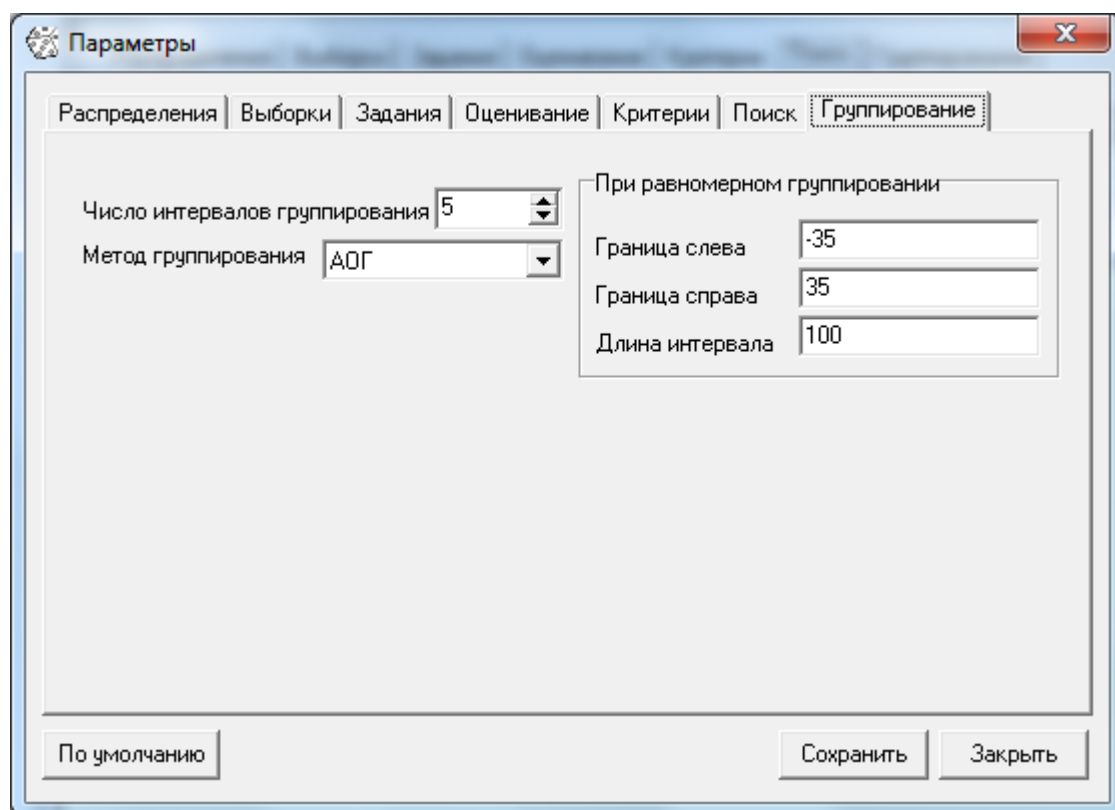
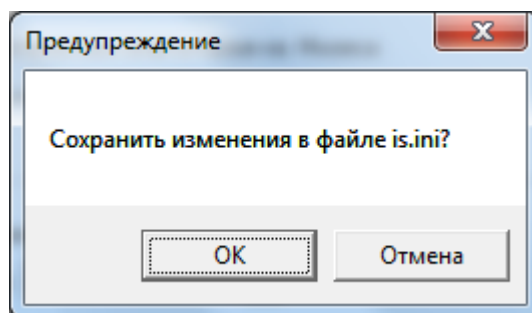
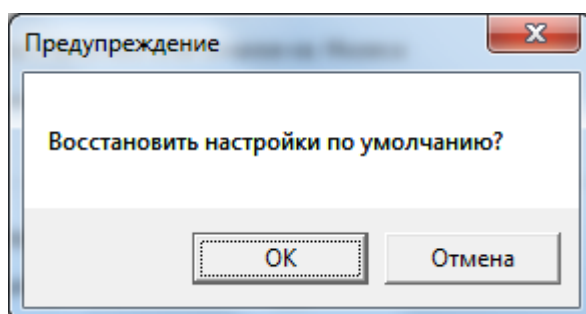


Рисунок 46 – Вкладка «Группирование» в форме «Параметры распределений»

Во всех реализованных вкладках данной формы присутствует клавиша «Сохранить», при нажатии которой пользователю предоставляется возможность перезаписать выбранные настройки в файл is.ini.



Также есть возможность откатить все изменения в файле is.ini и вернуть изначальные настройки нажатием клавиши «По умолчанию». При нажатии данной клавиши появляется следующее диалоговое окно.



7. Полезные ссылки для изучения подробностей о критериях

1. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход: Монография / Б.Ю. Лемешко, С.Б. Лемешко, С.Н. Постовалов, Е.В. Чимитова. – Новосибирск : Изд-во НГТУ, 2011. – 888 с. (серия «Монографии НГТУ»).

Текст: https://ami.nstu.ru/~headrd/seminar/publik_html/Statistical_Data_Analysis.pdf

2. Лемешко Б.Ю. Непараметрические критерии согласия: Руководство по применению: Монография / Б.Ю. Лемешко. – М.: ИНФРА-М, 2014. – 163 с. DOI: 10.12737/11873

Препринт: https://ami.nstu.ru/~headrd/seminar/publik_html/guid_nepar_gof_test.pdf

3. Лемешко Б.Ю. Критерии проверки отклонения распределения от нормального закона. Руководство по применению: Монография / Б.Ю. Лемешко. – М.: ИНФРА-М, 2015. – 160 с. – (Научная мысль). DOI: 10.12737/6086

Препринт: https://ami.nstu.ru/~headrd/seminar/publik_html/guid_normal_tets.pdf

4. Лемешко Б.Ю., Блинов П.Ю. Критерии проверки отклонения распределения от равномерного закона. Руководство по применению: Монография / Б.Ю. Лемешко, П.Ю. Блинов. – М.: ИНФРА-М, 2015. – 183 с. – (Научная мысль). DOI: 10.12737/11304

Препринт: https://ami.nstu.ru/~headrd/seminar/publik_html/test_random_lection.pdf

5. Лемешко Б.Ю. Критерии проверки гипотез об однородности. Руководство по применению : монография / Б.Ю. Лемешко. – М. : ИНФРА-М, 2017. – 208 с. – (Научная мысль). – ISBN: 978-5-16-012557-2 (print), 978-5-16-105463-5(online) DOI: 10.12737/22368

6. Лемешко Б.Ю. Критерии проверки гипотез об однородности. Руководство по применению : монография / Б.Ю. Лемешко. – 2-е изд., перераб. и доп. – Москва : ИНФРА-М, 2021. – 248 с. – (Научная мысль). – DOI [10.12737/986695](https://doi.org/10.12737/986695), ISBN 978-5-16-016336-9 (print), ISBN 978-5-16-108633-9 (online)

Расширенный препринт:

https://ami.nstu.ru/~headrd/seminar/publik_html/Guide_homogeneity.pdf

7. Лемешко Б.Ю. Критерии проверки отклонения от экспоненциального закона. Руководство по применению : монография / Б.Ю. Лемешко, П.Ю. Блинов. – Москва : ИНФРА-М, 2021. – 352 с. – (Научная мысль). –DOI [10.12737/1097477](https://doi.org/10.12737/1097477), ISBN 978-5-16-016328-4 (print), ISBN 978-5-16-108625-4 (online)

8. Лемешко Б.Ю., Веретельникова И.В. Критерии проверки гипотез о случайности и отсутствии тренда. Руководство по применению: Монография / Б.Ю. Лемешко, И.В. Веретельникова. – Москва : ИНФРА-М. 2021. – 221 с. – (Научная

мысль). DOI [10.12737/1587437](https://doi.org/10.12737/1587437) ISBN 978-5-16-017054-1 (print), ISBN 978-5-16-109774-8 (online)

9. Лемешко Б.Ю., Постовалов С.Н. [Компьютерные технологии анализа данных и исследования статистических закономерностей: Учебное пособие.](#) – Новосибирск: Изд-во НГТУ, 2004. – 119 с.

Текст: https://ami.nstu.ru/~headrd/seminar/publik_html/Pos_KTAD.pdf

Более обширную информацию смотри на: <https://ami.nstu.ru/~headrd/>
или <https://ami.nstu.ru/~headrd/seminar/start.htm>

С замечаниями и вопросами обращаться к:

Лемешко Борису Юрьевичу

E-mail: Lemeshko@ami.nstu.ru