Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего профессионального образования

«Волгоградский государственный технический университет»

(ВолгГТУ)

Кафедра: «Высшая математика»

Протокол лабораторной работы №2

по дисциплине «Теория вероятностей, математическая статистика и случайные процессы»

На тему: “Проверка статистических гипотез”

Подготовили студенты группы ИВТ – 360

ФИО студентов: Князев А. А.

Скориков А.В.

ФИО преподавателя: Андреева М.И.

г. Волгоград, 2014

**Теоретическая часть**

**Проверка гипотезы о нормальном распределении генеральной совокупности по критерию Пирсона**

***А. Эмпирическое распределение задано в виде последовательности равностоящих вариант и соответствующих им частот.***

Пусть эмпирическое распределение задано в виде последовательности равностоящих вариант и соответствующих им частот:



Требуется, использовать критерий Пирсона, проверить гипотезу о том, что генеральная совокупность ***X*** распределена нормально.

**Правило 1.**

Для того чтобы при заданном уровне значимости α проверить гипотезу о нормальном распределении генеральной совокупности, надо:

1. Вычислить непосредственно (при малом числе наблюдений) или упрощенным методом (при большом числе наблюдений), например методом произведений или сумм, выборочную среднюю и выборочное среднее квадратическое отклонение .

1. Вычислить теоретические частоты

где ***n –*** объем выборки (сумма всех частот), ***h –*** шаг (разность между двумя соседними вариантами),

1. Сравнить эмпирические и теоретические частоты с помощью критерия Пирсона.

а) Составить расчетную таблицу для определения наблюдаемого значения критерия



б) По таблице критических точек распределения , по заданному уровню значимости α и числу степеней свободы (***s*** – число групп выборки) находят критическую точку  правосторонней критической области.

Если  - нет оснований отвергнуть гипотезу о нормальном распределении генеральной совокупности. Другими словами, эмпирические и теоретические частоты различаются незначимо (случайно). Если  - гипотезу о нормальном распределении генеральной совокупности отвергают. Другими словами, эмпирические и теоретические частоты различаются значимо.

***Б. Эмпирическое распределение задано в виде последовательности интервалов одинаковой длины и соответствующих им частот.***

Пусть эмпирическое распределение задано в виде последовательности интервалов  и соответствующих им частот ***ni*** (***ni –*** сумма частот которые попали в ***i –*** ый интервал):



Требуется, использовать критерий Пирсона, проверить гипотезу о том, что генеральная совокупность ***X*** распределена нормально.

**Правило 2.**

Для того чтобы при заданном уровне значимости α проверить гипотезу о нормальном распределении генеральной совокупности, надо выполнить следующие действия:

1. Вычислить, выборочную среднюю и выборочное среднее квадратическое отклонение , причем в качестве вариант принимают среднее арифметическое концов интервала:

где – середина частичного интервала.

1. Пронормировать ***X,*** т.е. перейти к случайной величине , причем наименьшее значение полагают равным (), а наибольшее - ().
2. Вычислить теоретические частоты , где ***n –*** объем выборки (сумма всех частот), – вероятности попадания ***X*** в интервалы , – функция Лапласа.
3. Сравнить эмпирические и теоретические частоты с помощью критерия Пирсона.

а) Составить расчетную таблицу для определения наблюдаемого значения критерия



б) По таблице критических точек распределения , по заданному уровню значимости α и числу степеней свободы (***s*** – число групп выборки) находят критическую точку  правосторонней критической области.

Если  - нет оснований отвергнуть гипотезу о нормальном распределении генеральной совокупности. Другими словами, эмпирические и теоретические частоты различаются незначимо (случайно). Если  - гипотезу о нормальном распределении генеральной совокупности отвергают. Другими словами, эмпирические и теоретические частоты различаются значимо.

**Проверка гипотезы о биномиальном распределении генеральной совокупности по критерию Пирсона**

Пусть произведено ***n*** опытов. Каждый опыт состоит из ***N*** независимых испытаний, в каждом из которых вероятность события ***A*** одна и та же. Регистрируется число появлений события ***A*** в каждом опыте. В итоге получено следующее распределение дискретной случайной величины ***X*** – числа появлений события ***A*** (в первой строке указано число ***xi*** появлений события ***A*** в одном опыте; во второй строке – частота ***ni*** т.е. число опытов, в которых зарегистрировано ***xi*** появлений события ***A***):



Требуется, использовать критерий Пирсона, проверить гипотезу о том, что генеральная совокупность ***X*** распределена по биномиальному закону.

**Правило.**

Для того чтобы при заданном уровне значимости α проверить гипотезу о том, что дискретная случайная величина ***X*** (число появлений события ***A***) распределена по биномиальному закону, нужно:

1. Найти по формуле Бернулли вероятности ***Pi*** появления ровно ***i*** событий ***A*** в ***N*** испытаниях (***i*** = 0, 1, 2,…, ***s***, где ***s*** – максимальное число наблюдавшихся появлений события ***A*** в одном опыте, т.е. ***s ≤ N***).



1. Найти теоретические частоты , где ***n*** – число опытов.

3.1. Сравнить эмпирические и теоретические частоты с помощью критерия Пирсона, приняв число степеней свободы  (при этом предполагается, что вероятность ***Pi*** появления события ***A*** задана, т.е. не оценивалось по выборке)

3.2. Если же вероятность ***Pi*** была оценена по выборке, то .

**Сравнение двух дисперсий нормальных генеральных совокупностей**

По независимым выборкам, объемы которых ***n*1*, n*2*,*** извлеченным из нормальных генеральных совокупностей найдены исправленные выборочные дисперсии  и . Требуется сравнить эти дисперсии.

**Правило 1.**

Для того чтобы при заданном уровне значимости α проверить нулевую гипотезу  о равенстве генеральных дисперсий нормальных совокупностей при конкурирующей гипотезе  надо вычислить наблюдаемое значение критерия (отношение большей исправленной дисперсии к меньшей)



И по таблице критических точек Фишера-Снедекора, по заданному уровню значимости α и числам степеней свободы  (***k*1** – число степеней свободы большей исправленной дисперсии) найти критическую точку .

Если  - нет оснований отвергнуть нулевую гипотезу. Если  - нулевую гипотезу отвергают.

**Правило 2.**

При конкурирующей гипотезе  критическую точку  определяют по уровню значимости α/2 (вдвое меньше заданного) и числам степеней свободы ***k*1** и ***k*2** (***k*1** – число степеней свободы большей исправленной дисперсии).

Если  - нет оснований отвергнуть нулевую гипотезу. Если  - нулевую гипотезу отвергают.

**Практическая часть**

1. **Несколько слов о необходимом программном обеспечении**

Данная программа была написана на языке программирования C++ с применением библиотеки Qt 4.8.5 в среде разработки Microsoft Visual Studio 2008. Следовательно, она без проблем может работать на любых современных операционных системах семейства Windows (начиная с Windows XP).

Наличие в комплекте с основным файлом для запуска (TerVer\_Lab1.exe) ряда библиотек, позволяет сделать данную программу полностью независимой от программного обеспечения.

С точки зрения разработчика может показаться, что часть этих библиотек не является необходимой для нормальной работы программы, поскольку они уже входят в стандартный пакет операционной системы Windows. Тем не менее, это было сделано специально для того, чтобы полностью избавиться от любых возможных проблем в этой области.

Помимо этого хочется отметить необходимость наличия в папке с программой двух файлов: “ Table\_Laplace.txt ” и “ Table\_Pearson.txt ”. Эти файлы являются обязательными и их удаление, перемещение или какие-либо другие модификации (в том числе и изменение имени) могут привести к непредсказуемым последствиям, скорее всего программа потеряет свою работоспособность. Но, если учитывать тот факт, что эти программы находятся в одном каталоге с общим проектом, то при стандартной работе подобных неприятностей происходить вообще не должно.

Таким образом, для корректной работы программы пользователю необходимо просто скопировать папку с данной программой в любое удобное для него место на своем персональном/рабочем компьютере и использовать её по назначению.

1. **Описание программы**

Данная программа включает в себя две достаточно полезные функции: позволяет производить первичную статистическую обработку данных, и непосредственно реализует проверку статистических гипотез.

На первый взгляд программа может показаться достаточно сложной, из-за наличия большого количества окон и полей, но на самом деле это не так. Большая часть всех полей и окон заполняется программой автоматически, от пользователя всего лишь требуется предоставить достоверные статистические данные и минимальный набор характеристик для их описания (см. Приложение 1).

В общем случае программа имеет четыре страницы расположенных в одном окне. Каждая страница отвечает за определенный набор вычислений и имеет соответствующее исполняемым действиям название. Это страницы: “Первичная обработка статистических данных”, “Проверка гипотезы о нормальном распределении”, “Проверка гипотезы о биномиальном распределении”, “Сравнение дисперсий двух нормальных распределений”. Следовательно, это именно тот набор действий, которые пользователь может реализовать посредством использования данной программы.

Помимо этого программа имеет три дополнительных окна: “Таблица данных” – здесь отображаются экспериментальные данные, которые считываются из файла в случае их непосредственной загрузки, также это окно можно вызвать в любое время нажав на кнопку “Таблица обработки статистических данных”; окно “Справка” – здесь представлена общая информация необходимая пользователю для работы с данной программой (с этими данными желательно ознакомиться перед началом работы с программой); окно “Теоретический материал” – позволяет пользователю ознакомиться с теоретическим материалом соответствующим той странице, на которой он ведет свою работу на момент вызова этого окна. Также можно ознакомиться со всем теоретическим материалом по данной теме перемещаясь между страницами программы. В общем случае все эти окна носят справочный характер.

Её функционал позволяет пользователю работать именно так, как ему удобно, а не так как этого требует от него сама программа. Этот фактор значительно облегчает и ускоряет процесс взаимодействия пользователя и программы.

Таким образом, программа является простым, и в то же время удобным средством расчета.

1. **Примеры работы программы**

**Первичная обработка статистических данных**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 40,1 | 40,5 | 40,3 | 40,7 | 40,6 | 40,2 | 40,6 | 40,4 | 40,8 | 40,7 |
| 40,3 | 40,1 | 40,7 | 40,3 | 40,6 | 40,5 | 40,8 | 40,6 | 40,6 | 40,4 |
| 40,5 | 40,6 | 40,1 | 40,6 | 40,2 | 40,5 | 40,3 | 40,5 | 40,4 | 40,6 |
| 40,5 | 40,3 | 40,7 | 40,5 | 40,9 | 40,8 | 40,5 | 40,6 | 40,7 | 40,6 |
| 40,6 | 40,2 | 40,6 | 40,6 | 40,5 | 40,3 | 40,8 | 40,4 | 40,5 | 40,4 |
| 40,4 | 40,7 | 40,6 | 40,2 | 40,6 | 40,9 | 40,3 | 40,7 | 40,6 | 40,6 |
| 40,7 | 40,6 | 40,5 | 40,8 | 40,5 | 40,6 | 40,8 | 41,0 | 40,4 | 41,0 |
| 40,6 | 40,4 | 40,8 | 40,2 | 40,9 | 40,4 | 40,5 | 40,3 | 40,5 | 40,7 |
| 40,7 | 40,5 | 40,5 | 40,8 | 40,3 | 40,9 | 40,6 | 41,0 | 40,5 | 40,4 |
| 40,6 | 40,7 | 40,8 | 40,5 | 40,8 | 40,6 | 40,6 | 40,5 | 40,5 | 40,6 |

Таблица 1 - Вариант №7 из методического пособия «Методические указания к типовому расчёту "Обработка опытных данных в случае нормального закона распределения"»



Рисунок 1 – Окно «Таблица экспериментальных данных»

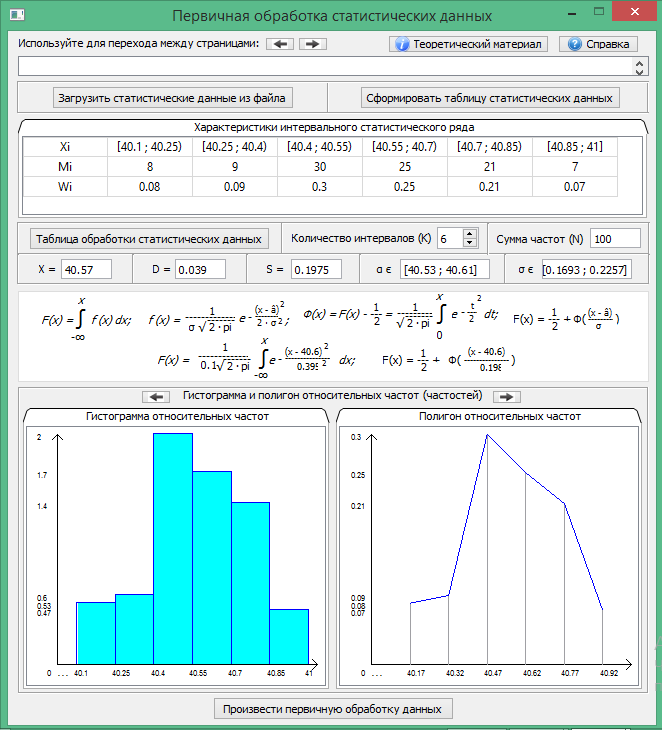


Рисунок 2 – Пример первичной обработки статистических данных

**Проверка гипотезы о распределении генеральной совокупности по нормальному закону**

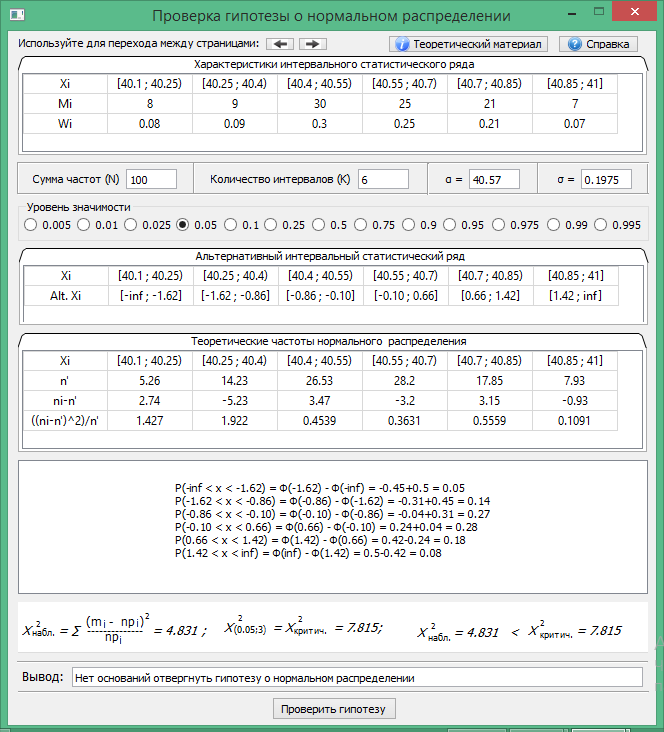


Рисунок 3 – Проверка гипотезы о распределении генеральной совокупности по нормальному закону

(на основе данных приложенных выше, см. стр.7)

**Проверка гипотезы о распределении генеральной совокупности по биномиальному закону**

Расчёт произведен по данным представленным в книге: Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике: Учеб.пособие – 12-е изд., перераб. - М.: Высшее образование, 2006.- 476 с. Задача № 652.

Задание: Произведено *n* = 100 опытов. Каждый опыт состоял из *N* = 10 испытаний, в каждом из которых вероятность *p* появления события *A* равна 0,3. В итоге получено следующее эмпирическое распределение.



Требуется при уровне значимости 0,05 проверить гипотезу о том, что дискретная случайная величина *X* (число появлений события *A*) распределена по биномиальному закону.

Ожидаемый результат:

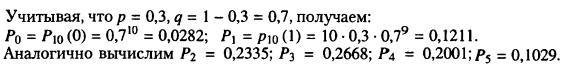


Рисунок 4 – Вероятность появления события *A* в *i* испытаниях из десяти

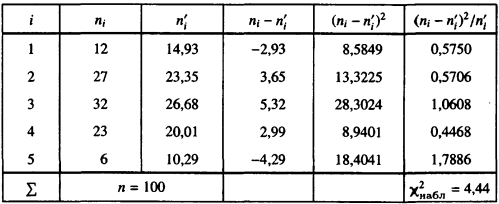


Рисунок 5 – Расчётная таблица



Рисунок 6 – Результат

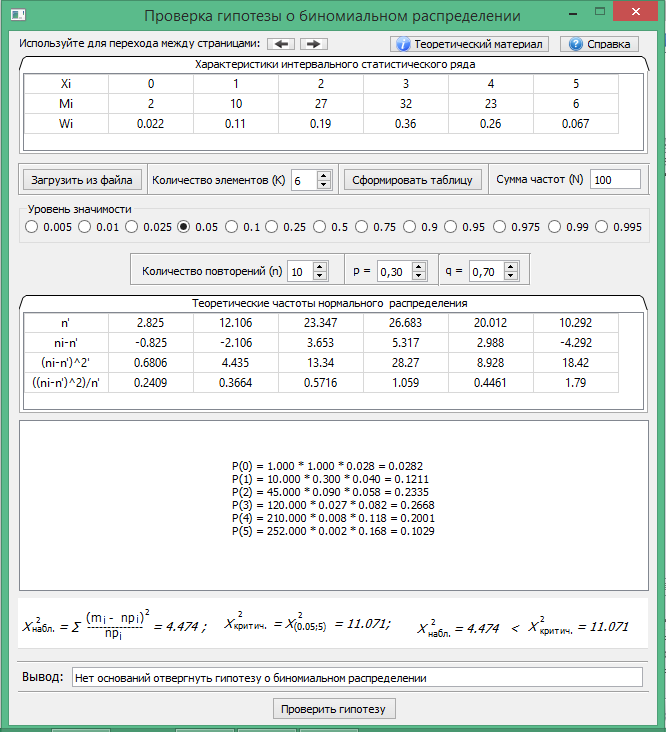


Рисунок 7 – Проверка гипотезы о распределении генеральной совокупности по биномиальному закону

**Сравнение двух дисперсий нормальных генеральных совокупностей**

Расчёт произведен по данным представленным в книге: Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике: Учеб.пособие – 12-е изд., перераб. - М.: Высшее образование, 2006.- 476 с. Задача № 559.

Задание: Для сравнения точности двух станков-автоматов взяты две пробы (выборки), объемы которых n1 = 10 и n2 = 8. В результате измерения контролируемого размера отобранных изделий получены следующие результаты:



Можно ли считать, что станки обладают одинаковой точностью [H0: D(X) = D(Y)], если принять уровень значимости α=0,1 и в качестве конкурирующей гипотезы [H1: D(X) ≠ D(Y)]

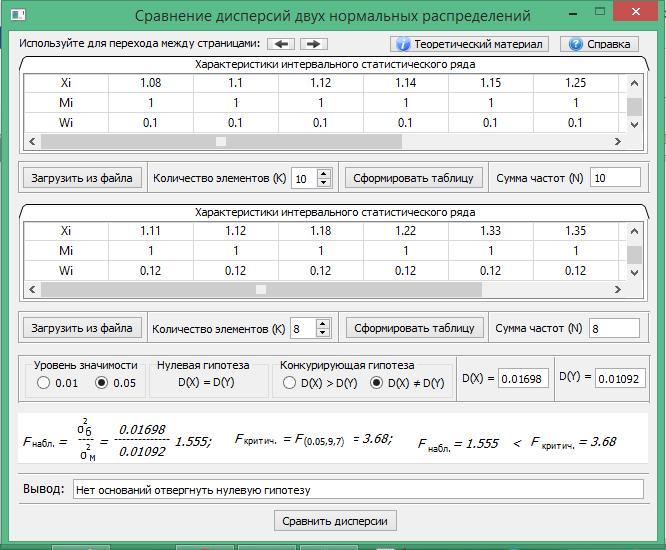


Рисунок 8 – Сравнение двух дисперсий нормальных генеральных совокупностей

**Приложение 1**

Справочные данные по использованию программы.

**1. Первый вход в программу**

Здравствуйте, уважаемый пользователь. Если Вы читаете данную главу то, скорее всего Вы сталкиваетесь с нашим приложением впервые. Поскольку в данной главе описываются элементарные действия работы из под ОС Windows, то если вы имеете соответствующие навыки можете перейти к следующей главе, в противном случае продолжайте чтение далее.

**Работа с кнопками**

Данное приложение решает следующие задачи:

* Осуществляет первичную обработку статистических данных;
* Осуществляет проверку гипотезы о распределении по нормальному закону;
* Осуществляет проверку гипотезы о биномиальном распределении;
* Осуществляет сравнение дисперсий двух нормальных распределений.

Все эти действия расположены на различных страницах, и потому первым (базовым) навыком, которому должен научиться пользователь – это перемещение между страницами. В этом нет ничего сложного! Существуют две кнопки  при помощи, которых Вы можете переходить на предыдущую и следующую страницу соответственно.

Следующим шагом является освоение работы с кнопками справочного и теоретического материала  и , здесь все также элементарно просто. И учитывая, что Вы уже читаете данную инструкцию, то наверняка выполнение подобных простейших действий не доставляет Вам никакого труда.

Поскольку с кнопкой  всё должно быть понятно (она вызывает окно с текущей инструкцией при нажатии), то стоит сказать несколько слов о кнопке 

Нажатие на эту кнопку вызывает окно «Теоретического материала», причем данное окно автоматически открывает именно ту страницу, информация на которой соответствует материалу заполняемой Вами на текущий момент страницы.

**Работа с файлами и таблицами**

Основной работой пользователя будет являться заполнение таблиц данных. В программе существует два пути решения этой задачи. Во-первых, это заполнение данных вручную, а во-вторых, это считывание данных из файла.

Первый способ осуществляется посредством введения данных о количестве предоставляемой пользователем информации (количестве интервалов или количестве элементов), для чего необходимо заполнить соответствующее поле . Затем необходимо нажать на кнопку . Таким образом, Вы получаете таблицу для ввода данных.

Для правильного (безошибочного) расчета в таблице необходимо заполнить две из трех строк: Xi и Mi, или Xi и Wi.

Альтернативным способом является заполнение данных из файла. Для этого необходимо нажать на кнопку . Для этого момента есть небольшое требование:

* Выбираемый файл должен быть текстовым в формате \*.txt;
* Данные в файле должны быть заполнены через один пробел, и представлять собой только множество действительных чисел (Обязательно!).

**2. Первичная обработка статистических данных**

Для первичной обработки данных пользователю необходимо выполнить следующую последовательность действий:

**1. Для данных считываемых из файла**

1.1 Загрузить статистические данные из файла. Кнопка 

1.2 Выбрать количество интервалов. Значение в .

1.3 Выбрать необходимый уровень значимости.

1.4 Нажать кнопку .

**2. Для данных набираемых вручную** Xi и Mi.

2.1 Выбрать количество интервалов. Значение в .

2.2 Нажать кнопку .

2.3 Заполнить таблицу. Важно, чтобы значения интервалов были помещены в скобки.

2.4 Выбрать необходимый уровень значимости.

2.5 Нажать кнопку .

**3. Для данных набираемых вручную** Xi и Wi.

3.1 Выбрать количество интервалов. Значение в .

3.2 Нажать кнопку .

3.3 Заполнить таблицу. Важно, чтобы значения интервалов были помещены в скобки.

3.4 Заполнить поле суммы частот. Значение в .

3.5 Выбрать необходимый уровень значимости.

3.6 Нажать кнопку .

**3. Проверка распределения по нормальному закону**

Если действия на предыдущей странице были выполнены правильно, то программа автоматически перенесет часть необходимых ей данных для расчета (Вы сможете увидеть заполненную таблицу и посчитанные величины в полях  и ). Все, что требуется от пользователя в этом случае, это нажать на кнопку .

**4. Проверка распределения по биномиальному закону**

Для проверки гипотезы о распределении по биномиальному закону пользователю необходимо выполнить следующую последовательность действий:

**1. Для данных считываемых из файла**

1.1 Загрузить статистические данные из файла. Кнопка .

1.2 Выбрать количество повторений. Значение в .

1.3 При необходимости (в зависимости от условия решаемой задачи), ввести значение вероятности появления или не появления событий, можно и того и другого (сумма p и q ≤ 1). Значения в полях  и/или .

1.4 Выбрать необходимый уровень значимости.

1.5 Нажать кнопку .

**2. Для данных набираемых вручную** Xi и Mi.

2.1 Выбрать количество элементов. Значение в .

2.2 Нажать кнопку .

2.3 Заполнить таблицу.

2.4 Выбрать количество повторений. Значение в .

2.5 При необходимости (в зависимости от условия решаемой задачи), ввести значение вероятности появления или не появления событий, можно и того и другого (сумма p и q ≤ 1). Значения в полях  и/или .

2.6 Выбрать необходимый уровень значимости.

2.7 Нажать кнопку .

**3. Для данных набираемых вручную** Xi и Wi.

3.1 Выбрать количество элементов. Значение в .

3.2 Нажать кнопку .

3.3 Заполнить таблицу.

3.4 Заполнить поле суммы частот. Значение в .

3.5 Выбрать количество повторений. Значение в .

3.6 При необходимости (в зависимости от условия решаемой задачи), ввести значение вероятности появления или не появления событий, можно и того и другого (сумма p и q ≤ 1). Значения в полях  и/или .

3.7 Выбрать необходимый уровень значимости.

3.8 Нажать кнопку .

**5. Сравнение дисперсий двух нормальных распределений**

Для сравнения дисперсий двух нормальных распределений необходимо выполнить следующую последовательность действий:

**1. Для данных считываемых из файла**

1.1 Загрузить статистические данные из файлов. Кнопка .

1.2 Выбрать необходимый уровень значимости. Один из двух: 

1.3 Указать конкурирующую гипотезу. Одну из двух: .

1.4 Нажать кнопку .

**2. Для данных набираемых вручную** Xi и Mi.

2.1 Выбрать количество элементов. Значение в .

2.2 Нажать кнопку .

2.3 Заполнить таблицу.

2.4 Выбрать необходимый уровень значимости. Один из двух: 

2.5 Указать конкурирующую гипотезу. Одну из двух: .

2.6 Нажать кнопку .

**3. Для данных набираемых вручную** Xi и Wi.

3.1 Выбрать количество элементов. Значение в .

3.2 Нажать кнопку .

3.3 Заполнить таблицу.

3.4 Заполнить поле суммы частот. Значение в .

3.5 Выбрать необходимый уровень значимости. Один из двух: 

3.6 Указать конкурирующую гипотезу. Одну из двух: .

3.7 Нажать кнопку .

**6. Подведение итогов**

Надеемся, что в ходе ознакомления с инструкцией у Вас не возникло каких-либо трудностей и всё прочитанное оказалось простым и понятным. Мы же в свою очередь желаем Вам приятной и легкой работы в плане применения нашей программы на практики, а также успешных и правильных расчетов!