МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение высшего образования «Самарский национальный исследовательский университет   
имени академика С.П. Королева»

(Самарский университет)

Институт информатики, математики и электроники

Факультет информатики

Кафедра программных систем

Дисциплина методы планирования эксперимента и статистической обработки информации

**ОТЧЁТ**

по лабораторной работе №1 по дисциплине «Методы планирования эксперимента и статистической обработки информации»

по теме «Дисперсионный анализ данных»

Студент группы № 6231-020402D

Проверил В. В. Любимов

Дата сдачи:

Оценка:

Самара  
2022

Постановка задачи

Задание к лабораторной работе:

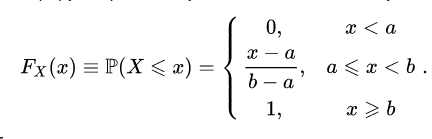
Провести дисперсионный анализ опытных данных с помощью программы MicrosoftOfficeExcel.

Теоретическая информация:

* 1. Равномерное распределение

Равномерное распределение в теории вероятностей – распределение случайной вещественной величины, принимающей значения, принадлежащие некоторому промежутку конечной длины, характеризующееся тем, что плотность вероятности внутри этого промежутка постоянна.

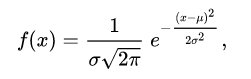
Функция распределения



* 1. Нормальное распределение

Нормальное распределение, также называемое распределением [Гаусса](https://ru.wikipedia.org/wiki/%D0%93%D0%B0%D1%83%D1%81%D1%81,_%D0%9A%D0%B0%D1%80%D0%BB_%D0%A4%D1%80%D0%B8%D0%B4%D1%80%D0%B8%D1%85) или Гаусса-[Лапласа](https://ru.wikipedia.org/wiki/%D0%9B%D0%B0%D0%BF%D0%BB%D0%B0%D1%81,_%D0%9F%D1%8C%D0%B5%D1%80-%D0%A1%D0%B8%D0%BC%D0%BE%D0%BD) – [распределение вероятностей](https://ru.wikipedia.org/wiki/%D0%A0%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5_%D0%B2%D0%B5%D1%80%D0%BE%D1%8F%D1%82%D0%BD%D0%BE%D1%81%D1%82%D0%B5%D0%B9), которое в одномерном случае задаётся функцией [плотности вероятности](https://ru.wikipedia.org/wiki/%D0%9F%D0%BB%D0%BE%D1%82%D0%BD%D0%BE%D1%81%D1%82%D1%8C_%D0%B2%D0%B5%D1%80%D0%BE%D1%8F%D1%82%D0%BD%D0%BE%D1%81%D1%82%D0%B8), совпадающей с [функцией Гаусса](https://ru.wikipedia.org/wiki/%D0%93%D0%B0%D1%83%D1%81%D1%81%D0%BE%D0%B2%D0%B0_%D1%84%D1%83%D0%BD%D0%BA%D1%86%D0%B8%D1%8F):

где параметр *μ* – [математическое ожидание](https://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%BE%D0%B5_%D0%BE%D0%B6%D0%B8%D0%B4%D0%B0%D0%BD%D0%B8%D0%B5) (среднее значение), [медиана](https://ru.wikipedia.org/wiki/%D0%9C%D0%B5%D0%B4%D0%B8%D0%B0%D0%BD%D0%B0_(%D1%81%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D0%BA%D0%B0)) и мода распределения, а параметр *σ* – [среднеквадратическое отклонение](https://ru.wikipedia.org/wiki/%D0%A1%D1%80%D0%B5%D0%B4%D0%BD%D0%B5%D0%BA%D0%B2%D0%B0%D0%B4%D1%80%D0%B0%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%BE%D0%B5_%D0%BE%D1%82%D0%BA%D0%BB%D0%BE%D0%BD%D0%B5%D0%BD%D0%B8%D0%B5) (*σ* ² – [дисперсия](https://ru.wikipedia.org/wiki/%D0%94%D0%B8%D1%81%D0%BF%D0%B5%D1%80%D1%81%D0%B8%D1%8F_%D1%81%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D0%BE%D0%B9_%D0%B2%D0%B5%D0%BB%D0%B8%D1%87%D0%B8%D0%BD%D1%8B)) распределения.

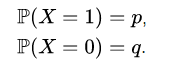


Стандартным нормальным распределением называется нормальное распределение с математическим ожиданием *μ*= 0 и стандартным отклонением *σ*= 1.

* 1. Распределение Бернулли

Распределение Бернулли в теории вероятностей и математической статистике – дискретное распределение вероятностей, моделирующее случайный эксперимент произвольной природы, при заранее известной вероятности успеха или неудачи.

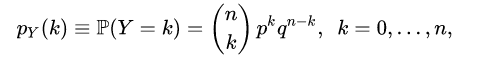
[Случайная величина](https://ru.wikipedia.org/wiki/%D0%A1%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D0%B0%D1%8F_%D0%B2%D0%B5%D0%BB%D0%B8%D1%87%D0%B8%D0%BD%D0%B0) X имеет распределение Бернулли, если она принимает всего два значения: 1 и 0 с вероятностями p и q = 1-p соответственно. Таким образом:



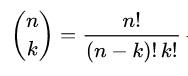
* 1. Биномиальное распределение

Биномиальное распределение в [теориивероятностей](https://ru.wikipedia.org/wiki/%D0%A2%D0%B5%D0%BE%D1%80%D0%B8%D1%8F_%D0%B2%D0%B5%D1%80%D0%BE%D1%8F%D1%82%D0%BD%D0%BE%D1%81%D1%82%D0%B5%D0%B9) – [распределение](https://ru.wikipedia.org/wiki/%D0%A0%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5_%D0%B2%D0%B5%D1%80%D0%BE%D1%8F%D1%82%D0%BD%D0%BE%D1%81%D1%82%D0%B5%D0%B9) количества «успехов» в последовательности из n [независимых](https://ru.wikipedia.org/wiki/%D0%9D%D0%B5%D0%B7%D0%B0%D0%B2%D0%B8%D1%81%D0%B8%D0%BC%D0%BE%D1%81%D1%82%D1%8C_(%D1%82%D0%B5%D0%BE%D1%80%D0%B8%D1%8F_%D0%B2%D0%B5%D1%80%D0%BE%D1%8F%D1%82%D0%BD%D0%BE%D1%81%D1%82%D0%B5%D0%B9)) [случайных экспериментов](https://ru.wikipedia.org/wiki/%D0%A1%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D1%8B%D0%B9_%D1%8D%D0%BA%D1%81%D0%BF%D0%B5%D1%80%D0%B8%D0%BC%D0%B5%D0%BD%D1%82), таких, что [вероятность](https://ru.wikipedia.org/wiki/%D0%92%D0%B5%D1%80%D0%BE%D1%8F%D1%82%D0%BD%D0%BE%D1%81%D1%82%D1%8C) «успеха» в каждом из них постоянна и равна p.

Функция распределения задается формулой:



где - биномиальный коэффициент



* 1. Распределение Пуассона

Распределение Пуассона – вероятностное распределение дискретного типа, моделирует случайную величину, представляющую собой число событий, произошедших за фиксированное время, при условии, что данные события происходят с некоторой фиксированной средней интенсивностью и независимо друг от друга.

Выберем фиксированное число L > 0 и определим [дискретное распределение](https://ru.wikipedia.org/wiki/%D0%A0%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5_%D0%B2%D0%B5%D1%80%D0%BE%D1%8F%D1%82%D0%BD%D0%BE%D1%81%D1%82%D0%B5%D0%B9), задаваемое следующей [функцией вероятности](https://ru.wikipedia.org/wiki/%D0%A4%D1%83%D0%BD%D0%BA%D1%86%D0%B8%D1%8F_%D0%B2%D0%B5%D1%80%D0%BE%D1%8F%D1%82%D0%BD%D0%BE%D1%81%D1%82%D0%B8):



где

* λ –математическое ожидание случайной величины (среднее количество событий за фиксированный промежуток времени),
* *e* –основание натурального логарифма.
  1. Однофакторный дисперсионный анализ

Однофакторный дисперсионный анализ – метод в математической статистике, направленный на поиск зависимостей в экспериментальных данных путём исследования значимости различий в средних значениях по единственной факторной (независимой) переменной. Дисперсионный анализ используется для проверки гипотезы о равенстве нескольких средних значений, соответствующих различным группам или уровням факторной переменной.

Этапы выполнения лабораторной работы:

1. Сгенерировать таблицу случайных чисел размерности 1000 разными законами распределения случайных величин. Были выбраны 5 законов распределения случайных величин:

* Равномерное распределение между 0 и 1;
* Нормальное распределение между 0 и 1;
* Бернулли с параметром p = 0,03;
* Пуассона с параметром лямбда = 4;
* Биноминальное с параметрами p = 0,05 и числом испытаний равным 50.

На рисунке 1 показана часть данных выборок, которые использовались для выполнения лабораторной работы.

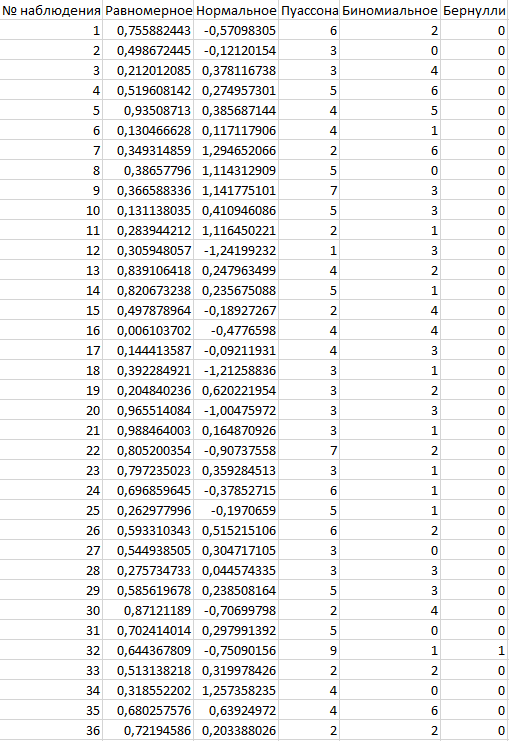


Рисунок 1 – Исходные данные лабораторной работы

1. Произвести однофакторный дисперсионный анализ сгенерированных законов распределения.

Ниже на рисунке 2 отображена статистика по дисперсионному анализу данных, сгенерированных по выбранным законам распределения.

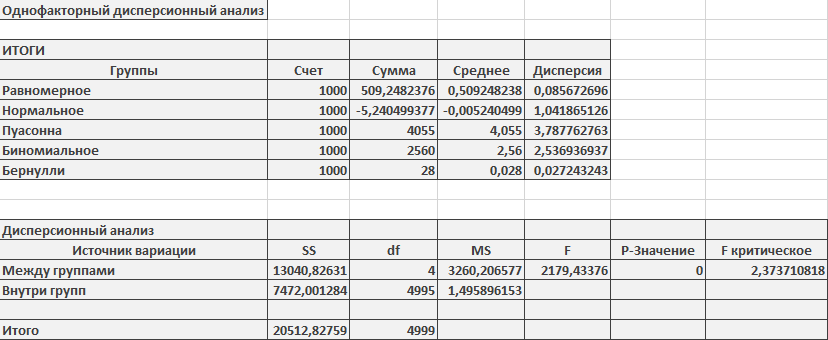


Рисунок 2 - Однофакторный дисперсионный анализ сгенерированных законов распределения

1. Сделать выводы и написать отчёт.

В диапазоне ячеек А13:G17 отображается информация, касающаяся существенности расхождений между группами данных. В строке 13 находятся названия параметров дисперсионного анализа. В строке 14 – результаты межгрупповой обработки, в строке 15 – результаты внутригрупповой обработки, а в строке 17 – суммы значений, выше указанных двух строк. В столбце SS расположены величины варьирования, т.е. суммы квадратов по всем отклонениям. Варьирование, как и дисперсия, характеризует разброс данных. По таблице можно заметить, что межгрупповой разброс значительно выше величины внутригруппового варьирования.

В столбце df находятся значения чисел степеней свободы. Данные числа указывают на количество независимых отклонений, по которым будет вычисляется дисперсия. Например, межгрупповое число степеней свободы равняется разности количеству групп данных и единицы. Чем больше число степеней свободы, тем выше надежность дисперсионных параметров. Данные степеней свобод в таблице показывают, что для внутригрупповых результатов надежность выше, чем для межгрупповых параметров.

В столбце MS расположены величины дисперсий воспроизводимости, которые определяются отношением варьирования и числа степеней свобод. Дисперсия характеризует степень разброса данных, но в отличие от величины варьирования, не имеет прямой тенденции увеличиваться с ростом числа степеней свобод. В ячейке D15 указана величина суммарной дисперсии воспроизводимости данных задачи 1, равной 1,4958. Таким образом, полученные с помощью программы MicrosoftExcel значения дисперсий соответствует ранее рассчитанным величинам. Из таблицы видно, что межгрупповая дисперсия значительно больше внутригрупповой дисперсии.

В столбце F находится значение F-статистики (F-критерий), вычисляемое отношением межгрупповой и внутригрупповой дисперсий. В столбце F критическое расположено F-критическое (табличное) значение, рассчитываемое по числу степеней свободы и величине Альфа. F-статистики и F-критическое значение используются в критерии Фишера. Если F-критерий больше значения F-критическое, то различие между группами данных носят неслучайный характер. В таблице результатов F-критерий превосходит величину F-табличное, что указывает на закономерность различий между группами.

В столбце Р-Значение находится значение вероятности того, что расхождение между группами случайно. Так как в таблице данная вероятность очень мала (равно нулю), то отклонение между группами, скорее всего, носит неслучайный характер.