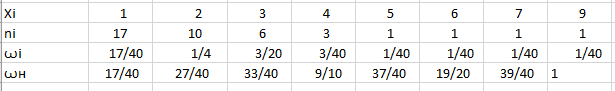
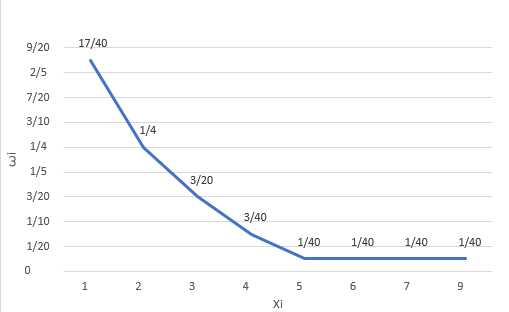
**Дискретные данные**

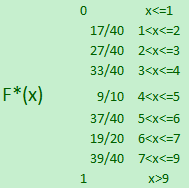
1. За набор данных выступает статистика моей активности на github. Выборка была разбита на дни. Из выборки исключены дни без активности. Статистика собрана самостоятельно. Все вычисления проводились в Excel, в том числе и для непрерывных данных.
2. Вариационный ряд, ряд распределения частот и ряд распределения частостей представлены ниже(а также в Excel файле):

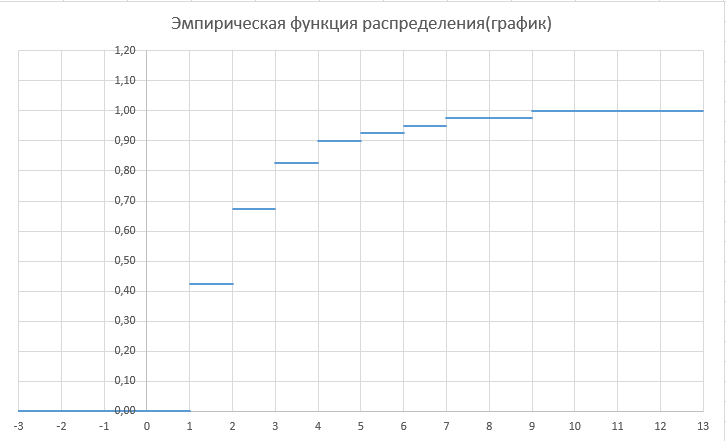


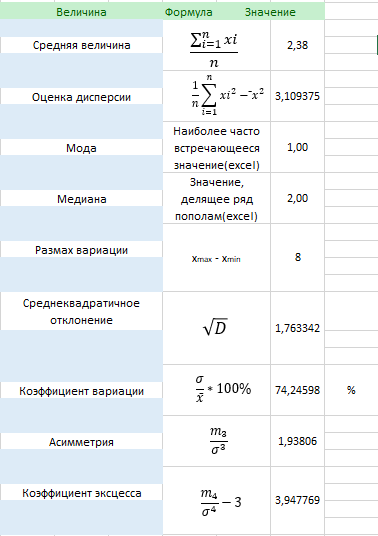
1. Полигон относительных частот:



1. Эмпирическая функция распределения:





1. Характеристики вариационного ряда: 
2. Объем выборки, xmin и xmax:





1. Вывод: одной из оценок математического ожидания такого ряда является генеральная средняя. Тут она выступает как средняя арифметическая (взвешенная).

1. Сделать вывод о данной совокупности легче всего по показателю вариации. Как следует из вычислений, показатель соответствует 74,246%. Существует следующий эмпирический ориентир:

- Если показатель вариации составляет примерно 30% и меньше, то статистическая совокупность может считаться однородной, что большинство вариант находится недалеко от средней, и найденное значение средней (наша оценка математического ожидания) хорошо характеризует центральную тенденцию совокупности.

- Если показатель вариации составляет существенно больше 30%, что является нашим случаем, то совокупность неоднородна, то есть, значительное количество вариант находится далеко от средней величины, и генеральная(выборочная) средняя плохо характеризует типичную варианту.

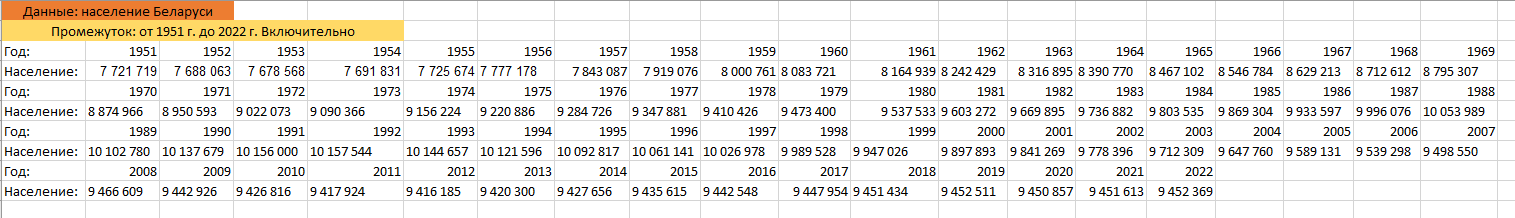
2. Рассмотрим коэффициент асимметрии. Он составляет примерно 1,938. Смысл таков, если коэффициент больше 0, то имеет место правосторонняя асимметрия (наш случай). В противовес распределение будет иметь левостороннюю асимметрию. Между тем, коэффициент оценить можно иначе. Для этого требуется посмотреть на среднюю величину и моду. Если средняя больше, чем мода, то значит коэффициент асимметрии положителен. Для оценки коэффициента принято следующее: если , то асимметрия незначительна. Если – умеренная, если – значительная. В данном случае, асимметрия значительна. Это и можно наблюдать на построенном графике.

3. Показатель дисперсии и стандартного отклонения в данной выборке не слишком велики, что указывает на то, что точки данных могут недалеко распространятся от среднего значения. Это также доказывает и построенный график частот.

4. Заключительным этапом будет оценка эксцесса. Если , то эмпирическое распределение является более высоким (островершинным) – относительно «эталонного» нормального распределения с параметрами и . В противном случае распределение более низкое и пологое. , а это говорит именно о первом варианте.

**Непрерывные данные**

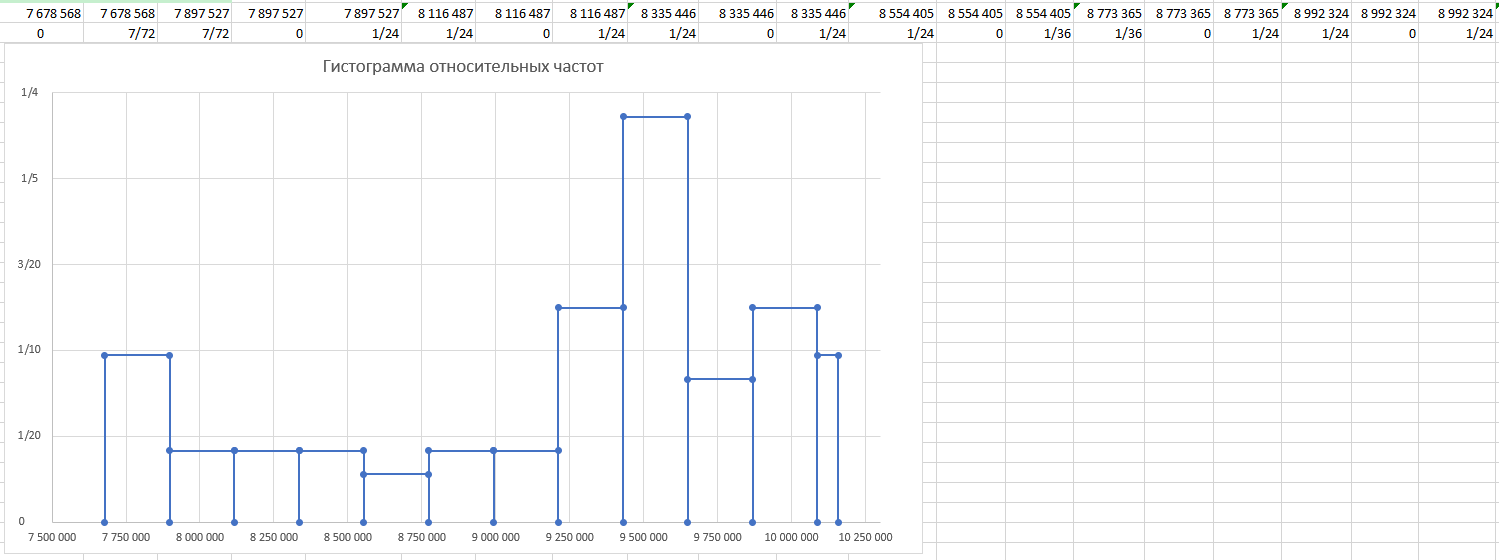
1. За набор данных выступает статистика населения Беларуси в период с 1951 по 2022 год. Используется равноинтервальная группировка, производимая на основании вычисленных оптимального количества интервалов (формула Стерджеса) и шага.
2. Данные, входящие в интервальный ряд:



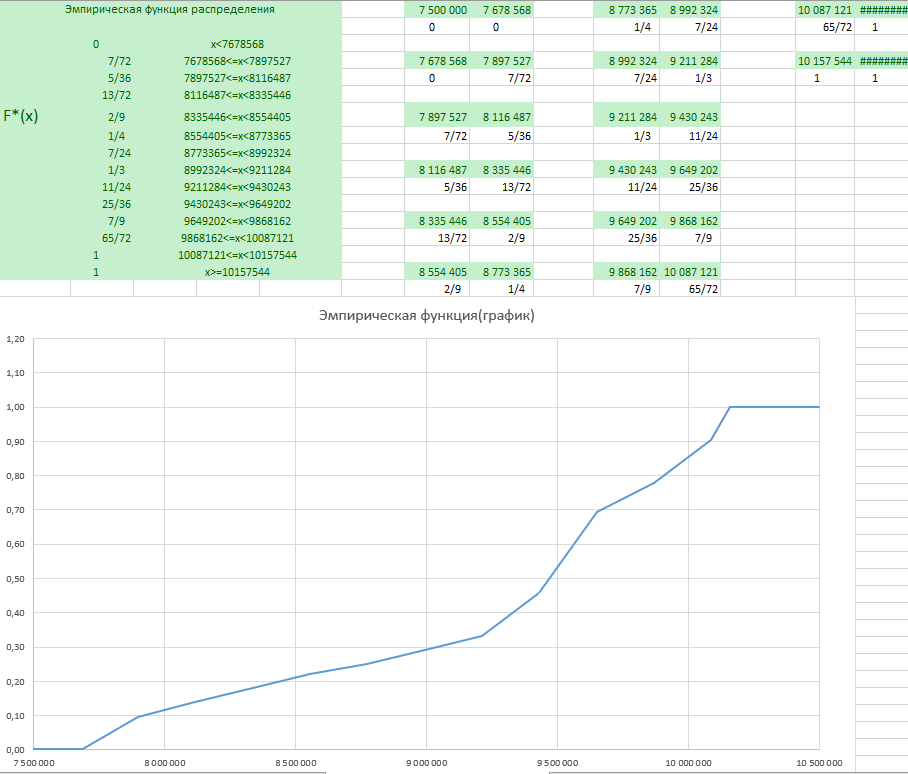
Сразу запишем и сумму значений:



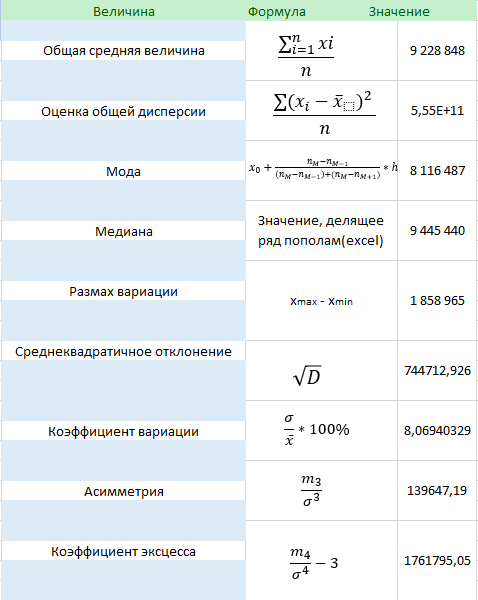
1. Гистограмма относительных частот:

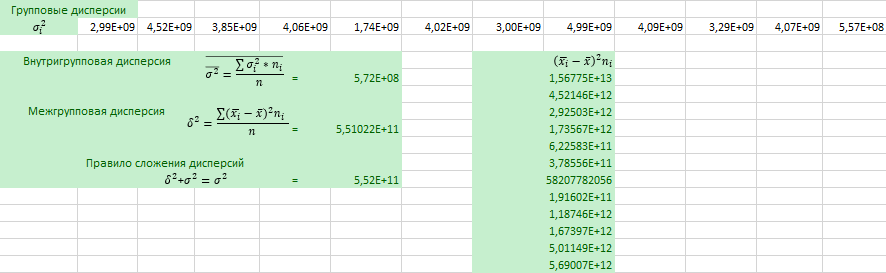
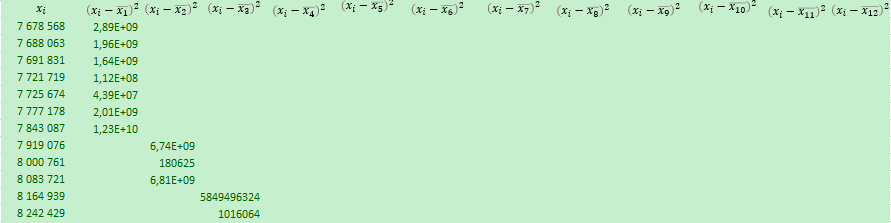


1. Эмпирическая функция распределения:



1. Характеристики:



1. Правило сложения дисперсий и часть использовавшейся для этого таблицы:
2. Объем выборки, xmin и xmax:

1. Вывод:

Судя по полученному коэффициенту вариации, полученная генеральная средняя хорошо характеризует центральную тенденцию совокупности.

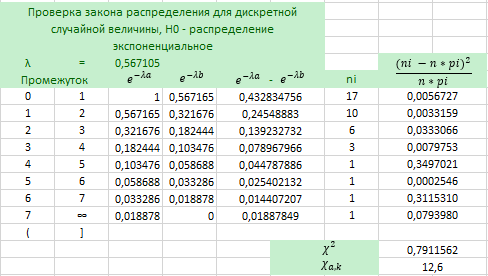
Полученный эксцесс говорит о большей «крутости» распределения по сравнению с нормальным.

Проверка правила сложения дисперсий дает понять, что выбранная группировка, а именно равноинтервальная, не подходит для этой выборки непрерывных данных.

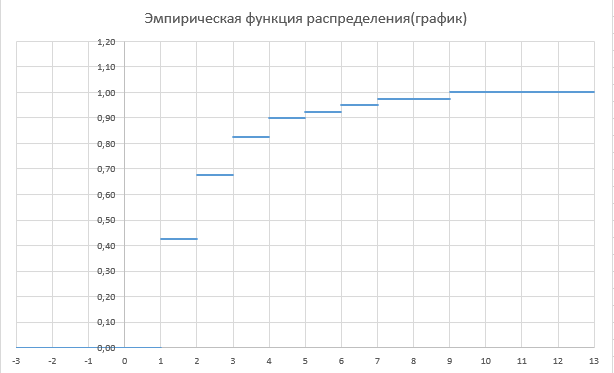
**Оценка принадлежности к распределению**

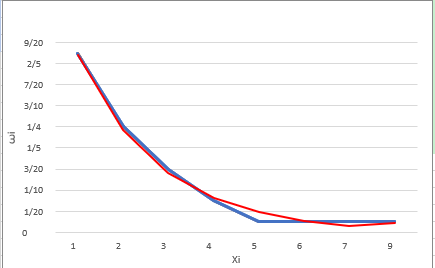
1. Дискретные данные

Для дискретных данных производилась оценка принадлежности к экспоненциальному распределению при помощи критерия согласия Пирсона, результат:

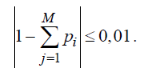


Как видно, полученное значение хи квадрат на порядок меньше критического, в связи с чем можем утверждать, что данные принадлежат к экспоненциальному распределению. Причем значение хи квадрат достаточно близко к нулю, что позволяет утверждать, что распределение практически полностью соответствует экспоненциальному. Результат подтверждается и тем, что вид графика эмпирической функции распределения имеет экспоненциальный вид:





Приближающая кривая(красная) на фоне полигона относительных частот

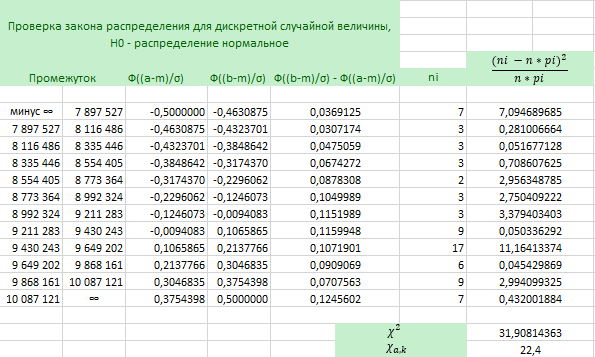
Проверка выполнения  :



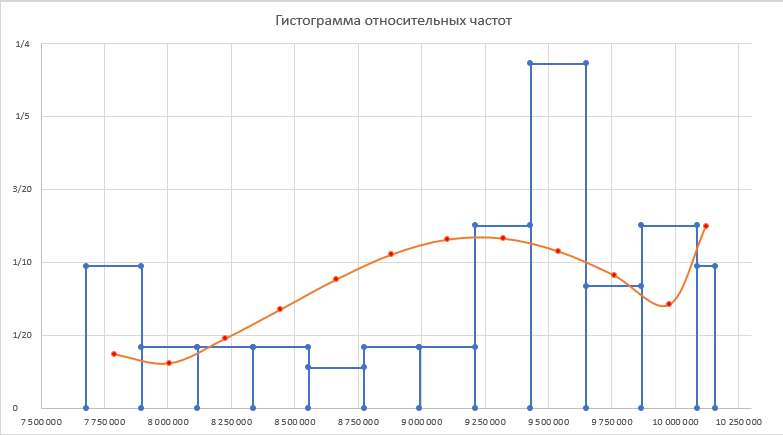
= 0

1. Непрерывные данные

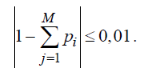
Для непрерывных данных производилась оценка принадлежности к нормальному распределению при помощи критерия согласия Пирсона, результат:



Как видно, полученное значение хи квадрат больше критического, в связи с чем можем утверждать, что данные не принадлежат к нормальному распределению. Однако значение хи квадрат не слишком значительно превышает критическое значение, в связи с чем можно утверждать, что хоть распределение и не является нормальным, оно достаточно близко к нему.



Приближающая кривая(оранжевая) на фоне гистограммы относительных частот

Проверка выполнения  :



= 0