Федеральное государственное бюджетное образовательное учреждение

высшего образования

«Саратовский государственный технический университет имени Гагарина Ю.А.»

Кафедра «Информационные системы и технологии»

# Отчет по лабораторной работе №4

по дисциплине

### «Основы интеллектуального анализа данных»

на тему

«Импорт данных, гистограмма распределения, выборка»

Выполнил: студент группы б1-ИФСТ-21

Яновский Евгений Валерьевич

Преподаватель:

Печенкин Виталий Владимирович

Саратов-2022

**Задание.**

**Первый этап. Формирование набора случайных данных в программе MS Excel.**

Для выполнения задания запустите программу MS Excel, установленную на компьютере.

В процессе работы сохраняйте файлы в папку, которая обсуждалась в разделе «Совет» выше под именем Report\_4\_FIO с соответствующим расширением. При написании фамилии используйте транслитерацию латинскими буквами. Например, если ваша фамилия «Иванов», имя файла будет Report\_4 \_Ivanov.

Как и на прошлых занятиях, мы используем ВАШ ИНДИВИДУАЛЬНЫЙ порядковый номер в журнале, обозначим его ***N***.

Для формирования набора данных, имеющего нормальное распределение, используем функцию MS Excel НОРМОБР(), которая возвращает нормально распределенную переменную для заданной вероятности для определенного **среднего значения** и **стандартного отклонения**. Синтаксис формулы выглядит следующим образом:

=НОРМОБР(вероятность; среднее\_значение; стандартное\_отклонение)

Создадим массив данных (в столбце) со средним значением ***N*** (ваш порядковый номер) и стандартным отклонением ***N/4***.на основе случайных чисел из диапазона [0, 1].

Формула для ячейки A1 будет иметь вид (если Ваш номер есть 22)

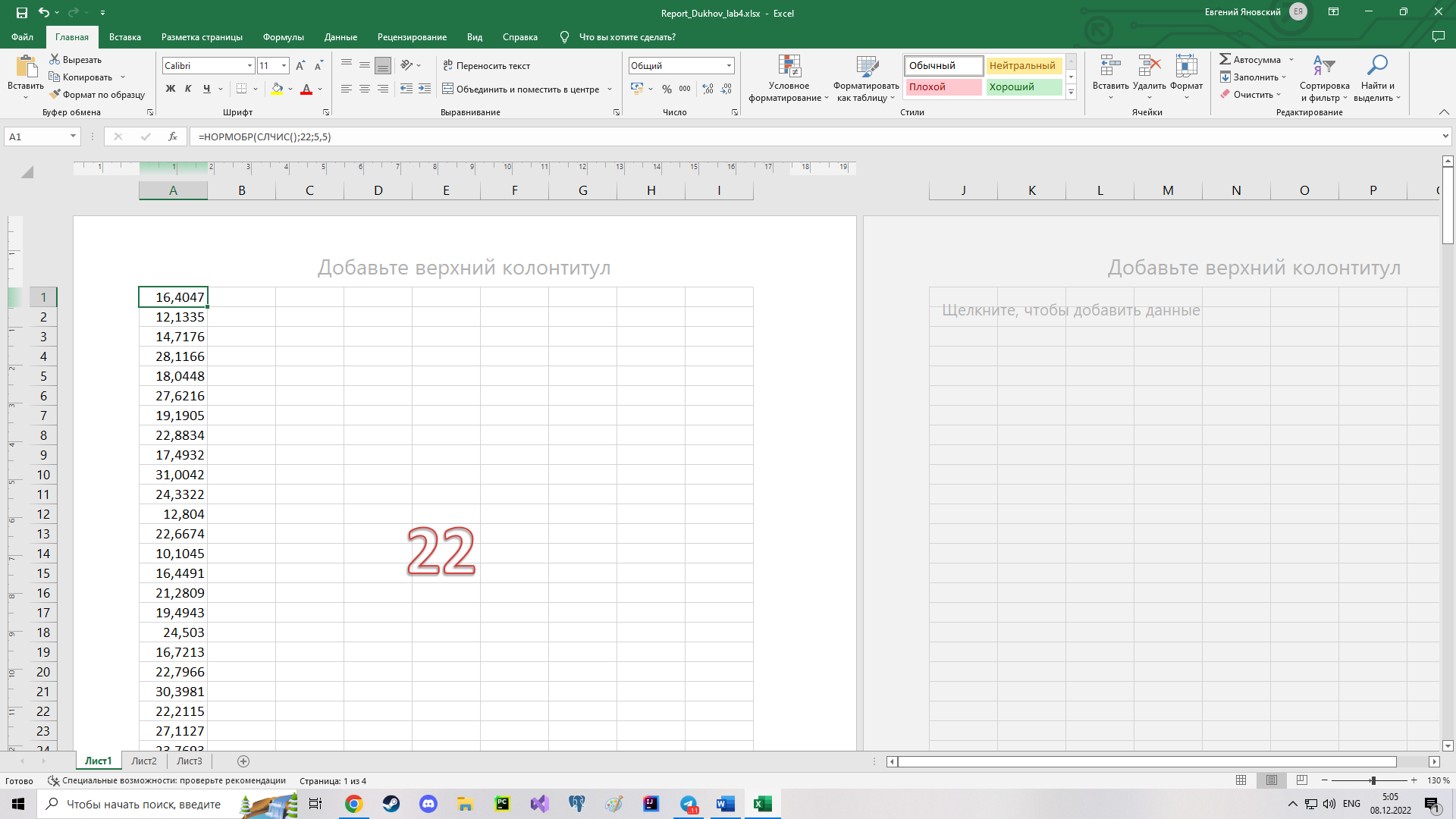
«=НОРМОБР(СЛЧИС(); 22; 4,5)»

Распространим это значение на 200 чисел по столбцу вниз.

Как на прошлых занятиях скопируем диапазон всех полученных значений и вставим обратно как значения с использованием специальной вставки. Это необходимо для того, чтобы предотвратить пересчёт значений при каждой инициализации данных.

Правее от столбца с данными вставьте надпись с вашим индивидуальным номером, так чтобы этот номер был виден на скриншоте экрана в финальном отчёте.

Скриншот экрана должен выглядеть примерно так



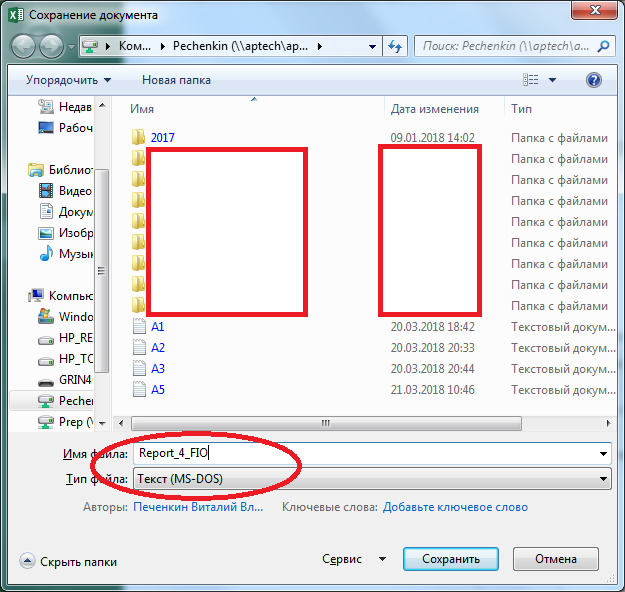
**Сохраните файл MS Excel. Скопируйте скриншот для финального отчёта.**

**Покажите результаты работы преподавателю.** Он отметит факт выполнения первой части задания

**Второй этап. Экспорт данных MS Excel в текстовый файл \*. txt**

Сохраните содержимое электронной таблицы в файл с расширением \*.txt (разделители запятые). Этот файл нам понадобится для импорта данных в программу PSPP.

Выбор имени файла и его типа должны соответствовать следующему диалогу

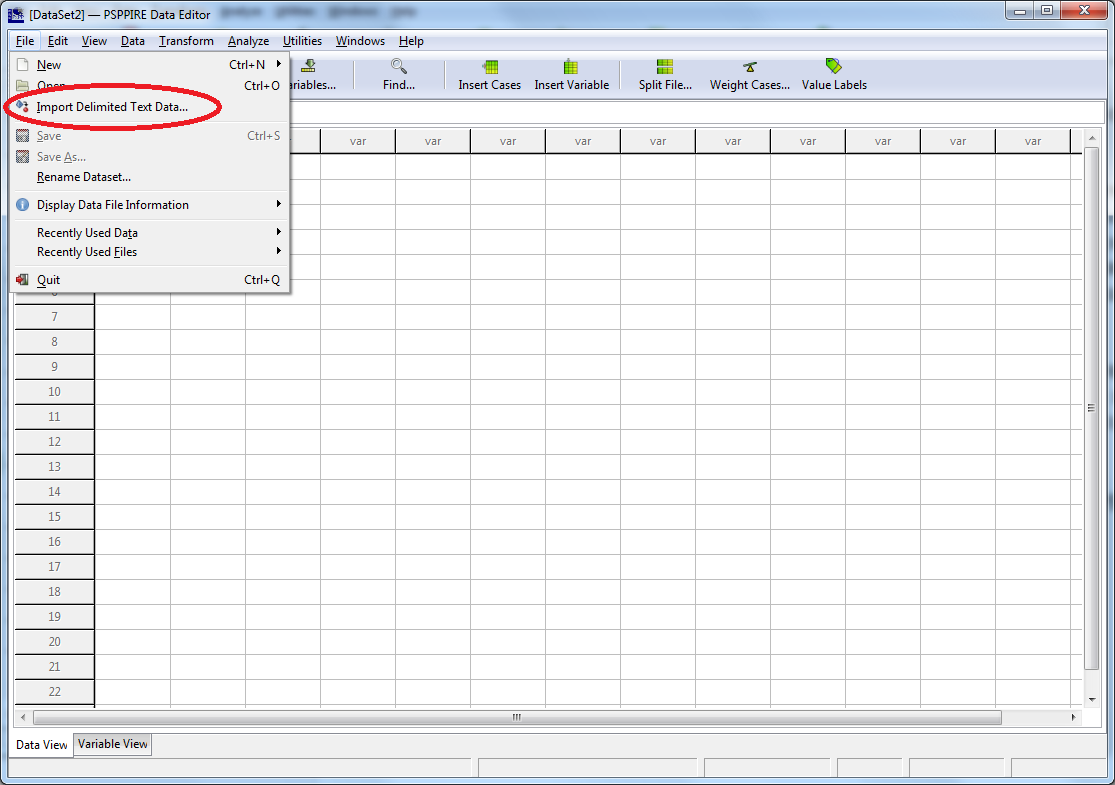


**Третий этап.** **Импорт данных из текстового файла в программу PSPP**

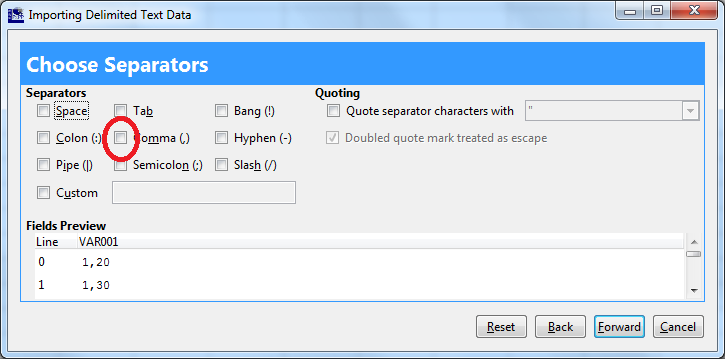
На этом этапе необходимо импортировать данные из текстового файл, полученного на предыдущем этапе в программу PSPP.

Поэкспериментируйте с процедурой импорта данных, чтобы добиться получения созданного в MS Excel массива в качестве данных в программе PSPP. Мы используем формат для импорта, который загружает текстовый файл (Text (\*.txt))

Используйте для импорта данных команду



Для того, чтобы запятая не считалась разделителем, измените значение разделителя в диалоге, просто снимите отметку в соответствующем чекбоксе. Если этого не сделать, значения в строке будут представлены как два различных числа, загружаемые в две переменные.

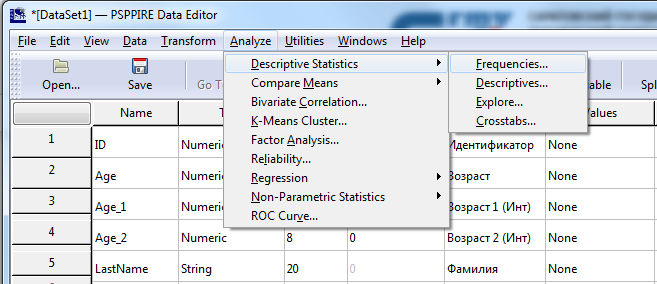


**Покажите результаты работы преподавателю.** Он отметит факт выполнения первой части задания

**Сохраните файл, используя стандартный формат данных программы PSPP.**

**Четвёртый этап. Построение гистограммы распределения исходных данных.**

Используя ресурсы интернет и систему меню программы PSPP постройте гистограмму распределения полученного массива данных. Параметры выводимого графика находятся на форме диалога, вызываемого командой, показанной на следующем рисунке (вызов параметров Chart кнопкой)



**Отобразите нормальную кривую** на гистограмме. Сделайте вывод о наличии нормального распределения для полученных данных.

Сделайте скриншот гистограммы и добавьте его к вашему отчёту.

**Пятый этап. Построение выборочной совокупности из имеющихся данных**

Используя систему подсказки выполните команду, которая отберёт из текущего набора данных 50%. Кейсы, не вошедшие в выборку, нужно удалить из набора (Используйте соответствующую опцию).

Для полученной выборочной совокупности вновь постройте гистограмму и скопируйте её скриншот в файл отчёта. **Скриншоты экрана.**

