**Федеральное государственное образовательное**

**бюджетное учреждение**

**высшего образования**

**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ**

**ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ**

**ФЕДЕРАЦИИ»**

**(Финансовый университет)**

**Факультет**

**информационных технологий и анализа больших данных**

**Кафедра «Бизнес-информатика»**

**Домашнее задание № 7**

«Решение задач экспертных оценок.»

Студенты группы БИ20-8:

Луканина Полина

Аверкин Никита

Филимонова Арина

Совин Владимир

Горшков Георгий

Киселева Евгения

Руководитель:

Аксенов Дмитрий Андреевич

**Москва 2022**

### Оглавление

[Оглавление 2](#_Toc104756542)

[1. Постановка задачи (физическая модель) 4](#_Toc104756543)

[2. Математическая модель 4](#_Toc104756544)

[3.1. Алгоритм в Python 9](#_Toc104756545)

[3.1.1. Описание входных данных. 9](#_Toc104756546)

[3.1.2. Описание алгоритма решения 11](#_Toc104756547)

[3.1.3. Описание выходных данных 12](#_Toc104756548)

[3.2. Алгоритм в Excel 12](#_Toc104756549)

[3.2.1. Описание входных данных. 12](#_Toc104756550)

[3.2.2. Описание алгоритма решения 13](#_Toc104756551)

[3.2.3. Описание выходных данных 18](#_Toc104756552)

[4. Вариант использования 18](#_Toc104756553)

[5. Архитектура решения 20](#_Toc104756554)

[5.1 Функции считывания информации 20](#_Toc104756555)

[5.2 Функции обработки информации 22](#_Toc104756556)

[5.3 Функции вывода информации 24](#_Toc104756557)

[6. Тестирование 24](#_Toc104756558)

[6.1. Тестирование Датасета №1: 24](#_Toc104756559)

[6.1.1 Метод Python: 24](#_Toc104756560)

[6.1.2 Метод Excel: 26](#_Toc104756561)

[6.2. Тестирование Датасета №2: 27](#_Toc104756562)

[6.2.1 Метод Python: 27](#_Toc104756563)

[6.2.2 Метод Excel: 28](#_Toc104756564)

[6.3. Тестирование Датасета №3: 29](#_Toc104756565)

[6.3.1 Метод Python: 29](#_Toc104756566)

[6.3.2 Метод Excel: 30](#_Toc104756567)

[6.4. Тестирование Датасета №4: 31](#_Toc104756568)

[6.4.1 Метод Python: 31](#_Toc104756569)

[6.4.2 Метод Excel: 32](#_Toc104756570)

[6.5. Тестирование Датасета №5: 33](#_Toc104756571)

[6.5.1 Метод Python: 33](#_Toc104756572)

[6.5.2 Метод Excel: 34](#_Toc104756573)

[7. Заключение: 35](#_Toc104756574)

# Постановка задачи (физическая модель)

Дубайская компания Dubai Port World (в дальнейшем DP World) является одним из крупнейших мировых портовых операторов. Компания оперирует 78 морскими терминалами и самым большим индустриальным парком мира Jebel Ali Free Zone.

Цель задания – улучшить клиентский сервис, а именно: удовлетворенность сроками и качеством работы, вежливостью и компетентностью персонала, переработать систему индивидуального подхода к каждому заказчику.

Для достижения цели провести опрос потенциальных и уже существующих клиентов, необходимо вставить в опросник такие критерии, как бережное отношение к грузу, качество обратной связи, доставка груза, срочная доставка, ценовая политика, индивидуальный подход к каждому клиенту, оплата, программа лояльности.

На основе полученных данных провести анализ бизнес-процесса убирая ложные данные из аналитики. Предоставить план действий по увеличению показателей.

# Математическая модель

Существует несколько семейств математических методов обработки экспертных оценок:

1. **Назначение баллов.** Применяется, когда можно сказать, что один вариант лучше другого и оценить, насколько

2. **Ранжирование.** Применяется, когда можно сказать, что один вариант лучше другого, но нельзя оценить насколько.

3.**Попарные сравнения (бинарные отношения).** Применяется, когда невозможно ранжировать все варианты по порядку, но можно попарно сравнить их между собой.

Среднее количество выборки – **математическое ожидание:**

Данный показатель отражает среднее значение результатов и ответов всех экспертов.

Степень рассогласованности мнений и ответов экспертов – **дисперсия по выборке:**

Если дисперсия маленькая – мнения и ответы экспертов согласованы, в противном случае, их мнения на счет оцениваемых предметов сильно отличаются. С помощью этого показателя можно определить, насколько правильно подобраны эксперты и верная ли перед ними поставлена задача.

Если дисперсия показывает, как именно и насколько сильно разбросаны значения, то **среднеквадратическое отклонение** показывает среднюю степень разброса значений:

Интервальная оценка баллов – **доверительный интервал**, например ±2𝜎:

Интервальная оценка отражает, насколько сильно разбросаны мнения экспертов, относительно среднего значения, например. На основе этой информации можно судить, корректно ли в итоге оценены сравниваемые объекты.

**Метод медианных рангов:**

применяется в том случае, если невозможно назначить баллы, но можно расставить по рангу – отранжировать объекты. Подходит для сравнения сложных объектов, для которых можно сказать, что один вариант лучше другого, но нельзя указать количественно на сколько.

**Медиана** – это некоторая отметка, делящая ранжированные данные (отсортированные по возрастанию или убыванию) на две равные части, и она, в отличие от математического ожидания, устойчива к аномальным значениям

**Метод средневзвешенных рангов:**

Средневзвешенные ранги определяются с учетом компетентности всех экспертов, которые рассчитываются по следующему алгоритму:

1. Определяется среднее арифметическое рангов кандидатов (матожидание):

2. По каждому эксперту рассчитывается степень отклонения его ответа от среднего значения М – то есть дисперсия мнений каждого эксперта:

3. Чем больше дисперсия, тем сильнее мнение эксперта отличается от мнения большинства. Следовательно, его компетентность в данном вопросе можно считать ниже, чем у коллег (исключая догму «согласованности»). Компетентность эксперта K будет обратно пропорциональна его дисперсии

4. Выразим значимость мнения каждого эксперта в процентах (нормируем) и приведем его к виду коэффициента компетентности k:

5. Далее найдем средневзвешенное мнение экспертов относительно кандидатов с учетом коэффициента их компетентности (важности их мнения) (таблица 19.14):

**Бинарные отношения**

**Метод попарных сравнений** - один из наиболее сложных методов. Применяется в случаях, когда объекты настолько комплексные и многомерные, что нет возможности отранжировать их по одному признаку, но есть возможность попарно сравнить их между собой. Каждый эксперт должен сравнить каждую пару вариантов друг с другом. Результат записывается в виде матриц бинарных отношений

Принцип применения метода, следующий: например, имеется два объекта с условными названиями а и b и метрика [0, 1]. В случае, если объект, а лучше или равен объекту b, метрика приобретает значение 1, в противном случае – значение 0:

Если сравниваемых объектов несколько 𝑎1, 𝑎2, …, 𝑎𝑛, то результатом их попарных сравнений будет являться матрица бинарных отношений:

‖𝑋𝑖𝑗‖ = {1, 𝑎𝑖 ≥ 𝑎𝑗 0, 𝑎𝑖 <𝑎𝑗

Далее необходимо найти некоторое усредненное значение мнений экспертов. Для того, чтобы сделать это на матрицах введем несколько новых понятий.

**Расстоянием Кемени** между бинарными отношениями X и Y называется число, характеризующее количество несовпадающих элементов матриц X и Y:

**Медианой Кемени** называется такая матрица А, суммарное расстояние Кемени от которой до всех остальных матриц 𝑋1, 𝑋2… 𝑋𝑚 является минимальным:

Таким образом, медиана Кемени А является той самой матрицей, которая находится на «геометрической середине» по отношению ко всем остальным матрицам 𝑋1, 𝑋2, 𝑋3, 𝑋4 и за счет этого является усредненным мнением всех экспертов.

Найти медиану Кемени можно с помощью задачи нелинейного программирования: Исходные данные: матрицы бинарных отношений 𝑋1, 𝑋2… 𝑋𝑚.

Искомая переменная: медиана Кемени А. Целевая функция – минимальное суммарное расстояние до всех матриц:

Ограничения:

Таким образом, в методе бинарных отношений фокус внимания эксперта сужен до решения простой задачи – какой из двух вариантов лучше. Это сделать проще, чем охватить область всех вариантов, еще и назначив каждому из них баллы или ранги. Поэтому метод бинарных отношений дает более точные ответы, когда множество альтернатив велико или используются сложные объекты, но трудозатратен в обработке.

# 3.1. Алгоритм в Python

## 3.1.1. Описание входных данных.

Формат входных данных определяется тем, что программа принимает только CSV файл, который содержит в себе массивы данных. Исходя из необходимости решить определенную задачу, а именно математическую обработку экспертных оценок или математическую обработку ранговых оценок.

**Пример входных данных в CSV файле:**

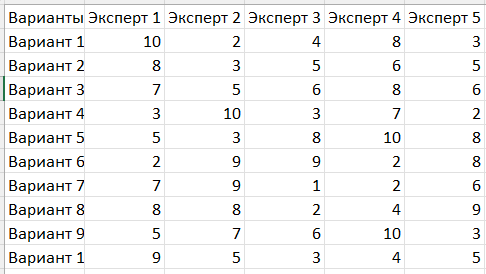


Рисунок / Для математической обработки экспертных оценок

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Для математической обработки ранговых оценок



Рисунок . Пример заполнения пути к файлу

**Пример импортированных данных в программу:**



Рисунок . Для математической обработки экспертных оценок

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Для математической обработки ранговых оценок

## 3.1.2. Описание алгоритма решения

После того как данные введены, программе необходимо преобразовать данные для дальнейшего использования.

Шаг 1: рассчитывается математическое ожидание по выборкам, состоящим из экспертных оценок по каждой позиции.

Шаг 2: рассчитывается дисперсия по выборкам, состоящим из экспертных оценок по каждой позиции.

Шаг 3: рассчитывается среднеквадратическое отклонение по выборкам, состоящим из экспертных оценок по каждой позиции.

Шаг 4: рассчитывается доверительный интервал по выборкам, состоящим из экспертных оценок по каждой позиции.

Шаг 5: рассчитывается медиана рангов.

Шаг 6: рассчитывается массив коэффициентов компетентности экспертов.

Шаг 7: определение средневзвешенных рангов с учетом коэффициентов компетентности.

## 3.1.3. Описание выходных данных

Исходя из того, какую задачу вы выберете решить, программа выдаст в конце либо математическую обработку экспертных оценок, либо математическую обработку ранговых оценок.

**Пример выходных данных:**

**Изображение выглядит как стол

Автоматически созданное описание**

Рисунок . Для математической обработки экспертных оценок

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Для математической обработки ранговых оценок

## 3.2. Алгоритм в Excel

### 3.2.1. Описание входных данных.

Формат входных данных представляет собой матрицы с экспертными оценками.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Пример входных данных

### 3.2.2. Описание алгоритма решения

После того как данные введены, программе необходимо преобразовать данные для дальнейшего использования.

Шаг 1: составляем матрицы бинарных отношений по всем экспертам, посредством выбора между двумя объектами выбора.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Матрицы бинарных отношений

Шаг 2: введение в матрицу медианы Кемени.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Медиана Кемени

Шаг 3: расчёт расстояния между матрицей медианы Кемени и остальными матрицами бинарных отношений по всем экспертам. При помощи функции СУММКВРАЗН, которая возвращает сумму квадратов разностей соответствующих значений в двух массивах. Первый массив данных относится к матрице бинарных отношений с экспертными оценками, а второй массив данных – это матрица медианы Камени.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Матрицы

Шаг 4: расчет суммы всех расстояний путем сложения.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Сумма расстояний

Шаг 5: заполнение нижней половины матрицы.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Матрица

Шаг 6: заполнение верхней половины матрицы с помощью поиска решения.

Изображение выглядит как текст

Автоматически созданное описание

Рисунок . Поиск решения

В окошко «оптимизировать целевую функцию» указывается ячейка со

значением целевой функции. И выбираем оптимизировать функцию до

минимума.

Затем заполняем ограничения, где указываем, что каждая строка верхней части матрицы должны быть бинарными.

Шаг 7: вычисление суммы значений по каждой столбцу матрицы медианы Камени.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Вычисление значений

### 3.2.3. Описание выходных данных

Выходными данными будет являться сумма значений по каждому столбцу, которая представляет собой место в рейтинге.



Рисунок . Рейтинг

# 4. Вариант использования

Существует единственный вариант использования кода – это загрузка файла. Данный вариант использования включает в себя ввод данных с помощью файла csv. Для того, чтобы ввести путь к файлу необходимо запустить программу и выбрать какую задачу будете решать (1, 2 или 3).

****

Рисунок . Выбор задачи

После этого появляется окно, в котором вводим путь к csv файлу. Например:



Рисунок 18. Выбор местонахождения csv файла

При выборе 1 и 2 типа задач после выбора csv. Файла программа сразу выдает решение.

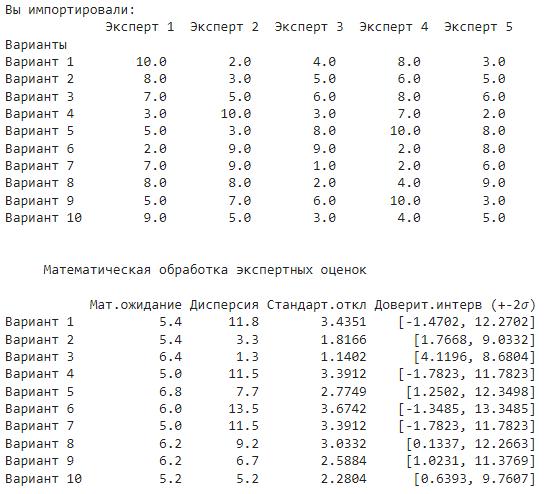


Рисунок . Решение

При выборе 3 типа задачи необходимо ввести количество экспертов и после этого для каждого подгрузить соответствующие файлы.

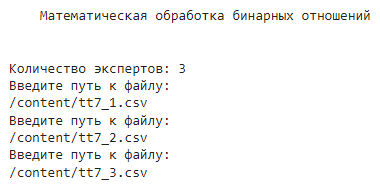


Рисунок . 3 тип задачи

# 5. Архитектура решения

Для решения задачи использовались методы (функции), которые можно разделить на 3 принципиальных кода.

## 5.1 Функции считывания информации

**Функция import\_csv:**

**Входные данные:**

* Нет входных данных.

**Выходные данные:**

* A – массив данных в формате dataframe.

**Переменные, затрагиваемые в ходе работы:**

* csv – матрица исходных данных.
* A – массив данных в формате dataframe.
* x – переменная, затрагиваемая в ходе работы.
* name\_columns – список значений столбца.
* name\_index – список значений индекса.

**Изображение выглядит как текст

Автоматически созданное описание**

Рисунок . Фрагмент кода

**Функция input\_way:**

**Входные данные:**

* Нет входных данных.

**Выходные данные:**

* Нет выходных данных.

**Переменные, затрагиваемые в ходе работы:**

* way - номер, решаемой задачи.

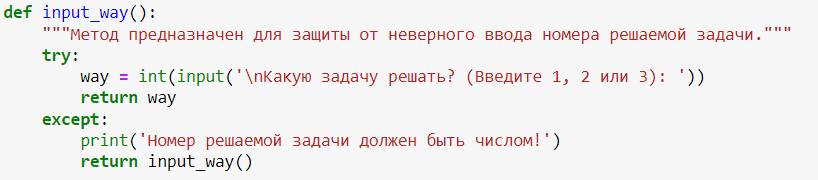


Рисунок . Фрагмент кода

## 5.2 Функции обработки информации

После того, как вы введете все необходимые данные, программа их получит и начнет первичную обработку.

Поскольку ввод данных осуществляется с помощью файла csv, то для начала файл необходимо открыть и прочитать программе, а затем разделить данные и также занести их в пустой словарь. Более того, программа образует список с названиями колонок массива данных, а также список с названиями строк. Также предусмотрена защита от неверного ввода номера решаемой задачи.

Функция expert\_opinions предназначена для математической обработки экспертных оценок. Внутри функции рассчитывается математическое ожидание, дисперсия, среднеквадратическое отклонение и доверительный интервал по выборкам, состоящих из экспертных оценок по каждой позиции.

Функция rank\_estimates предназначена для математической обработки ранговых оценок. Внутри функции рассчитывается медиана рангов, массив коэффициентов компетентности экспертов и определяется средневзвешенные ранги с учетом коэффициентов компетентности.

**Функция expert\_opinions:**

**Входные данные:**

* data – данные из файла csv.
* name\_index – индекс данных.

**Выходные данные:**

* data\_expert\_opinions – матрица значений математической обработки.

**Переменные, затрагиваемые в ходе работы:**

* mean – значения математического ожидания.
* dispersion – значения дисперсии.
* standard\_deviation – значения среднеквадратического отклонения.
* confidence\_interval – значения доверительного интервала.
* data – данные из файла csv.
* name\_index – индекс данных.
* data\_expert\_opinions – матрица значений математической обработки.

**Функция rank\_estimates:**

**Входные данные:**

* data – данные из файла csv.
* name\_index – индекс данных.
* name\_columns – названия столбцов.

**Выходные данные:**

* Нет выходных данных.

**Переменные, затрагиваемые в ходе работы:**

* data – данные из файла csv.
* name\_index – индекс данных.
* mean – значения математического ожидания.
* median – значения медианы рангов.
* dispersion – значения дисперсии.
* competence – список компетенций экспертов.
* competence\_str – список компетенций экспертов в процентах типа str
* competence\_df – таблица DataFrame с компетенциями.
* rank\_with\_competence – значения средневзвешенных рангов с учётом коэффициентов компетентности.

## 5.3 Функции вывода информации

Метод вывода информации

Что делает: осуществляет вывод необходимой информации

В данном методе осуществляется непосредственно вызов функций с помощью метода print ().

**Изображение выглядит как текст

Автоматически созданное описание**

Рисунок . Часть кода, отвечающая за вывод данных

# 6. Тестирование

Проведём тестирование нашей программы и сравним полученные показатели,

чтобы сделать вывод о предпочтительном варианте использования нашей программы

или Excel под условия заказчика.

## 6.1. Тестирование Датасета №1:

### 6.1.1 Метод Python:

Выбираем какую задачу решить, импортируем датасет и получаем результат:

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Решение 1 задачи

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Решение 2 задачи

### 6.1.2 Метод Excel:

Вводим данные датасета в необходимые поля, выделяем фрагмент данных, используем «Анализ данных» методом описательной статистики с группированием по строкам и получаем результат:

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Тестирование №1 в Excel

## 6.2. Тестирование Датасета №2:

### 6.2.1 Метод Python:

Выбираем какую задачу решить, импортируем датасет и получаем результат: Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Решение 1 задачи

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Решение 2 задачи

### 6.2.2 Метод Excel:

Вводим данные датасета в необходимые поля, выделяем фрагмент данных, используем «Анализ данных» методом описательной статистики с группированием по строкам и получаем результат:

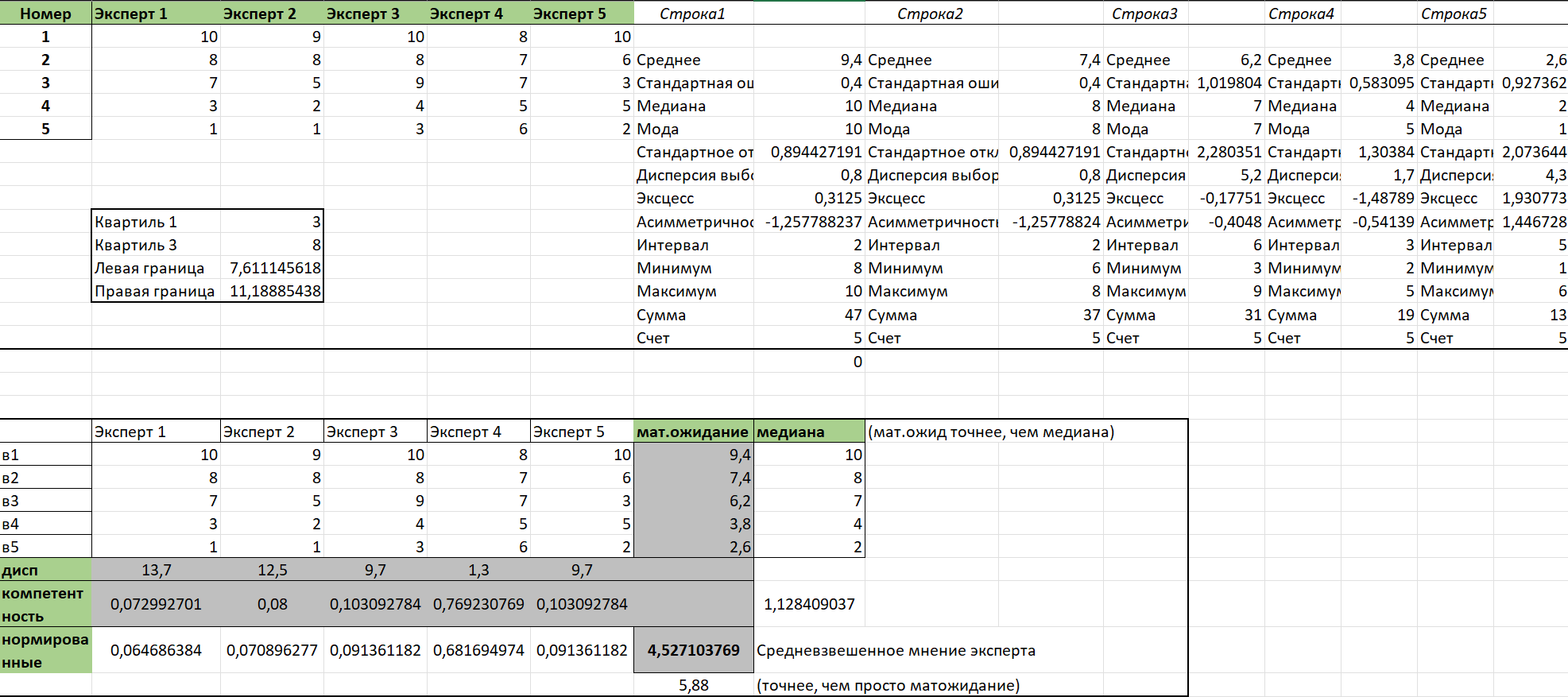


Рисунок . Тестирование №2 в Excel

## 6.3. Тестирование Датасета №3:

### 6.3.1 Метод Python:

Выбираем какую задачу решить, импортируем датасет и получаем результат:

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Решение 1 задачи

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Решение 2 задачи

### 6.3.2 Метод Excel:

Вводим данные датасета в необходимые поля, выделяем фрагмент данных, используем «Анализ данных» методом описательной статистики с группированием по строкам и получаем результат:

Изображение выглядит как текст, внутренний

Автоматически созданное описание

Рисунок . Тестирование №3 в Excel

## 6.4. Тестирование Датасета №4:

### 6.4.1 Метод Python:

Выбираем какую задачу решить, импортируем датасет и получаем результат:

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Решение 1 задачи

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Решение 2 задачи

### 6.4.2 Метод Excel:

Вводим данные датасета в необходимые поля, выделяем фрагмент данных, используем «Анализ данных» методом описательной статистики с группированием по строкам и получаем результат:

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Тестирование №4 в Excel

## 6.5. Тестирование Датасета №5:

### 6.5.1 Метод Python:

Выбираем какую задачу решить, импортируем датасет и получаем результат:

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Решение 1 задачи

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Решение 2 задачи

### 6.5.2 Метод Excel:

Вводим данные датасета в необходимые поля, выделяем фрагмент данных, используем «Анализ данных» методом описательной статистики с группированием по строкам и получаем результат:

Изображение выглядит как стол

Автоматически созданное описание

Рисунок . Тестирование №5 в Excel

# 7. Заключение:

Представленный нами код решает поставленную задачу. На основании тестирования данного алгоритма можно сделать вывод о том, что Python выводит самое оптимальное решение быстро и наглядно. Теперь сравним два алгоритма по критериям: эффективности, скорости использования алгоритма, простоты использования, надёжности в разрезе человеческого фактора и точности предоставляемого решения.

Таблица 1. Сравнение решения в Excel и Pythоn

|  |  |  |
| --- | --- | --- |
| **Критерий** | **Python** | **Excel** |
| Эффективность | Высокая | Высокая |
| Скорость использования алгоритма | Высокая | Низкая |
| Простота использования | Высокая | Средняя |
| Надёжность | Высокая | Средняя |
| Точность | Высокая | Высокая |

Мы считаем, что представленный рукописный код на языке Python лучше, потому что удобнее, быстрее и проще, имеет функцию импорта исходных данных. Улучшением кода может послужить добавление времени выполнения запроса, более детальной выводимой информации, ручного ввода данных и генерации случайных данных.