

СРАВНИТЕЛЬНЫЙ АНАЛИЗ РЕАЛИЗАЦИИ КОРРЕЛЯЦИОННОГО АНАЛИЗА В MICROSOFT EXCEL И LIBREOFFICE CALC

Пепельшев Д.И., Воробьев Д.А.

ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина», Екатеринбург, e-mail: dmitry.pepelyshev@urfu.ru

Важность корреляционного анализа трудно переоценить. Он лежит в основе многих исследовательских работ, бизнес-аналитики и прогнозирования. Понимание взаимосвязей между переменными позволяет делать обоснованные выводы о влиянии одного фактора на другой, прогнозировать будущие значения и принимать эффективные решения. Данная работа посвящена сравнительному анализу реализации корреляционного анализа в двух популярных электронных процессорах: Microsoft Excel и LibreOffice Calc. Цель исследования – выявить сходства и различия в методах и результатах расчета корреляционных коэффициентов в обоих приложениях, сравнить точность встроенных методов, а также оценить удобство использования интерфейсов для выполнения анализа. В ходе исследования рассматривается самый популярный тип корреляционного коэффициента, коэффициент корреляции Пирсона, который может быть рассчитан с помощью встроенных функций в обеих программах. В качестве набора данных использовались исторические данные о погоде в Екатеринбурге за декабрь 2024 г. Результаты помогут пользователям выбрать наиболее подходящий инструмент для своих аналитических задач. В заключение кратко описываются преимущества и недостатки каждой из программ, а также даются рекомендации по выбору оптимальной программы.

Ключевые слова: Microsoft Excel, LibreOffice Calc, корреляционный анализ, анализ, аналитика

COMPARATIVE ANALYSIS OF THE IMPLEMENTATION OF CORRELATION ANALYSIS IN MICROSOFT EXCEL AND LIBREOFFICE CALC

Pepelyshev D.I., Vorobev D.A.

Ural Federal University named after the first President of Russia B.N. Yeltsin, Yekaterinburg, e-mail: dmitry.pepelyshev@urfu.ru

The importance of correlation analysis is difficult to overestimate. It underlies many research works, business analytics and forecasting. Understanding the relationships between variables allows you to make informed conclusions about the influence of one factor on another, predict future values and make effective decisions. This work is devoted to a comparative analysis of the implementation of correlation analysis in two popular electronic processors: Microsoft Excel and LibreOffice Calc. The purpose of the study is to identify similarities and differences in the methods and results of calculating correlation coefficients in both applications, compare the accuracy of the built-in methods, and evaluate the ease of use of interfaces for performing analysis. The study examines the most popular type of correlation coefficient, the Pearson correlation coefficient, which can be calculated using built-in functions in both programs. Historical weather data for the city of Yekaterinburg for December 2024 was used as a data set. The results will help users choose the most suitable tool for their analytical tasks. The conclusion briefly describes the advantages and disadvantages of each program and provides recommendations for choosing the optimal program.

Keywords: Microsoft Excel, LibreOffice Calc, correlation analysis, comparative analysis, analytics

Введение

С ростом объемов данных, которые генерируются в различных сферах нашей жизни – от социальных сетей и Интернета вещей до промышленных приложений и научных исследований – становится критически важным использовать современные инструменты и технологии, способные эффективно обрабатывать и анализировать большие массивы информации [1, с. 44].

Проведение сравнительного анализа реализации корреляционного анализа в Libre Office Calc и Microsoft Excel становится актуальной задачей, поскольку выбор программного обеспечения может существенно повлиять на результаты исследования и работу пользователя.

Корреляционный анализ – это статистический метод, который используется для оценки и измерения степени взаимосвязи между двумя или более переменными. Он позволяет определить, насколько изменение одной переменной связано с изменением другой [2, с. 75].

Данный инструмент помогает выявить наличие и направление связи между переменными: положительная корреляция означает, что с увеличением одной переменной другая также увеличивается, в то время как отрицательная корреляция указывает на то, что увеличение одной переменной связано с уменьшением другой. Он часто применяется в различных областях, таких как экономика, психология,

социология и медицина, для выявления закономерностей и построения прогнозов [3, с. 100].

Основным показателем, используемым в корреляционном анализе, является коэффициент корреляции Пирсона (*), который варьируется в диапазоне от -1 до +1. Коэффициент +1 указывает на идеальную положительную корреляцию, -1 – на идеальную отрицательную корреляцию, а 0 говорит о том, что между переменными нет линейной зависимости [4].

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (*)$$

Цель исследования – выявить сходства и различия в методах и результатах расчета корреляционных коэффициентов в обоих приложениях, сравнить точность встроенных методов, а также оценить удобство использования интерфейсов для выполнения анализа.

Набор данных

В качестве набора данных для сравнения использовался архив погоды в Екатеринбурге за промежуток с 3 по 9 декабря 2024 г. Данные были взяты с открытого источника – сайта метеонюостей hmn.ru. Пример данных для анализа представлен в таблице.

Пример данных для анализа

| Дата | Время | Температура, °C | Относительная влажность воздуха | Дата | Время | Температура, °C | Относительная влажность воздуха |
|----------|-------|-----------------|---------------------------------|----------|-------|-----------------|---------------------------------|
| 03.12.24 | 02:00 | -1 | 81 | 06.12.24 | 02:00 | -5 | 70 |
| | 05:00 | 0 | 80 | | 05:00 | -6 | 71 |
| | 08:00 | -1 | 84 | | 08:00 | -7 | 78 |
| | 11:00 | 0 | 83 | | 11:00 | -8 | 77 |
| | 14:00 | 0 | 82 | | 14:00 | -6 | 70 |
| | 17:00 | 0 | 83 | | 17:00 | -9 | 82 |
| | 20:00 | 0 | 84 | | 20:00 | -9 | 83 |
| | 23:00 | 0 | 86 | 23:00 | -9 | 83 | |
| 04.12.24 | 02:00 | -1 | 89 | 07.12.24 | 02:00 | -9 | 83 |
| | 05:00 | -2 | 88 | | 05:00 | -10 | 84 |
| | 08:00 | -3 | 89 | | 08:00 | -10 | 83 |
| | 11:00 | -3 | 89 | | 11:00 | -10 | 79 |
| | 14:00 | -3 | 85 | | 14:00 | -9 | 67 |
| | 17:00 | -3 | 89 | | 17:00 | -11 | 74 |
| | 20:00 | -2 | 90 | | 20:00 | -12 | 80 |
| | 23:00 | -3 | 88 | 23:00 | -13 | 78 | |
| 05.12.24 | 02:00 | -3 | 86 | 08.12.24 | 02:00 | -13 | 80 |
| | 05:00 | -3 | 85 | | 05:00 | -13 | 77 |
| | 08:00 | -4 | 82 | | 08:00 | -13 | 77 |
| | 11:00 | -3 | 83 | | 11:00 | -13 | 72 |
| | 14:00 | -4 | 68 | | 14:00 | -12 | 69 |
| | 17:00 | -4 | 74 | | 17:00 | -15 | 77 |
| | 20:00 | -4 | 73 | | 20:00 | -15 | 82 |
| | 23:00 | -4 | 71 | 23:00 | -14 | 81 | |
| 09.12.24 | 02:00 | -12 | 79 | 09.12.24 | 14:00 | -6 | 42 |
| | 05:00 | -10 | 70 | | 17:00 | -6 | 44 |
| | 08:00 | -8 | 51 | | 20:00 | -6 | 66 |
| | 11:00 | -7 | 53 | | 23:00 | -7 | 74 |

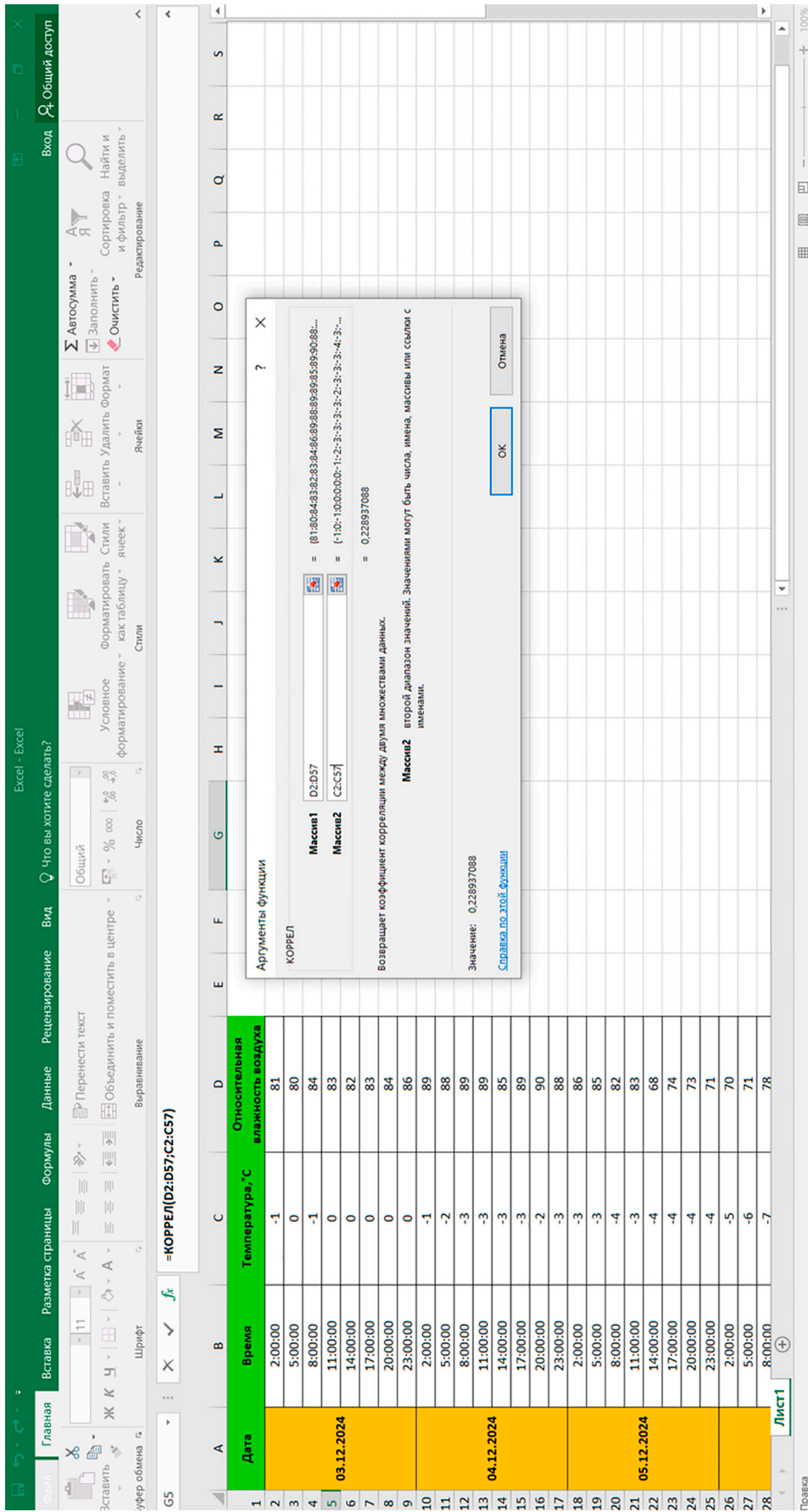


Рис. 1. Расчет коэффициента корреляции с помощью функции «КОРРЕЛ» в программе Microsoft Excel

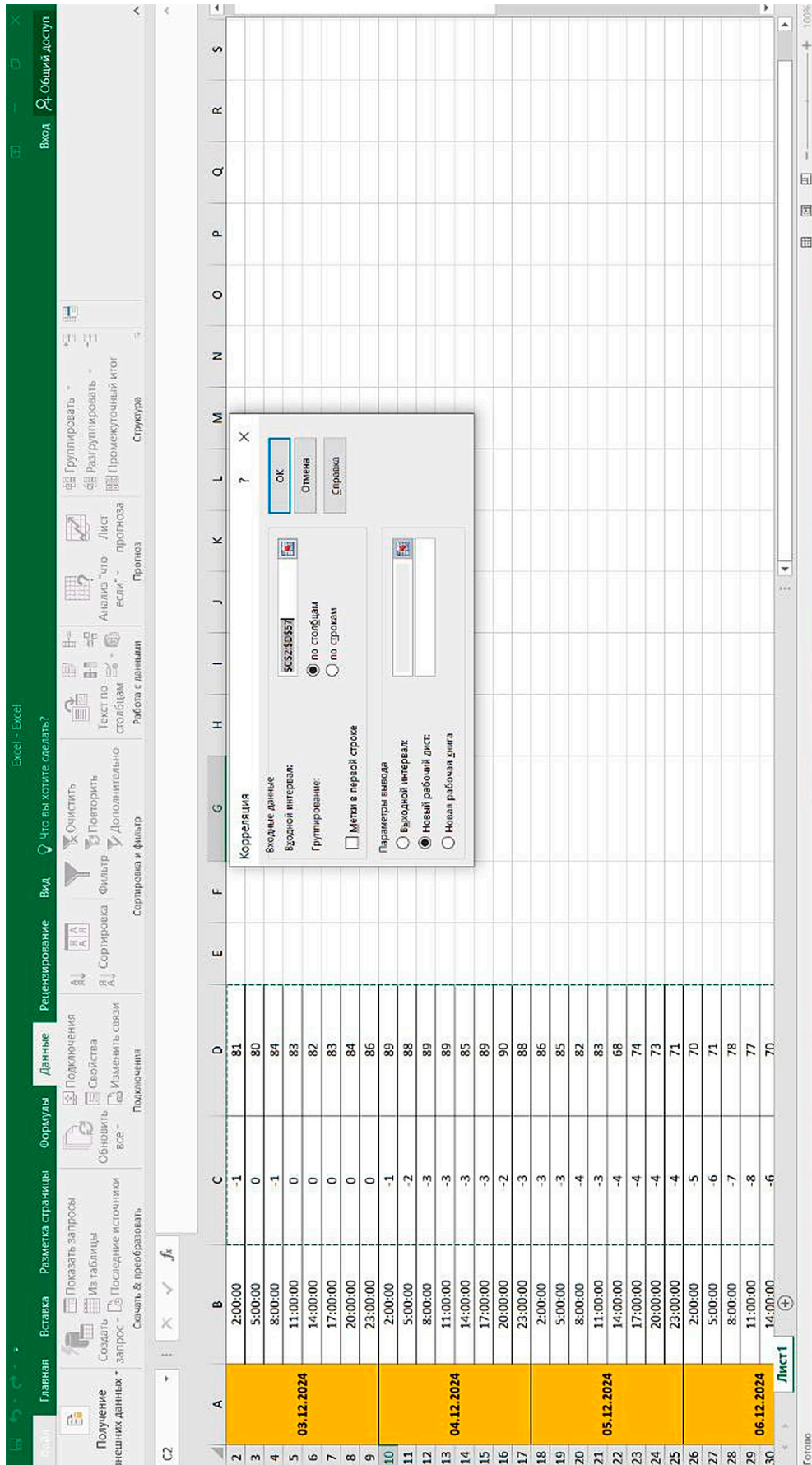


Рис. 2. Расчет коэффициента корреляции с помощью инструментов анализа в программе Microsoft Excel

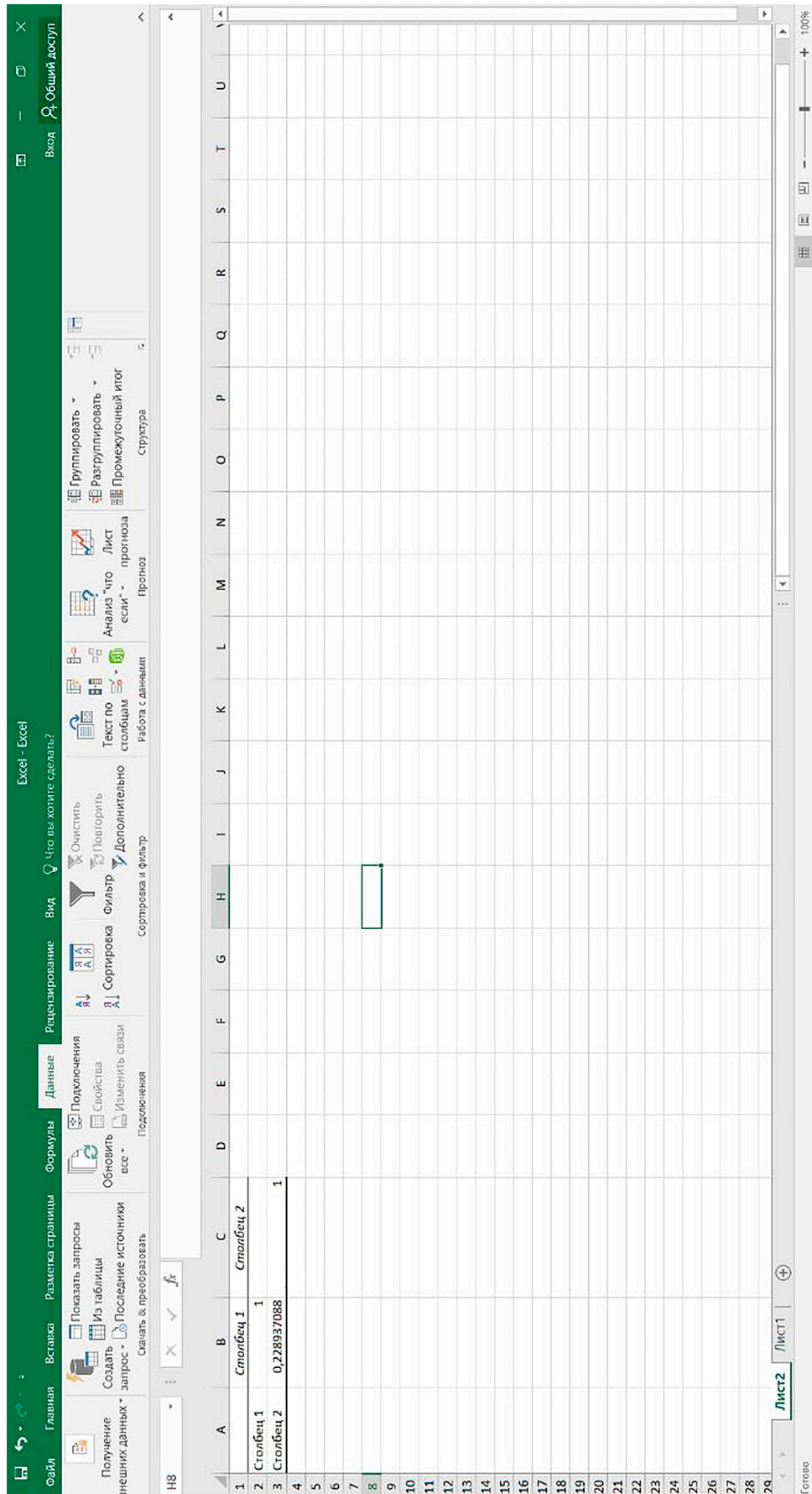


Рис. 3. Результат расчета в программе Microsoft Excel

LibreOffice Calc

Файл Правка Вид Вставка Формат Стили Лист Данные Сервис Окно Справка

| Дата | Время | Температура, °C | Относительная влажность воздуха |
|------------|----------|-----------------|---------------------------------|
| | 02:00:00 | -1 | 81 |
| | 05:00:00 | 0 | 80 |
| | 08:00:00 | -1 | 84 |
| 05.12.2024 | 11:00:00 | 0 | 83 |
| | 14:00:00 | 0 | 82 |
| | 17:00:00 | 0 | 83 |
| | 20:00:00 | 0 | 84 |
| | 23:00:00 | 0 | 86 |
| | 02:00:00 | -1 | 89 |
| | 05:00:00 | -2 | 88 |
| | 08:00:00 | -3 | 89 |
| 04.12.2024 | 11:00:00 | -3 | 89 |
| | 14:00:00 | -3 | 85 |
| | 17:00:00 | -3 | 89 |
| | 20:00:00 | -2 | 90 |
| | 23:00:00 | -3 | 88 |
| | 02:00:00 | -3 | 86 |
| | 05:00:00 | -3 | 85 |
| | 08:00:00 | -4 | 82 |
| 05.12.2024 | 11:00:00 | -3 | 83 |
| | 14:00:00 | -4 | 68 |
| | 17:00:00 | -4 | 74 |
| | 20:00:00 | -4 | 73 |
| | 23:00:00 | -4 | 71 |
| | 02:00:00 | -5 | 70 |
| | 05:00:00 | -6 | 71 |

Корреляция

Данные

Входной диапазон: \$Лист1.\$C\$2:\$D\$57

Результат в: \$F\$12

Группировать по Столбцам Строкам

Справка

OK Отменить

Рис. 4. Расчет коэффициента корреляции в программе Libreoffice Calc

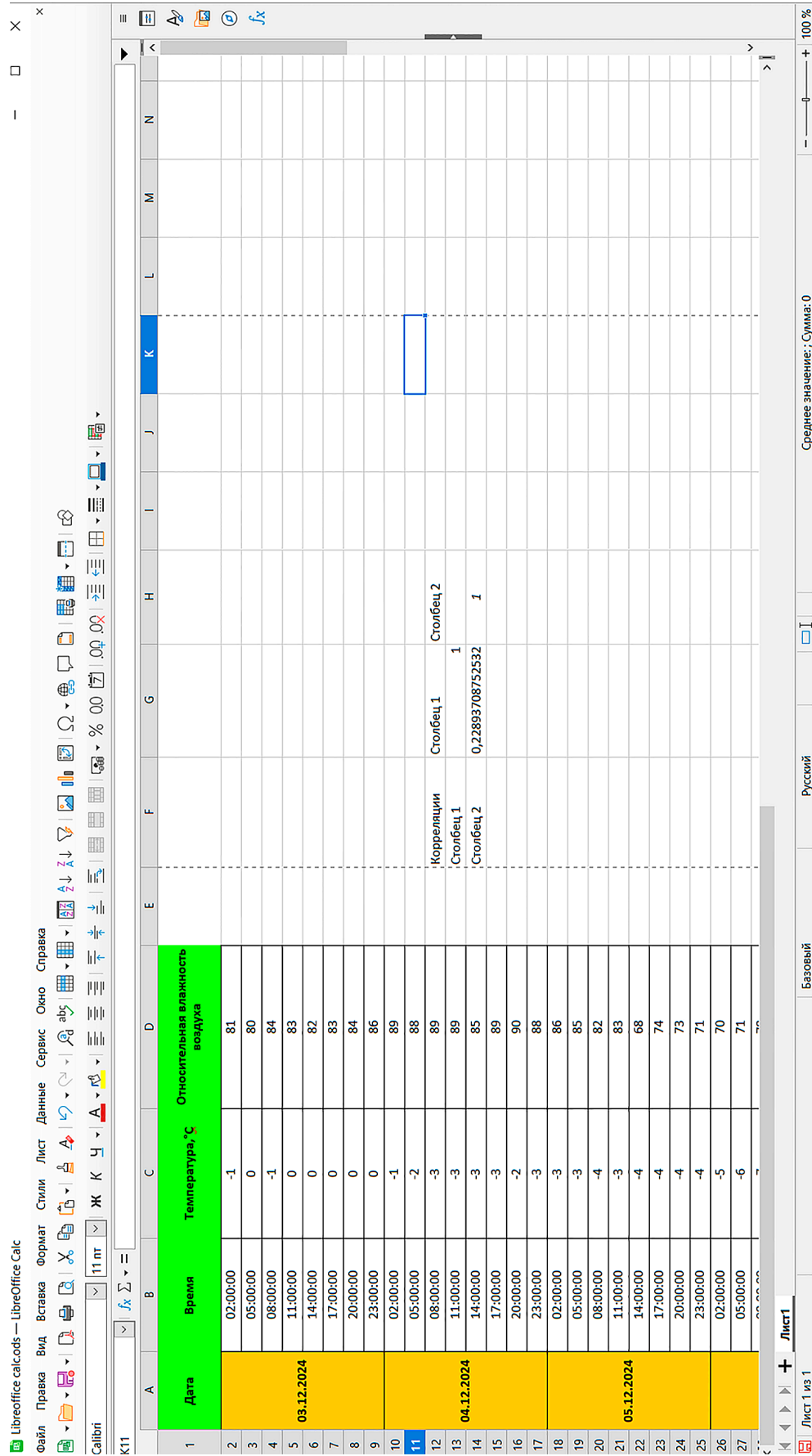


Рис. 5. Результат расчета в программе Libreoffice Calc

Для выявления взаимосвязи между анализируемыми данными были выбраны два самых значимых показателя – среднесуточная температура и относительная влажность воздуха.

Реализация в Microsoft Excel

Одним из способов, с помощью которого можно провести корреляционный анализ в Microsoft Excel, является использование функции «КОРРЕЛ». Для использования данной возможности необходимо открыть мастер функций и выбрать эту функцию. В качестве аргументов она принимает два массива данных, для которых будет определяться коэффициент корреляции [5, с. 38].

При использовании тестового набора данных результат получился примерно 0,22, что можно трактовать как слабая прямая зависимость. На рис. 1 представлен скриншот с результатом.

Кроме того, корреляцию можно вычислить с помощью одного из инструментов, который представлен в пакете анализа. Для этого нужно зайти в анализ данных и выбрать инструмент «корреляция». Далее выбирается нужный диапазон ячеек, который будет сравниваться (рис. 2, 3) [6, с. 129].

Как можно увидеть, на данных рисунках представлен коэффициент корреляции. Естественно, он оказался тот же, что и при использовании первого способа вычисления. Это объясняется тем, что оба варианта выполняют одни и те же вычисления, однако произвести их можно разными способами, что однозначно, можно отнести к плюсам программы Microsoft Excel, так как каждый исследователь сможет выбрать наиболее удобный для себя способ вычисления коэффициента корреляции.

Реализация в Libreoffice Calc

Для того, чтобы выполнить корреляцию в Libreoffice Calc, необходимо зайти в раздел «Данные» и выбрать из предложенного списка формул корреляцию. После этого нужно указать массив данных, влияние друг на друга которых будет рассматриваться. На рис. 4 представлен пример интерфейса для расчета коэффициента корреляции в программе Libreoffice Calc [7, с. 39].

Результат выполнения вставляется в выбранной ячейке таблицы, при этом выполняется вставка мини-таблицы, состоящей из нескольких столбцов и строк. Пример результата расчета коэффициента корреляции представлен на рис. 5.

Результат коэффициента корреляции оказался таким же, как и в Excel, с точностью до 9 знака, что является довольно высоким показателем. Из недостатков Libreoffice Calc можно отметить, что вставка результата осуществляется в ячейку в формате мини-таблицы, состоящей из нескольких столбцов и строк. Это плохо, так как новые данные могут заменить значения в существующих ячейках.

Заключение

Таким образом, сравнительный анализ реализации корреляционного анализа в Microsoft Excel и LibreOffice Calc выявил как сходства, так и различия в результатах расчета коэффициента корреляции Пирсона. Обе программы предоставляют функциональность для выполнения такого анализа, но незначительные различия в обработке данных могут приводить к небольшим отклонениям в полученных результатах.

Microsoft Excel, обладая более широким функционалом и возможностями, может быть предпочтительнее для сложных исследовательских задач, требующих, например, дополнительных визуализаций или углубленного статистического анализа.

LibreOffice Calc, являясь бесплатным и кроссплатформенным решением, остается удобной альтернативой для решения задач, не требующих высокой степени точности и детализации.

Список литературы

1. Гордиенко Е.П., Паненко Н.С. Современные технологии обработки и анализа больших данных в научных исследованиях // Актуальные проблемы железнодорожного транспорта. 2018. С. 44–48.
2. Баврина А.П., Борисов И.Б. Современные правила применения корреляционного анализа // Медицинский альманах. 2021. № 3 (68). С. 70–79.
3. Шишлянникова Л.М. Применение корреляционного анализа в психологии // Психологическая наука и образование. 2009. Т. 14, № 1. С. 98–107.
4. Кизянов А.О. Корреляция Пирсона на языке программирования Python // Постулат. 2018. № 8 (34). С. 14.
5. Слесарева Е.А. Корреляционный анализ как один из методов обработки результатов психологического исследования // Международный журнал психологии и педагогики в служебной деятельности. 2016. № 4. С. 38–40.
6. Береснева И.В. Применение методов корреляционного анализа в Microsoft Excel для выявления зависимостей между выборками значений двух случайных величин // Синтез науки и образования как инструмент решения глобальных проблем современности. 2023. С. 129–132.
7. Красников В.Л. Применение пакета LibreOffice для обработки экспериментальных данных по методу наименьших квадратов // Учебная физика. 2015. № 1. С. 39–40.