МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**Липецкий государственный технический университет**

Факультет автоматизации и информатики

Кафедра автоматизированных систем управления

Лабораторная работа №1

по дисциплине статистические методы.  
Работа со статическими данными

Студенты Станиславчук С.М.

Коновалов К.А.

Группа АС-21-1

Руководитель Рыжкова Д. В.

Липецк 2023 г.

2. Цель работы.

Сформировать из реальных данных, связанных по смыслу, массив. Вычислить основные коэффициенты и характеристики этих данных, произвести их корреляционный анализ.

Данные для выполнения работы взяты с сайтов: rosstat.gov.ru, ru.wikipedia.org.

3. Задание

1. Посчитать все выборочные значения коэффициентов корреляции (объясняющая переменная – объясняемая переменная).
2. Проверить значимость полученных коэффициентов.
3. Построить доверительные интервалы для выборочных коэффициентов корреляции (для одного коэффициента использовать преобразование z).
4. Проверить гипотезу об однородности для двух коэффициентов корреляции.
5. Посчитать корреляционное отношение (объясняющая переменная – объясняемая переменная).
6. Проверить значимость корреляционного отношения.
7. Построить доверительный интервал для корреляционного отношения.
8. Посчитать все частные коэффициенты корреляции.
9. Рассчитать выборочный множественный коэффициент корреляции.
10. Проверить значимость выборочного множественного коэффициента корреляции.
11. Построить доверительный интервал для выборочного множественного коэффициента корреляции.

Теоретическая часть

1. Индекс корреляции - наиболее общая характеристика степени тесноты связи между переменными. Для линейной связи его значение равно коэффициенту корреляции. Для нелинейных регрессионных моделей его значение используется для оценки значимости этих моделей по критерию Фишера.

2.1 Коэффициент корреляции – величина, характеризующая степень тесноты связи между случайными компонентами ξ и η, т.е. показывающая степень статистической зависимости между двумя числовыми переменными:

2.2 Выборочное значение коэффициента корреляции (т. е. статистическая оценка неизвестного значения *r*) подсчитывается по исходным статистическим данным:

2.3 Возникает необходимость при данном уровне значимости α проверить нулевую гипотезу H0={r=0} о равенстве нулю генерального коэффициента корреляции при конкурирующей гипотезе H1={rs≠0}.

В качестве критерия проверки нулевой гипотезы применяют случайную величину

Величина T при справедливости нулевой гипотезы имеет распределение Стьюдента с k=n-2 степенями свободы. Поэтому вычисляется эмпирическое значение критерия:

и по таблице критических точек распределения Стьюдента (см.рис.1) по выбранному уровню значимости α и числу степеней свободы k=n-2 находят критическую точку tкр(α;k).

         Если |Tэмп| > tкр, то нулевую гипотезу отвергают, и выборочный коэффициент корреляции значимо отличается от нуля, а X и Y коррелированы, т.е. связаны линейной зависимостью.

         Если |Tэмп| ≤ tкр, то нет оснований отвергать нулевую гипотезу и говорят, что выборочный коэффициент корреляции незначим, а X и Y не коррелированы, т.е. не связаны линейной зависимостью.

2.4 Доверительные интервалы для истинного значения коэффициента корреляции *r* строятся из условия нормальной распределенности. Концы интервала можно вычислять по приближенной формуле:

Концы доверительного интервала можно вычислять по приближенной формуле:

2.5 Для проверки гипотезы о том, что *k* рассмотренных генеральных совокупностей имеют один и тот же теоретический коэффициент корреляции используется следующий критерий однородности:

3.1 Корреляционное отношение – характеристика зависимости между случайными величинами. Оно несимметрично по отношению к исследуемым переменным и является неотрицательной величиной:

Эмпирическое корреляционное отношениеили Эмпирическое корреляционное отношение

3.2 Проверка значимости. Для построения соответствующего критерия воспользуемся фактом приближенной *F(k-1,n-k)* - распределенности случайной величины в предположении, что индекс корреляции равен нулю и что условные распределения зависимой переменной *η(х)* при любом фиксированном *х* описываются нормальным законом с постоянной дисперсией.

Если , то гипотеза об отсутствии корреляционной связи отвергается.

3.3 Доверительный интервал для истинного значения корреляционного отношения:

Где - распределенная по закону Фишера, случайная величина.

- число степеней свободы числителя

4.1 Коэффициент частной корреляции - измеряет тесноту линейной связи между отдельным фактором и результатом при устранении воздействия прочих факторов модели.

Т.о.

5.1 Множественный коэффициент корреляции - характеризует тесноту линейной корреляционной связи между одной случайной величиной и некоторым множеством случайных величин.

5.2 Свойства множественного коэффициента корреляции:

а)

б) Минимальное значение множественного коэффициента корреляции (*Rηξ=* 0) соответствует случаю полного отсутствия корреляционной связи.

в) Максимальное значение множественного коэффициента корреляции (*Rηξ*= 1) соответствует полному отсутствию варьирования «регрессионных остатков», что означает наличие чисто функциональной связи (т.е.не существует остатков).

5.3 Выборочное значение множественного коэффициента корреляции статистически отлично от нуля, следовательно, пользуются фактом *F(р, n-р-1)-*распределенность случайной величины.

Для проверки гипотезы о значимости выборочного множественного коэффициента корреляции используется следующая статистика:

Если , то гипотеза об отсутствии множественной корреляционной связи отвергается.

5.4 Доверительный интервал для истинного значения множественного коэффициента корреляции:

Практическая часть

1. Исходные данные

В качестве исходных данных нами была выбрана таблица – средняя продолжительность жизни по субъектам РФ. Часть таблицы представлена на Рис. 1

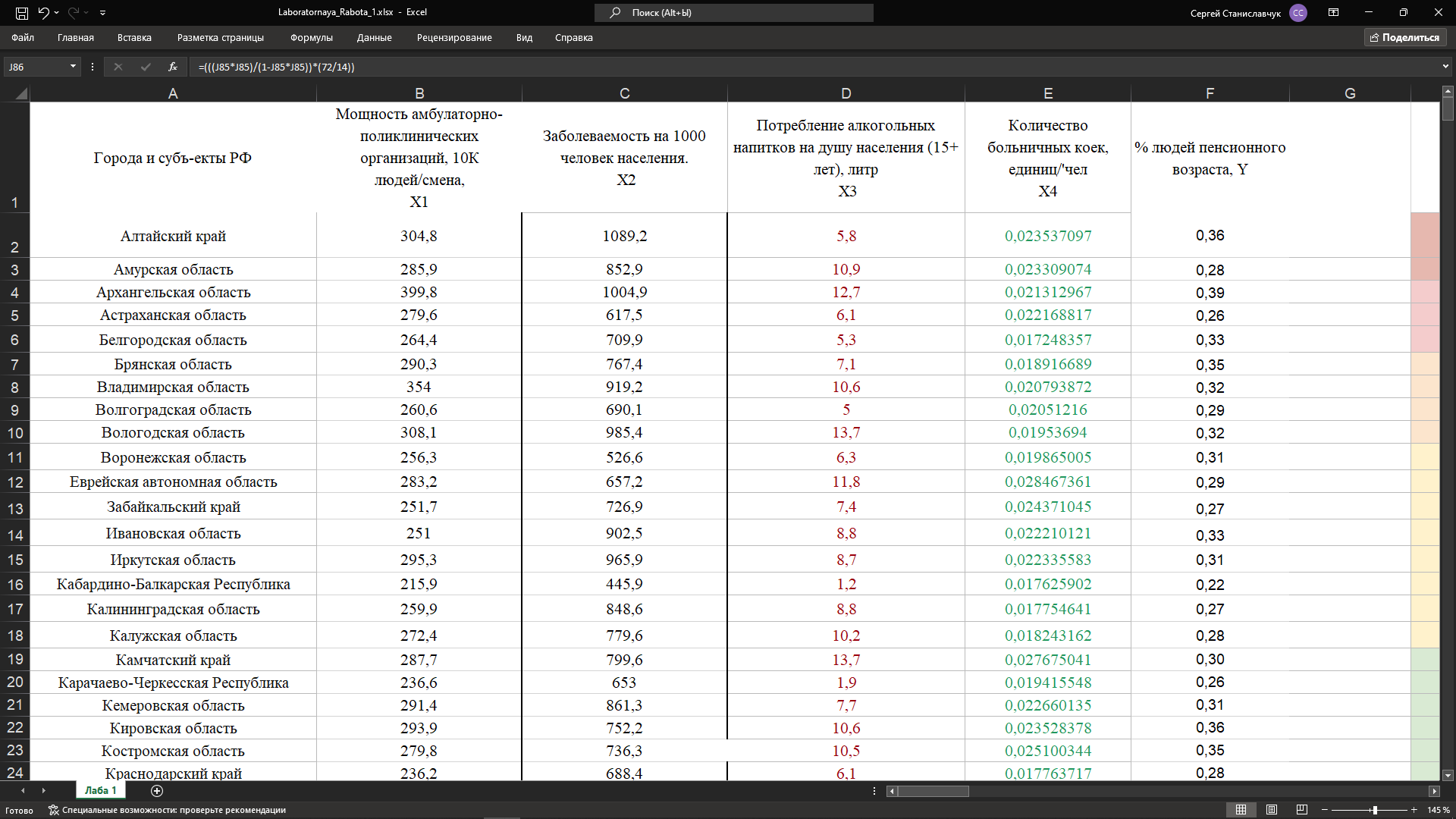


Рисунок 1. Исходные данные

2. Считаем выборочные коэффициенты корреляции. Результаты представлены на Рис. 2

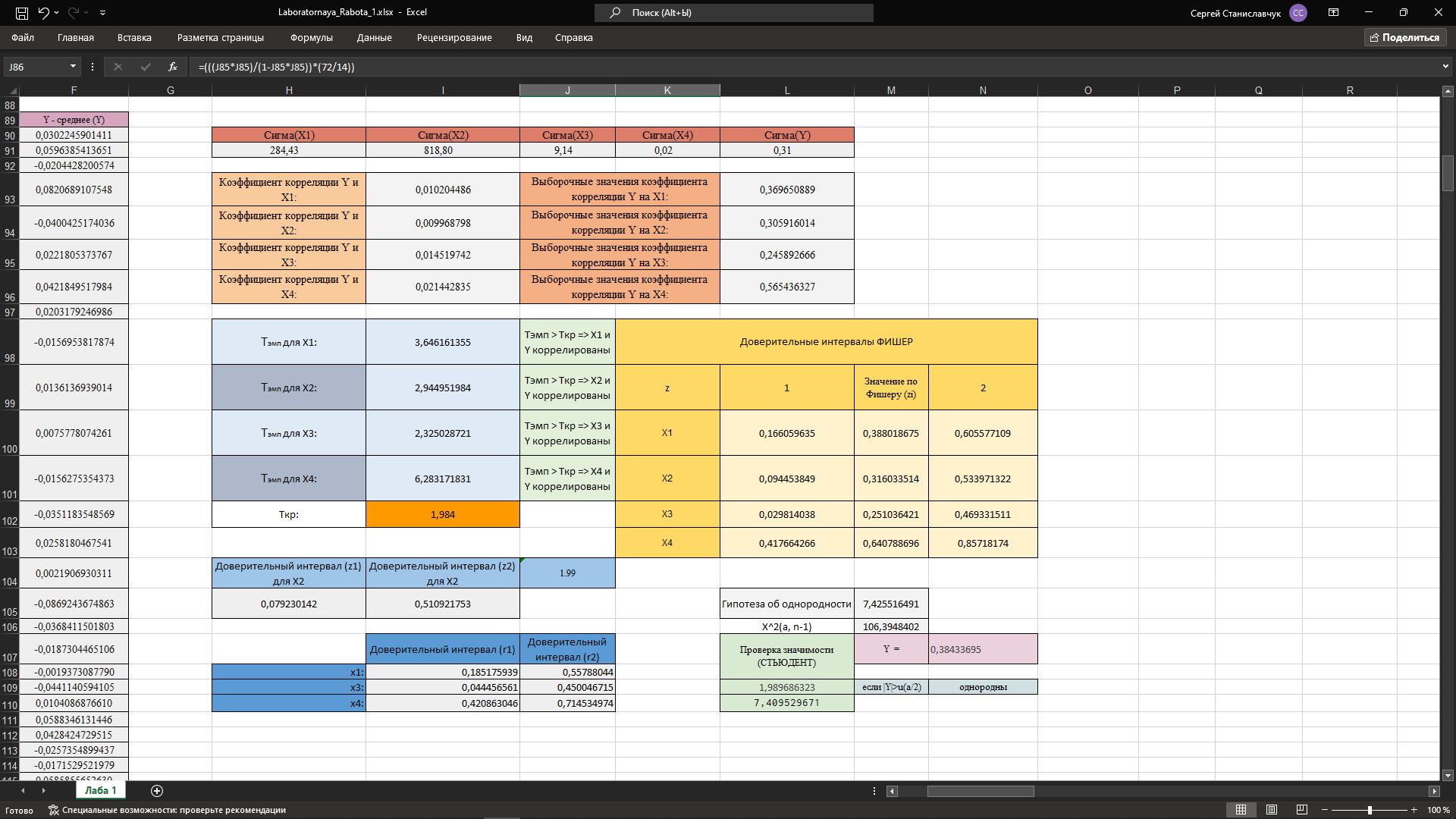


Рис. 2. Выборочные коэффициенты корреляции

3. Считаем доверительные интервалы для выборочного коэффициента корреляции (Рис. 3)

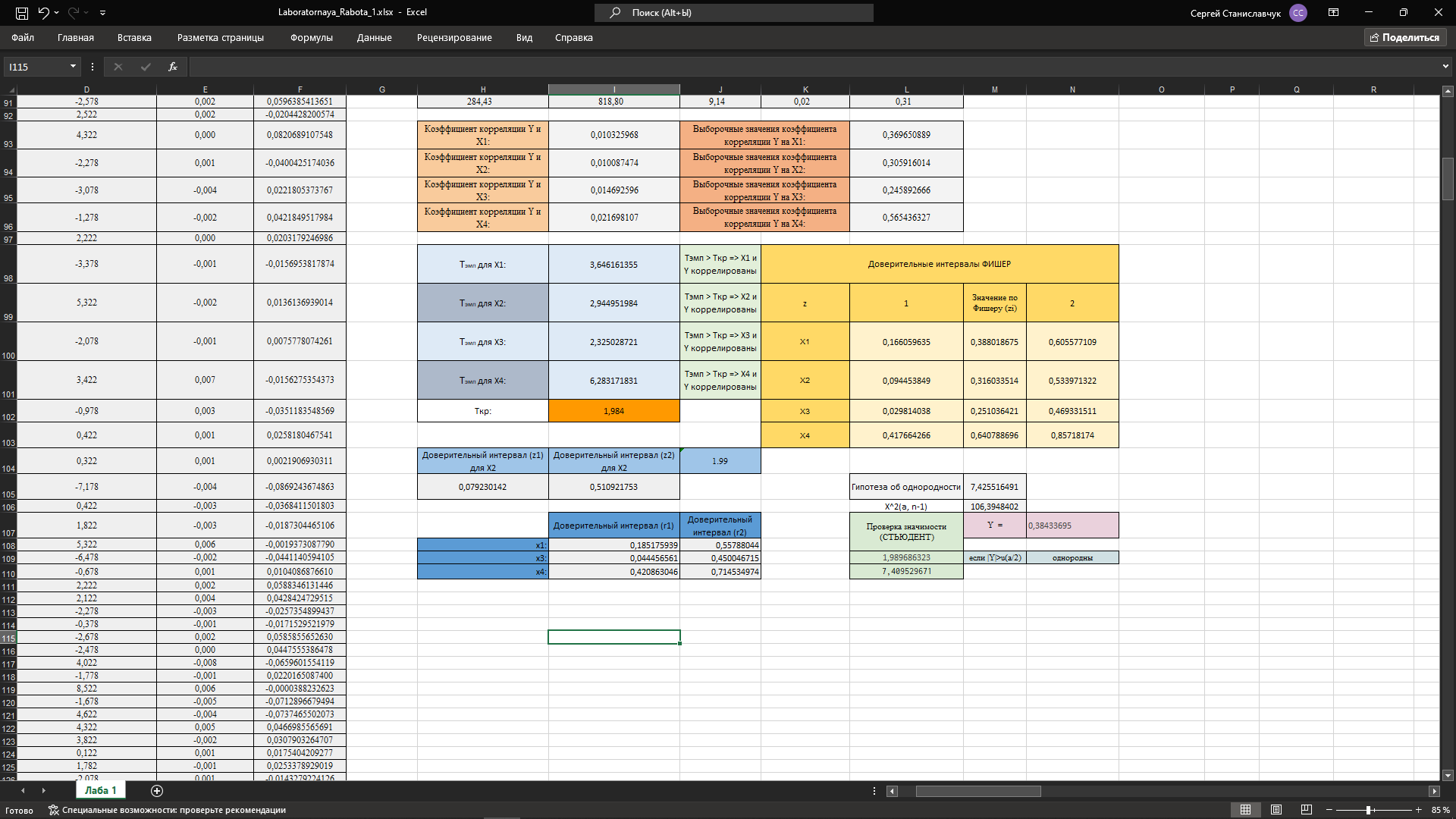


Рисунок 3. Доверительные интервалы

4. Выполнена проверка гипотезы об однородности двух коэффициентов корреляции, что и показано на Рис. 4.

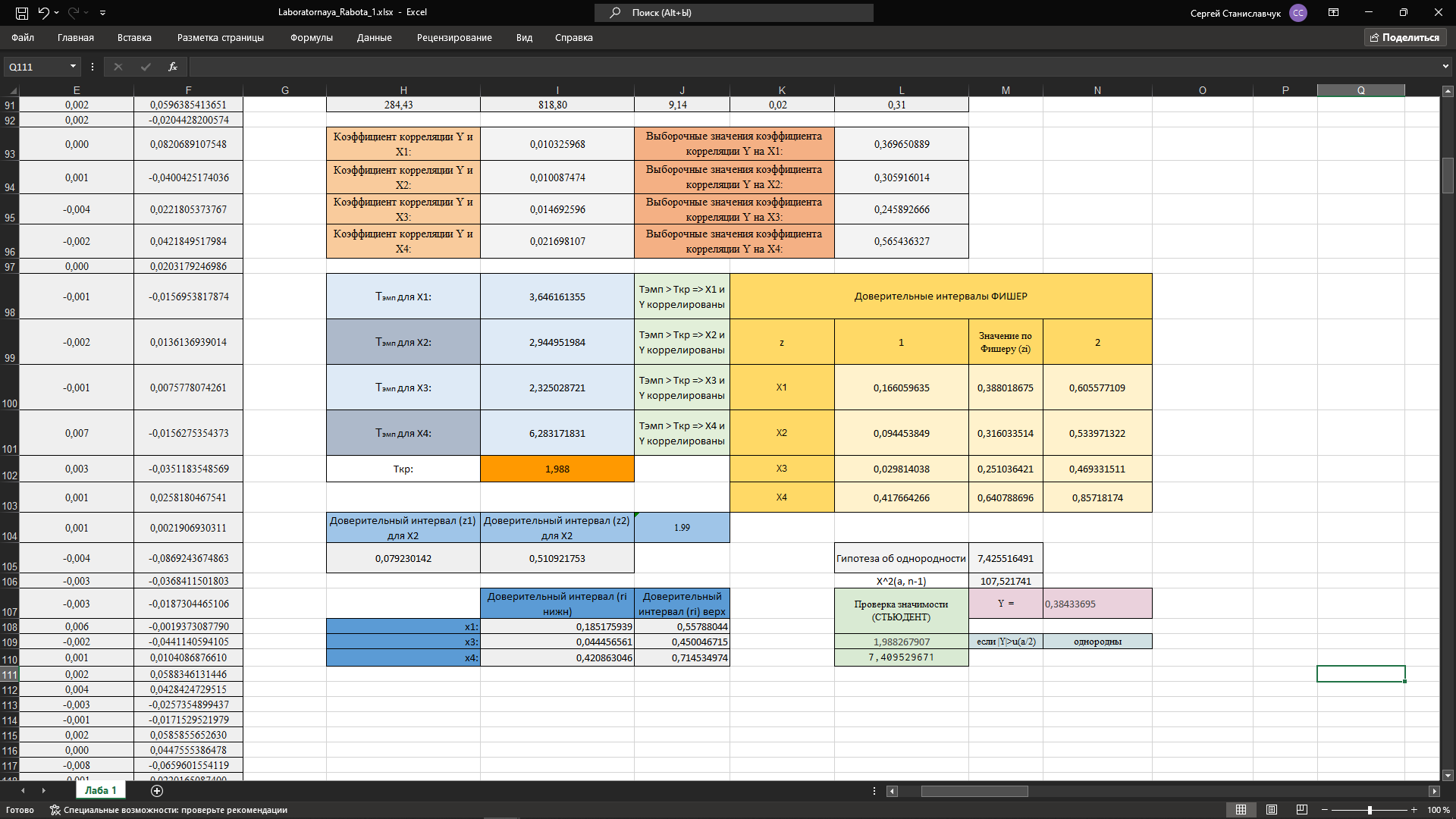
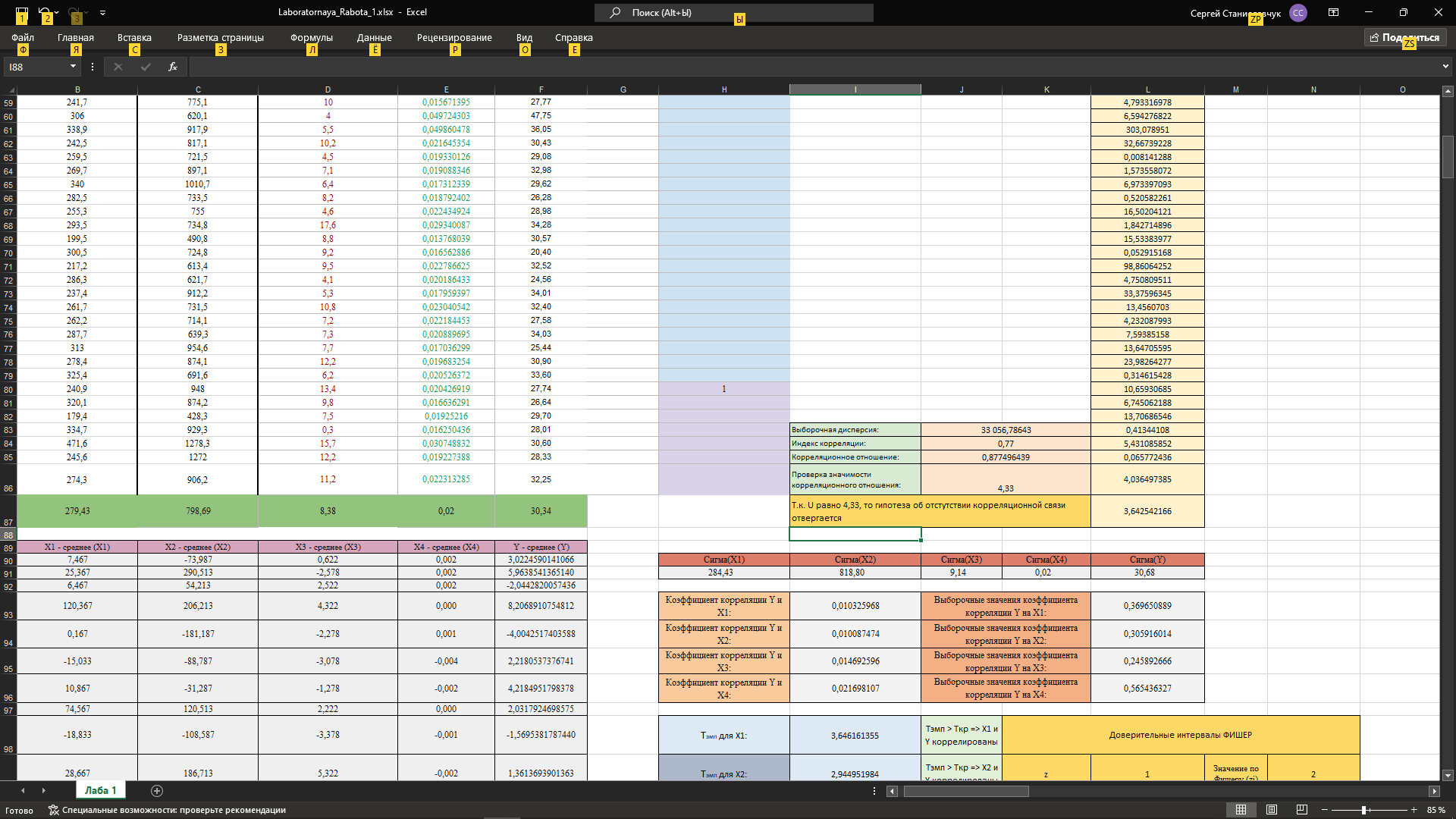


Рис. 4. Проверка гипотезы

5. Далее вычисляем корреляционное отношение. Для этого нам

понадобятся промежуточные вычисления, а именно: выборочная дисперсия, общая дисперсия, индекс корреляции (Рис. 5).



6. Считаем доверительный интервал для корреляционного отношения. Для этого предварительно вычисляем значимость коэффициента корреляционного отношения

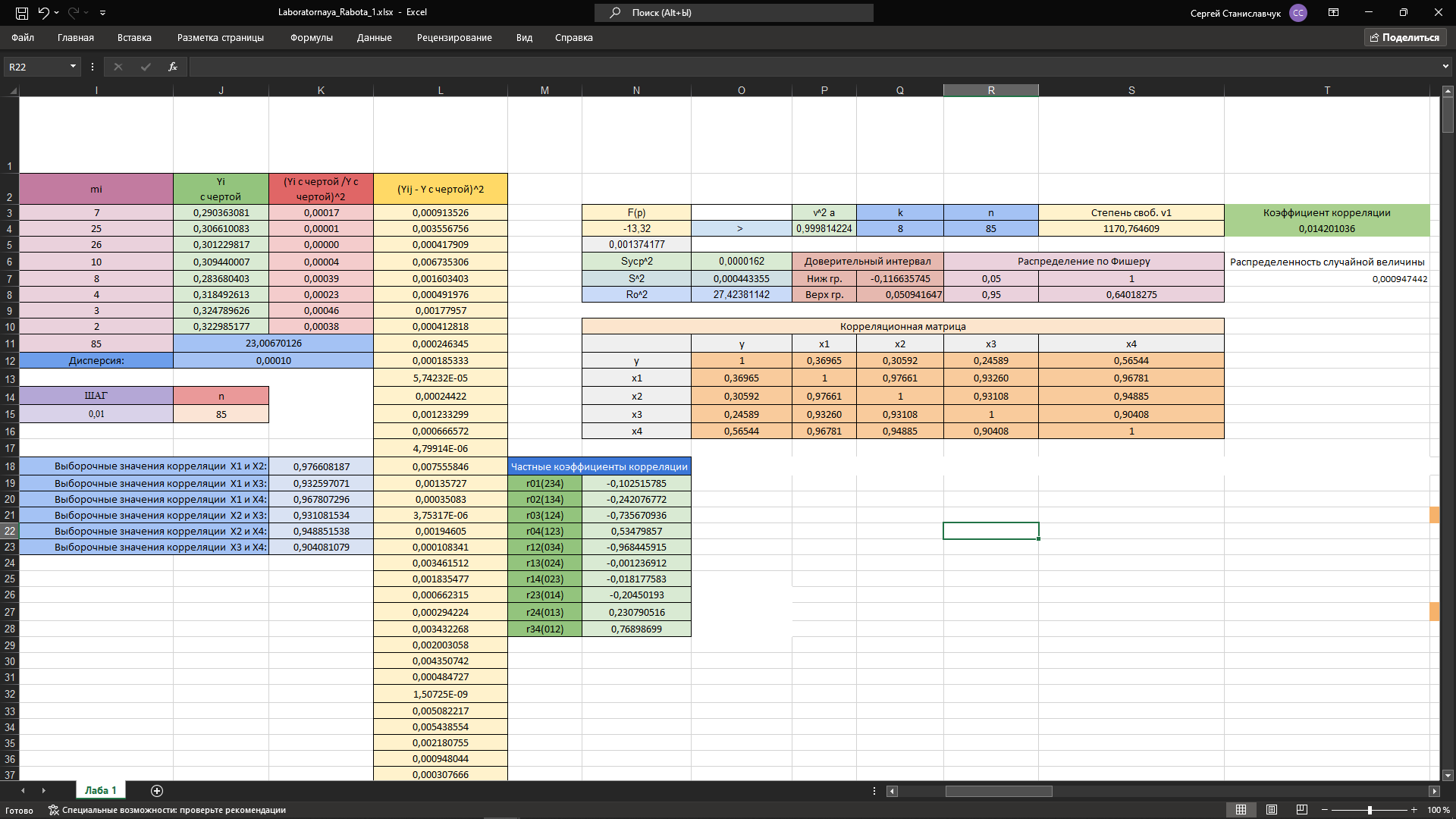


Рисунок – 6. Доверительный интервал9

7. Вычисляем все частные коэффициенты корреляции

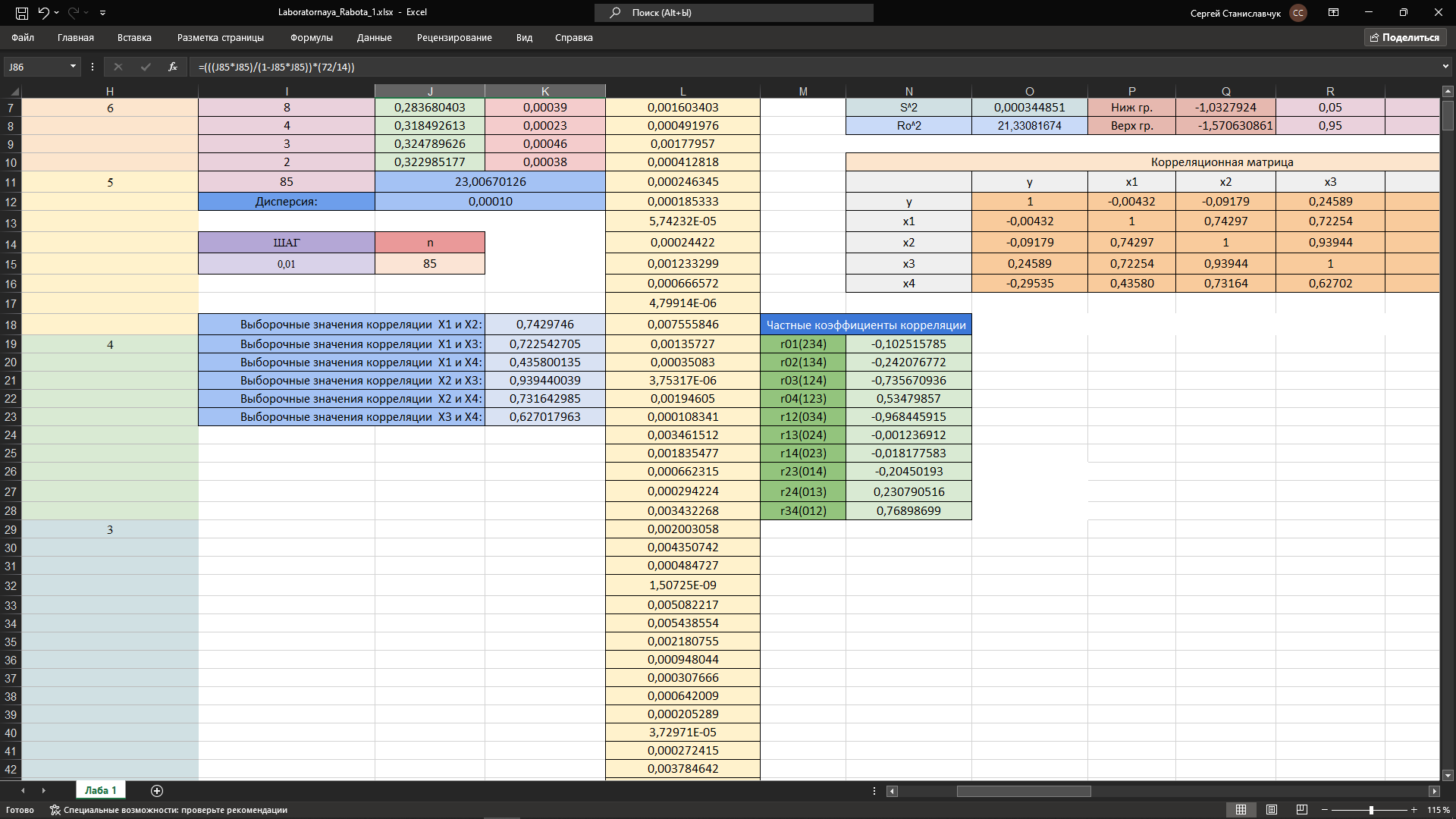


Рисунок – 7. Частные коэффициенты корреляции

8. Теперь вычисляем множественный коэффициент корреляции.

Проверим значимость выборочного множественного коэффициента корреляции. Построим доверительный интервал для выборочного коэффициента корреляции. Результаты представлены на Рис. 8.

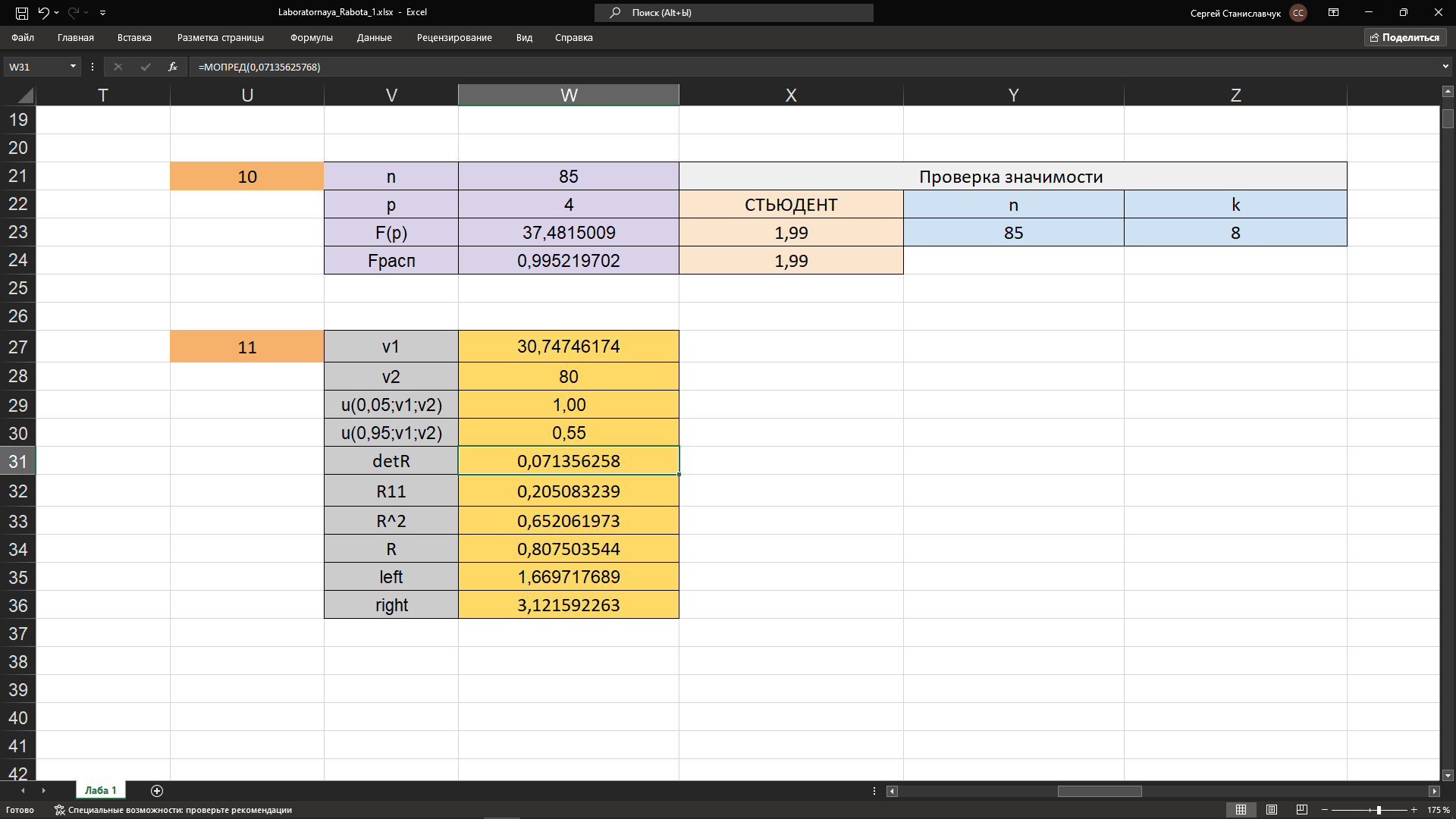


Рисунок 8. Результаты.

Вывод:

С помощью корреляционного анализа мы проверили статистическую таблицу государственной статистики о количестве населения пожилого возраста, на наличие в ней корреляционных и ковариационных связей и получили такие значения:

1. Выборочные значения коэффициентов корреляции: , , , . При проверке значимости полученных коэффициентов нулевую гипотезу не отвергаем.
2. Доверительные интервалы: , ,

.

1. При проверке гипотезы об однородности для двух коэффициентов корреляции получили значение 0,15, следовательно, различие статистически не значимо.
2. Корреляционное отношение равно 0,94.
3. При проверке значимости корреляционного отношения получаем значение U=4,33, следовательно, гипотеза об отсутствии корреляционной связи отвергается.
4. Доверительный интервал для корреляционного отношения: нижняя граница -0,11, верхняя – 0,05.
5. Частные коэффициенты корреляции представлены на Рис. 7.
6. Выборочный множественный коэффициент корреляции равен 0,80.
7. При проверке значимость выборочного множественного коэффициента корреляции приходим к выводу, что гипотеза об отсутствии множественной корреляционной связи отвергается.
8. Доверительный интервал для выборочного множественного коэффициента корреляции: нижняя граница 1,67, верхняя – 3,12.

Следовательно, данные связаны не сильно.