МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение высшего образования «Самарский национальный исследовательский университет   
имени академика С.П. Королева»

(Самарский университет)

Институт информатики, математики и электроники

Факультет информатики

Кафедра программных систем

Дисциплина методы планирования эксперимента и статистической обработки информации

**ОТЧЁТ**

по лабораторной работе №4 по дисциплине «Методы планирования эксперимента и статистической обработки информации»

по теме «Корреляционный и регрессионный анализ»

Студент группы № 6231-020402D В. А. Артамонов

Проверил В. В. Любимов

Дата сдачи:

Оценка:

Самара  
2022

Постановка задачи

Задание к лабораторной работе:

Провести регрессионный анализ данных с помощью программы MicrosoftOfficeExcel.

1. Постановка задачи, описание исходных данных

Коэффициент корреляции между двумя множествами данных позволяет найти программа Excel. Введем исходные данные на рабочем листе и с помощью мастера функций в категории Статистические выбираем функцию КОРРЕЛ. На экране появится диалог Аргумент функций. В окно Массив1вводим множество значений х, а в окно Массив2 множество у. Нажатием кнопки OK закрываем диалог и в заданной ячейке рабочего листа получаем величину коэффициента корреляции.

Программа MicrosoftOfficeExcel2007 позволяет одновременно провести дисперсионный, корреляционный и регрессионный анализ с оценкой значимости коэффициентов полученного уравнения регрессии. Проведем регрессионный анализ экспериментальных данных задачи 5.1с помощью программы Excel.

Выберем команду меню Данные, Анализ данных. На экране появится диалог Анализ данных. Щелкнем мышью на строке Регрессия в списке Инструменты анализа. Нажмем кнопку ОК, чтобы закрыть диалог Анализ данных. На экране появится диалог Регрессия для проведения регрессионного анализа данных.

Щелкнем мышью на поле ввода Входной интервал Y группы элементов управления Входные данные. Выделим диапазон ячеек С1:С4, чтобы указать первую группу данных. В поле Входной интервал Y группы Входные данные появится выбранный диапазон. Щелкнем мышью на поле ввода Входной интервал Х группы элементов управления Входные данные. Выделим диапазон ячеек В1:В4, чтобы указать первую группу данных. В поле Входной интервал Х группы Входные данные появится выбранный диапазон.

Установим флажок Метки в группе элементов управления Входные данные, т.к. указанные столбцы содержат названия. Если в группе элементов управления Параметры ввода не установлен переключатель Новый рабочий лист, то необходимо его установить, чтобы результаты регрессионного анализа были помещены на новый рабочий лист. Нажмем ОК, чтобы закрыть диалог Регрессия. На новом рабочем листе появятся результаты регрессионного анализа.

На рисунке 1 показана часть данных выборок, которые использовались для выполнения лабораторной работы.

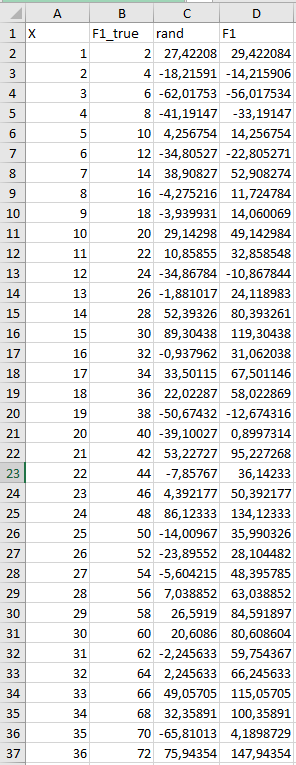


Рисунок 1 – Исходные данные лабораторной работы

1. Результаты

На рисунке 2 показаны результаты нахождения коэффициентов корреляции с помощью встроенной функции КОРЕЛЛ.

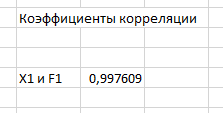


Рисунок 2 – Коэффициенты корреляции

На рисунке 3 показан регрессионный анализ для величин X1 и F1.

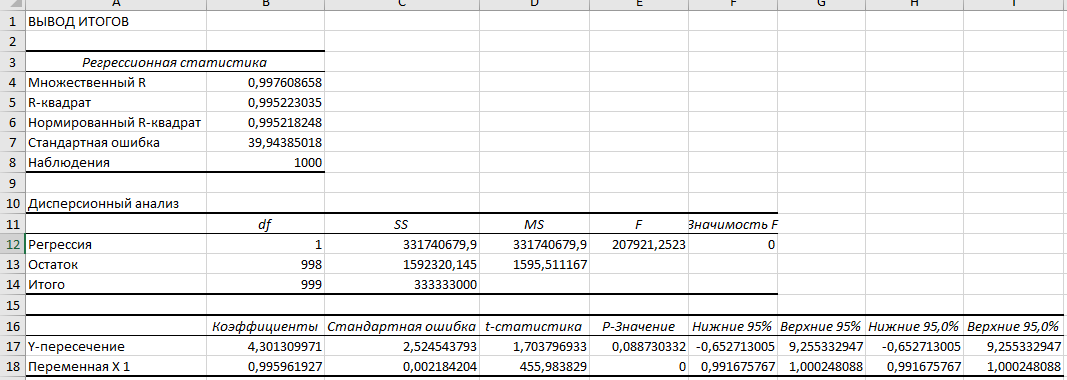


Рисунок 3 – Регрессионный анализ для величин X1 и F1

На рисунке 4 показано распределение точек исходной функции F1\_true и этих же точек со случайным смещением (F1).

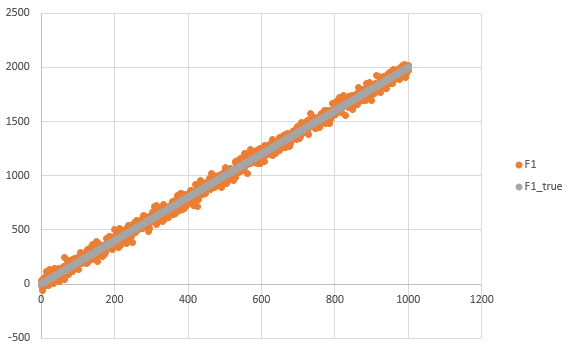


Рисунок 4 – Точки функций F1\_trueи F1

На рисунке 5 показано распределение точек функции F1 и функции, коэффициенты для которой найдены с помощью регрессионного анализа (F1\*).

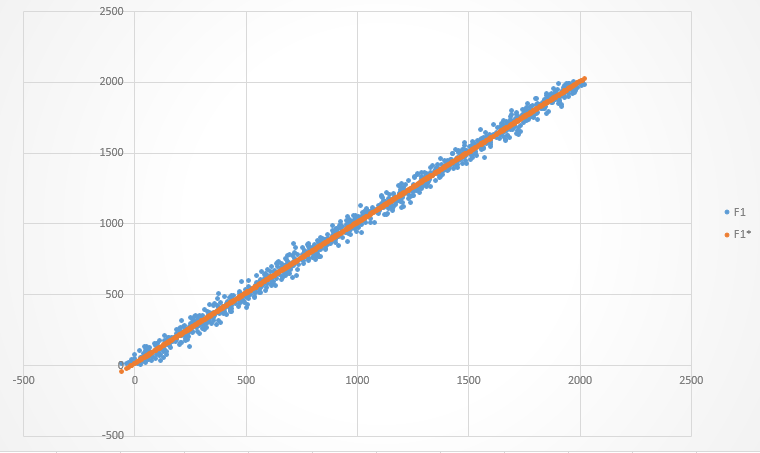


Рисунок 5 – Точки функций F1 и F1\*

Из рисунков 4 и 5 видно, что графики функций F1\_true и F1\* практически совпадают, что говорит об успешности проведения регрессионного анализа.

1. Выводы

В диапазоне ячеек В4:В8 расположены данные регрессионной статистики. В ячейке В5 находится коэффициент детерминированности 0,995, который вычисляется сравнением реальных данных и значений полученного уравнения. Если коэффициент близок к единице, то существует хорошая корреляция данных с уравнением, если коэффициент примерно равен нулю, то модель выбрана неудачно.

В ячейке В4 расположен коэффициент корреляции, равный квадратному корню из коэффициента детерминированности.

В ячейке В6находится нормированный коэффициент детерминированности. Данная величина нормирует простой коэффициент детерминированности по числу измерений.

В ячейке В7 расположена стандартная ошибка предсказания данных по уравнению регрессии.

В ячейке В8 находится количество измерений.

В диапазоне ячеек А11:F14 расположены результаты дисперсионного анализа регрессионных данных, которые аналогичны параметрам, рассмотренным в разделе 4.2. В строке 11находятся названия параметров, в строке 12–результаты, касающиеся регрессионного уравнения, в строке 13–данные, относящиеся к отклонению измерения от вычисленного по уравнению, в строке 14 – сумма строк 12 и 13.

В ячейке С12 расположена регрессионная сумма квадратов, т.е. сумма квадратов отклонений вычисленных значений от средней величины. В ячейке С13находится остаточная сумма квадратов, т.е. сумма квадратов отклонений измерений от вычисленных значений. Чем регрессионная сумма больше остаточной, тем лучше регрессионное уравнение описывает данные.

В столбце df расположены числа степеней свобод, которые определяются по количеству независимых элементов, требующихся для образования упоминавшихся сумм квадратов. В столбце MS находятся значения дисперсии.

В столбце F находятся значения F-статистики, в ячейке F12–вероятность того, что предлагаемое регрессионное уравнение не подходит для описания данных.

В диапазоне ячеек А16:I18 находятся коэффициенты регрессионного уравнения и их параметры. В строке 16расположены названия параметров, в строке 17–параметры, характеризующие свободный член уравнения, в строке 18–данные, представляющие коэффициент при независимой переменной в уравнении.

В столбце Коэффициенты расположены собственно коэффициенты уравнения.

В столбце Стандартная ошибка находятся стандартные ошибки или погрешности коэффициентов, которые подтверждают статическую значимость данных коэффициентов.

В столбце t-статистика расположены значения t-статистики, т.е. доли коэффициентов в стандартной погрешности. Чем больше отличается от нуля величина t-статистики, тем лучше.

В столбце Р-Значение расположены значения, характеризующие, насколько стандартную погрешность можно считать статистически значимой. В столбце Нижние 95% расположены нижние пределы коэффициентов с 95-процентным уровнем надежности. В столбце Верхние 95% содержаться верхние пределы коэффициентов уравнения с 95-процентным уровнем надежности.

Случай 1: X1 и F1.

Коэффициент корреляции (0.997) и коэффициент детерминированности (0.995) близки к единице, что говорит о хорошей корреляции между значениями Xи Y.

Регрессионная сумма квадратов (331740679,9) в несколько порядков превосходит остаточную сумму квадратов (1592320,145) что говорит о том, что регрессионное уравнение достаточно хорошо описывает данные.

Значимость Fполучилась равной 0, что говорит о том, что регрессионное уравнение идеально подходит для описания данных.

Регрессионный вид уравнения имеет вид:

Y = 4.301 + 0.996\*X

Стандартная ошибка для свободного члена (2.525) достаточно велика по отношению к своему коэффициенту (4.301), в то время как стандартная ошибка линейного члена (0.0022) членов вполне приемлема по отношению к своему коэффициенту (0.996).

С учётом допустимых интервалов для коэффициентов регрессионное уравнение выглядит следующим образом:

Y = 1.572 (±5.049) + 1.997 (±0.002) \*X

Из полученного уравнения видно, что значение коэффициента линейного члена существенно больше его доверительного интервала, что говорит о его значимости. Чего нельзя сказать о свободном члене.