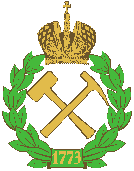
ПЕРВОЕ ВЫСШЕЕ ТЕХНИЧЕСКОЕ УЧЕБНОЕ ЗАВЕДЕНИЕ РОССИИ



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования

«САНКТ-ПЕТЕРБУРГСКИЙ ГОРНЫЙ УНИВЕРСИТЕТ»

Кафедра маркшейдерского дела

**Отчет по практической работе №2**

**«Регрессионный и корреляционный анализ»**

Вариант 8

Выполнил: студент гр. ГГ-18-2 \_\_\_\_\_\_\_\_ /Шевченко А.С./

(подпись) (Ф.И.О.)

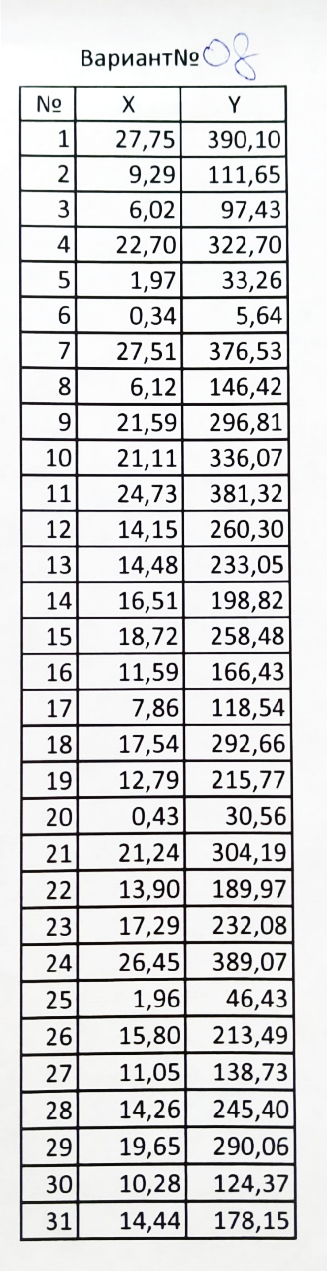
Проверил: доцент  **\_\_\_\_\_\_\_\_ /**Выстрчил М.Г./

(должность) (подпись) (Ф.И.О.)

Санкт-Петербург

2022

***Исходные данные:***



**Регрессионный анализ** – статистический аналитический метод, позволяющий вычислить предполагаемые отношения между зависимой переменной одной или несколькими независимыми переменными; устанавливает вид функциональной зависимости и подбирает функцию таким образом, чтобы она наилучшим образом определяла зависимость у от х. Цель регрессионного анализа – с помощью уравнения регрессии предсказать ожидаемое среднее значение результирующей переменной.

**Корреляционный анализ** – статистический метод, позволяющий с использованием коэффициентов корреляции определить, существует ли зависимость между переменными и насколько она сильна (есть ли связь между х и у). Корреляционная зависимость – это согласованные изменения двух (парная корреляционная связь) или большего количества признаков (множественная корреляционная связь). Суть ее заключается в том, что при изменении значения одной переменной происходит закономерное изменение (уменьшение или увеличение) другой(-их) переменной(-ых).

***1. Регрессионный анализ***

По исходным данным можно заметить, что с увеличением Х, растет У. И чтобы описать их зависимость какой-то функцией, необходимо провести прямую, которая бы наилучшим образом вписывалась в исходные значения. Критерием для такой прямой будет минимальная сумма квадратов отклонений [vv] = min.

Уравнение линейной регрессии имеет вид:

– уравненные параметры

– уравненное значение функции

Уравненное значение функции можно представить в виде:

– поправки в исходные значения функции

Отсюда выразим поправки как:

Уравненное значение функции можно так же представить в виде:

Где и – приближенные значения параметров, которые можно задать любыми значениями для линейной функции (для нелинейной их следует задавать максимально близкими к истинным)

- это свободный член (то, что хотим исправить), произвольно близкое к итоговому значению

Далее воспользуемся параметрическим способом уравнивания

Задача уравнивания состоит в определении поправок и параметров. Для этого применяем линеаризацию функций (приведение к линейному виду с помощью разложения в ряд Тейлора)

Где первые два слагаемых это результат разложения в ряд Тейлора

– вектор невязок свободных членов

Обозначим:

В итоге, получаем параметрические уравнения поправок:

Применяя условия наименьших квадратов поправок, запишем:

Для соблюдения условия минимума квадратов поправок достаточно взять из выражения частные производные по каждому параметру в отдельности и приравнять их к нулю. После нахождения производных по всем параметрам будем иметь систему нормальных уравнений:

Система параметрических уравнений поправок в матричной записи принимает вид

Входящие в это выражение матрицы в развернутом виде равны:

Матрица коэффициентов параметрических уравнений поправок :

|  |  |
| --- | --- |
| Вектор невязок свободных членов : | Вектор поправок : |

Матричная запись условия наименьших квадратов для равноточных измерений (матрица весов Р равна 1) имеет вид:

Матричная запись системы нормальных уравнений:

Где T находится как:

*–* матрица коэффициентов нормальных уравнений

- вектор свободных членов =

Исправленные значения имеют вид:



Рисунок 1 – Вычисление элементов в системе нормальных уравнений



Рисунок 2 – Расчет вектора поправок из системы нормальных уравнений

Уравнение регрессии:

***2. Корреляционный анализ***

Качество аппроксимации описывают через 2 параметра: коэффициент ковариации и коэффициент корреляции.

Ковариация оценивает силу линейной зависимости между двумя числовыми переменными X и Y. Знак ковариации указывает на вид линейной связи между рассматриваемыми величинами: если она > 0 - это означает прямую связь (при росте одной величины растет и другая), ковариация < 0 указывает на обратную связь. При ковариации = 0 линейная связь между переменными отсутствует.

Следовательно, с ростом X значения Y действительно увеличивается.

Коэффициент корреляции показывает тесноту линейной взаимосвязи и изменяется в диапазоне от -1 до 1. -1 означает полную (функциональную) линейную обратную взаимосвязь. 1 – полную (функциональную) линейную положительную взаимосвязь. 0 – отсутствие линейной корреляции (но не обязательно взаимосвязи).

Коэффициент корреляции:

Значение и близко к 1, следовательно, корреляция положительная, и значения X и Y обладают между собой тесной линейной связью.



Рисунок 3 – Расчет коэффициента ковариации



Рисунок 4 – Расчет коэффициента корреляции

Для оценки качества подбора уравнения регрессии определяется коэффициент детерминации.

Коэффициент детерминации – параметр, показывающий долю объясненной дисперсии. Чем ближе значение к 1, тем лучше регрессия описывает зависимость между результативным признаком и зависимой переменной.

Где рассчитывается как разница между значением у, полученным по уравнению регрессии, и исходным значением.

Так как значение близко к 1, то доля объясненной дисперсии зависимой переменной высока.

|  |  |
| --- | --- |
| Рисунок 4 – Значения У из уравнения регрессии | Рисунок 5 – Расчет коэффициента детерминации |

Рисунок 6 – Исходные данные, линейная регрессия и линия тренда с коэффициентом детерминации

***Вывод:*** по исходным экспериментальным данным в результате регрессионного анализа было получено уравнение линейной регрессии:

В результате корреляционного анализа был рассчитан коэффициент ковариации:

который указывает на сильную линейную зависимость между Х и У, а так же был рассчитан коэффициент корреляции:

который указывает на тесноту линейной зависимости между Х и У.

Для того чтобы выявить, какую долю дисперсии объясняет полученное уравнение регрессии, был рассчитан коэффициент детерминации:

что означает, что подобранное уравнение линейной регрессии достаточно хорошо описывает поведение исходных данных.