**МИНОБРНАУКИ РОССИИ**

**Санкт-Петербургский государственный**

**электротехнический университет**

**«ЛЭТИ» им. В.И. Ульянова (Ленина)**

**Кафедра математического обеспечения и применения ЭВМ**

отчет

**по лабораторной работе №5**

**по дисциплине «Статистические методы обработки экспериментальных данных»**

Тема: Элементы регрессионного анализа. Выборочные прямые среднеквадратической регрессии. Корреляционное отношение.

|  |  |  |
| --- | --- | --- |
| Студентка гр. 8382 |  | Звегинцева Е.Н. |
| Студент гр. 8382 |  | Мирончик П.Д. |
| Преподаватель |  | Середа А.-В.И. |

Санкт-Петербург

2022

**Цель работы.**

Ознакомление с основными положениями метода наименьших квадратов (МНК), со статистическими свойствами МНК оценок, с понятием функции регрессии и роли МНК в регрессионном анализе, с корреляционным отношением, как мерой тесноты произвольной (в том числе и линейной) корреляционной связи.

**Основные теоретические положения.**

Метод наименьших квадратов (МНК) — метод, основанный на поиске минимума суммы квадратов отклонений значений некоторых функций от заданного множества значений. МНК является одним из основных методов регрессионного анализа и применяется для оценки параметров регрессионных моделей на основе выборочных данных.



Т.е. необходимо найти такие значения переменных, чтобы значения функций были близки к .

Регрессионный анализ – это [статистический метод](https://ru.wikipedia.org/wiki/%D0%A1%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B5_%D0%BC%D0%B5%D1%82%D0%BE%D0%B4%D1%8B) исследования влияния одной или нескольких [независимых переменных](https://ru.wikipedia.org/wiki/%D0%9D%D0%B5%D0%B7%D0%B0%D0%B2%D0%B8%D1%81%D0%B8%D0%BC%D0%B0%D1%8F_%D0%B8_%D0%B7%D0%B0%D0%B2%D0%B8%D1%81%D0%B8%D0%BC%D0%B0%D1%8F_%D0%BF%D0%B5%D1%80%D0%B5%D0%BC%D0%B5%D0%BD%D0%BD%D1%8B%D0%B5) на [зависимую переменную](https://ru.wikipedia.org/wiki/%D0%9D%D0%B5%D0%B7%D0%B0%D0%B2%D0%B8%D1%81%D0%B8%D0%BC%D0%B0%D1%8F_%D0%B8_%D0%B7%D0%B0%D0%B2%D0%B8%D1%81%D0%B8%D0%BC%D0%B0%D1%8F_%D0%BF%D0%B5%D1%80%D0%B5%D0%BC%D0%B5%D0%BD%D0%BD%D1%8B%D0%B5) .

Линейные функции выборочной среднеквадратической регрессии:

где - выборочное среднее , соответственно, – выборочный коэффициент корреляции, и – статистические оценки среднеквадратических отклонений для выборок , .

Для оценки корреляционной зависимости между случайными величинами в общем, а не только линейной, может быть использовано так называемое корреляционной отношение.

Чтобы рассчитать выборочное корреляционное отношение нужно рассчитать внутригрупповую и межгрупповую дисперсии. Оценку общей дисперсии можно представить, как сумму:

Чтобы рассчитать выборочное корреляционное отношение к нужно рассчитать внутригрупповую и межгрупповую дисперсии.

Внутригрупповая дисперсия вычисляется по формуле (среднее дисперсий внутри каждой группы):

Межгрупповая дисперсия вычисляется по формуле (дисперсии средних значений внутри каждой группы относительно ):

Выборочное корреляционное отношение к определяется в соответствии с выражением:

И определяется следующими свойствами:

* – нет оснований предположить, что между величинами X и Y существует корреляционная зависимость
* – выборочные данные дают основание предположить, что между величинами X и Y существует функциональная зависимость
* – выборочные данные согласованы с предположением, что случайные величины X и Y связаны линейной корреляционной зависимостью.

**Постановка задачи.**

Для заданной двумерной выборки построить уравнения выборочных прямых среднеквадратической регрессии. Полученные линейные функции регрессии отобразить графически. Найти выборочное корреляционное отношение. Полученные результаты содержательно проинтерпретировать.

**Выполнение работы.**

Для выполнения данной работы была использована выборка, сформированная в первой лабораторной работе. Из лабораторных работ №2 и 4 берем выборочные параметры распределения величин *E* и *ν* (выборочное среднее, дисперсия, исправленная оценка среднеквадратического отклонения) и выборочный коэффициент корреляции, что представлены в таблице на рис.1.

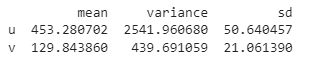


Рисунок 1 – Выборочные параметры из предыдущих лабораторных работ

Зависимость между двум величинами можно наблюдать на диаграмме рассеяния, представленной на рис.2.

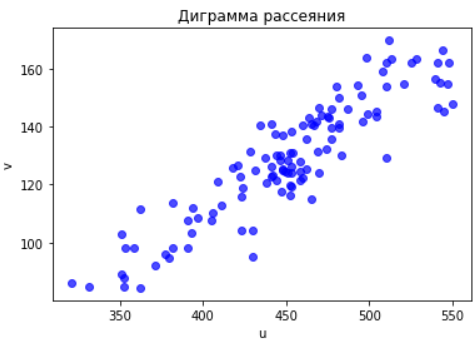


Рисунок 2 – Диаграмма рассеяния

Далее построим уравнение выборочной прямой среднеквадратичной регрессии, а также уравнение среднеквадратичной регрессии для величин *E* и *ν*:

Рассчитаем эти значения:



Данные регрессионные прямые можно отобразить на графике, представленном на рис.3. Точкой отмечена координата, которая является пересечением наших прямых.

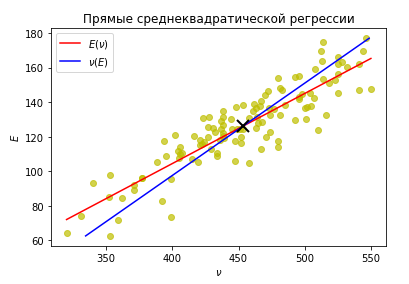
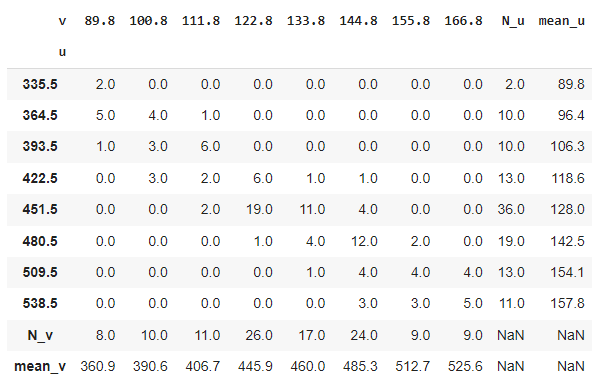


Рисунок 3 – график прямых среднеквадратичной регрессии

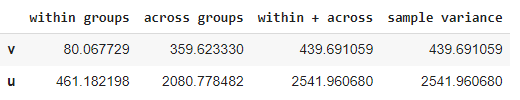
Далее можно найти дисперсию отклонений для величин *E* и  *v*, данная остаточная дисперсия равна:

Вычислим данные значения:

Можно наблюдать, что остаточная дисперсия величины *E* сильно меньше величины *v*, что можно наблюдать на рис.3, где по оси *E* точки меньше отступают от прямой *E(v)*, чем у второй величины.

Составим корреляционную таблицу для нахождения выборочного корреляционного отношения. Для фиксированных значений величин рассматриваем наблюдения (*E*, *v*) и (*v*, *E*) как группу и вычисляем средние значения величины в этой группе. Представим это в таблице.

Для каждой величины вычислим внутрегрупповую и межгрупповую дисперсии, а также их сумму:



Как мы можем наблюдать, сумма равна выборочной дисперсии, следовательно вычисления произведены верно.

Далее вычислим корреляционные отношения *E* к *v*, *v* к *E.*

Выборочное корреляционное отношение к определяется в соответствии с выражением:



= 0.904378

= 0.904750

Выборочное корреляционное отношение должно быть не меньше модуля выборочного коэффициента корреляции, который в нашем случае равен:

Как мы видим, значения корреляционных отношений соответствуют ожиданиям и близки к 1, следовательно между величинами присутствует сильная корреляционная зависимость.

Далее необходимо построить параболическую корреляционную кривую, определив значения коэффициентов с помощью МНК.

Для определения коэффициентов корреляционной кривой параболического вида была решена следующая система уравнений:

Система была решена с помощью написанной программы на языке Python матричным способом.В результате работы программы были получены следующие значения коэффициентов:



В связи с тем, что коэффициент *a* достаточно мал, то кривые будут близки к прямым, что можно увидеть на рис. 4.

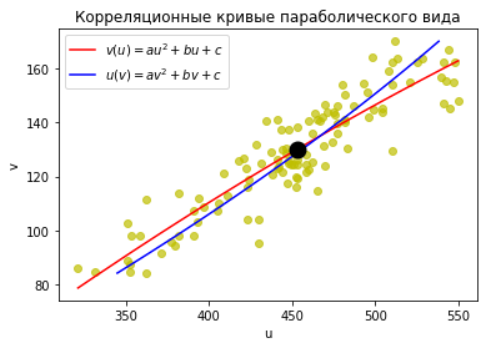
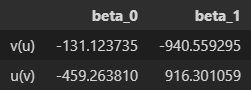


Рисунок 4 - Корреляционные кривые параболического вида

Согласно варианту, построим корреляционную кривую экспоненциальной функции . c

Для этого воспользуемся методом МНК и выразим с его помощью искомые переменные. После расчета на исходных данных получим следующие значения коэффициентов кривой обратной функции:



И построим график для найденных коэффициентов:

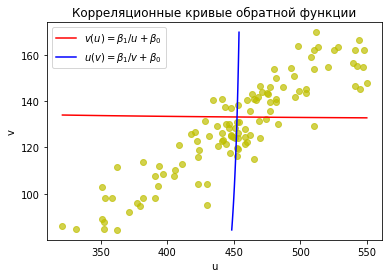


Рисунок 5 – Полученная кривая на множестве выборки

Видно, что полученные прямые очень плохо приближают значения распределения – вероятно, это связано с тем, что используемая функция не подходит для рассматриваемых данных, т.к. между распределениями величин *E* и *v* наблюдается прямая зависимость.

**Выводы.**

Была построена выборочная прямая линейной регрессии, коэффициенты которой вычислены с помощью метода наименьших квадратов. Было вычислено корреляционное отношение величины *E* к *ν* и величины *ν* к *E*, что подтвердило сильную корреляционную зависимость величин. Были построены корреляционные кривые параболического вида и обратной функции.