МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ

ФГБОУ ВО «АЛТАЙСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Институт цифровых технологий, электроники и физики (ИЦТЭФ)

Кафедра вычислительной техники и электроники (ВТиЭ)

Отчет по лабораторной работе № 4

**Корреляционный анализ экспериментальных данных**

(дисциплина «Метрология, стандартизация и сертификация»)

Выполнил студент 595 гр.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Д.В. Осипенко

Проверил: д.т.н, проф. каф. ВТиЭ

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Седалищев В.Н.

Лабораторная работа защищена

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2020 г.

Оценка \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Цель работы

1. Получить навыки проведения корреляционного анализа экспериментальных данных.

# Необходимые теоретические сведения

Диаграмма рассеяния (разброса) показывает взаимосвязь между двумя видами связанных данных и подтверждает их зависимость. Такими двумя видами данных могут быть характеристика качества и влияющий на неё фактор, две различных характеристики качества, два фактора, влияющих на одну характеристику качества, и т.д.

Для построения диаграммы рассеяния нужно не менее 30 пар данных . Оси и строят так, чтобы длины рабочих частей были примерно одинаковы. На диаграмму наносят точки , название диаграммы, число пар данных, названия осей. Точки, далеко отстоящие от основной группы, являются выбросами, и их исключают.



Возможны различные варианты скоплений точек. Пусть – выборка из наблюдений пары переменных . Для установления силы связи полезно вычислить коэффициент корреляции по формуле:



где и– выборочные средние, определяющиеся следующим образом:



Коэффициент корреляции используют только при линейной связи между величинами. Значение r находится в пределах от –1 до +1. Если r близко к 1, имеется сильная положительная корреляция (сильная связь между рядами данных). Если r близко к –1, имеется сильная отрицательная корреляция. При r, близком к 0, корреляция слабая (отсутствует). Если r близко к 0,6 (или –0,6), корреляционная зависимость считается существующей.

Характерные варианты скоплений точек показаны на рис. 1:

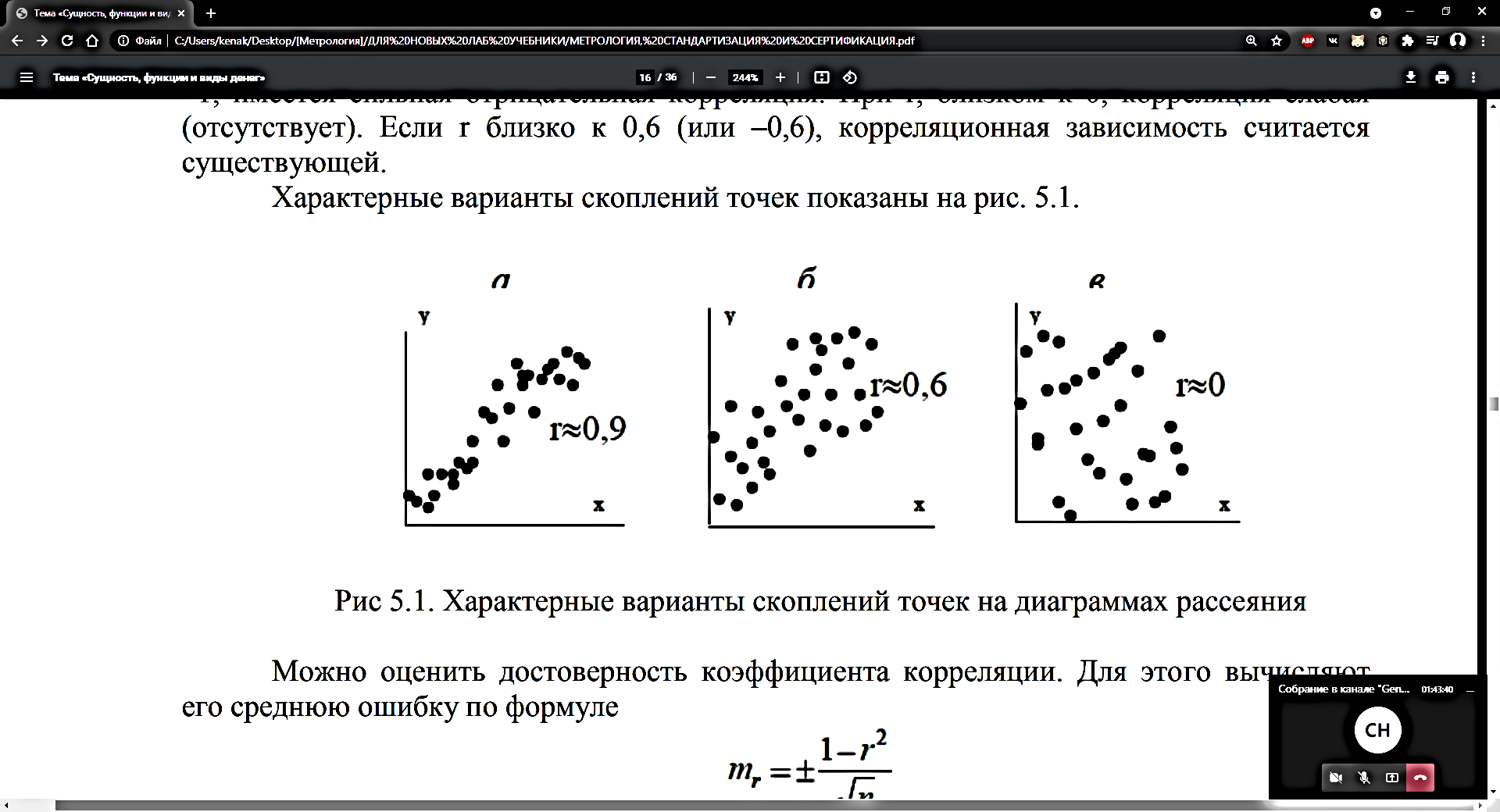


Рис. 1 Характерные варианты скоплений точек на диаграммах рассеяния

Можно оценить достоверность коэффициента корреляции. Для этого вычисляют его среднюю ошибку по формуле:



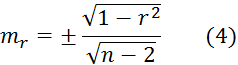
При коэффициент корреляции считается достоверным, т.е. связь доказана. А при . связь недостоверна.



В том случае, если число наблюдений меньше 100, но больше 30, точнее определять среднюю ошибку коэффициента корреляции, пользуясь формулой:



В случае малой выборки (число наблюдений меньше 30) для оценки достоверности коэффициента корреляции, т. е. для определения соответствия коэффициента корреляции, вычисленного по выборочным данным, действительным размерам связи в генеральной совокупности, средняя ошибка коэффициента корреляции определяется по формуле:



С достаточной надежностью о наличии той или иной степени связи можно утверждать только тогда, когда величина коэффициента корреляции превышает или равняется величине трех своих ошибок (). Обычно это отношение коэффициента корреляции к его средней ошибке обозначают буквой и называют *критерием достоверности*.



При коэффициент корреляции считается достоверным, т.е. связь доказана. А при . связь недостоверна.



# Схема подключения приборов

# Таблицы с полученными результатами измерений и расчетов

16 Вариант

|  |  |
| --- | --- |
| **X** | **Y** |
| **1729** | **1618,74** |
| **1794** | **1618,819** |
| **1748** | **1618,899** |
| **1582** | **1618,978** |
| **1590** | **1619,058** |
| **1628** | **1619,137** |
| **1798** | **1619,217** |
| **1342** | **1619,296** |
| **1652** | **1619,375** |
| **1302** | **1619,455** |
| **1805** | **1619,534** |
| **1889** | **1619,614** |
| **1572** | **1619,693** |
| **1619** | **1619,773** |
| **1894** | **1619,852** |
| **1631** | **1619,931** |
| **1471** | **1620,011** |
| **1468** | **1620,09** |
| **1534** | **1620,17** |
| **1607** | **1620,249** |
| **1370** | **1620,329** |
| **1431** | **1620,408** |
| **1395** | **1620,488** |
| **1374** | **1620,567** |
| **1745** | **1620,646** |
| **1721** | **1620,726** |
| **1812** | **1620,805** |
| **1617** | **1620,885** |
| **1316** | **1620,964** |
| **1899** | **1621,044** |
| **1833** | **1621,123** |
| **1322** | **1621,202** |
| **1877** | **1621,282** |
| **1585** | **1621,361** |
| **1408** | **1621,441** |
| **1785** | **1621,52** |
| **1884** | **1621,6** |
| **1719** | **1621,679** |
| **1586** | **1621,759** |
| **1352** | **1621,838** |
| **1346** | **1621,917** |
| **1727** | **1621,997** |
| **1703** | **1622,076** |
| **1806** | **1622,156** |
| **1492** | **1622,235** |
| **1759** | **1622,315** |
| **1614** | **1622,394** |
| **1707** | **1622,474** |
| **1810** | **1622,553** |
| **1423** | **1622,632** |
| **1312** | **1622,712** |
| **1675** | **1622,791** |
| **1635** | **1622,871** |
| **1451** | **1622,95** |
| **1891** | **1623,03** |
| **1304** | **1623,109** |
| **1872** | **1623,188** |
| **1677** | **1623,268** |
| **1754** | **1623,347** |
| **1621** | **1623,427** |

|  |  |
| --- | --- |
| **X\_avr** | **1621.0833333333333** |
| **Y\_avr** | **1621,083333** |
| **r** | **0,007519634** |
| **mr** | **0,130181549** |
| **t** | **0,057762669** |

# Вывод

В ходе данной работы была проведена статистическая обработка результатов пар наблюдений.

В итоге были рассчитан коэффициент корреляции изначальных данных и аппроксимирующей функции, равный 0.007, что близко к нулю. Из этого можно сделать вывод, что корреляция отсутствует. Также t = 0.05 < 3, из-за чего следует что связь не достоверна.

Приложение

**Текст программы:**

print('МСС: Лабораторная работа 3.\nВыполнил: Осипенко Данил Владимирович, студент 595 группы.\n')

import random

import numpy as np

import pandas as pd

import xlsxwriter

np.set\_printoptions(suppress=True)

y = np.array([1729, 1794, 1748, 1582, 1590, 1628, 1798, 1342, 1652, 1302, 1805, 1889, 1572, 1619, 1894, 1631, 1471, 1468, 1534, 1607, 1370, 1431, 1395, 1374, 1745, 1721, 1812, 1617, 1316, 1899, 1833, 1322, 1877, 1585, 1408, 1785, 1884, 1719, 1586, 1352, 1346, 1727, 1703, 1806, 1492, 1759, 1614, 1707, 1810, 1423, 1312, 1675, 1635, 1451, 1891, 1304, 1872, 1677, 1754, 1621])

x = np.arange(1,1+len(y))

n = 60

a = (n\*np.sum([x[i]\*y[i] for i in range(n)]) - np.sum(x)\*np.sum(y))/(n\*np.sum(x\*\*2) - np.sum(x)\*\*2)

b = (np.sum(x\*\*2)\*np.sum(y) - np.sum(x)\*np.sum([x[i]\*y[i] for i in range(n)]))/(n\*np.sum(x\*\*2) - np.sum(x)\*\*2)

y2 = np.array([a\*x[i]+b for i in range(n)])

So = np.sum(y\*\*2)/(n-2) - np.sum(y)\*\*2/(n\*(n-2)) - (n\*np.sum([x[i]\*y[i] for i in range(n)]) - np.sum(x)\*np.sum(y))\*\*2/(n\*(n-2)\*(n\*np.sum(x\*\*2) - np.sum(x)\*\*2))

Sa = So/(n\*np.sum(x\*\*2) - np.sum(x)\*\*2)

Sb = (So\*np.sum(x\*\*2))/(n\*np.sum(x\*\*2) - np.sum(x)\*\*2)

data\_pd = pd.DataFrame()

data\_pd['X'] = x

data\_pd['Y'] = y

data\_pd['Y2'] = y2

data\_pd['Y2+d'] = np.array([(a+np.sqrt(Sa))\*x[i]+(b+np.sqrt(Sb)) for i in range(n)])

data\_pd['Y2-d'] = np.array([(a-np.sqrt(Sa))\*x[i]+(b-np.sqrt(Sb)) for i in range(n)])

data\_pd2 = pd.DataFrame()

data\_pd2['a'] = ['b','So','Sa','Sb']

data\_pd2[f'{a}'] = [b,np.sqrt(So),np.sqrt(Sa),np.sqrt(Sb)]

data\_pd = pd.concat([data\_pd,data\_pd2],axis=1)

with pd.ExcelWriter('lab3.xlsx') as writer:

    data\_pd.to\_excel(writer,index=None, sheet\_name='lab3')