

## Отчет по лабораторной работе №3. Савва Даниил

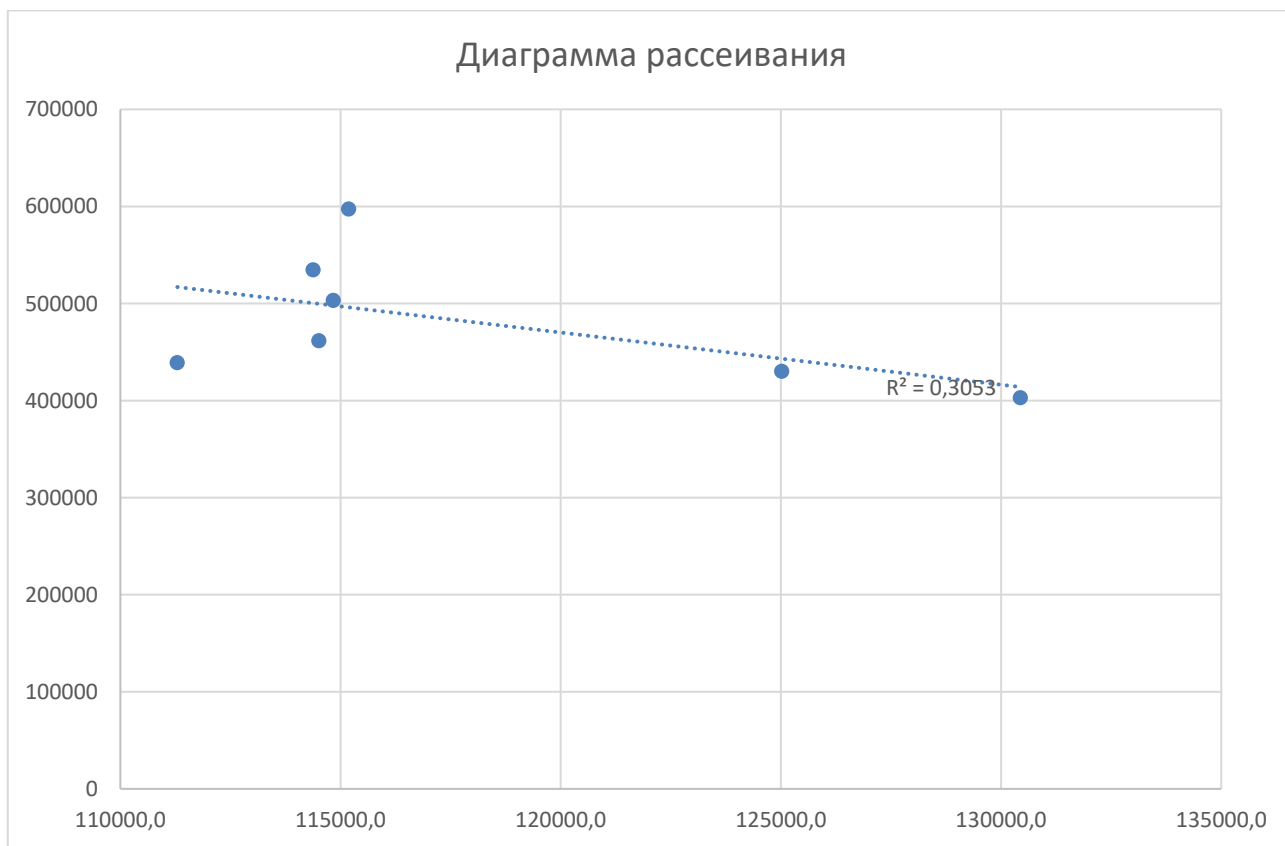
В данной лабораторной работе для исследования были использованы:

- количество заболевших и количество новорожденных за 2016-2022 годах,
- количество населения старше 16-ти лет по возрасту, полу и по состоянию в браке по субъектам российской федерации за 2020 год, количеством наблюдений около 100.

1. В данной части лабораторной работы происходит исследование предположительной зависимости влияние количества заболевших людей на количество новорожденных за период 2016-2022 годах (файл `stat_population_vs_illness.xlsx`). Расчеты проводились в Excel.

Полученный коэффициент корреляции равен  $-0,552533363$ , что означает слабую отрицательную корреляцию, а значит слабую зависимость данных друг от друга и другие факторы больше влияют на данные.

Была построена диаграмма рассеивания, проведена линия аппроксимации и вычислена величина достоверности аппроксимации  $R^2$ , который равен  $0.3053$ .



2. Для части 2.1, сделали случайную выборку по возрастам и мужчин, и женщин всего из датасета количества населения, возраста взяли по нарастающим, но из разных федеративных

округов. Для части 2.2, стратифицированной выборки, было выбрано население Белгородской области городское и сельское.

3. Данная часть работы была сделана с использованием python, код из main.py

```
import pandas as pd
from pandas import DataFrame
import numpy
import matplotlib.pyplot as plt
from scipy import stats

data_general: DataFrame = pd.read_excel('viborki.xlsx', header=0, sheet_name="Генеральная
совокупность")
data_random: DataFrame = pd.read_excel('viborki.xlsx', header=0, sheet_name="Случайная
выборка")
data_stratified: DataFrame = pd.read_excel('viborki.xlsx', header=0,
sheet_name="Белгородская область")

#print(data.head())

COLUMN_NAME: str = "Количество мужчин и женщин"

# Среднего значения по выборкам
print(f"Среднее значение (генеральная совокупность):
{data_general[COLUMN_NAME].mean():.2f}")
print(f"Среднее значение (случайная выборка): {data_random[COLUMN_NAME].mean():.2f}")
print(f"Среднее значение (стратифицированная выборка):
{data_stratified[COLUMN_NAME].mean():.2f}")

def confidence_interval(data: DataFrame, confidence: float) -> tuple:
    """Расчет доверительного интервала"""
    length: int = len(data)
    mean: float = numpy.mean(data)
    sem = stats.sem(data) # standard error of the mean
    margin = sem * stats.t.ppf((1 + confidence) / 2, length - 1)
    return mean - margin, mean + margin

# Доверительные интервалы для случайной выборки
print("\n")
print(f"Доверительный интервал - случайная выборка 90%:
{confidence_interval(data=data_random[COLUMN_NAME], confidence=0.90)}")
print(f"Доверительный интервал - случайная выборка 95%:
{confidence_interval(data=data_random[COLUMN_NAME], confidence=0.95)}")
print(f"Доверительный интервал - случайная выборка 99%:
{confidence_interval(data=data_random[COLUMN_NAME], confidence=0.99)}")

# Доверительные интервалы для стратифицированной выборки
print("\n")
print(f"Доверительный интервал - стратифицированная выборка 90%:
{confidence_interval(data=data_stratified[COLUMN_NAME], confidence=0.90)}")
```

```
print(f"Доверительный интервал - стратифицированная выборка 95%:  
{confidence_interval(data=data_stratified[COLUMN_NAME], confidence=0.95)}")  
print(f"Доверительный интервал - стратифицированная выборка 99%:  
{confidence_interval(data=data_stratified[COLUMN_NAME], confidence=0.99)}")
```

Среднее значение (генеральная совокупность): 8747999.86

Среднее значение (случайная выборка): 720177.57

Среднее значение (стратифицированная выборка): 92218.86

Доверительный интервал - случайная выборка 90%: 209448.59746567142, 1230906.5453914716

Доверительный интервал - случайная выборка 95%: 97137.27539240883, 1343217.8674647342

Доверительный интервал - случайная выборка 99%: -148548.8029722533, 1588903.9458293961

Доверительный интервал - стратифицированная выборка 90%: 77066.16071187195, 107371.55357384234

Доверительный интервал - стратифицированная выборка 95%: 73734.02281506569, 110703.6914706486

Доверительный интервал - стратифицированная выборка 99%: 66444.82100242407, 117992.89328329022

Из полученных данных можно предположить, что население по регионам российской федерации распределено крайне неравномерно