МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

ВЫСШЕГО ОБРАЗОВАНИЯ

«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ им. А.И. ГЕРЦЕНА»



Направление подготовки

09.03.01 – Информатика и вычислительная техника

Профиль «Технологии разработки программного обеспечения»

**Индивидуальное задание**

**“Исследование зависимости образования и возраста на уровень заработной платы”**

|  | Работу выполнили:  Шардт Максим,  очная форма обучения  курс: 2; группа: ИВТ-1.1 |
| --- | --- |
|  | Научный руководитель:  Профессор Власова Елена Зотиковна |

Санкт-Петербург

2023

## **Содержание**

[Содержание 2](#_oy55agxipus)

[Использованные формулы: 3](#_ndo2torp5lfy)

[Задание 1 5](#_y0kuels5jf3q)

[Решение 5](#_y0kuels5jf3q)

[Задание 2 8](#_y0kuels5jf3q)

[Решение 8](#_slud1tu1ao5s)

**Исследование зависимости образования и возраста на уровень заработной платы**

*Выполнил Шардт М.А.*

**Цель работы:** Вычислить значения коэффициентов корреляции. Построить корреляционное поле. Сделать предположение о форме и направлении взаимосвязи исследуемых показателей.

### *Использованные формулы:*

1. Коэффициент корреляции — параметр, который характеризует степень линейной взаимосвязи между двумя выборками, рассчитывается по формуле:

1. Коэффициент ранговой корреляции Спирмена:

,

где - квадраты разности рангов,

n - число наблюдений.

1. Статистика гипотезы Стьюдента:

Проведено исследование для изучения влияние возраста и образования на уровень заработной платы, в котором приняло участие 20 человек. Были собраны данные о уровне образования, каждому из уровней был присвоен собственный ранг для удобства вычислений:

| 1 | основное общее |
| --- | --- |
| 2 | среднее общее |
| 3 | среднее профессиональное |
| 4 | высшее профессиональное |

Таблица 1. Уровни образования

Результаты опроса приведены в таблице 2. n - порядковый номер участника, x1 - уровень образования, x2 - возраст, y - заработная плата.

| n | x1 | x2 | y |
| --- | --- | --- | --- |
| 1 | 4 | 20 | 45000 |
| 2 | 5 | 30 | 75000 |
| 3 | 3 | 45 | 43000 |
| 4 | 2 | 55 | 40000 |
| 5 | 1 | 19 | 47000 |
| 6 | 4 | 25 | 46000 |
| 7 | 5 | 35 | 80000 |
| 8 | 3 | 50 | 45000 |
| 9 | 2 | 60 | 41000 |
| 10 | 1 | 22 | 48000 |
| 11 | 4 | 28 | 47000 |
| 12 | 5 | 38 | 85000 |
| 13 | 3 | 52 | 46000 |
| 14 | 2 | 63 | 42000 |
| 15 | 1 | 21 | 48000 |
| 16 | 4 | 27 | 46000 |
| 17 | 5 | 37 | 82000 |
| 18 | 3 | 51 | 45000 |
| 19 | 2 | 62 | 41000 |
| 20 | 1 | 23 | 47000 |

Таблица 2. Результаты опроса

## Задание 1

Определить, существует ли связь между величиной y (зарплатой) и величинами x1 (уровень образования).

## Решение

Была создана вспомогательная таблица из исходных данных:

| n | x1i-x1 ср | (x1i-x1 ср)^2 | yi-y ср | (yi-y ср)^2 | (x1i-x1 ср)\*(yi-y ср) |
| --- | --- | --- | --- | --- | --- |
| 1 | 1 | 1 | -6000 | 36000000 | -6000 |
| 2 | 2 | 4 | 24000 | 576000000 | 48000 |
| 3 | 0 | 0 | -8000 | 64000000 | 0 |
| 4 | -1 | 1 | -11000 | 121000000 | 11000 |
| 5 | -2 | 4 | -4000 | 16000000 | 8000 |
| 6 | 1 | 1 | -5000 | 25000000 | -5000 |
| 7 | 2 | 4 | 29000 | 841000000 | 58000 |
| 8 | 0 | 0 | -6000 | 36000000 | 0 |
| 9 | -1 | 1 | -10000 | 100000000 | 10000 |
| 10 | -2 | 4 | -3000 | 9000000 | 6000 |
| 11 | 1 | 1 | -4000 | 16000000 | -4000 |
| 12 | 2 | 4 | 34000 | 1156000000 | 68000 |
| 13 | 0 | 0 | -5000 | 25000000 | 0 |
| 14 | -1 | 1 | -9000 | 81000000 | 9000 |
| 15 | -2 | 4 | -3000 | 9000000 | 6000 |
| 16 | 1 | 1 | -5000 | 25000000 | -5000 |
| 17 | 2 | 4 | 31000 | 961000000 | 62000 |
| 18 | 0 | 0 | -6000 | 36000000 | 0 |
| 19 | -1 | 1 | -10000 | 100000000 | 10000 |
| 20 | -2 | 4 | -4000 | 16000000 | 8000 |
| Среднее | 0 | 40 | 19000 | 4249000000 | 284000 |

Таблица 3. Вспомогательная таблица

1. Нулевая гипотеза – взаимосвязь между уровнем образования и зарплатой отсутствует.
2. Конкурирующая гипотеза – взаимосвязь между уровнем образования и зарплатой существует.

По таблице “Критические значения корреляции r Пирсона” было определено критическое значение для значения коэффициента корреляции, где n = 20, k = 18, P = 0.05.

(5)

После чего пользуясь данными таблицы были найдены коэффициенты корреляции:

Также был составлено корреляционное поле:

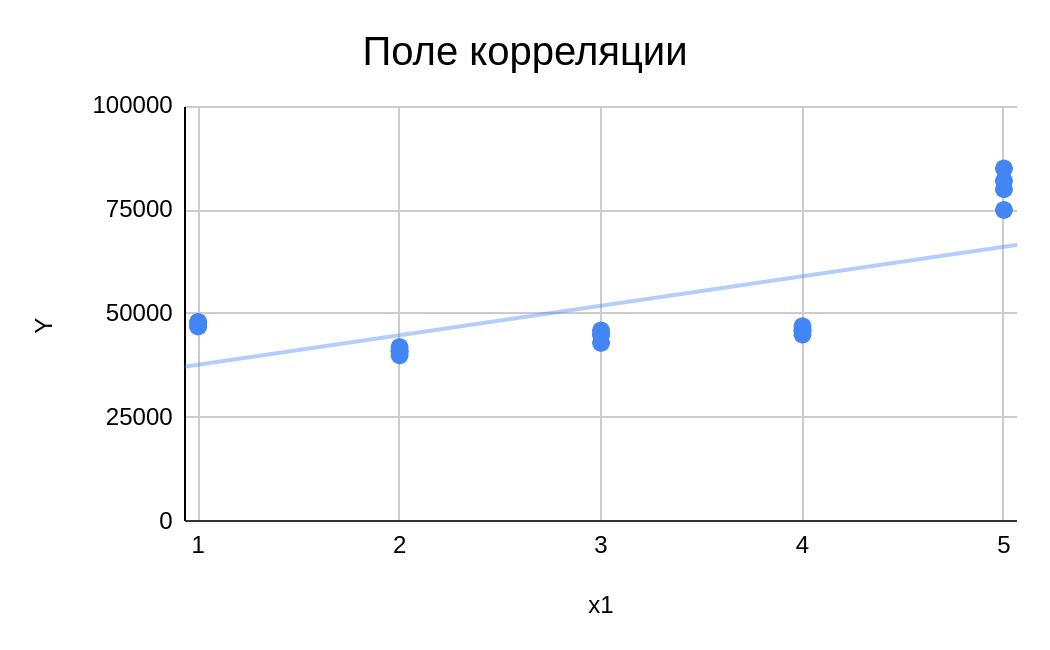


Рис 1. Корреляционное поле зарплаты от уровня образования

> , следовательно, гипотеза Н0 отвергается и принимается гипотеза H1. Также гипотеза Н1 соответствует наблюдаемым результатам корреляционного поля.

**Ответ:** Между уровнем образования и заработной платой существует положительная корреляция.

## Задание 2

Определить, существует ли связь между величиной y (зарплатой) и величинами x2 (возраст).

## Решение

В таблицах 2 и 3 содержатся все необходимые данные для вычислений.

1. Нулевая гипотеза – взаимосвязь между образованием и зарплатой отсутствует.
2. Конкурирующая гипотеза – взаимосвязь между образованием и зарплатой существует.

Пользуясь данными таблицы были найдены коэффициенты корреляции:

Также был составлено корреляционное поле:

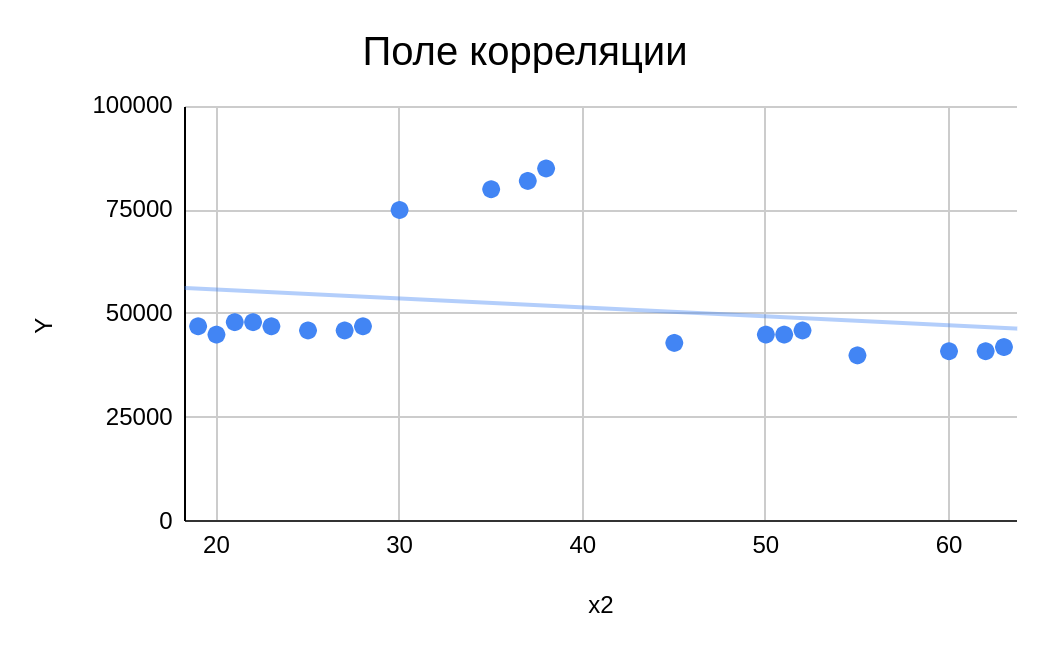


Рис 2. Корреляционное поле зарплаты от возраста

< (5), следовательно, гипотеза Н1 отвергается и принимается гипотеза H0. По корреляционному полю можно заметить скачок значений Y при значениях X2 от 30 до 40, что может свидетельствовать о нелинейной зависимости.

**Ответ:** Между зарплатой и возрастом корреляция не является существенно значимой. Однако, возможна нелинейная зависимость между этими величинами.