# ЛАБОРАТОРНАЯ РАБОТА №4
# «ИССЛЕДОВАНИЕ ИНСТРУМЕНТОВ ПРЕДСТАВЛЕНИЯ И СТАТИСТИЧЕСКОГО АНАЛИЗА ИНФОРМАЦИИ»

**Цель работы:**

Представить исходную информацию в виде функциональной (или корреляционной) таблицы y = f(x); изучить взаимосвязи эколого-экономических явлений, построить математическую модель эколого-экономической корреляции и исследовать приложения модели; привести статистическую оценку генеральной средней.

**Ход работы:**

Был получен вариант задания (рисунок 1).

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | 16 | 16 | 22 | 27 | 17 | 2 | 22 | 4 | 12 | 30 | 29 | 29 | 11 | 19 | 2 |
| $y_i$ | 360 | 530 | 710 | 790 | 590 | 620 | 740 | 680 | 380 | 590 | 170 | 240 | 230 | 970 | 450 |
| $x_i$ | 11 | 22 | 13 | 6 | 4 | 10 | 17 | 23 | 11 | 7 | 16 | 14 | 7 | 9 | 17 |
| $y_i$ | 130 | 740 | 420 | 250 | 250 | 360 | 180 | 240 | 430 | 200 | 910 | 600 | 310 | 140 | 290 |

Рисунок 1 – Вариант задания

Проранжируем по аргументу $x_i$ исходные данные (Таблица 1).

Таблица 1 – Исходные данные, проранжированные по аргументу x

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| xi | 2 | 2 | 4 | 4 | 6 | 7 | 7 | 9 | 10 | 11 | 11 | 11 | 12 | 13 | 14 |
| yi | 620 | 450 | 680 | 250 | 250 | 200 | 310 | 140 | 360 | 130 | 430 | 230 | 380 | 420 | 600 |
| xi | 16 | 16 | 16 | 17 | 17 | 17 | 19 | 22 | 22 | 22 | 23 | 27 | 29 | 29 | 30 |
| yi | 910 | 360 | 530 | 590 | 180 | 290 | 970 | 710 | 740 | 740 | 240 | 790 | 170 | 240 | 590 |

Количество интервалов m было определено по эмпирической формуле Стерджесса:

$$m = l + 3{,}32lg(n).$$

В рассматриваемом случае п = 30, lg(n) = 1,477, m = 5,9. Для удобства счета принимается m = 5. Вычислим длину интервала $l_x$ по формуле на рисунке 2:

$$l_x = \frac{x_{max} - x_{min}}{m}.$$

Рисунок 2 – Формула вычисления длины интервала

В данном случае размах выборки R = $x_{max} - x_{min}$ = 28, lx = 5,6. Поэтому границы интервалов будут такими:

1) первый интервал: 2…7,6; частота – 7;

2) второй интервал: 7,6…13,2; частота – 7;

3) третий интервал: 13,2…18,8; частота – 7;

4) четвёртый интервал: 18,8…24,4; частота – 5;

5) пятый интервал: 24,4…30; частота – 4.

Были вычеслены средние значения аргумента x и функции y для каждого из пяти интервалов путём деления суммы значений в интервале на их количество в интервале.

Также был расчитан процентный «штрафа Пигу» (ШП) по формуле, представленной на рисунке 3.

$$\overline{ШП}_j = \frac{\bar{y}_j}{\bar{x}_j} 100,\% .$$

Рисунок 3 – Формула «штрафа Пигу»

Была построена итоговая таблица 3 метода группировок.

Таблица 3 – Итоговая таблица метода группировок

| Интервалы по x | $\overline{x}_j$ млн. руб. | $\overline{y}_j$, тыс. руб. | ШП, % | $F_j$ |
|---|---|---|---|---|
| 2…7,6 | 4,571428571 | 394,2857143 | 8,625 | 7 |
| 7,6…13,2 | 11 | 298,5714286 | 2,714285714 | 7 |
| 13,2…18,8 | 16,14285714 | 494,2857143 | 3,061946903 | 7 |
| 18,8…24,4 | 21,6 | 680 | 3,148148148 | 5 |
| 24,4…30 | 28,75 | 447,5 | 1,556521739 | 4 |

Анализируя таблицу, можно сделать следующие выводы:

1) с ростом экономического оборота x экологический штраф y изменяется немонотонно (рисунок 4).
2) налог Пигу также изменяется немонотонно (рисунок 5).

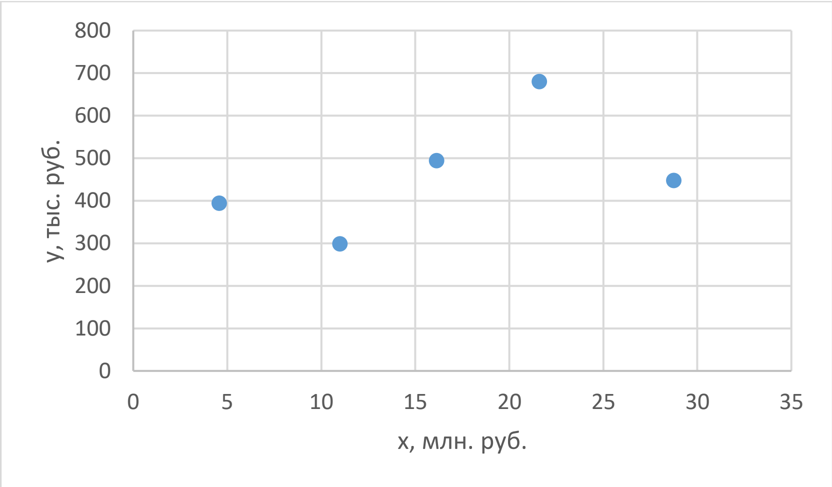


Рисунок 4 – Связь между экологическим штрафом и ежемесячным оборотом

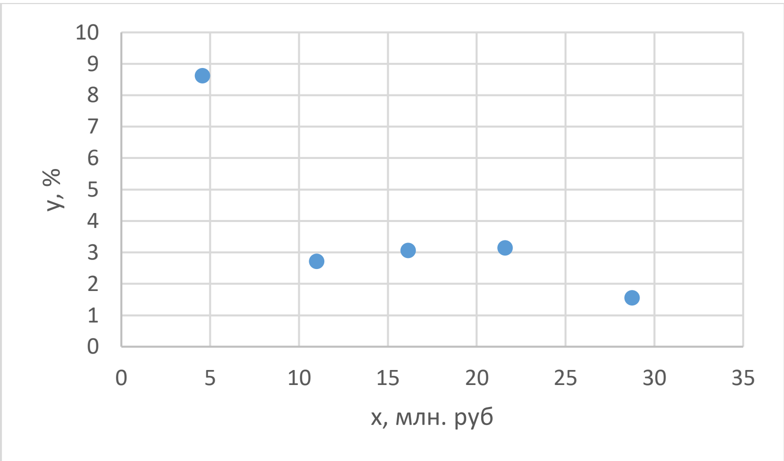


Рисунок 5 – Связь между зелёным «штрафом» (или налогом Пигу) и ежемесячным оборотом

Используя процедуру нахождения средней арифметической взвешенной величины (рисунок 6), найдём средние значения аргумента и функции.

$$\bar{x} = \frac{\sum x_j f_j}{n}$$

$$\bar{y} = \frac{\sum y_j f_j}{n}$$

Рисунок 6 – Процедура нахождения средней арифметической взвешенной величины

Были получены $\bar{x}$ = 14,83333 и $\bar{y}$ = 450. Средний налог Пигу находим по формуле, представленной на рисунке 7.

$$\overline{ШП} = 100\,\frac{\bar{y}}{\bar{x}}$$

Рисунок 7 – Средний налог Пигу

Получаем $\overline{\text{ШП}}$ = 3,033707865.

Далее было проведено выяснение взаимосвязи между аргументом x и функцией y с помощью прямолинейной зависимости по формуле на рисунке 8. Для этого также были определены коэффициенты $a_0$ и $a_1$ по формулам на рисунке 9. Была вычислена $s_{xy}$ – ковариация, учитывающая взаимовлияние функции и аргумента по формуле на рисунке 10, и $D_x$ – дисперсия по x, которая характеризует рассеивание случайной величины x вокруг своего среднего значения, по формуле на рисунке 11.

$$\tilde{y} = a_0 + a_1 x.$$

Рисунок 8 – Теоретическое значение экологического штрафа

$$a_1 = s_{xy} / \sigma_x^2; \quad a_0 = \bar{y} - a_1 \bar{x}$$

Рисунок 9 – Коэффициенты прямой линии

$$s_{xy} = \frac{\sum (x_j - \bar{x})(y_j - \bar{y}) f_j}{n}$$

Рисунок 10 – Ковариация, учитывающая взаимовлияние функции и аргумента

$$D_X = \frac{\sum (x_j - x)^2 f_j}{n}$$

Рисунок 11 – Дисперсия по x

Были получены следующие результаты:
$D_x$ = 61,85484; $S_{xy}$ = 537,1309524; $a_1$ = 8,683733421; $a_0$ = 321,1912876.

Найденная функциональная зависимость:

$$\tilde{y} = 321{,}19 + 8{,}68x$$

Были вычислены: коэффициент корреляции Пирсона по формуле, представленной на рисунке 12, дисперсия по функции y по формуле, показанной на рисунке 13, а также среднеквадратичные отклонения, которые являются квадратными корнями из дисперсий.

$$r_{xy} = \frac{s_{xy}}{\sigma_x \sigma_y}$$

Рисунок 12 – Коэффициент корреляции Пирсона

$$D_y = \frac{\sum (y_j - \bar{y})^2 f_j}{n}$$

Рисунок 13 – Дисперсия по функции y

Итого, искомые величины таковы:

$D_y$ = 15349,88095; $\sigma_y$ = 123,8946365; $\sigma_x$ = 7,864784884; $r_{xy}$ = 0,551240129.

Такое значение коэффициента корреляции Пирсона свидетельствует о среднем уровне корреляции.

По формуле, показанной на рисунке 14, была найдена средняя «эластичность» модели $\bar{Э}$ = 0,286241583. Её смысл таков: если аргумент x увеличить на 1%, то функция y в среднем изменится на 0,29%.

$$\bar{Э} = a_1 \frac{\bar{x}}{\bar{y}}$$

Рисунок 14 – Средняя «эластичность» модели

Далее с помощью формул Чебышева П.Л. и Ляпунова А.М. (рисунок 15), позволяющих определить интервалы изменения аргумента X и функции Y для «генерального» случая N = 1000 для требуемого уровня риска p, полученные результаты были перенесены на «генеральную совокупность»:

$$\bar{x} - t_p \frac{\sigma_x}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \le X \le \bar{x} + t_p \frac{\sigma_x}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$\bar{y} - t_p \frac{\sigma_y}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \le Y \le \bar{y} + t_p \frac{\sigma_y}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

Рисунок 15 – Формулы Чебышева П.Л. и Ляпунова А.М.

Если принять риск p = 5%, тогда $t_p$ = 2 (по таблице критических точек распределения Стьюдента), n = 30 изначально.

В итоге были получены следующие интервалы изменения аргумента X и функции Y:

12,00351 ≤ X ≤ 17,66316; 405,4215466 ≤ Y ≤ 494,5784534.

Пользуясь формулой, представленной на рисунке 7, были вычислены пределы изменения экологического штрафа Пигу:

2,800056976 ≤ ШП ≤ 3,377524992

Общие выводы:

1) с ростом экономического оборота x экологический штраф y изменяется немонотонно (рисунок 4);
2) налог Пигу также изменяется немонотонно (рисунок 5);
3) средний экономический налог составил 3,03%;
4) если аргумент x увеличить на 1%, то функция y в среднем изменится на 0,29%;
5) значение коэффициента корреляции, равное 0,55, свидетельствует о среднем уровне корреляции;
6) средний экологический штраф (штраф Пигу) для рассматриваемого региона на уровне риска 5% меняется в пределах 2,8…3,38.

**Вывод:**

В ходе лабораторной работы исходная информация была представлена в виде функциональной (или корреляционной) таблицы y = f(x). Была изучена взаимосвязь эколого-экономических явлений, также была построена математическая модель эколого-экономической корреляции и были исследованы приложения модели. Была приведена статистическая оценка генеральной средней.