

## ПРАКТИЧЕСКАЯ РАБОТА №10

**Тема:** Анализ связи между явлениями.

**Цель:** Научиться составлять уравнение регрессии, находить коэффициент корреляции, определять виды связи между явлениями.

**Оборудование:** вычислительная техника, чертежные инструменты.

### Теоретическое обоснование

Исследование объективно существующих связей между явлениями - важнейшая задача теории статистики. Социально-экономические явления представляют собой результат одновременного воздействия большого числа причин. При изучении этих явлений необходимо выявлять главные, основные причины, абстрагируясь от второстепенных.

Связи между признаками и явлениями, ввиду их большого разнообразия, классифицируются по ряду оснований. Признаки по их значению для изучения взаимосвязи делятся на два класса. Признаки, обуславливающие изменение других, связанных с ними признаков, называют **факторными**, или просто факторами. Признаки, изменяющиеся под действием факторных признаков, называют **результативными**. Связи между явлениями и их признаками классифицируются по степени тесноты, по направлению и по аналитическому выражению.

По степени тесноты связи различают количественные критерии оценки тесноты связи:

Значение $r$ – коэффициента корреляции	Связь
От 0 до $\pm 0,3$	Практически отсутствует Слабая Умеренная Сильная
От $\pm 0,3$ до $\pm 0,5$	
От $\pm 0,5$ до $\pm 0,7$	
От $\pm 0,7$ до $\pm 1$	

По направлению выделяют связь прямую и обратную. При прямой связи с увеличением или уменьшением значений факторного признака происходит увеличение или уменьшение значений результативного. Например, увеличение степени механизации труда способствует росту рентабельности строительного производства. В случае обратной связи значения результативного признака изменяются в противоположном направлении по сравнению с изменением факторного признака. Так, с увеличением уровня фондоотдачи снижается себестоимость единицы производимой продукции.

Значение $r(x,y)$	Связь	Интерпретация связи
$r = 0$	Отсутствует	Отсутствует линейная связь между величинами $x$ и $y$
$0 < r < 1$	Прямая	С увеличением $x$ величина $y$ в среднем увеличивается и наоборот
$-1 < r < 0$	Обратная	С увеличением $x$ величина $y$ в среднем уменьшается и наоборот
$r = +1$ $r = -1$	Функциональная	Каждому значению $x$ соответствует одно строго определенное значение величины $y$ и наоборот

Коэффициент корреляции рассчитывается по формуле:

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{S_x S_y},$$

$$\text{где } s_x = \sqrt{\overline{x^2} - (\bar{x})^2}, \quad s_y = \sqrt{\overline{y^2} - (\bar{y})^2}.$$

По аналитическому выражению выделяют связи прямолинейные (или просто линейные) и нелинейные (криволинейные). Если статистическая связь между явлениями приближенно выражена уравнением прямой линии, то ее называют линейной связью; если же она выражена уравнением какой-либо кривой линии (параболы, гиперболы: степенной, показательной, экспоненциальной и т. д.), то такую связь называют нелинейной или криволинейной.

Для выявления наличия связи, ее характера и направления в статистике используются методы: приведения параллельных данных, аналитических группировок, графических, корреляции и регрессии.

Метод приведения параллельных данных основан на сопоставлении двух или нескольких рядов статистических величин. Такое сопоставление позволяет установить наличие связи и получить представление о ее характере.

Статистическую связь между двумя признаками можно изобразить графически и по графику, который называется «Ломаная регрессии» судить о наличии, направлении и форме связи.

Парная регрессия характеризует связь между двумя признаками: результативным и факторным. Аналитически связь между ними описывается уравнениями:

$$\text{прямой } \bar{y}_x = a_0 + a_1x; \text{ параболы } \bar{y}_x = a_0 + a_1x + a_2x^2; \text{ гиперболы } \bar{y}_x = a_0 + a_1 \frac{1}{x} \text{ и т.д.}$$

Существуют более общие указания, позволяющие выявить уравнение связи аналитическим путем. Если результативный и факторный признаки возрастают одинаково, примерно в арифметической прогрессии, то это свидетельствует о наличии линейной связи между ними, а при обратной связи - гиперболической. Если результативный признак увеличивается в арифметической прогрессии, а факторный значительно быстрее, то используется параболическая или степенная функции.

**Оценка параметров уравнения регрессии**  $a_0, a_1, (a_2 - \text{в уравнении параболы второго порядка})$  осуществляется методом наименьших квадратов.

Система уравнений для нахождения параметров линейной парной регрессии методом наименьших квадратов имеет следующий вид:

$$\begin{cases} na_0 + a_1 \sum x = \sum y \\ a_0 \sum x + a_1 \sum x^2 = \sum xy \end{cases}$$

где  $n$  - объем исследуемой совокупности (число единиц наблюдения).

В уравнениях регрессии параметр  $a_0$  показывает усредненное влияние на результативный признак неучтенных (не выделенных для исследования) факторов; параметр  $a_1$  (а в уравнении параболы и  $a_2$ ) - коэффициент регрессии показывает, насколько изменяется в среднем значение результативного признака при изменении факторного на единицу его собственного измерения.

**Порядок выполнения**

- 1) Ознакомиться с теоретическим обоснованием.
- 2) Записать тему, цель, оборудование, задание в отчет.
- 3) Выполнить задание.
- 4) Сделать анализ полученных результатов.
- 5) Ответить на контрольные вопросы в отчете.
- 6) Сделать вывод по работе.

**Задание:** Используя различные статистические методы, установите направление и характер связи между двумя рядами статистических величин X и Y.

**Исходные данные:** В таблице приведены значения статистических величин X и Y.

<b>вариант</b>	<b>x</b>		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
	<b>0</b>	<b>y</b>	24	21	19	16	14	11	9	6	4
	<b>1</b>		5	6	9	10	14	17	15	20	23
	<b>2</b>		25	22	19	13	16	12	10	7	3
	<b>3</b>		3	7	10	12	16	13	19	22	25
	<b>4</b>		23	20	15	17	14	10	9	6	5
	<b>5</b>		6	8	10	13	16	20	18	21	24

**Ход работы:**

1. Выпишите данные своего варианта.
2. Постройте график взаимосвязи величин X и Y. По графику определить форму и направление связи величин X и Y.
3. Составьте расчетную таблицу вида:

	$x_i$	$y_i$	$x_i y_i$	$x_i^2$	$y_i^2$	$\bar{y}_x$
	1					
	2					
	3					
	4					
	5					
	6					
	7					
	8					
	9					
Сумма -	45					
Средняя -	5					

4. Найдите коэффициент корреляции и сделайте вывод о степени тесноты связи величин X и Y.
5. Найдите параметры уравнения регрессии методом наименьших квадратов.
6. Составьте уравнение регрессии и сделайте вывод об изменении результативного признака при изменении факторного на единицу его собственного измерения.

*Пример выполнения задания (рассмотрен вариант 0)*

X	1	2	3	4	5	6	7	8	9
Y	24	21	19	16	14	11	9	6	4

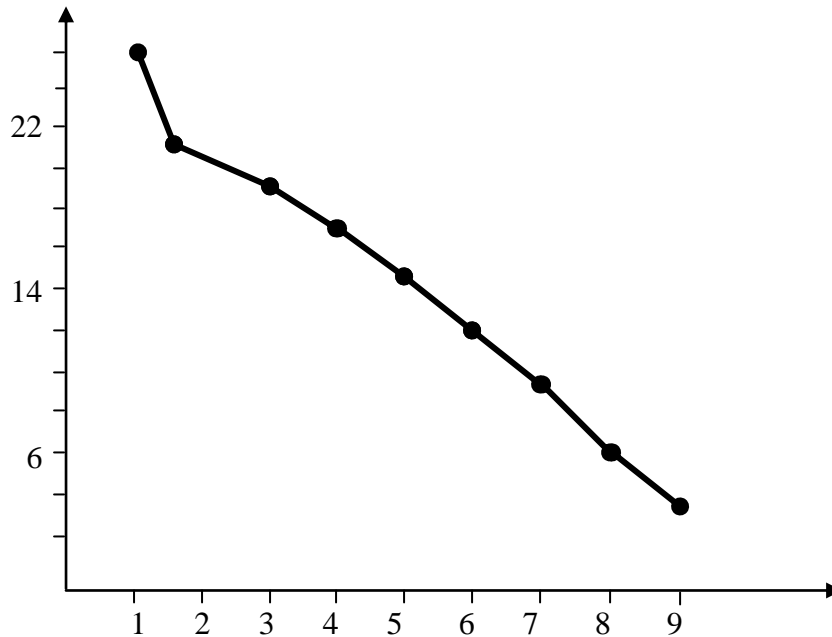


Рисунок 1 - График взаимосвязи величин X и Y

Используя для анализа графический, аналитический методы и метод приведения параллельных данных, можно сказать, что между величинами X и Y существует прямолинейная обратная взаимосвязь.

Составим расчетную таблицу:

	$x_i$	$y_i$	$x_i y_i$	$x_i^2$	$y_i^2$	$\bar{y}_x$
	1	24	24	1	576	24
	2	21	42	4	441	21
	3	19	57	9	361	19
	4	16	64	16	256	16
	5	14	70	25	196	14
	6	11	66	36	121	11
	7	9	63	49	81	9
	8	6	48	64	36	6
	9	4	36	81	16	4
Сумма -	45	124	70	285	2084	
Средняя -	5	13,8	52,2	37,7	231,6	

Рассчитаем коэффициент корреляции:

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{S_x S_y} = \frac{52,2 - 5 \times 13,8}{3,56 \times 6,87} = \frac{-16,8}{24,46} = -0,67$$

$$s_x = \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{37,7 - 25} = \sqrt{12,7} = 3,56$$

$$s_y = \sqrt{\overline{y^2} - (\bar{y})^2} = \sqrt{237,6 - 190,44} = \sqrt{47,16} = 6,87$$

Т.к. коэффициент корреляции меньше нуля, то связь между величинами X и Y обратная, т.е. с увеличением X величина Y в среднем уменьшается. По степени тесноты – связь умеренная.

Находим параметры уравнения регрессии, используя метод наименьших квадратов:

$$\begin{cases} na_0 + a_1 \sum x = \sum y \\ a_0 \sum x + a_1 \sum x^2 = \sum xy \end{cases} \quad \begin{cases} a_0 = \frac{124 - 45a_1}{9} = 13,8 - 5a_1 \\ 45(13,8 - 5a_1) + 285a_1 = 470 \\ 60a_1 = -151 \\ a_1 = -2,5 \\ a_0 = 13,8 - 5 \times (-2,5) = 26,3 \end{cases}$$

Т.к. связь между X и Y прямолинейная, то уравнение регрессии будет иметь вид:

$$\bar{y}_x = 26,3 - 2,5x$$

Т.е. при увеличении X на единицу, Y уменьшается в среднем на 2,5.

#### **Контрольные вопросы:**

- 1) Виды связей между явлениями.
- 2) Какие признаки анализируются при изучении взаимосвязи между явлениями?
- 3) В каких пределах изменяется коэффициент корреляции? Что он показывает?
- 4) Какие методы применяются для выявления наличия связи, ее характера и направления?
- 5) Что показывает коэффициент регрессии в уравнение регрессии?

#### **Литература:**

- 1) Мхитарян В.С. Статистика. М.: ИЦ «Академия», 2007
- 2) Практикум по теории статистики. Под ред. Шмойловой Р.А. М.: «Финансы и статистика», 1998

**Дисциплина:** Статистика

**Группа:** БУ-21

**Преподаватель:** Севостьянова Ю.С.

**e-mail:** seva.ju@yandex.ru

**Срок выполнения:** 27.05.2020 (выполнить в рукописном виде, отослать фото)