

## Тема 9. СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ СВЯЗИ МЕЖДУ ЯВЛЕНИЯМИ

1. Функциональные и статистические зависимости
2. Статистические методы выявления связи
3. Аналитическое выражение корреляционной зависимости
4. Оценка тесноты связи между явлениями

### 1. Функциональные и статистические зависимости

Формы проявления взаимосвязей наблюдаемых процессов и явлений классифицируются в статистике по ряду оснований.

По степени полноты выделяют функциональную (полную) и корреляционную (неполную) связи.

При функциональной связи каждому значению факторного признака  $X$  соответствует одно строго определенное значение результативного признака  $Y$  (например, прямо пропорциональная зависимость между производительностью труда и увеличением производства продукции).

Корреляционная связь между факторным признаком  $X$  и результативным  $Y$  проявляется не в каждом конкретном случае (строго функционально), а лишь в среднем по совокупности. При этом каждому значению фактора соответствует не одно значение результата, а распределение значений, варьирующих около средней величины (например, зависимость издержек обращения от объема товарооборота: помимо объема товарооборота ( $X$ ) на сумму издержек обращения ( $Y$ ) влияют и другие неучтенные факторы).

По направлению выделяют прямую и обратную связи.

Прямая – связь, при которой факторный и результативный признаки изменяются в одном и том же направлении – по мере увеличения или уменьшения факторного признака значения результативного соответственно увеличиваются или уменьшаются.

В случае обратной связи значения результативного признака изменяются под действием факторного, но в противоположном направлении по сравнению с изменением факторного признака (например, по мере снижения цены объем спроса, как правило, увеличивается).

По аналитическому выражению выделяют связи линейные и нелинейные. Статистическую связь называют линейной, если она может быть приближенно выражена математическим уравнением прямой линии. А если статистическая связь может быть выражена уравнением какой-либо кривой линии (параболы, гиперболы и т.д.), то ее называют нелинейной.

Если характеризовать связи с точки зрения количества взаимодействующих факторов, то связь двух признаков принято называть парной, связь более двух признаков – множественной.

ВОПРОС: Прямая зависимость между факторным и результативным признаками устанавливается, если:

имеется согласованность в изменении факторного и результативного признаков

с увеличением значений факторного признака увеличиваются значения результативного признака

с увеличением значений факторного признака уменьшаются значения результативного признака

## 2. Статистические методы выявления связи

В статистике используют следующие методы выявления взаимосвязей:

1. Метод сопоставления параллельных данных заключается в построении двух или нескольких рядов статистических величин, которые сравнивают между собой, что позволяет не только подтвердить связь, но и выявить ее направление.

2. Балансовый метод заключается в построении балансов-таблиц, в которых итог одной части равен итогу другой (например, баланс производства какого-либо продукта и его потребления).

3. Метод аналитических группировок (см. тему 2).

4. Графический метод предполагает построение корреляционного поля – графика, где по оси абсцисс откладываются значения  $X$ , а по оси ординат – значения  $Y$ . По расположению точек, их концентрации в определенном направлении можно судить о наличии или отсутствии связи, а также ее направлении.

5. Корреляционно-регрессионный анализ включает решение задач двух видов. Задачи собственно корреляционного анализа сводятся к измерению тесноты связи между варьирующими признаками. Задачи регрессионного анализа лежат в сфере установления формы зависимости, определения уравнения регрессии.

ВОПРОС: Задача корреляционного анализа – определение:  
формы зависимости между признаками  
среднего значения признака в совокупности  
степени рассеивания признака в совокупности  
тесноты связи между признаками

## 3. Аналитическое выражение корреляционной зависимости

Рассмотрим применение приемов корреляционного анализа на конкретном примере (табл. 1).

Таблица 1

Товарооборот и издержки обращения десяти предприятий (млн руб.)

Товарооборот	20	28	5	6	8	17	19	25	13	26
Издержки обращения	0,8	1,0	0,2	0,4	0,3	0,7	0,6	0,9	0,6	0,9

При рассмотрении вопросов подбора формы связи особое внимание уделяется глубокому теоретическому анализу изучаемого процесса, установлению причинно-следственных связей. Наиболее распространенным приемом выявления формы связи являются графические изображения.

Графический анализ исходных данных (рис. 1) показывает, что с увеличением товарооборота растут, как правило, и издержки обращения.

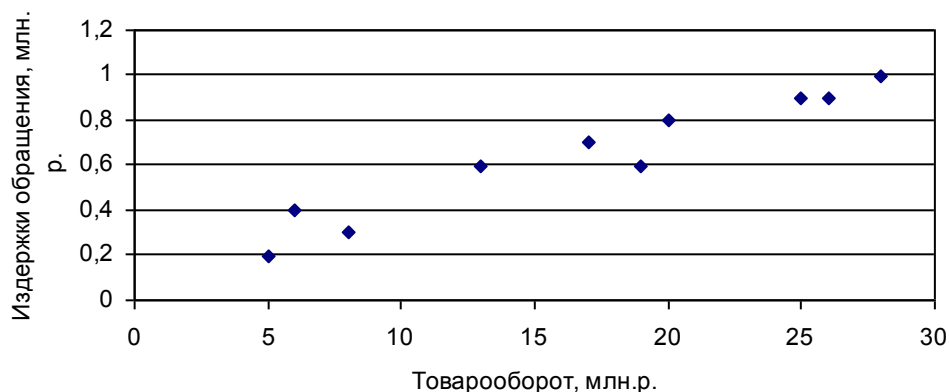


Рис. 1. Зависимость издержек обращения от товарооборота

После того как установлено, что зависимость между признаками есть, нужно установить теоретическую форму связи, т.е. вид математической функции  $f(x)$ , которая наилучшим образом описывает поведение изучаемого признака.

Допустим, что между рассматриваемыми показателями существует прямолинейная связь.

Уравнение линейной связи в общем виде можно записать так:

$$\bar{y}_x = a_0 + a_1 x.$$

Это уравнение, выражающее зависимость  $Y$  от  $X$ , называется уравнением регрессии.

Найти уравнение регрессии означает определить параметры  $a_1$  и  $a_0$ . Их оценивают при помощи метода наименьших квадратов, который дает следующую систему нормальных уравнений:

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

где  $x$  - значения факторного признака, в нашем примере суммы товарооборота (табл. 1, строка 1);

$y$  - значения результативного признака - суммы издержек обращения (табл. 1, строка 2);

$n$  - число парных значений факторного и результативного признаков.

Для составления и решения системы линейных уравнений, т.е. определения параметров воспользуемся данными табл. 2, в которой рассчитаем необходимые компоненты  $\sum x$ ,  $\sum y$ ,  $\sum x^2$ ,  $\sum xy$ , исходные данные предварительно ранжируем (располагаем по возрастанию значений факторного признака - товарооборота) и заполняем таблицу 2.

Таблица 2

Расчетная таблица параметров уравнения регрессии

Номер предприятия	Товарооборот, млн. руб. (x)	Издержки обращения, млн. руб. (y)	$x^2$	$xy$	$\bar{y}_x$	$y^2$
A	1	2	3	4	5	6
1	5	0,2	25	1,0	0,28	0,04

Окончание табл. 2						
A	1	2	3	4	5	6
2	6	0,4	36	2,4	0,31	0,16
3	8	0,3	64	2,4	0,37	0,09
4	13	0,6	169	7,8	0,52	0,36
5	17	0,7	289	11,9	0,65	0,49
6	19	0,6	361	11,4	0,71	0,36
7	20	0,8	400	16,0	0,74	0,64
8	25	0,9	625	22,5	0,90	0,81
9	26	0,9	676	23,4	0,93	0,81
10	28	1,0	784	28,0	0,99	1,00
Итого	167	6,4	3429	126,8	6,40	4,76

В гр.3 таблицы 2 вносим квадраты переменных  $x^2$  ( $5^2$ ,  $6^2$ ,  $8^2$  и т.д.), в гр.4 - произведение X на Y ( $5 \times 0,2$ ;  $6 \times 0,4$ ; и т.д.).

Итоговые показатели граф (1-4) подставляем в систему нормальных уравнений:

$$\begin{cases} na_0 + a_1 \sum x = \sum y & 10 a_0 + 167 a_1 = 6,4 & \left| \times 167 \right. \\ a_0 \sum x + a_1 \sum x^2 = \sum xy & 167 a_0 + 3249 a_1 = 126,8 & \left| \times 10 \right. \end{cases}$$

Каждый член первого уравнения умножаем на 167, второго - на 10 и из второго вычитаем первое:

$$\begin{array}{r} 1670 a_0 + 27889 a_1 = 1068,8 \\ 1670 a_0 + 32490 a_1 = 1268 \\ \hline 6404 a_1 = 199,2 \end{array}$$

Параметр  $a_1 = \frac{199,2}{6404} = 0,031$ . Подставим его значение в первое уравнение и найдем

параметр  $a_0$ :

$$\begin{aligned} 10a_0 + 0,031 \times 167 &= 6,4 \\ 10a_0 + 5,177 &= 6,4 \\ a_0 &= \frac{6,4 - 5,177}{10} = 0,12 \end{aligned}$$

Уравнение регрессии примет вид:  $\bar{y}_x = 0,12 + 0,031x$ . Подставляя в него значения  $x$ , найдем выравненные или теоретические значения  $\bar{y}_x$ . Так, при товарообороте 5 млн. руб. ( $x_1$ ) теоретическое значение суммы издержек составит:  $\bar{y}_{x_1} = 0,12 + 0,031 \times 5 = 0,28$ . При товарообороте 6 млн. руб.  $\bar{y}_{x_2} = 0,12 + 0,031 \times 6 = 0,31$  и т.д.

Теоретические значения помещены в таблице 2, гр.5. Сумма выравненных значений должна быть равна сумме фактических значений результативного признака ( $\sum \bar{y}_x \approx \sum y$ );  $6,4 = 6,4$ . Если такого равенства нет, то следует проверить правильность всех предшествующих расчетов.

Экономический смысл имеет параметр  $a_1$  - коэффициент регрессии, показывающий на сколько единиц в среднем изменится  $Y$  при увеличении  $X$  на единицу. В рассмотренном примере увеличение товарооборота на 1 млн руб. ведет в среднем к росту издержек обращения на 0,031 млн руб.

ВОПРОС: Коэффициент регрессии показывает:

как изменяется факторный признак с изменением результативного на единицу значение результативного показателя

тесноту связи

как изменяется результативный признак с изменением факторного на единицу

#### 4. Оценка тесноты связи между явлениями

В случае установления линейной зависимости между факторным ( $x$ ) и результативным ( $y$ ) признаками для оценки тесноты связи между ними рассчитывают линейный коэффициент корреляции:

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} = \frac{n \sum xy - \sum x \cdot \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2] \cdot [n \sum y^2 - (\sum y)^2]}}$$

Линейный коэффициент корреляции может принимать любые значения от -1 до +1. Чем ближе по модулю значение этого коэффициента к 1, тем более тесная связь предполагается между признаками  $X$  и  $Y$ . Если  $r_{xy} = 0$ , то это не всегда говорит об отсутствии связи вообще - часто это означает отсутствие линейной связи. В таком случае нужно использовать нелинейные зависимости (уравнение гиперболы, уравнение логарифмической кривой, экспоненциальную зависимость и др.).

Для качественной оценки тесноты связи между признаками используется шкала Чэддока (табл. 3).

Таблица 3

Оценка тесноты связи по шкале Чэддока

Показания тесноты связи	0,1 – 0,3	0,3 – 0,5	0,5 – 0,7	0,7 – 0,9	0,9 – 0,99
Характеристика тесноты связи	Слабая	Умеренная	Заметная	Высокая	Весьма высокая

Знак при линейном коэффициенте корреляции указывает направление связи: «+» - прямая связь; «-» - обратная связь.

Рассчитаем линейный коэффициент корреляции для рассмотренного примера, для этого воспользуемся данными итоговой строки таблицы 2:

$$\overline{xy} = \frac{\sum xy}{n} = \frac{126,8}{10} = 12,68;$$

$$\bar{x} = \frac{\sum x}{n} = \frac{167}{10} = 16,7;$$

$$\bar{y} = \frac{\sum y}{n} = \frac{6,4}{10} = 0,64.$$

Средние квадратические отклонения по признакам  $X$  и  $Y$  найдем по формулам:

$$\sigma_x = \sqrt{x^2 - (\bar{x})^2};$$

$$\sigma_y = \sqrt{y^2 - (\bar{y})^2}.$$

Среднюю величину из квадратов переменных  $X$  рассчитываем, подставив в формулы итоги гр.3,6 таблицы 2:

$$\overline{x^2} = \frac{\sum x^2}{n} = \frac{3429}{10} = 342,9; \quad \overline{y^2} = \frac{\sum y^2}{n} = \frac{4,76}{10} = 0,476 .$$

Следовательно, средние квадратические отклонения будут равны:

$$\sigma_x = \sqrt{342,9 - 16,7^2} = \sqrt{342,9 - 278,89} = 8;$$

$$\sigma_y = \sqrt{0,476 - 0,64^2} = \sqrt{0,476 - 0,4096} = 0,257 .$$

Линейный коэффициент корреляции составит:

$$r = \frac{12,68 - 16,7 \times 0,64}{8 \times 0,257} = \frac{12,68 - 10,686}{2,056} = \frac{1,992}{2,056} = 0,969 .$$

Согласно таблице Чэддока, при  $r = 0,969$  зависимость результативного признака от факторного очень высокая. Следовательно, найденное уравнение регрессии  $\bar{y}_x = 0,12 + 0,031x$  можно использовать для прогноза суммы издержек при наличии данных об изменении суммы товарооборота.

**ВОПРОС: Если коэффициент корреляции равен 0,78, связь между признаками:**

прямая тесная

прямая средняя

обратная сильная

обратная средняя