

Исследование
взаимосвязей
социально-
экономических
явлений

Методы изучения связи социальных явлений.

Важной задачей статистики является разработка методики статистической оценки социальных явлений, которая осложняется тем, что многие социальные явления не имеют количественной оценки. Количественная оценка связей социальных явлений осуществляется на основе расчета и анализа целого ряда коэффициентов.

Взаимосвязанные признаки:

а) факторные (под их воздействием изменяются другие признаки)

б) результативные

тесноты:

a)

функциональная

б)

статистическая

функциональная
связь — каждому значению
факторного признака
соответствует строго
определённое одно или
несколько значений
результативного признака

сбор зерна = средняя
урожайность * посевные
площади

Статистическая

связь – одному и тому же значению факторного признака может соответствовать несколько значений результативного признака. Проявляются только для большого числа единиц совокупности

Корреляционная
связь – соответствие одному
и тому же значению факторного
признака сколько угодно различных
значений результативного
признака. Связь прослеживается
лишь при достаточно большом
количестве наблюдений и лишь в
форме средней величины

Связь по направлению:

- **ПРЯМАЯ** — с увеличением или уменьшением значений факторного признака происходит увеличение или уменьшение значений результативного;
- **ОБРАТНАЯ** — значения результативного признака изменяются в противоположном направлении по сравнению с изменяющим факторного признака

По аналитическому выражению выделяют связь:

- **ЛИНЕЙНУЮ** – статистическая связь между явлениями приближенно выражена уравнением прямой линии;
- **НЕЛИНЕЙНУЮ** – статистическая связь выражена уравнением какой-либо кривой линии (параболы, гиперболы).

Основные
приемы
изучения
взаимосвязей

Метод сравнения параллельных рядов.

Параллельное(одновременное)приведение двух рядов данных, связь между которыми следует выявить.

Пример: выявление зависимости между объёмом производства и себестоимостью

Применение метода: ранжирование предприятий по объёму выпуска по возрастанию и прослеживание динамики себестоимости

Балансовый

метод

Изображение данных взаимосвязанных показателей в виде таблицы и их расположение следующего вида: итоги между отдельными её частями были равны (сбалансированы). Используется для характеристики взаимосвязи между производством и распределением продуктов, денежными доходами и расходами населения

Объёмы транспортных потоков между регионами (млн тонно-км)

Регионы отправ- ления	Регионы прибытия				Итого отправ- лено
	А	Б	В	Г	
А	25	75	80	100	280
Б	100	20	95	30	245
В	80	70	30	65	245
Г	70	45	55	35	205
Итого прибыло	275	210	260	230	975

Графический

метод - наглядное представление о наличии и направлении (прямая/обратная) взаимосвязей между признаками. Метод используется как самостоятельно, так и совместно с другими

Метод аналитической группировки

- 1. Группировка единиц совокупности по факторному признаку.
- 2. Расчет средней или относительной величины по результативному признаку для каждой группы.
- 3. Сопоставление рассчитанных изменений с изменениями факторного признака для выявления характера связи между ними

*Количество вкладчиков и средний
остаток вклада по филиалам
Сбербанка (тыс. руб.)*

<i>Число вкладчиков, человек</i>	<i>Средний остаток по вкладу, тыс. руб.</i>
<i>13 500</i>	<i>11,50</i>
<i>1 290</i>	<i>11,81</i>
<i>2 205</i>	<i>12,05</i>

Дисперсионный анализ

1. Определение вида признака-факторный/результативный
2. Группировка по факторному признаку
3. Расчет среднего значения факторного и результативного признака в группах
4. Выявление взаимосвязи между рассчитанными средними

Межгрупповая

дисперсия используется для
оценки тесноты связи по
результатам факторной
группировки

Сопоставление межгрупповой
дисперсии и общей дисперсии
дает характеристику тесноты
корреляционной связи между
признаками

Сопоставление межгрупповой дисперсии и общей дисперсии дает характеристику тесноты корреляционной связи между признаками –

Коэффициент
детерминации

$$\eta^2 \approx 0,5$$

- **тесная связь**

Корреляционное отношение $[0;1]$

Характеризует долю вариации результативного признака, вызванной действием факторного признака. Чем ближе корреляционное отношение к 1, тем большее влияние оказывает факторный признак на результативный

$\delta^2 = 0$ Если факторный
признак не влияет на
результативный
 $\eta = 0$

$\delta^2 = \sigma^2$ Если результативный
признак изменяется только
под воздействием одного
факторного признака

$\eta = 1$ Существование полной
связи

Уровень

значимости

- достаточно малое значение вероятности, отвечающее событиям, которые в данных условиях исследования будут считаться практически невозможными. Это указывает на ошибочность начального предположения

Число степеней
свободы факторной
дисперсии

$$k_1 = m - 1$$

m – число групп

Число степеней свободы
случайной дисперсии

$$k_2 = m - 1$$

m – число групп

n – число вариант

Критерий Фишера

- Проверка существенности связи.
Используется при распределении близком к нормальному.
- Отношение межгрупповой дисперсии к средней из внутригрупповых дисперсий, исчисленных с учетом числа степеней свободы

$$F = \frac{\delta^2}{n-1} : \frac{\overline{\sigma_i^2}}{n-m}$$

Корреляционно- регрессионный анализ

1. *Определение формы связи*
2. *Измерение тесноты связи*

1. Определение формы связи

- Нахождение уравнения регрессии
- Априорный теоретический анализ (с ростом факторного признака равномерно растет и результативный)
- Проверка априорного теоретического анализа с помощью графического анализа

2. Измерение тесноты связи

- Оценка и анализ полученных результатов при помощи показателей корреляционного анализа (коэффициенты детерминации, линейной и множественной корреляции)
- Проверка существования связи между изучаемыми признаками

Этапы корреляционного анализа

1. Предварительный анализ объекта исследования
2. Сбор и первичная обработка информации
3. Построение уравнения регрессии и определение его параметров
4. Проверка адекватности полученной модели

корреляционно- регрессионных связей

- Прямые/обратные
- Однофакторные (парная корреляция)/многофакторные
- Частичная связь
- Полное отсутствие связи

Уравнение регрессии (по аналитическому выражению)

- Прямойной*
- Криволинейной*

Прямолинейное уравнение

- *Величина явления изменяется приблизительно равномерно в соответствии с изменением величины влияющего фактора*

$$y = a_0 + a_1 x$$

- Линейное уравнение регрессии

a_1

Коэффициент регрессии. Показывает, на сколько в среднем отклоняется величина результативного признака Y при отклонении факторного признака X на одну единицу. При: $x = 0$

$$y = a_0$$

Криволинейное уравнение

- *Неравномерное изменение явления в связи с изменением величины влияющего фактора*

$$y = a_0 + a_1x + a_2x^2$$

- Уравнение параболы второго порядка

a_2 Параметр характеризует степень ускорения или замедления кривизны параболы и при :

$a_2 > 0$ Парабола имеет минимум

$a_2 < 0$ Парабола имеет максимум

a_1

Характеризует угол
наклона кривой

a_0

Характеризует начало
кривой

*Уравнение криволинейной
связи может быть выражено
и в виде
гиперболической функции*

$$y = a_0 + \frac{a_1}{x}$$

Парная корреляция

Аналитическое выражение связи

Аналитические методы -
основной способ изучения
связи. Различают
параметрические и
непараметрические методы

Корреляционный анализ

Между стоимостью основного капитала и выпуском продукции существует прямолинейная связь, выраженная уравнением прямой.

Найдите параметры a_0 и a_1

*Эта задача решается
методом наименьших
квадратов при помощи
системы нормальных
уравнений. Все расчеты
ведутся по данным
выборочного наблюдения*

Нахождение параметров
позволит определить
теоретические значения Y для
разных значений x_i . Причем a_0 и
 a_1 должны быть такими,
чтобы было достигнуто
максимальное приближение к
первоначальным значениям y
теоретических значений Y

ТЕСНОТЫ КОРРЕЛЯЦИОННОЙ

СВЯЗИ
ТЕСНОТА СВЯЗИ — степень связи между признаками при наличии корреляционной зависимости, когда средняя величина значений одного признака меняется в зависимости от изменения другого признака

Измерение тесноты связи

Для измерения тесноты
прямолинейной связи между
двумя признаками
используют линейный
коэффициент корреляции -

r

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\sigma_x^2 \cdot \sigma_y^2}}$$

umm

$$r_{xy} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \cdot \sum_i (y_i - \bar{y})^2}}$$

*Пределы изменения
парного коэффициента
корреляции*

$$-1 \leq r_{xy} \leq 1$$

Оценка линейного коэффициента корреляции

Значение линейного коэффициента связи	Характер связи	Интерпретация связи
$r = 0$	Отсутствует	–
$0 < r < 1$	Прямая	С увеличением X увеличивается Y
$-1 < r < 0$	Обратная	С увеличением X уменьшается Y, и наоборот
$r = 1$	Функциональная	Каждому значению факторного признака строго соответствует одно значение результативного признака

Шкала Чеддока

ТЕСНОТЫ СВЯЗИ.

- слабая – от 0,1 до 0,3;
- умеренная – от 0,3 до 0,5;
- заметная – от 0,5 до 0,7;
- высокая – от 0,7 до 0,9;
- весьма высокая (сильная) – от 0,9 до 1,0

Значимость линейного коэффициента корреляции проверяется на основе t -критерия Стьюдента:

$$t_p = \sqrt{\frac{r^2}{1-r^2} \cdot (n-2)} = \frac{|r|}{\sqrt{1-r^2}} \sqrt{n-2}$$

*При криволинейной и линейной
формах связи используется
индекс корреляции*

$$R = \sqrt{\frac{\sigma^2_T}{\sigma^2_{\text{общ}}}}$$

σ^2_T - Теоретическая
дисперсия

$\sigma^2_{общ}$ - Общая
дисперсия

*Пределы изменения
индекса корреляции по
своему абсолютному
значению*

$$0 \leq R \leq 1$$

*При функциональной
зависимости*

$$R=1$$

*При отсутствии
связи*

$$R=0$$

Коэффициент детерминации

- *Используется при любом количестве факторных признаков и при любой форме связи.*
- Характеризует роль факторной вариации в общей вариации и по посторению аналогичен корреляционному отношению.*
- Рассчитывается путем возведения в квадрат индекса корреляции*

$$R^2 = \sqrt{\frac{\sigma^2_T}{\sigma^2_{\text{общ}}}}$$

Множественн

ая

корреляция

*Множественная
корреляция
используется при
изучении, измерении
связи между
результативными
признаком, двумя и
более факторными*

Множественная корреляция определяет :

1. форму связи

2. тесноту связи

*3. влияние отдельных
факторов на общий результат*

1. Определение формы связи

- Сводится к нахождению уравнения связи y с факторами $x, z, w, \dots, v$

$$Y_{xz}^{z,w,\dots,v} = a_0 + a_1x + a_2z$$

- Линейное уравнение зависимости результативного от двух факторов x и z

- Для определения параметров a_0 , a_1 и a_2 по способу наименьших квадратов, необходимо решить следующую систему трех нормальных уравнений

$$\begin{cases} \sum y = na_0 + a_1 \sum x + a_2 \sum z \\ \sum yz = a_0 \sum x + a_1 \sum xz + a_2 \sum z^2 \\ \sum yx = a_0 \sum x + a_1 \sum x^2 + a_2 \sum xz \end{cases}$$

2. Измерение тесноты связи

- Производится на основе вариации результативного признака и правила сложения дисперсий:

$$\sigma_{\text{общ}}^2 = \sigma_T^2 + \sigma_{\text{ост}}^2$$

*Теоретическая дисперсия-
вариация теоретического
признака вокруг общей средней*

$$\sigma^2_T = \frac{\sum_{i=1}^n (Y_i^T - \bar{y})^2}{n}$$

Остаточная дисперсия -
среднее квадратическое
отклонение теоретического
признака от фактического

$$\sigma^2_{ост} = \frac{\sum_{i=1}^n (y_i - Y_i^T)}{n}$$

3. Расчет коэффициента множественной корреляции

Рассчитывается для определения тесноты связи результативного признака от двух факторных

$$R_{y,x,z} = \frac{r_{yx}^2 + r_{yz}^2 - 2r_{yx} \cdot r_{yz} \cdot r_{xz}}{1 - r_{xz}^2}$$

r_{xy} r_{xz} r_{zy} - парные коэффициенты корреляции

*Пределы изменения
коэффициента
множественной
корреляции*

$$0 \leq R_{y,x,z} \leq 1$$

*Если коэффициент
множественной корреляции
возвести в квадрат , то
получим совокупный
коэффициент детерминации,
который характеризует долю
вариации результативного
признака Y под воздействием
всех изучаемых факторных
признаков*

Частные коэффициенты корреляции

- Рассчитываются для определения тесноты связи между результативным признаком и одним из факторных при постоянных значениях прочих факторов

$$r_{xy(z)} = \frac{r_{xy} - r_{zy} \cdot r_{xz}}{(1 - r_{zy}^2)(1 - r_{xz}^2)}$$

Непараметрические показатели связи.

Ранговые коэффициенты.

В анализе социально-экономических явлений часто приходится прибегать к различным условным оценкам с помощью рангов, а взаимосвязь между отдельными признаками измерять с помощью непараметрических коэффициентов связи.

Ранжирование

– это процедура упорядочения объектов изучения, которая выполняется на основе предпочтения.

Ранг

– это порядковый номер значений признака, расположенных в порядке возрастания или убывания их величин.

Среди непараметрических методов
оценки тесноты связи наибольшее
значение имеют ранговые

коэффициенты Спирмена и Кендалла.

Эти коэффициенты могут быть
использованы для определения тесноты
связи как между количественными, так и
между качественными признаками при
условии, если их значения будут
упорядочены или проранжированы по
степени убывания или возрастания

Учебное

коэффициенты связи

Спирмена

Коэффициент ранговой корреляции Спирмена – это непараметрический метод, который используется с целью статистического изучения связи между явлениями

ранговой корреляции Спирмена

$$r = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

$\sum d^2$ - сумма квадратов разностей рангов,

n - число парных наблюдений

У инв коэффициент связи Кендалла -

коэффициент корреляции, определяющий степень соответствия упорядочения всех пар объектов по двум переменным:

$$\tau = \frac{2S}{n(n-1)}$$

n - число наблюдений, S - сумма разностей между числом последовательностей и числом инверсий по второму признаку

коэффициент ассоциации и КОНТИНГЕНЦИИ.

Применяются для определения тесноты связи двух качественных признаков, каждый из которых состоит только из двух групп. Для их вычисления строится таблица, которая показывает связь между двумя явлениями, каждое из которых должно быть альтернативным, т.е. состоящим из двух качественно отличных от друг друга значений признака.

*Таблица для вычисления
коэффициентов ассоциации
и контингенции*

a	b	a + b
c	d	c + d
a + c	b + d	a + b + c + d

Пример 1

По приведенным условным данным по отдельным отраслям промышленности в регионе рассчитать ранговые коэффициенты связи Спирмена и Кендалла

Отрасль промышленности	Уровень производительности труда, млн руб./чел.	Среднемесячная номинальная заработная плата, тыс. руб.
Электроэнергетика	1,127	10,96
Топливная	2,630	19,35
Черная металлургия	1,632	9,35
Цветная металлургия	1,155	13,45
Машиностроение	0,550	6,68

Решение

Число пар значений $n=5$. Ранжируем значения факторного и результативного признаков в порядке возрастания количественных величин (графы 4 и 5). Находим разности рангов $d_i = R_x - R_y$ (графа 6). Полученные разности рангов d_i возводим в квадрат, находим их сумму (графа 7)

Отрасль промышленности	X	Y	R_x	R_y	d_i	d_i^2
1	2	3	4	5	6	7
Электроэнергетика	1,127	10,96	2	3	-1	1
Топливная	2,630	19,35	5	5	0	0
Черная металлургия	1,632	9,35	4	2	2	4
Цветная металлургия	1,155	13,45	3	4	-1	1
Машиностроение	0,550	6,68	1	1	0	0
Итого	-	-	-	-	-	6

*Подставляем значения в формулу
коэффициента Спирмена*

$$\rho = 1 - \frac{6 \cdot \sum d_i^2}{n \cdot (n^2 - 1)} = 1 - \frac{6 \cdot 6}{5 \cdot (5^2 - 1)} = 1 - \frac{36}{5 \cdot 24} = 0,7$$

*При вычислении коэффициента
Кендалла значения факторного
признака предварительно ранжируем.
Значения результативного признака
записываем в соответствии с
исходными данными*

Отрасль промышленности	X	Y	R _x	R _y	P	Q
1	2	3	4	5	6	7
Машиностроение	0,550	6,68	1	1	4	0
Электроэнергетика	1,127	10,96	2	3	2	1
Цветная металлургия	1,155	13,45	3	4	1	1
Черная металлургия	1,632	9,35	4	2	1	0
Топливная	2,630	19,35	5	5	0	0
Итого					8	-2

Для каждого R_y определяем:

- ❖ число следующих за ним рангов, больших по значению, чем данный ранг. Общее число таких случаев учитывают со знаком «+» и обозначают буквой P (графа 6);
- ❖ число следующих за ним рангов, меньших по значению, чем данный ранг. Общее число таких случаев учитывают со знаком «-» и обозначают буквой Q (графа 7).

Вычисляем $S = P + Q = 8 + (-2) = 6$

*Подставим в формулу коэффициента
Кендалла полученные значения*

$$\tau = \frac{2 \cdot S}{n \cdot (n - 1)} = \frac{2 \cdot 6}{5 \cdot (5 - 1)} = 0,6$$

*Величины коэффициентов Спирмена и
Кендалла свидетельствуют о тесной
зависимости среднемесячной
заработной платы от уровня
производительности труда в
представленных отраслях экономики*

Пример 2

По приведенным ниже условным данным определить степень тесноты связи между успеваемостью студентов по математике и посещением занятий по этой же дисциплине

Группы студентов	Численность студентов - всего, чел.	Из них	
		Успешно сдали экзамен	Не сдали экзамен
Посещающие занятия	19	16	3
Не посещающие занятия	7	2	5

Решение

*Расчитаем коэффициенты ассоциации
и контингенции*

$$k_a = \frac{a \cdot d - b \cdot c}{a \cdot d + b \cdot c} = \frac{16 \cdot 5 - 3 \cdot 2}{16 \cdot 5 + 3 \cdot 2} = \frac{80 - 6}{80 + 6} = 0,86$$

$$\begin{aligned} k_k &= \frac{a \cdot d - b \cdot c}{\sqrt{(a + b) \cdot (b + d) \cdot (d + c) \cdot (c + a)}} = \\ &= \frac{16 \cdot 5 - 3 \cdot 2}{\sqrt{(16 + 3) \cdot (3 + 5) \cdot (5 + 2) \cdot (2 + 16)}} = 0,53 \end{aligned}$$

*Значения полученных
коэффициентов
свидетельствуют о тесной
связи между успешной
сдачей экзамена по
математике студентом и
посещением занятий по
этой же дисциплине, т.к.
 $k_a > 0,5$ или $k_k > 0,3$*

- Спасибо за
внимание