Тема 8: СТАТИСТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА СВЯЗЕЙ

1. Актуальность изучения взаимосвязей экономических явлений
2. Виды связей между признаками явлений
3. Парная линейная и нелинейная связи.
4. Множественная линейная и нелинейная связи.

1. Виды связей между признаками явлений

В статистике различают:

функциональную	**стохастическую**
Функциональной называют такую связь, при которой определенному значению факторного признака соответствует одно и только одно значение результативного признака. Функциональные связи между признаками изучаются в экономике посредством **индексного метода.**	**При *стохастической связи*** каждому отдельному значению факторного признака ***x*** отвечает определенное **множество** значений результативного признака ***y.***

парную	**множественную**
Изучение влияния одного факторного признака ***x*** на результирующий признак ***y***.	Изучение влияния нескольких факторных признаков ***x*** на результирующий признак ***y***.
прямая	**обратная**
с увеличением ***x*** увеличивается ***y***.	с увеличением ***x*** уменьшается ***y***.
линейная	**нелинейная**
значения признаков ***в первой степени***	значения признаков ***в любой степени***

2.Парная линейная и нелинейная связи.

- Частным случаем статистической связи является **корреляционная** связь.

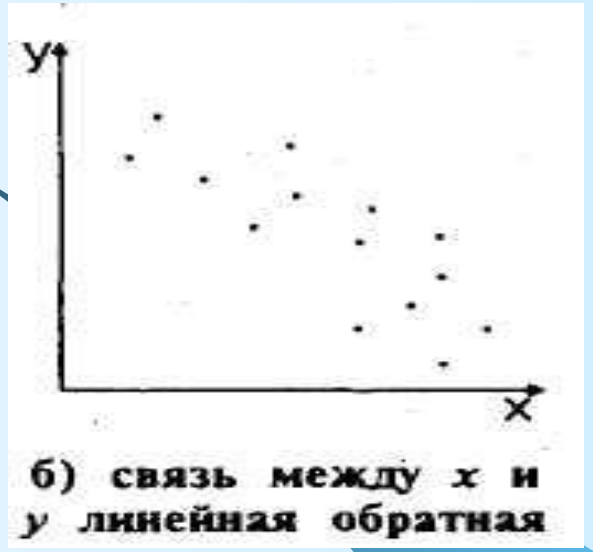
Корреляционная связь между признаками ***x*** и ***y (это связь в среднем: заданному значению x ставится в соответствие среднее значение y)*** записывается в виде *уравнения корреляционной связи,* или *уравнения регрессии:*

$$Y=f(x),$$

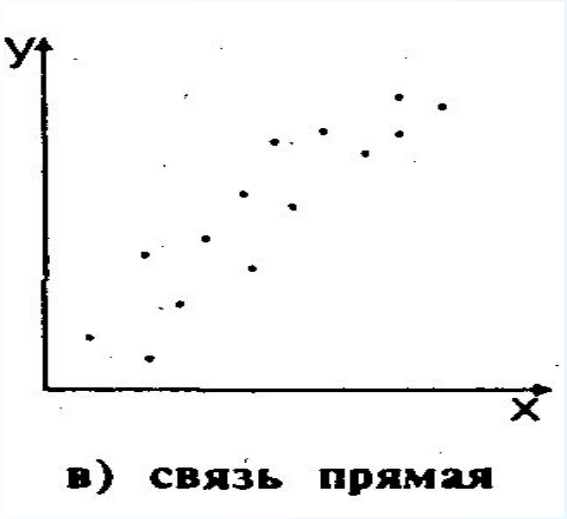
где f(x) — определенный вид функции корреляционной связи, которая описывает *линию регрессии.*

Графическое представление связи

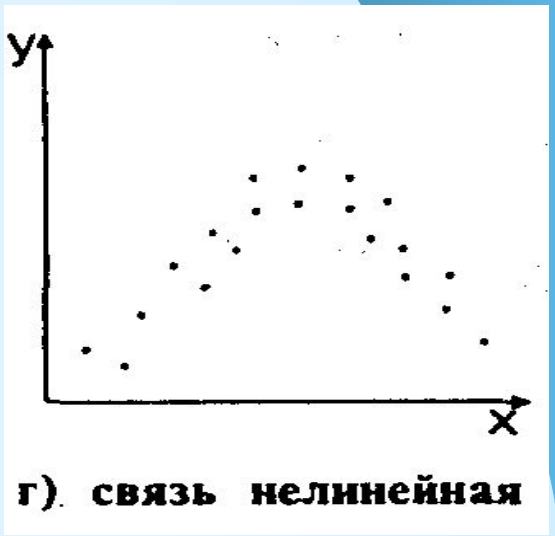
а) связь между x и y отсутствует



б) связь между x и y линейная обратная



в) связь прямая



г) связь нелинейная

Парная регрессия

Наиболее часто для характеристики корреляционной связи между признаками применяют такие *виды уравнений парной регрессии,* или корреляционных уравнений:

а) линейный $\bar{y}_x = a_0 + a_1 x$ (8.2)

б) параболический $\bar{y}_x = a_0 + a_1 x + a_2 x^2$ (8.3)

в) гиперболический $\bar{y}_x = a_0 + \frac{a_1}{x}$ (8.4)

г) степенной $\bar{y}_x = a_0 x^{a_1}$ (8.5)

и др.

- где a_0, a_1 — **параметры** уравнений регрессии, которые подлежат определению и находятся методом наименьших квадратов(МНК).

В случае линейной связи ее теснота измеряется с помощью **коэффициента парной корреляции и детерминации**:

$$r = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}}$$

$\mathbf{r^2}$ - коэффициент детерминации. Он показывает меру качества уравнения регрессии: чем ближе r^2 к 1, тем лучше регрессия описывает зависимость между x_i и y. Коэффициент детерминации может быть выражен в процентах.

Количественные критерии оценки тесноты связи

Величина коэффициента корреляции	Сила связи
До ±0,3	практически отсутствует
±0,3 – ±0,5	слабая
±0,5 – ±0,7	умеренная
+0,7 – ±1,0	сильная

Оценка линейного коэффициента корреляции

Значение линейного коэффициента связи	Характеристика 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

Матрица коэффициентов парной корреляции (общий вид)

Признаки	y	x_1	x_2	x_3	...	x_k
Y	1	r_{yx_1}	r_{yx_2}	r_{yx_3}	...	r_{yx_k}
X_1		1	$r_{x_1x_2}$	$r_{x_1x_3}$	...	$r_{x_1x_k}$
X_2			1	$r_{x_2x_3}$	...	$r_{x_2x_k}$
X_3				1	...	$r_{x_3x_k}$
.					...	
.					...	
.					...	
X_k					...	1

4 .Множественная линейная и нелинейная связи.

Если на результативный фактор влияет не один, а несколько факторов, то применяют

(не парную), а множественную регрессию.

Эта связь может быть выражена ***линейными и нелинейными функциями.***

Наиболее часто используемой является линейная функция – уравнение множественной линейной регрессии в виде:

$$\tilde{y}_{1,2....k} = a_0 + a_1 x + a_2 x_2 + ... + a_k x_k$$

где a_0,… a_k — **параметры** уравнений регрессии (находятся с помощью МНК). Они показывают, на сколько изменится ***y*** при изменении x_i ***на 1 единицу и при неизменных остальных факторах.***

Виды уравнений множественной регрессии:

1) линейная:

$$\tilde{y}_{1,2....k} = a_0 + a_1 x + a_2 x_2 + ... + a_k x_k$$

2) степенная:

$$\tilde{y}_{1,2....,k} = a_0 x_1^{a_1} * x_2^{a_2} * ... * x_k^{a_k}$$

3) показательная:

$$\tilde{y}_{1,2....k} = e^{a_0 + a_1 x + a_2 x_2 + ... + a_k x_k}$$

4) параболическая:

$$\tilde{y}_{1,2....k} = a_0 + a_1 x_1^2 + a_2 x_2^2 + ... + a_k x_k^2$$

5) гиперболическая:

$$\tilde{y}_{1,2....k} = a_0 + \frac{a_1}{x_1} + \frac{a_2}{x_2} + ... + \frac{a_k}{x_k}$$

Множественный коэффициент корреляции

- Теснота связи y со всей совокупностью факторов x_i определяется с помощью множественного коэффициента корреляции **R**

$$R = \sqrt{\frac{\sum (\tilde{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}} = \sqrt{\frac{\sigma^2_{\tilde{y}}}{\sigma^2_y}}$$

- Множественный коэффициент корреляции изменяется в пределах от 0 до 1 и по определению положителен: $0 \leq R \leq 1$.

В частном случае двухфакторной линейной регрессии можно использовать формулу(выраженную через парные коэффициенты корреляции:

$$R_{y/x_1x_2} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2}},$$

Матрица коэффициентов парной корреляции (общий вид)

Признаки	y	x_1	x_2	x_3	...	x_k
Y	1	r_{yx_1}	r_{yx_2}	r_{yx_3}	...	r_{yx_k}
X_1		1	$r_{x_1x_2}$	$r_{x_1x_3}$	...	$r_{x_1x_k}$
X_2			1	$r_{x_2x_3}$	...	$r_{x_2x_k}$
X_3				1	...	$r_{x_3x_k}$
.					...	
.					...	
.					...	
X_k					...	1

Коэффициент множественной детерминации показывает, в какой мере вариация результативного признака ***y*** определяется вариацией факторного признака ***x***.

Коэффициент детерминации принимает значение от 0 до 1.

$$R^2 = \frac{\sum(\tilde{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})} = \frac{\sigma^2_{\tilde{y}}}{\sigma^2_y}$$

5. Оценка и проверка качества модели
А). для парной связи

После установления тесноты связи дают оценку ***значимости связи*** между признаками.

Под термином ***«значимость связи»*** понимают оценку отклонения выборочных переменных от своих значений в генеральной совокупности посредством статистических критериев.

Оценку значимости связи осуществляют с использованием F-*критерия Фишера* и t-*критерия Стьюдента*.

Для парной регрессии (линейной и нелинейной) F-критерий Фишера рассчитывается по формуле:

$$F = \frac{\sum (Y - \bar{y})^2}{1} : \frac{\sum (y - Y)^2}{(n-2)},$$

где [1, n-2] – число степеней свободы числителя и знаменателя формулы.

Под термином ***«степень свободы»*** понимают целое число, которое показывает, сколько независимых элементов информации в переменных ***y*** нужно для суммы их квадратов, что объясняет соответствующую дисперсию: общую, межгрупповую, среднюю из групповых .

Для множественной регрессии степени свободы равны:

$$(k\ ;\ n-k-1)$$

Теоретическое значение (рассчитанное по формуле) F сравнивают с табличным (критическим) значением F*табл.*

Последнее выбирают из справочных математических таблиц F-*критерия Фишера* в зависимости от степеней свободы 1, *(n - 2)* и принятого уровня значимости ά(альфа). *(0,05 -5% вероятность допустимой ошибки)*

Если F > *Fтабл,* то **связь** между признаками признается **значимой.**

Для проверки значимости коэффициентов уравнения множественной регрессии a_i $(i=1,..,k)$ используют Критерий Стьюдента:

$$t_i = \frac{|a_i|}{\sqrt{\sigma^2_{a_i}}}$$

Коэффициенты уравнения (модели) **признаются статистически значимыми**, если $|t_i| > t(\acute{\alpha}; n-k-1)$.

Где: t (ά; n-k-1) - табличное значение**.**

ά - уровень значимости

n-k-1 - число степеней свободы, которое характеризует число свободно варьирующих элементов совокупности.

n – число наблюдений

k – число факторных признаков.

6. Изучение связи между качественными признаками

- Пример: Обработать данные социологического опроса работников предприятия.

Y \ X	Мужчины	Женщины	Итого
Имеют в/о	4	5	4+5
Без в/о	8	10	8+10
Итого	4+8	5+10	4+5+8+10

- где 4, 5,8,10 -частоты

Вычисление коэффициентов ассоциации и контингенции

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

Коэффициенты вычисляются по формулам:

ассоциации

$$Ka = \frac{ad - bc}{ad + bc}$$

и контингенции

$$Kк = \frac{ad - bc}{\sqrt{(a+b)\cdot(b+d)\cdot(a+c)\cdot(c+d)}}$$

Коэффициент контингенции всегда меньше коэффициента ассоциации.

Когда каждый из качественных признаков состоит более чем из двух групп, то для определения тесноты связи возможно применение **коэффициентов взаимной сопряженности Пирсона-Чупрова.** Эти коэффициенты вычисляются по следующим формулам:

$$K_n = \sqrt{\frac{\varphi^2}{1+\varphi^2}}; \quad K_ч = \sqrt{\frac{\varphi^2}{\sqrt{(K_1-1)\cdot(K_2-1)}}}.$$

где φ2 — показатель взаимной сопряженности;

φ — определяется как сумма отношений квадратов частот каждой клетки таблицы к произведению итоговых частот, соответствующего столбца и строки. Вычитая из этой суммы «1», получим величину φ 2:

$$\varphi^2 = \sum \frac{n_{xy}^2}{n_x n_y} - 1;$$

К1 - число значений (групп) первого признака;

К2 - число значений (групп) второго признака.

Чем ближе величина *Кп* и *Кч* к 1, тем теснее связь.

Ранговые коэффициенты связи

Среди непараметрических методов оценки тесноты связи ранжированных признаков наибольшее значение имеют ранговые коэффициенты Спирмена (ρ_{xy}) и Кендалла (τ_{xy}).

Эти коэффициенты могут быть использованы для определения тесноты связи как между количественными, так и между качественными признаками.

Коэффициент корреляции рангов (коэффициент Спирмена) рассчитывается по формуле

$$\rho_{xy} = 1 - \frac{6\sum d_i^2}{n(n^2-1)},$$

где d_i^2 $(R_{xj} - R_{yj})$- квадраты разности рангов;

n — количество единиц в ряду.

Коэффициент Спирмена принимает любые значения в интервале -1; 1.

Если d_i=0 p=1 –существует тесная прямая связь. Если первому рангу по размеру одного признака соответствует последний ранг по размеру второго признака, второму рангу – предпоследний ранг второго признака и т.п., то p = -1, и существует **тесная обратная связь. Если значение p близко к 0, то связь слабая или ее вообще нет.**

- **Алгоритм проведения корреляционно-регрессионного анализа.**

- отбор наиболее существенных данных для включения в корреляционно-регрессионные модели, дифференциация их на объясняющие и результативные признаки;
- выявление причин возникновения взаимосвязей между признаками, предварительный расчёт и анализ парных коэффициентов корреляции, построение матрицы коэффициентов множественной корреляции и оценка возможных вариантов группировки признаков для построения регрессионной модели;
- решение уравнения регрессии – вычисление коэффициентов уравнения регрессии и их смысловая интерпретация;
- статическая оценка достоверности параметров уравнения и общая оценка качества модели;
- практические выводы из анализа, применение результатов анализа для совершенствования планирования и управления экономическим процессом.