СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ВЗАИМОСВЯЗЕЙ

корреляционный и регрессионный анализ

Производственная функция

– экономико-статистическая модель связи, характеризующая изменение уровня результативных производственных показателей в зависимости от одного или ряда важнейших производственных факторов

Способы построения производственных функций

- аналитический – построение математического уравнения, моделирующего зависимость результативного экономического показателя от одного или ряда производственных факторов
- табличный - представление результативного показателя, соответствующего определенным значениям факторов, в виде таблицы
- графический - представление зависимости исследуемого показателя от фактора в виде графика

Классификация производственных функций

- по степени влияния человека на исследуемый результативный показатель
  - объективные
  - субъективные
  - объективно-субъективные
- по признаку сложности
  - простые - немногофакторные, элементарные зависимости
  - сложные - зависимости от целого ряда факторов
- по степени полноты учета факторных признаков
  - закрытые - простые детерминированные зависимости
  - открытые - сложные стохастические производственные функции

Классификация производственных функций

- по числу факторов, учтенных в модели
  - однофакторные
  - Многофакторные
- по виду математической модели
  - линейные
  - криволинейные
- по направлению влияния факторных признаков на зависимый показатель
  - прямые
  - обратные
  - комбинированные

Классификация производственных функций

- по виду ряда данных
    - Вариационные
    - Динамические
    - Вариационно-динамические
- по полноте учета информации
    - Выборочные
    - Генеральные
- по временному фактору
    - Однопериодные
    - Многопериодные
- по уровню управления
    - Межотраслевые
    - Отраслевые
    - Региональные
    - Межхозяйственные
    - Хозяйственные

Виды производственных функций

- Определяется видом уравнения, которое используется в качестве ее математической модели
- Одно и то же математическое уравнение как математическая модель может использоваться для построения нескольких (различных) зависимостей.
- Одна и та же производственная связь может имитироваться разными математическими уравнениями.
- Апробированные математические модели с изменением места и времени часто оказываются практически неприемлемыми.
- Идеальной производственной функцией следует считать ту, которая наиболее точно воспроизводит исследуемое явление или процесс. Но построение таких идеальных статистико-экономических моделей возможно только в простейших случаях.

Виды производственных функций

- однофакторная статистико-экономическая зависимость - линейная функция - прямая пропорциональная зависимость. Ее графиком является прямая, которая проходит через начало координат. Число (а) называется угловым коэффициентом прямой. С помощью этой производственной функции моделируют зависимость, например, стоимости продукции (у) от ее количества или цены ее единицы

$$y = ax$$

Виды производственных функций

- Линейная производственная функция моделирует зависимость, например, уровня оплаты труда (у) от его производительности (х).

$$y = a_0 + a_1 x$$

Виды производственных функций

- Парабола второго порядка - производственная функция, которую целесообразно использовать для моделирования зависимостей, имеющих одну экстремальную точку (минимума или максимума). Такой является, например, зависимость урожая культуры (y) от внесения удобрений.

$$y = a_0 + a_1 x + a_2 x^2$$

Виды производственных функций

- Целая рациональная функция, которую используют тогда, когда исследуемая зависимость содержит ряд экстремумов.

$$y = a_0 + a_1 x + a_2 x^2 + ... + a_n x^n$$

- Гипербола, сдвинутая по оси ординат на a_0 – которую используют для моделирования обратных пропорциональных зависимостей, например, издержек на единицу продукции (у) от производительности оборудования (х).

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

Виды производственных функций

$$y = a_0 + \frac{a_1}{x} + \frac{a_2}{x^2}$$

$$y = a_0 + \frac{a_1}{x} + \frac{a_2}{x^2} + \ldots + \frac{a_n}{x^n}$$

Виды производственных функций

- степенная функция

$$y = a_0 x^{a_1}$$

- показательная производственная функция – для анализа рядов динамики и уровня важнейших экономических параметров предприятия

$$y = a_0 a_1^{\,x}$$

Виды производственных функций

- Для моделирования периодических, сезонных колебаний, волнообразных процессов применяют различные тригонометрические уравнения. Простейшими из них являются уравнения синусоиды.

$$y = \sin x$$

$$y = a \sin x$$

$$y = a_0 + a_1 \sin x$$

Виды производственных функций

- многофакторная линейная функция

$$y = a_0 + a_1 x + a_2 x + \ldots + a_n x$$

- обратная многофакторной функции

$$y = a_0 + \frac{a_1}{x} + \frac{a_2}{x^2} + \ldots + \frac{a_n}{x^n}$$

Виды производственных функций

- многофакторная парабола второго порядка

$$y = a_0 + a_1 x_1 + a_2 x_1^2 + a_3 x_2 + a_4 x_2^2 + a_5 x_1 x_2$$

Направления использования производственных функций:

1. *Определение влияния различных факторов на анализируемые результативные показатели.*
2. *Поиск оптимального сочетания факторов, при котором зависимый показатель достигает экстремального (максимального или минимального) уровня.*
3. *Анализ, хозяйственных процессов предприятий, их подразделений и объединений, происходящие под воздействием как объективных, так и субъективных факторов, и результатов их деятельности, характеризующихся определенной системой показателей.*
4. *Прогнозирование и планирование уровня важнейших показателей производства.*

Направления использования производственных функций:

5. *Обработка информации. Экономической информацией называют информацию об общественных процессах производства, распределения, обмена и потребления материальных благ.*
6. *Обоснование нормативов.*
7. *Обоснование уровня оплаты труда.*

Классификация взаимосвязей

- теснота связи
  - функциональные (полные)
  - корреляционные (неполные)
- характер связи
  - прямые
  - обратные
- вид уравнения
  - линейные (прямолинейные)
  - нелинейные (криволинейные)
- количество факторов
  - однофакторные (парная зависимость)
  - многофакторные (множественная зависимость)

Задачи корреляционного анализа

- измерение тесноты связи между варьирующими признаками,
- определение неизвестных причинных связей,
- оценка факторов оказывающих наибольшее влияние на результативный признак.

Задачи регрессионного анализа

- установление формы зависимости,
- определение функции регрессии,
- использование уравнения для оценки неизвестных значений зависимой переменной.

Методы оценки тесноты связи

Количественная шкала	Порядковая шкала	Номинальная шкала
Линейный коэффициент корреляции	Коэффициент Спирмена	Коэффициент ассоциации и контингенции
Корреляционное отношение	Коэффициент Кенделла	Коэффициенты Пирсона и Чупрова

Виды шкал

- *Количественная – используется для описания количественных показателей;*
- *Номинальная – шкала наименований (атрибутивных и альтернативных признаков) – (= и ≠);*
- *Порядковая – применяется для измерения упорядоченности объектов по одному или нескольким признакам – (>, <, =).*

Линейный коэффициент корреляции

$$r_{xy} = \frac{n\sum x \cdot y - \sum x \cdot \sum y}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right] \cdot \left[n\sum y^2 - \left(\sum y\right)^2\right]}}$$

Линейный коэффициент корреляции

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

Линейный коэффициент корреляции

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y}$$

Линейный коэффициент корреляции

Величина коэффициента корреляции	Характер связи
До | ± 0,3 |	Практически отсутствует
| ± 0,3 | - | ± 0,5 |	Слабая
| ± 0,5 | - | ± 0,7 |	Умеренная
| ± 0,7 | - | ± 1,0 |	Сильная

Корреляционное отношение

$$\eta = \sqrt{\frac{\sigma_{yx}^2}{\sigma_y^2}}$$

σ_{yx}^2 *Характеризует вариацию результативного признака под влиянием факторного*

σ_{y}^2 *Характеризует вариацию результативного признака под влиянием всех факторов*

Ранговый коэффициент корреляции Спирмена

$$\rho = 1 - \frac{6}{n(n^2 - 1)} \cdot \sum_{k=1}^{n} (A_k - B_k)^2$$

$A_k - ранг\_k - того\_наблюдения\_по\_показателю\_x;$

$B_k - ранг\_k - того\_наблюдения\_по\_показателю\_y;$

$n - число\_пар\_наблюдений.$

Ранговый коэффициент корреляции Кенделла

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

S – сумма баллов, если баллом +1 оценивается пара рангов, имеющих по обоим показателям одинаковый порядок, а баллом -1 – пара с разным порядком.

Пример: баллы студентов по гуманитарным (x) и естественным (y) дисциплинам

x	Y	N_x	N_y	$D=N_x - N_y$	d^2	«+»	«-»
46	45						
60	69						
66	59						
68	49						
71	54						
78	70						
82	58						
90	75						

Пример: баллы студентов по гуманитарным (x) и естественным (y) дисциплинам

x	Y	N_x	N_y	$D=N_x - N_y$	d^2	«+»	«-»
46	45	1	1				
60	69	2	6				
66	59	3	5				
68	49	4	2				
71	54	5	3				
78	70	6	7				
82	58	7	4				
90	75	8	8				

Пример: баллы студентов по гуманитарным (x) и естественным (y) дисциплинам

x	Y	N_x	N_y	$D=N_x - N_y$	d^2	«+»	«-»
46	45	1	1	0	0		
60	69	2	6	-4	16		
66	59	3	5	-2	4		
68	49	4	2	2	4		
71	54	5	3	2	4		
78	70	6	7	-1	1		
82	58	7	4	3	9		
90	75	8	8	0	0		
					38		

Пример: баллы студентов по гуманитарным (x) и естественным (y) дисциплинам

x	Y	N_x	N_y	$D=N_x - N_y$	d^2	«+»	«-»
46	45	1	1	0	0	7	0
60	69	2	6	-4	16	2	4
66	59	3	5	-2	4	2	3
68	49	4	2	2	4	4	0
71	54	5	3	2	4	3	0
78	70	6	7	-1	1	1	1
82	58	7	4	3	9	1	0
90	75	8	8	0	0	-	-
					38	20	8

Ранговый коэффициент корреляции Спирмена

$$\rho = 1 - \frac{6*38}{8(64-1)} = 1 - 0.453 = 0.547$$

Ранговый коэффициент корреляции Кенделла

$$\tau = \frac{2(20-8)}{8(8-1)} = \frac{24}{56} = 0.429$$

Номинальные шкалы

X	*Y*		
	1	*2*	
1	*a*	*b*	*a+b*
2	*c*	*d*	*c+d*
	a+c	*b+d*	

Коэффициент ассоциации

$$k_a = \frac{a \cdot d - b \cdot c}{a \cdot d + b \cdot c}$$

Если $k_a > 0{,}5$*, то между признаками имеется существенная взаимосвязь*

Коэффициент контингенции

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

Если $k_k \geq 0{,}3$, то между признаками имеется существенная взаимосвязь

Пример

Потребление наркотиков	Семейное положение		Всего
	Замужем (женат)	Не замужем (холост)	
Потреблял	10,0	14,5	24,5
Не потреблял	2,5	4,5	7,0
Итого	12,5	19,0	31,5

Пример

$$K_a = \frac{10*4.5-14.5*2.5}{10*4.5+14.5*2.5} = 0.108$$

$$K_k = \frac{10*4.5-14.5*2.5}{\sqrt{(10+4.5)*(14.5+4.5)*(4.5+2.5)*(2.5+10)}} =$$

$$= 0.043$$

Коэффициенты сопряженности

Y / x	I	II	III	Всего
I	n_{yx}	n_{yx}	n_{yx}	n_x
II	n_{yx}	n_{yx}	n_{yx}	n_x
III	n_{yx}	n_{yx}	n_{yx}	n_x
Всего	n_y	n_y	n_y	n

Коэффициент сопряженности Пирсона

$$K_П = \sqrt{\frac{\varphi^2}{1+\varphi^2}}$$

где

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

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

Коэффициент сопряженности Чупрова

$$K_{Ч} = \sqrt{\frac{\varphi^2}{\sqrt{(m_1 - 1)\cdot(m_2 - 1)}}}$$

m_1 – количество градаций первого признака (число строк)

m_2 – количество градаций второго признака (число столбцов)

Пример

Форма собственности предприятия	Оценка уровня жизни				Итого
	Вполне удовле творен	Скорее удовлетво рен	Скорее не удовлетво рен	Совсем не удовлетворе н	
Государственная	31	35	35	35	136
Муниципальная	17	13	14	9	53
Смешанная	4	2	1	1	8
Частная	8	5	4	3	20
Итого	60	55	54	48	217

Пример

$$1+\varphi^2 = \frac{\frac{31^2}{60}+\frac{35^2}{55}+\frac{35^2}{54}+\frac{35^2}{48}}{136}+\frac{\frac{17^2}{60}+\frac{13^2}{55}+\frac{14^2}{54}+\frac{9^2}{48}}{53}+$$

$$+\frac{\frac{4^2}{60}+\frac{2^2}{55}+\frac{1^2}{54}+\frac{1^2}{48}}{8}+\frac{\frac{8^2}{60}+\frac{5^2}{55}+\frac{4^2}{54}+\frac{3^2}{48}}{20}=$$

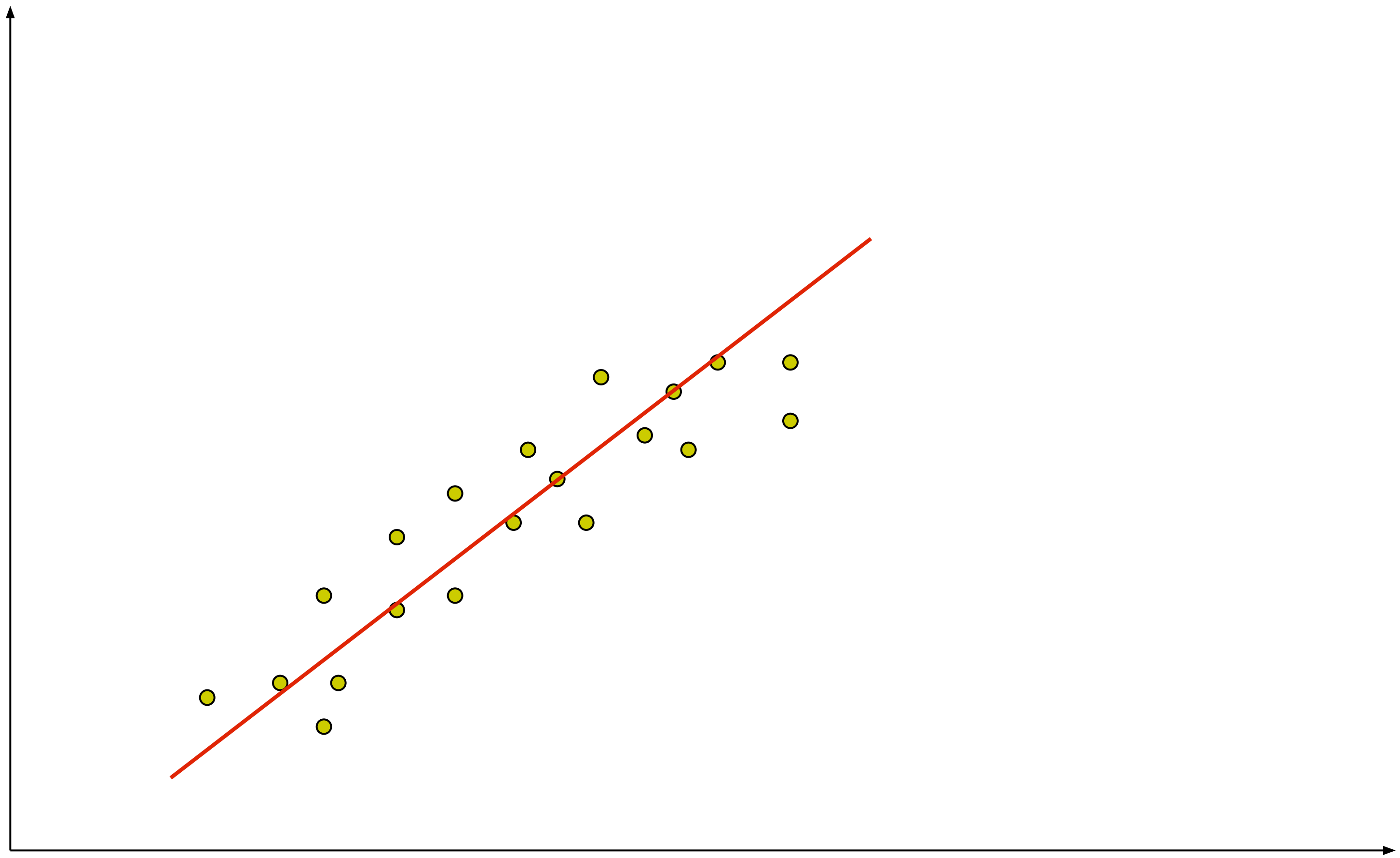
$$=0{,}636+0{,}249+0{,}047+0{,}100=1{,}032$$

Пример

$$K_{П} = \sqrt{\frac{0.032}{1.032}} = 0.176$$

$$K_{Ч} = \sqrt{\frac{0.032}{\sqrt{(4-1)\cdot(4-1)}}} = 0.103$$

Корреляционное поле

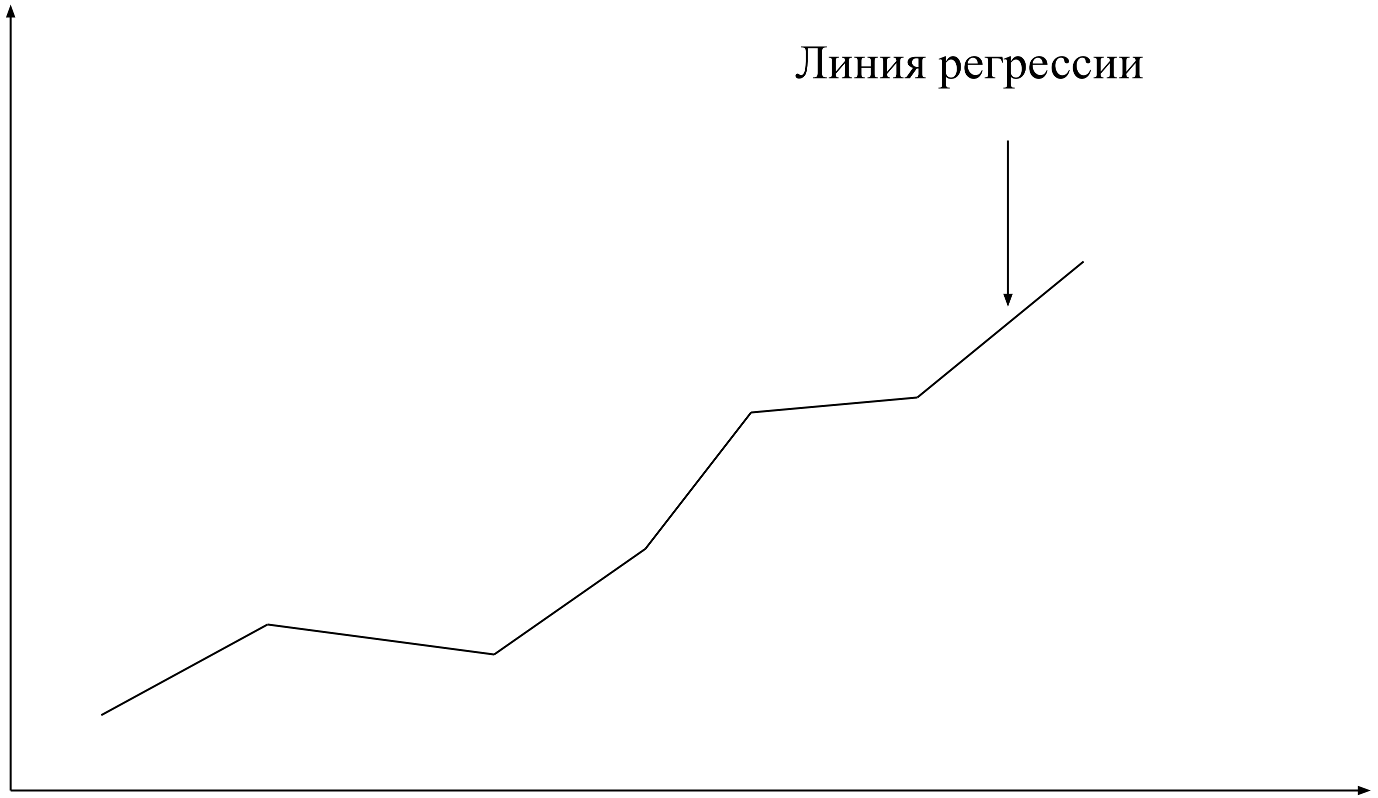


Корреляционная таблица

X \ Y	y_1	y_2	y_m	Итого
x_1	φ_{11}	φ_{12}	φ_{1n}	$\sum\varphi_{ij}$
x_2	φ_{21}	φ_{22}	φ_{2n}	$\sum\varphi_{ij}$
x_n	φ_{n1}	φ_{n2}	φ_{nm}	$\sum\varphi_{ij}$
Итого	$\sum\varphi_{ji}$	$\sum\varphi_{ji}$	$\sum\varphi_{ji}$	

Метод параллельных данных

Сопоставление двух или нескольких рядов статистических величин.



Метод параллельных данных

Номер студента	Балл в сессию, y	Кол-во пропущенных семинаров, x	Приведенные параллельные данные	
			x	y
1	5	1	***1***	***5***
2	3	8	***2***	***5***
3	4	3	***2***	***5***
4	4	5	***3***	***4***
5	3	8	***4***	***4***
6	2	10	***5***	***4***
7	5	2	***6***	***3***
8	4	4	***8***	***3***
9	5	2	***8***	***3***
10	3	6	***10***	***2***

Построение уравнения регрессии

Параметры уравнения регрессии определяют из так называемой системы нормальных уравнений, отвечающей требованию метода наименьших квадратов (МНК).

$$\sum\left(y - \overline{y}_x\right)^2 \rightarrow \min$$

Построение уравнения регрессии

Для линейной зависимости:

$$y = a_0 + a_1 x$$

$$S = \sum_{i=1}^{n} (y_x - y)^2 = \sum_{i=1}^{n} (a_0 + a_1 x - y)^2 \to \min$$

Построение уравнения регрессии

Для линейной зависимости:

$$\frac{dS}{da_0} = 2\sum(a_0 + a_1 - y) * 1 = 0$$

$$\frac{dS}{da_1} = 2\sum(a_0 + a_1 - y) * x = 0$$

Построение уравнения регрессии

Для линейной зависимости:

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

Построение уравнения регрессии

Для линейной зависимости:

$$a_0 = \frac{\sum x^2 \sum y - \sum x \sum xy}{n\sum x^2 - (\sum x^2)}$$

$$a_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x^2)}$$

Построение уравнения регрессии

Для параболы второго порядка:

$$y = a_0 + a_1 x + a_2 x^2$$

$$\begin{cases} na_0 + a_1 \sum x = \sum y \\ a_0 \sum x + a_1 \sum x^2 = \sum xy \\ a_0 \sum x^2 + a_1 \sum x^3 = \sum x^2 y \end{cases}$$

Построение уравнения регрессии

Для гиперболы:

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

$$\begin{cases} na_0 + a_1 \sum \frac{1}{x} = \sum y \\ a_0 \sum \frac{1}{x} + a_1 \sum \frac{1}{x^2} = \sum \frac{y}{x} \end{cases}$$

Аналитические характеристики производственных функций и их экономическая трактовка.

К числу важнейших аналитических характеристик относятся:

- ✔ *коэффициент детерминации,*
- ✔ *средняя и предельная эффективность ресурса,*
- ✔ *коэффициент эластичности,*
- ✔ *норма взаимозаменяемости факторов*

Аналитические характеристики производственных функций и их экономическая трактовка.

Коэффициент детерминации характеризует удельный вес факторного признака или признаков в общей вариации зависимого показателя.

$$K_Д = 100 \cdot r^2$$

<10%	Слабая
10% - 50%	Средняя
>50%	Сильная
100%	Полная

Аналитические характеристики производственных функций и их экономическая трактовка.

Средняя эффективность ресурса определяется путем деления соответствующей производственной функции на объем использованного ресурса. Эффективность измеряется в единицах результативного показателя в расчете на единицу ресурса.

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

Аналитические характеристики производственных функций и их экономическая трактовка.

Предельная эффективность ресурса измеряется в единицах зависимого показателя в расчете на единицу факторного признака.

$$Э_ч = \frac{dy}{dx} \qquad Э_ч = \frac{d(a_0 + a_1 x)}{dx} = a_1$$

Аналитические характеристики производственных функций и их экономическая трактовка.

Эластичность - рассчитывают путем умножения предельной эффективности ресурса на соотношение значений фактора и зависимого признака

$$\varepsilon = \frac{Э_x}{\bar{y}} \qquad \varepsilon = \frac{a_1 \bar{x}}{a_0 + a_1 \bar{x}}$$

Оценка значимости параметров взаимосвязи

Стандартная ошибка коэффициента корреляции:

$$\sigma_{r^{xy}} = \sqrt{\frac{1 - r_{xy}^2}{n - 2}}.$$

$$\sigma_{r^{xy}} < r_{xy}$$

Оценка значимости параметров взаимосвязи

Значимость r_{xy}

$$t_{расч} = r_{xy} \cdot \sqrt{\frac{n-2}{1-r_{xy}^2}},$$

Если $t_{расч}$ больше теоретического (табличного) значения критерия Стьюдента ($t_{табл}$) для заданного уровня вероятности и (n - 2) степеней свободы, то можно утверждать, что r_{xy} значимо.

Оценка значимости параметров взаимосвязи

Вывод о правильности выбора вида взаимосвязи и характеристику значимости всего уравнения регрессии получают с помощью F-критерия, вычисляя его расчетное значение.

Оценка значимости параметров взаимосвязи

$$F_{расч} = \frac{R^2(n-m)}{(1-R^2)(m-1)},$$

где n - число наблюдений;

m - число параметров уравнения регрессии.

$F_{расч}$ также должно быть больше $F_{теор}$ при $v_1 = (m - 1)$ и $v_2 = (n - m)$ степенях свободы.

Данные о выпуске продукции (х) и расходе условного топлива (у) по однотипным предприятиям

№	X	Y	X^2	Y^2	XY	$Y_{модель}$
1	5	4				
2	6	4				
3	8	6				
4	8	5				
5	10	7				
6	10	8				
7	14	8				
8	20	10				
9	20	12				
10	24	16				
Сумма	***125***	***80***				
Среднее	***12,5***	***8,0***				

Данные о выпуске продукции (x) и расходе условного топлива (y) по однотипным предприятиям

№	X	Y	X^2	Y^2	XY	$Y_{модель}$
1	5	4	25	16	20	
2	6	4	36	16	24	
3	8	6	64	36	48	
4	8	5	64	25	40	
5	10	7	100	49	70	
6	10	8	100	64	80	
7	14	8	196	64	112	
8	20	10	400	100	200	
9	20	12	400	144	240	
10	24	16	576	256	384	
Сумма	***125***	***80***	***1961***	***770***	***1218***	
Среднее	***12,5***	***8,0***	***196,1***	***77,0***	***121,8***	

Коэффициент корреляции

$$r_{xy} = \frac{n\sum x \cdot y - \sum x \cdot \sum y}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right] \cdot \left[n\sum y^2 - \left(\sum y\right)^2\right]}} =$$

$$= \frac{10 \cdot 1218 - 125 \cdot 80}{\sqrt{\left[10 \cdot 1961 - 125^2\right] \cdot \left[10 \cdot 770 - 80^2\right]}} = 0.96$$

Определение параметров уравнения регрессии

$$a_0 = \frac{\sum x^2 \sum y - \sum x \sum xy}{n\sum x^2 - (\sum x^2)} =$$

$$= \frac{1961 \cdot 80 - 125 \cdot 1218}{10 \cdot 1961 - 125^2} = 1.16$$

Определение параметров уравнения регрессии

$$a_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x^2)} =$$

$$= \frac{10 \cdot 1218 - 125 \cdot 80}{10 \cdot 1961 - 125^2} = 0.55$$

Уравнение регрессии

$$y = a_0 + a_1 x$$

$$y = 1.16 + 0.55 \cdot x$$

Данные о выпуске продукции (х) и расходе условного топлива (у) по однотипным предприятиям

№	X	Y	X^2	Y^2	XY	$Y_{модель}$
1	5	4	25	16	20	3,90
2	6	4	36	16	24	4,44
3	8	6	64	36	48	5,54
4	8	5	64	25	40	5,54
5	10	7	100	49	70	6,63
6	10	8	100	64	80	6,63
7	14	8	196	64	112	8,82
8	20	10	400	100	200	12,10
9	20	12	400	144	240	12,10
10	24	16	576	256	384	14,29
Сумма	***125***	***80***	***1961***	***770***	***1218***	***80,00***
Среднее	***12,5***	***8,0***	***196,1***	***77,0***	***121,8***	***8,0***

Аналитические характеристики производственных функций и их экономическая трактовка.

$$K_{Д} = 100 \cdot r^2 = 100 \cdot 0.96^2 = 91.74$$

$$\overline{Э} = \frac{\overline{y}}{\overline{x}} = \frac{8}{12.5} = 1.56$$

$$Э_{ч} = a_1 = 0.55$$

Аналитические характеристики производственных функций и их экономическая трактовка.

$$\varepsilon = \frac{Э_x}{\bar{y}} = \frac{0.55}{8} = 0.07$$

$$\sigma_{r^{xy}} = \sqrt{\frac{1 - r_{xy}^2}{n - 2}} = \sqrt{\frac{1 - 0.96^2}{10 - 2}} = 0.10$$