ЭКОНОМЕТРИКА

НИКИФОРОВ
СЕРГЕЙ
АЛЕКСЕЕВИЧ

ТЕОРИЯ

- Эконометрика изучает связи между явлениями. Связи могут быть **полными** – ***функциональными*** и **неполными** – ***корреляционными***. Для функциональной зависимости характерным является то, что изменение признака следствия целиком определяется изменением признаком фактора.

ТЕОРИЯ

- Например: площадь круга выражается формулой , $S = \pi R^2$ т.е. площадь круга изменяется от изменения квадрата ее радиуса.
- Корреляционные связи характеризуются тем, что величина признака следствия изменяется под влиянием нескольких факторов. При этом одни факторы оказывают влияние на все единицы изучаемой совокупности, а другие -

ТЕОРИЯ

- только на отдельные из них. Корреляционные связи проявляются отчетливо только в большом числе факторов, т.к. при этом сглаживаются индивидуальные особенности и второстепенные факторы. Например: анализируя производительность труда на предприятии, можно увидеть зависимость от ее уровня энерговооруженности труда.

ТЕОРИЯ

- Но производительность труда зависит и от других факторов: от режима работы предприятия, организации снабжения, квалификации работников и т.д. Поэтому зависимость производительности труда от уровня энерговооруженности труда не может быть полной, а является корреляционной.

ТЕОРИЯ

- Для выявления корреляционных зависимостей используют теоретическую формулу связи в виде математического уравнения, которое называется ***уравнением регрессии***. Регрессия – это зависимость среднего значения величины от другой величины или нескольких величин.

ТЕОРИЯ

- Уравнение регрессии может быть описано уравнением линейной связи
- $Y = a_0 + a_1x$,
- гиперболой $Y = a_0 + a_1 1/x$,
- параболой $Y = a_0 + a_1x + a_2x^2$.
- Если результативный признак (Y) с увеличением факторного признака (X) равномерно возрастает или убывает, то такая зависимость является линейной и выражается уравнением прямой.

ТЕОРИЯ

- Найти теоретическое уравнение связи – это значит рассчитать параметры прямой линии методом наименьших квадратов, который дает систему двух нормальных уравнений.

ТЕОРИЯ

- $Yx = a_0 + a_1 x$
- $na_0 + a_1\sum(x) = \sum(y)$
- $a_0\sum(x) + a_1\sum(x) = \sum(xy)$
- Yx – теоретическое значение результативного признака.
- Y – индивидуальное значение результативного признака.

ТЕОРИЯ

- n– число показателей.
- X - индивидуальное значение результативного признака.
- a_0,a_1 – параметры (коэффициенты) уравнения регрессии.

ТЕОРИЯ

- Теоретическое уравнение выражает функциональную зависимость (Y) от (X). Это возможно допустить, если прочие факторы, влияющие на (Y) не оказывают в данном случае существенного влияния.

ТЕОРИЯ

- Это бывает, когда корреляционная зависимость между (Y) и (X) высокая. В этом случае параметр (a_1) при (X) в уравнении регрессии приобретает большое практическое значение. Этот параметр, который называется ***коэффициентом регрессии***,
- характеризует, в какой мере увеличивается (Y) , с ростом величины (X).

ЗАДАЧА

- Имеются выборочные данные по однородным предприятиям: энерговооруженность труда одного рабочего (квт /час) и выпуск готовой продукции (шт).
- ОПРЕДЕЛИТЬ:
- 1. Факторные и результативные признаки.
- 2. Провести исследование взаимосвязи энерговооруженности и выпуска готовой продукции.

ЗАДАЧА

- 3. Построить уравнение регрессии и вычислить коэффициент регрессии.
- 4. Построить графики практической и теоретической линии регрессии.
- 5. Определить форму связи и измерить тесноту связи.
- 6. Провести оценку адекватности.

РЕШЕНИЕ

- 1. (X) – факторным признаком является энерговооруженность.
- (Y)– результативным признаком является выпуск готовой продукции.
- 2. Исходные данные поместим в следующую таблицу.

Номер анализа	X	Y	(X –X̄)	(X – X̄)²	X²	Y²	(XY)
1	1	25	-1	1	1	625	25
2	1	20	-1	1	1	400	20
3	1,5	20	-0,5	0,25	2,25	400	30
4	1,5	22	-0,5	0,25	2,25	484	33
5	2	25	0	0	4	625	50
6	2	28	0	0	4	784	56
7	2,5	30	+0,5	0,25	6,25	900	75
8		32	+0,5	0,25	6,25	1024	80
9		32	+1	1	9	1024	96
10	2,5	30	+1	1	9	900	90
ИТОГО	2	264		8	45	7166	

РЕШЕНИЕ

- 3. Первичная информация проверяется на однородность по признаку-фактору с помощью коэффициента вариации

$$\bar{x} = \frac{\sum x}{n} = \frac{20}{10} = 2$$

$$\delta = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} = \sqrt{\frac{8}{10}} = 0{,}894$$

$$v = \frac{\delta}{\bar{x}} = \frac{0{,}894}{2} = 0{,}447 \times 100\% = 44{,}7\%$$

РЕШЕНИЕ

- 4. Проверка первичной информации на нормальность распределения с помощью правила «трех сигм». Сущность правила заключается в том, что в интервал «трех сигм» должны попасть факторные признаки. Те показатели, которые больше или меньше интервала «трех сигм», удаляются из таблицы.

ИНТЕРВАЛЫ $\bar{X} \pm \check{G}$	ЧИСЛО ЕДИНИЦ ВХОДЯЩИХ В ИНТЕРВАЛ	УДЕЛЬНЫЙ ВЕС ЕДИНИЦ	УДЕЛЬНЫЙ ВЕС ПРИ НОРМАЛЬНОМ РАСПРЕДЕЛЕНИИ
$\bar{X} \pm 1\check{G}$ $(\bar{X} - 1\check{G}) – (\bar{X} + 1\check{G})$ (2 – 0,894) – (2 + 0,894) 1,106 – 2,894	6	60	68,3
$\bar{X} \pm 2\check{G}$ $(\bar{X} - 2\check{G}) – (\bar{X} + 2\check{G})$ (2 – 1,788) – (2 + 1,788) 0,212 – 3,788	10	100	95,4
$\bar{X} \pm 3\check{G}$ $(\bar{X} - 3\check{G}) – (\bar{X} + 3\check{G})$ (2 – 2,682) – (2 + 2,682) -0,683 – 4,682	10	100	99,7

РЕШЕНИЕ

- 5. Исключить из первичной информации резко выделяющиеся единицы, которые по признаку-фактору не попадают в интервал «трех сигм».
- Вывод: Резко выделяющихся единиц в первичной информации нет.

- 6. Для установления факта наличия связи производится аналитическая группировка по признаку-фактору. Построить интервальный ряд распределения.
- При этом формула для определения величины интервала имеет следующий вид:

$$i = \frac{x_{\max} - x_{\min}}{m} = \frac{3-1}{4} = 0{,}5$$

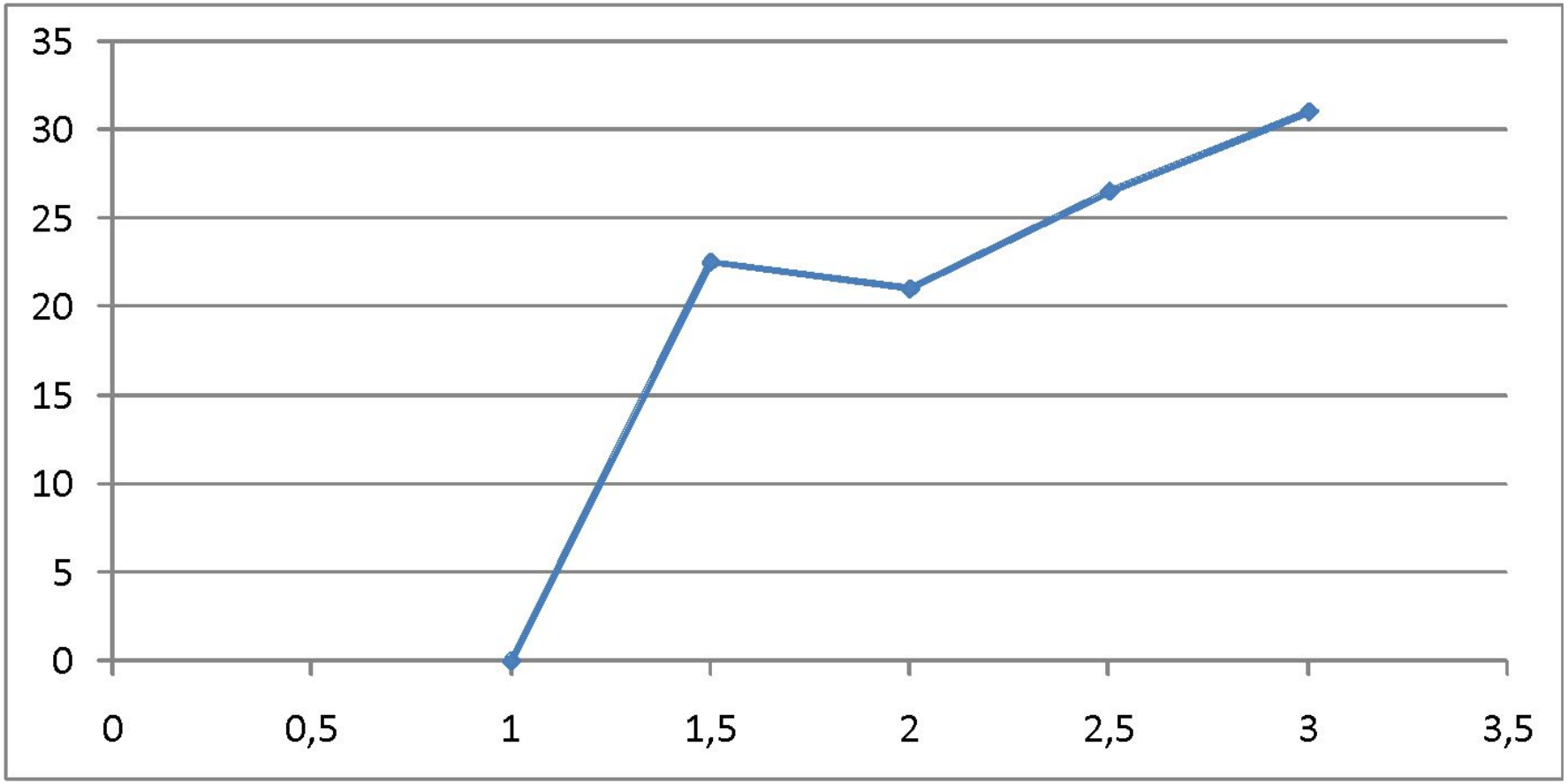
$$m \geq 4$$

№	интервалы X	Номер анализа	Число анализов	Y	∑Y	Ȳ
1	1 – 1,5	1, 2	2	20, 25	45	22,5
2	1,5 – 2	3, 4	2	20, 22	42	21
3	2 – 2,5	5, 6	2	28, 25	53	26,5
4	2,5 – 3	7, 8, 9, 10	4	30, 32, 30, 32	124	31
ИТОГО	–	–	10	–	264	–

РЕШЕНИЕ

- 7. Построить эмпирическую линию связи. По оси абсцисс откладываются значения интервалов факторного признака – (X) . По оси ординат откладываются значения средней величины результативного признак – ($\bar{Y}$).

ЭМПИРИЧЕСКАЯ ЛИНИЯ СВЯЗИ



- 8. Для измерения степени тесноты связи используется линейный коэффициент связи:

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

$$= \frac{555 - \frac{20 \cdot 264}{10}}{\sqrt{(45 - \frac{20^2}{10})(7166 - \frac{264^2}{10})}} = 0{,}86$$

РЕШЕНИЕ

- Полученное значение линейного коэффициента корреляции необходимо сравнить с табличными данными.

ПОКАЗАТЕЛЬ КОРРЕЛЯЦИИ	ТЕСНОТА СВЯЗИ
0	СВЯЗЬ ОТСУТСТВУЕТ
0,2 – 0,3	СЛАБАЯ
0,3 – 0,5	УМЕРЕННАЯ
0,5 – 0,7	ЗАМЕТНАЯ
0,7 – 0,99	ВЫСОКАЯ
1	ФУНКЦИОНАЛЬНАЯ

РЕШЕНИЕ

- Т.о. связь высокая. r = 0,86, а интервал связи (0,7 – 0,99).
- 9. Предположим, что между энерговооруженности труда и выпуском готовой продукции существует линейная корреляционная связь которую можно выразить уравнением прямой.
- Для этого составим новую таблицу.

№	X	Y	XY	X^2	Y^2	$(Y - \bar{Y})$	$(Y - \bar{Y})^2$	Yx	(Y –Yx)	$(Y - Yx)^2$
1	1	25	25	1	625	-1,4	1,96	21	+4	16
2	1	20	20	1	400	-6,4	40,96	21	-1	1
3	1,5	20	30	2,25	400	-6,4	40,96	23,7	-3,7	13,96
4	1,5	22	33	2,25	484	-4,4	19,36	23,7	-1,7	2,89
5	2	25	50	4	625	-1,4	1,96	26,4	-1,4	1,96
6	2	28	56	4	784	+1,6	2,56	26,4	+1,6	2,56
7	2,5	30	75	6,25	900	+3,6	12,96	29,1	+0,9	0,81
8	2,5	32	80	6,25	1024	+5,6	31,36	29,1	+2,9	8,41
9	3	32	96	9	1024	+5,6	31,36	31,8	+0,2	0,04
10	3	30	90	9	900	+3,6	12,96	31,8	-1,8	3,24
∑	20	264	555	45	7166	-	196,4	264	-	50,6

- Вычислим параметры прямой с помощью системы двух нормальных уравнений:
- $Yx = a_0 + a_1X$
- $\begin{cases} na_0 + a_1\Sigma(X) = \Sigma(Y) \\ a_0\sum(X) + a_1\sum(X^2) = \sum(XY) \end{cases}$
- $\begin{cases} 10a_0 + 20a_1 = 264 \\ 20a_0 + 45a_1 = 555 \end{cases}$
- $\begin{cases} 10a_0 + 20a_1 = 264 \times \{ (-2)\} \\ 20a_0 + 45a_1 = 555 \end{cases}$

РЕШЕНИЕ

- $\begin{cases} -20a_0 - 40a_1 = -528 \\ +20a_0 + 45a_1 = 555 \end{cases}$ +
- $5a_1 = 27$
- $a_1 = 5{,}4$
- $a_0 = 15{,}6$

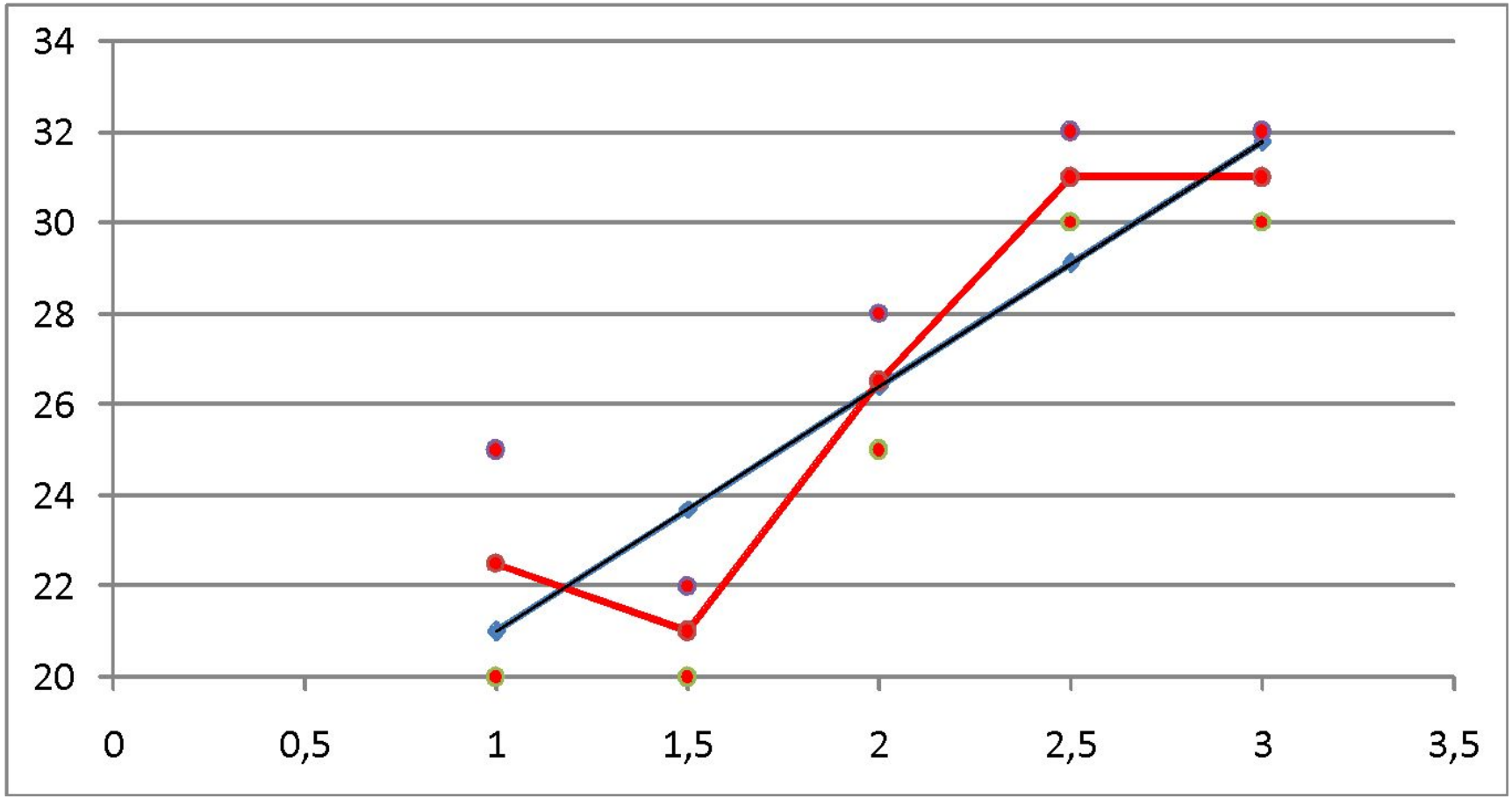
РЕШЕНИЕ

- Конечное уравнение следующее.
- $Yx = 15,6 + 5,4(X)$
- В уравнении регрессии коэффициент a_1 показывает, что с увеличением энерговооруженности труда одного рабочего на 1 (квт/час) выпуск готовой продукции возрастает на 5,4 шт.

РЕШЕНИЕ

- Построим графики практической и теоретической линии регрессии. По оси абсцисс отложим значения факторного признака (x) , по оси ординат (Yx) и (Y). Чтобы определить (Yx) в уравнение регрессии подставить значения (x) и занести в таблицу.

ЛИНИИ РЕГРЕССИИ



ИЗМЕРИТЬ ТЕСНОТУ СВЯЗИ

- 10. Одним из важнейших этапов исследования является измерение тесноты связи. Для этого применяют линейный коэффициент корреляции (r) и индекс корреляции (R). Индекс корреляции применяется для измерения тесноты связи между признаками при любой форме связи, как линейной, так и нелинейной.

- Но его можно вычислять только после того, как определена форма связи и вычислена теоретическая линия регрессии.

$$\delta^2_{(y-y_x)} = \frac{\sum (y - y_x)^2}{n} = \frac{50,6}{10} = 5,06$$

$$\bar{y} = \frac{\sum y}{n} = \frac{264}{10} = 26,4$$

$$\delta^2_y = \frac{\sum (y - \bar{y})^2}{n} = \frac{196,4}{10} = 19,64$$

$$R = \sqrt{\frac{\delta^2_y - \delta^2_{(y-y_x)}}{\delta^2_y}} = \sqrt{\frac{19,64 - 5,06}{19,64}} = \sqrt{0,742} = 0,86$$

- Индекс корреляции измеряется от 0 до 1. Чем ближе индекс к 1, тем теснее связь между признаками. Частным случаем индекса корреляции является коэффициент корреляции, который применяется только при линейной форме связи. В отличии от индекса корреляции линейный коэффициент корреляции показывает не только тесноту связи, но и направление связи (прямая или обратная) и измеряется от -1 до +1.

ВЫВОД

- Все показатели тесноты корреляционной связи показывают тесную связь между производительностью труда и энерговооруженностью труда. Т.к. R=r=0,86 то можно сделать заключение, что гипотеза о линейной форме связи подтверждена.

АДЕКВАТНОСТЬ МОДЕЛИ

- Проведем оценку адекватности регрессионной модели с помощью критерия Фишера.

$$\delta^2_{y_x} = \delta^2_y - \delta^2_{(y-y_x)} = 19{,}64 - 5{,}06 = 14{,}58$$

$$F_э = \frac{\delta^2_{y_x}}{\delta^2_{(y-y_x)}} \cdot \frac{n-m}{m-1} = \frac{14{,}58}{5{,}06} \cdot \frac{10-2}{2-1} = 23{,}048$$

ВЫВОД

- Табличное значение критерия Фишера равно (Fт = 20,20). Эмпирическое значение критерия Фишера (Fэ = 23,048) сравниваем с табличным.
- Если Fэ < Fт, то уравнение регрессии можно признать неадекватным.
- Если Fэ > Fт, то уравнение регрессии признается значимым. (23,048 > 20,20)
- Т.о. данная модель является адекватной.

КОНТРОЛЬНАЯ РАБОТА

	X	Y
1	НВ + 0,5	НВ + 10
2	НВ + 1	НВ + 5
3	НВ + 1,5	НВ + 15
4	НВ + 2	НВ + 20
5	НВ + 2,5	НВ + 25
ИТОГО Fт = 10,13		