Пример 2. По имеющимся данным о самой многочисленной группе рабочих предприятия (Таблица 10) определите тесноту связи между признаками *Среднемесячная производительность труда* и *Премия по итогам года* с помощью:

1) коэффициента Фехнера;
2) коэффициента ранговой корреляции Спирмена;
3) коэффициента линейной корреляции.

Решение.

Построим вспомогательную таблицу для определения необходимых значений (Таблица 73), где x – *Среднемесячная производительность труда* (факторный признак), y – *Премия по итогам года* (результативный признак).

Таблица 73

Таблица для расчетов показателей тесноты связи

	x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	ранг x	ранг y	D^2
1	6,5	15,7	– 0,6	– 0,6	0,36	0,36	0,36	1	2	1
2	6,6	15,5	– 0,5	– 0,8	0,40	0,25	0,64	2	1	1
3	6,8	16,2	– 0,3	– 0,1	0,03	0,09	0,01	3	7	16
4	6,9	16,1	– 0,2	– 0,2	0,04	0,04	0,04	4	6	4
5	7,0	15,9	– 0,1	– 0,4	0,04	0,01	0,16	5	4	1
6	7,0	15,8	– 0,1	– 0,5	0,05	0,01	0,25	5	3	4
7	7,1	17,6	0,0	1,3	0,00	0,00	1,69	6	13	49
8	7,1	16,4	0,0	0,1	0,00	0,00	0,01	6	9	9
9	7,2	16,5	0,1	0,2	0,02	0,01	0,04	7	10	9
10	7,3	16,4	0,2	0,1	0,02	0,04	0,01	8	9	1
11	7,4	16,0	0,3	– 0,3	– 0,09	0,09	0,09	9	5	16
12	7,5	16,7	0,4	0,4	0,16	0,16	0,16	10	11	1
13	7,5	16,3	0,4	0,0	0,00	0,16	0,00	10	8	4
14	7,6	17,2	0,5	0,9	0,45	0,25	0,81	11	12	1
$\sum$	99,5	228,3			1,49	1,47	4,27			117

1) $\bar{x} = \frac{\sum x}{n} = \frac{99,5}{14} = 7,1$; $\bar{y} = \frac{\sum y}{n} = \frac{228,3}{14} = 16,3$;

$K_{\Phi} = \frac{\sum C - \sum H}{n} = \frac{13-1}{12} = 0,857 \Rightarrow$ связь тесная, прямая.

2) $\rho = 1 - \frac{6\sum D^2}{n(n^2-1)} = 1 - \frac{6 \cdot 117}{14 \cdot (256-1)} = 0,803 \Rightarrow$ связь тесная, прямая.

3) $\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{1,47}{14}} = 0,324$; $\sigma_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n}} = \sqrt{\frac{4,27}{14}} = 0,552$;

$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{n\sigma_x \sigma_y} = \frac{1,49}{14 \cdot 0,324 \cdot 0,552} = 0,6 \Rightarrow$ связь заметная, прямая.

Пример 3. По данным о 30 работниках предприятия (Таблица 6) определить тесноту корреляционной связи между признаками *Среднемесячная производительность труда* и *Премия по итогам года* с использованием коэффициента детерминации и эмпирического корреляционного отношения.

Решение.

Эмпирический коэффициент детерминации η^2 оценивает, насколько вариация результативного признака y объясняется вариацией фактора x.

$$\eta^2 = \frac{\delta_x^2}{\sigma^2},$$

где σ^2 – общая дисперсия признака y, δ_x^2 – межгрупповая (факторная) дисперсия признака y.

Для расчета общей дисперсии построим вспомогательную таблицу (Таблица 74).

Таблица 74

Вспомогательная таблица для расчета общей дисперсии

Номер	y	y^2
1	15,7	246,49
2	18,0	324,00
3	12,1	146,41
4	13,8	190,44
5	15,5	240,25
6	17,9	320,41
7	12,8	163,84
8	14,2	201,64
9	15,9	252,81
10	17,6	309,76
11	18,2	331,24
12	13,0	169,00
13	16,5	272,25
14	16,2	262,44
15	16,7	278,89
16	14,6	213,16
17	14,8	219,04
18	16,1	259,21
19	16,7	278,89
20	15,8	249,64
21	16,4	268,96
22	15,0	225,00
23	16,5	272,25
24	18,5	342,25
25	16,4	268,96
26	16,0	256,00
27	19,1	364,81
28	16,3	265,69
29	19,6	384,16
30	17,2	295,84
Итого	483,1	7873,73

$$\overline{y^2} = \frac{\sum y^2}{n} = \frac{7873,73}{30} = 262,46; \quad \bar{y} = \frac{\sum y}{n} = \frac{483,1}{30} = 16,1;$$

$$\sigma^2 = \overline{y^2} - \bar{y}^2 = 262,46 - 16,1^2 = 3,25$$

Для расчета межгрупповой дисперсии необходимо также построить вспомогательную таблицу (Таблица 75). Столбцы 1–3 заполняются по результатам проведенной ранее аналитической группировки (Таблица 11).

Таблица 75

Вспомогательная таблица для расчета межгрупповой дисперсии

Группы рабочих по среднемесячной производительности труда, тыс. руб.	Численность рабочих, чел. f	Среднее значение премии по итогам года в группе, $\bar{y}_{гр}$	$\bar{y}_{гр} - \bar{y}$	$(\bar{y}_{гр} - \bar{y})^2 f$
1	2	3	4	5
4,1 – 5,3	4	13,1	– 3	36
5,3 – 6,5	5	14,9	– 1,2	7,2
6,5 – 7,7	14	16,3	0,2	0,56
7,7 – 8,9	4	17,8	1,7	11,56
8,9 – 10,1	3	19	2,9	25,23
Итого	30	16,1	0,6	80,55

$$\delta_x^2 = \frac{\sum(\bar{y}_{гр} - \bar{y})^2 f}{\sum f} = \frac{80,55}{30} = 2,69; \qquad \eta^2 = \frac{\delta_x^2}{\sigma^2} = \frac{2,69}{3,25} = 0,826;$$

$$\eta = \sqrt{\frac{\delta_x^2}{\sigma^2}} = \sqrt{0,826} = 0,91.$$

Таким образом, вариация премии по итогам года на 82,6% обусловлена вариацией среднемесячной производительности труда и на 17,4% – другими неучтенными факторами. Значение эмпирического коэффициента корреляции согласно шкале Чэддока свидетельствует о наличии весьма тесной связи между рассматриваемыми признаками.

Пример 4. Требуется измерить тесноту связи между прививками от гриппа и пониженной заболеваемостью от гриппа в группе случайно отобранных студентов (Таблица 76) с помощью:

1) коэффициента контингенции;
2) коэффициента ассоциации.

Таблица 76

Данные о заболеваемости гриппом

	Заболели	Не заболели	Итого
Привитые	10	20	30
Непривитые	15	5	20
Всего	25	25	50

Решение.

1) $K = \dfrac{ad - bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}} = \dfrac{10 \cdot 5 - 20 \cdot 15}{\sqrt{25 \cdot 25 \cdot 30 \cdot 20}} = -0{,}41$

2) $K = \dfrac{ad - bc}{ad + bc} = \dfrac{10 \cdot 5 - 20 \cdot 15}{10 \cdot 5 + 20 \cdot 15} = -0{,}71$

И в том, и в другом случае связь получилась обратной. Так как по модулю значения обоих коэффициентов превышает 0,3, то между качественными признаками *Наличие прививки от гриппа* и *Заболеваемость гриппом* существует корреляционная связь.

6.2. Построение линейной модели регрессии. Оценка ее адекватности

Уравнение однофакторной (парной) линейной корреляционной связи имеет вид:

$$\tilde{y} = a_0 + a_1 x,$$

где $\tilde{y}$ – теоретическое значение результативного признака, полученное по уравнению регрессии; a_0 и a_1 – коэффициенты (параметры) уравнения регрессии.

Коэффициент парной линейной регрессии a_1 показывает среднее изменение результативного признака y при изменении факторного признака x на одну единицу. Знак коэффициента a_1 указывает направление этого изменения.

Параметры уравнения a_0 и a_1 определяют методом наименьших квадратов по следующим формулам:

$$a_1 = \frac{n\sum yx - \sum y \sum x}{n\sum x^2 - (\sum x)^2}; \qquad a_0 = \frac{\sum y \sum x^2 - \sum yx \sum x}{n\sum x^2 - \sum x \sum y}.$$

Можно использовать и другие формулы (однако расчёты будут менее точными, поскольку в них присутствуют средние величины):

$$a_1 = \frac{\sum (y - \bar{y})(x - \bar{x})}{\sum (x - \bar{x})^2} \text{ или } a_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2}, \quad a_0 = \bar{y} - a_1 \cdot \bar{x}.$$

Определив значения a_0 и a_1, подставляют их в уравнение регрессии и находят значения $\tilde{y}$, которые зависят только от заданного значения x.

Для практического использования модели регрессии очень важна её ***адекватность*** – соответствие фактическим статистическим данным. Параметры уравнения регрессии, коэффициенты корреляции и детерминации могут быть искажены действием случайных факторов. Чтобы проверить, насколько эти показатели характерны для всей генеральной совокупности, не являются ли они результатом стечения случайных обстоятельств, необходимо оценить их значимость.

Значимость коэффициентов парной линейной регрессии и коэффициента линейной корреляции проверяют с помощью t – критерия Стьюдента (см. Приложение) в несколько этапов:

1. Вычисляют расчетные (фактические) значения t – критерия:

Таблица 77

Формулы для расчета фактических значений t–критерия Стьюдента

	Расчетные значения	Пояснения
Параметр a_0	$t_{a_0} = \lvert a_0 \rvert \cdot \dfrac{\sqrt{n-2}}{\sigma_{ост}}$	n – объём выборки $\sigma_{ост} = \sqrt{\dfrac{\sum(y-\tilde{y})^2}{n}}$ $\sigma_x = \sqrt{\dfrac{\sum(x-\bar{x})^2}{n}} = \sqrt{\dfrac{\sum x^2}{n} - \left(\dfrac{\sum x}{n}\right)^2}$
Параметр a_1	$t_{a_1} = \lvert a_1 \rvert \cdot \dfrac{\sqrt{n-2}}{\sigma_{ост}} \sigma_x$	
Линейный коэффициент корреляции	$t_r = \lvert r \rvert \cdot \sqrt{\dfrac{n-2}{1-r^2}}$	

2. Расчётные значения t–критерия сравнивают с критическими, которые находят по таблице Стьюдента при определённом уровне значимости α и числом степеней свободы вариации для прямой $\nu = n-2$. В социально-экономических исследованиях уровень значимости α обычно принимают равным 0,05 или 0,01. Если $t_{расч}$ > $t_{табл}$, то исследуемый показатель признаётся значимым, и гипотеза о его случайности отвергается.

При проверке адекватности модели определяется также теснота корреляционной связи между признаками x и y с помощью ***теоретического корреляционного отношения*** (***индекса корреляции R***):

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}},$$

где $\delta^2 = \dfrac{\sum(\tilde{y}-\bar{y})^2}{n}$ – межгрупповая дисперсия результативного признака y, обусловленная влиянием только фактора x; $\sigma^2 = \dfrac{\sum(y-\bar{y})^2}{n}$ – общая дисперсия признака y, обусловленная влиянием на y всех факторов, включая x.

Таким образом, получаем:

$$\eta = \sqrt{\frac{\sum(\tilde{y}-\bar{y})^2}{\sum(y-\bar{y})^2}}.$$

Теоретическое корреляционное отношение изменяется от 0 до 1. Близость к единице означает, что связь между признаками достаточно хорошо описывается полученным уравнением, следовательно модель считается адекватной фактическим данным.

Теоретический коэффициент детерминации η^2 (***индекс детерминации*** R^2) показывает, какая часть общей вариации расчетных значений признака y объясняется вариацией фактора x.

Если выполняется условие $\lvert \eta^2 - r^2 \rvert \le 0{,}1$, то зависимость между признаками можно считать прямолинейной. В этом случае теоретическое корреляционное отношение практически совпадает с линейным коэффициентом корреляции.

После проверки адекватности регрессионной модели необходимо провести экономическую интерпретацию параметров регрессии. Для интерпретации параметра a_1 используют ***коэффициент эластичности***, показывающий средние изменения результативного признака при изменении факторного признака на 1%:

$$K_э = a_1 \cdot \frac{\bar{x}}{\bar{y}}.$$

Пример 2. По исходным данным о товарообороте и издержках обращения 10 магазинов (Таблица 78) постройте линейную модель парной регрессии, определите тесноту связи с помощью линейного коэффициента корреляции и эмпирического корреляционного отношения. Оцените адекватность полученной модели и дайте экономическую интерпретацию.

Таблица 78

Расчетная таблица для построения линейной модели парной регрессии

№	Товарооборот, тыс. руб. (x)	Издержки обращения, тыс. руб. (y)	x^2	y^2	xy	$\tilde{y}$	$(\tilde{y}-\bar{y})^2$	$(y-\bar{y})^2$
1	480	30	230400	900	14400	26,2	36,00	4,84
2	510	25	260100	625	12750	28,0	17,64	51,84
3	530	31	280900	961	16430	29,2	9,00	1,44
4	540	28	291600	784	15120	29,9	5,29	17,64
5	570	29	324900	841	16530	31,7	0,25	10,24
6	590	32	348100	1024	18880	32,2	0,00	0,04
7	620	36	384400	1296	22320	34,7	6,25	14,44
8	640	36	409600	1296	23040	35,5	10,89	14,44
9	650	37	422500	1369	24050	36,5	18,49	23,04
10	660	38	435600	1444	25080	37,1	24,01	33,64
Σ	5790	322	3388100	10540	188600	321	127,82	171,6

Решение.

Из таблицы видно, что с ростом товарооборота в целом растут и издержки обращения.

Вычислим параметры уравнения регрессии:

$$a_1 = \frac{n\sum yx - \sum y \sum x}{n\sum x^2 - (\sum x)^2} = \frac{10 \cdot 188600 - 5790 \cdot 322}{10 \cdot 3388100 - 5790^2} = \frac{21620}{356900} = 0,061$$

$$a_0 = \bar{y} - a_1\bar{x} = 32,2 - 0,061 \cdot 579 = -2,887$$

Таким образом, уравнение регрессии имеет вид:

$$\tilde{y} = -2,887 + 0,061 \cdot x$$

При увеличении товарооборота на 1000 руб., издержки обращения увеличиваются на 61 р.

Рассчитаем по уравнению регрессии теоретические значения результативного признака $\tilde{y}$, линейный коэффициент корреляции и теоретическое корреляционное отношение:

$$r = \frac{\sum yx - \frac{\sum y \sum x}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}} = \frac{188600 - \frac{5790 \cdot 322}{10}}{\sqrt{\left(3388100 - \frac{5790^2}{10}\right)\left(10540 - \frac{322^2}{10}\right)}} =$$

$$= \frac{216,2}{247,4} = 0,876$$

$$\eta = \sqrt{\frac{\sum(\tilde{y} - \bar{y})^2}{\sum(y - \bar{y})^2}} = \sqrt{\frac{127,82}{171,6}} = 0,863$$

Полученные показатели согласно шкале Чэддока свидетельствуют о тесной связи между товарооборотом и издержками. Так как $r > 0$, $a > 0$, то связь прямая.

$\eta^2 = 0,745 \Rightarrow$ изменение издержек обращения на 74,5% зависит от изменения товарооборота и на 25,5% от других факторов.

$K_Э = a_1 \cdot \frac{\bar{x}}{\bar{y}} = 0,061 \cdot \frac{579}{32,2} = 1,1 \Rightarrow$ с увеличением товарооборота на 1% издержки возрастут на 1,1%.

6.3. Задачи для самостоятельного решения

1. Имеются данные по десяти небанковским кредитным организациям (Таблица 79):

Таблица 79

Организация	Собственный капитал, млн. руб.	Привлеченный капитал, млн. руб.
1	5	1,8
2	7	2,0
3	2	1,0
4	4	1,2
5	8	3,0
6	3	1,1
7	4	1,3
8	6	1,8
9	10	2,9
10	11	3,9

Для изучения связи между размером собственного капитала и привлеченным капиталом:

1) определите направление и форму связи с помощью метода параллельных рядов и графического метода;

2) рассчитайте линейный коэффициент корреляции, эмпирическое корреляционное отношение и коэффициент детерминации;

3) найдите параметры линейного уравнения регрессии;