

Задание 1.

По данным таблицы определить зависимость производительности труда (y) от фондоотдачи (x) предприятия «Рождественская звезда». Для этого

1) вычислить:

выборочные средние;

выборочную ковариацию между x и y ;

выборочную дисперсию для x и y ;

выборочный коэффициент корреляции между x и y ;

выборочный коэффициент линейной регрессии y на x и x на y ;

коэффициент детерминации;

среднюю ошибку аппроксимации;

средний коэффициент эластичности.

Решение:

В Excel составим вспомогательную таблицу :

	A	B	C	D	E	F
1	<i>N п.п.</i>	x	y	xy	x^2	y^2
2	1	1,08	7343	7930,4	1,1664	53919649
3	2	1,05	3993	4192,7	1,1025	15944049
4	3	0,99	5760	5702,4	0,9801	33177600
5	4	1,02	3000	3060	1,0404	9000000
6	5	0,98	5241	5136,2	0,9604	27468081
7	6	1,04	4500	4680	1,0816	20250000
8	7	1,03	4300	4429	1,0609	18490000
9	8	1,1	3213	3534,3	1,21	10323369
10	9	1,03	6743	6945,3	1,0609	45468049
11	10	0,89	5234	4658,3	0,7921	27394756
12	11	0,78	2500	1950	0,6084	6250000
13	12	0,99	3930	3890,7	0,9801	15444900
14	13	1,43	14333	20496	2,0449	205434889
15	14	1,03	6980	7189,4	1,0609	48720400
16	15	1,06	6740	7144,4	1,1236	45427600
17	сумма	15,5	83810	90939	16,2732	582713342
18	среднее	1,033	5587,33	6062,61	1,08	38847556,1
19	$n=$	15				

Вычислим количество измерений n . Для этого в ячейку **B19** поместим $=СЧЁТ(A2:A19)$.

С помощью функции Σ (Автосумма) на панели инструментов *Стандартная* найдем сумму всех x (ячейка **B17**) и y (ячейка **C17**).

Вычислим выборочные средние: $\bar{x}=1,033$; $\bar{y}=5587,33$. Таким образом, средняя фондоотдача по предприятию «Рождественская звезда» составит 1,033 руб., а средняя производительность труда – 5587,33 руб..

Для вычисления **выборочной ковариации** между x и y используем формулу $Cov(x, y) = \overline{xy} - \bar{y} \cdot \bar{x}$, т.е. в ячейку **B21** поместим **=D18-B18*C18** и получим 289,04 (рис. 1.1).

Выборочную дисперсию для x найдем по формуле $Var(x) = \overline{x^2} - \bar{x}^2$, т.е. в ячейку **B22** поместим **=E18-B18*B18** и получим 0,02 (рис. 1.1). Аналогично определяем $Var(y)=7629262,36$.

	A	B	C	D	E	F	G	H	I	J
1	N п.п.	x	y	xy	x ²	y ²	y [^]	y [^] -y	(y [^] -y)/y	(y [^] -y) ²
2	1	1,08	7343	7930,4	1,1664	53919649	6376,02	-966,98	0,13	935043,3
3	2	1,05	3993	4192,7	1,1025	15944049	5869,01	1876,01	0,47	3519407,7
4	3	0,99	5760	5702,4	0,9801	33177600	4854,98	-905,02	0,16	819064,7
5	4	1,02	3000	3060	1,0404	9000000	5361,99	2361,99	0,79	5579012,1
6	5	0,98	5241	5136,2	0,9604	27468081	4685,97	-555,03	0,11	308055,0
7	6	1,04	4500	4680	1,0816	20250000	5700,00	1200,00	0,27	1440008,1
8	7	1,03	4300	4429	1,0609	18490000	5531,00	1231,00	0,29	1515356,8
9	8	1,1	3213	3534,3	1,21	10323369	6714,03	3501,03	1,09	12257237,6
10	9	1,03	6743	6945,3	1,0609	45468049	5531,00	-1212,00	0,18	1468948,1
11	10	0,89	5234	4658,3	0,7921	27394756	3164,93	-2069,07	0,40	4281061,6
12	11	0,78	2500	1950	0,6084	6250000	1305,87	-1194,13	0,48	1425942,6
13	12	0,99	3930	3890,7	0,9801	15444900	4854,98	924,98	0,24	855584,4
14	13	1,43	14333	20496	2,0449	205434889	12291,20	-2041,80	0,14	4168943,1
15	14	1,03	6980	7189,4	1,0609	48720400	5531,00	-1449,00	0,21	2099605,9
16	15	1,06	6740	7144,4	1,1236	45427600	6038,01	-701,99	0,10	492785,0
17	сумма	15,5	83810	90939	16,2732	582713342	83810,00	0,00	5,04	41166055,9
18	среднее	1,033	5587,33	6062,61	1,08	38847556,1	5587,33	0,00	0,34	2744403,7
19	n=	15								
20			проверка							Результат выполнения функции ЛИНЕЙН
21	Cov(x,y)=	289,04	289,04		A _{ср} =	33,58%		16900,51	-11876,52	
22	Var(x)=	0,02	0,02		Э _{ср} =	3,126		3513,39	3659,46	
23	Var(y)=	7629262,36	7629262,36		S _{ост} =	1779,500		0,64	1779,50	
24	r _{xy} =	0,80	0,80		F _{табл} =	4,667		23,14	13,00	
25	b=	16900,51			F _{факт} =	23,139		73272879,4	41166055,9	
26	a=	-11876,52			y _p =	8206,912				
27	Var(y [^])=	4884858,63			t _{табл} =	2,160				
28	R ² =	0,64			Уравнение парной линейной регрессии: y [^] =-11876,52+16900,51x					
29	b(x,y)=	0,000038								

Рис. 1.1. Решение задания 1 в Excel

Выборочный коэффициент корреляции $0,7 < r_{xy} = 0,8 < 1$ указывает на тесную связь x и y , т.е. с увеличением фондоотдачи предприятия (x) увеличивается производительность труда (y).

Выборочный коэффициент линейной регрессии y на x :

$$b = b(y, x) = \frac{\text{Cov}(x, y)}{\text{Var}(x)} = 16900,51; \quad \text{параметр} \quad a = \bar{y} - b\bar{x} = -11876,52. \quad \text{Значит,}$$

уравнение парной линейной регрессии имеет вид $\hat{y} = 11876,52 + 16900,51 \cdot x$.

Коэффициент b показывает, что при увеличении фондоотдачи (x) на 1 руб. производительность труда (y) в среднем увеличится на 16900,51 руб..

Выборочный коэффициент линейной регрессии x на y

$$b(x, y) = \frac{\text{Cov}(x, y)}{\text{Var}(y)} = \frac{289,04}{7629262,36} = 0,000038.$$

Коэффициент $b(x, y)$ показывает, что при увеличении производительности труда (y) на 1 руб. фондоотдача (x) в среднем увеличится на 0,000038 руб.

Коэффициент детерминации. $R^2 = r_{xy}^2 = 0,64$.

Средняя ошибка аппроксимации.

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\%. \quad \bar{A} = 33,58\%.$$

Следовательно, в среднем теоретические значения \hat{y} отклоняются от фактических y на 33,58%.

Средний коэффициент эластичности $\bar{\varepsilon} = 3,126$ показывает, что с ростом размера фондоотдачи на 1% производительность труда возрастет на 3,126%.

2) на одном рисунке построить корреляционные поля и график уравнения линейной регрессии y на x .

Решение:

С помощью *Мастера диаграмм* строим корреляционные поля (выделяя столбцы со значениями x и y) и уравнение линейной регрессии (выделяя столбцы со значениями x и \hat{y}). Выбираем тип диаграммы – *Точечная* и, следуя рекомендациям *Мастера диаграмм*, вводим нужные параметры (название, подписи к осям, легенду и т.п.). В результате получим рис. 1.2.

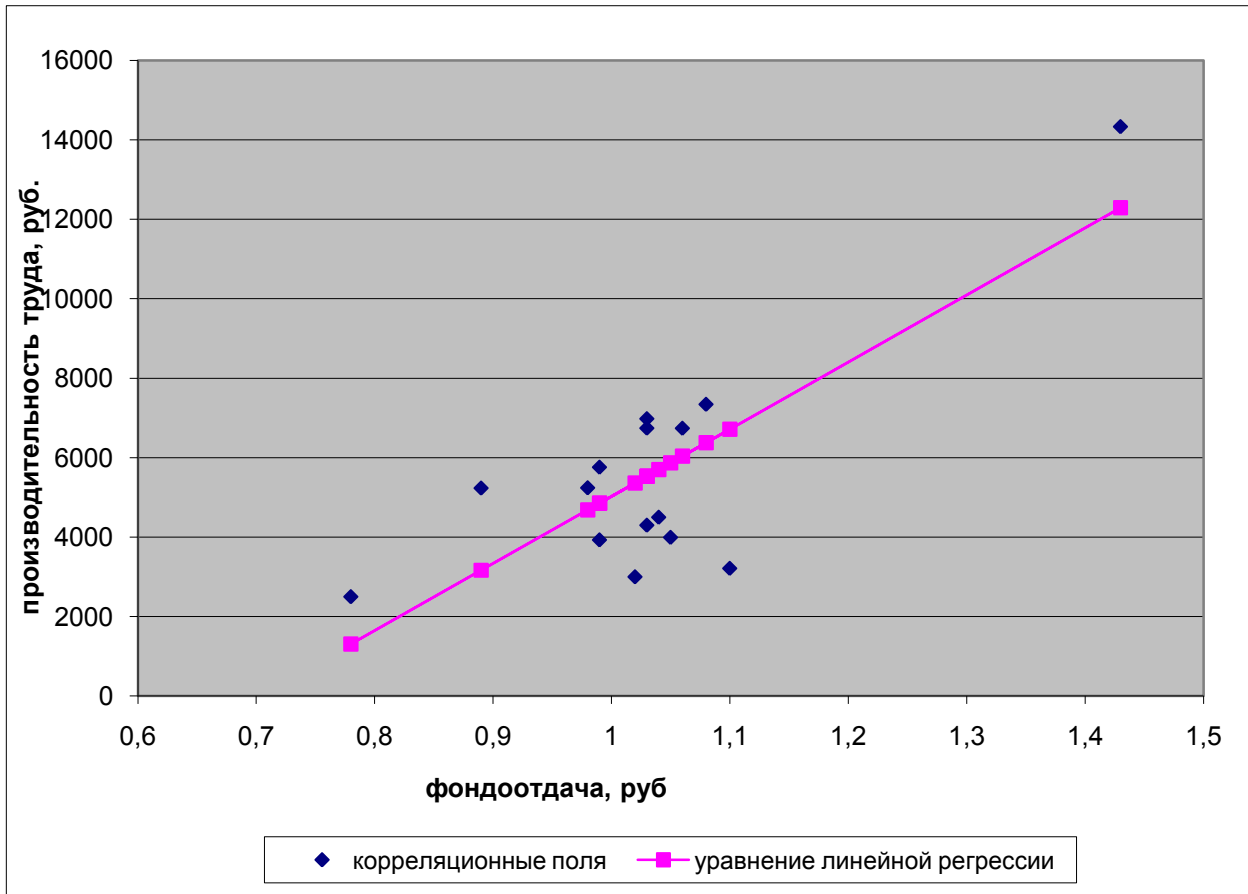


Рис. 1.2. График зависимости фондоотдачи от теоретических и фактических значений производительности труда.

3) сделать вывод о качестве модели и, если модель можно считать удачной, то рассчитать прогнозное значение y при увеличении значения фактора x на 15% от его среднего уровня.

Решение:

С помощью F -критерия Фишера оценим **значимость уравнения**

регрессии в целом: $F_{\text{факт}} = \frac{R^2}{1 - R^2} (n - 2) = 23,139.$

На уровне значимости 0,05 $F_{табл}=4,667$ определяем по таблице F -критерия Фишера (таблица 1 приложения).

Так как $F_{табл} < F_{факт}$, то уравнение регрессии значимо при $\alpha = 0,05$.

Коэффициент детерминации $R^2 = 0,64$ показывает, что изменение производительности труда (y) на 64% объясняется изменением фондоотдачи (x), а 36% приходится на долю неучтенных в модели факторов, что указывает на качество построенной регрессионной модели.

Рассчитаем **прогнозное значение** y_p путем подстановки в уравнение регрессии $\hat{y} = -11876,52 + 16900,51 \cdot x$ прогнозного значения фактора $x_p = \bar{x} \cdot 1,15 = 1,188$. Получим $y_p = 8206,912$. Следовательно, при фондоотдаче 1,188 руб. производительность труда составит 8206,912 руб.

Найдем $S_{ост} = \sigma_{ост} = 1779,5$ поместив в ячейку **F23 =КОРЕНЬ(J17/(B19-2))**.

Средняя стандартная ошибка прогноза:

$$m_{y_p} = 1779,5 \cdot \sqrt{1 + \frac{1}{15} + \frac{(1,033 - 1,188)^2}{15 \cdot 0,02}} = 1916,8437.$$

На уровне значимости $\alpha = 0,05$ либо по таблице t -критерия Стьюдента (таблица 2 приложения) либо с помощью встроенной статистической функции **СТЮДРАСПОБР** определим $t_{табл} = 2,16$ и вычислим предельную ошибку прогноза, которая в 95% случаев не будет превышать $\Delta_{y_p} = t_{табл} m_{y_p} = 4141,1$.

Доверительный интервал прогноза:

$$5587,33 - 4141,1 < \gamma_{y_p} < 5587,33 + 4141,1 \quad \text{или} \quad 1446,24 < \gamma_{y_p} < 7504,18.$$

Выполненный прогноз затрат на производительность труда оказался надежным ($1 - 0,05 = 0,95$), но неточным, так как диапазон верхней и нижней границ доверительного интервала составляет $D = \frac{7504,18}{1446,24} = 5,189$ раза. Это

произошло за счет малого объема наблюдений.

Задание 2.

На основании таблицы определить зависимость производительности труда (y) от фондоотдачи (x) и уровня рентабельности (p) предприятия «Рождественская звезда», т.е. рассчитать:

- 1) коэффициенты множественной линейной регрессии y на x_1 и x_2 .

Решение:

В Excel составим вспомогательную таблицу:

	A	B	C	D	E	F	G	H	I	J
1	i	x_1	x_2	y	yx_1	yx_2	x_1x_2	x_1^2	x_2^2	y^2
2	1	1,08	20,1	7343	7930,4	147594	21,7	1,17	404,0	53919649
3	2	1,05	12,9	3993	4192,7	51509,7	13,5	1,10	166,4	15944049
4	3	0,99	18	5760	5702,4	103680	17,8	0,98	324,0	33177600
5	4	1,02	11,7	3000	3060,0	35100	11,9	1,04	136,9	9000000
6	5	0,98	17,9	5241	5136,2	93813,9	17,5	0,96	320,4	27468081
7	6	1,04	16,8	4500	4680,0	75600	17,5	1,08	282,2	20250000
8	7	1,03	15,6	4300	4429,0	67080	16,1	1,06	243,4	18490000
9	8	1,1	14,3	3213	3534,3	45945,9	15,7	1,21	204,5	10323369
10	9	1,03	18,1	6743	6945,3	122048	18,6	1,06	327,6	45468049
11	10	0,89	17,6	5234	4658,3	92118,4	15,7	0,79	309,8	27394756
12	11	0,78	13	2500	1950,0	32500	10,1	0,61	169,0	6250000
13	12	0,99	14,2	3930	3890,7	55806	14,1	0,98	201,6	15444900
14	13	1,43	24,2	14333	20496,2	346859	34,6	2,04	585,6	205434889
15	14	1,03	20	6980	7189,4	139600	20,6	1,06	400,0	48720400
16	15	1,06	19,3	6740	7144,4	130082	20,5	1,12	372,5	45427600
17	Сумма	15,5	253,7	83810	90939,2	1539337	266,0	16,27	4448,0	582713342
18	Среднее	1,033	16,9	5587,3	6062,6	102622	17,7	1,08	296,5	38847556
19	n=	15								

Аналогично заданию 1 вычислим: $Cov(x_1, y)=289,036$; $Var(y)=7629262,4$; $Cov(x_2, y)=8122,042$; $Cov(x_1, x_2)=0,255$; $Var(x_1)=0,017$; $Var(x_2)=10,469$.

Затем найдем коэффициенты множественной линейной регрессии и оформим вывод результатов как на рис.2.1.

Например, для вычисления значения коэффициента b_1 в ячейку **F20** поместим формулу $=(B20*B24-B23*B22)/(B23*B24-B22*B22)$ и получим 8360,024. В ячейке **F21** найдем $b_2=571,842$ как $=(B21*B23-B20*B22)/(B23*B24-B22*B22)$. Коэффициент $a=-12723,1$ вычислим по формуле $=D18-B18*F20-$

C18*F21.

Уравнение множественной линейной регрессии примет вид:

$$\hat{y} = -12723,1 + 8360,024 \cdot x_1 + 571,842 \cdot x_2.$$

Таким образом, при увеличении фондоотдачи (x_1) на 1 руб. производительность труда (y) в среднем увеличится на 8360,024 руб., а при увеличении уровня рентабельности (x_2) на 1% производительность труда (y) увеличится в среднем на 571,842 руб.

	A	B	C	D	E	F	G	H	I	J	K
1	i	x_1	x_2	y	yx_1	yx_2	x_1x_2	x_1^2	x_2^2	y^2	
2	1	1,08	20,1	7343	7930,4	147594	21,7	1,17	404,0	53919649	
3	2	1,05	12,9	3993	4192,7	51509,7	13,5	1,10	166,4	15944049	
4	3	0,99	18	5760	5702,4	103680	17,8	0,98	324,0	33177600	
5	4	1,02	11,7	3000	3060,0	35100	11,9	1,04	136,9	9000000	
6	5	0,98	17,9	5241	5136,2	93813,9	17,5	0,96	320,4	27468081	
7	6	1,04	16,8	4500	4680,0	75600	17,5	1,08	282,2	20250000	
8	7	1,03	15,6	4300	4429,0	67080	16,1	1,06	243,4	18490000	
9	8	1,1	14,3	3213	3534,3	45945,9	15,7	1,21	204,5	10323369	
10	9	1,03	18,1	6743	6945,3	122048	18,6	1,06	327,6	45468049	
11	10	0,89	17,6	5234	4658,3	92118,4	15,7	0,79	309,8	27394756	
12	11	0,78	13	2500	1950,0	32500	10,1	0,61	169,0	6250000	
13	12	0,99	14,2	3930	3890,7	55806	14,1	0,98	201,6	15444900	
14	13	1,43	24,2	14333	20496,2	346859	34,6	2,04	585,6	205434889	
15	14	1,03	20	6980	7189,4	139600	20,6	1,06	400,0	48720400	
16	15	1,06	19,3	6740	7144,4	130082	20,5	1,12	372,5	45427600	
17	Сумма	15,5	253,7	83810	90939,2	1539337	266,0	16,27	4448,0	582713342	
18	Среднее	1,033	16,9	5587,3	6062,6	102622	17,7	1,08	296,5	38847556	
19	n=	15									
20	$Cov(x_1, y) =$	289,036	289,036		$b_1 =$	8360,024		$\Theta_{1cp} =$	1,546		
21	$Cov(x_2, y) =$	8122,042	8122,042		$b_2 =$	571,842		$\Theta_{2cp} =$	1,731		
22	$Cov(x_1, x_2) =$	0,255	0,255		$a =$	-12723,1					
23	$Var(x_1) =$	0,017	0,017		$r_{yx1 x2} =$	0,756		$F_{факт} =$	74,54	$F_{табл} =$	3,89
24	$Var(x_2) =$	10,469	10,469		$r_{yx2 x1} =$	0,890		$F_{x1} =$	16,04	$F_{x табл} =$	4,75
25	$Var(y) =$	7629262,4	7629262,4		$r_{x1x2 y} =$	-0,494		$F_{x2} =$	45,94		
26	$r_{yx1} =$	0,800	0,800		$R_{yx1x2} =$	0,962		$ t_{yx1} =$	4,810	$t_{табл} =$	2,16
27	$r_{yx2} =$	0,909	0,909		$R^2 =$	0,925		$ t_{yx2} =$	7,853		
28	$r_{x1x2} =$	0,604	0,604		$R^{\wedge 2} =$	0,913		$ t_{x1x2} =$	2,730		

Рис. 2.1. Решение примера 2.1 в Excel

2) парные коэффициенты корреляции, оценить их значимость на уровне 0,05 и пояснить их экономический смысл.

Решение:

Для вычисления **средних коэффициентов эластичности** воспользуемся формулой: $\bar{\varepsilon}_i = b_i \frac{\bar{x}_i}{\bar{y}}$. Вычисляем: $\bar{\varepsilon}_1 = 1,546$ и $\bar{\varepsilon}_2 = 1,731$. Т.е. увеличение только фондоотдачи (от своего среднего значения) или только уровня рентабельности на 1% увеличивает в среднем производительность труда на 1,546% или 1,731% соответственно. Таким образом, фактор x_2 оказывает большее влияние на результат, чем фактор x_1 .

Найдем значения **парных коэффициентов корреляции**:

$$r_{yx_1} = \frac{Cov(x_1, y)}{\sqrt{Var(x_1)Var(y)}} = 0,8; \quad r_{yx_2} = \frac{Cov(x_2, y)}{\sqrt{Var(x_2)Var(y)}} = 0,909;$$

$$r_{x_1x_2} = \frac{Cov(x_1, x_2)}{\sqrt{Var(x_1)Var(x_2)}} = 0,604.$$

Значения парных коэффициентов корреляции указывают на тесную связь y с x_2 и y с x_1 . В то же время межфакторная связь $r_{x_1x_2}$ средняя ($0,5 < r_{x_1x_2} = 0,604 < 0,7$), что говорит о том, что один из факторов является неинформативным, т.е. в модель необходимо включать или x_1 , или x_2 .

Значимость парных коэффициентов корреляции оценим с помощью t -критерия Стьюдента. $t_{табл} = 2,16$ определяем по таблице t -критерия Стьюдента.

Фактическое значение t -критерия Стьюдента для каждого парного

коэффициента определим по формулам: $t_{yx_1} = r_{yx_1} \sqrt{\frac{n-2}{1-r_{yx_1}^2}}$;

$t_{yx_2} = r_{yx_2} \sqrt{\frac{n-2}{1-r_{yx_2}^2}}$; $t_{x_1x_2} = r_{x_1x_2} \sqrt{\frac{n-2}{1-r_{x_1x_2}^2}}$. Получим $t_{yx_1} = 4,81$; $t_{yx_2} = 7,853$; $t_{x_1x_2}$

$= 2,730$.

Так как фактические значения t -статистики t_{yx_1} , t_{yx_2} , $t_{x_1x_2}$ превосходят табличное, то парные коэффициенты корреляции r_{yx_1} , r_{yx_2} , $r_{x_1x_2}$ неслучайно отличается от нуля и являются статистически значимыми.

3) частные коэффициенты корреляции и с их помощью оценить целесообразность включения факторов в уравнение регрессии.

Решение:

Вычислим **частные коэффициенты корреляции** по формулам:

$$r_{yx_1|x_2} = \frac{r_{yx_1} - r_{yx_2}r_{x_1x_2}}{\sqrt{(1-r_{yx_2}^2)(1-r_{x_1x_2}^2)}}; \quad r_{yx_2|x_1} = \frac{r_{yx_2} - r_{yx_1}r_{x_2x_1}}{\sqrt{(1-r_{yx_1}^2)(1-r_{x_1x_2}^2)}};$$
$$r_{x_1x_2|y} = \frac{r_{x_1x_2} - r_{x_1y}r_{x_2y}}{\sqrt{(1-r_{x_1y}^2)(1-r_{x_2y}^2)}}.$$

Получим $r_{yx_1|x_2}=0,756$; $r_{yx_2|x_1}=0,89$; $r_{x_1x_2|y}=-0,494$. Таким образом, фактор x_2 оказывает намного более сильное влияние на результат, чем x_1 .

Коэффициент корреляции $r_{yx_1|x_2}=0,756$ немногим отличается от соответствующего парного коэффициента $r_{yx_1}=0,8$, значит взаимосвязь признаков y и x_1 в некоторой степени обусловлена воздействием на них фиксируемой переменной x_2 .

Коэффициент корреляции $r_{yx_2|x_1}=0,89$ немногим отличается от соответствующего парного коэффициента $r_{yx_2}=0,909$, значит, взаимосвязь признаков y и x_2 в некоторой степени обусловлена воздействием на них фиксируемой переменной x_1 .

4) коэффициент множественной корреляции, множественный коэффициент детерминации, скорректированный коэффициент детерминации и

охарактеризовать степень совместного влияния факторов на результативный признак.

Решение:

Коэффициент множественной корреляции

$$R_{yx_1x_2} = \sqrt{1 - (1 - r_{yx_1}^2) \cdot (1 - r_{yx_2|x_1}^2)} = 0,962.$$

Следовательно, зависимость y от x_1 и x_2 характеризуется как средняя, в которой $R_{yx_1x_2}^2 = 92,5\%$ вариации производительности труда определяется вариацией учтенных в модели факторов: фондоотдачи и уровня рентабельности. Прочие факторы, не включенные в модель, составляют соответственно 7,5% от общей вариации y .

Скорректированный коэффициент множественной детерминации

$\hat{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-3} = 0,913$ указывает на тесную связь между результатом и признаками.

Оценим надежность уравнения регрессии в целом с помощью F -критерия Фишера. Вычислим $F_{факт} = \frac{R^2}{1 - R^2} \cdot \frac{(n-3)}{2} = 74,54$. $F_{табл} = 3,89$ определяем по таблице F -критерия Фишера (таблица 1 приложения) взяв $\alpha = 0,05$, $k_1 = 2$, $k_2 = 15 - 2 - 1 = 12$.

Так как фактическое значение больше табличного, то с вероятностью 95% делаем заключение о статистической значимости уравнения множественной линейной регрессии в целом.

Оценим целесообразность включения фактора x_1 после фактора x_2 и x_2 после x_1 с помощью частного F -критерия Фишера

$$F_{x_1} = \frac{R_{yx_1x_2}^2 - r_{yx_2}^2}{1 - R_{yx_1x_2}^2} \cdot (n-3) = 16,04; \quad F_{x_2} = \frac{R_{yx_1x_2}^2 - r_{yx_1}^2}{1 - R_{yx_1x_2}^2} \cdot (n-3) = 45,94.$$

$F_{табл}=4,75$ при $\alpha=0,05$, $k_1=1$, $k_2=12$. Так как $F_{x_1}=16,04 > F_{табл}=4,75$, а $F_{x_2}=45,94 > F_{табл}=4,75$, то включение факторов x_1, x_2 в модель статистически оправдано и коэффициенты чистой регрессии b_1, b_2 статистически значимы.