МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ЕЛЕЦКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМ. И. А. БУНИНА»

Н. Ю. Тимофеева

ЛАБОРАТОРНЫЙ ПРАКТИКУМ ПО ЭКОНОМЕТРИКЕ
(ПРОДВИНУТЫЙ УРОВЕНЬ)

Учебно-методическое пособие

Елец – 2017

УДК 330.4(075)
ББК 65в631я73
Т 41

Размещено на сайте по решению редакционно-издательского совета
Елецкого государственного университета имени И.А. Бунина
от 31.01.2017 г., протокол № 1

Рецензенты:

М. И. Шепелев, кандидат экономических наук, доцент
(Елецкий государственный университет им. И. А. Бунина),

Н. А. Фортунова, кандидат технически наук, доцент
(Елецкий государственный университет им. И. А. Бунина)

Н.Ю. Тимофеева

Т 41 Лабораторный практикум по эконометрике (продвинутый уровень). – Елец: Елецкий государственный университет им. И. А. Бунина, 2017. – 44 с.

Учебно-методическое пособие содержит комплекс лабораторных работ с многовариантными заданиями по эконометрике. Задания сопровождаются методическими комментариями и рекомендациями по решению средствами табличного редактора Microsoft Excel, а так же наглядными иллюстрациями и инструкциями по их выполнению.

Учебно-методическое пособие подготовлено для проведению лабораторных занятий в курсе «Эконометрика (продвинутый уровень)», а так же может быть использовано студентами экономических специальностей, аспирантами, преподавателями и сотрудниками экономических служб организаций.

УДК 330.4(075)
ББК 65в631я73

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Компьютерное моделирование экономических процессов становится не объемлемым элементом подготовки современного экономиста. Данное учебное пособие ориентировано на формирование у студентов навыков практического выполнения достаточно сложного комплекса расчетов по эконометрике и проведению с ними вычислительных экспериментов.

В учебно-методическом пособии рассмотрены вопросы построения экономико-математических моделей основных типов задач эконометрики и способы их решения средствами табличного редактора Microsoft Excel.

В пособие включены лабораторные задания по всем темам, предусмотренных рабочей программой курса «Эконометрика». Задания по каждой теме содержат справочную информацию по расчетным формулам и методам, используемым при выполнении заданий. Чтобы облегчить понимание и ускорить овладение учебным материалом, в начале каждой темы приведено подробное решение типового задания с соответствующим выводом результатов. Навыки, полученные при решении типового задания, закрепляются в процессе самостоятельной работы над выполнением лабораторных работ.

Задания лабораторного практикума могут выполняться как с использованием Excel, так и любого статистического или эконометрического пакета (STATISTICA, SPSS, STATS, STATGRAPHICS). Однако автор предусмотрел выполнение компьютерных типовых задач в среде табличного процессора Excel, как наиболее известной и доступной.

1. ЛИНЕЙНАЯ ПАРНАЯ РЕГРЕССИЯ

1.1. Методические рекомендации к решению типовых задач

Пример 1

По семи территориям Уральского района за 200Х г. известны значения двух признаков (таблица 1).

Таблица 1

Значения признаков по территориям Уральского района

Район	Расходы на покупку продовольственных товаров в общих расходах, %, *y*	Среднедневная заработная плата одного работающего, руб., *x*
Удмуртская республика	68,8	45,1
Свердловская обл.	61,2	59,0
Башкортостан	59,9	57,2
Челябинская обл.	56,7	61,8
Пермская обл.	55,0	58,8
Курганская обл.	54,3	47,2
Оренбургская обл.	49,3	55,2

Для характеристики зависимости *y* от *x* рассчитать параметры следующих функций: а) линейной; б) степенной; в) показательной; г) равносторонней гиперболы. Оценить каждую модель через среднюю ошибку аппроксимации А и F-критерий Фишера.

Решение

1а. Для расчета параметров a и b линейной регрессии $y = a + bx$ решаем систему нормальных уравнений относительно *a* и *b:*

$$\begin{cases} n \cdot a + b\sum x = \sum y \\ a\sum x + b\sum x^2 = \sum y \cdot x \end{cases} \quad (1)$$

По исходным данным рассчитаем нужные неизвестные: $\sum y$, $\sum x$, $\sum yx$, $\sum x^2$, $\sum y^2$. Данные расчет представлены на рисунке 1 и 2.

$$b = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\sigma_x^2} = \frac{3166{,}05 - 57{,}89 \cdot 54{,}9}{5{,}86^2} = -0{,}35.$$

$$a = \bar{y} - b \cdot \bar{x} = 57{,}89 - 0{,}35 \cdot 54{,}9 = 76{,}9$$

Уравнение регрессии: $\hat{y}(x)$= 76,88 - 0,35х. С увеличением среднедневной заработной платы на 1 руб. доля расходов на покупку продовольственных товаров снижается в среднем на 0,35 %. Рассчитаем линейный коэффициент парной корреляции:

$$r_{xy} = b \cdot \frac{\sigma(x)}{\sigma(y)} = -0{,}35 \cdot \frac{5{,}86}{5{,}74} = -0{,}354$$

Связь умеренная, обратная.

Определим коэффициент детерминации:

$$r^2_{xy} = (-0{,}354)^2 = 0{,}125$$

Вариация результата на 12,5% объясняется вариацией фактора х. Подставляя в уравнение регрессии фактические значения х, определим теоретические (расчетные) значения $\hat{y}(x)$. Найдем величину средней ошибки аппроксимации $\overline{Ai}$.

В среднем расчетные значения отклоняются от фактических на 8,1%.

Рассчитаем F-критерий:

$$F_{\text{факт}} = \frac{0{,}125}{0{,}873} \cdot 5 = 0{,}7$$

Поскольку 1 < F < ∞, следует рассмотреть F^{-1}

	A	B	C	D	E	F	G	H	I
1		y	x	yx	x²	y²	y^(x)	y-y^(x)	Ai
2	1	68,8	45,1	=B2*C2	=C2*C2	=B2*B2	=B$14+B$13*C2	=B2-G2	=ABS((B2-G2)/B2)*100
3	2	61,2	59	=B3*C3	=C3*C3	=B3*B3	=B$14+B$13*C3	=B3-G3	=ABS((B3-G3)/B3)*100
4	3	59,9	57,2	=B4*C4	=C4*C4	=B4*B4	=B$14+B$13*C4	=B4-G4	=ABS((B4-G4)/B4)*100
5	4	56,7	61,8	=B5*C5	=C5*C5	=B5*B5	=B$14+B$13*C5	=B5-G5	=ABS((B5-G5)/B5)*100
6	5	55	58,8	=B6*C6	=C6*C6	=B6*B6	=B$14+B$13*C6	=B6-G6	=ABS((B6-G6)/B6)*100
7	6	54,3	47,2	=B7*C7	=C7*C7	=B7*B7	=B$14+B$13*C7	=B7-G7	=ABS((B7-G7)/B7)*100
8	7	49,3	55,2	=B8*C8	=C8*C8	=B8*B8	=B$14+B$13*C8	=B8-G8	=ABS((B8-G8)/B8)*100
9	**итого**	=СУММ(B2:B8)	=СУММ(C2:C8)	=СУММ(D2:D8)	=СУММ(E2:E8)	=СУММ(F2:F8)	=СУММ(G2:G8)	=СУММ(H2:H8)	=СУММ(I2:I8)
10	**ср. знач**	=СРЗНАЧ(B2:B8)	=СРЗНАЧ(C2:C8)	=СРЗНАЧ(D2:D8)	=СРЗНАЧ(E2:E8)	=СРЗНАЧ(F2:F8)	x	x	=(I9*100%)/7
11	**δ**	=КОРЕНЬ(B12)	=КОРЕНЬ(C12)	x	x	x	x	x	x
12	**δ²**	=F10-B10*B10	=E10-C10*C10	x	x	x	x	x	x
13	**b**	=(D10-B10*C10)/C12							
14	**a**	=B10-B13*C10							
15	**r(xy)**	=B13*(C11/B11)							
16	**r²(xy)**	=B15*B15							
17	**Fфакт**	=(B16/(1-B16))*5							

Рис. 1. Фрагмент листа электронной таблицы Microsoft Excel с расчетными формулами примера 1а

	A	B	C	D	E	F	G	H	I
1		y	x	yx	x²	y²	y^(x)	y-y^(x)	Ai
2	1	68,8	45,1	3102,88	2034,01	4733,44	61,3	7,5	10,9
3	2	61,2	59,0	3610,80	3481,00	3745,44	56,5	4,7	7,7
4	3	59,9	57,2	3426,28	3271,84	3588,01	57,1	2,8	4,7
5	4	56,7	61,8	3504,06	3819,24	3214,89	55,5	1,2	2,1
6	5	55,0	58,8	3234,00	3457,44	3025,00	56,5	-1,5	2,8
7	6	54,3	47,2	2562,96	2227,84	2948,49	60,5	-6,2	11,5
8	7	49,3	55,2	2721,36	3047,04	2430,49	57,8	-8,5	17,2
9	итого	405,2	384,3	22162,34	21338,41	23685,76	405,2	0,0	57,0
10	ср. знач	57,89	54,90	3166,05	3048,34	3383,68	x	x	8,1
11	δ	5,74	5,86	x	x	x	x	x	x
12	δ²	32,92	34,33	x	x	x	x	x	x
13	b	-0,35							
14	a	76,9							
15	r(xy)	-0,353							
16	r²(xy)	0,125							
17	Fфакт	0,7							

Рис. 2. Фрагмент листа электронной таблицы Microsoft Excel с решением примера 1а

Полученное значение указывает на необходимость принять гипотезу о случайной природе выявленной зависимости и статистической незначимости параметров уравнения и показателя тесноты связи.

1б. Построению степенной модели $y = a \cdot x^b$ предшествует процедура линеаризации переменных. В примере линеаризация производится путем логарифмирования обеих частей уравнения:

$$lgy = lga + b \cdot 1gx$$

$$Y = C + b \cdot X,$$

где Y=lg y, X=lg x, C = lg *a*.

Расчет представлен на рисунке 3 и 4.

	A	B	C	D	E	F	G	H	I	J
1		y	x	yx	x²	y²	y^(x)	y-y^(x)	(y-y^(x))²	Ai
2	1	1,8376	1,6542	=B2*C2	=C2*C2	=B2*B2	=B$14+B$13*C2	=B2-G2	=H2*H2	=ABS((B2-G2)/B2)*100
3	2	1,7868	1,7709	=B3*C3	=C3*C3	=B3*B3	=B$14+B$13*C3	=B3-G3	=H3*H3	=ABS((B3-G3)/B3)*100
4	3	1,7774	1,7574	=B4*C4	=C4*C4	=B4*B4	=B$14+B$13*C4	=B4-G4	=H4*H4	=ABS((B4-G4)/B4)*100
5	4	1,7536	1,791	=B5*C5	=C5*C5	=B5*B5	=B$14+B$13*C5	=B5-G5	=H5*H5	=ABS((B5-G5)/B5)*100
6	5	1,7404	1,7694	=B6*C6	=C6*C6	=B6*B6	=B$14+B$13*C6	=B6-G6	=H6*H6	=ABS((B6-G6)/B6)*100
7	6	1,7348	1,6739	=B7*C7	=C7*C7	=B7*B7	=B$14+B$13*C7	=B7-G7	=H7*H7	=ABS((B7-G7)/B7)*100
8	7	1,6928	1,7419	=B8*C8	=C8*C8	=B8*B8	=B$14+B$13*C8	=B8-G8	=H8*H8	=ABS((B8-G8)/B8)*100
9	итого	=СУММ(B2:B8)	=СУММ(C2:C8)	=СУММ(D2:D8)	=СУММ(E2:E8)	=СУММ(F2:F8)	=СУММ(G2:G8)	=B9-G9	=H9*H9	=СУММ(J2:J8)
10	ср. знач	=СРЗНАЧ(B2:B8)	=СРЗНАЧ(C2:C8)	=СРЗНАЧ(D2:D8)	=СРЗНАЧ(E2:E8)	=СРЗНАЧ(F2:F8)	x	x		=(J9*100%)/7
11	δ	=КОРЕНЬ(B12)	=КОРЕНЬ(C12)	x	x	x	x	x		x
12	δ²	=F10-B10*B10	=E10-C10*C10	x	x	x	x	x		x
13	b	=(D10-B10*C10)/C12								
14	C	=B10-B13*C10								

Рис. 3. Фрагмент листа электронной таблицы Microsoft Excel с расчетными формулами примера 1б

	A	B	C	D	E	F	G	H	I	J
1		y	x	yx	x²	y²	y^(x)	y-y^(x)	(y-y^(x))²	Ai
2	1	1,8376	1,6542	3,0398	2,7364	3,3768	61,0000	7,8000	60,8400	11,30
3	2	1,7868	1,7709	3,1642	3,1361	3,1927	56,3000	4,9000	24,0100	8,00
4	3	1,7774	1,7574	3,1236	3,0885	3,1592	56,8000	3,1000	9,6100	5,20
5	4	1,7536	1,7910	3,1407	3,2077	3,0751	55,5000	1,2000	1,4400	2,10
6	5	1,7404	1,7694	3,0795	3,1308	3,0290	56,3000	-1,3000	1,6900	2,40
7	6	1,7348	1,6739	2,9039	2,8019	3,0095	60,2000	-5,9000	34,8100	10,90
8	7	1,6928	1,7419	2,9487	3,0342	2,8656	57,4000	-8,1000	65,6100	16,40
9	итого	12,3234	12,1587	21,4003	21,1355	21,7078	403,5000	1,7000	2,8900	53,60
10	ср. знач	1,7605	1,7370	3,0572	3,0194	3,1011	x	x	28,2871	8,00
11	ծ	0,0425	0,0484	x	x	x	x	x		x
12	ծ²	0,0018	0,0023	x	x	x	x	x		x
13	b	-0,2978								
14	C	2,2777								

Рис. 4. Фрагмент листа электронной таблицы Microsoft Excel с решением примера 1б

Рассчитаем *C* и *b:*

$$b = \frac{\overline{y\cdot x} - \bar{y}\cdot\bar{x}}{\sigma_{\bar{x}}^2} = \frac{3{,}06 - 1{,}76\cdot 1{,}74}{0{,}05^2} = -0{,}298,$$

$$C = \bar{y} - b\cdot\bar{x} = 1{,}7605 + 0{,}298\cdot 1{,}7170 = 2{,}278.$$

Получим линейное уравнение: $Y = 2{,}278 - 0{,}298\cdot X$.

Выполнив его потенцирование, получим:

$$\hat{y} = 10^{2{,}278}\cdot x^{-0{,}298} = 189{,}7\cdot x^{-0{,}298}$$

Подставляя в данное уравнение фактические значения x, получаем теоретические значения результата $\hat{y}(x)$. По ним рассчитаем показатели: тесноты связи - индекс корреляции ρ_{xy} и среднюю ошибку аппроксимации $\overline{A}i$:

$$\rho_{xy} = \sqrt{1 - \frac{\Sigma(y - \hat{y}(x))^2}{\Sigma(y - \bar{y})^2}} = \sqrt{1 - \frac{28{,}27}{32{,}92}} = 0{,}3758$$

$$\overline{A} = 8{,}0\%$$

Характеристики степенной модели указывают, что она несколько лучше линейной функции описывает взаимосвязь.

1в. Построению уравнения показательной кривой $y = a\cdot b^x$ предшествует процедура линеаризации переменных при логарифмировании обеих частей уравнения:

$$lg\, y = lg\, a + x\cdot lg\, b$$

$$Y = C + B\cdot x,$$

где Y = lg y, C = lg a, B = lg b.

Расчет представлен на рисунке 5 и 6.

	A	B	C	D	E	F	G	H	I	J
1		y	x	yx	x^2	y^2	y^(x)	y-y^(x)	$(y-y\hat{}(x))^2$	Ai
2	1	1,8376	45,1	=B2*C2	=C2*C2	=B2*B2	=$B$14+$B$13*C2	=B2-G2	=H2*H2	=ABS((B2-G2)/B2)*100
3	2	1,7868	59	=B3*C3	=C3*C3	=B3*B3	=$B$14+$B$13*C3	=B3-G3	=H3*H3	=ABS((B3-G3)/B3)*100
4	3	1,7774	57,2	=B4*C4	=C4*C4	=B4*B4	=$B$14+$B$13*C4	=B4-G4	=H4*H4	=ABS((B4-G4)/B4)*100
5	4	1,7536	61,8	=B5*C5	=C5*C5	=B5*B5	=$B$14+$B$13*C5	=B5-G5	=H5*H5	=ABS((B5-G5)/B5)*100
6	5	1,7404	58,8	=B6*C6	=C6*C6	=B6*B6	=$B$14+$B$13*C6	=B6-G6	=H6*H6	=ABS((B6-G6)/B6)*100
7	6	1,7348	47,2	=B7*C7	=C7*C7	=B7*B7	=$B$14+$B$13*C7	=B7-G7	=H7*H7	=ABS((B7-G7)/B7)*100
8	7	1,6928	55,2	=B8*C8	=C8*C8	=B8*B8	=$B$14+$B$13*C8	=B8-G8	=H8*H8	=ABS((B8-G8)/B8)*100
9	итого	=СУММ(B2:B8)	=СУММ(C2:C8)	=СУММ(D2:D8)	=СУММ(E2:E8)	=СУММ(F2:F8)	=$B$14+$B$13*C9	=B9-G9	=H9*H9	=СУММ(J2:J8)
10	ср. знач	=СРЗНАЧ(B2:B8)	=СРЗНАЧ(C2:C8)	=СРЗНАЧ(D2:D8)	=СРЗНАЧ(E2:E8)	=СРЗНАЧ(F2:F8)	x	x		=(J9*100%)/7
11	δ	=КОРЕНЬ(B12)	=КОРЕНЬ(C12)	x	x	x	x	x		x
12	$δ^2$	=F10-B10*B10	=E10-C10*C10	x	x	x	x	x		x
13	b	=(D10-B10*C10)/C12								
14	a	=B10-B13*C10								

Рис. 5. Фрагмент листа электронной таблицы Microsoft Excel с расчетными формулами примера 1в

	A	B	C	D	E	F	G	H	I	J
1		y	x	yx	x^2	y^2	y^(x)	y-y^(x)	$(y-y\hat{}(x))^2$	Ai
2	1	1,8376	45,1000	82,8758	2034,0100	3,3768	60,7000	8,1000	65,6100	11,80
3	2	1,7868	59,0000	105,4212	3481,0000	3,1927	56,4000	4,8000	23,0400	7,80
4	3	1,7774	57,2000	101,6673	3271,8400	3,1592	56,9000	3,0000	9,0000	5,00
5	4	1,7536	61,8000	108,3725	3819,2400	3,0751	55,5000	1,2000	1,4400	2,10
6	5	1,7404	58,8000	102,3355	3457,4400	3,0290	56,4000	-1,4000	1,9600	2,50
7	6	1,7348	47,2000	81,8826	2227,8400	3,0095	50,0000	-5,7000	32,4900	10,50
8	7	1,6928	55,2000	93,4426	3047,0400	2,8656	57,5000	-8,2000	67,2400	16,60
9	итого	12,3234	384,3000	675,9974	21338,4100	21,7078	403,4000	-1,8000	200,7800	56,30
10	ср. знач	1,7605	54,9000	96,5711	3048,3443	3,1011	x	x	28,6800	8,00
11	δ	0,0425	5,8595	x	x	x	x	x		x
12	$δ^2$	0,0018	34,3343	x	x	x	x	x		x
13	b	-0,0023								
14	a	1,8878								

Рис. 6. Фрагмент листа электронной таблицы Microsoft Excel с решением примера 1в

$$B = \frac{\overline{Y \cdot x} - \bar{Y} \cdot \bar{x}}{\sigma_x^2} = \frac{96{,}5711 - 1{,}7605 \cdot 54{,}9}{5{,}86^2} = -0{,}0023$$

$$A=\bar{Y}\text{-}B\cdot\bar{x}= 1{,}7605+0{,}0023\cdot 54{,}9=1{,}887.$$

Получено линейное уравнение: $\hat{Y}=1{,}887-0{,}0023\cdot x$.

Произведем потенцирование полученного уравнения и запишем его в обычной форме:

$\hat{y} = 10^{1{,}887} \cdot 10^{-0{,}0023x} = 77{,}1 \cdot 0{,}9947^x.$

Тесноту связи оценим через индекс корреляции ρ_{xy}:

$$\rho_{xy} = \sqrt{1 - \frac{\sum(y - \hat{y}(x))^2}{\sum(y - \bar{y})^2}} = \sqrt{1 - \frac{28{,}27}{32{,}92}} = 0{,}375$$

Связь умеренная. A = 8,0%, что говорит о повышенной ошибке аппроксимации, но в допустимых пределах. Показательная функция чуть хуже, чем степенная, она описывает изучаемую зависимость.

1г. Уравнение равносторонней гиперболы $y = a + b \cdot \frac{1}{x}$ линеаризуется при замене: $z = \frac{1}{x}$. Тогда $y = a + b \cdot z$. Для расчетов используем данные рис. 7 и 8.

	A	B	C	D	E	F	G	H	I	J
1		y	z	yz	z²	y²	y^(x)	y-y^(x)	(y-y^(x))²	Ai
2	1	68,8	0,0222	=B2*C2	=C2*C2	=B2*B2	=$B$14+$B$13*C2	=B2-G2	=H2*H2	=ABS((B2-G2)/B2)*100
3	2	61,2	0,0169	=B3*C3	=C3*C3	=B3*B3	=$B$14+$B$13*C3	=B3-G3	=H3*H3	=ABS((B3-G3)/B3)*100
4	3	59,9	0,0175	=B4*C4	=C4*C4	=B4*B4	=$B$14+$B$13*C4	=B4-G4	=H4*H4	=ABS((B4-G4)/B4)*100
5	4	56,7	0,0162	=B5*C5	=C5*C5	=B5*B5	=$B$14+$B$13*C5	=B5-G5	=H5*H5	=ABS((B5-G5)/B5)*100
6	5	55	0,017	=B6*C6	=C6*C6	=B6*B6	=$B$14+$B$13*C6	=B6-G6	=H6*H6	=ABS((B6-G6)/B6)*100
7	6	54,3	0,0212	=B7*C7	=C7*C7	=B7*B7	=$B$14+$B$13*C7	=B7-G7	=H7*H7	=ABS((B7-G7)/B7)*100
8	7	49,3	0,0181	=B8*C8	=C8*C8	=B8*B8	=$B$14+$B$13*C8	=B8-G8	=H8*H8	=ABS((B8-G8)/B8)*100
9	итого	=СУММ(B2:B8)	=СУММ(C2:C8)	=СУММ(D2:D8)	=СУММ(E2:E8)	=СУММ(F2:F8)	=СУММ(G2:G8)	=СУММ(H2:H8)	=СУММ(I2:I8)	=СУММ(J2:J8)
10	ср. знач	=СРЗНАЧ(B2:B8)	=СРЗНАЧ(C2:C8)	=СРЗНАЧ(D2:D8)	=СРЗНАЧ(E2:E8)	=СРЗНАЧ(F2:F8)	x	x	=СРЗНАЧ(I2:I8)	=(J9*100%)/7
11	δ	=КОРЕНЬ(B12)	=КОРЕНЬ(C12)	x	x	x	x	x		x
12	δ²	=F10-B10*B10	=E10-C10*C10	x	x	x	x	x		x
13	b	=(D10-B10*C10)/C12								
14	a	=B10-B13*C10								

Рис. 7. Фрагмент листа электронной таблицы Microsoft Excel с расчетными формулами примера 1г

	A	B	C	D	E	F	G	H	I	J
1		y	z	yz	z²	y²	y^(x)	y-y^(x)	(y-y^(x))²	Ai
2	1	68,8	0,0222	1,5274	0,0005	4733,4400	61,8	7,0	48,50	10,1
3	2	61,2	0,0169	1,0343	0,0003	3745,4400	56,3	4,9	24,37	8,1
4	3	59,9	0,0175	1,0483	0,0003	3588,0100	56,9	3,0	9,03	5,0
5	4	56,7	0,0162	0,9185	0,0003	3214,8900	55,5	1,2	1,37	2,1
6	5	55,0	0,0170	0,9350	0,0003	3025,0000	56,4	-1,4	1,87	2,5
7	6	54,3	0,0212	1,1512	0,0004	2948,4900	60,8	-6,5	42,05	11,9
8	7	49,3	0,0181	0,8923	0,0003	2430,4900	57,5	-8,2	67,65	16,7
9	итого	405,2	0,1291	7,5069	0,0024	23685,7600	405,2	0,0	194,85	56,4
10	ср. знач	57,9	0,0184	1,0724	0,0003	3383,6800	x	x	27,84	8,1
11	δ	5,7380	0,0021	x	x	x	x	x		x
12	δ²	32,9241	0,0000	x	x	x	x	x		x
13	b	1051,4367								
14	a	38,4942								

Рис. 8. Фрагмент листа электронной таблицы Microsoft Excel с решением примера 1г

Значение параметров регрессии *a* и *b* составили:

$$a = \bar{y} - b \cdot \bar{z} = 57,89 - 1051,4 \cdot 0,0184 = 38,5;$$

$$b=\frac{\overline{y\cdot z}-\bar{y}\cdot\bar{z}}{\sigma_z^2}=\frac{1{,}0723-57{,}9\cdot 0{,}0184}{0{,}002145^2}=1051{,}4.$$

Получено уравнение: $\hat{y}=38{,}5+1051{,}4\cdot\frac{1}{x}$.

Индекс корреляции: $\rho_{xy}=\sqrt{1-\frac{27{,}84}{32{,}92}}=0{,}3944$.

$\bar{A}=8{,}1\%$. По уравнению равносторонней гиперболы получено наибольшая оценка тесноты связи: $\rho_{xy}=0{,}3944$ (по сравнению с линейной, степенной и показательной регрессиями). $\bar{A}$ остается на допустимом уровне:

$$F_{\text{факт.}}=\frac{P_{xy}^2}{1-P_{xy}^2}\cdot\frac{n-m-1}{m}=\frac{0{,}1555}{0{,}8445}\cdot 5=0{,}92,$$

где $F_{\text{табл.}}=6{,}6>F_{\text{факт.}}$, $a=0{,}05$.

Следовательно, принимается гипотеза H_0 о статистически незначимых параметрах этого уравнения. Этот результат можно объяснить сравнительно невысокой теснотой выявленной зависимости и небольшим числом наблюдений.

Пример 2

По территориям региона приводятся данные за 200X г. (таблица 2).

Таблица 2

Значения признаков по территориям

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., *x*	Среднедневная заработная плата, руб., *y*
1	78	133
2	82	148
3	87	134
4	79	154
5	89	162
6	106	195
7	67	139
8	88	158
9	73	152
10	87	162
11	76	159
12	115	173

1. Построить линейное уравнение парной регрессии *y* от *x*.

2. Рассчитать линейный коэффициент парной корреляции и среднюю ошибку аппроксимации.

3. Оценить статистическую значимость параметров регрессии и корреляции.

4. Выполнить прогноз заработной платы *y* при прогнозном значении среднедушевого прожиточного минимумах, составляющем 107% от среднего уровня.

5. Оценить точность прогноза, рассчитав ошибку прогноза и его доверительный интервал.

Решение

1. Расчет параметров уравнения линейной регрессии представлен на рисунке 9 и 10.

	A	B	C	D	E	F	G	H	I
1		y	x	xy	x²	y²	y^	y-y^(x)	Ai
2	**1**	133	78	=B2*C2	=C2*C2	=B2*B2	=B$19+B$18*C2	=B2-G2	=ABS((B2-G2)/B2)*100
3	**2**	148	82	=B3*C3	=C3*C3	=B3*B3	=B$19+B$18*C3	=B3-G3	=ABS((B3-G3)/B3)*100
4	**3**	134	87	=B4*C4	=C4*C4	=B4*B4	=B$19+B$18*C4	=B4-G4	=ABS((B4-G4)/B4)*100
5	**4**	154	79	=B5*C5	=C5*C5	=B5*B5	=B$19+B$18*C5	=B5-G5	=ABS((B5-G5)/B5)*100
6	**5**	162	89	=B6*C6	=C6*C6	=B6*B6	=B$19+B$18*C6	=B6-G6	=ABS((B6-G6)/B6)*100
7	**6**	195	106	=B7*C7	=C7*C7	=B7*B7	=B$19+B$18*C7	=B7-G7	=ABS((B7-G7)/B7)*100
8	**7**	139	67	=B8*C8	=C8*C8	=B8*B8	=B$19+B$18*C8	=B8-G8	=ABS((B8-G8)/B8)*100
9	**8**	158	88	=B9*C9	=C9*C9	=B9*B9	=B$19+B$18*C9	=B9-G9	=ABS((B9-G9)/B9)*100
10	**9**	152	73	=B10*C10	=C10*C10	=B10*B10	=B$19+B$18*C10	=B10-G10	=ABS((B10-G10)/B10)*100
11	**10**	162	87	=B11*C11	=C11*C11	=B11*B11	=B$19+B$18*C11	=B11-G11	=ABS((B11-G11)/B11)*100
12	**11**	159	76	=B12*C12	=C12*C12	=B12*B12	=B$19+B$18*C12	=B12-G12	=ABS((B12-G12)/B12)*100
13	**12**	173	115	=B13*C13	=C13*C13	=B13*B13	=B$19+B$18*C13	=B13-G13	=ABS((B13-G13)/B13)*100
14	**Итого**	=СУММ(B2:B13)	=СУММ(C2:C13)	=СУММ(D2:D13)	=СУММ(E2:E13)	=СУММ(F2:F13)	=СУММ(G2:G13)	=СУММ(H2:H13)	=СУММ(I2:I13)
15	**Средн. Знач.**	=СРЗНАЧ(B2:B13)	=СРЗНАЧ(C2:C13)	=СРЗНАЧ(D2:D13)	=СРЗНАЧ(E2:E13)	=СРЗНАЧ(F2:F13)	x		=СРЗНАЧ(I2:I13)
16	**σ**	=КОРЕНЬ(B17)	=КОРЕНЬ(C17)	x	x	x	x		
17	**σ²**	=F15-B15*B15	=E15-C15*C15	x	x	x	x		
18	**б**	=(D15-B15*C15)/C17							
19	**a**	=B15-B18*C15							
20	**r(xy)**	=B18*(C16/B16)							
21	**r²(xy)**	=B20*B20							

Рис. 9. Фрагмент листа электронной таблицы Microsoft Excel с расчетными формулами примера 2

	A	B	C	D	E	F	G	H	I
1		**y**	**x**	**xy**	**x²**	**y²**	**y^**	**y-y^(x)**	**Ai**
2	**1**	133	78	10374,0	6084,0	17689,0	148,8	-15,8	11,8571942
3	**2**	148	82	12136,0	6724,0	21904,0	152,5	-4,5	3,00796657
4	**3**	134	87	11658,0	7569,0	17956,0	157,1	-23,1	17,2044353
5	**4**	154	79	12166,0	6241,0	23716,0	149,7	4,3	2,79837736
6	**5**	162	89	14418,0	7921,0	26244,0	158,9	3,1	1,91678741
7	**6**	195	106	20670,0	11236,0	38025,0	174,5	20,5	10,4912186
8	**7**	139	67	9313,0	4489,0	19321,0	138,6	0,4	0,25515667
9	**8**	158	88	13904,0	7744,0	24964,0	158,0	0,0	0,01621909
10	**9**	152	73	11096,0	5329,0	23104,0	144,2	7,8	5,15268714
11	**10**	162	87	14094,0	7569,0	26244,0	157,1	4,9	3,05312143
12	**11**	159	76	12084,0	5776,0	25281,0	146,9	12,1	7,59169358
13	**12**	173	115	19895,0	13225,0	29929,0	182,8	-9,8	5,67976807
14	**Итого**	1869,0	1027,0	161808,0	89907,0	294377,0	1869,0	0,0	69,0246254
15	**Средн. З**	155,8	85,6	13484,0	7492,3	24531,4	x		5,75205212
16	**σ**	16,5	13,0	x	x	x	x		
17	**σ²**	273,35	167,7	x	x	x	x		
18	**б**	0,920							
19	**a**	76,976							
20	**r(xy)**	0,721							
21	**r²(xy)**	0,520							

Рис.10. Фрагмент листа электронной таблицы Microsoft Excel с решением примера 2

$$b = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\sum x^2 - (\bar{x})^2} = \frac{13484 - 85{,}6 \cdot 155{,}8}{7492{,}3 - 85{,}6^2} = \frac{151{,}8}{164{,}94} = 0{,}92,$$

$$a = \bar{y} - b \cdot \bar{x} = 155{,}8 - 0{,}92 \cdot 85{,}6 = 77{,}0.$$

Получено уравнение регрессии: $y = 77{,}0 + 0{,}92 \cdot x$.

С увеличением среднедушевого прожиточного минимума на 1 руб. среднедневная заработная плата возрастает в среднем на 0,92 руб.

2. Тесноту линейной связи оценит коэффициент корреляции:

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = 0{,}92 \cdot \frac{12{,}95}{16{,}53} = 0{,}721, \qquad r_{xy}^2 = 0{,}52.$$

Это означает, что 52% вариации заработной платы (y) объясняется вариацией фактора x - среднедушевого прожиточного минимума. Качество модели определяет средняя ошибка аппроксимации:

$$\bar{A} = \frac{1}{n} \sum A_i = \frac{68{,}9}{12} = 5{,}7\%.$$

Качество построенной модели оценивается как хорошее, так как $\bar{A}$ не превышает 8 - 10%.

3. Оценку статистической значимости параметров регрессии проведем с помощью t-статистики Стьюдента и путем расчета доверительного интервала каждого из показателей.

Выдвигаем гипотезу H_0 о статистически незначимом отличии показателей от нуля: $a = b = r_{xy} = 0$.

$t_{\text{табл.}}$ для числа степеней свободы $df = n - 2 = 12 - 2 = 10$ и $\alpha = 0{,}05$ составит 2,23.

Определим случайные ошибки $m_a, m_b, m_{r(xy)}$:

$$m_a = 12{,}6 \frac{\sqrt{89907}}{12 \cdot 12{,}95} = 24{,}3;$$

$$m_b = \frac{12{,}6}{12{,}95 \cdot \sqrt{12}} = 0{,}281;$$

$$m_{r(xy)} = \sqrt{\frac{1 - 0{,}520}{12 - 2}} = 0{,}219.$$

Тогда

$$t_a = \frac{77}{24{,}3} = 3{,}2;$$

$$t_b = \frac{0{,}92}{0{,}281} = 3{,}3;$$

$$t_{r(xy)} = \frac{0{,}721}{0{,}219} = 3{,}3.$$

Фактические значения t-статистики превосходят табличные значения:

$$t_a = 3{,}2 > t_{табл.} = 2{,}3; t_b = 3{,}3 > t_{табл.} = 2{,}3; t_{r(xy)} = 3{,}3 > t_{табл.} = 2{,}3,$$

поэтому гипотеза H_0 отклоняется, т. е. a, b и r_{xy} не случайно отличаются от нуля, а статистически значимы.

Рассчитаем доверительный интервал для a и b. Для этого определим предельную ошибку для каждого показателя:

$\Delta_a = 2{,}23 \cdot 24{,}3 = 54;$

$\Delta_b = 2{,}23 \cdot 0{,}281 = 0{,}62.$

Доверительные интервалы:

$\gamma_a = a \pm \Delta_a = 77 \pm 54;$

$\gamma_{a\ min} = 77 - 54 = 23;$

$\gamma_{a\ max} = 77 + 54 = 131;$

$\gamma_b = b \pm \Delta_b = 0{,}92 \pm 0{,}62;$

$\gamma_{b\ min} = 0{,}92 - 0{,}62 = 0{,}30;$

$\gamma_{b\ max} = 0{,}92 + 0{,}62 = 1{,}54.$

Анализ верхней и нижней границ доверительных интервалов приводит к выводу о том, что с вероятностью $p = 1 - \alpha = 0{,}95$ параметры a и b, находясь в указанных границах, не принимают нулевых значений, т.е. не являются статистически незначимыми и существенно отличны от нуля.

4. Полученные оценки уравнения регрессии позволяют использовать его для прогноза. Если прогнозное значение прожиточного минимума составит: $x_p = \bar{x} \cdot 1{,}07 = 85{,}6 \cdot 1{,}07 = 91{,}6$ тыс. руб., тогда прогнозное значение прожиточного минимума составит:

$$\hat{y}_p = 77 + 0{,}92 \cdot 91{,}6 = 161 \text{ тыс. руб.}$$

5. Ошибка прогноза составит:

$$m_{\hat{y}(p)} = 12{,}6 \cdot \sqrt{1 + \frac{1}{12} + \frac{(91{,}6 - 85{,}6)^2}{12 \cdot 12{,}95^2}} = 13{,}2 \text{ тыс. руб.}$$

Предельная ошибка прогноза, которая в 95% случаев не будет превышена, составит:

$\Delta_{\hat{y}(p)} = t_{табл.} \cdot m_{\hat{y}(p)} = 2{,}23 \cdot 13{,}2 = 29{,}4.$

Доверительный интервал прогноза:

$\gamma_{\hat{y}(p)} = 91{,}6 \pm 29{,}4;$

$$\gamma_{\hat{y}(p)min} = 91{,}6 - 29{,}4 = 62{,}2 \text{ руб.};$$

$$\gamma_{\hat{y}(p)max} = 91{,}6 + 29{,}4 = 121 \text{руб.}$$

Выполненный прогноз среднемесячной заработной платы оказался надежным (р = 1 - α = 1 - 0,05 = 0,95), но неточным, так как диапазон верхней и нижней границ доверительного интервала D_γ составляет 1,95 раза:

$$D_\gamma = \frac{\gamma_{\hat{y}max}}{\gamma_{\hat{y}min}} = \frac{121}{62} = 1{,}95.$$

Пример 3

По группе предприятий, производящих однородную продукцию, известно, как зависит себестоимость единицы продукции *y* от факторов, приведенных в таблице 3.

Таблица 3

Зависимость себестоимость единицы продукции *y* от факторов

Признак-фактор	Уравнение парной регрессии	Среднее значение фактора
Объем производства, млн. руб., x_1	$\hat{y}(x_1) = 0{,}62 + 58{,}74 \cdot \frac{1}{x1}$	$\overline{x_1}$=2,64
Трудоемкость единицы прод., чел./час, x_2	$\hat{y}\ (x_2)$= 9,30+9,83·x_2	$\overline{x_2}$ = 1,38
Оптовая цена за 1 т энергоносителя, млн руб., x_3	$\hat{y}(x_3) = 11{,}75 + x_3^{1{,}6281}$	$\overline{x_3}$= 1,503
Доля прибыли, изымаемой государством, %, x_4	$\hat{y}(x_4) = 14{,}87 \cdot 1{,}016^{x_4}$	$\overline{x_4}$ = 26,3

1) Определить с помощью коэффициентов эластичности силу влияния каждого фактора на результат;

2) Ранжировать факторы по силе влияния.

Решение

1. Для уравнения равносторонней гиперболы

$$\hat{y}(x_1) = 0{,}62 + 58{,}74 \cdot \frac{1}{x_1}$$

$$\bar{Э}(yx_1) = f'(x_1)\frac{\overline{x_1}}{\overline{y}} = -\frac{b}{\bar{x}_1^2} \cdot \frac{\overline{x_1}}{a + b/\overline{x_1}} = -\frac{b}{b + a \cdot \overline{x_1}} = -\frac{58{,}74}{0{,}62 \cdot 2{,}64 + 58{,}74} = -0{,}973\%$$

Для уравнения прямой $\hat{y}\,(x2)$= 9,30+9,83·x_2:

$$\overline{Э}(yx_2) = f^{`}(x_2)\frac{\overline{x_2}}{\bar{y}} = \frac{b\cdot\overline{x_2}}{a+b\cdot\overline{x_2}} = \frac{9{,}83\cdot 1{,}38}{9{,}30+9{,}83\cdot 1{,}38} = 0{,}59\%$$

Для уравнения степенной зависимости

$$\hat{y}(x_3) = 11{,}75 + x_3^{1{,}6281}$$

$$\overline{Э}(yx_3) = f^{`}(x_3)\frac{\overline{x_3}}{\bar{y}} = a\cdot b\cdot \bar{x}_3^{b-1}\cdot\frac{\overline{x_3}}{a\cdot\bar{x}_3^{b}} = b = 1{,}63\%$$

Для уравнения показательной зависимости

$$\hat{y}(x4) = 14{,}87\cdot 1{,}016^{x4}$$

$$\overline{Э}(yx_4) = f^{`}(x_4)\frac{\overline{x_4}}{\bar{y}} = a\cdot$$

$$b^{\overline{x_4}}\cdot\ln b * \frac{\overline{x_4}}{a * b^{\overline{x_4}}} = \ln b\cdot\overline{x_4} = \ln 1{,}016\cdot 26{,}3 = 0{,}42\%$$

2. Сравнивая значения $\overline{Э}_{yxi}$, ранжируем x_j по силе их влияния на себестоимость единицы продукции:

а) $\overline{Э}_{yx3}$ = 1,63%; в) $\overline{Э}_{yx2}$ = 0,59%;

б) $\overline{Э}_{yx1}$= -0,973%; г) $\overline{Э}_{yx4}$ = 0,42%.

Для формирования уровня себестоимости продукции группы предприятий первоочередное значение имеют цены на энергоносители; в гораздо меньшей степени влияют трудоемкость продукции и отчисляемая часть прибыли. Фактором снижения себестоимости выступает размер производства: с ростом его на 1% себестоимость единицы продукции снижается на -0,97%.

Пример 4

Зависимость потребления продукта *А* от среднедушевого дохода по данным 20 семей характеризуется следующим образом: уравнение регрессии $\bar{y}_x = 2\cdot x^{0,3}$; индекс корреляции ρ_{xy} = 0,9; остаточная дисперсия $\sigma^2_{ост}$ = 0,06.

Провести дисперсионный анализ полученных результатов. **Решение**

Результаты дисперсионного анализа приведены в таблица 4.

Таблица 4

Результаты дисперсионного анализа

Вариация результата у	Число степеней свободы	Сумма квадратов отклонений, *S*	Дисперсия на одну степень свободы, *D*	$F_{факт}$	$F_{табл}$ a = 0,05, k_1 =1, k_2 = 18
Общая	$df = n - 1$ = 19	6,316	-	-	-
Факторная	k_1 = m= *1*	5,116	5,116	76,7	4,41
Остаточная	$k_2 = n - m - 1$ = 18	1,200	0,0667	-	-

$S_{ост} = \sigma^2_{ост} \cdot n = 0{,}06 \cdot 20 = 1{,}2;$
$S_{общ} = S_{ост}/(1 - p^2_{xy}) = 1{,}2/(1 - 0{,}81) = 6{,}316;$
$S_{факт} = 6{,}316 - 1{,}2 = 5{,}116$

$$F_{факт} = \frac{0{,}9^2}{1 - 0{,}9^2} \cdot \frac{18}{1} = 76{,}7.$$

В силу того, что $F_{факт} = 76{,}7 > F_{табл} = 4{,}4$, гипотеза о случайности различий факторной и остаточной дисперсий отклоняется. Эти различия существенны, статистически значимы, уравнение надежно, значимо, показатель тесноты связи надежен и отражает устойчивую зависимость потребления продукта А от среднедушевого дохода.

Задание 1

Определить зависимости между сменной добычей известняка на одного рабочего (переменная *Y*, измеряемая в тоннах) и мощностью пласта (переменная *X*, измеряемая в метрах), результаты которых представлены таблицей 5.

Таблица 5

Исходные данные задачи

Номер наблюдения (i)	Мощность пласта, м (x_i)	Добыча известняка на одного рабочего, т. (y_i)
1	9+n	5+n
2	12+n	10+n
3	13+n	10+n
4	10+n	7+n
5	9+n	5+n
6	9+n	6+n
7	10+n	6+n
8	10+n	5+n
9	9+n	6+n
10	13+n	8+n

Замечание. n – номер варианта студента.

1.2. Лабораторная работа № 1
Определение уравнения линейной регрессии

Цель работы. Определение уравнения линейной регрессии и вычисление коэффициентов уравнения линейной регрессии по пространственной выборке табл. 5

Ход работы.

1. Вычислить коэффициенты a и b линейной регрессии и записать уравнение $y = a + bx$, используя табличный процессор Excel.

2. Используя получено уравнение, определите производительность труда рабочего, если толщина угольного слоя равна:

а) 8.5 метров (интерполяция данных);

б) 14 метров (экстраполяция данных).

Пример решения исходного примера представлен на рис. 11.

1	**Лабораторная работа № 1**						
2	**Вычисление коэффициентов уравнения линейной регрессии**						
3							
4		Исходные данные					
5		x_i	y_i	x^2_i	$x_i * y_i$		
6		8	5	=B6^2	=B6*C6		
7		11	10	=B7^2	=B7*C7		
8		12	10	=B8^2	=B8*C8	*a*	=(E16-B16*C16)/(D16-B16:B16^2)
9		9	7	=B9^2	=B9*C9	*b*	=C16-G8*B16
10		8	5	=B10^2	=B10*C10		
11		8	6	=B11^2	=B11*C11		
12		9	6	=B12^2	=B12*C12		
13		9	5	=B13^2	=B13*C13		
14		8	6	=B14^2	=B14*C14		
15		12	8	=B15^2	=B15*C15		
16	Средние значения	*=СРЗНАЧ(B6:B15)*	*=СРЗНАЧ(C6:C15)*	*=СРЗНАЧ(D6:D15)*	*=СРЗНАЧ(E6:E15)*		
17							

Рис. 11. Фрагмент листа электронной таблицы Microsoft Excel с расчетными формулами вычисления коэффициентов линейной регрессии

1.3. Лабораторная работа № 2
Вычисление выборочного коэффициента корреляции

Цель работы. Вычисление выборочного коэффициента корреляции по пространственной выборке таб. 5.

Ход работы.

1. Вычислить выборочный коэффициент корреляции по пространственной выборке таб. 5, используя табличный процессор Excel.

Пример решения исходного примера представлен на рис. 12.

	A	B	C	D	E	F	G	H
1	**Лабораторная работа № 2**							
2	**Вычисление выборочного коэффициента корреляции**							
3								
4	Исходные данные							
5	x_i	y_i	x^2_i	y^2_i	$x_i y_i$			
6	8	5	64	25	40			
7	11	10	121	100	110			
8	12	10	144	100	120	s_X	*1,562*	
9	9	7	81	49	63	s_Y	*1,833*	
10	8	5	64	25	40			
11	8	6	64	36	48	r_{XY}	*0,866*	
12	9	6	81	36	54			
13	9	5	81	25	45			
14	8	6	64	36	48			
15	12	8	144	64	96			
16	*9,4*	*6,8*	*90,8*	*49,6*	*66,4*	Средние значения		
17								

Рис. 12. Фрагмент листа электронной таблицы Microsoft Excel с расчетными формулами вычисления коэффициента корреляции

1.4. Лабораторная работа № 3
Вычисление оценок дисперсий коэффициентов парной линейной регрессии

Цель работы. Вычислить оценки S_a^2, S_b^2 для дисперсий коэффициентов *a, b*, определенных в лабораторной работе № 1.

Ход работы.

1. Вычислить оценки S_a^2, S_b^2 для дисперсий коэффициентов *a, b* по пространственной выборке таб. 5., используя табличный процессор Excel.

Пример решения исходного примера представлен на рис. 13.

	A	B	C	D	E	F	G	H	I
1	**Лабораторная работа № 3**								
2	**Вычисление оценок дисперсий коэффициентов парной линейной регрессии**								
3									
4	*a*	-2,75							
5	*b*	1,016							
6	Исходные данные								
7	x_i	y_i	$\hat{y}_i$	$e_i = \hat{y}_i - y_i$	e^2_i	$(x_i - \bar{x})^2$			
8	8	5	5,378	0,378	0,143	1,96			
9	11	10	8,426	-1,574	2,477	2,56			
10	12	10	9,442	-0,558	0,311	6,76			
11	9	7	6,394	-0,606	0,367	0,16			
12	8	5	5,378	0,378	0,143	1,96			
13	8	6	5,378	-0,622	0,387	1,96			
14	9	6	6,394	0,394	0,155	0,16			
15	9	5	6,394	1,394	1,943	0,16			
16	8	6	5,378	-0,622	0,387	1,96			
17	12	8	9,442	1,442	2,079	6,76			
18	*9,4*	*6,8*			*8,393*	*24,40*			
19									
20									
21	$\bar{x}$	9,400							
22	$\bar{x^2}$	90,800		S^2a	*3,904*				
23	S^2	*1,049*		S^2b	*0,043*				
24									

Рис. 13. Фрагмент листа электронной таблицы Microsoft Excel с расчетными формулами вычисления оценок для дисперсий и коэффициентов

1.5. Лабораторная работа № 4
Функции excel для вычисления коэффициентов парной линейной регрессии

Цель работы. Вычислить коэффициенты уравнения линейной регрессии по пространственной выборке таб. 5, используя функции Excel.

Ход работы.

1. Вычислить коэффициенты уравнения линейной регрессии по пространственной выборке таб. 5. Используя следующие функции Excel:

Для вычисления коэффициента *a:*

ОТРЕЗОК (*диапазон_значений_ y ; диапазон_значений_ x*).

Для вычисления коэффициента *b:*

НАКЛОН (*диапазон_значений_ y ; диапазон_значений_ x*).

Для вычисления значения линейной парной регрессии при заданном значении независимой переменной (обозначена через z) :

ПРЕДСКАЗ (z ;*диапазон_значений_ y ;диапазон_значений_ x*).

Для вычисления оценки S для среднеквадратического отклонения σ возмущений ε_i и обращение имеет вид (YX – латинские буквы):

СТОШYX(*диапазон_значений_ y ; диапазон_значений_ x*).

2. Сравните вычисленные значения *a*, *b*,*S* с значениями, полученными в лабораторных работах № 1 и № 3.

Пример решения задания представлен на рис. 14.

	A	B	C
1	**Лабораторная работа № 4**		
2	**Функции Excel для вычисления коэффициентов парной линейной регрессии**		
3			
4	Исходные данные		
5	x_i	y_i	x^2_i
6	8	5	=ПРЕДСКАЗ(A6;$B$6:$B$15;$A$6:$A$15)
7	11	10	=ПРЕДСКАЗ(A7;$B$6:$B$15;$A$6:$A$15)
8	12	10	=ПРЕДСКАЗ(A8;$B$6:$B$15;$A$6:$A$15)
9	9	7	=ПРЕДСКАЗ(A9;$B$6:$B$15;$A$6:$A$15)
10	8	5	=ПРЕДСКАЗ(A10;$B$6:$B$15;$A$6:$A$15)
11	8	6	=ПРЕДСКАЗ(A11;$B$6:$B$15;$A$6:$A$15)
12	9	6	=ПРЕДСКАЗ(A12;$B$6:$B$15;$A$6:$A$15)
13	9	5	=ПРЕДСКАЗ(A13;$B$6:$B$15;$A$6:$A$15)
14	8	6	=ПРЕДСКАЗ(A14;$B$6:$B$15;$A$6:$A$15)
15	12	8	=ПРЕДСКАЗ(A15;$B$6:$B$15;$A$6:$A$15)
16			
17			
18		b_0	***=ОТРЕЗОК(B6:B15;A6:A15)***
19		b_1	***=НАКЛОН(B6:B15;A6:A15)***
20		S	***=СТОШYX(B6:B15;A6:A15)***
21			

Рис. 14. Фрагмент листа электронной таблицы Microsoft Excel с расчетными формулами

1.6. Лабораторная работа № 5
Построение интервальной оценки для функции парной линейной регрессии

Цель работы. Построение интервальной оценки для функции регрессии $f(x) = M(Y \mid x)$ с надежностью γ = 0.95, используя для этого уравнение регрессии $\hat{y}(x)$, построенное в лабораторной работе № 1.

Ход работы.

1. Построить интервальные оценки для функции регрессии $f(x) = M(Y \mid x)$ с надежностью γ = 0.95. Используя следующие функции Excel:

$$t(\gamma, n-2) = \text{СТЬЮДРАСПОБР}(1-\gamma; n-2).$$

Значения нижней y_i^H и верхней y_i^B границ интервала вычислить для $x = x_i, i = 1,\ldots,10$. Величины $\sum_{i=1}^{10}(x_i - \bar{x})^2$, s^2, $\bar{x}$ (ячейки B16:B18) и коэффициенты a, b (B1:B2) взяты из предыдущих лабораторных работ. Величина $t(0.95, 10-2)$ = СТЬЮДРАСПОБР(0.05;10 – 2) = 2.31.

Пример решения задания представлен на рис. 15.

1	**Лабораторная работа № 5**					
2	**Построение интервальной оценки для функции парной линейной регрессии**					
3						
4	*a*	-2,75				
5	*b*	1,016				
6	Исходные данные					
7	x_i	y_i	$\hat{y}_i$	$S^2_{\hat{y}}$	y^H_i	y^B_i
8	8	5	=$B$4+$B$5*A8	=$B$19*(1/10+(A8-$B	***=C8-2,31*КОРЕНЬ(***	***=C8+2,31*КОРЕНЬ(***
9	11	10	=$B$4+$B$5*A9	=$B$19*(1/10+(A9-$B	***=C9-2,31*КОРЕНЬ(***	***=C9+2,31*КОРЕНЬ(***
10	12	10	=$B$4+$B$5*A10	=$B$19*(1/10+(A10-$I	***=C10-2,31*КОРЕНЬ***	***=C10+2,31*КОРЕНЬ***
11	9	7	=$B$4+$B$5*A11	=$B$19*(1/10+(A11-$I	***=C11-2,31*КОРЕНЬ***	***=C11+2,31*КОРЕНЬ***
12	8	5	=$B$4+$B$5*A12	=$B$19*(1/10+(A12-$I	***=C12-2,31*КОРЕНЬ***	***=C12+2,31*КОРЕНЬ***
13	8	6	=$B$4+$B$5*A13	=$B$19*(1/10+(A13-$I	***=C13-2,31*КОРЕНЬ***	***=C13+2,31*КОРЕНЬ***
14	9	6	=$B$4+$B$5*A14	=$B$19*(1/10+(A14-$I	***=C14-2,31*КОРЕНЬ***	***=C14+2,31*КОРЕНЬ***
15	9	5	=$B$4+$B$5*A15	=$B$19*(1/10+(A15-$I	***=C15-2,31*КОРЕНЬ***	***=C15+2,31*КОРЕНЬ***
16	8	6	=$B$4+$B$5*A16	=$B$19*(1/10+(A16-$I	***=C16-2,31*КОРЕНЬ***	***=C16+2,31*КОРЕНЬ***
17	12	8	=$B$4+$B$5*A17	=$B$19*(1/10+(A17-$I	***=C17-2,31*КОРЕНЬ***	***=C17+2,31*КОРЕНЬ***
18						
19	s^2	**1,049**				
20	$\dot{x}$	**9,4**				
21	$\sum$	**24,4**				
22						

Рис.15. Фрагмент листа электронной таблицы Microsoft Excel с расчетными формулами интервальной оценки для $f(x) = M(Y \mid x)$

1.7. Лабораторная работа № 6
Проверка значимости уравнения линейной регрессии по критерию Фишера

Цель работы. По данным таблицы 5 оценить на уровне α = 0.05 значимость уравнения регрессии $\hat{y}(x) = -2.75 + 1.016 \cdot x$, построенного в лабораторной работе № 1.

Ход работы.

1. Оценить на уровне α = 0.05 значимость уравнения регрессии $\hat{y}(x) = -2.75 + 1.016 \cdot x$. Используя следующие функции Excel:

$$F_{1-\alpha;1;n-2} = \text{FРАСПОБР}(\alpha;1;n-2).$$

Значения Q_e, $Q_r = Q - Q_e$ и критерий F. В столбце D значения вычисляются по формуле $\hat{y}_i = \hat{y}(x_i) = -2.75 + 1.016 \cdot x_i$. Значения коэффициентов a и b взяты из лабораторной работы № 1.

2. Сделать вывод.

Пример решения задания представлен на рис. 16.

1	**Лабораторная работа № 6**				
2	**Проверка значимости уравнения линейной регрессии по критерию Фишера**				
3					
4	*a*	-2,75			
5	*b*	1,016			
6	Исходные данные				
7	x_i	y_i	$(y_i - \bar{Y})^2$	Yt	$(yj-yi)^2$
8	8	5	3,240	5,378	0,143
9	11	10	10,240	8,426	2,477
10	12	10	10,240	9,442	0,311
11	9	7	0,040	6,394	0,367
12	8	5	3,240	5,378	0,143
13	8	6	0,640	5,378	0,387
14	9	6	0,640	6,394	0,155
15	9	5	3,240	6,394	1,943
16	8	6	0,640	5,378	0,387
17	12	8	1,440	9,442	2,079
18		*6,800*	*33,600*		*8,393*
19		$\bar{Y}$	Q		Q_e
21					
22	$Q_r = Q - Q_e$		*25,207*		
23	F		*24,02498*		
24					

Рис. 16. Фрагмент листа электронной таблицы Microsoft Excel с расчетными формулами вычисления величины F – критерия

2. НЕЛИНЕЙНАЯ ПАРНАЯ РЕГРЕССИЯ

Задание 2. В таблице 6 приведены значения независимой переменной X (доход российской семьи в тыс. руб.) и значения зависимой переменной Y (доля расходов на товары длительного пользования в процентах от общей суммы расходов).

Таблица 6

Исходные данные задачи

Номер наблюдения (i)	Доход российской семьи в тыс. руб. (x_i)	Доля расходов на товары длительного пользования в процентах от общей суммы расходов), %. (y_i)
1	1+n	11+n
2	2+n	14,4+n
3	3+n	16,4+n
4	4+n	17,5+n
5	5+n	19,6+n
6	6+n	20,1+n

Замечание. n - номер варианта студента.

2.1. Лабораторная работа № 7
Построение нелинейной регрессии с использованием команды «добавить линию тренда»

Цель работы. Научится строить уравнение нелинейной регрессии с использованием команды «Добавить линию тренда».

Ход работы.

1. Используя пространственную выборку табл. 6 необходимо построить уравнение нелинейной регрессии вида $y = a \cdot x^b$ с использованием команды «Добавить линию тренда» и вычислить коэффициент детерминации R^2.

Команда «Добавить линию тренда». Используется для выделения тренда (медленных изменений) при анализе временных рядов. Однако эту команду можно использовать и для построения уравнения нелинейной регрессии, рассматривая в качестве времени t независимую переменную x.

Эта команда позволяет построить следующие уравнения регрессии:

Для построения одной из перечисленных регрессий необходимо выполнить следующие шаги:

Шаг 1. В выбранном листе Excel ввести по столбцам исходные данные $\{x_i, y_i\}, i = 1,2,\ldots,n$ (см. рис. 17).

Шаг 2. По этим данным построить график в декартовый системе координат (см. рис 17).

Шаг 3. Установить курсор на построенном графике, сделать щелчок правой кнопкой и в появившемся контекстном меню выполнить команду *Добавить линию тренда* (см. рис. 17).

Шаг 4. В появившемся диалоговом окне (см. рис. 18) активизировать закладку «Тип» и выбрать нужное уравнение регрессии.

Шаг 5. Активизировать закладку «Параметры» (см. рис. 18) и «включить» *необходимые для нас опции:*

- *«Показать уравнение на диаграмме»* - на диаграмме будет показано выбранное уравнение регрессии с вычисленным коэффициентами;
- *«Поместить на диаграмму величину достоверности аппроксимации (R^2)»* - на диаграмме будет показана значение коэффициент детерминации R^2 (для нелинейной регрессии -индекс детерминации. Если по построенному уравнению регрессии необходимо выполнить прогноз, то нужно указать число периодов прогноза (см. рис. 17).

Назначение других опций понятны из своих названий.

Шаг 6. После задания всех перечисленных опций щелкнуть на кнопке «ОК» и на диаграмме появиться формула построенного уравнения регрессии и значение индекса детерминации R^2 (рис. 18).

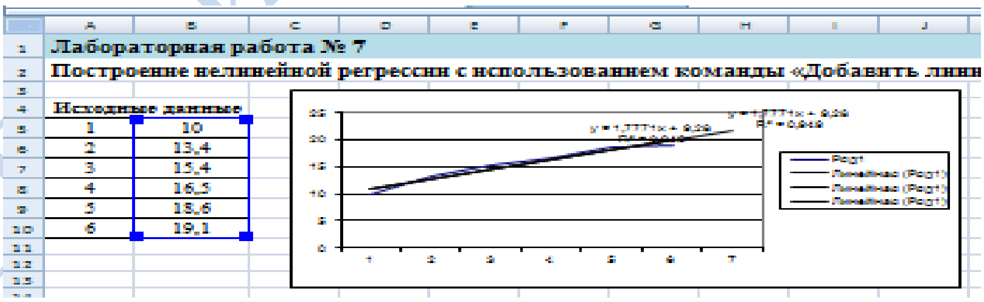


Рис. 17. Фрагмент листа электронной таблицы Microsoft Excel построение графика по исходным данным

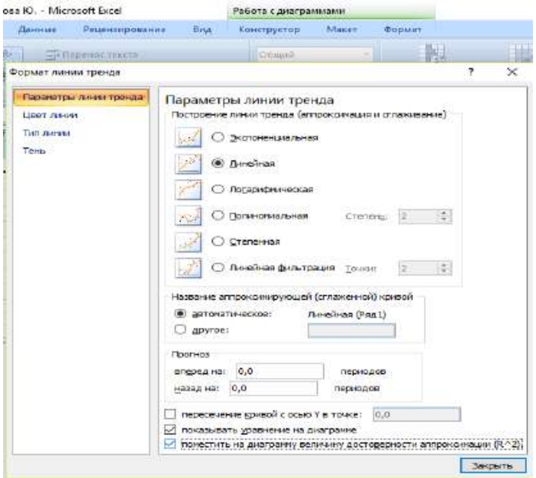


Рис. 18. Фрагмент листа электронной таблицы Microsoft Excel выбор вида уравнения регрессии

2.2. Лабораторная работа № 8
Выбор наилучшей нелинейной регрессии по приведенному коэффициенту детерминации

Цель работы. Научится строить различные виды уравнений нелинейной регрессии с использованием команды «Добавить линию тренда».

Ход работы.

1. Используя пространственную выборку таблицы 6 и команду «Добавить линию тренда» построить шесть уравнений нелинейной регрессии:

- линейную $y = a \cdot x^b$;
- полиноминальную $y = a + b_1 x + \ldots + b_k x^k$ ($k \le 6$)(при $m = 2$ и $m = 3$);
- логарифмическую $y = a + b_1 \ln x$;
- степенную $y = a \cdot x^{b_1}$;
- экспоненциальную $y = a \cdot e^{b_1 x}$.

2. Определить для каждого уравнения коэффициент детерминации R^2 (значение выводится), приведенный коэффициент детерминации $\hat{R}^2$ (значение вычисляется) и по максимальному значению $\hat{R}^2$ найти наилучшее и наихудшее уравнение нелинейной регрессии.

Пример решения задания представлен на рис. 19.

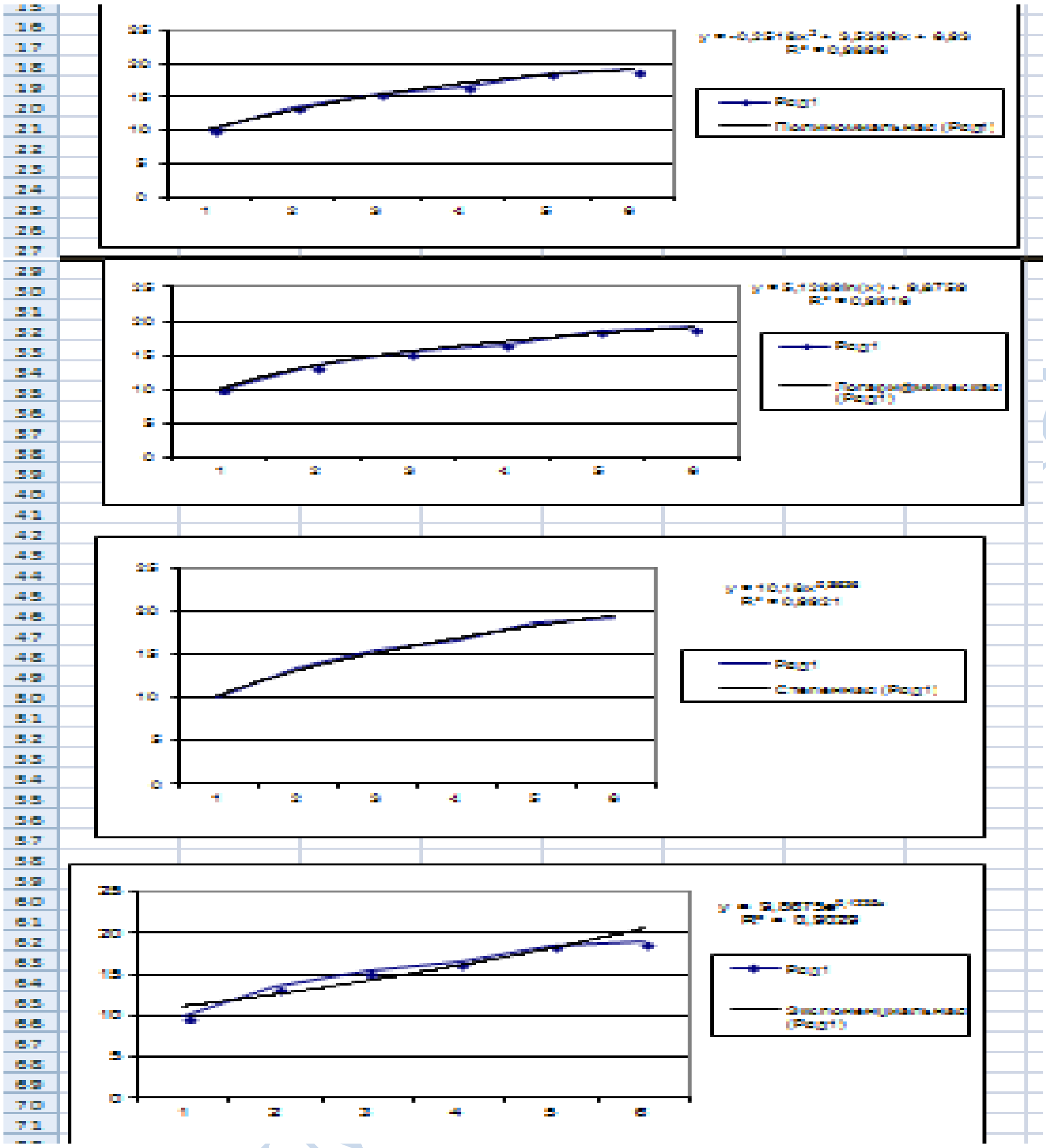


Рис. 19. Фрагмент листа электронной таблицы Microsoft Excel построение графиков по исходным данным

Замечание. Коэффициент детерминации R^2 характеризует близость построенной регрессии к исходным данным, которые содержат «нежелательную» случайную составляющую ε. Очевидно, что, построив по данным таб. 7 полином 5-ого порядка, получаем «идеальное» значение $R^2 = 1$, по такое уравнение содержит в себе не только независимую переменную X, но составляющую ε и это снижает точность использования построенного уравнения для прогноза. Поэтому при выборе уравнения регрессии надо учитывать не только величину R^2, но и «сложность» регрессионного уравнения, определяемое количеством коэффициентов уравнения. Такой учет удачно реализован в так называемом *приведенном коэффициенте детерминации:*

$$\hat{R}^2 = 1 - \frac{(n-1)\cdot Q_e}{(n-m)\cdot Q} = 1 - \frac{n-1}{n-m}\cdot(1-R^2)\,, \qquad (2)$$

где m - количество вычисляемых коэффициентов регрессии. Видно, что при неизменных Q_e, Q увеличение m уменьшает значение $\hat{R}^2$. Если количество коэффициентов у сравниваемых уравнений регрессии одинаково (например, $m = 2$), то отбор наилучшей регрессии можно осуществлять по величине R^2. Если в уравнениях регрессии меняется число коэффициентов, то такой отбор целесообразно по величине $\hat{R}^2$.

Значение коэффициентов детерминации R^2 и $\hat{R}^2$ представим в табл. 7.

Таблица 7

№	Уравнение	R^2	$\hat{R}^2$
1	$\hat{y} = 9.28 + 1.777x$	0.949	0.938
2	$\hat{y} = 9.8759 + 5.1289\cdot\ln x$	0.9916	0.9895
3	$\hat{y} = 6.93 + 3.5396x - 0.2518x^2$ (полиноминальная, $m = 2$)	0.9896	0.9827
4	$\hat{y} = 5.8333 + 4.9192\cdot x - 0.7087\cdot x^2 - 0.0435\cdot x^3$ (полиноминальная, $m = 3$)	0.9917	0.9792
5	$\hat{y} = 10.18x^{0.3626}$	0.9921	0.9901
6	$\hat{y} = 9.8675\cdot e^{0.1225x}$	0.9029	0,8786

В качестве «наилучшего» уравнения регрессии выбираем уравнение, имеющее наибольшую величину приведенный коэффициент детерминации $\hat{R}^2$. Из таб. 7 видно, что таким уравнением является степенная функции (в таблице строка с этой функцией выделена серым цветом).

2.3. Лабораторная работа № 9
Решение задач с помощью встроенных статистических функция

Цель работы. Научится строить различные виды уравнений линейной и нелинейной регрессии с использованием встроенных статистических функций

Ход работы.

1. Встроенная статистическая функция ЛИНЕЙН определяет параметры линейной регрессии $y = a + b\,x$.

Порядок вычисления следующий:

1) введите исходные данные или откройте существующий файл, содержащий анализируемые данные;

2) выделите область пустых ячеек 5х2 (5 строк, 2 столбца) для вывода результатов регрессионной статистики или область 1х2 - для получения только оценок коэффициентов регрессии;

3) активизируйте Мастер функций любым из способов:

а)в главном меню выберите **Вставка/Функция;**

б)на панели инструментов **Стандартная** щелкните по кнопке **Вставка функции;**

4) в окне Категория (рис. 20) выберите **Статистические,** в окне Функция - **ЛИНЕЙН.** Щелкните по кнопке **ОК;**

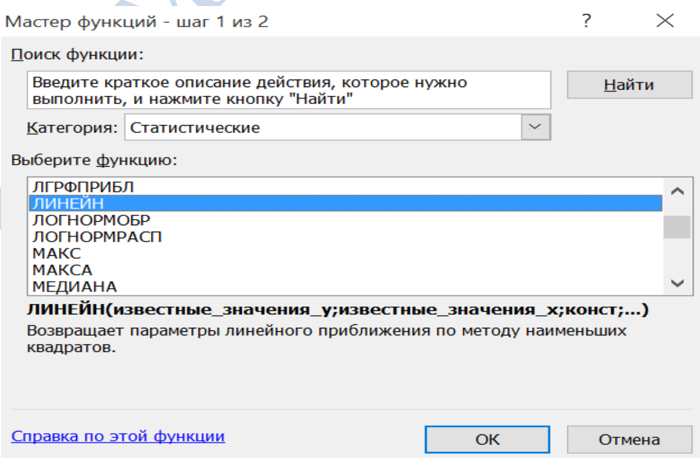


Рис. 20. Диалоговое окно «Мастер функций»

5) заполните аргументы функции (рис. 21):

Известные_значения_у - диапазон, содержащий данные результативного признака;

Известные_значения_х - диапазон, содержащий данные факторов независимого признака;

Константа - логическое значение, которое указывает на наличие или на отсутствие свободного члена в уравнении; если *Константа* = 1, то свободный член рассчитывается обычным образом, если *Константа* = 0, то свободный член равен 0; *Статистика* - логическое значение, которое указывает, выводить дополнительную информацию по регрессионному анализу или нет. Если *Статистика* = 1, то дополнительная информация выводится, если *Статистика* - 0, то выводятся только оценки параметров уравнения. Щелкните по кнопке ОК;

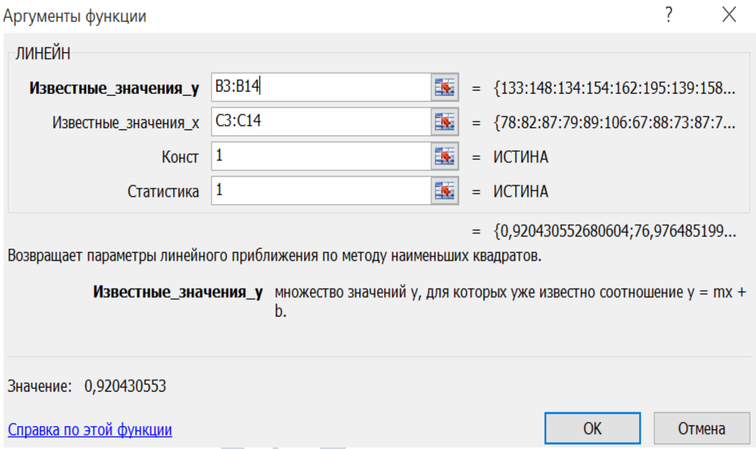


Рис. 21. Диалоговое окно ввода аргументов функции ЛИНЕЙН.

6) в левой верхней ячейке выделенной области появится первый элемент итоговой таблицы. Чтобы раскрыть всю таблицу, нажмите на клавишу <F2>, а затем - на комбинацию клавиш <CTRL>+<SHIFT>+<ENTER>.

Дополнительная регрессионная статистика будет выводиться в порядке, указанном в следующей схеме (таб. 8).

Таблица 8

Регрессионная статистика

Значение коэффициента *b*	Значение коэффициента *a*
Среднеквадратическое отклонение *b*	Среднеквадратическое отклонение *a*
Коэффициент детерминации R^2	Среднеквадратическое отклонение *y*
F-статистика	Число степеней свободы
Регрессионная сумма квадратов	Остаточная сумма квадратов

Для вычисления параметров экспоненциальной кривой $y = \alpha \cdot \beta^x$ в MS Excel применяется встроенная статистическая функция **ЛГРФПРИБЛ.** Порядок вычисления аналогичен применению функции **ЛИНЕЙН.**

Для данных из примера 2 результат вычисления функции **ЛИНЕЙН** представлен на рис. 22, функции **ЛГРФПРИБЛ -** на рис. 23.

	A	B	C	D	E	F	G	H
1	Территории региона	Прожит. мин.- x	Сред.мес. з/п - y					
2	1	78	133					
3	2	82	148					
4	3	87	134			Линейн		
5	4	79	154					
6	5	89	162			0,920431	76,97649	
7	6	106	195			0,279716	24,21156	
8	7	67	139			0,519877	12,54959	
9	8	88	158			10,82801	10	
10	9	73	152			1705,328	1574922	
11	10	87	162					
12	11	76	159					
13	12	115	173					

Рис. 22. Фрагмент листа электронной таблицы Microsoft Excel результатом вычисления функции ЛИНЕЙН

	A	B	C	D	E	F	G	H	I
1	Территории региона	Прожит. мин.- x	Сред.мес. з/п - y						
2	1	78	133						
3	2	82	148						
4	3	87	134			Лгрфприбл			
5	4	79	154						
6	5	89	162			1,005664	96,53277		
7	6	106	195			0,001791	0,154997		
8	7	67	139			0,498671	0,08034		
9	8	88	158			9,946979	10		
10	9	73	152			0,064202	0,064544		
11	10	87	162						
12	11	76	159						
13	12	115	173						

Рис. 23. Результат вычисления функции ЛГРФПРИБЛ

2.4. Лабораторная работа № 10
Решение задач с помощью с помощью инструмента *Анализа данных Регрессия*

Цель работы. Научится строить различные виды уравнений линейной и нелинейной регрессии с использованием встроенных статистических функций

Ход работы.

1. С помощью инструмента анализа данных Регрессия, помимо результатов регрессионной статистики, дисперсионного анализа и доверительных интервалов, можно получить остатки и графики подбора линии регрессии, остатков и нормальной вероятности Порядок действий следующий.

1) проверьте доступ к пакету анализа В главном меню последовательно выберите **Сервис /Надстройки**. Установите флажок Пакет анализа (рис. 24);

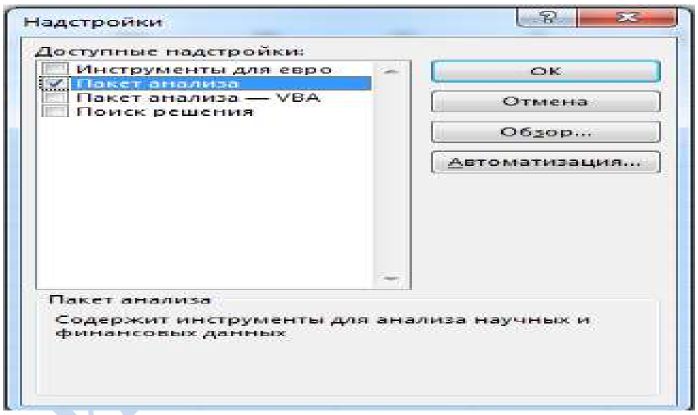


Рис. 24. Подключение надстройки Пакет анализа

2) в главном меню выберите **Сервис/Анализ данных/Регрессия.** Щелкните по кнопке **ОК**;

3) заполните диалоговое окно ввода данных и параметров вывода (рис.25).

Входной интервал Y – диапазон, содержащий данные результативного признака,

Входной интервал X – диапазон, содержащий данные факторов независимого признака;

Метки – флажок, который указывает, содержит ли первая строка названия столбцов или нет;

Константа – ноль – флажок, указывающий на наличие или отсутствие свободного члена в уравнении;

Выходной интервал – достаточно указать левую верхнюю ячейку будущего диапазона;

Новый рабочий лист – можно задать произвольное имя нового листа.

Если необходимо получить информацию и графики остатков, установите соответствующие флажки в диалоговом окне. Щелкните по кнопке **ОК**.

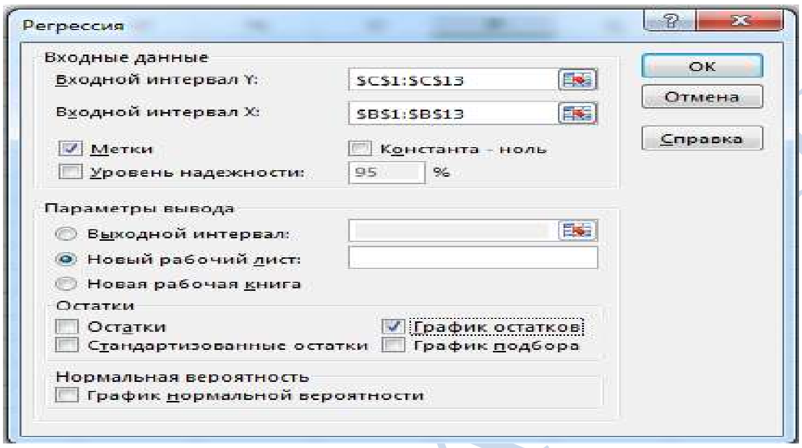
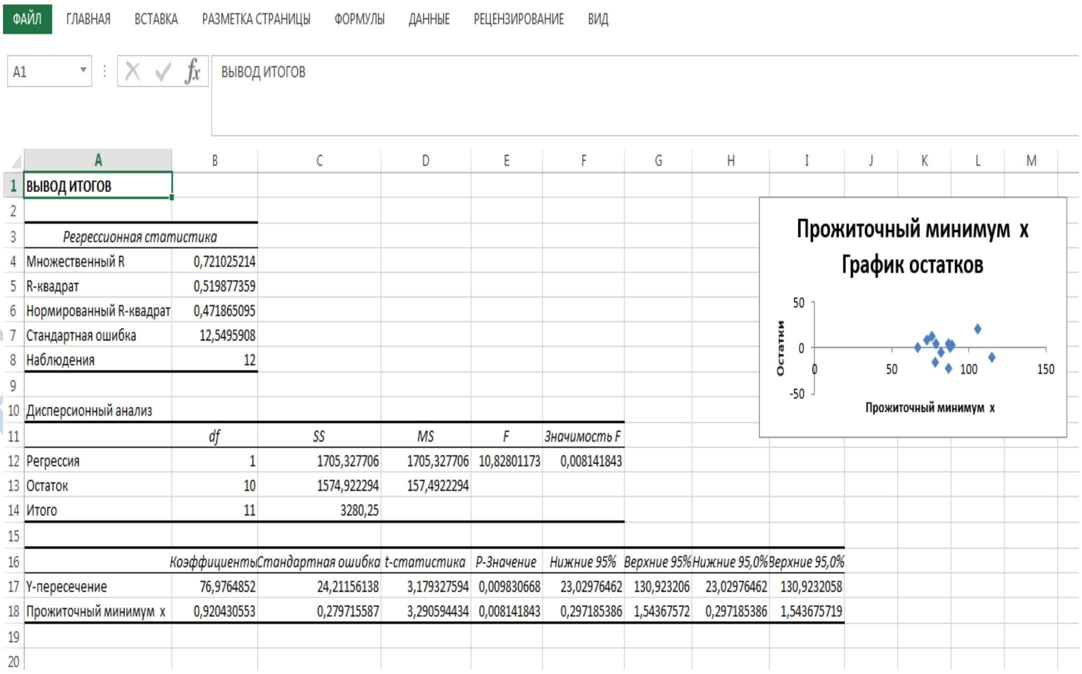


Рис. 26. Диалоговое окно ввода параметров инструмента Регрессия

Результаты регрессионного анализа для данных из примера 2 представлены на рис. 27.



ВЫВОД ИТОГОВ								
Регрессионная статистика								
Множественный R	0,721025214							
R-квадрат	0,519877359							
Нормированный R-квадрат	0,471865095							
Стандартная ошибка	12,5495908							
Наблюдения	12							
Дисперсионный анализ								
	df	*SS*	*MS*	*F*	*Значимость F*			
Регрессия	1	1705,327706	1705,327706	10,82801173	0,008141843			
Остаток	10	1574,922294	157,4922294					
Итого	11	3280,25						
	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
Y-пересечение	76,9764852	24,21156138	3,179327594	0,009830668	23,02976462	130,923206	23,02976462	130,9232058
Прожиточный минимум x	0,920430553	0,279715587	3,290594434	0,008141843	0,297185386	1,54367572	0,297185386	1,543675719

Рис. 27. Фрагмент листа электронной таблицы Microsoft Excel с результатом применения инструмента *Регрессия*

3. ЛИНЕЙНАЯ МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

Эта тема включает выполнение лабораторных работ, посвященных построению и исследованию уравнения линейной множественной регрессии вида

$$y = b_0 + b_1 x_1 + \ldots + b_к x_к \qquad (3)$$

Пространственная выборка для построения этого уравнения взята из следующего примера.

Задание 3. Определить зависимости между сменной добычей известняка на одного рабочего (переменная Y, измеряемая в тоннах) и мощностью пласта (переменная X_1, измеряется в метрах) и уровнем механизации работ (переменная X_2, измеряется в процентах), результаты которых представлены таблицей 9

Таблица 9

Исходные данные задачи

Номер наблюдения (i)	Мощность пласта, м (x_{i1})	Уровень механизации работ, % (x_{i2})	Добыча известняка на одного рабочего, т. (y_i)
1	9+n	6+n	5+n
2	12+n	9+n	10+n
3	13+n	9+n	10+n
4	10+n	6+n	7+n
5	9+n	8+n	5+n
6	9+n	9+n	6+n
7	10+n	7+n	6+n
8	10+n	5+n	5+n
9	9+n	6+n	6+n
10	13+n	8+n	8+n

Замечание. n - номер варианта студента.

3.1. Лабораторная работа № 11
Вычисление коэффициентов линейной множественной регрессии

Цель работы. Используя пространственную выборку таблицы 9 необходимо вычислить вектор коэффициентов $b = \begin{vmatrix} b_0 \\ b_1 \\ b_2 \end{vmatrix}$ уравнения регрессии (3).

Ход работы.

Вектор коэффициентов, найденный методом наименьших квадратов является решением следующей системы уравнений:

$$X^T Xb = X^T y,$$

где X - матрица размера 10×3, первый столбец которой составлен из 1, а другие два столбца составлены из значений x_{i1}, x_{i2}, т.е. матрица X имеет следующую структуру (символы … означают не отображенные элементы)

$$X = \begin{vmatrix} 1 & 8 & 5 \\ 1 & 11 & 8 \\ \dots & \dots & \dots \\ 1 & 12 & 7 \end{vmatrix},$$

а y - вектор, составленный из 10 значений y_i, т.е.

$$y = \begin{vmatrix} 5 \\ 10 \\ \vdots \\ 8 \end{vmatrix}.$$

Матрица $X^T X$ имеет обратную матрицу $\left(X^T X\right)^{-1}$ и тогда вектор коэффициентов вычисляется в виде:

$$b = A^{-1}(X^T y). \qquad (4)$$

Для реализации этой матричной формулы необходимо выполнить следующие операции: транспонирование; умножение матриц (частный случай – умножение матрицы на вектор); вычисление обратной матрицы. Все эти операции можно реализовать с помощью следующих для работы с матричных функциями Excel:

1) обратиться к ***Мастеру функций*** и выберем нужную категорию функций, затем указать имя функции и задать соответствующие диапазоны ячеек, или ввести с клавиатуры имя функции задать соответствующие диапазоны ячеек.

Транспонирование матрицы осуществляется с помощью функции ТРАНСП (категория функций – *Ссылки и массивы*). Обращение к функции имеет вид: ТРАНСП (*диапазон ячеек*), где параметр *диапазон ячеек* задает все элементы транспонируемой матрицы (или вектора).

Умножение матриц осуществляется с помощью функции МУМНОЖ (категория функций – *Математические*). Обращение к функции имеет вид:

МУМНОЖ(*диапазон_1;диапазон_2*),

где параметр *диапазон_1* задает элементы первой из перемножаемых матриц, а параметр *диапазон_2* – элементы второй матрицы. При этом перемножаемые матрицы должны иметь соответствующие размеры (если первая матрица $n \times k$, вторая - $k \times m$, то результатом будет матрица $n \times m$).

Обращение матрицы (вычисление обратной матрицы) осуществляется с помощью функции МОБР (категория функций – *Математические*). Обращение к функции имеет вид:

МОБР (*диапазон ячеек*),

где параметр *диапазон ячеек* задает все элементы обращаемой матрицы, которая должна быть квадратной и невырожденной.

Замечание. При использовании этих функций необходимо соблюдать следующий порядок действий:

- *выделить фрагмент ячеек*, в которые будет занесен результат выполнения матричных функций (при этом надо учитывать размеры исходных матриц);
- *ввести арифметическое выражение*, содержащее обращение к матричным функциям Excel;
- *одновременно нажать клавиши* [Ctrl], [Shift], [Enter]. Если этого не сделать, то *вычислится только один элемент* результирующей матрицы или вектора.

Сформируем матрицу X и вектор $У$ (см. рис. 28).

	A	B	C	D	E	F
2	**Вычисление коэффициентов линейной множественной регрессии**					
3						
4		1	9	6		5
5		1	12	9		10
6		1	13	9		10
7		1	10	6		7
8	X=	1	9	8	y=	5
9		1	9	9		6
10		1	10	7		6
11		1	10	5		5
12		1	9	6		6
13		1	13	8		8
14						
15		=МУМНОЖ(ТРАНСП(B4:D13);B4:D13)	=МУМНОЖ(ТРАН	=МУМНОЖ(ТРА		=МУМНОЖ(ТРАНСП(B4:D13);F4:F13)
16	$X^T \cdot X =$	=МУМНОЖ(ТРАНСП(B4:D13);B4:D13)	=МУМНОЖ(ТРАН	=МУМНОЖ(ТРА	$X^T \cdot y =$	=МУМНОЖ(ТРАНСП(B4:D13);F4:F13)
17		=МУМНОЖ(ТРАНСП(B4:D13);B4:D13)	=МУМНОЖ(ТРАН	=МУМНОЖ(ТРА		=МУМНОЖ(ТРАНСП(B4:D13);F4:F13)
18						
19						
20		=МОБР(B15:D17)	=МОБР(B15:D17)	=МОБР(B15:D17)		=МУМНОЖ(B20:D22;F15:F17)
21	$(X^\wedge T{\cdot}X)^\wedge(-1)=$	=МОБР(B15:D17)	=МОБР(B15:D17)	=МОБР(B15:D17)	b =	=МУМНОЖ(B20:D22;F15:F17)
22		=МОБР(B15:D17)	=МОБР(B15:D17)	=МОБР(B15:D17)		=МУМНОЖ(B20:D22;F15:F17)
23						
24	Уравнение регрессии: y(x1,x2)=-3,54+0,854x1+0,367x2					

Рис. 28. Вычисление коэффициентов множественной регрессии

Затем выполним формирование матрицы $X^T X$, вектора $X^T y$ и вычисление вектора $b = |b_0, b_1, b_2|^T$ по формуле (4). Все эти вычисления показаны на рис. 28. Получен вектор коэффициентов $b = \begin{vmatrix} -3.5393 \\ 0.8539 \\ 0.3670 \end{vmatrix}$ и тогда уравнение регрессии (3) примет вид: $\hat{y}(x_1, x_2) = -3.54 + 0.854 x_1 + 0.367 x_2$.

3.2 Лабораторная работа № 12
Вычисление коэффициентов линейной множественной регрессии и проверка значимости в режиме *Регрессия*

Цель работы. Используя пространственную выборку таблицы 9 и используя режим **Регрессия** необходимо вычислить вектор коэффициентов уравнения регрессии

$$\hat{y}(x_1, x_2) = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 . \qquad (4)$$

Ход работы. Первоначально введем в столбец С десять значений первой переменной, в столбец D - десять значений первой переменной (см. рис. 29), а в столбец F – десять значений зависимой переменной.

Вызовем режима *Регрессия* модуля *Анализ данных* (см. Лаб. раб. №10). Зададим необходимые параметры (см. рис. 29).

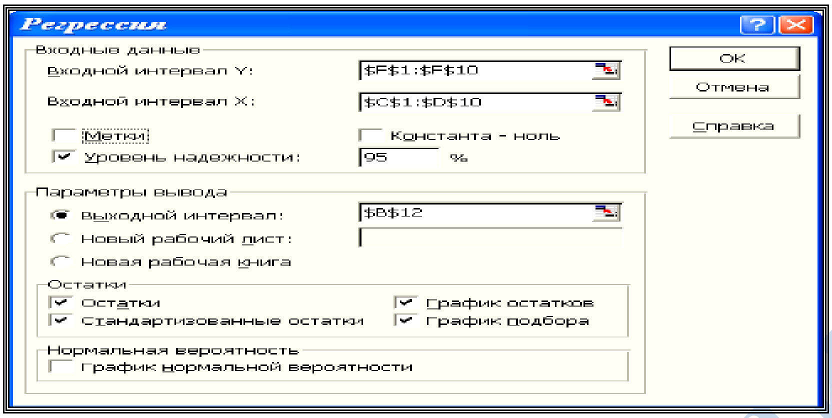


Рис. 29. Диалоговое окно режима *Регрессия*

Результаты работы приводятся рис. 30-35. Заметим, из-за большой «ширины» таблиц, в которых выводятся результаты работы режима *Регрессия,* часть результатов помещены в другие ячейки.

	A	B	C	D	E	F
16	ВЫВОД ИТОГОВ					
17						
18	*Регрессионная статистика*					
19	Множественный R	0,9009				
20	R-квадрат	0,8116				
21	Нормированный R-квадра	0,7578				
22	Стандартная ошибка	0,9509				
23	Наблюдения	10,0000				
24						
25	*Дисперсионный анализ*					
26		*df*	*SS*	*MS*	*F*	*Значимость F*
27	Регрессия	2	27,2704	13,6352	15,0794	0,0029
28	Остаток	7	6,3296	0,9042		
29	Итого	9	33,6000			

Рис. 30. Результаты работы режима *Регрессия*

Дадим краткую интерпретацию показателям, значения которых вычисляются в режиме *Регрессия.* Первоначально рассмотрим показатели, объединенные названием *Регрессионная статистика* (см. рис. 30).

Множественный R - корень квадратный из коэффициента детерминации.

$R-$*квадрат* – коэффициент детерминации R^2.

Нормированный $R-$*квадрат* – приведенный коэффициент детерминации $\hat{R}^2$ (см. формулу (2)).

Стандартная ошибка – оценка s для среднеквадратического отклонения σ.

Наблюдения – число наблюдений n.

Перейдем к показателям, объединенных названием *Дисперсионный анализ* (см. рис. 30).

Столбец df - число степеней свободы. Для строки *Регрессия* показатель равен числу независимых переменных $k_r = k = m-1$; для строки *Остаток* - равен $k_e = n-(k_r+1) = n-m$; для строки *Итого* – равен $k_r + k_e$.

Столбец SS – сумма квадратов отклонений. Для строки *Регрессия* показатель равен величине Q_r, т.е.

$$SS_r = Q_r = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2;$$

для строки *Остаток* - равен величине Q_e, т.е.

$$SS_e = Q_e = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2;$$

для строки *Итого* – равен $Q = Q_r + Q_e$.

Столбец $MS-$ дисперсии, вычисленные по формуле

$$MS = \frac{SS}{df},$$

т.е. дисперсия на одну степень свободы.

Столбец F – значение F_c, равное $F-$ критерию Фишера, вычисленного по формуле:

$$F_c = \frac{SS_r / k_r}{SS_e / k_e}.$$

Столбец значимость F - значение уровня значимости, соответствующее вычисленной величине $F-$критерия и равное вероятности $P(F(k_r, k_e) \geq F_c)$, где $F(k_r, k_e)$ - случайная величина, подчиняющаяся распределению Фишера с k_r, k_e степенями свободы. Эту вероятность можно также определить с помощью функции FРАСП($F_c; k_r; k_e$). *Если вероятность меньше уровня значимости* α *(обычно* $\alpha = 0.05$*), то построенная регрессия является значимой..*

Перейдем к следующей группе показателей, объединенных в таблице, показанной на рис. 31.

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*
Y-пересечение	-3,5393	1,9066	-1,8564	0,1058
Переменная X 1	0,8539	0,2205	3,8726	0,0061
Переменная X 2	0,3670	0,2429	1,5108	0,1746

Нижние 95%	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
-8,0477	0,9690	-8,0477	0,9690
0,3325	1,3753	0,3325	1,3753
-0,2074	0,9415	-0,2074	0,9415

Рис. 31. Фрагмент листа электронной таблицы Microsoft Excel с результатов работы режима *Регрессия*

Столбец Коэффициенты – вычисленные значения коэффициентов $b_0, b_1, ..., b_k$, расположенных сверху – вниз.

Столбец Стандартная ошибка – значения $s_{b_j}, j = 0,...,k$, вычисленные по формуле $s_{b_j} = \sqrt{s^2 \cdot \left\{ \left(X^T X \right)^{-1} \right\}_{j,j}}$.

Столбец $t -$ *статистика* – значения статистик T_{b_j}.

Столбец P – значение – содержит вероятности случайных событий $P(t(n-m) \geq T_{b_j})$, где $t(n-m) -$ случайная величина, подчиняющаяся распределению Стьюдента с $n-m$ степенями свободы.

Если эта вероятность меньше уровня значимости α*, то принимается гипотеза о значимости соответствующего коэффициента регрессии.*

Из рис. 31 видно, что значимым коэффициентом является только коэффициент b_1.

Столбцы Нижние 95% и Верхние 95% - соответственно нижние и верхние интервалы для оцениваемых коэффициентов β_j.

Перейдем к следующей группе показателей, объединенных в таблице, показанной на рис. 33.

ВЫВОД ОСТАТКА			
Наблюдение	***Предсказанное Y***	***Остатки***	***Стандартные остатки***
1,0000	5,1273	-0,1273	-0,1518
2,0000	8,7903	1,2097	1,4425
3,0000	9,6442	0,3558	0,4243
4,0000	5,9813	1,0187	1,2148
5,0000	5,8614	-0,8614	-1,0272
6,0000	6,2285	-0,2285	-0,2724
7,0000	6,3483	-0,3483	-0,4153
8,0000	5,6142	-0,6142	-0,7324
9,0000	5,1273	0,8727	1,0406
10,0000	9,2772	-1,2772	-1,5229

Рис. 33. Фрагмент листа электронной таблицы Microsoft Excel результатом работы режима *Регрессия*

Столбец Наблюдение – содержит номера наблюдений.

Столбец Предсказанное У – значения $\hat{y}_i$, вычисленные по построенному уравнению регрессии.

Столбец Остатки – значения невязок $y_i - \hat{y}_i$

В заключении рассмотрения результатов работы режима *Регрессия* приведем график невязок (на рисунке 34 невязки названы остатками) $y_i - \hat{y}_i$ при заданных значениях только второй переменной. Наличие чередующихся положительных и отрицательных значений невязок является косвенным признаком *отсутствия систематической ошибки* (неучтенной независимой переменной) в построенном уравнении регрессии.

Рис. 35. График невязок как функция переменной X_2

4. НЕЛИНЕЙНАЯ МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

Эта тема включает выполнение лабораторной работы, посвященных построению нелинейной множественной регрессии на примере производственная функция Кобба-Дугласа.

4.1. Лабораторная работа № 13 Вычисление коэффициентов нелинейной множественной регрессии для производственная функция Кобба-Дугласа

Цель работы. Используя пространственную выборку таблицы 10 и команду *Поиск решения*, построить нелинейную множественную регрессию для производственная функция Кобба-Дугласа.

Таблица 10

Q_i	658	1201	2428	4258	8096	9850
L_i	163	246	453	715	1084	1565
K_i	280	1168	3070	5586	9120	13990

Ход работы.

Производственная функция Кобба-Дугласа имеет вид:

$$Q = A \cdot K^{\beta_1} \cdot L^{\beta_2}, \quad (6)$$

где $Q -$ объем производства, $K -$ затраты капитала, затраты труда. Показатели β_1, β_2 являются коэффициентами частной эластичности производства Q соответственно по затратам капитала K и труда L. Это означает, что при увеличении одних только затрат капитала (труда) на 1% объем производства увеличивается на β_1 % (β_2 %). При этом имеет место ограничение

$$\beta_1 + \beta_2 = 1.$$

Нахождение оценок B, b_1, b_2 для коэффициентов A, β_1, β_2 нелинейной модели (5) будем осуществлять из решения следующей задачи условной минимизации:

$$\min\left[\sum_{i=1}^{n}\left(Q_i - B \cdot K_i^{b_1} \cdot L_i^{b_2}\right)^2\right] \quad (7)$$

при ограничении $$b_1 + b_2 = 1. \quad (8)$$

Для решения этой задачи используем команду *Поиск решения.* Первоначально введем в столбцы A, B, C значения $K_i, L_i, Q_i,\ i=1,\ldots,6$ (см. рис. 36). Затем в ячейках B13, B14, B15 зададим начальные («стартовые») значения искомых коэффициентов: $B=2,\ b_1=0.5,\ b_2=0.5$.

После этого в соответствующих ячейках столбца D вычислим значения $\hat{Q}_i = B\cdot K_i^{b_1}\cdot L_i^{b_2}$. В столбце E запрограммируем вычисления значений $(Q_i-\hat{Q}_i)^2$, а в ячейке E13 вычислим значения функционала

$$F(B,b_1,b_2)=\sum_{i=1}^{6}(Q_i-\hat{Q}_i)^2. \qquad (9)$$

После этих подготовительных вычислений для выполнения команды «Поиск решения» необходимо обратиться к пункту основного меню **Сервис** и в появившемся меню щелкнуть мышью на команде ***Поиск решения***. Затем в появившемся диалоговом окне выполнить следующие действия (см. рис. 36):

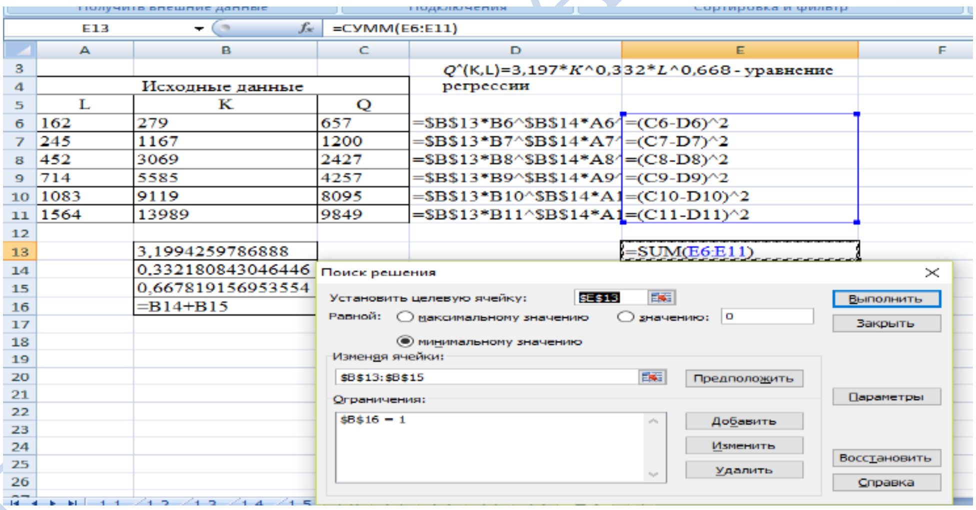


Рис. 37. Фрагмент листа электронной таблицы Microsoft Excel с решением задачи условной минимизации с помощью команды *Поиск решения*

- в поле ввода *Установить целевую ячейку* ввести адрес ячейки, в которой вычисляется значение минимизируемого функционала (в нашем примере – E13);
- включить опцию *Минимальное значение* (ищутся значения коэффициентов, при которых функционал достигает своего минимального значения);

•в поле ввода *Изменяя значения* ввести адреса ячеек, в которых находятся значения искомых коэффициентов (в нашем примере это ячейки B13:B15);

•щелкнув мышью на кнопке *Добавить* формируем ограничения на значения искомых коэффициентов (в нашем примере это условие (8)).

После задания параметров щелкаем на кнопке *Выполнить* и в ячейках B13, B14, B15 выводятся вычисленные значения коэффициентов, а в ячейке E13 – значение функционала (9) при этих значениях коэффициентов (см. рис. 37). Видно, что вычисленные значения коэффициентов $B = 3.197, b_1 = 0.332$, $b_2 = 0.668$ удовлетворяют ограничению (8).

Таким образом получено следующее уравнение регрессии:

$$\hat{Q}(K, L) = 3.197 \cdot K^{0.332} \cdot L^{0668}$$

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Елисеева И.И. Практикум по эконометрике [Текст]: учебное пособие / И.И. Елисеева, С.В. Курышева, Н.М. Гордеенко и др. – М.: Финансы и статистика, 2003. – 192 с.

2. Елисеева И.И. Эконометрика [Текст]: учебное пособие / И.И. Елисеева, С.В. Курышева, Н.М. Гордеенко и др. – М.: Финансы и статистика, 2002. – 344 с.

3. Тимофеева Н.Ю. Практикум по эконометрике (часть 1) / Н.Ю. Тимофеева. – Елец: Елецкий государственный университет им. И.А. Бунина, 2016. – 92 с.

4. Тимофеева Н.Ю. Практикум по построению Экономико-математических моделей управления производством / Н.Ю. Тимофеева. – Елец: Елецкий государственный университет им. И.А. Бунина, 2014. – 83 с.

5. Тимофеева Н.Ю. Практикум по построению Экономико-математических моделей прогнозирования деятельности предприятия (на основе нелинейного программирования) / Н.Ю. Тимофеева. – Елец: Елецкий государственный университет им. И.А. Бунина, 2015. – 53 с.

6. Тимофеева Н.Ю. Бюджетирование денежных средств предприятия с использованием моделей управления финансовым инвестиционным портфелем предприятия / Н.Ю. Тимофеева. – Елец: Елецкий государственный университет им. И.А. Бунина, 2016. – 120 с.

7. Эддоус М. Методы принятия решений [Текст]: учеб. пособие / М. Эддоус, Р. Стенсфилд. – М.: Аудит, ЮНИТИ, 2007.

8. Экономическое моделирование в Microsoft Excel [Текст]: учеб. пособие / Мур, Джеффри, Уэдерфорд, Ларри Р, и [др.]. – 6-е изд. / Пер. с англ. – М.: Издательский дом «Вильямс», 2004. – 1024 с.

Учебно-методическое издание

Наталья Юрьевна Тимофеева

ЛАБОРАТОРНЫЙ ПРАКТИКУМ ПО ЭКОНОМЕТРИКЕ
(ПРОДВИНУТЫЙ УРОВЕНЬ)

Учебно-методическое пособие

Технический редактор – О. А. Ядыкина
Книга издается в авторской редакции

Лицензия на издательскую деятельность
ИД № 06146. Дата выдачи 26.10.01.
Формат 60 х 84 /16. Гарнитура Times. Печать трафаретная.
Печ.л. 2,8 Уч.-изд.л. 2,6
Электронная версия

Размещено на сайте: http://elsu.ru/kaf/eeam/edu
Заказ 33

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Елецкий государственный университет им. И.А. Бунина»
399770, г. Елец, ул. Коммунаров, 28,1