

Эконометрика

Краткий курс лекций и порядок
выполнения практик
(для заочной формы обучения)

Преподаватель к.т.н., доцент

УРАЗБАХТИНА

Анжелика Юрьевна

Общие понятия эконометрики и эконометрических моделей

Эконометрика изучает конкретные количественные взаимосвязи экономических (производственных) объектов и процессов с помощью математических и статистических методов и моделей [1].

Взаимосвязи изучаются для выработки рекомендаций по управлению экономическими (производственными) системами/процессами или для выполнения прогнозов о состоянии этих систем/процессов в будущем.

В любой эконометрической (регрессионной) модели участвуют переменные:

- x – объясняющие, независимые, экзогенные переменные или факторные признаки. Это могут быть величины: погодные условия, цена, зарплата, процентные ставки, и пр. Может быть одна переменная x – тогда эконометрическая модель называется парной моделью регрессии; наличие x_1, x_2 и т.д. указывает на необходимость определения множественной модели регрессии. Если в качестве « x » выступает переменная «время», то ее обычно обозначают уже не как « x », а как « t ».
- $y(x)$ или $y(x_1, x_2, \dots)$ – результирующая, зависимая от x , эндогенная переменная или результативный признак. Это может быть: урожайность (зависит от погодных условий), уровень продаж (зависит от цены товара), доход по вкладу (зависит от процентных ставок), траты на товары не первой необходимости (зависит от зарплаты) и т.д. $y(t)$ – называется временным рядом. У временных рядов есть свойства, которые влияют на порядок формирования эконометрической модели временного ряда.

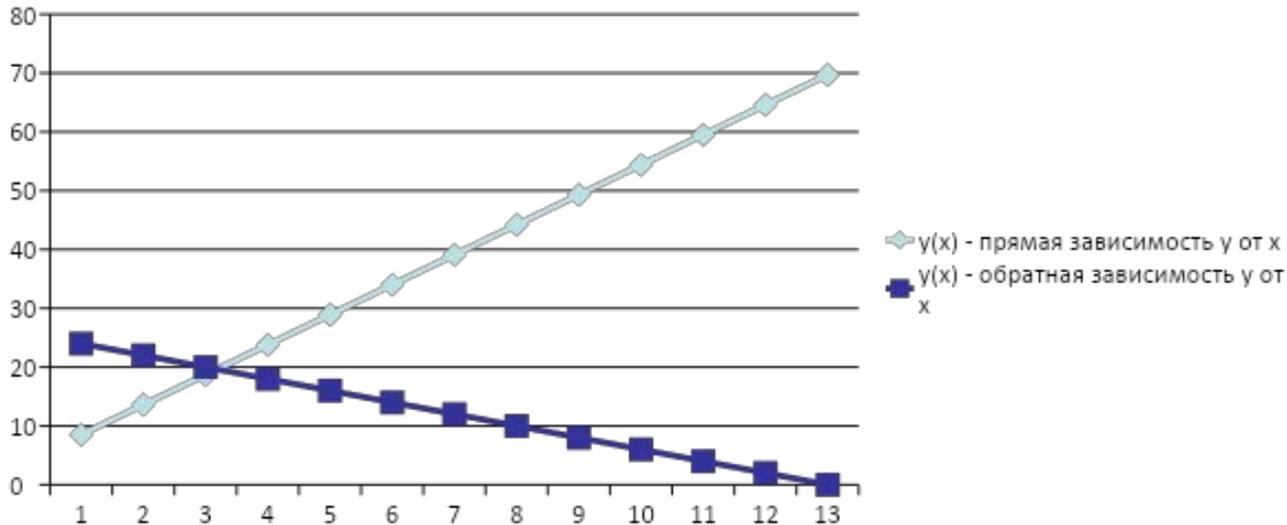
- Существуют также эконометрические модели, состоящие более чем из одного уравнения, их называют системами эконометрических уравнений.
- *Мы их рассмотрим позже.*

Линейная парная (эконометрическая) модель регрессии

- Регрессионный анализ (РА) - основной метод эконометрики.
- Основная задача РА заключается в исследовании зависимости результата работы некоторой системы y от различных факторов x (и/или t), и отображения этой зависимости в виде математической функции $y=f(x)$ или $y=f(t)$.
- Парная модель регрессии это зависимость между парой переменных y и x .
- Линейная парная модель регрессии – это использование линейной математической функции $y=f(x)=a+b*x$ или $y=f(t)=a+b*t$

Пример регрессионной линейной эконометрической модели

Запись модели в виде формулы $y(x)=a+b \cdot x$



Прямая зависимость: когда **x** возрастает, возрастает и **y** ; или **x** убывает – убывает и **y**; коэффициент корреляции положительный; тенденция **y** (тренд) к возрастанию.

Обратная зависимость: когда **x** возрастает, **y** - убывает ; или **x** убывает – **y** возрастает; коэффициент корреляции отрицательный; тенденция **y** (тренд) к убыванию.

Практика 1. Определение парной линейной модели регрессии, ее оценка и выполнение прогноза

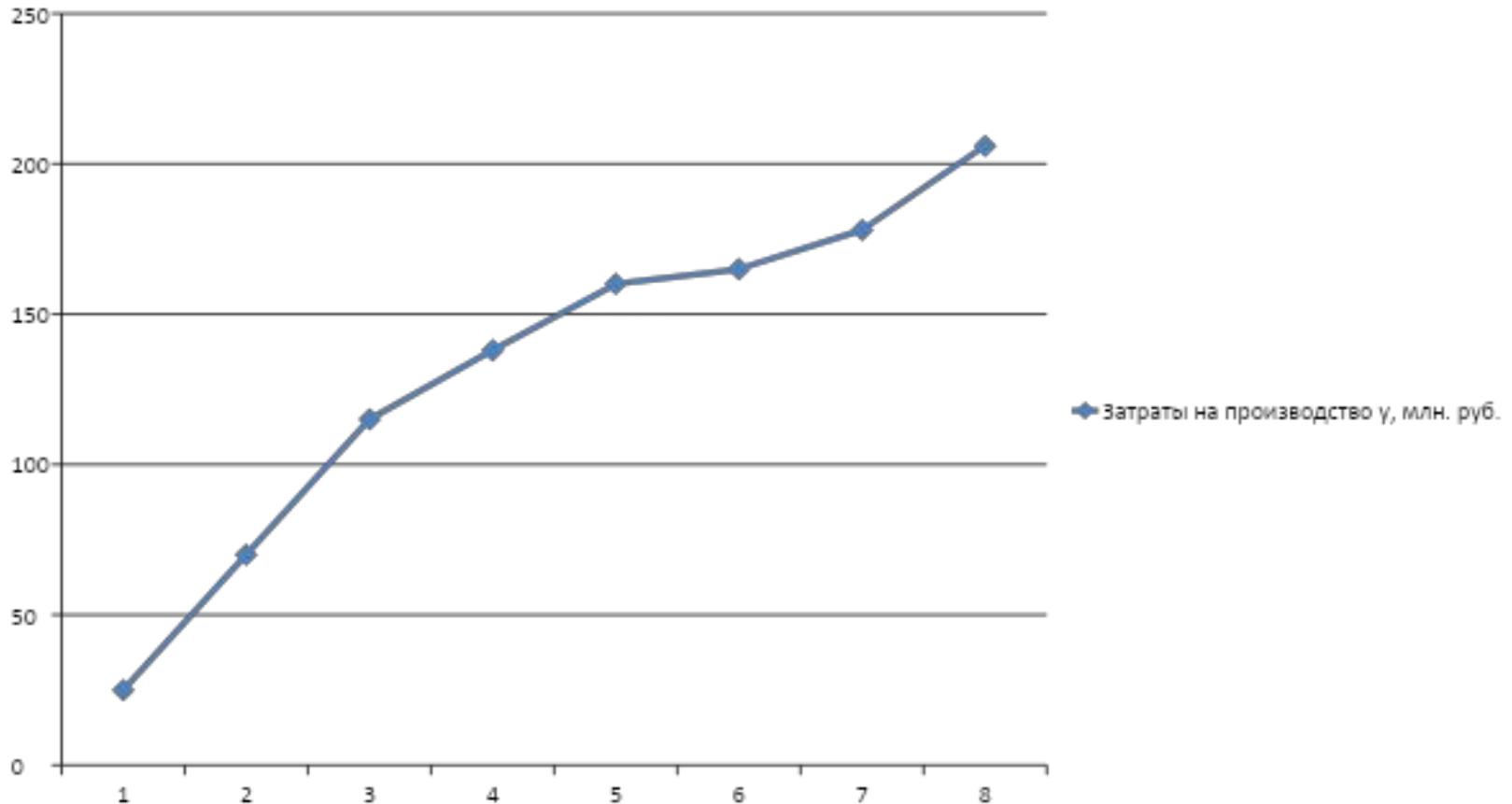
- 1. Пусть даны результаты статистических исследований

Затраты на производство y , млн. руб.	Выпуск продукции x , тыс. ед.
38	1
169	4,1
70	2
123	3,5
100	3
197	4,4
150	4
215	5

- 2. Сортировать всю заданную таблицу по возрастанию x .

Затраты на производство y , млн. руб.	Выпуск продукции x , тыс. ед.
38	1
70	2
100	3
123	3,5
150	4
169	4,1
197	4,4
215	5

3. Построить график y (убедиться, что график - почти прямая линия, только тогда правильно использовать **ЛИНЕЙНУЮ** эконометрическую модель $y=a+b*x$)



4. Определить степень зависимости **y** от **x** (оценить тесноту линейной связи) с помощью коэффициента корреляции ***r_{y,x}***.

Коэффициент парной линейной корреляции показывает силу связи между **y** и **x**.

Для нахождения этого коэффициента используем функцию MS Excel =КОРРЕЛ (все **y**; все **x**).

Если у вас установлен Open Office, то используйте функцию CORREL

		Затраты на производство у, млн. руб.	Выпуск продукции х, тыс. ед.			
1						
2	1	38	1			
3	2	70	2			
4	3	100	3			
5	4	123	3,5			
6	5	150	4			
7	6	169	4,1			
8	7	197	4,4			
9	8	215	5			
10						
11	Коэффициент корреляции $r_{y,x}$			=КОРРЕЛ(B2:B9;C2:C9)		
12						

Для качественной оценки коэффициента корреляции применяются различные шкалы, наиболее часто — шкала Чеддока. В зависимости от значения коэффициента корреляции связь может иметь одну из оценок:

- 0,1–0,3 — слабая;
- 0,3–0,5 — заметная;
- 0,5–0,7 — умеренная;
- 0,7–0,9 — высокая;
- 0,9–1,0 — весьма высокая.

7	6	169	4,1							
8	7	197	4,4							
9	8	215	5							
10	Сумма									
11	Среднее									
12						Комментарии:				
13	Коэффициент корреляции $r_{y,x}$			0,98	Весьма высокая зависимость между y и x					
14					Зависимость прямая					
15					Тенденция y к возрастанию					

- Модель определять целесообразно при $|r_{y,x}| > 0,5$.

Так как оценка тесноты связи с помощью коэффициента корреляции проводится, как правило, на основе более или менее ограниченной информации об изучаемом явлении, то возникает вопрос: насколько правомерно наше заключение по выборочным данным о наличии корреляционной связи в той генеральной совокупности, из которой была извлечена выборка?

В связи с этим и возникает необходимость **оценки существенности (значимости) линейного коэффициента корреляции**, дающая возможность распространить выводы по результатам выборки на генеральную совокупность. В зависимости от объема выборочной совокупности предлагаются различные методы оценки существенности линейного коэффициента корреляции.

Оценка значимости коэффициента корреляции при малых объемах выборки выполняется с использованием ***t*-критерия Стьюдента**. При этом фактическое (наблюдаемое) значение этого критерия определяется по формуле

$$t_{\text{набл}} = \sqrt{\frac{r_{y,x}^2}{1 - r_{y,x}^2}} (n - 2).$$

Вычисленное по этой формуле значение $t_{\text{набл}}$ сравнивается с критическим значением t -критерия, которое берется из таблицы значений t -критерия Стьюдента (см. Приложение) с учетом заданного уровня значимости α и числа степеней свободы $(n - 2)$.

Если $t_{\text{набл}} > t_{\text{табл}}$, то полученное значение коэффициента корреляции признается **значимым** (т.е. **нулевая гипотеза**, утверждающая равенство нулю коэффициента корреляции, **отвергается**). И таким образом делается вывод, что между исследуемыми переменными есть **тесная** статистическая взаимосвязь.

Если значение $r_{y,x}$ близко к нулю, связь между переменными **слабая**.

**Значения t -критерия Стьюдента (двуст
при уровнях значимости 0,10; 0,05**

Число степеней свободы	α			Число степеней свободы	0
	0,10	0,05	0,01		
1	6,3138	12,706	63,657	18	1,7
2	2,9200	4,3027	9,9248	19	1,7
3	2,3534	3,1825	5,8409	20	1,7
4	2,1318	2,7764	4,6041	21	1,7
5	2,0150	2,5706	4,0321	22	1,7
6	1,9432	2,4469	3,7074	23	1,7
7	1,8946	2,3646	3,4995	24	1,7

Буфер обмена		Шрифт		Выравнивание	
СУММ		=ЛИНЕЙН(B2:B9;C2:C9)			
	A	B	C	D	E
8	7	197	4,4		
9	8	215	5		
10	Сумма				
11	Среднее				
12					
13	Коэффициент корреляции $r_{y,x}$				0,98
14					
15					
16	a				
17	b	=ЛИНЕЙН(B2:B9;C2:C9)			

15		
16	a	
17	b	45,46
18		

- 5. Приступаем к нахождению неизвестных коэффициентов эконометрической модели:
- Сначала найдем b . Используем функцию =ЛИНЕЙН(все y ; все x).
- Если у вас Open Office используйте LINEST

- Теперь найдем **a**. Сначала найдем суммы **y** и **x**.
- Найдем среднее **x** ($x_{ср}$) и среднее **y** ($y_{ср}$)

	Затраты на производство у, млн. руб.	Выпуск продукции x, тыс. ед.
1	38	1
2	70	2
3	100	3
4	123	3,5
5	150	4
6	169	4,1
7	197	4,4
8	215	5
Сумма	1062	27
Среднее	132,75	3,375

- $a = y_{\text{ср}} - x_{\text{ср}} * b$

	A	B	C
5	4	123	3,5
6	5	150	4
7	6	169	4,1
8	7	197	4,4
9	8	215	5
10	Сумма	1062	27
11	Среднее	132,75	3,375
12			
13	Коэффициент корреляции $r_{y,x}$		
14			
15			
16		=B11-C11*B17	
17	b	45,46	

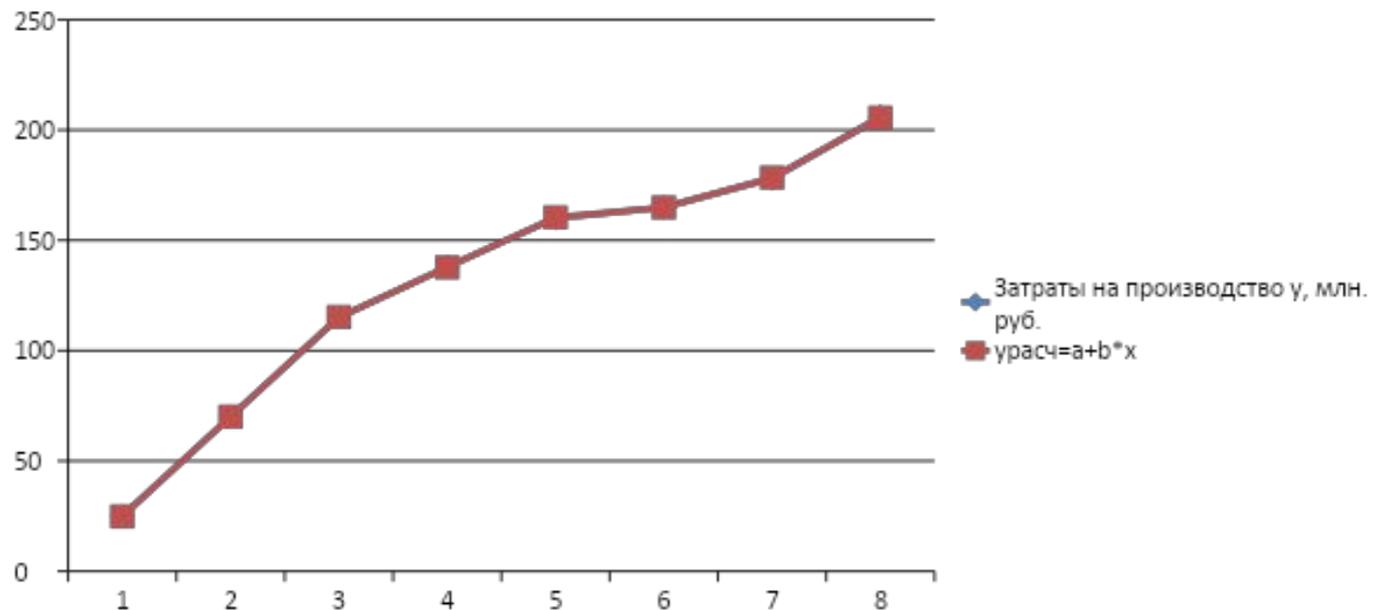
B16		fx	=B11-C11*B17
	A	B	
15			
16	a	-20,68	
17	b	45,46	

- 6. Вычислим значения в столбце **урасч** по формуле модели $a+b*x$
- Найдем сумму **урасч**

	B	C	D
Затраты на производство у, млн. руб.		Выпуск продукции х, тыс. ед.	урасч=a+b*x
10		2	=B\$16+\$B\$17*C2

	Затраты на производство у, млн. руб.	Выпуск продукции х, тыс. ед.	урасч=a+b*x
1	38	1	24,78
2	70	2	70,24
3	100	3	115,70
4	123	3,5	138,43
5	150	4	161,16
6	169	4,1	165,71
7	197	4,4	179,35
8	215	5	206,63
Сумма	1062	27	1062
Среднее	132,75	3,375	132,75

- 7. Построим на одном поле графики **у** и **урасч**.
Проведем первичную проверку качества модели:
 - 1) **сумма у** должна быть равна или очень близка к значению **СУММЫ урасч**
 - 2) график **у** и график **урасч** должны или совпадать (идеальная модель), или быть близки друг у к другу (чем больше совпадений/пересечений графиков – тем лучше модель)



8. Количественная характеристика модели по значению $E_{отн}$ – средней относительной ошибке аппроксимации:

$$E_{отн} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \cdot 100\%$$

Чем меньше рассеяние эмпирических точек вокруг теоретической линии регрессии, тем меньше средняя ошибка аппроксимации; $E_{отн} < 7\%$ свидетельствует о хорошем качестве модели.

После того как уравнение регрессии построено, выполняется проверка значимости построенного уравнения в целом и отдельных параметров.

- Так как $E_{отн}$ состоит из среднего от суммы дробей, эти дроби необходимо **ВЫЧИСЛИТЬ**

В	С	Д	Е
Затраты на производство у, млн. руб.	Выпуск продукции х, тыс. ед.	урасч= $a+b*x$	Дробь для $E_{отн}$ по формуле $ABS(y-урасч)*100/y$
38	1	24,78	$=ABS(B2-D2)*100/B2$
70	2	70,24	

	Затраты на производство у, млн. руб.	Выпуск продукции х, тыс. ед.	$у_{расч}=a+b*x$	Дробь для Еотн по формуле $ABS(y-у_{расч})*100/y$	
1	38	1	24,78	34,79	
2	70	2	70,24	0,34	
3	100	3	115,70	15,70	
4	123	3,5	138,43	12,55	
5	150	4	161,16	7,44	
6	169	4,1	165,71	1,95	
7	197	4,4	179,35	8,96	
8	215	5	206,63	3,90	
Сумма	1062	27	1062	85,63	
Средне е	132,75	3,375	132,75	10,70	<- Еотн

Средняя ошибка аппроксимации в данном примере 10,7%; что существенно больше допустимого значения в 7%.

Делаем вывод о плохом качестве модели.

Расчеты при таком выводе заканчивают.

Плохое качество модели может быть при:

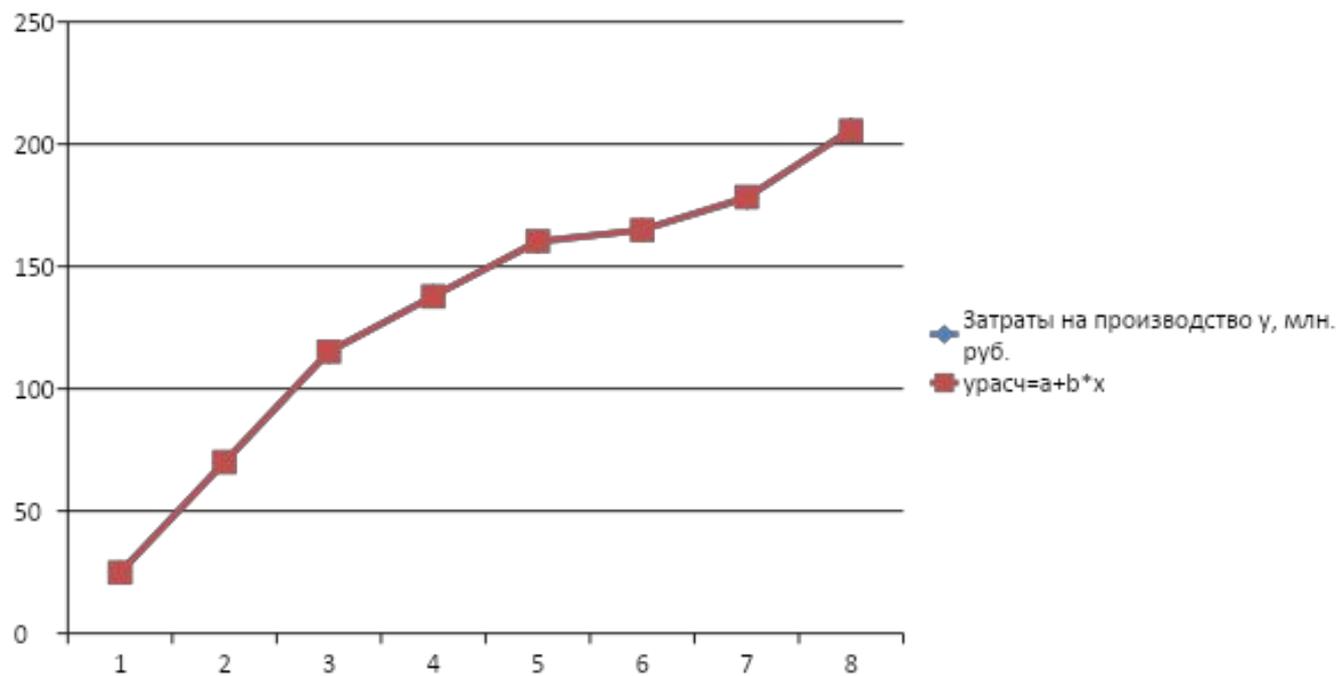
- 1) не правильном выборе формулы модели;
- 2) не корректных исходных данных x или y ;
- 3) не достаточном количестве наблюдений (для хорошей модели необходимо от 14 до 50 наблюдений)

- Изменим исходные данные. Лист EXCEL автоматически пересчитает все данные и построит новые графики.

	Затраты на производство у, млн. руб.	Выпуск продукции х, тыс. ед.	у _{расч} =a+b*x	Дробь для Еотн по формуле $\frac{ABS(y-y_{расч}) * 100}{y}$	
1	25	1	24,85	0,58	
2	70	2	70,02	0,03	
3	115	3	115,19	0,16	
4	138	3,5	137,77	0,17	
5	160	4	160,35	0,22	
6	165	4,1	164,87	0,08	
7	178	4,4	178,42	0,24	
8	206	5	205,52	0,23	
Сумма	1057	27	1057	1,71	
Среднее	132,125	3,375	132,125	0,21	<- Еотн

Коэффициент корреляции $r_{y,x}$					1,00
-------------------------------------	--	--	--	--	------

a	-20,31
b	45,17



Вычислим коэффициент Стьюдента $t_{набл}$ для вновь вычисленного коэффициента корреляции; получилось $t_{набл} > t_{табл}$, следовательно, значение коэффициента корреляции признаем значимым, и все выводы, основанные на нем – правильными.

Буфер обмена		Шрифт		Выравнивание		Число	
СУММ		✕ ✓ f_x		=КОРЕНЬ(E13^2*(8-2)/(1-E13^2))			
	A	B	C	D	E	F	G
12	Расчетный коэффициент Стьюдента $t_{набл}$				=КОРЕНЬ(E13^2*(8-2)/(1-E13^2))		

буфер обмена		шрифт		выравнивание	
J27		f_x			
	A	B	C	D	E
12	Расчетный коэффициент Стьюдента $t_{набл}$				478,17
13	Коэффициент корреляции $r_{y,x}$				0,99999
14	Коэффициент Стьюдента (табличный) t_{α}				2,4469
15	Стандартная ошибка S_e				0,3312

Эконометрическая линейная модель по новым данным практически идеальна (погрешность $E_{отн} = 0,21\%$), графики визуально совпадают. Такую модель можно использовать для прогнозирования.

- 9. Прогнозирование. Сначала определяем точку прогноза. Пусть требуется определить затраты на производство **упрогн**, если увеличить выпуск продукции до **хпрогн=6**.

	A	B
16	a	-20,31
17	b	45,17
18	хпрогн	6
19	упрогн	250,687678

Данный прогноз называется *точечным*. Значение независимой переменной $x_{\text{прогн}}$ не должно значительно отличаться от значений, входящих в выборку, по которой вычислено уравнение регрессии.

Вероятность реализации точечного прогноза теоретически равна нулю. Поэтому рассчитывается *средняя ошибка*, или *доверительный интервал*, прогноза с достаточно большой надежностью.

Доверительные интервалы зависят от следующих параметров:

- стандартная ошибка
- удаление $x_{\text{прогн}}$ от своего среднего значения \bar{x}
- количество наблюдений n ;
- уровень значимости прогноза α .

В частности, для прогноза будущие значения $y_{\text{прогн}}$ с вероятностью $(1 - \alpha)$ попадут в **доверительный интервал**

$$y_{\text{прогн}} \in \left[\hat{y}_{\text{прогн}} \pm S_e t_{\alpha} \sqrt{1 + \frac{1}{n} + \frac{(x_{\text{прогн}} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \right]$$

10. Вычисление доверительного интервала

10.1. Вычислить столбец значений $(x - x_{\text{ср}})^2$ и найти сумму этих значений

10.2. Определить по таблице (число степеней свободы для парной модели $n-2$; $\alpha = 0,05$) значение коэффициента Стьюдента t_{α}

10.3. Вычислить столбец ошибок в квадрате

$$e^2 = (y - \text{расч})^2$$

	Затраты на производство у, млн. руб.	Выпуск продукции х, тыс. ед.	урасч=a+b* х	Дробь для Еотн по формуле ABS(у-урасч)*100/ у		(х-хср)^2	е^2=(у-урасч)^2
1	25	1	24,85	0,58		5,64	0,02
2	70	2	70,02	0,03		1,89	0,00
3	115	3	115,19	0,16		0,14	0,04
4	138	3,5	137,77	0,17		0,02	0,05
5	160	4	160,35	0,22		0,39	0,13
6	165	4,1	164,87	0,08		0,53	0,02
7	178	4,4	178,42	0,24		1,05	0,18
8	206	5	205,52	0,23		2,64	0,23
Сумм а	1057	27	1057	1,71		12,30	0,66

10.4. Вычислить стандартную ошибку S_e

Для модели парной регрессии

$$S_e = \sqrt{\frac{1}{n-2} \sum_{i=1}^n e_i^2}$$

	A	B	C	D	E
14	Коэффициент Стьюдента t_α				2,4469
15	Стандартная ошибка S_e				0,3312

- Теперь имеются все данные для вычисления доверительного интервала U

буфер обмена | шрифт | выравнивание | число

СУММ $=E15*E14*КОРЕНЬ(1+1/8+(B18-C11)^2/G10)$

	A	B	C	D	E	F	G	H
10	Сумма	1057	27	1057	1,71		12,30	0,66
11	Среднее	132,125	3,375	132,125	0,21	<- E0тн		
12	Комментарии:							
13	Коэффициент корреляции $r_{y,x}$				1,00	Весьма высокая зависимость		
14	Коэффициент Стьюдента t_{α}				2,4469			
15	Стандартная ошибка Se				0,3312			
16	a	-20,31				В частности, для $\hat{y}_{пр}$ с вероятностью $(1 - \alpha)$		
17	b	45,17						
18	хпрогн	6						
19	упрогн	250,6877						
20	Доверительный интер	$=E15*E14*КОРЕНЬ(1+1/8+(B18-C11)^2/G10)$						
21					КОРЕНЬ(число)			

$$y_{\text{прогн}} \in \left[\hat{y}_{\text{пр}} \right]$$

8	хпрогн	6		
9	упрогн	250,6877		
0	Доверительный интервал U			1,05

Результат практики 1:

Определили линейную парную
эконометрическую модель

$у_{расч} = -20,31 + 45,17x$; модель хорошего
качества с относительной средней
погрешностью в 0,21%;

По модели выполнен прогноз: при
увеличении $x_{прогн}$ до 6; $у_{прогн}$
увеличится до $250,69 \pm 1,05$