

## МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ

по выполнению контрольных заданий  
по дисциплине «Эконометрика»  
для студентов заочного отделения  
экономического факультета

*Регрессионный анализ* заключается в определении аналитического выражения связи зависимой переменной  $y$  (называемой также *результативным признаком*) с независимыми переменными  $x_1, x_2, \dots, x_m$  (называемыми также *факторами*).

Форма связи результативного признака  $y$  с факторами  $x_1, x_2, \dots, x_m$  получила название *уравнения регрессии*. В зависимости от типа выбранного уравнения различают *линейную* и *нелинейную* регрессию (в последнем случае возможно дальнейшее уточнение: квадратичная, экспоненциальная, логарифмическая и т. д.).

В зависимости от числа взаимосвязанных признаков различают *парную* и *множественную* регрессию. Если исследуется связь между двумя признаками (результативным и факторным), то регрессия называется *парной*, если между тремя и более признаками - *множественной* (*многофакторной*) регрессией.

При изучении регрессии следует придерживаться определенной последовательности этапов:

1. Задание аналитической формы уравнения регрессии и определение параметров регрессии.
2. Определение в регрессии степени стохастической взаимосвязи результативного признака и факторов, проверка общего качества уравнения регрессии.
3. Проверка статистической значимости каждого параметра уравнения регрессии и определение их доверительных интервалов.

Основное содержание выделенных этапов рассмотрим на примере множественной линейной регрессии, реализованной в режиме «Регрессия» надстройки *Пакет анализа* Microsoft Excel.

Этап 1. Уравнение линейной парной регрессии имеет вид

$$y = a + bx + e$$

или

$$\hat{y} = a + bx,$$

где  $y$  – эмпирические (фактические) значения зависимой переменной;

$\hat{y}$  – теоретические значения зависимой переменной, получаемые подстановкой в уравнение регрессии соответствующих значений независимой переменной;

$e$  – случайная составляющая зависимой переменной (остаток), характеризующая влияние неучтенных факторов;

$x$  – значения независимой переменной;

$a$  и  $b$  – параметры уравнения регрессии.

Параметры уравнения регрессии могут быть определены *методом наименьших квадратов* (именно этот метод и используется в Microsoft Excel). Сущность данного метода заключается в нахождении параметров модели, при которых минимизируется сумма квадратов отклонений эмпирических (фактических) значений результативного признака от теоретических, полученных по выбранному уравнению регрессии, т.е.

$$S = \sum_{i=1}^n e^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - a - bx_i)^2 \rightarrow \min,$$

где  $n$  – число наблюдений.

Рассматривая  $S$  в качестве функции параметров  $a$  и  $b$  проводя математические преобразования (дифференцирование), получаем систему нормальных уравнений с 2 неизвестными (по числу параметров):

$$\begin{aligned} na + b \sum x &= \sum y; \\ a \sum x + b \sum x^2 &= \sum xy. \end{aligned}$$

Решив систему уравнений, находим значения параметров являющихся коэффициентами искомого теоретического уравнения регрессии:

$$\begin{aligned} a &= \frac{\sum y - b \sum x}{n} = \bar{y} - b\bar{x}, \\ b &= \frac{Cov(xy)}{s_x^2} = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\bar{x}^2 - \bar{x}^2} \end{aligned}$$

Этап 2. Для определения тесноты стохастической взаимосвязи результативного признака  $y$  и фактора  $x$  необходимо знать следующие дисперсии:

*общую дисперсию* зависимой переменной  $y$ :

$$s_y^2 = \frac{\sum (y - \bar{y})^2}{n - 1};$$

*факторную дисперсию* зависимой переменной, отображающую влияние независимой переменной  $X$ :

$$s_{\text{рег}}^2 = \sum (\hat{y} - \bar{y})^2;$$

*остаточную дисперсию*:

$$s_{\text{ост}}^2 = \frac{\sum (y - \hat{y})^2}{n - 2} = S^2.$$

Долю дисперсии, объясняемую регрессией, в общей дисперсии зависимой переменной  $y$ , характеризует коэффициент детерминации  $R^2$ :

$$R^2 = \frac{\sum (\hat{y} - \bar{y})^2}{\sum (y - \bar{y})^2} = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}.$$

Коэффициент детерминации – квадрат коэффициента парной корреляции  $r_{xy}$ :

$$r_{xy} = b \frac{s_x}{s_y} = \frac{Cov(xy)}{s_x s_y} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{s_x s_y}.$$

Так как в основе эконометрического метода лежит представление об анализируемой совокупности как о некоторой случайной выборке, возникает вопрос об адекватности построенного уравнения генеральным данным. Для этого проводится проверка статистической значимости коэффициента детерминации  $R^2$  на основе  $F$ -критерия Фишера:

$$F = \frac{s_{рег}^2}{s_{ост}^2} = \frac{R^2}{1 - R^2} (n - 2).$$

Гипотеза  $H_0: R^2 = 0$  о статистической незначимости коэффициента детерминации  $R^2$  (нулевая гипотеза) отвергается, если  $F > F_{\alpha, крит}$  при выбранном уровне значимости  $\alpha$ . В противном случае нулевая гипотеза не отвергается. Уровень значимости это вероятность отклонения правильной гипотезы. Обычно  $\alpha$  принимается на уровне 0,05; 0,01 или 0,10.

Этап 3. Для оценки статистической значимости параметров регрессии для каждого из них рассчитываются значения  $t$ -статистики Стьюдента и доверительные интервалы. Оценка значимости параметров регрессии с помощью  $t$ -критерия Стьюдента проводится по формулам:

$$t_a = \frac{a}{s_a};$$

$$t_b = \frac{b}{s_b}.$$

Здесь  $s_a$  и  $s_b$  – стандартные ошибки параметров регрессии, определяемые по формулам:

$$s_b = \frac{S}{\sqrt{\sum (x - \bar{x})^2}};$$

$$s_a = S \cdot \frac{\sqrt{\sum x^2}}{\sqrt{n \cdot \sum (x - \bar{x})^2}};$$

где  $S$  – стандартная ошибка уравнения регрессии, определяемая как

$$S = \sqrt{\frac{(y - \hat{y})^2}{n - 2}}.$$

Сравнивая фактическое и критическое значения  $t$ -статистики, отвергаем нулевую гипотезу при  $t > t_{\alpha, \text{крит}}$  и принимаем в противоположном случае.

Связь между  $F$ -критерием Фишера и  $t$ -критерием Стьюдента выражается равенством

$$t_b = \sqrt{F}.$$

Доверительные интервалы параметров регрессии определяются следующим образом.

$$a \pm \Delta a = a \pm t_{\alpha} s_a,$$

$$b \pm \Delta b = b \pm t_{\alpha} s_b.$$

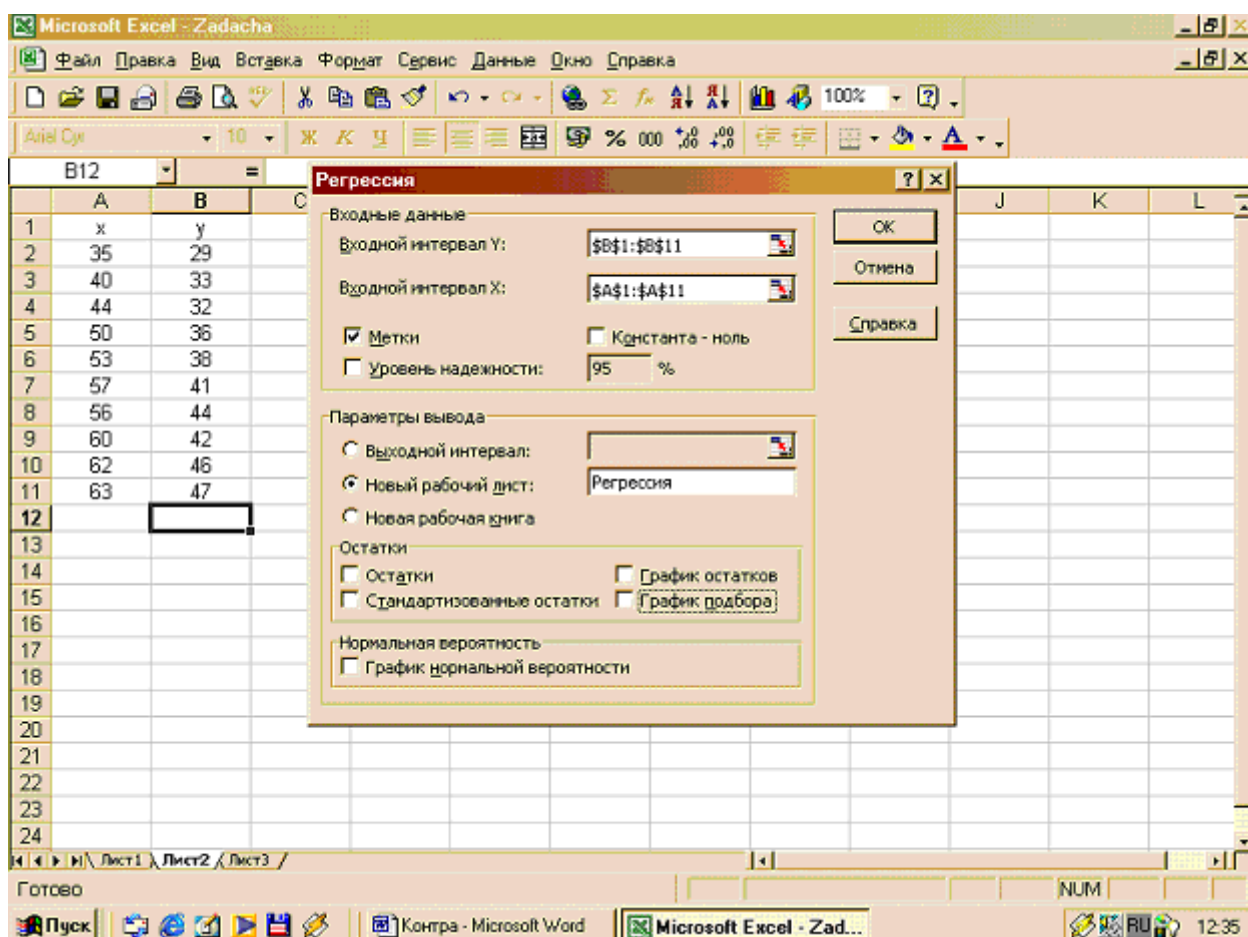
Здесь  $t_{\alpha}$  – значение  $t$ -статистики для выбранного уровня значимости  $\alpha$ . Величина  $p=1-\alpha$  называется доверительной вероятностью или уровнем надежности, нередко выражаемым в процентах. Это показатель, характеризует вероятность того, что теоретическое значение параметра регрессии будет находиться в полученном доверительном интервале.

Рассмотрим пример решения модели парной линейной регрессии с помощью надстройки «Пакет анализа» табличного процессора *Microsoft Excel*.

Имеются данные о цене акции в у. е. ( $y$ ) и объеме продаж в млн. у. е. ( $x$ ) одной из корпораций:

$x$	35	40	44	50	53	57	56	60	62	63
$y$	29	33	32	36	38	41	44	42	46	47

Введем их в таблицу *Excel* и выберем пункт меню «Регрессия» надстройки «Пакет анализа». Режим работы «Регрессия» служит для расчета параметров уравнения линейной регрессии и проверки его адекватности исследуемому процессу.



В диалоговом окне данного режима задаются следующие параметры:

1. *Входной интервал Y*— вводится ссылка на ячейки, содержащие данные по результативному признаку. Диапазон должен состоять из одного столбца.
2. *Входной интервал X*— вводится ссылка на ячейки, содержащие факторные признаки. Максимальное число входных диапазонов (столбцов) равно 16.
3. *Метки* — установленный флажок уведомляет о том, что в первой строке записаны названия переменных.
4. *Уровень надежности* — установите данный флажок в активное состояние, если в поле, расположенное напротив флажка, необходимо ввести уровень надежности, отличный от уровня 95%, применяемого по умолчанию. Установленный уровень надежности используется для проверки значимости коэффициента детерминации  $R^2$  и коэффициентов регрессии  $a$  и  $b$ . (Примечание: при неактивном флажке *Уровень надежности* в таблице параметров уравнения регрессии генерируются две одинаковые пары столбцов для границ доверительных интервалов).
5. *Константа-ноль* — установите данный флажок в активное состояние, если требуется, чтобы линия регрессии прошла через начало координат (т. е.  $a = 0$ ).
5. *Выходной интервал/Новый рабочий лист/Новая рабочая книга* — впишите имя рабочего листа, куда будут записаны результаты.
6. *Остатки* - установите данный флажок в активное состояние, если требуется включить в выходной диапазон столбец остатков.
7. *Стандартизованные остатки* — установите данный флажок в активное состояние, если требуется включить в выходной диапазон столбец стандартизованных остатков.
8. *График остатков* - установите данный флажок в активное состояние, если требуется вывести график остатков.

9. *График подбора* - установите данный флажок в активное состояние, если требуется вывести график подбора.

10. *График нормальной вероятности* - установите данный флажок в активное состояние, если требуется вывести график нормальной вероятности.

Результаты регрессионного анализа и соответствующие графики приведены ниже.

The screenshot shows the 'Data Analysis ToolPak' results in Excel. The 'РЕГРЕССИОННЫЙ АНАЛИЗ' (Regression Analysis) output is displayed on 'Лист1' (Sheet1). The results are organized into two main sections: 'РЕГРЕССИОННАЯ СТАТИСТИКА' (Regression Statistics) and 'ДИСПЕРСИОННЫЙ АНАЛИЗ' (ANOVA).

РЕГРЕССИОННАЯ СТАТИСТИКА						
Регрессионная статистика						
Множественный R	0,962731526					
R-квадрат	0,926851992					
Нормированный R-квадрат	0,917708491					
Стандартная ошибка	1,777637185					
Наблюдения	10					

ДИСПЕРСИОННЫЙ АНАЛИЗ						
	df	SS	MS	F	Значимость F	
Регрессия	1	320,3200483	320,3200483	101,3672976	8,06841E-06	
Остаток	8	25,27995169	3,159993961			
Итого	9	345,6				

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%
Y-пересечение	6,457004831	3,261225721	1,979931898	0,08305747	-1,06340003	13,97740961
x	0,621980676	0,061777163	10,06813278	8,06841E-06	0,479522192	0,76443916

Полученная регрессионная модель может быть записана следующим образом:

$$y = 6,457 + 0,622x + e; \quad R^2 = 0,927; \quad F = 101,367.$$

(3,261) (0,062)

Анализ результатов свидетельствует о наличии сильной регрессионной зависимости цены акции от объема продаж, а также статистической значимости параметра регрессии  $b$  и коэффициента детерминации  $R^2$ .

Прогнозное значение зависимой переменной  $y^*$  определяется путем подстановки в уравнение регрессии прогнозного значения  $x^*$ . Вычисляется стандартная ошибка прогноза  $\Delta y$ :

$$\Delta y = s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x - \bar{x})^2}},$$

и строится доверительный интервал прогноза  $(y^* - t_\alpha \Delta y; y^* + t_\alpha \Delta y)$ .

## КОНТРОЛЬНЫЕ ЗАДАНИЯ

### ЗАДАНИЕ 1

В предложенных задачах приведены результаты наблюдения за факторами  $x$  и результатами  $y$ . Требуется исследовать регрессионные зависимости  $y = f(x)$  в соответствии со следующей схемой:

1. Рассчитайте параметры линейной регрессии.
2. Оцените тесноту связи с помощью показателей корреляции и детерминации.
3. Оцените статистическую надежность результатов регрессионного моделирования с помощью F-критерия Фишера и коэффициентов регрессии с помощью t-критерия Стьюдента.
4. Рассчитайте прогнозное значение результата, если прогнозное значение фактора увеличится на 5% от среднего уровня. Определите доверительный интервал прогноза для уровней значимости  $\alpha = 0,1$ ;  $\alpha = 0,05$  и  $\alpha = 0,01$ .
5. Оформите результаты по каждому пункту в пояснительной записке. Сформулируйте выводы по проведенному регрессионному анализу.

#### Вариант 1

Администрация страховой компании приняла решение о введении нового вида услуг – страхование на случай пожара. С целью определения тарифов по выборке из 10 случаев пожаров анализируется зависимость стоимости ущерба, нанесенного пожаром от расстояния до ближайшей пожарной станции:

у - сумма ущерба, млн. руб.	26,2	17,8	31,3	23,1	27,5	36,0	14,1	22,3	19,6	31,3
х - Расстояние до ближайшей станции, км	3,4	1,8	4,6	2,3	3,1	5,5	0,7	3,0	2,6	4,3

#### Вариант 2

Торговцу нужно выяснить, как изменяется количество пучков салата, продаваемого ежедневно в розницу.

Имеются следующие сведения о количестве и цене.

у - Кол-во продаж за день	21	28	29	34	35	37	37	41	43	46
х - Цена, (у.е. за 1 пучок)	0,38	0,30	0,31	0,25	0,26	0,22	0,24	0,16	0,15	0,12

#### Вариант 3.

Проведено исследование, сколько сберегает население и сколько оно зарабатывает за год. Были получены следующие данные для случайно отобранных 12 чел.:

х - Доходы, тыс. у.е.	90	18	36	18	120	66	84	60	38	16	100	72
у – Сбереже- ния, у.е.	2000	200	500	500	2500	1800	1500	1500	600	400	1500	1600

#### Вариант 4

Требуется установить и исследовать зависимость между доходами на душу населения и индексом человеческого развития (ИЧР) на основании следующих данных:

Страна	х Душевой доход, долл.	у Индекс человеческого развития (ИЧР),
Объединенные Арабские Эмираты	1600	0,866
Таиланд	7100	0,833
Уругвай	6750	0,883
Ливия	6130	0,801
Колумбия	6110	0,848
Иордания	4190	0,730
Египет	3850	0,514
Марокко	3680	0,566
Перу	3650	0,717
Шри-Ланка	3280	0,711
Филиппины	2680	0,672
Боливия	2600	0,589
Китай	2600	0,626
Зимбабве	2200	0,513
Пакистан	2150	0,445
Уганда	1370	0,328
Нигерия	1350	0,393
Индия	1350	0,446



**Вариант 5.**

Для установления уровня ежедневных издержек, бухгалтер фирмы собрал следующие данные:

х - Выпуск, тыс. шт. в день	12	19	18	12	17	15	15	17	13	16	11	18	19
у - Издержк и, тыс. у.е. в день	3,2	4,1	3,8	2,9	3,8	3,6	3,5	3,9	3,3	3,7	2,8	4,0	4,2

**Вариант 6.**

Изучается влияние стоимости оборотных средств на величину валового дохода торговых предприятий. Для этого по 12 торговым предприятиям были получены данные, приведенные в таблице. Построить и исследовать регрессионную модель.

у Валовой доход за год, млн. руб.	х Среднегодовая стоимость оборотных средств, млн. руб.
203	105
63	56
45	54
113	63
121	28
88	50
110	54
56	42
80	36
237	106
160	88
75	46

**Вариант 7.**

Туристическую фирму крупного курортного города интересует связь между числом отпусков, останавливающихся в отелях и расходами на рекламу отелей. Взято случайное количество отелей – 10, сходных по размеру. Была собрана следующая информация за текущий сезон:

Отель	1	2	3	4	5	6	7	8	9	10
х Реклама, у.е.	9000	6000	10000	8000	7000	4000	3000	5000	8000	2000
у - Число гостей, чел.	1100	1200	1600	1300	1100	800	700	1100	1200	1000

**Вариант 8**

В компании 12 магазинов. Финансовый директор группы магазинов рассматривает возможность слияния мелких магазинов для увеличения прибыльности компании. Он предложил, что оборот магазинов вследствие слияния останется прежним. Постройте и исследуйте регрессионную модель. Ответьте на вопрос о целесообразности слияния мелких магазинов.

Магазин	1	2	3	4	5	6	7	8	9	10	11	12
Y - Годовая прибыль, млн. руб.	2	4	11	17	18	28	34	36	48	55	71	85
X - Оборот, млн. руб.	50	60	85	85	100	120	140	155	180	210	250	300

**Вариант 9.**

Изучается влияние стоимости внеоборотных активов на величину валового дохода торговых предприятий. Для этого по 12 торговым предприятиям были получены данные, приведенные в таблице. Построить и исследовать регрессионную модель.

Y- Валовой доход за год, млн. руб.	203	63	45	113	121	88	110	56	80	237	160	75
X- Среднегодовая стоимость, внеоборотных активов, млн.руб.	118	28	17	50	56	102	116	124	114	154	115	98

**Вариант 10.**

Требуется установить и исследовать зависимость между доходами на душу населения и индексом человеческой бедности (ИЧБ) на основании следующих данных:

Страна	x Душевой доход, долл.	y Индекс человеческой бедности (ИЧБ)
Объединенные Арабские Эмираты	1600	14,9
Таиланд	7100	11,7
Уругвай	6750	11,7
Ливия	6130	18,8
Колумбия	6110	10,7
Иордания	4190	10,9
Египет	3850	34,8
Марокко	3680	41,7
Перу	3650	22,8
Шри-Ланка	3280	20,7
Филиппины	2680	17,7
Боливия	2600	22,5
Китай	2600	17,5
Зимбабве	2200	17,3
Пакистан	2150	46,8
Уганда	1370	41,3
Нигерия	1350	41,6
Индия	1350	36,7