Министерство образования Российской Федерации

Магнитогорский Государственный технический университет
имени Г. И. Носова

Кафедра технологий обработки материалов

ПАРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

Методические указания

Магнитогорск

2017

1.ЦЕЛИ РАБОТЫ

В реальных условиях функционирования технических объектов и организационно-технических систем зависимость результатов функционирования (откликов) от управляемых и контролируемых воздействий (факторов) проявляется как опосредованная разнообразными случайными причинами (возмущениями). Подобные зависимости принято называть стохастическими.

Парный регрессионный анализ – это метод математической статистики, который позволяет найти наиболее точное и достоверное отображение (модель, аппроксимацию) стохастической зависимости между откликом Y и одним из факторов X.

Цели работы:

Изучение методики парного регрессионного анализа.

Приобретение навыков решения задачи парного регрессионного анализа в среде электронных таблиц *MS Excel* с применением статистических функций.

Приобретение навыков решения задачи парного регрессионного анализа в среде электронных таблиц *MS Excel* с применением инструмента «Линия тренда».

2.КРАТКИЕ СВЕДЕНИЯ ИЗ СТАТИСТИКИ

2.1.Содержание и допущения регрессионного анализа

Понятие стохастической зависимости в некотором смысле является обобщением понятия о зависимости функциональной. Для последней характерно, что каждому значению фактора (аргумента) соответствует совершенно определенное значение отклика. В случае стохастической зависимости при определенном значении x_i фактора X может наблюдаться множество значений отклика Y. В производственных условиях фактор также является случайной величиной, но при проведении регрессионного анализа полагают, что всякое его значение x_i неслучайно [1].

При проведении регрессионного анализа принимают следующие допущения [1, 2, 3].

1. Фактор измеряется с пренебрежимо малой ошибкой по сравнению с ошибкой определения отклика. Большая ошибка y_i объясняется наличием в каждом из наблюдений влияний на отклик нерегламентированных параметров.

2. Каждому значению фактора соответствуют значения отклика, представляющие собой независимые и нормально распределенные величины.
3. При проведении эксперимента с объемом выборки n при условии, что каждый опыт произведен m раз, выборочные дисперсии отклика однородны.

Учитывая возможные отклонения, модель связи некоторого значения отклика с соответствующим значением фактора может быть представлена в виде двух составляющих:

$$y_i = \varphi(x_i) + \varepsilon_i \,, \tag{2.1}$$

где $\varphi(x_i)$ - систематическая (объясненная) составляющая. Она обусловлена существованием зависимости между откликом и фактором;

ε_i - случайная составляющая. Она обусловлена разнообразными возмущениями и вызывает отклонения y_i от соответствующих реальной зависимости.

Относительно ε_i делают следующие предположения:

1. Это нормально распределенная случайная переменная.
2. $\mu(\varepsilon_i) = 0$ (математическое ожидание случайной составляющей равно нулю).
3. $\sigma(\varepsilon_i) = Const$ (дисперсия случайной составляющей постоянна).
4. В различных наблюдениях значения ε_i не зависят друг от друга.

Для построения парной регрессионной модели (иначе – парного уравнения регрессии или просто парной регрессии) необходимо решить две задачи - определить вид уравнения регрессии и оценить допустимость отображения исследуемой зависимости выбранным уравнением регрессии.

2.2.Определение вида уравнения парной регрессии

Задача определения вида уравнения регрессии состоит в определении систематической составляющей $\varphi(x)$. Однако, как уже указывалось ранее, истинные параметры (коэффициенты) этого уравнения не могут быть определены, поскольку используются выборки ограниченного объема ($n <<< \infty$). Поэтому могут быть найдены лишь оценки истинных параметров и действительная связь между откликом и фактором $y = \varphi(x)$ представляется оценкой (отображением) этой связи $\hat{y} = \hat{\varphi}(x)$. Именно данная оценка и является уравнением регрессии.

Для подбора уравнения $\hat{y} = \hat{\varphi}(x)$, которое наилучшим образом отображает стохастическую связь между откликом и фактором, используют метод наименьших квадратов (МНК). Согласно МНК наилучшей оценкой исследуемой зависимости является та, которая дает наименьшую сумму квадратов отклонений наблюдаемых значений отклика y_i от рассчитанных по уравнению регрессии $\hat{y}_i$ при тех же значениях фактора x_i (рис. 1). Это условие выражается следующим образом:

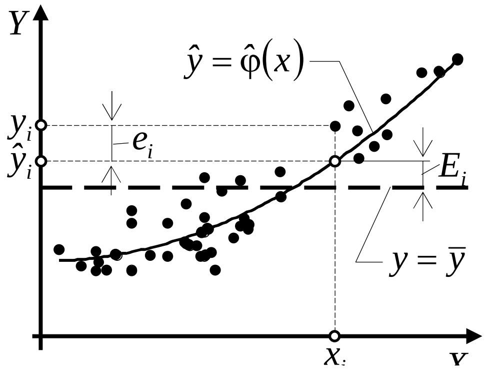


Рис. 1. Линия регрессии и ее оценивание

$$\sum_{i=1}^{n} e_i^{\,2} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = min. \qquad (2.2)$$

Оценки, полученные МНК, являются несмещенными, состоятельными и эффективными. Несмещенность означает, что математическое ожидание каждого из оцениваемых параметров равно соответствующему истинному значению. Состоятельность означает, что с увеличением числа наблюдений n оценки параметров все более концентрируются вокруг истинных значений (т. е. с возрастанием n дисперсии оценок стремятся к нулю). Эффективность означает, что оценки, полученные МНК, обладают наименьшей дисперсией по сравнению с оценками этих же параметров, полученными другими методами.

Задача определения коэффициентов уравнения регрессии сводится практически к определению минимума функции нескольких переменных и решена математической статистикой для линейного уравнения

$$y = \beta_0 + \beta_1 x \approx b_0 + b_1 x. \qquad (2.3)$$

Оценки коэффициентов $b_0 \approx \beta_0$ и $b_1 \approx \beta_1$ вычисляются по формулам [1, 2]:

$$b_1 = \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i - n \sum_{i=1}^{n} x_i y_i}{\left(\sum_{i=1}^{n} x_i\right)^2 - n \sum_{i=1}^{n} x_i^2}; \qquad (2.4)$$

$$b_0 = \frac{1}{n}\left(\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i\right). \qquad (2.5)$$

В *MS Excel* коэффициенты b_0 и b_1 линейной регрессии можно определить с использованием статистических функций НАКЛОН(OY;OX) и ОТРЕЗОК(OY;OX). Аргументы OY и OX - ссылки на ячейки, где записаны значения отклика и фактора соответственно.

При необходимости вычисления коэффициентов уравнений регрессии других видов поступают следующим образом [2, 3]. Сначала по экспериментальным данным x_i и y_i находят значения x_i' и y_i', которые соответствуют уравнению данного вида после его линеаризации. Далее используют формулы (2.4) и (2.5), расчеты по которым дадут коэффициенты именно лианеризованного уравнения (b_0' и b_1'). После этого находят коэффициенты реальных уравнений. Формулы для таких преобразований представлены в прил. 1.

2.3.Оценивание уравнения парной регрессии

Из различных уравнений регрессии наилучшим в смысле МНК считают то, которое обеспечивает минимум дисперсии фактических (полученных экспериментально) значений отклика относительно линии регрессии. Эту дисперсию называют остаточной или дисперсией относительно регрессии и определяют по формуле:

$$S_e^2 = \frac{1}{n-k} \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 , \qquad (2.6)$$

где k - число коэффициентов регрессии в уравнении.

Точность отображения (аппроксимации) исследуемой зависимости выбранным уравнением регрессии оценивают с помощью дисперсионного анализа. Для этого сравнивают дисперсию относительно линии регрессии (S_e^2) с оценкой дисперсии значений y_i относительно выборочного среднего фактических значений отклика $\bar{y}$:

$$S_E^2 = \frac{1}{k-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = \frac{1}{m}\sum_{i=1}^{n}\left(y_i - \frac{1}{n}\sum_{i=1}^{n} y_i\right)^2 . \qquad (2.7)$$

Величина S_E^2 характеризует (рис. 1) рассеяние y_i, обусловленное зависимостью отклика от факторов, и поэтому называется объясненной дисперсией. Остаточная дисперсия S_e^2 характеризует рассеяние y_i, вызванное случайными воздействиями (возмущениями). Очевидно, что связь между откликом и факторами в виде данного уравнения регрессии существует, если объясненная дисперсия существенно больше остаточной.

Чтобы выяснить, можно ли считать отличие рассматриваемых дисперсий существенным, выдвигают нулевую гипотезу об их равенстве $H_0 : S_E^2 = S_e^2$ и проверяют ее с использованием числа Фишера:

$$F = \frac{S_E^2}{S_e^2} . \qquad (2.8)$$

Гипотеза считается справедливой, если рассчитанное число Фишера не превышает табличного значения $F\left[\alpha, \nu_1, \nu_2\right]$ для заданного уровня значимости α. При выборе табличного значения важно помнить, что в данном случае $\nu_1 = \nu_E = k-1$ и

$\nu_2 = \nu_e = n - k$. В *MS Excel* табличное число Фишера может быть найдено с применением статистической функции:

FРАСПОБР(α ; $k-1$; $n-k$).

Справедливость гипотезы о равенстве остаточной и объясненной дисперсий означает, что выбранное уравнение регрессии нельзя принимать в качестве модели связи между откликом и фактором.

Если же $F > F[\alpha, k-1, n-k]$, то объясненная дисперсия существенно больше остаточной. А это означает, что между откликом и факторами существует взаимосвязь, которую с вероятностью α допустимо аппроксимировать рассматриваемым уравнением регрессии.

Из нескольких допустимых аппроксимаций наиболее точной, очевидно, будет та, для которой значение S_e^2 является наименьшим. Отсюда следует, что для наиболее точной модели $\hat{y}_i = \hat{\varphi}(x_i)$ различие расчетного и табличного чисел Фишера будет максимальным.

Во многих программных средствах, содержащих опции обработки данных (в том числе и в *MS Excel*), для оценивания качества аппроксимации предлагается параметр R^2 (коэффициент детерминации):

$$R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \, \frac{n-1}{n-k}. \qquad (2.9)$$

Коэффициент детерминации представляет собой долю дисперсии отклика, объясненную с помощью оцениваемого уравнения регрессии. Чем меньше остаточная дисперсия по отношению к дисперсии значений отклика относительно его среднего выборочного, тем точнее модель $\hat{y} = \hat{\varphi}(x)$ отображает изменчивость Y в связи с изменениями фактора X, тем больше значение R^2. Идеальная модель ($\sum_{i=1}^{n} e_i = 0$) обеспечивает $R^2 = 1$ и в этом случае необходимость в оценивании адекватности аппроксимации отпадает. В остальных случаях следует проверить условие отличия коэффициента детерминации от нуля:

$$F = \frac{R^2}{1-R^2}\,\frac{n-k}{k-1} > F\left[\alpha; k-1; n-k\right]. \tag{2.10}$$

Если при заданной доверительной вероятности указанное условие выполняется, можно считать, что R^2 значимо отличается от нуля и, следовательно, оцениваемое уравнение регрессии является достаточно точной аппроксимацией исследуемой зависимости. В противном случае необходимо признать, что аппроксимация не является адекватной и рассмотреть иной вариант уравнения регрессии.

3.ВЫПОЛНЕНИЕ РАБОТЫ

Работа включает выполнение парного регрессионного анализа в *MS Excel* двумя способами – с использованием статистических функций и с применением инструмента «Линия тренда». Порядок выполнения работы иллюстрируется на примере построения зависимости предела текучести стали от степени деформации при холодном пластическом деформировании.

В практике ОМД для отображения зависимости предела текучести от степени деформации при холодном пластическом деформировании наибольшее распространение получили выражения следующего вида [6, 7]:

$$\sigma_{\text{Т}} = \sigma_{\text{Т0}} + b_0\,\varepsilon^{b_1}; \tag{3.1}$$

$$\sigma_{\text{Т}} = \sigma_{\text{Т0}} + b_1\sqrt{\varepsilon} + b_0, \tag{3.2}$$

где $\sigma_{\text{Т0}}$ - предел текучести металла в отожженном (недеформированном) состоянии.

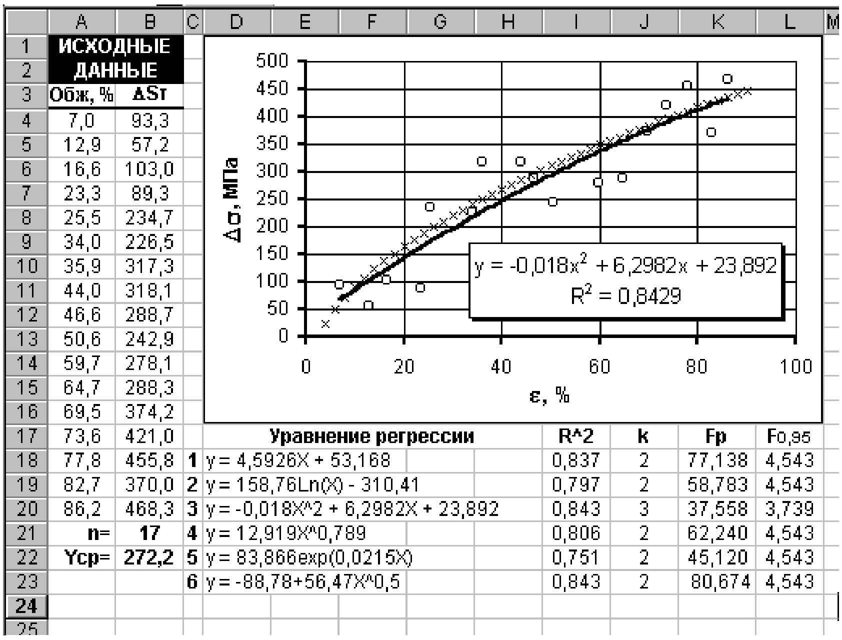
Оба выражения можно представить в иной форме:

$$\sigma_{\text{Т}} = \sigma_{\text{Т0}} + \Delta\sigma_{\text{Т}} \tag{3.3}$$

Тогда задача построения зависимости для расчета предела текучести сводится к задаче определения выражения, отображающего влияние степени деформации ε на приращение предела текучести $\Delta\sigma_{\text{Т}}$.

3.1. Построение аппроксимаций с применением инструмента «Линия тренда»

Сначала на рабочий лист (рис. 2) с клавиатуры заносятся исходные данные. На рис. 2 они расположены в ячейках А3:В20.



Обж, %	ΔSт		Уравнение регрессии	R^2	k	Fp	F0,95
7,0	93,3						
12,9	57,2						
16,6	103,0						
23,3	89,3						
25,5	234,7						
34,0	226,5						
35,9	317,3						
44,0	318,1						
46,6	288,7						
50,6	242,9						
59,7	278,1						
64,7	288,3						
69,5	374,2						
73,6	421,0						
77,8	455,8	1	y = 4,5926X + 53,168	0,837	2	77,138	4,543
82,7	370,0	2	y = 158,76Ln(X) - 310,41	0,797	2	58,783	4,543
86,2	468,3	3	y = -0,018X^2 + 6,2982X + 23,892	0,843	3	37,558	3,739
n=	17	4	y = 12,919X^0,789	0,806	2	62,240	4,543
Ycp=	272,2	5	y = 83,866exp(0,0215X)	0,751	2	45,120	4,543
		6	y = -88,78+56,47X^0,5	0,843	2	80,674	4,543

Рис.2. Фрагмент рабочего листа с исходными данными и результатами парного регрессионного анализа

В ячейке В21 определяется число наблюдений. При этом следует применить статистическую функцию СЧЕТ():

=СЧЕТ(В4:В20).

В ячейке В22 необходимо определить среднее выборочное значение отклика $\Delta\sigma_T$:

=СРЗНАЧ(В4:В22).

Далее, с использованием инструмента «Мастер диаграмм», исходные данные необходимо отобразить в виде диаграммы типа «Точечная». Диаграмма должна быть отформатирована таким образом, чтобы ее оформление соответствовало правилам графического представления данных.

После форматирования диаграммы, с помощью мыши выделить на диаграмме ряд данных. Через пункт <Диаграмма> в строке главного меню *MS Excel* , задать команду <Добавить линию тренда>. На экране появится диалоговое окно <*Линия тренда*> (вид окна и опции работы с линиями тренда представлены в прил. 2).

На закладке «Параметры» включить опции «Показывать уравнение на диаграмме» и «Поместить на диаграмму величину достоверности аппроксимации (R^2)». Затем на закладке «Тип» выбрать тип аппроксимации «Линейная». После нажатия кнопки <ОК> на диаграмме появятся линия тренда и надпись, содержащая уравнение регрессии и значение R^2. Автоматически фон надписи устанавливается прозрачным, что затрудняет чтение текста. Поэтому необходимо придать ей атрибут «С тенью» через закладку «Вид» окна <*Формат подписи данных*>, которое появиться, если выполнять форматирование надписи через пункт <**Формат**> главного меню.

Под диаграммой необходимо подготовить обозначения столбцов, в которых будет размещаться информация о рассмотренных уравнениях регресси. На рис. 2 такие обозначения записаны в ячейках D17:L17.

Уравнение регрессии и значение R^2 записываются с клавиатуры на основании надписи, которой сопровождается линия тренда. Число коэффициентов регрессии k также вводится с клавиатуры.

Рассчитанное число Фишера F_p вычисляется по формуле (2.10). Например, в ячейке K18:

=I18*($B$21-J18)/((1-I18)*(J18-1)).

Табличное число Фишера $F_{0,95}$ находится для доверительной вероятности 95% с использованием статистической функции FРАСПОБР(). Например, в ячейке L18:

=FРАСПОБР(0,05;J18-1;$B$21-J18).

Чтобы рассмотреть другие аппроксимации, доступные через инструмент «Линия тренда», необходимо выполнять форматирование линии тренда.

Форматирование может быть реализовано либо через контекстное меню, либо через пункт <Формат> главного меню. В этом случае на экране также появляется окно <*Линия тренда*>, но теперь достаточно указать другой тип аппроксимации на закладке «Тип». В результате на диаграмме вместо предыдущего тренда появиться новый, а его параметры отобразятся в надписи. **Внимание!** Если каждый из вариантов аппроксимации наносить на диаграмму как добавление линии тренда, то на диаграмме будут одновременно отображены все рассмотренные варианты.

В примере выполнения работы (рис. 2) аппроксимации, полученные с применением инструмента «Линия тренда», обозначены номерами с 1 по 5.

3.2. Построение аппроксимации с применением статистических функций

Строится аппроксимация вида $\Delta\sigma_т = b_0 + b_1\sqrt{\varepsilon}$. Исходными данными являются те же, которые подвергались анализу с использованием инструмента «Линия тренда». Пример оформления фрагмента рабочего листа представлен на рис. 3. Сначала необходимо выполнить линеаризацию исходных данных по формулам, приведенным в прил. 1. На рис. 3 результаты линеаризации расположены в ячейках N4:O20. Они получены путем преобразований $y_i' = y_i$ и $x_i' = x_i^m$ для m =0,5. Например, в ячейках N4 и O4 соответственно запрограммировано:

=B4 и =A4^$Q$3.

Затем следует определить коэффициенты линейной регрессии $y_i' = b_0' + b_1'x_i'$. В примере (рис. 3) значения b_0' и b_1' представлены в ячейках P5 и Q5 соответственно. Для их вычисления здесь использованы статистические функции:

=ОТРЕЗОК(N4:N20;O4:O20) и =НАКЛОН(N4:N20;O4:O20).

Далее необходимо осуществить переход к действительным коэффициентам регрессии. В соответствии с прил. 1 для рассматриваемой аппроксимации переход осуществляется с помощью преобразований $b_0 = b_0'$ и $b_1 = b_1'$. В примере действительные значения коэффициентов регрессии представлены в ячейках P7 и Q7. Здесь соответственно запрограммировано:

=P4 и =Q4.

После этого по уравнению, вид и коэффициенты которого известны, следует вычислить оценки отклика. Для рассматри-

	N	O	P	Q	R	S	T	V	W
1	АППРОКСИМАЦИЯ							ТОЧКИ	
2	Y=b0+b1*X^m							ЛИНИИ РЕГРЕССИИ	
3	**Y'=Y**	**X'=X^m**	**m=**	**0,5**	**ΔSт~**	**ei**	**Ei**	**Обж.%**	**ΔSт~, МПа**
4	93,3	2,24	**b0'**	**b1'**	37,5	55,8	-178,9	4	24,2
5	57,2	3,16	-88,78	56,47	89,8	-32,6	-214,9	6	49,5
6	103,0	3,87	**b0**	**b1**	129,9	-26,9	-169,2	8	70,9
7	89,3	4,47	-88,8	56,47	163,8	-74,5	-182,9	10	89,8
8	234,7	5,00			193,6	41,1	-37,5	12	106,8
9	226,5	5,48			220,5	6,0	-45,7	14	122,5
10	317,3	5,92			245,3	72,0	45,1	16	137,1
11	318,1	6,32			268,4	49,7	45,9	18	150,8
12	288,7	6,71			290,0	-1,3	16,6	20	163,8
13	242,9	7,07			310,5	-67,6	-29,3	22	176,1
14	278,1	7,42			330,0	-51,9	5,9	24	187,9
15	288,3	7,75			348,6	-60,3	16,2	26	199,2
16	374,2	8,06			366,5	7,7	102,0	28	210,0
17	421,0	8,37			383,7	37,3	148,9	30	220,5
18	455,8	8,66			400,3	55,6	183,7	32	230,7
19	370,0	8,94			416,3	-46,4	97,8	34	240,5
20	468,3	9,22			431,9	36,5	196,2	36	250,0
21						**38754**	**263662**	38	259,3
22								40	268,4
23								42	277,2
24								44	285,8
25								46	[illegible]
45								86	434,9
46								88	441,0
47								90	447,0
48									

Рис. 3. Фрагмент рабочего листа с результатами построения аппроксимации $\Delta\sigma_{\text{Т}} = b_0 + b_1\sqrt{\varepsilon}$ и расчетными данными для ее графической интерпретации

ваемого примера (рис. 3) оценки приращения предела текучести рассчитаны в ячейках R4:R20. Например, в ячейке R4:

=$P$7+$Q$7*A4^$Q$3.

Теперь необходимо вычислить отклонения $e_i = y_i - \hat{y}_i$ и $E_i = y_i - \bar{y}$ для всех n наблюдений, а также суммы квадратов отклонений $\sum_{i=1}^{n} e_i^2$ и $\sum_{i=1}^{n} E_i^2$. Для рассматриваемого примера квадраты отклонения рассчитаны в ячейках S4:S20 и T4:T20 соответственно. Например, в ячейках S4 и T4 записано:

=B4-S4 и =B4-$B$22.

Суммы квадратов отклонений вычисляются в ячейках S21 и T21 с использованием функции КВАДРАТОТКЛ ():

=КВАДРОТКЛ(S4:S20) и =КВАДРОТКЛ(T4:T20) .

Среди вариантов уравнений регрессии (рис. 2) рассматриваемая аппроксимация обозначена номером 6. Соответствующее ей значение R^2 рассчитано в ячейке I23 по формуле (2.9). При этом используются ссылки на суммы квадратов отклонений, вычисленные в ячейках S21 и T21:

=1-S21*($B$21-1)/(T21*($B$21-J23)) .

Значения рассчитанного и табличного чисел Фишера определяются также, как и в случае построения аппроксимаций с использованием инструмента «Линия тренда».

4.СОДЕРЖАНИЕ ВЫВОДОВ ПО РАБОТЕ

В выводах по работе необходимо ответить на следующие вопросы:

1. Связь между какими величинами анализировалась?
2. Сколько и какие аппроксимации были рассмотрены, какие из них являются статистически значимыми?
3. Какая из аппроксимаций является наилучшим отображением связи между анализируемыми параметрами и почему можно это утверждать?

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. Львовский Е.Н, Статистические методы построения эмпирических формул: Учебн. пособие для втузов, - 2-е изд., перераб. и доп.- М., Высш. шк . , 1988. - 239 с.
2. Математическая статистика: Учебник / Иванова В. М., Калинина В. Н., Нешумова Л. А. и др. – М., Высшая школа, 1981. – 371 с.
3. Четыркин Е. М., Калихман И. Л. Вероятность и статистика. – М., Финансы и статистика, 1982. – 320 с.
4. Ахназарова С.Л,, Кафаров В.В. Организация эксперимента в химии и химической технологии: Учебн. пособие для химико-технологических вузов.- М., Высшая школа , 1978, - 319 с.
5. Дьяконов В.П. Справочник по алгоритмам и программам на языке Бейсик для персональных ЭВМ: Справочник. - М., Наука. Гл. ред. физ.-мат. лит., 1987. - 240 с.

6. Третьяков А. В., Зюзин В. И. Механические свойства металлов и сплавов при обработке давлением. – М.: Металлургия, 1973. – 224 с.
7. Грудев А. П., Сигалов Ю. Б. Методика определения предела текучести металла при холодной прокатке с учетом влияния основных факторов деформации // Обработка металлов давлением: Науч. Тр. Днепропетровского металлургич. ин-та № 56. – М.: Металлургия, 1971. – С. 47-55.
8. Мелник М. Основы прикладной статистики: Пер. с англ. – М., Энергоатомиздат, 1983. – 413 с.

Содержание

ПРИЛОЖЕНИЕ 1

ФОРМУЛЫ ДЛЯ ВЫЧИСЛЕНИЯ КОЭФФИЦИЕНТОВ НЕКОТОРЫХ НЕЛИНЕЙНЫХ УРАВНЕНИЙ ПАРНОЙ РЕГРЕССИИ

Аппроксимация	Преобразование исходных данных		Преобразование коэффициентов	
	x_i'	y_i'	b_0	b_1
$b_0 + b_1 / x$	$1/x_i$	y_i	b_0'	b_1'
$b_0 + b_1 x^m$	x_i^m	y_i	b_0'	b_1'
$b_0 + b_1 \lg x$	$lg\, x_i$	y_i	b_0'	b_1'
$b_0 + b_1 \ln x$	$ln\, x_i$	y_i	b_0'	b_1'
$1/(b_0 + b_1 x)$	x_i	$1/y_i$	b_0'	b_1'
$x/(b_0 + b_1 x)$	x_i	x_i / y_i	b_0'	b_1'
$b_0 / (b_1 + x)$	x_i	$1/y_i$	$1/b_0'$	b_0' / b_1'
$b_0 x / (b_1 + x)$	$1/x_i$	$1/y_i$	$1/b_0'$	b_0' / b_1'
$b_0\, b_1{}^x$	x_i	$lg\, y_i$	$10\, b_0'$	$10\, b_1'$
$b_0\, x^{b_1}$	$ln\, x_i$	$ln\, y_i$	$\exp(b_0')$	b_1'
$b_0 \exp(b_1 x)$	x_i	$ln\, y_i$	$\exp(b_0')$	b_1'
$b_0 \exp(b_1 / x)$	$1/x_i$	$\ln y_i$	$\exp(b_0')$	b_1'

ПРИЛОЖЕНИЕ 2

ЛИНИЯ ТРЕНДА В *MS EXCEL* КАК ИНСТРУМЕНТ РЕГРЕССИОННОГО АНАЛИЗА

Понятие «Тренд» впервые стало применяться в экономическом анализе т. н. временных рядов. Временной ряд - упорядоченные во времени данные, а тренд представляет собой тенденцию этого ряда, обнаруживаемую при сглаживании краткосрочных колебаний. Графически тренд отображается линией, положение которой отыскивают по методу наименьших квадратов [8]. Таким образом, линия тренда, по существу, представляет собой линию регрессии.

В *MS Excel* предусмотрены следующие способы построения тренда [9]:

1. Использование маркера заполнения или команды «Прогрессия»;
2. Использование статистических функций рабочего листа ПРЕДСКАЗ(),ТЕНДЕНЦИЯ(),РОСТ(),ЛИНЕЙН(),ЛГРФПРИБЛ();
3. Добавление линии тренда к ряду данных на диаграмме.

Первыми двумя способами тренд может быть получен только в табличной форме. Третий способ наиболее пригоден для целей парного регрессионного анализа - он позволяет провести линию тренда на диаграмме, получить ее уравнение и оценку R^2.

Чтобы применить линию тренда для регрессионного анализа, сначала необходимо построить диаграмму. Затем необходимо выделить на диаграмме ряд данных и либо через контекстное меню, либо через пункт <Диаграмма> главного меню задать команду <Добавить линию тренда>.

На экране появится диалоговое окно, содержащее закладки «Тип» и «Параметры» (рис. П2).

Через закладку «Тип», щелчком мыши на соответствующем изображении, выбирается тип тренда (т. е. вариант линии регрессии). Различным типам тренда соответствуют следующие уравнения:

1. Линейный - $y = b_0 + b_1 x$.
2. Логарифмический - $y = b_0 + b_1 \, ln(x)$.
3. Полиномиальный(степень полинома не более 6) - $y = b_6 x^6 + ... + b_2 x^2 + b_1 x + b_0$.

4. Степенной - $y = b_0 \, x^{b_1}$.

5. Экспоненциальный - $y = b_0 \, exp\left(b_1\right)$.

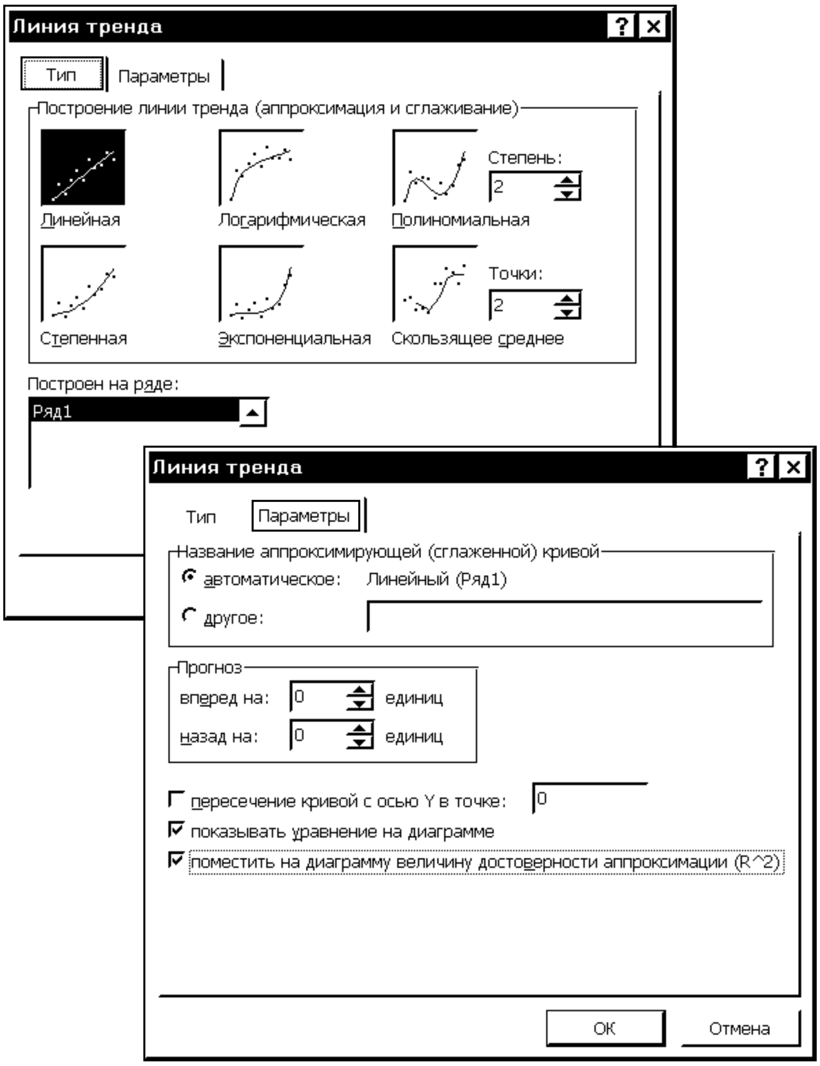


Рис. П2. Диалоговое окно «Линия тренда»

Тип тренда «Скользящее среднее» для целей регрессионного анализа не пригоден.

Через закладку «Параметры» задаются атрибуты, определяющие особенности представления линии тренда

Название аппроксимирующей (сглаженной) кривой — условное обозначение длиной до 256 символов, которое будет включено в легенду диаграммы. Если задана опция «Автоматическое», название линии тренда будет создано автоматически с указанием идентификатора ряда данных (например, «Ряд1») и типа тренда (например, «Линейный»). Чтобы задать оригинальное название необходимо выбрать опцию «Другое» и ввести его с клавиатуры.

Прогноз – указание о необходимости построения линии тренда за пределами варьирования независимой переменной (фактора) для анализируемого ряда данных. Атрибут «Вперед …» обеспечивает продолжение линии тренда в направлении возрастания независимой переменной, а «Назад …» - в направлении уменьшения.

Пересечение кривой с осью Y - значение зависимой переменной (отклика) y_0, которое соответствует точке пересечения оси ординат линией тренда. Задание атрибута фактически означает добавление к ряду анализируемых данных еще одной точки, для которой $y_i = y_0$ и $x_i = 0$. Это приведет и к модификации уравнения регрессии. Например, для линейного тренда вместо уравнения $y = b_0 + b_1 x$ будет получено $y = y_0 + c\,x$. Атрибут удобен для получения уравнения линии, которая должна проходить желаемым образом относительно начала координат. Однако при его использовании следует внимательно анализировать значения показателя R^2. В частности, может быть получено $R^2 < 0$, что не противоречит формуле (2.9). Но такой результат означает, что $\sum_{i=1}^{n} e_i^2$ > $\sum_{i=1}^{n} E_i^2$ и, следовательно, данную зависимость нельзя рассматривать как надежную аппроксимацию взаимосвязи между откликом и фактором.

Показать уравнение на диаграмме. Если опция задана, на диаграмму будет выведена надпись в виде уравнения регрессии.

Поместить на диаграмму величину достоверности аппроксимации. Если опция задана, надпись, содержащая уравнение регрессии, будет дополнена значением R^2.

Линии тренда могут быть построены для рядов данных, представленных на ненормированных плоских диаграммах с областями, линейчатых диаграммах, гистограммах, графиках, биржевых, точечных и пузырьковых диаграммах. Для регрессионного анализа рекомендуется применять диаграмму типа «Точечная».