

Дубинина Екатерина Викторовна

старший преподаватель, начальник отдела
учебно-методического обеспечения
ФГБОУ ВО «Российский государственный
социальный университет»
г. Москва

ЕЖЕДНЕВНАЯ МАГИЯ EXCEL. ВОССТАНОВЛЕНИЕ ПРОПУСКОВ В ДАННЫХ

Аннотация: в исследуемой статье рассматривается способ восстановления пропусков в данных с использованием аналитических моделей с применением MS Excel. Используется метод фиктивных переменных с построением регрессионной модели.

Ключевые слова: MS Excel, восстановление данных, регрессия.

Для управления любыми системами требуется мониторинг, сбор информации и оперативное наблюдение за характеристиками. Если процесс сбора данных не автоматизирован, то на этапе анализа могут возникать потеря и пропуски данных. В таких ситуациях требуется уточнение данных, однако если решение принимается в условиях ограниченного времени, то в качестве решения проблемы может быть использовано восстановление пропусков в данных с использованием аналитических моделей с применением MS Excel.

Ниже показано одно из возможных решений восстановления данных методом фиктивных переменных с построением регрессионной модели. В качестве примера будет использована статистика, отражающая динамику индекса РТС, а также стоимость акций Банка N.

Инструмент анализа «Регрессия» (рис. 1) применяется для подбора графика для набора наблюдений с помощью метода наименьших квадратов. Регрессия используется для анализа воздействия на отдельную зависимую переменную значений одной или нескольких независимых переменных [1].

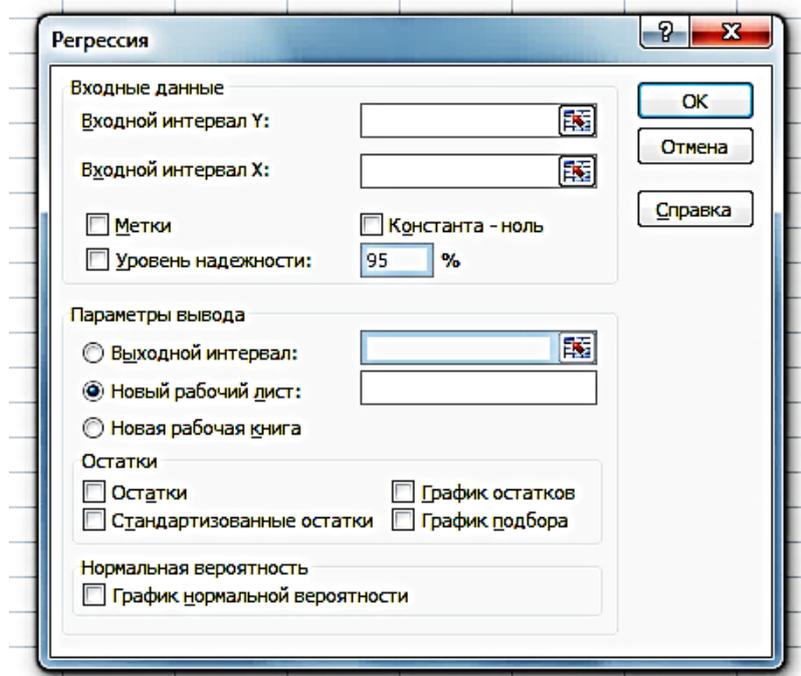


Рис. 1. Диалоговое окно Регрессия

Во *входной интервал Y* вводится ссылка на диапазон зависимых данных. Диапазон должен состоять из одного столбца.

Во *входной интервал X* вводится ссылка на диапазон независимых данных. Эти данные будут расположены слева направо в порядке возрастания. Максимальное число независимых переменных равно 16.

В графе *метки* устанавливается флажок, если первая строка или первый столбец входного диапазона содержит заголовки. Если флажок убрать, то подходящие заголовки для данных выходной таблицы будут созданы автоматически.

В графе *уровень надежности* устанавливается флажок, чтобы включить в выходную таблицу итогов дополнительный уровень. В соответствующее поле вводится уровень надежности, который следует применить, дополнительно к уровню 95%, применяемому по умолчанию.

В графе *константа – ноль* устанавливается флажок, чтобы линия регрессии прошла через начало координат.

В графе *остатки* устанавливается флажок для включения остатков в выходную таблицу.

В графе *стандартизированные остатки* устанавливается флажок для включения стандартизированных остатков в выходную таблицу.

При необходимости можно графически представить данные и вывести *график остатков* для каждой независимой переменной, *график зависимости* предсказанных значений от наблюдаемых, *график нормальной вероятности*.

Порядок работы в MS Excel.

1. Вначале в MS Excel вводятся исходные данные (табл. 1).
2. Затем вводятся фиктивные переменные (табл. 2).
3. Далее строится регрессионная модель (рис. 2).
4. Результаты регрессионного анализа вносятся в исходные данные (табл. 4).

В общем виде формула будет выглядеть как

$$y_t^{[1]} = b_0 + b_1 x_t + b_2 f_{t1}^{[1]} + b_3 f_{t2}^{[1]} + b_4 f_{t3}^{[1]} + b_5 f_{t4}^{[1]} + b_6 f_{t5}^{[1]} + \varepsilon_t^{[1]}$$

где, $y_t^{[1]}$ – стоимость акций Банка N;

x_t – величина индекса РТС;

b_0, b_1, \dots, b_6 – коэффициенты моделей;

$f_{t1}^{[1]}, \dots, f_{t5}^{[1]}$ – фиктивные переменные моделей;

$\varepsilon_t^{[1]}$ – ненаблюдаемые случайные составляющие моделей.

Таблица 1

Исходные данные для расчета в MS Excel

	А	В	С
1.	Дата	Стоимость акций Банка	Индекс РТС
2.	01.12.2016	20,8	685,51
3.	02.12.2016	21,01	733,93
4.	05.12.2016	<i>Нет данных</i>	746,03
5.	06.12.2016	24	748,62
6.	07.12.2016	22,69	740,47
7.	08.12.2016	24,05	760,58

8.	09.12.2016	27	810,9
9.	12.12.2016	28	817,41
10.	13.12.2016	30	814,67
11.	14.12.2016	<i>Нет данных</i>	807,61
12.	15.12.2016	<i>Нет данных</i>	805,85
13.

Таблица 2

Структура данных с фиктивными переменными

	A	B	C	D	E	F
	Дата	Стоимость акций Банка	Индекс РТС	Фиктивные переменные		
1.	01.12.2016	20,8	685,51	0	0	0
2.	02.12.2016	21,01	733,93	0	0	0
3.	05.12.2016	0	746,03	-1	0	0
4.	06.12.2016	24	748,62	0	0	0
5.	07.12.2016	22,69	740,47	0	0	0
6.	08.12.2016	24,05	760,58	0	0	0
7.	09.12.2016	27	810,9	0	0	0
8.	12.12.2016	28	817,41	0	0	0
9.	13.12.2016	30	814,67	0	0	0
10.	14.12.2016	0	807,61	0	-1	0
11.	15.12.2016	0	805,85	0	0	-1

Для построения моделей необходимо вызвать модуль Регрессия (Данные – Анализ – Анализ данных – Регрессия) и в открывшемся окне указать входные и выходные интервалы, где:

– во входной интервал Y вводится ссылка на диапазон (B2 : B12) в котором расположены сведения о стоимости акций компании в которых необходимо найти пропущенные сведения;

– во входной интервал X вводится ссылка на диапазон (C2 : F12) независимых данных (индекс РТС и введенные фиктивные переменные).

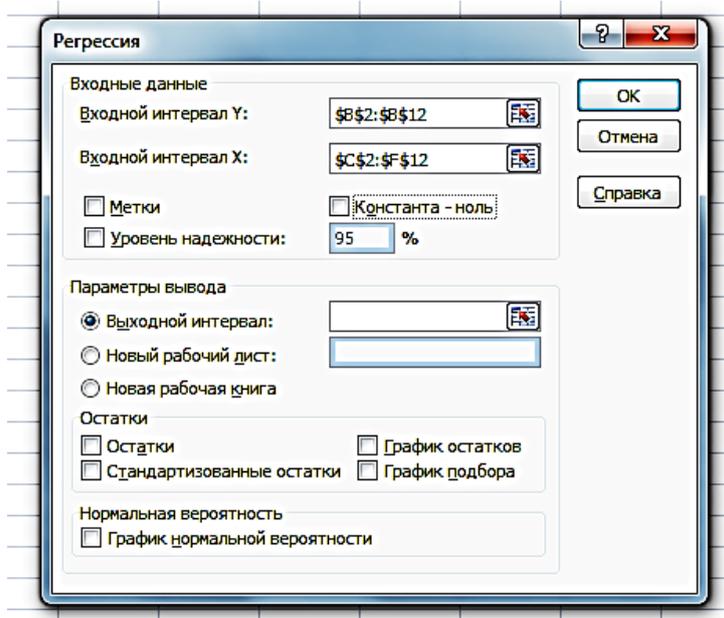


Рис. 2. Модуль Регрессия (Банк N)

Таблица 3

Результаты регрессионного анализа

<i>Вывод итогов</i>						
<i>Регрессионная статистика</i>						
Множественный R	0,979101					
R-квадрат	0,958639					
Нормированный R-квадрат	0,957759					
Стандартная ошибка	3,660986					
Наблюдения	193					
<i>Дисперсионный анализ</i>						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>	
Регрессия	4	58400,34	14600,08	1089,33	8,3E-129	
Остаток	188	2519,73	13,40282			
Итого	192	60920,06				
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
Y-пересечение	-32,4315	1,402475	-23,1245	7,17E-57	-35,1981	-29,6649
Переменная X 1	0,07409	0,001208	61,32688	3E-126	0,071707	0,076473
Переменная X 2	22,84204	3,701317	6,171327	4,07E-09	15,54059	30,14348
Переменная X 3	27,40451	3,69249	7,421688	3,88E-12	20,12047	34,68855
Переменная X 4	27,27411	3,692722	7,38591	4,79E-12	19,98962	34,55861

Восстановленные данные в MS Excel

	А	В	С
1.	Дата	Стоимость акций Банка	Индекс РТС
2.	01.12.2016	20,8	685,51
3.	02.12.2016	21,01	733,93
4.	05.12.2016	22,84	746,03
5.	06.12.2016	24	748,62
6.	07.12.2016	22,69	740,47
7.	08.12.2016	24,05	760,58
8.	09.12.2016	27	810,9
9.	12.12.2016	28	817,41
10.	13.12.2016	30	814,67
11.	14.12.2016	27,4	807,61
12.	15.12.2016	27,27	805,85

Список литературы

1. Использование пакета анализа [Электронный ресурс]. – Режим доступа: <https://support.office.com> (04.12.2016).